



HAL
open science

Qu'est-ce qu'une belle photo ? Essai sur l'esthétique en photographie numérique

Henri Maître

► **To cite this version:**

Henri Maître. Qu'est-ce qu'une belle photo ? Essai sur l'esthétique en photographie numérique. 2018.
hal-01864135

HAL Id: hal-01864135

<https://imt.hal.science/hal-01864135v1>

Preprint submitted on 29 Aug 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Qu'est-ce qu'une belle photo ?

Essai sur l'esthétique en photographie numérique

Henri Maître

Professeur émérite, LTCI - Télécom Paristech, 46 rue Barrault 75013 Paris

henri.maitre@telecom-paristech.fr

version du : 29 août 2018

RÉSUMÉ

Si la notion de qualité d'image est un sujet d'étude des traiteurs d'images depuis plus de quarante ans, les notions liées à l'esthétique des photographies ne sont apparues que depuis une dizaine d'années dans cette communauté. Mais avec l'automatisation des fonctions d'archivage et de recherche dans les très vastes bases de données numériques (big data), les études qui sont consacrées à ce thème se multiplient aujourd'hui, mettant à profit le développement des techniques d'apprentissage (machine learning) ainsi que la prolifération des sites spécialisés dans l'archivage des photos sur internet. Se pose alors fréquemment le problème de choisir parmi de nombreuses photos et le critère esthétique est alors fréquemment invoqué pour justifier un choix. Est-ce que ce critère est pertinent ? Est-il judicieusement mis en algorithme par les techniques automatiques ?

Quel espoir peut-on mettre dans des solutions qui ne font appel que très marginalement au jugement humain ? Quel profit tirent-elles de vingt siècles de littérature sur l'Esthétique, le Beau et l'Art tant en philosophie qu'en sociologie et en psychologie expérimentale ? Comment se relient-elles aux travaux menés parallèlement en physiologie de la perception et en neuro-biologie qui tentent d'expliquer le fonctionnement du cerveau et dont certains se sont dégagés vers la compréhension de notre jugement esthétique, se regroupant sous le nom de « neuro-esthétique » ?

On fait ici une présentation des diverses approches de l'esthétique des images telles qu'elles ont pu être proposées au cours des 20 derniers siècles, en dégageant les critères les plus opérationnels qui peuvent se prêter à une analyse automatique. Après ce bilan dans le champ de la philosophie, de la neuro-biologie et la psychologie de la perception, on aborde les méthodes qui cherchent à se substituer à l'expert humain. On examine en particulier les deux grandes voies qui ont conduit à des algorithmes : l'approche algébrique, et en particulier sa formulation par Birkhoff, s'appuyant sur une conception objectiviste de la beauté, et l'approche par apprentissage (machine learning), d'inspiration nettement plus subjectiviste. On insiste sur la seconde qui focalise aujourd'hui les attentions et on fait un bilan sur ses résultats, alors même que les travaux utilisant l'outil puissant des réseaux de neurones profonds n'ont encore fourni qu'un petit nombre de conclusions. On discute le choix de techniques à base de détection de paramètres et de classifieurs vs. les techniques à base de réseaux neuro-mimétiques profonds. On dresse un tableau critique des divers résultats aujourd'hui disponibles. On questionne la pertinence des bases de données utilisées ainsi que des « expertises » servant à l'apprentissage. On examine les pistes pour le futur et on donne quelques recommandations à ceux qui s'engagent dans cette voie.

ABSTRACT

Image quality has been introduced in image processing community studies at least 40 years ago, on the contrary to notions linked with aesthetical qualities of photographs which only appeared in the computer science literature in the last decade. However, because of the evolution of Information and Communication Technologies and its consequences on business and societal practices (storage on very large and distributed data bases, archival and retrieval functions, automatic learning processes) and because of the flourishing of photographic exchange sites, the domain received a growing attention in the last decade. Digital aesthetic image assessment is now rich of some tens of papers and "Image Beauty" is more and more chosen as a major criterion to select images in retrieval operations. Is this criterion mimicking human judgement relevant ? Is it correctly translated into algorithms ?

How reliable may be approaches which are only moderately supported by human expertise ? Which benefit is hired from 20 centuries of literature on aesthetics, beauty and art, in Philosophy, in Sociology and in experimental Psychology as well ? Which relations are weaved with the ongoing studies in Physiology and in Neuro-Biology that intend to analyze how our brain is working, some of them directed towards explaining our aesthetical judgement, now recognized under the name of "neuro-aesthetics" ?

This text proposes a review of the several ways to come up to image aesthetics, as proposed in the last 20 centuries. It tries to focus on the most operational criteria which may lead to automatic

implementation. This review being done in the field of philosophy, neuro-biology and psychology of perception, methods are presented which have been proposed to oust human expert. We separate the two main trends : from one side the algebraic way, following Birkhoff's formulation, based on an objectivist concept of beauty, and, on the other side the machine learning approach, much more subjectivist in its inspiration. We pay a greater attention to the later, which retains today most of the computer science community efforts. We present the major works on the subject along with their performances and we discuss their premisses, aware that the most efficient techniques, based on deep neural networks are still under development and far from having deployed their full potential. Methods based on handcrafted feature detection and classifiers are discussed versus end-to-end multi-layer neural nets. We present a critical landscape of the available results today. Relevance of the used data bases is questioned, as well as the associated expertise. Possible tracks for the future are examined and some recommendations are given to those who enter the domain.

Table des matières

1	Introduction	9
1.1	Pourquoi regardons-nous une image ?	9
1.1.1	Les mécanismes liés à l'intérêt	10
1.1.2	Les mécanismes liés à la surprise	10
1.2	Art, Beauté, Esthétique ... quel rapport ?	12
1.3	Notre projet	13
2	L'héritage des philosophes	17
2.1	L'approche objectiviste	18
2.1.1	Les esthétiques des Objectivistes	19
2.2	L'approche subjectiviste	23
2.3	Vers une synthèse ?	25
2.3.1	Un débat toujours vivant	25
2.3.2	L'esthétique interactionniste	26
2.3.3	La mise en place d'une esthétique interactionniste	27
2.3.4	La diversité des esthétiques	28
3	Neuro-esthétique	31
3.1	Les protocoles d'IRMf et la neuro-esthétique	31
3.2	IRMf et processeur de beauté	32
3.2.1	Le rôle du cortex préfrontal	32
3.2.2	Le rôle du cortex insulaire	33
3.2.3	Le rôle des aires visuelles	35
3.2.4	Le rôle de la mémoire et de la cognition	36
3.2.5	Le rôle de l' <i>embodiment</i>	36
3.2.6	Quel schéma d'ensemble ?	37
3.3	L'approche neuro-esthétique critiquée	38
4	Les règles de la photo	39
4.1	Le format ou rapport d'aspect	40

4.1.1	Que disent les statistiques?	41
4.1.2	Le nombre d'or	41
4.2	La composition	41
4.2.1	La règle du tiers (<i>Rule Of Third</i> , ROT)	42
4.2.2	Le centre de l'image	43
4.2.3	Quelles autres règles de composition?	44
4.3	Mise au point et spectre de densité de puissance	45
4.4	Niveaux de gris et histogramme	46
4.5	La couleur	46
4.5.1	Les préférences des couleurs isolées	47
4.5.2	Les préférences concernant les palettes colorées	48
4.6	Que dit la psycho-sociologie?	52
4.6.1	Mesurer la beauté	53
4.6.2	Images de la nature ou scènes artificielles?	54
4.6.3	Le rôle du contexte	55
4.6.4	La perception globale : unité et prototypalité	56
5	Approches algorithmiques	57
5.1	Les premiers essais : Ch. Henry	57
5.2	L'approche mathématique de G. D. Birkhoff	57
5.3	A la suite de G. Birkhoff	59
5.3.1	Les <i>designers</i>	59
5.3.2	P. Machado et A. Cardoso	59
5.3.3	Les travaux de J. Rigau <i>et al.</i>	59
5.4	Une approche algorithmique d'IA	60
6	Le contexte de l'IA	63
6.1	L'IA, pourquoi faire?	63
6.1.1	Les résultats attendus	64
6.1.2	Les expertises	64
6.2	Les bases de données	65
7	<i>Hand-crafted features</i>	71
7.1	Primitives multimedia : le système Acquine et les travaux de R. Datta <i>et al.</i>	72
7.2	Contours et distance chromatique : l'approche de Y. Ke <i>et al.</i>	73
7.3	Objet et fond	74
7.4	Des primitives de haut niveau : l'approche de S. Dhar <i>et al.</i>	75
7.5	La bibliothèque de primitives esthétiques de H.-H. Su <i>et al.</i>	76

7.6	Le système OSCAR de L. Yao <i>et al.</i>	77
7.7	Un système embarqué : l'approche de K.Y. Lo <i>et al.</i>	78
7.8	Descripteurs génériques de la vision : l'approche de L. Marchesotti <i>et al.</i>	79
7.9	Harmonie chromatique : les travaux de P. Lu <i>et al.</i>	79
7.10	Réseaux sociaux et <i>crowdsourcing</i> : l'approche de R. Schifanella <i>et al.</i>	81
7.11	Prise en compte de commentaires : les travaux de J. San Pedro <i>et al.</i>	81
8	Réseaux de neurones profonds	83
8.1	Les RNP dédiés à l'évaluation esthétique	84
8.1.1	Le système RAPID de X. Lu <i>et al.</i>	84
8.1.2	L'architecture multi-voies DMA-Net de X. Lu <i>et al.</i>	86
8.1.3	S'adapter à la taille de l'image : les travaux de L. Mai <i>et al.</i>	87
8.1.4	La beauté du web : travaux de M. Redi <i>et al.</i>	87
8.1.5	Le réseau siamois de S. Kong <i>et al.</i>	88
8.1.6	Utiliser l'information sémantique : les travaux de Kao <i>et al.</i>	89
8.2	Variantes autour de l'architecture de base des RNP	90
8.2.1	Une approche qui distingue la nature de la photo : les travaux de Y. Kao <i>et al.</i>	91
8.2.2	Comparer les photos entre-elles : la proposition de K. Schwarz <i>et al.</i>	91
8.2.3	A mi chemin entre classification et RNP : la méthode de Z. Wang <i>et al.</i>	92
8.2.4	Porter attention à la construction de l'image : le système A-Lamp-CNN	93
8.2.5	Exploiter la distribution des évaluations	94
8.2.6	Formuler une appréciation littérale	95
8.2.7	Personnaliser l'évaluation	96
9	Analyse critique de l'apprentissage	99
9.1	Popularité des études sur l'esthétique	99
9.2	Bilan des méthodes d'apprentissage	99
9.2.1	Quelles architectures?	100
9.2.2	Quelles performances?	101
9.2.3	S'interroger sur les hypothèses	103
9.2.4	De quelle image une belle image est-elle l'image?	104
9.3	Comment introduire de la "subjectivité" dans les RNP?	105
9.3.1	Multiplier les avis	105
9.3.2	Les systèmes de recommandation	106
10	En forme de conclusion	109
	Bibliographie	111

Chapitre 1

Introduction

1.1 Pourquoi regardons-nous une image ?

La prolifération des images dans le monde moderne nous place de plus en plus devant l'obligation de sélectionner des photos dans un vaste ensemble. Cette tâche était réservée aux professionnels de l'édition, de la communication, de l'archivage, aux photographes professionnels, aux collectionneurs, pour choisir les illustrations d'un article, la couverture d'un ouvrage, une affiche, pour résumer un évènement, un voyage. Elle incombe maintenant à chacun de nous, pour décider de ce qu'il conserve ou qu'il oublie, de ce qu'il envoie, de ce qu'il poste sur internet, de ce qu'il détruit irréversiblement ou de ce qu'il archive pour une improbable revisite. La gestion de ses archives photographiques est un casse-tête pour beaucoup d'entre nous¹. L'étape de sélection a toujours été réputée importante et délicate. En milieu professionnel, elle est confiée à des personnes distinguées soit pour leur expertise, soit pour leur autorité.

Arrêtons nous un instant pour nous interroger sur les mécanismes qui sont activés lors de ces décisions de tri et de sélection d'images et voyons la place qui peut être faite alors à la beauté de l'image.

Il nous apparaît que l'on peut regrouper les mécanismes de préférence lors de la sélection de photographies selon trois grandes classes :

- l'intérêt du document, c'est-à-dire sa capacité à mobiliser nos capacités d'attention en attachant le document à nos domaines familiers,
- la surprise, c'est-à-dire, au contraire du point précédent, sa capacité à renouveler notre expérience visuelle par une contribution inattendue,
- enfin la beauté, c'est-à-dire l'agrément qu'il nous procure, indépendamment de son contenu, par la disposition de ses composantes.

C'est ce dernier aspect qui nous intéresse dans ce texte, c'est lui que nous examinerons exclusivement. Dans ce cadre, les deux premiers relèvent des *impressions associatives* au sens de G.T. Fechner qui a soulevé cette distinction dès le *XIX^e* siècle [Fechner, 1871]. Il arrive fréquemment qu'un même document associe plusieurs de ces registres, renforçant l'attrait du document, mais rendant moins clair l'apport de chacun des registres d'attention. Les contributions d'intérêt, de surprise et de beauté semblent cependant évoluer, pour notre décision, dans des espaces indépendants ou orthogonaux, au sens de P. Gärdenfors [Gärdenfors, 2000], c'est-à-dire sans intime influence de l'un sur l'autre. Ils sont donc évalués séparément par notre conscience, puis probablement combinés en un score unique qui nous fait finalement préférer telle image à telle autre dans un choix heuristique difficilement exprimable mais qui pourrait suivre par exemple les schémas de décision empirique proposés par

1. Dans le domaine de la vision par ordinateur, divers travaux se sont intéressés aux critères qui nous font retenir une image que l'on appelle *interestingness* ou *memorability*; on lira par exemple [Isola et al., 2011, Gygli et al., 2013]). Ces articles qui privilégient pas l'appréciation esthétique ne seront pas suivis dans ce texte.

A. Tversky et D. Kahneman [Tversky and Kahneman, 1981]. Examinons brièvement, avant de les abandonner, les deux premiers domaines.

1.1.1 Les mécanismes liés à l'intérêt

Ils ont été particulièrement étudiés depuis les années 1920 par les psychologues expérimentaux qui ont produit des travaux souvent regroupés aujourd'hui sous le nom de "théorie de la pertinence" (*relevance theory*) [Sperber and Wilson, 2004].

L'intérêt² est une émotion, c'est-à-dire une expérience psycho-physiologique complexe qui peut résulter :

- de stimulus extérieurs (environnementaux se transmettant par nos sens) ou intérieurs (biochimiques, créés par notre cortex), elle est alors passive ;
- ou de raisonnements, c'est-à-dire de processus cognitifs qui permettent de déduire de nouveaux éléments de connaissance à partir d'un contexte particulier ; elle est alors active.

Les travaux de W. James [Lange and James, 1922] ont ouvert les études sur les relations entre émotion et évaluation (*appraisal*) d'une situation par l'homme. L'importance de l'incitation (*arousal*) et l'ordre des diverses étapes (incitation, émotion, évaluation) ont ainsi été bien documentés. Ces travaux ont servi de guide à de nombreuses études conduites un demi-siècle plus tard sur l'esthétique.

Mais c'est surtout D. Sperber et D. Wilson [Sperber and Wilson, 2004] qui ont permis de construire une théorie de la pertinence mobilisant l'ensemble du bagage cognitif du récepteur et non plus seulement ses connaissances linguistiques comme le suggérait W. James. Dans ce cadre, J.L. Dessalles [Dessalles, 2008] a proposé une mesure plus quantitative de la pertinence et l'a dotée de capacités prédictives par une modélisation mathématique originale.

Lorsqu'on cherche à appliquer cette théorie de la pertinence à l'intérêt éveillé par une collection d'images, il est utile de distinguer deux cadres extrêmes qui seront peut-être traités de façon différente.

L'intérêt universel fait référence aux thèmes particulièrement exposés dans la société et les médias et transmis par la culture : tel acteur ou tel sportif, telle voiture, monument ou événement explicitement et régulièrement couvert. Leur popularité pourrait se mesurer aujourd'hui avec les outils de médiamétrie [Dessalles, 2013] : fréquence d'exposition à la télévision, nombre d'instances sur internet, de " clics " ou de " likes " sur les réseaux sociaux, etc. (figure 1.1 à gauche).

L'intérêt personnel il rapproche des images en raison des liens étroits avec les thèmes de la vie individuelle et plus particulièrement personnelle : "ma" famille, "mon" travail, "ma" ville . . . La mesure de la pertinence prend alors les formes plus classiques proposées par exemple dans [Dessalles, 2013] mettant en jeu des variables d'espace et de temps à décroissance plus ou moins rapide et des relations d'ordre de proximité (par exemple dans un arbre de famille ou d'entreprise) (figure 1.1 à droite).

1.1.2 Les mécanismes liés à la surprise

Ces mécanismes fonctionnent un peu à l'inverse des précédents car ils sont d'autant plus importants que le document qui nous est présenté est plus éloigné de ce que nous connaissons (figure 1.2). La surprise³ peut se décliner selon des formes variées : l'humour, l'effroi, la perplexité, la tendresse, le dégoût . . .

On voit bien que la pertinence d'une part, la surprise d'autre part, sont des motivations différentes des qualités esthétiques pour retenir une photographie. Qu'elles opèrent conjointement pour expliquer

2. Les textes anglo-saxons utilisent les mots *relevance* ou *interestingness* à ce propos.

3. Le terme anglais consacré pour la surprise est *amazement*. Il couvre un champ plus large que le mot français et c'est bien toute l'étendue qu'il couvre que l'on souhaite représenter ici.



Figure 1.1 – La pertinence d’une image peut avoir pour référence le cadre général des connaissances d’une communauté (à gauche : des acteurs, des sportifs, des objets fréquemment présentés par exemple dans les journaux ou sur les réseaux sociaux), ou au contraire un cadre personnel lié à l’observateur (à droite : les membres d’une même famille, les lieux de vacances, les passe-temps, etc.).



Figure 1.2 – La surprise est un autre mécanisme qui nous amène à porter une attention particulière à une image. On s’attache alors à bien identifier les éléments particuliers de l’image qui sortent des schémas connus de notre représentation de l’univers.

un choix est aussi très probable⁴, comme nous le constatons sur les deux photos de la figure 1.3. Nous nous efforcerons cependant par la suite de les séparer, c’est-à-dire que nous essaierons d’évaluer la qualité esthétique à pertinence et surprise égales. Ce ne sera pas facile et il faudra retenir que dans de nombreuses circonstances les avis donnés lors de tests subjectifs n’auront pu échapper à une certaine confusion sur ce point. Cela sera particulièrement vrai lorsque l’on sollicitera des avis de personnes inconnues et distantes comme on le fait par exemple par des mesures sur le web.

4. Il est en particulier très intéressant de voir comment les rétrospectives régulièrement faites par les journaux professionnels sous des titres comme : “Les 100 plus belles photos de l’année” mêlent très largement aux photos esthétiquement remarquables les évocations de phénomènes exceptionnels du sport, de la vie politique ou de la marche du monde, retenues essentiellement pour leurs qualités mémorielles.

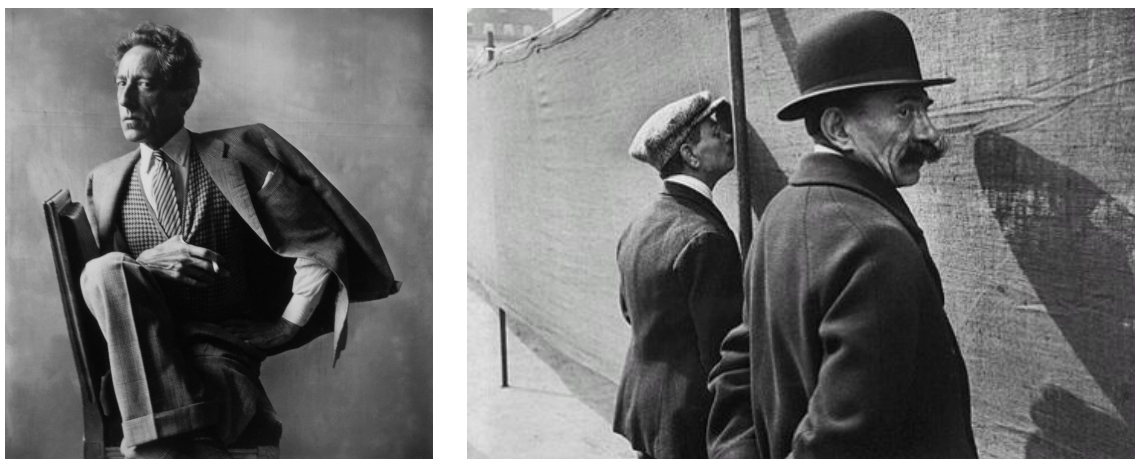


Figure 1.3 – Ces deux photographies sont indéniablement remarquables pour leurs qualités esthétiques. Celle de gauche (Portrait de Jean Cocteau, 1948 par Irving Penn) retient également notre attention parce que nous y reconnaissons un homme célèbre (stimulus d'intérêt de type universel). Celle de droite (Bruxelles, 1932 de Henri Cartier-Bresson) nous interpelle par l'interrogation qu'elle suscite (stimulus d'amazement).

1.2 Art, Beauté, Esthétique ... quel rapport ?

Nous ne serons dorénavant plus concernés que par les seuls éléments relevant de l'esthétique qui nous amènent à préférer telle ou telle image.

Avant d'explorer plus loin ces éléments, rappelons ici quelques définitions.

« **L'Esthétique** est la science qui a pour objet de rechercher et de déterminer les caractères du beau dans les productions de la nature ou de l'art » (définition de l'Académie française de 1835). En termes plus modernes, c'est « la branche de la philosophie concernée par :

- la nature de l'art, de la beauté et du goût,
- la création et l'appréciation de la beauté. »

« **La Beauté** est la qualité de ce qui est beau [...]. C'est le caractère de ce qui provoque l'admiration et l'émotion, par ses formes, ses proportions, ses rythmes, son harmonie » (définition de l'Académie française de 1986).

On préfère souvent aujourd'hui des définitions qui mettent plus l'accent d'une part sur l'origine du stimulus, d'autre part sur la forme de ses effets : « La Beauté est la caractéristique d'une personne ou d'un objet qui procure une expérience perceptuelle de plaisir ou de satisfaction ».

« **L'Art** est la gamme des activités de l'homme conduisant à la fabrication d'artéfacts dans le but qu'ils soient appréciés pour leur beauté ou pour leur pouvoir émotionnel ».

Rapprochant ces trois termes : Esthétique, Beauté et Art, nous voyons qu'ils couvrent bien notre champ d'étude, et comme ils ont les uns et les autres conduit à une abondante littérature, il serait tentant de les mobiliser tous les trois. Il faut cependant être plus prudent avec le troisième : "Art". Si, pendant de nombreux siècles, l'objectif de l'Art a été très majoritairement tendu vers la recherche de la Beauté⁵ il faut cependant examiner avec circonspection son évolution au cours du XIX^e siècle et en particulier considérer la "fracture du XX^e siècle" comme un changement définitif de paradigme. Il est assez clair que le second terme de la définition ci-dessus ("leur pouvoir émotionnel") est devenu l'enjeu premier de l'Art, bien avant "la Beauté". Rétrospectivement, il est aisé de retrouver des prémices de cette préoccupation bien avant les Surréalistes, puisque de nombreuses œuvres

5. Il nous faut ici restreindre notre propos à l'art occidental issu de l'école grecque. Les arts des autres sociétés, et en particulier ceux de l'Orient, relèvent probablement de tout autres intentionalités [Jullien, 2003].

éminemment artistiques (Bosch, Chardin, Goya ...) ne prétendent pas à être premièrement “belles” (figure 1.4).



Figure 1.4 – Nombre d’œuvres d’art ont été conçues sans que le critère de “Beauté” soit mis en avant. L’artiste recherche alors avant tout la “charge émotionnelle” de son message. Ces œuvres “à message” sont très minoritaires dans la période classique (représentée ici par J. Bosch, F. Goya et J.S. Chardin). Elles deviennent rapidement majoritaires au XX^e siècle (représenté ici par M. Duchamp, A. Warhol, E. Munch, F. Bacon) et il serait bien difficile d’évaluer les œuvres du XX^e siècle à la seule aune de leurs qualités esthétiques.

Cette rupture dans les objectifs mêmes de l’Art doit nous faire prendre avec une grande précaution certains textes pour guides dans notre quête du Beau. Cela est très vrai pour ce qui concerne la peinture, mais ça l’est aussi pour ce qui concerne la photographie qui aspire à juste titre au rang d’Art⁶. Il nous faudra suivre avec une grande prudence les recommandations des photographes, surtout les plus chevronnés, lorsqu’ils s’engageront dans des conseils où ne seront pas clairement démêlées les intentions à atteindre une plus grande esthétique et celles visant à une plus forte charge émotionnelle. Nous renvoyons le lecteur aux remarques faites plus haut en section 1.1. Nous noterons aussi qu’une définition claire de ce qu’est l’Art dans notre société (voir par exemple [Danto, 1992, Danto and Goehr, 2014]) nous sera particulièrement utile pour répondre à deux questions souvent entendues qui empoisonnent l’art contemporain :

- une copie de la Joconde est-elle aussi belle que l’original⁷ ?
- est-ce que la boîte de soupe Campbell d’A. Warhol est plus belle que celle de l’épicerie du coin et qu’est-ce qui en fait une œuvre d’art ?

1.3 Notre projet

Il nous faut expliquer tout d’abord pourquoi nous abordons ce sujet aujourd’hui. Les premières lignes de ce texte rappelaient combien l’image est aujourd’hui abondante dans notre société. Elles évoquaient la complexité de la tâche nécessaire de tri des masses de photographie ainsi que le rôle de l’esthétique, à côté d’autres émotions, l’intérêt et la surprise, pour assurer ce tri. Il nous faut

6. Que la photographie puisse accéder au rang d’art est aujourd’hui bien admis parmi les amateurs d’art. Une enquête d’octobre 2017 de la revue *Beaux Arts Magazine* indique qu’à la question *A l’aide de quelle forme d’expression les artistes réussissent-ils le plus souvent à incarner la beauté ?*, la photo est placée au premier rang pour 32 % des Français interrogés, tandis que la peinture ne recueille que 31 % des avis et la sculpture 15 %. On est bien loin de la situation constatée par P. Bourdieu [Bourdieu, 1965] qui qualifiait, au milieu du siècle dernier, la photographie en France d’ “art moyen”, non seulement pour la qualité de sa production mais également pour l’estime dont elle bénéficiait.

7. Cette question fut longuement débattue par G.T. Fechner à propos de la Madone de Dresde (voir [Vidal, 2011]). A contrario de la position exprimée par exemple dans [Danto, 1992] qui exprime un consensus souvent partagé aujourd’hui, Fechner avait soutenu l’étroite dépendance entre *Schönheit* (beauté) et *Echtheit* (authenticité), refusant toute beauté à la copie, aussi parfaite fût-elle.

alors aussi rendre compte des progrès qui ont été faits pour comprendre comment on juge de la qualité esthétique d'une photographie, puis des tentatives qui ont été faites pour nous faire aider dans cette tâche par des outils du traitement de l'information. Ainsi, pour la première fois, en 2018, un concours photographique international est arbitré parallèlement par deux jurys, l'un constitué d'experts photographes, l'autre d'un ordinateur doté d'un logiciel d'intelligence artificielle⁸. C'est également pour assister l'émergence de ces techniques automatiques d'évaluation de la beauté d'une photographie que nous avons entrepris cette étude et rassemblé dans un texte unique un ensemble de points de vue très complémentaires sur ce sujet.

Il nous semble que ces deux aspects : la compréhension des mécanismes mis en jeu par l'homme d'une part, le développement des outils d'assistance d'autre part, vont nécessairement de pair. C'était du moins la conviction profonde qui nous animait au lancement de cette étude. Mais alors même que ce travail avançait (soit de septembre 2016 à septembre 2018), des résultats n'ont cessé de s'accumuler qui semblent contredire cette idée : les réseaux de neurones profonds, quoiqu'encore balbutiants, font fi d'une compréhension des liens entre l'observateur et la photo, mais conduisent pourtant à d'étonnament bons jugements sur les qualités esthétiques de celle-ci (s'ils ne se substituent nullement aujourd'hui à l'œil d'un expert, ils surclassent nettement les autres outils préalablement développés).

Notre sujet a ainsi évolué alors que nous le traitions. Nous expliquons moins que nous le voulions initialement, mais nous témoignons plus précisément des démarches adoptées, nous rapportons les évolutions dans les approches, nous comparons les performances et nous soulignons les avancées et les faiblesses.

Notre travail est scindé en deux grandes parties.

Dans la première, nous essayons de faire un bilan des travaux entrepris dans divers domaines de la connaissance pour identifier les mécanismes de la perception esthétique. C'est tout d'abord l'énorme corpus de la littérature philosophique que nous chercherons à cartographier de façon très schématique au chapitre 2. Nous nous efforcerons de simplifier les très nombreux travaux qui courent depuis 25 siècles de façon à dégager des pistes universelles. Nous examinerons ensuite ce que nous disent les biologistes, et surtout les neurologues, des mécanismes de la perception esthétique. Le développement des outils modernes d'imagerie cérébrale a bouleversé notre connaissance de la cartographie des aires impliquées dans l'élaboration de nos émotions et de nos jugements. Les travaux qui relèvent du domaine de l'art, actuellement regroupés sous le terme de neuro-esthétique, seront présentés au chapitre 3. La photographie, au cours des 150 années de son histoire, a également développé un important corpus de recommandations et d'enseignements, souvent issus de la pratique de la prise de vue, mais fréquemment aussi étayés par des expérimentations et des démarches scientifiques s'appuyant sur les connaissances issues de la psychologie expérimentale. Ces résultats que les traiteurs d'image ont souvent suivis seront présentés au chapitre 4.

Dans la seconde partie nous examinerons les travaux mettant à contribution les outils issus des mathématiques pour mesurer la beauté. Tout d'abord, au chapitre 5 nous présenterons les techniques, parfois assez anciennes, qui proposent une formule complète (*closed form*) de la beauté (comme celle proposée par D. Birkhoff). Ces travaux, quoique très rares, se poursuivent encore de nos jours. Les trois chapitres suivants sont consacrés aux méthodes actuelles issues de l'intelligence artificielle. Nous ferons tout d'abord un examen global du contexte qui rend ces démarches possibles, en particulier des bases de données d'images disponibles sur le réseau, puis des expertises que l'on peut mobiliser pour assurer l'indispensable étape d'apprentissage automatique (chapitre 6). Le chapitre 7 détaillera la première famille des travaux conduits à partir d'une détection de primitives sensées contribuer à la qualité esthétique et leur classification par des techniques de *machine learning*. Ces travaux ont pris naissance vers l'an 2000. Le dernier chapitre couvrira les travaux les plus actuels, conduits essentiellement depuis 2014, effectués à l'aide de réseaux de neurones profonds et se passant donc de toute expertise humaine, hormis dans la constitution des sources d'images annotées (chapitre 8).

8. Il s'agit de la compétition "Spark" (*SPARK, a RenAIssance photographic contest*) de Huawei, construit autour de son modèle P20 Pro de téléphone portable, <https://consumer.huawei.com/uk/campaign/sparkrenaissance/> (juillet 2018).

Bien sûr nous apporterons une conclusion à ce travail, en remettant en perspective les contributions des divers chapitres mais avec le sentiment que nous ne sommes qu'au démarrage d'une époque nouvelle et qu'il est plus que probable que l'avenir se fera un malin plaisir de déjouer nos prévisions.

Chapitre 2

L'héritage des philosophes

L'importance accordée par les philosophes à la notion de Beau depuis l'Antiquité se lit à travers l'abondance des œuvres majeures que cette notion a engendrées. Les “travaux théoriques” de Platon, de Socrate et d'Aristote, appuyés sur les “travaux pratiques” d'Homère, de Phidias ou de Praxitèle, fournissent le cadre général de réflexion que la Grèce a transmis au monde romain, puis au Moyen-Âge et enfin à la Renaissance où il s'est vu accorder une place prééminente dans une vie artistique et intellectuelle particulièrement exubérante. Au Siècle des Lumières, c'est le monde des idées qui s'en empare à l'occasion de disputes particulièrement vigoureuses (Boileau, Locke, Diderot . . .) mais sans que le domaine des Arts le remette vraiment en cause. A la fin du *XVIII^e* et au *XIX^e* siècle, il n'est pas de grand philosophe qui ne se mêle au débat : Kant, Hegel, Schopenhauer, Nietzsche . . . Ils produisent sur ce sujet les traités essentiels de nos bibliothèques, cependant que les artistes entament de leur côté une révolution copernicienne dans la conception de leur art, emblématiquement portée par les Romantiques en littérature et par les Impressionnistes en peinture. Dans le domaine de l'Art, le *XX^e* siècle s'ouvre sur un indescriptible champ de bataille où s'affrontent les théoriciens retranchés derrière des positions passionnées : Heidegger, Adorno, Wittgenstein, Deleuze, Derrida, Foucault, Eco, Danto . . .¹, tandis que le monde des œuvres d'art est livré à une rare anarchie².

Il n'est pas question bien sûr de rapporter ici fidèlement tous ces apports. Nous en ferons un raccourci caricatural, extrêmement simplifié (et donc extrêmement critiquable), qui aura pour seul mérite de bien situer l'enjeu de notre étude et d'en placer certains termes en perspective. Nous introduirons ainsi deux points de vue de la conception esthétique, aux antipodes l'un de l'autre : l'approche “objectiviste” qui attribue le mérite de la beauté au seul objet qui la porte, et l'approche “subjectiviste” qui au contraire situe la beauté dans le ressenti de l'observateur. Nous montrerons comment la première a laissé la place à la seconde au cours des derniers siècles, puis comment les deux approches se sont tant bien que mal hybridées dans un mouvement que l'on appelle parfois “interactionniste”. Nous montrerons surtout quelles implications ces deux théories induisent dans notre projet de déterminer automatiquement la beauté d'une photographie et pourquoi elles sont immanquablement présentes dans les attendus des logiciels développés aujourd'hui.

Introduisons tout d'abord l'enjeu de cette classification en deux familles opposées, et pour cela nous citons D. Diderot qui en fixe clairement les termes dans l'Encyclopédie : *“Il est évident que S. Augustin avoit été beaucoup plus loin dans la recherche du beau que le philosophe Leibnizien³ : celui-ci semble prétendre d'abord qu'une chose est belle, parce qu'elle nous plaît ; au lieu qu'elle ne nous plaît que parce qu'elle est belle ; comme Platon & S. Augustin l'ont très-bien remarqué.”* Est-ce le Beau qui donne du plaisir, ou le plaisir qui crée le Beau ? Diderot, se référant aux Anciens, a choisi, mais il existe bien deux façons de prendre le problème, qui ne seront probablement pas équivalentes lorsqu'il

1. On consultera avec intérêt : [Gombrich, 1960, Danto, 1964, Tatarkiewicz, 1970, Zemach, 1987, Kemp, 1990, Solso, 1996, Morin, 2016]

2. On ne peut manquer de citer ici “la fin de l'histoire de l'art” annoncée par A. Danto en 1964 [Danto and Goehr, 2014].

3. Diderot parle ici du philosophe et mathématicien Christian Wolff (1679-1784).

faudra, comme c'est notre propos, examiner les moyens de faire évaluer le Beau par une machine.

Muni de ces deux référentiels philosophiques, nous pourrions alors poursuivre notre démarche de façon à déterminer les arguments de cette perception du Beau, essayant d'apporter, là encore, réponse à Diderot : *“Tout le monde raisonne du beau : on l'admire dans les ouvrages de la nature : on l'exige dans les productions des Arts : on accorde ou l'on refuse cette qualité à tout moment ; cependant si l'on demande aux hommes du goût le plus sûr & le plus exquis, quelle est son origine, sa nature, sa notion précise, sa véritable idée, son exacte définition ; si c'est quelque chose d'absolu ou de relatif ; s'il y a un beau essentiel, éternel, immuable, règle & modèle du beau subalterne ; ou s'il en est de la beauté comme des modes : on voit aussitôt les sentimens partagés ; & les uns avoient leur ignorance, les autres se jettent dans le scepticisme.*

Comment se fait-il que presque tous les hommes soient d'accord qu'il y a un beau ; qu'il y en ait tant entr'eux qui le sentent vivement où il est, & que si peu sachent ce que c'est ?” (in D. Diderot, Encyclopédie, entrée : Beau).

2.1 L'approche objectiviste

L'approche objectiviste est celle qui fut formalisée dès l'Antiquité grecque, celle de Platon⁴, Socrate et Aristote⁵, qui fut transmise au monde occidental, s'épanouit à la Renaissance⁶, prospéra dans l'époque Classique et tint le haut du pavé jusqu'au milieu du *XIX^e* siècle, au moins dans le milieu artistique, si ce n'est dans le milieu philosophique.

L'approche objectiviste affirme que la Beauté est une qualité de la personne, de l'objet, de la musique concernés. Cette Beauté est le reflet de propriétés universelles d'harmonie dans les proportions et dans les formes, propriétés qui ne nous sont pas exactement connues mais qui sont préexistantes à notre observation et s'imposent à notre jugement. La Beauté est à ce titre perçue également par tout observateur⁷, en tous temps et en tous lieux⁸. Ainsi, parmi d'autres, la Vénus de Milo, l'Apollon du Belvédère, le David de Michel-Ange ont été parés par les experts et historiens d'art les plus réputés de ces qualités qui les imposent à l'humanité.

L'approche objectiviste justifie pleinement que l'on s'attache à déterminer, sans l'aide d'un observateur, les qualités esthétiques d'un objet puisque celles-ci sont immanentes et s'expriment également, avec ou sans observateur. Cette intention est d'autant plus accessible que les règles de l'esthétique sont universelles et largement indépendantes de la forme prise par l'*artefact* : peinture, musique, architecture, sculpture. Ces règles, au sens antique du terme, résident dans l'harmonie des formes et des proportions et en particulier dans la consonnance du tout et des parties, reflet d'une harmonie céleste transcendente. Newton, découvrant le spectre lumineux, leur a adjoint une dimension supplémentaire : l'harmonie chromatique qui interviendra bien sûr dans notre propos.

E. Kant a conféré une forme plus moderne à l'interprétation objectiviste de la Beauté en en “humanisant” l'origine⁹. Il en fait le produit du sens esthétique de l'observateur plus que propriété de l'objet lui-même (répondant à un débat intense au *XVIII^e* siècle comme nous le verrons plus loin), mais ce sens esthétique qu'il défend s'impose à tous de façon universelle et procure un plaisir désintéressé (l'expérience de la beauté ne répond pas à un besoin ou une finalité précise). Dans l'esprit

4. Platon aurait écrit deux *Traité du Beau*, l'un dans Hippias Majeur, l'autre dans Phèdre, [Platon, 2011].

5. “... les formes les plus hautes du Beau sont l'ordre, la symétrie, le défini...”, Aristote, Métaphysique M 3.

6. “La Beauté est une certaine harmonie régulière entre toutes les parties d'une chose, harmonie d'une espèce telle que rien ne pourrait être ôté ou ajouté, ou changé en elle sans qu'elle ait aussitôt moins de charme.” L. B. Alberti, *L'Art d'Edifier*, Florence, 1485 [Alberti, 1992].

7. Selon les auteurs, cette universalité de perception de la beauté par tout observateur doit être modérée : dans la démocratie grecque, tous les hommes ne sont pas égaux et le discours des philosophes s'adresse aux seuls *kalos kagathos* ; à la Renaissance les textes s'adressent au petit nombre des lettrés et le temps des Lumières distingue l'Honnête Homme. Le langage de l'art est accessible à tout observateur pour peu qu'il soit “assez sensible” pour recevoir le message du philosophe.

8. On lit par exemple “La Beauté est goûtée de toutes les nations, à toutes les époques” dans [Hegel, 1838].

9. Les textes d'Emmanuel Kant se rapportant à l'esthétique sont en grande partie issus de *Critique de la faculté de juger*, de 1790 [Kant, 1985].

même du texte que nous entreprenons, Kant réserve à l'observateur la faculté de juger objectivement du beau et de faire connaître ce jugement par des arguments partageables. Nous verrons plus en détail le rôle de Kant en fin de ce chapitre car il tient une place particulière dans le débat objectiviste.

Les temps modernes n'ont pas oublié les interprétations objectivistes du Beau, en particulier au sein d'une école philosophique qualifiée de "réalisme esthétique" [Pouivet, 2006, Zemach, 2005]. Par exemple, dans [Zemach, 2005], E. Zemach accorde aux œuvres d'art des "prédicats" esthétiques (*l'élégance, l'harmonie* ...) qui dépendent de "prédicats" non-esthétiques liés à la forme, la couleur, etc. Cette dépendance est une sorte de corrélation univoque (du non-esthétique vers l'esthétique) qui permet à Zemach d'affirmer que la beauté est une propriété réelle de certaines œuvres (voir sur ce point la discussion de [Brehin, 2007]).

Nous verrons que l'approche objectiviste a irrigué tous les travaux d'évaluation non supervisée de la beauté des images. En particulier, les travaux conduisant à des formules complètes et explicites telles que nous les verrons au chapitre 5 (par exemple les approches de Birkhoff, de Moles, de Machado et Cardoso ...) ont cherché à décliner les deux concepts de *simplicité* et de *complexité*, très généralement associés à la Beauté dans les textes classiques. Les approches par apprentissage (chapitres 6, 7 et 8), plus modernes, font moins appel à des critères précis des textes anciens, mais elles reposent plus encore sur les présupposés d'universalité et d'intemporalité posés en principes par les objectivistes.

2.1.1 Les esthétiques des Objectivistes

Voyons comment diverses conceptions du Beau sont rangées sous l'étiquette objectiviste. On trouvera dans [Tatarkiewicz, 1970] de très larges développements d'où sont issues ces lignes.

L'esthétique de la période classique grecque : la Grèce hellénique

La période classique athénienne (*V^e* et *IV^e* siècle av JC) porte toutes les références ultimes à l'esthétique. Le Beau y est doté de vertus autant esthétiques que morales. Il est le produit de trois sources : "*la matière, fournie par la nature, le savoir, fruit de la tradition et le travail donné par l'artisan*"¹⁰ ([Tatarkiewicz, 1970], livre 1). La notion de création y est quasiment absente, la fidélité à la tradition y est dominante, tradition exprimée par les "canons" qui fixent des formes précises à l'objet. Les canons ont le plus souvent de lointaines sources religieuses et liturgiques, mais d'autres semblent bien n'avoir que des fondements esthétiques.

Les canons nous sont assez bien connus dans les domaines de l'architecture et de la statuaire où de nombreuses œuvres sont encore disponibles. Ils ont été très largement repris et commentés par Vitruve, à Rome, au premier siècle. Nous avons beaucoup moins de traces en peinture. Ils s'expriment par des règles, de simplicité¹¹, d'ordre¹², d'égalité¹³ et de symétrie en particulier et par des chiffres, rapports entre les grandeurs qui expriment des proportions et reflètent une harmonie¹⁴. La simplicité vise à garantir la "juste mesure", c'est-à-dire l'utilisation des seuls éléments nécessaires à l'expression. Plus les rapports sont simples, plus grand est l'effet esthétique.

Aristote introduit un concept supplémentaire qui résume les propriétés ci-dessus, celle de perceptibilité qui prône que les qualités du bel objet doivent apparaître naturellement et spontanément, en adaptant la taille de l'objet au champ de vue de l'observateur et en éliminant les artefacts superflus.

10. Pour les Grecs, les arts plastiques : peinture, architecture, sculpture ... sont œuvres d'artisans et non d'artistes comme l'est l'expression orale, dotée du pouvoir quasi-divinatoire de préparer l'avenir.

11. "*Si quelqu'un dépasse la mesure, les choses les plus agréables deviennent déplaisantes.*" Démocrite, rapporté par J. Stobaeus, Flor, III, 17, 38, frg. B 23 Diels, cité par [Tatarkiewicz, 1970].

12. "*L'ordre et les proportions sont beaux et utiles, tandis que le désordre et l'absence de proportions sont laids et inutiles.*" Ecole de Pythagore, rapporté par J. Stobaeus IV I, 40 H. frg. D 4 Diels, cité par [Tatarkiewicz, 1970].

13. "*Dans toute chose, l'égalité est juste, l'excès et le défaut ne le sont pas à mon avis.*" Démocrite, Democrites, Sent. 68, frg. B 102 Diels, cité par [Tatarkiewicz, 1970].

14. Cette vision pythagoricienne de la beauté est reprise dans les textes chrétiens écrits au début de notre ère : "*Omnia in mensura et numero et pondere disposuisti. (Tu as fait toute chose en accordant mesure, nombre et poids.)*", Livre de la Sagesse, XI, 21.

D'ailleurs Aristote soutient que la perception du beau s'impose à l'esprit sans justification, que ce n'est pas une propriété des sens ou du jugement¹⁵.

L'esthétique utilise les mathématiques, mais elle n'est pas d'essence abstraite. Elle puise ses sources dans la nature et dans l'univers et en particulier dans le corps humain qui, partie de l'univers, en reflète l'harmonie. Aujourd'hui, nous dirions que cette harmonie est fractale car elle concerne les proportions à toutes les échelles : celle du buste au corps entier, celle de la main au membre, celle des phalanges à la main, celle du nez au visage et ainsi de suite¹⁶. Les rapports ainsi préférés sont 1/2 ou 1/3, ou ceux construits à partir du triangle rectangle "parfait" de côtés (3,4,5) dont on déduit le nombre d'or $\frac{1+\sqrt{5}}{2} = 1,618$ ou le rapport approché 8/5 = 1,6. Mais ces proportions sont dans des rapports finalement assez nombreux. Par exemple, en architecture, ils doivent se conformer aux exigences de la mécanique ; ainsi, les colonnes des temples les plus hauts sont écartées d'un espace plus faible pour assurer la répartition de charges plus importantes et l'on trouve alors des rapports plus complexes.

Dans le domaine des couleurs, l'harmonie se trouve dans la composition de quatre primaires : le blanc, le noir, le rouge et le jaune (Démocrite, Empédocle)¹⁷.

Si les canons sont des contraintes supérieures, de petits écarts aux canons sont tolérés et même appréciés de certains¹⁸. Ils ont pour but de mieux faire percevoir les règles esthétiques par l'observateur. Ainsi il avait été noté que le parallélisme des colonnes d'un temple élevé donnent l'impression de diverger. Il est alors conseillé de rendre ces colonnes légèrement convergentes de façon qu'un observateur, au centre de la place, les voit parallèles. Vitruve rapporte que Démocrite usait de tels artifices pour exprimer le relief dans ses peintures.

La Grèce hellénistique et Rome

La Grèce hellénique a fait place, après les conquêtes d'Alexandre, à la Grèce hellénistique et au monde romain. Pour ce qui concerne les arts et na notion de beau, ils s'inscrivent plus ou moins l'une et l'autre dans la tradition classique. Quoique le contexte social soit complètement bouleversé avec l'émergence d'empires immenses, l'ouverture des frontières à de lointaines cultures très élaborées et l'apparition de nouveaux régimes politiques, les grandes lignes de l'esthétique de Platon et d'Aristote demeurent au centre de la *doxa* esthétique.

Néanmoins trois grands courants philosophiques se sont fait sentir qui traitent différemment de l'esthétique :

- les hédonistes (Epicure) : ils assimilent art et plaisir et symétriquement plaisir et art. Il n'y a pas d'autonomie de l'art et pas de règles esthétiques ;
- les sceptiques soulignent les désaccords et les contradictions des jugements sur la beauté et l'art ; en conséquence, théoriser sur l'art est impossible, inutile et dangereux ; ils réfutent les arguments platoniciens d'universalité de l'art ;
- les stoïciens (Zénon, Sénèque, Cicéron) subordonnent le beau aux valeurs morales ; ils reconnaissent une beauté morale et une beauté des sens, la seconde ayant généralement moins de valeur. Ils s'accordent aux critères platoniciens et leur font écho.

Dans la pratique, se sont les règles platoniciennes et péripatéticiennes qui sont le plus généralement suivies. L'influence des stoïciens se traduit par une plus grande souplesse dans la mise en œuvre des canons, qui régissent les grandes lignes des œuvres, tandis que les détails s'accrochent d'un *decorum*¹⁹

15. Diogène affirme que seul un aveugle peut s'interroger sur la raison de perdre du temps sur la notion de beau.

16. Pour cela, cette esthétique a été qualifiée d'"antropométrique" par E. Panofsky [Panofsky, 1983].

17. "Il y a quatre couleurs correspondant au nombre des éléments : blanc, noir, rouge et jaune." d'après Empédocle, rapporté par Aëtius, Plac. I, 15, 3 ; frg A 92 Diels, cité par [Tatarkiewicz, 1970].

18. Mais ce n'est pas le cas de Platon qui recommande le plus grand accord avec la tradition : "... en bref, ceux qui sont en charge de notre communauté doivent prendre garde à ne pas négliger la moindre consigne envers l'innovation à travers notre système d'éducation du corps ou de l'esprit ..." Platon, la République, 424 B, cité par [Tatarkiewicz, 1970].

19. "Ut enim in vita, sic in oratione nihil est difficilius quam quid deceat videre. Πρέπον appellant hoc Graeci ; nos dicamus sane decorum. De quo praeclare et multa praecipuntur et res est cognitione dignissima. Dans l'oratoire, comme dans la vie, il n'est rien de plus difficile que de reconnaître ce qui convient. Les Grecs l'appellent "prepon" tandis

qui peut être interprété comme une déclinaison locale et partielle du canon spécifique à chaque objet représenté.

Le Bas Moyen-Âge : Europe de l'Ouest et Byzance

Reprise par le monde chrétien naissant, l'esthétique platonicienne fait migrer la beauté de la nature au monde transcendant (Plotin, Pseudo-Denys l'Aréopagite). C'est le monde divin qui est paré de tous les attributs de la beauté et c'est dans sa représentation que l'artiste met en œuvre les recommandations d'harmonie, d'équilibre, de simplicité, ... C'est la période du néo-platonnisme.

De cette dimension divine, la beauté gagne une vertu nouvelle. Alors qu'elle s'imposait passivement à l'observateur, maintenant elle rayonne et participe à la diffusion de la parole. La notion de lumière y prend ainsi une place particulière avec ce qui l'accompagne de brillant. En contrepartie, elle perd en sensualité.

Plus à l'Est, dans cette autre partie de l'empire romain qui commence à se séparer de l'Occident, une figure nouvelle de la peinture émerge : l'icône prend à Byzance une place centrale ; seule l'image de Dieu ou de ses saints est digne d'effort elle reproduit un monde formellement beau et dépourvu des troubles terrestres qu'exprimerait une expression faciale. Les codes (transpositions des canons) s'imposent : œil en position centrale, figures allongées, couleurs artificielles, cadre dépourvu de contexte matériel ou réduit à des fragments symboliques de végétation, si ce n'est idéalisé par un disque d'or.

Enfin, en Afrique du Nord, Augustin, évêque d'Hippone, partant des textes platoniciens revus par Plotin, jette les bases de ce qui fera l'esthétique de l'Occident jusqu'à la Renaissance : la beauté est objective et indépendante de nous ; elle est créée par l'harmonie des parties, leur ordre et leur unité et cela doit être réglé par des nombres²⁰. Elle culmine dans la beauté divine, qui surpasse et inspire toute beauté. C'est aussi probablement lui qui introduit l'idée qu'il n'y a pas de valence négative du beau (le contraire du beau), mais simplement une absence de beau dans ce que nous désignons par laid²¹.

Le Haut Moyen-Âge

A partir de l'an 1000, c'est avec l'architecture religieuse que se manifeste la résurgence d'une vigoureuse expression esthétique : à travers le style roman tout d'abord, puis du style gothique. Art d'artisans bien sûr, mais aussi d'ingénieurs ; et donc art du Nombre dans la tradition pythagoricienne, mais aussi dans l'enseignement et le secret des ateliers et des guildes. Les règles de proportion, d'harmonie, de rythme²² sont retenues et transmises, mais les sources sont généralement omises ; les chiffres s'imposent par eux-mêmes comme un prérequis de beauté et non plus, comme à l'origine platonicienne, par leur rapport au corps humain, à la sagesse de la pensée sociale et collective. Les prouesses de la technique aidant, la lumière est amenée dans les édifices ; les arts ont leur inspiration au firmament. C'est R. Grosseteste qui fait le mieux la synthèse entre l'esthétique des nombres, de la

que nous disons "decorum". Il y a de nombreuses et excellentes règles qui méritent d'être connues.)", Cicéron, Orator, 21, 70, cité par [Tatarkiewicz, 1970].

20. "Omnia enim quanto magis moderata, speciosa, ordinata sunt, tanto magis utique bona sunt ; quanto autem minus moderata, minus speciosa, minus ordinata sunt, minus bona sunt. (En toute chose, plus il y a de mesure, de forme, de nombre, meilleure elle est ; et moins il y a de mesure, de forme, de nombre, moins elle est bonne.)", Augustin, De l'origine du Bien, 3, PL, 42, c, 554, cité par [Tatarkiewicz, 1970].

21. Il semble qu'il lui était difficile d'affirmer que dans un monde créé par Dieu il y ait eu création du laid. Ainsi, il écrit : "Cuique naturae non est malum nisi minui bono. (Aucune nature n'a de mal en elle, mais seulement un déficit de bien.)", Augustin, De l'origine du Bien, 14-17, PL, 42, c, 555-6, cité par [Tatarkiewicz, 1970]

22. "Parilitas autem dimensionis secundum aequalitatem, similitudinem, compositionem et modificatam et commensuratum congruentiam artium non minima pars pulchritudinis est. (L'unité de dimension établi sur les principes d'égalité, de similarité, d'arrangement approprié, d'adaptation et de commensuration des parties n'est pas le moindre facteur de beauté.)", Baudouin de Cantorbéry, Tractatus de salutatione angelica, PL 204, c, 469, cité par [Tatarkiewicz, 1970]

géométrie et de la lumière²³, dans une approche qui vise à donner une vision mathématique unifiée de la beauté [Tatarkiewicz, 1970].

Une distinction nouvelle dans l'esthétique est apportée par Alhazem, philosophe et mathématicien du monde arabo-musulman au virage de l'an 1000. Alhazem, fidèle propagateur des écrits helléniques et hellénistiques est aussi un excellent opticien. Son message diffuse lentement en Europe et il faut plusieurs siècles pour qu'il soit repris par la nouvelle école scholastique (E. Vitellion, R. Bacon). Il défend l'idée que l'oeil ne perçoit directement que la couleur et la lumière. Nous n'apprécions les autres composantes de l'esthétique (forme, composition, disposition) que par d'autres actions qui composent la perception dont la mémoire et la réflexion. C'est cette forme nouvelle - et singulièrement moderne - qui est à la base de la perception esthétique qui peut distinguer une beauté simple, immédiatement perçue par l'oeil seul, puis une plus complexe s'adressant à des formes elles-mêmes plus complexes²⁴. Si la distinction de deux types de beau est classique dans la pensée antique, elle y est fondée sur des principes philosophiques transcendants, tandis que la distinction proposée ici repose sur des mécanismes psycho-physiologiques inconnus mais pressentis. Il n'y a que deux propriétés qui participent au premier type de beauté (la lumière et la couleur), mais Alhazem et Vitellion recensent près de 20 propriétés contribuant à la seconde ([Tatarkiewicz, 1970], II, p. 266). Regrettons cependant que ces propriétés soient bien loin de fournir des règles opérationnelles de beauté. Notons enfin que ces philosophes demeurent fondamentalement objectivistes dans leur conception de la beauté. S'ils prennent en compte l'observateur, c'est pour comprendre les mécanismes qui le font sensible au beau.

G. d'Ockham traite peu de l'esthétique mais il se penche sur la nature de l'image et y décèle une dimension originale qui échappe à l'objet représenté et qui est le fruit de l'intention du dessinateur²⁵. C'est là aussi un concept nouveau qui sera particulièrement pertinent dans l'esthétique de la photographie où l'objet qui est fixe et le même pour tous pourra être "transcendé" en objet d'art par l'oeil du photographe. La subjectivité s'introduit alors, non pas dans l'observateur où nous l'attendions naturellement mais dans le créateur.

La Renaissance

C'est ce mouvement qui a porté la revisite de la pensée hellénique, puis sa reformulation en termes modernes et ensuite sa diffusion en l'Europe. Elle en a hérité pour partie son nom²⁶. Elle s'est développée aux XIV^e et XV^e siècles dans le Nord de la péninsule italienne, mais surtout à Florence, Venise puis Rome. L'héritage platonicien est le point de départ de la pensée esthétique, appuyé par les textes De Plotin et de Vitruve dans le domaine architectural et complété de notions issues d'Aristote. L'esthétique se décline sous la Renaissance vers de nombreux autres domaines de la vie courante (orfèvrerie, marquetterie, tissage, ...). Le rôle des sciences est particulièrement renforcé pour soutenir la notion de forme dans les arts visuels. La nature reste le modèle premier de la beauté et l'artiste doit s'efforcer de s'en approcher. Sa contribution se place souvent dans la subtilité et la pertinence des écarts qu'il apporte dans cette reproduction. La notion d'harmonie globale préside à la composition

23. "Solae quinque proportionales repertae his quattuor numeris : unum, duo, tris, quattuor, aptantur compositioni et concordiae stabilienti omne compositum. (La composition et l'harmonie de toute chose composée découle seulement des cinq proportions que l'on trouve entre les quatre nombres : un, deux, trois, quatre.)", Grosseteste, de la Lumière, (Baur, 58) et "lux est maxime pulchricativa et pulchritudinis manifestiva. (La lumière rend les choses belles et montre leur beauté au plus haut point.)", Grosseteste, Comm. in Div. Nom., IV (Pouillon 320), cités par [Tatarkiewicz, 1970].

24. "Omnis vera comprehensio formarum visibilium aut est per solam intuitionem, aut per intuitionem cum scientia praecedente. (Toute appréhension des formes visibles, se fonde soit sur une observation directe, soit sur une observation à travers une connaissance antérieure.)", Vitellion, Optique, III, 62 (Alhazem, II, 69), cité par [Tatarkiewicz, 1970].

25. "(a) Strictissime et sic imago est substantia formata ab artifice ad similitudinem alterius ... Sic accipiendo imaginem de ratione imaginis est, quod fiat ad imitationem illius, cuius est imago ; [...] .

(b) Alio modo accipitur imago pro tali formato, sive fiat ad imitationem alterius, sive non.

((a) Au sens strict, l'image est une substance faite par l'artiste à la ressemblance d'une autre ... Dans ce sens, il est important que l'image soit à l'imitation de ce qu'elle représente ; [...]. (b) Dans un autre sens, l'image est conçue comme toute chose faite par l'artiste, comme l'imitation d'une autre ou non.)", Ockham, Questions en IV, Sententiarum libros, I, d., 3, q. 10B, (Baudry), cité par [Tatarkiewicz, 1970].

26. "Renaissance" a deux sens, le premier exprime la mise au monde de l'homme nouveau, libéré des structures mentales du Moyen-Âge, le second la remise à l'honneur de la philosophie hellénique au sein de la société (voir [Tatarkiewicz, 1970], III, p. 32).

des œuvres picturales²⁷. La beauté est universelle et tout le monde doit en convenir²⁸.

C'est aussi au cours de la Renaissance que la place de l'artiste est vraiment reconnue à l'égal de celle d'un créateur²⁹. C'est aussi à cette période que l'on reconnaît que l'objectif de l'artiste est bien le plaisir qu'il tire d'une part de sa création, d'autre part de son partage avec les spectateurs.

2.2 L'approche subjectiviste

Le paradigme de la beauté universelle et intrinsèque à l'objet est remis en cause au *XVI^e* siècle. En fait elle n'avait jamais été totalement absente de critiques, par exemple celles des sophistes grecs³⁰, mais ces critiques prennent au *XVI^e* siècle une forme plus officielle. En 1588, William Shakespeare écrit : "*Beauty is bought by judgement of the eye*"³¹ et, en 1591, Giordano Bruno : "*Alia est pulchritudo unius speciei, alia alterius, alia unius generis, alia alius*"³², ou "*Nihil absolutum pulchrum*"³³.

En 1630, René Descartes exprime sans ambiguïté qu'il ne tient pas le Beau comme objet de rationalité partageable : "*Généralement ni le beau, ni l'agréable, ne signifient rien qu'un rapport de notre jugement à l'objet; et parce que les jugements des hommes sont si différents, on ne peut dire que le beau, ni l'agréable, aient aucune mesure déterminée*"³⁴. Mais c'est le philosophe John Locke qui a élaboré et installé la doctrine "subjectiviste". Il exprime de la façon la plus nette qu'il n'y a pas d'acquis dans la pensée humaine, que toute pensée naît de nos observations et que le Beau, comme *Qualité Seconde*, n'existe nulle part en dehors de celui qui le perçoit³⁵.

De nombreux arguments qu'il a utilisés se retrouvent en substance dans les querelles des Anciens et des Modernes, qui fait rage à ce temps dans toute l'Europe. Ses idées sont largement reprises au *XVIII^e* siècle dans des textes de Denis Diderot³⁶, d'Emmanuel Kant³⁷ puis de Johann Wolfgang von Goethe³⁸. Elles affirment que la Beauté est une expérience et non un attribut, que cette expérience est largement le fruit de l'émotion de l'individu, qu'elle est donc personnelle et non universelle, contingente et non immanente. Elle est éminemment fluctuante dans le temps, dépendant des "humeurs" de l'individu. Ce qui est beau ne le sera pas toujours et ce qui ne l'est pas aujourd'hui sera peut-être

27. "*La bellezza non consiste solo nelli colori, ma è una qualità que resulta dalla proporzione e corrispondenza delli membri e delle altre parti del corpo; tu non dirai che una donna sia bella per avere uno bello naso o belle mani, ma quando vi sono tutte le proporzioni. (La beauté ne consiste pas en couleurs, mais en une qualité qui résulte de la proportion et des correspondances des parties du corps. Tu ne diras pas qu'une femme est belle parce qu'elle a un beau nez ou de belles mains, mais seulement si tout est en proportion.)*", J. Savonarole, Sermon sur Ézéchiel, XXVIII, cité par [Tatarkiewicz, 1970].

28. "*Ma e'ci sono alcuni che non approvano simili cose et che dicono che ella è una certa varia oppenione, con la quale noi facciamo giudicio della bellezza, et di tutte le muraglie, e che la forma degli edificij se muta secundo il diletto, e il piacere di ciascuno, non si ristignendo dentro ad alcuni comandamenti de la arte. Comune diffeto de gli Ignoranti, é il dire che quelle cose che non sanno loro, non sieno. (Mais il y a ceux qui ne reconnaissent pas cela et qui disent qu'il y a diverses vues concernant la beauté et l'architecture et que la forme des édifices évolue selon les goûts et les préférences de chacun, sans être gouvernées par les commandements de l'art. C'est une erreur commune des ignorants qui disent que ce qu'ils ignorent n'existe pas.)*", L. B. Alberti, Della pittura, Lib. II, cité par [Tatarkiewicz, 1970].

29. "*La deità, ch'a la scientia del pittorefa che la mente del pittore si trasmuta in una similitudine di mente divina. (La divinité qui est la science de la peinture, transmute l'esprit du peintre en une ressemblance avec l'esprit divin.)*", Leonardo Da Vinci, Traité de la Peinture, frag. 280, cité par [Tatarkiewicz, 1970].

30. On lit ainsi dans les Dialexeis "*Si l'on demandait à tous les hommes de jeter en pile les choses qu'ils trouvent laides, puis de retirer de la pile les choses qu'ils trouvent belles, je pense que rien ne resterait sur la pile, car tous les gens diffèrent dans leurs opinions.*" (Dialexeis, 2, 8, rapporté dans [Tatarkiewicz, 1970]).

31. Shakespeare dans *Love's labour lost*, 1588.

32. "Les divers genres et types de beauté différent les uns des autres", G. Bruno, De vinculis in genere, III, 638, cité par [Tatarkiewicz, 1970].

33. "Rien n'est absolument beau." G. Bruno, De vinculis in genere, III, 637, cité par [Tatarkiewicz, 1970].

34. Lettres à Mersenne du 18 mars 1630 in Œuvres et Lettres, Gallimard, Paris, 1953, rapporté dans [Malinowski-Charles, 2004].

35. Ces idées de Locke font l'objet de l'*Essai sur l'entendement humain* de 1689.

36. Les textes de Diderot concernant la critique esthétique sont regroupés dans la rubrique *Les Salons* de [Diderot, 1769].

37. Voir en particulier [Kant, 1985].

38. Les idées de Goethe se retrouvent dans "*De l'expérience considérée comme médiatrice entre l'objet et le sujet*" de 1792 et plus encore dans "*Le traité des couleurs*" de 1810.

beau demain. L'idée est remarquablement résumée dans la formule célèbre : “*La Beauté est dans l'œil de l'observateur*”³⁹.

Il ne faudrait pas croire cependant que cette transition de l'objectivisme au subjectivisme s'est faite d'un coup. Elle a été mûrie par d'intéressantes réflexions qui tentent de concilier les deux concepts. La plus aboutie est celle de Jean-Pierre de Crousaz, philosophe suisse, auteur d'un *Traité du Beau* [DeCrousaz, 1724] qui reconnaît au Beau deux natures, l'une qui relève des idées (et donc du monde platonicien), l'autre des sentiments⁴⁰. Cette double hérédité accordée au Beau fait dire à Syliane Malinowski-Charles : “*le pas franchi par Crousaz est significatif et il peut être vu comme l'emblème d'une transition majeure entre le paradigme rationaliste (qu'on pourrait dire aussi objectiviste) et le paradigme empiriste ou subjectiviste*” [Malinowski-Charles, 2004]. De Crousaz, explorant plus à fond la notion d'esthétique, s'accorde sur des critères très proches de ceux que mettra en équation Birkhoff, qui mêlent intimement complexité et simplicité⁴¹. Les idées “translationnelles” de Crousaz seront reprises par Jean-Baptiste Du Bos et par Francis Hutcheson [Jullien, 2017] et développées par le père André⁴².

Malgré ces essais pour ménager la vision platonicienne du Beau, le *XIX^e* siècle et la période Romantique ont donné à l'interprétation subjectiviste ses plus belles pages. La Beauté a besoin d'un objet, mais sa relation à l'objet est ténue et peut éventuellement se reporter sur un autre objet si le premier s'éloigne ou disparaît. Si poètes et musiciens sont les premiers à suivre cette voie, les peintres s'en emparent également. Dans la société, la rupture avec l'École Classique est brutale et les deux courants sont très largement parallèles et concurrents, chacun “tenant” sa vérité. Le succès des Impressionnistes sur les Pompiers en France consacre le renversement des tendances dans le domaine des arts picturaux. L'éclosion ensuite à travers le monde de nombreux autres mouvements artistiques : Pointillisme, Fauvisme, Cubisme, Surréalisme ... installe durablement cette approche centrée sur le spectateur au cœur de la démarche de l'artiste, tandis que se maintiennent ici et là des îlots de classicisme.

L'émergence de l'école psychanalytique à la fin du siècle ne peut que renforcer cette interprétation subjectiviste de la Beauté, enfonçant de plus ses ressorts dans le brouillard de l'inconscient⁴³. Suivant J.P. Changeux [Changeux, 2016], on pourrait dire que le subjectivisme a lentement migré d'un sensualisme (à la façon de Condillac) et d'un empirisme issu des philosophes britanniques (Locke, Hume), vers une cérébralisation, domaine où se retrouvent les neuro-biologistes et les informaticiens tenant de l'intelligence artificielle.

La Beauté, que l'on cherchait dans l'objet, est donc à trouver dans l'observateur : dans son œil pensait on, il y a deux siècles ; plutôt dans son cerveau, dit-on aujourd'hui. Mais les pistes pour la cerner sont diverses et passablement obscures.

Sous les efforts de courants de pensée très variés et sous l'inspiration des diverses sciences qui se sont emparées du problème, de nombreux travaux originaux ont vu le jour. La psychologie expérimentale, appuyée sur des techniques mises en place par G.T. Fechner en 1871, permet de tester divers paramètres de la perception visuelle susceptible de valider l'universalité du beau ; en pratique elle démontre surtout l'étendue du relatif lié à l'éducation, la culture et l'humeur. Au milieu du *XX^e* siècle, D.E. Berlyne [Berlyne, 1970, Berlyne, 1971] a perfectionné les méthodes de l'esthétique expérimentale et a inséré la beauté dans le cadre formel de la psycho-biologie ; F. Molnar [Molnar, 1974, Molnar, 1997] s'est attaché particulièrement à lier esthétique et perception, tandis que R.L. Solso [Solso, 1996] a

39. “*Beauty is in the eye of the beholder*”, phrase de Margaret Wolfe Hungerford dans Molly Bawn (1878) souvent attribuée à Oscar Wilde - on ne prête qu'aux riches !

40. De Crousaz écrit ainsi : “*Quand on demande ce que c'est que le Beau, on ne prétend pas parler d'un objet qui existe hors de nous et séparé de tout autre, comme quand on demande ce que c'est qu'un cheval ou ce que c'est un arbre.*”, dans [DeCrousaz, 1724].

41. “*le beau est créé par l'unité au sein de la diversité*”, dans [DeCrousaz, 1724], prenant ses inspirations dans la beauté des figures mathématiques : triangle, carré, polygone

42. “*Je dis qu'il y a un beau essentiel, & indépendant de toute institution, même divine : qu'il y a un beau naturel, & indépendant de l'opinion des hommes : enfin qu'il y a une espèce de beau, d'institution humaine, & qui est arbitraire jusqu'à un certain point.*”, dans [André, 1759], pp 4.

43. En 1930, Sigmund Freud écrit : “*... la beauté dérive du domaine de la sensibilité sexuelle.*” S. Freud, *La Malaise dans la culture*, Presses Universitaires de France, Paris, 1985.

plutôt cherché ses liens avec la psychologie et la cognition.

L'étape suivante dans cette voie est franchie avec l'avènement des grands travaux de neuro-biologie, à la fin du *XX^e* siècle. Ils permettent de compléter notre connaissance du fonctionnement du système visuel. Les travaux de physiologie du milieu du *XX^e* siècle avaient permis de bien comprendre le rôle des voies visuelles et du cortex postérieur. La neuro-biologie nous éclaire aujourd'hui sur le rôle du cortex préfrontal lors des tâches d'évaluation et lors de la prise de décision sous l'effet d'une émotion issue de notre perception. Sous l'éclairage des nouvelles techniques d'imagerie cérébrale, un certain nombre de philosophes, d'historiens d'art ou d'artistes, associés à des neuro-biologistes, se sont regroupés dans un volet nouveau de l'esthétique, qu'ils ont appelé la **neuro-esthétique**. La neuro-esthétique a suscité une très abondante littérature depuis 30 ans, parmi laquelle les travaux séminaux de [Zeki, 1999, Livingstone, 2002, Kawabata and Zeki, 2004], ceux plus récents de [Cupchnik et al., 2009, Di Dio and Vittorio, 2009] ou ceux de [Di Dio et al., 2011, Ishizu and Zeki, 2011, Kuehn and Gallinat, 2012]. Ces travaux feront l'objet des développements de la section 3.

Ces résultats ont durablement ancré l'interprétation de la Beauté dans un contexte subjectiviste. Ils soutiennent des conclusions qui, dans leurs formes extrêmes, suggèrent que :

- la Beauté est une émotion personnelle, c'est-à-dire une expérience psycho-physiologique complexe en réaction aux influences biochimiques et environnementales transmises par nos sens ;
- le jugement esthétique (*esthetic appraisal*) est distinct des autres formes de jugement fournies par notre cortex (par exemple le jugement sur la nourriture) car la beauté n'est pas vitale pour l'être humain (à l'inverse par exemple de la nourriture) et nous n'avons aucun "intérêt" physiologique à l'exprimer⁴⁴ [Beardsley, 1966] ;
- en cela, la matière artistique pourrait être traitée par des circuits différents et spécialisés auxquels est parfois donné le nom "d'aires hédoniques"⁴⁵.

2.3 Vers une synthèse des démarches objectivistes et subjectivistes ?

2.3.1 Un débat toujours vivant

Les travaux entrepris depuis un siècle dans le champ de la psychologie expérimentale et les résultats les plus récents des méthodes d'investigation en neuro-biologie ont bien établi l'importance des phénomènes subjectifs d'appréciation des qualités esthétiques d'une image. Mais par ailleurs, de nombreux témoignages attestent de l'existence de consensus objectifs sur la beauté de certains objets, ce consensus indiquant que des "indices de beauté" émanent de cet objet qui sont très largement reconnus et actifs sur une large part de la population.

Il semble aujourd'hui assez largement partagé que les objets reconnus comme "beaux" possèdent des propriétés intrinsèques qui favorisent cette reconnaissance pour tout observateur (voir par exemple la figure 2.2), mais que l'observateur reste celui qui décide *in fine* de ce qu'il juge beau ou non. On remet ainsi au centre de l'interprétation du Beau des visions que l'on a déjà rencontrées chez de Crousaz ou chez Kant qui tentent d'accomoder le subjectivisme qui triomphe alors et l'objectivisme de longue tradition⁴⁶.

44. Cette remarque laisse la place à une explication où la Beauté n'est pas d'essence matérielle mais relève d'une inspiration supérieure.

45. Ainsi, F. Hedgeson considère t'il que le Beau est ce qui est saisi par notre sens interne du beau, comme le visible est ce qui est saisi par notre œil. Il attribue ainsi à l'Homme des organes de perception hédonique que de nombreux travaux ultérieurs chercheront à localiser (voir section 3.2).

46. Il serait erroné de penser qu'une approche conciliant objectivisme et subjectivisme émerge au temps des Lumières. Elle toujours été présente dans les débats sur l'esthétique. Citons l'exemple de Guillaume d'Auvergne au début du *XIII^e* siècle. Il écrit successivement : "*Hoc indubitantur essentialiter pulchrum est et decorum, et vjus essentia et quidditas ejus est pulchritudo.*" (En ce qui est indubitablement essentiellement beau et élégant, l'essence profonde et la qualité est la beauté.) Du bien et du Mal, 206, (Pouillon, 317), et *Ipsum quod est in se decorum, est turpe ad aliquid, aut cum alio, aut in alio.*" (Une chose belle en elle-même peut être laide près d'un autre objet, ou avec un autre objet ou dans un autre objet.), Du bien et du Mal, 207, (Pouillon, 317), cités par [Tatarkiewicz, 1970].

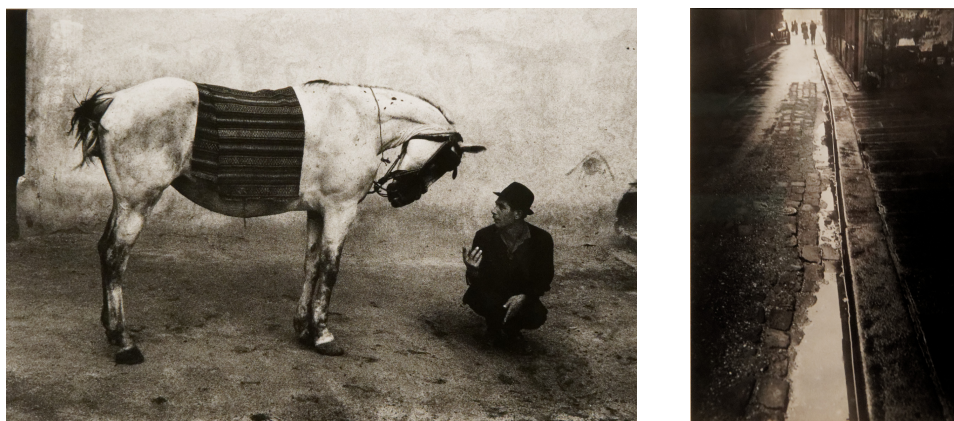


Figure 2.1 – Pourquoi une photo est-elle belle ? On pourrait dire de la photo de gauche (Romania de Josef Koudelka) que sa beauté est issue de son objet ; ici le cheval, lui-même est beau. Mais outre que cette remarque ne fait que déplacer notre interrogation sur les attributs de la beauté, elle ne supprime pas l'essence esthétique de la photo qui s'impose nettement. Sur la photo de droite (Rive gauche de Gordon Parks), l'argument précédent ne tient pas car l'objet lui-même (une ruelle à contre-jour) ne tire ses qualités esthétiques que de l'image qui en est faite, de sa composition et de ses jeux de lumières.

Examinons par exemple la conception kantienne, issue de l'expérience esthétique, qui considère le jugement esthétique comme à la fois universel et subjectif. Kant, pour mobiliser l'universel, a recours à ce qu'il appelle le "sens commun" (*sensus communis*), un jugement partagé par tous, non sur des fondements rationnels comme le sens commun de Descartes, non sur des critères de majorité d'avis comme dans l'interprétation des stoïciens, mais par une évidence spontanée, universelle et transcendentale qui le rend perceptible immédiatement et par tous⁴⁷. Nous voyons un objet beau, non seulement parce qu'il est beau pour nous, mais aussi parce que nous considérons qu'il doit être beau par tout autre⁴⁸.

2.3.2 L'esthétique interactionniste

La démarche de Kant a été fortement revisitée et son recours à la transcendance du jugement esthétique par le sens commun est généralement abandonné pour faire place à des interprétations où le beau, porté par des indices visuels objectifs, évoque une appréciation hédonique subjective par une adaptation particulière des signaux transmis par les voies primaires aux aires supérieures [Reber et al., 2004, Ingarden, 2011]. Cette approche se reconnaît sous le terme d'interactionniste. Elle porte en elle aujourd'hui une bonne part des éléments du débat sur l'explication du sentiment de beauté.

Pour elle, le plaisir hédoniste ressenti par l'observateur naît de la facilité avec laquelle l'observateur perçoit le message contenu dans les formes convoyées par l'objet, dans une interaction phénoménologique [Merleau-Ponty, 1945, Dufrenne, 1967].

Les indices de beauté seront alors plus ou moins efficaces pour créer l'émotion et le plaisir qui accompagnent le sentiment de beauté. Cette efficacité dépend tout d'abord de la force des signaux (et l'on évoque ici tous les arguments platoniciens concernant en particulier la simplicité [Berlyne, 1971],

47. On lira avec intérêt sur ce sujet le blog de S. Manon, <http://www.philolog.fr/le-sens-commun/>.

48. " (...) sous cette expression de *sensus communis* on doit comprendre l'Idée d'un sens commun à tous, c'est-à-dire d'une faculté de juger, qui, dans sa réflexion, tient compte en pensant (a priori) du mode de représentation de tout autre homme, afin de rattacher son jugement à la raison humaine tout entière et échapper, ce faisant, à l'illusion, résultant de conditions subjectives et particulières pouvant aisément être tenues pour objectives, qui exerceraient une influence néfaste sur le jugement. [...] on y parvient en écartant autant que possible ce qui dans l'état représentatif est matière, c'est-à-dire sensation, et en prêtant uniquement attention aux caractéristiques formelles de sa représentation ou de son état représentatif. " E. Kant, *Critique de la faculté de juger*, paragraphe 40, (Traduction de A. Philonenko).

mais aussi les propositions de la Gestalt [Reber et al., 2004, Moshagen and Thielsch, 2010] recouvertes du terme de *Prägnanz*⁴⁹. Mais elle dépend tout autant d'un grand nombre de paramètres internes à l'observateur, certains relevant de son humeur, d'autres de son tempérament, de sa culture, de son entraînement, d'autres enfin du contexte, tous facteurs qui agissent concurremment sur ses émotions par des effets de masquage ou au contraire d'amplification.

2.3.3 La mise en place d'une esthétique interactionniste



Figure 2.2 – *Ces photos sont-elles belles ? C'est à cette question qu'essaie de répondre ce texte. Plaçant un vaste jury devant cette interrogation, un certain consensus se manifeste souvent, qui conforte la démarche de ceux qui pensent mesurer cette beauté par ordinateur. Si tout le monde perçoit dans les deux images de gauche des critères de beauté plus évidents que dans l'image de droite, généralement perçue comme ordinaire, pourquoi l'ordinateur ne les détecterait-il pas ? La photo de gauche (Poppy d'Irving Penn) est signée d'un photographe célèbre et l'on pourrait objecter qu'elle bénéficie de cette référence (c'est l'effet de "signature" analysé par exemple dans [Changeux, 2016]), mais la photo du centre est anonyme (issue du web) et n'est appréciée que par ses qualités propres. Par ailleurs ces trois photos représentent des fleurs et héritent donc du même contexte culturel et sémantique pour la plupart des observateurs. Enfin, elles ne font appel à aucun des ressorts qui mettent en jeu de l'intérêt ou de la surprise, autres ingrédients potentiels de différenciation comme nous l'avons vu en section 1.1.*

Comment sont transmis à notre conscience ces indices, les "propriétés" intrinsèques du bel objet ? Bien sûr ce sera avant tout par nos organes de perception et donc, pour une image, par la vue. Il est donc judicieux de confier au traitement d'images le soin de détecter ces indices s'ils existent. Comment, une fois détectés, sont-ils exprimés ? Cela est beaucoup plus complexe. S'il ne faut pas négliger les étapes de traitement des images par les aires visuelles primaires, il faut aussi accorder beaucoup d'attention aux traitements postérieurs jusqu'aux aires qui président à la décision, sans négliger l'importance des circuits latéraux, en particulier ceux de la mémoire, éventuellement stimulés par les signaux visuels, mais contribuant à l'aide d'apports cognitifs à la sensation finale.

On comprend que certaines étapes du traitement de l'information par le cerveau (et le système visuel prend toute sa place comme composante du système cérébral) seront amenées à réaliser des opérations automatiques, tandis que d'autres mettront en jeu des actions réfléchies. Peu d'auteurs se sont attachés à proposer pour ces traitements un organigramme complet du traitement esthétique. La figure 2.3, issue de [Leder et al., 2004], en propose un, illustrant une modélisation psychologique du processus d'évaluation esthétique d'une œuvre d'art. L'évaluation esthétique d'une œuvre d'art fait beaucoup plus appel à des tâches cognitives (reconnaissance du style, reconnaissance de l'artiste) que l'évaluation de photographies, mais elle nous propose tout de même un modèle intéressant pour notre propos. Selon H. Leder *et al.*, elle se diviserait en cinq étapes : les deux premières concernent essentiellement les voies visuelles et sont surtout réflexes (pour nous donc vraiment objectivistes), les trois suivantes sont en charge de la reconnaissance du contenu, de l'interprétation globale et de l'évaluation ; celles-ci sont cognitives : elles utilisent notre expertise du domaine, nos centres d'intérêt, nos goûts,

49. *Prägnanz* est le terme choisi par M. Wertheimer pour exprimer la facilité de perception d'une figure en raison des relations particulières entre ses composantes : symétrie, parallélisme, proximité, etc. [Wertheimer, 1922]

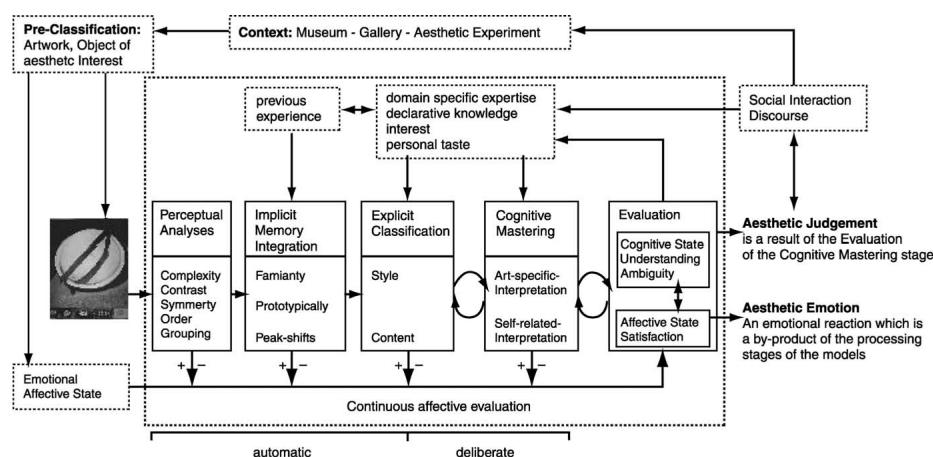


Figure 2.3 – Un modèle de la perception esthétique d’après [Leder et al., 2004]. On distingue dans ce modèle deux sorties distinctes : l’une correspond à l’évaluation esthétique, l’autre à l’émotion esthétique. Selon les auteurs, les deux opérations sont élaborées en parallèle, suivant des circuits distincts et ne sont donc pas réductibles l’une à l’autre. Ils considèrent cependant que le plaisir esthétique (au cœur de l’émotion) est d’autant plus important que la compréhension de l’œuvre (un des points majeurs de l’évaluation) est plus forte. L’étude est conduite sur des œuvres de l’art contemporain et il n’est pas évident que les mêmes arguments s’appliquent aux photographies.

mais aussi les éléments de contexte qui accompagnent l’expérience et qui sont fort importants pour estimer une œuvre d’art (l’ambiance d’un musée ou l’effet “signature” déjà évoqué). Celles-là seraient alors plutôt subjectivistes. Les grandes lignes du schéma de la figure 2.3 pourraient certainement inspirer un schéma à peine différent qui serait adapté à la photographie.

L’objet du prochain chapitre est de faire le point sur ce que l’on sait aujourd’hui de ces traitements à l’aide des outils de la neuro-biologie en essayant de vérifier si certaines aires cérébrales auxquelles on attribue certaines fonctions précises sont sollicitées au cours de l’analyse d’une photographie. Notons que le modèle présenté sur la figure 2.3 a été élaboré avant que ne soit disponible la plupart des résultats qui seront présentés dans ce chapitre.

2.3.4 La diversité des esthétiques

Mais le XX^e siècle, puis le XXI^e siècle ont vu de nombreuses écoles de pensée s’attacher à déterminer des facteurs particulièrement importants lors de l’évaluation esthétique, ces facteurs s’écartant souvent très largement de l’antagonisme objectiviste/subjectiviste qui portera la suite de cet ouvrage. Nous ne pouvons pas les ignorer complètement à ce point et nous les évoquons brièvement ici.

Ainsi, nous avons cité l’esthétique réaliste de E. Siegel, défendue par E. Zemach, ainsi que l’esthétique empirique de D.E. Berlyne (voir la section 2.1). Très importante aussi a été l’esthétique fonctionnelle qui rattache la beauté à la capacité de l’objet à remplir un rôle dans la société. Elle puise ses sources très loin dans la pensée antique des Stoïciens qui soulignaient l’importance de la fonctionnalité dans la perception du beau. Elle s’est trouvée au centre de la vie intellectuelle (Blau Reiter, Suprématisme, ...) avec une forte synergie avec le monde moderne naissant, l’avènement d’une architecture et d’une industrie fortement outillées au sein de la société du XX^e siècle, donnant naissance en particulier au design industriel (H. van de Velde, W. Gropius).

Une autre famille importante de travaux concerne l’esthétique dite contextualisée, ou historique,

illustrée par exemple par E.H. Gombrich et A.C. Danto⁵⁰. Dans l'esthétique historique, on considère qu'une grande part des qualités artistiques de l'œuvre provient d'une connaissance du contexte de sa création permettant à celui qui connaît ce contexte "historique" d'apprécier l'intention artistique d'où découlerait le jugement de beau, beaucoup plus qu'au naïf. Cette hypothèse nous renvoie, dans une certaine mesure, aux remarques faites au paragraphe 1.1 sur le rôle de la "pertinence" dans notre intérêt pour un document. Dans [Bullot and Reber, 2013], une synthèse est tentée entre l'interprétation contextualisée et une approche purement subjective (celle de la neuro-esthétique), synthèse baptisée esthétique psycho-historique

Mais d'autres démarches plus radicales encore situent l'esthétique dans le champ de la sémiotique (la science des symboles et de leur signifié) relevant exclusivement de l'activité cognitive⁵¹, (N. Goodman, G. Genette).

50. L'esthétique historique concerne avant tout l'art puisqu'elle accorde une grande place à l'acte de création et à l'auteur ; elle s'adresse à ce titre à la photo d'art à laquelle s'appliquent les principaux arguments développés par ses défenseurs.

51. "... les qualités d'une œuvre d'art, celles de notre appréhension, sont des produits de l'expérience, de la capacité à voir, à appliquer des catégories, des produits de la compétence." (N. Goodman [Goodman, 1990]) et "*Ce n'est pas l'objet qui rend la relation esthétique, mais la relation qui rend l'objet esthétique*" (G. Genette [Genette, 1997]), cités par R. Pouivet [Pouivet, 2014].)

Chapitre 3

La neuro-biologie ou l'arbitre des consciences

Les études qui ont permis de connaître les mécanismes de la vision humaine sont multiples. Les études cliniques portant sur les blessés lors d'interventions chirurgicales, complétées parfois des études anatomiques *post mortem*, ont ouvert la voie permettant de localiser des régions importantes pour certaines étapes de perception, d'analyse ou d'interprétation du signal visuel. Puis des travaux d'électrophysiologie sur l'homme (électro-rétinographie, électro-oculométrie, potentiels évoqués), mais surtout ceux conduits par des techniques plus invasives sur les oiseaux, les chats et les singes, ont permis de vérifier *in vivo* la validité des modèles de fonctionnement des voies optiques et des aires primaires de la vision.

Il a fallu cependant attendre la disponibilité des techniques d'électro-encéphalographie (EEG), de magnéto-encéphalographie (MEG), puis de l'imagerie par résonance magnétique (IRM) et de la tomographie par émission de positons (TEP), pour accéder au fonctionnement des aires supérieures de la vision, en temps (presque) réel.

3.1 Les protocoles d'IRMf et la neuro-esthétique

Ce sont surtout les travaux menés en résonance magnétique fonctionnelle (IRMf) qui ont permis de faire les progrès les plus décisifs dans ce domaine. Rappelons que cette technique, mise en place dans les années 1990, repose sur la détection de la consommation d'oxygène (que l'on rattache à l'activité circulatoire) par les zones cérébrales actives. Cette consommation est faible et se traduit par une très faible variation du signal détecté par IRM (de l'ordre du pourcent). Pour cela, les protocoles employés nécessitent que le sujet examiné se livre plusieurs fois de suite au même exercice, de façon que l'on puisse intégrer plusieurs mesures. Comme il faut aussi que ces protocoles permettent d'isoler les signaux recherchés, il convient d'alterner les étapes de détection avec des étapes conçues pour identifier la consommation d'oxygène au repos, ou celle due à d'autres activités cérébrales du sujet, indépendantes du stimulus de test. La réponse recherchée est alors obtenue par différence entre ces deux séries d'expériences. Voyons sur l'exemple d'une expérience fictive comment fonctionnent ces protocoles. Supposons que l'on cherche à identifier si une couleur, le rose, dispose d'un circuit particulier de traitement dans notre cerveau. On produit une série d'images très variées, certaines contenant des plages roses, d'autres n'en contenant pas. On les présente successivement, dans un ordre aléatoire, pendant quelques secondes, à un observateur placé dans un imageur par résonance magnétique. S'il existe dans le cerveau un "processeur de rose", alors on s'attend à voir apparaître une zone active dans toutes les images IRMf produites lors de la présentation d'une image contenant du "rose". Ce processeur sera spécifique du rose si la même expérience répétée avec les plages roses converties en violet "n'allument" pas les zones précédemment actives.

Pour généraliser ces résultats, il convient de répéter chaque expérience sur plusieurs sujets pour s'assurer d'un consensus sur la réponse, puis de moyenner les réponses obtenues. Mais ces moyennes devant prendre en compte la variabilité des morphologies cérébrales, on procède à un recalage de chaque carte cérébrale dans un repère de référence (par exemple le repère de Talairach), avant de procéder au moyennage.

Ces protocoles complexes sont bien sûr l'objet de nombreuses discussions et critiques, et tous les résultats doivent être examinés avec une grande circonspection. Cela explique que de nombreuses études aient été menées avec des objectifs assez proches mais des protocoles différents. C'est en particulier le cas des études qui s'attachent à suivre les processus cérébraux associés à des signaux en fonction de leurs qualités esthétiques. Ces travaux ont été regroupés depuis une dizaine d'années dans une branche des sciences spécifique, dénommée "neuro-esthétique" [Skov, 2009, Vidal, 2011] et dédiée spécifiquement aux bases neurales de la beauté dans les diverses modalités de la perception : la vision bien sûr (avec la peinture, la photographie, l'architecture, la statuaire), mais aussi l'audition (avec la musique et le chant) et aussi, dans de très rares cas, le goût et l'olfaction. Un objectif important de la neuro-esthétique pour de nombreux auteurs a été de "naturaliser" les circuits d'évaluation de la beauté, c'est-à-dire de montrer qu'ils utilisent des fonctions du cortex déjà établies dans le cerveau pour d'autres fonctionnalités et en particulier des fonctionnalités cruciales pour la survie de l'espèce (ce que les jugements esthétiques ne sont pas) [Brown et al., 2011], ces fonctionnalités cruciales pouvant justifier l'existence de ces structures dans l'espèce dans un cadre darwinien de l'évolution.

3.2 Que nous dit l'IRMf des "processeurs de beauté" du cerveau ?

[Que dit l'IRMf ?]

Nous ne saurions évoquer ici tous les travaux consacrés à la neuro-esthétique. Une revue de 2009, [Di Dio and Vittorio, 2009], en citait une trentaine, une autre [Chatterjee and Vartanian, 2016] près de 200, en 2016, mais une recherche en 2017 sur le mot *neuroaesthetics* en rapporte près de 3000. Nous ne visiterons pas ces travaux séparément. Nous isolerons les principales zones cérébrales concernées en renvoyant le lecteur à quelques articles qui expliquent comment elles ont été mises en évidence.

3.2.1 Le rôle du cortex préfrontal

Le cortex préfrontal (en couleur sur la figure 3.2) a été le premier candidat à la fonction de "centre hédonique" dans des travaux qui ont souligné très tôt son implication lors de l'observation d'œuvres d'art [Kawabata and Zeki, 2004, Jacobsen et al., 2006, Cela-Conde et al., 2011]. S'il s'est vu adjoindre ensuite d'autres aires très importantes, il demeure l'une des zones maîtresses de la vision esthétique. L'OFC (cortex orbito-frontal) est connu pour son rôle important dans les tâches de décision. C'est un point de passage de la plupart des voies perceptuelles et un élément important du système en charge de l'évaluation (*appraisal*) [Kringelbach, 2005, Wallis, 2007] ainsi que de la régulation des "récompenses" (*rewards*). A ces titres, il pourrait être une pièce centrale de la fonction de plaisir [Berridge and Kringelbach, 2013].

Nous illustrerons ce rôle par les travaux d'Ishizu et Zeki [Ishizu and Zeki, 2011] qui constituent un socle pour les études de neuro-esthétique et ont permis à leurs auteurs de proposer une "théorie de la beauté" qui sert souvent de référence. Dans leurs expériences, des sujets sont exposés dans un imageur IRMf à une série de stimulus (des photos et des musiques), classés préalablement en "beaux", "indifférents" ou "laid". Le dépouillement consiste alors à comparer les zones activées selon la "qualité" du stimulus.

Les résultats ont montré une activité systématique et préférentielle de la partie centrale du cortex orbito-frontal (OFC) chaque fois qu'un document "beau" est présenté plutôt qu'un document "indifférent" ou "laid" (figure 3.1). Ces résultats, déjà cités par ailleurs, ont été régulièrement retrouvés

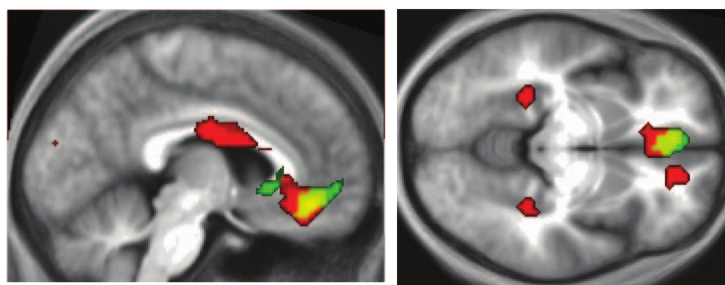


Figure 3.1 – Dans l’expérience en référence [Ishizu and Zeki, 2011], Ishizu et Zeki ont montré que seule la partie centrale du cortex orbito-frontal (la tache à droite dans l’image gauche) était systématiquement activée lors d’expériences avec des signaux visuels ou acoustiques porteurs de qualités esthétiques. Les noyaux caudés peuvent également être actifs dans certaines expériences. Expériences conduites avec 21 sujets. Les zones en rouge sont excitées par un signal visuel, les zones en vert par un signal acoustique tandis que les zones en jaune sont excitées par les deux types de signaux.

(voir la synthèse de [Brown et al., 2011]).

Par ailleurs, pour autant qu’il est possible de le déduire d’un petit nombre d’expériences, l’ampleur de l’activation est directement fonction de la beauté du document telle qu’elle a été appréciée lors de la phase préalable de calibrage.

Ces expériences semblent donc ancrer biologiquement les notions d’esthétique et de plaisir, un point régulièrement abordé par la philosophie du Beau.

Enfin, il apparaît qu’aucune activation significative ne permet de distinguer la “Laidéur”. Ce point tente de répondre à une question importante en philosophie de l’esthétique concernant l’existence ou non d’une valence négative à la Beauté¹.

L’étude citée identifie d’autres régions également activées, en particulier les noyaux caudés (dont le rôle pour réguler les émotions est connu) ou le gyrus supra-marginal (impliqué dans la perception de l’espace), mais de façon non systématique. Nous les retrouverons par ailleurs.

3.2.2 Le rôle du cortex insulaire

Les travaux de S. Brown *et al.* [Brown et al., 2011] ont fait la synthèse des études disponibles en neuro-esthétique en 2011 de façon à retrouver, si elles existent, les aires communes à l’évaluation esthétique à travers les diverses modalités de perception (vision, audition, goût et olfaction). C’est donc là aussi la recherche de la “fonction hédonique” qui est en jeu, mais que l’on cherche à “factoriser” pour toutes les modalités. La synthèse a porté sur 93 études de neuro-imagerie (IRMF ou TEP exclusivement) concernées par l’étude des zones du cerveau actives lors de l’exposition à des stimulus esthétiques positifs dans les quatre domaines de perception. Les articles concernant la vision étaient nettement majoritaires et nous insisterons ici sur les résultats qui les concernent. Toutes ces études portaient sur des personnes saines et fournissaient des résultats pouvant être rapportés dans un même atlas de référence (Talairach ou MNI)².

1. Beaucoup de traités de l’esthétique ne reconnaissent pas de valence négative à la Beauté, tandis que le sens commun accepterait la candidature de la “Laidéur” dans ce rôle. Mais il faudrait alors admettre des attributs de “Laidéur”, ce qui serait distinct de l’absence de critères de “Beauté”, position la plus communément adoptée. Notons donc ici que la neuro-biologie mettant en évidence des zones d’évaluation aussi sensibles aux valences positives que négatives (notamment le cortex orbito-frontal) tente d’apporter des réponses à cette question.

2. En raison de la diversité géométrique des anatomies cérébrales, il n’est pas possible de superposer les aires fonctionnellement identiques d’individus différents sans procéder à des transformations géométriques. L’atlas de Talairach est la référence la plus ancienne [Talairach and Tournoux, 1988] ; elle est très adaptée pour les structures superficielles. On lui préfère parfois l’atlas MNI (du nom du Montreal Neurological Institute de l’université McGill) plus représentatif de la variété des populations et plus adapté aux structures profondes [Evans et al., 1992].

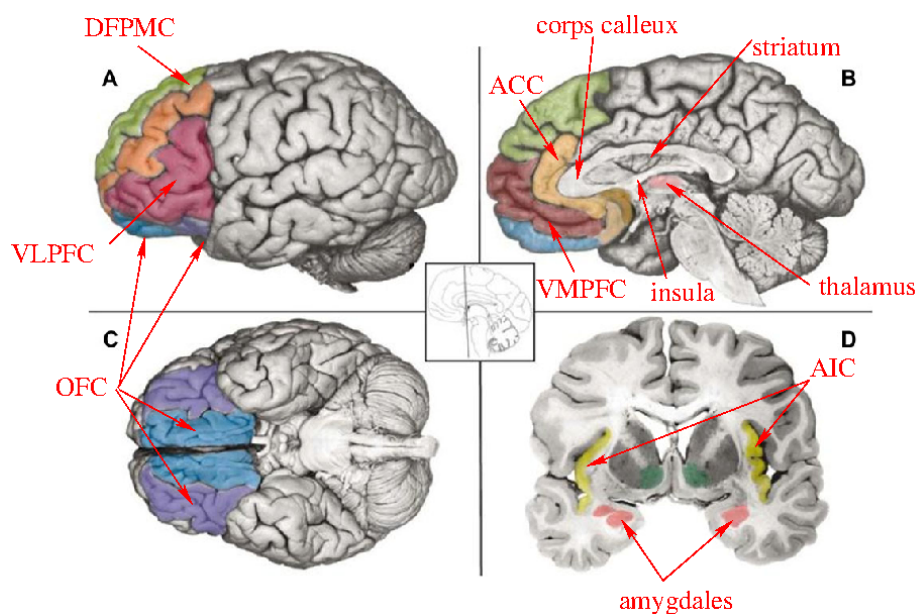


Figure 3.2 – Les aires concernées lors de l’expression d’émotions, d’après [Chatterjee and Vartanian, 2016]. Le cortex orbito-frontal (OFC) est la zone en violet des vues A et C. Le cortex insulaire (AIC) est en jaune sur la vue D. Il est relié de façon très étroite aux aires latérales de l’OFC. Les circuits viscéro-moteurs comportent la zone ventrale du cortex ventro-médian préfrontal (VMPMC) (en bleu sur A, B et C). Le VMPMC est étroitement lié à l’amygdale (en rose dans la vue D). Le cortex cingulaire antérieur (ACC) est en beige sur B, en avant du corps calleux. La zone dorsale du cortex préfrontal est associée au contrôle des états mentaux : partie dorsale du VMPMC, zone frontale (en marron), cortex dorso-médian préfrontal (DMPFC) (en vert en A et B). Le cortex ventro-latéral préfrontal (VLPFC) est en rouge sur la vue A. Le thalamus est en rose au centre de la vue B et le gyrus frontal moyen est en orange.

L’étude a consisté à examiner la conjonction des activations entre auteurs et entre modalités dans l’ensemble des études, à l’aide de protocoles bien établis³.

Bien que l’on retrouve le rôle important de l’OFC dans la plupart des observations (figure 3.3), il n’est pas engagé dans le jugement esthétique en olfaction ou en goût et, pour les auteurs de [Brown et al., 2011], il n’est donc pas systématiquement associé au jugement esthétique. Les auteurs ne lui attribuent donc pas la primauté de la fonction hédonique.

Cette étude conclut au contraire à l’importance première du lobe droit du cortex insulaire antérieur, siège d’une très importante activation lors d’un stimulus esthétique quelle que soit la modalité (audition, vision, goût, olfaction). Ce serait la seule aire spécifiquement active pour toutes les modalités. Seule cette aire serait systématiquement activée par les stimulus esthétiques positifs et il semblerait que ce soit là que se situent les mécanismes les plus spécifiques liés à l’appréciation esthétique.

On retrouve des conclusions concordantes dans [Cupchnik et al., 2009] (figure 3.4).

Cette aire est cependant activée également dans une grande variété de situations où la dimension esthétique n’est pas engagée et son rôle exact est encore très confus. Elle est en particulier réputée pour être active dans les situations de jugement négatif (dégoût, douleur, empathie, etc.), ce qui nous éloigne sensiblement de nos conclusions. Elle participe aussi à la conscience introspective (maintien de l’état du corps humain, homéostasie, conscience de soi, etc.) activités qui ne nous rapprochent pas beaucoup du jugement esthétique. Enfin elle semble avoir un rôle régulateur dans l’expression des émotions par l’amygdale, ce qui est probablement plus proche de notre sujet.

Dans les travaux de [Lacey et al., 2011], on précise l’importance des zones en charge des ”récompenses” : striatum ventral, hypothalamus et OFC lors de l’examen d’œuvres d’art, et l’on propose quelques

3. Le protocole choisi est GingerALE : ALE=Activation Likelihood Estimation.

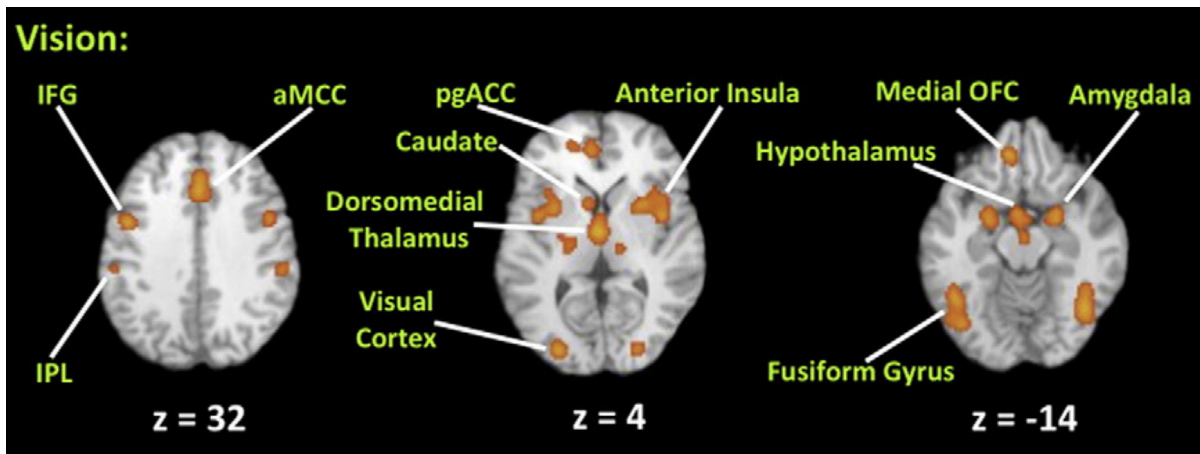


Figure 3.3 – Les principaux foyers d’activité de la vision lors des expériences rapportées dans [Brown et al., 2011]. Le repère choisi est celui de Talairach. La présentation est faite dans des coupes axiales dont la coordonnée verticale est indiquée par z . Les zones actives ont une probabilité seuillée à 5%. Les abréviations signifient : IFG = gyrus frontal inférieur, IPL = lobe pariétal inférieur, aMCC = cortex cingulaire médian antérieur, pgACC = cortex cingulaire prégenouillé, OFC = cortex orbito-frontal. Caudate désigne le noyau caudé.

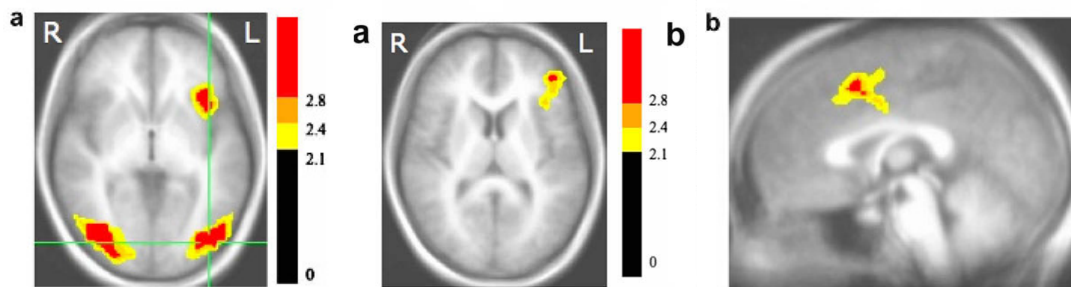


Figure 3.4 – L’expérience conduite dans [Cupchik et al., 2009] offre des résultats d’expérimentation avec trois types de stimulus différents : i) un champ visuel quasiment vide (fond uniforme marqué de 4 croix), ii) des images quelconques et iii) des reproductions de peintures. On voit à gauche les aires activées spécifiquement en contrastant les images (ordinaires et artistiques réunies) par rapport au champ vide : on distingue, dans la partie inférieure du cerveau, l’activation des aires visuelles tertiaires (gyrus fusiforme inférieur), et dans la partie supérieure le cortex insulaire antérieur. Dans les deux images de droite, on distingue les aires cérébrales activées de façon discriminante lors de l’observation d’images “esthétiques” versus d’images ordinaires.

schémas de connectivité entre ces zones.

Selon [Cupchik et al., 2009], la coordination entre OFC et AIC (cortex insulaire antérieur) serait alors complexe, l’OFC, en charge des “récompenses” liées à la perception (en particulier visuelle), garderait la “mémoire” de ces récompenses au cours de l’expérience de façon à maintenir l’attractivité esthétique à travers l’AIC jusqu’à la prise de décision.

3.2.3 Le rôle des aires visuelles

Ces études ne passent pas sous silence l’activité des aires supérieures visuelles et l’on dispose de très nombreux travaux pour attester de leur engagement dans la vision des œuvres d’art (par exemple sur la figure 3.4 à gauche on a noté l’engagement des aires visuelles tertiaires).

Dans [Di Dio et al., 2007], un observateur était placé dans les conditions d’observation de statues dans un musée. Ces statues avaient soit les proportions d’origine, soit des proportions déformées.

L'étude a mis en évidence le rôle des aires visuelles : les aires temporales ainsi que le cortex occipital latéral, l'un et l'autre connus pour traiter des images du corps humain. Dans [Cupchik et al., 2009], une comparaison a été faite entre l'observation des peintures à contours nets de celle des peintures à contours doux. La motivation de cette étude était de mettre en évidence le possible rôle de l'attention active dans l'interprétation de l'œuvre d'art, puisque l'information de contour, importante pour la reconnaissance des formes, n'est pas disponible à la sortie des voies optiques en l'absence de bords nets. L'étude montre alors une distinction au niveau du lobule pariétal supérieur, aire en charge, entre autres, de l'orientation spatiale, indiquant, conformément à l'hypothèse des auteurs, que le cortex est bien engagé dans une tâche de complétion du message pictoral si l'image source ne porte pas toutes les informations (figure 3.5).

Dans [Vartanian and Skov, 2014] une synthèse de 15 études confirme l'enrôlement systématique des lobes antérieurs temporaux, des gyrus fusiformes et du gyrus parahippocampique qui joue un rôle dans la reconnaissance de l'environnement.



Figure 3.5 – Dans cette expérience issue de [Cupchik et al., 2009], on distingue les réponses des aires spécifiquement activées lors de la vision d'images artistiques “à bords doux” (dont un exemple est au centre) versus d'images à “bord raides” (dont un exemple est à droite). C'est le lobule pariétal supérieur qui est préférentiellement actif dans le premier cas. Cette aire est en charge de fonctions de localisation spatiale dans l'interprétation de scènes.

3.2.4 Le rôle de la mémoire et de la cognition

Ces mêmes études concluent souvent à l'activation d'aires impliquées dans la mémorisation. Ainsi, dans l'étude déjà citée [Di Dio et al., 2007], les auteurs identifient l'action des aires pariétales médianes et du lobe préfrontal, zones impliquées dans la fonction de mémoire. Cette expérience plaide donc aussi pour un enrôlement des fonctions cognitives lors du traitement des statues de référence. De plus, l'examen des enregistrements en présence des objets “beaux” a montré une suractivation de l'amygdale, région aux rôles multiples mais souvent engagée pour son rôle de valuation positive ou négative lors des épreuves d'apprentissage et pour son rôle dans l'expression des sentiments.

De même, la méta-étude de [Vartanian and Skov, 2014] conclut à l'activation très générale du lobe antérieur temporal dont le rôle a été identifié comme important pour la mémorisation sémantique mais aussi pour l'intégration conceptuelle lors de la reconnaissance des formes. De leurs observations, les auteurs de [Di Dio et al., 2007] déduisent que l'interprétation des œuvres d'art est d'une part “poussée par les données” (rôle des voies optiques), d'autre part “tirée par les émotions” (rôle de l'AFC et de l'AIC).

3.2.5 Le rôle de l'*embodiment*

Si les travaux précédents plaident pour la prise en compte de la mémoire, de l'expérience et *in fine* de la cognition dans l'évaluation esthétique, des travaux ont au contraire suggéré la mise en œuvre de processus très différents.

C'est le cas des travaux présentés dans [Freedberg and Gallese, 2007]. Les auteurs partent de l'hypothèse que des œuvres de contenus très différents peuvent mobiliser des aires cérébrales très différentes. Cette approche semble particulièrement adaptée à la photographie qui reproduit des scènes réelles, au risque de mêler des phénomènes que nous avons tenté de distinguer en section 1.1.

Les auteurs montrent effectivement la variété des aires concernées, souvent liées à la cognition ou à la mémorisation (comme ci-dessus), mais de façon plus originale ils trouvent que l'observation de l'œuvre peut affecter des sites qui, chez l'observateur, feraient écho à ceux qui ont pu être touchés par l'artiste soit lors de son inspiration soit lors de la réalisation de l'œuvre⁴. La réponse envoyée par ces sites va contribuer activement à l'élaboration de notre émotion, suivant une démarche connue sous le nom d'*embodiment*, d'incarnation ou de personnification.

Cette interprétation s'appuie sur la théorie encore débattue des neurones miroirs qui permettent de maintenir en soi une représentation de son propre corps et, par transition, du monde, aboutissant au développement de l'empathie classiquement exprimée lors de l'observation de ces œuvres. Ainsi les aires de notre cortex ventral prémoteur en charge des membres seraient sollicitées en observant une photographie d'athlète et interviendraient alors dans notre sensation autant, et peut-être plus, que les zones en charge des connaissances ou des souvenirs qui sont aussi invoquées (lobes temporaux, amygdale et hippocampe). Cette interprétation, testée sur des œuvres de Goya, Michelange et Le Caravage, a également été vérifiée sur des œuvres abstraites (Pollock, Fontana). Elle a été soutenue expérimentalement dans [Di Dio et al., 2007] et plus récemment dans [Ticini et al., 2014], et corroborerait des propositions avancées par ailleurs sur le fonctionnement de notre jugement [Danmasio, 1994].

3.2.6 Quel schéma d'ensemble ?

On ne peut manquer de garder une vision un peu confuse de ces travaux assez complexes : [Cela-Conde et al., 2011]. Essayons ici de voir en conclusion les principales fonctions sollicitées.

- les aires visuelles, zones occipitales, zones latérales inférieures, cortex insulaire, lobule pariétal supérieur sont bien sûr actives par la tâche même de vision, mais aussi par le contenu selon des aires spécifiques : formes, couleurs, mouvements, visages ;
- le cortex orbito-frontal en charge de l'évaluation, régulateur des "récompenses" et du plaisir, est une pièce importante qui semble contrôler notre décision ;
- le cortex insulaire, qui contrôle nos émotions, est lui aussi une pièce importante et quasiment toujours impliquée lors de l'observation d'une œuvre d'art ;
- les zones engagées dans les opérations de cognition (amygdale) et de mémoire (aires pariétales médianes, lobe préfrontal) sont sollicitées ;
- des aires en charge du contrôle prémoteur sont convoquées de façon spécifique dans des situations de forte empathie et d'*embodiment* (cortex ventral prémoteur, lobes temporaux, hippocampe).

On trouve dans [Brown et al., 2011] et dans [Vartanian and Skov, 2014] des schémas qui tentent de figurer le réseau des activations lors d'une exposition à une œuvre d'art, mettant en relation les trois grandes fonctionnalités : cognition, émotion, contrôle sensori-moteur. Ils sont encore très spéculatifs. Par ailleurs, J.P. Changeux propose un schéma assez complet de la progression des signaux depuis l'entrée dans le système visuel jusqu'à l'accès à la conscience "*au-delà des cortex inféro-temporal et pariétal*" ([Changeux, 2016], pages 166-173). Ce schéma représente bien les aiguillages qui orientent les diverses composantes du signal : couleur, contours, fond/forme, visages . . . vers des zones spécialisées, puis la mobilisation des aires en charge de la mémoire, du signifié, des émotions . . . , et couvre bien les fonctions d'observation consciente. Il est moins précis sur l'accès à "*l'espace de travail neuronal conscient qui s'appuie sur un réseau de neurones, dont les axones à longue distance réunissent les régions préfrontale, pariéto-temporale et cingulaire [. . . dont] l'ignition fait intervenir une rétroaction descendante en boucle [. . .] Cette ignition serait la signature objective de l'efficacité esthétique de l'œuvre d'art*" ([Changeux, 2016], pages 169). L'objet même que nous nous fixions en début de ce chapitre est bien effectivement isolé là, mais dans une forme beaucoup trop conceptuelle et

4. Par exemple un paysage maritime peut évoquer les mêmes souvenirs de vacances à l'observateur qu'au peintre.

hypothétique pour convaincre. Il est indispensable que l'on progresse dans nos capacités d'examen des activités cérébrales pour confirmer ces réflexions. Cela ne pourra se faire sans des avancées importantes dans la sensibilité et la rapidité de détection des instruments de mesure (IRMf, MEG) et l'on peut imaginer que l'avènement d'outils plus performants ouvrira alors de nouvelles interrogations qui pourront conduire à d'autres interprétations.

3.3 L'approche neuro-esthétique critiquée

L'importance considérable des travaux consacrés aux bases neurales de l'esthétique, et plus spécifiquement à l'approche neuro-esthétique, ainsi que le soutien considérable qu'elle a reçu dans le monde scientifique [Changeux, 2008, Changeux, 2016, Boulez et al., 2014, Le Bihan, 2012] ne doit pas cacher des critiques importantes qui lui sont adressées. Ces critiques s'adressent en général à toutes les approches de l'imagerie cérébrale [Uttal, 2002, Dupont, 2015] et visent à la fois son objet et ses méthodes. Pour la neuro-esthétique, on les trouve bien rassemblées dans [Vidal, 2011, Vidal, 2012], ainsi que des éléments de réponse dans [Pearce et al., 2016].

Les premières critiques concernent l'objet des travaux de neuro-esthétique. De nombreuses études définissent mal l'objectif de leurs expériences et confondent ou mélangent des notions différentes : beau, art, intérêt, innovation ou singularité ... Les expériences ignorent souvent des dimensions déterminantes : culture, éducation, valorisation sociale, etc. Ces dimensions sont depuis de nombreuses années les composantes identifiées par les sciences sociales, l'écologie, la psychologie. Si les neurologues ne les prennent pas assez en compte, ce n'est point tant qu'ils les ignorent, mais soit ils les considèrent - sans démonstration - comme intégrées à leur démarche, soit ils les jugent trop complexes pour être convenablement traitées et préfèrent alors les exclure totalement du champ de leurs travaux.

Sur l'objet, des critiques beaucoup plus radicales sont exprimées, prenant en compte l'évolution sociale liée au contexte multi-culturel contemporain dans [Brown and Dissanayake, 2009]. On oppose alors l'interprétation du terme *esthétique*, issue du siècle des Lumières et attachée à la réponse émotionnelle résultat de la contemplation des œuvres d'art (celle aussi que poursuivent les neuro-esthètes) et le sens historique associé à tout système de valeur lié à l'appréciation de la beauté, sens que l'on retrouve en particulier dans l'œuvre de Darwin où il s'applique à des gents animales.

Sur la méthode ensuite, on reproche aux neuro-esthètes, non sans pertinence, de mettre en évidence des "corrélations" (telle zone est active lorsque tel stimulus apparaît) et d'en déduire souvent hâtivement des modèles de fonctionnement sous-tendant des causalités. Or les fonctions cérébrales sont distribuées dans le cortex et simultanément diverses aires sont sollicitées à l'occasion de divers traitements. L'analyse a posteriori de cartes d'activité cérébrale, ignorant la diversité des mécanismes sous-jacents et cherchant l'explication la plus concise entre objet et appréciation esthétique, est tentée de donner des raccourcis d'interprétation comme : « *la beauté est une qualité des objets qui est corrélée avec l'activité de cortex préfrontal médian par l'intermédiaire des sens* » [Ishizu and Zeki, 2011].

Chapitre 4

Que disent les manuels de photo ? Qu'en pensent les psychologues expérimentaux ?

Abandonnons le domaine de la neuro-biologie et tournons nous maintenant vers les très abondantes bibliothèques des photographes (ou des artistes peintres), pour voir quelles recommandations ils nous font pour composer une belle photo.

Ce qui frappe tout d'abord, c'est le faible nombre de conseils "opérationnels" qui sont fournis. Les ouvrages se répartissent généralement en deux familles : ceux qui essaient de faire partager une expérience esthétique et ceux qui transmettent une leçon technique.

Les premiers s'intéressent généralement à l'objet (ou au sujet) de la photographie : ils abordent des questions de forme, de lumière, de composition, de couleur, ... à la façon des traités d'esthétique en peinture. Comme eux, ils ont pour but d'éduquer le photographe, c'est-à-dire de le doter de la sensibilité qui l'amènera à la production d'une belle photo, mais celle-ci est définie en termes très génériques qui font référence à un style et bien rarement à des règles. Il est très difficile d'utiliser de tels ouvrages pour répondre à la question : "Cette photo est-elle belle ?" ; tout au plus pourrions nous répondre à la question : "Cette image répond-elle aux critères de ... ?"

Les seconds s'attachent généralement à lier les réglages de l'appareil aux effets que l'on désire obtenir : le choix d'une émulsion ou d'une sensibilité, d'un diaphragme, d'un temps d'exposition, d'une lentille ou d'un éclairage, peut-être de filtres, etc.

Si les manuels de photographie donnent peu de règles utilisables pour décider de la beauté d'une image, les traiteurs d'images, et en particulier ceux qui s'intéressent à l'image animée, ont élaboré des méthodes permettant de noter la qualité d'une image ou d'une séquence d'images (voir chapitre 6 de [Maitre, 2015]). La qualité d'image est alors un attribut technique qui regroupe des aspects particuliers du signal :

1. le contraste ou la dynamique et la diversité des couleurs,
2. la mise au point en particulier sur les contours et la finesse des détails dans les textures,
3. l'absence de bruit sur les plages uniformes et la subtilité des modelés sur les zones faiblement variables,
4. éventuellement les distorsions géométriques et colorimétriques.

Ces qualités techniques permettent de préférer un mode de transmission, de codage ou d'affichage à un autre. Elles ont conduit à des algorithmes opérationnels, par exemple : [Mittal et al., 2013, Moorthy and Bovik, 2011, Charrier et al., 2012, Chandler, 2013]. Mais en n'accordant aucun rôle aux qualités esthétiques de l'image, elles ne répondent pas à nos attentes (voir la figure 4.1). Notons

cependant que la proximité des deux notions : beauté d’une part, qualité d’autre part, a conduit certains auteurs à traiter les deux problèmes avec les mêmes outils [Talebi and Milanfar, 2017].

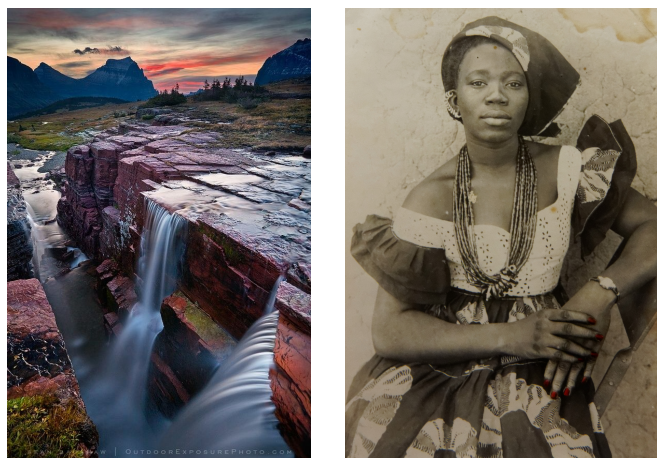


Figure 4.1 – *Qualité de la photographie et Beauté de la photographie. L’image de gauche (Triple falls in Glacier National Park, Montana de Sean Bagshaw) présente tous les attributs de qualité : résolution, mise au point, palette chromatique, nuances, textures. Elle a été distinguée dans la communauté très technophile du site Pinterest pour sa beauté ; il n’est pas sûr pourtant qu’elle rassemble les suffrages de tous les experts en ce qui concerne son esthétique. A droite, au contraire, cette photo (Sans titre de Seydou Keïta) a souffert d’un environnement très médiocre tant à la prise de vue qu’à l’archivage qui ont considérablement nui à sa qualité technique. Elle jouit cependant d’une indéniable reconnaissance quant à son esthétique.*

Notre travail dans ce chapitre consiste à collecter les critères qui sont usuellement proposés pour réaliser de belles photos, à les discuter d’un point de vue des théories disponibles et à voir comment ils ont été transposés en pratique. Nous distinguerons successivement les critères objectifs (le format, la composition, les contrastes, les propriétés spectrales, la palette chromatique . . .) qui se mesurent directement dans l’image, puis ceux, subjectifs qui feront intensivement appel à des observateurs pour être vérifiés. Pour chaque rubrique, nous nous efforçons de confronter trois aspects : tout d’abord les règles issues du domaine de la photographie ou de l’esthétique picturale, puis les mécanismes qui peuvent être mis en jeu pour les expliquer et que justifieraient des études de psycho-physiologie expérimentale et enfin des vérifications statistiques faites à posteriori sur de grands corpus de documents. Malheureusement, dans de nombreuses situations il n’est pas possible de recouper tous ces avis.

4.1 Le format ou rapport d’aspect

Les photographies sont produites selon des formats¹ très variables (voir par exemple le chapitre 3 de [Maitre, 2015]). Le format carré est celui qui tire le meilleur parti de l’objectif photographique. Le format 2/3 a été posé en standard de fait pour des raisons commerciales du temps du film en bobine, le format 4/3 s’est imposé pour des raisons identiques lors de l’avènement des écrans d’ordinateur, considéré avant tout comme un outil de travail à l’époque de son développement. Le format 16/9 a accompagné l’essor des salles de cinéma afin de satisfaire le confort d’une observation à distance. Chaque fabricant propose, de plus, un format particulier lié à la géométrie de ses capteurs : 2/3, 4/3, 5/3, 16/9 . . . , tandis que les fabricants de papier photographique disposent de leurs propres formats, issus souvent des normes de l’imprimerie. Si l’on ajoute que la photographie peut se présenter en format “*portrait*” ou en format “*paysage*” selon le besoin, ce que ne peut faire ni le cinéma ni la vidéo, on voit qu’on dispose d’un très vaste champ du possible.

1. Le mot “format” est ici pris au sens géométrique de rapport largeur/hauteur (rapport d’aspect ou *aspect ratio*). Ce terme a un deuxième sens qui indique le codage adopté pour représenter l’image (JPEG ou TIFF par exemple).

Il n’y a pas vraiment de recommandations esthétiques sur ce point et les ouvrages prônent une grande souplesse. Hormis ceux qui recommandent de suivre la règle du nombre d’or, les autres stipulent plutôt d’adapter le format à la scène représentée sans entrer dans une prescription précise². Ainsi les marines sont très allongées, de même que les scènes de bataille et les plans de villes. Au contraire les natures mortes sont plus compactes, ainsi que les scènes d’intérieur ou de rues. Dans l’autre sens les portraits en pied gagnent à être très hauts tandis que les bustes seuls sont plus ramassés.

4.1.1 Que disent les statistiques ?

Plusieurs travaux se sont attachés à mesurer la répartition des divers formats parmi les œuvres (picturales surtout) réputées belles. Le rapport le plus probable des dimensions sur un grand nombre d’œuvres dépend du sujet, conformément aux recommandations d’Arnheim. Statistiquement, les tableaux verticaux sont préférentiellement dans le rapport 5/4, tandis que les horizontaux sont plutôt dans le rapport 4/3 [Arnheim, 1983]. Des résultats semblables ont été retrouvés dans des collections de photographies d’art, avec pour dimensions privilégiées les rapports 1/1, 2/3 et 3/4 [Arbellini, 2017], rapports typiques des formats de films.

4.1.2 Le nombre d’or

Il vaut $\Phi = \frac{1+\sqrt{5}}{2} \simeq 1,618$ et se trouve au cœur de nombreuses propriétés mathématiques. Il a fait l’objet de très nombreux travaux, d’une part pour tenter de justifier ses qualités esthétiques et donc son emploi depuis l’Antiquité, d’autre part pour le retrouver dans les constructions de la nature [Ghyka, 1931]. Des résultats un peu hâtifs de G. Fechner semblaient en avoir établi l’intérêt d’un point de vue perceptuel et, à la suite des publications de A. Zeising, puis de M. Ghyka, Φ s’est vu promu au titre de “divine proportion” et pour cela abondamment utilisé par de nombreux artistes au cours des deux derniers siècles (Seurat, Le Corbusier, Duchamp, Pissaro, Picabia, Léger, Dali).

Des travaux postérieurs ont largement montré l’arbitraire de beaucoup des vertus esthétiques attribuées à Φ [Markowsky, 1992, Neveux, 1995, Livio, 2008]. Nous retiendrons en particulier ce qui concerne l’aspect général d’une photographie : le nombre d’or n’est pas le rapport préféré pour le grand côté et le petit côté d’une surface rectangulaire quelconque (cette proposition est celle qu’a tenté de prouver Fechner, mais son résultat est entaché d’un biais méthodologique [Markowsky, 1992]) et le rapport le plus probable de ces dimensions sur un grand nombre d’œuvres considérées comme belles n’est pas égal à Φ .

4.2 La composition

La composition est le choix de la disposition des divers motifs d’intérêt au sein de la photographie. Son rôle est d’offrir une ligne de lecture des éléments composant la scène [Bouleau, 2014, Crettez, 2017]. L’impact de la composition d’une photographie sur l’observateur peut être analysé au moyen de techniques d’oculométrie qui enregistrent le parcours de l’œil examinant l’image, l’enchaînement de ses visites et la durée de chaque visite [Yarbus, 1967] (voir la figure 4.2). Quoique F. Molnar ait pu distinguer “bonnes” et “mauvaises” compositions par l’examen de ces parcours (une bonne composition conduit à peu de transitions d’un site à l’autre et à de plus longues fixations [Molnar, 1977]), il demeure difficile d’aller beaucoup plus loin dans le jugement esthétique d’une œuvre par cette technique. On aimerait pouvoir coupler les examens oculométriques avec des enregistrements neurologiques, par exemple par IRMf, de façon à attacher les activations cérébrales à tel ou tel détail. Malheureusement les constantes de temps de ces deux dispositifs sont aujourd’hui incompatibles.

2. Ce qu’Arnheim exprime par : “*The more visual affinity there is between frame and picture, the more they can*



Figure 4.2 – Oculométrie sur une peinture de Vladimir Repin (*La visite inattendue*), à gauche. L'un des parcours du regard obtenus par A. Yarbus, à droite. Yarbus a donné plusieurs consignes à l'observateur afin de répondre à diverses questions : par exemple Quel est l'âge des personnes ? ou Quels vêtements portent-ils ? Chaque consigne conduit à des parcours différents. Ce parcours est obtenu sans consigne et correspond vraisemblablement assez bien à celui que l'on adopterait dans un musée. Notons que ce parcours n'est pas recalé sur l'image, contrairement à ceux que l'on obtient aujourd'hui. [Yarbus, 1967].

Notons cependant que les études oculométriques ont apporté un résultat annexe intéressant notre étude en mettant en évidence une différence dans la façon dont les experts en art ou les non-experts examinent une œuvre. Elles concluent que l'appréciation d'une œuvre d'art est modifiée par l'apprentissage, et par là le jugement esthétique qui lui est porté [Molnar, 1981, Nodine et al., 1993]. Cependant [Rosenberg and Klein, 2015] avance une interprétation différente, rattachant les écarts constatés à des différences de motivation entre les observateurs.

4.2.1 La règle du tiers (*Rule Of Third, ROT*)

C'est en fait la seule règle de photographie qui soit véritablement universelle ; on la trouve dans tous les manuels de photographie, même si ses origines remontent à la peinture du XVIII^e siècle [Amirshahi et al., 2014]. Elle prône que les zones d'intérêt de l'image doivent se trouver de préférence sur les deux horizontales ou sur les deux verticales qui partagent en trois les deux dimensions de la photographie (voir les images de la figure 4.3).

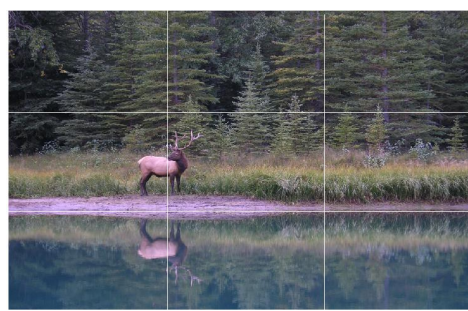
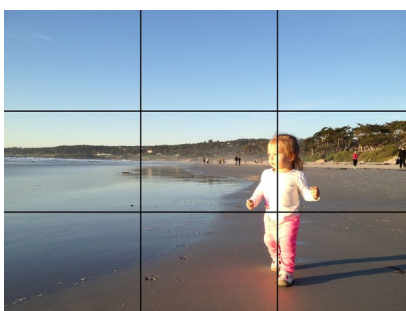


Figure 4.3 – La règle du tiers indique comment l'on doit disposer les zones d'attention au sein de l'image. Divisant longueur et largeur en trois, c'est sur les 4 lignes ainsi tracées qu'il faut placer les objets d'intérêt.

Plusieurs études ont été consacrées à la vérification de la règle du tiers. La plus complète propose, dans [Amirshahi et al., 2014], de tester la validité de cette règle sur une base d'images judicieusement choisies (les images "belles" sont sélectionnées parmi les mieux notées du site Photo.net), soit par un groupe d'observateurs, soit à l'aide d'un algorithme qui utilise comme zone d'intérêt une zone de

influence each other" [Arnheim, 1983].

forte saillance. Les conclusions de cette étude, présentées sur la figure 4.4, font clairement apparaître que la règle du tiers n'est majoritairement pas respectée par les "belles" photos, ni plus ni moins que par les photos ordinaires.

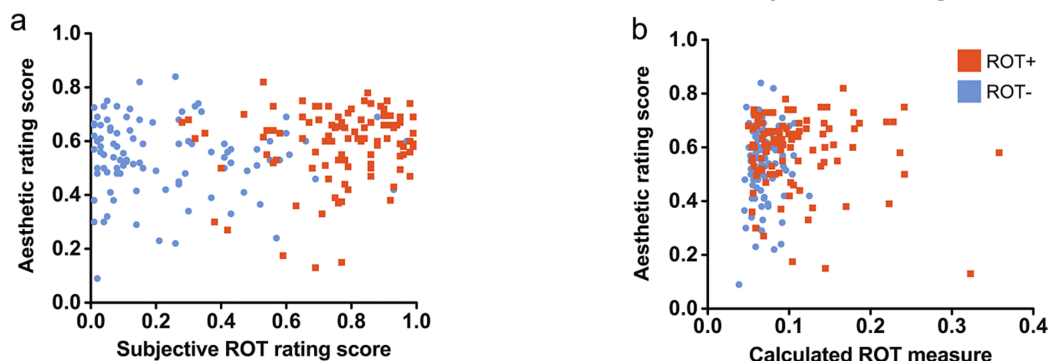
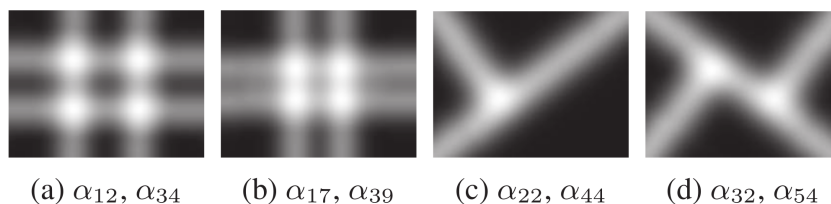


Figure 4.4 – Dans les résultats de l'expérience rapportée dans [Amirshahi et al., 2014], on observe la distribution des appréciations esthétiques de photographies, soit "belles" (notées ROT+), soit "ordinaires" (notées ROT-), en fonction de leur évaluation au regard de la règle du tiers. A gauche les résultats sont obtenus pour une ROT estimée par un observateur humain, à droite par un algorithme s'appuyant sur un calcul de saillance à l'intérieur de l'image tel que proposé dans [Mai et al., 2011].

Notons enfin que l'on trouve dans [Obrador et al., 2010] l'usage de proportions fondées sur le nombre d'or pour placer les points d'intérêt en place du tiers et également des triangles d'or pour placer des droites (voir la figure 4.5). Cet article montrerait que l'usage de ces règles de composition serait efficace pour sélectionner les "belles" images.



(a) α_{12}, α_{34} (b) α_{17}, α_{39} (c) α_{22}, α_{44} (d) α_{32}, α_{54}

Figure 4.5 – Les masques permettant de détecter la composition de l'image selon [Obrador et al., 2010]. De gauche à droite : le masque de la règle du tiers, puis le masque de la moyenne d'or, puis un masque de l'un des triangles d'or, enfin un masque combinant deux triangles d'or. Chaque masque est obtenu par convolution d'une ligne ou plusieurs lignes avec un noyau gaussien. Il y en a 55 en tout, désignés ici par α_n .

4.2.2 Le centre de l'image

Mettre le point d'intérêt au centre de l'image est en opposition avec la règle précédente. C'est cependant l'une des règles proposées par le psychologue D. Ross en 1907. R. Arnheim a développé cette proposition [Arnheim, 1954, Arnheim, 1983]. Il distingue le centre géométrique du cadre et le centre de masse de l'image. Lorsque ces deux points se confondent, la construction apparaît équilibrée. Sinon il s'exerce une force de recentrage potentiellement nuisible à l'esthétique. Il suggère également que les axes horizontaux et verticaux, éventuellement diagonaux, ainsi que les coins constituent d'autres lieux de rappel de l'équilibre concurrentement du centre. Une étude [McManus et al., 2011] s'est attachée à vérifier statistiquement cette propriété, sans vraiment convaincre³. D'autres études ont proposé des formules différentes pour mesurer l'équilibre apparent d'une composition [Hubner and Fillinger, 2016].

3. L'étude adopte une définition trop simple de la masse à partir des niveaux de gris et ne tient pas compte du biais statistique introduit par une image de taille bornée. De plus elle teste plutôt la qualité de l'observateur que celle du

Elles pèchent par des protocoles de vérification utilisant des images très simples (formes géométriques noires sur fond blanc) et leurs résultats sont difficilement généralisables aux photographies.

Pour assouplir et généraliser un peu le rôle du centre de l'image, il a été proposé de définir la "focalité" comme l'existence dans l'image d'un point unique vers lequel converge le regard [Ulrich, 1983]. Une forte focalité est appréciée dans les images. Ce point est souvent en relation avec la présence de lignes fuyantes (voir ci-dessous).

4.2.3 Quelles autres règles de composition ?

Beaucoup d'autres éléments sont aussi présentés dans les manuels de photographie comme importants pour l'évaluation esthétique de l'image, mais à notre connaissance, ces propriétés n'ont pas fait l'objet d'études systématiques.

L'espace négatif On désigne parfois sous les termes d'espaces "positifs" et "négatifs", les portions du champ de la photographie utilisées pour représenter les sujets principaux (ceux que l'on photographie) d'une part, et le fond ou l'environnement d'autre part. Quoique l'espace négatif ne soit pas primordial dans beaucoup de photos, il est recommandé de lui attribuer une place importante, mettant en valeur l'objet principal, et une place d'autant plus grande que l'espace positif est en mouvement ; il sera alors plus ample dans la direction du mouvement (figure 4.6 au centre).

Cette règle très empirique a des conséquences importantes pour définir le format de l'image (sujet abordé plus haut) et amène à choisir un format global faisant écho à l'aspect des objets d'intérêt (un format allongé pour un lac ou une chaîne de montagne et plutôt carré pour un visage).

Les objets répétés et périodiques. Au contraire du cas précédent, on recommande souvent que ces espaces positifs ne soient accompagnés d'aucun espace négatif par un cadrage très serré, limité au seul espace positif, lorsque l'objet est régulier et que cette régularité est l'objet de l'image (c'est ce qui est illustré sur la figure 4.6 à gauche).

Cadrage des corps humains. Il est recommandé de ne couper les corps humains par un cadrage serré qu'en trois positions : à mi-poitrine, à la taille ou à mi-cuisse. Les autres césures sont réputées disgracieuses. Notons cependant que cette recommandation, très imprécise, laisse place à des choix très personnels.

Traitement des fuyantes. Lorsque l'image comporte des parallèles fuyantes dans l'espace objet, leur point de convergence dans l'image, virtuel ou réel est un point d'attraction de l'observation. Il doit être placé judicieusement soit s'il est le sujet principal de la photo, sur une ligne de tiers (voir la section 4.2.1), soit s'il appartient à l'espace négatif de façon à aider l'observation de l'espace positif (figure 4.6 à droite).

Usage des diagonales. Les diagonales constituent des lignes de force sur lesquelles on recommande souvent de s'appuyer soit pour introduire une profondeur dans la scène, soit pour accompagner un mouvement.

document. Elle conclut cependant que la taille et la couleur des stimulus sont de faible importance pour la sensation d'équilibre comparées à leur position, contrairement à une proposition d'Arnheim [McManus et al., 1985].

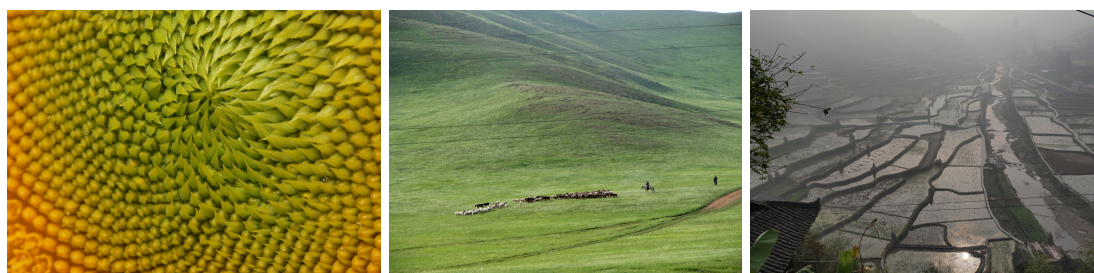


Figure 4.6 – A gauche, sur un objet très régulier, le cadrage serré s'impose tandis que sur une scène en mouvement (au centre), l'espace négatif doit être ample. A droite, le point de convergence des fuyantes est le point fort de l'image.

4.3 Mise au point et spectre de densité de puissance

[Mise au point] La mise au point est un élément très important dans l'évaluation de la **qualité technique** d'une photographie ([Maitre, 2015], chap 6). Sa relation à la beauté de l'image est complexe :

1. un défaut de mise au point est immédiatement perçu, et généralement de façon très négative, surtout s'il affecte l'objet principal ; pour de nombreuses scènes (extérieurs, nature), on attend une même mise au point sur tous les plans de la scène.
2. des photographes peuvent néanmoins utiliser le flou de mise au point pour exprimer des ambiances particulières : irréalité, distance, oubli, mystère⁴ ;
3. le traitement des arrière-plans s'accompagne souvent d'un flou important de façon à mettre en évidence l'objet d'intérêt. C'est une règle particulièrement importante pour les portraits, les photos de fleurs ou d'animaux. On apprécie alors souvent que la profondeur de champ soit très faible (figure 4.7, à gauche).

La recherche d'une très bonne résolution a souvent été un objectif des photographes⁵. L'avènement des images numériques avec des capteurs qui ne satisfont pas aux conditions d'échantillonnage de Shannon⁶ a introduit le défaut opposé : le repliement de spectre (*aliasing*). Des dispositifs ont été mis en place pour réduire ce défaut et les boîtes à outils proposent des logiciels pour agir sur le spectre de densité de puissance du signal (modification de l'acutance⁷ des photos). Les détails fins interviennent dans le piqué⁸ de la photo et contribuent, avec les textures, à la richesse de l'image. Les textures sont une partie importante de l'image, quoiqu'on les remarque peu. Ce sont des variations des valeurs de l'image qui "remplissent" les surfaces entre les contours et permettent de distinguer la nature des objets : herbe, peau, tissu, crépi... L'impact des textures et des détails fins apparaît dans la décroissance du spectre de densité de puissance de la photographie. Des études ont montré que pour respecter une bonne hiérarchie des échelles, il convenait que la décroissance du spectre de densité de puissance suive une loi en $1/f^2$ (où f exprime la fréquence spatiale) [Alvarez et al., 1999, Gousseau and Roueff, 2007]. Le système visuel humain semble particulièrement

4. Notons l'emploi systématique du flou dans le cinéma et Noir et Blanc pour adoucir les visages de femmes.

5. On rappelle en particulier l'éphémère mouvement d'esthétique photographique dénommé $f/64$ qui prônait l'usage exclusif de cette ouverture très petite et très contraignante, mais qui garantissait une excellente mise au point dans tout le champ.

6. Le théorème de Shannon fixe la distance qui doit séparer les échantillons d'un signal lors de sa numérisation pour ne pas introduire de défaut dans le signal ainsi créé. Le théorème de Shannon permet d'avoir un signal discret équivalent au signal continu original. Les conditions de validité de ce théorème portent sur l'extension du spectre de densité de puissance du signal continu. Les capteurs CMOS et CCD, constitués d'une mosaïque de photo-récepteurs, ne peuvent pas satisfaire aux conditions d'échantillonnage de Shannon (voir [Maitre, 2015], chapitre 2).

7. L'acutance est une mesure de la qualité d'une image qui prend en compte à la fois la fonction de transfert de l'appareil photo et la sensibilité du système visuel humain dans les conditions d'observation (voir le chapitre 6 de [Maitre, 2015]).

8. Le piqué est un terme de technique photographique qui désigne un bon rendu des détails fins accompagné d'une bonne résolution de l'image (voir [Maitre, 2015], page 220). L'optimisation de l'acutance est une façon de maximiser le piqué d'une photo donnée.

adapté à ce type de signal [Brady and Field, 2000]. Des études ont montré que c'était bien le type de propriétés que l'on retrouve dans les images esthétiquement plaisantes [Koch et al., 2010]. De plus il a été montré [Schweinhart and Essock, 2013] que l'on appréciait que cette décroissance en $1/f^2$ soit régulière dans toutes les directions de l'espace, tandis que des écarts à cette règle conduisaient à des images "inconfortables" [Fernandez and Wilkins, 2008].



Figure 4.7 – Quelques images qui s'écartent des règles générales de la photographie. A gauche, La mise au point n'est faite que sur la fleur, le reste de l'image est flou. Au centre, image en low keys n'utilisant que des teintes sombres. A droite image en high keys n'utilisant que des fortes luminosités.

4.4 Niveaux de gris et histogramme

Il est souvent recommandé d'utiliser largement la gamme de luminosité disponible en évitant les zones saturées (claires ou sombres), ainsi que des plages non utilisées de l'histogramme [Kardan et al., 2017]. D'ailleurs, la plupart des caméras disposent par défaut de cette fonction d'étirement du contraste. Néanmoins de fameux contre-exemples offrent des dynamiques très faibles, l'ensemble de la photo se faisant par une modulation subtile des niveaux de gris (*low-keys* pour les images très sombres, *high-keys* pour les images très claires (figure 4.7)). Ainsi, il n'y a pas de recommandation impérative sur l'allure générale de la répartition des niveaux de gris (ou de la luminosité dans le cas de photos couleur). L'expérience montre que les histogrammes les plus variés peuvent accompagner de belles photos. Il n'y a donc pas de recommandation pour la valeur moyenne et pour l'écart-type de cette distribution de niveaux de gris. Surprenamment, des études ont montré que ces distributions vérifiaient des propriétés au troisième ordre (la moyenne μ et la variance σ étant les deux premiers). Si p est la distribution des niveaux de gris i , le biais γ est défini par :

$$\gamma = \frac{1}{N} \sum_{i=1}^N \frac{(p(i) - \mu)^3}{\sigma^3} \quad (4.1)$$

où N est le nombre de niveaux de gris de l'image.

On trouve dans [Attewell and Baddeley, 2007] une étude montrant que l'on préfère les images à biais négatif plutôt que celles à biais nul (comme dans les distributions gaussiennes) ou, à fortiori, à biais positif (figure 4.8). Les conclusions de [Motoyoshi et al., 2007] sont un peu différentes mais confirment l'importance du biais dans la qualité perçue. Cette sensibilité au moment d'ordre 3 serait expliquée par des mécanismes particuliers d'adaptation des cellules visuelles [Laughlin, 1987, Brady and Field, 2000, Motoyoshi et al., 2007].

4.5 La couleur

La couleur est bien sûr un attribut très important de la photographie. S'il n'est pas question de restreindre les "belles" photos aux photos en couleur, il est important de souligner qu'il n'est pas concevable qu'une photo en couleur soit jugée "belle" sans que l'on porte un avis positif sur la



Figure 4.8 – Le rôle du biais (moment d'ordre 3) dans l'apparence d'une image. Les photos du haut ont un biais négatif tandis que celles du bas, avec une moyenne et une variance identiques, ont un biais positif [Attewell and Baddeley, 2007].

palette des couleurs qu'elle utilise. En cela, il semble que l'esthétique des couleurs mérite un traitement particulier. Les progrès constants faits par les scientifiques pour relier les phénomènes physiques à nos sensations ont donné lieu à un perpétuel renouvellement de notre interprétation de nos préférences en matière chromatique. Il nous faut distinguer en ce point deux grands volets bien distincts. Le premier concerne l'appréciation que l'on a pour les couleurs prises isolément, la seconde concerne l'harmonie colorée, c'est-à-dire la façon dont est perçu un champ composé de plusieurs teintes juxtaposées.

4.5.1 Les préférences des couleurs isolées

Il y a peu à dire en fait des règles générales de préférences en matière de couleurs. L'adage "des goûts et des couleurs, on ne dispute pas" laisse à penser qu'il n'y a ni règle ni rationalité dans le choix des couleurs. De très nombreuses études ont cependant été conduites pour essayer de déterminer certaines tendances générales et pour les justifier.

Une revue très complète est proposée dans [Hurlberg and Ling, 2012] des divers travaux menés depuis un siècle sur les préférences en matière de couleurs. Il apparaît très généralement que les couleurs *bleu* et *bleu-vert* sont généralement préférées tandis que les teintes *jaune* et *jaune-vert* sont les moins appréciées [McManus et al., 1981, Ou et al., 2004]. Cet ordre de préférence est assez globalement indépendant des caractéristiques des observateurs (origine, sexe, âge, éducation), quoique de subtiles dépendances aient pu être parfois détectées (voir par exemple [Hurlberg and Ling, 2007, Taylor et al., 2013]). Il est à noter que ces préférences s'appliquent à des populations importantes mais souffrent de nombreux écarts individuels.

Des tentatives pour proposer des modèles mathématiques de prédiction de l'appréciation d'une couleur ont été faites, mais elles ne sont pas vraiment concluantes :

- en décomposant le signal à partir de nos connaissances du système visuel [Ling and Hurlberg, 2007]. Exploitant la représentation LMS, [Vienot and Le Rohellec, 2012], on utilise alors un modèle à base de quatre variables : *l'intensité*, *la saturation*, *le contraste vert-rouge L-M* et *le contraste jaune-bleu S-(l+M)*, les deux derniers facteurs expliquant 70 % des préférences ;
- à partir de critères de psychologie affective [Machajdik and Hanbury, 2010]. Les trois grandeurs *plaisir*, *excitation* et *dominance* s'expriment à partir de la luminance Y et de la saturation s

par :

$$\begin{aligned} plaisir &= 0,69Y + 0,22s \\ excitation &= 0,31Y + 0,60s \\ dominance &= 0,76Y + 0,32s \end{aligned} \quad (4.2)$$

- en choisissant des pondérations émotionnelles des couleurs *activité colorée, poids et chaleur* [Ou et al., 2004]. Les émotions sont au nombre de 10, chacune dotée d'une polarité positive/négative, parmi lesquelles : *lourd/léger, moderne/classique, masculin/féminin, dur/doux*, etc.
- à partir d'arguments écologiques [Palmer, 2010] (voir la figure 4.9), s'appuyant sur 4 catégories : *couleurs saturées, légères, sombres et voilées*.

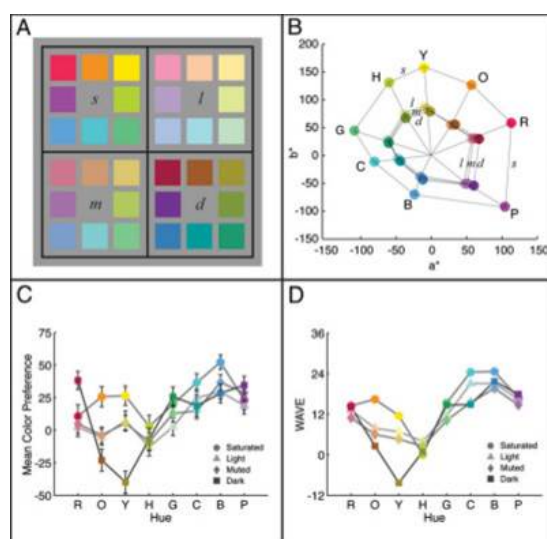


Figure 4.9 – Mise en évidence des préférences chromatiques. En A, les diverses couleurs utilisées lors de l'étude [Palmer, 2010]. On juge indépendamment des petites plages colorées, bien séparées. En B leur position dans le plan a^*b^* de l'espace Lab. En C le résultat de l'appréciation par les 48 observateurs (notée de -5 à +50), en D l'évaluation à l'aide du modèle WAVE (weighted affective valence estimate) proposé dans [Palmer, 2010], en assez bon accord avec l'expérience.

4.5.2 Les préférences concernant les palettes colorées

Si la littérature hésite à s'engager sur les recommandations concernant les couleurs prises isolément, elle est plus riche sur les propositions concernant l'assemblage de plusieurs couleurs dans une composition, ce que l'on appelle généralement la palette.

L'harmonie des couleurs s'est développée pendant vingt siècles au sein de l'harmonie platonicienne sur des bases très incertaines où, par exemple, on décomposait l'octave allant du noir au blanc en intervalles simples, de tierce, ou de quarte. La décomposition de la lumière à l'aide du prisme, mise en évidence par Newton, a ouvert la porte à des interprétations plus réalistes utilisant soit des échelles linéaires, soit des "roues chromatiques" (figure 4.10 à gauche) où étaient placées en opposition des couleurs complémentaires (Newton, Goethe, Runge, Bezold, Chevreul). Néanmoins une grande confusion⁹ s'est maintenue, en particulier dans le milieu artistique, tant que l'on n'a pas distingué synthèse additive (telle que réalisée par Newton) et synthèse soustractive (obtenue par mélange de pigments par les peintres). Ainsi, si E. Chevreul est considéré à juste titre comme celui qui a le premier très largement exploré la richesse des harmonies chromatiques nécessaires à la réalisation des

9. Cette confusion a porté essentiellement sur le nombre des couleurs primaires et sur leur choix.

tapisseries¹⁰, son interprétation à partir des “complémentaires” est aujourd’hui réfutée. Une autre approche globale de la palette est celle d’Itten qui prône la recherche d’un équilibre. Cet équilibre peut être atteint avec un champ gris, ou en combinant des couleurs complémentaires. L’harmonie idéale est atteinte lorsque le bon équilibre est trouvé [Itten, 1973].

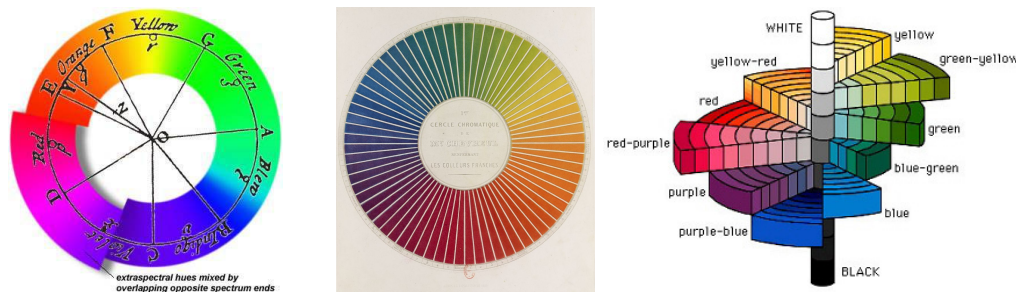


Figure 4.10 – A gauche, la première roue des couleurs telle que vue par Newton. Au centre, un cercle chromatique de Chevreul (source : Gallica/BNF), à droite (simplifié) : le nuancier de Munsell qui sépare les diverses plages colorées selon des degrés de différence perceptuelle expérimentalement contrôlée.

C’est à partir des travaux de A. Munsell, en 1905 (et à l’aide de son atlas rationnel des couleurs - figure 4.10 à droite), puis grâce aux travaux de normalisation de la CIE en 1931 (en particulier grâce à la mise en place de l’espace chromatique CIELab ainsi que des métriques associées), et enfin à ceux de D.B. Judd, en 1943 (faisant le lien entre les intervalles expérimentaux de Munsell et la métrique de CIELab), qu’ont pu être menées des études systématiques et reproductibles d’harmonie chromatique.

L’harmonie selon Moon et Spencer :

Ce sont les travaux de P. Moon et D.E. Spencer qui constituent la référence scientifique la plus fréquemment invoquée lorsque l’on cherche les fondements mathématiques de l’harmonie colorée [Moon and Spencer, 1944c]. Notons cependant tout de suite que cette théorie donne les outils géométriques pour manipuler les couleurs ainsi que les règles pour évaluer la qualité esthétique d’une palette, mais qu’elle ne justifie les recommandations qu’elle fait que par l’accumulation des écrits antérieurs d’autres auteurs (Chevreul, Oswald, Bezold, Brücke, Field). Les auteurs n’invoquent ni arguments physiologiques, ni résultats de psychologie expérimentale. Pour les auteurs, les règles qu’ils proposent rendent compte d’une pratique universellement partagée dans les milieux artistiques. Elles mettent sous forme mathématique un savoir universel qu’ils considèrent avéré. Que disent les règles de l’harmonie chromatique de Moon et Spencer ? Elles recommandent de placer toute couleur dans un espace ω chromatique tel que le nuancier de Munsell, c’est-à-dire défini en coordonnées cylindriques par la luminosité, la teinte et la saturation (figure 4.11 à gauche). Les dix teintes de base de Munsell sont réparties de façon uniforme et définissent autant de secteurs du cylindre. Dans cette représentation ω , l’association de couleurs est harmonieuse si :

1. les relations entre deux couleurs quelconques ne sont jamais ambiguës,
2. les points chromatiques représentant ces couleurs sont disposés selon une distribution géométrique simple dans ω

Il y a ambiguïté de deux couleurs si leur teinte appartient à deux secteurs de teintes voisines (figure 4.11 à droite). Selon la loi 1, on autorise des teintes à être soit proches (mais pas trop) soit en contraste chromatique. Selon la loi 2, on peut faire varier deux ou trois variables à condition de laisser les points de couleur dans des plans ou sur des droites de l’espace ω . Il existe donc une infinité d’harmonies, les

10. Chevreul a dépassé les travaux antérieurs (ceux de Goethe par exemple) qui ne considéraient que la teinte comme composante de l’harmonie, en distinguant les harmonies de teinte et les harmonies de contraste, qui font appel explicitement à la saturation et l’intensité en plus de la teinte [Chevreul, 1864]. En cela il a préparé les travaux de Moon et Spencer.

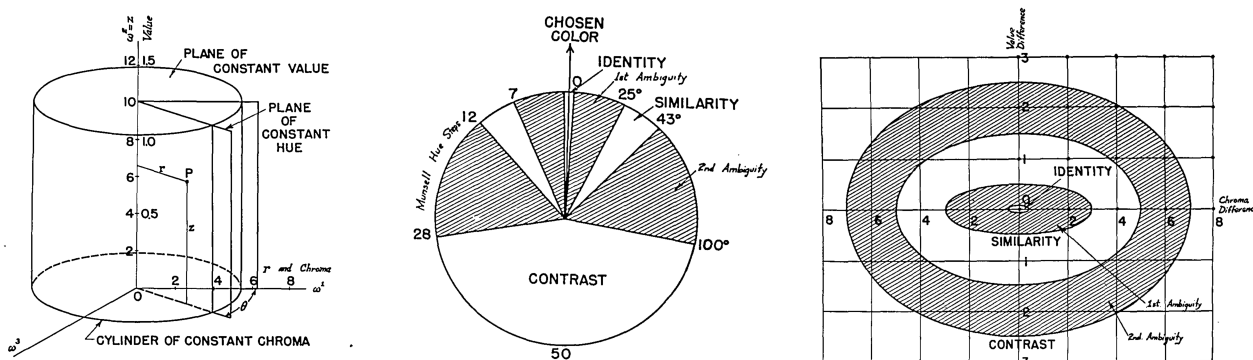


Figure 4.11 – Les travaux de Moon et Spencer : à gauche, l'espace de représentation d'une palette de couleurs. Pour les 2 dessins de droite, les zones hachurées représentent les zones interdites tandis que les zones blanches correspondent aux couleurs qui sont en relation harmonieuse. Au centre à luminance constante ; à droite à teinte constante, d'après [Moon and Spencer, 1944c].

auteurs se gardant de les hiérarchiser. De plus, les auteurs proposent une classification à deux classes (harmonieuse ou dis-harmonieuse), reconnaissant cependant que la frontière est probablement floue entre les deux et qu'il est probablement possible de trouver des gradations dans l'harmonie.

Moon et Spencer proposent des valeurs numériques précises pour définir les zones permises et interdites (celles précisées sur les figures 4.11).

De ces postulats, Moon et Spencer tirent une classification méthodique des harmonies qu'ils explorent systématiquement. Trois classes principales sont distinguées selon le nombre de variables que l'on s'autorise à laisser libre. La première classe est ainsi obtenue en laissant se déplacer les points de couleur selon les 3 variables luminance, teinte ou saturation. Lorsque 2 variables sont libres, les points de couleurs peuvent décrire des formes plus complexes et plus encore dans le cas à trois variables.

Notons qu'ils se sont plus intéressés à un examen systématique de toutes les configurations compatibles avec les deux postulats qu'ils ont établis en début d'étude qu'à une vérification de leurs qualités esthétiques. Ils ont cependant mené des expériences avec la majorité des harmonies qu'ils proposent¹¹ et leur reconnaissent généralement d'intéressantes qualités, quoiqu'ils distinguent des configurations plus plaisantes, plus éclatantes ou plus subtiles.

Ces travaux ont été complétés d'études prenant en compte la taille des zones, ainsi que leur distance à un point d'accommodation chromatique [Moon and Spencer, 1944b]. Ces travaux peuvent se résumer ainsi :

1. une balance plaisante de n zones colorées est obtenue lorsque les moments colorés¹² de ces zones sont égaux dans l'espace ω ;
2. un autre équilibre plaisant est obtenu si ces moments colorés sont dans des rapports simples.

Comme on le voit, l'approche de Moon et Spencer suit une démarche "objectiviste" de l'esthétique et laisse très peu de place à l'observateur dans l'affirmation de ses préférences. D'ailleurs, les auteurs ont affirmé cette approche en traduisant dans le formalisme objectiviste de Birkhoff (tel que nous l'avons présenté en section 5.2) leur expression de l'harmonie chromatique [Moon and Spencer, 1944a]. La formulation qui en résulte, qui prend en compte des coefficients de pondération heuristiquement déterminés, est très confuse et difficilement exploitable.

11. Ces expériences étaient réalisées à l'aide d'assemblages de papiers colorés, suivant le nuancier de Munsell et disponibles dans le commerce.

12. Le moment coloré est le produit de la surface de la zone par sa distance au point d'adaptation chromatique (souvent sur l'axe des teintes neutres, mais potentiellement déplacé lorsque la scène a une couleur dominante), distance mesurée à l'aide des métriques de Munsell (exprimées en distances jnd (*just noticeable difference*)).

TABLE IV. Classification of harmonies.

Class I. One variable	
1. Variable z	
A. Achromatic	(a) Two color-points (b) Three points (c) More than three points
B. Chromatic	(a) Two points (b) Three points (c) More than three points
2. Variable r	(a) Two points (b) Three points (c) More than three points
3. Variable θ	(a) Two points (b) Three points (c) More than three points
Class II. Two variables	
1. In a plane of constant θ	A. Points on a straight line B. Triangles C. Rectangles D. Points on a circle
2. In a plane of constant z	(a) Two points (b) Three points on isosceles triangle (c) Five points on two triangles (d) n points on a circle with center on neutral axis (e) n points on a circle with center at a chromatic point
3. In a cylinder of constant r	(a) Two points (b) n points on an ellipse with center on neutral axis
Class III. Three variables	
1.	With reference to planes of constant θ
2.	With reference to planes of constant z
3.	With reference to cylinders of constant r
4.	With reference to tilted planes

TABLE V. Class I. Harmonies in one variable.

1A. Variable z , achromatic harmonies	
(a) Two color-points	(1) Harmony of analogy. Difference in z is between 4 and 12 ($\frac{1}{2}$ to $1\frac{1}{2}$ Munsell value steps). Example: $N 7, N 8$. (2) Harmony of contrast. Difference in z is between 20 and 80 ($2\frac{1}{2}$ to 10 Munsell value steps). Examples: $N 3, N 6; N 2, N 6; N 1, N 9$.
(b) Three points	(1) Small equal steps. Difference in z is between 4 and 12. Example: $N 4, N 5, N 6$. (2) Large equal steps. Difference in z is between 20 and 80. Example: $N 2, N 5, N 8$. (3) Large and small steps. Examples: $N 2, N 3, N 7; N 3, N 7, N 8$.
(c) More than three points	An obvious extension of the classification for three colors.
1B. Variable z , chromatic harmonies	
(a) Two points	(1) Harmony of analogy Examples: $R 6/8, 5/8; B 6/6, 7/6$. (2) Harmony of contrast Examples: $P 6/8, 3/8; RP 8/6, 3/6$.
(b) Three points	(1) Small steps Examples: $CY 7/8, 6/8, 5/8; Y 8/6, 7/6, 6/6$. (2) Large steps Example: $R 8/4, 5/4, 2/4$. (3) Large and small steps Examples: $BG 7/4, 6/4, 3/4; YR 8/4, 4/4, 3/4$.
(c) More than three color-points	

TABLE VI. Class I. Harmonies in one variable.

1B. Variable z , chromatic harmonies	
(a) Two points	(1) Harmony of analogy Examples: $R 6/8, 5/8; B 6/6, 7/6$. (2) Harmony of contrast Examples: $P 6/8, 3/8; RP 8/6, 3/6$.
(b) Three points	(1) Small steps Examples: $CY 7/8, 6/8, 5/8; Y 8/6, 7/6, 6/6$. (2) Large steps Example: $R 8/4, 5/4, 2/4$. (3) Large and small steps Examples: $BG 7/4, 6/4, 3/4; YR 8/4, 4/4, 3/4$.
(c) More than three color-points	

Figure 4.12 – Les divers types d’harmonie chromatique selon Moon et Spencer. A gauche l’ensemble des harmonies, à droite, celles à une seule variable libre, donc de type 1. D’après [Moon and Spencer, 1944c].

L’harmonie de Matsuda :

Y. Matsuda a proposé en 1995 une extension de l’approche de Moon et Spencer. Se plaçant dans l’espace CIE HSV¹³, cette approche reconnaît, comme chez Moon et Spencer, des configurations harmonieuses (voir figure 4.13). Mais celles-ci sont distribuées différemment et peuvent rendre compte d’harmonies à plus de deux couleurs. La mesure de l’harmonie globale se fait par sommation de toutes les configurations compatibles avec les masques donnés, chaque configuration étant chargée d’un poids égal au produit des saturations, donnant ainsi un rôle plus important aux teintes vives. Notons qu’il n’y a aucune démonstration de la légitimité de cette approche dans les travaux de Matsuda, mais il s’appuie sur une longue pratique expérimentale.

Des travaux plus récents ont cherché à déterminer les règles d’une harmonie universelle à l’aide de techniques d’apprentissage et de bases de données du web [Lu et al., 2014a, Lu et al., 2016]. Nous les examinerons plus loin (section 7.9).

Des harmonies qui ne sont pas universelles :

Depuis cinquante ans, de nombreux travaux s’inscrivent en faux face à l’hypothèse de couleurs universellement appréciées. De nombreux auteurs soulignent l’importance de la personnalité, de la culture, de l’éducation (voir par exemple [Mehrabian, 1977, Smith, 2008]).

Dans cette voie, Schloss et Palmer [Schloss and Palmer, 2011] ont revisité la notion d’harmonie chromatique en tempérant l’approche objectiviste. S’appuyant sur une analogie avec la musique¹⁴, ils

13. L’espace HSV (*Hue*, *Saturation*, *Value*) est un espace chromatique beaucoup utilisé dans le monde de l’informatique, retenu par la Commission Internationale de l’Eclairage (CIE).

14. L’analogie musicale proposée est de comparer des musiques de Mozart et de Stravinsky. Si l’on peut préférer Stravinsky à Mozart (critères individuels), on convient généralement de trouver Mozart plus harmonieux.

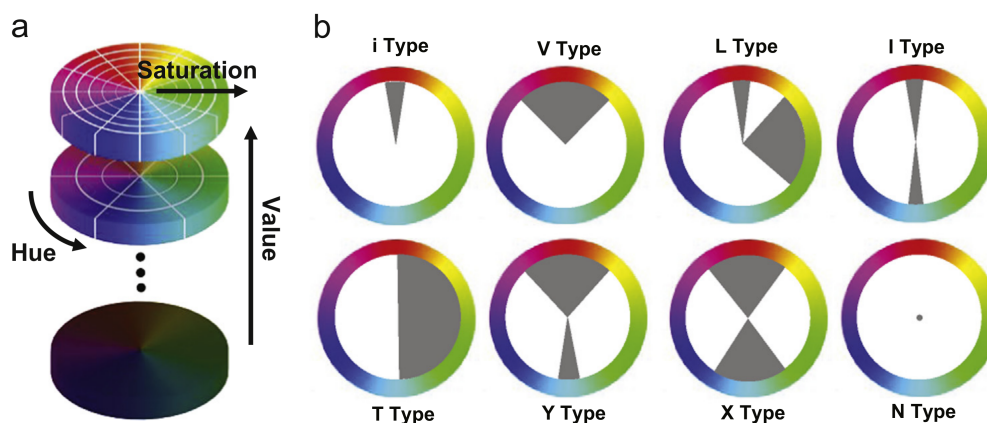


Figure 4.13 – Les configurations retenues par Matsuda. Le cercle chromatique peut librement tourner autour de son centre. Les configurations associent la teinte et la saturation. D’après une figure de [Lu et al., 2015a].

ont pris en compte trois critères de qualité lors de l’appréciation d’une association de deux couleurs :

1. la *préférence* de l’observateur pour l’association des deux couleurs,
2. la perception de l’*harmonie* au sein de cette association,
3. l’appréciation de l’*accord* de ces deux couleurs avec les autres teintes considérées comme un fond.

La distinction entre *préférence d’une paire* et *harmonie d’une paire* leur semble nécessaire comme elle l’est pour distinguer des sons *harmonieux* de sons *agréables*. C’est sur cette distinction que peuvent se placer des différences inter-individus (attachées à l’éducation ou à la personnalité). Néanmoins, leurs conclusions ne laissent pas une grande place à cette distinction qui apparaît en pratique trop subtile pour se refléter nettement dans leurs expériences.

Poussant plus loin encore le refus de l’objectivisme, Z. O’Connor, dans l’intention de personnaliser complètement la mesure de l’harmonie [O’Connor, 2010] a retenu plusieurs facteurs particuliers qui interviennent dans la perception de la couleur : les différences individuelles (genre, âge, personnalité . . .), l’expérience culturelle, le contexte, les effets perceptuels et le temps (influence de la mode). Pour elle, ces facteurs s’additionnent et affectent multiplicativement une harmonie universelle (comme celle qu’aurait permis de calculer Moon et Spencer ou Matsuda) selon une formule du type :

$$\mathcal{H}_{O'Connor} = [\Phi_{individuel} + \Phi_{culturel} + \Phi_{contexte} + \Phi_{mode} + \dots] \mathcal{H}_{Matsuda} \quad (4.3)$$

Bien sûr une telle formule nécessite que de nombreux paramètres soient connus pour pouvoir être utilisée opérationnellement, ce qui est bien rarement le cas. On n’est donc pas surpris de peu rencontrer ces formules dans les travaux conduits sur d’importantes populations.

4.6 Que dit la psycho-sociologie ?

Nous avons examiné très en détail les propriétés des photographies telles que les manipulent les traiteurs d’image. Ces grandeurs sont directement issues de l’image et se prêtent bien à une interprétation objectiviste de l’esthétique puisqu’elles donnent à évaluer des grandeurs également disponibles pour tous.

D’autres études ont adopté des chemins différents, en questionnant les observateurs sur des aspects moins facilement mesurables de l’image et qui s’insèrent mieux dans une approche subjectiviste. Ces travaux ont pour méthodologie des enquêtes généralement conduites dans le cadre des départements de psychologie, de sociologie ou d’histoire de l’art. Ils offrent des résultats qui éclairent d’un jour très

différent notre rapport au beau et ces résultats sont généralement très difficiles à inclure dans des approches automatisées car ils font référence à des concepts très mal modélisés de façon opérationnelle pour un informaticien. Nous ne pouvons cependant les ignorer.

4.6.1 Mesurer la beauté

Réponses à des enquêtes, les résultats obtenus reposent sur l'avis de l'observateur. Il est important alors que l'on comprenne bien ce que sera la réponse de l'observateur et, ceci fixé, que l'on trouve des mesures permettant de transporter les réponses d'observateurs différents dans un cadre où l'on pourra les comparer.

La dimension du Beau

Des travaux assez anciens se sont posé la question de l'espace dans lequel évolue le Beau. Est-ce une grandeur scalaire ? A-t-elle plusieurs dimensions ? Cet aspect a été abordé à l'aide des outils de la psychologie expérimentale, déjà anciens. Ils ont cherché à déterminer une échelle globale de mesure de l'appréciation esthétique et à la "dimension" du jugement esthétique [Dewar, 1937, Eysenck, 1939]. Ils s'appuient sur l'ordonnement d'une collection d'œuvres d'art par des expérimentateurs plutôt novices et par l'examen croisé de leurs résultats. Les œuvres choisies sont des reproductions de peintures, de vases, de fleurs, etc. qui doivent être ordonnées, à l'intérieur d'une même thématique, en fonction de leurs seules qualités esthétiques. Après une analyse statistique des résultats (plus de 300 observateurs sont employés) qui visent à éliminer l'observateur et à n'examiner que les relations entre œuvres, les auteurs concluent : "*L'analyse des facteurs suggère qu'un seul facteur général du goût artistique est responsable des classements artistiques.*" [Dewar, 1937]. Ce facteur est assimilé donc à la "Beauté" ¹⁵ et il lui est attribué une dimension 1. La seconde étude nuance ce résultat en montrant qu'un second facteur intervient si l'on introduit des documents abstraits dans le test (dessins, couvertures de livres, broderies). Ce second facteur disperse les résultats selon un axe allant du formel au représentatif [Eysenck, 1939] et les auteurs concluent que la dimension est alors 2.

A notre connaissance ces travaux n'ont été utilisés que pour concevoir des tests d'aptitude aux études artistiques mais n'ont pas été employés pour des études de mesure des qualités esthétiques.

Une échelle du Beau

Receemment, des travaux ont souligné la difficulté de travailler en esthétique sans disposer d'une mesure opérationnelle de ce qu'est le beau. En effet, la plupart des études ignorent ce problème en se reposant sur l'avis supposé sincère des expérimentateurs et sur un sens commun supposé universel.

Les travaux récents qui ont tenté d'aborder ce problème s'appuient généralement sur une étape intermédiaire, la recherche d'une mesure du plaisir hédonique, sensé refléter l'émotion éprouvée par l'observateur et donc mesurer directement le degré de beauté. Dans le domaine des arts visuels, ces travaux sont particulièrement balbutiants et leurs résultats sont loin de faire l'unanimité [Augustin et al., 2011, Hayn-Leichsenring et al., 2017, Blijlevens et al., 2017, Augustin et al., 2018] ; Ils ont plus de succès dans les champs du design ou de la conception des interfaces homme-machine, domaine où l'on sait mieux mesurer d'autres aspects liés à l'utilité fonctionnelle.

15. Il est intéressant de noter ici combien cette notion est fidèle à l'idée objectiviste d'un beau unique et universel. Au contraire, elle s'oppose frontalement à l'interprétation subjectiviste qu'exprime par exemple Xénophon, sophiste de l'école de Socrate, qui affirme : "*Un beau lutteur est différent d'un beau coureur, un bouclier qui est beau pour la défense est complètement différent d'un javelot, beau pour la vitesse et le lancé.*" (Xénophon, Commentaires, III,8,4, rapporté par [Tatarkiewicz, 1970]).

4.6.2 Images de la nature ou scènes artificielles ?

Dans le registre désigné parfois sous le nom d'esthétique environnementale [Ulrich, 1983], quelques études de psychologie expérimentale se sont attachées à rechercher des critères d'appréciation esthétique des photographies combinant une multitude de facteurs soit de bas niveau soit de haut niveau. L'idée directrice est de déterminer ce qui nous permet de décider qu'une image représente une scène naturelle (jardins, parcs, montagnes) ou un paysage en partie construit par l'homme (rues, bâtiments). Ils s'appuient sur des travaux déjà anciens [Kaplan et al., 1972], qui ont montré que l'aspect naturel est un critère important du jugement esthétique car il est bénéfique pour l'attention, la mémoire et généralement pour stimuler les mécanismes impliqués dans l'appréciation esthétique¹⁶

Ainsi, dans [Kardan et al., 2017], on compare le rôle des primitives de bas niveau dans la reconnaissance des scènes naturelles anthropisées. A partir de primitives issues de l'étalement de l'histogramme, de la palette des couleurs, de la distribution des contours ou de l'entropie, les auteurs montrent que l'on peut raisonnablement prédire si la scène est naturelle ou anthropisée et en déduire le jugement esthétique d'un observateur. On vérifie au passage que, à complexité équivalente, les scènes de nature sont préférées aux paysages artificiels.

Les trois primitives les plus significatives sont :

- l'absence de segments droits dans les contours, les ombres et les silhouettes,
- une dominante chromatique plutôt portée par l'antagonisme jaune-vert que par l'antagonisme bleu-rouge,
- une grande dynamique dans la saturation chromatique.

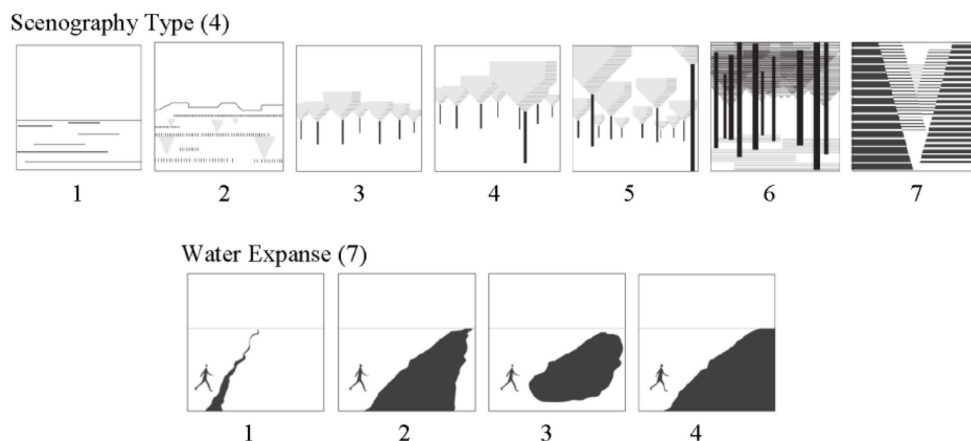


Figure 4.14 – Deux exemples des primitives de haut niveau proposées par [Hunter and Askarinejad, 2015] et utilisées dans [Ibarra et al., 2017] pour mesurer les propriétés “naturelles” ou “artificielles” d’une scène. La scénographie décrit la répartition d’ensemble des diverses masses de l’image, tandis que la mesure nommée Water Expanse s’intéresse à la disposition d’une masse d’eau. Les autres mesures également significative pour mesurer l’aspect naturel d’où découlerait la préférence esthétique sont : la distribution des bâtiments, les surfaces construites pour permettre le mouvement, la géométrie de la ligne d’horizon et l’ondulation maximale de la ligne d’horizon (qui correspond à sa variation totale).

Les auteurs de [Kardan et al., 2017] laissent une remarque en suspens. Lorsque l’on choisit des

16. Les travaux de [Kaplan et al., 1972] ont été revisités dans [Kotabe et al., 2017] qui cherche à expliquer notre préférence pour les scènes naturelles. H.P. Kotabe et ses collègues aboutissent à une conclusion paradoxale : les scènes naturelles sont effectivement préférées, malgré leurs propriétés “désordonnées”, ce qui est contradictoire avec notre préférence vérifiée assez régulièrement pour l’ordre [Berlyne, 1971, Ulrich, 1983]. Ils testent trois hypothèses concurrentes pour résoudre ce paradoxe :

1. la préférence pour les scènes naturelles est plus forte que l’aversion pour le désordre,
2. le désordre n’intervient pas dans l’évaluation de scènes naturelles (masquage par la sémantique),
3. le désordre est apprécié dans les scènes naturelles.

Ils concluent que la première hypothèse est la plus probable.

photos où les primitives ci-dessus ne sont pas déterminantes, on parvient cependant à conclure entre images naturelles et anthropisées, mais au prix d'un temps de réaction plus long qui laisse penser que d'autres mécanismes que le système visuel sont actifs.

Dans [Ibarra et al., 2017], l'étude précédente est complétée par la prise en compte non seulement de primitives de bas niveau, mais aussi de primitives de haut niveau. Celles-ci sont choisies parmi celles proposées par [Hunter and Askarinejad, 2015] et traduisent des aspects d'une grande variété thématique (végétation, bâti, voies de circulation) et d'une grande richesse sémantique (obstacle à la pénétration, franchissement, organisation ...). Ces primitives sont au nombre de 62; 2 sont présentées sur la figure 4.14. Les auteurs montrent que six d'entre elles permettent, mieux encore que les primitives de bas niveau, de prédire le jugement esthétique d'un observateur. La séparation entre naturel et anthropisé se fait à près de 70 % sur ces critères plutôt que sur les indices de bas niveau, tandis la prédiction esthétique les utilise à plus de 50 %.

Ces études sont conduites sur quelques centaines de photos et soumises au jugement d'une cinquantaine d'évaluateurs expérimentés. Les traitements de bas niveaux sont assurés par un logiciel de traitement d'images tandis que les mesures de haut niveau sont le résultat de la décision d'un expert et sont mesurées numériquement par l'intermédiaire d'une tablette graphique.

4.6.3 Le rôle du contexte

Ces travaux trouvent leur place au sein d'une interprétation historique de l'esthétique (voir section 2.3.4). Ils visent à mesurer l'influence sur l'évaluation d'une œuvre, des informations qui peuvent accompagner son examen.

Les travaux de V. Swami [Swami, 2013] se sont attachés à déterminer l'influence du type de connaissances attachées à une œuvre sur deux propriétés particulières de l'observateur : d'une part sa capacité à comprendre l'œuvre, d'autre part le jugement esthétique qu'il porte sur celle-ci. Les expériences conçues procèdent en collectant les opinions de quatre cohortes d'expérimentateurs confrontés à la visualisation de peintures abstraites, statistiquement équivalentes, mais dotées d'informations préalables différentes. Les réponses attendues permettent de tester tout d'abord le niveau de compréhension de l'œuvre, puis de mesurer son évaluation esthétique.

L'expérience est constituée de trois tests différents :

- Tout d'abord, devant un lot de peintures de Max Ernst, les cohortes sont dotées, I_1) soit d'aucune information, I_2) soit du titre de l'œuvre¹⁷, I_3) soit du titre et d'une présentation de l'artiste mentionnant quelques traits de sa méthode artistique (mais sans référence à l'œuvre présentée), I_4) soit, en plus des informations précédentes, d'une description de l'œuvre et d'une note sur l'intention de l'artiste pour la produire.

Les conclusions de cette étude montrent que, si la capacité à comprendre l'œuvre progresse régulièrement de I_1 à I_4 , la valeur esthétique attribuée à l'œuvre est identique pour I_1 , I_2 et I_3 et ne progresse très sensiblement que pour la cohorte I_4 . Cette expérience montre que la valeur esthétique est dépendante d'une connaissance sur la manière et l'intention de cette œuvre.

- Dans cette seconde étude, on cherche à déterminer l'influence de l'information additionnelle (de type I_4 ci-dessus) dans le cas d'œuvres abstraites ou représentationnelles (des peintures de Picasso issues soit de sa période cubiste dénotées P_c (donc abstraites) soit de sa période Bleue, dénotées P_b). Pour cela on réduit les tests aux expériences I_1 et I_4 ci-dessus, faites sur les informations P_c ou P_b .

Les résultats montrent que l'information contextuelle I_4 est très importante pour améliorer la compréhension de P_c , mais peu utile pour la compréhension de P_b . Cette information permet également de faire croître très sensiblement le jugement esthétique de P_c et plus modestement celui de P_b .

- Dans une troisième expérience, l'auteur souhaite tester le rôle d'une information non-pertinente

17. D'autres études ont été faites sur le rôle du seul titre dans l'appréciation esthétique [Leder et al., 2006].

dans l'interprétation et le jugement de l'œuvre. Pour cela il reprend l'expérience 1 d'évaluation d'œuvres de M. Ernst ci-dessus, avec deux cohortes recevant soit I_1 soit I_4 , puis deux autres cohortes qui reçoivent l'information P_c et P_b ci-dessus (exposant la technique de Picasso), plus enfin une cohorte utilisant l'information P_b mais avec un autre nom que Picasso comme auteur. Les résultats montrent que comme dans la première expérience, l'information exacte contribue très largement aussi bien à la compréhension qu'à l'appréciation. Au contraire, les fausses informations ne contribuent ni à l'une ni à l'autre : il n'y a pas d'évolution sensible des notes attribuées.

On déduit de ces travaux que l'appréciation esthétique dépend de la connaissance que l'on a sur l'intention et la manière de l'auteur lors de sa création. Ces résultats se transportent probablement dans le domaine de la photographie, mais il faut prévoir que leur rôle est important. Probablement faudrait-il mesurer de façon identique l'apport d'information sur le contexte historico-géographique de la prise de vue, mais il n'est pas clair que l'accroissement de l'intérêt porté à l'image ne serait pas dû principalement à un effet de *relevance* (voir le paragraphe 1.1).

4.6.4 La perception globale : unité et prototypalité

On doit aux écoles de *design* des études plus formelles sur des éléments qui influencent le jugement esthétique d'une scène. Ces travaux ont été orientés vers la compréhension des avis portés sur les œuvres créées par l'homme : mobilier, architecture d'intérieur, haute-couture, mais rien ne s'oppose à ce que leurs résultats soit importés dans le champ des beaux-arts et de la photographie.

Ils ont permis de dégager des critères spécifiques et nous en retiendrons deux qui nous semblent particulièrement pertinents : l'unité et la prototypicalité.

Unité : L'unité est relation qui relie les diverses parties de la scène, en particulier en mettant en relation d'harmonie des caractéristiques visuelles locales saillantes. Une photo exprimera des qualités d'unité si les diverses parties expriment les mêmes propriétés perceptives : contraste, finesse des détails, mise au point, palette de couleurs, etc. Les travaux [Bell et al., 1991] et [Veryzer Jr and Hutchinson, 1998] apportent des éclairages sur le rôle de l'unité dans l'appréciation d'une scène. Ces études suggèrent que les objets esthétiquement attractifs présentent une forte unité, tandis que les objets à faible unité reçoivent des avis esthétiques plus faibles.

Prototypicalité : La prototypicalité est la propriété d'un objet qui le rattache à une catégorie ou aux valeurs moyennes des objets de cette catégorie. Certains domaines de la photos s'exposent plus à cette propriété : le nu, les portraits, les photos d'animaux, les couchers de soleil. Les études conduites sur la prototypicalité dans le domaine du design aboutissent à une conclusion assez unanime que plus de prototypicalité conduit à plus d'esthétique [Barsalou, 1985], ce qui est cependant assez paradoxal. Certes, en s'approchant de la forme type on réunit très probablement le plus grand nombre de traits appréciés dans chaque expression de la forme. Certes on a l'impression que l'on retrouve dans le prototype les "canons" de la beauté aristotélicienne. Certes on s'attend avec des formes proches du prototype à offrir des lignes familières à l'observateur et donc à lui faciliter le travail. Mais cependant on écarte par là-même des éléments justement au cœur de l'œuvre artistique et qui en sont usuellement inséparables : l'innovation et la création. Ces résultats sont cependant confirmés par de nombreuses expériences [Whitfield and Slatter, 1979, Veryzer Jr and Hutchinson, 1998] quoiqu'aucune n'ait été conduite dans le domaine de l'image ou de la photographie où la notion de prototype mériterait d'être développée¹⁸ ?

18. Une exception doit être faite cependant dans le domaine de l'esthétique des portraits où des travaux se sont attachés à définir un visage moyen et où il a été montré que ce visage est particulièrement attractif. On est cependant loin des critères de beauté tels que ceux adoptés dans cet article.

Chapitre 5

Les approches algorithmiques de calcul de la beauté

Nous allons examiner maintenant les travaux qui ont proposé de faire calculer la beauté d'une photographie (ou d'une image, ou d'une peinture) par un programme et non plus par un humain.

Ces travaux se divisent en deux grandes parties très inégales : la première portée par ce chapitre est consacrée aux méthodes algorithmiques qui proposent de confronter la photo à un calcul dont les paramètres sont connus et universels. Ces méthodes s'appuient donc beaucoup sur l'approche objectiviste. La seconde partie traite des méthodes qui s'appuient sur un apprentissage à partir d'exemples notés par un observateur humain. Cette partie sera abordée en trois chapitres. Le premier examine les bases de données qui servent à l'apprentissage ainsi que l'expertise utilisée en référence. Le suivant examine les méthodes développées de 2000 à 2014 qui utilisent des primitives extraites des images, des règles mises en œuvre des classifieurs agissant sur les primitives. Le chapitre suivant est consacré aux réseaux de neurones profonds qui ont été utilisés depuis 2015.

5.1 Les premiers essais : Ch. Henry

Les premières contributions à cette détermination automatique de la beauté sont dues à Charles Henry [Henry, 1885, Henry, 1891]. Identifiant très lucidement que la détermination exacte des modifications physico-chimiques corrélatives aux perceptions esthétiques était hors de la portée de la science de son temps, il a proposé un cadre général pour évaluer les liaisons entre excitation et sensation, quelle que soit la modalité de perception. Fondée à la fois sur l'harmonie "antique" des proportions (en particulier les 12 premières puissances positives et négatives de $3/2$ et les polygones inscrits réguliers), ainsi que sur les théories thermodynamiques de Van't Hoff et le principe de moindre action, cette théorie introduit un principe unificateur de *dynamogénie* qui nous semble aujourd'hui assez confus [Henry, 1889]. Il semble avoir retrouvé ainsi par les mathématiques des évaluations esthétiques des harmonies chromatiques corroborées par les travaux de Chevreul et d'Helmholtz, des formes [Henry, 1895] ainsi que des musiques telles qu'enseignées par Rameau. Les travaux d'Henry, quoiqu'ils furent très bien reçus en leur temps, n'ont pas été repris à sa mort.

5.2 L'approche mathématique de G. D. Birkhoff

Le mathématicien et probabiliste Georges D. Birkhoff a adopté une approche tout à fait différente, mais elle aussi empreinte de l'idée fondamentale que la beauté réside dans l'objet et qu'elle peut se mesurer par une combinaison judicieuse de ses attributs [Birkhoff, 1933]. Il propose pour cela une

formule de la beauté qu’il teste assez systématiquement sur des données très variées. Cette formule relie trois grandeurs :

- une grandeur, notée \mathcal{M} , qui mesure la *beauté* de l’objet, c’est-à-dire sa valeur esthétique et qui apparaît comme une récompense de l’effort qui est réclamé pour examiner l’objet¹ ;
- une grandeur, notée \mathcal{O} , qu’il appelle l’*ordre* et qui mesure les qualités d’harmonie, de régularité et de symétrie indispensables à l’expression esthétique ;
- une grandeur, notée \mathcal{C} , qu’il appelle la *complexité* et qui correspond à l’effort d’attention ci-dessus.

Il propose alors pour conjecture :

$$\mathcal{M} = \frac{\mathcal{O}}{\mathcal{C}} \quad (5.1)$$

Il passera une bonne partie des dernières années de sa vie à vérifier la validité de cette conjecture dans diverses situations : en poésie, en musique, en dessin, en sculpture. Elle exprime la beauté comme une sorte de “récompense” à un effort consenti par l’observateur pour intégrer la richesse de la scène. Notons qu’il ne dispose à cette époque d’outillage ni pour décrire des formes (cela viendra plus tard avec les outils de traitement d’images et de reconnaissance des formes), ni pour exprimer la complexité (cela viendra avec les travaux de Shannon et de Kolmogoroff). Enfin il ne peut s’appuyer sur les travaux de psychologie expérimentale de l’école de la Gestalt qui sont publiés dans des cercles assez étroits dans cette période [Wertheimer, 1922, Wertheimer, 1923] et lui sont encore inconnus. Birkhoff est donc amené à exprimer de façon très spécifique les termes de son équation pour chaque problème traité, ce qui réduit considérablement la portée de ses travaux : pour chaque figure (vase, frise), il lui faut créer un “dictionnaire” des primitives, puis recenser les propriétés qui créent un “ordre” parmi ces formes (voir figure 5.1). On trouvera également dans [Delahaye, 2015] une discussion des résultats de Birkhoff.



Figure 5.1 – Birkhoff a expérimenté la validité de son équation sur des vases chinois en porcelaine. La complexité de la forme s’exprimait à partir de la diversité des plans tangents, des courbures et des symétries dans divers plans, ainsi que par la position des points d’inflexion et des extrêmes de courbure. Le résultat de ses mesures de beauté était finalement comparé au prix affiché de ces vases dans des salles de vente, considéré comme l’expression objective de leur beauté. L’accord des deux mesures l’a conforté dans ses choix (source [Birkhoff, 1933]).

1. Cette attention doit aller jusqu’à l’empathie de l’observateur avec son objet pour que s’exprime pleinement la dimension esthétique de l’objet. Cela est directement dans la ligne des travaux du psychologue T. Lipps qui a inspiré Birkhoff [Lipps, 1903].

5.3 A la suite de G. Birkhoff

5.3.1 Les *designers*

Après-guerre, sous l'impulsion des grands courants de design industriel, les formulations de Birkhoff inspirent plusieurs chercheurs qui prennent en compte à la fois les progrès faits en psychologie expérimentale et en théorie de l'information. Ce sont en particulier A. Moles en France [Moles, 1957] et M. Bense en Allemagne qui apportent sur ces points les contributions les plus originales [Bense, 1969]. Ainsi, Bense propose de transformer la formule 5.1 en utilisant l'entropie de Shannon H appliquée à la seule couleur de l'image. L'ordre \mathcal{O} s'exprime alors comme la réduction d'entropie entre la palette complète équi-répartie H_{max} et la palette effectivement utilisée pour une œuvre donnée, H_o , conduisant à une formule de beauté :

$$\mathcal{M}_B = \frac{H_{max} - H_o}{H_{max}} \quad (5.2)$$

5.3.2 P. Machado et A. Cardoso

Plus tard P. Machado et A. Cardoso [Machado and Cardoso, 1998] proposent une formule “à la Birkhoff” (dont il semble cependant qu'ils n'aient pas eu connaissance), où la complexité \mathcal{C} et l'ordre \mathcal{O} sont remplacés respectivement par la *complexité de traitement* PC , et la *complexité de l'image* IC . Après quelques manipulations, la formule 5.1 devient :

$$\mathcal{M}_{MC} = \frac{IC^a}{(PC(t_0)PC(t_1))^b \left[\frac{PC(t_1) - PC(t_0)}{PC(t_1)} \right]^c} \quad (5.3)$$

où les exposants a , b , et c sont arbitraires et les instants t_0 et t_1 expriment le temps d'observation. Les auteurs évaluent enfin la différence entre la complexité de l'image et la complexité de traitement par le système visuel à l'aide de la différence entre image originale et image compressée (par un codage JPEG ou un codage fractal). Ces travaux sont remarquables car ils proposent une intéressante vérification expérimentale à partir d'une base d'images de tests de capacités artistiques. Par ailleurs, ils introduisent explicitement la durée d'observation $t_1 - t_0$ comme paramètre de la beauté, une hypothèse qui sera également au cœur des travaux de J. Schmidhuber.

D'autres auteurs ont exploité plus à fond la théorie de la complexité pour en dériver des mesures de \mathcal{O} et \mathcal{C} en utilisant les travaux de Kolmogorov et de Levin sur la complexité [Koshelev et al., 1998].

5.3.3 Les travaux de J. Rigau *et al.*

Il faudra cependant attendre plusieurs années pour que ces travaux soient adaptés à une étude plus construite des œuvres peintes. J. Rigau, M. Feixas et M. Sbert ont produit le travail le plus accompli à partir de la formulation de Birkhoff, appliquée à l'évaluation des qualités esthétiques de peintures des XIX^e et XX^e siècles [Rigau et al., 2008]. Pour cela ils ont mobilisé les outils de traitement des images (segmentation par zones, classification des couleurs) et la théorie de la complexité.

Plusieurs formulations sont proposées. Tout d'abord la complexité de Kolmogorov K peut avantageusement être employée dans l'expression 5.2 :

$$\mathcal{M}_{RFB_1} = \frac{nH_{max} - K}{nH_{max}} \quad (5.4)$$

où n est le nombre de pixels de l'image. La complexité de Komogorov K étant non calculable, on l'approche en utilisant les algorithmes de compression du commerce, en l'occurrence JPEG (comme

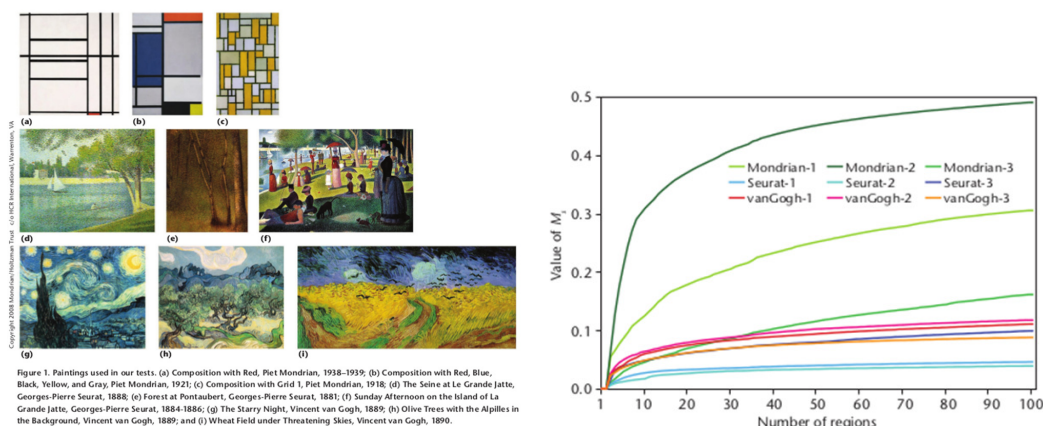


Figure 5.2 – Dans leur travail, Rigau et al. ont choisi une base expérimentale de peintures très variées qu'ils ont décomposées en plages colorées. La mesure de beauté de Birkhoff est déduite d'un calcul complexe mettant en jeu la complexité de Kolmogorov. A droite, l'évolution de la beauté de Birkhoff en fonction du nombre de plages retenues (source [Rigau et al., 2008]).

dans [Machado and Cardoso, 1998]). La formule ci-dessus est ensuite raffinée en utilisant l'entropie physique H_o de Zureck qui conduit à :

$$\mathcal{M}_{RFB_2} = \frac{nH_o - K}{nH_o} \quad (5.5)$$

Finalement les auteurs construisent un algorithme de segmentation de l'image par arbre, en ν zones, sur un critère de couleur et expriment la variation de la beauté de l'approximation $B(\nu)$ obtenue, représentée par la description $\hat{R}(\nu)$ par la formule (illustrée sur la figure 5.2 à droite) :

$$\mathcal{M}_{RFB_3}(\nu) = \frac{H[B(\nu), \hat{R}(\nu)] - H[B(\nu)|\hat{R}(\nu)]}{H[B(\nu)]} \quad (5.6)$$

où $H(x, y)$ exprime l'entropie jointe des processus x et y et $H(x|y)$ l'entropie conditionnelle de x sachant y .

Ces travaux méritent une expérimentation plus complète. Tout d'abord, il semble difficile de traiter de la même façon des styles très différents en raison du fort contexte culturel qui ne manque pas de masquer les seules évaluations esthétiques comme nous l'avons souligné en introduction. Par ailleurs, la prise en compte des propriétés spécifiques du système visuel humain semble indispensable lors de l'évaluation de la "complexité" de l'image, la métrique de l'espace chromatique et les effets de masquage en particulier.

5.4 Une approche algorithmique d'intelligence artificielle

Les travaux de J. Schmidhuber ne sont pas liés aux précédents en ce qu'ils ont recours à une formulation originale de la beauté [Schmidhuber, 1997]. Ils ne sont pas non-plus spécifiquement applicables dans un contexte objectiviste puisque leur auteur défend l'idée que l'évaluation de la beauté dépend de l'observateur et des conditions précises de l'observation. Il nous semble cependant que l'approche algorithmique proposée, ainsi que le poids accordé à la simplicité dans la mesure de la beauté les placent dans la lignée de ces travaux que nous venons de voir, plus que dans l'esprit des méthodes d'apprentissage que nous verrons plus loin.

J. Schmidhuber choisit une approche de théorie de l'information et de codage, qu'il appelle le modèle du "cerveau paresseux", où l'image, notée D est transmise à l'observateur qui s'en fait une

représentation D' par un algorithme de codage noté C . D, D' et C peuvent être décrits par des chaînes de symboles, conformément à la représentation des objets informatiques proposée par Turing. C représente, à un instant donné, toutes les connaissances que l'observateur peut avoir acquises sur D . Si $D' < D$, alors D est redondant par rapport aux connaissances de l'observateur et il est donc compressible au sens de la théorie de l'information. Pour J. Schmidhuber, il y a beauté s'il y a une forte probabilité de D' conditionnelle à D sachant C [Schmidhuber, 2007]. La beauté est directement mesurée par le gain en compression obtenu par l'observateur pour passer de D à D' . La beauté est donc d'autant plus grande que l'on dispose des règles pour interpréter le schéma visuel qui nous est proposé et en faire une description simple au regard de nos connaissances.

Ce schéma semble adapté pour décider, en présence d'un grand nombre de stimuli D_1, D_2, D_3, \dots lequel nous préférons, mais il ne semble pas à même de nous guider dans un choix d'action devant des stimuli très différents. J. Schmidhuber, propose, dans cette situation, de définir la démarche esthétique comme celle qui, conduit à un progrès dans la capacité de compression. Dans le but de modéliser cette démarche d'apprentissage, il a recours à la théorie de Kolmogorov pour caractériser la complexité du compresseur, ou plus précisément la variation de complexité du compresseur au cours du temps, par exemple dans une expérience d'apprentissage par renforcement. En l'absence de toute autre récompense extérieure, le comportement optimal consiste alors à optimiser ce qu'il appelle la récompense de "curiosité" qui s'exprime en fonction des connaissances C_t et C_{t-1} que l'on aurait si l'on devait observer D_t en utilisant la connaissance de D'_t ou non. L'implémentation de cette stratégie - très théorique - utilise abondamment les schémas de prédiction de Solomonoff et les machines virtuelles de Gödel.

En pratique, la mise en œuvre de cette approche demeure très complexe et, comme dans le cas des expériences de D. Birkhoff, repose sur une détection *ad hoc* des meilleurs schémas de codage (figure 5.3). De plus J. Schmidhuber concède lui-même que de nombreux schémas de représentation qui devraient conduire à des codage très simples ne fournissent pas des résultats remarquablement heureux. Le modèle de J. Schmidhuber semble trouver une bonne confirmation dans les études faites sur les visages moyens. Plusieurs travaux ont souligné que les "visages moyens" obtenus par exemple pour les travaux de reconnaissance spatiale, bénéficiaient d'un très bon jugement esthétique. Or ce sont probablement ceux que nous apprenons par défaut au cours de notre vie.

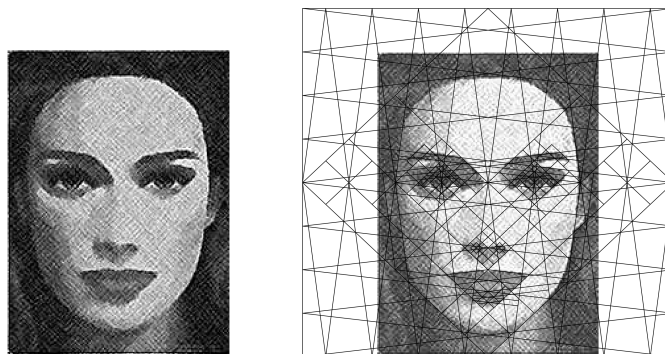


Figure 5.3 – L'image de gauche est belle car elle est simple. J. Schmidhuber propose de la déterminer à partir d'un schéma itératif très simple qui aboutit au réseau des lignes de l'image de droite. Les points les plus remarquables se trouvent représentés de façon très compacte sur ce maillage adapté (d'après [Schmidhuber, 2007]).

Chapitre 6

Le Saint-Graal de la Société du Numérique : l'intelligence artificielle

Les trois chapitres qui commencent ici seront consacrés aux méthodes dites par apprentissage automatique. Le principe de toutes les approches que nous verrons dans les deux chapitres suivants, repose sur l'hypothèse que les règles qui président à l'évaluation d'une photo peuvent être déduites de l'observation d'un grand nombre d'images qui ont été évaluées par des experts humains. Implicitement, on fait donc l'hypothèse que les règles sont universelles et intemporelles. Si l'on n'introduit pas d'autres règles, on est donc pleinement plongé dans le cadre objectiviste.

Nous verrons que deux familles successives de méthodes ont été proposées. Tout d'abord ont été développés des algorithmes de classification, assez simples, utilisant des informations sélectionnées dans l'image, le plus souvent choisies pour leur rôle dans le jugement esthétique. Puis les méthodes à base de réseau de neurones profonds les ont progressivement évincés à partir de 2015 ; ils ont fait porter tout l'effort sur la puissance du système de classification (le réseau profond) en laissant de côté l'étape de sélection de primitives. En contre-partie, ils ont réclamé de très grosses bases de données pour assurer un apprentissage de qualité.

Nous allons nous consacrer, dans le présent chapitre, à l'examen des sources d'images et des sources d'expertises communes à ces deux familles de travaux, ainsi qu'à l'ensemble des considérations qui sont partagées par ces techniques dites "par apprentissage"¹

6.1 L'Intelligence Artificielle (IA), pourquoi faire ?

Devant la difficulté des approches de psycho-physiologie ou de neuro-biologie de l'esthétique des images, de nombreux travaux ont été proposés afin d'aborder ce problème par les techniques informatiques. Ces travaux s'appuient sur la conjonction de trois progrès importants :

1. le développement de méthodes d'apprentissage (couramment dénommées *machine learning*),
2. l'accès aisé à de grandes quantités d'images mises à disposition sur le web, bases de données accompagnées d'éléments de jugement de qualité,
3. la disposition de processeurs puissants à même de traiter de très grosses quantités d'information.

1. La méthode de Schmidhuber, vue au chapitre précédent (paragraphe 5.4), se présente également comme une méthode d'IA et peut être vue comme une méthode d'apprentissage, mais l'objectif n'est pas d'apprendre des critères d'esthétique des images, mais pour chaque image, le comportement, dans le temps, des connaissances de l'observateur [Schmidhuber, 2007].

Nous séparerons les méthodes d'apprentissage intensif utilisées pour les études d'esthétique en deux familles principales :

- La première (chapitre 7) regroupe les **classifieurs fonctionnant sur des primitives extraites des images**. Particulièrement utilisées ces dernières années, les méthodes à base de machine SVM (*Support Vector Machine*), de mise en œuvre assez simple, conduisent à des temps d'apprentissage raisonnables avec des bases d'apprentissage assez réduites (fort pouvoir de généralisation). Leurs performances dépendent beaucoup de la qualité des primitives qui sont utilisées. Elles sont souvent bonnes, mais il est assez difficile d'apprendre des relations subtiles (relations hiérarchiques, exceptions, classification sémantiquement complexe, etc.).
- Depuis quelques années on leur préfère des classifieurs de type **RNP "Réseaux de Neurones Profonds"** (*deep neural network*) (chapitre 8). Les RNP sont des classifieurs "de bout en bout", c'est-à-dire qu'ils prennent en entrée l'image elle-même et donnent en sortie la classe ou la mesure que l'on attend. Les RNP sont de mise en œuvre beaucoup plus complexe car ils enchaînent des couches successives (de 5 à 15) de traitements selon des architectures savamment agencées. Ils nécessitent également de très longues étapes d'apprentissage de façon à configurer chaque couche de classifieur (apprentissage des dépendances entre couches). Pour cela, il faut disposer de très nombreuses données d'apprentissage. En contrepartie, les RNP offrent des performances remarquables, même dans les cas très difficiles et nécessitent très peu d'expertise.

6.1.1 Les résultats attendus

Tous les travaux qui s'intéressent à l'esthétique des photographies n'ont pas le même objectif. Ainsi, certains ne cherchent qu'à séparer les images médiocres des bonnes photos (c'est-à-dire de qualité utilisable par exemple en tant qu'illustration). D'autres cherchent à distinguer les photos d'amateurs des photos de professionnels en vue par exemple d'une commercialisation. Certains cherchent à classer, sur un critère esthétique, un ensemble limité de photos pour constituer un résumé, un album ou un répertoire. C'est en particulier le cas des systèmes qui se proposent en assistance du photographe, pour décider du meilleur cadrage d'un ensemble de photos. Cela peut se faire à posteriori sur une station de travail, mais c'est de plus en plus proposé comme système embarqué [Lo et al., 2013, Schwarz et al., 2016] Enfin d'autres s'attachent à attribuer une note dans un référentiel universel (de 0 à 100, ou sur 5 niveaux).

Un petit nombre de travaux cherchent à "expliquer" les facteurs de leur jugement en fournissant, par exemple, des notes partielles sur divers critères : *composition, couleur, contraste, etc.* ou des cartes qui indiquent les zones où les critères ont été détectés.

Enfin quelques travaux s'appuient sur les informations complémentaires que l'on peut trouver dans les bases de données sur la thématique dont relève l'image, par exemple *portrait, paysages* . . . de façon à adapter la note en privilégiant certains critères. Lorsque la thématique n'est pas donnée, il peut arriver que l'on demande à l'algorithme de la déterminer avant de procéder à l'évaluation. Enfin dans un petit nombre d'études, on associe à l'image des descriptifs d'ambiance, liées à des émotions : *bonheur, effroi, surprise* . . . pour aider à l'évaluation. Ces émotions sont alors données par des photo-interprètes quoique, de plus en plus, des algorithmes se proposent de les estimer lorsque l'image comporte par exemple des visages en gros plan.

6.1.2 Les expertises

Les méthodes de *machine learning* ont recours à des experts lors de la phase initiale d'apprentissage qui permet de fixer les nombreux paramètres des algorithmes. Le choix de ces experts est crucial puisque c'est leur façon de décider que cherchent à imiter tous les algorithmes. Contrairement à ce qui se faisait pour la conception d'autres outils de décision automatique (comme les systèmes experts, maintenant obsolètes), les experts ne participent pas directement à la construction des méthodes de *machine learning*. Ces nouvelles méthodes ne nécessitent que la connaissance du résultat de l'expertise sur une grande variété de cas et ignorent tout de la démarche qui a conduit à cette expertise.

De grandes différences existent entre les études sur ce point. Visitons les rapidement.

Les expertises les plus reconnues sont issues du **domaine professionnel de la photographie** : des critiques d'art, des journalistes, des photographes professionnels, des jurys de concours photographiques. Ces expertises sont disponibles sur les sites spécialisés, dans les revues en ligne, mais aussi à l'occasion de compétitions souvent largement médiatisées. Si les avis ainsi exprimés sont généralement peu contestables, ils nécessitent souvent une étape d'interprétation car ils s'expriment rarement sous une forme immédiatement exploitable (on souhaiterait idéalement un avis sous forme de note dans un référentiel absolu). Cette étape est très réductrice et il n'est pas clair qu'elle ne dénature pas exagérément ces expertises. De plus, ces avis peuvent refléter des partis pris d'école ou de tendance, soit parce que l'expert en est imprégné, soit parce que le cadre de la compétition le réclame. De plus, ainsi que nous l'avons dit en section 1.1 sur les diverses raisons de regarder une photographie, il est parfois délicat de faire la part de ce qui provient, dans un commentaire élogieux, des seuls critères esthétiques. Notons en particulier que la photographie, suivant la tendance générale de l'art au *XX^e* siècle a, aujourd'hui, tendance à privilégier le "pouvoir émotionnel" sur le "plaisir esthétique" comme nous l'avons exprimé en section 1.2, et les avis des experts reflètent souvent cette tendance (voir la note 4 du paragraphe 1.1).

Moins rares que les avis des experts, mais aussi plus discutés sont les avis des **amateurs éclairés** qui sont généralement réunis sur des sites spécialisés comme DPChallenge. Leur très grand nombre permet de conduire des études statistiques sérieuses. La grande diversité des origines des "experts" permet d'assurer une meilleure couverture des opinions sur des tranches d'âge, des origines socio-professionnelles, des contextes géographiques et culturels différents. Les avis sont généralement un peu argumentés mais se traduisent aisément en score sans risque d'altérer l'avis exprimé. Lorsque les évaluations sont faites lors de "challenges", on doit cependant souligner tous les biais des votes ouverts (comportements de groupe, bourrage d'urnes, etc.) [Reagle, 2013].

Plusieurs études se sont appuyées sur de **petites cohortes d'experts** constituées généralement d'étudiants en beaux-arts ou en traitement des images, entraînés spécifiquement pour évaluer selon des protocoles bien établis des bases d'images préparées. Ces travaux ne garantissent ni la qualité des avis des experts professionnels, ni l'étendue des résultats des amateurs éclairés, mais ils assurent une homogénéité et une cohérence des résultats auxquelles les travaux précédents ne peuvent prétendre.

Enfin quelques études se fondent sur des "**non-experts**" en prenant des avis, souvent binaires (*j'aime* ou *je n'aime pas*), directement sur les réseaux sociaux généralistes ou spécialisés pour la diffusion de photos (Instagram, Flickr). Une autre tendance très en vogue est de faire appel aux internautes à l'occasion de petites compétitions lors de campagnes de **crowdsourcing**. Dans les deux cas, on essaie de déterminer un avis "moyen" parmi la population en réduisant notablement les attentes sur les critères d'esthétique ainsi dégagés.

6.2 Les bases de données

Il est très facile de trouver beaucoup d'images sur le web mais il est beaucoup plus difficile de trouver les informations associées qui sont indispensables à l'apprentissage, et en l'occurrence des informations fiables sur la qualité esthétique des photos. Les bases de données utilisées font donc souvent l'objet de compromis entre des avis très peu fiables mais très nombreux, issus du grand public, et des avis, précieux mais très rares, issus de professionnels dans la photo d'art (voir table 6.1).

Les sites professionnels donnent d'excellentes sources de photographie, de très haute qualité, mais pour la difficulté que nous avons soulignée ci-dessus d'exploiter les expertises, ils sont finalement peu utilisés directement pour constituer des bases volumineuses. Leur mérite principal est de fournir d'excellents exemples associés à des commentaires très instructifs.

Les sites d'amateurs éclairés sont la source première d'images. Ils offrent de très grandes quantités de photos, souvent de bonne qualité. Ils permettent d'éviter l'écueil de la "tentation artistique"

de façon à privilégier les qualités esthétiques, ils ne sont cependant pas à l'abri de critiques. En particulier, dans un mouvement que l'on observe bien pour tous les réseaux sociaux, ils ont tendance à privilégier des comportements "de clans". Par exemple, on ne peut manquer d'être frappé sur certains sites par l'importance des effets spéciaux (matériels ou logiciels) mis en œuvre pour "améliorer" l'image : accroissement du piqué, du contraste, filtrage du bruit, saturation des couleurs, distorsions géométriques, etc., reflétant une esthétique véritablement technophile bien loin des critères classiques des pères de la photographie (on se reportera pour exemple à la figure 4.1 à gauche). Enfin, les jugements portés ne sont pas à l'abri de critiques liés à la nature informelle de ces consultations [Reagle, 2013].

Les bases de données généralistes d'images, comme Flickr, Tumblr, Picasa, Instagram ou ImageNet, sont également une source permettant de disposer d'une très large variété d'images, mais leur qualité est parfois médiocre et les éléments associés qui permettraient de servir d'expertise ne reposent le plus souvent pas sur des critères esthétiques. Ces sources "primaires" ne sont pas toujours utilisées en l'état pour alimenter les études sur l'esthétique. Elles sont parfois filtrées pour fournir les bases de données opérationnelles de ces études. Présentons ces bases opérationnelles ici :

- **Flickr**² : quoique source primaire d'images, Flickr a aussi été utilisée comme base de données pour des études sur l'esthétique dans les travaux de [San Pedro and Siersdorfer, 2009, San Pedro et al., 2012], à partir de la notion d'attrait (*attractiveness*) qui mêle de façon un peu confuse des critères de beauté et de popularité. L'attrait est mesuré par le décompte des annotations faites par les internautes : "*attractive*" ou "*unattractive*".
- **Photo.net**³ : c'est un site d'amateurs éclairés et de professionnels de la photo qui regroupe probablement plus d'un million de photos de bonne qualité, classées en 38 catégories. Les photos reçoivent des notes de 1 à 7 selon les deux critères d'esthétique et d'originalité (voir figure 7.2 à gauche). L'accès aux photos peut se faire par ordre de note moyenne décroissante.
- **DPChallenge**⁴ est aussi un site d'amateurs éclairés et de professionnels. Il comporte plus de 650 000 photos généralement de bonne qualité. Il donne une large place aux concours photos, jugés par les internautes. Les notes vont de 1 à 10. Chaque photo reçoit en moyenne plus de 200 notes. Les très nombreux concours sont identifiés par un thème qui peut servir d'identifiant (près de 1000 thèmes aujourd'hui). DPChallenge est la base primaire essentielle des études sur l'esthétique ; elle a donné naissance en particulier à la base AVA, qui est la plus utilisée. Outre le grand nombre d'images de bonne qualité, on apprécie qu'elle offre un large pannel de notes, ainsi que des commentaires littéraires sur de nombreuses images. Des études se sont attachées à bien connaître les propriétés statistiques des appréciations portées sur les images de DPChallenge. La répartition des notes est pratiquement gaussienne autour d'une moyenne un peu supérieure à 5 (voir figure 6.1 à gauche). L'écart-type des notes est lui-même assez gaussien. Les commentaires sont répartis de façon plutôt poissonnienne, avec un mode au voisinage de 15 commentaires par photo (voir la figure 6.1).
La dispersion des notes pour une même photo est reliée à des termes qui expriment soit l'originalité de la photo, soit la personnalisation de l'avis exprimé [Kim et al., 2015]. Elle reflète assez bien ce que certains auteurs appellent la "non-conventionalité" de la photo, c'est-à-dire sa capacité à sortir des canons classiques [Talebi and Milanfar, 2017].
- **CUHK-DB**⁵ développée par l'université de Hong-Kong, elle est issue de DPChallenge en vue de soutenir les études sur l'esthétique. Elle a été obtenue en sélectionnant des images ayant reçu une note soit appartenant aux 10 % meilleures soit aux 10 % moins bonnes. Les deux classes sont donc très peu ambiguës et les images "moyennes" ne participent pas à l'apprentissage.
- **ImageCLEF**⁶ est issue de la base de données MyFlickr issue elle-même de Flickr (ce sont donc des images ordinaires sans qualité esthétique particulière). Elle a été développée pour tester les algorithmes d'indexation et de fouille d'images. Elle comporte un million de photos, avec des annotations sur le contenu (44 tags sémantiques) ainsi que sur les émotions associées aux photos.

2. Flickr : url=<http://www.flickr.com/>.

3. Photo.net : url=<http://www.photo.net>.

4. DPChallenge : url=<http://www.dpchallenge.com>.

5. CUHK-DB : url=<http://mmlab.ie.cuhk.edu.hk/CUHKPQ/Dataset.htm>.

6. ImageCLEF : url=<http://www.imageclef.org/>.

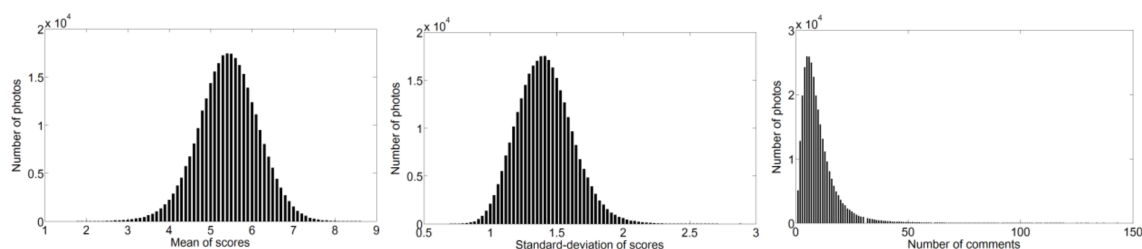


Figure 6.1 – Dispersion des notes moyennes attribuées à chaque photo de la base de données DPChallenge (d’après [Kim et al., 2015]). A gauche, distribution sur toutes les photos des moyennes des notes attribuées à chaque photo ; au centre distribution des variances des notes attribuées à chaque photo ; à droite répartition du nombre de commentaires littéraires par photo.

- **AVA**⁷ *Aesthetic Visual Analysis* est une base de données d’images spécialement développée pour les études d’esthétique. Elle est devenue progressivement la source principale des travaux utilisant des réseaux profonds. Elle est issue de DPChallenge et rassemble 250 000 images, regroupées en 60 catégories, avec une abondante indexation en scores esthétiques d’une part (plus de 200 notes en moyenne par photo), en style photographique (14), en contenu (44 tags sémantiques) ainsi qu’en émotions associées aux photos d’autre-part [Murray et al., 2012]. Elle hérite donc des propriétés statistiques de la base DPChallenge [Talebi and Milanfar, 2017].

On doit s’interroger sur la qualité des images de la base AVA. Nous l’avons dit, AVA est déduite de DPChallenge qui offre des images de très bonne qualité, à haute résolution (généralement de l’ordre de 4000 pixels par ligne). Il n’est cependant pas clair que toutes les études menées avec AVA bénéficient de cette qualité. En effet, la base disponible sur le site AVA est constituée d’images réduites à 640 pixels dans leur dimension la plus grande. Certes il est possible d’accéder à l’image source par un lien qui est fourni dans AVA, mais il n’est pas clair que cette opération de téléchargement de chaque image à haute résolution soit faite systématiquement par les études qui s’appuient sur AVA, et ce d’autant plus que DPChallenge met en place des mécanisme assez efficace de protection des images sources.

Deux types d’adaptations ont été faites avec AVA de façon à créer des ensembles d’apprentissage où les images sont nettement séparées sur des critères esthétiques. D’une part on a séparé un ensemble d’images, réputées belles, d’un ensemble d’images, réputées moins belles, par un seuil $\pm\delta$ autour de la valeur moyenne 5,5 et défini ainsi des ensembles d’apprentissages divers pour des δ de 0, 1 ou 2. D’autre part, on a réduit la base AVA aux seules 10 % d’images les plus belles et 10 % d’images les moins belles. Cette nouvelle base est appelée AVA2 et ne comporte donc plus que 50 000 images environ.

Style	Number	Style	Number
Complementary Colors	949	Duotones	1, 301
High Dynamic Range	396	Image Grain	840
Light on White	1,199	Long Exposure	845
Macro	1,698	Motion Blur	609
Negative Image	959	Rule of Thirds	1,031
Shallow DOF	710	Silhouettes	1,389
Soft Focus	1,479	Vanishing Point	674

Figure 6.2 – Les 14 attributs de style utilisés dans la base de données AVA et leur nombre d’occurrences, d’après [Wang et al., 2016]

- **AADB**⁸ La base *Aesthetics with Attributes Database* est issue du site Flickr. Elle contient 10 000 images annotées par 5 opérateurs qui leur ont attribué une note d’esthétique (sur 5 niveaux) ainsi qu’une note pour chacun de 11 attributs descriptifs de qualité issus d’une liste fournie par des professionnels de la photo (*intérêt du contenu, éclairage, harmonie des couleurs, etc.*).

7. AVA url=<http://www.computervisiononline.com/dataset/1105138637>.

8. AADB url=<http://www.ics.uci.edu/skong2/aesthetics.html>.

La moyenne des 5 notes d'esthétique est prise comme note de référence pour l'esthétique, mais on peut aussi suivre un opérateur unique de façon à conserver éventuellement une esthétique cohérente [Kong et al., 2016].

- **BEAUTY**⁹ est une base de 15 000 images issues de Flickr, mêlant des images de qualité très variée. L'indexation de ces images en fonction de leur qualité est confiée à la communauté des internautes : elle se fait par *crowdsourcing*. Des précautions sont prises pour éviter des dispersions trop grandes des avis, ainsi que des biais dans les résultats statistiques. Pour le premier point les seuls internautes mis à contribution sont issus d'un nombre restreint de pays à forte homogénéité culturelle et leurs résultats sont à posteriori filtrés pour écarter les comportements par trop singuliers. Pour le second point, le stock des images traitées est partagé en trois grandes familles de qualités différentes et les représentations statistiques des échantillons issus de ces familles sont notablement différentes. Enfin des critères différents sont retenus pour décrire les quatre classes : “personnes”, “nature”, “animaux” et “ville” [Schifanella et al., 2015].
- **La base de M. Redi et al.** comporte également des images (100 000) de toutes qualités, mais issues du web à partir de 8 000 requêtes en tirant aléatoirement 20 images par requête (cela évite de n'utiliser qu'une source potentiellement biaisée). La base comporte ainsi 85 % de photos et 15 % de graphiques, dessins, images de synthèse, etc. Ces images sont annotées par des professionnels selon deux indices : la *qualité esthétique* (sur 4 niveaux) et le *type* (“photo” ou “non-photo”) [Redi et al., 2017]. Elle fait également l'objet d'une évaluation publique par le Turc Mécanique d'Amazon, cette évaluation se montrant plus générale, mais moins fiable.
- **La base de l'Université de Tübingen** [Schwarz et al., 2016] est constituée à partir des images de Flickr et comporte 380 000 images. L'annotation varie de 0 à 1. Elle est obtenue en prenant en compte le nombre de vues (\mathcal{V}_i) (c'est-à-dire de clics sur le lien de l'image) et du nombre d'avis favorables (\mathcal{F}_i) (mesuré par le nombre de clics sur la case *favor*) par la formule :

$$S(i) = \alpha \frac{\log(\mathcal{F}_i)}{\log(\mathcal{V}_i)} \quad (6.1)$$

où α permet de fixer à 1 la mesure de la meilleure image. Seules les images telles que \mathcal{F}_i est supérieur à 1 sont gardées dans le décompte. La mesure de qualité est donc celle d'observateurs néophytes mais très nombreux (en moyenne 6 800 votes par photo). La distribution des votes est pratiquement plate entre 0 et 1.

- **La base de l'Université de Iéna : JenAesthetics**¹⁰ La base de données constituée dans le groupe de Computer Vision de l'Université de Iéna [Amirshahi et al., 2013], comme la suivante que nous examinerons ci-dessous, n'est pas une base de données de photos, mais de peintures. Elle a cependant été constituée dans le but précis de comprendre les ressorts du jugement esthétique et de confronter les résultats trouvés à l'épreuve de techniques d'apprentissage automatique. En cela elle rejoint notre propos. Elle est constituée de 1628 photographies de haute qualité d'œuvres d'art collectées à l'occasion du *Google Art Project*, couvrant tous les styles de peinture, des Primitifs à l'art contemporain. Chaque œuvre d'art est accompagnée d'une vingtaine d'évaluations conduites par un panel de 130 observateurs, non spécialistes de l'art, interrogés sur leur appréciation de cinq critères : l'esthétique, la beauté¹¹, la couleur, la composition et le contenu. L'appréciation se fait à l'aide d'un curseur continu entre 0 et 1.
- **La base de données BAM! : Behance Artistic Media dataset**¹² Cette base de données a été constituée à partir du site Behance¹³, un portefeuille de sites spécialisés dans la présentation de professionnels dans le domaine de la production multimédia. On y trouve donc des photos, mais également des dessins, des peintures et des vidéos qui ne relèvent pas directement de notre propos. Behance comporte plus de 65 millions de documents; BAM! en a sélectionné 1,9 millions, dans les catégories suivantes : images de synthèse, bandes dessinées et dessins

9. BEAUTY url=<http://www.di.unito.it/~schifane/beauty-icvsm15/>.

10. JenAesthetics url=<http://www.inf-cv.uni-jena.de/en/jenaesthetics>.

11. Il n'est pas précisé comment les auteurs distinguent esthétique et beauté, et les résultats statistiques montrent que ces variables sont finalement fortement corrélées.

12. BAM! Behance Artistic Media dataset url=<https://bam-dataset.org/>.

13. Behance : url=<https://behance.net/>.

critères	Photo .net	DP Challenge	Image CLEF	BEAUTY	AADB	Redi	AVA	Uni Tübingen
taille ($\times 1000$)	> 1000	650	1000	15	10	100	250	380
qualité esthétique	haute	haute	faible	faible	faible	faible	haute (?)	faible
score esthétique	7	10		3	5	4	2	continu
classes sémantiques	38	\neq thèmes	44	4	non	2	44	non
étiquettes de style	non	non	non	non	11	non	14	non
annotation	non	oui	oui	non	oui	oui	oui	non
origine			Flickr	Flickr	Flickr	web	DP Chal.	Flickr

Table 6.1 – Quelques propriétés des bases de données les plus utilisées pour l’étude esthétique des images. La taille est exprimée en milliers d’images. La qualité esthétique est considérée comme “haute” si les images proviennent de professionnels ou d’amateurs éclairés, comme “faible” si elles sont issues de réseaux sociaux. Le nombre associé au score esthétique indique le nombre de niveaux du résultat de l’évaluation. Les 3 lignes suivantes indiquent quel type d’information est associé à chaque image : le nombre de classes sémantiques, l’existence de caractéristique de style, la disponibilité d’une autre annotation (par exemple les émotions associées).

graphiques et peintures. A chaque image sont attribuées des étiquettes exprimant l’émotion (calme, heureux, triste, inquiet) et la catégorie de l’objet (seuls sont retenues les scènes couvertes par les termes : bicyclette, oiseau, bâtiment, voiture, chat, chien, fleur, personne et arbre). Les annotations ont été obtenues en deux étapes : tout d’abord par *crowdsourcing* d’un nombre limité d’images, puis par classification automatique itérative [Wilber et al., 2017].

- **La base FACD** (*filter aesthetic comparison dataset*¹⁴) C’est une base tirée de AVA pour comparer les qualités esthétiques des divers traitements très souvent utilisés pour “améliorer” les photos déposées sur les sites comme Instagram, [Sun et al., 2017]. Elle est composée de 1260 images originales auxquelles ont été appliqués 22 traitements issus, soit des bibliothèques d’Instagram, soit de la boîte à outil GIMP. Ces traitements procèdent essentiellement à des changements de contraste et des colorisations. Chaque image est affectée à une catégorie parmi 8 (animal, flore, paysage, architecture, nourriture, portrait, scène urbaine, et nature morte), également peuplées. La base de données comporte également 42 240 paires d’images auxquelles sont associées des jugements de préférence attribués par des observateurs humains à l’aide du Turc Mécanique d’Amazon.
- Enfin, **des sites de vente d’œuvres d’art en ligne** : de nombreux sites sont spécialisés dans la vente d’œuvres d’art en ligne ; ils offrent souvent, à côté de peintures ou de sculptures, des photographies (Artspace, Artsper, ArtPhoto, KazoArt, Pixopolitan, Singulart, Ugalery ... par exemple). Ils proposent ainsi des catalogues de quelques centaines à quelques milliers de photos qui se distinguent par leur qualité artistique affirmée. Ces sites regroupent les photos par catégories (paysages, portraits, natures mortes), ils mentionnent les auteurs et annotent souvent les documents de commentaires qui peuvent être utilisés comme méta-data dans des classifications. Les ressources offertes par ces collections sont par exemple discutées dans [Messina et al., 2018]. Notons que ces sites proposent souvent des œuvres beaucoup plus engagées artistiquement que les sites d’amateurs comme DPChallenge et font référence à des critères esthétiques que l’on pourrait qualifier de “savants” en ce qu’ils mobilisent une technique photographique très avancée.

14. FACD (*filter aesthetic comparison dataset*) url=<https://wtwilsonsun.github.io/FACD/>

Chapitre 7

Les méthodes de classification à base de primitives

Les photos artistiques sont des documents de grande dimension. Plus que les autres photographies, elles réclament qu'on les traite avec toute leur résolution et toute la dimension de leurs couleurs. Elles ne se prêtent pas bien à des opérations de sous-échantillonnage ou de codage. Les méthodes présentées ici vont donc choisir d'extraire de l'image un ensemble de primitives, censé retenir les principales qualités de la photo initiale. Les méthodes que nous verrons se distinguent par le choix des primitives retenues. Une approche employée dans le domaine de l'imagerie multimédia consiste à extraire des caractéristiques connues pour bien refléter les propriétés d'une grande diversité de photos : descriptif de l'histogramme de luminosité ou des plans couleurs, distribution spatiale des principales plages colorées, moments géométriques des régions lors d'une décomposition hiérarchique, indicateurs de textures, etc. D'autres approches s'appuient sur des représentations utilisées en vision par ordinateur pour retrouver des formes et des objets : liste de points caractéristiques (SIFT, SURF, Harris, etc.). On peut également s'appuyer sur les recommandations faites dans le domaine de la photo d'art (voir section 4), pour ne conserver que les primitives qui reflètent des critères exprimés de beauté. C'est alors ce que l'on dénomme couramment dans la littérature par les primitives "artisanales" (*hand-crafted*) qui visent à refléter les recommandations des experts. Dans le tableau présenté sur la figure 7.1 on observe un mélange de ces divers types de primitives.

Néanmoins, quelles que soient les primitives choisies, le schéma qu'adoptent ces travaux se fait selon deux étapes :

1. une étape d'apprentissage :
 - (a) une source d'images est identifiée, fournissant pour chaque image une évaluation de la qualité esthétique ;
 - (b) des primitives sont extraites de chacune des images de façon à mettre en avant des propriétés discriminantes des qualités esthétiques des images ;
 - (c) un algorithme d'apprentissage supervisé est mis en place qui détermine l'association judicieuse des primitives de façon à rendre compte du classement esthétique ;
2. une étape de notation d'une photo dont les qualités esthétiques sont inconnues :
 - (a) l'image inconnue est décomposée dans les mêmes primitives que celles utilisées lors de l'apprentissage ;
 - (b) un classifieur, identique à celui qui a été entraîné, mais dont les paramètres sont maintenant fixés, fournit une estimation de la qualité de l'image.

Name	Description
Brightness AVG and STD	(1) Average and standard deviation of the brightness, using the V channel in the HSV space.
Color Variance	(1) Variance of colors in the LAB space.
Contrast	(1) Width of the middle 96% mass of the histogram of the V channel in the HSV space.
#Edges and #Edges L, R, T, B, C	(1) We split the canny map into 16×16 blocks and we compute the number of blocks containing more than 10% of edges. We also compute this number on the left, right, top, bottom and center regions of the image.
Hue Count	(1) Approximation of the number of unique hues [18].
Saturation AVG and STD	(1) Average and standard deviation of the saturation.
Sharpness	(1) Variance of the Laplacian. [20]
Distance to the Center	(2) Distance of the salient region to the center of the image.
Rule of Thirds	(2) Shortest distance of the salient region to a power point.
Salient Hue, Brightness and Saturation	(2) Average hue, brightness and saturation of the salient region.
Salient Sharpness	(2) Sharpness of the salient region.
Salient Size	(2) Size of the salient region.
Salient LOC	(2) We split the image into nine equal parts, and compute the proportion of the salient region in each part. LOC can then take nine values: Top-Left, Middle-Left, Bottom-Right...
Color Difference	(3) Difference of colors in the LAB space between the salient object and the background.
Hue, Saturation and Brightness Difference	(3) Difference of hue, saturation and brightness between the salient region and the background.
Sharpness Difference	(3) Difference of sharpness between the salient region and the background.

Figure 7.1 – Quelques primitives couramment utilisées pour mesurer la beauté d'une photographie. Certaines sont des primitives génériques du traitement de l'image (comme la luminosité ou le contraste), d'autres relèvent plus de critères esthétiques (distance au centre, règle du tiers) (d'après [Simond et al., 2015]).

7.1 Primitives multimedia : le système Acquine et les travaux de R. Datta *et al.*

Ces travaux sont parmi les premiers à avoir abordé le problème de détermination des qualités esthétiques des photos par ordinateur [Datta et al., 2006, Datta et al., 2008, Datta and Wang, 2010]. Ils ont abouti à un service en ligne d'évaluation de la qualité d'une photographie : ACQUINE¹.

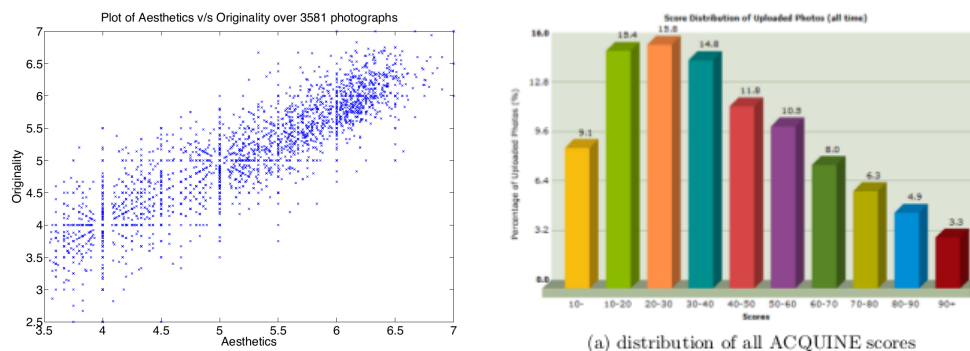


Figure 7.2 – A gauche, relation entre les notes d'esthétique et d'originalité dans la base *Photo.net* (d'après [Datta et al., 2006]). A droite distribution des scores d'Acquine sur l'ensemble des photos traitées (150 000 environ) : les très bonnes notes sont rares et les notes les plus probables sont faibles (d'après [Datta and Wang, 2010]).

R. Datta a choisi des photos issues du site *Photo.net*. Seules ont été considérées les images ayant reçu au moins deux évaluations supérieures à 6 sur le seul critère de l'esthétique. Ainsi environ 3 800 images ont été retenues.

1. Disponible à l'adresse : <http://infolab.stanford.edu/~wangz/project/imsearch/Aesthetics/MIR10/>.

Le résultat attendu de l'algorithme est en deux parties : une classification à deux classes : *low* et *high*, définies à partir d'images de notes inférieures à 4,2 d'une part et supérieures à 5,6 d'autre part, et une note individuelle de 0 à 100, croissante avec la qualité esthétique de l'image.

Les primitives extraites des images sont au nombre de 56 :

- le niveau de gris moyen,
- une mesure de coloration est déduite d'une quantification grossière de l'espace RVB en 64 cubes, puis d'un calcul de distance (dans l'espace Luv) entre cette image et une image de couleurs équi-réparties ;
- les valeurs moyennes de saturation et de teinte ;
- 3 mesures donnant l'accord à la règle du tiers dans les 3 composantes HSV ;
- 2 indicateurs de *familiarité* qui comparent l'image à un ensemble de référence et calculent la distance moyenne aux 20 ou 100 plus proches ;
- 12 termes de texture sont déduits de décompositions en ondelettes des composantes HSV ;
- la taille et le facteur d'aspect sont également retenus ;
- une décomposition par régions (classification par *k*-moyennes dans l'espace chromatique) ne retient que les 5 régions les plus importantes qui sont décrites par leur taille et leur point de couleur moyen. De plus 2 indicateurs expriment la distribution des teintes et leur complémentarité sur la roue des couleurs ;
- 3 indicateurs de profondeur de champ permettant de distinguer les images nettes sur fond flou ;
- enfin un facteur qui mesure la convexité et la régularité des formes.

Le classifieur utilisé est un SVM classique, entraîné par permutation croisée de l'ensemble d'apprentissage et de l'ensemble de test après sélection des primitives (30 sur 58 sont conservées). Enfin une régression linéaire sur l'expression polynomiale des facteurs retenus permet de déduire une formule approchée de l'esthétique de l'image à partir des seules valeurs des primitives.

Notons que le taux de bonne classification *low* et *high* est de l'ordre de 63 %, ce qui reste assez modeste, avec une variance de l'ordre de 0,5 sur la notation à 7 niveaux. Il est difficile de tirer des conclusions du choix des primitives par les algorithmes de sélection : arriveraient en premiers la description des régions, le niveau de gris moyen et la convexité des formes, mais ces conclusions ne sont pas vraiment convaincantes.

Le système a été mis en expérimentation en ligne sous le nom d'ACQUINE² qui a connu un certain succès chez les photographes amateurs (voir les résultats de la figure 7.2 à droite). Parmi les plus intéressants résultats, notons qu'un forum a été investi par les internautes qui ont cherché à retrouver les critères de notation du moteur d'inférence en soumettant des successions d'images de façon à avoir soit le score le plus élevé, soit le plus faible.

7.2 Contours et distance chromatique : l'approche de Y. Ke et al.

A peu près en même temps que les travaux précédents ont été proposés les travaux de Y. Ke [Ke et al., 2006] dont l'objectif est d'affecter toute photo à l'une des deux classes *Amateur* ou *Professionnel*. Notons tout de suite que d'autres critères entrent en jeu dans cette classification que les critères esthétiques, puisque certains éléments relèvent plus spécifiquement de ce que l'on appelle *la qualité* de l'image.

La base d'images utilisée est celle de *DPChallenge.com* ; seules les photos ayant reçu au moins 100 voix sont prises en compte et la moyenne des notes reçues leur est attribuée comme note unique. Les seules photos appartenant aux 10 % les mieux notées sont conservées pour former la classe P : *Professionnel* et les 10 % inférieurs pour la classe A : *Amateur*, constituant ainsi une base de données B de 6 000 photos (voir figure 7.3).

2. ACQUINE = *aesthetic quality inference engine* [Datta and Wang, 2010].

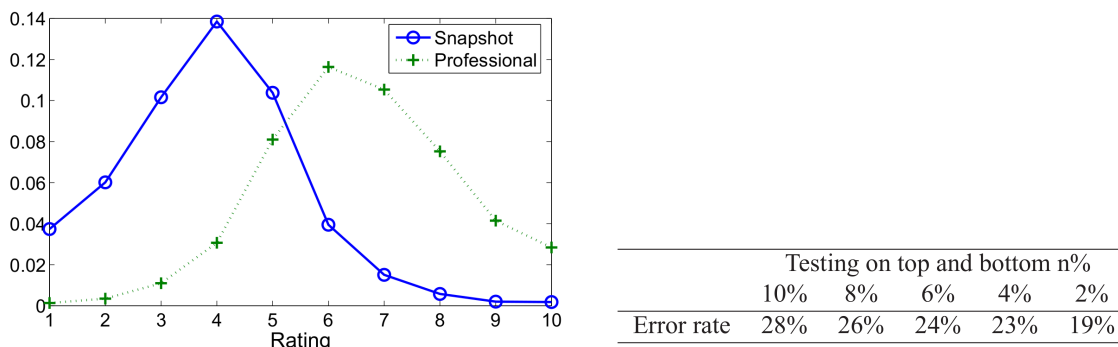


Figure 7.3 – *A gauche, les images P ou A ne sont pas totalement séparées dans les notes qu’elles reçoivent sur DPContest.com. A droite, en choisissant de plus en plus strictement (de gauche à droite) les photos sélectionnées comme P ou A dans la base de données, on améliore un peu les performances du classifieur, mais pas beaucoup, (d’après [Ke et al., 2006]).*

Pour être normalisées, les images sont ramenées à un format carré. Les primitives qui ont été utilisées relèvent du domaine du traitement de l’image :

- une carte des contours détectés par un laplacien est comparée à la carte moyenne des contours des deux répartitions P et A ; une boîte englobante de la majorité des contours est évaluée et rapportée à la taille de l’image (les “belles” images ont tendance à avoir leurs contours concentrés loin des bords) ;
- après une quantification des plans de couleurs, on recherche les n images de B dont les histogrammes sont les plus proches les uns des autres (aux k-plus-proches-voisins). Si n_P et n_A sont les images issues de P et de A, on utilise $n_P - n_A$ comme indicateur de qualité ;
- un indice de teinte est dérivé d’un décompte du nombre de couleurs dont la teinte est supérieure à 5 % de la dynamique des teintes ;
- un indice de contraste et un indice de brillance.

De ces primitives sont tirées cinq mesures de qualité qui sont combinées au sein d’un classifieur bayésien naïf, éventuellement suivi d’un classifieur Real-AdaBoost.

Les résultats obtenus par cette méthode restent modestes et les enseignements sur les primitives les plus importantes pour le jugement esthétique demeurent rares.

7.3 Objet et fond : l’approche de Y. Luo et X. Tang

Ces travaux sont rapportés dans [Luo and Tang, 2008]. Ils se mettent dans le prolongement des travaux de Y. Ke *et al.*, mais adoptent une stratégie fondée sur la distinction entre l’objet principal de la photo et son fond. La séparation s’appuie sur une hypothèse que l’on pourra mettre en question : le sujet est généralement clair et au point tandis que le fond est plus sombre et flou. L’algorithme de détection reprend donc ce principe et mesure localement diverses statistiques de l’intensité et de ses dérivées horizontales et verticales.

Les auteurs peuvent alors mettre en œuvre des critères différents sur l’objet et sur son fond, ainsi que sur le contraste qui les différencie. Ainsi le fond gagnera-t-il à être assez uniforme avec une palette chromatique réduite, tandis que l’objet bénéficiera d’une plus grande richesse harmonique mesurée dans l’espace HSV par des histogrammes marginaux. La qualité de la palette chromatique sera obtenue en comparant ces histogrammes marginaux à ceux des “belles” images de la base d’apprentissage les plus proches. Enfin des règles de disposition spatiale sont également mesurées, comme la règle du tiers comptée à partir du centre de gravité de l’objet.

La même base de données d’apprentissage issue de DPChallenge, modifiée par Y. Ke *et al.* est

utilisée et l'apprentissage se fait sur 3 000 photos de bonne qualité et 3 000 mauvaises photos. Le classement se fait soit par Bayes, soit par SVM suivi d'AdaBoost. Les résultats, sur les mêmes images, sont nettement supérieurs à ceux de Y. Ke *et al.* (figure 7.4). Les auteurs attribuent ce gain au mérite des primitives mieux adaptées à la photographie.

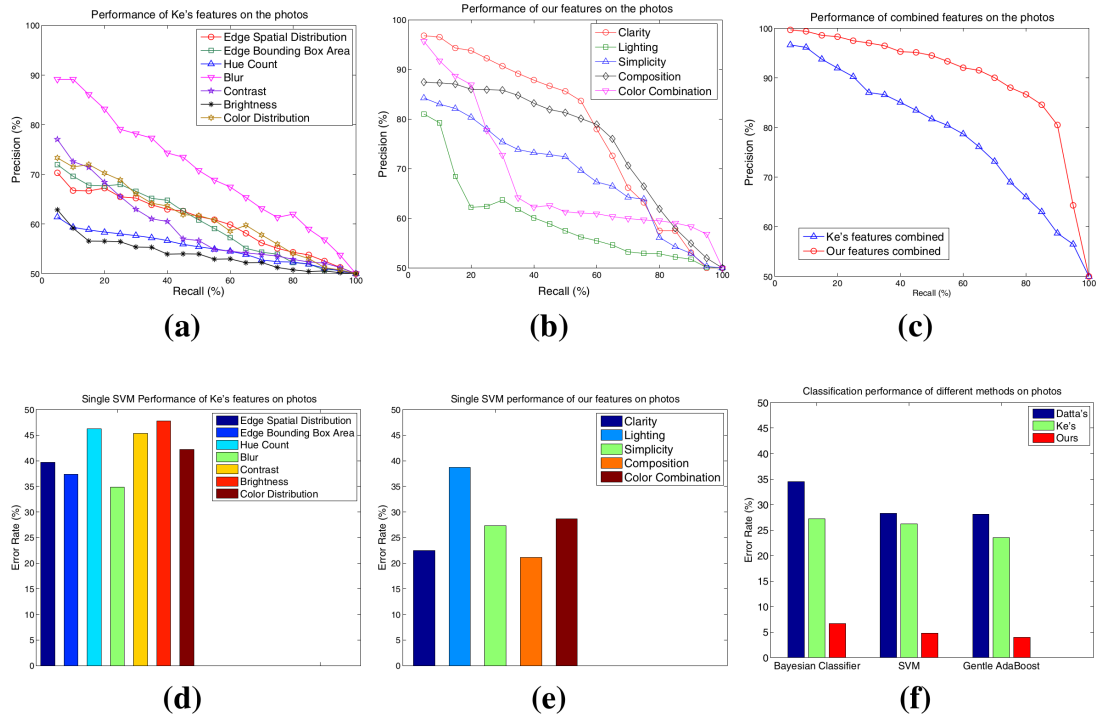


Figure 7.4 – *Comparaison des résultats (Précision en fonction du Rappel) de Y. Ke et de Y. Luo. Classification bayésienne (a) avec les primitives de Ke, (b) avec les primitives de Luo, (c) avec les primitives combinées; (d) et (e) rôle des primitives de Ke et de Luo dans la classification (pour Ke, c'est l'intensité qui est la grandeur dominante, pour Luo c'est l'éclairage); (f) performances des divers classificateurs (les primitives de Luo induit une plus faible erreur que celles de Ke ou de Datta, quel que soit le classifieur) (d'après [Luo and Tang, 2008]).*

7.4 Des primitives de haut niveau : l'approche de S. Dhar *et al.*

[Primitives de haut niveau] Ces travaux se placent également en parallèle de ceux de Y. Ke, mais les auteurs ont choisi d'utiliser des primitives de haut niveau, décrivant explicitement la composition de l'image, son contenu et l'illumination de façon interprétable par un interprète humain [Dhar et al., 2011]. La figure 7.5 indique de quelle façon ces primitives sont calculées à partir des primitives de Y. Ke. La détection de nombreuses primitives est complexe (détection des visages, des animaux, reconnaissance des types de scène - montagne, intérieur ...) et nécessitent la mise en œuvre de sous-systèmes de classification (17 SVM sont utilisés pour apprendre des tâches particulières), avec leurs propres bases d'apprentissage (par exemple un système pour chacun des 10 types d'animaux identifiés). Ils retiennent ainsi 26 attributs.

L'étude montre ensuite les bénéfices de cette lourde approche lors de deux tâches particulières : l'une (qui nous concerne) sur l'évaluation de la qualité esthétique, l'autre sur l'évaluation de l'intérêt de l'image. Donnons ici les résultats sur la seule tâche d'évaluation esthétique. Ils sont regroupés sur

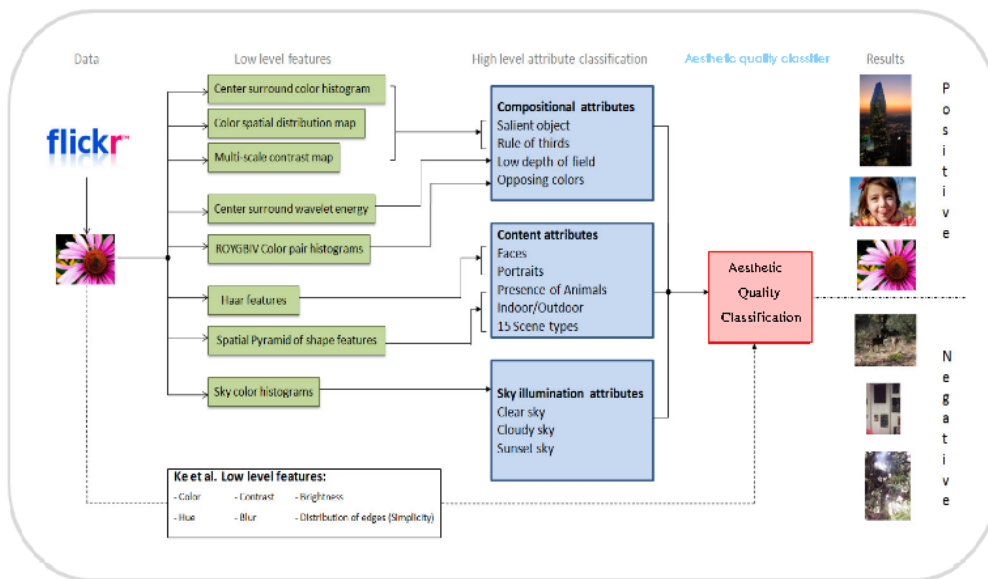


Figure 7.5 – Le schéma de calcul des primitives de haut niveau de S. Dhar et al. utilise les détections faites au bas niveau par le pré-traitement de Y. Ke (le schéma ci-dessus décrit l'évaluation de "l'intérêt" de l'image, mais un schéma identique permet d'évaluer "l'esthétique". Il ne diffère pas de celui-ci dans son architecture mais subit un entraînement différent de façon que la dernière couche de neurones fournisse un score d'esthétique et non d'intérêt (d'après [Dhar et al., 2011]).

la figure 7.6. Les auteurs préconisent l'emploi d'un classifieur SVM qui donne les meilleurs résultats sur une base 16 000 images issues de DPChallenge.

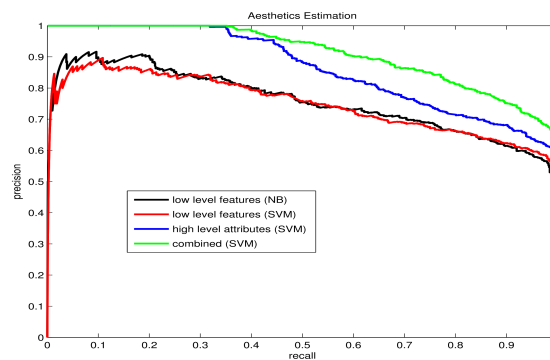


Figure 7.6 – Comparaison des résultats de l'approche de Y. Ke et de l'approche de Y. Dhar en termes de Rappel et Précision. Les courbes marquées low level ont été obtenues avec les primitives de Y. Ke. La courbe marquée high level utilise les primitives de Dahr et la courbe marquée combined utilise toutes les primitives. L'expérience avec les primitives de bas niveau a été conduite avec le classifieur recommandé par Y. Ke (Bayes naïf, noté NB), ou avec un classifieur SVM (d'après [Dhar et al., 2011]).

7.5 La bibliothèque de primitives esthétiques de H.-H. Su et al.

Dans les travaux présentés dans [Su et al., 2012], on se propose d'apprendre un catalogue de configurations spécifiques de photographies esthétiques à partir de belles photographies de DPChallenge.

Pour cela on décompose les photos selon des mailles fixes, déterminées une fois pour toutes, reflétant les principales échelles de composition des photos (figure 7.7). Sur ces maillages sont mesurées des statistiques particulières de certaines primitives (moyenne et contraste entre les mesures faites sur les sites noirs et les sites blancs de la figure 7.7). Ces primitives concernent la couleur (RVB ou HSV), la distribution des contours (détecteur HOG (*histogram of oriented gradient* [Dalal and Triggs, 2005])), une mesure de texture dénommée LBP et une mesure de salience. Le calcul des diverses configurations possibles, par exemple dans la subdivision (d) de 6×6 , est obtenu par un tri des 36 fenêtres de mesure, permettant d'obtenir très rapidement les configurations qui optimisent les contrastes et leur distance aux fenêtres de la bibliothèque.

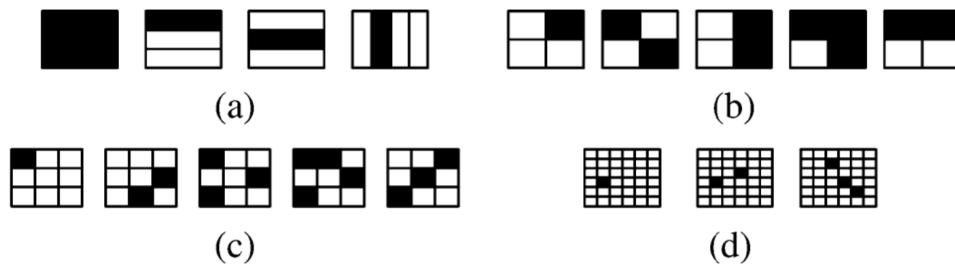


Figure 7.7 – Les fenêtres d'analyse proposées par H.-H. Su et al., qui fourniront la bibliothèque des primitives esthétiques. L'image est divisée selon les 4 configurations décrites dans les sous-figures (a) à (d). Pour chaque configuration sont calculées les propriétés particulières des zones noires par rapport aux zones blanches (d'après [Su et al., 2012]).

Cette technique se montre particulièrement efficace pour proposer un recadrage de l'image autour des zones les plus intéressantes, ce qui est l'objectif principal de ce travail.

7.6 Le système OSCAR de L. Yao et al.

Le système OSCAR (*on-site composition and aesthetic feedback through exemplars*) [Yao et al., 2012] développé à l'université de Pennsylvania se place dans la lignée des travaux de R. Datta et al. [Datta et al., 2008, Datta and Wang, 2010], mais s'en dégage par une architecture audacieuse et une approche originale. Le système n'a malheureusement pas été conduit à son terme, laissant plusieurs pans de développement annoncés mais non réalisés et un ensemble assez incomplet.

L'idée de base consiste à porter assistance à un photographe (muni d'un simple téléphone mobile) en le mettant en relation lors de sa prise de vue avec un système centralisé qui lui transmet des recommandations sur les paramètres à employer pour la photo qu'il souhaite prendre. Le schéma global du système est présenté sur la figure 7.8. Le système centralisé dispose d'une base de données d'images (issues de Photo.net), dotées d'une évaluation selon des critères esthétiques.

L'utilisateur transmet l'image qu'il souhaite prendre dans le nuage (*cloud*) où une analyse est faite de l'image sous plusieurs aspects :

- les couleurs sont extraites par classification dans l'espace HSV ;
- des descripteurs de compositions sont extraits en nombre de 5 : organisation diagonale, verticale, horizontale, centrale ou texture. Ces descripteurs s'appuient sur des détecteurs de contours suivis de filtrage par pertinence ;
- des éléments de Gestalt sont utilisés pour extraire des motifs significatifs sur des critères de parallélisme, similarité et continuité ;
- le bruit granulaire est estimé dans l'image source.

Muni de ces éléments de description, les images les plus semblables sur chacun de ces critères séparément, sont sélectionnées par le serveur qui les propose comme exemples à l'utilisateur. La sélection est faite par des processeurs en parallèle. Le moteur ACQUINE de Datta est l'un d'eux,

ainsi que le classifieur SIMPLiCity.

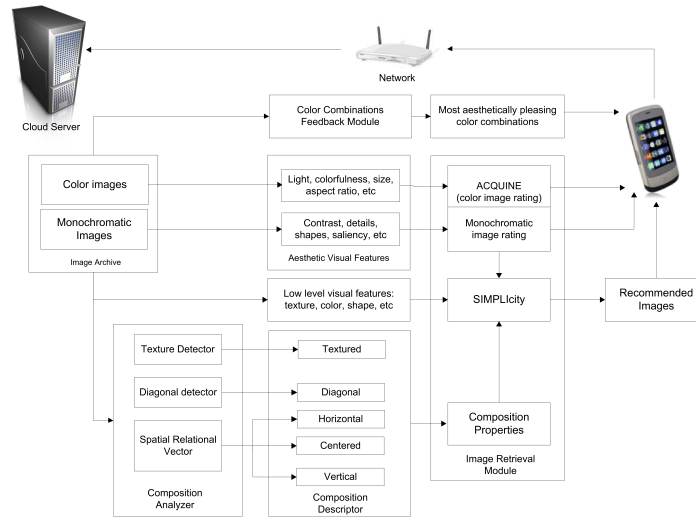


Figure 7.8 – Schéma de principe du système OSCAR, d’après [Yao et al., 2012]. Ce système fait une grande place aux règles de composition et d’harmonie chromatique pour évaluer la qualité des images de la base de données qui seront fournies en exemple pour l’éducation du photographe.

7.7 Un système embarqué : l’approche de K.Y. Lo et al.

Dans les articles [Lo et al., 2012, Lo et al., 2013], une technique de détermination de la qualité esthétique est développée, avec l’objectif de l’embarquer dans un appareil, c’est-à-dire d’assurer une complexité de calcul modeste et une faible consommation électrique. Ces travaux sont à rapprocher de ceux de L. Yao et al. (vus ci-dessus) ou ceux de W.N. Wang et al. [Wang et al., 2014] qui cherchent également à embarquer le système, d’évaluation dans un téléphone mobile. Ils ont donc recours à la puissance du *cloud* pour y transporter les plus gros calculs, ne gardant dans le téléphone que l’interface avec l’utilisateur. Les techniques employées par K.Y. Lo et al. sont assez proches de celles de l’étude de S. Dhar et al. (section 7.4). On attribue 5 notes à chaque image jugeant : la couleur, la saturation, la composition, la richesse et le contraste (voir figure 7.9 à gauche). Un apprentissage est fait sur 2 classes : l’une des images de haute qualité (la classe dénotée *P* ci-dessus), une autre sur la classe de basse qualité (la classe dénotée *A* ci-dessus). Le classifieur est un SVM.

Les primitives détectées sont :

- des descripteurs de couleur semblables à ceux utilisés dans l’étude précédente, construits sur les histogrammes fortement quantifiés (voir figure 7.9 à droite) ;
- des indicateurs de teinte, saturation et luminance obtenus à partir de la moyenne des plans HSV déduits de l’image, puis un indicateur global égal à la somme des 3 moyennes précédentes ;
- des indicateurs de contours sur les 4 canaux H, S, V et H+S+V ;
- des indicateurs de texture calculés après division de l’image en tiers horizontalement et verticalement et intégration des intensités d’une part et des gradients d’autre part dans chacune de ces bandes ;
- un indicateur de flou, un indicateur de canal noir et un décompte des zéros de l’histogramme HSV quantifié.

C’est donc au total une trentaine de primitives qui sont utilisées. La base de données d’images est celle de l’Université de Hong-Kong. Plus de 10 000 images sont utilisées. Les images sont de très petite taille (moins de 250 000 pixels). Elles comportent des photos de types “professionnel” et “amateur”, classées sous 7 rubriques différentes selon leur thématique.

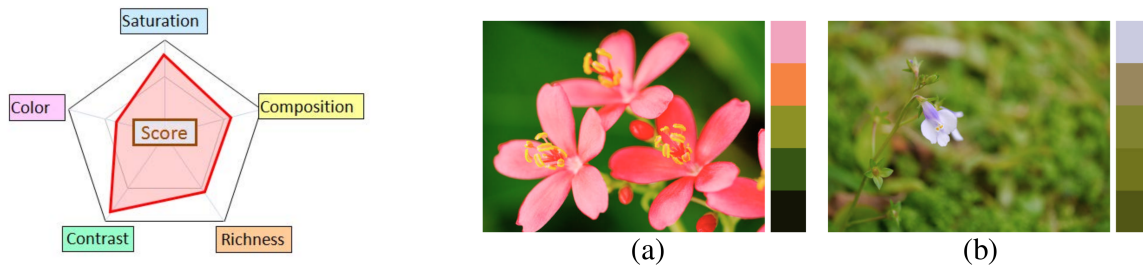


Figure 7.9 – L’approche de Lo et al.. A gauche, la note finale se déduit de 5 notes partielles dont l’importance varie selon la catégorie des images traitées. A droite, la palette est réduite à 5 teintes dominantes et permet de distinguer les images avec une harmonie agréable (en (a)) et celles moins plaisantes (en (b)), (d’après [Lo et al., 2012]).

7.8 Descripteurs génériques de la vision : l’approche de L. Marchesotti *et al.*

Ces travaux ont été présentés dans [Marchesotti et al., 2011, Marchesotti et al., 2013]. Sans chercher à coder des “règles de beauté”, les auteurs utilisent des primitives de description génériques développées pour l’indexation et la fouille d’images. Ce sont soit des sacs de mots visuels (BOVW = *Bags of visual words*) soit des vecteurs de Fisher (FV). Ils opèrent au second niveau de représentation du signal, s’appuyant sur des primitives de bas niveau (en l’occurrence des SIFT³ d’une part et des descripteurs de couleur d’autre part). Les SIFT et les descripteurs de couleur sont traités séparément. Ils sont réduits par une analyse en composantes principales (PCA) en 64 dimensions puis regroupés dans des mots obtenus en concaténant en une chaîne unique soit leur nombre d’occurrences (BOVW) soit leurs statistiques (à l’aide d’une modélisation par mélange de gaussiennes) jusqu’à l’ordre 2 (FV).

Une information sur la composition de l’image peut être donnée par représentation pyramidale mais ne semble pas donner de meilleures performances.

La classification se fait par SVM et l’apprentissage à l’aide d’un algorithme stochastique de descente de gradient. La classification est conduite séparément sur les SIFT et sur la couleur, et leurs scores sont ensuite moyennés. Comme précédemment, deux classes sont recherchées : *Low* et *High*. La base de données utilisée est tout d’abord celle de *Photo.net* un peu modifiée pour en retirer les cadres de certaines images (résultant en une sélection de 3 200 photos environ). Les auteurs ont ensuite utilisé la base CUHK (12 000 images). Ces travaux ont finalement contribué à la création de la base de données annotée AVA, spécialement dédiée à l’étude des propriétés esthétiques des images [Murray et al., 2012].

En conclusion de ces travaux, les auteurs soulignent les très bonnes performances des primitives génériques associées à une classification de second niveau (en particulier les FV). Cette approche permet de dépasser les performances données par les seules primitives “esthétiques” en confiant au classifieur FV le soin d’incorporer des propriétés de haut niveau dans la décision. Notons enfin que cette technique de recherche associant esthétique et sémantique a fait l’objet d’un brevet [Murray et al., 2013].

7.9 Harmonie chromatique : les travaux de P. Lu *et al.*

Les travaux présentés dans [Lu et al., 2014a, Lu et al., 2015a, Lu et al., 2016] ont un objectif plus réduit que les précédents puisqu’ils visent uniquement à déterminer l’esthétique de la palette d’une

3. Les SIFT (*Scale-invariant feature transform*) sont des détecteurs de points caractéristiques multi-échelle.

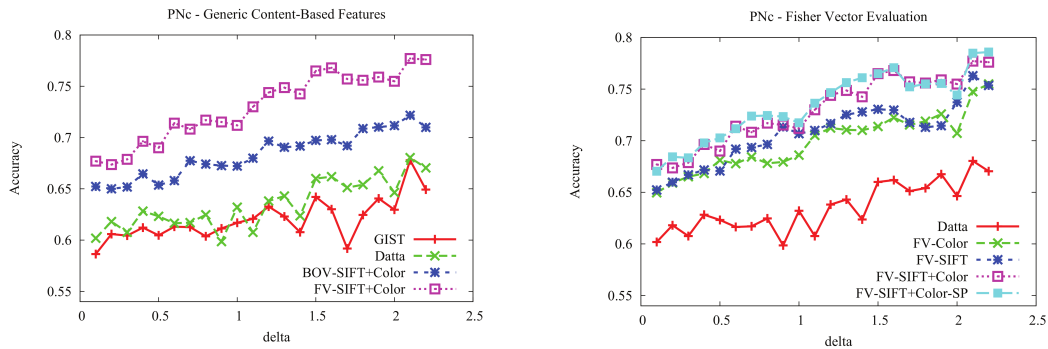


Figure 7.10 – L’approche de L. Marchesotti et al. A gauche, comparaison des performances de reconnaissance sur la base de données Photo.net, des diverses stratégies de représentation des images. La courbe notée “Datta” est celle décrite en section 7.1. La variable “delta” portée en abscisse correspond à l’écart entre les notes des images “Low” et “High” imposé lors de l’apprentissage. A droite, rôle des descripteurs SIFT ou de couleur dans la performance du classifieur : ils sont assez complémentaires. La courbe marquée SP utilise une décomposition pyramidale. Pour une plus grande complexité, elle n’offre pas des performances supérieures, d’après [Marchesotti et al., 2011, Marchesotti et al., 2013].

photographie.

P. Lu *et al.* ont choisi d’apprendre par une technique automatique les règles de l’harmonie. Ils comparent un apprentissage automatique avec les règles d’harmonie de Moon et Spencer et de Matsuda présentées dans la section 4.5.2.

Pour cela, ils s’appuient sur la base de données annotée AVA [Murray et al., 2012] (250 000 images), triée par grandes thématiques (fleurs, paysages, etc.). Huit thématiques sont considérées. Pour chaque thématique, ils distinguent deux ensembles : celui des images harmonieuses et celui des images banales. Puis, après avoir décomposé chaque image en une mosaïque de vignettes, ils utilisent une approche bayésienne (la méthode *Latent Dirichlet Allocation* (LDA)) pour déterminer les associations de vignettes les plus fréquemment attribuées aux “belles” images. Pour cela, ils emploient un régulariseur laplacien, (normalisé ou non), puis une régression par LASSO pour estimer le score d’esthétique de l’image [Lu et al., 2016].

Leurs résultats montrent des performances de reconnaissance des “belles” images supérieures à celles des performances obtenues avec les critères de Moon et Spencer, mais surtout avec les critères de Matsuda (voir la figure 4.13) qui apparaissent les plus mauvais (figure 7.11).

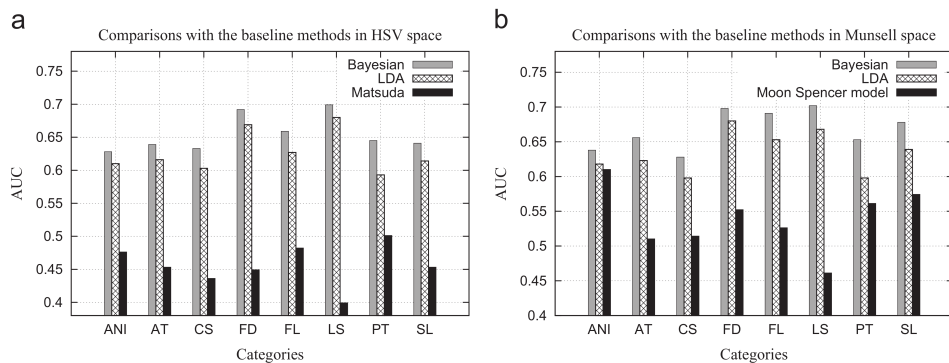


Figure 7.11 – Apprentissage de l’harmonie des couleurs sur la base AVA. A gauche avec les critères de Matsuda dans l’espace HSV, à droite avec les critères de Moon et Spencer dans l’espace de Munsell, d’après [Lu et al., 2014a].

7.10 Réseaux sociaux et *crowdsourcing* : l'approche de R. Schifanella *et al.*

Les travaux présentés dans [Schifanella et al., 2015] se distinguent des précédents sur plusieurs points. Tout d'abord, ils font appel à des images ordinaires, la base BEAUTY issue du réseau social Flickr indexée en fonction de leur qualité par la communauté des internautes et *crowdsourcing* avec de grandes précautions pour éviter des biais dans les résultats statistiques (voir 6.2). La question posée : *How beautiful is this picture ?* se réfère aux seules qualités esthétiques et non au contenu de l'image. La note est attendue parmi 5 : *inacceptable, avec des défauts, ordinaire, professionnelle, exceptionnelle*. Enfin des critères différents sont utilisés pour les diverses classes : "personnes", "nature", "animaux" et "ville".

Les primitives utilisées sont des primitives visuelles de couleur (à partir des critères issus de la psychologie affective (voir la section 4.5.1 [Machajdik and Hanbury, 2010]), de texture et de disposition spatiale (règle du tiers, distribution des gradients). La classification se fait par régression quadratique partielle (PLSR = *Partial least square regression*).

La mesure de qualité obtenue pour une image inconnue à l'aide du classifieur précédent peut être comparée à la note de popularité de l'image sur Flickr. Elle ne montre aucune relation. Elle peut également être comparée à celle obtenue par des experts (la méthode de Marchesotti, section 7.8). Elle montre alors une tendance assez nette à surnoter les photos.

7.11 Prise en compte de commentaires : les travaux de J. San Pedro *et al.*

Une approche très originale mérite ici d'être évoquée quoiqu'elle nous éloigne sensiblement des méthodes automatiques d'évaluation de la qualité à partir des propriétés iconographique. C'est celle rapportée dans [San Pedro et al., 2012]. L'idée développée consiste non pas à traiter l'image mais à examiner les commentaires qui accompagnent l'image. Les auteurs utilisent DPChallenge qui fournit à la fois les images et les commentaires associés déposés par les internautes.

Pour ce faire les auteurs identifient des mots de vocabulaire dotés de significations variées : certains désignent des propriétés de l'image : "éclairage", "composition", "couleurs", d'autres des valeurs positives ou négatives de jugement des internautes. Collectant toutes les opinions portées sur la photo, les auteurs déterminent, pour chaque attribut, le sentiment qui ressort globalement de ces avis. Cela est fait à l'aide d'un processus markovien qui estime un état caché (la qualité de la photo). L'apprentissage des dépendances entre critères et beauté est fait de façon supervisée par maximisation d'entropie selon le schéma popularisé par le système de traitement du langage naturel *OpinionMiner* [Jin et al., 2009].

Cette stratégie est employée pour réordonner les résultats d'une requête sur une base d'images, de façon à placer en tête des retours les images les mieux classées par les critères esthétiques ainsi dégagés. Une courte expérimentation montre que la sélection s'appuyant sur l'analyse des commentaires est plus pertinente qu'une sélection sur les seuls mots clefs de la requête ou qu'une sélection utilisant des critères esthétiques tirés de l'image (suivant les algorithmes proposés dans [Datta et al., 2006], section 7.1 et dans [San Pedro and Siersdorfer, 2009]).

Chapitre 8

Les systèmes à base de Réseaux de Neurons Profonds

Les techniques à base de Réseaux de Neurons Profonds sont l’objet d’une grande attention depuis quelques années chaque fois que l’on cherche à reproduire une prise de décision complexe ordinairement assurée par un expert. L’évaluation esthétique relève tout à fait de ce critère et l’on a donc vu plusieurs travaux s’engouffrer dans cette voie. La référence [Deng et al., 2017b] donne une bonne revue des travaux disponibles jusqu’aux premiers mois de 2017.

Comme nous l’avons dit les RNP proposent une approche “de bout en bout” pour laquelle l’image à évaluer est directement mise en entrée du système. La couche d’entrée dispose donc d’un neurone par pixel, ou plutôt de 3 neurones par pixel puisque l’on traite généralement des images en couleur. Les premières couches (en nombre variable, de 3 à 5), sont traditionnellement des couches de neurones convolutifs¹ qui permettent de réduire progressivement la taille des cartes traitées en reliant un neurone de la couche n à un petit nombre (3×3 , 7×7) neurones de la couche $n - 1$. On peut aussi réduire la dimension des couches par des opérations de mutualisation (*pooling*) qui remplacent le sous-échantillonnage des couches convolutives par d’autres fonctions de décimation (fonction “sup”, “médian” ou “inf”). Les couches suivantes sont souvent totalement connectées, de façon que chaque neurone de la couche n soit lié à tous les neurones de la couche $n - 1$. La dernière couche comporte autant de neurones que de classes traitées (par exemple 2 classes pour un jugement “Amateur” ou “Professionnel”, ou 5 classes pour un jugement à 5 niveaux). Les architectures utilisées sont celles disponibles en ligne : AlexNet, ResNet, VGG, Inception ou MobileNet par exemple.

Les limitations des RNP pour l’évaluation de l’esthétique des images sont étroitement liées à l’exigence d’avoir des images de bonne qualité disposant d’une finesse de détails et d’une richesse de champ compatible avec la qualité des photos professionnelles. Cela devrait se traduire en pratique par :

1. une couche d’entrée de plusieurs millions de neurones, généralement triplée pour rendre compte de la couleur ;
2. un nombre de connexions considérable à l’intérieur du réseau, conduisant à des temps de calcul extrêmement longs pour l’apprentissage des paramètres et l’obligation de disposer de bases de données annotées gigantesques pour conduire cet apprentissage.

Les systèmes opérationnels en 2017 fonctionnent surtout sur des imageries de taille réduite (généralement inférieure à 256×256) et ne rendent compte de la finesse à haute résolution, lorsqu’ils s’en préoccupent, que par des pirouettes (voir les sections 8.1.1 et 8.1.2 par exemple). Seul le travail présenté en section 8.1.3 aborde ce problème de front. Pourtant, une petite taille fait perdre d’importants détails pour juger la qualité de l’image et l’on ne peut guère faire confiance à des résultats obtenus sur de si médiocres représentations.

1. Les réseaux convolutifs sont désignés dans la littérature par l’acronyme CNN (*convolutionnal neural network*).

Au-delà de cet aspect lié à la taille de l'image, se pose le problème de son format (ou facteur d'aspect). Les images peuvent se présenter selon des formats très différents tandis que le réseau de neurones est par nature de taille fixée (généralement carrée que nous désignerons par $N \times N$). Il est classique de répondre à ce problème en choisissant l'une des 4 pistes suivantes (illustrées sur la figure 8.1) :

1. en sélectionnant une sous-image $N \times N$ au sein de l'image (opération de recadrage, *cropping*) ;
2. en déformant l'image pour la ramener au format $N \times N$ (opération de rééchantillonnage) ;
3. en réduisant l'image de façon que sa plus grande dimension vaille N et en complétant la plus petite dimension par l'image repliée (opération de remplissage, *padding*) ;
4. en utilisant une version inversée de l'image dans l'opération de *padding* précédente, de façon à éviter les discontinuités lors du repliement (opération de remplissage après retournement, *padding/flipping*).

D'un point de vue de la conservation du signal, le remplissage après retournement préserve le mieux à la fois les propriétés géométriques et les propriétés spectrales.

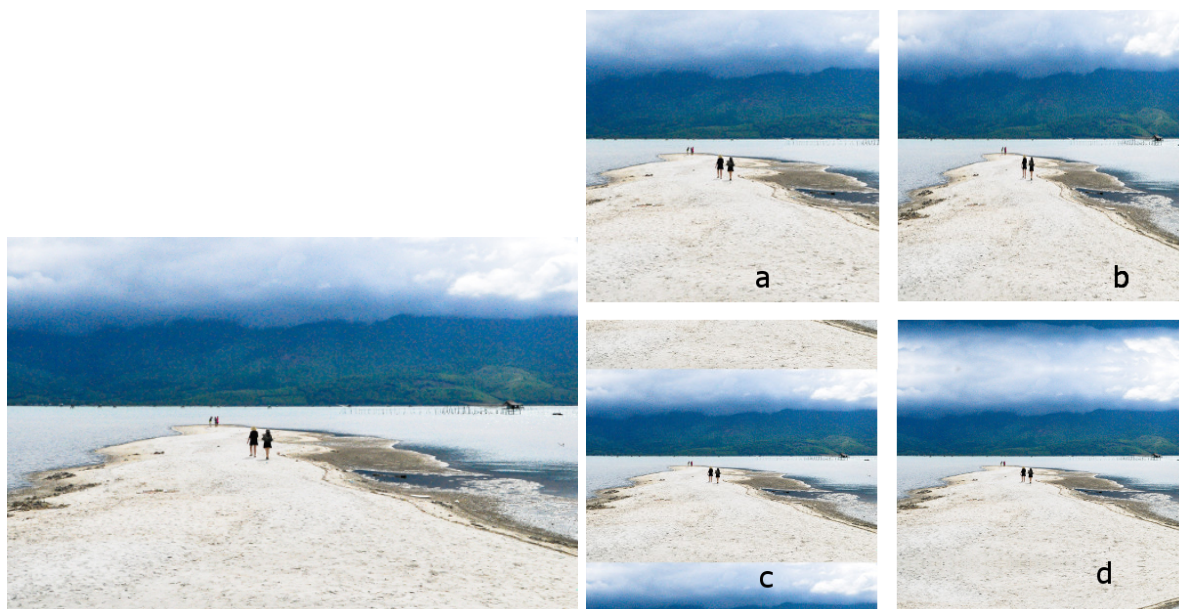


Figure 8.1 – On peut adapter une image rectangulaire (à gauche) à la couche d'entrée carrée d'un RNP, en a) par recadrage, *cropping*, en b) par rééchantillonnage affine, en c) par remplissage, *padding*, en d) par remplissage après retournement, *padding after flipping*). C'est cette dernière solution qu'il faut privilégier.

Examinons maintenant quelques uns de ces travaux.

8.1 Les RNP dédiés à l'évaluation esthétique

8.1.1 Le système RAPID de X. Lu *et al.*

Dans [Lu et al., 2014b, Lu et al., 2014c, Lu et al., 2015b] est présenté le premier système utilisant l'apprentissage profond (RNP (réseaux de neurones profonds) ou DNN (*Deep Neural Network*)) : RAPID (*R*ating *P*ictorial *a*esthetics using *D*eep *l*earning). L'usage des DNN doit permettre de supprimer l'étape de choix des primitives placées à l'entrée des systèmes de classification, en nourrissant directement le classifieur par l'image. Celle-ci est alors traitée par une succession de réseaux convolutifs qui réduisent progressivement la taille du signal avant de le traiter par des couches de réseaux totalement

δ	[Marchesotti et al., 2013]	SCNN	AVG-SCNN	DCNN	RDCNN
0	66.7 %	71.20 %	69.91 %	73.25 %	74.46 %
1	67 %	68.63 %	71.26 %	73.05 %	73.70 %

Table 8.1 – Résultats de l'apprentissage par RNP. Le paramètre δ est le même que le delta de la figure 7.10. Le réseau SCNN est à une seule colonne. AVG-SCNN est le résultat du moyennage de 2 réseaux indépendants. Le réseau DCNN est celui qui utilise 2 colonnes couplées à la dernière couche, tandis que le réseau RDCNN est celui qui est régularisé par l'information de "style", d'après [Lu et al., 2014b].

connectés. Ainsi le RNP extrait lui-même les primitives les plus pertinentes. Le travail présenté dans [Lu et al., 2014c] prend en compte de façon très originale la taille et la diversité des formats d'images.

Comme dans tous les autres systèmes qui utilisent l'architecture VGG-Net², la taille est ramenée à la dimension 256×256 , sur 3 canaux, par un zoom et l'image est normalisée en taille et en aspect en la complétant par des zéros. Cela constituera le RNP principal de traitement. Il est alors utilisé un second RNP pour traiter une vignette, elle aussi de taille $256 \times 256 \times 3$, dont la position est tirée aléatoirement, à pleine résolution, dans l'image originale. Pour rendre compte de la richesse de l'image, lors de l'apprentissage, 50 vignettes sont ainsi tirées. Chaque vignette, associée au canal principal, sera alors traitée par le double réseau et l'on retient celle qui donne le meilleur score. Ces deux RNP partagent finalement la dernière couche totalement connectée des réseaux pour fournir un score unique : 0 ou 1 (figure 8.2).

Ces travaux exploitent la base de données AVA pour l'apprentissage (voir section 6.2).

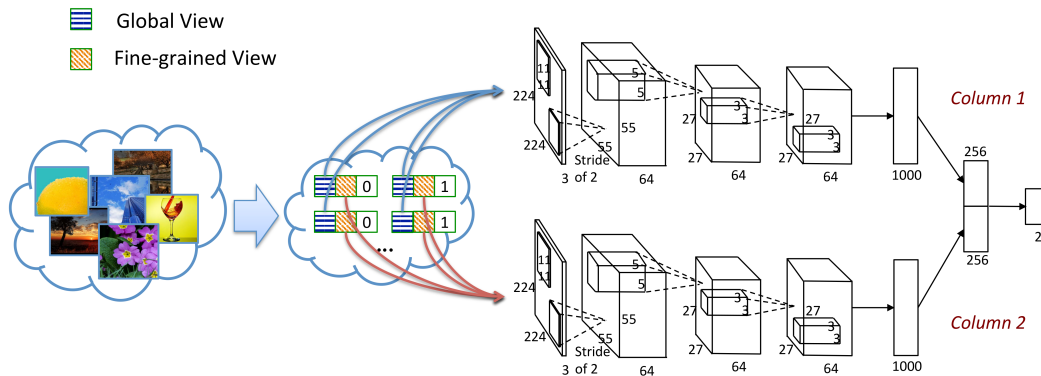


Figure 8.2 – Réseau neuro-mimétique profond à deux colonnes proposé par X. Lu et al. afin de réaliser l'apprentissage. Les deux colonnes sont identiques. L'une est nourrie par l'image réduite et normalisée, l'autre par la vignette extraite de façon aléatoire de l'image d'origine, à pleine résolution. Les 4 premières couches sont convolutives (avec des noyaux de $11 \times 11 \times 3$, $5 \times 5 \times 64$ et deux fois $3 \times 3 \times 64$). La réduction de dimension des couches totalement connectées se fait par mutualisation par maximum (max-pooling). La première a 1000 neurones, la seconde 256. Enfin la sortie est à 2 états. D'après [Lu et al., 2014c].

Les auteurs disposant d'informations de style³ sur certaines images de la base de données font alors un premier apprentissage du style (au moyen d'un RNP semblable au précédent), puis, utilisant l'information de style comme élément de régularisation, procèdent à la détermination de l'appréciation

2. VGG-Net est une architecture développée par le Visual Geometry Group de l'Université d'Oxford. Il se distingue par des convolutions limitées à des noyaux 3×3 . Il a obtenu la seconde place à la compétition ILSVRC-2014 (*ImageNet Large Scale Visual Recognition Competition*). VGG-Net est maintenant un produit de Torch, association de développeurs de logiciels d'intelligence artificielle issus de diverses sociétés des GAFA. VGG16, souvent utilisé pour les études sur l'esthétique offre 16 couches de traitement

3. Le style porte sur des propriétés particulières de l'image, annotées manuellement : présence de couleurs complémentaires, images à deux tons, images en Noir et Blanc, images HDR, photo macro, bougé, déplacement, etc.

esthétique.

Les résultats (table 8.1) permettent de valider le choix de l'architecture ainsi que le protocole d'apprentissage et l'utilisation d'un vecteur de style pour contraindre la solution. Ces résultats sont supérieurs à ceux présentés dans [Marchesotti et al., 2013] pour un coût d'implémentation cependant considérablement accru. Sur la base AVA, ils donnent un accord à plus de 73% avec la notation à 2 classes "Low" et "High".

8.1.2 L'architecture multi-voies DMA-Net de X. Lu *et al.*

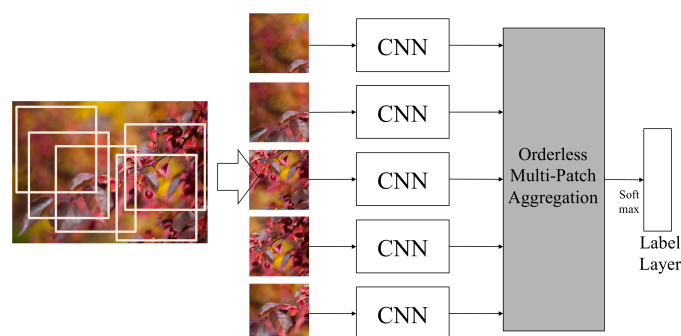


Figure 8.3 – Schéma d'apprentissage par réseau de neurones profond multi-voies. Afin de simplifier l'apprentissage, les divers CNN sont identiques. L'étape de fusion (Orderless Multi-Patch aggregation) est décrite sur la figure 8.4. Le choix de sous-images aléatoires se montre, en pratique, préférable à un positionnement prédéfini, d'après [Lu et al., 2015b].

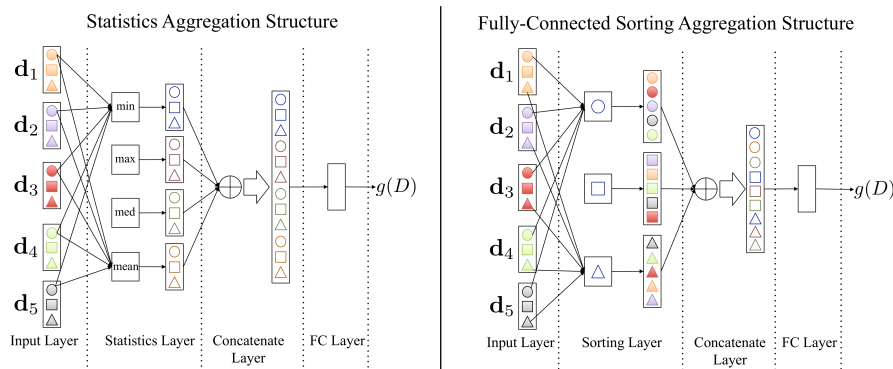


Figure 8.4 – Fusion des réseaux multi-voies (Orderless Multi-Patch aggregation) de la figure 8.3. Les réseaux neuro-mimétiques étant sensibles à l'ordre des données en entrée, il est souhaitable d'introduire une couche assurant l'indépendance du résultat à l'ordre de choix des sous-fenêtres avant l'étape de fusion. Cela peut être fait soit en fusionnant tous les neurones de même rang par un opérateur indépendant de l'ordre (max, min, médian) (à gauche), soit en triant les données de ces neurones de façon à les ordonner (à droite), d'après [Lu et al., 2015b].

L'architecture de RAPID est ensuite étendue dans [Lu et al., 2015b] en remplaçant l'architecture à deux voies par une architecture multi-voies, conformément au schéma de la figure 8.3. Elle prend le nom de DMA-Net (*Deep multipatch aggregative network*). On adopte un schéma à cinq voies, chaque voie étant une sous-image à pleine résolution tirée aléatoirement au sein de l'image à classer. Afin de simplifier les modules d'apprentissage, tous les CNN traitant une sous-image ont des paramètres identiques. Les cinq voies sont ensuite fusionnées selon deux stratégies qui donnent en pratique des résultats assez semblables : soit un opérateur statistique indépendant de l'ordre de ses

Méthode	Précision	Mesure-F	Score AUC
VGG avec recadrage	71.2 %	0.83	0.66
VGG avec chgt d'échelle	73.8 %	0.83	0.74
VGG avec remplissage	72.9 %	0.83	0.73
SPP-CNN	76.06 %	0.84	0.77
MNA-CNN	77.1 %	0.85	0.79

Table 8.2 – Comparaison des performances d'évaluation des images de la base AVA selon le type d'entrée adopté : images réduites par recadrage (*cropping*), changement d'échelle ou remplissage (*padding*) (en haut), à l'aide d'un réseau à multi-voies (SPP est le réseau original de [He et al., 2015b], MNA est celui de L. Mai et al.). Les deux mesures (Mesure-F et AUC) sont déduites de la courbe ROC de reconnaissance, d'après [Mai et al., 2016].

entrées (moyenne, max ou médian), soit une couche de neurones prenant en entrée des vecteurs triés (figure 8.4).

En 2017, le système DMA-Net présenté dans [Lu et al., 2015b] sert de référence pour la plupart des travaux conduits sur la qualité esthétique des photos par RNP.

8.1.3 S'adapter à la taille de l'image : les travaux de L. Mai *et al.*

Ces travaux se sont eux aussi attaqués au problème du format d'entrée de l'image [Mai et al., 2016]. Ils ont choisi une solution complexe mais efficace en adoptant une approche par décomposition pyramidale d'une part et par parallélisation des voies de traitement d'autre part.

L'idée de base consiste à faire précéder les couches de traitement classiques d'un RNP d'une couche de convolution et de mutualisation *pooling*. Pour s'adapter aux diverses tailles possibles d'images, les auteurs proposent de mettre ainsi en parallèle cinq RNP, avec des facteurs de décimation de 12×12 , 9×9 , 6×6 , 4×4 et 2×2 respectivement (voir la figure 8.5 à gauche)⁴. Par ailleurs, le rapport d'aspect de l'image pouvant ne pas être le même que celui (généralement carré) de la couche d'entrée du RNP, ils proposent une technique de mutualisation adaptative, dérivée, comme la précédente, des travaux de [He et al., 2015b]. L'architecture ainsi conçue est appelée MNA-CNN (*Multi-Net adaptive pooling CNN*) ; elle est réalisée à l'aide d'un réseau VGG-Net (voir note 2), pré-entraîné à l'aide du modèle Zoo de BVLC/CAFFE⁵

Ils adoptent ensuite une amélioration supplémentaire en prenant en compte le contenu sémantique de l'image. Cela est fait en ajoutant un canal supplémentaire, en pratique, celui développé par Google (Places 205-GoogLeNet)⁶, dont ils réduisent le nombre de classes à 7 (au lieu de 205) : *humains, plantes, architecture, paysage, nature-morte, animal et nuit* (figure 8.5 à droite).

L'expérimentation est conduite sur la base AVA. Les résultats ainsi obtenus sont meilleurs que ceux donnés par les méthodes classiques de recadrage *cropping*, de sous-échantillonnage, de mise à l'échelle ou de pavage (voir la table 8.2). Ils montrent aussi le bénéfice du canal de catégorisation.

8.1.4 La beauté du web : travaux de M. Redi *et al.*

Présentée dans [Redi et al., 2017], cette démarche a pour objectif de trouver des critères de beauté, non pas dans les photos de sites spécialisés, mais dans les images ordinaires issues du web qui comporte

4. Il semble que l'on puisse ainsi traiter des images dont la dimension maximale serait $12 \times 256 = 3072$ pixels dans la grande dimension, mais cela n'est pas précisé dans l'article.

5. CAFFE est une plateforme logicielle de réseau neuromimétique profond, développée par Berkeley AI Research (BAIR) et créée par Yangqing Jia au BVLC (Berkeley Vision and Learning Center).

6. GoogLeNet est un réseau à 22 couches développé par Google, 205 indique le nombre de classes de la couche de sortie. GoogLeNet a remporté la compétition ILSVRC-2014.

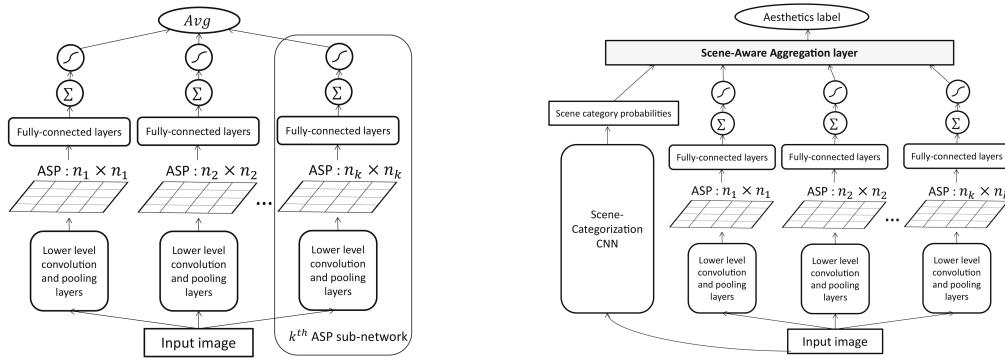


Figure 8.5 – Les deux architectures proposées par L. Mai et al.. A gauche pour traiter le problème d’adaptation au format de l’image, cinq réseaux sont mis en parallèle, précédés chacun d’une couche (convolution + mutualisation) de décimation variable. A droite un canal supplémentaire est ajouté pour prendre en compte la catégorie de l’image d’entrée (d’après [Mai et al., 2016]).

beaucoup plus d’images médiocres, voire de dessin ou de graphiques, que les sites de photo. Elle a pour cela utilisé une base spécifique (voir la section 6.2).

Les auteurs souhaitent également comparer les performances de trois chaînes de traitement différentes :

1. une méthode à base de primitives et d’un classifieur par forêts aléatoires (*random forest*). Les primitives choisies couvrent assez largement les descripteurs d’images : couleur, texture, contraste, mais aussi des primitives sur la composition : présence d’objets, symétries et unicité. Ils sont au nombre de 27 ;
2. un réseau de neurones profond à 3 sorties : “haute-qualité”, “qualité ordinaire” et “basse qualité”, qui est entraîné entièrement à partir de la base de données issue du web ;
3. un réseau de neurones profond pré-entraîné à partir d’une base de données AVA (images essentiellement “esthétiques”) selon le protocole de [Lu et al., 2015b], puis finement réglé à l’aide d’images de la base web (donc avec des images de mauvaise qualité).

Ces travaux confirment que quelques primitives permettent de distinguer avec une raisonnable fiabilité, “photos” et “non-photos” de façon à adapter, ensuite, l’évaluation de la qualité esthétique à cette information. Ils montrent que les meilleures performances sont obtenues à l’aide du réseau de neurones profond entraîné à l’aide du web (et non de la base AVA). Ils montrent aussi que les performances du classifieur sur primitives sont très proches des meilleures. Enfin, ils montrent que, si un réseau entraîné avec les seules images de bonne qualité de la base AVA donne des résultats médiocres (en particulier à cause des erreurs commises sur les images “non-photo”), il est possible de l’améliorer sensiblement par raffinement d’apprentissage (sans toutefois atteindre les performances d’un réseau entraîné sur une base web).

8.1.5 Le réseau siamois de S. Kong et al.

Ce travail (présenté dans [Kong et al., 2016]) propose d’ambitieux objectifs pour vérifier les capacités d’évaluation esthétique des RNP. Pour cela il développe tout d’abord une base de données originale (AADB, voir la section 6.2), elle aussi renseignée par *crowdsourcing* avec une évaluation de la qualité, une catégorisation et des attributs de qualité esthétique. A la différence des autres bases de données AADB conserve l’identité des évaluateurs, photo par photo.

Le schéma adopté est un RNP travaillant sur une image réduite (256×256), éventuellement repliée, sans canal de traitement de l’information à pleine résolution. Le réseau est amélioré par un montage “siamois” [Chopra et al., 2005] qui procède par traitement parallèle (avec des RNP semblables) de paires d’images qui ont été jugées de façon à peu près équivalente sur le critère d’esthétique (figure 8.6 à gauche).

Les auteurs testent tout d'abord le rôle des diverses informations dans les performances du système et pour cela ils développent une architecture capable de suppléer l'absence d'information, sur la catégorie par exemple, ou sur les attributs (figure 8.6). Ils montrent que toutes ces informations sont utiles mais ne sont pas indispensables. Les conclusions de l'expérimentation montrent de plus

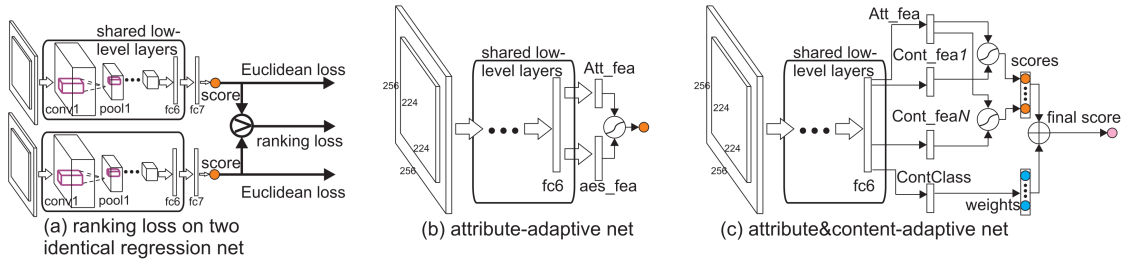


Figure 8.6 – Les divers éléments d'architecture utilisés dans [Kong et al., 2016]. A gauche, le réseau “siamois” qui traite deux images en parallèle avec les mêmes architectures. Au centre, un réseau entraîné de façon à donner simultanément une évaluation de l'esthétique (en bas) et une détection de primitives (en haut), ces deux informations étant combinées pour fournir l'évaluation finale. A droite, une architecture où l'on détecte à la fois les primitives et le contenu.

l'importance de l'observateur ayant annoté l'image dans le classement, ce que les auteurs mettent sur le compte d'une “même esthétique”, mais qui peut aussi être mis au compte de la fiabilité réduite du *crowdsourcing*. Elles montrent aussi la très faible généralisation des réseaux ainsi entraînés car les performances d'apprentissage et de reconnaissance croisées sur les bases AADB et AVA se révèlent très médiocres. Enfin les auteurs souhaitaient proposer une note “continue” entre les photos laides et les belles photos, par opposition aux approches binaires, mais aucun résultat n'est présenté sur ce point.

Une configuration assez semblable à celle des réseaux siamois, une configuration en réseaux “adverses” est utilisée dans [Deng et al., 2017a] pour améliorer une image initialement médiocre par un changement de contraste (séparément sur les canaux *Lab*) et par un recadrage. Un réseau adverse est un réseau génératif qui modifie des paramètres de l'image d'entrée de l'une des deux voies d'un réseau siamois de façon à obtenir des sorties identiques à une couche d'évaluation. Pour cela on met en opposition une image réputée belle et l'image médiocre et l'on fait évoluer les paramètres de la seconde jusqu'à rendre les sorties indiscernables.

8.1.6 Utiliser l'information sémantique : les travaux de Kao et al.

Présentés dans [Kao et al., 2017], ces travaux se proposent de démontrer que l'évaluation esthétique de la photo bénéficie d'une connaissance du contenu sémantique. Ils proposent une architecture par RNP qui accomplit simultanément les deux tâches de détermination du contenu sémantique et de l'évaluation des qualités esthétiques. Ces travaux sont conduits à l'aide du réseau décrit en figure 8.7. C'est un schéma classique de MT-CNN (*Multi task convolutional neural net*) qui est suivi d'une étape d'apprentissage des dépendances de la note esthétique en fonction de la classification sémantique.

Les images sont ici encore de taille modeste (256×256). Elles sont issues des bases AVA qui servent à l'apprentissage ; les résultats d'apprentissage ont été transportés avec succès à des images issues de Photo.net.

Une image x_n ($n \in 1, \dots, N$) est décrite par un attribut d'évaluation esthétique y_n qui peut prendre C valeurs et des attributs sémantiques z_n dont les valeurs z_n^m ($m = 1, \dots, M$) sont soit 0 soit 1 selon que l'image appartient aux images de sémantique m ou non. La base de données (X, Y, Z) est donc constituée de N triplets $\{x_n, y_n, z_n^m\}$. La première étape confiée au MT-CNN se fait à l'aide d'une régression logistique multinomiale. Elle est suivie par une étape de prise en compte de la dépendance

δ	MT-CNN 1	MTRL-CNN
0	76.15 %	79.08 %
1	75.90 %	77.71 %

Table 8.3 – Résultats comparatifs du traitement avec le couplage entre sémantique et esthétique et sans ce couplage. MT-CNN correspond au premier étage de CNN, sans le couplage (celui de la figure 8.7). MTRL-CNN est le réseau complet avec couplage de Kao et al..Le paramètre δ est le même que le delta en abscisse de la figure 7.10, d’après [Kao et al., 2017].

explicite entre sémantique et esthétique par l’intermédiaire de la matrice de covariance Ω :

$$\Omega = \frac{(W^T W)^{1/2}}{\text{Tr}[(W^T W)^{1/2}]} \quad (8.1)$$

où W est le vecteur des paramètres de la couche finale (figure 8.7). L’optimisation se fait alors en alternant rétro-propagation du gradient à Ω constant pour calculer les W et recalcul de Ω à partir de ces valeurs de W . Cette méthode prend le nom de MTRL-CNN (*Multi Task Relationship Learning Convolutional Neural Net*).

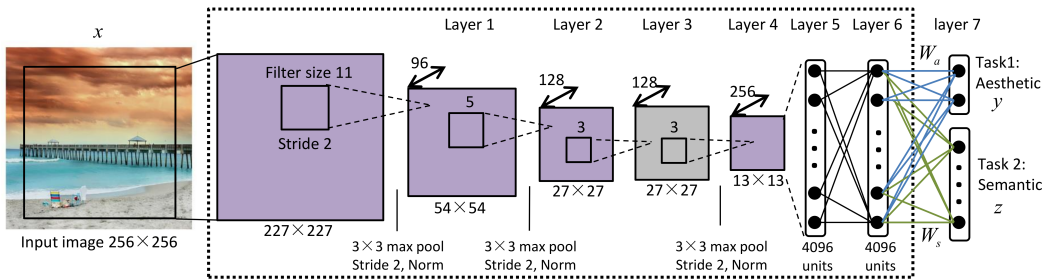


Figure 8.7 – L’architecture utilisée dans la première étape des travaux de Kao et al. On reconnaît quatre couches de réseau convolutif avec décimation, suivies de trois couches totalement connectées, la dernière étant constituée de deux réseaux spécialisés, l’un pour attribuer une note esthétique, l’autre pour reconnaître le contenu sémantique (d’après [Kao et al., 2017]).

L’expérimentation est conduite sur la base AVA dont sont tirées 185 000 images avec au moins un identifiant sémantique d’appartenance à l’une des 29 classes sémantiques disponibles ($M = 29$). L’entraînement se fait sur 165 000 images et la vérification sur 20 000. Deux valeurs d’esthétique seulement sont utilisées ($C = 2$). Les images AVA étant notées sur 10, l’apprentissage utilise un terme δ pour séparer autour de la valeur 5 les belles ($> 5 + \delta$) et les vilaines images ($< 5 - \delta$). δ vaudra 0 ou 1.

Les résultats obtenus dans ces conditions montrent un taux de bonne classification dans l’une des deux classes esthétiques supérieur à 75 % (voir table 8.3). Le bénéfice de l’information sémantique est toujours vérifié, mais modéré (de l’ordre de 2 à 3 %).

8.2 Variantes autour de l’architecture de base des RNP

De très nombreux travaux sont consacrés aux techniques d’évaluation de la qualité esthétique par RNP. Si les architectures de base sont celles que nous avons examinées ci-dessus, une tendance pourrait apparaître dans les derniers travaux à remplacer les couches totalement connectées par des couches purement convolutives. Par ailleurs, un effort est fait pour proposer des interprétations des toute dernières couches pour justifier la classification faite [Jin et al., 2016a, Kairanbay et al., 2017,

Murray and Gordo, 2017] en présentant des cartes d'activation superposées aux images. Ces résultats sont encore loin d'être convaincants.

D'autres approches visent à simplifier les architectures [Kairanbay et al., 2017] ou à rendre plus simple l'apprentissage [Srivastava and Kant, 2018] (en l'occurrence en proposant un apprentissage *from scratch*). Nous allons voir ici plus en détail quelques-unes de ces tendances.

8.2.1 Une approche qui distingue la nature de la photo : les travaux de Y. Kao *et al.*

La méthode développée par Y. Kao, K. Huang et S. Maybank [Kao et al., 2016], propose d'associer à la photographie sa classe : "Scène", "Objet" ou "Texture" puis, en fonction de la catégorie trouvée, de l'orienter vers 3 réseaux d'architectures un peu différentes.

Ainsi, pour la classe "Scène" un seul réseau classique est utilisé, doté de 5 couches convolutives, suivies de 2 couches totalement connectées. Pour la classe "Objet", le traitement est conduit en parallèle sur deux réseaux, l'un traitant l'image globale, l'autre focalisé sur la zone de plus grande saillance. Les deux canaux passent par une couche totalement connectée, puis sont regroupés avant de traverser une nouvelle couche totalement connectée. Enfin, dans le cas d'une image de "Texture", on sélectionne 16 sous-images par une fenêtre glissante, chaque fenêtre passe par un réseau un peu raccourci puis les 16 résultats sont ensuite moyennés.

Le réseau fournit deux résultats. Tout d'abord par un classement sur 1 bit, une appartenance ou non à la classe des "belles images", puis, par une régression utilisant la somme des carrés des sorties de la toute-dernière couche totalement connectée, une note continue d'évaluation de la qualité esthétique de la photo.

Le réseau ainsi développé est appelé A&C CNN : (*Aesthetic and Category*). Ses résultats sont assez bons, sans vraiment justifier l'accroissement de la complexité du à la séparation en trois catégories.

8.2.2 Comparer les photos entre-elles : la proposition de K. Schwarz *et al.*

Nous avons cité la base de données de l'Université de Tübingen, obtenue en recueillant les avis favorables des visiteurs du site Flickr. Elle a été conçue pour mener une démarche originale de détermination de l'esthétique des images à partir de l'opinion - non spécialiste - d'un grand nombre d'internautes [Schwarz et al., 2016]. Rappelons pour cela que les images se sont vu attribuer une note de qualité, continue, entre 0 et 1 (voir l'équation 6.1 et la figure 8.8). Les auteurs entreprennent alors un apprentissage automatique, à partir d'un réseau profond (architecture ResNet-50⁷ et logiciel TensorFlow⁸) qui utilise les qualités esthétiques non pas pour classer les images ("Belle" ou "Pas belle") mais pour les ordonner par comparaison 2 à 2 (réseaux siamois [Chopra et al., 2005], comme dans l'étude précédente), ou 3 à 3 par les réseaux triplets [Hoffer and Ailon, 2015]. Cet objectif (classer les images par rang d'esthétique) est, de l'avis des auteurs, plus proche de la façon de fonctionner du jugement humain.

Les qualités de ce *ranking* sont vérifiées par recours à un vote massif sur internet pour contrôler que deux images placées dans une relation d'ordre particulière par le RNP se retrouvent bien ordonnées de la même façon par les internautes. Les auteurs montrent qu'ils peuvent cependant déduire de cet ordonnancement une classification aux performances très semblables à celles des autres techniques.

Les auteurs utilisent finalement leur logiciel pour 3 applications différentes :

- pour ordonner des photos dans un album,
- pour extraire la meilleure photo d'une séquence vidéo,

7. ResNet est le réseau de neurones convolutif de l'équipe Microsoft qui a gagné la compétition ILSRVC 2015 [He et al., 2015a]). ResNet-50 est l'une des versions de ce réseau.

8. TensorFlow est une librairie de logiciels libres de calcul utilisant les graphes de flots de données. Il a été développé par les équipes du projet Google Brain au sein du centre de recherche en Intelligence Artificielle de Google.

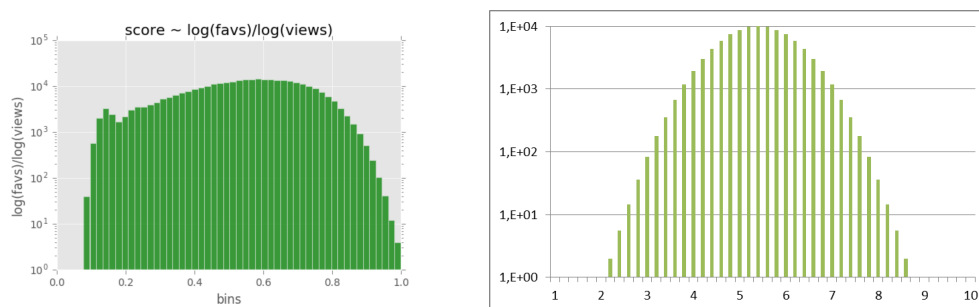


Figure 8.8 – A gauche, répartition des 350 000 notes (entre 0 et 1) de la base de données de Tübingen (d’après [Schwarz et al., 2016]). Les notes sont calculées à partir de la formule (6.1) qui tire profit des avis des internautes. On note que la distribution est presque plate, distribution très différente de celle des images de la base AVA, pratiquement gaussienne, représentée ici à droite également en échelle logarithmique, obtenue à partir de la figure 6.1).

- pour sélectionner la meilleure sous-image dans une scène large.

8.2.3 A mi chemin entre classification et RNP : la méthode de Z. Wang et al.

Ces travaux [Wang et al., 2016] essaient de coupler l’approche par primitives sélectionnées et par RNP. Faisant fonctionner en parallèle des voies de sélection de primitives, ils se placent sous l’analogie du cerveau humain pour justifier la démarche adoptée et son nom (*Brain-inspired-CNN*), et plus particulièrement dans la lignée du modèle cérébral proposé par A. Chatterjee [Chatterjee, 2003]. Ils divisent ainsi le traitement en deux niveaux :

- dans un premier niveau, des détecteurs sont menés en parallèle sur des primitives élémentaires : la couleur (directement mesurée par ses composantes HSV), puis chacune des 14 informations de style photographique renseignée dans la base AVA (voir la figure 6.2), chaque information étant prise en charge par un RNP composé de 4 couches convolutives et 2 couches totalement connectées) ;
- dans un second niveau, les sorties des détecteurs précédents servent d’entrée à un réseau à 128 entrées, 4 couches convolutives et 2 couches totalement connectées (voir figure 8.9).

La sortie finale consiste en deux valeurs : la moyenne et la variance de la gaussienne qui reproduit la courbe de distribution des avis des membres du jury AVA (voir la figure 6.1)⁹. Ces valeurs sont obtenues en remplaçant la fonction de sortie *softmax* par le calcul d’une divergence de Kullback-Leibler.

Résultat intéressant de cette étude, alors que l’on s’attendrait à ce que l’apprentissage de la distribution des notes permette un classement selon une échelle continue, les auteurs insistent sur le rôle particulier de cette distribution pour améliorer la qualité du classement binaire : ”belle image “ versus ”vilaine image“.

En annexe de ces travaux, les auteurs ont conduit une étude pour voir quelles transformations d’images préservent la note d’évaluation donnée par le réseau. Leur objectif était de multiplier la base d’apprentissage en démultipliant le nombre d’images annotées par des transformations respectueuses de l’évaluation. Pour cela ils accordent une note (*LP* dans le tableau 8.4) qui décroît de 1 à 0 lorsque la distorsion réduit l’évaluation. On en déduit que seuls la réflexion, le changement d’échelle et le bruit faible assurent une bonne conservation de la note. Par suite, une campagne de tests montre que l’augmentation de données d’apprentissage à l’aide de ses 3 transformations permet de gagner 2 ou 3 % dans les performances.

9. Les auteurs notent cependant que le choix d’une distribution gaussienne est mal adapté aux photos de très bonne ou de très mauvaise qualités dont les notes ont une distribution très dissymétrique.

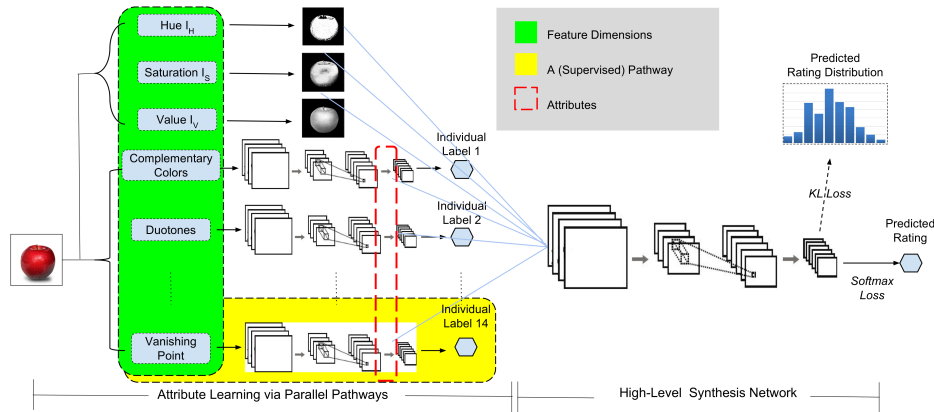


Figure 8.9 – Architecture du système BDN de Z. Wang et al., inspiré du cerveau humain (d’après [Wang et al., 2016]). En entrée, il met en œuvre un ensemble de détecteurs en parallèle, sensés détecter des primitives indépendamment les unes des autres. Hormis les 3 premiers qui mesurent les primitives chromatiques HSV, les autres canaux font appels à des RNP. Les sorties de cet étage sont elles-mêmes envoyées sur un autre RNP, dit de haut-niveau. Le réseau prédit également la distribution des notes (similaire à celle que donnerait AVA) à l’aide d’une fonction de coût de Kullback-Leibler.

transformation	Description	Note LP
réflexion	symétrie verticale	0,99
changement d’échelle aléatoire	choix d’un facteur unique dans $[0,9, 1, 1]$	0,94
écrasement/étirement	choix de 2 facteurs dans $[0,9, 1, 1]$	0,55
bruit faible	bruit gaussien $\in N(0, 5)$	0,87
bruit fort	bruit gaussien $\in N(0, 30)$	0,63
modification de la couleur	distorsion indépendante des canaux RVB	0,10
rotation	transformation affine aléatoire	0,26

Table 8.4 – Effet de diverses transformations sur le maintien de la note d’appréciation d’une photographie. Le facteur LP vaut 1 lorsque la note demeure inchangée. L’évaluation ne résiste qu’à la réflexion et au changement d’échelle, ainsi, dans une moindre mesure au bruit de faible valeur [Wang et al., 2016].

Le système proposé a été implémenté à l’aide du package Cuda-Convnet¹⁰

8.2.4 Porter attention à la construction de l’image : le système A-Lamp-CNN

Ces travaux se placent dans la lignée du système DMA-Net (voir section 8.1.2) qui utilise plusieurs voies en parallèle. Il a été présenté dans [Ma et al., 2017], son nom (A-Lamp CNN, *Adaptive layout-aware multi-patch*) exprime ses propriétés essentielles : d’une part il procède à une estimation de la composition (*layout*), d’autre part il mène en parallèle une mesure de l’esthétique sur plusieurs zones examinées (*multi-patch*) (voir la figure 8.10). L’analyse de la composition s’inspire de nombreux travaux consacrés à la distribution de zones d’intérêt dans une image (par exemple [Liu et al., 2010] ou [Obrador et al., 2010]). Elle utilise un réseau de détection adaptative de zones d’intérêt dont les résultats alimenteront également l’autre voie de traitement. Chaque zone d’intérêt, définie par sa boîte englobante, est un nœud du graphe de représentation, auquel sont affectés des attributs locaux et des attributs globaux. Les attributs locaux décrivent la position relative d’une zone d’intérêt par

10. Cuda-Convnet est un package développé par l’Université de Toronto à partir de 2011 permettant de simuler tout type de RNP. Il est écrit en C++ et parallélisé en Cuda [Krizhevsky et al., 2012].

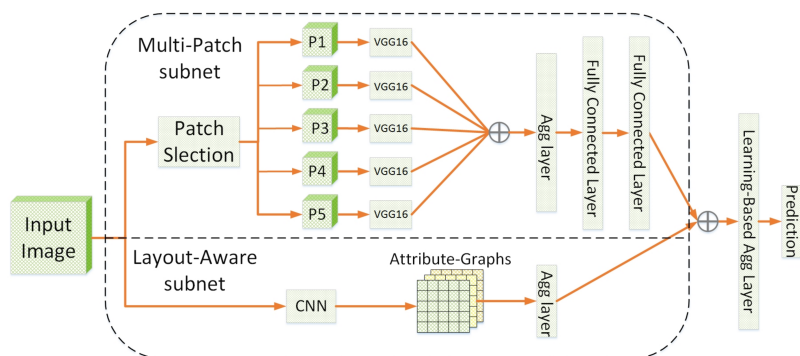


Figure 8.10 – Architecture du système A-Lamp mettant en évidence, dans la partie supérieure, le sous-réseau multi-patch qui traite en parallèle des sous-images à pleine résolution et dans la partie inférieure, le réseau traitant la composition (d’après [Ma et al., 2017]). Un réseau de détection adaptative de zones d’intérêt (pas représenté sur cette figure) est placé immédiatement après la lecture de l’image d’entrée et alimente les deux parties du traitement.

rapport à chacune des autres (distance, orientation du lien, recouvrement), tandis que les attributs globaux représentent la position de chaque nœud dans la scène.

La voie multi-patch est identique au DMA-Net de [Lu et al., 2014a] ou de [Lu et al., 2015a], mais au lieu d’utiliser des sous-images en position fixe ou en position aléatoire, la voie de traitement parallèle en pleine résolution traite les sous-images sélectionnées par le réseau de détection adaptative.

Les auteurs démontrent que cette voie multi-patch utilisant des zones pré-sélectionnées est préférable aux versions à sous-images fixes et aléatoires. Ils montrent aussi que le réseau complet, avec ses deux voies, donne de meilleurs résultats de classement binaire (“belle” ou “pas-belle”) que les autres architectures proposées antérieurement.

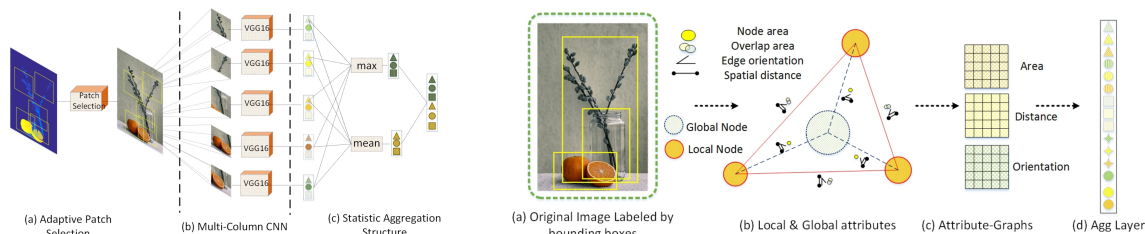


Figure 8.11 – A gauche le sous-réseau multi-colonne traite en parallèle, et en pleine résolution, les zones d’intérêt détectées par le réseau de détection adaptative de régions. A droite, le réseau traitant la composition à l’aide du graphe d’attributs (d’après [Ma et al., 2017]).

Ces travaux utilisent l’architecture Multi-Patch VGG16 (voir note 2) pré-entraîné sur ImageNet et le logiciel CAFFE (voir note 5). La détection préalable des zones d’intérêt suit les travaux de [Zhang et al., 2016].

8.2.5 Exploiter la distribution des évaluations

H. Talebi et P. Milanfar se sont penchés sur l’évaluation de l’esthétique d’une part et de la qualité des photos d’autre part. Nous ne nous intéressons ici qu’aux travaux concernant l’esthétique. Ils ont développé un système appelé NIMA (*Neural image assessment*) qui attribue à chaque photo une note entre 1 et 10, plus précise que la simple affectation à une classe “belle” ou “pas belle” [Talebi and Milanfar, 2017].

Pour cela, ils exploitent la totalité de la base AVA, sans en écarter les photos de qualité moyenne comme la plupart des autres auteurs. De plus ils exploitent une information disponible dans cette base mais qui n'a pas beaucoup été utilisée par les autres travaux¹¹ : la distribution des notes affectées à chaque photo (en moyenne 200 notes, voir la figure 6.1 à gauche). Lors de l'étape d'apprentissage des poids du réseau, c'est vers l'estimation de cette distribution qu'ils cherchent à converger, à l'aide d'une optimisation par une méthode de transport optimal (*Earth Mover's Distance*). La note unique finalement retenue pour une photo est, comme dans les autres études, la moyenne des notes données par cette distribution, comprise donc entre 1 et 10. Les résultats obtenus montrent un accord assez satisfaisant avec la note des juges humains (une erreur moyenne de 5 % et max de 12 % sur les 27 exemples présentés) quoique la corrélation moyenne sur la base AVA soit médiocre (de l'ordre de 0,6). Repassant à une classification en deux classes par seuillage, ils donnent des performances de plus de 80 % de bonnes réponses, proches de l'état de l'art.

La bibliothèque logicielle utilisée est TensorFlow (voir note 8) et le RNP est pré-entraîné sur la base ImageNet de façon à accélérer la phase d'apprentissage. Trois architectures différentes sont comparées : VGG16 (voir note 2), Inception-v2¹² et Mobile-Net¹³. Elles donnent des résultats assez semblables, les meilleurs étant cependant obtenus avec Inception-v2.

Une approche très semblable est proposée par [Murray and Gordo, 2017]. Elle permet aussi de prédire :

- la classification en deux classes par seuillage à 5,
- la note esthétique comme moyenne de la distribution des scores,
- la distribution des scores (répartie sur 10 niveaux).

La méthode d'optimisation adoptée se distingue par un critère exploitant une norme de Huber sur les distributions (quadratique en dessous d'un seuil, linéaire au dessus).

8.2.6 Formuler une appréciation littérale

Dans le système appelé NAIR *Neural Aesthetic Image Retriever* [Wang et al., 2018], une tentative est faite d'associer dans un même système deux objectifs :

- d'une part séparer les images en haute et basse qualité comme la plupart des autres systèmes,
- d'autre part associer un commentaire littéral explicitant les raisons de la décision.

Ce second objectif ressemble à celui que nous avons décrit au paragraphe 7.11 et mis en œuvre à l'aide de primitives et de classifieur dans [San Pedro et al., 2012].

Pour la première partie, on utilise la base de données AVA et un réseau classique (modèle Inception-v3 et logiciel TensorFlow). La seconde partie tire profit des nombreux commentaires associés à la base AVA et regroupés dans un corpus appelé AVA-Comments, qui couvre plus de 20 000 photos. A l'aide de ce corpus, on entraîne un réseau capable de créer un commentaire littéral inspiré par la description de l'image, suivant le schéma proposé dans [Vinyals et al., 2015]. La tâche est délicate car les avis sont rédigés en langage naturel, ils comportent des termes non normalisés et des valences positives ou négatives issues soit du vocabulaire, soit de la syntaxe.

Nous notons, en examinant la figure 8.12, que ce très louable effort de fournir une justification de l'évaluation reste à améliorer, les termes employés étant peu précis et pas très pertinents.

11. Cette information a été prise en compte dans les travaux de B. Jin, M.V. Ortiz-Segovia et S. Süsstrunk [Jin et al., 2016a] qui attachent à la variance des notes affectées à une photo une notion de difficulté de jugement esthétique (plus la variance est forte, plus les avis des juges sont dispersés). Dans cet article, les auteurs proposent, de plus, de pondérer les échantillons par un poids inversement proportionnel à la fréquence de la note attribuée, de façon à éviter un sur-apprentissage des photos de qualité moyenne, ce que la plupart des auteurs traitent en excluant, lors de l'apprentissage, les photos dont les notes sont comprises entre 4 et 6 (poids binaires).

12. Inception-v2 exploite une parallélisation des étapes de convolution et de mutualisation et remplace la couche de connexion complète par une mutualisation par maximum (*max pooling*), réduisant ainsi notablement le nombre de paramètres à apprendre [Szegedy et al., 2015].

13. Mobile-Net est un RNP dédié aux architectures embarquées de vision. Il se distingue par les étapes de convolution qui sont remplacées par des filtres séparables, particulièrement rapides [Howard et al., 2017].

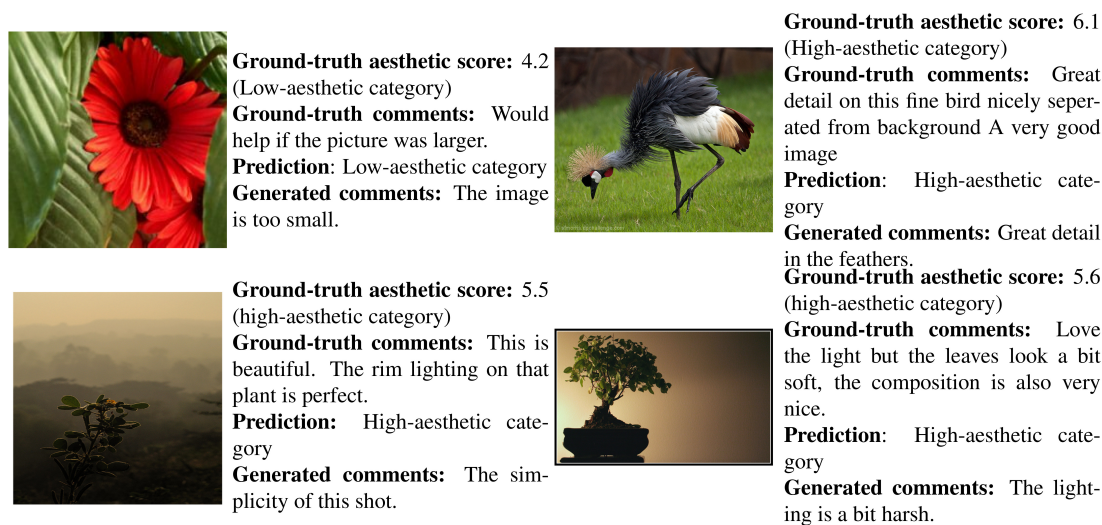


Figure 8.12 – Résultats d'évaluation de l'esthétique de quatre images de la base AVA par le système NAIR : appréciation de l'esthétique d'une part et expression littérale de l'évaluation d'autre part, comparées l'une et l'autre à leur "vérité" dans la base AVA (d'après [Wang et al., 2018]).

Des travaux utilisant les mêmes données sont présentés dans [Hii et al., 2017]. Ils tirent également parti de AVA-Comments, mais avec comme objectif seulement d'améliorer la classification binaire. La méthode de représentation par mots utilise l'algorithme GloVe (*global vector*), proposé par [Pennington et al., 2014]. Chaque image est décrite par 100 mots appartenant au vocabulaire complet de 20 000 mots extrait des commentaires. Un réseau récurrent à deux couches est alors utilisé, actif sur les termes textuels. Ses résultats sont concaténés à ceux du réseau multiGAP utilisé pour la partie visuelle. L'ensemble fournit un classifieur à 2 états.

8.2.7 Personnaliser l'évaluation

Le travail présenté dans [Park et al., 2017] par K. Park *et al.* est singulier et particulièrement intéressant car il cherche à donner une appréciation particulière, conforme aux choix d'un observateur précis.

Le principe consiste à utiliser un algorithme d'ordre et non de classification binaire. L'ordre prend en compte deux critères simultanément : une préférence générale en deux classes ("beau"/"pasbeau", comme dans la plupart des algorithmes) et des goûts particuliers obtenus à partir de réponses faites par l'observateur à un petit jeu de test. L'objectif est alors de fournir une classification qui respecte autant que possible l'avis général, mais en prenant en compte l'avis particulier.

La préférence individuelle se construit à partir, tout d'abord, du choix fait par l'observateur d'un échantillon préféré au sein d'une dizaine de paires, puis que par l'ordonnement complet d'un ensemble de 10 images. La préférence est ensuite propagée à la base de données par des techniques de plus proches voisins.

L'implémentation logicielle est assez complexe. Tout d'abord on choisit de décrire chaque échantillon par un vecteur de dimension 4 096 obtenu à la sortie de l'avant dernière couche d'un RNP classique [Krizhevsky et al., 2012]. Puis on recherche une optimisation selon deux critères l'un exprimant l'accord à la classification globale, l'autre la régression selon les préférences individuelles. L'algorithme original, inspiré du *support vector regression* (SVR) d'une part et du *ranking support vector machine* (R-SVM) d'autre part est appelé *support vector regression* (SVR). Il est testé sur la base AVA. Les premiers résultats laissent place à de possibles progrès. Il semble en particulier nécessaire de bien comprendre comment les goûts individuels peuvent être généralisés de façon convaincante à partir

d'un tout petit nombre d'apprentissages (condition nécessaire pour rendre le projet acceptable par l'utilisateur).

Chapitre 9

Analyse critique des techniques d'apprentissage automatique

9.1 Popularité des études sur l'esthétique

Nous allons tout d'abord rapprocher deux chiffres : d'une part celui des articles publiés sur le thème de l'esthétique 9.1 à gauche, d'autre part celui des publications qui associent les deux mots à priori assez antithétiques "Esthétique" et "Ordinateur" figure 9.1 à droite. Si les études sur l'esthétique restent très nombreuses (plus de 50 000 par ans depuis plus de 10 ans), les travaux mêlant esthétique et ordinateur croissent de façon beaucoup plus forte. Ils constituent plus de la moitié des travaux sur l'esthétique en 2018. Il y a donc aujourd'hui une bonne acceptation du milieu intellectuel à associer ces deux concepts dans un même raisonnement, ce qui était très rare il y a 15 ans.

Comme nous l'avons vu par ailleurs, même si quelques travaux sont encore publiés sur des approches algorithmiques (essentiellement issues des travaux de Birkhoff), depuis 15 ans ce sont les approches par apprentissage qui mobilisent la communauté.

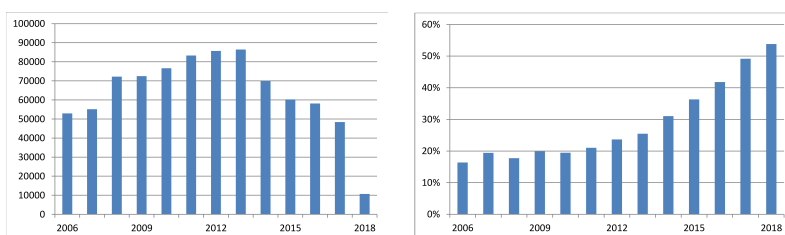


Figure 9.1 – *A gauche, recensement des références trouvées lors de la requête "Aesthetics" dans Google Scholar, par année de publication. A droite pourcentage des publications précédentes qui répondent au double critère "Aesthetics + Computer".*

9.2 Bilan des méthodes d'apprentissage

Ces systèmes font l'objet du tableau 9.1 où sont comparés certains de leurs attributs. Il montre bien comment les systèmes à base de primitives et de classificateurs, qui ont ouvert le chemin de l'esthétique par ordinateur, ont progressivement cédé la place aux techniques à base de réseaux de neurones. Notons cependant que des techniques spécifiques de classification peuvent encore parfois être employées pour des tâches particulières en aval d'un traitement par RNP comme dans [Park et al., 2017].

référence	base de données	type d'évaluation	primitives	classifieur
[Ke et al., 2006]	DPChallenge	2 classes	attributs génériques	Bayes, AdaBoost
[Luo and Tang, 2008]	DPChallenge	2 classes	primitives photo	Bayes, SVM, AdaBoost
[Datta and Wang, 2010]	Photo.net	2 classes+note	primitives photo	SVM
[Marchesotti et al., 2011]	Photo.net+CUHK	note	attributs génériques	PCA+FV+SVM
[Dhar et al., 2011]	DPChallenge	2 classes	primitives haut niveau	SVM
[Yao et al., 2012]	Photo.net	recommandations	composition,texture,color	ACQUINE+ SIMPLicity
[Lo et al., 2012]	CUHK	2 classes	primitives photo	SVM par catégorie
[San Pedro et al., 2012]	DPChallenge	note continue	opinions internautes	SV- ϵ
[Su et al., 2012]	DPChallenge	2 classes	bibliothèque esthétique	AdaBoost
[Lu et al., 2014a]	AVA	2 classes	HSV ou Munsell	LDA + Lasso
[Lu et al., 2014b]	AVA	2 classes	photo + ss-photo	RNP - SCNN + RD-CNN
[Schifanella et al., 2015]	BEAUTY	5 niveaux	primitives photo	PLSR
[Jin et al., 2016b]	AVA	2 classes	image réduite	RNP = ILGNet
[Kong et al., 2016]	AADB + AVA	2 classes	image réduite + catégorie	RNP "siamois"
[Mai et al., 2016]	AVA	2 classes	pyramide + info class.	RNP multitâche MNA
[Schwarz et al., 2016]	Tübingen	classe ment 2 à 2	image : 1 piste	RNP siamois + triplets
[Kao et al., 2016]	AVA	2 classes + distrib.	image + 3 catégories	A&C RNP
[Wang et al., 2016]	AVA - AVA2	2 classes + distrib.	histo. de notes.	Brain-inspired RNP
[Redi et al., 2017]	AVA \rightarrow Redi	3 classes	apprentissage images web	RNP
[Kao et al., 2017]	AVA	2 classes	esthétique + catégorie	RNP multitâche
[Kairanbay et al., 2017]	AVA - Comments	2 classes	images + style	RNP + couches GAP
[Ma et al., 2017]	AVA	2 classes	composition + ss-images	A-Lamp, multitâche
[Murray and Gordo, 2017]	AVA	note continue	histo. de notes.	RNP= ResNet
[Park et al., 2017]	AVA + interact.	choix personnalisé	préférences personnelles	RNP +SVR
[Talebi and Milanfar, 2017]	AVA	note continue	histo. de notes.+catégorie	RNP multitâche
[Srivastava and Kant, 2018]	AVA	2 classes	image réduite+couleur LAB	RNP = ILGNet
[Wang et al., 2018]	AVA - Comments	explications	AVA-Reviews	RNP = Inception

Table 9.1 – Les divers systèmes d'apprentissage automatique de qualité de photographie dans l'ordre de leur apparition. On recouvre ici sous le terme "primitives photo" les primitives qui sont spécifiquement déduites de règles photographiques : règle du tiers, mise au point sur l'objet et flou du fond, importance de la construction, etc. Les primitives de "haut niveau" sont calculées dans l'image, tandis que les "catégories" sont issues des données, souvent comme métadonnées, parfois calculées par le RNP. Les études utilisant des RNP sont présentées sur un fond grisé.

9.2.1 Quelles architectures ?

Sans surprise, nous constatons que les travaux présentés se contentent souvent d'exploiter des architectures classiques ayant fait leurs preuves par ailleurs dans les compétitions internationales de reconnaissance des formes. Les architectures ainsi que les logiciels sont alors les meilleurs à l'instant t et la génération qui triomphera à la prochaine compétition les remplacera. Pour les études sur l'esthétique, une attention particulière est portée sur le format des images et concerne aussi bien la résolution que le rapport d'aspect, mais là aussi les solutions adoptées sont celles retenues pour d'autres problèmes. Les auteurs ne semblent pas rencontrer de problème à traiter des images de taille très modeste (la base AVA universellement utilisée pour l'apprentissage ne dispose que de réductions des images originales de DPChallenge de taille 640×480 .)

Certains développements des RNP sont suivis avec un grand intérêt par la communauté de détermination de l'esthétique par RNP : les réseaux multitâches et les réseaux siamois surtout, mais aussi les réseaux récurrents. On met ainsi en parallèle des tâches très différentes comme celles qui traitent l'image elle-même d'une part et celles qui traitent les métadonnées ou qui cherchent à les déduire de l'image : catégorie, style, contexte, etc.

Au contraire de la tendance observée en 2018 dans les applications les plus générales des RNP qui privilégient une multiplication de couches convolutives, l'utilisation conjointe de couches convolutives et de couches totalement connectées demeure la règle dans les études sur l'esthétique et les architectures adoptées sont quasiment toutes "en entonnoir" plutôt qu'en "yoyo" comme c'est la tendance dans d'autres domaines.

L'apprentissage du réseau reste un point délicat. Compte tenu de la grande profondeur des réseaux actuellement employés, les bases de données dédiées à l'esthétique apparaissent trop réduites pour

δ	SVM+Gauss. mixt. [Murray et al., 2012]	SCNN [Lu et al., 2014b]	RD-CNN [Lu et al., 2014b]	DMA-Net [Lu et al., 2015b]	MNA-CNN [Mai et al., 2016]
0	66.7 %	71.20 %	74.46 %	75.41 %	77.4 %
1	67.0 %	68.63 %	77.70 %	-	76.5 %

δ	ILGNet [Jin et al., 2016b]	BDN [Wang et al., 2016]	MTRL-CNN [Kao et al., 2017]	A-LAMP [Ma et al., 2017]	NIMA [Talebi and Milanfar, 2017]
0	79.25 %	76.80 %	79.08 %	81.7 %	81.5 %
1	-	76.04 %	77.71 %	-	-

Table 9.2 – Résultats comparatifs de diverses méthodes d'évaluation esthétique sur la base de données AVA, appliquées à une classification à deux classes “haute qualité” / “basse qualité”. La méthode A-LAMP [Ma et al., 2017] donne les meilleurs résultats (en 2017). Le paramètre δ est le même que le delta en abscisse de la figure 7.10, d'après [Kao et al., 2017].

conduire un apprentissage de bout en bout (c'est en particulier le cas de AVA). Le recours à des apprentissages “universels” effectués sur des bases d'images banalisées pour des tâches très génériques de reconnaissance des formes est donc aujourd'hui généralisé. Contrairement à ce que l'on pourrait craindre, cela ne semble pas affecter les performances des études sur l'esthétique..

9.2.2 Quelles performances ?

La communauté utilise aujourd'hui la classification à 2 classes (“belle-image” vs. “pas belle image”) sur la base AVA comme banc de test des performances des systèmes¹.

On discutera plus loin de la pertinence de ce critère.

Ces résultats sont présentés sur le tableau 9.2 qui est repris sur la figure 9.2

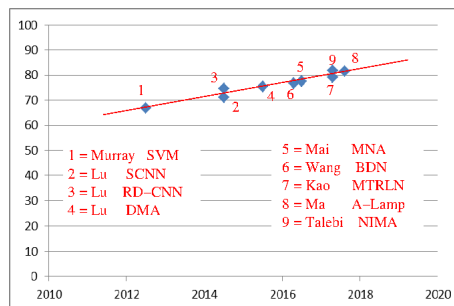


Figure 9.2 – Performances d'évaluation esthétique à 2 classes (belle / pas-belle) des systèmes, sur la base AVA, avec apprentissage automatique, depuis 10 ans. Notons que seul le système de Murray et al. a été retenu dans ce “podium” qui regroupe les meilleurs systèmes, comme système utilisant une sélection de primitives suivie d'une classification. Tous les autres utilisent des RNP.

On constate que ces performances de classification sont aujourd'hui bonnes, statistiquement, pour de grands ensembles de photos (plus de 75 % de bonne classification entre “haute qualité” et “basse qualité” sur des ensembles de plusieurs centaines de milliers d'images). Par ailleurs qu'avec quelques connaissances supplémentaires sur la catégorie de la scène observée ou sur le style de la photographie, on peut porter ce taux de reconnaissance à 85 %. Ces travaux étant encore très préliminaires, on peut s'attendre à ce qu'ils croissent encore à court terme.

On constate aussi que les méthodes proposées s'adaptent bien à une grande diversité de photos qui se distinguent aussi bien par leur style que par leur format.

1. On a cependant signalé les variantes possibles de ce critère : d'une part AVA a donné naissance à plusieurs sous-bases selon qu'une image est ou non associée à certaines méta-data, d'autre-part, le critère de “belle-image” est souvent lié à un paramètre (soit la distance entre classes, soit le pourcentage d'images retenues (cf. section 6.2)).

Les études qui visent à transporter les apprentissages d'une base de données vers une autre concluent plutôt à une bonne robustesse de ces apprentissages [Kao et al., 2017] quoique des travaux donnent des résultats opposés [Kong et al., 2016].

On ne peut cependant manquer de souligner une critique majeure de ces résultats. La classification à deux classes est un problème intellectuellement peu excitant et socio-économiquement sans grand intérêt. Son attrait essentiel est d'offrir un juge de paix aux algorithmes. La plupart des travaux qui l'utilisent se fixent initialement un objectif plus ambitieux soit de conclure à une évaluation fine et non binaire de la mesure esthétique, soit de comparer entre elles un petit ensemble d'images, soit enfin de dégager des éléments permettant d'expliquer ou de justifier une notation.

Examinons ces autres objectifs et la façon dont ils ont été abordés.

Evaluation continue : hormis l'éphémère système ACQUINE [Datta and Wang, 2010], aucun système ne propose une note fiable, reflétant une évaluation continue de l'esthétique. Plusieurs systèmes se proposent de prédire une répartition de notes telle que pourrait la fournir un jury AVA et de là en déduisent une évaluation moyenne [Talebi and Milanfar, 2017, Murray and Gordo, 2017], mais ces résultats sont plus présentés comme des sous-produits, éventuellement exploitables, que comme la sortie principale du système. La comparaison, sur une base d'images tests issue d'AVA, de la note moyenne déduite du jugement du RNP et de la note moyenne de l'évaluation AVA, laisse la place à des progrès importants.

Comparaison entre images : c'est aussi un objectif souvent évoqué mais très peu traité. C'est cependant explicitement l'objectif de l'étude présentée dans [Schwarz et al., 2016]. La méthode proposée est très intéressante, les résultats préliminaires présentés sont encourageants mais au prix d'une complexité certaine du modèle mis en place. Il sera particulièrement intéressant de suivre sur ce point les résultats de la compétition SPARK² dont nous avons parlé plus haut, où, pour la première fois, un algorithme sera confronté à un jury de photographes professionnels pour dépouiller une compétition.

Justifier la notation : les techniques à base de classifieurs et de primitives sont assez bien adaptées pour dégager les causes essentielles de l'estimation esthétique puisque l'on peut tracer parmi les critères adoptés ceux qui sont effectivement actifs. Malheureusement ils ont montré des performances trop faibles pour que cette piste soit poursuivie. Les méthodes à base de RNP tournent le dos à une justification de leurs résultats puisqu'elles cachent leurs décisions dans l'épaisseur des couches successives. Traditionnellement, on essaie tout de même d'expliquer les décisions en révélant des couches proches de la sortie où s'affichent les zones activées juste avant la décision. C'est ce que font, sans grand succès³, plusieurs études [Murray and Gordo, 2017, Kairanbay et al., 2017, Jin et al., 2016a]. Plus raisonnables sont les arguments des travaux qui s'appuient sur des expressions littérales, comme ceux présentés dans [San Pedro et al., 2012, Wang et al., 2018]. Leurs résultats sont cependant encore loin de convaincre. Si la méthodologie mise en place semble tout à fait adaptée, on peut s'interroger sur leur base de travail, les avis des "experts" de la base AVA. Est-ce une matière assez fiable et assez bien formalisée pour soutenir le raisonnement proposé ?

2. "Spark" (Le concours *SPARK, a RenAIssance photographic contest*) de Huawei, en promotion de son modèle P20 Pro de téléphone portable, <https://consumer.huawei.com/uk/campaign/sparkrenaissance/> (juillet 2018).

3. Lorsque l'on examine attentivement la localisation des zones actives, il apparaît de nombreuses situations où ces zones sont "à côté" des éléments qui devraient retenir l'attention. Bien sûr ce ne sont pas les cas présentés dans les articles qui illustrent ce défaut, encore que l'on peut longuement discuter les résultats de [Kairanbay et al., 2017] ou de [Jin et al., 2016a].

9.2.3 S'interroger sur les hypothèses

On ne peut manquer, à ce point de bilan, de revenir sur les hypothèses qui soutiennent ces travaux. Nous regretterons tout d'abord que ces hypothèses soient très rarement formulées et bien sûr encore moins justifiées. L'argument de la supériorité des RNP pour les tâches de reconnaissance des formes et d'évaluation dans les domaines les plus divers tient trop souvent lieu de justificatif aux travaux entrepris en esthétique à l'aide de ces outils.

L'hypothèse la plus cachée est celle d'une esthétique objective et universelle. Il est pourtant clair dans tous ces travaux (hormis quelques très rares exceptions sur lesquelles nous reviendrons) que les auteurs mettent en place une estimation valable pour tous et en tout lieu, indépendamment de l'observateur, de sa culture, de son tempérament, de son humeur de ses émotions du moment. Ce jugement se constitue à partir d'une connaissance, reflet elle-même d'une appréciation universelle attachée à l'image et à elle seule. Ces informations sont tout entières contenues dans la base de données. Nous sommes très exactement dans un contexte que les informaticiens nomment "*data-driven*", le programmeur s'efforçant d'être aussi transparent que possible de façon que les conclusions soient universelles. Ce parti-pris, en esthétique, relève clairement de l'objectivisme tel que nous l'avons résumé au paragraphe 2.1.

Cette interprétation s'inscrit ainsi clairement à contre-sens des approches que l'on développe dans d'autres domaines de la connaissance, que ce soit en neuro-biologie (chapitre 3), en psychologie, sociale ou expérimentale (chapitre 4), en philosophie (paragraphe 2.2). Si le subjectivisme, position diamétralement opposée à l'objectivisme, n'est plus aujourd'hui la règle comme il a pu l'être en philosophie et malgré ses nombreux supporters [Zeki, 1999, Ishizu and Zeki, 2011, Changeux, 2016], il est quand même communément mis à contribution (en particulier dans l'interprétation interactionniste qui rassemble aujourd'hui beaucoup de supporters [Reber et al., 2004, Ingarden, 2011] qui prônent que l'observateur avec ses états de conscience et d'attention est un déterminant majeur de l'appréciation esthétique qui "exploite" le signal envoyé par l'objet, l'autre déterminant.

Les approches proposées en intelligence artificielle pour aborder l'évaluation de l'esthétique des photos ne laissent aujourd'hui pratiquement aucune place au subjectivisme. Elles évacuent ainsi totalement des notions pourtant reconnues comme très importantes de sensibilité, d'éducation, de culture, de contexte.

Soulignons pourtant deux tentatives de laisser une petite marge d'influence à la personnalité de l'observateur dans les travaux.

- Dans les travaux de [Kong et al., 2016], on garde trace des évaluateurs ayant apporté une note et l'on est ainsi capable de décliner une esthétique "à la façon de" tel ou tel expert. Ainsi, le subjectivisme est rattrapé en plaçant un observateur au sein d'une boîte particulière, en compétition avec d'autres observateurs dotés de sensibilités différentes. L'utilisateur dispose alors de la capacité de reconnaître "son" esthétique en identifiant le Beau qui lui convient à celui reconnu comme tel par l'un des observateurs. C'est une première façon de reconnaître des points de vue un peu différents sur un œuvre.
- Les travaux de [Park et al., 2017] vont plus loin car ils laissent la place à une préférence personnelle en mêlant, après les couches de réseau de neurones, les résultats de classification obtenue (non supervisée et universelle) et les résultats d'une évaluation subjective faite après une étape d'interaction avec l'utilisateur. C'est donc une démarche qui, comme celles qui mettaient en œuvre des classifieurs et des primitives, pourrait prendre en charge la richesse et l'autonomie de l'observateur. Notons cependant qu'il reste beaucoup de travail à faire pour atteindre des résultats convaincants avec cette approche, tant il est peu clair ce qui fait la différence entre un observateur et un autre. Il semblerait que dans cette direction, les résultats les plus convaincants devraient hériter des approches de profilage actuellement exploitant toute la richesse des données personnelles disponibles sur les réseaux sociaux. Nous examinerons cette piste plus loin.

9.2.4 De quelle image une belle image est-elle l'image ?

Reprenons ici les mécanismes mis en œuvre dans les études les plus avancées aujourd'hui qui impliquent massivement les RNP et essayons de tirer quelques conclusions des choix retenus.

C'est donc dans la base de données qui sert d'apprentissage que se trouve toute "l'expertise esthétique" déployée dans le RNP et c'est aujourd'hui la base AVA qui est universellement retenue pour cela. Rappelons qu'elle est constituée de plus de 250 000 photos, issues du site DPChallenge. A chaque photo sont associées d'une part des méta-données, parfois sommaires, d'autre part des appréciations données par des internautes sous forme de notes et, parfois de commentaires littéraires.

Nous ferons peu de remarques sur les images elles-mêmes. A notre connaissance il n'y a pas d'étude qualifiant cette base de données (excepté l'article où elle fut présentée [Murray et al., 2012]). Les images proposées sont généralement de bonne qualité technique, quoique de taille très modeste, couvrent un très large champ de thématiques, tout à fait compatible avec ce que l'on trouve dans les galeries d'art, les musées ou les collections : portraits, paysages, natures mortes, scènes de foule, scènes d'intérieur ou de plein air, etc. Cette base de données se distingue à ce titre de bases orientées vers la vision par ordinateur comme par exemple ImageNet ou ImageCLEF ou des entrepôts d'images à finalité sociale (Flicker, Instagram), et cela nous semble un point très positif. Il est bien évidemment très difficile de qualifier la représentativité statistique des thématiques, des styles, des compositions proposées. Il serait aussi difficile de fixer un objectif idéal de représentativité sans arrêter clairement le cadre de l'esthétique que nous souhaitons caractériser. Notons cependant que sont très peu représentées des scènes abstraites ou proches de l'abstraction qui constituent une part importante des œuvres d'artistes photographiques (Atkins, Ray, Bernhard). Notons également que très peu d'œuvres utilisent les critères des premiers temps de la photographie, comme ceux que l'on trouve dans les œuvres d'Adget, Cameron, Kertesz, Carjat, dont l'esthétique est souvent dictée par les contraintes techniques, mais qui ont progressivement gagné un public expert. On sait combien importante est la représentativité des bases d'apprentissage dans le résultat des RNP. Il nous semble qu'il y a des enseignements à tirer de la représentativité de la base AVA sur ce point, mais que ces études sont à faire. Notons que l'étude entreprise dans [Murray et al., 2012] pour identifier chaque image dans un référentiel psychologique (en l'occurrence le modèle d'affects circomplexe de Russell⁴ [Russell, 1980]) est une excellente piste qui mériterait d'être approfondie et conduite par des équipes pluridisciplinaires.

Nous serons plus critiques sur les évaluations associées à chaque photo. On sait qu'elles sont nombreuses (entre 100 et 500 selon la photo, plus de 200 en moyenne). On ne connaît pas individuellement ceux qui les ont formulées, et l'on ne dispose pas statistiquement de beaucoup d'information sur la communauté dont ils sont issus. On sait que DPChallenge est fréquenté par un public d'internautes amateurs souvent éclairés, parfois professionnels de la photographie. Leurs profils montrent une grande variété d'âge et de genre, mais il n'est pas clair qu'ils présentent une grande diversité d'origine géographique et culturelle. Il semble qu'il y ait une très forte représentation des villes moyennes de l'Amérique du Nord ainsi que de l'Australie, que les quadras ou quinquas, les mères de famille, les jeunes adolescents, tous disposant d'un peu de temps libre, soient bien représentés. Ils présentent une bonne familiarité avec la photographie et la plupart d'entre eux exposent également leurs photos. Ils contribuent très activement aussi bien aux divers concours qu'à l'évaluation, apportant souvent leurs appréciations à plusieurs milliers de photos. Mais les commentaires qu'ils postent sont trop souvent très vagues et peu informatifs, voire anecdotiques, laissant penser qu'ils sont peu familiers de l'argumentation esthétique où n'en maîtrisent pas les usages. Très majoritairement, les commentaires littéraires sont très positifs et donc pour les seules images bien notées. Ils se résument alors bien souvent à "J'aime!". Il est très rarement fait allusion à des critères esthétiques, ou alors sous une forme très générale : "Très belles couleurs" ou "Bonne maîtrise de l'espace". Les très belles photos de nature (couchers de soleil, forêts brumeuses, plages, sommets enneigés), de fleurs ou d'animaux, sont

4. Le modèle d'affects circomplexe de Russell propose de distribuer les états psychologiques de façon ordonnée sur une circonférence organisée autour de huit affects principaux, régulièrement distribués sur le cercle : le plaisir, l'excitation, l'incitation, la détresse, le déplaisir, la dépression, la torpeur et la relaxation. Il propose alors de rapporter toutes les autres émotions dans ce référentiel, en coordonnées polaires.

très sûrement plébiscitées de même que les portraits de vieillards ou d’enfants s’ils sortent des photos d’identité. Beaucoup de photos sont appréciées pour leur sujet cocasse, humoristique ou grave, plus que pour leurs qualités esthétiques comme en attestent les commentaires. Les photos abstraites dont l’interprétation n’est pas immédiate sont généralement mal cotées ou font l’objet d’avis très dispersés.

On ne peut manquer, en suivant les débats qui accompagnent les évaluations sur le site DPChallenge, d’être frappé par le rôle très important que prend le sujet de la photo au regard de la forme de cette photo. La dimension narrative de la photo est incontestablement une caractéristique importante dans la balance de l’évaluation, renforcée d’une part par le titre de la photo, qui est mis en bonne place et prend souvent une tournure très descriptive, et d’autre part par l’intitulé de la compétition à laquelle se présente la photo, cet intitulé s’imposant à l’évidence comme un cadre de lecture et d’interprétation de la photographie. Tout ceci nous renvoie aux remarques faites au paragraphe 1.1 où nous examinons les raisons d’apprécier une image ne relevant pas de l’appréciation esthétique.

Retenons enfin que l’évaluation finale dans AVA est généralement réduite à la moyenne des évaluations et l’on voit que ce qui va servir à entraîner le RNP est une “vérité” bien particulière, une sorte de beauté statistique, reflet d’un consensus raisonnable d’un petit nombre de “jurés” qui sans être tirés aléatoirement auraient du mal à être considérés comme des parangons de l’esthétique photographique.

Ainsi, le RNP, s’il fait bien son travail (et tout porte à croire qu’à brève échéance il le fera) qualifiera avec une grande fiabilité (95 % ? 98 % ?) les “belles photos”, c’est-à-dire des photos conformes au goût moyen d’un petit groupe d’aficionados, rassemblés autour d’une passion commune pour la photographie. Nous revoilà plongés dans l’univers de Bourdieu, plus que dans celui d’Arnheim ou de Berlyne.

9.3 Comment introduire de la “subjectivité” dans les RNP ?

Devant ces remarques, on peut chercher de trois façons différentes à échapper à ce jugement unique et “objectif” :

1. soit en poursuivant la piste de l’individualisation telle que l’ont proposée par exemple Park *et al.* [Park et al., 2017], mais nous avons vu que cette piste nécessite que l’on améliore les capacités des RNP à apprendre avec un tout petit nombre d’exemples, ce qui n’est pas envisagé à court terme ;
2. soit en multipliant les modèles de sensibilité, de façon que l’utilisateur puisse choisir “son” esthétique ;
3. soit en prenant en compte explicitement la spécificité de l’utilisateur lors de l’expérience, c’est ce que recouvre généralement le terme de “recommandation”.

Nous explorons ici les deux dernières pistes.

9.3.1 Multiplier les avis

Une première façon de briser l’objectivité esthétique du RNP est d’introduire explicitement une subjectivité à l’aide d’experts en charge de représenter des voies concurrentes d’esthétique. Ainsi, on porterait une plus grande attention, au cours de la phase d’apprentissage, à certaines qualités de l’image qui s’adresseraient en priorité à tel ou tel groupe d’observateurs en fonction de sa sensibilité, sa culture, son éducation.

Cette démarche a en fait été timidement abordée dans S. Kong *et al.* [Kong et al., 2016] qui confient l’évaluation de leur base AADB à un petit nombre de vrais experts dont ils gardent trace dans l’expertise. Chaque utilisateur est alors amené à se comparer à tous les experts et celui qui lui ressemble le plus servira seul à constituer la base d’apprentissage utilisée pour prédire la note de chaque nouvelle photo. Notons que les auteurs n’ont pas cherché à identifier des courants esthétiques

précis, ni même à mettre en valeur les écarts de points de vue entre experts. Les différences observées sont le résultat du choix initial (arbitraire?) d'experts qui manifestent naturellement des écarts non recherchés. On pourrait concevoir une attitude plus systématique de désignation d'experts, chacun reflétant un courant particulier de l'esthétique si l'on disposait d'une cartographie raisonnablement consensuelle de ces courants, ce qui à notre connaissance n'existe pas.

Des travaux cependant s'engagent dans cette voie. Ils concernent le jugement esthétique en photographie et sont particulièrement ambitieux. Ils ont pour objectif de déterminer ces communautés aux goûts plus ou moins homogènes dans des populations extrêmement variées (comme par exemple chez les abonnés à Flickr) [Hong et al., 2016]. Ces travaux, encore en cours, utilisent pour chaque image deux types de descripteurs : des descripteurs sémantiques d'une part (issus de tags fournis avec chaque image), des descripteurs de bas niveau d'autre part (couleur, contraste, contours, textures). Dans l'espace de très grande dimension de ces descripteurs, chaque image est traitée par un algorithme d'allocation latente de Dirichlet (LDA), comme un mot. Les mots sont regroupés en phrases (une phrase égale un photographe) qui fournissent donc des sujets (*topics*) latents et définissent des communautés (un *topic* est une communauté) qui partagent des sujets voisins dans l'espace des descripteurs. Ce regroupement dans le graphe des sujets se fait par un algorithme de *graph shift*. Si l'outillage mathématique mis en jeu est considérable, les premiers résultats ne sont encore aujourd'hui guère convaincants. Le poids respectifs des divers descripteurs est délicat à trouver et il est difficile de distinguer ce qui est une tendance esthétique et ce qui est un thème photographique et bien difficile de caractériser de vraies différences esthétiques dans des distinctions encore très grossières (voir la figure 9.3).

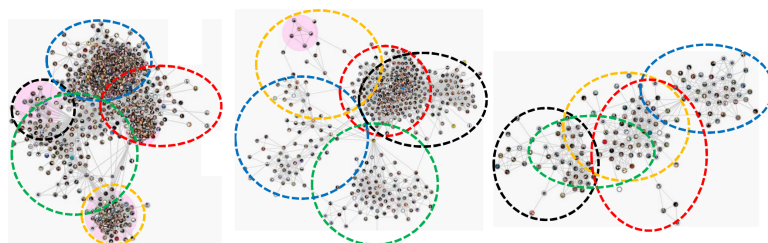


Figure 9.3 – Détection de communautés dans des nuages de descripteurs par une technique d'allocation latente de Dirichlet puis partitionnement de graphe. Les communautés sont identifiées par des ellipses de couleur. En rouge : les “designers”, en bleu : “les couleurs”, en vert : “l’architecture”, en jaune : “le Noir et Blanc”. Les 3 représentations sont des projections particulières du nuage, d’après [Hong et al., 2016]. On voit qu’il y a un important recouvrement entre ces communautés et les thématiques des photos.

Notons que dans la démarche rapportée dans [Kong et al., 2016] on peut reprocher une vision un peu brutale des préférences esthétiques où chaque utilisateur apprécie toute œuvre conformément aux préférences d'un expert et un seul. Ce schéma est probablement meilleur que celui qui aligne tout point de vue sur un point de vue unique mais il ne résiste probablement pas à une analyse très fine.

9.3.2 Les systèmes de recommandation

L'autre solution consiste à mettre en place un système de recommandation pour l'esthétique photographique. Les systèmes de recommandation qui permettent de conseiller à un client d'autres produits susceptibles de lui convenir également, sont opérationnels dans de nombreux domaines (librairie, musique, commerce de détail...). Ceux qui fonctionnent bien s'appuient sur des descriptions sémantiques riches des objets (telles qu'on peut les obtenir par exemple dans le résumé d'un livre ou dans un catalogue descriptif des produits offerts par un magasin). Connaissant les choix antérieurs de l'utilisateur, pour peu qu'ils soient assez nombreux et variés, on est alors capable de définir par un vecteur d'attributs, son “portrait” d'acheteur. Ce portrait est alors comparé à des portraits-types

obtenus soit à l’aide de modèles théoriques, soit par apprentissage sur un grand nombre d’utilisateurs. Les portraits-types disposent d’“avis” sur tous les produits (souvent binaires : “j’aime” ou “je n’aime pas”). Par ailleurs on dispose de statistiques sur l’achat conjointement de tel produit et de tel autre. Ce sont alors les avis les mieux notés par le portrait-type le plus proche de l’utilisateur des produits les plus fréquemment acquis qui lui sont suggérés, éventuellement filtrés à partir de toute connaissance dont on dispose des écarts entre le portrait-type et l’utilisateur.

Dans le domaine du cinéma [Deldjoo et al., 2016, Elahi et al., 2017, Deldjoo et al., 2018] on trouve des travaux qui semblent assez proches du domaine de la photographie d’art. On pourrait espérer tirer profit de l’analyse qu’ils font du signal d’image (choix des contrastes, des couleurs, des cadrages), pour la transposer à l’image fixe. Malheureusement cela semble encore difficile car le rôle de l’image est très secondaire pour la recommandation en matière de films. Celle-ci repose beaucoup sur une information sémantique riche dont ne dispose pas la photographie (le script, la production, le casting, ainsi que l’enchaînement des images, des changements de plans, la bande sonore . . .).

Des travaux récents ont pris pour objet la recommandation dans le domaine de l’art et concernent surtout la vente en ligne de peintures [Dominguez et al., 2017, Messina et al., 2017, Benouaret, 2017, Messina et al., 2018]. Ils nous semblent pouvoir mieux se transposer au domaine de la photographie que ceux concernant le cinéma. Dans le domaine de l’art, on est particulièrement sensible à des aspects complémentaires concernant d’une part le style de l’œuvre, d’autre part le rôle personnel de l’artiste, aspects qui interviennent de la même façon en photographie. Certaines particularités fortes du marché de l’art les distinguent cependant, comme par exemple le fait qu’une œuvre d’art n’existe souvent qu’à un exemplaire et qu’il n’est donc pas possible de donner sur elle l’avis de plusieurs acheteurs puisqu’elle disparaît de la vente après le premier achat. Ces travaux sur la recommandation en matière d’art sont cependant encore peu convaincants. Ces travaux, s’appuyant sur les bases de données d’une galerie d’art en ligne, ont confirmé que, comme pour l’évaluation de la qualité, les approches de recommandation à partir des propriétés de bas niveaux de l’œuvre qui utilisent des RNP sont plus efficaces que celles qui s’appuient sur des classifieurs avec primitives dédiées [Dominguez et al., 2017]. Le rôle de ces primitives visuelles demeure cependant très marginal au regard du rôle de certaines primitives symboliques. Ainsi on n’est pas surpris que la méta-donnée “nom de l’artiste” prenne une place beaucoup plus importante dans la recommandation finale [Messina et al., 2018].

Un système de recommandation a également été développé pour assister les recherches au sein de la base de données BAM! du site artistique Behance (voir paragraphe 6.2) [He et al., 2016]. Il utilise deux modèles concurremment, l’un exploite une chaîne de Markov personnalisée pour maintenir au cours des requêtes le profil des préférences de l’utilisateur, l’autre modélise les apparences visuelles des divers éléments de la scène. Ils créent ainsi un système qu’ils appellent VISTA+. Là encore ce sont des algorithmes de recherche de connaissances latentes qui sont mis en œuvre. Elles doivent permettre de définir : la relation d’un observateur à une œuvre, la relation d’un observateur aux autres observateurs, la relation de dépendance entre les œuvres, et enfin l’évolution temporelle des préférences de l’utilisateur au cours de ses visites.

Signalons enfin, le domaine de la photographie, des travaux assez spécialisés qui se sont intéressés à un aspect très particulier de la recommandation, celle qui préside au choix des filtres d’amélioration des photographies sur les réseaux sociaux. Ces filtres qui agissent sur le contraste et la colorisation de l’image sont en petit nombre (une vingtaine que l’on trouve dans les boîtes à outils des réseaux sociaux), la tâche de recommandation est alors raisonnablement complexe, puisqu’elle consiste à proposer à un utilisateur donné, pour une image particulière, un ou deux filtres adaptés à ses goûts et ses images [Sun et al., 2017]. Le logiciel proposé qui fonctionne par un apprentissage par RNP de préférence sur des couples d’images traitées par des logiciels différents, fonctionne relativement bien dans 80 % des cas.

Chapitre 10

En forme de conclusion

Comment aborder aujourd’hui la relation de l’esthétique et de la photographie, alors même que la photographie vit en ce début du *XXI^e* siècle un cycle nouveau de son histoire brève mais mouvementée, cycle lié à l’évolution de la technique de prise de vue et de diffusion ?

Constatons tout d’abord que l’image photographique est de plein droit objet de l’esthétique et que nul ne lui dénie aujourd’hui, comme l’a fait P. Bourdieu il y a 50 ans, une place parmi les autres arts : peinture, sculpture, musique Selon de nombreux critères, la photographie est même devenue l’un des arts les plus dynamiques et les plus respectés [Fried, 2008, Hariman and Lucaites, 2016, Cotton, 2009].

Notons ensuite que, plus que les autres modalités de l’art, la photographie se prête aux techniques modernes d’investigation, d’une part parce qu’elle est aisément archivable et manipulable par l’ordinateur, d’autre part parce qu’elle a trouvé, au sein de la société du numérique, une place qui la rend très facilement accessible en très grandes quantités et une forme directement intégrée aux chaînes de traitement de l’information.

Nous avons présenté une analyse de l’esthétique en photographie selon divers registres, espérant pouvoir réunir ces champs dans une même conclusion. Ce n’est malheureusement pas le cas car il y a bien peu de passerelles entre ces divers pans de connaissance, où chacun se retranche volontiers derrière son expertise (voir par exemple [Gombrich, 2000, Ramachandran and Freeman, 2001, Bullo et Reber, 2013]) et c’est en déclinant chacun des registres où nous l’avons présentée que l’on peut analyser les progrès des relations entre photographie et esthétique.

Dans le domaine philosophique, l’esthétique est confrontée, depuis 50 ans au moins, au bouleversement de la notion même d’art [Arnheim, 1986, Danto, 1992]. Ce bouleversement a amené à l’avant-scène des œuvres où la notion d’esthétique était pour le moins secondaire, si ce n’est même étrangère. Il s’est accompagné d’une grande confusion dans la pensée occidentale [Danto and Goehr, 2014]. Parallèlement, l’irruption des arts “primitifs”, puis l’ouverture internationale du marché de l’art et son exposition à des traditions et des cultures très variées ont alimenté une réflexion sur la contingence de la beauté et sur le poids considérable des cultures dans sa définition [Wascheck, 2000, Brown and Dissanayake, 2009], cette intrusion de la culture ne rendant pour autant la notion de beauté plus malléable. Cet état de chose s’étend à l’image tout autant qu’à la peinture ou la sculpture.

Dans le domaine de la neuro-biologie, le champ de la neuro-esthétique est toujours en quête de réponses convaincantes à la chronologie des mécanismes de l’appréciation esthétique [Zeki, 1999]. Si un petit nombre d’hypothèses ont été écartées (comme celles du traitement par les aires dédiées d’un hypothétique cortex hédonique), de grandes incertitudes demeurent sur la sélection des flux montants du cortex visuel, ainsi que sur les relations avec les aires en charge du souvenir, du raisonnement ou de l’émotion, ce qui laisse l’interprétation très largement spéculative [Changeux, 2016]. La nécessité de disposer de meilleurs outils d’introspection fonctionnelle (comme par exemple l’IRM de diffusion,

qui met en évidence les connexions entre aires actives) se fait cruellement sentir. De même des outils rapides permettant de relier les signaux issus des voies visuelles aux activations des zones préfrontales seraient indispensables pour étayer des hypothèses de mécanismes d'évaluation esthétique convaincants.

Les expériences de psycho-physiologie expérimentale, très bien adaptées à de nombreuses études sur l'appareil visuel, sur la perception et sur l'interprétation de formes simples sont aujourd'hui beaucoup trop simplifiées dans leurs protocoles expérimentaux pour pouvoir débroussailler utilement le foisonnement des interrogations du praticien ou du théoricien de la photographie confronté à la notion du beau. Les édifices théoriques élaborés par la Gestalt Theorie [Wertheimer, 1938] peinent à se retrouver dans des recommandations faites en termes sémantiquement trop généraux ("zone d'intérêt", "sujet principal", "personnage central" . . .), ou dans des concepts par trop géométriques ("règle du tiers", "centre de fixation" . . .). La vérification à posteriori de ces recommandations (lorsqu'elles ont pu être formulées) par le biais de statistiques conduites après traitement d'images serait certainement un acquis expérimental très appréciable. Ces travaux manquent cependant aujourd'hui de protocoles universellement reconnus, de corpus de données partagés et acceptés et souffrent donc de capacités modestes de conviction.

Enfin, les dernières venues des études sur l'esthétique des photographies, celles qui s'appuient sur les méthodes d'apprentissage automatique et les bases de données du web, s'installent de façon iconoclaste et exagérément aculturée dans le paysage, fortes de potentiels développements sur le marché des nouvelles technologies et de la réputation soigneusement entretenue d'ultime progrès d'une "intelligence artificielle", qui a du mal à échapper à son image de propagande socio-économique. Ces approches sont encore bien loin de convaincre malgré les énormes ressources qu'elles ont su drainer. Les travaux encore préliminaires sont, au mieux, "prometteurs" (ce qui dans le domaine scientifique indique qu'ils doivent faire leurs preuves), au pire, bien décevants. Il est plus juste de dire qu'ils offrent des résultats intéressants mais intrigants dont on ne sait pas bien quoi faire en l'état.

S'il nous faut identifier des raisons de ces blocages dans nos démarches, un élément semble devoir être pointé du doigt que la plupart des auteurs se gardent bien d'aborder, il s'agit de la relation très intime de l'esthétique et de la culture. Niée dans les approches objectivistes, cette relation est immanente dans les approches subjectivistes, mais elle introduit une nouvelle variable, en l'occurrence la culture elle-même dont la nature est au moins aussi complexe et protéiforme que l'esthétique. Si la notion de culture semble détenir de nombreuses clefs pour expliquer l'esthétique, elle apparaît elle-même terriblement hermétique à nos outils actuels d'introspection et semble en l'occurrence remplacer un mystère très grand par un mystère plus grand encore.

Bibliographie

- [Alberti, 1992] Alberti, L. (1992). *De pictura*. Editions Macula.
- [Alvarez et al., 1999] Alvarez, L., Gousseau, Y., and Morel, J. M. (1999). The size of objects in natural and artificial images. *Avances in Imaging & Electron Physics*, 111 :167–242.
- [Amirshahi et al., 2013] Amirshahi, S., Denzler, J., and Redies, C. (2013). Jenaesthetics—a public dataset of paintings for aesthetic research. In *European Conference on Computer Vision (Workshop)*, Jena, Germany.
- [Amirshahi et al., 2014] Amirshahi, S., Hayn-Leichsenring, G., Denzler, J., and Redies, C. (2014). Evaluating the rule of thirds in photographs and paintings. *Art & Perception*, 2(1–2) :163–182.
- [André, 1759] André, J. (1759). *Essai sur le Beau*. Schneider, J.H., Amsterdam.
- [Arbellini, 2017] Arbellini, J. (2017). Les photographies d’art ont-elles un format compatible avec le nombre d’or ? Student project, Institut Villebon-Georges Charpak, Telecom-ParisTech Dept. IDS.
- [Arnheim, 1954] Arnheim, R. (1954). *Art and Visual Perception : A Psychology of the Creative Eye*. University of California Press, Los Angeles (CA).
- [Arnheim, 1983] Arnheim, R. (1983). *The power of the center. A study of composition of the visual arts*. Univ. of California Press.
- [Arnheim, 1986] Arnheim, R. (1986). *New essays on the psychology of art*. University of California Press, Oakland (CAL).
- [Attewell and Baddeley, 2007] Attewell, D. and Baddeley, R. (2007). The distribution of reflectances within the visual environment. *Vision Research*, 47(4) :548–554.
- [Augustin et al., 2011] Augustin, M. D., Carbon, C., and Wagemans, J. (2011). Measuring aesthetics impression. In *ECVP’11*, Toulouse.
- [Augustin et al., 2018] Augustin, M. D., Wagemans, J., Carbon, C. C., Holmes, T., Kapoula, Z., and Roberts, M. N. (2018). Measuring aesthetic impressions. (under revision).
- [Barsalou, 1985] Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of experimental psychology : learning, memory, and cognition*, 11(4) :629.
- [Beardsley, 1966] Beardsley, M. (1966). *Aesthetics from classical Greece to the present*. McMillan, New-York.
- [Bell et al., 1991] Bell, S. S., Holbrook, M. B., and Solomon, M. R. (1991). Combining esthetic and social value to explain preferences for product styles with the incorporation of personality and ensemble effects. *Bell, S. S., Holbrook, M. B., & Solomon, M. R. (1991). Combining esthetic and social value to explain preferences for product styles with the incorporation of personality and ensemble effects. Journal of Social behavior and Personality*, 6(6), 243., 6(6) :243.
- [Benouaret, 2017] Benouaret, I. (2017). *Un système de recommandation contextuel et composite pour la visite personnalisée de sites culturels*. PhD thesis, Thèse de l’Université de Technologie de Compiègne (France).
- [Bense, 1969] Bense, M. (1969). *Einführung in die informationstheoretische Ästhetik,. Grundlegung und Anwendung in der Texttheorie*. Rowoldt Taschenbuch Verlag.

- [Berlyne, 1970] Berlyne, D. (1970). Novelty, complexity and hedonic values. *Perception and Psychophysics*, 8 :279–290.
- [Berlyne, 1971] Berlyne, D. (1971). *Aesthetics and Psychobiology*, volume 336. Appleton-Century Croft, New-York.
- [Berridge and Kringelbach, 2013] Berridge, K. and Kringelbach, M. (2013). Neuroscience of affect : brain mechanisms of pleasure and displeasure. *Current Opin Neurobiology*, 23(3) :294–303.
- [Birkhoff, 1933] Birkhoff, G. (1933). *Aesthetic Measure*. Harvard University Press, Cambridge (USA).
- [Blijlevens et al., 2017] Blijlevens, J., Thurgood, C., Hekkert, P., Chen, L. L., Leder, H., and Whitfield, T. W. (2017). The aesthetic pleasure in design scale : The development of a scale to measure aesthetic pleasure for designed artifacts. *Psychology of Aesthetics, Creativity, and the Arts*, 11, 11(1) :86–97.
- [Bouveau, 2014] Bouveau, C. (2014). *The Painter's Secret Geometry : A Study of Composition in Art*. Allegro editions.
- [Boulez et al., 2014] Boulez, P., Changeux, J., and Manoury, P. (2014). *Les neurones enchantés*. Odile Jacob, Paris.
- [Bourdieu, 1965] Bourdieu, P. (1965). *Un art moyen : essai sur les usages sociaux de la photographie*. Les Editions de Minuit, Paris.
- [Brady and Field, 2000] Brady, N. and Field, D. (2000). Local contrast in natural images : normalization and coding efficiency. *Perception*, 29 :1041–1055.
- [Brehin, 2007] Brehin, Y. (2007). Une lecture de la "beauté réelle" d'Eddy Zemach. *Marges : revue d'art contemporain*, 6 :112–119.
- [Brown and Dissanayake, 2009] Brown, S. and Dissanayake, E. (2009). The Arts are more than Aesthetics : Neuroaesthetics as narrow aesthetics. *Neuroaesthetics*, pages 43–57.
- [Brown et al., 2011] Brown, S., Gao, X., Tisdelle, L., Eickoff, S., and Lotti, M. (2011). Naturalizing aesthetics : Brain areas for aesthetic appraisal across sensory modalities. *Neuroimages*, 58 :250–258.
- [Bullot and Reber, 2013] Bullot, N.-J. and Reber, R. (2013). The artful mind meets art history : Towards a psycho-historical framework for the science of art appreciation. *Behavioural and brain sciences*, 36 :123–180.
- [Cela-Conde et al., 2011] Cela-Conde, C., Agnati, L., Huston, J., Mora, F., and Nadal, M. (2011). The neural foundations of aesthetic appreciation. *Progress in Neurobiology*, 94 :39–48.
- [Chandler, 2013] Chandler, R. (2013). Seven challenges in image quality assessment : past, present and future research. *ISRN Signal Processing*, 2013 :1–53.
- [Changeux, 2008] Changeux, J. (2008). *Du vrai, du beau, du bien : une nouvelle approche neuronale*. Odile Jacob.
- [Changeux, 2016] Changeux, J. (2016). *La beauté dans le cerveau*. Odile Jacob, Paris.
- [Charrier et al., 2012] Charrier, C., Lezoray, O., and Lebrun, G. (2012). Machine learning to design full-reference image quality assessment algorithm. *Signal Processing : Image Communications*, 27-3 :209–219.
- [Chatterjee, 2003] Chatterjee, A. (2003). Prospects for a cognitive neuroscience of visual aesthetics. *Bulletin of Psychology and the Arts*, 4(2) :55–60.
- [Chatterjee and Vartanian, 2016] Chatterjee, A. and Vartanian, O. (2016). Neuroscience of aesthetics. *Annals of the New-York Academy of Sciences*, 1369 :172–194.
- [Chevreul, 1864] Chevreul, E. (1864). *Des couleurs et leurs applications aux arts industriels à l'aide des cercles chromatiques*. J.B. Baillières et Fils, Paris.
- [Chopra et al., 2005] Chopra, S., Hadsell, R., and LeCun, Y. (2005). Learning a similarity metric discriminatively with application to face verification. In Soc., I. C., editor, *CVPR'05*, volume 1.
- [Cotton, 2009] Cotton, C. (2009). *The photograph as contemporary art*, volume 1. London : Thames & Hudson.

- [Crettez, 2017] Crettez, J. (2017). *Les supports de la géométrie interne des peintres*. ISTE Editions, London.
- [Cupchnik et al., 2009] Cupchnik, G., Vartanian, O., Crawley, A., and Mikulis, D. (2009). Viewing artworks : Contributions of cognitive control and perceptual facilitation for aesthetic experience. *Brain and Cognition*, 70 :84–91.
- [Dalal and Triggs, 2005] Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *CVPR conference Computer Vision and Pattern Recognition*, pages 886–893, San-Diego Etats-Unis.
- [Danmasio, 1994] Danmasio, A. (1994). *Descartes' error : emotion, reason and the human brain*. Grosset/Putnam.
- [Danto, 1964] Danto, A. (1964). The artworld. *The journal of philosophy*, 61(19) :571–584.
- [Danto, 1992] Danto, A. (1992). *Beyond the Brillo Box : The Visual Arts in Post-Historical Perspective*. University of California Press.
- [Danto and Goehr, 2014] Danto, A. and Goehr, L. (2014). *After the End of Art : Contemporary Art and the Pale of History*. Princeton Classics.
- [Datta et al., 2006] Datta, R., Joshi, D., Li, J., and Wang, J. (2006). Studying aesthetics in photographic images using a computational approach. In *Computer Vision, ECCV 2006*, volume 3953 of *Lecture Notes in Computer Science*, pages 288–301.
- [Datta et al., 2008] Datta, R., Li, J., and Wang, J. Z. (2008). Algorithmic inferencing of aesthetics and emotion in natural images : An exposition. In IEEE, editor, *15th IEEE International Conference on Image Processing.*, pages 105–108.
- [Datta and Wang, 2010] Datta, R. and Wang, J. (2010). ACQUINE : Aesthetic quality inference engine - real-time automatic rating of photo-aesthetics. In ACM, editor, *MIR'10*.
- [DeCrousaz, 1724] DeCrousaz, J. (1715-1724). *Traité du beau, où l'on montre en quoi consiste ce que l'on nomme ainsi, par des exemples tirés de la plupart des arts et des sciences.*. Fayard (janvier 1985).
- [Delahaye, 2015] Delahaye, J. P. (2015). La beauté mise en formules. *Pour la Science*, 455 :78–83.
- [Deldjoo et al., 2018] Deldjoo, Y., Constantin, M. G., Ionescu, B., Schedl, M., and Cremonesi, P. (2018). Mmtf-14k : A multifaceted movie trailer feature dataset for recommendation and retrieval. In *Proceedings of the 9th ACM Multimedia Systems Conference*, pages 450–455.
- [Deldjoo et al., 2016] Deldjoo, Y., Elahi, M., Cremonesi, P., Garzotto, F., Piazzolla, P., and Quadrana, M. (2016). Content-based video recommendation system based on stylistic visual features. *Journal on Data Semantics.*, 5(2) :99–113.
- [Deng et al., 2017a] Deng, Y., Loy, C. C., and Tang, X. . (2017a). Aesthetic-driven image enhancement by adversarial learning. arXiv preprint arXiv :1707.05251.
- [Deng et al., 2017b] Deng, Y., Loy, C. C., and Tang, X. (2017b). Image aesthetic assessment : An experimental survey. *IEEE Signal Processing Magazine*, 34(4) :80–106.
- [Dessalles, 2008] Dessalles, J. (2008). *La pertinence et ses origines cognitives*. Hermès- Sciences, Paris.
- [Dessalles, 2013] Dessalles, J. (2013). Algorithmic simplicity and relevance. In Dowe, D., editor, *Algorithmic probability and friends*, volume LNAI, pages 119–130, Berlin. Springer-Verlag.
- [Dewar, 1937] Dewar, H. (1937). A comparison of tests of artistic appreciation. *British Psychological Society*, pages 29–49.
- [Dhar et al., 2011] Dhar, S., Ordonez, V., and Berg, T. (2011). High level describable attributes for predicting aesthetics and interestingness. In on, I. C., editor, *Computer Vision and Pattern Recognition (CVPR), 2011*, pages 1657–1664.
- [Di Dio et al., 2011] Di Dio, C., Canessa, N., Cappa, S., and Rizzolatti, G. (2011). Specificity of esthetic experience for artworks : an fmri study. *Frontiers in human Neuroscience*, 5(139) :1–14.

- [Di Dio et al., 2007] Di Dio, C., Macaluso, E., and Rizzolatti, G. (2007). The golden beauty : brain response to classical and renaissance sculptures. *PLoS one*, 2(11) :e1201.
- [Di Dio and Vittorio, 2009] Di Dio, C. and Vittorio, G. (2009). Neuroaesthetics : a review. *Current Opinion in Neurobiology*, 19 :682–687.
- [Diderot, 1769] Diderot, D. (1769). *Correspondance littéraire*. Furne, Paris.
- [Dominguez et al., 2017] Dominguez, V., Messina, P., Parra, D., Mery, D., Trattner, C., and Soto, A. (2017). Comparing neural and attractiveness-based visual features for artwork recommendation. In *Proceedings of the 2nd ACM Workshop on Deep Learning for Recommender Systems*, pages 55–59.
- [Dufrenne, 1967] Dufrenne, M. (1967). *Phénoménologie de l'expérience esthétique*. Presses Universitaires de France, Paris.
- [Dupont, 2015] Dupont, J. (2015). Que faire de l'imagerie cérébrale ? - Territoires anciens et nouveaux d'une technologie. *Comptes Rendus Biologie*, 8-9 :607–612.
- [Elahi et al., 2017] Elahi, M., Deldjoo, Y., Bakhshandegan Moghaddam, F., Cella, L., Cereda, S., and Cremonesi, P. (2017). Exploring the semantic gap for movie recommendations. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pages 326–330.
- [Evans et al., 1992] Evans, A., Collins, D., and Milner, B. (1992). An MRI-based stereotactic atlas from 250 young normal subjects. *Journ of Soc. Neurosci.*, 18 :412.
- [Eysenck, 1939] Eysenck, H. (1939). The general factor in aesthetic judgements. *British Psychological Society*, pages 94–102.
- [Fechner, 1871] Fechner, G. (1871). *Zur Experimentalen Aesthetik*. Hirtzel, S., Leipzig.
- [Fernandez and Wilkins, 2008] Fernandez, D. and Wilkins, A. (2008). Uncomfortable images in art and nature. *Perception*, 37(7) :1098–1113.
- [Freedberg and Gallese, 2007] Freedberg, D. and Gallese, V. (2007). Motion, emotion and empathy in esthetic experience. *Trends in Cognitive Sciences*, 11(5) :197–203.
- [Fried, 2008] Fried, M. (2008). *Why photography matters as art as never before*. New Haven : Yale University Press.
- [Gärdenfors, 2000] Gärdenfors, P. (2000). *Conceptual Spaces, the Geometry of Thought*. MIT Press, a Bradford book, Cambridge (Mass.).
- [Genette, 1997] Genette, G. (1997). *La relation esthétique*. Seuil, Paris.
- [Ghyka, 1931] Ghyka, M. (1931). *Le nombre d'or*. Gallimard, Paris.
- [Gombrich, 1960] Gombrich, E. H. (1960). *Art and Illusion : a study in the psychology of pictorial representations*. Princeton University Press, Wahington D.C.
- [Gombrich, 2000] Gombrich, E. H. (2000). Concerning 'the science of art' : Commentary on ramachandran and hirstein. *Journal of consciousness studies*, 7(8-9) :17–17.
- [Goodman, 1990] Goodman, N. (1990). *Langages de l'art*. Eds. J. Chambon, Paris.
- [Gousseau and Roueff, 2007] Gousseau, Y. and Roueff, F. (2007). Modeling occlusion and scaling in natural images. *SIAM Journal of Multiscale Modeling and Simulation*, 6(1) :105–134.
- [Gygli et al., 2013] Gygli, M., Grabner, H., Riemenschneider, H., Nater, F., and Van Gool, L. (2013). The interestingness of images. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1633–1640.
- [Hariman and Lucaites, 2016] Hariman, R. and Lucaites, J. L. (2016). Photography : The abundant art. *Photography and Culture*, 9(1) :39–58.
- [Hayn-Leichsenring et al., 2017] Hayn-Leichsenring, G. U., Lehmann, T., and Redies, C. (2017). Subjective ratings of beauty and aesthetics : correlations with statistical image properties in western oil paintings. *i-Perception*, 8(3).
- [He et al., 2015a] He, K., Zhang, X., Ren, S., and Sun, J. (2015a). Deep residual learning for image recognition. ArXiv arXiv :1512.03385, Microsoft Beijing.
- [He et al., 2015b] He, K., Zhang, X., Ren, S., and Sun, J. (2015b). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans on PAMI*, 37(9) :1904–1907.

- [He et al., 2016] He, R., Fang, C., Wang, Z., and McAuley, J. (2016). Vista : a visually, socially, and temporally-aware model for artistic recommendation. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 309–316.
- [Hegel, 1838] Hegel, G. (1835-1838). *Vorlesungen über die Ästhetik*.
- [Henry, 1885] Henry, C. (1885). *Introduction à une esthétique scientifique*. Revue contemporaine, Paris.
- [Henry, 1889] Henry, C. (1889). *Éléments d'une théorie générale de la dynamogénie, autrement dit, du contraste, du rythme et de la mesure avec applications spéciales aux sensations visuelles et auditives*. Charles Verdin, Paris.
- [Henry, 1891] Henry, C. (1891). *Physiologie générale des sensations et esthétique*, pages 1–71. Notices sur les travaux scientifiques de M. Charles Henry. Imprimerie des sciences mathématiques et physiques, Via Lombardia, Rome.
- [Henry, 1895] Henry, C. (1895). Quelques aperçus sur l'Esthétique des formes. *La Revue Blanche*, pages 6–61.
- [Hii et al., 2017] Hii, Y. L., See, J., Kairanbay, M., and Wong, L. K. (2017). Multigap : Multi-pooled inception network with text augmentation for aesthetic prediction of photographs. In *IEEE International Conference on Image Processing (ICIP)*, pages 1722–1726.
- [Hoffer and Ailon, 2015] Hoffer, E. and Ailon, N. (2015). Deep metric learning using triplet network. In *Lectures Notes in Computer Sciences N 9370*.
- [Hong et al., 2016] Hong, R., Zhang, L., and Tao, D. (2016). Unified photo enhancement by discovering aesthetic communities from flickr. *IEEE transactions on Image Processing*, 25(3), 25(3) :1124–1135.
- [Howard et al., 2017] Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and H., A. (2017). Mobilenets : Efficient convolutional neural networks for mobile vision applications. arXiv :1704.04861.
- [Hubner and Fillinger, 2016] Hubner, R. and Fillinger, M. (2016). Comparaison of objective measures for predicting perceptual balance and visual aesthetic preferences. *Frontiers in Psychology*, 7(335) :1 :15.
- [Hunter and Askarinejad, 2015] Hunter, M. R. and Askarinejad, A. (2015). Designer's approach for scene selection in tests of preference and restoration along a continuum of natural to manmade environments. *6, 1228.*, 6 :1228.
- [Hurlberg and Ling, 2007] Hurlberg, A. and Ling, Y. (2007). Biological components of sex differences in color preference. *Current Biology*, 17(16) :R623–R625.
- [Hurlberg and Ling, 2012] Hurlberg, A. and Ling, Y. (2012). Understanding colour perception and preference. In Best, J., editor, *Colour design, theories and applications*, volume 12, chapter 5, pages 129–158. Woodhead Publishing.
- [Ibarra et al., 2017] Ibarra, F. F., Kardan, O., Hunter, M. R., Kotabe, H. P., Meyer, F. A., and Berman, M. G. (2017). Image feature types and their predictions of aesthetic preference and naturalness. *Frontiers in psychology*, 8 :632–648.
- [Ingarden, 2011] Ingarden, R. (2011). *Esthétique et ontologie de l'œuvre d'art. Choix de textes 1937-1969*. J. Vrin, Paris.
- [Ishizu and Zeki, 2011] Ishizu, T. and Zeki, S. (2011). Toward a brain-based theory of beauty. *PLOS One*.
- [Isola et al., 2011] Isola, P., Xiao, J., Torralba, A., and Oliva, A. (2011). What makes an image memorable? In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 145–152.
- [Itten, 1973] Itten, J. (1973). *The art of color*. Wiley, New-York.
- [Jacobsen et al., 2006] Jacobsen, T., Schubotz, R., Höfel, L., and Cramon, D. (2006). Brain correlates of aesthetic judgment of beauty. *Neuroimage*, 29 :276–285.
- [Jin et al., 2016a] Jin, B., Ortiz-Segovia, V., and Süsstrunk (2016a). Image aesthetic predictors based on weighted CNNs. In *ICIP, Int. Conf. on Image Processing 2016*, volume 1, pages 2291–2295.

- [Jin et al., 2009] Jin, W., Ho, H., and Srihari, R. (2009). OpinionMiner : a novel machine learning system for web opinion mining and extraction. In *Proc ACM SIGKDD, KDD'09*, pages 1195–1204, New-York.
- [Jin et al., 2016b] Jin, X., Chi, J., Peng, S., Tian, Y., Ye, C., and Li, X. (2016b). Deep image aesthetics classification using inception modules and fine-tuning connected layer. In *8th IEEE International Conference on Wireless Communications & Signal Processing (WCSP)*, pages 1–6.
- [Jullien, 2017] Jullien, C. (2017). Les mathématiques : le langage de la beauté. *Maths Langages express*, pages 93–98.
- [Jullien, 2003] Jullien, F. (2003). *La grande image n'a pas de forme*. Editions du Seuil.
- [Kairanbay et al., 2017] Kairanbay, M., See, J., Wong, L. K., and Hii, Y. L. (2017). Filling the gaps : Reducing the complexity of networks for multi-attribute image aesthetic prediction. In *IEEE International Conference on Image Processing (ICIP)*, pages 3051–3055.
- [Kant, 1985] Kant, E. (1985). *Critique de la faculté de juger*. Gallimard, Folio-Essais, Paris.
- [Kao et al., 2017] Kao, Y., He, R., and Huang, K. ((2017)). Deep aesthetic quality assessment with semantic information. *IEEE Transactions on Image Processing*, 26(3) :1482–1495.
- [Kao et al., 2016] Kao, Y., Huang, K., and Maybank, S. (2016). Hierarchical aesthetic quality assessment using deep convolutional neural networks. *Signal Processing : Image Communication*, 47 :500 :510.
- [Kaplan et al., 1972] Kaplan, S., Kaplan, R., and Wendt, J. S. (1972). Rated preference and complexity for natural and urban visual material. *Perception & Psychophysics*, 12(4) :354–356.
- [Kardan et al., 2017] Kardan, O., Demiralp, E., Hout, M. C., Hunter, M. R., Karimi, H., Hanayik, T. and Yourganov, G., Jonides, J., and Berman, M. G. (2017). Is the preference of natural versus man-made scenes driven by bottom-up processing of the visual features of nature? *Frontiers in psychology*, 6 :471–484.
- [Kawabata and Zeki, 2004] Kawabata, H. and Zeki, S. (2004). Neural correlates of beauty. *Journal of Neurophysiology*, 91(4) :1699–1705.
- [Ke et al., 2006] Ke, Y., Tang, X., and Jing, F. (2006). The design of high-level features for photo quality assessment. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 419–426.
- [Kemp, 1990] Kemp, M. (1990). *The Science of Art, Optical themes in western art from Brunelleschi to Seurat*. Yale university Press.
- [Kim et al., 2015] Kim, W. H., Choi, J. H., and Lee, J. S. (2015). Subjectivity in aesthetic quality assessment of digital photographs : Analysis of user comments. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 983–986.
- [Koch et al., 2010] Koch, M., Denzler, J., and Redies, C. (2010). 1/f 2 characteristics and isotropy in the fourier power spectra of visual art, cartoons, comics, mangas, and different categories of photographs.
- [Kong et al., 2016] Kong, S., Shen, X., Lin, Z., Mech, R., and Fowlkes, C. (2016). Photo aesthetics ranking network with attributes and content adaptation. In *ECCV European Conference on Computer Vision*, pages 662–679. Springer International Publishing.
- [Koshelev et al., 1998] Koshelev, M., Kreinovich, V., and Yann, Y. (1998). Towards the use of aesthetics in decision making : Kolmogorov complexity formalizes Birkhoff's idea. In *Digitalcom-mons@UTEP*.
- [Kotabe et al., 2017] Kotabe, H. P., Kardan, O., and Berman, M. G. (2017). The nature-disorder paradox : A perceptual study on how nature is disorderly yet aesthetically preferred. *Journal of Experimental Psychology : General*, 146(8), 1126., 145(8) :1126–1142.
- [Kringelbach, 2005] Kringelbach, M. (2005). The human orbito-frontal cortex : linking reward to hedonic experience. *Nat. Rev Neurosciences*, 6 :691–702.

- [Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.
- [Kuehn and Gallinat, 2012] Kuehn, S. and Gallinat, J. (2012). The neural correlates of subjective pleasantness. *Neuroimage*, 61 :289–294.
- [Lacey et al., 2011] Lacey, S., Hagtvædt, H., Patrick, V. M., Anderson, A., Stilla, R., Deshpande, G., HU, X., Sato, J., Reddy, S., and Sathian, K. (2011). Art for reward’s sake : Visual art recruits the ventral striatum. *Neuroimage*, 55(1) :420–433.
- [Lange and James, 1922] Lange, C. and James, W. (1922). *The emotions*. Williams & Wilkins.
- [Laughlin, 1987] Laughlin, S. (1987). Form and function in retinal processing. *Trends in Neurosciences*, 10(11) :478–483.
- [Le Bihan, 2012] Le Bihan, D. (2012). *Le cerveau de cristal*. Odile Jacob, Paris.
- [Leder et al., 2004] Leder, H., Belke, B., Oeberst, A., and Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgements. *British Journal of Psychology*, 95 :489–508.
- [Leder et al., 2006] Leder, H., Carbon, C. C., and Ripsas, A. L. (2006). Entitling art : Influence of title information on understanding and appreciation of paintings. *Acta psychologica*, 121(2) :176–198.
- [Ling and Hurlberg, 2007] Ling, Y. and Hurlberg, A. (2007). A new model for color preference : Universality and individuality. In *Color and Imaging Conference*, pages 8–11.
- [Lipps, 1903] Lipps, T. (1903). *Ästhetik : Psychologie des Schönen und der Kunst*. Voss, L. Hamburg.
- [Liu et al., 2010] Liu, L., Chen, R., Wolf, L., and Cohen-Or, D. (2010). Optimizing photo composition. *Computer Graphics Forum*.
- [Livingstone, 2002] Livingstone, M. (2002). *Vision and Art : the biology of Seeing*. Harry N. Abrams. Inc., Publishers.
- [Livio, 2008] Livio, M. (2008). *The golden ratio : the story of Phi, the world’s most astonishing number*. Broadway Books.
- [Lo et al., 2012] Lo, K., Liu, K., and Chen, C. (2012). Assessment of photo aesthetics with efficiency. In *IEEE International Conference on Pattern Recognition (ICPR)*,, pages 2186–2189.
- [Lo et al., 2013] Lo, K., Liu, K., and Chen, C. (2013). Intelligent photographing interface with on-device aesthetic quality assessment. In Parks, J. and Kim, J., editors, *ACCV 2012 Workshop*, volume LNCS 7729, pages 533–544, Berlin. Springer Verlag.
- [Lu et al., 2014a] Lu, P., Kuang, Z., Peng, X., and Li, R. (2014a). Discovering harmony : a hierarchical colour harmony model for aesthetics assesment. In *Asian Conference on Computer Vision*, pages 452–467. Springer Int.
- [Lu et al., 2015a] Lu, P., Peng, X., Li, R., and Wang, X. (2015a). Towards aesthetics of image : A Bayesian framework for color harmony modeling. *Signal Processing : Image Communication*, 39 :487–498.
- [Lu et al., 2016] Lu, P., Peng, X., Zhu, X., and Li, R. (2016). An EL-LDA based general color harmony model for photo aesthetics assessment. *Signal Processing*, 120 :731–745.
- [Lu et al., 2014b] Lu, X., Lin, Z., Jin, H., Yang, J., and Wang, J. (2014b). RAPID : Rating pictorial aesthetics using Deep Learning. In ACM, editor, *22nd Int. Conf on Multimedia MM’14*, pages 457–466, Orlando, Fl.
- [Lu et al., 2014c] Lu, X., Lin, Z., Jin, H., Yang, J., and Wang, J. Z. (2014c). Rating image aesthetics using deep learning. *IEEE Transactions on Multimedia*, 17(11) :2921–2034.
- [Lu et al., 2015b] Lu, X., Lin, Z., Shen, X., Mech, R., and Wang, J. Z. (2015b). Deep multi-patch aggregation network for image style, aesthetics, and quality estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 990–998.
- [Luo and Tang, 2008] Luo, Y. and Tang, X. (2008). Photo and video quality evaluation : Focusing on the subject. In Forsyth, D. and Zisserman, A., editors, *ECCV 2008*, volume LNCS 5304, pages 386–399, Berlin-Heidelberg. Springer-Verlag.

- [Ma et al., 2017] Ma, S., Liu, J., and Chen, C. (2017). A-lamp : Adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment. arXiv :1704.00248.
- [Machado and Cardoso, 1998] Machado, P. and Cardoso, A. (1998). Computing aesthetics. In Springer, editor, *Brazilian symposium on Artificial Intelligence*, pages 219–228, Berlin, Heidelberg.
- [Machajdik and Hanbury, 2010] Machajdik, J. and Hanbury, A. (2010). Affective image classification using features inspired by psychology and art theory. In *MM'10, ACM Conference*, pages 83–92, Firenze (It.).
- [Mai et al., 2011] Mai, L., Le, H., Niu, Y., and Liu, F. (2011). Rule of thirds detection from photograph. In *IEEE Int. Symp on Multimedia (ISM)*, pages 91–96.
- [Mai et al., 2016] Mai, L., Niu, Y., and Liu, F. (2016). Composition preserving deep photo aesthetic assessment. In *Proc. IEEE Conf. Comp. Vision and Pattern Recognition*, pages 497–506.
- [Maitre, 2015] Maitre, H. (2015). *Du photon au pixel : l'appareil photographique numérique*. ISTE Editions, Londres, Royaume Uni.
- [Malinowski-Charles, 2004] Malinowski-Charles, S. (2004). Entre rationalisme et subjectivisme : l'esthétique de jean-pierre de crousaz. *Revue de théologie et de philosophie*, 136 :7–21.
- [Marchesotti et al., 2011] Marchesotti, L., Perronnin, F., Larlus, D., and Czurka, G. (2011). Assessing the aesthetic quality of photographs using generic image descriptors. *Int. Conf. on Computer Vision*, pages 1784–1791.
- [Marchesotti et al., 2013] Marchesotti, L., Perronnin, F., and Meylan, F. (2013). Learning beautiful (and ugly) attributes. In *BMVC*, volume 7, pages 1–11.
- [Markowsky, 1992] Markowsky, G. (1992). Misconceptions about the golden ratio. *The College of Mathematics*, 23(1) :2–19.
- [McManus et al., 1985] McManus, I., Edmonson, D., and Rodger, J. (1985). Balance in picture. *British J. of Psychology*, 76 :311–324.
- [McManus et al., 1981] McManus, I., Jones, A., and Cottrell, J. (1981). The aesthetic of colour. *Perception*, 10 :651–666.
- [McManus et al., 2011] McManus, I., Stöver, K., and Kim, D. (2011). Arnheim's gestalt theory of visual balance : Examining the compositional structure of art photographs and abstract images. *i-Perception*, 2(6) :615–647.
- [Mehrabian, 1977] Mehrabian, A. (1977). Individual differences in stimulus screening and arousability. *J. Personality and Social Psychology*, 45 :237–250.
- [Merleau-Ponty, 1945] Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. NRF, Gallimard, Paris.
- [Messina et al., 2017] Messina, P., Dominguez, V., Parra, D., Trattner, C., and Soto, A. (2017). Exploring content-based artwork recommendation with metadata and visual features. arXiv preprint arXiv :1706.05786.
- [Messina et al., 2018] Messina, P., Dominguez, V., Parra, D., Trattner, C., and Soto, A. (2018). Content-based artwork recommendation : integrating painting metadata with neural and manually-engineered visual features. *User Modeling and User-Adapted Interaction*, Springer, pages 1–40.
- [Mittal et al., 2013] Mittal, A., Soundararajan, R., and Bovik, A. C. (2013). Making a "complete blind" image quality analyzer. *IEEE Signal Processing Letters*, 20 (3) :209–212.
- [Moles, 1957] Moles, A. A. (1957). Théorie de l'information et perception esthétique. *Revue Philosophique de la France et de l'Etranger et Flammarion*, pages 233–242.
- [Molnar, 1974] Molnar, F. (1974). Experimental aesthetics or the science of art. *Leonardo*, 7(1) :23–26.
- [Molnar, 1977] Molnar, F. (1977). Mouvement des yeux et l'hypothèse des explorations épistémiques et diversives. *Journal National de la Société Française de Psychologie*.
- [Molnar, 1981] Molnar, F. (1981). *Towards Science in Art*, pages 385–413. Advances in intrinsic motivation and aesthetics. Hill, A., ed., Plenum Press, New-York.

- [Molnar, 1997] Molnar, F. (1997). A science of vision for visual art. *Leonardo*, 30(3) :225–232.
- [Moon and Spencer, 1944a] Moon, P. and Spencer, D. (1944a). Aesthetic measure applied to color harmony. *Journal Optical Society of America*, 34 :234–242.
- [Moon and Spencer, 1944b] Moon, P. and Spencer, D. (1944b). Area in color harmony. *Journal Optical Society of America*, 34 :93–103.
- [Moon and Spencer, 1944c] Moon, P. and Spencer, D. (1944c). Geometric formulation of classical color harmony. *Journal of the Optical Society of America*, 34(1) :46–63.
- [Moorthy and Bovik, 2011] Moorthy, A. K. and Bovik, A. C. (2011). Blind image quality assessment : From natural scene statistics to perceptual quality. *IEEE transactions on Image Processing*, 20 (12) :3350–3364.
- [Morin, 2016] Morin, E. (2016). *Sur l'esthétique*. Robert Laffont.
- [Moshagen and Thielsch, 2010] Moshagen, M. and Thielsch, M. T. (2010). Facets of visual aesthetics. *International journal of human-computer studies*, 66(10) :689–709.
- [Motoyoshi et al., 2007] Motoyoshi, I., Ishida, S., Sharan, L., and Adelson, E. (2007). Image statistics and the perception of surface qualities. *Nature*, 447 :206–209.
- [Murray and Gordo, 2017] Murray, N. and Gordo, A. (2017). A deep architecture for unified aesthetic prediction. arXiv preprint arXiv :1708.04890.
- [Murray et al., 2012] Murray, N., Marchesotti, L., and Perronnin, F. (2012). AVA : a large -scale database for for aesthetic visual analysis. In *CVPR : Intern. Conf. on Comp. Vision and Pattern Recognition*, pages 2408–2415.
- [Murray et al., 2013] Murray, N., Marchesotti, L., and Perronnin, F. (2013). Methods and systems for ranking images using semantic and aesthetic models (us patent -20140351264).
- [Neveux, 1995] Neveux, M. (1995). *Le nombre d'or, radiographie d'un mythe*. Editions du Seuil, Paris.
- [Nodine et al., 1993] Nodine, C., Locher, P., and Krupinski, E. (1993). The role of formal art training on perception and aesthetic judgement of art compositions. *Leonardo*, 26(3) :219–227.
- [Obrador et al., 2010] Obrador, P., Schmitt-Hackenberg, L., and Oliver, N. (2010). The role of image composition in image aesthetics. In *Proc of the 17th IEEE Int. Conf. on Image Processing*, pages 3185–3188, Hong-Kong.
- [O'Connor, 2010] O'Connor, Z. (2010). Colour harmony revisited. *Color Research & Applications*, 35(4) :267–273.
- [Ou et al., 2004] Ou, L., Luo, M., Woodcock, A., and Wright, A. (2004). A study of colour emotion and colour preference : Part i colour emotions of single colours. *Color research and applications*, 29(3) :232–240.
- [Palmer, 2010] Palmer, S. E. and Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences, USA*, 107(19) :8877–8882.
- [Panofsky, 1983] Panofsky, E. (1983). *Meaning in the visual arts*. University of Chicago Press.
- [Park et al., 2017] Park, K., Hong, S., Baek, M., and Han, B. (2017). Personalized image aesthetic quality assessment by joint regression and ranking. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1206–1214.
- [Pearce et al., 2016] Pearce, M. T., Zaidel, D. W., Vartanian, O., Skov, M., Leder, H., Chatterjee, A., and Nadal, M. (2016). Neuroaesthetics : the cognitive neuroscience of aesthetic experience. *Perspectives on Psychological Science*, 11(2) :265–279.
- [Pennington et al., 2014] Pennington, J., Socher, R., and Manning, C. (2014). Glove : Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.
- [Platon, 2011] Platon (2011). *Platon, œuvres complètes*. Flammarion.
- [Pouivet, 2006] Pouivet, R. (2006). *Le réalisme esthétique*. Presses universitaires de France, Paris.

- [Pouivet, 2014] Pouivet, R. (2014). Goodman et la reconception de l'esthétique. *Rue Descartes*, 80 :4–19.
- [Ramachandran and Freeman, 2001] Ramachandran, V. S. and Freeman, A. (2001). Sharpening up 'the science of art'. *Journal of Consciousness Studies*, 8(1) :9–30.
- [Reagle, 2013] Reagle, J. (2013). Revenge rating and tweak critique at photo.net. personal website : <http://reagle.org/joseph/2013/photo.net.html>.
- [Reber et al., 2004] Reber, R., Schwarz, N., and Winkielman, P. (2004). Processing fluency and aesthetic pleasure : Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, 8(4) :364–382.
- [Redi et al., 2017] Redi, M., Liu, F., and O'Hare, N. (2017). Bidding the aesthetics gap : The wild beauty of web imaginary. In ACM, editor, *ICMR'17 conference*, pages 242–250, Bucarest.
- [Rigau et al., 2008] Rigau, J., Feixas, M., and Sbert, M. (2008). Information aesthetic measures. *IEEE Computer Graphics and Applications*, 2 :24–34.
- [Rosenberg and Klein, 2015] Rosenberg, R. and Klein, C. (2015). *Art, Aesthetics and the brain*, chapter 5 - The moving eye of the beholder : Eye tracking and the perception of paintings, pages 79–108. Oxford University Press.
- [Russell, 1980] Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6) :1161–1178.
- [San Pedro and Siersdorfer, 2009] San Pedro, J. and Siersdorfer, S. (2009). Ranking and classifying attractiveness of photos in folksonomies. In *Proc. ACM conf on World Wide Web WWW'09*, pages 771–780, New-York.
- [San Pedro et al., 2012] San Pedro, J., Yeh, T., and Oliver, N. (2012). Leveraging user comments for aesthetic aware image search reranking. In ACM, editor, *Proceedings of the 21st international conference on World Wide Web*, pages 439–448.
- [Schifanella et al., 2015] Schifanella, R., Reddi, M., and Aiello, L. (2015). An image is worth more than thousand favourites : surfacing the hidden beauty of Flickr pictures. ArXiv : 1505.03358v2.
- [Schloss and Palmer, 2011] Schloss, K. B. and Palmer, S. E. (2011). Aesthetic response to colour combinations : preference, harmony and similarity. *Attention Perceptual Psychophysics*, 73 :551–571.
- [Schmidhuber, 1997] Schmidhuber, J. (1997). Low-complexity art. *Leonardo, Journal of the International Society for the Arts, Sciences, and Technology*, 30(2) :97–103.
- [Schmidhuber, 2007] Schmidhuber, J. (2007). Simple algorithmic principles of discovery, subjective beauty, selective attention, curiosity and creativity. arXiv : 0709.0674.
- [Schwarz et al., 2016] Schwarz, K., Wieschollek, F., and Lensch, H. (2016). Will people like your image? arXiv 1611-05203, University of Tübingen.
- [Schweinhart and Esock, 2013] Schweinhart, A. and Esock, E. (2013). Structural content in paintings : Artists overregularize oriented content of paintings relative to the typical natural scene bias. *Perception*, 42 :1311–1332.
- [Simond et al., 2015] Simond, F., Arvanitopoulos, N., and Süsstrunk, S. (2015). Image aesthetics depends on context. In *IEEE International Conference on Image Processing*, pages 3788–3792. IEEE.
- [Skov, 2009] Skov, M. (2009). Neuroaesthetic problems : a framework for neuroaesthetic research. In Baywood, editor, *Neuroaesthetic*, pages 9–36, Amityville (N.Y).
- [Smith, 2008] Smith, D. (2008). Color-person-environment relationship. *Color Research & Applications*, 33 :312–319.
- [Solso, 1996] Solso, R. (1996). *Cognition and the visual Arts*. The MIT Press, Cambridge (Mass.), 5th edition.
- [Sperber and Wilson, 2004] Sperber, D. and Wilson, D. (2004). Relevance theory. In Horn, L.R. & Ward, G., editor, *The Handbook of Pragmatics*, pages 607–632. Oxford : Blackwell.

- [Srivastava and Kant, 2018] Srivastava, M. M. and Kant, S. (2018). Visual aesthetic analysis using deep neural network : model and techniques to increase accuracy without transfer learning. arXiv preprint arXiv :1712.03382.
- [Su et al., 2012] Su, H. H., Chen, T. W., Kao, C. C., Hsu, W. H., and Chien, S. Y. (2012). Preference-aware view recommendation system for scenic photos based on bag-of-aesthetics-preserving features. *IEEE Transactions on Multimedia*, 14(3) :833–843.
- [Sun et al., 2017] Sun, W. T., Chao, T. H., Kuo, Y. H., and Hsu, W. H. (2017). Photo filter recommendation by category-aware aesthetic learning. *IEEE Transactions on Multimedia*, 19(8) :1870–1880.
- [Swami, 2013] Swami, V. (2013). Context matters : Investigating the impact of contextual information on aesthetic appreciation of paintings by Max Ernst and Pablo Picasso. *Psychology of Aesthetics, Creativity, and the Arts*, 7(3) :285–296.
- [Szegedy et al., 2015] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2015). Rethinking the inception architecture for computer vision. arXiv :1512.00567.
- [Talairach and Tournoux, 1988] Talairach, J. and Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain :3-dimensional proportional system-An approach to cerebral imaging*. Thieme Medical Publishers, New-York.
- [Talebi and Milanfar, 2017] Talebi, H. and Milanfar, P. (2017). NIMA : Neural image assessment. arXiv : 1709.0541v1.
- [Tatarkiewicz, 1970] Tatarkiewicz, W. (1970). *History of aesthetics*. Mouton ed.
- [Taylor et al., 2013] Taylor, C., Clifford, A., and Franklin, A. (2013). Color preferences are not universal. *Journal of experimental psychology*, 142(4) :1015–1027.
- [Ticini et al., 2014] Ticini, L. F., Rachman, L., Pelletier, J., and Dubal, S. (2014). Enhancing aesthetic appreciation by priming canvases with actions that match the artist’s painting style. *Frontiers in human neuroscience*, 8 :391.
- [Tversky and Kahneman, 1981] Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211 :451–458.
- [Ulrich, 1983] Ulrich, R. S. (1983). *Behavior and the natural environment*, chapter 3 - Aesthetic and affective response to natural environment, pages 85–125. Springer, Boston, MA.
- [Uttal, 2002] Uttal, W. (2002). Precis of the new phrenology : the limits of localizing processes in the brain. *Brain and Mind*, 3 :221–228.
- [Vartanian and Skov, 2014] Vartanian, O. and Skov, M. (2014). Neural correlates of viewing paintings : evidence from a quantitative meta analysis of fmrimaging data. *Brain Cogn.*, 87 :52–56.
- [Veryzer Jr and Hutchinson, 1998] Veryzer Jr, R. W. and Hutchinson, J. W. (1998). The influence of unity and prototypicality on aesthetic responses to new product designs. *Journal of consumer research*, 24(4) :374–394.
- [Vidal, 2011] Vidal, F. (2011). La neuroesthétique, esthétique scientiste. *Revue d’histoire des sciences humaines*, 25(2) :239–264.
- [Vidal, 2012] Vidal, F. (2012). Neuroaesthetics : Getting rid of art and beauty. *Biosocieties*, 7(2) :209.
- [Vienot and Le Rohellec, 2012] Vienot, F. and Le Rohellec, J. (2012). Colorimétrie et physiologie : la spécification LMS. In Fernandez-Maloinie, C., Robert-Inacio, F., and Macaire, L., editors, *Couleur numérique : acquisition, perception, codage et rendu*, chapter 1, pages 19–40. Hermès-Lavoisier, Paris.
- [Vinyals et al., 2015] Vinyals, O., Toshev, A., Bengio, S., and Erhan, D. (2015). Show and tell : A neural image caption generator. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3156–3164.
- [Wallis, 2007] Wallis, J. (2007). Orbitofrontal cortex and its contribution to decision making. *Annu. Rev. Neurosciences*, 30 :31–56.
- [Wang et al., 2014] Wang, W., Liu, J., Zhao, W., and Li, J. (2014). A system of image aesthetic classification and evaluation using cloud computing. In Li, S., Liu, C., and Wang, Y., editors, *Chinese Conference on Pattern Recognition. CCPR 2014.*, volume 443, Communications in Computer and Information Science, pages 183–195. Springer, Berlin, Heidelberg.

- [Wang et al., 2018] Wang, W., Yang, S., Zhang, W., and Zhang, J. (2018). Neural aesthetic image reviewer. arXiv preprint arXiv :1802.10240.
- [Wang et al., 2016] Wang, Z., Chang, S., Dolcos, F., Beck, D., and Huang, T. (2016). Brain-inspired deep network for image aesthetic assesment. arXiv : 1601.04155v2.
- [Wascheck, 2000] Wascheck, M. (2000). *Le chef d'œuvre : un fait culturel*, chapter 2, pages 25–46. Qu'est-ce qu'un Chef d'œuvre ? Gallimard (Paris).
- [Wertheimer, 1922] Wertheimer, M. (1922). Untersuchungen zur Lehre von der Gestalt - i - prinzipelle Bemerkungen. In *Psychologische Forschung*, volume 1, pages 47–58.
- [Wertheimer, 1923] Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt - ii - laws of organization in perceptual forms. In *Psychologische Forschung*, volume 4, pages 301–350.
- [Wertheimer, 1938] Wertheimer, M. (1938). *Laws of of organization in perceptual forms*. Ellis, W.D., A source book of Gestalt psychology, London.
- [Whitfield and Slatter, 1979] Whitfield, T. W. and Slatter, P. E. (1979). The effects of categorization and prototypicality on aesthetic choice in a furniture selection task. *British Journal of Psychology*, 70(1) :65–75.
- [Wilber et al., 2017] Wilber, M. J., Fang, C., Jin, H., Hertzmann, A., Collomosse, J., and Belongie, S. (2017). Bam! the Behance artistic media dataset for recognition beyond photography. In *Proc. Int. Conf. Comp. Vision*, volume 1, page 4.
- [Yao et al., 2012] Yao, L., Suryanarayan, P., Qiao, M., Wang, J. Z., and Li, J. (2012). Oscar : On-site composition and aesthetics feedback through exemplars for photographers. *International Journal of Computer Vision*, 96(3) :353–383.
- [Yarbus, 1967] Yarbus, A. (1967). *Eye movements and vision*. Plenum Press.
- [Zeki, 1999] Zeki, S. (1999). *Inner vision : An exploration of art and the brain*. Oxford University Press.
- [Zemach, 1987] Zemach, E. M. (1987). Aesthetic properties, aesthetic laws and aesthetic principles. *The journal of aesthetics and art criticism*, 46(1) :239 :251.
- [Zemach, 2005] Zemach, E. M. (2005). *La beauté réelle : une défense du réalisme esthétique*. Presses Universitaires de Rennes ; collection "Aesthetica".
- [Zhang et al., 2016] Zhang, J., Sclaroff, S., Lin, Z., Shen, X., Price, B., and Mech, R. (2016). Unconstrained salient object detection via proposal subset optimization. In *IEEE Conf. on Comp. Vision and Pattern Recognition (CVPR)*, pages 5733–5742.

Index

- A-Lamp, méthode d'évaluation, 93
- AADB, (base d'images), 68
- A&C, méthode d'évaluation, 91
- accutance, 45
- Acquine, méthode d'évaluation, 73
- affect circomplexe, 104
- aires visuelles, 35
- Alberti, L.B., 23
- Alberti,L.B., 18
- AlexNet, architecture de RNP, 83
- Alhazem, 22
- amazement*, 10
- appraisal*, 25
- apprentissage numérique, 64
- approche
 - objectiviste, 18
 - subjectiviste, 23
- Arnheim, R., 41, 44
- art, 12
- artisanales (primitives), 71
- attrait, 66
- Augustin d'Hippone, 21
- AVA, (base d'images), 67
- AVA-Comments (base de données), 95
- AVA2, (base d'images), 67

- bag of visual words* (BOVW), 79
- BAM (base de données d'œuvres d'art), 68
- bases de données, 65
 - Photo.Net, 66
 - AADB, 68
 - AVA, 67
 - AVA2, 67
 - BAM, 68
 - BEAUTY, 68
 - CHUK-BD, 66
 - DPChallenge, 66
 - FACT, 69
 - Flickr, 66
 - ImageCLEF, 66
 - ImageNet, 66
 - JenAesthetics, 68
 - Uni Tübingen, 68
- beauté, 12
- BEAUTY, (base d'images), 68
- Bense, M., 59

- Berlyne, D. E., 25
- big data*, 64
- Birkhoff, D., 57
- Bourdieu, P., 13
- brain-inspired* RNP, méthode d'évaluation, 92
- Brown, S., 33, 37

- cadrage, 44
- CAFFE, environnement de RNP, 87
- canon (esthétique), 19
- centre hédonique, 32
- Changeux, J.-P., 24, 38
- Chevreul, E., 49, 57
- CHUK-DB, (base d'images), 66
- CIELab (métrique), 49
- classification, 71
- cognition, (aires de la), 36
- composition, 41, 44
- cortex
 - insulaire, 33
 - préfrontal, 32
- couleur, 47
- Cuda-Convnet, 93
- Cupchnik, G., 35

- d'Ockam, G., 22
- de Crousaz, J.-P., 24
- Dessalles, J.L. , 10
- Di Dio, C., 36
- diagonales, 44
- Diderot, D., 18, 24
- DMA-Net, méthode d'évaluation, 87
- DPChallenge, (base d'images), 66
- Du Bos, J.-B., 24

- échantillonnage de Shannon, 46
- embodiment*, 36
- esthétique, 12
 - contextualisée ou historique, 29
 - empirique, 28
 - fonctionnelle, 28
 - interactionniste, 26
 - psycho-historique, 29
 - réaliste, 28
- experts, 65

- FACD, (base de données), 69
 Fechner, G.T., 10, 13, 41
 flou, 45
 format d'aspect, 83
 format d'image, 40
 fuyantes, 44
- Gärdenfors, U., 10
 Gestalt, (théorie de la), 58
 Goethe, J.W., 49
 GoogLeNet, réseau de neurones, 87
- hand-crafted, (features)*, 71
 hédonique (centre), 32
 Henry, C., 57
high-keys, 46
 histogramme, 46
 Hutcheson, F., 24
- icône, 21
 ImageCLEF, (base d'images), 66
 incarnation (*embodiment*), 36
 Inception, réseau de neurones, 95
 IRMF, imagerie par résonance magnétique fonctionnelle, 31
 Ishizu, T., 32
 Itten, J., 49
- James, W., 10
 JenAesthetics (base de peintures), 68
- Kant, E., 19, 24, 25
- Leonardo da Vinci, 23
 Locke, J., 24
low-keys, 46
- machine learning*, 64
 masses de données, 64
 Matsuda, Y., 51
 mémoire, (aires de la), 36
 méthode d'évaluation
 A-Lamp, 93
 A&C, 91
 Acquine, 73
 brain-inspired RNP, 92
 DMA-Net, 87
 MNA, 87
 MTRL-CNN, 90
 NAIR, 95
 NIMA, 94
 OSCAR, 77
 RAPID, 85
 MNA, méthode d'évaluation, 87
 Mobile-Net, réseau de neurones, 95
 Moles, A., 59
- Molnar, F., 25, 41
 Moon, P. et Spencer, D.E., 49
 MTRL-CNN, méthode d'évaluation, 90
 Munsell, A., 49
 mutualisation (*pooling*), 83
- NAIR, méthode d'évaluation, 95
 neuro-biologie, 31
 neuro-esthétique, 25
 neurones miroirs, 37
 NIMA, méthode d'évaluation, 94
 nombre d'or, 41
- objectiviste, (approche), 18
 OSCAR, méthode d'évaluation, 77
- palette chromatique, 48
 perceptibilité, 20
 personnification (*embodiment*), 36
 pertinence, 10
 Photo.net, (base d'images), 66
 piqué, 45
 Platon, 17, 18
Prägnanz, 27
 primitives, 71
 artisanales, 71
 de haut niveau, 75
 psychologie expérimentale, 39
- RAPID, méthode d'évaluation, 85
 rapport d'aspect, 40
 recommandation (système de), 106
 règle du tiers, 42
 réseau de neurones
 adverse, 89
 Inception, 95
 Mobile-Net, 95
 profonds (RNP), 64, 83
 ResNet, 91
 siamois, 89
 VGG, 85, 87, 95
 ResNet, réseau de neurones, 91
 résonance magnétique fonctionnelle (IRMF), 31
reward, 32
 RNP : réseau de neurones profonds, 64, 83
- Saint-Augustin, 17
 Schmidhuber, J., 60
 Shannon, (échantillonnage, théorème), 46
 Socrate, 17, 18
 Solso, R.L., 25
 SPARK, *photographic contest*, 14, 102
 spectre de densité de puissance, 45
 Sperber, D. and Wilson, D., 10
 subjectiviste, (approche), 23

TensorFlow, bibliothèque logicielle, 95
texture, 45
théorème de Shannon, 46
Tversky, A. et Kahneman, D., 10

Vartanian, O., 36
VGG, réseau de neurones, 85, 87, 95
VISTA+, (système de recommandation), 107

Yarbus, A., 41

Zeki, S., 32
Zemach, E., 19