

# Juger du Beau avec subjectivité : le défi de l'esthétique computationnelle

## Assessment of Beauty with subjectivity: a challenge for computational aesthetics

Henri Maître<sup>1</sup>

<sup>1</sup>LTCI, Télécom Paris, Institut Polytechnique de Paris, France, henri.maitre@telecom-paris.fr

**RÉSUMÉ.** Les techniques à base d'intelligence artificielle dont l'objectif est d'évaluer automatiquement la qualité esthétique d'une photographie, ont reçu une attention notable ces dernières années et peuvent se targuer de performances prometteuses. On constate cependant que la plupart d'entre elles souffrent de limitations en raison de leur paradigme de base emprunté à l'esthétique platonicienne, qui attribue tous les critères de beauté à l'objet ou à la personne belle. A partir du très volumineux corpus consacré à l'esthétique depuis 25 siècles, ces limitations auraient pu être anticipées. Le reproche le plus fréquemment exprimé est que le jugement porté sur l'image ne prend pas en compte l'observateur et sa subjectivité. Sans surprise, plusieurs travaux très récents s'attaquent ce point délicat, adoptant des approches diverses. Nous les discutons ici.

**ABSTRACT.** Artificial Intelligence based photography beauty assessment have received a great attention in the last 25 years. They may now claim noticeable performances in replacing the human observer. However, they face limitations which are rooted in the basic choices of the machine learning stage, borrowed from the old Platonism, i.e. the poor place left to a specific observer in the assessment value. Several tracks are explored to short-cut these limitations, based on very different approaches. This paper is presenting an overview of these proposals to re-introduce subjectivity in computational aesthetic assessment and to discuss their foundations.

**MOTS-CLÉS.** Esthétique, beauté, photographie, réseaux de neurones, subjectivité, évaluation.

**KEYWORDS.** Aesthetics, beauty assessment, artificial intelligence, subjectivity, photography, recommendation.

La beauté est souvent regardée comme un attribut mineur et superflu d'un monde grave. L'intérêt qu'on y porte est fréquemment tenu pour futile et mondain. Le jugement de Pâris nous rappelle qu'il a pu pourtant avoir de lourdes conséquences en dehors du champ de l'esthétique. On n'est pas surpris alors devant l'ampleur des commentaires qui se sont attachés, tout au long des siècles, à la désignation d'Aphrodite, bien au-delà de l'enjeu du bénéfice d'une pomme. "Jugement de goûts" dirait Kant. Voir ! Ce n'était pas la mode du temps des Grecs. La beauté était affaire d'harmonie et de symétrie (non pas notre symétrie géométrique, reflet à l'identique, mais une proportion accordée entre le tout et les parties). Le jugement n'y était pour rien ; le beau était affaire de nombres et s'imposait à chacun. Nul besoin d'un pâtre pour en décider. Voir sur ce point encore, puisque ce pâtre, (par hasard ?) était lui-même fort beau et réputé amateur d'un sexe qui était encore faible ! On a dit qu'il y avait d'autres valeurs que la beauté dans les arguments du débat : des promesses de victoires militaires, des espoirs de règne magnifique, l'amour exceptionnel de la femme la plus enviée. Ces arguments contournent l'obstacle, comme le fait l'argutie qu'il fallait en venir à la guerre et que le chemin le plus court ne pouvait qu'être le meilleur. On sait le choix de Pâris, mais comment a-t-il tranché ? A-t-il mesuré pour chaque déesse l'accord de ses formes aux canons de Polyclète ? A-t-il, comme le pensent Descartes ou Hegel, usé de réflexion pour conclure rationnellement sur les arguments de ses sens ? S'est-t-il plutôt, comme le suggère Kant, laissé envahir par l'intuition d'un sens commun impartialement éprouvé ? Ou envahi par la jouissance dionysiaque des trois déesses comme l'aurait fait Nietzsche ?

Vingt-cinq siècles d'esthétique en débattent longuement entre Objectivistes ("l'objet est beau par ses

vertus propres") et Subjectivistes ("l'objet est beau dès que je le vois beau"). Les psychologues expérimentaux [Fechner, 1871, Leder et al., 2004, Hurlberg and Ling, 2012], les sociologues [Elridge, 2015, Ray, 2020], puis les neuro-biologistes [Di Dio and Vittorio, 2009, Brown et al., 2011, Ishizu and Zeki, 2011] ont apporté leurs voix à ce débat et depuis peu les informaticiens se sont joints à eux, qui ont fait du jugement esthétique un nouveau défi pour leurs algorithmes déjà fortement sollicités par le jeu d'échecs, la prédiction boursière ou le diagnostic médical.

## 1. Que fait l'informatique pour la beauté ?

Les premières propositions d'utiliser les sciences dures au service de l'esthétique remontent à Charles Henry, mais son *Introduction à une esthétique scientifique* de 1885 n'a guère fait d'émules. Il n'en est pas de même des travaux du mathématicien George Birkhoff [Birkhoff, 1933] qui ont donné naissance à une longue lignée de recherches attachées à formaliser une "équation de la beauté", travaux qui ont bénéficié de divers progrès de nos connaissances au cours du XX<sup>e</sup> siècle : théorie de la Gestalt, morphologie mathématique, reconnaissance des formes, traitement des images [Eysenck, 1941, Moles, 1957, Bense, 1969, Rigau et al., 2008].

Nous allons ici nous intéresser spécifiquement à la mesure de la beauté des photographies pour trois raisons essentiellement. D'une part les photos ont pris dans notre société une place exceptionnelle en raison des progrès et de la prolifération des systèmes d'acquisition ainsi que de la popularité des réseaux sociaux. D'autre part la photographie a accédé à la reconnaissance d'un art de plein droit doté de tous les attributs exigibles : musées, expositions, experts, marché et cotation, revues . . . Enfin, une forte demande sociale et sociétale appelle la mise en place de techniques d'évaluation rapides et efficaces de la beauté des photographies, tant pour les besoins individuels de gestion de ses archives personnelles que pour des secteurs économiques qui font très abondamment appel aux images : édition, publicité, agences de voyage, d'immobilier, de mode . . .

Nous allons nous pencher particulièrement sur ces approches, nées à la fin du siècle dernier, qui, au lieu de mettre en place une formule explicite de la beauté comme l'ont fait Birkhoff et ses successeurs, s'appuient sur les techniques d'apprentissage automatique et d'intelligence artificielle et mettent à profit des bases de données d'images annotées par des jugements esthétiques. Deux séries de travaux se sont succédé, explorant cette piste. Tout d'abord, de 1994 à 2015 environ, des travaux pionniers ont mis en œuvre la détection de primitives pertinentes (appelées souvent *handmade features*<sup>1</sup>), et de techniques simples et robustes de classification. La plupart de ces travaux utilisaient des connaissances explicites des règles de la photographie et des recommandations formulées par les spécialistes (composition de la scène, mise au point, harmonie des couleurs . . .) [Datta and Wang, 2010, Luo and Tang, 2008, Dhar et al., 2011] mais des approches moins expertes ont mis à profit les primitives issues du traitement de l'image (textures, contours . . .) [Lo et al., 2013] ou de la vision par ordinateur (SIFT, SURF . . .) [Marchesotti et al., 2011, Marchesotti et al., 2015] avec des performances aussi bonnes. Ces techniques à base de primitives ont laissé la place, vers 2015, à des techniques à base de réseaux de neurones profonds, qui font l'impasse sur l'extraction de primitives et ignorent les savoirs qui ont conduit à les choisir. Les images à tester sont

1. *Handmade features* pourrait se traduire par primitives artisanales ; ce sont des configurations particulières des pixels (lignes, contours, coins, couleurs . . .) qui seraient des signatures locales de certaines recommandations photographiques (règle du tiers, contraste objet/fond, distribution des points d'intérêt, construction, etc.).

placées à l'entrée du réseau qui a été entraîné pour évaluer l'esthétique de l'image et le résultat s'affiche en sortie : *image belle* ou *image banale* dans le cas fréquent d'un classement binaire, ou évaluation selon une échelle de 1 à 10 dans les cas d'un jugement continu. Les performances des techniques à base de réseaux de neurones profonds ont très largement surpassé celles des primitives artisanales et aujourd'hui ces seules techniques survivent pour proposer à l'utilisateur d'éliminer des images médiocres de grandes collections ou de sélectionner celles qui émergent.

Revenant au débat qui fait l'objet des premières lignes de ce texte, il faut constater que l'approche "objectiviste" s'impose dans ces deux démarches. Le diagnostic de l'algorithme, quel qu'il soit, est universel : l'image est belle ou non, indépendamment de l'observateur, par ses seules propriétés internes : sa composition, ses textures, ses couleurs, l'assemblage de ses lignes, etc. L'observateur n'a pas de rôle dans cette décision et ne sait quels critères l'ont emporté. L'algorithme adopte une esthétique platonicienne : la beauté appartient à l'image, non à celui qui en fait l'expérience.

Expliquons la méthodologie suivie par les auteurs de ces travaux. Tout d'abord une collection d'images est rassemblée, chacune - et cela est important - dotée d'une évaluation censée exprimer sa qualité esthétique. Les sites spécialisés en photographie ont été abondamment utilisés à cet effet (Flickr, Photo.net, Instagram, DpChallenge). Selon le public auquel ils s'adressent (grand public, amateurs éclairés, professionnels), ils proposent des photos de toute qualité et de toute nature : du novice à l'expert, du reportage, de la mode, de l'art, du sport, etc. Ils sont fréquemment associés de jugements, soit sous forme de votations lors de compétitions thématiques, soit d'expertises de professionnels, soit encore de simples commentaires postés par des internautes. Ainsi ont été créées quelques bases de données rassemblant au moins des dizaines de milliers de photos auxquelles sont attachées leur note et parfois les commentaires qu'ont posté les internautes. Elles ont pour noms AVA, AADB ou BEAUTY (voir le tableau 1.1) [Murray et al., 2012, Schifanella et al., 2015, Redi et al., 2017].

Ces bases de données servent à entraîner un réseau neuromimétique, généralement l'un de ceux qui ont fait leurs preuves dans les compétitions de reconnaissance des formes (ResNet, VGG, Mobile-Net, GoogLeNet, Inception) en présentant successivement chaque photo à l'entrée du réseau et en imposant sa note associée en sortie. Après avoir figé les paramètres du réseau, on peut procéder à l'évaluation d'une photo inconnue en la plaçant à l'entrée du réseau. On lit alors en sortie son évaluation.

## 1.1. *Les bases de données et l'expertise*

Elles sont brièvement présentées dans le tableau 1.1. et plus longuement détaillées dans [Maître, 2021].

C'est sur la base de données AVA que les algorithmes sont généralement testés [Murray et al., 2012]. AVA regroupe plus de 250 000 photos, chacune accompagnée de notes comprises entre 0 et 10 attribuées par les internautes et issues du site DpChallenge un site spécifiquement dédié aux photographes. Seules les photos ayant reçu au moins 200 notes sont retenues dans AVA, et à chacune est généralement attribuée pour note définitive la moyenne de ces notes.

De la base AVA sont déduites la base AVA-2 et la base CUHK-DB en ne retenant que les meilleures et les plus mauvaises images, ainsi que la base AVA-PD (AVA-Photographer Demographic) en ne retenant que les seules images de certains photographes dont on connaît quelques informations (genre, âge, pays).

BEAUTY et la base de Redi et ont été évaluées par *crowdsourcing*, c'est-à-dire par une consultation

critères	BEAUTY	AADB	Base de Redi	AVA	AVA-2 CUHK-DB	AVA-PD	Uni Tübingen	Psycho Flickr	Flickr AES
taille ( $\times 1000$ )	15	10	100	250	50	120	380	60	40
qualité esthétique	faible	faible	faible	haute	haute	haute	faible	haute	faible
score esthétique	3	5	4	10	10 / 2	10	10	5	5
classes sémantiques	4	non	oui	44	44	44	non	oui	oui
étiquettes de style	non	11	non	14	14	14	non	non	non
annotation	non	oui	oui	oui	oui	oui	non	non	oui
origine	Flickr	Flickr	web	DP Chal.	AVA	AVA	Flickr	Flickr	Flickr

**Tableau 1.1.:** Quelques bases de données fréquemment utilisées pour l'étude esthétique des images avec certaines de leurs propriétés. 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 *score* représente l'excursion de la notation. Les *annotations* sont composées des commentaires littéraux qui accompagnent parfois les avis.

rémunérée des internautes. AADB a été évaluée par 5 experts qui ont attribué le score esthétique lorsque leurs avis convergeaient raisonnablement. Flickr-AES a sélectionné des photos de Flickr qui ont été annotées sur des critères esthétiques, par *crowdsourcing*. Psycho-Flickr a sélectionné des images de Flickr produites par des photographes (300 photographes « professionnels » de Flickr) pour lesquels ont été dressés les profils psychologiques (émetteur et récepteur) à l'aide des Cinq Majeures (voir ci-dessous).

## 1.2. Résultats

Les stratégies divergent quant aux résultats recherchés. Certains algorithmes visent à reproduire une évaluation continue entre 0 et 10, certains cherchent à retrouver la distribution des évaluations qu'ont données les internautes, d'autres enfin - et ce sont les plus nombreux - proposent un choix binaire : image « belle » ou « pas-belle » et là encore divers choix sont faits. On choisit souvent de séparer « belles » ou « pas-belles » selon la note attribuée, à l'aide d'un paramètre  $\delta$  par rapport à la valeur moyenne  $M$  de l'ensemble des 250 000 photos ; si la note est supérieure à  $M + \delta$  l'image est belle, si elle est inférieure à  $M - \delta$ , elle n'est pas belle. Lorsque  $\delta = 0$ , plusieurs équipes obtiennent plus de 80 % de résultats corrects sur la base AVA [Sheng et al., 2018, Ma et al., 2017, Talebi and Milanfar, 2017, Deng et al., 2017b] (en séparant bien sûr les photos servant à l'apprentissage de celles qui serviront à la vérification) et ces performances augmentent encore si l'on utilise, en plus de l'image, quelque information sémantique sur le thème concerné par la photographie (portrait, paysage, vie quotidienne, etc.) de façon à spécialiser un peu le traitement.

Ces résultats sont indéniablement intéressants puisqu'ils permettent de trier rapidement, parmi des milliers de photographies, celles qui présentent de bonnes qualités esthétiques, sans trop en laisser de côté, mais ils s'exposent cependant à des critiques diverses. Elles peuvent concerner la pertinence des bases d'apprentissage qui reflètent un genre photographique très actuel répondant à des codes sociaux précis (abondance des photos d'animaux, des portraits, ...) et utilisant une forme rendue très accessible par la technologie qui fait souvent appel à des jeux de couleurs, de haute résolution, d'incrustations ou d'effets spéciaux. Elles peuvent concerner les avis donnés qui souffrent, comme la plupart des avis collectés sur internet, de biais et de dispersion [Reagle, 2013, Pasquier et al., 2014, Cochoy, 2011]. Mais un reproche plus profond s'attache à l'hypothèse que toute photo est jugée identiquement, belle ou non,

par tous les observateurs. Cette hypothèse objectiviste que la base de données annotée des jugements des internautes reflète une propriété caractéristique de chaque photo, propriété qui conduit à un « jugement de goût » unique, partagé par tous, qui permet de paramétrier un réseau capable de reproduire ce « jugement de goût » universel, fait l'objet d'objections nombreuses et a suscité nombre de travaux qui ont cherché à y échapper. Ce sont ces tentatives que nous voulons décrire ici qui ont pour but d'introduire de la subjectivité dans cette automatisation de l'esthétique.

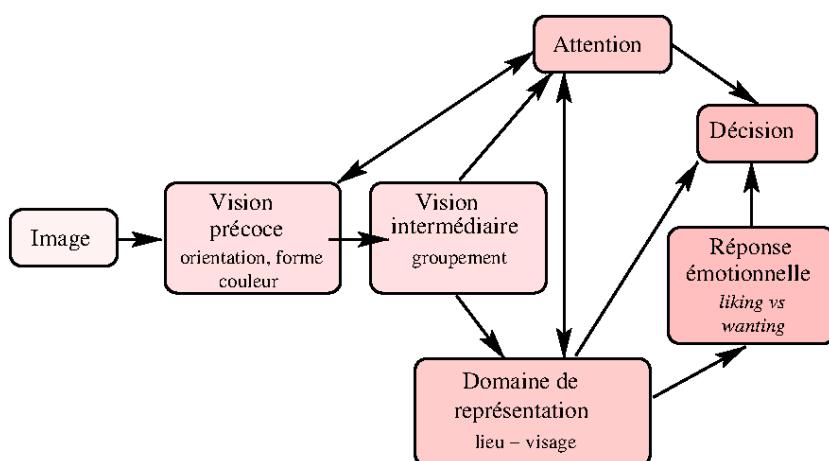
## 2. La subjectivité du goût, c'est quoi ?

Mais il nous faut tout d'abord essayer de cerner un peu ces concepts d'objectivité et de subjectivité qui se fondent subtilement l'un en l'autre. Tout commence avec la vision. Il est bien établi aujourd'hui que c'est de l'objet que partent les rayons lumineux qui atteignent la rétine et qu'ils ne jaillissent pas de l'œil comme on a pu le croire un temps. Ce sont eux qui portent le signal de beauté dans une vision platonicienne. Dans cette pensée, les rayons « sont » beaux et le seront pour tout observateur : la Victoire de Samothrace reste belle une fois les portes du Louvre fermées. Captés par la rétine où ils reçoivent de premiers traitements, ces signaux suivent alors les nerfs optiques et rejoignent les aires visuelles où ils sont mis en forme selon des traitements aujourd'hui assez bien connus : sélections chromatiques, sélections fréquentielles spatiales et temporelles, filtrages, amplifications, groupements multiéchelles, etc. C'est la « vision simple » de Dretske [Dretske, 1969]. Le modèle computationnel de Marr en rend compte assez bien dans son *primal sketch* [Marr, 1982], tandis que la formulation psycho-physiologique de la Gestalttheorie en explique de nombreux mécanismes [Wertheimer, 1938]. A ce point, nous perdons un peu trace de leur parcours et nous n'avons plus qu'une information *a posteriori* sur leur action dans les diverses aires du cerveau. S'il y a beauté, on convient aujourd'hui de le décider car il y a plaisir et activation de zones en charge de récompense (probablement dans le cortex préfrontal) [Berridge and Kringelbach, 2013], mais beaucoup d'autres aires sont également activées, qui sont ordinairement concernées lorsqu'on sollicite la mémoire, la prise de décision, l'anticipation de l'action, etc., [Brown et al., 2011, Chatterjee and Vartanian, 2016]. On distinguerait ce plaisir issu de la beauté car à la différence de beaucoup d'autres il ne semble pas provenir de la satisfaction d'un besoin de notre être : faim, soif, désir sexuel ...

Comment est-on passé de ce signal optique à ce plaisir attesté par notre conscience ? Comment les associations de lignes, de formes, de couleurs deviennent-elles pour nous harmonie, élégance, grâce, poésie ? Comment passe-t-on d'une information objective, mesurable, partageable et potentiellement universelle à un ressenti intime, éminemment personnel et, disons-le, subjectif ? Dans cette transition, les avis divergent largement et les théories se contredisent. De Kant et Schopenhauer qui affirment que cette perception du beau est spontanée, intuitive, impartageable, inaccessible au raisonnement et néanmoins commune à tous, de Descartes et Hegel qui font de cette conviction le fruit de la raison, de Henry ou Séailles qui invoquent la spécificité de la « machine » cérébrale à extraire le beau comme l'oreille capte le son, de Bell, Zemach ou Arnheim qui voient, à travers l'assemblage des signaux optiques la construction de « formes signifiantes », de « prédictats esthétiques », de « qualia » qui, sous des conditions favorables seront décodés justement par des consciences convenablement équipées.

L'affaire est complexe et la compréhension de ce qui reste pour nous une alchimie est lointaine encore. Les schémas qui tentent d'associer nos diverses facultés dans l'élaboration de cette conscience sont

nombreux [Chatterjee, 2004, Brown et al., 2011, Leder et al., 2004, Redies, 2015, Koelsch et al., 2015, Hsu, 2009] mais encore largement débattus (voir les figures 1 et 2). Dans tous ces schémas, le jugement du goût est en grande dépendance d'éléments étrangers à l'image perçue, qu'il s'agisse des éléments de contexte transmis par nos sens (ambiance, cadre), des éléments de contexte de notre état interne à l'instant particulier (humeur) ou de façon plus permanente (tempérament), ainsi que des éléments acquis : notre éducation, notre culture, notre expérience, l'information dont on dispose sur l'objet observé, etc, tous ces éléments que l'on doit appeler subjectifs puisqu'ils sont étroitement attachés à un observateur, et que certains disent uniquement acquis, tandis que d'autres supposent en partie innés. Ces schémas accordent très généralement le rôle de jugement du beau aux couches les plus évoluées de notre système cérébral. Seul, peut-être, le modèle neuro-géométrique de Petitot [Petitot, 2008] permet d'expliquer notre capacité d'intuition ou de « survenance » qui semble au cœur de nombreuses expériences esthétiques par des structures câblées très proches des aires visuelles.

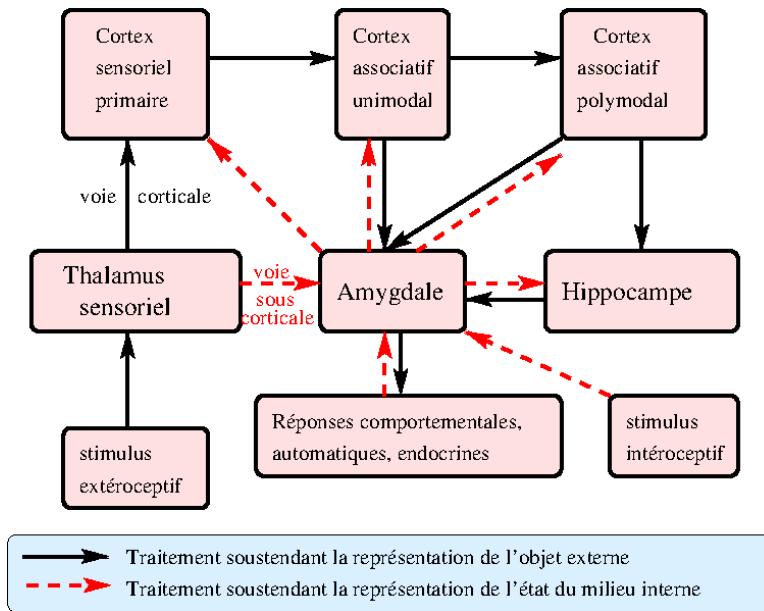


**Figure 1.:** L'un des modèles fonctionnels de la perception esthétique : celui de A. Chatterjee. Deux fonctions différentes concernent l'une l'attention, l'autre la représentation. Elles sont en interaction bi-univoque. L'attention réagit sur l'étage de vision précoce en dirigeant l'exploration. Elle est en partie influencée par la réponse émotionnelle qu'elle contribue à élaborer et elle agit sur notre décision en particulier à travers les réactions réflexes. La représentation est en relation avec nos modèles du monde et intervient également dans la prise de décision (la justification du jugement esthétique) et la réponse émotionnelle (adapté de [Chatterjee, 2004]).

### 3. Quelles approches pour introduire de la subjectivité dans le jugement de goût ?

En l'absence d'un schéma subjectiviste définitivement établi, il n'est pas simple de proposer un algorithme pour accomplir ces tâches de jugement de goût. En comparaison, l'approche objectiviste qui attribue tous les mérites du beau au seul objet et fait partager à tous les observateurs un identique jugement semble triviale, surtout si l'on dispose d'une machine « magique », le réseau neuro-mimétique profond apte à reproduire toute déduction pour peu qu'on la nourrisse d'assez d'exemples.

Voyons, à travers les travaux récents, comment peut être abordée cette subjectivisation du jugement du beau par ordinateur.



**Figure 2.:** Un autre modèle de perception esthétique, celui de Li-Hsiang Hsu, déduit du modèle de A. Damasio. C'est le stimulus extéroceptif (par l'intermédiaire des voies visuelles) qui constitue l'entrée du signal qui distribue les informations vers une représentation du monde extérieur d'une part (la perception physique de l'objet) et vers une représentation proprioceptive (modification de l'état de l'observateur) qui exprime les émotions ressenties à la vue de l'objet. La voie sous-corticale, vers le système limbique (hippocampe, amygdale, gyrus singulaire, hypothalamus), est courte, tandis que l'autre, vers le néo-cortex (vers les zones occipitales des aires visuelles d'une part, vers les zones préfrontales et orbitofrontales d'autre part) est longue. C'est la voie cognitive qui est associée à une valuation (d'après [Hsu, 2009]). Notons que les deux modèles présentés ici sont encore en concurrence avec d'autres et qu'aucun n'a véritablement été validé.

### 3.1. Les systèmes de recommandation

Les techniques les plus répandues aujourd'hui pour adapter une offre à une demande non explicitement formulée est de recourir aux systèmes de recommandation. Ils sont particulièrement efficaces pour proposer à un client des produits répondant à ses goûts dans le domaine du cinéma, de la musique, de l'achat en règle générale et du loisir en particulier [Deldjoo et al., 2016, Elahi et al., 2017, Deldjoo et al., 2018]. Pour cela on recueille des achats passés du demandeur, ainsi que toute information pertinente sur son âge, son lieu de vie, son statut social, voire ses goûts dans des domaines connexes. A partir de ces informations on classe ce client dans un groupe dont les pratiques sont semblables. On peut alors lui proposer des offres que ce groupe a unanimement appréciées, ou, si on souhaite prédire ses réactions à venir sur un produit particulier, on lui attribuera l'avis moyen des autres membres du groupe ou celui de l'autre client qui lui ressemblera le plus au sein du groupe.

Ces techniques sont mal adaptées pour prédire le jugement de goût en matière de photographie à moins d'avoir pu constituer auparavant des mesures de goûts de grandes populations sur de mêmes images. Rappelons que, si la base de données AVA collecte des appréciations, elle ne sait pas qui les attribue et n'est pas à même de construire ces groupes aux goûts semblables. Par ailleurs il est difficile de recueillir les éléments pertinents des goûts esthétiques d'une personne sans la soumettre à un questionnement qui n'aurait pas vraiment sa place ici. Extraire de quelques images déjà appréciées des caractéristiques permettant une catégorisation rapide de l'observateur, est autrement difficile que le faire à partir d'une liste de films ou de morceaux de musiques. Dans ces derniers cas, la catégorie s'appuiera probablement sur des propriétés sémantiquement très riches : sujet du film, nom du producteur, des acteurs, type d'action, ou du morceau musical : genre, interprète, orchestration, etc., propriétés auxquelles on ne peut généralement pas accéder de manière directe.

ment accéder à partir de simples photos.

Dans le domaine de la vente d'œuvres d'art en ligne, des services de recommandation existent également [Dominguez et al., 2017, Benouaret, 2017, Messina et al., 2018, He et al., 2016], qui semblent s'approcher de notre problème. Dans ces systèmes, il apparaît cependant que, si la peinture ou la photo entrent bien en ligne de compte dans la recommandation par leur style et leur composition, le rôle de ces informations issues de l'image reste mineur face au rôle considérable de données plus sémantiques comme le nom de l'auteur, sa cote dans le marché, le prix de l'œuvre, ou les achats antérieurs du client.

### **3.2. Rechercher un profil social**

Dans les systèmes décrits ci-dessus, le profil recherché est avant tout un profil social : âge, genre, profession, environnement familial, hobbies, etc. [Kosinski et al., 2013, Kosinski et al., 2014] ; il est très probable que ces déterminants interviennent dans les goûts de l'usager, mais peu d'études ont aujourd'hui établi de liens exploitables pour en déduire les préférences esthétiques de l'internaute. Certains chercheurs prennent en compte les pratiques de l'internaute sur les réseaux sociaux pour modifier l'ordre des images proposées par les moteurs de recherche à ses requêtes [Cui et al., 2014]. Aucun critère de goût n'est ici utilisé, seules les pratiques (adhésions, connexions, activités) sont prises en compte.

Les travaux rapportés dans [Kairanbay et al., 2019] montrent quelle information on peut tirer d'images postées sur son site par un photographe. Pour cela, ils exploitent un sous-ensemble de photos, très spécifique, issu de la base AVA, pour lequel on dispose exceptionnellement de données sociales (âge, genre, pays d'origine) sur les photographes qui les ont prises (la base appelée AVA-PD, *Photographer Demographic*). Les auteurs montrent alors qu'un profil réduit à ces seules données peut être en partie déduit des photos qu'il poste. Ils montrent ensuite que l'on peut, dans une certaine mesure, prédire l'évaluation qu'il portera à une photo donnée à partir des photos que poste ce photographe. Pour obtenir ce résultat, les auteurs adoptent un schéma classique d'évaluation du goût par transfert d'apprentissage complété d'une étape de réglage fin (*fine tuning*) à l'aide des images postées par le photographe.

D'autres auteurs ont cherché à exploiter les images postées sur les réseaux sociaux (le plus souvent Flickr, Instagram ou Pinterest) par un internaute pour cerner ses centres d'intérêt en examinant le style et le contenu de ces images [Lovato et al., 2013, Yang et al., 2015, You et al., 2016]. Ces travaux diffèrent par leurs approches, mettant l'accent soit sur les caractéristiques de l'image, soit sur ses thématiques. Ils conduisent à des résultats assez convaincants quant à la capacité des techniques d'apprentissage à prédire l'intérêt que portera un photographe à une photo donnée, mais ces travaux ne montrent pas le rôle du jugement esthétique dans cet intérêt.

### **3.3. Définir un profil psychologique de l'utilisateur**

#### **3.3.1. Personnalité et Cinq Majeures**

Le profil psychologique de l'observateur est souvent considéré comme l'un des déterminants fondamentaux du goût esthétique, avant la culture, l'humeur ou le contexte [Konecni, 1979, Jacobsen, 2010]. De nombreuses études classiques se sont penchées sur la catégorisation des profils de personnalité [Eysenck, 1991], d'autres sur des façons rapides de déterminer le profil d'un sujet au moyen de tests aussi

courts que possibles [John et al., 1991, Costa and McCrae, 1992, Rammstedt and John, 2007]. L'usage de cinq dimensions majeures de la personnalité (les Big Five [Goldberg, 1990]) s'est progressivement imposé, en particulier pour les études en ligne. Les cinq dimensions les plus importantes sont (en respectant le jargon qui tente de s'installer) : l'Ouverture d'esprit, la Conscienciosité, l'Extraversion, l'Agréabilité, le Névrosisme.

L'image, et en particulier l'image postée sur son site par un internaute, est une façon objective d'exposer sa personnalité. Elle peut être analysée, comme tout échange lors d'interactions entre humains, à l'aide de modèles statistiques éprouvés [Brunswick, 1956]. Ainsi les auteurs de [Cristani et al., 2013] ont associé des photos postées et le profil psychologique du photographe. Ils ont tout d'abord constitué, à l'aide d'un protocole supervisé assez complexe, la base de données, PsychoFlickr, associant les images postées par 300 usagers « professionnels » du réseau Flickr, ainsi que les profils psychologiques « auteur » et le profil « récepteur » pour chacun de ces usagers. Ces profils sont constitués des Cinq Majeures. Le profil émetteur représente ce que l'auteur pense présenter, tandis que le profil récepteur est ce que perçoit un visiteur du site. Puis, à partir de primitives extraites des photos du site, on procède à l'apprentissage d'un classifieur qui relie propriétés des images et profil psychologique du photographe. Ce classifieur permet alors de dresser le profil d'un internaute inconnu à partir des photos qu'il a postées.

Ces travaux, menés à l'aide de primitives artisanales et de régression, ont été poursuivis et étendus dans [Segalin et al., 2016, Segalin et al., 2017] avec des réseaux de neurones, améliorant en particulier la qualité des prédictions.

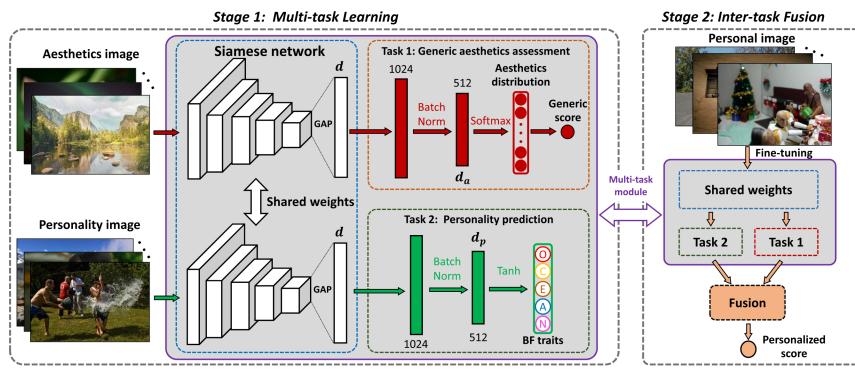
### 3.3.2. Cinq Majeures et esthétique

Dans [Li et al., 2020], on trouve comment mettre à profit un profil individuel pour obtenir un jugement esthétique subjectif. Lors d'une phase d'apprentissage, deux réseaux sont mis en parallèle, l'un chargé de donner une note d'esthétique, « générique » selon les auteurs, (« objective » selon nos termes), l'autre de mesurer le profil d'un utilisateur à partir de sa représentation sur les Cinq Majeures. Ces réseaux sont identiques dans leurs premières couches (réseaux siamois) mais se distinguent dans les couches de sortie. Le premier est entraîné à l'aide de la base universelle AVA pour donner une note esthétique à toute image en entrée. Le second est entraîné à l'aide de PsychoFlickr. Il apprend à noter le profil d'un photographe d'après les images qu'il a prises, mais aussi, aidé du premier réseau, à déterminer comment ce profil intervient dans le jugement esthétique (figure 3).

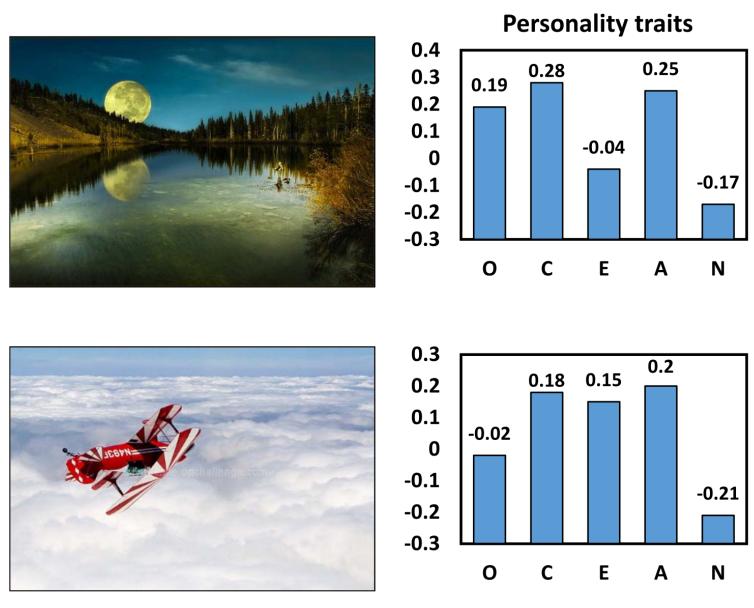
Lors de l'étape d'évaluation subjective d'une nouvelle photo, une note d'esthétique « générique » est mesurée dans la première branche, tandis que le profil de l'internaute à partir de ses photos est déterminé dans la seconde branche. Les sorties des deux réseaux sont ensuite combinées : la note générique est modifiée linéairement par cinq termes correctifs introduisant la contribution subjective propre au profil de l'internaute. Chaque terme exprime la contribution d'une majeure comme produit d'une sensibilité à cette majeure de l'observateur par un poids de la majeure dans l'image (figure 4).

## 3.4. Apprendre les goûts de l'utilisateur par des tests

Les travaux présentés jusque-là exploitent des photos postées par l'utilisateur reflétant ses goûts esthétiques. Cette situation est cependant assez rare car peu d'utilisateurs dont on veut prédire le jugement



**Figure 3.:** Architecture du système développé par Li *et al.*. Il utilise deux branches d'un réseau siamois : sur celle du haut est mesurée une beauté générique identique pour tous les observateurs ; sur celle du bas est déterminé le profil en termes de Cinq Majeures (représentées par les 5 notes OCEAN), puis les deux évaluations sont fusionnées dans une étape qui fournit un score unique personnalisé [Li et al., 2020].



**Figure 4.:** Poids des Majeures dans des images : deux images projetées dans l'espace des Cinq Majeures. Les 5 notes sont comprises entre -4 et +4 : O = Ouverture d'esprit, C = Conscienciosité, E = Extraversion, A = Agréabilité, N = Névrosisme, d'après [Li et al., 2020].

disposent de cette source d'information. Le recours à une étape préalable de tests (comme on l'a fait par exemple pour déterminer les Cinq Majeures) est parfois proposé. Ces tests consistent très généralement à juger un petit nombre d'images. Comme ces tests sont fastidieux, on cherche à les réduire. La littérature retient aujourd'hui deux niveaux de tests : l'un rapide se fait à partir de 10 images, l'autre, plus lourd, avec 100 images. Les résultats ne concordent pas sur l'intérêt de mettre en œuvre un test lourd. Si l'intuition semblerait favoriser des tests longs, certains auteurs ont constaté que leurs performances se dégradent vite, probablement en raison de la chute d'attention de l'observateur.

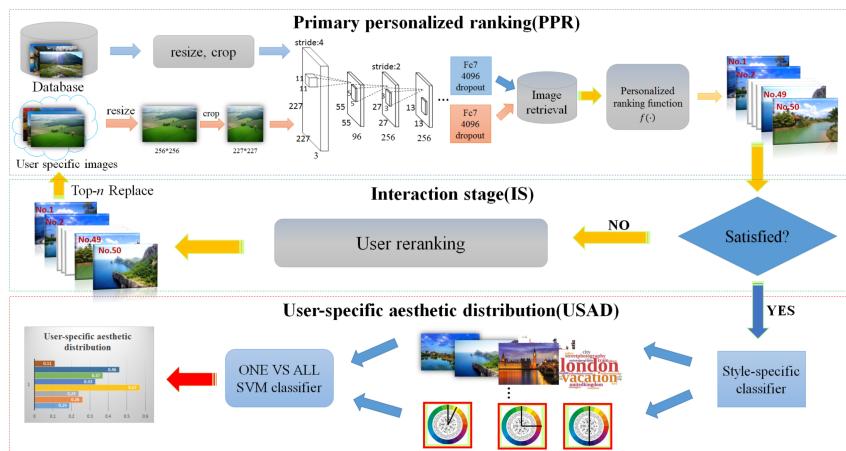
L'idée de nombreux travaux qui procèdent de cette façon est de spécialiser un réseau de beauté « générique » en réalisant un réglage fin des dernières couches neuronales à l'aide des résultats des tests. On obtient ainsi, à faible coût, un réseau de jugement esthétique, spécialisé dans le goût de l'utilisateur qui a répondu au test.

Dans [Zhu et al., 2020], on tire profit d'une technique de méta-apprentissage par optimisation. Pour cela il nous faut disposer d'une base de données annotée, pour laquelle on a gardé trace des personnes qui l'ont annotée (c'est le cas de deux bases Flickr-AES et AADB). Un réseau est entraîné pour chacun des observateurs et l'on détermine ainsi d'une part les règles partagées par tous les observateurs pour juger les images et d'autre part les paramètres qui permettent de raffiner ce modèle pour chaque observateur. Lorsqu'un nouvel usager inconnu se présente, on détermine à l'aide d'un jeu de tests de quel observateur connu il s'approche et on le fait bénéficier du réseau spécialisé par cet observateur pour juger de nouvelles photos. Des tests d'évaluation de la méthode montrent que l'on atteint 70 % de classification correcte à 2 classes lorsque l'apprentissage se fait avec une série de 100 images de test, mais le taux de bonnes réponses chute à 56 % avec 10 images seulement. En comparaison, des travaux comparables mais utilisant une factorisation matricielle et l'analyse latente de Dirichlet à la place du réseau de neurones, plafonnent à 52% dans ces deux situations (annotation de 10 ou de 100 images) [O'Donovan et al., 2014].

Les travaux présentés dans [Park et al., 2017] proposent à l'observateur de classer les images de test par préférence (et non un choix binaire belle/non belle). Pour accélérer cette étape, on lui offre des choix binaires : choisir la plus belle au sein de paires d'images (une dizaine de paires), puis un algorithme ordonne l'ensemble des images de test de façon à respecter ces choix. Ce tri accompli sur le petit nombre d'images de test est propagé sur la totalité de la base d'apprentissage selon un principe de plus proche voisin. Enfin, pour évaluer une nouvelle image, un compromis est fait entre une note objective obtenue par régression à la sortie d'un classifieur générique et un classement par ordonnancement selon l'ordre personnel de l'observateur. Ce compromis est géré par une fonction objective unique optimisant conjointement la partie subjective et la partie objective du système.

Dans [Lv et al., 2018], on procède à l'apprentissage à travers un petit nombre d'images proposées par l'utilisateur. Un réseau permet alors d'extraire de la base de données d'images celles qui s'approchent le plus des choix de l'utilisateur. Par l'intermédiaire d'une boucle de réaction interactive (apprentissage par renforcement) l'utilisateur peut améliorer itérativement ce choix jusqu'à la constitution d'une base de données spécialisée rendant compte de ses goûts esthétiques. L'enchaînement de 3 à 5 étapes de corrections semble suffisante pour un apprentissage efficace ; la phase d'interaction est donc relativement courte. Cette base de données est alors disponible pour créer un réseau spécialisé, adapté à l'esthétique de l'utilisateur (figure 5). Il semble que les résultats (mesurés par la corrélation entre l'ordonnancement d'une série de photos par le réseau et par l'observateur) soient bons (supérieurs à 0,8) et qu'un

apprentissage sur un tout petit nombre d'images soit préférable à un long apprentissage. Ce n'est pas le cas des travaux présentés dans [Ren et al., 2017], assez semblables dans leur objectif, mais utilisant une fonction discriminante (ici un *support vector regressor* utilisant des fonctions de base radiales) au lieu d'un réseau. Contrairement au cas précédent, il apparaît que les performances croissent régulièrement et sensiblement avec le nombre d'images utilisées pour qualifier l'utilisateur, au moins jusqu'à 100. Pour prendre en compte cette propriété, les auteurs développent une approche incrémentale qui s'améliore au fur et à mesure des réponses de l'utilisateur.



**Figure 5.:** Le système développé par Lv *et al.* utilise une boucle d'apprentissage qui peut être parcourue itérativement. Le système fournit à l'utilisateur des images de la base de données annotées proches d'un petit jeu d'images appréciées de l'utilisateur. L'utilisateur confirme la pertinence de ces images ou les rejette, de façon que l'ensemble finalement retenu corresponde bien à ses goûts. Ce sont ces images qui vont servir à entraîner le système [Lv et al., 2018].

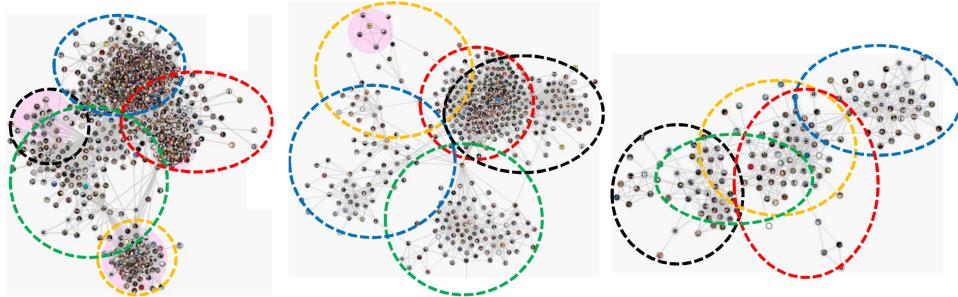
Certains travaux combinent des tests sur les goûts de l'observateur (apprentissage par un petit jeu de tests où les images sont ordonnées) et une analyse de son profil social à partir de ses connexions sur le réseau [Deng et al., 2017a]. L'image à évaluer reçoit deux notes, l'une reflétant les goûts personnels (déterminés par ses réponses aux tests et sa participation aux réseaux sociaux), l'autre pour sa conformité au goût générique. Ces deux évaluations sont combinées de façon assez complexe.

### 3.5. Multiplier les expertises concurrentes

Une 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. C'est une voie qui est également en partie explorée dans les travaux de [Zhu et al., 2020] vus plus haut.

Cette démarche a été abordée par S. Kong *et al.* [Kong et al., 2016] qui confient l'évaluation de leur base AADB à un petit nombre de vrais experts d'esthétique dont ils gardent trace dans l'expertise. Chaque expert exprime une sensibilité différente. On peut ainsi être amené à n'utiliser que les notes d'un seul expert pour évaluer une image (si un internaute se sent en plus grande concordance avec l'un d'eux) ou au contraire plusieurs d'entre eux.

Plus ambitieux sont les travaux rapportés dans [Hong et al., 2016] qui essaient de déterminer des communautés de goûts parmi la population des abonnés à Flickr. Pour cela sont utilisés, pour chaque image, deux types de descripteurs : des descripteurs sémantiques d'une part (issus de légendes fournies avec chaque image), des descripteurs de bas niveau d'autre part (couleur, contraste, contours, textures). Dans



**Figure 6.:** 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.

l’espace de ces descripteurs, chaque image est traitée comme un mot par un algorithme d’allocation latente de Dirichlet (LDA). Les mots sont regroupés en « phrases » (une phrase représente un photographe) qui fournissent donc des sujets (*topics*) latents définissant des communautés (un *topic* représente une communauté). Les communautés partagent des sujets voisins dans l’espace des descripteurs. Si la démarche est très intéressante, l’outillage mathématique mis en jeu est considérable et les premiers résultats ne sont pas totalement convaincants. Le poids respectifs des divers descripteurs est délicat à trouver et il est difficile de séparer les rôles des goûts esthétiques et de l’intérêt pour un thème photographique (voir la figure 6).

On aimerait pouvoir regrouper ces deux dernières approches pour construire des communautés incontestables autour d’experts pertinemment choisis pour exprimer des tendances esthétiques bien établies mais cette démarche de classification des esthétiques dans le domaine de la photographie ne semble pas avoir à ce jour été entreprise.

### 3.6. Conclusion

La démarche adoptée par la communauté d’esthétique computationnelle pour échapper à un cadre purement objectiviste résulte des critiques qu’ont justement rencontrées, malgré des résultats évidents, les travaux menés dans ce cadre objectiviste à l’aide de méthodes à base de réseaux de neurones profonds entraînés sur de vastes bases établies une fois pour toutes.

La multiplicité et la diversité des approches engagées pour apporter une note de subjectivité dans le jugement de goût reflètent la complexité du projet. L’étendue et la variété des déterminants du jugement esthétique, la méconnaissance de leurs rôles et de leurs influences réciproques rendent les démarches extrêmement incertaines et expliquent ces essais tâtonnants, généralement sensés mais finalement bien peu convaincants. L’absence de référentiel unanimement accepté, aussi bien dans le champ de l’esthétique philosophique que dans ceux de la neuro-biologie ou de la psychologie expérimentale, laisse l’informaticien désorienté face aux voies à suivre. Les démarches engagées sont parfois habiles, elles sont toujours techniquement très élaborées ; elles convoquent des schémas d’intelligence artificielle souvent très raffinés, mais elles reflètent aussi un parti pris pour des pistes étroites qui contribuent probablement au choix, mais n’expliquent que très imparfairement le jugement de goût. Ainsi, elles peinent à convaincre. Leurs

prémisses esthétiques sont trop fréquemment indigentes. Leurs références s'alignent souvent sur des pratiques sociologiques dont la validité dans le champ du goût n'est pas établie car les autres déterminants du choix n'ont pas été suffisamment écartés. Les critères de succès qu'elles affichent nous interpellent autant par leur modeste signification que par l'étroitesse de leur validité. On n'a certes pas épousé toutes les ressources que la technique ne cesse de produire dans le champ des algorithmes mais on attend surtout les résultats que la neuro-biologie pourra dégager avec les outils modernes de l'exploration fonctionnelle en réduisant les échelles temporelles des mesures, en accroissant les sensibilités, en étendant la portée des suivis des signaux cérébraux de façon à mieux comprendre la chronologie de la décision humaine.

## Bibliographie

- 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, M. (1969). *Einführung in die informationstheoretische Ästhetik. Grundlegung und Anwendung in der Texttheorie*. Rowoldt Taschenbuch Verlag.
- Berridge, K. and Kringelbach, M. (2013). Neuroscience of affect : brain mechanisms of pleasure and displeasure. *Current Opinion in Neurobiology*, 23(3) :294–303.
- Birkhoff, G. (1933). *Aesthetic Measure*. Harvard University Press, Cambridge (USA).
- 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.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. University of California Press.
- Chatterjee, A. (2004). Prospects for a cognitive neuroscience of visual aesthetics. *Bulletin of psychology and the arts*, 4(2) :55–60.
- Chatterjee, A. and Vartanian, O. (2016). Neuroscience of aesthetics. *Annals of the New-York Academy of Sciences*, 1369 :172–194.
- Cochoy, F. (2011). *De la curiosité, l'art et la séduction marchande*. Armand Colin, Paris.
- Costa, P. and McCrae, R. (1992). *NEO-PI-R Professional manual*. Odessa : Psychological Assessment Resources.
- Cristani, M., Vinciarelli, A., Segalin, C., and Perina, A. (2013). Unveiling the multimedia unconscious : Implicit cognitive processes and multimedia content analysis. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 213–222.
- Cui, P., Liu, S., Zhu, W., Luan, H., Chua, T., and Yang, S. (2014). Social-sensed image search. *ACM Transactions on Information Systems (TOIS)*, 32(2)(2) :1–23.
- Datta, R. and Wang, J. (2010). ACQUINE : Aesthetic quality inference engine - real-time automatic rating of photo-aesthetics. In ACM, editor, *MIR'10*.
- 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, 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, X., Cui, C., Fang, H., Nie, X., and Yin, Y. (2017a). Personalized image aesthetics assessment. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 2043–2046.
- Deng, Y., Loy, C. C., and Tang, X. (2017b). Image aesthetic assessment : An experimental survey. *IEEE Signal Processing Magazine*, 34(4) :80–106.
- 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, C. and Vittorio, G. (2009). Neuroaesthetics : a review. *Current Opinion in Neurobiology*, 19 :682–687.
- 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.

- Dretske, F. (1969). *Seeing and Knowing*.
- 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.
- Elridge, A. (2015). Photography and sociology, an exercise in serendipity.
- Eysenck, H. (1991). Dimensions of personality : 16, 5 or 3 ?—criteria for a taxonomic paradigm. *Personality and individual differences*, 12(8)(8) :773–790.
- Eysenck, H. J. (1941). The empirical determination of an aesthetic formula. *Psychological Review*, 48(1) :83–92.
- Fechner, G. (1871). *Zur Experimentalen Aesthetik*. Hirtzel, S., Leipzig.
- Goldberg, L. (1990). An alternative “description of personality” : The big-five factor structure. *Journal of Personality and Social Psychology*, 59 :1216–1229.
- 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.
- 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.
- Hsu, L. (2009). *Le visible et l'expression. Etude sur la relation intersubjective entre perception visuelle, sentiment esthétique et forme picturale*. PhD thesis, Ecole des Hautes Etudes en Sciences Sociales, Paris. HAL id : tel00401739.
- 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.
- Ishizu, T. and Zeki, S. (2011). Toward a brain-based theory of beauty. *PLOS One*.
- Jacobsen, T. (2010). Beauty and the brain : culture, history and individual differences in aesthetic appreciation. *Journal of anatomy*, 216(2) :184–191.
- John, O., Donahue, E., and Kentle, R. (1991). Big five inventory. *Journal of Personality and Social Psychology*.
- Kairanbay, M., See, J., and Wong, L. K. (2019). Beauty is in the eye of the beholder : Demographically oriented analysis of aesthetics in photographs. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 15(2s) :1–21.
- Koelsch, S., Jacobs, A. M., Menninghaus, W., Liebal, K., Klann-Delius, G., von Scheve, C., and Gebauer, G. (2015). The quartet theory of human emotions : an integrative and neurofunctional model. *Physics of life reviews*, 13 :1–27.
- Konecni, V. (1979). Determinants of aesthetic preference and effects of exposure to aesthetic stimuli : social, emotional and cognitive factors. *Prog Exp Pers Res.*, 9 :149–197.
- 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.
- Kosinski, M., Bachrach, Y., Kohli, P., Stillwell, D., and Graepel, T. (2014). Manifestations of user personality in website choice and behaviour on online social networks. *Machine learning*, 95(3)(3) :357–380.
- Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. In *Proceedings of the national academy of sciences*, number 110(15), pages 5802–5805.
- 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.
- Li, L., Zhu, H., Zhao, S., Ding, G., and Lin, W. (2020). Personality-assisted multi-task learning for generic and personalized image aesthetics assessment. *IEEE Transactions on Image Processing*, (29) :3898–3910.
- 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.
- Lovato, P., Perina, A., Cheng, D. S., Segalin, C., Sebe, N., and Cristani, M. (2013). We like it ! mapping image preferences on the counting grid. In *2013 IEEE International Conference on Image Processing*, pages 2892–2896.
- 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.
- Lv, P., Wang, M., Xu, Y., Peng, Z., Sun, J., Su, S., and Xu, M. (2018). Usar : an interactive user-specific aesthetic ranking framework for images. In *Proceedings of the 26th ACM international conference on Multimedia*, pages 1328–1336.
- 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.

- Maître, H. (2021). *Qu'est-ce qu'une belle photo ? Essai sur l'esthétique en photographie numérique*. ISTE - Arts et Sciences - (à paraître, printemps 2021).
- Marchesotti, L., Murray, N., and Perronnin, F. (2015). Discovering beautiful attributes for aesthetic image analysis. *Int. J. of Computer Vision*, 113 :248–266.
- 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.
- Marr, D. (1982). *Vision : A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman and Company, 1982.
- 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.
- 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.
- 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.
- O'Donovan, P., Agarwala, A., and Hertzmann, A. (2014). Collaborative filtering of color aesthetics. In *Proceedings of the Workshop on Computational Aesthetics*, pages 33–40.
- 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.
- Pasquier, D., Beaudoin, V., and Legon, T. (2014). "Moi je lui donne 5/5", *Paradoxes de la critique amateur en ligne*. Presses des Mines (Paris).
- Petitot, J. (2008). *Neurogéométrie de la vision*. Les Editions de l'Ecole Polytechnique.
- Rammstedt, B. and John, O. (2007). Measuring personality in one minute or less : A 10-item short version of the Big Five inventory in English and German. *Journal of research in Personality*, 41(1)(41(1)) :203–212.
- Ray, L. (2020). Social theory, photography and the visual aesthetic of cultural modernity. *Cultural Sociology*, (1749975520910589).
- Reagle, J. (2013). Revenge rating and tweak critique at photo.net. personal website : <http://reagle.org/joseph/2013/photo.net.html>.
- Redi, M., Liu, F., and O'Hare, N. (2017). Bridging the aesthetics gap : The wild beauty of web imaginary. In ACM, editor, *ICMR'17 conference*, pages 242–250, Bucarest.
- Redies, C. (2015). Combining universal beauty and cultural context in a unifying model of visual aesthetic experience. *Frontiers in human neuroscience*, 9 :218.
- Ren, J., Shen, X., Lin, Z., Mech, R., and Foran, D. J. (2017). Personalized image aesthetics. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 638–647.
- Rigau, J., Feixas, M., and Sbert, M. (2008). Information aesthetic measures. *IEEE Computer Graphics and Applications*, 2 :24–34.
- Schifanella, R., Redi, M., and Aiello, L. (2015). An image is worth more than thousand favourites : surfacing the hidden beauty of Flickr pictures. ArXiv : 1505.03358v2.
- Segalin, C., Cheng, D., and Cristani, M. (2017). Social profiling through image understanding : Personality inference using convolutional neural networks. *Computer Vision and Image Understanding*, 156 :34–50.
- Segalin, C., Cristani, M., Perina, A., and Vinciarelli, A. (2016). The pictures we like are our image : continuous mapping of favorite pictures into self-assessed and at- tributed personality traits. In *IEEE Trans. Affect. Comput.*, volume 99, pages 1–1.
- Sheng, K., Dong, W., Ma, C., Mei, X., Huang, F., and Hu, B. G. (2018). Attention-based multi-patch aggregation for image aesthetic assessment. In *Proceedings of the 26th ACM international conference on Multimedia*, pages 879–886.
- Talebi, H. and Milanfar, P. (2017). NIMA : Neural image assessment. arXiv : 1709.0541v1.
- Wertheimer, M. (1938). *Laws of organization in perceptual forms*. Ellis, W.D., A source book of Gestalt psychology, London.
- Yang, L., Hsieh, C. K., and Estrin, D. (2015). Beyond classification : Latent user interests profiling from visual contents analysis. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 1410–1416.

You, Q., Bhatia, S., and Luo, J. (2016). A picture tells a thousand words—about you ! user interest profiling from user generated visual content. *Signal Processing*, 124 :45–53.

Zhu, H., Li, L., Wu, J., Zhao, S.and Ding, G., and Shi, G. (2020). Personalized image aesthetics assessment via meta-learning with bilevel gradient optimization. *IEEE Transactions on Cybernetics*.