Assessment of Beauty with subjectivity
A challenge for computational aesthetics

Henri Maître
LTCI, Télécom Paris, Institut Polytechnique de Paris, France,
henri.maitre@telecom-paris.fr

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 let 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.

Beauty is often regarded as a minor and superfluous attribute of a serious world. The interest that one carries there is frequently held for futile and worldly. The judgment of Pâris reminds us that it could however have heavy consequences outside the field of aesthetics. It is not surprising, then, to see the extent of the comments that have been made throughout the centuries throughout the centuries, to the designation of Aphrodite, well beyond the stake of the profit of an apple. "Judgment of taste" would say Kant. Even so ! It was not the fashion in the time of the Greeks. Beauty was a matter of harmony and symmetry (not our geometrical symmetry, identical reflection, but a proportion granted between the whole and the parts). Judgment had nothing to do with it ; beauty was a matter of numbers and was imposed on everyone. No need for a shepherd to decide. Even on this point, since this shepherd, (by chance ?) was himself very beautiful and reputed to be a lover of a gender that was still weak ! It has been said that there were other values than beauty in the arguments of the debate : promises of military victories, hopes of magnificent reign, the exceptional love of the most envied woman. These arguments circumvent the obstacle, as does the argument that it was necessary to come to the war and that the shortest way could only be the best. One knows the choice of Pâris, but how did he decide ? Did he measure for each goddess the agreement of its forms with the canons of Polyclète ? Did he, as Descartes or Hegel think, use reflection to conclude rationally on the arguments of his senses ? Has he rather, as Kant suggests, let himself be invaded by the intuition of a common sense impartially experienced ? Or invaded by the Dionysian enjoyment of the three goddesses as Nietszche would have done ?

Twenty-five centuries of aesthetics debate at length between Objectivists ("the object is beautiful by its own virtues") and Subjectivists ("the object is beautiful as soon as I see it beautiful"). The experimental psychologists [Fechner, 1871, Leder et al., 2004, Hurlberg and Ling, 2012], sociologists [Elridge, 2015, Ray, 2020], then neuro-biologists [Di Dio and Vittorio, 2009, Brown et al., 2011, Ishizu and Zeki, 2011] have brought their voices to this debate and recently computer scientists have joined them, who have made the aesthetic judgment a new challenge for their algorithms, which are already highly solicited by the game of chess, stock market prediction or medical diagnosis.

1 What can computer science do for beauty ?

The first proposals to use the hard sciences in the service of aesthetics date back to Charles Henry, but his *Introduction to a Scientific Aesthetic* of 1885 has hardly been emulated. The same cannot be said of the work of the mathematician George Birkhoff [Birkhoff, 1933], who gave birth to a long line of research aimed at formalizing an "equation of beauty". The "equation of beauty" works benefited from various progresses of our knowledge during the XX^th century : Gestalttheorie, mathematical morphology pattern recognition, image processing [Eysenck, 1941, Moles, 1957, Bense, 1969, Rigau et al., 2008].

We are going to focus here on the measurement of beauty in photographs for three main reasons. On the one hand, photographs have taken an exceptional place in our society because of the progress and proliferation of acquisition systems as well as the popularity of social networks. On the other hand, photography has gained the recognition of an art in its own right with all the required attributes : museums, exhibitions, experts, market and quotation, magazines . . . Finally, a strong social and societal demand calls for the implementation of rapid and efficient techniques of evaluation of the beauty of photographs, as much for the individual needs of management of its personal archives as for economic sectors that rely heavily on images : publishing, advertising, travel, real estate and fashion agencies . . .

We are going to look particularly at these approaches, born at the end of the last century, which, instead of setting up an explicit formula of beauty as Birkhoff and his successors did, rely on machine learning and artificial intelligence techniques and make use of image databases annotated with aesthetic judgments. Two series of works followed one another, exploring this track. First, from 1994 to 2015 approximately, pioneer works have implemented the detection of relevant primitives (often called *handmade features*) and simple and robust classification techniques. Most of these works used explicit knowledge of the rules of photography and the recommendations made by specialists (scene composition, focus, color harmony) but less expert approaches have taken advantage of primitives image processing (textures, contours) or computer vision (SIFT, SURF, etc.) with equally good performances. These primitive-based techniques have given way, around 2015, to deep neural network-based techniques, which do not consider the extraction of primitives and ignore the knowledge that led to their selection.

The images to be tested are placed at the input of the network which has been trained to evaluate the aesthetics of the image and the result is displayed at the output : *a beautiful image* or *a banal image* in the frequent case of a binary classification, or evaluation on a scale of 1 to 10 in the case of a continuous judgment. The performances of deep neural network techniques have largely surpassed those of hand-crafted primitives and today these techniques alone survive to propose the user to eliminate poor images from large collections or to select those that emerge.

Returning to the debate which is the subject of the first lines of this text, it must be

criteria	BEAUTY	AADB	Redi's base	AVA	AVA-2 CUHK-DB	AVA-PD	Uni Tübingen	Psycho Flickr	Flickr AES
size (×1000)	15	10	100	250	50	120	380	60	40
aesthetical quality	weak	weak	weak	high	high	high	weak	high	weak
aesthetic mark	3	5	4	10	10 / 2	10	10	5	5
semantic classes	4	no	yes	44	44	44	no	yes	yes
style labels	no	11	no	14	14	14	no	no	no
annotation	no	yes	yes	yes	yes	yes	no	no	yes
origine	Flickr	Flickr	web	DP Chal.	AVA	AVA	Flickr	Flickr	Flickr

TABLE 1 – *Some frequently used databases for the aesthetic study of images with some of their properties. The* size *is expressed in thousands of images. The* aesthetic quality *is considered as "high" if the images come from professionals or enlightened amateurs, as "low" if they come from social networks. The* score *represents the excursion of the rating. The* annotations *are composed of the literal comments that sometimes accompany the reviews.*

noted that the "objectivist" approach imposes itself in these two approaches. The diagnosis of the algorithm, whatever it is, is universal : the image is beautiful or not, independently of the observer, by its only internal properties : its composition, its textures, its colors, the assembly of its lines, etc. The observer has no role in this decision and does not know which criteria have prevailed. The algorithm adopts a Platonic aesthetic : beauty belongs to the image not to the one who experiences it.

Let us explain the methodology followed by the authors of this work. First of all, a collection of images is gathered, each one - and this is important - endowed with an evaluation supposed to express its aesthetic quality. Sites specialized in photography have been used extensively for this purpose (Flickr, Photo.net, Instagram, DpChallenge). Depending on the audience they address (general public, enlightened amateurs, professionals), they offer photos of any quality and nature : from novice to expert, reportage, fashion, art, sports, etc. They are frequently associated with judgments, either in the form of votes during thematic competitions, or expert opinions of professionals, or simple comments posted by Internet users. Thus, some databases have been created, gathering at least tens of thousands of photos to which are attached their rating and sometimes comments posted by Internet users. Their names are AVA, AADB or BEAUTY (see Table 1.1) [Murray et al., 2012, Schifanella et al., 2015, Redi et al., 2017].

These databases are used to train a neuromimetic network, usually one of those proven in pattern recognition competitions (ResNet, VGG, Mobile-Net, GoogLeNet, Inception) by successively presenting each photo at the input of the network and imposing its associated score at the output. After having frozen the network parameters, we can proceed to the evaluation of an unknown photo by placing it at the input of the network. We then read its evaluation at the output.

1.1 Databases and expertise

They are briefly presented in Table 1.1. and more extensively detailed in [Maitre, 2021].

It is on the AVA database that the algorithms are usually tested. AVA gathers more than 250 000 photos, each accompanied by scores between 0 and 10 assigned by Internet users and from the DpChallenge site, a site specifically dedicated to photographers. Only the photos having received at least 200 marks are retained in AVA, and to each one is the average of these scores generally given as the final score.

From the AVA database, the AVA-2 database and the CUHK-DB database are deducted by retaining only the best and the worst images, as well as the base AVA-PD (AVA-Photographer Demographic) by retaining only the images of some photographers for whom we know some information (gender, age, country).

BEAUTY and the Redi database were evaluated by crowdsourcing, i.e. by a paid consultation of Internet users. AADB was evaluated by 5 experts who assigned the aesthetic score when their opinions reasonably converged. Flickr-AES selected photos from Flickr that were annotated on aesthetic criteria, by crowdsourcing too. Psycho-Flickr selected images from Flickr produced by photographers (300 "professional" photographers from Flickr) for whom the psychological profiles (sender and receiver) were drawn up using the Big Five (see below).

1.2 Results

The strategies differ in terms of the results sought. Some algorithms aim to reproduce a continuous evaluation between 0 and 10, some seek to find the distribution of evaluations given by Internet users, others - and these are the most numerous - propose a binary choice : "beautiful" or "not beautiful" image and there again various choices are made. One often chooses to separate "beautiful" or "not-beautiful" according to the note attributed, with the help of a parameter δ compared to the average value M of the whole of the 250 000 photos ; if the note is higher than $M + \delta$ the image is beautiful, if it is lower than $M - \delta$, it is not beautiful. When $\delta = 0$, several teams obtain more than 80 % of correct results on the AVA base (separating of course the pictures used for learning from those used for verification) and these performances increase even more if we use, in addition to the image, some semantic information on the theme (portrait, landscape, daily life, etc.) in order to specialize the processing a bit [Sheng et al., 2018, Ma et al., 2017, Talebi and Milanfar, 2017, Deng et al., 2017].

These results are undeniably interesting since they allow us to quickly sort out, among thousands of photographs, those which present good aesthetic qualities, without leaving too many aside, but they are however exposed to various criticisms. They can concern the relevance of the learning bases which reflect a very current photographic genre responding to precise social codes (abundance of photos of animals, portraits, ...) and using a form made very accessible by the technology which often calls upon games of colors, high resolution, inlays or special effects. They can concern the opinions given which suffer, like most opinions collected on the internet, from bias and dispersion [Reagle, 2013, Pasquier et al., 2014, Cochoy, 2011]. But a deeper reproach is attached to the assumption that any photo is judged identically, beautiful or not, by all observers. This objectivist assumption that the annotated database of Internet users' judgments reflects a characteristic property of each photo, a property that leads to a unique judgment of taste, shared by all, which makes it possible to parameterize a network capable of reproducing this "universal" taste judgment, is the object of numerous objections and has given rise to a number of works that have to escape from it. It is these attempts that we want to describe here which aim at introducing subjectivity in this automation of the aesthetic.

2 What is the subjectivity of taste ?

But first of all we must try to define a little these concepts of objectivity and subjectivity which are subtly merged into each other. Everything starts with the vision. It is well established today that it is from the object that the light rays that reach the retina leave and that they do not come from the eye as was thought for a while. They are the ones who carry the signal of beauty in a Platonic vision. In this thought, the rays are "beautiful" and will be for any observer : the Victory of Samothrace remains beautiful once the doors of the Louvre are closed. Captured by the retina where they receive initial processing, these signals then follow the optic nerves and reach the visual areas where they are shaped according to processes that are now fairly well known : chromatic selections, spatial and temporal frequency selections, filtering, amplifications, multi-scale groupings, etc. This is the "simple vision" of Dretske [Dretske, 1969]. Marr's computational model accounts for it quite well in his *primal sketch*, while the psycho-physiological formulation of Gestalttheorie [Wertheimer, 1938] explains many of its mechanisms. At this point, we lose a little trace of their course and we have only a posteriori information on their action in the various areas of the brain. If there is beauty, we agree today to decide it because there is pleasure and activation of areas in charge of reward (probably in the prefrontal cortex) [Berridge and Kringelbach, 2013], but many other areas are also activated, which are usually involved when memory, decision making, anticipation of action, etc. One would distinguish this pleasure from beauty because unlike many others it does not seem to come from the satisfaction of a need of our being : hunger, thirst, sexual desire...

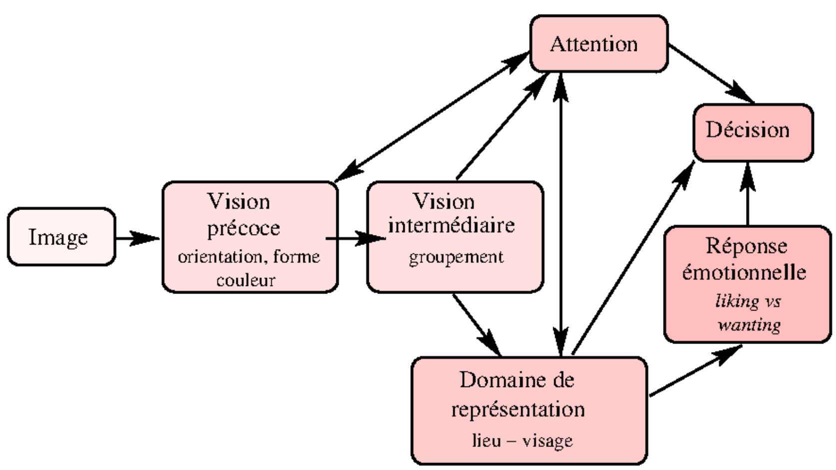

FIGURE 1 – *One of the functional models of aesthetic perception : that of A. Chatterjee. Two different functions concern one the attention, the other the representation. They are in bi-univocal interaction. The attention reacts on the early vision stage by directing the exploration. It is partly influenced by the emotional response that it contributes to elaborate and it acts on our decision in particular through the reflex reactions. The representation is in relation with our models of the world and intervenes also in the decision making (the justification of the aesthetic judgment) and the emotional response (adapted from [Chatterjee, 2004]).*

How did we go from this optical signal to this pleasure attested by our conscience? How do the associations of lines, shapes, colors become for us harmony, elegance, grace, poetry? How does one pass from an objective, measurable, shareable and potentially uni-

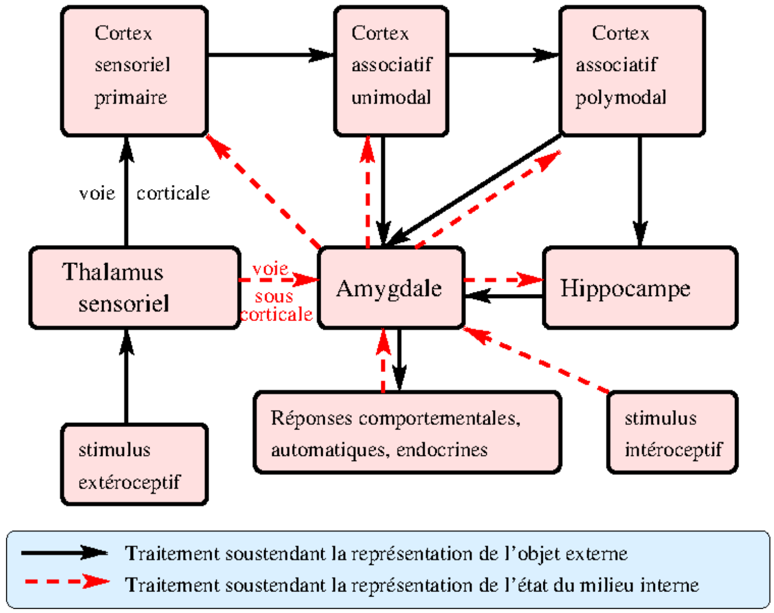

FIGURE 2 – *Another model of aesthetic perception, that of Li-Hsiang Hsu, deduced from A. Damasio's model. It is the exteroceptive stimulus (by the intermediary of the visual ways) which constitutes the entry of the signal which distributes the information towards a representation of the external world on the one hand (the physical perception of the object) and towards a proprioceptive representation (modification of the observer's state) which expresses the emotions felt at the sight of the object. The subcortical pathway, towards the limbic system (hippocampus, amygdala, singular gyrus, hypothalamus), is short, while the other, towards the neo-cortex (towards the occipital areas of the visual areas on the one hand, towards the pre-frontal and orbitofrontal areas on the other hand) is long. This is the cognitive pathway that is associated with a valuation (according to [Hsu, 2009]). Note that the two models presented here are still in competition with others and that none of them has really been validated.*

versal information to an intimate, eminently personal and, let's say it, subjective feeling? In this transition, opinions diverge widely and theories contradict each other. From Kant and Schopenhauer who affirm that this perception of the beautiful is spontaneous, intuitive, impartable, inaccessible to the reasoning and nevertheless common to all, from Descartes and Hegel who make of this conviction the fruit of the reason, of Henry or Séailles who invoke the specificity of the "cerebral-machine" to extract the beautiful as the ear captures the sound, of Bell, Zemach or Arnheim who see, through the assembly of the optical signals the construction of "signifying forms", of "aesthetic predicates", of "qualia" which, under favorable conditions will be decoded precisely by consciences suitably equipped.

The matter is complex and the comprehension of what remains for us an alchemy is still far. The schemes which try to associate our various faculties in the elaboration of this consciousness are numerous [Chatterjee, 2004, Brown et al., 2011, Leder et al., 2004, Redies, 2015, Koelsch et al., 2015, Hsu, 2009] but still widely debated (see figures 1 and 2). In all these schemes, the judgment of taste is highly dependent on elements foreign to the perceived image, whether they are elements of context transmitted by our senses ("atmosphere", setting), elements of context of our internal state at the particular moment (mood) or more permanently (temperament), as well as acquired elements : our education, our culture, our experience, the information we have about the observed object, etc., all these elements that we must call subjective since they are closely attached to an observer, and that some say are only acquired, while others assume are partly innate. These schemes grant very generally the role of judgment of the beautiful to the most evolved layers of our cerebral system. Perhaps only Petitot's neuro-geometric model allows us to explain our capacity of intuition or of "survenance" which seems to be at the heart of many aesthetic experiences by wired structures very close to the visual areas [Petitot, 2008].

3 What approaches to introduce subjectivity into the judgment of taste ?

In the absence of a definitively established subjectivist scheme, it is not easy to propose an algorithm to accomplish these tasks of judgment of taste. In comparison, the objectivist approach which attributes all the merits of beauty to the only object and makes all the observers share an identical judgement seems trivial, especially if we have a magical machine, the deep neuro-mimetic network able to reproduce any deduction provided that it is fed with enough examples.

Let us see, through recent works, how this subjectivization of the judgment of beauty by computer can be approached.

3.1 The recommendation systems

The most widespread techniques today to adapt an offer to a not explicitly formulated request is to resort to the recommendation systems. They are particularly effective for proposing products to a customer that correspond to his/her tastes in the field of cinema, music, online shopping in general and leisure in particular [Deldjoo et al., 2016, Elahi et al., 2017, Deldjoo et al., 2018]. For this purpose we collect past purchases of the of the applicant, as well as any relevant information on his age, his place of life, his social status, and even his tastes in related fields. Based on this information, we classify this client in a group with similar practices. We can then propose offers that this group has unanimously appreciated, or, if we wish to predict his future reactions, to a particular product, we will attribute to him the average opinion of the other members of the group or that of the other customer who is most similar to him within the group.

These techniques are ill-suited to predicting taste judgments in photography unless we have previously been able to build up measures of taste in large populations over a longer period of time on the same images. Let us recall that, if the AVA database collects appreciations, it does not know who attributes them and is not able to construct these groups with similar tastes. In addition, it is difficult to collect the relevant elements of a person's aesthetic tastes without subjecting them to questioning, which would not really be appropriate here. To extract from a few images already appreciated the characteristics allowing a fast categorization of the observer, is otherwise difficult than to do it from a list of films or pieces of music. In the latter case, the category will probably be based on very rich semantic properties : subject of the film, name of the producer, of the actors, type of action, or of the musical piece : genre, performer, orchestration, etc., properties that cannot usually be accessed from simple photos.

In the field of the sale of works of art on line, recommendation services also exist [Dominguez et al., 2017, Benouaret, 2017, Messina et al., 2018, He et al., 2016], which seem to approach our problem. In these systems, it appears however that, if the painting or the photo are indeed taken into account in the recommendation by their style and their composition, the role of this information resulting from the image remains minor compared to the considerable role of more semantic data like the name of the author, his quotation in the market, the price of the work, or the previous purchases of the customer.

3.2 Search a social profile

In the systems described above, the profile sought is above all a social profile : age, gender, profession, family environment, hobbies, etc. [Kosinski et al., 2013, Kosinski et al., 2014] It is very likely that these determinants intervene in the user's tastes, but few studies have today established exploitable links to deduce the aesthetic preferences of the Internet user. Some researchers take into account the practices of the Internet user on social networks to modify the order of the images proposed by the search engines to his requests [Cui et al., 2014]. No criteria of taste is used here, only the practices (memberships, connections, activities) are taken into account.

The work reported in [Kairanbay et al., 2019] shows what information can be drawn from images posted on its site by a photographer. To do so, they exploit a very specific subset of photos from the AVA database, for which social data (age, gender, country of origin) on the photographers who took them are exceptionally available (the database called AVA-PD, *Photographer Demographic*). The authors then show that a profile reduced to these data alone can be partly deduced from the photos he/she posts. They then show that we can, to a certain extent, predict the evaluation he/she will give to a given photo from the photos that this photographer posts. To obtain this result, the authors adopt a classical scheme of taste evaluation with CNN by learning transfer completed by a fine tuning step using the pictures posted by the photographer.

Other authors have sought to exploit images posted on social networks (most often Flickr, Instagram or Pinterest) by an Internet user to identify his or her interests by examining the style and content of these images [Lovato et al., 2013, Yang et al., 2015, You et al., 2016]. These works differ in their approaches, focusing either on the characteristics of the image or on its themes. They lead to fairly convincing results regarding the ability of learning techniques to predict a photographer's interest in a given picture, but these works do not show the role of aesthetic judgment in this interest.

3.3 Defining a psychological profile of the user

3.3.1 Personality and Big Five

The psychological profile of the observer is often considered one of the fundamental determinants of aesthetic taste, ahead of culture, mood, or context [Konecni, 1979, Jacobsen, 2010]. Many classic studies have focused on categorizing personality profiles [Eysenck, 1991], others on quick ways to determine a subject's profile using the shortest possible tests [John et al., 1991, Costa and McCrae, 1992, Rammstedt and John, 2007]. The use of five major personality dimensions (the Big Five [Goldberg, 1990]) has become increasingly popular, especially for online studies. The five most important dimensions are (respecting the jargon that is trying to settle in) : Open-mindedness, Conscientiousness, Extraversion, Agreeableness, Neuroticism.

The image, and in particular the image posted on his site by an Internet user, is an objective way to expose his/her personality. It can be analyzed, like any exchange during interactions between humans, with the help of proven statistical models [Brunswick, 1956]. Thus the authors of [Cristani et al., 2013] have associated posted photos and the psychological profile of the photographer. They first constituted, with the help of a rather complex supervised protocol, the database, PsychoFlickr, associating the images posted by 300 "professional" users of the Flickr network, as well as the psychological profiles of the author and the receiver for each of these users. These profiles are made up of the Big Five. The sender profile represents what the author thinks he/she is presenting, while the receiver profile is what a visitor to the site perceives. Then, from the primitives extracted from the photos of the site, we proceed to the learning of a classifier that links the properties of the images and the psychological profile of the photographer. This classifier then allows us to draw up a profile of an unknown Internet user from the photos he has posted.

This work, carried out with handcrafted primitives and regression, has been continued and extended in [Segalin et al., 2016, Segalin et al., 2017] with neural networks, improving in particular the quality of predictions.

3.3.2 Big Five and aesthetics

In [Li et al., 2020], we find how to leverage an individual profile to obtain a subjective aesthetic judgment. During a learning phase, two networks are put in parallel, one in charge of giving an aesthetics score, "generic" according to the authors, ("objective" according to our terms), the other to measure the profile of a user from his representation on the Big Fives. These networks are identical in their first layers (Siamese networks) but differ in their output layers. The first one is trained with the universal AVA base to give an aesthetic score to any input image. The second one is trained using PsychoFlickr. It learns to rate a photographer's profile based on the images he or she has taken, but also, aided by the first network, to determine how that profile is involved in aesthetic judgment (Figure 3).

In the subjective evaluation step of a new photo, a "generic" aesthetic score is measured in the first branch, while the profile of the Internet user from his photos is determined in the second branch. The outputs of the two networks are then combined : the generic score is linearly modified by five corrective terms introducing the subjective contribution of the user's profile. Each term expresses the contribution of a major as a product of a sensitivity to this major of the observer by a weight of the major in the image (Figure 4).

3.4 Learn the user's taste through testing

The works presented so far exploit photos posted by the user reflecting his aesthetic tastes. However, this situation is quite rare because few users whose judgment we want to predict have this source of information. The use of a preliminary stage of tests (as we did for example to determine the Big Five) is sometimes proposed. These tests very generally consist in judging a small number of images. As these tests are tedious, we try to reduce them. The literature now considers two levels of tests : a quick test using 10 images, and a

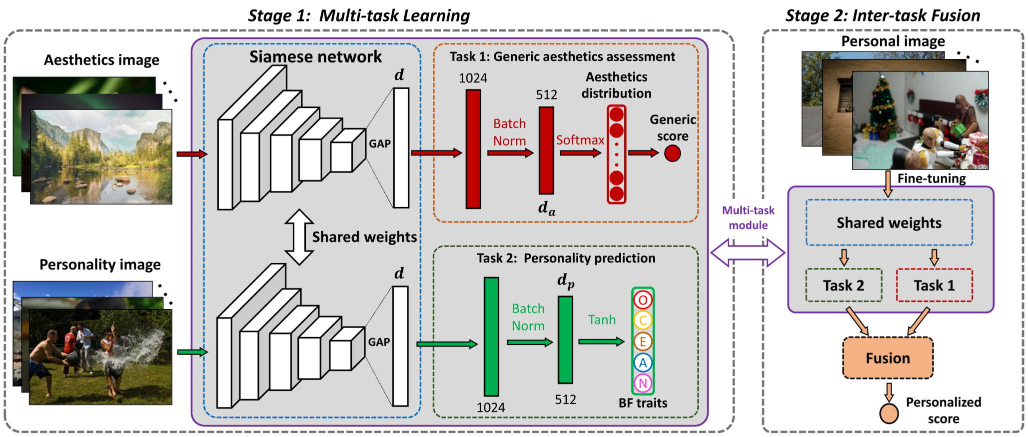

FIGURE 3 – *Architecture of the system developed by Li* et al.. *It uses two branches of a Siamese network : on the upper branch, a generic beauty identical for all observers is measured ; on the lower one, the profile is determined in terms of the Big Five (represented by the 5 OCEAN scores), then the two evaluations are merged in a step that provides a single personalized score [Li et al., 2020].*

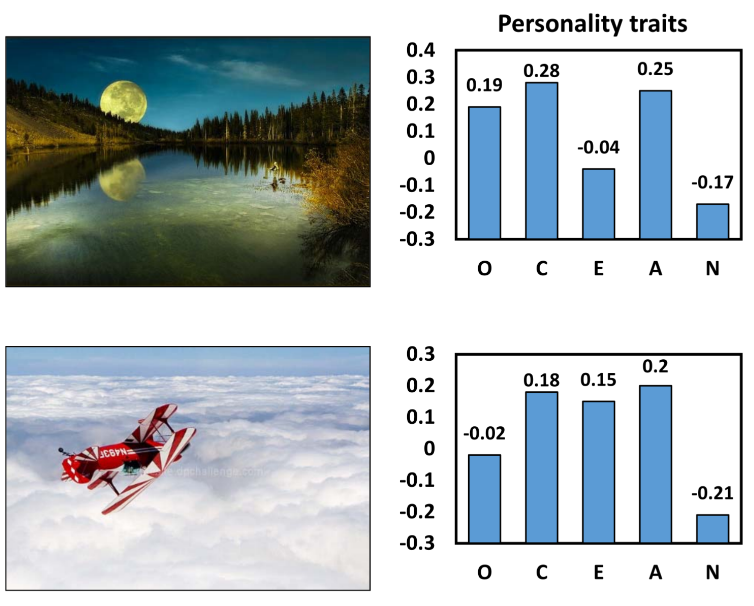

FIGURE 4 – *Weights of the Big Five : two images projected in the Big Five space. The 5 notes ly in the range -4 +4 : O = Open-mindedness, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Neuroticism, [Li et al., 2020].*

heavier test using 100 images. The results do not agree on the interest of implementing a heavy test. Although intuition would seem to favor long tests, some authors have noted that their performance deteriorates quickly, probably because of the observer's loss of attention.

The idea of many works that proceed in this way is to specialize a "generic" beauty network by fine-tuning the last neural layers with the help of test results. We thus obtain, at low cost, an aesthetic judgment network, specialized in the taste of the user who answered the test.

In [Zhu et al., 2020], we take advantage of a meta-learning technique by optimization. To do so, we need to have an annotated database, for which we have kept track of the people who annotated it (this is the case for two databases Flickr-AES and AADB). A network is trained for each of the observers and we determine on the one hand the rules shared by all the observers to judge the images and on the other hand the parameters that allow to refine this model for each observer. When a new unknown user comes along, we determine with the help of a set of tests which known observer he is approaching and we let him benefit from the network specialized by this observer to judge new pictures. Evaluation tests of the method show that 70 % of correct 2-class classification is achieved when learning with a set of 100 test images, but the rate of correct answers drops to 56 % with only 10 images. In comparison, comparable work using matrix factorization and latent Dirichlet analysis instead of the neural network reaches a ceiling of 52 % in both situations (annotation of 10 or 100 images) [O'Donovan et al., 2014].

The work presented in [Park et al., 2017] proposes to the observer to classify the test images by preference (and not a binary choice beautiful/not beautiful). To speed up this step, the observer is offered binary choices : choose the most beautiful one within pairs of images (about ten pairs), then an algorithm orders the set of test images in a way that respects these choices. This sorting, performed on the small number of test images, is propagated on the whole learning base according to a nearest neighbor principle. Finally, to evaluate a new image, a compromise is made between an objective score obtained by regression on the output of a generic classifier and a ranking by ordering according to the observer's personal order. This trade-off is managed by a unique objective function optimizing both the subjective and the objective parts of the system.

In [Lv et al., 2018], we proceed to the learning through a small number of images proposed by the user. A network then allows to extract from the image database the ones that are the closest to the user's choices. Through an interactive feedback loop (reinforcement learning), the user can iteratively improve this choice until a specialized database is built that reflects his aesthetic tastes. The sequence of 3 to 5 correction steps seems to be sufficient for an efficient learning ; the interaction phase is therefore relatively short. This database is then available to create a specialized network, adapted to the aesthetics of the user (figure 5). It seems that the results (measured by the correlation between the ordering of a series of photos by the network and by the observer) are good (above 0.8) and that learning on a very small number of images is preferable to a long one. This is not the case of the work presented in [Ren et al., 2017], quite similar in its objective, but using a discriminant function (here a *support vector regressor* using radial basis functions) instead of a network. Contrary to the previous case, it appears that the performance grows steadily and significantly with the number of images used to qualify the user, at least up to 100. To take this property into account, the authors develop an incremental approach that improves with the user's responses.

Some works combine tests on the observer's tastes (learning by a small set of tests where the images are ordered) and an analysis of his social profile from his connections on the network. The image to be evaluated receives two scores, one reflecting the personal taste (determined by his answers to the tests and his participation in the social networks), the other for its conformity to the generic taste. These two evaluations are combined in a rather complex way.

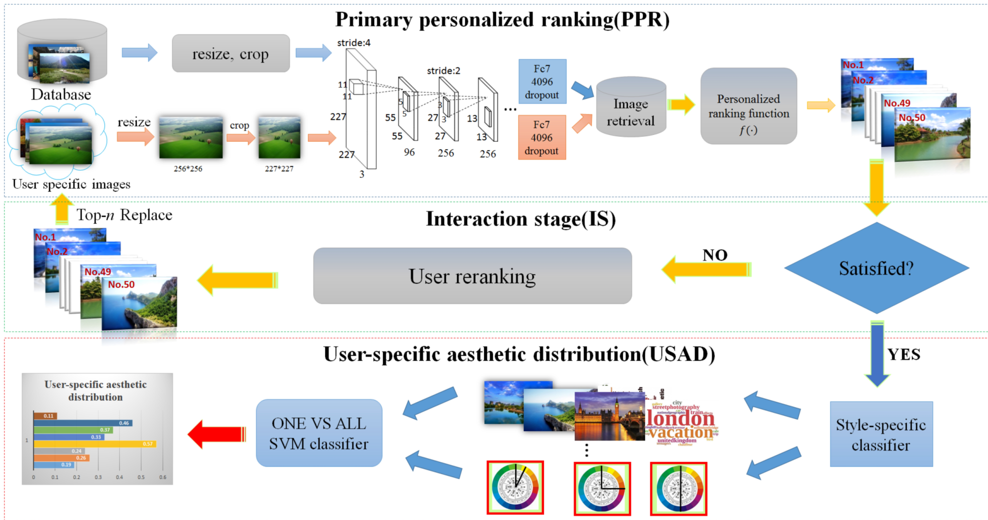

FIGURE 5 – *The system developed by Lv* et al. *uses a learning loop that can be iterated through. The system provides the user with images from the annotated database that are close to a small set of images that the user likes. The user confirms the relevance of these images or rejects them, so that the set finally selected corresponds well to his tastes. These are the images that will be used to train the system [Lv et al., 2018].*

3.5 Multiplying competing expertise

One way to break down the aesthetic objectivity of RNP is to explicitly introduce subjectivity with the help of experts in charge of representing competing aesthetic paths. This is a path that is also partly explored in the work of [Zhu et al., 2020] seen above.

This approach has been addressed by S. Kong et al. [Kong et al., 2016] who entrust the evaluation of their AADB base to a small number of real aesthetic experts whose expertise they keep track of. Each expert expresses a different sensitivity. Thus, one can use only the notes of one expert to evaluate an image (if an Internet user feels more in agreement with one of them) or on the contrary several of them.

More ambitious are the works reported in [Hong et al., 2016] which try to determine communities of tastes among the population of Flickr subscribers. For this purpose, two types of descriptors are used for each image : semantic descriptors on the one hand (from the captions provided with each image), and low-level descriptors on the other hand (color, contrast, contours, textures). In the space of these descriptors, each image is treated as a word by a latent allocation algorithm of Dirichlet allocation algorithm (LDA). The words are grouped into phrases (a phrase represents a photographer) which thus provide latent topics defining communities (a topic represents a community). Communities share neighboring topics in the descriptor space. If the approach is very interesting, the mathematical tools involved are considerable and the first results are not totally convincing. The respective weights of the various descriptors are difficult to find and it is difficult to separate the roles of aesthetic tastes and interest for a photographic theme (see figure 6).

We would like to be able to group these last two approaches to build unquestionable communities around experts pertinently chosen to express well established aesthetic tendencies, but this approach of classification of aesthetics in the field of photography does not seem to have been undertaken to date.

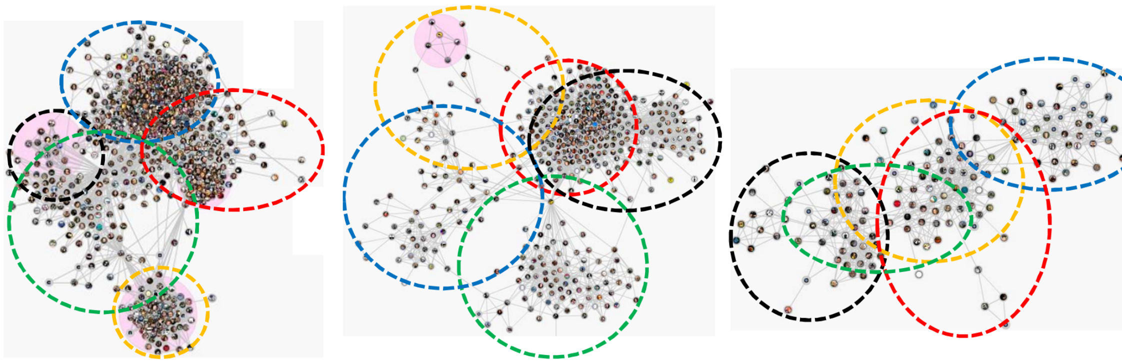

FIGURE 6 – Detection of communities in descriptor clouds using a latent Dirichlet allocation technique and graph partitioning of the graph. Communiti*es are identified by colored ellipses. In red : "designers", in blue : "colors", in green : architecture", in yellow : "Black and White". The 3 representations are particular projections of the cloud, according to [Hong et al., 2016]. We see that there is an important overlap between these communities and the themes of the photos.*

3.6 Conclusion

The approach adopted by the computational aesthetics community to escape from a purely objectivist framework is the result of the criticisms that, despite obviously positive results, the work carried out within this objectivist framework using methods based on deep neural networks trained on large bases established once and for all, have met.

The multiplicity and diversity of the approaches used to bring a note of subjectivity to the judgment of taste reflect the complexity of the project. The extent and the variety of the determinants of the aesthetic judgment, the ignorance of their roles and of their reciprocal influences make the approaches extremely uncertain and explain these groping attempts, generally sensible but finally very little convincing. The absence of a unanimously accepted frame of reference, as much in the field of philosophical aesthetics as in that of neuro-biology or experimental psychology, leaves the computer scientist disoriented as to the paths to follow. The approaches taken are sometimes skilful, they are always technically very elaborate ; they call upon artificial intelligence schemes that are often very refined, but they also reflect a bias towards narrow paths that probably contribute to the choice, but only imperfectly explain the judgment of taste. Thus, they struggle to convince. Their aesthetic premises are too frequently indigent. Their references are often aligned with sociological practices whose validity in the field of taste has not been established because the other determinants of choice have not been sufficiently ruled out. The criteria of success that they display challenge us as much by their modest significance as by the narrowness of their validity. We have certainly not exhausted all the resources that the technique keeps producing in the field of algorithms, but we are waiting for the results that neurobiology will be able to produce with the modern tools of functional exploration by reducing the temporal scales of the measurements, by increasing the sensitivities, by extending the range of the follow-up of the cerebral signals so as to better understand the chronology of the human decision.

Translated with the help of www.DeepL.com/Translator (free version) from **Juger du Beau avec subjectivité : le défi de l'esthétique computationnelle** *Arts et sciences, N :4 Vol : 4 Date : 2020/12/21 - DOI : 10.21494/ISTE.OP.2020.0588*

Références

[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). *Einfürung in die informationstheoretische Ästhetik,. Grundlegung und Andwendung in der Texttheorie.* Rowoldt Taschenbuch Verlag.

[Berridge and Kringelbach, 2013] Berridge, K. and Kringelbach, M. (2013). Neuroscience of affect : brain mechanisms of pleasure and displeasure. *Current Opinion in Neurobiology*, 23(3) :294–303.

[Birkhoff, 1933] Birkhoff, G. (1933). *Aesthetic Measure.* Harvard University Press, Cambridge (USA).

[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.

[Brunswick, 1956] Brunswick, E. (1956). *Perception and the representative design of psychological experiments.* University of California Press.

[Chatterjee, 2004] Chatterjee, A. (2004). Prospects for a cognitive neuroscience of visual aesthetics. *Bulletin of psychology and the arts*, 4(2) :55–60.

[Cochoy, 2011] Cochoy, F. (2011). *De la curiositÃ©, l'art et la sÃ©duction marchande.* Armand Colin, Paris.

[Costa and McCrae, 1992] Costa, P. and McCrae, R. (1992). *NEO-PI-R Professional manual.* Odessa : Psychological AssessmentResources.

[Cristani et al., 2013] 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 et al., 2014] 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.

[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., 2017] Deng, Y., Loy, C. C., and Tang, X. (2017). Image aesthetic assessment : An experimental survey. *IEEE Signal Processing Magazine*, 34(4) :80–106.

[Di Dio and Vittorio, 2009] Di Dio, C. and Vittorio, G. (2009). Neuroaesthetics : a review. *Current Opinion in Neurobiology*, 19 :682–687.

[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.

[Dretske, 1969] Dretske, F. (1969). *Seeing and Knowing.*

[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.

[Elridge, 2015] Elridge, A. (2015). Photography and sociology, an exercise in serendipity.

[Eysenck, 1991] Eysenck, H. (1991). Dimensions of personality : 16, 5 or 3âcriteria for a taxonomic paradigm. *Personality and individual differences*, 12(8)(8) :773–790.

[Eysenck, 1941] Eysenck, H. J. (1941). The empirical determination of an aesthetic formula. *Psychological Review*, 48(1) :83–92.

[Fechner, 1871] Fechner, G. (1871). *Zur Experimentalen Aesthetik.* Hirtzel, S., Leipzig.

[Goldberg, 1990] Goldberg, L. (1990). An alternative âdescription of personalityâ : The big-five factor structure. *Journal of Personality and Social Psychology*, 59 :1216–1229.

[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.

[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.

[Hsu, 2009] 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 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.

[Ishizu and Zeki, 2011] Ishizu, T. and Zeki, S. (2011). Toward a brain-based theory of beauty. PLOS One.

[Jacobsen, 2010] Jacobsen, T. (2010). Beauty and the brain : culture, history and individual differences in aesthetic appreciation. *Journal of anatomy*, 216(2) :184–191.

[John et al., 1991] John, O., Donahue, E., and Kentle, R. (1991). Big five inventory. *Journal of Personality and Social Psychology.*

[Kairanbay et al., 2019] 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 et al., 2015] 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, 1979] Konecni, V. (1979). Determinants of aesthetic preference and effects of exposure to aesthetic stimuli : social, emotional and cognitive factors. *Prog Exp Pers Res.*, 9 :149â197.

[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.

[Kosinski et al., 2014] 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 et al., 2013] 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 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.

[Li et al., 2020] 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.

[Lovato et al., 2013] 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.

[Lv et al., 2018] 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 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.

[Maitre, 2021] Maitre, H. (2021). *L'esthétique en photographie numérique*. ISTE.

[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.

[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.

[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.

[O'Donovan et al., 2014] 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 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.

[Pasquier et al., 2014] Pasquier, D., Beaudoin, V., and Legon, T. (2014). *"Moi je lui donne 5/5", Paradoxes de la critique amateur en ligne*. Presses des Mines (Paros).

[Petitot, 2008] Petitot, J. (2008). *Neurogéométrie de la vision*. Les Editions de l'Ecole Polytechnique.

[Rammstedt and John, 2007] 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, 2020] Ray, L. (2020). Social theory, photography and the visual aesthetic of cultural modernity. *Cultural Sociology*, (1749975520910589).

[Reagle, 2013] Reagle, J. (2013). Revenge rating and tweak critique at photo.net. personal website : http ://reagle.org/joseph/2013/photo.net.html.

[Redi et al., 2017] 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, 2015] Redies, C. (2015). Combining universal beauty and cultural context in a unifying model of visual aesthetic experience. *Frontiers in human neuroscience*, 9 :218.

[Ren et al., 2017] 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 et al., 2008] Rigau, J., Feixas, M., and Sbert, M. (2008). Information aesthetic measures. *IEEE Computer Graphics and Applications*, 2 :24–34.

[Schifanella et al., 2015] 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 et al., 2017] 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 et al., 2016] 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 et al., 2018] 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 and Milanfar, 2017] Talebi, H. and Milanfar, P. (2017). NIMA : Neural image assessment. arXiv : 1709.0541v1.

[Wertheimer, 1938] Wertheimer, M. (1938). *Laws of of organization in perceptual forms.* Ellis, W.D., A source book of Gestalt psychology, London.

[Yang et al., 2015] 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 et al., 2016] You, Q., Bhatia, S., and Luo, J. (2016). A picture tells a thousand wordsâabout you ! user interest profiling from user generated visual content. *Signal Processing*, 124 :45–53.

[Zhu et al., 2020] 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.*