UNIVERSITY OF PLYMOUTH

COMPUTATIONAL MODELLING OF HUMAN AESTHETIC PREFERENCES IN THE VISUAL DOMAIN:

A BRAIN-INSPIRED APPROACH

François Lemarchand

A thesis submitted to the University of Plymouth

in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

2018

Copyright Statement

Abstract

Following the rise of neuroaesthetics as a research domain, computational aesthetics has also known a regain in popularity over the past decade with many works using novel computer vision and machine learning techniques to evaluate the aesthetic value of visual information. This thesis presents a new approach where low-level features inspired from the human visual system are extracted from images to train a machine learning-based system to classify visual information depending on its aesthetics, regardless of the type of visual media. Extensive tests are developed to highlight strengths and weaknesses of such low-level features while establishing good practices in the domain of study of computational aesthetics. The aesthetic classification system is not only tested on the most widely used dataset of photographs, called AVA, on which it is trained initially, but also on other photographic datasets to evaluate the robustness of the learnt aesthetic preferences over other rating communities.

The system is then assessed in terms of aesthetic classification on other types of visual media to investigate whether the learnt aesthetic preferences represent photography rules or more general aesthetic rules. The skill transfer from aesthetic classification of photos to videos demonstrates a satisfying correct classification rate of videos without any prior training on the test set created by Tzelepis et al. Moreover, the initial photograph classifier can also be used on feature films to investigate the classifier's learnt visual preferences, due to films providing a large number of frames easily labellable. The study on aesthetic classification of videos concludes with a case study on the work by an online content creator. The classifier recognised a significantly greater percentage of aesthetically high frames in videos filmed in studios than on-the-go. The results obtained across datasets containing videos of diverse natures manifest the extent of the system's aesthetic knowledge.

To conclude, the evolution of low-level visual features is studied in popular culture such as in paintings and brand logos. The work attempts to link aesthetic preferences during contemplation tasks such as aesthetic rating of photographs with preferred low-level visual features in art creation. It questions whether favoured visual features usage varies over the life of a painter, implicitly showing a relationship with artistic expertise. Findings display significant changes in use of universally preferred features over influential

abstract painters' careers such an increase in cardinal lines and the colour blue; changes that were not observed in landscape painters. Regarding brand logos, only a few features evolved in a significant manner, most of them being colour-related features.

Despite the incredible amount of data available online, phenomena developing over an entire life are still complicated to study. These computational experiments show that simple approaches focusing on the fundamentals instead of high-level measures allow to analyse artists' visual preferences, as well as extract a community's visual preferences from photos or videos while limiting impact from cultural and personal experiences.

Table of Contents

Chapter 1 Introduction 1

1.1. Problem formulation 2

1.2. Clarifying the vocabulary: Beauty, aesthetics and visual preferences 3

1.3. Structure of the thesis 4

Chapter 2 Human Visual Preferences and Aesthetics: A Literature Review 6

2.1. The origins of the study of aesthetics and visual preferences 6

2.2. From visual processing to aesthetic pleasantness 8

2.2.1. Visual structure and composition 9

2.2.2. Visual complexity 13

2.2.3. Familiarity with the visual stimuli and conceptual association 15

2.2.4. Modelling aesthetic judgement 17

2.3. Computational processing of visual information 20

2.3.1. Image Processing 21

2.3.2. Visual Feature Extraction for Aesthetic Measures 23

2.3.3. Advanced Feature Extraction Algorithms: Visual Descriptors 24

2.3.4. Binary classifiers: Support Vector Machine & Multilayer Perceptrons 26

2.3.5. Deep Learning 29

Chapter 3 Theoretical and Conceptual Frameworks 35

3.1. Trends in the study of human visual preferences 35

3.2. Situating the thesis in the domain of study of human visual preferences 39

3.3. Methods 42

Chapter 4 Aesthetic Classification of Images Using Brain-Inspired Features 46

4.1. Dataset configurations in related works 50

4.1.1. Description of the first dataset of photographs with aesthetic ratings (Datta et al.'s dataset) 50

4.1.2. Description of a large-scale dataset of photographs with aesthetic ratings (AVA Dataset) 50

4.1.3. Computational Aesthetic Classification 51

4.2. Method: Brain-inspired visual feature extraction and machine learning 53

4.2.1. Gradient orientation distribution 53

4.2.2. Distribution of local curvature 54

4.2.3. Global symmetry 55

4.2.4. Colour: hue, saturation, and brightness 56

4.2.5. Parameter selection for feature extraction algorithms 57

4.2.6. Neural network architectures and parameters 58

4.3. Machine learning models for aesthetic classification 61

4.3.1. Classification results 61

4.3.2. Locating the border between aesthetic classes 62

4.3.3. Cross-dataset performance and comparison with existing works 65

4.3.4. Visual preferences of the aesthetic classifier and differences between the average features of each class 68
4.4. Discussion 71
4.5. Conclusion 73
Chapter 5 Computational Aesthetic Classification Beyond Photographs 76
5.1. Training of the aesthetic classifier 77
5.2. Experiment 1: Aesthetic classification of video clips by Tzelepis et al. 79
5.3. Experiment 2: Using the trained aesthetic classifier to evaluate a film’s frames 82
5.4. Experiment 3: Aesthetic prediction on YouTube content creators 87
5.5. Conclusion 89
Chapter 6 Analysis of Visual Features in Paintings and Logos 91
6.1. Evolution of low-level visual features in influential painters’ careers 91
6.2. Extracting low-level visual features from paintings 93
6.3. Preliminary tests: use of cardinal lines over an artist’s career 95
6.4. Data filtering 96
6.5. Dataset composition 98
6.6. Method 101
6.7. Results 102
6.8. Discussion 109
6.9. Applying low-level feature analysis on brand logos’ history 111
6.9.1. Dataset 113

6.9.2. Method ... 113

6.9.3. Results ... 114

6.9.4. Discussion on low-level features in logo designing ... 118

6.10. Conclusion ... 119

Chapter 7 Proof of Concept: Psychological Games and Real-World Applications with an Aesthetic Classifier ... 120

7.1. Human performances on aesthetic classification tasks ... 120

7.2. PikPik: Android app for aesthetic filtering of photographs ... 123

7.3. Discussion ... 126

Chapter 8 Discussion and Conclusion ... 127

8.1. From low-level visual features extraction to experimental tool ... 127

8.2. Successes and shortcomings ... 128

8.3. Thesis contributions ... 132

8.4. Future work ... 135

Bibliography ... 136

List of Figures

Figure 2.1: Video quality available over time on Youtube 22

Figure 2.2: Number of transistors per CPU since the 1970s 23

Figure 2.3: SVM representation of inputs and transform into high-dimensional space. 27

Figure 2.4: Basic architecture of a multilayer perceptron. 28

Figure 2.5: Representation of a neural network with and without dropout 31

Figure 2.6: Kernel processing the input volume into the convolutional layer. 33

Figure 2.7: Connections in a traditional neural network and CNN 33

Figure 4.1: Extraction of orientation and curvature distributions. 55

Figure 4.2: Representation of the 6CDNN architecture. 60

Figure 4.3: Performance of the aesthetic classifier over different ranges of ratings 64

Figure 4.4: Performance of the aesthetic classifier exclusive ranges of ratings. 65

Figure 4.5: Differences between the average features representing orientations of aesthetically high and low images for Datta et al. and AVA datasets. 70

Figure 4.6: Differences between the average features representing curvature of aesthetically high and low images for Datta et al. and AVA datasets 71

Figure 4.7: Differences between the average features representing colour hues of aesthetically high and low images for Datta et al. and AVA datasets. 71

Figure 5.1: Mean percentage of aesthetically high frames detected in a video by the aesthetic classifier depending on the average human rating of the given video. 81

Figure 5.2: Relationship between brightness and aesthetic prediction in the film The Hateful Eight (2015) 86

Figure 5.3: Aesthetic prediction over time, averaged using a sliding window of 500 frames of the film, Django Unchained (2012). 87

Figure 5.4: Percentages of aesthetically high frames in each video of the different seasons of vlogging by Casey Neistat. 89

Figure 6.1: Percentage of lines with cardinal orientations in paintings by Wassily Kandinsky and Pablo Picasso through their respective career. 96

Figure 6.2: Average representation of specific features in landscape photos, landscape paintings, and abstract paintings. 103

Figure 6.3: Differences in the distribution of gradient orientations between early and late works in landscape and abstract paintings.. 104

Figure 6.4: Differences in the distribution of curvatures between early and late abstract paintings. 106

Figure 6.5: Differences in the distribution of colour hue between early and late works in landscape and abstract painting. 107

Figure 6.6: Differences in the distribution of colour saturation between early and late abstract paintings. 108

Figure 6.7: Differences in the distribution of colour brightness between early and late abstract paintings. 108

Figure 6.8: Percentage of use of different visual features in an average painting by a 20th-century abstract artist 109

Figure 6.9: Evolution of orientations in logos over two different time settings.. 115

Figure 6.10: Evolution of curvatures in logos over two different time settings. 115

Figure 6.11: Evolution of colour hues in logos over two different time settings. 116

Figure 6.12: Evolution of colour saturation in logos over two different time settings. 116

Figure 6.13: Evolution of colour brightness in logos over two different time settings. 116

Figure 7.1: Screenshot of the interface used in the classification task for human participants to challenge the aesthetic classifier.. 121

Figure 7.2: Screenshot displaying the participant and the aesthetic classifier's scores 122

Figure 7.3: Screenshots of the initial page of PikPik. 123

Figure 7.4: Screenshots of the selection process in PikPik.. 125

Figure 8.1: Model of aesthetic judgement. 131

List of Tables

Table 4.1: Parameters for the deep neural architectures. 60

Table 4.2: Performance comparison on different datasets and machine learning algorithms... 62

Table 4.3: Proposed aesthetic classifier compared against previous works 68

Table 5.1: Aesthetic classification on the top-n percent highest-rated videos. 82

Table 6.1: Abstract artists studied. 99

Table 6.2: Landscape painters studied. 100

Table 6.3. Number of images and average pixel size in existing datasets. 105

Table 6.4: List of companies and brands studied. 117

Table 7.1: Performance comparison between human and aesthetic classifier when guessing the community's preferences for photographs 122

Acknowledgements

I would like to thank my supervisor, Roman Borisyuk, for his dedication, his time, and his ability to always see the bigger picture over the last four years. Roman taught me how essential it is to build a good story to convey ideas, which I believe had a significant impact on the work presented in this thesis. I also owe a great amount of gratitude to Sue Denham for always being so enthusiastic and ready to discuss new ideas. I must also thank my examiners, Thomas Wennekers and Alex Forsythe, for their helpful and constructive suggestions.

Big thanks must go to all the Cognovo fellows. Their presence made everything possible, and the solidarity helped to stay afloat in moments of doubts. The conversations regularly reached new levels of madness, leading to some brilliant ideas which, from time to time, concluded in productive collaborations. A special thank goes to Frank Loesche and David Bridges who have helped to proofread the work presented in this thesis.

I am thankful to my family for always staying positive and be interested in the latest developments of my thesis. Of course, I would like to thank my partner, Jenna, for her love and her continued support, which were crucial over the last few months of writing this thesis.

Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Doctoral College Quality Sub-Committee.

Work submitted for this research degree at the University of Plymouth has not formed part of any other degree either at the University of Plymouth or at another establishment.

This work was financed with by a fellowship from the Marie Curie Initial Training Network FP7-PEOPLE-2013-ITN, CogNovo, grant number 604764 The CogNovo doctoral programme included a series of multi-disciplinary seminars and conferences that required active participation, and often, presentation of the current work under development.

Publications:

Lemarchand, F., (2018). "Evolution of Low-Level Visual Features in Influential Painters' Careers". (Submitted)

Lemarchand, F., (2018). "Fundamental visual features for aesthetic classification of photographs across datasets". *Pattern Recognition Letters*, *112*, pp.9-17.

Lemarchand, F., (2017). "From Computational Aesthetic Prediction for Images to Films and Online Videos". *Avant*, *8*, pp.69-78.

Lemarchand, F., (2015), "Visual Arts Creation Assisted by BICASSO: Brain-Inspired Computationally Aesthetic Selective Savant & Observer", *Conference Proceedings of ISEA (International Symposium on Electronic Art)*.

Presentations at conferences:

Lemarchand, F., (2017), "Computational Prediction of Aesthetics in Movies and Online Videos", *Off the Lip 2017*, Plymouth, United Kingdom.

Lemarchand, F., (2016), "Dataset Selection & Aesthetic Measures for Computational Evaluation of Visual Art" (Poster), *The Neural Basis for Creativity Workshop*, Plymouth, United Kingdom.

Lemarchand, F., (2015), "Visual Arts Creation Assisted by BICASSO: Brain-Inspired Computationally Aesthetic Selective Savant & Observer", *ISEA2015*, Vancouver, Canada.

Public engagement:

Lemarchand F., (2016). "Fitting a Population's Visual Preferences into an Artificial Intelligence", Off the Lip 2016, Plymouth, United Kingdom.

Bridges D., Sonne Kristensen M., Lemarchand F., Loesche F., (2016), "The Exciting Synesthesia Machine", Off the Lip 2016, Plymouth, United Kingdom.

Conferences and workshops attended:

June 2017: Hack The Brain 2017 (Dublin Science Gallery, Ireland)

April 2015: Entrepreneurship workshop (CogNovo, Plymouth University, UK).

March 2015: Integrated Systems Neuroscience Workshop 2015 (Manchester University)

January 2015: Public outreach and social innovation workshop (CogNovo, Plymouth University, UK).

November 2014: EEG workshop (CogNovo, Plymouth University, UK).

September 2014: Computational modelling workshop (CogNovo, Plymouth University, UK).

June 2014: Experimental methods workshop (CogNovo, Plymouth University, UK).

June 2014: International Conference of Computational Creativity 2014, Ljubljana, Slovenia.

April 2014: Research methods workshop (CogNovo, Plymouth University, UK).

Peer-reviewing:

August 2017: Off The Lip 2017 (Plymouth, UK)

March 2017: H-WORKLOAD - 1st International Symposium on Human Mental Workload: Models and Applications (Dublin, Ireland)

Other:

Qualcomm Artificial Intelligence Developer Contest Finalist with the Android project "Aesthetic Filtering for Photos", 2017.

Co-Lead of the ColLaboratoire Summer School project "Remapping the Sensorium: Do you hear what I see?" with Frank Loesche, David Bridges, Susan L. Denham, Michael Sonne Kristensen and Thomas Wennekers, 2016.

Word count of main body of thesis: 33,950

Signature:

Date:

Chapter 1

Introduction

In the study of human visual preferences, experiments of behavioural psychology and neuroscience have been dominating the study of the phenomenon and have more recently been placed under the banner of neuroaesthetics (Chatterjee, 2011). Meanwhile, a subfield of artificial intelligence called computational aesthetics has been gaining momentum and has provided more insights into the understanding of human preferences. Historically, the domain is thought to originate in David Birkhoff's attempt to define a theoretical measure of aesthetic taking order and complexity into account (Birkhoff, 1933). With the increase in computational resources over the last decades, the field has evolved from its purely mathematical roots into computational experiments manipulating large datasets. Machine learning systems now train to predict human visual preferences and aesthetic judgements. The most prominent datasets used for research in visual aesthetics are assembled using community-based photography websites, implying thousands of pictures provided with hundreds of aesthetic ratings (Datta et al., 2006; Murray et al., 2012). While the datasets are not conceived in a strictly controlled environment as the ones in traditional psychology and neuroscience experiments, it offers opportunities to investigate human visual preferences and aesthetics on a larger scale.

1.1. Problem formulation

The **aim** of this thesis is to select a set of features represented in the human early visual system to evaluate the effect of low-level visual information (orientation, curvature, colour, etc.) in aesthetic judgement. The selected features should be implemented and fed into different machine learning algorithms to classify visual information based on its aesthetic value. As state-of-the-art solutions currently focus on binary classification, the proposed systems will be referred to as aesthetic classifiers. The selected features should efficiently represent human preferences and contribute heavily to the proposed aesthetic classifier's state-of-the-art performances. It should extend state-of-the-art works due to the used low-level features which receive limited impact from cultural and personal experiences. The robustness of such aesthetic classifier should display cross-media capabilities, allowing investigations on the links of aesthetic criteria between different visual media. It should also make it possible to examine creative processes such as painting or logo designing, establishing a basis for future studies of aesthetics across datasets and visual media. Therefore, the brain-inspired features and the resulting aesthetic classifier should present a strong case that aesthetic evaluation is possible even when significantly discarding high-level information.

The **scope** of this work is designed to be as broad as possible and is one of the main strengths of the proposed system, due to its different sources of inspiration from neuroaesthetics to computer vision. Indeed, the ambition behind the designed aesthetic classifier is to offer behavioural analyses while providing automated aesthetic predictions, making it attractive for developers and researchers in recommendation systems. From a computational resource perspective, the proposed feature extraction algorithms and the deep neural network should be able to run on any computer and even on the most recent phones. However, the proposed aesthetic classifier may be limited by the number of low-level features presented in this proof of concept, meaning it is acceptable to have future implementations outperforming it.

1.2. Clarifying the vocabulary: Beauty, aesthetics and visual preferences

This thesis contains vocabulary issued from neuroaesthetics and computational aesthetics. Due to the multiple disciplines involved, a significant dilemma relies on the fact that each subject has its definition and interpretation of the vocabulary. To begin with one of the most recurrent words in the study of aesthetics, the word “beauty” is used in various ways despite the strong cultural and philosophical meaning it conveys. For instance, neuroscientists and psychologists would define a stimulus as “beautiful” when the participant can report it as such, or the visual information triggers some observable positive physiological rewards (Ishizu & Zeki, 2011). In the meantime, artistic practices would tend to not limit themselves to the visual stimuli and define the whole experience of seeing a visual stimulus as “beautiful”, including the feeling of being confronted to the physical object that a painting or a photograph is and the eventual social context. Therefore, the word "beauty" is not used in this thesis to avoid any ambiguity.

To be as accurate as possible, two principal terms are used to replace the concept of “beauty”. The first term is “aesthetic pleasantness”, which aims to emphasise the relationship between the set of rules and features included in the visual stimulus and the positive reward felt by the average observer. The second term is “visual preferences”, which is used when discussing how the observer perceives the visual stimulus and how specific groups of people are stimulated by specific visual features. While not limited to those two expressions, they illustrate how particular attention is given to dissociate visual stimuli and observers.

1.3. Structure of the thesis

The work developed in this thesis anchors itself deeply into empirical findings in the study of aesthetics. Chapter 2 introduces the historical developments leading to the study of human visual preferences, before reviewing the different trends that have been appearing along the last few decades regarding aesthetics in the domain of neuroscience of vision, psychology, and psychophysics. It is then followed by an overview of the different computer vision tools used in the field of computational aesthetics.

Following the literature review, Chapter 3 clarifies how the different approaches interact with each other in the domain of study of human visual preferences and their shortcomings are enumerated. It allows to address the theoretical framework motivating the work presented in this thesis, as well as some general methods that will be used in the following chapters.

Chapter 4 presents how a set of low-level features have been selected and developed. The set of features is then coupled with machine learning algorithms to reproduce human aesthetic judgements on datasets of photographs. This first computational experiment applies strict machine learning methodology and establishes a framework to compare state-of-the-art systems in computational aesthetics. The proposed aesthetic classifier is also tested for cross-dataset performances, with no notable work offering comparable results in the past.

The demonstration that low-level features offer state-of-the-art performances when coupled with deep learning allows to move onto a new type of computational experiments in computational aesthetics. Chapter 5 exposes how the aesthetic classifier is used to categorise videos by assessing frames one by one and attempts cross-media aesthetic classification.

After achieving state-of-the-art levels of classification on photographs and satisfying cross-media performances on videos, Chapter 6 presents how the set of features is used to investigate human preferences on highly diverse types of visual information such as paintings and brand logos. The experiment attempts to link the set of features used for aesthetic classification with possible visual features used by influential artists and companies dominating their industries over the years.

To finish, Chapter 7 addresses possible real-world applications of such works while relying on prototypes, before summarising all the results and assessing the contributions to the study of human visual preferences and aesthetics in Chapter 8.

Chapter 2

Human Visual Preferences and Aesthetics: A Literature Review

2.1. The origins of the study of aesthetics and visual preferences

Artistic values and trends across centuries have heavily influenced the conversation on aesthetics. One of the first documented discussions on aesthetic experience originates back to ancient Greece with Plato and Aristotle addressing mimesis (Beardsley, 1975). Mimesis hypothesises that artistic beauty results from a representation of the real world. Plato firmly stated that art is a simple imitation of reality (Plato, 2003) from which it borrows its aesthetic qualities. Aristotle agreed to the idea of mimesis but rejected Plato's opinion about art, arguing that art provides added value to the aesthetic attributes of the depicted object (Aristotle, 1996). Ancient Greece was a period rich in technical innovations for the artistic world and where works of art were primarily valued by their fidelity to represent real-world scenes. This approach brought an increasing flow of mathematical knowledge into the artistic world, which led to the creation of sophisticated technical methods to draw perfect shapes or credible perspectives (Casalderrey, 2017). The idea was later revisited and reached its peak with Renaissance paintings displaying photorealistic renders. These artworks left little room for technical improvements in the reproduction of the real-world and consequently triggered a new exploratory phase in search of the meaning of aesthetics. Immanuel Kant in 1781 with the *Kritik der reinen Vernunft* (Critique Of Pure Reason) described the action of aesthetic judgement as a subjective phenomenon. Regardless of the reflections expressed by philosophers, it appears that the creative artistic process was merely addressed and the product remained the main subject. With Kant, the aesthetic quality of an object and the observer's opinion are dissociated (Kant, 1999). Kant emphasised that aesthetic judgements are made without expecting a reward from the assessed object, as pure aesthetic evaluation should not take the functions of an object into account. Therefore, by interpreting the real world as a functionless medium, true beauty can be appreciated without any ambiguity or expectations.

Coinciding with art evolving towards modern art in the middle of the 19th century, the birth of modern psychology motivated the first empirical study about perception and aesthetics in 1860 with *Elemente der Psychophysik* (Elements of psychophysics) by Gustav Fechner (Fechner, 1860). Fechner's work marked the creation of psychophysics, the field of study exploring the relationship between visual stimuli and the observers' experience or behaviour. Contradicting previous philosophical discussions on beauty, it focused on defining key elements of visual aesthetics such as colours or shapes in a bottom-up manner (Fechner, 1876). Fechner's bottom-up perspective on the study of perception was challenged by Gestalt psychology, stipulating that features are always processed as a whole (Ehrenfels, 1890; Koffka, 1922; Wertheimer, 1923). Despite being initially developed to investigate visual perception, both psychophysics and Gestalt psychology are often sources of inspiration and cited by studies examining aesthetic experiences. It was only in 1999 that Semir Zeki and Vilayanur S. Ramachandran attempted to connect the "science of the beautiful" to the neurological knowledge at the time, both pioneering the field of neuroaesthetics (Ramachandran & Hirstein, 1999; Zeki, 1999). While the first statements were mainly theoretical and partly speculative, Zeki and his colleague Hideaki Kawabata applied their expertise in the neuroscience of vision and established the first study using functional magnetic resonance imaging (fMRI) to investigate correlations between activated brain areas during aesthetic experiences (Kawabata & Zeki, 2004).

This brief introduction traces the study of aesthetic experiences and human visual preferences back to philosophical discussions in ancient Greece. Following the Renaissance period and with the emergence of psychological studies, scientific methods were increasingly used to study the topic. This section highlights how it only recently became its own research domain, instead of being a by-product of research on perception, vision, or philosophy of the arts.

2.2. From visual processing to aesthetic pleasantness

In both theoretical papers marking the birth of neuroaesthetics by Zeki and Ramachandran, a number of laws are suggested to categorise the types of aesthetics experiences in their relation to the human visual system (Ramachandran & Hirstein, 1999; Zeki, 1999). Zeki hypothesised that the visual brain aims at abstracting visual information to reduce noise and easily interpret the surrounding environment. Furthermore, Ramachandran and Hirstein developed eight laws to categorise artistic experiences: peak shift principle, grouping, isolating, contrast, symmetry, generic viewpoint, Bayesian logic of perception, and metaphors. Inspired from psychophysics and Gestalt psychology, the principles suggested by Zeki and Ramachandran can themselves be classified into three types. The principles focus on the structure of the visual stimuli, the level of complexity that leads to the understanding of the visual scene, and the association of the visual information with the observer's previous knowledge. It highlights primary factors in aesthetic experiences and by consequence, in aesthetic judgement. Therefore, the problem can be simplified and reformulated with the three main factors which are structural properties, complexity and familiarity.

Building on the previous discussion, the following three definitions are used throughout this thesis. Firstly, a study is interpreted as investigating the impact of visual structure on aesthetic experiences if the visual stimuli presented to participants contain specific colour or line combinations, or a number of repetitions such as recursive features, or symmetries (reflectional, rotational and translational). Secondly, a study is understood as targeting visual complexity if the stimuli can be characterised by the density and variety of visual components. Thirdly, familiarity depends on the intensity and frequency of previous exposures to specific visual features or stimuli. Consequently, it varies across observers and may be seen as a purely subjective experience.

This literature review develops each of the previously mentioned factors. Hereby it avoids grouping findings on philosophical or neurological levels as well as observations in specific contexts. Contextualised discussions, for example on landscape or human faces, are intentionally excluded from the following discussion to offer more general conclusions and avoid topics with cultural and individual preferences.

2.2.1. Visual structure and composition

The impact on aesthetic judgement by low-level visual components and their distribution across a two-dimensional space have been extensively studied in Gestalt and psychophysics studies. Regarding colours, it is important to note that studies on colour preference tend to use HSB (Hue, Saturation, Brightness) encoding, instead of the RGB (Red, Green, Blue) encoding. HSB encoding was introduced as a colour model in alignment with perception (Joblove & Greenberg, 1978). It is, therefore, used in applications involving users, unlike RGB encoding which is less intuitive and harder to manipulate due to being a mix of the additive primary colours.

In the human retina, two types of photoreceptor cells are sensitive to light, cones and rods. Cone cells act as colour receptors and exist in three different types. Each type of cone cell responds to a specific wavelength range, which corresponds roughly to red, green and blue (Kandel et al., 2000). The difference between the information received by these three types of cones then allows to identify perceived colours. A variety of fMRI experiments on humans and more invasive studies on other mammals with similar neurobiological structures lead to building robust hypotheses on colour processing at deeper neural levels. Areas in the visual cortex are numbered from V1 to V5 and are functional units within the occipital lobe towards the back of the head. The wavelengths received from the retina are thought to be registered and differentiated in the primary visual cortex, V1, with groups of cells, called blobs, also sensitive to specific wavelengths, before projecting onto V2, itself connected onto V4 (Kandel et al., 2000; Sincich & Horton, 2005). However, recent reviews question the blob-based structure in V1, due to evidence weakening the hypothesis that neural cells within blobs are sensitive to colours (Conway et al., 2010; Solomon & Lennie, 2007). While the visual area V4 was initially considered as the centre for colour processing, this hypothesis was dismissed due to its involvement in other aspect of visual processing such as shape information (Brouwer & Heeger, 2009; Hadjikhani et al., 1998; Heywood et al., 1992; Motter, 1994; Zeki & Marini, 1998). V4 appears to be involved in feature selectivity, contributing to visual scene understanding by highlighting salient features, but also to a more active attentional mechanism (Roe et al., 2012).

More recently, investigations on colour preferences have been led mostly through psychological experiments. As reviewed by Palmer et al., several studies have found conclusive results in the average adult regarding a positive preference for shades of blue and a negative preference for shades of yellow and orange (Palmer et al., 2013). Preferences are not limited to individual colours, as some colour combinations are known to be favoured by people (Ou et al., 2004b). The presence of other colours can impact the perception of individual colours, as well as the entire visual scene (Locher et al., 2005). Nonetheless, there are still significant variations in colour preferences across cultures and individuals (Ou et al., 2004a). For this reason, studies on infants shape a promising field on potential hardwired preferences in the human visual system. A study on infants by Franklin et al. demonstrates an innate positive colour preference for red hues and negative preference for green hues, with no difference between sexes, unlike in adults (Franklin et al., 2010).

While the understanding of colour processing on a neurobiological level is still limited, many studies have also been conducted to improve the understanding of edge and shape processing. Experiments on mammals have demonstrated orientation selectivity in low-level visual areas, with clear preferences for cardinal orientations (Blasdel, 1992; Chapman & Bonhoeffer, 1998). Similarly to the blobs of neural cells sensitive to specific colours, other groups of cells in V1, called interblobs, show preferences for orientation but can still be stimulated by some colours (Dow, 2002; Kandel et al., 2000; Solomon & Lennie, 2007). As for colours, the visual area V4 is involved in feature selection, but it is also sensitive to orientations, curves and some basic shapes (Roe et al., 2012). Despite the uncertainties behind the neural mechanisms composing the human visual system, psychological experiments confirm and emphasise the existence of such orientation preferences in humans (Girshick et al., 2011; Latto & Russell-Duff, 2002). This type of experiments also allows to go further and investigate the preference in shapes on a higher level, showing positive responses to smooth shape over sharp corners (Bertamini et al., 2016; Munar et al., 2015; Silvia & Barona, 2009).

The composition of a visual stimulus requires observation on different scales to fully comprehend its structural properties. Gaze and attention to local features in visual information are studied using eye tracking, due to eye muscles activity potentially interfering with the recording of brain waves. Training in the arts shows to profoundly influence gaze efficiency and attention time to local features (Antes & Kristjanson, 1991; Koide et al., 2015; Vogt & Magnussen, 2007). For this reason, studies on structural properties focus on global particularities such as global symmetry, showing consistent results across people. Global symmetry is proven to prime positively aesthetic judgements in subjects across cultures from an early age, suggesting a hardwired mechanism in the visual system at birth (Huang et al., 2018). Pecchinenda et al. found that displaying symmetric patterns for only 75ms was sufficient to significant prime decision positively in aesthetic judgement (Pecchinenda et al., 2014). The brief display of the stimuli can be considered as a proof that symmetry detection is an unintentional behaviour, even though it has previously been argued that symmetry detection does not happen spontaneously. Another study strengthens the hypothesis of spontaneous processing by observing an Event-Related Potential (ERP) component occurring after subjects were exposed to symmetric visual patterns (Höfel & Jacobsen, 2007). The ERP component has also been observed with visual stimuli containing rotational symmetry, despite reflectional symmetry being preferred by the human visual system (Makin et al., 2012). While the results confirm the spontaneous nature of symmetry detection, it also points out that aesthetic evaluation is an intentional phenomenon as another ERP only appears in tasks where the subject is attending to the stimuli. In terms of electrophysiological responses, simple aesthetic contemplation is illustrated by a lateralised late positivity (440-880ms) while aesthetic evaluation showed an additional early frontocentral negativity (300-400ms) (Makin et al., 2012). Evidence of specific sub-processes for symmetry detection also exists at the functional level. The V1 area in the human visual system seems to encode information about the axis or the centre of symmetry, while higher visual areas such as V3a, V4, V7 and the lateral occipital cortex show significant activations when exposed to symmetric patterns (Bona et al., 2014; Sasaki et al., 2005; van der Zwan et al., 1998).

2.2.2. Visual complexity

The distinction between structure and complexity is vague, especially when addressing topics such as structural symmetry due to its direct impact on complexity (Gartus & Leder, 2017). Visual complexity is complicated to define and anchor in a specific domain of research, due to being a concept of a higher, and therefore, more abstract level. A measure for visual complexity takes into account the variety and density of elements on different scales, as well as the variety of these elements' characteristics. Early experiments are rarely reproducible due to subjective ratings based on unknown criteria. Berlyne first establishes a relationship between complexity and ratings of "interestingness" and "pleasingness", hypothesising a threshold where complexity becomes too high for visual information to be aesthetically pleasant (Berlyne, 1970, 1971). The use of subjective human ratings for visual complexity with vague criteria has been heavily challenged, with Rump suggesting that a more accurate vocabulary needs to be employed and potentially forgetting the term "visual complexity" (Rump, 1968). To solve many debates regarding the veracity of Berlyne's findings, Nadal designed an experiment to investigate different meanings of complexity, concluding that arguments regarding the relationship between complexity and preference are due to vocabulary issues (Nadal et al., 2010). While most studies use biased metrics such as average ratings using a jury composed of only a few members, more mathematical and computational solutions are now emerging to improve reproducibility. For instance, image compression algorithms prove to be a reliable indicator of human ratings of visual complexity due to their optimisation of memory by finding patterns and areas of perceptually similar colours in images (Donderi, 2006; Forsythe et al., 2011; Machado et al., 2015; Romero et al., 2012). The use of image compression algorithms is particularly appreciated as they are mostly developed by a third-party organisation focusing on digital images and data compression ratio, and avoiding any potential bias from researchers running experiments. Fractal dimensionality also provides an objective measure of visual complexity. A fractal is a mathematical pattern often observed in nature, such as plants during their growth, but it also emerges in many artworks without intent from the artists (Forsythe et al., 2011; Taylor et al., 2005).

Visual complexity brings new elements into the investigation of human aesthetic judgements. Measures of visual complexity have successfully been used in multiple applied studies looking at first impressions of visual content such as advertising or web page designing (Huhmann, 2003; Pieters et al., 2010; Reinecke et al., 2013; Tuch et al., 2012). With the experiments focusing on specific scenarios, visual complexity appears to grow linearly with cognitive load, and therefore, could give indications of processing fluency (Harper et al., 2009). A phenomenon similar to the affective aspect of a subjective Eureka experience can be observed during aesthetic judgement when recognising familiar shapes in images initially confusing to the observers (Muth & Carbon, 2013). Palumbo et al. demonstrate that complex abstract patterns can make observers underestimate time durations, possibly implying that processing fluency could be influenced by non-artistic stimuli (Palumbo et al., 2014). Processing fluency is the assessment of the ease by which information is assimilated, either on a time or complexity scale. While Reber et al.'s theory suggests that processing fluency is tightly linked with aesthetic pleasantness, little is known about the actual mechanism at the origin of this relationship (Reber, 2012; Reber et al., 2004). Reber argues that perceptual and conceptual fluency are both fed into the aesthetic judgement process, moderated by the expectations of processing complexity of the visual stimuli. In other words, a simple stimulus triggers no pleasure due to the expectation and the actual processing being fluent, but a complex stimulus that is processed fluently is rewarding. Therefore, Reber suggests that aesthetic pleasantness is a purely subjective experience that does not uniquely rely on the perception of the objective features of a stimulus, but also on the conceptual interpretation bringing cultural and personal experience into the decision process.

2.2.3. Familiarity with the visual stimuli and conceptual association

Conceptual processing fluency plays a role in aesthetic pleasantness (Reber et al., 2004). Both perceptual and conceptual fluency prove to be the source of positive emotions, triggering a sense of familiarity towards a visual stimulus (Lanska et al., 2014). Forsythe et al. also define familiarity as the stage in visual processing following complexity estimation and structural understanding (Forsythe et al., 2008, 2011). Familiarity creates a bias in the judgement of complexity, and therefore, indirectly affects aesthetic experiences. Such an impact was exposed by studies comparing abstract and representational artworks. Zeki et al. first addressed the work of influential and successful artist, Francis Bacon, who balanced his work between abstraction and teasing of familiar items such as faces (Zeki & Ishizu, 2013). Interestingly, conceptual stability appears to be preferred over unstable contexts (Muth et al., 2016). The plausibility of scenes and well-chosen titles contribute to aesthetic pleasure (Sammartino & Palmer, 2012).

An object recognition task in paintings using event-related fMRI not only displays activation in higher-level visual regions when looking at representational art, but it also triggers stronger activation in the fusiform gyrus and temporoparietal junction (TPJ), brain regions implicated in object recognition and local and global searches (Fairhall & Ishai, 2008). Another study using transcranial magnetic stimulation (TMS) demonstrates that interfering with activity in the lateral occipital cortex (LOC) leads to a loss in aesthetic appreciation of representational paintings (Cattaneo et al., 2015). Cattaneo et al. did not, however, observe the same effect with abstract paintings. The activation of the LOC during object recognition implies that semantic associations positively influence aesthetic judgement. Nonetheless, it can also be argued that the semantic content of representative paintings is positively biased, meaning that no strong negative emotion could be triggered, for instance by disgusting or violent visual content. Another TMS experiment run by the same research group focused on the involvement of the prefrontal cortex (PFC) in aesthetic judgement of faces (Ferrari et al., 2015). The results indicate that stimulation of the dorsolateral prefrontal cortex before exposure to a face positively primes the participant. Studies investigating brain activity during tasks with representational and abstract visual information allow to pinpoint significant differences in regions involved and potential influences of object recognition and memory on aesthetic judgement.

Of course, it is to be expected that visual preferences in individuals can be altered by their personal and cultural experiences, as well as art training. Silvia attempts to connect the sciences of aesthetics and emotions and argues that emotions in aesthetic experiences are diverse. Consequently, conceptual association is likely to be the stage where the most variance emerge across people (Silvia, 2012). People tend to especially disagree when judging pictures containing controversial or negative semantics (Cooper & Silvia, 2009; Silvia & Brown, 2007). For example, Silvia demonstrates some similar findings regarding preferences in complexity, where art experts have more interest for complex pictures despite expressing similar emotions as lay people (Silvia, 2006). Another study also shows that the strong structural particularity that is symmetry is preferred over asymmetry in the same manner by both art experts and non-specialists in implicit preference tests, but not in explicit statements by participants (Weichselbaum et al., 2018). It contributes to the idea of a robust priming on the visual system, but also that personal experience and training can overwrite this bias to some extent at a later stage. Moreover, the effect of the context surrounding the whole aesthetic experience on the emotional response is still controversial. Experiments taking place in an artistic environment find that positive emotions are either attenuated or stronger, while negative emotions stay unchanged (Brieber et al., 2015; Gerger et al., 2014). With personal experiences being complicated to control in experimental settings, research in aesthetic judgement across different art expertise levels allow to hypothesise that preferences related to perception are the most stable across people.

2.2.4. Modelling aesthetic judgement

Berlyne’s theory suggests that interest and preference grow curvilinearly with the complexity of visual stimuli up to a threshold from where this correlation weakens (Berlyne, 1970, 1971). Perceived complexity emerges from different elements in visual information, which is processed through different stages with some level of parallelism (Zeki, 2015). The different visual preferences that exist at the neural level in the processing of colours, edges, and shapes are overwritten by personal or cultural experiences. Nevertheless, these hardwired preferences appear to positively prime observers in their aesthetic judgements. A recent review by Che et al. puts the common view of visual preferences across cultures in perspective, with empirical studies strongly indicating that perceptual preferences shape individuals’ aesthetic preferences (Che et al., 2018). Moreover, Che et al. emphasise the robustness of perceptual preferences by mentioning how such preferences are not only shared across humans, but also across animals, and especially mammals. As put forward by Silvia, research on the relationship between aesthetics and emotions is still, similarly to neuroaesthetics, in its early stage (Silvia, 2012). Silvia underlines that the effect of both positive and negative emotions should receive more attention. While Silvia's argument is valid, adopting a bottom-up approach and evaluate the influence of perception on aesthetic experiences may be a priority based on the previous suggestion by Che et al. (2018). For modelling purposes, factors of aesthetic experiences with the smallest variation across populations are essential to successful development. It is necessary to isolate the different elements even further and simplify the problem for further investigation in human aesthetic judgement. Considering that aesthetic contemplation and judgement can both be identified as creative activities, it can be assumed that they can be classified according to Melvin Rhodes' 4 Ps of creativity: Person, Process, Product and Press (which stands for the context) (Rhodes, 1961). Applying such segregation of the different components involved in the aesthetic judgement decision process strengthens previous conclusions. Indeed, the product can be assessed through direct visual observation and computer vision descriptors, and the process has been investigated with great progress being been made in the neuroscience and cognitive of aesthetic experience. However, the role of the person and the surrounding context are difficult to dissociate from each other, mainly because they vary hugely.

With evidence pointing at a strong relationship between complexity estimation at the beginning of the aesthetic judgement process and the final decision, Graf et al. introduced an estimation of processing complexity of the visual system in their fluency-based model under the name of "fluency estimation" (Chipman, 1977; Forsythe et al., 2008; Graf & Landwehr, 2015). It implies that when first attending a visual stimulus, the human visual system estimates the resources required for processing. If the stimulus is processed more fluently than expected, for example, because of a specific structure in the stimulus, the system can produce a reward. Nonetheless, it is important to note that the complexity estimation theory has little empirical support from neuroscience.

In the model proposed by Redies, the physical and cultural environment of the subject are represented (Redies, 2015). The model is especially interesting for its representation of the whole process as a gradient shared between perceptual and cognitive processing, respectively representing the processing of the form and semantic content in visual stimuli. Chatterjee also made it a pivotal point of its model of aesthetic experience, labelling the main three factors as sensory-motor, knowledge-meaning, and emotion-valuation (Chatterjee & Vartanian, 2014). This categorisation emphasises the full body experience that is provided by aesthetic pleasantness. However, as discussed previously, influences caused by personal and cultural experiences can lead to an increasing number of different factors to take into account.

Despite all the visual processing happening prior to aesthetic appreciation, the latter is still heavily investigated to pinpoint a possible brain area weighing the many factors to reach a decision. As a whole, experiments in this emerging domain are rarely consistent with each other and consequently, it is difficult to make definite conclusions. However, it is suggested that the results are complementary and that the lack of consistency is due to subtle differences in vocabulary in the tasks, such as asking participants to judge whether a visual stimulus is “beautiful or not beautiful” or “beautiful or ugly” (Cela-Conde et al., 2011; Nadal et al., 2008). For instance, Munar et al. observed significant differences in brain activity 400ms after exposure to “beautiful” or “not beautiful” visual stimuli (Munar et al., 2012). On a neurobiological level, Kawabata and Zeki found differences in activation of the Orbito-Frontal Cortex (OFC) and the motor cortex depending on whether paintings are qualified as “beautiful” or “ugly” (Kawabata & Zeki, 2004). The results were confirmed by Ishizu and Zeki’s study, which revealed additional subcortical regions but came to the conclusion that the role of these areas is not straightforward to determine, due to processes such as decision-making, judgement, and aesthetic experience being tricky to dissociate (Ishizu & Zeki, 2013). Also studying aesthetic experiences triggered by paintings, Vartanian and Goel obtained results showing reduced activity in the right caudate nucleus as preference decreased, while the bilateral occipital gyri, left cingulate sulcus, and bilateral fusiform gyri displayed stronger activations as preference increased (Vartanian & Goel, 2004). Therefore, it is suggested that these brain areas are involved in the evaluation of stimuli with emotional content and potential rewards.

To conclude, this section has for main role to define the most stable sub-processes contributing to the final decision in aesthetic judgement in order to develop computational models. While an increasing number of studies on complexity, emotions and aesthetic judgement show influences on the final outcome, findings regarding preferences related to early visual processing appear more consistent across the different areas of study. However, it is essential to note that the perception of motion and depth were both intentionally left out from this review to focus exclusively on two-dimensional static images.

2.3. Computational processing of visual information

This section introduces traditional and more recent computational tools used in computer vision. The goal of this section is to establish the general direction of advances in computer vision, before developing on feature extraction and machine learning algorithms used in their entirety or as inspiration in this thesis. A strong focus is brought on the efficiency and sustainability of solutions in a domain under pressure of an exploding number of technological innovations.

2.3.1. Image Processing

With computing power respecting Moore's law and internet speed growing at a steady pace, digital visual content is more present than ever in everyday life. From live video streaming on websites such as Twitch.tv to real-time video alteration on smartphone application such as Snapchat, not only the distribution of visual content has reached quality levels where further improvements are barely perceivable to the naked eye, but the content can also be transformed in real-time. Image processing and computer vision as a research domain have incredibly expanded with ever more complex and critical problems given to machines. When selecting an algorithm for image processing purposes, three main factors are taken into account: image pixel size, computer power available and task complexity. In many scenarios of image processing, information is extracted from pixels individually or by applying filters or templates on pixel areas. While no norm has been established regarding the size or format of digital photographs, a good indicator of the number of pixels in digital visual content and potential datasets mined on the internet is the maximum video quality offered by the streaming website, YouTube. Before activating videos of higher quality on the website, YouTube has to evaluate the potential level of adoption. For example, it must consider the number of video content creators who would produce in such resolution, but also whether the average broadband can support this level of data transfer and the average computer can display the video. As shown in Figure 2.1, the maximum number of pixels in 2005 is dwarfed by the current format. Despite the incredible variations possible between the different visual media, evaluating the order of magnitude of the average number of pixels in visual content online can help to anticipate the effort required to process visual content.

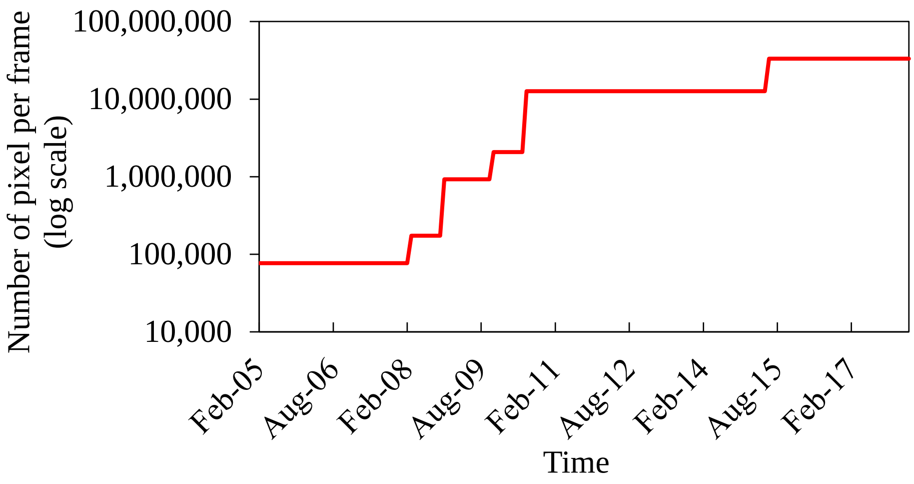


Figure 2.1: Number of pixels in the highest video quality available over time on the video streaming website, YouTube (data retrieved from http://en.wikipedia.org/wiki/YouTube, 2018).

According to Moore's law, the cost of making transistors for computational units is set to reduce significantly over time, making computing power grow in an exponential manner (Moore, 1965). The prediction made by Gordon Moore, co-founder of Intel Corporation, appeared to fulfil itself until more recently. As the exponential growth of the number of transistors in Central Processing Units (CPU) can be observed in Figure 2.2, the exponential trend has only been saved by the development of multi-core CPUs despite the actual single-core frequency losing momentum (Rupp et al., 2018). As the average number of pixels in images climbs at a seemingly exponential rate, a possible end to the trend respecting Moore's law could cause severe processing issues. Despite both predictions showing risks of plateauing on the long term, it does illustrate how the design of computer vision algorithms should take into account this relationship and focus on the efficiency and scalability of visual information processing to keep the whole computer vision ecosystem sustainable on the long term.

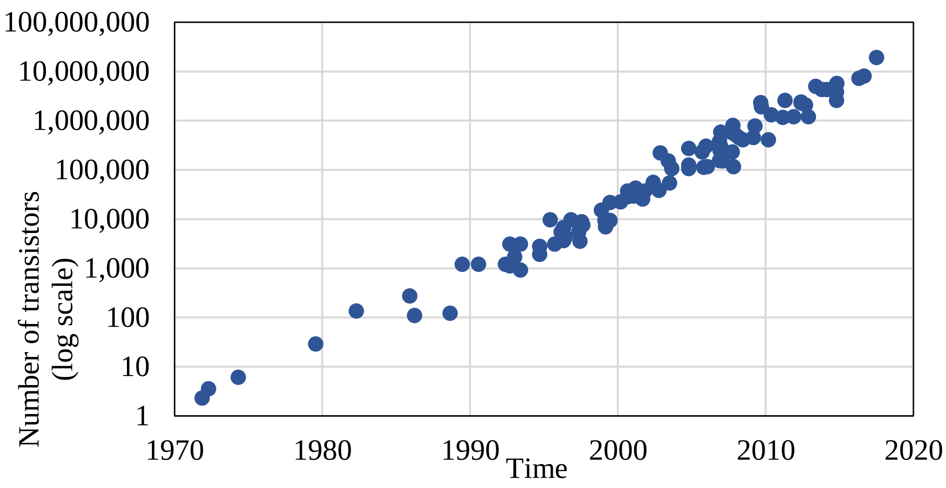


Figure 2.2: Number of transistors per CPU since the 1970s. (Rupp et al., 2018)

2.3.2. Visual Feature Extraction for Aesthetic Measures

Many machine learning and statistical studies develop their own handmade features to use as metrics for a specific task. The metrics are usually inspired from different domains of research such as humanities or behavioural sciences. The main advantage of such approach is that it allows to reduce considerably the dimensionality of the visual information being processed. First stated by Bellman, the curse of dimensionality is a recurrent problem in machine learning assuming that data quickly become noisy when working with a high number of dimensions, which then requires to increase the size of training sets (Bellman, 1957).

For example, Datta et al. developed a straightforward score for light exposure in photographs by averaging thousands of values representing the brightness of coloured pixels in an image into a single score (Datta et al., 2006). In another example, Lo et al. proposed to score the “color palette” by using a weighted k-means clustering algorithm to extract the 5 dominant colours, before scoring it by measuring the closeness of the dominant colours with photographs in the training set (Lo et al., 2012). While much more complex aesthetic measures exist and attempt to be as loyal as possible to the rules of aesthetics, such scoring systems can be biased by the authors of the algorithms (Romero et al., 2012; Sun et al., 2018). Indeed, rare are the rules of aesthetics purely and strictly based on a mathematical formula (Casalderrey, 2017). The grey areas offer freedom of interpretation, which may lead the developer to feed their cultural and personal experience into the algorithm. Moreover, using such methods to avoid the curse of dimensionality and make machine learning easier can significantly affect performances in the other way due to potential oversimplification of the information.

Regardless of the performance of such methods, Romero et al. (2012) display an evolutionary process to develop and select measures of complexity for aesthetic classification. The paper exposes the results of different possible measures of complexity but also investigates the gain provided by each component to the final results. After ranking the many features by their potential information gain using a Support Vector Machine, the paper goes further and highlights the information overlap of the different features by showing how coupling many features provide minimal performance improvements.

2.3.3. Advanced Feature Extraction Algorithms: Visual Descriptors

On the other end of the spectrum of visual feature extraction, many algorithms have been designed to retrieve as much information as possible regarding low-level visual features. Most of the well-known feature extraction algorithms, such the Scale Invariant Feature Transform (SIFT) or the Histograms of Oriented Gradient (HOG), have initially been designed for object recognition purposes (Dalal & Triggs, 2005; Lowe, 1999). The sets of features extracted by these algorithms are qualified as visual descriptors, as they define basic characteristics such as shape and colour for instance. SIFT marked the beginning of a new era in computer vision with a series of efficient feature extractor for images in uncontrolled environments. Indeed, SIFT aimed at providing robust feature detection independent from translation, scaling and orientation with reasonable flexibility for noise and illumination. SIFT uses the Difference of Gaussians to efficiently highlight major features in different scale spaces. The Difference of Gaussians is achieved by subtracting two versions of the same image with different levels of Gaussian blur. After noise-cleaning and locating keypoints, each keypoint is characterised by its estimated orientation and gradient magnitude, both contained into what is called a visual descriptor. With a number of keypoints on the order of the thousands, SIFT has demonstrated an ability to correctly match 78.6% of features in images even after distortions, orientation changes and added noise (Lowe, 1999).

Over the years, many other algorithms using the keypoint approach have built onto the base provided by SIFT and improved on robustness and efficiency, such as SURF (Speeded Up Robust Features) and more recently ORB (Oriented FAST and Rotated BRIEF – two other algorithms) (Bay et al., 2006; Rublee et al., 2011). With datasets reaching millions of images, Puna Turcot and SIFT's creator, David Lowe addressed efficiency issues in the representation of the extracted information and proposed a method to reduce the number of local visual features extracted per image (Turcot & Lowe, 2009). Following a bag-of-word framework, the method consists of combining local visual features to define "visual words" which have their usefulness assessed by their frequency of occurrence in images of the dataset.

With keypoint-based approaches dominating the domain of image recognition, the Histograms of Oriented Gradient (HOG) algorithm provided an alternative to the traditional use of visual descriptors and was initially developed to detect humans. HOG is based on a simple design where an image is considered as a grid of pixels, itself divided into 16x16-pixel sub-areas called "blocks" used for local illumination normalisation. Each block contains four 8x8-pixel sub-areas called "cells", and within each cell, an estimated orientation is given for every pixel using the grayscale difference of its surrounding pixels. A histogram with a pre-defined number of bins is then processed to visualise the different categories of orientations present in the cell.

Despite its simple design, HOG outperformed keypoint-based solutions in human detection tasks due to its grid covering the whole image, preventing any error in the detection of keypoints. In a human detection scenario, HOG also has the advantage of being sensitive to orientation and proportion (Dalal & Triggs, 2005).

2.3.4. Binary classifiers: Support Vector Machine & Multilayer Perceptrons

The extraction and scoring of visual information is the first step in many computational aesthetics experiments. The extracted information then needs to be fed into machine learning algorithms to learn its relationship to aesthetic preferences. SVMs (Support Vector Machine) and Multilayer Perceptrons are among the most popular supervised machine learning algorithms used for binary classification tasks. Both algorithms owe their popularity and adoption to the wide range of libraries which include them such as scikit-learn for Python or WEKA for Java. Their global adoption is also due to their ease to train and run, making them perfect for building prototypes. Most current computers can run these algorithms even in the greediest settings, with the only possible limitation caused by the time required for training.

Cortes and Vapnik first introduced SVM in 1995 as a method to differentiate two sets of items (Cortes & Vapnik, 1995). In the simplest case, data points belonging to two distinct classes are mapped in a dimensional feature space and the SVM algorithm attempts to find the most optimal hyperplane to segregate the two sets of items. The algorithm searches for the data points in each class which are the closest to each other in the dimensional space to calculate the support vector, creating a border in the space where the decision is the most critical and ambiguous. While SVM initially assumes that the two sets of data points are linearly separable and maximises the margin between the decision hyperplane and the two sets of data points, it can also adopt a soft margin hyperplane where it calculates the decision hyperplane with the least errors using the hinge loss function. To solve non-linear problems, the "kernel trick" was developed to transform the input data into a higher-dimensional space (Boser et al., 1992). As illustrated in Figure 2.3, while the data points cannot be linearly separated in the input space, a maximum-margin hyperplane can be found after transformation into a higher-dimensional feature space, despite this same hyperplane not appearing linear when represented in the original input space. SVMs can also be tuned using regularisation parameters aiming to improve generalisation. The regularisation parameters allow to modify the amount of allowed misclassification in the training phase and the width of the margin surrounding the decision hyperplane.

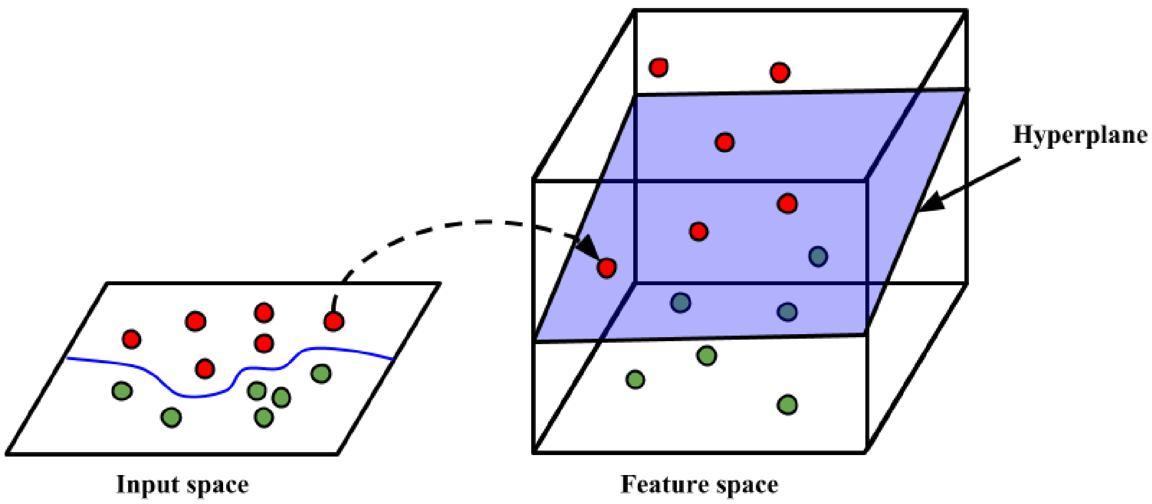


Figure 2.3: Representation of inputs and the transform into a high-dimensional space.

Following a connectionist approach, the multilayer perceptron (MLP) is often considered as the most basic working solution of neural networks. MLPs are composed of at least three interconnected layers of artificial neurons (also called units), where input and output layers are connected by one or more hidden layers, as represented in Figure 2.4. Each neuron on a layer is connected to all neurons of the following layer, and each connection is characterised by its strength, called weight. Training of such architecture was made possible by the development of the back-propagation algorithm (Rumelhart et al., 1986). The algorithm allows the network to learn by feeding data with an expected outcome and spread back potential errors across the layers during the training phase to tune the weights of the connections between neurons. Thanks to the use of hidden layers, a multilayer perceptron can model non-linear relationships, which is not possible with single-layer perceptron.

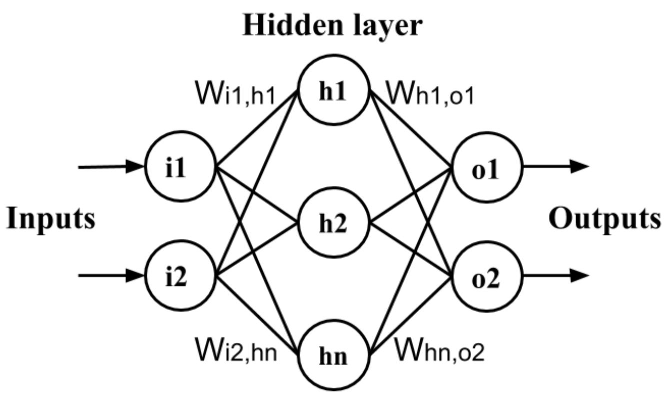


Figure 2.4: Basic architecture of a multilayer perceptron.

SVM and MLP have continuously been opposed and compared with respect to their theoretical principles, but no winner has ever been truly declared (Collobert & Bengio, 2004; Suykens, 2001). While SVM is praised for always finding the global minimum solution and being accurate with high-dimensional data, MLP is seen as a more practical solution due to the fixed size of the models with a small number of neurons and, therefore, being a more scalable (Osowski et al., 2004).

2.3.5. Deep Learning

Deep learning is a subfield of machine learning attracting the most attention in the last decade by overtaking traditional methods in pattern recognition contests such as MNIST (Schmidhuber, 2015). Moreover, the academic community contributing to research in deep learning is extraordinarily active on social media and conferences, making the field expand at an even faster rate. Emerging from a combination of larger datasets, growing GPU processing capabilities and machine learning innovations allowing neural networks of a much larger scale to learn, deep learning englobes several types of deep neural networks with different architectures and learning methods. A deep neural network can adopt an architecture as simple as a multilayer perceptron with more than one hidden layer of units. However, even in its most basic form, deep neural networks can include a high number of regularisation strategies, which helps to avoid overfitting the data and by consequence, improve generalisation (Goodfellow et al., 2017).

While some regularisation methods are not exclusive to deep learning, their use is nowadays more critical in this domain due to the machine learning algorithms previously mentioned now being mainly used for prototyping. One of the most straightforward regularisation techniques, called early stopping, consists of continually making a copy of the most performant configurations on the validation set during the training phase. The reason for such practice is that excessive training has been shown to lead to overfitting of the training data, causing performance loss on the final test set. Speaking of training data, due to the facilitation in processing large datasets containing millions of entries, it has also become common practice to generate additional training items to improve generalisation. New entries can be created by adding noise to the inputs such as, for visual information, create symmetric images or modify the colour settings of the original image.

Other regularisation methods are implemented within the cost function which allows to calculate the errors between expected and actual outcomes. The cost function is an indicator of the progress made during training and is aimed to be as small as possible. For instance, the L2 parameter penalty, also called weight decay, allows to reduce the weights' magnitude and keep them as close as possible to a given value, which is zero in the average case. L2 regularisation is achieved by adding the square of the sum of all weights to the cost function. Considered as having similar effects as early stopping, L2 regularisation limits weight divergence by forcing them to stay as small as possible and therefore, reduce influence from potential noise. Another method, called L1 regularisation, uses a similar approach and adds the sum of absolute values of the weight to the cost function. While it may appear to have the same effect as L2 regularisation, L1 regularisation shrink the weights by a constant amount at each step while L2 regularisation shrinks them proportionally. Therefore, L1 tends to highlight the most robust connections and select the most relevant inputs, but it can also cause some weights to reach an absolute zero and not reach the most optimal solution.

One of the most recent regularisation techniques developed, the dropout algorithm consists of randomly removing a chosen percentage of units in the input and hidden layers at each training step (Srivastava et al., 2014). As represented in Figure 2.5, selected units have their weights set to zero for the given training step. Such process makes it possible to train the ensemble of all possible subnetworks within the trained neural architecture, improving generalisation.

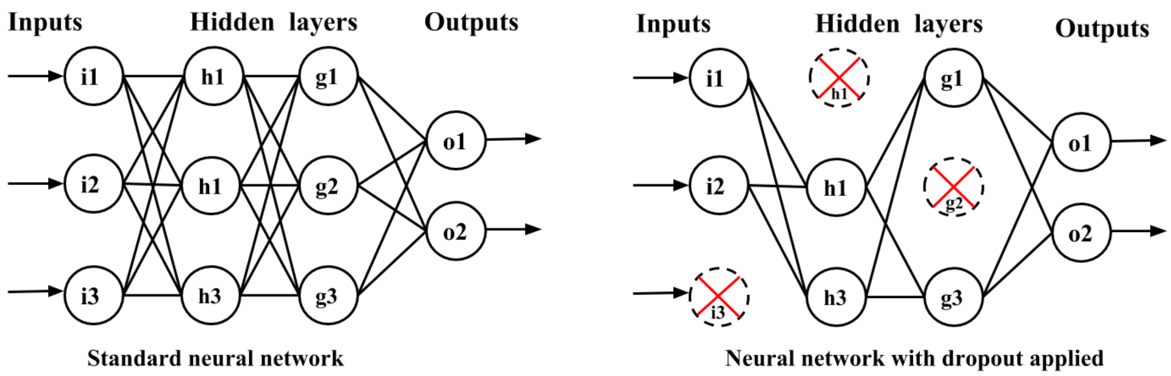


Figure 2.5: Representation of a neural network in its standard setting and with dropout applied.

With the constant flow of innovations, deep learning has made it easier to learn from a gigantic amount of data with a high number of features. It has led to a regain in popularity of the previously developed convolutional neural networks (CNN), which have become unprecedently dominant in the domain of computer vision. By taking inspiration from simple and complex cells, Fukushima had already introduced the concept of convolution and average pooling into a machine learning model called the neocognitron (Fukushima, 1980). The model was the first to implement a receptive field coupled to a filter moving across it to learn visual patterns. At the time, the concept did not spread due to the lack of direct usability, similarly to deep learning in its early years. However, Yann LeCun successfully revived the idea by publishing a large dataset for handwritten digit recognition and establishing new state-of-the-art systems, showing the potential and reliability of CNNs for automated postcode reading (LeCun et al., 1998, 1999, 1995).

CNNs allow to process a high number of inputs while keeping some level of spatial information, which makes it perfect for image processing as a raw image can count millions of pixels, each represented by 3 values (RGB or HSB in most cases, and therefore shaping a 3-dimensional input). CNNs take their name from the use of convolution to process information in at least one layer. A typical CNN layer is composed of an input layer, a convolutional layer, and a pooling layer.

Taking the example of an image, the input takes the shape of a 3-dimensional volume (width, height, and number of channels). A finite number of filters, also called kernels, is applied on the input volume as shown in Figure 2.6. Each kernel computes the dot product using the values of the filter and the input for each local area (including in depth) across the width and height of the input. Therefore, an activation map is created for each filter. The weights of each kernel are tuned during training, so the most relevant features are highlighted when filtering information. Such process leads to neural connections in CNNs having sparse interactions, unlike traditional neural network layers which are fully connected. As displayed in Figure 2.7, a given neuron on the first layer only propagates its information towards a certain number of neurons of the following layer. However, neurons on deeper levels still indirectly receive information from the first layer, meaning the middle layer have some spatial information from the input layer while feeding the information through the network in a more efficient manner.

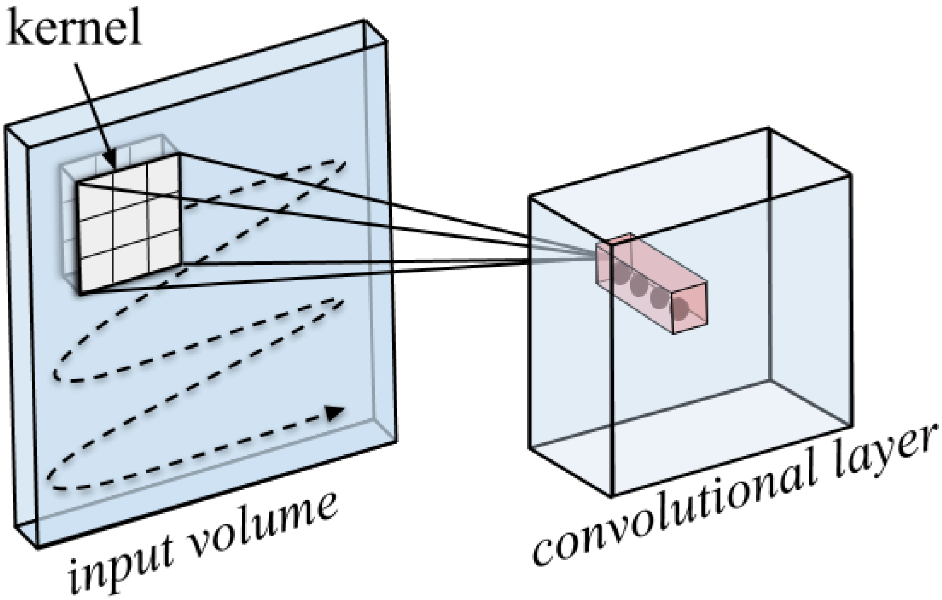


Figure 2.6: Kernel processing the input volume into the convolutional layer.

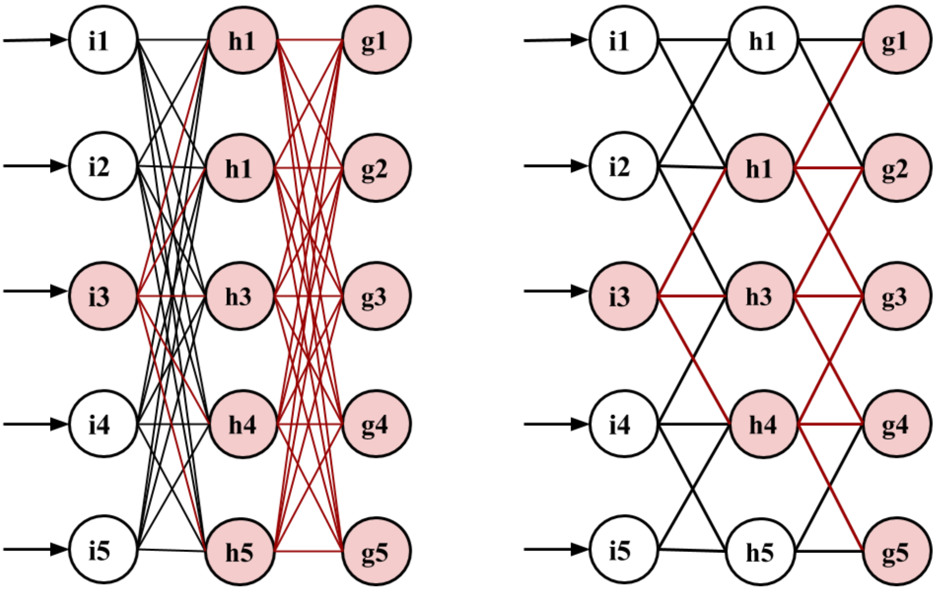


Figure 2.7: Differences in connections between a traditional fully-connected neural network (left) and a convolutional layer using a kernel of width 3 (right).

Another key feature of CNNs is the pooling layer which allows downsampling the information with minimal loss. It applies max pooling filters on the previous layer where each filter only keeps the input with the highest value from the previous layer. In most cases, several layers of convolutional and pooling layers are set to extract features before feeding them into a traditional fully-connected layer of neurons.

To conclude, this overview exposes the robust mechanisms of deep learning and its most important architecture for computer vision. While deep learning brings efficient and performant solutions, the learning process is complicated to investigate, which can turn out to be problematic in a context where one of the objectives is to learn about relationships within the training data.

Chapter 3

Theoretical and Conceptual Frameworks

As developed in Chapter 2, distinct trends are progressing in the domain of study of human visual preferences. This chapter establishes typical characteristics of these current trends before comparing the approaches. The position of the thesis in this environment is then discussed by introducing its framework and methodology.

3.1. Trends in the study of human visual preferences

In order to compare the past and current investigations in human visual preferences and understand the evolution of the field of study, it is necessary to highlight common characteristics due to the wide range of disciplines involved. To identify and compare the different trends, it is proposed to represent them as evolving in a two-dimensional space, with one dimension being fidelity to empirical knowledge of the human visual system and the second dimension being generalisability of the findings.

The first dimension is the level of fidelity to the knowledge of the visual system. For instance, the domain of neuroaesthetics in its early years represents one of the extremes on this axis as its initial objective was to link the idea of “beauty” to empirical studies on the human visual system. As first demonstrated in the study by Kawabata and Zeki (2004), brain regions can be associated with aesthetic judgment. While such an association is established through rigorous methodology, it suffers from similar limitations as traditional neuroscience experiments. The interpretation of results relies heavily on the previously known roles of brain regions, leading to the conception of neural and cognitive models which can overfit the previously observed phenomena. As mentioned in Chapter 2, Nadal et al. (2008) mention that the lack of reproducibility in neuroaesthetics could be simply caused by the vocabulary used to formulate the tasks composing experiments. If the directives for a task can impact the findings to such an extent, it appears risky and too early to base new models on neuroaesthetics findings. Therefore, it can be argued that sub-domains of neuroscience can be speculative in their early years due to the low numbers of tested scenarios and that models issued from neuroaesthetics experiments are still fragile. Hence, the conception of a model of aesthetic judgement may benefit from relying solely on fundamental knowledge in the neuroscience of vision instead of more novel but also less verified findings.

At the opposite end of the axis representing fidelity to neuroscientific knowledge, many works such as Reber at al. (2004) develop theories accentuating on the differences across people and the subjectivity of aesthetic experiences. As part of the aesthetic judgment process is subjective due to cultural and personal experiences, it is currently impossible to investigate such differences at the neurological level. When studying the subjective aspect of aesthetic appreciation, discussions often focus on individuals and address renown artists with exceptional aptitudes to create aesthetically appealing works of art. While some experiments may be able to pinpoint neurological mechanisms observed to help artistic creation, there is still no way to objectively and quantitively assess the impact on aesthetic judgment during visual art creation of a well-developed brain area or a condition such as dementia (B. L. Miller et al., 1998; Z. A. Miller & Miller, 2012). Even though Semir Zeki contributed to the first experiments with strict neuroscience methodology, he also is the main author of published essays exploring the links between artists and neuroscience by attempting to reverse-engineer the brains of selected artists from their works of art (Zeki, 2002; Zeki & Ishizu, 2013). This top-down approach aims to determine brain areas that may have a role in artistic creation. It is, however, not relying on any quantifiable data from the selected “subjects”, making the conclusions of such study weak from a neuroscience perspective despite having some philosophical value. Zeki explains that the studied artists do not appear to appeal to the people’s visual preferences but to emotions such as romantic love or conceptual ambiguity. This shows that, while there is no existing framework to study human visual preferences during creative processes, essays about gifted individuals may lead to hints about the motivation behind works of art and potentially link concepts to aesthetic pleasantness.

The second proposed dimension for studies of human visual preferences is the generalisability of findings and proposed models. Whether it is an experiment, a cognitive model or a computational model, the usefulness of the outcome to feed theories of aesthetic appreciation varies greatly. In a domain where objectivity is required to move forward, an individual’s subjective judgment cannot be ignored, as much as the context of the task. Ramachandran et al. (1999) aimed to link neurological mechanisms with philosophical theories and observations on aesthetics to establish general rules. While such approaches focus exclusively on perceptual preferences of the human visual brain, more recent cognitive models include concepts of emotions and pleasure (Chatterjee & Vartanian, 2014; Che et al., 2018; Graf & Landwehr, 2015; Redies, 2015). Indeed, these models take into account that contextual, cultural and personal backgrounds can overwrite the final decision in aesthetic appreciation, making them a much more robust and generalised model (Weichselbaum et al., 2018). In terms of computational modelling, implementations of the measure of visual complexity prove to be successful at determining aesthetic goodness in visual information across many scenarios. Despite being a higher-level measure with no empirical basis of such process in the human brain, measures of visual complexity such as the JPEG compression algorithm indirectly include elements recognised as predictors of good aesthetics (colour repetitions, entropy/novelty, etc.) (Forsythe et al., 2011; Romero et al., 2012; Wallace, 1991). Visual complexity has not only proven to be a great indicator aesthetic pleasantness in computational experiments but it also appears to indicate aesthetic changes in visual art creation over time (Forsythe et al., 2017; Romero et al., 2012). It implies that it is a reliable tool as an objective measure of aesthetic pleasantness despite lacking neuroscience evidence.

On the other hand, Zeki explored, once again, another facet of the domain by considering aesthetic experiences as a human behaviour in response to a very specific situation or objects such as exposure to paintings or mathematical formula (Kawabata & Zeki, 2004; Zeki et al., 2014). Other neuroaesthetics studies have been targeting, for example, the very specific phenomenon of aesthetic appreciation of architecture (Ma et al., 2015; Vartanian et al., 2013). While these examples can be deemed scientifically accurate, they teach us more about the interactions with a particular type of stimuli than about aesthetic judgement as a phenomenon. These types of studies have been criticised because of their narrow view of aesthetic experiences and their results being difficult to interpret (Marin, 2015). Interestingly, a majority of the computational models of aesthetic experiences can be classified as adopting a similar narrow vision due to the lack of cross-dataset or cross-media tests. Paradoxically, some computational models fully assume the very specialised training process, claiming that the aesthetic pleasantness of visual features is strictly context-dependent (Kao et al., 2017; Simond et al., 2015; Tian et al., 2015). By consequence, it implies that these models are highly dependent on their semantic recognition performance.

To conclude, the presented analysis of the trends in investigations on human visual preferences utilises fidelity to neuroscience knowledge and generalisability of the outcomes as two dimensions to locate the works in the space and compare the different approaches. When analysing the works at both extremes of the two dimensions suggested, it can be observed that neuroscientists and cognitive scientists advance on all fronts with experiments and model suggestions. However, there is a striking lack of computational models with plausible neuroscience inspiration and demonstrated learning of general human aesthetic preferences. Indeed, most models are specialised for particular contexts or used semantic recognition to improve aesthetic classification scores. Such evolution in the domain is explained by the fact that only one large-scale dataset for aesthetic classification exists, while datasets for semantic recognition as ImageNet or YouTube-8M profit from millions of entries (Abu-El-Haija et al., 2016; Deng et al., 2009). The only area which links the experimental study of human visual preferences to computational models is the implementation of visual complexity but it still lacks a strong neuroscientific basis. Therefore, it can be concluded that the domain of research in human visual preferences severely miss works which would include theoretical knowledge into models, which would allow reverse-engineering and feed novel approaches to neuroscience.

3.2. Situating the thesis in the domain of study of human visual preferences

The field of computational aesthetics has attempted to find visual features or machine learning systems allowing to classify images depending on aesthetic criteria. The outcomes of such works mainly consist of tools to sort or automatically retouch photographs, but there is no major takeaway issued from computational aesthetics in terms of machine learning knowledge. Moreover, very little work has been done to connect the outcomes to neuroscience and psychology knowledge of aesthetic appreciation, underlining that computational aesthetics has not shown any major contribution to the study of human visual preferences.

The work presented in this thesis places itself in the lineage of the initial works in neuroaesthetics due to its inspiration from the neuroscience of vision to investigate human visual preferences. The computational experiments developed in this thesis can individually be considered as part of the field of computational aesthetics due to its heavy use of machine learning techniques and large-scale data analysis. When considering the thesis as a whole, its aim is to link empirical knowledge about human visual preferences with computational aesthetics in order to acquire feedback for future neuroscience and cognitive models. Furthermore, multiple experiments will be set to validate that the aesthetic preferences learnt by the proposed computational models represent actual human visual preferences and do not solely portray preferences for specific photographic features or a subset of visual stimuli. On a machine learning point of view, the presented work will differentiate itself from previous works in the domain based on its biological inspiration, with hopes that the biological basis leads to great efficiency. Indeed, defining a set of features to extract relevant visual information for aesthetic judgement may provide better efficiency in terms of machine learning. Moreover, it offers the possibility to control the inputs and make sure that the models strictly learn from the low-level visual features, which also represents a risk of overall performance loss.

This presented work is taking a position on several levels, it first takes a theoretical ground in neuroaesthetics and neuroscience of vision to define low-level visual features known for being preferred by the human visual system. Limiting the machine learning techniques to existing algorithms is necessary to focus on a set of brain-inspired features which questions the origins of aesthetic judgment. Nevertheless, going against the trend of using convolutional neural networks to solve computational aesthetic problems may allow to challenge its dominance. Once the aesthetic classifier tested against existing works, its cross-media capability will allow to investigate whether aesthetic preferences can be generalised across different types of visual stimuli. To finish, an experimental framework will be set using the set of features to analyse aesthetic changes in the careers of visual artists. It is hoped to develop new measures characterising changes in a similar way as visual complexity measures but with ease of interpretation thanks to its stronger inspiration from the human visual system. All these positions are illustrated in the following questions that will be answered over the course of this thesis:

1. Can human aesthetic preferences be learnt from rated photographs retrieved from the internet?
2. Can an aesthetic classifier using a pre-defined set of low-level visual features rival traditional classifiers retrieving and learning from all available information, including contextual metadata and semantic information?
3. Beside photography, do other media offer a possibility to assess the completeness of an aesthetic classifier's features?
4. Are aesthetic preferences in the visual domain during contemplation and creation comparable?

3.3. Methods

3.3.1. Standard practices in computational aesthetics

As mentioned in Chapter 2, while research in human visual preferences has been developing for centuries, computational models making direct use of experimental data to learn and attempt to predict aesthetic judgement is still a novel approach. Indeed, only a handful of public datasets have been made available and widely used. The first published and used dataset for computational aesthetics was created by Datta et al (Datta et al., 2006). The dataset contains over 20,278 images and is issued from a community-driven website focussing on photography named photo.net. In this dataset, each photograph is provided with the average user rating of aesthetics on a scale from 1 to 7, and the corresponding rating distribution. Due to the number of images being too small to train state-of-the-art machine learning models, images are segregated between two classes to simplify the prediction problem. As machine learning algorithms can be trained either for regression or classification, classification is usually chosen for new or hard problems as it allows to divide the data into specific subsets making the task noticeably easier. For example, Datta et al. (2006) demonstrate that picking subsets of images from the very top and bottom of the rating distribution improves classification performances significantly. The creation of large-scale datasets such as the Aesthetic Visual Analysis (AVA) dataset has significantly increased performance but no standard has been established for subset selection in classification tasks (Murray et al., 2012). The difference in subset selection among the current works leads to future work requiring a throughout test over several subsets to afford comparisons. In the dataset created by Tang et al. (2013) called CUHK, images are not classified through data mining over the internet but by making members of a pre-defined jury label images individually. While it means that the task of aesthetic judgment is more controlled, it makes the data collection process less scalable, as, for example, only 10 reviewers were questioned to build the CUHK dataset. Moreover, the jury was asked to classify images between two classes unlike in previous datasets, which matched the standard binary classification task in computational aesthetics. It is, however, more complicated to obtain a reliable representation of human visual preferences with a small-sized jury using a binary scale.

The size of the dataset also has a key role when building a machine learning model. In machine learning, the standard practice is to split the dataset between three distinct parts, the training set, validation set and test set (Alpaydin, 2009). In the learning phase, the training set is used purely to teach the machine learning system to recognise patterns while the validation set is used to calibrate hyperparameters and to evaluate whether any additional learning step could improve performances. Once the hyperparameters are chosen and the system reached its maximum performance, it is finally assessed over the test set which had been kept aside. Splitting a dataset to create the three sets is complicated as a balance has to be found between the learning potential of a larger training set and the improved representation of the data from larger validation and test sets. Reasonably large datasets, such as the MNIST handwritten digit database, provide a great number of samples in regards to the complexity of the problem, allowing to keep a representative chunk of the data for validation and testing without impacting the learning process (LeCun et al., 1998). A simple separation of the dataset is not advised for smaller datasets due to risks of poor performance or unrepresentative sets. A common practice is to use a technique called, *k*-fold cross-validation, which consists of randomly shuffling the data before sampling separate test sets by splitting the dataset into *k* equal groups (Kohavi, 1995). For each of the *k* groups, the leftovers from the test set selection are used to train the model before evaluating it over the corresponding test set. After repeating the operation for each group, all the scores are averaged, therefore covering the entirety of the dataset and ensuring that the results are not due to simply overfitting a model to a small subset of data. Due to the small size of Datta et al.'s dataset in comparison to the current machine learning standards, a cross-validation will be used (Datta et al., 2006). In regard to the AVA dataset, the 255,530 images are considered to be a large enough sample and, therefore, a simple percentage split of the dataset should suffice to provide representative results (Murray et al., 2012). As these datasets are finite and used as benchmarks for particular tasks across the field, selecting the right testing method is, therefore, key to obtain an outcome easily evaluated by peers.

Traditional computational models such as multilayer perceptrons or SVMs will be trained and tested using the implementations from WEKA, a user-friendly data mining software with a great number of machine learning algorithms pre-implemented and easily tuneable (Hall et al., 2009). Deep learning models will be developed using Tensorflow, a machine learning library specialised in deep learning and compatible with several programming languages such as Python (Abadi et al., 2016). The use of a specific library is necessary due to memory optimisation issues caused by a large number of units in deep learning models, as well as the huge amount of data required for training. Interestingly, no standard has been defined regarding specific hardware requirements for a model so future works can offer a comparison in terms of computational greed, either it is in terms of time, parallel-process potential or memory usage.

3.3.2. Statistical analysis

To accentuate the disruption with previous works in computational aesthetics, the set of low-level features extracted from the studied images are compared between the two aesthetic classes created within every used dataset. By studying the difference in presence of low-level features, it helps to improve the understanding surrounding aesthetic judgment and how difficult of a task it is as a computational problem. Many works have studied rating distributions or the impact of semantic content on rating but machine learning-based models for aesthetic prediction have never attempted to quantify the difficulty of the task from a human point of view (Kong et al., 2016; Murray et al., 2012; Tang et al., 2013). For every dataset used in this thesis, images are classified into two categories and the difference in presence of each extracted low-level feature between the two classes is tested for statistical significance. For example, images in datasets for aesthetic classification will be segregated into, either, aesthetically poor or aesthetically high. It is important to note that these two classes will be generated without selecting subsets or excluding any outliers. Welch's t-test for unequal variances will be used to verify whether the means of each class are significantly different, on the condition that the resulting p-value is below 0.05 (Welch, 1947). Taking into account the fact that the datasets used in this thesis contain thousands of items and can be considered as large, normality can be assumed as stated by the Central Limit Theorem (Lumley et al., 2002). The statistical test is supported by Cohen's d to evaluate the effect size of the studied phenomena (J. Cohen, 1988). Following Cohen's guidelines on the comparison of means between two groups, a score of 0.2 can be interpreted as a small effect, 0.5 as a medium-sized effect, and 0.8 as a large effect. All operations regarding statistical analysis will be done using the implementation from the Scikit-learn package for Python (Pedregosa et al., 2011).

Chapter 4

Aesthetic Classification of Images Using Brain-Inspired Features

Predicting people's interest for visual information has seen an exponential popularisation, mainly due to the emergence of social networks and the need to find relevant content among the constant information feed. Computational aesthetic classification was first attempted in Datta et al.'s paper in 2006 (Datta et al., 2006), and it has since attracted more researchers thanks to new larger datasets sometimes provided with semantic tags. Datasets are retrieved initially from community-based photography websites and consist of images, mainly photographs, associated with ratings given by users. Users are asked to rate images using graduated scales depending on photographic rules, general aesthetics and for some, their coherence in a pre-defined context. In a typical binary classification task regarding aesthetics, images are split into two categories using the average of all the user ratings collected, which can vary from a dozen to hundreds depending on the dataset.

Datta et al.'s work used measures based on photography rules (Datta et al., 2006). While it shows positive results, this type of measure may suffer from a bias caused by the subjective interpretation of photography rules by their designers and algorithm developers. Indeed, some measures such as the rule of thirds or golden ratio can be interpreted differently and compress a specific artistic criterion into one number certainly affects data accuracy. The photography rules are inspired from artistic practices of the western world and directly assume preferred visual features and preferred feature combinations. In the footsteps of Datta et al.'s work, further attempts using Support Vector Machines (SVM) and multilayer perceptrons aimed at improving classification results by applying measures modelling photography rules but also computer vision algorithms originally used for recognition tasks. For example, descriptors have allowed efficient encoding of visual information by relying on algorithms inspired by the human visual cortex (Marchesotti et al., 2011). Visual complexity was also introduced as a good predictor of aesthetics with estimations calculated using image compression algorithms or fractal dimensionality analysis (Romero et al., 2012). Those measures can, however, make data less malleable for further investigations.

With the creation of the Aesthetic Visual Analysis (AVA) dataset, approaches to aesthetic classification became more machine learning-oriented with the apparition of deep neural networks (Murray et al., 2012). In recent papers by Lu et al. and by Wang et al. (Lu et al., 2014; Wang et al., 2016), brain-inspired architectures with parallel processing columns are designed so that the deep neural networks learn preferred aesthetic features. While the architectures are themselves brain-inspired, the inputs to the neural networks are focusing on colours channels, photography rules, and raw images. Parallel architectures have produced some of the best results in terms of rate of correct aesthetic classification. Lu et al. and some recent works on the AVA dataset have used Convolutional Neural Networks (CNN/Conv. Net.) with raw images as the input, partly because extracting visual features from a large dataset of images can take a considerable amount of time (Kao et al., 2016; Lu et al., 2014; Murray et al., 2012). All the previous solutions including warping or cropping images, Mai et al. avoided denaturing images by using multiple sub-networks of different spatial pooling sizes, leading to improved performance (Mai et al., 2016). While CNNs and descriptors have shown to be efficient and displayed good performances, it is complicated to evaluate and dissociate the contributions from low-level visual features and the semantic contained in visual information to the learning process. In point of fact, contexts and semantics have been shown to influence strongly aesthetic judgements of images and computational classifying systems' performance (Kao et al., 2016, 2017; Simond et al., 2015). Exploiting semantic information to the maximum, Tian et al. developed a query-dependent aesthetic model which retrieves and builds a training subset of photographs with textual tags matching the tested photograph using the AVA dataset (Tian et al., 2015). This computational experiment shows that context recognition using pre-defined textual tags highly increases performances. It highlights questions regarding the boundary between the two families of aesthetic models which are universal and query-dependent. Most CNN-based aesthetic classifiers are considered as universal aesthetic models due to the training and testing phases being independent of each other. However, it can be argued that raw images provide more hints of semantic information than engineered features or photography-based measures, meaning that aesthetic classifiers could also be characterised by their potential to recognise contexts. To conclude on related works, a recently issued paper by Sun et al. has adopted a similar perceptual approach to the one presented here. It estimates visual complexity for aesthetic prediction using distributions of edges and colours with many additional measures of textures and edges (Sun et al., 2018).

The described approach introduces features inspired by the ones appearing in the human visual process, such as colour, shape, depth, and motion (Kandel et al., 2000). Considering that depth of field and motion detection in static images can only be assumed, it was decided to focus only on shape and colour. The suggested features provide an explicit representation of real-world visual information as inputs to the classifier, as well as some level of abstraction of visual information to investigate the necessity of semantics to aesthetics. Only the percentage distributions of the extracted features are used for classification, which reduces the impact of contexts and semantics on the learning of aesthetic preferences by removing semantical hints from spatial organisation. As shown in Sammartino and Palmer's paper, spatial composition can have a real impact on aesthetic perception if a context is known (Sammartino & Palmer, 2012). Based on low-level visual features processed in the early human visual system, the distributed measures permit to analyse preferences for specific orientations, curvatures, and colours. The fact that the distributions of the previous features are independent of each other allows to relate the results directly to existing neuroscience and psychophysics experiments, as well as removing the effect of spatial composition. A limited representation of spatial composition is included in the aesthetic classification process with a measure of global reflectional symmetry.

This chapter offers an approach to aesthetic classification where percentage distributions of low-level features are used to learn aesthetic preferences. It allows to encode visual information into a low number of features fed as only 114 inputs to the classifier, against several hundred or thousand inputs for latest similar works. The classifier maintains state-of-the-art results even across datasets, hinting at possible cross-media aesthetic classification. Cross-dataset tests are also used to evaluate the assimilation of the population's aesthetic preferences even when different rating scales or communities were inquired. Feature analysis allows to question whether the aesthetic classifier learns about photography rules or more general human aesthetic preferences, as for example, horizontal lines or the colour blue have shown to be preferred.

4.1. Dataset configurations in related works

4.1.1. Description of the first dataset of photographs with aesthetic ratings (Datta et al.'s dataset)

The dataset by Datta et al. consists of photographs posted on the website photo.net, and each image is linked to an average aesthetic rating provided by the community (Datta et al., 2006). Users could post photographs, and provide both comments and ratings (using a scale from 1 to 7) on a voluntary basis. Photographs were not required to have a specific context or background, but the community of photo.net appears to appreciate technical information such as camera settings (camera model, focal length, exposure time, etc.). High ratings on a photograph would only lead to potential exposure on the website's homepage, meaning that it may not trigger strong competitive motives. While the dataset was built using the rating system previously described, the current version of photo.net does not operate with ratings anymore. The rating system has been replaced by a "like" feature, implying that viral content may have now more potential to be promoted than photographs of high quality. Not all images were available to download at the time of the study, and only images with at least 10 user ratings were used. The initial dataset contained 20,278 images and the final total of available images is 17,453. To be able to compare our results to previous classifiers, only photographs in the dataset with the 10% best and 10% worst ratings were kept (10% of the initial 20,278 images).

4.1.2. Description of a large-scale dataset of photographs with aesthetic ratings (AVA Dataset)

The AVA Dataset is initially a set of 255,530 photographs extracted from the website DPchallenge.com (Murray et al., 2012). While users provided ratings on a voluntary basis as in Datta et al.'s dataset, all posted photographs are initially taking part in contests with pre-defined themes, where the winner only gains increased attention from the community. Such context means that this dataset may be more affected by competitive rating behaviours. All 255,530 images were available for download and the same classification rule was applied as for the dataset by Datta et al. The top and bottom 10% of the rating distribution were extracted, resulting in 25,553 aesthetically high and poor images. The rating scale is from 1 to 10 with the median average rating for the entire dataset of 5.35, while the mean of average ratings is of 5.33.

4.1.3. Computational Aesthetic Classification

Since the practice of computational aesthetic classification is still in its early years, it has not been regulated and computational experiments are difficult to compare with each other. It is mainly due to researchers having different backgrounds and publications aiming at different audiences (computer vision, machine learning, human-machine interaction, neuroaesthetics, etc.). For instance, the aesthetic classifier RAPID by Lu et al. has been used as a reference for aesthetic classification results over the AVA dataset (Lu et al., 2014). Despite designing an innovating deep-neural architecture and demonstrating impressive results, the classification rates may be biased by the distribution of the training set. The rating value that represents the border defining whether an image is labelled as aesthetically good or aesthetically poor is selected by researchers and can lead to unbalanced classes. Selecting 5.0 as the class border on a 1 to 10 scale generated a training set containing 167,000 highly aesthetic images and 68,000 lower aesthetic images. It means that each class has a respective distribution of 71.06% and 28.93%. The same distribution was calculated for the testing set and the rate of correct classification reached 74.46%. In case of unbalanced datasets, artificial neural networks tend to learn the probability of distribution and give the most represented class as output (Lawrence et al., 2012). Therefore, assessing the effectiveness of such computational experiments is problematic, particularly when the percentage of correct classification for each class is not explicitly stated. Despite Sun et al. exposing issues with class borders, tests on samples with various rating gaps do not always present balanced samples between classes, results are not reported for both aesthetic classes and tests on smaller samples were not reproduced several times using different training samples, leading to less significant results (Sun et al., 2018). Another issue encountered in aesthetic classifier comparison is the pre-selection of datasets, where a subset is extracted out of original datasets to focus on landscapes or portraits for example. Better results are expected from such practice due to the diminished diversity of images, meaning less aesthetic rules to learn and an increased ease to recognise patterns. For example, Tan et al. and Lee et al. reaching respectively 87.10% and 87.98% of correct classification on different subsets of photos of landscapes and nature from the AVA dataset (Lee et al., 2017; Tan et al., 2017). Following a similar approach of developing an aesthetic classifier for a specific photographic theme, the query-dependent aesthetic model by Tian et al. is not directly comparable to universal aesthetic models due to the use of tags to actively recognised the semantic content of photographs (Tian et al., 2015). Regarding pre-selection, borders surrounding photographs are sometimes cropped out or images containing borders excluded. While highly aesthetic images and borders seem to correlate, this bias may be explained by the fact that borders may have a role in aesthetic appreciation of photographs. To conclude about suggested good practices for easier comparisons, it is essential to make aesthetic classification results more trustworthy to assess aesthetic classifiers based on their inputs and designs rather than the percentage of correct classification only. One proposed solution is to balance classes and when not possible, provide the percentage of correct classification for each class. Then, classifiers designed for particular subsets should be tested against the entire original datasets, to shape a link with existing works and to avoid implications that a proposed method is better when only assessed over a particular subset.

4.2. Method: Brain-inspired visual feature extraction and machine learning

This section describes the set of low-level visual features extracted from images and the different machine learning algorithms that will be combined to learn aesthetic preferences. The algorithms to extract this new set of visual features have been designed and developed as part of this thesis. However, as the machine learning algorithms are issued from third-party implementations, the description will be focusing on the architecture and hyperparameters selected.

4.2.1. Gradient orientation distribution

The first set of features extracted represents orientation preferences in the visual system. Proofs of preferences for specific orientations do not only exist in humans, but also in other mammals such as ferrets or monkeys (Blasdel, 1992; Chapman & Bonhoeffer, 1998; Girshick et al., 2011). The human visual system has demonstrated preferences for cardinal orientations (horizontal and vertical) and those preferences have an impact on aesthetic evaluation.

To observe the impact of orientation distribution in images as a factor of aesthetic quality, a customised version of the Histograms of Oriented Gradients algorithm, also called HOG, is employed to estimate the dominant orientation of gradients present in each area of an image (Dalal & Triggs, 2005). When processing the original HOG algorithm, a grid formed of 16x16 pixels squares, called cells, is applied onto the selected image and a histogram of oriented gradients is then processed for every cell. A modification of the original HOG algorithm has been made so the dominant orientation out of 32 possible orientations is detected for each cell before calculating the percentage distribution of orientations for the entire image. The 32 orientations are spread evenly over 180° and each orientation is distributed in bins representing a 5.625° angle. The conversion to percentage is made to normalise the results considering that the pictures do not have the same number of pixels and that cells are defined by a number of pixels.

4.2.2. Distribution of local curvature

In the same manner as orientation preferences, some studies have investigated preferences for curvature in visual stimuli. For instance, it has been demonstrated that humans, as well as great apes, have preferences for rounded corners over squared corners (Munar et al., 2015). It implies that the preference emerges from some neurological structure in the primate brain and it was suggested to be due to a negative response to sharp corners. Moreover, this negative response would be triggered by activity in the amygdala, causing fear and arousal (Bar & Neta, 2007). However, another recent study has shown through several experiments that curved lines are also visually preferred over straight lines, and visual information complexity had no impact in this aesthetic judgement (Bertamini et al., 2016). Therefore, it can be suggested that curved lines and rounded corners are favoured over sharp corners and straight lines, without the preference having emerged from a dislike for sharp corners.

When processing HOG of an image, the estimated orientation of each pixel is calculated by combining the differences between greyscale values of horizontally and vertically adjacent pixels, as depicted in Figure 4.1 (Dalal & Triggs, 2005). Therefore, the same process used to detect orientation is applied a second time for each pixel, combining the differences between estimated orientations of adjacent pixels to detect orientation changes. It can be considered as a measure of curvature due to the analogy with the approximation of the second derivative. This process averages the change of gradient orientation between the pixels situated vertically and horizontally of the selected pixel. Again, similarly to the processing of gradient orientation, the dominant type of curvature is selected for each 32 by 32 pixel cell, before calculating the percentage distribution of the different types of curvature. The feature extracted represents the distribution of the most dominant changes in gradient orientation for each 32 by 32 pixel cells (assuming that the cells are organised under the form of a grid). To summarise, the image is split into squares of pixels where the dominant type of curvature is extracted, and the percentage distribution of those different types of curvature is then calculated.

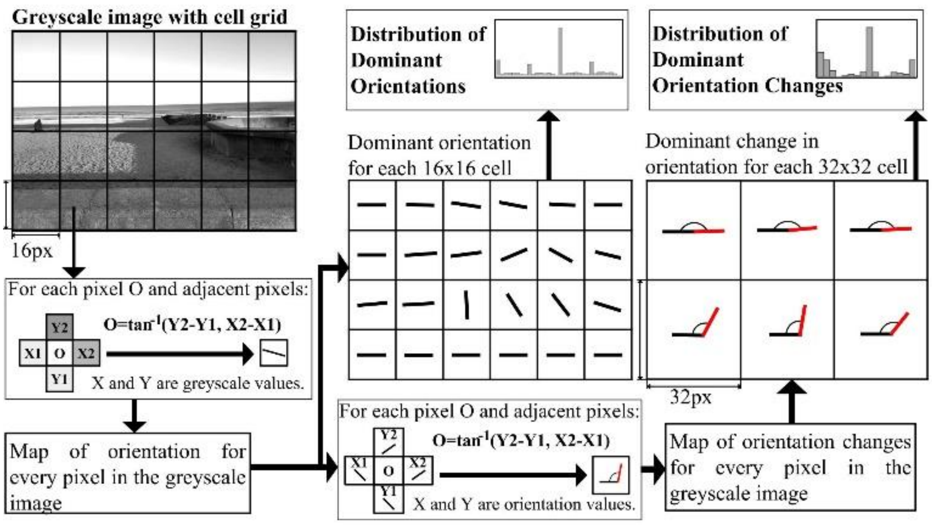


Figure 4.1: Extraction of orientation and curvature distributions. The data displayed in this figure is purely illustrative.

4.2.3. Global symmetry

Studies in neuroaesthetics have shown that symmetric patterns can prime positively viewers (Pecchinenda et al., 2014). However, not all types of symmetry have the same effect. Electroencephalograms (EEG) studies have looked at reflectional symmetry and displayed stronger reactions than for rotational or translational symmetry, implying an ease to detect this particular type of symmetry (Makin et al., 2012). It is then suggested that symmetry is unconsciously processed and is one of the many factors involved in aesthetic pleasantness.

The measure representing global symmetry consists of using the HOG algorithm on all 4 quarters of each image (split through the middle vertically and horizontally), and then subtracting respective histograms from one corner to another (vertically, horizontally and diagonally). All the values of the resulting histogram are then averaged. The lower the average is, the more likely the image is to contain reflectional symmetry.

4.2.4. Colour: hue, saturation, and brightness

Colours are processed in the retina and also later in the primary visual pathway. Knowing that people tend to consciously have favourite colours, it seems logical that colours have a direct impact on our appreciation of visual information. However, as demonstrated by Ou et al., colours meanings can change depending on the gender and cultural background on the viewer (Ou et al., 2004a). While those results may underline the impact of one's personality on visual preferences, it also shows that specific colours are preferred among populations. As shown in Romero et al. computational experiment, colour channels for brightness, hue and saturation can be used as reliable predictors of aesthetic appreciation, with up to 75.81% of correct classification on the Datta et al. dataset (Romero et al., 2012).

In a similar approach as the other measures, colour hue, saturation, and brightness (HSB) were extracted for each pixel, before calculating the respective percentage distributions for each colour characteristic.

4.2.5. Parameter selection for feature extraction algorithms

As mentioned previously, the features extracted from the photographs are designed to represent visual features in the early human visual process while providing some level of data abstraction. It intends to prevent the classifier from relying extensively on hints of context or semantic content present in images. As discussed by Hughes, the peaking phenomenon states that for a finite data sample, feeding additional features into a pattern recogniser becomes counterproductive past a certain point and may result in more classification errors (Hughes, 1968; Sima & Dougherty, 2008; Trunk, 1979). This means that the number of features must be carefully selected to keep superficial information as low as possible. On the other hand, it may reduce feature extraction and training times. The algorithms for extraction of orientation and curvature distributions used in this paper have been developed and calibrated on Datta et al.'s dataset. In this case, increasing the number of distribution bins for these two types of features leads to increased feature extraction time, as well as training time. A series of computational experiments is conducted with an increasing number of bins to find the saturation point where additional bins do not improve classification performances. It was observed that the number of bins providing the best classification scores is 32 bins for orientation distribution. Following the same process, the number of bins for curvature distribution displaying the best classification results is 16 bins. Regarding the features representing colours, the number of bins only affects training time and classification performances due to the algorithm extracting colour distribution having a negligible processing time. Applying the same method as for orientation and curvature, the number of distribution bins selected for each colour characteristic (hue, saturation, brightness) is 20 bins, as classification performances do not increase with additional bins. Finally, the number of features representing symmetry is fixed due to the number of symmetries possible between the 4 quarters of an image. Therefore, the total number of features fed into the machine learning algorithms reaches 114 inputs (32 for orientation distribution, 16 for curvature distribution, 6 representing symmetry, 20 for colour hue distribution, 20 for colour saturation distribution, 20 for colour brightness distribution).

4.2.6. Neural network architectures and parameters

In most previous works focusing on the visual feature extraction approach, the most represented machine learning algorithms used by the community have been Support Vector Machines (SVM) and Multilayer Perceptrons. To link previous works and recent approaches using deep learning, SVMs and multilayer perceptrons are studied, as well as two different deep neural architectures.

Concerning the more traditional machine learning algorithms, SVMs and multilayer perceptrons experiments were run thanks to the implementations from WEKA, a data mining software with pre-implemented machine learning algorithms (Hall et al., 2009). The SVM selected for the computational experiments utilises a polynomial kernel and the following parameters: $C=10.0$, $\varepsilon=10^{-12}$. The multilayer perceptron is conceived, simply, of a single hidden layer with 11 units, a learning rate set to 0.1 and a momentum set to 0.

Regarding the two deep learning models, the code was developed in Python, which allowed the use of Google's deep learning library, Tensorflow (Abadi et al., 2016). The first architecture tested is a traditional fully-connected deep neural network (DNN) with 3 hidden layers, respectively 250, 140 and 80 units. The second one is a novel architecture presented in Figure 4.2, comparable to the one used in the computational experiment by Lu et al., is based on multi-column deep neural networks, since it has been suggested that visual information may be processed in parallel (Lu et al., 2014; Zeki, 2015). It is referred to as 6CDNN, for 6-Column Deep Neural Network. It could possibly increase performances as aesthetic measures tend to overwrite each other during training, leading to minor measures having a disproportionally small impact on the outcome. The architecture consists of separating the network's layers depending on the aesthetic measure they represent. Therefore, each measure is trained separately before converging towards a common final layer and has the opportunity to influence the final decision. The two deep neural networks are set with the hyperparameters described in Table 4.1, with regularisation achieved by using dropout and early stopping. Despite the similarities in structures and hyperparameters in the two architectures, both have been conceived and tuned by trial and error to reach the best correct classification scores on separate validation sets.

Finally, a Convolutional Neural Network is used as a control and is fed raw RGB images instead of the proposed set of features. The role of this classifier is not to challenge state-of-the-art results but to offer a comparison between the solutions presented in this chapter and machine learning algorithms not requiring prior data filtering. Images are only warped into 128x128 inputs with the 3 RGB channels. The CNN is composed of 2 convolutional layers of respectively 32 kernels of size 5x5x3 and 64 kernels of size 3x3x32, followed by a fully-connected layer of 1024 units providing with the classification decision as output. Both convolutional layers are followed by a max pooling layer. The CNN used similar hyperparameters as the two other deep neural networks, except for a learning rate of 0.001 and a dropout probability of 0.5.

To refer to previous works using the Datta et al. dataset, it was decided to first extract visual features from this dataset, before working with a similarly sized sample from the AVA dataset, and finally, using the entire AVA dataset. Results are processed using an average of 10 distinct 5-fold cross-validation runs for the two smaller datasets, while an average of 10 runs of percentage split (60% training, 10% validation, 30% testing) is used for the AVA dataset.

Parameters	**Choice**
Learning rate	0.01
Optimiser	Adam
Activation function (hidden layers)	Rectified Linear Units (ReLU)
Activation function (output layer)	Softmax
Dropout probability (input layer)	0.2
Dropout probability (hidden layers)	0.3

Table 4.1: Parameters for DNN and 6CDNN, created using Tensorflow.

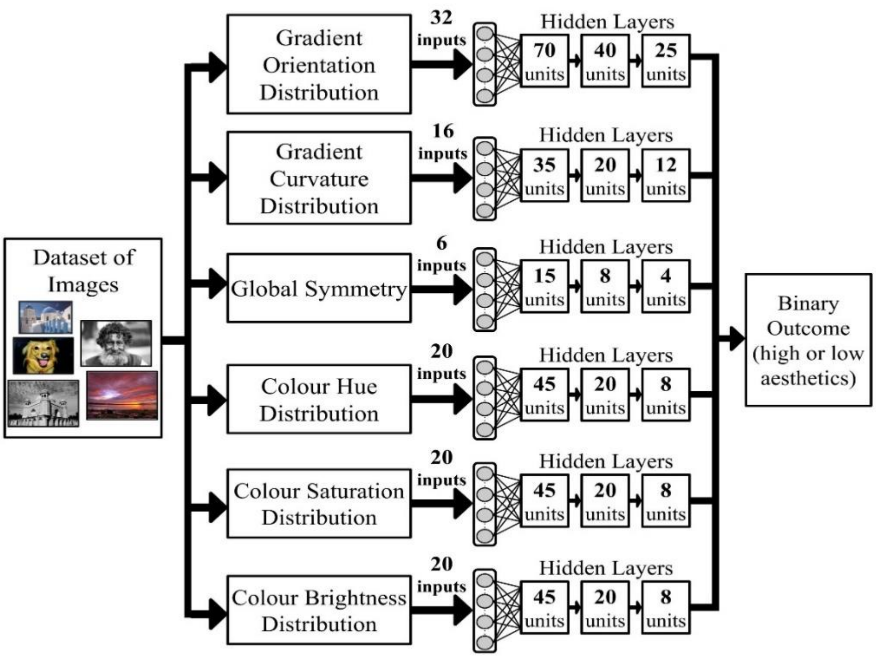


Figure 4.2: Representation of the 6CDNN architecture.

4.3. Machine learning models for aesthetic classification

4.3.1. Classification results

The main reason for running this first computational experiment is to evaluate whether the size of the dataset by Datta et al. that is used in many previous works is large enough to learn aesthetic preferences from low-level features. Results for the Datta et al. dataset using traditional classifiers such as support vector machines (SVM) and multilayer perceptrons do not surpass state-of-the-art results (Table 4.2). While SVM appears to perform as well as the other proposed neural network architectures on the smaller datasets, it does not benefit from a larger dataset. Overall, both deep neural architectures not only outperform Datta et al.'s features but also outperform all other architectures tested on all datasets, with a significant improvement seen when using the entire AVA dataset (Datta et al., 2006). The difference between the correct rates of classification on the small sample and the entire AVA dataset implies that the whole dataset is more representative of users' average visual preferences and allows better generalisation altogether. Nonetheless, when comparing to the convolutional neural network, it confirms that measures representing photography rules or low-level visual features may be more efficient on small datasets. The CNN shows the highest performance increase between the small sample and the entire AVA dataset, meaning it was the most affected by the lack of data.

To finish, the DNN and 6CDNN architectures showed similar results, implying that the hidden layers' segregation to simulate parallel processing had little effect on aesthetic classification.

Classifiers	Dataset								
	Datta et al.			AVA (4000-sample)			AVA		
	Avg	Good	Poor	Avg	Good	Poor	Avg	Good	Poor
SVM	69.76	67.90	71.61	69.58	69.55	69.59	70.82	69.60	72.08
Multilayer Perceptron	68.49	66.63	70.34	69.00	69.39	68.63	74.07	76.14	72.00
DNN	71.63	70.10	71.58	71.87	73.90	69.84	78.81	79.76	77.86
6CDNN	71.31	67.63	74.99	71.15	71.10	71.19	77.46	80.97	73.98
Conv. Net.	61.43	61.62	61.25	58.01	48.32	67.70	69.83	70.23	69.42

Table 4.2: Results for each dataset with different classifiers and percentages of correct classification for aesthetically good, aesthetically poor images and total average.

4.3.2. Locating the border between aesthetic classes

The previous results were obtained by selecting only images in the top and bottom 10% of the rating distribution of each dataset. Instead of arbitrarily selecting borders for good and poor aesthetic classes, the border is implicitly drawn by selecting equally sized samples of images in both the top and bottom end of the rating distribution. For all further computational experiments, the AVA dataset is selected due to the amount of data available and classification is attempted using the more traditional fully-connected DNN architecture.

As mentioned previously, aesthetic classifiers have been comparable only on their percentages of correct classification. The following computational experiment shows further details about the impact of the gap in ratings between the two aesthetic classes. The samples of images represent 10% of the top and bottom of the rating distribution, with a sample expansion of 10% towards the median rating for each test. As it can be observed in Figure 4.3, in the case with the largest gap, images with poor aesthetic proved to be slightly more challenging to classify than highly aesthetic images.

Regarding the gap between the two aesthetic classes, the first 20% of images on each side of the median rating cannot be classified accurately. The rates of correct classification neighbouring chance (50% for binary classification tasks), it is possible that visual features present in the highest and lowest aesthetic images generate confusion when classifying images close to the border (Figure 4.4). Moreover, this poor performance could imply that this part of the dataset is represented by another aesthetic class.

As the previous results of binary classification showed that the classifier performs at chance level around the class border, it could hint at another class including images of “average” aesthetic level in the middle of the rating distribution, as well as it could simply be caused by the complexity and ambiguity of classifying this type of images. The implementation of an additional class for regular images proves to be inconclusive. Despite all three classes being balanced during training, no image is classed as regular and aesthetically high and low images are privileged. The classification of aesthetically high and low images reaches similar correct classification rates as in the binary task while regular images are entirely ignored by the classifier. It indicates that the training for recognition of regular images is ineffective and it solely creates ambiguity between classes and around class borders.

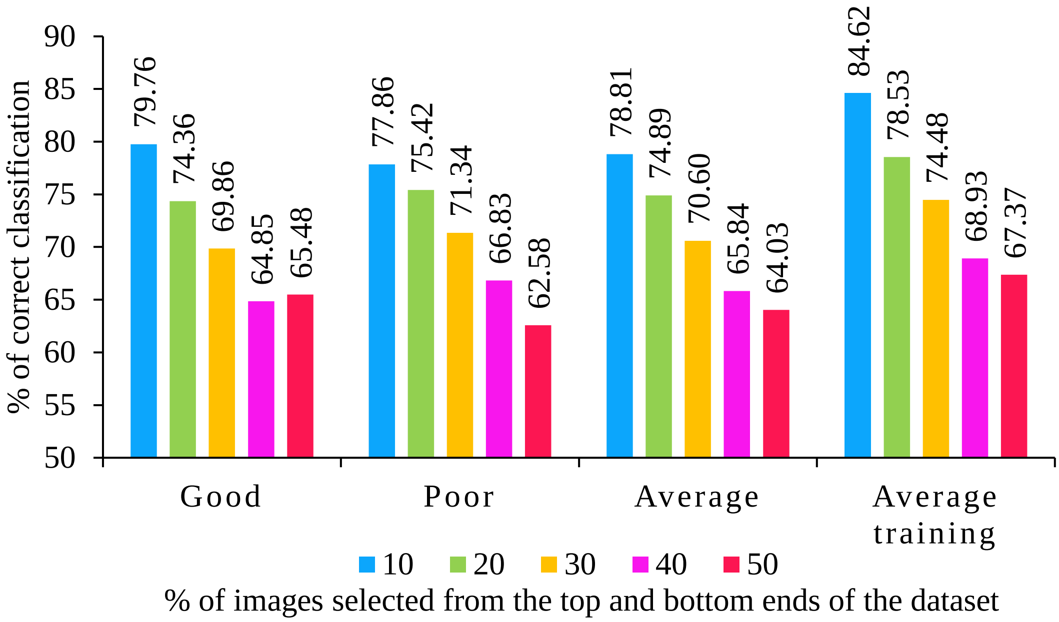
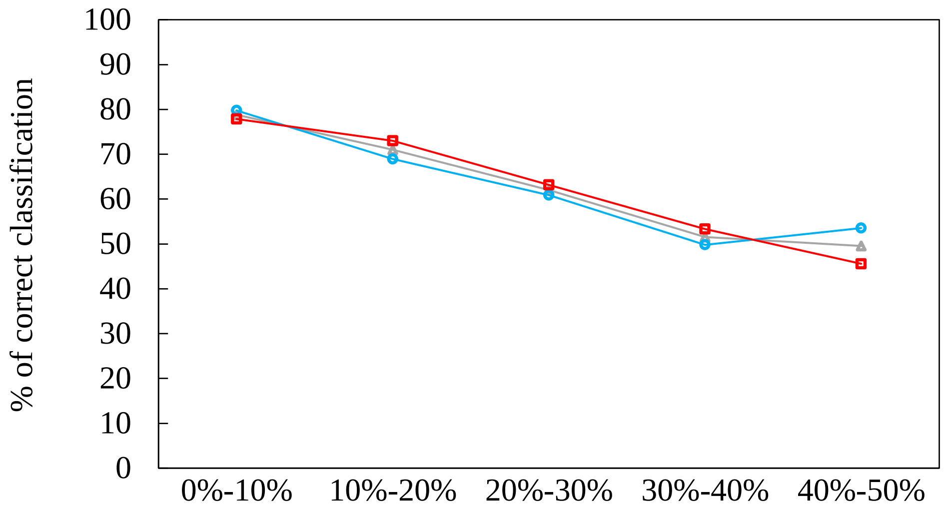


Figure 4.3: Percentage of correct classification depending on the percentage of images selected from the top or bottom of the aesthetic rating distribution also called respectively aesthetically good and poor images. For example, 50% represents images from one end of the rating distribution to the median rating.



	0%-10%	10%-20%	20%-30%	30%-40%	40%-50%
Avg	78.81	70.97	62.01	51.56	49.55
Good	79.76	68.96	60.87	49.80	53.52
Poor	77.86	72.98	63.16	53.32	45.58

Samples of images from the top and bottom

Figure 4.4: Percentage of correct classification where samples only represent 10% of the total images (around 25,000 images) on each side of the median rating. Each test uses two samples of aesthetically good and aesthetically poor images, selected from the top and bottom end of the rating distribution. While results represent correct classification for the test set on a 10% sample, the training process took into account all images above the bottom limit (meaning that the result for the sample of images from 20% to 30% on both sides of the median rating actually trained on all images in the top and bottom 30%).

4.3.3. Cross-dataset performance and comparison with existing works

As mentioned initially, comparing current results with existing studies is a complex task due to the number of datasets and the various class borders. In Table 4.3, the performance of the proposed classifier and the convolutional neural network are affected in a similar manner by the different dataset settings. It can also be observed that the proposed classifier displays similar results as Sun et al. obtained when splitting the AVA dataset using the median rating. However, while the proposed solution and the tested convolutional neural network both perform slightly better when shifting the class border to mid-scale, Sun et al.'s results dramatically escalate. Due to the proposed solution being trained on a balanced dataset unlike previous works, it seems that all results obtained using unbalanced training sets and mid-scale class border are too biased for comparison. To finish on the comparison between Sun et al. and the proposed solution, Sun et al. gave classification results for a subset of around 40k photos with each half issued from the top and bottom of the AVA dataset. Despite having more fundamental features, the proposed method provides significantly better results. Beside Sun et al.'s results, only works by Lo et al. (2012) and Mavridaki et al. (2015) have been found to be directly comparable due to their training and testing phases run on balanced datasets, and the proposed solution is shown to outperform both.

When comparing the efficiency of other existing classification systems, Mavridaki et al.'s (2015) system utilises 1323 inputs against only 114 inputs for the proposed classifier. While Lo et al.'s classifier only requires 24 inputs, its results on the AVA dataset are far below the proposed classifier's results. Nonetheless, it does display a shorter processing time with an averaged 0.26s per image of width and height below 480pixels, against an averaged 0.41s for the proposed method (configuration: Linux, Intel i5-4210U CPU at 2.40GHz, 16GB RAM). Sun et al.'s experiment is the most competitive in terms of approach and results but averages a processing time of 33.45s per image over 500 images of 640 by 480 pixels, while the proposed method reaches a much shorter 1.04s. Therefore, the proposed method is an attractive compromise between good efficiency compared to works with related approaches and higher accuracy than other previous systems focusing on efficiency.

After selecting AVA as the best dataset for training, the results shown in Table 4.3 demonstrate satisfying cross-dataset performance when testing the "AVA-trained" classifier over the Datta and CUHK datasets. It brings additional evidence that the reported level of performance is not due to overfitting the data (Datta et al., 2006; Tang et al., 2013). While the CUHK dataset was not initially selected due to its arbitrary photograph categorisation, aesthetic classification tests were run to link the comparison from the results on the AVA dataset to other previous studies while demonstrating the cross-dataset capabilities of the proposed solution. Despite a non-negligible performance loss, the results are reasonably close to comparable works which have the advantage of being trained and tested on the same photograph datasets. While all known works show similar classification rates on aesthetically high and low images when testing on the native dataset, significant differences could be seen when testing the proposed classifier on Datta et al.'s (high: 89.79%, low: 50.76%) and CUHK datasets (high: 60.14%, low: 91.68%). This imbalance in performance between the two aesthetic classes may imply that the community of photo.net is more elitist than the DPchallenge.com community, while the arbitrary categorisation by the authors of the CUHK dataset show to be more tolerant.

Classifiers	**Datasets tested**						
	(Datta et al., 2006)	AVA (Murray et al., 2012)					CUHK (Tang et al., 2013)
	Mid-scale ± 0.8	Mid-scale	Median rating	20% top & bottom	10% top & bottom	20k top & bottom	Authors' categorisation
Proposed classifier	70.28*	65.84**	64.03	74.89	78.81	79.80	75.91*
Conv. Net.	64.84*	58.31**	56.75	64.50	69.83	71.17	62.83*
(Datta et al., 2006)	70.19	-	-	-	-	-	-
(Romero et al., 2012)	75.81	-	-	-	-	-	-
(Lo et al. 2012)	-	-	-	62.14	66.60	-	77.25
(Mavridaki et al. 2015)	-	-	-	74.35	77.08	-	82.41
(Murray et al. 2012)	-	66.7	-	-	-	-	-
(Lu et al., 2014)	-	74.46	-	-	-	-	-
(Wang et al., 2016)	-	76.8	-	-	-	-	-
(Sun et al. 2017)	-	73.41	64.24	-	-	76.28	-

*classifier originally trained on the AVA dataset   ** mid-scale border with equally distributed classes in the training set.

Table 4.3: Trained only on the AVA dataset, the proposed classifier and a convolutional neural network used as control are both tested on 3 datasets and compared against previous works trained and tested on the same datasets, with occasionally, different border between "good" and "poor" aesthetic classes

4.3.4. Visual preferences of the aesthetic classifier and differences between the average features of each class

One additional advantage to using such sets of low-level visual features is that it makes datasets for computational aesthetics easily understandable and interpretable due to its direct relationship with the human visual system. The features can be manipulated to study the differences between aesthetic classes in a dataset. It allows to investigate the visual preferences of a rating community, creating the possibility of comparing visual preferences across communities. For both classes of the AVA dataset, average features of over 127,000 images are calculated to compare their values and highlight the best factors of good aesthetics. The same process is also applied to the Datta et al. dataset to confirm whether the extracted features are representative of human visual preferences across datasets. In both datasets, vertical lines represent 16% of all line orientations with no significant change (using an independent samples t-test) between the two aesthetic categories. As presented in Figure 4.5, a significant increase is observed in horizontal lines ($t(254522)=28.79$, $p<.001$, $d=0.11$) from aesthetically poor ($M=20.56\%$, $S=.11$) to good photographs ($M=21.88\%$, $S=.12$) for the AVA dataset, as well as a similar significant increase in the Datta et al. dataset ($t(17288)=12.24$, $p<.001$, $d=0.18$) from aesthetically poor ($M=20.41\%$, $S=.10$) to good photographs ($M=22.39\%$, $S=.11$). While it could mean that horizontal lines are dominantly detected by the HOG algorithm in aesthetically pleasant images, it could also imply a visual preference for horizontal lines within the online community who rated the datasets. In Figure 4.6, a strong preference is observed in the AVA dataset for gradient orientation changes of around 45° within a 32x32 pixels square ($t(255030)=19.15$, $p<.001$, $d=0.076$), as well as a preference for 90° gradient orientation change ($t(255511)=27.20$, $p<.001$, $d=0.11$). Again, a similar observation was made for the Datta et al. dataset, with an increase approaching significance for 45°-orientation changes ($t(17144)=1.58$, $p<.08$, $d=.027$) and a significant increase of 90°-orientation changes ($t(17108)=10.87$, $p<.001$, $d=.16$). Regarding colours in the AVA dataset, Figure 4.7 displays a significant growth in the shares of red ($t(255396)=16.04$, $p<.001$, $d=0.063$) and blue ($t(254349)=7.20$, $p<.001$, $d=0.029$) in pleasing images, with an opposite and significant trend for yellow ($t(255331)=16.72$, $p<.001$, $d=0.066$). While aesthetically pleasing images of the Datta et al. dataset display a significant increase of blue ($t(16989)=2.45$, $p=.01$, $d=0.04$), no evolution was observed for yellow and the colour red was also significantly less present ($t(16872)=5.15$, $p<.001$, $d=0.074$). Previous behavioural experiments on the aesthetics of colours displayed an average preference for blue and dislike for yellow (Ball, 1965; McManus et al., 1982;

Palmer et al., 2013), which both appear in the AVA dataset in Figure 4.7. However, only the visual preference for blue was seen in the Datta et al. dataset. The presence of the colour red is also contrasted between the AVA and Datta datasets, with respectively 27.87% and 35.74% in bad photos against 29.91% and 33.15% in good photos, showing some possible conversion towards a right distribution of red. Moreover, the AVA dataset showed a preference for proportionally more unsaturated images, but an inverse tendency has seen in the Datta et al. dataset. The features also reveal preferences for proportionally darker images in both datasets. It is also possible that the colours with a hue value of 0° are on the black and white spectrum and may relate to the distribution of greyscale images in the datasets and in both aesthetic classes.

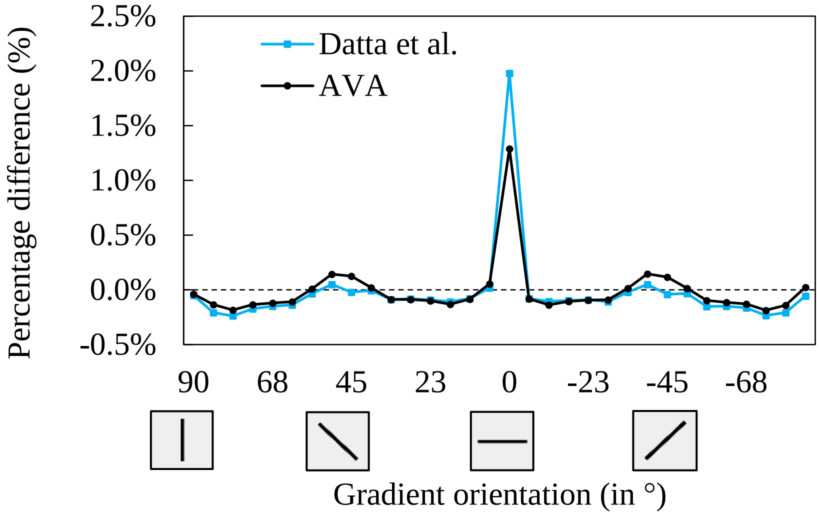


Figure 4.5: Differences between the average features representing orientations of aesthetically high and low images for Datta et al. and AVA datasets. The illustrations below the x-axis are examples of orientations within a cell for the given orientation (16 by 16 pixels).

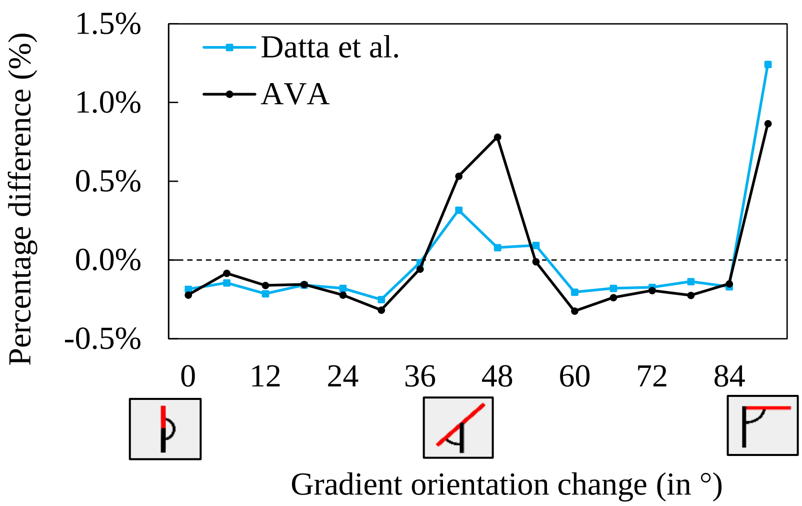


Figure 4.6: Differences between the average features representing curvature of aesthetically high and low images for Datta et al. and AVA datasets. The x-axis represents the average change of gradient orientation within a cell. The illustrations below the axis are examples of orientation change within a cell for the given change of orientation.

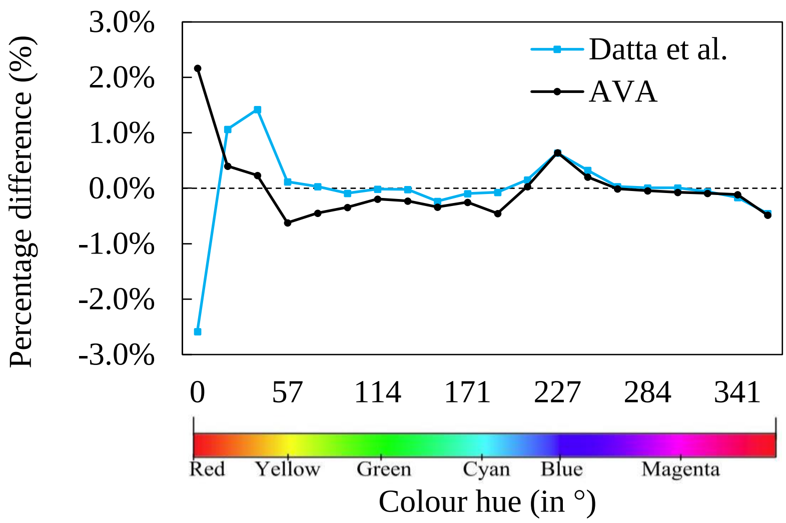


Figure 4.7: Differences between the average features representing colour hues of aesthetically high and low images for Datta et al. and AVA datasets.

4.4. Discussion

Based the previous results shown in Table 4.3, the proposed classifier appears to be a good compromise between performance and efficiency, while displaying robust cross-dataset capabilities. The results in cross-dataset classification show an imbalance in the correct classification of aesthetically low and aesthetically high photographs which may illustrate the visual preferences of the community rating these smaller datasets. In agreement with Kong et al.'s conclusion on the poor skill transferability across datasets between AVA and their own dataset, the rate of correct classification per class may be influenced by a difference in raters' aesthetic taste and the different types of images composing the dataset (Kong et al., 2016). It is suggested that the robust cross-dataset performances are due to the low level of the presented features, allowing to learn visual preferences that are more universal than photographic rules. Furthermore, using low-level features allows a focus on universal human visual preferences partly removed from cultural or personal biases. As the community of DPchallenge.com appears to be mainly westerners with at least 73.45% of visitors connected from English-speaking countries such as the United States, United Kingdom and Canada (data collected from similarweb.com for June 2017), this hypothesis could be confirmed by testing the aesthetic classifier with a dataset rated by people with a drastically different culture of aesthetics, such as East Asian populations for example.

From a method point of view, the previous computational experiments emphasise the importance of reporting performance per class and different ranges of image samples to afford better comparison between aesthetic classifiers in the future. While the features and the classifier presented here seem to perform well on images at the top and bottom end of the rating distribution, it barely performed better than chance on the 20% of images on both sides of the median rating, which represent around 100,000 images of the AVA dataset. It is likely that images from this sample contain features that are too ambiguous to be distinguished. However, another potential source of classification error for any image of the dataset is due to the nature of the dataset and the rating process. The current data collection process assumes that some wisdom of the crowd phenomenon happens for aesthetic ratings, but it is plausible that the community can occasionally make mistakes

(Mollick & Nanda, 2015). The community of DPchallenge.com is asked to rate images on their aesthetic qualities, but it seems natural that many images are rated due to their intense semantic content, or due to the photographer's popularity. Estimating the impact of those biases on aesthetic ratings is complicated, future datasets could require people to rate the meaningfulness of an image to assess the probability of an image to cause controversy.

While the extracted features seem to represent some visual preferences accurately, it is hard to tell whether they truly model aesthetic preferences of the human visual system in general or only aesthetic preferences for photographs. For instance, with the AVA dataset, the horizontal orientation preference is demonstrated by the extracted features while the vertical orientation preference is missing. Similarly for colours, the appreciation for the colour blue and the dislike for yellow are both represented in the extracted features as mentioned in the literature (Ball, 1965; McManus et al., 1982; Palmer et al., 2013), but an undocumented preference for red can also be observed. Despite the results being statistically significant, it is important to point out that the effect sizes, represented by Cohen's *d*, indicate that the phenomena are relatively small (J. Cohen, 1988). Nevertheless, limited effect sizes do not imply that the observed phenomena are not impactful as the combination of the tested variables in this experiment could contribute to the final decision in aesthetic judgement.

4.5. Conclusion

This chapter shows that simple brain-inspired measures of visual information can still perform as well as methods following a machine learning approach for computational aesthetic classification of photographs while providing analysable results. The features extracted from images in the AVA dataset demonstrated visual preferences in the rating community that were observed in previous experiments. It also shows that datasets such as Datta et al.'s may be too small to determine aesthetic rules from low-level visual features. The tested deep neural architectures provide better results than traditional machine learning algorithms such as SVM and multilayer perceptron, but the minimal difference between DNN and 6CDNN also points out that parallel processing for aesthetic classification may not be the most optimal solution. Considering that aesthetic evaluation is a by-product and not a direct goal of the visual system, it is also complicated to make any conclusion about the architecture of our visual system yet. In the future, other features could be added to complete the existing set with, for example, a feature representing motion in images. Motion has been proven to be correlated with aesthetic pleasantness, even in static 2D images (Thakral et al., 2012). One way to estimate motion in images would be to detect the amount of blur. By starting with low-level features, it makes the set of features extremely modular and it is now simple to build up onto the current proposed solution.

To conclude, this biologically inspired design allows to avoid overlaps in feature extraction, and despite requiring a considerable amount of data to learn to distinguish good and poor aesthetic characteristics in photographs, the present results demonstrate that teaching simple visual preferences to a classifier can outperform complex photography rules. Moreover, the set of features has been designed to carry limited semantical hints from spatial organisation, which questions the contribution of such information to the generalisation of aesthetic preferences for a universal aesthetic model. While some query-dependent aesthetic classifiers have shown that semantics and contextual information can boost performance, the aesthetic preferences learnt by the proposed classifier illustrates empirical data and reaches state-of-the-art results compared to other universal aesthetic models. It challenges the received idea that semantics embedded in visual information and its aesthetic value are tightly related. It is still complicated to quantify the influence of semantic information in aesthetic judgement, but it could be beneficial to include photographic style recognition tests alongside aesthetic classification tests when introducing new aesthetic classifiers in the future.

Chapter 5

Computational Aesthetic Classification Beyond Photographs

The domain of computational aesthetics traditionally focuses on digitised photographs. As the first dataset by Datta et al. (2006) was built at a time when the infrastructure of internet was significantly less performant, photographs appeared as the ideal medium to build the first datasets. The popularisation of high-speed internet has facilitated the use and sharing of bandwidth-greedy media, allowing the development of new datasets composed of high definition videos. Moreover, the increasing use of high definition videos makes investigations in visual preferences strongly relevant to filter, as for example, hundreds of hours of videos are uploaded to the streaming website, YouTube, every minute. Even though recommendation systems already suggest videos based on textual tags, speech analysis, or semantic analysis, little has been done to offer aesthetic-based filters. The current lack of video suggestions using aesthetic criteria is mostly due to the limited size of existing datasets. Despite recent efforts to build large datasets of videos, such as YouTube 8M, no video dataset for aesthetic video classification reaches similar scales in terms of number of items as existing datasets for computational aesthetic classification of photographs (Abu-El-Haija et al., 2016). The largest dataset of videos with aesthetic ratings known to date is the recently published dataset by Tzelepis et al., which is composed of 700 short videos collected on YouTube and matched with aesthetic ratings (Tzelepis et al., 2016).

While previous works have focused on computational aesthetic classification of short videos, “The Colors of Motions” by Charlie Clark illustrated the change in dominant colours over several feature films (Clark, 2014; Niu & Liu, 2012; Tzelepis et al., 2016; Yang et al., 2011). Moreover, Jason Schulman’s “Photographs of Films” offers novel ways of looking into the aesthetics of films as they overlap all frames from a film to obtain a single merged image (Shulman, 2017). The computational system previously proposed was developed and trained to classify photographs depending on their aesthetics, whereas this chapter introduces the cross-media capabilities of this aesthetic classifier on the video dataset published by Tzelepis et al. In addition to the tests on Tzelepis et al.’s dataset, the aesthetic classifier is used on films to observe particular aesthetic patterns over time and points out the potential weaknesses and strengths of such classifier on both photographs and videos. At the end of the chapter, the classifier is tested using YouTube videos by Casey Neistat, a filmmaker and daily vlogger (Neistat, 2016). In the form of a case study, potential links between aesthetic prediction and video quality are investigated by looking at the evolution of aesthetics across years of work.

5.1. Training of the aesthetic classifier

In order to compare the behaviour and performance of aesthetic classification systems in photographs and videos, the machine learning-based system previously designed to classify images based on aesthetics that achieves state-of-the-art results on different datasets is selected. The aesthetic classifier is first trained on a large scale photograph dataset called AVA (Murray et al., 2012). The AVA dataset is superior for learning aesthetic preferences as it provides one rating per image, compared to only one rating per video in Tzelepis et al.'s dataset, which can be considered as a collection of sequences of still images. As stated previously, visual information is defined as aesthetically pleasant if it has the potential to induce a positive response among the average observer, which is represented in existing datasets by the rating community's self-reports. The AVA dataset is also superior in terms of representation of human visual preferences, as every image has received at least a hundred ratings according to aesthetic criteria, against only 5 ratings per video for Tzelepis et al.'s dataset. Training for aesthetic classification per image (and therefore per frame) allows a deeper understanding of videos as sequences and scenes can be isolated and analysed. In fact, aesthetic classification systems are usually trained and tested with still images, mainly due to the complexity of collecting aesthetic ratings for video streams.

Previous works have proven to be effective aesthetic classification solutions with, for example, algorithms scoring images based on photography rules (rule of thirds, leading lines, etc.), or more computation-based approaches such as image descriptors and convolutional neural networks linking visual features to expected classifications (Datta et al., 2006; Lu et al., 2014; Marchesotti et al., 2011; Romero et al., 2012). As a reminder, the aesthetic classifier used in this chapter extracts measures of orientation distribution, curvature distribution, HSB colour distribution (Hue, Saturation, Brightness), and reflectional symmetry on cardinal and diagonal axes. A deep neural network composed of 3 hidden layers is then used to learn visual preferences and obtain state-of-the-art results across several datasets such as Datta et al., CUHK and AVA (Datta et al., 2006; Murray et al., 2012; Tang et al., 2013). This classifier is selected due to its cross-dataset performances and the fact that the low-level visual features extracted illustrate fundamental preferences in the human visual system. Therefore, it is suggested that low-level visual preferences can provide better cross-media performance as they tend to be less influenced than higher-level preferences by cultural and personal experiences.

5.2. Experiment 1: Aesthetic classification of video clips by Tzelepis et al.

5.2.1. Method

Aesthetic classes in the AVA dataset are defined for each photograph by the average rating of all the human aesthetic ratings provided with the dataset. As no ground truth is available for the individual frames composing the videos of Tzelepis et al.'s dataset, all further predictions on new images or video frames are considered as a display of the aesthetic preferences of AVA dataset's rating community. The dataset of Tzelepis et al. is composed of 700 short videos downloaded from YouTube and rated in terms of aesthetic quality by a jury composed of 5 people. Due to each member of the jury giving a binary answer for each video, a video is associated with an average rating between 0 and 1. In this computational experiment, the previously trained aesthetic classifier is used to categorise each individual frame as aesthetically low or aesthetically high in a video stream. This allows to obtain the number of aesthetic high frames over the whole length of a video. It will be represented as a percentage of aesthetically high frames due to videos not being of the same length.

Once all videos have their percentage of aesthetically high frames calculated by running the classifier on every frame, it allows to compare the average percentage of aesthetically high frames depending on the jury's rating. Such comparison highlights whether the aesthetic classifier is able to dissociate aesthetically low and high video frames despite its training of photographs. Therefore, it is hypothesised that videos that received a higher rating from the human jury contain a higher percentage of aesthetically high frames.

The percentage of aesthetically high frames is then used as an input to a multilayer perceptron that is trained using the videos' aesthetic ratings. While using a multilayer perceptron may be excessive for such a simple classification problem with one input, it will potentially allow to easily implement additional measures in the future. The method allows comparison with the existing video aesthetic classifier designed by Tzelepis et al. The results are compared in terms of accuracy (correct predictions out of all videos tested) and precision (correct predictions out of positive samples tested). The precision and accuracy are averaged out of 1,000 repetitions and the training set (300 videos) and testing set (400 videos) are randomly sampled for each repetition.

5.2.2. Results

When comparing the percentage of aesthetically high frames in function of the rating given by the human jury, the two values appear to be strongly linked. The linear regression model displayed in Figure 5.1 presents a significant increasing slope of 0.12 ($t(698)=5.11$, $p<.001$), meaning that a greater number of aesthetically high frames are detected in the best-rated videos.

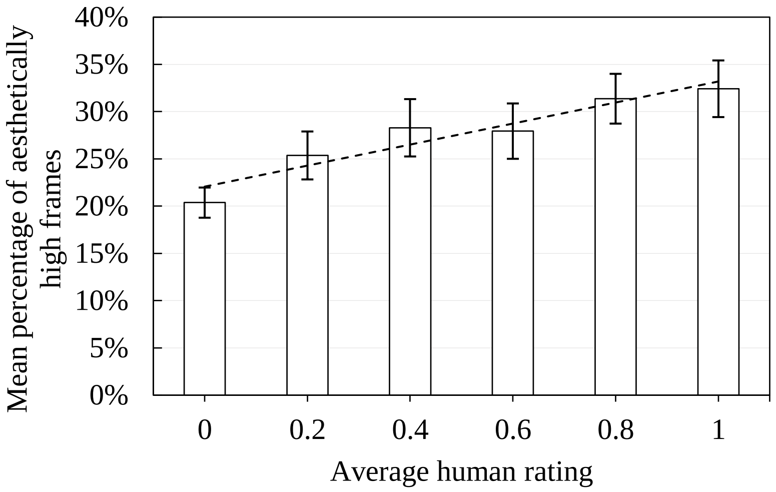


Figure 5.1: Mean percentage of aesthetically high frames detected in a video by the aesthetic classifier depending on the average human rating of the given video.

Despite a considerable difference in performance with the original results achieved by Tzelepis' solution, the proposed solution achieves results significantly above chance, as seen in Table 5.1. This transfer of skill from photograph to video classification demonstrates the cross-media capabilities of the aesthetic classifier and that the percentage of aesthetically high frames in a video can be a relatively efficient predictor of aesthetic pleasantness. This task also allowed to test the proposed binary classifier against another version which gave a continuous output. While the first version decides between the two aesthetic classes (low and high aesthetics), the second version estimates aesthetics on a scale from 1 to 10, as given in the AVA dataset. However, the percentage of aesthetically high frames estimated by the second version did not show any relationship with the human ratings, and the classification of videos depending on aesthetics was only slightly above chance, implying that the binary classification version is much more reliable, even though it has a limited scale.

	(Tzelepis et al., 2016)	**Proposed classifier**
Top 5%	82.00	64.54
Top 10%	82.00	64.74
Top 15%	83.33	64.89
Top 20%	81.50	64.80
Accuracy	68.14	54.60
Precision	69.97	56.38

Table 5.1: Precision for the top-n (5,10,15,20) percent highest-rated videos with average accuracy and precision (in %).

5.3. Experiment 2: Using the trained aesthetic classifier to evaluate a film's frames

5.3.1. Method

Due to the limited number of datasets available for aesthetic classification of videos, the aesthetic classifier is tested using feature films. The high number of frames per second present in videos allows to have several images depicting the same visual content from possibly different points of view. Most recent films last between 90 and 180 minutes at a rate of 24 frames per second on average, which means that extracting visual features of all frames implies an extensive amount of processing. The image resolution of the processed films is 720p (1280×720 pixels), which offers a good compromise between a reasonable image size for meaningful feature extraction and processing speed. Only one frame per second is extracted in order to limit the number of frames to process.

In the following computational experiment, two different types of measures are tested, both based on the proposed aesthetic classifier. The first measure is the percentage of aesthetically high frames over an entire film, like the one tested in the previous experiment. Several films from different directors are evaluated with the aesthetic classifier in order to investigate whether some known characteristics of films can be linked to its percentage of aesthetically high frames. Testing the measure on films also allows to verify whether it works as well as the video clips in the previous experiment, as it can be expected that feature films are highly aesthetic.

The second measure represents the aesthetic average over time. Also calculated from the predictions of the aesthetic classifier, it applies a moving average on the series of binary predictions resulting from the application of the aesthetic classifier on the film's frames. This moving average does not only estimate the aesthetic quality of the visual content over time, but the different points of view observed across frames allow to distinguish sequences containing frames with a normal aesthetic level, despite the fact that the binary classification focuses on low or high levels. Indeed, sequences in which the distribution of the frames' classes is close to chance (50% low, 50% high) implicitly shows that the frames are close to average levels of visual aesthetics, based on the previously learnt visual preferences. This provides additional information regarding the classifier's confidence in its decision; in a binary classification task, this is a significant advantage compared to other existing classifiers. Nonetheless, biases in aesthetic classification may appear due to differences between the norms of photography and videography. Furthermore, in comparison to photographs, videos include additional semantic content due to auditory and motion information. This may mean that even self-reports may not correlate with the aesthetic classifier's predictions, as it focuses purely on visual information.

5.3.2. Results

Scoring feature films by different directors using the aesthetic classifier, Wes Anderson, with his focus on symmetry, has made films reaching rates of aesthetically high frames such as 56.16% for *The Grand Budapest Hotel*, 22.0% for *Moonrise Kingdom*, 20.60% for *Fantastic Mr. Fox* and 58.83% for *The Royal Tenenbaums*. On another hand, Stanley Kubrick, known for his shots in depth, directed *Full Metal* Jacket which presents 12.10% of aesthetically high frames, *A Clockwork Orange* with 17.75%, *The Shining* with 14.12% and *Space Odyssey* with 16.21%. Although percentages of aesthetically high frames possibly indicate some of the aesthetic classifier's visual preferences, reducing a whole feature film to a single score is a highly limiting analysis. Due to having too few films to obtain significant statistics, further investigations on the aesthetic classifier's preferences between the two film directors is difficult, particularly with potential influences from film types or years of release. It is, however, intriguing that the percentage of aesthetically high frames varies so much in comparison to the short videos by Tzelepis et al. (2016) from the previous experiment.

Moving averages of aesthetic prediction are then used to provide insight into how aesthetics may evolve through a feature film. Entire films are processed such as for example, Tarantino films, and some interesting patterns are observed. One film, *The Hateful Eight* (2015), particularly stands out because of the aesthetic prediction averaging zero in the second part of the film despite a normal amount of aesthetically high frames in the first part in comparison to other films. This drop seems to be correlating with the switch from outdoor scenes to indoor scenes in the film, and it was confirmed by showing a strong positive correlation ($r=.82$, $p<.001$) over the entire film between aesthetic prediction and the feature representing the distribution of pixels with normal brightness values (Figure 5.2). Considering the number of other features involved in the processing of aesthetic prediction, this high correlation score shows that the aesthetic classifier trained on photos is extremely biased by the high level of darkness present in the film. The results displayed by the classifier support the assumption that using photographs for training will create a bias in the learnt aesthetic preferences as for example, high levels of darkness are more acceptable to a human eye watching films due to motion.

In another example (Figure 5.3- left), the aesthetic prediction curve of *Django Unchained* (2012) shows a vertical symmetry centred on the middle of the film. The pattern is relevant knowing that Quentin Tarantino designed the film with two parts. It can be speculated that Tarantino knowingly wrote the scenario and organised camera shots to generate a symmetric pattern between those two parts. After removing the credits, the axis of symmetry was found in order to compare the two parts using this axis as a splitting point. As shown in Figure 5.3 (right), when mirroring the aesthetic prediction over time of the second part over the y-axis, a strong correlation (r=.85, p<.001) is found between the two parts that contained over 4,000 frames. Such a strong correlation score seems to indicate intentions from the film director. As the curve of aesthetic prediction does not appear to correlate with the different types of shot scales and the types of scenes (dialogue, action, etc.), it implies that the apparent pattern must be generated by an abnormal value in one of the features, similarly to the outdoor-indoor scene observation made for *The Hateful Eight* film. As suggested by the field of film theory and its auteur theory, the visual aspect of films can be heavily influenced by their director as they represent the major creative force in control of all audio and visual elements (Stam, 2017). While no definite factor is identified as the origin, such pattern may be influenced by easily manipulable features such as symmetry and colours during filming or altered in postproduction by Robert Richardson, the film’s cinematographer.

Relying on IMDb.com's ratings, films considered as of poor quality are also processed such as *Birdemic*, *Batman and Robin* or *Kill Bill*. Not all analysed films display interesting moving averages of aesthetic prediction, but all cases appear to be influenced by the beginnings and ends of sequences present in a film. Considering that the aesthetic classifier is trained on photographs, the influence of film sequences on the aesthetic curve may be due to static dialogue scenes complying more with the rules of good aesthetics in photography than dynamic action scenes. The presented examples expose two advantages of such experiments. First, it allows to test visual preferences of the trained classifier and evaluate the extent of the cross-media capabilities of a photograph-trained aesthetic classifier. Second, it allows to analyse and investigate emerging patterns and styles generated by film directors.

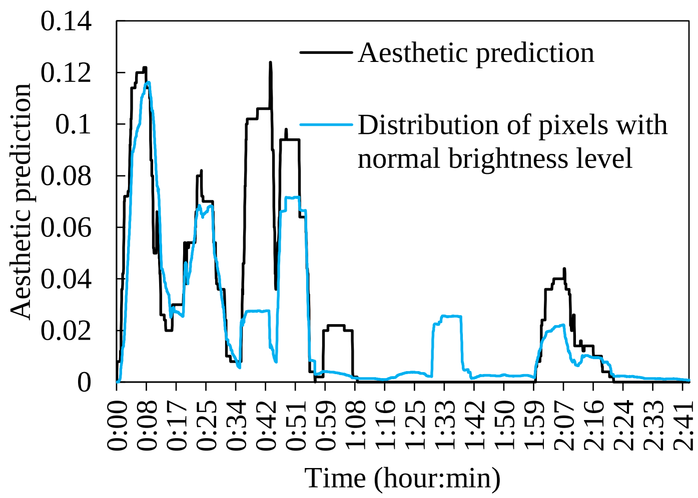


Figure 5.2: Plot displaying the correlation between brightness and aesthetic prediction in the film *The Hateful Eight* (2015).

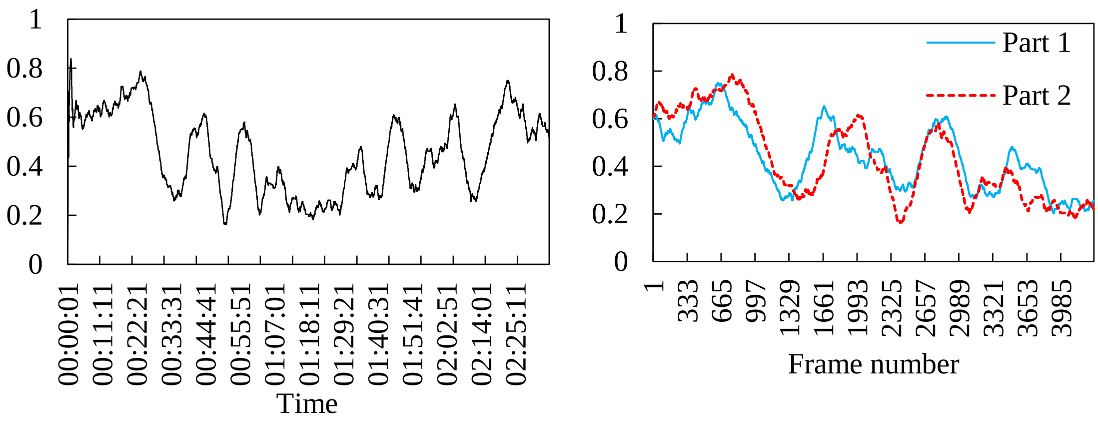


Figure 5.3: Aesthetic prediction over time, represented by a moving average over 500 frames. *Left*: Comparison of the aesthetic prediction over time (implied by frame number) of the first half of Django Unchained and the mirrored aesthetic prediction of the second part on the y-axis. *Right*: Aesthetic prediction of the film, Django Unchained (2012) directed by Quentin Tarantino.

5.4. Experiment 3: Aesthetic prediction on YouTube content creators

While the previous analyses on films are an indirect attempt at investigating patterns related to film directors' creative process (e.g. Quentin Tarantino), the abundant video content provided by the internet, and especially YouTube, allows to study this further. With some video content creators producing up to one video per day, it is now possible to look at aesthetic prediction of videos for one producer over time, and possibly over a career. For this task, the filmmaker and YouTube video creator Casey Neistat is selected due to being one of the first vloggers with over 650 vlogging videos online to this date and having a recognised interest in videography (Neistat, 2016). As shown in Figure 5.4, percentages of aesthetically high frames for each video vary through the different seasons of vlogging by Casey Neistat, but the average aesthetic remains steady over time with an average of 21% of aesthetically high frames per video. To provide ground truth to the aesthetic predictions, all 650 videos were manually labelled depending on whether the video was shot in studio or outdoor. Over the first season, 25% of the videos were entirely shot in studio and contain 26.4% of aesthetically high frames in average, against only 20% of aesthetically high frames for the rest of the videos which were shot spontaneously. The difference in the detection of aesthetically high frames between videos filmed in studio and on-the-go appears to be statistically significant ($p<.001$). Overall, videos reaching over 60% and up to 77% of aesthetically high frames were exclusively shot in studio with professional lighting and framing, emphasising the accuracy of the classifier on highly aesthetic videos. Surprisingly, no correlation exists between the percentage of aesthetically high frames categorised by the aesthetic classifier and any of the metadata variables available indicating the popularity of a video, such as the number of views or the average user rating. The measure is a satisfying but still fragile predictor for aesthetic video quality. It can be suggested that video content creators could use such tools to improve their contents by selecting particular sequences or shooting their videos in different conditions to match the suggested standards emerging from the aesthetic classifier's training mimicking a population's preferences.

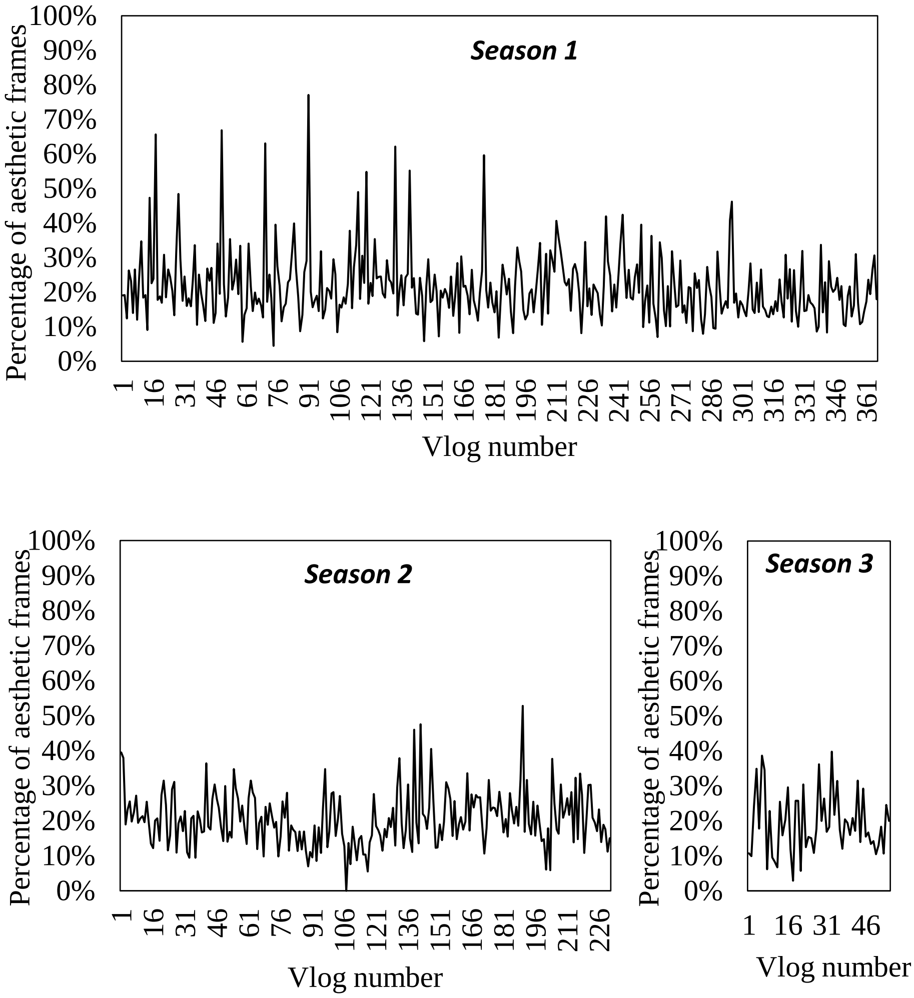


Figure 5.4: Percentages of aesthetically high frames in each video of the different seasons of vlogging by Casey Neistat.

5.5. Conclusion

This chapter demonstrates that a computational aesthetic classifier trained on photographs can also be used for classification of videos. Despite showing humble performances on the classification of videos, the results demonstrate the cross-media capabilities of the original classifier, particularly when focusing exclusively on the visual material while videos include audio which could provide valuable data. Moreover, processing frames individually and observing them as sequences across films allows to learn more about the content creator's decisions and the aesthetic classifier's preferences, which is a new approach in the domain of computational aesthetic prediction of videos. The first two tests in this chapter focused on the creative product as a whole and in its content respectively. The third experiment on vlogging videos establishes a new method to investigate video content creators over time. The approach offered by this chapter is novel due to its focus on meaningful decisions from the content creator rather than traditional computational classification tasks. One application of this approach is to support video content creators in decisions based on aesthetic criteria during the editing and postproduction process, giving them an immediate estimation of their audience's visual appreciation. Furthermore, it could be developed into assisting technology for visually impaired people willing to share their experiences and communicate through videos.

Chapter 6

Analysis of Visual Features in Paintings and Logos

6.1. Evolution of low-level visual features in influential painters' careers

In recent years, the contribution of computer vision to the understanding of visual art contemplation has shown rapid growth. Larger and larger datasets of photographs have been manipulated to investigate the influence of particular visual parameters over human aesthetic preferences (Datta et al., 2006; Murray et al., 2012; Romero et al., 2012). While datasets of photographs can provide a large amount of data, it is far more complicated to put together datasets composed of other visual media, such as films or paintings, as they generally require more resources and time to produce. The digitisation processes and sharing of films and paintings over the internet are not as straightforward as for photographs due to copyrights and file sizes, making datasets harder to assemble and distribute. Nonetheless, photographs associated with their human aesthetic ratings only allow the investigation of art contemplation.

Strongly linked to the phenomenon of contemplation, visual art creation has received little attention from the domain of computer vision. Some preliminary works have attempted to establish guidelines to assess paintings from technical and emotional point of views, but no outstanding results were found regarding key factors for aesthetic pleasantness specifically in paintings (Stork, 2009; Yanulevskaya et al., 2012). The current and quickly evolving trend in painting analysis using computer vision algorithms is aimed at automating artist or style classification of paintings, as well as generation of paintings according to a given artist or style (Arora & Elgammal, 2012; ING et al., 2016; Saleh et al., 2016).

In order to investigate the evolution of aesthetics in paintings, three main types of measures are used: gradient orientation distribution, curvature distribution, and colour distribution. Previous neuroscience and psychological experiments, as well as empirical art studies, have demonstrated significant preferences for cardinal lines (horizontal and vertical orientations) in terms of visual processing but possibly also in term of aesthetics (Blasdel, 1992; Chapman & Bonhoeffer, 1998; Girshick et al., 2011; Latto & Russell-Duff, 2002). Regarding the shape of edges, it has been shown that smooth curves were preferred over straight lines and angles (Bertamini et al., 2016; Munar et al., 2015). Studies looking at human preferences for colours have concluded a universal dislike for shades of orange and yellow, and a preference for shades of cyan, blue and purple (Ball, 1965; McManus et al., 1982; Ou et al., 2004a; Shimamura & Palmer, 2012). High levels of saturation and brightness have been demonstrated as preferred, regardless of the colour's hue (Camgö et al., 2002).

The hypothesis is that abstract painters have significantly more aesthetic changes along their careers compared to representational painters. The greater amplitude in changes may be due to the lack of constraints linked to the representation of real-world scenes, allowing for more experimentations. The second hypothesis investigated in this chapter is that visual features preferred by the human visual system (cardinal lines, curves, shades of blue) are used in significantly different quantities when influential abstract artists gain painting expertise or experiment with new styles, possibly emphasising a link between preferences in art contemplation and creation. Of course, this does not imply that all artists are consciously aiming at producing more aesthetically pleasing works but that the freedom of aesthetic experimentation leads to more aesthetically pleasant choices to the average observer. Moreover, while this study primarily appears to focus on expertise in artists, it also provides preliminary research in user perception using digitised artworks. This chapter attempts to reflect on the fact that many paintings were initially created to be experienced within a real-world environment and now find their audience in another context and format. To proceed, two new datasets of digitised paintings are conceived. The first dataset consists of influential abstract artists of the western world from the 20th century, while the second dataset is composed of landscape painters from the 19th century. Low-level features are extracted for each painting, using the extraction algorithms previously developed and described in Chapter 3. The features allow the comparison of aesthetic modification over the career of an artist.

6.2. Extracting low-level visual features from paintings

Inspired by features processed in the human visual system, algorithms have been developed to extract low-level visual information. It allows analysis from a basic aesthetic point of view while providing some level of abstraction from context and semantic by removing hints from spatial organisation. Originally, the visual features extractor was built for computational aesthetics research, meaning that the extracted visual information was used to categorise photographs depending on their aesthetics and determine possible preferences from the human visual system. The training process involved the AVA dataset, which includes over 250,000 photographs each provided with a rating from at least 100 people (Murray et al., 2012). However, in the specific case of paintings, it is complicated to obtain aesthetic ratings from a population over thousands of works due to their heavy cultural meanings and popularity causing a bias. The extracted visual features are analysed to widen the understanding of aesthetic preferences when acquiring painting expertise.

An algorithm based on the Histogram of Oriented Gradients (HOG) algorithm is used to detect the dominant orientation (in degrees from horizontal) in each area (shaped as multiple 8x8 pixel squares organised as a grid) of an image, before calculating the percentage distribution for each orientation (out of 32 possible orientations detected with ~5.6° accuracy)(Dalal & Triggs, 2005). With the HOG-based algorithm providing an estimated gradient orientation for each pixel, it is then used to calculate the difference in estimated orientation of pixels surrounding a given pixel. This can be considered as an approximation of curvature for a given pixel space. Once applied to all pixels in a pixel area, it allows to determine the dominant category of orientation change for the pixel area. Similarly to orientation distribution, this process is applied to each pixel area covering the image (organised as a grid of pixel squares), and the different categories of changes in orientation, occasionally called curves as a simplification, are distributed into 16 bins before calculating the percentage distribution for each category.

Colour distribution was simply processed by extracting the hue, saturation, and brightness (HSB) values of each pixel in an image. For each image, hue, saturation and brightness distributions are then calculated and shared between 20 bins. By storing and analysing hue, saturation and brightness separately, results can easily be compared to previous empirical findings. However, information is lost regarding the relationships between these 3 characteristics, meaning that it is impossible to know if specific levels of saturation and brightness are favoured by artists for a given hue, and vice versa.

6.3. Preliminary tests: use of cardinal lines over an artist's career

Pilot tests are run on datasets consisting of digitised paintings by popular artists, Wassily Kandinsky and Pablo Picasso. Their careers are different on many levels, Kandinsky being professionally active for 48 years while Picasso dedicated his entire life to it, totalling 83 active years, with respectively 502 and 1065 paintings available and dated online. Kandinsky's paintings were downloaded from wassilykandinsky.net, and Picasso's works were retrieved from wikiart.org ("WassilyKandinsky.net," 2018; "WikiArt," 2018).

Tests on Wassily Kandinsky's work demonstrates an increase in the presence of cardinal orientation of lines in his paintings, which is not observed in Picasso's work (Figure 6.1). The use of cardinal lines by Picasso fluctuates between 15% and 25% while Kandinsky shows a progression from 20% to up to 38%. Picasso had a longer career with more renowned works of art, which may bias the comparison as the accuracy of each data point relies on those two parameters. The developing style of the artist can also be an important source of aesthetic changes, meaning drawing any conclusion on a larger scale is not possible at this stage. To investigate the broader picture, two datasets are, therefore, assembled with two-dimensional visual arts by many renowned artists from the 19-20th century era of western tradition.

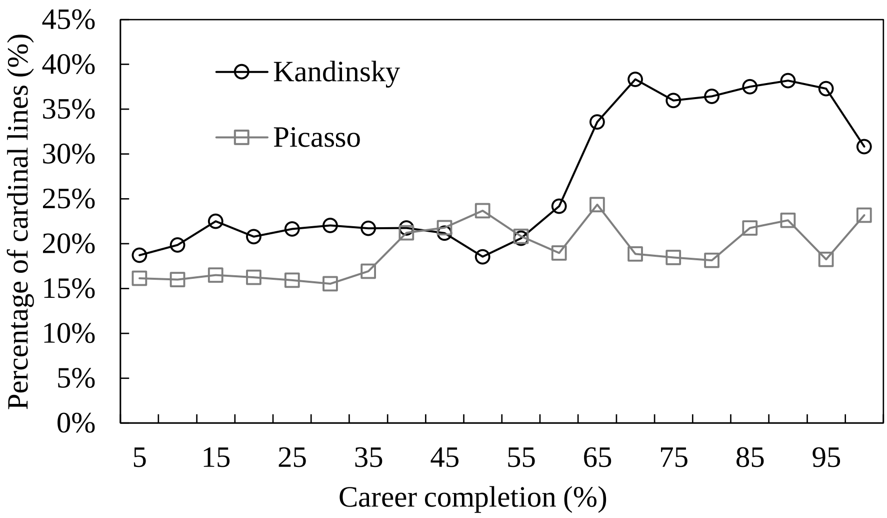


Figure 6.1: Percentage of lines with cardinal orientations in paintings by Wassily Kandinsky and Pablo Picasso through their respective career. Each plot starts at the beginning of their career and stops at their last painting, meaning that the data points are not plotted over the same time frame in real life. The number of paintings for each data point also varies due to the different numbers of total paintings for each artist.

6.4. Data filtering

When building a dataset of digitised artworks, one of the main issues is that the selected period profoundly influences the quality of the digitisation process. It is partly due to the varying preservation of the works but also to the attention received and technology used during the scanning process. The selected period must include a large number of artists with overlapping active years, who produced many paintings in order to obtain significant results from the experiment. Moreover, it also requires a broad enough time span for the artists to modify their painting styles or expertise level. The technology involved in the colouration of painting and the making of canvases have also changed dramatically over the centuries, with some colours being more expensive to produce (Barnett et al., 2006). Regarding art movements, influential artists in representational arts and abstract arts are selected due to the movements being highly different and to the proximity of both time periods. It allows to control for any bias caused by the nature of the subjects represented in works of art.

Studying the evolution of a phenomenon taking place over a lifetime, mental and physical health problems must be controlled as they can significantly alter artistic work and progression. Mental and physical health problems have a demonstrated impact on the visual content of an artist's work (Forsythe et al., 2017). Therefore, artists with diagnosed diseases known to have modified their artistic process are discarded in order to not influence the outcome of this study. For instance, Van Gogh's work has been demonstrated to be highly influenced by his mental health difficulties (Aragón et al., 2008). Experts have also debated over the possible impact of Alzheimer's disease on Willem De Kooning's work. Some of his most renowned paintings were actually created after he had been diagnosed with Alzheimer's disease (B. L. Miller et al., 1998; B. L. Miller & Hou, 2004). Henri Matisse is discarded from the list of studied artists due to the fact that he was left seriously weakened by cancer. The disease triggered the production of his collages, such as *The Snail*, thus being an example of physical health issues that altered an artist's style (Stirling, 1993). Nonetheless, Salvador Dalí, who had Parkinson's disease in his final years, is kept in the dataset due to the proportionally small number of paintings from this period present in the dataset (Gibson, 1997). The symptoms appeared only in 1980 while the last painting available on wikiart.org dated from 1983. Despite controlling for diseases that could influence artists in their practice development, artists diagnosed with depression are not excluded from the dataset as there is a known history of depression among artistic populations and the disease has no direct physiological effect on the human visual system (Sussman, 2007).

Regardless of time periods and health issues, complementary criteria are needed to label an artist as influential or successful. While the Google Search Ranking allowed us to establish the initial list of "famous abstract painters", abstract artists were required to be recognised by the Museum of Modern Arts (MoMA), in New York, United States. Of course, not all abstract artists fitting the multiple criteria had online digitised paintings available due to copyright issues, or to an extremely low number of available works of art on wikiart.org ("WikiArt," 2018).

6.5. Dataset composition

Two datasets representing abstract and landscape painters are assembled. The first dataset includes 50 famous abstract painters of the western world from the 20th century and a total of 6,348 dated paintings (Table 6.1). With an average life expectancy of 70 years and average career length of 44 years, the first artist started his career in 1890 while the latest active artist of the dataset was until 2011, therefore covering a time window of 121 years. The number of paintings per artist ranges from 17 to 1067 with a median of 63. A second dataset composed of landscape painters from the 19th century is created to offer a comparison with artists focusing on real-world aesthetics (Table 6.2). 3,312 paintings from 33 different artists were retrieved, with a median of 64 paintings per artist and between 16 and 446 paintings per artist. This second dataset is not covering the same period as the first dataset due to most of the influential landscape painters with digitised works available online living in the 19th century. With an average life expectancy of 68 years and average career length of 38 years, the first artist started their career in 1820 while the latest active artist of the dataset was until 1942, which covers 122 years of landscape painting. Hence, both datasets present relatively similar parameters despite not being in the same time period. They also contain a significant majority of male painters of 90% and 97% for the dataset of abstract and landscape painters, respectively.

Table 6.1: Abstract artists present in the dataset.

Name	Number of paintings	Birth	Death	Career Start	Career End	Art movement
Alfred Manessier	48	1911	1993	1935	1982	Art Informel
Andy Warhol	134	1928	1987	1954	1987	Pop Art
Antoni Tapies	54	1923	2012	1945	2011	Art Informel
Arshile Gorky	62	1904	1948	1922	1948	Surrealism
Asger Jorn	38	1914	1973	1940	1972	Art Informel
Barnett Newman	76	1905	1970	1944	1970	Abs. Expressionism
Clyfford Still	45	1904	1980	1934	1976	Abs. Expressionism
Cy Twombly	98	1928	2011	1948	2011	Abs. Expressionism
Elaine De Kooning	33	1918	1989	1917	1964	Abs. Expressionism
Fernand Léger	315	1881	1955	1900	1955	Cubism
Francis Picabia	100	1879	1953	1898	1951	Dada
Franz Kline	20	1910	1962	1940	1958	Abs. Expressionism
Franz Marc	115	1880	1916	1902	1914	Expressionism
Giacomo Balla	64	1871	1958	1900	1925	Futurism
Hans Hartung	21	1904	1989	1921	1989	Tachisme
Hans Hofmann	100	1880	1966	1902	1965	Abs. Expressionism
Jackson Pollock	78	1912	1956	1934	1953	Abs. Expressionism
Jean Arp	43	1886	1966	1912	1966	Abstract Art
Jean Dubuffet	53	1901	1985	1942	1985	Haute Pâte
Jean Michel Basquiat	134	1960	1988	1980	1988	Neo-Expressionism
Joan Miro	186	1893	1983	1912	1983	Surrealism
Joan Mitchell	62	1925	1992	1950	1992	Abs. Expressionism
Josef Albers	53	1888	1976	1915	1976	Constructivism
Karel Appel	22	1921	2006	1946	2005	Art Informel
Kazimir Malevich	291	1879	1935	1900	1934	Suprematism
Lee Krasner	22	1908	1984	1938	1974	Abs. Expressionism
Louis Marcoussis	45	1878	1941	1914	1941	Cubism
Marcel Duchamp	78	1887	1968	1901	1968	Dada
Mark Rothko	131	1903	1970	1925	1970	Abs. Expressionism
Marsden Hartley	31	1877	1943	1908	1943	Abstract Art
Max Ernst	343	1891	1976	1909	1975	Dada
Morris Louis	110	1912	1962	1948	1962	Abs. Expressionism
Natalia Goncharova	55	1881	1962	1900	1935	Cubo-Futurism
Nicolas De Stael	30	1914	1955	1947	1955	Art Informel
Pablo Picasso	1067	1881	1973	1890	1972	Cubism
Paul Klee	195	1879	1940	1903	1940	Expressionism
Philip Guston	55	1913	1980	1930	1980	Abs. Expressionism
Piet Mondrian	65	1872	1944	1892	1944	De Stijl
Robert Delaunay	24	1885	1941	1904	1940	Orphism(Simultanism)
Roberto Matta	17	1911	2002	1936	2002	Surrealism
Roy Lichtenstein	112	1923	1997	1951	1997	Pop Art
Salvador Dali	1047	1904	1989	1917	1983	Surrealism
Sonia Delaunay	43	1885	1979	1907	1972	Orphism(Simultanism)
Stuart Davis	19	1892	1964	1912	1964	Cubism
Theo Van Doesburg	152	1883	1931	1899	1931	De Stijl
Umberto Boccioni	83	1882	1916	1902	1916	Futurism
Wassily Kandinsky	216	1866	1944	1896	1944	Expressionism
Willi Baumeister	82	1889	1955	1905	1955	Abstract Art
William Scott	41	1913	1989	1938	1982	Abs. Impressionism
Wyndham Lewis	40	1882	1957	1897	1946	Vorticism

Table 6.2: Landscape painters present in the dataset.

Name	Number of paintings	Birth	Death	Career Start	Career End	Art movement
Albert Bierstadt	188	1830	1902	1850	1900	Romanticism
Aleksey Savrasov	221	1830	1897	1840	1894	Realism
Alfred Sisley	446	1839	1899	1865	1897	Impressionism
Anna Ostroumova-Lebedeva	81	1871	1955	1900	1942	Art Nouveau
Arkhip Kuindzhi	131	1842	1910	1869	1908	Realism
Armand Guillaumin	56	1841	1927	1867	1917	Impressionism
Charles-François Daubigny	60	1817	1878	1844	1878	Realism
David Bates	23	1840	1921	1873	1907	Realism
David Johnson	65	1827	1908	1851	1890	Romanticism
Eugene Von Guerard	73	1811	1901	1852	1882	Romanticism
Franklin Carmichael	20	1890	1945	1920	1939	Art Nouveau
Frederic Edwin Church	45	1826	1900	1847	1891	Romanticism
Fyodor Vasilyev	92	1850	1873	1863	1873	Realism
Gustave Loiseau	168	1865	1935	1889	1930	Post-Impressionism
Hans Heysen	20	1877	1968	1904	1929	Realism
Homer Watson	16	1855	1936	1879	1932	Realism
Isaac Levitan	406	1860	1900	1875	1900	Realism
Ivan Shishkin	370	1832	1898	1854	1898	Realism
James E. H. Macdonald	33	1873	1932	1909	1932	Art Nouveau
Jose Maria Velasco	112	1840	1912	1860	1911	Realism
Joseph Farquharson	32	1846	1935	1867	1915	Realism
Jules Dupré	18	1811	1889	1835	1870	Realism
Knud Baade	22	1808	1879	1828	1879	Romanticism
Martin Johnson Heade	38	1819	1904	1840	1904	Realism
Richard Parkes Bonington	39	1802	1828	1820	1828	Romanticism
Robert Julian Onderdonk	70	1882	1922	1901	1922	Impressionism
Theodore Clement Steele	64	1847	1926	1882	1922	Impressionism
Théodore Rousseau	79	1812	1867	1829	1867	Realism
Thomas Cole	127	1801	1848	1825	1848	Romanticism
Thomas Moran	36	1837	1926	1855	1926	Romanticism
Willard Metcalf	109	1858	1925	1877	1924	Impressionism
William Hart	29	1823	1894	1849	1881	Romanticism
William Leighton Leitch	23	1804	1883	1835	1882	Romanticism

6.6. Method

Visual features are extracted from all the paintings in each dataset. The visual features represent the distribution of gradient orientation, the distribution of changes in gradient orientation and HSB-encoded colours. As the number of paintings per artist varies, visual features are averaged for the paintings created in the first and second half of the career of each artist. This method attributes as much importance to each artist, regardless of the number of known paintings they produced. Splitting an artist's career into 2 periods allows for an accurate average of visual features in each period while smaller periods, similar to regression analysis, would mean highly variable values due to painters sometimes not producing any work for years. It is then possible to compare average visual features between the two halves of a career, the breakpoint being the year of the median of the first and last painting ever dated for a given artist (Equation 1).

In order to test for statistical significance of the different visual features related to human visual preferences, some extracted features are selected and categorised. A paired sample t-test is run for each category of features to compare the features used in the first and second half of the painters' careers. The two tested samples represent average features for both periods. The categories represent the percentage of cardinally oriented lines, straight lines, curved lines, shades of blue, shades of orange and yellow, desaturated colours and brighter colours. The category for cardinal lines incorporates extracted features representing horizontal and vertical lines. The categories for straight lines and curves are opposites, with all lines with an orientation change different from 0° considered as curves. The category for shades of blue include features representing hues between 198° and 252°, while the shades of orange and yellow consist of hues from 18° to 72°. The categories regarding colour saturation and brightness incorporate respectively pixels with a saturation greater than 0.90 and pixels with brightness over 0.90.

To finish, one of the reasons for using such lists of artists is due to the availability of the digitised works online, as well as their reputation. It is therefore essential to be careful with the results as the selected artists were extraordinarily successful and influential, and so, may not be representative of the average artist. Moreover, while the datasets of landscape and abstract painters are both composed of a large number of subjects, it is complicated to establish what exact painting styles and what distribution of painting styles would represent perfectly the average influential artist from the western world.

$$\delta_F = \frac{1}{N_a} \sum_{a=0}^{N_a} \left(\frac{1}{N_{a,e}} \sum_{i=0}^{N_{a,e}} F(P_{a,e,i}) - \frac{1}{N_{a,l}} \sum_{i=0}^{N_{a,l}} F(P_{a,l,i}) \right)$$

Equation 1. Equation describing the process of calculating the average difference for a given feature $\boldsymbol{\delta_F}$. $\boldsymbol{N_a}$: Number of artists; $\boldsymbol{N_{a,e}}$: Number of early paintings for an artist a; $\boldsymbol{N_{a,l}}$: Number of late paintings for an artist *a*; F(x): Feature for painting *x*; $\boldsymbol{P_{a,x,i}}$: Painting of the artist numbered a during the period *x* (*e*: early; *l*: late) and their painting numbered *i*.

6.7. Results

Before looking at the two datasets of paintings side by side, an additional comparison is set to evaluate the impact of the types of representation, which are photographs and paintings, on the presence of the studied visual features. Figure 6.2 illustrates how both datasets of landscape and abstract paintings contain different percentages of the particular sets of features related to visual preferences, with another set of 1,000 landscape photographs from the dataset by Datta et al. used as a comparison (Datta et al., 2006). While it is expected to have differences in low-level features distributions between photos and paintings of natural landscapes, it is surprising that abstract paintings have closer distributions to landscape photos than landscape paintings. It demonstrates that the differences in production and aesthetic rules in media, such as photographs and paintings, have a significant role in the distributions of low-level features, emphasising why comparison can only be made with datasets of the same visual medium.

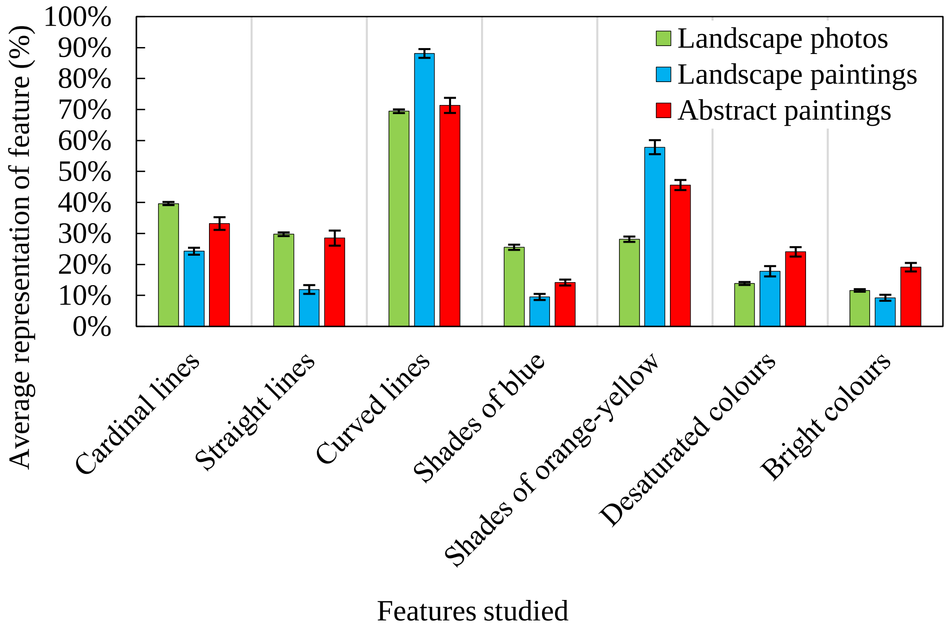


Figure 6.2: Average representation of specific features in landscape photos, landscape paintings, and abstract paintings.

While it has been shown that gradient orientations and the quantity of cardinal lines influence the aesthetic assessment of photographs, results similar to these findings are shown regarding the evolution of gradient orientations distribution in paintings over famous abstract painters' careers in Figure 6.3. The average differences in gradient orientation distribution over an abstract painter's career display a significant increase of 3.9% in vertical lines from early (M=.15, SD=.06) to late paintings (M=.19, SD=.08); $t(49)$=4.78, p<.001, d=0.54. It is paired with a significant increase of 3.3% in horizontal lines from early (M=.14, SD=.05) to late paintings (M=.17, SD=.07); t(49)=5.21, p<.001, d=0.52. Overall, there is a significant increase in cardinal lines use, displayed in Figure

6.8, between early (*M*=.30, *SD*=.10) and late paintings (*M*=.37, *SD*=.14); *t*(49)=5.53, *p*<.001, *d*=0.57. However, when comparing to landscape painters, no evolution is observed except for a small significant loss in vertical lines use (*t*(32)=2.71, *p*=.01, *d*=0.31). The current results support the hypothesis that artistic expertise in abstract visual art influences the usage of preferred visual features in paintings.

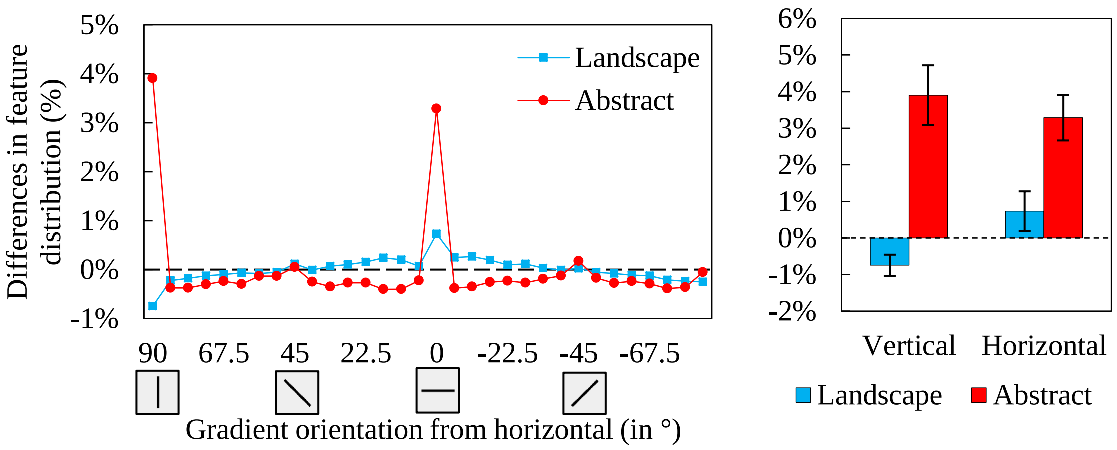


Figure 6.3: Differences in the distribution of gradient orientations between early and late works in landscape and abstract paintings. Left: Plot with all gradient orientations. Right: Plot for cardinal orientations with error bars representing standard error.

Designing a measure of curvature for digital images raised an issue regarding the scale of the visual representation. In a digital image, curves can be interpreted as a change in gradient orientation over a certain pixel area. A fixed size is selected for all local pixel areas observed within an image. Curvature distribution is calculated by detecting the dominant type of curvature in each local pixel area. This allows the comparison of items depending on how they are displayed on a screen, rather than depending on the scale of representation. Moreover, the pixel sizes of images between the two categories do not significantly vary for the average artist (*t*(49)=0.59, *p*=0.55). In Table 6.3, it is shown that existing image datasets used for aesthetic analysis in photographs are similar in pixel size. However, the average landscape and abstract paintings downloaded from wikiart.org contains 2.46 times more pixels on average. The defined local pixel area is then modified from 32x32pixels to 80x80 in order to make the results potentially comparable with other datasets in the future. The two different local pixel area sizes are both tested for comparison, but no significant change is noticed in terms of results.

Figure 6.4 displays results for abstract painters only, as no significant evolution was observed for landscape painters regarding curvature distribution. There is a 6.7% increase of flat angle (0° orientation change within the observed pixel area) and a 1.9% decrease of 48° orientation change within the observed pixel area over an average abstract artist's career. As illustrated in Figure 6.8 summarising major results, when comparing the progress in the use of straight lines and curves over a career, there is a significant increase of straight lines from early (M=.25, SD=.13) to late paintings (M=.32, SD=.18); $t(49)$=4.44, p<.001, d=0.49. There is also a significant decrease in the amount of curves in early (M=.75, SD=.13) and late paintings (M=.67, SD=.18); $t(49)$=4.50, p<.001, d=0.50. While the results contribute to the hypothesis that abstract painting expertise impacts the distributions of low-level visual features contained in works of art, the category of visual features displaying an increase in use is not known to be visually or aesthetically preferred by the human visual system.

Table 6.3. Description of datasets in terms of number of images and their average pixel size.

Dataset source	Number of images	Average width (pixels)	Average height (pixels)
Photo.net	17,453	611.62	567.46
DPChallenge.com	255,529	606.99	534.37
Wikiart.org	9,660	912.63	873.95

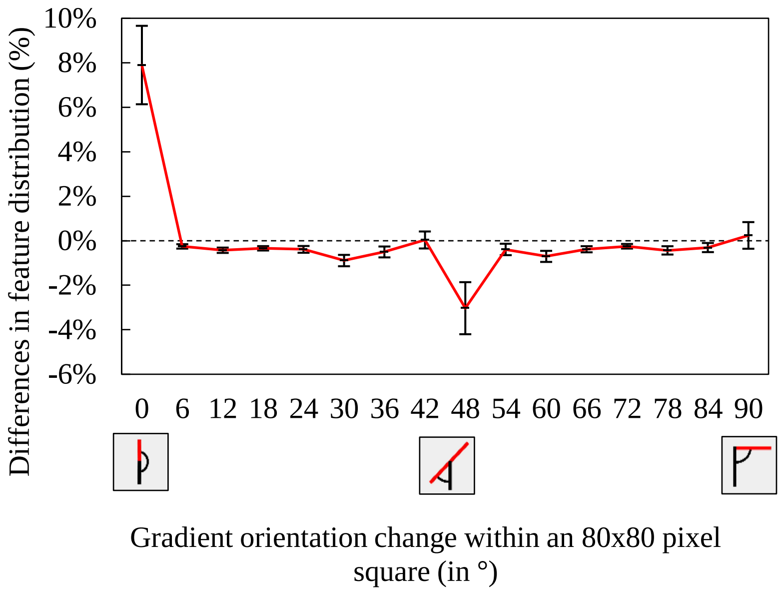


Figure 6.4: Differences in the distribution of gradient orientation changes between early and late abstract paintings. The pair of sets is composed of paintings of the beginning of famous abstract artists' careers against the end of their career.

In Figure 6.5, it can be observed that the amount of orange and yellow colours decreases by over 4.0% in landscape paintings and 6.1% in abstract paintings, while shades of blue are increased by 2.6% in landscape paintings and 4.4% in abstract paintings. As Figure 6.8 shows about abstract paintings, there is a significant increase between the amount of shades of blue in early (M=.12, SD=.06) and late paintings (M=.16, SD=.10); $t(49)$=2.67, p=.01, d=0.49. There is also a significant loss observed in the amount of shades of orange and yellow from early (M=.50, SD=.13) to late paintings (M=.43, SD=0.13); $t(49)$=2.80, p<.01, d=0.46.

With regard to saturation and brightness in landscape painting, no significant finding is observed. Meanwhile, in abstract works, no significant difference can be observed (p>.05) regarding high levels of saturation for colours (Figure 6.6). Colours with low saturation (below 0.10) are therefore tested, showing an increase approaching significance between early (M=.15, SD=.08) and late paintings (M=.19, SD=.10);

t(49)=2.32, *p*=.02, *d*=0.42. Moreover, a significant increase is displayed in the number of pixels with bright colours between early (*M*=0.16, *SD*=0.09) and late paintings (*M*=.21, *SD*=.12); *t*(49)=2.79, *p*<.01, *d*=0.45 (Figure 6.7). To summarise, hue and brightness evolutions over time follow the predictions. The only significant result for saturation is a rise for highly desaturated colours over time, leading to more pixels belonging to the black and white spectrum. In regards to the effect size represented by Cohen's *d*, all significant results also display effect sizes that can be defined as medium-sized, according to Cohen's guidelines (J. Cohen, 1988).

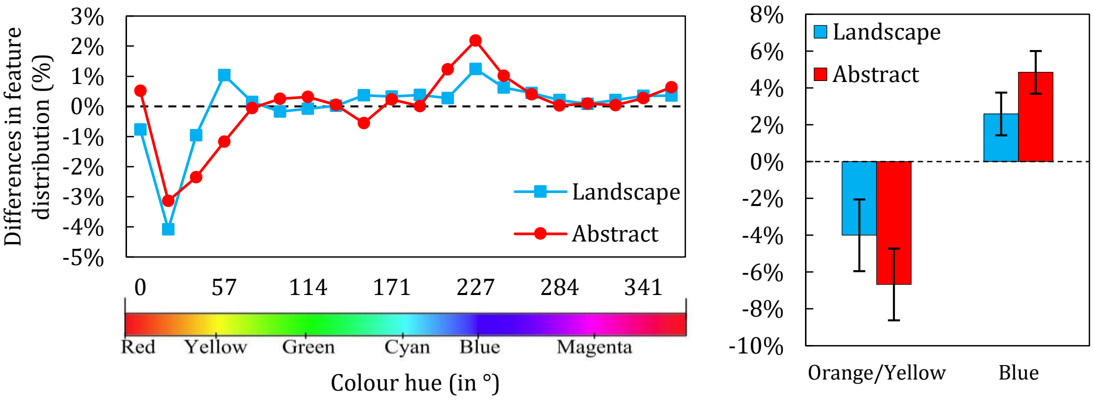


Figure 6.5: Differences in the distribution of colour hue between early and late works in landscape and abstract painting. Left: Plot with all colour hues. Right: Plot for shades of orange/yellow and shades of blue with error bars representing standard error. The colour hue scale given is set with constant saturation=1 and brightness=1.

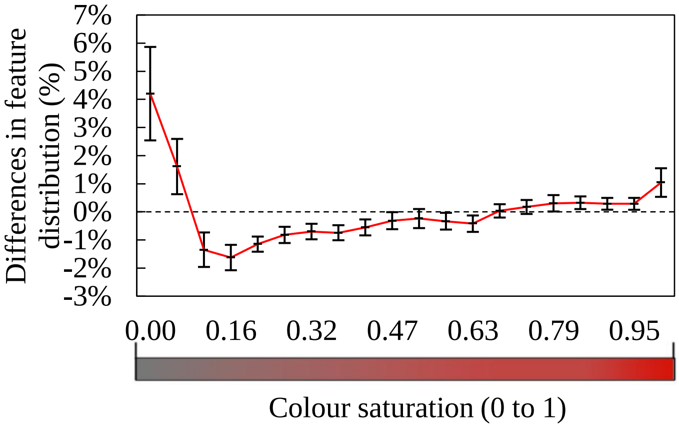


Figure 6.6: Differences in the distribution of colour saturation between early and late abstract paintings. The colour saturation scale given is set with constant hue=0 and brightness=0.6.

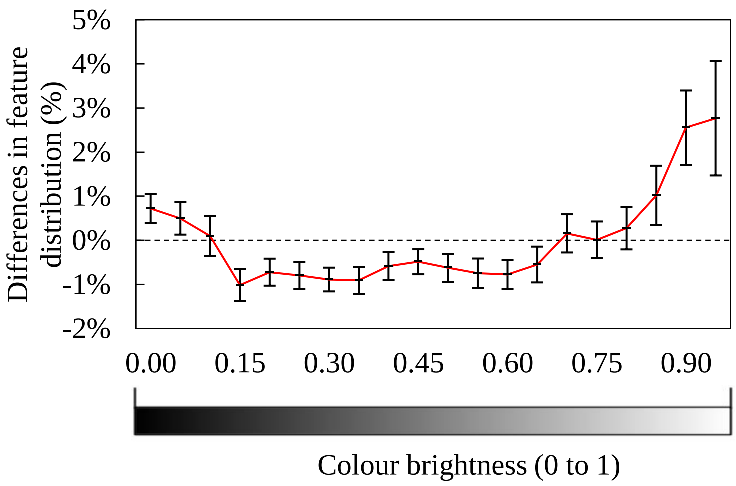


Figure 6.7: Differences in the distribution of colour brightness between early and late abstract paintings. The colour brightness scale given is set with constant hue=0 and saturation=0.

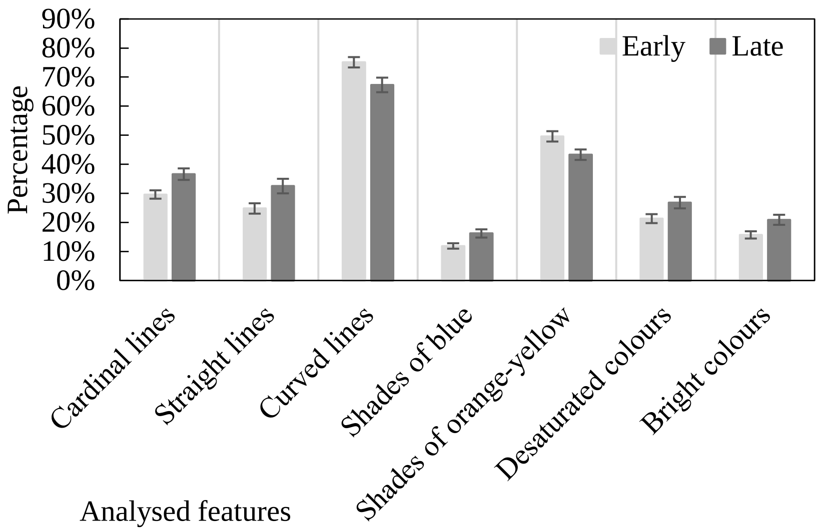


Figure 6.8: Percentage of use of different visual features in an average painting by a 20th-century abstract artist. All early-late differences displayed are statistically significant. Error bars represent standard error.

6.8. Discussion

Overall, the results show that landscape painters have little aesthetic changes in their painting style across their careers, which is in clear contrast with the progression of abstract painters. The aesthetic experimentations in paintings by influential abstract artists across their careers match some of the common human visual preferences pointed out by previous neuroscience, psychology, psychophysics and empirical arts studies. While orientation preferences, colour hues and colour brightness align with previous findings, it is more complicated to interpret the results of curvature distribution and colour saturation along the careers of abstract artists (Bertamini et al., 2016; Blasdel, 1992; Girshick et al., 2011; McManus et al., 1982; Munar et al., 2015; Ou et al., 2004a). Flat angles and 48° changes in orientation within an 80x80 pixel square are the two dominant types of orientation changes overall. The increasing amount of straight lines coupled to the loss of curved lines may go against the previous suggestion that aesthetic experimentations by abstract artists are purely based on aesthetic pleasantness.

Considering that the extracted features are measured as a percentage over the surface of an image, curved lines represent a high percentage of all lines as observed in Figure 6.8. This can imply that aesthetic experimentations do not always consist of increasing the amount of preferred visual elements. Otherwise, it may mean that artists converge towards a distribution of straight lines and curved lines being more balanced and therefore more pleasing to the eye.

The relationship between abstract painting expertise and human visual preferences can be translated in two ways. Firstly, the development and expertise in the practice of abstract painting allow the creation of works of art with a greater quantity of specific aesthetically pleasant features. Secondly, it could mean that some kind of feedback process allowed the selected artists to be more consciously aware of the visual preferences of their audiences. As demonstrated in a recent study looking at the relationship between colour composition in paintings and their observers' preferences, when asked to select missing colours of a pre-existent painting, observers unknowingly select the same colour as the original artist despite seeing the paintings for the first time (Nascimento et al., 2017). This study also shows that it was true even in abstract paintings with unnatural colour composition, underlining the fact that expert abstract painters manage to agree with lay people and visually please their audience. While the development of such skill is expected, it is surprising that a dataset of 50 of the most successful abstract artists displays such compliance to the potential preferences of their audience.

The dataset also plays a vital role for understanding the results, as it was built to list two-dimensional visual arts created by humans, with, in the case of abstract painters, works of art distinguishing themselves from real-world representations. In consequence, it proves that evolution of visual features over time is not directly linked to external perturbations from the outside world or technical innovation in representational arts (e.g., new techniques for perspective drawing or human proportions). It is also not common practice for abstract painters to be especially focused on aesthetic quality, meaning that the evolution of visual features may be a by-product of their painting expertise or the development of their practice.

Regardless of all conclusions in respect of human perception, the outcome of this computational experiment could help to design an artificially intelligent system estimating the level of expertise of the author of some visual content, whether it is a physical or a digital work of art. Moreover, it could also provide guidelines to creative computational systems and eventually provide a new way of testing whether those systems are gaining expertise or experimenting new styles by checking if the visual features' evolution match human aesthetic preferences.

6.9. Applying low-level feature analysis on brand logos' history

After leading investigations about the important factors in aesthetic judgement of photographs and the evolution of low-level features in abstract and landscape paintings through an artist's career, a similar low-level feature analysis is created with brand logos. Logos are designed by companies to display the personality of the brand, by including its name in a stylised font and often accompanied by a symbol. To clarify on the terminology and the studied objects in this paper, the word "logo" is used to define the brand name in stylised letters, as well as any additional symbol (Brownlee, 2014). The designs are meant to be functional to fit on diverse backgrounds and supports while attracting the attention of the audience and staying distinguishable to increase the value of the brand and the performance of the company (Park et al., 2013; Schechter, 2010). As companies get a wider audience and spread around the world, more abstract ways of displaying brands are required for countries with different alphabets, leading to the popularisation of symbols representing brands. In recent years, logos have been subjects to more scrutinised designing process, fed by studies looking at the effects of aesthetics on consumer perception. Pieters et al. present a study of complexity in advertising and introduced two categories which are feature complexity and design complexity (Pieters et al., 2010). Feature complexity is about colours and edges while design complexity is about global structures such as symmetry. In accordance with the trend, Pieters et al. demonstrate that feature complexity reduces likability and attention while design complexity was correlated with appreciation. Grinsven et al. complete further testing on design complexity and finds that simple logos are easier to recognise on the short term, but more complex logos end up more likeable on the long-term (van Grinsven & Das, 2016). One particular study explores how colour hue, saturation, and brightness influence brand personality (Labrecque & Milne, 2012). It exposes that some colours score exceptionally high for some brand personality. For example, red scores high for excitement, blue for competence, black for sophistication and brown for ruggedness.

One of the main issue when attempting to compare the evolution of logos of different companies is that the numbers of logo changes across time vary, as well as the date of logo changes. Therefore, some old companies developed in the 19th century such a Skoda have known only 6 logos in the company's history, in contrast to companies such as Smart, set up in 1993, and with the exact same number of logo changes. This issue arises due to significantly more funding being dedicated to neuromarketing since the 1990s. It is, therefore, logical to have more logo changes as companies are relying on science to improve their image through advertisement.

6.9.1. Dataset

Digitised logos from 55 companies are retrieved from the internet with a total of 399 items with an average size of 198 by 140 pixels (Table 6.4). All logos are provided with the year they came into use, while companies without accurate entry dates for their logos are discarded. Companies have an average of 7.25 logos across their history. While no existing dataset was available and ancient logos are not widely published on the internet or in books, studied companies are primarily selected due to the availability of the data. In the case where a company presents several options of alphabets for their logos, the Latin alphabet version of the logo is used. When multiple logos (logotype, logomark or combined version) were available, the version containing the brand name and a symbol is selected as it contains more visual content and is more likely to express the aesthetic chart of the company.

6.9.2. Method

The method consists of establishing two periods for which features present in logos are averaged, to then analyse the difference between the two periods. Two different types of tests are run. The first one consists of splitting data using the median year of each company's history. The median year corresponds to the year halfway between the first and the last logo update. Following this method, the average split year for all companies is 1968. As additional information, the very first logo in the list was established in 1850, and the latest one was produced in 2016.

A second test, aiming at highlighting eventual effects from neuromarketing studies, computer use and the internet in the 1990s, split the data on the year 1990. The split year being later than in the first method, some differences can be expected in the presence of features due to the use of computers in logo design. Before 1990, the studied companies have had 4.82 logo changes on average, against only 2.44 between 1990 and the present days. Of course, all companies studied had logos established before and after the 1990 split to make the comparison fair between the two tests.

When talking about both conditions, logos before the year split are defined as "early" while logos updated after the year split are defined as "late".

6.9.3. Results

In the first condition, no statistical significance is found in the use of cardinal lines despite the differences observed in Figure 6.9. Regarding Figure 6.10, the changes in gradient orientation are not shown to be statistically significant. In terms of colour hues which are represented in Figure 6.11, a significant decrease in red between early (M=66.15, SD=25.99) and late logos (M=56.78, SD=21.53) is observed ($t(54)$=2.04, p<.05, d=0.40). It is intriguing that another shade of red, itself represented on the end of the colour hue scale, significantly increases between early (M=2.8, SD=6.12) and late logos (M=6.37, SD=11.61) ($t(54)$=2.0, p<.05, d=0.38). An increase in shades of blue (from 189 to 245) between early logos (M=8.43,SD=11.96) and late logos (M=14.59,SD=14.46) is observed and statistically significant ($t(54)$=2.41, p=.02, d=0.46). Regarding saturation (Figure 6.12), both extremes on the saturation scale seem to display changes, but the evolution of the feature representing low-saturated colours is not statistically significant. The feature representing highly saturated pixel colours increases between early (M=3.37,SD=5.30) and late logos (M=6.8,SD=7.83) ($t(54)$=2.67, p<.01, d=0.52). To finish, no significant differences in brightness is detected, as illustrated in Figure 6.13.

In the second condition, none of the orientation and curvature preferences observed display statistical significance. Again, as in the first condition, statistical significance for colour hues are similar. The presence of red pixels decreases from early

(*M*=64.24,*SD*=24.54) to late logos (*M*=55.10,*SD*=20.80) ($t(54)=2.08$, $p<.05$, d=40). Shades of blue also increase from early (*M*=9.43,*SD*=11.71) to late logos (*M*=16.44,*SD*=16.12) ($t(54)=2.59$, $p=.01$, $d=0.49$). The feature representing highly saturated pixel colours increases between early (*M*=3.84,*SD*=4.93) and late logos (*M*=8.33,*SD*=9.81) ($t(54)=3.01$, $p<.01$, $d=0.57$). Surprisingly, the only difference obtained is in the percentage of dark pixels, with a decrease between early (*M*=6.41, *SD*=7.52) and late logos (*M*=2.99, *SD*=5.20) ($t(54)=2.74$, $p<.01$, $d=0.52$).

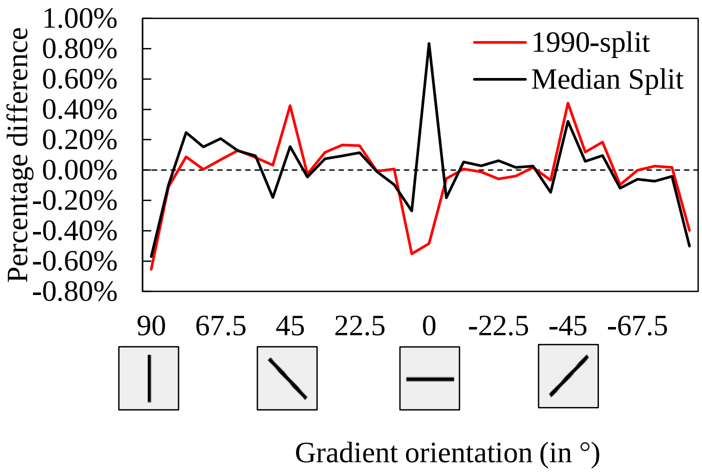


Figure 6.9: Evolution of gradient orientations in logos over two different time settings.

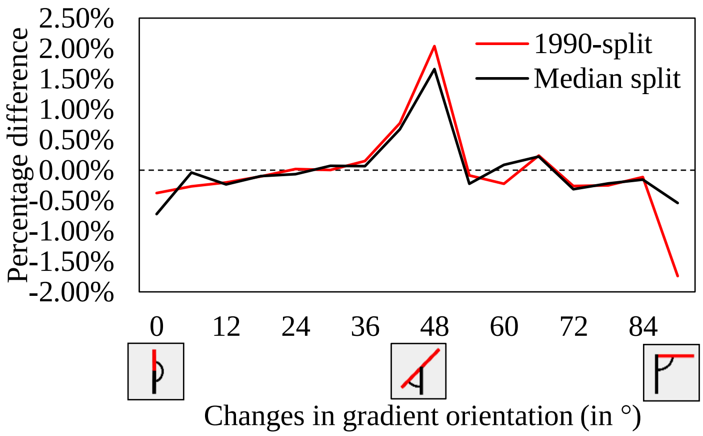


Figure 6.10: Evolution of gradient orientation changes in logos over two different time settings.

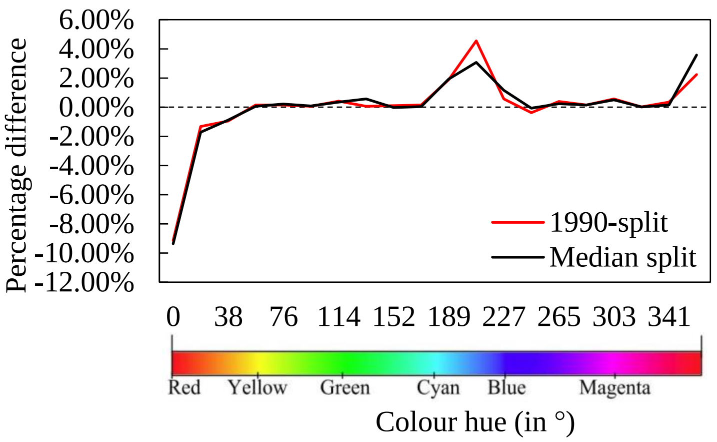


Figure 6.11: Evolution of colour hues in logos over two different time settings.

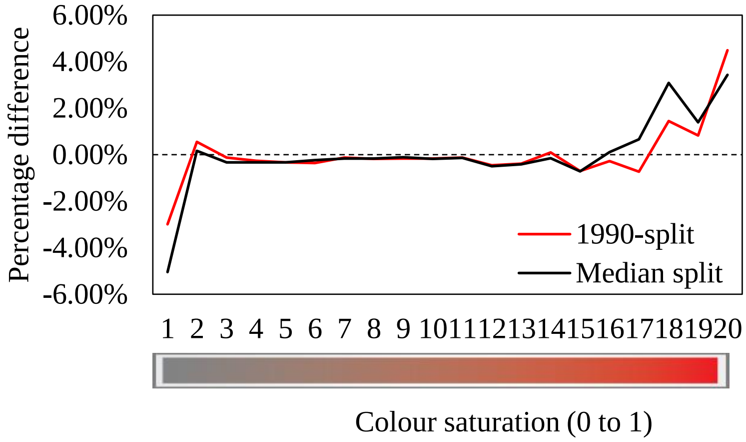


Figure 6.12: Evolution of colour saturation in logos over two different time settings.

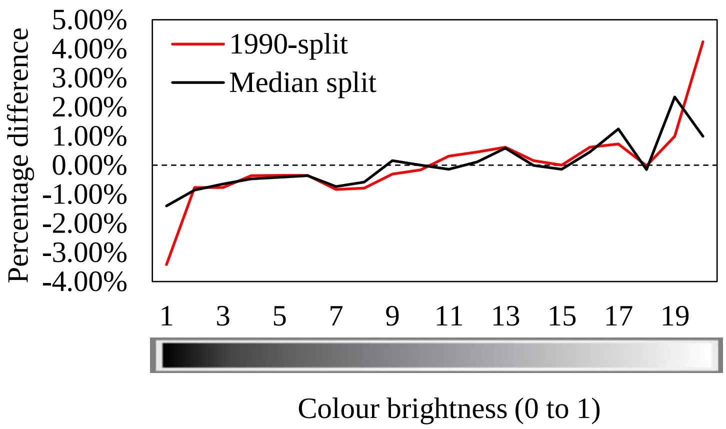


Figure 6.13: Evolution of colour brightness in logos over two different time settings.

Table 6.4: List of companies and brands studied.

Company Name	First published logo	Last published logo	Number of logos
Adidas	1949	2005	4
AmericanAirlines	1934	2013	5
Apple	1976	2014	7
ATT	1889	2005	8
Audi	1932	2009	2
BMW	1916	2000	6
Bouygues	1972	2015	3
BP	1921	2000	7
Buick	1904	2002	15
Chevrolet	1911	2011	9
CocaCola	1886	2007	12
DELL	1984	2010	3
DisneyChannel	1983	2014	8
Doritos	1964	2013	8
Dove	1955	2003	4
Fanta	1940	2016	9
Fiat	1899	2006	10
Ford	1903	2003	8
GeneralElectric	1892	2004	8
HP	1954	2012	5
Intel	1968	2005	2
KFC	1952	2006	5
KLM	1919	2011	9
Kodak	1907	2006	7
Lays	1965	2007	5
Lego	1934	1998	11
Mazda	1934	1998	7
McDonald's	1940	2003	7
Mercedes	1902	2011	8
Microsoft	1975	2014	6
MountainDew	1950	2008	5
Nestle	1868	1995	5
Nike	1971	1995	4
Nokia	1865	1992	5
Opel	1900	2002	9
Pepsi	1898	2008	11
Peugeot	1850	2010	10
Reebok	1958	2008	8
Renault	1900	2015	11
Seat	1950	2012	7
Shell	1900	1999	10
Siemens	1899	2001	9
Skoda	1895	2011	6
Starbucks	1971	2011	4
Target	1902	2004	8
Total	1954	2003	5
Toyota	1955	2012	14
Umbro	1924	1994	8
UPS	1916	2003	4
Visa	1976	2014	5
Vodafone	1984	2006	4
Volswagen	1937	2012	12
Walmart	1962	2008	6
Xerox	1906	2008	11
Yamaha	1898	2016	10

6.9.4. Discussion on low-level features in logo designing

Comparing to the evolution of the distribution of the selected low-level features in abstract painter's careers, it is surprising to see so little differences over 55 companies across over a century of logo designing. Despite a recent tendency to simplify logos to improve recognition, it does not appear that the distribution of edges in logos has changed much. Only colour hues and saturation levels display alterations in both conditions, with darkness level significantly decreasing in the second condition underlining modification in the last couple of decades. Both conditions show a sharp decrease in shades of red and a small but significant increase in blue. Knowing that red is matched with excitement for consumers and blue with trust, it could be explained by the fact that companies adopt aesthetic charts on a long-term approach instead of looking for consumers to buy impulsively (Labrecque & Milne, 2012). Due to the HSB colour information being decomposed to make the analysis more manageable, it is not possible to verify if the decreasing shade of red characterised by the 0° on the hue scale is typically saturated and bright. It could, therefore, mean that the decrease in red is correlated with low-saturated colours. Indeed, the decrease in red hue is by about 10%, which is far higher than the significant drop in saturation and brightness. A strong and significant correlation between the feature representing red and the feature representing low saturation is found in the images; (r=.87,p<.001).

While previous papers looked at preferences of low-level visual features in photographs, videos, and paintings, investigating such evolution of visual features and applying a similar approach to logos is complicated for many reasons. First, building the dataset from scratch is influenced by the availability of the data, and it is possible that studied companies are the ones with the highest investment in public outreach, meaning a particular focus on successful marketing strategies rather than the global evolution of low-level visual features across all logos. Moreover, logos often being conceived by design teams instead of individual artists and each iteration being treated as an independent project, it is essential to keep in mind that this computational experiment does not investigate into individual visual preferences in the creative process but rather in the society's visual preferences in the creative process.

6.10. Conclusion

These two computational experiments offer insight into how artistic expertise and aesthetics evolve over time in popular culture. The results contribute to the hypothesis that visual art creation in influential abstract artists is subject to influences from human visual and aesthetic preferences, despite an unnatural context where aesthetic pleasantness is not always the primary goal. While most visual features preferred by humans seem to appear in more significant quantities as a painter gains experience and experiment new styles, the evolution of the number of curved lines and desaturated colours goes against this hypothesis. Preferred features that do not evolve as expected over the artists' careers are already present in high quantities in early career. It can be suggested that visually preferred features do not maintain their aesthetic attribute past a specified distribution and that artists look for a balance between aesthetically pleasing and less stimulating visual features. Future works could also include different artistic movements or offer a comparison with oriental artists to strengthen the findings while highlighting the effects of cultural background on visual artists. While the results on logo designing do not allow to make any definite conclusion, it is interesting to see how little significant changes have been achieved in the domain in terms of low-level aesthetics.

Chapter 7

Proof of Concept: Psychological Games and Real-World Applications with an Aesthetic Classifier

7.1. Human performances on aesthetic classification tasks

To comprehend the order of magnitude of the aesthetic classifier's performances, a brief experiment is set up to compare an aesthetic classifier with human performances. An aesthetic classifier is trained using 225,000 images of the AVA dataset (Murray et al., 2012). As stated in previous chapters, each image is rated by at least 100 people according to its aesthetics. Images are classified as aesthetically high or aesthetically low depending on the average of the community's ratings and 25,000 images are kept as test images and are not shown to the aesthetic classifier. Participants are shown 20 images from the test set and asked to match the rating community's opinion. For each image, feedback is given with both the answers from the aesthetic classifier and the community, as displayed in Figure 7.1. After the last answer is given, the scores for both the participant and the aesthetic classifier are disclosed, as shown in Figure 7.2. This pilot experiment was run during Off The Lip 2016 in Plymouth, UK, and 10 volunteers participated while being engaged in a discussion over the different implications of automated aesthetic classification. Off The Lip 2016 was a public engagement event that aimed at introducing to the public the research produced by fellows and investigators of the CogNovo doctoral programme.

Figure 7.1: Screenshot of the interface used in the classification task for human participants to challenge the aesthetic classifier at Off The Lip 2016, Plymouth University, United Kingdom.

Despite the challenge between human and machine only being a pilot, the results shown in Table 7.1 lead to believe that the classifier performs at similar levels as the average person. However, it is important to keep in mind that the participants were randomly selected and volunteered, meaning they may not have any expertise in aesthetics appreciation but still express interest in such a task. Also, the results displayed in Table 7.1 are slightly higher than previously reported, which is assumed to be due to a sample too small to be fully representative. Therefore, it would be interesting to control for art expertise in future iterations of such experiments. Nevertheless, when looking at performances on individual images, human participants appear better at classifying images incorporating aesthetic criteria not treated by the classifier. For example, it was observed that some aesthetically good images with blur and reflections are rightly classified by people but are not appreciated by the aesthetic classifier.

Table 7.1: Correct classification rates by the average human participant and the aesthetic classifier on images issued from different parts of the rating distribution

	80% of images around the median rating	**10% of images in the top and bottom of the rating distribution**
Average participant	64.20%	76.32%
Aesthetic classifier	66.05%	84.21%

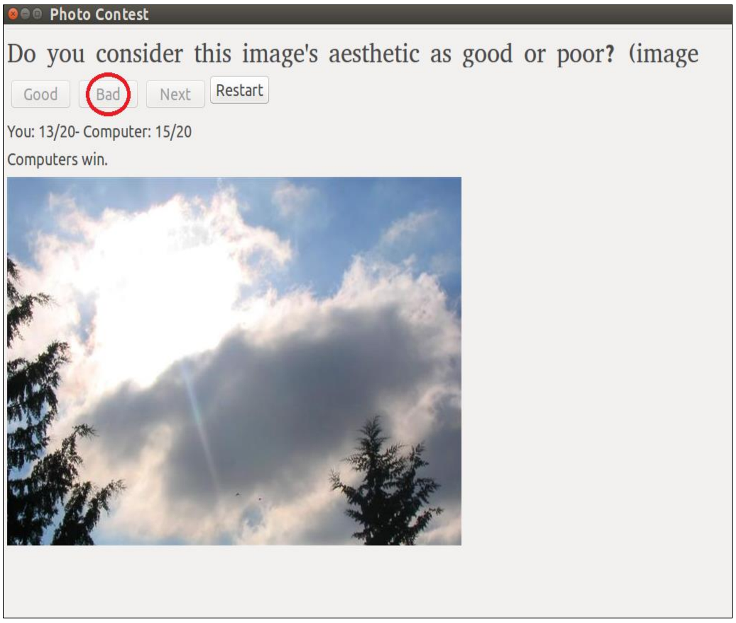


Figure 7.2: Screenshot of the final page displaying the participant and the aesthetic classifier's scores, compared to the DPChallenge.com community's aesthetic judgements.

7.2.PikPik: Android app for aesthetic filtering of photographs

The computational aesthetic research conducted throughout the thesis has been adapted into an Android app. The smartphone application was an entry for an artificial intelligence contest organised by Qualcomm. The app, named PikPik (for "Picture Picker"), aims at reducing the long chore of sorting photographs on smartphones by applying visual preferences learnt from the AVA dataset, but also by trying to define a user profile. The app is one of the first using TensorFlow for Android, which is a deep learning library optimised for the Qualcomm Snapdragon 835, provided by the contest's organisers. Therefore, the app development was not only a technological challenge but also a programming challenge, due to the limited documentation available online at the time, except a couple of tutorials.

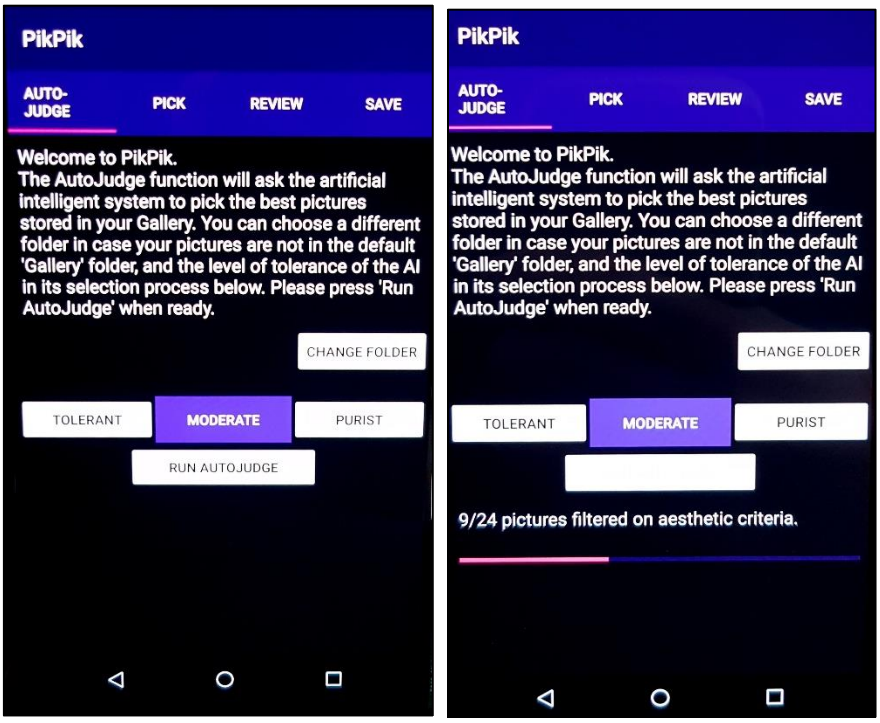


Figure 7.3: Screenshots of the initial page of PikPik. The user can select a folder to have its images filtered by the aesthetic classifier.

In comparison to the traditional task of binary aesthetic classification studied in the domain of computational aesthetics, the app mainly aims at having strong confidence when rejecting images instead of being trained to be both good at predicting aesthetically high and aesthetically low photographs. Therefore, the aesthetic classifier selected to be integrated into the app displays the highest precision on the validation set, instead of the highest accuracy. Precision is a ratio of true positives over all positives, while accuracy is a ratio of all trues over all items. It implies that the deep neural network is chosen for their high classification scores at detecting images with high aesthetics. All photographs selected by the classifier are automatically included in the final selection, without any possible alterations from the user. However, rejected photographs are available for review, in order to save potentially misclassified photographs. In this case, photographs are not only misclassified due to learning errors but also due to differences between the taught visual preferences generalised from the DPchallenge.com community and the subjective visual preferences of the user. While it may seem unfair to the user to not be able to review the photographs accepted by the classifier, it is the founding stone that allows the sorting process to be faster and hopefully, more efficient. Indeed, the app is designed to be reliable on a majority of photographs (in this case, the most aesthetically pleasing photographs) instead of being performant at classifying photographs of all aesthetic classes, so the user’s opinion may only be needed on the most ambiguous or poorly looking photographs. Depending on the parameters of the aesthetic classifier and also the average quality of the photographs shot by the user, it can reduce the number of images requiring the user’s opinions significantly.

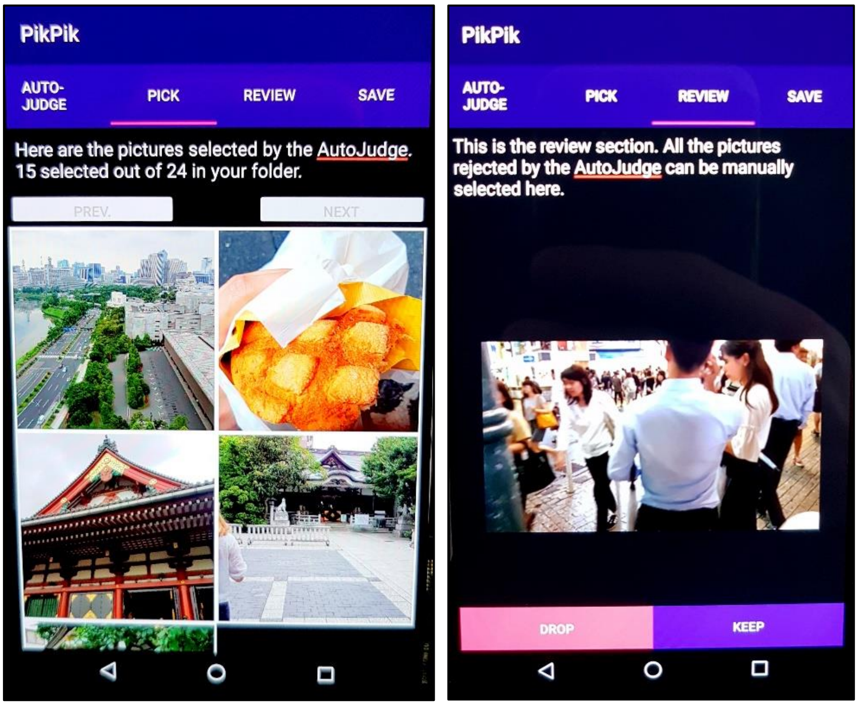


Figure 7.4: Screenshots of the page showing the images selected by the classifier (left) and the page allowing the user to confirm or disagree with the rejected photographs (right).

The platform on which the app is developed also presents a major limitation in terms of machine learning. Indeed, Tensorflow for Android requires deep neural networks to be trained and saved on a machine compatible with the traditional version of Tensorflow. The lack of neural network training possibilities on Tensorflow for Android implies that the software design needed some adaptability to the user's preferences. Consequently, the app is provided with 3 different classifiers that were trained on 3 subsets of the AVA datasets, which have higher or lower ratings representing the border with the two aesthetic categories. As a result, it makes the different classifiers more or less selective. In the review process, decisions from all 3 different classifiers are compared to the user's feedback, regardless of the classifier initially selected. The app can then suggest whether one of the classifiers would be more appropriate to the user's preferences for future use, without necessitating training a new deep neural network.

7.3. Discussion

The extracted features and the deep neural network developed in this thesis allow to efficiently classify photographs to state-of-the-art levels in the research domain of computational aesthetics, while also providing a tool for individual users. Some companies such as Xerox or regaind.io (bought by Apple in 2017) have been working on such implementations but only behind closed doors, implying an impossibility to compare academic results to products to be released shortly on the market (Marchesotti et al., 2011; Murray et al., 2012). One of the aims of human classification tasks and the PikPik app is to attempt linking results expressed in terms of correct classification rates to the potential satisfaction of a user on such automated photographs filtering tool. However, this task is difficult as the results presented in academic works are quantified over a population, while commercially-aimed products focus on a single user. Through the development of both projects and the feedback from participants, it appears that low-level visual features can only illustrate a person's preferences to some extent, due to a high variance in terms of aesthetic preferences across people. To conclude, the set of low-level visual features allows to train an ensemble of aesthetic classifiers which present itself as a robust tool to predict the preferences of the average individual without any prior learning of this specific individual.

Chapter 8

Discussion and Conclusion

8.1. From low-level visual features extraction to experimental tool

Relying on findings in neuroscience, psychology and psychophysics, the literature suggests that features in the early visual system may help to investigate human visual preferences due to diminished personal and cultural influences. On the other hand, computer algorithms to extract visual information from images were developed with a focus on a limited collection of photographs. Through the thesis, increments are made to ensure the performance of the aesthetic classifier on unseen datasets and unseen types of media. Because visual preferences are also displayed in creative products, datasets of paintings and logos were assembled for this thesis. While it was not possible to apply the proposed aesthetic classifier on such type of items due to the lack of reliable ground-truthing, particular attention was brought to the evolution in use of the visual features across time and creative persons. The aesthetic classifier has been implemented into a phone application to offer quick sorting of photographs based on aesthetics. Both the developed features and the aesthetic classifier can be considered as experimental tools for investigating visual preferences among individuals, whether they are creating or judging visual content, over different time spans.

8.2. Successes and shortcomings

The computational experiments exposed throughout this thesis have demonstrated either equal or better results than other approaches to aesthetic classification, mainly thanks to its inspiration from the human visual system and its preferences highlighted by the literature. While it is the most attractive aspect of the thesis from the perspective of the computational aesthetics and computer vision community, this achievement was only a by-product of the initial plan to improve the understanding of human visual preferences and create a link between neuroaesthetics and computational modelling (Lemarchand, 2017, 2018). The designed set of features and its coupling to a fully-connected deep neural network challenges existing works. Considering that other existing classifiers with state-of-the-art results rely on convolutional neural networks which take inspiration from the simple and complex cells in the human visual system, it can be suggested that brain-inspired solutions lead to particularly efficient systems (Fukushima, 1980; Goodfellow et al., 2017). The main difference between existing CNN-based classifiers and the solution introduced in this thesis is that CNNs figure out key patterns, without the patterns having to match directly with aesthetic pleasantness and possibly recognising visual features present in objects likely to appear in aesthetically pleasant images. The proposed system, however, learns from the pre-selected set of features represented as percentage distribution, certifying a controlled and efficient learning process. Due to the little amount of innovation on the deep learning techniques, the results can be fully attributed to the set of brain-inspired features. This approach not only succeeded to reach state-of-the-art results, but it confirms that computational aesthetics should reconnect with the initial ideology shared by neuroaesthetics.

Following the successful results on the classification of images issued from the AVA dataset, additional tests were arranged to evaluate the cross-dataset and cross-media capabilities (Murray et al., 2012). After training on the AVA dataset, cross-dataset tests were run on Datta et al. and the CUHK datasets, matching results from existing classifiers which were entirely tuned and trained on these datasets (Datta et al., 2006; Tang et al., 2013). Despite the average correct classification rate matching performances by existing systems, there are large differences between correct classification rates of aesthetically low and aesthetically high photographs. It indicates possible variations in terms of culture and expertise among the respective rating community of each dataset. Therefore, it would be useful for future datasets to control for the level of expertise and cultural background of the individuals rating the photographs. While performance on the dataset of videos by Tzelepis et al. showed to be satisfying but limited, it is essential to take into consideration the fact that the simple model used to adapt the proposed classifier to assess videos is potentially a source of performance loss (Tzelepis et al., 2016). A possible way to improve the results would be by using neural architectures optimised for time series data such as Long Short-Term Memory networks (LSTMs) (Hochreiter & Schmidhuber, 1997). This computational experiment also hints that the set of low-level visual features could be a valuable tool to detect aesthetic changes in videos, despite missing essential information about motion. Another advantage of having a pre-defined set of visual features is that it allows to run statistical analysis over the features present across the datasets and easily interpret the outcome.

Following this proof of robustness of the proposed system on images sourced from different provenances and types of media, the set of features was used to analyse potential aesthetic changes over the length of famous abstract artists. Several significant changes in preferred low-level visual features have been observed in popular abstract artists' careers such as an increase in cardinally-oriented lines or the colour blue. The goal of such experiment was not only to define whether artists used a great amount of visually preferred features as they gain experience but also to determine whether the proposed set of low-level visual features can link the knowledge of aesthetic judgement during contemplation and creation. Along the different data analysis, it is essential to note that other statistical tests may have been better suited than Welch's t-test as the features may be dependent on each other. Due to the study focussing on the distribution of different types of features, features such as colours must be implicitly linked to some extent.

From a computational aesthetics perspective, this new and promising brain-inspired approach confirms that human visual preferences can be learnt from photographs that have simply been rated by online users. Even though the tests introduced along the chapters still lack comparable works, it is hoped that cross-datasets and cross-media tests such as the test on videos by Tzelepis et al. (2016) become a new standardised test. As the field has seen correct classification rates on the AVA dataset slowly reaching a plateau, it can be expected that studies approaching computational aesthetics as a pure machine learning problem will disappear unless a larger dataset with a controlled rating process is published in the near future (Lo et al., 2012; Lu et al., 2014; Mavridaki & Mezaris, 2015; Wang et al., 2016). While a lack of new dataset may negatively affect the field, it is hoped that it could incite future works to focus on possible brain-inspired solutions to improve their systems' efficiency, as for example, in the recent work by Sun et al. (2018).

Based on the computational experiments in this thesis, as well as the existing literature underlined in Chapter 2 and 3, a model is proposed in Figure 8.1 to reflect on the outcome of this thesis. Also inspired by other known models, this simple model has for main objective to highlight how the demonstrated results take place in the multi-disciplinary field of study of human visual preferences. It takes inspiration from Redies et al. (2015) which represent aesthetic experience as a perceptual and cognitive phenomenon. Despite the fact that the set of low-level visual features introduced in this thesis attempt to ease the influence from semantic and contextual information on the final judgment of the model, including it into this model is a statement that the cognitive and emotional aspect of aesthetic appreciation should not be omitted. While the presented computational model performed well with perceptual information only, the model's performances are only valid in comparison to the average opinion of a group of people. As addressed by Reber (2012), cultural and personal experiences is one of the main sources of variation in aesthetic judgment across individuals. The model proposed by Che et al. (2018) also argues that perceptual preferences are much more similar across individuals and cultures. Therefore, the computational aesthetics models that have been criticised previously for not focusing on brain-inspired solutions and not contributing to a better understanding of human visual preferences, may appear to be a good solution to build systems prediction individuals' preferences instead of a group's preferences (Kao et al., 2016; Simond et al., 2015; Tian et al., 2015).

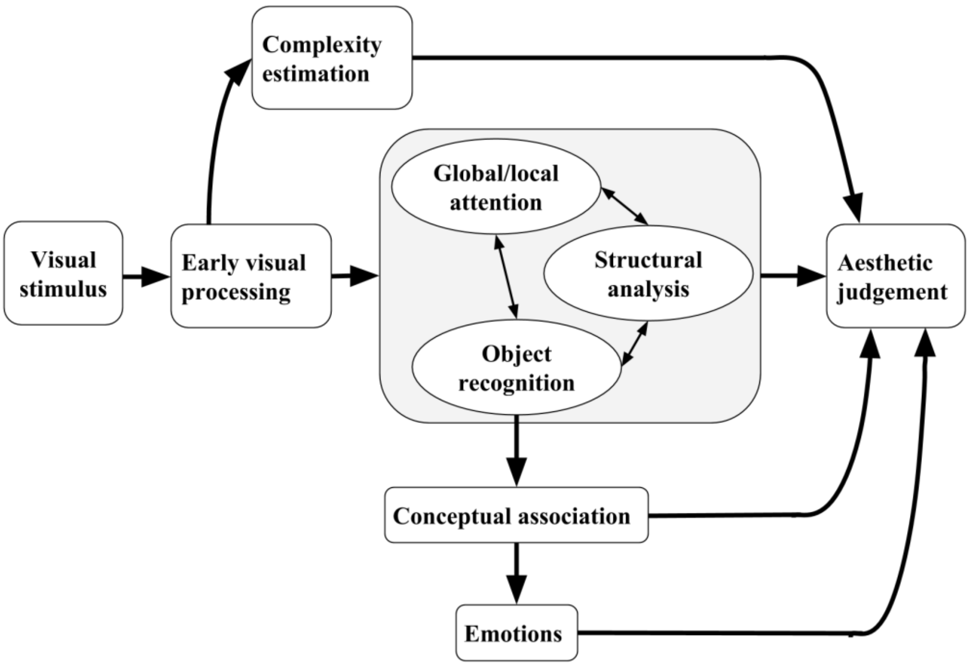


Figure 8.1: Model of aesthetic judgement.

The concept of visual complexity is included in the model despite its weak neuroscientific background, due to the multiple demonstrations of its relationship with aesthetics in many scenarios (Forsythe et al., 2011; Osborne & Farley, 1970; Reinecke et al., 2013; Romero et al., 2012). As suggested by Forsythe et al. (2011), complexity estimation may happen during early visual processing. Such measure happening early in visual processing, and therefore, in the aesthetic judgment process, could give expectations on how fast a visual stimulus should be processed, linking to the concept of processing fluency. This leads to hypothesise that the feeling of processing fluency may be triggered by the difference between complexity estimation and the actual difficulty to perceptually and cognitively process the visual stimulus.

To conclude, while the field of neuroaesthetics has been focusing on locating brain regions related to aesthetic appreciation, little work has been done from a bottom-up approach to study human visual preferences using the knowledge issued of the neuroscience of vision. The set of low-level features designed and tested through this thesis shows that computational aesthetics, as a field, would benefit from taking inspiration from the visual brain.

8.3. Thesis contributions

A new approach to computational aesthetics confirms a reliable link between a set of brain-inspired visual features and human visual preferences. The features can be coupled with various machine learning algorithms to investigate human visual preferences. It not only challenges the performance of state-of-the-art aesthetic classifiers, but it also offers fast visual information extraction and efficient encoding of human visual preferences into 114 features. The brain-inspired features demonstrate that the distribution of low-level visual information can provide sufficient hints of aesthetics value to learn human visual preferences, questioning the role of semantic information in aesthetic judgement. Aspects of this work were published in Lemarchand (2018).

The array of tests and analysis applied onto the proposed aesthetic classifier allows comparison between machine learning solutions and contributes towards a standard in the domain of computational aesthetics. Hereby, it addresses frequent and risky shortcuts in existing studies. The additional tests presented in the thesis improves the evaluation and comparison of the different photography datasets, as well as the aesthetic classifiers' performances on them (Datta et al., 2006; Murray et al., 2012). It allows to observe the aesthetic classifier's performance over different parts of the datasets, highlighting that a number of images seem virtually impossible to classify in a binary choice situation. A corresponding argument was published in Lemarchand (2018).

Extracting low-level visual features from images allows to train an aesthetic classifier with cross-media capabilities, with the classification of videos for example. The proposed aesthetic classifier shows satisfying performances when classifying videos, despite not reaching the classification levels demonstrated by Tzelepis et al. with a system designed and trained specifically on videos (Tzelepis et al., 2016). It demonstrates that an aesthetic classifier designed for static scenes provide a reasonable indicator of aesthetic quality in a dynamic setting. Such cross-media tests also provide qualitative feedback on the proposed aesthetic classifier's preferences by comparing aesthetic predictions with specific characteristics known to appear in selected films. This work was published as Lemarchand (2017).

An experimental framework to study basic visual preferences in creative processes in the visual domain on a long-term scale. The experiment consists of a statistical analysis on abstract and landscape paintings which attempts to build a bridge between aesthetic judgement processes modelled in the domain of computational aesthetics and creative artistic processes. With the increasing numbers of artificially intelligent agents imitating artistic processes such as AARON or the Painting Fool, investigations on low-level visual features provide additional information on cognitive processes to design future artistic AIs (H. Cohen, 2002; Colton, 2012). As most of the known artistic AIs in the computational creativity literature are product-focused, further understanding of the creative process regarding visual aesthetics would help to make AI-generated works of art with a focus on the modelling of the cognitive process.

The pilot experiment and prototype phone application provide quantitative and qualitative investigation tools to learn more about human visual preferences in individuals. Both accentuate how artificial the binary classification task attempted by the computational aesthetics community is. In fact, it helps to reflect on the ethical issues posed when learning visual preferences of individual, as it could be used as biometrics due to the subjectivity of preferences in any given person. Furthermore, the prototype phone application demonstrates that powerful neural network processing can be implemented locally on phone hardware, consequently addressing data leaks in client-server protocols and users' privacy concerns.

8.4. Future work

In the current state of this work and taking the most recent literature into consideration, three main promising opportunities are available to pursue the investigations:

- Use the proposed low-level visual features as a base and implement higher-level features, leading towards the improved aesthetic classification of images containing blur and depth of field or dynamic visual information such as videos.
- Develop a large-scale dataset with a controlled cultural background to confirm whether the findings of this thesis are specific to the western world. Finalising and releasing the photograph aesthetic filter application, for example on the Android platform, could provide valuable feedback from various cultures.
- Develop a creative artificial intelligent agent to generate visual art using extensively the existing literature of studies on human visual preferences and aesthetics. This could offer an alternative to traditional rule-based art producing systems or deep learning solutions training on large datasets of paintings.

Bibliography

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., & Zheng, X. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *ArXiv Preprint*, 1603.04467.

Abu-El-Haija, S., Kothari, N., Lee, J., Natsev, P., Toderici, G., Varadarajan, B., & Vijayanarasimhan, S. (2016). YouTube-8M: A Large-Scale Video Classification Benchmark. *ArXiv Preprint*, 1609.08675.

Alpaydin, E. (2009). *Introduction to machine learning*. MIT Press.

Antes, J. R., & Kristjanson, A. F. (1991). Discriminating Artists from Nonartists by Their Eye-Fixation Patterns. *Perceptual and Motor Skills*, *73*(3), 893–894. https://doi.org/10.2466/pms.1991.73.3.893

Aragón, J. L., Naumis, G. G., Bai, M., Torres, M., & Maini, P. K. (2008). Turbulent Luminance in Impassioned van Gogh Paintings. *Journal of Mathematical Imaging and Vision*, *30*, 275–283.

Aristotle. (1996). *Poetics (Translated by Malcolm Heath. Originally published 335BC)*. Penguin Classics.

Arora, R. S., & Elgammal, A. (2012). Towards Automated Classification of Fine-art

Painting Style: a Comparative Study. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR 2012)* (pp. 3541–3544). IEEE.

Ball, V. (1965). The Aesthetics of Color: A Review of Fifty Years of Experimentation. *The Journal of Aesthetics and Art Criticism*, *23*(4), 441–452.

Bar, M., & Neta, M. (2007). Visual elements of subjective preference modulate amygdala activation. *Neuropsychologia*, *45*(10), 2191–2200.

Barnett, J. R., Miller, S., & Pearce, E. (2006). Colour and art: A brief history of pigments. *Optics & Laser Technology*, *38*(4), 445–453.

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded Up Robust Features. In *Proceedings of the European Conference on Computer Vision (ECCV 2006)* (pp. 404–417). Springer, Berlin, Heidelberg.

Beardsley, M. C. (1975). *Aesthetics from classical Greece to the present : a short history* (38th ed.). University of Alabama Press.

Bellman, R. (1957). *Dynamic Programming*. Princeton University Press.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics*, *8*(5), 279–286.

Berlyne, D. E. (1971). *Aesthetics and Psychobiology*. New York: Appleton-Century-Crofts.

Bertamini, M., Palumbo, L., Gheorghes, T. N., & Galatsidas, M. (2016). Do observers like curvature or do they dislike angularity? *British Journal of Psychology*, *107*(1), 154–178.

Birkhoff, G. D. (1933). *Aesthetic measure*. Humphrey Milford, Oxford Univ. Press.

Blasdel, G. G. (1992). Orientation selectivity, preference, and continuity in monkey striate cortex. *Journal of Neuroscience*, *12*(8), 3139–3161.

Bona, S., Herbert, A., Toneatto, C., Silvanto, J., & Cattaneo, Z. (2014). The causal role of the lateral occipital complex in visual mirror symmetry detection and grouping: An fMRI-guided TMS study. *Cortex*, *51*, 46–55.

Boser, B., Guyon, I., & Vapnik, V. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth Annual Workshop on Computational Learning*

Theory (COLT 1992) (pp. 144–152). ACM Press.

Brieber, D., Nadal, M., & Leder, H. (2015). In the white cube: Museum context enhances the valuation and memory of art. *Acta Psychologica*, *154*, 36–42.

Brouwer, G. J., & Heeger, D. J. (2009). Decoding and Reconstructing Color from Responses in Human Visual Cortex. *Journal of Neuroscience*, *29*(44), 13992–14003.

Brownlee, J. (2014). What's The Difference Between A Logo And A Symbol? Retrieved August 23, 2017, from https://www.fastcodesign.com/3034007/whats-the-difference-between-a-logo-and-a-symbol

Camgö, N. N., Yener, C., & Gü Venç, D. (2002). Effects of Hue, Saturation, and Brightness on Preference. *Color Research & Application*, *27*(3), 199–207.

Casalderrey, F. M. (2017). *Playing with the Senses: Art through mathematical eyes*. National Geographic.

Cattaneo, Z., Lega, C., Ferrari, C., Vecchi, T., Cela-Conde, C. J., Silvanta, J., & Nadal, M. (2015). The role of the lateral occipital cortex in aesthetic appreciation of representational and abstract paintings: A TMS study. *Brain and Cognition*, *95*, 44–53.

Cela-Conde, C. J., Agnati, L., Huston, J. P., Mora, F., & Nadal, M. (2011). The neural foundations of aesthetic appreciation. *Progress in Neurobiology*, *94*(1), 39–48.

Chapman, B., & Bonhoeffer, T. (1998). Overrepresentation of horizontal and vertical orientation preferences in developing ferret area 17. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(5), 2609–2614.

Chatterjee, A. (2011). Neuroaesthetics: A Coming of Age Story. *Journal of Cognitive Neuroscience*, *23*(1), 53–62.

Chatterjee, A., & Vartanian, O. (2014). Neuroaesthetics. *Trends in Cognitive Sciences*, *18*(7), 370–375.

Che, J., Sun, X., Gallardo, V., & Nadal, M. (2018). Cross-cultural empirical aesthetics. In *Progress in Brain Research* (pp. 77–103). Elsevier.

Chipman, S. F. (1977). Complexity and structure in visual patterns. *Journal of*

Experimental Psychology: General, *106*(3), 269–301.

Clark, C. (2014). The Colors Of Motion. Retrieved August 30, 2017, from http://thecolorsofmotion.com/films

Cohen, H. (2002). A Self-Defining Game for One Player: On the Nature of Creativity and the Possibility of Creative Computer Programs. *Leonardo*, *35*(1), 59–64.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Collobert, R., & Bengio, S. (2004). Links between perceptrons, MLPs and SVMs. In *Proceedings of the 21st International Conference on Machine Learning (ICML 2004)* (p. 23). New York, New York, USA: ACM Press.

Colton, S. (2012). The Painting Fool: Stories from Building an Automated Painter. In *Computers and Creativity* (pp. 3–38). Berlin, Heidelberg: Springer Berlin Heidelberg.

Conway, B. R., Chatterjee, S., Field, G. D., Horwitz, Gregory D. Johnson, Elizabeth N. Koida, K., & Mancuso, K. (2010). Advances in color science: from retina to behavior. *Journal of Neuroscience, 30*(45), 14955–14963.

Cooper, J. M., & Silvia, P. J. (2009). Opposing Art: Rejection as an Action Tendency of Hostile Aesthetic Emotions. *Empirical Studies of the Arts*, *27*(1), 109–126.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297.

Dalal, N., & Triggs, B. (2005). Histogram of Oriented Gradients for Human Detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)* (Vol. 1, pp. 886–893). IEEE.

Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2006). Studying aesthetics in photographic images using a computational approach. In *Proceedings of the European Conference on Computer Vision (ECCV 2006)* (pp. 288–301). Springer, Berlin, Heidelberg.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009)* (pp. 248–255). IEEE.

Donderi, D. C. (2006). An Information Theory Analysis of Visual Complexity and Dissimilarity. *Perception*, *35*(6), 823–835.

Dow, B. M. (2002). Orientation and Color Columns in Monkey Visual Cortex. *Cerebral Cortex*, *12*(10), 1005–1015.

Ehrenfels, C. Von. (1890). Über Gestaltqualitäten. *Vierteljahrsschrift Für Wissenschaftliche Philosophie*, *14*(3), 249–292.

Fairhall, S. L., & Ishai, A. (2008). Neural correlates of object indeterminacy in art compositions. *Consciousness and Cognition*, *17*(3), 923–932.

Fechner, G. (1860). *Elemente der Psychophysik*. Leipzig, Germany: Breitkopf & Härtel.

Fechner, G. (1876). *Vorschule der Aesthetik*. Leipzig, Germany: Breitkopf & Härtel.

Ferrari, C., Lega, C., Tamietto, M., Nadal, M., & Cattaneo, Z. (2015). I find you more attractive… after (prefrontal cortex) stimulation. *Neuropsychologia*, *72*, 87–93.

Forsythe, A., Mulhern, G., & Sawey, M. (2008). Confounds in pictorial sets: The role of complexity and familiarity in basic-level picture processing. *Behavior Research Methods*, *40*(1), 116–129.

Forsythe, A., Nadal, M., Sheehy, N., Cela-Conde, C., & Sawey, M. (2011). Predicting beauty: Fractal dimension and visual complexity in art. *British Journal of Psychology*, *102*(1), 49–70.

Forsythe, A., Williams, T., & Reilly, R. G. (2017). What paint can tell us: A fractal analysis of neurological changes in seven artists. *Neuropsychology*, *31*(1), 1–10.

Franklin, A., Bevis, L., Ling, Y., & Hurlbert, A. (2010). Biological components of colour preference in infancy. *Developmental Science*, *13*(2), 346–354.

Fukushima, K. (1980). Neocognitron. *Biological Cybernetics*, *36*(4), 93–202.

Gartus, A., & Leder, H. (2017). Predicting perceived visual complexity of abstract patterns using computational measures: The influence of mirror symmetry on complexity perception. *PLoS ONE*, *12*(11), e0185276.

Gerger, G., Leder, H., & Kremer, A. (2014). Context effects on emotional and aesthetic evaluations of artworks and IAPS pictures. *Acta Psychologica*, *151*, 174–183.

Gibson, I. (1997). *The shameful life of Salvador Dalí*. Faber and Faber.

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*(7), 926–932.

Goodfellow, I., Bengio, Y., & Courville, A. (2017). *Deep learning*. MIT Press.

Graf, L. K. M., & Landwehr, J. R. (2015). A Dual-Process Perspective on Fluency-Based Aesthetics. *Personality and Social Psychology Review*, *19*(4), 395–410.

Hadjikhani, N., Liu, A. K., Dale, A. M., Cavanagh, P., & Tootell, R. B. H. (1998). Retinotopy and color sensitivity in human visual cortical area V8. *Nature Neuroscience*, *1*(3), 235–241.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter*, *11*(1), 10–18.

Harper, S., Michailidou, E., & Stevens, R. (2009). Toward a definition of visual complexity as an implicit measure of cognitive load. *ACM Transactions on Applied Perception (TAP)*, *6*(2), 10.

Heywood, C. A., Gadotti, A., & Cowey, A. (1992). Cortical area V4 and its role in the perception of color. *Journal of Neuroscience*, *12*(10), 4056–4065.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780.

Höfel, L., & Jacobsen, T. (2007). Electrophysiological indices of processing aesthetics: Spontaneous or intentional processes? *International Journal of Psychophysiology*, *65*(1), 20–31.

Huang, Y., Xue, X., Spelke, E., Huang, L., Zheng, W., & Peng, K. (2018). The aesthetic preference for symmetry dissociates from early-emerging attention to symmetry. *Scientific Reports*, *8*(1), 6263.

Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, *14*(1), 55–63.

Huhmann, B. A. (2003). Visual complexity in banner ads. *Visual Communication*

Quarterly, *10*(3), 10–17.

ING, Microsoft, & TU Delft. (2016). The Next Rembrandt. Retrieved January 30, 2017, from https://www.nextrembrandt.com/

Ishizu, T., & Zeki, S. (2011). Toward A Brain-Based Theory of Beauty. *PLoS ONE*, *6*(7), e21852.

Ishizu, T., & Zeki, S. (2013). The brain's specialized systems for aesthetic and perceptual judgment. *European Journal of Neuroscience*, *37*(9), 1413–1420.

Joblove, G. H., & Greenberg, D. (1978). Color spaces for computer graphics. In *Proceedings of the ACM SIGGRAPH Computer Graphics (SIGGRAPH 1978)* (Vol. 12, pp. 20–25). ACM.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of Neural Science* (Vol. 4). New York: McGraw-hill.

Kant, I. (1999). *Critique of pure reason (Translated and edited by Paul Guyer and Allen W. Wood. Originally published 1781)*. Cambridge University Press.

Kao, Y., He, R., & Huang, K. (2016). Visual Aesthetic Quality Assessment with Multi-task Deep Learning. *ArXiv Preprint*, 1604.04970.

Kao, Y., He, R., & Huang, K. (2017). Deep Aesthetic Quality Assessment With Semantic Information. *IEEE Transactions on Image Processing*, *26*(3), 1482–1495.

Kawabata, H., & Zeki, S. (2004). Neural Correlates of Beauty. *Journal of Neurophysiology*, *91*(4), 1699–1705. https://doi.org/10.1152/jn.00696.2003

Koffka, K. (1922). Perception: an introduction to the Gestalt-Theorie. *Psychological Bulletin*, *19*(10), 531–585.

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 1995)* (pp. 1137-1145).

Koide, N., Kubo, T., Nishida, S., Shibata, T., & Ikeda, K. (2015). Art Expertise Reduces Influence of Visual Salience on Fixation in Viewing Abstract-Paintings. *PLoS ONE*, *10*(2), e0117696.

Kong, S., Shen, X., Lin, Z., Mech, R., & Fowlkes, C. (2016). Photo Aesthetics Ranking

Network with Attributes and Content Adaptation. In *Proceedings of the European Conference on Computer Vision (ECCV 2016)* (pp. 662–679). Springer International Publishing.

Labrecque, L. I., & Milne, G. R. (2012). Exciting red and competent blue: the importance of color in marketing. *Journal of the Academy of Marketing Science*, *40*(5), 711–727.

Lanska, M., Olds, J. M., & Westerman, D. L. (2014). Fluency effects in recognition memory: Are perceptual fluency and conceptual fluency interchangeable? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 1–11.

Latto, R., & Russell-Duff, K. (2002). An Oblique Effect in the Selection of Line Orientation by Twentieth Century Painters. *Empirical Studies of the Arts*, *20*(1), 49–60.

Lawrence, S., Burns, I., Back, A., Tsoi, A. C., & Giles, C. L. (2012). Neural network classification and prior class probabilities. In *Neural networks: Tricks of the trade* (pp. 295–309). Springer, Berlin, Heidelberg.

LeCun, Y., Cortes, C., & Burges, C. J. C. (1998). The MNIST database of handwritten digits. Retrieved from http://yann.lecun.com/exdb/mnist/

LeCun, Y., Haffner, P., Bottou, L., & Bengio, Y. (1999). Object Recognition with Gradient-Based Learning. In *Shape, contour and grouping in computer vision* (pp. 319–345). Springer, Berlin, Heidelberg.

LeCun, Y., Jackel, L. D., Bottou, L., Cortes, C., Denker, J. S., Drucker, H., Guyon, I., Muller, U. A., Sackinger, E., Simard, P., & Vapnik, V. (1995). Learning algorithms for classification: A comparison on handwritten digit recognition. In *Neural networks: the statistical mechanics perspective* (pp. 261–276).

Lee, H.-J., Hong, K.-S., Kang, H., & Lee, S. (2017). Photo Aesthetics Analysis via DCNN Feature Encoding. *IEEE Transactions on Multimedia*, *19*(8), 1921–1932.

Lemarchand, F. (2017). From Computational Aesthetic Prediction for Images to Films and Online Videos. *AVANT*, *8*, 69–78.

Lemarchand, F. (2018). Fundamental visual features for aesthetic classification of photographs across datasets. *Pattern Recognition Letters*, *112, 9-17*.

Lo, K.-Y., Liu, K.-H., & Chen, C.-S. (2012). Assessment of Photo Aesthetics With Efficiency. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR 2012)* (pp. 2186–2189). IEEE.

Locher, P., Overbeeke, K., & Stappers, P. J. (2005). Spatial Balance of Color Triads in the Abstract Art of Piet Mondrian. *Perception*, *34*(2), 169–189.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV 1999)* (pp. 1150–1157). IEEE.

Lu, X., Lin, Z., Jin, H., Yang, J., & Wang, J. Z. (2014). RAPID: Rating Pictorial Aesthetics using Deep Learning. In *Proceedings of the 22nd ACM International Conference on Multimedia (MM 2014)* (pp. 457–466). ACM.

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review of Public Health*, *23*(1), 151–169.

Ma, Q., Hu, L., & Wang, X. (2015). Emotion and novelty processing in an implicit aesthetic experience of architectures. *NeuroReport*, *26*(5), 279–284.

Machado, P., Romero, J., Nadal, M., Santos, A., Correia, J., & Carballal, A. (2015). Computerized measures of visual complexity. *Acta Psychologica*, *160*, 43–57.

Mai, L., Jin, H., & Liu, F. (2016). Composition-Preserving Deep Photo Aesthetics Assessment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)* (pp. 497–506). IEEE.

Makin, A. D. J., Wilton, M. M., Pecchinenda, A., & Bertamini, M. (2012). Symmetry perception and affective responses: A combined EEG/EMG study. *Neuropsychologia*, *50*(14), 3250–3261.

Marchesotti, L., Perronnin, F., Larlus, D., & Csurka, G. (2011). Assessing the aesthetic quality of photographs using generic image descriptors. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV 2011)* (pp. 1784–1791). IEEE.

Marin, M. M. (2015). Crossing boundaries: toward a general model of neuroaesthetics. *Frontiers in Human Neuroscience*, *9*, 443.

Mavridaki, E., & Mezaris, V. (2015). A comprehensive aesthetic quality assessment method for natural images using basic rules of photography. In *Proceedings of the IEEE International Conference on Image Processing (ICIP 2015)* (pp. 887–891). IEEE.

McManus, I. C., Jones, A. L., & Cottrell, J. (1982). The aesthetics of colour. *Perception*, *10*(6), 651–666.

Miller, B. L., Cummings, J., Mishkin, F., Boone, K., Prince, F., Ponton, M., & Cotman, C. (1998). Emergence of artistic talent in frontotemporal dementia. *Neurology*, *51*(4), 978–982.

Miller, B. L., & Hou, C. E. (2004). Portraits of Artists. *Archives of Neurology*, *61*(6), 842.

Miller, Z. A., & Miller, B. L. (2012). A cognitive and behavioral neurological approach to aesthetics. In *Aesthetic science: Connecting mind, brain, and experience* (pp. 357–374). Oxford University Press.

Mollick, E., & Nanda, R. (2015). Wisdom or Madness? Comparing Crowds with Expert Evaluation in Funding the Arts. *Management Science*, *62*(6), 1533–1553.

Moore, G. E. (1965). Cramming More Components onto Integrated Circuits. *Electronics*, *38*(8), 114–117.

Motter, B. C. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *Journal of Neuroscience*, *14*(4), 2178–2189.

Munar, E., Gomez-Puerto, G., Call, J., & Nadal, M. (2015). Common visual preference for curved contours in humans and great apes. *PLoS ONE*, *10*(11), e0141106.

Munar, E., Nadal, M., Castellanos, N. P., Flexas, A., Maestú, F., Mirasso, C., & Cela-Conde, C. J. (2012). Aesthetic appreciation: event-related field and time-frequency analyses. *Frontiers in Human Neuroscience*, *5*, 185.

Murray, N., Marchesotti, L., & Perronnin, F. (2012). AVA: A large-scale database for aesthetic visual analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2012)* (pp. 2408–2415). IEEE.

Muth, C., & Carbon, C.-C. (2013). The Aesthetic Aha: On the pleasure of having insights into Gestalt. *Acta Psychologica*, *144*(1), 25–30.

Muth, C., Raab, M. H., & Carbon, C.-C. (2016). Semantic Stability is More Pleasurable in Unstable Episodic Contexts. On the Relevance of Perceptual Challenge in Art Appreciation. *Frontiers in Human Neuroscience*, *10*, 43.

Nadal, M., Munar, E., Capó, M. À., Rosselló, J., & Cela-Conde, C. J. (2008). Towards a framework for the study of the neural correlates of aesthetic preference. *Spatial Vision*, *21*(3), 379–396.

Nadal, M., Munar, E., Marty, G., & Cela-Conde, C. J. (2010). Visual Complexity and Beauty Appreciation: Explaining the Divergence of Results. *Empirical Studies of the Arts*, *28*(2), 173–191.

Nascimento, S. M. C., Linhares, J. M. M., Montagner, C., João, C. A. R., Amano, K., Alfaro, C., & Bailão, A. (2017). The colors of paintings and viewers' preferences. *Vision Research*, *130*, 76–84.

Neistat, C. (2016). Casey Neistat's YouTube video channel. Retrieved from www.youtube.com/user/caseyneistat/videos

Niu, Y., & Liu, F. (2012). What Makes a Professional Video? A Computational Aesthetics Approach. *IEEE Transactions on Circuits and Systems for Video Technology*, *22*(7), 1037–1049.

Osborne, J. W., & Farley, F. H. (1970). The relationship between aesthetic preference and visual complexity in absract art. *Psychonomic Science*, *19*(2), 69–70.

Osowski, S., Siwek, K., & Markiewicz, T. (2004). MLP and SVM Networks - a Comparative Study. In *Proceedings of the 6th Nordic Signal Processing Symposium (NORSIG 2004)* (pp. 37–40).

Ou, L.-C., Luo, M. R., Woodcock, A., & Wright, A. (2004a). A study of colour emotion and colour preference. Part I: Colour emotions for single colours. *Color Research & Application*, *29*(3), 232–240.

Ou, L.-C., Luo, M. R., Woodcock, A., & Wright, A. (2004b). A study of colour emotion and colour preference. Part II: Colour emotions for two-colour combinations. *Color Research & Application*, *29*(4), 292–298.

Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual Aesthetics and Human Preference. *Annual Review of Psychology*, *64*(1), 77–107.

Palumbo, L., Ogden, R., Makin, A. D. J., & Bertamini, M. (2014). Examining visual complexity and its influence on perceived duration. *Journal of Vision*, *14*(14), 1–18.

Park, C. W., Eisingerich, A. B., Pol, G., & Park, J. W. (2013). The role of brand logos in firm performance. *Journal of Business Research*, *66*(2), 180–187.

Pecchinenda, A., Bertamini, M., Makin, A. D. J., & Ruta, N. (2014). The pleasantness of visual symmetry: Always, never or sometimes. *PLoS ONE*, *9*(3), e92685.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pieters, R., Wedel, M., & Batra, R. (2010). The Stopping Power of Advertising: Measures and Effects of Visual Complexity. *Journal of Marketing*, *74*(5), 48–60.

Plato. (2003). *The Republic (Translated by H.D.P. Lee. Originally published 380BC)*. Penguin Classics.

Ramachandran, V. S., & Hirstein, W. (1999). The science of art: a neurological theory of aesthetic experience. *Journal of Consciousness Studies*, *6*(6–7), 15–51.

Reber, R. (2012). Processing fluency, aesthetic pleasure, and culturally shared taste. In *Aesthetic science: Connecting mind, brain, and experience* (pp. 223–249). Oxford University Press.

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience? *Personality and Social Psychology Review*, *8*(4), 364–382.

Redies, C. (2015). Combining universal beauty and cultural context in a unifying model of visual aesthetic experience. *Frontiers in Human Neuroscience*, *9*, 218.

Reinecke, K., Yeh, T., Miratrix, L., Mardiko, R., Zhao, Y., Liu, J., & Gajos, K. Z. (2013). Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2013)* (pp. 2049–2058). ACM Press.

Rhodes, M. (1961). An Analysis of Creativity. *The Phi Delta Kappan*, *42*(7), 305–310.

Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., Lu, H., & Vanduffel, W. (2012). Toward a Unified Theory of Visual Area V4. *Neuron*, *74*(1), 12–29.

Romero, J., Machado, P., Carballal, A., & Santos, A. (2012). Using complexity estimates in aesthetic image classification. *Journal of Mathematics and the Arts*, *6*(2–3), 125–136.

Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV 2011)* (pp. 2564–2571). IEEE.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning Internal Representations by Error Propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 1, pp. 318–362). Cambridge, MA: Bradford Books.

Rump, E. E. (1968). Is there a general factor of preference for complexity? *Perception & Psychophysics*, *3*(5), 346–348.

Rupp, K., Horowitz, M., Labonte, F., Shacham, O., Olukotun, K., Hammond, L., & Batten, C. (2018). 42 Years of Microprocessor Trend Data. Retrieved April 6, 2018, from https://www.karlrupp.net/2018/02/42-years-of-microprocessor-trend-data/

Saleh, B., Ravneet, K. A., Arora, S., & Elgammal, A. (2016). Toward Automated Discovery of Artistic Influence. *Multimedia Tools and Applications*, *75*(7), 3565–3591.

Sammartino, J., & Palmer, S. E. (2012). Aesthetic issues in spatial composition: Representational fit and the role of semantic context. *Perception*, *41*(12), 1434–1457.

Sasaki, Y., Vanduffel, W., Knutsen, T., Tyler, C., & Tootell, R. (2005). Symmetry activates extrastriate visual cortex in human and nonhuman primates. *Proceedings of the National Academy of Sciences*, *102*(8), 3159–3163.

Schechter, A. H. (2010). Measuring the Value of Corporate and Brand Logos. *Design Management Journal (Former Series)*, *4*(1), 33–39.

Schmidhuber, J. (2015). Deep Learning in Neural Networks: An Overview. *Neural Networks, 61*, 85–117.

Shimamura, A. P., & Palmer, S. E. (2012). *Aesthetic science : connecting minds, brains, and experience*. Oxford University Press.

Shulman, J. (2017). Photographs of Films. Retrieved August 30, 2017, from http://www.jasonshulmanstudio.com/photographs-of-films/

Silvia, P. J. (2006). Artistic Training and Interest in Visual Art: Applying the Appraisal Model of Aesthetic Emotions. *Empirical Studies of the Arts, 24*(2), 139–161.

Silvia, P. J. (2012). Human emotions and aesthetic experience: An overview of empirical aesthetics. In *Aesthetic science: Connecting mind, brain, and experience* (pp. 250–275). Oxford University Press.

Silvia, P. J., & Barona, C. M. (2009). Do People Prefer Curved Objects? Angularity, Expertise, and Aesthetic Preference. *Empirical Studies of the Arts*, *27*(1), 25–42.

Silvia, P. J., & Brown, E. M. (2007). Anger, disgust, and the negative aesthetic emotions: Expanding an appraisal model of aesthetic experience. *Psychology of Aesthetics, Creativity, and the Arts*, *1*(2), 100–106.

Sima, C., & Dougherty, E. R. (2008). The peaking phenomenon in the presence of feature-selection. *Pattern Recognition Letters*, *29*(11), 1667–1674.

Simond, F., Arvanitopoulos, N., & Susstrunk, S. (2015). Image aesthetics depends on context. In *Proceedings of the IEEE International Conference on Image Processing (ICIP 2015)* (pp. 3788–3792). IEEE.

Sincich, L. C., & Horton, J. C. (2005). The Circuitry of V1 and V2: Integration of Color, Form, and Motion. *Annual Review of Neuroscience*, *28*(1), 303–326.

Solomon, S. G., & Lennie, P. (2007). The machinery of colour vision. *Nature Reviews Neuroscience*, *8*(4), 276–286.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research, 15*(1), 1929–1958.

Stam, R. (2017). *Film Theory: An Introduction*. John Wiley & Sons, Incorporated.

Stirling, G. (1993). The Impact of Surgery on Fine Art. *Perspectives in Biology and Medicine*, *37*(1), 67–73.

Stork, D. G. (2009). Computer Vision and Computer Graphics Analysis of Paintings and Drawings: An Introduction to the Literature. In *Proceedings of the International Conference on Computer Analysis of Images and Patterns (CAIP 2009)* (pp. 9–24). Springer Berlin Heidelberg.

Sun, L., Yamasaki, T., & Aizawa, K. (2018). Photo aesthetic quality estimation using visual complexity features. *Multimedia Tools and Applications, 77*(5), 5189–5213.

Sussman, A. (2007). Mental Illness and Creativity: A Neurological View of the "Tortured Artist." *Stanford Journal of Neuroscience*, *1*(1), 21–24.

Suykens, J. A. K. (2001). Nonlinear modelling and support vector machines. In *Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference (IMTC 2001)* (Vol. 1, pp. 287–294). IEEE.

Tan, Y., Tang, P., Zhou, Y., Luo, W., Kang, Y., & Li, G. (2017). Photograph aesthetical evaluation and classification with deep convolutional neural networks. *Neurocomputing*, *228*, 165–175.

Tang, X., Luo, W., & Wang, X. (2013). Content-Based Photo Quality Assessment. *IEEE Transactions on Multimedia*, *15*(8), 1930–1943.

Taylor, R. P., Spehar, B., Clifford, C. W. G., Hagerhall, C. M., Purcell, T., & Martin, T. P. (2005). Perceptual and Physiological Responses to the Visual Complexity of Fractal Patterns. *Nonlinear Dynamics, Psychology, and Life Sciences*, *9*(1), 89–114.

Thakral, P. P., Moo, L. R., & Slotnick, S. D. (2012). A neural mechanism for aesthetic experience. *NeuroReport*, *23*(5), 310–313.

Tian, X., Dong, Z., Yang, K., & Mei, T. (2015). Query-Dependent Aesthetic Model With Deep Learning for Photo Quality Assessment. *IEEE Transactions on Multimedia*, *17*(11), 2035–2048.

Trunk, G. V. (1979). A Problem of Dimensionality: A Simple Example. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 1*(3), 306–307.

Tuch, A. N., Presslaber, E. E., Stöcklin, M., Opwis, K., & Bargas-Avila, J. A. (2012).

The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments. *International Journal of Human-Computer Studies*, *70*(11), 794–811.

Turcot, P., & Lowe, D. G. (2009). Better matching with fewer features: The selection of useful features in large database recognition problems. In *Proceedings of the 12th IEEE International Conference on Computer Vision Workshops (ICCV 2009 Workshops)* (pp. 2109–2116). IEEE.

Tzelepis, C., Mavridaki, E., Mezaris, V., & Patras, I. (2016). Video aesthetic quality assessment using kernel Support Vector Machine with isotropic Gaussian sample uncertainty (KSVM-IGSU). In *Proceedings of the IEEE International Conference on Image Processing (ICIP 2016)*. IEEE.

van der Zwan, R., Leo, E., Joung, W., Latimer, C., & Wenderoth, P. (1998). Evidence that both area V1 and extrastriate visual cortex contribute to symmetry perception. *Current Biology*, *8*(15), 889–892.

van Grinsven, B., & Das, E. (2016). Logo design in marketing communications: Brand logo complexity moderates exposure effects on brand recognition and brand attitude. *Journal of Marketing Communications*, *22*(3), 256–270.

Vartanian, O., & Goel, V. (2004). Neuroanatomical correlates of aesthetic preference for paintings. *Neuroreport*, *15*(5), 893–897.

Vartanian, O., Navarrete, G., Chatterjee, A., Fich, L. B., Leder, H., Modroño, C., Nadal, M., Rostrup, N., & Skov, M. (2013). Impact of contour on aesthetic judgments and approach-avoidance decisions in architecture. *Proceedings of the National Academy of Sciences*, *110*(Supplement 2), 10446–10453.

Vogt, S., & Magnussen, S. (2007). Expertise in Pictorial Perception: Eye-Movement Patterns and Visual Memory in Artists and Laymen. *Perception*, *36*(1), 91–100.

Wallace, G. K. (1991). The JPEG still picture compression standard. *Communications of the ACM*, *34*(4), 30–44.

Wang, Z., Chang, S., Dolcos, F., Beck, D., Liu, D., & Huang, T. (2016). Brain-Inspired Deep Networks for Image Aesthetics Assessment. *ArXiv Preprint*, 1601.04155.

WassilyKandinsky.net. (2018). Retrieved from wassilykandinsky.net

Weichselbaum, H., Leder, H., & Ansorge, U. (2018). Implicit and Explicit Evaluation of Visual Symmetry as a Function of Art Expertise. *I-Perception*, *9*(2), 204166951876146.

Welch, B. L. (1947). The generalization of Student's problem when several different population variances are involved. *Biometrika*, *34*(1/2), 28–35.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. II. *Psychologische Forschung*, *4*(1), 301–350.

WikiArt. (2018). Retrieved from www.wikiart.org

Yang, C.-Y., Yeh, H.-H., & Chen, C.-S. (2011). Video aesthetic quality assessment by combining semantically independent and dependent features. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011)* (pp. 1165–1168).

Yanulevskaya, V., Uijlings, J., Bruni, E., Sartori, A., Zamboni, E., Bacci, F., Melcher, D., & Sebe, N. (2012). In the eye of the beholder. In *Proceedings of the 20th ACM International Conference on Multimedia (MM 2012)* (p. 349). ACM Press.

Zeki, S. (1999). *Inner Vision: An Exploration of Art and the Brain*. Oxford University Press.

Zeki, S. (2002). Neural concept formation & art : Dante, Michelangelo, Wagner. *Journal of Consciousness Studies*, *9*(3), 53–79.

Zeki, S. (2015). A massively asynchronous, parallel brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1668), 20140174.

Zeki, S., & Ishizu, T. (2013). The "Visual Shock" of Francis Bacon: an essay in neuroesthetics. *Frontiers in Human Neuroscience*, *7*, 850.

Zeki, S., & Marini, L. (1998). Three cortical stages of colour processing in the human brain. *Brain : A Journal of Neurology*, *121*(9), 1669–1685.

Zeki, S., Romaya, J. P., Benincasa, D. M. T., & Atiyah, M. F. (2014). The experience of mathematical beauty and its neural correlates. *Frontiers in Human Neuroscience*, *8*, 68. https://doi.org/10.3389/fnhum.2014.00068