Affective Analysis of Abstract Paintings Using Statistical Analysis

and Art Theory

Andreza Sartori

Advisor: Prof. Nicu Sebe

Doctoral School in Information and Communication Technology

Department of Information Engineering and Computer Science (DISI)

University of Trento

April 2015

The emotion of beauty is always obscured by the appearance of the object. Therefore, the object must be eliminated from the picture.

Piet Mondrian

Acknowledgements

Abstract

This research thesis aims to provide a novel approach to Emotion Recognition of Images: based on empirical studies, we employ the state-of-the-art computer vision techniques in order to understand what makes an abstract artwork emotional. We identify and quantify the emotional regions of abstract paintings. We also investigate the contributions of the main aspects present on abstract artworks (i.e., colour, shape and texture) to automatically predict emotional valence of them. By using eye-tracking recordings we investigate the link between the detected emotional content and the way people look at abstract paintings. We apply a bottom-up saliency model to compare with eye-tracking in order to predict the emotional salient regions of abstract paintings. Finally, we use the metadata associated to the paintings (e.g., title, description and/or artist statement) and correlate it with the emotional responses of the paintings. This research opens opportunity to understand why an abstract painting is perceived as emotional from global and local scales. Moreover, this work provides to art historians and art researches with a new perspective on the analysis of abstract paintings.

Publications

This thesis is based on the following publications:

- Chapter 2 and 3:

  - **A. Sartori**, V. Yanulevskaya, A. A. Salah, J. Uijlings, E. Bruni and N. Sebe. Affective Analysis of Professional and Amateurs Abstract Paintings Using Statistical Analysis and Art Theory. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2015. (in Press)

  - V. Yanulevskaya, J. Uijlings, E. Bruni, **A. Sartori**, E. Zamboni, F. Bacci, D. Melcher, N. Sebe. In the eye of the beholder: employing statistical analysis and eye tracking for analyzing abstract paintings. *ACM International Conference on Multimedia* pages 349-358, 2012.

  - V. Yanulevskaya, E. Bruni, J. Uijlings, **A. Sartori**, E. Zamboni, F. Bacci, D. Melcher, N. Sebe. Automatic Analysis of Emotions Conveyed by Abstract Painting. *Visual Science of Art Conference (VSAC)*, 2012.

- Chapter 4:

  - **A. Sartori**, B. Senyazar, A. A. Salah, N. Sebe, A. Salah. Emotions in Abstract Art: Does Texture Matter? *International Conference on Image Analysis and Processing (ICIAP)*, 2015. (Submitted)

- Chapter 6:

  - **A. Sartori**, Y. Yan, G.Ozbal, A. A. Salah, A. Salah, N. Sebe. Looking at Mondrian's Victory Boogie-Woogie: What do I feel? *International Joint Conference on Artificial Intelligence (IJCAI)*, 2015. (Submitted)

The following are the papers published during the course of the Ph.D but not included in this thesis:

- **A. Sartori**. Affective Analysis of Abstract Paintings Using Statistical Analysis and Art Theory. Doctoral Consortium. *In Proceedings of the 16th International Conference on Multimodal Interaction (ICMI)*, 2014.

Contents

Chapter 1

Introduction

Images play a key role in everyday life. They can affect people on emotional level and strongly influence the interpretation, memorization, and reaction of people. Indeed, there is an ever-growing digital production of user-generated images and consequently a huge amount of images available online. This growing number of digitized images has raised the request for automated tools to support efficient search of images. Therefore, the automatic categorization of emotions conveyed by images can be a powerful support in many fields, such as advertisement, psychological and cognitive studies, education, journalism, photography and the analysis of art theories. The study of emotions conveyed by images and artworks has been an active topic not only in psychology and visual art research, but also from a computational perspective. However, there is a *gap* between the human perception of the emotions of the images and the interpretation produced by a machine. This issue is currently one of the main challenge of multimedia and computer vision community.

It is known that the various art movements and artists have used different tools and techniques to communicate, to express their ideas and emotions in their artworks. The abstract art, art movement that started in the early 20th century, shifted the emphasis from depicting the reality to exploring different means to evoking emotions in artworks. The abstract artists use the relationship between colour, shapes and textures to convey emotions in a "non-figurative" way: "artists sought to express in their work only internal truths, renouncing in consequence all consideration of external form" Kandinsky [1914]. It is therefore intriguing to understand what makes an abstract artwork emotional and likewise, if a computer algorithm can learn what makes

an abstract artwork emotional.

In this thesis[1], we incorporate the extended studies of art theory and psychology to computer vision and machine learning techniques to automatically recognize positive or negative emotions conveyed in abstract paintings. Specifically, we analyze the patterns of the low-level features (i.e., colour, shape and texture), which are particularly used by artists on abstract paintings to depict emotion. Based on two empirical studies of emotional valence, our approach provides a significant contribution on identifying and quantifying emotional valence regions of abstract paintings. In addition, our study of viewer's eye movements while looking at abstract paintings, contributes to recognize the regions (positive or negative) at which people prefer to look.

In order to understand the specific aspects of each feature, we investigate the contribution of texture and colour combinations taken independently. We made an empirical study on a gray scale version of the paintings and explore the contribution of texture to the classification of emotions elicited by a painting. Moreover, we predict the emotional salient regions of the abstract paintings by applying a bottom-up saliency model. We propose a novel framework to recognize colour combinations of abstract paintings and classify them in positive and negative emotions from a more global perspective. Finally, we introduce a multimodal approach, that combines both visual and metadata (i.e., title, description and/or artist statement) features, in which we investigate how the metadata associated to the abstract painting can be a significant feature to the automatic emotion recognition.

To summarize, the contributions of this thesis are:

- We collected and made available two datasets of abstract paintings. Each dataset is composed of 500 abstract artworks, and they are distinguishable by the professional background of the artists. These datasets can be useful for diverse studies not only for the computer vision and multimedia communities, but also on psychology and cognitive science for instance.

- We apply two methods of annotation (Absolute and Relative Scale) to access the positive and negative emotional responses from people. The results from the annotations are also made publicly available.

---

[1]This research thesis was supported by the the Semantics and Knowledge Innovation Lab (SKIL) of Telecom Italia.

- We use the state-of-the-art computer vision techniques to quantify and identify which part of the image evokes positive or negative emotion. The quantification of evidence for positive and negative emotions gave the possibility to make an in-deep analysis of the abstract paintings and to investigate if there is a positive attentional bias when people look at abstract paintings.

- We provide a user study on the emotional response of texture on abstract paintings and derive texture parameters using an inverse-Perlin parametrization to map the paintings in a psychophysical model of emotions. We also investigate the saliency distribution in the abstract paintings, and show that computational models of bottom-up attention can be used to predict emotional valence in a parsimonious manner.

- We propose a framework to recognize the colour combinations portrayed in the abstract paintings from a more global perspective that improves our approach to the classification of emotions elicited by abstract paintings.

- We present a multimodal approach to emotion recognition of abstract paintings by including the metadata associated to the painting as feature to the automatic emotion classification. We propose a late fusion with weighted linear combination and a novel joint flexible Schatten p-norm model which can exploit the sharing patterns between visual and textual information for emotion analysis of paintings.

Therefore, we believe that our approach can enrich computer vision techniques by understanding emotional patterns and structures of images. Thus, this study offers a useful addition not only to the computer vision and multimedia community, but also from the psychological and art theory point of view.

1.1   Structure of the Thesis

In the next chapters of this thesis we will give a detailed description of our research approach. This thesis is structured as follows:

- In Chapter 2 we investigate the relationship between emotions and artworks and discuss why we use abstract art as reference to our study. In addition, we detail on how we collect the two datasets used in this thesis: one set of professional artists from the Museum of Modern and Contemporary Art of Trento and Rovereto (MART)[1], and the second collection of amateur artists from deviantArt, an online social network site dedicated to user-generated art. Moreover, in this chapter, we describe how we address the issue of measuring emotions evoked by abstract paintings by performing two methods of annotations, Absolute and Relative Scale, to obtain the emotional valence (positive or negative) responses from people. The results from these two methods of annotations are used as ground truth to validate our approaches in this thesis.

- In Chapter 3 we analyse how a painting evokes a positive or negative emotion by determining where the classification evidence resides. We train a Bag-of-Visual-Words model to predict if a painting evokes positive or negative emotions and then backprojected the classification evidence to visualize which parts of the paintings convey the positive or the negative emotion. In addition, we present a eye-tracking study, in which we correlated the localized emotion evidence with eye movements to investigate which part (positive or negative) people prefer to focus on while observing the abstract paintings.

- In Chapter 4 we explore the contribution of texture on the emotional responses to abstract artworks. We report a Relative Scale annotation on MART dataset with a gray-scale version of the paintings. Moreover, we propose a method to derive a small set of features (Perlin parameters) from an image to represent its overall texture. Furthermore, we apply a state-of-the-art of bottom-up saliency model to the abstract paintings and we present an analysis of the relation between its predictions and the human fixations.

- In Chapter 5 we investigate the contribution of colour combinations in a painting. We provide a framework that recognizes two-colours combinations portrayed on the abstract paintings from a more global perspective. We map the paintings

---

[1]The photographs from Mart dataset used in the thesis are property of MART Museum. All rights are reserved by MART Museum.

into Colour Name features and cluster them to generate our Colour Palette, and then, we use the state-of-the-art segmentation technique to generate the colour co-occurrence features in order to classify the paintings in positive and negative emotions.

- In Chapter 6 we explore the influence of the metadata (i.e., titles, description and/or artist's statement) associated with the abstract painting and investigate how this can be a significant feature to the automatic emotion recognition of abstract paintings. In this chapter, we propose a multimodal approach which combines both visual and metadata features in order to improve the machine performance. In particular, we present two methods to combine visual and textual features: the first is a late fusion based on weighted linear combination, and the second is a novel joint flexible Schatten p-norm model which can exploit the sharing patterns between visual and textual information for abstract painting emotion analysis. Moreover, a qualitative analysis is presented on the cases in which metadata help improving the machine performance.

Chapter 2

Emotion Recognition of Professional and Amateur Abstract Paintings[1]

In this chapter we provide the necessary background concerning the study of emotions and the relation between emotion and artworks. We describe the two collections of abstract paintings used in this thesis, and explain why we consider only abstract artworks in our study. In addition, we describe the problem of how challenging is to measure emotional responses evoked by images and artworks and how we address this issue. We applied two methods of annotations, in which we consider the Absolute Scale and Relative Scale study to evaluate the paintings from the emotional point of view, that is detailed in Section 2.5.

2.1 Emotions and Feelings

The human behaviour, communication and interaction, is mostly driven by emotions. In general, emotion can be described by the physiological, psychological and cognitive responses, induced respectively by the body reactions, the brain activity and the interpretation that leads to emotional responses in a particular situation.

Philosophers and psychologists generated different 'emotion theories', based on cognitive, psychological and physiological effects which attempts to understand and explain why humans experience emotion. The *James-Lange theory*, proposed by the

---

[1]This study was published in Sartori *et al.* [2015c]; Yanulevskaya *et al.* [2012a,b].

psychologists William James and Carl Lange, states that emotions are an effect of the physiological interpretation to a particular event. James affirms that "we feel sorry because we cry, angry because we strike, afraid because we tremble, and not that we cry, strike, or tremble, because we are sorry, angry, or fearful, as the case may be" James [1884]. In other words, this theory suggests that the emotional reaction depends on the physiological responses. Another well-known theory of emotion suggested by Walter Cannon and Philip Bard (*Cannon-Bard theory*), assumes that the emotional and physiological reactions to a situation happens simultaneously Cannon [1927]. Differently, the *theory of Schachter-Singer* affirms that "an emotional state may be considered a function of a state of physiological arousal and of a cognition appropriate to this state of arousal." Schachter & Singer [1962]. In other words, this theory assumes that the physical reaction is interpreted and identified in order to experience the emotion. These and many other theories of emotion from the last century are still today debated and supports this area of study Calvo & D'Mello [2010]. Even thought we are aware that a formal theory of emotion still cannot be proven Scherer [2005], based on these theories, we interpret the term 'Emotion' as the cognitive, physiologic and psychological effect caused by human mind and behaviour in a determined context.

In recent years Damasio [1999] conducted neurological studies on the difference between feelings and emotions. He affirms that these terms are usually interchanged and states that emotions happen unconsciously while feelings, instead, are the conscious reaction after the physical change: "This separation is difficult to envision, not only because the traditional meanings of the words block our view, but because we tend to be conscious of our feelings. There is, however, no evidence that we are conscious of all our feelings, and much to suggest that we are not." Damasio [1999].

In the empirical investigation of the psychophysiological elements of aesthetics, Berlyne [1971] points out that "Art is commonly (but not invariably) expected to evoke 'emotions' or 'feelings'. The word 'feeling' has been used in a number of distinct senses, both by psychologists and by laymen, and since, in one of its senses, it is more or less synonymous with 'emotion', we may as well confine ourselves to the latter term." Indeed, the use of terms feeling and emotion, and their relation to art, is still today an open discussion in different research communities. In this work, we apply the study of emotions together with art theory to automatically recognize and classify

emotions evoked by abstract paintings.[1]

2.1.1 Abstract Art and Emotions

During the centuries humans used visual art to communicate, to describe the historical evolution and to express emotions and ideas. The different art movements and artists apply various tools and techniques to convey emotional meaning for artworks. In particular artworks can be considered as a combination of what the artist meant to depict in concert with the appreciation of the artwork beauty and the emotional interpretation from the viewer.

Several psychological approaches attempt to explain art appreciation Bullot & Reber [2013], human perception Cupchik et al. [2009]; Hagtvedt et al. [2008], aesthetics judgments Mather [2012] and emotions Leder et al. [2014]; Silvia [2005] concerning artworks. Hagtvedt et al. [2008] developed a model based on the cognitive and emotional elements that are stimulated by artworks. They employed a structured equation model that integrates the cognitive and emotional elements in the evaluation process and concluded that perception and emotion of visual art depend on factors such as curiosity and aesthetic appeal, complexity, typicality or familiarity. Cupchik et al. [2009] analysed how cognitive control and perceptual facilitation contribute to aesthetics perception together with the experience of emotion, measuring the brain activities by using functional MRI (Magnetic Resonance Imaging). The authors conclude that "aesthetic experience is a function of the interaction between top-down orienting of attention and bottom-up perceptual input". Similar works which discuss perception include Mather [2012]; McManus et al. [2011]; Pelowski & Akiba [2011] and emotional responses to visual art Leder et al. [2014]; Silvia [2005]; Tan & Ferguson [2014]. In a psychological research, Leder et al. [2012] analysed how art is appreciated on the basis of several variables including: the level of expertise in art of the participants, the different classes of artworks (Abstract, Modern and Classic), the measure of emotion (valence and arousal) and the artwork comprehension. Based on empirical studies, Leder

---

[1]We use the word 'emotion' in this thesis following the terminology in the computational literature, but in general 'emotion' may be too strong a word for what we are assessing, and 'feeling' (pleasant-unpleasant) is more appropriate in this context. We use the term 'emotion to avoid misunderstandings from researchers inside the computational community. However, we are aware that these terms may have different meanings in other communities.

et al. [2012] concluded that emotion is the strongest predictor of art appreciation, independently of styles of art and the expertise of people. The work of Leder *et al.* [2012], demonstrates the importance of emotions on the analysis of artworks.

The visual art have been constructing and varying according the historical context and cultures. Indeed, the need of representing and describing the nature such as houses, landscapes, people, animals, have decrease with the emergence of the photography, cinema and press illustration, leading to a progressive growth of the abstract art movement. Jackson Pollock point out that "The modern artist is living in a mechanical age and we have a mechanical means of representing objects in nature such as the camera and photograph" Ross [1990]. In fact, a new way of art expression arises, the abstract art, in which artists express only internal feelings in their artworks, abandoning the external form Kandinsky [1914]

The abstract art movement became known in late of 1911, when some artists such as Vasily Kandinsky and Frantiek Kupka started exhibit artworks with no recognizable content Dickerman *et al.* [2012]. Afterwards, artists initiated to express emotions with forms outer of any visual world references. Indeed, Abstract Artworks exploit the relationships between colours, lines and textures to depict the reality of the artwork in a "non-figurative" way.

The founding director of MoMA (The Museum of Modern Art), Alfred H. Barr Jr., in the mid '30s, illustrated on the exhibition of "Cubism and Abstract Art", the historical development of the cubism and abstract art Dickerman *et al.* [2012]. The Barr's diagram (Figure 2.1), divide the abstract art two general categories: non-geometrical and geometrical abstract art.

The Non-geometrical abstract art, also defined as Lyrical Abstraction, is based on colour expression and it is characterized by dripped paintings, strong contour line and colour forms. It sought an approach to music, where the expressiveness of the sounds turned into artistic language. The Barr's diagram presented in Figure 2.1 remarks the influence of the Fauvism, Expressionism, Futurism, Dadaism and Surrealism art styles on Non-geometrical abstract art. The Fauvism represents artworks in a distinctive variety of colours, "undisguised brushstrokes and high-keyed, vibrant colors directly from the tube" Rewald [October 2004]. In Expressionism instead, spontaneous brushstrokes and vibrant colours are used to express feelings. The Futurism exploits dynamic forms, while the Dadaism and Surrealism express the irrational instincts with forms that are
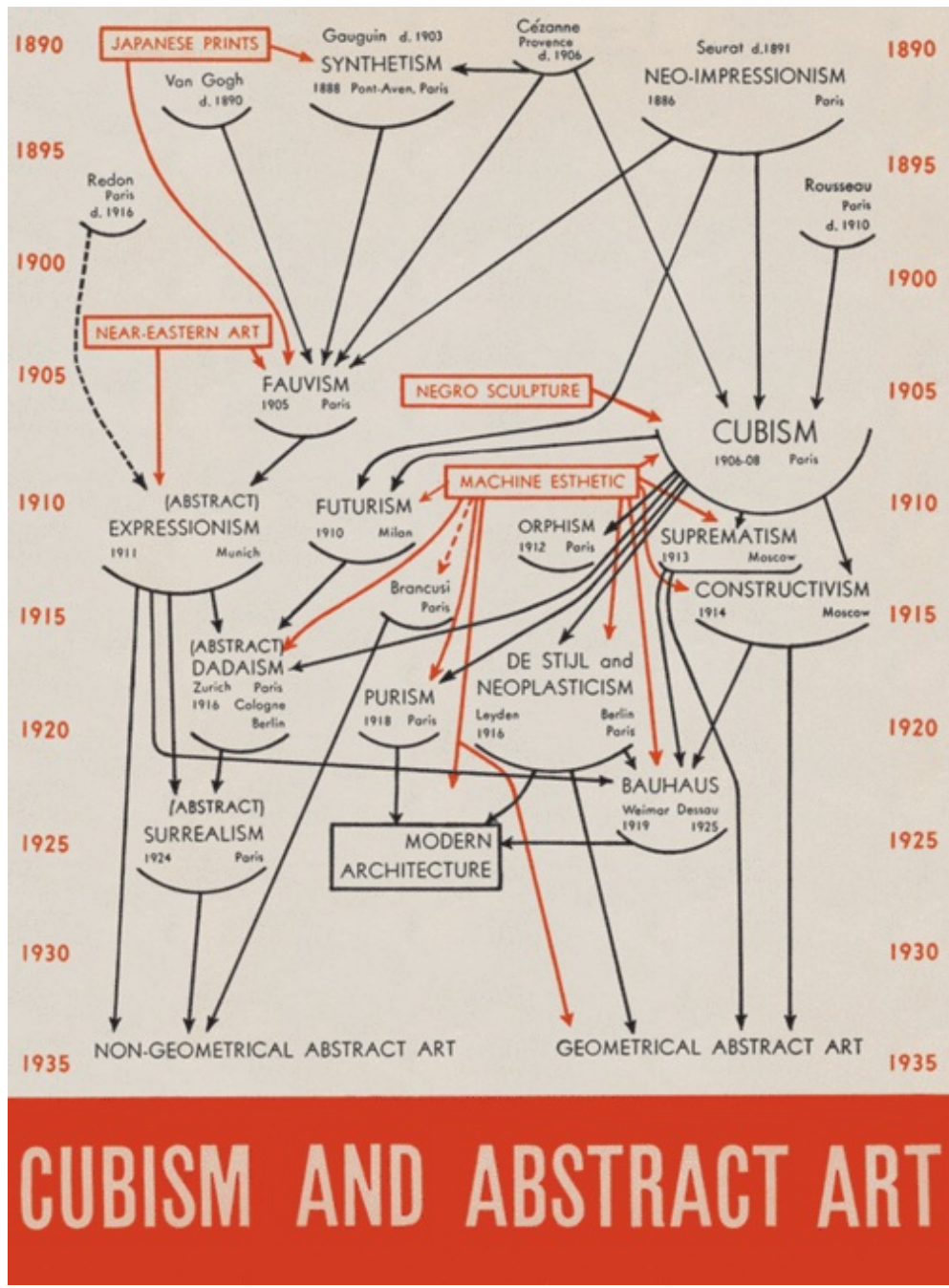
Figure 2.1: Barr's diagram of the historical development of the cubism and abstract art. Cubism and Abstract Art, 1936 Dickerman *et al.* [2012].

outside of the aesthetic patterns Dorfles & Vettese [2001].

The above mentioned art styles were based on the aspects of the emotional expression, with instinctive brushstrokes and a mixture of colours. Differently, the Geometrical Abstract art is based on the rational thinking and characterized by simple geometric forms (i.e., line, square, triangle, etc.) and uniform colours Dorfles & Vettese [2001]. On the Geometrical Abstraction, Barr's diagram outlines the influence of the Neo-impressionism, Cubism, Suprematism and Constructivism. Neo-impressionism, known also as Pointillism, depicts an accurate use of colours into points which allows the viewer's eye do a colour mixing. Cubism is characterized by geometric static forms and tends to a deformation of the subject. Constructivism combined the dynamism of Futurism and the geometry of Cubism. Suprematism employs the geometric shapes (i.e., squares, circles and triangles), in a limited range of colours Dorfles & Vettese [2001].

In this study we used both abstract art segments (Geometrical and Non-Geometrical) to evaluate emotional valence responses. Figure 2.2 provide some examples of Geometrical and Non-Geometrical abstract paintings that compose our dataset of professional abstract paintings.

Figure 2.2: The first two paintings from te left: examples of Non-Geometrical Abstract paintings ("Kleine Welten VIII", 1922, Kandinskij, Wasilij; "Concetto spaziale", 1956, Lucio Fontana). The third and fourth paintings: examples of Geometrical Abstract paintings ("Composizione(Sequenza)", 1967, Aldo Schmid; "Study for Homage to the Square: Still Remembered", 1954-1956, Josef Albers). ©*MART - Archivio fotografico e Mediateca.*

2.2   Why to Focus on Abstract Art?

Abstract art revolutionized the visual art, which lost its most common task, the one of representing or narrating world's objects, nature and facts. Artists use non-figurative elements to convey emotions (i.e. colours, lines, shapes, and textures) and, unlike figurative art, abstract art does not depict objects and scenes that are familiar to our visual system. Even though, abstract art became a popular art movement on its time and it is still successful on museum exhibitions. Indeed, it is therefore intriguing to see what fascinates people while they are looking at abstract artworks.

In his *Defense of Abstract Art*, Moholy-Nagy comments on how abstract art focused on articulating basic means of visual impact by using "shape, size, position, direction, point, line, plane, colour, rhythm". The main reason of this focus was the need to experiment with basic shapes to see what kind of emotional impact they would have Moholy-Nagy [1945]. As Melcher & Bacci [2013] state, from the perspective of visual neuroscience, the physical reasons of this emotional impact is not hard to trace: our visual system is built to perceive certain visual features including contours, shapes, colours, depth, texture and movement. All these building blocks (except movement) are also the building blocks of abstract art. Hubel [1998]; Zeki [1992] observed that when looking at a visual stimulus, our visual system initially deconstructs the visual scene into elementary components (such as colours, lines, simple shapes, etc), then, it reconstruct to more complex objects and scenes, as described in the work of Aviv [2014]; van Paasschen *et al.* [2014]. Hence, there are many studies show that subjects, even children, agree on the emotional and expressive nature of abstract shapes and colors Blank *et al.* [1984]; Leder *et al.* [2011]; Takahashi & Ejima [2013]; Winner *et al.* [1983].

In particular, abstract artists were concerned about the basic elements of visual art and how these elements and their various compositions affect the viewer. They created theories, based on intuition, observation, experiments, and expertise, to derive rules that describe this process Arnheim [2004]; Itten [1974]; Tomas [1969], and applied these studies to their artworks. Indeed, abstract artwork offers a good ground to study the visual fundamentals and the emotions they generate in the viewer.

The semantic content (or top-down information), present on figurative art, leads to an interpretation of what is depicted. Conversely, in abstract art, the information

conveyed is bottom-up or low-level, i.e., it is free from the common restrictions that the visual system is used to. Kandinsky described the effect of colours on the spirit of the observer "not a 'subjective' effect, but an objective one, determined by the characteristics of the colours and their interactions" Veronesi [1968]. El Lissitzky and Ilya Ehrenburg founded a journal in Berlin that aimed to countering the "tyranny of the subjective in art in favour of the systematization of the means of expression to produce results that are universally comprehensible." Dickerman *et al.* [2012]. In fact, the lack of semantic content makes abstract artworks an interesting study material as it is possible to investigate aspects and patterns that are not present in figurative art. Moreover, it can induce similar emotions in the observers, as there is consistency in the way people interpret colours and patterns when looking at paintings.

Kandinsky describes the abstract art as a revolution of the spirit, where the artist expresses through colors and shapes their inner need. Sadler, in the introduction of the book of Kandinsky [1914], affirms that "he has broken down the barrier between music and painting, and has isolated the pure emotion which, for want of a better name, we call the artistic emotion." Another notorious abstract artist, Jackson Pollock affirms that the artist "is working and expressing an inner world - in other words - expressing the energy, the motion and the other inner forces[...] the modern artist is working with space and time, and expressing his feelings rather than illustrating" Ross [1990]. Therefore, the abstract art presents an interesting challenge to the study of emotions, as it stands as a pure instrument of expression, in which forms and colors are the most important and the only characteristics of emotion expression.

2.3 The Datasets of Professional And Amateur Abstract Artworks

In this section we present the two collections of abstract paintings used as element to validate our methods in this thesis. These two datasets, each one composed by 500 abstract artworks, distinguish each other from the origin of the artists. The first dataset is a collection of 500 abstract paintings from notable artists of the Museum of Modern and Contemporary Art of Trento and Rovereto (MART). The second dataset is composed by 500 abstract paintings of amateur artists from deviantArt, an online

social network site devoted to user-generated art.

2.3.1 The MART Dataset: A collection of Professional Abstract Artworks

The Museum of Modern and Contemporary Art of Trento and Rovereto (MART) is one of the most important centre of modern and contemporary art in Italy with a substantial archive from the twentieth century art to the contemporary. The collection of the MART contains over 20,000 works of art including paintings, drawings, engravings and sculptures. Out of the electronic archive of MART, with the guidance of an art historian, 500 images of abstract paintings were selected Melcher & Bacci [2013]; Sartori *et al.* [2015c]; Yanulevskaya *et al.* [2012b].

The selected artworks date from 1913 to 2008 and belong to 78 artists, mainly from Italy, but also from Russia, Europe and USA. Among these artists, the most notable ones are Wassily Kandinsky, Josef Albers, Paul Klee, Luigi Veronesi, Aldo Schmid, Carlo Belli and Luigi Senesi. In particular, Kandinsky, Albers and Klee, discussed and formulated some of the art theories of abstractionism. Their artworks are a reflection of their principles and theories on colour, lines, shapes and textures. On this dataset[1] we applied two empirical studies to evaluate the emotional response from people for each selected artwork: respectively Absolute and Relative Scale annotations.

2.3.2 The deviantArt Dataset: A collection of Amateurs Abstract Artworks

DeviantArt (dA) is an online social network website devoted to creation, sharing, dissemination and marketing of user-generated artworks. It is one of the largest online art communities with more than 220 million artworks and 30 million registered users. The site hosts amateur and professional artists as well as art connoisseurs, critics and layman alike.

DeviantArt (dA), is an online space reserved for artworks only. For example, unlike Flickr, which as an online social network website have the biggest image collection

---

[1]The images and the emotional scores of MART dataset have been made available online: http://disi.unitn.it/~sartori/datasets/mart-dataset/

and membership number, dA does not allow for batch uploading of images. Instead, dA members, who are called '*deviants*' in dA jargon, have to oblige to certain rules when sharing an image: they have to give a title and a description of their artwork, as well as to indicate which category the artwork belongs to. dA offers a complex categorization system which grows as the dA archive grows. These strict rules invite members to choose each image with care, and to ponder on its artistic nature. Through this intricate upload process, dA also makes sure that the artworks in its archive have comprehensive metadata and are linked via a categorization system. These factors turn dA into a valuable source of (mostly amateur) artworks.

The categorization system of the dA makes possible to crawl artworks in a specific genre only. For the purposes of our study we have crawled the dA site the images that are shared under the category Traditional Art/Paintings/Abstract Art, and downloaded initially 8,000 artworks.

Out of these artworks, we first manually excluded all the artworks that have recognizable objects. To further downsize our data to a manageable number of 500[1], we made use of additional information from each image that is available on dA website. For each downloaded abstract painting, were available the following statistics from the site: the number of pageviews, the number of comments and the number of favourites.

Favourites is a 'like' system in dA: every member can add a *deviation*, which in dA jargon means the deviant artworks, to their 'favorites' by hitting a 'plus' button. Once a *deviation* is favoured, it is added to the homepage of the deviant who favors it. Through this system, each deviant generates a gallery of artworks, and displays them on his dA homepage. The site statistics bestows a certain status to *deviants*, i.e. members with higher number of pageviews/comments/favorites are more successful and popular in dA community. This so called 'power' status furthermore gives an influence to successful members: if a popular member favours a *deviation*, this will probably cause an increase of traffic to such an artwork, likely increasing its pageviews and other statistics.

We checked the correlation among different information attributes of the *deviations* and their success in terms of favourites and comments. It results that the number of comments and favourites are correlated, while the the number of pageviews and the

---

[1]We selected only 500 paintings to construct the deviantArt dataset in order to make an impartial comparison with the MART dataset.

time of upload are not correlated. In particular, if a member takes the time to comment on a *deviation*, this is probably because the *deviation* has an impact on that member, and it is likely that they also favours the *deviation*. On the other hand, the time of upload does not influences the quantity of comments, favorites and pageviews of a *deviation* as the number of the pageviews is probably dependent on the social position of each member: the more popular a member is, the more pageview numbers they will get.

Since there is no correlation between the time of upload of a *deviation* and its site statistics, we disregarded the time of creation/upload while downsampling our dataset. In order to have a distribution of the most/least popular painters, we decided to use the number of favorites, and select the most favourite paintings, the least favourite and randomly in the middle. Some of these 500 chosen paintings[1] have the same author. In total, there are 406 different authors for the 500 paintings.

2.4 Emotion Recognition: How Can Emotions Be Measured?

In recent years the study of emotions has become a relevant research topic in many fields. Several works had observed that emotions are directly associated with human attention Vuilleumier & Huang [2009], perception and memory Phelps [2006]. The advancement of computer technologies and the improvement of tools and sensors to detect human physiological signals increased the interest in detecting emotional information. Moreover, it has been shown that emotion recognition allows for improvements in human-computer interaction Sebe *et al.* [2005], image retrieval Joshi *et al.* [2011] and enhances computers' abilities to make decisions Picard [1995]. For this reason, the extensive psychological and philosophical research in emotions have been incorporated into the engineering and computer science research, which leverage the approaches to machine learning and computer vision techniques to identify automatically emotions conveyed by images, video, voice and text Calvo & D'Mello [2010].

However, extracting emotion automatically is considered by the research commu-

---

[1]The images and the emotional scores of deviantArt dataset have been made available online: http://disi.unitn.it/~sartori/datasets/deviantart-dataset/

nity a challenging task, due to the difficulty on measuring or defining emotion. It is difficult to understand how to deal with human personality, the social and cultural environment, background knowledge, life experiences and past emotions. Moreover, the expressions and perceptions of human emotions are ambiguous. Humans are able to blend and mask emotions, as well as the emotions can exhibit conflicting (i.e., positive and negative) aspects of the expression, such as weeping for joy. In addition, it is difficult to differentiate between emotions that may be very similar, such as irritation and anger Mower *et al.* [2011].

Some of the psychological and neurological works assumes that emotions can be measurable and physiologically distinct. The current research on emotion recognition generally uses two of the most popular psychological approaches to classify emotions: discrete and dimensional. The discrete approach is commonly based on Ekman's work, who categorizes emotion in six basic emotional groups: happiness, sadness, anger, fear, surprise and disgust Ekman [1993]. In the dimensional approach, based on Russell [1980], emotions are categorized along *arousal* and *valence* dimensions. Arousal delimitate the intensity of emotion, and valence measures the positive (e.g., joy, pleasure, etc.) and negative (e.g., sadness, disgust, etc.) emotions. Some studies use an additional dimension, i.e., *control* or *dominance*, which ranges from "no control" to "full control". However the *control* or *dominance* dimension is not common used in affective multimedia content analysis Hanjalic [2006].

Several works that deal with emotional analysis of images and artworks reports that using the dimensional approach has as advantages the disambiguation of the emotional states and the consistency with the way the brain processes emotion Hanjalic [2006]; Leder *et al.* [2012, 2014]; Lu *et al.* [2012]; Pihko *et al.* [2011]. Nicolaou *et al.* [2011] points out that the automatic emotion recognition models "are deemed unrealistic as they are unable to capture the non-basic and subtle affective states exhibited by humans in everyday interactions". For this reason the actual research in this field usually adopts a dimensional approach that are "sufficient for capturing most of the affective variability: valence and arousal (V-A), signifying respectively, how negative/positive and active/inactive an emotional state is." Nicolaou *et al.* [2011]. According to Leder *et al.* [2012] it is difficult to evaluate the effects of arousal in art appreciation: "arousal might also be closely connected to emotional valence and, therefore, affect appreciation in a very similar way. In this case, emotion and arousal would show high in-

terrelations." Csíkszentmihályi & Robinson [1990] remarks that emotion is the most important factor of the aesthetic experience; people tend to associate artworks with emotional valence. By following the prominent psychological and neuropsychological studies in emotional preferences of artworks abovementioned, we can conclude that positive-negative emotional valence can disambiguate the affective states and is adequate to quantify emotional responses on abstract paintings. In this study we apply the dimensional approach of emotion, using positive and negative valence to quantify the emotional message evoked by abstract paintings.

Even though, labeling abstract paintings as positive/negative emotions is an arduous task. For this reason we study two different approaches to collect the ground truth of emotions elicited by abstract paintings that are presented in the following section.

2.5 Annotation Methods

In this section we describe the two methods of annotations used to measure the emotional responses on abstract paintings. We employ an Absolute and Relative Scale approach to collect the ground truth of positive and negative emotions evoked by the abstract artworks. In the Absolute Scale method we ask subjects to rank images on a Likert scale from 1-to-7, while in the Relative Scale method, pairs of paintings were given to the subjects and were asked to choose the most positive one.

The reason we choose the Absolute Scale method to annotate the abstract paintings in positive and negative valence mainly to allow the subjects, in a somehow specific way, to evaluate emotionally the abstract paintings between the two broad valence emotional categories. The choice of applying a Relative Scale method to emotionally evaluate abstract paintings is that we believe that judging between two paintings, placed side by side, is a relatively straightforward setting, which may result more consistent results. Indeed, the latter, demonstrated to be more adequate to our purpose that is discussed in details in Section 2.6.3.

2.5.1 Absolute Scale Annotation

To annotate the paintings on the Absolute Scale method, people were asked to judge according to a Likert scale of 1-to-7-points, where 1 meant a highly negative emotion

and 7 meant a highly positive emotion. The participants got the following instructions: "You are asked to judge all the paintings that will be presented. Let your instinct guide you and follow your first impression of the painting." Moreover, was emphasize to the participants that they were not asked to state whether they 'like' the artwork or not, although only to self state the degree of positive/negative emotion aroused while they are seeing the artwork. As there was not a fixed time limit to evaluate the paintings, during the experiment, the participants were encouraged to score the paintings as fast as possible, in order to get instinctive emotions and make sure that they may not be influenced by a possible prior knowledge about the painting. Also, the paintings were shown to the user without any title/artist information, to make sure that the participants were not influenced by the painting's title or by its author.

In the experiment, the paintings were randomly divided in five groups of equally distributed 100 paintings each. Each participant was exposed exclusively to one group of paintings and rated 100 paintings. In total, each artwork received 20 judgments from 20 different subjects. From these 20 judgments per painting, we calculated the average score and obtained one score per painting. We presented the paintings on a grey background, as grey is usually considered to be a neutral colour, with no impact on other colours. After every ten images, a grey slide was presented in order to give participants the opportunity to rest.

2.5.2 Relative Scale Annotation

On the Relative Scale method, paintings were shown in pairs to the subjects and were asked to choose the more positive painting in any given pair. However, if we compare each painting in the dataset with all other paintings, we will need $N * (N - 1) * 0.5$ comparisons where N is the number of paintings in the dataset. In our case, $N = 500$ which makes 124,750 comparisons. Instead, we used the TrueSkill ranking system Herbrich & Graepel [2006] to get a reliable ranking by optimally sampling pairs of paintings. The TrueSkill ranking system was developed by Microsoft Research for Xbox Live to identify and rank the skills of the players. It allows to match players with similar skills for a new game. The TrueSkill is a Bayesian rating system that infers the marginal belief distribution over the skill of each player. It calculates the average skill of a player and the system's certainty associated with it. It was shown Herbrich &

Graepel [2006] that 15 games are enough for a reliable estimate. Therefore, by using TrueSkill, we need to annotate only $N * 15 * 0.5$ pairs of paintings in order to obtain a reliable emotional score, which means that we need only 3,500 comparisons instead of 124,750.

We used the TrueSkill implementation of Moser [2010]. At the beginning of the annotation, all paintings have the same 'skills'. During each comparison, we assume that the painting which is chosen as more positive wins the 'game', and we update the ranking of the compared paintings accordingly. Afterwards, the paintings with similar rankings are compared, until each painting is compared with at least 15 other paintings. The final 'skills' are considered as emotional scores, the lower values correspond to negative emotions and higher values to positive emotions.

During the experiment, the subjects were presented with an interface with two paintings next to each other together with buttons, which allowed them to select one of the paintings. The participant got the following instructions: "Which painting in the pair looks more positive to you? Let your instinct guide you and follow your first impression of the paintings. You will see two abstract paintings next to each other, please, choose the one which evokes more positive emotions. Notice that the task is not about 'liking' one image more than another. Sometimes a very beautiful painting might be very sad and even fearful, in this case the painting should be still considered as negative."

2.6 Applying the Annotation Methods on MART Dataset

2.6.1 Absolute Scale Annotation Procedure

On the Absolute Scale method, a total of 100 people were recruited for the annotation, including visitors of the museum, teachers in primary schools, museum curators at MART and students. The subjects were Italians and the experiment was made in Italian language. From the 100 subjects, 74 were females and 26 males, with ages between 18 to 65 years old ($M = 39.87$). The participants reported to visit from a minimum of 1 to a maximum of 100 museums per year ($M = 5.5$). All subjects participated in our

experiment voluntarily without getting any reward.

Besides recording the annotations, we have also checked the preferential biases according to gender (74 females versus 26 males), age (48 people under 40 years old versus 52 over 40 years old) and art background. The 31 participants who declared to visit more than 5 museums per year are considered to be knowledgeable in art, whereas the rest are not (69 people). We observed that there is not a significant difference in the emotional evaluation of the paintings between these groups (p values of a double t-test ≥ 0.460).

2.6.2 Relative Scale Annotation Procedure

For the Relative Scale method of annotation, 25 subjects participated, in which 11 were females and 14 males. Each person annotated from 19 to 356 pairs of paintings, with a mean of 145. The participants were anonymous and they participated in our experiment voluntarily, without getting any reward. The annotation was done online. There was no time-limit to the annotation procedure: each participant was free to annotate whenever they had time to do it.

2.6.3 Comparison between Absolute and Relative Scales on MART Dataset

In this Section we compare the emotional evaluation between the two different annotation methods, i.e., Absolute and Relative Scale, applied on MART dataset. Figure 2.3 presents montage of MART dataset arranged by the annotation scores each painting has received from subjects. The Figure 2.3 provides two different orderings of paintings: one according to Absolute Scale scores (left column), and one according to Relative Scale scores (middle column). Each montage is assembled with the thumbnails of each painting, ordered from the most negative to the most positive according to the mean emotion scores they have received. The black line is the threshold dividing the dataset in positive and negative paintings. Each montage is a matrix of 50*10, where 1st row/1st column contains the most negative painting, and 50th row/10th column contains the most positive painting. So, the painting on the second row/first column is ranked 11th negative and likewise, the third row/first column is 21th and so on.

Figure 2.3 gives an insight into the different rankings obtained by using the Absolute and Relative Scale annotations on MART dataset. The Absolute Scale method provides to the subject the choice to rate the paintings in a scale of 1-7 of valence emotions from 'more negative' to the 'more positive'. The advantage of this method of annotation is that people can give a degree (or a specific rating) of the two broad emotional valence dimension (negative and positive). Moreover, this method allows to evaluate the agreement between annotators for each of the five groups of paintings.

To calculate the agreement between the annotators we use the Intraclass Correlation Coefficient (ICC) of Shrout & Fleiss [1979]. The ICC is a measure that estimates the inter-rater reliability, i.e., gives the degree of agreement between raters. The rule of thumb of the ICC coefficient is: 0 to 0.2 indicates poor agreement; 0.3 to 0.4 indicates fair agreement; 0.5 to 0.6 indicates moderate agreement; 0.7 to 0.8 indicates strong agreement; and higher than 0.8 indicates almost perfect agreement.

According to Shrout & Fleiss [1979], the reliability measure can be defined in three cases, depending of the study design:

- One-Way Analysis of Variance: each target is measured by a different set of raters from a randomly selected population of raters.

- Two-Way Random Analysis of Variance: each target is measured by the same raters from a larger population of random raters.

- Two-Way Mixed Analysis of Variance: each target is measured by the same raters and they are only raters of interest.

Moreover, Shrout & Fleiss [1979] outlines that for each of the above mentioned cases, the ICC can be applied in two ways: on a single rating or the mean of several ratings, in which totalize six different types of the ICC.

Since each group of paintings was rated by the same subjects, we estimate the reliability of the mean of several ratings on the two-way random intra-class correlation (ICC) to analyse the rating scores from each subject for each painting.

To calculate the ICC, we followed the method of van Paasschen *et al.* [2014], which first converted the ratings scores to standardized z-scores to correct the bias across subjects. The ICC coefficient with 95% of confidence interval for each group of data has

in the most of cases an almost perfect agreement (ICC Group 1= 0.847, CI= 0.800-0.887; ICC Group 2=0.841, CI: 0.792-0.883; ICC Group 3=0.726, CI=0.641-0.798; ICC Group 4=0.853, CI=0.807-0.891; ICC Group 5=0.879, CI=0.842-0.911). This result suggest strong agreement on positive and negative ratings for different paintings between participants, which shows that people are consistent when emotionally evaluate abstract paintings.

Afterwards, as suggested by van Paasschen *et al.* [2014] we verify the frequency of the mean rating of each artwork, in order to observe if the agreement scores is not a reflection of a vast amount of neutral ratings. For this calculation, we consider that mean scores lower than 3.5 as low valence rating, higher than 4.5 as high valence rating and rating between 3.5 and 4.5 as neutral. As a result 67.2% of the artworks received a mean valence rating that was either lower than 3.5 (total of 80 paintings) or higher than 4.5 (a total of 256 paintings). For 32.8% of paintings (a total of 164 paintings) has a neutral mean valence rating. We then replicate the ICC coefficient excluding the artworks with neutral mean ratings. The high consistency ratings remained high for valence scores, except for Group 3 which has fair agreement (ICC Group 1= 0.777, IC= 0.695-0.845; ICC Group 2=0.726, CI=0.622-0.812; ICC Group 3= 0.385, IC=0.163-0.570; ICC Group 4= 0.771, IC=0.677-0.848; ICC Group 5=0.765, IC=0.670-0.842) with 95% of confidence interval. By analysing the paintings that composed the Group 3 of annotation, we observe that they are mostly composed by neutral colours (i.e., shades of grey and brown) or composed by both light and dark colours, which may be the reason that people were not consistent as they were in the other groups.

The Relative Scale method allows people to choose the more positive paintings out of a pair. This method has as advantage that is easier to judge between two paintings that are placed side by side than to give an absolute value. Another advantage of this method of annotation is that it was done online and people were free to annotate as much as they felt comfortable, and they could annotate any time they wanted. In addition, The Relative Scale setup requires shorter time to accomplish the entire annotation than Absolute Scale, as well as is less costly (e.g., there is no need to have one person to control the experiment). As we conduct the study via Internet, we were also able to collect the emotional data from a diverse population which encompasses a wide social and cultural environment. On the other hand, we do not control the conditions that these annotations are collected.

As the rating system determines annotation sampling online and never presents the same pair of images to two different annotators, we have no overlapping annotations. This makes the computation of an inter-annotator agreement impossible. We have, however, matched the True Skill ratings obtained at the end of the user study to the individual annotations of the pairs, presented to all the annotators. The results show about 80% agreement, which is effectively the mean human performance for the valence classification task, and represents an upper bound for our automatic algorithm. This result shows that also Relative Scale is an effetive method for the evaluation of positive and negative emotions in abstract paintings.

2.7 Applying the Annotation Method on deviantArt Dataset

Given the comparison previously reported, we assume that Relative Scale method is more suitable for our purpose, and for this reason, on deviantArt dataset we apply only the Relative Scale method for annotation.

In total 60 people participated, 27 females and 33 males. Each participant annotated from 2 to 436 pairs of paintings, with a mean of 63.0. The participants were anonymous and agreed to be part of our experiment voluntarily, without getting any reward. Also for deviantArt we matched the True Skill ratings obtained at the end of the user study to the individual annotations of the pairs, presented to all the annotators, in which show about 74% agreement. Figure 2.3 shows the montage of dA dataset with Relative Scale method (right collumn). The paintings are ordered from the most negative to the most positive according to the emotion scores they have received.

2.8 Discussion

In this chapter we discussed the study of emotions and their relation to abstract artworks. In our study we used two different collections of abstract paintings: one Professional set from MART Museum and one Amateur set from deviantArt, a user-generated art website. Moreover, we discuss the difficulty of measuring emotional responses and address this problem by analysing two different methods of annotation. We conducted an Absolute and Relative Scale study to evaluate the paintings from the emotional point

of view.

The Figure 2.3 shows both collections as a montage arranged by the annotation scores each painting has received from subjects. A comparison between MART and dA datasets is also imminent: MART dataset is compiled from paintings with extremes whereas dA paintings in general have an even distribution of colour and saturation. From very dark paintings to very light paintings, MART dataset offers many different types of abstract art examples. In contrast, dA collection seems to be drawn with the same colour palette by the same artist. This is a striking difference, and could be explained by taking into account two different factors: The first is the dimension of time. MART dataset covers the whole 20th Century, whereas the paintings from deviantArt are all recent creations. The second reason is the impact of digital tools. Even though our dA dataset has paintings that are generated via traditional means (i.e. with traditional drawing tools such as pen, paint, paper, canvas etc.), the artists working with traditional tools also make use of digital tools, either during their creation process of traditional artworks, or they separately produce digitally created artworks as well.

We observe in Figure 2.3 that for both datasets, the paintings are mostly ordered by a gradient of colours from the most negative to the most positive paintings. Where the most negative are the paintings with dark colours and the most positive are the paintings with lighter colours. The results of the annotation analysis provides a conclusion that for both methods of annotation, Absolute and Relative, people are consistent on evaluating in positive/negative valence emotions in abstract paintings.

We used the results of both Absolute and Relative Scale annotations as ground truths to a classification model that will be described in Chapter 3, in which we obtained higher accuracy when the ground truth was created using a relative ranking. Indeed , we use the Relative Scale ground truth to the study of Texture, Colour Combination and Titles, respectively presented in Chapters 4, 5 and 6.
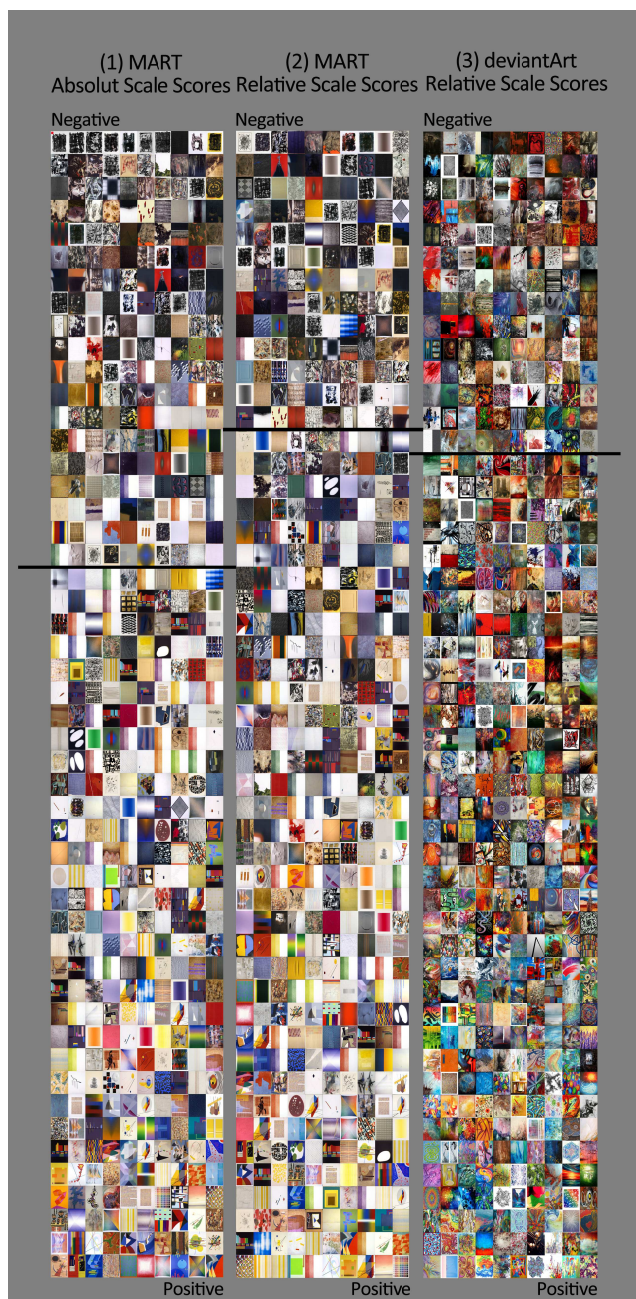
Figure 2.3: Paintings from MART and deviantArt datasets ordered from most negative (on top) to most positive (on bottom): (1) MART-Absolute Scale, (2) MART-Relative Scale (©*MART - Archivio fotografico e Mediateca*) and (3) deviantArt - Relative Scale (*Courtesy of deviantArt*).

Chapter 3

Statistical Analysis for Assessing Emotions[1]

In this chapter we turn to the exploration of the content within the abstract paintings. Specifically, we present an analysis of why/how a painting stimulates a certain emotion by determining where the classification evidence resides. To this effect, we train a Bag-of-Visual-Words model to predict if a painting evokes positive or negative emotions. We use the two collections of abstract paintings presented in the Section 2.3, the MART dataset and the deviantArt dataset. These two sets, which comes from different (i.e. professional and amateur) backgrounds, have one important aspect in common: in all paintings, the (professional and the amateur) artists use only basic visual stimuli such as forms and colour for their compositions. These visual stimuli offer properties that are well captured by Bag-of-Visual-Words. We validate our method with the two ground truths of Absolute and Relative Scale detailed in Section 2.5. We furthermore make use of the method in Uijlings *et al.* [2012] to dissect the classification machinery to determine which parts of an image evokes what emotions. Moreover, as an application, we test attentional bias, which states that people prefer to look at the positive parts of an image, regardless of the overall image tendency towards negativity or positivity.

To summarize, the contributions presented in this chapter are: (1) we use computer vision techniques to quantify which parts of the image evoke positive and negative emotion; (2) we present an in-depth analysis investigating whether there is a positive

---

[1]This study was published in Sartori *et al.* [2015c]; Yanulevskaya *et al.* [2012b].

attentional bias when people look at abstract paintings; (3) we test our approach using two ways to annotate emotions, i.e. Absolute and Relative Scale; described in Section 2.5; and (4) we compare the classification results of the two different datasets, i.e. paintings from professional and amateur artists detailed in Section 2.3.

3.1 Related Work

The recent works that observe emotional responses evoked by images deal with interdisciplinary fields such as art, aesthetics and psychology Joshi *et al.* [2011]. In particular, there is an increase of interest of computer vision and multimedia research on the automatic emotion recognition of artworks. As introduced on the Chapter 2, the main reason of this interest relies on that artworks are specifically composed to convey emotions.

The progress of computer vision research has become mature enough to predict emotions Machajdik & Hanbury [2010]; Yanulevskaya *et al.* [2008], aesthetics Marchesotti *et al.* [2011], and interestingness Isola *et al.* [2011] of images, paintings, and even web pages Wu *et al.* [2011]. Yanulevskaya *et al.* [2008] proposed an emotion categorization system based on the assessment of local image statistics followed by supervised learning of emotion categories using Support Vector Machines. This system was trained on the International Affective Picture System, which is a standard emotion evoking image set in psychology Lang *et al.* [1999], and was then applied to a collection of masterpieces. The work by Machajdik & Hanbury [2010] combined low-level features with concepts from psychology and art theory for categorization of emotion in affective images and art works. Surprisingly, they obtained better accuracy in categorizing affective images that are semantically rich in comparison with abstract paintings that are non-figurative and thus semantics free. Lu *et al.* [2012] used a statistical analysis to investigate how shapes are related to emotions in natural images. Recently, Zhao *et al.* [2014] extracted emotion features from images based on art principles, including balance, emphasis, harmony, variety, gradation, and movement. They also used the IAPS dataset of Lang *et al.* [1999] to emotionally classify the images, and concluded that these features can improve the performance of emotion recognition in images. Another works which deal with automatic emotion recognition in images include Jia *et al.* [2012]; Li *et al.* [2012]; Solli & Lenz [2009]; Wei-ning *et al.* [2006].

The abovementioned works, demonstrate that it is possible to use computer vision techniques for emotion recognition.

Understanding how people observe paintings is essential for determining the features that evoke emotions in the viewer's mind. Analysing eye movements and fixations has been used as an important tool to understand the human perception of an artwork and the features that are closely related to a specific emotion Pihko *et al.* [2011]. Moreover, emotional stimuli were shown to attract attention. Subramanian *et al.* [2010, 2011] demonstrated how eye movements can be used to understand social and affective scenes. Recently, Liu *et al.* [2014] demonstrated that the use of emotion information improves the gaze density estimation. In this study, we investigate the link between emotional content and the way people look at abstract paintings. In particular, we take a close look at the way positive elements of a painting may influence the emotion perceived by the viewer.

3.2 Proposed Method

Given the above exposed advances in computer vision related to images and art, we aim for a better understanding of why a painting evokes a certain emotion by determining where the classification evidence resides. In particular, we use a state-of-the-art Bag-of-Visual-Words classification framework which largely follows Uijlings *et al.* [2010] in order to learn the difference between positive and negative paintings.

3.2.1 General Bag-of-Visual-Words Framework

The concept of Bag-of-Visual-Words is inspired by a method used in text analysis for retrieving or classifying documents Sivic & Zisserman [2003]. In the case of Bag-of-Visual-Words, an image is divided into small image patches. Each of these patches is matched to a visual word from a previously learned visual dictionary. A visual word can be thought of as a specific image pattern, such as "green stripes" or "black spots on a white background". The final image representation is a frequency histogram of the visual words present in the image Csurka *et al.* [2004]. The standard pipeline to form a so-called "visual vocabulary" consists of (1) collecting a large sample of features from a representative corpus of images, and (2) quantizing the feature space according to

their statistics. Often, k-means clustering is used for the quantization. In that case, the visual "words" are the k cluster centers. Once the vocabulary is established, the corpus of sampled features can be discarded. Then, the features of the new image can be translated into visual words by determining which visual word they are nearest to in the feature space (e.g., based on the Euclidean distance between the cluster centers and each descriptor feature) Csurka *et al.* [2004]. Finally, each image is represented as a histogram of its visual words.

3.2.2 LAB and SIFT Descriptors

We use low-level features (colour, texture and shape) to measure the emotional information of abstract artworks. To extract these features, we represent paintings with LAB-based colour visual words and SIFT-based texture visual words.

LAB

The LAB colour space is a 3 dimension space where L defines the lightness and *a* and *b* defines the colour-opponent dimensions, based on CIE XYZ colour space coordinates. In LAB, luminance corresponds closely to brightness as recorded by the brain-eye system, and chrominance (red-green and yellow-blue) axes mimic the oppositional color sensations the retina reports to the brain Szeliski [2010]. In order to map LAB descriptors to visual words, we quantize the colour space into 343 different colours by dividing uniformly each colour into 7 different bins.

SIFT

The Scale-Invariant Feature Transform, or SIFT, proposed by Lowe [2004] describes the local distribution of contours, edges and textures within images. The SIFT descriptor is invariant to image scale and rotation and is efficient in matching across noise, affine distortion and illumination changes. In this work, local patches of 16-by-16 pixels were taken and sampled at every single pixel. From these patches we extract grey-scale SIFT descriptors and two colour variants, RGB-SIFT and RGI-SIFT as recommended by van de Sande *et al.* [2010]. To create a visual vocabulary we quantize 250,000 randomly selected SIFT descriptors into 4096 clusters using a hierarchical

implementation of k-means clustering Vedaldi & Fulkerson [2010].

3.2.3 Classification

The extracted representations for abstract paintings are used to train a classifier to distinguish between the positive and negative emotions. We use a Support Vector Machine classifier with a Histogram Intersection kernel for supervised learning of emotions, using the fast approximation of Maji *et al.* [2008]. We run two separate frameworks, one based on LAB descriptors, and another one based on SIFT descriptors. To combine both frameworks, we use late fusion, i.e. we average the scores of the two frameworks.

3.2.4 Backprojection

Backprojection (BP) is a technique which determines the relative contribution that comes from each pixel of the image to the classification task. In our case, using BP we are able to detect those visual words which convey the most positive or negative information according to our classifier. In other words, this technique permits us to label a certain region as being "positive" or "negative", as we might expect human subjects to do. To implement BP, we follow Uijlings *et al.* [2012]. Consider the classification function for the Histogram Intersection kernel, defined as

$$h(\mathbf{x}) = b + \sum_{j=1}^{m} \alpha_j t_j \left(\sum_{i=1}^{n} \min(x_i, z_{ij}) \right), \tag{3.1}$$

where $\mathbf{x} = \{x_1, \ldots, x_n\}$ is the vector to be classified, $\mathbf{z}_j = \{z_{1j}, \ldots, z_{nj}\}$ is the j-th support vector, α_j is its corresponding positive weight, $t_j \in \{-1, 1\}$ is its corresponding label, m is the number of support vectors, and $\sum_{i=1}^{n} \min(x_i, z_i)$ is the Histogram Intersection kernel function. The outer sum of (3.1) is over the m support vectors. However, for determining the contribution of visual words to the classification, we need the outer sum to be over the n visual words. This is possible because the Histogram Intersection kernel is an *additive* kernel Maji *et al.* [2008] where the inner sum

can be brought outside, leading to:

$$h(\mathbf{x}) \;=\; b + \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_j t_j \min(x_i, z_{ij}) \qquad (3.2)$$

$$\;=\; b + \sum_{i=1}^{n} w_i. \qquad (3.3)$$

The evidence per visual word channel is represented therefore by the weights w_i, which are equally distributed over the number of words of that type in the painting. By keeping the locations of each visual word, we can then backproject this evidence into the painting.

3.3 Emotion Recognition For Professional And Amateur Abstract paintings

In this section we describe how we apply our classification method on the two datasets of abstract paintings introduced in Section 2.5. On MART dataset we present a comparison of the classification results by using as ground truth the outputs of the Absolute and Relative Scale annotation methods. On the deviantArt dataset we apply our classification method using as ground truth the output of the Relative Scale annotation method. Moreover, we demonstrate how scene analysis and machine learning techniques can be used not only to differentiate between emotionally positive and negative abstract paintings, but also to identify emotional parts of the paintings.

3.3.1 MART Dataset: Classification Results of Absolute and Relative Scale

To construct the ground truth of the Absolute Scale annotation method, for each painting the average of all available scores is computed. The paintings with average scores lower or equal to 4 we define as negative, and paintings with average scores higher than 4 as positive. In total, 183 paintings were assigned to the negative class, and 317 to the positive class.

To construct the ground truth of the Relative Scale annotation method, we needed to define a threshold to separate positive and negative paintings. As each painting was compared 15 times, we assume that the threshold should be related to one of the paintings which was chosen 8 times as more positive. Therefore, we consider the TrueSkill ranking values of these paintings, which are the ratings resulted from the comparisons of paintings by using the TrueSkill algorithm. We sort these results and find the largest increase. The TrueSkill ranking value of the corresponding painting is considered as the threshold (i.e., the value equal to 25.1). The paintings with TrueSkill ranking equal or lower than the threshold are defined as negative and paintings with ranking higher than the threshold are defined as positive. In total, we obtain 131 paintings in the negative class and 369 in the positive class. Figure 2.3 displays the difference between (1) Absolute (left column) and (2) Relative (middle column) Scale annotations. In both annotation methods, subjects have tagged paintings with darker colours, rough texture and sharp lines as negative while most of the paintings with light colours and geometrical shapes were classified as positive.

Although the paintings have different ranks when the two annotation methods are used, Figure 2.3 shows that many paintings have similar ranking. From the 500 paintings, we observe that 128 were rated differently by people while evaluating emotionally the paintings with the Relative and Absolute Scale methods. We observe that most of the paintings with no agreement of positive and negative judgments when using the two annotation methods are composed mainly of colours considered as neutral such as beige, white, brown, grey. Most of these paintings are represented by single colours, without lines or shapes.

For both annotation approaches we tested our model using a two-fold cross-validation repeated 1,000 times. The average accuracy of both Absolute Scale and Relative Scale is reported in Table 3.1. In order to analyse the contribution of each type of visual words we separately evaluate the accuracy of the system based on LAB visual words and on SIFT visual words. Then we combine both systems by averaging their classification scores.

Overall, the proposed methods perform significantly better than chance level. We observe that in the case of Absolute Scale LAB visual words are more effective for emotional classification of abstract paintings compared to SIFT visual words, the accuracy is 0.747 versus 0.740 respectively. The same situation is observed for Relative

	LAB	SIFT	LAB+SIFT
Accuracy Absolute Scale	0.747	0.740	0.764
Accuracy Relative Scale	0.768	0.764	0.772

Table 3.1: Classification accuracy of LAB and SIFT Visual Words

Scale, in which the accuracy of LAB visual words is 0.768 while the accuracy of SIFT visual words is 0.764. The combination of LAB and SIFT visual words raises the accuracy to 0.764 for Absolute Scale and 0.772 for Relative Scale. This indicates that both colour-based LAB visual words and texture-based SIFT visual words are important for emotion recognition.

We observe that both Absolute and Relative Scale annotation methods can provide a reliable ground-truth for a positive/negative emotion recognition task in abstract paintings. Furthermore, it can be inferred that there is a significant difference in the accuracy of these two methods (the p value of a double t-test $< 10^{-29}$), as well as, there is an increase in the accuracy of the classification from 0.764 to 0.772.

3.3.2 Classification Using Global Configurations on MART Dataset

As an additional application, we applied Gist descriptors to investigate if an approach that analyse the structure of the abstract paintings as a whole would improve our models. The Gist descriptor, proposed by Oliva & Torralba [2001], efficiently describes the image scene based in a low dimensional representation. In Oliva & Torralba [2001], they represent the structure of a scene using a set of perceptual dimensions that are related to the *shape of the scene*: naturalness, openness, roughness, expansion, ruggedness. In their holistic approach, they apply oriented Gabor filters over different scales and average the filter energy in each bin. The Gist descriptor is very similar to SIFT descriptor, but instead of being employed at local patches it is applied at the entire image at once.

To extract Gist, we used a publicly available code[1], in which a vector of dimension 6144 is computed by averaging each filter output over a 32 x 32 grid. We tested this approach in two-fold cross-validation repeated 1,000 times. As a result, we have ob-

---

[1]Gist features were constructed using the code: http://people.csail.mit.edu/torralba/code/spatialenvelope/

34

tained 0.661 correct classification rate when the Gist descriptor was used as feature for Absolute Scale and 0.740 for Relative Scale, as shown in Table 3.2. The combination of Gist descriptor and LAB visual words decrease the accuracy to 0.742 for Absolute Scale and 0.762 for Relative Scale.

	LAB	GIST	LAB+GIST
Accuracy Absolute Scale	0.747	0.661	0.742
Accuracy Relative Scale	0.768	0.740	0.762

Table 3.2: Classification accuracy of LAB Visual Words and Gist Descriptor

As a result, we observe that the local representations of the abstract paintings, i.e., using SIFT Visual Words, obtain significantly better results on the emotion classification than using a global representation of the image, or Gist descriptor, on both Absolute and Relative Scales.

3.3.3 MART Dataset: Backprojection

In this Section we visualize how the classifier "sees" paintings while deciding whether they are positive or negative for both Absolute and Relative Scales annotations. We use the method described in Section 3.2.4 and backproject the relative contribution which comes from each pixel in the painting. In the Figures 3.1, 3.2, 3.3 and 3.4, we present some of the results by using the backprojection technique. In the backprojections, yellow colour corresponds to the positive emotional evidence and blue colour corresponds to the negative emotional evidence. The results corresponding to 4 paintings classified as highly positive for Absolute and Relative Scale methods are shown in Figures 3.1 and 3.2 respectively. Two paintings, the second and the third from the left of Absolute Scale (Figure 3.1) are also represented as highly positive for Relative Scale annotation (first and third painting from the left of Figure 3.2). The 4 paintings classified as highly negative for both methods are shown in Figures 3.3 and 3.4 respectively. The second painting (from the left) on Absolute Scale is also considered highly negative for Relative Scale (last figure). As it can be seen in these figures, two paintings are highly positive and one painting is highly negative for both methods on the classification results.

The contributions of LAB visual words are displayed in the second rows. The background colours typically yield neutral information, which is coherent with their purpose to be a background for more emotional parts of the painting. Dark colours such as black, dark blue and dark green have mostly negative responses, while red, yellow, and light blue often evoke positive emotions. Moreover, we observe that for the positive images there are more positive parts of backprojection on the Relative Scale than the Absolute Scale. The third rows in Figures 3.1 and 3.2, as well as in Figures 3.3 and 3.4 demonstrate the emotional evidence from SIFT visual words. Positive emotions often come from straight lines and smooth curves. In contrast, chaotic textured regions often evoke negative emotions even if their colour is considered positive. For negative paintings we observe that in the case of Relative Scale there are more negative parts of backprojection than on the Absolute Scale. The last rows show the combined evidence from LAB and SIFT visual words: evidence coming from colour and texture of the painting is effectively mixed together for both annotation approaches.

In a closer analysis on the results of both Absolute and Relative Scale we observe that in general most of the paintings that show agreement between the computational classification and humans' judgements come from Geometrical Abstraction Style. The Geometrical Abstraction Style, as introduced on Chapter 2 Section 2.1.1, is composed of geometric forms (i.e., line, square, triangle, etc.) and uniform colours. The paintings in which there is no agreement between humans' judgement and the computational classification belong mostly to the Lyrical Abstraction Style. The Lyrical Abstraction Style, also detailed on Section 2.1.1, influenced by expressionism, is composed by lines and colours inspired by the artist instinct which is detached from forms that humans are used to see.

In addition, the paintings exhibiting differences between the human and machine annotation results in the case of Absolute Scale have the mean scores near the threshold and the standard deviation is high. This means that many people scored these paintings as positive, however a high number of people considered them as negative as well. The high variance of human emotional evaluation may be influenced by the familiarity of forms and shapes. The study of Leder *et al.* [2014] who analyses the positive-negative emotional effects of art expertise, showed that familiar artworks were experienced as being more positive than unfamiliar artworks. Even the artworks with negative ratings considered as familiar, were experienced as less negative. Since most of the paint-

Figure 3.1: Visualisations of pixel-wise contributions to the classification of highly positive paintings using Absolute Scale of MART dataset. The original paintings are shown in the first row. The second row displays the contribution of LAB visual words, followed by the contribution of SIFT visual words in row 3, and their combination in row 4. Yellow colour corresponds to the positive emotional evidence and blue colour corresponds to the negative emotional evidence. ©*MART - Archivio fotografico e Mediateca.*
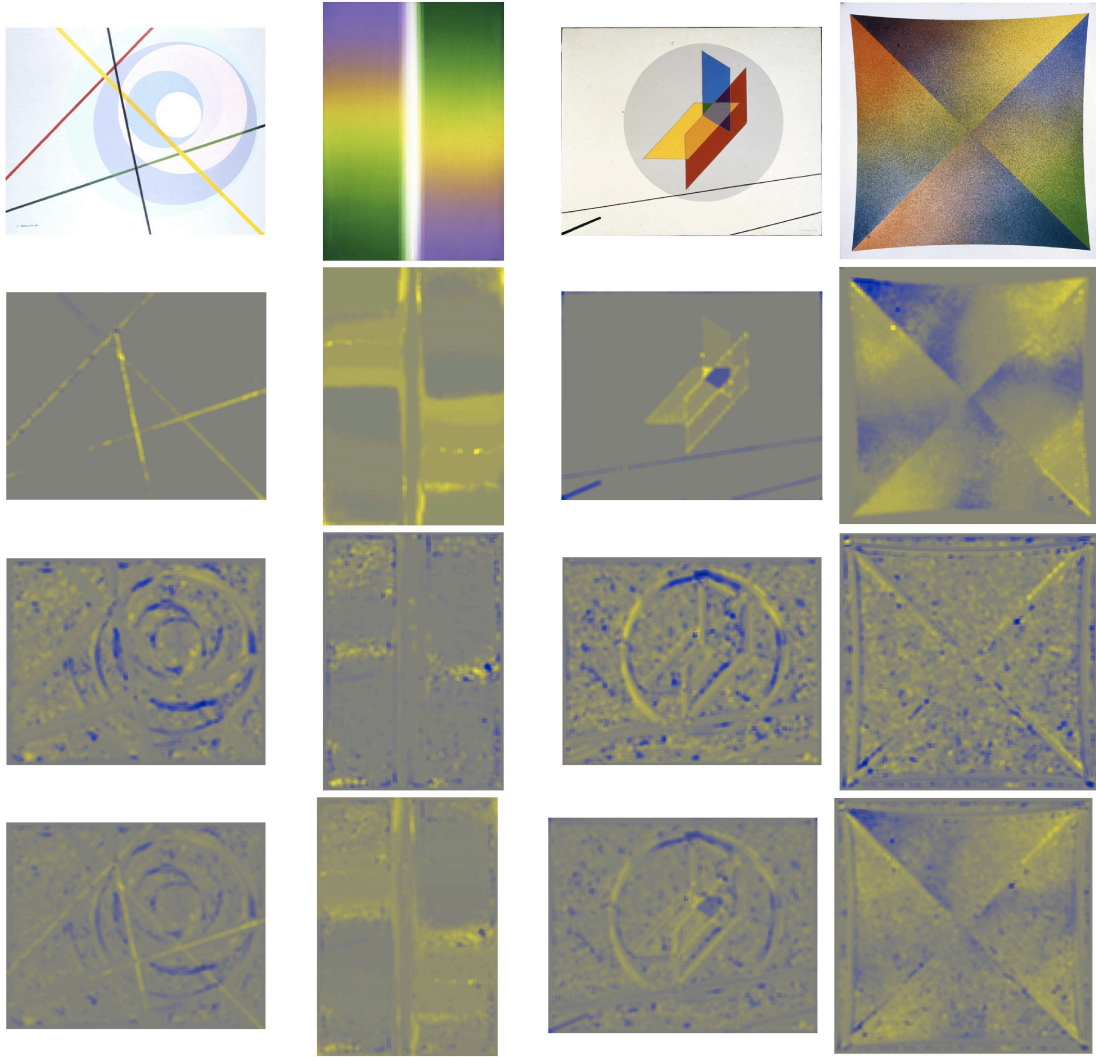
Figure 3.2: Visualisations of pixel-wise contributions to the classification of highly positive paintings using Relative Scale of MART dataset. The original paintings are shown in the first row. The second row displays the contribution of LAB visual words, followed by the contribution of SIFT visual words in row 3, and their combination in row 4. Yellow colour corresponds to the positive emotional evidence and blue colour corresponds to the negative emotional evidence. ©*MART - Archivio fotografico e Mediateca.*
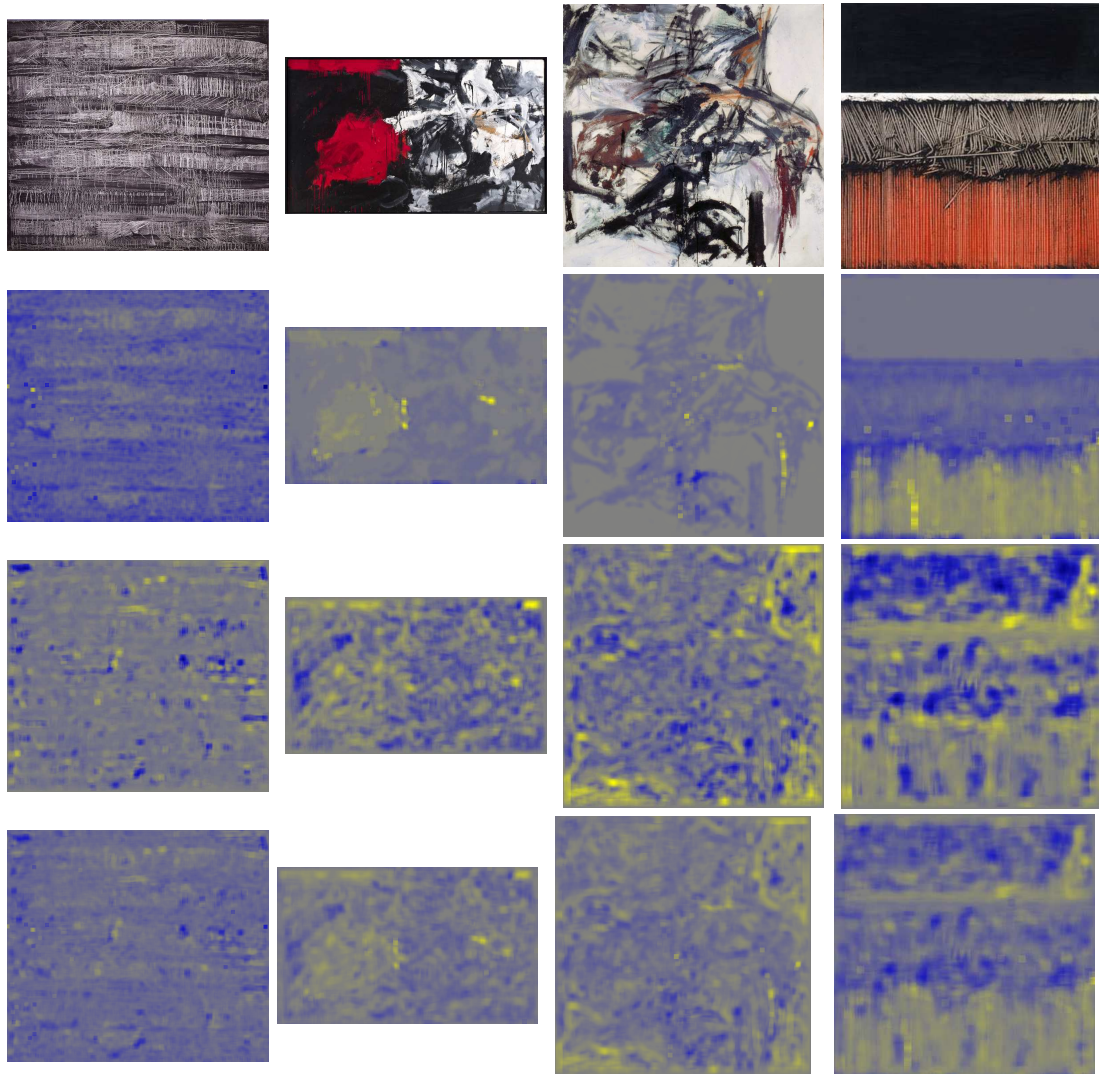
Figure 3.3: Visualisations of pixel-wise contributions to the classification of highly negative paintings using Absolute Scale of MART dataset. The original paintings are shown in the first row. The second row displays the contribution of LAB visual words, followed by the contribution of SIFT visual words in row 3, and their combination in row 4. Yellow colour corresponds to the positive emotional evidence and blue colour corresponds to the negative emotional evidence. ©*MART - Archivio fotografico e Mediateca.*
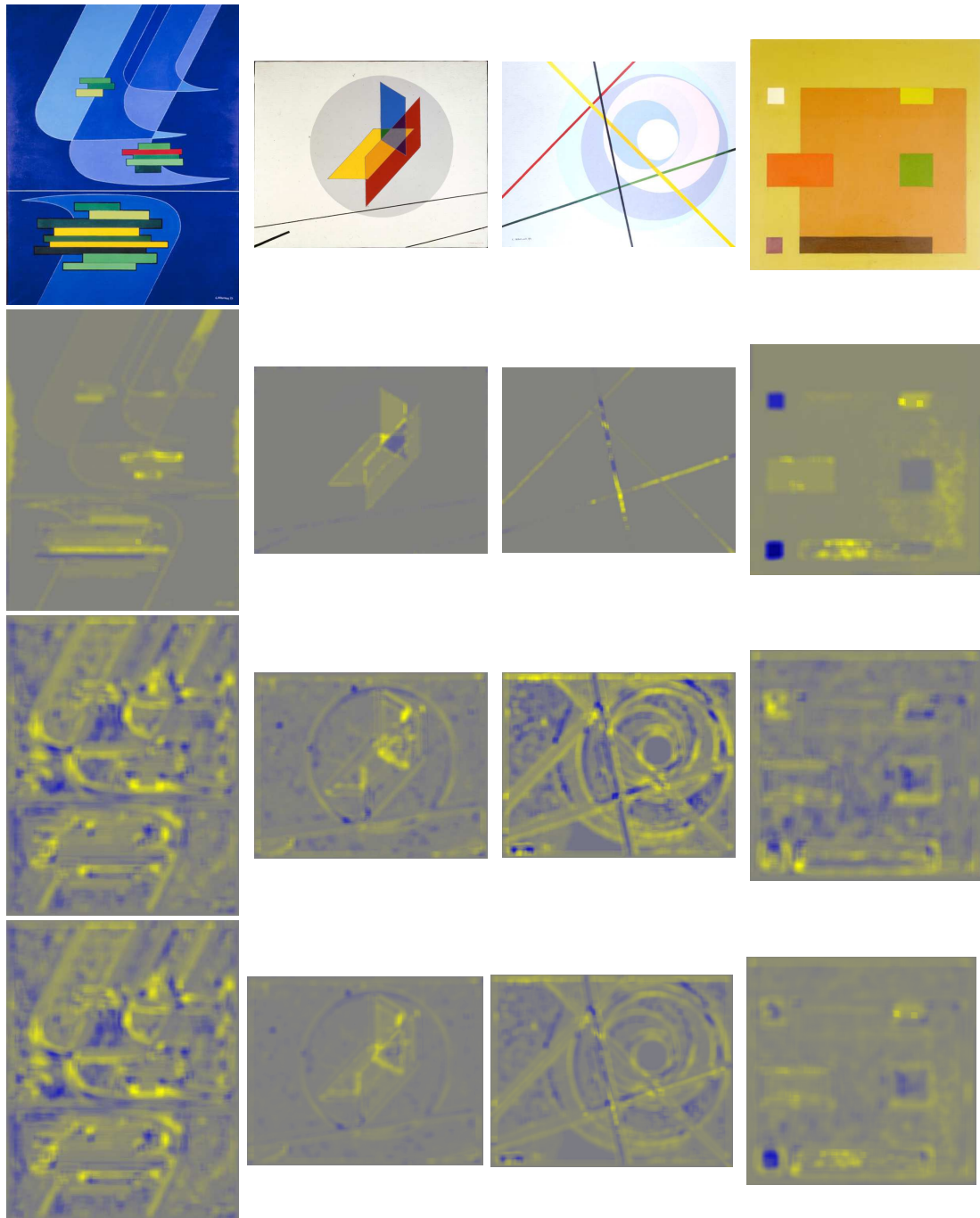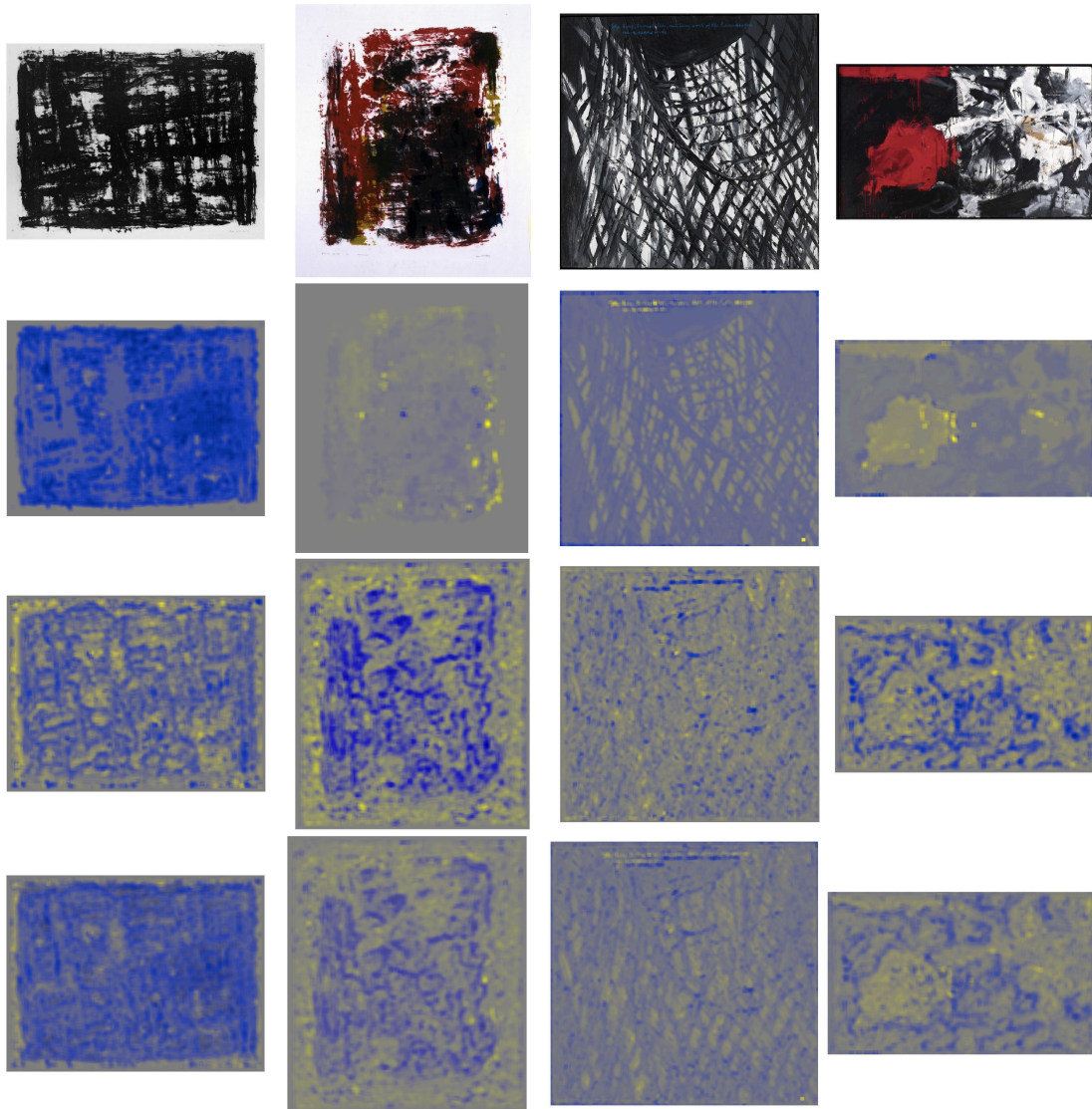
Figure 3.4: Visualisations of pixel-wise contributions to the classification of highly negative paintings using Relative Scale of MART dataset. The original paintings are shown in the first row. The second row displays the contribution of LAB visual words, followed by the contribution of SIFT visual words in row 3, and their combination in row 4. Yellow colour corresponds to the positive emotional evidence and blue colour corresponds to the negative emotional evidence. ©*MART - Archivio fotografico e Mediateca.*

ings with high variance are from the Lyrical Abstraction Style, the forms and shapes may not be familiar to some people and this may trigger a negative rating. With the backprojection visualizations we observe for instance that the paintings which people rated as positive and the machine as negative are composed by straight lines and smooth shapes. These features are usually interpreted by the machine as positive, but since these paintings are composed of dark colors (i.e. negative parts) the machine is likely to classify them as negative. Similar observations can be inferred in the case of paintings that have different annotation results made by humans and machines in the Relative Scale framework. Figure 3.5 provides one example of when the classifier has made the mistake abovementioned.
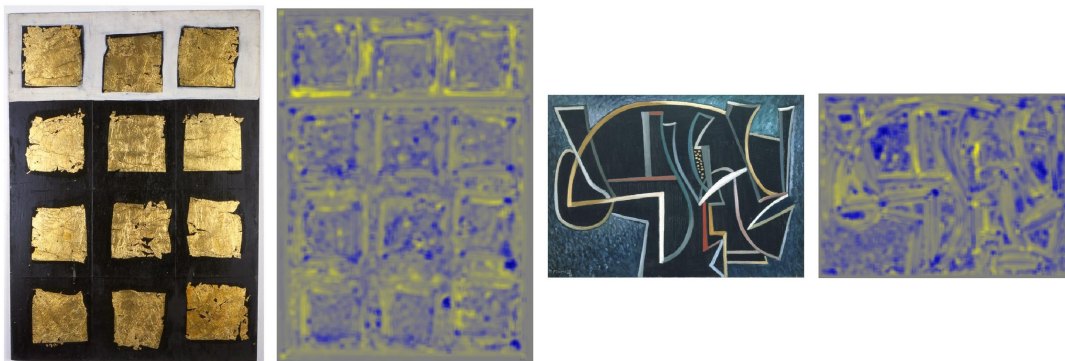
Figure 3.5: Example of when the classifier makes mistake: paintings rated by people as positive but the machine classify as negative. The second and fourth image from the left are the visualisations of pixel-wise contribution to the classification for the combination of LAB and SIFT visual words. Yellow colour corresponds to the positive emotional evidence and blue colour corresponds to the negative emotional evidence. ©*MART - Archivio fotografico e Mediateca.*

3.3.4 deviantArt Dataset: Classification Results of Relative Scale

To construct the ground truth of deviantArt dataset, we followed the analysis of Relative Scale explained in Section 3.3.1 to define the threshold. We analysed the ranking values of deviantArt paintings and selected a painting whose ranking can separate the dataset in positive paintings and negative paintings. We considered the paintings with ranking equal or lower than 21.0 as negative and paintings with ranking higher than 21.0 as positive. As a result, 140 paintings were assigned to the negative class and the

360 to the positive class.

To verify if our classification method is effective in classifying the deviantArt dataset paintings in positive and negative emotions, we apply the same classification method described in Section 3.3.1. We tested our model using a two-fold cross-validation repeated 1,000 times. We evaluate separately the accuracy of both LAB and SIFT visual words, to analyse the contribution of each type of visual word. Then, we combined the classification scores of both LAB and SIFT visual words and average them, as reported in Table 3.3.

	LAB	SIFT	LAB+SIFT
Accuracy	0.769	0.749	0.770

Table 3.3: Classification accuracy of LAB and SIFT Visual Words.

We observe that also in the case of the deviantArt dataset, LAB visual words are more effective than SIFT visual words on emotional classification. The accuracy is 0.769 versus 0.749 respectively. The combination of LAB and SIFT visual words raises the accuracy to 0.770. This result indicates that the combination of colors and textures can provide better accuracy than in the case of using each of them independently.

3.3.5 Classification Using Global Configurations on deviantArt Dataset

We also analysed the structure of the deviantArt paintings as a whole by applying Gist descriptors presented in Section 3.3.2. The accuracy when the Gist descriptor was used as feature is 0.718, as shown in Table 3.4. The combination of Gist descriptor and LAB visual words decrease the accuracy to 0.741.

	LAB	GIST	LAB+GIST
Accuracy	0.769	0.718	0.741

Table 3.4: Classification accuracy of LAB Visual Words and Gist Descriptor.

Also in the case of the deviantArt dataset, using SIFT Visual Words obtain significantly better results on the emotion classification than using a global representation of the image, i.e., Gist descriptor.

3.3.6 deviantArt Dataset: Backprojection

Also for deviantArt dataset we used the backprojection technique described in Section 3.2.4, to understand which is the most relevant positive and negative information in the classification task. We backproject the relative contribution coming from each pixel in the painting. The results for the 4 paintings classified as highly positive and the 4 paintings classified as highly negative are shown in Figures 3.6 and 3.7 respectively. The second row displays the contributions of LAB visual words. Also in this case the dark colours, such as black and dark blue yield mostly negative response, and light colours, such as green, orange, yellow and red, evoke positive emotions. The emotional evidence of SIFT visual words can be seen in the third row of Figures 3.6 and 3.7. In the backprojection visualizations we observe that chaotic textured regions often correspond to negative emotional evidence, while straight lines, defined contours and smooth curves mostly correspond to positive emotional evidence, even in paintings which are considered negative. The last row shows the combined evidence from LAB and SIFT visual words, in which the evidence of colour and texture is mixed together.

In a comparison of both datasets of abstract paintings, we observed that MART paintings are mostly composed of neat colours and geometrical shapes while deviantArt paintings mostly present blended colours and undefined shapes. Even with these pattern differences of colours and shapes present in MART and deviantArt compositions, our classification method was able to identify similar positive and negative patterns for both set of abstract paintings. By analyzing these results we can conclude that both LAB and SIFT visual words are relevant in recognizing emotions in both professional and amateur abstract paintings, giving similar detected positive and negative parts.

3.4 Eye Movements and Emotional Content

The content of an image is one of the aspects determining how people observe it. For this reason, we believe that the emotional content present in abstract paintings also influences the behaviour of the viewer. We hypothesize that, in general, people prefer to focus on the emotionally positive parts over the emotionally negative ones. To test this hypothesis, we recorded eye-gaze movements and correlate them with the positive and negative parts of the paintings as determined by our backprojection.
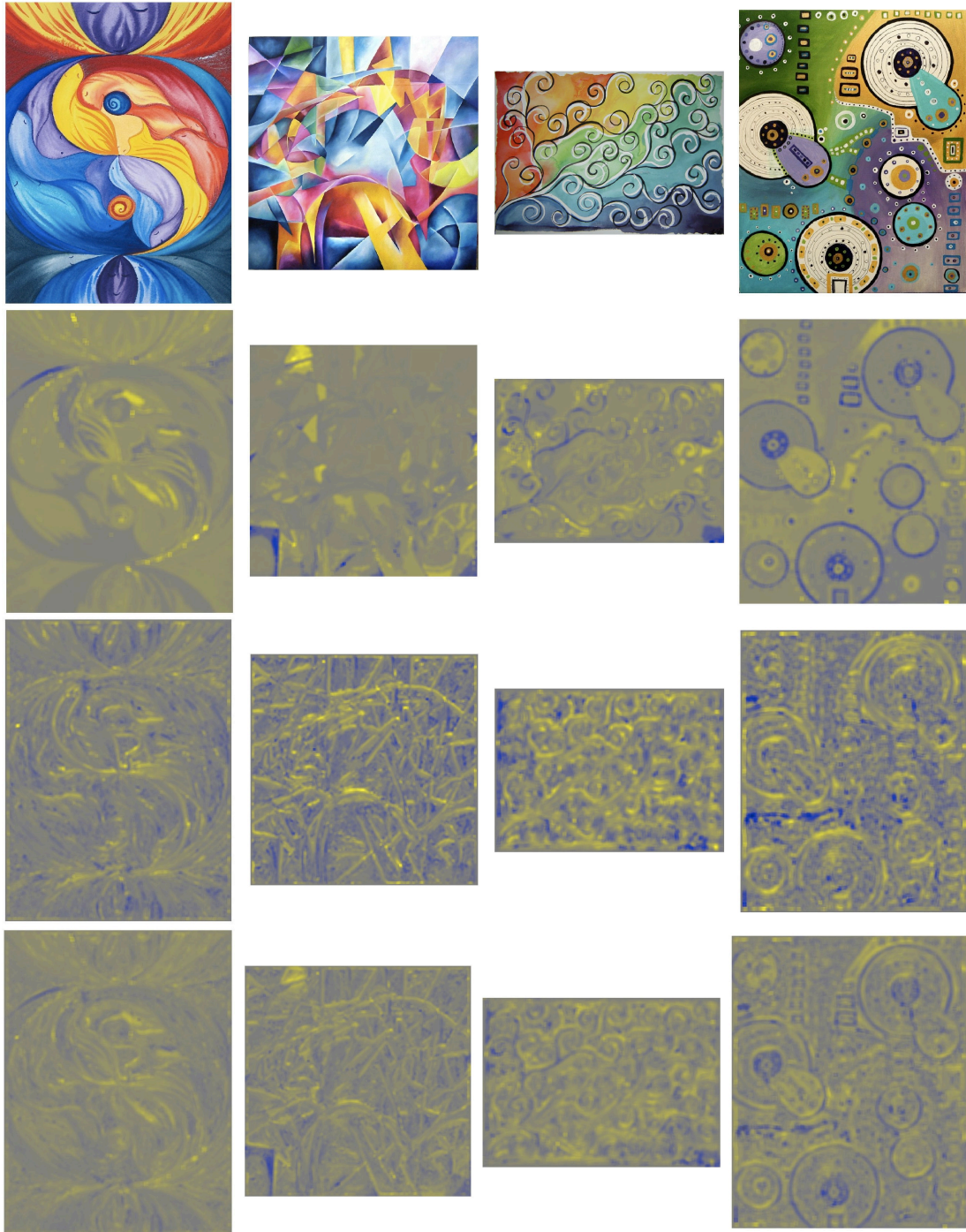
Figure 3.6: Visualisations of pixel-wise contributions to the classification of highly positive paintings of deviantArt dataset. The original paintings are shown in the first row. The second row displays the contribution of LAB visual words, followed by the contribution of SIFT visual words in row 3, and their combination in row 4. Yellow colour corresponds to the positive emotional evidence and blue colour corresponds to the negative emotional evidence. (*Courtesy of deviantArt.* Artists respectively: http://{oceanpeace, coconut-artist, klbailey, numb-synapse}.deviantart.com.)
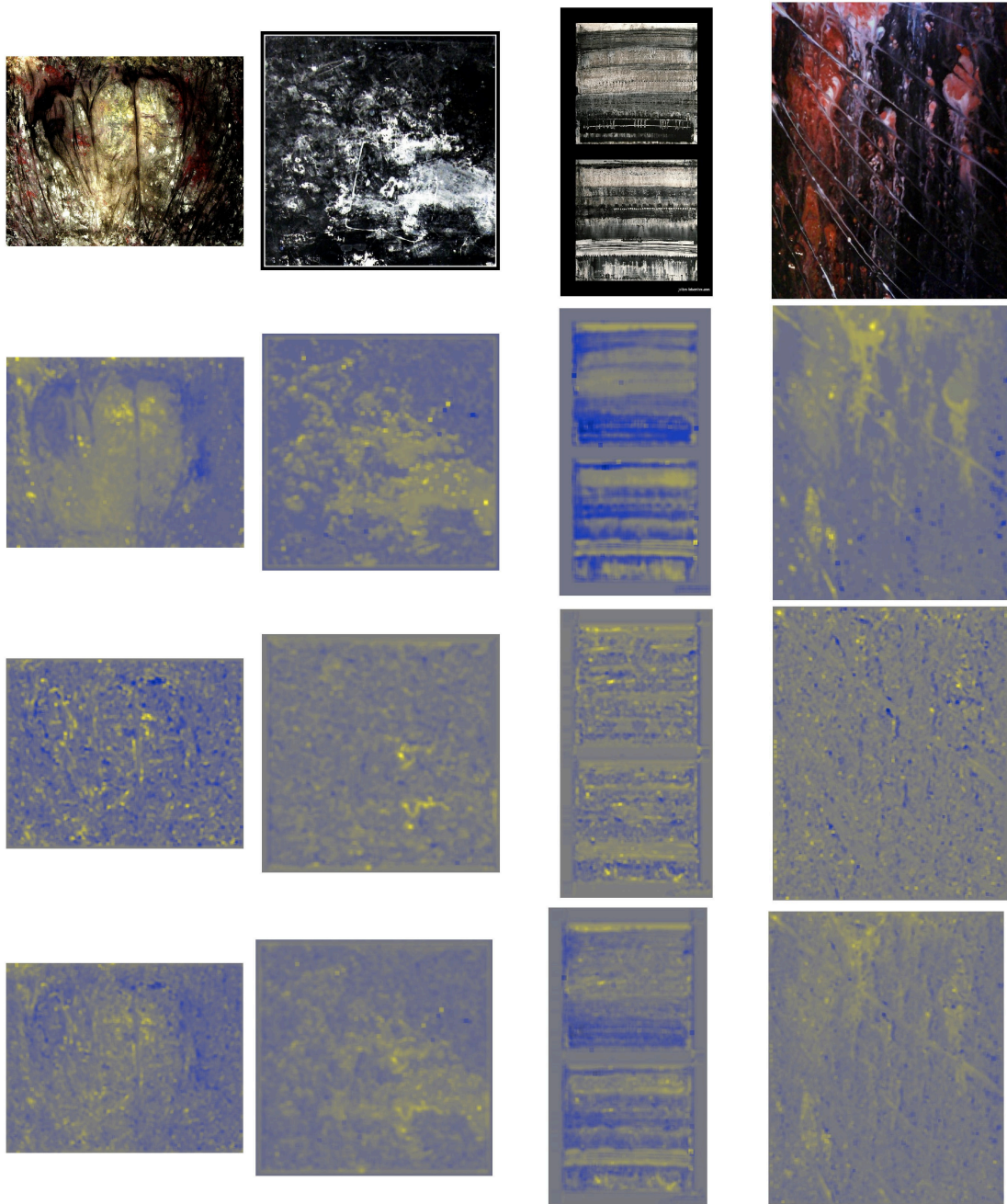
Figure 3.7: Visualisations of pixel-wise contributions to the classification of highly negative paintings of deviantArt dataset. The original paintings are shown in the first row. The second row displays the contribution of LAB visual words, followed by the contribution of SIFT visual words in row 3, and their combination in row 4. Yellow colour corresponds to the positive emotional evidence and blue colour corresponds to the negative emotional evidence. (*Courtesy of deviantArt.* Artists respectively: http://{diceglia, epaf, volute, satiricmilk}.deviantart.com.)

In order to ensure variability of emotional content, in this experiment we selected paintings with positive, negative, and neutral parts at the same time. Specifically, for each painting we required that at least 25% of its pixels yield positive contribution to the classification. Similarly, at least 25% of pixels of the same painting are required to yield negative contribution to the classification. In this way 110 paintings were chosen, in which 71 paintings were scored by subjects as positive and 39 as negative in the case of Absolute Scale, and 87 paintings were scored as positive and 23 as negative in the case of Relative Scale.

3.4.1 Eye Movement Recording

Eye movements were recorded using an EyeLink 1000 Tower Mount that samples pupil positions at 1000 Hz. Subjects were seated in a darkened room at 85 cm from a computer monitor and used a chin-rest so that their head position was stable. To calibrate the eye positions and to validate the calibration, subjects were asked to focus on 12 fixation spots on the screen, which appeared one by one in random order. The mean spatial accuracy of the eye tracker calibration was $0.36°$, with a standard deviation of $0.07°$, where one visual angle $1°$ equals to 30 pixels in our experimental settings.

During the experiment, stimuli were presented on a full size 23 inch screen (ASUS VG236H, 1920x1080 resolution) for 7 seconds. After each presentation of the stimulus, a grey background was shown for 0.5 seconds to prevent after-image effects. The order in which stimuli were displayed was randomized for every observer. Fixation locations and durations were calculated online by the eye tracker. The MATLAB psychophysics toolbox was used for stimulus presentation Brainard [1997]. The fixation locations and durations were calculated by the eye tracker software EyeLink Data Viewer.

We recorded eye movements of 9 observers while they were looking at the artworks from the MART dataset, 5 males and 4 females with ages between 24 to 31 years old ($M = 28.56$). Each participant had normal or corrected to normal vision. No participant had formal art training and all were naive to the purpose of the experiment. Their instructions were to freely look at the paintings like the way they would do it in a museum.

3.4.2 Data Analysis

According to Locher *et al.* [2007], viewers of both representational and abstract art use the first two seconds to do a sweep of the image, analysing its "gist". Only after this first explorative stage, viewers tend to focus on finer details. It has been shown that during this two second period, bottom-up saliency plays an important role in the allocation of eye movements Bruce & Tsotsos [2009]; Tatler *et al.* [2005]; Yanulevskaya *et al.* [2011]; Zhao & Koch [2011]. We are interested in the influence of high-level information such as emotional content to the gaze patterns. Therefore, in our analysis we consider eye movements which occur between the third and seventh seconds.

To quantify the relationship between emotional content detected by the classifier and human's gaze patterns, we compared the emotional evidence that comes from fixated and non-fixated locations. In particular, we averaged pixel-wise contributions to the classifier from the region around the fixated locations and compared them with the averaged pixel-wise contributions from the region around non-fixated locations. The regions around the fixations are taken to be fovea sized ($1°$, i.e. 30x30 pixels), as this is the area sampled in high resolution by the human eye. Now if people prefer to focus on those parts of the painting containing positive information, then emotional statistics of fixated regions should be higher than the ones of non-fixated regions.

To determine the non-fixated locations for a painting, we followed Reinagel & Zador [1999]; Tatler *et al.* [2005] and sampled from the fixated locations recorded while viewing other paintings. In this process, we required the same amount of fixated and non-fixated locations per painting, where non-fixated locations should be at least $1°$ (30 pixels) apart from the fixated locations. This guarantees similar distributions of fixated and non-fixated regions Tatler *et al.* [2005].

We define Positive Attentional Bias (**PAB**) the situation when the average contribution to the classifier from the fixated regions is *higher* than the average contribution to the classifier from the non-fixated regions. We define Negative Attentional Bias (**NAB**) to be the opposite condition. When emotional statistics of fixated and non-fixated regions do not differ significantly according to the 2 sample t-test, the attentional bias is neutral to positive and negative emotions. To test if people prefer to focus on emotionally positive visual information, we calculate how many abstract paintings evoke PAB. To backproject emotional evidence, we use our best performing classifier according

	LAB+SIFT	#	PAB	NAB	Neutral
Absolute Scale	All paintings	110 (100%)	82 (74%)	26 (24%)	2 (2%)
	Pos paintings	71 (100%)	59 (83%)	11 (16%)	1 (1%)
	Neg paintings	39 (100%)	23 (59%)	15 (38%)	1 (3%)
	LAB+SIFT	#	PAB	NAB	Neutral
Relative Scale	All paintings	110 (100%)	77 (70%)	31 (28%)	2 (2%)
	Pos paintings	87 (100%)	63 (73%)	22 (25%)	2 (2%)
	Neg paintings	23 (100%)	14 (61%)	9 (39%)	0

Table 3.5: Distribution of paintings according to "emotions" of the attentional bias for Absolute Scale and Relative Scale ground-truth of MART dataset.

to Table 3.1 which combines LAB and SIFT descriptors. Table 3.5 contains averaged results over 50 independent samplings of non-fixated locations for Absolute Scale and Relative Scale annotations. From our results we observe that general tendencies hold for both types of ground truth. Only 2 paintings have an attentional bias for emotionally neutral regions, whereas 70% of considered paintings shows positive attentional bias for Relative Scale annotation. As can be expected, PAB holds for most of the emotionally positive paintings. Moreover, even while observing negative paintings, in 61% of cases people still prefer to look at positive parts.

By analysing the fixation duration, we also investigate if people prefer to look longer at the positive parts of the painting. We compared the emotional evidence that comes from the fixated locations and the fixation durations of those locations. The region with fixation duration higher than the total average of fixation duration (323 ms) is considered as long fixation region, and the fixation duration lower than the overall average of fixation duration as short fixation region. We consider that people prefer to look longer at the positive parts when the average contribution to the classifier from the long fixation region is higher than the average contribution to the classifier from the short fixation region. In total, for 71 paintings (65%) people looked longer at positive parts. From these 71 paintings, 49 paintings were evaluated by people as positive and 22 as negative for the Absolute Scale annotation, and 62 paintings were evaluated as positive and 9 as negative for the Relative Scale annotation. Figure 3.8 illustrates some examples of paintings together with the recorded fixated locations.
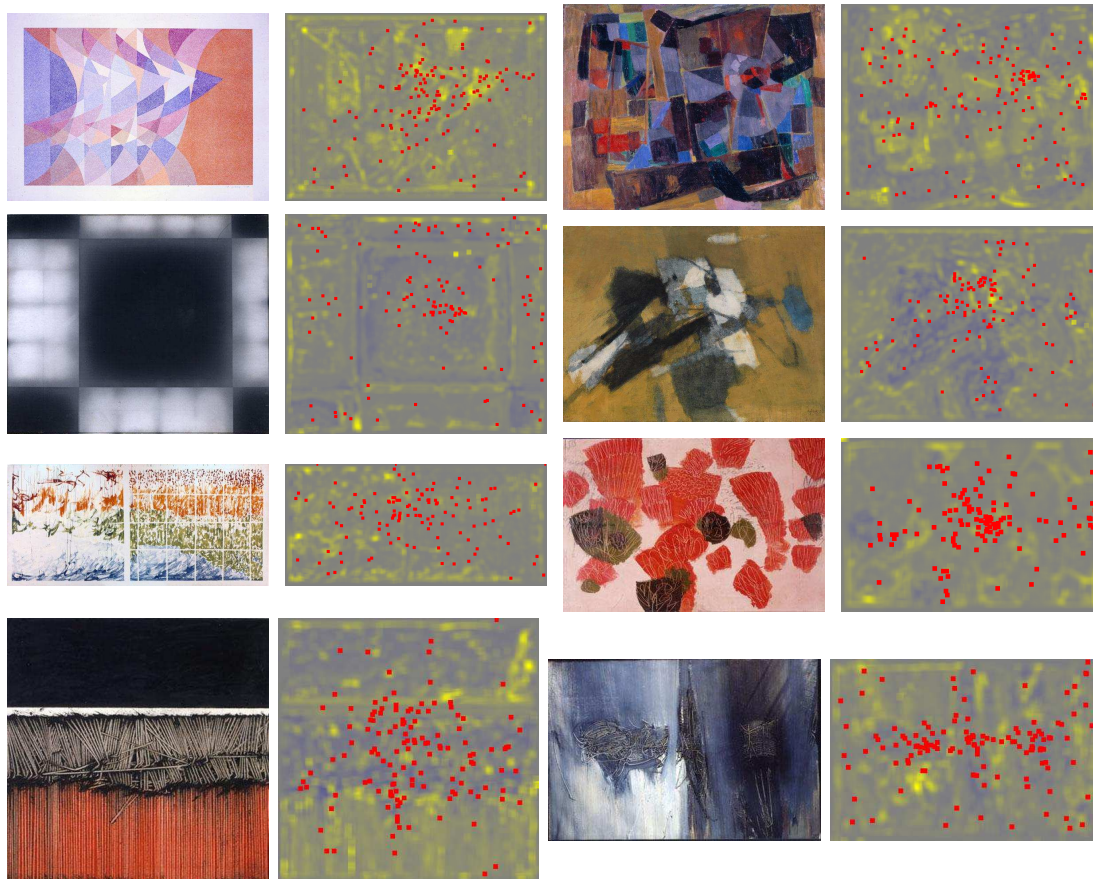
Figure 3.8: Paintings with superimposed fixated locations: the first row displays positive paintings with PAB, the second row displays negative paintings with PAB, the third row displays positive paintings with NAB, and the last row displays negative paintings with NAB. ©*MART - Archivio fotografico e Mediateca.*

3.5 Comparison between Amateur and Professional Abstract Paintings Classification Results

In this Section we compare the emotional evaluation between the two different datasets of abstract paintings: the paintings from professional artists (MART Museum), and the paintings from amateur artists (deviantArt). From our classification results we observe that the proposed method is effective in the classification of positive and negative emotions for both professional and amateur's abstract paintings. By applying the Relative Scale method to construct the ground-truth for both datasets we achieve an overall ac-

curacy of 0.77, which is significantly better than the chance level. Interestingly, the majority of the paintings from the MART dataset authored by notable artists such as Wassily Kandinsky, Josef Albers, Paul Klee, Aldo Schmid, Luigi Veronesi, Carlo Belli and Lucio Fontana have been evaluated similarly by the humans and the machine. These artworks are a result of their study on colours, lines, shapes and textures. Particularly, Wassily Kandinsky, Josef Albers and Paul Klee were the most influential artists of abstract art, having published detailed studies about colours and shapes.

We observe that LAB visual words are more effective in the emotional classification than SIFT visual words. Indeed, as shown in Figure 2.3, both datasets are mostly ordered by people in a gradient order of colours from the most negative to the most positive, such that the most negative paintings are those having dark colours and the most positive paintings are the ones having light colours. However, the combination of LAB visual words and SIFT visual words increases the accuracy of the classification for both datasets, which demonstrates that both colour-based and texture-based visual words are relevant to recognizing emotions in abstract paintings. By analysing the results of backprojection, we observe that in the deviantArt dataset the paintings classified as positive contain defined lines and high colour contrasts. In the MART dataset, geometrical shapes and uniform colours are found in most of the positive paintings. The negative paintings of MART and deviantArt dataset are composed by undefined lines, rough texture and a mixture of dark colours. Such characteristics may not be familiar to the beholder, and thus may provoke the negative emotions.

3.6 Discussions

We propose an effective approach for understanding which part of an abstract painting evokes positive or negative emotion through advanced computer vision techniques. We apply our approach in two collections of abstract paintings: one of professional set from MART Museum and one amateur set from deviantArt, a user-generated art website. We applied the ground truths (Absolute and Relative Scale) presented in the previous chapter to our classification model, in which we obtained higher accuracy when the ground truth was created using a relative ranking.

For classification of both datasets we trained a Bag-of-Visual-Words system and then backprojected the classification evidence to visualize which parts of the paint-

ings conveyed positive and negative emotions. The results show that our classification method is effective in classifying both Professional and Amateur abstract paintings in positive and negative emotions. By using backprojections visualizations we observe that even though the two datasets do not present similar characteristics, the machine was able to find similar characteristics that are crucial for the emotional classification process. We qualitatively analysed that the results follow long-known observations in art: bright colours evoke positive emotions; dark colours tend to evoke negative emotions. Smooth lines are generally positive. Chaotic texture is generally negative. Moreover, we correlate the localized emotion evidence with eye movements, in order to identify the parts people prefer to focus on while observing abstract paintings. Our results demonstrate that there is a positive attentional bias, even in paintings with a negative emotional content, i.e., people prefer to look at the positive parts of these paintings. In addition, we observe that in 65% of the paintings people look longer at the positive parts. The ability of localising the emotional evidence in abstract paintings opens up a new toolbox for art historians, art critics and artists to revisit art theories with a new perspective.

With this analysis we observed the importance to look into each feature (colour and textures) and make an in-deep study of how these features affect emotionally people on abstract painting. Indeed, on Chapter 4 and 5 we investigate which are the specific aspects of each feature, in order to enhance the understanding of why abstract paintings conveys emotion.

Chapter 4

Contribution of Texture[1]

This chapter presents how we assess the contribution of texture to automatically predict emotional valence of abstract paintings. In this study we establish the ground truth with a user study of Relative Scale method (detailed in Section 2.5.2) that was conducted over the gray-scale version of MART dataset. In addition, we use a study from the psychophysics literature that looked into the contribution of texture in the perception of different emotions about an image, including masculinity-femininity, heaviness-lightness, and hardness-softness perception. In Lucassen *et al.* [2011], synthetic and parameterized textures created with Perlin noise were used to conduct a user study. The authors created a model that maps the Perlin parameters of the image to the feeling it induces. In this study, we derive these parameters for MART artworks by an inverse-Perlin parametrization, and use the psychophysical model proposed in Lucassen *et al.* [2011] to characterize texture. To obtain this parametrization, we derive MR8 texture feature descriptors Varma & Zisserman [2005], and train a classifier that maps these features to the respective Perlin parameters.

We postulate that the contribution of colour to the feelings created by the artwork is more prominent than texture, but a combination of colour and texture may improve our models. We illustrate that the coloured and non-coloured versions of artworks have low, but visible correlation in the feelings they invoke in their viewers.

Moreover, in this study we look at image saliency to analyse qualitatively what features people associate with positive and negative emotions. We replicate the eye

---

[1]This study was submitted as Sartori *et al.* [2015a]

gaze study conducted with MART artworks reported in Section 3.4, but this time with non-coloured images. Then, we contrast the eye gaze study of both coloured and non-coloured versions of MART dataset with the predictions of a computational bottom-up saliency model. The high predictive power of the computational model establishes that the bottom-up estimation is reliable, and we are justified in our usage of abstract art for reducing the top-down effects. Furthermore, it suggests that by disabling the colour in the computational model, we can predict the eye gaze patterns of viewers in non-coloured images, and perform an analysis based on this.

To summarize, the main contributions presented in this study are: (1) we study the role of texture on positive and negative emotions induced by abstract paintings; (2) we propose an inverse-Perlin parametrization to enable a psychophysical model on emotions induced by the images; and (3) we apply a state-of-the-art of bottom-up saliency model to abstract paintings and we present an analysis of the relation between its predictions and the human fixations.

4.1   Related Work

The texture is an important element on the emotion expression of images and it is widely used in art, design and photography. On his book *"Point and Line to Plane"*, Kandinsky describes texture as the way the elements are combined with each other and depends of three factors Kandinsky [1979]:

1. the character of the given space (e.g., smooth, rough, flat, plastic, etc.),

2. the type of tool (e.g., the brush of various types, or other tools), and

3. the manner of application, (e.g., the colour may be laid on loosely, compactly, by stippling, by spraying, etc., depending upon its consistency).

The visual texture is an intuitive part of the human sensory input and it can influence the human perception and emotions. Indeed, texture has been subject of study in many areas including psychophysics and computer science. Simmons & Russell [2008] investigated the emotional effect of ten different visual textures on colours. They reported that, depending on a certain texture class, add texture to colours change significantly the humans raking of unpleasant colours. Kim *et al.* [2006] extracted

colour, texture and pattern features from textile images to predict humans emotions. The authors show that, their proposed textile indexing system was effective for predicting human emotions based on textile images. Thumfart *et al.* [2011] used a layered prediction model to predict the human aesthetic judgments given a computational texture feature. These and others studies (e.g., Leung & Malik [2001]; Tamura *et al.* [1978]) aim to simulate the human perception on texture in order to improve the computational representations for classification and segmentation.

Several recent works in computer vision that focuses on emotion recognition, including Jia *et al.* [2012]; Machajdik & Hanbury [2010]; Yanulevskaya *et al.* [2008]; Zhao *et al.* [2014] has used texture as one of the features to recognize emotions in images and artworks. However, these and the other studies mostly rely on colour in predicting affective content. There are few studies that investigates the contribute of texture independent of colour for this problem. In this study we investigate the role of texture, regardless to the colours, on the positive and negative emotions on abstract paintings.

4.2 Texture Analysis

Texture can give the beholder a 'visual sense' of the artwork composition, which is almost multimodal, in that it can evoke tactile associations. Some artists, such as Vincent van Gogh, Lucio Fontana, Anish Kapoor, use texture associations with great effect. In abstract art, feelings of hardness, softness, smoothness, etc. can be created via texture, and we postulate that texture is an integral element of the feeling induced by the artwork (regardless of the artist intentions). In the Figure 4.1 shows some examples of paintings from MART dataset in which the brushstrokes technique used by the artist creates a textured pattern.

In this section we present the analysis on how texture affect the emotions elicited by abstract paintings. We have conducted the Relative Scale annotation method to collect ground truth of positive and negative emotions evoked by abstract paintings with no hue value. In addition, in this section we describe our method for obtaining inverse Perlin parameters from the images, and we provide a qualitative analysis of the ensuing emotion classification.

Figure 4.1: Example of paintings in which the brushstrokes technique used by the artist depicts a textured pattern. Paintings from MART dataset composed by (from the left to right): Senesi Luigi, "Circolarità rotatoria", 1978; Lucio Fontana, "Concetto spaziale Oro", (1960-1961); Aldo Schmid, "Struttura colore n. 2", 1972; Lucio Fontana, "Concetto spaziale", 1954. ©*MART - Archivio fotografico e Mediateca.*

4.2.1 User Study for Assessing Emotions Induced by Paintings

The method to annotate the paintings in positive and negative emotions and the scoring procedure follows the Relative Scale annotation approach presented in Section 2.5.2, which is based on the TrueSkill ranking system of Herbrich & Graepel [2006] and the implementation of Moser [2010]. The annotation was made online and was asked to choose the more positive painting out of two. The MART dataset, described on Section 2.3.1, is used for the user study. For studying texture independent from the colour, we have prepared a gray-scale version of MART dataset.[1]

On the annotation of gray-scale version of MART paintings participated in total 55 subjects, 22 females and 33 males, respectively. Each subject annotated from 5 to 334 pairs of paintings, 67 pairs of paintings on average. The subjects participated voluntarily and they were free to annotate any time they wanted. To analyse the annotation results, we have matched the TrueSkill ratings obtained at the end of the user study to the individual annotations of the pairs, presented to all the annotators. The results show about 79% agreement, which is effectively the mean human performance for the valence classification task, and represents an upper bound for the automatic algorithm.

The distribution of gray-scale paintings from the most negative to the most positive

---

[1]This annotation is publicly available: http://disi.unitn.it/~sartori/datasets/mart-dataset/

annotation scores is illustrated in Fig. 4.2 (a). The distribution is a matrix of 50×10, where the paintings are sorted in reading order from the most negative to most positive. From the annotation results of the gray-scale dataset, we observe that lightness is a determinant factor of positive and negative emotions. Paintings with low value of overall intensity (nearly black) are considered as more negative. Johannes Itten postulates that neutral gray is a characterless, mute, indifferent, achromatic colour, and the number of distinguishable shades of gray depends on the sensitivity of the eye Itten [1974]. This may be the reason of the mixed distribution. Another factor effecting the ranking from negative to positive seems to be in the structural qualities of the paintings, i.e. some very dark paintings with light spots are considered positive, probably due to their composition and the way they use light colors and/or whites. We also note that the ordering of the paintings is quite different compared to the ordering induced by coloured images. Compared to the distribution of coloured images, one can say that greyscale images' ranking are more dependent to the composition of the paintings, whereas in coloured images colour itself plays a much more important role in determining the overall positive/negativeness of the painting.

4.2.2 Inverse Perlin Parametrization

A psychophysical experiment assessing the effect of texture on color emotion was conducted by Lucassen *et al.* [2011]. They presented the subjects with textured image patches in varying colours and asked the subjects to place those images on four emotion scales: warmness-coolness, masculinity-femininity, hardness-softness and heaviness-lightness. The texture samples were created with a parametric model, called Perlin noise. As a result of their experiment, the authors performed a regression analysis and established parametric relations between the Perlin images and the subject classifications. We give the ensuing three formulae for the mentioned emotion scales (warmness-coolness, which depends on colour, is not computed for texture) in Table 4.1. The computation of these scales for a given image can provide us with a parsimonious and informative representation of the texture features.

Perlin images are textures created using Perlin noise, which is a method to create a parameterized, pseudo-randomized noise by taking the sum of different noise sig-

(a) Emotion scores from negative to positive

(b) Emotion scores from masculine to feminine

(c) Emotion scores from hard to soft

(d) Emotion scores from heavy to light
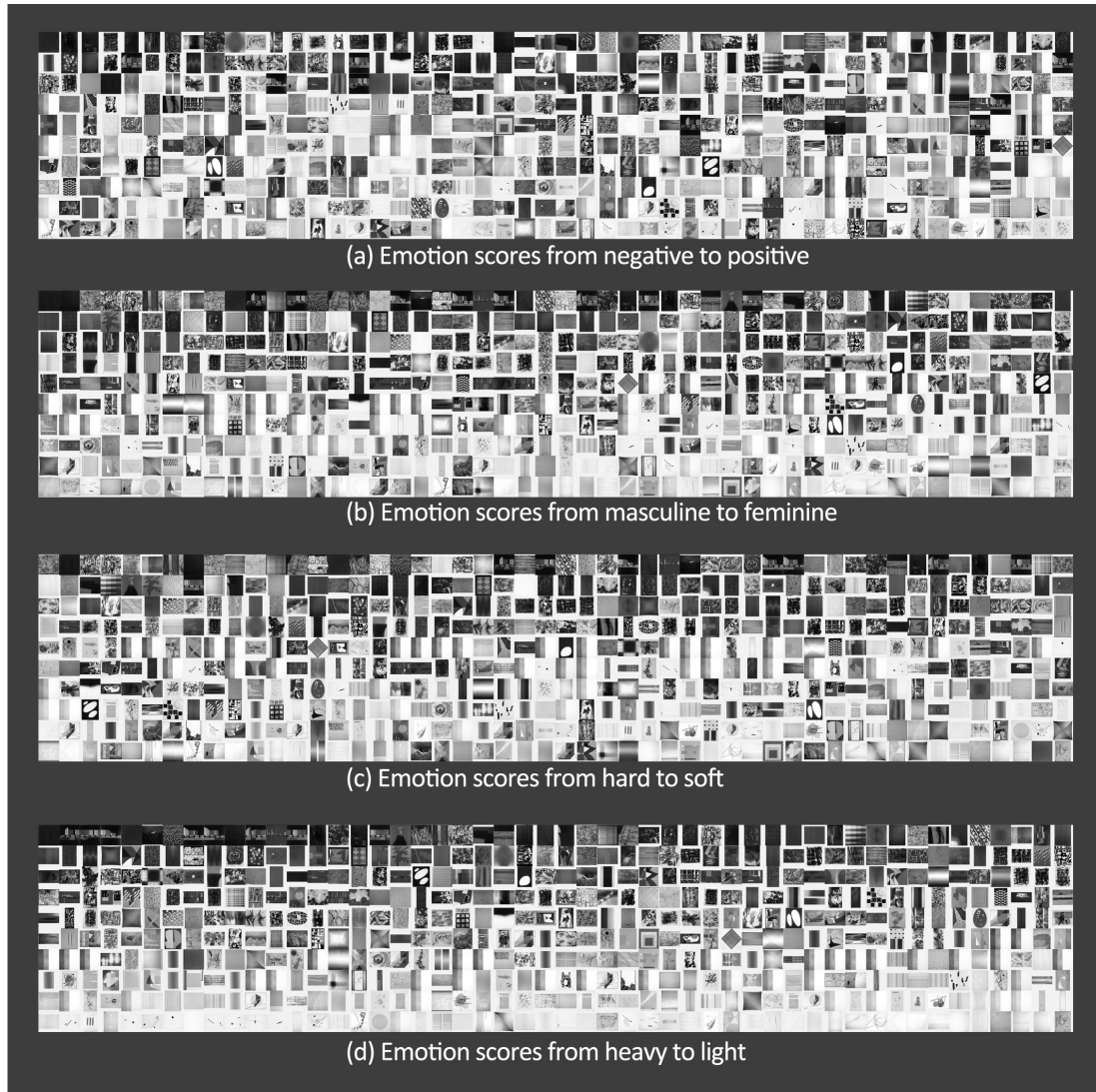
Figure 4.2: The entire MART dataset depicted as thumbnails. The paintings in (a) are ordered by TrueSkill scores from the most negative to the most positive. The next three images are ordered by computing Perlin parameters and using the psychophysical model: (b) from masculine to feminine, (c) from hard to soft and (d) from heavy to light. ©*MART - Archivio fotografico e Mediateca.*

nals Perlin [1985]. We used the same open-source libnoise library[1] used by Lucassen *et al.* [2011] to generate those images. They have four parameters: octave, frequency,

---

[1] http://libnoise.sourceforge.net/

Table 4.1: Texture emotion scales based on Perlin parameters (from Lucassen *et al.* [2011]). Emotion scales are MF: masculinity-femininity, HS: hardness-softness, HL: heaviness-lightness, given as a function of mean intensity (L) and parameters of Perlin noise texture (oct: octaves, freq: frequency, pers: persistence, lac: lacunarity).

Emotion Scale	Function Predicting Absolute Scale Values
MF	$101.36 + 9.27L^{0.1} - 30.06oct^{0.05} - 6.06freq^{0.3} - 53.38pers^{0.1} - 25.15lac^{0.1}$
HS	$116.12 + 6.10L^{0.1} - 32.30oct^{0.05} - 13.13freq^{0.1} - 48.81pers^{0.1} - 29.33lac^{0.1}$
HL	$42.67 + 0.064L - 12.46oct^{0.05} - 11.35freq^{0.1} - 5.84pers^{0.5} - 17.23lac^{0.05}$

persistence and lacunarity. Octave is the number of noise functions added together, frequency is the base frequency, persistence is the number which determines the amplitude of the each successive functions and lacunarity is the number that sets the frequency of the noise functions. More comprehensive explanations can be gathered from libnoise library. Fig. 4.3 shows some samples of Perlin images.

To determine the Perlin parameters for any given image, we developed an inverse Perlin method, which consists of extraction of texture features of Perlin images and machine learning to give us Perlin parameters from those features. To extract texture features from the raw images, we used the MR8 texture descriptors proposed by Varma & Zisserman [2005]. These features are low dimensional, include rotational invariance properties, and allow us to perform parameter interpolation from a single image. We used the same MR8 filter bank of Varma & Zisserman [2005] that consists of one bar filter, a set of edge filters at three scales and six orientations, a Gaussian filter and a Laplacian of Gaussian filter. The maximum responses of each orientation is collected and together with 2 isotropic filters an 8 dimensional feature vector is obtained. In this way the dimensions of filter responses are reduced from 38 to 8.

To create the MR8 texture descriptors, we create a gallery of Perlin images, and take the convolution of the these images with the designated filter bank. The responses are clustered, called textons, using the K-Means clustering algorithm Duda *et al.* [2001]. From a given image, the textons are extracted and their histograms are formed, which are then used as feature vectors in the classifiers. Essentially, what we propose is to create a rich training set of Perlin images (i.e. the gallery), and derive the
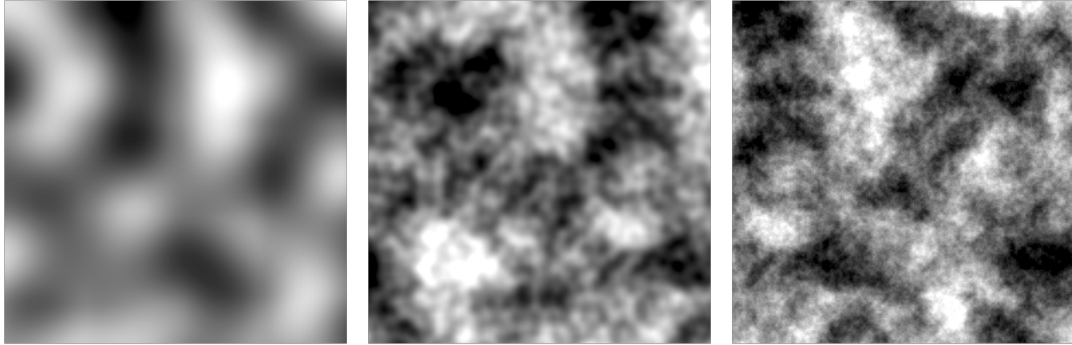
Figure 4.3: Examples of textures created with Perlin noise. These so-called Perlin images differ by just their octave parameter (1, 3, and 5, respectively), which has the most visible effect of all the four parameters. (Their remaining parameters are set as frequency=0.5, persistence=0.6, and lacunarity=2.5).

texton histograms for each such image. Then for a given probe image, we extract its texton histogram, and use it as an input to a classifier that outputs the corresponding Perlin parameters.

For training the classifier, we created eight images with different appearances for each Perlin parameter set used by Lucassen *et al.* [2011] in their experiment. This gave us a total of 840 Perlin images, whose parameters span the three emotion scales described in the previous section. For each image, we applied the filters, put filter responses from every image together and normalized them, following the procedure proposed by Varma & Zisserman [2005]. Then we computed 250 cluster centers (i.e. textons), which are the features we used for the training phase. For every image, we labelled their filter responses to the closest texton using the Euclidean distance. We created histograms of the textons and normalized them by dividing every bin by the total number of responses of each image.

We used Extreme Learning Machines (ELM), a single-hidden layer feedforward network implementation for classifying the Perlin parameters Huang *et al.* [2004]. We trained four networks using histograms to predict four Perlin parameters and tested their accuracy with cross-validation. Then using those parameters and the formulas of Table 4.1, we calculated the predicted emotion scores for each image. Figure 4.4 shows the result of ELM prediction on Heavy-Light emotion scores for a set of syn-

thesized Perlin images, and demonstrates that the classifier successfully estimates the parameters, and subsequently, the emotion scores are calculated accurately.
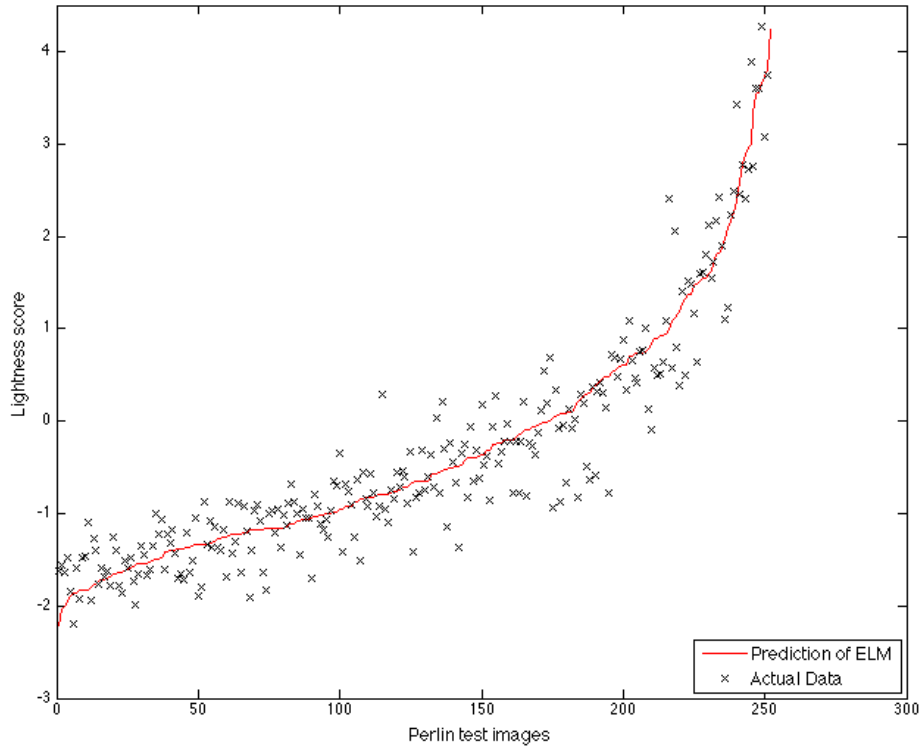
Figure 4.4: Heavy-Light emotion scores of a set of Perlin images. Red line shows the scores calculated using the predicted Perlin parameters by ELM. Black crosses are the scores calculated using the actual parameters.

The results of MART dataset by computing Perlin parameter and using the psychophysical model are shown in the three bottom images of Figure 4.2, in which the MART paintings are sorted according to the emotion scores. The paintings sorted from heavy to light (Fig. 4.2(d)) follow a luminosity pattern. The distribution of hard to soft (Fig. 4.2(c)) has a similar pattern: in general, the darker a painting, the harder it is scored. However, 'texture' itself plays a great importance in the distribution of images from 'hard' to 'soft'. This observation is in agreement with the results of Lucassen *et al.* [2011] which reports the hard-soft scales independency of colour and its domi-

nance by texture. Hence the parameters that bring texture to the fore, such as brush-strokes, complexity, composition become essential elements. Paintings with chaotic structure and granular patterns are ranked as hard, whereas paintings with geometrical structure, neat lines and a look of 'matte' finish are ranked as soft. We observe a similar pattern in the distribution of masculine-feminine (Fig. 4.2(b)): dark and chaotic paintings seem to be more masculine, whereas light and simple designs are feminine. This observation again supports the findings of Lucassen *et al.* [2011], when they report the effect of texture as being most influential in the hard-soft scale, followed by masculine-feminine, and then heavy-light.

4.2.3 Coloured vs. Gray-scale Images

The Relative Scale user study we conducted produced valence ratings of paintings viewed in gray-scale. We have contrasted these valence ratings with the Relative Scale rating of the coloured versions of MART dataset, reported in Section 2.5.2.

Even though the annotation approach is essentially the same in both studies (but with different sets of subjects performing the evaluation), we could observe that there was a low correlation between the preferences of subjects (correlation coefficient: 0.3674). This result both stresses the relative importance of colour in valence judgments, and also shows the magnitude of the contribution due to texture.

Fig.4.5 displays the relation between the results of these two sets of annotations. We sort the paintings by the annotated valence in gray-scale, and show these values in a black line that monotonically increases from 4 to 47. The scattered points are the valence annotations for the same images, but made on coloured paintings. The regression line, shown in red, shows the small but positive correlation between the two annotations. These results suggest that texture by itself has a smaller contribution for the positive and negative emotional valence of coloured paintings compared to the contribution of colour. The emotional valence of gray-scale images (including black and white photographs) should be assessed independently.
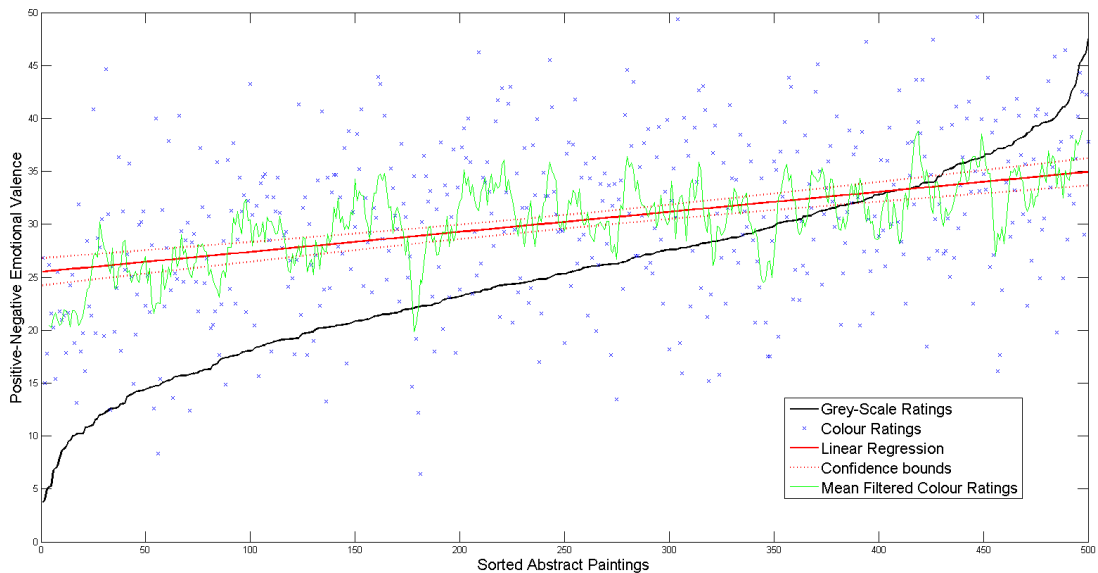
Figure 4.5: Correlation between annotations of gray-scale and coloured paintings. The x axis shows the paintings index sorted by valence. The black line indicates the valence ranking on gray-scale dataset. The blue crosses are the rankings for these images on coloured dataset. The green line illustrates the mean filtered results on the coloured set, and the red line shows the linear regression fit to the ratings on the coloured set. The small positive correlation manifests itself in the slope of the red line.

4.3   Visual Attention and Emotional Content

We now turn to the exploration of content within paintings. When making a valence judgment, different parts of a painting may influence a subject in different ways. One way of analysing which parts are most influential in the evaluation is to track the subject's gaze while he or she looks at the painting. Another way is to use a computational model that can simulate the gaze behaviour of a subject. In this section, we report an eye tracking study on non-coloured version of MART dataset paintings that replicates the eye tracking study described in Section 3.4. Moreover, we describe how we apply the saliency models to understand if computational models of bottom-up attention can be used to predict emotional valence.

4.3.1 Eye Movements

To investigate the link between the emotional content and the way people look at abstract paintings with no colour, we replicate the eye-tracking study reported in Section 3.4 on the gray-scale version of the same 110 selected abstract paintings.

We recorded eye movements of 12 participants (8 male and 4 female), who had normal or corrected to normal vision. To present the stimuli and to calculate the fixation locations and durations, we used the *ASL Eye-Trac 6* software in a full-size 19 inch screen (ASUS VW192T+, 1680x1050 resolution). To calibrate the eye positions and to validate the calibration, subjects were asked to focus on 9 fixation spots on the screen, which appeared one by one in random order. Each painting was randomly shown during 7 seconds to every observer and a noise image was placed between the paintings.

To analyse the data we follow the same procedure as presented on Section 3.4.2. The results shows a positive attention bias (63%) when people look at the abstract paintings with no colour. Only 4 paintings presents a neutral attention bias. In Table 4.2, is possible to see also that there is a Positive Attention Bias (PAB) mosttly for positive paintings and even on the negative paintings people prefer to look at the positive parts (50%).

LAB+SIFT	#	PAB	NAB	Neutral
All paintings	110 (100%)	69 (63%)	37 (34%)	4 (3%)
Pos paintings	82 (100%)	51 (62%)	31 (38%)	0 (0%)
Neg paintings	28 (100%)	14 (50%)	10 (36%)	4 (14%)

Table 4.2: Distribution of non-coloured paintings according to "emotions" of the attentional bias for Relative Scale ground-truth.

4.3.2 Visual Attention

It is postulated that attention is driven by a combination of bottom-up and top-down components, the former being data-driven, and the latter depending on higher level cognitive factors like intentions, context, knowledge, and such Frintrop [2011]; Itti & Koch [2000]. Most computational models of attention are bottom-up, as the top-down

part is extremely complex and very difficult to model. If our premises hold, we expect that the visual attention while viewing abstract paintings will mostly be a bottom-up process, especially in the absence of narratives that may direct the subject to look at an image in a particular way.

To validate this, we have compared the predictions of a successful computational saliency model, implemented by Walther & Koch [2006], with eye-tracking results obtained on coloured version of the MART dataset reported on Section 3.4, and with eye-tracking results obtained on the gray-scale version of MART paintings described in the previous section.

We believe that inspecting the distribution of saliency over the images in our study can also be potentially informative. The saliency algorithm of Walther & Koch [2006] is partly a reimplementation of the iNVT toolkit of Itti *et al.* [1998]. This saliency model is one of the best known bottom-up computational attention models, and it is based on Treisman and Gelade's feature integration theory Treisman & Gelade [1980]. In this model, different channels of features (including colour, intensity, edges, etc.) are evaluated over the visual scene independently before being integrated into a saliency master map that shows what part of the visual scene is interesting.

We first establish whether computational models do a good job of predicting the gaze patterns of subjects on abstract paintings. We use the Fixation Analysis Software of Le Meur & Baccino [2012] to compare predictions of the computational model of attention (i.e. using Dirk and Koch's software) with the eye fixation ground truth data of MART dataset for coloured paintings and the ground truth results collected for gray-scale paintings.

The Area Under the Curve (AUC) measure, computed from the ROC curve, is the suggested method of assessing similarity of these streams. An AUC value of 0.50 indicates random performance, where 1.00 denotes a perfect match. The average AUC between eye-fixation ground truth and simulations of the computational model of attention were 0.93 for colored paintings and 0.94 for gray-scale paintings. These results indicate that 1) the bottom-up computational approach is a good substitute for the eye tracking study, and can be used to find the salient locations of abstract paintings; 2) there is little difference in salient locations for coloured and gray-scale paintings. While colour is the dominant modality for inducing emotions, texture dominates in guiding the attention of the subject.

4.4 Classifying Positive and Negative Images

In this section we implement a classifier to automatically estimate the emotional valence of a given painting. In this study we also apply the bag-of-words paradigm based on SIFT descriptors (detailed in Section 3.2.2), extracted from either a dense grid over the image, or just from salient locations. We train a Support Vector Machine with a histogram intersection kernel for supervised learning, using the fast approximation of Maji *et al.* [2008]. The TrueSkill ratings are used as ground truth labels.

For testing the approach, a 5×2 cross validation setup is used, where the images are assigned to folds randomly. We have obtained a 73.0% correct classification rate when the entire images are used, which is close to the 79.0% human classification rate reported in Section 4.2.1. Using the computational saliency model described in the previous section, and extracting SIFT features from only the top 2.5% most salient locations in the image with the same experimental setup gives a classification rate of 73.9%. The difference is not significant (as established by a t-test), but a much smaller portion of the painting is evaluated for reaching this accuracy. Fig. 4.6 illustrates some examples of paintings and their saliency maps, with recorded fixations superimposed on them. Using the inverse Perlin coefficients, and the emotional scales of Lucassen *et al.* [2011] that are derived from these coefficients with a Support Vector Machine classifier directly results in an accuracy of 62.0%. The primary reason is that the latter set of descriptors are very parsimonious, compared to the SIFT-based descriptors, and while useful in qualitative evaluation, do not contain enough discriminatory power for automatic classification.

4.5 Discussions

In this chapter we present our study of how texture affects the positive and negative emotions invoked by abstract paintings. To this effect, we conducted a Relative Scale annotation to establish a ground truth of emotional valence for 500 abstract paintings with no hue value. The analysis of annotations revealed that people reacted differently to gray-scale versions of paintings when annotating the emotional content. Our experiments confirmed that colour played a greater role than texture in emotional assessment, yet the effect of texture was visible.
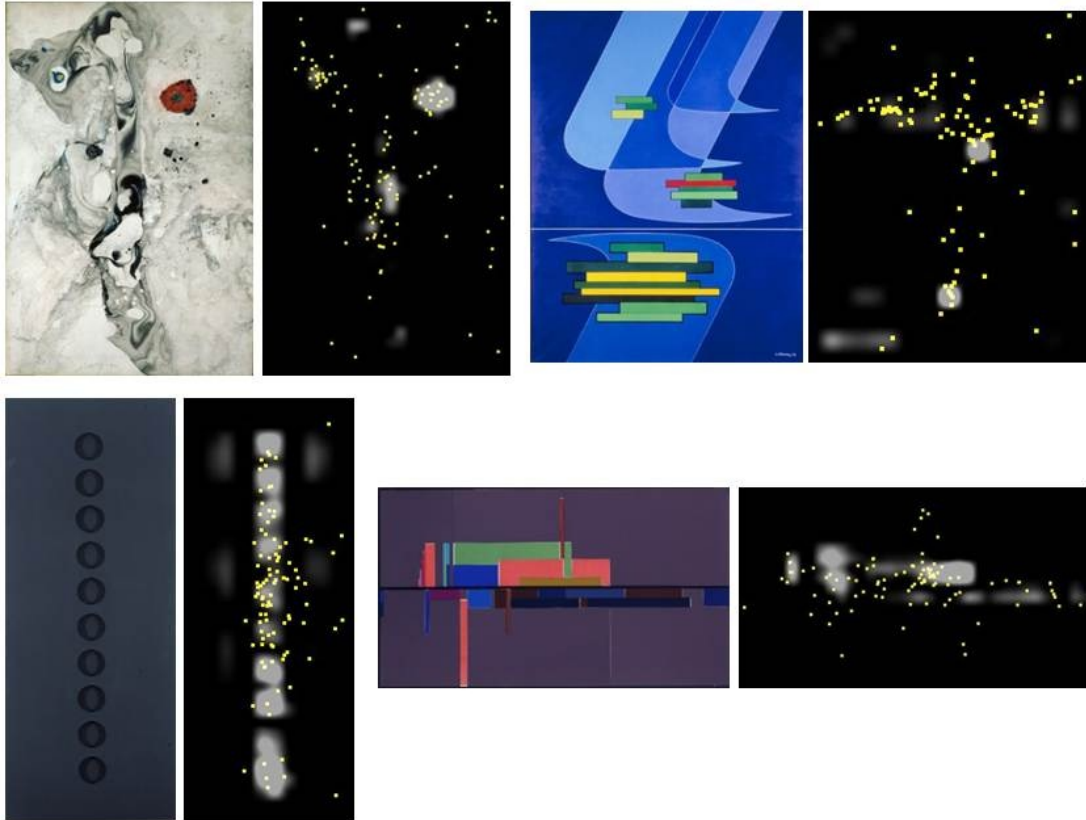
Figure 4.6: Saliency maps of paintings with superimposed fixated locations. ©*MART - Archivio fotografico e Mediateca.*

We proposed an inverse-Perlin parametrization method to map a given image to a four-dimensional representation of its overall texture content. We have used this parametrization in a psychophysical model to depict the paintings in several emotional scales. A quantitative analysis showed that texture was especially influential in giving an impression of hardness and softness.

Existing computational models of attention use colour and texture jointly to predict where the bottom-up attention of a subject will be directed, and as we have observed via eye tracking studies, they make a good job of predicting attention in abstract paintings. It is quite interesting that these models, when used on gray-scale versions of the paintings, predict very similar saliency maps. This may be partly due to the abstract nature of the images, and partly due the prominence of texture-based channels over

colour-based channels in the computational approach. However, the user study we report in this work established that the emotions induced by gray-scale images are quite different than the emotions induced by their coloured counterparts.

Thus, in the next Chapter, we investigate the contribution of two-colour combinations in evoking positive or negative emotions in abstract paintings.

Chapter 5

Contribution of Colour Combinations

The previous chapter showed that the contribution of colour is more prominent to emotional response than texture. Indeed, colours have a strong influence on humans' perception and it is commonly used to transmit information. The different amount of one or more colours conveys different messages which are closely connected with the mental and emotional states Ivanova *et al.* [2012].

Therefore, in this Chapter we present our study on the contributions of colour combinations. Specifically, we show how we use the state-of-the-art computer vision techniques to understand the contribution of two-colour combinations in evoking positive or negative emotions in abstract paintings. Our study is based on Itten's theory of colour expression, according to which it is observed that it is difficult to define the expressive properties of a colour without relating it to other colours Itten [1974]. Indeed, in his theory of colour expression, Itten observed and defined the emotional aspects of two-colour combinations. Using Itten's theory of colour expression Itten [1974], we investigate how colour combinations can be exploited to improve machine learning approaches for classifying emotions elicited by abstract paintings. In addition we discuss how the insight gained can, potentially, be used to revise and enhance Iten's theory.

5.1 Related Work

5.1.1 Colour Theory

Colour is one of the most salient elements of an image and it is the subject of study in many areas, including physics, psychology, philosophy, neuroscience, art, computing and many others. These and other fields attempt to understand the nature of colour, how it is constituted, how people perceive it, and how it influences people's life. Initially, Newton [1704] observed the physical aspects behind the perception of colour and discovered that if some wavelengths of light are reflected better than others, then the object appears colored: "When light falls on an object, some is absorbed. The light that isn't absorbed is reflected off the object's surface; this is the light we see." Livingstone [2002]. Opposed to Newton, Goethe proposed a colour wheel based on physiological experiences (e.g., after images) and psychological aspects (e.g., moral associations, symbolic and mystic use of colours) to describe complementary colours: "The chromatic circle... [is] arranged in a general way according to the natural order... for the colours diametrically opposed to each other in this diagram are those which reciprocally evoke each other in the eye. Thus, yellow demands purple; orange, blue; red, green; and vice versa: thus... all intermediate gradations reciprocally evoke each other; the simpler colour demanding the compound, and vice versa." von Goethe & Eastlake [1840]. These and subsequent physiological and psychological theories of the effects of colour, were used by famous painters and are used still today in order to improve the aesthetic aspects of paintings.

In abstract paintings, colours have an essential role. They are particularly studied and used by artists to convey the emotional message to the beholder. Artists such as Kandinsky, Itten, Albers and others, studied how the presence of colours, in different proportions and combinations, affects the viewer emotionally and applied these studies to their artworks. Kandinsky [1914] describes the emotional effects of individual colours, affirming that it depends on two factors: a physical impression, "one of pleasure and contentment at the varied and beautiful colours" and a psychic effect, in which colors "produce a corresponding spiritual vibration, and it is only as a step towards this spiritual vibration that the elementary physical impression is of importance" Kandinsky [1914].

More recently, Johannes Itten, a notable expressionist painter, theorist and professor of the Bauhaus School[1] studied and taught the aesthetic and expressive aspects of colours and its interpretation. In his studies of the relationship between the chromatic scale and the spiritual evolution, he provided a set of rules of colours and colour combinations, which he called the "objective principles of colours" Itten [1974]. Indeed, Itten affirms that "The concept of color harmony should be removed from the realm of subjective attitude into that of objective principle." The fact that many paintings induce similar emotions into the viewer, indicates that there may be consistency on the way people perceive colours on abstract paintings.

Moreover, Itten observed that it is difficult to define the expressive properties of a colour without relating it to other colours Itten [1974]. Indeed, he formalized theories on the emotional effects of two-colour combinations and their properties to generate harmonious artworks Itten [1974]. He describes the emotional values of the colour combination hues defined on the 12-color circle. Itten's colour wheel is widely used in art studies and is composed of 12 hues of primaries, secondary and tertiary colours which are respectively (Figure 5.1): yellow, red, blue, orange, green, violet, yellow-orange, red-orange, red-violet, blue-violet, blue-green and yellow-green. For Itten the two diametrically opposed colours in the wheel are complementary, which form a harmonious dyad, for instance red/green, blue/orange, yellow/ violet, etc. These 12 hues are varied by five levels of luminance and three levels of saturation and are represented by him in a Colour Sphere with a total of 180 colours. In the Colour Sphere, the 12 hues are located in the equatorial zone, the luminance varies along the meridians, and the horizontal sections contains the degrees of saturation: "If I use the color sphere, I can get an indefinite number of harmonious dyads. The only requirement is that the two tones be symmetrical with respect to the center of the sphere. Thus if I take a tint of red, the corresponding green must be shaded in the same degree as the red is lightened" Itten [1974].

---

[1]The Bauhaus School was an German art school which has "...strong influence of scientific studies of color on the development of abstraction..." Dickerman *et al.* [2012]
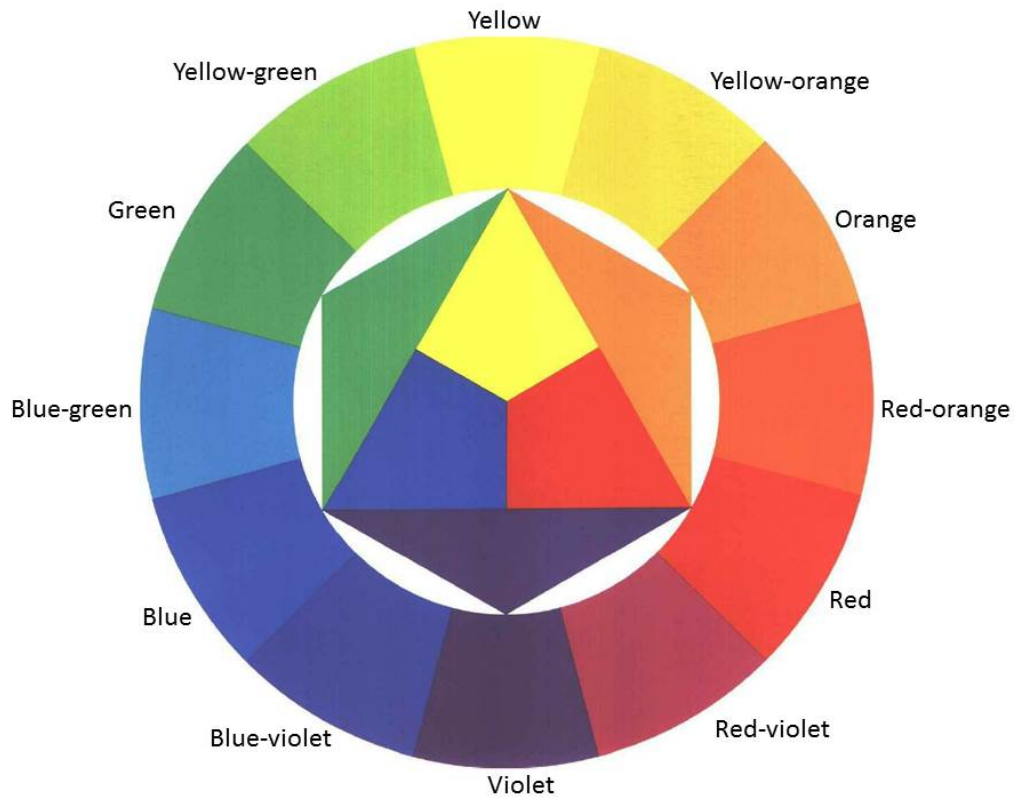
Figure 5.1: Itten's Colour Sphere. Extracted from Itten [1974]

5.1.2 Emotion Recognition

Recently there is a growing interest from different research communities in understanding the emotional response of the viewer while seeing colours in images, artworks, etc. A psychological study on the effects of colours on emotions based on Pleasure, Arousal and Dominance model concludes that brighter colours are more pleasant, less arousing, and induce less dominance than the darker colours Valdez & Mehrabian [1994]. Ou *et al.* investigate how eleven emotional scales are associated with three colour-emotion factors (i.e., colour activity, colour weight and colour heat) of single colours Ou *et al.* [2004a] and two-colour combinations Ou *et al.* [2004b] determined by means of the factor analysis method. These studies show that there is consistency in the way people perceive colours. From a computational perspective, colour, along with shape and texture has been extensively used as a feature in the research that focuses on art-

works Machajdik & Hanbury [2010]; Yanulevskaya *et al.* [2008]; Zhao *et al.* [2014]. In this Chapter, we present an analysis of the emotional effects of colour combinations portrayed in abstract paintings.

5.1.3 Datasets and Ground Truth Collection

We conduct our analysis on the two sets of 500 abstract paintings each: MART dataset that is described in Section 2.3.1, and deviantArt dataset detailed in Section 2.3.2. For this study, we use the ground truth from Relative Scale method of annotation for both MART and deviantArt datasets. The method of Relative Scale annotation and how we construct the ground truth is detailed in Section 2.5.2, in which the TrueSkill ratings are used as ground truth labels.

5.2 Proposed Method

In this section we provide details on how we use two-colour combinations to automatically classify abstract paintings in positive and negative emotions. Specifically, we analyse the contribution of adjacent colours in a painting and use segmentation technique to generate the colour combinations features to emotionally classify abstract paintings.

The system flow of our proposed method is illustrated in Figure 5.2. We first collect the database and the corresponding ground truth of positive and negative emotions evoked by abstract paintings. Afterwards, we use the method of van de Weijer *et al.* [2009] to map the paintings into Colour Name features and then cluster the entire dataset into 180 clusters (i.e., the largest number of colours in Itten's model) which were used as our *Colour Palette*. Then, we calculate the closest value between the Colour Palette and each pixel of the painting, and generate a new painting to segment. As our aim is to find colours that are next to each other (i.e., the colour combination), for each segment, we find its neighbour and select the most frequent colour for both segments. Finally, a histogram with all colour combinations found in the image is generated. After all features vectors are computed, we split the data in half into training and test set. We train the classifier with the appropriate labels of positive and negative emotions and during the evaluation we compare the automatic assigned class with the
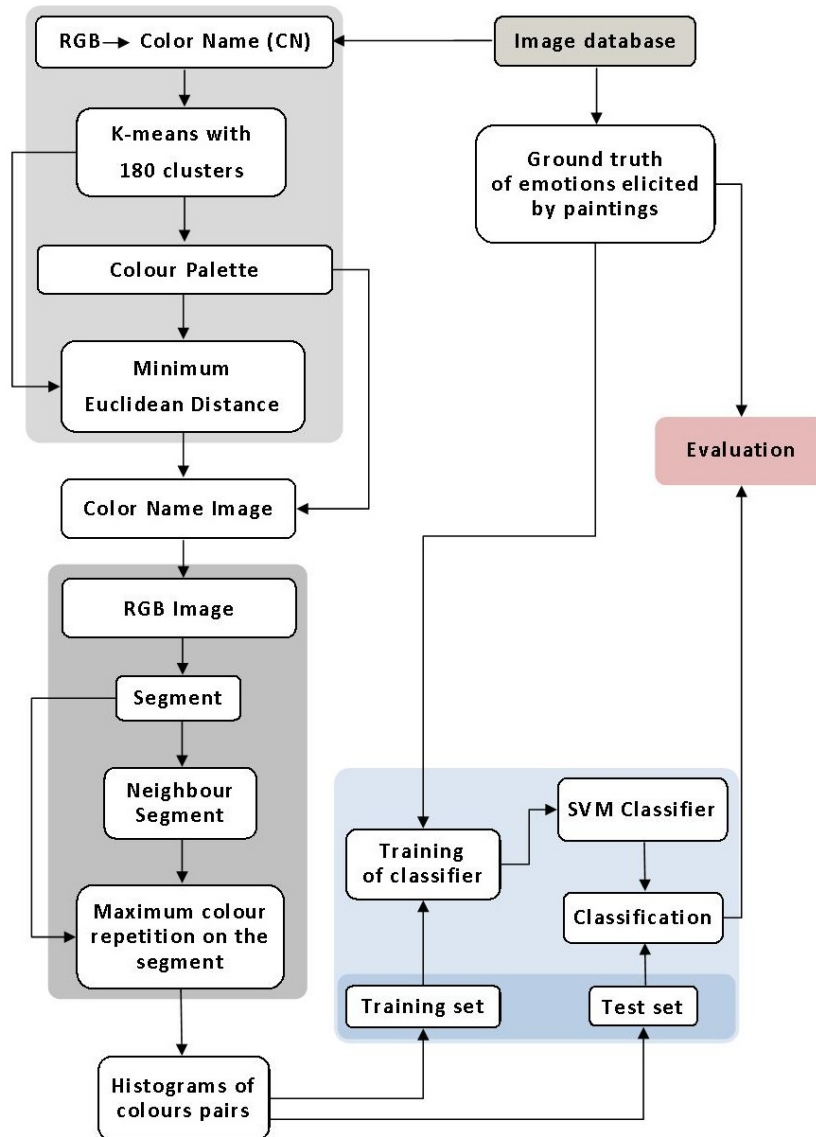
Figure 5.2: System Flow.

ground truth labels.

5.2.1 Colour Features

To analyse the contribution of colour combinations, we compare the results of the colour combination framework with the LAB-based Colour Visual Words presented in

Section 3.2.2. Specifically, LAB Visual Words are based on local features that are discretized into a visual vocabulary of colours (i.e., LAB Colour Space) and represent the paintings as histograms over visual words. In Chapter 3 we observe that LAB Visual Words are more effective for emotional classification of abstract paintings compared to SIFT Visual Words.

In this study, we propose a novel framework to recognize two-colour combinations of abstract paintings and classify them in positive and negative emotions from a more global perspective. We use the Colour Name method of van de Weijer *et al.* [2009] as feature to learn the colours descriptions of abstract paintings and to generate our *colour palette*. van de Weijer *et al.* [2009] describe colors through the linguistic labels that humans use to express the colours in the world (i.e., 'blue', 'red', 'yellow', etc.). Therefore, we use the method of van de Weijer *et al.* [2009] as a feature to map the RGB colors from the abstract paintings into Colour Names Features (CNF). We used the default parameters of van de Weijer *et al.* [2009] method which calculates the mean of patches with 21 pixels each to assign the nearest colour to each pixel on the image.

Afterwards, we use the k-means algorithm to cluster the CNF from the entire dataset of abstract paintings in 180 distinct colours (i.e., k clusters are equal to 180). We select 180 colours matching the largest number of colours from Itten's colour model introduced in Itten [1974].

Once we identified the 180 k cluster centers, the minimum Euclidean distance is calculated in order to find the Colour Name Feature for the respective cluster center, and then we select the RGB values from the Colour Name Feature to generate a *Colour Palette* which contains both Colour Name Feature and its respective RGB values. Figures 5.3 and 5.4 displays our Colour Palette with the 180 colours respectively for MART and deviantArt datasets, that were used as reference to the colour co-occurrence calculation.

Afterwards, for each pixel in the image, that was before converted into CNF, we calculate the closest value of the CNF Colour Palette. Then, for the respective colour (i.e., CNF Colour Palette) we select the RGB Colour Palette and assign this colour to a new image. This new image, with a RGB Colour Palette, is used to segment in order to find the colour patches that are near to each other in the image. Figure 5.5 provides one example from MART dataset of the original painting and the new painting transformed into Colour Name Features.
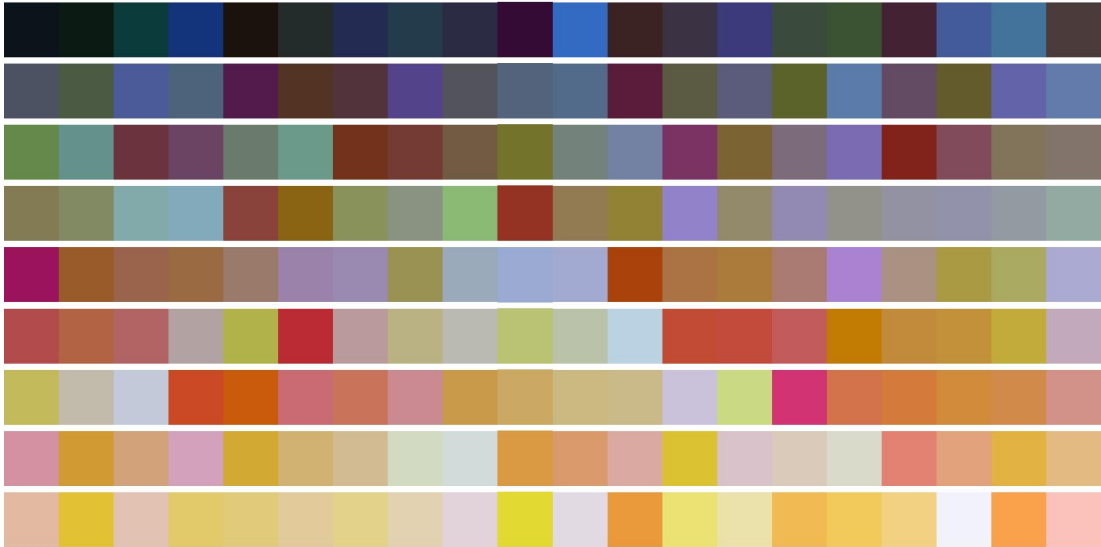
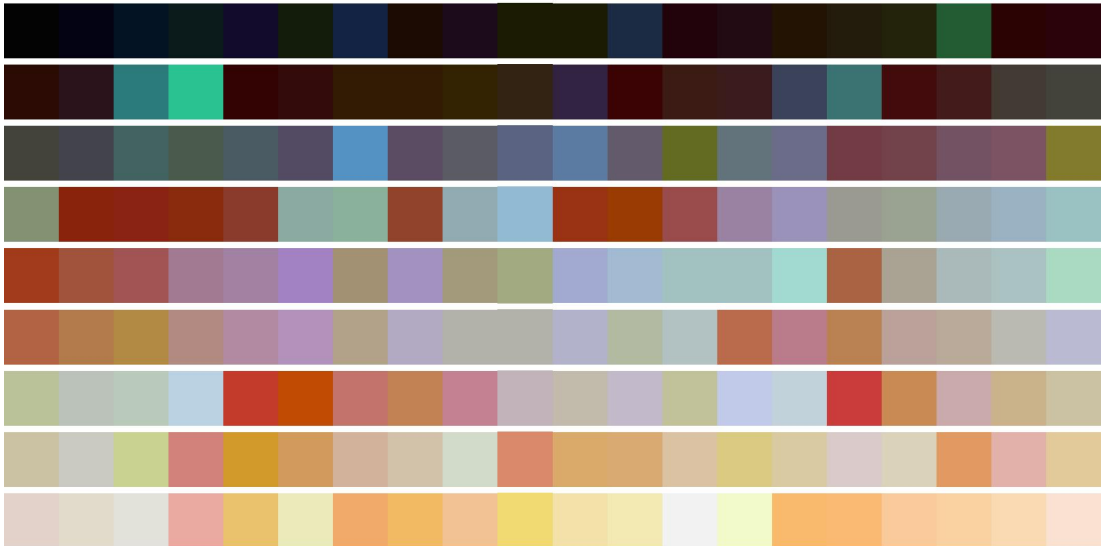Figure 5.3: Colour Palette extracted from MART dataset with 180 colours.

Figure 5.4: Colour Palette extracted from deviantArt dataset with 180 colours.

Figure 5.5: MART paintings transformed in Colour Name Feature. ©*MART - Archivio fotografico e Mediateca.*

5.2.2 Segmentation

To segment the paintings we use the graph-based image segmentation method of Felzenszwalb & Huttenlocher [2004]. Their method effectively measures the evidence for a boundary between two regions, which corresponds to our purpose of finding the adjacent colours in the paintings. We use the publicly available algorithm by the authors with the parameters $\sigma = 0.8$ that is the Gaussian filter to smooth the image, $k = 100$ which is used to compute the threshold function.

For each segment of the image we find the maximum repetition of one colour and assign that colour to the segment. Then, we find the neighbour segment and apply the same procedure. Afterwards, a histogram of all colour combinations occurring in a painting is generated. The histograms of colour combinations for all paintings are used as feature vectors to the classifier. Figure 5.6 provides an example of the segmentation applied in this study to find the adjoin colours.
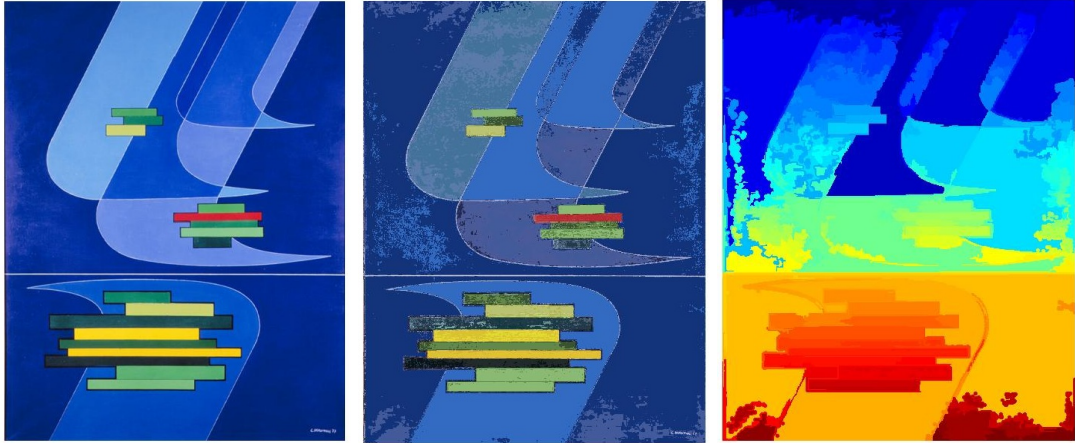
Figure 5.6: Example of our process in finding colour combinations. From the left: (1) the original image; (2) The image repainted with our colour palette; (3) Visualization of all segments of the image. In (3) each colour represents one segment. ©*MART - Archivio fotografico e Mediateca.*

5.2.3 Classification

In order to compare our results with the LAB Colour-based Visual Word, we apply the same method for classification used in Section 3.2.3. We train a Support Vector Machine with a histogram intersection kernel for supervised learning, using the fast approximation of Maji *et al.* [2008]. The TrueSkill ratings are used as ground truth labels. We test our model in a 2-fold cross-validation setup, where the images are assigned to folds randomly, and repeated 1,000 times.

5.3 Evaluation and Results

In this section we present the experimental results on the positive and negative emotional classification of abstract paintings. Table 5.1 shows the classification accuracy results on Colour Combination approach and on LAB Visual Words.

Using the colour combinations framework we have obtained a 78.1% correct classification rate for MART dataset. This result shows that our method of using colour combinations as feature improves the performance of the classification. Particularly, most of the paintings on MART dataset is composed by uniform colours (i.e., 781.808 unique colours) which make simpler the assess to the Colour Name Features.

On the other hand, the paintings on deviantArt dataset were represented with a great amount of colours (i.e., over 4.4 million of unique colours), which made more difficult to determine the Colour Name Features. To solve this problem, we split the set into smaller blocks of 100,000 random colors and apply k-means on each block to get 100 centroids per block. As a result, we got 44.400 centroids from the blocks and then we clustered it using the 180 k clusters. Although deviantArt dataset presents a large amount of colours, the resulting Colour Palette seems consistent with the colours represented in this dataset. However, this may be the reason that the classification accuracy (i.e., 76.6% correct classification rate) is slightly lower than LAB Visual Words that has 76.9% correct classification rate.

Table 5.1: Classification accuracy comparison.

Accuracy	LAB Visual Words	Colour Combination
MART	0.768	0.781
deviantArt	0.769	0.766

5.4 Discussions

In this Chapter we analyse the contributions of colour combinations to automatically predict emotional valence in abstract paintings. We used the art theory of Itten [1974], which points out that colors are rarely experienced isolated. Indeed, Itten [1974] describes the expressive properties of colours on two-colour combinations.

By applying Colour Name Features to the datasets of abstract paintings, we were able to generate a new colour palette which effectively represents the colours in both datasets (i.e., MART and deviantArt) of abstract artworks. Our approach on the occurrence of colour combinations in a painting show that colour combination are an important feature, improving the accuracy of MART dataset. Even thought our approach did not improve the classification rate of deviantArt dataset, it shows to be an important approach for understanding of colours represented in the image. For instance, we could observe how different the datasets are in terms of colours.

In the future work, we are planning to make an in-deep analysis of these new colour palettes based on empirical studies. Specifically, we aim to generate Mondrian-like paintings from the Neo-plasticism style using the colours obtained from the datasets and invite people to emotionally judge these simulated paintings. We choose the paintings from Piet Mondrian, as he was a notorious Dutch painter and contributor to the De Stijl, a Dutch artistic movement. The Neo-plasticism style, considered as "the new way of treating the form", is essentially composed by horizontal and vertical straight black lines filled with primary colors (Figure 5.7 show one example of Mondrian's painting). Mondrian paintings employ two of the seven color contrasts from Itten theory: the contrast of hue and the contrast of proportion. "Mondrian confined himself to the fundamental colors yellow, red, blue, white and black. Each of these colors has its unique character and special weight. The position of each color is very important, and so is its orientation, horizontal or vertical." Itten [1974].
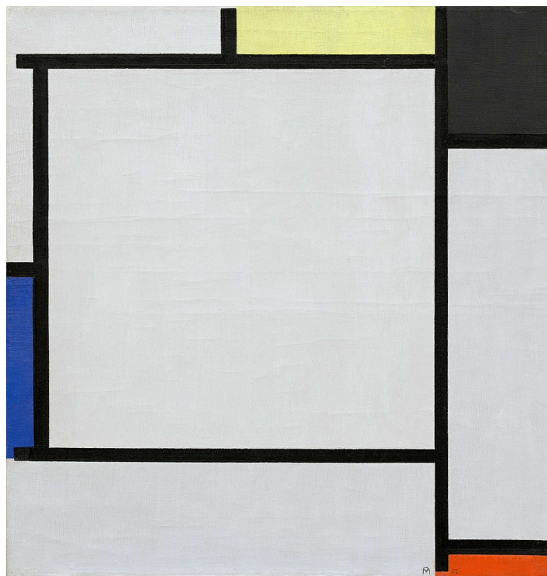
Figure 5.7: Piet Mondrian, Tableau 2, 1922. Extracted from: Solomon R. Guggenheim Museum, New York (http://www.guggenheim.org/new-york/collections/collection-online/artwork/3013)

With the emotional responses from the simulated paintings and the theory of art, we plan to generate new features based on Itten's theory, such as, the amount of colours (i.e., contrast of proportion), the position (i.e., the composition) and the distances between colours.

Chapter 6

Contribution of Metadata[1]

The last Chapter of this thesis, we present an analysis of the influence of the metadata (i.e., titles, description and/or artist's statement) associated with the abstract painting and investigate how it can be a significant feature to the automatic emotion recognition of abstract paintings.

Specifically, we extract the corresponding metadata for the paintings and apply sentiment analysis to detect the emotional meaning of these texts. We use the method detailed in Section 3.2, based on colour, shapes and textures to augment computation of the emotional information of the artwork. Finally, we propose two approaches to combine textual features with visual ones, to improve the performance of affective image classification. The first approach is based on weighted linear combination. In the second approach, we propose a novel joint flexible Schatten p-norm model which can exploit the sharing patterns between visual and textual information for the emotion analysis of paintings. We apply our multimodal approach on the two datasets of abstract paintings described in Section 2.5: the MART dataset and devianArt (dA) dataset.

To summarize, the main contributions presented in this chapter are: (1) we study the role of metadata on positive and negative feelings induced by abstract paintings; (2) we analyse the contribution of metadata on the affective image classification and we propose a novel joint flexible Schatten p-norm model which can exploit the sharing patterns between visual and textual information for emotion analysis of paintings; (3)

---

[1]This study was submitted as Sartori *et al.* [2015b]

we apply our approach on the two datasets of abstract artworks and make a qualitative analysis.

6.1   Related Work

The relationship between paintings, emotions and art appreciation has been extensively studied in many fields. Psychological studies on aesthetics and emotional responses to art have shown that titles and descriptions influence the perceptual experience of paintings. These works demonstrated that people describe paintings differently after they read the title Franklin *et al.* [1993]. Also, they postulate that title and description of paintings can improve the understanding Leder *et al.* [2006] and the aesthetic evaluation Hristova *et al.* [2011]; Millis [2001]. The influence of title and description is even more crucial for abstract artworks where the visual clues are open to interpretation.

One of the initial works is from Franklin *et al.* [1993], which observed that titles help to guide the interpretation of artworks. Millis [2001] investigate the effects of metaphorical (or elaborative) and descriptive titles on the understanding, the level of emotional responses, interest, enjoyment, and elicited thoughts of artworks. He observed that metaphorical titles, increases the aesthetic experience more than no title or descriptive titles. In a similar study, Leder *et al.* [2006] remarks that elaborative titles increase the understanding of abstract paintings, but do not affect their appreciation. Hristova *et al.* [2011] investigated how the style of the paintings and the titles can influence the fixation duration and saccade amplitude of the viewers. They observed that, when the viewers are presented with the title of a painting while looking at the artwork, the average number of saccades of each viewer to observe the work tend to decline. This finding draws the conclusion that titles lead to a more focal processing of the paintings. These works demonstrate that the metadata influences the perception of artworks. In this study, we use metadata as an additional feature to improve the affective classification of abstract paintings.

The various computational approaches have shown the importance of detecting emotion in textual information Balahur *et al.* [2011]; Strapparava & Mihalcea [2008]; Strapparava *et al.* [2011]. The sentiment analysis technique which uses the computational linguistics to extract the human sentiment in text, is increasingly used in many areas, such as marketing and social networks to improve their products and/or services.

Recent works, including Gong *et al.* [2014]; Hwang & Grauman [2012]; Wang *et al.* [2009], empathize the importance of using textual information associated with images for creating stronger models for image classification. Liu *et al.* [2011] proposed two methods to extract emotional meaning of text, provided by images on the Web in the form of tags, and combine them with visual features. They demonstrated how textual information associated with images can improve affective image classification. In our work, we use a multimodal approach to investigate how the title, descriptions and/or artists statement, associated with paintings can lead to a stronger model to compute affective analysis of abstract paintings.

6.2 Textual Features for Assessing Emotions in Abstract Paintings

In this section we describe how the sentimental features were extracted from the metadata (i.e., titles, descriptions, artist statement) associated with a painting.

6.2.1 Datasets and Ground Truth Collection

We conduct our analysis on the two sets of abstract paintings: MART and deviantArt datasets, that are reported in Section 2.3. For this study, we use the ground truth from the Relative Scale method of annotation for both MART and deviantArt datasets. The method of Relative Scale annotation is reported in Section 2.5.2, and how we construct the ground truth for MART and deviantArt datasets is detailed respectively in Sections and 3.3.1 and 3.3.4. The TrueSkill ratings are used as ground truth labels.

6.2.2 Text Data Selection

To obtain the descriptions and/or artist statements for MART dataset, we searched through museums and art related websites, as well as the official pages of the artists. In total, we found descriptions and/or artist statements for 158 paintings. Most of these descriptions were in Italian, as the artists were Italian. The titles of the paintings were already provided by the dataset.

In deviantArt dataset, the artists themselves provided an opinion and/or a description on/about their paintings. From those, we selected 158 descriptions which represent the artists' intention in detail. The titles and descriptions of this dataset are in English.

Figure 6.1 provides two examples of titles and descriptions for paintings in the MART (top) and deviantArt (bottom) datasets.

Kleine Welten I, 1922.

In *Kleine Welten* (Small worlds) Kandinsky demonstrates the different effects of drypoint, lithography, and woodcut, providing four examples of each technique. As suggested by the portfolio's title, each abstract image is a world unto itself; meaning is generated exclusively through the interplay of line, plane, and color and the specific properties of the given medium. [...] Drypoint expressed passion and haste, foregrounding line and point. The limited number of impressions that could be pulled from a metal drypoint plate made it an "aristocratic" medium. Woodcut was more egalitarian in that it allowed for greater edition sizes; it also best conveyed planar relationships. Lithography was the most painterly, and its unlimited reproducibility made it the most "democratic," a quality that led Kandinsky to proclaim it the medium of his time.

Visual Score: 0.2573 (positive sentiment)
Title Score: -0.0556 (negative sentiment)
Description Score – average: 0.0224 (positive sentiment)

Alive, 2006.

"Life can be such a mess sometimes. But it can't be that bad. It's good to be alive. We have the capability to survive."

Visual Score: 0.2234 (positive sentiment)
Title Score: 0.3926 (positive sentiment)
Description Score – average: 0.2073 (positive sentiment)

Figure 6.1: Example of titles and descriptions on abstract paintings with their respective emotional scores. The painting on the top is from MART dataset (©*MART - Archivio fotografico e Mediateca*) and the description was extracted from The Museum of Modern Art ("MoMA"): http://www.moma.org/collection/object. php?object_id=68437. The painting on the bottom is from deviantArt dataset. (*Courtesy of deviantArt.*).

6.2.3 Applying Sentiment Analysis

To determine the sentiment conveyed by the title and the description of a painting in the deviantArt dataset, we tokenize and part-of-speech (POS) tag the related text with Stanford Parser Klein & Manning [2003], and we use WordNet Miller [1995] for lemmatization. For each title or description we simply average the scores coming from SentiWords Guerini *et al.* [2013] based on the lemma and POS information. SentiWords is a high coverage resource associating approximately 155,000 English lemma-POS pairs to sentiment scores between -1 (very negative) and 1 (very positive).

As (to the best of our knowledge) there is no resource for sentiment scores in Italian, we utilize a multilingual lexical database called MultiWordNet[1] Artale *et al.* [1997] for the painting descriptions in the MART dataset. More specifically, we first lemmatize and POS tag the Italian titles and descriptions with TextPro Pianta *et al.* [2008]. Then, with the help of MultiWordNet we find all the English synsets associated with the Italian lemma and POS pairs. We calculate the sentiment score of a lemma-POS pair as the average of the scores provided in SentiWordNet Baccianella & Sebastiani [2010] for each synset that it is connected to. SentiWordNet is a lexical resource built for supporting sentiment classification and opinion mining applications and it provides three sentiment scores (i.e., positivity, negativity, objectivity) for each WordNet synset. We consider the difference between the positivity and negativity score as the final sentiment score of a synset. As we did for deviantArt dataset, we average the scores of each lemma-POS pair to determine the overall emotion conveyed by a painting title or description.

6.3 Fusion of Visual and Textual Features

In this section we describe the classification approach to automatically estimate the emotional valence of a given painting, and provide details on how we used both late fusion and joint flexible Schatten p-norm learning to combine visual and textual features.

---

[1]http://multiwordnet.fbk.eu/

6.3.1 Visual Features

To extract the visual features, we use the method reported in Section 3.2. We used a standard bag-of-words paradigm to extract LAB-based colour visual words and SIFT-based texture visual words. We train a Support Vector Machine with a histogram intersection kernel for supervised learning, using the fast approximation of Maji *et al.* [2008]. For testing the approach, a 5-fold cross-validation setup is used, where the images are assigned to folds randomly. We run two separate frameworks (Lab descriptors and SIFT descriptors) and then average the scores to combine them.

6.3.2 Late Fusion with Weighted Linear Combination

To combine visual and textual features we use late fusion with Weighted Linear Combination. To calculate the Weighted Linear Combination we use the following equation:

$$WLC = a * T + (1 - a) * V \tag{6.1}$$

where T is the sentiment score from text and V is the visual score (the fusion of LAB and SIFT Visual Words) and $a \in [0, 1]$ is a parameter. To choose the parameter a we analyze the performance of the decision fusion (Visual + Text) and set a to the value resulting in the highest mean performance.

6.3.3 Estimating Paintings' Emotion through Joint Flexible Schatten *p*-norm Learning

The drawback of late fusion with weighted linear combination approach is that it considers visual and texture information separately before classification. This may not capture well the emotion of the viewers about a painting because the viewers usually judge a painting by looking at it and considering its title and available description at the same time. To better exploit the visual and textual information for the emotion analysis of a painting, we propose a novel joint flexible Schatten p-norm model.

Assuming that $\mathbf{x_i}$ is a d-dimensional feature vector, $\mathbf{X_a} = \{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_{n_a}}\} \in \mathbb{R}^{d \times n_a}$ is the feature matrix from visual information and $\mathbf{X_b} = \{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_{n_b}}\} \in \mathbb{R}^{d \times n_b}$ is the corresponding feature matrix from textual information. $\mathbf{Y_a} =$

$\{\mathbf{y_1}, \mathbf{y_2}, ..., \mathbf{y_{n_a}}\} \in I\!R^{m \times n_a}$ and $\mathbf{Y_b} = \{\mathbf{y_1}, \mathbf{y_2}, ..., \mathbf{y_{n_b}}\} \in I\!R^{m \times n_b}$ are label information matrices with m classes. $\mathbf{y_i} = [0, ..., 0, 1, 0, ..., 0]$ (the position of non-zero element indicates the emotion class). To extract the painting's emotion both from visual and texture information, we propose the following optimization problem:

$$\min_{\mathbf{U}, \mathbf{V_a}, \mathbf{V_b}} \|\mathbf{Y_a} - (\mathbf{U} + \mathbf{V_a})\mathbf{X_a}\|_F^2 + \|\mathbf{Y_b} - (\mathbf{U} + \mathbf{V_b})\mathbf{X_b}\|_F^2 \\ + \lambda_1 \|\mathbf{U}\|_F^2 + \lambda_2\, rank(\mathbf{V_a},\ \mathbf{V_b}) \tag{6.2}$$

where $\mathbf{U} \in I\!R^{m \times d}$ are the common patterns shared across visual and textual information. $\mathbf{V_a} \in I\!R^{m \times d}$ and $\mathbf{V_b} \in I\!R^{m \times d}$ are the individual patterns for visual and textual information respectively. In Eqn.(6.2), the first two terms are loss functions and the regularization term $\|\mathbf{U}\|_F^2$ is used for preventing overfitting. One way to capture the relationship is to constrain the models from both visual and textual information to share a low-dimensional subspace resulting in the rank minimization problem. The last term of rank regularization captures the correlation of the visual and textual information. However, the rank minimization problem is in general NP-hard. One popular approach is to replace the rank function by the trace norm (or nuclear norm).

Inspired by Nie *et al.* [2012], we adopt the Schatten p-norm instead of the traditional trace norm to discover the low rank matrix. The Schatten p-norm ($0 < p < \inf$) of a matrix $\mathbf{M}$ is defined as $\|\mathbf{M}\|_{S_p} = (Trace((\mathbf{M'M})^{\frac{p}{2}}))^{\frac{1}{p}}$. If $p = 1$, it reduces to the trace norm. Eqn.(6.2) can be reformulated as:

$$\min_{\mathbf{U}, \mathbf{V_a}, \mathbf{V_b}} \|\mathbf{Y_a} - (\mathbf{U} + \mathbf{V_a})\mathbf{X_a}\|_F^2 + \|\mathbf{Y_b} - (\mathbf{U} + \mathbf{V_b})\mathbf{X_b}\|_F^2 \\ + \lambda_1 \|\mathbf{U}\|_F^2 + \lambda_2 \|[\mathbf{V_a};\ \mathbf{V_b}]\|_{S_p}^p \tag{6.3}$$

Based on Nie *et al.* [2012], $\|\mathbf{V}\|_{S_p}^p = trace(\mathbf{V'VD})$, where $\mathbf{D}$ is the diagonal matrix $\mathbf{D} = \frac{2}{p}(\mathbf{V'V})^{\frac{p-2}{2}}$. Then if we define $\mathbf{V} = [\mathbf{V_a}; \mathbf{V_b}]$ and $\mathbf{Y} = [\mathbf{Y_a}; \mathbf{Y_b}]$, Eqn.(6.2) becomes:

$$\min_{\mathbf{U}, \mathbf{V}} \|\mathbf{Y} - (\mathbf{U} + \mathbf{V})\mathbf{X}\|_F^2 + \lambda_1 \|\mathbf{U}\|_F^2 + \lambda_2 trace(\mathbf{V'VD}) \tag{6.4}$$

Optimization: Since Eqn.(6.4) involves the Schatten p-norm which is non-smooth and cannot be solved in a closed form, we adopt the alternating minimization algorithm to optimize the objective function with respect to $\mathbf{U}$ and $\mathbf{V}$ respectively in the following steps:

(I). Fix $\mathbf{V}$, optimize $\mathbf{U}$. By setting the derivative of Eqn.(6.4) *w.r.t.* $\mathbf{U}$ to zero, we have:

$$\frac{\partial}{\partial \mathbf{U}} = 2(\mathbf{UX} + \mathbf{VX} - \mathbf{Y})\mathbf{X}^T + 2\lambda_1 \mathbf{U}^T = 0$$

$$\mathbf{U} = (\mathbf{YX}^T - \mathbf{VXX}^T)(\mathbf{XX}^T + \lambda_1 \mathbf{I})^{-1}$$

(II). Fix $\mathbf{U}$, optimize $\mathbf{V}$. By setting the derivative of Eqn.(6.4) *w.r.t.* $\mathbf{V}$ to zero, we have:

$$\frac{\partial}{\partial \mathbf{V}} = 2(\mathbf{UX} + \mathbf{VX} - \mathbf{Y})\mathbf{X}^T + \lambda_2(\mathbf{VD}^T + \mathbf{VD}) = 0$$

Since $\mathbf{D}$ is a diagonal matrix, $\mathbf{D}^T = \mathbf{D}$, we have

$$\mathbf{V} = (\mathbf{YX}^T - \mathbf{UXX}^T)(\mathbf{XX}^T + \lambda_2 \mathbf{D})^{-1}$$

According to the optimization, we propose Algorithm 1 to solve the objective function of Eqn.(6.4).

Algorithm 1: An efficient iterative algorithm to solve the optimization problem in Eq.(6.4).

Input:
    The training data $\mathbf{X_a} \in \mathbb{R}^{d \times n_a}, \mathbf{X_b} \in \mathbb{R}^{d \times n_b}$;
    The training data labels $\mathbf{Y_a} \in \mathbb{R}^{m \times n_a}, \mathbf{Y_b} \in \mathbb{R}^{m \times n_b}$;
    Parameters $\lambda_1, \lambda_2, 0 \leq p \leq 2$.
Output:
    Optimized $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{m \times d}$.
 1: Set $t = 0$, randomly initialize $\mathbf{D}_t \in \mathbb{R}^{d \times d}$;
 2: **repeat**
    Fix $\mathbf{V}$, Optimize $\mathbf{U}$: $\mathbf{U}_t = (\mathbf{YX}^T - \mathbf{V}_t\mathbf{XX}^T)(\mathbf{XX}^T + \lambda_1 \mathbf{I})^{-1}$;
    Fix $\mathbf{U}$, Optimize $\mathbf{V}$: $\mathbf{V}_t = (\mathbf{YX}^T - \mathbf{U}_t\mathbf{XX}^T)(\mathbf{XX}^T + \lambda_2 \mathbf{D}_t)^{-1}$;
    Update $\mathbf{D}_{t+1} = \frac{p}{2}(\mathbf{V}_{t+1}^T \mathbf{V}_{t+1})^{\frac{p-2}{2}}$;
    $t = t + 1$.
     until *Convergence*;

6.4 Experiments

In this section we present the experimental results on the classification of the artworks with their corresponding textual information into positive and negative emotional categories.

6.4.1 Setup

For the late fusion with the weighted linear combination strategy, the weight a is tuned within the range $\{0.1, 0.2, 0.3, ..., 1\}$. For the joint flexible Schatten p-norm model, the parameters λ_1 and λ_2 are tuned within the range $\{10^{-3}, 10^{-2}, ..., 10^3\}$ and p is tuned within the range $\{0.01, 0.05, 0.1, 0.2, 0.5, 0.8, 1\}$. 5-fold cross-validations are used in all experiments and the best performance is reported.

6.4.2 Results

Since the metadata associated with a painting contains title and description, we perform three types of experiments: (i) visual + title; (ii) visual + descriptions; (iii) visual + title + descriptions. Table 6.1 shows the classification accuracy results using only title information. Since there are descriptions available only for 158 paintings, we consider this set to perform the experiments (ii) and (iii).

Table 6.2 shows the classification accuracy results based only on descriptions information and Table 6.3 shows the classification accuracy results based on both paintings' titles and descriptions information. From the results, it can be observed that the combination of the textual and visual features improves the performance of the classification. Moreover, we observe that more than 3% improvement in accuracy is achieved by sharing patterns between visual and textual information through the joint flexible Schatten

	Visual	Title	Late Fusion	Schatten p-norm
MART	0.753 +/-0.043	0.582 +/-0.044	0.755 +/-0.043	0.783 +/-0.024
deviantArt	0.763 +/-0.038	0.674 +/-0.038	0.776 +/-0.038	0.801 +/-0.028

Table 6.1: Comparison of classification accuracies considering only *title*. (All paintings are considered)

	Visual	Description	Late Fusion	Schatten p-norm
MART	0.664 +/-0.035	0.594 +/-0.024	0.666 +/-0.037	0.701 +/-0.031
deviantArt	0.717 +/-0.024	0.683 +/-0.023	0.721 +/-0.021	0.742 +/-0.019

Table 6.2: Comparison of classification accuracies considering only *description*. (158 paintings with description are considered)

p-norm model compared with the late fusion strategy, which shows the effectiveness of our proposed method.

With these results, we investigated in which cases the title helps more the emotional classification of abstract paintings. The textual features increase or even correct the classification results mostly in paintings displaying mixed visual cues. For instance, as observed in Sections 3.3.3 and 3.3.6, light colors (such as, red, white, yellow, pink, green) and smooth shapes usually generate a positive emotion, whereas dark colors (such as, brown, black, grey) and sharp edges usually generate a negative emotion. The paintings with both light and dark colors, and/or smooth and sharp shapes are harder to classify. In such cases using titles and descriptions is expected to increase the classification accuracy.

	Visual	Title + Description	Late Fusion	Schatten p-norm
MART	0.664 +/-0.031	0.474 +/-0.033	0.664 +/-0.023	0.697 +/-0.022
deviantArt	0.717 +/-0.029	0.658 +/-0.031	0.725 +/-0.033	0.751 +/-0.021

Table 6.3: Comparison of classification accuracies considering *title* + *description*. (158 paintings with description are considered)

The Figure 6.2 provides example of paintings together with their respective titles from each dataset. These paintings (from MART dataset on the left, dA dataset on the right) are classified as positive when we use only the visual features for classification. However, they are classified as negative when we use only the title information. When we apply the late fusion of visual and textual features, the results are consistent with the ground truth, classifying these paintings as negative.

We also observe that both datasets, the paintings for which the descriptions helped the classification are mostly considered positive by people. Examples of such painting from each dataset are provided in Figure 6.3. Both paintings, as they are composed by

Figure 6.2: Two examples where descriptions helped the classification algorithm. The painting on the left is from MART dataset: Spazio inquieto (Restless Space), 1953. Emilio Vedova (©*MART - Archivio fotografico e Mediateca*). The description was extracted from: http://www.gagosian.com/exhibitions/september-20-2008--cecily-brown. The painting on the right is from deviantArt dataset: Conflict, 2013. Georgiana Beligan, http://phylactos.deviantart.com/ (*Courtesy of deviantArt.*)

dark colours, i.e., dark green and dark blue, are considered as negative when using only visual features. However, when the descriptions are combined with the visual features the outcome is positive, which is consistent with human annotation.

By generally comparing both datasets, we observe that adding text as a feature is more effective for deviantArt, than for MART. One reason may be that the descriptions from deviantArt dataset are made by the artists themselves. In addition, only a few paintings are untitled on deviantArt dataset (only 1%). On the contrary, on MART dataset 10.4% of the paintings have no titles and the descriptions are usually made by others (i.e., art historians, art critics, etc.).

Finally, we also study the parameter sensitivity of the proposed method in Fig.6.4. Here, we fix $\lambda_1 = 1$ and analyze the regularization parameters λ_2 and p. As shown in Fig.6.4, we observe that the proposed method is more sensitive to p compared with λ_2, which confirms the importance of flexibility of our proposed Schatten p-norm model.

"I'm more interested in sublimation. I love the way Francis Bacon talked about the grin without the cat, the sensation without the boredom of its conveyance...I've always wanted to be able to convey figurative imagery in a kind of shorthand, to get it across in as direct a way as possible. I want there to be a human presence without having to depict it in full." Cecily Brown

Inspired in part by Mary Webb's Precious Bane (1924), the Sarn Mere paintings evoke an imaginary place, a lake where all manner of dark happenings transpire.

Just a feeling, just a paradox, in its small size it is an invaluable piece oft art for me.

Figure 6.3: Example where descriptions help on the classification. The painting on the top is from MART dataset: Study for Sarn Mere II, 2008. Cecily Brown (©*MART - Archivio fotografico e Mediateca*). The painting on the bottom is from deviantArt dataset: Rot (red), 2009. Sias Est, http://excymiir.deviantart.com (*Courtesy of deviantArt.*)

6.5 Discussions

In this study we investigate how the metadata associated with abstract paintings is an important element to automatically identify the positive and negative emotions evoked by these paintings. We have proposed a multimodal approach, in which we employed computer vision techniques and sentiment analysis to learn statistical patterns correlated to the valence emotions conveyed by abstract paintings. We propose two approaches to combine visual and textual features. The first is late fusion based on weighted linear combination and the second is a novel joint flexible Schatten p-norm model which can exploit the sharing patterns between visual and textual information for sentiment analysis of abstract paintings. In addition, we evaluate this method in the two datasets of abstract artworks described in Section 2.5, consisting of professional
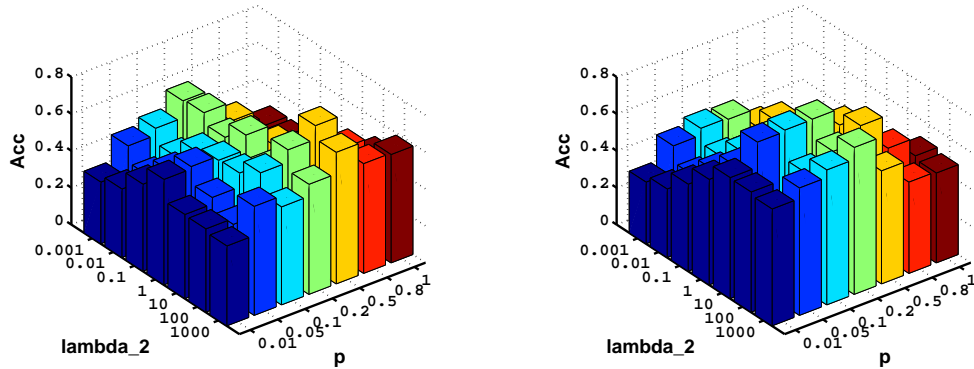
Figure 6.4: Sensitivity study of parameters on MART dataset (left) and deviantArt dataset (right).

and amateur paintings respectively. We observed that using textual features (i.e., titles, descriptions and/or artist statements) improves the accuracy in both approaches of combination.

A qualitative analysis showed that the combination of metadata associated with the paintings can help computers to efficiently recognize positive and negative emotions in the cases when the painting composition is mixed. For instance, when the painting is composed by dark colours (which conveys negative emotions) and light colours (which conveys positive emotions) at the same time.

Chapter 7

Conclusions

This research thesis aimed to understand what makes an abstract artwork emotional. We made an in-deep analysis of the effects of the low level features (i.e., colour, shapes and texture) in the positive and negative emotions evoked on abstract artworks. This analysis allows us to develop effective methods to predict emotional valence (positive and negative) of paintings with low semantic content.

Specifically, we conducted 2 different user studies to emotionally evaluate abstract paintings, Absolute and Relative Scale annotation, in two distinct collections of abstract paintings, a professional collection from the MART Museum and an amateur set from the deviantArt social network. The results showed that both annotation methods are efficient to evaluate positive and negative emotions on the abstract paintings. However, the Relative Scale is a more straightforward setting on the evaluation process and less time consuming. We applied advanced computer vision techniques to classify the emotional valence of these abstract paintings. By training a Bag-of-Visual-Words system and by backprojecting the classification evidence, we visualized which parts of the paintings conveyed positive and negative emotions. The results follow the long-known observations in art: (1) bright colours evoke positive emotions, dark colours tend to evoke negative emotions; (2) smooth lines are generally positive and chaotic texture is generally negative.

Analyzing the correlation between the localized emotional evidence and the eye fixations we were able to estimate where people prefer to look. Our results show that there is a positive attention bias when people look at abstract paintings, even in the non-coloured version of the paintings. Moreover, we investigated the saliency distri-

bution of these paintings and compare with corresponding eye fixations, showing that computational models of bottom-up attention are effective to predict emotional valence in abstract paintings. The studies of preferential biases and saliency distribution can be useful to enhance the comprehension between the viewer and the image, which can be applied in studies of aesthetics and interestingness of images.

We observed the importance to understand the specific aspects of each feature used by abstract artists to convey emotions. For this reason, we investigated the contribution of texture and colour combinations taken independently. With the use of inverse-Perlin parametrization method in a psychophysical model to map the paintings in emotional scales we observed that texture was especially influential in giving an impression of hardness and softness, i.e., the paintings with chaotic structure and granular patterns are ranked as hard, whereas paintings with geometrical structure, neat lines and a look of 'matte' finish are ranked as soft. The study on the contribution of texture confirmed that colour plays a greater role than texture in emotional assessment, yet the effect of texture was visible. Indeed, based on art theory, we explore the contributions of two-colours combinations on a more global framework. The results shows that colour combinations are an effective feature for emotional classification of abstract paintings. Moreover, this study opens opportunities to analyse other aspects related to colours in the artworks, such as the amount of colours, the position and the distances between colours. We plan to investigate these aspects in the future. Finally, we observed that combining textual with visual features, in our case the metadata (i.e., titles, descriptions and/or artist statement) associated with the paintings, can help computers to efficiently recognize positive and negative emotions in the cases when the painting composition is mixed, for instance, when the painting is composed by dark colours (which conveys negative emotions) and light colours (which conveys positive emotions) at the same time.

The localization of emotional parts in abstract paintings opens up a new toolbox for art historians, art critics and artists to revisit art theories with a new perspective. On a practical note, the knowledge of which painting evokes which emotion can support museum curators to improve the arrangement of the paintings in the museum according to the emotional experience they want to promote. This not only would improve the visiting experience, but also would generate a new way to engage and motivate visitors. Moreover, we believe that this work is a useful addition to various compu-

tational problems, such as image retrieval systems and user recommendation systems. For instance when curators need to deal with a with huge amount of paintings, as is the case with deviantArt that has about 220 million of individual pieces of art, they may need to filter out certain images or locate some relevant ones. In addition, this work could be useful to art consumers in searching for specific artworks (image retrieval) and for receiving recommendations based on the artworks they had previously liked (recommendation systems).

To further extend this work, we plan to do a user analysis for the emotional valence of MART and deviantArt datasets with their respective titles and descriptions in order to improve both textual and visual features. We also plan to make an eye-tracking study on the metadata associated with the painting, in order to see the gaze pattern between the abstract paintings and the textual information. Moreover, we aim to apply a domain specific lexicon extracted from a collection of abstract paintings to strengthen the textual features. Further, we plan to make an in-deep analysis on the composition of the paintings by employing the rules of colour harmony of Itten [1974] together with the geometrical principles of Kandinsky [1979] in order to improve our algorithms.

References

ARNHEIM, R. (2004). *Art and Visual Perception: A Psychology of the Creative Eye*. University of California Press. 12

ARTALE, A., MAGNINI, B. & STRAPPARAVA, C. (1997). Wordnet for italian and its use for lexical discrimination. In *AI* IA 97: Advances in Artificial Intelligence*, 346–356. 84

AVIV, V. (2014). What does the brain tell us about abstract art? *Frontiers in Human Neuroscience*, **8, Article 85**. 12

BACCIANELLA, A.E.S. & SEBASTIANI, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, vol. 10. 84

BALAHUR, A., HERMIDA, J.M., MONTOYO, A. & MUNOZ, R. (2011). Emotinet: a knowledge base for emotion detection in text built on the appraisal theories. In *Proceedings of the 16th international conference on Natural language processing and information systems*, NLDB'11, 27–39. 81

BERLYNE, D. (1971). *Aesthetics and psychobiology*. Century psychology series. 7

BLANK, P., MASSEY, C., GARDNER, H. & WINNER, E. (1984). Perceiving what paintings express. *Cognitive processes in the perception of art*, 127–143. 12

BRAINARD, D.H. (1997). The psychophysics toolbox. *Spatial Vision*, **10**. 46

BRUCE, N.D.B. & TSOTSOS, J.K. (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision*, **9**, 1–24. 47

BULLOT, N.J. & REBER, R. (2013). The artful mind meets art history: Toward a psycho-historical framework for the science of art appreciation. *Behavioral and Brain Sciences*, **36**, 123–137. 8

CALVO, R.A. & D'MELLO, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, **1**, 18–37. 7, 16

CANNON, W.B. (1927). The james-lange theory of emotions: A critical examination and an alternative theory. *The American Journal of Psychology*, **Vol. 39, No. 1/4**, 106–124. 7

CSÍKSZENTMIHÁLYI, M. & ROBINSON, R. (1990). *The Art of Seeing: An Interpretation of the Aesthetic Encounter*. Getty Trust Publications: Getty Education Institute for the Arts Series, J. Paul Getty Museum. 18

CSURKA, G., DANCE, C., FAN, L., WILLAMOWSKI, J. & BRAY, C. (2004). Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision*, vol. 1. 29, 30

CUPCHIK, G.C., VARTANIAN, O., CRAWLEY, A. & MIKULIS, D.J. (2009). Viewing artworks: Contributions of cognitive control and perceptual facilitation to aesthetic experience. *Brain and Cognition*, **70**, 84–91. 8

DAMASIO, A. (1999). *The Feeling of what Happens: Body and Emotion in the Making of Consciousness*. 7

DICKERMAN, L., AFFRON, M. & MUSEUM OF MODERN ART (NEW YORK, N. (2012). *Inventing Abstraction, 1910-1925: How a Radical Idea Changed Modern Art*. Museum of Modern Art. 9, 10, 13, 70

DORFLES, G. & VETTESE, A. (2001). *Arti visive. Protagonisti e movimenti. VOL.3A: Il Novecento.*. 11

DUDA, R., HART, P. & STORK, D. (2001). Pattern classification second edition john wiley & sons. *New York*. 58

EKMAN, P. (1993). Facial expression and emotion. *American Psychologist*, **48(4)**, 384–392. 17

FELZENSZWALB, P.F. & HUTTENLOCHER, D.P. (2004). Efficient graph-based image segmentation. *Int. J. Comput. Vision*, **59**, 167–181. 76

FRANKLIN, M.B., BECKLEN, R.C. & DOYLE, C.L. (1993). The influence of titles on how paintings are seen. *Leonardo*, **26, No. 2**, 103–108. 81

FRINTROP, S. (2011). *Computer Analysis of Human Behavior*, chap. Computational Visual Attention. Computer Analysis of Human Behavior, Advances in Pattern Recognition. Springer. 63

GONG, Y., KE, Q., ISARD, M. & LAZEBNIK, S. (2014). A multi-view embedding space for modeling internet images, tags, and their semantics. *IJCV*, **106**, 210–233. 82

GUERINI, M., GATTI, L. & TURCHI, M. (2013). Sentiment analysis: How to derive prior polarities from sentiwordnet. In *EMNLP*, 1259–1269. 84

HAGTVEDT, H., HAGTVEDT, R. & PATRICK, V. (2008). The perception and evaluation of visual art. *Empirical Studies of the Arts*, **26(2)**, 197–218. 8

HANJALIC, A. (2006). Extracting moods from pictures and sounds: towards truly personalized tv. *Signal Processing Magazine, IEEE*, **23**, 90 –100. 17

HERBRICH, R. & GRAEPEL, T. (2006). Trueskill(tm): A bayesian skill rating system. *no. MSR-TR-2006-80*. 19, 55

HRISTOVA, E., GEORGIEVA, S. & GRINBERG, M. (2011). Top-down influences on eye-movements during painting perception: the effect of task and titles. In *Proceedings of the Third COST 2102 international training school conference on Toward autonomous, adaptive, and context-aware multimodal interfaces: theoretical and practical issues*, 104–115, Springer-Verlag. 81

HUANG, G.B., ZHU, Q.Y. & SIEW, C.K. (2004). Extreme learning machine: a new learning scheme of feedforward neural networks. In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, vol. 2, 985–990, IEEE. 59

HUBEL, D.H. (1998). *Eye, brain, and vision*. Scientific American Library : Distributed by W.H. Freeman New York. 12

HWANG, S.J. & GRAUMAN, K. (2012). Learning the relative importance of objects from tagged images for retrieval and cross-modal search. *IJCV*, **100**, 134–153. 82

ISOLA, P., XIAO, J., TORRALBA, A. & OLIVA, A. (2011). What makes an image memorable? In *IEEE Conference on Computer Vision and Pattern Recognition*, 145–152. 28

ITTEN, J. (1974). *The Art of Color: The Subjective Experience and Objective Rationale of Color*. A VNR book, Wiley. 12, 56, 68, 70, 71, 74, 78, 79, 95

ITTI, L. & KOCH, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, **40**, 1489–1506. 63

ITTI, L., KOCH, C. & NIEBUR, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**. 64

IVANOVA, K., DOBREVA, M., STANCHEV, P. & TOTKOV, G., eds. (2012). *Access to Digital Cultural Heritage: Innovative Applications of Automated Metadata Generation.*, chap. APICAS - Content-Based Image Retrieval in Art Image Collections Utilizing Colour Semantics, 153–202. Univ. Publ. House Paisii Hilendarsk. 68

JAMES, W. (1884). *What is an emotion?*. Mind, 9, 188-205. 7

JIA, J., WU, S., WANG, X., HU, P., CAI, L. & TANG, J. (2012). Can we understand van gogh's mood?: Learning to infer affects from images in social networks. In *Proceedings of the 20th ACM International Conference on Multimedia*, 857–860, ACM. 28, 54

JOSHI, D., DATTA, R., FEDOROVSKAYA, E.A., LUONG, Q.T., WANG, J.Z., LI, J. & LUO, J. (2011). Aesthetics and emotions in images. *IEEE Signal Process. Mag.*, **28**, 94–115. 16, 28

KANDINSKY, W. (1914). *Concerning the Spiritual in Art*. Dover Books on Art History Series, Dover. 1, 9, 13, 69

KANDINSKY, W. (1979). *Point and Line to Plane*. Dover Publications. 53, 95

KIM, S.J., KIM, E.Y., JEONG, K. & KIM, J.I. (2006). Emotion-based textile indexing using colors, texture and patterns. In *Advances in Visual Computing*, 9–18. 53

KLEIN, D. & MANNING, C.D. (2003). Accurate unlexicalized parsing. In *41st Annual Meeting on Association for Computational Linguistics-Volume 1*, 423–430. 84

LANG, P.J., BRADLEY, M.M. & CUTHBERT, B.N. (1999). International affective picture system (iaps): Technical manual and affective ratings. 28

LE MEUR, O. & BACCINO, T. (2012). Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior Research Methods*, **45**, 251–266. 64

LEDER, H., CARBON, C.C. & RIPSAS, A.L. (2006). Entitling art: Influence of title information on understanding and appreciation of paintings. *Acta psychologica*, **121(2)**, 176–98. 81

LEDER, H., TINIO, P.P. & BAR, M. (2011). Emotional valence modulates the preference for curved objects. *Perception-London*, **40**, 649. 12

LEDER, H., GERGER, G., DRESSLER, S.G. & SCHABMANN, A. (2012). How art is appreciated. *Psychology of Aesthetics, Creativity, and the Arts.*, **Vol 6(1)**. 8, 9, 17

LEDER, H., GERGER, G., BRIEBER, D. & SCHWARZ, N. (2014). What makes an art expert? emotion and evaluation in art appreciation. *Cognition & Emotion*, **28**, 1137–1147. 8, 17, 36

LEUNG, T. & MALIK, J. (2001). Representing and recognizing the visual appearance of materials using three-dimensional textons. 29–44. 54

LI, B., XIONG, W., HU, W. & DING, X. (2012). Context-aware affective images classification based on bilayer sparse representation. In *Proceedings of the 20th ACM International Conference on Multimedia*, 721–724, ACM. 28

LIU, H., XU, M., HE, X. & WANG, J. (2014). Estimate gaze density by incorporating emotion. In *Proceedings of the ACM International Conference on Multimedia*, 1113–1116, ACM. 29

LIU, N., DELLANDREA, E., TELLEZ, B. & CHEN, L. (2011). Associating textual features with visual ones to improve affective image classification. In *International Conference on Affective Computing and Intelligent Interaction (ACII)*. 82

LIVINGSTONE, M. (2002). *Vision and art: the biology of seeing*. Harry N. Abrams. 69

LOCHER, P., KRUPINSKI, E.A., MELLO-THOMS, C. & NODINE, C.F. (2007). Visual interest in pictorial art during an aesthetic experience. *Spatial Vision*, **21**, 55–77. 47

LOWE, D.G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, **60**, 91–110. 30

LU, X., SURYANARAYAN, P., ADAMS JR, R., NEWMAN, J. & WANG, J. (2012). On shape and the computability of emotions. In *Proceedings of the ACM Multimedia Conference*. 17, 28

LUCASSEN, M.P., GEVERS, T. & GIJSENIJ, A. (2011). Texture affects color emotion. *Color Research & Application*, **36**, 426–436. 52, 56, 57, 58, 59, 60, 61, 65

MACHAJDIK, J. & HANBURY, A. (2010). Affective image classification using features inspired by psychology and art theory. In *Proceedings of the international conference on Multimedia*, 83–92, ACM. 28, 54, 72

MAJI, S., BERG, A.C. & MALIK, J. (2008). Classification using intersection kernel support vector machines is efficient. In *IEEE Conference on Computer Vision and Pattern Recognition*. 31, 65, 77, 85

MARCHESOTTI, L., PERRONNIN, F., LARLUS, D. & CSURKA, G. (2011). Assessing the aesthetic quality of photographs using generic image descriptors. In *International Conference on Computer Vision*. 28

MATHER, G. (2012). Aesthetic judgement of orientation in modern art. *i-Perception*, **3**, 18–24. 8

MCMANUS, I.C., STÖVER, K. & KIM, D. (2011). Arnheim's gestalt theory of visual balance: Examining the compositional structure of art photographs and abstract images. *i-Perception*, **2**, 615–647. 8

MELCHER, D. & BACCI, F. (2013). Chapter 10 - perception of emotion in abstract artworks: A multidisciplinary approach. In *The Fine Arts, Neurology, and Neuroscience New Discoveries and Changing Landscapes*, vol. 204 of *Progress in Brain Research*, 191 – 216, Elsevier. 12, 14

MILLER, G.A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, **38**, 39–41. 84

MILLIS, K. (2001). Making meaning brings pleasure: the influence of titles on aesthetic experiences. *Emotion*, **1**, 320. 81

MOHOLY-NAGY, L. (1945). In Defense of "Abstract" Art. *The Journal of Aesthetics and Art Criticism*, **4**, 74–76. 12

MOSER, J. (2010). True skil library. https://github.com/moserware/Skills/. 20, 55

MOWER, E., MATARIC, M.J. & NARAYANAN, S.S. (2011). A framework for automatic human emotion classification using emotion profiles. *IEEE Transactions on Audio, Speech & Language Processing*, **19**, 1057–1070. 17

NEWTON, I. (1704). *Opticks: Or, a Treatise of the Reflections, Refractions, Inflexions and Colours of Light. Also Two Treatises of the Species and Magnitude of Curvilinear Figures*. S. Smith, and B. Walford. 69

NICOLAOU, M.A., GUNES, H. & PANTIC, M. (2011). A multi-layer hybrid framework for dimensional emotion classification. In *Proceedings of the 19th ACM International Conference on Multimedia*, 933–936, ACM. 17

NIE, F., HUANG, H. & DING, C. (2012). Low-rank matrix recovery via efficient schatten p-norm minimization. In *AAAI*, 655–661. 86

OLIVA, A. & TORRALBA, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, **42**, 145–175. 34

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**, 232–240. 71

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**, 292–298. 71

PELOWSKI, M. & AKIBA, F. (2011). A model of art perception, evaluation and emotion in transformative aesthetic experience. *New Ideas in Psychology*, **29**, 80–97. 8

PERLIN, K. (1985). An image synthesizer. *ACM Siggraph Computer Graphics*, **19**, 287–296. 57

PHELPS, E.A. (2006). Emotion and cognition: Insights from studies of the human amygdala. *Annual Review of Psychology*, **57**, 27–53. 16

PIANTA, E., GIRARDI, C. & ZANOLI, R. (2008). The textpro tool suite. In *LREC*, European Language Resources Association. 84

PICARD, R.W. (1995). Affective computing. *Perceptual Computing Section, Media Laboratory, Massachusetts Institute of Technology*. 16

PIHKO, E., VIRTANEN, A., SAARINEN, V.M., PANNASCH, S., HIRVENKARI, L., TOSSAVAINEN, T., HAAPALA, A. & HARI, R. (2011). Experiencing art: the influence of expertise and painting abstraction level. *Frontiers in human neuroscience*, **5**. 17, 29

REINAGEL, P. & ZADOR, A. (1999). Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems*, **10**. 47

REWALD, S. (October 2004). Fauvism. *In Heilbrunn Timeline of Art History. New York: The Metropolitan Museum of Art, 2000. http://www.metmuseum.org/toah/hd/fauv/hd$_f$auv.htm.* 9

ROSS, C. (1990). *Abstract expressionism: creators and critics : an anthology.* Abrams. 9, 13

RUSSELL, J. (1980). A circumplex model of affect. *Journal of personality and social psychology*, **39**, 1161–1178. 17

SARTORI, A., SENYAZAR, B., SALAH, A.A., SEBE, N. & SALAH., A. (2015a). Emotions in abstract art: Does texture matter? In *International Conference on Image Analysis and Processing (ICIAP)*. 52

SARTORI, A., YAN, Y., OZBAL, G., SALAH, A.A., SALAH, A. & SEBE., N. (2015b). Looking at mondrian's victory boogie-woogie: What do i feel? In *International Joint Conference on Artificial Intelligence (IJCAI)*. 80

SARTORI, A., YANULEVSKAYA, V., SALAH, A.A., UIJLINGS, J., BRUNI, E. & SEBE., N. (2015c). Affective analysis of professional and amateurs abstract paintings using statistical analysis and art theory. *ACM Transactions on Interactive Intelligent Systems (TiiS)*. 6, 14, 27

SCHACHTER, S. & SINGER, J. (1962). *Psychological Review*, **69**, 379–399. 7

SCHERER, K. (2005). What are emotions? and how can they be measured? *Social Science Information*. 7

SEBE, N., COHEN, I. & HUANG, T.S. (2005). Multimodal emotion recognition. *Handbook of Pattern Recognition and Computer Vision*, 387–410. 16

SHROUT, P. & FLEISS, J. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, **86**, 420–428. 22

SILVIA, P.J. (2005). Emotional responses to art : From collation and arousal to cognition and emotion. *Review of general psychology*, **9**, 342–357. 8

SIMMONS, D.R. & RUSSELL, C. (2008). Visual texture affects the perceived unpleasantness of colours. *Perception*, **37**, 146–146. 53

SIVIC, J. & ZISSERMAN, A. (2003). Video Google: A text retrieval approach to object matching in videos. In *International Conference on Computer Vision*. 29

SOLLI, M. & LENZ, R. (2009). Color based bags-of-emotions. In *Proceedings of the 13th International Conference on Computer Analysis of Images and Patterns*, 573–580. 28

STRAPPARAVA, C. & MIHALCEA, R. (2008). Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on Applied computing*, 1556–1560, ACM. 81

STRAPPARAVA, C., GUERINI, M. & ZBAL, G. (2011). Persuasive language and virality in social networks. In *Proceedings of the 4th international conference on Affective computing and intelligent interaction*, ACII'11, 357–366, Springer-Verlag. 81

SUBRAMANIAN, R., KATTI, H., SEBE, N., KANKANHALLI, M. & CHUA, T.S. (2010). An eye-fixation database for saliency detection in images. *European Conference on Computer Vision*. 29

SUBRAMANIAN, R., YANULEVSKAYA, V. & SEBE, N. (2011). Can computers learn from humans to see better? Inferring scene semantics from viewers' eye movements. In *ACM Multimedia*, 33–42. 29

SZELISKI, R. (2010). Computer vision: Algorithms and applications. 30

TAKAHASHI, S. & EJIMA, Y. (2013). Contextual information processing of brain in art appreciation. *Behavioral and Brain Sciences*, **36**, 158–159. 12

TAMURA, H., MORI, S. & YAMAWAKI, T. (1978). Textural features corresponding to visual perception. *Systems, Man and Cybernetics, IEEE Transactions on*, **8**, 460–473. 54

TAN, C. & FERGUSON, S. (2014). The role of emotions in art evaluation. In L. Candy & S. Ferguson, eds., *Interactive Experience in the Digital Age*, Springer Series on Cultural Computing, 139–152, Springer International Publishing. 8

TATLER, B.W., BADDELEY, R.J. & GILCHRIST, I.D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, **45**. 47

THUMFART, S., JACOBS, R.H., LUGHOFER, E., EITZINGER, C., CORNELISSEN, F.W., GROISSBOECK, W. & RICHTER, R. (2011). Modeling human aesthetic perception of visual textures. *ACM Transactions on Applied Perception (TAP)*, **8**, 27. 54

TOMAS, V. (1969). Kandinsky's theory of painting. *The British Journal of Aesthetics*, **9**, 19–38. 12

TREISMAN, A.M. & GELADE, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, **12**. 64

UIJLINGS, J., SMEULDERS, A. & SCHA, R. (2012). The Visual Extent of an Object. *International Journal of Computer Vision*. 27, 31

UIJLINGS, J.R.R., SMEULDERS, A.W.M. & SCHA, R.J.H. (2010). Real-time visual concept classification. *IEEE Transactions on Multimedia*, **12**, 665–681. 29

VALDEZ, P. & MEHRABIAN, A. (1994). Effects of color on emotions. *Journal of Experimental Psychology: General*, **123**, 394. 71

VAN DE SANDE, K.E.A., GEVERS, T. & SNOEK, C.G.M. (2010). Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32**. 30

VAN DE WEIJER, J., SCHMID, C., VERBEEK, J. & LARLUS, D. (2009). Learning color names for real-world applications. *IEEE Transactions on Image Processing*, **18**, 1512–1523. 72, 74

VAN PAASSCHEN, J., ZAMBONI, E., BACCI, F. & MELCHER, D. (2014). Consistent emotions elicited by low-level visual features in abstract art. *Art & Perception, Issue 1-2*, **2**, 99 118. 12, 22, 23

VARMA, M. & ZISSERMAN, A. (2005). A statistical approach to texture classification from single images. *International Journal of Computer Vision*, **62**, 61–81. 52, 58, 59

VEDALDI, A. & FULKERSON, B. (2010). VLFeat – an open and portable library of computer vision algorithms. In *ACM Multimedia*. 31

VERONESI, L. (1968). *Luigi Veronesi. Con una antologia di scritti*. Torino. 13

VON GOETHE, J. & EASTLAKE, C. (1840). *Goethe's Theory of Colours*. J. Murray. 69

VUILLEUMIER, P. & HUANG, Y.M. (2009). Emotional attention uncovering the mechanisms of affective biases in perception. *Psychological Science*, **Vol. 18, No. 3.**, 148–152. 16

WALTHER, D. & KOCH, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, **19**, 1395–1407. 64

WANG, G., HOIEM, D. & FORSYTH, D. (2009). Building text features for object image classification. In *Conference on Computer Vision and Pattern Recognition, CVPR*, 1367–1374, IEEE. 82

WEI-NING, W., YING-LIN, Y. & SHENG-MING, J. (2006). Image retrieval by emotional semantics: A study of emotional space and feature extraction. In *Systems, Man and Cybernetics, 2006. SMC '06. IEEE International Conference on*, vol. 4. 28

WINNER, E., BLANK, P., MASSEY, C. & GARDNER, H. (1983). Childrens sensitivity to aesthetic properties of line drawings. In *The acquisition of symbolic skills*, 97–104, Springer. 12

WU, O., CHEN, Y., LI, B. & HU, W. (2011). Evaluating the visual quality of web pages using a computational aesthetic approach. In *Association of Computing Machinery conference on Web search and data mining*, 337–346. 28

YANULEVSKAYA, V., GEMERT, J.V., ROTH, K., HERBOLD, A., SEBE, N. & GEUSEBROEK, J. (2008). Emotional valence categorization using holistic image features. In *IEEE International Conference on Image Processing*, 101–104. 28, 54, 72

YANULEVSKAYA, V., MARSMAN, J.B., CORNELISSEN, F. & GEUSEBROEK, J.M. (2011). An image statistics based model for fixation prediction. *Cognitive Computation*, **3**. 47

YANULEVSKAYA, V., BRUNI, E., UIJLINGS, J.R., SARTORI, A., ZAMBONI, E., BACCI, F., MELCHER, D. & SEBE, N. (2012a). Automatic analysis of emotions conveyed by abstract painting. *1st Visual Science of Art Conference*. 6

YANULEVSKAYA, V., UIJLINGS, J., BRUNI, E., SARTORI, A., ZAMBONI, E., BACCI, F., MELCHER, D. & SEBE., N. (2012b). In the eye of the beholder: employing statistical analysis and eye tracking for analyzing abstract paintings. In *ACM Multimedia*. 6, 14, 27

ZEKI, S. (1992). The visual image in mind and brain. *Scientific American*, **267**, 68 – 76. 12

ZHAO, Q. & KOCH, C. (2011). Learning a saliency map using fixated locations in natural scenes. *Journal of vision*, **11**. 47

ZHAO, S., GAO, Y., JIANG, X., YAO, H., CHUA, T.S. & SUN, X. (2014). Exploring principles-of-art features for image emotion recognition. In *Proceedings of the ACM International Conference on Multimedia*, 47–56, ACM. 28, 54, 72