



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Affective Analysis of Professional and Amateur Abstract Paintings Using Statistical Analysis and Art Theory

### Citation for published version:

Sartori, A, Yanulevskaya, V, Salah, AA, Uijlings, J, Bruni, E & Sebe, N 2015, 'Affective Analysis of Professional and Amateur Abstract Paintings Using Statistical Analysis and Art Theory', *ACM Transactions on Interactive Intelligent Systems*, vol. 5, no. 2, 8. <https://doi.org/10.1145/2768209>

### Digital Object Identifier (DOI):

[10.1145/2768209](https://doi.org/10.1145/2768209)

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### Document Version:

Peer reviewed version

### Published In:

ACM Transactions on Interactive Intelligent Systems

### General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Affective Analysis of Professional and Amateur Abstract Paintings Using Statistical Analysis and Art Theory

ANDREZA SARTORI, DISI, University of Trento, Italy & Telecom Italia - SKIL, Trento, Italy

VICTORIA YANULEVSKAYA, DISI, University of Trento, Italy

ALMILA AKDAG SALAH, University of Amsterdam, the Netherlands

JASPER UIJLINGS, CALVIN Group, University of Edinburgh, United Kingdom

ELIA BRUNI, Free University of Bozen, Bolzano, Italy

NICU SEBE, DISI, University of Trento, Italy

When artists express their feelings through the artworks they create, it is believed that the resulting works transform into objects with 'emotions', capable of conveying the artists' mood to the audience. There is little to no dispute about this belief: Regardless of the artwork, genre, time and origin of creation, people from different backgrounds are able to read the emotional messages. This holds true even for the most abstract paintings. Could this idea be applied to machines as well, i.e. can machines learn what makes an artwork 'emotional'? In this work, we employ a state-of-the-art recognition system to learn which statistical patterns are associated with positive and negative emotions on two different datasets that comprise professional and amateur abstract artworks. Moreover, we analyse and compare two different annotation methods in order to establish the ground truth of positive and negative emotions in abstract art. Additionally, we use computer vision techniques to quantify which parts of a painting evoke positive and negative emotions. We also demonstrate how the quantification of evidence for positive and negative emotions can be used to predict which parts of a painting people prefer to focus on. This method opens new opportunities of research on why a specific painting is perceived as emotional at global and local scales.

Categories and Subject Descriptors: J.5 [Computer Applications]: Arts and Humanities - fine arts; I.2 [Artificial Intelligence]: Vision and Scene Understanding; H.3.1 [Information Search and Retrieval]: Content Analysis and Indexing

General Terms: Algorithms, Experimentation, Theory

Additional Key Words and Phrases: Visual Art, Abstract Paintings, Emotion Recognition, Eye tracking

## 1. INTRODUCTION

A known neurologist, Semir Zeki, claims that in some sense artists (of the 20<sup>th</sup> century) work like neurologists, since they study the brain "with techniques unique to them" [Zeki and Nash 1999]. Indeed, with the turn of the century and the break from the old, i.e. realistic/naturalistic traditions in art, a new approach arose in Western Art, where artists were concerned more and more with capturing the universals rather than the particulars. The most ardent movement in this approach was Abstract Art. Many abstract paintings evoke emotions purely through the basic elements, i.e., non-figurative, that compose a painting: colours, lines, shapes, and textures. Moholy-Nagy emphasized the importance of "visual fundamentals" over the realistic rendering of the world and advocated "the supremacy of colour over story" [Moholy-Nagy 1945].

---

Corresponding authors e-mail: andreza.sartori@disi.unitn.it; yanulevskaya@disi.unitn.it; alelma@ucla.edu; jrr.ujlings@ed.ac.uk; elia.bruni@unibz.it; sebe@disi.unitn.it.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© YYYY ACM 2160-6455/YYYY/01-ARTA \$15.00

DOI: <http://dx.doi.org/10.1145/0000000.0000000>

Kandinsky described the effect of colours on the spirit of the observer as “not a ‘subjective’ effect, but an objective one, determined by the characteristics of the colours and their interactions” [Veronesi 1968].

Abstract artists like Kandinsky, Moholy-Nagy, Klee, Mondrian and others were concerned with the question of what are the basic elements of visual art, and how these elements and their various compositions affect the viewer. They have extensively debated these questions, written about them, and tested their theories in their artworks. Hence, abstract art offers a good ground for someone who wants to study visual fundamentals and the emotions they generate in the viewer. Therefore, an important question arises: which part of a painting, i.e. which fundamentals, evoke what kind of emotion? In this paper we aim to answer this question using state-of-the-art computer vision techniques.

Recent research shows that computer vision has become mature enough to predict emotions [Machajdik and 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]. Surprisingly, both [Marchesotti *et al.* 2011] and [Isola *et al.* 2011] showed that the popular Bag-of-Visual-Words framework [Sivic and Zisserman 2003] performed better in predicting what is aesthetically pleasing than using art theory features such as the golden rule.

In this paper we explore why a painting evokes a certain emotion by determining where the classification evidence resides. To this effect, we train a Bag-of-Visual-Words (BoVWs) model to predict if a painting evokes positive or negative emotions. We use two different collections of abstract paintings: (1) a professional set compiled from 500 paintings of Museum of Modern and Contemporary Art of Trento and Rovereto’s (MART) collection, and (2) a set of 500 paintings collected from deviantArt, an online social network site devoted to user-generated art. These two sets, albeit coming from different backgrounds (i.e. professional and amateur), have one important thing in common: in all paintings, both types of artists use only basic visual stimuli such as forms and colour in their compositions. These visual stimuli are well captured by the BoVWs model. Furthermore, we use the method defined in [Uijlings *et al.* 2012] to dissect the classification machinery to understand which part of an image evokes what emotion. The quantification of evidence for positive and negative emotions opens up a new toolbox for both researchers and theoreticians of art. It provides more quantitative data and hence objective grounds in order to formulate and verify theories on art, aesthetics, and visualization.

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. Understanding human engagement in art helps us unravel the effect artworks have, as well as analyse the various kinds of visual creations from interface design to information visualization and score their effectiveness in terms of emotions they evoke.

We also collect data through two different annotation methods to set the ground truth of positive and negative emotions. We use an Absolute Scale method and we ask subjects to rank images on a Likert scale from 1-to-7. Moreover, we use a Relative Scale method, where the subjects were given pairs of paintings and were asked to choose the most positive one.

To summarize, the contributions of this work are the following: (1) we conducted a Relative Scale study to evaluate paintings from the emotional point of view and apply the TrueSkill method to optimally sample pairs of paintings; (2) we use computer vision techniques to quantify which parts of the image evoke positive and negative emotion; (3) we present an in-depth analysis investigating whether there is a positive attentional bias when people look at abstract paintings; (4) we test our approach using

two ways to annotate emotions, i.e. Absolute and Relative Scale; (5) we compare two different datasets, i.e. paintings from professional and amateur artists and finally (6) we propose a new method of analysis of artworks useful for art researchers and for visualization theories.

The rest of the paper is structured as follows: Section 2 presents previous studies on art theories as well as the existing emotion recognition approaches using computer vision techniques and relate these approaches to our work. Section 3 presents in details our proposed approach. In Section 4, we describe the two datasets of professional and amateur abstract paintings and also the Absolute and Relative Scale methods employed to annotate artworks as emotionally positive and negative. In Section 5, we demonstrate how scene analysis and machine learning techniques can be used to identify emotional parts from professional and amateur collections of abstract paintings. Section 6 presents an eye tracking study showing that, in general, people prefer to focus on the positive parts of the paintings. In Section 7 a comparison between the results obtained using our approach on each of the two datasets is presented. Our last Section is devoted to the final discussion and conclusions.

## 2. RELATED WORK

### 2.1. Art Theory

In a *Saturday Evening Post* article, E. H. Gombrich, wrote the story of modern art, musing on how shocking it was for the audience to witness the replacement of representational tradition in visual art with art movements reducing the art of drawing to basic shapes and colours. Their first reaction was to condemn the modernist painters with incompetence, but with time, Gombrich notes, that “outcry subsided. What remained was the conviction that the artist knew more about seeing than the layman” [Gombrich 1961]. Zeki, clearly follows this conviction when he states that modern artist had an intrinsic understanding of how visual brain functioned. In a way, while eliminating certain visual fundamentals and focusing only on certain elements, modern artists were “exploiting the characteristics of the parallel processing-perceptual systems of the brain to create their works, sometimes even restricting themselves largely or wholly to one system” [Zeki 1999]. In this sense, abstract art offers an exquisite dataset to demonstrate how artists have experimented with “art of painting in which only aesthetic elements seem to be present” [Schapiro 1937].

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 and Bacci [2012] 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 maybe movement) are also the building blocks of abstract art. Hence, it is no wonder that many studies show that subjects, even children, agree on the emotional and expressive nature of abstract shapes and colours [Winner et al. 1983; Blank et al. 1984; Leder et al. 2011; Takahashi and Ejima 2013].

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 integrating these 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 [Hagtvedt et al. 2008]. Pelowski and Akiba [2011] developed a five-stage model of art perception that relates art-viewing to the viewer’s



personality. McManus et al. [2011] used Arnheim's Gestalt theory of visual balance to examine the visual composition of art photographs and abstract images. Similar works which deal with perception include [Cupchik et al. 2009; Mather 2012] and emotional responses to art [Silvia 2005]. In this paper, we propose a novel method for analysing emotions in art: we train a computer vision system to predict emotions in paintings and then inspect in details its decision such that we can see exactly which parts of the image are responsible for which emotion. This provides a new angle to test and verify art theories.

## 2.2. Emotion Recognition

Emotion is an intuitive characteristic of human behaviour and an essential element of human communication and interactions. Humans can autonomously perceive and evaluate emotional signals of their visual surroundings. This becomes a difficult task for machines. Recent works that observe emotional responses invoked by pictures employ interdisciplinary information from art, aesthetics and psychology [Joshi et al. 2011]. These aspects are important in order to comprehend the interaction between features of the scene and the viewer, as well as the physical, social and historical contexts where the visual experience occurs.

The study of aesthetics and affect (emotion and feelings) related to art started to receive attention in the last years [Silvia 2005]. These studies try to understand how people perceive and feel while looking at the artworks. Damasio [1999] observes the difference between the terms feeling and emotion and affirms that these terms are usually interchangeable. He describes that emotions happen unconsciously. Instead feelings 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.<sup>1</sup>

In the recent years, computational techniques working on diverse problems related to aesthetics and emotions of artworks have been developed. The possibility of digitizing artworks from the museums and galleries and the improvement of tools to detect physiological human signals, increased the interest of the computational community in detecting emotional information. They recognize that emotion recognition allows improvements in human-computer interaction [Sebe et al. 2005], image retrieval [Joshi et al. 2011] and enhances computers' abilities to make decisions [Picard 1995].

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 and Hanbury [2010] combined low-level features with concepts from psy-

<sup>1</sup>We use the word 'emotion' 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 in the computational community. However, we are aware that these terms may have different meanings in other communities.

chology 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. 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 paper, 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. PROPOSED METHOD

In this work we use a state-of-the-art Bag-of-Visual-Words classification framework which follows [Uijlings et al. 2010] in order to learn the difference between positive and negative paintings. The method proposed in this paper is based on our previous work presented in [Yanulevskaya et al. 2012].

#### 3.1. 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 and 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. Descriptors

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

**3.2.1. 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 colour 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.

**3.2.2. 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 and Fulkerson 2010].

### 3.3. Classification

The extracted representations of 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.

### 3.4. 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

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

where  $\mathbf{x} = \{x_1, \dots, x_n\}$  is the vector to be classified,  $\mathbf{z}_j = \{z_{1j}, \dots, z_{nj}\}$  is the  $j$ -th support vector,  $\alpha_j$  is its corresponding positive weight,  $t_j \in \{-1, 1\}$  is its corresponding label,  $m$  is the number of support vectors, and  $\sum_{i=1}^n \min(x_i, z_{ij})$  is the Histogram Intersection kernel function. The outer sum of (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:

$$h(\mathbf{x}) = b + \sum_{i=1}^n \sum_{j=1}^m \alpha_j t_j \min(x_i, z_{ij}) \quad (2)$$

$$= b + \sum_{i=1}^n w_i. \quad (3)$$

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.

#### 4. GROUND TRUTH FOR EMOTION RECOGNITION EVOKED BY PROFESSIONAL AND AMATEUR ABSTRACT ARTWORKS

In this paper, we are interested in analysing whether abstract paintings convey positive or negative emotions. According to Leder et al. [2012] it is difficult to evaluate the effects of arousal in art appreciation. Moreover, “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 interrelations.” Csíkszentmihályi and Robinson [1990] point out that emotion is the most important factor of the aesthetic experience; people tend to associate artworks with emotional valence. Therefore, for annotating the abstract paintings we follow the standard procedure of [Silvia 2006; Leder et al. 2011; Pihko et al. 2011; Leder et al. 2012; Leder et al. 2014] and consider positive-negative valence, introduced by Russell [1980], in emotional judgments. By following the prominent psychological and neuropsychological studies in emotional preferences of artworks cited above, we can conclude that positive-negative emotional valence can disambiguate the affective states and is adequate to quantify emotional responses on abstract paintings.

However, labelling abstract paintings as positive/negative emotions is an arduous task. In Sections 4.2 and 4.3 we consider two different approaches to collect ‘ground truth’ emotions elicited by a collection of professional abstract paintings. We compare an ‘Absolute Scale’ annotation, in which people were asked to use the Likert scale to judge paintings presented one by one, with a ‘Relative Scale’ annotation, in which people were asked to choose the more positive painting from a pair of paintings. Moreover, in Section 4.4 we evaluate an amateur set of abstract paintings by using the Relative Scale annotation.

##### 4.1. MART Dataset: A collection of Professional Abstract Artworks

The Museum of Modern and Contemporary Art of Trento and Rovereto (MART) has a substantial archive of modern and contemporary art. The permanent collection of MART contains more than 20,000 artworks, including paintings, drawings, engravings, and sculptures. Out of this collection, with the guidance of an art historian, 500 images of abstract paintings were selected [Yanulevskaya et al. 2012; Melcher and Bacci 2012]. For the purposes of this study, we used the digital versions of the artworks, which reside in the electronic archive of MART. The selected artworks date from 1913 to 2008. The chosen paintings 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. These artists, especially Kandinsky, Albers and Klee, discussed and formulated some of the art theories on abstractionism. Their artworks are a reflection of their principles and theories on colour, lines, shapes and textures. We applied two empirical methods to evaluate the emotional response on this dataset: Absolute Scale and Relative Scale annotation.

##### 4.2. Absolute Scale Annotation

This Section describes the Absolute Scale approach to collect the ground truth of positive and negative emotions evoked by abstract artworks. This study was published in [Yanulevskaya et al. 2012; Melcher and Bacci 2012].

**4.2.1. Participants.** We recruited a total of 100 people 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.

**4.2.2. Scoring Procedure.** For the annotation, 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.” 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. 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 the participants were not using 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 its author.

For the experiment, the paintings have been randomly divided in five groups of 100 paintings each, and only one group was presented to a single subject. Each subject rated 100 paintings and each artwork received 20 judgments from 20 different subjects. From these 20 judgements per painting, we calculated the average score in order to obtain 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.

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  $\geq 0.460$ ).

In Figure 8, the first group of paintings on the left illustrates the distribution of paintings from the most negative (average score of 1.95) to the most positive (average score of 6.2). This distribution of artworks was composed by averaging the scores given by all participants. The distribution showed in Figure 8 illustrates that people follow some trends in categorizing abstract paintings emotionally. The images and the emotional scores have been made available online.<sup>2</sup>

### 4.3. Relative Scale Annotation

In this Section we describe the Relative Scale approach to collect the ground truth of positive and negative emotions for MART dataset. This time, subjects were shown paintings in pairs and were asked to choose the more positive painting in any given pair. The detailed description of this annotation method is described below.

**4.3.1. Participants.** There were 25 subjects participating in the experiment (11 females and 14 males). Each person annotated from 19 to 356 pairs of paintings, 145 paintings on average. The participants were anonymous and they participated in our experiment voluntarily, without getting any reward. The annotation was done online. There was no

<sup>2</sup><http://disi.unitn.it/~sartori/datasets/mart-dataset/>

time-limit to the annotation procedure: each participant was free to annotate whenever he/she had time to do it.

**4.3.2. Scoring Procedure.** 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.”

If we compare each painting in the dataset with all the 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 et al. 2007] 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 matching 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 et al. 2007] 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.

#### 4.4. 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.

On this dataset, we apply the Relative Scale annotation to collect the emotional judgments from people. We compare the differences between the professional MART and the amateur dA datasets in Section 7, both in terms of the types of paintings and their emotional contents. Below we detail the selection process of the abstract paintings in order to have a dataset of Amateur paintings.

**4.4.1. deviantArt Dataset Selection.** The categorization system of the deviantArt makes it possible to crawl artworks in a specific genre only. For the purpose of our study, we needed to collect amateur artworks of Abstract Art. Thus, we have crawled the dA site for images that are shared under the category Traditional Art/Paintings/Abstract Art, and downloaded initially 8,000 artworks. Out of these artworks, we have manually eliminated the artworks that have recognizable objects. To further downsize our data to a manageable number of 500, we made use of additional information from each image that is available on dA website.

For each abstract painting downloaded, we had all the statistics of the site: the number of pageviews, the number of comments and the number of favourites. Favourites are a 'like' system in dA: every member can add a deviation to her 'favorites' by hitting a 'plus' button. Once a deviation is favoured, they are added to the homepage of the deviant who favours 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/favourites are more successful and popular in the dA community. Furthermore, this so called 'power' status makes some members become the more influential in this community: if a popular member favours a deviation, this act will probably direct some traffic to this artwork increasing its pageviews and other statistics.

First we postulated that there is a correlation between the number of pageviews, comments and favourites, and these numbers are affected by the time of upload of each deviation. In other words, we thought that the older a deviation is, the more comments, favourites and pageviews it has received. However, when we checked the correlation between the number of pageviews and the time of upload, we could not confirm our hypothesis. Instead, we found that only the number of comments and favourites are correlated. 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 he/she also favours the deviation. In contrast, pageviews show how many people have visited/seen the deviation. 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 he/she 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 favourites, and select the most of favourite paintings, the least favourite and randomly in the middle. Some of these 500 chosen paintings have the same author. In total, there are 406 different authors for the 500 paintings.

**4.4.2. Participants.** For the annotation of deviantArt dataset 60 people participated, 27 females and 33 males. Each participant annotated from 2 to 436 pairs of paintings, 63 paintings on average. The participants were anonymous and agreed to be part of our experiment voluntarily, without getting any reward.

**4.4.3. Scoring Procedure.** To annotate the deviantArt dataset, we follow the Relative Scale approach described in details in Section 4.2.2. We used the Microsoft TrueSkill algorithm [Herbrich et al. 2007] to generate pairs of paintings where each painting was compared with other 15 paintings. The annotation was done online. Each participant was given the option to participate to multiple sessions, and there was no time-limit set on any session.

## 5. EMOTION RECOGNITION FOR PROFESSIONAL AND AMATEUR ABSTRACT PAINTINGS

We apply our classification method on the two datasets of abstract paintings introduced in Section 4. 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.

### 5.1. 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 are defined 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 8 visualizes the difference between (1) Absolute and (2) Relative 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 8 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 I. In order to analyse the contribution of each type of visual words we evaluate separately 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.

Table I. Classification accuracy comparison.

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

Overall, the proposed method performs 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 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.

*5.1.1. 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 3.4 and backproject the relative contribution which comes from each pixel in the painting. In the Figures 1, 2, 3 and 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 1 and 2 respectively. Two paintings, the second and the third from the left of Absolute Scale (Figure 1) are also represented as highly positive for Relative Scale annotation (first and third painting from the left of Figure 2). The 4 paintings classified as highly negative for both methods are shown in Figures 3 and 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 1 and 2, as well as in Figures 3 and 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 reach agreement between the computational classification and humans’ judgements come from the Geometrical Abstraction Style. The Geometrical Abstraction Style 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, 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.

Moreover, 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 ex-

perceived 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 paintings 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.

## 5.2. deviantArt: Classification Results of Relative Scale

In order to construct the ground truth of deviantArt, we followed the analysis of Relative Scale explained in Section 5.1 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 the 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 paintings in the deviantArt dataset in positive and negative emotions, we apply the same classification method described in Section 5.1. We tested our model using a two-fold cross-validation that was repeated 1,000 times. We evaluated 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 II.

Table II. Classification accuracy.

|          | LAB   | SIFT  | LAB+SIFT |
|----------|-------|-------|----------|
| Accuracy | 0.769 | 0.749 | 0.770    |

We observe that also in the case of the deviantArt dataset, LAB visual words are more effective than SIFT visual words in 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.

*5.2.1. Backprojection.* We used the backprojection technique, described in Section 3.4, 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 5 and 6 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 5 and 6. 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.

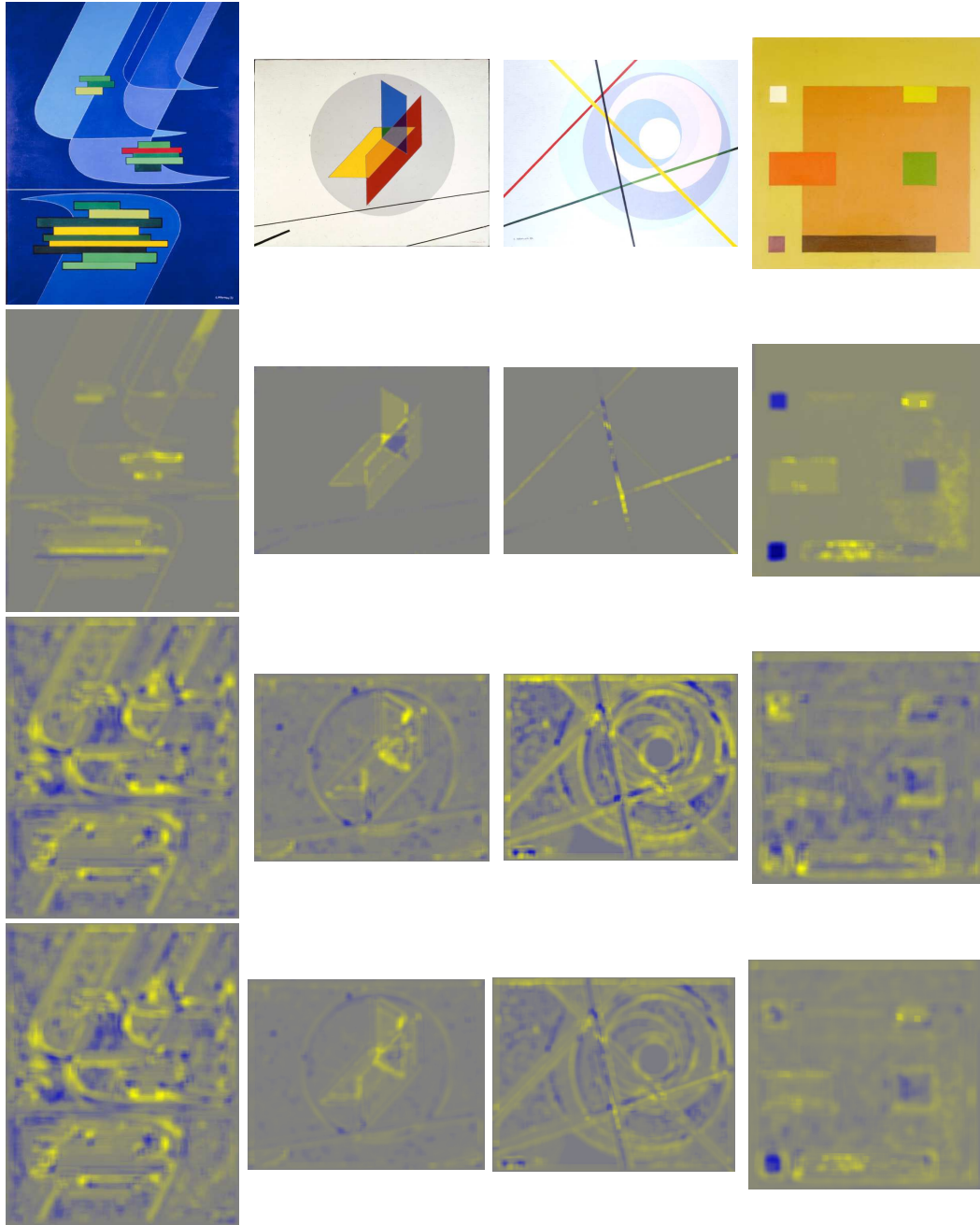


Fig. 1. Visualisations of pixel-wise contributions to the classification of highly positive paintings using Absolute Scale of MART dataset. The original paintings are shown in the first row. The second row displays the contribution of LAB visual words, followed by the contribution of SIFT visual words 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. (Courtesy of MART photographic archive, Rovereto.)

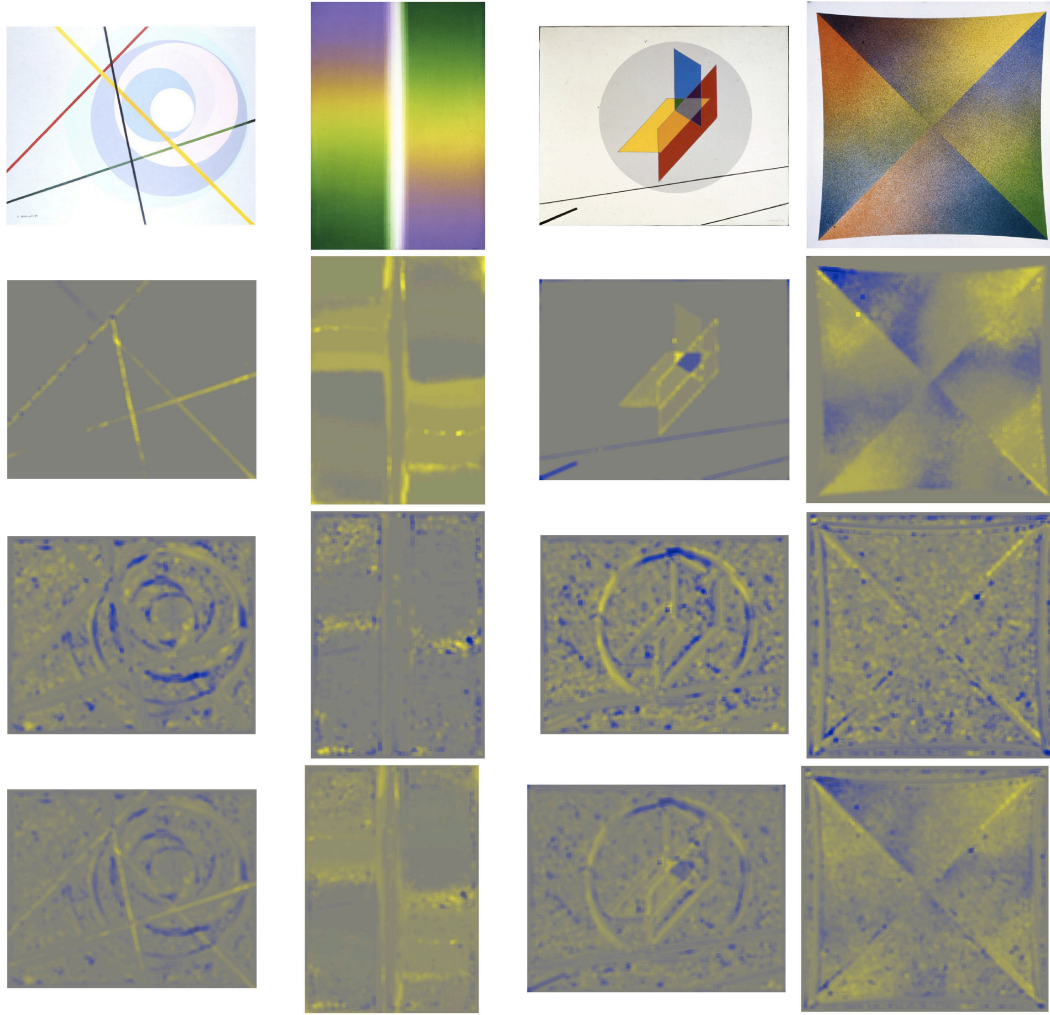


Fig. 2. Visualisations of pixel-wise contributions to the classification of highly positive paintings using Relative Scale of MART dataset. The original paintings are shown in the first row. The second row displays the contribution of LAB visual words, followed by the contribution of SIFT visual words 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. (Courtesy of MART photographic archive, Rovereto.)

By comparing 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.



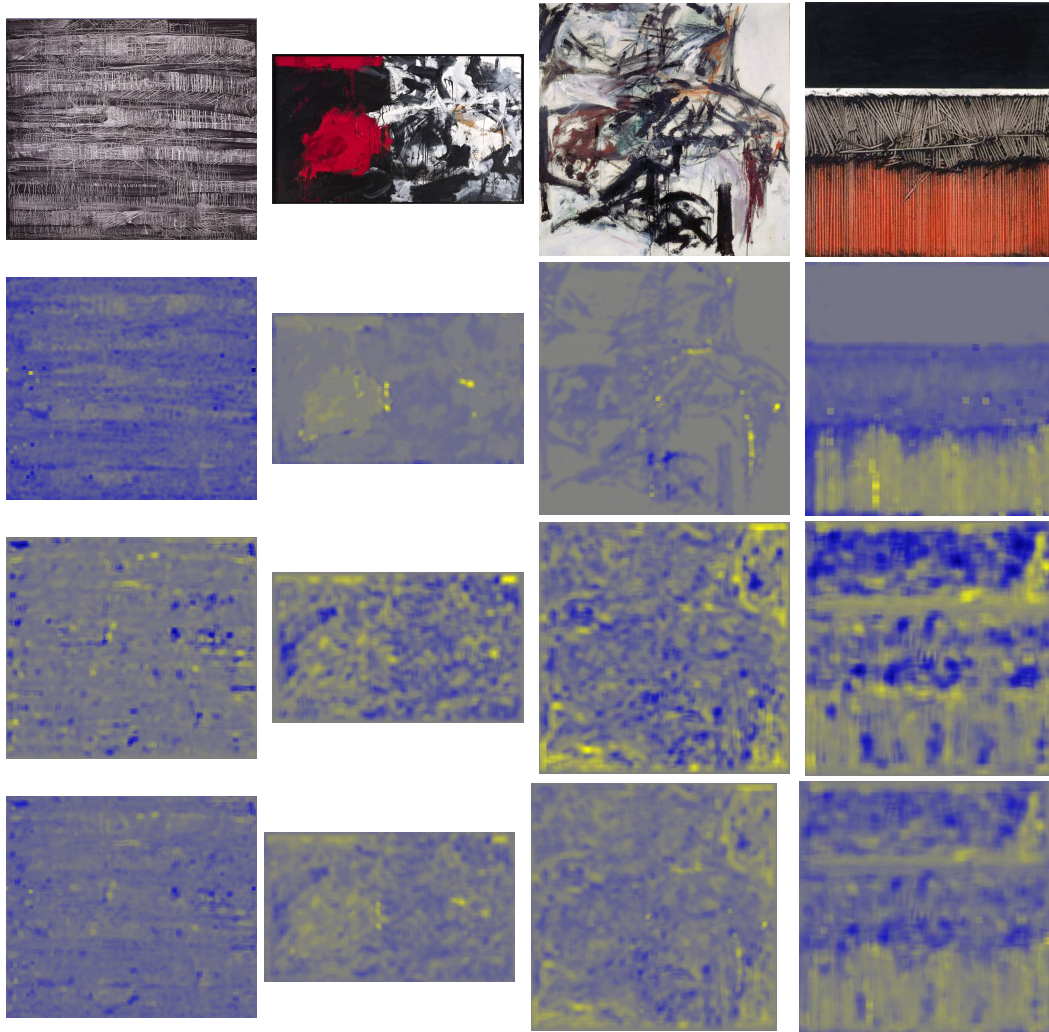


Fig. 3. Visualisations of pixel-wise contributions to the classification of highly negative paintings using Absolute Scale of MART dataset. The original paintings are shown in the first row. The second row displays the contribution of LAB visual words, followed by the contribution of SIFT visual words 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. (Courtesy of *MART photographic archive, Rovereto.*)

## 6. 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.

We recorded eye movements of 9 observers while they were looking at the artworks from the MART dataset. In order to ensure variability of emotional content, in this experiment we selected paintings with positive, negative, and neutral parts at the same

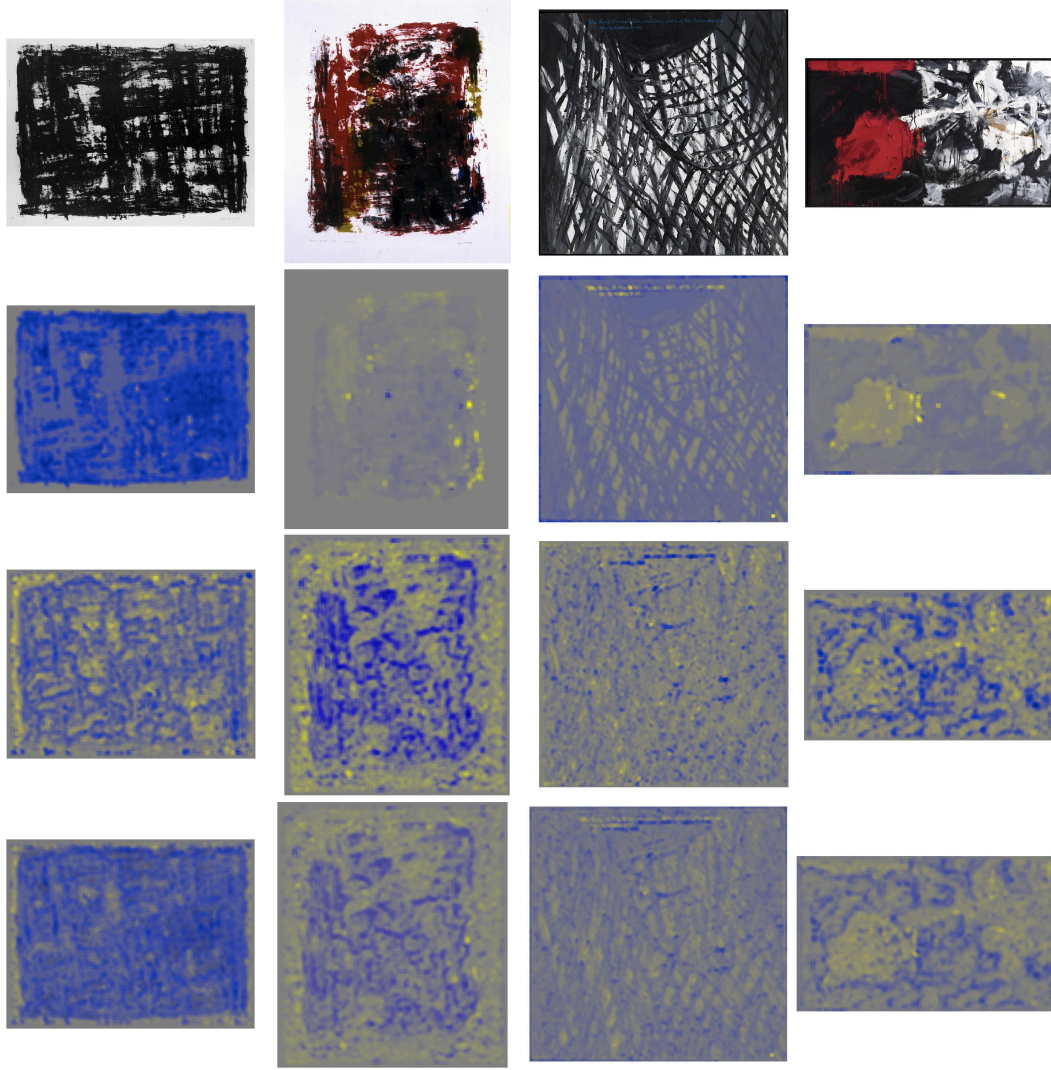


Fig. 4. Visualisations of pixel-wise contributions to the classification of highly negative paintings using Relative Scale of MART dataset. The original paintings are shown in the first row. The second row displays the contribution of LAB visual words, followed by the contribution of SIFT visual words 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. (Courtesy of MART photographic archive, Rovereto.)

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.



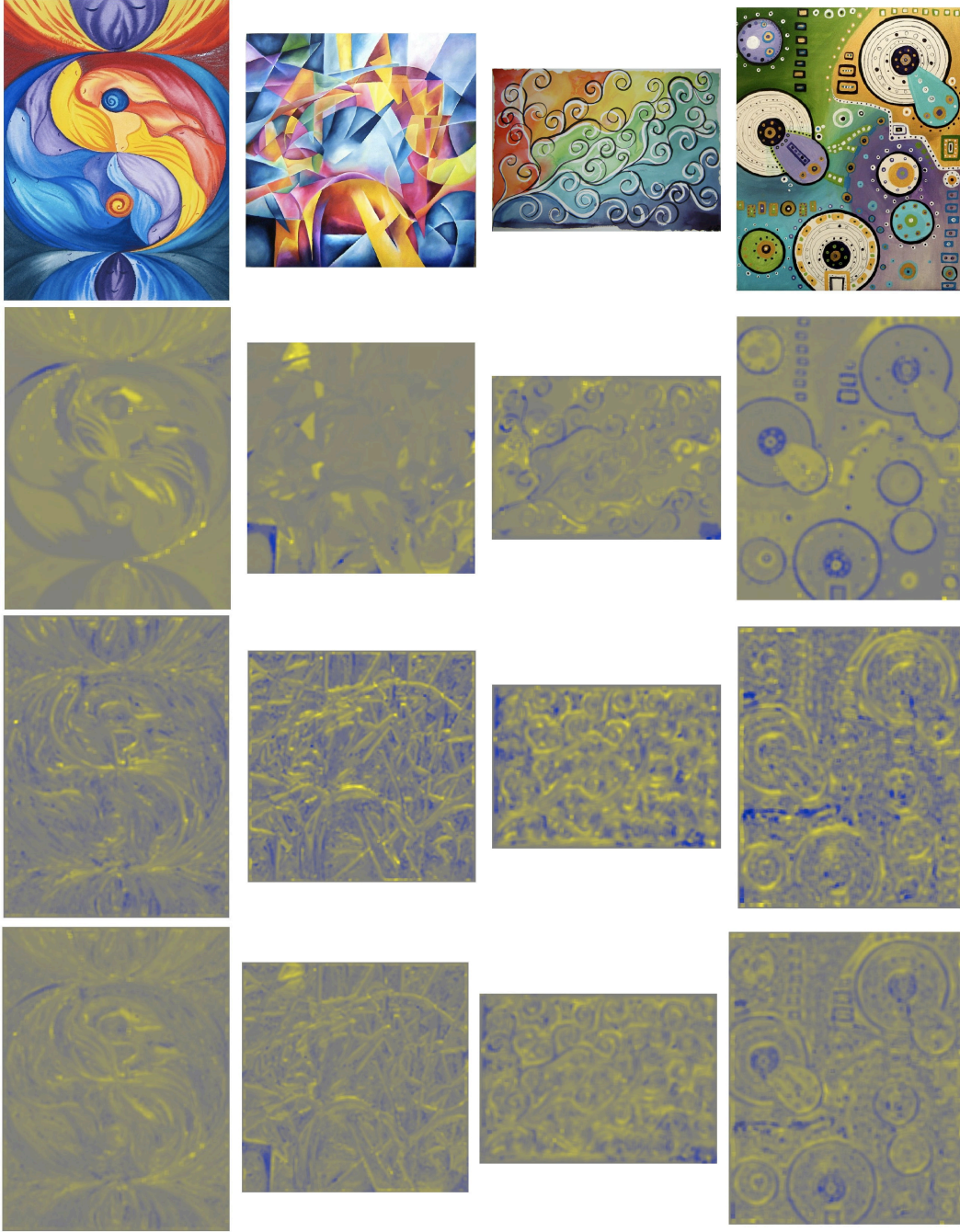


Fig. 5. Visualisations of pixel-wise contributions to the classification of highly positive paintings of deviantArt dataset. The original paintings are shown in the first row. The second row displays the contribution of LAB visual words, followed by the contribution of SIFT visual words in row 3, and 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: <http://oceanpeace.deviantart.com>; <http://coconut-artist.deviantart.com>; <http://klbailey.deviantart.com>; <http://numb-synapse.deviantart.com>.)

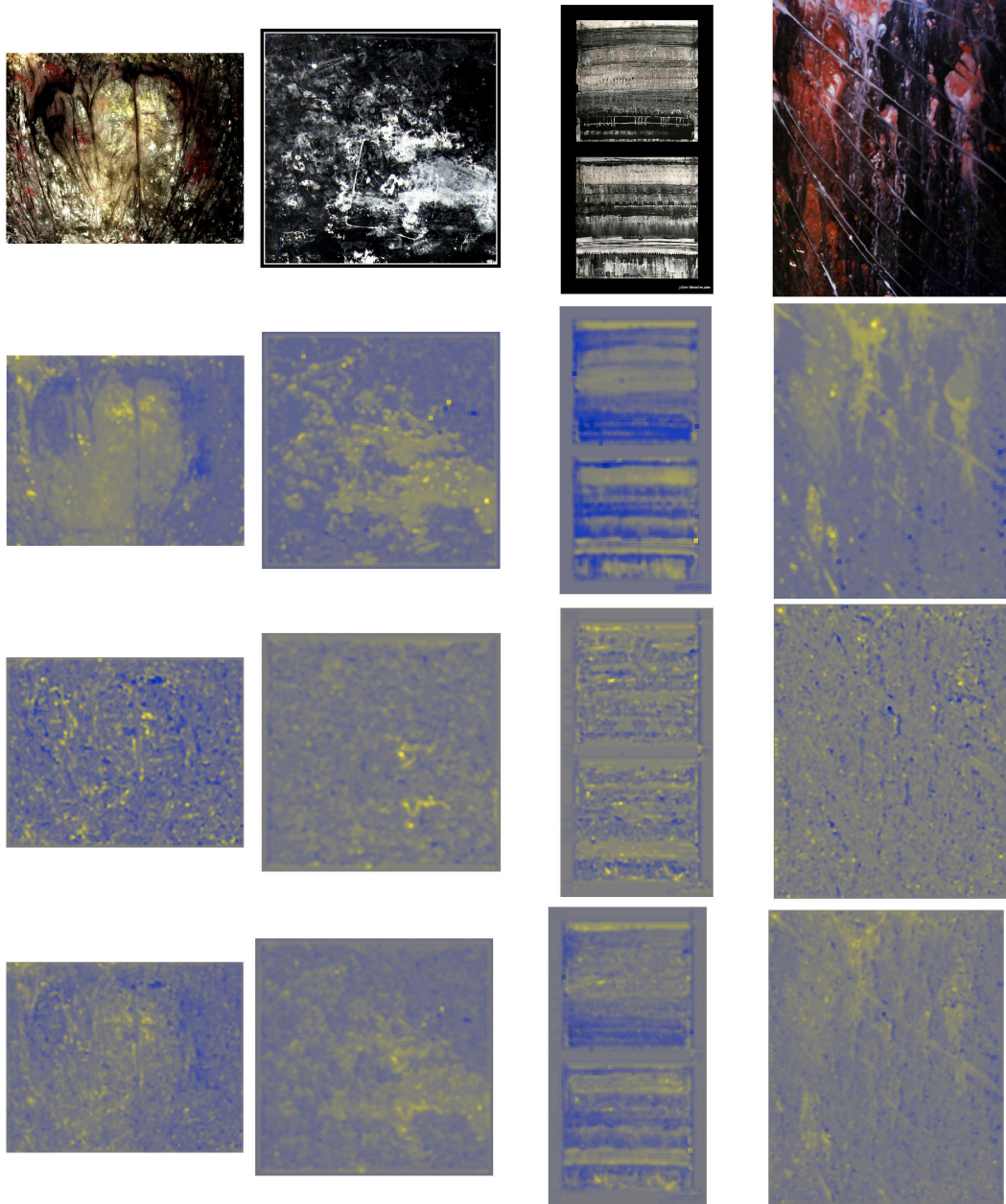


Fig. 6. Visualisations of pixel-wise contributions to the classification of highly negative paintings of deviantArt dataset. The original paintings are shown in the first row. The second row displays the contribution of LAB visual words, followed by the contribution of SIFT visual words in row 3, and 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: <http://diceglia.deviantart.com>; <http://epaf.deviantart.com>; <http://volute.deviantart.com>; <http://satiricmilk.deviantart.com>.)



## 6.1. 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  $0.36^\circ$ , with a standard deviation of  $0.07^\circ$ , where one visual angle  $1^\circ$  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.

*6.1.1. Participants.* Nine participants took part in the experiment, 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.

## 6.2. 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 and Tsotsos 2009; Tatler et al. 2005; Yanulevskaya et al. 2011; Zhao and Koch 2011]. In this work 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. Particularly, 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 ( $1^\circ$ , 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 and Zador 1999; Tatler et al. 2005] and sampled from the fixated locations recorded while viewing another 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  $1^\circ$  (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**) to be 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

Table III. Distribution of paintings according to “emotions” of the attentional bias for Absolute and Relative Scale ground-truth.

|                       | LAB+SIFT      | #          | PAB      | NAB      | Neutral |
|-----------------------|---------------|------------|----------|----------|---------|
| <b>Absolute Scale</b> | 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 |
| <b>Relative Scale</b> | 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 (0%)  |

on emotionally positive visual information, we calculate how many abstract paintings evoke PAB. To backproject emotional evidence, we use our best performing classifier according to Table I which combines LAB and SIFT descriptors. Table III 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 7 illustrates some examples of paintings together with the recorded fixated locations.

## 7. AMATEUR VERSUS PROFESSIONAL ABSTRACT PAINTINGS

Several psychological approaches attempt to explain art appreciation, human perception, aesthetics judgments and emotions concerning artworks. A recent psychological research [Leder et al. 2012] analyses how art is appreciated based on several variables including: the level of expertise of the participants in art, the different classes of artworks (Abstract, Modern and Classic), the measure of emotion (valence and arousal) and the artwork comprehension. Leder et al. [2012] conclude that emotion is the strongest predictor of art appreciation, independent of art styles and the expertise of people. Based on empirical studies, the authors demonstrated the importance of emotions in the analysis of artworks.

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). Figure 8 shows both collections as a montage arranged by the annotation scores each painting has received from subjects. In the case of MART dataset, we have two different orderings of paintings: one according to Absolute Scale scores (left column), and one according to Relative Scale scores (middle column). Since the deviantArt dataset is annotated with the Rel-

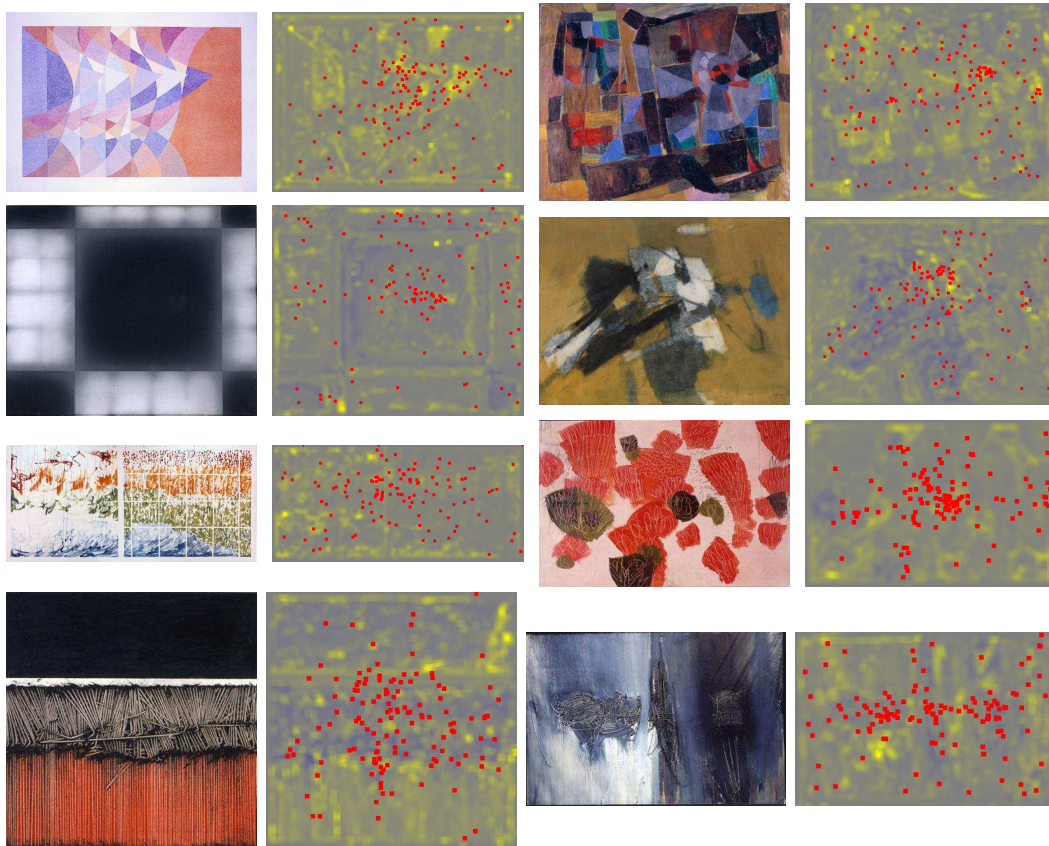


Fig. 7. Paintings with superimposed fixated locations: the first row displays positive paintings with PAB, the second row displays negative paintings with PAB, the third row displays positive paintings with NAB, and the last row displays negative paintings with NAB. (Courtesy of MART photographic archive, Rovereto.)

ative Scale method, only the ordering obtained with this method is displayed (right 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  $50 \times 10$ , 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 8 gives an insight into the different rankings obtained by using the Absolute and Relative Scale annotations. A comparison between MART and deviantArt datasets is imminent: MART dataset is compiled from paintings with extreme distribution of colour and saturation, whereas deviantArt paintings in general have an even distribution of colour and saturation. MART dataset offers many different types of abstract art examples that range from very dark-colored paintings to very light-colored paintings. In contrast, deviantArt 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 20<sup>th</sup> Century, whereas the paintings from deviantArt are all recent creations. The second reason is the impact of digital tools. Even though our deviantArt dataset

has paintings 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 by separately producing digitally created artworks.

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 accuracy 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 8, 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.

## 8. CONCLUSIONS

This paper proposes 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 on two different collections of abstract paintings: one dataset being composed of professional paintings obtained from the digital collection of MART Museum and a second dataset composed of amateur paintings obtained from deviantArt, a user-generated art website. We employed two empirical methods of annotation to evaluate positive and negative emotions evoked by abstract paintings: Absolute and Relative Scale. To conduct the Relative Scale method, we used the TrueSkill ranking system to sample the paintings pairs in an optimal way. We used both annotations into our classification model, and we obtained higher accuracy when the ground truth was created using the Relative Scale 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 paintings 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

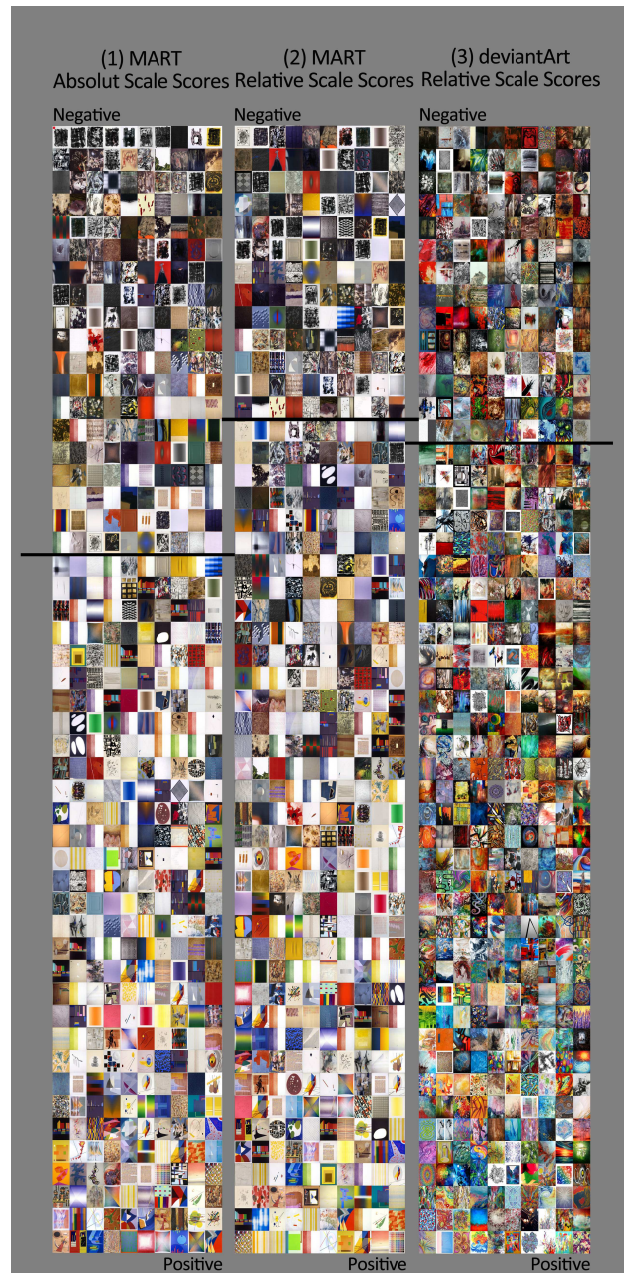


Fig. 8. Paintings from MART and deviantArt datasets ordered from most negative to most positive: (1) MART-Absolute Scale, (2) MART-Relative Scale (*Courtesy of MART photographic archive, Rovereto,*) and (3) deviantArt - Relative Scale. *Courtesy of deviantArt.*)



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 to identify the evidence for positive and negative emotions can be applied to the study of attention by analysing the salient distribution of the paintings. We are planning to do this in the future work.

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. From a practical point of view, knowing whether a painting evokes positive or negative emotions 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 computational problems, such as image retrieval systems and user recommendation systems. For instance when curators need to deal with a huge amount of paintings, as it 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 images. In addition, this work could be useful to art consumers when searching for specific artworks (image retrieval) and for receiving recommendations based on the artworks they had previously liked (recommendation systems).

## ACKNOWLEDGMENTS

The authors would like to thank Francesca Bacci, David Melcher and Elisa Zamboni for providing information on the MART Dataset. We would like to thank also the MART Museum (Museum of Modern and Contemporary Art of Trento and Rovereto) which provided us with their collection of abstract paintings.

This work was supported by the Italian Ministry of University and Research (MIUR) through the “Active Ageing at Home” and the FIRB S-PATTERNS projects.

## REFERENCES

- D. E. Berlyne. 1971. *Aesthetics and psychobiology*.
- P. Blank, C. Massey, H. Gardner, and E. Winner. 1984. Perceiving what paintings express. *Cognitive Processes in the Perception of Art* (1984), 127–143.
- D. H. Brainard. 1997. The psychophysics toolbox. *Spatial Vision* 10, 4 (1997).
- N. Bruce and J. Tsotsos. 2009. Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision* 9, 3 (2009).
- M. Csikszentmihályi and R.E. Robinson. 1990. *The Art of Seeing: An Interpretation of the Aesthetic Encounter*. J. Paul Getty Museum.
- G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. 2004. Visual categorization with bags of keypoints. In *Workshop Statistical Learning in Computer Vision, ECCV*, Vol. 1. 1–2.
- G. C. Cupchik, O. Vartanian, A. Crawley, and D. J. Mikulis. 2009. Viewing artworks: Contributions of cognitive control and perceptual facilitation to aesthetic experience. *Brain and Cognition* 70 (2009), 84–91. Issue 1.
- A. R. Damasio. 1999. *The Feeling of what Happens: Body and Emotion in the Making of Consciousness*.
- E. H. Gombrich. 1961. *How to Read a Painting*. (1961).
- H. Hagtvædt, R. Hagtvædt, and V. Patrick. 2008. The Perception and Evaluation of Visual Art. *Empirical Studies of the Arts* 26(2) (2008), 197–218.
- R. Herbrich, T. Minka, and T. Graepel. 2007. TrueSkill(TM): A Bayesian Skill Rating System. (2007), 569–576.
- P. Isola, J. Xiao, A. Torralba, and A. Oliva. 2011. What makes an image memorable?. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 145–152.

- D. Joshi, R. Datta, E. A. Fedorovskaya, Quang-Tuan Luong, J. Z. Wang, Jia Li, and Jiebo Luo. 2011. Aesthetics and Emotions in Images. *IEEE Signal Processing Magazine* 28, 5 (2011), 94–115.
- P. J. Lang, M. M. Bradley, and B. N. Cuthbert. 1999. International affective picture system (IAPS): Technical manual and affective ratings. (1999).
- H. Leder, G. Gerger, D. Brieber, and N. Schwarz. 2014. What makes an art expert? Emotion and evaluation in art appreciation. *Cognition & Emotion* 28, 6 (2014), 1137–1147.
- H. Leder, G. Gerger, S. G. Dressler, and A. Schabmann. 2012. How art is appreciated. *Psychology of Aesthetics, Creativity, and the Arts*. Vol 6(1) (2012).
- H. Leder, P.P. Tinio, and M. Bar. 2011. Emotional valence modulates the preference for curved objects. *Perception* 40(6):649-55 (2011).
- H. Liu, Min Xu, X. He, and J. Wang. 2014. Estimate Gaze Density by Incorporating Emotion. In *Proceedings of the ACM International Conference on Multimedia*. 1113–1116.
- P. Locher, E. A. Krupinski, C. Mello-Thoms, and C. F. Nodine. 2007. Visual interest in pictorial art during an aesthetic experience. *Spatial Vision* 21, 1-2 (2007), 55–77.
- D. G. Lowe. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60, 2 (2004), 91–110.
- X. Lu, P. Suryanarayan, R.B. Adams Jr, J.L.M.G. Newman, and J.Z. Wang. 2012. On Shape and the Computability of Emotions. In *Proceedings of the ACM International Conference on Multimedia*. 229–238.
- J. Machajdik and A. Hanbury. 2010. Affective image classification using features inspired by psychology and art theory. In *Proceedings of the ACM International Conference on Multimedia*. 83–92.
- S. Maji, A. C. Berg, and J. Malik. 2008. Classification using Intersection Kernel Support Vector Machines is Efficient. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1–8.
- L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka. 2011. Assessing the aesthetic quality of photographs using generic image descriptors. In *IEEE International Conference on Computer Vision (ICCV)*. 1784–1791.
- G. Mather. 2012. Aesthetic judgement of orientation in modern art. *i-Perception* 3, 1 (2012), 18–24.
- I. C. McManus, K. Stöver, and D. Kim. 2011. Arnheim’s Gestalt theory of visual balance: Examining the compositional structure of art photographs and abstract images. *i-Perception* 2 (2011), 615–647. Issue 6.
- D. Melcher and F. Bacci. 2012. 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. 191–216.
- L. Moholy-Nagy. 1945. In Defense of “Abstract” Art. *The Journal of Aesthetics and Art Criticism* 4, 2 (1945), 74–76.
- J. Moser. 2010. True Skil library. <https://github.com/moserware/Skills/>. (2010).
- M. Pelowski and F. Akiba. 2011. A model of art perception, evaluation and emotion in transformative aesthetic experience. *New Ideas in Psychology* 29 (2011), 80–97. Issue 2.
- R. W. Picard. 1995. Affective Computing. *Perceptual Computing Section, Media Laboratory, Massachusetts Institute of Technology* (1995).
- E. Pihko, A. Virtanen, V. Saarinen, S. Pannasch, L. Hirvenkari, T. Tossavainen, A. Haapala, and R. Hari. 2011. Experiencing Art: The Influence of Expertise and Painting Abstraction Level. *Frontiers in Human Neuroscience* 5 (2011).
- P. Reinagel and A. Zador. 1999. Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems* 10, 4 (1999).
- J. A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39, 6 (1980), 1161–1178.
- M. Schapiro. 1937. *Nature of abstract art*. American Marxist Association.
- N. Sebe, I. Cohen, and T. S. Huang. 2005. Multimodal Emotion Recognition. *Handbook of Pattern Recognition and Computer Vision* 4 (2005), 387–419.
- P. J. Silvia. 2005. Emotional responses to art : From collation and arousal to cognition and emotion. *Review of general psychology* 9, 4 (2005), 342–357.
- P. J. Silvia. 2006. Artistic training and interest in visual art: Applying the appraisal model of aesthetic emotions. *Empirical Studies of the Arts* 24, 2 (2006), 139–161.
- J. Sivic and A. Zisserman. 2003. Video Google: A Text Retrieval Approach to Object Matching in Videos. In *Proceedings of the International Conference on Computer Vision (ICCV)*. 1470–1477.
- R. Subramanian, H. Katti, N. Sebe, M. S. Kankanhalli, and C. Tat-Seng. 2010. An Eye Fixation Database for Saliency Detection in Images. In *European Conference on Computer Vision (ECCV)*. 30–43.

- R. Subramanian, V. Yanulevskaya, and N. Sebe. 2011. Can computers learn from humans to see better?: inferring scene semantics from viewers' eye movements. In *Proceedings of the ACM International Conference on Multimedia*. 33–42.
- R. Szeliski. 2010. *Computer Vision: Algorithms and Applications*. (2010).
- S. Takahashi and Y. Ejima. 2013. Contextual information processing of brain in art appreciation. *Behavioral and Brain Sciences* 36, 2 (2013), 158–159.
- B. W. Tatler, R. J. Baddeley, and I. D. Gilchrist. 2005. Visual correlates of fixation selection: Effects of scale and time. *Vision Research* 45, 5 (2005), 643–659.
- J. R. R. Uijlings, A. W. M. Smeulders, and R. J. H. Scha. 2010. Real-time Visual Concept Classification. *IEEE Transactions on Multimedia* 12, 7 (2010), 665–681.
- J. R. R. Uijlings, A. W. M. Smeulders, and R. J. H. Scha. 2012. The Visual Extent of an Object. *International Journal of Computer Vision* 96, 1 (2012), 46–63.
- K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. 2010. Evaluating Color Descriptors for Object and Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 9 (2010), 1582–1596.
- A. Vedaldi and B. Fulkerson. 2010. VLFeat: An open and portable library of computer vision algorithms. In *Proceedings of the ACM International Conference on Multimedia*. 1469–1472.
- L. Veronesi. 1968. *Luigi Veronesi. Con una antologia di scritti*. Torino.
- E. Winner, P. Blank, C. Massey, and H. Gardner. 1983. Children's sensitivity to aesthetic properties of line drawings. In *The acquisition of symbolic skills*. Springer, 97–104.
- O. Wu, Y. Chen, B. Li, and W. Hu. 2011. Evaluating the visual quality of web pages using a computational aesthetic approach. In *ACM Conference on Web Search and Data Mining*. 337–346.
- V. Yanulevskaya, J. V. Gemert, K. Roth, A. Herbold, N. Sebe, and J. Geusebroek. 2008. Emotional valence categorization using holistic image features. In *IEEE International Conference on Image Processing*. 101–104.
- V. Yanulevskaya, J. B. Marsman, F. Cornelissen, and J. M. Geusebroek. 2011. An image statistics based model for fixation prediction. *Cognitive Computation* 3, 1 (2011), 94–104.
- V. Yanulevskaya, J. Uijlings, E. Bruni, A. Sartori, E. Zamboni, F. Bacci, D. Melcher, and N. Sebe. 2012. In the eye of the beholder: employing statistical analysis and eye tracking for analyzing abstract paintings. In *Proceedings of the ACM International Conference on Multimedia*. 349–358.
- S. Zeki. 1999. Art and the brain. *Journal of Consciousness Studies* 6, 6-7 (1999), 76–96.
- Semir Zeki and John Nash. 1999. *Inner Vision: An Exploration of Art and the Brain*. Vol. 415. Oxford University Press Oxford.
- Q. Zhao and C. Koch. 2011. Learning a saliency map using fixated locations in natural scenes. *Journal of vision* 11, 3 (2011), 9.
- S. Zhao, Y. Gao, X. Jiang, H. Yao, T. Chua, and X. Sun. 2014. Exploring Principles-of-Art Features For Image Emotion Recognition. In *Proceedings of the ACM International Conference on Multimedia*. 47–56.