In the Eye of the Beholder: Employing Statistical Analysis and Eye Tracking for Analyzing Abstract Paintings

Victoria Yanulevskaya[1], Jasper Uijlings[1], Elia Bruni[2], Andreza Sartori[1,3],
Elisa Zamboni[2], Francesca Bacci[4], David Melcher[2], Nicu Sebe[1]

[1] DISI, University of Trento, Italy;
{yanulevskaya, jrr, andreza.sartori, sebe}@disi.unitn.it

[2] CiMEC, University of Trento, Italy;
{elia.bruni, david.melcher}@unitn.it, elisa.zamboni@student.unitn.it

[3] Telecom Italia - SKIL, Trento, Italy

[4] Museum of Art of Trento and Rovereto, Rovereto, Italy;
f.bacci@mart.tn.it

ABSTRACT

Most artworks are explicitly created to evoke a strong emotional response. During the centuries there were several art movements which employed different techniques to achieve emotional expressions conveyed by artworks. Yet people were always consistently able to read the emotional messages even from the most abstract paintings. Can a machine learn what makes an artwork emotional? In this work, we consider a set of 500 abstract paintings from Museum of Modern and Contemporary Art of Trento and Rovereto (MART), where each painting was scored as carrying a positive or negative response on a Likert scale of 1-7. We employ a state-of-the-art recognition system to learn which statistical patterns are associated with positive and negative emotions. Additionally, we dissect the classification machinery to determine which parts of an image evokes what emotions. This opens new opportunities to research *why* a specific painting is perceived as emotional. We also demonstrate how quantification of evidence for positive and negative emotions can be used to predict the way in which people observe paintings.

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

MM'12, October 29–November 2, 2012, Nara, Japan.
Copyright 2012 ACM 978-1-4503-1089-5/12/10 ...$15.00.

Keywords

Visual Art, Abstract Paintings, Emotion Recognition, Eye tracking

1. INTRODUCTION

Most art is created to evoke emotions. These emotions are not only contained within the narrative of the art piece, as for example in religious art, but also in the very techniques used to create it. In many abstract paintings the idea is to evoke emotions purely through the (non-figurative) elements that make a painting: colours, lines, shapes, and textures. For example, 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 colors and their interactions" [32]. The fact that for many paintings observers report similar emotions, shows that indeed there is consistency in the way people interpret colours and patterns when looking at these paintings. An interesting question is then which part of a painting evokes 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 the aesthetics [16], interestingness [8], and emotions [14, 34] of images, paintings, and even web pages [33]. Surprisingly, both [16] and [8] showed that the popular Bag-of-Visual-Words framework performed better in aesthetics prediction than the theory inspired features such as the golden rule. Hence, while improving prediction performance, the lack of success for the theory inspired features made the reasons *why* a specific image was deemed positive or beautiful less clear, especially as the exact workings of Bag-of-Visual-Words are little understood.

In this paper we aim for a better understanding of why a painting evokes a certain emotion by determining where the classification evidence resides. In particular, we train a Bag-of-Visual-Words model to predict if a painting evokes positive or negative emotions. We use a collection of abstract paintings which evoke emotions purely through geometry and colour, exactly the properties that are well captured by Bag-of-Words. Afterwards, we use the method in [29] to dissect the classification machinery to determine which parts of an image evokes what emotions.

The quantification of evidence for positive and negative emotions opens up a new toolbox for both researchers and artists alike. It allows art researchers to provide more quantitative data and hence objective grounds for forming and verifying their art theories. For artists it can help them better understand why their paintings evoke certain emotions and how they can make emotional content even stronger. As an application in this paper we test the positive attentional bias, which states that people prefer to look at the positive parts of an image, regardless if the overall image tends towards negativity or positivity.

To summarize, our main contributions are: (1) we use computer vision to help quantify which part of the image evokes what emotion; (2) we present an in-depth analysis to investigate if there is a positive attentional bias when looking at abstract paintings; and finally (4) we propose a new method of analysis useful for both art researchers and artists alike.

The rest of the paper is structured as follows. Section 2 presents the related work covering the art theory aspect as well as the existing emotion recognition approaches using computer vision techniques. Section 3 presents the details of our proposed approach. In Section 4, 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. In Section 5, we present an eye tracking study to prove that in general, people prefer to focus on the positive parts of the paintings. Section 6 presents the final discussions and conclusions.

2. RELATED WORK

2.1 Art Theory

Arnheim defines art as "the product of organisms and therefore probably neither more nor less complex than these organisms themselves" [1]. The discipline of art history encompasses the study of development of art works, in a specific period of time, and the explanation of how to interpret these based on historical context and styles. The evolution of art works has seen many painting styles that have become the main pillars of the art philosophy over time [9]. Furthermore, artists have employed the use of several features and techniques to evoke specific emotions in the viewer through their art. Many works try to comprehend the perception, evaluation and the consequent emotion that is evoked in the viewer by a piece of art. Hagtvedt *et al.* [7] developed a model based on the cognitive and emotional elements that are stimulated by a piece of art. In addition, they employed a structured equation model that integrates these elements in the evaluation process. They concluded that perception and emotion of a visual art piece depend on factors such as curiosity and aesthetic appeal, complexity, typicality or familiarity [7]. Pelowski and Akiba [18] developed a five-stage model of art perception that relates art-viewing to the viewer's personality. Recently, McManus *et al.* [17] used the Arnheim's Gestalt theory of visual balance in order to examine the visual composition of art photographs and abstract images. Similar works which deal with perception include [4, 10] and emotional responses to art [22]. 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 dissect its decision such that we can see exactly what parts of the image are responsible for this emotion. This provides a new angle to test and verify art theories.

2.2 Emotion Recognition

Humans can autonomously perceive and evaluate emotional signals of visual surroundings. However, it is a difficult task for machines to recognize emotions automatically. Recent works that observed emotional responses invoked by pictures deal with interdisciplinary fields such as art, aesthetics and psychology [9]. 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.

Several recent works have focused on emotion recognition in art. Yanulevskaya *et al.* [34] proposed an emotion categorization system which is based on the assessment of local image statistics followed by supervised learning of emotion categories using Support Vector Machines. Their system was trained on the International Affective Picture System, which is a standard emotion evoking image set in psychology [11], and was then applied to a collection of masterpieces. The work by Machajdik and Hanbury [14] 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 categorization of affective images which are semantically rich in comparison with abstract paintings which are non-figurative and thus semantics free. These works demonstrate that it is possible to use computer vision techniques for emotion recognition. In our analysis, we concentrate on the basis of why such categorization happens.

The comprehension of how people observe paintings is essential for the realization of the features contained within an image 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 and the features contained in an artwork that are closely related to a specific emotion [19]. Moreover, emotional stimuli are shown to attract attention. Recently, Subramanian *et al.* [24, 25] demonstrated how eye movements can be used in order to understand social and affective scenes. In this paper we investigate the link between emotional content and the way people look at abstract paintings. Particularly, we hypothesize that people prefer to focus on the positive parts of the paintings.

3. PROPOSED METHOD

To learn the difference between positive and negative paintings, we use a state-of-the-art Bag-of-Visual-Words classification framework which largely follows [28].

3.1 General Bag-of-Visual-Words Framework

For paintings representation we use a standard bag-of-visual-words framework, inspired by NLP [23]. This approach relies on the notion of a common vocabulary of "visual words" that can serve as discrete representations for a collection of images. 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. Typically, 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 a new image can be translated into words by determining which visual word they are nearest to in the feature space (i.e., based on the Euclidean distance between the cluster centers and each descriptor feature) [6]. Finally, each image is represented as a histogram of its visual words.

3.2 Descriptors

To take into account two most important characteristics of abstract art works, in our analysis we represent paintings with colour based LAB visual words and texture based SIFT visual words.

3.2.1 LAB

The LAB color space plots image data in 3 dimensions along 3 independent (orthogonal) axes, L for brightness and a and b for the color-opponent dimensions, based on nonlinearly compressed CIE XYZ color space coordinates. One problem with the CIE XYZ color system, is that colorimetric distances between the individual colors do not correspond to perceived color differences. In LAB, luminance corresponds closely to brightness as recorded by the brain-eye system; chrominance (red-green and yellow-blue) axes mimic the oppositional color sensations the retina reports to the brain [26]. To map LAB descriptors to visual words, we quantize the color space into 343 different colours by uniformly dividing each colour channel into 7 different levels.

3.2.2 SIFT

The SIFT descriptor proposed by Lowe [13] describes the local shape of a region using edge orientation histograms. It is derivative-based and contains local spatial information. Particularly, SIFT descriptors capture contours, edges, and textures within images. They demonstrate invariance to image scale transformations, and also provide a robust matching across affine distortion, noise and change in illumination.

In this paper, we take local patches of 16- by-16 pixels which are 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.* [30]. To create a visual vocabulary we quantize 250,000 randomly selected SIFT descriptors into 4096 clusters using a hierarchical version of k-means clustering [31].

3.3 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 [15] for supervised learning of emotions. 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 allows to determine the relative contribution which comes from each pixel in the image to the classification task. In our case, thanks to 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 [29]. Consider the classification function for the Histogram Intersection kernel, defined as

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

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. In this function the outer sum is over the m support vectors. However, for determining which visual word contributes how much 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 [15] where the inner sum can be brought outside, leading to:

$$\begin{aligned} h(\mathbf{x}) &= b + \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_j t_j \min(x_i, z_{ij}) && (2) \\ &= b + \sum_{i=1}^{n} w_i. && (3) \end{aligned}$$

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. EMOTION RECOGNITION FOR ABSTRACT PAINTINGS

In this section 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.

4.1 MART Dataset

4.1.1 The painting collection

We selected 500 images of abstract paintings from the art collection of the MART museum in Rovereto, Italy (the digitalized images of the artworks are contained in the electronic archive of MART). The artworks were realized between 1913 and 2008 by Italian as well as European and American artists. The collection presents some characteristics that are interesting from the point of view of art history. In particular, there are four authors whose work is well represented in the group of paintings which were used in this experiment: Carlo Belli, Aldo Schmid, Luigi Senesi, and Luigi Veronesi. These artists are particularly interesting because their work is not the result of an improvisation but it is part of a deep theoretical reflection on the elements that make a painting (colors, lines, shapes, textures). They all wrote their theoretical reflections declaring the principles that inspired and shaped their artworks.

4.1.2 Participants

We recruited a total of 100 people (74 females, 26 males; ages 18-65. $M = 39.87$). Most participants were teachers in primary schools, operators at MART, visitors of the museum, and students. They 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.

4.1.3 Scoring procedure

In this paper, we are interested in analysing whether positive or negative feelings are conveyed by abstract paintings. Therefore, we follow the standard procedure [5, 20] and consider positive-negative valenced emotional judgements. In our experiment, we asked the participants to look at the MART paintings and report their emotions elicited by them. Particularly, 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." The judgements were given according to a Likert scale of 1-to-7-points, where number 1 meant a highly negative emotion and 7 meant a highly positive emotion. The scale was presented on a separated sheet of paper, although paintings were accessed through a code to prevent subjects from getting influenced by the artwork's title or its author. While the participants did not have a fixed time window, they were explicitly instructed to score the paintings as fast as possible in order to encourage the rise of very instinctive emotions, and to prevent that previous knowledge about the artworks could come into play. On average, people spent about 9 seconds per painting to make their judgement.

During the experiment, the images of the paintings were randomly divided into five groups, 100 works each, and only one group was presented to a single subject. Therefore, each subject rated 100 paintings and each artwork received 20 judgments from 20 different subjects. We presented the paintings on a gray background, because gray is usually considered to be a neutral color, with no impact to the other colors (and withcoming emotions). After every ten images, a gray slide was presented in order to give participants the opportunity to rest.

To check for preferential biases, we analyse the data according to the gender of the participants (74 females versus 26 males), the age of the participants (48 people under 40 years old versus 52 people over 40 year old), and the art background. The participants that visited more than 5 museums per year are considered to be knowledgable in art (31 people), whereas the rest are not (69 people). We did not observe significant difference in judgement within all these three groups (p values of a double t-test ≥ 0.4596).

Figure 1 illustrates the distribution of paintings from the most negative (average score of 1.95) to the most positive (average score of 6.2). The images, emotional scores and recorded eye-movements described in Section 5 have been made available online[1].

4.2 Classification Results

To construct a ground truth, for each painting the average of all available scores is computed. We define paintings with average scores lower or equal to 4 as negative, and paintings with average scores higher than 4 as positive. As a result, 183 paintings were assigned to the negative class, and 317 to the positive class. We tested our model on a two-fold cross-validation repeated ten times. The training/testing sets are from a random 50% and 50% split of the MART dataset. The average accuracy is reported in Table 1.

Table 1: Classification accuracy.

	LAB	SIFT	LAB+SIFT
Accuracy	0.76	0.73	0.78

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. Overall, the proposed method performs significantly better than chance level which reaches accuracy of 0.63 in our experimental settings. We observe that LAB visual words are more effective for emotional classification of abstract paintings compared to SIFT visual words, the accuracy is 0.76 versus 0.73 respectively. The combination of LAB and SIFT visual words raises the accuracy to 0.78. This indicates that both colour-based LAB visual words and texture-based SIFT visual words are important for emotion recognition.

4.3 Backprojections

In this section we visualize how the classifier "sees" paintings while deciding whether they are positive or negative. We use the method described in Section 3.4 and backproject the relative contribution which comes from each pixel in the painting. The results for 4 paintings which are classified as highly positive and 4 paintings which are classified highly negative are shown in Figures 2 and 3 respectively. 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 have mostly negative response, while red, yellow, and blue often evoke positive emotions.

The third rows in Figures 2 and 3 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 on the positive side. The last rows show the combined evidence from LAB and SIFT visual words. Here evidence coming from colour and texture of the painting is effectively mixed together.

5. EYE MOVEMENTS AND EMOTIONAL CONTENT

In the previous section we showed that we can classify MART paintings as being positive and negative and determine which parts of the image evoke what emotion. As a practical application, in this section we use this to investigate the link between emotional content and the way people look at abstract paintings. We hypothesize that, in general, people prefer to focus on the positive parts of the paintings.

To check our hypothesis, we recorded eye movements of 9 observers while they viewed art works from the MART dataset. In order to ensure variability of emotional con-

[1] http://disi.unitn.it/∼yanulevskaya/mart.html

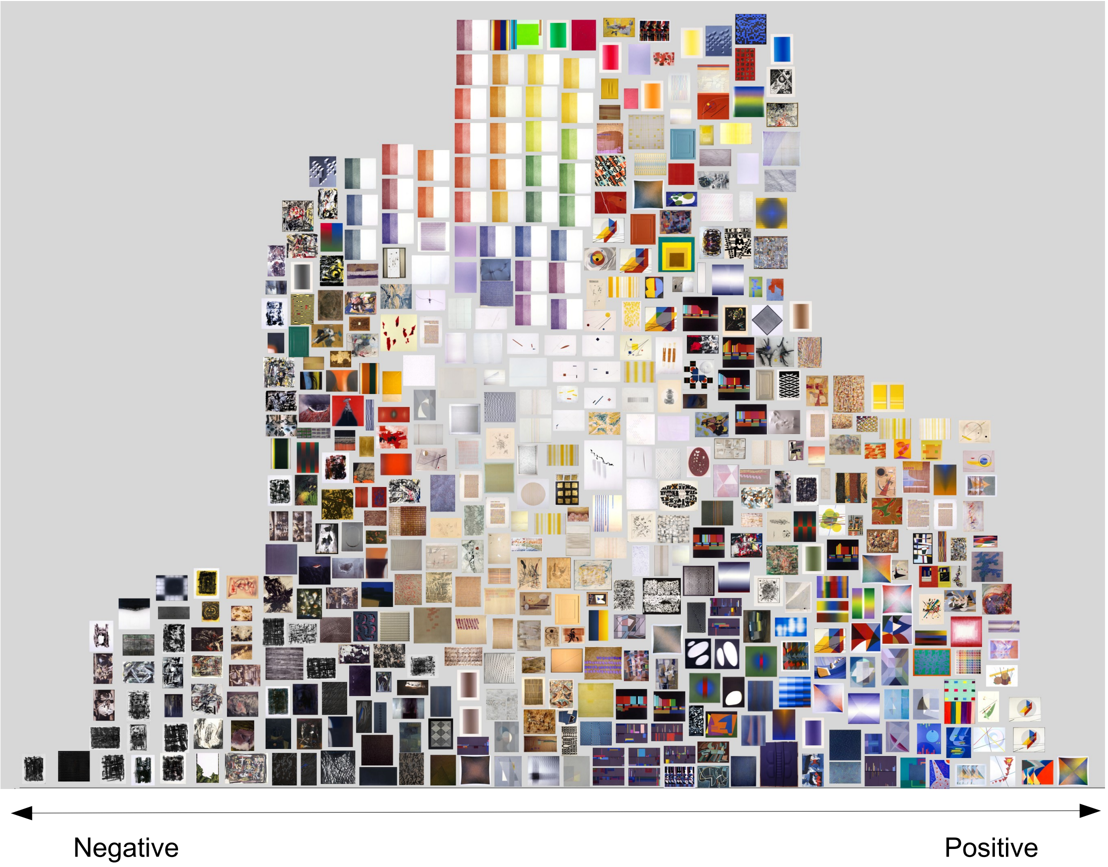

Figure 1: Paintings from the MART collection ordered by human scores.

tent, for this experiment we selected paintings which contain 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, where 71 paintings are scored as positive by humans and 39 are scored as negative.

5.1 Eye Movement Recording

Eye movements were recorded using an EyeLink 1000 Tower Mount which 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 23 inch screen (ASUS VG236H, 1920x1080 resolution) for 7 seconds. After each presentation of the stimulus, a gray background was shown for 0.5 second 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 [2].

Nine participants took part in the experiment, each 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 as they would do it in a museum.

5.2 Data Analysis

According to Locher *et al.* [12], viewers of art (no matter if representational or abstract) occupy the first two seconds to do a sweep of the image, analyzing its "gist". Only after this first explorative stage, viewers tend to focus on finer details. Moreover, it has been shown that during this period, bottom-up saliency plays an important role in the allocation of eye movements [3, 27, 35, 36]. In this paper we are interested in the influence on gaze patterns of such a high level information as emotional content. Therefore, in our analysis we consider eye movements which occur between the second and seventh seconds.

To quantify the relationship between emotional content detected by the classifier as traced with backprojection technique and human's gaze pattern, we compared emotional evidence which 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 it 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 which is sampled in high resolution by the human eye. Now if people prefer to focus on those parts of the painting which contain positive infor-

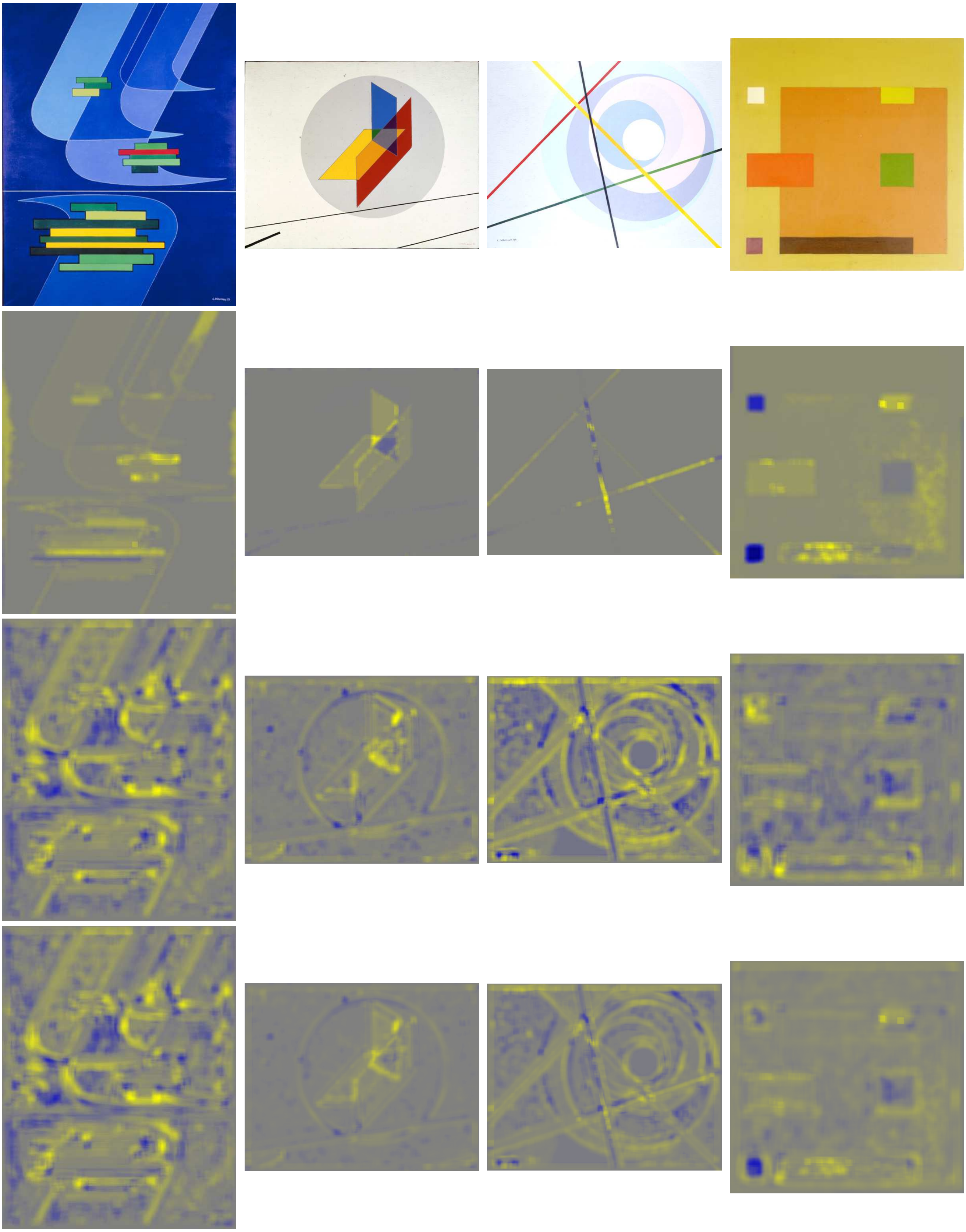

Figure 2: Visualisations of pixel-wise contribution to the classification for highly positive paintings. The original paintings are shown in the first row. The second row displays the contribution of LAB visual words, followed by the the contribution of SIFT visual words in row 3, and their combination 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)

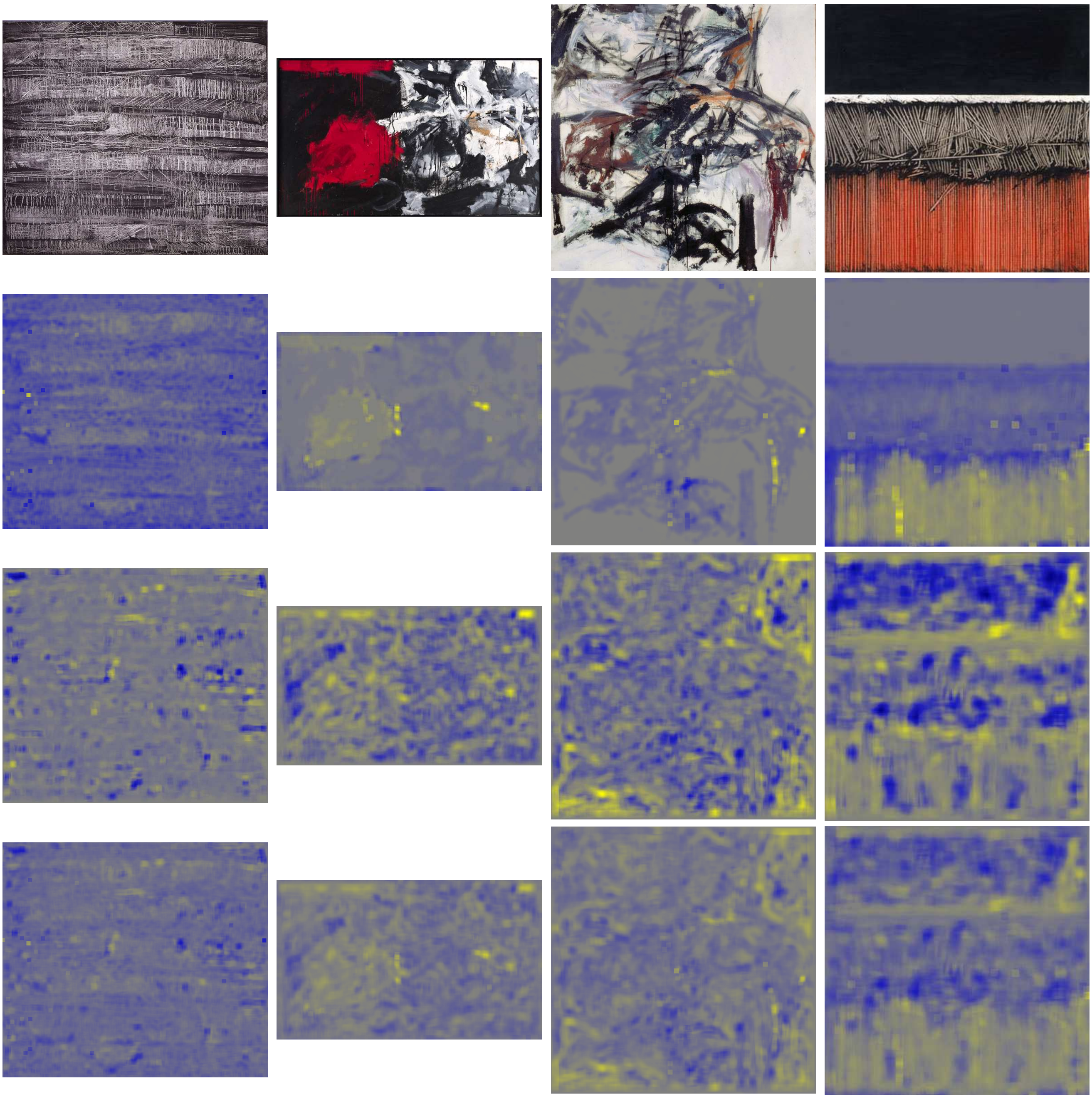

Figure 3: Visualisations of pixel-wise contribution to the classification for highly negative paintings. The original paintings are shown in the first row. The second row displays the contribution of LAB visual words, followed by the the contribution of SIFT visual words in row 3, and their combination 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)

mation, 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 [21, 27] 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° (30 pixels) apart from the fixated locations. This guarantees similar distributions of fixated and non-fixated regions [27].

Table 2: Distribution of paintings according to "emotions" of the attentional bias.

LAB+SIFT	#	PAB	NAB	Neutral
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%)

We call Positive Attentional Bias (**PAB**) the situation when average contribution to the classifier from the fixated regions is *higher* than average contribution to the classifier from the non-fixated regions. We call Negative Attentional Bias (**NAB**) 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 evidences, we use our best performing classifier according to Table 1 which combines LAB and SIFT descriptors. Table 2 contains averaged results over 50 independent samplings of non-fixated locations. Only for 2 paintings attentional bias demonstrates neutral behaviour, whereas 74% of considered paintings show the positive attentional bias. As it can be expected, PAB holds for most of the emotionally positive paintings (83%). Interestingly, even while observing negative paintings, in 59% of cases people still prefer to look at positive parts. Therefore, we conclude that at least for this dataset, the positive attentional bias holds. Figure 4 illustrates several examples of paintings together with the recorded fixated locations.

6. CONCLUSION

In this paper we sought for a better understanding of which part of an image evoked what kind of emotions through advanced computer vision techniques. We trained a Bag-of-Visual-Words system to distinguish between positive and negative abstract paintings. Afterwards we backprojected the classification contributions back to the painting to have an accurate visualization of which part of the painting conveyed positive emotions and negative emotions. At first, qualitative analysis showed 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.

The ability of localising the evidence for emotions opens up a new toolbox for researchers and artists for the verification and formulation of art theories. Theoretically, the technique is not limited to emotions but can be applied to better study aspects such as aesthetics and interestingness as well. We are planning to do so in future work.

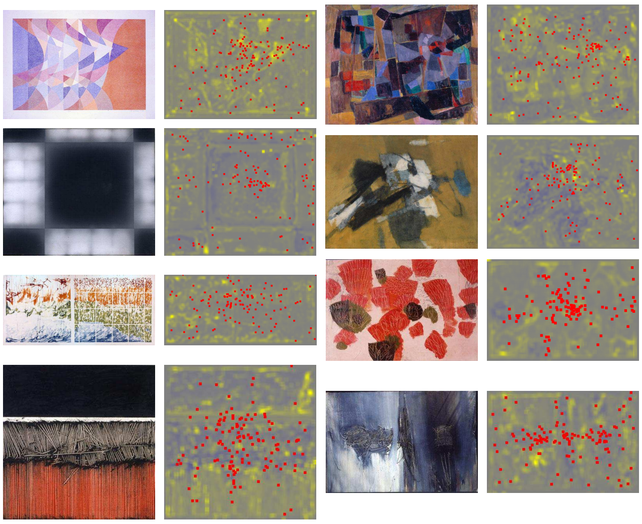

Figure 4: Paintings with superimposed fixated locations: the first row displays positive paintings with PAB, the second row displays negative paintings with PAB, the second row displays positive paintings with NAB, and the last row displays negative paintings with NAB. (Courtesy of MART photographic archive, Rovereto)

As a practical application, we used the localisation of the emotion evidence in combination with recorded eye movements to test where people prefer to focus while observing abstract art works. We hypothesized that positive parts attract most of the attention. Indeed, our results demonstrated that positive attentional bias holds for 74% of considered paintings, and even in paintings with a negative emotional content people prefer to look at their positive parts.

7. ACKNOWLEDGMENTS

Victoria Yanulevskaya, Jasper Uijlings, and Nicu Sebe were partially supported by the FIRB S-PATTERNS project. Francesca Bacci was supported by a CARITRO (Fondazione Cassa di Risparmio di Trento e Rovereto) grant. The authors would like to thank MART photographic archive for allowing the use of the images for this study.

8. REFERENCES

[1] R. Arnheim. *Art and Visual Perception: A Psychology of the Creative Eye*. University of California Press, Nov. 2004.

[2] D. H. Brainard. The psychophysics toolbox. *Spatial Vision*, 10(4), 1997.

[3] N. Bruce and J. Tsotsos. Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision*, 9(3), 2009.

[4] G. C. Cupchik, O. Vartanian, A. Crawley, and D. J. Mikulis. Viewing artworks: Contributions of cognitive control and perceptual facilitation to aesthetic experience. *Brain and Cognition*, 70:84–91, 2009.

[5] M. Dellagiacoma, P. Zontone, G. Boato, and L. Albertazzi. Emotion based classification of natural

images. In *Int. workshop on Detecting and Exploiting Cultural Diversity on the Social Web*, pages 17–22, 2011.

[6] K. Grauman and B. Leibe. *Visual Object Recognition.* Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2011.

[7] H. Hagtvedt, R. Hagtvedt, and V. Patrick. The perception and evaluation of visual art. *Empirical Studies of the Arts*, 26(2):197–218, 2008.

[8] P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an image memorable? In *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[9] D. Joshi, R. Datta, E. A. Fedorovskaya, Q.-T. Luong, J. Z. Wang, J. Li, and J. Luo. Aesthetics and emotions in images. *IEEE Signal Process. Mag.*, 28(5):94–115, 2011.

[10] D. Joshi, R. Datta, E. A. Fedorovskaya, Q.-T. Luong, J. Z. Wang, J. Li, and J. Luo. Aesthetic judgement of orientation in modern art. *i-Perception*, 3:18–24, 2012.

[11] P. J. Lang, M. M. Bradley, and B. N. Cuthbert. International affective picture system (iaps): Technical manual and affective ratings. 1999.

[12] P. Locher, E. A. Krupinski, C. Mello-Thoms, and C. F. Nodine. Visual interest in pictorial art during an aesthetic experience. *Spatial Vision*, 21(1-2):55–77, 2007.

[13] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vision*, 60:91–110, 2004.

[14] J. Machajdik and A. Hanbury. Affective image classification using features inspired by psychology and art theory. In *ACM Multimedia*, pages 83–92, 2010.

[15] S. Maji, A. C. Berg, and J. Malik. Classification using intersection kernel support vector machines is efficient. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[16] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka. Assessing the aesthetic quality of photographs using generic image descriptors. In *International Conference on Computer Vision*, 2011.

[17] I. C. McManus, K. Stöver, and D. Kim. Arnheim's gestalt theory of visual balance: Examining the compositional structure of art photographs and abstract images. *i-Perception*, 2:615–647, 2011.

[18] M. Pelowski and F. Akiba. A model of art perception, evaluation and emotion in transformative aesthetic experience. *New Ideas in Psychology*, 29:80–97, 2011.

[19] E. Pihko, A. Virtanen, V. Saarinen, S. Pannasch, L. Hirvenkari, T. Tossavainen, A. Haapala, and R. Hari. Experiencing art: The influence of expertise and painting abstraction level. *Front Hum Neurosci*, 5, 2011.

[20] E. Pihko, A. Virtanen, V. M. Saarinen, S. Pannasch, L. Hirvenkari, T. Tossavainen, A. Haapala, and R. Hari. Experiencing art: the influence of expertise and painting abstraction level. *Frontiers in human neuroscience*, 5, 2011.

[21] P. Reinagel and A. Zador. Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems*, 10(4), 1999.

[22] P. J. Silvia. Emotional responses to art : From collation and arousal to cognition and emotion. *Review of general psychology*, 9(4):342–357, 2005.

[23] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *Proceedings of the International Conference on Computer Vision*, pages 1470–1477, Oct. 2003.

[24] R. Subramanian, H. Katti, N. Sebe, M. Kankanhalli, and T. S. Chua. An eye-fixation database for saliency detection in images. *European Conference on Computer Vision*, 2010.

[25] R. Subramanian, V. Yanulevskaya, and N. Sebe. Can computers learn from humans to see better? Inferring scene semantics from viewers' eye movements. In *ACM Multimedia*, pages 33–42, 2011.

[26] R. Szeliski. Computer vision: Algorithms and applications. 2010.

[27] B. W. Tatler, R. J. Baddeley, and I. D. Gilchrist. Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, 45(5), 2005.

[28] J. R. R. Uijlings, A. W. M. Smeulders, and R. J. H. Scha. Real-time Visual Concept Classification. *IEEE Transactions on Multimedia*, 12, 2010.

[29] J. R. R. Uijlings, A. W. M. Smeulders, and R. J. H. Scha. The Visual Extent of an Object. *International Journal of Computer Vision*, 2012.

[30] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9), 2010.

[31] A. Vedaldi and B. Fulkerson. VLFeat – an open and portable library of computer vision algorithms. In *ACM Multimedia*, 2010.

[32] L. Veronesi. *Luigi Veronesi. Con una antologia di scritti.* Torino, 1968.

[33] O. Wu, Y. Chen, B. Li, and W. Hu. Evaluating the visual quality of web pages using a computational aesthetic approach. In *Association of Computing Machinery conference on Web search and data mining*, pages 337–346, 2011.

[34] V. Yanulevskaya, J. V. Gemert, K. Roth, A. Herbold, N. Sebe, and J. Geusebroek. Emotional valence categorization using holistic image features. In *IEEE International Conference on Image Processing*, pages 101–104, 2008.

[35] V. Yanulevskaya, J. B. Marsman, F. Cornelissen, and J. M. Geusebroek. An image statistics based model for fixation prediction. *Cognitive Computation*, 3(1), 2011.

[36] Q. Zhao and C. Koch. Learning a saliency map using fixated locations in natural scenes. *Journal of vision*, 11(3), 2011.