JenAesthetics Subjective Dataset: Analyzing Paintings by Subjective Scores

Seyed Ali Amirshahi[1,2(✉)], Gregor Uwe Hayn-Leichsenring[2], Joachim Denzler[1], and Christoph Redies[2]

[1] Computer Vision Group, Friedrich Schiller University Jena, Jena, Germany
{seyed-ali.amirshahi,joachim.denzler}@uni-jena.de
http://www.inf-cv.uni-jena.de
[2] Experimental Aesthetics Group, Institute of Anatomy I, Jena University Hospital, Jena, Germany
{Gregor.Hayn-Leichsenring,christoph.redies}@med.uni-jena.de
http://www.anatomie1.uniklinikum-jena.de

Abstract. Over the last few years, researchers from the computer vision and image processing community have joined other research groups in searching for the bases of aesthetic judgment of paintings and photographs. One of the most important issues, which has hampered research in the case of paintings compared to photographs, is the lack of subjective datasets available for public use. This issue has not only been mentioned in different publications, but was also widely discussed at different conferences and workshops. In the current work, we perform a subjective test on a recently released dataset of aesthetic paintings. The subjective test not only collects scores based on the subjective aesthetic quality, but also on other properties that have been linked to aesthetic judgment.

Keywords: Computational aesthetics · Aesthetic · Beauty · Color · Content · Composition · Paintings · Subjective dataset · JenAesthetics dataset

1 Introduction

In recent years, there has been a growing interest in the topic of aesthetic quality assessment of paintings and photographs in the computer vision and image processing community. This interest has resulted in what is now known as computational aesthetics [12]. Numerous workshops, conferences and special sessions dealing with this topic have attracted researchers in the past few years [2–10,17–20,22–24,28–30,39]. Due to the nature of research in this field, further progress depends on the availability of datasets for analysis.

Over the years, most research in this field has focused on proposing new methods to evaluate different aesthetic properties [2–11,17,18,20,21,23,24,28–30,38,39]. Although these methods reached interesting results, the lack of a common dataset prevented different methods and approaches to be comparable to one another.

L. Agapito et al. (Eds.): ECCV 2014 Workshops, Part I, LNCS 8925, pp. 3–19, 2015.
DOI: 10.1007/978-3-319-16178-5_1

Like in other fields of research that deal with quality assessment of stimuli, such as image and video quality assessment, subjective datasets [32–34,36,37] play an important role for research. Subjective datasets provide researchers with scores given by observers with regard to different properties of a stimuli. Thanks to the many photo-sharing websites nowadays used by professional and amateur photographers, several subjective datasets [8,17,19,20,22,23] covering different types and styles of photographs have been introduced to the public. These websites provide the user with a large number of photographs that have been rated subjectively by the community of photographers. The procedure for collecting such datasets is inexpensive both in the sense of time consumption and financial cost. It should be mentioned that a drawback of these datasets is that the scoring of images is not done in a standardized format. This means that the subjective scores were likely given under various viewing condition using different displaying devices. To try to prevent such issues, in the field of image and video quality, subjective tests are normally collected using specific standards such as those described in [13]. Unfortunately, in the aesthetic quality assessment community, there is no specific standard agreed among different research groups with regard to how subjective tests should be performed as of yet.

Unlike for photographs, there has been no public dataset of paintings with subjective scores until recently. Last year, two small subjective datasets have been introduced to the community [2,39]. However, these datasets fall short of corresponding to the needs of the community, which we will describe in the next section.

In this paper, we take advantage of the JenAesthetics dataset [1,5,15], which is available for public use, and perform a subjective test to evaluate different properties of the paintings in this dataset. The JenAesthetics dataset is one of the largest publicly available datasets and covers a wide range of different styles, art periods, and subject matters [1,5]. The images in this dataset are colored oil paintings that are all on show in museums and were scanned at a high resolution. The present study will combine the objective data previously provided in [5] with different subjective scores.

The next sections of this article are as follow: Section 2 introduces the previous subjective datasets. Section 3 describes the JenAesthetics subjective dataset. Section 4 evaluates the subjective scores provided by the observers. Finally, Section 5 gives a short conclusion and proposes possible future work to extend this dataset.

2 Previous Work

Before the introduction of the two mentioned datasets [2,39] (Sections 2.1 and 2.2, respectively), other researchers [4,7,10,18,21,24,28] have gathered their own datasets. This was done either by scanning high-quality art books, by ordering digital samples from museums, or by using their own personal collections. Unfortunately, there is no possibility to release these datasets to other research groups due to copyright restrictions, making different approaches incomparable to one another.

Table 1. Comparison of different properties between the JenAesthetics subjective dataset and the JenAestheticsβ [2] and the MART [39] subjective datasets. NA, not assessed.

Properties	JenAesthetics	JenAestheticsβ [2]	MART[39]
Number of images	1628	281	500
Number of observers per painting	19 - 21	49	20
Total number of observers	134	49	100
Scores of individual observers	yes	no	yes
Color images	yes	yes	majority
Average image size (pixels)	4583×4434	2489×2517	513×523
Number of properties evaluated	5	1 (beauty)	1 (emotion)
Rating scale	continuous, 1-100	ordinal, 1-4	ordinal, 1-7
Art periods/styles	11	NA	1
Number of subject matters	16	NA	1
Number of artists	410	36	78

In the following subsections, we will give a short summary on the two available subjective datasets [2, 39]. Table 1 lists different properties of the available subjective datasets.

2.1 JenAestheticsβ [2]

This dataset, which was introduced in 2013, consists of 281 paintings of different subject matters and art styles. A positive aspect of this dataset is the high number of observers who rated the images. The subjective scoring in this dataset has been done based on a scale of 1-4. Subjective scores show that paintings with bluish or greenish colors are generally given higher subjective scores compared to paintings with brownish or dark colors [2]. This result confirms findings by Palmer and Schloss [25]. Compared to the JenAesthetics dataset [1, 5, 15], this dataset does not provide information on the art periods which the paintings belong to, and no subject matters are assigned to the paintings.

2.2 MART Dataset [39]

This subjective dataset of paintings consists of 500 abstract paintings produced between 1913 and 2008. The paintings were selected from the MART museum in Rovereto, Italy. The images in this dataset were divided into 5 subsets, each consisting of 100 images. 20 observers rated each subset based on a 7-point rating scale. The observers were mostly female (74 females, 26 males) and on average visited 5.5 museums per year. The observers were allowed to spend as much time as they wanted to see and observe a painting before giving a score, but they were advised to rate the paintings in the fastest possible manner. 11 images are in a monochrome format (Table 1). Also, the average pixel size of the images is relatively small compared to the JenAesthetics and the JenAestheticsβ [2] datasets. Unlike these two datasets, the paintings of the MART dataset belong to a single art period/style (i.e., abstract art).

3 JenAesthetics Subjective Dataset

In collecting the JenAesthetics dataset [1,5,15], Amirshahi et al. took advantage of the fact that the Google Art Project (http://www.googleartproject.com/) has released a large number of high-quality scanned versions of artworks for public use. Although the artworks in this dataset are mostly from famous painters, this does not guarantee that they will be ranked highly by non-expert observers. The non-expert observers who participated in our experiment were not familiar with most of the paintings and/or painters. The importance of familiarity in evaluating the quality of a photograph or painting has been noticed in different studies [8,18]. A painting is labelled as familiar in the JenAesthetics subjective dataset if the observer believes that he/she has previously seen the painting or if they know the painter. Moreover, as it will be discussed in Section 4, some famous paintings are not among paintings with the highest subjective scores. This implies that the observers are not necessarily biased towards famous paintings.

As mentioned previously, there is a lack of standards for subjective tests in the field of computational aesthetics. We believe that the following issues have to be taken into account when performing a subjective test of paintings.

1. Tests should be carried out under standard viewing conditions in a controlled environment. This will ensure that the observers are viewing all paintings under the same condition so that the scores are comparable.
2. The observers should not be familiar with the paintings. Different approaches have been taken to prevent the subjective scores from being biased towards a familiar painting. For example, Li et al. [18] removed scores given by observers when they expressed that they were familiar with the painting shown. In our work, we found that the observers were not familiar with the vast majority of paintings evaluated.
3. Multiple properties should be assessed and not just one. For example, if we evaluate the aesthetic quality as well as the observers' liking of the colors, composition, and content of the paintings, as done for the JenAesthetics subjective dataset in the present study, we can correlate each preference with the aesthetic scores given by the observer (see Section 4.2).
4. The visual ability of the observer should be taken into account. Results provided from observers with visual impairment should be treated differently compared to other observers.

In the following sections, we will first provide information on why specific questions/properties were evaluated by the observers (see Section 3.1). We will then describe the experimental procedure (see Section 3.2).

Table 2. Questions that the observers were asked for each property in the JenAesthetics subjective dataset. The two terms visible on the rating scale which correspond to the highest and lowest possible scores are also shown.

Property	Question asked	Left side	Right side
Aesthetic quality	How aesthetic is the image?	not aesthetic	aesthetic
Beauty	How beautiful is the image?	not beautiful	beautiful
Liking of color	Do you like the color of the image?	no	yes
Liking of content	Do you like the content of the image?	no	yes
Liking of composition	Do you like the composition of the image?	no	yes
Knowing the artist	Do you know the artist?	no	yes
Familiarity with the painting	Are you familiar with this painting?	no	yes

3.1 Properties Evaluated

Table 2 gives an overview on the properties evaluated in the JenAesthetics subjective dataset. The main goal for the dataset is to collect subjective scores related to the aesthetic quality of paintings. Previous works have taken different approaches to reach this goal. While in the case of the JenAestheticsβ dataset Amirshahi et al. [2] used beauty as their measure, Li et al. [19] asked their subjects to give their general opinion about the painting. For the MART dataset [39], the observers were asked to give a score with regard to their emotion towards the paintings. In this experiment, the lowest score represented the most negative emotion while the highest score represented the most positive emotion. In the JenAesthetics subjective dataset, we evaluate subjective scores with regard to both aesthetic quality and beauty. The two properties are compared in Section 4.2.

We also evaluated three other properties (i.e., the liking of color, composition and content of the paintings) and studied the relationship between these properties and the aesthetic and beauty scores (see Section 4.2). Previously, Amirshahi et al. [2] and Yanulevskaya et al. [39] used simple color features to predict the aesthetic quality of paintings with a high accuracy. Other works such as [25,26,31] have also focused on the importance of color when evaluating the aesthetic quality of images. The composition of an image (for example the rule of thirds) plays an important role in the aesthetic quality of paintings and photographs according to several studies [3,8,16,18,20,35,38]. Finally, it is well known that the content of a painting and/or photograph can influence the subjective rating of aesthetic quality. In the JenAesthetics dataset, the content of paintings is represented by the subject matter.

Table 3. Characteristics of the JenAesthetics subjective dataset.

Attribute	Value
Number of participants	134
Number of observers after removing clickers and people with color blindness	129
Number of observation sessions	190
Age range of the observers	19 to 42 years
Mean age	25.3 years
Male / female	70 / 59
Right / left-handed	119 / 10
With / without glasses	64 / 65
Interested / not interested in art	90 / 39
Nationality	15 different countries
Nationality represented most frequently	103 from Germany

To evaluate whether the subjective scores are in any way biased by being familiar with a painting, the observers' familiarity towards each painting was assessed. We also asked the observers whether he/she knew the painter. A similar approach was taken by Li et al. [18].

3.2 Experimental Procedure

Participants 134 participants attended this study; sixty-seven of them took part in two observation sessions (leading to a total of 201 sessions). Most of them were students, in particular of natural sciences, but other fields of studies and professions were reported also. However, no participant was a student of arts, art history, or any related field. All participants declared having normal or corrected-to-normal visual acuity and gave their written informed consent after receiving an explanation of the procedures. The consent allows us to use and share their subjective scores. Each participant was tested for color blindness using the Ishihara test [14]. Data from observers who were color blind were excluded from the analysis. See Table 3 for additional data on the participants.

Stimuli We used the 1628 art images from the JenAesthetics database [1] as stimuli. In every session, a subset of 163 images were rated. Works from 410 painters are available in this dataset. The dataset covers paintings from 11 art periods (Renaissance, Baroque, Classicism, Romanticism, Realism, Impressionism, etc.). Each painting in the dataset is tagged with up to three different subject matters. These subject matters (16 in total) include abstract, landscape, still life, portrait, nude, urban scene, and etc.

Procedure The experiment was performed using the PsychoPy [27] program (version 1.77.01) on a BenQ T221W widescreen monitor with a resolution of 1680×1050 pixels (WSXGA+). The monitors where calibrated with a colorimeter (X-Rite EODIS3 i1Display Pro) using the same calibration profile in order to create similar conditions for all observers.

For presentation, each image was scaled so that the longer side of the image was 800 pixels on the screen. The images were placed in the middle of the screen on a black background (see Figure 1). Images were presented with a size of 20.5 cm (longest side) on the computer screen, corresponding to about 19.4 degrees of visual angle (at a viewing distance about 60 cm).

First, twenty images from the dataset that were not used in the rating experiment were presented for three seconds each to get the observer used to the data.

Then, 163 images that were selected randomly from the dataset were presented to the observer in a random order. Careful attention was taken so that no two subsets where identical. In total, each painting was rated by 19 to 21 observers. The participants were asked to rate the images on seven properties using a sliding bar located on the bottom of the screen. Internally, the sliding bar was binned into 100 equal intervals. Accordingly, the ratings obtained (see Figures 3-8) ranged from 0 to 100. The rated properties were "Aesthetic quality", "Beauty", "Liking of color", "Liking of content", "Liking of composition", "Knowing the artist", and "Familiarity with the painting" (see Table 2 for details on the presented questions). As shown in Figure 1, the questions were presented above the sliding bar. The terms presented in Table 2 indicated the range for the rating at each end of the bar. The observers where instructed that the mentioned phrases on the scoring bar were to represent the two extreme cases for the scores and that their scores would not be treated on a binary scale. The participants had no time restrictions for answering each question. After rating, the next question appeared. The image was visible until the last rating was given. Participants who attended a second session were provided with a new randomly selected set of images that shared no images with the images shown in the first session, in which they participated.

4 Analysis of the Subjective Scores

The first step in analyzing the subjective scores gathered was to remove the scores that were provided by observers in an improper manner. These scores were mainly provided by what will be referred from here on as clickers. Clickers are observers who provide their results by randomly clicking the score bar, independent of the image content or the question asked. The random clicking of the score bar is mostly performed at a high speed resulting in short response times (Figure 2(a)). Also, clickers tend not to move their mouse for a few questions before moving their mouse to another position (Figure 2(b)). Subjective scores for each property were calculated after removing the scores provided by the clickers. After removing the clickers and the scores provided by people who were color blind, the total number of observers was 129.

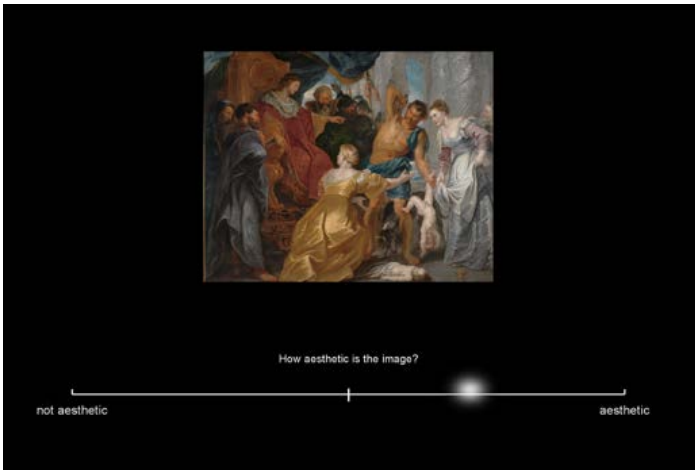


Fig. 1. Screen-shot from the subjective test for one of the assessed properties (aesthetic quality). The question regarding the assessed property is represented under the painting. Painting by Peter Paul Rubens, about 1617.

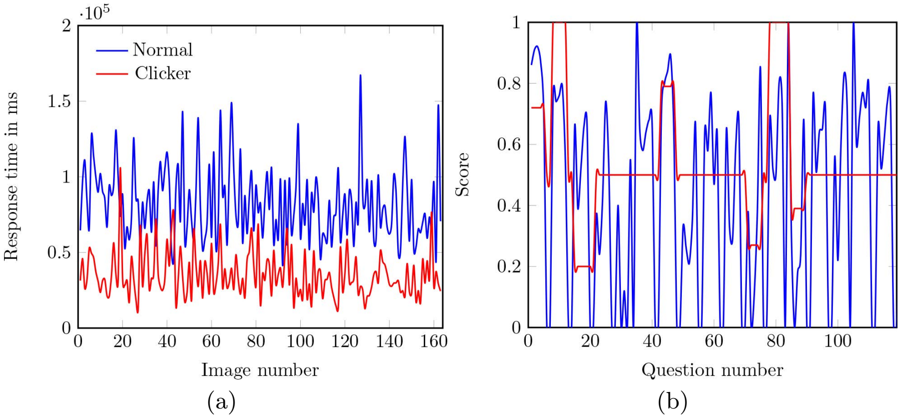
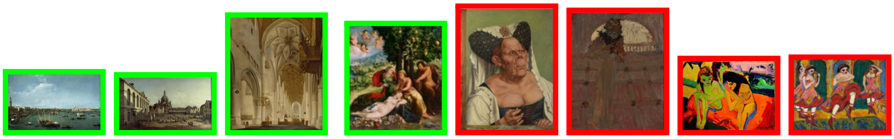


Fig. 2. Comparing results from a clicker and a normal observer. (a) Response time spent on each image. (b) Scores given to the seven properties for 17 paintings selected randomly for the same observers whose response times are shown in (a).

4.1 Calculating the Scores

After removing the clickers from the obtained data, the final step in producing a subjective datatest was to calculate a score for each property for each painting. Among the different options available, we decided that calculating the median value between the scores would be the best possible option. This is mainly to take into account the small chance that some scores are given in an incorrect way. For instance, the score might have been given by accidentally clicking the



(a) 92 (b) 91 (c) 89 (d) 85 (e) 12 (f) 21 (g) 21 (h) 24

Fig. 3. The four paintings ranked highest for their aesthetic quality are marked by a green border ((a)-(d)) and the four paintings ranked lowest by a red border ((e)-(h)). The scores given to each painting is presented below each image. (a) Antonio Canaletto, 1738, (b) Antonio Canaletto, 1749, (c) Pieter Jansz Saenredam, 1648, (d) Dosso Dossi,1524, (e) Quentin Matsys, 1513, (f) Édouard Vuillard, 1900, (g) Ernst Kirchner, 1910, (h) Ernst Kirchner, 1920.

(a) 95 (b) 92 (c) 91 (d) 91 (e) 2 (f) 10 (g) 13 (h) 13

Fig. 4. The four paintings ranked highest for their beauty are marked by a green border ((a)-(d)) and the four paintings ranked lowest with a red border ((e)-(h)). The scores given to each painting is presented below each image. (a) Edmund C. Tarbell, 1892, (b) Antonio Canaletto, 1738, (c) Félix Ziem, 1850, (d) John Constable, 1816, (e) Quentin Matsys, 1513, (f) Ernst Kirchner, 1910, (g) Francisco Goya, 1812, (h) Édouard Vuillard, 1900.

score bar. Using the median scores provides us with a better chance to remove these outliers and achieve a more accurate score. Figures 3-7 represents the four highest rated paintings (marked by a green border) and the four lowest rated paintings (marked by a red border) for the first five properties introduced in Table 2. Figure 8 represents the distribution of the scores for each property. As shown in the figure, the median value of the subjective scores for all properties is around the mid-point of the score range. Note that the subjective scores cover a wide range of the score bar.

As mentioned before (Table 2), the observers were asked two more questions with regard to the familiarity of the paintings and the painter who created the painting. Results revealed that a majority of the paintings neither looked familiar nor did the observers know the painter (in both cases, 99% had a score of less than 10%). This finding suggests that the results for the other five properties cannot have been substantially influenced by familiarity of the observers with the paintings.

(a) 98 (b) 96 (c) 95 (d) 93 (e) 14 (f) 16 (g) 17 (h) 18

Fig. 5. The four paintings ranked highest for their liking of color are marked by a green border ((a)-(d)) and the four paintings ranked lowest with a red border ((e)-(h)). The scores given to each painting is presented below each image. (a) Edmund C. Tarbell, 1892, (b) P. C. Skovgaard, 1857, (c) Childe Hassam, 1913, (d) Antonio Canaletto, 1738, (e) Edgar Degas, 1890, (f) Isidre Nonell, 1903, (g) Pierre Puvis de Chavannes, 1881, (h) Ernst Kirchner, 1920.

(a) 95 (b) 95 (c) 94 (d) 93 (e) 11 (f) 13 (g) 13 (h) 14

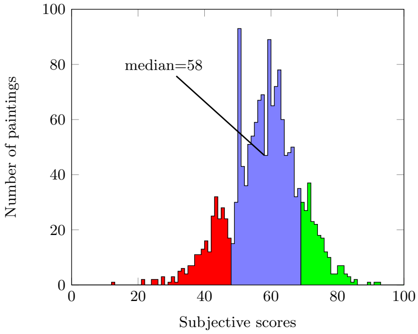
Fig. 6. The four paintings ranked highest for their liking of composition are marked by a green border ((a)-(d)) and the four paintings ranked lowest with a red border ((e)-(h)). The scores given to each painting is presented below each image. (a) Antonio Canaletto, 1738, (b) Johan Christian Dahl, 1839, (c) Viktor Vasnetsov, 1881, (d) John Singleton Copley, 1765, (e) Émile Bernard, 1892, (f) Paul Cézanne, 1877, (g) Isidre Nonell, 1903, (h) Marcus Gheeraerts the Younger, 1591.

(a) 97 (b) 93 (c) 93 (d) 93 (e) 3 (f) 7 (g) 10 (h) 11

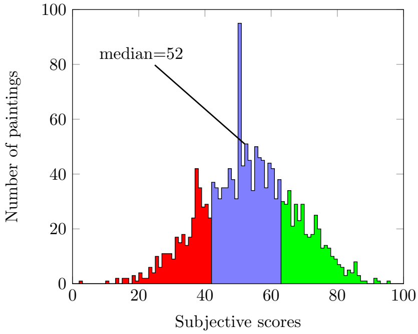
Fig. 7. The four paintings ranked highest for their liking of content are marked by a green border ((a)-(d)) and the four paintings ranked lowest with a red border ((e)-(h)). The scores given to each painting is presented below each image. (a) Johan Christian Dahl, 1839, (b) Childe Hassam, 1913, (c) Anton Mauve, 1887, (d) David Teniers the Younger, 1652, (e) Quentin Matsys, 1513, (f) Felice Boselli, 1690, (g) Jusepe de Ribera, 1621, (h) Abraham Staphorst, 1665.

4.2 Relationships Between Subjective Scores

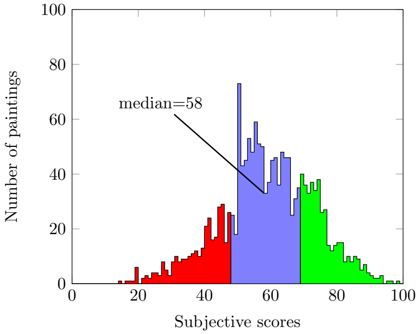
Next, we investigated the relationships between the subjective scores of the different properties by calculating the Spearman correlation coefficient.



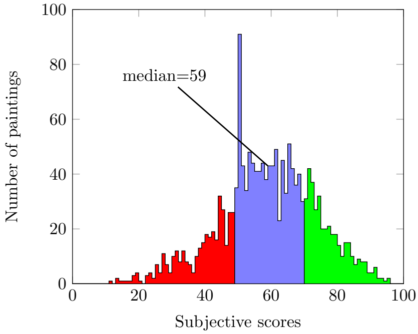
(a) aesthetic quality



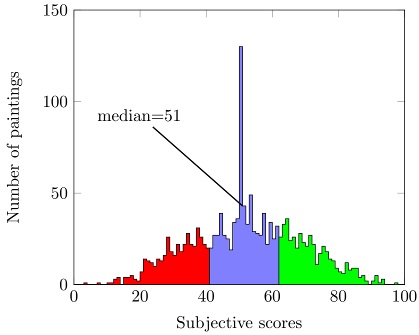
(b) beauty



(c) liking of color



(d) liking of composition



(e) liking of content

Fig. 8. Histograms representing the score distribution of the median score for different properties evaluated in the JenAesthetics subjective dataset. Images defined as having high quality are shown in green and images defined as having low quality in red (see Section 4.2). The blue values represent intermediate values with the median value indicated.

Table 4. ρ values calculated for Spearman correlation between subjective scores for different properties. All values are significantly different from zero ($p < 0.01$).

Properties	Aesthetic quality	Beauty	Liking of color	Liking of composition	Liking of content
Aesthetic quality	1				
Beauty	.7802	1			
Liking of color	.6676	.7237	1		
Liking of composition	.7114	.7642	.6506	1	
Liking of content	.6216	.8110	.5945	.7010	1

The following findings deserve further comments:

- The highest correlation is seen between liking of content and beauty. The fact that the content and subject matter of a painting plays a crucial role for how an observer evaluates a painting was mentioned previously in [1]. Other works emphasize this fact for other stimuli, such as [16] for webpages.
- The second highest correlation is between subjective aesthetic quality and beauty scores. Keeping in mind that the Oxford dictionary defines aesthetic as "concerned with beauty or the appreciation of beauty", such a high correlation is not a surprise.
- Previous studies have related different composition techniques such as the rule of thirds, the golden ratio, etc. to the beauty and aesthetic quality of paintings and photographs [3,8,18,20,26,35,38]. In the present study, the correlation between the liking of composition and beauty is the third highest correlation, and the correlation between aesthetic quality and the liking of composition is the fifth highest correlation.
- With regard to the liking of color, studies such as [2,25,26,31] have emphasized the importance of color on subjective aesthetic and beauty scores. This aspect was also seen in the correlation between the scores given to the liking of color and ratings for beauty and aesthetic quality.
- The correlation between the beauty scores and the mentioned three properties (liking of color, content, and composition) are among the highest (fourth, third, and first, respectively). In contrast, the correlation for aesthetic quality with the three properties is not as high as that of beauty (seventh, fifth, and ninth, respectively).

We also implemented a five-fold cross validation classifier using a linear SVM. This was done to enable users of the dataset to compare the present performance of a classifier with their own classifiers based on the subjective scores provided in the present work.

For the each property in our classification, we divide the images into two groups (high quality and low quality). The assigning of the groups is done based on the subjective score for each image. If the image has a subjective score greater than $Median(allscores) + 10$, the image will be labelled as high quality and if the subjective score is lower than $Median(allscores) - 10$, it will be labelled as low quality. The other remaining properties are used as features in our classifier. Average recognition rates of this classification procedure are listed in Table 5 for different scenarios. From this table we can conclude that:

– High classification rates were found between subjective aesthetic quality and beauty scores and the other three properties. This finding was previously seen for the correlation rates (Table 4) and supports the notion that these properties (liking of color, composition and content) are closely related to the aesthetic quality and beauty judgement (see Section 3.1).
– Using the subjective scores provided for the liking of color and content together in our feature vector resulted in high recognition rates for beauty. As mentioned above, a close relation of color liking and beauty perception has been previously pointed out in the literature.
– Similar to the correlation rates shown in Table 4, the lowest classification rate is for the case in which either the liking of color or the liking of content is used as a feature to classify the other property. This result is not surprising since the two mentioned properties are usually not related to one another in paintings.

5 Discussion and Future Work

In this paper, we present subjective ratings of the previously introduced JenAesthetics dataset. We hope that such a public dataset of paintings along with their subjective scores will provide a significant contribution to the computational aesthetic community. The lack of a publicly available subjective dataset of paintings has been mentioned numerous times in different publications and/or meetings. The subjective dataset comprises scores for five different properties (aesthetic, beauty, and liking of color, composition, and content). The scores were gathered by performing 190 observation sessions by 129 observers. The results show that the properties assessed are highly correlated with one another. It was interesting to see that the subjective scores related to the liking of color, composition and content had a higher correlation with beauty scores than with aesthetic scores. This finding shows that a high aesthetic quality of a painting does not necessarily mean that the color, content, or composition are pleasing to the observer as well. The fact that the subjective scores for beauty and aesthetic quality were highly correlated confirms findings from previous studies [1–3,8,16,18,20,25,26,31,35,38].

Compared to previous datasets, the JenAesthetics subjective dataset contains a larger number of paintings and covers a wider range of different subject matters, styles and art periods. It also evaluates different properties providing the user with many different scenarios to test and evaluate.

In the future, we will increase the number of images assessed in our subjective dataset. Also, we are planning to extract additional features from the images. Finding the relationship between scores provided in different subgroups, such as males and females, could be an interesting topic of investigation.

Table 5. Average classification rate of five-fold cross validation using different subjective scores as features in a linear SVM. In the first column, A corresponds to subjective scores for aesthetic quality, B for beauty, CL for liking of color, CM for liking of composition, and CN for liking of content.

Feature vector	Aesthetic	Beauty	Color	Composition	Content
A	–	76.37%	73.50%	74.09%	71.99%
B	76.53%	–	76.16%	78.10%	82.77%
CL	72.82%	76.40%	–	73.56%	70.19%
CM	74.59%	77.07%	73.77%	–	76.63%
CN	73.16%	79.80%	69.19%	73.87%	–
A, B	–	–	77.68%	80.63%	82.23%
A, CL	–	82.62%	–	76.19%	75.19%
A, CM	–	82.74%	74.73%	–	77.27%
A, CN	–	85.66%	74.61%	79.79%	–
B, CL	80.96%	–	–	80.02%	82.09%
B, CM	81.54%	–	77.81%	–	84.20%
B, CN	80.23%	–	77.62%	79.86%	–
CL, CM	78.60%	82.36%	–	–	78.13%
CL, CN	77.28%	86.11%	–	78.92%	–
CM, CN	76.46%	84.75%	73.28%	–	–
A, B,CL	–	–	–	80.47%	83.05%
A, B, CM	–	–	77.96%	–	84.76%
A, B, CN	–	–	77.68%	81.18%	–
A, CL, CM	–	85.04%	–	–	79.00%
A, CL, CN	–	88.30%	–	80.59%	–
A, CM, CN	–	87.90%	77.33%	–	–
B, CL, CM	81.56%	–	–	–	84.07%
B, CL, CN	81.03%	–	–	80.23%	–
B, CM, CN	81.59%	–	77.85%	–	–
CL, CM, CN	79.07%	87.42%	–	–	–
A, B, CL, CM	–	–	–	–	84.77%
A, B, CL, CN	–	–	–	81.35%	–
A, B, CM, CN	–	–	78.22%	–	–
A, CL, CM, CN	–	88.68%	–	–	–
B, CL, CM, CN	81.64%	–	–	–	–

References

1. Amirshahi, S.A., Denzler, J., Redies, C.: JenAesthetics-a public dataset of paintings for aesthetic research. Computer Vision Group, University of Jena Germany, Tech. rep. (2013)
2. Amirshahi, S.A., Hayn-Leichsenring, G.U., Denzler, J., Redies, C.: Color: A crucial factor for aesthetic quality assessment in a subjective database of paintings. In: 12th Congress of the International Colour Association (AIC). Newcastle, UK (July 2013)
3. Amirshahi, S.A., Hayn-Leichsenring, G.U., Denzler, J., Redies, C.: Evaluating the rule of thirds in photographs and paintings. Art & Perception **2**(1–2), 163–182 (2014)
4. Amirshahi, S.A., Koch, M., Denzler, J., Redies, C.: PHOG analysis of self-similarity in aesthetic images. In: IS&T/SPIE Electronic Imaging, pp. 82911J–82911J. International Society for Optics and Photonics (2012)
5. Amirshahi, S.A., Redies, C., Denzler, J.: How self-similar are artworks at different levels of spatial resolution? In: Proceedings of the Symposium on Computational Aesthetics, pp. 93–100. ACM (2013)
6. Bhattacharya, S., Sukthankar, R., Shah, M.: A framework for photo-quality assessment and enhancement based on visual aesthetics. In: Proceedings of the International Conference on Multimedia, pp. 271–280. ACM (2010)
7. Condorovici, R.G., Florea, C., Vrânceanu, R., Vertan, C.: Perceptually-inspired artistic genre identification system in digitized painting collections. In: Kämäräinen, J.-K., Koskela, M. (eds.) SCIA 2013. LNCS, vol. 7944, pp. 687–696. Springer, Heidelberg (2013)
8. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Studying aesthetics in photographic images using a computational approach. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 288–301. Springer, Heidelberg (2006)
9. Datta, R., Li, J., Wang, J.Z.: Algorithmic inferencing of aesthetics and emotion in natural images: An exposition. In: IEEE 15th International Conference on Image Processing (ICIP), pp. 105–108. IEEE (2008)
10. Demetriou, M.L., Hardeberg, J.Y., Adelmann, G.: Computer-aided reclamation of lost art. In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) ECCV 2012 Ws/Demos, Part I. LNCS, vol. 7583, pp. 551–560. Springer, Heidelberg (2012)
11. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 248–255. IEEE (2009)
12. Hoenig, F.: Defining computational aesthetics. In: Proceedings of the First Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging, pp. 13–18. Eurographics Association (2005)
13. International Telecommunication Union: Recommendation ITU-R BT.500-11: Methodology for the subjective assessment of the quality of television pictures. Tech. rep., International Telecommunication Union/ITU Radiocommunication Sector (2009)
14. Ishihara, S.: Test for colour-blindness. Hongo Harukicho, Tokyo (1917)
15. JenAesthetics: JenAesthetics dataset (2013). http://www.inf-cv.uni-jena.de/en/jenaesthetics
16. Karvonen, K.: The beauty of simplicity. In: Proceedings on the 2000 Conference on Universal Usability, pp. 85–90. ACM (2000)

17. Ke, Y., Tang, X., Jing, F.: The design of high-level features for photo quality assessment. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 419–426. IEEE (2006)
18. Li, C., Chen, T.: Aesthetic visual quality assessment of paintings. IEEE Journal of Selected Topics in Signal Processing **3**(2), 236–252 (2009)
19. Luo, W., Wang, X., Tang, X.: Content-based photo quality assessment. In: IEEE International Conference on Computer Vision (ICCV), pp. 2206–2213. IEEE (2011)
20. Mai, L., Le, H., Niu, Y., Liu, F.: Rule of thirds detection from photograph. In: IEEE International Symposium on Multimedia (ISM), pp. 91–96. IEEE (2011)
21. Mallon, B., Redies, C., Hayn-Leichsenring, G.U.: Beauty in abstract paintings: Perceptual contrast and statistical properties. Frontiers in Human Neuroscience **8**(161) (2014)
22. Müller, H., Clough, P., Deselaers, T., Caputo, B. (eds.): Imageclef - Experimental evaluation in visual information retrieval, vol. 32. The Information Retrieval Series (2010)
23. Murray, N., Marchesotti, L., Perronnin, F.: Ava: A large-scale database for aesthetic visual analysis. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2408–2415. IEEE (2012)
24. Oncu, A.I., Deger, F., Hardeberg, J.Y.: Evaluation of digital inpainting quality in the context of artwork restoration. In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) ECCV 2012 Ws/Demos, Part I. LNCS, vol. 7583, pp. 561–570. Springer, Heidelberg (2012)
25. Palmer, S.E., Schloss, K.B.: An ecological valence theory of human color preference. Proceedings of the National Academy of Sciences USA **107**(19), 8877–8882 (2010)
26. Palmer, S.E., Schloss, K.B., Sammartino, J.: Visual aesthetics and human preference. Annual Review of Psychology **64**, 77–107 (2013)
27. Peirce, J.W.: PsychoPy-Psychophysics software in Python. Journal of Neuroscience Methods **162**(1–2), 8–13 (2007). http://www.sciencedirect.com/science/article/pii/S0165027006005772
28. Redies, C., Amirshahi, S.A., Koch, M., Denzler, J.: PHOG-Derived aesthetic measures applied to color photographs of artworks, natural scenes and objects. In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) ECCV 2012 Ws/Demos, Part I. LNCS, vol. 7583, pp. 522–531. Springer, Heidelberg (2012)
29. Redies, C., Hänisch, J., Blickhan, M., Denzler, J.: Artists portray human faces with the fourier statistics of complex natural scenes. Network: Computation in Neural Systems **18**(3), 235–248 (2007)
30. Redies, C., Hasenstein, J., Denzler, J., et al.: Fractal-like image statistics in visual art: similarity to natural scenes. Spatial Vision **21**(1–2), 137–148 (2007)
31. Schloss, K.B., Palmer, S.E.: An ecological valence theory of human color preferences. Journal of Vision **9**(8), 358–358 (2009)
32. Seshadrinathan, K., Soundararajan, R., Bovik, A.C., Cormack, L.K.: A subjective study to evaluate video quality assessment algorithms. In: IS&T/SPIE Electronic Imaging, pp. 75270H–75270H. International Society for Optics and Photonics (2010)
33. Seshadrinathan, K., Soundararajan, R., Bovik, A.C., Cormack, L.K.: Study of subjective and objective quality assessment of video. IEEE Transactions on Image Processing **19**(6), 1427–1441 (2010)
34. Sheikh, H.R., Sabir, M.F., Bovik, A.C.: A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Transactions on Image Processing **15**(11), 3440–3451 (2006)

35. Svobodova, K., Sklenicka, P., Molnarova, K., Vojar, J.: Does the composition of landscape photographs affect visual preferences? The rule of the golden section and the position of the horizon. Journal of Environmental Psychology **38**, 143–152 (2014)
36. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)
37. Liu, X., Pedersen, M., Hardeberg, J.Y.: CID:IQ – A New Image Quality Database. In: Elmoataz, A., Lezoray, O., Nouboud, F., Mammass, D. (eds.) ICISP 2014. LNCS, vol. 8509, pp. 193–202. Springer, Heidelberg (2014)
38. Xue, S.F., Lin, Q., Tretter, D.R., Lee, S., Pizlo, Z., Allebach, J.: Investigation of the role of aesthetics in differentiating between photographs taken by amateur and professional photographers. In: IS&T/SPIE Electronic Imaging, pp. 83020D–83020D. International Society for Optics and Photonics (2012)
39. Yanulevskaya, V., Uijlings, J., Bruni, E., Sartori, A., Zamboni, E., Bacci, F., Melcher, D., Sebe, N.: In the eye of the beholder: employing statistical analysis and eye tracking for analyzing abstract paintings. In: Proceedings of the 20th ACM International Conference on Multimedia, pp. 349–358. ACM (2012)