Visual Properties Differentiating Art from Real Scenes

Alex Leykin, Florin Cutzu, and Riad Hammoud
email: [florin, oleykin, rhammoud]@cs.indiana.edu
Dept of Computer Science
Indiana University
Bloomington IN 47405
USA

October 18, 2002

Contents

Abstract

We addressed the problem of automatically differentiating photographs of real scenes from photographs of paintings. We found that photographs differ from paintings in their color, edge, and texture properties. Based on these features, we trained a classifier to separate a database of 12,000 images downloaded from the web into photographs and paintings. Single features result in $70-83\%$ performance, whereas with a neural net classifier correct rates were around 92%.

1 Introduction

1.1 Problem statement

The goal of the present work was the determination of the image features distinguishing photographs of real-world, three-dimensional, scenes from (photographs of) paintings and the development of a classifier system for their automatic differentiation.

In the context of this paper, the class "painting" included not only conventional canvas paintings, but also frescoes and murals (see Figure 1). Line (pencil or ink) drawings (see Figure 1) as well as computer-generated images were excluded. No restrictions were imposed on the historical period or on the style of the painting.

The class "photograph" included exclusively color photographs of three-dimensional real-world scenes.

The problem of distinguishing paintings from photographs is non-trivial even for a human observer, as can be appreciated from the examples shown

in Figure 2. We note that the painting in the bottom right corner was classified as photograph by our algorithm.

In fact, photographs can be considered as a special subclass of the paintings class: photographs are photorealistic paintings. Thus, the problem can be posed more generally as determining the degree of perceptual photorealism of an image. Given an input image, the classifier proposed in this paper outputs a number $\in [0, 1]$ which can be interpreted as a measure of the degree of photorealism of the image.

From a theoretical standpoint, the problem of separating photographs from paintings is interesting because it constitutes a first attempt at revealing the features of real-world images that are mis-represented in hand-crafted images. From a practical standpoint, our results are useful for the automatic classification of images in large electronic-form art collections, such as those maintained by many museums. A special application is in distinguishing pornographic images from nude paintings. Distinguishing paintings from photographs is important for web browser blocking software, which currently blocks not only pornography (photographs) but also artistic images of the human body (paintings).

1.2 Related work

To our knowledge, the present study is the first to address the problem of photograph-painting discrimination. This problem is related thematically to other work on broad image classification: city images vs. landscapes [12], indoor vs. outdoor [11], and photographs vs. graphics [7] differentiation.

Distinguishing photographs from paintings is, however, more difficult than the above classifications due to the generality of the problem. One difficulty is that that are no constraints on the *image content* of either class, such as those successfully exploited in differentiating city images from landscapes or indoor from outdoor images.

The problem of distinguishing computer-generated graphics from photographs is closest to the problem considered here, and their relation will be discussed in more detail in Section 5. At this point, it suffices to note that the differences between (especially realistic) paintings and photographs are subtler than the differences between graphics and photographs; in addition, the definition of computer-generated graphics used in [7] is restrictive and allowed the use of powerful constraints that are not applicable to the paintings class.

2 The image set

The image set used in this study consisted of 6000 photographs and 6000 paintings. The definition of painting and photograph in the context of this paper was given in Section 1.1.

The paintings were obtained from two main sources. 3000 paintings were downloaded from the Indiana University Department of the History of Art DIDO Image Bank[1], 2000 were obtained from the Artchive art database[2], and 1000 from a variety of other web sites.

2000 photographs were downloaded from freefoto.com, and the rest were downloaded from a variety of other web sites.

The paintings in our database were of a wide variety of artistic styles and historical periods, from Byzantine Art and Renaissance to Modernism (cubism, surrealism, pop art, etc).

The photographs were also very varied in content–including animals, humans, city scenes and landscapes, indoor scenes.

Image resolution was typical of web-available images. Mean image size was for paintings 534×497 pixels and standard deviation 171×143 pixels. For photographs mean image size was 568×506 pixels and standard deviation 144×92 pixels.

Certain rules were followed when selecting the images included in the database: (1) no monochromatic images were used; all our images had a color resolution of 8 bits per color channel (2) frames and borders were removed (3) no photographs altered by filters or special effects were included (4) no computer generated images were used (5) no images with large areas overlayed with text were used.

3 Distinguishing features

Based upon the visual inspection of a large number of photographs and paintings, we defined several image features for which paintings and photographs differ significantly.

Four features, defined in Sections 3.1–3.4 are color-based, and one is image intensity-based (Section 3.8).

[1] www.dlib.indiana.edu/collections/dido

[2] The Artchive CD-ROM is available from www.artchive.com

3.1 Color edges vs. intensity edges

We observed that while the removal of color information (conversion to gray-scale) leaves most edges in photographs intact, it eliminates many of the perceptual edges in paintings. More generally, it appears that the removal of color eliminates more visual information from a painting than from a photograph of a real scene.

In a photograph of a real-world scene, the variation of image intensity is substantial and systematic, being the the result of the interaction of light with surfaces of various reflectances and orientations. In the real world, color is not essential for recognition and navigation and color-blind visual systems can function quite well.

Painters, however, use primarily color—rather than systematic changes of image intensity—to represent different objects and object regions.

Edges are essential image features, in that they convey a large amount of visual information. Edges in photographs are of many different types: occlusion edges, edges induced by surface property (texture or color) changes, cast shadow edges. In most cases, however, the surfaces meeting at the edge have different material or geometrical (orientation) properties, resulting in a difference in the intensity (and possibly color) of the reflected light. One exception to this rule is represented by edges delimiting regions painted in different colors on a flat surface—as on billboards or in paintings on building walls for example; in effect, such cases are paintings within photographs of real world scenes. On the contrary, in paintings, adjacent regions tend to differ in their hue, change often not accompanied by an edge-like change in image intensity.

The above observations led to the following hypotheses:

(1) Perceptual edges in photographs are, largely, intensity edges. These intensity edges tend to be at the same time color edges and there are few "pure" color edges – color, not intensity edges.

(2) Many of the perceptual edges in paintings are pure color edges, as they result from color changes that are not accompanied by concomitant edge-like intensity changes.

A quantitative criterion was developed. Consider a color input image—painting or photograph.

The intensity edges were obtained by converting the image to gray-scale and applying the Canny edge detector [13].

Then, image intensity information was removed by dividing the R, G, and B image components by the image intensity at each pixel, resulting in normalized RGB components: $R_n = \frac{R}{I}, G_n = \frac{G}{I}, B_n = \frac{B}{I}$, where $I \approx$

$0.3R + 0.6G + 0.1B$ is image intensity.

The color edges of the resulting "intensity-free" color image were determined applyn the Canny edge detector to the three color channels and fusing the resulting edges.

Two type of edge pixels were then determined, as follows:

(1) The edge pixels that were intensity but not color edge (pure intensity edge pixels). Hue does not change substantially across a pure intensity edge. For a given input image, E_c denotes the number of pure intensity-edge pixels divided by the total number of edge pixels:

$$E_g = \frac{\text{\# pixels: intensity, not color edge}}{\text{total number of edge pixels}}.$$

Our hypothesis is that E_g is larger for photographs.

(2) The edge pixels that are color edge but not intensity edge (pure color edge pixels). Hue, but not image intensity, changes across a pure color edge. Let E_g denote the proportion of pure color-edge pixels :

$$E_c = \frac{\text{\# pixels: color, not intensity edge}}{\text{total number of edge pixels}}.$$

Our hypothesis is that E_c is larger for paintings.

3.1.1 Single-feature discrimination performance: finding the optimal threshold

We determined the discrimination power of the two edge-derived features, considered separately.

The feature under consideration was measured for all photographs and all paintings in the database, and a threshold value, optimizing the separation between the two classes, was determined.

The optimal threshold was chosen so that it *minimized the maximum* of the two misclassification rates—for photographs and for paintings. Note that choosing the threshold so that it maximizes the total number of correctly classified images, although possibly yielding more correctly classified images, does not ensure balanced error rates for the two classes.

Also note that using a single threshold for discriminating between two classes in 1-D feature space is only the simplest method; a more general method would employ multiple thresholds, resulting in more than one interval per class.

The painting-photograph discrimination results, using edge features, are listed in Table 1.

As expected, paintings have more pure-color edges, and photographs have more pure-intensity edges. E_g is more discriminative than E_c.

E_c and E_g are not independent features: as can be expected from their definition, they are negatively correlated to a significant extent. The Pearson correlation coefficients of E_c and E_g are as follows: -0.80 over the photograph set, -0.74 over the painting set, -0.79 over the entire image database.

Given the strong correlation between E_c and E_g, the superior discrimination power of E_g (see Table 1), we decided to to discard E_c and employ E_g as the sole edge-based feature.

3.1.2 Intensity edges in paintings and photographs are structurally similar

We examined the spatial variation of image intensity in the vicinity of intensity edges in paintings and photographs. The intensity edges were determined by applying the Canny edge detector to both paintings and photographs followed their conversion to gray-scale.

In one experiment we examined the one-dimensional change of image intensity along a direction orthogonal to the intensity edge (i.e. along the image gradient), on a distance of 20 pixels of either side of the edge. We did not find significant differences between paintings and photographs in the shape of these image intensity profiles.

In a second experiment ellipsoidal Gabor kernels were positioned on the intensity edges, such that the long axis of the Gabor ellipse coincided with the intensity edge. Ten Gabor kernels of sizes from 5 to 25 pixels were employed, resulting in a 10-dimensional Gabor response vector. We did not find significant differences in the Gabor filter output between paintings and photographs.

This negative finding has to be interpreted with caution — it is possible that the differences between intensity edges in paintings and photographs are not observable at the modest resolutions of our image set.

3.2 Spatial variation of color

Our observations indicated that color changes to a larger extent from pixel to pixel in paintings than in photographs. This difference was quantified as follows.

The color of a pixel is determined by the ratios of its red, green and blue values, in other words by the orientation of its RGB vector. The norm of

this vector is not relevant for our purposes. Given an input image, its R, G and B channels were normalized by division by the norm of the RGB vector at each pixel.

Each of the thus-normalized R, G, and B-channel images can be viewed as a function (a surface) defined over the image plane. Consider a location in the image plane — a pixel. If in the neighborhood of this location the R, G, and B surfaces are parallel or nearly so, there is no substantial color change in that neighborhood. If there is a non-zero angle among any two of these surfaces, color changes qualitatively. At each pixel, for each of the R, G, and B surfaces, we determined the orientation of the plane that best fitted (in a least-squares sense) a 5×5 neighborhood centered on the pixel of interest, in the R, G, and B domains respectively. Thus, at each pixel, one obtains three normals: n_R, n_G, n_B, one for each color channel. The sum of the areas of the facets of the pyramid determined by these normals was taken as a measure of the local variation of color around the pixel.

Let R denote the average of this quantity taken over all image pixels. R should be, on the average, larger for paintings than for photographs.

Discrimination performance We determined the photograph-painting discrimination performance using R as the sole feature and an optimal threshold for R, computed as described in Section 3.1.1.

The miss rate rate for paintings was 36.14, the miss rate for photographs was 36.13, with most paintings above the threshold and most photographs below the threshold.

3.3 Number of unique colors

Paintings appear to contain more unique colors, i.e., to have a larger color palette than photographs. We used this characteristic to help differentiate between the two image classes.

For all images in our database, the color resolution was of 256 levels for each color channel. Thus, there are 256^3 possible colors, a number much larger than the number of pixels in a typical image.

Given an input image the number of unique colors was determined by counting the distinct RGB triplets. To reduce the impact of noise, a color triplet was counted only if it appeared in more than 10 of the image pixels.

The number of unique colors was normalized by the total number of pixels, resulting in a measure, denoted U, of the richness of the color palette of the image. U should be, on the average, for paintings than for photographs.

Discrimination performance We determined the photograph-painting discrimination performance using R as the sole feature and an optimal threshold for R, computed as described in Section 3.1.1.

The miss rate rate for paintings was 37.40, the miss rate for photographs was 37.413, with most paintings being above the threshold and most photographs being below the threshold.

3.4 Pixel saturation

We observed that paintings tend to contain a larger percentage of pixels with highly saturated colors than photographs in general, and photographs of natural objects and scenes in particular. Photographs, on the other hand, contain more unsaturated pixels than paintings.

This can be seen in Figure 3, which displays the mean saturation histograms derived from all paintings and all photographs in our datasets.

These characteristics were captured quantitatively. The input images were transformed from RGB to HSV (hue-saturation-value) color space, and their saturation histogram were determined, using a fixed number of bins, n. Consider the ratio, S, between the count in the highest bin (bin n) and the lowest bin (bin 1): S measures the ratio between the number of highly saturated and highly unsaturated pixels in the image. Our hypothesis was that S is, on the average, larger for paintings than for photographs.

Discrimination performance We determined the photograph-painting discrimination performance using S as the sole feature and an optimal threshold for S, computed as described in Section 3.1.1.

The miss rate rate for paintings was 37.93, the miss rate for photographs was 37.923, with most paintings being above the threshold and most photographs being below the threshold.

Hue and RGB histograms We found experimentally that the hue (or full RGB) histograms are quite useful in distinguishing between photographs and paintings; for example, the hue corresponding to the color of the sky was quite characteristic of outdoor photographs. However, since hue is image content-dependent to a large degree, we decided against using hue histograms (or RGB histograms) in our classifiers, as our intention was to distinguish paintings from photographs in a *image content-independent manner*.

3.5 Relations among the scalar-valued features: E_g, U, R, S

In the preceding section we introduced four simple, scalar-valued image features. The immediate question arises whether these features capture genuinely different image properties or there is substantial redundancy in their encoding of the images.

Two measures of redundancy were measured: pairwise feature correlation and the singular values of the feature covariance matrix.

3.5.1 Feature correlation

We calculated the Pearson correlation coefficients ρ for all pairs of scalar-valued color-based features, considering the paintings and photographs image sets separately. The correlation coefficients, shown in Table 2 separately for paintings and photographs, indicate that the different color-based features were not correlated significantly.

3.5.2 Eigenvalues of feature covariance matrix

Consider a d-dimensional feature space, and a "cloud" of n points in this space. If all d singular values of the $d \times d$ covariance matrix of the point cloud are significant (compared to the sum of all singular values), it follows that the data points are not confined to some linear subspace[3] of the d-dimensional feature space; in other words, there are no linear dependencies among the d features.

In our case, we have a 4-dimensional feature space corresponding to the color-based features described above. We computed three 4×4 covariance matrices, one for the paintings data set, one for the photograph data set, and one for the joint photograph-paintings data set. All covariance matrices were calculated on centered data, i.e. each feature was centered on its mean value. The eigenvalues of the paintings covariance matrix are: 0.16, 0.06, 0.01, 0.004, 0.002. The eigenvalues of the photograph covariance matrix are: 0.13, 0.03, 0.02, 0.002, 0.001.

Two observations can be made. First, the smallest eigenvalue is in both cases significant, indicating that the point clouds are truly four-dimensional. Second, the eigenvalues of the paintings-derived covariance matrix are significantly larger than for the photograph data set, indicating that there is more variability in the paintings data set.

[3]However, the points may be confined to a non-linear subspace–for example the surface of a sphere (a 2-D subspace) in 3-D space.

3.5.3 Principal components

For visualization purposes, we determined the principal components of the common painting and photograph data set encoded in the space of the four simple color-based features described above. Figure 4 displays separately the painting and the photograph subsets in the same space–the space spanned by the first two principal components.

The examination of Figure 4 leads to the interesting observation that the photographs overlap a subclass of the paintings: the photograph data set (at least in the space spanned by the first two principal components) coincides with the right "lobe" of the paintings point cloud. This observation is in accord with the larger variability of the paintings class indicated by the eigenvalues listed in the preceding section, and with the observation that photographs can be construed as extremely realistic paintings.

3.6 Classification in the space of the scalar-valued features

We used a neural network classifier to perform painting-photograph discrimination in the space of the scalar-based features. A perceptron with six sigmoidal units in its unique hidden layer was employed.

The performance of this classifier was evaluated as follows. We partitioned the paintings and photographs sets into 6 parts (non-overlapping subsets) of 1000 elements each. By pairing all photograph parts with all painting parts, 36 training sets were generated. Thus, a training set consisted of 1000 paintings and 1000 photographs, and the corresponding test set consisted of 5000 paintings and 5000 photographs. 36 networks were trained and tested, one for each training set. Due to the small size of the network, the convergence of the backpropagation calculation was quite rapid in almost all cases, and usually, four to seven re-initializations of the optimization were sufficient for deriving an effective network.

On the average, the networks correctly classified 71.1% of the photographs and 72.2% of the paintings in the test set, with a standard deviation of 4%, respectively, 5%.

3.7 Pixel distribution in RGBXY space

An image pixel is a point in 3-D RGB space, and the image is a point cloud in this space. The shape of this point cloud depends on the color richness of the image. The RGB clouds of color-poor images (photographs, mostly) are restricted to subspaces of the 3-D space, having the appearance of cylinders—indicating that color variability in the image is essentially

one-dimensional or planes—indicating that color variability in the image is essentially bi-dimensional. The RGB clouds of color-rich images (paintings, mostly) are fully 3-D and cannot be approximated well by a 1-D or 2-D subspace.

The linear dimensionality of the RGB cloud is summarized by the singular values of the 3×3 covariance matrix of the RGB point cloud. If the RGB cloud is essentially one-dimensional (cylindrical), the second and the third singular values are negligible compared to the first. If the RGB cloud is essentially two-dimensional (a flat point cloud), the third singular value is negligible.

One can enhance this representation by adding the two spatial coordinates, x and y to the RGB vector of each image pixel, resulting in a 5-dimensional, joint color-location space we call RGBXY. An image is a cloud of points in this space. The singular values $s_{1,2,3,4,5}$ of the 5×5 covariance matrix of the RGBXY point cloud describe the variability of the image pixels in both color space as well as across the plane of the image. Typically, paintings use both a larger color palette and have larger spatial variation of color, resulting in larger singular values for the covariance matrix.

The above considerations led to representing each image by a 5-dimensional vector $\mathbf{s}$ of the singular values of its RGBXY pixel covariance matrix.

3.7.1 Paintings and photographs in RGBXY space

For visualization purposes, we determined the principal components of the common painting and photograph data set encoded in the space of the five singular values of the RGBXY covariance matrix. Figure 5 displays separately the painting and the photograph subsets in the same space–the space spanned by the first two principal components.

The examination of Figure 5 reconfirms the previously-made observation that photographs appear to be a special case of paintings: the photograph point cloud has less variance and partially overlaps (at least in the space spanned by the first two principal components) with a portion of the paintings point cloud. This observation is also supported by the larger singular values of the painting point cloud (5.03, 0.21, 0.1, 0.08, 0.002) compared to those of the photograph point cloud (4.15, 0.12, 0.08, 0.03, 0.003).

3.7.2 Classification using the singular values of the RGBXY covariance matrix

As explained in the preceding section, the singular values of the covariance matrix of the image pixels represented in RGBXY space summarize the spatial variation of image color.

We used a neural network classifier to perform painting-photograph discrimination in the 5-dimensional space of the singular values. A perceptron with six sigmoidal units in its unique hidden layer was employed.

The performance of this classifier was evaluated as follows. We partitioned the paintings and photographs into 6 parts (non-overlapping subsets) of 1000 elements each. By pairing all photograph parts with all painting parts, 36 training sets were generated. Thus, a training set consisted of 1000 paintings and 1000 photographs, and the corresponding test set consisted of 5000 paintings and 5000 photographs. 36 networks were trained and tested, one for each of training set.

On the average, the networks correctly classified 81.12% of the photographs and 81.26% of the paintings in the test set, with a standard deviation of 3%, respectively, 3% percent.

The convergence of the backpropagation calculation was quite rapid in almost all cases, and usually, four to seven re-initializations of the optimization were sufficient for deriving a well-performing network.

3.8 Texture

All of the features described in the preceding section use color to distinguish between paintings and photographs. To increase discrimination accuracy, it is desirable to derive a feature that is color-independent—that is, a feature that can be computed from image intensity alone. Image texture is an obvious choice.

Following the idea described in [4] we used the statistics of Gabor filter outputs to encode the texture properties of the filtered image. Gabor filters can be considered orientation and scale-adjustable edge detectors. The mean and the standard deviation of the outputs of Gabor filters of various scales and orientations can be used to summarize the underlying texture information [4].

Our Gabor kernels were circularly symmetric, and were constrained to have the same number of oscillations within the Gaussian window at all frequencies—consequently, higher frequency filters had smaller spatial extent. We used four scales and four orientations (0, 90, 45, 135 degrees),

resulting in 16 Gabor kernels. The images were converted to gray-scale and convolved with the Gabor kernels. For each image we calculated the mean and the standard deviation of the Gabor responses across image locations for each of the 16 scale-orientation value pairs, obtaining a feature vector of dimension 32.

To estimate their painting-photograph discriminability potential, we calculated the means and the standard deviations of the features over all paintings and all photographs. Figure 7 displays the results. Interestingly, photographs tend to have more energy at horizontal and vertical orientations at all scales, while paintings have more energy at diagonal (45 and 135 degrees) orientations.

3.8.1 Classification using the Gabor feature vectors

As explained in the preceding section, the directional and scale properties of the texture of images were encoded by 32-dimensional feature vectors. We used a neural network to perform painting-photograph discrimination in this space. A perceptron with five sigmoidal units in its unique hidden layer was employed. Classifier performance was evaluated as follows. We partitioned the paintings and photographs into 6 parts (non-overlapping subsets) of 1000 elements each. By pairing all photograph parts with all painting parts, 36 training sets were generated. Thus, a training set consisted of 1000 paintings and 1000 photographs, and the corresponding test set consisted of 5000 paintings and 5000 photographs. 100 networks were trained and tested, one for each of training set.

On the average, the networks correctly classified 78% of the photographs and 79% of the paintings in the test set, with a standard deviaton of 4%, respectively, 5% percent.

The convergence of the backpropagation calculation was quite rapid in almost all cases, and usually, two to five re-initializations were sufficient for obtaining a good network.

4 Discrimination using multiple classifiers

In the preceding sections we described the classification performance of three classifiers: one for the space of the scalar-valued features (Section 3.6), one for the space of the singular values of the RGBXY covariance matrix (Section 3.7.2) and one for the space of the Gabor descriptors (Section 3.8.1).

We found that the most effective method of combining these classifiers is to simply average their outputs–the "committees" of neural networks idea

(see for example [14]). An individual classifier outputs a number between 0 (perfect painting) and 1 (perfect photograph). Thus, if for a given input image, the average of the outputs of the three classifiers was ≤ 0.5, it was classified as a painting; otherwise it was considered a photograph.

4.1 Painting-photograph discrimination performance

To evaluate the performance of this combination of the individual classifiers, we partitioned the painting and photograph sets into 6 equal parts each. By pairing all photograph parts with all painting parts, 36 training sets were generated. A training set consisted of 1000 paintings and 1000 photographs, and the corresponding test set consisted of the remaining 5000 paintings and 5000 photographs. Each of the three classifiers were trained on the same training set, and their individual and their average performance was measured on the same test set. This procedure was repeated for all available training and testing sets.

Classifier performance is described in Table 3. The averaged (combined) classifier significantly outperforms the individual classifiers for both paintings and photographs. This improvement is to expected, since each classifier works in a different feature space.

4.2 Illustrating classifier performance

In the following two sections, we illustrate with examples the performance of our classifier. We selected the best-performing classifier from the set of classifiers from which the statistics Table 3 were derived, and we studied its performance on its test set. The following two sections illustrate classifier behavior.

4.2.1 Typical photographs and paintings

For an input image, the output of the combined classifier is a number $\in [0, 1]$, 0 corresponding to a perfect painting and 1 to a perfect photograph; in other words, classifier output can be interpreted as the degree of photorealism of the input image.

In this section, we illustrate the behavior of the combined classifier by displaying images for which classifiers output was very close to 0 (≤ 0.1) or to 1 (≥ 0.9). Thus, these are images that our classifier considers to be typical paintings and photographs. We note that the error rate is very low (under 4%) for these output values.

Figures 5, 5 display several typical paintings. Note the variety of styles of these paintings: one is tempted to conclude that the features the classifiers use capture the essence of "paintingness" of an image.

Figures 5 display examples of typical photographs. We note that these tend to be typical, not artistic or otherwise unusual photographs.

4.2.2 Misclassified images

The mistakes made by our classifier were interesting, in that they seemed to reflect the degree of perceptual photorealism of the input image.

Figures 5, 5, 5 display paintings that were incorrectly classified as photographs. Note that most of these incorrectly classified paintings look quite photorealistic at a local level, even if their *content* is not realistic.

Figures 5, 5, 5 display photographs that were incorrectly classified as paintings. These photographs correspond, by and large, to vividly colored objects—which sometimes are painted 3-D objects— or to blurry or "artistic" photographs, or to photographs take under unusual illumination conditions.

5 Discussion

We presented an image classification system that discriminates paintings from photographs. This image classification problem is challenging and interesting, as it is very general and must be performed in image-content-indpendent fashion. Using low-level image features, and a relatively small training set, we achieved discrimination performance levels of over 90%.

Most of our features use color in one way or another. The Gabor feature are the only ones that use exclusively image intensities, and taken in isolation are not sufficient for accurate discrimination. Thus, color is critical for the good performance of our classifier. This appears to be different from human classification, since human can effortlessly discriminate paintings from photgraphs in gray-scale images. However, it is possible that human classification relies heavily on image content, and thus is not affected by loss of color information. To elucidate this point, we are planning to conduct psychophysical experiments on scrambled *gray-level* images. If the removal of color information affects the photorealism ratings significantly, it will mean that color is critical for human observers also.

It is easy to convince oneself that reducing image size (by smoothing and subsampling) makes the perceptual painting/photograph discrimination

more difficult if the paintings have "realistic" content. Thus, it is reasonable to expect that the discrimination performance of our classifier will also improve with increasing image resolution—hypothesis that we are planning to verify in future work.

In our study we employed images of modest resolution, typical for web-available images. Certain differences between paintings and photographs might be observable only at high resolutions. Specifically, although we did not observe any differences in the edge structure of paintings and photographs in our images, we suspect that the intensity edges in paintings are subtly different from intensity edges in photographs. In future work, we plan to study this issue on high-resolution images.

References

[1] J.G. Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Res*, 20:847–856, 1980.

[2] J.G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A.*, 2(7):1160–1169, 1985.

[3] J. Malik and P. Perona. Preattentive texture-discrimination with early vision mechanisms. *J. Opt. Soc. Am.*, 7:923–932, 1990.

[4] B.S. Manjunath and W.Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transaction on Pattern Analysis an Machine Intelligence*, 18(8):837–842, 1996.

[5] S. Marcelja. Mathematical description of the responses of simple cortical cells. *J. Opt. Sco. Am.*, 70:1297–1300, 1980.

[6] J.F.C. Wanderley and M. H. Fisher. Multiscale color invariants based on the human visual system. *IEEE Transactions on Image Processing*, 2001. http://www.ene.unb.br/ juliana/publications.html.

[7] V. Athitsos, M.J. Swain, and C. Frankel. Distinguishing photographs and graphics on the World Wide Web. Workshop on Content-Based Access of Image and Video Libraries (CBAIVL '97) Puerto Rico, 1997.

[8] J.B.Subirana-Vilanova and K.K. Sung. Multi-scale vector-ridge-detection for perceptual organization without edges, 1993. A.I. Memo 1318, MIT Artificial Intelligence Laboratory, Cambridge, MA, USA.

[9] N.C. Rowe and B. Frew. Automatic caption localization for photographs on world wide web pages. *Inform Process Manag*, 34:95–107, 1998.

[10] J.R. Smith and S.F. Chang. Searching for images and videos on the world-wide web. Tr 459-96-25, Center for Telecommunications Research, Columbia University, 1996.

[11] M. Szummer and R. W. Picard. Indoor-outdoor image classification. In *IEEE International Workshop on Content-based Access of Image and Video Databases, in conjunction with CAIVD'98*, pages 42–51, 1998.

[12] A. Vailaya, A. K. Jain, and H.-J. Zhang. On image classification: City vs. landscapes. *International Journal of Pattern Recognition*, 31:1921–1936, 1998.

[13] J. F. Canny. A computational approach to edge detection *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 8:679–697,1986.

[14] C. M. Bishop. Neural networks for pattern recognition The Clarendon Press Oxford University Press, New York, 1995

Tables

Feature	P miss rate	Ph miss rate	order
E_c	37.37	37.36	P > Ph
E_g	33.34	33.34	P < Ph

Table 1: Painting-photograph discrimination performance based on each of the two edge features. P denotes paintings, Ph denotes photographs. For each feature, paintings were separated from photographs using an optimal threshold. The miss rate is defined as the proportion of images incorrectly classified. The last column indicates the order of the classes with respect to the threshold.

Feature	E_c	R	U	S
E_c	1.00; 1.00	0.01; 0.13	0.10; -0.13	0.45; 0.52
R	0.01; 0.13	1.00; 1.00	0.43; 0.25	0.33; 0.44
U	0.10; -0.13	0.43; 0.25	1.00; 1.00	0.28; 0.17
S	0.45; 0.52	0.33; 0.44	0.28; 0.17	1.00; 1.00

Table 2: Correlation coefficients for all feature pairs, calculated over all photographs and all paintings. Each entry in the table lists first the correlation coefficient calculated over photographs, followed by the correlation coefficient for paintings.

Classifier	P hit rate ($\mu \pm \sigma$)	Ph: hit rate ($\mu \pm \sigma$)
C_1	72.2±5%	71.1±4%
C_2	81.3±3%	81.1±3%
C_3	79±5%	78±4%
C	94.5±3%	92.5±2%

Table 3: Classification performance: the mean and the standard deviation of the hit rates over the 100 testing sets. C_1 is the classifier operating in the space of the scalar-valued features. C_2 is the classifier for RGBXY space, and C_3 is the classifier for Gabor space. C is the average classifier. P denotes paintings, Ph denotes photographs.

Figures

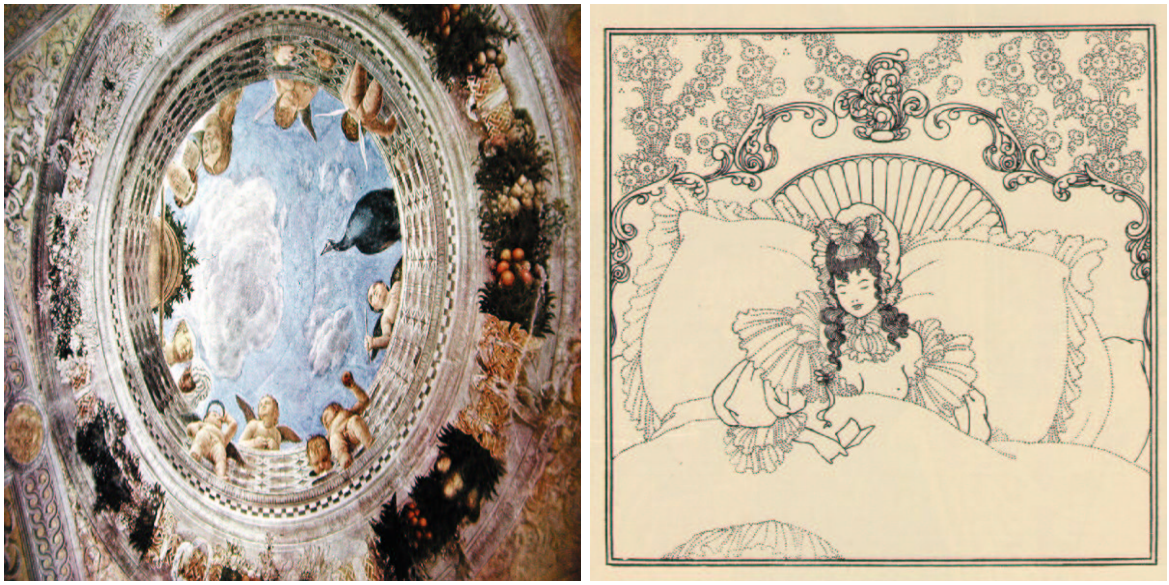

Figure 1: Murals (left) were included in the class "paintings". Line drawings (right) were excluded.

Figure 2: Visually differentiating paintings from photographs can be a non-trivial task. *Left:* Photographs. *Right:* Paintings.

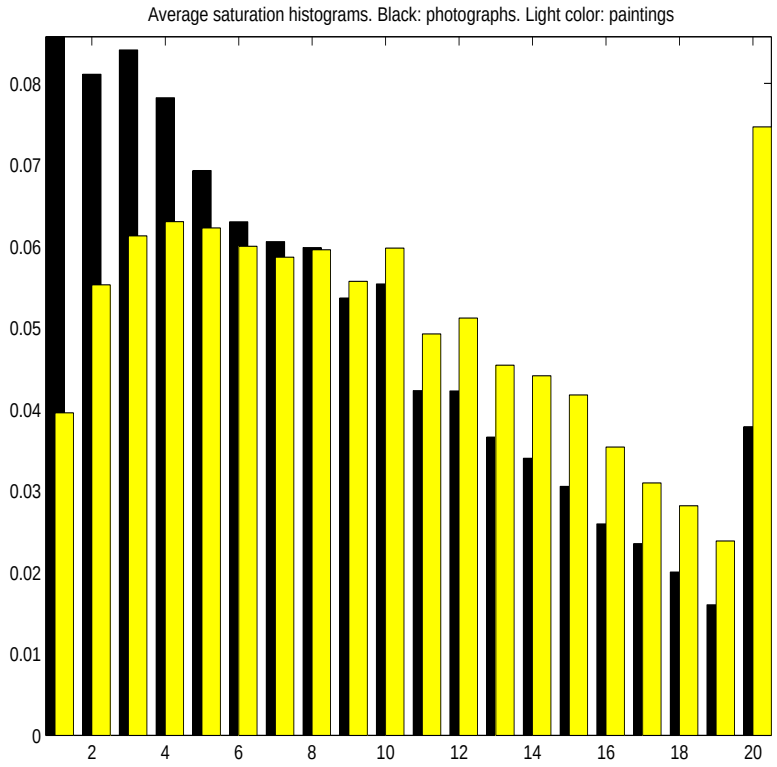

Figure 3: The mean saturation histogram for photographs (black) and paintings (light gray). 20 bins were used. Photographs have more unsaturated pixels, paintings have more highly saturated pixels.

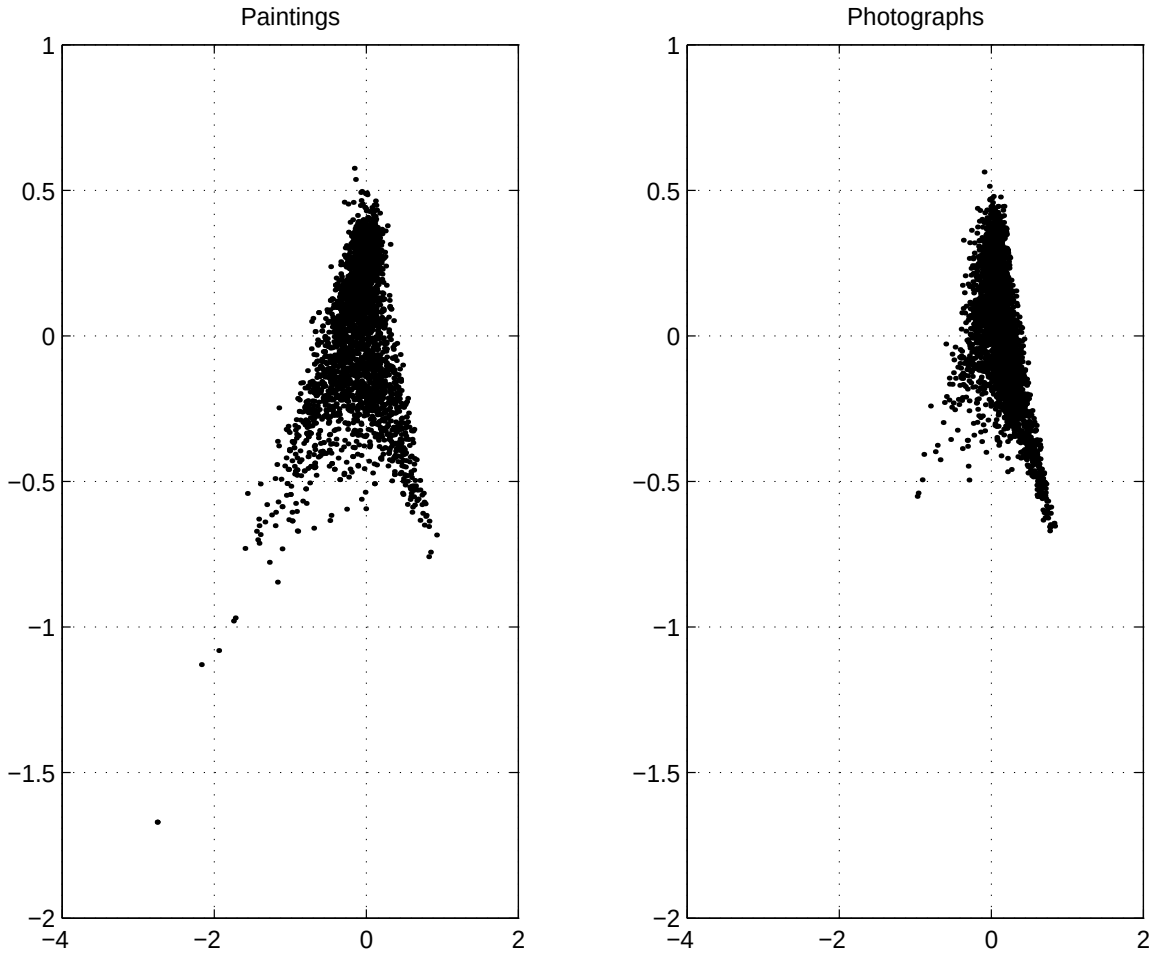

Figure 4: Painting and photograph data points represented separately in the same two-dimensional space of the first two principal components of the common painting-photograph image set. LEFT: paintings RIGHT: photographs.

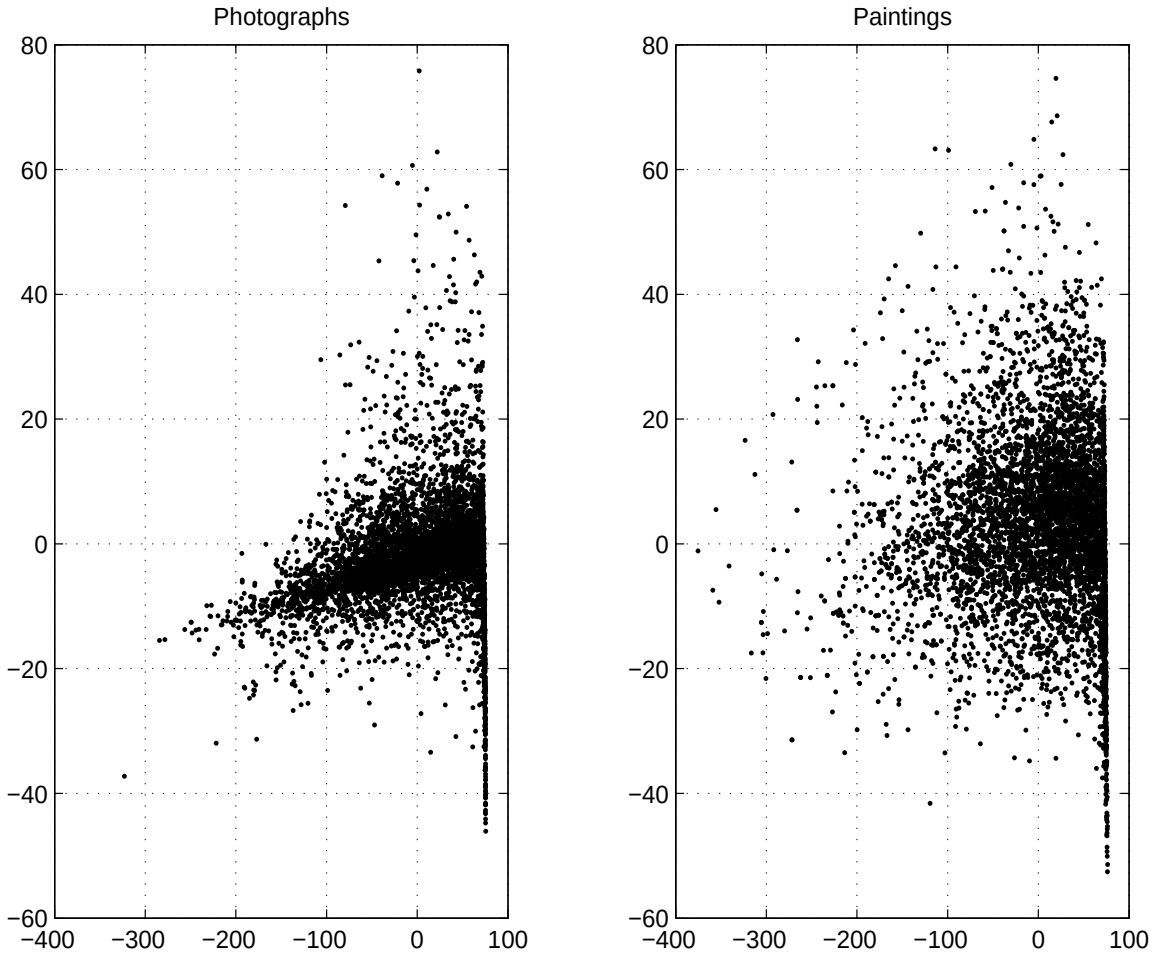

Figure 5: RGBXY space: painting and photograph data points represented separately in the same two-dimensional space of the first two principal components of the common painting-photograph image set. LEFT: photographs RIGHT: paintings.

1	1	1
1	-8	1
1	1	1

Figure 6: Laplacean mask

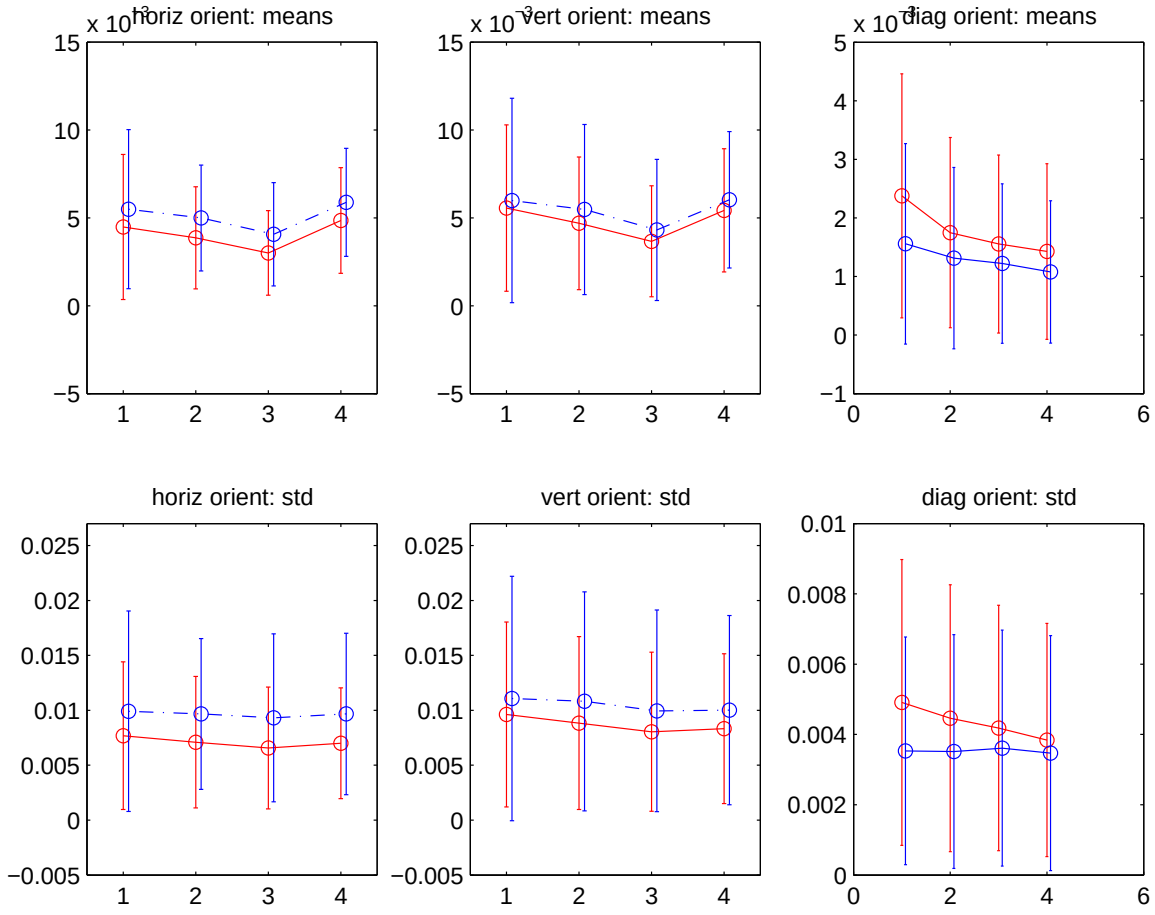

Figure 7: Errorbar plots illustrating the dependence of the image-mean and image-standard deviation of Gabor ouputs on filter scale and orientation for the painting (red lines) and photograph (interrupted blue lines) image sets. TOP LEFT: Horizontal orientation. Errorbar plot representation of the image-set-mean and image-set-standard deviation of the image-mean of Gabor filter output magnitude as a function of filter scale. Errobars represent the standard deviations determined across images, expressing inter-image variability. The plots for the paintings set are in red, for the photographs set, in blue. TOP MIDDLE: Corresponding plots for the vertical orientation. TOP RIGHT: Corresponding plots for the diagonal orientations: the data for 45° and 135° are presented together. BOTTOM LEFT: Horizontal orientation. Errorbar plot representation of the image-set-mean and image-set-standard deviation of the image-standard-deviation of Gabor filter output magnitude as a function of filter scale. Errobars represent the standard deviations determined across images, expressing inter-image variability. BOTTOM MIDDLE: Corresponding plots for the vertical orientation. BOTTOM RIGHT: Corresponding plots for the diagonal orientations: the data for 45° and 135° are presented together.

Figure 8: Images rated as typical paintings. Classifier output is displayed above each image. An output of 1 is a perfect photograph

Figure 9: Images rated as typical paintings. Classifier output is displayed above each image. An output of 1 is a perfect photograph

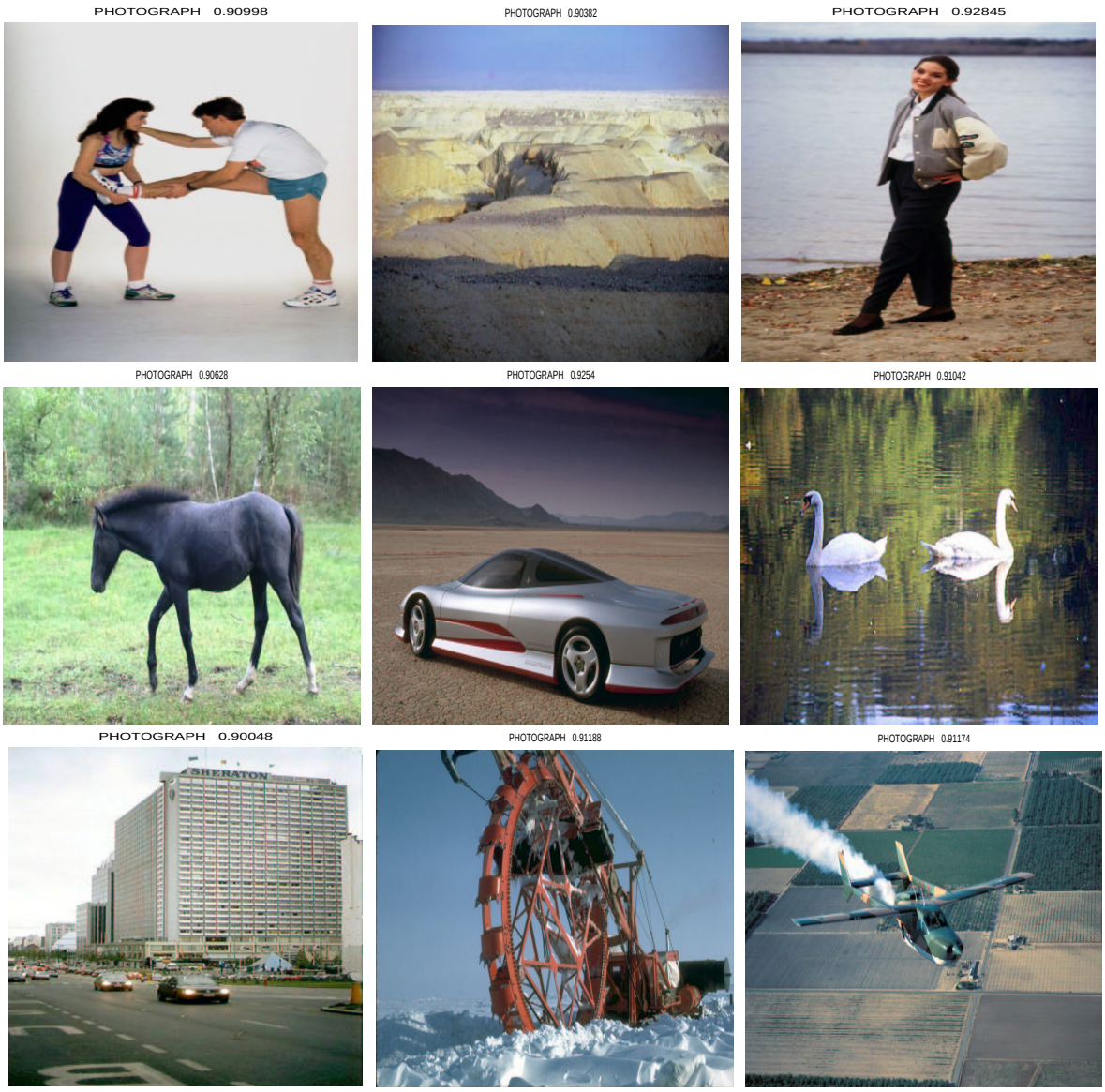

Figure 10: Images rated as typical photographs. Classifier output is displayed above each image. An output of 1 is a perfect photograph

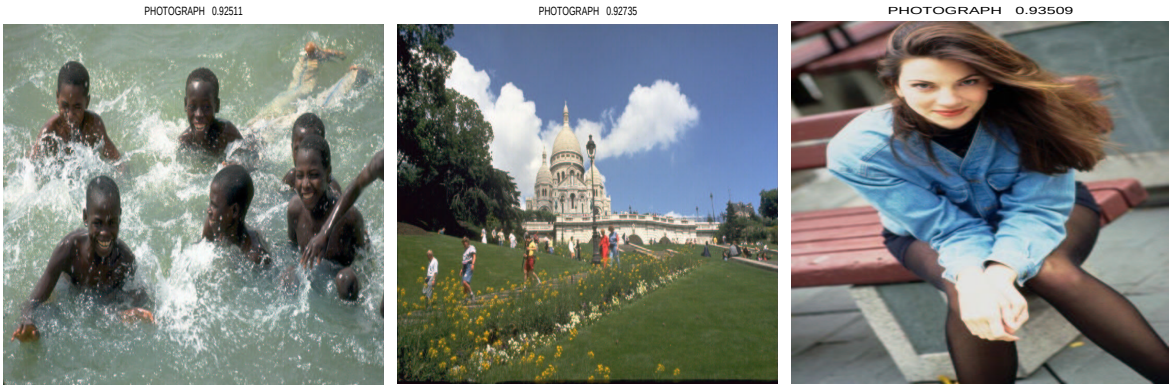

Figure 11: Images rated as typical photographs. Classifier output is displayed above each image. An output of 1 is a perfect photograph

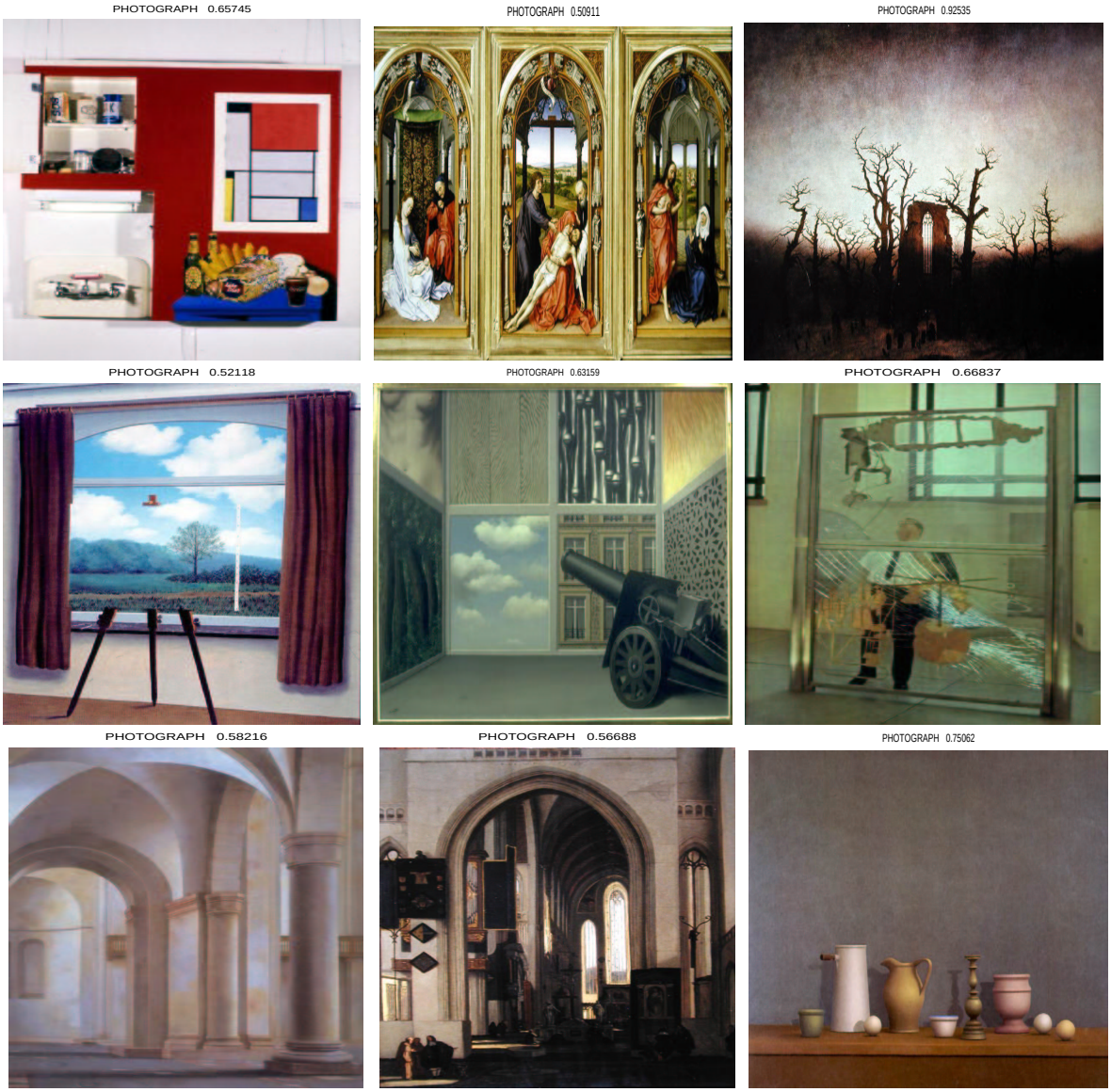

Figure 12: Paintings classified as photographs. Classifier output is displayed above each image. An output of 1 is a perfect photograph

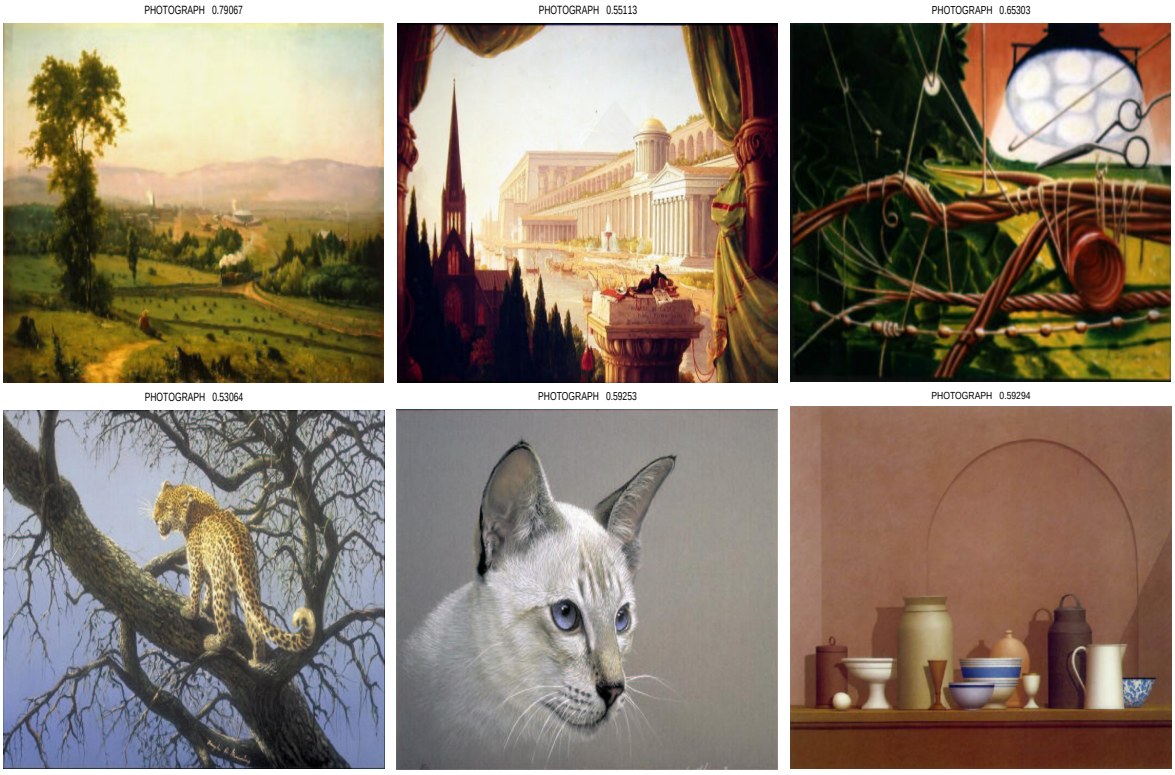

Figure 13: Paintings classified as photographs. Classifier output is displayed above each image. An output of 1 is a perfect photograph

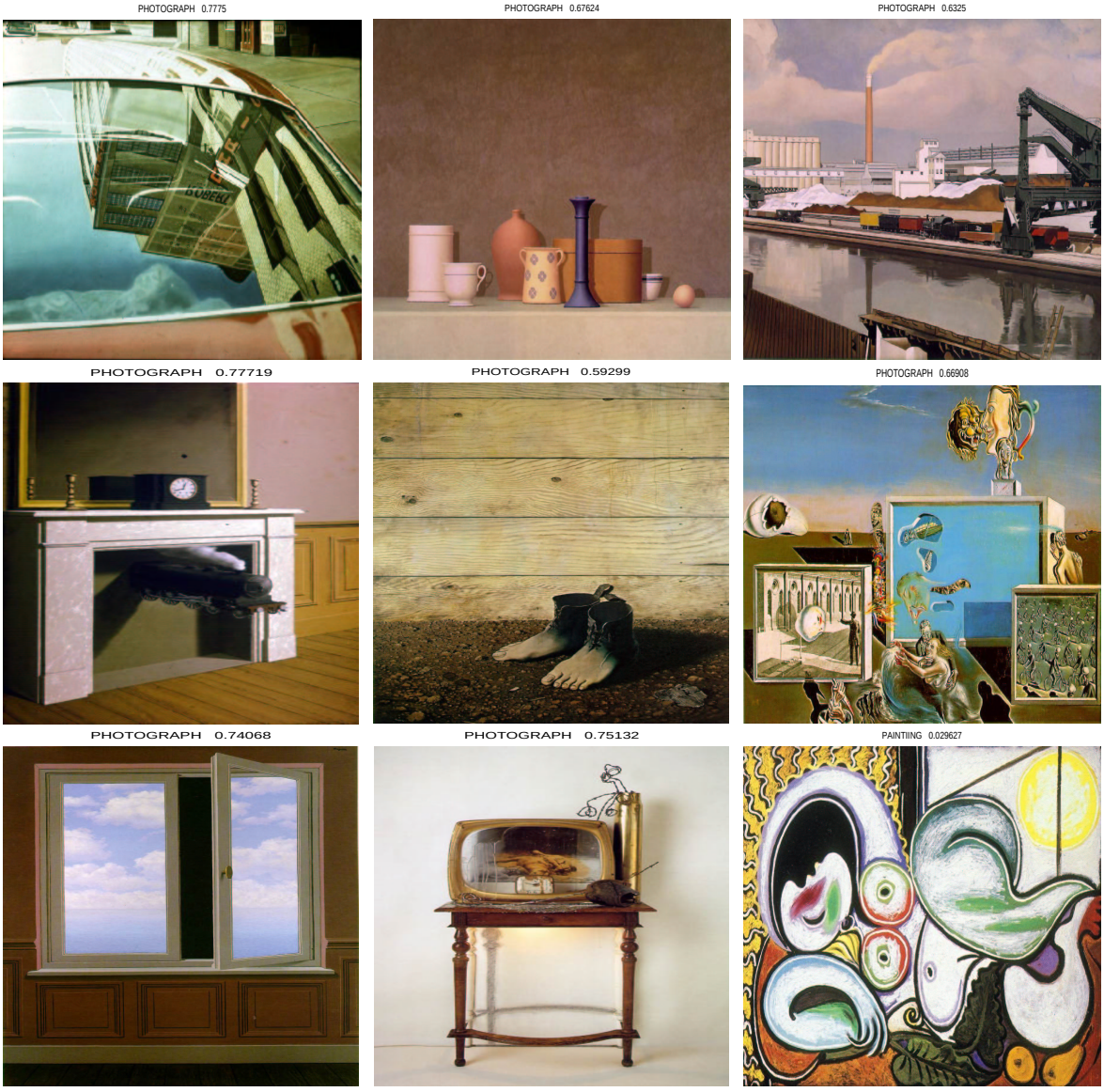

Figure 14: Paintings classified as photographs. Classifier output is displayed above each image. An output of 1 is a perfect photograph

Figure 15: Photographs classified as paintings. Classifier output is displayed above each image. An output of 0 is a perfect painting

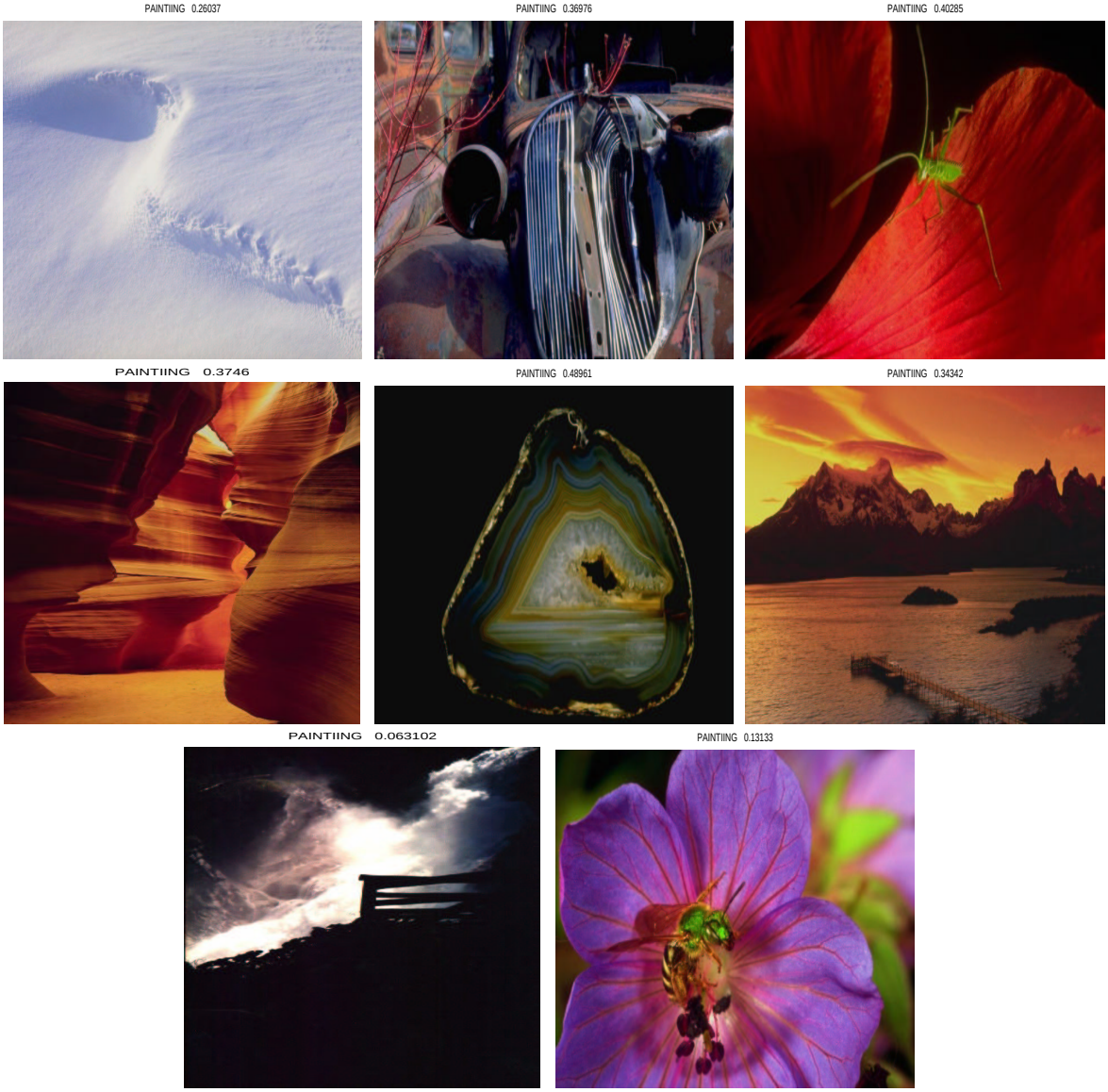

Figure 16: Photographs classified as paintings. Classifier output is displayed above each image. An output of 0 is a perfect painting

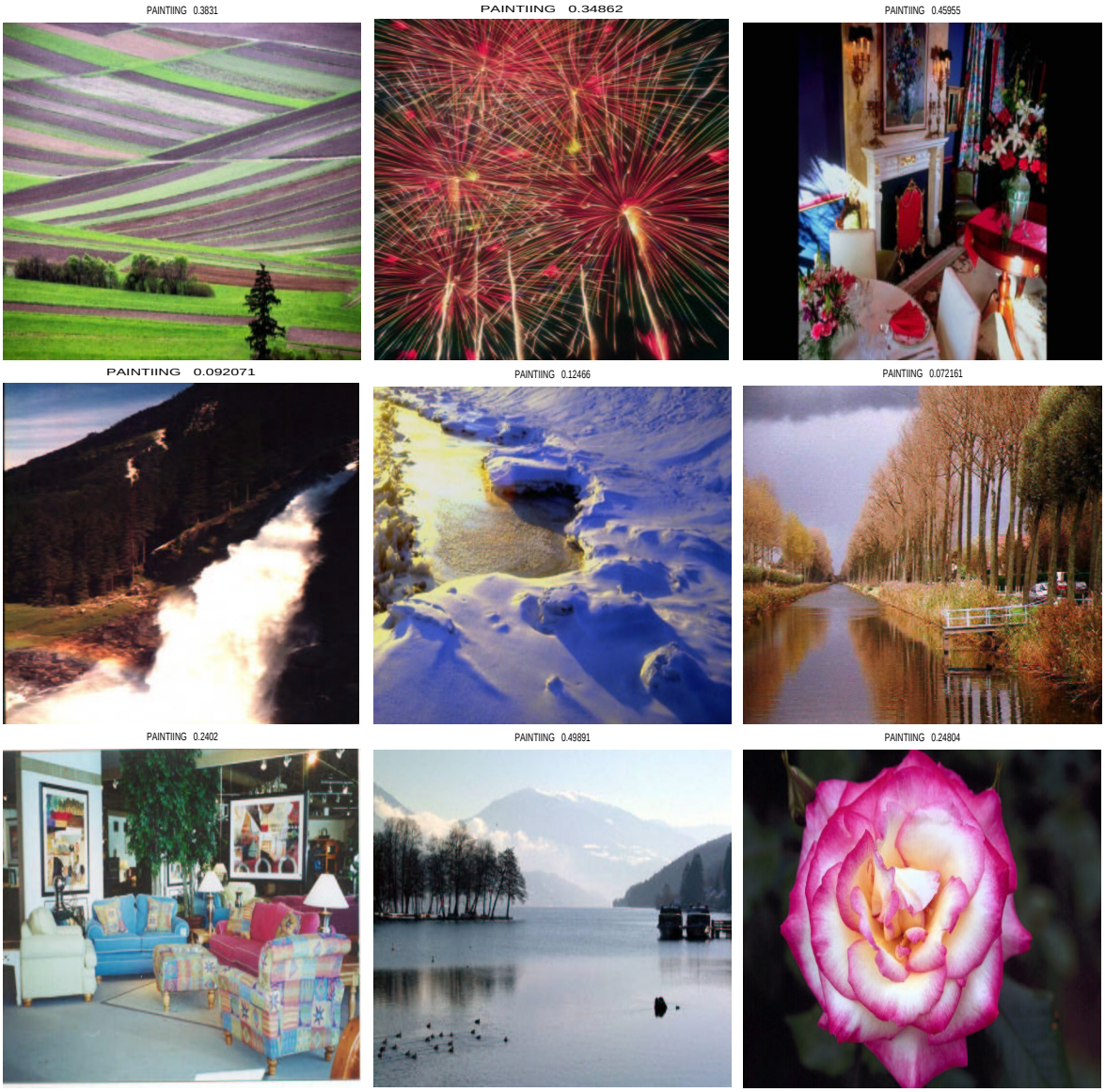

Figure 17: Photographs classified as paintings. Classifier output is displayed above each image. An output of 0 is a perfect painting

Figure 18: LEFT: scrambled painting. RIGHT: scrambled photograph.