

A Deep Approach for Classifying Artistic Media from Artworks

Heekyung Yang¹ and Kyungha Min^{2*}

¹Dept. of Computer Science, Graduate School, Sangmyung Univ.,
Hongji-dong, Jongro-gu, Seoul, Korea
[e-mail: yhk775206@naver.com]

²Dept. of Computer Science, Sangmyung Univ.,
Hongji-dong, Jongro-gu, Seoul, Korea
[e-mail: minkh@smu.ac.kr]

*Corresponding author: Kyungha Min

Received October 22, 2018; accepted November 17, 2018; published May 31, 2019

Abstract

We present a deep CNN-based approach for classifying artistic media from artwork images. We aim to classify most frequently used artistic media including oilpaint brush, watercolor brush, pencil and pastel, etc. For this purpose, we extend VGGNet, one of the most widely used CNN structure, by substituting its last layer with a fully convolutional layer, which reveals class activation map (CAM), the region of classification. We build two artwork image datasets: *YMSet* that collects more than 4K artwork images for four most frequently used artistic media from various internet websites and *WikiSet* that collects almost 9K artwork images for ten most frequently used media from WikiArt. We execute a human baseline experiment to compare the classification performance. Through our experiments, we conclude that our classifier is superior in classifying artistic media to human.

Keywords: CNN, VGGNet, classification, artistic media, artwork

1. Introduction

Recent progress of deep convolutional neural network (CNN) accelerates the study on understanding artwork images. Most of the study concentrate on recognizing and classifying the artistic style from the artworks. However, we understand that recognizing and classifying the artistic media used to create an artwork is a fundamental step for understanding the artwork. The artistic media used to create artworks include pencils of various H's and B's, pastels, diverse oilpaint and watercolor brushes, various papers and canvas, etc. Among the various artistic media, we aim to recognize and classify the stroke-based media that deposit pigments on a surface through a series of strokes. The media such as canvas, paper, panel or wood are excluded.

The key challenge in classifying artistic media is to define a set of features that represents the difference of the media explicitly. Artistic medium produces distinctive stroke marks on a surface. The difference of the marks comes from the process of depositing pigments. For example, pencils and pastels deposit pigments via the abrasion of the media with the surface and brushes leave the pigments via water or oil. Our visual recognition system can distinguish the media used to create an artwork, but it is difficult to define the features that make us distinguish them explicitly. For example, we can easily distinguish artworks created by oilpaint brush and color pencil, but we cannot explain what makes us distinguish them. Furthermore, the differences between the pairs of the media are not equal. For example, the pair of pencil and watercolor brush is an easily distinguishable, but the pair of pencil and pastel is very confusing.

For the first step of classification, we survey Wikiart.org, the famous visual artwork image encyclopedia website (www.wikiart.org) to select ten most frequently used artistic media for the classical artworks. Among the ten media, we select four more frequently used media in both classical art and contemporary art. As a second step, we build two artwork image datasets for our research. An artwork image dataset collected from WikiArt.org is denoted as *WikiSet*, which is composed of almost 9K images for ten media, and a dataset collected from various internet websites is denoted as *YMS**et*, which is composed of more than 4K images for the four more frequently used media. The ten media for classical artwork and four media for contemporary artwork are illustrated in [Fig. 1](#).

In the third step, we devise a deep convolutional neural network (CNN)-based approach for classifying artistic media from artworks. Deep CNN is a state-of-the-art technique in recognizing and classifying objects embedded in various images. The modern deep CNN structures such as AlexNet, VGGNet, GoogLeNet and ResNet show good performance in classifying objects whose categories are not defined explicitly. We employ a pre-trained network which has been used for classifying objects. We add several layers to the network and train the newly added layers to construct a classifier. Among the many deep CNNs, we choose VGGNet [18], which is very frequently used in various applications such as texture transfer and image synthesis [7, 19]. VGGNet is distinguished from other networks in its simple structure which is composed of ordinary convolution, pooling and fully connected layers.

To understand the process of classification, we substitute the fully connected layers in VGGNet with fully convolutional layers. In object localization and detection research, the

fully convolutional layers identify the region that plays dominant clue for a classification. The fully convolutional layer is an alternative component for specific goals such as segmentation [13], localization [22] and detection [4]. This component generates class activation map (CAM), which is known as heat map, that notifies impact area based on which the classifier predicts. We compare the results of two networks: the original VGGNet and the VGGNet whose fully connected layers are substituted with fully convolutional layers.

Our contributions are listed as follows.

- (1) Even though the classification of the media is essential, we have not found a serious technique that computationally classifies important stroke-based media from an artwork. We present a deep CNN-based approach to classify the stroke-based artistic media from artwork images.
- (2) We execute a human baseline experiment to compare the performances of human and deep CNN-based classifier. Through this experiment, the performance of deep CNN-based classifier is proved to be compatible with that of human.
- (3) We visualize the clue to the classification by employing a fully convolutional layer as the last layer of our classifier. We compare the region of an image, which is the clue of classification with that of human baseline experiment, and present an understanding of classification process.

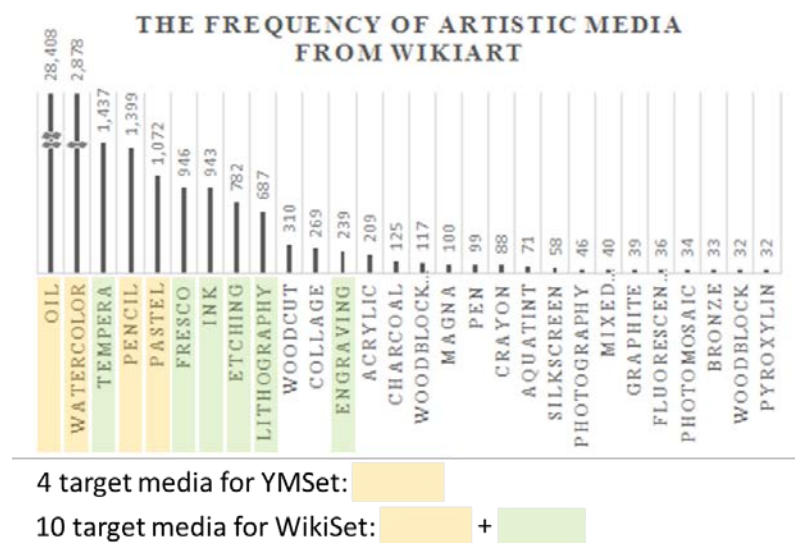


Fig. 1. The frequency of artistic media from WikiArt: Four target media for YMSets are illustrated in yellow and ten media for WikiSet in yellow and green.

2. Related Work

2.1 Estimation of aesthetic quality of photographs

In the research of estimating aesthetic quality, photographs have been received more attention than artworks. The research proposed a variety of estimation metrics from simple features such as colorfulness, saturation, rule-of-thirds and depth of field [5] to complicated features such as layout or configuration, objects or scene types and lighting conditions [6].

Isola et al. [9] predicted human judgements of image memorability using various high-level features. Gygli et al. [8] defined interestingness elements as aesthetics, unusualness and general preferences and estimated them by making feature set. Borth et al. [1] analyzed sentiment by training object classifier on adjective-noun pairs. However, they did not consider estimation of artworks. Since artworks have huge subjectivity and diversity in both their expression and evaluation, different metrics with ones of photographs are needed.

2.2 Classification of artistic styles

Since style is one of important elements of artworks, techniques recognizing or classifying it from artworks have been studied. Keren et al. [11] regarded styles as painters and suggested categorizing method according to the painters. They used DCT transform coefficients as their features for classifying 5 painters in their 30 artworks. Shamir et al. [16] included schools of arts addition to painters in style. And they used 11 features such as histogram and edge statistics for classifying 3 schools of arts and 9 painters in about 60 artworks per painter. However, in both of the researches, the number of painters and artworks is too small and the data is restricted within easily distinguishable paintings such as Van Gogh's and Dali's.

Karayev et al. [10] constructed bigger data set with 80K artworks than previous works and proposed classification techniques for more various 20 historical styles. According to their experiments, AlexNet, one of deep networks, extracts the best features among popular 6 features for style classification performance. So they give us reason that deep network is a good structure for classifying artworks as well as photographs. After AlexNet, deep networks of improved performance have been proposed in computer vision community. However, they did not deal with artistic media.

Recently, the development of multi-layer perceptron structure such as CNN is cooperated with support vector machine (SVM) technique to present an automatic recognition technique for artistic genre of paintings [3]. Chu and Wu [2] employed correlation feature extracted by deep CNN to classify image style. They argue that the vectorized correlation-based scheme presents better performance than the CNN-based schemes. Matsuo and Yanai [14] presented a style matrix, which is constructed from PCA dimension reduction on the Gram matrix [7]. The style matrix is used for image and style retrieval on image database. Shamir et al. [17] presented a recognition scheme for abstract expressionism by evaluating the relation between the level of artistic and visual contents. Lecoutre et al. [12] employed deep residual network for recognizing artistic style of artwork images in WikiArt database. Sun et al. [21] presented a two-pass CNN structure that extracts both object feature and texture feature. The texture pathway combines the Gram matrix of the features in object pathway to improve the performance. They constructed their structure based on AlexNet and VGG-19.

2.3 Classification of artistic media

Mensink et al. [15] developed an automatic method classifying photographs of collections in Rijksmuseum in Netherlands. They defined recognizing target as artist, type, material and creation year. However, artistic media depositing pigments were not the main consideration in their artistic material. Their materials included paper, wood, silver and only three easy distinguishable pairs such as oil, ink and watercolor. Also, they used SIFT feature which is strong in pattern-based classification, whereas not appropriate for media pair with similar

pattern such as pencil and pastel. We select features extracted from the deep network which are the best for classification of both photographs and artworks.

3. Collecting dataset

The category of the artistic media we aim to recognize covers oilpaint brush, watercolor brush, tempera, pencil, pastel, fresco, ink, etching, lithography and engraving. Among the ten media above, we furthermore select four most frequently used media as oilpaint brush, watercolor brush, pencil and pastel. We omit tempera, which is very similar to oilpaint, since it is not used nowadays. We collect 8,597 artwork images from WikiArt website for ten media and 4,139 real artwork images from internet websites for four media (see Fig. 2). Most of the collected artwork images are color images. Among the collected dataset, we assign 70% of them for training, 15% for validation and 15% for test.

The collected raw artwork images, however, can be inadequate for training our classifier, since the media texture is not observed or some obstacle elements exist. Therefore, we preprocess them by excluding the images where the media texture is not observed or by cropping the elements that disturb the training.

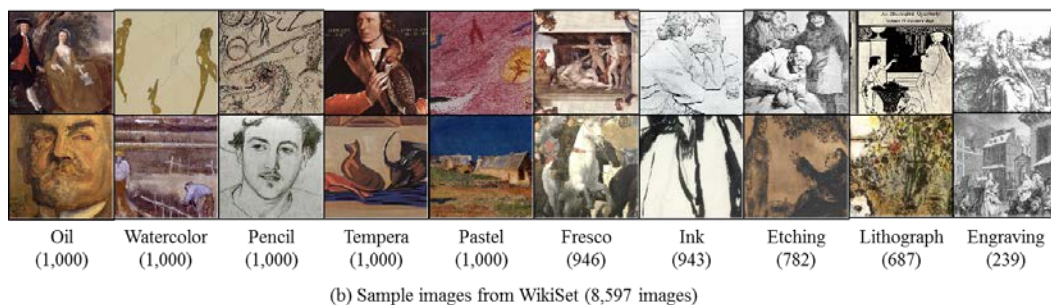
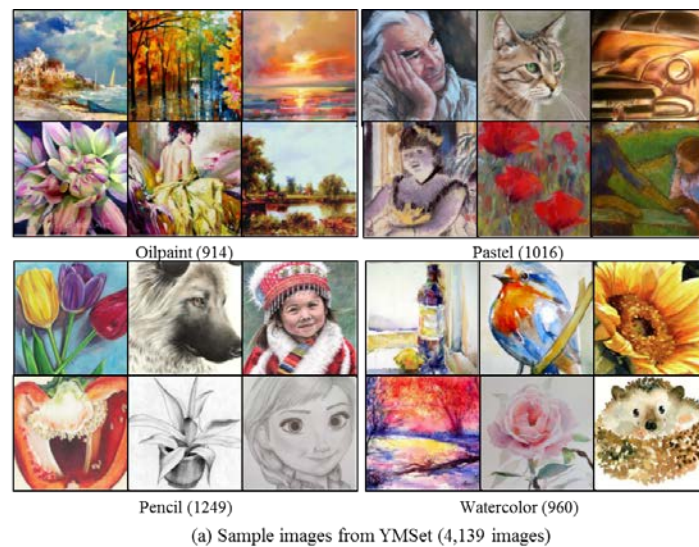


Fig. 2. Sample artwork images from two datasets.

We list the preprocessing operations as follows:

- ① Excluding images of extremely high and low resolutions. In these images, the media texture hardly appear (see Fig. 3 (a)).
- ② Cropping Images containing those elements irrelevant to the media texture such as blank, watermark, eraser and clip, etc (see Fig. 3 (b)).



(a) High resolutional and low resolutional artwork images



(b) Cropping irrelevant elements



(c) Historical paintings from WikiArt



(d) Hyperrealism artworks

(e) Rough sketch artworks

Fig. 3. Examples of artworks images to avoid.

- ③ Excluding images of historical paintings from WikiArt for YMSet (see [Fig. 3 \(c\)](#)).
- ④ Excluding images of hyperrealism artwork. Since these artworks pursue photorealism, the media texture hardly appears (see [Fig. 3 \(d\)](#)).
- ⑤ Excluding images of rough sketches. In these artworks, paper texture dominates media texture (see [Fig. 3 \(e\)](#)).
- ⑥ Excluding images of the artworks created by an identical artist. The artworks created by an identical artist can overfit our classifier to his/her drawing style.

4. Deep convolutional classifier

We compare two classifiers: the one that ends with fully connected layers and the other that ends with fully convolutional layers.

4.1 Classifier with fully connected layer

Our classifier employs the structure of classifiers based on deep convolutional neural network. Our classifier is constructed using a pre-trained 19-layer VGGNet whose last layer is a fully connected layer, which composed of two layers, each of which has 512 units and 1024 units, respectively (see [Fig. 4 \(a\)](#)). Most of the training on our classifier is executed on the new fully convolutional layer. The collected real artwork dataset is organized into three groups: training, validation and test.

The training dataset, which is 75% of the collected real artwork dataset, is to update parameters of the network such as weights and biases in the newly added fully convolution layer in forward and backward propagations. Validation dataset, which is 15%, is to estimate the accuracy of our classifier after the training phase. According to the estimated accuracy, we correct the hyper-parameters of our classifier such as learning rate, number of epochs, number of layers and units, etc. The classifier is trained for each correction in order to improve the accuracy. Test dataset, which is 15%, is to estimate the accuracy of the classifier after training and validation phases.

4.2 Classifier with fully convolutional layer

Another classifier is constructed using a pre-trained 19-layer VGGNet whose last layer is substituted with a fully convolutional layer (See [Fig. 4 \(b\)](#)). We employ the localization scheme proposed by Zhou et al. [22]. The fully convolutional layer is composed of a convolutional layer with 512 feature maps and a fully connected layer of 512 units. In each layer, we use global average pooling instead of widely used max pooling. In recognizing objects, Zhou et al. showed that the fully convolutional layer is better than the fully connected layer using ILSVRC dataset [22]. We execute the same comparison on the dataset collected from real artworks.

Regularization is one of the key techniques in training a deep network to prevent overfitting. We use the dropout technique presented by Srivastava et al. for the regularization [20]. We assign the values of the hyper-parameters of our classifier as follows: 0.001 for learning rate, 100 for training epochs, and 32 for batch size.

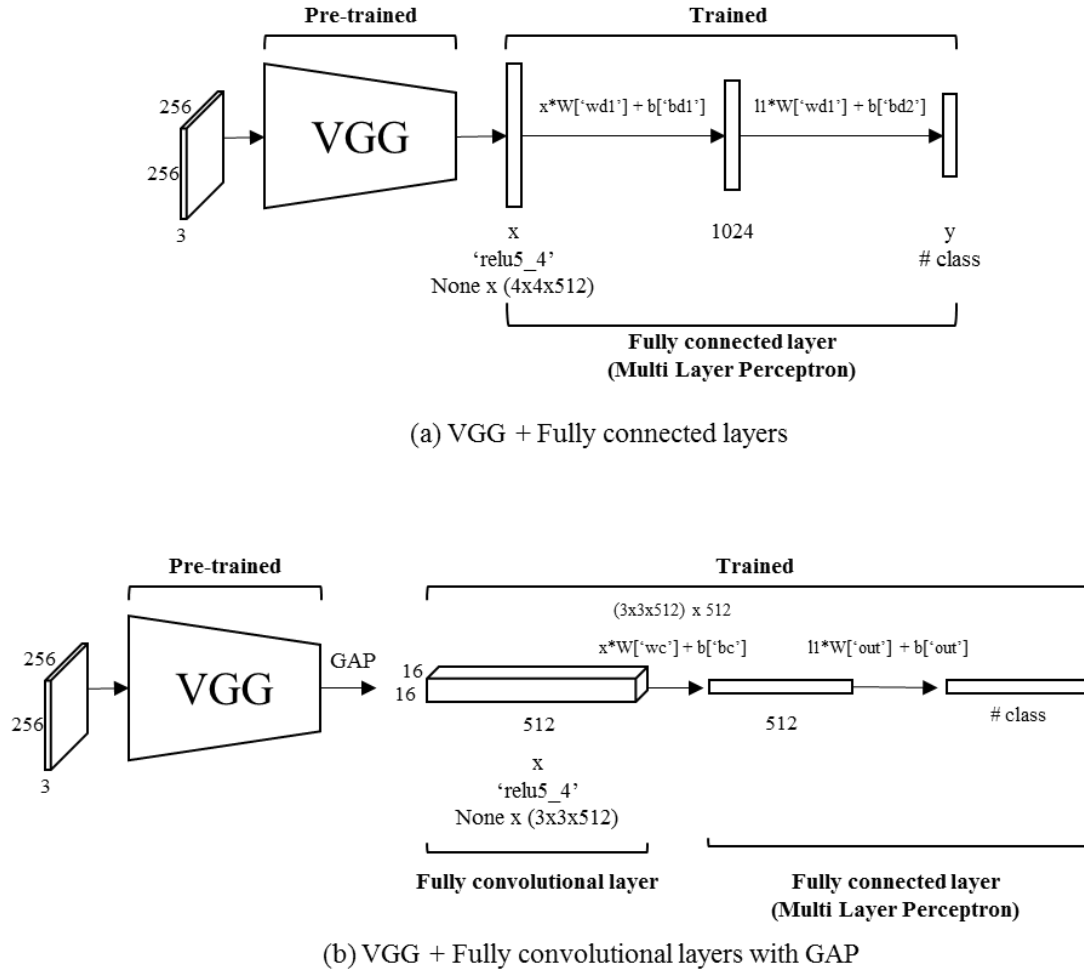


Fig. 4. The structure of VGGNet-based classifiers

5. Experiment and analysis

5.1 Experiment and results

We implement our classifier on a PC with Pentium 7 CPU, 32 GByte main memory and nVidia Titan GPU. Our software environment is Python with GPU-accelerated Tensorflow libraries. Our classifier is constructed based on VGGNet structure. We compare two structures for the last layer of our classifier: a fully connected layer and a fully convolutional layer with global average pooling (GAP). We use F1 score and accuracy to evaluate the performance of our classifier. The F1 score and accuracy are estimated on the test dataset of 582 real artwork images, which are composed of 143 oilpaints, 150 pastel paintings, 144 pencil drawings and 145 watercolor paintings. Since the performance of the classifier depends on its initial parameters, we estimate the performance for five times and select the parameter set with the parameters of the best case. We suggest the results of the five experiments in [Table 1](#), where the fully convolutional layer with GAP shows better F1 score and accuracy than the fully convolutional layer. The fully connected one takes 5~6 minutes for training and the fully convolutional one about one hour. We choose the fully convolutional layer structure for the

last layer of our classifier according to the following two reasons. First, the fully convolutional layer structure shows better performance than the fully connected layer structure. Second, the fully convolutional layer extracts the class activation map (CAM), which visualizes the background of prediction by representing the importance of the pixels with different colors. Our result is presented in the confusion matrix presented in Fig. 5. We also present F1 score from the matrix in the column of Classifier 600 in Table 2.

Table 1. The comparison of five experiments for determining optimal parameters.

Metric \ Test order	F1 Score		Accuracy	
	Connected	Convolutional	Connected	Convolutional
1	0.83	0.91	0.88	0.94
2	0.84	0.86	0.89	0.9
3	0.83	0.83	0.88	0.89
4	0.82	0.87	0.87	0.91
5	0.79	0.88	0.86	0.92
Avg.	0.82	0.87	0.88	0.91

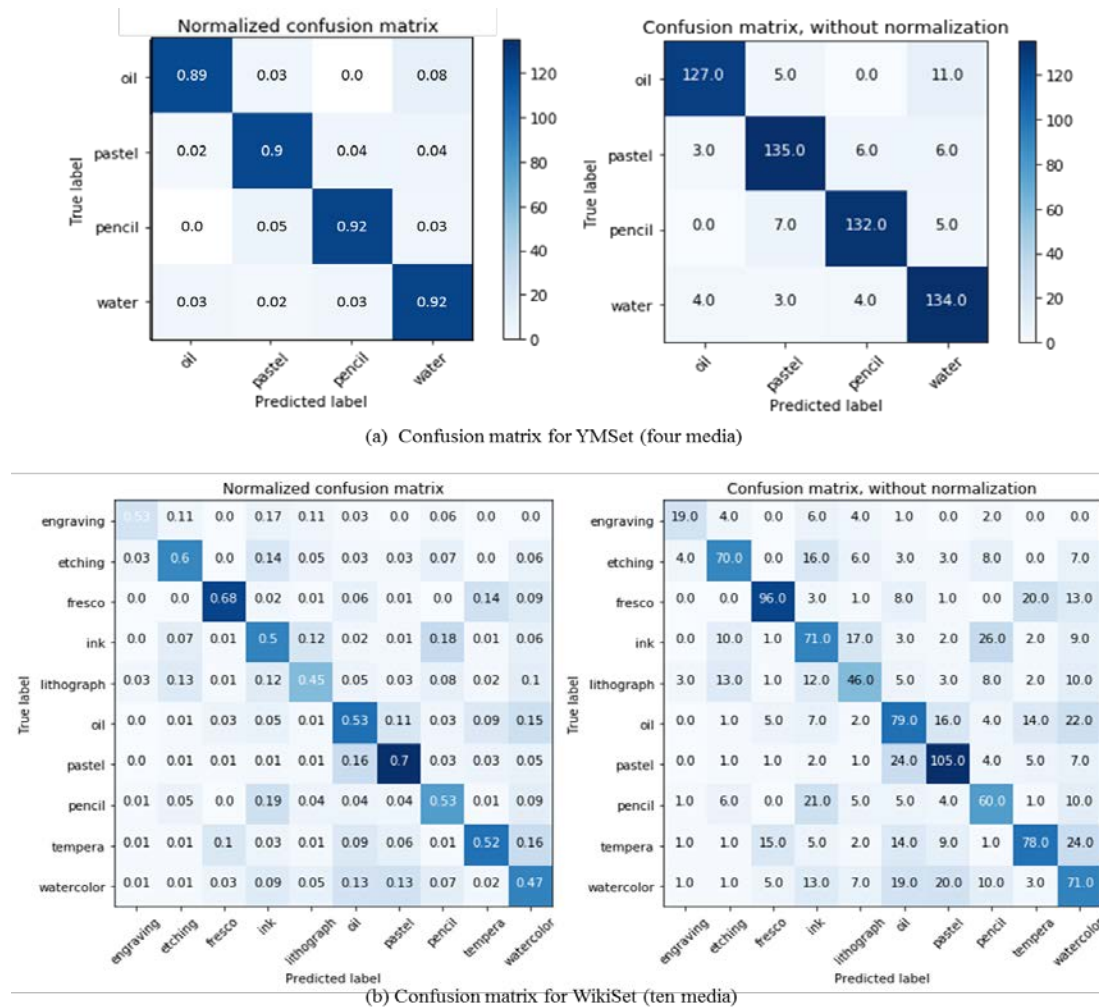


Fig. 5. The confusion matrices for our experiment.

5.2 Comparison with human baseline

Human baseline is another important factor in estimating the performance of artificial classifiers. We hire 20 human subjects whose ages are in twenties and thirties and arrange them in two groups: a training group and a non-training group. Each group has 10 subjects. A subject is given 100 questions, which show an artwork image and four selection samples for the media: oilpaint, pastel, pencil and watercolor. Each subject is given the same set of questions. For each question, we employ a voting strategy that chooses the most frequently chosen sample as the answer of the question. For the training group, we give 100 questions with answers before the test. The questions in the training set are different from those in the test set. Since they can check the correctness of their selections, they are trained for classifying the media used in an artwork. [Fig. 6](#) illustrates the overview of human baseline experiment. [Fig. 7](#) illustrates an example of a question and the process of voting.

Table 2. F1 scores for experiments

	Classifier 1200								
	acc	prec	rec	f1		acc	prec	rec	f1
Oil	0.71	0.49	0.53	0.51	Fresco	0.87	0.77	0.68	0.72
Watercolor	0.86	0.41	0.47	0.44	Ink	0.77	0.46	0.50	0.48
Pencil	0.73	0.49	0.53	0.51	Etching	0.92	0.65	0.60	0.62
Tempera	0.78	0.62	0.52	0.57	Lithograph	0.76	0.51	0.45	0.47
Pastel	0.70	0.64	0.70	0.67	Engraving	0.98	0.66	0.53	0.58
Overall	0.81	0.57	0.55	0.56					

(a) F1 score for the test set from WikiArt (1200 images)

	Classifier 600				Classifier 100				Trained Human 100				Non-trained Human 100			
	acc	prec	rec	f1	acc	prec	rec	f1	acc	prec	rec	f1	acc	prec	rec	f1
Oilpaint	0.96	0.95	0.89	0.92	0.96	0.95	0.86	0.90	0.92	0.82	0.82	0.82	0.91	0.71	1.00	0.83
Pastel	0.92	0.90	0.90	0.90	0.91	0.93	0.89	0.91	0.89	0.88	0.79	0.83	0.85	0.88	0.54	0.67
Pencil	0.92	0.93	0.92	0.92	0.89	0.96	0.83	0.89	0.94	0.93	0.93	0.93	0.93	0.87	0.90	0.88
watercolor	0.94	0.86	0.92	0.89	0.93	0.76	1.00	0.86	0.97	0.88	1.00	0.94	0.99	0.96	1.00	0.98
Overall	0.94	0.91	0.91	0.91	0.92	0.90	0.90	0.89	0.93	0.88	0.88	0.88	0.92	0.85	0.86	0.84

acc: accuracy, prec: precision, rec: recall, f1: f1 score

(b) Classifier 600 is from our classifier with dataset of 582 images, and Classifier 100 is from dataset of 100 images. The same 100 images are used for Trained Human 100 and Non-trained Human 100.

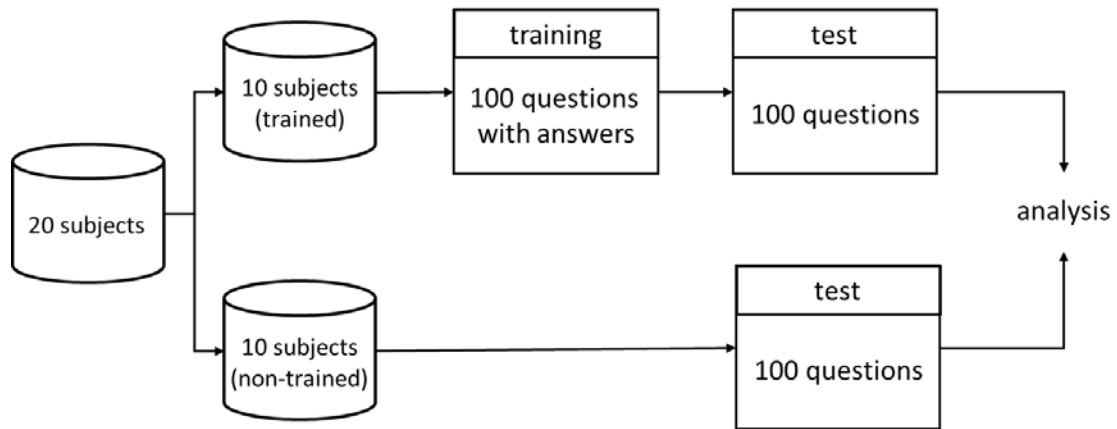



Fig. 6. The structure of human baseline experiment.

In the training stage, they repeat the process 5 times, where they select media for 20 artworks and check their own errors. We limit the number of real artworks in the test only 100, which avoids the subjects' tedium. The images in the questions are the same images used for the classifier. Each image is 256 x 256 resolution image center-cropped from an artwork. We test our classifier with the same test set for the human baseline estimation. We present human baseline estimation results using F1 score (See Table 2) and confusion matrix (See Fig. 8). In Table 2, we apply the same dataset to our classifier to compare the results. From the human baseline experiment, we conclude that the trained group shows better performance than the non-trained group and that our classifier shows better performance than the trained group.

5.3 Analysis


In addition to the analysis on the individual media, we analyze the results of our classifier in the following viewpoints. The metrics we use for the analysis are confusion matrix, F1 score and CAM.




① Oilpaint ② Pastel ③ Pencil ④ Watercolor

(a) An example of question: A figure with four candidates

#1	① Oilpaint	② Pastel	③ Pencil	④ Watercolor
#2	① Oilpaint	② Pastel	③ Pencil	④ Watercolor
#3	① Oilpaint	② Pastel	⑥ Pencil	④ Watercolor
#4	① Oilpaint	② Pastel	③ Pencil	④ Watercolor
#5	① Oilpaint	② Pastel	③ Pencil	④ Watercolor
#6	① Oilpaint	② Pastel	③ Pencil	④ Watercolor
#7	① Oilpaint	② Pastel	③ Pencil	④ Watercolor
#8	① Oilpaint	② Pastel	③ Pencil	④ Watercolor
#9	① Oilpaint	② Pastel	⑥ Pencil	④ Watercolor
#10	① Oilpaint	② Pastel	③ Pencil	④ Watercolor


Count answers

① 4	② 3	③ 2	④ 1
-----	-----	-----	-----


Select the best answer

User selection: ① Oilpaint

(b) The process of voting: Sampling 10 answers from the subjects and counting answers to select the best answer

Fig. 7. An example of a question for human baseline experiment.

5.3.1 F1 score-based analysis

We analyze best-classified media and worst-classified media using F1 score. Since the F1 score is the harmonic averaging of precision and recall, a medium that has higher F1 score is regarded as the media that is more correctly classified. The low F1 score means that our classifier tends to misclassify the medium to other medium. According to the F1 scores of the experiments, the oilpaint and pencil are the best media and the watercolor is the worst predicted by our classifier with 600 dataset (See [Table 2](#)). It is interesting to note that our classifier with 100 dataset classifies oilpaint and pastel better than pencil and watercolor. In the human baseline experiment, human subjects classify oilpaint and pastel better than pencil and watercolor.

5.3.2 Confusion matrix-based analysis

We analyze the most confusing pair of media using confusion matrix. We measure the most confusing pair by adding the misclassification probabilities in the confusion matrix, which are the diagonal elements (See [Table 3](#)). In our analysis, oilpaint-watercolor and pastel-pencil pairs are the most confusing pairs by our classifier and oilpaint-pastel and oilpaint-pencil pairs are the least confusing pairs. For human baseline experiments, oilpaint-pastel and pastel-pencil pairs are the most confusing and oilpaint-pencil and pastel-water are the least confusing. It is interesting to note that oilpaint-pencil pair is least confusing pair for classifier and human baseline and pastel-pencil pair is most confusing pair.

We also look into CAM for the reason of confusion. According to their CAM in [Fig. 9 \(a\)](#), the oil paint that is misclassified as watercolor shows high background reasoning regions on the area of flat stroke texture and low saturation. Another most confusing pair is pencil and pastel. According to their CAM in [Fig. 9 \(b\)](#), the confused artwork shows high CAM value in rubbed or monochrome regions. The most confused pairs are very difficult to classify correctly, since the physics of the confused media resembled each other. The flatly placed oilpaint strokes resembles the color textures expressed by watercolor painting and the rubbed pencil strokes and pastel strokes are very similar.

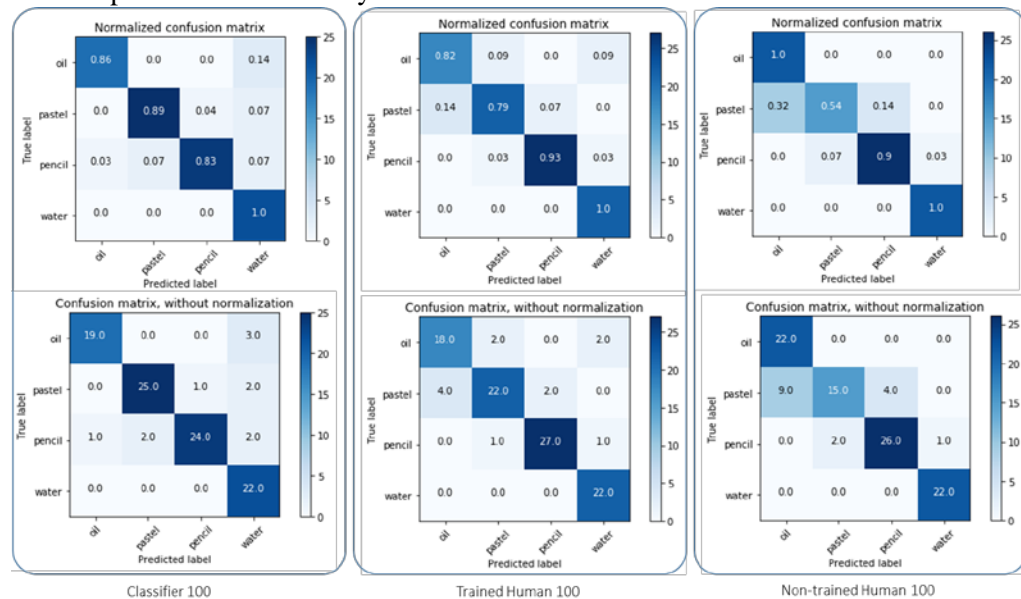


Fig. 8. Confusion matrix of human baseline experiments and our classifier which uses the same dataset with human baseline experiment.

5.4 Limitation

The following cases are the limitations for our classifier.

- (1) Our classifier misclassifies those media whose physical properties are similar. For example, pencil and pastel that deposits pigments on a paper surface through the abrasion of the media are the most confusing pair. Furthermore, oilpaint and watercolor that use brush to deposit pigments are another confusing pair. Since our classifier classifies media from the stroke marks they produce on a paper surface, the pair of media that share similar physical properties tends to be misclassified.
- (2) Another limitation is an artwork drawn by mixed media. Our classifier misclassifies watercolor artworks whose preliminary sketch is drawn using pencil (see Fig. 10). Since the pencil strokes used for the sketch are apparent in the watercolor painting, our classifier confuses to classify such artworks.

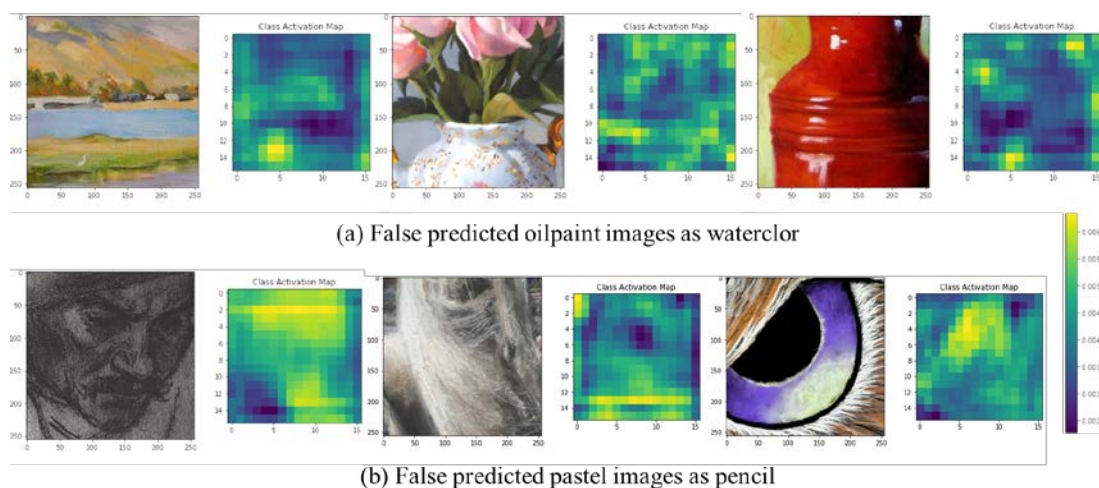


Fig. 9. The class activation map (CAM) of the most confusing pairs (Yellow regions are highly confounded regions).

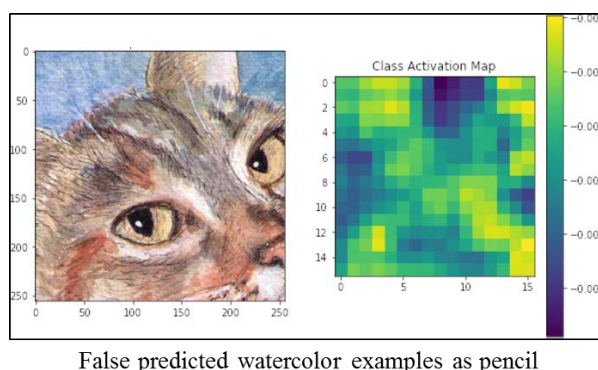


Fig. 10. Limitation caused by mixed media.

6. Conclusion and future work

We present a deep CNN-based approach for classifying stroke-based artistic media from artwork images. We employ VGGNet, one of the most prominent and widely-used CNN, to classify media and compare our result with that of human baseline experiment. To our experiment, we conclude that the performance of object classifying deep CNN is very effective for classifying artistic media and that the classifier shows better performance than human.

Our future research plan is to employ various CNN structures to show that the progress of object classifying CNN is effective in the classification of artistic media. Another research plan is to analyze human recognition and classification process and to apply the process for recognizing and classifying artistic media. Our third plan is to extend our classification structure to understanding the comprehending artworks.

Acknowledgement

This research was supported by a research grant from Sangmyung Univ. in 2017.

Reference

- [1] Borth, D., Ji, R., Chen, T., Breuel, T., and Chang, S. F., "Large-scale visual sentiment ontology and detectors using adjective noun pairs," in *Proc. of the 21st ACM international conference on Multimedia*, pp. 223-232, 2013. [Article \(CrossRef Link\)](#).
- [2] Chu, W.-T. and Wu, Y.-L., "Deep correlation features for image style classification," in *Proc. of ACM Multimedia*, pp. 402-406, 2016. [Article \(CrossRef Link\)](#).
- [3] Condorovici, R., Florea, C. and Vertan, C., "Automatically classifying paintings with perceptual inspired descriptors," *Journal of Visual Communication and Image Representation*, vol. 26, pp. 222-230, 2015. [Article \(CrossRef Link\)](#).
- [4] Dai, J., Li, Y., He, K., and Sun, J., "R-fcn: Object detection via region-based fully convolutional networks," in *Advances in neural information processing systems*, pp. 379-387, 2016. [Article \(CrossRef Link\)](#).
- [5] Datta, R., Joshi, D., Li, J., and Wang, J., "Studying aesthetics in photographic images using a computational approach," in *Proc. of European Conference on Computer Vision*, pp. 288-301, 2006. [Article \(CrossRef Link\)](#).
- [6] Dhar, S., Ordonez, V. and Berg, T., "High level describable attributes for predicting aesthetics and interestingness," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1657-1664, 2011. [Article \(CrossRef Link\)](#).
- [7] Gatys, L., Ecker, A., Bethge, M., "Image style transfer using convolutional neural networks," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2414-2423, 2016. [Article \(CrossRef Link\)](#).
- [8] Gygli, M., Grabner, H., Riemenschneider, H., Nater, F., and Van Gool, L., "The interestingness of images," in *Proc. of the IEEE International Conference on Computer Vision*, pp. 1633-1640, 2013. [Article \(CrossRef Link\)](#).
- [9] Isola, P., Xiao, J., Torralba, A., and Oliva, A., "What makes an image memorable?," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 145-152, 2011. [Article \(CrossRef Link\)](#).
- [10] Karayev, S., Trentacoste, M., Han, H., Agarwala, A., Darrell, T., Hertzmann, A., and Winnemoeller, H., "Recognizing image style," in *Proc. of the British Machine Vision Conference 2014*, pp. 1-20, 2014. [Article \(CrossRef Link\)](#).

- [11] Keren, D., "Painter identification using local features and naive Bayes," in *Proc. of ICPR 2002*, pp. 474-477, 2002. [Article \(CrossRef Link\)](#).
- [12] Lecoutre, A., Negrevergne, B. and Yger, F., "Recognizing art styles automatically in painting with deep learning," in *Proc. of Asian Conference on Machine Learning*, pp. 327-342, 2017. [Article \(CrossRef Link\)](#).
- [13] Long, J., Shelhamer, E. and Darrell, T., "Fully convolutional networks for semantic segmentation," in *Proc. of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 3431-3440, 2015. [Article \(CrossRef Link\)](#).
- [14] Matsuo, S. and Yanai, K., "CNN-based style vector for style image retrieval," in *Proc. of International Conference on Multimedia Retrieval*, pp. 309-312, 2016. [Article \(CrossRef Link\)](#).
- [15] Mensink, T. and van Gemert, J., "The Rijksmuseum Challenge: Museum-centered visual recognition," in *Proc. of ICMR*, pp. 451, 2014. [Article \(CrossRef Link\)](#).
- [16] Shamir, L., Macura, T., Orlov, N., Eckley, D., and Goldberg, I., "Impressionism, Expressionism, Surrealism: Automated recognition of painters and schools of art," *ACM Trans. Applied Perc.*, vol. 7(2), pp. 8, 2010. [Article \(CrossRef Link\)](#).
- [17] Shamir, L., Nissel, J. and Winner, E., "Distinguishing between abstract art by artists vs. children and animals: Comparison between human and machine perception," *ACM Transactions on Applied Perception*, Vol. 13, No. 3, pp. 17, 2016. [Article \(CrossRef Link\)](#).
- [18] Simonyan, K., and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014. [Article \(CrossRef Link\)](#).
- [19] Selim, A., Mohamed E., and Linda D., "Painting style transfer for head portraits using convolutional neural networks," *ACM Transactions on Graphics*, vol. 35(4), pp. 129, 2016. [Article \(CrossRef Link\)](#).
- [20] Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R., "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15(1), pp. 1929-1958, 2014. [Article \(CrossRef Link\)](#).
- [21] Sun, T., Wang, Y., Yang, J. and Hu, X., "Convolution neural networks with two pathways for image style recognition," *IEEE Transactions on Image Processing*, vol. 26(9), pp. 4102-4113, 2017. [Article \(CrossRef Link\)](#).
- [22] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A., "Learning deep features for discriminative localization," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2921-2929, 2016. [Article \(CrossRef Link\)](#).



Heekyung Yang received her B.S. degree and M.S. degree from Sangmyung University, Seoul, Korea, in 2010 and 2012, respectively. She is currently a Ph. D. candidate in the same college. Her major is computer graphics, especially NPR(non-photorealistic rendering). She is interested in machine learning, image processing, volume rendering and medical rendering.



Kyungha Min received his MS in Computer Science from KAIST in 1992. He received his BS and Ph.D in Computer Science and Engineering from POSTECH in 1994 and 2000, respectively. His main research interests are computer graphics and image processing.