Style Classification in Fine-Art Paintings Using Convolutional Neural Network

Ting Pan [*1] **Miaomiao Niu** [*1]

Abstract

Painting appreciation requires people to have certain amount of background information like art history, art genres, painting styles, distinguished painters, and etc. Such requirement creates a barrier for people who have not received relevant education to appreciate paintings in their life. To make painting appreciation more accessible to everyone, we decide to use machine learning algorithm to help people understand the painting context. There are many aspects that we can explore to describe the context. For example, we can identify the main color, painter, genres, styles, and even more abstract concept like composition. To make the project more feasible within few weeks, we simplify this task by building a model to predict the painting style since each style has its own special characteristics which are necessary for people to understand the context. There are two steps in the training process of our model. We use VGG-16 to extract painting features, then we input the extracted features together with style labels to the classification model. Finally, we deploy the machine learning model using Python Flask to our demo product which allows users can upload a picture and get the corresponding painting style together with the explanatory text of style characteristics. In our user scenario, we consider possible needs of visually impaired people and use text-to-speech API to generate audio. In this paper, we present the workflow of training and tuning our machine learning model for painting style classification. We experiment how the number of styles in the classification model will affect the accuracy, and find that the accuracy will drop when the number of prediction class increases.

1. Introduction

Nowadays there are growing demands for machine learning applications in accessibility. We can see some products which use image recognition to help visually impaired people know the physical environment around them. Also, there are some products that translate speaking into text simultaneously so that hearing-impaired people are able to know the content. However, current machine-learning-based products for people with accessibility needs are far from enough. For example, for visually impaired people who are curious about fine-art paintings, it's difficult to perceive the content even if they are allowed to touch these art pieces. This is where our motivation comes from and we want to use machine learning to extract features from painting and compile them into descriptive text information. When we try to describe painting content, it includes lots of information like color, style, genre, characters, painter and etc. To make our project more feasible, we decide to narrow down the scope and focus on style classification. Distinguishing different styles is important in art appreciation and each style is associated with background information like history, art movement, and representative artists. If we are able to train a machine learning model that can accurately predict corresponding painting style, we move a great step towards the goal of describing painting content to visually impaired people. In addition, the usefulness of style classification is not limited to visually impaired people. For common people who are not exposed to formal art education, it's difficult for them to distinguish and name different painting styles. Allowing users to know the style by simply uploading an image will make painting appreciation more accessible to everyone.

Before training our style classification model, we study and examine some formal approaches. Most of the previous work only utilizes simple features such as color, shades, and combination of them or focuses on identifying the artist from the painting. Besides, the accuracy achieved is not satisfying. Khan (Khan et al., 2010) uses a bag-of-word (BoW) approach with color and shades features to infer the painter. BoW approach is inadequate for recognizing the genre of the painting because it can only identify the painter. Falomir (Falomir et al., 2018) introduces QArt-Learn approach which uses K-Nearest Neighbor and Support Vector Machines to analyzes color to describe the style of paintings. This method focuses on only three styles(Baroque, Impres-

[*]Equal contribution [1]School of Information, University of Texas at Austin, TX, United States. Correspondence to: Ting Pan <ting.pan@utexas.edu>, Miaomiao Niu <miaomiao_niu@utexas.edu>.

sionism, and Post-Impressionism). We would like to explore to extend our model to more styles by analyzing more features like shades, texture, and edges. Arora (Arora & Elgammal, 2012) presents a comparative study of different classification methodologies which mainly use Intermediate level features and Semantic-level features. Although these models do not use low-level features like color, light, shades, and texture, the highest accuracy is only 65.4%.

For the dataset preparation, we choose paintings from wikiart which are already classified into different styles to create our own dataset. We compare different pre-trained convolutional neural network models and decide to use VGG16 to extract image features because it is widely used for large-scale image recognition and has relatively high efficiency. Different to some methods which focus on low-level features like color (Tominaga, 1992) or inferring the painter first to decide painting style (Keren, 2002), our approach uses high-level features for feature extraction. In addition, our approach strikes a balance between the number of feature extraction and accuracy to offer low computational and data costs. We explore the actual application scenario, implement our research findings and deploy our machine learning model to a real product. For example, we develop a web application which allows users to upload an image and obtain the corresponding painting style. We use wikipedia API and text-to-speech API to provide functions like offering style characteristics and real-time audio output so that people can strengthen their comprehension of certain styles.

2. Related Topics

Color classification Tominaga describes a color classification method that partitions a color image into a set of uniform color regions(Tominaga, 1992). Chai addresses an image classification technique that uses the Bayes decision rule for minimum cost to classify pixels into skin color and non-skin color(Chai & Bouzerdoum, 2000). Tominage proposes a color classification method that partitions color image data into a set of uniform color regions is described(Tominaga, 1990). Pietikainen considers a pattern recognition approach to accurate camera-based color measurements(Pietikainen et al., 1996). Our work is different from these works because color classification is a part of painting style classification. As a significant low-level feature, color identification is important in our work and there are a lot of work done in this field. And we will also use other medium-level and high-level features to improve our method.

Classification in Sculpture and Music Using quantitative rather than visual methods, Gansell approaches the levantine ivory sculptures classification task by exploiting computational methods from machine learning(Gansell et al., 2008). In this paper, Basili investigates the impact of machine learning algorithms in the development of automatic music classification models aiming to capture genres distinctions(Basili et al., 2004). Li proposes a new feature extraction method for music genre classification, DWCHs. DWCHs stands for Daubechies Wavelet Coefficient Histograms(Li et al., 2003). Our work is different from these works because we focus on paintings classification and these works are about music and sculpture. We could learn experiences from these works since we all apply classification approach to Fine-Art fields. However, there are a lot of gaps in our methods due to the difference of researching object.

Painter identification Widjaja presents a method for identifying painters using color profiles of skin patches in painting images(Widjaja et al., 2003). Khan uses the bag-of-word approach which basically describes the statistics of small image patches to infer the painter from a painting(Khan et al., 2010). The goal of Keren is to offer a framework for image classification ”by type”. This is accomplished by using local features, and by using the naive Bayes classifier(Keren, 2002). Our work is different from these works because we will directly identify the style of paintings instead of inferring the painters from paintings. Recognizing style is a more challenging task than painter identification.

Deep learning approaches in style classification Sandoval (Sandoval et al., 2019) introduces a new, two-stage image classification approach aiming to improve the style classification accuracy. At the first stage, the proposed approach divides the input image into five patches and applies a deep convolutional neural network (CNN) to train and classify each patch individually (Sandoval et al., 2019). At the second stage, the outcomes from the individual five patches are fused in the decision-making module, which applies a shallow neural network trained on the probability vectors given by the first-stage classifier (Sandoval et al., 2019). Tan (Tan et al., 2016) present a study on large-scale classification of fine-art paintings using the Deep Convolutional Network using wikiart dataset and achieve state-of-the-art results (68%) in overall performance. Rodriguez (Rodriguez et al., 2018) propose an approach that is based on transfer learning and classification of sub-regions or patches of the painting . Experimental validation based on two standard art classification datasets and six different pre-trained convolutional neural network (CNN) models (AlexNet, VGG-16, VGG-19, GoogLeNet, ResNet-50 and Inceptionv3) shows that the proposed approach outperforms the baseline techniques and offers low computational and data costs. (Rodriguez et al., 2018). Our work is different from these works because we will use different pre-trained models and art classification dataset scale. And we aim to achieve an accuracy higher than 68%.

Automatic description generation Guu (Guu et al., 2018) propose a new generative language model for sentences that first samples a prototype sentence from the training corpus and then edits it into a new sentence. Yao (Yao et al., 2017) present Long Short-Term Memory with Attributes (LSTM-A) - a novel architecture that integrates attributes into the successful Convolutional Neural Networks (CNNs) plus Recurrent Neural Networks (RNNs) image captioning framework, by training them in an end-to-end manner. Devlin (Devlin et al., 2015) compare the merits of two different language modeling approaches for the first time by using the same state-of-the-art CNN as input. Our work is different from these works because we not only rely on extracted features from image to generate a descriptive sentence, but also include useful external resources like wikipedia. After extracting features from paintings, we use them as keywords to search for broader information which could be useful for our users.

3. Methods

Our proposed approach for painting style classification mainly has two steps. First of all, we use VGG-16 and remove its fully connected layer to extract image features. The feature extraction process is shown in Figure 1. VGG-16 is a pre-trained model which is widely used in large-scale image recognition. Biggest benefit of using the VGG-16 pre-trained model is that it uses almost negligible time to train the dense layer with greater accuracy (Gupta et al., 2017). After generating image features using VGG-16, we apply them to different classification methods which include artificial neural network, SVM, and logistic regression as shown in Figure 2. We also add fully connected layer and softmax classification layer to form the complete VGG-16 model to conduct classification. Each classification method is optimized to maximize the overall style recognition accuracy. We test the performance of different classifier and the one with best performance is used in our final model.

4. Experimental Design

We retrieve our dataset from wikiart which has a collection of more than 80,000 fine-art paintings and 27 different styles. We download the dataset using the resources described in Tan's paper (Tan et al., 2016) and choose some styles of the original dataset to create new datasets so that it will be suitable for our experiment. As shown in table 1, the complete dataset contains 27 styles and each style contains different number of paintings. We choose 8 styles that each contains more than 4000 paintings and create a small dataset as shown in table 2 to identify best classification method. We also generate two training datasets of 2 styles and 4 styles as shown in table 3 and 4. We compare the accuracy of our model with different types of training datasets and study how the different types of training data can affect the accuracy. Classification accuracy is our evaluation metric. We use the percentage of correct predictions to evaluate the performance of different models. The model with highest accuracy will be our final model.

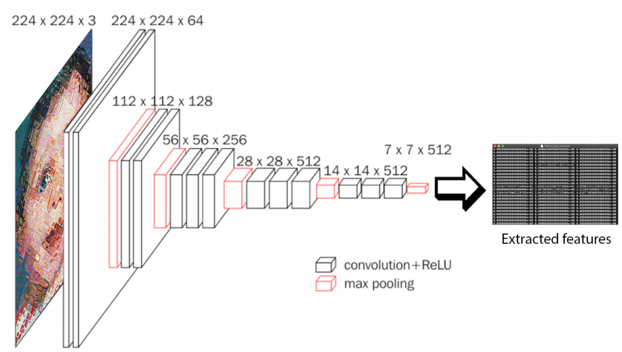

Figure 1. VGG-16 (remove fully connected layer) to extract image feature.

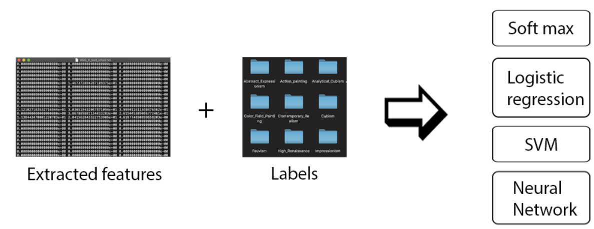

Figure 2. Input image features and labels to classifiers.

4.1. Experiment 1: VGG16-ANN with 8 styles dataset

In this experiment, we use artificial neural network as classifier and optimize its parameters to compare its performance with other classification methods.

Datasets. We use the dataset of 8 styles which consists of eight painting styles as shown in table 2. And we tune the parameters to achieve best performance.

Baselines. A baseline test is performed using the 8 styles dataset as inputs to the original artificial neural network model without tuning parameter.

Evaluation Metrics. We use classification accuracy because it is straightforward to present the model's performance, and make sure it is consistent among all experiments.

Table 1. Complete dataset.

Index	Style	Number
0	Abstract Expressionism	2783
1	Action painting	99
2	Analytical Cubism	111
3	Art Nouveau	4335
4	Baroque	4242
5	Color Field Painting	1616
6	Contemporary Realism	482
7	Cubism	2236
8	Early Renaissance	1392
9	Expressionism	6737
10	Fauvism	935
11	High Renaissance	1344
12	Impressionism	13061
13	Mannerism Late Renaissance	1280
14	Minimalism	1338
15	Naive Art Primitivism	2406
16	New Realism	315
17	Northern Renaissance	2553
18	Pointillism	514
19	Pop Art	1484
20	Post Impressionism	6452
21	Realism	10734
22	Rococo	2090
23	Romanticism	7020
24	Symbolism	4529
25	Synthetic Cubism	217
26	Ukiyo e	1168

Table 2. 8 style dataset.

Index	Style	Number
3	Art Nouveau	4335
4	Baroque	4242
9	Expressionism	6737
12	Impressionism	13061
20	Post Impressionism	6452
21	Realism	10734
23	Romanticism	7020
24	Symbolism	4529

Table 3. 4 style dataset.

Index	Style	Number
9	Expressionism	6737
12	Impressionism	13061
20	Post Impressionism	6452
21	Realism	10734

Table 4. 2 style dataset.

Index	Style	Number
3	Art Nouveau	4335
4	Baroque	4242

4.2. Experiment 2: VGG16-PCA-SVM with 8 styles dataset

In this experiment, we use SVM as classifier. We use PCA to reduce dimensionality and compare its performance with other classification methods.

Datasets. We use the dataset of 8 styles which consists of eight painting styles as shown in table 2. And we tune the parameters to achieve best performance.

Baselines. A baseline test is performed using the 8 styles dataset as inputs to the original SVM model without tuning parameter.

Evaluation Metrics. We use classification accuracy because it is straightforward to present the model's performance, and make sure it is consistent among all experiments.

4.3. Experiment 3: VGG16-Logistic regression with 8 styles dataset

In this experiment, we use Logistic Regression as classifier to compare its performance with other classification methods. And we tune the parameters to achieve best performance.

Datasets. For this experiment, we use the 8 style dataset as shown in table 2.

Baselines. A baseline test is performed using 8 styles dataset as inputs to the logistic regression model without tuning parameter.

Evaluation Metrics. We use classification accuracy because it is straightforward to present the model's performance, and make sure it is consistent among all experiments.

4.4. Experiment 4: VGG16 with 8 styles dataset

We add fully connected layer and softmax classification layer to classify the features generated before. Then we compare its performance with other classification methods that using 8 styles dataset.

Datasets. We use the 8 styles dataset as shown in table 2.

Baselines. A baseline test is performed using the 8 styles dataset as inputs to the VGG16 model.

Evaluation Metrics. We use classification accuracy because it is straightforward to present the model's performance, and make sure it is consistent among all experiments.

4.5. Experiment 5: VGG16 with 4 styles dataset

After experimenting on the 8 styles dataset, we find that the accuracy is not satisfying enough and we want figure out the reason. That's why we decide to do more experiments using 4 styles and 2 styles datasets as shown in table 3 and table 4. In this experiment, we add fully connected layer and softmax classification layer to classify the features generated before. And we use the 4 styles dataset and measure its classification accuracy.

Datasets. We use the 4 styles dataset as shown in table 3.

Baselines. A baseline test is performed using the 4 styles dataset as inputs to VGG16 model.

Evaluation Metrics. We use classification accuracy because it is straightforward to present the model's performance, and make sure it is consistent among all experiments.

4.6. Experiment 6: VGG16 with 2 styles dataset

We add fully connected layer and softmax classification layer to classify the features generated before. Different to previous experiments, this time we use the dataset that only has two styles.

Datasets. 2 styles dataset as shown in table 4.

Baselines. A baseline test is performed using the 2 styles dataset as inputs to the VGG16 model without tuning parameter.

Evaluation Metrics. We use classification accuracy because it is straightforward to present the model's performance, and make sure it is consistent among all experiments.

5. Experimental Results

Firstly, we only use the 8 styles dataset to measure the performance of different models. Among the first four experiments, we find that VGG16 - Logistic regression model performs best and it has an accuracy of 0.34913. So we decide to deploy this model in our product. The classification accuracy of each model is presented in the table 5.

Table 5. Experimental findings.

Dataset	Model	Accuracy
8 Styles Dataset	VGG16-ANN	0.34913
	VGG16-PCA-SVM	0.22881
	VGG16-Logistic	0.34980
	VGG16	0.31326
4 Styles Dataset	VGG16	0.55783
2 Styles Dataset	VGG16	0.864

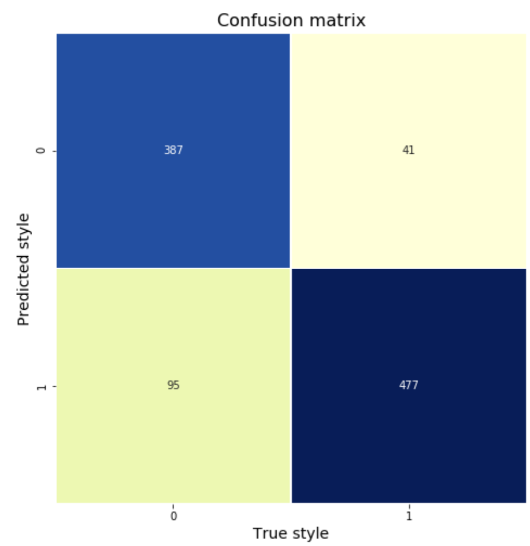

Figure 3. Confusion matrix of 2 styles.

Secondly, we conduct another two comparative experiments and find that as the number of styles increases, the accuracy drops obviously. When we use VGG16 model to conduct classification, we achieve accuracy of 0.864 on 2 styles dataset and accuracy of 0.55783 on 4 styles dataset. By contrast, we only get an accuracy of 0.31326 when we classify 8 styles. Figure 3 shows the confusion matrix of 2 styles dataset. Figure 4 shows the confusion matrix of classifying 4 styles and Figure 5 presents the confusion matrix of classifying 8 styles.

In our experiment, we have not explored the reason for the difference of accuracy when using dataset that consists of different styles. In the future work, we would also like to apply more pre-trained CNN models, such as AlexNet, VGG-19, GoogLeNet, ResNet-50 and Inceptionv3 and compare their performances with VGG-16. We are

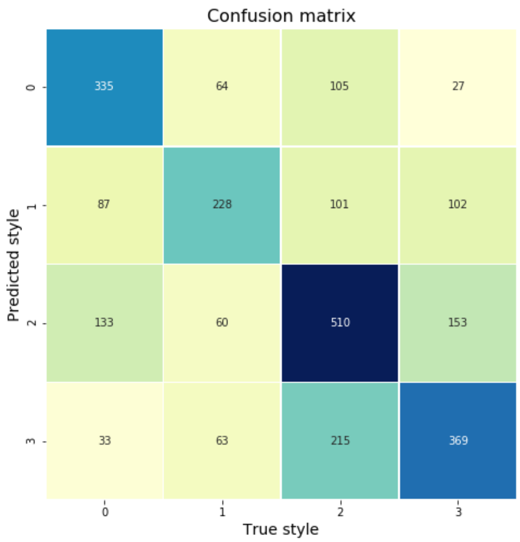

Figure 4. Confusion matrix of 4 styles.

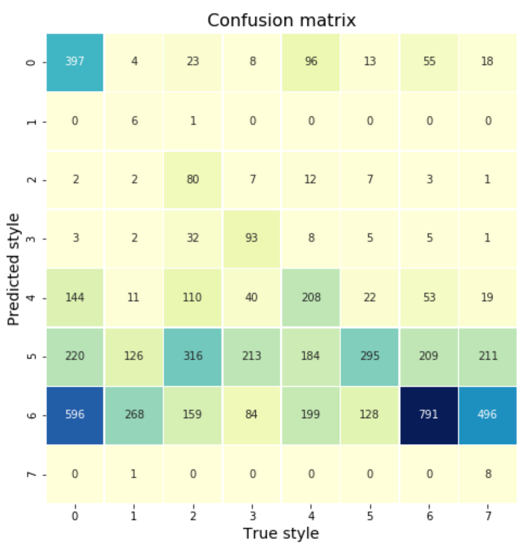

Figure 5. Confusion matrix of 8 styles.

interested in developing a better model to classify paintings with higher accuracy.

In addition to training the machine learning model for painting style classification, we deploy the machine learning model that is trained by 8 style dataset and develop a web

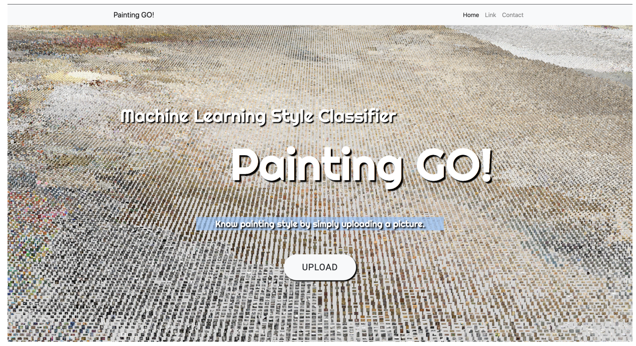

Figure 6. User interface of uploading picture.

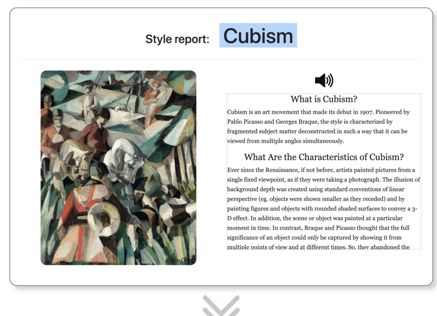

Figure 7. User interface of prediction result.

application that enables users to know the painting style by simply uploading a picture as shown in Figure 6. Users can even learn more about certain painting styles since after identifying the painting style, our web application will provide style characteristics as shown in Figure 7 and more paintings that have the same style as shown in Figure 8.

6. Conclusions

In our work, we conduct experiments on painting style classification using wikiart dataset. We compare the performance of different models including VGG16-ANN, VGG16-PCA-SVM, VGG16-Logistic regression and VGG16. VGG16-Logistic regression has the highest accuracy. We also test VGG16's performance using multiple datasets of different size. As the number of styles in the dataset increases, the classification accuracy declines. Common people usually has 60% accuracy in identifying the painting style. When classify two painting styles, our approach shows that it outperforms common people. However, when the number of styles increases, the accuracy of our approach drops which is what we need to improve in the future.

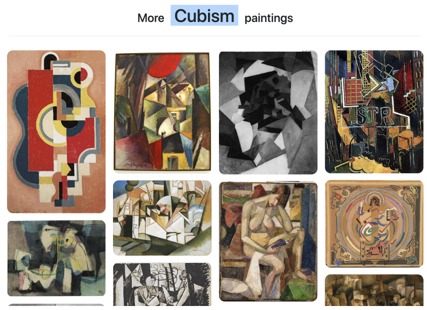

Figure 8. User interface of learning more paintings.

In the future, we want to explore more possibilities of our current project. First, we would like to work on optimization of the machine learning models to achieve higher accuracy. Second, we want to try more ways to extract image feature and translate extracted features into comprehensive language. For example, we would like to research on how people perceive the painting content in terms of people's imagination and different age groups. It will contribute to realizing the goal of making painting appreciation accessible to everyone. Last but not least, we want to test some not well-know paintings, like paintings composed by ourselves, to see how our machine learning model will predict the style. It could be an interesting research about potential biases of our machine learning algorithm.

References

Arora, R. S. and Elgammal, A. Towards automated classification of fine-art painting style: A comparative study. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pp. 3541–3544. IEEE, 2012.

Basili, R., Serafini, A., and Stellato, A. Classification of musical genre: a machine learning approach. In *ISMIR*, 2004.

Chai, D. and Bouzerdoum, A. A bayesian approach to skin color classification in ycbcr color space. In *2000 TENCON Proceedings. Intelligent Systems and Technologies for the New Millennium (Cat. No. 00CH37119)*, volume 2, pp. 421–424. IEEE, 2000.

Devlin, J., Cheng, H., Fang, H., Gupta, S., Deng, L., He, X., Zweig, G., and Mitchell, M. Language models for image captioning: The quirks and what works. *arXiv preprint arXiv:1505.01809*, 2015.

Falomir, Z., Museros, L., Sanz, I., and Gonzalez-Abril, L. Categorizing paintings in art styles based on qualitative color descriptors, quantitative global features and machine learning (qart-learn). *Expert Systems with Applications*, 97:83–94, 2018.

Gansell, A. R., Tamaru, I. K., Jakulin, A., and Wiggins, C. H. Predicting regional classification of levantine ivory sculptures: a machine learning approach. *arXiv preprint arXiv:0806.4642*, 2008.

Gupta, D., Jain, S., Shaikh, F., and Singh, G. Transfer learning & the art of using pre-trained models in deep learning. *Analytics Vidhya*, 2017.

Guu, K., Hashimoto, T. B., Oren, Y., and Liang, P. Generating sentences by editing prototypes. *Transactions of the Association of Computational Linguistics*, 6:437–450, 2018.

Keren, D. Painter identification using local features and naive bayes. In *Object recognition supported by user interaction for service robots*, volume 2, pp. 474–477. IEEE, 2002.

Khan, F. S., van de Weijer, J., and Vanrell, M. Who painted this painting. In *2010 CREATE Conference*, pp. 329–333, 2010.

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Li, T., Ogihara, M., and Li, Q. A comparative study on content-based music genre classification. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pp. 282–289. ACM, 2003.

Pietikainen, M., Nieminen, S., Marszalec, E., and Ojala, T. Accurate color discrimination with classification based on feature distributions. In *Proceedings of 13th International Conference on Pattern Recognition*, volume 3, pp. 833–838. IEEE, 1996.

Rodriguez, C. S., Lech, M., and Pirogova, E. Classification of style in fine-art paintings using transfer learning and weighted image patches. In *2018 12th International Conference on Signal Processing and Communication Systems (ICSPCS)*, pp. 1–7. IEEE, 2018.

Sandoval, C., Pirogova, E., and Lech, M. Two-stage deep learning approach to the classification of fine-art paintings. *IEEE Access*, 2019.

Tan, W. R., Chan, C. S., Aguirre, H. E., and Tanaka, K. Ceci n'est pas une pipe: A deep convolutional network for fine-art paintings classification. In *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 3703–3707. IEEE, 2016.

Tominaga, S. A color classification method for color images using a uniform color space. In *[1990] Proceedings. 10th International Conference on Pattern Recognition*, volume 1, pp. 803–807. IEEE, 1990.

Tominaga, S. Color classification of natural color images. *Color Research & Application*, 17(4):230–239, 1992.

Widjaja, I., Leow, W. K., and Wu, F.-C. Identifying painters from color profiles of skin patches in painting images. In *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)*, volume 1, pp. I–845. IEEE, 2003.

Yao, T., Pan, Y., Li, Y., Qiu, Z., and Mei, T. Boosting image captioning with attributes. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4894–4902, 2017.