Multi-scale Painter Classification

Irina M. Ciortan[1], Yoko Arteaga[1,2], Sony George[1], and Jon Y. Hardeberg[1]

[1] NTNU - Norwegian University of Science and Technology, 2815 Gjøvik, Norway
irina-mihaela.ciortan@ntnu.no
[2] C2RMF - Center for Research and Restoration of the Museums of France, Paris, France

Abstract. The characterization of a painter's style is useful for a series of applications, such as documenting art history, planning style-aware conservation and restoration, and discarding forgery attempts. In this work, we propose a method to assign paintings to the right artist with two strategies: traditional machine learning and deep learning. In particular, we quantify the visual characteristics of a painting at multiple scales, covering low-level as well as mid-level features (pyramid of histogram of oriented gradients, residual convolutional neural network features). We focus on coeval artists, representing Impressionism, Expressionism and Cubism art periods. Our results are consistent with state-of-the-art findings in art and computer vision literature.

Keywords: Art Attribution; Multi-scale Classification; Machine Learning; Pyramid of Histogram of Oriented Gradients; Residual Neural Network.

1 Introduction

In the field of art history, the identification of artists, and moreover the attribution of a disputed artwork to a particular artist is a complex task with well-recognised challenges. Specialised art historians and curators would usually consult a number of sources in an attempt to discern between artworks. First-hand, they would recur to visual examination to appreciate the stylistic coherence of the paintings themselves, but also to historical archives to verify provenance. Analytical methods from conservation science are carried out in order to obtain a more detailed, evidence-based understanding of artists' materials, artworks' age and execution, stylistic technique, and any changes the artworks may have undergone [8].

In the light of the digital era, "computational connoisseurship" [11] has gained stronger credibility as there have been more and more cases where computational tools have aided art attribution. To name only a few: automation of canvas thread count and weave pattern analysis of Vermeer paintings [23], watermark identification in Rembrandt's etchings, and photographic paper classification [11].

2 Related Work

Painter identification has been addressed through computational digital image analysis in many scientific works over the last two decades [7]. While the first attempts were designing handcrafted features in a traditional machine learning setting, the take off and increasing accessibility of convolutional neural networks (CNN) has greatly enabled the application of deep learning for artworks analysis.

Machine Learning Approaches. As one of the earliest works on painter identification, Keren [18] trained a Naive Bayes classifier to identify 5 different painters (Rembrandt, van Gogh, Picasso, Magritte, and Dali). They used discrete cosine transform (DCT) coefficients as features and obtained 86% of accuracy. In another work, Widjaja et al. [29] conducted an artist identification for Nude paintings. They exploited skin color as feature and support vector machines (SVM) as classifier to attribute paintings to Rubens, Michelangelo, Ingres, and Botticelli. Wavelets derived with a multi-resolution hidden Markov model (HMM) have been used in Li and Wang's research [22] on grayscale images of traditional Chinese ink paintings to recognize between 5 artists.

In the attempt to identify 5 artists (Cezanne, Van Gogh, Monet, Picasso, Dali), Blessing and Wen [5] tested 15 simple and advanced features, including color histograms, histogram of oriented gradients (HOG) [9], local binary patterns (LBP), and Dense SIFT (for the full list please refer to the paper). The features were employed to assign paintings to one of two artists (binary classification), as well as one out of all 5 artists (multi-class classification). For both classification tasks, the highest performance was given by HOG features. In the binary case, the accuracy ranges from 90.2% (Picasso/Cezanne) to 95.9% (Van Gogh/Cezanne), while in the multi-categorical case, the average overall accuracy was 82%. With higher number of classes, the accuracy given the same feature, decreases due to increased complexity and variation.

Keshvari and Chalechale [19] use traditional machine learning to classify 5 Iranian painters in binary and multi-class fashion. In total, 6 features are defined: 256-bin histogram, 18-bin histogram, 16-bin histogram, Gabor filter, HOG, LBP, concatenated HOG and LBP. For the binary classification, the HOG features were the most discriminative, while for the multi-class task, LBP was overall outperforming HOG. Moreover, two models, namely SVM and K-nearest neighbor (KNN) are compared, SVM giving the best accuracy.

In brief, as mentioned in [7], comparing these identification methods is not an easy task, as different datasets are used. It can be concluded that most of the features and descriptors in these studies are mainly low level features (color histograms, edge histograms and texture descriptors). Particularly, HOG has been found to have a good discerning power [19], [5]. Furthermore, in these studies, multiple classifiers are tested out.

Deep Learning Approaches. Towards recognizing 91 painters based on multi-class image classification, Bianco et al. [4] propose a network that is sensitive to multi-scale features by training in parallel 3 branches that have 3 different inputs, as follows: two random crops of 227x227 pixels extracted after resizing the original image to 512x512 resolution and another 227x227 crop after resizing the original image to 256x256. The network is formed by residual blocks, similar to those of Residual Neural Network (ResNet) [12], with an inversion in the order of the batch normalization layers and the summation of the skip connection layer. The model achieves a 78.8% accuracy as the average of a 10-fold classification experiment. In line with the findings of [27], the authors of [4] confirm that the highest confusion happens for contemporaneous and coeval artists. Viswanathan [28] compared 3 CNN designs based on their accuracy in assigning unseen painting to 57 artists: a simple CNN with only 2 convolutional layers trained from scratch, the Residual Neural Network architecture with 18 convolutional layers (ResNet18), trained from scratch on the painting images and ResNet18 pretrained on ImageNet dataset [25] fine tuned on the artworks images. The fine tuning experiment outperformed the train-from-scratch designs, reaching a top-1 accuracy of 77% and a top-3 accuracy of 89.8% on the test images. With the intention of mapping art periods and style in a coherent and automatic manner using CNN features, Elgammal et al. [10] compare the suitability of state-of-the-art deep learning architectures (AlexNet [21], VGG[26] and ResNet[12]), in a transfer learning setup and a train-from-scratch setup. The overall best performance is given by the transfer learning methodology, where the highest accuracy (63.7%) is given by ResNet152. Bai et al.[2] proposed a Multi-layer Feature Fusion DenseNet[13] to combine shallow features with deep features. They trained the system for 23 painters and reached a multi-categorical classification accuracy of 86.83%. Within the Kaggle platform, a solution was proposed to recognize artists from a subset of the "Best Artworks of All Time" using ResNet50, reporting a multi-class accuracy of 84% on the validation set[17] for artists broadly spread over various historical periods.

As in the case of traditional machine learning approaches, the inconsistency of the studied datasets, makes it difficult to compare all the above described deep learning methods on a one-to-one correspondence basis. A common finding is that fine tuning outperforms learning from scratch methodologies. The reason for this is the comparatively small number of images of artworks to the datasets available for natural images. In addition to the dataset dimensionality, there also seems to be a lack of a unique, standardized and holistic image test-bed for paintings. Complexity of multi-class painter identification, especially if the painters represent similar art periods, is a challenge repeatedly pointed out in the literature for both machine learning and deep learning techniques. Another agreed upon intuition is that multi-scale analysis provides a more thorough understanding of the visual characteristics of a certain artist [14], [4], [2].

3 Method

In this work, we have addressed painter identification as a digital image classification task solved using traditional machine learning and deep learning. Following the insights from the literature review, we opted to deepen the analysis on HOG features and fine tuning ResNet, by testing both approaches at multiple scales.

3.1 Pyramid of Histograms of Oriented Gradients

In the first approach, a conventional machine learning approach is exploited where several statistics are extracted from Pyramid of Histograms of Oriented Gradients (PHOG) descriptors and multiple different classifiers are used to identify the painter. PHOG [6] computes HOG features at various sub-regions of an image that follow a pyramidal distribution. At each level, the image is split into 4^L sub-regions, where L is the level . Thus, level 0 corresponds to the initial image, level 1 corresponds to the initial image split into 4 sub-regions, level 2 to 16 sub-regions and so on and so forth. For each sub-region, HOG is computed on its gradient representation and then all the HOG vectors are concatenated into a single array, namely PHOG.

Based on PHOG features, several statistics were proposed by Amirshahi et al. [1] to assess the aesthetic quality of natural images and artworks: weighted self-similarity, complexity and anisotropy. Self-similarity is based on the notion that natural scenes, as well as artworks have a self-similar structure (fractal-like) in the Fourier domain.

To measure self similarity, first the the histogram intersection kernel HIK is computed:

$$HIK(h,h') = \Sigma_{i=1}^{m} min(h(i), h(i')) \tag{1}$$

where $h(i)$ and $h(i')$ represent the HOG features of two sub-regions in an image. HIK is unity for identical histograms.

From HIK, the self-similarity $SefS$ of an image I can be calculated with respect to the level at which HOG is being assessed, L:

$$SefS(I,L) = median(HIK(h(S), h(Ref(S)))) \tag{2}$$

where $h(S)$ is the HOG vector of a sub-region S in the image I at level L and $h(Ref(S))$ corresponds to the HOG vector of the reference region that the sub-region is compared to. The formula 2 outputs a number that assesses the self-similarity of an image at a certain level - the higher the number, the higher the self-similarity. However, the concept of self-similarity should take into account various levels of resolutions in order to verify for the fractal-like properties of natural images [1]. Hence, the weighted self-similarity was proposed, which balances the importance of scale for the self-similarity of an image:

$$WSefS(I) = \frac{1 - \sigma(SefS(I))}{\sum_{L=1}^{l} \frac{1}{L}} \cdot \sum_{L=1}^{l} \frac{1}{L} \cdot SefS(I,L) \tag{3}$$

where l is the total number of levels in the PHOG pyramid, $SefS(I)$ is the vector that concatenates the self-similarities for all levels $L = 1, ..., l$, and σ represents the standard deviation. The weighted self-similarity assigns greater importance to lower levels in the pyramid.

Anisotropy refers to the heterogeneity of an image across its orientations and it was found to be a property coherent among certain image categories

[20]. Mathematically, the anisotropy of an image is given by the variance of the gradient strength in the HOG vector across its bin entries [1]:

$$AnI(L) = \sigma^2(H(L)) \tag{4}$$

AnI represents the anisotropy in the image at level L, where $H(L)$ corresponds to the concatenated HOG vector at level L in the PHOG pyramid, and σ^2 is the variance. In aesthetical studies, complexity was identified by psychologist Berlyne [3] as crucial for the appreciation of artworks: there is a certain threshold of complexity up to which images are considered puzzling enough. While an image that is too simple might induce boredom in an observer, a highly complex image might lead to frustration and difficulty to interpreting its meaning. The complexity Co of an image can be computed as the mean of the image gradient G across all orientations in the x, y directions [1].

$$Co(I) = \frac{1}{N \cdot M} \sum_{x=1}^{M} \sum_{y=1}^{N} G(x, y) \tag{5}$$

In this equation, N and M are the height and width of the gradient G of an image I. Since image gradients represent the changes of lightness in an image, it is assumed that calculating the mean gradient over the luminance channel will give a good prediction of image complexity. The higher the mean absolute gradient, the more complex an image is. As opposed to weighted self-similarity and anisotropy, the complexity of an image is a global measure, extracted for the full image resolution and doesn't take sub-regions into account.

3.2 Residual Neural Network

As mentioned in the literature review, ResNet, pretrained on ImageNet dataset [25], has shown a good performance for artist identification tasks in a transfer learning setup. The ResNet architecture [12] contains modules of residual blocks, where each block is composed of convolutional layers, batch normalization layers and non-linear activation (ReLu) function layers. The most innovative component of the ResNet architecture is the use of skip connections that solves the problem of vanishing gradients in deep convolutional neural networks. The way a residual connection (also known as skip or shortcut connections) works is that it jumps over some layers, facilitating the forward propagation of the input through the network and thus, allowing for deeper architecture without placing a high burden on the training speed.

ResNet comes in shallower and deeper configurations according to the cardinality of the convolutional layers, leading to ResNetX series, where X stands for the number of core layers: ResNet18, ResNet50, ResNet101, etc... A higher cardinality of layers is proportional to the level of features that the networks is learning. Since we are interested in characterizing the artists' style and not in recognizing high-level semantics in the paintings, we assume that low-level and mid-level features are sufficient for our classification tasks. For this reason and in order to mitigate overfitting, we don't choose very deep architectures of ResNet.

4 Experimental Design

4.1 Dataset

As far as dataset is concerned, the collection of previous works show a heterogeneity of number of painters, where some painters are over-represented and others, under-represented. For example, none of the works include the Norwegian painter Edvard Munch. In our formulation of artist recognition, we narrowed down the search to late 19^{th} century to first half of the 20^{th} century art periods: Symbolism, Impressionim, Expressionism, Cubism; and to 8 famous representatives: Alfred Sisley (259 images), Edgar Degas (702 images), Edward Munch (382 images), Marc Chagall (239 images), Pablo Picasso (439 images), Paul Gauguin (311 images), Pierre-Auguste Renoir (336 images) and Vincent Van Gogh (877 images). Some of these painters have inspired each other and might have reiterated the same themes, which adds complexity to the task of differentiating among them. Most of the images were taken from the Kaggle dataset "Best Artworks of All Time" [16], but because there were few images for Edward Munch in this dataset, additional images of his paintings were downloaded from WikiCommons.

Given the sparsity of artist identification works in the literature that consider the Norwegian painter Edvard Munch into account, we have defined two classification experiments: binary classification between Munch and all the other painters, and multi-category classification on all painters against all painters. To the best of our knowledge, Edward Munch is underrepresented in most of the previous classification works and it is for this reason that we chose to revolve the binary categorization around Munch/not Munch.

Table 1. Test set accuracy for the Munch vs Others and All vs All classification tasks based on anisotropy, complexity, parent and ground self-similarity for multiscale PHOG pyramid, trained with SVM and KNN classifiers. While for the binary task, there are models that achieve an accuracy higher than 50%, in the multiclass case, the models have a rather poor performance.There is an overall slight improvement when splitting the image into 5 levels instead of 3.

		Fine KNN	**Medium KNN**	**Coarse KNN**	**Cosine KNN**	**Cubic KNN**	**Weighted KNN**	**Linear SVM**	**Quadratic SVM**	**Fine Gaussian SVM**	**Medium Gaussian SVM**	**Coarse Gaussian SVM**
PHOG 3 levels	**Binary**	**85.00 %**	76.60 %	65.70 %	69.10 %	77.10 %	70.90 %	61.70 %	67.80 %	**73.10 %**	64.50 %	63.90 %
	Multiclass	27.10 %	**31.60 %**	28.20 %	29.10 %	31.20 %	29.80 %	33.60 %	34.50 %	33.80 %	**36.90 %**	32.20 %
PHOG 4 levels	**Binary**	**84.50 %**	73.40 %	65.90 %	70.00 %	73.90 %	67.30 %	62.00 %	57.60 %	**73.20 %**	65.20 %	64.60 %
	Multiclass	30.10 %	31.40 %	30.20 %	31.10 %	**33.20 %**	29.90 %	34.60 %	38.10 %	37.30 %	**39.40 %**	33.90 %
PHOG 5 levels	**Binary**	**85.70 %**	75.50 %	65.60 %	70.50 %	76.10 %	69.90 %	61.60 %	63.10 %	**74.30 %**	64.90 %	64.40 %
	Multiclass	32.90 %	33.50 %	29.80 %	31.80 %	**34.30 %**	31.20 %	33.50 %	38.60 %	37.10 %	**39.00 %**	33.90 %

4.2 Specifications of Classification Experiments

We split our dataset into 80% training set and 20% test set, where 20% of the training data was held out for validation.

In the handcrafted classification approach, PHOG-based features (self-similarity, complexity and anisotropy) were calculated for the whole dataset using a Matlab implementation [24]. We compute anisotropy and self-similarity for 3 different scales of the PHOG pyramid (3,4,5). We extract the self-similarity with respect to the ground image, as well as the parent region and compute its average over the studied levels according to Eq. 3. This process results into a 4-feature vector for each level. Then, using the Matlab Classification Learner toolbox, SVM and KNN classifiers are trained with these features as input.

For the deep learning approach, we fine-tuned 3 configurations of ResNet, all pretrained on the ImageNet dataset. Even though our deep learning approach is similar to [17], we would like to clarify that we implemented different training strategies (freezing the pre-trained weights in a one-stage training process), we applied multi-scale analysis, we selected different painters for comparison, who were contemporaneous and similar in their styles, and we reported the results for the test instead of the validation set as in [17]. Given the rather small number of images in our dataset, we took several measures to prevent overfitting. Firstly, we didn't choose the very deep configurations of ResNetX, selecting ResNet18, ResNet34 and ResNet50, since higher layer depths leverage the risk of overfitting. Secondly, we tried different data augmentation options: random horizontal flip and zoom-in effect up to 20% so that we get a slight variation in scale without losing meaning and semantics of the image. However, horizontal flip was decreasing the performance of the trained models and we believe it is due to the fact that the horizontal orientations are idiosyncrasies of the painters' techniques, an insight confirmed as well by [27]. Nonetheless, we preserved the zoom-in factor, since it adds multi-scale variation that we try to account for in our work. Thirdly, for all three ResNet configurations, we froze the pretrained weights for all convolutional layers. Conversely, the fully connected layer was trained to match our data, where each class had different weights proportional to the number of available images, as a way to compensate for the unbalanced dataset. Lastly, we trained the network using Adam optimizer for a maximum number of 50 epochs, with early stop option (patience 20 epochs) and learning rate of 10^{-4} (that decays with a factor of 10^{-1} if the cross-entropy loss on the validation set has stopped improving for 5 epochs). We set up the deep learning experiments in Python programming language, using Tensorflow and Keras libraries on a CUDA-enabled GPU machine.

4.3 Discussion of results

In agreement with the findings of previous works in the literature, our classification methods give overall higher accuracy for the binary task than for the multi-categorical task. Tables 1 and 2 summarize our results. The accuracy values show that the deep learning approach outperforms the handcrafted features

Table 2. Test set accuracy for the Munch vs Others and All vs All classification tasks based on fine-tuned ResNet of depths. While it seems that the highest layer depth brings the highest accuracy, the difference is not higher than 2 units.

	ResNet18	**ResNet34**	**ResNet50**
Binary	93.66 %	93.79 %	**95.72 %**
Multi-class	78.93 %	77.39 %	**79.07 %**

based on PHOG representation. As a matter of fact, in the multi-class experiment, the trained machine learning models don't converge, as they don't achieve an accuracy higher than 50%. It is noteworthy to mention that the SVM and KNN classifiers were trained separately on the handcrafted features extracted for different levels (3,4,5) in the pyramid, and the highest accuracy was obtained by level 5, even though the difference among the layers is rather small (in some cases less than unit). This might be due to the fact that the higher level encompasses all the previous lower levels. Nonetheless, the outperformance of level 5 PHOG features suggests that including a larger range of scales covers more thoroughly the visual characteristics of an image.

Aiming for features at various scales seem to impact the performance of the pretrained ResNet as well. The highest accuracy is given by the 50-layer configuration, with the mention that the differences between various layer configurations do not exceed 2 units. ResNet50 manages to correctly distinguish Munch from other artists in 73% of the cases, leaving a 27% false negative rate. Fig. 1 highlights 4 randomly picked test images along with their predictions. The two Munch paintings containing human poses are correctly identified, while the landscape painting is a false negative. This could be related to the painting theme itself (natural scenery are sometimes more commonly depicted in arts) that makes the artists be more similar in some cases (natural scenery) than others (portraits), resulting in similar features as processed by the classification models. Moreover, in their early career, it is common that most artists start painting in a classic rendition, before defining their own style signature.

The confusion matrix of the multi-categorical classification with ResNet50, shown in Fig. 2 displays no red off-diagonal values, meaning that in average, the painters get correctly attributed with a rate of more than 50%. However the shades of lighter blue in the first column (corresponding to Van Gogh) indicate that many painters get misattributed to Van Gogh. A possible reason for this could be the unbalanced dataset (where Van Gogh has the highest number of training images), that the prior class weights didn't manage to fully account for. Several randomly selected images are illustrated in Fig. 3 and they represent correct predictions. Chagall is mistaken for Picasso with a false negative rate of 19% (see Fig. 4). Munch is confused 12% of the cases with Van Gogh and 8% of the cases with Degas. Interestingly, the similarities between Munch and Van Gogh have been previously emphasized in the world of art [15], meaning that our model is able to detect an acknowledged resemblance. Fig. 5 illustrates 4 Munch vs Others predictions by fine-tuned ResNet50. While the two paintings contain-

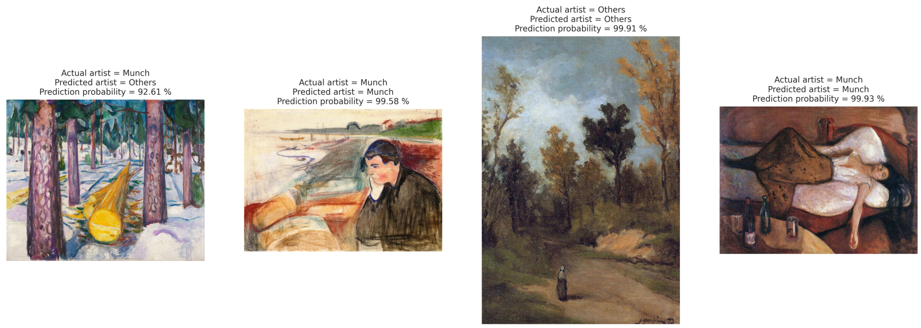

Fig. 1. Munch vs Others predictions by fine-tuned ResNet50. While the two paintings containing human poses get correctly identified as Munch, the landscape painting gets misattributed to Van Gogh with 92% probability.

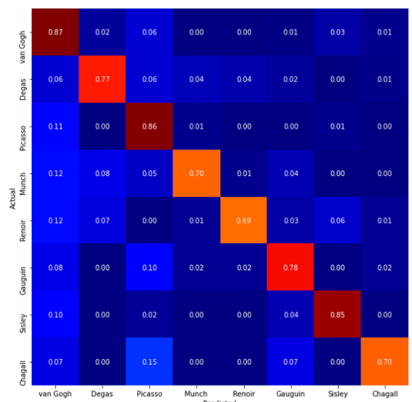

Fig. 2. All vs All: confusion matrix on the fine-tuned ResNet50. Van Gogh is recognized with the highest true positive rate (87%), while Munch and Chagall are accurately identified in 70% of the cases.

ing human characters get correctly identified as Munch, the landscape painting gets misattributed to Others (Van Gogh) with 92% probability. These similarities are also in agreement with the PHOG-based features. All the 3 features (ground self-similarity, anisotropy and complexity) are more similar among the two artists for the landscape rendition than they are for the human portrayal.

5 Conclusion

This article shows how computer vision learning-based techniques, using multi-scale pyramid of histogram of oriented gradients and pretrained residual neural network features, are able to distinguish coeval painters who adopted similar art styles. Even though many computer vision methods have been designed for

Fig. 3. Examples of correct artist attributions, as predicted by fine-tuned ResNet50.

Fig. 4. Examples of wrong artist attributions, as predicted by fine-tuned ResNet50.

art attribution, the problem has not yet converged to a fully optimized solution that, as of now, offers full credibility stand-alone. Nonetheless, classification procedures can be a valuable clue as long as they are supported by other art forensic methods.

Acknowledgments. This paper is a spin-off of the group project deliverable submitted for the DT8121 Colour Imaging course, completed in December 2020

Fig. 5. The self-portraits get correctly classified, while the landscapes get misattributed by the finetuned ResNet50 binary model. This complies with PHOG-based features. From left to right, weighted ground self-similarity, anisotropy, complexity at level 5: a) A portrait by Edward Munch: 0.85, 12.36, $0.92 \cdot 10^{-5}$; b) A portrait by Vincent Van Gogh: 0.64, 14.57, $2.38 \cdot 10^{-5}$; c) A landscape by Munch: 0.73, 11.27,$2.24 \cdot 10^{-5}$; d)A landscape by Van Gogh: 0.71, 12.91, $2.94 \cdot 10^{-5}$

at NTNU. Thus, we would like to acknowledge the other members of the group: Majid Ansari-Asl, Agnese Babini and Jan Cutajar.

References

1. Seyed Ali Amirshahi, Michael Koch, Joachim Denzler, and Christoph Redies. PHOG analysis of self-similarity in aesthetic images. In Bernice E Rogowitz, Thrasyvoulos N Pappas, and Huib de Ridder, editors, *Human Vision and Electronic Imaging XVII*, volume 8291, pages 450–459. International Society for Optics and Photonics, SPIE, 2012.
2. Rui Bai, Huang Ling, Zhao Kai, Donglian Qi, and Qiao Wang. Author recognition of fine-art paintings. In *2019 Chinese Control Conference (CCC)*, pages 8513–8518. IEEE, 2019.
3. Daniel E Berlyne. *Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation.* Hemisphere, 1974.
4. Simone Bianco, Davide Mazzini, and Raimondo Schettini. Deep multibranch neural network for painting categorization. In Sebastiano Battiato, Giovanni Gallo, Raimondo Schettini, and Filippo Stanco, editors, *Image Analysis and Processing - ICIAP 2017*, Lecture Notes in Computer Science, page 414–423. Springer International Publishing, 2017.
5. Alexander Blessing and Kai Wen. Using machine learning for identification of art paintings. *Technical report*, 2010.
6. Anna Bosch, Andrew Zisserman, and Xavier Munoz. Representing shape with a spatial pyramid kernel. In *Proceedings of the 6th ACM international conference on Image and video retrieval*, pages 401–408, 2007.
7. Anselm Brachmann and Christoph Redies. Computational and experimental approaches to visual aesthetics. *Frontiers in computational neuroscience*, 11:102, 2017.
8. Paul Craddock. *Scientific investigation of copies, fakes and forgeries.* Routledge, 2009.

9. Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. IEEE, 2005.
10. Ahmed Elgammal, Bingchen Liu, Diana Kim, Mohamed Elhoseiny, and Marian Mazzone. The shape of art history in the eyes of the machine. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.
11. Margaret Holben Ellis and C. Richard Johnson. Computational connoisseurship: Enhanced examination using automated image analysis. *Visual Resources*, 35(1–2):125–140, Apr 2019.
12. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
13. Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
14. C. R. Johnson, E. Hendriks, I. J. Berezhnoy, E. Brevdo, S. M. Hughes, I. Daubechies, J. Li, E. Postma, and J. Z. Wang. Image processing for artist identification. *IEEE Signal Processing Magazine*, 25(4):37–48, Jul 2008.
15. Jonathan Jones. Side by side, Edvard Munch and Vincent van Gogh scream the birth of expressionism. *The Guardian*, Sep 2015.
16. Kaggle. Collection of Paintings of the 50 Most Influential Artists of All Time, 2020. https://kaggle.com/ikarus777/best-artworks-of-all-time.
17. Kaggle. Explore and run machine learning code with Kaggle Notebooks — Using data from Best Artworks of All Time, 2020. https://kaggle.com/supratimhaldar/deepartist-identify-artist-from-art.
18. Daniel Keren. Painter identification using local features and naive Bayes. In *Object recognition supported by user interaction for service robots*, volume 2, pages 474–477. IEEE, 2002.
19. S Keshvari and A Chalechale. Stylometry of Painting Using Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP). *Tabriz Journal of Electrical Engineering*, 47(3):1195–1204, 2017.
20. Michael Koch, Joachim Denzler, and Christoph Redies. 1/f2 characteristics and isotropy in the fourier power spectra of visual art, cartoons, comics, mangas, and different categories of photographs. *PLoS one*, 5(8):e12268, 2010.
21. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.
22. Jia Li and James Ze Wang. Studying digital imagery of ancient paintings by mixtures of stochastic models. *IEEE Transactions on Image Processing*, 13(3):340–353, 2004.
23. Walter Liedtke, C. Richard Johnson, and Don H. Johnson. Canvas Matches in Vermeer: A Case Study in the Computer Analysis of Fabric Supports. *Metropolitan Museum Journal*, 47:101–108, Jan 2012.
24. Christoph Redies. Matlab code to calculate PHOG self-similarity, complexity and anisotropy, 2020. https://osf.io/csvta/.
25. Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.
26. Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

27. Nanne van Noord, Ella Hendriks, and Eric Postma. Toward discovery of the artist's style: Learning to recognize artists by their artworks. *IEEE Signal Processing Magazine*, 32(4):46–54, Jul 2015.
28. Nitin Viswanathan. Artist identification with convolutional neural networks. *Standford193CS231N Report*, 2017.
29. Ivo Widjaja, Wee Kheng Leow, and Fang-Cheng Wu. Identifying painters from color profiles of skin patches in painting images. In *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)*, volume 1, pages I–845. IEEE, 2003.