Neural Style Representations and the Large-Scale Classification of Artistic Style

Jeremiah W. Johnson
Department of Applied Sciences & Engineering
University of New Hampshire
Manchester, NH
jeremiah.johnson@unh.edu

***Abstract*—The artistic style of a painting can be sensed by the average observer, but algorithmically classifying the artistic style of an artwork is a difficult problem. The recently introduced neural-style algorithm uses features constructed from the low-level activations of a pretrained convolutional neural network to merge the artistic style of one image or set of images with the content of another. This paper investigates the effectiveness of various representations based on the neural style algorithm for use in algorithmically classifying the artistic style of paintings. This approach is compared with other neural network based approaches to artistic style classification. Results that are competitive with other recent work on this challenging problem are obtained.**

Keywords*—*Artificial intelligence; neural network; style transfer; deep learning; computer vision; machine learning

I. Introduction

Any observer can sense the artistic style of painting, even if it takes training to articulate it. To an art historian, the artistic style is the primary means of classifying the painting [1]. However, artistic style is not a well defined concept. Artistic style can be loosely defined as "…a distinctive manner which permits the grouping of works into related categories" [2]. Algorithmically classifying the artistic style of an artwork is a challenging problem which may include analysis of features such as the painting's color, its texture, and its subject matter, or none of those at all. Detecting the style of a digitized image of a painting poses additional challenges raised by the digitization process, which itself has consequences that may affect the ability of a machine to correctly detect artistic style. For instance, textures may be affected by the resolution of the digitization. Despite these challenges, intelligent systems for detecting artistic style would be useful for identification and retrieval of images of a similar style.

In this paper, the neural-style algorithm introduced in [3] is adapted to obtain neural-style representations of digitized images of artwork for large-scale artistic style classification. The results obtained are compared with other recent approaches to large-scale artistic style classification, showing competitive performance.

II. Related Work

Algorithmic determination of artistic style in paintings has only been considered sporadically in the past. Examples of early efforts at style classification are [4] and [5], where the datasets used are quite small, and only a handful of very distinct artistic style categories considered. Several complex models are constructed in [6] by hand-engineering features on a large dataset similar to the one used for this work. And in [7], it is demonstrated that convolutional neural networks are effective models for detecting image style in general, including artistic style in paintings, via transfer learning. In [6] and [7], the number of artistic style categories is held to 25 and 27 respectively.

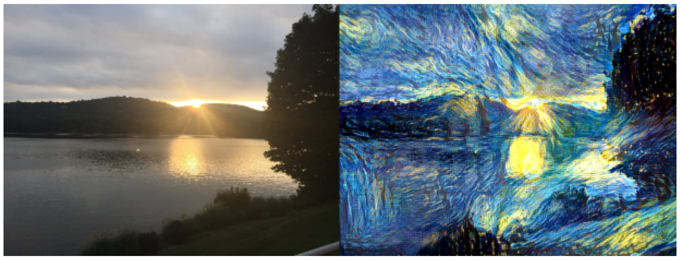

Fig. 1. Original image on the left, after application of the 'neural-style' algorithm (style image 'Starry Night', by Van Gogh) on the right.

In the paper "A Neural Algorithm of Artistic Style", it is demonstrated that the correlations between the low-level feature activations in a deep convolutional neural network encode sufficient information about the style of the input image to permit a transfer of the visual style of the input image onto a new image via an algorithm informally referred to as the "neural-style" algorithm [3]. An example of the output of this algorithm is presented in Fig. I. Several authors have built upon the work of Gatys et. al. in the past year [8], [9], [10]. These investigations have primarily focused on ways to improve either the quality of the style transfer or the efficiency of the algorithm. To the best of our knowledge the only other look at the use of the style representation of an image as a classifier is in [11], where the authors take an approach similar to that taken here, but with a much smaller dataset and a much smaller set of style categories, and without comparison to current deep neural network based approaches.

III. Data and Methods

A. Data

The data used for this investigation consists of 76449 digitized images of fine art paintings. The vast majority of the images were originally obtained from http://www.wikiart.org, the largest online repository of fine-art paintings. For conve-

TABLE I. BASELINE RESULTS

Model	Accuracy (top 1%)
Convolutional Neural Network	27.47
Pretrained Residual Neural Network	36.99

nience, a prepackaged set of images sourced and prepared by Kiri Nichols and hosted by the data-science competition website http://www.kaggle.com was used for the experiments documented in this paper. A stratified 10% of the dataset was held out for validation purposes. A more fine-grained set of style categories for classification than has been used in previous work on image style was chosen, as finer classification is likely necessary for practical application. Here 70 distinct style categories are used, the maximum amount possible with the current dataset while maintaining at least 100 observations of each style category. This noticeably increases the complexity of the classification task as many of the class boundaries are not well-defined, the classes are unbalanced, and there are not nearly as many examples of each of the artistic styles as in previous attempts at large-scale artistic style classification.

B. The Neural Style Algorithm

The primary insight in the neural-style algorithm outlined by Gatys et. al. is that the correlations between low-level feature activations in a convolutional neural network capture information about the style of the image, while higher-level feature activations capture information about the content of the image. Thus, to construct an image $\mathbf{x}$ that merges both the style of an image $\mathbf{a}$ and the content of an image $\mathbf{p}$, an image is initialized as white noise and the following two loss functions are simultaneously minimized:

$$\mathcal{L}_{content}(\mathbf{p},\mathbf{x}) = \sum_{l\in L_{content}} \frac{1}{N_l M_l} \sum_{i,j} \left(F^l_{ij} - P^l_{ij}\right)^2, \qquad (1)$$

and

$$\mathcal{L}_{style}(\mathbf{a},\mathbf{x}) = \sum_{l\in L_{style}} \frac{1}{N_l^2 M_l^2} \sum_{i,j} (G^l_{ij} - A^l_{ij})^2, \qquad (2)$$

where N_l is the number of filters in the layer, M_l is the spatial dimensionality of the feature map, $\mathbf{F}^l$ and $\mathbf{P}^l$ represent the feature maps extracted by the network at layer l from the images $\mathbf{x}$ and $\mathbf{p}$ respectively, and letting $\mathbf{S}^l$ represent the feature maps extracted by the network at layer l from the image $\mathbf{a}$, $G^l_{ij} = \sum_{k=1}^{M_l} F^l_{ik} F^l_{jk}$ and $A^l_{ij} = \sum_{k=1}^{M_l} S^l_{ik} S^l_{jk}$. That is, the style loss, which encodes the images style, is a loss taken over Gram matrices for filter activations.

C. Style Classification

To establish a baseline for style classification, a single convolutional neural network was first trained from scratch. The network has a uniform structure consisting of convolutional layers with 3x3 kernels and leaky ReLUs activations ($\alpha = 0.333$). Between every pair of convolutional layers is a fractional max pooling layer with a 3x3 kernel. Fractional max-pooling is used as given the relatively small size of the dataset, the more commonly used average or max-pooling operations would lead to rapid data loss and a relatively shallow network [12]. The convolutional layer sizes are $3 \to 32 \to 96 \to 128 \to 160 \to 192 \to 224$, followed by a fully-connected layer and 70-way softmax. 10% dropout is applied to the fully connected layer. Aside from mean normalization and horizontal flips, the data were not augmented in any way. The model was trained over 55 epochs using stochastic gradient descent and achieved a top 1% accuracy of 27.468%.

We then finetuned a pretrained object classification model for style classification. The model used was a residual neural network with 50 layers pretrained on the ImageNet 2015 dataset. There are two motivating factors for choosing to finetune this network. The first is that residual networks currently exhibit the best results on object recognition tasks, and previous work on style classification suggests that a network trained for the task of object recognition and then finetuned for image style detection will perform the task well [13], [7]. The second and more interesting reason from the standpoint of artistic style classification is that the architecture of a residual neural network makes the outputs of lower levels of the network available unadulterated to higher levels in the network. In this way, residual networks have been noted to function similar to a Long Short-Term Memory network without gates [14]. For style classification, this is particularly appealing as a means of allowing the higher levels in the net to consider both lower-level features and higher-level features when forming an artistic style classification, where the style may very much be determined by the lower-level features. The residual neural network model obtained top-1% accuracy of 36.985%. Baseline results are summarized in Table I.

To determine whether or not the style representation encoded in the Gram matrices for a given image has any power as a classifier, the feature activations at layers ReLU1_1, ReLU2_1, ReLU3_1, ReLU4_1, and ReLU5_1 were extracted from the nineteen-layer convolutional neural network developed by the Visual Geometry Group at the University of Oxford, the so-called VGG-19 model, for the paintings described above [15], and the Gram matrices of these activations were calculated. The model and layers used were chosen based on the quality of the style transfers obtained by [3] using this network and layers. The weights for the pretrained VGG-19 model was obtained from the Caffe Model Zoo [16]. The Gram matrices were then reshaped to account for symmetry, producing a total of 304,416 distinct features per image. This is nearly a factor of four greater that the total number of observations in the dataset.

Analyzing the style representation was approached in two ways. First, the full feature vector was normalized and then passed to a single-layer linear classifier which was trained using Adam over 55 epochs, producing a top 1% accuracy of 13.23% [17].

Random forest classifiers were then built on the individual Gram matrices extracted from the activations of the network. The dimensionality of the Gram matrices post-reshaping is 2016, 8128, 32640, 130816, and 130816 respectively. Considered separately, the random forest classifiers built on the first three of these style representations performed better than the linear classifier based on the full style representation and better than the baseline convolutional neural network, with top-

TABLE II. STYLE REPRESENTATION RESULTS

Model	Accuracy (top 1%)
Full Style Representation - Linear Classifier	13.21
ReLU1_1 Random Forest	27.84
ReLU2_1 Random Forest	28.97
ReLU3_1 Random Forest	33.46
ReLU4_1 Random Forest	9.79
ReLU5_1 Random forest	10.18

1% accuracies of 27.84%, 28.97%, and 33.46%. The random forests built on the latter two layers performed considerably worse. The results are presented in Table II.

In contrast to results reported in [11], we observed a significant loss in accuracy when dimensionality reduction was even lightly utilized on the Gram matrix representations extracted individually. For instance, performing PCA while preserving 90% of the variance in the data from the layer ReLU1_1 style representation reduced the accuracy of the random forest model on that layer from 27.84% to 17%, perhaps due to the use of a larger, less balanced, and less homogeneous dataset. No significant gains were observed when the data were normalized.

IV. CONCLUSION & FUTURE WORK

The neural-style representation of an artwork offers competitive performance as an artistic style classifier, with top 1% accuracy comparable to results presented in [6]. Nevertheless, these experiments demonstrate that a modern deep neural network, when pretrained for a vision task and finetuned for artistic style classification, obtains superior results. The best results obtained using the neural-style representation of an artwork were obtained when models suitable for high-dimensional nonlinear data were constructed individually on the first three Gram matrices that form the building blocks of the neural-style representation.

Despite the aesthetically pleasing results that can be obtained using the neural-style algorithm for style transfer, it appears that the various neural-style representations described in this paper do not fully encode the art-historical definition of artistic style. However, it is clear that this information is relevant and has some predictive ability. Understanding the strengths of this method and improving on these results is a goal of future work.

ACKNOWLEDGMENTS

The author would like to thank NVIDIA for GPU donation to support this research, Wikiart.org for providing many of the images, the website Kaggle.com for hosting the data, and Kiri Nichols for sourcing the data.

REFERENCES

[1] B. Lang, *The Concept of Style*. Cornell University Press, 1987.

[2] E. Fernie, *Art History and its Methods: A Critical Anthology*. London: Phaidon, 1995.

[3] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2414–2423.

[4] D. Keren, "Recognizing image style and activities in video using local features and naive bayes," *Pattern Recogn. Lett.*, vol. 24, no. 16, pp. 2913–2922, Dec 2003. [Online]. Available: http://dx.doi.org/10.1016/S0167-8655(03)00152-1

[5] L. Shamir, T. Macura, N. Orlov, D. M. Eckley, and I. G. Goldberg, "Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art," *ACM Trans. Appl. Percept.*, vol. 7, no. 2, pp. 8:1–8:17, feb 2010. [Online]. Available: http://doi.acm.org/10.1145/1670671.1670672

[6] B. Saleh and A. Elgammal, "Large-scale classification of fine-art paintings: Learning the right metric on the right feature," in *International Conference on Data Mining Workshops*. IEEE, 2015.

[7] S. Karayev, A. Hertzmann, H. Winnemoeller, A. Agarwala, and T. Darrell, "Recognizing image style," *CoRR*, vol. abs/1311.3715, 2013. [Online]. Available: http://arxiv.org/abs/1311.3715

[8] M. Ruder, A. Dosovitskiy, and T. Brox, "Artistic style transfer for videos," *arXiv preprint arXiv:1604.08610*, 2016.

[9] R. Novak and Y. Nikulin, "Improving the neural algorithm of artistic style," *CoRR*, vol. abs/1605.04603, 2016. [Online]. Available: http://arxiv.org/abs/1605.04603

[10] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*, 2016.

[11] S. Matsuo and K. Yanai, "Cnn-based style vector for style image retrieval," in *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, ser. ICMR '16. New York, NY, USA: ACM, 2016, pp. 309–312. [Online]. Available: http://doi.acm.org/10.1145/2911996.2912057

[12] B. Graham, "Fractional max-pooling," *CoRR*, vol. abs/1412.6071, 2014. [Online]. Available: http://arxiv.org/abs/1412.6071

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[14] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Training very deep networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 2377–2385.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: http://arxiv.org/abs/1409.1556

[16] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

[17] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, 2014.