NEURAL STYLE REPRESENTATIONS OF FINE ART

Jeremiah Johnson
Department of Applied Engineering and Sciences
University of New Hampshire
Manchester, NH 03101 USA
{jeremiah.johnson}@unh.edu

ABSTRACT

The artistic style of a painting is a subtle aesthetic judgment used by art historians for grouping and classifying artwork. The neural style algorithm introduced by Gatys et al. (2016) substantially succeeds in image style transfer, the task of merging the style of one image with the content of another. This work investigates the effectiveness of a style representation derived from the neural style algorithm for classifying paintings according to their artistic style.

1 INTRODUCTION

A casual observer can sense the artistic style of painting, even if it takes formal training to articulate it. Artistic style is the primary means of classifying a painting Lang (1987). However, artistic style is not well defined; an authoritative treatment only goes so far as to describe it as "... a distinctive manner which permits the grouping of works into related categories" Fernie (1995). Thus, algorithmically determining 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 perhaps none of those at all. Detecting the artistic style of a digitized image of a painting poses additional challenges raised by the digitization process, which itself has consequences that may adversely affect the ability of a machine to correctly classify the artwork; for instance, textures may be affected by the resolution of the digitization (Polatkan et al. (2009)).

Convolutional neural networks have shown to be very effective at large-scale classification tasks such as image recognition and object detection (Krizhevsky et al. (2012), Simonyan & Zisserman (2014), He et al. (2015)). In this paper we adapt the neural-style algorithm introduced in Gatys et al. (2016) for image style tranfer to produce a 'neural style' representation of a digitized image of an artwork that incorporates only low-level information from a convolutional neural network, with the aim of focusing the representation on the stylistic aspects of the artwork rather than on the artwork's content. This representation is then used as input into a classification algorithm to determine the artistic style of the artwork. We apply classification via style representation to a database of approximately 75000 digitized images of paintings across 70 distinct artistic style categories, and obtain promising classification results.

Figure 1: Content image, left, style image 'Starry Night', by Van Gogh, center, new image generated by the 'neural-style' algorithm, right.

1.1 THE NEURAL STYLE ALGORITHM

The key insight of Gatys et al. (2016) is that style and content in an image can be separated to a certain extent. The correlations between low-level activations in a convolutional neural network effectively capture information about the stylistic features of the image such as coloration and texture, while the higher-level activations capture information about the image's content. Thus, one may construct an image $\mathbf{x}$ that merges both the style of an image $\mathbf{a}$ and the content of an image $\mathbf{p}$ by initializing an image as white noise and then simultaneously minimizing the following two loss functions:

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

and

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

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 activations at layer l from the images $\mathbf{x}$ and $\mathbf{p}$ respectively, and letting $\mathbf{S}^l$ represent the activations at layer l from the image $\mathbf{a}$, $G_{ij}^l = \sum_{k=1}^{M_l} F_{ik}^l F_{jk}^l$ and $A_{ij}^l = \sum_{k=1}^{M_l} S_{ik}^l S_{jk}^l$. The loss function in equation 2 captures the difference in style between the target image and the style image; in this construction, the Gram matrices of the activations of the style image and the target image can be seen as encoding their respective styles.

1.2 THE NEURAL STYLE REPRESENTATION

The neural style representation of an image is obtained by passing the image through a pre-trained convolutional neural network and extracting the activations at various early layers in the network. The representation obtained will depend not only on the image, but also on the network used and the layers at which the activations are extracted. Gatys et al. (2016) demonstrated successful image style transfer using the VGG-19 network (Simonyan & Zisserman (2014)) and activations at layers ReLU$*$_1, for $* = 1,\ldots,5$, so we follow course here and extract the same activations. We then calculate the Gram matrices of the layerwise activations and reshape them into vectors, taking advantage of the symmetry of the Gram matrix to resize them to length $n(n-1)/2$ for a layer with n activations. The size of the representation produced is quite large; for instance, layers ReLU4_1 and ReLU5_1 each have 512 activations, resulting in a representation from each of those layers of size $(512 \times 511)/2 = 130,816$.

2 ARTISTIC STYLE CLASSIFICATION

The data used for this investigation consists of 76449 digitized images of fine art paintings, each of which falls into one of 70 artistic style categories. Images and categories were selected so that no artistic style category contained less than 100 samples, and a stratified 10% of the data was held out for validation purposes. For convenience we utilize a set of images sourced and prepared by Kiri Nichols for a data-science competition hosted by the website `http://www.kaggle.com`. The vast majority of the images were originally obtained from `http://www.wikiart.org`.

We consider two approaches to artistic style classification using the representation obtained in section 1.2. The first is a fully-connected linear classifier trained online over 55 epochs using the Adam algorithm, yielding a top-1% accuracy of 13.23% (Kingma & Ba (2014)).

The second approach to artistic style classification is to take the layerwise representations and use them individually as input to a classification algorithm. Given the high dimensionality of the full representation, this is a more flexible approach, and yields significantly higher performance when the representations are input to a minimally tuned random forest classifier (Breiman (2001)). Results are summarized in Table 1. Note that the higher-dimensional layers lead to significantly weaker representations, suggesting that the full representation might be improved by omitting the last two layers.

Table 1: Style Representation Results

Model	Accuracy (top 1%)
Full Style Representation	13.21
ReLU1_1 Representation	27.84
ReLU2_1 Representation	28.97
ReLU3_1 Representation	33.46
ReLU4_1 Representation	9.79
ReLU5_1 Representation	10.18

3 Conclusions and Future Work

The fuzzy concept of artistic style in paintings is still quite difficult to capture algorithmically. The neural style representation discussed here incorporates only information produced at the early layers in a deep convolutional network, and quickly reaches very high dimensionality. This limits the representation to containing only low-level information such as texture and color. Improving on artistic style classification likely requires the incorporation of higher-level content and context-specific information. In related experiments, we have seen that a residual neural network (He et al. (2015)) pre-trained for object detection on the Imagenet database can be finetuned on this dataset for artistic style classification to produce a top-1% accuracy of 36.99%. We believe that the increase in accuracy is due not only to the valuable pre-training of the weights, but also to the fact that the residual network makes some valuable low-level information available to the higher layers in the network unadulterated (Srivastava et al. (2015)). A possible avenue for future improvement is to incorporate the Gram matrix calculation into the connections between layers in the residual neural network.

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

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324. URL http://dx.doi.org/10.1023/A:1010933404324.

Eric Fernie. *Art History and its Methods: A Critical Anthology*. London: Phaidon, 1995.

Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2414–2423, 2016.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. URL http://arxiv.org/abs/1512.03385.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, 2014.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*, pp. 1097–1105, 2012.

Berel Lang. *The Concept of Style*. Cornell University Press, 1987. ISBN 9780801494390.

G. Polatkan, S. Jafarpour, A. Brasoveanu, S. Hughes, and I. Daubechies. Detection of forgery in paintings using supervised learning. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pp. 2921–2924, Nov 2009. doi: 10.1109/ICIP.2009.5413338.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014. URL `http://arxiv.org/abs/1409.1556`.

Rupesh K Srivastava, Klaus Greff, and Jürgen Schmidhuber. Training very deep networks. In *Advances in Neural Information Processing Systems*, pp. 2377–2385, 2015.