9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Modeling Culture with AI: Understanding Photographic Style with Deep Learning

Keywords: Deep Learning, Computer Vision, Artistic Style, Culture, Photography

Extended Abstract

Introduction

In recent years, deep neural nets have become extremely useful in artificial intelligence and machine learning applications due to technical and algorithmic innovations that have allowed researchers to achieve greater accuracy in more complicated tasks such as vision detection and speech recognition. One of these areas has been in advancements in computer vision and artistic style. This paper contributes to the deep learning literature on artistic style by utilizing a unique dataset of art photographers and their work to better understand how deep neural nets can classify and learn based on artistic style.

To the best of our knowledge, this is the first paper and project in the deep learning literature that examines style in artistic and fine art photography, using a dataset of over 20,000 photographs by art photographers whose work is collected by major museums and studied by students in fine art institutions. Utilizing an original dataset assembled, collected, and cleaned by the research team, we are able to extend and connect some of the previous research done on artist style and deep learning to the medium of artistic photography.

Specifically, we use a ResNet-18 model [1] with transfer learning over a training set consisting of images by various art photographers. Using a softmax classifier, we then predict on a separate test set, the actual artist based on our model's learning. We run various experiments to first build the most accurate classification model, and then we use this version of our model on five variations of our dataset, each containing a different number of artists. From these results, we examine both the accuracy scores and the corresponding confusion matrix to better understand how our model is predicting, and we interpret these findings in the discussion and conclusion.

This project offers the following contributions: we show that in datasets containing a few artists (6) and a sufficient number of training samples (over 200), our model can achieve test accuracy rates above 90%. Even when the model is expanded to our full dataset of 511 artists and with limited training samples, we achieve an accuracy rate of 44% which is close to an almost even 50/50 chance of being correct despite there being over 500 artists to choose from. In examining the corresponding confusion matrix for each variation, we see that classification errors occur more frequently among artists with similar style rather than over content, and we also observe a few surprises in what the models are able to accurately predict and what they do less well on, verifying some assumptions about how our models are learning.

Methods

The network architecture used is based on a ResNet-18 architecture that starts with pre-trained weights from ImageNet [2]. The final fully-connected layer is replaced with a new layer to calculate a score for each artist in our dataset instead of a score for ImageNet classes. We use a softmax classifier with cross-entropy loss: $L_i = -log(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}})$. L_i is the loss for example i in the training minibatch (size 32), f is the score for a particular class calculated by the network,

j is one of the possible classes, which depends on the number of artists in the variation of the dataset we were using. For training, we first held the weights of our pre-trained network constant for 10 epochs, and then allowed all the weights throughout the entire network to update, training for an additional 10 epochs. Accuracy charts were created after the first 10 epochs, and a second accuracy chart was made after the additional 10 epochs to show the effects of fine tuning, and our final accuracy scores reflect this version of our model architecture. Confusion matrices were created from the accuracy results and were used to interpret how style was being applied by the models. (See **Figure 4**).

Results and Discussion

Quantitatively, accuracy increases as the number of artists decreases, and at very low numbers, there is very high accuracy, at over 90%. Despite best efforts to address overfitting with various regularization techniques, our results with the 6-artist sample give us confidence that increasing the number of training examples to over 250 in future studies should help address some of the overfit issues.

Qualitatively, this is where our results are much more informative. In a thorough examination of the various confusion matrix outputs, a few trends emerged that highlighted the power of neural nets in learning artistic style. The three main findings include:

1. Certain photographers whose work would be categorized by human experts as having a distinct and strong style, enjoyed very high accuracy rates despite the increase in the number of artists in the dataset. (See **Figure 1**).

2. When more artists were compared against each other, the errors made by the model often happened in cases where style was similar, not just because the artists had the same content. (See **Figure 2**).

3. There were unique cases where a human expert would have no trouble in classifying a photograph but the model struggled more due to the lack of stylistic information in the image. (See **Figure 3**).

In sum, a qualitative analysis shows that deep neural nets utilize a high degree of understanding style in their models. Artists who are well known for having strong distinctive artistic styles have high rates of accuracy regardless of the number of other artists they are being compared against. Some artists who overlap heavily within sub-genres of style and would be curated by art experts as being stylistically very similar, are often confused by our models, especially when there are more artists. And finally, artists who are known for mimicking the style of other artists confuse the neural nets, despite cues that human experts may perceive (such as the same person appearing in all of the images).

References

[1] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
[2] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M. and Berg, A.C., 2015. Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 115(3), pp.211-252.

Figures

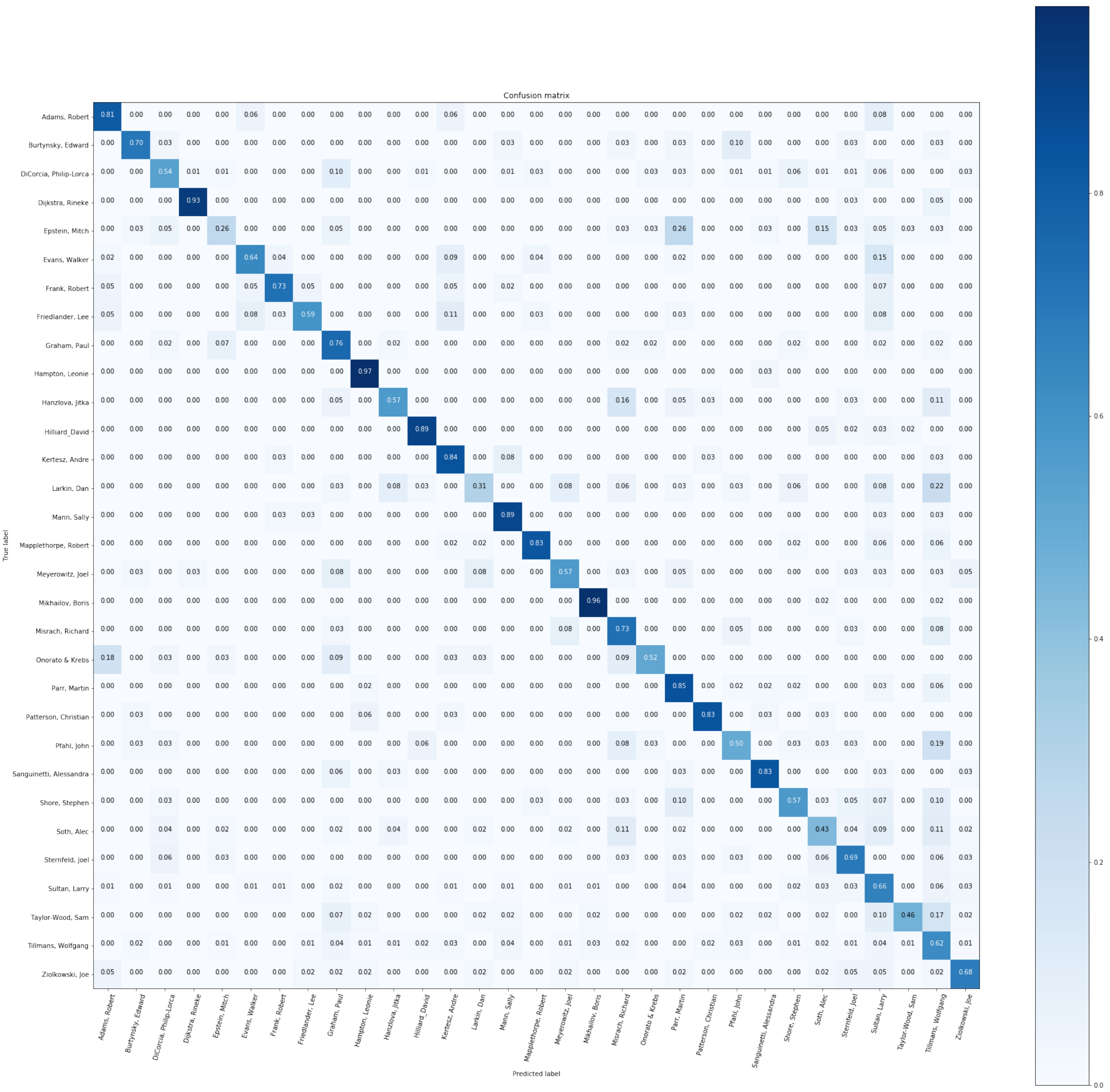
The following sets of figures show examples from the corresponding results and discussion section:

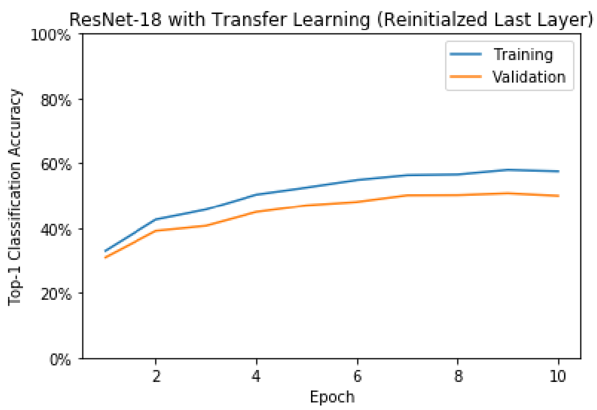
Figure 1: Three different photographs above by the Dutch artist, Rineke Dijkstra, whose work is renown in the art world for having a very distinct and formal style. CNNs were able to predict her work with extremely high accuracy in all the various experiments done classifying her work. For example, the 31-artist model predicted her work at a 93% accuracy rate, while the 110-artist model predicted her work at an 88% rate.

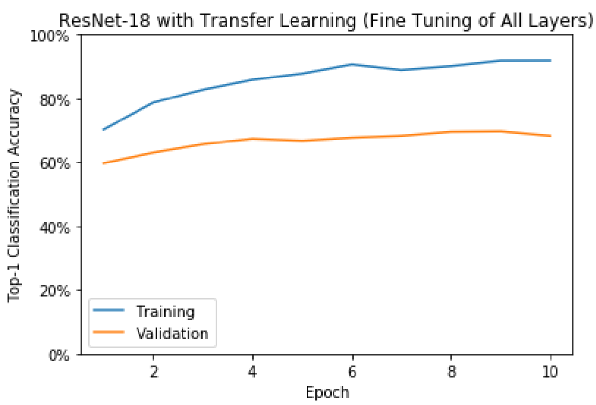
Figure 2: The above left photograph is by the artist Sam Taylor-Wood while the above right image is by Philip-Lorca DiCorcia. When there are fewer artists in the dataset, the neural nets do a relatively good job at predicting each artist individually. However, as the dataset grows, the model begins to confuse these two artists more (0.24 normalized error in the 110 artist dataset). This is noteworthy because both artists come from the same genre of photography called "Documentary Fiction," a style known for expressive moods and emotion, using carefully constructed lighting, actors, and staged imagery to construct a fictionalized image rather than one found in regular documentary photography, and the two artists are often grouped together in stylistic genre by museum curators and other experts.

Figure 3: The above 6 images are all from the artist Cindy Sherman, who mimics the aesthetics and style of various artists and photography genres. However, she is the subject of all her photographs which makes it easy for a museum curator to spot her images (as long as they recognize her and know about her work). Our model in the 110-artist experiment only predicted her work at a 50% accuracy rate.

Figure 4. Confusion Matrix and Accuracy Charts from 31 Artist Experiment These figures are examples of one of the experiments done with the dataset and show how darker shades within the confusion matrix indicate where the model confused one artist with another.







These two accuracy charts represent a total of 20 epochs. The left chart represents the first 10 epochs with just the pre-trained weights from ImageNet, while the chart on the right represents an additional 10 epochs where we fine tuned the model by allowing all the weights to update.