DeepArt: Identify Artist from Painting

Kai Qu, Daniel Xiao, Haoran Zhu
Department of Computer Science, University of Virginia, Charlottesville, VA 22904
[kq4ff, zx8yz, hz3fr]@virginia.edu

Abstract

Artists have different styles that are unique to themselves. While some styles are much different from each other, others are subtle and hard to tell apart. We are curious whether computers can identify the artists based on particular paintings they created. In this project, we adapt and fine-tune different pretrained models (ResNet50, GoogLeNet, VGGNet11) to a dataset that contains artworks of 50 the most influential artists of all time and compare the results to the baseline model (ResNet18). Furthermore, we trained another identical ResNet50 Model (call it ResNet50_2). However, during the training of ResNet50_2, we freeze deeper layers to only train shallow layers to capture the overall styles of different paintings, comparing this result to the one merely by fine-tuning approach. Our ResNet 50_2 model outperforms all other pretrained models by at least 1.5% *in all metrics, which validates our speculation that shallow layers play a more important role in understanding the overall style of paintings than deep layers.*

1. Introduction

Identifying the artist based on his or her artistic works is a challenging task for computers. Different from most classification tasks (for example, identify objects on images in the COCO dataset) we encountered, the goal here is not to identify the objects in a painting but to understand the overall style of the painting and establish the connection between style and artist. Moreover, since the classification problem in this project is in a different category from the problems we did in assignments, we believe that this project will give us opportunities of exploring areas that are outside of coursework and deepening our understanding of the application of machine learning in computer vision.

Our dataset is obtained from a publicly available task on Kaggle, where the creator of the dataset scrapes images of artworks from Art Challenge[1]. We apply several models including ResNet, GoogleNet, and VGGNet that are pretratined on ImageNet 1000 classes. Moreover, we employ another ResNet50 model, where we mainly focus on training shallow layers rather than deep layers. Eventually, we compare the performance of different models. From the observation, we come up with hypothesis that relates the performance of models to the number of layers and the nature of architecture. Also, we conclude that shallow layers in deep learning models are mainly responsible for understanding the overall style of paintings.

[1] http://artchallenge.ru

2. Related Work

Early work: There have been research focusing on classification based on the overall "style" of objects. For example, [2] designed a model that was used to identify fish species based on features such as color and texture. However, most of these work were done before proposals of modern deep network architectures such as GoogleNet and ResNet and only used feedforward networks in their classification [5, 6]. Our work employed some of the most popular modern architectures in computer vision and investigated on the roles and functions of shallow and deep layers in the architecture.

Residual Network: Before the occurrence of residual network [1], deep neural networks usually suffer from gradient vanishing problem. As network depth increases, accuracy gets saturated and then degrades rapidly. ResNet solved this problem by introducing skip connection. By stacking identity mappings upon the current neural network, the skip connections between layers add the outputs from previous layers to the outputs of stacked layers. Therefore, the gradients can flow throughout the entire network and will not disappeared in the middle of gradient descend.

VGGNet: VGGNet [3] is based on the idea that with a given receptive field, multiple stacked smaller size kernels are better than one large-size kernel. Therefore, VGGNet is composed of multiple modules of 3x3 kernel-sized filters one after another. With this structure, the depth of the network is vastly increased, while the number of parameters is reduced significantly and the receptive field is kept the same.

GoogLeNet: Consisted of 22 layers of deep convolu-

Method	F1(%)	Precision(%)	Recall(%)	Validation Accuracy(%)
ResNet18	63.34	66.37	66.44	66.45
ResNet50	72.30	73.18	73.19	73.20
VGGNet11	49.51	54	50.80	50.80
GoogLeNet	76.21	76.64	76.53	76.54
ResNet50_2	**77.84**	**78.55**	**78.02**	**79.62**

Table 1. Training results of different models.

tional neural network, GoogLeNet [7] also utilizes the idea of deep neural network to improve its performance. The network uses a CNN inspired by LeNet [4] but implemented a novel element called inception module. This module is based on several very small convolutions to reduce the number of parameters.

3. Methodology

We first split the raw dataset into training set and validation set by a ratio of 4:1. An 180-degree rotated copy of every painting was then added into the dataset. After data augmentation, the dataset contains 16896 paintings in total. Paintings in both sets are grouped under different subfolders based on their creators and labelled properly. Image pixels were normalized and all images were cropped to a 224x224 resolution by performing random crops.

Models we used in our work were fine-tuned ResNet18, VGGNet11, ResNet50, GoogleNet, and ResNet50_2. All of them were pretrained in the ILSVRC task on ImageNet. We trained all models using stochastic gradient descent with momentum 0.9 and a mini-batch size of 64. All models except ResNet50_2 were trained for 60 epochs. Regarding ResNet50_2, we first trained the full model for 10 epochs. Then we froze all layers other than the first 50 layers and trained the model for another 50 epochs. The learning rate was set to 0.001 for all training. Cross-entropy loss was used in training to compute loss.

4. Results

From Table.1 and Figure.1 and 2, despite having the same architecture, ResNet50 achieves much better results than ResNet18. A reasonable explanation would be that ResNet50 has more layers. Moreover, the nature of different architecture also seems to exert great influence on the results. While GoogLeNet (22 layers) has much fewer layers than ResNet50, the former manages to achieve higher accuracy. However, further research is needed to explain how the nature of these architectures influences models' prediction accuracy.

Among all models, ResNet50_2 achieves the highest validation accuracy. However, with the same number of epochs, the accuracy from training full ResNet50 model is lower than the one from GoogLeNet. Since the only difference between ResNet50 and ResNet50_2 is the training procedure, we conclude that shallow layers in the model play more important roles in understanding the overall style of paintings than deep layers. Therefore, we hypothesize that shallow layers in these models are mainly responsible for understanding the overall style of an image while deep layers are responsible for detecting objects.

Eventually, from Figure.1, we can see that the training and validation accuracy of ResNet50_2 are remains basically stable at 88% and 77%, respectively. This somewhat low accuracy is expected as an artist might have paintings in different styles, which creates overlap of styles among different paintings. Thus, even if a model successfully identifies the style of a painting, there might be multiple artists who employ the same style. We expect the accuracy to rise if we try to match paintings to styles. While we haven't implemented this idea yet, we do have clear directions about how to do this. What's more, the original dataset is imbalanced. For example, Vincent van Gogh has 877 artworks in the dataset while Jackson Pollock only has 24 paintings. Such imbalance certainly increases the error but it is an inevitable trade-off for having a reasonable size of dataset.

5. Conclusion & Future Work

In this project, we investigate the performance of different pretrained models on the task of identifying the creators of artworks. By designing a training procedure that focuses on training shallow layers, our ResNet50_2 model outperforms all other pretrained models by at least 1.5% in all metrics. We then conclude that shallow layers in a deep learning model plays more important roles in understanding the overall styles of paintings. We suspect that the low prediction accuracy on validation set is due to the imbalance of dataset and the overlap of styles among different artists, which bewilders models in the process of learning. For the future work, we will alleviate the imbalance by scraping more paintings for artists whose artworks take relatively small portions in the original dataset. Moreover, we will design style-matching models that match paintings to artistic styles and compare the performance with current models. We also plan to use style-matching model as a cross-checking tool in identifying creator of paintings to improve the accuracy of our current models.

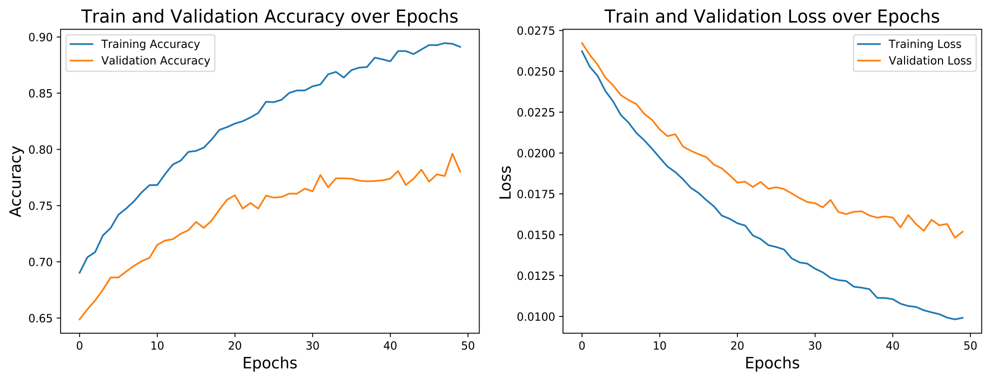

Figure 1. Loss and Accuracy of ResNet50_2 in Shallow-Layer Training

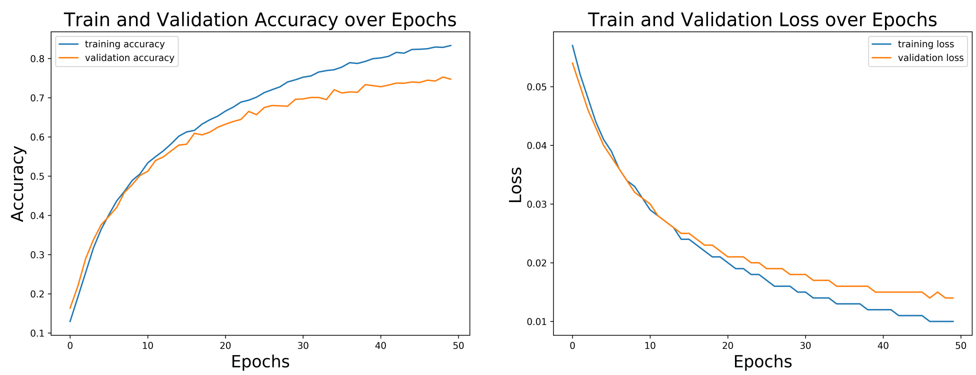

Figure 2. Loss and Accuracy of ResNet50

References

[1] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.

[2] J. Hu, D. Li, Q. Duan, Y. Han, G. Chen, and X. Si. Fish species classification by color, texture and multi-class support vector machine using computer vision. 88:133–140, Oct 2012.

[3] A. Z. Karen Simonyan. Very deep convolutional networks for large-scale image recognition, 2014.

[4] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, Nov 1998.

[5] F. Pedreschi, D. Mery, A. Bunger, and V. Yanez. Computer vision classification of potato chips by color. In *Journal of Food Process Engineering*, volume 34, pages 1714–1728, Sep 2011.

[6] O. Söderkvist. Computer vision classification of leaves from swedish trees. 2001.

[7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions, 2014.