Classify Fine-Art Painting Style with Convolution Neural Network

Xiaoxue Ma
New York University
xm576@nyu.edu

Yicheng Pu
New York University
yp653@nyu.edu

Lige Wang
New York University
lw2427@nyu.edu

1 Introduction

As art museums and fine arts collections become increasingly digitized, the automated tool to classify the art style of a painting is of great necessity and importance for both museum curators and visitors. The visual style, however, is not rigorously and universally defined. Besides, recognizing the painting styles requires high-level perception and training about artistic-related knowledges. Instead of depending on the abilities of human, it is worth to build an art classifier to automate the style labeling process. In this project, we dealt with 30 fine art styles to try to cover more kinds of features, such as: objects, color, lines, shapes, brush stroke, texture, etc. Instead of using hand-crafted low-level descriptors, we turned to apply a range of deep learning algorithms for this classification problem.

2 Literature Review

For image classification, many deep learning models achieved great performance in these years. Inception[8] model is one of the most successful deep learning model in this field. It was designed for Imagenet Challenge, in which models are trained to classify images belong to 1000 kinds of categories. In 2015 Imagenet Challenge, an ensemble model of four Inception models was ranked at the second place. Among many deep learning models for image classification, Inception V3 is not only noted for high performance, but also low computational cost. It only has 25 million parameters while Alexnet[3] has 60 million and VGGNet[7] employs around 240 million parameters. Inception Model saved many parameters by replace a $N \times N$ filter with a stacking of $1 \times N$ and $N \times 1$ filters. The results show that a large, deep convolutional neural network is capable of achieving record-breaking results on a highly challenging dataset using purely supervised learning.

However, for this specific task of visual art style classification, current research do not use pure neural network models, but a combination of neural networks and other machine learning approaches.

Bar, Levy and Wolf[1] developed automatic style classification algorithm using features extracted from low-level descriptors, PiCoDes[2], and deep network. Low-level descriptors cover Edge texture information, color histogram, Local binary patterns, etc. The conclusion is that deep features outperform the hand-crafted low-level descriptors in distinguishing styles.

Saleh and Elgammal[6] used a methodology called feature-fusion, which extracts different kinds of features from a image and concatenate them together as the final feature vector. After the feature-fusion, they used Metric Learning, which learns the correct metric to evaluate the similarity between two objects, followed by metric-fusion, which uses multiple metrics to get a final feature vector. At the very end, they used a SVM to classify styles, genres and artists.

Lu et al[5] designed a deep multi-patch aggregation network allowing the model to be trained by multiple patches generated randomly from on image. The aggregation of those patches is supported by two novel network layer, which is statistics layer and sorting layer. Comparing with traditional ConvNet, they obtained a increase of 13% in accuracy.

The Inception V3 model achieved top performance in Imagenet Challenge, in which models are trained to classify images belong to 1000 kinds of categories. In 2015 Imagenet Challenge, an ensemble model of four Inception models was ranked at the second place. Among many deep learning models for image classification, Inception V3 is not only noted for high performance, but also low computational cost. It only has 25 million parameters while Alexnet has 60 million and VGGNet employs around 240 million parameters. Inception Model saved many parameters by replace a $N \times N$ filter with a stacking of $1 \times N$ and $N \times 1$ filters.

3 Task

This project aims to develop several algorithms of identifying art styles, apply them on large-scale image dataset and compare their performances. We'll use accuracy, precision/recall as evaluation metrics. We found this task pretty interesting because such kind of image classification is not about a specific object — i.e., classify if it is a car in the image, but about art styles such as: impressionism, baroque, cubism, and etc, which could be very abstract. From our literature review, most existing researches heavily rely on handcrafted feature extractors, it gives us a chance to explore more deep learning approaches and create innovative solutions.

4 Dataset

We scraped a subset of WikiArt via Python as our image dataset. WikiArt is a comprehensive and well-structured online repository of visual arts, storing digitized paintings in categories of different styles, genres, etc. Considering that number of paintings in different styles could vary a lot, we only collected styles with more than 1000 paintings. The dataset covers 30 styles, within each style, we randomly kept 1000 pictures to make it a balanced dataset. All the pictures are cropped and resized to size (512,512).
The styles set covers Early Renaissance, Op Art, Naïve Art (Primitivism), Expressionism, Magic Realism, Northern Renaissance, Rococo, Neo Expressionism, Ukiyo-e, Academicism, Art Nouveau (Modern), Pop Art, High Renaissance, Minimalism, Mannerism (Late Renaissance), Art Informel, Neoclassicism, Color Field Painting, Symbolism, Realism, Romanticism, Surrealism, Cubism, Impressionism, Baroque, Conceptual Art, Abstract Expressionism, Lyrical Abstraction, Post Impressionism, Abstract Art.
It is quite comprehensive as we included styles with both remarkable distinctions and strong similarities. The former one enables our classifier to learn better, the later one increases the difficulties of giving good performance.

5 Local Invariant Features

A lot of features in art styles are local invariant. For example, the intense color variation of impressionism1. Comparing the Impressionism example from the other two examples, we can see the color variation of Impressionism is much more dense than Academicism or Realism, in any part of the painting except the background part. Such features make it possible to use multi-patch trick2, which was proposed by Lu et al[5]. Randomly select small patches from a large image helps the neural network to learn local invariant features. Based on this characteristic,we also designed a structure that could capture these low-level feature quickly for our own model SenseNet.

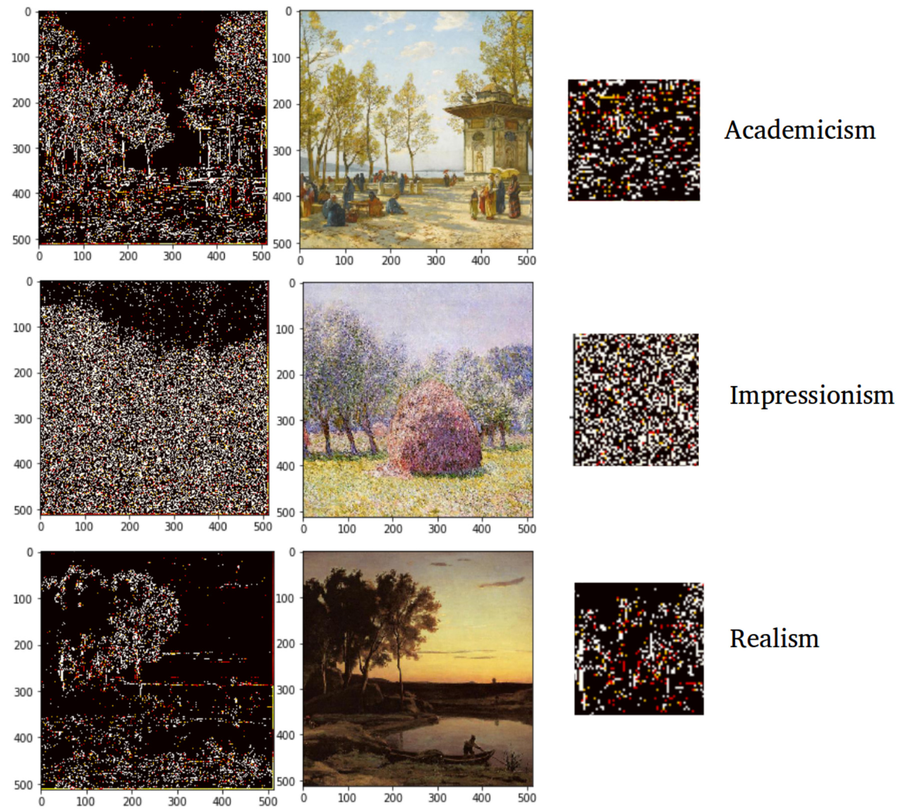

Figure 1: Color Variation of Different Styles

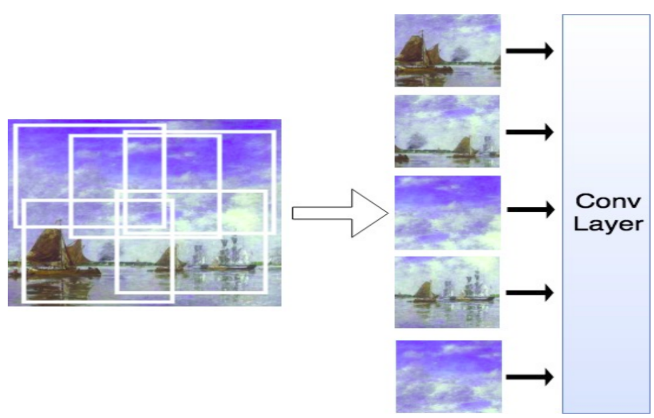

Figure 2: Multi-patch Trick

6 Modeling

6.1 Baseline model - LeNet

We built a baseline CNN model with LeNet-5[4], which has 2 2D-convolution layers, 2 max-pooling layers and 3 fully connected layers. Such a simple CNN network shows the very basic capability of deep neural network in this task. We trained the model with 75% of data, which has 22500 images in total, each style has 750 training samples. And the left 25% data to test. Both training and testing data are balanced . During the experiment, we set the stride size to 3 to reduce model complexity and save computational resources, and kept all other default settings the same. The baseline model achieved an accuracy of 10.77%. Considering that we did not do any hyper-parameter tuning and there are 30 classes, we conclude that CNN could be helpful to this task.

6.2 SenseNet

We built our own model based on the characteristics of art style classification task. Unlike object classification, art styles are abstract and does not have many structured features, most features scatter all over the painting, like we mentioned in last section. Based on the principle of capturing local-invariant features from low-level abstraction, we designed the SenseNet, which is shown in Figure3 below. We use three different filters in parallel to capture features after one layer of 1×1 convolution, then do a 64×64 max-pooling to find the region with strong feature representation, finally concatenate the average value of pooling result in each filter. To capture higher-level features, we added another column of 4-layer convolutions together with max-pooling layers. The original SenseNet got an accuracy of 22%. We then applied multi-patch trick on it, but the performance did not improve. We also tuned the number of filters, and it shows 64 filters is good enough.

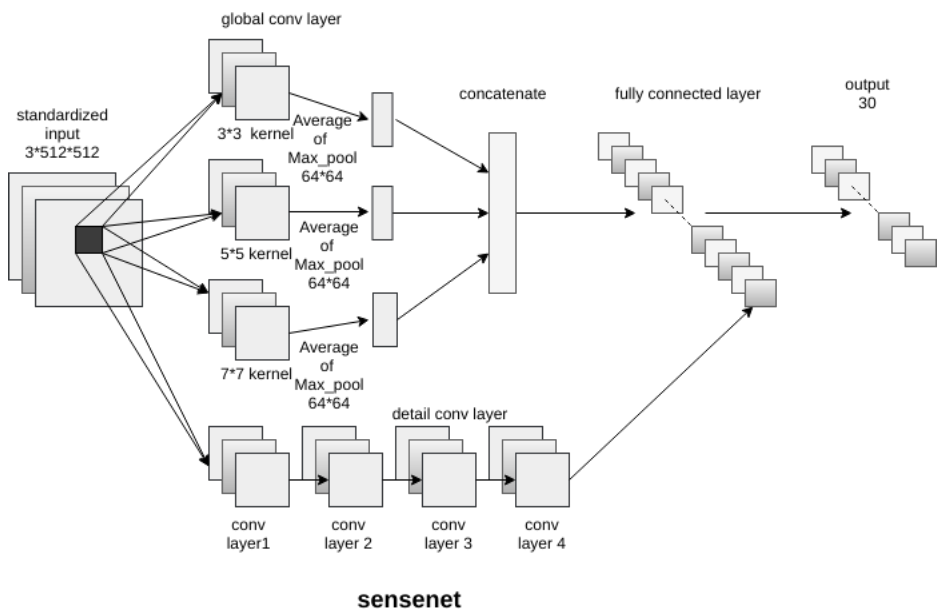

Figure 3: sensenet

6.3 Inception

Inception[8] V3 is a deep convolution neural work composed of couple convolution layers and inception cells. Inception cells are a multi-column convolution layer, for example, one column of 3×3 filters, one column of 1×3 and 3×1 filters and one column of 1×1 filters. Such structures save the usage of parameters but preserves high performance. Inception V3 has 5 convolution layers and 11 inception cells. We implemented Inception V3 for this task to see if a much deeper structure would perform better, also Inception's structure enables it to contain more filters. The raw Inception V3 achieved an accuracy around 27%. However, when we used the pre-trained Inception V3 from imagenet and only tune the fully connected layer, it achieved an accuracy of 34%.

Methods	Accuracy
LeNet	11%
$\text{LeNet}_{patch128}$	18%
$\text{LeNet}_{patch256}$	23%
SenseNet	22.2%
$\text{SenseNet}_{patch256}$	17.2%
Inception	27.0%
$\text{Inception}_{pretrained}$	34.4%

Table 1: Summary of Experiments

7 Conclusion And Future Work

Except the above experiments in Table 1, we have also applied multi-patch on Inception and combined the SenseNet with Inception, but we have not seen any improvement on these architectures. In conclusion, we believe our assumption about the characteristics of art style classification is partially correct. Multi-patch tricks and global feature extractors do help in this task. However, deeper architecture is also important. In the next step, we will keep trying to combine deeper architecture with SenseNet, and also pre-train the model on Imagenet to learn from more data.

References

[1] Yaniv Bar, Noga Levy, and Lior Wolf. "Classification of artistic styles using binarized features derived from a deep neural network". In: *Workshop at the European Conference on Computer Vision*. Springer. 2014, pp. 71–84.

[2] Alessandro Bergamo, Lorenzo Torresani, and Andrew W Fitzgibbon. "Picodes: Learning a compact code for novel-category recognition". In: *Advances in Neural Information Processing Systems*. 2011, pp. 2088–2096.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.

[4] Yann LeCun et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

[5] Xin Lu et al. "Deep multi-patch aggregation network for image style, aesthetics, and quality estimation". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2015, pp. 990–998.

[6] Babak Saleh and Ahmed Elgammal. "Large-scale classification of fine-art paintings: Learning the right metric on the right feature". In: *arXiv preprint arXiv:1505.00855* (2015).

[7] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).

[8] Christian Szegedy et al. "Rethinking the inception architecture for computer vision". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 2818–2826.