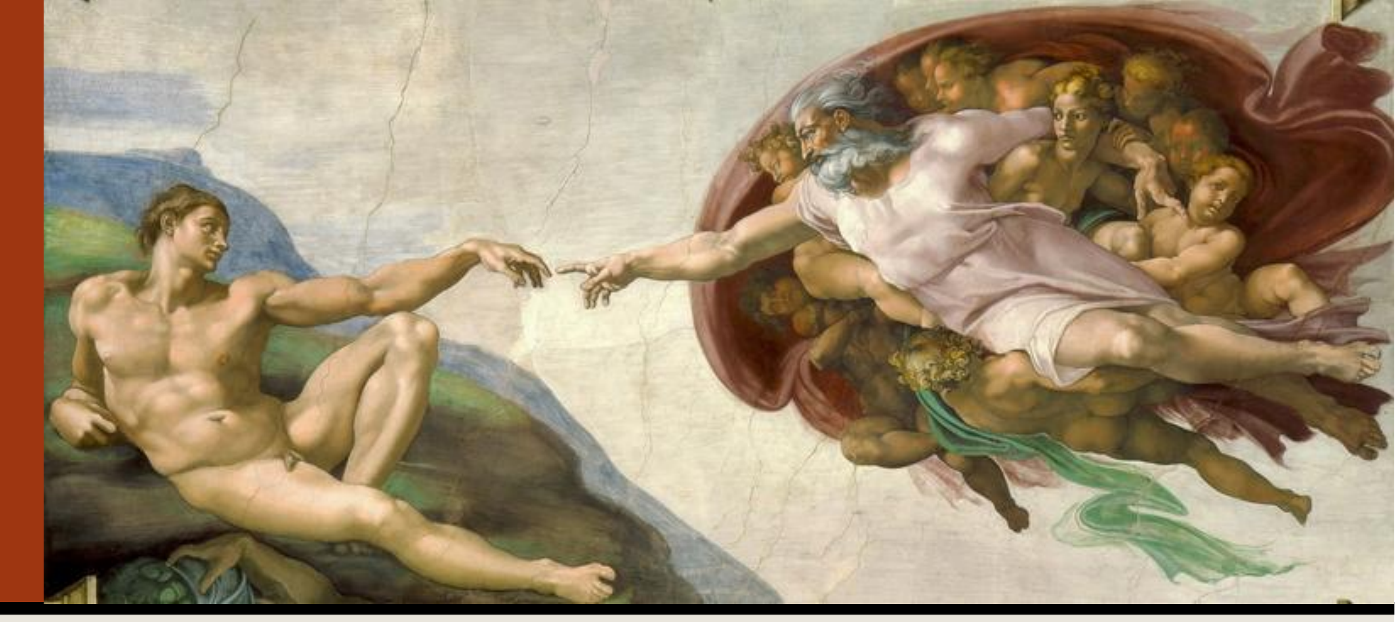
Classify Fine-Art Painting Style with Convolutional Neural Network

Yicheng Pu; Lige Wang; Xiaoxue Ma

New York University, Center for Data Science



ABSTRACT

In this project, we dealt with 30 fine-art styles covering Expressionism, Rococo, Cubism, Impressionism, Abstract Art, etc. Instead of using hand-crafted low-level descriptors, we turned to apply a range of deep learning algorithms for this classification problem.

We implemented Lenet-5, Inception and a novel architecture created by ourselves for this task.

Through the experiments we have verified our assumptions about art styles' characteristics, also found that deep neural networks, especially those pre-trained on large dataset, would be very helpful for the performance..

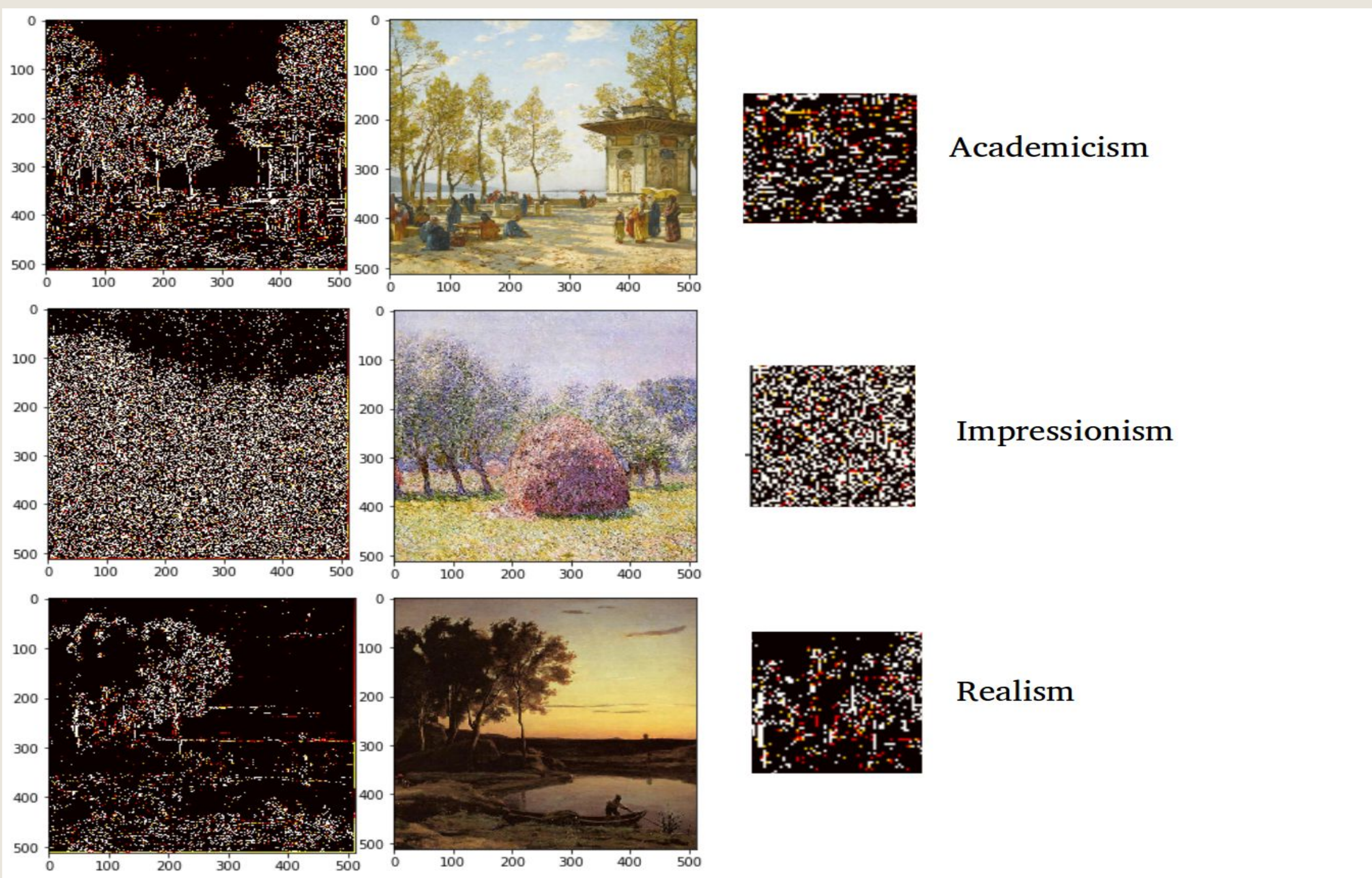
NYU CENTER FOR DATA SCIENCE

BASELINE MODEL

LeNet

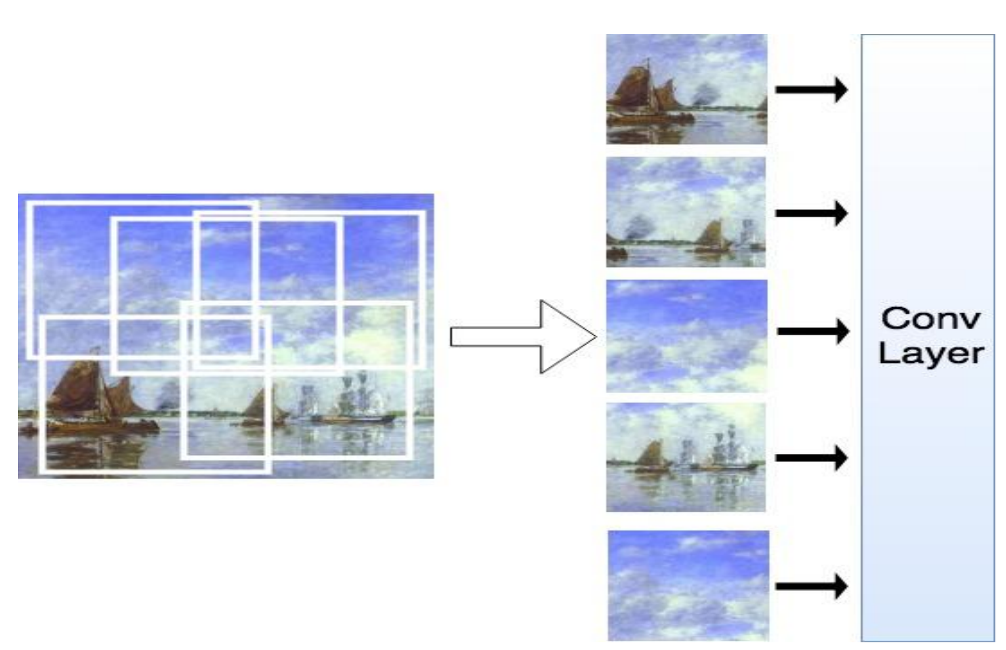
We built a baseline CNN model with Lenet-5, which has 2 2D-convolution layers, 2 max-pooling layers and 3 fully connected layers. We trained the model with 75% of data and the left 25% data to test. Both training and testing data are balanced. We set the stride size to 3 to reduce model complexity and save computational resource, kept all other defaulted. The baseline model achieved an accuracy of 10.77%. After integrating multi-patch, we got accuracy of 18% with patch size 128*128, and 23% with size 256*256.

LOCAL INVARIANT FEATURES



Color Variation of Different Art Styles

A lot of features in art styles are local invariant. For example, the intense color variation of impressionism. Such features make it possible to use multi-patch trick, also motivate us to design a structure that could capture these low-level feature quickly.



Multi-Patch Trick

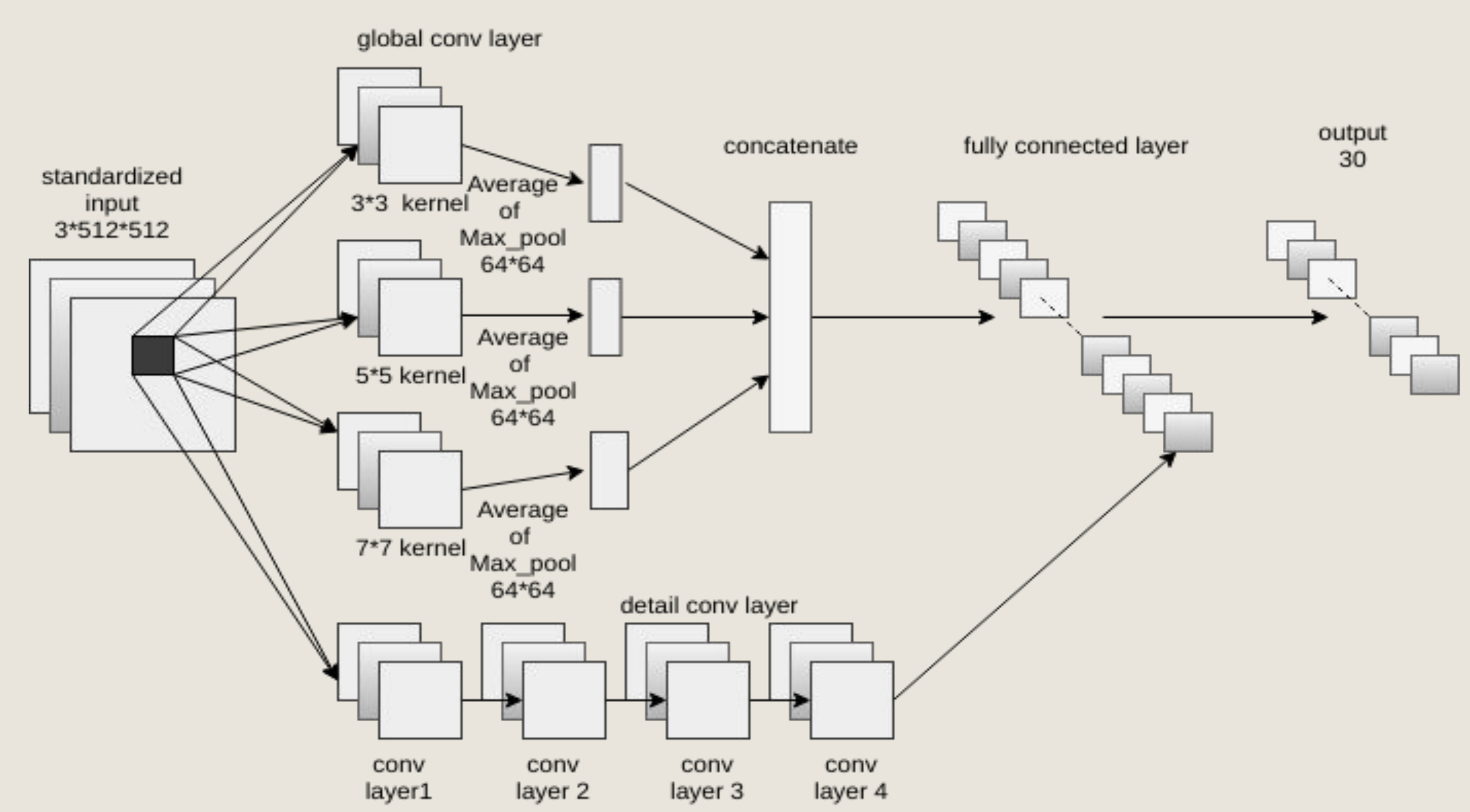
MODELS

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 Figure 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.



sensenet

Inception

Inception 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%.

RESULTS

MODELS	ACCURACY
LeNet	11%
LeNet (patch size 128)	18%
LeNet (patch size 256)	23%
SenseNet	22.2%
SenseNet (patch size 256)	17.2%
Inception	27.0%
Inception (pre-trained)	34.4%

CONCLUSIONS

Except the above experiments, 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] 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.

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

[3] 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.

[4] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: arXiv preprint arXiv:1409.1556 (2014). [5] 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.