A simple CNN model for painting classification solving few-shot learning problem

Ze Qian, Huiling Liang,Jinge Guo, Haotian Xue,Xu He, Tianyu Huang, Xinqi Liu

Abstract

The ABSTRACT is to be in fully-justified italicized text, at the top of the left-hand column, below the author and affiliation information. Use the word "Abstract" as the title, in 12-point Times, boldface type, centered relative to the column, initially capitalized. The abstract is to be in 10-point, single-spaced type. Leave two blank lines after the Abstract, then begin the main text. Look at previous CVPR abstracts to get a feel for style and length.

1. Introduction

CNN, short for Convolutional Neural Networks, is a specific kind of Multilayer perception, brought up by Yann LeCun in 1998. The success of CNN should be remained to its special application of local connection and shared weight, which reduces the quantities of weight to optimize the network on the one hand, and reduces the risk of overfitting on the other hand simultaneously.

CNN is one of the neural network, its weight sharing network makes it more like a biological neural network, meaning that it reduces the complexity of network model, meanwhile cutting the number of weight. This advantage is presented more obviously when the input of network is multidimensional images. It makes images directly as the input of the network, and avoids the complex process of feature extraction and data rebuilt.

CNN have great advantages over traditional techniques that it has remarkable abilities of false tolerance, it is capable of parallel processing and self studying. It could handle with complex environment information, and when solving problems that the background information is not known or the reasoning rules are not sure, it enables the samples to have reasonable deficiency or distortion. Meanwhile its running is high but it still have remarkable resolving power.

2. Related Work

Convolution neural networks have been widely used in a variety of computer vision tasks and have achieved remarkable performance in recent studies. However, algorithmic determination of artistic style in paintings has only been partly figured out. In [3] and [5] there are early examples of style classification with quite small datasets and several complex models are built in [4] by hand-engineering features on a larger dataset.

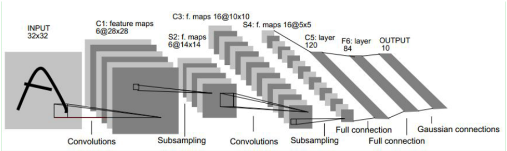


Figure 1. a tranditional structure of CNN Network

As is loosely described as ".. a distinctive manner which permits the grouping of works into related categories"[1], artistic style is hard to define. Therefore, algorithmically determining the artistic style of an artwork is a challenging problem that may include analysis of features like the paintings color, texture and subject matter. Additionally, digitization process might cause extra problems for our computer to correctly recognize the artistic style in paintings. For example, textures may get influenced by the resolution of the digitization. In spite of these challenges, intelligent systems for detecting artistic style would greatly contribute to identification and retrieval of artworks of a similar style.

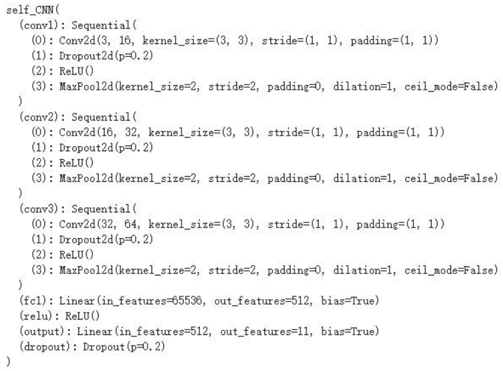
Our work is also related to overfitting.

Overfitting.Overfitting refers to the problem that a model after training can be fitted better than other models on training data, but it can not fit data well on data sets . The method of avoiding overfitting can be listed as :1. using simpler model structure2.regularization such as using L1-norm and L2-norm 3.data augmentation 4.dropout[6]5.early stopping6.ensemble Learning

Data augmentation.Data augmentation has been widely used in computer vision and natural language processing. When more training data is not available, transformations to the existing training data which reflect the variation found in images can synthetically increase the generalization ability of the model,thus increasing the accuracy in the test set.Existing methods of data augmentation can be roughly divided into three parts:1.geometric transformations such as image flip,image crop and image rotate[7] 2.add noise to image 3. change the color of the image appropriately

3. Model

Our network structure is shown in figure 2. We use three convolutional layers. Each layer includes a convolutional layer, an activation layer(ReLU) and a max-pooling layer. Then we use two fully connected layer to determine the final output based on the features extracted with convolutional network.



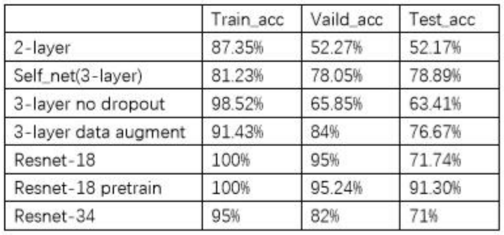
```
self_CNN(
  (conv1): Sequential(
    (0): Conv2d(3, 16, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (1): Dropout2d(p=0.2)
    (2): ReLU()
    (3): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  )
  (conv2): Sequential(
    (0): Conv2d(16, 32, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (1): Dropout2d(p=0.2)
    (2): ReLU()
    (3): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  )
  (conv3): Sequential(
    (0): Conv2d(32, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (1): Dropout2d(p=0.2)
    (2): ReLU()
    (3): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  )
  (fc1): Linear(in_features=65536, out_features=512, bias=True)
  (relu): ReLU()
  (output): Linear(in_features=512, out_features=11, bias=True)
  (dropout): Dropout(p=0.2)
)
```

Figure 2. our CNN Network structure

4. Experiment

4.1. outcome

our outcome is shown is figure 3



	Train_acc	Vaild_acc	Test_acc
2-layer	87.35%	52.27%	52.17%
Self_net(3-layer)	81.23%	78.05%	78.89%
3-layer no dropout	98.52%	65.85%	63.41%
3-layer data augment	91.43%	84%	76.67%
Resnet-18	100%	95%	71.74%
Resnet-18 pretrain	100%	95.24%	91.30%
Resnet-34	95%	82%	71%

Figure 3. chart for comparison

4.2. contrast and analysis

4.2.1 2 layer

The number of layers in CNN affects the predictive performance of the network. Deeper networks can extract more hidden features but take longer training time. We compared two and three convolutional layers convolutional neural networks. Obviously, In the two-layer CNN, as the number of training rounds increases, although the accuracy rate of the training set has been increasing, the ACC of the validation set has not been significantly improved.

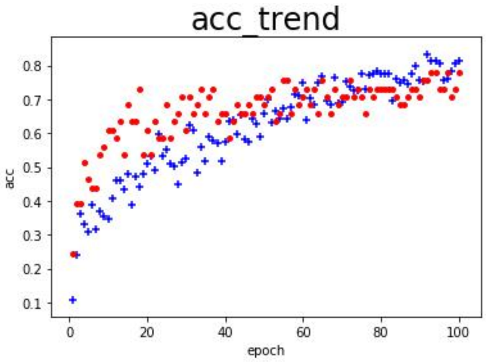


Figure 4. performance of our CNN Network

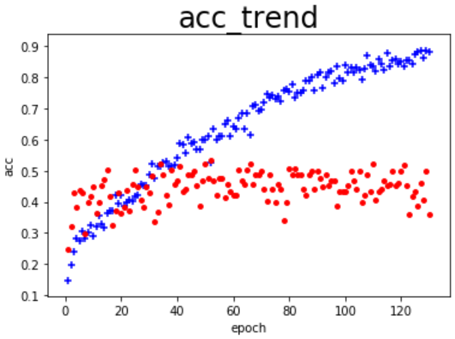


Figure 5. 2-layer-CNN

4.2.2 dropout

Dropout means that during the training process of deep learning network, the neural network unit is temporarily discarded from the network at a certain probability. Note that it is temporary. For stochastic gradient descent, each mini-batch is training different networks because it is randomly discarded. When the model is being trained, it is getting fitter to the training data because we are training the model according to the training data, which is called overfitting. Overfitting may leads to a consequence that the model doesnt fits other data well.

We're going to take a neuron and ignore it with a certain probability, while it's conducting forward, as is shown in the image below. It means we train a completely different models each epoch, before we add them up and get an average, which is our final model.

Whats more, for a neural network with N nodes, with dropout, it can be regarded as a set of 2n models, but the number of parameters to be trained at this time is unchanged, which frees up the problem of time consuming. According to the graph below, we can come to an conclusion that with the help of dropout, the model is fitting the training data more slowly while it fits the testing data better eventually. The model without dropout, however, doesnt perform well in the end.

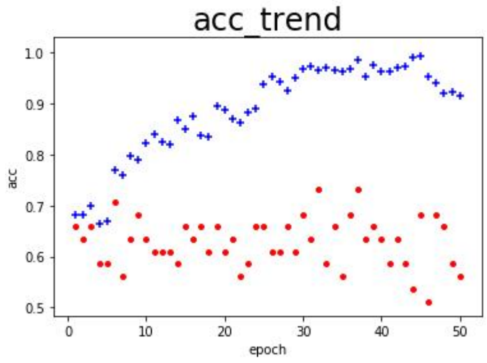


Figure 6. 3-layer without dropout

4.2.3 data augmentation

To enhance our data set, we use discoloration, adjusting brightness, flipping, zooming, rotating and so on. However, the result is contrary to our expectation. With our enhanced data set, the overfitting occurred even earlier, and the final accuracy rate is similar to the raw data set, which is around 76%.

augmentation.jpg

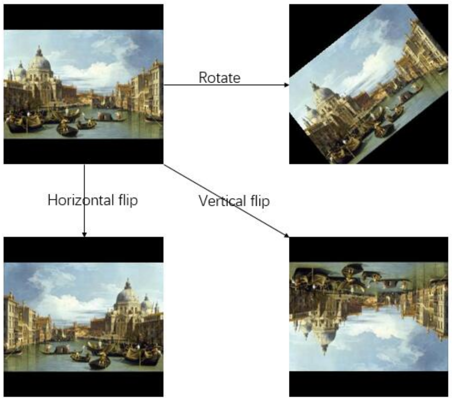


Figure 7. data augmentation methods

5. other's model:Resnet

To enhance the accuracy, the common practice is adding network layers. However, over adding could lead to a degradation problem, especially when the size of training data is small. So, we need to avoid identity mapping during the process of adding network layers, and one solution is the ResNet. The principle is:The residual learning unit establishes a direct association channel between input and output through the introduction of Identity mapping, so that the powerful reference layer concentrates on learning the residual between input and output, the structure is shown in Figure 8. We use the Resnet-18 and Resnet-34 to train our model and the result are shown below.

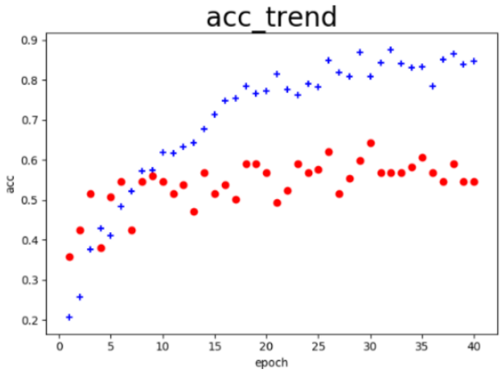
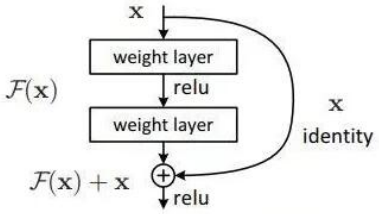


Figure 8. with data augmentation


x
weight layer
relu
weight layer
$\mathcal{F}(\mathbf{x})$
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

net.jpg

Figure 9. Residual network: a building block

5.1. Resnet-18

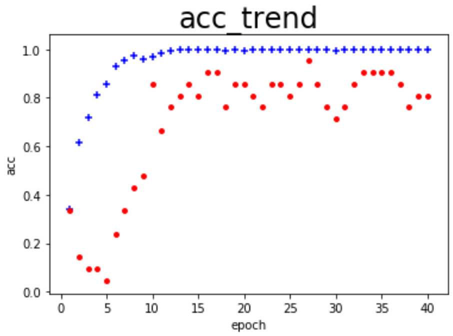


Figure 10. resnet-18-without-pretrain

The pre-trained model was trained on the ImageNet dataset. Therefore, it has a high accuracy rate for classifying objects on an image. In the test set, most of the objects depicted in the paintings of different authors are different, so the pre-training model can obtain high accuracy rate after several rounds of training.

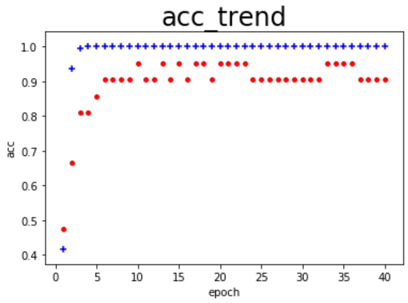


Figure 11. resnet-18-with-pretrain

5.2. Resnet-34

The performance on Resnet-34 is not very good. According to our analysis, the reason is that the network is so deep and we do not have large dataset to support the training of deep network.

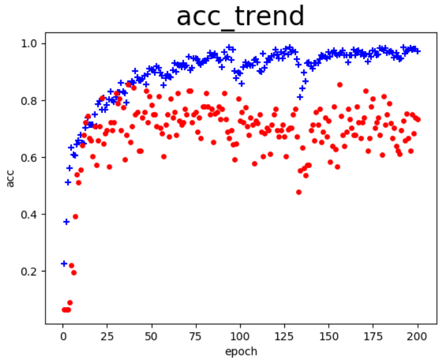


Figure 12. resnet-34-without-pretrain

6. Conclusion

We implement a three-layer CNN and tried to solve the problem of identifying the author of the painting. And we compare this model with several existing mature models such as resnet18, resnet34. A better model does not improve the accuracy rate of the test set significantly while pre-trained models work best because the data set is too small.

References

[1] FERNIE, E. Art History and its Methods: A Critical Anthology. London: Phaidon, 1995.

[2] KARAYEV, S., HERTZMANN, A., WINNEMOELLER, H., AGARWALA, A., AND DARRELL, T. Recognizing image style. CoRR abs/1311.3715 (2013).

[3] KEREN, D. Recognizing image style and activities in video using local features and naive bayes. Pattern Recogn. Lett. 24, 16 (Dec 2003), 29132922.

[4] SALEH, B., AND ELGAMMAL, A. Large-scale classification of fine-art paintings: Learning the right metric on the right feature. In International Conference on Data Mining Workshops (2015), IEEE.

[5] SHAMIR, L., MACURA, T., ORLOV, N., ECKLEY, D. M., AND GOLDBERG, I. G. Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art. ACM Trans. Appl. Percept. 7, 2 (feb 2010), 8:18:17.

[6] G.E.Hinton.Improving neural networks by preventing co-adaptation of feature detectors. arXiv:1207.0580(2012)

[7] Andrew G. Howard. Some Improvements on Deep Convolutional Neural Network Based Image Classification. arXiv:1312.5402(2013)