Surface Blemishes of Aluminum Material Image Recognition Based on Transfer Learning

Hao Ming and Ke Xu*
College of Computer Science, South-Central University for Nationalities, Wuhan 430074, Hubei, China.
Email: monkeydlufyaha@foxmail.com, cole.xu@gmail.com

Abstract. Image datasets in the field of industrial manufacturing usually have the problem of uneven distribution of samples. To solve this problem, this paper utilizes transfer learning to recognize surface blemishes of Aluminum material image. In order to get rid of the disadvantage of VGGNet too many parameters, this paper combines VGGNet with Network-in-Network to generate a new model. The results of experiments is shown that it has achieved significant improvement by using transfer learning. Additionally, the new model also achieves a better performance and the number of parameters of the new model is much smaller than that of the original VGGNet.

1. Introduction
In the production process of aluminium profiles, due to various factors, the surface of aluminium profiles will produce cracks, peeling, scratches and other defects, which will seriously affect the quality of aluminium profiles. Nevertheless, the surface of aluminium profile itself will contain grain, and the distinction between defects is not high. Traditional manual naked eye inspection is very laborious and cannot accurately judge the surface defects. In order to deal with this problem, deep learning can be applied to classify these aluminium profiles. But the defective aluminium profile is far less than the normal aluminium profile, and the sample distribution of aluminium profile datasets is uneven. Transfer learning[1] have a pretty good performance in this problem. Take the most common example of image recognition and train a neural network. To identify different breeds of cats, if you train from zero, you need millions of levels of even and tagged data. If using transfer learning, only thousands of pictures are enough, even these thousands of pictures, using a mature network like Inception[2] or vgg16[3], which is published by Google.

The VGGNet model is the second place in the ILSVRC competition in 2014 and it performance better in transfer learning task than GoogleLeNet[4] which is the top place in the ILSVRC competition in 2014, however, it has a large number of parameters that cause long training, large memory overhead and overfitting. Accordingly, his paper combines VGGNet model with NIN (network in network)[5] to get a new model which has only one tenth of parameters of VGGNet model and is as the same deep as vgg16. Finally, this paper uses loss, accuracy, ROC, AUC and other indicators to test the effect of the model[6].

2. Related Works
In recent years, deep learning has achieved very good results[7], surpassing most other machine learning algorithms, but deep learning needs a large quantity of balanced data, sometimes we cannot obtain a good many of balanced data, then the effect of deep learning is not so good any more. The size and balance of datasets has always been a hard problem in the field of deep learning. Nonetheless,
Unlike previous isolated neural network models, transfer learning enables these isolated neural network models to be connected together. The new neural network model does not need to start from zero, but can learn on the basis of the existing trained model, which makes it possible to have a great performance of the neural network model even trained on a very small and uneven dataset[1]. Transfer learning enables us to achieve excellent results when datasets are not perfect enough.

Image set classification has been applied in many real-world scenarios, such as classification from personal albums, surveillance video[8], multi-view camera networks[9]. And the deeper structure of the neural network model began to appear. The deeper network structure can make the model have better non-linear expression ability, but the deeper model will lead to the disappearance of gradient or gradient explosion, so this paper does not directly replace the ordinary convolution layer with the MLP layer, but the block of convolution layers in the VGGNet model. This makes the new model have the advantage of vgg network deep enough, and has no shortcomings of vgg network computing too much.

ROC analysis technology has roots in statistical decision theory in the 1950s to explain the relationship between classification hit rate and false alarm rate. Spackman then introduced ROC analysis technology into machine learning field, and explained ROC curve value estimation and comparison algorithm. Later, ROC analysis technology has been applied more and more in the field of machine learning in recent years[6]. ROC curve enables us to see the effect of the model more intuitively.

3. Methods

3.1. Model Structure

The new model VGG-NIN combines normal vgg16 model with NIN(network in network). The main idea of NIN is replacing the normal convolutional layer with MLP(Multi-Layer Perception) which is shown in figure 1 and the fully connected layer is replaced with GLP(global average pooling).

![Multi-Layer Perception](image)

**Figure 1.** Multi-Layer Perception

MLP is a 3x3 convolutional layer and two 1x1 convolutional layers and in the classification layer, GLP is interpreted more easily and less inclined to overfitting. The structure of vgg16 model is shown in figure 2.
To get the VGG-NIN model we replace the 2\textsuperscript{nd}, 4\textsuperscript{th}, 6\textsuperscript{th}, 7\textsuperscript{th}, 9\textsuperscript{th}, 11\textsuperscript{th} and 12\textsuperscript{th} 3x3 convolutional layers with 1x1 convolutional layers and GLP take place of three fully connected layers. The structure of VGG-NIN model is shown in figure 3.

Because pictures have a very large size (2560, 1920) and defects in some images are distributed around images. We can't cut the image directly to a relatively small pixel, which will lose many important features. And this paper chooses to add an average pooling layer before each model to extracted those important features distributed around images.

3.2. Strategies of Model Transfer Learning

Deep learning usually requires a large amount of data, and the uneven distribution of samples may lead to poor performance of the model. Taking aluminum profile dataset as an example, if the classifier is trained from zero, the model may not achieve the expected results due to the uneven distribution of samples. So we can consider training the network on large-scale datasets such as ImageNet, and then transfer the model to our dataset.

4. Experiments and results analysis

4.1. Dataset Introduction

Aluminium profile dataset whose example is shown in figure 4 used in the Guangdong Industrial Intelligence Big Data Innovation Competition held in Alibaba. This dataset includes 12 classifications, including normal aluminium profiles and 10 different defective aluminium profiles, as well as one other class that includes many other defects. And this dataset contains 2376 images with a pixel size of (2590,1920), as shown in the figures. Moreover, the sample distribution of the dataset is extremely uneven, for example, there are 904 normal aluminium profiles and only a few dozen defective aluminium profiles.
4.2. Training model
Firstly, because sample distribution is uneven, data augmentation is carried out for classifications with very few images to balance the sample distribution as far as possible. And then normalizing the whole dataset. Secondly, average pooling layer with 5x5 filter and 5 strides is added before all the models to reduce the image dimension for the large size of images. After image preprocessing, training vgg16 model and VGGNet model combined with NIN (network in network), and these two models are both into two cases which are training directly and transferring models trained on ImageNet.

4.3. result of this experiment
The dataset is divided into training set and test set according to the ratio of 9 to 1. Results of different models are arranged in table 1, and four ROC curve diagrams are shown in figure 5.

| Model            | loss  | accuracy | precision | recall  | F1     | AUC     | parameters       |
|------------------|-------|----------|-----------|---------|--------|---------|------------------|
| Vgg16            | 2.22  | 0.429    | 0.704     | 0.778   | 0.739  | 0.804   | 24,170,060       |
| Vgg16(transfer)  | 2.18  | 0.617    | 0.763     | 0.812   | 0.787  | 0.890   | 24,170,060       |
| VGG-NIN          | 2.32  | 0.577    | 0.720     | 0.805   | 0.760  | 0.837   | 2,275,084        |
| VGG-NIN(transfer)| 1.28  | 0.730    | 0.822     | 0.887   | 0.853  | 0.934   | 2,275,084        |

It can be seen from the four groups of data in the table above that the model using transfer learning is much better than the model without transfer learning, for example, the accuracy of the vgg16 model without transfer learning is only 0.429 in the test set, but the accuracy of the vgg16 model with transfer learning is 0.617 in the test set, and the indicators of the model with transfer learning are all better than those without transfer learning. Moreover, the vgg16 model combined with NIN (network in network) has also become more excellent, the loss has dropped from 2.32 to 1.28, and the various indicators of the VGG-NIN model are also greater than the ordinary vgg16 model. In addition, the number of parameters in the VGG-NIN model which is 2,275,084 is less than one tenth of that in the ordinary vgg16 model which is 24,170,060, so it greatly shortens the training time and reduces the memory overhead.
As can be seen from the above four ROC curves, whether or not transfer learning is used, the area under the ROC curve whose abbreviation is AUC of the VGG-NIN model is much larger than that AUC of the VGG-NIN. In both cases, AUC increased by 0.4, from 0.80 to 0.84 without using transfer learning, and from 0.89 to 0.93 after using transfer learning.

5. Conclusion
Aiming at the problem of uneven distribution of samples in aluminium profile image datasets, this paper proposes a model based on transfer learning which highly improves the performance of the model. In addition, modifying the traditional vgg16 model via combining it with NIN can greatly reduced the number of parameters, and the new model also achieved better results.

6. Acknowledge
This research is supported by the Natural science foundation of Hubei province (No. 2017CFC840), Supported by “the Fundamental Research Funds for the Central Universities”, South-Central University for Nationalities (CZY19012), and the 2018 college student innovation and entrepreneurship training program of South-Central University for Nationalities (agricultural products
trading platform based on block-chain). Ming is currently an undergraduate, Ke Xu is the corresponding author and an instructor in SCUN.

7. Reference

[1] Pan S J , Yang Q. A Survey on Transfer Learning. IEEE Transactions on Knowledge & Data Engineering, 2010, 22(10):1345-1359.

[2] Szegedy C , Vanhoucke V , Ioffe S , et al. Rethinking the Inception Architecture for Computer Vision. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, IEEE Computer Society, 2016: 2818-2826.

[3] Simonyan K , Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. Computer Science, CoRR abs/1409.1556, 2014.

[4] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, IEEE Computer Society, 2015: 1-9.

[5] Lin M , Chen Q , Yan S. Network In Network. CoRR abs/1312.4400, 2013.

[6] Bradley A P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition, 1997, 30(7): 1145-1159.

[7] Wan J , Wang D , Hoi C H , et al. Deep Learning for Content-Based Image Retrieval: A Comprehensive Study. Proceedings of the ACM International Conference on Multimedia, Orlando, FL, USA, ACM, 2014: 157-166.

[8] Teng K , Wang J , Xu M , et al. Mask Assisted Object Coding with Deep Learning for Object Retrieval in Surveillance Videos. Proceedings of the ACM International Conference on Multimedia, FL, USA, 2014: 1109-1112.

[9] Ma L , Stückler, Jörg, Kerl C, et al. Multi-View Deep Learning for Consistent Semantic Mapping with RGB-D Cameras. Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vancouver, BC, Canada, 2017: 598-605.