Research of Image Recognition and Classification Based on NIN Model

Liu Mengxi\textsuperscript{1,2}, Ju Yongfeng\textsuperscript{1}, Song Jiuxu\textsuperscript{2} and Wang Zheng\textsuperscript{2}

\textsuperscript{1}Shaanxi Key Laboratory of Measurement and Control Technology for Oil and Gas Wells, Xi’an, China; \\
\textsuperscript{2}School of Electronic and Control Engineering, Chang’an University, Xi’an, China; \\
E-mail:26702163@qq.com

Abstract. Focus on the weak feature expression ability of traditional Convolution Neural Network (CNN) in image recognition and classification, the CNN has been improved and optimized by introduction of Mlpconv layer to form a Network in Network (NIN). The NIN enhances feature extraction and abstract expression of the local patch; get better performance of the image recognition. The experiments of weld image recognition and classification show that the optimized CNN can improve the feature expression ability of the whole network by reducing the number of parameters, obtain higher recognition and classification precision, and avoid the fitting of the network model effectively. The CNN achieve improvement in both performance and efficiency.

1. Introduction

As a method of accurate measurement and numeral expression, Digital Image Processing (DIP) is rapidly applied to industrial production such as material meso spatial structure and geometrical form\cite{1}. Due to the advantages of simple to operation and easy to obtain, X-ray weld flaw detection are widely used. The X-ray image has a high value of judgment. In order to ensure the quality of welded structures, it is necessary to identify and classify defects through X-ray images of welds.

With the technology progresses, image recognition and classification of automatic detection technology has made great progress. Deep learning has been successfully applied in the field of the computer vision such as image classification, target detection, image retrieval, image segmentation and other tasks. The most commonly used structure of deep learning includes Convolution Neural Network(CNN)\cite{2}, Deep Belief Network(DBN)\cite{3}, AUTO-ENCODERS(AE)\cite{4}, Generative Adversarial Net(GAN)\cite{5}. At present, many applications of image recognition and classification are carried out around the optimized deep network structure, which provide a way for the weld defects recognition and classification.

In 2014, Lin M proposed the Network in Network (NIN)\cite{6}, which improved the traditional CNN by introducing the mlpconv layer, using fewer parameters to achieve higher accuracy rate on data sets as CIFAR-10, CIFAR-100 etc. Based on the preprocessing of X-ray digital Weld image, a high precision recognition and classification of weld defects is realized by NIN network structure.

2. Image Preprocessing
Due to many parameters and the environmental interference, weld cannot avoid defects, which are mainly two types: round defects and linear defects, shown in Figure 1, and weld defects can be subdivided into porosity, cracks, slag inclusions, and non-fusion, incomplete penetration.

![Figure 1. Weld Defect Type](image)

By the influence of weld defect type, size, position and other factors, the original image may loss vital information according to the general image processing method. The paper proposed a method: First, the original image is processed into an image less than 256x256. Considering the image may distort, longer side of the image transform into 256, shorter side carries on the equal-ratio transformation [7]. Then the image is divided into four vertices and centers of five images, each image is 227×227 size., which is shown in Figure 2:

![Figure 2. Weld defect image segmentation diagram](image)

For this method, the picture can achieve full coverage without omission feature information while the number of samples can be increased so that correct rate of identification can be ensured. The specific transformation algorithm is is shown in Table 1:

| Image preprocessing algorithm |
|------------------------------|
| **Input:** InImage           |
| **Output:** OutImage         |
| **Step1:** Compute the image length and width, the longer side denoted as ma, the shorter side denoted as mi; |
| **Step2:** If ma > 256, go to the equal proportion conversion by 256/mi, The image denoted as Tr.; |
| **Step3:** The image Tr is intercepted from four vertices and centers of 227 × 227 image blocks, and 5 training images generated. which denoted as OutImage[0]; OutImage[1] ; OutImage[2] ; OutImage[3] ; OutImage[4]; |
| **Step4:** Output outImage   |
3. NIN construction

For the CNN's data patch, convolution kernel is essentially a Generalized Linear Model (GLM)[8], which has weak ability to express the feature and the lower level of abstraction. When the concept sample is linearly measurable, GLM will achieve good results and complete a good abstraction, so the traditional CNN can get good training results for linear and measurable samples.

In the actual process, the data under the same concept is usually the nonlinear, so the abstract expression of these concepts is often a highly nonlinear equation of input. Traditional CNN will not be able to do any better abstraction of all layer features; Therefore, GLM should be replaced by a more effective nonlinear function approximation to improve the abstract expression of the local model. As a general function approximation, Multi-Layer Perceptron (MLP)[9] is also a deep model, which can be compatible with CNN. With BP algorithm, Mlp is consistent with the spirit of feature reuse[10]. Based on the above principles, the network is improved and optimized through using the Mlpconv network layer instead of the traditional convolution layer to achieve the approximation of nonlinear equations. The improved network structure is shown in Figure 3, in the case of cross channel; Mlpconv is equivalent to the convolution layer added 1x1 convolution layer. The 1x1 convolution layer is adopted to realize cross channel interaction and information integration while carry out the dimensionality reduction of the convolution kernel channel number according to the need[11].

![Convolution layer optimization diagram](image)

*Figure 3. Convolution layer optimization diagram*

For traditional linear convolution layer, the specific formula is:

\[ f_{i,j,k} = \max(W_k^T x_{i,j}, 0) \]  \hspace{1cm} (1)

Here \((i, j)\) is the pixel index in the feature map, \(x_{i,j}\) stands for the input patch centered at location \((i, j)\), and \(k\) is used to index the channels of the feature map.

For single channel mlpcovn layer, the specific formula is:

\[
\begin{align*}
    f_{i,j,k}^1 &= \max(W_k^T x_{i,j} + b_k, 0) \\
    f_{i,j,k}^n &= \max(W_k^n f_{i,j}^{n-1} + b_k, 0)
\end{align*}
\]  \hspace{1cm} (2)

Here \(n\) is the number of layers in the multilayer perceptron. Rectified linear unit is used as the activation function in the multilayer perceptron. Since it is multichannel convolution, so the \(W_k^n\) and \(x\) are three-dimensional, using ReLU as Mlp activation function.

In the above formula, the formula (1) is the same as the traditional convolution, that is, the convolution kernel and the patch convolution of the characteristic graph are output through ReLU, \(f_{i,j,k}^1\) is essentially a value of \([1, 1]\), and all the \(k\) convolution results are combined as \(f_{i,j}^1\), which is a \([1, 1]\) vector of output channel, the vector is output to the next layer, and the next layer starts with MLP.
From the perspective of the cross channel pooling, the formula (2) is equivalent to the Cascade Cross Channel Parameter Pooling (CCCP) of traditional convolution layer. Each pooling layer is actually a weighted linear reorganization of the input feature graph, and then passes through a ReLU. These cross channel feature graphs pooling iterate over the channel in the next layer, which allows for complex and learning interactions across the channel information.

![Figure 4. NIN structure diagram](image)

According to the method above, the improved network is shown in Figure 4, the NIN is a four-layer convolution neural network, and the main parameters are shown in the Table. For the FC layers is very easy to fitting and and relies heavily on the dropout to generalization, the global average pool itself has a generalization that can naturally prevent the entire structure from being fitted together. Global Average Pooling is used to replace the FC layer for output. In the last layer of mlpcon. the spatial average of the feature map is used as the basis of classification. The final vector input softmax for classification.

| Layer | Convn.kernel | Channels | Pooling | Input |
|-------|--------------|----------|---------|-------|
| 1     | 11×11        | 96       | 3×3     | 227×227→55×55 |
|       | Stride4 padding0 |          | Stride2 padding0 |       |
| 2     | 5×5          | 256      | 3×3     | 27×27  |
|       | Stride1 padding2 |        | Stride2 padding0 |       |
| 3     | 3×3          | 384      | 3×3     | 13×13  |
|       | Stride1 padding1 |      | Stride2 padding0 |       |
| 4     | 3×3          | 1024→1000| 6×6     | 6×6    |

| Layer | Convn.kernel | Channels | Pooling | Input |
|-------|--------------|----------|---------|-------|
| 1     | 11×11        | 96       | 3×3     | 227×227→55×55 |
|       | Stride4 padding0 |          | Stride2 padding0 |       |
| 2     | 5×5          | 256      | 3×3     | 27×27  |
|       | Stride1 padding2 |        | Stride2 padding0 |       |
| 3     | 3×3          | 384      | 3×3     | 13×13  |
|       | Stride1 padding1 |      | Stride2 padding0 |       |
| 4     | 3×3          | 1024→1000| 6×6     | 6×6    |

### 4. Numerical experiments

The pictures in the experiment are selected from the steel tube weld X-ray image data from the industrial production, and the image size is randomly selected. According to the algorithm of the article, the processed image 227x227 size is used as input of the NIN network. The experimental parameters of NIN are shown in the Table 2. At the same time, to verify the fitting performance, classification experiments select 500 training samples and test samples of each, using Softmax to classify. The traditional CNN and NIN to the weld image experiment results which include circular defect and lineshape defect are shown in the Figure 5 and Figure 6:
It is known from the Figure 5 and Figure 6 that for the error rate of weld defects, the difference between CNN and NIN is not obvious, but the convergence rate of NIN is much faster than that of CNN. NIN linear convolution covers all potential concepts without a supercomplete convolution kernel, which greatly reduce the number of parameters, enhancing the network performance. In addition, the error rate of NIN test sample and training sample tends to be consistent, which avoids the overfitting.

In the more detailed classification experiment, the defect types are porosity (1), cracks (2), slag inclusions (3), non-fusion (4), and incomplete penetration (5). The experimental results are shown in the Table 3. From the Table 3, for the slag inclusions, the recognition rate of NIN is lower than that of CNN. For other weld defects, porosity, cracks, non-fusion, and incomplete penetration, NIN can classify weld defects accurately and achieve well fitting. In general, the classification recognition effect of NIN is ideal.

5. Conclusion
The improved CNN overcomes the problem that the linear convolution cannot fit the expression highly non-linear characteristic and linear convolution cannot ensure coverage of all potential concepts.
unless an ultra complete convolution kernel. Because the 1x1 convolution kernel is adopted, the information interaction and learning between channels are strengthened. The global average pooling enhanced correlation between feature map and corresponding classification. The network training performance has been greatly improved and has avoided the fitting effectively.

Table 3. Weld defects classification results

| Defect type | Testing sample recognition rate (%) | Training sample recognition rate(%) |
|-------------|-------------------------------------|------------------------------------|
|             | CNN       | NIN   | CNN       | NIN   |
| 1           | 92.49     | 95.89 | 92.89     | 96.20 |
| 2           | 93.89     | 97.01 | 95.29     | 95.97 |
| 3           | 94.18     | 93.14 | 96.19     | 92.98 |
| 4           | 93.56     | 94.65 | 94.26     | 94.97 |
| 5           | 92.36     | 95.23 | 95.22     | 96.08 |

References
[1] Liao C W, Yu J H and Tarng Y S 2010 On-line full scan inspection of particle size and shape using digital image processing Particuology, 8(3) 286-92.
[2] Krizhevsky A, Sutskever I and Hinton G E 2012 ImageNet classification with deep convolutional neural networks International Conference on Neural Information Processing Systems 60 1097-105
[3] G E Hinton and R Salakhutdinov 2006 Reducing the dimensionality of data with neural networks Science 313(5786) 504-7
[4] Stolcke A, Coccaro N, Bates R, Ess-Dykema C V and Ries K. 2000 Dialogue act modeling for automatic tagging and recognition of conversational speech. Computational Linguistics 26(3) 339-73
[5] Goodfellow I J, Pougetabadie J, Mirza M.,Xu B, Wardefarley D and Ozair S. 2014 Generative adversarial networks Advances in Neural Information Processing Systems 3 2672-80
[6] Lin M, Chen Q and Yan S 2013 Network in network Computer Science
[7] Jia Y, Shelhamer E, Donahue J, Karayev S, Long J and Girshick R 2014 Caffe: Convolutional Architecture for Fast Feature Embedding Acm International Conference on Multimedia 675-78
[8] Babadi B, Casti A, Xiao Y, Kaplan E. and Paninski L. 2010 A generalized linear model of the impact of direct and indirect inputs to the lateral geniculate nucleus Journal of Vision 10(10) 22
[9] Zhang Z, Lyons M, Schuster M and Akamatsu S 1998 Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron IEEE International Conference on Automatic Face and Gesture Recognition 1998 Proceedings 454-59
[10] Mannion M, Kaindl H and Savolainen J 2017 Product Line Strategies and Feature Reuse The, International Systems and Software Product Line Conference 252.
[11] Cascianelli S, Bello-Cerezo R, Bianconi F, Fravolini M L, Belal M and Palumbo B Intelligent Interactive Multimedia Systems and Services (Springer) p 169