Optimization of Deep Convolution Neural Network Based on Sparse DropConnect

Liu Mengxi¹, ², *, Song Jixiu², Wang Zheng² and Ju Yongfeng¹

¹School of Electronic and Control Engineering, Chang’an University, Xi’an, 710064, China; ²Shaanxi Key Laboratory of Measurement and Control Technology for Oil and Gas Wells, Xi’an, 710065, China

*E-mail: 26702163@qq.com

Abstract: According to the problem of overfitting in the traditional Convolutional Neural Networks (CNN) with large samples, an improved sparse DropConnect algorithm is put forward. With proposed algorithm, the CNN is improved and optimized. The performance of the network in image recognition is verified. Experiments of weld image recognition and classification indicate that the selective ability of the CNN on the sparse feature is improved. High recognition accuracy and classification accuracy of test sample are achieved. Over-fitting in the traditional network model can be avoided. More important, the feature sparsity and discriminative ability of the CNN are increased.

1. Introduction

Deep Convolutional Neural Networks (CNN), as one of the deep learning models, is suitable for processing of large-scale image data. With regard to CNN feature extraction ability, rapid increase and decrease can be achieved by adjustment of the number of network layers and the number of neurons per layer [1]. In terms of application, when passing each layer, CNN input data will obtain corresponding significant features of observation data, which include significant features such as translation, scaling, and rotation invariance. Then, by means of layer-by-layer extraction, advanced features of the image can be extracted at a high level and successfully applied to target detection, image classification, speech recognition and other fields [2]. At present, many studies on application scenarios for image recognition and classification are centered on the optimization of deep network structure. there are improved RCNN [3], Faster-RCNN [4], and YOLO [5] network structures that break the limitations of traditional Convolutonal Neural Networks through area selection.

However, large image data sets have a great influence on the images recognition speed, and the network scale of CNN also increases, making the abstract relationship between training samples and features more and more complex. The BP algorithm enables the CNN to start from a randomly initialized starting point and achieves a near-perfect recognition result on a limited training sample after a sufficient number of iterations. Due the feature extractor is trained too well for the training sample, there are deviations from the image features in the testing sample, resulting in overfitting [6]. This requires CNN to be optimized, considering enhancement of the network performance by means of improving the sparsity of the CNN.

2. Optimization design of sparse CNN
In order to prevent over-fitting between training sample and test sample, this paper introduces DropConnect [7] algorithm, which schematic diagram as shown in the figure 1. In the DropConnect algorithm, the connection weights between the two networks are zeroed by a certain probability, and the neuron weight mask discard is generated by a Bernoulli variable, which is independent of the neuron weights and activation values. Compared with the DropOut algorithm, the DropConnect algorithm has better accuracy and efficiency in image recognition and classification.

![DropConnect structure diagram](image)

**Figure 1.** DropConnect structure diagram

2.1 DropConnect sparsity investigation

To introduce a sparse mechanism in the DropConnect algorithm, all weights are dropped with the same probability. If different neurons allocate different drop probabilities according to the sparsity of the activation value, the weight mask becomes the function of the activation value of the neuron. The improvement enables the network model to dynamically adjust the drop probability of the weight value during the training process according to the sparsity of the output value of the network layer, and thus dynamically adjust the statistical model of the network, allowing the neuron connection with sparse feature extraction capability to have greater probability of being activated, eventually leading more sparse features of network model extraction [8].

The Sparse DropConnect training algorithm can be described as:

\[
\text{mask}^{(1)}_{ij} \sim \text{Uniform}(0.5) \\
\rho^{(l)}_{ij} = p_{\text{drop}}(y^{(l-1)}_j) \\
\tilde{W}^{(l)}_{ij} = \tilde{W}^{(l)} \circ (\text{mask}^{(l)}_{ij} > \rho^{(l)}_{ij}) \\
z_i = \sum_{j=1}^{g} \tilde{W}^{(l)}_{ij} y^{(l-1)}_j + b_i^{(l)} \\
y^{(l)} = f(z_i) \\
\sum_M f((M \circ W)x) \approx f(\sum_M (M \circ W)x)
\]

Where, \( W^{(l)}_{ij} \) is for the connection weight between the \( j \) neuron in the \( l-1 \) layer and the \( i \) neuron in the \( l \) layer; \( p^{(l)}_{\text{drop}}(\cdot) \) is for the drop probability of the weight associated with the weight \( W^{(l)}_{ij} \), and the more sparse neuron connection has a lower drop probability; \( p_{\text{drop}}(\cdot) \) is for the calculation function of weight drop probability; The sparsity measurement method used in this paper is based on the relationship between \( L_1 \) norm and \( L_2 \) norm [9]. The specific formula is:

\[
\text{sparseness}(x) = \frac{\sqrt{n - (\sum_i |x_i|)} / \sqrt{\sum_i x_i^2}}{\sqrt{n - 1}}
\]
Where, \( x \) is the input feature vector, \( n \) is the number of elements in the vector \( x \). After normalization processing, the output value of the sparsity formula is in the range \([0, 1.0]\).

2.2 CNN optimization.

The deep Convolutional Neural Network with improved structure is shown in Figure 2.

![Figure 2. Structure diagram of deep Convolutional Neural Network](image)

In the network, the settings of COVN layer, Pooling layer, and FC layer are shown in Table 1. The last layer uses the Softmax function to output the probability of the category that each dimension image belongs to. After addition of Sparse DropConnect to the FC layer, the initial learning rate is 0.001, the momentum is 0.9, the weight attenuation is 0.0005, and the number of iterations is 500 [10].

| Layer No. | COVN layer | Pooling layer | FC layer |
|-----------|------------|---------------|----------|
| 1         | 64×11×11   | 2×2           | 4096     |
|           | stride1 pad0 | maxpool       | softmax:5 |
| 2         | 192×3×3    | 2×2           | 4096     |
|           | stride2 pad2 | maxpool       | softmax:2 |
| 3         | 384×3×3    |               |          |
| 4         | 256×3×3    |               |          |
| 5         | 256×3×3    | 2×2           | 4096     |
|           | stride1 pad0 | maxpool       |          |

3. Numerical experiments

To verify the algorithm proposed in this paper, X-ray image data of steel pipe welds taken in industrial production are selected for the test to effectively identify and classify weld defects images. The size of the image is randomly selected. To maximum retention the characteristic information, when the image is processed, the longer side of the length and width is converted to 256 pixels, and the other side is subjected to the equal proportion conversion [11], which results in the other side being less than 256 pixels. For the shorter side, the original image is centered, and both sides are overturned to a size of 256 pixels. On the one hand, the method ensures that the processed image is still 256 × 256 in size, and on the other hand, it will not cause the loss of welds defects information in the image. The specific transformation algorithm is shown below:

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 \( 256/ma \), then step 4; or Step 3:

Step3: Overturn \( ma \), until 256;
Step4: Overturn mi, until 256;  
Step5: Output outImage

The experiment uses the CNN structures and parameters proposed in Section 2.2, and uses Softmax to connect and implement image classification. To better verify the fitting, the ratio of the training sample to the testing sample in the original defect sample data is set to 1:1, 500 samples respectively.

To study the influence of sparsity on CNN, Sparse DropConnect and DropConnect after the FC layer of CNN were respectively used to perform image recognition performance experiments. The results are shown in the figure 3. The figure shows the experimental results of the variation of recognition performance with sparsity. From the figure 3, it can be seen that with the change to sparsity, CNN, to which DropConnect was added, experience large fluctuations in recognition performance, while Sparse DropConnect is less affected by sparsity. This is because sparse action in Sparse DropConnect is also in fact a kind of regularization operation. It has a similar effect in preventing model over-matching and therefore has less fluctuation.

Figure 3. Relationship between sparsity and error rate

In the image recognition fitting experiment, comparison experiments were performed using the Sparse DropConnect, DropConnect, DropOut, and non-regularization measures after the FC layer of CNN. The experimental results are shown in Figure 4. As seen from Figure 4, models that did not use the regularization method soon experienced overfitting during the training process. While the models using the Sparse DropConnect, DropConnect, and DropOut methods slowed down, they eventually achieved better recognition results. The rate of convergence of Sparse DropConnect model is slower than DropOut, slightly faster than DropConnect, and has the lowest error rate of the test set, avoiding overfitting very well.
The main types of weld defects included: porosity (1), cracks (2), slag inclusions (3), and non-fusion (4), incomplete penetration (5). In the classification and recognition experiment for weld defects, comparison experiments were performed using the Sparse DropConnect, DropConnect, Dropout, and non-regularization measures respectively. The experimental results are shown in Table 2, the results show that: CNN under Sparse DropConnect has certain advantages, mainly the accuracy of the two training and test samples higher than that of other methods and the data of two samples tending to be consistent, showing a better goodness of fit.

Table 2. Weld defects classification results

| Defect type | Testing sample recognition rate(%) | Training sample recognition rate(%) |
|-------------|-----------------------------------|-----------------------------------|
|             | No-Drop | Dropout | Drop Connect | Sparse DropConnect | No-Drop | Dropout | Drop Connect | Sparse DropConnect |
| 1           | 89.45   | 92.36   | 91.43        | 96.77              | 88.45   | 91.45   | 92.22        | 96.98              |
| 2           | 89.44   | 92.45   | 92.16        | 95.26              | 90.44   | 93.89   | 95.29        | 95.86              |
| 3           | 90.78   | 92.78   | 91.78        | 95.73              | 93.29   | 94.46   | 89.19        | 96.03              |
| 4           | 90.86   | 91.89   | 89.86        | 93.00              | 92.57   | 95.19   | 92.49        | 93.16              |
| 5           | 90.28   | 91.59   | 92.36        | 92.86              | 91.45   | 95.28   | 95.22        | 93.10              |

Compared with other methods, the CNN network under Sparse DropConnect in the test has higher recognition and classification of the weld defects images. The first reason is that the deep network structure of CNN determines that the transmission of data information is made on the bottom-to-top and layer-by-layer basis, retaining important information as much as possible and representing it in multiple levels, so it has a better recognition accuracy than the shallow network; secondly, the introduction of sparse DropConnect in the CNN network can obtain sparse representation of the target, reduce information processing amount, and simplify the learning and training process, greatly improving the accuracy and adaptability of defect feature classification.

4. Conclusion

As a kind of deep learning method, CNN has its better image feature extraction and information expression capability. It expands and optimizes the the CNN structure by sparse DropConnect algorithm added after the FC, increasing the ability of network to select sparse features. The optimized CNN ensures higher recognition accuracy under the precondition of increase the sparsity of the output features. As the data set of the weld image collected by the industry used in this experiment is not big enough, further experimental studies are needed for more data sets under the CNN.
References

[1] Simard P Y, D Steinkraus, and J C Platt 2003 Best practices for convolutional neural networks applied to visual document analysis Proceedings IEEE International Conference on Document Analysis and Recognition 958

[2] Soatto S and Chiuso A 2016 Visual representations: defining properties and deep approximations Eprint Arxiv 2018 161-9

[3] Ioffe S and Szegedy C 2015 Batch normalization: accelerating deep network training by reducing internal covariate shift International Conference on International Conference on Machine Learning 448-56

[4] Ren S, He K, Girshick R and Sun J 2017 Faster r-cnn: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39(6) 1137-49.

[5] Redmon J, Divvala S, Girshick R and Farhadi A 2016 You Only Look Once: Unified, Real-Time Object Detection. IEEE Conference on Computer Vision and Pattern Recognition 779-88

[6] Gong Y, Wang L, Guo R and Lazebnik S 2014 Multi-scale orderless pooling of deep convolutional activation features European Conference on Computer Vision 8695 392-407.

[7] Pennec X 2008 Statistical computing on manifolds: from riemannian geometry to computational anatomy. Lecture Notes in Computer Science 5416 347-86.

[8] Goodfellow I J, Bulatov Y, Ibarz J, Arnoud S and Shet V 2013 Multi-digit number recognition from street view imagery using deep convolutional neural networks Computer Science 29

[9] Zha S, Luister F, Andrews W, Srivastava N and Salakhutdinov R 2015 Exploiting image-trained cnn architectures for unconstrained video classification Computer Science 106

[10] Zhang N, Paluri M, Ranzato M, Darrell T and Bourdev L 2014 PANDA: Pose Aligned Networks for Deep Attribute Modeling. Computer Vision and Pattern Recognition 1637-44

[11] Mendel K R, Li H, Sheth D and Giger M L 2018 Transfer learning with convolutional neural networks for lesion classification on clinical breast tomosynthesis Computer-Aided Diagnosis 28