

Research on image recognition method based on LMAL and VGG-16

Ying Cao^{a,*}, Runlong Gu^b, Chenghua Huang^c

^aGansu Vocational College of Architecture, Lanzhou 730050, China; ^bLanzhou Resources & Environment Voc-Tech University, Lanzhou 730050, China; ^cLanzhou Vocational Technical College, Lanzhou 730070, China

ABSTRACT

In deep convolutional neural networks, the traditional Softmax loss function lacks the ability to distinguish similar classes. In order to solve this problem, the idea of increasing the inter-class spacing and reducing the intra-class spacing is widely recognized. The Large Margin Angular Loss (LMAL) loss function is introduced to reduce the intra-class spacing by L2-standardization of the features and weight vectors of the Softmax loss function. At the same time, LMAL also has a good ability to distinguish deeper features. Combined the LMAL loss function and the VGG-16 model, the results on three independent data sets show that the image recognition accuracy of the improved model has been significantly improved.

Keywords: Convolutional neural network, image recognition, LMAL, VGG-16

1. INTRODUCTION

In recent years, with the rise of the field of artificial intelligence, the research of deep learning in academia and industry has developed rapidly, and the research of related computer vision technology has also made great progress. Starting from the Perceptron¹ proposed by Rosenblatt as the first-generation computer vision system, Fukushima et al. were inspired by the connection pattern of the cat's brain nerves and proposed Neocognitron². In subsequent studies, researchers will propagate backwards. The algorithm³ combined with Neocognitron finally proposed the concept of Convolutional Neural Network (CNN)⁴⁻⁵. Convolutional neural network (CNN) has been proven to be an effective model for solving various visual tasks⁶⁻⁸. It implements the function of a feature extractor through the interleaved stack of convolutional layers and a series of nonlinear and sub-sampling layers. This makes CNN a powerful network for describing images. Compared with other neural networks, the convolutional neural network greatly reduces the number of connections and weights between neurons in each layer due to the inspiration of the visual circuit structure of the brain nerves, which reduces the training difficulty and computing time of the model, and also reduces the probability of overfitting.

The most direct way to improve the performance of a deep neural network is to increase the size of the network, which includes increasing the depth of the network and the number of units in each layer. Especially when considering the availability of large amounts of labeled training data, this is a simple and safe way to train higher quality models. However, this simple solution has two main disadvantages. One is that it will increase the probability of overfitting the model, which may be a big bottleneck when there are more and more data; the other is that it will increase computing resources⁹. Therefore, this paper adopts the VGG16¹⁰ model proposed by Simonyan et al. as the experimental model. 16 represents the number of network layers, including convolutional layer, fully connected layer and Softmax layer.

The function of the loss function is to reflect the difference between the predicted data and the actual data. It is a way to measure the performance of the model. At the same time, the loss function is also a key to improving the accuracy of image recognition and classification. While researchers continue to propose new models, the loss function is also constantly evolving, such as Softmax Loss¹¹, Contrastive Loss¹², Triplet Loss¹³, SphereFace¹⁴, InsightFace¹⁵, etc. This article will show the common loss functions at this stage and analyze their advantages. Inferiority, the loss function is finally combined with the model, and the accuracy of the model is analyzed through the test of a large number of data sets.

The second chapter of this article introduces three neural network models, the third chapter compares and analyzes the

*gsjyxxzxcy@163.com

advantages and disadvantages of the three loss functions, and the fourth chapter is the part of experimental analysis and discussion.

2. CONVOLUTIONAL NEURAL NETWORKS

The VGG-16 model¹⁰ is a deep convolutional neural network jointly developed by the University of Oxford’s Visual Geometry Group and Google’s DeepMind Department, as shown in Figure 1. The depth of the VGG network is determined to be 16-19 layers after many studies. This article uses a 16-layer structure. Compared with the previous network structure with excellent performance, the VGG network has a significant drop in test error rate, and won the championship of the positioning project and the runner-up of the classification project in the ILSVRC-2014 competition. Therefore, the VGG network is widely used for image recognition and classification tasks.

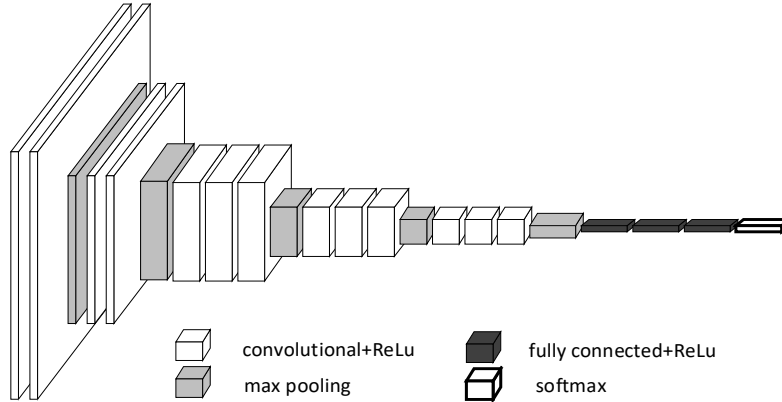


Figure 1. VGG-16 architecture.

In the VGG-16 model, a 3*3 convolution kernel and a 2*2 pooling kernel are used, which is also a major feature of the excellent performance of the model. It can be seen from the figure that the first 13 layers of the VGG-16 model are a stack of convolutional layers, the last three layers are fully connected layers, and the last is the Softmax layer.

3. LOSS FUNCTION

As mentioned in this paper, the loss function is the reflection of the fitting degree of the model to the data. The better the fitting degree, the smaller the value of the loss function, and vice versa. Loss function also plays a great role in depth feature learning, so the selection of loss function is also very important for image recognition and classification.

3.1. Large margin cosine loss

Liu¹⁶ introduces angle margin. When Softmax classifies, there will be an obvious decision boundary. The points near the decision boundary will reduce the generalization ability and robustness of the model. Therefore Wang et al.¹⁷ proposed a Large Margin Cosine Loss (LMCL) loss function.

$$L_{LMCL} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i})-m)}}{e^{s(\cos(\theta_{y_i})-m)} + \sum_{j \neq y_i} e^{s \cos(\theta_j)}} \quad (1)$$

$$W = \frac{W^*}{\|W^*\|}, \quad X = \frac{X^*}{\|X^*\|} \quad (2)$$

$$\cos(\theta_j) = W_j^T X_i \quad (3)$$

where N represents the number of training samples, X_i represents the i -th eigenvector corresponding to class y_i , W_j

represents the weight vector of class j , θ_j is the angle between W_j and x_i , m represents the cosine margin, $s = \|x\|$.

The proposed LMCL loss function greatly improves the performance of the model, eliminates the need to set the tricky super parameter in the implementation, and makes it easier to fit without Softmax supervision.

3.2. Large margin angular loss

Cosine margin is characterized by one-to-one mapping from cosine space to angle space, but the margin of cosine space is different from that of angle space. In fact, the geometric representation of angle margin is clearer than cosine margin, and the boundary in angle space corresponds to the arc distance on hyperspherical aggregate, as shown in Figure 2. Take the binary classification problem as an example, Figure 2 shows the intuitive correspondence between the angle margin and the arc edge of the hypersphere¹⁵.

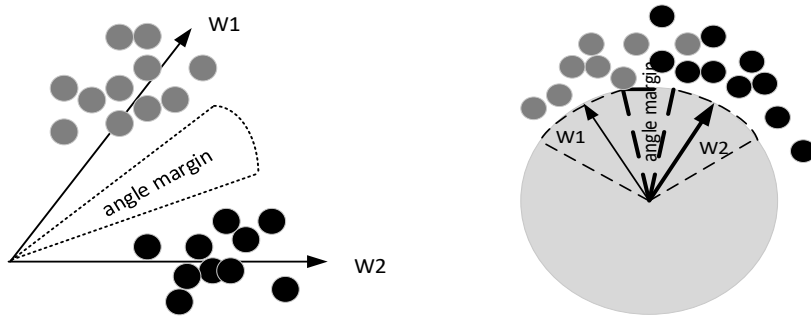


Figure 2. Margin interpretation of LMAL diagram.

Therefore, on the basis of LMCL¹⁵, a Large Margin Angular Loss (LMAL) loss function is proposed:

$$L_{LMAL} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j \neq y_i} e^{s \cos \theta_j}} \quad (4)$$

$$\theta \in [0, \pi - m] \quad (5)$$

The meaning of variables in the formula is the same as that in equation (1). One of the advantages of LMAL is its clear geometric interpretation, as shown in Figure 3.

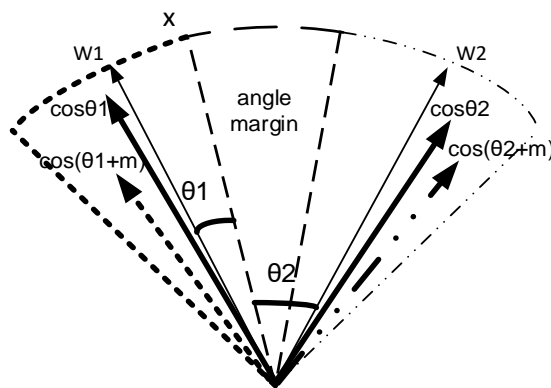


Figure 3. Geometrical interpretation of LMAL diagram.

4. EXPERIMENTAL ANALYSIS AND DISCUSSION

As mentioned above, vgg16 model and LMAL loss function are selected for training on the data set Imagenet. Through the target logit curves of the three loss functions, as shown in Figure 4, we can see that when $\theta \in [30^\circ, 90^\circ]$. The target logit curve of LMAL is lower than that of LMCL. Therefore, according to equation (4), LMAL has a stricter margin in this interval than LMCL. According to the experimental analysis, the model performance can not be significantly improved when $\theta < 30^\circ$. According to Deng et al.¹⁵, when $\theta \in [60^\circ, 90^\circ]$, the training will not converge, while when $\theta \in [30^\circ, 60^\circ]$, the model performance can be effectively improved.

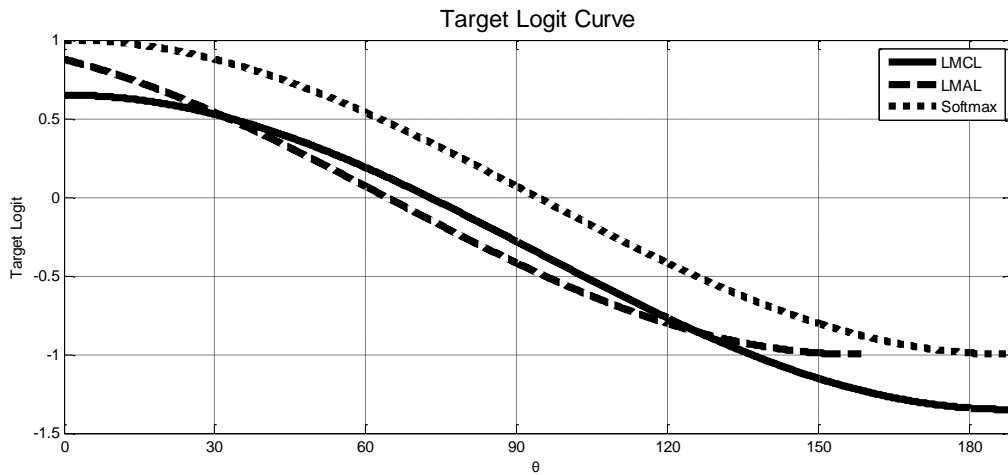












Figure 4. Target logit curve.

The model is tested with Pascal visual object classes 2007 (voc-2007), Pascal visual object classes 2012 (voc-2012). Voc-2007 includes 20 categories, with a total of 9963 drawings; Voc-2012 includes 20 object classes and 10 action classes, with a total of 17125 drawings. 10 samples are selected from three data sets respectively, as shown in Table 1.

Table 1. VOC-2012 top-1 test sample.

Image	Test result	Accuracy (%)	Image	Test result	Accuracy (%)
	Mountain bike	0.61760551		Persian cat	0.65801388
	Ocean liner	0.99372792		Racing car	0.87700474
	Bullet train	0.99832124		Airliner	0.45872828
	French bulldog	0.99862778		Yawl	0.85090107
	Motor scooter	0.43986964		Ski	0.92091626

Through the test of three data sets, the accuracy of vgg-16 model based on angle margin loss is shown in Table 2. It can be seen from Table 2 that among the test accuracy of the three data sets, LMAL has significantly improved compared with softmax, but it also reflects that the image recognition rate will be greatly affected when the image background, illumination and shooting angle are different, which is also the deficiency of this paper and one of the improvement directions in the future.

Table 2. Independent datasets' top-1 accuracy (%).

Model \ Test dataset	VOC-2007	VOC-2012	Caltech-256
VGG-16 with LMAL	79.15	72.48	82.27
VGG-16 with Softmax	75.27	70.59	77.07

5. CONCLUSION

This paper analyzes three common loss functions and models in image recognition, and proposes to apply LMAL loss function to image recognition and classification model VGG-16, which optimizes the potential problem of insufficient ability to reduce intra class gap when using softmax loss function. Through the integration of the improved loss function and the model, good results have been achieved in the experiments on the independent data sets voc-2007, voc-2012 and caltech-256, and the performance of the model has been improved to a certain extent. The future improvement direction of this paper is to solve the recognition and classification ability when the picture background is complex.

REFERENCES

- [1] Rosenblatt, F., [The Perceptron: A Perceiving and Recognizing Automaton], Technical Report, 85-460-1 (1957).
- [2] Fukushima, K., "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, 36(4), 193-202 (1980).
- [3] Rumelhart, D. E., Hinton, G. E. and Williams, R. J., "Learning representations by back-propagating errors," *Nature*, 323(6088), 533-536 (1986).
- [4] LeCun, Y., Boser, B. and Denker, J. S., "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, 1(4), 541-551 (1989).
- [5] Lang, K. J., Waibel, A. H. and Hinton, G. E., "A time delay neural network architecture for speech recognition," *Neural Networks*, 3(1), 23-43 (1988).
- [6] Krizhevsky, A., Sutskever, I. and Hinton, G. E., "ImageNet classification with deep convolutional neural networks," *NIPS*, (2012).
- [7] Ren, S., He, K., Girshick, R. and Sun, J., "Faster R-CNN: Towards real-time object detection with region proposal networks," *NIPS*, (2015).
- [8] Toshev, A. and Szegedy, C., "DeepPose: Human pose estimation via deep neural networks," *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 1653-1660 (2014).
- [9] Szegedy, C., Liu, W., Jia, Y. Q., et al., "Going deeper with convolutions," *arXiv:1409:4842v1*, (2014).
- [10] Simonyan, K. and Zisserman, A., "Very deep convolutional networks," *arXiv preprint arXiv:1409.1556*, (2015).
- [11] Cao, Q., Shen, L., Xie, W., Parkhi, O. M. and Zisserman, A., "Vggface2: A dataset for recognising faces across pose and age," *arXiv:1710.08092*, (2017).
- [12] Sun, Y., Chen, Y., Wang, X. and Tang, X., "Deep learning face representation by joint identification-verification," *Proc. of 27th Inter. Conf. on Advances in Neural Information Processing Systems*, 1988-1996 (2014).
- [13] Schroff, F., Kalenichenko, D. and Philbin, J., "Facenet: A unified embedding for face recognition and clustering," *CVPR*, 815-823 (2015).
- [14] Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B. and Song, L., "Sphereface: Deep hypersphere embedding for face recognition," *CVPR*, 212-220 (2017).
- [15] Deng, J. K., Guo, J. and Zafeiriou, S., "Additive angular margin loss for deep face recognition," *arXiv:1801.07698*, (2018).
- [16] Liu, W., Wen, Y., Yu, Z. and Yang, M., "Large-margin softmax loss for convolutional neural networks," *ICML*, 507-516 (2016).
- [17] Wang, F., Cheng, J., Liu, W. and Liu, H., "Additive margin softmax for face verification," *arXiv:1801.0559*, (2018).