CNN Based Image Classification

Xiaoling Long
SIST
ShanghaiTech Univ.
longxl@shanghaitech.edu.cn

Second Author
SIST
ShanghaiTech Univ.
secondauthor@i2.org

Second Author
SIST
ShanghaiTech Univ.
secondauthor@i2.org

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

Please follow the steps outlined below when submitting your manuscript to the IEEE Computer Society Press. This style guide now has several important modifications (for example, you are no longer warned against the use of sticky tape to attach your artwork to the paper), so all authors should read this new version.

2. CNN Part

3. CNN Based Image classification

The ImageNet Large Scale Visual Recognition Challenge(ILSRC) is a benchmark in object category classification and detection on 1000-classes and millions of images. After AlexNet achieved hugu success in ILSRC-2012, there are various variations of AlexNet [4] and many other types of ConvNet for image classification. Since that, ConvNet is widely used for image classification. [1] illustrate beriefly what is ConvNet, the components of ConvNet, the activation function in ConvNet, from LeNet to ResNet bunch of successful ConvNet and some open issues on CNN based image classification.

AlexNet [4] brought Convolutional neural network into ILSRC. In this implementation, it contains 8 layers 5 convolutional and 3 fully-conneted. The main features of this network's architecture are ReLU [6] as activation function, overlapping pooling and skills for reducing overfitting. Based on ReLU and overlapping pooling, the network err rate has more or less reduction, and ReLU network learn several times faster than other saturating acitvation function such as tanh neurons network. Overfitting is common issue for machine learning, it uses data augmentation and dropout to avoid overfit. Dropout is a skill to reduce argument or increase hypothesis. There are many discuss about this. As to data augmentation, it enlarge the dataset by cropping 224×224 patched from the original image as well as these patchs' horizontal refliction. At the end averaging the predications as final socre.

[2] explored the generalization ability of ConvNet features, releasing DeCAF. Traditional image classifition pipeline is extracting feature, building bag of feature then put into classifier. [13] propose a CNN based feature extractor. This is an unsurpervising learning ConvNet or in other words, input image is also the kind of ground truth. After feature extracting, the final result can classify by any classifier. This strategy is used as a tool for visualizing and undestanding how ConvNet works [12].ConvNet have an impressive classification performance. However there is no clear understanding for why this work. [12] propose a architecture for visualizing and undestanding how ConvNet works.

OverFeat [7] is a integrated recognition, localization and detection. This network uses CNN extract feature from image and then perform classification and localization and detection. Multi-scale classification brought up in [7] to increase accuracy.

"Networ In Network" [5] propsed a new deep network structure. Different from conventional convolution layer, it brings up a new Mlpconv layer. This Mlpconv layer consist of sliding multilayer perceptron(MLP) window. In stead of fully-connected layer at the top of network, global average pooling is used to produce the resulting vector fed directly into the softmax layer. Verified by experiments, this NIN structure indeed works well on some benchmark datasets, and global average pooling can be regarded as regularizer. THis glolbal average pooling has no parameter. This stratety is used widely afterwards. 1×1 convolution conception proposed in [5] is used in GoogLeNet for dimension reduction.

AlexNet make a great success in image classification. Afterwards many various Network appear. GoogLeNet [11] proposed by google is a new level of oganization in the form of the " Inception module". This is a multi-scale arcthitecture. With the limitation of computational resource, it perform a 1×1 convolution to dimension reduction. Auxiliary classifier is also a brilliant strategy.This smart design makes a great success in ILSRC-2014. At the same time, the widely used ConvNet architecture VGGNet [8] won the first place in *Classification* + *Localization competition*. It adds the number of layers up to $16 - 19$. Instead of 7×7 convolution filter in [8], it uses 3×3 as convolution filter. After multiple layers, it can get similar effect as 7×7 one. This design significantly reduces the parameters, and then reduces the overfitting. It also means the number of layers significantly increases. Altering convolutional layers and poolint layer became a common used Network architecture.

As the depth of ConvNet increasing, training gets more and more difficult. The training of very deep network becomes a open issue in CNN. Highway Networks [9, 10] propose *information highways* which allow unimpeded information flow across several layers. The *transform gate* $T(x, W_T)$ and the *carry gate* $C(x, W_C)$ proposed for decided how much flow pass through to output. The new model given by

$$y_{output} = H(x, W_H) \cdot T(x, W_T) + x \cdot C(x, W_C) \quad (1)$$

For simplicty, [3] set $C = 1 - T$. This design make training hundreds of layers be possible and the err rate just has slightly increase. This architecture promote the success of ResNet [3]. ResNet has similar structure as deep plain network stacked by dozens of convolution layers followed by global pooling layer and 1 fully-connected layer. except shortcup connection. This design has a residual representation which called deep residual learning. This archicture keeps parameter less than VGG-19 model even the network has 152 layers. This smart design make ResNet won first place in ILSRC-15.

4. Application

5. conclusion

References

[1] N. Aloysius and M. Geetha. A review on deep convolutional neural networks. In *Communication and Signal Processing (ICCSP), 2017 International Conference on*, pages 0588–0592. IEEE, 2017.

[2] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In *International conference on machine learning*, pages 647–655, 2014.

[3] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[5] M. Lin, Q. Chen, and S. Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.

[6] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

[7] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*, 2013.

[8] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[9] R. K. Srivastava, K. Greff, and J. Schmidhuber. Highway networks. *arXiv preprint arXiv:1505.00387*, 2015.

[10] R. K. Srivastava, K. Greff, and J. Schmidhuber. Training very deep networks. In *Advances in neural information processing systems*, pages 2377–2385, 2015.

[11] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, et al. Going deeper with convolutions. Cvpr, 2015.

[12] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.

[13] M. D. Zeiler, G. W. Taylor, and R. Fergus. Adaptive deconvolutional networks for mid and high level feature learning. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2018–2025. IEEE, 2011.