Research on Data Augmentation for Image Classification Based on Convolution Neural Networks

Jia Shijie
Electronic&information college
Dalian Jiaotong University
Dalian, China
jsj@djtu.edu.cn

Wang Ping
Electronic&information college
Dalian Jiaotong University
Dalian, China
jsj@djtu.edu.cn

Jia Peiyi
School of Mechatronics& Information Engineering
Shandong University
Weihai, China
pp_luck@163.com

Hu Siping
Electronic&information college
Dalian Jiaotong University
Dalian, China
jsj@djtu.edu.cn

Abstract*—The performance of deep convolution neural networks will be further enhanced with the expansion of the training data set. For the image classification tasks, it is necessary to expand the insufficient training image samples through various data augmentation methods.This paper explores the impact of various data augmentation methods on image classification tasks with deep convolution Neural network, in which Alexnet is employed as the pre-training network model and a subset of CIFAR10 and ImageNet (10 categories) are selected as the original data set. The data augmentation methods used in this paper include: GAN/WGAN, Flipping, Cropping, Shifting, PCA jittering, Color jittering, Noise, Rotation, and some combinations. Experimental results show that, under the same condition of multiple increasing, the performance evaluation on small-scale data sets is more obvious, the four individual methods (Cropping, Flipping, WGAN, Rotation) perform generally better than others, and some appropriate combination methods are slightly more effective than the individuals.

***Keywords*—Image Classification; Data Augmentation; Convolution Neural network**

I. INTRODUCTION

In recent years, deep convolution neural network[1]has made a great breakthrough on image classification tasks[2-5]. However, it requires a large amount of tagged data to train the deep convolution models to avoid overfitting[6], which is hard to meet in practical applications. In the case of insufficient training data, the regularization technologies are commonly used to prevent overfitting, such as Dropout [7], BN (Batch normalization) [8]. Data augmentation, which refers to the process of creating new similar samples to the training set, can be regarded as one kind of regularization technology[6]. For the tasks of image classification, data augmentation are commonly used in the pioneer works. For example, the famous AlexNet [2] employed random crop, horizontal flip, and PCA jitting.

In the recent days, Joseph Lemley, etc.al. proposed a smart augmentation method[6], which works by creating a network that learns how to generate augmented data during the training process of a target network in a way that reduces that networks loss; VGG[4] and ResNet[5] employed scale jittering, while GooLeNet[3] employed scale and aspect ratio augmentation transformation. Data Augmentation methods have been widely used in deep learning, and selection of appropriate data augmentation strategies is even more important than choosing a network structure [9]. However, for data augmentation techniques, the lack of necessary research has long remained in the intuition and experience stage, there is no one general choice strategy in applications.

This paper attempts to explore the data augmentation techniques for image classification tasks with deep convolution neural network. The main concerns are as follows:(1) What are the differences of the impacts of different data enhancement methods on classification performances?(2) For different scales of training set, what are the differences of the impacts of data enhancement techniques on classification performances? (3) What are the differences of the impacts on the promotion of the classification performance between any single type and combinations in the case of the same expansion of the data volume?

Based on the above problem, this paper employes Alexnet as the pre-training network model and selectes a subset of CIFAR10 and ImageNet (10 categories) as the original data set. The training set is grouped into different scales (small , medium and large), the data augmentation methods include: GAN/WGAN, Flipping, Cropping, Shifting, PCA jittering, Color jittering, Noise, Rotation,and some combinations.

The structure of this paper is as follows: Part 1describes the various data enhancement methods. Part 2 gives a biref introduction to the deep convolution neural network, the third part is the experiments and results analysis, the fourth part gives a brief summarization.

II. Data augmentation method

A. Unsupervised data augmentation

The so-called unsupervised data augmentation means that the augmentation methods are not related to data labels[8]. For image classification tasks, some categroy-free image transformation methods are employed to generate new samples from the training set. The commonly-used image transformation methods are listed blow:

1) Flipping. Flip the image in the horizontal direction.
2) Rotation. Rotate the image at random orientation.
3) Cropping. Crop a part from the original image and resize the cropped image to the specific resolution(if neccessary).

This work is supported by Liaoning Provincial Natural Science Foundation(201602118).

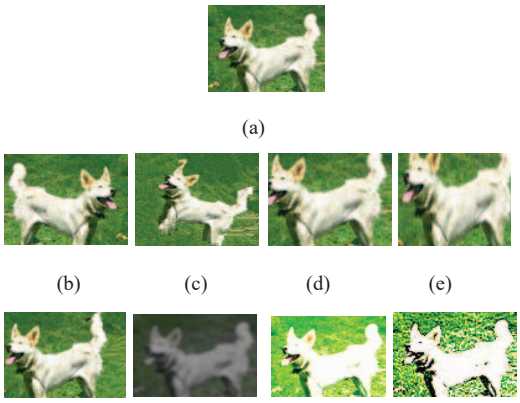
4) Shifting. The image is shifted to the left or right, and the translation range and step length can be specified manually to change the location of the image content.

5) Color jittering. Change the random factors of color saturation, brightness and contrast in the image color space.

6) Noise. Add random perturbation(noise) to RGB channels of each pixel in the image. The commonly-used noise is gaussian noise.

7)PCA jittering[2]. Perform PCA on the image to get the principal component, which is then added to the original image with a gaussian disturbance of (0, 0.1) to generate the new image.

The unsupervised data augmentation methods described above are shown in Fig 1.



(a)

(b) (c) (d) (e)

(f) (g) (h) (i)

(a) original image (b) flipping (c) rotation (d)cropping (e)random-cropping (f)shifting (g)noise (h)color-jittering (i)PCA-jittering

Fig1. Original and transformed images

B. *Supervised Data Augmentation*

The so-called supervised data augmentation means that the augmentation methods is related to the data labels[8]. For the image classification tasks, each augmentation image sample is generated with the specific category in the training set. GAN(Generative Adversarial Networks) and its improved methods can be categorized into the supervised methods.

The GAN model is composed of a generative model G and a discriminative model D. In the training process, G is taught to map from a latent space to a particular data distribution of interest, and D is simultaneously taught to discriminate between instances from the true data distribution and synthesized instances produced by G.The objective function is:

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1-D(G(z)))](\quad (1)$$

Where x denotes the real image, z denotes the noise of the input G network, $G(z)$ denotes the image generated by the G network, $D(x)$ and $D(G(z))$ denotes the probability of x and $G(z)$ as a real image by D, respectively.

The original GAN model employs KL as the distance measurement, which results in unsteady gradient, and hard to generate diversity samples. WGAN (Wasserstein GAN)[15] introduce Wasserstein distance to solve the problem of training instability. The relationship of W distance, the number of iterations and the generated images is shown in Figure 2:

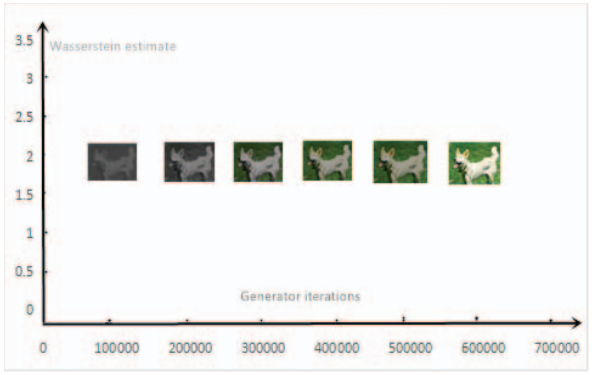


Figure2 W distance, the number of iterations and the generated images

III. Deep convolution neural network

CNN (Convolutional Neural Network) is inspired by the biological natural visual cognitive mechanism, which is composed of the input layer, the convolutional layers, the pool layers and the full connection layers and the output layer. The characteristics of the convolution neural network are embodied in two aspects: (1) the connection between the neurons in the convolutional layer is non-fully connected; (2) the weights of the connections between certain neurons is shared in the same layer. The sparse connection and weight-sharing design reduces the complexity of the network.

In 1959, Hubel & Wiese [10] found that its unique network structure can effectively reduce the complexity of the feedback neural network when studying the local sensitive and directional selection of neurons in the cortical cortex. Inspired by this work, Kunihiko Fukushima made the predecessor of CNN in 1980-Neocognitron [11]. In the 1990s, LeCun [12] proposed a multi-layer artificial neural network (LeNet-5) to achieve handwritten digital classification. The breakthrough took place in 2012. Krizhevsky et.al proposed the CNN model-AlexNet to get the championship in the ILSVRC-2012 image classification competition, its top-5 test error rate got 15.3%, which promoted 40% than the traditional methods. On the basis of the AlexNet, some more complicate deep CNN models were proposed, such as ZFNet [13], VGGNet [5], GoogleNet [3] and ResNet [14].

In this paper, the classical CNN model Alexnet is used to study the effect of data augmentation on image classification tasks. The structure of AlexNet network is shown in Figure 3. Other than the input and output layers, Alexnet contains five convolutional layers, three pooling layers and the three fully-connected layers. The output of the last fully-connected layer is fed to a 1000-way softmax which produces a distribution over the 1000 class labels.

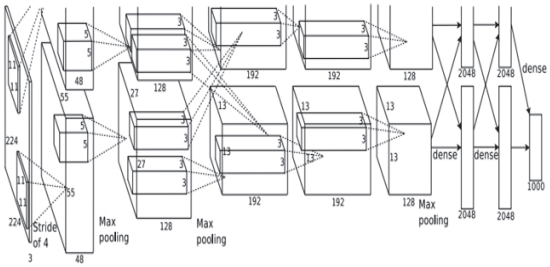


Figure 3 AlexNet network structure

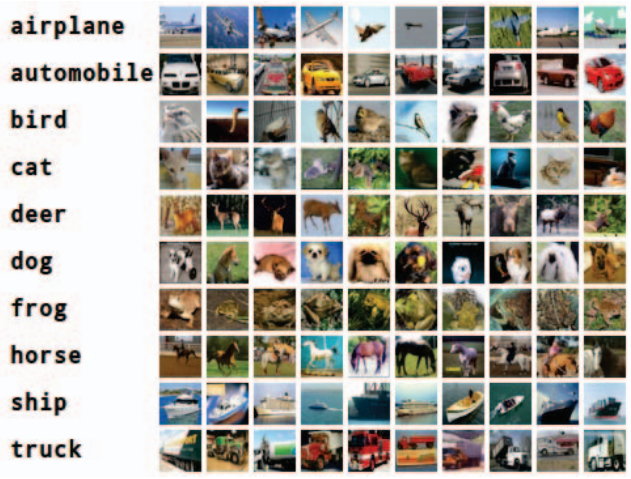
IV. Experiments and results analysis

A. Experimental setup

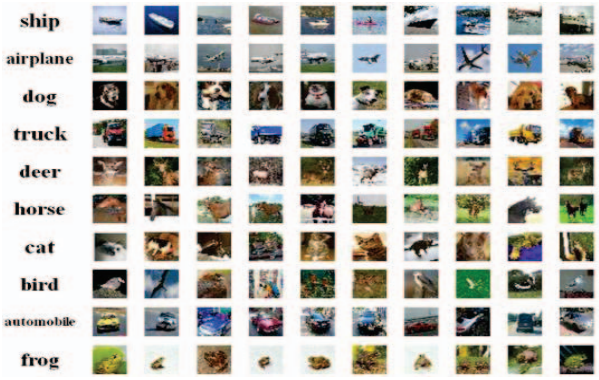
The main hardware and software used in this experiment are as follows:

- CPU: Intel (R) Core (TM) i7-6700, clocked at 3.4 GHz.
- Memory: 16GB.
- Operationg system: Linux UbunTu14.04.
- Development language: python3.5.
- Deep learning development platform: Tensorflow 1.0.

The two datasets used in the experiment are taken from a subset of CIFAR10 and ImageNet, respectively. The CIFAR10 dataset contains 6000 color images of 32*32 resolution, which are divided into 10 categories, including aircraft, car, bird, cat, deer, dog, frog, horse, ship, truck, etc. Corresponding to the 10 categories above, 6000 images are randomly selected from ImageNet and all images are resized to 224*224. The experimental image datasets are shown in Figure 4.



(a)CIFAR10



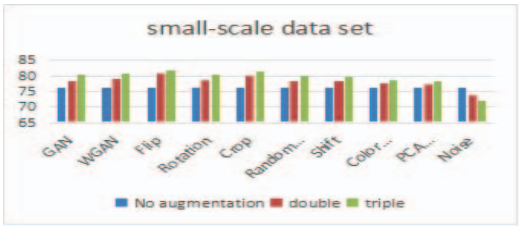
(b)Subset of ImageNet

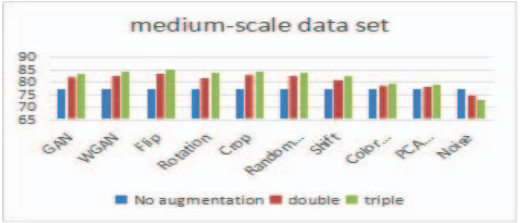
Figure 4 Samples of the experimental dataset

In this paper, all the experiments are carried out on the two dateset independently. Different scales of the training set would bring different effects with the same data enhancement method. In order to verify the effect of various data enhancement methods under different training data sets, three scales of training datasets are employed: ① small-scale training set: a total of 2000 training samples with 200 samples each category; ② medium-scale training set: a total of 10,000 training samples with 1000 samples each category; ③ large-scale data set: a total of 50000 training samples with 5000 each category. The test set comprises of a total of 10,000 images with 1000 images each category, which is not intersect with the training set.

B. Experimental content

Each augmentation method listed above is employed to generate new samples with the quantity of one or two times the original training set, respectively. Afer training with ①the original training set (No augment) ②the original training set plus the same size of the generated samples(Double) ③the original training set plus the double size of the generated samples(Triple). The test results are shown in Fig.5 and Fig.6.





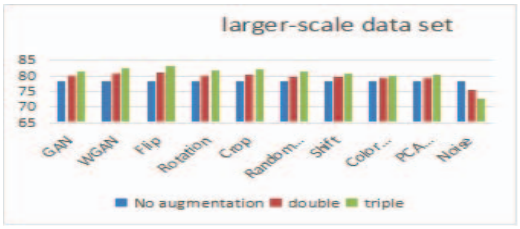


Figure 5 The test results on CIFAR10

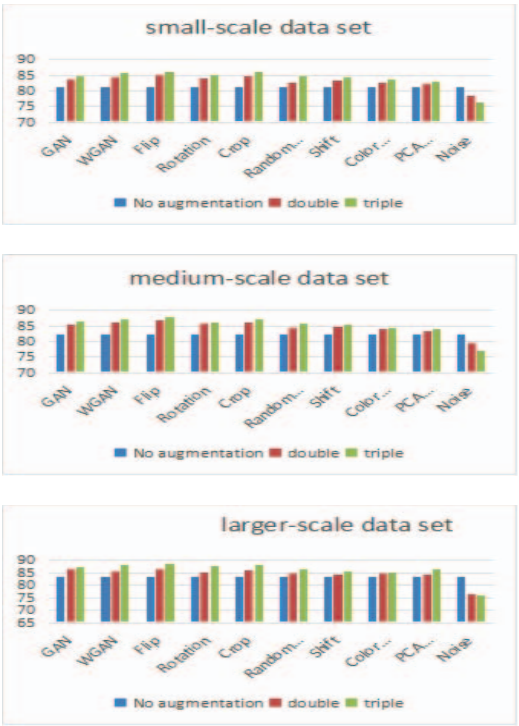


Figure 6 The test results on the subset of ImageNet

From the experimental results shown in Figure 5 and Figure 6, some conclusions can be listed blow:

1) Compared to the model trained with the unenhanced dataset, the models trained with enhanced training data set mostly perform better(except for adding noise), and the more augmentation samples added to the original training set, the higher classification accuracy the trained model achieves.

2) For the same enhencement method, the smaller the scale of the original training set, the better the enhancement effect is.

3) Compared to the other enhencement methods, WGAN, Cropping, Ratotion, Flipping are more effective. The remaining experiments will focus on the comparison of the four methods and their combinations.

Figures 7, 8 and Figures 9, 10 illustrate the test accuracies of six pair combinations and four triple combiantions of WGAN, Cropping, Ratotion, Flipping with different scales of traing set and different augmentation volumes(No augmentation, double, triple). The augmentation samples are evenly generated by each individual method.

Figure 7 Test results of six pair combinations on CIFAR10

Figure 8 Test results of six pair combinations on ImageNet subset

Figure 9 Test results of four triple combinations on CIFAR10

Figure 10 Test results of four triple combinations on ImageNet subset

From Figure 7-10, we can conclude that:

(1)On the condition of same volumes of augmentation, both the pair combination and the triple combination perform better than any individual method. For example, compaired to Flipping, WGAN, Cropping, the combination methods (Flipping +Cropping, Flipping+WGAN, WGAN+Cropping) improve 1.6%, 2%, 1.5% with triple augmentation on the small-scale CIFAR10 training set, respectivelty.

(2)Flipping+Cropping and Flipping+WGAN are the best pair combinations among the six ones, which improve 3% , 3.5% on CIFAR10 and 2% , 2.5% on ImageNet subset with triple augmentation on small-scale training set, respectivelty.

(3)The overall performance of triple combinations is prior to that of the pair combinations. However, some triple combinations may bring performance degradation. For example, for triple augmentation on small-scale CIFAR training set, compared to Flipping+Cropping, the test accuracy of Flipping+ Cropping+Rotation increases 0.9% while WGAN+ Flipping + Cropping decreases 1%.

V. Summary

This paper mainly discusses the data augmentation methods for image classification with deep convolution neural networks. Through the experiment results on CIFAR10 and ImageNet subset, this paper compares and analyzes the effects of various data augmentation methods and their combinations on different training scales. Subsequent research will further explore the effects of data augmentation in terms of large categories, complex network models and unbalanced training data.

References

[1] J. Lemley, S. Bazrafkan, and P. Corcoran, “Deep learning for consumer devices and services: Pushing the limits for machine learning, artificial intelligence, and computer vision.” IEEE Consumer Electronics Magazine, vol. 6, no. 2, pp. 48-56, 2017.

[2] Alex Krizhevsky,Ilya Sutskever,Geoffrey E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks ”, Advances in Neural Information Processing Systems 25, (NIPS 2012),pp.1-9.

[3] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich. “Going Deeper With Convolutions”. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1-9.

[4] Karen Simonyan, Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition” , Computer Science, 2014.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. “Deep Residual Learning for Image Recognition” The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.

[6] Joseph Lemley,Shabab Bazrafkan and Peter Corcoran. “Smart Augmentation Learning an Optimal Data Augmentation Strategy”, arXiv:1703.08383v1 [cs.AI] 24 Mar, 2017.

[7] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting.” Journal of Machine Learning Research, vol. 15, no. 1,pp. 1929-1958, 2014.

[8] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," Proceedings of Machine Learning Research. vol. 37, 2015. pp.:448-456.

[9] I.Goodfellow,Y.Bengio,and A.Courville, "Deep Learning. " MIT Press, 2016, http://www.deeplearningbook.org.

[10] D. H. Hubel, T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex", Journal of Physiology , vol 195,no 1, pp. 215–243, 1968.

[11] K. Fukushima, Neocognitron. "A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position", Biological Cybernetics, vol 36, no 4, pp.193-202, 1980.

[12] B. B. Le Cun, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel, "Handwritten digit recognition with a back-propagation network", Advances in Neural Information Processing Systems, pp. 396-404, 1990.

[13] M. D. Zeiler, R. Fergus, "Visualizing and understanding convolutional networks", European Conference on Computer Vision , pp. 818-833, 2014.

[14] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition". The IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778. 2015

[15] Martin Arjovsky, Soumith Chintala, and LeonBottou.Wasserstein GAN.arXiv:1701.07875v2 [stat.ML], 9 Mar 2017.