Proceedings of the Fifth IIEEJ International Workshop
on Image Electronics and Visual Computing 2017
Da Nang, Vietnam, February 28- March 3, 2017

DATA AUGMENTATION USING DCGAN FOR BREED IDENTIFICATION

Hiroki Watabe Hiroshi Watanabe
Graduate School of Fundamental Science and Engineering, Waseda University

ABSTRACT

In general, CNN designed for image recognition requires a lot of images at its training phase. However, it is sometimes difficult to provide enough number of images for CNN training. In this research, we apply DCGAN to CNN's data augmentation. Through the experiments, it is shown that the proposed data augmentation is more effective than the standard data augmentation. In our experiments, both realistic and unrealistic images are generated. We investigate the effect of the subjective quality of the generated images. The quality of generated images affects the performance of breed identification and it is better to use realistic images.

1. INTRODUCTION

Many people freely take pictures using smartphones and digital cameras. They upload pictures to SNS, such as Facebook and Twitter. Animal picture is one of them and people like taking their pet pictures. These pictures are tagged to be searched effectively on the Web. Breed information is important tag for pet pictures. Pet animals such as dogs and cats often have local minor differences between varieties. Therefore, breed identification is a kind of fine-grained recognition and it is sometimes more difficult than object recognition.

In recent years, Deep Learning has been focused on in various fields, such as image recognition and natural language process. In particular, Convolutional Neural Network (CNN) shows the high performance in image recognition field [1][2]. CNN's performance is affected by network structure, its parameters, and the number of training images. Data augmentation is a popular method to increase training data and improve CNN's performance [3]. In the field of automatic image generation, Deep Convolutional Generative Adversarial Network (DCGAN) is known to generate high quality images [4]. Coevolution between generator and discriminator leads to obtain the high-performance image generator.

In this paper, we propose the data augmentation which add similar images generated using DCGAN to training dataset. We evaluate this data augmentation technique to a small number of cat and dog dataset. In addition, we investigate the effect of generated images quality on breed identification.

2. DEEP LEARNING

2.1 CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Network (CNN) is the most powerful tool in image recognition tasks. CNN is composed of several layers, such as convolution, pooling and fully connected layers. CNN extracts the feature by repeating convolution and pooling alternately. This feature often works better than human-made features, such as SIFT and HOG [5]. By advancing the training steps, CNN makes the features to recognize images itself.

2.2 DATA AUGMENTATION

In many case, available images are insufficient to train CNN. Data augmentation is one popular method to make up for shortage of data. Data augmentation is executed by transforming available images using label-preserving transformations. For example, affine transformation (translation, rotation, inversion), brightness change, Gaussian noise addition, and so on. Data augmentation reduces over fitting and improves the accuracy and the robustness of a classifier [1].

2.3 DCGAN

Generative Adversarial Network (GAN) is a framework which consists of generative model (generator) and discriminative model (discriminator) [6]. Generator and discriminator play a minimax two-player game. The aim of generator is to capture the data distribution p_g which is similar to the training data distribution p_{data}. That of discriminator is to estimate the probability whether input image came from the training data. Generator transforms random noise vector z into an image $G(\boldsymbol{z})$. Discriminator calculates the probability $D(\boldsymbol{x})$ that input $\boldsymbol{x}$ came from the training data rather than generator. It means if input $\boldsymbol{x}$ came from the training data, $D(\boldsymbol{x}) = 1$, and else, $D(\boldsymbol{x}) = 0$. We train generator and discriminator by solving the objective function $V(G, D)$.

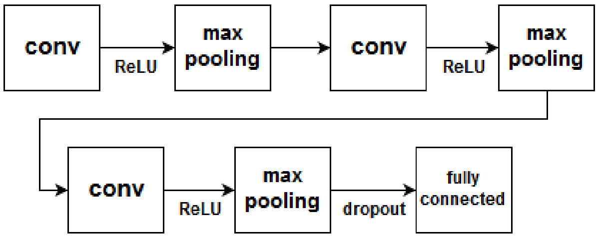

Figure 1: the architecture of our CNN for breed identification

$$\min_{G}\max_{D} V(G,D) = \mathbb{E}_{x\sim p_{data}}[\log D(\boldsymbol{x})] + \mathbb{E}_{z\sim noise}[\log(1 - D(\boldsymbol{G}(\boldsymbol{z})))] \quad (1)$$

After training, generator obtains a superior data distribution and becomes possible to generate images such that discriminator cannot judge whether it is real or not.

In the development of the GAN, scaled up using special CNN is called Deep Convolutional Generative Adversarial Network (DCGAN) [4]. CNN which is used in DCGAN is based on the following key ideas.

1. Use the all convolutional network, which is replaced pooling layers by convolutional layers with striding [7].

2. Remove fully connected hidden layers for deeper architectures.

3. Use Batch Normalization, which is to perform normalization for each batch so that the average becomes 0 and the variance becomes 1 [8].

4. Use ReLU activation in generator and Leaky ReLU activation in discriminator.

DCGAN can generate higher quality images than GAN by these ideas.

3. PROPOSED APPROACH

We apply DCGAN to CNN's data augmentation. We sometimes have to use a small number of dataset for CNN. With more images, CNN would show higher performance. Thus, we add similar images which are created by DCGAN to the CNN's training dataset as data augmentation. DCGAN's generator creates images which discriminator's CNN cannot distinguish whether they are real or not. Therefore, we presume generated images have similar CNN features and work as similar images. We consider generated images by DCGAN assist a small number of dataset. Proposed approach is performed in the following steps.

1. Train DCGAN for each class. For example, if we want to classify 10 classes, we train 10 DCGAN models.

2. Generate images by trained DCGAN models.

3. Add generated images to the original dataset.

This is the flow of our data augmentation (Proposal 1). In addition, we consider the combination of standard data augmentation and our proposed method (Proposal 2). In this paper, we use translation, rotation, inversion, brightness change, and Gaussian noise addition as standard data augmentation.

Proposal 1: add DCGAN's generated images to original dataset

Proposal 2: Proposal 1 × standard data augmentation

Table 1: result of experiments which we used original dataset or generated images as training dataset

	Accuracy (%)	
Dataset	**cat**	**dog**
Original	81.9	80.8
Generated	77.6	71.2

4. RESULTS AND DISCUSSION

In this paper, we proposed CNN's data augmentation using DCGAN. We evaluated the proposed method with cat and dog breed identification as a target. Cat and dog images were taken from The Oxford-IIIT-Pet dataset [9]. The number of training images is 100 per one class and that of test is the same. The cat and dog breeds are 12. Simple CNN architecture is used for pet's breed identification. It consists of three convolution layers and three max pooling layers and one fully connected layer (Figure 1).

4.1 GENERATED IMAGES

We trained DCGAN for each cat and dog breeds. Figure 2 shows the examples of generated images. In this experiment, many unrealistic images which were not able to be seen cat or dog were generated. The lack of training images for DCGAN caused this emergence of many unrealistic images.

Table 2: Result of comparison between the standard data augmentation methods and Proposal 1

method	Accuracy (%)
No DA	69.6
Translation	73.8
Rotation	73.7
Inversion	73.8
Brightness change	70.9
Gaussian noise Addition	69.8
Proposal 1	75.4

Before the evaluation of proposed method, we investigated whether generated images had features to help breed identification or not. We trained CNN using only DCGAN's generated images. We used 400 generated images per one class. Translation, rotation, inversion, brightness change, and Gaussian noise addition are applied as data augmentation. Table.1 shows the result. Both for cat and dog, they show over 70 % accuracy, which is close to the case using original data. Therefore, we were convinced that generated images had similar features to original images for CNN.

4.2 EVALUATION

First, we compared the standard data augmentation methods (translation, rotation, inversion, brightness change, and Gaussian noise addition) and proposed data augmentation to add generated images (Proposal 1) for cat breed identification. We added 400 generated images per one class in proposed method. Table 2 shows the result of the experiments. In the standard data augmentation method, translation and inversion which are affine transformations showed the greatest effect with 73.8%. Meanwhile, Proposal 1 showed 75.4% which was better than others.

Table 3: Result of evaluation experiment for our proposed methods

	Accuracy (%)	
method	cat	dog
Standard DA	81.9	80.8
Proposal 1	75.4	72.8
Proposal 2	82.7	82.1

Next, we evaluated the proposed method compared with the combination of the standard data augmentation method. Table 3 shows the result of the experiment. The combination of standard data augmentation method showed 81.9% for cat breed identification and 80.8% for dog breed identification. Proposal 2 showed 82.7% for cat and 82.1% for dog, which were the best accuracies of the three. It shows that our Proposal 2 is more effective than the standard data augmentation.

4.3 QUALITY OF GENERATED IMAGES

In our experiments, various quality images were generated. We picked up realistic and unrealistic images by our subjective evaluation. The criterion was whether we could see them as cat or dog. Then, we examined the contribution of the effectivity by realistic and unrealistic images. We performed the additional three experiments for cat and dog. (1) Add realistic 100 images per one

Figure 2: Examples of generated cat and dog images
The upper left side images are generated realistic cat images and the upper right side ones are generated unrealistic cat images and the downer left side ones are generated realistic dog images and the downer right side ones are generated unrealistic dog images.

Table 4: Result of experiments about the contribution of the effectivity by realistic and unrealistic images

	Accuracy (%)	
Quality	**cat**	**dog**
Realistic	83.2	81.8
Unrealistic	81.8	78.7
Random	82.3	81.2

class. (2) Add unrealistic 100 images per one class. (3) Add random quality 100 images per one class. Table.4 shows the results of these experiments. In both pet cases, adding realistic images showed higher performances than adding unrealistic images. Moreover, it also confirmed that using realistic images was better than using random quality images. This indicates that it is important for our method to use realistic images.

The influence of the generated dog images quality was greater than that of cat. We consider that this is due to the difference of face shape between cat and dog. Cat has uniform faces; most cats have standing ears on a round face. In contrast, dog has various face shapes, e.g. standing ears and dropping ears, long muzzle and short muzzle. It is believed that the face shape of dog is more important for breed identification than the cat's case. In unrealistic images, the face shapes often collapse. Thus, the generated images quality is greatly affected in the dog's case.

5. CONCLUSION

In this study, we applied DCGAN to CNN's data augmentation. Our approach demonstrated 82.7% accuracy for cat and 82.1% accuracy for dog, which were better than the standard data augmentation. Thus, we conclude that our method which is combination of the standard data augmentation and generated images addition is effective for CNN's pet identification. In our research, both realistic and unrealistic images are generated. We confirmed the subjective quality of generated images affected the performance of pet breed identification. It was also found that using realistic images was more effective than using unrealistic images. Therefore, further studies are needed in order to obtain realistic generated images. If we use more realistic images, results of our proposed method are still better.

6. REFFERENCE

[1] A. Krizhevsky, I. Sutskever, G. Hinton: "ImageNet classification with deep convolutional neural networks", In NIPS, pp. 1106-1114 (2012).

[2] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna: "Rethinking the inception architecture for computer vision." *arXiv preprint arXiv:1512.00567* (2015).

[3] D. Cire¸san, U. Meier, J. Schmidhuber: "Multi-column deep neural networks for image classification", IEEE Conference on Computer Vision and Pattern Recognition (2012).

[4] A. Radford, L. Metz, S. Chintala: "Unsupervised representation learning with deep convolutional generative adversarial networks", In 4th Int. Conf. on Learning Representations (2016).

[5] A. Sharif Razavian, H. Azizpour, J. Sullivan, S. Carlsson: "CNN features off-the-shelf: an astounding baseline for recognition.", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2014).

[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio: "Generative adversarial nets", In NIPS, p.p. 2672-2680 (2014).

[7] J.T. Springenberg, A. Dosovitskiy, T. Brox, M. Riedmiller: "Striving for simplicity: The all convolutional net", In ICLR (2015).

[8] S. loffe, C. Szegedy: "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", International Conference on Machine Learning Representations (2015).

[9] O. M. Parkhi, A. Vedaldi, A. Zisserman, C. V. Jawahar: "Cats and Dogs", IEEE Conference on Computer Vision and Pattern Recognition (2012).