

# Convolutional neural network for classifying cartoon images augmented by DCGAN

Yuzhi Hu

The Department of Software, Nanjing University, 22 Hankou Road, Nanjing, Jiangsu Province, 210008, China

15120140218@xs.hnit.edu.cn

**Abstract.** Convolutional Neural Network (CNN) tend to have better results on large data sets and poor performance on small data sets, so the data augmentation is crucial for a CNN to get better performance based on the dataset with limited size. In this paper, Deep Convolution Generative Adversarial Network (DCGAN) was used to augment data to make the AlexNet perform better on an image classification task with small data sets. AlexNet was trained on a small anime face training set with only 160 samples to determine whether the anime face was male or female, and then tested its accuracy on a test set with 240 sample. Then, a pre-trained DCGAN was transferred to train on the male and female training sets respectively. And 2 DCGANs were obtained, one could generate male cartoon faces and another could generate female cartoon faces. The images generated by DCGANs were put in train set, which was used to train AlexNet again and the result was recorded. Other data augmentation methods such as cutout, cutmix and Noise Injection were compared as well. Finally, it is found that AlexNet has the best performance when using the DCGAN augmentation method, which can significantly improve the verification accuracy of the model.

**Keywords:** Data Augmentation, CNN, DCGAN, AlexNet.

## 1. Introduction

Convolutional Neural Network (CNN) tend to have better results on large data sets and poor performance on small data sets, which has been proved in many tasks [1-4]. Only when feeding the model enough samples, then it will recognize sufficient features. However, large datasets are not available everywhere, there may be some situations where people can only obtain some small data sets e.g. some rare animal photos. At this time, how to use a small data set to train a CNN with superb performance is a challenging issue. Fortunately, data augmentation can expand the data set. It can generate new data from the original data then expand the size of the data set, so that models can have enough training data to improve their robustness.

In the early days, the operation of data augmentation was mainly based on geometric transformation methods e.g. modifying and combining existing images. For example, related studies have shown some operations like flipping, rotation, scaling and cropping can improve the robustness of the model and prevent overfitting. In addition, existing studies have found several ways to augment data. In [1], the CutMix augmentation strategy was suggested by the authors. In this method, patches are cut and pasted between training images, and the ground truth labels are blended in proportion to the area of the patches.

CutMix routinely outperforms state-of-the-art augmentation techniques on CIFAR and ImageNet classification tasks, as well as on the ImageNet weakly-supervised localization challenge, by making effective use of training pixels and maintaining the regularization effect of regional dropout. In [6], the authors added random Gaussian noise to the image to create a new sample, and then add the new sample to the data set. They compared the performance of the original data set and the augmented data set on CNN, and the experimental results proved that the data augmentation operation increased the robustness and accuracy of the model. Furthermore, in [7], the authors randomly selected two samples from the training data, then averaged each pixel of the two samples to get a new sample, and finally added the new sample to the data set. By combining each sample with the other, they could eventually get  $N^2$  new samples from  $N$  training samples. The results show that this simple data augmentation technique can improve the classification accuracy of all test data sets. However, the data augmentation techniques mentioned above do not generate new additional features, but only transform operations on the existing images, which usually bring a very limited improvement in model performance.

In this research, a DCGAN was used to carry out data augmentation for a small animation image dataset and expand the size of the dataset [8]. The DCGAN used in this research is a pre-trained model on a large animation image dataset. This study took a small data set then used it to train AlexNet to tell whether a person is male or female [9]. DCGAN was transfer to this small data set for training then augment the dataset. Finally, this study used the new data set to train the AlexNet. The training results were compared with traditional data augmentation operations. The results show that the DCGAN data augmentation method used in this research has better performance than other data augmentation methods, making the accuracy of the model reach the highest level, and improving the robustness of the model.

## 2. Method

### 2.1. Model Data set and preprocessing

The data set of this paper uses the RBG cartoon images collected from the Internet. Then 200 male cartoon images and 200 female cartoon images were manually selected as the data set, whose sizes is 256x256. The total sample number is 400. The male and female image samples were divided into the train set and the test set according to the ratio of 4:6, that is, 240 samples were randomly selected to join the test set, and the remaining samples are the train set. Ratio of male and female in each set was 1:1. The data were normalized to [0, 1] before the training. Figure 1 shows the sample data in the collected dataset.

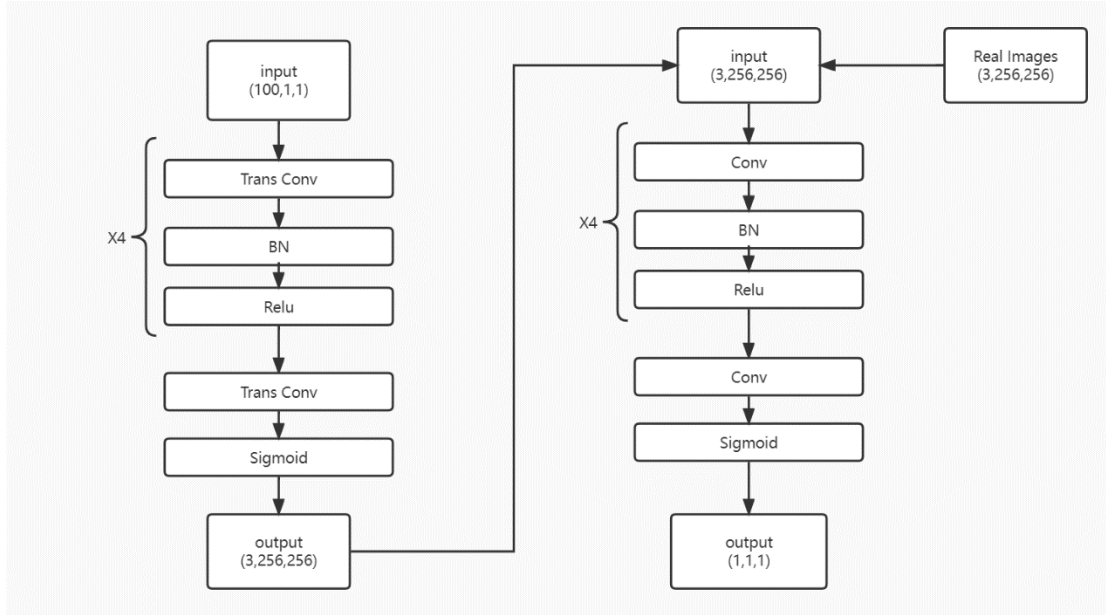


**Figure 1.** The sample data in the collected dataset.

### 2.2. DCGAN

For the specific model design, Deep Convolutional Generative Adversarial Network (DCGAN) that was built to generate 96x96 resolution images. DCGAN is a generative adversarial network, that is, it learns the generative model of data distribution by means of adversarial. The adversarial here refers to the competition between the generator network and the discriminator network. The generative network tries its best to generate realistic samples, while the discriminant network tries its best to judge whether the sample is a real sample or a generated fake sample. This study adjusted the specific parameters and re-design the model, improving it into a network that can produce higher-resolution images of 256x256. The generator of DCGAN was a 5-layer convolutional neural network, and the discriminator was also a

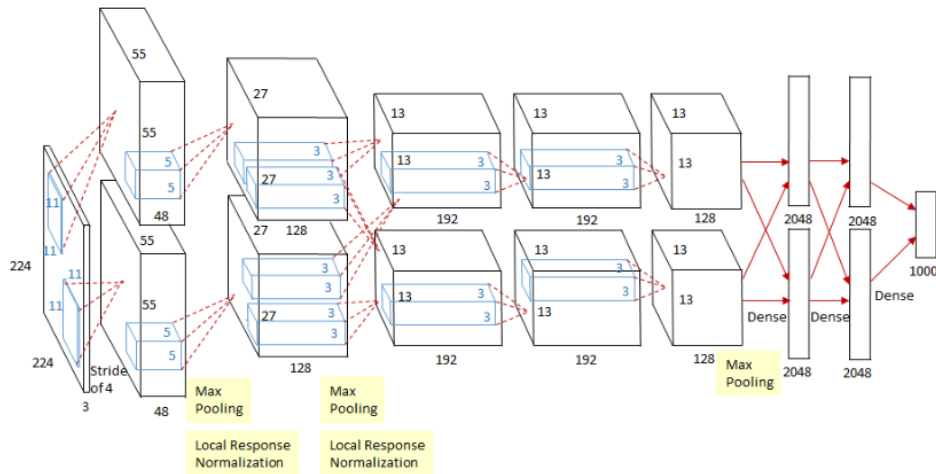
5-layer convolutional neural network, with a convolution kernel size of 3x3 or 4x4. Figure 2 presents the structure of the DCGAN used in this study.



**Figure 2.** The structure of the DCGAN used in this study.

### 2.3. AlexNet

AlexNet shown in Figure 3 is an image classification network. It was the champion network in the 2012 ISLVR 2012 (ImageNet Large Scale Visual Recognition Challenge) competition, with a classification accuracy of 80%+ from the traditional 70%+. It uses ReLU as the activation function and local response normalization (LRN). In the first two layers of the fully connected layer, the network uses Dropout random deactivated neuron operation to reduce overfitting.



**Figure 3.** The structure of the AlexNet [9].

### 2.4. Implementation Details

**2.4.1. DCGAN.** Particularly, Binary Cross Entropy loss (BCE) was used as the loss function. Moreover, the discriminator loss is a BCE loss between the true samples and true labels plus the BCE loss between

the fake samples and fake labels. The generator loss is a BCE loss between of true labels and fake samples. In this way, the discriminator will learn to classify the true and fake images, while the generator will try to output images that cannot be easily classified by the discriminator.

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

There are two terms in this total loss. In this instance, x stands for the actual image, z for the noise input to the G network, and G(z) for the image produced by the generator.

The likelihood that the discriminator will decide whether or not the true picture is real is represented by D(x) (since x is real, the closer this value is to 1 for D, the better). The likelihood that the discriminator will determine if the image produced by G is real is given by D(G(z)).

This research used pytorch to build the DCGAN, using the Adam optimizer with a learning rate of 0.0001 for the generator and a learning rate of 0.00002 for the discriminator. Batch size was set to 4 for the training set. The DCGAN used in this research is a pre-trained model on a large animation image dataset, then the DCGAN was transfer to this small data set for training. Two DCGANs were trained in this paper, one for generating the male image and one for generating the female image.

#### 2.4.2. AlexNet

$$J(x) = \frac{1}{m} \sum_{i=0}^m \sum_{j=1}^k 1_{\{y^{(i)} = j\}} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{1 \leq l \leq k} (\theta_l^T x^{(i)})} \quad (2)$$

The loss function employed in this paper is the cross entropy. In addition, pytorch was used to build AlexNet and Adam optimizer was used for optimization. The learning rate was selected as 0.0002 and batch size was set to 4. Softmax regression is used in the output layer. For it is a binary classification, there are only two nodes in the output layer to output the probability of each class respectively. Max epoch was set to 150.

Then AlexNet will be trained on 5 training sets, and finally test the results in a unified verification set and calculate the accuracy for comparison. The five training sets are the original data set, the data set augmented by DCGAN, the data set augmented by Cutout, the data set augmented by Cutmix, and the data set augmented by Gaussian noise injection augmented.






### 3. Result and Discussion

The results of model can be found in Table 1, Figure 4 and Figure 5.

After many attempts, batch size was finally set to 4 in this study. Different data augmentation methods that achieved highest test accuracy in different epochs are recorded in Table 1, and the highest accuracy is in the range of 70-150 Epochs. From the final result, without using data augmentation, test accuracy of Alexnet on this data set was 87.1%. Using DCGAN as data augmentation made AlexNet achieve the highest accuracy of 90.4% in this task, which increased test accuracy by 3.3%. Secondly, Cutmix increased the test accuracy by 1.2%, and then Noise Injection increased the test accuracy by 0.8%. Cutout method did not improve the test accuracy, instead reduced the test accuracy by 0.4%.

As can be seen from Table 1, the method of data augmentation using DCGAN has achieved the highest test accuracy, while the traditional method of data augmentation often has limited improvement in test accuracy, and some methods even reduce the accuracy. It can be observed from the table that the value of training accuracy is relatively low under the traditional augmentation method. One possible reason is that the traditional augmentation techniques do not generate new additional features, but only transform operations on the existing images, which usually bring a very limited improvement in model performance. Another possible reason is that the main role of traditional data augmentation is to prevent the overfitting of the model and improve the robustness of the model. AlexNet had a feature that randomly inactivating part of neurons in the full connectivity layer, which has a good performance of preventing overfitting. Improvements through traditional data augmentation are likely to be further tempered.

**Table 1.** Performance Based on Augmentation Methods.

Augmentation Method		Training loss	Training Accuracy	Testing Loss	Testing Accuracy
None		0.216	85% (0.00%)	0.436	87.1% (baseline)
DCGAN		0.211	89.7% (+4.7%)	0.356	90.4% (+3.3%)
Cutout [10]		0.238	84.3% (-0.7%)	0.378	86.7% (-0.4%)
Cutmix [1]		0.234	84.5% (-0.5%)	0.396	88.3% (+1.2%)
Noise Injection [6]		0.289	85.2% (+0.2%)	0.387	87.9% (+0.8%)



**Figure 4.** Part of Male Image Generated by DCGAN.



**Figure 5.** Part of Female Image Generated by DCGAN.

#### 4. Conclusion

This research discusses the problem that the performance of classification networks on small data sets tend to be bad, and gives a better solution of extending data sets. In this study, two DCGANs were used

to augment the data set for Alexnet in a classification task, which is to distinguish whether a character in the cartoon image is male or female. One DCGAN is used to augment the male samples and one is used to extend the female samples and then add them up to the train set. In comparison with other data augmentation methods such as Cutout, Cutmix and noise injection, the method using DCGANs had the highest accuracy and performance. However, the quality of the image generated by DCGAN is unsatisfied. There is still a big difference between generated image and the original image in the data set. In further research, some better or more advanced generating models may be utilized to augment the data.

## References

- [1] Samee N A et al. 2022 A Hybrid Deep Transfer Learning of CNN-Based LR-PCA for Breast Lesion Diagnosis via Medical Breast Mammograms Sensors 22(13) 4938
- [2] Jia Q et al. 2022 Bitr-unet: a cnn-transformer combined network for mri brain tumor segmentation In International MICCAI Brainlesion Workshop pp. 3-14 Springer Cham
- [3] Yu Q et al. 2022 Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training Biomedical Signal Processing and Control 72 103323
- [4] Chen S et al. 2022 An end-to-end approach to segmentation in medical images with CNN and posterior-CRF Medical Image Analysis 76 102311
- [5] Yun S 2019 et al. Cutmix: Regularization strategy to train strong classifiers with localizable features Proceedings of the IEEE/CVF international conference on computer vision
- [6] Moreno-Barea Francisco J et al. 2018 Forward noise adjustment scheme for data augmentation IEEE symposium series on computational intelligence (SSCI) IEEE
- [7] Inoue H 2018 Data augmentation by pairing samples for images classification arXiv preprint arXiv:1801.02929
- [8] Radford A 2015 Unsupervised representation learning with deep convolutional generative adversarial networks arXiv preprint arXiv:1511.06434
- [9] Krizhevsky A et al. 2017 Imagenet classification with deep convolutional neural networks Communications of the ACM 60(6): 84-90
- [10] DeVries T et al. 2017 Improved regularization of convolutional neural networks with cutout arXiv preprint arXiv:1708.04552.