# FACE GENERATION BASED ON GENERATIVE ADVERSARIAL NETWORK FOR DATA AUGMENTATION

[1]*Ssu-Ming Wang* (王思敏) [1]*Chiou-Shann Fuh* (傅楸善)

[1] Graduate Institute of Biomedical Electronics and Bioinformatics,
National Taiwan University, Taipei, Taiwan,
E-mail: r07945029@ntu.edu.tw fuh@csie.ntu.edu.tw

**ABSTRACT**

In fields such as face recognition and image semantic segmentation, data augmentation is kept an important for training Deep Neural Network (DNN) with larger dataset and without overfitting. In this research, we propose an Auxiliary Classifier Generative Adversarial Network based generator for face generation by using a set of real human face, CelebA Dataset which labels are up to 40 classes, as the training data. The result shows that the generator can produce the realistic face by the model, however, the background environment looks blurring in testing image samples. In spite of the blurring background, Generative Adversarial Network provides a state-of-art approach to deep learning programmers for artificial data augmentation.

***Keywords:*** *Face generation, Data augmentation, Generative Adversarial Network.*

## 1. INTRODUCTION

Data augmentation is known to be important for training deep neural network (DNN) in fields such as face recognition, visual context, and image semantic segmentation, etc. [1]

In statistics, overfitting is "the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably". [2] Generally, for an overfitting model, the outcomes of score metrics prediction will be well in the training data while being terrible in validation data. Lack of the training data amount is the main reason that leads to overfitting in training process according to the following points. First, the degree of freedom for parameters of a complex DNN model will probably exceed the information contained in the data with less training data. Second, some specialized and random characteristics will be fitted during the training process. Hence, some skills are needed to avoid the situation, for example, cross validation, early stopping, and model comparison, etc. By means of applying the above approaches, one can avoid overfitting as well as keep the testing results on expected standard level. However, further improvement of the training and testing quality can be implemented by data augmentation, a technique of artificially increasing the data amount by data generator.

The related work included works proposed by Nikita Dvornik and Julien Maira in 2019 and Hiroshi Inoue in 2018. In "On the Importance of Visual Context for Data Augmentation in Scene Understanding" proposed by Nikita Dvornik and Julien Maira, an explicit context model by using a convolutional neural network was suggested to predicts whether an image region is suitable for placing a given object or not. [2] On the other hands, in "Data Augmentation by Pairing Samples for Images Classification" proposed by Hiroshi Inoue, a technique named SamplePairing was suggested to synthesize a new sample from one image by overlaying another image randomly chosen from the training data. [3]

In this research, a realistic face generator based on Generative Adversarial Network (GAN) with ResNet architecture is implemented which can be applied in fields such as face recognition and face emotion detecting. The face generator is built by an Auxiliary Classifier GAN (ACGAN) which is composed of a generator and a discriminator with ResNet architecture. The discriminator contains San auxiliary classifier and an adversarial net for back propagation of labels and generated images losses. Data and labels used for training model are from Large-scale CelebFaces Attributes (CelebA) Dataset (http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html).

The outcomes imply that the ACGAN based face generator can produce realistic face image, however the complicated environment background ware not able to painting by the generator. Even so, the result still shows the feasibility to artificially generate multi-label face images by GAN and provides a state-of-art approach of data augmentation for programmers.

## 2. METHOD

### 2.1. Data Descriptions

CelebFaces Attributes Dataset (CelebA) is a large-scale face attributes dataset with more than 200K celebrity images, each with 40 attribute annotations. The images in this dataset cover large pose variations and background clutter. CelebA has large diversities, large quantities, and rich annotations, including 10,177 numbers of identities, 202,599 numbers of face images, and 5 landmark locations, 40 binary attributes annotations per image [4]. The main part used in this research contains the face images, on which an example is shown in Figure 1, and the corresponding 40 classes of binary label, including 5 o' Clock Shadow, Arched Eyebrows, Attractive, Bags Under Eyes, Bald, Bangs, Big Lips, Big Nose, Black Hair, Blond Hair, Blurry, Brown Hair, Bushy Eyebrows, Chubby, Double Chin, Eyeglasses, Goatee, Gray Hair, Heavy Makeup, High Cheekbones, Male, Mouth Slightly Open, Mustache, Narrow Eyes, No Beard, Oval Face, Pale Skin, Pointy Nose, Receding Hairline, Rosy Cheeks, Sideburns, Smiling, Straight Hair, Wavy Hair, Wearing Earrings, Wearing Hat, Wearing Lipstick, Wearing Necklace, Wearing Necktie, and Young.

Fig. 1: An example of face image in CelebA dataset.

### 2.2. Auxiliary Classifier Generative Adversarial Network (ACGAN)

In conditional GAN (cGAN) model, labels will be input to the discriminator to increase the discriminability. Different from the approach, ACGAN introduces an auxiliary classifier to assist classification. The overall structure of ACGAN is shown in Figure 2.

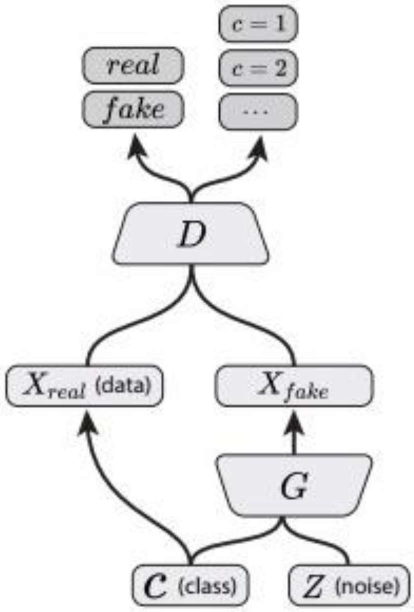


Fig. 2: Structure of ACGAN.

The loss of the discriminator will be split into adversarial loss and auxiliary loss to evaluate the facticity of the fake generated images from real images and the accuracy of label classification separately. The chosen optimizer and loss of the model are Adam and Binary Cross Entropy Loss (BCELoss).

### 2.3. Wasserstein Auxiliary Classifier Generative Adversarial Network (WACGAN)

In WACGAN model, the Wasserstein loss was introduced to calculate and propagate loss term. The Wasserstein loss term is shown in Figure 3.

| GAN | DISCRIMINATOR LOSS |
|---|---|
| WGAN | $\mathcal{L}_{\mathrm{D}}^{\mathrm{WGAN}} = -\mathbb{E}_{x\sim p_d}[D(x)] + \mathbb{E}_{\hat{x}\sim p_g}[D(\hat{x})]$ |

| GAN | GENERATOR LOSS |
|---|---|
| WGAN | $\mathcal{L}_{\mathrm{G}}^{\mathrm{WGAN}} = -\mathbb{E}_{\hat{x}\sim p_g}[D(\hat{x})]$ |

Fig. 3: Generator and discriminator loss function of Wasserstein GAN.

The main improvements of Wasserstein loss are replacing the log operation to mean operation from loss calculation, removing the Sigmoid layer from the discriminator, and clipping the updating weights to a certain range to satisfy the Lipschitz continuity condition. The introduction of Wasserstein loss seems to be able to solve the issue of unstable training process by means of measuring the distance between the generated data distribution and the real data distribution instead of the original cross entropy method which is not suitable for measuring the distance between distribution with disjoint parts.

### 2.4. ResNet

The architecture of the generator and discriminator of ACGAN and WACGAN model are based on ResNet. Inside generator and discriminator are two Resnet with 2 layers and 4 layers separately. Each layer of generator consists of 2 blocks with sequence of convolutional layers, batch normalization layers, leaky ReLU layers, convolutional layers, batch normalization layers, and leaky ReLU layers in each block. On the other hand, each layer of discriminator consists of 2 blocks with sequence of convolutional layers, batch normalization layers, leaky ReLU layers, and dropout layers in each block.
In generator, input noise will be fed into a linear layer first. Afterward, the output of the linear layer will be concatenated with the input label and fed into the Resnet of generator. The output will eventually be fed to a convolutional layer and activated by Tanh.
In discriminator, input image will be fed into the Resnet of generator. The output will then be fed into adversarial layer and auxiliary layer which are both composed of a Linear layer and a Sigmoid layer with different output sizes for validity and label.
The hyperparameters of the models are shown in Tables 1, 2, 3, and 4.

Table 1: Hyperparameters of ResNet-based image generator.

| Hyperparameters | size |
|---|---|
| Batch Normalization Layer | 0.8 |
| Block convolution Layer of Layer 1 | 128 |
| Block convolution Layer of Layer 2 | 64 |
| Leaky ReLU Layer | 0.2 |
| Convolution Layer 3 | 3 |

Table 2: Hyperparameters of ResNet-based Discriminator.

| Hyperparameters | size |
|---|---|
| Batch Normalization Layer | 0.8 |
| Block convolution Layer 1 | 16 |
| Block convolution Layer 2 | 32 |
| Block convolution Layer 3 | 64 |
| Block convolution Layer 4 | 128 |
| Leaky ReLU Layer | 0.2 |
| Dropout Layer | 0.25 |

Table 3: Hyperparameters of Adversarial Classifier of Discriminator.

| Hyperparameters | size |
|---|---|
| Flatten Layer | None |
| Linear Layer | 1 |
| Sigmoid Layer | None |

Table 4: Hyperparameters of Auxiliary Classifier of Discriminator.

| Hyperparameters | size |
|---|---|
| Flatten Layer | None |
| Linear Layer | Label Classes |
| Sigmoid Layer | None |

### 2.5. Score Metrics

The score metrics in this research is Fréchet Inception Distance (FID) score. FID score is based on calculating the distance between the real picture and the false picture at the feature level. With the FID score, the training set of the generated model can be different from that of Inception set. Besides, the generated data and the real data are all used in calculation of FID score, thus, the results are more reasonable. Finally, FID score prediction will not optimize the final generated output image. Therefore, the output image distortion can be avoided.

## 3. RESULTS AND DISCUSSION

The Experiments cover the ResNet-based ACGAN and MACGAN generator comparison.
The result of ACGAN generator is shown in Figures 4, 5, and 6. Figure 6 shows the producing result of first epoch. From Figure 6, we can observe that the face contour of generated image had been distinguished with the unclear facial features. However, after about 6 epochs of training, the results in Figures 4 and 5 imply that the ACGAN model can realize the face image data generation with clear facial features and contour.
From the observation from Figures 4, 5, and 6, some works still needed to be implemented in the future work to make the improvement to the current model such as fixing the blurring environment problem and edge enhancement. Also, painting fixation can be applied to the generated images to make the photograph more complete. The unclear background and not welled distinguished edges of facial features are inferred to the numerous label classes and the lack of labels related to the background environment information leading to the higher difficulty of background producing from the random noise.


Fig. 4: Multiple face image results generated by ACGAN face generator.


Fig. 5: More detailed results of Fig. 4.

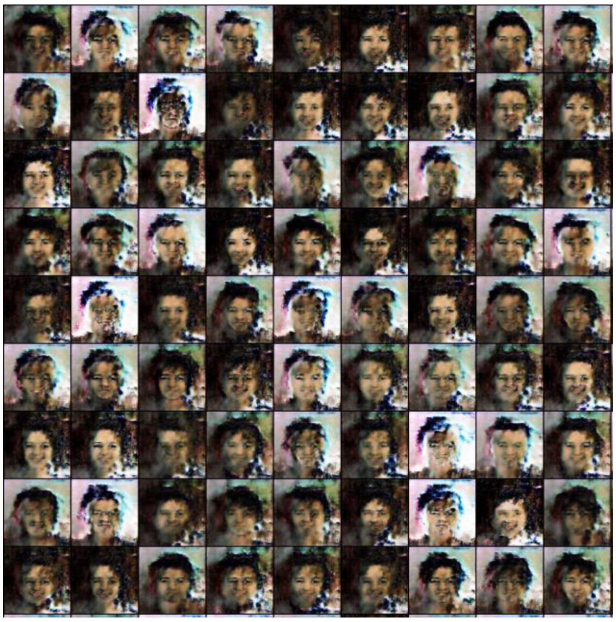
Fig. 6: Multiple face image results generated by ACGAN face generator in the first epoch.

The result of WACGAN generator is shown in Figures 7, 8, 9, and 10. Same as the ACGAN model generation task outcomes, Figure 9 shows the result of WACGAN in the first epoch. From Figure 9, we can observe that the face contour of generated image had been distinguished with the unclear facial features, and after about 4 epochs of training, the results in Figures 7 and 8 also imply that the WACGAN model can realize the face image data generation with clear facial features and contour.

However, the outcome in Figure 10 implies the limitation of the current model that after too many training epochs, the current model will lose the capability of telling the label classes and cause the generation of the confusing face images.

Fig. 7: Multiple face image results generated by WACGAN face generator.

Fig. 8: More detailed results of Fig. 7.

Fig. 9: Multiple face image results generated by WACGAN face generator in the first epoch.

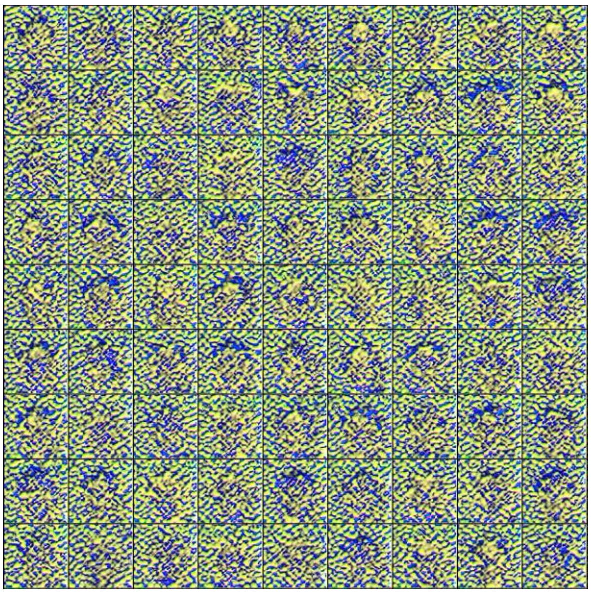

Fig. 10: Multiple face image results generated by WACGAN face generator after too many epochs.

For ensuring the reason leading to image generation error of WACGAN model, discriminator and generator loss plots of ACGAN model and WACGAN model are shown in Figures 11, 12, 13, and 14. In ACGAN model, the loss function of the generator is the BCELoss calculated by the fake image and real image, and the loss function of the discriminator is composed of adversarial loss and auxiliary loss which are both calculated with BCELoss function according to the predicting outcomes of adversarial classifier and auxiliary classifier.
Comparing the two loss graphs, we can observe that the ACGAN loss keeps above the zero line and decrease slowly during the training process in Figures 11 and 12.

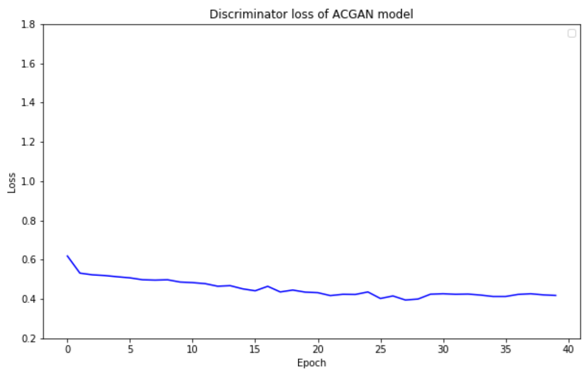


Fig. 11: Discriminator loss of ACGAN face generator in training process.

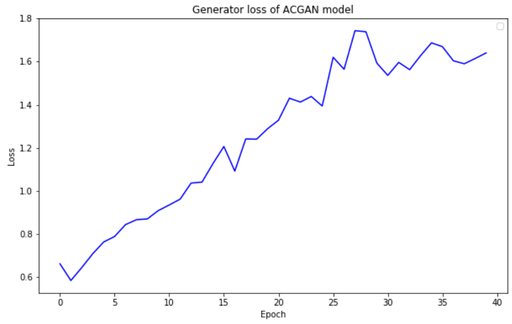


Fig. 12: Generator loss of ACGAN face generator in training process.

Compared with the ACGAN loss in Figures 11 and 12, however, the WACGAN loss shows that the loss value decreases slowly at the beginning of the training in Figures 13 and 14. After the loss becomes negative, the loss decreases dramatically linearly until loss value reaches -0.18, then keep unchanged.

Corresponding to the training epoch and image generation time, the point that discriminator loss value becomes unchanged is close to the image generation error time point. Thus, we suppose that the backpropagation of the negative discriminator loss accumulates the error since about epoch 11 and eventually causes the error of confusing face image generation after epoch 13.

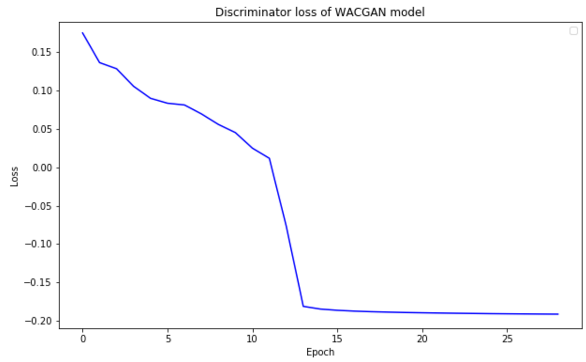


Fig. 13: Discriminator loss of WACGAN face generator in training process.

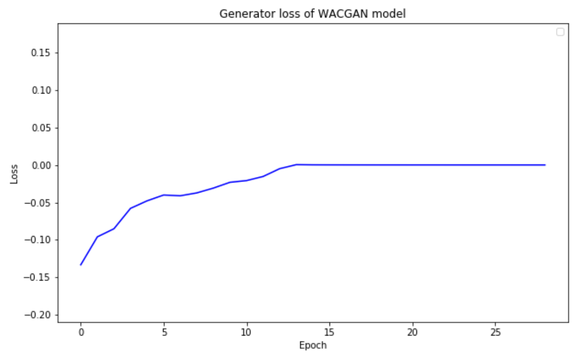


Fig. 14: Generator loss of WACGAN face generator in training process.

Besides, from the two loss plots, we can observe that the generator loss keeps increasing since the beginning of the training, no matter the ACGAN model or WACGAN model is. The reason for this confusing outcome is that the calculation of the loss of generator is based on the loss of discriminator. With the decreasing value of discriminator, loss of generator will become large more easily. Thus, a better method to evaluate and visualize the training quality such as G*D, etc, is needed to discuss in the future.

The FID score prediction results of ACGAN and WACGAN are shown in Figures 15 and 16.

Comparing Figures 15 and 16, the results imply that in the beginning of the training, WACGAN model shows the better result in FID prediction. While in the prediction of FID score in every 5 epochs, the results show increasing FID score. Referring to the results from loss plots and FID prediction plots, we find that the improvement of the classification capability of discriminator in ACGAN and WACGAN model is needed to decrease the FID which means the better result in image generation.

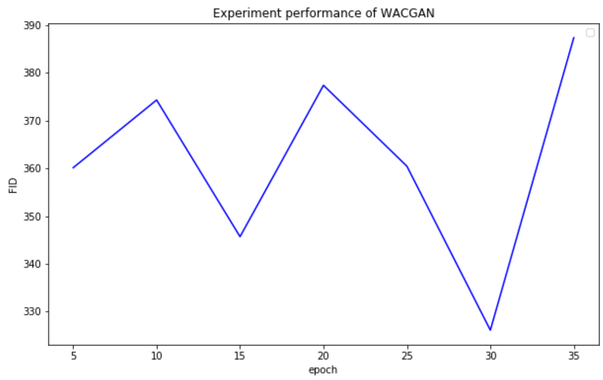


Fig. 15: FID score prediction of ACGAN model.

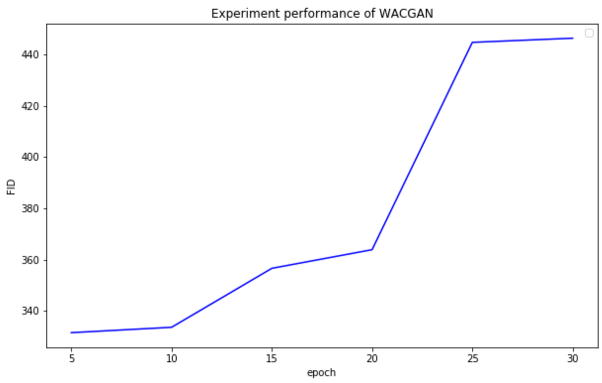


Fig. 16: FID score prediction of WACGAN model.

To verify if the current discriminator could work on the less label classification task with better performance, the same experiment has implemented on the Cartoon Set, which has only 15 classes of label. The generated images in Figures 17, 18, 19, and 20 and the FID score calculation in Figures 21 and 22 confirm the ACGAN and WACGAN discriminators capability in less attributes classification task and imply the need of improving the current discriminator model for classifying numerous attributes. Also, lack of label and complexity of environment background also has an effect on the image generation task and is needed to be solved in the future work.

Fig. 17: Multiple face image results generated by ACGAN face generator on Cartoon Set.

Fig. 18: More detailed results of Fig. 17.

Fig. 19: Multiple face image results generated by WACGAN face generator on Cartoon Set.

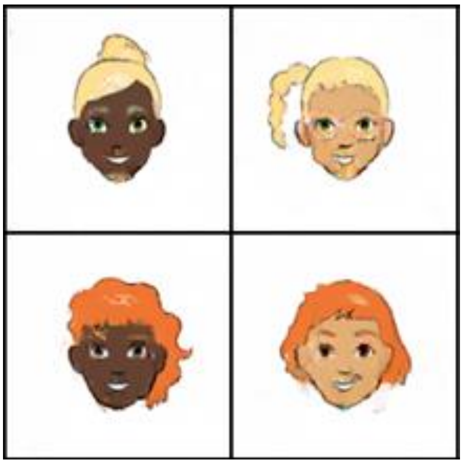

Fig. 20: More detailed results of Fig. 19.

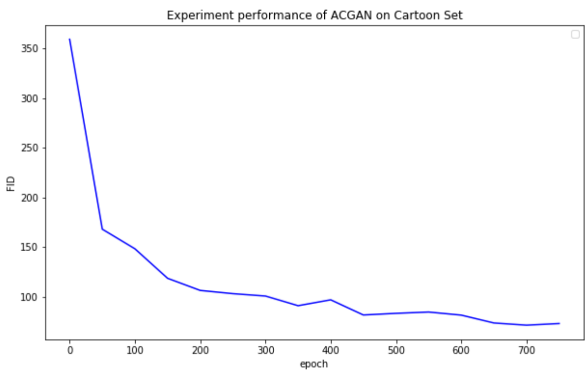


Fig. 21: FID score prediction of ACGAN model on Cartoon Set.

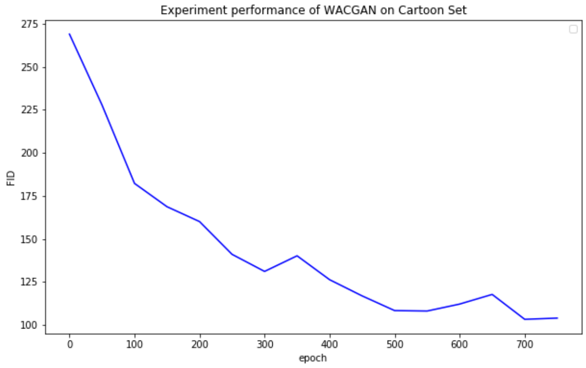


Fig. 22: FID score prediction of WACGAN model on Cartoon Set.

In conclusion, ACGAN and WACGAN models have potential to provide an approach in data augmentation to generate more data than the traditional methods such as rotate, flip, and blur. In the future, we can focus on the improvement of the discriminator performance on the numerous labels classification, fixing the loss problem of WACGAN model, and other works such as edge enhancement and unclear background problem.

## 4. REFERENCES


[1] Definition of "overfitting" at OxfordDictionaries.com: this definition is specifically for statistics.

[2] On the Importance of Visual Context for Data Augmentation in Scene Understanding Nikita Dvornik, Julien Mairal, Senior Member, IEEE, and Cordelia Schmid, Fellow, IEEE

[3] Data Augmentation by Pairing Samples for Images Classification

[4] http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html