Fine-grained Multi-attribute Adversarial Learning for Face Generation of Age, Gender and Ethnicity

Lipeng Wan[1+], Jun Wan[2+], Yi Jin[1*], Zichang Tan[2], Stan Z. Li[2]
[1] Beijing Jiaotong University
{lipengwan, yjin}@bjtu.edu.cn
[2] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
{jun.wan, tanzichang2016}@ia.ac.cn

Abstract

Since the Generative Adversarial Network (GAN) was proposed, facial image generation used for face recognition has been studied in recent two years. However, there are few GAN-based methods applied for fine-grained facial attribute analysis, such as face generation with precise age. In this paper, fine-grained multi-attribute GAN (FM-GAN) is presented, which can generate fine-grained face image under specific multiply attributes, such as 30-year-old white man. It shows that the proposed FM-GAN with fine-grained multi-label conditions is better than conditional GAN (c-GAN) in terms of image visual fidelity. Besides, synthetic images generated by FM-GAN are used for data augmentation for face attribute analysis. Experiments also demonstrate that synthetic images can assist the CNN training and relieve the problem of insufficient data.

1. Introduction

Facial attributes analysis is an active research topic in the pattern recognition for many years. However, for a long time lack of sufficient training data was one of the main challenges, especially in age estimation [4]. At that time, collecting face images of each age in the human lifetime from the same people is quite difficult. So many aging datasets appear to exist serious imbalanced problem [12] and lack of samples. Recently most methods proposed focus on learning label distribution [14], local regions of faces from limited samples [15] and generating new data. Although there has been quite a few methods of generating face images to supply data, more precise generation of fine-grained attributes is precious. Besides, in real life age is not an independent attribute and has some relevance to other facial attributes (e.g. gender, ethnicity and so on). Generation and modification of these attributes require combination of local changes and global changes. The presented traditional face generation of different attributes is limited to modeling of progressing pattern without considering global facial changes. So jointly analyzing face generation of multi-attributes and focusing on synthesis of fine-grained ages are necessary for lots of facial works in the future.

In recent two years, natural image generation has been developed to a new height by Generative Adversarial Networks(GAN) proposed by Ian Goodfellow [5] which will be introduced in Section 2 for details. This model has been verified to be able to produce images with fairly high visual fidelity [11] and learn abundant representations from training samples [8], like learning pose representation for frontal facial synthesis [3]. After that, lots of works focus on employing GAN to face applications. However on the face aging and generation of different ages problem, most methods simplify the problems by only making generated face older or younger [10], or dividing the range of ages into several groups to generate face images of different stages [1].

In contrast, we propose a novel method for face generation of fine-grained multi-attribute. Our model could generate realistic face images of multi-attribute including gender, ethnicity and fine-grained ages according to the provided conditions. Meanwhile, we use synthetic samples as supplement for original dataset and solve data augmentation problem to some extent.

The summary of contributions of our work is the following:

- We propose a novel FM-GAN for face generation of multi-attribute including fine-grained synthesis of different ages. Synthetic images perform great visual fidelity, and representations of gender, ethnicity and age are perfectly disentangled from other variations.
- We propose to enlarge the MORPH Album II dataset [12] with our generated samples and ap-

*:Corresponding author; +:Equal contribution; This work was done by Lipeng Wan who visited the lab of Center for Biometrics and Security Research, NLPR, CASIA.

ply the new dataset to assist age estimation training, which achieves good performance on MORPH Album II dataset. Synthetic dataset could be used as supplement data to augment other dataset without influencing its performance.

The rest of the paper is organized as follows. Related works are reviewed in Sec. 2. Sec. 3 will emphasize on our proposed method. Then, experiments are performed in Sec. 4 to evaluate our method. Finally, conclusion and future works are drawn in Sec. 5.

2. Related Work

Generative Adversarial Networks (GAN) As introduced by Ian Goodfellow et al. [5], Normal GAN consists of a generator D and a discriminator G that compete with each other in a two-player minimax game. G learns a mapping from probability distribution $P_z(z)$ of latent vectors in low-dimension manifold to data $P_{data}(x)$ in high-dimension manifold and synthesizes face images $\hat{x} = G(z)$, where $z \sim N(0,1)$ is random latent vector, as real as possible to fool D while D tries to distinguish the generated images $\hat{x}$ from training images x. When G outputs images that D could not judge whether it's real or fake, then a good enough GAN model gets trained. The adversarial functions can be described as below:

$$
\begin{aligned}
argmin_{G} max_{D} V(D,G) = E_{x \sim p_{data}(x)}[\log D(x)] + \\
E_{z \sim p_z(z)}[\log(1 - D(G(z)))]
\end{aligned}
\tag{1}
$$

More recent works on GAN focus on face applications, such as Face aging [1], Face modification [10], Frontal face synthesis for recognition [7, 6]. In contrast. we propose an extended GAN mainly for studying face generation of fine-grained multi-attribute. By synthesizing facial attributes dataset, synthetic images could supply other dataset from the perspective of data augmentation.

cGAN versus AC-GAN Most proposed methods about adding extra information to GAN are based on two models: conditional GAN (cGAN) [8] and auxiliary classifier GAN (AC-GAN) [9]. The former is implemented by supplying both generator and discriminator with class labels in order to learn conditional distribution. The latter tasks discriminator as an auxiliary classifier to output the predicted conditions, and the generator could be seen as an decoder to map current conditional vector and noise vector to a synthetic face image. So the whole process is a conditional reconstruction. Both methods have its own advantages on conditional face generation. Considering both methods, our FM-GAN is proposed and modified from AC-GAN for face generation of fine-grained multi-attribute.

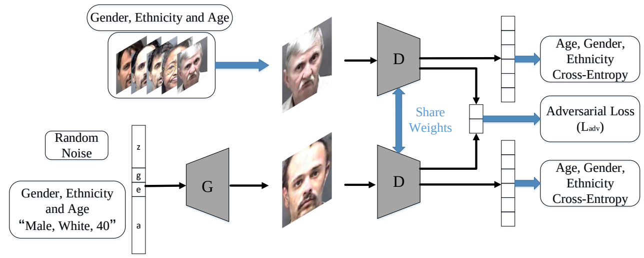

Figure 1. The overall architecture of our FM-GAN.

3. The Proposed Method

The overall architecture of our FM-GAN can be seen in Fig. 1. In the following content, we will focus on introducing the implementations of FM-GAN and describe it in the form of an algorithm.

3.1. Fine-grained Multi-attribute GAN (FM-GAN)

The aim of multi-attributes facial synthesis is to produce realistic and sufficient face samples based on MORPH-II dataset [12] and assist age estimation classifier training. In order to achieve such networks, we adopt a extension of the generative adversarial networks to multi-attribute setting. The crucial problem is how to lead fine-grained side information into GAN. Compared with cGAN [8] which directly inserting labels into discriminator, tasking GAN with conditional reconstruction is the better way. In the process of conditional reconstruction, the discriminator D is tasked as an multi-attribute classifier networks to output the predicted conditional information [9].

Given sufficient training faces $\{x_i, {y_i}^g, {y_i}^e, {y_i}^a i = 1...n\}$, where n is the number of images in our training set, g, e, a represent the gender, ethnicity and age label, respectively. Before being supplied to generator, age should be transformed to a one-hot vector with N_a dimensions. N_a means the number of fine-grained categories. Meanwhile, we make an assumption that generated faces and training faces have hidden labels. All training faces are labeled with 1 while all synthetic images are labeled with 0.

The discriminator should not only learn to classify the real face images to its hidden label 1 and the synthetic face images to 0, which is described as distinguishing real face images from synthetic face images in previous GAN work, but also learn multiple labels distribution and classify real face images to its corresponding multiply classes through training. Its parameters are optimized by minimizing adversarial loss $L_{adv}(D)$ and softmax cross-entropy loss $L_{atr}(D)$. For any training sample (x, y^a, y^g, y^e) and synthetic sample $\hat{x} = G(z, a, g, e)$, the optimization problem can be formulated as below:

$$
\begin{aligned}
L_{adv}(D) &= E_{x\sim p_{data}(x)}[\log D(x)] \\
&\quad + E_{\substack{z\sim p_z(z), g\sim p_g(g),\\ e\sim p_e(e), a\sim p_a(a)}}[\log(1-D(G(z,g,e,a)))] \\
L_{atr}(D) &= E_{x,y\sim p_{data}(x,y)}[\log D_{y^a}^{\hat{y^a}}(x) + \log D_{y^g}^{\hat{y^g}}(x) + \log D_{y^e}^{\hat{y^e}}(x)] \\
\max_D V_D(D,G) &= L_{adv}(D) + L_{atr}(D)
\end{aligned}
\tag{2}
$$

where z is the random noise, g, e, a respectively stands for class of gender, ethnicity and age sampled from label distribution as the input to generator. $L_{adv}(D)$ is introduced to distinguish real face images ${x_i}_{i=1}^n$ from synthesized ones ${x'}_{i=1}^n$. $\hat{y^a}, \hat{y^g}, \hat{y^e}$ are the output of discriminator for training images as a auxiliary classifier .

Following the training pace of discriminator, generator G is updated to synthesize realistic face $\hat{x}$ of specific gender g, ethnicity e and age a in the input of generator which could fool discriminator to classify them to the real label 1. The optimization formulations of generator are listed as followed:

$$
\begin{aligned}
L_{adv}(G) &= E_{\substack{z\sim p_z(z), g\sim p_g(g),\\ e\sim p_e(e), a\sim p_a(a)}}[\log(D(G(z,g,e,a)))] \\
L_{atr}(G) &= E_{\substack{z\sim p_z(z), g\sim p_g(g),\\ e\sim p_e(e), a\sim p_a(a)}}[\log D_a^{a'}(G(z,g,e,a)) + \\
&\qquad \log D_g^{g'}(G(z,g,e,a)) + \log D_e^{e'}(G(z,g,e,a))] \\
\max_G V_G(D,G) &= L_{adv}(G) + L_{atr}(G)
\end{aligned}
\tag{3}
$$

where a', g', e' are the output of discriminator for generated images as a auxiliary classifier.

3.2. Learning Strategy

The Algorithm 1 summarizes the training procedure. After initializing the input of generator and discriminator (lines 2,3) we generate faces of specific multiple attributes. The generated images and real images are supplied to discriminator. $y^{g'}, y^{e'}, y^{a'}$ encode real images to predicted gender, ethnicity and age. s_r estimates real samples' probability. g', e', a' encode generated images to predicted gender, ethnicity and age. s_f estimates synthetic samples' probability. Lines 8,10,11 indicate taking a gradient step to optimize GAN.

4. Experiment

In this section, we will introduce three datasets on which all of our following experiments are carried out and describe the implementation details during training GAN, especially some tricks to optimize the training process. Last is to evaluate the performance of our FM-GAN and verify our demonstration.

Algorithm 1 The FM-GAN with gender, ethnicity and age representations learning strategy

Input: Minibatch Images: $x = \{x_i, y_i^g, y_i^e, y_i^a\}_{i=0}^{m-1}$, Latent Representation vector: $z = \{z_i\}_{i=0}^{m-1}$ Gender, Ethnicity and Age Representation vector batch: $g = \{g_i\}_{i=0}^{m-1}$, $e = \{e_i\}_{i=0}^{m-1}$, $a = \{a_i\}_{i=0}^{m-1}$, Batchsize: m, learning rate $\lambda = 0.0002$.
Output: Generated Images: $x' = \{x_i{}'\}_{i=0}^{m-1}$

1: **while** not converge **do**
2: $z \sim U(-1,1)^Z$, {Draw sample of random noise}
3: $(g, r, a) \sim p_{data}(g, r, a)$, {Draw specific label from labels distribution}
4: $x' \leftarrow G(z, g, r, a)$, {Decode vector forward through generator}
5: $(s_r, y^{g'}, y^{e'}, y^{a'}) \leftarrow D(x)$
6: $(s_f, g', e', a') \leftarrow D(x')$
7: $L_D \leftarrow \log(s_r) + \log(1-s_f) + \log_{y^g}(y^{g'}) + \log_{y^e}(y^{e'}) + \log_{y^a}(y^{a'})$
8: $D \leftarrow D - \lambda \cdot \frac{\partial L_D}{\partial D}$, {update discriminator}
9: $L_G \leftarrow \log(s_f) + \log_g(g') + \log_e(e') + \log_a(a')$
10: $G \leftarrow G - \lambda \cdot \frac{\partial L_G}{\partial G}$, {update generator}
11: $G \leftarrow G - \lambda \cdot \frac{\partial L_G}{\partial G}$, {update generator twice}
12: **end while**

4.1. Dataset

MORPH Album II [12] is one of largest datasets widely used for human facial age estimation. All samples of dataset are under age, gender and ethnicity variations in controlled environment, containing 55244 images of 19598 subjects in which most are nearly frontal faces or some having poses within $\pm 30°$. During the stage of training GAN, we find that sufficient samples are necessary. Lack of training data severely influence the generative results. So the whole dataset is fully used for training GAN. We follow the work [15] to split it into three non-overlapped subsets $S1$, $S2$, $S3$ randomly. The details of these subsets are described in the test protocols[1] provided by Yi et al [15] and Tan et al. [13]. Before training, we first align and crop all samples $(x_i, y_i)_{i=1}^n$ to resolution of $128 \times 128 \times 3$ according to the distance between eyes and nose to let whole head with hair The final samples are shown in Fig. 2. Details of three attributes y in the MORPH II dataset used illustrate as followed: age labels y^a are from 16 to 77, gender y^g with only two labels (e.g.male, female), ethnicity labels y^e are Black or White.

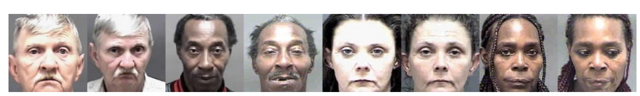

Figure 2. Some MORPH II examples of resolution $128 \times 128 \times 3$ after being preprocessed. Each couple of faces are of the different gender, ethnicity, age and identity.

Besides Morph Album II, our method has also been run on CACD [2] and FG-NET databases. The Cross-Age Celebrity Dataset(CACD) is the largest public cross-age database, which contains more than 160 thousand images from 2000 celebrities, with age ranging from 14 to 62. The

[1] http://www.cbsr.ia.ac.cn/users/dyi/agr.html

FG-NET database contains 1002 color or grayscale face images of 82 subjects, with ages from 0 to 70. These images are taken in a totally uncontrolled environment with large variations of lighting, poses and expressions. Both two datasets have only age labelled. So only the performance of age attribute is experimented on these two databases.

4.2. Implementation details

For FG-NET, the target model we will optimize is firstly pre-trained on LMDB-WIKI database which is another large age database. Due to lack of samples for training, some operations like images flipping, rotating and noising are applied to face images for data augmentation. For CACD, the target model is a modified VGGNet pre-trained on ImageNet dataset. Since the input of both pre-trained models is 224×224, the networks of GAN should be also changed to generate face images of the same size. For both databases, all face images are aligned and cropped to a view of size 224×224.

The key implementation is how to set the size of a batch. If the size of a batch is set too small or too large, GAN probably will not converge and generate complete face at all. In our work, we skillfully set the size of a batch in accordance with the number of classes in ages. It could help reasonably allocate the scale of samples referring to gender, ethnicity and age distribution in a batch. In this work, we set the size of a batch to $N^a = 64$ and let the specific input of generator be the same as the label distribution imported into discriminator. It will greatly stabilize the training process of GAN by balancing the update speed of generator and discriminator. It's not recommended to control age distribution of training samples for generator and discriminator. There exists high probability of over-fitting in the severely imbalanced condition.

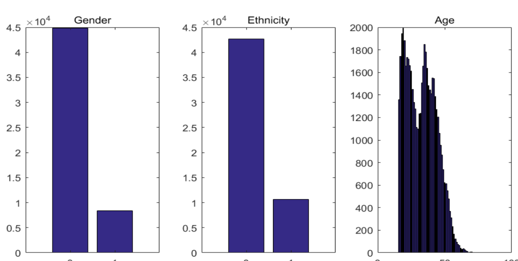

Figure 3. Labels distribution with three attributes in MORPH II dataset. The distribution of gender, ethnicity and age is all extreme. In the left sub table, $[0, 1]$ stands for $[male, female]$. In the mid table, $[0, 1]$ stands for $[Black, White]$.

Our model is extensively modified from a publicly available implementation of DC-GAN using Tensorflow[2]. The random noise z is set to a 100-dim Uniform vector. Images intensities are also linearly scaled to the range of $[-1, 1]$.

[2]https://github.com/carpedm20/DCGAN-tensorflow

Following the optimization strategy in [11], all weights in the networks are initialized from a zero-centered normal distribution with a standard deviation of 0.02. Adam optimizer is used with a learning rate of 0.0002 and momentum 0.5 in initial training. At the stage of representation learning during which network parameters need subtle adjustment, learning rate is reset to a little smaller. The detail of the networks for 128×128 generation is presented in Tab. 1.

Table 1. FM-GAN network architecture

Generator			Discriminator		
Layer	Filter Size	Output Size	Layer	Filter Size	Output Size
FC1		4x4x1024	Conv1	5x5/2	64x64x64
Fconv1	5x5/2	8x8x512	Conv2	5x5/2	32x32x128
Fconv2	5x5/2	16x16x256	Conv3	5x5/2	16x16x256
Fconv3	5x5/2	32x32x128	Conv4	5x5/2	8x8x512
Fconv4	5x5/2	64x64x64	Conv5	5x5/2	4x4x1024
Fconv5	5x5/2	128x128x3	FC1		Ng+Nr+Na+1

4.3. Model Evaluation

Besides synthesizing realistic face images of specific attributes, the main goal of our FM-GAN is to supply the generated images to original MORPH II dataset and assist its training on the task of age estimation. The visual quality of synthetic samples directly affect the performance on age estimation to a great extent. Therefore, in this following content, we firstly compare cGAN with our FM-GAN, and then concentrate on measuring the quality of synthesized samples and evaluating demonstration of assisting MORPH II dataset training.

4.3.1 cGAN versus FM-GAN

We conduct the comparative experiments based on the same network structure mentioned in Sec. 4.2 to compare our FM-GAN with cGAN [8]. Their comparison is shown in Fig. 4.

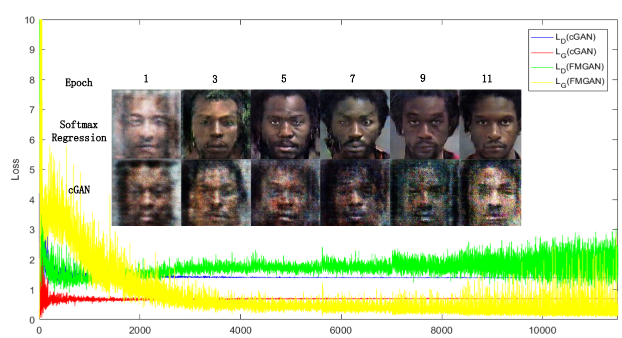

Figure 4. Comparison of generative performance between FM-GAN and cGAN during training. In the training process, one epoch equals to 1000 iterations.

Although both methods quickly converge at nearly the same pace, where green and yellow lines stand for the discriminative and generative loss curve of FM-GAN, red and blue lines stand for the corresponding loss curve of cGAN, with the training process proceeds FM-GAN has a clear process of face generation and synthesize photorealistic face images. However as for cGAN, no matter how to adjust hyperparameters in the training, the model can only generate the blurry outline of faces. It can be summed up that compared with cGAN, the implementation of FM-GAN is more suitable for achieving face generation of find-grained multi-attribute.

4.3.2 Face Synthesis by FM-GAN

Fig.5 illustrates some representative synthetic samples drawn from different attributes. Each sample corresponds to a latent vector z sampled randomly and specific labels, gender y^g, ethnicity y^e and age y^a. Our results display outstanding effect in terms of images' diversity and quality.

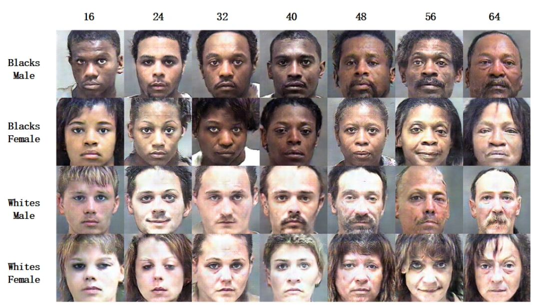

Figure 5. Examples of 128×128 synthetic images generated by our FM-GAN with the noise data randomly sampled for all faces and varying gender, ethnicity g sampled in [Male,Female], e sampled in [Black, White], and age sampled with aging process.

By assigning initial latent approximations z arbitrary fixed value, varied gender, ethnicity and ages could be observed. Except for gender, ethnicity and age, all the other facial features we have not considered and even the background factors like illumination and scene are controlled by latent noise vector z. Fig. 6 shows that image information encoded by conditions determining gender, ethnicity and age is perfectly disentangled and shows appealing effect to human eyes. In each row faces are listed in order of aging from younger to older with fixed identity, gender and ethnicity. In each column shifts are applied to gender and ethnicity with same identity and fixed age. The shifts cause noticeable effect on facial features meanwhile it is evident that slightly shifted conditions of gender and ethnicity have not influenced generation of similar-looking faces.

In order to objectively measure quality of synthetic face images and accuracy of corresponding attributes generation, we respectively use S1 and synthetic dataset of the same amount and attributes distribution with S1 (named G1 in the following paper) to train two classifiers based on modified AlexNet provided by Tan et al. [13] and evaluate their performance on S2+S3 set. Tab. 2 shows the comparison

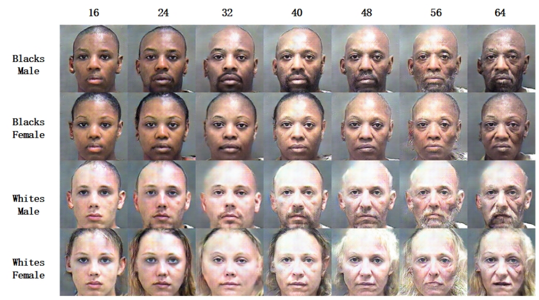

Figure 6. Examples of 128×128 samples with fixed identity(noise) and varying $g, e \in [0, 1]$ respectively in the vertical ,and $a \in [16, 24, 32, 40, 48, 56, 64]$ in the horizon

of performance between MORPH II samples and synthetic samples at the resolution of 128×128.

Table 2. Synthetic Performance on Morph-II test set

Synthetic Resolution	128x128		
Criteria	MAE	Accuracy	
Attribute	age	gender	race
MORPH-II (baseline)	3.851798	0.985435	0.970392
Synthetic images	7.369065	0.941347	0.973449

As shown in Tab. 2, the estimation accuracy of gender and ethnicity is close to the performance of real images, respectively reaching 94.1% and 97.3%. The result of age is not satisfied. Observing the generated images of each age, we make assumptions that poor generated images of older ages may be responsible for this results. These poor generated faces bring lots of noise to the training process.

In order to verify the assumption, we divide the test set into two subsets, young subset P_{S0} and old subset P_{S1}. The dividing point is set to 40. The young subset P_{S0} is composed of ages from 16 to 39. The other subset is ages from 40 to 77. We separately test the pre-trained model on two subsets and record their performance. Tab. 3 verifies our analysis. P_{S0} display better performance which approaches original real dataset while the second P_{S1} gets a worse result than the first group and the whole synthetic set. On the whole, There is still plenty of room for improving fine-grained generation.

Table 3. Synthetic performance on subsets of S2+S3

Dataset	age	gender	race
P_{S0}	6.139406	0.934923	0.974145
P_{S1}	10.331515	0.955781	0.975692

4.3.3 Data Augmentation with Synthetic Images

To further verify our demonstration that largening MORPH II dataset with synthetic samples could improve the performance of age estimation, different scales of samples in accordance with the distribution of age are taken out from G1 set(Our generated dataset) and added to S1 set for joint training. In fact, most samples taken out are from young categories having larger proportion than others. So this experiment mainly emphasizes on the contribution from generated faces of young ages before the age of 40.

Table 4. Performance of different scaled synthetic images added to MORPH II

Scale	Age	Race	Gender
Baseline	3.851798	0.985435	0.970392
0.2	3.821998	0.979125	0.962331
0.4	3.790397	0.980650	0.959916
0.6	3.786226	0.977389	0.962496
0.8	3.787912	0.977389	0.964794
1.0	3.773675	0.980181	0.959235

Tab. 4 shows the performance comparison of different scale of G1 set added to S1 set. All results are tested on S2+S3 test set. Baseline is the performance of original S1 without addition. Except for the experiment of baseline being trained individually from the beginning, the following experiments of different scales are implemented by finetuning the pre-trained model of baseline. Final results show that augmenting original MORPH II dataset with generated images obviously improve its performance on age estimation. With the increase of supplementary synthetic images, MAE keeps decreasing and becomes better. These results are sufficient to prove that generated faces can be used to solve data augmentation.

The same process of verification is applied to FG-NET and CACD. The similar results are shown in the Tab.5 below. Finetuning the model with generated images do increase the performance of model and solve generalization of model.

Table 5. Performance of age on CACD and FG-NET datasets

Scale	CACD+euclidean	FG-NET
Baseline	5.315590	6.402655
0.5	5.230732	6.149355
1.0	5.199142	6.075456

5. Conclusions and Future Works

In this paper, we proposed an novel model for face generation of multiple attributes gender, ethnicity and fine-grained age and verify the demonstration that generated images can be supplied to MORPH dataset and greatly improve its performance. Realistic synthetic images from FM-GAN could solve the data augmentation problem to some extent. Of course, generated data could not be only employed to MORPH II dataset. It could be expanded to any dataset which is restricted to missing data.

Our work largely depends on generation of high-quality face image of fine-grained multi-attribute. We believe that, the synthesis of high quality human face with fine-grained ages requires both generator and discriminator to have a certain degree of improvement. The discriminator should have ability to learn age label distribution with limited imbalanced dataset. The generator should generate not only the whole face but also every crucial part to precisely generate corresponding features of its age. The generated faces of close ages should follow the natural order of aging. For examples, two faces have disparity of less than one or two years old, their difference on face appears at local part like wrinkles around eyes. Generating high quality face image of precise multiple attributes will be our future works on generative adversarial networks.

6. ACKNOWLEDGMENTS

This work was partially supported by the National Key Research and Development Plan (Grant No. 2016YFC0801002), the Chinese National Natural Science Foundation Projects ♯61502491, ♯61403024, ♯51505004, Science and Technology Development Fund of Macau (No. 112/2014/A3, 151/2017/A, 152/2017/A).

References

[1] G. Antipov, M. Baccouche, and J.-L. Dugelay. Face aging with conditional generative adversarial networks. *arXiv:1702.01983*, 2017.

[2] B.-C. Chen, C.-S. Chen, and W. H. Hsu. Cross-age reference coding for age-invariant face recognition and retrieval. In *European Conference on Computer Vision*, pages 768–783. Springer, 2014.

[3] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *NIPS*, pages 2172–2180, 2016.

[4] X. Geng, K. Smith-Miles, and Z. H. Zhou. Facial age estimation by learning from label distributions. In *AAAI*, pages 451–456, 2010.

[5] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *NIPS*, pages 2672–2680, 2014.

[6] R. Huang, S. Zhang, T. Li, and R. He. Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. *arXiv:1704.04086*, 2017.

[7] T. Luan, X. Yin, and X. Liu. Disentangled representation learning gan for pose-invariant face recognition. In *CVPR*, 2017.

[8] M. Mirza and S. Osindero. Conditional generative adversarial nets. *arXiv:1411.1784*, 2014.

[9] A. Odena, C. Olah, and J. Shlens. Conditional image synthesis with auxiliary classifier gans. *arXiv:1610.09585*, 2016.

[10] G. Perarnau, J. van de Weijer, B. Raducanu, and J. M. Álvarez. Invertible conditional gans for image editing. *arXiv:1611.06355*, 2016.

[11] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv:1511.06434*, 2015.

[12] A. W. Rawls. Morph: development and optimization of a longitudinal age progression database. In *Joint Cost 2101 and 2102 International Conference on Biometric Id Management and Multimodal Communication*, pages 17–24, 2009.

[13] Z. Tan, S. Zhou, J. Wan, Z. Lei, and S. Z. Li. Age estimation based on a single network with soft softmax of aging modeling. In *ACCV*, pages 203–216, 2016.

[14] X. Yang, B. B. Gao, C. Xing, and Z. W. Huo. Deep label distribution learning for apparent age estimation. In *ICCVW*, pages 344–350, 2015.

[15] D. Yi, Z. Lei, and S. Z. Li. Age estimation by multi-scale convolutional network. In *ACCV*, pages 144–158, 2014.