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] School of Computer and Information Technology, Beijing Jiaotong University, China
{lipengwan, yjin}@bjtu.edu.cn
[2] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
{jun.wan, tanzichang2016}@ia.ac.cn, szli@nlpr.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 (cGAN) 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 [3]. 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 [11] and lack of samples. Recently most methods proposed focus on learning label distribution, local regions of faces from limited samples 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 [4] 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 [10] and learn abundant representations from training samples [7], like learning pose representation for frontal facial synthesis [6]. 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 [9], or dividing the range of ages into several groups to generate face images in 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 [11] with our generated samples and apply the

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

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. [4], vanilla 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-dimensional manifold to data $P_{data}(x)$ in high-dimensional 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:

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

More recent works on GAN focus on face applications, such as Face aging [1], Face modification [9], Frontal face synthesis for recognition [6, 5]. 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) [7] and auxiliary classifier GAN (AC-GAN) [8]. 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.

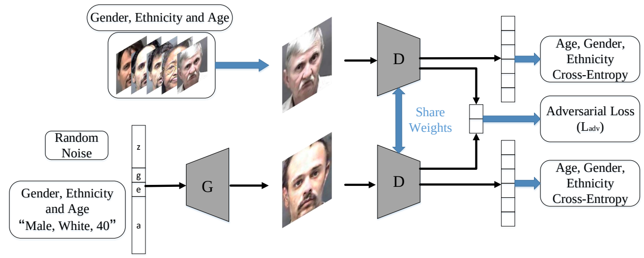

Figure 1. The overall architecture of our FM-GAN.

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-attribute facial synthesis is to produce realistic and sufficient face samples based on MORPH-II dataset [11] 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 [7] which directly insert 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 [8].

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.

The discriminator should not only learn to distinguish synthetic face images from real face images, 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 and softmax cross-entropy loss. 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:

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

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. Adversarial loss 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 specified in the input of generator which could fool discriminator to classify them to the real images. The optimization formulations of generator are listed as followed:

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

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 input into discriminator. $y^{g\prime}, y^{e\prime}, y^{a\prime}$ 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.

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: $(s_r, y^{g\prime}, y^{e\prime}, y^{a\prime}) \leftarrow D(x)$
6: $(s_f, g', e', a') \leftarrow D(x')$
7: $L_D \leftarrow \log(s_r) + \log(1 - s_f) + \log_{y^g}(y^{g\prime}) + \log_{y^e}(y^{e\prime}) + \log_{y^a}(y^{a\prime})$
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. 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.

4.1. Dataset

MORPH Album II [11] 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. The whole database is split into three subsets $S1$, $S2$ and $S3$. Then they are assembled to two couples of non-overlapped subsets $S1$ and $S2+S3$ as Test set, $S2$ and $S1+S3$ as Test set. The details of these subsets are described in the test protocols[1] provided by Yi et al [14] and Tan et al. [12]. Before training, we first align and crop all samples $(x_i, y_i)_{i=1}^n$ to resolution of 128×128 according to the distance between eyes and nose to keep 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.

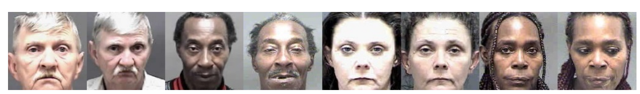

Figure 2. Some MORPH II examples of resolution 128×128 after being preprocessed. Each couple of faces are of the different gender, ethnicity, age and identity.

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 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. When evaluating on FG-NET, leave-one-person-out(LOPO) cross validation strategy is taken and averaging performance over the 82 splits is reported. When evaluating on CACD, the

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

database is split into three subset: 1800 noisy celebrities for training; 80 cleaned celebrities for validation and 120 cleaned celebrities for testing.

4.2. Implementation details

In this work, we set the size of a batch to $N^a = 64$ and let the input of generator be the same as the label distribution of training samples imported into discriminator. It will greatly stabilize the training process of GAN by balancing the update speed of generator and discriminator when updating the cross entropy loss of multiple attributes. 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. The structure of generator and discriminator being built symmetrically will stabilize the training process when updating the adversarial loss for both models. In addition, owing to lack of samples for training in FG-NET, some operations like images flipping, rotating and noising are applied to face images for data augmentation before training GAN. For both CACD and FG-NET, all face images are aligned and cropped to a view of size 128×128 like MORPH II.

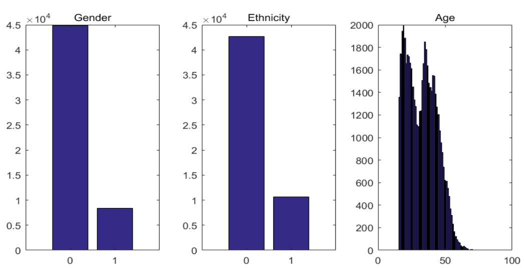

Figure 3. Labels distribution with three attributes in MORPH II dataset. The distribution of gender, ethnicity and age is all extreme. In the left sub table, $[0, 1]$ stands for $[male, female]$. In the mid table, $[0, 1]$ stands for $[Black, White]$.

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 $[-1, 1]$. Following the optimization strategy in [10], 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.

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

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 [7]. Their comparison is shown in Fig. 4.

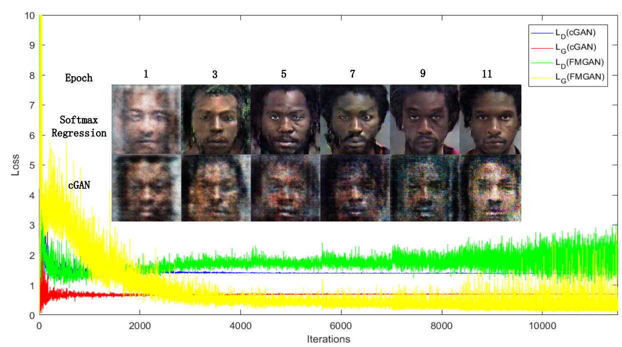

Figure 4. Comparison of generative performance between FM-GAN and cGAN during training. In the training process, one epoch equals to 1000 iterations.

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.

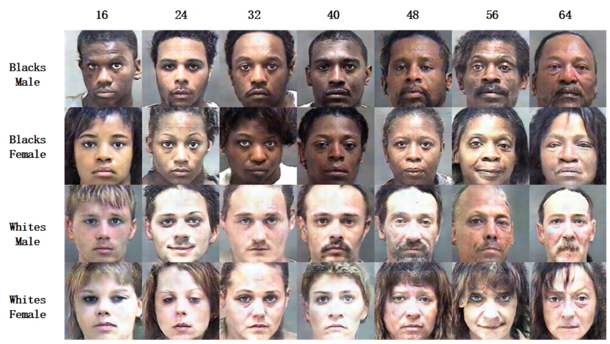

Figure 5. Examples of 128×128 synthetic images generated by our FM-GAN with the noise data randomly sampled for all faces and varying gender g sampled in [Male,Female], ethnicity e sampled in [Black, White], and age sampled with aging process.

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.

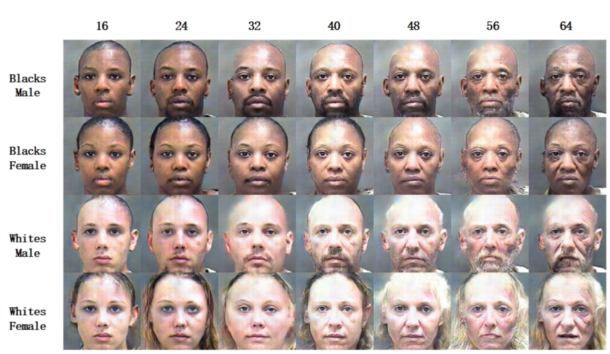

Figure 6. Examples of 128×128 samples with fixed identity(noise) and varying $g, e \in [0, 1]$ respectively in the vertical ,and $a \in [16, 24, 32, 40, 48, 56, 64]$ in the horizon

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. [12] and evaluate their performance on corresponding test set. Tab. 2 shows the comparison 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.

By analyzing the performance of synthetic dataset on each age, it is found that MAE in the range of young ages show better performance than in the range of old ages which nearly approaches the performance of original MORPH II dataset. The average MAE before the age of 40 is 6.139406 while MAE behind the age of 40 gets 10.331515. From the perspective of synthetic quality, older faces perform worse visual fidelity than young faces. On the whole, There is still plenty of room for improving fine-grained generation.

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 in the range of young ages.

Tab. 3 shows the performance comparison of different scale of G1 set added to S1 set. All results are tested on corresponding test set. Baseline is the performance of original train set without additional generated samples trained on AlexNet. 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

Table 3. Performance of different numbers of synthetic images added to MORPH Album II

Method	Age	Race	Gender
Base(AlexNet)	3.851798	0.985435	0.970392
scale:0.2	3.821998	0.979125	0.962331
scale:0.4	3.790397	0.980650	0.959916
scale:0.6	3.786226	0.977389	0.962496
scale:0.8	3.787912	0.977389	0.964794
scale:1.0	3.773675	0.980181	0.959235
CNN[13]	4.60	-	-
Multi-Scale CNN[14]	3.63	-	-
Soft softmax[12]	3.14	-	-

images obviously improve its performance on age estimation and has scarcely little impact on gender and ethnicity 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. Some models like Multi-scale CNN[14] and Soft softmax[12] which have been proposed and trained before have images with larger resolutions as input which could not be used for fine-tuning but becomes our works in the future.

The same process of verification is applied to FG-NET and CACD. The results are shown in the Tab.4 below. It should be noted that the results of CACD and FG-NET are directly using datasets for training VGGNet and AlexNet on age estimation without any other pre-trained initialization. The performance on CACD is not ideal for the reason of bad generation of face samples. Some databases like CACD and FG-NET could get better performance on age estimation if experimented on some pre-trained models.

Table 4. Performance of age on CACD and FG-NET datasets

Method	CACD	FG-NET
Baseline	5.605837	6.402655
scale:0.5	5.805193	6.149355
scale:1.0	5.964726	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. 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. In the future, on the one hand, face images with larger resolutions, especially e.g 224×224, are essential because many existing models in age estimation are experimented on the input of resolution 224×224. Fine-tuning on these models and compare the performance with them are the final goal. On the other hand, face images should be generated with higher quality and precise age. while generating complete faces, changes on face should be more consistent with the process of aging.

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 *ECCV*, pages 768–783. Springer, 2014.

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

[4] 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.

[5] 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.

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

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

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

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

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

[11] 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.

[12] 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.

[13] M. Yang, S. Zhu, F. Lv, and K. Yu. Correspondence driven adaptation for human profile recognition. In *CVPR, 2011 IEEE Conference on*, pages 505–512. IEEE, 2011.

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