Face Aging using Conditional GANs

Ashfaque Ahmad
Reg. no: 11704664
Computer Science & Engineering, Lovely Professional University, Punjab.
Submitted to: Usha Mittal
ashfak987ali@gmail.com

Abstract— Deep learning based approaches has gained very optimistic results in face recognition area. Face recognition is become very effective research topic and has a of a number of attainments. Also there are some researches for periocular recognition to overcome limitations of entire face recognition. The challenges of aging in periocular recognition has not gained attention after its achievements. Deep learning approaches are used to overcome many challenges of face recognition such as pose, expression, illumination and aging. Periocular images recognition under less restricted environments is the problem researchers faced in face recognition. But proposed approach has a new structure that can get efficient periocular recognition. This work focuses on the aging face recognition problems of entire face image based on a deep learning method, in particular, convolutional neural network. The proposed methodology gives a deep learning based approach for periocular recognition subject to aging. Using a CNN feature extraction and classification characteristic of deep learning gives an accurate and efficient recognition rate as compared to conventional method.

1. INTRODUCTION

Face Aging, a.k.a. age synthesis and age progression can be defined as aesthetically rendering an image of a face with natural aging and rejuvenating effects on the individual face.

Traditional face aging approaches can be split into 2 types:

- prototyping
- modeling

1.1 Prototyping approach

➤ Estimate average faces within predefined age groups

➤ The discrepancies between these faces constitute the aging patterns which are then used to transform an input face image into the target age group.

Pros - Are simple and fast
Cons - As they are based on general rules, they totally discard the personalized info which often results in unrealistic images.

1.2 Modeling approach

➤ Employ parametric models to simulate the aging mechanisms of muscles, skin and skull of a particular individual

Both these approaches generally require face aging sequences of the same person with wide range of ages which are **very costly to collect**.
These approaches are helpful when we want to model the aging patterns sans natural human facial features (its personality traits, facial expression, possible facial accessories, etc.)

Anyway, in most of the real-life use cases, *face aging must be combined with other alterations to the face*, e.g. adding sunglasses or moustaches.
Such non-trivial modifications call for global generative models of human faces.

This is where the Age-cGAN model comes in.
You see, vanilla GANs are explicitly trained to generate the most realistic and credible images which should be hard to distinguish from real data.

Obviously, since their inception, GANs have been utilized to perform modifications on photos of human faces, viz. changing the hair colour, adding spectacles and even the cool application of binary aging (making the face look younger or older without using particular age categories to classify the face into).

But a familiar problem of these GAN-based methods for face modification is that, the original person's identity is often lost in the modified image.

The Age-cGAN model focuses on identity-preserving face aging. The original paper made 2 new contributions to this field:

1. The Age-cGAN (Age Conditional GAN) was the first GAN to generate high quality artificial images within defined age categories.

2. The authors proposed a novel **latent vector optimization** approach which allows Age-cGAN to reconstruct an input face image, preserving the identity of the original person.

2. *Introducing Conditional GANs for Face Aging*

Conditional GANs (cGANs) extend the idea of plain GANs, allowing us to control the output of the generator network**.** We know that face aging involves changing the face of a person, as the person grows older or younger, making no changes to their identity.

In most other models (GANs included), the identity or appearance of a person is lost by 50% as facial expressions and superficial accessories, such as spectacles and facial hair, are not taken into consideration.

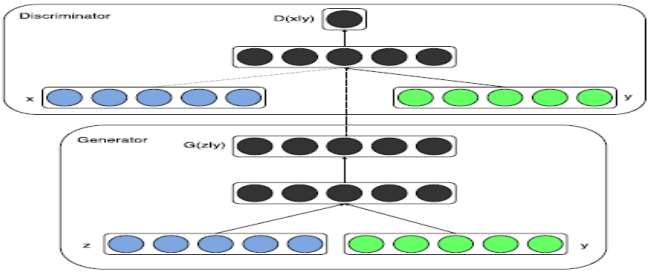
Age-cGANs consider all of these attributes, and that's why they are better than simple GANs in this aspect.

3. *Understanding cGANs*

GANs can be extended to a conditional both provided both the generator and discriminator networks are conditioned on some extra information, y.
y can be any kind of additional information, including class labels and integer data.
In practice, y can be any information related to the target face image - facial pose, level of illumination, etc.
The conditioning can be performed by feeding y into both the generator and discriminator as an additional input layer.
Some drawbacks of vanilla GANS:

1. Users have no control over the category of the images generated. When we add a condition *y* to the generator, we can only generate images of a specific category, using *y*.
2. Thus, they can learn only 1 category and it's highly tedious to construct GANs for multiple categories.

A cGAN overcomes these difficulties and can be used to generate *multi-modal* models with different conditions for various categories.
Below is the architecture of a cGAN:



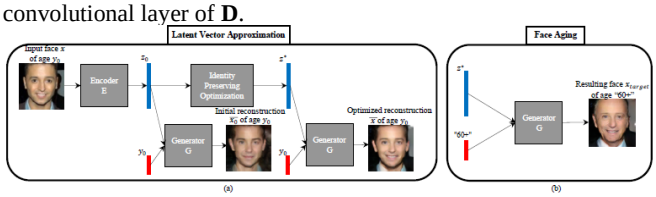
Let us understand the working of a cGAN from this architecture -

In the generator, the prior input noise pz(z)pzz, and y are combined to give a joint hidden representation, the adversarial training framework making it easy to compose this hidden representation by allowing a lot of flexibility.
In the discriminator, x and y are presented as inputs and to a discriminative function (represented technically by a Multilayer Perceptron).
The objective function of a 2-player minimax game would be as

minG maxD V(D, G) = Ex~pdata(x)[log D(x|y)] + Ex~pz(z)[log(1 − D(G(z|y)))]minG maxD VD, G = Ex~pdataxlog Dx|y + Ex~pzzlo g1 - DGz|y
Here, G is the generator network and D is the discriminator network.
The loss for the discriminator is log D(x|y)log Dx|y and the loss for the generator is log(1 − D(G(z|y)))log1 - DGz|y.
G(z|y)Gz|y is modeling the distribution of our data given z and y.
z is a prior noise distribution of a dimension 100 drawn from a normal distribution.

4. Architecture of the Age-cGAN

The Age-cGAN model proposed by the authors in the original paper used the same design for the Generator **G** and the Discriminator **D** as in **Radford, et al. (2016)**, "Unsupervised representation learning with deep convolutional generative adversarial networks", the paper which introduced DCGANs.
In adherence to **Perarnau, et al. (2016)**, "Invertible conditional GANs for image editing", the authors add the conditional information at the input of **G** and at the first convolutional layer of **D**.



The picture shows the face aging method used in Age-cGANs.

a) depicts an approximation of the latent vector to reconstruct the input image
b) illustrates switching the age condition at the input of the generator **G** to perform face aging.
Diving deeper into more technical details, now.

The Age-cGAN consists of 4 networks - an encoder, the FaceNet, a generator network and a discriminator network.

The functions of each of the 4 networks -

a) **Encoder** - Helps us learn the inverse mapping of input face images and the age condition with the latent vector z0z0.
b) **FaceNet** - It is a face recognition network which learns the difference between an input image **x** and a reconstructed image ˜xx~.

c) **Generator network** - Takes a hidden (latent) representation consisting of a face image and a condition vector, and generates an image.
d) **Discriminator network** - Discriminates between the real and fake images.
cGANs have a little drawback - they can't learn the task of inverse mapping an input image **x** with attributes **y** to a latent vector **z**, which is necessary for the image reconstruction: x = G(z, y)x = Gz, y.
To solve this problem, we use an encoder network, which we can train to approximate the inverse mapping of input images **x**.

A) The encoder network

- Its main purpose is to generate a latent vector of the provided images, that we can use to generate face images at a target age.
- Basically, it takes an image of a dimension of (64, 64, 3) and converts it into a 100-dimensional vector.
- The architecture is a deep convolutional neural network (CNN) - containing 4 convolutional blocks and 2 fully connected (dense) layers.

B) The generator network

- The primary objective is to generate an image having the dimension (64, 64, 3).
- It takes a 100-dimensional latent vector (from the encoder) and some extra information **y**, and tries to generate realistic images.
- The architecture here again, is a deep CNN made up of dense, upsampling, and Conv layers.

C) The discriminator network

- What the discriminator network does is identify whether the provided image is fake or real. Simple.
- It does this by passing the image through a series of downsampling layers and some classification layers, i.e. it predicts whether the image is fake or real.
- Like the 2 networks before, this network is another deep CNN.

D) Face recognition network

- The FaceNet has the main task of recognizing a person's identity in a given image.
- To build the model, we will be using the pre-trained Inception-ResNet-v2 model without the fully connected layers.
- The pretrained Inception-ResNet-v2 network, once provided with an image, returns the corresponding embedding.

5. Stages of the Age-cGAN

The Age-cGAN model has multiple stages of training. The Age-cGAN has 4 networks, which get trained as 3 stages which are as follows:

1) Conditional GAN training
2) Initial latent vector optimization
3) Latent vector optimization

A) Conditional GAN Training

➤ This is the first stage in the training of a conditional GAN.
➤ In this stage, we train both the generator and the discriminator networks.
➤ Once the generator network is trained, it can generate blurred images of a face.
➤ This stage is comparable to training a vanilla GAN, where we train both the networks simultaneously.

B) Initial latent vector approximation

➤ Initial latent vector approximation is a method to estimate a latent vector to optimize the reconstruction of face images.
➤ To approximate a latent vector, we have an encoder network.
➤ We train the encoder network on the generated images and real images.
➤ Once trained, it will start generating latent vectors from the learned distribution.
➤The training objective function for training the encoder network is the Euclidean distance loss.

Although GANs are no doubt one of the most powerful generative models at present, they cannot exactly reproduce the details of all real-life face images with their infinite possibilities of minor facial details, superficial accessories, backgrounds, etc.
It is common fact that, a natural input face image can be rather **approximated** than **exactly reconstructed** with Age-cGAN.

C) Latent vector optimization

➤ During latent vector optimization, we optimize the encoder network and the generator network simultaneously.
➤ The equation used for this purpose is as follows:

z∗IP = argminz‖FR(x) − FR(˜x)‖L2zIP* = argminz‖FRx - FR x~‖L2

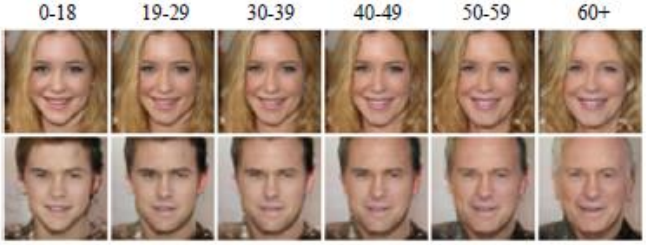
This equation follows the novel **"Identity Preserving"** latent vector optimization approach followed by the authors in their paper.

The key idea is simple - given a face recognition neural network **FR** which is able to recognize a person's identity in an input face image **x**, the difference between the identities in the original and reconstructed images **x** and ˜xx~ can be expressed as the Euclidean distance between the corresponding embeddings **FR(x)** and **FR(**˜xx~**)**.

6. Experiments on the Age-cGAN model

The authors trained the Age-cGAN on the **IMDB-Wiki_cleaned [3]** dataset containing around 120,000 images, which is a subset of the public **IMDB-Wiki** dataset **[4]**. 110,000 images were used for training of the Age-cGAN model and the remaining 10,000 were used for the evaluation of identity-preserving face reconstruction.

A) Age-Conditioned Face Generation



The above image illustrates examples of synthetic images generated by the Age-cGAN model using 2 random latent vectors **z** (rows), conditioned on the respective age categories **y** (columns).

7. Observations & Results

Reconstruction type	FR score
Initial Reconstruction (z_0)	53.2%
"Pixelwise" Optimization (z^*_{pixel})	59.8%
"Identity-Preserving" Optimization (z^*_{IP})	**82.9%**

Table 1. "OpenFace" Face Recognition (FR) scores on three compared types of face reconstruction.

➢ Table 1 presents the percentages of "OpenFace" positive outputs (i.e. when the software believed that a face image and its reconstruction belonged to the same person).

The results confirm the visual observations presented above.

➢ Initial reconstructions allow "OpenFace" to recognize the original person in only half of the test examples.

➢ This number is slightly increased by "Pixelwise" optimization but the improvement is very slight.

➢ Conversely, "Identity-Preserving" optimization approach preserves the individual's identities far better, giving the best face recognition performance of **82.9%**.

Now we shall cover the basic implementation of all the 4 networks - encoder, generator, discriminator and face recognition - using the Keras library**.**

8. Interesting applications of Age-cGAN

I. Cross-age face recognition

➤ This can be incorporated into security applications as means of authorization e.g. smartphone unlocking or desktop unlocking.

➤ Current face recongnition system suffer from the problem of necessary updation with passage of time.

➤ With Age-cGAN models, the lifespan of cross-age face recognition systems will be much longer.

II. Finding lost children

➤ As the age of a child increases, their facial features change, and it becomes much harder to identify them.

➤ An Age-cGAN model can simulate a person's face at a specified age easily.

III. Entertainment

➤ For e.g. in mobile applications, to show and share a person's pictures at a specified age.

The famous phone application FaceApp is the best example for this use case.

IV. Visual effects in movies

➤ It requires lots of visual effects and to manually simulate a person's face when they are older. This is a very tedious and lengthy process too.

➤ Age-cGAN models can speed up this process and reduce the costs of creating and simulating faces significantly.

9. References

[1] Radford, et al. (2016), "Unsupervised representation learning with deep convolutional generative adversarial networks"
[2] Perarnau, et al. (2016), "Invertible conditional GANs for image editing"
[3] Antipov, et al. (2016), "Apparent age estimation from face images combining general and children-specialized deep learning models"
[4] Rothe, et al. (2015), "DEX: Deep EXpectation of apparent age from a single image"
[5] Ludwiczuk, et al. (2016), "Openface: A general-purpose face recognition library with mobile applications"
[6] Szegedy, et al. (2016), "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning"
[7] Antipov, et al. (2017),