Age Progression/Regression by Conditional Adversarial Autoencoder

Michał Ostyk-Narbutt (1854051)
Vision and Perception Final Project
Prof. Fiora Pirri – Edoardo Alati

September 30, 2019

1 Introduction

For this project I decided to use the paper titled *Age Progression/Regression by Conditional Adversarial Autoencoder* which is an attempt of improving the currently used techinques for face-aging using Generative Adverserial Networks (GANs). These former methods are incapable of inferring the age i.e. 80+ given a current picture at an unknown age and a series of pictures at a very young age i.e. 5. The reason for this is that they most likely rely on learning transformations between age groups and require paired labeled samples in addition to the labeled query image. This paper approaches the problem from a different perspective without using paired samples. Namely, a conditional adversarial autoencoder (CAAE) is proposed that learns a face manifold, travereses it, and is able to obtain a smooth age progression and regression simultaneously.

2 Paper summary

2.1 Introduction

The main focus of this paper is automatic face progression and regression, otherwise known as face aging and rejuvenation from the generative modelling perspective. The aim revolves around rendering faces at different ages but preserving personazied features. Several factors of a face come into play i.e. expression, pose, resolution, illumination etc.

Most existing (pre-2017) require the availability of paired samples, and even long ranging life spain of images, and mostly focus on age progression. The large majority of solutions revolve around surface based modeling which focus on simulating textute removal/addition to faces. There is also a fair trade-off between preserving personality and at the same time inducing severe ghosting artifacts.

This paper proposes a conditional adversarial autoencoder (CAAE) network to learn the face manifold. By controlling the age attribute, it will be flexible to achieve age progression and regression at the same time.There are four benefits to CAAEs:

1. they both age progression and regression while generating photo-realistic face images

2. they do not use the popular group based learning hence paired smples are not required making this solution more flexible
3. the disentanglement of age and personality in the latent vector space helps preserving personality while avoiding the ghosting artifacts.
4. robust against variations in pose, expression, and occlusion

2.2 Generative Adversarial Networks

One of the biggest issues of GANs is that the training process is unstable, and generated images are noisy. Prior research on GANs has helped amend these problems with variations such Conditional GANs which transformed learning from unsupervised to semi-supervised. Deep-conv GAN introduced convolutional and deconvolutions layers to the generator and discrimiator networks. The undesirable and uncontrollable inherent property of GANs has to be controlled to ensure that the output face looks as the same person, preserving personality traits.

2.3 Traversing on the Manifold

Since modelling and subsequent of a high-dimensional manifold is difficult, a mapping is learned between it and a lower-dimensional space, called the *latent space*. As shown on Figure 1, the faces x_1 and x_2 are first mapped to the latent space by a convolutional Encoder E which extracts personal features z_1 and z_2 respectively. These are concatenated with age labels l_1 and l_2 to create points. These are disentangled in the latent sapce, hence aging preserves personality. Then through another mapping via a Generator G, those points are mapped onto the manifold $\mathcal{M}$, generating a images. Conversly this procedure is applied to optain regression and progression.

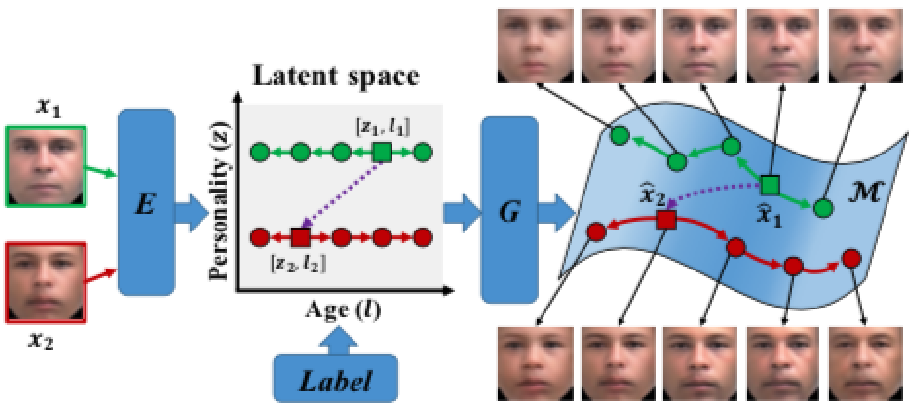

Figure 1: Face images are asssumed to lie on a manifold (M) , and images are clustered according to their ages and personality by a different direction.

2.4 CAAE network pipeline

2.4.1 architecture

Figure 2 present the structure of the CAEE network. The encoder E maps the input face to a vector z (personality). Concatenating the label l (age) to z, the new latent vector $[z, l]$ is fed to the generator G. Both the encoder and the generator are updated based on the $L2$ loss between the input and output faces. The discriminator D_z imposes the uniform distribution on z and the discriminator D_img forces the output face to be photo-realistic and plausible for a given age label.

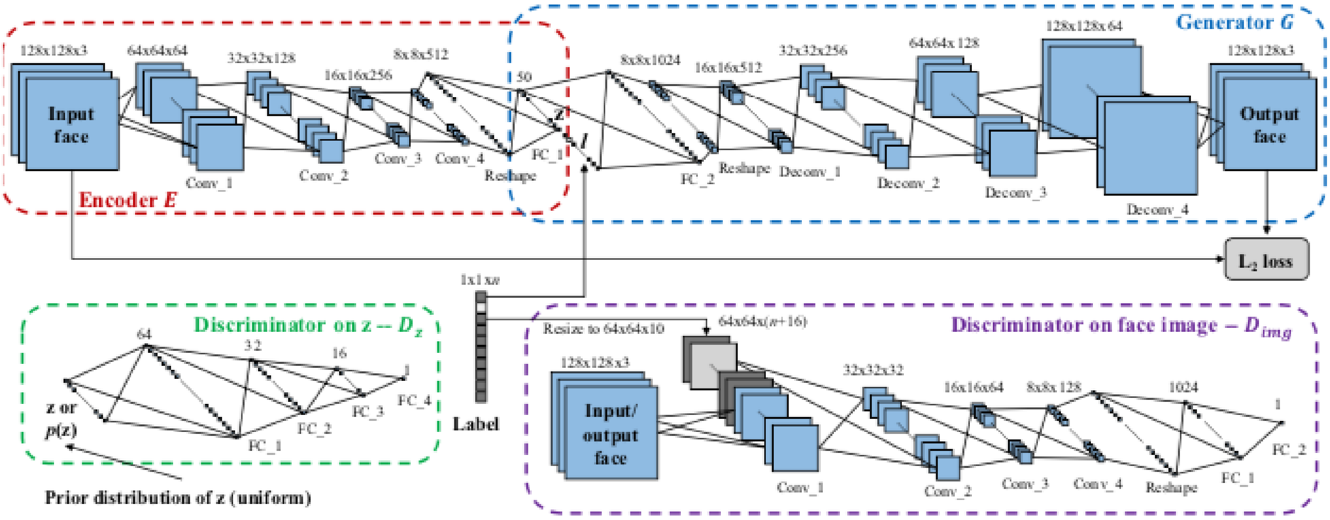

Figure 2: Architecture of the network

3 Approach

3.1 Tools

The key tools for this project were tensorflow-gpu version 1.14, which was setup on a virtual machine (VM) instance on google cloud platform (GCP).

3.2 Dataset

The *All-Age-Faces Dataset* [4] used was different than the dataset used in the original paper. It differed by using a single race (mostly Asian) and also augmented (slightly rotated images). This dataset compromises of 13000 images of aligned faces with 7000 being female and the remaining 5000, male.

The faces were divided into ten categories: 0–5, 6–10, 11–15, 16–20, 21–30, 31–40, 41–50, 51–60, 61– 70, and 71–80.

3.3 Training and Evaluation

Training was performed twice using learning rates $\alpha = 0.0002$ and $\alpha = 0.0005$ with the former yieldied better results. After 50 epochs of learning semi-realistic faces were obtained. During testing on the Encoder and Generator are active. The Encoder E will map the image to z, then by concatenting the age label to z, the generator G will generator a hopefully realistic face whilst preserving personality. The results are shown in Figure 3. The $EG - loss$ decreasing is a good indicator of the CAAE network performing better and reaching a stable rate of image generation. Other losses in the Figure 3 fluctuate more which could be an indicator of unstableness on those corresponding sub-networks.

The original github repository can be found her [2] wheras my viariation of it (along with the trained model) can be found here [3].

4 Results

Figure 4 illustrates examples of the reconstruction and age progression/regression testing. The Reconstruction loss shows promising results, however on the other hand the age progression performs good but could be improved using a more diverse dataset (other races than just Asian) and also a much much larger sample size.

(a) Reconstruction

(b) Testing (aging)

Figure 4: The first row in the reconstruction results (a) are testing samples that yield the testing results (b) in the age ascending order from top to bottom.

5 Conclusions

- CAAE is a network that achieves face aging and regression in one unified system.
- Personality traits are preserved meanwhile traversing different age categories (up and down) the manifold is possible and plausible.
- Four sub-networks are trained: E, G, D_z and D_img with each serving a diffrent role.
- The framework could potentially be used for other image generation tasks, where charecteristics of a generated image are controlled by a condition label.
- better results can be achieved with a more diverse dataset using techinques like population based training for finding the optimal hyperparameters.

References

[1] Zhifei Zhang, Yang Song, and Hairong Qi. "Age Progression Regression by Conditional Adversarial Autoencoder." IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. http://web.eecs.utk.edu/~zzhang61/docs/papers/2017_CVPR_Age.pdf

[2] Original repository https://github.com/ZZUTK/Face-Aging-CAAE

[3] transformed code + model https://github.com/Ostyk/Face-Aging-CAAE-project

[4] Exploiting effective facial patches for robust gender recognition https://github.com/JingchunCheng/All-Age-Faces-Dataset

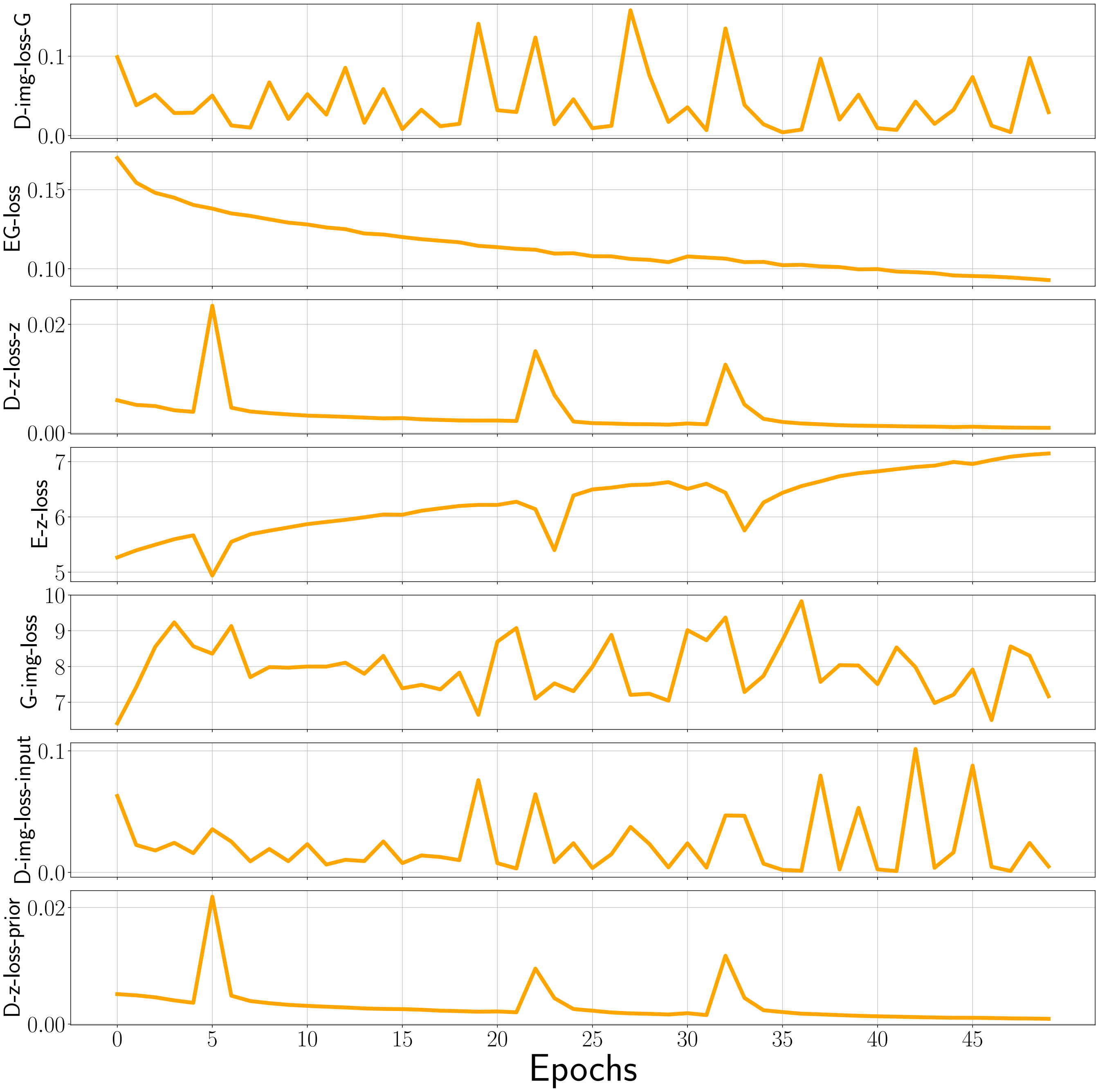

Figure 3: Training results loses over 50 epochs taken from tensorboard