Facial Expression Recognition and Synthesis using Generative Adversarial Network

Project Plan

Supervisor: Chan K P

Members: Wang Hanyan, Lo Man Chun

Introduction

Facial expression synthesis is the process of rendering face images with target expressions while preserving identity information. The generated images can be used in various areas including affective interaction, data augmentation, artificial agent, etc. Existing methods are usually conducted in a sequence to sequence manner which requires the availability of a video of varying facial expressions of a subject. However, the lack of such data in real life limits the application of these methods. Synthesizing facial expressions on limited faces remains an open problem.

Facial Expression Recognition means to distinguish facial expressions from face images, including happiness, sadness, anger, fear, disgust, surprise and neutral. It is important in many areas such as mental disease diagnosis and human social interaction detection. Although the system in the laboratory has very high accuracy, when it comes to the real world, the accuracy is only approximately 50% due to illumination variation, head pose and subject-dependance etc.

In recent years, the generative adversarial nets(GANs) [5] have received substantial attention. GAN consists of two models: (1) Generator G that produces images based on random noises and (2) Discriminator D that distinguishes real sample images between fake images generated by G. GAN establishes a minimax adversarial game between G and D, making the generated images look more and more real. Different variations of GANs have been applied successfully to age progression/regression [2], image super-resolution [7], image to image translation [6], etc.

In this project, we will use the idea of GAN to synthesize face images, i.e. given a subject *I* and a facial expression *E*, we should generate an image with identity *I* and expression *E*. Also, we will use GAN to recognize the facial expression of face images, i.e. given an image with identity *I* and expression *E*, we should distinguish the facial expression *E*. This project plan will propose the objectives of the study, discuss the approaches to achieve them and set up a schedule for conducting the research.

Objectives

The project aims to generate faces with target expressions and recognize facial expressions using the idea of GANs.

As for the synthesis part. The first goal is to try state of the art GANs to produce photorealistic faces and implement existing architectures like [1], [2] to synthesize faces on public datasets. The intermediate goal is to modify the architecture and loss function to achieve better qualitative results on unknown subjects. The final objective is to apply the model to videos, i.e. given a video of subject A and subject B. The model should transfer the facial expressions of B to A while preserving the identity of A. This objective is far more difficult than the above goals as it needs to ensure the continuity of frames in the video and the computing resource will be much higher.

As for the recognition part, to modify the discriminator to have higher accuracy in distinguishing the facial expression of public datasets and the synthesized faces under different variations such as illumination, head pose and subject-dependance. If time permits, a survey of quantitative results of the generated faces will be conducted.

Methodology

In the project, we will use HKU GPU farm since training GANs requires intense computing resources and HKU provides the platform where software is already installed. We will write Jupyter Notebook for training and demonstration purposes since it is free and easy to use. As for the frameworks and libraries, we will choose TensorFlow and Keras because they are well documented and there are enough tutorials online. Concerning the first goal, we will clone repositories from GitHub and run them on public datasets which are available online. For the intermediate goal, we need to dig into the source code and modify it. Meanwhile, we will read the latest papers related to the GANs area and integrate ideas into our model. This is a time-consuming process because we don't know whether the modification will work or not. We will upload the deliverables to the final year project server for others to see.

In the synthesis part, we will implement existing models and try on different public datasets. One architecture we will use is ICface [8] (see Figure 1). It uses the idea of encoder, GAN and several losses to reenact faces. According to the authors, it outperforms another popular face manipulation model GANimation [4]. We will first train on this model to see its effect.

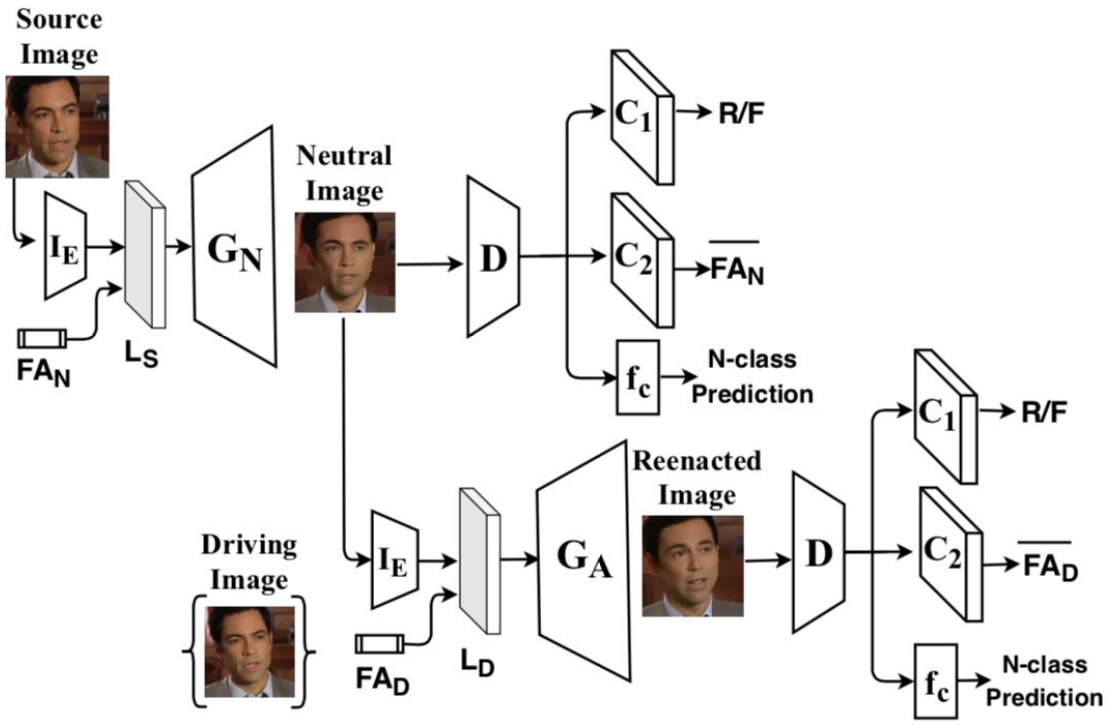

Fig. 1: [8] The whole model consists of four parts. The first is the image encoder I_E, it encodes a source image to latent space feature representation before feeding to a generator. The second is a neutralizer G_N, it accepts the concatenation of feature representation and a neutral attribute vector FA_N and outputs a neutral image with the front pose. The third is a generator G_A, which generates the reenacted image based on the feature representation of the neutral image and target attribute vector FA_D. The last part is a discriminator D, it distinguishes whether the reenacted image is real or fake.

Project schedule

29 Septemeber 2019	Deliverables of Phase 1 (Inception) • Detailed project plan • Project web page
13-17 January 2020	First presentation
2 February 2020	Deliverables of Phase 2 • Preliminary implementation • Detailed interim report
19 April 2020	Deliverables of Phase 3 • Finalized tested implementation • Final report
20-24 April 2020	Final presentation
5 May 2020	Project exhibition

Reference

[1] Photorealistic Facial Expression Synthesis by the Conditional Difference Adversarial Autoencoder
[2] Age Progression/Regression by Conditional Adversarial Autoencoder
[3] Geometry-Contrastive GAN for Facial Expression Transfer
[4] GANimation: Anatomically-aware Facial Animation from a Single Image
[5] I. Goodfellow et al, "Generative adversarial nets," in Advances in Neural Information Processing Systems, 2014, pp. 2672-2680.
[6] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-toimage translation using cycle-consistent adversarial networks,” arXiv preprint arXiv:1703.10593v6, 2017.
[7] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, ´A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang et al., “Photorealistic single image super-resolution using a generative adversarial network,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017, pp. 105–114.
[8] Tripathy, Juho, Rahtu, and Esa, “ICface: Interpretable and Controllable Face Reenactment Using GANs,” arXiv.org, 03-Apr-2019. [Online]. Available: https://arxiv.org/abs/1904.01909.