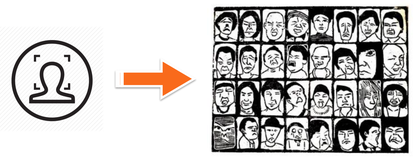
Facial Expression Generation System

Final Project - EECS 442 Fall 2018

Chengcheng Jin, Qiucheng Wu, Shucheng Zhong, Tongan Cai
University of Michigan – Department of Computer Science



Abstract

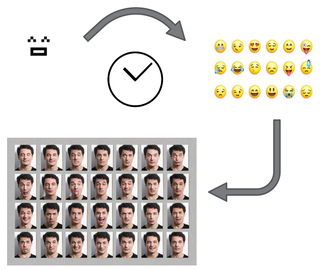
Face, as one of the most important first-glance impressions of human nature, has always been of great interests to Computer Vision. There're hundreds of thousands of face related databases that specialists in Computer Vision are constructing and utilizing. Facial expression is also a well-studied topic for Computer Vision as it has fascinating variations. In this project, we developed a novel system for facial expression generation with the help of Haar-Cascades Face Detection—which is used for collecting source images, and Generative Adverbial Networks—which was trained on both Cohn-Kanade Dataset and Japanese Female Facial Expression (JAFFE) dataset for comparative study. The collected images from detection module are them transferred and synthesized by the GAN module. Our tests on both modules are showing good performance and the system can achieve fair generation results from our input, and we're still working to perfect the fusion of two modules. We also realized the potential of the system can be real-time facial expression generation, for which we tested on several scenarios. Our system shows its promising value in generating facial expression in an interactive and efficient way, and a good prospect to be utilized for different purposes.

Initiatives

Using emotions has been a fashion in online communication for a long time. With the development of the digital technology, the emotion icons are evolving as well. Although these emotions are interesting, different users have to use the same series of emotions, which eliminates creativity and brings potential problems.

A natural thought is to replace the emotion icons with user emotions. However, there is a large need for user emotions every day, and it will be really a tedious job for users to take photos every time.
Therefore, with our knowledge in computer vision, we purpose a Facial Expression Generation System to solve the problem, which synthesizes arbitrary human emotions.

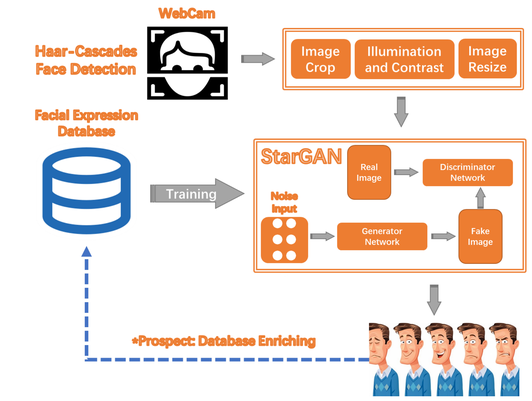
With single human image as input, our system can generate other human emotions within a short time. Therefore, our users can obtain and use their private customized emotions. When in production, this project can become the next generation of online emotions.



Methodology

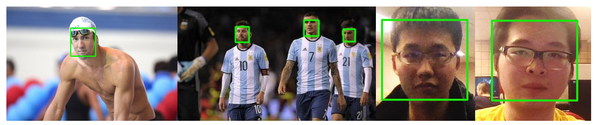
- System Flow

Our Facial Expression Generation System runs like follows:



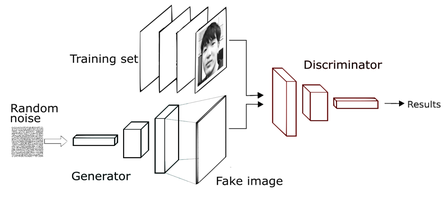
- Haar-Cascades Face Detection

Our team is developing a face detection interface to capture visitors' face photos from laptop's webcam and use them as the input for the next module based on Haar-Cascades. When detecting different kinds of features, Haar-Cascades will group the features into different stages and whenever it fails, the image will be regarded as negative. This algorithm has fast calculation speed and under good illumination condition, the accuracy of this algorithm is above 80%. The following is some detection output results of the software.

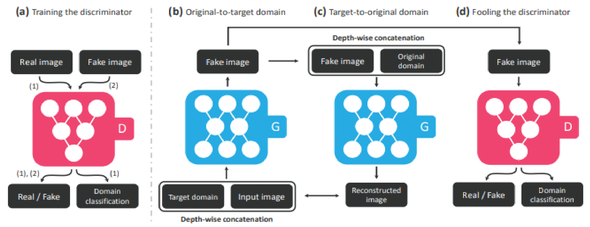


- Generative Adversarial Networks

We tried to use Generative Adversarial Networks (GAN) to generate our facial expression images. The basic idea of GAN is using a discriminator and a Generator. The job of the generator is generating fake images from original image, while the discriminator aims to tell the fake images from the real ones. By optimizing the generator and discriminator, generator can generate fake images that are hard to be distinguished from the real ones.



Traditional GANs has to train different model for each expression. Hence we took the reference of StarGAN. StarGAN is a 'Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation', which means it only needs one pair of Discriminator and Generator to complete the task. In order to achieve that property, we modified the loss function of GAN. Besides the loss of difference of real and fake images, we also measured the loss of classification error of generating wrong expressions. Additionally, a cycle-consistence loss was introduced to avoid the contents of fake images being changed from original image. The workflow is shown below:



Prospect

We trained our models on a laptop with Intel Core i7-4790 CPU, 16G Ram and a NVIDIA GTX 1070 GPU. We realized that we can further develop our future project in the following aspects:

- Real Time Facial Expression Generation

As we observe in above pictures, when we are trying to use photos detected by web camera as input to our generator, the result does not correspond to expectations. The underlying problems might be lacking large scale dataset and computational resources. Therefore, with more facial expression datasets like RaFD or FENET and high performance computing facilities, we expect the generator to perform better in real time.

- Synthesize Dataset for researches

A big number of researches in computer vision are related to facial expressions, bringing an urge need for large amount of input data. The future product can solve the problems by generating many kinds of facial expressions from a small dataset and, in return, enlarges the data amount for further researches.

Reference & Support

Y. Choi, M. Choi, M. Kim, J. Ha, S. Kim and J. Choo, "StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.*

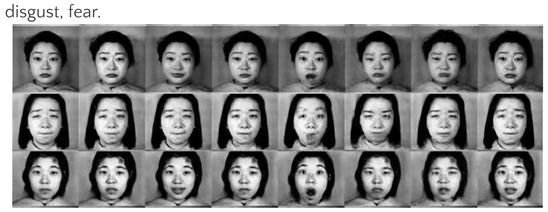
I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, pp. 2672-2680, 2014.

P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2001.*

Results

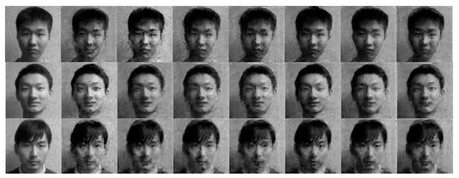
For the face detection function, the system gives a good performance.
The testing result of Cohn-Kanade Dataset at 100,000 iterations. From left to right is original image, neutral, anger, contempt, disgust, fear, happy, sadness, surprise.

The testing result of Japan female dataset at 100,000 iterations. From left to right is original image, neutral, happy, sadness, surprise, anger, disgust, fear.



Finally, we try to integrate our face detector and expression generator and results are shown below. We use the model from Japan Female dataset.

From left to right is original image, neutral, happy, sadness, surprise, anger, disgust, fear.



The results are rather awful compared to the result from the testing of our model. Some expressions are more obvious such as anger. We think the root cause is the lack of data that cannot cover large cases of potential input.