# Controlling the Features on Images Generated by a GAN

## Session : August - December 2020


B Akhil
PES1201802061
akhil.b2111@gmail.com

Surya Dutta
PES1201800674
suryadutta240@gmail.com

Hemanth R
PES1201802038
rhemanth818@gmail.com

Dr.T S Chandar
chandarts@pes.edu



**Abstract**

Generative Adversarial Networks(GANs) have been widely used to generate photo realistic images but their capabilities in image-to-image translations have not been explored properly. Applying such capabilities of GANs in the context of human faces can be very useful in building more robust models for various purposes such as face detection or emotion detection. This study concentrates on synthesis of conditional facial expressions. We present a CycleGAN model for learning to translate an image from a domain conditioned on a given emotion of facial expression to a different emotion of facial expression in the absence of paired examples.

## Contents

## 1 Motivation

Detection of human emotions has been long explored because of its applicability in various domains such as assisted living, health monitoring and real time crowd behavior tracking. Deep neural networks require an abundance of data, and its most notable successes have been in areas where this condition is met. A category of major success is supervised learning where large labeled datasets exist. GANs provide a mechanism to create such labelled data.
Photo realistic facial expression synthesis can be widely applied to face recognition, entertainment, virtual and augmented reality and computer graphics but this is much more challenging in part due to the scarcity of large labelled paired datasets I.e. where the same person is observed with different facial expressions.

## 2 Problem Statement

We propose a model which can control the features on the images produced by a GAN. We try to control the facial expressions in the images produced by the GAN.

## 3 Literature Survey

### 3.1 Generative Adversarial Networks (GANs)

In 2014, Goodfellow introduced the notion of Generative Adversarial Networks (GANs) which offered a new level of impact for generative models. Under the GAN framework, a generative model competes with a discriminative adversary. The discriminative model determines whether a sample is generated or a data example, while the generative model attempts to fool it. These adversarial models are trained simultaneously. In some ways, this model can be interpreted as a two player minimax (zero sum) game.

### 3.2 StyleGAN

As we know , GANs are effective in generating high quality images. Most of recent improvements have been made to discriminative models in an effort to train more effective generator models, although less effort has been put into improving the generator models itself. The StyleGAN is an extension to the GAN architecture that proposes large changes to the generator model, including the use of a mapping network to map points in latent space to an intermediate latent space, the use of the intermediate latent space to control style at each point in the generator model, and the introduction to noise as a source of variation at each point in the generator model. Due to these, The resulting model is capable not only of generating impressively photorealistic high-quality photos of faces, but also offers control over the style of the generated image at different levels of detail through varying the style vectors and noise. The use of this style vector gives control over the style of the generated image. Stochastic variation is introduced through noise added at each point in the generator model. Thus, styleGAN has heavily influenced works on synthesizing human faces.

### 3.3 Conditional GANs

The development of Conditional-GANs by Mirza (2014) suggested that we could take advantage of natural one-to-many mappings to produce a conditional predictive distribution for the discriminator D/generator G network described above. This framework consists of both D/G models conditioned on some auxiliary information y. This is represented by the following objective function: minGmaxD(D, G) = Ex pdata [log(D(x|y)] + Ez pz(z) [log(1 - D(G(z|y)))] Conditional GANs extended GANs conditioning the generator or discriminator on some extra information.

### 3.4 CycleGANs

The CycleGAN model was described by Jun-Yan Zhu, in the 2017 paper titled "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks." The benefit of the CycleGAN model is that it can be trained without paired examples. That is, it does not require examples of photographs before and after the translation in order to train the model. This provided a more effective way than adversarial autoencoders as paired examples are not needed for training unlike the former. Generator models are regularized to not just create new images in the target domain, but instead translate more reconstructed versions of the input images from the source domain.

### 3.5 Datasets

FER2013 available online at Kaggle consists of 28,709/7,178 train/test 48x48 pixel grayscale images of faces annotated with the emotion of facial expression as one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy. 4=Sad. 5=Surprise, 6=Neutral). It is a dataset of unpaired images.

## 4 Methodology/Working Principle

### 4.1 Conditional Cycle-Consistent GANs

Condition the GAN on the emotion of the facial expressions.. Our goal is to learn a mapping function between a domain X and itself conditioned on the emotions of facial expressions Y = 0, 1,..., k. Hence, given an image to x0 belongs to X annotated with y0 belongs Y and given a desired emotion of facial expression y1, we want to translate x0 into x1 having expression y1, i.e. x1 = G (x0 \ y1), where G is the conditional mapping we want to learn or generator. Also, we split such generator into Genc, the encoder responsible to encode a face image into its latent representation, i.e. z0 = Genc (x0), and Gdec, i.e. decoder responsible to perform the image-to-image translation given the desired facial expression label and the latent representation of the image.
x1 = Gdec (z0/y1) = Gdec (Genc (x) | y). In addition, we would like to introduce one adversarial discriminator D(x/y) to distinguish true images conditioned on true facial expressions and translated images conditioned on desired facial expressions.

### 4.2 Network Architecture

We adopt U-NET generator where for Genc encodes face images into latent vectors through four convolutional layers and Gdec takes as inputs 1 x 1 x 512 latent vectors and one-hot label vectors to decode them through three deconvolutional layers. . For the multi-task discriminator we use four convolutional layers followed by fully connected layers and sigmoid activations.

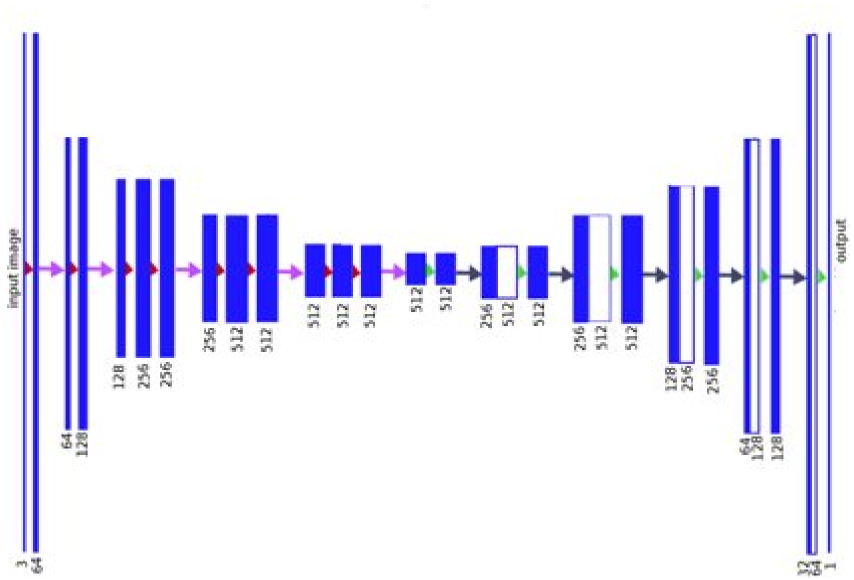
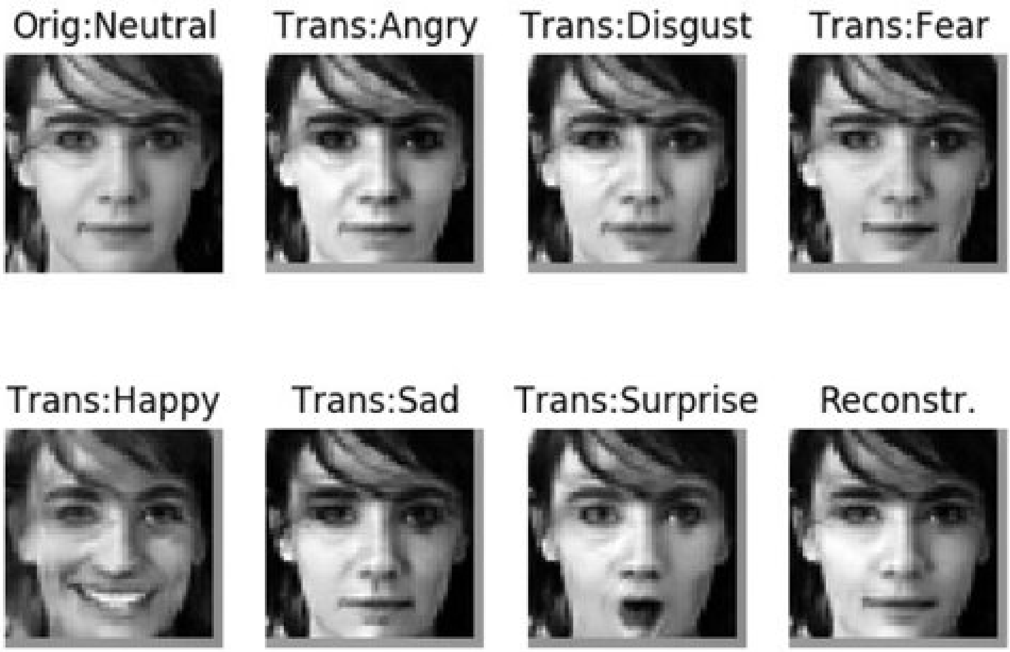



Figure 1: Block diagram for generator model

### 4.3 Training Details

For each train image x0 E X annotated with facial expression yo E Y, we extract the latent vector zo = Genc (xo), and then generate all the possible 6 face images conditioning on the remaining 6 class labels to train the discriminator and the generator.

## 5 Results and Discussions




The results are qualitatively observed after training the model for 150 epochs.
As visible from the image the model synthesises a good estimation of human faces with different facial expressions. One observation from the results is that the estimation for happy, angry, surprise expression is much better compared to that of other expressions and the estimation of Disgust is with the lowest accuracy. This can be attributed to the imbalanced nature of the dataset.

## 6 Conclusions

We provide a CycleGAN model to synthesise realistic face images conditioned on the facial expressions in the absence of paired examples. Qualitative results for the same are provided.

## 7 Future Scope

Improvements can be done to the network architecture to make the model more efficient like having progressive growing generator models to produce better outputs. By using a more balanced dataset the results in every class of emotion can be seen more vividly. Research is being done to translate human movement to Augmented reality.

## References


[1] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, *Generative adversarial nets.* Montreal,QC H3C 3J7,2014. https://arxiv.org/abs/1406.2661

[2] Alec Radford, Luke Metz, Soumith Chintala, *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks.*[v2] ICLR 2016. https://arxiv.org/abs/1511.06434

[3] Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen, *Progressive Growing of GANs for Improved Quality, Stability, and Variation.* [v3] ICLR Feb 2018. https://arxiv.org/abs/1710.10196

[4] Mehdi Mirza, Simon Osindero, *Conditional Generative adversarial networks.* ICCV 2014. https://arxiv.org/abs/1411.1784

[5] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros, *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks* [v7] ICCV August 2020. https://arxiv.org/abs/1703.10593

[6] https://machinelearningmastery.com/what-are-generative-adversarial-networks-

[7] https://towardsdatascience.com/understanding-generative-adversarial-networks-gans-cd6e4651a29

[8] https://blog.usejournal.com/the-rise-of-generative-adversarial-networks-be52d424e517


## 8 Appendices

This is the GitHub link of our project work - https://github.com/surya-dutta/Mini_project_GANs.git