Sketch to Realistic Face Generation with Facial Expression Synthesis

Guides

Dr. Shiv Ram Dubey
Dr. Snehasis Mukherjee

Upendra Reddy
(S20160010038)

K. Sai Prasanna Kumar
(S20160010047)

OUTLINE

- Introduction
- Literature review
- Deliverables
- Previous Work Done
- Work Done
- Results
- Dataset
- Summary & Conclusion

INTRODUCTION

- **Aim:** To create a facial expression modifier from sketch using a deep learning method called Generative Adversarial Networks (GANs)[1][2] trained using images of different Facial Expressions and sketches.
- Image to Image translation is task of changing a particular aspect of a given image to another.
    - Example, changing the facial expression of a person or changing physical features of a person.
- StarGAN [4] ,Pix2Pix[8] can be useful in above translation for different domains like facial attribute transfer and a facial expression synthesis tasks.
- This modifier is trained on multiple datasets on different domains (disgusted, angry and fearful, neutral, happy, sad, surprised) within a single network.

DELIVERABLES

Module-1 → 1. Implementing DcGAN
2. Dataset Generation

Module-2 → Facial Expression Synthesis using StarGAN

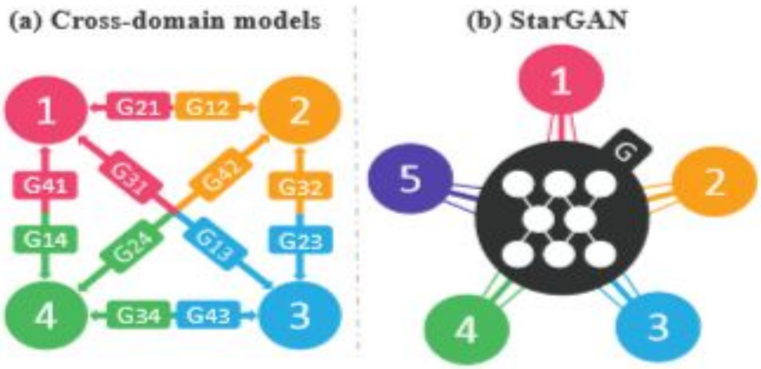
Module-3 → Extending StarGAN architecture

Module-4 → Sketch to Face expression synthesis

LITERATURE REVIEW

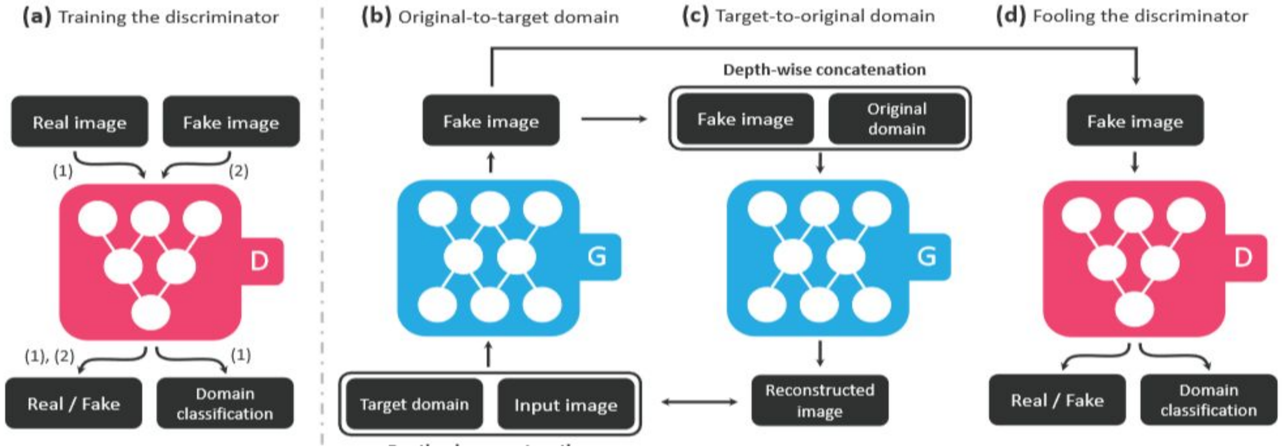
Cross Domain models vs StarGAN model

- Previous implementations like Pix2Pix[9] have leveraged the power of cGAN's to learn mappings from one domain to domain but However, existing models are both inefficient and ineffective in such multi-domain image translation tasks.
- The idea of StarGAN[5] is instead of learning a fixed translation, StarGAN generator takes both image and domain information as inputs during training.
- To achieve this, we train generator to translate an input image x into an output image y conditioned on the target domain label c,G(x,c)→y.



Previous Work Done

- StarGAN has the generator network composed of two convolutional layers with the stride size of two for downsampling, six residual blocks, and two transposed convolutional layers with the stride size of two for upsampling.
- Instance normalization [7] for the generator is used but no normalization is used for the discriminator.
- PatchGANs is used for the discriminator network, which classifies whether local image patches are real or fake.

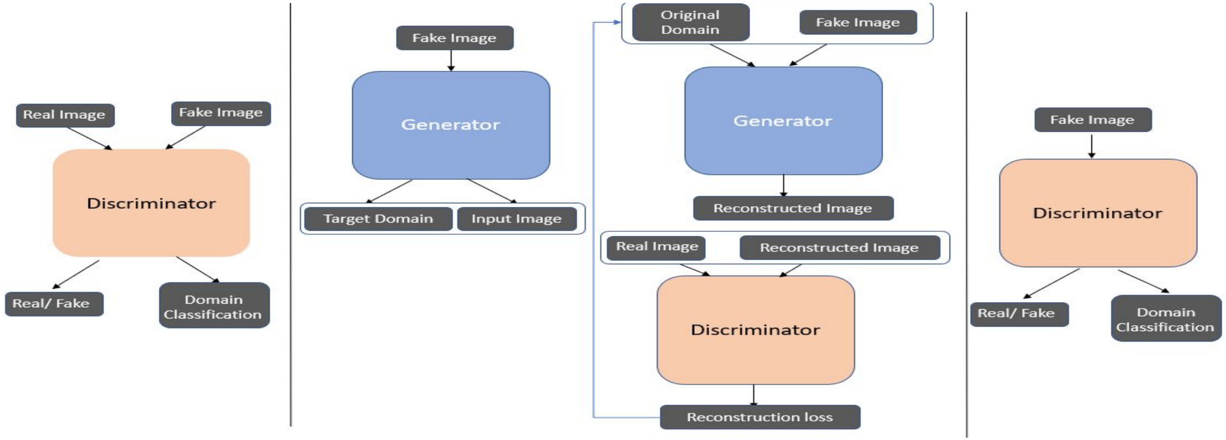


[5]

Previous Work Done

Extended StarGAN model:

- One of the bigger challenges is to preserve input image features in the reconstructed image.
- In the new architecture for preserving the identity of input image in the reconstructed image we are using a Discriminator between input generator and reconstruction image generator.



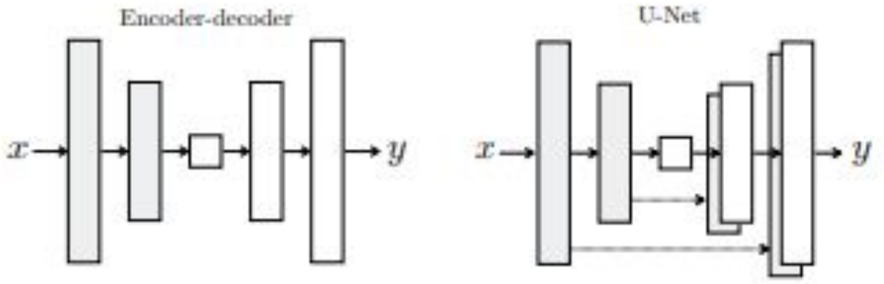
Results

Results Obtained on Extended StarGAN model

Work Done

Pix2pix Model:-

- Supervised approach that learns mappings from Pair-wise images using appropriate loss functions for image to image translation.
- Generator in Pix2Pix:
    - Uses U-Net shaped Encoder Decoder Network with skip connections between mirrored layers in Encoder-Decoder Network.
    - Useful to deal with low-level information shared between input and output



[8]

Work Done

- Discriminator in Pix2Pix:
  - Uses PatchGAN architecture – that only penalizes structure at the scale of patches and It is a form of texture/Style loss.
  - PatchGAN has fewer parameters, runs faster, and can be applied to arbitrarily large images.

Loss Functions:-

- Objective of conditional GAN is

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z))],$$ [8]

- Final objective with L1 loss-

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G).$$ [8]

Datasets

Facial Expression Dataset:

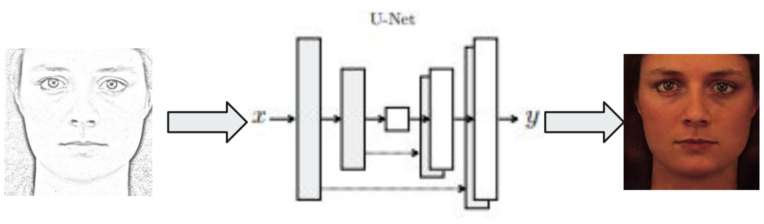
- This Dataset has been created by combining datasets KDEF, cohn-kanade[2], MUG [3] with all images being all front facing with labelled expressions.
- Totally, there are about 1200 images with expressions: neutral, happy, sad, surprised,disgusted, angry and fear.
- Likewise, all of them are preprocessed into the size of 256x256 with face being detected by dlib frontal face detector and cropping into the required size.
- 90% of data is used for training and 10% is used for testing.

Sketch-Image Dataset:

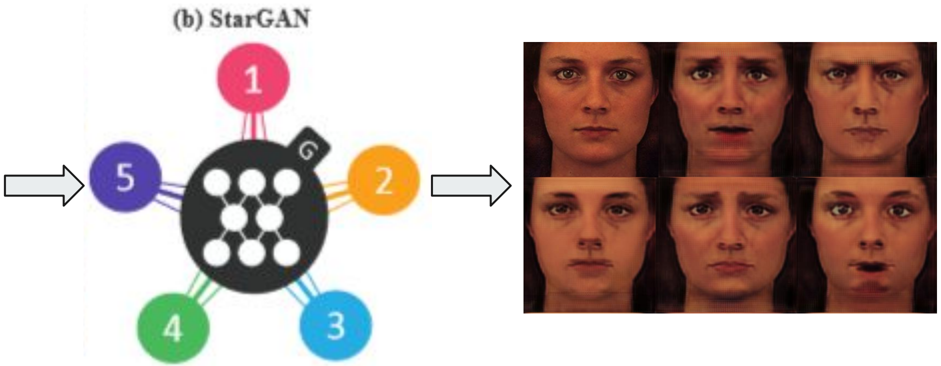
- Combined datasets of KDEF and CUHK images with photo taken in frontal pose with neutral expression along with sketches drawn by an artist.
- There are about 328 images in combined dataset with photos and sketches and all are pre-processed into size of 256x256 as above mentioned.

Training

- The Sketch - Image Dataset is used to train the pix2pix model. The Dataset contains 328 Sketch - Image pairs of realistic faces.
- The pix2pix model is trained for 300 epochs of the sketch - Image dataset on Tesla K20 GPU.
- The pix2pix model generates realistic face images on giving Sketch image as Input.
- The output of pixpix model which is a realistic neutral face image is used as input to extended stargan model which generates realistic facial expressions.



[8]



[5]

Results:-

<u>Result</u>:- **Sketch to Realistic Facial expression synthesis**

Results

- A face expression classifier is trained to classify the generated face expression with the target expressions.
- This classifier is used to classify the generated face expressions from the initial starGAN model. The classifier has performed well in classifying the face expressions and has 87.23 % accuracy for classifying expressions on the testing data.
- Since the modified StarGAN model is just preserving the input facial features. There's no drastic improvement in accuracy in classifying the generated expressions by the classifier.

References

1. Ian Goodfellow. NIPS 2016 tutorial: Generative adversarial networks. CoRR, abs/1701.00160, 2017
2. T. Kanade, J. F. Cohn, and Yingli Tian. Comprehensive database for facial expression analysis. In Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition.
3. https://mug.ee.auth.gr/fed/
4. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in Advances in neural information processing systems, 2014, pp. 2672–2680.
5. StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation Yunjey Choi, Minje Choi. Munyoung Kim, Jung-Woo Ha. Sunghun Ki, Jaegul Choo
6. Conditional Generative Adversarial Nets Mehdi Mirza. Simon Osindero.
7. Sergey Ioe and Christian Szegedy. Batchnormalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32Nd International Conference

References

8. Pix2Pix Image-to-Image Translation with Conditional Adversarial Networks, Isola, Phillip and Zhu, Jun-Yan and Zhou, Tinghui and Efros, Alexei A

9. X. Wang and X. Tang, “Face Photo-Sketch Synthesis and Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 31, 2009.

THANK YOU