Sketch2Face: Using CycleGAN to Produce Photo-Like Images from Unpaired Sketches

Jerry Meng (zmeng90@stanford.edu), Connor Quinn (ctquinn@stanford.edu), Danny Takeuchi (dtakeuch@stanford.edu)

Problem Definition

In both police TV dramas and real world criminal cases, law enforcement often only possesses witness description-based sketches to guide their search for potentially harmful individuals. This motivated us to create a deep learning framework that will reverse engineer photo-like images from relatively basic facial sketches. We employed the recent Cycle Generative Adversarial Network (CycleGAN) technique in the hope that the adversarial nature of the GAN will generate even more realistic photos, succeeding in training such a model with unpaired sketches and photos.

Dataset and Features

- 501 Image-sketch pairs to run our neural style transfer and Pix2Pix baselines from the CUHK Face Sketch Database (CUFS), all in 256x256 resolution. We used these sketches to power our final CycleGAN as well
- 1800 randomly sampled 256x256 resolution face photos from Nvidia's Flickr-Faces-HQ (FFHQ) faces dataset
- We used our images as raw RGB input to our CycleGAN, but employed a Mask-RCNN to crop out backgrounds from the sketches and photos.

Results

- To the right are two test set inferences.

- While not our emphasis, we also performed quantitative analysis on the two models via inception scores:

Input Pix2Pix Gen Default

Model: Mean, Std. dev

Pix2Pix Gen: 1.62, 0.205

Default Gen: 1.60, 0.223

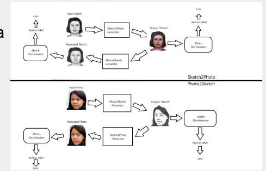
$$IS(G) = e^{E_{x \sim p_g}[KL(p(y|x)||p(y))]}$$

Models

Neural Style Transfer: Our baseline is a paired photo-sketch Neural Style Transfer network, which transfers the style of a photo to the paired sketch. This model struggled to color sketches' faces with consistent skin tones.

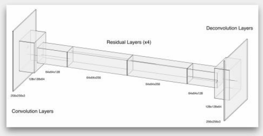
Conditional GAN (a.k.a. Pix2Pix): Using the same paired data, this model was able to very effectively translate paired sketches to images. However, with only a small dataset of sketches from one artist, we concluded the model simply overfit the training images.

CycleGAN: To solve this paired data scarcity problem, we employed a CycleGAN as our final model, as it learns from unpaired photos and sketches. CycleGANs process two categories of data, X and Y. Both have a generator which tries to map X to Y or vice versa, and a discriminator that tries to tell it apart (top image). We created two types of generators:

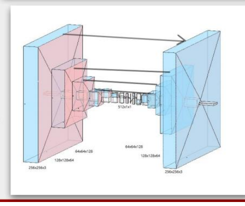


$$J = J^{(D1)} + J^{(G1)} + J^{(D2)} + J^{(G2)} + \lambda J^{cycle}$$

1. The original CycleGAN version (middle image)
2. A custom-made Pix2Pix version (bottom image)



- Both are based on a set of convolutions followed by a set of deconvolutions, with residual skips.



- We also tested various "PatchGAN" discriminator sizes from the original CycleGAN paper.

Analysis & Discussion

Model Performance:

- From a purely inception score-based viewpoint, the Default and Pix2Pix versions of the CycleGAN yielded similar performance.

- The Pix2Pix architecture trained more quickly however, requiring about half the steps to achieve solid performance, and generally yielded more realistic skin colorations.

- The Default architecture trained more slowly, but outputted lusher skin tones and avoided background fragmentation more consistently.

Error Analysis:

- Despite cropping out backgrounds with Mask-RCNN, background fragmentation persisted in both models due to imperfect cropping

- Red eyes and eye-area makeup also seemed to persist into the generated photos from the training set

Conclusion & Future Work

- We are satisfied with our model's performance, though also recognize the potential for improvement in generated image quality.
- To improve this project, we'd expand out dataset and potentially look into super resolution GANs as a means of sharpening our generated images.
- Beyond this, we are interested in less obviously paired translation tasks using CycleGANs, such as generating parental estimates of a given face.

References

- Jun-Yan Zhu*, Taesung Park*, Phillip Isola, and Alexei A. Efros. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", ICCV, 2017.
- Isola, Phillip, et al. Image-to-Image Translation with Conditional Adversarial Networks. BAIR Laboratory, 2017.