CSE455 final report

Shengjian Chen, Hung Lo

1. Introduction

In this final project, we first build a small dataset of Shiba Inu with around 300 images. Then we try out different GANs to make use of such a small dataset to get reasonable results. Basically, we have tried 3 ideas: providing extract information like edges with Pix2Pix, transferring human features to dogs with AttGAN and training on single image with SinGAN. Code for this project is available at https://github.com/tony23545/Shiba_GAN.

2. Dataset

There are two parts in the dataset. Images in first part are crawled from Flickr with whole body and different backgrounds. The second part are from the same dog on Instagram with only faces. We develop a script to manually crop all images to 128*128 with dogs at the center.

Fig 1. Left: sample image from the dataset. Right: DCGAN results.

We firstly train a DCGAN on our dataset as a baseline. We find that this network generates nothing interesting but some low quality images.

3. Pix2Pix

To make use of our small dataset. One way we can think about is to provide extract information. For example, edges. We adapt the Pytorch implementation of Pix2Pix[1] to our dataset. We use two kinds of edge detectors: Holistically-Nested Edge Detection (Hed) from [4] and Canny edge detector discussed in class. We then train the model on two parts of the dataset with different edge detectors respectively. Fig 2 illustrates results from different settings.

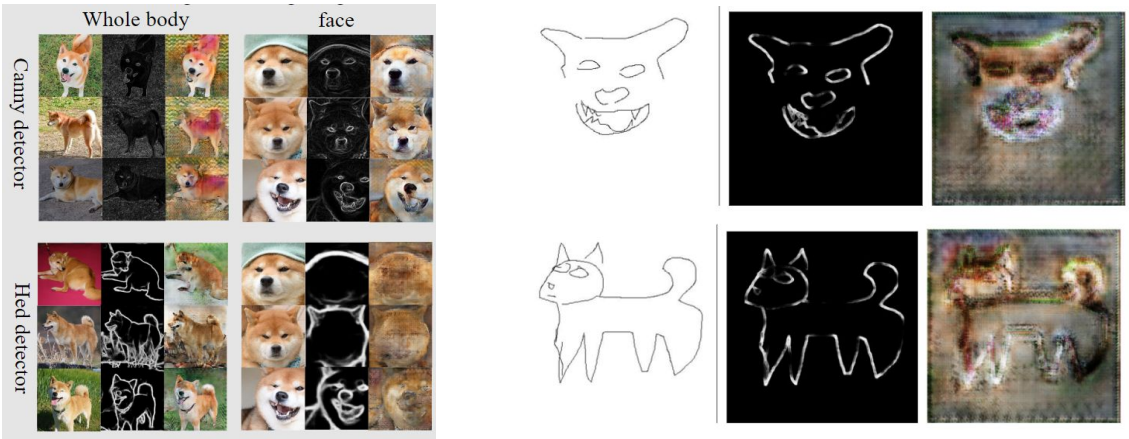

Fig 2.Left: Results of Pix2Pix. Right: Customized input.

We also develop a UI to receive customized input edge image from users and generate dog image with trained model.

Comparing the results above, we find that Canny detector provides more weak edges and capture details quite well, so it works well of face datasets. On the other hand, it will fail in whole body datasets, because it is too sensitive to the noise in the background. However, the Hed detector works in contrast. It dismisses weak edge in the picture and amplify the main edge. Therefore, the whole body will perform well, but face is bad due to less edge in face. In future research, we can try to combine these two edge detectors to increase the outcome performance. Results of the customized inputs are much worse than the training images. We think human drawings are too abstract and lack of details compared with the training edge images. So the decrease in the performance is expected.

4. Other GANs

AttGAN [2]

This GAN is usually used for human facial expression editing. We download the pretrained weights trained on CelebA dataset and apply to our dog dataset. We find that some of the attributes can be directly transferred to dogs from human reasonably.

Fig 3. AttGAN results.

SinGAN [3]

SinGAN goes to an extreme to train GAN with single image by digging deep into different scales of features. Images used in the original paper usually contains large landscape but contents in our images are so simple that the model doesn't work well. But one good thing is that it generally really high resolution images. We also try to extend single image to multiple images but see no improvement.

Fig 4. SinGAN result. Left: training image. Right: sample output.

Reference

[1] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In CVPR, 2017. 2

[2] Zhenliang He, Wangmeng Zuo, Meina Kan, Shiguang Shan, and Xilin Chen. Arbitrary facial attribute editing: Only change what you want. arXiv:1711.10678, 2017. 2

[3] Tamar Rott Shaham, Tali Dekel, and Tomer Michaeli. SinGAN: Learning generative model from a single natural image. In ICCV, 2019. 3

[4] S. Xie and Z. Tu. Holistically-nested edge detection. In ICCV, 2015. 1, 2, 4.