Final Project presentation Notes

Hi!! How are you guys doing? I am Yunting Chiu. Me and my teammate, Huong Doan(洞)

欸pic = epoch

**Start**

What does image inpainting mean?

Image Inpainting is a task of reconstructing missing regions in an image.

Which is the correct image inpainting in our model?

The answer is all, because PICNet can produce 50 different images from a single image.

**Introduction**

Early methods mainly focus on the similarity between the outputs and the ground truth. There are too many methods focusing on reconstructing the original image during the training. (For example, many methods' final goal is to recover images that are similar to the ground truth).

That is, one image only generate one output

(PICNet introduces random noises for generating diverse results with a deep generative network.

The challenge is : During the training, only one "ground truth" is available. How does the discriminator work?

Solution: It ranks the top 10 samples from the discriminator, which is close to the ground truth.)

PICNet’s final goal is to generate a wide range of possible results with different structure, color, and texture

**Related Works - T-MAD**

T-MAD is another image inpainting model. Also, it is a newer paper than PICNet.

* Structural Similarity Index (SSIM) - the higher the better
* Peak Signal-to-Noise Ratio (PSNR - the higher the better)

**Methodology**

Ig: the original image (inpainting image)

Im (the masked partial image)

Ic: complement partial image : included the original missing pixels

so Ic is the converse of Im, we will show the outputs of Im and IC in the following slide.

(In order to have a distribution to sample from, the probability of taking the sample is given Im then to take Ic )

This neural network consists of two parallel pipelines. The **yellow** line (reconstructive) merges data from Im and Ic, which are only used for training purposes. The blue (generative) pipeline estimates the conditional distribution of hidden regions, which can then be sampled during testing.

**Approach - Datasets**

Flickr-Faces-HQ Dataset (FFHQ): Flickr is a online album

The dataset consists of 70,000 high-quality face images from low-quality to 1024×1024 resolution

This dataset contains people of different ages, races, and dress styles, such as glasses, sunglasses, and hats.

(Selfie)

**Art work images:**

The authors used most of the popular image’s datasets (CelebA, Paris, Places2, and ImageNet) and they basically just did not use the art images so the art images will be interesting for us to do image inpainting.

Because the original images are large in size, we must first downsample them.

**Strength**

2. We want to verify it because we can see the outputs of people's faces are excellent in their paper, particularly in large-hole input.

3. Outputs are natural, realistic-looking. Even though our latest in FFHQ dataset’s epoch is 41. (two eyes, one mouth, two ears….but unclear)

4. One image can generate 50 diverse images with plausible content

**Limitation**

1. We should modify before running the code. There is one SyntaxError.

**async is no longer in cude()**

non\_blocking (bool):

If True and the source is in pinned memory, the copy will be asynchronous with respect to the host. Otherwise, the argument has no effect. Default: False.

1. free colab will be crashed after 12 hours, colab pro will be crashed after 24 hours
2. The maximum epoch is 41. I have tried more than 20 times to run the training model. Usually, the epoch does not surpass 30.
3. Celeba is one of the datasets used by the authors, and it only has 24813 images compared to ours... .....the training pace is slow. Normally, one epoch takes about 2-3 hours.
4. Based on our computing environment, the small image size is better than a high quality image. That’s why we need to downsample Art work’s images at the beginning.

Houng is going to take more detail of our experiments

Last slide of mine:

rec = reconstruction image

discriminator is a classifier