Progress update - I'm Something of a Painter Myself

Xutao Cao (xutaoc2) & Dingjia Chen (dingjia2)

1. Datasets

- monet_jpg - 300 Monet paintings sized 256x256 in JPEG format
- monet_tfrec - 300 Monet paintings sized 256x256 in TFRecord format
- photo_jpg - 7028 photos sized 256x256 in JPEG format
- photo_tfrec - 7028 photos sized 256x256 in TFRecord format

We downloaded the datasets from Kaggle (https://www.kaggle.com/competitions/gan-getting-started/data) and used them to train our model.

2. Workflow

- Generate Dataloader to extract data from the datasets
- Construct the CycleGAN model
- Training for 150 epochs
- Print out predictions and tune the parameters accordingly

3. Current works

We implemented a CycleGAN model using residual blocks as generator from scratch. We ran the model under the Monet paintings and actual photos provided by Kaggle and get the following training curve.

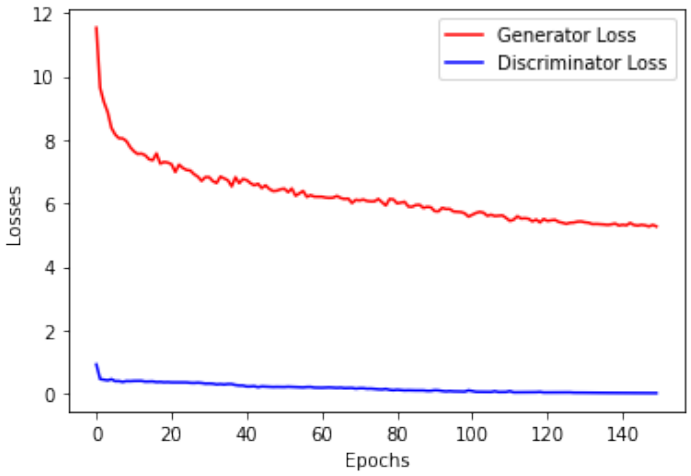

Figure 1 Generator Loss and Discriminator Loss for 150 training epochs

The generated Monet images are like below.

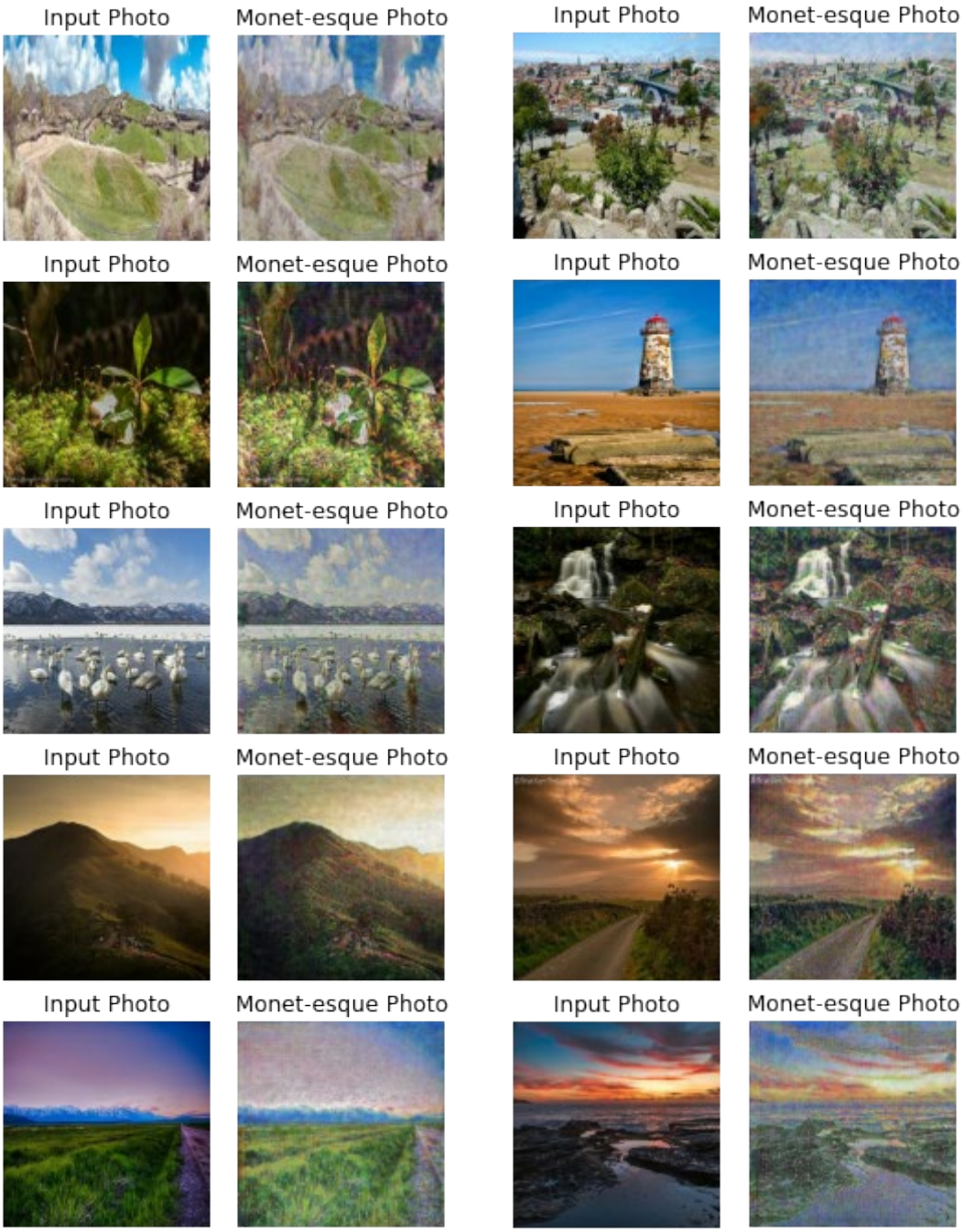

Figure 2 Sample Output for our CycleGAN

4. Future works

- Try Using different structures as the generator in the CycleGan
- Using FID or MiFID to evaluate the performance of our Gan model

- Try other GAN models on this task, like styleGAN
- Submit our final results to Kaggle and write the final report

5. Evaluation

FID (Fréchet Inception Distance)

$$FID = \|\mu_r + \mu_g\| + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2})$$

FID and IS (Inception Score) are widely used methods to evaluate GANs. FID is calculated by computing the Fréchet distance between two Gaussians fitted to feature representations of the Inception network.

MiFID (Memorization-informed FID)

MiFID is a modified version of FID, it also takes the training sample memorization into account.

$$d_{ij} = 1 - \cos(f_{gi}, f_{rj}) = 1 - \frac{f_{gi} \cdot f_{rj}}{|f_{gi}||f_{rj}|}$$

f_g represent the generated images in feature space (defined in pre-trained networks).

f_r represent the real images in feature space.

f_{gi} and f_{rj} represent the i^{th} and j^{th} vectors of f_g and f_r respectively.

$$d = \frac{1}{N}\sum_i \min_j d_{ij}$$

defines the minimum distance of a certain generated image (i) across all real images (j), then averaged across all the generated images.

$$d_{thr} \begin{cases} d, if\ d < \epsilon \\ 1, otherwise \end{cases}$$

defines the threshold of the weight only applies when the (d) is below a certain empirically determined threshold.

Finally, this memorization term is applied to the FID:

$$MiFID = FID \cdot \frac{1}{d_{thr}}$$