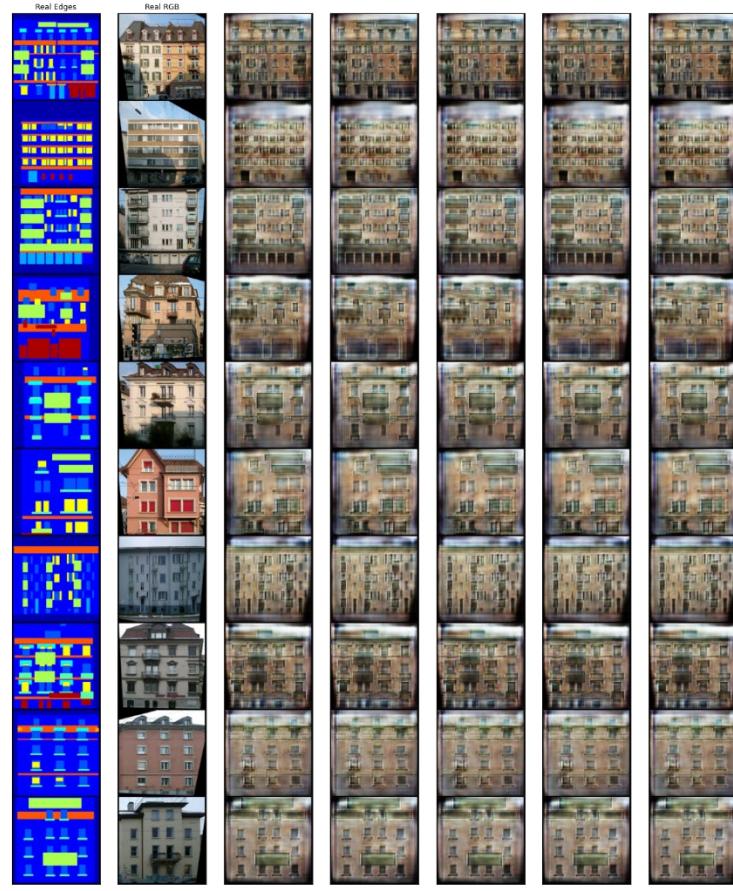


Evaluation

1. Qualitative evaluation



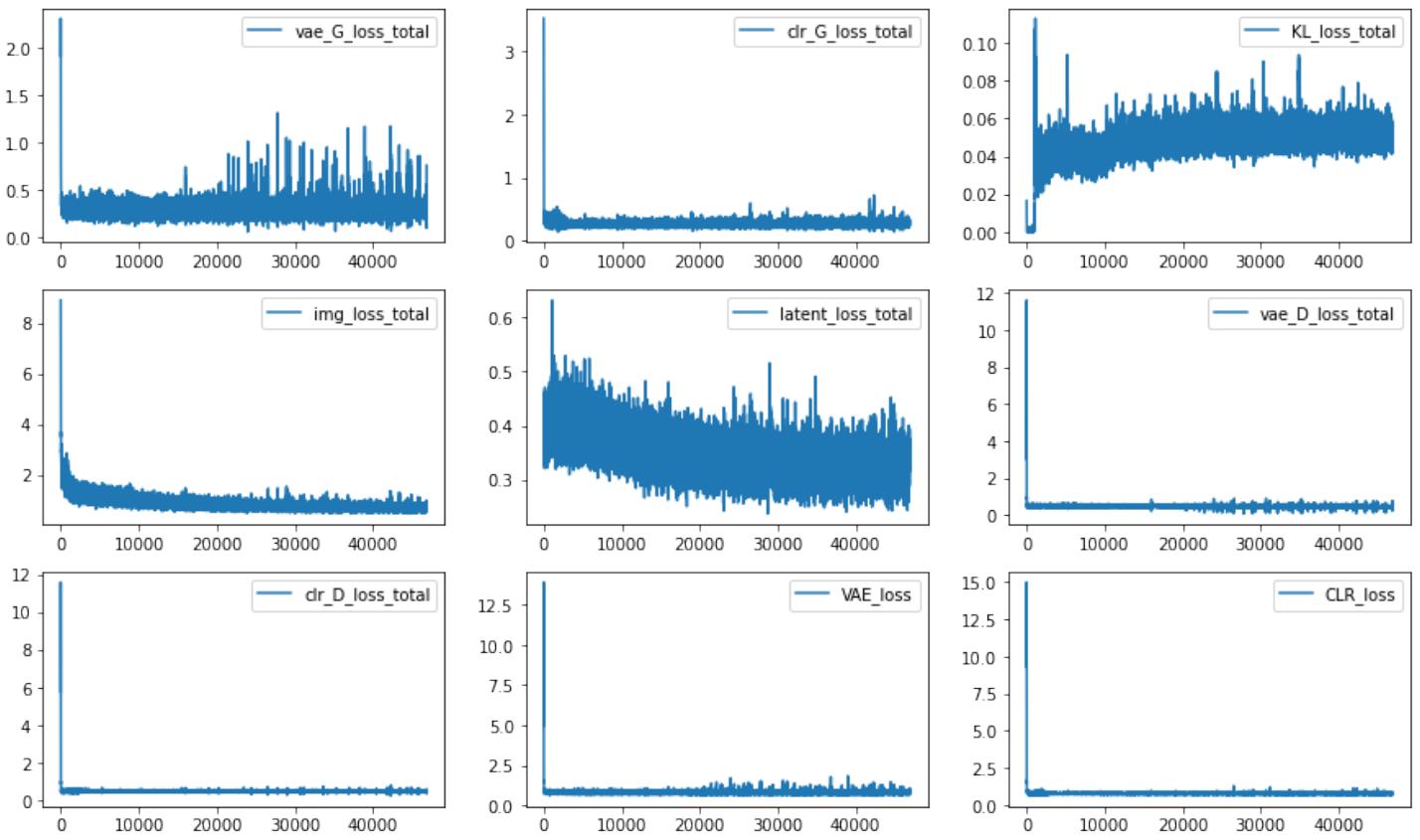


Inference result of ResNet based generator model in Map2Satellite dataset



Inference result of ResNet based generator model in Label2Facade dataset

Training loss curve



Training loss curve of major nine term in objective function

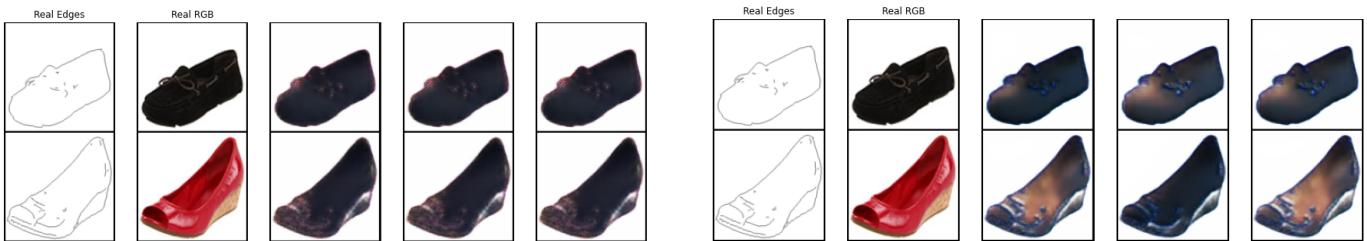
2. Quantitative Evaluation

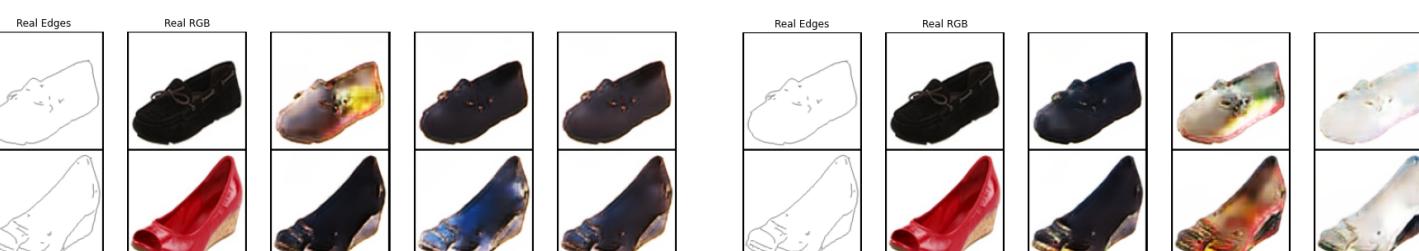
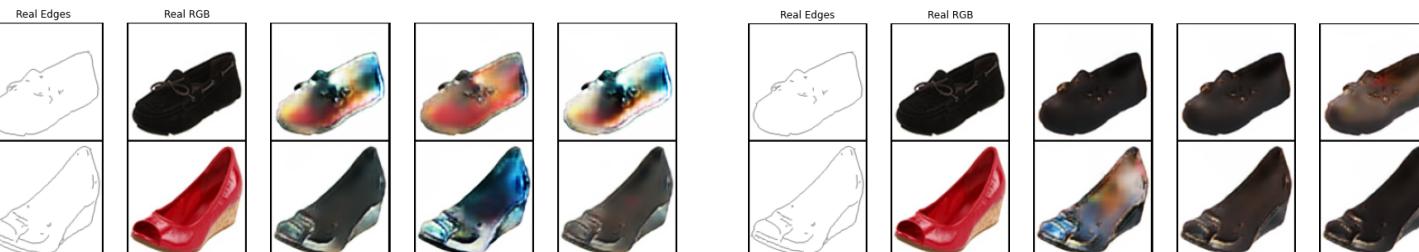
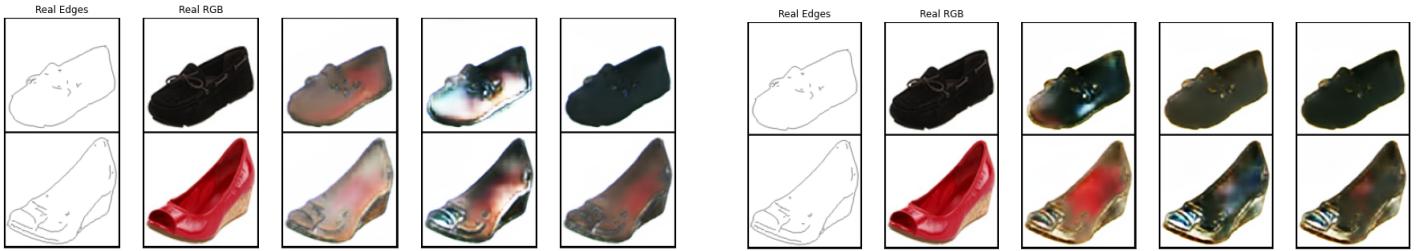
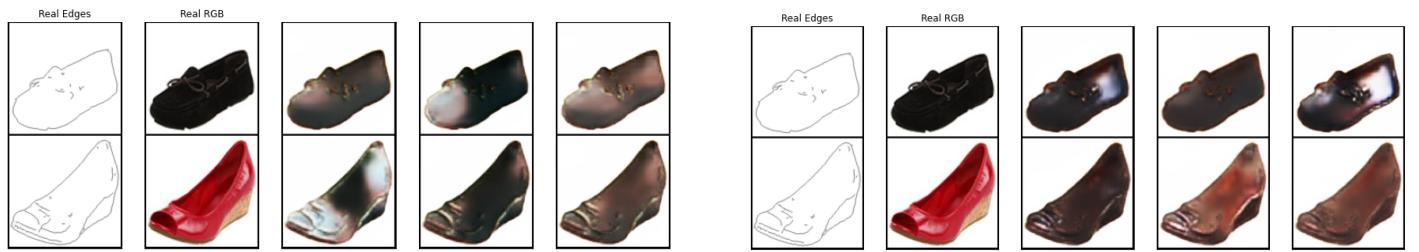
- FID Score

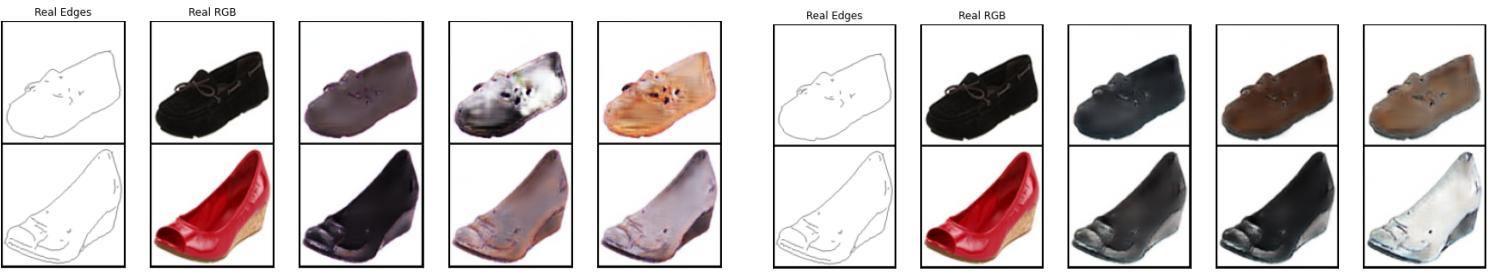
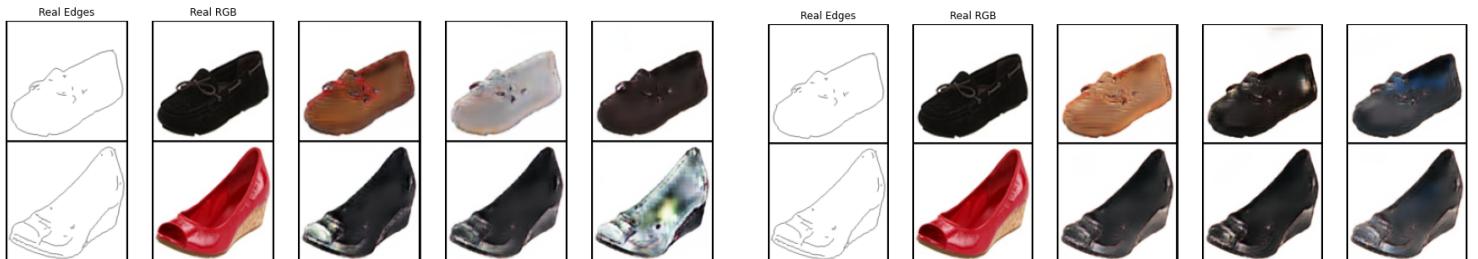
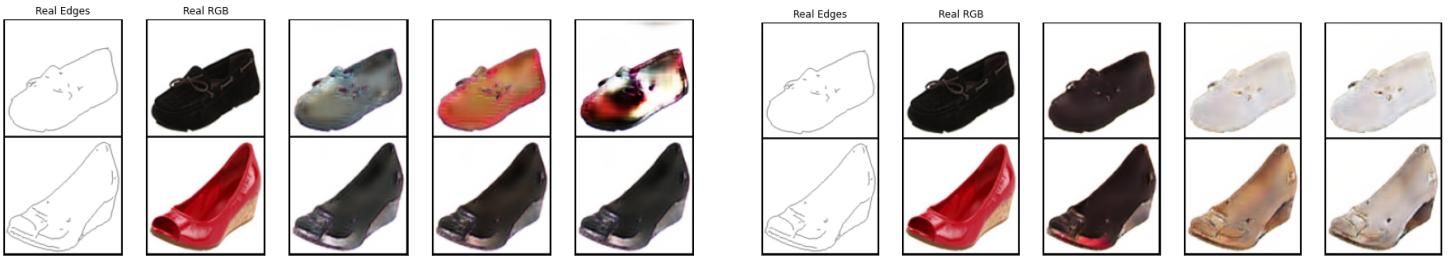
To evaluate the photo-realistic quality, we have chosen to evaluate the FID score between 200 generated images and real images in the validation set. It measures the distance between real image set and generated image set. By embedding both the real image and generated image into a feature space given by the output of final avgpool layer of Inception Net v3 model, we can extract the image distribution mean and variance statistics. Then we can calculate the Frechet distance between these two Gaussian distributions. Our model has reached a FID score of 88.00.

- PIPS metric

To evaluate the generated image diversity from our BicycleGAN model, we have adopted the Learned Perceptual Image Patch Similarity (LPIPS) metric which computes the distance between generated image patches. According to our ResNet based generator model it can achieve LPIPS score of 0.158.







Training visualization over 20 epochs