CIS 6800 Final Project Evaluation Results

Yukai Yang, Zhuolun Zhao

December 20, 2022

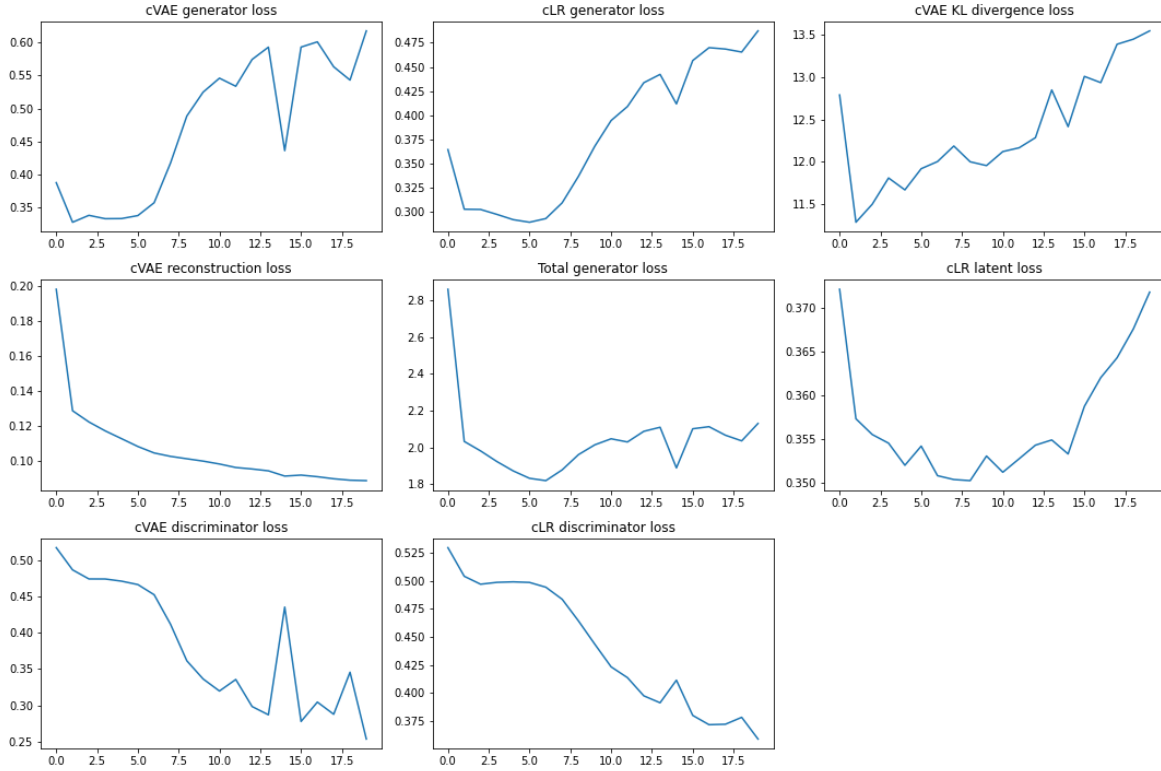## 1 Training Curves and Visualization

Figure 1: Training loss curves of BicycleGAN

We trained our model for 20 epochs. The total generator loss is the aggregate sum of cVAE generator loss, cLR generator loss, KL divergence loss, and reconstruction loss.

(a) Epoch 1   (b) Epoch 2   (c) Epoch 3   (d) Epoch 4   (e) Epoch 5

(f) Epoch 6   (g) Epoch 7   (h) Epoch 8   (i) Epoch 9   (j) Epoch 10

(k) Epoch 11   (l) Epoch 12   (m) Epoch 13   (n) Epoch 14   (o) Epoch 15

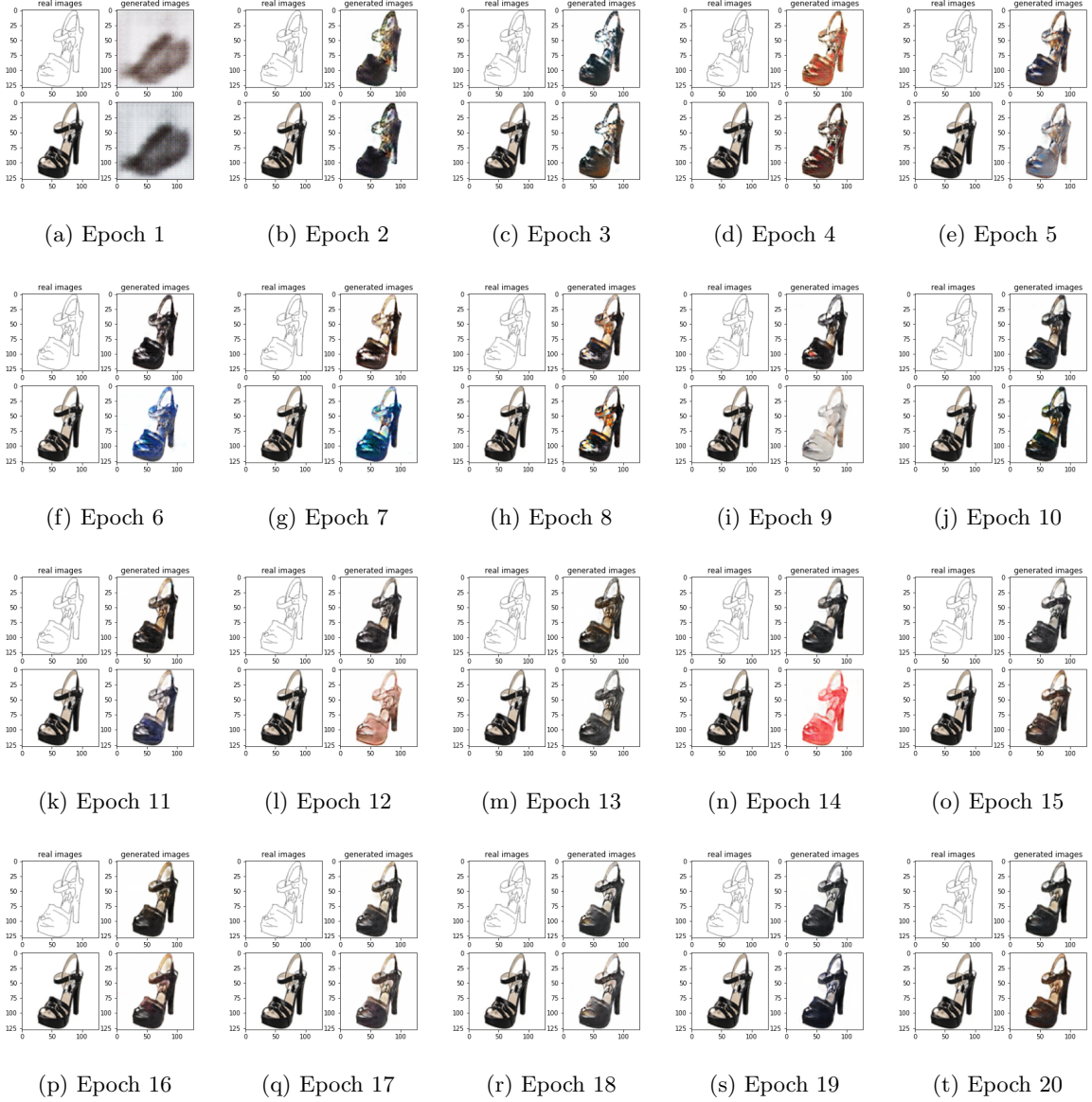(p) Epoch 16   (q) Epoch 17   (r) Epoch 18   (s) Epoch 19   (t) Epoch 20

Figure 2: Visualization of real and generated shoe images during training. Images in top left, bottom left, top right, bottom right order are real A, real B, encoded B, and random B.

The visual quality of the generated images improve over the course of training. In particular, the encoded B output by cVAE-GAN becomes more similar to real B while random B remain diverse. This is expected due to the reconstruction loss and the fact that cVAE-GAN has prior information about B from its sampled latent encoding.
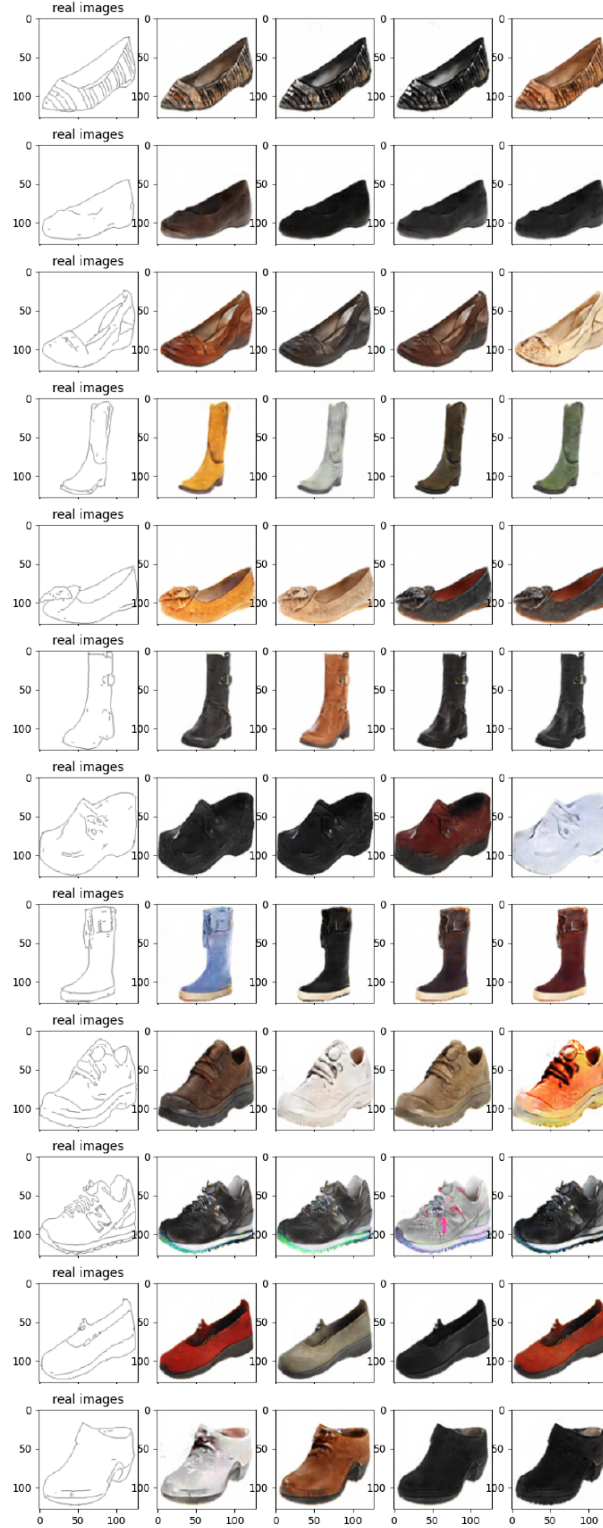
2 Qualitative Evaluation

Figure 3: Inference results of multiple random latent vectors

We selected different image A's from our test set and draw random latent vectors to generate image B's. Please refer to section 5.1 in the main report for discussions about the results.

3 Quantitative Evaluation

Our metrics are computed on the entire test set. Across our experiments, the best FID score we obtained is 78.84 and the best LPIPS score is 0.216. Please refer to section 5.2 in the main report for discussions about the results.

Model	FID ↓	LPIPS ↑
baseline	**78.84**	0.110
sep-disc	79.44	0.142
sep-multi-disc	95.76	0.133
sep-disc + U-Net-gen	95.34	**0.216**
sep-disc + U-Net-In-gen	84.36	0.123

Table 1: Comparison of generation realism and diversity across different model choices.