CSE 803: Project

Image-to-Image Translation (Sketch to Image)

Asadullah Hill Galib

Motivation

Image-to-Image Translation with Conditional Adversarial Networks

Phillip Isola Jun-Yan Zhu Tinghui Zhou Alexei A. Efros

Berkeley AI Research (BAIR) Laboratory, UC Berkeley

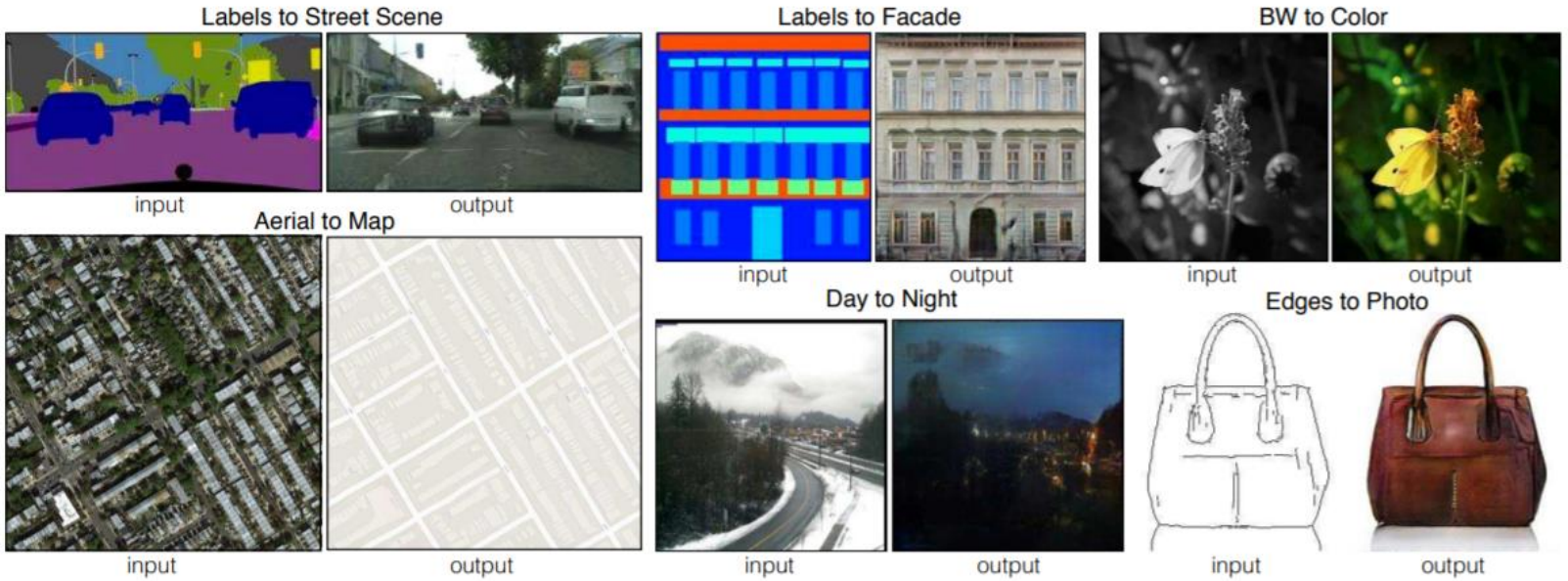
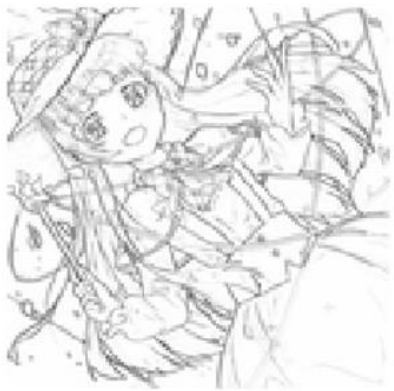


Figure 1: Many problems in image processing, graphics, and vision involve translating an input image into a corresponding output image. These problems are often treated with application-specific algorithms, even though the setting is always the same: map pixels to pixels. Conditional adversarial nets are a general-purpose solution that appears to work well on a wide variety of these problems. Here we show results of the method on several. In each case we use the same architecture and objective, and simply train on different data.

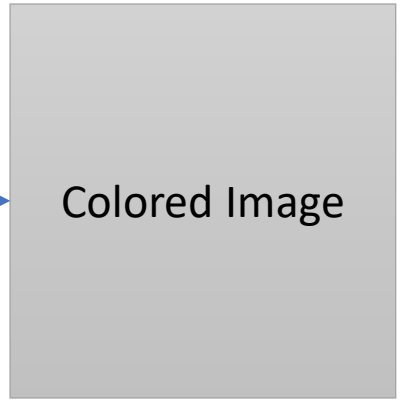
Source: Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134).

Problem Definition

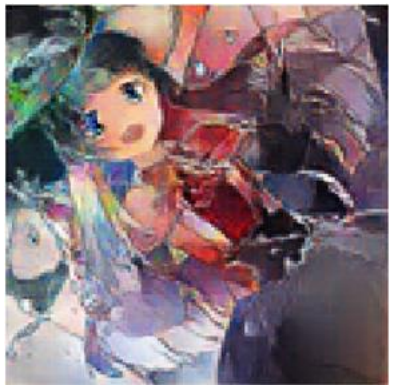
- Sketch to Image Generation
  - Given sketch-image pairs, train model to generate colored images
  - Using U-net Generator and conditional GAN



Sketch

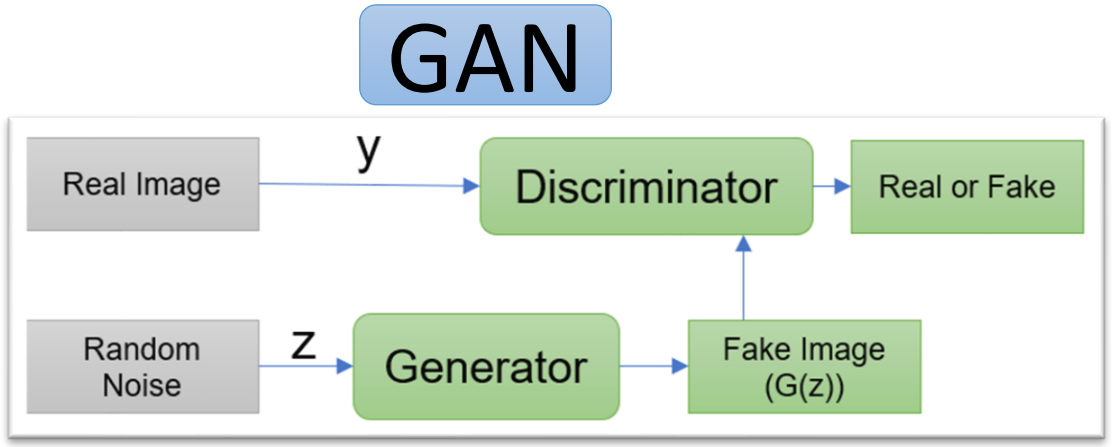


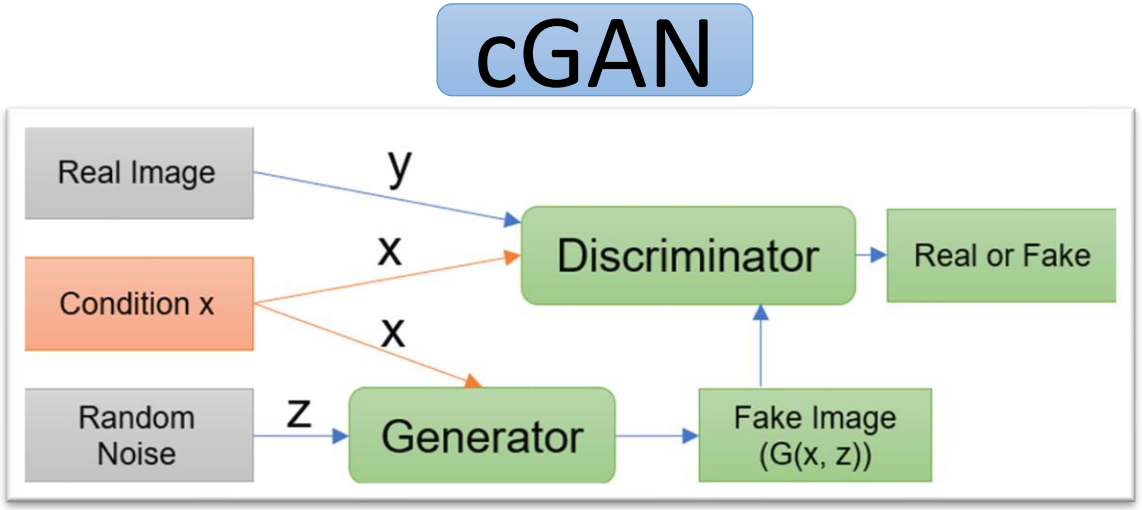
Generated Image



Ground Truth

Architecture (GAN vs cGAN)

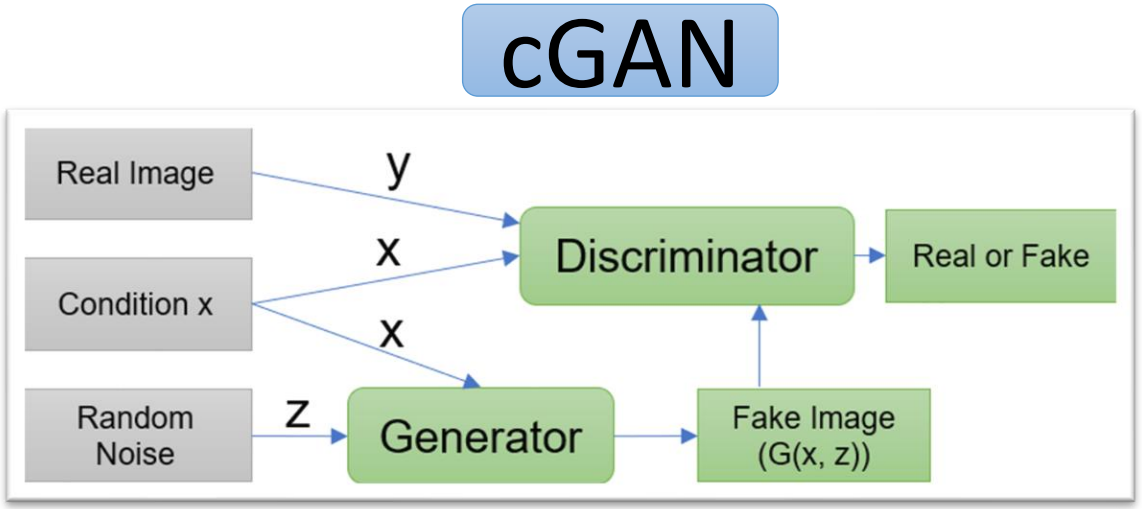




$$\mathcal{L}_{GAN}(G, D) = \mathbb{E}_y[\log D(y)] + \mathbb{E}_z[\log(1 - D(G(z)))]$$

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))]$$

Architecture (cGAN)

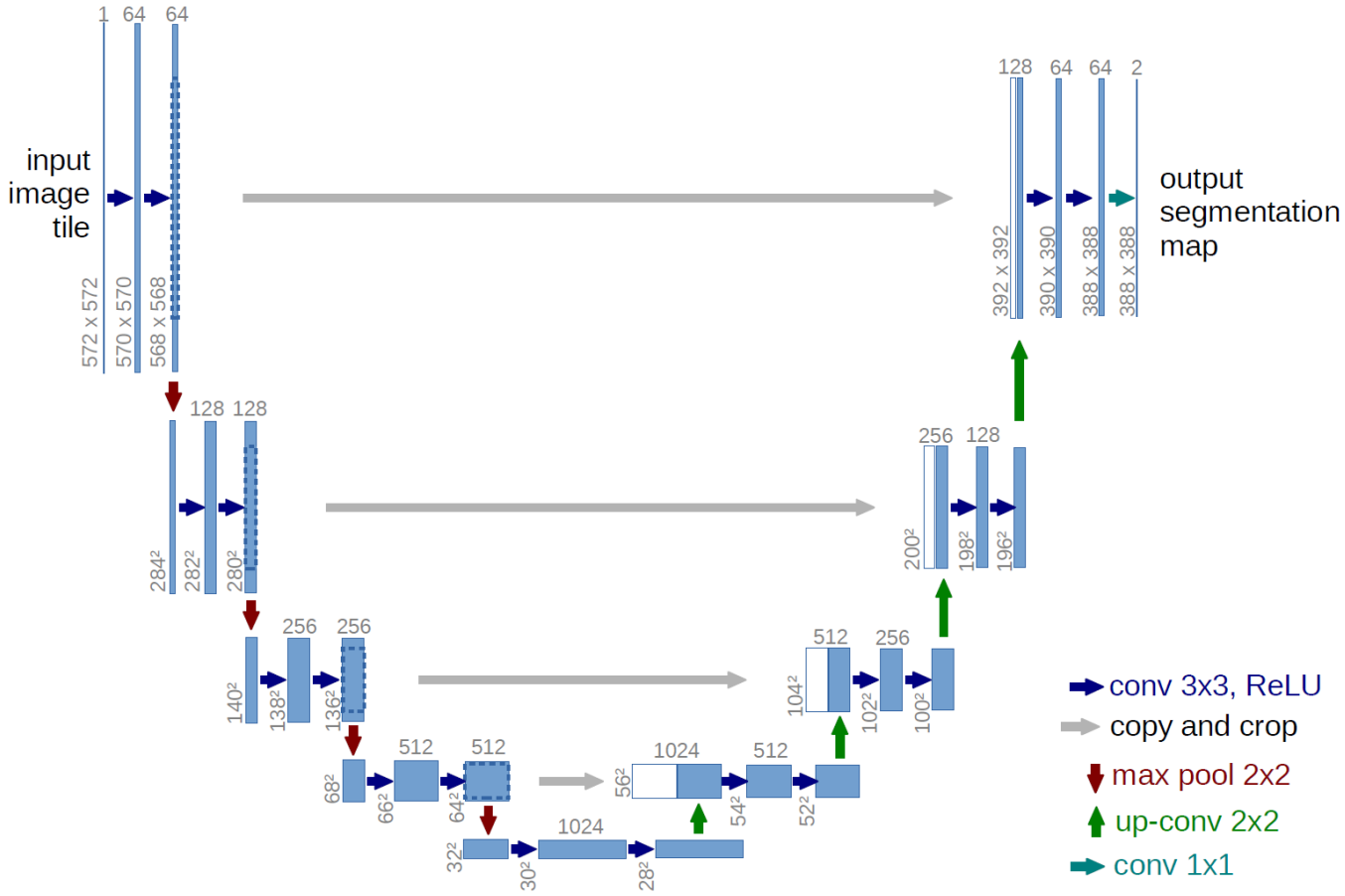


$$\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{x,y}[\log D(x,y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x,z)))]$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[||y - G(x,z)||_1]$$

$$\mathcal{L}_{Total}(G,D) = \mathcal{L}_{cGAN}(G,D) + \lambda\mathcal{L}_{L1}(G)$$

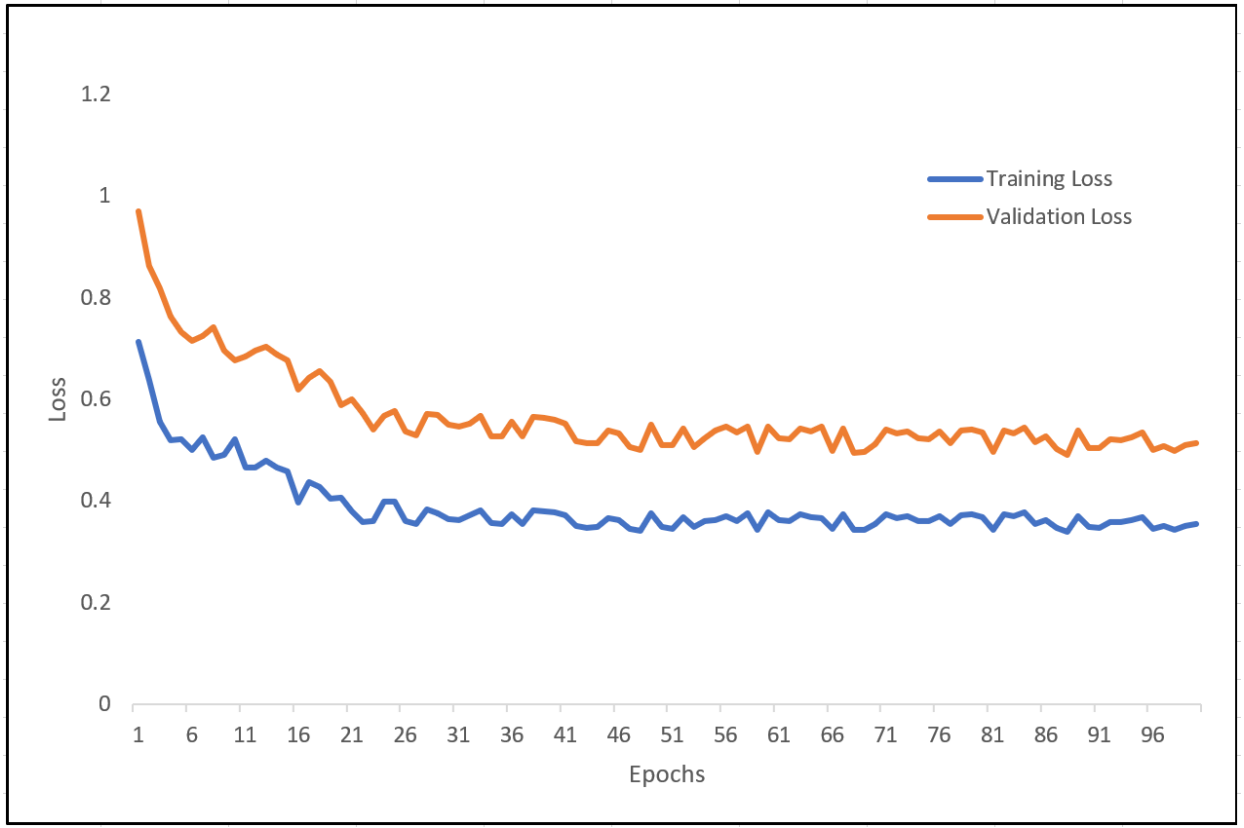
Architecture (U-net-based Generator)



Source: Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.

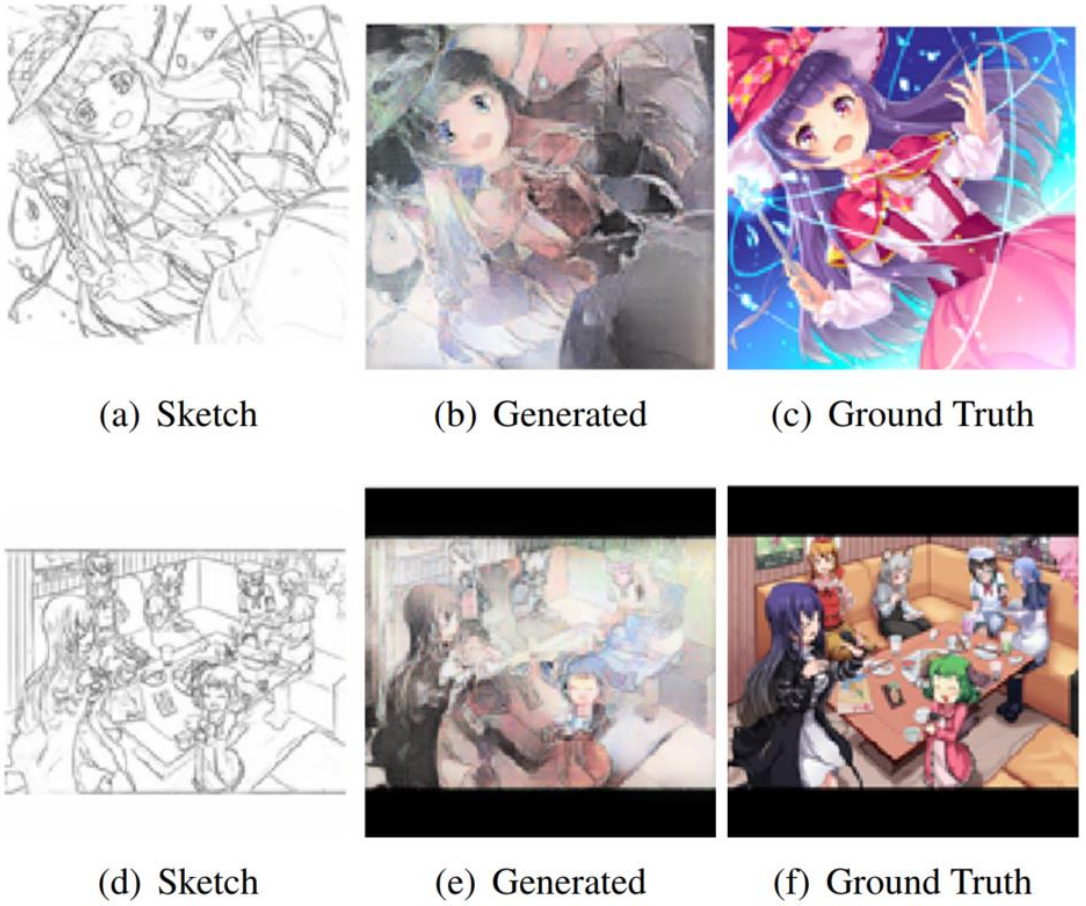
Results (U-net-based Generator)

Learning Curve:



Results (U-net-based Generator)

Sketch to Image using U-net-based Generator:

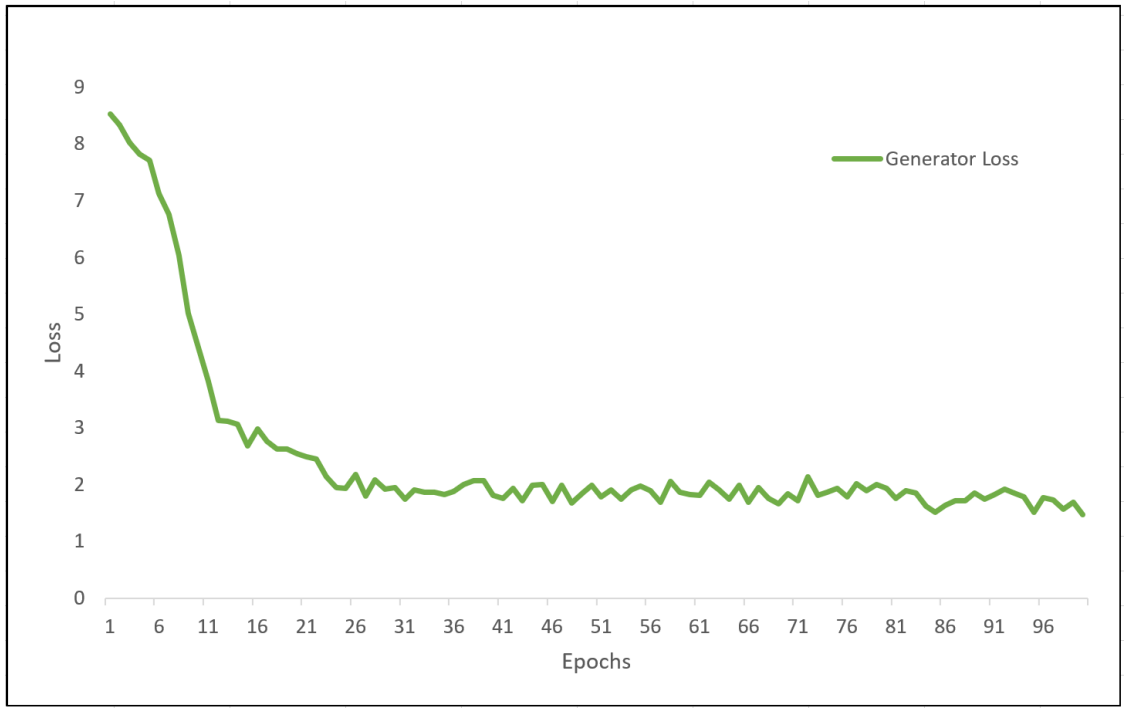


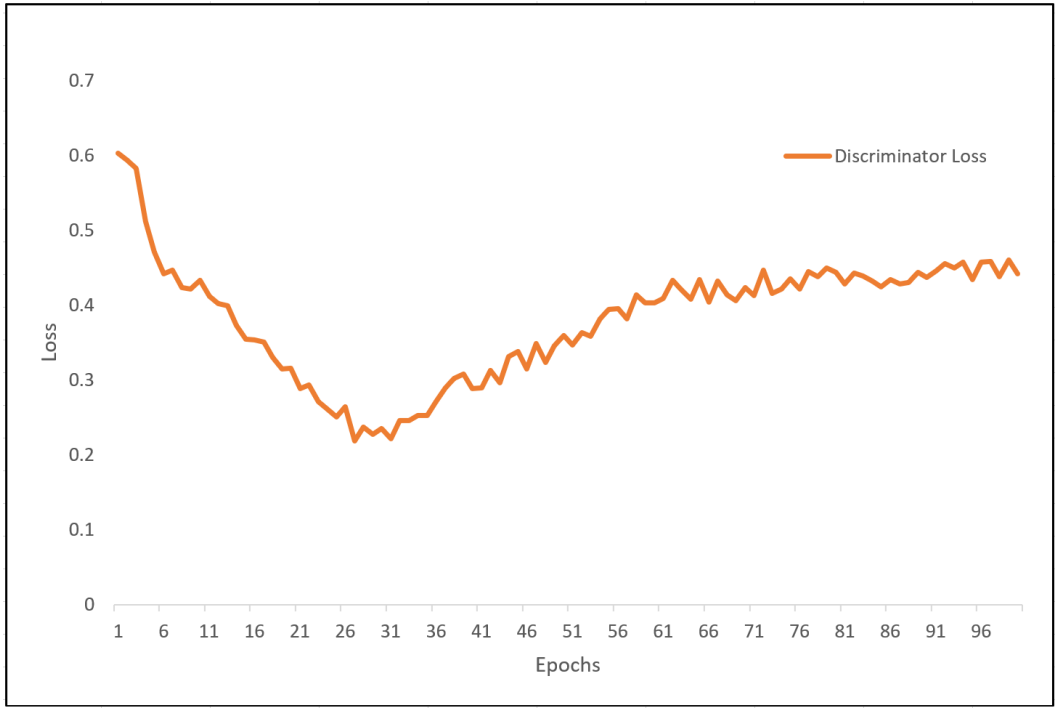
(a) Sketch (b) Generated (c) Ground Truth

(d) Sketch (e) Generated (f) Ground Truth

Results (cGAN)

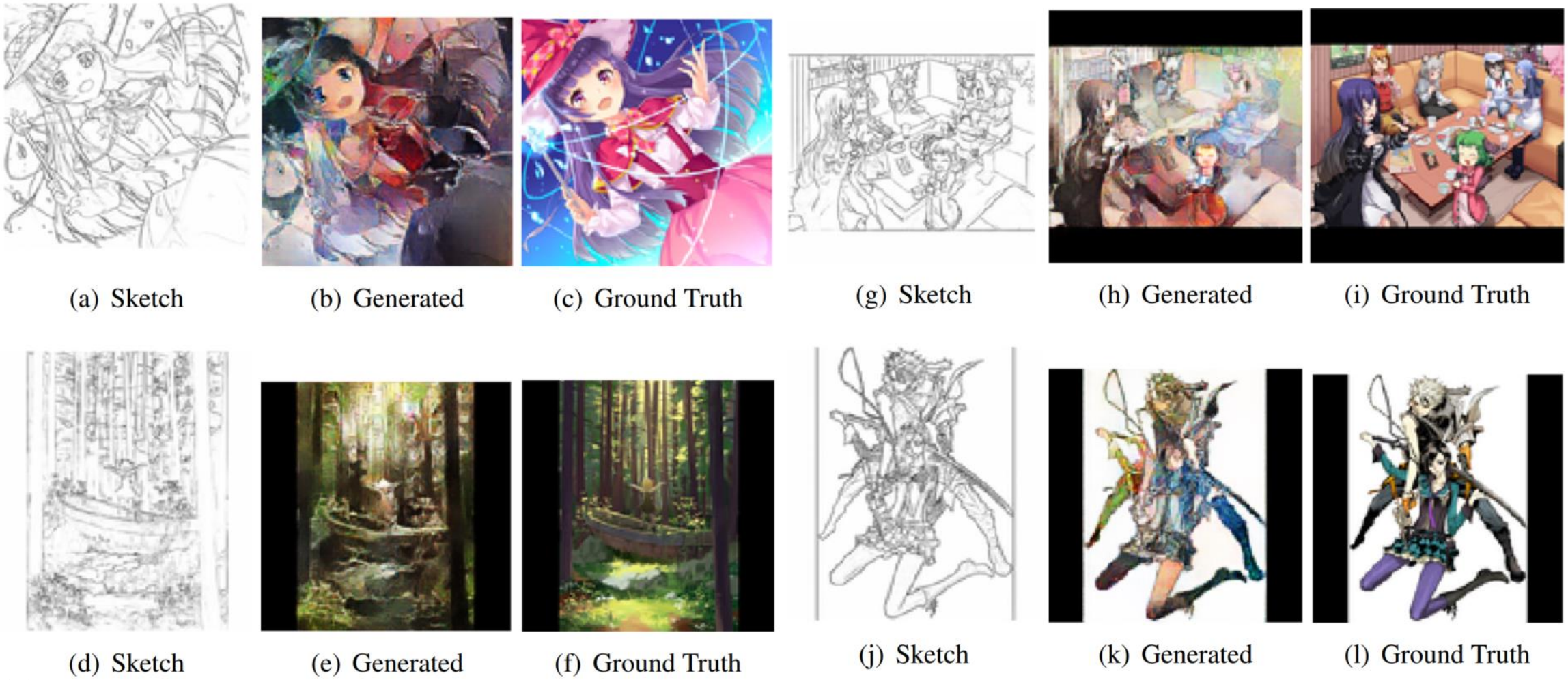
Learning Curve of Generator and Discriminator:





Results (cGAN)

Sketch to Image using cGAN architecture:



(a) Sketch (b) Generated (c) Ground Truth

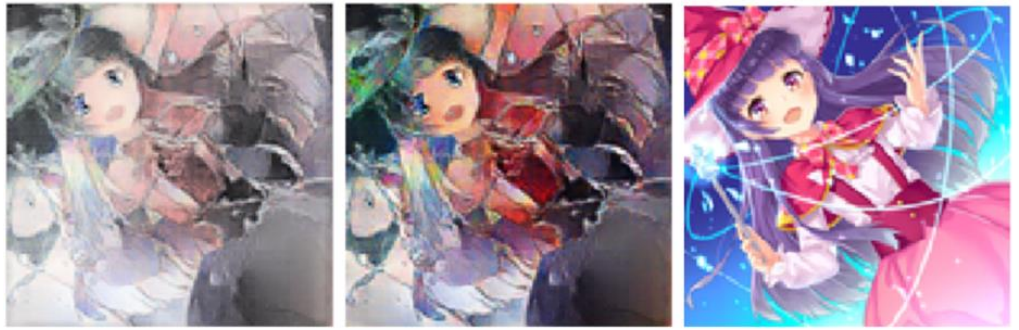
(g) Sketch (h) Generated (i) Ground Truth

(d) Sketch (e) Generated (f) Ground Truth

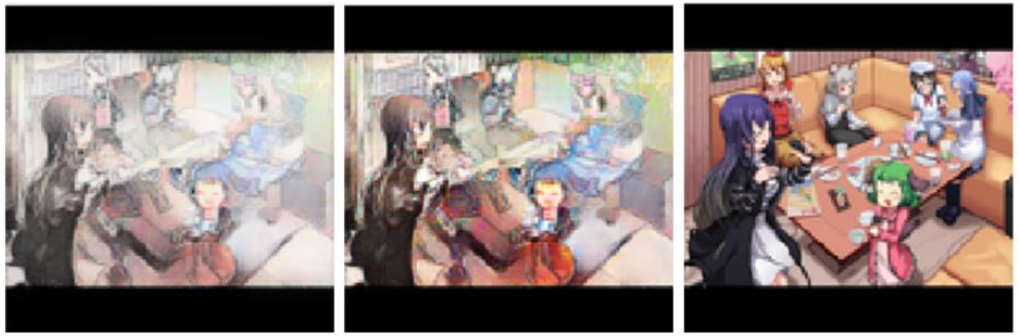
(j) Sketch (k) Generated (l) Ground Truth

Results (U-net-based Generator vs cGAN)

Comparison between U-net Generator (without discriminator) and cGAN architecture (with discriminator):



(a) U-net (b) cGAN (c) Ground Truth



(d) U-net (e) cGAN (f) Ground Truth

(g) U-net (h) cGAN (i) Ground Truth

Thank you