

Samsung Innovation Campus

Doodle Art: Sketch to Artwork Converter

Sandip Katel

Sharad Pokharel

Yujal Shrestha

Problem Overview

Challenges in Sketch Translation

- Lack of Details
- Variation in Drawing Quality
- Ambiguous shapes

Why Traditional Methods Fail

- Manual effort is time consuming and requires artistic skills
- Rule-based techniques break down on complex scenes
- CNNs alone produce blurry results without adversarial pressure

Why cGANs Work Better

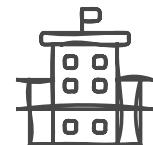
- Adversarial training produces realism matching natural images
- Conditional input (cGAN) enforces structure aligned to the sketch
- U-Net skip connections preserve spatial edge information

Objectives

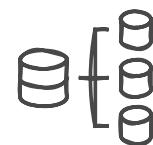
Core Development Goals

Main Objective:

Develop a system that generates high-quality, realistic colored images from sketches



Implement Conditional GAN Architecture



Train Model Using Paired Datasets

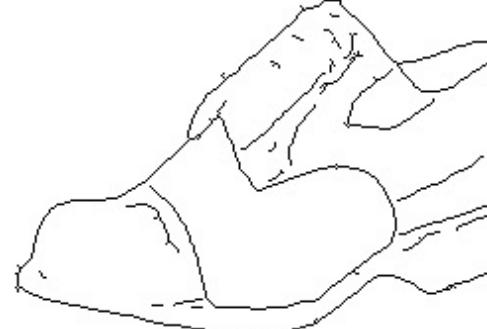


Evaluate System Performance Quantitatively

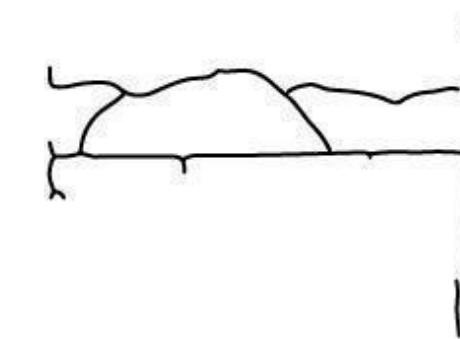
Methodology

1. Datasets used

- Object dataset: edges2shoes/edges2bags paired datasets



- Scene dataset: Mountains, hills, lakes



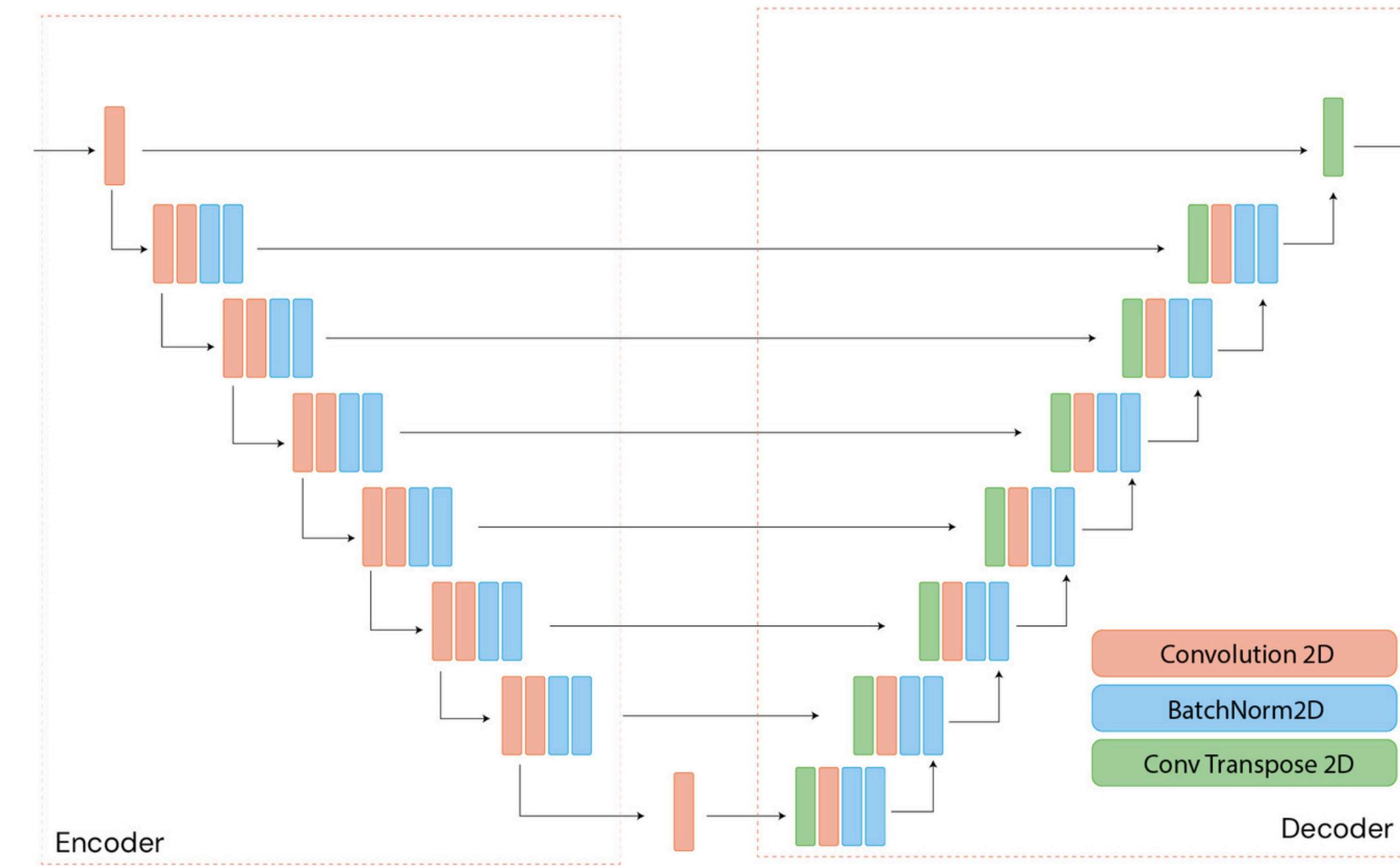
2. Preprocessing Steps

- **Image Splitting:** Each combined image is divided into input (sketch) and target halves.
- **Pairwise Augmentation:** Apply identical transformations (flip, jitter, etc.) to both images to maintain alignment.
- **Normalization:** Pixel values normalized to the range [-1, 1] for stable GAN training.
- **Tensor Conversion:** Images are converted to PyTorch tensors for model compatibility.

Model Architecture

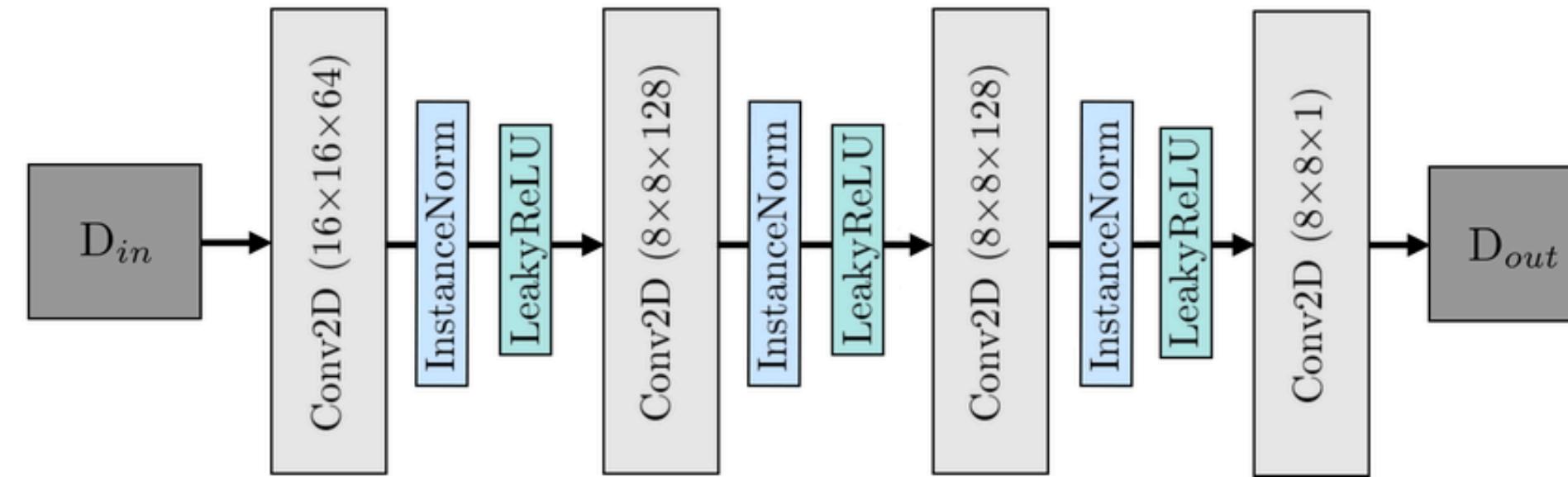
1. Generator

- U-Net consisting of an encoder and a decoder with skip connections between corresponding layers



2. Discriminator

- PatchGAN to evaluate the realism of local image patches rather than the entire image

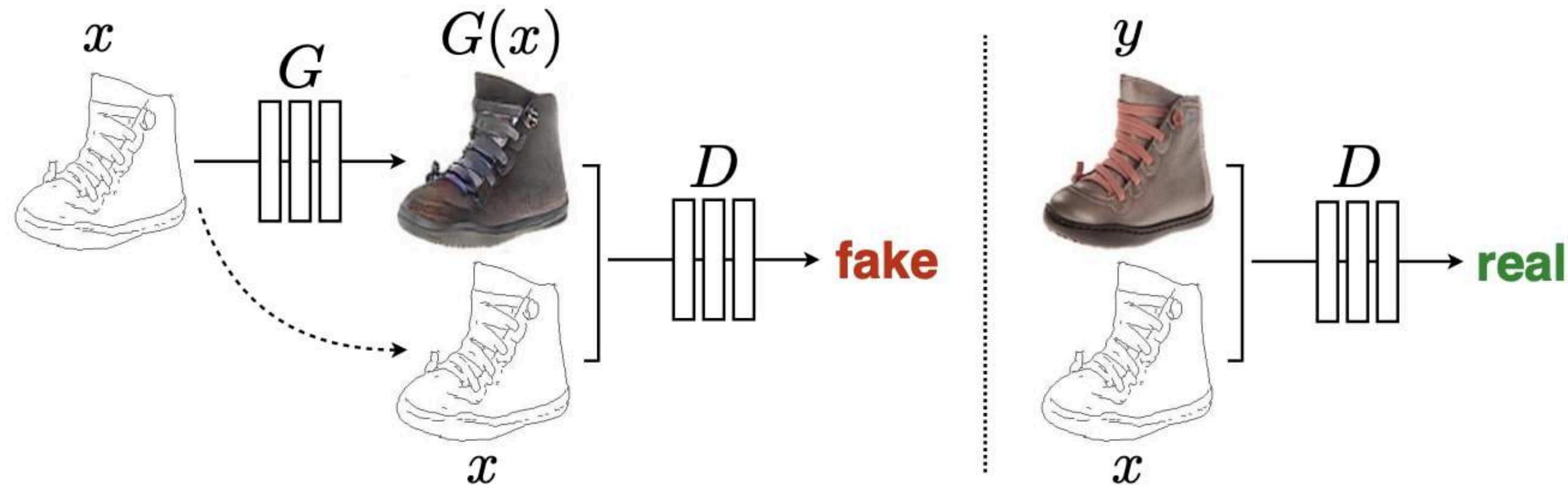


3. Loss Functions

- **Adversarial Loss:** Encourages the generator to produce outputs indistinguishable from real images.
- **L1 Loss (Reconstruction Loss):** To make the generated image match the ground truth image globally.
- **Perceptual Loss:** Evaluate high-level semantic similarity between generated and target images

Model Architecture

Core system



Pix2Pix cGAN Architecture

Training Setup

Training Hyperparameters

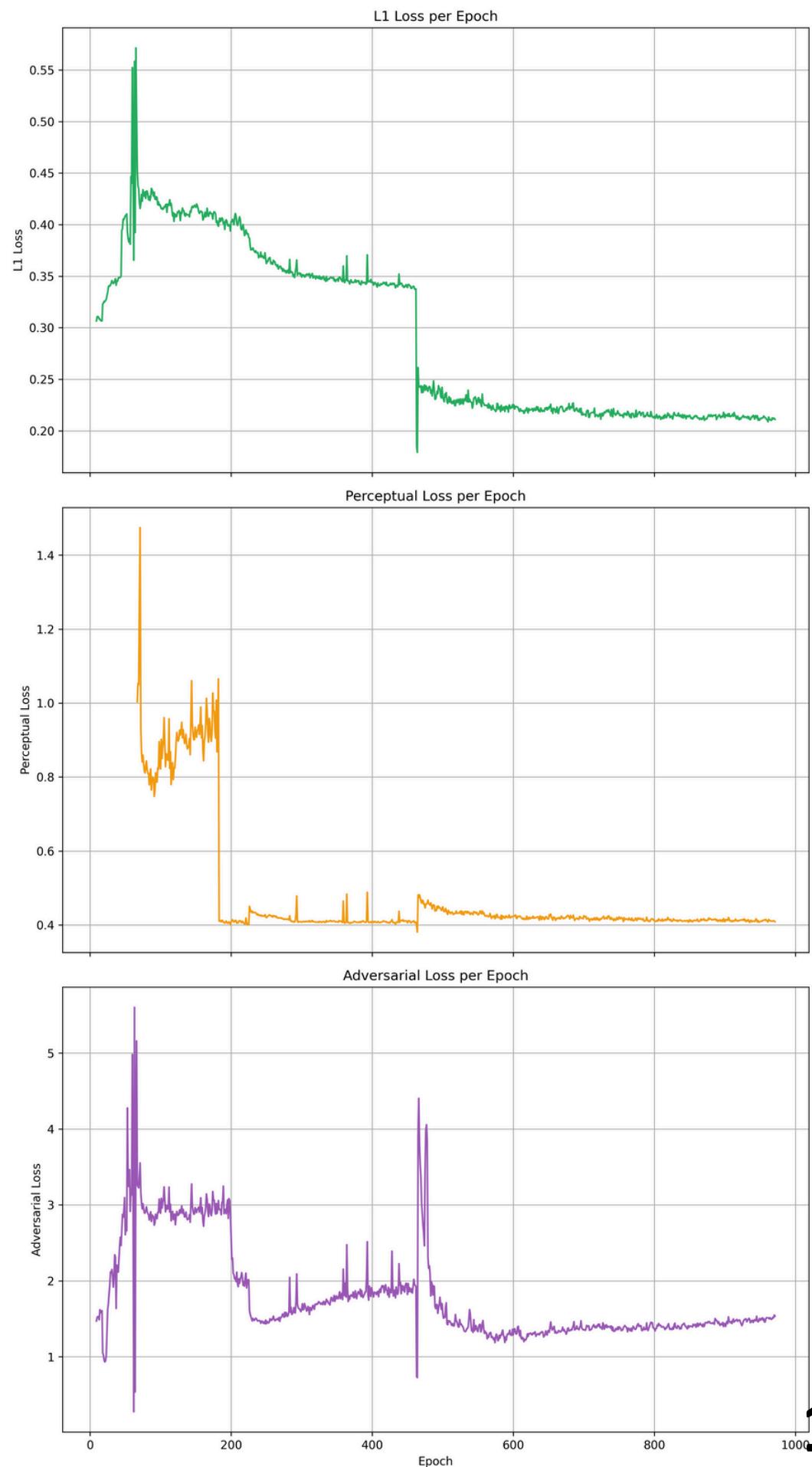
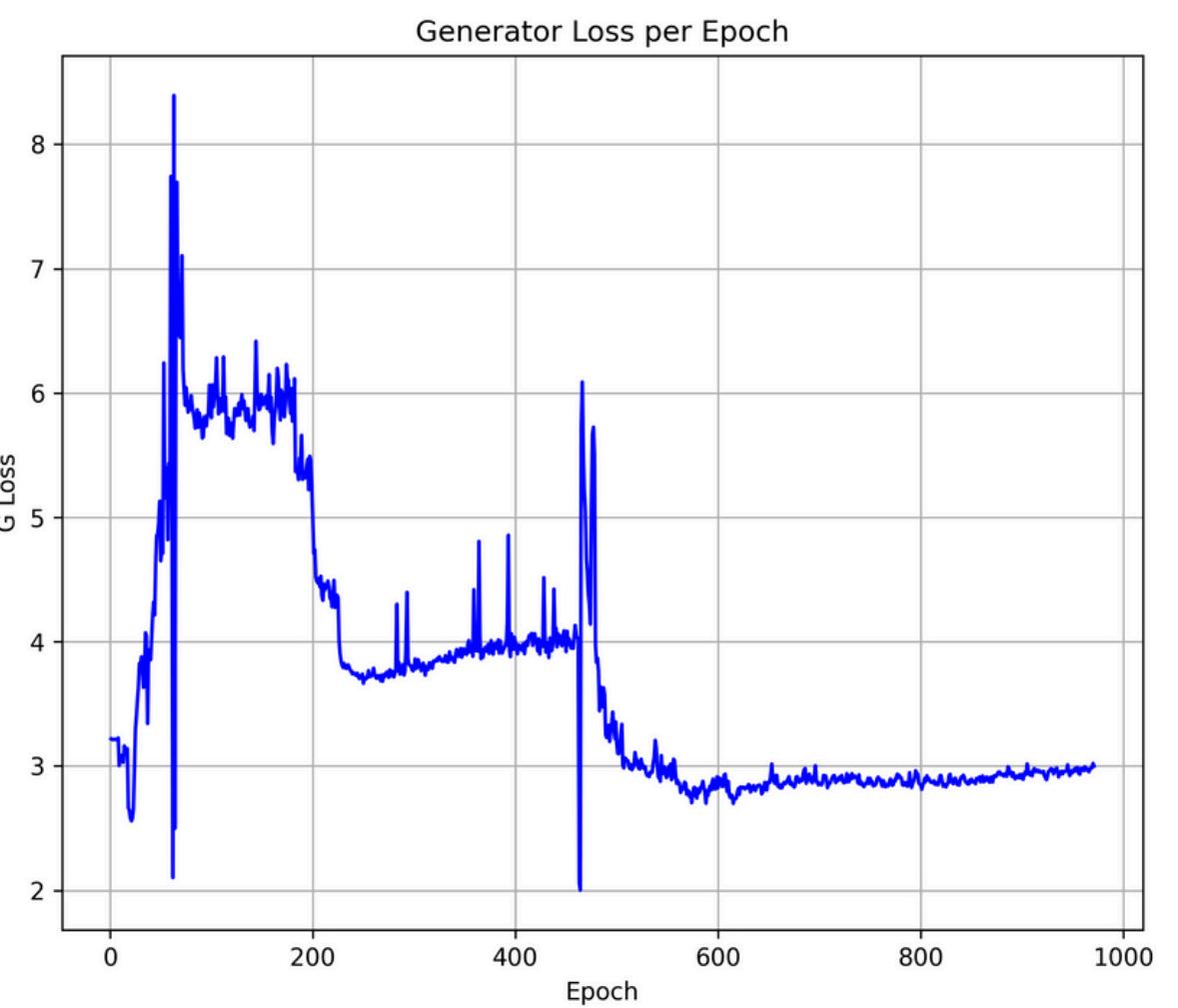
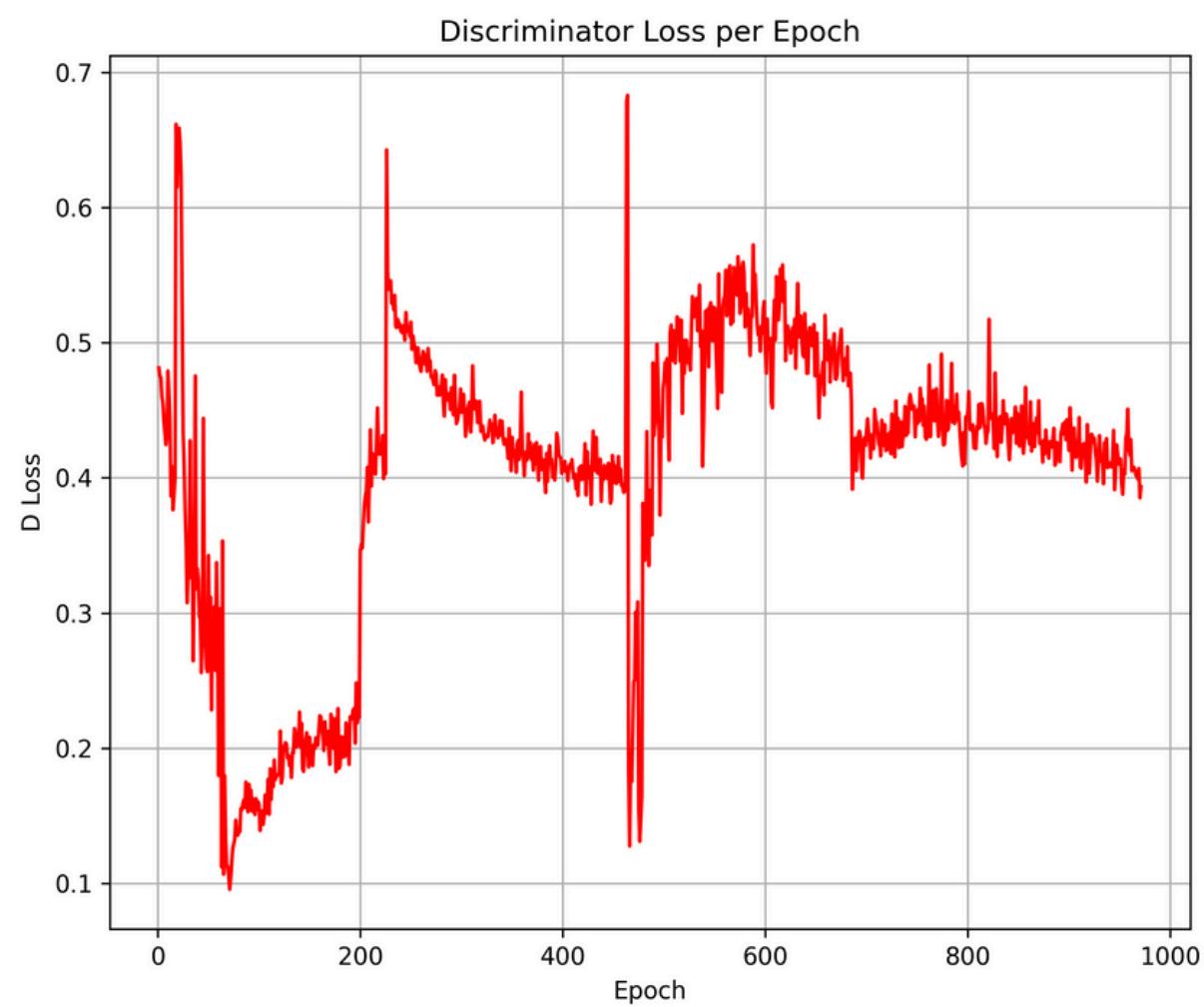
- Batch size: 16–32
- Learning rates:
 - Gen: 2×10^{-4} ($\beta_1=0.5$)
 - Disc: 1×10^{-4} ($\beta_1=0.9$)
- Epochs:
 - Object: 500 Pix2Pix + 500 DeepBlocks
 - Scene: 420 DeepBlocks
- Optimizer: Adam
- Image size: 256×256

Loss Functions

- Total Generator Loss = Adv + $\lambda_1 * L1 + \lambda_2 * \text{Perceptual}$
- Discriminator Loss = BCE

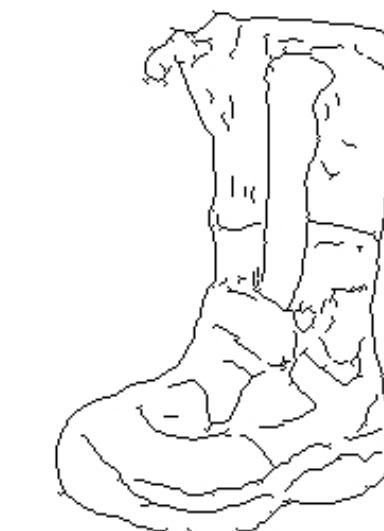
Results

Object Model: Training Loss Curves



Results

Object Model: Qualitative Results



Sketch

Generated

Original

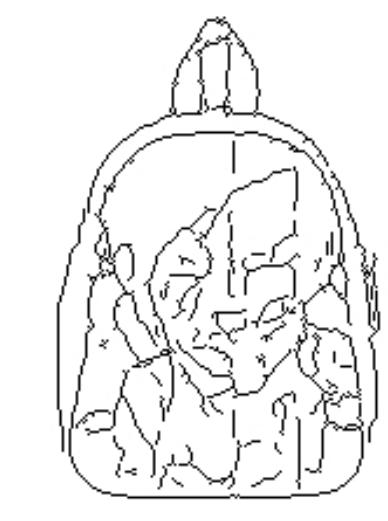
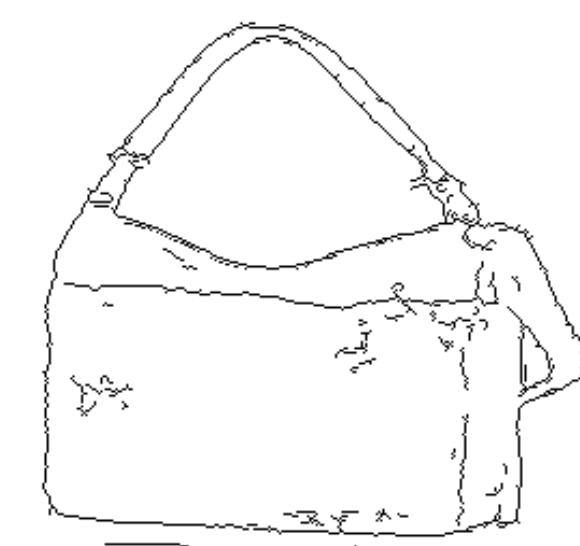
Sketch

Generated

Original

Results

Object Model: Qualitative Results



Sketch

Generated

Original

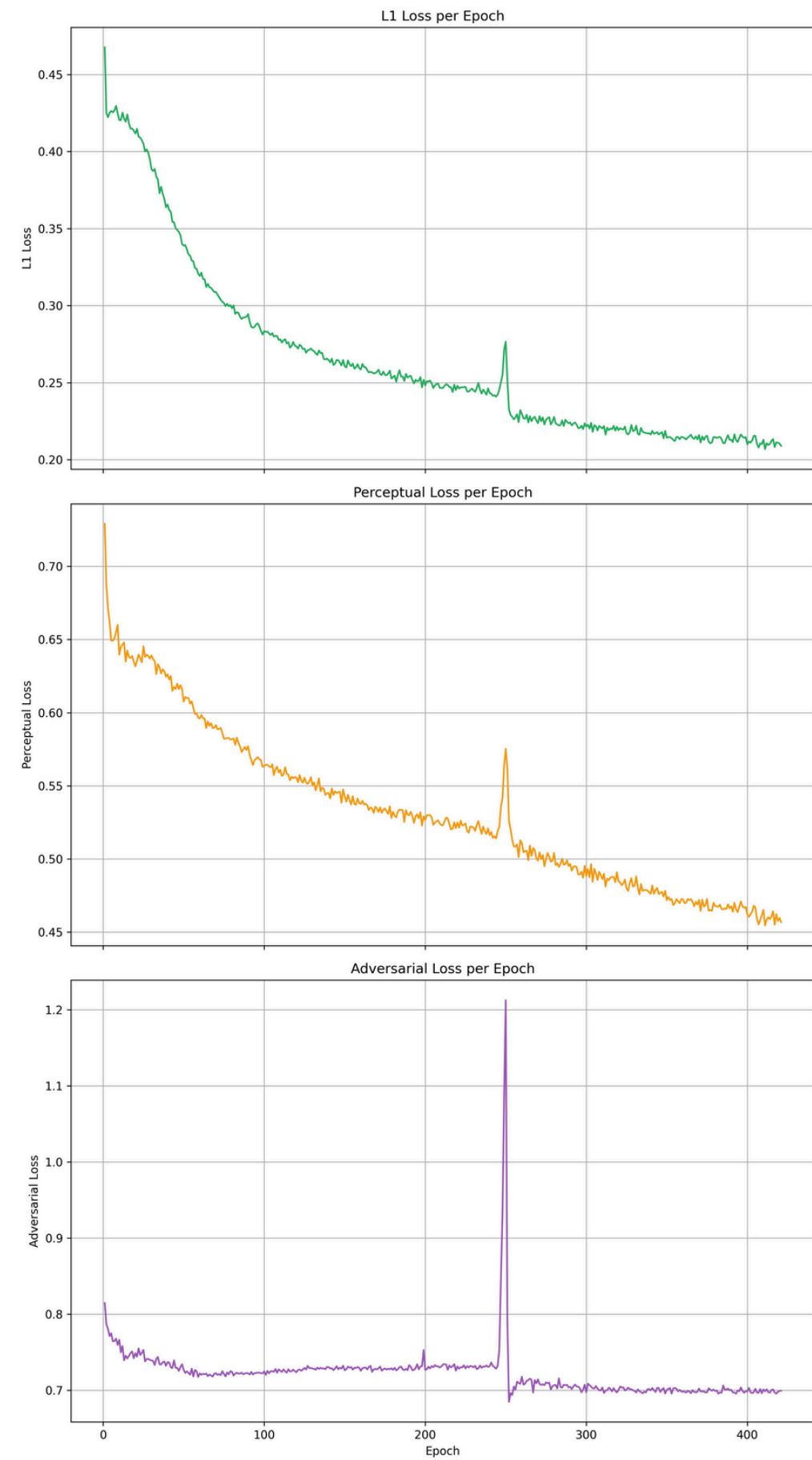
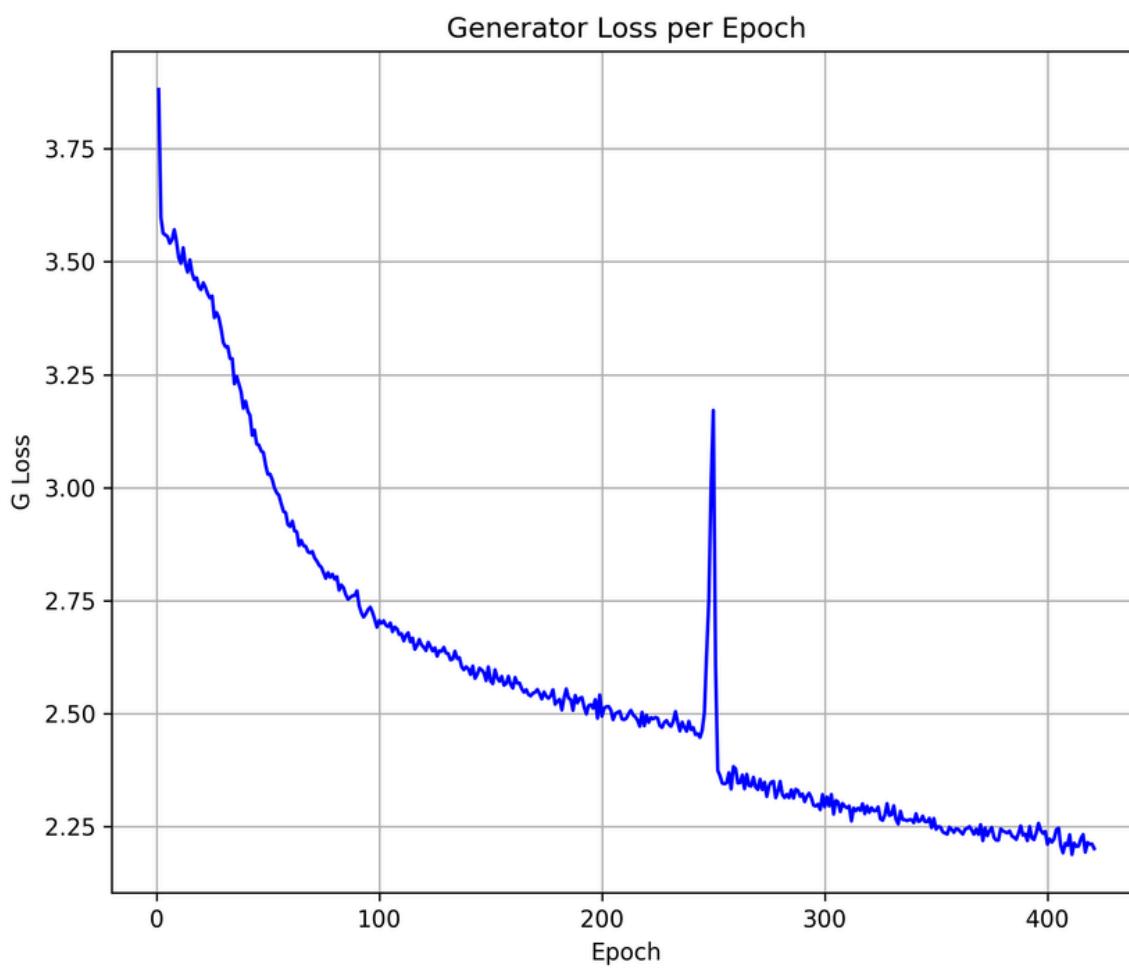
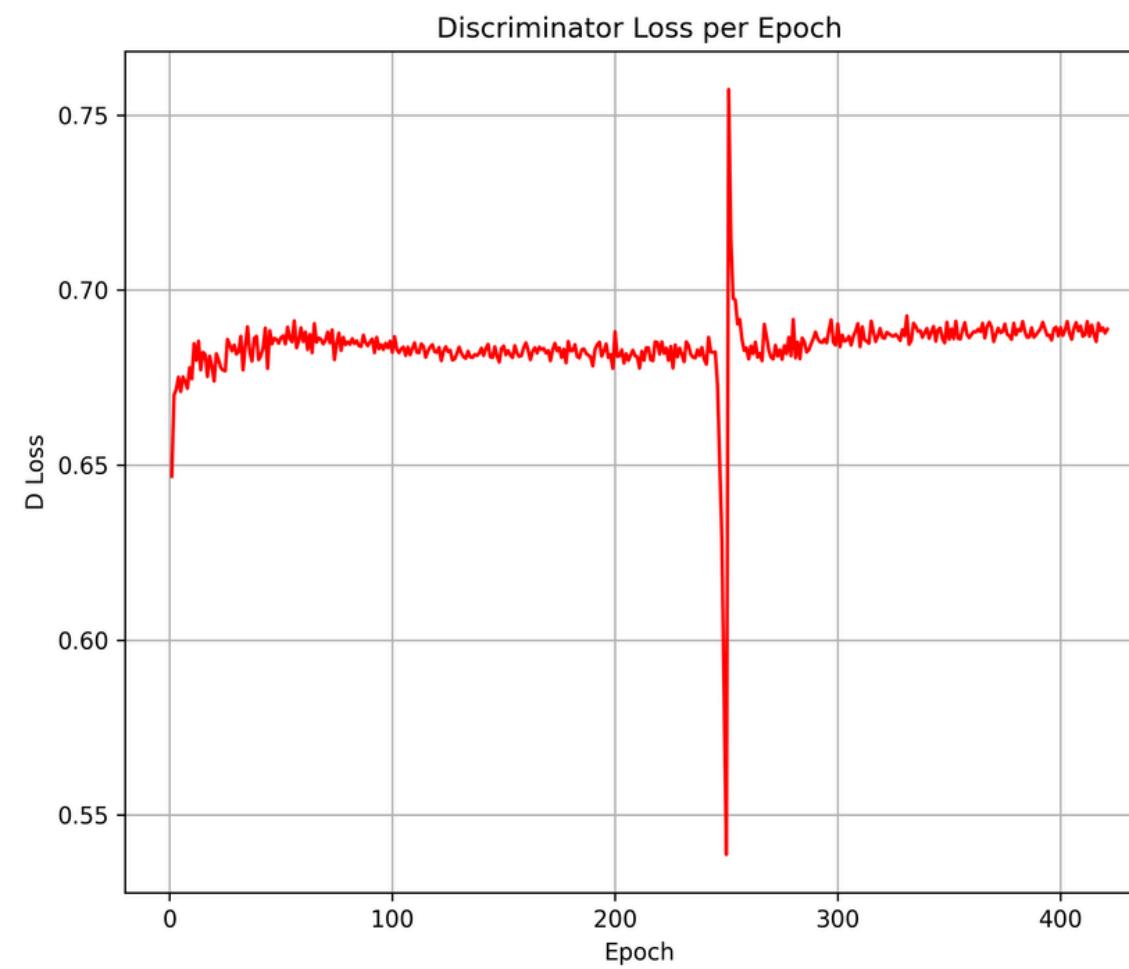
Sketch

Generated

Original

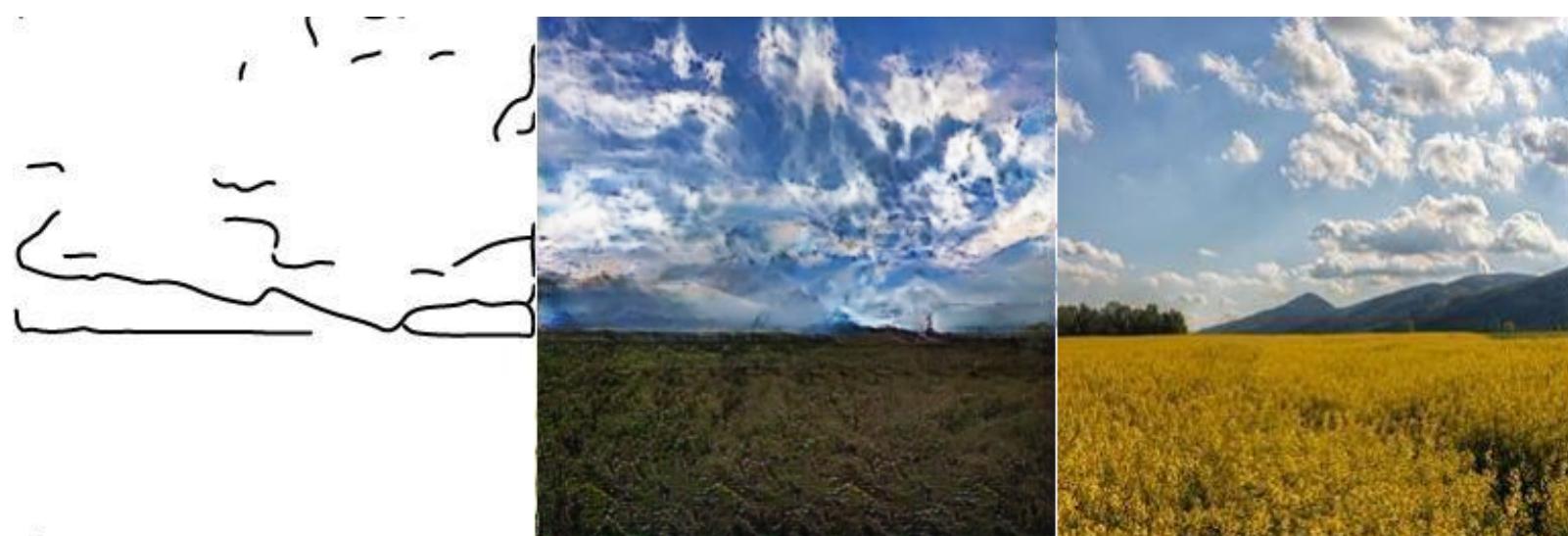
Results

Scene Model: Training Loss Curves



Results

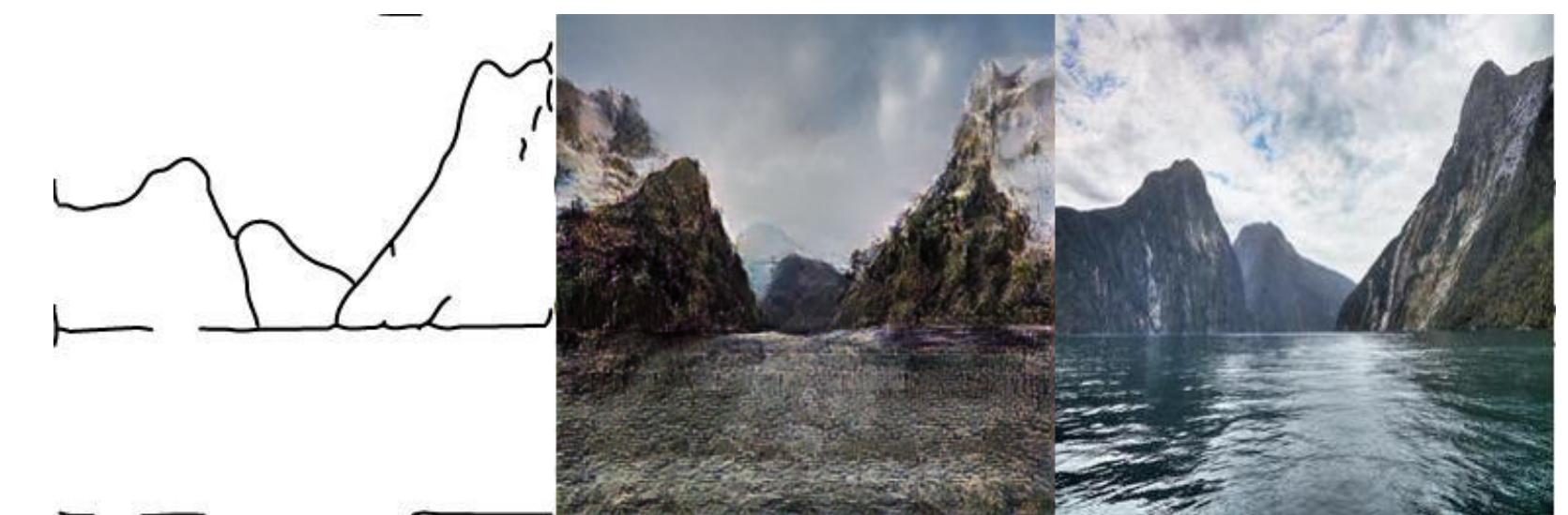
Scene Model: Qualitative Results



Sketch

Generated

Original



Sketch

Generated

Original

D E M O

Quantitative Comparison

| Measures | Object Model | Scene Model |
|---|--------------|-------------|
| SSIM (Closer to 1 the better) | 0.535 | 0.279 |
| PSNR (30dB - 50dB for good image) | 10.9 dB | 12.16 dB |
| FID (close to 0 the better; 35 for good image) | 120 | 88.96 |
| LPIPS (below 0.1 or even lower) | 0.44 | 0.593 |

Discussion & Conclusion

Key Observations

- DeepBlocks improved texture realism
- LPIPS boosted sharpness + perceptual quality
- Scene model harder than object model
- Discriminator stabilized well after 100+ epochs

Challenges

- Long training time
- Limited paired datasets
- Sensitive to hyperparameters
- Some outputs blurry in detailed regions

Discussion & Conclusion

Conclusion

- Successful application of Pix2Pix with enhancements
- Two well-performing models
- Achieved realistic image synthesis from sketches

Future Work

- Train on more diverse datasets
- Use attention U-Net
- Try alternative sketch-to-image approaches
- Improve frontend interactivity

Thank You

Any Queries?

Sandip Katel
078BCT077

Sharad Pokharel
078BCT083

Yujal Shrestha
078BCT096