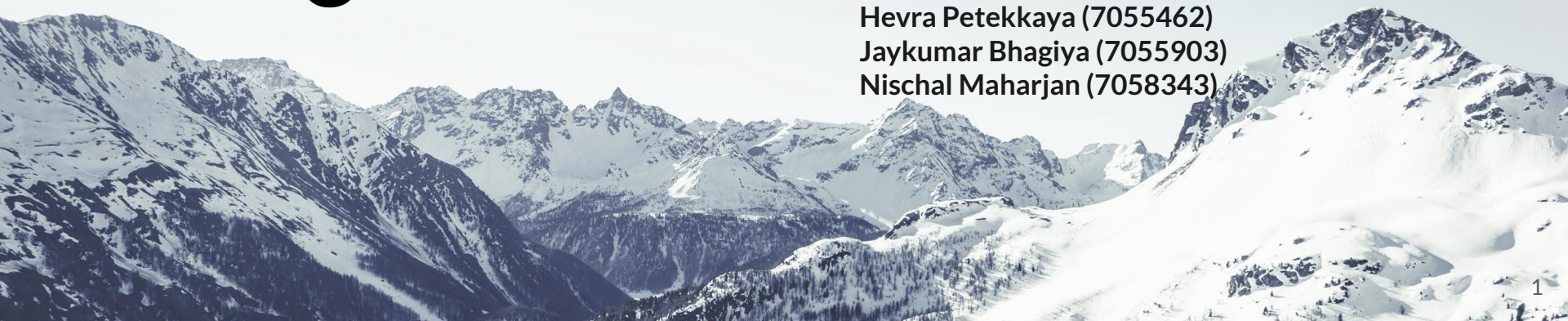
Generation of Real World Images from Simulated Images

Members:
Hevra Petekkaya (7055462)
Jaykumar Bhagiya (7055903)
Nischal Maharjan (7058343)



OUTLINE

1 Introduction

2 Literature Review

3 Image Translation Pipeline

4 Vision Transformer

5 Swin Transformer

6 Results

Motivation

- Need of huge amount of data

- **Distribution shift** occurs when trained on simulated data

Problem Statement

- Given simulated images, generate realistic looking images

Simulated image

Real image

Dataset

- We have used Kitti and Virtual Kitti Dataset to generate image pairs

- 2126 image pairs

Source

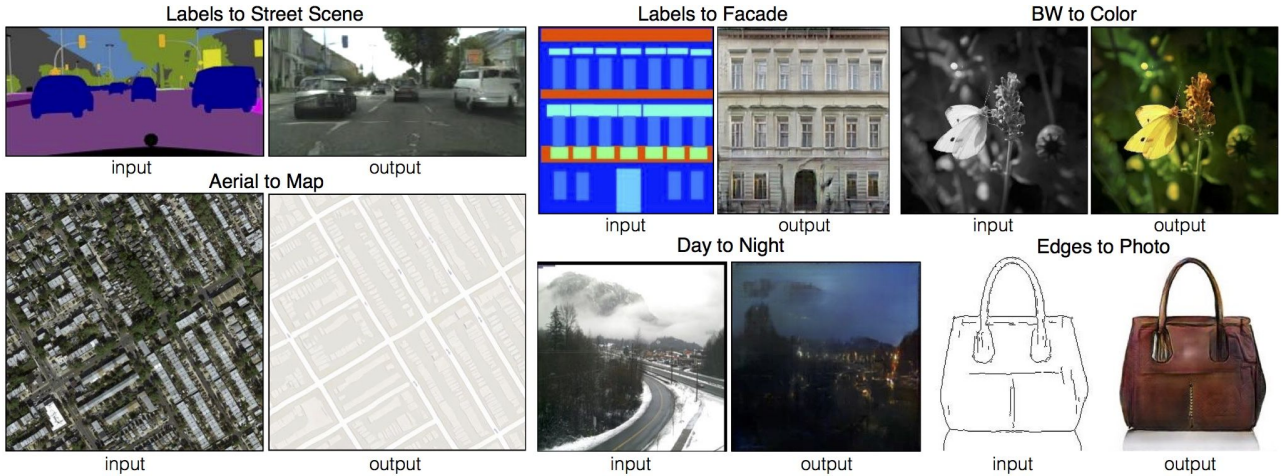
Source

Source

Image Pairs

Literature Review

- We are trying to achieve **Image Translation**



Pix2pix : Image-to-Image Translation with Conditional Adversarial Networks Source

Literature Review

- Autoencoder with **euclidean loss**
- Usually Blurry image
- Hence use adversarial loss

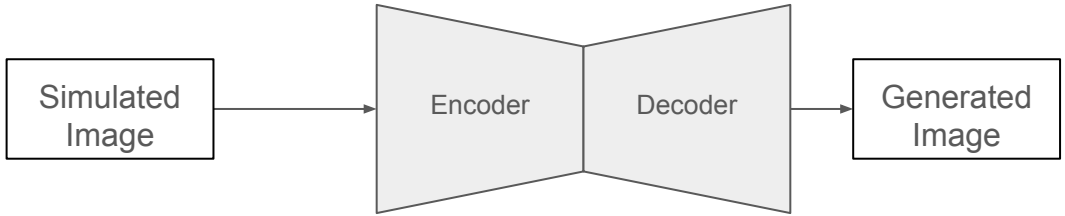
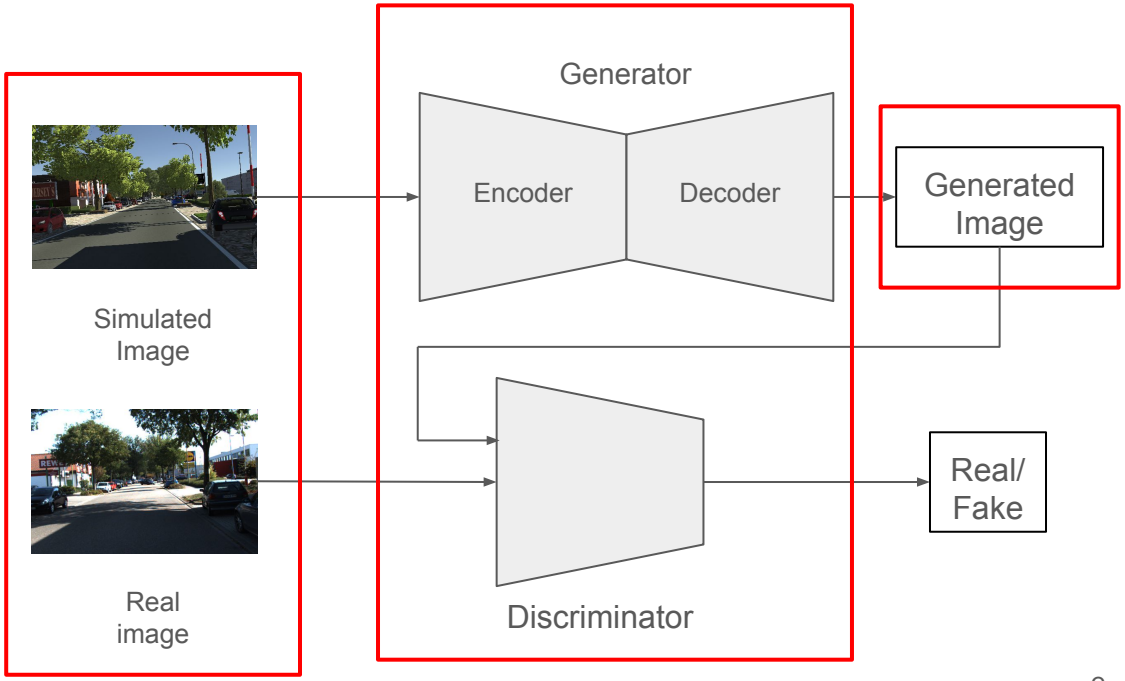
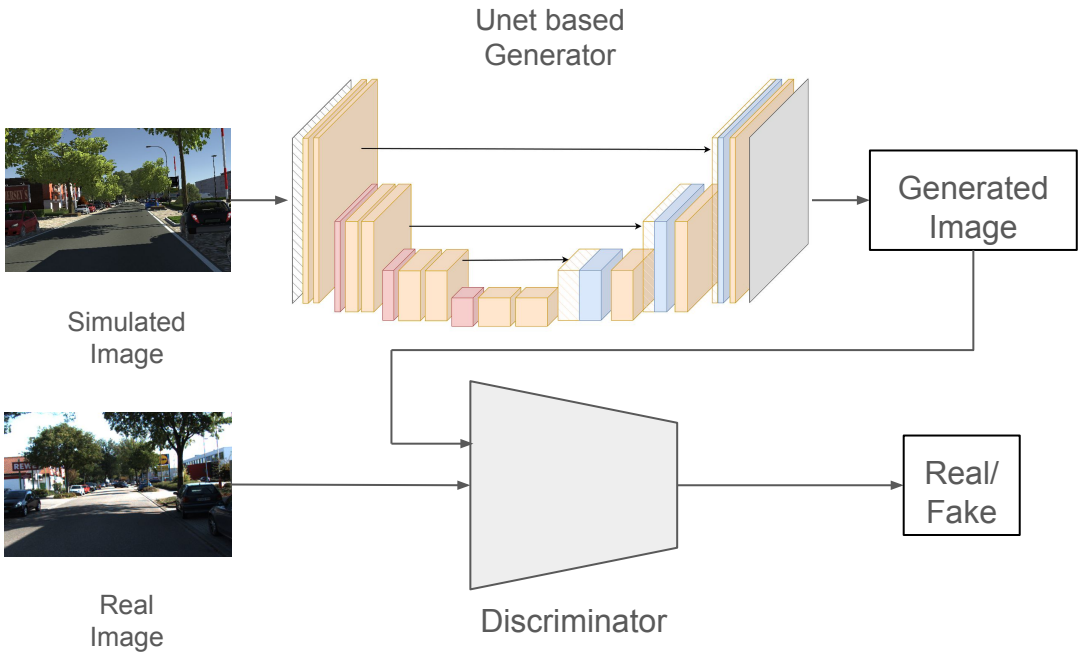


Image Translation Pipeline

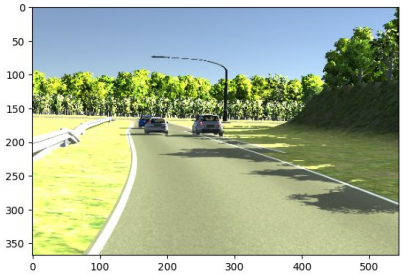
- **Input:** Image pairs
- **Output:** Generated image
- **Architecture:** Generator and Discriminator



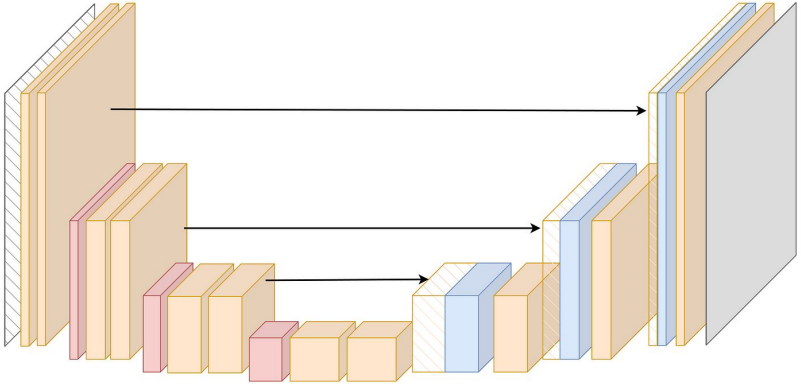
Unet based approach



Unet based Generator



Synthetic Image

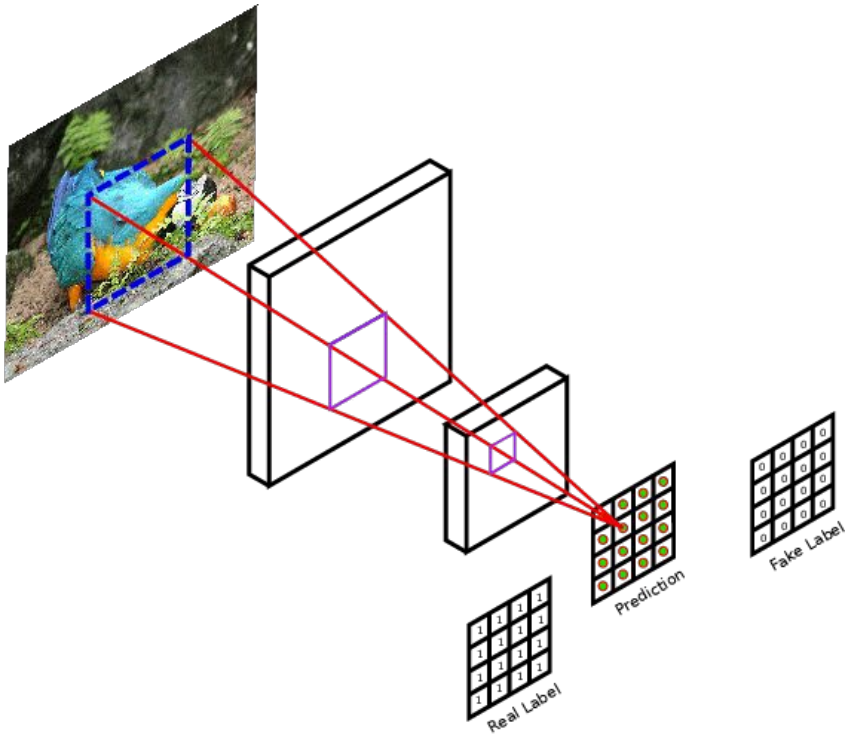


Source

Realistic Looking Image

Discriminator: Patch Gan

- Instead of single output for whole image we have outputs for **each patch**
- Improves the quality of generated images
- Reduces computational burden



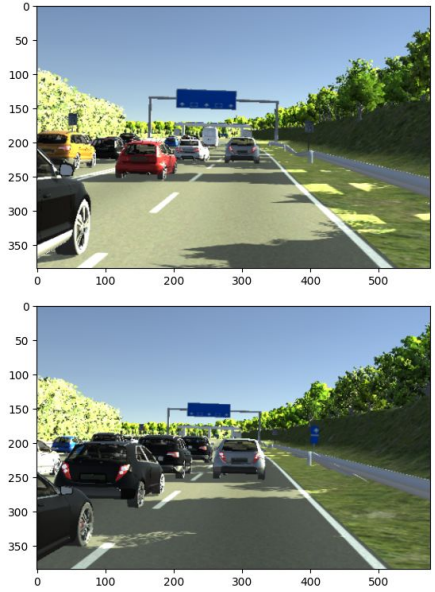
Source

Results

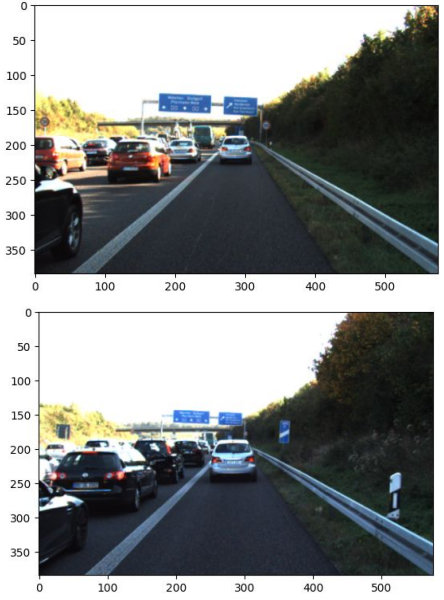
Unet Based	RMSE↓	Perceptual↓	Inception↑	FID Value↓
Unet GAN	**0.180**	**0.0482**	**3.5±0.24**	**259.86**
Unet GAN Aug	0.196	0.0502	2.73±0.09	307.30

Generated Images

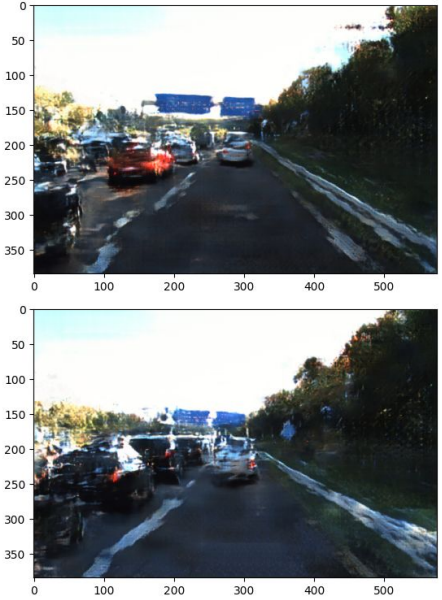
Input Image



Ground truth

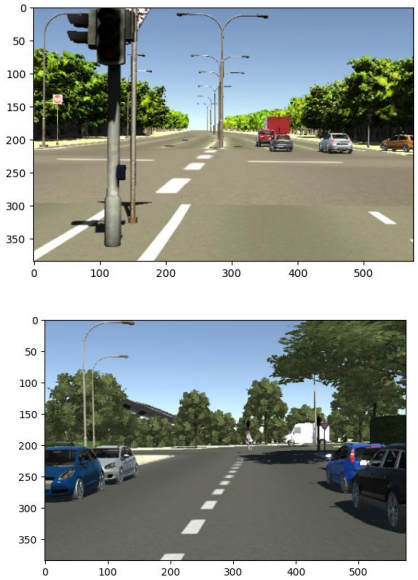


Generated Image

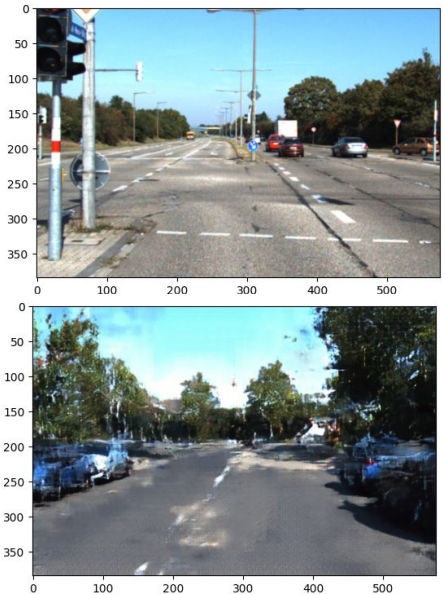


Generated Images

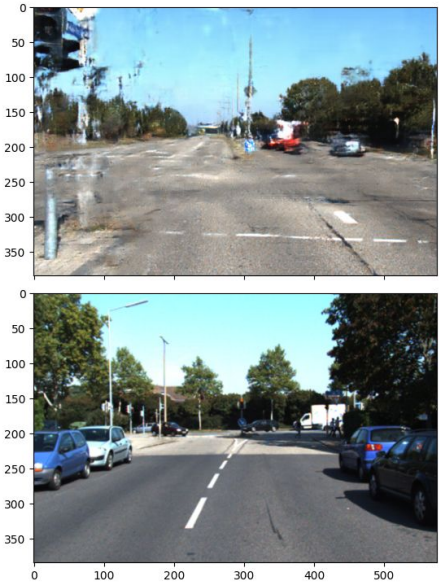
Input Image



Ground truth

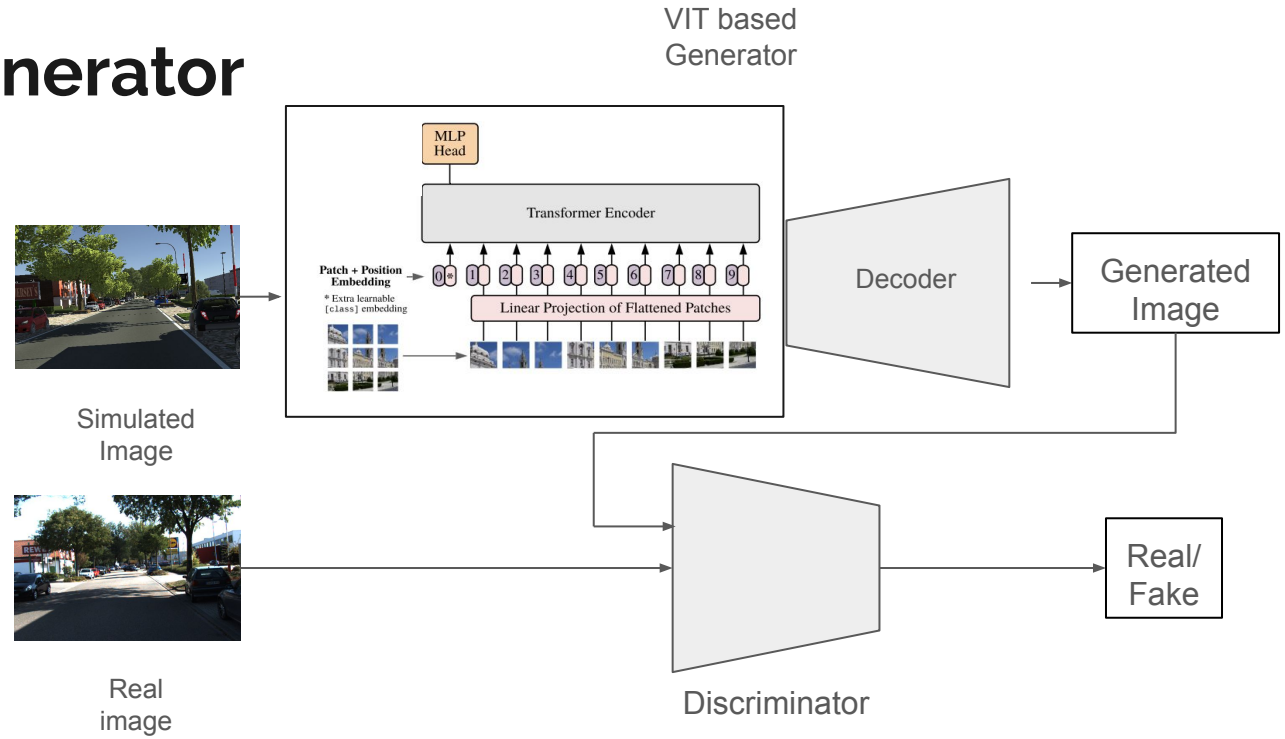


Generated Image



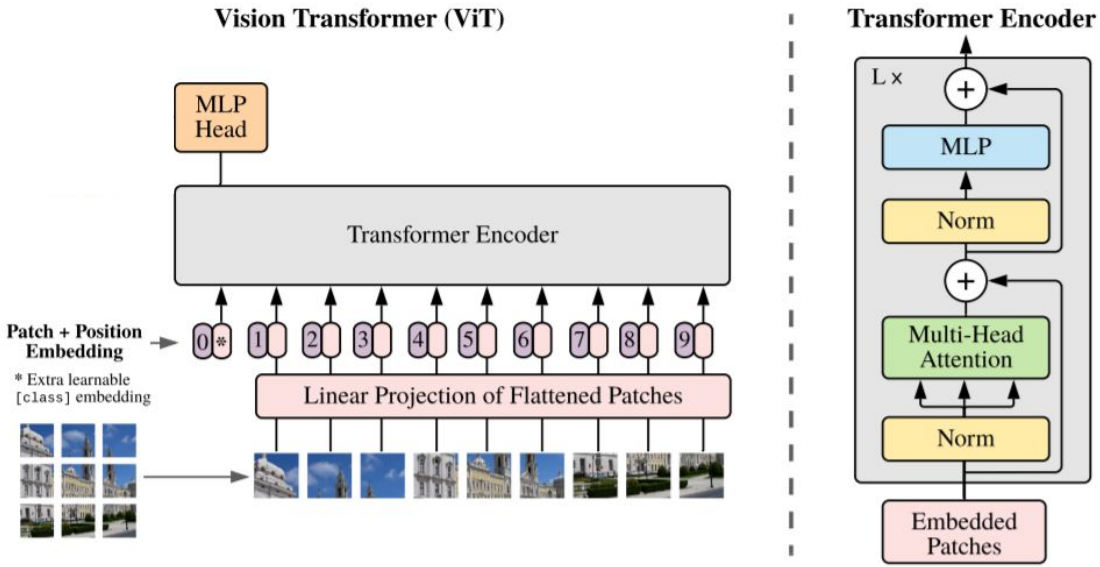
VIT based Generator

- VIT model as Encoder
- Generates image latent representation



VIT based Generator

- Uses Attention Mechanism



Source

Decoder in Generator

- We use decoder similar to that of Unet architecture but without skip connection

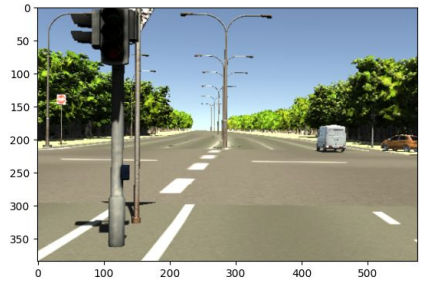
Results

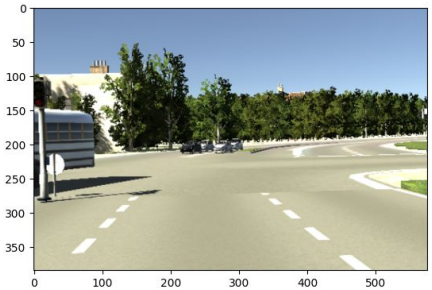
Model (Patch size 16)	RMSE	Perceptual	Inception	FID Value
ViTGAN-Complex	**0.151**	0.0388	3.01±0.25	**210.33**
ViTGAN-Color	0.169	**0.0321**	2.66±0.18	267.12
ViTGAN-Aug	0.165	0.0365	**3.10±0.29**	280.90

Model (Patch size)	RMSE	Perceptual	Inception	FID Value
ViTGAN-8	**0.144**	**0.0316**	**3.19±0.19**	**191.85**
ViTGAN-16	0.154	0.0515	2.80±0.12	252.30
ViTGAN-32	0.162	0.0412	2.99±0.39	283.78

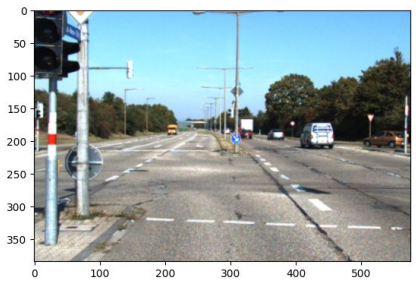
Generated Images

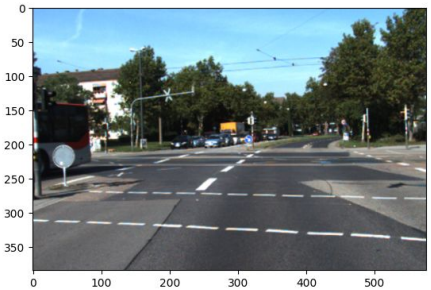
Input Image



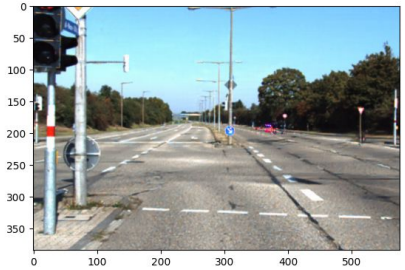


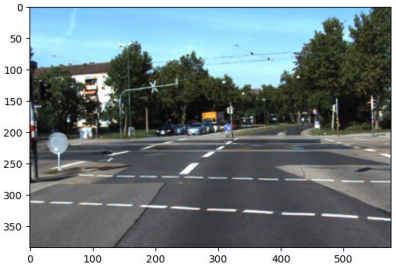
Ground truth





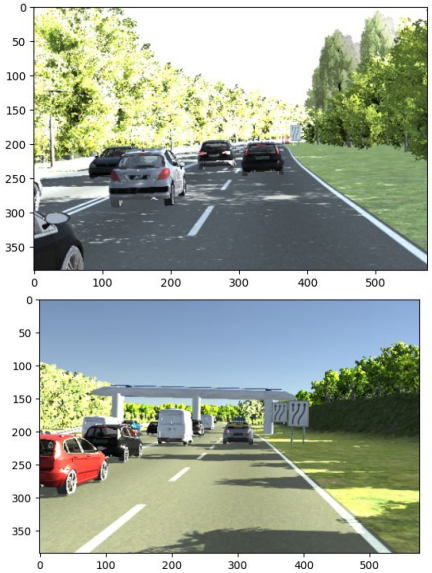
Generated Image



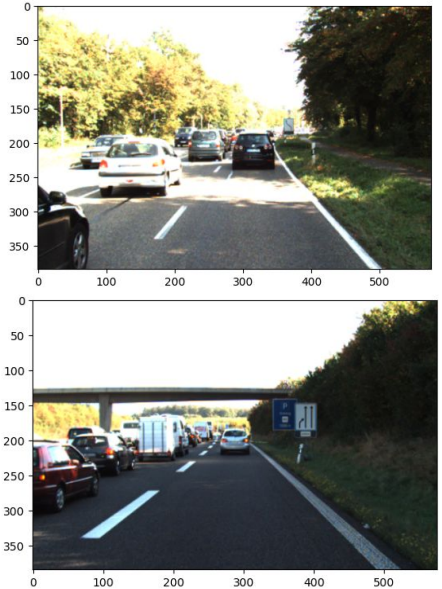


Generated Images

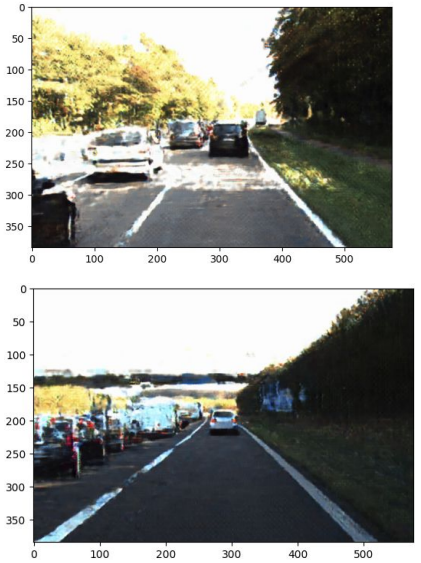
Input Image



Ground truth

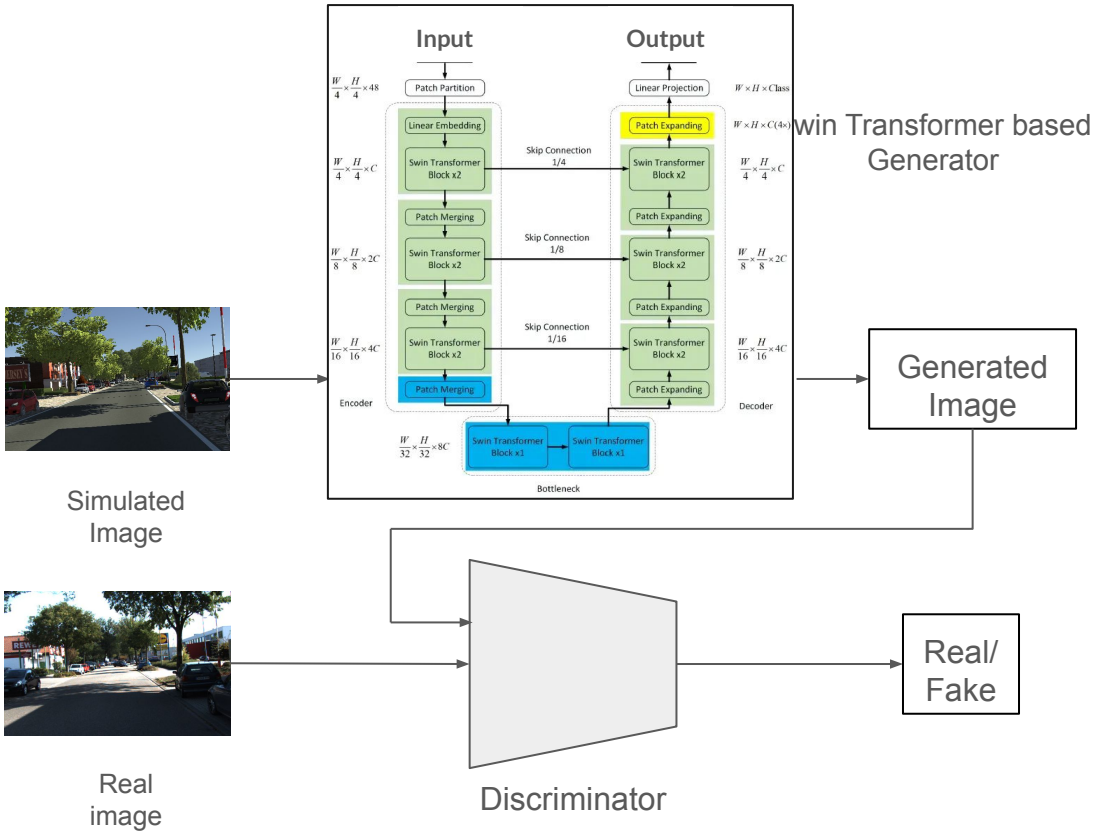


Generated Image



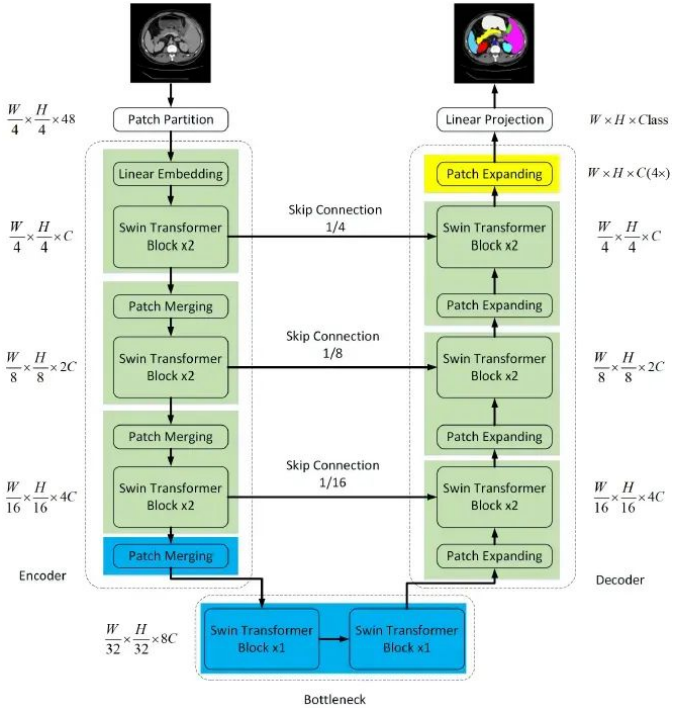
Swin Transformer based Generator

- Use of Swin transformer for generator



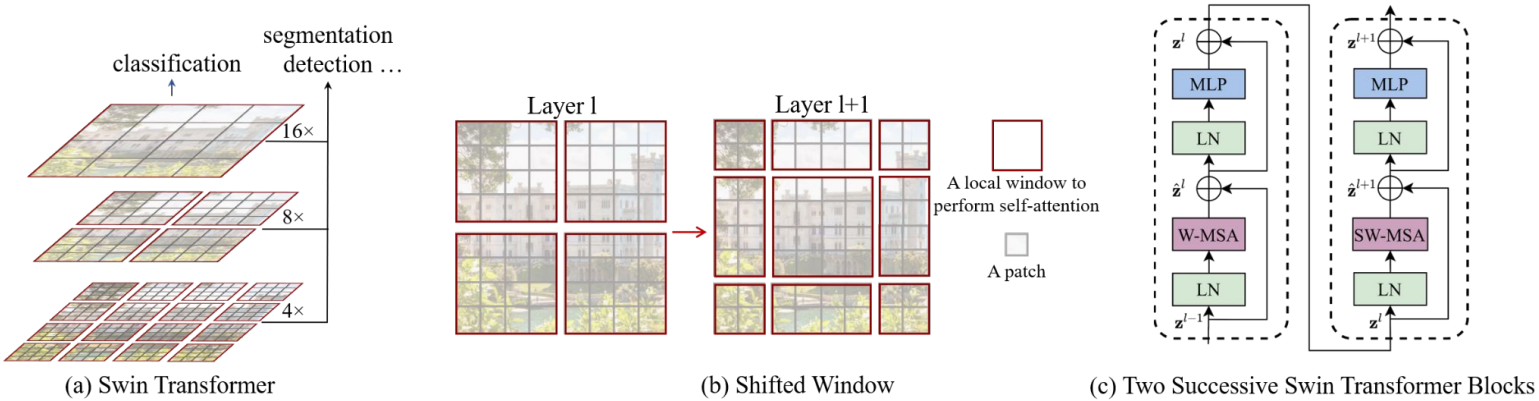
Swin Transformer based Generator

- Variant of VIT with hierarchical structure



Source

Swin Transformer based Generator

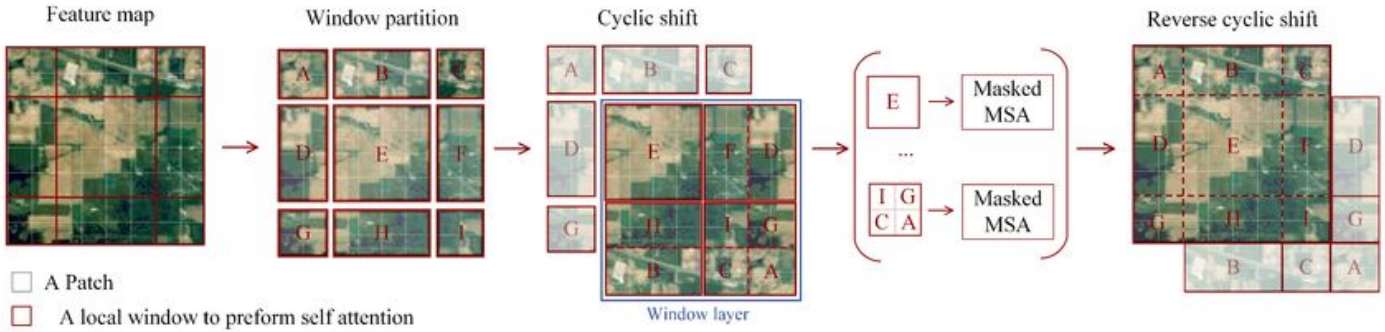


(a) Swin Transformer (b) Shifted Window (c) Two Successive Swin Transformer Blocks

Source

Uses concept of shifted window

Swin Transformer based Generator



Source

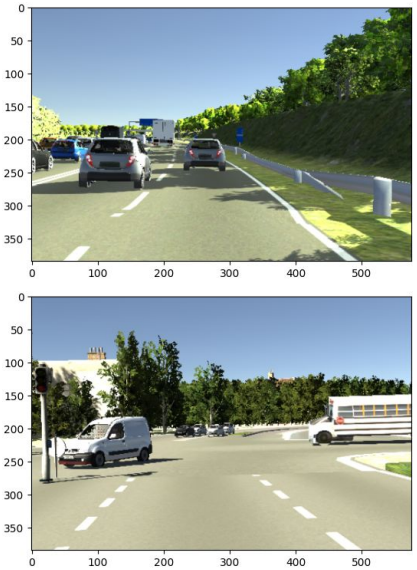
Uses Cyclic shift instead of padding for efficient computation

Results

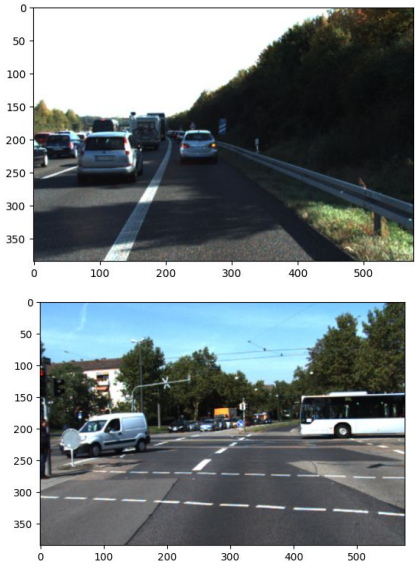
Swin (Window)	RMSE	Perceptual	Inception	FID Value
Swin (6, 6)	0.224	0.0600	1.79±0.10	435.18
Swin (12, 6)	**0.204**	**0.0419**	**3.24±0.09**	429.64
Swin-Color (12, 6)	0.236	0.0484	2.10±0.12	**416.57**

Generated Images

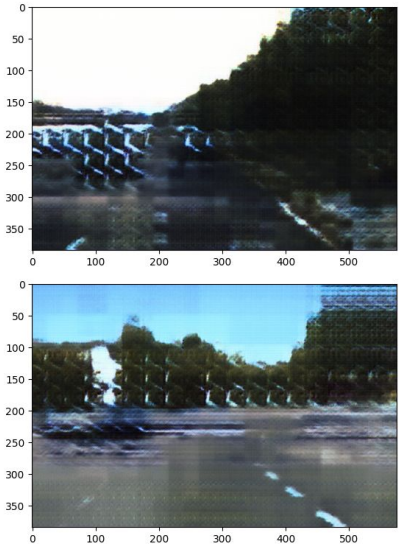
Input Image



Ground truth

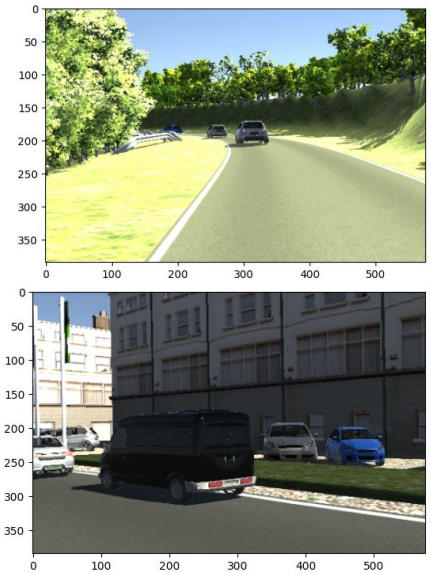


Generated Image

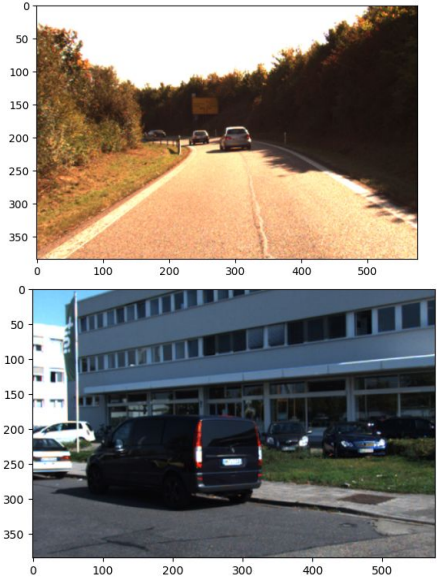


Generated Images

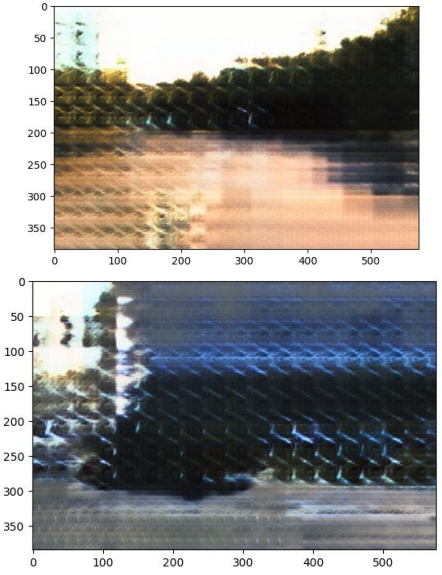
Input Image



Ground truth



Generated Image



Comparison

Model	RMSE	Perceptual	Inception	FID Value
ViTGAN-8	**0.144**	**0.0316**	3.19±0.19	**191.85**
Swin (12, 6)	0.204	0.0419	3.24±0.09	429.64
Unet GAN	0.180	0.0482	**3.5±0.24**	259.86

Thank you.