Deep Learning-Based Automatic Map Generation from Satellite Imagery

Project Guide :- **Prof. Surajit Manna**

Presented By -

Soumik Das
Arghadip Biswas
Santu Maity

Introduction

The rise of satellite and aerial imaging has unlocked unprecedented Earth observation capabilities, producing vast amounts of high-resolution data. However, manually interpreting this data is slow, inefficient, and error-prone. Various sectors like urban planning, agriculture, and disaster response increasingly rely on accurate spatial information. This project aims to automate the conversion of raw satellite imagery into structured map formats, enabling faster and more reliable geospatial analysis.

Objective

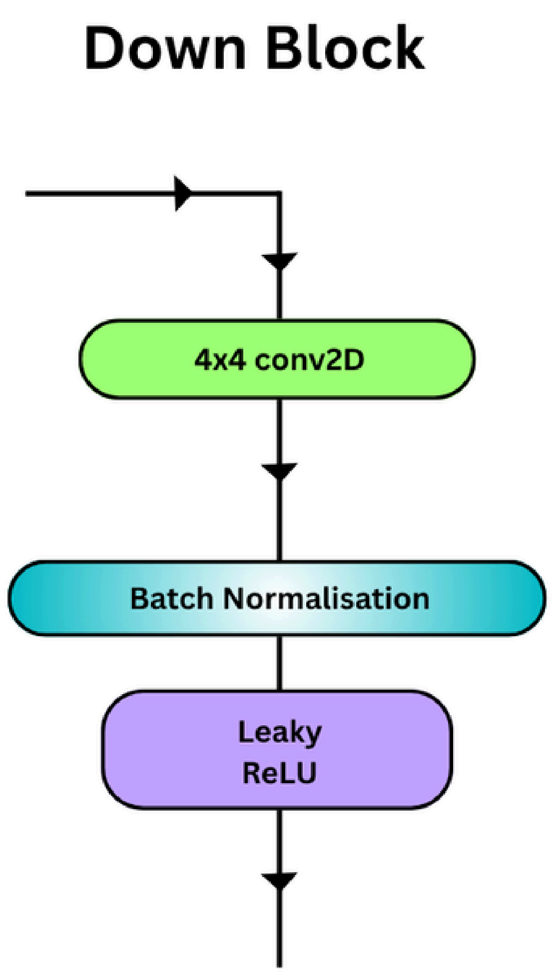
- To develop an automated pipeline for converting satellite and aerial images into stylized map representations.
- To leverage Generative Adversarial Networks (GANs), specifically the Pix2Pix architecture, for initial image-to-image translation.
- To demonstrate the practical utility of the developed system in generating high-quality map data for potential applications in urban planning, infrastructure management, and geospatial analysis

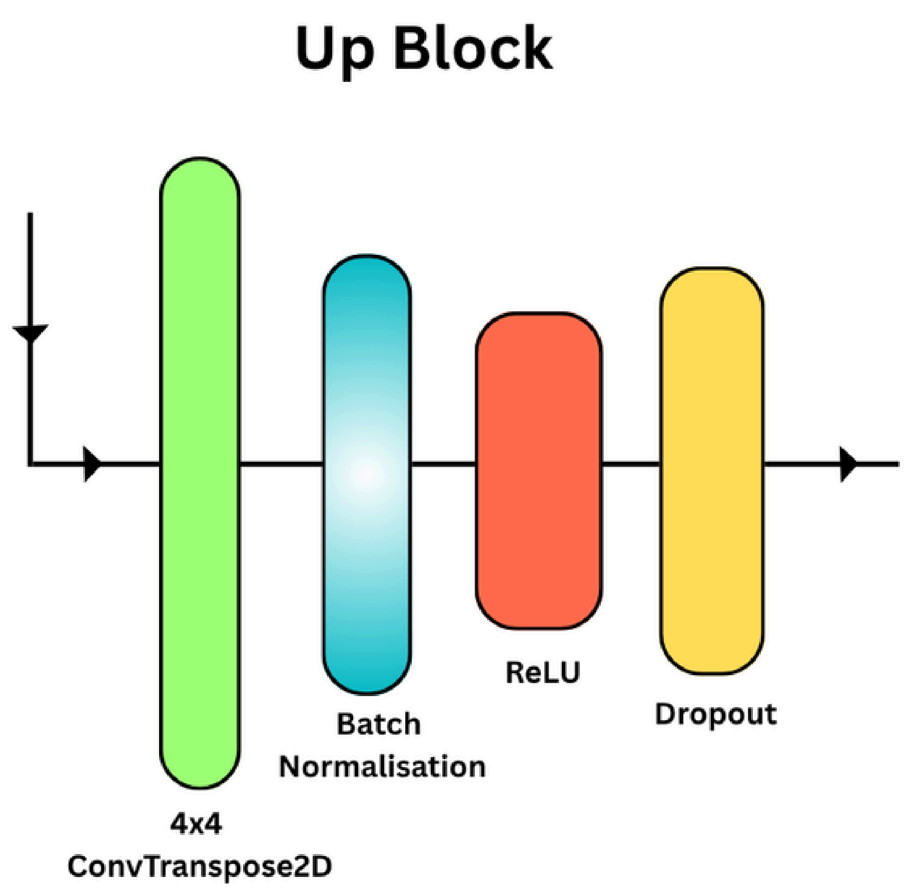
Dataset and Preprocessing

In this project, two key datasets were utilized to train and evaluate the image-to-map generation model. The **SpaceNet** Dataset provided high-resolution satellite images of diverse urban environments, offering realistic and complex visual data essential for learning spatial structures. These images captured various building layouts, roads, and landscapes, making them ideal for training a mapping system.

Alongside SpaceNet, the **Pix2Pix** Map Dataset was employed, which contains paired satellite images and corresponding stylized map renderings. This alignment enabled supervised learning where the model could learn pixel-wise mappings from raw satellite visuals to cartographic representations. Prior to training, all images underwent normalization and preprocessing to ensure consistency in scale, lighting, and contrast, thereby improving the model's ability to generalize across different geographic regions.

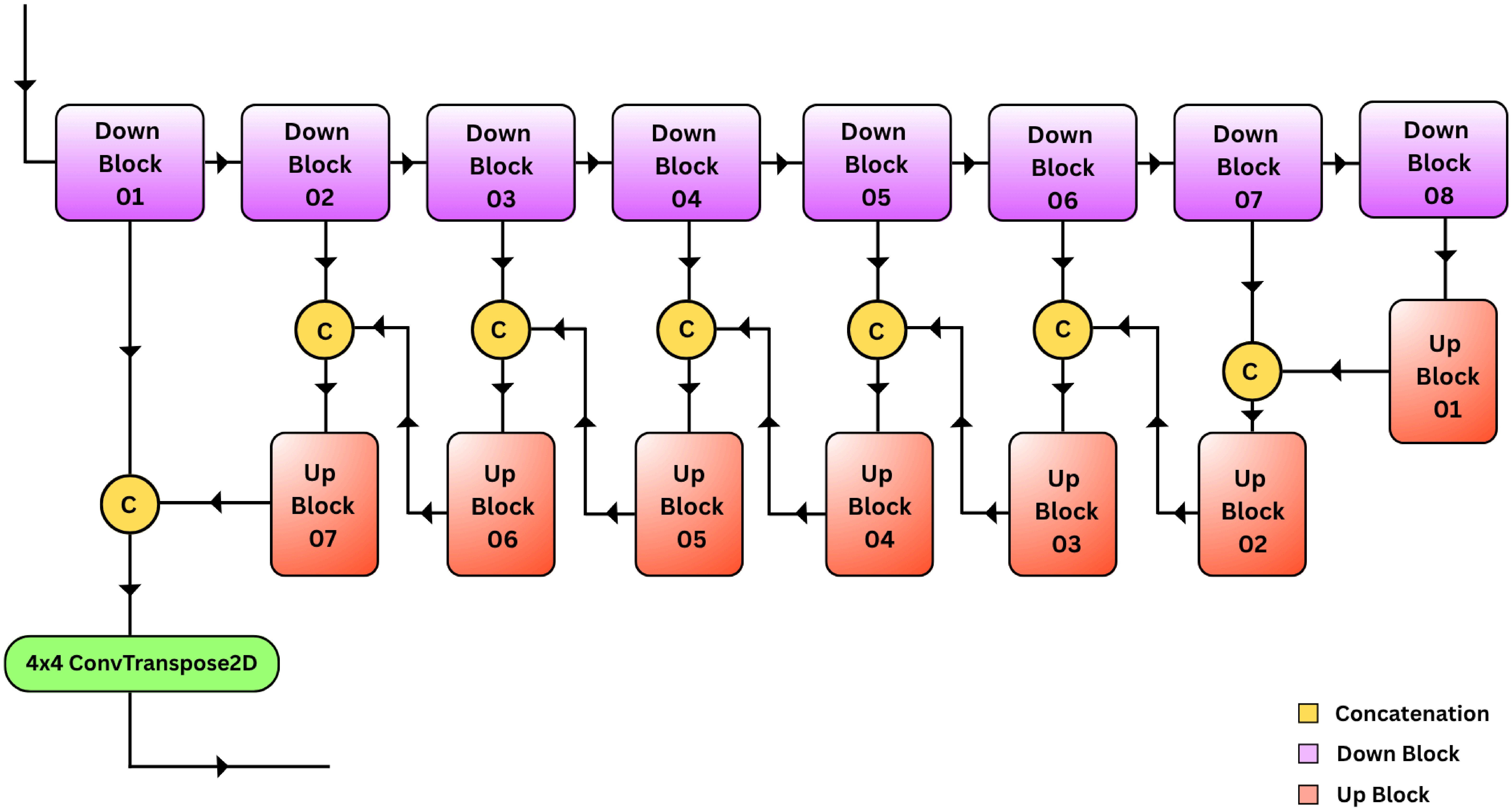
Blocks used in our Model





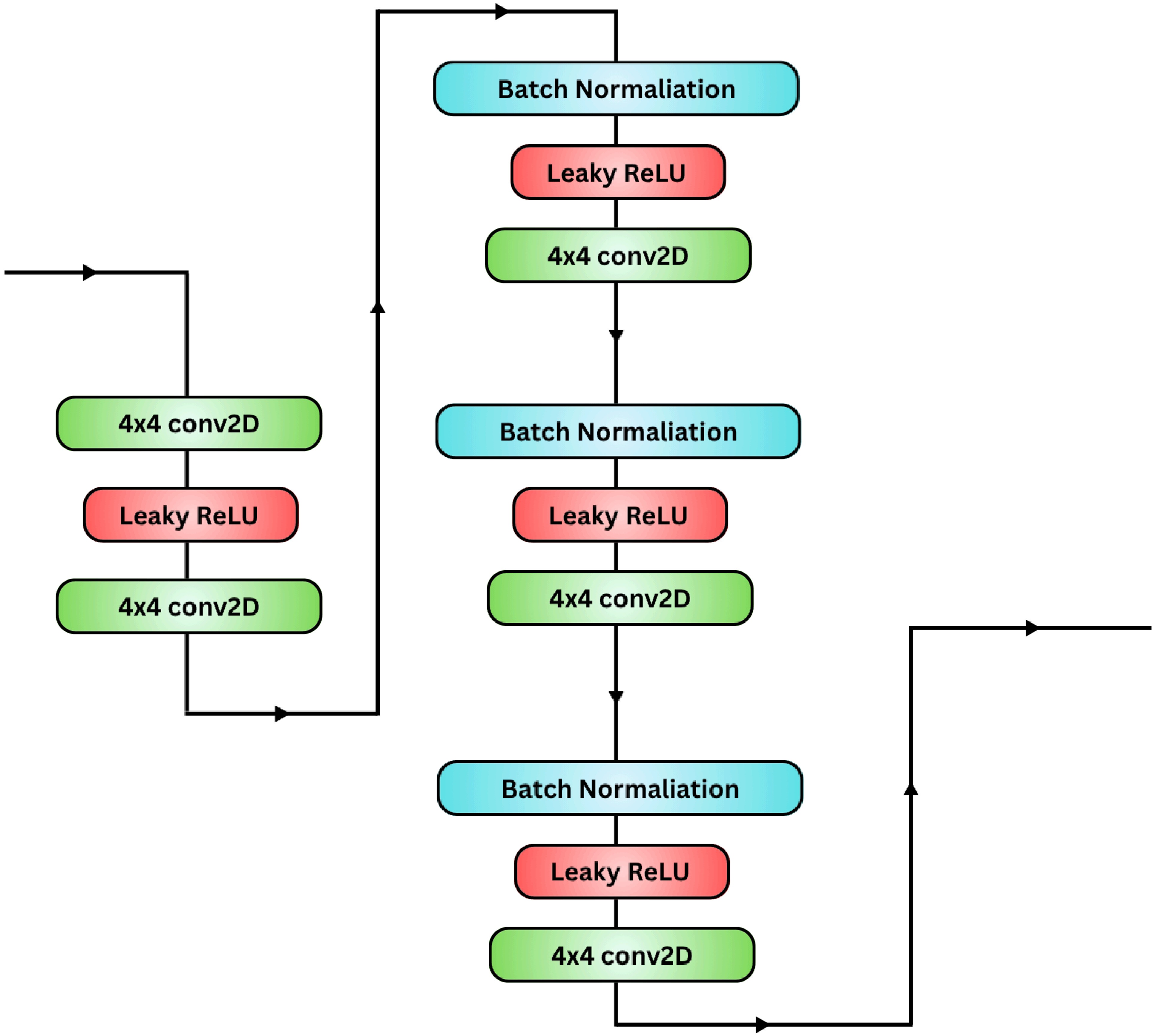
This Up Block and Down Block are used to Contruct the GeneratorUNet Architecture of the Generative Adversial Network or GAN

Architecture of the GeneratorUNet



Architecture of Patch Discriminator

The Discriminator in our model evaluates the realism of generated maps by classifying individual patches rather than the entire image. This approach focuses on local structures, helping the model learn fine-grained details like edges and textures. By comparing real map patches with those from the generated output, PatchGAN effectively guides the generator toward producing sharper and more realistic maps. Its localized feedback mechanism ensures that both global layout and local accuracy are preserved during training.



Refinement Process

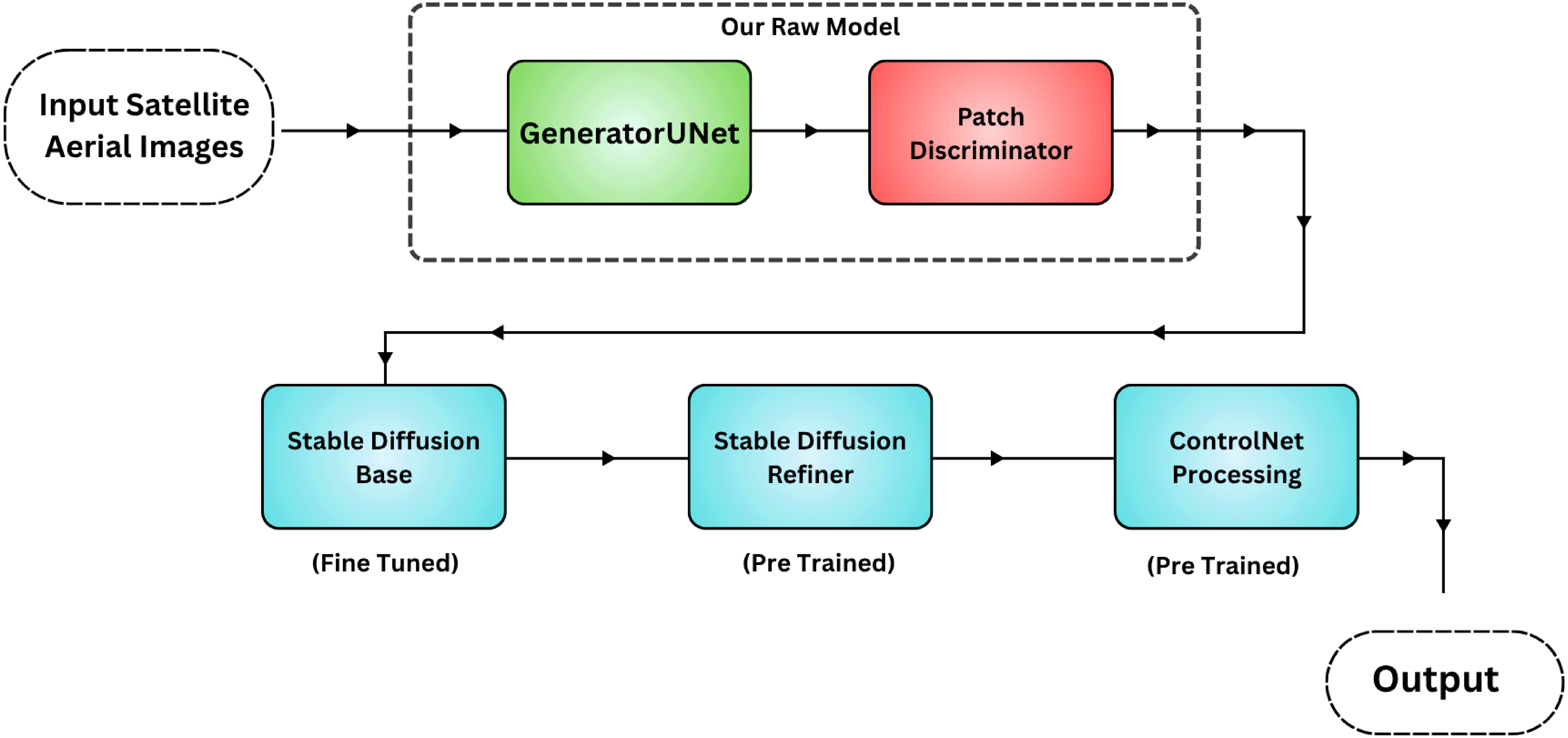
After the initial map is generated by the custom GAN architecture, it is passed through a fine-tuned **Stable Diffusion Base model** to enhance its cartographic style. This model was pre-trained and then specifically fine-tuned using actual Google Maps-style images, enabling it to learn the unique noise patterns and stylistic traits of professional maps. The refinement process uses a diffusion-based approach where the model learns to denoise the GAN-generated image by minimizing the difference between predicted noise and the noise patterns learned from the fine-tuning dataset. This stage significantly improves the overall map clarity, texture consistency, and stylistic alignment with real-world map representations.

Final Enhancement Models

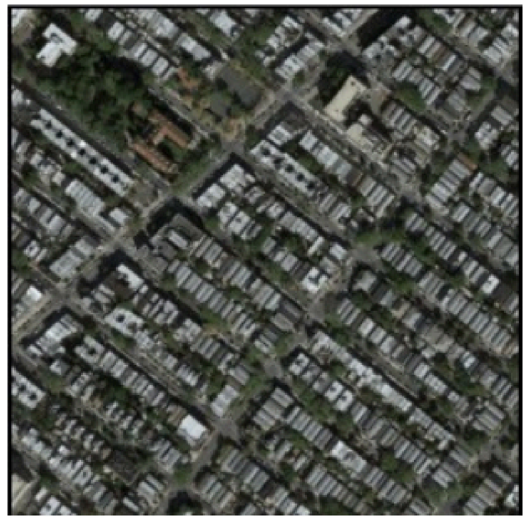
The output from the base refinement stage is then fed into **Stable Diffusion XL Refiner**, a high-capacity model designed to further polish visual details. This refiner improves edge sharpness, reduces blur and noise, and enhances the overall visual fidelity of the map. Following this, the image undergoes a final transformation using **ControlNet**, a model that focuses on structural guidance.

ControlNet extracts and reinforces important features such as road layouts, building boundaries, and landscape edges using techniques like Canny edge detection. By anchoring the refinement to these critical structures, ControlNet ensures the map is not only visually clean but also geometrically accurate and practically usable for navigation and planning purposes.

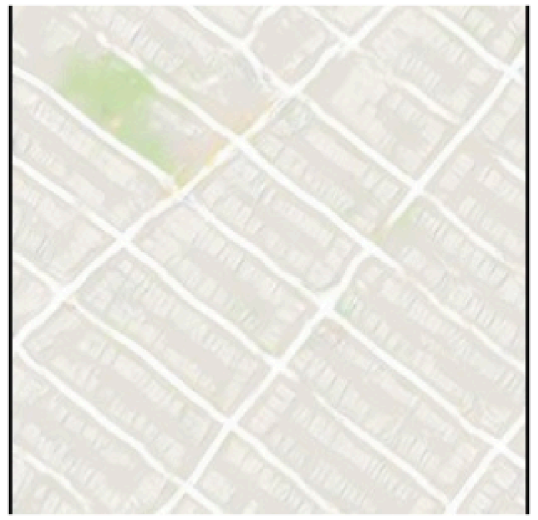
Our Model Pipeline



Model Analysis



Satellite Images



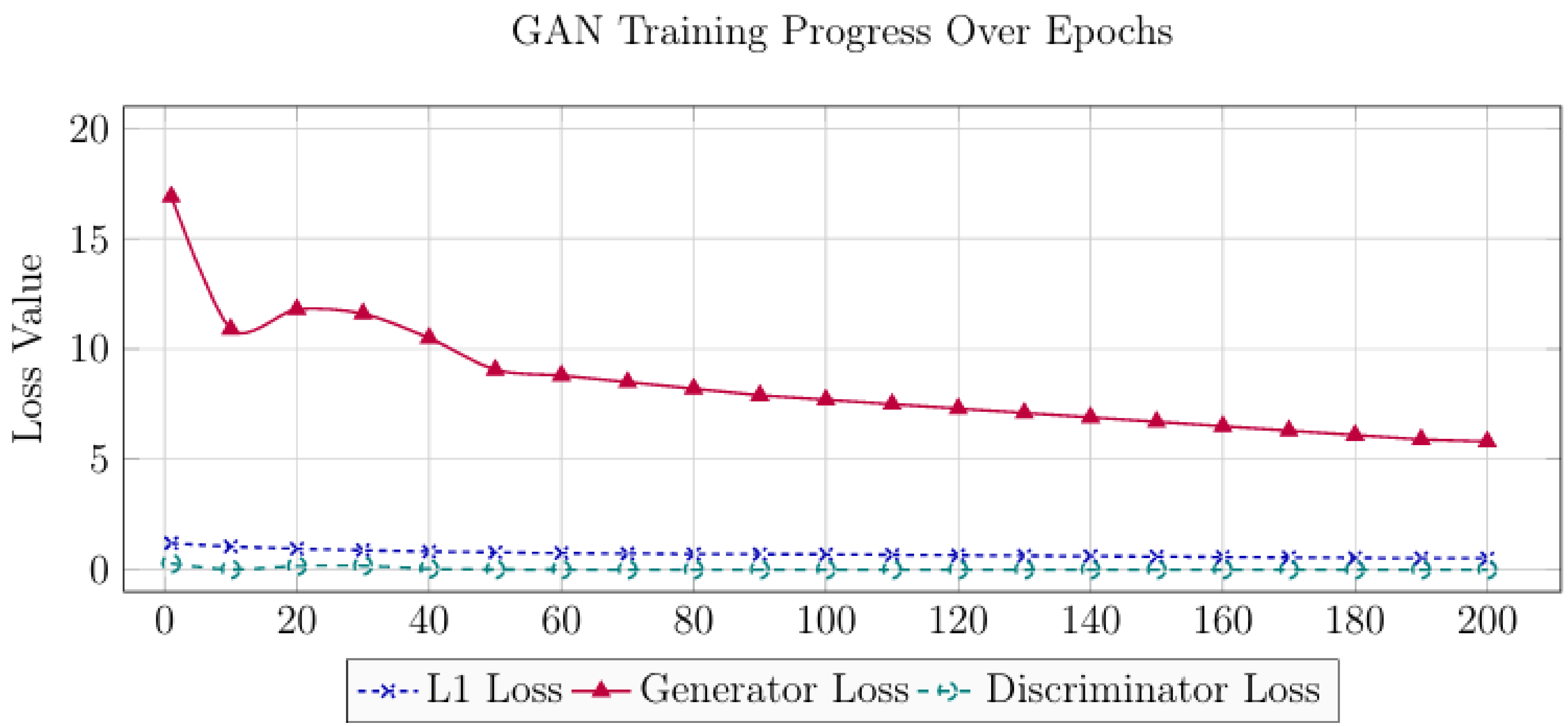
Output of
GeneratorUNet +
Patch Discriminator

Output of
Stable Diffusion
Base (Fine Tuned)

Output of
Stable Diffusion
Refiner (Fine Tuned)

Output of
ControlNet
Processing (Fine Tuned)

Evaluation Metrics



- L1 Loss: Measures pixel-wise difference between generated and real maps; decreases steadily as output becomes more accurate.
- Discriminator Loss: Fluctuates as it learns to distinguish real vs. fake images; stabilizes when generator improves.
- Generator Loss: Initially high; reduces over time as generator produces more realistic maps that fool the discriminator.

Comparative Analysis

Method	Architecture	PSNR (dB)	SSIM	FID	Key Strengths	Limitations
Goodfellow et al. (2014)	Vanilla GAN	18-20	0.55	90-100	Introduced GAN framework	Struggles with fine-grained detail
Isola et al. (2017)	Pix2Pix (cGAN + U-Net)	21-23	0.68	70	Supervised image-to-image translation	Output often blurry; lacks refinement
Ronneberger et al. (2015)	U-Net (Segmentation)	22-24 (on segmentation)	0.70	N/A	Semantic segmentation excellence	Not suitable for realistic image synthesis
Our Proposed Method	**GAN (Pix2Pix-inspired) + SD-Base + SD-XL Refiner + ControlNet**	**28.5**	**0.87**	**22.7**	**Stylized, smooth, and structurally accurate maps; integrates multiple refinement stages**	**Slightly longer inference time; dependent on multiple models**

System Requirments

GPUs:

- NVIDIA Tesla P100 (Kaggle)
- NVIDIA T4 (Kaggle Dual GPU Environment)
- NVIDIA RTX 3060 (Local Machine)

CPU:

- Intel Core i7

Frameworks & Libraries:

- PyTorch, OpenCV, Albumentations for model development and preprocessing
- Diffusers, HuggingFace Transformers, ControlNet APIs for map refinement and diffusion-based generation

Future Works

- Future work for this project involves enhancing its practical usability and scalability. One major goal is to integrate real-time satellite data streams, enabling up-to-date and dynamic map generation for time-sensitive applications like disaster monitoring or urban planning.
- Additionally, extracting vector data (e.g., roads, buildings) from the generated maps will enable seamless compatibility with Geographic Information System (GIS) software, greatly expanding its utility.
- To ensure broader accessibility, model optimization for edge devices and mobile platforms is also planned, involving techniques such as model pruning, quantization, and lightweight architecture design to enable fast and efficient inference in low-resource environments.

Conclusion

This project presents a robust satellite-to-map image translation pipeline combining GAN-based generation with ControlNet-Canny and Stable Diffusion XL refinement. The approach delivers high-quality, geometrically accurate, and stylistically polished maps. By leveraging Canny edge detection and spatial conditioning, the system achieves superior detail and visual appeal. It holds strong potential for real-world applications in urban planning, GIS, and disaster management.

THANK YOU...