TRANSFORMING SATELLITE IMAGES INTO GOOGLE MAPS FORMAT

Salma Elmalky
Student# 1010855371
sal.elmalky@mail.utoronto.ca

Muhammad Ahmed
Student# 1010603839
muhd.ahmed@mail.utoronto.ca

Mohammed Suwan
Student# 1009842093
mohammed.suwan@mail.utoronto.ca

Chenhan Xu
Student# 1010874047
chenhan.xu@mail.utoronto.ca

ABSTRACT

The methods of image-to-image translations for utilizing the ever-growing supply of satellite images has been constantly explored in recent years. Deep neural networks are used to classify and collect information from images. While the purpose of maps remains fundamental for our society, different research has been conducted to manipulate the image-to-image translation for generating maps from aerial photos and satellite images. The purpose of this paper is to build on the research by looking into conditional generative adversarial networks (cGANs) to transfer Pix2Pix images and output a Google-like format of maps. Using the network the model is trained on map and satellite pairs to output clearly visual geographical mappings. Within the network will be a U-net generator and PatchGAN discriminator both following a convolutional neural network (CNNs) architecture to evidently classify and output images. The research and project aims to showcase the effective methods of artificial intelligence on transferring satellite data into user-friendly maps.

1 INTRODUCTION

Maps are an integral part of society since their creation, used to plan infrastructural projects, environmental responses, navigation, exploration, and much more. Initially, Google Maps relied on aerial images to capture and transfer geospatial information globally. However, through the application of deep learning for converting satellite images to a map format, more cost-effective and reliable data collection methods were utilized. This deep learning process has been in use since 1960, while its field of study continues to expand for further improved ways of image-to-image translation.

The project of translating satellite images into a Google Maps format will focus on utilizing deep neural networks. Specifically, the Generative Adversarial Networks (GANs) that combine two convolutional Neural Networks (CNNs) to classify global locations from satellite images into a readable map format. The goal of our project is to aid in processing the growing availability of satellite images from Google Earth and satellites alike to produce clear, readable maps of sufficient accuracy and variety. Enabling easy user usability that is practical and efficient.

As aforementioned, the team will utilize conditional GANs split into a generator (Unet) and a discriminator (PatchGAN), comprised of CNNs. By inputting paired supervised data of maps, to train our model to properly output the correct map of a given location through data augmentation. The neural network should facilitate accurate satellite-to-map conversion, enabling an effective methodology of updating and producing global maps.

2 ILLUSTRATION

Figure 1 below illustrates our project model. The satellite image dataset undergoes data processing and is then passed into a conditional GAN architecture consisting of a U-Net generator, which produces a map image, and a PatchGAN discriminator is trained on pairs of real input images and either real or generated output images.

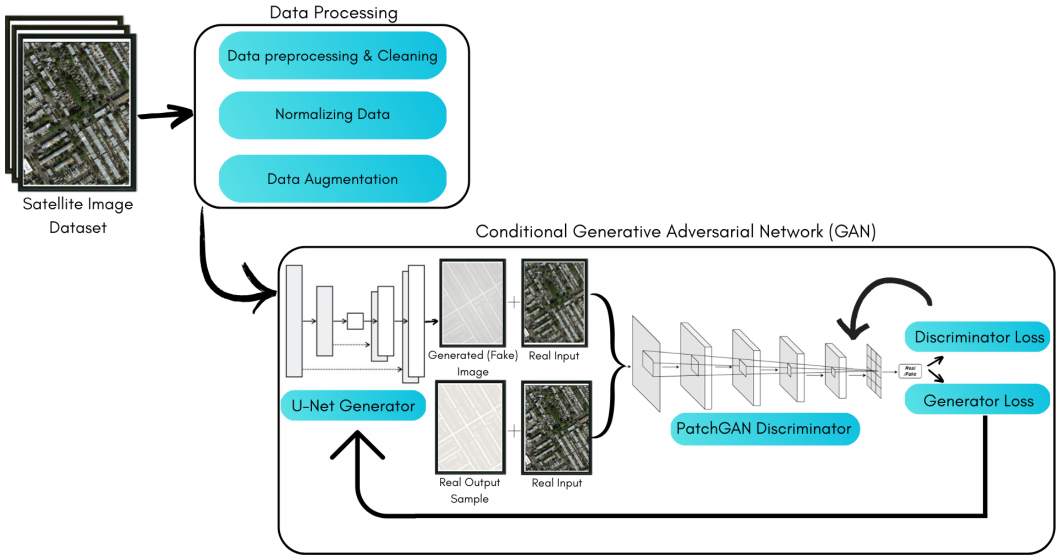

Figure 1: Project Overview

3 BACKGROUND RELATED WORK

The process of translating satellite images into a map format is being constantly researched, specifically through the use of Pix2Pix GANs to do so. Throughout recent years, more developments have been made related to this topic, considering different methods of manipulating neural networks to achieve the best results of image-to-image translation models. The following section elaborates on some of the previous models and research produced within this area.

3.1 SATELLITE IMAGE TO MAP TRANSFORMATION USING GENERATIVE ADVERSARIAL NETWORK

Pokhrel et al.(2023) of Tribhuvan University, Nepal, and the Asian Institute of Technology, Thailand, developed a generative adversarial network (GAN) model including two parts: a generator and a discriminator. The latter looks at examples generated, while the generator produces new data samples. The specific GAN model was built using Pix2Pix, a conditional GAN to allow for the conversion of satellite images to maps. Their model was trained up to 80,000 iteration steps with a batch size of one. Producing an SSIM (Structural Similarity Index between -1 and 1) score of 0.75 for their Unet with Receptive Fields Blocks model that reduces multiple small kernels compared to one large kernel.

3.2 IMAGE-TO-IMAGE TRANSLATION WITH CONDITIONAL ADVERSARIAL NETWORKS

Isola et al.(2018) from the Berkeley AI Research (BAIR) Laboratory, UC Berkeley. Their paper introduced Pix2pix and focuses on demonstrating the loss function of training a model on image-to-image translation with a conditional adversarial network. Highlighting its effective ways of classifying land objects and redesigning figures from edge maps. Their model utilizes a Unet generator as well as a convolutional PatchGAN discriminator, comparing their L1 + cGAN model to the L1 and L2 models. Amongst their experiments of map-to-map conversions was their maps to aerial photos using data from Google Maps. They validated their data through two methods, one being the ”real vs. fake” perceptual study on Amazon Mechanical Turk (AMT) to analyze visual accuracy, and then the realistic qualities of images produced by a fully convolutional network (FCN) score where the higher number is optimum. For their map experiment, their results for AMT were 18.9 percent compared to the L1 model, which only achieved 0.8.

3.3 Transforming Satellite imagery into vector maps using modified GANS

Bashir et al. (2024) used modified GANS to produce a vector tile map from satellite images, as well as conventional neural networks (CNNs). Their architecture for translating the images was HierarchicalPix, utilising two frameworks on a local and global level. The generator they used was the Unet++, while the discriminator was PatchGan, slightly modifying their CNN blocks. This model resulted in a pixel-level accuracy of 61.04 percent and an SSIM (Similarity Index between -1 and 1) score of 0.75.

3.4 Satellite image to map conversion and land cover using deep learning

Gayathri et al.(2025) used conditional generative adversarial networks (cGAN), Pix2Pix, to produce a model allowing for satellite-to-map conversion. By providing dual functionality, including image-to-image conversion, as well as land classification. The generation (U-Net) is used to extract information from the input images, while the PatchGAN discriminator evaluates those patches. Resulting in the system processing a clear map image in under 10 seconds per image (on average) after training for 800 epochs.

3.5 CNN-based operational approach for land cover mapping in China.

Zhao et al.(2020) research provides a 4-part methodology split into classification, data sources, training sample selection, and training/ inference. Comparing their own deep learning model to a highly accepted GLC30 land cover map used in industry. The model was trained using 20,000 samples for 300 epochs. Resulting in an accuracy within the range of 74.25-78.87 percent compared to the GLC30 model.

4 Data processing

The Pix2Pix Kaggle dataset (2018,2020) is widely used in image-to-image translation networks with 4 sets of images. The “maps” folder in particular, contains 1096 training sets and 1098 validation sets as coloured images (in RBG). Each image is 1200 by 600 pixels, containing the satellite image (left side, 600 by 600) and the corresponding map (right side, 600 by 600). The training set contains different landscapes, including cities, rivers, the ocean, forests, and their combinations.

4.1 Data preprocessing and cleaning

The images will be cropped and labeled to separate the input and ground truth. Then, they will be converted into numpy arrays and stored in Python. The input will be resized further to fit in our model as well as reducing noise with a filtering technique .

4.2 Normalizing Data

During training, the data should be normalized after each filtering kernel to avoid bias towards specific features. Centering the image to zero would ensure the model learns important features rather than noise and prevents gradient vanishing to some extent.

4.3 Data Augmentation

To expand on the training dataset and avoid bias from the input, data augmentation is used to create more diversity. Rotation, filtering, cropping, and mirroring can be applied to the training set such that the model will be more robust and less likely affected by noises.

5 ARCHITECTURE

We will be using a Conditional GAN architecture based on the Pix2Pix neural network known for image-to-image translation (2018). Our model will comprise a U-Net Generator and PatchGAN Discriminator, both built on Convolutional Neural Networks (CNNs).

The U-net Generator is a CNN-based architecture that consists of an encoder and decoder with skip connections. It takes in an input (real) satellite image and generates an output map image. The PatchGAN Discriminator takes in two input pairs: one consisting of the real satellite image and its corresponding real map output, and the other consisting of the real satellite image and the generated map. The discriminator identifies if an image pair is real or fake using CNN layers.

During the training/validation stage, the model is tuned and trained based on the loss evaluations. The Discriminator's hyperparameters will be updated using the discriminator loss. Moreover, the Generator's hyperparameters will be updated using GAN loss to deceive discriminator and L1 loss(mean absolute error) to reduce the difference between real map output image and generated image. Figure 2 below showcases the proposed architecture of our model.

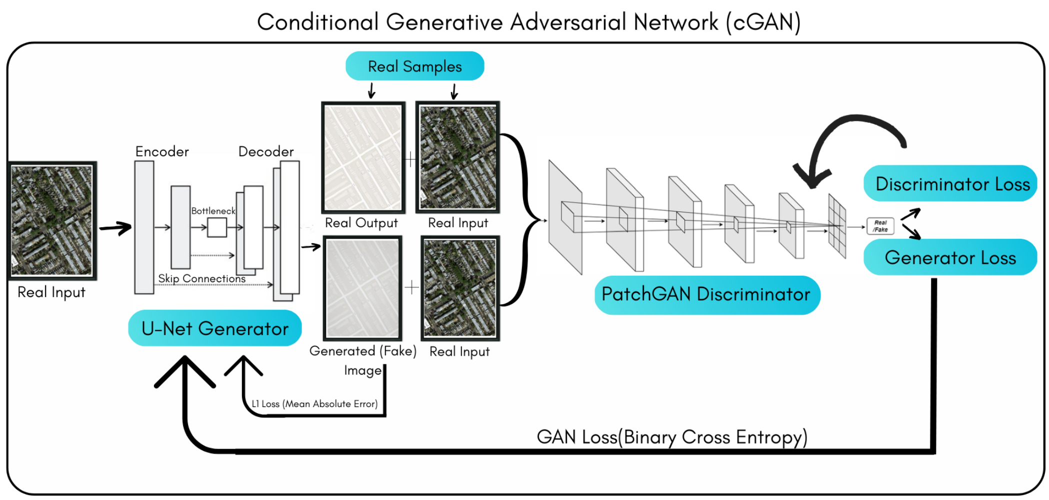

Figure 2: Proposed Architecture

6 BASELINE MODEL

The baseline model we will be using to compare our model to is the Canny edge detector[12], an algorithm developed by John F. Canny in 1986. In this operator, the input image is blurred to reduce noise then gradients in the image will be examined to find edges. This model will not be able to identify buildings, roads or rivers in the output. The result from this model will be a non-labeled map containing its main features.

7 ETHICAL CONSIDERATIONS

Considering that the model to translate satellite images into a Google Maps format requires the Maps Pix2Pix dataset, which is split into training and validation, there are some underlying ethical considerations to be taken into account. Primarily, the resolution of the images in the dataset is 1200x600, which is considerably low resolution, especially for the level of detail required in a city map. This could lead to issues of inaccuracy during inference, as the ground truth labels themselves are quite hard to read because of the low resolution.

Amongst the issues of low resolution is the sizing of the images, since all the data in the training and validation sets are 256x256, this means that any images processed during inference must be cropped and resized beforehand. In addition to the delay of preprocessing, there is also the potential issue of overfitting that can occur because the model is trained for a specific size and look of data.

The chosen data set, Maps, represents areas in New York City. This raises the concern of bias within the data, as while the model may be trained efficiently, different cities globally have different road and city structures. Certain neighborhoods and city layouts may be underrepresented due to this bias. Changes in city layouts and designs would also cause a rise for concern since the model would've been only trained on one city standard at a given time.

The three issues concluded above all lead to the ethical concern, which is the potential lack of accuracy. With an inaccurate map due to the underrepresentation of neighborhoods, it could lead to fatalities with a false positive when planning an infrastructure project. In addition, the issue of overfitting, as well as, low resolution only furthers the potential concern of low quality and inaccurate satellite to map translation. Considering the purpose of the model, alongside incorrect decision making, this could also lead to misdirection and problems in navigation, as the maps produced have the intended use of being wholly accurate sources of data.

8 Project Plan

This project plan outlines how team will collaboratively develop a Pix2Pix GAN to translate satellite images into map renderings. It details task assignments, internal deadlines, communication methods, and strategies for effective coordination. The goal is to ensure smooth progress, balanced workload, and clear accountability throughout the development process.

8.1 Team Coordination Plan

The team will hold weekly Saturday meetings at approximately 11:30 AM ET (Toronto Time) to accommodate the varying time zones of the team members. These meetings are to review progress, plan upcoming work, resolve any collaboration challenges, and assign tasks. The tasks are divided as follows in Figure 3, for the upcoming milestone, the progress report. Team members are free to divide their own subtasks.

2	**Progress Report**					
2.1	Preprocessing Data	Daisy	6/14/25	6/27/25	13	0%
2.2	Create script	Mohammed	6/14/25	6/27/25	13	0%
2.2	Generator Code	Muhammed Ahmed	6/27/25	7/4/25	7	0%
2.4	Discriminator Code	Salma	6/27/25	7/4/25	7	0%
2.5	Training and Validation	Mohammed	7/4/25	7/11/25	7	0%

Figure 3: Task Distribution

8.2 Tracking Progress

Progress will be tracked using a Gantt chart on a shared Google Sheets document, where team members are free to edit and manage their own subtasks and where major milestones are recorded. Using a Gantt chart allows the team to adapt dynamically to the varying schedules of the team members as well as ensuring accountability and transparency. An image of the task distribution for the proposal is shown in google sheets as displayed in Figure 4.

Milestone #	TASK TITLE	TASK OWNER	START DATE	DUE DATE	DURATION	% OF TASK COMPLETE
1	Project Proposal					
1.1	Introduction	Salma	6/8/25	6/10/25	2	100%
1.2	Illustration	Muhammed Ahmed	6/9/25	6/12/25	3	100%
1.3	Background and Related Work	Salma	6/9/25	6/12/25	3	100%
1.4	Data Preprocessing	Daisy	6/11/25	6/13/25	2	100%
1.5	Architecture	Muhammed Ahmed	6/10/25	6/11/25	1	100%
1.6	Baseline Model	Daisy	6/10/25	6/13/25	3	100%
1.7	Ethical Considerations	Salma	6/10/25	6/11/25	1	100%
1.8	Project Plan	Mohammed	6/11/25	6/13/25	2	100%
1.9	Risk Register	Mohammed	6/9/25	6/10/25	1	100%

Figure 4: Progress Tracking

8.3 Communication and Collaboration Tools

All team members will communicate through Discord for quick messaging and Google Meet for weekly meetings. Google Sheets will be used for task distribution and tracking the team's progress to ensure abiding by internal deadlines. For version control, we will use GitHub, where each member works on their local devices before pushing onto a remote repository.

Table 1: Communication Tool

Tool	**Purpose**
Google Sheets	Task tracking, timeline
Github	Code Collaboration, version control
Discord	Daily Communication
Google Meet	Weekly meetings

9 Risk Register

Table 2 describes the various risks associated with the project and the likelihood of these risks taking place given background research and possible contingency plans we can implement.

Table 2: Risk Management Plan

Risk	**Likelihood**	**Solutions**
The model takes too long to train	70% likelihood, given research on other related work that uses pix2pix GANs.	Utilize a summarized performance function to periodically plot our generated images while training, ensuring that the generated images are heading in the right direction, rather than waiting for them to converge after training is complete.
Bias towards specific city layouts and landscapes	20% likelihood, given the number of training data used and the fact that it contains diverse landscapes.	Utilize multiple datasets to train. Furthermore, augmenting the training dataset introduces more diversity.
Incorrect preprocessing	30% likelihood, image pairs have to be perfectly aligned pixel-wise, any misalignment caused by human error in preprocessing can degrade training performance	Check image pair dimensions and pixel alignment before training. Furthermore, using sanity checks will allow us to visually confirm proper matching and alignment.
Merge Conflicts	60% likelihood, some team members are inexperienced with git and may accidentally overwrite others' work.	Descriptive commit messages ensure clear communication and traceability. Frequent git pushes ensure that team members' work is regularly backed up and correctly integrated.

10 Link to Github

The project's GitHub repository link: `https://github.com/Muhammadahmed-1/APS360_Group12`

REFERENCES

Sunil Belde. Noise removal in images using deep learning models, 2021. URL https://medium.com/analytics-vidhya/noise-removal-in-images-using-deep-learning-models-3972544372d2.

Canva. Canva, 2025. URL https://www.canva.com/.

A. Cijov. Pix2pix maps, 2020. URL https://www.kaggle.com/datasets/alincijov/pix2pix-maps.

Geeks for Geeks. Implement canny edge detector in python using opencv, 2025. URL https://www.geeksforgeeks.org/machine-learning/implement-canny-edge-detector-in-python-using-opencv/.

R. Giri, B. R. Lamichhane, and B. Pokhrel. Sketch to image translation using generative adversarial network. *Journal of Engineering Sciences*, 2(1):70–75, 2023. URL https://nepjol.info/index.php/jes2/article/view/60397.

IARJSET. Satellite image to map conversion and land cover analysis using deep learning. *IARJSET International Advanced Research Journal in Science*, 12(5), 2025. doi: 10.17148/IARJSET.2025.125343. URL https://iarjset.com/papers/satellite-image-to-map-conversion-and-land-cover-analysis-using-deep-learning/.

P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *Berkeley AI Research (BAIR) Laboratory, UC Berkeley*, 2018. URL https://arxiv.org/pdf/1611.07004.

P. Kashyap. Image normalization in pytorch: From tensor conversion to scaling, 2024. URL https://medium.com/@piyushkashyap045/image-normalization-in-pytorch-from-tensor-conversion-to-scaling-3951b6337bc8.

A. Taparia, A. K. Bashir, Y. Zhu, T. R. Gdekallu, and K. Nath. Transforming satellite imagery into vector maps using modified gans. *Alexandria Engineering Journal*, 109:792–806, 2024. URL https://www.sciencedirect.com/science/article/pii/S1110016824010986?via%3Dihub.

V. Tiwari. Pix2pix dataset, 2018a. URL https://www.kaggle.com/datasets/vikramtiwari/pix2pix-dataset/data.

V. Tiwari. Pix2pix dataset. 2018b. URL https://www.kaggle.com/datasets/vikramtiwari/pix2pix-dataset/data.

X. Zhao, L. Gao, Z. Chen, and B. Zhang. An cnn-based operational approach for land cover mapping in china. In *IOP Conference Series: Earth and Environmental Science*, volume 502, 2020. doi: 10.1088/1755-1315/502/1/012036. URL https://iopscience.iop.org/article/10.1088/1755-1315/502/1/012036.