Generating maps using satellite images

Author: Jan Koten

CTU FIT

Project Repository: `https://github.com/kotenjan/map`

29. 11. 2023

Contents

Chapter 1

Introduction to Image-to-Image Translation with GANs

1.1 Overview of Generative Adversarial Networks

Generative Adversarial Networks (GANs) have revolutionized the field of machine learning by introducing a novel approach to generative modeling. The core concept of GANs involves two neural networks, a generator and a discriminator, that are trained simultaneously through adversarial processes. This section will delve into the foundational principles of GANs, their evolution, and how they have become a cornerstone in the field of image generation and translation.

1.1.1 Foundation and Development

Introduced by Ian Goodfellow and his colleagues in 2014, GANs have undergone significant advancements. From their initial conception to current sophisticated implementations, GANs have shown remarkable capabilities in various domains, particularly in generating realistic images.

1.1.2 Mechanics of GANs

The mechanics of GANs involve a game-theoretic approach where the generator aims to create data indistinguishable from real data, while the discriminator strives to differentiate between real and generated data. This adversarial competition drives both networks to improve continuously, leading to the generation of highly realistic data.

1.2 Image-to-Image Translation with GANs

Image-to-image translation is a specific application of GANs where the goal is to learn a mapping from an input image to an output image. This process has a wide range of applications, from style transfer to enhancing image resolution, and more pertinently, translating satellite imagery to map representations.

1.2.1 Concept and Applications

In image-to-image translation, GANs are trained to convert images from one domain to another while preserving certain characteristics of the input images. Applications include transforming photographs to artistic styles, converting day images to night, and in this thesis, translating satellite images to map views.

Chapter 2

Understanding DualGAN Architecture

2.1 Overview of DualGANs

Dual Generative Adversarial Networks (DualGANs) represent a novel approach in the field of image-to-image translation. They are particularly effective for tasks where paired images from two different domains are involved. In the context of translating satellite images to maps images, DualGANs offer a robust framework for handling such complex transformations.

2.1.1 The Dual Nature of DualGANs

DualGANs operate on a principle of dual learning mechanisms. This involves two GANs being trained simultaneously: one for the forward translation (e.g., satellite to map) and another for the backward translation (e.g., map to satellite). This dual training mechanism ensures a higher level of accuracy and consistency in the translation process.

2.2 DualGAN Architecture Components

The architecture of DualGANs is crucial for understanding their functionality and effectiveness in image-to-image translation tasks.

2.2.1 Generators in DualGANs

In DualGANs, each generator is responsible for translating images from one domain to another. For instance, one generator translates satellite images into maps, while the other does the reverse.

2.2.2 Discriminators in DualGANs

The discriminators in a DualGAN setup play the role of differentiating between the generated (fake) images and real images from each domain. Their primary function is to guide the generators towards producing more realistic and convincing translations.

2.3 Training and Loss Functions

The training process of DualGANs is intricate, involving alternating updates to both GANs using adversarial and cycle consistency losses, ensuring both realistic translation and retention of original image features. Adversarial loss aims to optimize realism and cycle consistency loss is aimed at maintaining spatial accuracy in translations.

Figure 2.1 showcases a DualGAN's translation from a satellite image to a map representation, illustrating the model's capability to maintain accuracy and detail.

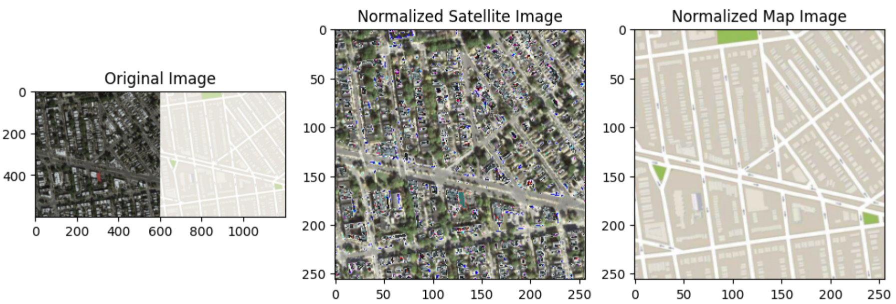

Figure 2.1: Example of an image-to-image translation from a satellite image to a map representation.

2.3.1 Adversarial and Cycle Consistency Loss

Adversarial loss ensures that the generated images are indistinguishable from real images, while cycle consistency loss ensures that an image translated from one domain to the other and back retains its original features.

Chapter 3

Data Preprocessing

3.1 Dataset Preparation

`SatelliteToMapDataset`, a pivotal class in our project, prepares the dataset for training the DualGAN model. It processes image pairs, each containing a satellite image and its corresponding map image, essential for the model's learning process.

3.1.1 Image Pair Extraction and Augmentation

Each image file in the dataset is split into two: the left half being the satellite image, and the right half the map image. The class also applies a series of augmentations such as flips, rotations, and color jittering. These augmentations are crucial for enhancing the model's robustness and preventing overfitting by introducing diverse training data.

3.1.2 Normalization and Training Visualization

The images undergo normalization to standardize input values, typically to the range [-1, 1], for efficient training. Figure 3.1 showcases the training progress, illustrating the transformations done to the dataset.

This chapter outlines the `SatelliteToMapDataset` class's role in preparing the dataset, emphasizing its contribution to the DualGAN model's successful training in image-to-image translation.

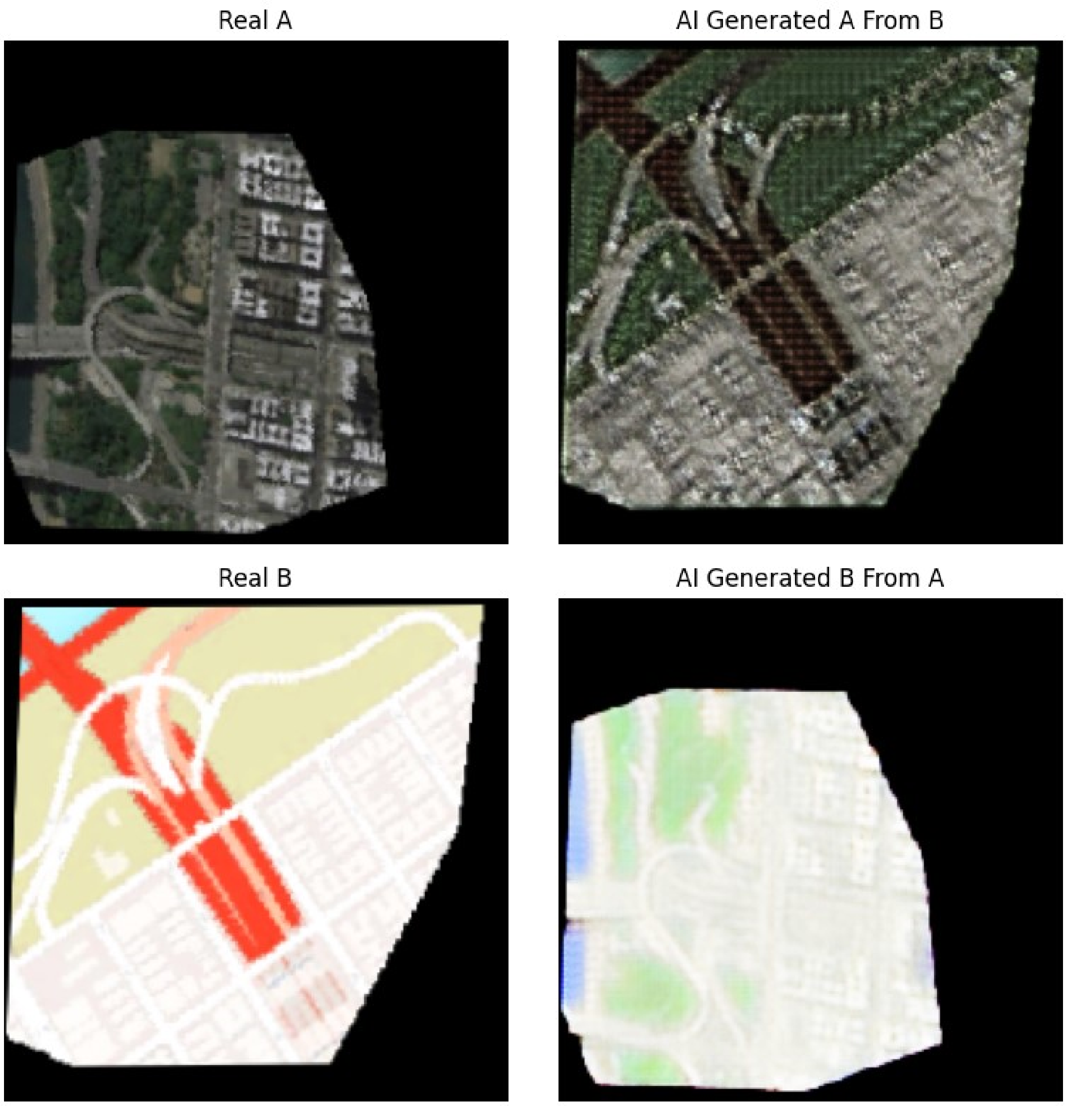

Figure 3.1: Visualization of satellite images and their map translations during training.

Chapter 4

Advanced Techniques and Optimization

This chapter delves into the advanced techniques and optimization strategies implemented in the DualGAN model for the satellite image to Google Map translation. It explores the nuances of enhancing the model's capabilities and the fine-tuning processes that significantly contribute to the accuracy and efficiency of the image-to-image translation.

4.1 Enhancements in Generator and Discriminator Architecture

4.1.1 Generator Architecture Improvements

The generator's architecture was enhanced by increasing its depth, allowing it to capture more complex features. Attention mechanisms were incorporated to direct the network's focus on salient features of satellite images. Advanced normalization techniques, like BatchNorm and GroupNorm, were used for stabilizing the training. Additionally, skip connections were implemented to preserve spatial information crucial for accurate image translation.

4.1.2 Discriminator Architecture Optimization

For the discriminator, the implementation of a PatchGAN architecture proved effective. This architecture classifies patches of an image as real or fake, enabling a more detailed assessment of image authenticity. Spectral normalization was used to stabilize the training by normalizing the weights of the layers. Different convolution types, including dilated and depthwise separable convolutions, were experimented with to enhance efficiency and effectiveness.

4.2 Training Process and Data Augmentation

4.2.1 Hyperparameter Tuning and Loss Function Adjustments

Hyperparameter tuning was a critical aspect of the training, involving adjustments to learning rates, batch sizes, and loss function weights. Experimenting with various loss functions, including perceptual loss, was key to enhancing model output.

4.3 Progressive Growing and Model Optimization

The training employed progressive growing, starting with low-resolution images and gradually moving to higher resolutions. This approach stabilized training and aided in learning complex image hierarchies.

4.4 Training Evolution and Results

The model's evolution was monitored, with initial predictions (Figure 4.1) being less refined and later stages (Figure 4.2) showing marked improvements in translation accuracy.

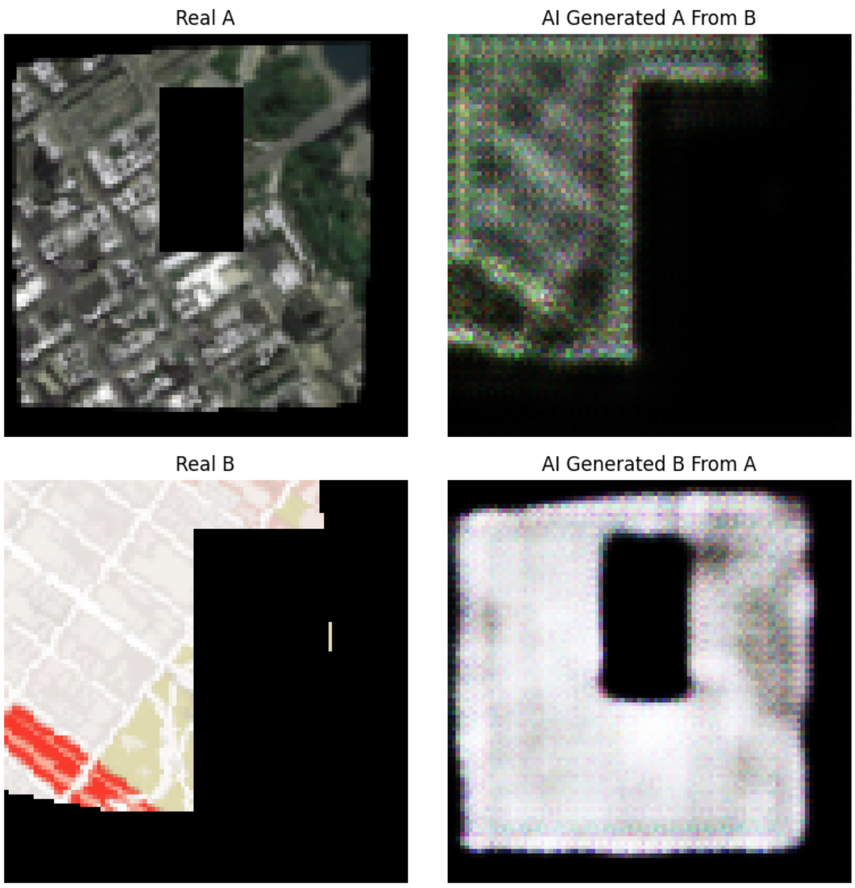

Figure 4.1: Early stage training results showing input and predictions

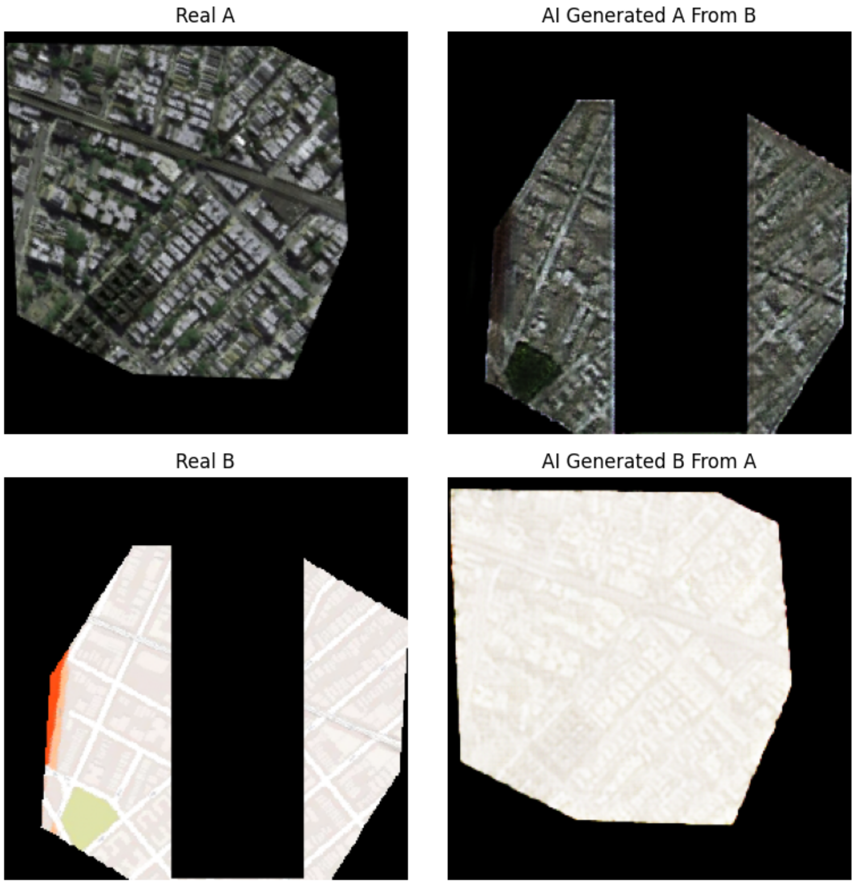

Figure 4.2: Later stage training results with refined input and predictions

4.5 Conclusion

Integrating advanced techniques and optimization strategies significantly enhanced the DualGAN model's quality and accuracy in translating satellite to map images. Fine-t

Chapter 5

Evaluation, Testing, and Results

5.1 Evaluation Methodology

This chapter outlines the evaluation methods for the image-to-image translation model, assessing map image quality and the role of combined loss functions in model performance with a Structural Similarity Index (SSIM).

5.2 SSIM in Model Evaluation

5.2.1 SSIM Implementation

SSIM was employed for validation data evaluation, focusing on texture, contrast, and luminance changes to measure translation accuracy against ground truth maps.

5.2.2 Loss Function Analysis

The model's effectiveness was further analyzed through a combination of adversarial and cycle consistency losses, crucial for realistic and coherent map generation.

5.3 Results and Performance Analysis

5.3.1 Accuracy Metrics

Training and validation accuracy, documented over 90 epochs, are illustrated in Figures 5.1 and 5.2. These showcase the model's progressive improvement, reflected in rising SSIM scores and decreasing loss values.

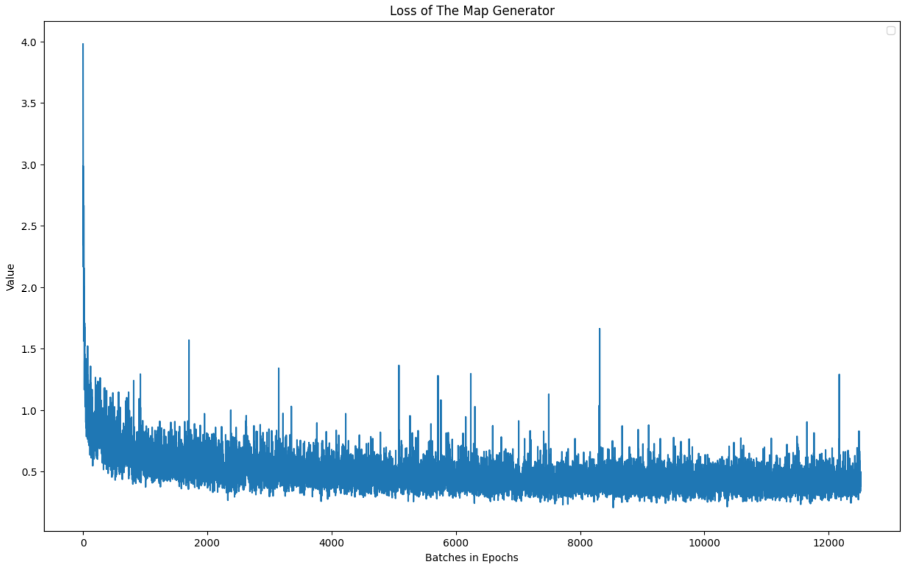

Figure 5.1: Evolution of Training Accuracy Across 90 Epochs

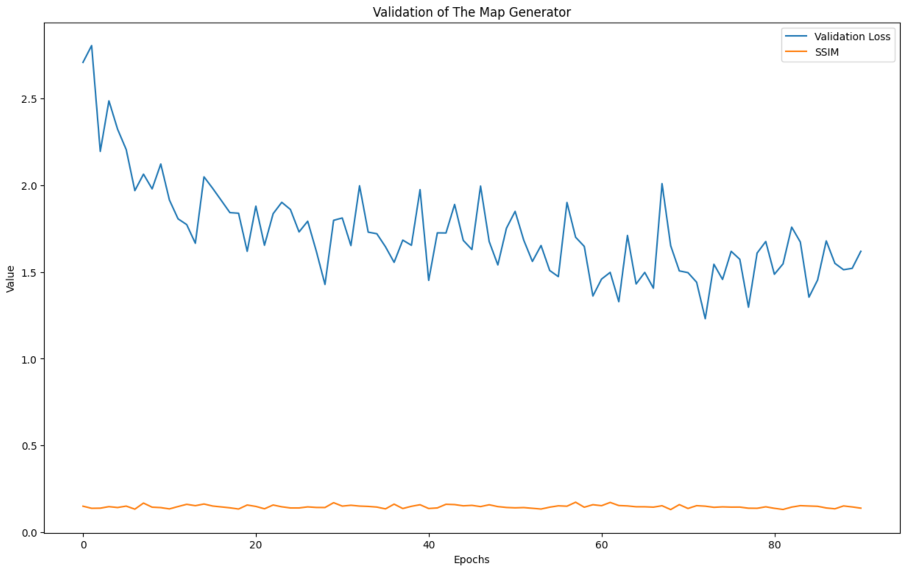

Figure 5.2: Evolution of Validation Accuracy Across 90 Epochs

Chapter 6

Conclusion

6.1 Challenges with Global Generator

The initial approach using a global generator for map creation revealed a significant limitation: the colors of different elements in the map (such as houses and roads) were not distinctly differentiated. This led to a lack of clarity and detail in the generated maps, which is crucial for realistic and usable map generation.

6.2 Proposed Method for Element-Specific Translation

To overcome this limitation, a novel approach is proposed: splitting the map into separate binary maps, where each map represents a specific element (e.g., houses, roads) with the element positions marked as 1 and the rest as 0. These binary maps can then serve as inputs to a model designed to construct the final, detailed map. This method is expected to enhance the differentiation and clarity of various map elements, leading to more accurate and visually distinct translations.

6.3 Result

The combination of SSIM evaluation and the analysis of loss functions provides a comprehensive view of the model's capabilities and areas for enhancement. The proposed approach of element-specific translation is expected to further refine the model's performance, particularly in terms of color differentiation and detail representation in the generated maps.