Generating maps using satellite images

Author: Jan Koten

CTU FIT

Project Repository:

`https://gitlab.fit.cvut.cz/kotenjan/ni_mvi_gan`

22. 12. 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

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.

Figure 2.1 shows training data for DualGAN's translation from a satellite image to a map representation. Its task is to maintain accuracy and detail.
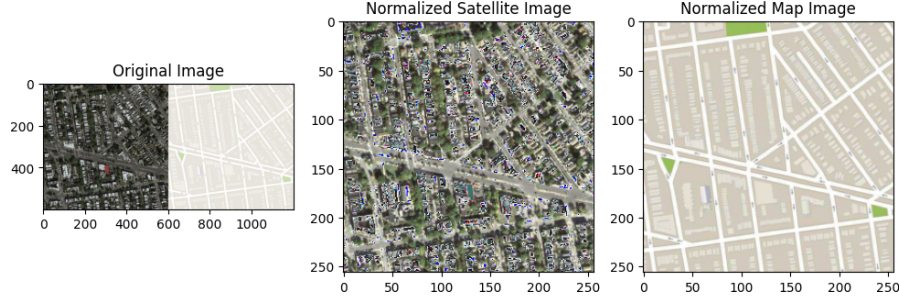

Figure 2.1: Training data for an image-to-image translation from a satellite image to a map representation.

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

Mask DualGAN Approach

This chapter presents the detailed methodology of the dual-generation, yet another approach for map generation from satellite images. This novel method involves two key stages: the generation of object-specific masks and the subsequent map creation from these masks.

3.1 Theoretical Background

The dual-generation approach hypothesizes that dividing map generation into two phases—first, identifying and masking individual objects in satellite images, and then, using these masks for map creation—improves the final map's accuracy and detail.

3.2 Algorithm and Implementation

3.2.1 Mask Generation

The mask generation process involves identifying the most frequent colors in the satellite images, representing different objects such as buildings, roads, and vegetation. These colors are determined using a statistical analysis of the entire dataset. The process is as follows:

1. Analyze the dataset to find the n most frequent distinct colors.

2. For each satellite image, map the pixels to the nearest of these n colors using the nearest neighbors algorithm.

3. Generate a mask for each object type, where each color channel corresponds to a mask of a different set of objects.

3.2.2 Map Generation from Masks

In the second phase, the generated masks are used to create the final map. This involves a separate generative model that interprets the masks and reconstructs a detailed map, preserving the spatial relationships and characteristics of each object.

Chapter 4

Data Preprocessing

4.1 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.

4.1.1 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 satellite images also undergo several preprocessing steps to optimize them for the generation process for either of the two model architectures. These steps include:

- Downscaling the image in early training epochs to stabilize the model.

- Normalization of the color space of the images.

- Applying filters, such as rotation, flips and crops to enhance specific features relevant for object identification and prevent overfitting.

- The map image is also split into individual object masks in case of the Mask Generator architecture

4.1.2 Training Visualization

After the data undergoes all of the preprocessing steps, the training phase is initiated. Figures 4.1 and 4.2 showcase 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 4.1: Visualization of satellite images and their map translations during training of the standard DualGAN architecture.
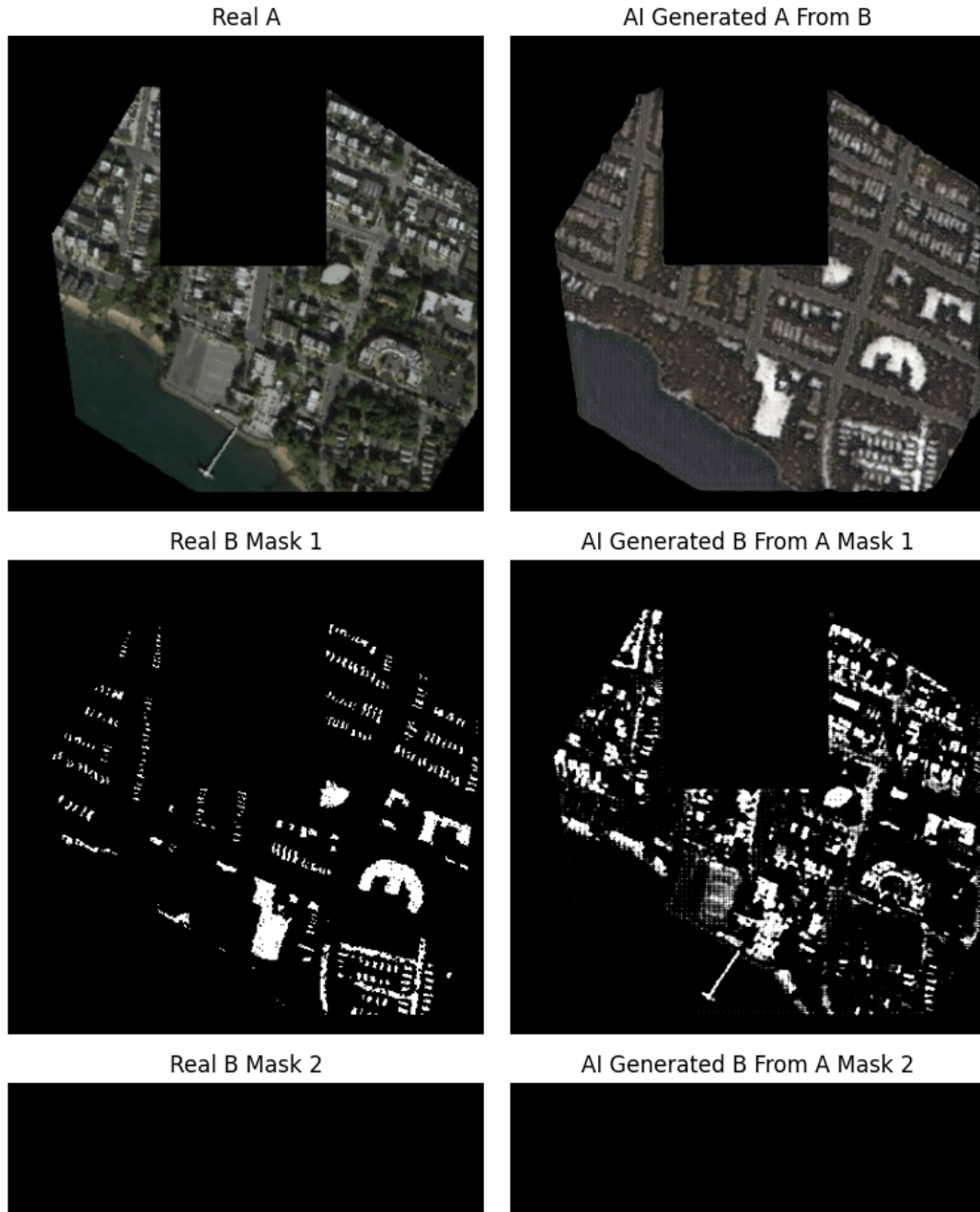
Figure 4.2: Visualization of satellite images and their mask translations during training of the Mask Generating architecture.

Chapter 5

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.

5.1 Enhancements in Generator and Discriminator Architecture

5.1.1 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.

5.1.2 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.

5.2   Training Process and Data Augmentation

5.2.1   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.

5.3   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.

5.4   Training Evolution and Results

The model's evolution was monitored, with initial predictions (Figure 5.1) being less refined and later stages (Figure 5.2) showing marked improvements in translation accuracy.
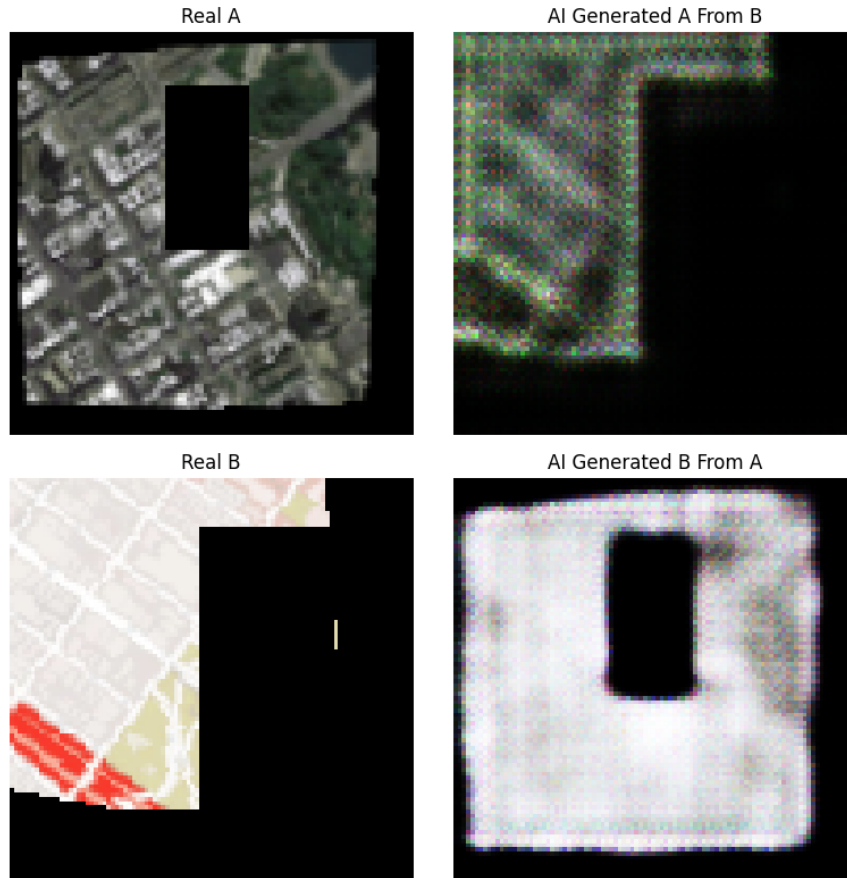

Figure 5.1: Early stage training results showing input and predictions

Figure 5.2: Later stage training results with refined input and predictions

5.5 Conclusion

Integrating advanced techniques and optimization strategies significantly enhanced the DualGAN model's quality and accuracy in translating satellite to map images.

Chapter 6

Evaluation, Testing, and Results

6.1 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). The generation with Masks was abandoned early on as it did not yield any major improvements over the standard DualGAN approach, which it's not being discussed.

6.2 SSIM in Model Evaluation

6.2.1 SSIM Implementation

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

6.2.2 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.

6.3 Results and Performance Analysis

6.3.1 Accuracy Metrics

Training and validation accuracy, documented over 300 epochs, are illustrated in Figures 6.1 and 6.2. These showcase the model's progressive improvement, reflected in rising SSIM scores and decreasing loss values.
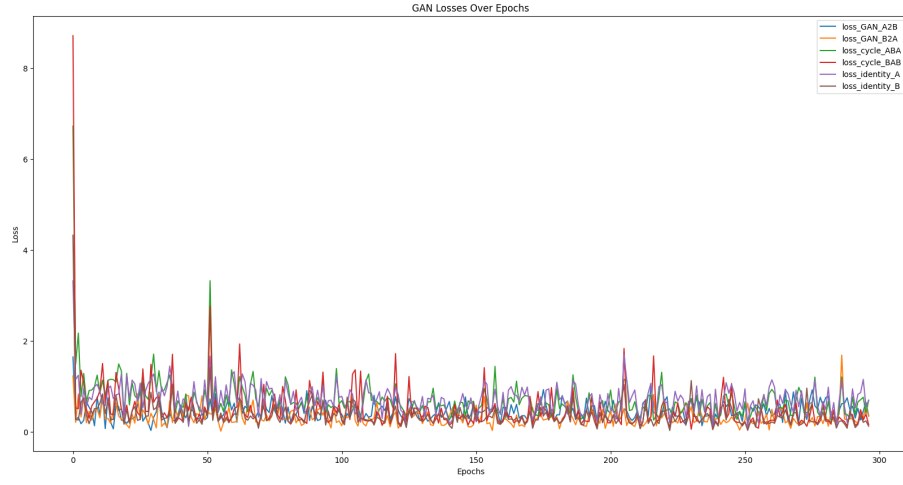
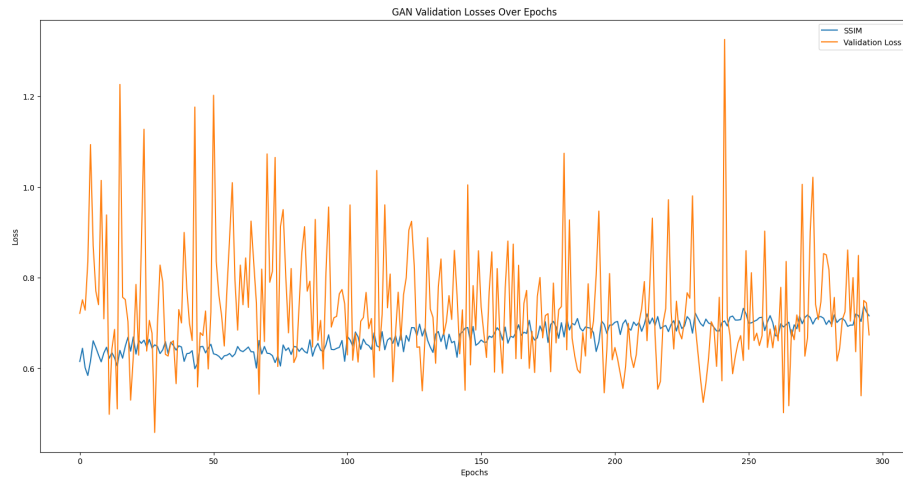Figure 6.1: Evolution of Training Accuracy Across 300 Epochs

Figure 6.2: Evolution of Validation Accuracy Across 300 Epochs

Chapter 7

Conclusion

7.1   Challenges with Global Generator

The initial approach using a global generator for map creation yielded results with impressive accuracy, however the colors of unique elements with occasional occurrence were not distinctly differentiated. This led to an occasional lack of clarity and detail in a negligible amount of generated maps. We assume that this problem could be solved simply by having more data or introducing more layers to the generators, which would increase the training complexity above our current hardware capabilities.

7.2   Proposed Method for Element-Specific Translation

To overcome this limitation, a novel approach was 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 then served as inputs to a model designed to construct the final, detailed map. This method enhanced the differentiation and clarity of individual map elements at an expense of the general quality of the final map.

7.3   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 generator enhancement is expected to further refine the model's performance, particularly in terms of color differentiation and detail representation in the generated maps.

References

1. **Deep Learning for Understanding Satellite Imagery: An Experimental Survey**
   URL: `https://www.frontiersin.org/articles/10.3389/frai.2020.534696/full`
   Overview: Examines automated translation of satellite imagery into maps using deep learning, highlighting enhanced U-Net and Mask R-CNN models with custom training methods for building position identification in satellite images.

2. **Satellite Image Deep Learning Techniques**
   GitHub Repository: `https://github.com/satellite-image-deep-learning/techniques`
   Overview: A repository documenting deep learning techniques for satellite imagery analysis, covering object detection, segmentation, and classification.

3. **Satellite Imagery Segmentation with Convolutional Neural Networks**
   URL: `https://joshting.medium.com/satellite-imagery-segmentation-with-convolutional-neural-networks-f9254de3b907`
   Overview: Describes a project using CNNs for land cover classification in satellite images, implementing U-Net and ResNet50_U-Net architectures with notable validation dice coefficients.

4. **DeepSat V2: Feature Augmented CNNs for Satellite Image Classification**
   URL: `https://cmp.felk.cvut.cz/~reinsmic/F3-DP-2017-Kunc-Vladimir-Deep_neural_network_for_satellite_image_classification_using_OpenStreetMap.pdf`
   Overview: Presents an end-to-end framework augmenting CNNs with hand-crafted features for satellite image classification, achieving high accuracies on SAT-4 and SAT-6 datasets.

5. **DualGAN: Unsupervised Dual Learning for Image-to-Image Translation**
   URL: `https://arxiv.org/abs/1704.02510`
   Overview: Introduces DualGAN for unsupervised image-to-image translation using two GANs, enabling effective translation across various image types without labeled data.

6. **DualGAN Study for Image-to-Image Translation**
   URL: `https://openaccess.thecvf.com/content_ICCV_2017/papers/Yi_DualGAN_Unsupervised_Dual_ICCV_2017_paper.pdf`
   Overview: Explores DualGAN for unsupervised image translation using two

GANs in dual learning, demonstrating its effectiveness in photo enhancement, style transfer, and object transfiguration.

7. **DualGAN Project**
   GitHub Repository: `https://github.com/duxingren14/DualGAN`
   Overview: Implements DualGAN for image-to-image translation, showing high-quality translations and capabilities in photo enhancement and style transfer.

8. **How CycleGAN Works**
   URL: `https://developers.arcgis.com/python/guide/how-cyclegan-works/`
   Overview: Details CycleGAN's approach to image-to-image translation using unpaired datasets, extending Pix2Pix architecture for applications like converting RGB imagery and generating map routes from satellite imagery.