Project Report: Satellite Image to Map Route Conversion Using U-Net and StyleGAN in a Pix2Pix Framework

1. Introduction

The project aimed to develop an advanced image-to-image translation model that could effectively convert satellite images into detailed map routes. This task falls under the broader domain of image translation, where the objective is to transform images from one domain (satellite images) to another (map routes). The chosen methodology involved leveraging U-Net architecture, enhanced by a PatchGAN discriminator, within a pix2pix framework. Additionally, StyleGAN techniques were employed to refine the generated images and improve the model's overall performance.

2. Background and Motivation

Satellite imagery offers a wealth of information, but its raw form is often difficult to interpret directly for specific applications such as navigation. By translating satellite images into map routes, the data becomes more accessible and practical for tasks like urban planning, navigation, and geographic information systems (GIS). The pix2pix framework is particularly well-suited for this translation task due to its ability to learn a mapping from input images to output images using paired datasets.

3. Methodology

3.1 U-Net Architecture for Image Translation

The U-Net architecture was chosen for its strong performance in segmentation tasks, which are analogous to the task of converting satellite images to maps. U-Net is a type of convolutional neural network (CNN) that is particularly effective for image translation tasks due to its encoder-decoder structure with skip connections. The encoder compresses the input image into a lower-dimensional representation, while the decoder reconstructs the image in the target domain, in this case, map routes.

- Encoder: The encoder consists of a series of convolutional layers with down-sampling, which captures the context and features from the satellite images.

- Decoder: The decoder up-samples these features to generate the output map routes.
- Skip Connections: To preserve spatial information, skip connections are used between corresponding layers in the encoder and decoder, which help in maintaining the finer details of the input image in the output.

3.2 PatchGAN Discriminator

To enhance the quality of the generated images, a PatchGAN discriminator was employed. Unlike traditional discriminators that classify the entire image, PatchGAN classifies patches of the image, which allows it to focus on local structures and texture details. This approach is beneficial for ensuring that the generated map routes are not only structurally accurate but also visually coherent.

- Local Focus: PatchGAN assesses the realism of small patches (e.g., 70x70 pixels) within the image, making it particularly effective at ensuring texture consistency.
- Global Coherence: By applying this local patch evaluation across the entire image, PatchGAN helps the model maintain global coherence, ensuring that the generated routes align well with the input satellite image.

3.3 Implementation of StyleGAN Techniques

In addition to the core pix2pix framework, several advanced techniques from StyleGAN were integrated to further refine the image generation process:

- Adaptive Instance Normalization (AdaIN: AdaIN was used to adjust the mean and variance of the style (map routes) based on the content (satellite images), allowing for more flexible and diverse outputs.
- Random Noise Injection: Random noise was introduced into the model during training to increase the variability of the generated images, making the model more robust to different types of input satellite images.
- Progressive Learning: The model was trained progressively, starting with lower-resolution images and gradually increasing the resolution. This technique helps in stabilizing the training process and improves the quality of the generated high-resolution images.
- Truncation Trick: A truncation trick was used during the generation phase to control the trade-off between image diversity and quality. By limiting the latent space sampling, the model was able to produce more coherent and higher-quality map routes.

4. Dataset and Preprocessing

The dataset consisted of paired satellite images and their corresponding map routes. Data preprocessing was crucial to ensure that the images fed into the model were of consistent quality and size. Steps included:

- Image Resizing: All images were resized to a standard size suitable for the U-Net architecture.
- Normalization: Pixel values were normalized to a range of [0, 1] to facilitate model training.
- Data Augmentation: To improve model robustness, data augmentation techniques such as rotation, flipping, and scaling were applied to the training images.

5. Training Process

The training process involved two primary steps: training the generator (U-Net) and the discriminator (PatchGAN) in an adversarial manner. The generator aimed to create realistic map routes from satellite images, while the discriminator attempted to distinguish between real and generated routes.
Used Google Colab GPU for training the model.

- **Loss Function**: A combination of adversarial loss and L1 loss was used. The adversarial loss encouraged the generator to produce realistic images, while the L1 loss ensured that the generated images were close to the ground truth in pixel space.
- **Training Stability**: Progressive learning and regularization techniques were employed to prevent mode collapse and ensure stable training.

6. Results and Evaluation

The model's performance was evaluated based on several metrics, including structural similarity index (SSIM), peak signal-to-noise ratio (PSNR), and visual inspection. The U-Net with PatchGAN and StyleGAN enhancements outperformed baseline models, producing high-quality map routes with accurate details.

- **Quantitative Metrics**: The model achieved a high SSIM score, indicating that the generated map routes closely resembled the ground truth.
- **Qualitative Analysis**: The map routes generated by the model were visually coherent and aligned well with the input satellite images, with clear roads and boundaries.

![Sample
Results](https://media.geeksforgeeks.org/wp-content/uploads/20200522231055/pix2pix-min.png
)

7. Conclusion and Future Work

The project successfully demonstrated the effectiveness of combining U-Net architecture with a PatchGAN discriminator in a pix2pix framework for converting satellite images into map routes. The integration of StyleGAN techniques further enhanced the quality and diversity of the generated images. Future work could involve exploring different GAN architectures, incorporating more advanced data augmentation techniques, and extending the model to other image translation tasks.

8. References

- "Image to Image Translation using Pix2Pix," GeeksforGeeks. Available: https://www.geeksforgeeks.org/image-to-image-translation-using-pix2pix/

9. GitHub Repo Link

MohakVyas/Pix2Pix-Image-translation (github.com)