Project 2 : Road Segmentation

Aurélien Ferlay

December 21, 2023

Introduction

This project focuses on the challenging task of extracting roads from satellite images, a key component in modern mapping and navigation systems. The dataset for this project consists of high-resolution satellite/aerial images sourced from GoogleMaps. Accompanying these images are corresponding ground-truth images where each pixel is meticulously labeled as either ‘road‘ or ‘background‘.

The primary objective of this project is to develop and train a machine learning model capable of accurately segmenting roads in these satellite images. In essence, the model is expected to classify each pixel in an image as ‘road=1‘ or ‘background=0‘. Achieving high precision in this segmentation task is crucial, as it directly impacts the effectiveness and reliability of a variety of applications, ranging from autonomous vehicle navigation to urban planning and infrastructure development.

The project leverages deep learning techniques, specifically a U-Net architecture with a ResNet50 encoder, to handle the segmentation task. This choice is motivated by the U-Net's proven efficacy in medical image segmentation, which shares similarities with our task in terms of the need for high precision and the ability to handle complex spatial structures. The project includes phases of data preparation, model training, and post-processing with morphological filters to refine the segmentation results.

Cropping strategy

Cropping approach: The initial dataset consists of 100 satellite images, each measuring 400x400 pixels. To enhance the dataset, we applied a strategic cropping technique. This method involves dividing each image into smaller sections, or ”crops,” which are then used as separate training samples. The primary objectives of this approach are:

- **Preserving Image Quality:** By maintaining high-resolution crops, we ensure that important details in the images are not lost.
- **Increasing Dataset Size:** Cropping allows us to significantly expand the number of training samples from the limited original dataset.
- **Improving Computational Efficiency:** Smaller image sizes reduce the computational burden during training.

Calculation of Cropped Images: The number of cropped images generated from each original image is determined by the formula:

$$\text{Number of Crops} = \left(\frac{400 - \text{crop\_size}}{n} + 1\right)^2$$

where crop_size is the dimension of the cropped image, and n is the step size. This calculation takes into account the overlap between adjacent crops. With a step size of 144 pixels, the formula provides a balance between thorough coverage of the original image and minimizing data redundancy.

(a) Original Image (400x400)

(b) Cropped Images (256x256 each)

Figure 1: Illustration of the cropping process applied to images and masks.

The dataset is divided into training (70%), validation (20%), and testing (10%) subsets before cropping. This division ensures a diverse range of data for training and evaluation purposes. Cropping is then applied consistently across all subsets.

Cropping Outcome: By applying our strategic cropping technique, we obtained 400 images and their associated ground truths from the original dataset. This substantial increase in data quantity provides a more diverse and comprehensive set for training our model.

Data Augmentation

Applied Transformations: The following augmentation techniques were employed:

- **Rotational Augmentation:** Images were rotated by various angles to simulate different orientations. This helps the model learn to recognize roads regardless of their orientation in the image.

- **Flipping:** Both horizontal and vertical flips were applied to the images, creating mirror images. This aids in training the model to understand symmetrical patterns and structures.

- **Color Jitter:** Adjustments in brightness, contrast, saturation, and hue were made to mimic varying lighting conditions and color variations that can occur in real-world scenarios.

- **Random Cropping:** In addition to the fixed cropping technique, random cropping was employed to introduce more variability in the positioning of roads within the image frame.

Model Architecture and Training

The segmentation task is approached using a U-Net model with a ResNet50 encoder. The model benefits from the encoder's pre-training on ImageNet, gaining robust feature extraction capabilities. The training leverages PyTorch Lightning for an efficient workflow, employing a binary cross-entropy loss function and an Adam optimizer. The learning rate and other hyperparameters are carefully tuned for optimal performance.

Figure 2: Training and validation loss over epochs.

Post-processing Techniques

The raw output masks from the model are subjected to post-processing for refinement. This includes:

- Binarization based on an empirically determined threshold.

- Application of morphological operations (such as closure) to enhance road structures and reduce noise

Model Evaluation

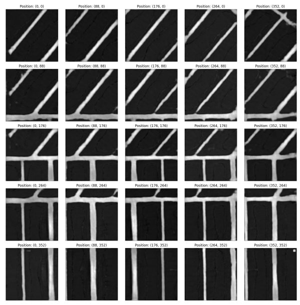
Given the class imbalance observed in the dataset, accuracy alone is not a sufficient metric. Therefore, the model's performance is evaluated using F1-score.

Visualization of Results

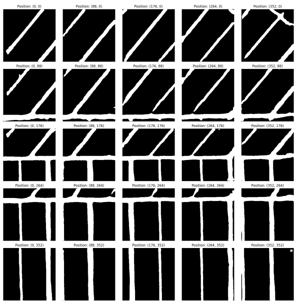
Below are the stages of the process from the original image to the final predicted mask.

(a) Original Image

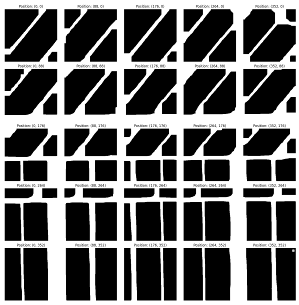
(b) Cropped Image



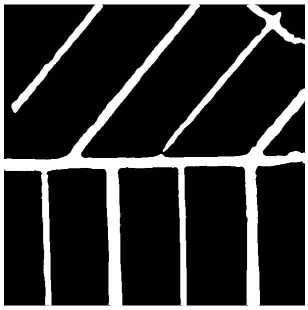
(c) 5x5 Predictions



(d) Post-Binarization



(e) Post-Filtering



(f) Post Reassembling

Figure 3: Stages of image processing and mask prediction.

Conclusion and Future Work

Our project successfully applied a U-Net model with a ResNet-50 encoder for road segmentation in satellite images, achieving an F1-score of 0.874 on AIcrowd (pseudo: aferlay). The cropping technique used for data augmentation proved effective, but future work could explore other architectures and encoders, more diverse datasets, and advanced augmentation methods. Further improvements might also include testing the model in various geographical locations and refining post-processing techniques.

Ethical Risks

The primary ethical concern of our project is the potential misuse of technology in surveillance, raising privacy issues. To mitigate this, we adhered to data privacy guidelines and ensured our dataset did not contain sensitive information. Future developments should continue to emphasize ethical considerations such as privacy, fairness, and transparency.

References

[1] *Segmentation Models PyTorch Documentation*, `https://segmentation-modelspytorch.readthedocs.io/en/latest/`, Accessed 2023.