CS-433 Road segmentation

Vincent Roduit
EPFL
Electrical Engineering
Sciper: 325140

Yannis Laaroussi
EPFL
Computer Science
Sciper: 369854

Fabio Palmisano
EPFL
Electrical Engineering
Sciper: 296708

***Abstract*—Road segmentation is a prevalent area within computer vision. The primary objective of this project is to discern roads within Google Maps satellite images by distinguishing them from the surrounding environment. Various models and data processing techniques are employed to explore and identify optimal solutions for achieving accurate road segmentation results.**

I. Introduction

The purpose of this work is to create a model capable of recognizing roads from Google Maps satellite images. This is a common problem in computer vision and image recognition. The main difficulty of this problem lies in the fact that the model has only the satellite view of the roads, and these can sometimes be obstructed by trees or other objects. Furthermore, roads often have the same color as buildings and parking areas, making the task very complex to solve. This paper will present the issues encountered during the process and propose solutions to address those problems.

II. Data Analysis

Before delving into building models, it is essential to examine the data to understand the problem. The training dataset consists of a hundred satellite images, each being an RGB image with dimensions of 400×400 pixels. Along with these satellite images, there is a set of ground truth images of the same size but in black and white. These images serve as labels, where a white pixel indicates a part of a road, and a black pixel is for the background. In order to predict the roads, the image is segmented into patches of size 16×16 pixels. A patch is predicted as a road if the mean of the ground truth patch is above a certain threshold, which is set to 0.25 (meaning that at least 25% of the patch is a road). The representation of one such image is depicted in Figure 1. This example illustrates the main challenges. Upon closer inspection around the coordinates (125, 325), it can be observed that some trees cover the road. However, this segment should be predicted as a road when considering its corresponding ground truth image.

In addition to the challenges discussed in Section I, distinguishing roads from areas such as parking lots or sidewalks can be challenging, even for humans. Another issue with the dataset is the prevalence of vertically or horizontally oriented roads, limiting its generalizability. Addressing these challenges will be a focus of upcoming sections.

Regarding the test set, it is composed of fifty images of size 608×608 pixels. The final prediction has to be done on patches of size 16×16 next to each other.

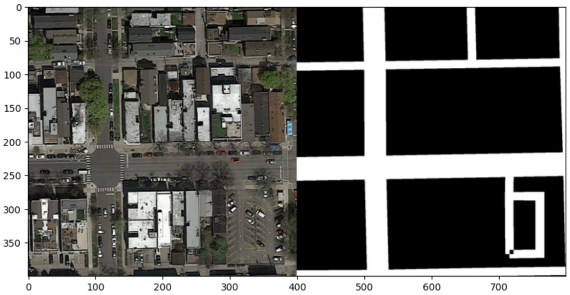


Fig. 1. Example of satellite image (satImage_035.png) with its ground truth

III. Data preprocessing

An important part of the project, which is a main factor to improve the results is the data preprocessing. Two approaches has been done: A basic processing and an advanced processing. The advantages and disadvantages of both classes is discussed in the following sections.

A. Basic preprocessing

The initial attempt to build a dataset involved categorizing images into two sets: a training set and a validation set. The construction of patches was accomplished by cropping the images into adjacent squares with dimensions of 16×16 pixels. As a result, each image yields $\frac{400^2}{16^2} = 625$ patches. The labels corresponding to these patches are generated based on the method outlined in Section II. An illustrative patch is showcased in Figure 2. As a human, it is not clear to decide whether the patch is a road or not and the question: ”Is it the same for the computer ?” answered in the section V, is legit.

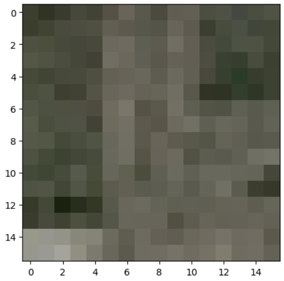


Fig. 2. Example of a patch of size 16x16

A significant challenge associated with this processing lies in the limitation of the training data, which is capped at 625 times the number of images used for training. This limitation gives rise to an additional issue: achieving dataset balance requires either discarding certain background samples (resulting in the loss of total samples) or replicating existing road samples. These challenges collectively weaken the efficacy of this approach in addressing the problem, prompting consideration of a more sophisticated processing approach. Section V provides additional evidence supporting the assertion that this processing method lacks efficiency.

B. *Advanced preprocessing*

To address the issues highlighted in III-A, a novel approach has been introduced. Following the same principles as the basic processing, the dataset is divided into distinct groups for training and validation. However, a significant modification lies in the patch creation process. As detailed in paper [2], a variety of data augmentation techniques can be employed to enhance the robustness and performance of the model. Unlike the conventional method of selecting adjacent patches, the new approach involves randomly choosing the center of the patch. This alteration allows for the generation of a vast array of diverse samples, preventing the occurrence of duplicate identical samples and enabling a perfectly balanced representation of the two classes (road and background) by taking the same number of patch representing a road and a background.

The second major enhancement involves the incorporation of rotated images. By introducing rotated images, roads are generated in orientations other than purely vertical or horizontal. Rotation is applied at angles that are multiples of 45°. This addition not only diversifies the dataset but also contributes to a more robust model. An illustrative example of a rotated image can be observed in Figure 3.

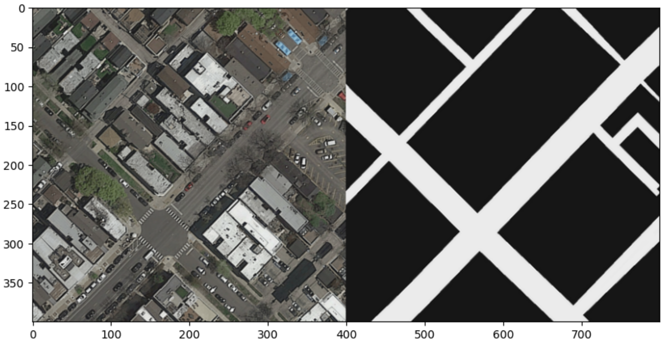
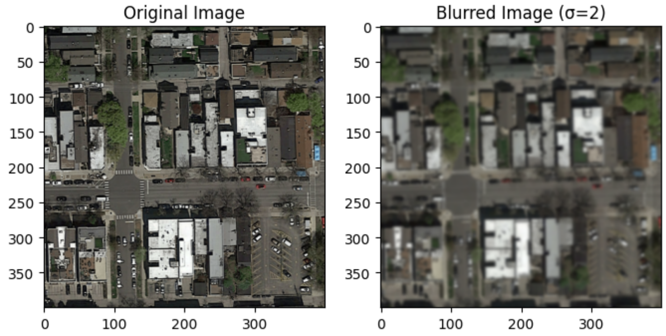


Fig. 3. Rotated image (satImage_035.png) with its groundtruth

By incorporating image rotation, it is important to address the potential issue of padding the resulting image with black pixels. To mitigate this concern, the ‘mode=’mirror’‘ argument from the ‘scipy.ndimage.rotate‘[1] module is employed. This setting ensures that the corners of the image are padded with mirrored pixels from the original image, effectively addressing the problem of road orthogonality. Notably, rotated images are generated with only a 10% probability, maintaining a controlled augmentation strategy.

A complementary data augmentation technique involves image blurring. With the same probability as rotation, images undergo blurring using a Gaussian filter[2] with a standard deviation of $\sigma = 2$. The outcome of this transformation is visualized in Figure 4.


Original Image

Blurred Image (σ=2)

Fig. 4. Blured image (satImage_035.png) with its groundtruth

Lastly, an optimization is made regarding the size of the patches fed into the neural network described in section IV. While predictions are required on patches of size 16×16, it is not a necessity to exclusively input patches of the same size. As discussed earlier in III-A, if discerning whether a patch represents a road is challenging for humans, the algorithm might face a similar challenge. The results section provides conclusive insights into this aspect. Multiple patch sizes are taken into consideration, selected to be multiples of 16. However, increasing the patch size introduces a challenge: a portion of the original image becomes inaccessible. For example, the patch located in the left corner of the image with coordinates [1:16;1:16] cannot be achieved in the augmented space, as the augmented patch must be centered within the original patch. To address this issue, the image needs to be padded using the following formula:

$$\text{size\_padding} = \left\lfloor \frac{\text{aug\_patch\_size} - \text{patch\_size}}{2} \right\rfloor$$

A patch of size 128×128 is illustrated in Figure 5. This patch represents the same fragment as the one shown in Figure 2.

IV. Models

According to the paper [1] and the course, the convolution neural network has been demonstrated to be an efficient model to solve this task due to its ability to extract features from images data. Therefore, the decision was made to build two distinct models. The first one intends to be very simple with very few layers while the second is more large in order to capture more features. The general structure shown in Figure 6 for both models involves applying a successive set of layers

[1] https://docs.scipy.org/doc/scipy/reference/generated/scipy.ndimage.rotate.html

[2] https://docs.scipy.org/doc/scipy/reference/generated/scipy.ndimage.gaussian_filter.html

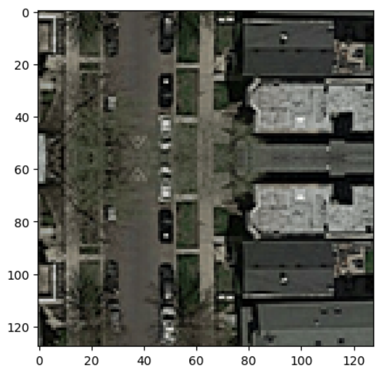


Fig. 5. Example of a patch of size 128x128 pixels

(resp. convolution, activation function and pooling) multiple times. Each convolution intends to extract the features from the input image. Furthermore, the activation layer helps to capture non-linearity. Finally, the pooling reduces the spatial dimension. After the feature learning part, the classification is done by a fully connected neural network with one layer. In order to get a probability of a patch to be a road or background, the BCEWithLogitsLoss[3] loss function from PyTorch combines the sigmoid activation function and the binary cross-entropy loss, which similarly to logistic regression, produces a probability of the output to be in class 1 (road) or 0 (background).

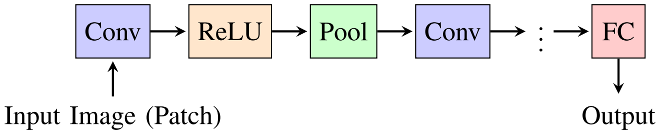


Fig. 6. Convolutional Neural Network Architecture

A. Basic model

The basic model, shown in Figure 7, is composed of two convolutional layers with a kernel size of 3 and increase the number of channels up to 32 (initially 3 for RGB). Additionally, the activation function is the leaky rectified linear unit (LeakyReLU[4]) with a negative slope of 0.1 and the pooling layer is a maximum pooling of kernel size 2.

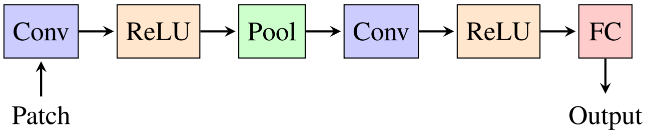


Fig. 7. Basic Convolutional Neural Network

B. Advanced model

The structure of the advanced model, illustrated in Figure 8, closely resembles that of the basic model, with the key difference being its increased depth. In the advanced model, the depth is achieved by executing additional convolutional operations, ultimately reaching 128 filters. However, a challenge often encountered in deep neural networks is overfitting. To mitigate this, the advanced model incorporates extra dropout layers between select convolutional and activation layers. These dropout layers are introduced to decrease the likelihood of overfitting. Each dropout layer has a dropout probability set to 0.1, indicating the likelihood of a node being dropped.

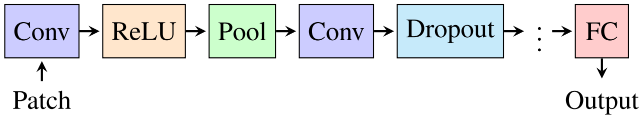


Fig. 8. Advanced Convolutional Neural Network

The primary advantage of this model, in comparison to the previous one, lies in its enhanced feature-capturing capability. This improvement is attributed to the utilization of 128 filters, which is four times greater than that of the basic model. It is important to note that, given the incorporation of multiple max-pooling layers (5 in total), it becomes imperative to ensure that the image patches are of a minimum size, calculated as $2^5 \times 2^5 = 32 \times 32$ pixels.

V. Results

In this section, the focus is based on a comprehensive exploration of various parameters to optimize the performance of two models: the basic and advanced CNN, as outlined earlier. The objective is to see the impact of different configurations on the road segmentation task. Parameters such as patch size, optimizer, and threshold are systematically varied in an effort to discern their influence on model outcomes.

To reiterate, the F1 score, symbolizing the balance between precision and recall, serves as the primary metric in our evaluation. A higher F1-Score, approaching 1, signifies a more robust model for road segmentation. While accuracy is a secondary metric contributing to the overall assessment of model performance, it is important to note that, particularly when working with unbalanced data, the F1 score holds greater significance and provides a more representative measure of model effectiveness than accuracy.

The outcomes of this investigation reveal a significant improvement in performance and accuracy with the increase of the patch size up to 128. The augmentation in patch size corresponds to improved capabilities exhibited by the model, particularly in the context of road segmentation where larger patches enable the model to capture more contextual information and intricate details within the images.

Furthermore, the application of color standardization refines the models, resulting in a noticeable improvement of the F1 score. Subsequent exploration involves investigating

[3] https://pytorch.org/docs/stable/generated/torch.nn.BCEWithLogitsLoss.html

[4] https://pytorch.org/docs/stable/generated/torch.nn.LeakyReLU.html

the impact of adjustments to the threshold, highlighting its comparatively lesser influence when contrasted with variations in patch size. Despite this, a noteworthy observation emerges, indicating a slightly superior result with a threshold set at 0.25.

Additionally, a series of experiments with different optimizer is conducted, confirming the consistent favorable outcomes achieved with the Adam Optimizer across various configurations. The superiority of the Adam optimizer in deep learning tasks is attributed to its adaptive learning rate and momentum capabilities.

In the final stages of the analysis, attempts were made to incorporate blurred images into the model. These efforts resulted in a slight performance increase, indicating that the introduction of blur has an impact on the overall effectiveness of the model. In the end, it emerged as the best-performing model, affirming that the data augmentation implemented successfully achieved the intended goal of enhancing performance.

The results exhibit slight variations from those obtained on AICrowd, demonstrating marginal improvement in the AICrowd evaluation.

The following tables provide a briefly summary of all trained models and their parameters, along with corresponding accuracy and F1 scores. Additionally, results from AICROWD are presented for comprehensive comparison.

TABLE I
DIFFERENT MODELS USED

ID	Model	Patch size	Optimizer	Threshold	Other
1	Basic	16	Adam	0.25	Basic Processing
2	Basic	16	Adam	0.25	-
3	Basic	32	Adam	0.25	-
4	Basic	64	Adam	0.25	-
5	Adv.	64	Adam	0.25	-
6	Adv.	128	Adam	0.25	-
7	Adv.	128	Adam	0.25	Color
8	Adv.	128	AdamW	0.25	Color
9	Adv.	128	Nesterov	0.25	Color
10	Adv.	128	Adam	0.3	Color
11	Adv.	128	Adam	0.2	Color
12	Adv.	128	Adam	0.2	Color + blur
13	Adv.	128	Adam	0.25	Color + blur

TABLE II
RESULTS OF THE DIFFERENT MODELS

ID	F1 score	Acc.	AIcrowd F1	AIcrowd acc.
1	0.648	0.836	-	-
2	0.679	0.800	-	-
3	0.720	0.832	-	-
4	0.743	0.850	0.783	0.876
5	0.792	0.889		-
6	0.833	0.912	0.857	0.923
7	0.857	0.924	0.860	0.926
8	0.851	0.921	-	-
9	0.826	0.907	-	-
10	0.853	0.922	-	-
11	0.853	0.922	0.866	0.928
12	0.856	0.923	0.870	0.930
13	0.856	0.919	0.861	0.926

Based on this tables, the optimal model is the advanced CNN configuration with a threshold of 0.20, utilizing the Adam optimizer, the blur of the images and a patch size of 128. This model achieves an accuracy of 0.923 and an F1-Score of 0.856, where the AICrowd evaluation yields a slightly higher F1-Score of 0.870. An example of a prediction made by this model is exhibited in 9.

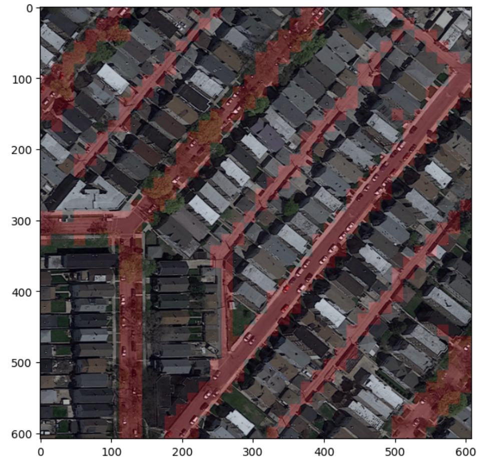


Fig. 9. Example of a solution produced by model 12 (test_48.png)

VI. SUMMARY

This paper highlights the challenges of road segmentation and demonstrates the efficiency of deep learning, particularly convolutional neural networks (CNNs), in solving this task. Choosing the right parameters and layers for the neural network is crucial for building a powerful model. The results reveal significant differences in the quality of predictions among different neural networks based on their structures. Data preprocessing, including augmentation techniques, contributes substantially to improving predictions. Combining these steps yields a solution that accurately distinguishes between roads and backgrounds.

VII. ETHICS

In contemporary machine learning projects like the Road Segmentation Challenge, the evaluation of ethical risks is taking importance. The lecture provided a brief introduction on how to navigate these risks using a helpful canvas. This is the responsibility of the designer of the machine learning project to assess potential impacts and address them responsibly.

Focusing specifically on the Road Segmentation Challenge, one identifiable ethical risk involves safety issues. Through stakeholder analysis, both indirect and direct stakeholders can be identified. Indirect stakeholders, who are impacted by the system but do not directly interact with it, include the public, Government, environment. Direct stakeholders, such as admins, end-users, or contractors, have direct interactions with the system.

Taking the example of safety issues, consider the Government as an indirect stakeholder. Government entities may be concerned with laws related to the security and privacy of Google Earth Images. On the other hand, end-users, as direct stakeholders, may face safety issues due to inaccurate road maps, navigation errors, or potential safety hazards.

Focusing on the specific ethical risk of safety issues, imagine deploying our model in an autonomous car with a 93% accuracy. This implies that 1 in 10 times, the car may misinterpret the presence of a road, potentially leading to accidents, injuries, or even fatalities. This underscore why accuracy is not the primary metric in this challenge; instead, the F1 score is prioritized. The F1 score assigns different weights to various types of false predictions, providing a more nuanced evaluation than accuracy, which simply indicates whether a prediction is correct or not, without considering the impact of a false prediction.

In this case, modifying aspects of the project proved challenging due to the limited number of images received. The finite dataset covered in the project may be expanded with a more extensive image set, and employing deeper neural networks could potentially yield more precise models, thereby reducing the identified ethical risk.

References

[1] Yecheng Lyu and Xinming Huang. Road segmentation using cnn with gru. *arXiv preprint arXiv:1804.05164*, 2018.

[2] Jesus Muñoz-Bulnes, Carlos Fernandez, Ignacio Parra, David Fernández-Llorca, and Miguel A. Sotelo. Deep fully convolutional networks with random data augmentation for enhanced generalization in road detection. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pages 366–371, 2017. doi:10.1109/ITSC.2017.8317901.

Appendix A: Libraries

The python libraries used for this project are:

- Numpy
- os
- re
- argparse
- sklearn
- PyTorch
- Pillow
- Matplotlib

Appendix B: AICrowd

The submission ID for the best model is : #247426