Bird's-Eye Mapper

Extracting Roads from Satellite Images

Authors:

Sanad Alali, Ahmad Alhayek, Mohammad Dahleh

Qusay Abusalem, Hasan Alhalabi

Supervisor: Dr. Tamam Al Sarhan

Course: Deep Learning

Department: Artificial Intelligence Department

Institution: University of Jordan

Date: December 20, 2024

ABSTRACT

The Bird's-Eye Mapper project tackles the critical challenge of extracting road networks from satellite imagery by employing deep learning techniques that surpass traditional mapping inefficiencies. Combining U-Net and Pix2Pix architectures, the project achieves precise road segmentation while addressing occlusions, diverse terrains, and limited annotated datasets. U-Net delivers pixel-level accuracy, and Pix2Pix generates synthetic training data to enhance model generalization. Data augmentation techniques, simulating real-world conditions, further boost performance, culminating in an **F1-score** of **0.992**. Bird's-Eye Mapper takes satellite data and turns it into easy-to-understand maps, helping communities improve roads, respond to emergencies, and plan smarter for the future.

I. Introduction

The extraction of roadway networks from satellite imagery has emerged as a critical component in shaping the trajectory of urban planning, transportation infrastructures, and emergency response frameworks. While traditional mapping techniques provide essential datasets, they often demonstrate significant inefficiencies, manifesting as both time-intensive and resource-demanding processes, in addition to their inability to adapt to rapidly evolving environments. Satellite imagery offers a groundbreaking alternative, providing extensive and nuanced views of the terrestrial surface. Nonetheless, the inherent intricacies associated with this medium—such as feature overlap, irregular roadway formations, and diverse environmental settings—require the formulation of innovative methodologies. With the progression of deep learning technologies, particularly through the utilization of convolutional neural networks (CNNs), there exists a unique opportunity to fundamentally revolutionize road extraction methodologies. These advanced algorithms are equipped to analyze complex visual data, discern intricate roadway attributes, and generate maps with remarkable precision. This research aims to address the persistent challenges of occlusion, noise, and scalability, thereby enabling automated, efficient, and adaptable road extraction techniques that convert raw data into actionable intelligence for a globally interconnected society.

A. Background and Context

The escalating demand for accurate and adaptable road mapping has eclipsed the limitations of conventional methodologies, which often manifest as excessively time-intensive and resource-intensive amidst the rapid proliferation of urban environments and shifting geographical landscapes. Satellite imagery, with its unmatched breadth and detail, presents a transformative opportunity for the systematic mapping of road infrastructures. However, its complexity—marked by obstructions, heterogeneous terrains, and variable resolutions—necessitates solutions that extend beyond traditional methodologies. Bird's-Eye Mapper emerges from this necessity, amalgamating cutting-edge deep learning techniques with geospatial intelligence to address these obstacles. By reimagining the conversion of unrefined satellite data into functional maps, it sets a novel standard for global connectivity and data-driven developmental strategies.

1. Challenges & Problems

- **Obstacles and Noise:** Imagine trying to draw a clear map of roads in a busy, messy picture where trees, buildings, and shadows are blocking your view. That's what our model faces when analyzing satellite images. Roads are not always easy to spot because things like top of them. The model sometimes gets confused and

https://github.com/SanaD-03-alali/Image-Segmentation-using-Unet ends up thinking these obstacles are part of the road or misses the roads entirely. It's like trying to connect the dots in a picture, but some dots are hidden or smudged. This makes the process of identifying roads accurately a big challenge.

- **Generalization Across Different Areas:** Imagine if you learned how to spot roads in sunny cities, but then someone asks you to do the same thing in a snowy village. It would be tricky because everything looks different—the colors, the shapes, even the patterns. Our model faces the same issue when it's trained on satellite images from one place and then must work on images from a completely different region. The images can look very different because of weather, landscape, or even the type of satellite taking the picture. This means the model must learn how to work well in many different environments, which is a lot to ask!

- **Limited Annotated Data:** The scarcity of annotated data significantly hampers the development of accurate road detection models. Manual labeling of satellite images is labor-intensive and costly, leading to limited datasets that may not fully represent the diversity of road types and environmental conditions. [1]

- **Diverse Road Types:** The wide range of road types, from urban highways to rural paths, presents challenges in segmentation. Differences in road width, texture, and material composition, along with environmental factors like shadows and vegetation, complicate the accurate delineation of roads in satellite imagery. [2]

These challenges underscore the importance of robust, reliable systems capable of managing recognition tasks through the use of advanced AI and technologies.

2. Our Vision

At Bird's-Eye Mapper, our objective is to revolutionize geospatial intelligence by offering an avant-garde platform that enables accurate and automated extraction of road networks from satellite imagery. We strive to empower urban planners and transportation authorities with instantaneous mapping solutions that effectively reconcile the gap between raw satellite data and actionable insights. By leveraging advanced machine learning and deep learning techniques, Bird's-Eye Mapper seeks to reassess the methodologies through which infrastructure data is gathered and utilized, thereby facilitating sustainable development and the creation of more intelligent urban landscapes. Our primary aim is to make geospatial technology accessible, reliable, and impactful in nurturing a more interconnected global community.

II. Related work

A. Machine Learning Project Road Segmentation

In the study conducted by Lam et al., a robust machine learning approach was employed to segment roads from satellite images using convolutional neural networks (CNNs) and U-Net architectures. The project involved extensive data augmentation, expanding an initial dataset of 100 images through rotations, noise addition, and lighting variations to address challenges such as angled roads and varying conditions. The U-Net model achieved a maximum F1-score of 0.905 on test data, outperforming traditional CNNs by effectively leveraging its encoder-decoder structure for pixel-wise classification. This research highlights the importance of tailored data augmentation and model optimization in addressing the complexities of road segmentation from aerial imagery, serving as a critical benchmark for similar tasks**.** [3]

B. Machine Learning CS-433 - Class Project 2 - Road Segmentation

Braz et al. explored the use of a U-Net architecture for road segmentation from satellite images, leveraging its encoder-decoder structure for pixel-wise classification. Their approach emphasized the importance of data augmentation, including 45° rotations and flips, to address the limited dataset and improve predictions for diagonal roads. They achieved an F1-score of 88.4% on the AI crowd dataset test, showcasing the model's effectiveness after training for just 35 epochs using an RTX 2060 GPU. The implementation also integrated batch normalization for improved performance and applied post-processing techniques, such as morphological operators, to enhance prediction continuity. Their findings underscore the critical role of tailored data preprocessing and optimization in achieving accurate road segmentation. [4]

III. Methodology

In this section, we introduce the detailed implementation of our road segmentation project, utilizing two advanced architectures: UNet and Pix2Pix. These models were chosen for their complementary strengths, offering high accuracy in pixel-level classification and the ability to synthesize data for enhanced generalization. Below, we outline the methodology, beginning with UNet, followed by Pix2Pix, and describe their significance in the context of our project.

A. Models Used

1. U-Net: Convolutional Networks for Image Segmentation

The network architecture is illustrated in Figure 1 below. It consists of a contracting path (left side) and an expansive path (right side). The contracting path follows the typical architecture of a convolutional network. It consists of the repeated application of two 3x3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling. At each downsampling step, the number of feature channels is doubled.

Every step in the expansive path consists of an upsampling of the feature map followed by a 2x2 convolution (“up-convolution”) that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU. The cropping is necessary due to the loss of border pixels in every convolution. At the final layer, a 1x1 convolution is used to map each 64-component feature vector to the desired number of classes. In total, the network has 23 convolutional layers. [5]

UNet's ability to process entire images instead of patches addresses the redundancy issue and significantly reduces computational complexity, making it a practical choice for segmentation tasks. In our project, UNet was trained on a combination of original and augmented datasets, achieving high segmentation accuracy with minimal loss.

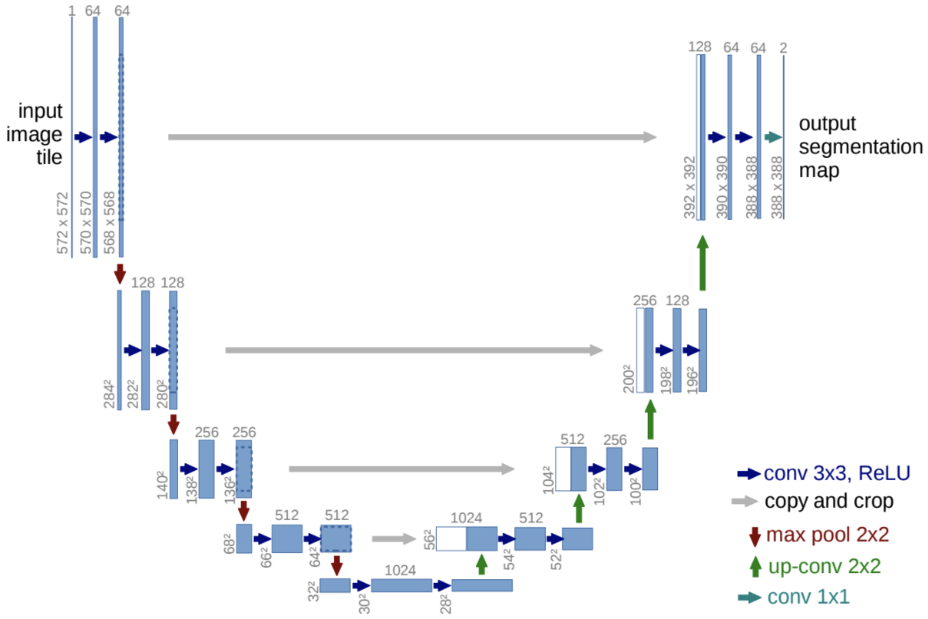

Figure 1: U-net architecture (example for 32×32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

2. Pix2Pix: Conditional Adversarial Networks for Image-to-Image Translation

Pix2Pix leverages conditional adversarial networks (cGANs) as a general-purpose solution for image-to-image translation tasks, where the objective is to map input images to corresponding output images **[6]**. Unlike traditional supervised learning approaches that rely on pre-defined loss functions, cGANs dynamically learn a loss function tailored to the data, ensuring sharper and more realistic outputs. This adaptability is critical for tasks such as semantic segmentation, edge-to-photo synthesis, and colorization, where preserving fine-grained details and context is paramount.

The architecture of Pix2Pix consists of two main components: a U-Net-based generator and a PatchGAN discriminator. The generator uses the contracting-expanding design of U-Net to effectively preserve spatial information during transformation. This ensures high-quality mapping between input and output domains. Meanwhile, the PatchGAN discriminator focuses on distinguishing real from fake data at the scale of image patches rather than the entire image. This patch-based approach emphasizes high-frequency detail preservation, a key feature for generating photorealistic textures and structures. **[6]**

As shown in Figure 2 below, Pix2Pix demonstrates its versatility across a range of applications, including translating between maps and aerial photographs at 512×512 resolution. Although trained on images at 256×256 resolution, the model generalizes effectively to larger inputs, a testament to its robust design. In the illustrated example, Pix2Pix achieves impressive results in maintaining geographical features, urban layouts, and fine details in both translation directions. The output showcases the ability to handle structured and unstructured transformations, critical for real-world applications such as satellite imagery analysis and geographic visualization.

In the context of our road segmentation project, Pix2Pix was employed to augment the dataset with synthetic examples, mitigating the challenges posed by limited labeled data. This augmentation enriched the model’s training distribution, enhancing its robustness and improving segmentation accuracy. By generating realistic road networks from sparse input images, Pix2Pix played a complementary role alongside UNet, creating a comprehensive pipeline for precise and reliable road extraction.

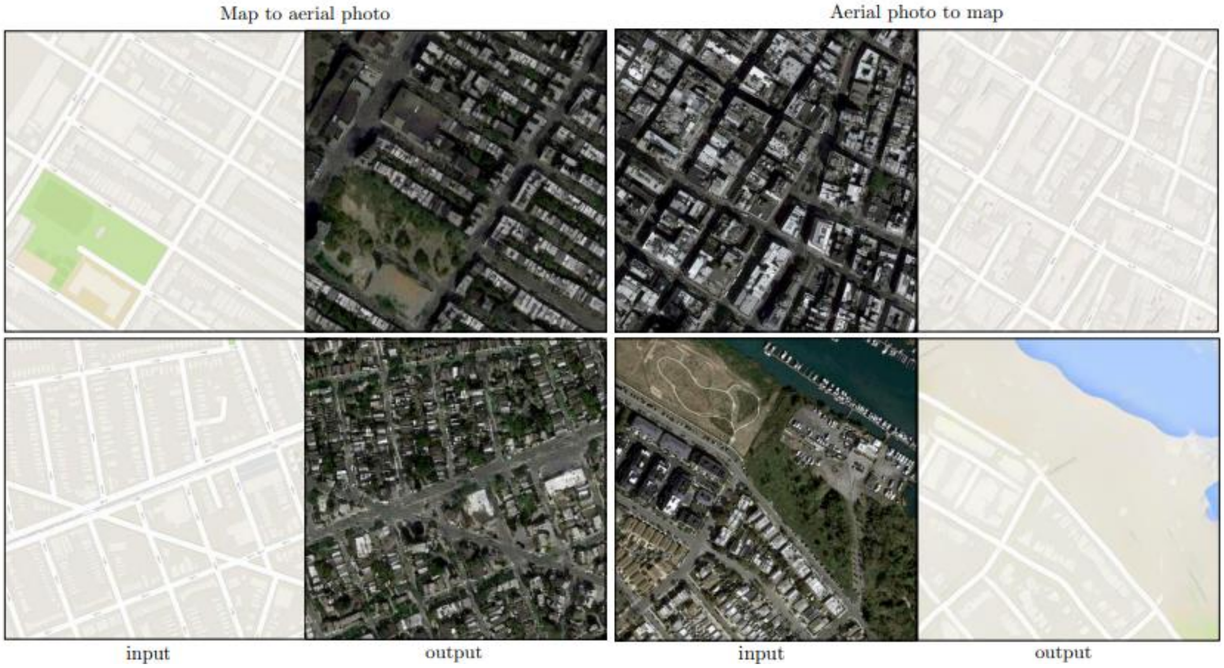

Figure 2: Example results on Google Maps at 512x512 resolution (model was trained on images at 256 × 256 resolution and run convolutionally on the larger images at test time). Contrast adjusted for clarity.

IV. Experiments and Results

In this section, we present the experimental setup and results of our road segmentation project. Initially, the models were trained on a dataset consisting of 100 training images and 50 testing images without any augmentation. To enhance the model's generalization capability and address the challenge of limited data, we subsequently applied various data augmentation techniques. These augmentations aimed to simulate diverse real-world conditions and improve the robustness of the models. The results of the experiments, including the evaluation metrics and visual outputs, are detailed in the subsections below.

A. Data Augmentation

To tackle the challenge of limited training data and improve the road segmentation model, we carefully designed a step-by-step data augmentation process across multiple trials. Each trial built upon the previous one, gradually enriching the dataset with more variety and complexity to help the model learn better and generalize effectively.

- **Trial 1: Baseline Training with Original Data**

Our first step was to train the model using the original dataset of 100 training images and 50 testing images. This served as a baseline to measure the model's performance without any augmentation. As expected, the small size and uniformity of the dataset limited the model's ability to generalize to new, unseen data.

- **Trial 2: Generating New Data with Pix2Pix**

To address the limitations of the baseline, we introduced synthetic data generated by the Pix2Pix model. By transforming map-style inputs into realistic aerial images, Pix2Pix provided additional training samples, adding new perspectives and variations. This enriched dataset made a significant difference, enabling the model to perform better by filling in the gaps left by the original dataset.

- **Trial 3: Basic Augmentation Techniques**

Building on the improved dataset from Trial 2, we applied a series of basic augmentation techniques to add even more variety. These included:

  - Random flips (horizontal and vertical).
  - Adjustments to brightness, contrast, saturation, and hue.
  - Random affine transformations like slight translations, scaling, and rotations.
  - Gaussian blur with varying intensities.

These transformations mimicked diverse environmental conditions, making the dataset more dynamic. For this trial, the model was trained for 10 epochs, marking the beginning of integrating systematic augmentations. Figure 3 shows how the images look like when applying augmentations.

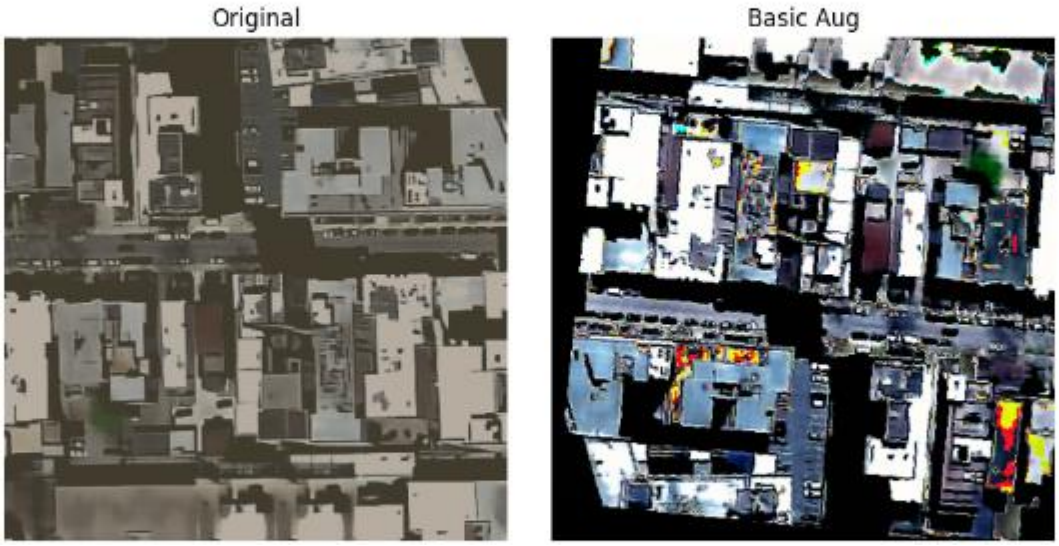

Figure 3: The picture on the left is the original image, and the one on the right is when the basic data augmentation techniques applied

- **Trial 4: Extending Training with Basic Augmentation**

In this trial, we kept the same set of basic augmentation techniques from Trial 3 but increased the number of training epochs from 10 to 20. This additional training time allowed the model to better absorb the patterns introduced by the augmented data, resulting in a more refined understanding of diverse scenarios. This trial showed how extended training can maximize the benefits of even simple augmentations.

- **Trial 5: Combining Basic and Advanced Augmentation Techniques**

For the final trial, we took the augmentation process a step further by combining basic techniques with advanced ones. These included:

- **Elastic Transformations**: Random distortions that mimic natural deformations caused by terrain or lens effects, improving robustness to irregular road shapes.
- **Shadow/Cloud Overlays**: Simulations of shadows or clouds as semi-transparent layers to better handle real-world satellite imagery conditions. Figure 4 below shows the augmentations applied.

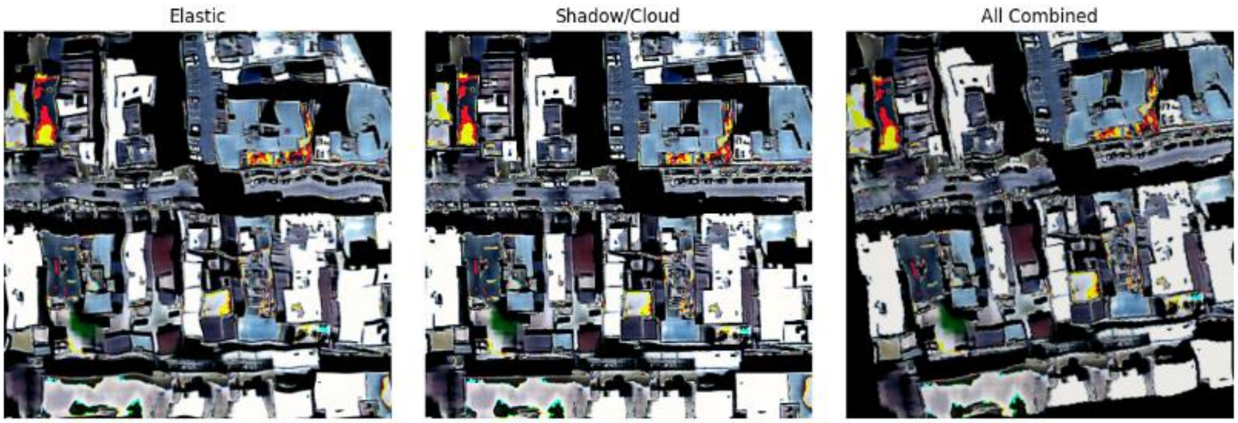

Figure 4: The picture on the left is when we applied the elastic transformation, and the one in the middle is when we applied the shadow/cloud overlay, and the last one is when we combined all techniques at once

The combination of these advanced techniques with the existing basic transformations created a diverse and comprehensive training dataset. With extended training epochs, the model became highly adaptable to complex and unpredictable real-world scenarios, delivering its best performance in this trial. Through this methodical augmentation process, we observed steady improvements in the model's ability to segment roads accurately. The results of each trial will be discussed in the following subsection.

B. Results

This subsection provides an overview of the outcomes from the various trials conducted during the project. Each trial tested different approaches and augmentation techniques, allowing us to observe how these changes affected the model's performance. Key metrics, such as F1 score, training and validation loss, and learning rate, were used to evaluate the effectiveness of each approach. Throughout all trials, we consistently used the Adam optimizer and a batch size of 4 for testing.

Table 1 below summarizes the results:

Table 1: Our Experiment Results

Trial	**Image Size**	**Batch Size**	**Epochs**	**Learning Rate**	**Test F1 Score**	**Train Loss**	**Val Loss**
Trial 1: Original Data	416x416	4	15	0.0001	0.96	0.1035	0.1962
Trial 2: Pix2Pix-Generated Data	416x416	32	10	0.0001	0.982	0.0459	0.0455
Trial 3: Basic Augmentation	416x416	32	10	0.0001	0.932	0.3047	0.3102
Trial 4: Basic Augmentation + Epochs	416x416	32	20	0.0001	0.979	0.0854	0.1214
Trial 5: Advanced Augmentation	416x416	32	20	0.0001	**0.992**	0.1300	0.2249

The table above highlights the gradual improvements observed as more advanced augmentation techniques and extended training durations were introduced. Notably, **Trial 5** achieved the highest **F1 score** of **0.992**, indicating a significant improvement in model performance. By iteratively refining the training process, including the use of advanced augmentation methods, we achieved notable gains in model accuracy and robustness. Further details on the augmentation methods used in each trial can be found in the Data Augmentation subsection.

In **Trial 2,** the Pix2Pix model played a vital role in creating synthetic data, significantly enriching the training dataset. The training process for Pix2Pix used a batch size of 4,3000 epochs, a learning rate of 0.0001, and a lambda pixel value of 50. During training, the discriminator loss (D_loss) was 0.0001, while the generator loss (G_loss) reached 14.0379.

Although these hyperparameters contributed to generating high-quality image translations and enhancing model performance, slight overfitting was observed in the last two trials. This indicates room for improvement, which could be addressed by incorporating regularization techniques or using a learning rate scheduler to better manage training dynamics.

V. Conclusion

In this project, we focused on using GANs for data augmentation and U-Net for image segmentation to extract roads from satellite images. The results were promising, showing that the model can accurately detect roads in complex imagery, even with limited labeled data. This could significantly benefit fields like urban planning, infrastructure development, and disaster response. However, while the approach shows strong potential, there is still room for improvement, especially in model generalization and handling diverse, real-world data. The project has provided valuable insights into satellite image analysis, setting the stage for future refinements and more scalable solutions.

VI. Limitations

While our approach shows promise, it has several limitations. The model’s performance depends on the quality and variety of the original training data, which is currently inefficient and insufficient. The testing dataset is also small, limiting the model’s ability to generalize to new, unseen images, especially in areas with unique road patterns. Moreover, the computational requirements of GANs and U-Net make real-time deployment challenging. The model focuses only on road detection without considering different road types or traffic factors, which could improve its practical use.

References

[1] Y. &. A. H. Nachmany, "Detecting Roads from Satellite Imagery in the Developing World.," 2019.

[2] Y. a. X. Z. a. F. Y. a. C. Z. Xu, "Road Extraction from High-Resolution Remote Sensing Imagery Using Deep Learning," 2018.

[3] [Online]. Available: https://github.com/zghonda/EPFL-Machine-Learning-Road Segmentation/blob/master/ML_Road_Segmentation_Project.pdf.

[4] [Online]. Available: https://github.com/LucasBrazCappelo/ML_EPFL_Project_2/blob/main/report/BRAZ_DURAND_NICOLLE_Project2_Road_Segmentation_ML_EPFL.pdf.

[5] O. a. F. P. a. B. T. Ronneberger, "U-net: Convolutional networks for biomedical image segmentation," pp. 234--241, 2015.

[6] J.-Y. Z. T. Z. A. A. E. Phillip Isola, "Image-to-Image Translation with Conditional Adversarial Networks," 2016.