Road Segmentation

Amine Bengelloun, Mustapha Yassine Wahidy, Amine Bousseta

December 19, 2024

Abstract

Semantic segmentation of roads from aerial imagery is crucial for applications such as autonomous driving and urban planning. This project explores various architectures, including a custom U-Net, advanced models from the `smp` library [1], and a novel SPIN-based approach integrating spatial and interaction reasoning. The SPIN Model, inspired by [2], achieves superior performance with an F1 score of 0.879, demonstrating its ability to capture road connectivity. This report presents our methodology, results, and insights into the trade-offs of these architectures for road segmentation.

I. Introduction

Road segmentation involves identifying roads in aerial or satellite images, assigning a label (0 or 1) to every pixel to classify it as part of a road (1) or background (0). This task is crucial for autonomous driving, disaster response, and infrastructure development.

In this project, we address this challenge by exploring multiple deep learning architectures:

1. **Custom U-Net:** A baseline model implemented from scratch.
2. **Advanced Architectures:** Feature Pyramid Network (FPN), U-Net using the `smp` library.
3. **SPIN-Based Models:** A novel integration of SPIN modules inspired by the SPIN Road Mapper framework [2].

Our objectives include comparing the performance of these architectures using metrics such as F1 score and accuracy while understanding their strengths, limitations, and trade-offs. This report documents our approach, implementation, and findings.

II. Dataset and Preprocessing

The dataset used in this project consists of high-resolution aerial images, making it well-suited for road segmentation tasks. The original dataset is divided as follows:

- **Training set:** 100 images, each paired with a corresponding binary mask.
- **Test set:** 50 images, provided without labels for evaluation.

Each image is of size 400 × 400 pixels with 3 channels, representing RGB color. These images capture aerial views of roads and surrounding landscapes.

To optimize the training process, we further split the training set into:

- **Training subset:** 90% of the original training data.
- **Validation subset:** 10% of the original training data.

This 90/10 split was used to monitor model performance and prevent overfitting during training. Both subsets maintain the original resolution and characteristics of the dataset.

Preprocessing Pipeline

To ensure consistent and effective model training, the following preprocessing steps were applied:

1. **Normalization:** Pixel values were scaled to the range $[0, 1]$ to enhance numerical stability.
2. **Data augmentation:** Techniques such as flipping, rotation, and cropping were applied to increase data variability and robustness.

These steps ensured a high-quality input dataset for training and evaluation, supporting the robust performance of the models implemented in this project.

III. Models

In this project, we explored a progression of architectures to perform semantic segmentation for road detection. Starting from a simple implementation, we incrementally adopted more advanced techniques, utilizing both custom-built and library-based models.

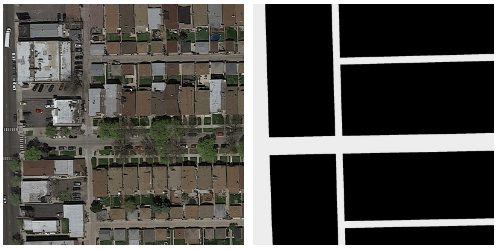

Figure 1: (Left) Example aerial image from the dataset. (Right) Corresponding segmentation mask.

Custom U-Net Implementation

The U-Net architecture, known for its effectiveness in biomedical image segmentation, served as our baseline. We implemented it from scratch, ensuring a clear understanding of its symmetric encoder-decoder structure with skip connections. The model was trained using binary cross-entropy loss and optimized with the Adam optimizer. Key hyperparameters included:

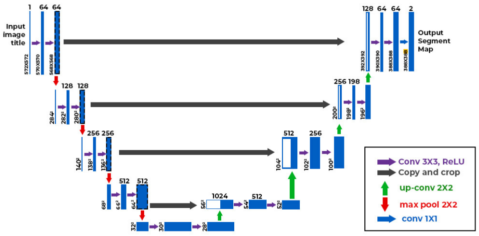

Figure 2: Architecture of the U-Net model used as a baseline. It consists of a symmetric encoder-decoder structure with skip connections.

Advanced Architectures with smp

To build upon the baseline, we utilized the `smp` library, implementing a variety of state-of-the-art architectures:

- **Feature Pyramid Network (FPN):** Designed for multiscale feature aggregation.
- **U-Net:** Enhanced versions of U-Net with denser skip connections for improved feature propagation.

All these architectures employed ResNet50 as their encoder, with some leveraging pretrained weights on ImageNet. The models were fine-tuned for the road segmentation task, using the same training settings as the custom U-Net.

SPIN Model

The SPIN (Spatial and Interaction Space Reasoning) model enhances road segmentation by addressing key challenges such as disconnected road segments, multi-scale variability, and complex spatial relationships. Building on the SPIN Road Mapper framework [2], we implemented a more simple and flexible architecture that incorporates spatial attention and graph-based reasoning mechanisms for improved segmentation performance.

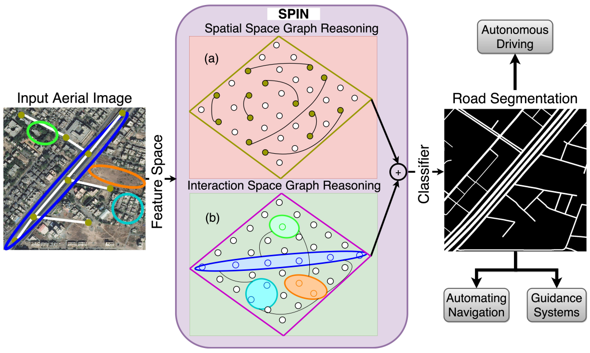

Figure 3: SPIN Module Workflow.

Generalized SPIN-Based Architecture

We implemented a modular framework that supports a variety of backbone models from the `torchvision.models.segmentation` library. This flexibility allows us to experiment with and optimize different pretrained models for the task at hand.

Key components of the generalized SPIN-based architecture include:

- **Backbone:** Configurable feature extractor (e.g., ResNet50 or ResNet101) using pretrained models from the `torchvision.models.segmentation` library.
- **FPN Decoder:** A Feature Pyramid Network decoder that aggregates multi-scale features from the backbone to handle the variability in road shapes and sizes.
- **SPIN Pyramid:** A multi-scale reasoning module integrating spatial attention and graph-based reasoning to refine features and improve road connectivity.
- **Segmentation Head:** A 1 × 1 convolutional layer that outputs the final binary segmentation mask.

SPIN Module and SPIN Pyramid

The SPIN module, a core component of the architecture, refines features using spatial attention and interaction space reasoning. It operates at multiple scales within the SPIN Pyramid, enabling enhanced multi-scale reasoning.

1. **Spatial Attention:** A 7 × 7 convolution generates an attention map that highlights road-like regions while suppressing noise.

2. **Graph-Based Reasoning:** Models long-range dependencies between spatial regions by representing the feature map as a graph, improving connectivity in road segmentation.

3. **Feature Fusion:** Combines the outputs of spatial attention and graph reasoning to produce refined feature maps.

The SPIN Pyramid applies these modules at three scales—original, half, and quarter resolutions—before upsampling and aggregating the outputs, ensuring robustness to roads of varying widths and orientations.

Key Adaptations

Compared to the original SPIN Road Mapper framework, the following changes were introduced:

- **Backbone Agnosticism:** The reliance on specific backbones was replaced with a configurable architecture supporting multiple models.

- **Simplified Graph Reasoning:** The graph reasoning layer was adjusted to balance computational complexity and performance.

- **Integration of FPN Decoder:** An FPN decoder replaced the original custom decoder for improved feature aggregation across scales.

These adaptations make the architecture more versatile while maintaining its ability to model road connectivity and fine-grained details.

IV. Training and Evaluation

Training Details

- **Optimizer:** Adam

- **Device:** Apple MPS, CUDA GPU or CPU.

Hyperparameter Tuning

- **Learning Rate Scheduling:** We observed that the SPIN model's training loss plateaued at around 0.04 when using a constant learning rate. To address this, we employed the `ReduceLROnPlateau` scheduler, which dynamically reduces the learning rate when the validation loss stops improving. This approach helped the model escape potential local minima and ensured a smoother convergence.

- **Batch Size:** Given the limited size of the training set, the batch size was set to 16.

- **Early Stopping:** To prevent overfitting, an early stopping mechanism was implemented during training.

- **Loss Function:** A combined loss function comprising Binary Cross-Entropy (BCE) and Dice Loss was used to address class imbalance. This combination improved segmentation accuracy, particularly for narrow roads and fine-grained details.

Evaluation Metrics

Performance was evaluated using:

- **F1 Score:** Measures the balance between precision and recall.

- **Accuracy:** Percentage of correctly classified pixels.

V. Performance and Ablation Study

The SPIN Road Mapper was trained with Binary Cross-Entropy (BCE) combined with Dice Loss to handle class imbalance. Using Adam optimization with a learning rate of 1×10^{-4} and a batch size of 16, the model achieved state-of-the-art performance.

Model	F1-Score	Accuracy
Handcrafted U-Net	0.745	0.861
`smp` U-Net (pre-trained weights)	0.783	0.878
SPIN + DeepLabV3+	0.872	0.930
SPIN + FCN8	**0.879**	**0.933**

Table 1: Performance metrics for the implemented models.

The full SPIN model achieves the highest performance, benefiting from the synergy of spatial attention, graph reasoning, and multi-scale learning.

VI. Visual Results

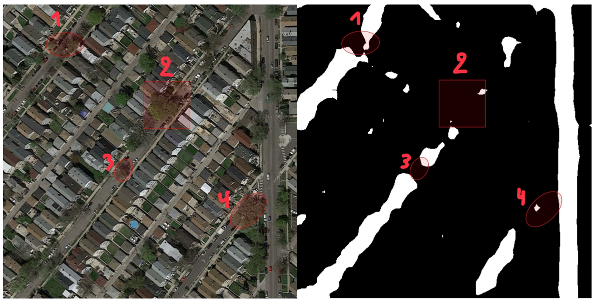

Figure 4: Segmentation Results. (Left) Test Dataset Image, (Right) Predicted mask. The SPIN Road Mapper demonstrates superior road connectivity and segmentation precision.

The SPIN Road Mapper represents a significant advancement in road segmentation, leveraging innovative spatial and interaction reasoning to achieve state-of-the-art results.

VII. Results and Analysis

This section presents the quantitative and qualitative results of the models implemented, followed by a comparative analysis of their performance. All models were evaluated using F1-score and accuracy metrics.

Analysis and Discussion

The results demonstrate a clear progression in performance:

- **Handcrafted U-Net:** Achieved moderate performance but struggled with boundary precision and small object segmentation.
- `smp U-Net:` Showed improved F1-score and accuracy due to the use of pretrained weights, which enhanced feature extraction.
- **SPIN + DeepLabV3+:** Benefited from the Atrous Spatial Pyramid Pooling (ASPP) module, leading to a significant increase in segmentation quality.
- **SPIN + FCN8:** Delivered the best results, achieving the highest F1-score and accuracy due to its optimized spatial pyramid design.

Despite the overall success, challenges remain in areas such as misclassification in occluded regions and handling edge cases in class imbalance. These findings suggest potential avenues for further improvements and optimizations.

VIII. Ethical Risks

In this project, we considered the ethical implications associated with the deployment of semantic segmentation models, particularly in the context of road segmentation for autonomous vehicles. The following aspects were analyzed:

VIII. Identified Ethical Risk

One potential ethical risk is the **misuse of segmentation models for surveillance applications**, where aerial imagery could be exploited for unauthorized monitoring of individuals or private properties. This raises concerns related to:

- **Privacy violations:** Unauthorized access to sensitive geographic or personal information.
- **Discrimination:** Potential bias in data collection or model predictions, disproportionately impacting certain communities.

Stakeholders Impacted

The stakeholders affected by this risk include:

- **Individuals:** Whose privacy may be compromised by improper use of segmentation technology.
- **Organizations:** Responsible for adhering to ethical standards and ensuring fair use of the technology.
- **Regulators:** Tasked with creating and enforcing policies to prevent misuse.

Evaluation and Mitigation

To address this risk, we conducted the following assessments:

- Reviewed privacy policies and ethical guidelines in aerial imagery analysis.
- Ensured the dataset used contains no personally identifiable information.
- Limited the scope of our project to road segmentation tasks, avoiding features that could directly identify individuals or properties.

While these steps mitigate the immediate risks, future work should include:

- Developing privacy-preserving methods, such as obfuscating sensitive data.
- Advocating for standardized ethical frameworks in the deployment of semantic segmentation models.

VIII. Concluding Remarks on Ethical Considerations

By restricting the scope of this project and ensuring compliance with privacy guidelines, we minimized potential ethical risks. However, ongoing vigilance and collaboration with policymakers and ethicists are essential to maintain the responsible use of this technology.

IX. Conclusion

This project investigated semantic segmentation for road detection using a progression of architectures, from a custom U-Net to advanced models like FPN, U-Net++, and SPINFCN8. The SPINFCN8 model achieved the highest performance, demonstrating the value of advanced architectural elements and pretrained encoders.

Despite the overall success, challenges such as handling occlusions and class imbalance remain. Future work could focus on improving robustness through better loss functions, expanding the dataset, and enabling real-time processing.

Ethical risks, including privacy violations and potential misuse for missile navigation, were evaluated and mitigated by restricting the scope to civilian applications and adhering to ethical guidelines.

In summary, this project highlights the effectiveness of modern segmentation models for road detection and underscores the importance of responsible AI development.

References

[1] OpenMMLab, *Segmentation Models PyTorch Library (smp)*, [Online]. Available: `https://github.com/qubvel/segmentation_models.pytorch`

[2] SPINpaper, *SPIN Road Mapper Paper: Spatial and Interaction Space Graph Reasoning for Road Extraction from Aerial Images*, [Online]. Available: `https://arxiv.org/abs/2109.07701`

[3] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.

[4] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. Yuille, *DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[5] EPFL CEDE, *Digital Ethics Canvas*, [Online]. Available: `https://www.epfl.ch/education/educational-initiatives/cede/training-and-support/digital-ethics/a-visual-tool-for-assessing-ethical-risks/`

[6] M. Paul, R. Mondal, R. Krishnamurthy, and B. Mitra, *SPIN Road Mapper: Spatial and Interaction Space Reasoning for Road Segmentation*, in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2021. [Online]. Available: `https://arxiv.org/abs/2109.07701`

[7] PyTorch Documentation, [Online]. Available: `https://pytorch.org/docs/`