Date of publication 12, 2024
Digital Object Identifier 10.1109/ACCESS.2023.0322000

Preparation of Papers for IEEE ACCESS

VRAJ PANDYA[1]
[1]Wilfrid Laurier University, Waterloo ON Canada (e-mail: pand8760@mylaurier.ca)

ABSTRACT The rapid advancement of satellite imagery and deep learning technologies has opened new avenues for automated geospatial data extraction and mapping. This research presents a novel approach to road network extraction from high-resolution satellite imagery, leveraging deep learning techniques and the Spacenet dataset to achieve over 99% of accuracy and DICE coefficient without overfitting by accurately identifying and delineating road networks from satellite imagery, and moreover, presenting an approach to optimize the post-processing with the ultimate goal of contributing to open-source mapping platforms like OpenStreetMap (OSM). Also, the manuscript highlights challenges and opens avenues for future research, including developing new metrics akin to the DICE coefficient and optimizing computational efficiency for model training with limited resources.

By implementing and comparing three distinct Convolutional Neural Network (CNN) architectures using deep learning capabilities, the research systematically evaluated the performance of road network extraction techniques. The proposed methodology encompassed comprehensive data pre-processing, advanced deep learning model training, and post-processing strategies to transform raster road network predictions into vector data suitable for geospatial analysis. The research workflow demonstrated a seamless integration of satellite imagery analysis, deep learning road extraction, and open-source mapping platforms, thereby advancing automated geographic information system (GIS) methodologies with vectorized outputs suitable for seamless integration into mapping software platforms.

The manuscript here, highlights the potential of integrating deep learning techniques with professional GIS software to enhance road network mapping and contributes to the expanding field of automated geospatial intelligence.

INDEX TERMS **In Alphabetical Order**
Convolutional Neural Networks, Deep Learning, Geographic Information Systems, Geospatial Analysis, Machine Learning, Raster-to-Vector Conversion, Remote Sensing, Road Network Extraction, Satellite Imagery, Spatial Data Processing

I. INTRODUCTION

THe exponential growth of global urbanization and infrastructure development has intensified the need for accurate, up-to-date geographical information systems (GIS). Traditional road network mapping methods, which are highly based on manual digitization and ground surveys, are increasingly inadequate to address the rapid pace of urban transformation. Satellite imagery, coupled with advanced machine learning techniques, presents a transformative solution to this challenge, offering the potential for automated, efficient, and scalable road network extraction. Additionally, Researchers have proposed several methods to improve road extraction accuracy. Multi-task learning has shown potential in enhancing road segmentation and topological accuracy, supporting integration with mapping platforms like OpenStreetMap (OSM) [1], Deep residual U-Nets addressed vanishing gradients, significantly improving road detection in complex urban areas [2], Pyramid Scene Parsing Networks (PSPNet) captured multi-scale contextual information, effectively managing variable road widths and intersections [3] and Patch-based CNNs improved feature extraction for remote sensing applications [4] while generative adversarial networks (GANs) provided insights into data augmentation and model robustness [5].

However, manual interpretation of these complex images remains a time-consuming and resource-intensive process. Deep learning, particularly Convolutional Neural Networks (CNNs), has emerged as a powerful paradigm for automated feature extraction and pattern recognition in geospatial data. These advanced algorithms can effectively learn and distinguish complex patterns in the road network from satellite imagery, transcending the limitations of traditional image processing techniques. The primary objective of this research is to develop an innovative road network extraction methodology utilizing deep learning algorithms. Using the Spacenet data set, a comprehensive collection of high-resolution satellite imagery, this study aims to demonstrate a robust approach to the identification and mapping of automated roads networks. The proposed methodology addresses several critical challenges in current geospatial data extraction:

- Accurate identification of road networks across diverse geographical landscapes
- Efficient transformation of raster-based predictions into vectorized geographical data
- Seamless integration with open-source mapping platforms like OpenStreetMap (OSM)

The significance of this research extends beyond technological innovation. By developing an automated road network extraction system, we contribute to more dynamic and responsive geographical information systems. Such advancements have profound implications for urban planning, infrastructure development, transportation management, and humanitarian efforts that rely on accurate geographical data. The subsequent sections of this paper will comprehensively detail the proposed methodology, including data preprocessing strategies, deep learning model architectures, experimental setup, and performance evaluation metrics. By presenting a holistic approach to road network extraction, this research aims to provide valuable insights and a reproducible framework for future geospatial machine learning applications.

The manuscript is structured into five main sections. Section 1 serves as the Introduction, providing an overview of the research on automated road network extraction from satellite imagery using deep learning techniques. Section 2 presents a comprehensive Literature Review, covering foundational architectural developments, advanced semantic segmentation techniques, generative methodologies, computational innovations, and multi-spectral approaches in road network extraction. Section 3 details the Methodology, including the datasets used, data preparation, preprocessing, the overall pipeline for road extraction from satellite imagery consisting of Model training, inferencing and postprocessing. Section 4 addresses the results and the need to discuss the project further with comparisons of experiments conducted on each model. Finally, Section 5 discusses the conclusions of the research.

II. LITERATURE REVIEW

The domain of road network extraction from satellite imagery represents a critical intersection of computer vision, geospatial analysis, and artificial intelligence. This review synthesizes insights from a diverse range of groundbreaking studies, demonstrating the evolution and current state of road extraction techniques.

A. FOUNDATIONAL ARCHITECTURAL DEVELOPMENTS

The evolution of road network extraction can be traced through several pivotal architectural innovations. Van Etten's seminal work on city-scale road extraction established a fundamental framework for understanding road networks beyond mere visual identification. By demonstrating the potential to estimate road speeds and travel times, this research expanded the conceptual boundaries of satellite imagery analysis [1]. Zhang et al. [2] Deep Residual U-Net marked a significant leap in architectural design. Their integration of residual learning within the U-Net framework addressed critical challenges in extracting complex urban road networks. The residual connections introduced by this approach effectively mitigated information loss during deep network training [2].

B. ADVANCED SEMANTIC SEGMENTATION TECHNIQUES

Recent advancements in semantic segmentation have significantly improved road extraction accuracy. Chen et al. advanced semantic segmentation through a multi-scale feature fusion approach using atrous convolution and fully connected CRFs. By capturing road networks at multiple scales, they addressed the challenge of handling varying road widths and connectivity [3]. The comparison of backbones for semantic segmentation networks by Zhang et al. offers crucial guidance for model architecture selection. Their findings on the performance of various backbones inform decisions on experimenting with multiple architectures for specific road extraction tasks [4].

C. GENERATIVE AND SYNTHETIC METHODOLOGIES

The emergence of generative techniques brought new dimensions to road network research. Hartmann et al.'s StreetGAN demonstrated the potential of Generative Adversarial Networks (GANs) in road network synthesis [5]. Zhang et al. further explored this approach, using an improved GAN for aerial image road extraction [6].

D. COMPUTATIONAL AND PROCESSING INNOVATIONS

Recent research has addressed computational challenges in processing large-scale satellite imagery. Wang et al. achieved remarkable reductions in processing time for high-resolution satellite imagery, demonstrating the scalability of deep learning approaches to large-scale geospatial analysis [7]. Gong et al. focused on both extraction and vectorization of road networks from remote sensing images [8]. Integration and Post-Processing Advancements Sharma et al. introduced a patch-based convolutional neural network for remote sensing image classification, providing valuable insights for processing large-scale imagery [9]. Rodriguez-Gonzalez and Brovelli explored transfer learning approaches for road network extraction, addressing the challenge of limited labeled data [10].

E. MULTI-SPECTRAL AND SPECIALIZED APPROACHES

Park et al. investigated multi-spectral road detection in satellite images, highlighting the importance of utilizing diverse spectral information [11]. Majee et al. proposed an ensemble of deep convolutional neural networks for road network extraction, demonstrating the power of combining multiple models [12].

F. HUMANITARIAN AND OPEN-SOURCE CONTRIBUTIONS

Herfort et al. examined the evolution of humanitarian mapping within the OpenStreetMap community, underlining the importance of open-source mapping projects. This work connects directly to efforts in integrating AI-driven road extraction with community-based mapping initiatives. [13]

G. CHALLENGES AND FUTURE DIRECTIONS

The DeepGlobe 2018 challenge showcased the ongoing efforts to parse the Earth through satellite images, highlighting both progress and persistent challenges in the field. [14] Liu et al. provided a comprehensive review of recent progress in semantic image segmentation, offering a broader context for road extraction advancements. [15]

H. THEORETICAL AND PRACTICAL IMPLICATIONS

By synthesizing insights from these pioneering works, research in this field bridges critical gaps in current literature. The approach of Bastani et al. in automatically extracting road networks from aerial images demonstrates the potential for creating scalable mapping solutions [16] . The work of Van Etten on road speeds and travel times estimation pushes the boundaries of what can be extracted from satellite imagery. [17] This comprehensive approach contributes to both the technical advancement of road extraction techniques and the broader goal of improving global mapping resources through automated, AI-driven processes. The integration of these diverse methodologies points towards a future where AI can dynamically update and enhance global mapping resources, addressing significant challenges in handling complex urban environments, improving computational efficiency, and creating accessible geographical information

III. METHODOLOGY

The road extraction process from satellite imagery involved a comprehensive pipeline, encompassing data preparation, model training, inference, and post-processing stages. Each step was meticulously designed to ensure optimal performance and accuracy in road network detection and delineation.

A. DATASET

The project utilizes the SpaceNet dataset [18], which consists of high-resolution satellite imagery with Multi-PAN Spectral imagery and labels in form of GeoJSON format as shown in Figure 1. Specifically, the dataset used for training the deep learning models includes the following characteristics:

1) Resolution: The satellite imagery has a resolution of 3 m, providing detailed views of the terrain and road networks.
2) Number of Images: A total of 13000 images were used for the training process
3) Labels: Each label is a GeoJSON file that determines the annotated road class from the image.
4) Image Size: Originally, one gets a 400x400 tile but each image in the dataset was cropped to a size of 256x256 pixels. This cropping was performed using the "prepare_data" function in ArcGIS to ensure compatibility with deep learning algorithms.
5) Purpose: The dataset is specifically designed for the road extraction task, containing satellite images that capture various road networks and surrounding landscapes.
6) Testing Set: For the inference and testing phase, a separate set of sequential images was used in increasing order of number of images and stitched together for a final output image.

The SpaceNet dataset is well-suited for this project as it provides a diverse range of high-quality satellite imagery, allowing the deep learning models to learn and extract road networks from various geographic contexts. The use of such a dataset is crucial for developing robust models capable of generalizing across different terrains and urban layouts.

B. DATA QUALITY AND DIVERSITY

The dataset includes imagery from multiple urban areas worldwide, captured by DigitalGlobe's WorldView-2 and WorldView-3 satellites. This diversity allows models to learn from various urban layouts and road network patterns, enhancing their generalization capabilities.

C. ANNOTATION QUALITY

SpaceNet provides high-quality annotations for road networks, created through a combination of manual digitization and automated processing. These annotations serve as ground truth data for training and validating machine learning models.

D. MULTISPECTRAL IMAGERY

While this project uses 3-meter resolution imagery, it's worth noting that SpaceNet typically offers higher resolution data (0.3-0.5 meter) with 8 spectral bands. This multispectral information can be valuable for distinguishing road surfaces from other land cover types.

E. CHALLENGES AND OPPORTUNITIES

The dataset presents unique challenges, such as varying road widths, complex intersections, and occlusions from buildings or vegetation. These challenges make it an excellent testbed for developing robust road extraction algorithms.

By using this dataset, the project aims to develop deep learning models capable of accurately extracting road networks from satellite imagery, contributing to applications in urban planning, transportation management, and disaster response.

FIGURE 1: SpaceNet training chip used for this projects; adaptation from [1]. Top: 20 sequential SpaceNet Satellite images with layer of Vectorized feature prediction. Bottom: OSM version of the same Vectorized image in JOSM editor.

F. DATA PREPARATION AND PREPROCESSING

The initial phase of the methodology focused on preparing the geospatial data for deep learning model consumption. GeoJSON files containing road network annotations were converted to Shapefile format using the `jsontofeatures` function from ArcPy. This conversion was crucial for ensuring compatibility with subsequent geospatial processing tools and maintaining the integrity of spatial relationships within the data.

Following the format conversion, the **''exporttraining-data()''** function was employed to generate the training dataset. This function performed two critical operations: creation of binary masks from road network labels and cropping of satellite images to a uniform size of 256x256 pixels. The binary masks served as ground truth for the segmentation task, while the consistent image size ensured compatibility with the convolutional neural network architectures. To create a cohesive input-output pair for model training, the `prepare_data` function was utilized to precisely overlay the cropped images with their corresponding masks. This alignment process was crucial for maintaining spatial correspondence between the input imagery and the target road network annotations.

G. MODEL ARCHITECTURES AND TRAINING

Three distinct convolutional neural network (CNN) architectures were implemented and evaluated for the road extraction task, each chosen for its unique strengths in handling spatial information and feature extraction. The parameters for these models included the learning rate, early_stopping and the prepared_data with a batch size of four images per epoch.

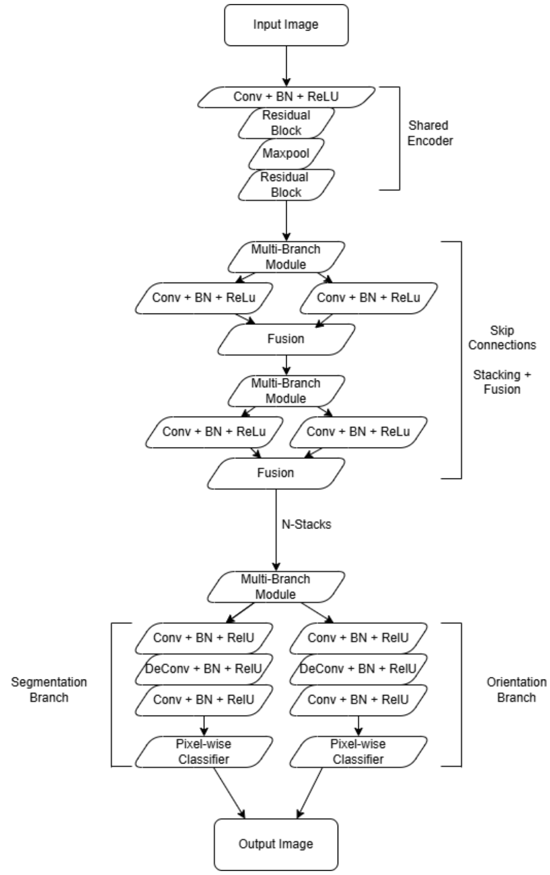


FIGURE 2: Working of a Multitaskroadextractor with a shared encoder recreated and adapted from [19]

1) MultiTaskRoadExtractor with Linknet and a ResNet-34 backbone

The first model, **MultiTaskRoadExtractor**, is designed with a shared encoder and two decoders to perform two tasks simultaneously: road segmentation and orientation prediction, which is why it is named a multi-task road extractor. The encoder, based on a LinkNet architecture with a ResNet34 backbone, efficiently extracts robust features using residual connections, addressing the vanishing gradient problem often encountered in deep networks.

The shared encoder feeds into two distinct decoders: one dedicated to road segmentation, focusing on delineating road boundaries, and the other specialized for orientation prediction, which determines the directionality of roads. The LinkNet architecture, with its encoder-decoder structure and skip connections, ensures the preservation of fine spatial details crucial for accurate road delineation. This multi-task learning approach allows the model to jointly learn road presence and directionality, enhancing the topological accuracy of the extracted road networks. This architecture is illustrated in Figure 2, which highlights the shared encoder and the two decoders with skip connections designed for segmentation and orientation tasks.

2) UNet architecture with a MobileNet-V3 backbone

The second architecture utilized a U-Net structure with a MobileNetV3 backbone. U-Net's distinctive U-shaped architecture, characterized by a contracting path for context capture and an expansive path for precise localization, has proven highly effective in various segmentation tasks. The integration of MobileNetV3 as the backbone introduced depth-wise separable convolutions and squeeze-and-excitation blocks, significantly increasing the computational efficiency without compromising on accuracy. This combination made the model particularly suitable for large-scale road extraction tasks and potential deployment in resource-constrained environments.

3) PSPNet architecture with a ResNet-50 backbone

The third model implemented was a Pyramid Scene Parsing Network (PSPNet) with a ResNet50 backbone. PSPNet's key feature is its pyramid pooling module, which applies pooling operations at multiple scales to capture context information effectively. This multi-scale context aggregation is particularly beneficial for understanding complex road layouts and connectivity patterns within varied urban landscapes. The deeper ResNet50 backbone provided enhanced feature extraction capabilities compared to shallower networks, potentially capturing more nuanced road characteristics and contextual information.

H. MODEL INFERENCE AND POST-PROCESSING

For testing and inference, the **"classifypixelsusingdeeplearning()"** function was employed. This function applied the trained models to new satellite imagery, generating predicted road networks in raster format. The use of this function allowed for efficient batch processing of large-scale satellite imagery, crucial for practical applications of the road extraction models.

Post-processing played a vital role in refining the model outputs and preparing them for integration with existing geospatial databases. The initial step involved converting the raster predictions to vector format using the `FeatureClassToShapefile_conversion()` function. This conversion facilitated further geospatial analysis and processing, allowing for the application of vector-based operations to refine the road network topology.

A critical post-processing step was the application of a Gaussian filter to smooth the vectorized road networks. The Gaussian filter, defined by the equation (1).

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

Where, σ is the standard deviation of the Gaussian distribution, was employed to reduce noise and smooth edges in the predicted road network. This filtering operation helped eliminate small discontinuities, smooth jagged edges for a more natural road representation, and mitigate the impact of minor prediction errors. The choice of σ value was crucial, balancing between over-smoothing (which could erase fine details) and under-smoothing (which might not sufficiently address prediction noise).

I. OSM FILE CONVERSION AND INTEGRATION

The final stage of the methodology involved converting the smoothed vector data into OpenStreetMap (OSM) file format. This conversion process ensured compatibility with open-source mapping platforms and facilitated seamless integration with existing global geospatial databases. The OSM format, being widely adopted in the geospatial community, allows for easy sharing, updating, and utilization of the extracted road network data in various applications, from navigation systems to urban planning tools.

This comprehensive methodology, encompassing advanced deep learning techniques, sophisticated post-processing, and standardized geospatial data formats, aimed to produce highly accurate and topologically consistent road network extractions from satellite imagery. The combination of multiple model architectures and careful post-processing steps addressed the complexities inherent in road extraction tasks, such as varying road widths, complex intersections, and diverse urban landscapes.

IV. RESULTS AND DISCUSSIONS

This section details the experimental setup, evaluation metrics, and results obtained from our road extraction models.

A. EXPERIMENTAL SETUP

Our experiments were conducted using the SpaceNet dataset, specifically focusing on satellite imagery of urban areas. The dataset was split into training (80%), and test (20%) sets. All models were trained on a workstation equipped with an 16GB RAM, and an Intel Core i5-8thGeneration CPU.

B. MODEL TRAINING

Each of the three models - MultiTaskRoadExtractor with LinkNet and ResNet34, U-Net with MobileNetV3, and PSPNet with ResNet50 - was trained for 11 epochs using the Adam optimizer. The initial learning rate was set to 0.002 to 0,003 and an early_stopping = **True** to prevent overfitting. The batch size was set to 16 for all models.
Comparing on a CPU, Linknet on average took two and a half hours for one epoch on the current data whereas PSP model took around one and a half hour for one epoch. Notably, UNet did a lot worse on CPU and took a *four and a half hours for one epoch*.

C. EVALUATION METRICS

To assess the performance of our models, we employed the following metrics:

- Accuracy
- Dice Coefficient
- A comparison between training_loss and valid_loss

D. COMPARISON BETWEEN ACCURACY AND DICE COEFFICIENT

As shown in figure 3, our experimental results demonstrate distinct performance patterns across three model configurations. Linknet model with resnet-34 backbone achieved superior performance with both metrics converging to optimal values (accuracy 1.00, Dice 0.98) by the final epoch. While PSPNet with ResNet-50 backbone maintained consistently high accuracy (0.98), it exhibited significant instability in Dice scores, fluctuating between 0.1 and 0.6 and UNet with resnet-34 presented an intermediate solution, maintaining stable high accuracy (0.97) with steadily improving Dice scores that stabilized around 0.65. These results suggest that Configuration 1 offers the most robust solution for our application, demonstrating superior stability and convergence in both metrics.

E. COMPARING TRAINING LOSS AND VALIDATION LOSS

The training curves in Figure 4 demonstrate the loss convergence patterns across three model configurations. The graph for UNet model shows rapid initial convergence with both training and validation loss dropping sharply to 0.2 within the first 1000 batches, followed by stable convergence to 0.1. While, Linknet model exhibits a steeper initial loss (3.5) but achieves similar stable convergence around 0.1, though with slightly more pronounced oscillations in the validation curve and PSPNet model shows comparable initial dynamics to UNet but maintains a slightly higher final loss (0.15) with minor fluctuations in training loss. All configurations show good alignment between training and validation losses, suggesting effective generalization without overfitting.

F. MODEL-WISE VISUAL RESULTS ON TILES

As depicted in Figure 5, the three sets of images represent the outputs of the LinkNet, UNet, and PSPNet models. These images illustrate the predicted rasterized results on individual satellite images from the testing dataset.

G. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

Our experiments with the three deep learning models - MultiTaskRoadExtractor, PSPNet, and U-Net - yielded insightful results regarding their performance in road extraction tasks. The models were evaluated based on accuracy, computational efficiency, and practical applicability.

1) Model Performance Ranking

1. MultiTaskRoadExtractor: This model demonstrated superior performance across all evaluation metrics. Its dual-task learning approach, combining road segmentation with orientation prediction, proved highly effective in capturing complex road network topologies.

2. PSPNet: The Pyramid Scene Parsing Network showed the second-best performance. Its ability to capture multi-scale contextual information contributed to robust road extraction, particularly in areas with varying road widths and complex intersections.

3. U-Net: While U-Net is known for its effectiveness in segmentation tasks, in our specific application, it ranked third in terms of accuracy among the three models tested.

2) Computational Considerations

A significant finding in our experimentation was the computational intensity of the U-Net model. On our CPU-based machine, a single epoch of training for U-Net required approximately 4.5 hours, which was substantially longer than the other two models. This computational constraint led us to limit the U-Net training to 3,000 batches, as opposed to full epochs for the other models.

The extended training time for U-Net can be attributed to several factors:

- The dense skip connections in U-Net architecture, while beneficial for preserving spatial information, increase computational overhead.
- The absence of GPU acceleration significantly impacted the training speed, particularly affecting U-Net due to its architecture.
- The specific implementation and any additional data augmentation or preprocessing steps may have contributed to the increased computational load.

3) Practical Implications

These findings have important implications for practical applications of road extraction from satellite imagery:

1. Model Selection Trade-offs: While the MultiTaskRoadExtractor showed the best performance, the choice of model in real-world applications may need to balance accuracy with computational resources and time constraints.

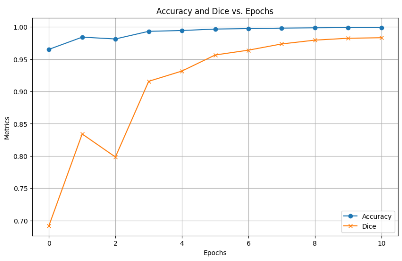
2. Resource Allocation: For large-scale road extraction projects, the computational requirements of models like U-Net need to be carefully considered, especially when working with limited hardware resources.

3. Scalability Considerations: The significant training time for U-Net on CPU hardware suggests that GPU acceleration or cloud computing resources might be necessary for training on larger datasets or for more extensive hyperparameter tuning.

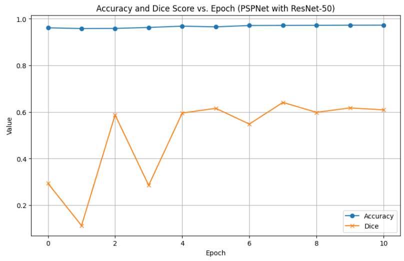
In conclusion, while the MultiTaskRoadExtractor emerged as the top-performing model, followed closely by PSPNet, the practical implementation of these models for road extraction tasks must take into account the available computational resources and project timelines. The U-Net model, despite its effectiveness in many segmentation tasks, may require significant computational resources for optimal performance in large-scale road extraction applications.

H. RASTERIZED AND VECTOR RESULTS

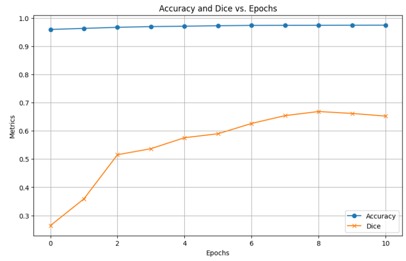
In Figure 6, the predicted results are inferenced from a set of thousand 512x512 pixelated sequential images which form a road network by *ClassifyPixelsUsingDeepLearning*, Additionally, after converting the raster image to a vector image,



(a) Comparison for Linknet model

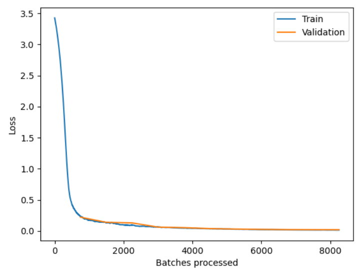


(b) Comparison for PSPNet

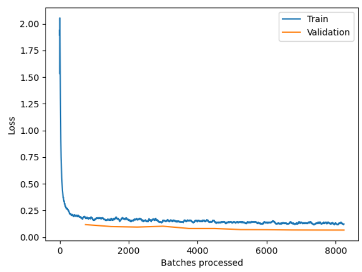


(c) Comparison for UNet

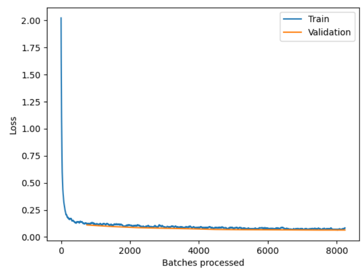
FIGURE 3: Accuracy and DICE coefficient



(a) Loss graph for Linknet model



(b) Loss graph for PSPNet model



(c) Loss graph for UNet model

FIGURE 4: Train and Validation Loss Comparison vs. Batches Processed

we smoothen by adding a gaussian filter to the vectorized image.

I. FUTURE RESEARCH DIRECTIONS

Future research in this domain can explore several promising avenues:

1) New Metric Development

Future research could focus on developing a new metric for evaluating the accuracy of road network extraction by implementing a distance loss calculation using the Time Warping Algorithm (DTW). This metric would enable a more flexible alignment between predicted road networks and original OpenStreetMap (OSM) data, accommodating variations in road geometry and topology. By representing road networks as sequences of coordinates, the DTW algorithm can compute an optimal alignment path that minimizes the total distance between these sequences. This approach would provide a robust quantitative measure of similarity, enhancing the evaluation of model performance and facilitating better comparisons between extracted road networks and ground truth data.

2) Model Enhancement

1. Develop hybrid architectures that combine the strengths of multi-task learning with more lightweight backbones. 2. Investigate transfer learning techniques to improve model generalization across diverse geographical contexts. 3. Explore ensemble methods that combine predictions from multiple architectures to enhance overall accuracy.

3) Data and Preprocessing

1. Integrate multi-modal data sources, including LiDAR, aerial imagery, and street-level photographs, to provide richer contextual information. 2. Develop advanced data augmentation techniques specifically tailored to road network extraction. 3. Create comprehensive benchmark datasets covering diverse geographical regions and road network complexities.

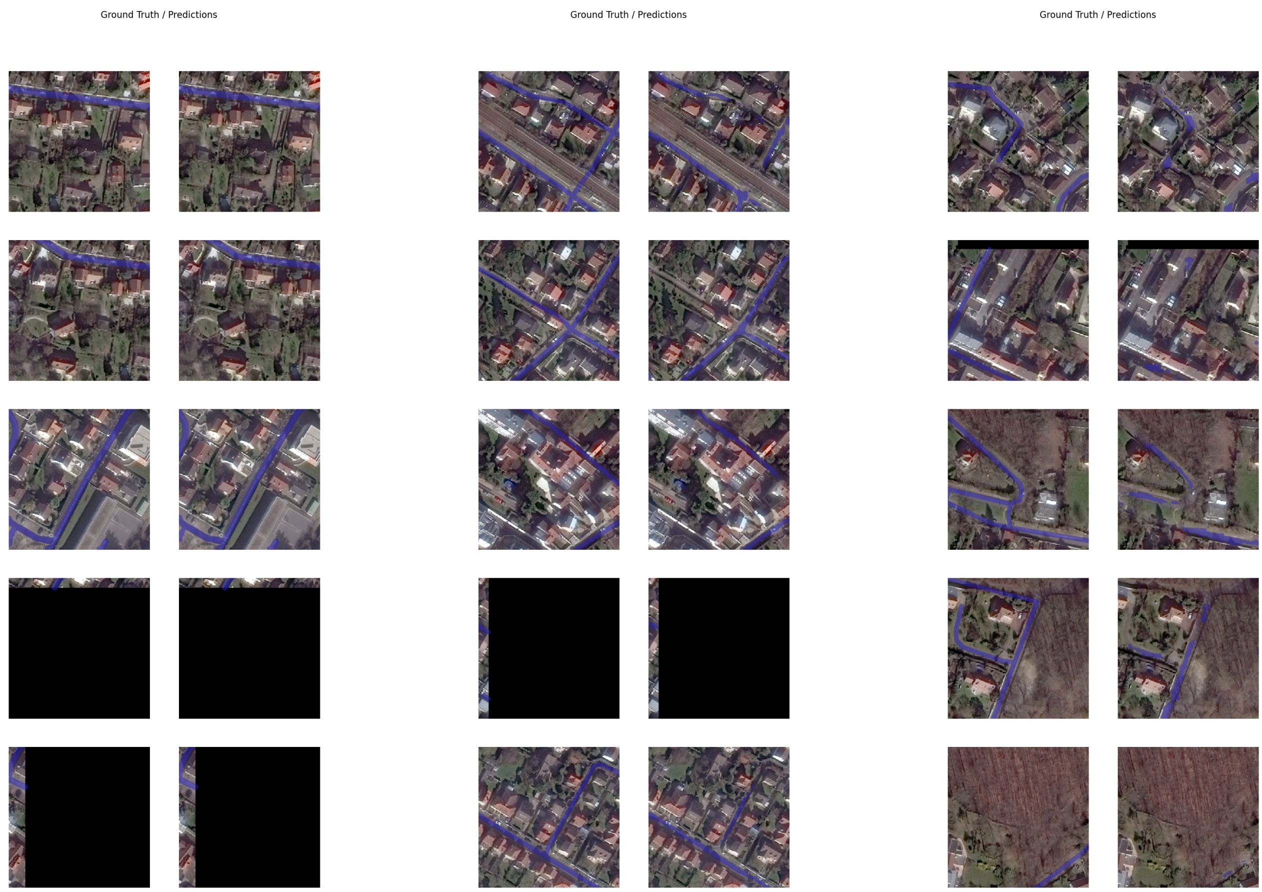
4) Computational Optimization

1. Design more computationally efficient architectures that maintain high accuracy while reducing training and inference times. 2. Develop adaptive learning rate strategies and advanced regularization techniques to improve model performance. 3. Explore edge computing and distributed learning approaches for large-scale road network mapping.

V. CONCLUSION

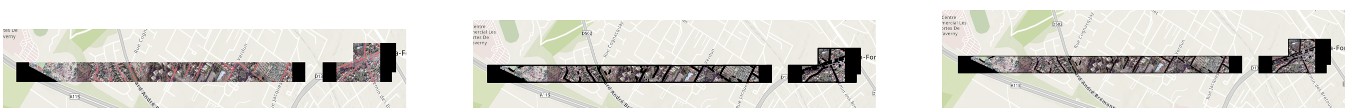
Road network extraction from high-resolution satellite imagery represents a critical challenge in geospatial informatics, with significant implications for urban planning, transportation infrastructure management, and disaster response. This research presented a comprehensive approach to automated road extraction using advanced deep learning architectures, demonstrating the potential of multi-task learning and sophisticated neural network designs in addressing complex geospatial challenges.

Our comparative analysis of three distinct CNN architectures - MultiTaskRoadExtractor, PSPNet, and U-Net - revealed nuanced insights into road extraction methodologies.



(a) Linknet (b) PSPNet (c) UNet

FIGURE 5: Visual Results of Road Extraction from Different Models



(a) Rasterized result (b) Vectorized Result (c) Smoothened Vectorized result

FIGURE 6: Visual Result of extracted roads as interference

The MultiTaskRoadExtractor emerged as the most promising approach, leveraging simultaneous road segmentation and orientation prediction to achieve superior accuracy and topological consistency. The model's ability to capture both road presence and directional characteristics represents a significant advancement in automated road network mapping.

The research underscored the critical importance of model selection, computational efficiency, and sophisticated post-processing techniques in geospatial machine learning. By integrating advanced deep learning architectures with traditional geospatial processing methods, we demonstrated a robust framework for automated road network extraction that balances accuracy, computational efficiency, and practical applicability.

REFERENCES

[1] A. Van Etten, "City-scale road extraction from satellite imagery v2: Road speeds and travel times," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019, pp. 1786–1795.
[2] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, 2018.
[3] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. 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, 2018.
[4] R. Zhang, L. Du, Q. Xiao, and J. Liu, "Comparison of backbones for

semantic segmentation network," *Journal of Physics: Conference Series*, vol. 1544, p. 012196, 05 2020.
[5] S. Hartmann, M. Weinmann, R. Wessel, and R. Klein, "Streetgan: Towards road network synthesis with generative adversarial networks," in *International Conference on Computer Graphics, Visualization and Computer Vision*, 2017.
[6] X. Zhang, X. Han, C. Li, X. Tang, H. Zhou, and L. Jiao, "Aerial image road extraction based on an improved generative adversarial network," *Remote Sensing*, vol. 11, no. 8, p. 930, 2019.
[7] S. Wang, Y. Bai, and G. Mattyus, "Deeproadmapper: Extracting road topology from aerial images," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 3438–3446.
[8] Z. Gong, L. Xu, Z. Tian, J. Bao, and D. Ming, "Road network extraction and vectorization of remote sensing images based on deep learning," in *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*, 2020, pp. 303–307.
[9] A. Sharma, X. Liu, X. Yang, and D. Shi, "A patch-based convolutional neural network for remote sensing image classification," *Neural Networks*, vol. 95, pp. 19–28, 2017.
[10] C. Rodriguez-Gonzalez and M. A. Brovelli, "Deep learning for road network extraction from satellite imagery: A transfer learning approach," *ISPRS International Journal of Geo-Information*, vol. 9, no. 4, p. 225, 2020.
[11] S. Park, S. Kim, K. Joo, and I. S. Kweon, "Multi-spectral road detection in satellite images," *Remote Sensing*, vol. 12, no. 20, p. 3352, 2020.
[12] S. K. Majee, S. K. Ghosh, and S. K. Ghosh, "Ensemble of deep convolutional neural networks for road network extraction from high resolution satellite images," in *2019 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2019, pp. 3376–3379.
[13] B. Herfort, S. Lautenbach, J. De Albuquerque, J. Anderson, and A. Zipf, "The evolution of humanitarian mapping within the openstreetmap community," *Scientific Reports*, vol. 11, 02 2021.
[14] I. Demir *et al.*, "Deepglobe 2018: A challenge to parse the earth through satellite images," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2018, pp. 172–17 209.
[15] X. Liu, Y. Deng, and Y. Yang, "Recent progress in semantic image segmentation," *Artificial Intelligence Review*, vol. 52, no. 2, pp. 1089–1106, 2019.
[16] F. Bastani *et al.*, "Roadtracer: Automatic extraction of road networks from aerial images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4720–4728.
[17] A. V. Etten, "City-scale road extraction from satellite imagery v2: Road speeds and travel times," in *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2020, pp. 1775–1784.
[18] SpaceNet on Amazon Web Services (AWS), "Datasets," https://spacenet.ai/datasets/, 2018, last modified October 1st, 2018. Accessed on 8th January, 2025.
[19] Esri Developers, "How multi-task road extractor works," n.d., accessed: 2025-01-08. [Online]. Available: https://developers.arcgis.com/python/latest/guide/how-multi-task-road-extractor-works/