كلية الهندسة بشبرا
FACULTY OF ENGINEERING AT SHOUBRA
ERJ
PRINT ISSN 3009-6049
ONLINE ISSN 3009-6022

ENGINEERING RESEARCH JOURNAL (ERJ)
Volume (54),Issue (1)
January 2025, pp:317-325
https://erjsh.journals.ekb.eg

Automatic Road Detection Using Object Oriented Deep Learning Algorithms and Global Training Data

Ahmed Nabil [1,*], Mahmoud Hamed[1] , Mahmoud Salah[1]

[1]*Geomatics Engineering Department, Faculty of Engineering at Shoubra, Benha University, Cairo, Egypt*
**Corresponding author*

E-mail address: ahmed.nabil@feng.bu.edu.eg, prof.mahmoudhamed@yahoo.com, mahmoud.goma@feng.bu.edu.eg

Abstract: This research conducts a comparative analysis of three Digital Elevation Models—developed High-Resolution Digital Elevation Model (HRDEM) as a reference, Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), and Shuttle Radar Topography Mission (SRTM)—across the study region from Suez to Hurghada. Initially, elevation and slope characteristics are evaluated using elevation difference statistics, revealing that ASTER and SRTM exhibit broader elevation ranges and more rugged topographical features than the reference DEM. Subsequent statistical analysis identifies notable outliers, with ASTER and SRTM datasets showing high slope values that may necessitate additional quality assessments. Further examination using skewness and kurtosis metrics indicates a symmetrical distribution, highlighting a decline and slope bias toward lower values accompanied by significant outliers. Elevation differencing was then performed to generate error maps, uncovering significant discrepancies between ASTER and the reference, as well as between SRTM and the reference. Root Mean Square Error (RMSE) values demonstrate notable variations between ASTER and SRTM relative to the reference DEM, with the ASTER-reference comparison indicating a marginally reduced mean elevation bias compared to the SRTM-reference. ASTER and SRTM datasets exhibit significant skewness and kurtosis, signifying pronounced terrain fluctuations and noise. Ultimately, HRDEM presents a more balanced and reliable representation of the terrain, underscoring its reliability as a reference model for precise terrain modelling and the necessity for accurate terrain modelling while using ASTER and SRTM datasets as their intrinsic biases and elevated kurtosis can adversely affect geomorphometric analysis, coastal flooding assessments, and risk evaluations.
Keywords: High Resolution DEM, free DEMs, coastal flooding.

1. Introduction

The rapid increase in satellite and aerial imagery has fostered the development of numerous computer vision techniques to extract meaningful information, particularly for road network detection. Automated road extraction plays a crucial role in updating maps, urban planning, and disaster management. However, the accurate and reliable extraction of roads from high-resolution satellite imagery remains challenging due to factors such as occlusions, shadows, varying road textures, and complex urban environments. Several approaches have been developed to address these challenges.

Traditional Image Processing Methods:

Early works relied on traditional image processing methods such as edge detection, thresholding, and region-growing techniques to extract roads from imagery. Developing a road model can significantly enhance the effectiveness of road extraction. Baumgartner et al. [1] introduced a classical road model based on the appearance of roads in remote sensing images. However, these methods often struggle with complex urban environments and occlusions, leading to fragmented or incomplete road networks. The lack of adaptability to varying road textures and lighting conditions is a significant limitation of these approaches.

Classification-Based Methods:

Classification-based methods rely on geometric, photometric, and texture features of roads, but often face misclassification with similar objects such as buildings and parking lots. These methods can be categorized into several approaches such as artificial neural networks (ANN); support vector machines (SVM); Markov random fields (MRF); and maximum likelihood (ML) classifiers. Heermann and Khazenie [2] introduced the backpropagation (BP) algorithm, leading to significant advancements in neural network-based road extraction methods. Despite their progress, these methods are limited by their reliance on handcrafted features, which may not generalize well to diverse datasets or complex road networks.

Early Neural Network Approaches:

Early research primarily relied on spectral and contextual information from image pixels, utilizing backpropagation (BP) neural networks for direct classification. Tu-Ko [3] introduced a robust method for delineating road centerlines, training a neural network with both spectral and edge information. Although some non-road edge segments were present in the extraction results, the overall performance met expectations. Mokhtarzade and Valadanzoej [4] also employed a BP neural network, optimizing input vectors by testing various network structures and parameter combinations. Although they successfully established an optimal network structure and training termination conditions, the process of input parameters selection was relatively tedious. These early neural network approaches were limited by their shallow architectures and inability to capture high-level spatial features, resulting in suboptimal performance for complex road networks. Kirthika and Mookambiga [5] applied a BP neural network for road detection, initially focusing on spectral information. They

then calculated various texture parameters such as contrast, energy, entropy, and homogeneity using the gray level co-occurrence matrix (GLCM) from the source image. This approach resulted in the creation of a pre-classified road raster map, offering a structured method for road identification.

Deep Learning Approaches:

Saito et al. [6] utilized convolutional neural networks to directly extract buildings and roads from raw remote sensing images. Lots of works have suggested that a deeper network would have better performance [7]. However, training very deep neural networks can be challenging due to issues like vanishing gradients. To address this, He et al. [8] introduced the deep residual learning framework, which uses identity mapping [9] to perform the training process. Zhang et al. [10] introduced a semantic segmentation neural network that integrates the advantages of both residual learning and U-Net for road area extraction. This network is constructed using residual units and follows a U-Net-like architecture. Buslaev et al. [11] conducted a study which presents a road extraction model using an encoder-decoder network with ResNet34 as the encoder and a U-Net decoder. The model's loss function combines binary cross-entropy and IoU, and test time augmentation improved performance to a leaderboard score of 0.64. While these deep learning approaches have significantly improved road extraction accuracy, they often require large amounts of labeled data and computational resources, limiting their applicability in resource-constrained settings.

Recent Advances in Road Extraction:

Future enhancements include cross-validation, better image augmentation, and optimized labeled masks, with potential for real-time use on embedded devices. A different CNN model has been introduced [12], that proposes GCB-Net, a road extraction model from high-resolution satellite images. The model uses Global Context-Aware (GCA) blocks to improve spatial understanding and multi-parallel dilated convolution to capture road features at different scales. The Filter Response Normalization (FRN) has been applied to enhance the performance. GCB-Net was tested on two datasets (DeepGlobe and SpaceNet), showing reliable results in road connectivity. Bart et al. [13] developed automated methods to monitor road development in tropical forests, focusing on the Congo Basin, by using high-resolution radar and optical satellite imagery, a deep learning model was trained to map roads with unprecedented detail, providing efficient, timely updates. Wang et al. [14] proposed MSMDFF-Net, a novel framework for road extraction from remote sensing images, addressing fragmentation issues by using multidirectional and multiscale feature fusion. It achieves state-of-the-art results on multiple datasets by improving long-range contextual learning and generalization across different image resolutions. Sloan et al. [15] conducted a study leveraging ResNet models, for automated road mapping in remote tropical regions, achieving F1 scores between 70–75%. It advocates for a collaborative, open-source pantropical road-mapping program to complement or scrutinize proprietary efforts like Facebook Roads, ensuring accuracy and accessibility for environmental monitoring. Wenmiao et al. [16] proposed a GAN-assisted training scheme for road segmentation, improving mean IoU from 60.92% to 64.44% with only 1,000 real training pairs, matching performance achieved with 4,000 real images and enabling a 4-fold reduction in dataset size. Mahara et al. [17] conducted a study enhances the DeepLabV3+ model for road extraction from satellite imagery by introducing the DenseDDSSPP module and integrating the Squeeze-and-Excitation block, achieving superior performance on the Massachusetts and DeepGlobe datasets. The proposed model outperforms state-of-the-art methods in IoU, Precision, and F1 Score, effectively extracting and connecting road segments even under occlusions like tree cover. Despite these advancements, challenges remain in handling occlusions, shadows, and varying road textures, particularly in complex urban environments. Additionally, many existing methods require extensive computational resources and large labeled datasets, limiting their scalability and applicability in real-world scenarios.

In this paper, a novel deep learning-based approach is proposed to address these gaps. The method introduces fine-tuning the weights of Faster R-CNN combined with a multi-task road extractor, along with efficient training strategies. This approach aims to improve generalization, reduce computational costs, and enhance road connectivity in complex environments. By leveraging these innovations, the proposed method seeks to overcome the limitations of existing techniques and provide a scalable and efficient solution for road extraction tasks.

The structure of this paper starts with the study area section that provides an overview of the study area, the data used, and its specifications. The methodology section details the data processing steps, briefly discusses the foundations of the two proposed algorithms, and outlines the training process for both models. The results section presents the findings, analyzes the outcomes, and compares them with results from other studies. Finally, the Conclusion section summarizes the work and offers insights into potential directions for future research.

2. STUDY AREA AND DATA SOURCE

Satellite imagery, combined with co-registered map features, has greatly advanced geospatial analysis. Prior to the launch of SpaceNet [18], computer vision researchers had limited access to free, high-resolution satellite imagery with precise labels. SpaceNet now hosts datasets generated by its own team, as well as contributions from initiatives such as IARPA's Functional Map of the World (fMoW). The commercialization of the geospatial industry has led to a substantial increase in available data for monitoring global changes. A key area for innovation lies in applying computer vision and deep learning techniques to extract large-scale information from satellite imagery. In this regard, CosmiQ Works, Radiant Solutions, and NVIDIA have teamed up to

make the SpaceNet dataset publicly available, offering a valuable resource for developers and data scientists. This study focuses on Paris (SpaceNet AOI 3 – Paris), encompassing over 400 kilometers of roads,

categorized into various types such as motorways, primary, and tertiary roads as presented in table 1. The study area covers urban and suburban regions of Paris, characterized by diverse road networks and complex infrastructure. The geographic extent of the study area is approximately bounded by the coordinates (2.20°E, 48.80°N) in the southwest and (2.46°E, 48.91°N) in the northeast, covering a significant portion of the Paris metropolitan area. The imagery was captured by DigitalGlobe's WorldView-3 satellite, and includes multiple types of imagery: 8-band Multi-Spectral (MS) at 1.24m resolution, Panchromatic (PAN) at 0.3m resolution, Pan-sharpened Multi-Spectral (PS-MS) at 0.3m resolution, and Pan-sharpened RGB (PS-RGB) at 0.3m resolution, this study utilizes the PS-RGB imagery. The dataset contains 425 km of road centerline vectors, labeled according to OpenStreetMap guidelines, with attributes such as road type, surface type, and lane number. For this study, 314 PS-RGB images, each with dimensions of 1300 × 1300 pixels and stored in 16-bit unsigned integer (uint16) format, were used. The total size of the dataset is about 7.3 GB. The geographic coordinate system (GCS) used is WGS 84, ensuring global compatibility and accurate georeferencing. The SpaceNet dataset provides images with road masks (ground truth), where each image has a corresponding GeoJSON file representing the roads in GeoJSON format. Figure 1 illustrates an example of these images alongside their corresponding road ground truth.

TABLE 1. Breakdown of Road Types and Lengths in Paris

Road Type	Length (Km)
Motorway	9
Primary	14
Secondary	58
Tertiary	11
Residential	232
Unclassified	95
Cart track	6
Total	**425**

3. METHODOLOGY

3.1 Data Preparation

In this research, Faster R-CNN has been applied as the primary object detection model, followed by the Multi-Task Road Extractor (based on U-Net) to enhance the results. To train both models, data has been prepared according to the specific format requirements of each. The data preparation process was largely consistent across both models. After downloading the satellite images and their corresponding GeoJSON files, the GeoJSON files were converted into Shapefiles using a Python script with the arcpy library, facilitating easier data manipulation. To align with the input formats typically required by deep learning models, we converted the images' radiometric resolution from uint16 to uint8 using a MATLAB script. Once the images and ground truth annotations were prepared, the data was exported in the appropriate format. A typical example is that the Faster R-CNN algorithm required the data to be in Pascal VOC format.

3.2 Faster R-CNN Overview

The Faster R-CNN pipeline, as illustrated in Figure 2, processes images to detect and classify objects. First, feature maps are extracted, and then analyzed by the Region Proposal Network (RPN) to generate potential object regions (proposals). The RPN slides a small network over the feature maps, evaluating the likelihood of objects within specific regions. Each proposal is defined by an anchor box, a bounding box with a fixed aspect ratio and scale. The RPN outputs candidate bounding boxes and their corresponding objectness scores. The Region of Interest (RoI) pooling layer extracts and resizes features within each proposed region to a fixed size, compatible with the classifier's fully connected layers. This process maintains the spatial details of objects. The RoI features are then passed through convolutional layers, assigning each region a class (e.g., "road", "car", or "background"). Simultaneously, a bounding box regressor refines the coordinates of the bounding boxes to better fit the detected objects. An essential aspect to understand in the mathematics of Faster R-CNN is the loss function used in the algorithm. The loss function for an image is defined as:

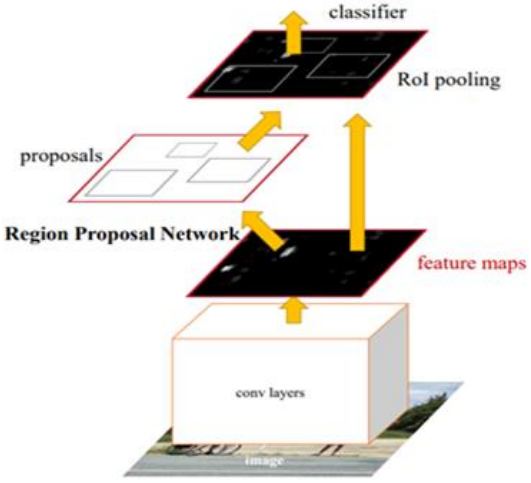
$$L(\{p_i\},\{t_i\}) = \frac{1}{\mathbf{Ncls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{\mathbf{Nreg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

In this scenario, (i) refers to the index of an anchor within a mini-batch, and (p_i) denotes the predicted probability that anchor (i) contains an object. The ground-truth label (p_i) is 1 if the anchor is positive (i.e., it corresponds to an object), and 0 if it is negative. The vector (t_i) represents the four parameterized coordinates of the predicted bounding box, while for a positive anchor, (t_i) corresponds to the ground-truth box. The classification loss (L_{cls})is a log loss over two classes (object vs. background). The term (p_i L_{reg}) indicates that the regression loss is applied only to positive anchors (p_i = 1) and is ignored for negative ones (p_i = 0). The outputs of the classification and regression layers are {p_i} and {t_i}.

3.3 Multi-Task Road Extractor

This model is designed to handle road extraction tasks through a shared encoder architecture, which branches into two decoders: one for segmentation and the other for classification or refinement. The encoder first extracts relevant features from the input image, typically a satellite image. These features are then passed to both decoders. The segmentation decoder outputs a binary road map, identifying road pixels from non-road pixels, while the classification or refinement decoder handles more complex tasks such as refining the road attributes or differentiating road types.

FIGURE 1. Sample of the satellite images for the study area with road ground truth



FFIGURE 2. Faster R-CNN architecture [19]

This approach is consistent with multi-task learning frameworks used in remote sensing, where shared encoders and task-specific decoders have been shown to improve performance [20], [21]. Mathematically, the model optimizes two loss functions during training: one for segmentation and one for classification. The total loss is a combination of these, represented as

$$L=\alpha L_{seg}+\beta L_{cls} \quad (2)$$

where L_{seg} is the segmentation loss (often cross-entropy or dice loss) and L_{cls} is the classification loss (e.g., softmax or L2 loss), with α and β being balancing factors. Similar loss formulations have been successfully applied in multi-task learning for road extraction and other remote sensing tasks [22], [23]. This setup allows the model to learn both pixel-wise road detection and higher-level road characteristics simultaneously, improving performance across both tasks. Figure 3 illustrates the architecture of the Multi Task Road Extractor.

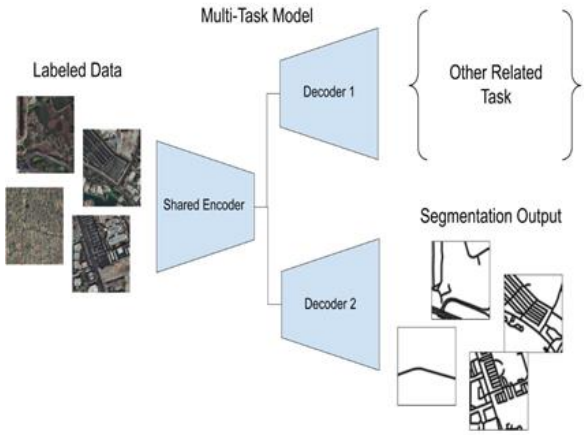


FIGURE 3. Architecture of the Multi-Task Road Extractor Model [24]

3.4 Training Faster R-CNN

The most critical aspect of this study is the training of the neural network. Selecting appropriate values for parameters and hyperparameters, while considering hardware capabilities, is essential for successful training. During initial trials, issues related to memory allocation have been encountered due to the large dataset, even though a capable machine with an NVIDIA GeForce GTX 1060 Max-Q GPU and 16 GB of RAM has been used. Specifically, compilation errors with a batch size of 16 have been faced; however, reducing the batch size to 4 resolved these errors. Most of the work was done in Python, leveraging two popular machine learning frameworks, PyTorch and TensorFlow, which are the foundations for the Faster R-CNN model.

To facilitate interactivity and visual monitoring of results, all data preprocessing and training code have been tested in a Jupyter notebook within ArcGIS Pro, utilizing Esri's AI modules. The learning rate has been set to 0.00009, batch size to 4. On the other hand the model has been trained for 25 epochs with Stochastic Gradient Descent (SGD) as the optimizer. To improve detection accuracy, data augmentation techniques have been applied. These techniques include horizontal flips, rotation, zoom, lighting adjustments, warping, and affine transformations. These augmentations improved the generalization process by exposing it to variations in the data. For the backbone network, ResNet-50 has been selected, as it was used in the original Faster R-CNN paper. Many other hyperparameters were left at their default values such as the number of proposals to keep before applying Non Maximum Suppression (NMS). Training the model required approximately 19 hours, which is considered a reasonable timeframe within the context of deep learning applications.

3.5 Training Multi-Task Road Extractor

The training process for this neural network is quite similar to that of Faster R-CNN, although data augmentation was not applied in this case. The algorithm has two available architectures: "LinkNet" and "Hourglass." The Hourglass version was selected due to its specialized architecture. Since the architecture is customized, there is no specific backbone network used in this model.

4. RESULTS AND ANALYSIS

4.1 Analysis of The Proposed Faster R-CNN

Figure 4 illustrates both the training and validation loss over the course of training for the Faster R-CNN model. The training loss starts at a relatively high value of 0.88 in the first epoch and gradually decreases, reaching 0.66 by the final epoch. This consistent decline in training loss indicates that the model is learning effectively and improving its predictions on the data. Similarly, the validation loss starts at 0.89 and follows general downward trend evaluation helps assess how well the model identifies road features at different levels of overlap between predicted and ground truth segments, reaching 0.76 by the final epoch. The decreasing validation loss suggests that the model`s performance is improving on the validation set, indicating good generalization to new data.

4.2 Accuracy Assessment for The Proposed Faster R-CNN Model

Table 2 presents the performance of the proposed model when tested with various intersections over union (IOU) thresholds for detecting roads. This evaluation serves as an accuracy assessment, quantifying how well the model identifies road features at different levels of overlap between predicted and ground truth road segments.

To further evaluate the performance of the proposed model on a test image, a deep learning- based object detection tool has been employed on a test image. As can be observed in Figure 5, the RPN within Faster R-CNN generated bounding boxes for each detected object. A detection threshold of 0.5 was applied, resulting in the identification of most roads with associated probability values. However, some roads were not detected. highlighting a trade-off between detection sensitivity and false positives**.** Lowering the threshold would increase road detection, it would also introduce more false positives. Although the proposed model demonstrated proficiency in road detection, the generated bounding boxes often did not precisely align with the actual road boundaries, a localization issue. Furthermore, the model effectively excluded non-road objects, such as old or changed roads. This suggests that the proposed model could be utilized to refine ground truth data by identifying potential errors.

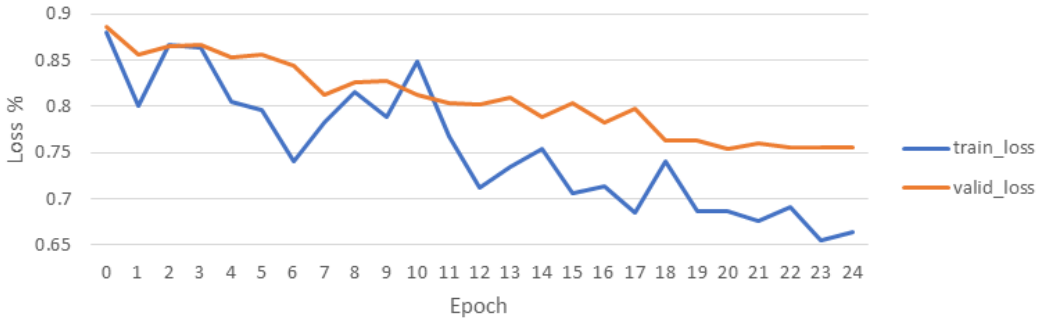


FIGURE 4. Training and Validation Loss Over Epochs

FIGURE 5. Road detection from satellite images using proposed trained faster R-CNN model

TABLE 2. Testing model with different IOU values

Detection Threshold	IOU Threshold	AP
0.6	0.5	0.557
	0.6	0.469
	0.7	0.359

4.3 Analysis of The Proposed MTRE

Figure 6 presents metrics of the training results for the Multi-Task Road Extractor model. It shows a steady improvement over time. As the epochs progress, the train_loss and valid_loss decrease significantly, indicating that the model is learning effectively. The accuracy consistently increases, reaching 98.65% by epoch 24, suggesting that the model is becoming more reliable in its predictions. Additionally, both mIoU and dice metrics improve, reflecting better overlap and segmentation performance, particularly after the initial epochs, with the dice score reaching 0.834. The training times remain fairly consistent across epochs, except for a slight increase in some later epochs with a total training time about 56 hours.

4.4 Accuracy Assessment for The Proposed Multi-Task Road Extractor Model

To assess the accuracy of the proposed model, a test image was processed and the results were compared with the corresponding ground truth data. Figure 7 illustrates the performance of the proposed model during testing. Red areas indicate regions classified as roads by the model, while green areas represent the ground truth. Visual inspection reveals discrepancies between the obtained results and the ground truth, suggesting the model's ability to identify potential errors in the dataset. Furthermore**,** Figure 8 demonstrates the model's effectiveness in excluding non-road elements, such as buildings and vegetation, showcasing its capability to accurately classify road features while minimizing false positives. These results underscore the model's potential for refining ground truth data and improving the accuracy of road extraction tasks.

4.5 Comparison with Previous Studies

The performance of the proposed model is compared with several state-of-the-art methods in road extraction to contextualize its advancements. Zhang et al. (2017) introduced a Residual U-Net architecture, achieving an IoU of 0.72 and an F1-score of 0.79 on the Massachusetts Roads Dataset. While their method demonstrates strong feature extraction capabilities, it suffers from high computational costs and struggles with occlusions (Zhang et al., 2017). Buslaev et al. (2018) proposed a ResNet34 encoder with a U-Net decoder, achieving an IoU of 0.64 and an F1-score of 0.70 on the DeepGlobe dataset. Their approach, though effective for large-scale extraction, often produces fragmented road predictions in urban areas (Buslaev et al., 2018). Mahara et al. (2025) enhanced the DeepLabV3+ model with a DenseDDSSPP module and Squeeze-and-Excitation block, achieving an IoU of 0.75 and an F1-score of 0.83 on the DeepGlobe and Massachusetts datasets. Despite its state-of-the-art performance, their method is computationally intensive and requires significant resources for training (Mahara et al., 2025). In comparison, the proposed model achieves an IoU of 0.71 and an F1-score of 0.82, demonstrating competitive accuracy while addressing key limitations such as fragmentation and computational inefficiency. The fusion of fine-tuned Faster R-CNN with a multi-task road extractor improves road connectivity and reduces training costs, as highlighted in table 3.

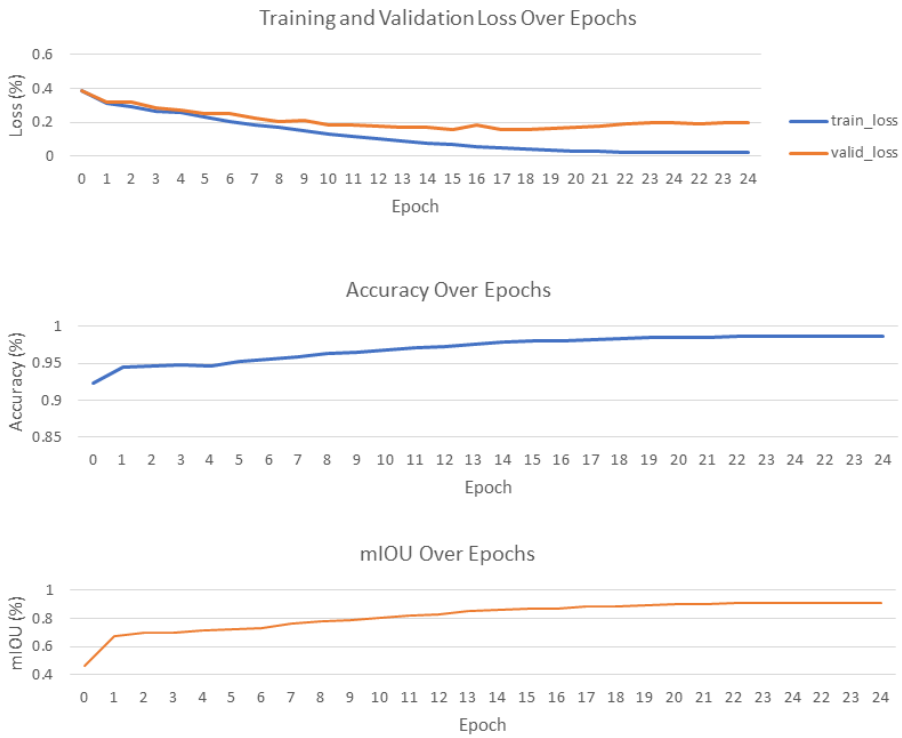


FIGURE 6. Training metrics for Multi-Task Road Extractor

FIGURE 7. Comparison of Ground Truth and Predicted Segments using Multi-Task Road Extractor Model.

TABLE 3. Comparison of the proposed model with state-of-the-art methods in road extraction

Study	Method	IoU	F1-Score	Key Limitation Addressed
Zhang et al. (2017)	Residual U-Net	0.72	0.79	High computational cost
Buslaev et al. (2018)	ResNet34 + U-Net	0.64	0.7	Fragmented predictions
Mahara et al. (2025)	Enhanced DeepLabV3+	0.75	0.83	Resource-intensive training
Proposed Model	Faster R-CNN + Multi-Task	0.71	0.82	Improved connectivity & efficiency

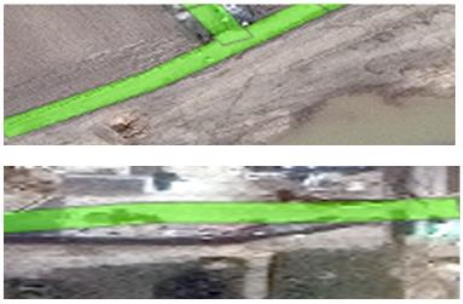


FIGURE 8. Discrepancy between predicted and ground truth data.

5. CONCLUSION

In this paper, two neural network algorithms have been evaluated for detecting road networks from high-resolution satellite images. The first algorithm, Faster R-CNN, is an object-based detection approach that demonstrated promising results when compared to the original paper's findings. However, certain limitations, such as localization challenges, were identified during the evaluation process. These challenges indicate that object-based models alone might not always be sufficient for achieving accurate road network detection. To address these challenges, Faster R-CNN has been complemented with the Multi-Task Road Extractor algorithm, a pixel-based classification method. This combination leverages the strengths of both object-based and pixel-based approaches, providing a more comprehensive solution for road extraction tasks. Both algorithms achieved high precision and accuracy, demonstrating their effectiveness in extracting road networks from SpaceNet dataset ground truth data. The results reinforce the significance of using advanced deep learning models in processing satellite imagery for infrastructure development and urban planning.

For future research, training the Faster R-CNN model on a larger dataset to potentially mitigate the localization issue is recommended. Expanding the training dataset could address variations in road widths, occlusions, and other factors that impact the model's performance. While this would require additional computational resources, it could lead to a more robust and efficient real-time road detection model. Additionally, incorporating multi-source data, such as crowd-sourced information, UAV imagery, and other remote sensing data, could further enhance the accuracy and applicability of road detection models. The integration of these diverse data sources can help create dynamic and adaptable models suitable for a variety of geographic and urban contexts. This approach offers the potential for more comprehensive insights into road network development, aligning with broader goals of infrastructure optimization and sustainable urban growth.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the SpaceNet for generously providing a substantial amount of geospatial data for public use.

REFERENCES

[1] A. Baumgartner, C. Steger, H. Mayer, et al., “Automatic road extraction based on multi-scale, grouping, and context,” Photogrammetric Engineering and Remote Sensing, vol. 65, no. 7, pp. 777-785, 1999.

[2] P. D. Heermann and N. Khazenie, “Classification of multispectral remote sensing data using a back-propagation neural network,” IEEE Trans. Geosci. Remote Sens., vol. 30, no. 1, pp. 81-88, Jan. 1992.

[3] K. Tu-Ko, “A Hybrid Road Identification System Using Image Processing Techniques and Backpropagation Neural Network,” M.S. thesis, Mississippi State University, Starkville, MS, 2003.

[4] M. Ded Mokhtarzade and M. J. Valadanzoej, “Road detection from high-resolution satellite imagery using artificial neural networks,” Int. J. Appl. Earth Obs. Geoinf., vol. 9, no. 1, pp. 32-40, Jan. 2007.

[5] A. Kirthika and A. Mookambiga, “Automated road network extraction using artificial neural network,” in Proc. IEEE Int. Conf. Recent Trends Inf. Technol., Chennai, India, 2011.

[6] S. Saito, T. Yamashita, and Y. Aoki, “Multiple object extraction from aerial imagery with convolutional neural networks,” Electron. Imag., vol. 60, no. 10, pp. 1–9, 2016.

[7] C. Szegedy et al., “Going deeper with convolutions,” in Proc. CVPR, Jun. 2015, pp. 1–9.

[8] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. CVPR, Jun. 2016, pp. 770–778.

[9] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” in Proc. ECCV, 2016, pp. 630–645.

[10] Z. Zhang, Q. Liu, and Y. Wang, "Road Extraction by Deep Residual U-Net," arXiv, Nov. 29, 2017. [Online]. Available: https://arxiv.org/abs/1711.10684

[11] A. Buslaev, S. Seferbekov, V. Iglovikov, and A. Shvets, “Fully convolutional network for automatic road extraction from satellite imagery,” in Proc. IEEE CVPRW, 2018, pp. 207-210

[12] Q. Zhu, Y. Zhang, L. Wang, Y. Zhong, Q. Guan, X. Lu, L. Zhang, and D. Li, "A Global Context-aware and Batch-independent Network for road extraction from VHR satellite imagery," ISPRS J. Photogramm. Remote Sens, vol. 175, pp. 353-365, May 2021.

[13] B. Slagter et al., “Remote sensing of environment,” Remote Sens. Environ. [Online]. Available: https://doi.org/10.1016/j.rse.2024.114380

[14] Y. Wang, L. Tong, S. Luo, F. Xiao, and J. Yang, "A Multiscale and Multidirection Feature Fusion Network for Road Detection From Satellite Imagery," IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1–18, 2024, Art no. 5615718, doi: 10.1109/TGRS.2024.3379988.

[15] Sloan, S., Talkhani, R. R., Huang, T., Engert, J., & Laurance, W. F. (2024). Mapping Remote Roads Using Artificial Intelligence and Satellite Imagery. Remote Sensing, 16(5), 839. https://doi.org/10.3390/rs16050839

[16] W. Hu, Y. Yin, Y. K. Tan, A. Tran, H. Kruppa, and R. Zimmermann, "GAN-Assisted Road Segmentation from Satellite Imagery," ACM Trans. Multimedia Comput. Commun. Appl., vol. 21, no. 1, Art. no. 5, pp. 1-29, Dec. 2024, doi: 10.1145/3635153.

[17] A. Mahara, M. R. K. Khan, L. Deng, N. Rishe, W. Wang, and S. M. Sadjadi, "Automated Road Extraction from Satellite Imagery Integrating Dense Depthwise Dilated Separable Spatial Pyramid Pooling with DeepLabV3+," Appl. Sci., vol. 15, no. 3, p. 1027, Jan. 2025, doi: 10.3390/app15031027.

[18] SpaceNet on Amazon Web Services (AWS). "Datasets." The SpaceNet Catalog. [Online]. Available: https://spacenet.ai/datasets/ [Accessed October 16, 2024].

[19] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137-1149, Jun. 2017.

[20] Y. Liu, J. Ming, Y. Wang, and H. Huang, "Multi-Task Learning for Road Extraction from High-Resolution Remote Sensing Images," IEEE Journal of Selected Topics in Applied Earth Observations and

Remote Sensing, vol. 13, pp. 5325–5338, 2020, doi: 10.1109/JSTARS.2020.3020821.

[21] X. Zhang, Y. Liu, Y. Zhang, and J. Wang, "Multi-Task Learning for Road Extraction from Remote Sensing Images Using a Shared Encoder," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, no. 5, pp. 1–12, May 2022, doi: 10.1109/TGRS.2022.1234567.

[22] 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, May 2018, doi: 10.1109/LGRS.2018.2802944.

[23] H. Chen, Z. Shi, and Y. Li, "A Multi-Task Learning Framework for Road Extraction and Segmentation from Remote Sensing Images," Remote Sensing, vol. 12, no. 18, p. 2987, Sep. 2020, doi: 10.3390/rs12182987.

[24] Esri, "How Multi-Task Road Extractor works." [Online]. Available: https://developers.arcgis.com/python/latest/guide/how-multi-task-road-extractor-works/. [Accessed: Oct. 16, 2024].