iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: III Month of publication: March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78115

Towards Accurate and Scalable Road Extraction from Satellite Imagery Using Deep Neural Networks

Ketan Kanjiya[1], Piyush Sonani[2], Upendrasinh Zala[3]
[1]*Chief Research Officer, Kshatrainfotech Pvt Ltd, Ahmedabad, Gujarat, India*
[2]*Chief Technology Officer, Kshatrainfotech Pvt Ltd, Ahmedabad, Gujarat, India*
[3]*Chief Executive Officer, Kshatrainfotech Pvt Ltd, Ahmedabad, Gujarat, India*

Abstract: Accurate extraction of road networks from high resolution satellite imagery is essential for various geospatial applications, including urban planning, autonomous navigation, disaster management, and geographic information system (GIS). However, automatic road segmentation remains challenging due to occlusions, spectral similarities with surrounding structures, and complex road geometries. This study proposes a deep learning based framework for road extraction using a transformer based semantic segmentation architecture. The model employs a SegFormer-B1 backbone to effectively capture both local spatial features and long range contextual information from satellite images. Experiments were conducted on the DeepGlobe Road Extraction dataset consisting of high resolution satellite imagery and corresponding road masks. The proposed approach achieved strong segmentation performance with a Mean Intersection over Union (IoU) of 0.7879, Dice score of 0.8632, and pixel accuracy of 97.8%. Qualitative results further demonstrate that the model successfully captures major road structures while maintaining good alignment with ground truth annotations. The results highlight the effectiveness of transformer based architectures for road extraction tasks and their potential for large scale geospatial mapping applications.
Keywords: Road Extraction, Satellite Imagery, Deep Learning, Remote Sensing, Semantic Segmentation

I. INTRODUCTION

Accurate road extraction from high resolution satellite imagery plays a crucial role in modern geospatial applications, including autonomous navigation systems, urban infrastructure planning, smart city development, disaster response, and real-time geographic information system (GIS) updates. Road networks form the structural foundation of transportation systems and are indispensable for intelligent mobility, spatial analytics, and emergency logistics. Recent advances in aerospace engineering and remote sensing technologies have enabled the acquisition of ultra high resolution satellite images that capture fine-grained spatial details such as pavement textures, lane markings, road boundaries, and material variations. These rich visual cues provide unprecedented opportunities for precise digital mapping and large scale geospatial intelligence. Despite the availability of high quality imagery, automatic road extraction remains a challenging and unresolved problem. Roads are frequently occluded by trees, buildings, vehicles, and shadows, resulting in discontinuities and broken network structures. In dense urban environments, spectral similarity between road surfaces and surrounding objects such as rooftops, parking areas, and bare land makes pixel-level discrimination highly ambiguous. Furthermore, the structural complexity of road networks characterized by variable widths, multi-scale geometries, irregular curvatures, and complex intersections, introduces significant difficulties for maintaining both geometric precision and topological continuity. These factors jointly degrade segmentation accuracy and reduce the reliability of extracted road networks for downstream applications. Modern deep learning based segmentation frameworks have significantly improved automatic road extraction by enabling data driven feature learning and hierarchical semantic representation. However, most existing models face a fundamental trade-off between preserving fine-grained boundary details and maintaining long range structural connectivity. Architectures that emphasize local feature learning often produce fragmented road segments in occluded regions, while models designed to capture global context may blur road boundaries and lose small scale structural details. This imbalance between local precision and global continuity remains a major limitation in high resolution road extraction. Moreover, many current methods struggle to generalize across diverse geographic environments, varying illumination conditions, and heterogeneous land cover distributions. Robust extraction frameworks must not only achieve high segmentation accuracy but also maintain structural consistency, continuity, and topological correctness under complex real-world conditions.

Addressing these challenges requires architectures that can simultaneously model multi-scale spatial features, refine boundary localization, and preserve long range connectivity. To address these limitations, this paper proposes a novel deep learning framework for accurate road extraction from high resolution satellite images. This study explores the capability of a SegFormer based architecture to capture both local spatial features and global contextual relationships for road segmentation. By jointly optimizing local detail preservation and long range structural reasoning, the framework aims to achieve robust and consistent performance across diverse urban and rural scenarios, providing a scalable and reliable solution for next generation geospatial intelligence systems.

II. LITERATURE REVIEW

The evolution of road extraction methodologies reflects a transition from handcrafted feature based approaches to data driven deep learning frameworks. Early methods relied on traditional machine learning techniques such as Support Vector Machines, Conditional Random Fields, and clustering algorithms including K-means, which utilized manually designed shallow features such as spectral intensity, texture descriptors, and geometric primitives [1-2].

Heuristic approaches, including template matching and active contour models, were widely explored due to their conceptual simplicity, but they required extensive manual tuning and exhibited limited robustness in complex urban environments where road surfaces share spectral similarities with buildings and parking lots [3]. Object-oriented techniques such as watershed segmentation attempted to delineate road structures through region based partitioning strategies, yet these methods were highly sensitive to noise and prone to over segmentation [4-5].

With the advancement of artificial intelligence, deep learning introduced a paradigm shift in road extraction by enabling automatic hierarchical feature learning and data driven representation modeling [6]. Initial deep learning approaches employed patch based convolutional neural networks for local region classification, which suffered from high computational costs and fragmented prediction artifacts commonly referred to as “salt-and-pepper” noise [7].

The emergence of Fully Convolutional Networks marked a major breakthrough by enabling end to end pixel-wise prediction through the replacement of fully connected layers with convolutional layers and the use of upsampling operations to restore spatial resolution [8-9].

Subsequently, encoder-decoder architectures became the dominant design paradigm for semantic segmentation in remote sensing applications [10-11]. U-Net introduced symmetric skip connections that fuse shallow spatial features with deep semantic representations, significantly improving boundary localization and small road target detection [12]. To facilitate deeper architectures, residual learning mechanisms were incorporated in models such as ResUnet, improving gradient propagation and training stability while maintaining parameter efficiency [13].

LinkNet further enhanced computational efficiency by optimizing feature bypass connections, while D-LinkNet employed cascaded dilated convolutions to expand receptive fields and capture richer contextual information without reducing feature map resolution [14].

Recent research has increasingly focused on modeling long range dependencies and structural continuity in road networks, particularly in occluded and cluttered environments [15]. Multi-scale feature fusion strategies and graph based modeling techniques have been introduced to capture contextual dependencies across large spatial extents [14]. DeepLab variants utilize Atrous Spatial Pyramid Pooling (ASPP) to aggregate multi-scale contextual information, enabling effective handling of road width variations and scale diversity [16]. Architectures such as AD-RoadNet introduce auxiliary decoders to decouple connectivity reasoning from multi-scale representation learning, supporting improved inference across occluded roads [17].

Attention mechanisms have further refined feature representation by selectively emphasizing salient road features and suppressing background interference [18-19]. Modules such as Squeeze-and-Excitation, Efficient Channel Attention, and Convolutional Block Attention Module enhance channel-wise and spatial feature discrimination, contributing to improved boundary refinement and segmentation accuracy [20].

More recently, Transformer based architectures have emerged as powerful tools for capturing long range dependencies through self-attention mechanisms, overcoming the locality constraints of conventional CNNs [21]. Hybrid designs such as HDDNet employ heterogeneous dual-decoder and dual-path structures to jointly model fine-grained textures and global topological continuity across complex satellite imagery scenes [22].

III.DATASET

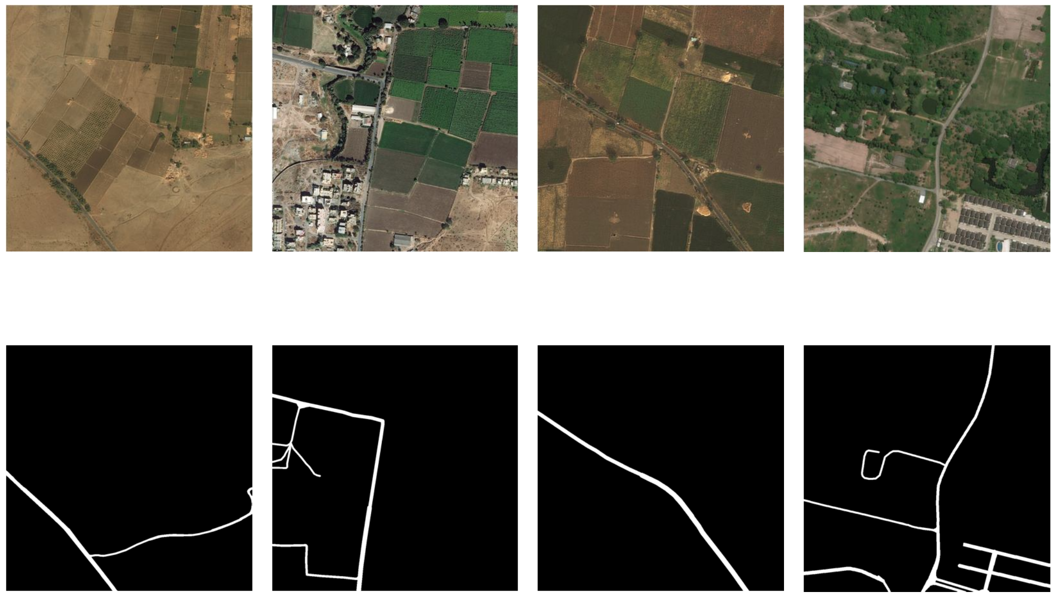


Fig. 1 Sample images from the dataset: (a) satellite images and (b) corresponding ground-truth road segmentation masks.

This study utilizes the DeepGlobe Road Extraction dataset. The dataset consists of high resolution satellite imagery captured by DigitalGlobe satellites with a spatial resolution of 50 cm per pixel, enabling detailed observation of road structures across diverse environments. The training set contains 6,226 RGB satellite images, each with a resolution of 1024 × 1024 pixels, along with corresponding ground truth segmentation masks. In addition, the dataset provides 1,243 validation images and 1,101 test images, although ground truth masks are not available for the test set. A sample of the dataset images and their corresponding ground truth masks is shown in Fig. 1.

Each satellite image is paired with a binary road mask, where road pixels are represented in white and background pixels in black. Since the mask values are not always strictly 0 or 255, a threshold of 128 is applied for binarization during preprocessing.

IV.METHODOLOGY

This study proposes a deep learning based framework for automatic road extraction from satellite imagery. The overall pipeline consists of data preprocessing, augmentation, model training, and evaluation using a transformer based semantic segmentation architecture.

A. Data Preprocessing

The DeepGlobe Road Extraction dataset is prepared by creating training and validation splits from the provided training data. A 90:10 split is used to divide the dataset into training and validation subsets. Each satellite image is paired with a corresponding segmentation mask. Since the masks are provided as RGB images, they are converted into binary class labels representing road and background. The masks are then transformed into a one-hot encoded representation to make them suitable for training the segmentation model.

To improve model stability and generalization, input images are normalized using ImageNet mean and standard deviation values, and both images and masks are converted into tensor format for processing by the deep learning framework.

To further increase the diversity of training samples and reduce overfitting, data augmentation techniques are applied during training. Specifically, horizontal and vertical flipping are used with a probability of 0.5. These geometric transformations allow the model to learn road patterns under different orientations, improving robustness to variations in satellite imagery.

B. Model Architecture

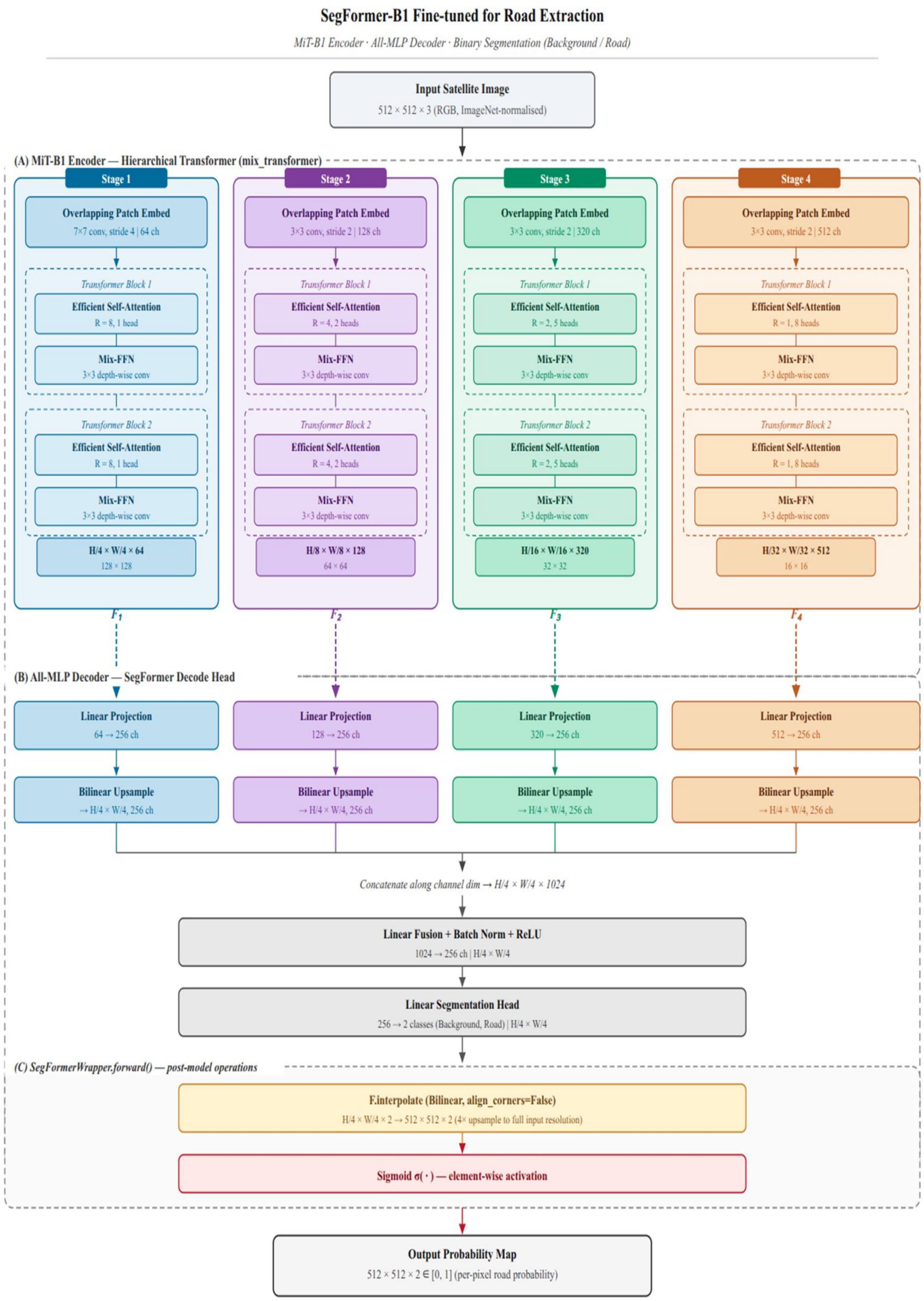


Fig. 2 Architecture of the SegFormer-B1 based road extraction model.

As shown in Fig. 2, the proposed framework employs a SegFormer based semantic segmentation model. SegFormer is a transformer based architecture designed for efficient and accurate semantic segmentation. It combines a hierarchical transformer encoder with a lightweight segmentation decoder to capture both local and global contextual information.

In this work, the SegFormer-B1 backbone pretrained on the ADE20K dataset is utilized as the encoder. The pretrained weights allow the model to leverage rich visual representations learned from large scale datasets. The model is adapted for binary road segmentation by modifying the output layer to predict two classes: road and background. The predicted segmentation maps are upsampled to match the original input resolution using bilinear interpolation.

C. Training Strategy

The model is trained using the Dice loss function, which is particularly effective for segmentation tasks with class imbalance. Optimization is performed using the Adam optimizer with a learning rate of 8×10^{-5}. The model is trained for 20 epochs with a batch size of 4.

D. Experimental Setup

All experiments were conducted on a high performance system running Ubuntu 24.04.1 LTS, equipped with an AMD Ryzen 7 5700X CPU, 64 GB RAM, and an NVIDIA GeForce RTX 4090 GPU. The implementation was carried out using the PyTorch deep learning framework, with additional libraries including Hugging Face Transformers, segmentation-models-pytorch, and Albumentations for model implementation, data processing, and augmentation.

V. RESULTS

A. Quantitative Evaluation Metrics

The performance of the proposed SegFormer-based road extraction model was evaluated using several commonly used semantic segmentation metrics, including Precision, Recall, F1-score, Intersection over Union (IoU), Dice score, and Pixel Accuracy. These metrics provide a comprehensive assessment of the model's ability to correctly identify road pixels while minimizing false detections.

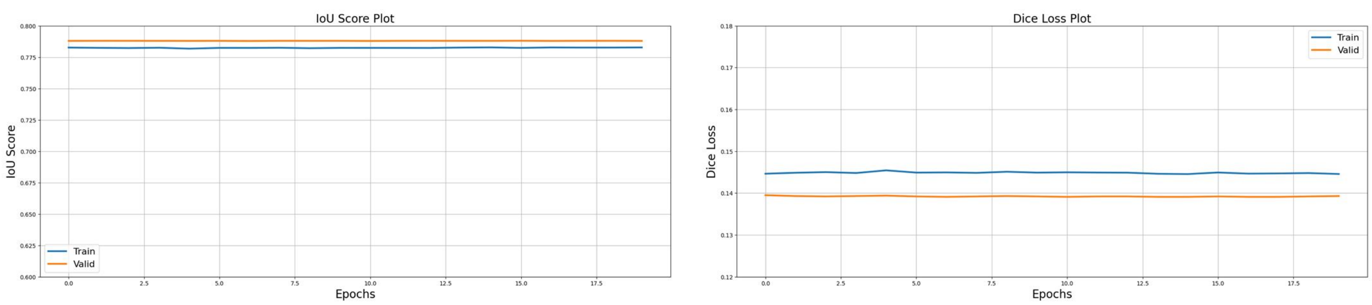
Table 1. Quantitative performance results

Metric	Score
Precision	0.8617
Recall	0.8784
F1 Score	0.8632
Mean IoU	0.7879
Dice Score	0.8632
Pixel Accuracy	0.9780
Dice Loss	0.1399

The proposed model achieved a Mean IoU score of 0.7879, indicating a strong overlap between predicted road regions and ground truth masks. The Dice score of 0.8632 further confirms the effectiveness of the segmentation model in accurately identifying road pixels. Additionally, a pixel accuracy of 97.8% demonstrates that the model correctly classifies the majority of pixels in the satellite imagery.

B. Training Performance Analysis

To analyze the training stability of the proposed model, the IoU score and Dice loss were monitored across training epochs.



(a) (b)

Fig. 3 Training performance of the SegFormer-based road segmentation model showing (a) IoU score vs. epochs and (b) Dice loss vs. epochs.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

Figure 3(a) illustrates the variation of IoU score during training. The model achieves a stable IoU score of approximately 0.782-0.783, indicating consistent segmentation performance across epochs. Figure 3(b) shows the Dice loss progression during training. The Dice loss remains around 0.144, demonstrating that the model converges early and maintains stable training behavior. The minimal fluctuation in both IoU and Dice loss suggests that the model reaches convergence early and does not suffer from significant overfitting.

C. Qualitative Results

To qualitatively evaluate the segmentation performance, several prediction samples are shown in Figure 4. Each example includes the original satellite image, the corresponding ground truth mask, the predicted segmentation mask, and the predicted road probability heatmap.

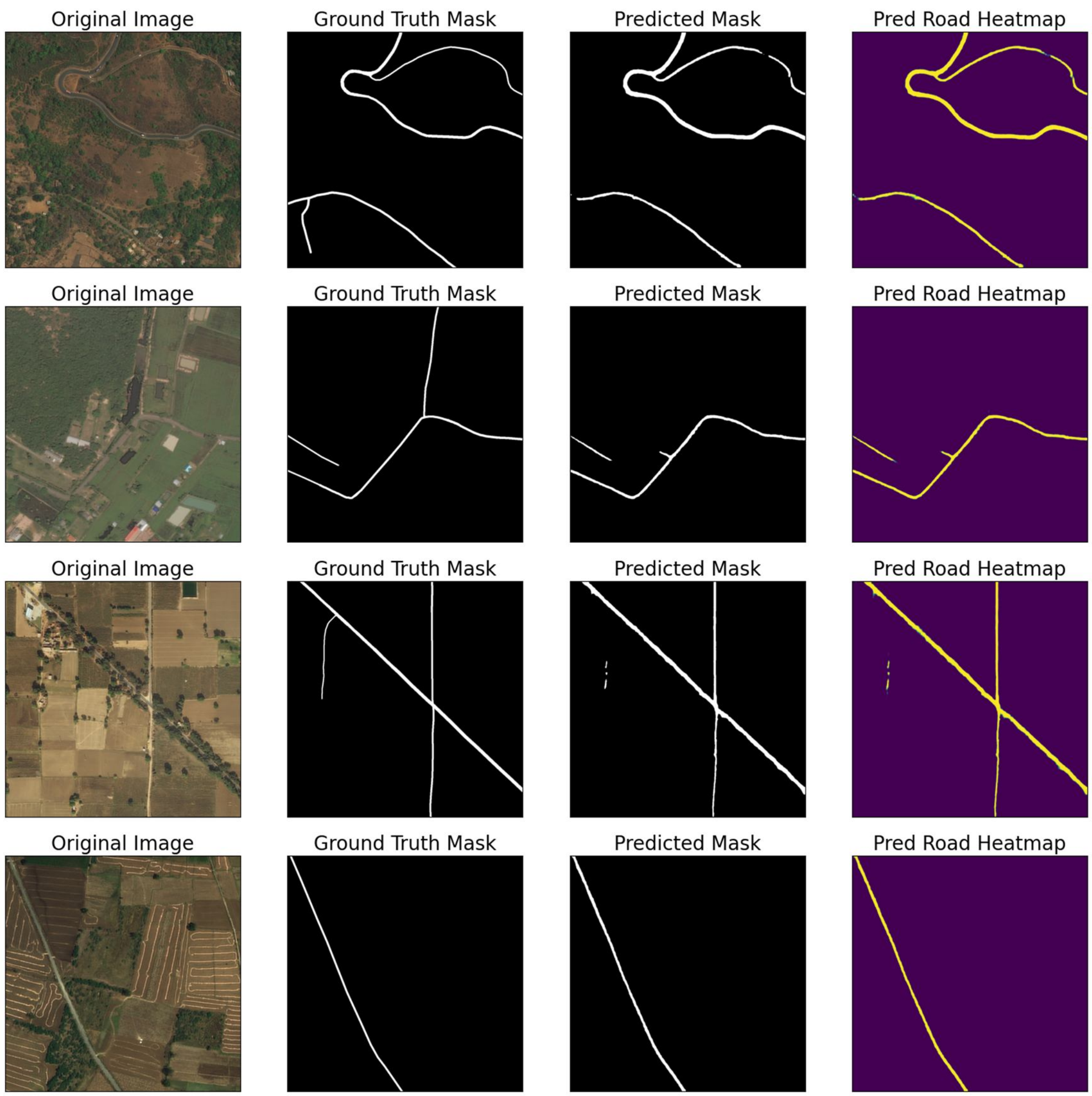


Fig. 4 Qualitative road extraction results showing the original satellite image, ground truth mask, predicted segmentation mask, and predicted road probability heatmap.

The visual results demonstrate that the model successfully captures major road structures and maintains good alignment with ground truth annotations. The predicted heatmaps further highlight the confidence of the model in identifying road regions.

VI. CONCLUSION

This study presented a deep learning framework for automatic road extraction from high resolution satellite imagery using a SegFormer based semantic segmentation model. The transformer based architecture effectively captures both local spatial features and global contextual information required for accurate road segmentation. Experiments on the DeepGlobe Road Extraction dataset demonstrate strong performance, achieving a Mean IoU score of 0.7879, Dice score of 0.8632, and pixel accuracy of 97.8%. These results indicate that the model can reliably identify road structures while maintaining consistent segmentation across complex satellite scenes. Qualitative analysis further shows that the model successfully delineates major road networks with good alignment to ground truth masks. Future work will focus on improving the detection of narrow and occluded roads by incorporating multi-scale feature fusion, boundary refinement techniques, and topology aware learning strategies to enhance segmentation accuracy and structural continuity.

REFERENCES

[1] H. R. R. Bakhtiari, A. Abdollahi, and H. Rezaeian, "Semi automatic road extraction from digital images," The Egyptian Journal of Remote Sensing and Space Sciences, vol. 20, no. 1, pp. 117–123, 2017, doi: 10.1016/j.ejrs.2017.03.001.

[2] Y. Wu, Q. Zhao, Z. Shen, and Y. Li, "Clustering point process based network topology structure constrained urban road extraction from remote sensing images," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 2087–2098, 2022, doi: 10.1109/JSTARS.2022.3151757.

[3] Y. Lv, Z. Yu, Z. Yin, and J. Qin, "Remote sensing image road segmentation based on conditions perceived 3D UX-Net," Journal of the Indian Society of Remote Sensing, vol. 53, no. 12, pp. 4005–4015, 2025, doi: 10.1007/s12524-025-02158-3.

[4] Y. Wang, H. Huang, and B. Wu, "Evaluating the potential of SDGSAT-1 glimmer imagery for urban road detection," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 18, pp. 785–794, 2025, doi: 10.1109/JSTARS.2024.3502218.

[5] P. Kumar, A. S. Mathias, and A. N. Padmasali, "A YOLOv8 segmentation approach for detecting unstructured road boundaries in rural terrains using dash-camera," IEEE Access, vol. 13, pp. 213429–213438, 2025, doi: 10.1109/ACCESS.2025.3645678.

[6] K. Zhang, A. As'arry, X. Shen, A. A. Hairuddin, M. K. Hassan, L. Zhu, and W. Qin, "DSWFNet: Dual-branch fusion of spatial and wavelet features for road extraction from remote sensing images," Scientific Reports, vol. 16, no. 1, p. 3966, 2025, doi: 10.1038/s41598-025-34091-3.

[7] R. Lian, W. Wang, N. Mustafa, and L. Huang, "Road extraction methods in high-resolution remote sensing images: A comprehensive review," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 5489–5507, 2020, doi: 10.1109/JSTARS.2020.3023549.

[8] A. Abdollahi, B. Pradhan, N. Shukla, S. Chakraborty, and A. Alamri, "Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review," Remote Sensing, vol. 12, no. 9, p. 1444, 2020, doi: 10.3390/rs12091444.

[9] Z. Hong, D. Ming, K. Zhou, Y. Guo, and T. Lu, "Road extraction from a high spatial resolution remote sensing image based on richer convolutional features," IEEE Access, vol. 6, pp. 46988–47000, 2018, doi: 10.1109/ACCESS.2018.2867210.

[10] X. Kuang, F. Cheng, C. Wu, H. Lei, and Z. Zhang, "BIR-Net: A lightweight and efficient bilateral interaction road extraction network," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 14194–14207, 2024, doi: 10.1109/JSTARS.2024.3439267.

[11] Q. Guo and Z. Wang, "A self-supervised learning framework for road centerline extraction from high-resolution remote sensing images," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 4451–4461, 2020, doi: 10.1109/JSTARS.2020.3014242.

[12] Y. Xu, Z. Xie, Y. Feng, and Z. Chen, "Road extraction from high-resolution remote sensing imagery using deep learning," Remote Sensing, vol. 10, no. 9, p. 1461, 2018, doi: 10.3390/rs10091461.

[13] H. S. Munawar, A. W. A. Hammad, S. T. Waller, D. Shahzad, and M. R. Islam, "Road network detection from aerial imagery of urban areas using deep ResUNet in combination with the B-snake algorithm," Human-Centric Intelligent Systems, vol. 3, no. 1, pp. 37–46, 2023, doi: 10.1007/s44230-023-00015-5.

[14] R. Liu, J. Wu, W. Lu, Q. Miao, H. Zhang, X. Liu, Z. Lu, and L. Li, "A review of deep learning-based methods for road extraction from high-resolution remote sensing images," Remote Sensing, vol. 16, no. 12, p. 2056, 2024, doi: 10.3390/rs16122056.

[15] W. Wang, N. Yang, Y. Zhang, F. Wang, T. Cao, and P. Eklund, "A review of road extraction from remote sensing images," Journal of Traffic and Transportation Engineering (English Edition), vol. 3, no. 3, pp. 271–282, 2016, doi: 10.1016/j.jtte.2016.05.005.

[16] Y. Zhang, N. Sun, L. Chen, L. Liu, and H. Zhu, "FAMNet: Lightweight road extraction network with fused attention and multilevel cascaded ASPP," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 18, pp. 25616–25629, 2025, doi: 10.1109/JSTARS.2025.3614661.

[17] Z. Luo, K. Zhou, Y. Tan, X. Wang, R. Zhu, and L. Zhang, "AD-RoadNet: An auxiliary-decoding road extraction network improving connectivity while preserving multiscale road details," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 8049–8062, 2023, doi: 10.1109/JSTARS.2023.3289583.

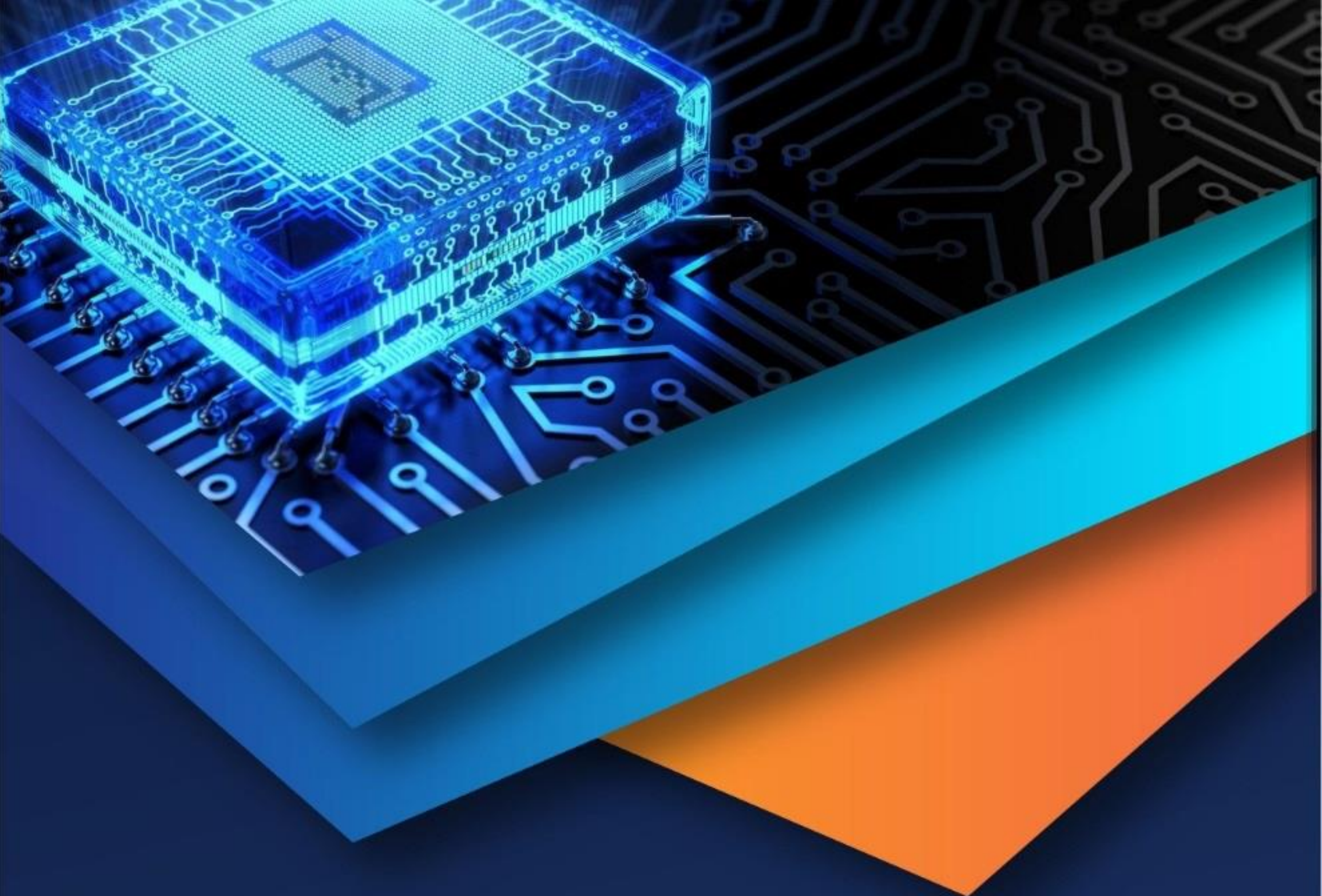
[18] S. Qu, G. Liu, X. Zhang, and Y. Liu, "Heterogeneous dual-decoder network for road extraction in remote sensing images," Scientific Reports, vol. 15, no. 1, p. 31619, 2025, doi: 10.1038/s41598-025-17445-9.

[19] S. Mo, Y. Shi, Q. Yuan, and M. Li, "A survey of deep learning road extraction algorithms using high-resolution remote sensing images," Sensors, vol. 24, no. 5, p. 1708, 2024, doi: 10.3390/s24051708.

[20] J. Ren, L. Liu, Z. Xia, and Y. Liu, "AED-Net: A high-resolution remote sensing image road extraction method integrating atrous spatial pyramid pooling and efficient channel attention mechanism," IEEE Access, vol. 13, pp. 61067–61077, 2025, doi: 10.1109/ACCESS.2025.3548262.

[21] H. Bai, C. Ren, Z. Huang, and Y. Gu, "A dynamic attention mechanism for road extraction from high-resolution remote sensing imagery using feature fusion," Scientific Reports, vol. 15, no. 1, p. 17556, 2025, doi: 10.1038/s41598-025-02267-6.

[22] L. Guo, X. Bai, H. Huo, Z. Wu, W. Zhang, and C. Wang, "High-resolution road information extraction based on an improved dual-path neural network," IEEE Geoscience and Remote Sensing Letters, vol. 21, pp. 1–5, 2024, doi: 10.1109/LGRS.2024.3453429.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY