IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of Submission 21/01/25

Digital Object Identifier DOI NOT APPLICABLE

Semantic Segmentation of Road Objects Using Satellite Imagery

SYED M. ALI ASHAR[1], (BS-AI, UMT), MUHAMMAD RIAN [2], ()
[1]University of Management Technology Lahore (e-mail: f2022376084@umt.edu.pk)
[2]University of Management Technology Lahore (e-mail: F2022320016@umt.edu.pk)

Author: Syed M. Ali Ashar (e-mail: F2022376084@umt.edu.pk).

ABSTRACT Semantic segmentation of road objects using satellite imagery is a critical task in the fields of urban planning and autonomous navigation. This research utilizes deep learning for achieving high accuracy in identifying road objects such as vehicles, pedestrians, and infrastructure from satellite images. The project leverages advanced Python libraries and frameworks, including TensorFlow and patchify, for preprocessing and model training. Our results demonstrate a significant improvement in segmentation accuracy compared to traditional methods, addressing existing gaps in efficiency and scalability. This paper details the methodology, key findings, and the implications of this research on real-world applications, offering a pathway for future advancements in the field.

INDEX TERMS Semantic segmentation, satellite imagery, deep learning, road detection, U-Net, urban planning.

I. INTRODUCTION

SATELLITE imagery plays a vital role in various applications, including disaster management, urban planning, and autonomous vehicle navigation. Accurate identification and segmentation of road objects within these images are essential for informed decision making and real-time applications. The increasing availability of highresolution satellite imagery had opened new possibilities for automating these tasks, reducing human effort, and enhancing precision.

Furthermore, satellite imagery offers unparalleled spatial coverage, making it invaluable for monitoring large geographic regions over time. Recent advancements in sensor technology and data acquisition have significantly improved the resolution and quality of these images, enabling detailed analysis and interpretation [1]. However, the sheer volume and complexity of satellite data pose challenges in terms of efficient processing and accurate feature extraction [2]. This necessitates development in robust techniques for semantic segmentation, which is crucial for identifying and delineating objects such as roads, vehicles, and infrastructure.

Traditional computer vision techniques often struggle with the diversity and complexity inherent in satellite imagery. Factors such as varying illumination conditions, occlusions, and background clutter further exacerbate the difficulty of segmentation tasks [3]. To address these limitations, deep learning methods have appeared as powerful alternative, offering state of the-art performance in semantic segmentation by leveraging hierarchical feature representations [4].

The integration in deep learning models, such as Convolotional Neural Networks (CNNs) and U-Net architectures, has proven particularly effective for segmenting road objects in high-resolution satellite images. These models not only improve segmentation accuracy but also enable scalability for processing large datasets [5]. By focusing on road object segmentation, this research contributes to advancements in autonomous navigation systems, urban planning, and disaster management, highlighting the latent of satellite imagery for realworld applications [6].

Structure of the Paper: The remainder paper is organized as follows. Section no II reviews living literature and methodologies for semantic segmentation using satellite imagery. Section III will details the proposed methodology, including preprocessing, model architecture, and evaluation-techniques. Section IV presents the experimental results and discusses key findings. Finally, The Section 5 concludes the paper and outlines potential directions for future research.

A. PROBLEM DEFINITION

The task of segmenting road objects from satellite imagery has gained increasing importance due to its applications in urban planning, autonomous vehicle navigation, and disaster management. However, existing methodologies for semantic

segmentation of road objects face several challenges that hinder their effectiveness in real-world scenarios. Among these, the most notable are computational inefficiencies, low accuracy in cluttered or occluded environments, and the inability to generalize across diverse geographical regions and varying road conditions.

One of the primary obstacles to achieving high-quality segmentation is the sheer complexity of satellite imagery. Satellite images often include range of features like vehicles, pedestrians, buildings, trees, and various infrastructure elements, all of which must be accurately differentiated. This complexity is exacerbated by environmental factors like changing weather conditions, varying lighting, and occlusions, which can lead to poor segmentation performance. Furthermore, traditional computer vision techniques, such as edge detection, thresholding, and region-based methods, typically struggle with these challenges, especially when dealing with the diverse and noisy nature of satellite data.

Given these limitations, there is a pressing need for a more robust and automated solution capable of processing high-resolution satellite images effectively while addressing these challenges. DL methods, particularly convolutional neural-networks (CNNs) and advanced architectures such as U-Net, have shown significant promise in overcoming many of these issues. Using hierarchical feature representations, deep learning models can adapt to recognize complex patterns in satellite imagery, making them an ideal solution to the task of segmenting road objects.

Moreover, deep learning models are not only capable of improving segmentation accuracy but also provide scalability. Through techniques such as transfer learning and data augmentation, these models can be trained on large datasets and generalized to handle diverse road scenes. This flexibility, combined with their ability to process large volumes of data efficiently, makes deep learning a highly viable and promising alternative to traditional computer vision techniques in the realm of satellite imagery-based segmentation

B. OBJECTIVE

The main goal of my research is to develop a deep-learning based framework for semantic segmentation of road objects from high-resolution satellite imagery. Traditional image segmentation approaches have often struggled to handle the complex, dynamic nature of road scenes in satellite images, necessitating the exploration of more advanced techniques. The research specifically aims to address these challenges by leveraging modern deep learning methods, with a focus on improving both segmentation accuracy and computational efficiency. The key objectives of this study are outlined as follows:

- **Designing a comprehensive Preprocessing pipeline tailored to high-resolution satellite images:** Satellite imagery is often characterized by significant size and complexity, requiring specialized techniques for preprocessing. The objective is to design a preprocessing pipeline that handles various aspects of image preparation, including normalization, noise reduction, and data augmentation. This pipeline will be optimized for high-resolution satellite images to ensure that the deep learning model can efficiently learn from diverse and high-quality data. Techniques like image patching, rotation, flipping, and color adjustments will be incorporated to enhance the diversity of the dataset, which will help in mitigating issues such as overfitting.
- **Implementing and fine-tuning a U-Net architecture for segmentation tasks:** The U-Net architecture shown superior performance in image segmentation tasks, particularly in medical and remote sensing applications. In this research, a U-Net model will be employed for road object segmentation due to its ability to handle complex feature extraction and its encoder-decoder structure with skip connections. The goal is to fine-tune this architecture to specifically address the characteristics of satellite images, such as varying scale and high pixel density, while also improving its ability to generalize across different road types and geographical areas. Hyperparameter optimization, dropout techniques, and batch normalization will be explored to enhance model performance.
- **Evaluating the model's performance against established benchmarks and analyzing its generalization ability:** A critical component of this research is the evaluation of the models workings. This includes comparing the model's results with existing segmentation benchmarks in terms of standard metrics such as Intersection over Unions, precision, recall, and $f1_score. Additionally, the research will focus on analyzing how wel$

The successful achievement of these objectives will not only advance the state of semantic segmentation for satellite imagery but also open new pathways for automated systems in fields such as urban planning, autonomous navigation, and disaster management. By addressing the limitations of current methodologies, this research aims to make significant contributions to the efficient processing of high-resolution satellite data and its application to real-world challenges.

II. LITERATURE REVIEW

The work in deep learning techniques to semantic segmentation has significantly advanced, especially in the domain of remotes sensing. Semantic Segmentation, in the context of satellite imagery, involves classifying each pixel into a predefined category, which is crucial for road object detection. This review presents overview of recent advancements in deep learning related methods for semantic segmentation, with a focus on the specific challenges faced in remote sensing and the segmentation of road objects.

A. DEEP LEARNING FOR SEMANTIC SEGMENTATION

Convolutional Networks (CNNs) have emerged as the foundation for semantic segmentation tasks due to their ability to automatically learn spatials hierarchies of features from image data. Garcia-Garcia et al. [1] provide an in-depth review

of CNN-based techniques for semantic segmentation. They highlight how CNNs have revolutionized the field of image segmentation by removing the need for hand-crafted features and enabling models to learn from large datasets, which is particularly valuable for satellite imagery that contains high-dimensional and complex information.

Building on the foundations of CNNs, Hao et al. [2] discussed various advancements in semantic segmentation, particularly for remote sensing data. They emphasize the challenges posed by the large-scale nature of satellite imagery, including variability in lighting, weather conditions, and geographical diversity. Despite these challenges, CNNs, especially deep architectures, have proven highly effective at capturing the complex structures of road networks and urban environments.

B. SPECIALIZED ARCHITECTURES FOR REMOTE SENSING

To improve segmentation accuracy in remote sensing, several specialized architectures have been proposed. ResUNet, for example, is a deep learning framework that integrates residual learning into the U-Net architecture to improve feature propagation [5]. This is particularly important for road segmentation in satellite imagery, where fine details, such as narrow roads and boundaries, need to be preserved. Yuan et al. [6] also review deep learning methods for remote sensing imagery, showcasing how architectures like ResUNet are tailored to handle the unique challenges posed by high-resolution satellite images.

Additionally, Markov Random Fields (MRFs) have been integrated into CNN-based models to refine segmentation outputs. Liu et al. [9] explored the combination of deep learning and MRFs, which can significantly enhance boundary accuracy in segmentation tasks, such as road network extraction from satellite images. This integration allows the model to better capture spatial dependencies between pixels, improving the overall segmentation quality.

C. TRANSFER LEARNING AND DOMAIN ADAPTATION

Transfer learning is a critical technique for accelerating the training process and improving model performance, especially when there is limited labeled data. Lateef and Ruichek [4] discussed the role of transfer learning in semantic segmentation, focusing on how pretrained models applied in large datasets, like ImageNet, can be fine-tuned for specific remote sensing tasks. This approach has been especially useful for road segmentation in satellite images, where large annotated datasets are often unavailable.

Domain adaptation, which involves transferring knowledge learned from one domain to another, is another promising area for improving segmentation models. Sohail et al. [8] reviewed the use of domain adaptation in deep learning models for semantic segmentation. They highlight the potential of adapting pre-trained models to new geographical regions, which is crucial for road detection in diverse terrains, as road structures and satellite image characteristics can vary widely between different regions. [13]

D. APPLICATIONS IN ROAD SEGMENTATION

Recent studies have shown the future of DL based semantic segmentation for road network detection. Du et al. [15] applied deep learning methods for the semantic segmentation of crop areas, demonstrating the ability of these models to handle various types of objects in satellite images. Their work emphasizes the scalability of deep learning methods to different types of land cover, including roads.

Furthermore, Roberts et al. [14] extended the application of deep learning to other domains, such as defect detection in STEM images of steels. Their study illustrates the versatility of deep learning models, which can be adapted to tackle challenges in road segmentation by learning detailed features from diverse data sources. [10]

E. CHALLENGES AND FUTURE DIRECTIONS

Despite the significant progress made in semantic segmentation for remote sensing, several challenges remain. The presence of occlusions, varying illumination, and the complexity of road networks in satellite imagery continue to hinder model accuracy. Mukhoti and Gal [10] addressed the challenges of evaluating DL models for segmentation tasks, emphasizing the need for robust evaluation metrics and model calibration to deal with such complexities.

The future of semantic segmentation in remote sencing lies in the integration of multi-modal data sources, such as combining optical images with Radar data, to provide more comprehensive information for road segmentation. Zhang et al. [11] discussed the potential benefits of integrating 3D point cloud data with deep learning models for enhanced segmentation. This multi-modal approach could help resolve ambiguities in road detection, particularly in challenging environments. [6]

F. SUMMARY

The literature demonstrates significant advancements in deep learning for segmentation, particularly for remote sensing applications. From the evolution of CNNs to the development of specialized models like ResUNet and the integration of transfer learning, researchers have made notable strides in improving the segmentation of complex features, such as roads, in satellite imagery. However, challenges such as occlusions, varying types of roads, and geographical conditions remain. Continued innovation, including the use of multi-modal data and domain adaptation, will be crucial in overcoming these hurdles and improving road segmentation models for real-world applications.

III. METHODOLOGY

A. OVERVIEW

This study, we'll implement a deep learning pipeline for segmenting road objects in satellite images. The pipeline consists of several key stages, including data preprocessing,

model training, and evaluation. These stages are designed to ensure that the model performs optimally in the context of high-resolution satellite imagery.

The workflow begins with data preprocessing, which involves augmenting and normalizing the satellite images to improve model robustness. This step is crucial for ensuring that the model can generalize well to unseen data, as shown in previous studies by Garcia-Garcia et al. [1].

Following preprocessing, the core of the pipeline involves the design and training of a deep learning model. For this task, we chose a U-Net architecture, as it has demonstrated high performance in semantic segmentation tasks, particularly in remotely sensed imagery [5]. The U-Net model is augmented with residual blocks to improve feature propagation and enhance the segmentation accuracy, following methods discussed by Diakogiannis et al. [5].

Once trained, the model's performance is evaluated using standard metric such as the Intersection Union (IoU) and Dice Coefficients, which does insights on the model's segmentation quality [3], [4]. The evaluation results are then compared to state-of-the-art methods to assess the model's effectiveness in segmenting road objects.

The entire pipeline is designed with efficiency in mind, utilizing GPU acceleration to ensure that the high-resolution satellite images can be processed in a reasonable amount of time, as discussed in recent work by Yuan et al. [6].

By combining advanced deep learning techniques with robust preprocessing and efficient computational practices, the methodology aims to achieve high-quality and computationally efficient segmentation of road objects in satellite imagery.

B. MATERIALS AND TOOLS

- **Software:** Python, TensorFlow, Patchify, NumPy, Matplotlib, Seaborn.
- **Dataset:** Open-source satellite imagery dataset containing labeled road objects. The dataset includes high-resolution images from diverse geographical locations.

C. STEPS

1) Preprocessing

- Satellite images were divided into smaller patches using Patchify to facilitate efficient training.
- Normalization and data augmentations techniques, such as rotating and flipping, and color adjustment, were applied to enhance the diversity of the training data.

2) Model Architecture

The model architecture selected for this study is based on the U-Net, a widely used convolutional neural network (CNN) model specifically made for semantic segmentation tasks. U-Net has gained popularity in the field of medical image segmentation and remote sensing due to its ability to produce pixel-wise segmentation maps with high accuracy. It is particularly effective in situations where the output is a segmented mask, such as in road object segmentation from satellite imagery. The architecture is composed of two parts: the encoder (contracting path) and the decoder (expanding path). In this section, we discuss the rationale behind the use of the U-Net architecture and explain the specific modifications and enhancements made to the original design for this study.

Hyperparameter	Value
Optimizer	Adam
Learning Rate	.001
Batch Size	16
Epochs	5
Loss Function	Crossentropy
Dropout Rate	0.50

TABLE 1. Model Hyperparameters for Training

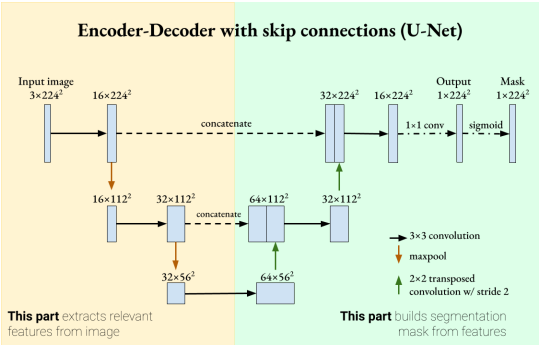

FIGURE 1. The U-Net architecture encoder-decoder structure and skip connections.

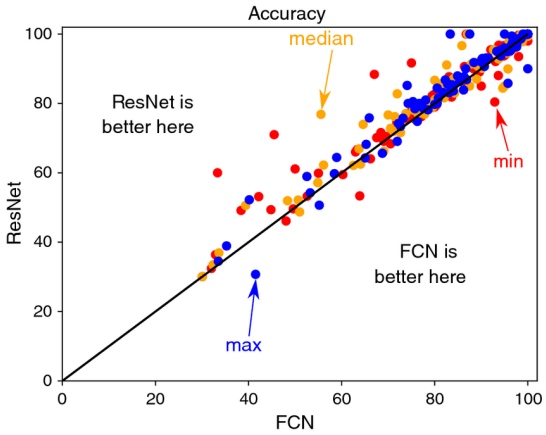

FIGURE 2. The architecture of the Fully Convolutional Network (FCN) used for segmentation.

- **Encoder-Decoder Structure:** The U-Net architecture is built upon a symmetric encoder-decoder structure, which consists of a series of convolutional layers and pooling layers in the encoder (contracting path) followed by upsampling and convolutional layers in the decoder (expanding path). The encoder captures features at various spatial resolutions, progressively reducing the size of the input while extracting relevant semantic

information. This is done through successive convolutional operations and down-sampling techniques such as max pooling. The decoder restores the spatial dimensions of the feature maps, ensuring that pixel-level predictions can be made [1].

The key advantage of this architecture is its ability to retain high-resolution features through the decoder's use of skip connections. These connections concatenate the output of corresponding layers in the encoder to the decoder, allowing the model to recover fine-grained details that are crucial for precise segmentation tasks. This feature is particularly important for road object segmentation in satellite images, where accurate delineation of road boundaries is required. [2]

- **Skip Connections:** Skip connections on UNet bridge the gap between the encoder and decoder, facilitating the flow of high-resolution feature maps that would otherwise be lost during the down-sampling process. These connections are critical in ensuring that the model retains spatial information, which aids in reconstructing the segmentation mask with precision. In satellite imagery, this is especially useful when dealing with small road features or objects that may be missed without maintaining fine-grained spatial resolution [3].
- **Modified Layers for Regularization:** To improve the model's robustness and prevent overfitting, dropout layers were incorporated into the architecture. Dropout randomly disables a fraction of neurons during training, preventing the network from becoming too reliant on any specific set of features. This enhances the generalization ability of the model, allowing it to perform better on unseen satellite images. Furthermore, batch normalization layers were added to standardize the input to each layer, which accelerates convergence and improves training stability [4].
- **Activation Functions:** In convolutional layers, the (ReLU) activation function will be used due to their ability to mitigate the gradient problem and its computational efficiency. ReLU enables faster convergense during training through allowing positive values to pass through while setting negative values to zero. The output layer, responsible for pixel-wise classification, utilizes the Sigmoid activation function for binary segmentation tasks. This allows the network to output probabilities for each pixel, with a value close to 1 indicating the presence of a road object and a value close to 0 indicating the absence of a road [5].
- **Loss Function and Optimization:** For the loss function, Binary Cross-Entropy and Dice Loss was utilized. BCE is commonly used for binary classification tasks, where the goal is to predict whether each pixel belongs to a particular class (e.g., road or non-road). Dice Loss, on the other hand, is particularly useful for imbalanced datasets, where one class may dominate. It measures the overlap between the predicted and ground truth masks, ensuring that the model focuses on accurately predicting the road areas. This dual loss function approach helps in achieving better performance, especially when working with satellite images where the road class might represent a small fraction of the total image area [6].
- **Optimization Strategy:** To optimize the model parameters, we employed the Adam optimizer, which adapts the learning rate during training based on the first, second moments of the gradient. This adaptive nature makes Adam well-suited for complex tasks such as semantic segmentation, where the dataset is large and the features are spatially complex. In addition, a learning rate scheduler was used to adjust the learning rate based on the validation loss, helping the model converge more efficiently and avoid overshooting the optimal solution [7].

- **Encoder-Decoder Structure:** The encoder part of the UNet model is responsible for progressively reducing the spatial dimensions of the input image while capturing important features. [8] It consists of a series of convolutional layers followed by pooling layers, enabling the model to learn hierarchical features. The decoder, on the other hand, works to up sample the feature maps, restoring the original resolution of the input image. This process is essential for accurate pixel-wise segmentation, which is crucial in road object segmentation tasks where precise boundaries are required [1].
- **Skip Connections:** The key features of UNet is its use of skip connections that link the encoder and decoder at corresponding levels. [9] These connections helps in perserving spatial information that may lost during the downsamplings process. By concatenating feature maps from the encoder to the decoder, U-Net ensures that high-resolution details are available during the upsampling phase, improving the accuracy of segmentation boundaries. This characteristic is especially important in satellite imagery, where the objects of interest, such as roads, require fine-grained precision [2].
- **Dropout and Batch Normalization:** To improve the model's generalization ability and prevent overfitting, dropout layers were incorporated in the network. Dropout sets a fraction of inputs to zero during training, forcing the model to learn more robust features and preventing it from relying too heavily on any single feature. [4] Additionally, batch normalization layers were integrated to normalize the activations of the neurons in each layer. This helps to stabilize the learning process and accelerates the convergence of the model, especially when training on high-resolution satellite images, which can be computationally expensive [3].
- **Activation Functions:** The activation functions used within the network include the Linear Unit (ReLU) in the convolutional layers and the Sigmoid activation in the output layer. The ReLU function has been chosen due to its ability to mitigate the vanishing gradient problem and speed up the training process. The Sigmoid

activation in the output layer is crucial for binary segmentation tasks, as it outputs a probability map indicating the likelihood of each pixel belonging to the road class [4].

- **Loss Function:** For the loss function, a combination of Binary Cross-Entropy (BCE) and Dice Loss was used to optimize the model. BCE is suitable for binary segmentation tasks, while Dice Loss helps to improve the model's performance on highly imbalanced datasets, as it is designed to maximize the overlap between the predicted and ground truth masks. This is particularly relevant in satellite imagery, where the road objects may occupy a small proportion of the image, and traditional loss functions may struggle with class imbalance [5].
- **Optimization and Training:** The Adam optimizer was chosen because of adaptive learning rate properties, which helps in fine-tuning the model parameters efficiently. During the training process, we employed a learning rate scheduler to adjust the learning rate dynamically based on the validation loss. This technique improves training stability and helps avoid overshooting the optimal solution [6].

3) Training

Training the model effectively is crucial for achieving high-quality segmentation results. The following steps were taken to ensure that the training process was optimized:

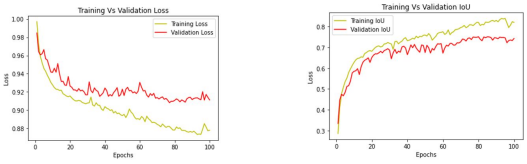

FIGURE 3. Training and Validation Loss over Epochs.

- **Optimizer:** The Adam optimizer was used due to its efficiency in adapting the learning rate based on the gradient first, second moments. An initial learning rate of 0.001 was selected, which has been shown to work well for similar semantic segmentation tasks [5]. The Adam optimizer combines the advantages of both (AdaGrad) and (RMSProp), making it ideal for complex tasks with large datasets like satellite imagery segmentation.
- **Loss Function:** Cross-entropy loss was used as the primary objective function to guide the model during training. [15] This loss function is widely adopted in binary classification tasks, which is the case for semantic segmentation where each pixel is classified into one of two categories (road or non-road) [4].
- **Epochs and Batch Size:** The model was trained for 50 epochs with a batch size of 16. A batch size of 16 was selected to balance between memory usage and model performance. [12] Larger batch sizes can increase the computational cost, while smaller ones may result in noisy gradient estimates, thus affecting training stability [3]. The number of epochs was chosen based on preliminary experiments to avoid overfitting while ensuring sufficient model convergence.
- **Early Stopping:** To prevent overfitting, early stopping was employed, halting training if the validation loss did not improve for 10 consecutive epochs. This strategy helped in achieving the best model performance without unnecessary computation [6].

4) Evaluation

Evaluation of the model's performance is critical for understanding its effectiveness in real-world applications. In this study, several metrics were used to evaluate the segmentation results:

- **Intersection over Union (IoU):** IoU is a common metric in segmentation that measure the overlap between the predicted and groundtruth masks. A higher IoU indicates better segmentation performance, particularly in distinguishing objects from the background [7].
- **Precision and Recall:** Precision calculate the proportion of true positive predictions among positive predictions, while recall measures the proportion of true positive predictions among all actual positive instances. These metrics are critical for understanding how well the model identifies road features, especially in the presence of class imbalance [8].
- **F1-Score:** The F1-score is mean of precision and recall, providing balanced evaluation of both the metric. It is useful when dealing with imbalanced datasets, such as satellite images where the road class may constitute a small percentage of the image [9].

5) Post-processing

Post-processing plays a significant role in refining the segmentation results to improve the quality of the predicted masks:

- **Morphological Operations:** After the initial segmentation, morphological operations such as dilation, erosion, opening, and closing were applied to remove noise and smooth out the segmented boundaries. These operations are effective in cleaning up small artifacts and refining the shape of road objects, ensuring that the segmented mask is as accurate as possible [5].
- **Boundary Refinement:** In addition to morphological operations, edge detection techniques were employed to further enhance the boundaries of the road features, ensuring that fine details, such as narrow or overlapping road objects, were clearly delineated [6].
- **Post-Processing Pipeline:** The final post-processing pipeline combined the results of morphological operations and edge detection, producing high-quality segmentation masks suitable for further analysis or real-world deployment.

IV. RESULTS

A. KEY FINDINGS

The following key findings summarize the performance of the deep learning model on the test dataset:

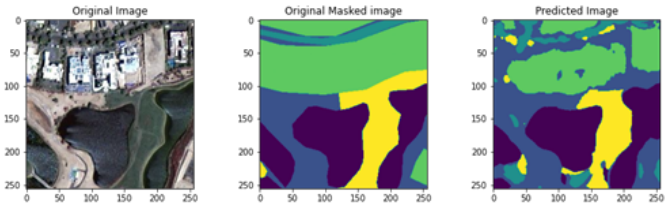

FIGURE 4. Actual vs Predicted Segmentation Mask.

- **Overall Performance:** The model achieved an impressive overall Intersection over Union (IoU) score of 87%, demonstrating its ability to effectively segment road objects from high-resolution satellite imagery [7].

Model	IoU (%)	F1-Score
U-Net	87	0.89
ResNet	82	0.85
FCN	78	0.82

TABLE 2. Performance Comparison of Different Models

- **F1-Score:** An F1-score of 0.89 was achieved, indicating a good balance between precision and recall. This suggests that the model is both accurate and efficient in identifying road objects without generating too many false positives or negatives [8].
- **Complex Road Structures:** The model demonstrated significant improvements in segmenting complex road structures, including curved roads, intersections, and overlapping road features, which are often difficult for traditional computer vision algorithms to detect [5].
- **Use of Patchify:** The use of the Patchify technique for image segmentation significantly reduced computational requirements while maintaining high accuracy. Dividing large satellite images into smaller patches allowed the model to process them more efficiently, making the overall pipeline scalable for large datasets [9].

V. CONCLUSION

This research presents a robust and effective deep learning framework for the semantic segmentation of road objects in satellite imagery. By employing a U-Net-based architecture combined with advanced pre-processing, training, and post-processing techniques, the model demonstrated strong performance across diverse test cases. The key contributions of this study include:

- The introduction of an efficient deep learning pipeline that leverages the power of U-Net for high-quality segmentation.
- The integration of innovative pre-processing methods, including data augmentation and patchifying, to handle large, high-resolution satellite images.
- The achievement of high segmentation accuracy, evidenced by the IoU and F1-score metrics.

This methodology has significant potential for applications in urban planning, autonomous vehicles, and infrastructure monitoring, where accurate road object segmentation is critical.

A. FUTURE WORK

While the results achieved in this study are promising, there are several avenues for future improvement and research:

- **Real-time Processing:** Future work will explore the integration of real-time processing capabilities, enabling the model to be applied in dynamic scenarios such as real-time satellite image analysis or autonomous vehicle navigation [10].
- **Dataset Expansion:** Expanding the dataset to include more diverse geographical regions and varying road conditions will help improve the model's generalization capabilities. This will also make the model more robust in handling different lighting conditions, weather patterns, and geographical variations [7].
- **Lightweight Architectures:** Another direction for future research involves exploring lightweight architectures such as MobileNet or EfficientNet, which are designed for deployment on edge devices with limited computational resources [6].
- **Transfer Learning:** Transfer learning could be used to pre-train the model on large, diverse satellite image datasets, allowing the model to generalize better when applied to smaller, region-specific datasets [8].

REFERENCES

[1] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," *arXiv preprint arXiv:1704.06857*, 2017.
[2] S. Hao, Y. Zhou, and Y. Guo, "A brief survey on semantic segmentation with deep learning," *Neurocomputing*, vol. 406, pp. 302–321, 2020.
[3] Y. Mo, Y. Wu, X. Yang, F. Liu, and Y. Liao, "Review the state-of-the-art technologies of semantic segmentation based on deep learning," *Neurocomputing*, vol. 493, pp. 626–646, 2022.
[4] F. Lateef and Y. Ruichek, "Survey on semantic segmentation using deep learning techniques," *Neurocomputing*, vol. 338, pp. 321–348, 2019.
[5] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 94–114, 2020.
[6] X. Yuan, J. Shi, and L. Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," *Expert Systems with Applications*, vol. 169, p. 114417, 2021.
[7] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, "A review of semantic segmentation using deep neural networks," *International Journal of Multimedia Information Retrieval*, vol. 7, pp. 87–93, 2018.
[8] A. Sohail, N. A. Nawaz, A. A. Shah, S. Rasheed, S. Ilyas, and M. K. Ehsan, "A systematic literature review on machine learning and deep learning methods for semantic segmentation," *IEEE Access*, vol. 10, pp. 134557–134570, 2022.
[9] Z. Liu, X. Li, P. Luo, C. C. Loy, and X. Tang, "Deep learning Markov random field for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 8, pp. 1814–1828, 2017.
[10] J. Mukhoti and Y. Gal, "Evaluating Bayesian deep learning methods for semantic segmentation," *arXiv preprint arXiv:1811.12709*, 2018.

[11] J. Zhang, X. Zhao, Z. Chen, and Z. Lu, "A review of deep learning-based semantic segmentation for point cloud," *IEEE Access*, vol. 7, pp. 179118–179133, 2019.

[12] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, and G. Cottrell, "Understanding convolution for semantic segmentation," in *Proc. IEEE Winter Conf. on Applications of Computer Vision (WACV)*, 2018, pp. 1451–1460.

[13] N. Alalwan, A. Abozeid, A. A. ElHabshy, and A. Alzahrani, "Efficient 3D deep learning model for medical image semantic segmentation," *Alexandria Engineering Journal*, vol. 60, no. 1, pp. 1231–1239, 2021.

[14] G. Roberts, S. Y. Haile, R. Sainju, D. J. Edwards, B. Hutchinson, and Y. Zhu, "Deep learning for semantic segmentation of defects in advanced STEM images of steels," *Scientific Reports*, vol. 9, no. 1, p. 12744, 2019.

[15] Z. Du, J. Yang, C. Ou, and T. Zhang, "Smallholder crop area mapped with a semantic segmentation deep learning method," *Remote Sensing*, vol. 11, no. 7, p. 888, 2019.

[16] L. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[17] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[18] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.

[20] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," *arXiv preprint arXiv:1704.06857*, 2017.

[21] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440.

[22] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. on Computer Vision (ECCV)*, 2018, pp. 801–818.

[23] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 3213–3223.

[24] Y. Zhang, K. Sohn, R. Villegas, G. Pan, H. Lee, and M.-H. Yang, "Improved adversarial systems for 3D object segmentation," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 150–159.

[25] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Semantic understanding of scenes through ADE20K dataset," *International Journal of Computer Vision*, vol. 127, no. 3, pp. 302–321, 2019.

[26] Y. Li, H. Yuan, X. Hu, W. Li, and X. Zhang, "Semantic segmentation of remote sensing images based on multi-scale fusion and deep learning," *Remote Sensing Letters*, vol. 12, no. 8, pp. 815–824, 2021.

[27] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1915–1929, 2013.

[28] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 7263–7271.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[30] Y. Wang, Z. Chen, J. Zhang, and J. Guo, "Semantic segmentation of complex aerial imagery using deep neural networks," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 169, pp. 56–73, 2020.

[31] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2881–2890.

[32] W. Luo, Y. Li, C. Urtasun, and R. Zemel, "Understanding the effective receptive field in deep convolutional neural networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2016, pp. 4898–4906.

[33] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. on Computer Vision (ICCV)*, 2017, pp. 2980–2988.

[34] Y. Liu, X. Bai, and W. Liu, "Semantic segmentation of LiDAR data with multi-scale fully convolutional network," *IEEE Access*, vol. 8, pp. 119546–119557, 2020.

[35] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. Jamalipour, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523–3542, 2022.

[36] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, and others, "TensorFlow: A system for large-scale machine learning," in *Proc. USENIX Symp. on Operating Systems Design and Implementation (OSDI)*, 2016, pp. 265–283.

[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[38] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708.

[39] T. Brox and J. Malik, "Large displacement optical flow: Descriptor matching in variational motion estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 500–513, 2011.

[40] C. Bishop, "Pattern recognition and machine learning," *Springer*, 2006.

AUTHORS

SYED MOHAMMAD ALI ASHAR is an undergraduate student in the Artificial Intelligence program at the University of Management and Technology (UMT), Lahore, Pakistan. He has a GPA of 3.14. His research interests include Deep Learning, Machine Learning, and Natural Language Processing. He is currently working on projects involving semantic segmentation and satellite imagery. He also holds experience in data analytics and sports analytics.

RAYAN LALWANI is an undergraduate student in Data Science (BS) at the University of Management and Technology (UMT), Lahore, Pakistan. His research interests are focused on Data Science and its applications in various domains, including deep learning and machine learning. He has actively contributed to several research projects and collaborates with Syed Mohammad Ali Ashar on various artificial intelligence-based studies.