Road Detection from Satellite Imagery Using a U-Net Convolutional Neural Network

Original Scientific Paper

Asmaa Abdul Jabbar*
Mustansiriyah University,
College of Science, Department of Computer Science,
Baghdad, Iraq
asmaasadiq@uomustansiriyah.edu.iq

Rana Lateef
College of Science, Department of Cybersecuirty Science,
Baghdad, Iraq
taught.rana.abdalrahman@baghdadcollege.edu.iq, ranacsbaghdad@gmail.com

*Corresponding author

Abstract *– The extraction of information from remote sensing data is essential for numerous fields, including urban planning, transportation and traffic management, disaster response, and monitoring environmental changes. Automatic extraction of road networks from satellite images remains a significant challenge due to their diverse structures and scales. Deep learning models that are based on convolutional neural networks have demonstrated exceptional performance in this semantic segmentation task. In this work, a U-Net model has been developed to accurately extract different types of roads from high-resolution satellite imagery. The model follows classic encoder-decoder architecture with skip connections, trained and tested on the DeepGlobe Road Extraction dataset. The encoder utilizes convolutional and max-pooling layers to capture context, while the decoder employs transposed convolutions for precise localization, leveraging skip connections to recover spatial detail. Quantitative evaluations on this benchmark establish that our model achieves a higher IoU (66.62%) and Precision (87.01%) than existing state-of-the-art methods, while maintaining a comparable F1-Score (71.11%), indicating superior detection accuracy with fewer false positives. The results confirm the effectiveness of the proposed approach for robust road extraction.*

Keywords: *Remote sensing,Road extraction,U-Net,Deep Learning*

Received: October 4, 2025; Received in revised form: November 24, 2025; Accepted: December 16, 2025

1. INTRODUCTION

As remote sensing technology has advanced, remotely sensed data have become a valuable resource, and greater amounts of high-quality imagery have been obtained. Also, High-resolution remote sensing image gathering cycles are getting shorter, providing a wealth of data for automated road extraction. A vital component of both urban and rural infrastructure, road networks are essential for fostering economic growth and raising citizens' standards of living in several domains, such as urban planning, traffic monitoring, disaster emergency response, and environmental monitoring, so an accurate road information extraction has substantial practical application value [1, 2].

Over the years, the road network has facilitated the demands of autonomous map generation for transportation system control and route planning, particularly in isolated and inaccessible locations. All of these requirements relate to the precise, automated road detection system. Navigation technologies, transportation planning, and mapping all benefit from accurate road detection from satellite and aerial photography [3, 4].

High-resolution satellite images (HRSI) are particularly useful for extracting relevant information, such as roads, because they capture finer details. However, auxiliary color bands increase the data volume. Furthermore, noise and errors can be introduced into high-resolution images by adjacent objects, such as trees, buildings, and vehicles, as well as spectral variations. All these factors reduced the accuracy of road extraction in HRSI, prompting researchers to develop numerous computational intelligence-based methods for extracting different road types [5, 6]. Because satellite data is collected from above, another issue is that local road characteristics can be obscured by additional obstacles such as shadows, clouds, trees, or buildings [7].

The key contribution of this work is the development and comprehensive evaluation of a U-Net-based model for road extraction from high-resolution satellite imagery. This work reaffirms the power of well-designed encoder-decoder networks with skip connections for complex geospatial segmentation tasks.

The remainder of this work is structured as follows: the next section reviews related works that utilized the DeepGlobe Road dataset. The third section provides an overview of the U-Net, followed by a description of the proposed U-Net architecture for this study. The proposed model is detailed in section four, while section five presents the dataset and evaluation metrics. Finally, sections six discuss the experimental results and conclusions.

2. RELATED WORK

The revolution in deep neural networks has recently led to advancements in almost every facet of traditional computer vision, including classification, detection, and semantic segmentation. Furthermore, deep convolutional neural networks have been adopted for remote sensing tasks, obviously improving performance [8]. In this context, many studies have addressed the extraction of roads from satellite and non-satellite images using various datasets. In this section earlier approaches in this area will introduce with focusing on satellite imagery and the DeepGlobe road dataset.

In [9], Hou *et* al. suggested a (Conditionally –Unet)C-Unet framework to enhance road extraction from high-resolution remote sensing images. The proposed model utilizes a four-stage process: firstly, an initial segmentation is performed using a standard UNet to capture prominent road features; secondly, the most confidently predicted road areas from the first stage are removed using a fixed threshold. Then, in the third stage, the finer road segments missed by the first network are extracted using the Multi-scale Dense Dilated UNet (MD-UNet). Finally, the complete and refined segmentation map is generated by fusing the results from both the first and third steps. The Massachusetts Road dataset was used to evaluate the proposed model. The main advantages are the innovative use of deliberate erasure to guide complementary learning, the integration of multi-scale dilated convolutions to capture long-range contextual information without losing resolution, and its demonstrated superior accuracy in segmenting thin and complex road networks. However, the multi-stage increase in architectural complexity and computational cost is considered a disadvantage to the proposed system.

In [7], Al-Iiedane *et* al. introduced an automated road extraction from high-resolution satellite imagery using an enhanced DeepLabV3+ model. The authors proposed a Dense Depthwise Dilated Separable Spatial Pyramid Pooling (DenseDDSSPP) module that replaces the standard ASPP module to address challenges such as multi-scale road structures and occlusions. Also, a Squeeze-and-Excitation (SE) block is integrated into the decoder to serve as a channel-wise attention mechanism, helping the model focus on the most relevant features for road extraction. The work was assessed on two datasets (the Massachusetts and DeepGlobe road datasets), and the results indicate that the submitted work is superior to a group of other works in connecting road segments with high accuracy.

The authors in [10] proposed an enhanced semantic segmentation method for road extraction by utilizing a Swin Transformer as a backbone for feature extraction and two parallel decoding branches.The initial road detection is performed in the first branch, then the second branch predicts the road's directional angle at each pixel. The refined segmentation is produced by fusing the two feature sets (the initial mask and angle map). The DeepGlobe road dataset is used to assess the results of the proposed approach. The results indicate the approach's ability to produce roads that are more geometrically consistent, with fewer fractures, smoother edges, and more uniform width. The mian disadvantage is the complexity and computational cost introduced by a large number of parameters and the additional angle prediction task, which requires careful loss function design and hyperparameter tuning, potentially increasing training complexity.

In [11], Akhtarmanesh *et* al. introduced an Attention-Assisted UNet model to address the significant class imbalance inherent in road extraction. The standard U-Net architecture was enhanced by integrating attention blocks into the decoder, which act as a soft attention mechanism, learning to weight feature maps from the encoder to focus on road-containing regions while suppressing irrelevant background information. A preprocessing operation was performed on the DeepGlobe dataset, which includ excluding patches that contained only or predominantly background, and implementing rotational augmentations to create a robust training dataset. The advantages of the suggested approach are its comprehensive analysis of metric errors in identifying failure modes, such occlusions and unclear road edges, in the resulting maps. On the other hand, the metrics analysis indicates the shortcomings of the proposed model, for example, rivers and airstrips appearing as roads. This is due to limited data and the need for further training.

In [12], Pahlevani *et* al. introduced an enhanced model, U-Net (EU-Net), to extract roads by integrating the pre-trained VGG19 architecture into the U-Net encoder. The pre-trained VGG19, with proven capabilities for extracting hierarchical features and initialized with ImageNet weights, was utilized to deliver more effective encoding of the satellite images than the base U-Net. The DeepGlobe Road dataset was used to evaluate the proposed model, and the results presented clear improvements over the base model. The advantages of the proposed method are the speed of training and the balance between performance and computational efficiency, while the key disadvantages are common challenges in method extraction, such as in complex scenes and occlusions.

In [13], Mahara *et* al. presented an improved DeepLabV3+ model for automatic road extraction from high-resolution satellite imagery. The authors proposed a module called Dense Depthwise Dilated Separable Spatial Pyramid Pooling (DenseDDSSPP) which replaces the conventional ASPP module to address issues such as multi-scale road construction and occlusions. This module uses a densely connected combination of depthwise separable and dilated convolutions to capture multi-scale contextual information efficiently. Also, the decoder incorporates a Squeeze-and-Excitation (SE) block, a channel-wise attention mechanism that helps the model focus on the most relevant features for road extraction. The suggested model was assessed using both the Massachusetts and DeepGlobe road datasets using the Xception network as the foundation. The results show that it is more computationally efficient, requiring 62% fewer operations and fewer parameters than the original DeepLabV3+ architecture, and that it beats various state-of-the-art models on important metrics, which include IOU and Precision metric. Even when road parts are partially obscured, as by trees, the model excels at precisely connecting them.

3. U-NET MODEL

Deep learning models are used in many fields, including autonomous driving, medical applications, and road extraction from satellite image segmentation [14]. The structure of U-Net, an enhanced fully convolutional network, resembles the letter "U". Fig. 1 shows the detailed architecture of U-Net, which consists of an encoder and a decoder. In the encoder, image features are extracted by convolutional layers and down-sampling layers (or pooling), while the decoder performs up-sampling of these features, enabling precise localization. Additionally, cross-layer connections between corresponding encoder and decoder layers can help the up-sampling process recover important image details. [2, 15].

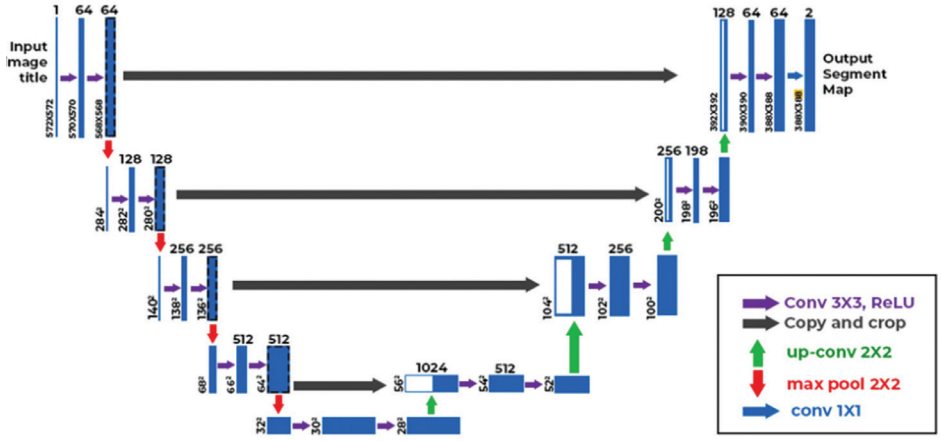

Fig. 1. The U-Net architecture [15]

4. THE PROPOSED MODEL

Fig. 2 illustrates an overview of the proposed U-Net: The input will be raw images of road scenes (street or road views), along with a corresponding mask indicating which parts of the original images are roads. During training, the U-Net learns to map the original images to their corresponding masks. The model adjusts its parameters to minimize the difference between its predicted and ground-truth masks. The output of this model is the predicted masks generated by the U-Net that indicate road areas in the original images. Generally, the model consists of three major parts:

Encoder: this part includes four blocks, each comprising a 2D convolutional mask followed by ReLU to extract hierarchical features; then MaxPooling is implemented to downsample feature maps to reduce spatial dimensions while retaining critical features.

Decoder: This part also includes four blocks, each block includes Conv2dTranspose (or transposed convolution) followed by a concatenate to merge up-sampled features with the corresponding encoder's feature map, then two Conv2D + ReLU layers are employed for feature refinement by rebuilding spatial context and reducing channel depth (e.g., 512 → 256).

Bottleneck: This part serves as a bridge between the encoder (downsampling path) and the decoder (upsampling path). It is located at the lowest level of the U-Net (between the last encoder layer and the first decoder layer). The purpose of the bottleneck is to capture the most abstract/global features (e.g., road topology, intersections) before upsampling, to prevent information loss and ensure the decoder can reconstruct spatial details via skip connections.

Skip connection: In the proposed model, four skip connections are used to ensure that fine details are not lost. Skip connections can introduce finer details by reintroducing them into the decoder. In the model's flowchart, skip connections are shown in different colors and these colors are identical in the encoder and decoder, indicating that detailed features are reintroduced from the encoder to the decoder. Fig. 3 shows these connections separately.

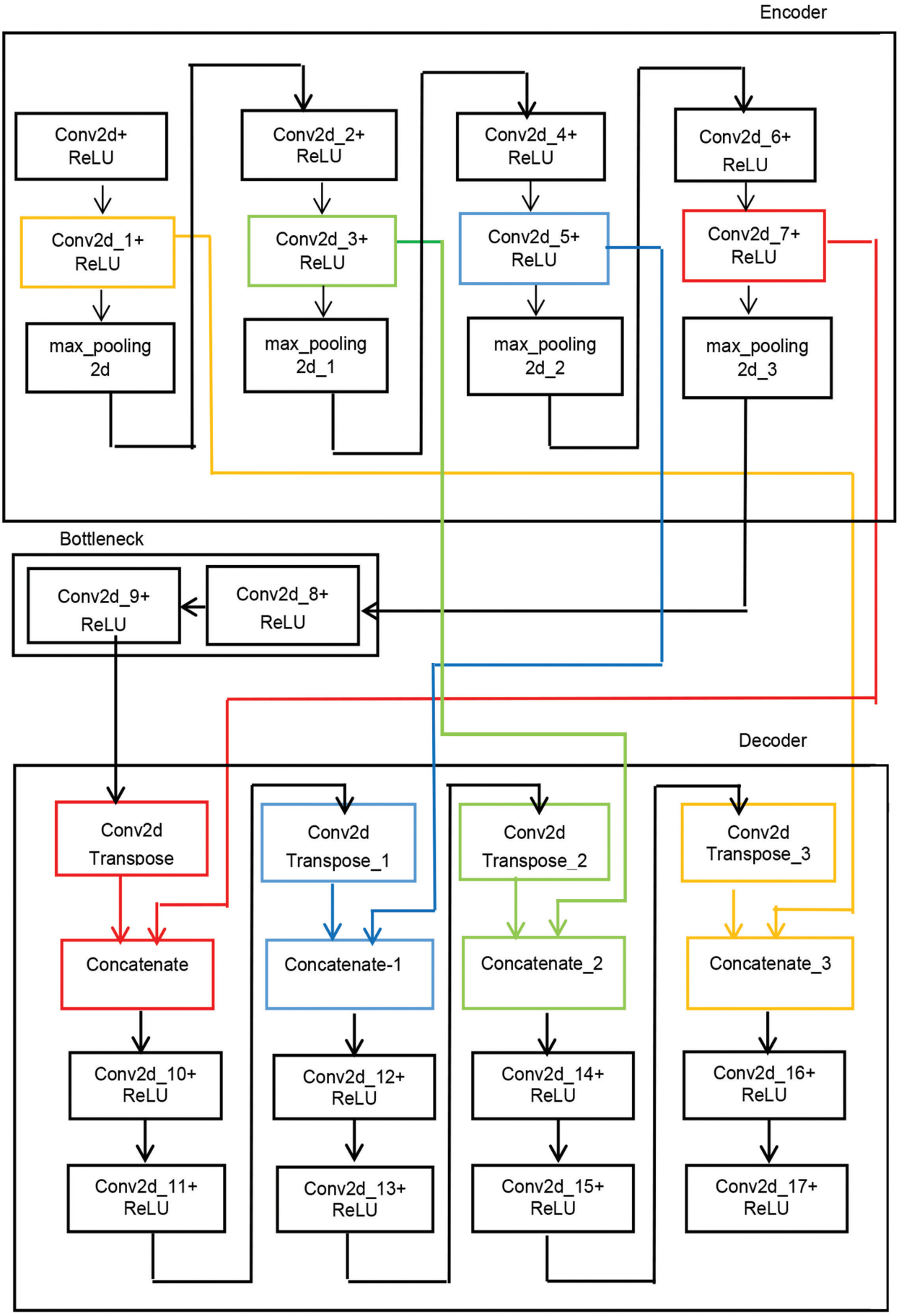

Fig. 2. The structure of the proposed model

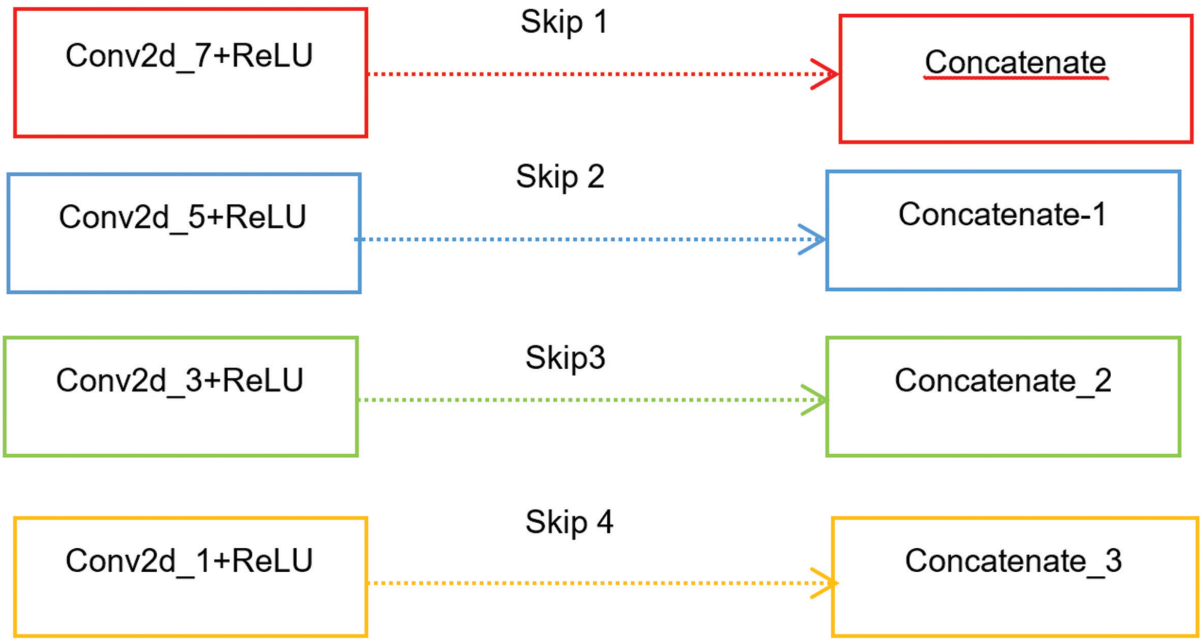

Fig. 3. The skip connection of the proposed model

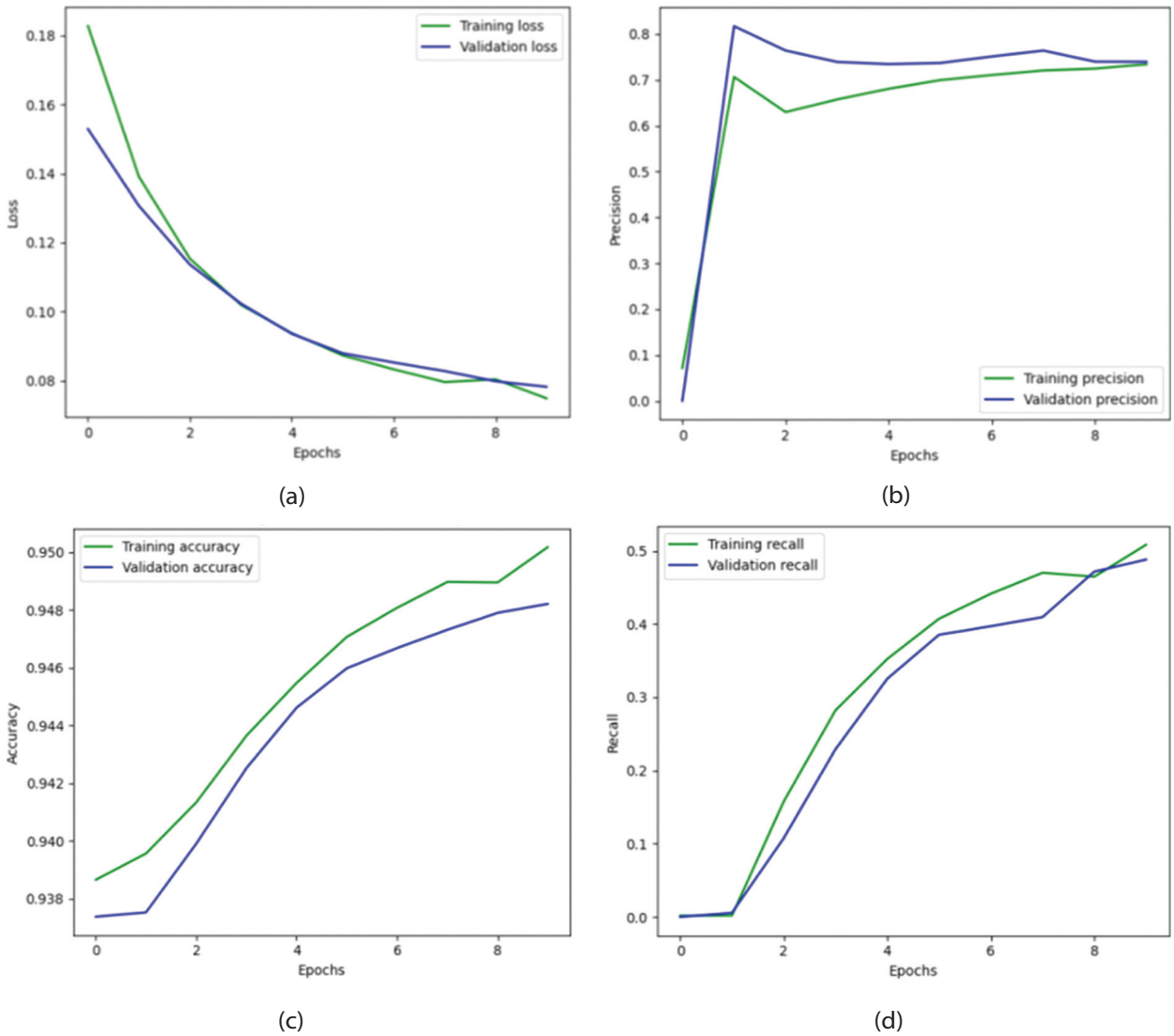

Fig. 4. The performance curves in the training phase. (**a**) Training and validation Accuracy, (**b**) Training and validation Loss, (**c**) Training and validation Precision, and (**d**) Training and validation Recall

5. DATASETS AND EVALUATION METRICS

To test the proposed approach, the Road images dataset, created with DeepGlobe Road Extraction, Alcrowd road segmentation images, and self-annotated images, is used. The images include an original image and its corresponding mask, with the same filenames but in different file formats (PNG and JPEG). The dataset consists of 13.004 road images with their corresponding masks, serving as training images, 1.234 images as validation images, and 1.101 images as test images [13].

Also, to assess the quality of the outcomes from the suggested model for road extraction, pixel-based performance metrics are used, such as IOU, precision, and F1 score. All these metrics depend on the following concepts [16]:

False Positive (FP): An instance of "over-detection," in which an actual non-road pixel is incorrectly identified as a road pixel.

False Negative (FN): An instance of "under-detection," in which an actual road pixel is incorrectly identified as a non-road pixel.

True Positive (TP): A correct detection, where an actual road pixel is accurately identified as a road pixel.

True Negative (TN): A correct non-detection, where an actual non-road pixel is accurately identified as a non-road pixel.

So, IOU, precision, and F1 score utilize TP, FP, FN, and TN to evaluate the model's performance effectively.

Intersection Over Union (IOU): geometrically, it is the overlap between the predicted and ground truth (or actual) road pixels. It measures the ratio of the area of overlap to the area of their union, providing a quantifiable metric for evaluating the accuracy of the predicted segmentation using the following equation [16, 17]:

$$IOU = \frac{TP}{TP + FP + FN} \quad (1)$$

Precision: This metric measures the ratio of correctly extracted road pixels (True Positives) among all extracted road pixels (the sum of True Positives and False Positives). It indicates the model's correctness in positive predictions [18, 19]. Mathematically, it can be expressed as:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

F1 Score: The harmonic mean of Precision and Recall is the F1 Score; it gauges how accurate optimistic predictions are. It can be mathematically represented and simplified as [13]:

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (3)$$

6. EXPERIMENTAL RESULTS AND DISCUSSION

After training on the corresponding training set, we evaluate the performance of the suggested U-Net model on the test set of the DeepGlobe Road dataset. The proposed model was trained using the Adam optimizer with a learning rate of 0.001, a batch size of 16 for 10 epochs. The loss function was binary cross-entropy with dice loss.

Fig. 4 presents the performance curves during the training phase (training and validation images): accuracy, precision, loss, and recall. The training and validation accuracy curves increase sharply and then plateau at an exceptionally high value, nearly 95%. The training and validation loss curves decrease rapidly and smoothly over the epochs, converging to a very low value (~0.08). Finally, both precision and recall curves start low and increase steadily. Like the other metrics, the training and validation curves for each metric are tightly aligned.

Fig. 5 demonstrates the qualitative results in our model's proficiency in predicting roads across diverse schemes, including urban landscapes, vegetated areas, along with complex intersections. This performance is directly attributable to the U-Net architecture. The encoder-decoder structure effectively captures hierarchical features, allowing the model to generalize across different environments. Remarkably, skip connections enable precise localization of roads at multiple scales by reintegrating high-resolution spatial details from the encoder into the decoder upsampling path. This allows for the accurate segmentation of both wide highways and narrow paths. Moreover, the bottleneck layer's large receptive field provides the necessary contextual information to correctly identify and outline complex structures like intersections, as it can comprehend the broader spatial relationships within the scene.

For rigorous comparison, a quantitative evaluation has been performed using the same dataset, with the metrics (IOU, Precision, and F1 score) compared against a recent study (2024 and 2025). The results in Table 1 highlight that our model achieves:

Higher IoU: (64.62% vs. 62.82% and 46%), indicating improved overlap between predicted and ground-truth road segments.

Superior Precision (87.01% vs. 80.36% and 72%), reflecting fewer false positives. Comparable F1 Score (71.11% vs. 71.83%), suggesting balanced precision-recall trade-off.

The numeric results indicate an improvement in metrics, due to:

- Skip Connections in Decoder: By incorporating skip connections in the U-Net decoder, the model concatenates the output with high-resolution features from the encoder. This integration combines coarse semantic information from the bottleneck with fine spatial details extracted from the encoder.

- Enhanced Skip Connections: To further enhance the skip connections, features from earlier encoder layers, such as $Conv2d_1$ and $Conv2d_3$, are concatenated to the decoder. This modification, for instance, by adjusting $Concatenate_3$ to include multi-scale features from both $Conv2d_1$ and $Conv2d_3$, helps in preserving high-resolution details. This preservation of detailed information contributes to more precise road boundary detection.

Fig. 5. The results of the proposed model on different images. (**a**), (**b**), (**c**), and (**d**) represent the satellite images (different urban and residential areas) with the road maps obtained from the proposed model. (**e**), (**f**) (**g**), and (**h**) represent the satellite images (different agricultural and rural areas) with the road maps obtained from the proposed model

Table 1. The results of quantitative evaluation of the proposed model

Study	IOU	Precision	F1 Score
Mahara et al. 2025 [13]	62.82	80.36	71.83
Pahlevani et al. 2024 [12]	46.00	72.00	----
Proposed Model	66.619	87.01	71.11

To study the impact of skipping connections on the proposed model's performance, skip connections are gradually removed, and the effects on performance metrics are summarized in Table 2.

Table 2. The impact of skipping connections of the proposed model

Model Configuration	IOU	Precision	F1 Score
With four skip connection	66.61	87.01	71.83
Remove Skip1	62.00	86.70	68.08
Remove Skip1&2	59.10	86.00	63.23
Remove Skip1&2&3	54.20	85.10	56.22

The results of the metrics in Table 2 indicate that:

Progressive Degradation in Performance: The model with full skip connections achieves the best performance, underscoring the importance of retaining all skip connections; removing them leads to a significant decline in IoU, Precision, and F1 Score.

Cumulative Impact: Each additional skip connection removed results in a further drop in performance, indicating a cumulative contribution of skip connections to model accuracy. Removing the first skip connection leads to a slight decrease in the metrics, indicating that although one connection was removed, there is a noticeable impact on the numerical results. Also, when removing the first and second skip connections, the results indicate a continued decrease in the metrics, suggesting a cumulative effect of each connection on the model's ability to perform road extraction accurately. In addition, removing three skip connections (1, 2, and 3) further demonstrates the cumulative effect of these connections on the accuracy of the extraction method metrics.

Observed Evidence from Results: Table (2) illustrates a reduction in IoU, Precision, and F1 Score as skip connections have been introduced, which underscores their role in enhancing the model's ability to discern road features accurately. For example, the highest F1 Score with skip connections (71.83) was significantly better than the lowest score (56.22) when all skip connections were removed. This contrast points up the contribution of skip connections to overall segmentation accuracy.

Also, the number of key points can be perceived about the impact of skip connections:

Preservation of Spatial Information: Skip connections enable the U-Net architecture to concatenate high-resolution features from the encoder with upsampled features in the decoder. This is critical for tasks like road extraction, where fine spatial resolution is essential for accurate delineation.

Enhanced Gradient Flow: By introducing skip connections, better gradient propagation can be facilitated during training. As a result, the model can learn more effectively, leading to improved performance metrics.

Ultimately, to assess the efficiency of the proposed model, U-Net++ has been implemented to extract roads from the same dataset using the same training setup as in the proposed model.

As shown in Table 3, the results indicate a close match between the two networks, with U-Net++ demonstrating a slight advantage in the Intersection over Union (IoU) metric

Table 3. Results of the comparison between the proposed method and U-Net++

Metric	Proposed U-Net	U-Net++
IoU	66.61	68.01
Precision	87.01	87.23
F1 Score	71.11	71.31

To further improve upon this model, future work will focus on enhancing the visual quality and structural accuracy of the extracted road networks. A natural progression of this research is to adopt a Generative Adversarial Network (GAN) framework after obtaining the initial segmentation from the U-Net; a GAN can be employed to refine the output. This can enhance the quality of the segmentation mask generated by U-Net. It is also possible to evaluate the proposed model on another dataset, such as the Massachusetts Roads Dataset, to assess its generalization and accuracy.

Acknowledgment: The author's thanks the Department of Computer Science", "College of Science,Mustansiriyah University, for supporting this work.

REFERENCES

[1] R. Kamiya, K. Hotta, K. Oda, S. Kakuta, "Road Detection from Satellite Images by Improving U-Net with Difference of Features", Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods, Funchal, Madeira, Portugal, 16-18 January 2018, pp. 603-607.

[2] M. J. Khan, P. P. Singh, "Advanced road extraction using CNN-based U-Net model and satellite imagery", Advances in Electrical Engineering, Electronics and Energy, Vol. 5, 2023, p. 100244.

[3] M. I. Abdulrahman, M. A. Shareef, A. A. Al-Attar, "Deep Learning (CNN) for Detecting Road Infrastructure in Old Mosul City Using High-Resolution Aerial Imagery", Baghdad Science Journal, Vol. 22, No. 3, 2025.

[4] Y. Xu, Z. Xie, Y. Feng, Z. Chen, "Road extraction from high-resolution remote sensing imagery using deep learning", Remote Sensing, Vol. 10, No. 9, 2018, p. 1461.

[5] A. Y. Noori, S. H. Shaker, R. A. Azeez, "Semantic Segmentation of Urban Street Scenes Using Deep Learning", Webology, Vol. 19, No. 1, 2022, pp. 2294-2306.

[6] R. Liu, J. Wu, W. Lu, Q. Miao, H. Zhang, X. Liu, Z. Lu, L. Li, "A review of deep Learning-Based methods for road extraction from High-Resolution remote sensing images", Remote Sensing, Vol. 16, No. 2, 2024, p. 2056.

[7] H. A. Al-liedane, A. I. Mahameed, "Satellite images for roads using transfer learning", Measurement: Sensors, Vol. 27, 2023, p. 100775.

[8] G. Cheng, C. Wu, Q. Huang, Y. Meng, J. Chen, D. Yan, "Recognizing road from satellite images by structured neural network", Neurocomputing, Vol. 356, 2019, pp. 131-141.

[9] Y. Hou, Z. Liu, T. Zhang, Y. Li, "C-UNet: Complement UNet for remote sensing road extraction", Sensors, Vol. 21, 2021, p. 2153.

[10] S. Xiong, C. Ma, G. Yang, Y. Song, S. Liang, J. Feng, "Semantic segmentation of remote sensing imagery for road extraction via joint angle prediction: comparisons to deep learning", Frontiers in Earth Science, Vol. 11, 2023.

[11] A. Akhtarmanesh, D. Abbasi-Moghadam, A. Sharifi, M. H. Yadkouri, A. Tariq, L. Lu, "Road extraction from satellite images using attention-assisted UNet", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 17, 2023, pp. 1126-1136.

[12] M. Pahlevani, F. Z. Pahlevan, R. Rastgoo, "Expanded U-Net Model for Road Extraction from Satellite Images", Modeling and Simulation in Electrical and Electronics Engineering, Vol. 4, 2024, pp. 39-45.

[13] A. Mahara, M. R. K. Khan, L. Deng, N. Rishe, W. Wang, S. M. Sadjadi, "Automated Road Extraction from Satellite Imagery Integrating Dense Depthwise Dilated Separable Spatial Pyramid Pooling with DeepLabV3+", Applied Sciences, Vol. 15, 2025, p. 1027.

[14] A. Maurya, P. Mittal, R. Kumar, "A modified u-net-based architecture for segmentation of satellite images on a novel dataset", Ecological Informatics, Vol. 75, 2023, p. 102078.

[15] S. K. Hussein, K. H. Ali, "Semantic segmentation of aerial images using u-net architecture", Iraqi

Journal for Electrical and Electronic Engineering, Vol. 18, 2021, pp. 58-63.

[16] V. Yerram, H. Takeshita, Y. Iwahori, Y. Hayashi, M. K. Bhuyan, S. Fukui, B. Kijsirikul, A. Wang, "Extraction and calculation of roadway area from satellite images using improved deep learning model and post-processing", Journal of Imaging, Vol. 8, No. 5, 2022, p. 124.

[17] J. M. Adam, W. Liu, Y. Zang, M. K. Afzal, S. A. Bello, A. U. Muhammad, C. Wang, J. Li, "Deep learning-based semantic segmentation of urban-scale 3D meshes in remote sensing: A survey", International Journal of Applied Earth Observation and Geoinformation, Vol. 121, 2023, p. 103365.

[18] Q. Xu, C. Long, L. Yu, C. Zhang, "Road extraction with satellite images and partial road maps", IEEE Transactions on Geoscience and Remote Sensing, Vol. 61, 2023, pp. 1-14.

[19] A. H. Alwan, S. A. Ali, A. T. Hashim, "Medical Image Segmentation Using Enhanced Residual U-Net Architecture", Mathematical Modelling of Engineering Problems, Vol. 11, 2024, pp. 507-516.