Expanded U-Net Model for Road Extraction from Satellite Images

Mahdi Pahlevani[1], Fatemeh Zahra Pahlevan[1], and Razieh Rastgoo[1,*]

[1] Electrical and Computer Engineering Department, Semnan University

***Abstract*-- Reliable extraction of information from aerial images is a challenging issue with many practical applications. One of the specific challenges within this problem is the automatic detection of roads. Due to the presence of shadows, obstructions, and a wide variety of non-road objects, this task is considered as a complex problem in computer vision. Despite the previous efforts in the field of automatic road detection, there is still room for improving in this area. This paper aims to enhance detection accuracy by proposing a model for road segmentation in satellite images based on image segmentation techniques. To this end, we introduce a novel model, namely Expanded U-Net (EU-Net) by embedding the VGG19 layers to the base U-Net model. Evaluation results on the DeepGlobe Road Extraction dataset indicate enhancements in results compared to a base U-Net model.**

***Keywords*-- U-Net, VGG, Road Extraction, Image Segmentation, Satellite Images.**

I. Introduction

Roads are crucial man-made features that need to be accurately extracted for map production from satellite images. With advancements in satellite technology, both in spatial and spectral resolution, satellite images have significantly improved and are readily accessible over short time intervals. Therefore, Automatic road extraction has become a major challenge in remote sensing and photogrammetry. Moreover, having an up-to-date road map is of great importance for providing many essential services. For instance, a city needs accurate road maps for routing emergency vehicles, while a navigation system based on GPS relies on the same information to provide users with the best routes [1]. Since new roads are often constructed, keeping road maps up-to-date is a significant challenge. Currently, road maps are constructed and updated based on high-resolution aerial images. Given that very large areas need to be taken into consideration, the updating process is both costly and time-consuming [2]. Moreover, the topology of roads is complex in satellites. Generally, current models in road extraction from satellite images can be categorized into two groups: conventional models and deep learning models. The first group employs conventional machine learning models for image processing and road detection. The proposed models in this group have mainly focused on road segmentation, multiresolution analysis, and multiresolution combination. While promising results have been obtained using these models, the challenges of handcrafted features and filters in these models remain. On the other hand, the second group, uses the recent advances in Deep Learning models. Considering the strong capability of Convolution Neural Network (CNN) in image processing, this model is one of the most-used models of deep learning models in road detection from satellite images. Although deep learning techniques are generally more efficient than classical methods, they face challenges in accurately extracting road segments under complex conditions, particularly when occlusion impedes visibility.

Over the years, Artificial Intelligence, especially deep learning techniques, has obtained state-of-the-art performance in many tasks [3-12]. Among the deep learning models, the U-Net architecture has particularly demonstrated remarkable success in accurately delineating roads from images. Recent studies have proposed various variants and improvements to the U-Net model, enhancing its performance in road extraction tasks [13, 14]. Rather than using the standard rectified linear unit (ReLU) activation function, researchers have enhanced their proposed technique by employing the exponential linear unit (ELU) activation function, which boosts overall effectiveness by eliminating erroneous road pixels. Additionally, they have augmented the training data by progressively rotating images in eight phases. This suggested technique outperforms older, state-of-the-art methods for extracting roads from remotely sensed imagery. In another study, a nonlocal link network incorporating nonlocal blocks (NLBs) was utilized, further advancing the capabilities of deep learning algorithms, particularly in image processing. After preprocessing the dataset to address issues like the correlation of spatial and geometric information based on different road structures and properties, a U-Net architecture-based CNN model was used to extract roads from remotely sensed data. Aiming to enhance the detection accuracy of road detection in satellite images, we propose a novel model, namely the Expanded U-Net (EU-Net), for automatic road detection by embedding the VGG model in the U-Net architecture. More specifically, we incorporated pre-trained VGG19 blocks into the U-Net structure. This modified model exhibits accelerated learning compared to the original U-Net model at the same convergence point. Ultimately, this enhancement results in improved performance during prolonged training phases. Our contributions can be listed as follows:

- Model: Aiming to enhance the detection accuracy of road detection in satellite images, we propose a novel model,

[*] Corresponding Author Email Address: rrastgoo@semnan.ac.ir

namely the Expanded U-Net (EU-Net), for automatic road detection by embedding the VGG model in the U-Net architecture. To the best of our knowledge, this is the first time that such a model has been proposed using the combination of U-Net and VGG for road segmentation from satellite images.

- Performance: The proposed model is evaluated using different evaluation metrics on the DeepGlobe Road Extraction dataset, outperforming the base U-Net model.

The remainder of the paper is arranged as follows. Section 2 provides a brief review of previous studies on identifying distracted driver behavior using machine learning algorithms. Section 3 presents a brief review of some related concepts used in the model. Details of the proposed model are described in section 4. Results are discussed in section 5. Finally, section 6 concludes this work with future opportunities.

II. Related Work

In this section, a brief review of the related work in road detection will be presented. Learning-based approaches for road detection are not entirely new, and researchers have suggested some models to predict whether a specific area in an image is a road or not using features extracted from its surroundings. While these models show progress, they have also encountered several main challenges, such as limited training data, feature deficiency, spatial dependencies, and synthesis labels [15]. Aiming to tackle these challenges and improve the area, we briefly review recent works in road detection from satellite images. Some studies have explored the extraction of road networks from satellite imagery, employing advanced techniques such as the U-Net architecture. For instance, Ying Wang et al. introduce the dual-decoder-U-Net (DDU-Net) model for improved small-sized road extraction from High-resolution Remote Sensing Images (HRSIs) [16]. An end-to-end change detection method employing the U-Net++ architecture is introduced in [17], addressing error accumulation issues by enabling the direct learning of change maps from co-registered image pairs. This method leads to superior performance on very high-resolution satellite image datasets. A unique approach for detecting changes in high-resolution images is introduced by J—Chen *et al*, known as dual attentive fully convolutional Siamese networks. By incorporating a dual attention mechanism, the model captures long-range dependencies to enhance feature representations, addressing the robustness issue against pseudo-change information prevalent in existing methods [18]. Some studies have been conducted in the realm of image segmentation using generative models. For instance, Wang et al. explore image super-resolution techniques, utilizing deep learning advancements. The proposed Ultra-Dense GAN (udGAN), integrating DenseNet, achieves superior perceptual and quantitative results in extensive experiments on benchmark datasets and real-world satellite imagery, demonstrating enhanced super-resolution performance [19]. Zeng et al. conducted a comprehensive review of methods for detecting vegetation phenology through the analysis of time-series multispectral remote sensing imagery. General land surface phenology stages and advanced species-specific phonological stages are covered, with a detailed discussion of common data processing methods [20]. A fully attentional model for general Satellite Image Time Series (SITS) processing based on the Vision Transformer (ViT) has been proposed by Tarasiou et al. The model's discriminative power is enhanced by the introduction of two novel mechanisms, namely acquisition-time-specific temporal positional encodings and multiple learnable class tokens. The effects of all novel design choices are evaluated through an extensive ablation study [21]. In this paper, we introduced the EU-Net model for satellite image segmentation, aiming to achieve higher speed and accuracy compared to the fundamental U-Net architecture.

III. A Brief Review of the Related Concepts

In this section, we present a brief review of the adopted deep learning models: U-Net and VGG19.

A. *U-Net Model*

An encoder-decoder CNN, known as U-Net [22], is utilized extensively in diverse domains, including medical imaging [23], autonomous driving [24], and satellite imaging [25]. The understanding of U-Net's segmentation process is deemed essential, considering that subsequent architectures are formulated based on the same underlying principles. The objective of this investigation is to elucidate how image segmentation is carried out by U-Net. Emphasis will be placed on the application of U-Net to the task of brain image segmentation [26], thereby contributing to an enhanced comprehension of its operational mechanisms. The fundamental U-Net model that is recommended for our image segmentation task is shown in Figure 1.

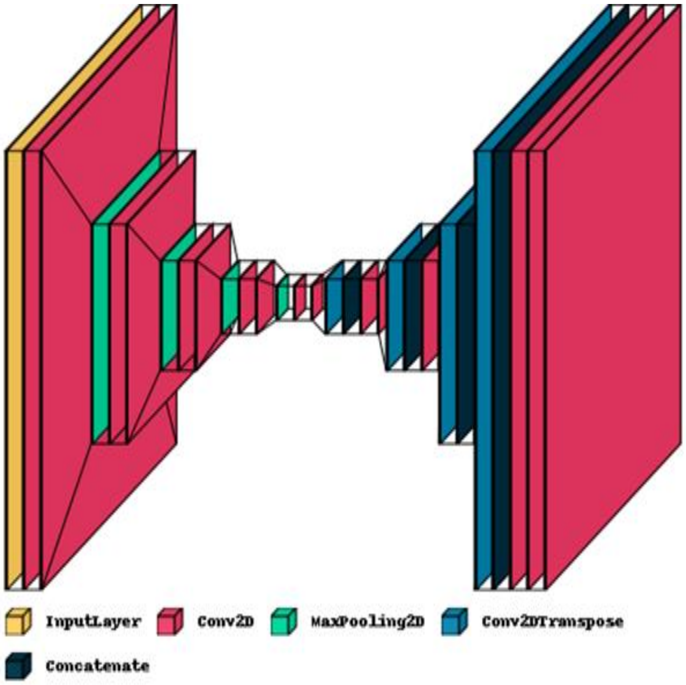

Fig. 1. Basic U-Net Model adopted on our dataset for image segmentation

B. VGG19 model

VGG19 is a deep CNN architecture, widely used for image classification tasks [27]. It consists of 19 layers, including 16 convolutional and 3 fully connected layers, organized sequentially. VGG19 is characterized by its simplicity and uniform architecture, employing 3x3 convolutional filters throughout the network. It has shown impressive performance in various computer vision tasks, making it a popular choice in deep learning research and applications. The model's depth and straightforward structure contribute to its effectiveness in capturing complex hierarchical features from input images [28]. Figure 2 demonstrates VGG19 blocks.

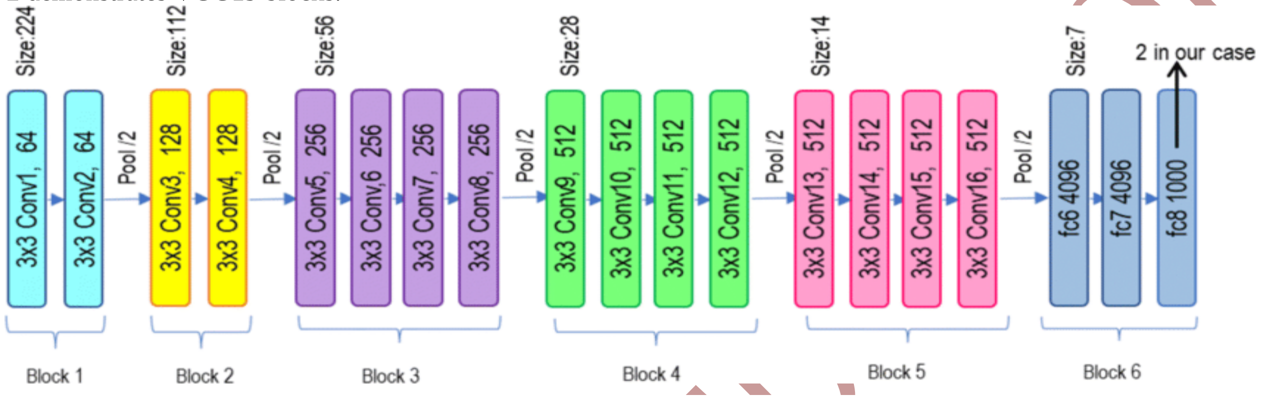

Fig. 2. VGG19 blocks

IV. THE PROPOSED APPROACH

In this section, we present the proposed model in detail. Figure 3 shows an overview of the proposed model, including four main steps: data pre-processing, model training, model prediction, and model evaluation. Details of these steps are explained in the following.

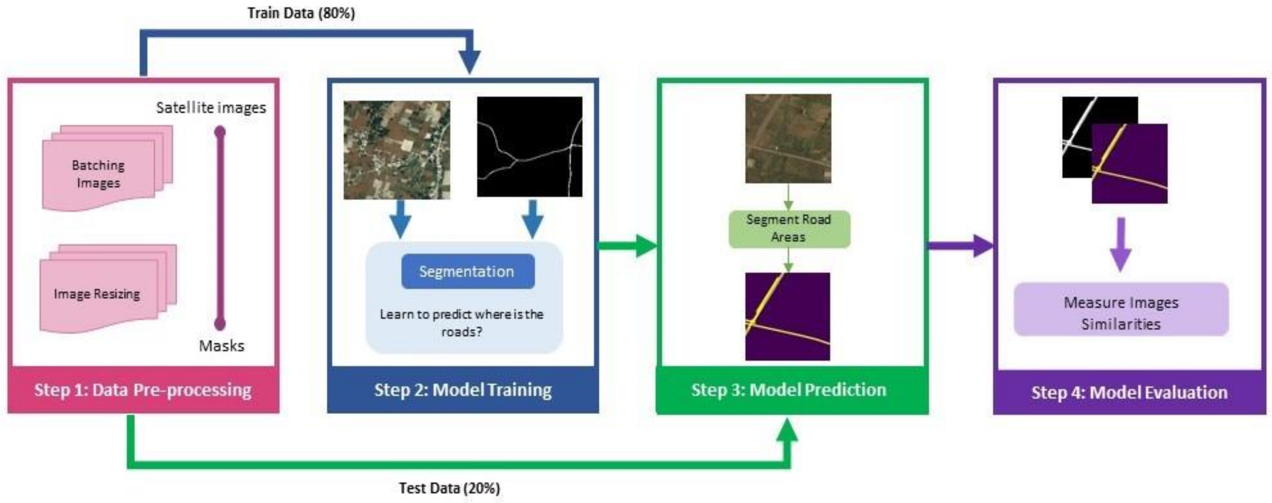

Fig. 3. An overview of the proposed approach, including four main steps: data pre-processing, model training, model prediction, and model evaluation.

A. Data pre-processing

In this initial phase, we apply various preprocessing techniques to the input data, such as resizing. Furthermore, to address resource limitations effectively, we organize the data into batches. This batching strategy optimizes the utilization of available resources throughout the experimental procedures, promoting efficiency in processing. Subsequently, the images undergo segmentation into distinct sets for training, validation, and testing purposes. This segmentation facilitates systematic evaluation and validation of our methodologies across diverse datasets, enhancing the robustness of our experimental outcomes.

B. Model

We have proposed a novel model, entitled EU-Net, which integrates VGG layers into its encoder section. This architectural innovation has yielded notable improvements in convergence during segmentation tasks, outperforming the standard U-Net architecture. Unlike the standard U-Net model, our approach exhibits variability in the number of layers within each decoder step, a consequence of the differing layer counts in the convolutional blocks of the VGG architecture. Furthermore, our model's bridge section incorporates the final convolutional block of the VGG model, enhancing feature representation. Although the decoder section of our model shares similarities with the basic U-Net architecture, a significant distinction lies in the depth of convolution layers. Specifically, while the conventional U-Net employs three convolution layers in its decoder section, our model integrates four convolution layers, providing enhanced feature extraction and representation capabilities. Figure 4 illustrates the overall architecture of the proposed model.

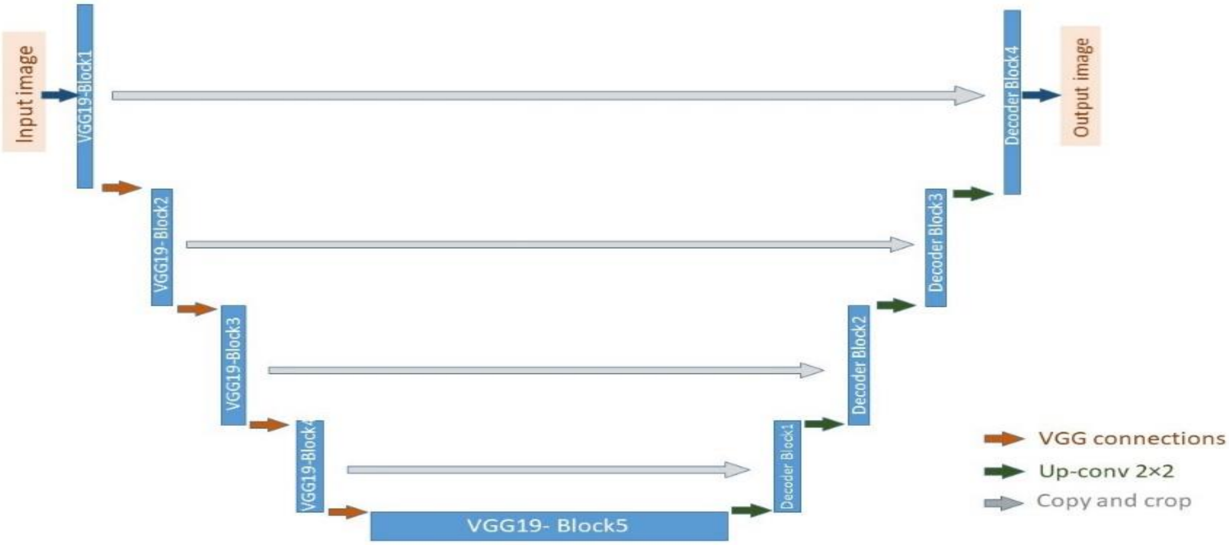

Fig. 4. Overall architecture of EU-Net model

For road segmentation in the U-Net architecture, the VGG network was chosen as the feature extractor for several significant reasons:

1. Proven Performance in Vision tests: VGG's deep architecture with modest convolutional filters (3x3) has contributed significantly to its consistently strong performance in picture segmentation and classification tests. Its simplicity and capacity to extract detailed hierarchical characteristics from satellite data make it a potent contender for road segmentation.
2. Benefits of Transfer Learning: Pre-trained VGG models on massive datasets such as ImageNet offer a substantial feature generalization advantage. The model can benefit from strong low-level and mid-level feature representations by employing these pre-trained weights. These representations are essential for detecting roads in satellite photos, where datasets may be more specialized or smaller.
3. Compatibility with U-Net design: Because of its downsampling properties, which match the encoder-decoder design of U-Net, the structure of VGG is especially well-suited to U-Net. Together, these synergies enable VGG to capture multi-scale features that are critical for semantic segmentation, particularly in tasks requiring the preservation of fine details and global context, such as road detection.

While other models for feature extraction, such as ResNet, DenseNet, or EfficientNet, might also be taken into consideration, VGG provides the best trade-off between computational efficiency and performance. For this specific application, for instance, ResNet and DenseNet add more depth and complexity, which could result in higher computational costs without appreciable gains in performance. The goal of this research was to create an efficient and successful road segmentation model, and VGG offers the tools needed to achieve this goal.

The main contribution of this work is the creation of the Expanded U-Net (EU-Net) model, which improves road segmentation in satellite pictures by integrating VGG19 layers into the conventional U-Net design. The VGG model and its linked layers are primarily used for the input picture encoding. This enables the model to make use of strong, hierarchical feature representations that have been acquired from a sizable dataset (ImageNet). One of the main issues with satellite data is reliably detecting road boundaries amidst varied terrains and occlusions. This integration greatly enhances the model's ability to differentiate roads from various background features.

In particular, we focused on the decoder section's architecture to make sure it could dynamically adapt and function as best it could for our particular goals. Four convolutional layers, carefully sized to seamlessly concatenate with the preceding layers, make up the decoder part. This improvement enables more accurate segmentation maps, collecting finer details that are essential for distinguishing between different road structures. Additionally, after every convolutional operation, we incorporated a dropout layer to mitigate the danger of overfitting. Furthermore, a batch normalization layer was added after dropout to handle possible problems with inflating gradients and stabilize the learning process. Considering that the dataset contains a lot of areas with black masks, these careful procedures are necessary to ensure reliable performance and efficient learning. Furthermore, the EU-Net architecture uses pre-trained VGG19 weights in addition to these architectural advances, which speeds up the training process and allows the model to converge more quickly than with the conventional U-Net. Overall, the combination of VGG19 and U-Net in the EU-Net model outperforms the basic U-net in road segmentation and provides a more effective and efficient solution for real-world applications in satellite image analysis.

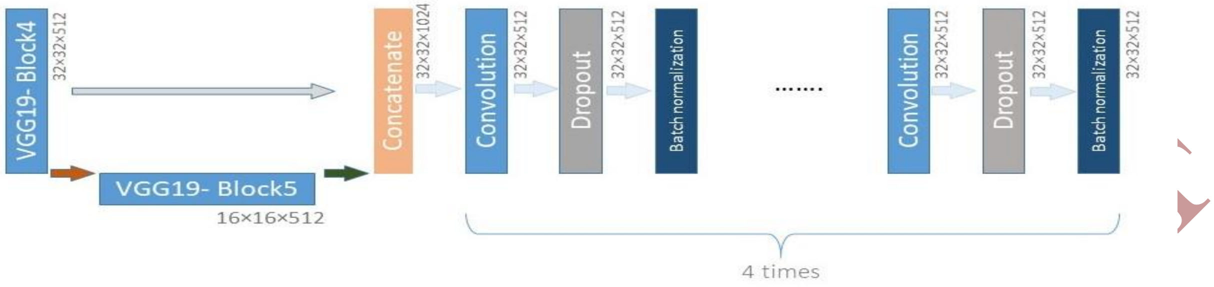

Fig. 5. The first decoder layer of proposed model in details

Figure 5 provides a comprehensive depiction of the intricacies within the initial decoder of the model. Each decoder receives inputs from both the encoder layers and the final decoder block. Subsequently, four convolutional layers, along with normalization and dropout layers, are applied. Notably, in the first decoder block, inputs comprise the bottleneck and the last encoder block. The output of the bridge undergoes upsampling and concatenation with the VGG19-Block4 output. The subsequent decoder blocks follow a similar structure to this block, albeit with adjustments in input and output sizes.

C. *Segmentation*

The EU-Net architecture, tailored for image segmentation, integrates an encoder path for feature extraction and a decoder path for upsampling. A bottleneck layer serves as a bridge between the two preserving spatial information. Skip connections concatenate features from corresponding encoder and decoder layers to retain fine-grained details. The final layer employs a Softmax activation function to generate a segmentation map. Training is conducted with proper loss-function to optimize model performance. This design allows EU-Net to effectively capture both local and global information, making it particularly suitable for our task.

D. *Model evaluation*

Throughout both the training and testing phases, the model undergoes evaluation using various metrics. Among these, IoU[1] stands out as the primary metric utilized to gauge image similarities in segmentation tasks [29].

IoU: By measuring the ratio between the intersection and the union of two sets, the Intersection over Union (IoU) is calculated. Here is how the IoU formula can be stated:

$$IoU(X,Y) = \frac{|A \cap B|}{|A \cup B|} \qquad (1)$$

The loss function employed in this model is the Dice loss, which correlates with the Dice coefficient metric.

Dice coefficient: The relative size of the intersection of two sets, X and Y, about the overall size of the sets is captured by this coefficient [30]. As shown in Equation (2).

$$Dice(X,Y) = \frac{2 * |A \cap B|}{|A| + |B|} \qquad (2)$$

Moreover, the model undergoes evaluation using other established metrics such as precision and recall, which are commonly employed in assessing segmentation performance.

V. Experimental Analysis and Results

In this section, we explain the details of the obtained results corresponding to the proposed model. To further evaluate the performance of the proposed EU-Net model, experiments were conducted using ResNet-based weights as an alternative feature extractor. This comparison aims to understand how different architectures impact segmentation accuracy, computational efficiency, and overall performance. To this end, the implementation details, dataset, split strategy of the dataset, evaluation metrics, ablation analysis, comparison with base model, and discussion on the results are discussed in details.

A. *Implementation details*

The implementation of the proposed model has been done in the Python programming language, TensorFlow library, and Google Colab platform. As previously mentioned, we used the EU-Net model, which is designed for image segmentation tasks. In the model, we employed the Shallow and VGG19 encoder, and the model weights are initialized with ImageNet weights. The activation function used is the sigmoid function in the last layer and the other layers are the ReLU activation function. Additionally, the optimizer used is Adam with an initial learning rate of 0.001. Also, we set the reduce learning rate function to have dynamic learning rate. The patience for

[1] *Intersection over Union*

the reducing learning rate is 5 epochs. Additionally, the early stopping function was set to stop the learning process when the learning process has not improved.

B. Dataset

In this research, the DeepGlobe Road Extraction dataset has been utilized [31]. This dataset includes 8570 different images of satellite maps. Each image is accompanied by its corresponding mask, which is used for evaluating accuracy. The images have been collected from various locations around the world. Some samples from the dataset are shown in Figure 6.

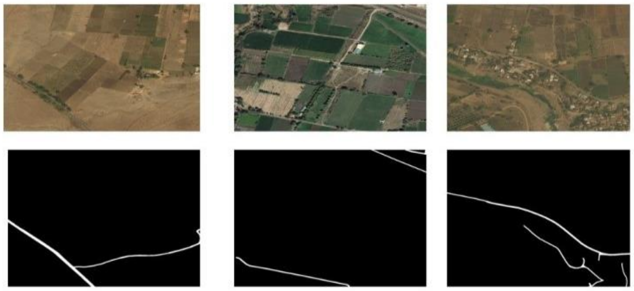

Fig. 6. Some samples of Dataset

C. Preprocessing and Splitting Strategy

The preprocessing steps for the images in the DeepGlobe Road Extraction dataset are as follows:

1. Both the input images and corresponding masks undergo a resizing process, wherein they are adjusted to dimensions of 256x256 pixels. This resizing ensures uniformity in data dimensions, facilitating consistency and ease of processing across the dataset.

2. The images are then split into training, validation, and test sets. Specifically, there are 6,427 samples for the training set, 1,285 samples for the validation set, and 1,114 samples for the test set. Then images were batched with the size of 8.

D. Ablation analysis

The ablation study was conducted to evaluate the performance improvements obtained by integrating different feature extraction methods into the U-Net architecture. Specifically, we analyzed three configurations: the basic U-Net, U-Net enhanced with VGG19 weights (referred to as EU-Net), and U-Net integrated with ResNet weights. The objective was to determine the effect of using pre-trained weights from different architectures on segmentation accuracy, training efficiency, and model robustness.

The training results, summarized in **Table 1**, reveal that both the EU-Net (with VGG19) and U-Net with ResNet weights outperform the basic U-Net model in terms of segmentation metrics. By utilizing pre-trained feature extractors, these models could leverage strong low-level and mid-level feature representations, improving their ability to identify road structures across varying conditions in satellite imagery. However, there are distinct differences in performance between the VGG19 and ResNet versions.

One of the key benefits observed with EU-Net was its faster convergence compared to the basic U-Net and ResNet-enhanced U-Net. As seen in **Table 2**, the VGG19 model's simpler, more straightforward architecture allowed for a more efficient use of computational resources, leading to reduced training time while maintaining high accuracy. In contrast, the ResNet-based model, though effective, required more epochs to reach similar accuracy levels due to its deeper and more complex structure, which may introduce additional computational overhead. The outcomes are summarized in **Table 1** and **Table 2**, which show the performance of the three models during training and validation.

TABEL 1
Outcomes from both models during training

Measure	*Basic U-Net model*	*Expanded U-Net*	*ResNet U-net*
Training IoU	0.80	0.92	0.84
Training Loss	0.20	0.10	0.17
Validation IoU	0.63	0.75	0.69
Precision	0.75	0.80	0.76
Recall	0.65	0.73	0.67
Validation Loss	0.39	0.29	0.34

TABEL 2
Outcomes from the evaluation of both models on the test dataset

Measure	*Basic U-Net model*	*Expanded U-Net*	*ResNet U-net*
IoU	0.46	0.56	0.51
Dice Loss	0.39	0.30	0.33
Precision	0.72	0.79	0.74
Recall	0.63	0.70	0.68

The U-Net with VGG19 weights showed significant improvements in both training and validation metrics, as presented in **Table 1** and **Table 2**, indicating that the VGG-based feature extractor effectively captures the necessary spatial features while maintaining a balance between depth and computational complexity. This balance is crucial for ensuring that the model remains feasible for real-world applications that require efficient training and fast inference times.

The ResNet-based U-Net also demonstrated competitive performance, particularly in capturing more complex patterns due to its deeper architecture. However, this came at the cost of increased training time (as seen in **Table 1**) and slightly lower validation accuracy (**Table 2**). The additional depth in ResNet allows it to learn more sophisticated features but also introduces the risk of overfitting, which may explain the marginally lower validation IoU compared to the VGG-based model.

E. Discussion

In this paper, we proposed a new model based on the U-Net architecture to segment satellite images into road regions, enabling the extraction of valuable information for analysis and interpretation. One of the key challenges encountered during the segmentation process with the basic U-Net model was the lack of meaningful feature extraction during the training, particularly in encoding input images. To address this, we introduced two enhanced versions: the EU-Net model, which integrates VGG19 weights, and an alternative model incorporating ResNet weights. The EU-Net model (with VGG19 weights) achieved higher IoU, precision, and recall scores compared to the ResNet-based U-Net, while also demonstrating faster convergence times. This suggests that the simpler, consistent architecture of VGG19 is well-suited for this application, as it allows for effective multi-scale feature extraction without excessive computational demands. Moreover, the use of pre-trained VGG19 weights facilitated quicker learning during the training phase, as evidenced by the reduced training time compared to the ResNet-enhanced model. On the other hand, the U-Net with ResNet weights also showed competitive performance, particularly in capturing more complex patterns. The deeper ResNet architecture enabled the model to learn sophisticated features, which improved the segmentation of roads in challenging conditions. However, this added complexity came at the cost of increased computational resources and longer training times, which may limit its practicality for real-time or resource-constrained applications. The comparison indicates that while both advanced models provide significant improvements over the basic U-Net, the EU-Net (with VGG19) offers a more balanced trade-off between performance and efficiency. Its ability to achieve high accuracy with lower computational costs makes it an attractive choice for road segmentation tasks in satellite imagery, where processing speed and resource usage are critical considerations. Overall, integrating pre-trained weights from different architectures has been shown to enhance the performance of U-Net-based models. The EU-Net, leveraging VGG19, achieved superior results across multiple metrics, demonstrating its effectiveness in practical applications. Future work could explore ways to optimize the ResNet-based model to reduce its computational overhead or develop hybrid approaches that combine the strengths of both VGG19 and ResNet architectures, potentially yielding even better segmentation performance.

VI. Conclusion

In this paper, we proposed the EU-Net model for road segmentation in satellite images. To this end, the VGG layers, based on ImageNet weights, are embedded in EU-Net. Evaluation results of the proposed model on the DeepGlobe Road Extraction dataset using different evaluation metrics confirm the superiority of the proposed EU-Net compared to the base U-Net model. The proposed model can be used for developing real-time systems to detect various needs in satellite images. While the presented model provides accurate results, the implementation of a real-time platform can further enhance the performance by enabling faster and more responsive analysis. In the future, we plan to extend our work to identify different objects and road networks in video data, which will allow for continuous monitoring and dynamic scene understanding. Furthermore, to improve segmentation accuracy, we aim to integrate the U-Net architecture with morphological processing techniques. This approach will enhance the model's ability to refine segmentation boundaries, especially in challenging scenarios with complex textures, occlusions, or noisy backgrounds. By leveraging morphological operations, we expect to better delineate objects and structures, leading to more precise and reliable detection in a wider range of satellite imagery applications.

References

[1] World health organization (2020). https://www.who.int/news-room/fact-sheets/detail/road traffic-injuries

[2] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: ICLR. (2015)

[3] Majidi, N., Kiani, K., Rastgoo, R.: A deep model for super-resolution enhancement from a single image. Journal of AI and Data Mining, vol. 8, no. 4, pp. 451-460, 2020.

[4] Rastgoo, R., Sattari Naeini, V.: A neurofuzzy QoS-aware routing protocol for smart grids. 2014 22nd Iranian Conference on Electrical Engineering (ICEE), pp. 1080-1084, 2014.

[5] Kiani, K., Hematpour, R., Rastgoo, R.: Automatic grayscale image colorization using a deep hybrid model. Journal of AI and Data Mining, vol. 9, no. 3, pp. 321-328, 2021.

[6] Rastgoo, R., Kiani, K., Escalera, S., Sabokrou, M.: Multi-modal zero-shot dynamic hand gesture recognition. Expert Systems with Applications, vol. 247, pp. 123349, 2024.

[7] Rastgoo, R., Kiani, K.: Face recognition using fine-tuning of Deep Convolutional Neural Network and transfer learning. Journal of Modeling in Engineering, vol. 17, no. 58, pp. 103-111, 2019.

[8] Rastgoo, R., Sattari-Naeini, V.: Tuning parameters of the QoS-aware routing protocol for smart grids using genetic algorithm. Applied Artificial Intelligence, vol. 30, no. 1, pp. 52-76, 2016.

[9] Rastgoo, R., Sattari-Naeini, V.: Gsomcr: Multi-constraint genetic-optimized qos-aware routing protocol for smart grids. Iranian Journal of Science and Technology, Transactions of Electrical Engineering, vol. 42, pp. 185-194, 2018.

[10] Zarbafi, S., Kiani, K., Rastgoo, R.: Spoken Persian digits recognition using deep learning. Journal of Modeling in Engineering, vol. 21, no. 74, pp. 163-172, 2023.

[11] Bagherzadeh, F., Rastgoo, R.: Deepfake image detection using a deep hybrid convolutional neural network. Journal of Modeling in Engineering, vol. 21, no. 75, pp. 19-28, 2023.

[12] Ahmadi, AM., Kiani, K., Rastgoo, R.: A Transformer-based model for abnormal activity recognition in video. Journal of Modeling in Engineering, vol. 22, no. 76, pp. 213-221, 2024.

[13] Mottaghi, R., Chen, X., Liu, X., Cho, N.G., Lee, S.W., Fidler, S., Urtasun, R., Yuille, A.: The role of context for object detection and semantic segmentation in the wild. In: CVPR. (2014)

[14] Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y.: Overfeat: Integrated recognition, localization and detection using convolutional networks. In: ICLR. (2014)

[15] Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: CVPR. (2015)

[16] Y. Wang *et al*., "DDU-Net: Dual-Decoder-U-Net for Road Extraction Using High-Resolution Remote Sensing Images," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-12, 2022, Art no. 4412612, doi: 10.1109/TGRS.2022.3197546.

[17] Peng, D.; Zhang, Y.; Guan, H. End-to-End Change Detection for High Resolution Satellite Images Using Improved U-Net++. *Remote Sens.* 2019, *11*, 1382. https://doi.org/10.3390/rs11111382

[18] J. Chen *et al*., "DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images," in *IEEE Journal of Selected Topics in Applied Earth*

Observations and Remote Sensing, vol. 14, pp. 1194-1206, 2021, doi: 10.1109/JSTARS.2020.3037893.

[19] Wang, Z., Jiang, K., Yi, P., Han, Z., & He, Z. (2020). Ultra-dense GAN for satellite imagery super-resolution. *Neurocomputing*, *398*, 328-337.

[20] Zeng, L., Wardlow, B. D., Xiang, D., Hu, S., & Li, D. (2020). A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote Sensing of Environment*, *237*, 111511.

[21] Tarasiou, M., Chavez, E., & Zafeiriou, S. (2023). ViTs for SITS: Vision Transformers for Satellite Image Time Series. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10418-10428).

[22] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18* (pp. 234-241). Springer International Publishing.

[23] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., & Wang, M. (2022, October). Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision* (pp. 205-218). Cham: Springer Nature Switzerland.

[24] Hoang, T. N., Nguyen, H. V. N., Nguyen, K. H., & Quach, L. D. (2023). Lane Road Segmentation Based on Improved UNet Architecture for Autonomous Driving. *International Journal of Advanced Computer Science and Applications*, *14*(7).

[25] Singh, N. J., & Nongmeikapam, K. (2023). Semantic segmentation of satellite images using deep-UNet. *Arabian Journal for Science and Engineering*, *48*(2), 1193-1205.

[26] Aghalari, M., Aghagolzadeh, A., & Ezoji, M. (2021). Brain tumor image segmentation via asymmetric/symmetric UNet based on two-pathway-residual blocks. *Biomedical signal processing and control*, *69*, 102841.

[27] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[28] Thakur, P. S., Sheorey, T., & Ojha, A. (2023). VGG-ICNN: A Lightweight CNN model for crop disease identification. *Multimedia Tools and Applications*, *82*(1), 497-520.

[29] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The Pascal visual object classes (voc) chal- lenge. International Journal of Computer Vision, 88(2):303– 338, June 2010

[30] W. R. Crum, O. Camara, and D. L. G. Hill, “Generalized overlap measures for evaluation and validation in medical image analysis.,” IEEE Trans. Med. Imaging, vol. 25, no. 11, pp. 1451– 61, Nov. 2006.

[31] Demir, Ilke, et al. "Deepglobe 2018: A challenge to parse the earth through satellite images." *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2018.