Engineering and Technology Journal e-ISSN: 2456-3358

Volume 11 Issue 06 June-2026, Page No.-10788-10793

DOI: 10.47191/etj/v11i06.55, I.F. – 8.595

Deep Learning-Based Semantic Segmentation Models for Crane Detection in Post-Earthquake UAV Imagery

Salih Enes Şeker[1], İlyas Çankaya[2]

[1,2]Department of Electrical and Electronics Engineering, Ankara Yildirim Beyazit University

ABSTRACT: The rapid identification of heavy machinery in post-disaster environments can significantly support recovery and reconstruction operations. This study presents a comparative evaluation of deep learning based semantic segmentation models for crane segmentation in UAV imagery acquired after the February 6, 2023 earthquakes in Kahramanmaraş, Türkiye. Five semantic segmentation architectures, namely SegFormer, DeepLabV3+, SegNet, FCN and U-Net, were trained and evaluated under identical experimental conditions. The results showed that DeepLabV3+ achieved the best overall quantitative performance with an IoU of 66.1%, mIoU of 81.0%, and Dice score of 78.9%. The findings demonstrate that modern segmentation architectures provide effective solutions for crane segmentation in complex post-earthquake UAV imagery and have significant potential for supporting disaster recovery and situational awareness applications.

KEYWORDS: Semantic segmentation, UAV Imagery, deep learning, post-earthquake, crane

I. INTRODUCTION

Rapid and accurate assessment of affected areas is crucial for search and rescue activities, damage evaluation, and reconstruction planning. Traditional ground-based inspections are often time-consuming, labor-intensive, and sometimes impossible due to accessibility constraints in heavily damaged regions. Consequently, Unmanned Aerial Vehicles (UAVs) have emerged as an effective tool for post-disaster monitoring because they can rapidly acquire high-resolution imagery over large areas while minimizing risks to response personnel [1].

Recent advances in computer vision and deep learning have further enhanced the value of UAV imagery by enabling automatic object detection, semantic segmentation, and damage assessment [2]. Deep learning based approaches have been successfully applied to identify damaged buildings, roads, debris fields, and other critical infrastructure elements from aerial imagery with high accuracy [3]. In particular, semantic segmentation methods provide pixel-level information that allows detailed analysis of disaster environments, making them highly suitable for emergency management and reconstruction applications.

Beyond damage assessment, the identification of critical construction and rescue equipment within disaster zones has become increasingly important [4]. Among these assets, mobile and tower cranes play a fundamental role during post-earthquake recovery operations. Cranes are extensively utilized for debris removal, transportation of heavy materials, stabilization of damaged structures, and reconstruction activities in densely affected urban environments. The availability and spatial distribution of cranes within disaster areas can provide valuable information regarding ongoing rescue and recovery efforts, resource allocation, and reconstruction progress. Despite the growing body of research on post-disaster damage detection and UAV-based object recognition, studies specifically focusing on crane detection in post-earthquake UAV imagery remain limited.

Motivated by the limitations and research gaps identified in the existing literature, this study presents a comparative evaluation of five semantic segmentation architectures, namely SegFormer, DeepLabV3+, SegNet, FCN and U-Net for crane segmentation in post-earthquake UAV imagery. The performance of the evaluated models was analyzed using both quantitative metrics and qualitative visual assessments to provide a comprehensive comparison. The remainder of this paper is organized as follows. Section 2 summarizes the literature. Section 3 presents he investigated segmentation models and dataset. Section 4 introduces the evaluation metrics, experimental setup, and comparative results along with a detailed discussion. Finally, Section 5 summarizes the main findings and outlines directions for future research.

II. LITERATURE REVIEW

Calantropio et al. investigated automated post-earthquake building damage assessment using high-resolution UAV imagery collected after the 2016 Central Italy earthquake [5]. The authors employed a deep learning based image classification framework to distinguish damaged and undamaged buildings from aerial images, aiming to accelerate post-disaster inspection processes. Experimental results demonstrated a validation accuracy exceeding 90%, while class-level precision, recall, and F1-score analyses

confirmed the effectiveness of UAV based deep learning approaches for rapid damage assessment. Takhtkeshha et al. proposed a self-supervised deep learning framework for post-earthquake damage detection using very high resolution UAV imagery acquired after the Sarpol-e Zahab earthquake in Iran [6]. Unlike conventional supervised approaches, the method reduced the dependence on large annotated datasets by exploiting self-supervised feature learning. The proposed framework achieved competitive damage detection performance with overall accuracies approaching 90%, demonstrating its suitability for rapid disaster response applications. Jozi examined the use of UAV photogrammetry and deep learning techniques for post-disaster building damage assessment [7]. The study integrated image processing workflows with segmentation based deep neural networks to identify damaged structures from aerial imagery and evaluate structural conditions. Experimental analyses reported IoU values above 70% for several damage categories, highlighting the effectiveness of UAV datasets for automated disaster assessment. Albeaino et al. conducted a comprehensive review of UAV applications in architecture, engineering, and construction (AEC), focusing on progress monitoring, safety management, equipment tracking, and post-disaster inspection [8]. The study synthesized more than 200 publications and identified object detection and computer vision as rapidly growing research directions. Their review highlighted the increasing adoption of deep learning techniques for construction equipment monitoring and automated site analysis. He et al. developed a YOLOv11-based object segmentation framework for intelligent recognition of construction-site elements, including heavy machinery and equipment [9]. The proposed model combined object detection and pixel-level segmentation to improve recognition accuracy in complex construction environments. Experimental evaluations reported mAP values above 85% and demonstrated strong segmentation performance for construction-related objects, indicating the potential applicability of segmentation models to crane identification tasks. Gu et al. reviewed recent advances in post-disaster building damage assessment using remote sensing imagery from satellites and UAV platforms [10]. The study compared conventional machine learning approaches with modern CNN and Transformer based architectures, focusing on damage classification, object detection, and semantic segmentation tasks. The authors concluded that deep learning based methods consistently outperform traditional approaches and that UAV imagery provides significant advantages for detailed post-disaster analysis. Al Shafian and Hu presented a decade-long review of machine learning and remote sensing applications in post-disaster building damage assessment [11]. The study analyzed research trends, datasets, and methodological developments involving UAVs, satellite imagery, and deep neural networks. Their findings showed a substantial increase in segmentation-based approaches and highlighted UAV imagery as one of the most influential data sources for future disaster management systems. Xiao et al. investigated the integration of robotic cranes, computer vision, and reinforcement learning to optimize material supply operations during post-earthquake reconstruction [12]. Therefore, this study aims to perform crane segmentation using UAV imagery collected after the 2023 Kahramanmaraş earthquakes and to compare the performance of five widely used semantic segmentation architectures, namely SegFormer, DeepLabV3+, SegNet, FCN and U-Net. By focusing on crane identification rather than conventional building damage assessment, the study seeks to address a relatively unexplored research gap and contribute to the development of intelligent monitoring systems for post-disaster recovery management.

III.METHODOLOY

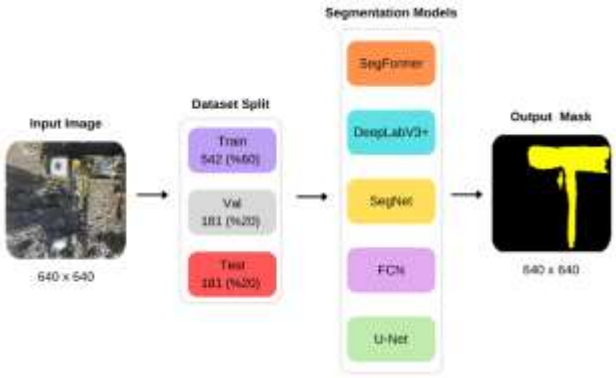

Figure 1. Workflow of the paper

To segment cranes in the post-earthquake UAV imagery, five semantic segmentation architectures were evaluated, spanning both convolutional and transformer-based paradigms. SegFormer-B0 is a lightweight transformer-based model that employs a hierarchical Mix Transformer (MiT) encoder to generate multi-scale features without relying on positional encodings, paired with a simple all-MLP decoder that aggregates information from different encoder stages to produce the final segmentation map, making it computationally efficient relative to its accuracy [13]. DeepLabv3+ is a convolutional encoder-decoder model that incorporates atrous (dilated) convolutions and an Atrous Spatial Pyramid Pooling (ASPP) module to capture multi-scale contextual information without losing spatial resolution, followed by a decoder module that refines object boundaries by fusing encoder outputs with low-level features [14]. The remaining three models share a VGG16 backbone but differ in their decoding strategies: SegNet mirrors the VGG16 encoder with a symmetric decoder that performs upsampling using the max-pooling indices stored during encoding, allowing efficient reconstruction of spatial detail with a relatively small memory footprint [15]; FCN replaces the fully connected layers of VGG16 with convolutional layers and uses learned deconvolution (transposed convolution) operations combined with skip connections to fuse coarse semantic information from deeper layers with finer spatial

information from shallower layers [16]; and U-Net adopts a fully symmetric encoder-decoder structure in which the VGG16 encoder progressively downsamples the input to extract contextual features, while the decoder upsamples these features and concatenates them with corresponding encoder feature maps via skip connections, enabling precise localization of object boundaries even for small or irregularly shaped targets such as cranes [17]. Figure 1 shows the overall workflow of this paper.

A. Dataset

The dataset used in this study consists of post-earthquake UAV imagery acquired in Kahramanmaraş, Türkiye, following the February 6, 2023 earthquakes. Since the objective of this study was crane segmentation, image patches containing crane instances were extracted from the original UAV imagery. A total of 904 image crops with a spatial resolution of 640 × 640 pixels were generated and manually annotated using pixel-level segmentation masks. In Figure 2, representative image samples used in this study are presented. The resulting dataset was divided into training, validation, and test subsets with proportions of 60%, 20%, and 20%, respectively.

Figure 2. Representative samples from the crane segmentation dataset

IV. RESULTS AND DISCUSSIONS

A. *Performance Metrics*

The performance of the segmentation models was evaluated using Intersection over Union (IoU), Mean Intersection over Union (mIoU), Dice Coefficient, Precision, Recall, and Mean Pixel Accuracy (mPA). IoU measures the overlap between the predicted and ground-truth segmentation masks, while mIoU represents the average IoU across all classes. The Dice coefficient quantifies the similarity between predicted and reference masks and is particularly suitable for segmentation tasks involving imbalanced class distributions. Precision and Recall were used to assess the ability of the models to correctly identify crane pixels while minimizing false detections and missed detections, respectively. Finally, mPA was employed to evaluate the average pixel-level classification accuracy across all classes. Together, these metrics provide a comprehensive assessment of segmentation performance from both region-overlap and pixel-classification perspectives.

B. Experimental Setup

All experiments were conducted under identical training conditions to ensure a fair comparison among the evaluated segmentation architectures. The input image size was fixed at 640 × 640 pixels, the batch size was set to 8, and the initial learning rate was set to 1×10^{-4}. Each model was trained for a maximum of 100 epochs, while an early stopping strategy with a patience value of 20 epochs was employed to prevent overfitting and terminate traning when no further improvement was observed on the validation set. The same experimental setup was consistently applied to SegFormer, DeepLabV3+, U-Net, SegNet, and FCN models.

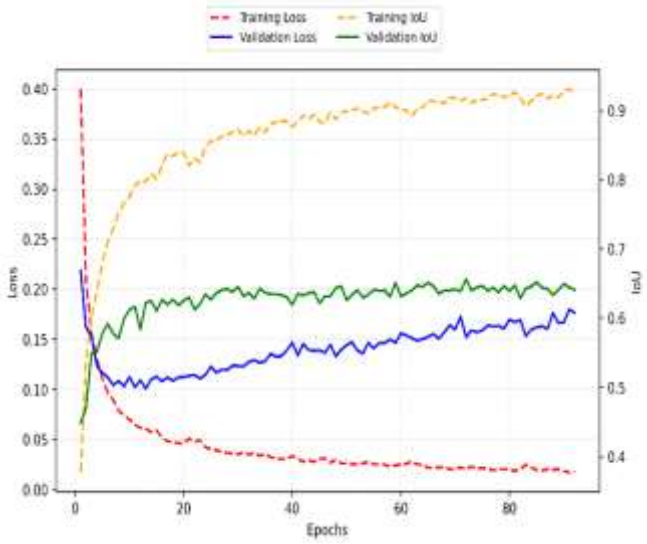

Figure 3. Training and validation loss and IoU curves of the SegFormer model.

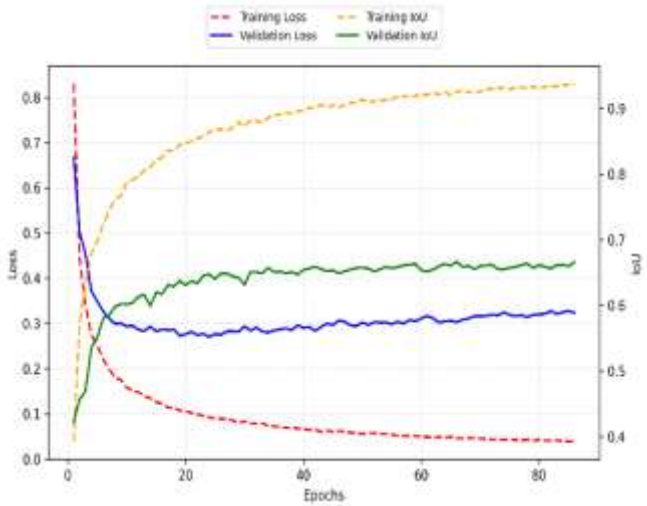

Figure 4. Training and validation loss and IoU curves of the DeepLabV3+ model.

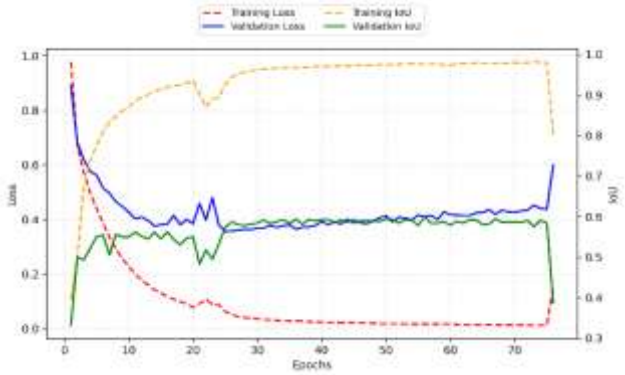

Figure 5. Training and validation loss and IoU curves of the SegNet model.

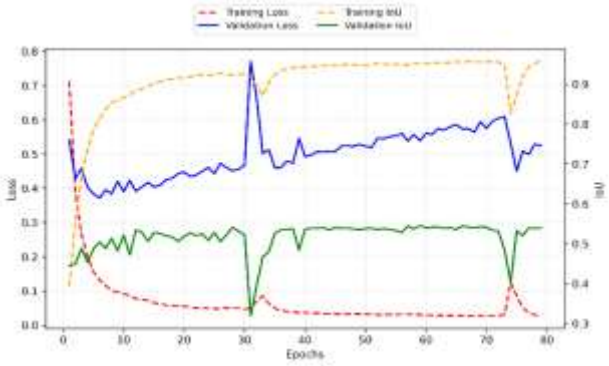

Figure 6. Training and validation loss and IoU curves of the FCN model.

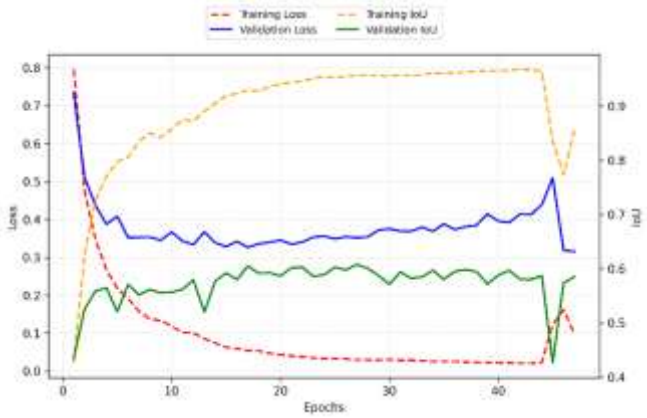

Figure 7. Training and validation loss and IoU curves of the U-Net model.

Figure 3 to 7 presents the training and validation curves of the evaluated segmentation models. For all models, the training loss consistently decreased while the training IoU increased throughout the learning process, indicating successful convergence and effective feature learning from the training data.

C. Qualitative Results

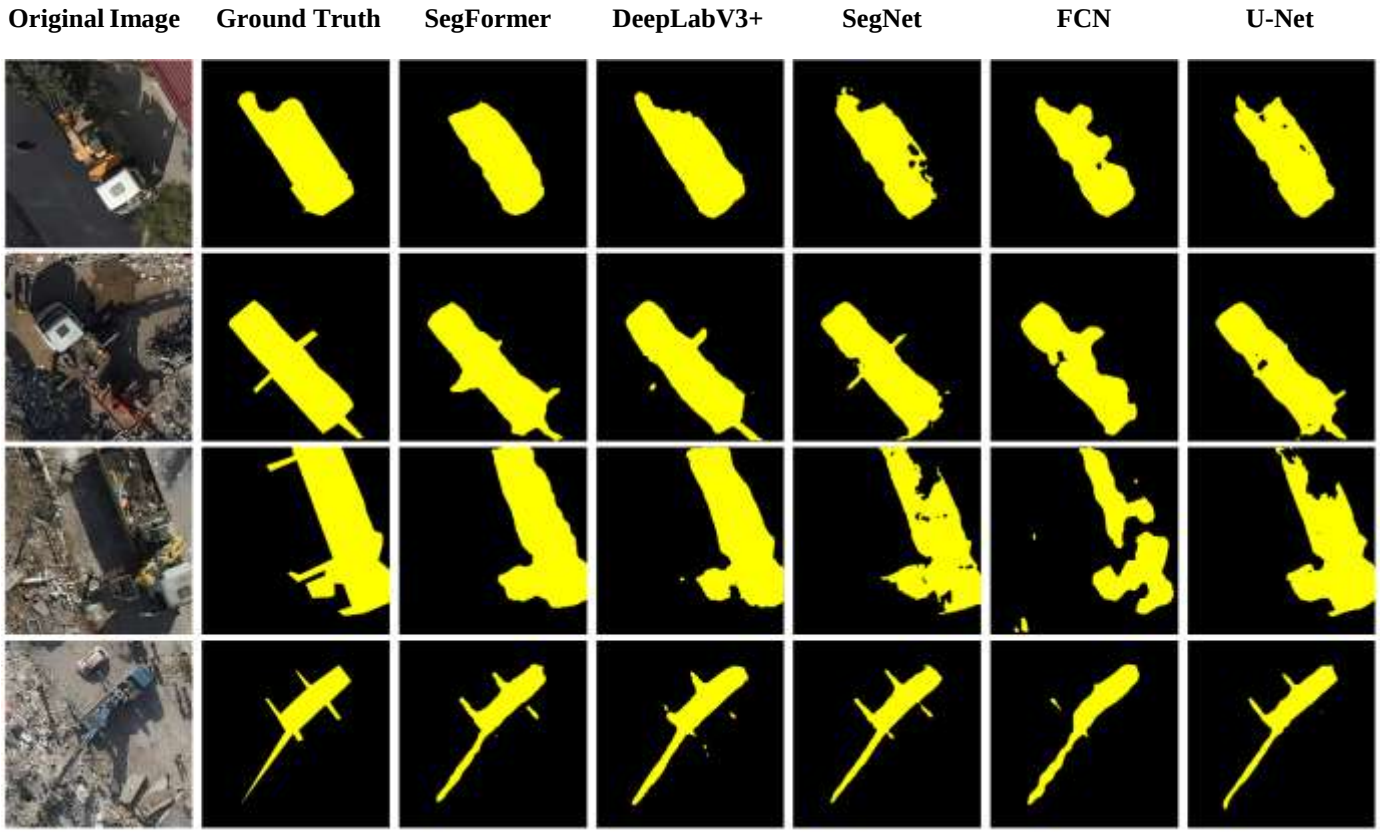

Figure 8. Visual comparison of segmentation results produced by the evaluated models.

Figure 8 presents qualitative segmentation results for four representative test images. The first and second columns show the original UAV images and corresponding ground-truth masks, respectively, while the remaining columns

illustrate the predictions generated by SegFormer, DeepLabV3+, SegNet, FCN, and U-Net. The examples provide a visual comparison of each model's ability to delineate crane structures with varying shapes and geometric complexities.

Overall, SegFormer produced the most accurate and visually consistent segmentation masks, showing the closest correspondence to the ground truth. The model effectively preserved object boundaries, thin structural components, and fine geometric details while maintaining compact and continuous object regions. DeepLabV3+ also demonstrated strong performance, generating smooth and coherent masks with accurate object boundaries, although some thin crane components were occasionally omitted.

SegNet preserved several fine structural details but frequently produced internal holes and fragmented regions within the segmented objects. In contrast, U-Net generated cleaner and more homogeneous masks with fewer internal gaps; however, some slender crane components were smoothed out or missed. FCN exhibited the weakest qualitative performance, often producing oversimplified object shapes and failing to capture important structural details.

Overall, the qualitative observations are consistent with the quantitative results. SegFormer achieved the best visual agreement with the ground-truth annotations, followed closely by DeepLabV3+. SegNet and U-Net provided competitive results with different trade-offs between detail preservation and mask continuity, whereas FCN showed clear limitations in complex UAV-based crane segmentation tasks.

D. Qualitative Results

Table 1 summarizes the quantitative performance of the evaluated segmentation models on the test set. The results indicate clear performance differences among the architectures, with DeepLabV3+ and SegFormer achieving the best overall results, while FCN obtained the lowest scores across most metrics.

DeepLabV3+ delivered the highest segmentation accuracy, achieving an IoU of 66.1%, mIoU of 81.0%, Dice coefficient of 78.9%, and Precision of 80.8%. These results indicate accurate object boundary delineation and a low false-positive rate. SegFormer ranked second with an IoU of 63.8% and a Dice coefficient of 77.0%, but achieved the highest Recall of 78.1% and mPA of 87.9%, demonstrating a stronger ability to capture crane pixels and preserve complete object structures.

The comparison between DeepLabV3+ and SegFormer highlights a trade-off between precision and completeness. DeepLabV3+ produced more conservative and accurate predictions, achieving a Precision of 80.8% and an IoU of 66.1%, whereas SegFormer showed a greater tendency to retain fine structural details and small object components, reflected by its higher Recall of 78.1% and mPA of 87.9%. This observation is consistent with the qualitative results presented in Figure 8.

Among the CNN-based methods, U-Net and SegNet achieved similar performance levels. U-Net obtained a Precision of 80.6% and a Recall of 70.0%, indicating cleaner masks with some missed object regions. In contrast, SegNet achieved a Recall of 72.0% and a Precision of 76.1%, reflecting a tendency to preserve more object pixels while introducing additional segmentation artifacts.

FCN produced the weakest results, with an IoU of 55.1% and a Dice coefficient of 70.2%, highlighting the limitations of early fully convolutional architectures for complex UAV-based crane segmentation tasks.

Overall, the quantitative results are consistent with the qualitative analysis. DeepLabV3+ achieved the best overall segmentation performance, while SegFormer demonstrated the strongest capability for preserving complete object structures and fine details. U-Net and SegNet provided competitive intermediate performance, whereas FCN consistently lagged behind the more advanced architectures.

Table I. Quantitative performance comparison of the evaluated segmentation models

Model	*IoU*	*mIoU*	*Dice*	*Precision*	*Recall*	*mPA*
SegFormer-B0	0.638	0.796	0.770	0.766	**0.781**	**0.879**
DeepLabV3+	**0.661**	**0.810**	**0.789**	**0.808**	0.775	0.878
SegNet	0.593	0.771	0.736	0.761	0.720	0.849
FCN	0.551	0.746	0.702	0.725	0.693	0.834
U-Net	0.605	0.778	0.744	0.806	0.700	0.842

V. CONCLUSION

This study presented a comparative evaluation of five semantic segmentation architectures. Both quantitative and qualitative analyses demonstrated that DeepLabV3+ and SegFormer consistently outperformed the remaining models. DeepLabV3+ achieved the highest overall quantitative performance, obtaining the best IoU, mIoU, Dice, and Precision scores, while SegFormer produced the highest Recall and mPA values and generated segmentation masks that most closely resembled the ground-truth annotations in the visual assessment. U-Net and SegNet achieved competitive intermediate results with different trade-offs between mask continuity and detail preservation, whereas FCN exhibited the lowest performance across all evaluation metrics. Overall, the findings indicate that modern transformer-based and advanced encoder–decoder architectures provide substantial advantages over traditional fully convolutional networks for UAV based crane

segmentation in complex post-disaster environments. Future work may focus on expanding the dataset with UAV imagery collected from different disaster-affected regions in Türkiye as well as from other countries to improve the generalization capability and robustness of the models.

ACKNOWLEDGMENT

I would like to thank Dr. Ali Değirmenci for his valuable guidance, constructive comments, and continuous support throughout this study.

REFERENCES

1. Giordan, D., Adams, M. S., Aicardi, I., Alicandro, M., Allasia, P., Baldo, M., ... & Troilo, F. (2020). The use of unmanned aerial vehicles (UAVs) for engineering geology applications. *Bulletin of Engineering Geology and the Environment*, *79*, 3437-3481.
2. Tang, G., Ni, J., Zhao, Y., Gu, Y., & Cao, W. (2023). A survey of object detection for UAVs based on deep learning. Remote Sensing, 16(1), 149.
3. Wang, L., Wu, J., Yang, Y., Tang, R., & Ya, R. (2024). Deep learning models for hazard-damaged building detection using remote sensing datasets: A comprehensive review. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, 15301-15318.
4. Ivanov, M. L., & Georgiev, M. P. (2025). Innovations, applications, and future perspectives in geospatial information visualization for disaster response: insights from the 2023 Kahramanmaras Earthquake urban search and rescue operations. *Natural Hazards*, *121*(19), 22787-22816.
5. Calantropio, A., Chiabrando, F., Codastefano, M., & Bourke, E. (2021). Deep learning for automatic building damage assessment: application in post-disaster scenarios using UAV data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *1*, 113-120.
6. Takhtkeshha, N., Mohammadzadeh, A., & Salehi, B. (2022). A rapid self-supervised deep-learning-based method for post-earthquake damage detection using UAV data (case study: Sarpol-e Zahab, Iran). *Remote Sensing*, *15*(1), 123.
7. Jozi, S. D. (2024). UAV-based Post-disaster Damage Assessment of Buildings Using Image Processing.
8. Albeaino, G., Gheisari, M., & Franz, B. W. (2019). A systematic review of unmanned aerial vehicle application areas and technologies in the AEC domain. *Journal of information technology in construction*, *24*.
9. He, L., Zhou, Y., Liu, L., & Ma, J. (2024). Research and application of YOLOv11-based object segmentation in intelligent recognition at construction sites. *Buildings*, *14*(12), 3777.
10. Gu, J., Xie, Z., Zhang, J., & He, X. (2024). Advances in rapid damage identification methods for post-disaster regional buildings based on remote sensing images: A survey. *Buildings*, *14*(4), 898.
11. Al Shafian, S., & Hu, D. (2024). Integrating machine learning and remote sensing in disaster management: A decadal review of post-disaster building damage assessment. *Buildings*, *14*(8), 2344.
12. Xiao, Y., Yang, T. Y., Pan, X., Xie, F., & Chen, Z. (2023). A reinforcement learning based construction material supply strategy using robotic crane and computer vision for building reconstruction after an earthquake. *arXiv preprint arXiv:2308.16280*.
13. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, *34*, 12077-12090.
14. Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 801-818).
15. Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, *39*(12), 2481-2495.
16. Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 3431-3440)
17. Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Cham: Springer international publishing.