Chapter 7

Semantic Segmentation of Green Areas in Satellite Images Using Deep Learning Models

Fatma Nur Kılıçkaya[1]

Murat Taşyürek[2]

Celal Öztürk[3]

Abstract

Accurate identification and segmentation of green areas from satellite imagery are important for environmental management, urban planning, and sustainable development. Accurate determination and calculation of the hectares of green areas are particularly important for monitoring natural events such as forest fires, land-use changes, and climate-related disasters. Such analysis directly expedites the processes of damage assessment, risk evaluation, and response. However, conventional approaches are inadequate for modeling complex spatial structures in high-resolution satellite data that require automated analysis. In this study, a dataset consisting of 8028 satellite images of Kayseri province, manually labelled as polygons on the MakeSense platform, is used. Six classes were identified for green areas: dense vegetation, light vegetation, wooded areas, object surroundings, and similar classes. We converted polygon labels to pixel-level masks and treated the task as a semantic segmentation problem. In this work, a comparative evaluation of the U-Net and DeepLabV3+ models, both based on the convolutional neural network (CNN), was carried out against the transformer-based SegFormer architecture. The performance values obtained from the experimental studies on real satellite images indicate that the reorganized 4-class structure achieved better results than the initial 6-class structure. In the 6-class structure, the best mIoU value was obtained by U-Net and DeepLabV3+ with 0.51, while the highest Dice and F1-score values were achieved by U-Net and SegFormer with 0.66. DeepLabV3+ achieved the best

1 Kayseri University, Technical Sciences Vocational School, Department of Computer Technologies-Computer Programming, Türkiye
Erciyes University, Faculty of Engineering, Department of Computer Engineering, Turkey

2 Kayseri University, Faculty of Engineering, Department of Computer Engineering, Turkey

3 Erciyes University, Faculty of Engineering, Department of Software Engineering, Turkey

https://doi.org/10.58830/ozgur.pub1404.c5657

Pixel Accuracy and Precision, with 0.74 and 0.70, respectively. In contrast, the 4-class structure yielded higher, more balanced performance across all models. DeepLabV3+ and SegFormer achieved the highest mIoU value of 0.57, while SegFormer obtained the best Dice, Recall, and F1-score values with 0.72. Pixel Accuracy reached 0.76 for all models in the 4-class structure. These findings show that reducing the class number from six to four improved the segmentation performance by decreasing confusion among visually similar green-area classes. The results indicate that deep learning-based approaches provide an effective solution for automatically detecting and classifying green areas in satellite images. The proposed method holds significant potential, particularly for rapid, accurate analysis of damage after forest fires and for monitoring green area loss. In this respect, the study has the potential to contribute to large-scale environmental monitoring and decision support systems.

Keywords: *Large-scale Satellite Imagery, Green Area Classification, Transformer-Based Deep Learning, Semantic Segmentation, Remote Sensing*

INTRODUCTION

With increasing population concentration in urban areas, environmental problems can arise. The growing urban population is threatened by global warming, air and water pollution. Rapid urbanization is predicted to reduce people's contact with nature. As of 2019, half of the world's population lives in cities, and this figure is expected to reach approximately 70% by 2050 (Gascon et al, 2015). This situation is thought to have far-reaching effects on the physical and mental health of urban dwellers (Zhang et al, 2023). Urban greenery not only reflects landscape character but also plays an important role in counteracting the negative environmental impacts of human activities. Urban greenery provides sustainable ecological benefits, improves the physiological health of urban dwellers, and reduces the risk of diseases associated with environmental regulations (Stubbings et al., 2019). In addition to urban green spaces, wooded areas are also important for maintaining ecological balance. Wooded areas perform critical ecological functions in terms of sustaining biodiversity, supporting carbon sequestration, reducing soil erosion, and protecting natural habitats (Diener et al, 2021). Natural disasters such as forest fires, floods, droughts, and landslides are among the significant environmental events that threaten the sustainability of green areas. Forest fires, in particular, cause the rapid destruction of large wooded areas; leaving long-term negative impacts on biodiversity, soil structure, the carbon cycle, and ecosystem balance (Kumar et al, 2019; Liu et al, 2014). The loss of vegetation after a fire leads not only to a decrease in forest cover but also to an increased risk of erosion, habitat fragmentation, and degradation of natural habitats. Therefore, wooded areas and other green regions need to be monitored

regularly. Accessing and monitoring these green areas requires manpower and financial resources. Satellite imagery provides a valuable data source for determining the location of wooded and green areas over large areas, tracking their changes, and conducting segmentation-based analyses. There are a lot of studies for identification and monitoring of green spaces in the literature. Taşyürek et al. (Taşyürek et al, 2024) proposed a method called DGAG to identify green areas and to determine geographical location information from large scale satellite images. Shahtahmassebi et al. (Shahtahmassebi et al, 2021) reviewed comprehensively remote sensing methods for identification, monitoring and evaluation of urban green areas, who pointed out satellite images are an important data source in green space studies. Liu et al. (Liu et al, 2019) proposed an approach to automatically extract urban green areas based on GF-2 satellite images by using a semantic segmentation model DeepLabV3+. Edra et al (Edra et al, 2023) applied a segmentation approach based on the U-Net architecture for a pixel-level segmentation of urban green areas in PlanetScope satellite images. In a study using very high-resolution satellite imagery and deep learning to examine changes in urban forest cover, Javed et al. (2023) revealed the temporal traceability of wooded areas. Estacio et al. (2023) proposed an open-source method for automatic extraction of urban green areas from satellite images, and demonstrated the applicability of this method on the green area mapping studies. Zahir et al. (2024) explored remote sensing monitoring of urban green areas based on vegetation indices derived from Sentinel-2 multiband satellite imagery. Guo et al. (2025) used a SegFormer-based semantic segmentation model to automatically extract urban green areas from BJ-2 remote sensing imagery. When these studies are evaluated as a whole, it is seen that satellite imagery, remote sensing techniques, and deep learning-based segmentation models offer effective approaches for the automatic detection, monitoring, and mapping of urban green spaces, wooded areas, and vegetation changes. However, automatically separating green spaces in satellite imagery is not always easy. Different types of green spaces may exhibit similar color, texture, and density characteristics. This situation can cause models to mix some classes, especially in segmentation studies with a large number of classes. This study aims to investigate the effect of the re-arranged 4-class structure on the model performance by changing the 6-class label structure used for green space segmentation in satellite imagery. For this purpose, the models U-Net, DeepLabV3+, and SegFormer were comparatively evaluated.

MATERIAL AND METHODS

Material

The material used in this study was obtained from the Kayseri Metropolitan Municipality's 2020 satellite imagery at https://tile.kayseri.bel.tr/tilecache/Cache/2020U YDU3857/z/x/y.jpeg. Among the large volume of satellite images, a single ECW-compressed image occupies approximately 20 GB on the computer. Due to this problem, preprocessing steps were applied to the images. In this process, noise was removed from the images, accuracy was improved, and the images were made suitable for analysis. Subsequently, the large-scale image data is integrated into the open-source GeoServer platform and published there via WMS services in accordance with standards. However, the WMS format is not suitable for direct processing by deep learning-based CNN models. Therefore, using the TileCache system offered under the BSD license, the WMS outputs are converted into tile sub-images with dimensions of 256×256 pixels. This tile structure allows for the hierarchical reproduction of images at each zoom level (z). The first level contains a single image; the second level contains 4; the third level contains 16 images; and this structure continues with 4^z sub-images. TileCache stores these sub-images in the z/x/y format.{mime} directory structure; the MIME type is usually JPEG or PNG. Thanks to this structure, a single large satellite image can be directly processed by CNN-based models, which are then broken down into numerous small sub-images containing positional information.

Table 1. TileCache Configuration

PARAMETER	VALUE
TYPE	WMS
URL	http://localhost:8082/geoserver/wms?layers=2020uydu_3857
EXTENSİON	png
SİZE	256,256
SRS	EPSG:3857
LEVELS	22
BBOX	-20037508.34, -20037508.34, 20037508.34, 20037508.34
MAXRESOLUTİON	156543.0339
EXTENT_TYPE	loose

Table 1 shows information about the layers provided by GeoServer. At this stage, it is necessary to define the specific GeoServer layer to be accessed. Information about the GeoServer service is expressed as "type," and GeoServer provides data via WMS (Web Map Service). The TileCache system converts the data obtained from this service into images of predefined dimensions.

As a result of the preprocessing steps, 8028 satellite images were obtained. 6823 of these images were labeled. Of the labeled data, 5619 images were split into a training dataset, 1204 into a validation dataset, and the remaining 1205 into a test dataset. Table 2 presents the proportional results of the data.

Table 2. Dataset Distribution Analysis

Data Set Section	*Number of Images*	*Percentage of Total (%)*
Total satellite imagery	8028	100%
Labeled imagery	6823	85%
Training dataset	5619	70%
Validation dataset	1204	15%
Test dataset	1205	15%

Methods

Labeling Structure

The satellite images were initially evaluated into six classes using the Makesense.ai platform. This structure aimed to provide a more detailed representation of green spaces based on their density, texture, and environmental characteristics. It was observed that some classes in the six-class structure had very similar visual characteristics. In particular, there were similarities in color, texture, and density between dense green space types and tree-covered green space subclasses. Therefore, to increase intra-class consistency and reduce inter-class confusion, semantically similar classes were merged. Figure 1 shows the class names and the merged classes.

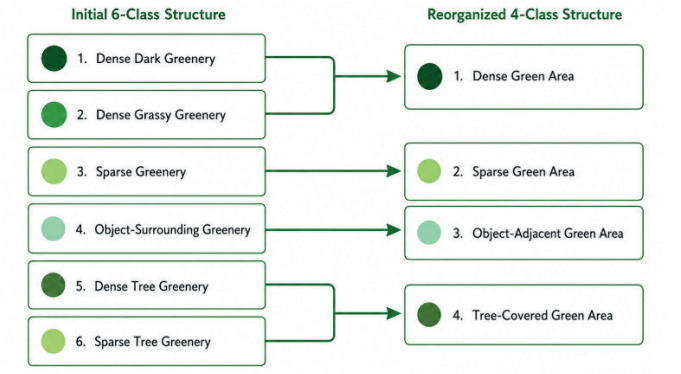


Figure 1. Class Reorganization from 6 Classes to 4 classes

Semantic Segmentation Models Used

Following preprocessing, the satellite images were divided into tile-based patches and subsequently used to train the U-Net, DeepLabV3+, and SegFormer architectures for green area segmentation. In this context, three widely used approaches in the literature over the past few years have been preferred. The first of these approaches is U-Net, a basic segmentation model that is particularly successful, especially on limited datasets, thanks to its encoder-decoder structure and skip connections. The second approach is DeepLabV3+, which captures multi-scale contextual information through its Atrous convolution and encoder-decoder structure. The third approach is the SegFormer model, which better learns global contextual information through its Transformer-based architecture and represents modern semantic segmentation approaches. After the training process, these architectures perform object detection on tiled images; the obtained outputs show the classification results. The models shown in Figure 2 were evaluated on both the initial 6-class and the rearranged 4-class label structures. Thus, the effect of the number of classes and the class rearrangement approach on different semantic segmentation architectures was analyzed comparatively.

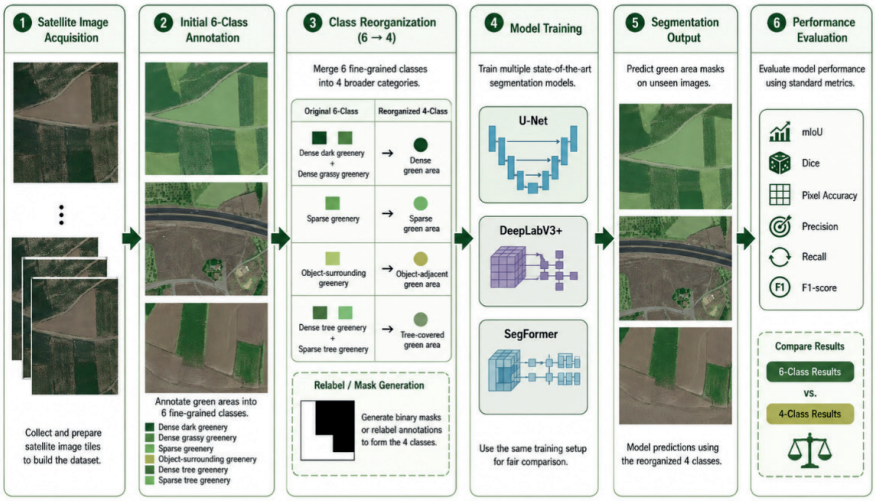


Figure 2. General architecture of the proposed green space segmentation.

Performance Appraisal Metrics

he performance evaluation utilized mIoU, Dice coefficient, Pixel Accuracy, Precision, Recall, and F1-score metrics.

Pixel Accuracy, given in Equation 1, gives the percentage of pixels correctly classified among all pixels. It shows the overall accuracy; however, it may not be sufficient on its own when there is class imbalance.

$$\text{Pixel Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

Precision, as defined in Equation 2, is the fraction of pixels predicted as positive that are positive. It is used to evaluate the impact of false positive predictions on model performance.

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

Recall, as given in Equation 3, is the fraction of the pixels that are positive that are correctly detected by the model. It reveals the extent of missing regions within the target class.

$$\text{Recall} = TP / (TP + FN) \quad (3)$$

F1-score, given in Equation 4, is the harmonic mean of the Precision and Recall values. It provides a balanced performance interpretation of false-positive and false-negative predictions.

$$\text{F1-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

mIoU given in Equation 5 is the average of the IoU values of all classes. It is the basic metric used to evaluate segmentation performance on a class-by-class and holistic basis.

$$\begin{aligned} IoU_i &= TP_i / (TP_i + FP_i + FN_i) \\ mIoU &= (1 / C) \times \Sigma_{i=1}^{c} IoU_i \end{aligned} \quad (5)$$

The Dice coefficient given in Equation 6 is an overlap metric that measures the similarity between the predicted and actual masks. It is a sensitive indicator for evaluating small or limited target areas.

$$Dice = 2TP / (2TP + FP + FN) \quad (6)$$

RESULTS

This study aims to automatically identify green areas from satellite images through semantic segmentation. To this end, satellite images with different land cover characteristics were used, and pixel-level labeled mask data were prepared for each image. Initially, green areas were evaluated into six classes based on visual density, texture, vegetation structure, and environmental location characteristics. These classes were defined as dense dark greenery, dense grassy greenery, sparse greenery, object-surrounding greenery, dense tree greenery, and sparse tree greenery. The goal was to represent green areas within a more detailed class structure. However, because some green area types in satellite images exhibit very similar color, texture, and density characteristics, segmentation models can struggle to distinguish between them. In particular, the visual similarities between dense dark green and dense grassy green classes, and between dense tree and sparse tree green classes, increase the risk of interclass confusion. Therefore, in this study, the initial six-class label structure was reorganized considering semantic and visual similarities, and a new four-class label structure was created. In this context, dense dark greenery and dense grassy greenery classes have been combined under the "dense green area" class, while dense tree greenery and sparse tree greenery classes have been combined under the "tree-covered green area" class. Sparse green area and object adjacent green area classes have been left as separate classes, as they retain their unique characteristics.

As shown in Figure 3, the confusion matrices were used to analyze class-wise prediction performance and to identify the degree of misclassification among visually similar green-area classes in the 6-class and 4-class segmentation structures.

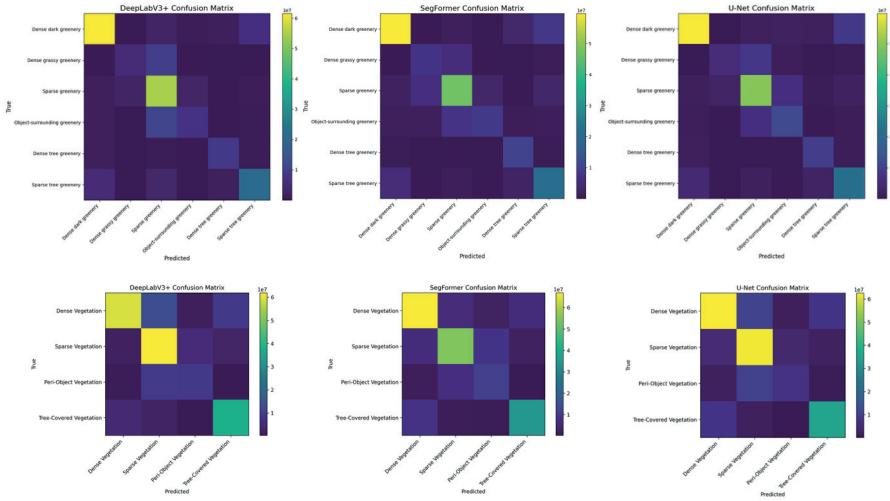


Figure 3. Confusion matrix results of the U-Net, DeepLabV3+, and SegFormer models under the initial 6-class and reorganized 4-class green area segmentation schemes

The performance of the models was quantitatively assessed using widely accepted evaluation metrics for semantic segmentation. This means that the rearranged 4-class structure gave more successful and balanced results than the 6-class structure. From Table 3, we can see that the best performance was achieved by the 6-class SegFormer and U-Net structure, with an F1 score of 66%, while the 4-class structure achieved an F1 score of 72% with SegFormer. Segmentation performance improved with fewer classes, especially reducing confusion among visually similar green space types.

Table 3. Model Performance Comparison for 4-Class and 6-Class Segmentation

Class	Model	mIoU	Dice	Pixel Accuracy	Precision	Recall	F1-score
4 class	**U-Net**	0.56	0.70	0.76	0.72	0.69	0.70
4 class	**DeepLabV3+**	0.57	0.71	0.76	0.72	0.71	0.71
4 class	**SegFormer**	0.57	0.72	0.76	0.71	0.72	0.72
6 class	**U-Net**	0.51	0.66	0.73	0.68	0.67	0.66
6 class	**DeepLabV3+**	0.51	0.65	0.74	0.70	0.64	0.65
6 class	**SegFormer**	0.50	0.66	0.72	0.66	0.67	0.66

An increase in mIoU, Dice, and Pixel Accuracy values was observed in the 4-class structure. This result shows that segmentation success depends

not only on the model architecture but also on the proper design of the label class structure.

DISCUSSION AND CONCLUSION

This study presents a comparative evaluation of the restructuring of the class structure for automatic segmentation of green areas in satellite imagery. The proposed approach can be considered an effective strategy to reduce class confusion and improve segmentation performance, especially in datasets with visually and semantically similar green-area subclasses. Although the 6-class structure provides a more detailed representation, some green area types in satellite images are visually very similar to one another. Therefore, it is difficult for the models to clearly distinguish between dense foliage, grassland, and wooded area subtypes. In the 4-class structure, intraclass consistency is increased by combining semantically similar classes, the learning process of the models is facilitated, and performance values are improved. These findings show that not only more advanced model architectures but also correct class definition and labeling strategies are critical in semantic segmentation studies. Furthermore, comparing the performance of U-Net, DeepLabV3+, and SegFormer models across different class structures shows that appropriate model and labeling strategies should be considered together in green area segmentation studies. Compared with previous studies that mainly focused on the extraction or segmentation of green areas from satellite images, the present study additionally investigates the effect of class structure on segmentation performance. These findings are consistent with previous studies demonstrating the effectiveness of deep learning models in green area extraction from satellite imagery. In this respect, the study offers a methodological framework for decision-support processes related to remote-sensing-based green space monitoring, urban planning, environmental sustainability, and the protection of wooded areas. In future studies, the generalizability of the proposed class-arrangement approach can be evaluated across different urban regions, seasonal images, and larger datasets.

References

Diener, A., & Mudu, P. (2021). How can vegetation protect us from air pollution? A critical review on green spaces' mitigation abilities for air-borne particles from a public health perspective-with implications for urban planning. *Science of the Total Environment, 796*, 148605.

Edra, A. P., Suharyadi, Farda, N. M., & Arjasakusuma, S. (2023). Urban green space segmentation in PlanetScope satellite images based on the improvement of U-Net architecture. *AIP Conference Proceedings, 2865*, 020008.

Estacio, I. (2024). Open-source automatic extraction of Urban Green Space. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-2-2024*, 65–72.

Gascon, M., Triguero-Mas, M., Martínez, D., Dadvand, P., Forns, J., Plasència, A., & Nieuwenhuijsen, M. J. (2015). Mental health benefits of long-term exposure to residential green and blue spaces: A systematic review. *International Journal of Environmental Research and Public Health, 12*(4), 4354–4379.

Guo, Y., Shao, Q., Wang, X., & Liu, F. (2025). Urban green space extraction from BJ-2 remote sensing image based SegFormer semantic segmentation model. *Canadian Journal of Remote Sensing, 51*(1).

Javed, A., Kim, T., Lee, C., Oh, J., & Han, Y. (2023). Deep learning-based detection of urban forest cover change along with overall urban changes using very-high-resolution satellite images. *Remote Sensing, 15*(17), 4285.

Kumar, P., Druckman, A., Gallagher, J., Gatersleben, B., Allison, S., Eisenman, T. S., ... & Morawska, L. (2019). The nexus between air pollution, green infrastructure and human health. *Environment International, 133*, 105181.

Liu, H. L., & Shen, Y. S. (2014). The impact of green space changes on air pollution and microclimates: A case study of the Taipei metropolitan area. *Sustainability, 6*(12), 8827–8855.

Liu, W., Yue, A., Shi, W., Ji, J., & Deng, R. (2019). An automatic extraction architecture of urban green space based on DeepLabv3plus semantic segmentation model. *2019 IEEE 4th International Conference on Image, Vision and Computing*, 311–315.

Shahtahmassebi, A. R., Li, C., Fan, Y., Wu, Y., Gan, M., Wang, K., ... & Blackburn, G. A. (2021). Remote sensing of urban green spaces: A review. *Urban Forestry & Urban Greening, 57*, 126946.

Stubbings, P., Peskett, J., Rowe, F., & Arribas-Bel, D. (2019). A hierarchical urban forest index using street-level imagery and deep learning. *Remote Sensing, 11*(12), 1395.

Taşyürek, M., Kılıçkaya, F. N., & Öztürk, C. (2024). DGAG: A novel approach that detects green areas with geolocation from large-scale satellite imaging

instance segmentation and GIS application. *Transactions in GIS, 28*(8), 2742–2762.

Zahir, I. L. M. (2024). Monitoring urban green spaces using remotely sensed vegetation indices from Sentinel-2 multispectral imagery. *Procedia Computer Science*.

Zhang, L., Wang, L., Wu, J., Li, P., Dong, J., & Wang, T. (2023). Decoding urban green spaces: Deep learning and Google street view measure greening structures. *Urban Forestry & Urban Greening, 87*, 128028.

Acknowledgment

The authors would like to acknowledge TÜBİTAK Career Development Program (3501), Project No. 122E315, for its support in the development of the dataset and research infrastructure related to this study.

Conflict of Interest

The authors have declared that there is no conflict of interest

Author Contributions

All authors contributed to the conception and design of the study. Data preparation, annotation organization, and class reorganization from six classes to four classes were carried out collaboratively. The implementation and evaluation of U-Net, DeepLabV3+, and SegFormer models were performed as part of the experimental process. All authors contributed to the analysis and interpretation of the results, manuscript writing, critical revision, and approval of the final manuscript.