Deep Learning Based Semantic Segmentation and Explainability Analysis for Building Footprint Extraction Using High-Resolution Remote Sensing Imagery

Kavzoglu T.*, Yilmaz E.O. and Teke A.

Department of Geomatics Engineering, Gebze Technical University, Kocaeli, Turkey

*kavzoglu@gtu.edu.tr

***Abstract*:** *In the increasingly urbanized world, the capacity to accurately map the urban environment is of paramount importance. Building footprint extraction, a process that involves the automatic identification of building outlines from remote sensing imagery, has emerged as a foundational technology of profound importance across numerous sectors. The process of transforming pixels into precise vector polygons enables the extraction of a substantial body of geospatial intelligence, which has been demonstrated to facilitate advancements in domains, including urban planning and disaster management. Although deep learning shows promise in building footprint extraction, challenges remain due to variations in building types and complex backgrounds. The progress realized in building footprint mapping directly supports two Sustainable Development Goals (SDGs), namely "Sustainable Cities and Communities (SDG 11)", which promotes planned and sustainable urban development, and "Climate Action (SDG 13)", which is critical for identifying buildings in disaster-prone areas. In this study, two semantic segmentation models, DeepLabV3+ and PSPNet, were assessed on high-resolution SPOT 6/7 imagery covering the Pyrénées-Orientales region of France. A comparison was conducted between the performance of models using accuracy metrics. DeepLabV3+ produced an IoU score of 0.9541 and an accuracy of 0.9762, while the PSPNet recorded 0.9463 for IoU and 0.9720 for accuracy. The efficacy of the DeepLabV3+ architecture was demonstrated in the identification of large and regular building types. In addition, the decision-making process of models was analyzed for explainability using eXplainable Artificial Intelligence (XAI) techniques (i.e., GradientSHAP) with XAI metrics. Apart from accuracy assessment, the GradientSHAP analysis revealed that the DeepLabV3+ model was more sensitive to building boundaries. The PSPNet model exhibited a more scattered and inconsistent performance in the XAI output. In summary, the best results for identifying building footprints from SPOT 6/7 imagery were obtained through the integration of the DeepLabV3+ model with the GradientSHAP method. Quantitative XAI analysis revealed that DeepLabV3+ outperformed PSPNet in faithfulness estimate and relevance mass accuracy, while PSPNet produced more selective but less stable explanations with higher sparseness values. The findings clarify how XAI contributes to the improvement of reliability and transparency in building footprint extraction.*

Keywords: Building footprint detection, DeepLabV3+, explainable AI, GradientSHAP, PSPNet.

Introduction

Urbanization is considered one of the most remarkable spatial transformation processes of the 21st century (Gao & O'Neill, 2020). According to the report by the United Nations (2019), over half of the world's population resides in urban areas. Also, this percentage is expected to rise to nearly 68% by 2050. The swift pace of urban expansion is placing cities under increasing strain, thereby necessitating well-structured spatial planning (Salem & Tsurusaki, 2024; Mahtta et al., 2022). Achieving these planning objectives, however, critically depends on the availability of reliable building footprint information in rapidly urbanizing regions. This is because up-to-date building footprint information has become essential for infrastructure management, energy modeling, and reducing the risk of disasters (Pesaresi et al., 2016; Li et al., 2020).

The delineation of building footprints can be achieved through the utilization of remote sensing technology. However, there are some problems, such as the diversity of building types, variation in structural characteristics, and complexity of environmental factors. Overcoming these problems can be critical, contributing to reaching the goals set by the SDGs. Among them, "Sustainable Cities and Communities (SDG 11)" highlights the importance of reliable spatial information for monitoring urban growth (United Nations, 2015; UN-Habitat, 2020). Likewise, "Climate Action (SDG 13)" points out the need to monitor vulnerable infrastructure to reduce risks (IPCC, 2022; Gupta et al., 2019). In other words, rapidly identifying structures exposed to hazards, such as earthquakes or floods, provides important data support for disaster risk management (Chen et al., 2020; Yang et al., 2025).

In recent years, significant advances in artificial intelligence (AI) and remote sensing fields have created significant opportunities for the identification of building footprints (Shi et al., 2022). These methods deliver substantially higher accuracy compared to traditional image processing techniques. Also, they have driven major advances in building detection (Zhu et al., 2017). Semantic segmentation models, including U-Net, SegNet, LinkNet, DeepLab, and PSPNet, have proven particularly effective in detecting building footprints of different sizes and materials (Wang et al., 2022; Abdollahi & Pradhan, 2021). However, there are still some challenges in the field of building extraction. First of all, it is important to note that buildings in different regions vary significantly in terms of size, shape, and construction materials. Furthermore, the boundaries of buildings may be obscured by intricate background elements comprising vegetation and shadows. To address these limitations, there is a necessity to develop more robust deep learning approaches (Lup et al., 2021; Chang et al., 2025).

Convolutional neural network (CNN) architectures have demonstrated strong capabilities in capturing multi-scale spatial features (Liu et al., 2021). Parallel to these developments, increasing attention has been devoted to model interpretability. Despite their high performance and capacity, deep learning models often work as a "black box". For users to trust the outcomes and for policy makers to rely on them, it is necessary to understand the features or pixels that shape the decisions of models (Kavzoglu et al., 2025). At this point, explainable AI (XAI) methods provide a valuable means of enhancing the interpretability of model decisions (Yilmaz & Kavzoglu, 2024). The employment of the XAI methods simplifies the visualization of the areas that models focus on and enhances the transparency of decision-making processes (Samek et al., 2021). However, it is important to note that reliance on visualizations only provides a partial solution. The employment of quantitative XAI metrics is also necessary to assess the reliability, accuracy, and selectivity of explanations (Miró-Nicolau et al., 2025). The metrics in question have been developed for the purpose of systematically revealing whether the model faithfully reflects its decisions, how focused it is on relevant domains, and how selective the explanations are. It can thus be concluded that explanations become analytically assessable as a consequence of this process, rather than serving merely as visual aids (Bommer et al., 2024).

This research seeks to evaluate the ability of DeepLabV3+ and PSPNet architectures, and their performance was also tested using accuracy assessment metrics (i.e., Intersection over Union (IoU), F-score, overall accuracy, precision, and recall). Alongside quantitative performance, the decision transparency of the models was assessed using the XAI method of GradientSHAP. In addition, the results of the XAI method were quantitatively evaluated through specific XAI metrics. Thus, this study not only evaluates the accuracy of DL models but also integrates explainable AI, providing two contributions, namely methodological robustness and interpretability for practical urban applications. It is expected that the findings of the study will contribute to urban planning and disaster risk management applications.

Study Area & Dataset

The study area, the Pyrenees-Orientales region of France, was chosen because the high-resolution satellite images and building footprint vector data for this region are freely available (AIRBUS Defense and Space, 2023). In addition, the region has both dispersed rural settlements and compact urban fabric in coastal cities, providing a significant advantage in terms of geographical diversity. The data set consists of SPOT 6/7 orthorectified multispectral satellite imagery acquired on July 8, 2023. The image contains 1.5 m panchromatic and 6 m resolution multispectral bands (blue, green, red, near infrared). All bands were sharpened to 1.5 m resolution by the PAN-sharpening process performed by AIRBUS. Red-green-blue bands were used for building footprint extraction, while the near infrared band was excluded. On the other hand, the labels of building footprint were provided by the Center for Urban Scientific Expertise through the GEOSUD/DINAMIS portal (THISME, 2020). Their vector data were generated using automated deep learning methods on SPOT 6/7 images, achieving an F-score of 69%. A visual inspection also confirmed the consistency between the vector data and the satellite imagery, with no substantial discrepancies observed (Figure 1).

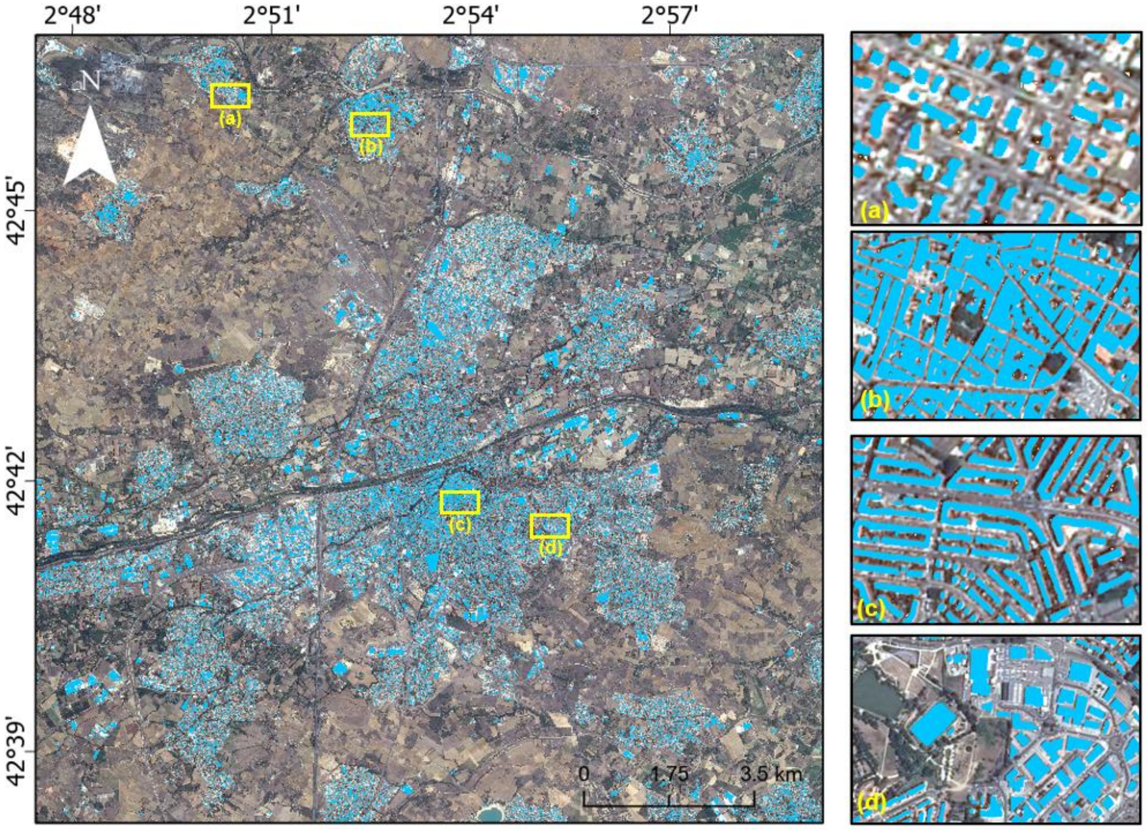

Figure 1: Example of the satellite image and building footprint data for the study area.

Methodology

Deep learning, as a subfield of machine learning, focuses on learning hierarchical representations from data through artificial neural networks (Luo et al., 2021). It has demonstrated remarkable success across a wide range of application domains, including image processing. Its effectiveness can largely be attributed to its ability to deliver high levels of accuracy while maintaining strong generalization across different datasets. Two widely used deep learning architectures, including DeepLabV3+ and PSPNet, were employed for building boundary extraction from high-resolution satellite images in this study.

IoU, F-Score, accuracy, dice loss, precision, and recall metrics were used to evaluate the performance of deep learning models, while overall accuracy is defined as the proportion of all pixels that are correctly predicted, precision is the ratio of true positives to predicted positives, and recall is the ratio of true positives to all relevant instances. The F-score represents the harmonic means of precision and recall values. Moreover, the term IoU is used to denote the ratio of the intersection of the predicted building footprint regions and the actual data to unity. Finally, a loss function is employed using the Dice coefficient. After the accuracy assessment, a comparative analysis was conducted, including interpretability assessments based on the XAI approach with its metrics.

a. DeeplabV3+ architecture

DeepLabV3+ is a deep learning model used for pixel-level semantic segmentation (Chen et al., 2018; Wang et al., 2024). Operations such as contextual information identification are effectively performed because of the features of this architecture. It includes three key modules, namely an encoder, an Atrous Spatial Pyramid Pooling module (ASPP), and a decoder. The encoder extracts feature-rich representations. ASPP module enhances the ability of models to capture multi-scale contextual information by using parallel convolution layers with different dilation rates (Mahara et al., 2025). Thus, objects of different sizes can be effectively recognized. Finally, the decoder enables more precise extraction of object boundaries by combining the low-resolution features obtained from the encoder with high-resolution details. However, the relatively high computational cost requires the use of powerful hardware resources (Chen et al., 2017).

b. PSPNet architecture

The Pyramid Scene Parsing Network (PSPNet), introduced in 2017, has demonstrated success in some segmentation tasks (Chen et al., 2021). Its architecture typically uses a ResNet-like backbone to extract high-level feature maps. The significant component of the PSPNet model is pooling windows at different scales. This is achieved by dividing the input image into subregions and then performing average pooling. These multi-resolution features are then upsampled and combined with the input feature maps. This preserves local details and integrates broad context information into the model. In the final stage, the final segmentation map is generated using several convolution blocks. PSPNet has the advantage of effectively utilizing contextual information, especially in complex scenes. However, it has some limitations in terms of precisely segmenting very large or very small objects together, and the additional computational cost (Zhao et al., 2017).

c. eXplainable Artificial Intelligence

In deep learning applications, gaining a thorough understanding of the parameters within a model is an extremely challenging process. These "black box" models process high-dimensional datasets containing many hidden layers and millions of parameters to solve complex problems (Ozupek et al., 2025). Consequently, numerous studies are being conducted to understand the internal dynamics of these models (e.g., Teke and Kavzoglu, 2024; Kavzoglu et al., 2025). Various XAI methodologies (e.g., GradientShap) have been developed to make models more explainable through algorithms (Metsch & Hauschild, 2025). The GradientShap method was used to explain the models in the study. It is a hybrid methodology that combines SHAP theory with gradient-based attribution techniques. It enhances Integrated Gradients using multiple baselines, where the contribution of input features is calculated for each baseline and subsequently averaged following the Shapley value principle (Lundberg & Lee, 2017; Kawauchi & Fuse, 2022; Yılmaz et al., 2025). This process ensures that feature attributions are not only sensitive to gradient information but also adhere to the fairness principle of SHAP, where each feature's contribution is considered across all possible subsets. The GradientShap algorithm offers the advantage of producing explanations that are both more robust and more generalizable. This is primarily because multiple references help reduce bias and increase attribution stability. However, the method requires more computational resources.

XAI methods are widely used to explain the decision processes of deep learning models, and assessing the quality of these explanations is an important research area. Correct and incorrect detections of building footprints were evaluated in the qualitative assessment process, and five key metrics were employed for quantitative analysis, which are faithfulness estimate, faithfulness correlation, relevance mass accuracy, relevance rank accuracy, and sparseness. A comparative evaluation of the methods was conducted by measuring the reliability, meaningfulness, and accuracy of the explanations using these metrics. Explanations of these metrics are provided below.

- **Faithfulness estimate:** XAI measures how accurately the importance values assigned to the features in the explanation reflect their contribution to the model output (Alvarez-Melis & Jaakkola, 2018). High fidelity is achieved if removing a feature significantly reduces the model output. It expresses the relationship between the explanation and the model output using the correlation coefficient. High values indicate strong agreement, while low values indicate that the explanation does not adequately reflect the model.

- **Faithfulness correlation:** XAI measures how faithful the explanation is to the model and calculates the correlation between the importance scores in the explanation and the changes in the model output (Bhatt et al., 2020). Removing high-importance features is expected to result in a large change in model output, while removing low-importance features is expected to result in a small change. This metric is calculated using Pearson correlation on randomly selected feature subsets.
- **Relevance mass accuracy:** It evaluates the extent to which the total contribution values in the explanation fall within the correct region. A high value indicates that the model truly explains its decisions over relevant areas (Arras et al., 2022).
- **Relevance rank accuracy:** It measures the success of the XAI annotation in detecting important regions at the correct location. A high value indicates that the important pixels are largely located within the correct segment (Arras et al., 2022).
- **Sparseness:** XAI measures how many features the explanation's importance scores are concentrated on. A high value indicates that the explanation focuses on a small number of features, while a low value indicates that the importance scores are spread across many features (Chalasani et al., 2020).

Results and Discussion

Building footprint extraction was performed in this study using high-resolution SPOT 6/7 images covering the Pyrénées-Orientales region of France. The satellite images and images containing building labels were divided into sections as 256×256-pixel image-mask pairs with 64-pixel overlap. Also, the label-image pairs without buildings were removed from the dataset. For training and accuracy assessment stages, the dataset was divided into three parts: 60% training (4858 samples), 20% validation (1620 samples), and 20% testing (1620 samples). To prevent overfitting and to enhance the generalization ability of the deep learning models, data augmentation techniques were applied during the training stage. Specifically, each training image was subjected to geometric transformations, including horizontal flipping, vertical flipping, and 90-degree random rotations with predefined probabilities. These augmentation strategies increased the diversity of the dataset, enabling the models to better learn structural variations and different orientations of buildings. It is important to note that augmentation was applied exclusively to the training set, whereas the validation and test sets were kept unchanged to guarantee a fair and unbiased evaluation of model performance (Yilmaz & Kavzoglu, 2025).

Before the training of the DeepLabV3+ and PSPNet models, the hyperparameters were determined. ResNet50 was selected as the encoder for both architectures. The Sigmoid activation function was employed, with Dice loss specified as the loss function. During the training procedure, a batch size of 16 was employed, whereas a batch size of 4 was utilized during the validation process. The Adam algorithm was implemented as the optimization function, with a learning rate of 0.00005 and 30 epochs.

The DeepLabV3+ model demonstrated a high success rate, achieving an IoU of 95.41%, accuracy of 97.62%, and F-score of 97.62% (Table 1). On the other hand, the Dice loss is 0.0242, indicating that the building boundaries are slightly widened, especially in densely built-up areas. Nevertheless, the model produced balanced and reliable outputs in terms of both precision and recall. On the other hand, the PSPNet model demonstrated suboptimal performance. The IoU of 94.63% and accuracy of 97.20% demonstrate that the model fulfils the fundamental segmentation task. However, the Dice loss is higher at 0.0285. This finding suggests the presence of deficiencies, particularly regarding the identification of small-scale structures and the occurrence of false positive classifications. It is evident that the pyramid pooling module of PSPNet has clear merits, particularly in its ability to integrate extensive contextual information into the model. However, the module also shows limitations when applied to structures that are small or irregular in shape.

Table 1: Performance comparison of deep learning models obtained for the study area.

Models	**Dice Loss**	**IoU**	**Recall**	**Precision**	**Accuracy**	**F-score**
DeepLabV3+	0.0170	0.9541	0.9757	0.9767	0.9762	0.9762
PSPNet	0.0741	0.9463	0.9713	0.9726	0.9720	0.9720

The prediction outputs of both deep learning models, considering the sample images, are shown in Figure 2. When the images are analyzed, it is seen that the DeepLabV3+ model could separate building boundaries more sharply and completely. It is noteworthy that it produces more accurate results than PSPNet in the separation of building clusters, especially in areas with compact urban texture. Although PSPNet can delineate general building zones by leveraging the contextual information provided by the pyramid pooling module, it occasionally misclassifies non-building areas as buildings and fails to capture certain small-scale structures. In addition, its predictions often produce blurred and merged building boundaries, which reduces segmentation precision. Consequently, the model exhibits lower performance in irregular and heterogeneous urban environments.

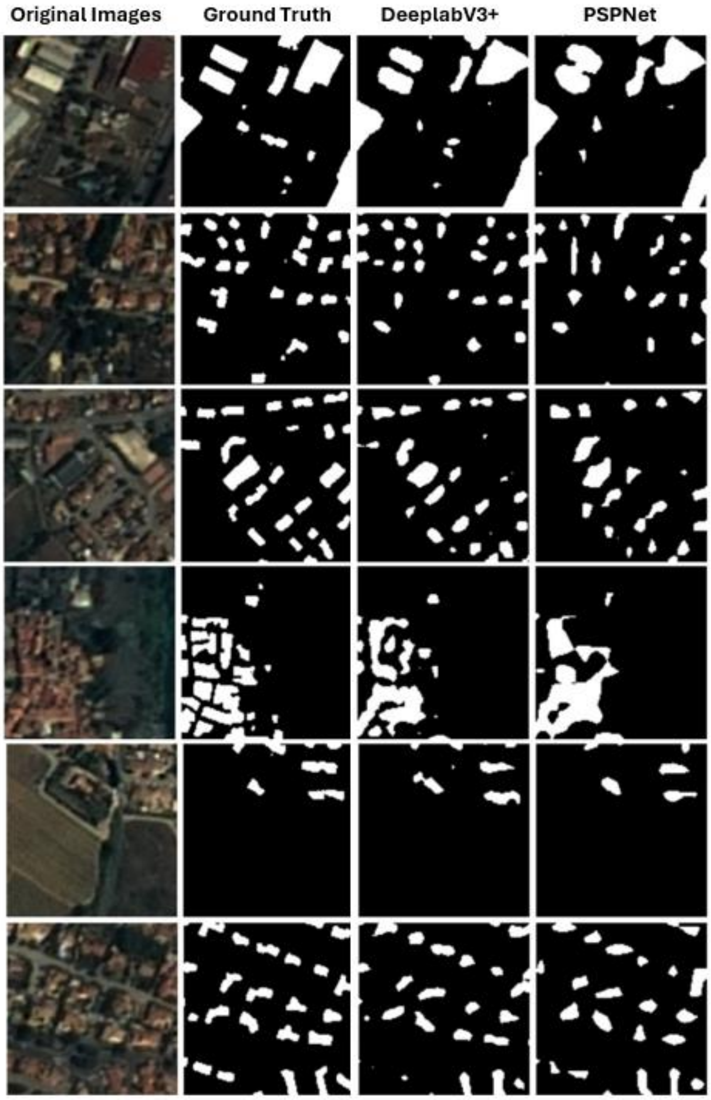

Figure 2: Prediction outputs of deep learning models of the study area.

For the study area, the decision mechanisms of DeepLabV3+ and PSPNet models were analyzed with the GradientShap method (Figure 3). The outputs show that both models focus on building zones at different levels. In the DeepLabV3+ model, GradientShap maps reveal that building boundaries are focused more intensely. Large and medium-sized buildings were detected with high accuracy by the model, and attention regions were particularly concentrated on the edges of buildings. This suggests that DeepLabV3+ employs a more selective and structurally consistent attention mechanism during the decision-making process. In contrast, the GradientShap outputs of PSPNet appeared more dispersed, with decision regions less clearly defined than those of DeepLabV3+. Although certain building groups exhibited identifiable attention regions, the explanations remained limited due to the model's weaker segmentation performance. For small-scale buildings in particular, the focus was insufficient, and false positives were frequently observed. Overall, the GradientShap analysis demonstrated that DeepLabV3+ could delineate building boundaries with greater reliability, whereas PSPNet, despite its ability to capture contextual information, provided restricted explanatory power for small and irregular structures.

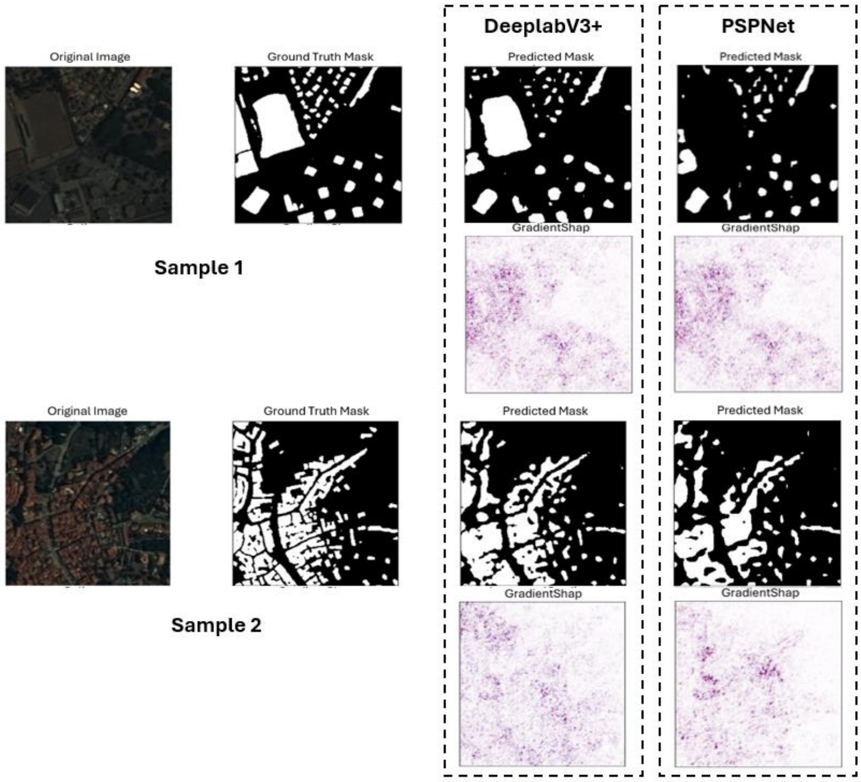

Figure 3: GradientShap maps with prediction masks of deep learning models for selected samples.

The XAI metrics used for quantitative analysis were also estimated for the two models employed in this study (Table 2). A radar chart is also plotted to facilitate a holistic comparison of the metrics for better comparison of the model robustness (Figure 4). The results reveal that DeepLabV3+ showed significantly higher values, especially in the second sample (0.247), in terms of the faithfulness estimate. This demonstrates that DeepLabV3+ is more successful in reflecting model decisions.

Table 2: XAI metric comparison of DeepLabV3+ and PSPNet models for two samples.

XAI Metric	DeepLabV3+		PSPNet	
	Sample 1	Sample 2	Sample 1	Sample 2
Faithfulness estimate	0,046	0,247	0,024	0,053
Faithfulness correlation	0,068	0,046	0,141	0,032
Relevance mass accuracy	0,126	0,517	0,386	0,052
Relevance rank accuracy	0,319	0,371	0,307	0,397
Sparseness	0,574	0,556	0,675	0,617

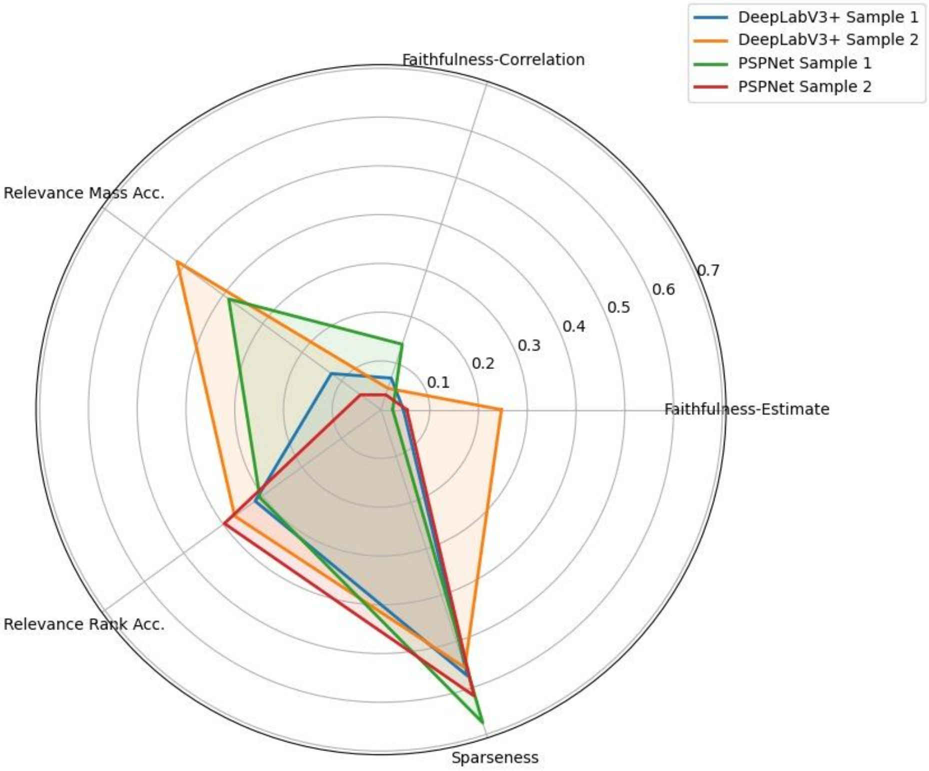

Figure 4: Radar plot of XAI metrics for DeepLabV3+ and PSPNet models.

For the faithfulness correlation metric, PSPNet achieved a better correlation in the first sample (0.141), while lower values were obtained in other cases. Furthermore, the relevance mass accuracy metric provides the most discriminatory results among the models. DeepLabV3+ demonstrated a high performance of 0.517 in the second sample, while PSPNet remained significantly lower (0.052) in the second sample. On the other hand, relevance rank accuracy values were similar for both models, with PSPNet achieving the highest performance in the second sample (0.397). The sparseness values for PSPNet are higher (0.675 and 0.617), indicating more selective explanations. The radar chart shown in Figure 4 also clearly demonstrates these trends. In summary, DeepLabV3+ particularly stands out in the faithfulness estimate and relevance mass accuracy metrics, while PSPNet, with its higher sparseness, offers more sparse but consistent explanations.

Conclusions and Recommendations

This study examined the extraction of building footprints from high-resolution SPOT 6/7 imagery of the Pyrénées-Orientales region in France using two widely adopted deep learning-based semantic segmentation models, namely DeepLabV3+ and PSPNet. The findings confirmed that both architectures can produce reliable outputs for building segmentation tasks; however, their effectiveness varied according to the spatial characteristics of the study area and the structural design of the models. Furthermore, quantitative evaluation confirmed the superior performance of the DeepLabV3+ model compared to PSPNet.

This study emphasizes the importance of evaluating building segmentation models not only in terms of accuracy but also in terms of interpretability. GradientShape analysis has shown that DeepLabV3+ produces more focused and consistent attention maps, particularly around building boundaries, and that this is consistent with its strong quantitative results. Overall, the findings reveal a clear connection between higher segmentation accuracy and more reliable interpretability. Furthermore, the XAI method used was evaluated with XAI metrics. The faithfulness estimate and relevance mass accuracy metrics show that DeepLabV3+ exhibits high accuracy performance with focused and consistent explanations; the sparseness metric shows that PSPNet produces more selective but scattered and inconsistent explanations, yielding relatively weak results. These findings reveal a strong correlation between model accuracy and interpretability. The higher the segmentation performance, the clearer and more reliable the explanations become. Consequently, the DeepLabV3+ model outperforms PSPNet not only in terms of accuracy but also in terms of interpretability.

The combination of strong evaluation scores and clearer GradientShap explanations indicates that this architecture holds strong potential for extracting building footprints from high-resolution satellite images. More broadly, the study highlights the value of integrating XAI into geospatial deep learning. Such integration can improve transparency and trust in automated mapping results. In addition, it provides practical benefits for urban planning and disaster risk management, supporting the broader objectives of SDG 11 and SDG 13.

Acknowledgments

This study was supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK) under Project No: 124Y240.

References

Abdollahi, A., & Pradhan, B. (2021). Integrating semantic edges and segmentation information for building extraction from aerial images using UNet. *Machine Learning with Applications*, *6*, 100194. https://doi.org/10.1016/j.mlwa.2021.100194

AIRBUS Defence and Space. (2023). *SPOT Ortho - Pyrénées Orientales France*. https://intelligence.airbus.com/imagery/sample-imagery/spot-2-ortho-pyrenees-orientales-france-july-2023/

Alvarez-Melis, D., & Jaakkola, T. (2018). Towards robust interpretability with self-explaining neural networks. In Advances in Neural Information Processing Systems (Vol. 31). https://doi.org/10.48550/arXiv.1806.07538

Arras, L., Osman, A., & Samek, W. (2022). CLEVR-XAI: A benchmark dataset for the ground truth evaluation of neural network explanations. *Information Fusion, 81,* 14–40. https://doi.org/10.1016/j.inffus.2021.11.012

Bhatt, U., Weller, A., & Moura, J. M. (2020). Evaluating and aggregating feature-based model explanations. *arXiv preprint arXiv:2005.00631.*

Bommer, P. L., Kretschmer, M., Hedström, A., Bareeva, D., & Höhne, M.-C. (2024). Finding the right XAI method—A guide for the evaluation and comparison of explanation methods in climate science. *Artificial Intelligence for the Earth Systems (AIES), 3*(3). https://doi.org/10.1175/AIES-D-23-0074.1

Chalasani, P., Chen, J., Chowdhury, A. R., Wu, X., & Jha, S. (2020). Concise explanations of neural networks using adversarial training. In *Proceedings of the International Conference on Machine Learning* (pp. 1383–1391). PMLR. https://doi.org/10.48550/arXiv.1810.06583

Chang, J., He, X., Song, D., Li, P., Qiao, M. & Cheng, X. (2025). A multi-scale attention network for building extraction from high-resolution remote sensing images. *Scientific Report,* 15, 24938. https://doi.org/10.1038/s41598-025-09086-9

Chen, C., Deng, J., & Lv, N. (2020). Illegal constructions detection in remote sensing images based on multi-scale semantic segmentation. In *2020 IEEE International Conference on Smart Internet of Things (SmartIoT)* (pp. 300–303). IEEE. https://doi.org/10.1109/SmartIoT49966.2020.00053

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision – ECCV 2018. Lecture Notes in Computer Science* (Vol. 11211, pp. 833–851). Springer. https://doi.org/10.1007/978-3-030-01234-2_49

Chen, S., Song, Y., Su, J., Fang, Y., Shen, L., Mi, Z., & Su, B. (2021). Segmentation of field grape bunches via an improved pyramid scene parsing network. *International journal of* *agricultural and biological engineering*, *14*(6), 185-194. https://doi.org/10.25165/j.ijabe.20211406.6903

Gao, J., O'Neill, B.C. Mapping global urban land for the 21st century with data-driven simulations and Shared Socioeconomic Pathways. *Nature Communications*, 11, 2302 (2020). https://doi.org/10.1038/s41467-020-15788-7

Gupta, R., Hosfelt, R., Sajeev, S., Patel, N., Goodman, B., Doshi, J., Heim, E., Choset, H., & Gaston, M. (2019). Building footprint extraction from satellite images in humanitarian contexts using deep learning. *Remote Sensing, 11*(6), 740. https://doi.org/10.3390/rs11060740

Intergovernmental Panel on Climate Change (IPCC). (2022). *Climate Change 2022: Impacts, adaptation and vulnerability*. Cambridge University Press.

Kavzoglu, T., Uzun, Y. K., Berkan, E., & Yilmaz, E. O. (2025). Global-scale explainable AI assessment for OBIA-based classification using Deep Learning and Machine Learning methods. *Advances in Geodesy and Geoinformation*, e62-e62. https://doi.org/10.24425/agg.2025.150692

Kawauchi, H., & Fuse, T. (2022). SHAP-based interpretable object detection method for satellite imagery. *Remote Sensing, 14*(9), 1970. https://doi.org/10.3390/rs14091970

Li, W., Gong, P., & Liang, L. (2020). A 30-year (1984–2013) record of annual urban dynamics of the United States from Landsat data. *Remote Sensing of Environment, 241*, 111731. https://doi.org/10.1016/j.rse.2020.111731

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 10012–10022). IEEE. https://doi.org/10.1109/ICCV48922.2021.00986

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4765–4774). Curran Associates, Inc.

Luo, L., Li, P., & Yan, X. (2021). Deep Learning-Based Building Extraction from Remote Sensing Images: A Comprehensive Review. *Energies*, *14*(23), 7982. https://doi.org/10.3390/en14237982

Mahara, A., Khan, M. R. K., Deng, L., Rishe, N., Wang, W., & Sadjadi, S. M. (2025). Automated Road Extraction from Satellite Imagery Integrating Dense Depthwise Dilated Separable Spatial Pyramid Pooling with DeepLabV3+. *Applied Sciences*, *15*(3), 1027. https://doi.org/10.3390/app15031027

Mahtta, R., Fragkias, M., Güneralp, B., Mahendra, A., Reba, M., Wentz, E. A., & Seto, K. C. (2022). Urban land expansion: The role of population and other drivers. *Nature Reviews Earth & Environment, 3*(7), 529–543. https://doi.org/10.1038/s42949-022-00048-y

Metsch, J. M., & Hauschild, A. C. (2025). BenchXAI: Comprehensive benchmarking of post-hoc explainable AI methods on multi-modal biomedical data. *Computers in Biology and Medicine*, *191*, 110124. https://doi.org/10.1016/j.compbiomed.2025.110124

Miró-Nicolau, M., Jaume-i-Capó, A., & Moyà-Alcover, G. (2024). A comprehensive study on fidelity metrics for XAI. *Information Processing & Management, 62*(1), 103900. https://doi.org/10.1016/j.ipm.2024.103900

Ozupek, E., Teke, A., Celik, N., & Kavzoglu, T. (2025). Explainable artificial intelligence to explore the intrinsic characteristics of climatic parameters governing meteorological drought forecasting: opening the black box. *Stochastic Environmental Research and Risk Assessment*, 39, 3201–3222. https://doi.org/10.1007/s00477-025-03007-y

Pesaresi, M., Ehrlich, D., Ferri, S., Florczyk, A. J., Freire, S., Halkia, M., Julea, A., Kemper, T., Soille, P., & Syrris, V. (2016). A global human settlement layer from optical HR/VHR RS data: Concept and first results. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9*(6), 1978–1996. https://doi.org/10.1109/JSTARS.2016.2518468

Salem, M., & Tsurusaki, N. (2024). Impacts of Rapid Urban Expansion on Peri-Urban Landscapes in the Global South: Insights from Landscape Metrics in Greater Cairo. *Sustainability, 16*(6), 2316. https://doi.org/10.3390/su16062316

Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K. R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE, 109*(3), 247–278. https://doi.org/10.1109/JPROC.2021.3060483

Teke, A., & Kavzoglu, T. (2024). Exploring the decision-making process of ensemble learning algorithms in landslide susceptibility mapping: Insights from local and global explainable AI analyses. *Advances in Space Research, 74*(8), 3765–3785. https://doi.org/10.1016/j.asr.2024.06.082

THISME. (2020). *Buildings footprint from Spot-6 and Spot-7 images*. https://thisme.cines.teledetection.fr/home

UN-Habitat. (2020). *World cities report 2020: The value of sustainable urbanization*. United Nations Human Settlements Programme.

United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*. United Nations. https://sdgs.un.org/2030agenda

United Nations. (2019). *World urbanization prospects: The 2018 revision*. United Nations, Department of Economic and Social Affairs, Population Division.

Wang, Y., Yang, L., Liu, X., & Yan, P. (2024). An improved semantic segmentation algorithm for high-resolution remote sensing images based on DeepLabv3+. *Scientific reports*, *14*(1), 9716. https://doi.org/10.1038/s41598-024-60375-1

Wang, Y., Zeng, X., Liao, X., & Zhuang, D. (2022). B-FGC-Net: A Building Extraction Network from High Resolution Remote Sensing Imagery. *Remote Sensing*, *14*(2), 269. https://doi.org/10.3390/rs14020269

Yang, D., Gao, X., Yang, Y., Guo, K., & Xu, L. (2025). Advances and future prospects in building extraction from high-resolution remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18*(8), 6994–7016. https://doi.org/10.1109/JSTARS.2025.3538662

Yilmaz, E. O., & Kavzoglu, T. (2024). Burned area detection with Sentinel-2A data: Using deep learning techniques with explainable artificial intelligence. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-5-2024,* 251–257. https://doi.org/10.5194/isprs-annals-X-5-2024-251-2024

Yilmaz, E. O., & Kavzoglu, T. (2025). DeepSwinLite: A Swin Transformer-Based Light Deep Learning Model for Building Extraction Using VHR Aerial Imagery. *Remote Sensing*, *17*(18), 3146. https://doi.org/10.3390/rs17183146

Yılmaz, E. Ö., Teke, A., & Kavzoğlu, T. (2025). A performance analysis of U-Net and U-Net++ in building footprint extraction using XAI. *2025 33rd Signal Processing and*

Communications Applications Conference (SIU), Şile, İstanbul, Türkiye, 1–4. https://doi.org/10.1109/SIU66497.2025.11111828

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2881–2890). IEEE. https://doi.org/10.1109/CVPR.2017.660

Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine, 5*(4), 8–36. https://doi.org/10.1109/MGRS.2017.2762307