Turkish Journal of Remote Sensing and GIS
Türk Uzaktan Algılama ve CBS Dergisi

Turk J Remote Sens GIS, March 2025, 6(1): 119-130
Dergi Anasayfası: http://www.dergipark.org.tr/en/pub/rsgis

Vision Foundation Models and Rule-Based Approaches for Roof Surface Segmentation and Photovoltaic Potential Analysis in Urban Areas

Kentsel Alanlarda Çatı Yüzey Segmentasyonu ve Fotovoltaik Potansiyel Analizinde Görsel Temel Model ve Kural Tabanlı Yaklaşımlar

Samed Özdemir[1]*, Ahmet Yavuzdoğan[1]

[1]*Gümüşhane University, Faculty of Engineering and Natural Sciences, Geomatics Engineering, Gümüşhane/ Türkiye.*

ORIGINAL PAPER

***Corresponding author:**
Samed Özdemir
samed.ozdemir@gumushane.edu.tr

doi: 10.48123/rsgis.1606873

Article history
Received: 24.12.2024
Accepted: 06.03.2025
Published: 26.03.2025

Abstract

This study presents two methods for rapidly and effectively determining the photovoltaic (PV) potential of building roofs in urban areas using aerial photographs and point cloud data. In the first method, the Segment Anything Model (SAM) and Contrastive Language Image Pre-Training (CLIP) models are used to detect roof surfaces and obstacles from aerial photographs. In the second method, the Random Sample Consensus (RANSAC) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithms are employed to identify roof surfaces from Light Detection and Ranging (LiDAR) point clouds. Through the first proposed method, the performance of current deep learning approaches in 2.5D PV potential analysis is investigated, while the second approach examines the performance of 3D PV potential analysis compared to the 2D approach. In PV potential analysis, the Photovoltaic Geographical Information System (PVGIS) Application Programming Interface (API) was utilized. The analysis is conducted based on roof parameters obtained through both proposed methods. In building detection, the first approach achieved an Intersection over Union (IoU) score of 94.29%, whereas the second approach attained an IoU score of 91.23%.

***Keywords**: Deep learning, Photovoltaic potential, Point cloud, Roof segments, Semantic segmentation*

Özet

Bu çalışma, kentsel alanlarda bina çatılarının fotovoltaik (FV) potansiyelinin hava fotoğrafları ve nokta bulutu verileri üzerinden hızlı ve etkin bir şekilde belirlenmesi için iki yöntem sunulmaktadır. İlk yöntemde, hava fotoğraflarından çatı yüzeyleri ve engellerin tespiti için Segment Anything Model (SAM) ve Contrastive Language Image Pre-Training (CLIP) modelleri kullanılmaktadır. İkinci yöntemde ise Light Detection and Ranging (LiDAR) nokta bulutlarından çatı yüzeylerinin tespitinde Random Sample Consensus (RANSAC) ve Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algoritmaları kullanılmaktadır. Çalışmada önerilen ilk yöntem sayesinde güncel derin öğrenme yaklaşımlarının 2.5B FV potansiyel analizindeki başarımı araştırılırken, önerilen ikinci yaklaşım ile 3B FV potansiyel analizinin 2B yaklaşıma göre başarımı ele alınmaktadır. FV potansiyel analizinde, PhotoVoltaic Geographical Information System (PVGIS) Application Programming Interface (API)'si kullanılmıştır. Önerilen her iki yöntemle elde edilen çatı parametreleri üzerinden analiz edilmektedir. Bina tespitinde, ilk yaklaşım %94.29 IoU skoru elde ederken ikinci yaklaşım ile elde edilen IoU skoru %91.23 olmuştur.

***Anahtar kelimeler**: Çatı yüzeyi, Derin öğrenme, Fotovoltaik potansiyel, Nokta bulutu, Semantik segmentasyon*

1. Introduction

Cities, now home to over 55% of the global population, are central to addressing the challenges posed by energy consumption and carbon dioxide (CO_2) emissions, which they contribute to at an alarming rate—approximately two-thirds of the global total. The rapid increase in these emissions is largely driven by population growth, urbanization, and increased economic activities within these densely populated areas (Ranalder et al., 2021). In response to these challenges, numerous cities and local governments have set ambitious goals to develop net zero-carbon emission buildings, aiming for greater energy self-sufficiency (Li & Han, 2022). The increasing feasibility of photovoltaic (PV) systems is supported by the widespread availability of unused urban rooftops, combined with declining costs and simplified installation processes, making solar energy a key solution to urban energy demands (Minelli et al., 2023, Huang et al., 2022). Due to the significant potential of rooftop solar PV systems, accurately estimating the total installed solar PV capacity and power generation is essential for guiding policymakers and stakeholders (Mao et al., 2023).

Satellite and aerial imagery play a crucial role in various geospatial applications, including the identification of land cover types (Kavzoglu et al., 2024; Sahak et al., 2023; Topaloğlu et al., 2022; Yagmur et al., 2022), road mapping (Ozturk et al., 2022) and building extraction (Comert & Kaplan, 2018; Tonbul & Kavzoglu, 2020). The introduction of LiDAR technology opened new frontiers in the remote sensing community, enabling more precise and efficient road mapping (Ma et al., 2022; Ozturk et al., 2023; Zhao et al., 2019), forestry (Özdemir et al., 2021), land cover (Akumu & Dennis, 2023; Hansch & Hellwich, 2021) mapping applications. In recent years, the growing demand for PV energy production has made PV potential analysis a critical area of research where advancements in the remote sensing technology made it a very viable tool for renewable energy applications (Huang et al., 2022; Lan et al., 2021; Stack & Narine, 2022; Zhong et al., 2021). PV potential studies can be broadly classified into 2D, 2.5D, and 3D.based on the data type used (Özdemir et al., 2023). 2D approaches utilize remotely sensed imagery acquired from aerial or satellite platforms to detect and segment roof surfaces using image processing or machine learning techniques. This approach assumes all building roofs as flat and uninterrupted surfaces, which leads to an overestimation of PV potential and makes it unsuitable for accurate applications. In contrast, 2.5D approaches aim to mitigate the inaccuracies of 2D methods by employing Digital Surface Models (DSMs) derived from point clouds or 3D city models. Similar to 2D approaches, building roofs are extracted using image processing or machine learning techniques. The extracted roof boundaries are then matched with the DSM to acquire tilt and azimuth values, which are subsequently used for PV potential estimation. While 2.5D techniques offer a more accurate PV potential estimation compared to 2D approaches, converting 3D data to 2.5D results in information loss, along with increased processing time. 3D approaches, on the other hand, employ 3D city models or point clouds directly without any data conversion, resulting in more accurate and effective PV potential estimates. However, acquiring and processing LiDAR data is a complex task due to the irregular nature of point clouds, and not all city-scale 3D models are readily available.

This study presents two approaches for analyzing rooftop PV potential, one utilizing 2.5D deep learning techniques and the other employing 3D point clouds. The proposed methods were tested on the ISPRS Vaihingen dataset. In the first approach, zero-shot segmentation of the aerial imagery of the study area was carried out with Segment Anything Model (SAM), which is a state of the art Visual Foundation Model (VFM). Subsequently, the Contrastive Language-Image Pre-Training (CLIP) model was employed to zero-shot labelling of the extracted image segments. SAM was then re-applied to each building segment to detect rooftop surfaces and any obstacles on them. The second approach involved using LiDAR point clouds of the study area, where roof surfaces were identified through the RANSAC and DBSCAN algorithms employing an open source QGIS plugin called Solar Potential Analyzer (SPAN) developed by (Özdemir et al., 2023). In determining roof surface parameters, the first method calculates slope, aspect, and area of individual rooftops using the DSM, while the second method directly derives these parameters from the point cloud data. For the PV potential analysis of roof surfaces, the PVGIS database (https://re.jrc.ec.europa.eu/pvg_tools/en/) was utilized. Queries generated based on the calculated rooftop parameters and geographic locations were sent through an API to the PVGIS database to retrieve the daily, monthly, and annual PV production data for each rooftop surface. In the roof segmentation stage, the first approach using on SAM and CLIP achieved an IoU score of 97.94% on the first study area and 94.29% IoU score in the second study area. For the second approach, IoU scores of 76.04% were achieved on the first study area, followed by scores of 91.23% on the second study area. Both approaches yielded successful results in PV potential analysis and demonstrated scalability, making them suitable for applications ranging from individual buildings to city-wide scales.

2. Material and methods

This section provides a detailed description of the methods and datasets used in the segmentation of rooftop surfaces and the analysis of PV potential from aerial imagery and point clouds. Initially, the study areas and datasets employed

for PV potential analysis are introduced. Following this, the algorithms and deep learning models utilized for rooftop surface segmentation from aerial imagery and point clouds are presented, along with a step-by-step explanation of the segmentation process. The accuracy metrics used for evaluating the segmentation results are also introduced. Finally, the methodology employed for conducting the PV potential analysis of the identified rooftop surfaces is described.

2.1. Dataset and Study Area

The ISPRS Vaihingen dataset was collected during flights conducted in 2008 in the Vaihingen region of Germany by the German Association of Photogrammetry and Remote Sensing (DGPF) to test digital aerial cameras (Cramer, 2010). The dataset consists of BGNIR (Blue, Green, and Near-Infrared) orthophotos with a spatial resolution of 8 cm, a LiDAR point cloud with a density of approximately 6.7 points/m², and a DSM with a spatial resolution of 25 cm derived from the point cloud. For this study, two sub-areas from the Vaihingen dataset, referred to as Area 1 and Area 2, were selected (Figure 1). Area 1, measuring 76 × 70 meters, contains six buildings with similar roof types but varying sizes and few rooftop details. In contrast, Area 2, measuring 86 × 64 meters, contains four buildings with more complex rooftop surfaces and numerous rooftop details compared to Area 1.

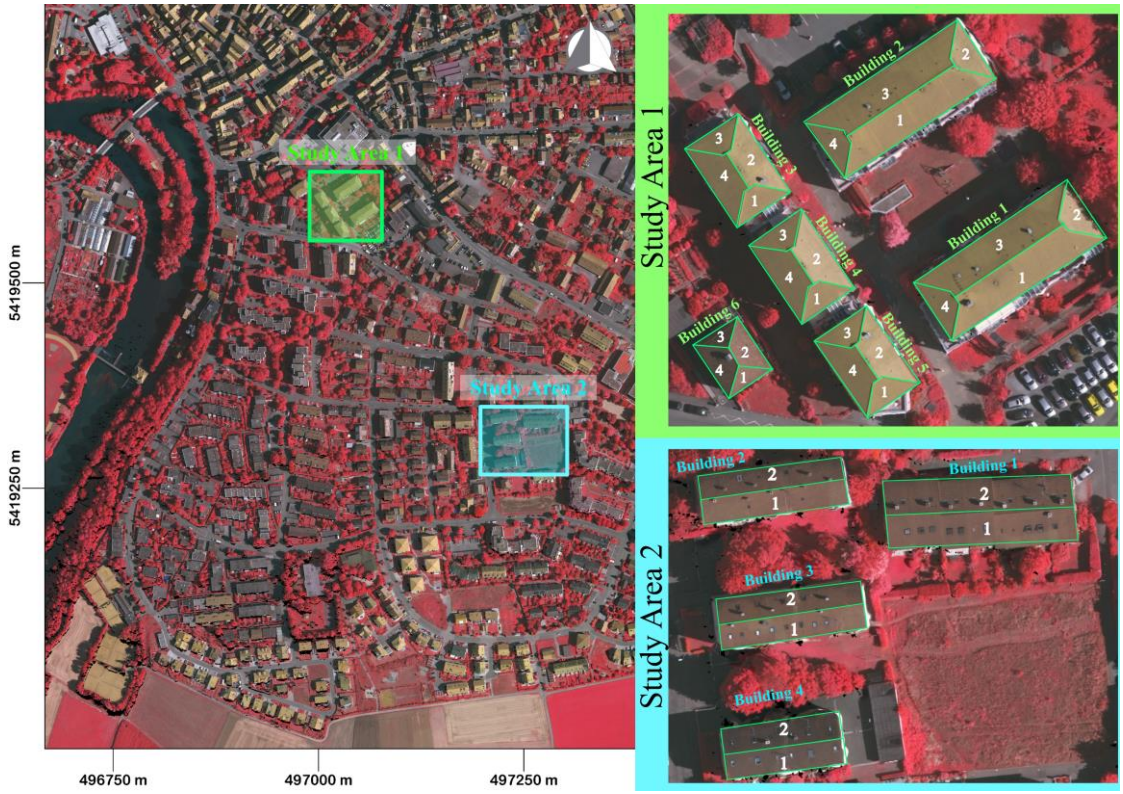

Figure 1. Selected study areas from the ISPRS Vaihingen dataset

2.2. Roof Surface Segmentation Pipeline

Two distinct methods have been proposed for the detection of roof surfaces. The first method involves using Vision Foundation Models (VFM) on aerial photographs to identify roof surfaces, employing the Segment Anything model and CLIP VFM during this phase. The second method detects roof surfaces from aerial LiDAR point clouds using rule-based approaches with RANSAC and DBSCAN algorithms. The PV potential of the identified roof surfaces has been calculated using the PVGIS database. The proposed methods are detailed in the following sections with a comprehensive analysis of each approach's methodology and outcomes.

2.2.1. Roof Surface Segmentation from Aerial Imagery

Orthophoto images have been segmented individually for each roof surface using the "Segment Anything Model" and "Contrastive Language-Image Pre-Training (CLIP)" models. The "Segment Anything Model" has been developed by META AI as a foundational model for image segmentation, trained on the SA-1B dataset (Kirillov et al., 2023).

The SAM model, incorporating a Vision Transformer (ViT), is capable of generating segmentation masks for any type of object within an image. The operation of the SAM model for image segmentation involves either placing points on the object to be segmented or drawing a bounding box around it. Additionally, for fully automated segmentation, the model can also perform this process by placing points on the image at predetermined intervals. Fundamentally, the SAM model comprises three main components: an Image Encoder, a Prompt Encoder, and a Mask Decoder (Figure 2).

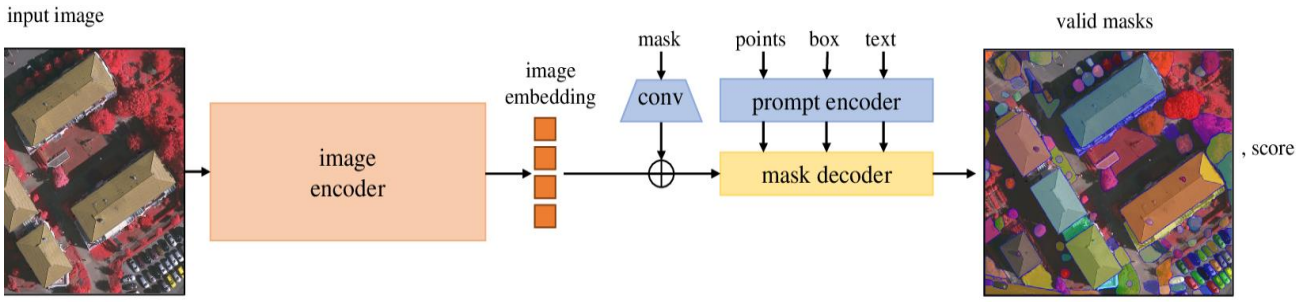

Figure 2. Segment anything model architecture (Kirillov et al., 2023)

CLIP (Contrastive Language–Image Pretraining), developed by OpenAI, is a deep learning model capable of learning relationships between language and image data (Radford et al., 2021). The CLIP model is designed to establish meaningful connections between text and images and to apply this knowledge across broad applications. One of the key features of the CLIP model is zero-shot learning, allowing it to adapt to new tasks that it has not encountered during training. The CLIP model can perform image-text matching using instructions given in natural language. The architecture of CLIP utilizes two separate networks to process language and image data: a Vision Transformer (ViT) model and a language model (a Transformer-based text encoder) (Figure 3). CLIP enhances the learning of semantic congruence between images and associated text by bringing the image and its related text closer together. The training of the CLIP model employs a contrastive loss function that aims to maximize the difference between positive and negative examples. A positive example represents a match between an image and a correctly describing text, while a negative example denotes a match between an image and a randomly selected misleading text.

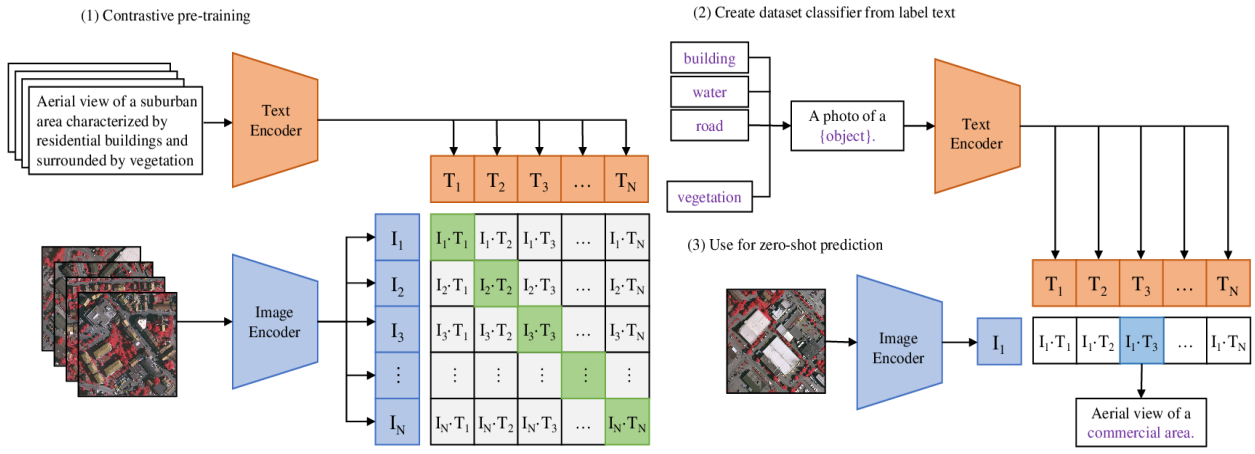

Figure 3. CLIP model architecture (Radford et al., 2021)

The aerial photographs of the study area were conditioned to segment all objects within the photographs using the SAM model. Subsequently, all the generated segments were classified into two categories—building and non-building—based on the query “Is this a building?” sent to the CLIP model. Segments labeled as buildings were further segmented from the aerial photographs and then re-submitted to the SAM model to ensure segmentation of all objects again. Segments smaller than 100×100 pixels were discarded, leaving only roof surfaces. After the segmentation process of buildings and roof surfaces from the images was completed, azimuth, slope (through DSM), and area information for each roof surface were calculated. Using the Global Horizontal Irradiance (GHI) data obtained via the PVGIS API, a PV potential estimate was performed with a module developed in a Python environment. The workflow of the applied approach is depicted in Figure 4.

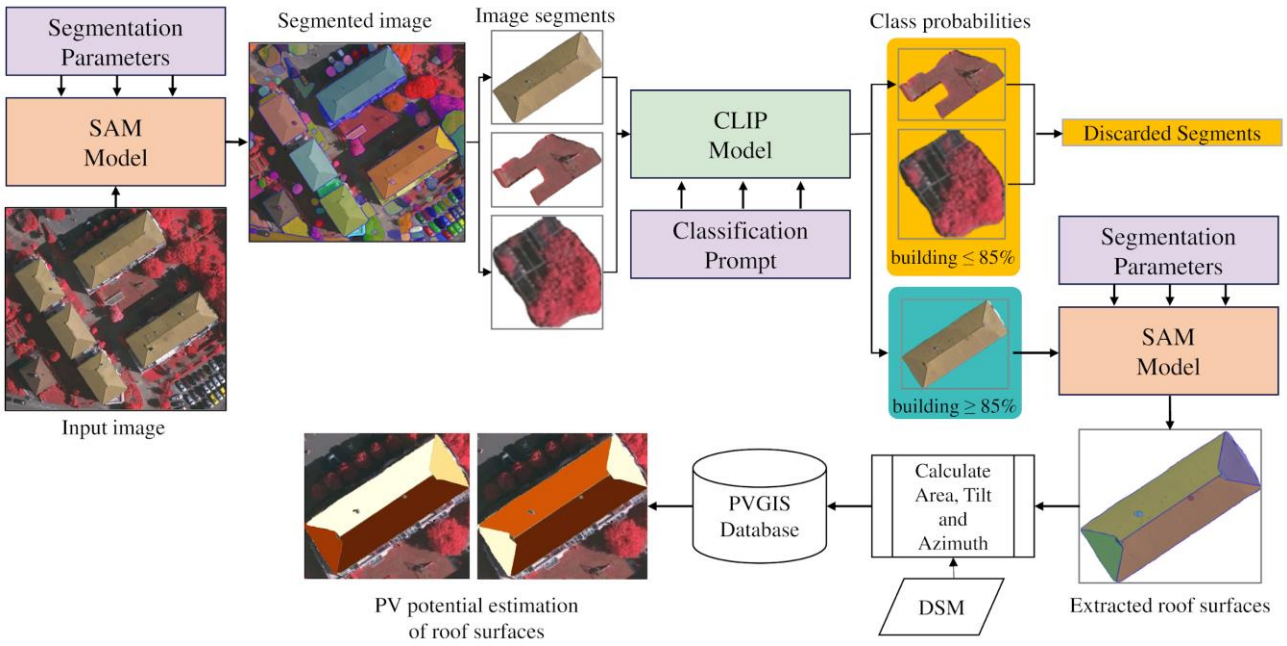

Figure 4. Detection of roof surfaces from aerial photographs and calculation of PV potential

2.2.2. Roof Surface Segmentation from Point Clouds

In the segmentation of roof surfaces from point clouds, we employed SPAN (Özdemir et al., 2023) which has a simple yet highly effective two-step methodology using the well-known and widely used algorithms: Random Sample Consensus (RANSAC) (Fischler & Bolles, 1981) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al., 1996) (Figure 5).

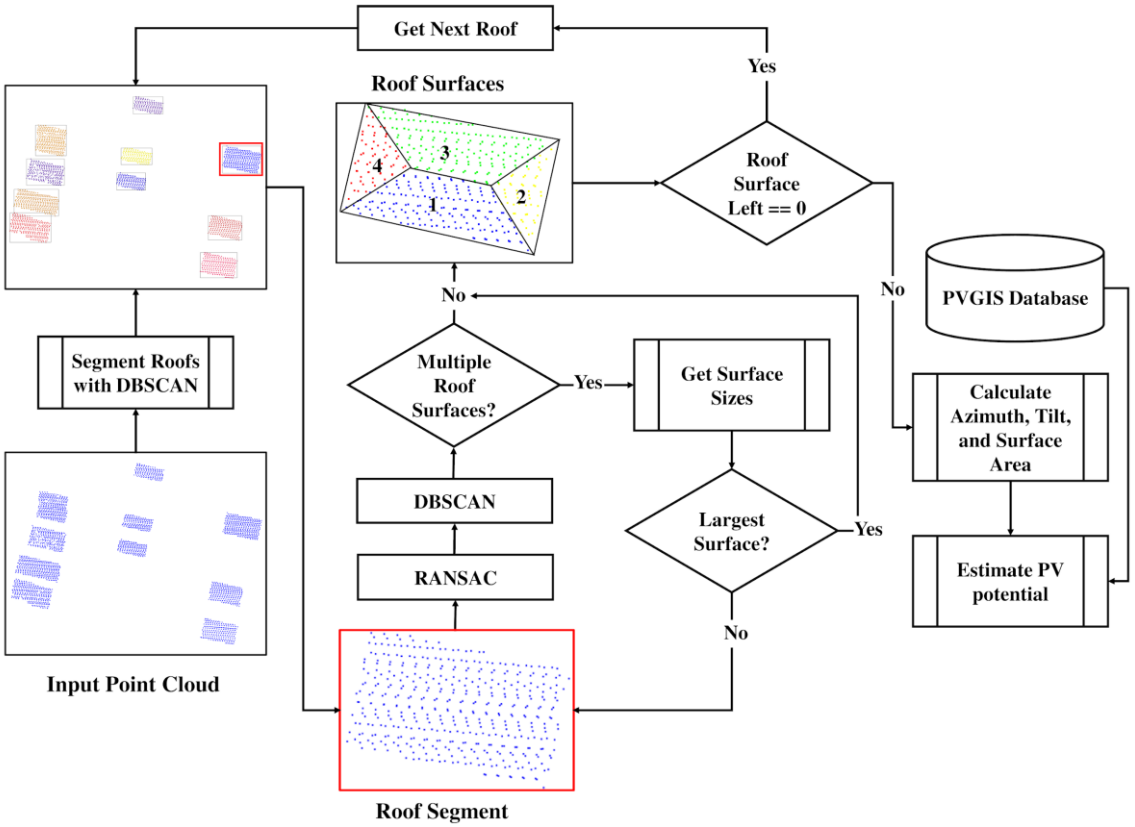

Figure 5. Detection of roof surfaces from point clouds and calculation of PV potential (Özdemir et al., 2023)

In the first step of the SPAN plugin, the input LiDAR point cloud is segmented on a building-by-building basis with the help of the DBSCAN clustering algorithm. This process enables the effective and efficient identification and differentiation of each building at an urban scale. After the building segments are created, the roof surfaces of each building are detected using the RANSAC algorithm. During the phase of separating roof surfaces with the RANSAC algorithm, planes fitting different surfaces of the roof segments are created. However, occasionally, one of the planes generated by RANSAC may overlap with multiple roof segments. In such cases, the DBSCAN clustering algorithm intervenes to determine the largest segment among the overlapping ones. Subsequently, the remaining points are returned to the original point cloud and reprocessed in the next iteration. It's worth noting that the SPAN plugin can only process point clouds containing roof points, which requires prior segmentation of the point clouds.

2.3. PV Potential Analysis

There are various databases, software, and web services available for PV potential analysis, such as PVGIS, Solar Monkey (SM), and Global Solar Atlas. In this study, the PVGIS database developed by the European Commission was utilized, which has been proven to produce results very close to on-site observations (Moudrý et al., 2019; Psiloglou et al., 2020; Suri et al., 2008). PVGIS is a comprehensive and up-to-date database for estimating PV potential, taking into account factors such as surface slope, azimuth, solar panel technology, system loss, and the peak power of the PV panel. The PVGIS database also has the capability to respond to queries containing information such as geographic location, slope, and orientation sent over its Application Programming Interface (API), providing data on Global Horizontal Irradiance (GHI) and PV production.

In this study, the PV potential analysis of roof surfaces detected from aerial photographs and point clouds was conducted using the PVGIS database. For both detection methods, a query file was created by adding solar radiation database information, PV technology, installed peak power, system loss, and economic parameters to the tilt, azimuth, and geographic location information of the detected roof surfaces, and sent to the database. PVGIS servers respond to the query by creating a JSON file that provides comprehensive PV information for each roof surface. The PV potential of each roof surface is calculated using the PV information and the roof surface area obtained from the JSON file.

2.4. Accuracy Assessment

The evaluation of the proposed approaches was conducted by comparing automatically extracted roofs from both aerial imagery and point clouds with reference data, utilizing overall accuracy, precision, recall, F1-score and IoU metrics as shown in Equations (1), (2), (3), (4) and (5).

$$OverallAccuracy = \frac{(TruePositive + TrueNegative)}{NumberofSamples} \quad (1)$$

$$Precision = \frac{TruePositive}{(TruePositive + FalsePositive)} \quad (2)$$

$$Recall = \frac{TruePositive}{(TruePositive + FalseNegative)} \quad (3)$$

$$F1Score = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \quad (4)$$

$$IoU = \frac{TruePositive}{(TruePositive + FalsePositive + FalseNegative)} \quad (5)$$

In Equations (1), (2), and (3), an element identified as an object that matches an object in the reference data is classified as a true positive. An element that matches an object in the reference but is mistakenly classified as background is deemed a false negative. Conversely, an element incorrectly identified as an object when it does not match any object in the reference is considered a false positive. Lastly, a true negative is defined as an element that is correctly identified as background in both the classified and reference data. The Intersection over Union (IoU) metric, also known as the Jaccard Index, quantifies how much the predicted object overlaps with the ground truth object (5).

3. Results

In this section, we thoroughly examine the segmentation results obtained from both developed approaches. Additionally, we evaluate the PV potential analysis outcomes, offering insights into the effectiveness and precision of the methodologies used.

3.1. Roof Surface Segmentation from Aerial Imagery with VFMs

Although the SAM model is trained solely on the RGB (Red, Green, Blue) color space, the segmentation carried out on aerial photographs in the NIR-RGB (Near-Infrared, Green, Blue) color space from the ISPRS Vaihingen dataset has successfully identified many segments according to the model's internal performance metrics, IoU and stability score. The internally calculated segmentation accuracies of the SAM model are approximately 99% IoU and ~98% Stability Score for Study Area 1 and 2. These results demonstrate that the SAM model can effectively detect object segmentation boundaries even in different color spaces.

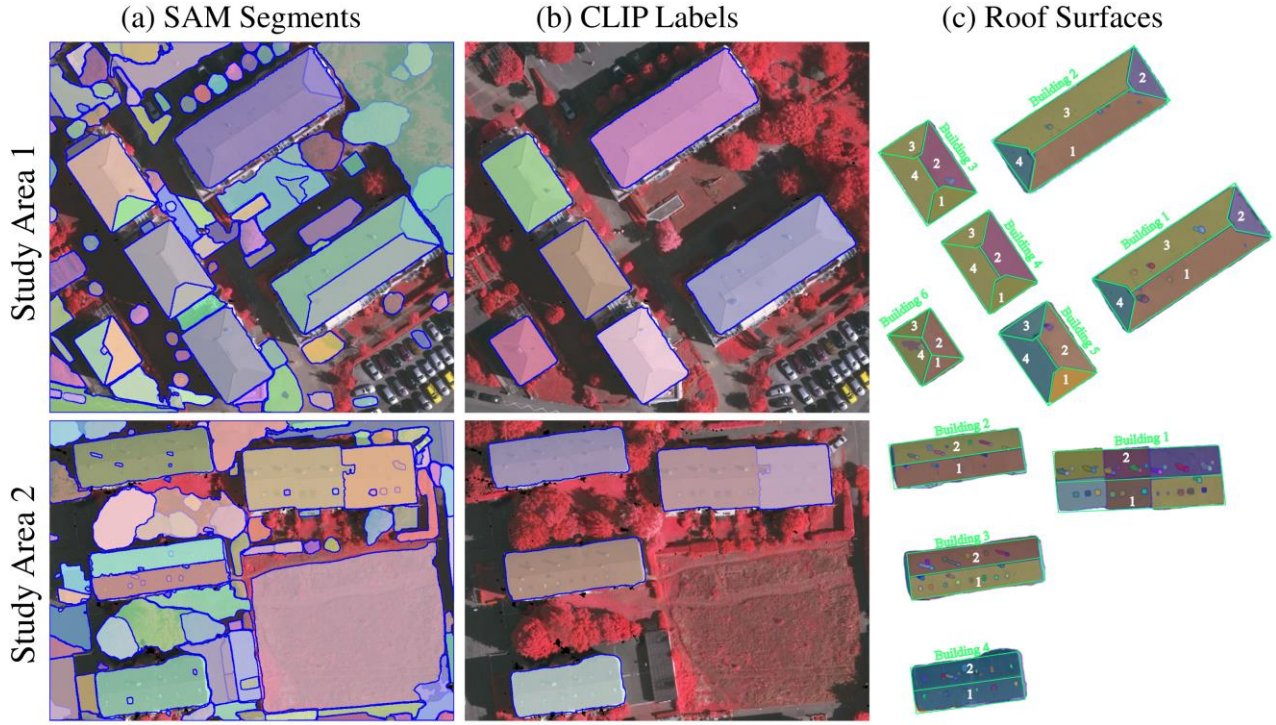

Figure 6. Roof segmentation results: (a) SAM segmentation results, (b) Extracted segments labelled with CLIP, (c) Final roof surfaces extracted with SAM

However, since the SAM model does not possess semantic information about the segments it produces, the identified segments need to be semantically interpreted. At this stage of semantically labeling the SAM segments, the CLIP model was utilized. With CLIP, it was verified whether a segment produced by the SAM model belonged to the roof class or not. Like the SAM model, the CLIP model is also trained only on RGB color space images. However, as shown in Figure 6, CLIP has successfully labeled building segments with approximately 99% Text Probability Score. Moreover, at this stage, it is more clearly visible that the SAM model can almost perfectly detect the actual building boundaries.

Table 1. Roof segmentation accuracy of the SAM and CLIP models

	Study Area 1	Study Area 2
Precision	98.71%	95.00%
Recall	99.21%	99.21%
F1 Score	98.96 %	97.06%
IoU	97.94%	94.29%
Overall Accuracy	99.48%	98.65%

In the last step of the analysis, segments identified as buildings by the CLIP model were reintroduced into the SAM model to carry out the detection of individual roof surfaces and any obstacles present on them. The detected obstacles on the roof surfaces were subsequently removed from the PV potential analysis to improve the accuracy of the estimation. The input images were exclusively composed of building roofs, and the SAM model was specifically conditioned to segment every detail within these images. Under these conditions, the SAM model demonstrated remarkable segmentation capabilities, achieving an overall accuracy exceeding 92.5%, an IoU of 91.2%, and an F1 score of 97.6%.

Table 2. Roof surface segmentation accuracy of the SAM model

	Study Area 1	Study Area 2
Precision	100%	100%
Recall	96.08%	95.36%
F1 Score	97.99%	97.61%
IoU	96.08%	91.23%
Overall Accuracy	92.54%	95.36%

3.2. Roof Surface Segmentation from Point Cloud using Rule Based Approaches

The proposed two-step approach for extracting roof surfaces from point clouds has achieved successful results in identifying roof surfaces within the study areas (Figure 7). This method demonstrated significant effectiveness, achieving an overall accuracy of over 92.5% and an F1 score of more than 90.7%. Additionally, with an accuracy of over 76.0% in the IoU metric, this approach has proven to be highly effective for roof surface segmentation from point clouds (Table 3).

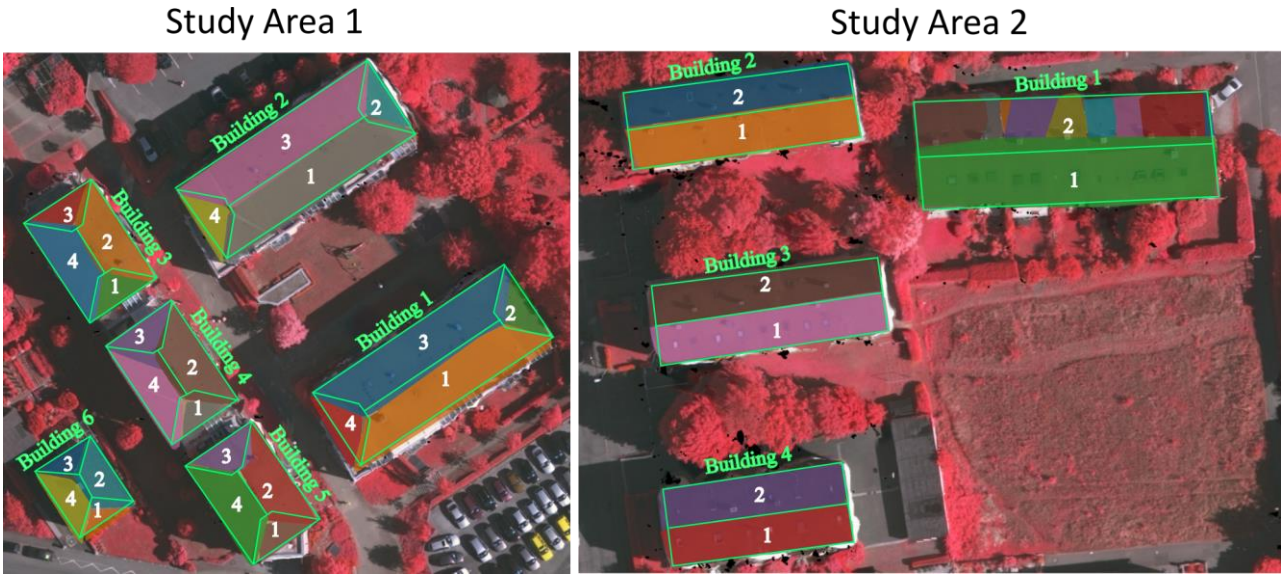

Figure 7. Point cloud roof surface segmentation results

Table 3. Accuracy assessment of roof surface extraction from point cloud

	Study Area 1	Study Area 2
Precision	84.52%	100%
Recall	84.52%	95.36%
F1 Score	90.76%	97.61%
IoU	76.04%	91.23%
Overall Accuracy	92.54%	95.36%

3.3. Photovoltaic Potential Analysis

The PV potential analysis was conducted using the PVGIS database through the PVGIS API. The tilt and azimuth information for each roof surface was calculated from the roof surfaces extracted using the two proposed approaches. For roof surfaces detected from aerial photographs, the tilt and azimuth data were calculated using the DSM, whereas for roof surfaces identified from point clouds, these measurements were directly computed from the point cloud data itself. In the queries sent to the PVGIS database, the PVGIS-SARAH 2 was selected as the solar radiation database. Crystalline Silicon panels, commonly used in current PV panel technology (Ballif et al., 2022; Kettle et al., 2022), were chosen as the installed PV panel. The system loss parameter was set at the default value of 14%. In the queries sent to the PVGIS database, geographic location data for each roof surface was also included. Additionally, the effect of the horizon was considered in calculating the annual PV potential.

In the final stage of the analysis using the SAM model, the roof surfaces obtained were overlaid with the DSM corresponding to the study areas to extract elevation data for each roof surface from the DSM. Subsequently, this elevation data was utilized to separately calculate the slope and azimuth for each of the roof surfaces. The derived slope and azimuth data were then employed to estimate the annual irradiation received by the roof surfaces and the potential PV production values, utilizing the PVGIS database via the PVGIS API (Table 4).

To avoid presenting multiple results of the same roof surface in Table 4 due to the fragmented segmentation of the LiDAR point cloud in Study Area 2 second roof surface of Building 1 we opted to merge the PV estimation results to provide a clearer presentation of the PV potential estimation.

Table 4. Yearly PV output (kWh) of extracted roof surfaces with the SPAN and SAM+CLIP approach

		Study Area 1				Study Area 2			
		SAM+CLIP	**SPAN**	**Difference (%)**		**SAM+CLIP**	**SPAN**	**Difference (%)**
Building 1	Surface 1	47608,47	46748,48	1,84	**Building 1**	Surface 1	77220,97	83968,96	-8,04
	Surface 2	9414,16	6329,61	48,73		Surface 2	64663,95	47604,74	35,84
	Surface 3	43545,17	47245,33	-7,83		**Total**	141884,92	131573,70	7,84
	Surface 4	9857,67	8317,07	18,52					
	Total	110425,47	108640,49	1,64					
					Building 2	Surface 1	36731,13	40929,42	-10,26
Building 2	Surface 1	47608,47	48581,24	-2,00		Surface 2	42866,74	36607,20	17,10
	Surface 2	9414,16	6726,08	39,96		**Total**	79597,87	77536,62	2,66
	Surface 3	43545,17	46039,98	-5,42					
	Surface 4	9857,67	7967,23	23,73	**Building 3**	Surface 1	36518,81	35596,52	2,59
	Total	110425,47	109314,53	1,02		Surface 2	43365,19	45332,24	-4,34
						Total	79884,00	80928,76	-1,29
Building 3	Surface 1	9521,69	6617,57	43,89					
	Surface 2	17180,84	19324,83	-11,09	**Building 4**	Surface 1	55807,46	28638,36	94,87
	Surface 3	8709,03	5180,83	68,10		Surface 2	-	30269,61	
	Surface 4	17990,26	18521,33	-2,87		**Total**	55807,46	58907,97	-5,26
	Total	53401,82	49644,57	7,57					
Building 4	Surface 1	9527,27	5681,14	67,70					
	Surface 2	17249,18	18942,32	-8,94					
	Surface 3	8756,91	8934,96	-1,99					
	Surface 4	18016,21	14327,13	25,75					
	Total	53549,57	47885,55	11,83					
Building 5	Surface 1	9521,69	5536,21	71,99					
	Surface 2	17180,84	16605,51	3,46					
	Surface 3	8709,03	7416,15	17,43					
	Surface 4	17990,26	18855,86	-4,59					
	Total	53401,82	48413,72	10,30					
Building 6	Surface 1	13238,77	2713,95	387,81					
	Surface 2	-	8968,40	-					
	Surface 3	13099,01	2894,75	352,51					
	Surface 4	-	6880,43	-					
	Total	26337,78	21457,53	22,74					

For both study areas, the two approaches yielded similar results in PV potential estimation regarding overall building PV production. However, inspection of individual roof surfaces revealed greater differences in estimated PV production for some roofs. These discrepancies arise from the characteristics of the data and the methods applied. Point cloud data can more accurately determine the height of objects in the scene, enabling precise determination of azimuth and tilt values. Conversely, accurately identifying object boundaries with linear formations, such as buildings, from point clouds is challenging due to point sampling deficiency. Additionally, fitting a plane to a roof surface using the RANSAC algorithm sometimes results in surfaces that do not exactly match the actual roof. In image-based roof surface extraction, limitations such as occlusions and shadows affect accuracy. Occlusions lead to inevitable loss of surface boundaries, while shadowed roof segments were relatively successfully handled by the SAM. Despite these setbacks, both approaches estimated PV potential values within a 20% discrepancy, with most buildings exhibiting even lower differences.

4. Discussion

The integration of the SAM and CLIP models has enabled for accurate detection of building rooftops, their individual surfaces, and obstacles such as chimneys and windows, thereby providing a more precise basis for PV potential analysis. The SAM, trained on terrestrial RGB imagery, was successfully applied to aerial imagery in BGNIR format and achieved notable accuracy metrics for two different study areas. Specifically, for Study Area 1 and Study Area 2, roof and obstacle segmentation were completed with 92.54% and 95.36% overall accuracy, 96.08% and 91.23% IoU, and 97.99% and 97.61% F1 scores, respectively. However, obstacles like chimneys sometimes included shadowed areas, slightly reducing the accuracy of surface area calculations. Optimal parameterization of the SAM model and the resolution of the input images, as well as careful prompt selection for the CLIP model, were found to be critical for achieving these performance levels. In terms of runtime, SAM required approximately one minute per image to complete segmentation using a T4 GPU (including module and image loading), while Open CLIP completed segmentation in about two seconds per image segment on an A100 GPU and seven seconds on a T4 GPU.

The RANSAC algorithm within the SPAN plugin was similarly effective for point cloud segmentation. It achieved comparable accuracy and quality metrics—with 92.54% and 95.36% overall accuracy, 76.04% and 91.23% IoU, and 90.76% and 97.61% F1 scores for Study Area 1 and Study Area 2, respectively—demonstrating that it is both a reliable and computationally efficient method. However, factors like low point density and outliers on certain building roofs can negatively impact its accuracy. Results may be improved by employing higher-density point clouds or pre-processing steps to remove outliers. Unlike many existing methods that treat an entire building as a single roof surface and thus rely on a uniform tilt angle and azimuth, our approach evaluates each roof surface separately. This refinement enhances the accuracy of PV potential analyses and highlights the ability of the SPAN plugin for more accurate and detailed rooftop segmentation.

5. Conclusion

The proposed methods demonstrated a high level of effectiveness in PV potential analysis derived from both aerial imagery and LiDAR point clouds. The approach utilizing the SAM and CLIP successfully delineated rooftop surfaces and identified obstacles, enabling a more accurate estimation of PV potential. This outcome highlights the capacity of VFMs to support complex tasks, such as obstacle detection on rooftop surfaces. Similarly, the point cloud-based method efficiently extracted roof surfaces, reinforcing its potential as a reliable and robust alternative.

Future research will focus on refining VFM-based rooftop extraction and minimizing reliance on auxiliary data for determining surface azimuth and tilt. This objective will be pursued by integrating VFM-based depth estimation methods. In parallel, considering the increasing accessibility and precision of LiDAR data, particularly its capacity to penetrate canopy cover, we plan to enhance the generalizability of the proposed workflow through the incorporation of deep learning strategies. Such improvements are expected to further advance the accuracy, scalability, and adaptability of PV potential assessments in diverse urban environments.

References

Akumu, C. E., & Dennis, S. (2023). Exploring the addition of airborne LiDAR-DEM and derived TPI for urban land cover and land use classification and mapping. *Photogrammetric Engineering & Remote Sensing, 89*(1), 19–26.

Ballif, C., Haug, F.-J., Boccard, M., Verlinden, P. J., & Hahn, G. (2022). Status and perspectives of crystalline silicon photovoltaics in research and industry. *Nature Reviews Materials, 7*(8), 597–616.

Comert, R., & Kaplan, O. (2018). Object-based building extraction and building period estimation from unmanned aerial vehicle data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 4*(3), 71–76. https://doi.org/10.5194/isprs-annals-IV-3-71-2018

Cramer, M. (2010). The DGPF-test on digital airborne camera evaluation: Overview and test design. *Photogrammetrie-Fernerkundung-Geoinformation*, *2*, 73–82.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996, August 2–6). *A density-based algorithm for discovering clusters in large spatial databases with noise* [Conference presentation]. 2nd International Conference on Knowledge Discovery and Data Mining, Portland, Oregon.

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM, 24*(6), 381–395.

Hansch, R., & Hellwich, O. (2021). Fusion of multispectral LiDAR, hyperspectral, and RGB data for urban land cover classification. *IEEE Geoscience and Remote Sensing Letters, 18*(2), 366–370.

Huang, X., Hayashi, K., Matsumoto, T., Tao, L., Huang, Y., & Tomino, Y. (2022). Estimation of rooftop solar power potential by comparing solar radiation data and remote sensing data—A case study in Aichi, Japan. *Remote Sensing, 14*(7), Article 1742. https://doi.org/10.3390/rs14071742

Kavzoglu, T., Colkesen, I., Atesoglu, A., Tonbul, H., Yilmaz, E. O., Ozlusoylu, S., & Ozturk, M. Y. (2024). Construction and implementation of a poplar spectral library based on phenological stages for land cover classification using high-resolution satellite images. *International Journal of Remote Sensing, 45*(6), 2049–2072.

Kettle, J., Aghaei, M., Ahmad, S., Fairbrother, A., Irvine, S., Jacobsson, J. J., Kazim, S., Kazukauskas, V., Lamb, D., Lobato, K., Mousdis, G. A., Oreski, G., Reinders, A., Schmitz, J., Yilmaz, P., & Theelen, M. J. (2022). Review of technology-specific degradation in crystalline silicon, cadmium telluride, copper indium gallium selenide, dye-sensitised, organic and perovskite solar cells in photovoltaic modules: Understanding how reliability improvements in mature technologies can enhance emerging technologies. *Progress in Photovoltaics: Research and Applications, 30*(12), 1365–1392. https://doi.org/10.1002/pip.3577

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). *Segment anything*. arXiv. http://arxiv.org/abs/2304.02643

Lan, H., Gou, Z., & Xie, X. (2021). A simplified evaluation method of rooftop solar energy potential based on image semantic segmentation of urban streetscapes. *Solar Energy, 230*, 912–924.

Li, S. Y., & Han, J. Y. (2022). The impact of shadow covering on the rooftop solar photovoltaic system for evaluating self-sufficiency rate in the concept of nearly zero energy building. *Sustainable Cities and Society*, *80*, Article 103821. https://doi.org/10.1016/j.scs.2022.103821

Ma, H., Ma, H., Zhang, L., Liu, K., & Luo, W. (2022). Extracting urban road footprints from airborne LiDAR point clouds with PointNet++ and two-step post-processing. *Remote Sensing, 14*(3), Article 789. https://doi.org/10.3390/rs14030789

Mao, H., Chen, X., Luo, Y., Deng, J., Tian, Z., Yu, J., Xiao, Y., & Fan, J. (2023). Advances and prospects on estimating solar photovoltaic installation capacity and potential based on satellite and aerial images. *Renewable and Sustainable Energy Reviews, 179*, Article 113276. https://doi.org/10.1016/j.rser.2023.113276

Minelli, F., D'Agostino, D., Migliozzi, M., Minichiello, F., & D'Agostino, P. (2023). PhloVer: a modular and integrated tracking photovoltaic shading device for sustainable large urban spaces—preliminary study and prototyping. *Energies*, *16*(15), Article 5786. https://doi.org/10.3390/en16155786

Moudrý, V., Beková, A., & Lagner, O. (2019). Evaluation of a high-resolution UAV imagery model for rooftop solar irradiation estimates. *Remote Sensing Letters, 10*(11), 1077–1085.

Özdemir, S., Akbulut, Z., Karsli, F., & Acar, H. (2021). Automatic extraction of trees by using multiple return properties of the LiDAR point cloud. *International Journal of Engineering and Geosciences, 6*(1), 20–26.

Özdemir, S., Yavuzdoğan, A., Bilgilioğlu, B. B., & Akbulut, Z. (2023). SPAN: An open-source plugin for photovoltaic potential estimation of individual roof segments using point cloud data. *Renewable Energy, 216*, Article 119022. https://doi.org/10.1016/j.renene.2023.119022

Ozturk, O., Isik, M. S., Sariturk, B., & Seker, D. Z. (2022). Generation of Istanbul road data set using Google Map API for deep learning-based segmentation. *International Journal of Remote Sensing, 43*(8), 2793–2812.

Ozturk, O., Isik, M. S., Kada, M., & Seker, D. Z. (2023). Improving road segmentation by combining satellite images and LiDAR data with a feature-wise fusion strategy. *Applied Sciences, 13*(10), Article 6161. https://doi.org/10.3390/app13106161

Psiloglou, B. E., Kambezidis, H. D., Kaskaoutis, D. G., Karagiannis, D., & Polo, J. M. (2020). Comparison between MRM simulations, CAMS and PVGIS databases with measured solar radiation components at the Methoni station, Greece. *Renewable Energy, 146*, 1372–1391. https://doi.org/10.1016/j.renene.2019.07.064

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). *Learning transferable visual models from natural language supervision*. arXiv. http://arxiv.org/abs/2103.00020

Ranalder, L., Busch, H., Hansen, T., Brommer, M., Couture, T., Gibb, D., ... & Sverrisson, F. (2021). *Renewables in cities: 2021 global status report*. REN21 Secretariat. https://www.ren21.net/wp-content/uploads/2019/05/GSR2021_Full_Report.pdf

Sahak, A. S., Karsli, F., Gormus, E. T., & Ahmadi, K. (2023). Seasonal monitoring of urban heat island based on the relationship between land surface temperature and land use/cover: A case study of Kabul City, Afghanistan. *Earth Science Informatics, 16*(1), 845–861.

Stack, V., & Narine, L. L. (2022). Sustainability at Auburn University: Assessing rooftop solar energy potential for electricity generation with remote sensing and GIS in a Southern US campus. *Sustainability, 14*(2), Article 626. https://doi.org/10.3390/su14020626

Suri, M., Huld, T., Cebecauer, T., & Dunlop, E. D. (2008). Geographic aspects of photovoltaics in Europe: Contribution of the PVGIS website. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 1*(1), 34–41. https://doi.org/10.1109/JSTARS.2008.2001431

Tonbul, H., & Kavzoglu, T. (2020). Semi-automatic building extraction from WorldView-2 imagery using Taguchi optimization. *Photogrammetric Engineering & Remote Sensing, 86*(9), 547–555.

Topaloğlu, R. H., Aksu, G. A., Ghale, Y. A. G., & Sertel, E. (2022). High-resolution land use and land cover change analysis using GEOBIA and landscape metrics: A case of Istanbul, Turkey. *Geocarto International, 37*(25), 9071–9097.

Yagmur, N., Dervisoglu, A., & Bilgilioğlu, B. (2022). Assessment of rapid urbanization effects with remote sensing techniques. In M. Ben Ahmed, A. A. Boudhir, İ. R. Karaş, V. Jain, & S. Mellouli (Eds.), *Innovations in Smart Cities Applications* (pp. 571–585). https://doi.org/10.1007/978-3-030-94191-8_46

Zhao, J., He, X., Li, J., Feng, T., Ye, C., & Xiong, L. (2019). Automatic vector-based road structure mapping using multibeam LiDAR. *Remote Sensing, 11*(14), Article 14. https://doi.org/10.3390/rs11141726

Zhong, T., Zhang, Z., Chen, M., Zhang, K., Zhou, Z., Zhu, R., Wang, Y., Lü, G., & Yan, J. (2021). A city-scale estimation of rooftop solar photovoltaic potential based on deep learning. *Applied Energy, 298*, Article 117132. https://doi.org/10.1016/j.apenergy.2021.117132