Optimizing Photovoltaic Detection in High-Resolution Satellite Imagery Using GIS, DeepLabv3+, and Transformer-Based Models: A Case Study of the Marrakesh-Safi Region

Saad Farah [a*ID], Nora Nait [b], Hachem Saadaoui [b], Hatim Lechgar [a], Mehdi Maanan [b], Mohamed Maanan [c,ID], Anass Rghioui [a,ID], Hassan Rhinane [b]

[a] LaGeS-SGEO Laboratory, Hassania School of Public Works, Casablanca, Morocco, Emails: farah.saad.cedoc@ehtp.ac.ma (SF), h.lechgar@gmail.com (HL) , a.rghioui@ehtp.ac.ma (AR)
[b] Geoscience Laboratory, Hassan II University, Casablanca, Morocco, Emails: saadaoui@greenenergypark.ma (HS), smouni@greenenergypark.ma (MS), mehdi.maanan@gmail.com (MeM), h.rhinane@gmail.com (HR)
[c] UMR 6554 CNRS LETG-Nantes Laboratory, Institute of Geography and Planning, Nantes University,44312 Nantes, France, Email: Mohamed.Maanan@univ-nantes.fr (MoM)

*Corresponding author, email: farah.saad.cedoc@ehtp.ac.ma , phone number: +212689479013

Keywords: Photovoltaic (PV) Systems, Semantic Segmentation, Advanced deep learning, Satellite Imagery, Marrakesh-Safi Region.

Abstract

Solar energy has become a major contributor to global renewable energy strategies, offering a sustainable alternative to fossil fuels. Photovoltaic (PV) systems, which convert sunlight into electricity, play a central role in this transition. As the demand for large-scale solar energy projects grows, Geographic Information Systems (GIS) and advanced deep learning models have become critical for accurately detecting and mapping PV installations, particularly from satellite imagery. However, challenges remain, especially in regions with suboptimal satellite data quality. This study focuses on the Marrakesh-Safi region of Morocco, where the potential for solar energy is high but hindered by limitations in available satellite imagery. We employ advanced transformer-based models, including **Mask2Former**, **SegFormer**, and **DeepLabV3+**, to enhance the semantic segmentation of PV systems from high-resolution satellite images. By integrating GIS with these deep learning models, we aim to improve the accuracy and scalability of PV detection, even in complex and diverse geographical settings. Our methodology involves training and testing these models on annotated satellite imagery, with performance evaluated using key metrics such as Intersection over Union (IoU), precision, recall, and F1 score. Mask2Former achieved notable results with a recall of **0.95** and an F1 score of **0.936**, excelling in the detection of smaller and more complex PV layouts. DeepLabV3+ demonstrated strong overall performance, with an IoU of **0.89** and precision of **0.93**, while also being the most computationally efficient model, processing **28** samples per second. This research highlights the effectiveness of integrating GIS with deep learning, particularly transformer-based architectures, for the accurate detection and mapping of PV systems. The results contribute to the broader efforts in renewable energy optimization, supporting more efficient solar energy deployment, especially in regions like Morocco where data quality poses significant challenges.

1. Introduction

The transition to renewable energy sources is now more urgent than ever, driven by the need to reduce carbon emissions and combat the environmental impacts of fossil fuels (Kleebauer et al., 2023). Renewable energy, particularly solar, has become a leading solution due to advancements in photovoltaic technology and its ability to harness abundant solar resources. Accurate segmentation and detection of photovoltaic systems from aerial and satellite imagery are critical for optimizing solar energy deployment and planning (García et al., 2024). Photovoltaic systems, which directly convert sunlight into electricity, have emerged as pivotal components in global strategies to accelerate the adoption of clean energy technologies (Kleebauer et al., 2023).

Geographic Information Systems (GIS) and deep learning have become crucial tools in optimizing green energy systems. GIS provides essential spatial data for identifying optimal sites for solar panel installation by analyzing environmental factors like land use, topography, and solar irradiance (Verso et al., 2015). Deep learning enhances this process by automating the detection and classification of land areas suitable for solar energy through advanced image recognition, thus improving the precision and scalability of site assessments (Zhu et al., 2023). Together, GIS and deep learning form a powerful synergy that accelerates the integration of solar energy, reducing time and cost for large-scale PV deployment across different regions.

In Morocco, the high potential for solar energy is often hindered by significant challenges related to the quality and resolution of satellite imagery available for photovoltaic (PV) detection. Despite its vast solar irradiance, making it an ideal candidate for large-scale solar energy projects, the limitations of the available satellite imagery create difficulties in accurately identifying and segmenting PV installations. This issue is particularly pronounced for smaller systems or those that are integrated into complex urban and rural landscapes, where shadows, building structures, and vegetation can obscure solar panels, leading to misclassification or omission of PV systems (Zefri et al., 2018). Moreover, the country's reliance on older, low-resolution satellite data complicates large-scale PV mapping efforts, as this imagery lacks the necessary detail to distinguish between PV panels and other reflective surfaces, such as metal rooftops or light-colored terrain.

These challenges can, however, be addressed through the application of advanced GIS and deep learning techniques, which are capable of extracting more accurate and detailed information from satellite images, even with varying spatial resolutions. GIS offers a robust platform for integrating various data layers, such as elevation, land use, and solar irradiance, which are critical for optimizing the siting of PV systems. When combined with deep learning, particularly convolutional neural networks (CNNs) and transformers, these systems can significantly enhance image interpretation by automatically detecting and classifying objects within satellite imagery. Deep learning models trained on large datasets can learn to differentiate between solar panels and other objects, even under challenging conditions (J. Wang et al., 2022).

Specifically, semantic segmentation techniques, which assign a class label to each pixel in an image, are particularly effective in compensating for the low quality of input data. By using these techniques, researchers can create detailed maps that highlight the presence of PV systems across wide geographical areas, even when the original satellite data is of low resolution. Furthermore, advanced methods such as multi-scale image analysis can enhance the resolution of certain regions in the imagery, providing more accurate and reliable information for PV detection. These technologies not only improve the precision of solar panel identification but also support more efficient and strategic deployment of solar energy infrastructure in Morocco, where the potential for solar energy remains largely untapped due to current technical limitations (Verso et al., 2015).

Recent advances in deep learning, particularly in computer vision, have further enhanced the accuracy of PV detection from satellite imagery. Traditionally, convolutional neural networks (CNNs) were used for tasks like segmentation, but they are limited by their inability to capture long-range dependencies and global context effectively (Dosovitskiy et al., 2021). Transformer-based models have proven to be more effective than traditional models by using attention mechanisms to capture global context across entire images. This makes transformers particularly suitable for complex tasks like photovoltaic (PV) panel segmentation, where subtle differences between solar panels and surrounding surfaces must be accurately detected. The attention mechanism allows these models to focus on relevant parts of the image, significantly enhancing segmentation precision and overall performance (Cheng et al., 2022; Dosovitskiy et al., 2021).

Among the most promising transformer-based models are Mask2Former, Segformer, and DeepLabv3+, which have shown superior performance in semantic segmentation tasks. Mask2Former excels in its ability to generalize across various segmentation tasks, while Segformer combines the strengths of transformers with high accuracy in dense prediction tasks (Cheng et al., 2022; Xie et al., 2021). DeepLabv3+, although originally based on CNNs, has incorporated attention mechanisms to improve its segmentation capabilities (Chen et al., 2018). By upgrading from CNNs to transformer-based models, our research aims to improve the accuracy and robustness of PV detection from satellite images, addressing the inherent challenges of low-quality imagery in regions like Morocco.

So, integrating GIS, deep learning, and transformer-based models offers a robust framework for improving the identification and mapping of photovoltaic systems from satellite images, even in regions where data quality is suboptimal (Verso et al., 2015; J. Wang et al., 2022). This study pushes the boundaries of current methodologies by leveraging advanced transformer architectures to enhance semantic segmentation for PV panel detection, contributing to the broader field of green energy optimization (Cheng et al., 2022; Dosovitskiy et al., 2021; Xie et al., 2021). By addressing the challenges of low-resolution satellite imagery, especially in regions like Morocco, this approach improves the accuracy and scalability of PV detection, paving the way for more efficient and sustainable solar energy deployment worldwide (Zefri et al., 2018).

2. Related works

The detection and mapping of photovoltaic (PV) systems using satellite imagery has garnered significant attention in recent years, particularly with the rise of high-resolution satellite data and advanced deep-learning techniques. In this section, we review the key studies and methodologies that have influenced the field, focusing on Geographic Information Systems (GIS), deep learning, semantic segmentation, and the role of convolutional neural networks (CNNs) and transformer-based models in solar panel detection.

2.1. GIS and Remote Sensing for Renewable Energy

Geographic Information Systems (GIS) and remote sensing have long played a central role in the planning and deployment of renewable energy systems. Researchers such as (Sørensen, 2001) demonstrated the use of GIS for optimizing the siting of solar panels by integrating spatial data on solar irradiance, land use, and environmental constraints. By providing detailed spatial data, GIS allows for efficient solar energy planning, especially in large-scale projects where geographic factors heavily influence energy potential. Moreover, (Gassar & Cha, 2021) expanded on the integration of GIS with remote sensing for regional assessments of solar potential, indicating that combining spatial data with machine learning could further improve decision-making processes.

2.2. Deep Learning in PV Detection

Deep learning has become an essential tool for automating the detection and classification of photovoltaic systems from satellite imagery. Convolutional neural networks (CNNs) have been widely used for object detection tasks due to their powerful feature extraction capabilities. For instance, (Ioannou & Myronidis, 2021) applied CNNs to automatically detect solar panels from satellite images, demonstrating high levels of accuracy. Similarly, (Vlaminck et al., 2022) used CNN-based models to analyze satellite images and segment regions containing PV systems. These early studies showed the effectiveness of CNNs in handling high-resolution images and in classifying various land cover types, including PV installations.

2.3. Semantic Segmentation for Photovoltaic Systems

Semantic segmentation techniques have further advanced the precision of PV detection, allowing for pixel-level classification of satellite imagery. One of the most widely used models in this domain is DeepLabv3+, which has been applied to a range of remote sensing tasks, including solar panel detection. (Chen et al., 2018) introduced DeepLabv3+, enhancing the feature extraction process through atrous convolution, which captures fine-grained details essential for segmenting objects like solar panels. The ability of DeepLabv3+ to detect small and dispersed solar panels in complex landscapes has been confirmed by studies such as (Costa et al., 2021), which demonstrated its effectiveness in various urban and rural settings. Moreover, by incorporating multi-scale image analysis, DeepLabv3+ can handle high-resolution satellite images, enabling more accurate mapping of PV installations.

2.4. Transformer-Based Models in Remote Sensing

While CNNs have traditionally dominated image segmentation tasks, transformer-based models have recently emerged as powerful alternatives due to their ability to capture long-range dependencies and global context. (Dosovitskiy et al., 2021) introduced Vision Transformers (ViTs), which apply self-attention mechanisms across the entire image, significantly improving the accuracy of segmentation tasks compared to CNNs. In the context of PV detection, transformers allow models to focus on specific features of solar panels across various spatial

resolutions, making them highly effective for large-scale satellite image analysis.

Several transformer-based architectures, such as Mask2Former and Segformer, have been proposed for semantic segmentation. (Cheng et al., 2022) introduced Mask2Former, which generalizes across various segmentation tasks, offering high flexibility and performance. Meanwhile, (Xie et al., 2021) presented Segformer, a hybrid model that combines the simplicity of transformers with the robustness of dense prediction tasks, making it well-suited for complex satellite imagery analysis. These models provide state-of-the-art performance in solar panel detection, particularly when dealing with high-resolution images where subtle differences between solar panels and other reflective surfaces need to be captured.

2.5. Hybrid Approaches Combining CNNs and Transformers

Recent research has explored hybrid approaches that integrate both CNNs and transformers to take advantage of the strengths of both architectures. For instance, (Ziyabari et al., 2022) They proposed a hybrid method for solar irradiance forecasting, leveraging CNNs for local feature extraction and transformers to capture long-term dependencies, which can be adapted for solar panel detection tasks. (Liang et al., 2023) developed a hybrid CNN-transformer model for remote sensing, where CNNs handle feature extraction and transformers capture the global context, improving segmentation performance in complex urban environments. Similarly, (L. Wang et al., 2021) demonstrated that combining CNNs and transformers helps improve the segmentation of urban scenes with varying resolutions, an approach applicable to satellite imagery used for PV detection. Hybrid models like these are increasingly being employed to manage varying spatial scales in satellite images, enhancing the accuracy of photovoltaic panel detection in diverse and challenging environments.

2.6. Photovoltaic Detection in the Moroccan Context

In the specific context of Morocco, studies have highlighted the unique challenges associated with using satellite imagery for PV detection. (Zefri et al., 2018) emphasized the need for high-resolution UAV-based thermal and visual data to overcome the limitations of traditional satellite imagery. This method provides more accurate inspections, especially in environments like Morocco, where older satellite data may lack the resolution necessary to detect smaller, distributed solar panels. The integration of UAV technology, Geographic Information Systems (GIS), and advanced image processing techniques, such as deep learning models including CNNs and transformers, has been identified as a promising solution for overcoming these challenges. These approaches enable more detailed and precise mapping and detection of PV installations, particularly in regions with suboptimal data quality, further improving the accuracy of photovoltaic detection.

3. Study Area: Marrakesh-Safi Region

3.1. General Characteristics of the Region

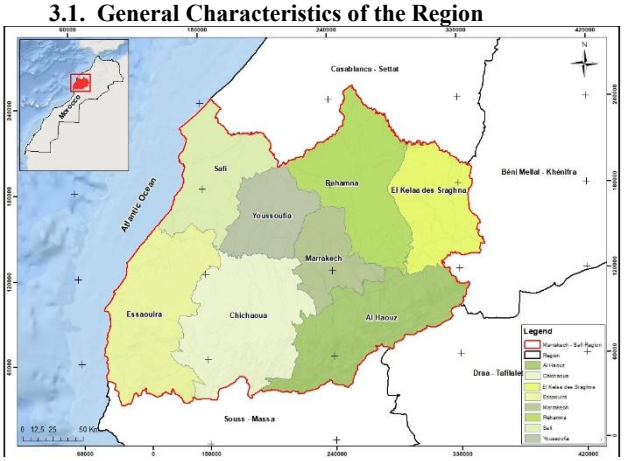

Figure 1. Study area. Marrakesh-Safi Region

The Marrakesh-Safi region, situated in central Morocco, is renowned for its high solar irradiance, making it an ideal area for solar energy projects. Spanning approximately 41.404 km² (Étude Sur Le Tissu Entrepreneurial Région de Marrakech-Safi, 2024), the region is diverse, featuring plains, coastal areas, and mountainous terrains that present both opportunities and challenges for photovoltaic (PV) deployment (HCP, 2014). Its semi-arid climate, with solar radiation levels between 5 and 6 kWh/m² per day, offers an optimal environment for year-round solar energy generation (Ouchani et al., 2021). Key urban centers such as Marrakesh are surrounded by a mixture of rural and urban landscapes, providing varied sites for both large-scale solar farms and distributed PV systems. However, the region's topographical diversity, particularly in the High Atlas Mountains, complicates PV deployment due to infrastructure limitations and challenging terrain (Ouchani et al., 2021).

Within this region, the city of Ben Guerir stands out as a significant player in Morocco's solar energy landscape. Located roughly 70 km north of Marrakesh, Ben Guerir is home to the Mohammed VI Green City and the Green Energy Park, one of Morocco's most advanced research centers for renewable energy. The Green Energy Park, part of the Mohammed VI Polytechnic University, plays a crucial role in research and development related to solar energy technologies, including the testing and deployment of PV systems (Oufadel et al., 2022). This city has become a focal point for Morocco's efforts to develop a sustainable energy sector and a center for innovation in solar energy, contributing to the national goal of increasing the share of renewable energy to 52% by 2030 (Moroccan Ministry of Energy, 2019).

Ben Guerir's flat terrain and expansive open spaces make it an ideal location for large-scale PV projects. However, as in the broader Marrakesh-Safi region, challenges remain in accurately detecting and mapping photovoltaic installations from satellite imagery. The combination of urban and rural landscapes, along with varying spatial resolutions in available satellite data, complicates the accurate identification of PV systems. (Faouzi et al., 2023) addressed these challenges using high-resolution satellite imagery and advanced deep-learning models, comparing different photovoltaic systems across multiple Moroccan sites. Their research demonstrates the use of advanced tools like PVsyst to enhance accuracy in detecting and assessing PV installations.

3.2. Solar Irradiance Patterns in the Study Region

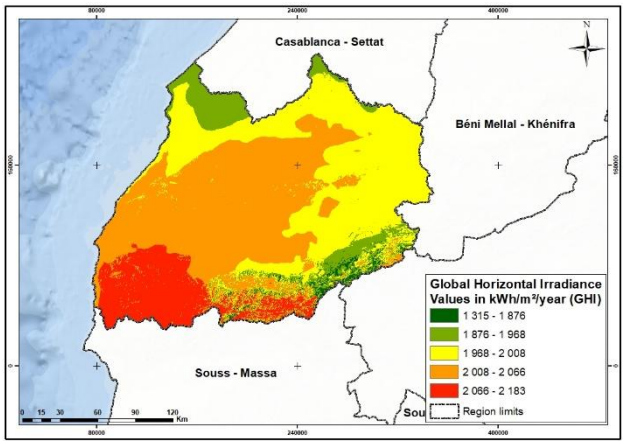

Figure 2. Solar Irradiance Map (GHI)

The Solar Irradiance Map (GHI) illustrates the distribution of solar energy across a 41.404 km² region using Global Horizontal Irradiance (GHI) data sourced from the Global Solar Atlas. Derived from long-term averages spanning 1994 to 2018, the map highlights the total solar radiation received on a horizontal surface, expressed in kWh/m²/year. GHI values in the region range from 769.95 to 2789.41 kWh/m²/year, indicating significant spatial variation influenced by factors such as geography, altitude, and atmospheric conditions. This variation is critical for identifying zones with high solar potential. The primary objective of the map is to support the planning and implementation of renewable energy projects, especially solar-powered irrigation systems, by providing a visual representation of solar energy availability. As such, it serves as a key decision-making tool for policymakers, engineers, and agricultural planners aiming to promote sustainable and efficient solar energy utilization in the region.

The Marrakech-Safi region was chosen as the study area due to its high solar energy potential, geographic diversity, and national relevance in Morocco's renewable energy strategy. These characteristics make it an ideal environment for testing and optimizing advanced photovoltaic detection techniques using high-resolution satellite imagery, GIS, and deep learning models such as DeepLabv3+ and Transformer-based architectures.

4. Methodology

In this workflow, we present a methodology for the semantic segmentation of photovoltaic (PV) panels from satellite images using advanced deep-learning models.

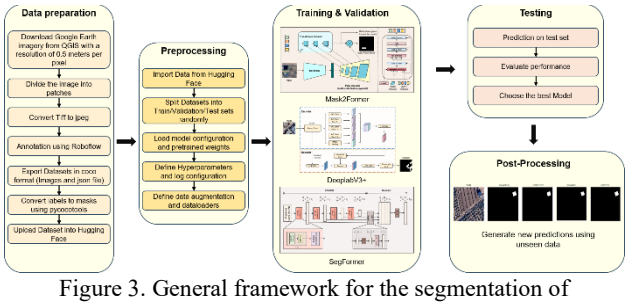

Figure 3. General framework for the segmentation of photovoltaic panels

4.1. Data preparation

The data preparation phase starts with downloading high-resolution satellite imagery from Google Earth using QGIS, with a resolution set to 0.5 meters per pixel. These large images are then divided into smaller patches to make them more manageable for further processing. The image format is converted from TIFF, commonly used in geographic data, to JPEG, a compressed format more compatible with many machine learning workflows. After conversion, the images undergo annotation using a tool like Roboflow, where specific objects of interest (such as solar panels) are labeled. These labeled datasets are then exported in the COCO format, which includes both the images and their corresponding annotation files in JSON format, making them suitable for supervised learning tasks. Following this, labels are converted into segmentation masks using pycocotools, enabling the model to understand where objects are located in the image. Finally, the dataset is uploaded to the Hugging Face platform for model training and validation.

4.2. Preprocessing

Once the dataset is uploaded, the preprocessing phase begins with importing the data from Hugging Face. The dataset is split randomly into training, validation, and test sets to ensure proper model training and evaluation. The next step involves loading model configurations and pre-trained weights, leveraging transfer learning to reduce training time and improve performance by starting from a model already trained on a similar task. Hyperparameters such as learning rate, batch size, and number of epochs are defined, alongside the logging configuration to track the training progress. To enhance the robustness of the model, data augmentation techniques, such as rotating and flipping the images, are applied, increasing the variety of training examples. Dataloaders are also configured to efficiently feed the data into the model during the training process.

4.3. Training &Validation

The training and validation phase leverages three advanced model architectures: Mask2Former, DeepLabV3+, and SegFormer.

4.3.1. Mask2Former:

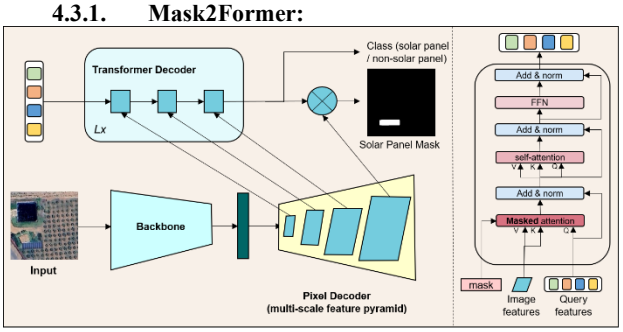

Figure 4. Illustration of Mask2Former Architecture

Firstly, we introduce the Masked-attention Mask Transformer (Mask2Former), which serves as the first model for comparison. The Mask2Former architecture, a variation of the Mask Transformer, is designed for segmentation tasks such as identifying solar panels in satellite images. The process begins with the image being passed through a backbone network, such as ResNet or a Vision Transformer (ViT), which extracts hierarchical feature maps at different scales (Dosovitskiy et al., 2021). These features are then processed through a pixel decoder, which aggregates multi-scale information to create a feature pyramid, capturing both small and large object features in the image (Cheng et al., 2022). The core of the architecture is the Transformer Decoder, which utilizes self-attention and cross-attention mechanisms to compute relationships between different parts of the image and between learned query features. A critical component of this architecture is masked attention, which enables the model to focus selectively on relevant image regions

by applying masks to ignore irrelevant parts. This mechanism significantly improves the segmentation process, as it helps to localize objects more precisely and efficiently by guiding the model to attend only to the regions containing potential objects of interest (Cheng et al., 2022). The decoder contains multiple layers that include normalization, feed-forward networks, and masked attention, allowing the model to progressively refine segmentation results. The final output is a set of object masks that classify pixels into categories such as solar panel and non-solar panel regions, producing a detailed solar panel mask. Optimized through a loss function that refines class distinctions, this architecture excels in handling multi-scale features and using attention mechanisms to deliver precise segmentation outputs, particularly for challenging tasks like solar panel identification in remote sensing imagery.

In this study, we utilized three different backbones Swin Tiny, Swin Small, and Swin Base within the Mask2Former model to evaluate their impact on the semantic segmentation of photovoltaic panels. These Swin Transformer-based backbones provide varying levels of complexity, with Swin Tiny offering a lightweight configuration, Swin Small delivering a balance between performance and efficiency, and Swin Base incorporating a larger model capacity for enhanced feature extraction.

4.3.2. DeepLabV3+:

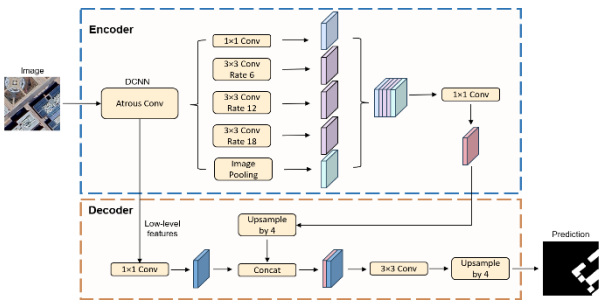

Figure 5. Illustration of DeepLabv3+ Architecture

The second model for comparison used in this study is the DeeplabV3+, The DeepLabv3+ architecture is a highly effective model for semantic segmentation that leverages the strengths of atrous convolutions and an encoder-decoder structure. In the encoder phase, it uses a deep convolutional neural network (DCNN) backbone, often Xception, to extract multi-scale features from the input image. Atrous (dilated) convolutions, applied at different rates (6, 12, 18), allow the model to capture contextual information at multiple scales without reducing the resolution, thus maintaining the spatial precision of the feature maps (Chen et al., 2018). In addition, global image pooling is performed to capture the global context of the entire scene, ensuring that the model has a comprehensive understanding of larger structures. After aggregating these multi-scale features, a 1x1 convolution is used to fuse them into a compact representation. The decoder then integrates the fine details from earlier layers of the network with the high-level semantic features from the encoder by concatenating them. A 3x3 convolution is applied to refine the merged features, followed by an upsampling operation to restore the output to the original resolution. This combination of multi-scale feature extraction and refinement allows DeepLabv3+ to generate accurate, high-resolution segmentation maps with well-defined object boundaries, making it highly effective for various semantic segmentation tasks (Chen et al., 2018).

In our experiments with DeepLabV3+, we employed the Se ResNeXt50 32x4d backbone, a powerful architecture known for its ability to capture rich feature representations through its use of group convolutions and channel-wise attention. This backbone is particularly effective in handling complex image segmentation tasks due to its enhanced capacity for extracting spatial and contextual information. By integrating Se ResNeXt50 32x4d with DeepLabV3+, we leveraged its strengths in segmenting photovoltaic panels from satellite imagery, ensuring high accuracy and efficient processing, especially in detecting larger and well-defined objects.

4.3.3. SegFormer:

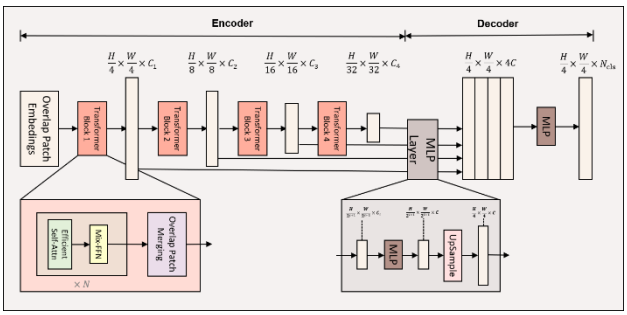

Figure 6. Illustration of SegFormer Architecture

Lastly, we incorporate the SegFormer as the third model used in this study. The SegFormer architecture is a transformer-based model designed for semantic segmentation that achieves efficiency and high performance by simplifying the segmentation process through a lightweight yet powerful structure. The model begins by dividing the input image into overlapping patches, which helps retain important boundary information. These patches are projected into feature embeddings using convolutional layers and are passed through several Transformer Blocks that operate at different resolutions. Each block contains an Efficient Self-Attention mechanism, which captures long-range dependencies, and a Mix-Feedforward Network (Mix-FFN) that combines depth-wise separable convolutions with fully connected layers to efficiently capture both local and global features (Xie et al., 2021). As the resolution decreases in each transformer block, Overlap Patch Merging occurs, preserving important features while maintaining computational efficiency. The model's decoder consists of simple MLP layers that aggregate and upsample the multi-scale feature maps generated by the encoder, restoring the spatial resolution of the output to match the input. This structure enables SegFormer to generate accurate segmentation masks while avoiding the need for complex decoder mechanisms seen in other models. By integrating global and local contexts using transformers and efficient design principles, SegFormer achieves state-of-the-art performance in semantic segmentation tasks with minimal computational overhead, making it ideal for real-time applications (Xie et al., 2021).

In this study, we employed the SegFormer model with three different MiT (Mix Transformer) backbones MiT-B0, MiT-B1, and MiT-B2 to assess their performance in the semantic segmentation of photovoltaic panels. These backbones offer varying levels of complexity, with MiT-B0 being the smallest and most lightweight, while MiT-B1 and MiT-B2 progressively increase in model capacity and feature extraction capabilities. The MiT architecture is designed to efficiently balance performance and computational cost, making it suitable for a wide range of segmentation tasks. By testing SegFormer with these backbones, we aimed to evaluate how different configurations impact both the accuracy and ability of the model to generalize across unseen satellite imagery.

4.4. Testing

After training, the models are tested using a separate test set to evaluate their performance on unseen data. The models make predictions based on the test images, and their outputs are compared to the ground truth labels. Performance metrics (Table 1), such as Accuracy, Precision, Recall, F1 score, and Intersection over Union (IoU), are used to determine how well each model performs. Based on these evaluations, the best model is selected for future use, ensuring that only the highest-performing model proceeds to the next phase. This testing phase is crucial as it demonstrates the model's ability to generalize to new, unseen examples, which is essential for real-world applications.

Accuracy assessment	Evaluation Metrics
Precision	$\frac{True\ Positive}{True\ Positive + False\ Positive}$
Recall	$\frac{True\ Positive}{True\ Positive + False\ Negative}$
F1 score	$\frac{2 \times True\ Positive}{True\ Positive + 0.5 \times (False\ Positive + False\ Negative)}$
Intesection over Union (IoU)	$\frac{Pixel\ Overlap}{Pixel\ Union}$

Table 1. Evaluation of Models Performance Using Various Metrics

4.4.1 Precision: Precision assesses the model's ability to make accurate positive predictions. It quantifies the ratio of true positive predictions to the total positive predictions made by the model.

4.4.2 Recall: Recall, also known as sensitivity or true positive rate, evaluates the model's capacity to identify all relevant instances. It calculates the ratio of true positive predictions to the total actual positive instances in the dataset.

4.4.3 F1-Score: The F1-Score is a balanced metric that considers both precision and recall. It offers a harmonized measure of the model's accuracy in capturing positive instances while minimizing false positives and false negatives.

4.4.4 Intersection-over-Union (IoU): Notable, IoU holds particular importance in evaluating the performance of semantic segmentation models due to its ability to address the class imbalance issue. IoU quantifies the similarity between the predicted area and the corresponding ground-truth region for an object.

4.5. Post-Processing

In the post-processing phase, the best-performing model is used to generate predictions on entirely new, unseen data. This phase involves applying the trained model to real-world scenarios where the model is expected to identify and segment objects of interest, such as solar panels, from satellite imagery. The results are visualized by comparing the model's predictions with the original imagery, showing how well the model generalizes beyond the test data. This final phase is critical for deploying the model in real-world applications, ensuring that it can be reliably used for tasks such as monitoring solar panel installations or other environmental features in satellite images.

5. Experimental results and discussion

5.1. Comparison of the plots

5.1.1. Mask2Former:

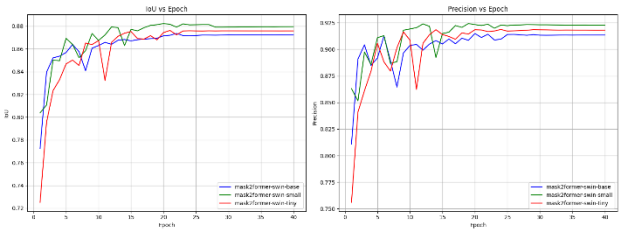

Figure 7. Comparison of IoU and Precision Metrics Across Mask2Former Variants over Epochs

5.1.1.1 IoU vs Epoch: The left plot illustrates the Intersection over Union (IoU) as a function of the number of epochs for three variations of Mask2Former, using different Swin Transformer backbones: Swin Base, Swin Small, and Swin Tiny. Initially, all models exhibit a rapid increase in IoU, particularly during the first 5 epochs, indicating that the models quickly learn to capture the relevant features for semantic segmentation. The Swin Base and Swin Small backbones show a slightly more stable improvement compared to Swin Tiny, which experiences more fluctuations, especially between epochs 5 and 15. By epoch 20, all models stabilize, with Swin Small achieving the highest final IoU of around 0.88, followed closely by Swin Base at approximately 0.87. Swin Tiny slightly lags behind but still performs robustly around 0.86.

5.1.1.2 Precision vs Epoch: The right plot shows the precision of each Mask2Former variant across 40 epochs. Similar to the IoU plot, there is a sharp increase in precision within the first 5 epochs, with all models reaching a precision greater than 0.85. The Swin Base and Swin Small models demonstrate slightly more stable precision improvements, while Swin Tiny shows minor instability during the initial epochs but converges after epoch 10. After epoch 15, all models maintain a high level of precision, with Swin Small marginally outperforming the other variants with a final precision of approximately 0.925. Swin Base and Swin Tiny achieve slightly lower but impressive precision scores around 0.92.

Overall, the results show that all three variations of Mask2Former perform well, with Swin Small slightly outperforming the others in both IoU and precision. The consistent performance after epoch 20 suggests that the models have fully converged by this point, with little improvement thereafter.

5.1.2. SegFormer:

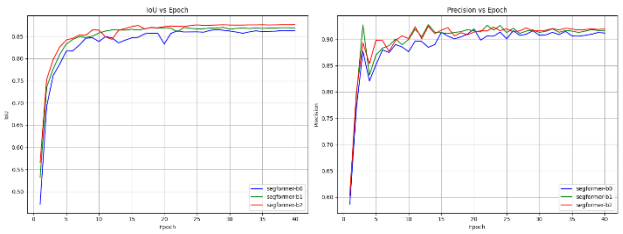

Figure 8. Performance Evaluation of SegFormer Variants using IoU and Precision Metrics Across Epochs

5.1.2.1 IoU vs Epoch: The left plot illustrates the IoU (Intersection over Union) as a function of the number of epochs for three variations of SegFormer using different MiT backbones: MiT-B0, MiT-B1, and MiT-B2. All three models show a sharp improvement in IoU during the initial 5 epochs, reaching above 0.80. After this rapid increase, the IoU continues to rise more gradually, with the models stabilizing around epoch 10. The MiT-B1 and MiT-B2 backbones perform similarly, converging

at an IoU of approximately 0.86 by epoch 40, with MiT-B1 showing a slightly higher IoU at times. The MiT-B0 backbone lags slightly behind, converging at around 0.85. Overall, all models demonstrate good convergence, with MiT-B1 and MiT-B2 providing slightly better IoU performance than MiT-B0.

5.1.2.2 Precision vs Epoch: The right plot shows the evolution of precision across 40 epochs for the three SegFormer variants. All models exhibit a significant improvement in precision within the first 5 epochs, quickly exceeding 0.90. Similar to the IoU plot, the models converge relatively early, around epoch 10, and maintain high precision throughout the remaining epochs. The MiT-B1 and MiT-B2 backbones demonstrate slightly better precision, maintaining values just above 0.91, while MiT-B0 stabilizes at around 0.90. While all models show some minor fluctuations in precision, especially during the early epochs, they generally remain highly stable after epoch 10.
The results indicate that all three SegFormer variations achieve strong and stable performance in both IoU and precision. MiT-B1 and MiT-B2 exhibit similar behavior, both outperforming MiT-B0 in terms of IoU and precision. The models converge early, within the first 10 epochs, and maintain high levels of segmentation accuracy throughout the rest of the training process, with precision exceeding 0.90 and IoU reaching 0.86 for the top-performing backbones. These results suggest that SegFormer with larger backbones (MiT-B1 and MiT-B2) provides a slight edge in performance for this task.

5.1.3. DeepLabV3+:

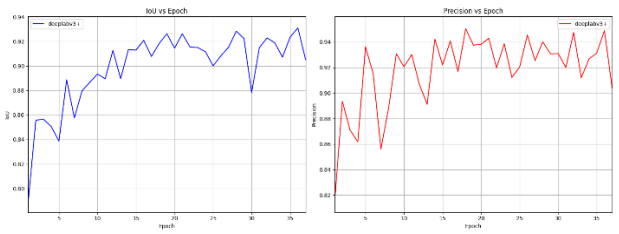

Figure 9. IoU and Precision Metrics Over Epochs for DeepLabv3+ Model Performance Evaluation

5.1.3.1 IoU vs Epoch: The left plot shows the evolution of the Intersection over Union (IoU) score over 35 epochs for the DeepLabV3+ model. The IoU begins at a relatively low value of around 0.80, rapidly improving in the first 5 epochs, where it reaches over 0.90. After this sharp increase, the IoU fluctuates slightly but continues to improve gradually throughout training, showing consistent values between 0.91 and 0.93 after epoch 10. Occasional dips in performance are observed, especially around epochs 25 and 30, but the model generally maintains high IoU values. The overall trend is positive, with the model stabilizing around an IoU of 0.92–0.93 in the later epochs, indicating that DeepLabV3+ effectively segments photovoltaic panels by the end of training.

5.1.3.2 Precision vs Epoch: The right plot illustrates the precision across 40 epochs for DeepLabV3+. Like the IoU curve, the precision initially starts at around 0.82 and quickly rises to exceed 0.90 by the 5th epoch. The plot reveals notable variability in precision between epochs 5 and 15, where it fluctuates between 0.88 and 0.94. However, after epoch 15, the precision stabilizes, oscillating between 0.92 and 0.94 with periodic dips. The fluctuations become less pronounced in the later epochs, suggesting the model achieves high precision in its predictions but exhibits some instability across epochs. The final precision consistently remains in the 0.92–0.94 range, reflecting the model's strong ability to make accurate positive identifications of photovoltaic panels in the images.
Overall, DeepLabV3+ demonstrates strong segmentation capabilities, with both IoU and precision stabilizing at high values after the initial training period. Despite some variability during the mid-epochs, the model achieves consistent performance, with IoU around 0.92–0.93 and precision in the 0.92–0.94 range, indicating reliable segmentation of the target objects. The occasional dips in both metrics suggest some sensitivity to specific batches or conditions, but the model performs well overall by the end of training.

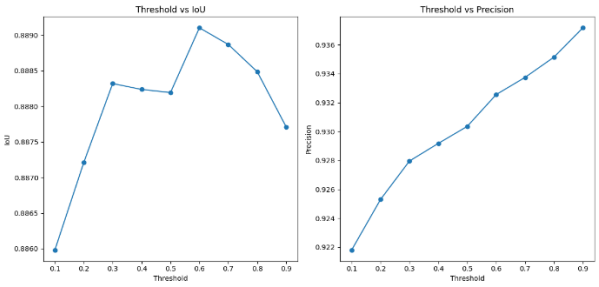

Figure 10: Effect of Threshold Values on IoU and Precision for DeepLabv3+ Model

5.1.3.3 Threshold vs IoU: The left plot depicts the relationship between the threshold value and Intersection over Union (IoU) for the DeepLabV3+ model. The IoU generally increases as the threshold is raised from 0.1 to 0.5. The model reaches its peak IoU value of approximately 0.889 when the threshold is around 0.6. After this point, the IoU starts to decrease, showing a downward trend as the threshold continues to increase up to 0.9. This indicates that, for this model, setting the threshold too high can negatively impact its performance, as the model becomes more conservative in its predictions, leading to a decrease in IoU. The optimal threshold for maximizing IoU appears to be around 0.6, suggesting a balance between identifying true positives and avoiding false positives.

5.1.3.4 Threshold vs Precision: The right plot shows the precision as a function of the threshold for DeepLabV3+. As expected, the precision increases consistently with higher thresholds, ranging from about 0.922 at a threshold of 0.1 to 0.937 at a threshold of 0.9. This trend indicates that as the threshold increases, the model becomes more confident in its positive predictions, reducing the number of false positives and, consequently, improving precision. However, this increase in precision may come at the cost of recall, as higher thresholds generally lead to fewer positive predictions overall, potentially missing some true positives.
These plots highlight the trade-off between IoU and precision with varying thresholds. While precision improves as the threshold increases, the IoU follows a more complex pattern, peaking at around a threshold of 0.6 before declining. This suggests that a threshold in the range of 0.5 – 0.6 provides a good balance between precision and IoU, ensuring strong segmentation performance while maintaining high confidence in the model's predictions.

5.1.4. Comparison of All Models:

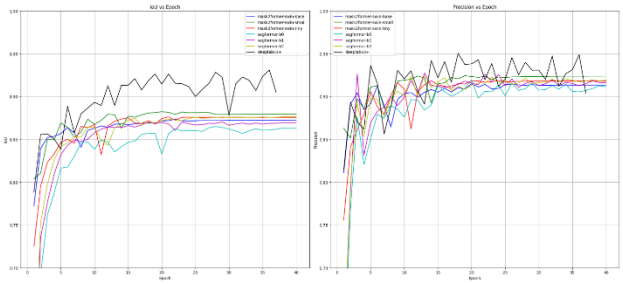

Figure 11. Comparative Analysis of IoU and Precision Metrics Across Epochs for Various Segmentation Models

5.1.4.1 IoU vs Epoch: The left plot shows the evolution of IoU (Intersection over Union) for all models (Mask2Former, SegFormer, and DeepLabV3+) across 40 epochs. DeepLabV3+ stands out, achieving the highest IoU values consistently, peaking at approximately 0.93 and exhibiting strong stability throughout the training. This model consistently outperforms the others after epoch 10, with noticeable fluctuations but a consistently higher IoU. Among the Mask2Former variants, Swin Small shows the best performance, reaching an IoU of around 0.88 by the end of the training, while Swin Base and Swin Tiny slightly lag behind with values around 0.87 and 0.86 respectively. For the SegFormer variants, MiT-B1 and MiT-B2 achieve similar performance, converging at an IoU of approximately 0.86, while MiT-B0 falls slightly behind at 0.85. Overall, DeepLabV3+ demonstrates superior IoU performance, followed by Mask2Former with the Swin Small backbone, with SegFormer and Mask2Former's smaller backbones trailing slightly.

5.1.4.2 Precision vs Epoch: The right plot illustrates the precision across 40 epochs for all models. Again, DeepLabV3+ achieves the highest precision, consistently hovering around 0.94 after epoch 10, indicating that this model is highly accurate in its predictions with few false positives. The Swin Small variant of Mask2Former also performs well, reaching precision values of approximately 0.93, closely followed by Swin Tiny and Swin Base variants, both achieving precision values around 0.92. SegFormer, with its MiT-B1 and MiT-B2 backbones, stabilizes around 0.91, while MiT-B0 reaches precision of approximately 0.90. Precision fluctuations are present in all models during the early epochs, but they stabilize around epoch 10, with DeepLabV3+ maintaining the highest precision across the training process.

DeepLabV3+ clearly stands out in terms of both IoU and precision, outperforming the other models by a significant margin. It achieves the best balance between high accuracy and robust segmentation performance, with values consistently above 0.92 for both metrics. Mask2Former, particularly with the Swin Small backbone, delivers the second-best performance, achieving competitive precision and IoU scores, while Swin Base and Swin Tiny show slightly lower but still robust results. SegFormer, although slightly behind Mask2Former, still offers strong performance, with MiT-B1 and MiT-B2 delivering similar results, slightly outperforming MiT-B0. The comparison shows a clear trade-off between the models, with DeepLabV3+ excelling in overall performance, while Mask2Former and SegFormer offer viable alternatives depending on the desired balance between computational cost and accuracy.

5.2. Comparative models

The following models were trained with 4 batches:

Model	Backbone	IoU	Precision	Recall	F1 score	Samples/Second (GPU)
Mask2Former	Swin Tiny	0.87	0.92	0.95	0.93	11
	Swin Small	0.88	0.92	**0.95**	**0.936**	9
	Swin Base	0.87	0.91	0.94	0.92	8
SegFormer	Mit-B0	0.85	0.89	0.94	0.92	11
	Mit-B1	0.86	0.92	0.93	0.92	8
	Mit-B2	0.86	0.90	0.95	0.93	5
DeeplabV3+	Se ResNext50 32x4d	**0.89**	**0.93**	0.92	0.92	28

Table 2. Performance Metrics and Processing Speed of Segmentation Models with Various Backbones

The table 2 presents the performance of the three deep learning models Mask2Former, SegFormer, and DeepLabV3+ on the semantic segmentation task for photovoltaic panels, evaluated using Intersection over Union (IoU), Precision, Recall, F1-score, and processing speed (samples per second on GPU).

Mask2Former, utilizing various Swin Transformer backbones, achieved competitive results, with the best performance observed with the Swin small backbone, attaining an IoU of 0.88, a recall of 0.95, and an F1 score of 0.936. The Swin Tiny variant also performed well, with an IoU of 0.87 and an F1 score of 0.93, demonstrating the model's capacity to generalize across different backbones. Notably, while the Swin small backbone excelled in recall, the processing speed slightly decreased as backbone complexity increased, ranging from 11 samples/second (Swin Tiny) to 8 samples/second (Swin base).

SegFormer, evaluated with the MiT-B0, MiT-B1, and MiT-B2 backbones, showed slightly lower IoU and F1 scores compared to Mask2Former. The highest IoU achieved was 0.86 with both MiT-B1 and MiT-B2, while the F1 score remained consistently at 0.92 to 0.93 across all backbones. However, SegFormer's performance on GPU throughput was notably slower, particularly for the larger MiT-B2 backbone, with a throughput of only 5 samples/second.

DeepLabV3+, using the Se resnext50 32x4d backbone, delivered the highest IoU of 0.89 and also achieved strong precision (0.93), but it had a slightly lower recall (0.92) and F1 score (0.92) compared to Mask2Former's best configuration. Despite this, DeepLabV3+ was significantly faster in processing, handling 28 samples/second, making it a more efficient model in terms of computational cost, particularly in real-time or large-scale applications.

Overall, while Mask2Former with the Swin small backbone delivered the best F1 score, DeepLabV3+ proved superior in computational efficiency. These results suggest a trade-off between segmentation performance and processing speed depending on the model and backbone configuration.

5.3. Visual results

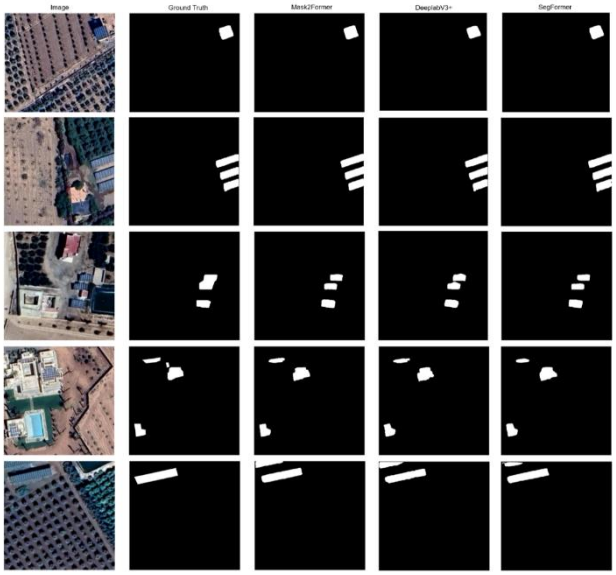

Figure 12. Comparative Visualization of Solar Panel Segmentation Using Mask2Former, DeepLabv3+, and SegFormer Models

The visual results display predictions from Mask2Former, DeepLabV3+, and SegFormer, compared to the ground truth for unseen data. Each model successfully detects solar panels in various settings. In simpler cases, such as the first row with a single solar panel and the fifth row with a large, well-defined panel in an agricultural field, all models accurately match the ground truth with minimal variation. However, in more complex scenarios, such as the third row with smaller solar panels near buildings, and the fourth row with scattered panels in a residential area, there are slight differences in performance. Mask2Former tends to provide more precise segmentation, particularly with smaller objects and complex environments. DeepLabV3+ generally performs well but sometimes exhibits slight under-segmentation or less accurate boundary detection for smaller objects. SegFormer offers robust results but occasionally misses finer details or shows slight under-segmentation in more challenging cases. Overall, all three models demonstrate strong performance, with Mask2Former excelling in precision and handling of smaller, complex objects, while DeepLabV3+ and SegFormer provide reliable results, particularly for larger, well-defined solar panels.

6. Discussion

The results presented in this study underscore the effectiveness of integrating GIS with deep learning—particularly transformer-based architectures—for enhancing the detection and semantic segmentation of photovoltaic (PV) systems from high-resolution satellite imagery. The comparative evaluation of three advanced segmentation models—Mask2Former, DeepLabV3+, and SegFormer—revealed notable trade-offs between segmentation performance and computational efficiency, which are critical considerations for real-world applications in regions such as the Marrakesh-Safi area.

Among the tested models, DeepLabV3+ demonstrated superior overall performance in terms of Intersection over Union (IoU) and precision, achieving values of 0.89 and 0.93 respectively. It also stood out as the most computationally efficient, with a processing speed of 28 samples per second, making it particularly suitable for large-scale or time-sensitive PV mapping applications. However, its recall (0.92) was slightly lower compared to Mask2Former, suggesting that while DeepLabV3+ is precise, it may occasionally miss smaller or more obscured PV instances.

Conversely, Mask2Former, especially with the Swin Small backbone, achieved the highest recall (0.95) and F1 score (0.936), indicating its strength in accurately detecting a broader range of PV system variations, including small-scale and irregularly shaped installations. This aligns with visual inspection results, where Mask2Former exhibited finer segmentation boundaries and better performance in cluttered or complex scenes such as residential rooftops or mixed urban-rural areas. The architecture's use of masked attention mechanisms and multi-scale feature decoding appears particularly advantageous in handling the inherent variability in PV appearance across satellite images. However, this accuracy comes at a computational cost, as processing speed was lower compared to DeepLabV3+, particularly for the more complex Swin Base backbone (8 samples/sec).

SegFormer offered a balanced performance, with precision and IoU values consistently above 0.90 and 0.86 respectively. The MiT-B1 and MiT-B2 variants performed comparably, suggesting diminishing returns beyond a certain backbone size. Although it lagged slightly behind Mask2Former in handling finer details, SegFormer remains a strong candidate for deployment where computational resources are limited, or where real-time inference is less critical.

Importantly, the study demonstrates that model selection should be context-driven. For example, DeepLabV3+ may be ideal for rapid mapping over large areas with well-defined installations (e.g., solar farms), while Mask2Former is better suited to detailed surveys in heterogeneous environments (e.g., urban neighborhoods or agricultural zones). In the context of the Marrakesh-Safi region, with its combination of high solar potential, urban-rural diversity, and data quality constraints, this adaptability is crucial.

Another key observation lies in the effect of thresholding strategies on model performance. As shown with DeepLabV3+, varying the threshold impacted the balance between IoU and precision, with an optimal trade-off found around 0.6. This highlights the importance of post-training calibration for maximizing real-world utility.

This work also emphasizes the importance of GIS integration, not only as a supporting tool for site selection but as a strategic component of the modeling pipeline. By embedding geospatial context into the data preprocessing and analysis stages, GIS enables more targeted deployment of deep learning models and enhances their relevance for policy-making and infrastructure planning.

Finally, while the results are promising, several challenges remain. The models' performance may vary when exposed to different satellite sources, seasonal conditions, or geographic contexts. Moreover, despite robust performance on the Marrakesh-Safi dataset, generalizability to other regions with different architectural styles or lower-quality data remains to be fully validated. Future work should consider domain adaptation strategies, multimodal data fusion (e.g., thermal imagery, LiDAR), and active learning to reduce annotation costs and improve scalability.

7. Conclusion

In this study, we explored the use of advanced deep learning models Mask2Former, DeepLabV3+, and SegFormer for the semantic segmentation of photovoltaic (PV) panels from satellite imagery. The results demonstrated that all three models are highly capable of accurately detecting solar panels, with Mask2Former excelling in more complex environments and smaller object detection, while DeepLabV3+ and SegFormer provided strong performance, particularly with larger and well-

defined panels. Through a combination of IoU and precision analysis, as well as visual inspection, we concluded that the models are robust enough to generalize well on unseen data, making them suitable for practical applications in solar energy analysis.

Looking forward, the next step for this research is to scale the methodology and apply it to satellite imagery across all of Morocco. By leveraging these segmentation models, we can accurately identify existing photovoltaic installations and calculate the potential solar energy generation for the country. This study can contribute significantly to solar energy mapping, allowing for the identification of high-potential areas for new solar projects. Moreover, this work can serve as a foundation for aiding future photovoltaic installation projects, enabling efficient planning and deployment of solar infrastructure based on data-driven insights. By integrating this analysis into broader renewable energy initiatives, we can support Morocco's transition toward sustainable energy solutions and contribute to global efforts in addressing energy challenges.

References

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*. https://github.com/tensorflow/models/tree/master/

Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., & Girdhar, R. (2022). *Masked-attention Mask Transformer for Universal Image Segmentation*. https://bowenc0221.github.io/mask2former

Costa, M. V. C. V. da, Carvalho, O. L. F. de, Orlandi, A. G., Hirata, I., Albuquerque, A. O. de, Silva, F. V. e, Guimarães, R. F., Gomes, R. A. T., & Júnior, O. A. de C. (2021). Remote Sensing for Monitoring Photovoltaic Solar Plants in Brazil Using Deep Semantic Segmentation. *Energies*, *14*(10), 2960. https://doi.org/10.3390/en14102960

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE*. https://github.com/

Étude sur le tissu entrepreneurial Région de Marrakech-Safi. (2024).

Faouzi, Y., El Fatni, O., Maftouh, A., Laarabi, B., & Barhdadi, A. (2023). Comparative Analysis of Energy Production and Simulation Software for Photovoltaic Systems in Multiple Moroccan Sites. *International Journal of Engineering Research in Africa*, *67*, 189–206. https://doi.org/10.4028/p-j5vhDU

García, G., Aparcedo, A., Nayak, G. K., Ahmed, T., Shah, M., & Li, M. (2024). Generalized deep learning model for photovoltaic module segmentation from satellite and aerial imagery. *Solar Energy*, *274*. https://doi.org/10.1016/j.solener.2024.112539

Gassar, A. A. A., & Cha, S. H. (2021). Review of geographic information systems-based rooftop solar photovoltaic potential estimation approaches at urban scales. *Applied Energy*, *291*, 116817. https://doi.org/10.1016/j.apenergy.2021.116817

Ioannou, K., & Myronidis, D. (2021). Automatic Detection of Photovoltaic Farms Using Satellite Imagery and Convolutional Neural Networks. *Sustainability*, *13*(9), 5323. https://doi.org/10.3390/su13095323

Kleebauer, M., Marz, C., Reudenbach, C., & Braun, M. (2023). Multi-Resolution Segmentation of Solar Photovoltaic Systems Using Deep Learning. *Remote Sensing*, *15*(24), 5687. https://doi.org/10.3390/rs15245687

Liang, S., Hua, Z., & Li, J. (2023). Hybrid transformer-CNN networks using superpixel segmentation for remote sensing building change detection. *International Journal of Remote Sensing*, *44*(8), 2754–2780. https://doi.org/10.1080/01431161.2023.2208711

Ouchani, F., Jbaihi, O., Alami Merrouni, A., Maaroufi, M., & Ghennioui, A. (2021). Yield analysis and economic assessment for GIS-mapping of large scale solar PV potential and integration in Morocco. *Sustainable Energy Technologies and Assessments*, *47*, 101540. https://doi.org/10.1016/j.seta.2021.101540

Oufadel, A., Ydrissi, M. EL, Berkia, Z., El Ghennioui, H., & Aicha, A. H. (2022). Optimization of the energy lost in the current limitation mode of the In grid-connected photovoltaic systems: a case study of the PV installation of the Mohammed 6 polytechnic campus of Rabat. *2022 9th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, 1–5. https://doi.org/10.1109/WINCOM55661.2022.9966418

Sørensen, B. (2001). GIS management of solar resource data. *Solar Energy Materials and Solar Cells*, *67*(1–4), 503–509. https://doi.org/10.1016/S0927-0248(00)00319-6

Verso, A., Martin, A., Amador, J., & Dominguez, J. (2015). GIS-based method to evaluate the photovoltaic potential in the urban environments: The particular case of Miraflores de la Sierra. *Solar Energy*, *117*, 236–245. https://doi.org/10.1016/j.solener.2015.04.018

Vlaminck, M., Heidbuchel, R., Philips, W., & Luong, H. (2022). Region-Based CNN for Anomaly Detection in PV Power Plants Using Aerial Imagery. *Sensors*, *22*(3), 1244. https://doi.org/10.3390/s22031244

Wang, J., Bi, L., Sun, P., Jiao, X., Ma, X., Lei, X., & Luo, Y. (2022). Deep-Learning-Based Automatic Detection of Photovoltaic Cell Defects in Electroluminescence Images. *Sensors*, *23*(1), 297. https://doi.org/10.3390/s23010297

Wang, L., Fang, S., Zhang, C., Li, R., & Duan, C. (2021). *Efficient Hybrid Transformer: Learning Global-local Context for Urban Sence Segmentation*.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). *SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers*.

Zefri, Y., ElKettani, A., Sebari, I., & Ait Lamallam, S. (2018). Thermal Infrared and Visual Inspection of Photovoltaic Installations by UAV Photogrammetry—Application Case: Morocco. *Drones*, *2*(4), 41. https://doi.org/10.3390/drones2040041

Zhu, R., Guo, D., Wong, M. S., Qian, Z., Chen, M., Yang, B., Chen, B., Zhang, H., You, L., Heo, J., & Yan, J. (2023). Deep solar PV refiner: A detail-oriented deep learning network for refined segmentation of photovoltaic areas from satellite imagery. *International Journal of Applied Earth Observation and Geoinformation*, *116*, 103134. https://doi.org/10.1016/j.jag.2022.103134

Ziyabari, S., Du, L., & Biswas, S. K. (2022). Short-term Solar Irradiance Forecasting based on Self-Attentive Transformers. *2022 IEEE Power & Energy Society General Meeting (PESGM)*, 1–5. https://doi.org/10.1109/PESGM48719.2022.9916713

Site institutionnel du Haut-Commissariat au Plan du Royaume du Maroc. (n.d.). Site Institutionnel Du Haut-Commissariat Au Plan Du Royaume Du Maroc. https://www.hcp.ma/

Rapports. (n.d.). https://www.mem.gov.ma/Pages/rapports.aspx