Learning and Nonlinear Models - Journal of the Brazilian Society on Artificial Intelligence (SBIA), Vol. 23, Iss. 2, pp. 59–74, 2025
https://doi.org/10.21528/lnlm-vol23-no2-art5

HOTSPOT DETECTION IN PHOTOVOLTAIC MODULES USING U-NET SEGMENTATION ON THERMAL INFRARED IMAGERY

Lucas Leonardo M. Carvalho
Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET/RJ
lucas.carvalho.3@aluno.cefet-rj.br

Jaqueline Neves Silva
Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET/RJ
jaqueline.neves@aluno.cefet-rj.br

Ronney A. M. Boloy
Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET/RJ
ronney.boloy@cefet-rj.br

Leandro Almeida Vasconcelos
Instituto Federal do Norte de Minas Gerais - IFNMG, Campus Montes Claros
leandro.vasconcelos@ifnmg.edu.br

Gabriel Matos Araujo
Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET/RJ
gabriel.araujo@cefet-rj.br

Milena Faria Pinto
Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET/RJ
milena.pinto@cefet-rj.br

Abstract – The reliable operation of photovoltaic (PV) systems depends on the early detection of failures such as hotspots, which are localized overheating regions typically caused by partial shading, cell mismatch, soiling, or internal defects. In these regions, affected cells operate in reverse bias and dissipate energy as heat instead of generating electricity, leading to efficiency losses, accelerated material degradation, and, in severe cases, irreversible damage or fire hazards. Conventional inspection methods, including visual evaluation and electrical testing, are time-consuming, labor-intensive, and unsuitable for large-scale solar plants. The use of Unmanned Aerial Vehicles (UAVs) equipped with thermal cameras has emerged as a promising alternative for autonomous inspection, enhancing operational efficiency and reducing human risks. In this work, we propose a hotspot detection approach based on a U-Net segmentation network applied to thermal infrared imagery of PV modules. The model was trained on a dataset that combined publicly available and laboratory-acquired images, utilizing preprocessing and augmentation strategies to enhance robustness. Experimental results demonstrate that the proposed method effectively identifies hotspot regions with precise delineation of their morphology and spatial distribution, outperforming bounding-box-based approaches such as YOLO in terms of fine-grained localization. This level of detail is crucial for predictive maintenance, as it enables the accurate measurement of hotspot size and shape. The findings highlight the potential of integrating UAV-based thermal inspection with deep learning segmentation models as a scalable and reliable solution for autonomous PV system monitoring and maintenance.

Keywords – Deep Learning, hotspot detection, infrared imagery, U-Net, photovoltaic systems.

1 Introduction

The growing demand for renewable energy sources has favored the expansion of systems such as wind turbines and photovoltaic solar panels. These technologies play a central role in the global energy transition by reducing dependence on fossil fuels and mitigating the environmental impacts associated with greenhouse gas (GHG) emissions. Specifically regarding solar energy, photovoltaic installations have shown remarkable growth. According to IRENA [1], the global installed capacity of photovoltaic solar energy reached 1.4 TW in 2023 and is expected to rise to approximately 5.4 TW by 2030, aligning with the goal of limiting global warming to 1.5 °C.

Despite their importance, photovoltaic systems are subject to failures and degradations, such as cracks, fractures, delamination, shading, and particularly hotspots. Hotspots are localized overheating regions usually caused by partial shading, dirt accumulation, manufacturing defects, or internal failures. In these areas, the affected cells behave as resistive loads, dissipating energy as heat rather than electricity, resulting in efficiency losses, accelerated material aging, and, in critical cases, irreversible

https://doi.org/10.21528/lnlm-vol23-no2-art5

module damage or fire hazards. [2] highlights that thermal cycles induced by hotspots substantially increase degradation, while [3] emphasizes the necessity of accurate detection and suppression to ensure both performance and operational safety.

Traditional inspection methods, such as visual evaluation or electrical analysis, are often time-consuming and impractical in large-scale photovoltaic plants. [4] traditionally present best practices for preventive maintenance in photovoltaic systems. They highlight that thermographic inspection, carried out with infrared cameras, is essential for detecting hotspots, which must be promptly corrected. However, this hotspot mapping is performed annually, according to the best practices recommended by the authors, and its analysis is conducted manually, which is time-consuming and subject to human error. In studies such as that of [5], even when aerial surveys with UAVs are used for data acquisition, the process still heavily depends on experts who examine high-resolution thermal images, identify anomalies such as hotspots, cracks, and delamination, and subsequently perform verification, troubleshooting, and corrective maintenance. While this approach enables detailed diagnostics, it remains time-consuming, labor-intensive, and subject to human limitations. Taken together, these studies show that, although effective, conventional methods of hotspot detection and inspection remain highly dependent on human work and specialized expertise, factors that have driven the development of automated approaches.

To overcome these limitations, detection systems based on Artificial Intelligence (AI) have been increasingly adopted. Among these approaches, Deep Neural Networks (DNNs) stand out due to their ability to learn hierarchical representations from complex data, extracting information from simple features to highly sophisticated patterns. Their architecture, composed of multiple hidden layers, enables nonlinear transformations that increase abstraction capacity, making them effective for computer vision and image analysis tasks, [6] and [7]. Recent studies also point to advances in efficiency and adaptability, such as energy-aware models and optimization-driven learning strategies, [8] and [9]. In addition to these advances in efficiency, object detection models have also achieved remarkable progress, particularly with approaches such as YOLO (You Only Look Once), which have reshaped real-time analysis in computer vision applications.

According to [10], compared to traditional image processing, deep learning methods do not require the definition of a large number of fixed thresholds in detection tasks and demonstrate greater robustness in complex scenarios. The authors note that deep learning approaches employed in detection tasks primarily encompass image classification, object detection, and semantic segmentation. While image classification methods can distinguish the categories of an entire image, the purpose of object detection is to detect and classify individual objects in real time.

Among the most widely adopted methods, YOLO stands out as a deep learning model that reformulates the detection task as a single regression from the image to bounding box coordinates and class probabilities, ensuring both speed and accuracy in practical applications. According to [11], YOLO provides a robust and efficient solution for detecting thermal anomalies in photovoltaic systems, outperforming traditional methods in automation, speed, and scalability. For the authors, this model strikes a balance between detection accuracy and inference speed, which is crucial for real-time thermal anomaly monitoring. Complementarily, [12] highlights that this methodology optimizes the detection of objects of different sizes and proportions, ensuring robust performance in varied scenarios. The strength of this methodology also lies in its ability to reach high accuracy levels in segmentation-oriented tasks.

Although YOLO is widely used and presents significant advantages for real-time applications, its main limitation is that it provides only bounding boxes for the detected objects. In situations where it is necessary to understand the exact shape of an element or anomaly, detection may be insufficient. At this point, the importance of image segmentation arises, which can be understood as an evolution in the granularity of analysis. According to [11], unlike object detection that simply provides a rectangular bounding box around an object, segmentation enables the identification of the object's exact shape, often pixel by pixel. This allows for precise contour mapping of anomalies (such as hotspots) rather than merely outlining a rectangular area that encompasses them. Furthermore, segmentation provides detailed information about the object's shape and area, enabling more in-depth analyses such as accurate defect measurement.

Within this context, image segmentation plays a fundamental role by enabling the automatic separation of regions of interest, such as defects or abnormal hotspots, from the background. By dividing images into meaningful components based on thermal intensity, contrast, or spatial patterns, segmentation facilitates the subsequent application of deep learning models for defect detection and classification, [13]. A key model in this domain is the U-Net, proposed by [14], initially developed for biomedical image segmentation. Its U-shaped architecture, composed of a contracting path (encoder) to capture context and an expansive path (decoder) to enable precise localization, introduced innovations such as skip connections, which preserve spatial details during reconstruction. Since its introduction, U-Net has become a benchmark in segmentation tasks across multiple domains. Recent adaptations, such as MC-CRSA U-Net [15], Half-UNet [16], and MEDU-Net+ [17], demonstrate its versatility and enduring relevance, particularly in applications that require the accurate and efficient analysis of complex image data, including the thermal inspection of photovoltaic modules.

In recent years, several studies have explored the potential of the U-Net in applications involving photovoltaic systems and aerial inspections with UAVs. The work of Ruan et al. [18] proposed a framework based on U2-Net, a derivative of the original U-Net, for the segmentation of photovoltaic modules from infrared images captured by drones, combined with YOLOv4 for defect detection at the module level. Similarly, [19] applied a U-Net to identify and classify hotspots in photovoltaic panels using high-resolution thermal images acquired by UAVs, enabling the analysis of faults such as cracks, delamination, and soiling. From a more geospatial perspective, [20] employed a U-Net with a ResNet50 backbone to detect small residential solar installations in aerial images collected by drones, highlighting the model's versatility for solar mapping tasks. [21], the authors developed SPF-Net, a U-Net-based architecture combined with InceptionV3, applied directly to the detection and classification of solar panel defects such as soiling, delamination, and electrical damage. Collectively, these studies highlight the relevance of the

https://doi.org/10.21528/lnlm-vol23-no2-art5

Table 1: Comparison of U-Net-based approaches for photovoltaic inspection.

Study	Application Focus	Data	Hotspot-Specific	Pixel-Level Delineation	IEC Alignment
Ruan et al. [18]	Module segmentation + defect detection	thermal	Partial	No (module-level)	No
Liu et al. [19]	Hotspot detection	thermal	Yes	Yes	No
IJGI [20]	Solar panel mapping	RGB	No	No	No
Rudro et al. [21]	Multi-defect classification	thermal	Partial	Partial	No
Proposed Work	**Hotspot morphological** segmentation	**thermal**	**Yes**	**Yes (fine-grained)**	**Yes**

U-Net as a support tool for inspecting and monitoring photovoltaic systems, particularly when integrated with aerial sensing technologies. In addition to providing accurate delineation of hotspot regions, the proposed segmentation framework can also contribute to standardized fault classification in photovoltaic inspections. The IEC TS 62446-3:2017 [22] establishes procedures for thermographic assessment of PV systems, specifying how thermal anomalies should be identified and categorized into distinct classes according to their characteristic patterns and temperature difference thresholds.

To clarify the contribution of this work in relation to existing U-Net-based photovoltaic inspection studies, Table 1 presents a structured comparison of representative approaches. The comparison highlights differences in methodological focus, segmentation granularity, and alignment with standardized thermographic inspection protocols. Although Liu et al. [19] also applied U-Net to infrared PV hotspot detection, their framework combines segmentation with HSV-based post-processing and predefined bright-spot classification mechanisms for maintenance decision support. In contrast, the present work places pixel-level segmentation at the center of the analytical framework, directly modeling hotspot morphology from thermal patterns without auxiliary color-space heuristics. Beyond detection performance, this study emphasizes geometric characterization of hotspot contours, i.e., including shape and spatial extent, and explicitly discusses their alignment with IEC TS 62446-3:2017 thermographic inspection criteria. This focus on morphology-driven, standards-oriented segmentation distinguishes the proposed framework from prior detection-oriented approaches.

1.1 Main Contributions

This work extends previous research by introducing a U-Net-based segmentation framework designed explicitly for hotspot detection in thermal infrared images of photovoltaic (PV) modules to facilitate the autonomous monitoring and surveillance of large-scale photovoltaic fields by utilizing UAVs. Earlier studies have demonstrated the use of U-Net variants or hybrid models for PV inspection. Most approaches either focus on module-level segmentation or rely on bounding-box detectors, such as YOLO, which provide only coarse localization of anomalies. In contrast, our approach leverages pixel-level segmentation to delineate the shape, size, and spatial distribution of hotspots, allowing for a more detailed analysis than bounding-box-based methods. Moreover, by integrating UAV-based data acquisition with a lightweight U-Net architecture, our framework addresses the need for scalable, real-time inspection of large-scale PV installations. The main contributions of this work can be summarized as follows:

- A segmentation-driven detection framework that uses a U-Net-based computer vision system for pixel-level localization of hotspots.
- Capability to estimate not only the presence but also the size, shape, and distribution of hotspots, which is critical for predictive maintenance and fault classification.

1.2 Organization

This paper is organized into three main sections. Section 2 provides a detailed overview of the proposed methodology, including the operation of the U-Net, its architecture, dataset processing techniques, and training configurations, as well as the setup used for data processing and network training. Finally, Section 4 concludes the work, summarizing the key findings and outlining potential directions for future research.

2 Methodology

The proposed methodology consists of: (i) image acquisition via Unmanned Aerial Vehicle (UAV); (ii) data preprocessing (radiometric normalization, resizing, panel-region extraction, and augmentation); (iii) U-Net segmentation in the obtained data; (iv) hotspot detection via segmented mask. This process is illustrated in Figure 1.

https://doi.org/10.21528/lnlm-vol23-no2-art5

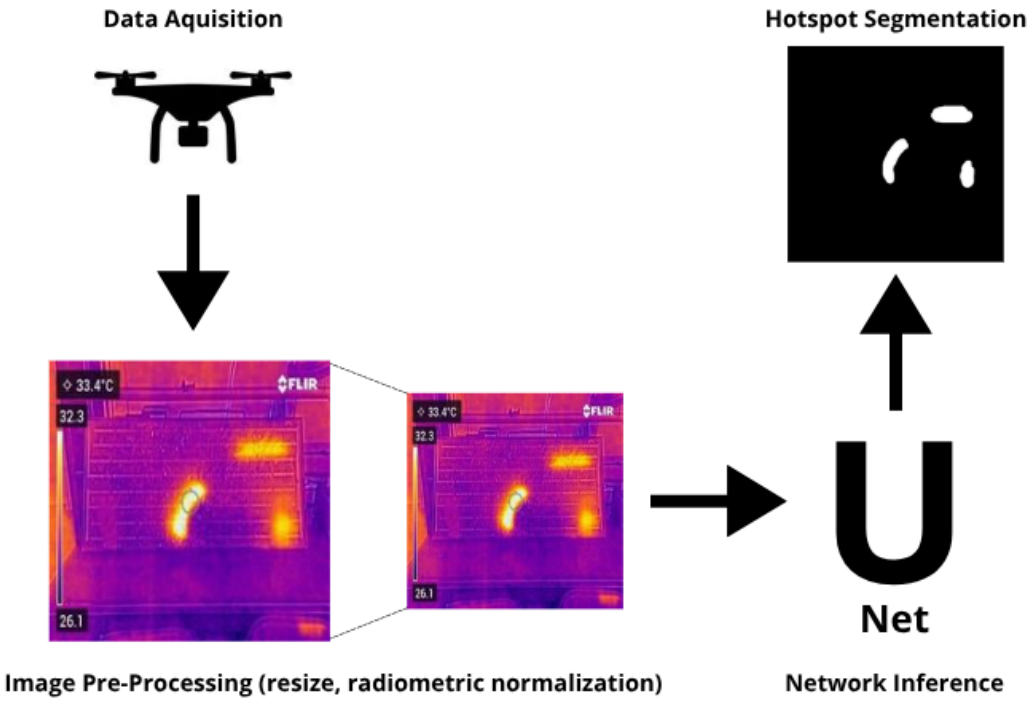


Figure 1: Proposed methodology architecture.

2.1 U-Net Architecture

Segmentation enables precise localization and per-instance shape and size estimation of hotspots, overcoming the limitations of bounding-box detectors (e.g., overlapping hotspots on adjacent cells, partial occlusions, and the need for fine-grained contours). The U-Net architecture, initially proposed for biomedical image segmentation, consists of a contracting path (encoder) and an expansive path (decoder). The encoder progressively reduces the spatial resolution of the input while increasing the number of feature maps, thereby capturing high-level contextual information [23]. In this work, ResNet-34 was selected as the encoder backbone for the U-Net. This choice was made because its depth provides an excellent tradeoff between performance and computational cost for low-end hardware, making it a classical choice for tasks such as segmentation. It also eliminates the need for manual configuration of the network's hidden layers, neurons, and activation functions, which can save a significant amount of time in future work. Additionally, the selected ResNet-34 was pre-trained on ImageNet, which initializes the network with pre-set weights and biases, enabling the use of transfer learning. This allows the model to leverage basic visual knowledge about colors, shapes, textures, and other patterns, avoiding the need to learn every characteristic from scratch.

Figure 2 provides an overview of the U-Net model. Each stage of the encoder progressively reduces the spatial resolution while increasing the feature depth. This enables the extraction of high-level contextual representations. The bottleneck serves as the transition between the encoder and decoder, capturing compact latent features. In the decoder, successive upsampling layers reconstruct the spatial resolution, with skip connections from the encoder ensuring the preservation of fine-grained spatial information. Finally, a 1×1 convolution layer produces the binary segmentation mask, distinguishing hotspot regions from the background.

This design allows the model to remain lightweight while retaining the ability to capture both global context and local thermal anomalies, making it suitable for on-board or near-edge inference in UAV-based inspections. The main hyperparameters adopted for training the U-Net are summarized in Table 2.

The hyperparameters were selected based on preliminary experiments and established practices for U-Net architectures with pretrained encoders. An exploratory analysis was conducted by varying the learning rate (10^{-3}, 10^{-4}, and 10^{-5}), loss functions (Binary Cross-Entropy and Dice Loss), and batch sizes (8, 16, and 32). Learning rates higher than 10^{-4} led to instability in validation loss, while lower values did not yield significant performance improvements. Dice Loss demonstrated superior balance between precision and recall compared to Binary Cross-Entropy, particularly due to the inherent class imbalance between background and hotspot pixels. A batch size of 16 provided the best trade-off between convergence stability and GPU memory constraints. Besides, early stopping, monitored by the F1-score, was adopted to automatically select the best generalization point, thereby preventing overfitting.

2.2 Dataset

The dataset used in this study combines images from the publicly available dataset of the American University of Sharjah [24] with images collected in the Control and Automation Laboratory (Laboratório de Controle e Automação - LACEA) from the

https://doi.org/10.21528/lnlm-vol23-no2-art5

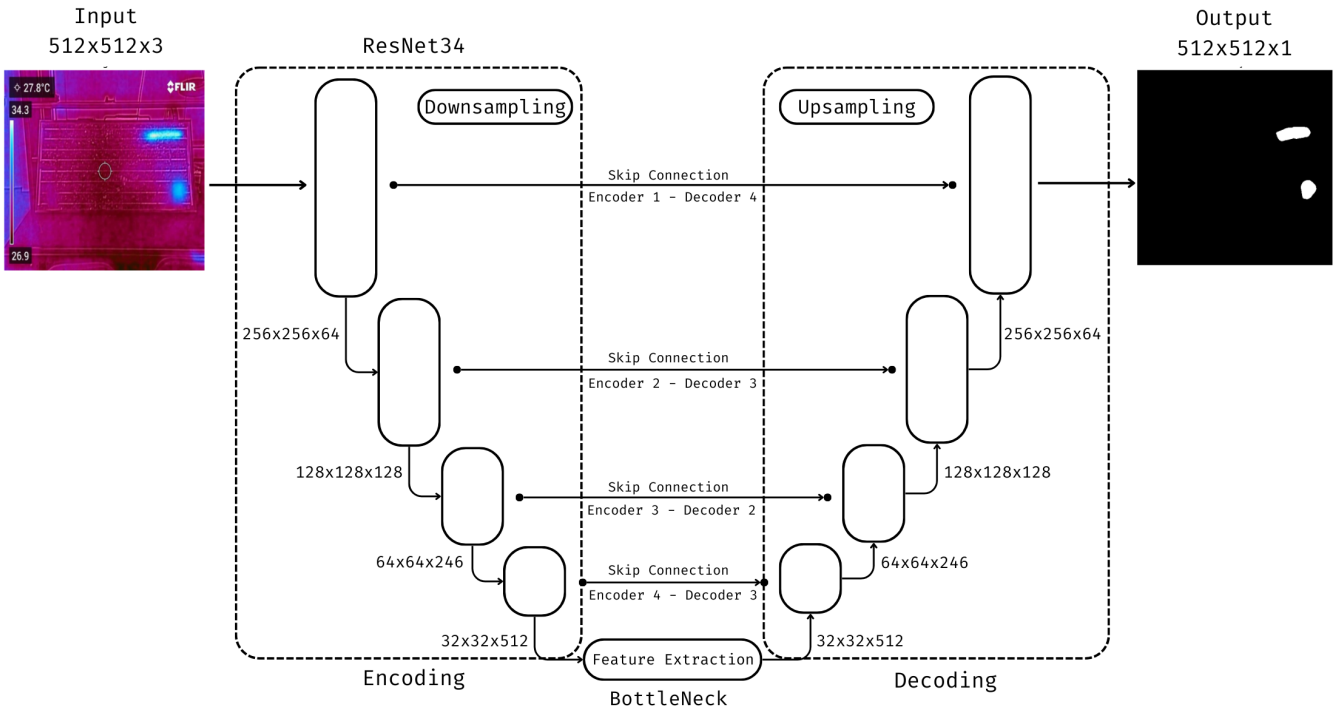


Figure 2: U-Net architecture overview.

Table 2: Hyperparameters used in the U-Net model.

Hyperparameter	**Value**
Encoder backbone	ResNet34 (pre-trained on ImageNet)
Input channels	3 (RGB thermal image representation)
Output classes	1 (binary segmentation: hotspot/background)
Loss function	Dice Loss (binary mode)
Optimizer	Adam
Learning rate	1×10^{-4}
Batch size	16
Training epochs	Max 999 (early stopping applied)
Early stopping patience	10 epochs
Device	NVIDIA Tesla T4 GPU (15GB RAM)

Federal Center of Technological Education Celso Suckow da Fonseca (CEFET/RJ). These collected images were obtained using a FLIR C5 thermal camera, capturing photos under different lighting conditions and angles, and they are available in the following link [25]. The base dataset consists of 500 thermal images of solar panels with identified hotspots [11]. Figure 3 illustrates representative samples from the dataset used in this study.

This dataset was originally made to work with YOLO, so the present labels were not in the format to be used in the U-Net. Then, to avoid the need to annotate the entire dataset, a pseudo-labeling technique was implemented. This semi-supervised method involved training an initial network with a small amount of manually labeled data and then utilizing its predictions to retrain the model or train another network, combining the predictions with the data used in the first training to create a better model that incorporates both datasets. The entire labeling process, illustrated in Figure 4, was executed as follows: i) a manually labeled dataset is created with a portion of the original dataset image, ii) the images and masks in this dataset are pre-processed (resized to 512x512) to fit in the network requirements, iii) the network is prepared, being configured its parameters, weights and bias, iv) the network is trained using the current data, v) the model is evaluated based on its training performance being analyzed metrics such as precision, recall and intersection over union (IoU), vi) predictions are extracted from unlabeled data,

https://doi.org/10.21528/lnlm-vol23-no2-art5

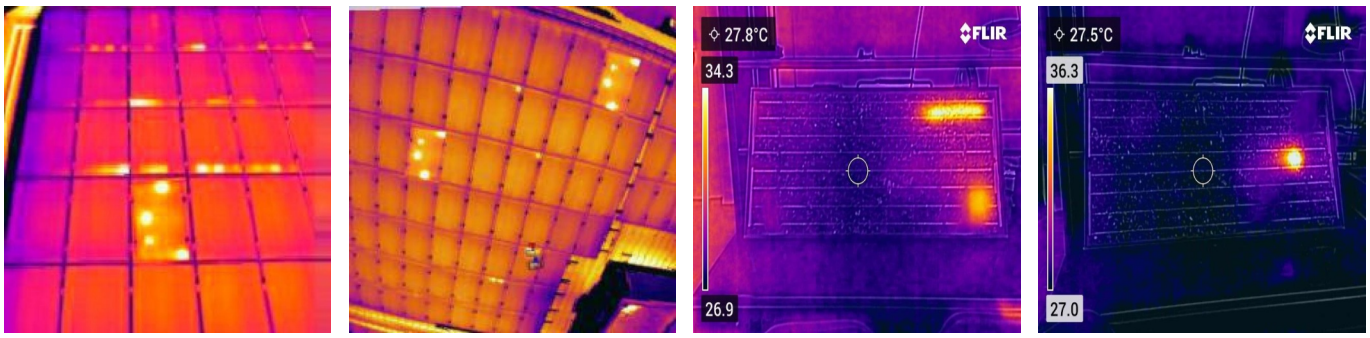


Figure 3: Dataset images. The images were acquired in the laboratory LACEA, along with images from [24].

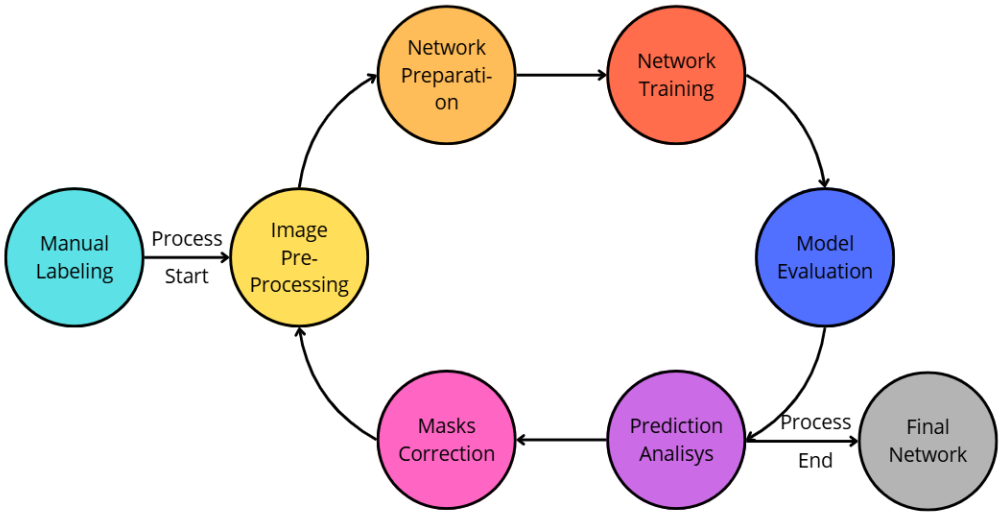


Figure 4: Labeling processing cycle of pseudo-labeling.

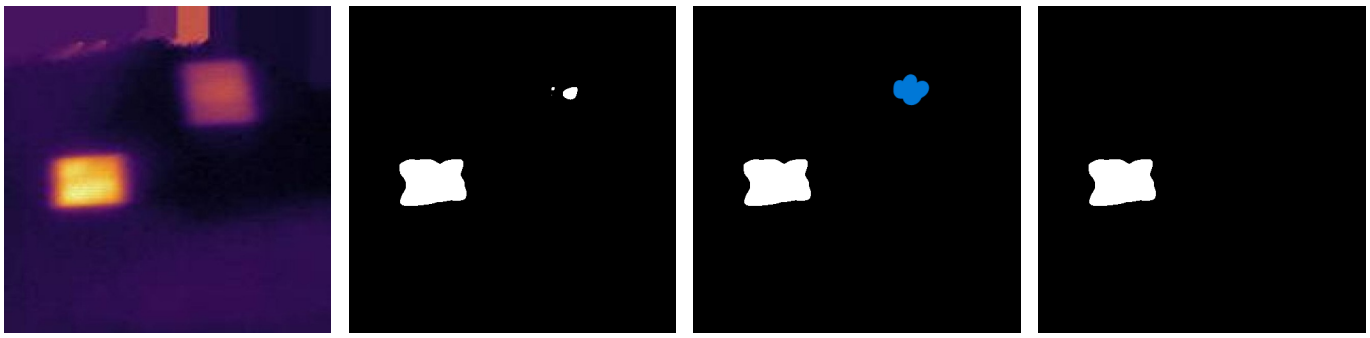
vii) extracted predictions are manually corrected or discarded, viii) The corrected masks are pre-processed, and the cycles repeat if there is still data to be labeled. If there is no unlabeled data, ix) the dataset is ready to be used in the training of the final network.

From the total dataset of 500 images, 100 images (20%) were manually annotated and used to train the initial model. The trained model was then employed to generate pseudo-labels for the remaining 400 images (80%). This resulted in a semi-supervised learning configuration with a manual-to-pseudo-label proportion of approximately 1:4. The generated pseudo-labels were subsequently subjected to a manual quality control procedure. Each predicted mask was visually inspected and evaluated according to its agreement with the underlying image content. Masks exhibiting high agreement (approximately $\geq$80% spatial correspondence with the expected anatomical/target region) were retained and, when minor local inaccuracies were present, manually refined as shown in Figure 5. Conversely, masks with substantial segmentation errors (i.e., estimated agreement below 80%) were discarded from the pseudo-labeled set. This procedure ensured that only high-quality pseudo-labels were incorporated into the final training dataset. After this analysis, the new annotated images were used to expand the manually annotated dataset and train a new network. This process was repeated 3 times until the entire original dataset was annotated with quality masks to be used in the final training of our network. After this curation process, 487 images were retained for training, corresponding to a pseudo-label rejection rate of 3.25% (13 out of 400 pseudo-labeled images).

The entire pseudo-labeling process was made without a limited number of training epochs. Instead, an early stopping parameter with a limit of 10 epochs and controlled by the F1 score metric (weighted average between precision and recall) was implemented. The choice for the F1 score as the early stopping parameter was due to this metric's meaning of how well the model is balancing the true positive prediction with the total predictions.

The dataset was divided into training and validation subsets to ensure a balanced evaluation of the proposed segmentation framework. The thermal images and their corresponding masks were split into two partitions: one used for model optimization and the other for performance monitoring during training. Specifically, 66.7% of the images were allocated for training, 22.2% for validation, and 11.1% for testing, ensuring a proportional distribution of hotspot and non-hotspot samples across the splits. The training set was used to update model weights through backpropagation, while the validation set was reserved for adjusting hyperparameters, early stopping, and preventing overfitting. Images were processed in batches of sixteen, with shuffling applied only to the training subset to improve generalization. The test set, unseen during training, was later used for final performance assessment.

https://doi.org/10.21528/lnlm-vol23-no2-art5



(a) (b) (c) (d)

Figure 5: Pseudo-labels correction process: (a) Original thermal image, (b) Predicted mask, (c) Mask manual correction, (d) Corrected mask

Before training, all thermal images were normalized to a consistent radiometric range and resized to ensure compatibility with the network input dimensions of 512x512 pixels. To improve generalization and reduce overfitting, data augmentation techniques were applied, including random horizontal and vertical flips, small rotations, and adjustments of brightness and contrast. These transformations mimic the variability encountered in real PV field inspections, including different camera angles and varying illumination conditions.

3 Results and Discussion

3.1 Setup

All the tests were run in an environment consisting of a system with 12GB of RAM and an NVIDIA Tesla T4 GPU with 15GB of RAM. The dataset was divided into 300, 150, and 50 images for training, validation, and testing, respectively. The training process had no fixed epoch limit, relying instead on an early stopping strategy with a patience factor of 10 epochs, and a batch size of 16 images was employed. To avoid configuring the entire network from scratch and manually configuring its layers, transfer learning was applied by selecting ResNet34 as the encoder of the U-Net, initialized with pre-trained ImageNet weights.

To evaluate the model's effectiveness and generalization capability, six metrics were collected and analyzed: (i) train loss; (ii) validation loss; (iii) precision; (iv) recall; (v) F1 score; (vi) IoU. Each one of these metrics says something about a crucial characteristic of the network and might be thoroughly analyzed to ensure the network quality.

3.2 Results and Performance Analysis

First, we have the train loss curve. This metric indicates whether the model is learning from the training data and how effectively it generalizes from this part of the dataset. This metric is expected to start high and decrease as the epochs progress. Figure 6 allows us to visualize that the train loss curve gets really low from the 100th epoch, asymptotically tending to 0. From the training log, we see that in the best epoch, the train loss reaches a mark of 4.03%, indicating that the model learns excellently from the training data.

Next, we have the validation loss metric. In this process, we feed the model with data that is not in the training dataset. This is one of the most important steps in network training because it tells how well the patterns that the model learns in the training process generalize to unseen data. Furthermore, this is used to correct the network's neurons and bias, instructing the model via backpropagation on whether it needs to search for more intrinsic patterns in the training dataset to help the model generalize better. It's expected from this metric the same behavior as the train loss, starts high and gets lower as the epochs go by. Figure 7 demonstrates the generalization capability achieved by the model as the training evolves without the presence of overfitting in the training. The training log shows that at the best epoch, the model achieves a validation loss value of 13.0%, meaning that the model has a very effective training.

The following metrics to be evaluated are the precision, recall, and F1-score. These three metrics are put together because they are directly connected. The precision is a metric that indicates the percentage of correctly predicted pixels in each prediction, i.e., the amount of True Positives (TP) present in each supposed positive prediction. The recall, on the other hand, is a metric that shows the amount of pixels that the model correctly classifies as positive that the model correctly classifies as positive. In other words, it indicates the number of pixels that should be TP the model marks as TP. Unfortunately, there is a problem with these two metrics alone: the precision tends to start very low and the recall very high, which might confuse the data analysis. This happens because at the start of the training, the model tends to classify almost every pixel as a positive one, which causes it to classify every TP pixel as a positive, causing the recall to start high but also causing the precision to be low, because there are a lot of False Positives (FP) in the first predictions.

To balance the divergence between precision and recall, the F1-score metric is established. This metric is the harmonic mean between precision and recall, which balances the divergence between these two metrics and provides a more realistic visualization of their evolution during training. Figure 8 shows the evolution of the precision curve during the advance of the training epochs,

https://doi.org/10.21528/lnlm-vol23-no2-art5

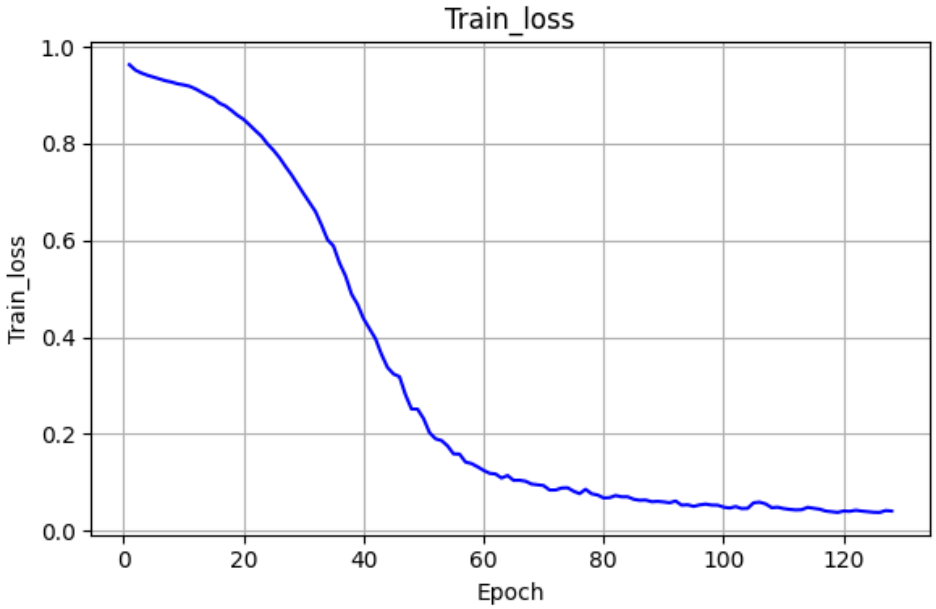


Figure 6: Training loss curve showing the convergence of the model

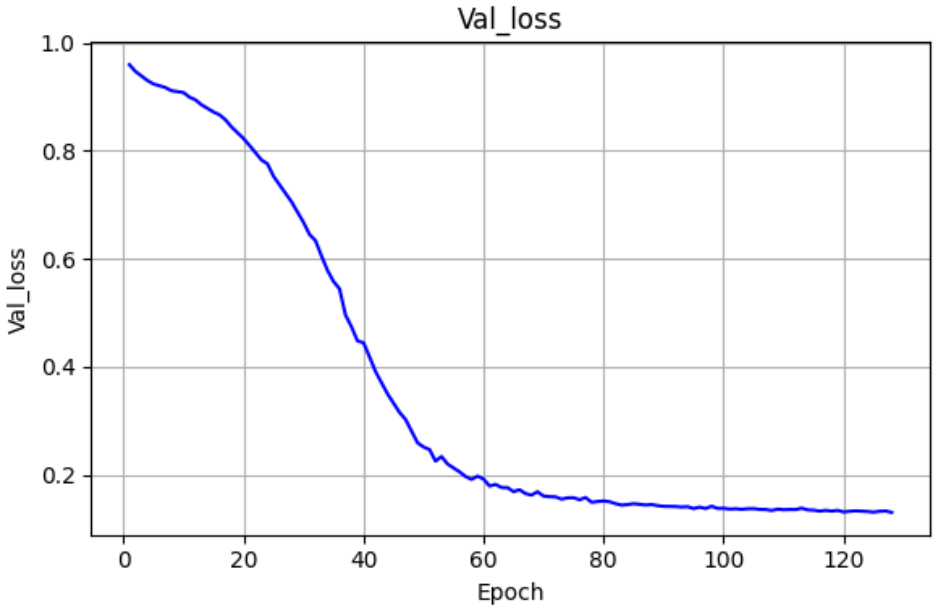


Figure 7: Validation loss curve indicating the generalization capability of the model

showing that the model achieves very high ratios at the end of the training. The Figure 9 indicates the recall ratios as the train progresses, indicating a very high accuracy in detecting the hotspots. Finally, Figure 10 illustrates the balance between these two metrics as the training goes by. The train log indicates that the model achieves the best epoch values of 87.90%, 90.05%, and 88.96% for precision, recall, and F1-score, respectively.

Finally, we have the last evaluation metric: the IoU. This metric is of extreme importance because it goes beyond the model accuracy, measuring how many pixels belonging to the prediction are present in the label. In tasks such as image segmentation, it is even more critical because it is necessary to know the exact shape of the prediction. For hotspot detection, the IoU will indicate not only the localization of the hotspot in the camera frame but also its shape, which is crucial for understanding how these form and how to avoid them. Figure 11 indicates the percentage of the predicted masks that overlap with the labels during the network training. The training log indicates that at the best epoch, the IoU achieved a mean of 74.64%, showing a great percentage of overlap between prediction and label.

Analysis of the loss, precision, recall, and IoU curves has already demonstrated the network's good performance, both in terms of learning and generalization. To complement this assessment, Table 3 summarizes the values obtained in the best epochs, providing a consolidated numerical view of the results.

https://doi.org/10.21528/lnlm-vol23-no2-art5

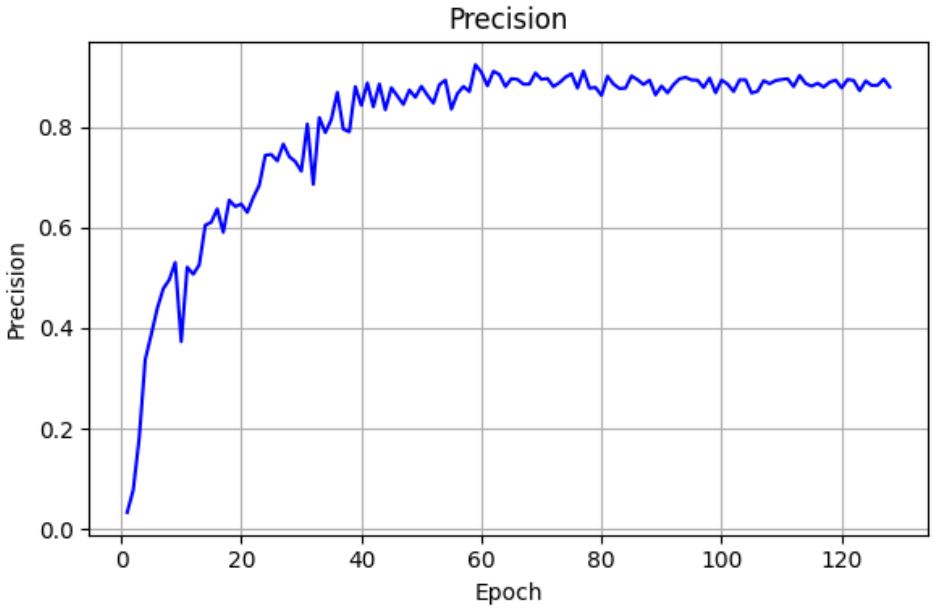


Figure 8: Precision curve obtained during the evaluation of the model.

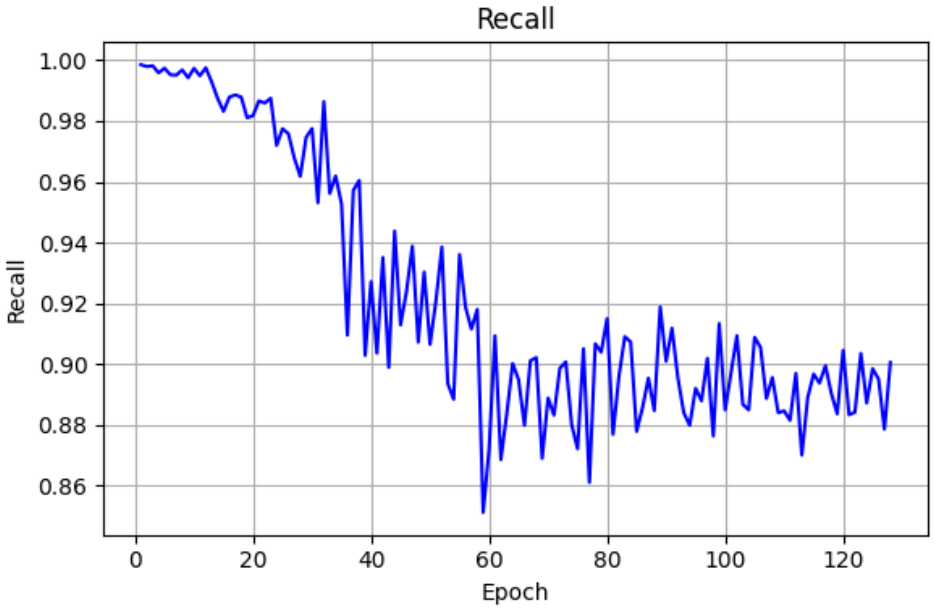


Figure 9: Recall curve at different decision thresholds.

Metric	Value
Training Loss (best epoch)	4.03%
Validation Loss (best epoch)	13.0%
Precision	87.90%
Recall	90.05%
F1-Score	88.96%
Intersection over Union (IoU)	74.64%

Table 3: Performance metrics of the U-Net model for hotspot detection in photovoltaic modules.

To acquire a wider view of the model's capability, we can use the confusion matrix. The confusion matrix highlights the classifier's behavior, showing a high accuracy rate for the Background class and a really good performance in identifying the target class (i.e., hotspot). It is a tool used to evaluate the performance of segmentation models. It organizes the model's predictions into four categories: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). In this study, TP represents correctly identified hotspot pixels, TN corresponds to correctly classified background pixels, FP indicates

https://doi.org/10.21528/lnlm-vol23-no2-art5

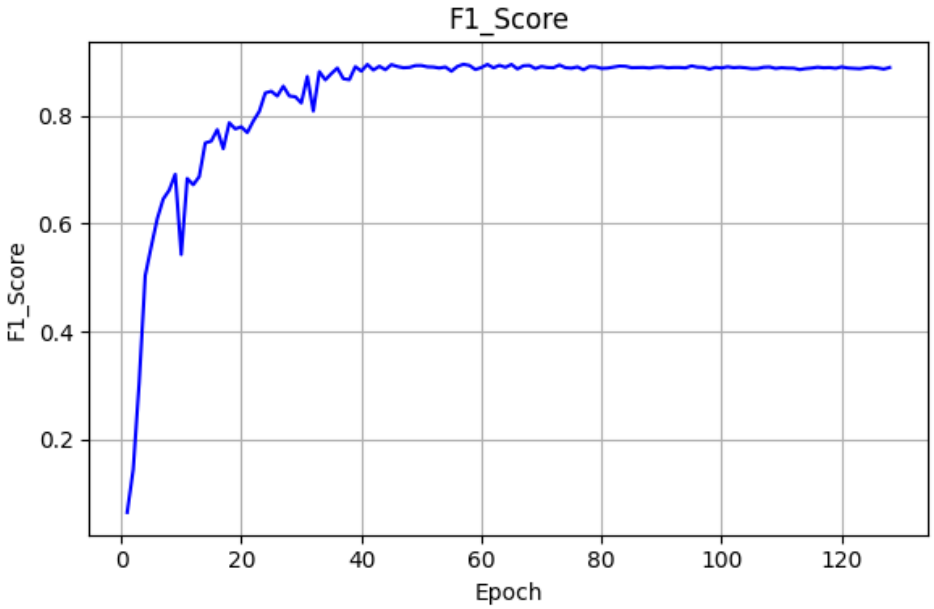


Figure 10: Balance between precision and recall in F1-Score

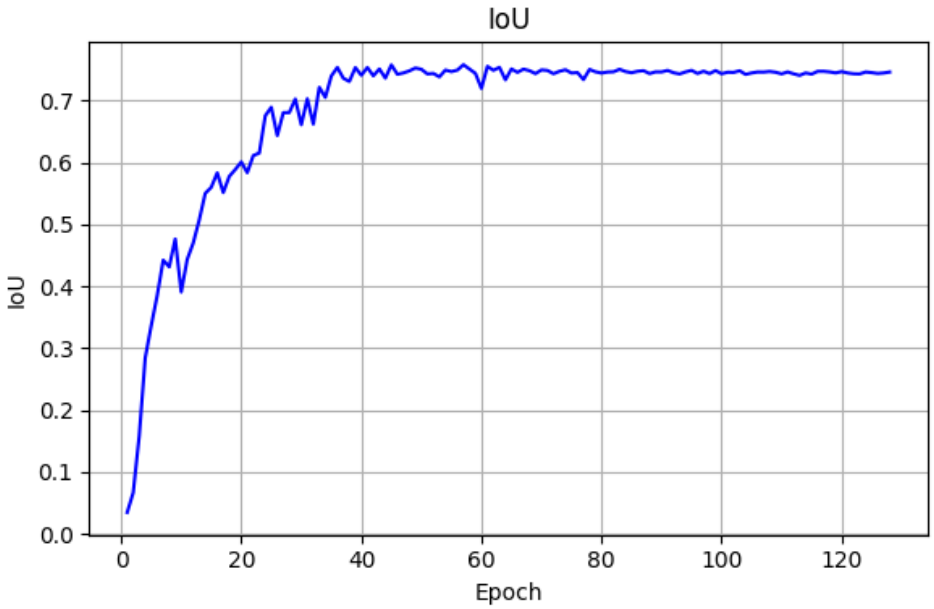


Figure 11: Intersection over Union curve indicating the mean percentage of overlap between prediction and mask.

Table 4: Confusion matrix showing the segmentation performance of the network.

True / Predicted	Background	Hotspot
Background	99.83%	0.17%
Hotspot	9.95%	90.05%

background pixels misclassified as hotspots, and FN are hotspot pixels misclassified as background. By arranging predictions in this table, the confusion matrix allows a more detailed analysis of the model's performance and supports the calculation of evaluation metrics such as accuracy, precision, recall, specificity, and F1-score. This result indicates that the model achieves a very high accuracy when classifying the background class (99.83%) and also in the correct segmentation of hotspot pixels (90.05%), with a misclassified total percentage of 10.12% (false negative + false positive), as shown in Table 4. This data highlights the efficiency of the pseudo-labeling technique, the quality of the chosen dataset, and the good selection of training hyperparameters, showing that the model efficiently converged to a very high accuracy stage.

Table 5 summarizes the quantitative evaluation of our model on small and large hotspot instances. Small hotspots achieved higher instance-level IoU (0.5560) and Precision-Recall AUC (0.8027) compared with large hotspots (IoU: 0.3305, Precision-

https://doi.org/10.21528/lnlm-vol23-no2-art5

Hotspot Size	IoU (instance)	FN Rate	Precision-Recall AUC
Small	0.5560	0.1329	0.8027
Large	0.3305	0.2745	0.4738

Table 5: Comparison of model performance between small and large hotspot instances.

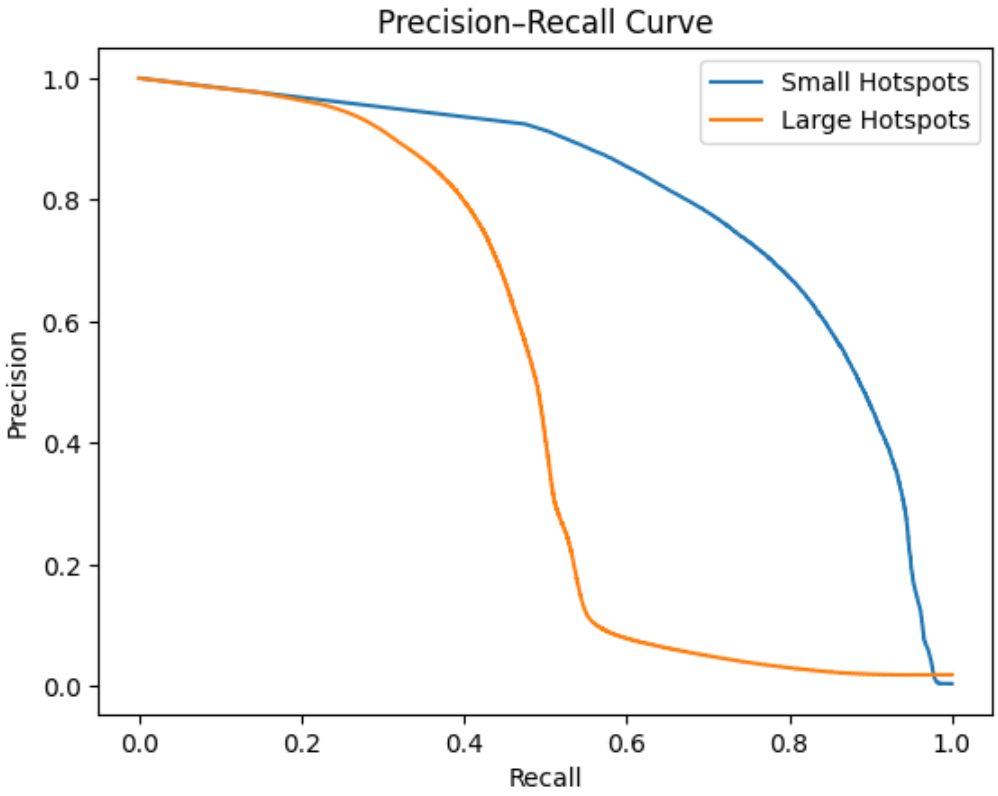


Figure 12: Precision-Recall curve comparison between small and large hotspots.

Recall AUC: 0.4738). The false negative rate was low for both categories, with 13.3% for small and 27.5% for large hotspots, indicating that most anomalous regions were detected.

The higher IoU and Precision-Recall AUC observed for small hotspots can be explained by the relative impact of minor boundary errors: small errors on small hotspots typically affect fewer pixels proportionally, preserving high precision and recall. In contrast, large hotspots are more sensitive to small segmentation inaccuracies along their boundaries, which results in lower IoU and PR AUC despite the model correctly identifying most of the hotspot region.

Figure 12 illustrates the precision-recall curves for small and large hotspots on the same image, highlighting this difference in sensitivity. The relatively low false negative rate across both categories suggests that critical hotspots are rarely missed, with small false negatives having minimal effect on small hotspots but potentially more impact on large hotspots due to their spatial extent.

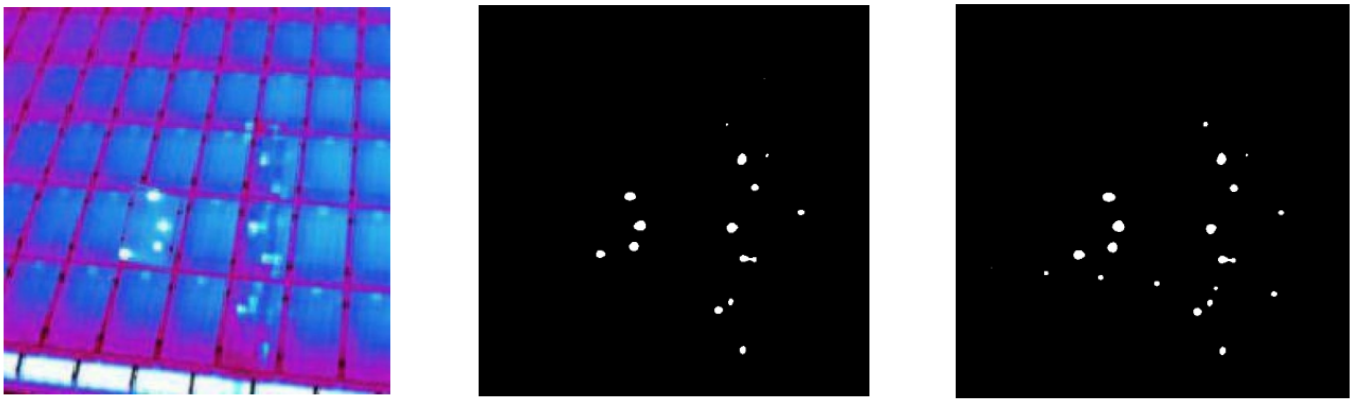
Figure 13 illustrates the prediction result, evidencing the model's ability to identify the anomalous regions correctly. It is possible to verify that the model successfully identifies hotspot regions with good delineation of their shapes. Compared with bounding-box-based approaches such as YOLO, segmentation enables finer-grained localization, allowing not only detection but also per-instance analysis of hotspot size, morphology, and spatial distribution across PV modules.

3.3 Comparison with YOLO-Based Object Detection

To further contextualize the performance of the proposed segmentation approach, we compared the U-Net results with a YOLOv11n (Nano) detection model trained on the same dataset. Figure 14 illustrates the training progression of YOLOv11n, while Figures 15 and 16 depict its precision–recall curves and F1 score evolution, respectively. This comparison highlights differences in detection granularity, localization accuracy, and overall ability to capture hotspot regions between bounding-box-based detection and pixel-level segmentation. All comparisons are performed on the same test dataset, ensuring a consistent evaluation for both networks. The metrics considered in this analysis include frames per second (FPS) and computational cost, providing insight into both real-time applicability and resource efficiency of the two approaches. Note that additional implementation details and extended analysis of the YOLO-based framework can be found in [11].

The comparative evaluation between U-Net and YOLO is summarized in Table 6. As observed, the two networks achieved

Learning and Nonlinear Models - Journal of the Brazilian Society on Artificial Intelligence (SBIA), Vol. 23, Iss. 2, pp. 59–74, 2025
https://doi.org/10.21528/lnlm-vol23-no2-art5



(a) (b) (c)

Figure 13: Prediction results: (a) Dataset image; (b) Ground truth; (c) Predicted hotspot.

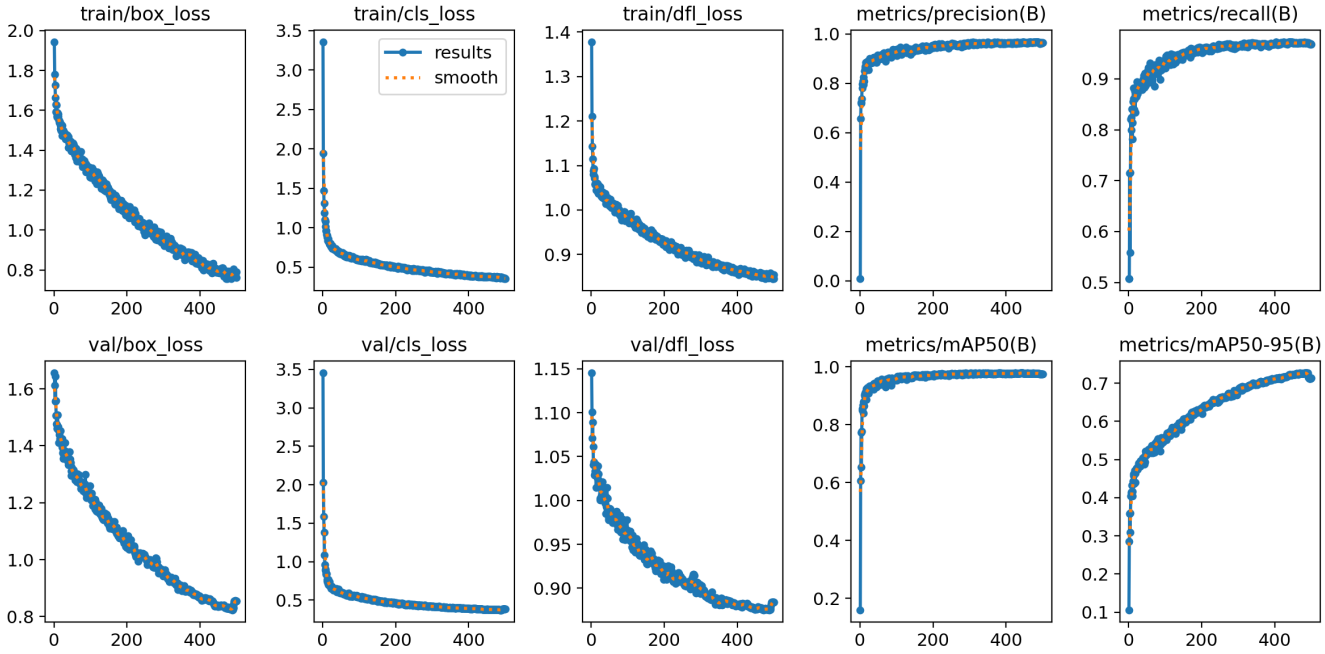


Figure 14: YOLOv11n train and validation losses results

Model	Precision	Recall	IoU	FPS	RAM (MB)
U-Net	0.786	0.635	0.539	10.54	288.87
YOLOv11n	0.695	0.797	0.526	31.90	259.42

Table 6: Comparison between U-Net and YOLO performance and computational cost.

closely aligned performance in terms of precision, recall, and IoU. U-Net obtained a precision of 0.786 and a recall of 0.635, while YOLO achieved 0.695 and 0.797, respectively. Similarly, the IoU values are highly comparable (0.539 for U-Net and 0.526 for YOLO). This proximity in segmentation and detection performance was expected, since both models were trained on the same dataset under similar training conditions, ensuring a consistent experimental setup.

In contrast, the computational metrics reveal a more pronounced difference between the two approaches. YOLO achieved 31.90 FPS, while U-Net reached 10.54 FPS, meaning that YOLO operates at approximately three times the inference speed of U-Net. This substantial difference is consistent with the architectural nature of each method, i.e., YOLO performs object detection using bounding boxes, which is computationally less intensive than the dense pixel-level processing required by U-Net for semantic segmentation. A similar trend is observed in memory consumption, where YOLO requires less RAM than U-Net. These results confirm that, while both models provide comparable detection capability, YOLO offers significantly higher computational efficiency, whereas U-Net provides more detailed pixel-wise representations at the cost of increased processing time and memory usage.

Figure 17 further illustrates the qualitative prediction capabilities of both networks. As observed, U-Net and YOLO produce

https://doi.org/10.21528/lnlm-vol23-no2-art5

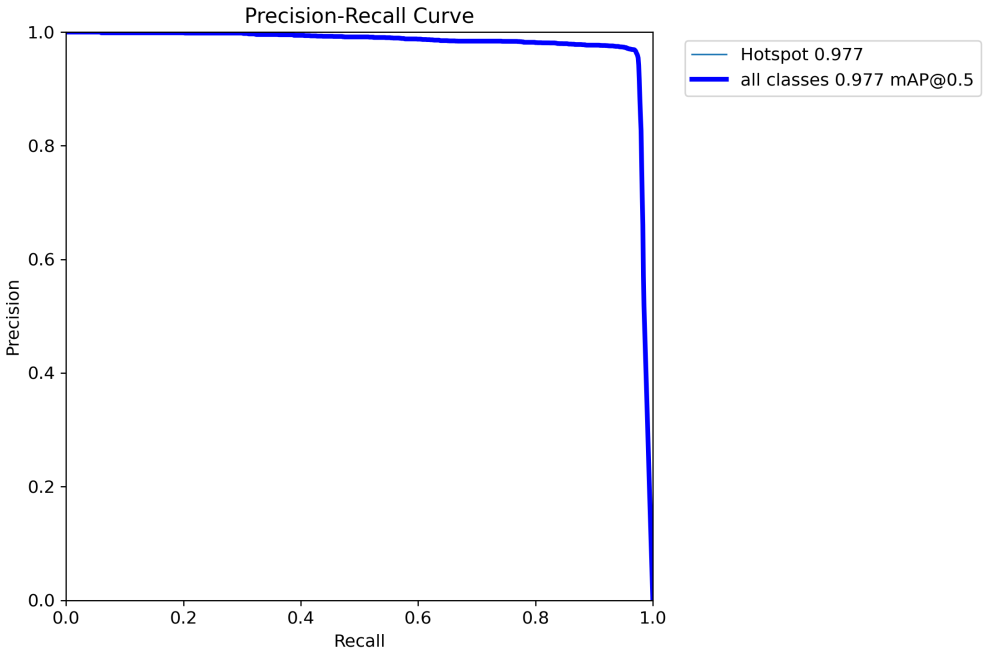


Figure 15: YOLOv11n Precision-Recall curve

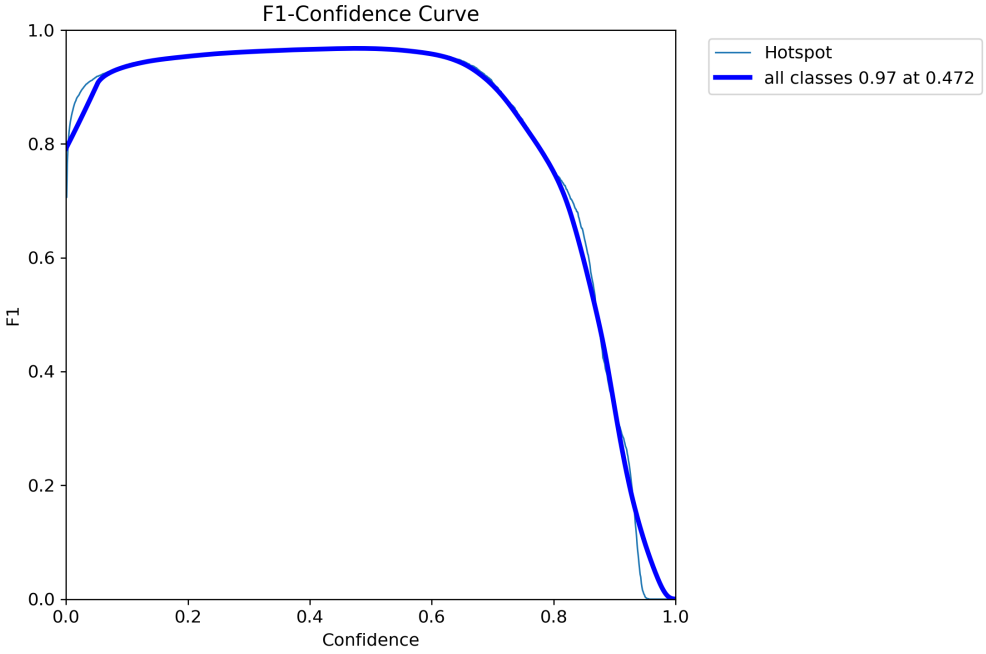
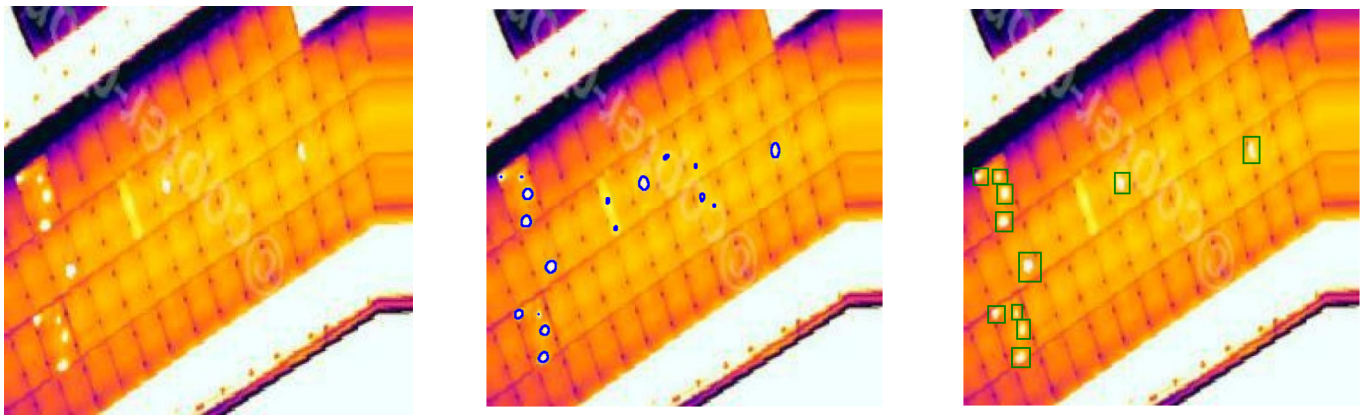


Figure 16: YOLOv11n F1-Score results

visually comparable hotspot detections across the evaluated samples, reinforcing the quantitative results previously discussed. Although the two approaches rely on fundamentally different detection strategies, the predicted regions show high correspondence in spatial coverage and hotspot identification. These visual examples confirm that both models achieved similar detection performance when trained on the same dataset under comparable conditions, with differences primarily arising from their distinct representation mechanisms rather than from discrepancies in learning quality.

Although the quantitative results are closely aligned and YOLO demonstrates superior computational performance, it inherently lacks the ability to capture the precise morphology of hotspot regions. As a bounding-box-based detector, YOLO localizes areas of interest but does not provide detailed information about their internal structure or exact spatial boundaries. This limitation is particularly relevant in scenarios where the geometric characteristics of hotspots, such as shape irregularities, spread patterns, or boundary definitions, are essential for diagnosing their underlying causes. In contrast, U-Net performs pixel-level segmentation, enabling accurate delineation of hotspot contours and fine-grained structural analysis. Therefore, while YOLO of-

https://doi.org/10.21528/lnlm-vol23-no2-art5



(a) (b) (c)

Figure 17: Prediction comparison results: (a) Original image; (b) U-Net prediction; (c) YOLO prediction.

fers advantages in speed and efficiency, U-Net becomes especially valuable when detailed morphological information is required for deeper interpretation and analysis of the detected regions.

Although the quantitative results are closely aligned and YOLO demonstrates superior computational performance, it inherently lacks the ability to capture the precise morphology of hotspot regions. As a bounding-box-based detector, YOLO localizes areas of interest but does not provide detailed information about their internal structure or exact spatial boundaries. This limitation is particularly relevant in photovoltaic systems, where the geometric characteristics of hotspots, such as contour irregularity, spatial spread across cells, and boundary definition, are directly associated with specific failure mechanisms, including cell cracks, interconnection faults, and localized degradation.

3.4 Discussion

The numerical results obtained with the U-Net model, such as a precision of 87.90%, recall of 90.05%, F1-score of 88.96%, and an IoU of 74.64%, demonstrate the ability of the framework to provide accurate pixel-level segmentation of hotspot regions. Besides, these metrics highlight the potential of segmentation methods to serve as a quantitative basis for standardized fault classification procedures.

When compared with related U-Net-based approaches, important differences emerges. Ruan et al. [18] combined U2-Net for module segmentation with YOLOv4 for defect detection at the module level. While effective for identifying defective modules, their approach does not focus on pixel-level hotspot delineation, nor does it report IoU values specifically for hotspot segmentation. In contrast, the present work performs direct hotspot segmentation and quantifies spatial overlap through IoU (74.64%), enabling geometric measurement of anomaly extent. Ren et al. [20] employed a U-Net with a ResNet50 backbone for detecting small solar installations in UAV RGB imagery, reporting an Average Precision of approximately 0.80 in real-world scenarios. However, their objective was panel detection rather than thermographic hotspot segmentation. The proposed method instead targets pixel-level thermal anomaly delineation, achieving a recall of 90.05%, which is particularly relevant in preventive maintenance applications where minimizing missed hotspots is critical. Rudro et al. [21] combined U-Net with InceptionV3 for multi-defect classification, emphasizing categorical accuracy rather than contour-level segmentation quality. In contrast, the present work explicitly reports IoU (74.64%) and confusion matrix metrics, including 99.83% correct background classification and 90.05% hotspot pixel detection, demonstrating both high sensitivity and controlled false positives.

Compared to bounding-box-based approaches such as YOLO-based thermal anomaly detectors [11], segmentation provides a substantial advantage in terms of spatial characterization. While YOLO achieves high detection speed, it produces rectangular regions that do not accurately represent hotspot geometry. The present segmentation masks allow direct measurement of hotspot area, perimeter, and spatial distribution within a PV module, which are parameters directly relevant to IEC TS 62446-3:2017 classification criteria.

In this sense, beyond quantitative performance, IEC TS 62446-3:2017 [22] establishes detailed guidelines for outdoor infrared thermography of PV modules and plants, including requirements for equipment, environmental conditions, and standardized methods for the evaluation of thermal anomalies. Annex C of the document provides a comprehensive matrix of thermal abnormalities, including representative thermographic patterns and temperature difference thresholds for various defect classes.

Typical anomalies identified during infrared inspections include isolated hotspots, bypassed substrings, and characteristic patchwork patterns. These anomalies are classified into specific Classes of Abnormality (CoA) according to their thermal signature and severity, which determines the urgency and type of maintenance action required. By enabling precise delineation of hotspot contours, the U-Net segmentation framework can accurately estimate parameters such as geometric size, morphology, and thermal extent. These are essential descriptors for the allocation of anomalies to the correct CoA, as defined by the IEC standard.

Integrating segmentation-based outputs with the IEC classification scheme opens the way to standardized and automated thermal inspection workflows. This integration not only assists repeatability and objectivity but also reduces dependence on

manual expert evaluation. Besides, it provides a scalable methodology for inspecting large-scale PV plants in accordance with international guidelines, thus bridging advanced deep learning techniques with established industrial practices.

4 Conclusions and Future Work

This work presented the implementation and evaluation of a U-Net-based segmentation approach for detecting hotspots in photovoltaic modules using thermal infrared imagery. The results indicated the potential of the approach as a reliable tool for large-scale PV inspection through the use of UAVs and deep learning techniques. Both quantitative results, expressed through precision, recall, and the confusion matrix, and qualitative analyses on real thermal images showed the future potential of the U-Net-based model. This study evidences the quality of the dataset and pseudo-labeling technique by the results obtained from the networks' training, showing that the model achieves a high generalization capability and correct shape detection of the hotspot class. It illustrates the U-Net's precision when classifying and segmenting classes, achieving high shape accuracy consistent with manually labeled data. However, some well-known limitations of natural segmentation tools were observed. Since the network needs to read and classify each pixel, it suffers from a lack of execution time, making it unsuitable for real-time applications.

Despite these promising outcomes, certain limitations inherent to pixel-wise segmentation were observed. Since the network must process every pixel, the inference time remains a bottleneck, preventing direct application in real-time UAV inspections. Future work should therefore concentrate on strategies to achieve near real-time performance, including the design of lightweight architectures and the adoption of model compression techniques such as quantization and pruning, making the approach viable for embedded hardware platforms. Another important direction involves strengthening the model's robustness to domain variability, ensuring consistent performance under different illumination conditions, weather scenarios, and levels of panel soiling. Expanding the framework beyond hotspot detection to incorporate other anomalies such as cracks, delamination, and dirt accumulation would further enhance its utility for predictive maintenance. Finally, integrating the segmentation model with autonomous UAV systems opens the possibility of end-to-end inspection pipelines where detection, prioritization, and navigation decisions are performed jointly, paving the way for fully automated and scalable PV monitoring solutions. An additional limitation of the present study is that the experimental validation is conducted using a single thermal camera model and a single photovoltaic plant. Although the obtained results demonstrate promising performance, future investigations should evaluate the proposed framework using different thermal imaging sensors and distinct photovoltaic installations. Such experiments would allow a more comprehensive assessment of the model's robustness, transferability, and generalization capability across different hardware characteristics and environmental conditions.

Acknowledgments

The authors thank CEFET-RJ, IFNMG, and the Brazilian research agencies CAPES, CNPq, and FAPERJ for supporting this work. This work was carried out with the support of the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

References

[1] IRENA — International Renewable Energy Agency. "World Energy Transitions Outlook 2024: 1.5°C Pathway". `https://www.irena.org/publications`, 2024.

[2] M. Dhimish, M. Theristis and V. D'Alessandro. "Photovoltaic hotspots: A mitigation technique and its thermal cycle". *Optics & Laser Technology*, vol. 173, pp. 110220, 2024.

[3] Q. Ge, Z. Li, Z. Sun, J. Xu, H. Long and T. Sun. "Low Resistance Hot-Spot Diagnosis and Suppression of Photovoltaic Module Based on I-U Characteristic Analysis". *Energies*, vol. 15, no. 11, 2022.

[4] W. A. d. Souza, R. C. R. Souza and A. M. Minori. "Boas práticas de manutenção preventiva em sistemas fotovoltaicos". *Brazilian Journal of Development*, vol. 5, no. 8, pp. 12779–12791, 2019 2019.

[5] A. M. Muhmad Kamarulzaman, W. S. Wan Mohd Jaafar, N. A. Abu Bakar, K. Chew, N. I. Khairi Anuar, A. I. Anuar, A. F. Mat Som, S. Mohamad Zamri and M. Z. H. Zainorzuli. "Comprehensive Analysis of Solar Photovoltaic System Defects and Inspection Techniques in Tropical Environments Using Thermal UAV: Study Case in Marang, Terengganu". In *E3S Web of Conferences, 6th International Conference on Science and Technology Applications in Climate Change (STACLIM 2024), Volume 599*, p. 03003 (10 pages), January 2024.

[6] N. Sindhwani, R. Anand, S. Meivel, R. Shukla, M. P. Yadav and V. Yadav. "Performance Analysis of Deep Neural Networks Using Computer Vision". *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, vol. 8, no. 29, pp. e3, 2021.

[7] T. K. Johnsen, I. Harshbarger and M. Levorato. "An Overview of Adaptive Dynamic Deep Neural Networks via Slimmable and Gated Architectures". In *2024 15th International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 252–256, 2024.

[8] A. Goel, C. Tung, Y.-H. Lu and G. K. Thiruvathukal. "A Survey of Methods for Low-Power Deep Learning and Computer Vision". In *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*, pp. 1–6, 2020.

[9] K. Chandrashekhar, P. Leela, K. R. Madhavi, J. Avanija and N. Tangudu. "Grey Wolf Optimizer and Deep Neural Network based Feature Selection and Classification in Medical Image Analysis". In *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 957–960, 2024.

[10] D.-d. Zuo, J.-q. Ge and Y.-m. Yu. "Surface Defect Detection of Button Cells Based on the Fusion of Object Detection and Semantic Segmentation". *Arabian Journal for Science and Engineering*, January 2025.

[11] E. V. B. Santana, M. F. Pinto, Y. N. da Silva, L. A. Vasconcelos, T. M. B. Santos and B. G. Antunes. "Automated Hotspot Detection in Photovoltaic Panels Using YOLOv8-Based Thermal Image Analysis". In *Anais do Simpósio Brasileiro de Automação Inteligente (SBAI)*, São João del Rei, Brasil, 2025.

[12] J. Amorim, A. Neto, R. Chaves, A. Zachi, J. Gouvêa, F. Andrade and M. Pinto. "Collaborative Inspection of Solar Panel Farms Using YOLOv5-Based Computer Vision and UGV-UAV Integration". *Journal of Intelligent & Robotic Systems*, vol. 111, no. — (article number 66), pp. —, May 2025.

[13] F. Wang, Z. Wang, Z. Chen, D. Zhu, X. Gong and W. Cong. "An Edge-Guided Deep Learning Solar Panel Hotspot Thermal Image Segmentation Algorithm". *Applied Sciences*, vol. 13, no. 19, 2023.

[14] O. Ronneberger, P. Fischer and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, volume 9351, pp. 234–241, 2015.

[15] H. Chen and K. Kim. "Multi-Convolutional Channel Residual Spatial Attention U-Net for Industrial and Medical Image Segmentation". *IEEE Access*, vol. 12, pp. 76089–76101, 2024.

[16] H. Lu, Y. She, J. Tie and S. Xu. "Half-UNet: A Simplified U-Net Architecture for Medical Image Segmentation". *Frontiers in Neuroinformatics*, vol. 16, pp. 911679, 2022.

[17] Z. Yang, X. Sun, Y. Yang and X. Wu. "MEDU-Net+: A Novel Improved U-Net Based on Multi-Scale Encoder-Decoder for Medical Image Segmentation". *KSII Transactions on Internet and Information Systems*, vol. 18, no. 7, pp. 1706–1725, July 2024.

[18] C. Ruan, W. Tang, X. Hu and W. Yan. "Deep Learning-based Method for PV Panels Segmentation and Defects Detection with Infrared Images". In *Proceedings of the 2021 China Automation Congress (CAC)*, pp. 7166–7171, 2021.

[19] J. Liu and N. Ji. "A bright spot detection and analysis method for infrared photovoltaic panels based on image processing". *Frontiers in Energy Research*, vol. 10, no. —, pp. 978247, 2023.

[20] S. Ren, J. Malof, R. Fetter, R. Beach, J. Rineer and K. Bradbury. "Utilizing Geospatial Data for Assessing Energy Security: Mapping Small Solar Home Systems Using Unmanned Aerial Vehicles and Deep Learning". *ISPRS International Journal of Geo-Information*, vol. 11, no. 4, 2022.

[21] R. A. M. Rudro, K. Nur, M. F. A. Al Sohan, M. F. Mridha, S. Alfarhood, M. Safran and K. Kanagarathinam. "SPF-Net: Solar panel fault detection using U-Net based deep learning image classification". *Energy Reports*, vol. 12, pp. 1580–1594, July 2024.

[22] International Electrotechnical Commission. "IEC TS 62446-3:2017 — Photovoltaic (PV) systems — Requirements for testing, documentation and maintenance — Part 3: Photovoltaic modules and plants — Outdoor infrared thermography". Technical report, IEC, Geneva, Switzerland, 2017.

[23] G. D. Figueiredo da Silva, J. T. Dias, L. Faletti Almeida and M. Faria Pinto. "Analysis of Neural Network Architectures for Semantic Segmentation of Seismic Data". In *Science and Information Conference*, pp. 643–658. Springer, 2023.

[24] A. U. of Sharjah. "Hotspot detection in solar panel Dataset". `https://universe.roboflow.com/american-university-of-sharjah/hotspot-detection-in-solar-panel`, mar 2023. visited on 2025-02-19.

[25] E. V. B. Santana. "wTCC_painel_programa10: Detector de Hotspots em Painéis Solares usando YOLOv8". `https://github.com/eduvbs/wTCC_painel_programa10/tree/main`, 2025. acessado em 03 de setembro de 2025.