



OPEN Optimized YOLO based model for photovoltaic defect detection in electroluminescence images

Achit Mohamed¹, Yassa Nacera¹, Bouzida Ahcene¹, Ali Teta², El Ouanas Belabbaci³, Abdelaziz Rabehi⁴, Yousef A. Alsabah⁵✉ & Mohamed Benghanem⁶

Ensuring the reliability of photovoltaic (PV) systems requires efficient defect detection to maintain optimal energy production. Deep learning-based object detection models have demonstrated remarkable performance in automating this process. In this study, PV-YOLOv12n is introduced as an optimized variant of YOLOv12n, tailored for defect detection in electroluminescence (EL) images of PV panels. The modifications incorporate an A2C2f module at the P5 scale (1024, True), which enhances feature extraction by prioritizing critical defect regions. This improvement significantly boosts recall and precision for detecting large cracks, significant dislocations, and widespread material inconsistencies. Experimental results on the PVEL-AD and Roboflow datasets demonstrate superior detection performance. PV-YOLOv12n achieves a mAP@50 of 0.91 on both datasets, surpassing the baseline YOLOv12n (mAP@50 of 0.90 and 0.88 for PVEL-AD and Roboflow, respectively). Additionally, mAP@50-95 increases to 0.58 on PVEL-AD and 0.75 on Roboflow, highlighting improved generalization. Despite these improvements, inference speed remains efficient at 4.24 ms for PVEL-AD and 4.43 ms for Roboflow, ensuring suitability for real-time applications. These results validate the effectiveness of PV-YOLOv12n in detecting critical PV panel defects, supporting its deployment in large-scale solar farm inspections.

Keywords Photovoltaic defect detection, Electroluminescence imaging, Deep learning, YOLOv12n, Feature extraction

The widespread integration of photovoltaic (PV) systems as a cornerstone of renewable energy infrastructure has underscored the critical importance of their systematic maintenance to ensure optimal performance and longevity¹. PV panels, continuously exposed to harsh environmental elements, are susceptible to a spectrum of defects, including but not limited to soiling from dust and debris², physical damage such as cracks and micro-cracks³, hotspots, delamination, and bird droppings, all of which can significantly curtail energy conversion efficiency and operational lifespan⁴. Failure to promptly identify and mitigate these issues can escalate into substantial long-term degradation, diminished power output, and inflated operational expenditures. While traditional manual inspection has long been the standard, its inherent labor intensity, time consumption, and potential for human error render it increasingly impractical and inefficient, especially for expansive utility-scale solar farms^{5,6}. This paradigm necessitates a shift towards automated defect detection systems capable of delivering scalable, rapid, and reliable assessments, a challenge that captivates researchers in renewable energy, computer vision, and artificial intelligence⁷. Automated PV defect detection, primarily relying on the analysis of visual or thermal imagery, presents a complex computer vision task. The visual data captured from PV panels is rich with information, but its effective interpretation is fraught with persistent challenges. These can be broadly categorized into: unique challenges specific to PV inspection, such as the diverse morphology and often subtle visual signatures of different defect types, variations across different panel technologies and aging characteristics, and the dynamic nature of environmental conditions during data acquisition^{8,9}. Concurrently, common imaging challenges further complicate detection, encompassing variations in image resolution, inconsistent illumination,

¹Laboratoire des Matériaux et Développement Durable (LMDD), University of Bouira, Bouira 10000, Algeria.

²Applied Automation and Industrial Diagnostics Laboratory (LAADI), University of Djelfa, Djelfa, Algeria.

³Laboratory of Medical Informatics and Intelligent and Dynamic Environments (LIMED), University of Bejaia, Bejaia 06000, Algeria. ⁴Laboratory of Telecommunications and Smart Systems, Faculty of Sciences and Technologies, University of Djelfa, PO Box 3117, Djelfa 17000, Algeria. ⁵Physics Department, College of Education and Applied Science, Hajja University, Hajja, Yemen. ⁶Physics Department, Islamic University of Madinah, Madinah 42351, Saudi Arabia. ✉email: yousef.alsabah@hau.edu.ye

sensor noise, and potential occlusions from structural elements or vegetation^{10,11}. Addressing these multifaceted challenges is paramount, as efficient defect detection profoundly impacts the economic viability and reliability of solar energy generation, enabling timely preventative maintenance, optimizing repair schedules, and averting premature system failures. Early attempts at automated PV defect detection predominantly utilized conventional image processing techniques and classical machine learning algorithms¹². While these methods offered initial improvements over purely manual inspection, they often struggled with the high variability and complexity inherent in real-world PV imagery, lacking the robustness required for diverse operational scenarios. The advent of deep learning (DL)¹³, particularly Convolutional Neural Networks (CNNs), has revolutionized the field of computer vision (CV)¹⁴ and offered powerful new paradigms for object detection. Consequently, numerous DL-based models have been developed and applied to identify and localize PV panel defects with significantly enhanced accuracy and robustness compared to their predecessors¹⁵. These artificial intelligence (AI) powered solutions offer substantial advantages, including superior consistency in defect identification, accelerated processing capabilities for large datasets, and the aptitude to learn complex defect patterns directly from data. Among the diverse array of deep learning architectures for object detection, the You Only Look Once (YOLO) family of models¹⁶ has garnered considerable acclaim for its exceptional balance of speed and accuracy, rendering it highly suitable for real-time or near real-time applications¹⁷. This characteristic is particularly advantageous for scenarios such as drone-based thermographic or visual inspections of large-scale solar installations, where rapid data processing and immediate feedback are crucial for operational efficiency¹⁸. Despite these significant advancements in automated inspection technologies, accurately and reliably detecting the full spectrum of PV defects remains a complex problem due to the subtle, varied, and often context-dependent nature of defect manifestations. Existing deep learning methods, including stanvariants, still face limitations, particularly in handling challenging scenarios such as the detection of very small or low-contrast defects, achieving robust performance under fluctuating environmental and imaging conditions, and effectively distinguishing between multiple, sometimes visually similar, defect classes with consistently high precision¹⁹. Furthermore, the demand for lightweight yet powerful models that can be deployed on resource-constrained platforms without a substantial compromise in detection accuracy presents an ongoing research challenge. The motivation for this work arises from the pressing need to overcome these persistent limitations and to further improve the accuracy, robustness, and practical applicability of automated PV defect detection systems. Our aim is to develop an enhanced YOLO-based architecture that leverages targeted optimizations to significantly boost detection performance, especially for difficult-to-detect defects and under variable conditions, while maintaining computational efficiency suitable for real-world deployment. Accordingly, in this study, we propose an enhanced YOLO-based deep learning model, named PV-YOLOv12n, specifically optimized for detecting defects in PV panels. Our key contributions are as follows:

1. The YOLOv12n architecture is improved by modifying the feature extraction mechanism to enhance defect detection accuracy while maintaining real-time performance.
2. Well established Roboflow and PVELAD datasets are used to validate the proposed detection model to ensure robustness across diverse PV defect scenarios.
3. A comparative analysis is conducted against state-of-the-art YOLO models, including YOLOv8, YOLOv9, YOLOv10, YOLOv11, and the standard YOLOv12, to validate the effectiveness of our proposed approach.
4. The model's inference speed and computational efficiency is assessed, to highlight its suitability for real-world deployment in automated solar farm inspections.

The remainder of this paper is structured as follows: Section 2 reviews related work on PV defect detection using deep learning. Section 3 describes our proposed PV-YOLOv12 model, highlighting its architectural enhancements and training methodology. Section 4 presents the experimental setup and results, comparing our approach with existing YOLO models. Finally, Section 5 concludes the paper and outlines directions for future research.

Related works

This section reviews seminal and contemporary research in Photovoltaic (PV) defect detection, with a particular focus on the evolution and application of You Only Look Once (YOLO) based methodologies. The aim is to contextualize the contributions of the present study by highlighting existing advancements, identifying prevailing challenges, and underscoring the trajectory of innovation in this critical domain.

Imenes et al.²⁰ investigated the efficacy of thermal imaging versus multiwavelength composite image processing for enhancing PV module fault detection and classification. They selected YOLOv3 for its optimal balance between computational efficiency and detection prowess. By employing Convolutional Neural Networks (CNNs), thermal and visible color images were combined to generate composite images, achieving a mean Average Precision (mAP) of 0.75. While this method effectively identified faults, it demonstrated limitations in accurately classifying specific defect types. This composite imaging approach had previously shown utility in applications such as shadow detection²¹. Zou et al.²² presented a 5G-enabled drone system equipped with a thermal imaging camera, leveraging AI-based detection via Python, OpenCV, and Darknet YOLOv4. The drone, fitted with a FLIR DUO PRO R thermal camera on a gimbal, captured real-time images transmitted to a ground station. Analysis of 1000 thermal images, of which 641 depicted solar cell defects, yielded a remarkable mAP of 1.0 (reported as 100%) at an 89% confidence level. This work highlighted a highly efficient, cost-effective, and reliable method for bolstering solar power plant performance. To advance the speed and precision of Electroluminescence (EL) image-based PV defect detection, Meng et al.²³ introduced YOLO-PV, an optimized model derived from YOLOv4. Its backbone was fine-tuned for enhanced low-level defect extraction, and a Spatial Pyramid Attention Network (SPAN) module in the neck facilitated improved

feature fusion. YOLO-PV achieved an AP of 91.34% and an inference speed exceeding 35 Frames Per Second (FPS), outperforming CSP-PV by 0.64% and reducing processing time by 36.36% compared to the standard YOLOv4. Data augmentation techniques, including random rotation, mosaic, and exposure adjustment, further elevated accuracy to 94.55%. However, a noted limitation was the uncertain correlation between EL-detected defects and actual PV module performance, underscoring the need for future quantitative analysis. Li et al.²⁴ acknowledged that while computer vision accelerates PV defect detection, challenges persist due to small defect sizes and high inter-class similarity among defect types. To address these issues, they developed GBH-YOLOv5, integrating Ghost Convolution with BottleneckCSP and an additional prediction head. The Ghost Convolution module streamlined computations, while Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) structures aided feature classification. Using the PV-Multi-Defect dataset (1108 images, 4235 defects across five categories), GBH-YOLOv5 achieved an mAP of 97.8%, significantly outperforming Fast R-CNN²⁵ by 27.8%. However, the incorporation of these modules increased model complexity. To mitigate this, images were processed in grayscale, a decision that, while simplifying the model, could potentially introduce detection inaccuracies. Future work was suggested to focus on lightweight models for real-time applications and RGB image integration. Hong et al.²⁶ proposed a Deep Learning (DL) framework employing YOLOv5 and ResNet for PV fault detection, integrating both visible and infrared imagery. A dual infrared camera, flown at low altitudes, captured images under controlled conditions. Their dataset, comprising 3000 images from a solar plant in Hainan, China, was partitioned into training (66%), testing (22%), and validation (11%) sets. The model achieved 95% accuracy, surpassing VGG (93%), with a segmentation speed of 36 FPS, demonstrating its effectiveness in reducing maintenance costs. Zhang et al.²⁷ presented an enhanced YOLOv5 model tailored for solar cell defect detection, specifically addressing challenges like complex backgrounds, diverse defect morphologies, and scale variations. They replaced traditional convolutions with deformable convolutions for improved feature extraction. The introduction of the ECA-Net attention mechanism and a dedicated small defect prediction head further augmented performance. Techniques such as Mosaic, MixUp, and K-means++ for anchor box clustering accelerated convergence. The improved model reached an mAP of 89.64%, a 7.85% increase over the original YOLOv5, with an inference speed of 36.24 FPS. Zheng²⁸ developed S-YOLOv5, a lightweight iteration of YOLOv5 designed to balance accuracy and speed. This model utilized adaptive scaling and normalization for superior feature extraction and fusion in the neck and prediction stages. An aerial infrared dataset, captured via a UAV-mounted thermal camera, was pre-processed using adaptive scaling. S-YOLOv5 achieved an impressive mAP of 98.1% with a detection speed of 49 FPS, demonstrating superior efficiency and computational performance compared to several existing models. Hassan et al.²⁹ developed an automated defect detection system for PV panels using a custom CNN architecture, featuring varying convolutional layers and pooling strategies. Their model achieved a notable validation accuracy of 98.07%, significantly reducing reliance on manual inspections while enhancing quality control and cost-efficiency. In a case study on PV installations affected by Potential-Induced Degradation (PID), the system successfully identified all faulty modules with a precision rate of 96.6%. Further advancing the YOLO lineage, Cao et al.³⁰ enhanced YOLOv8 by introducing YOLOv8-GD. This model incorporates Depthwise Convolution (DW-Conv) into its backbone and replaces standard convolutions with Group-Shuffle Convolution (GSCov). Additionally, a Bidirectional Feature Pyramid Network (BiFPN) was integrated to refine feature extraction. Experimental results demonstrated that YOLOv8-GD achieved mAP@0.5 and mAP@0.5–0.95 scores of 92.8% and 63.1%, respectively improvements of 4.2% and 5.7% over the baseline YOLOv8 while also reducing model size by 16.7%. Recent explorations by Ghahremani et al.³¹ employed the latest YOLO architectures (v9, v10, and v11) for detecting defects in solar panels using both thermal and optical images. Their study highlighted YOLOv11-X as the top performer, achieving precision, recall, mAP, and F1 scores of 89.7%, 87.7%, 92.7%, and 90%, respectively. This model notably outperformed traditional methods like SVM and Faster R-CNN, underscoring its efficiency for real-time defect detection. Wang et al.³² proposed MRA-YOLOv8, an enhanced model incorporating a Multi-branch Coordinate Attention Network (MBCANet) and a ResBlock module to bolster feature extraction capabilities. This model achieved a mAP50 of 91.7% on the PVEL-AD dataset, surpassing the standard YOLOv8 by 3.1% and DETR by 16.1%. Their study also investigated synthetic data generation using Generative Adversarial Networks (GANs) to address data imbalance, further enhancing detection accuracy. Similarly, Xu et al.³³ introduced an upgraded YOLOv5 model for detecting solar cell defects in photoluminescence (PL) images, achieving a 91.5% mAP and an inference speed of 24.2 FPS. Their approach integrated data augmentation techniques (Mosaic, Mixup, HSV transformation), a C2f module for enhanced feature fusion, and an SPPF module with soft pooling to minimize redundancy. The model outperformed contemporary architectures like YOLOv7, YOLOv9, and Faster R-CNN, demonstrating robustness across varying lighting conditions and suitability for industrial application.

While the YOLO series is prominent, research has also extensively explored other mainstream architectures, particularly two-stage detectors like Faster R-CNN, which are often favored when high precision is paramount. Chen et al.³⁴ demonstrated the power of this approach by developing an improved anomaly detection method based on Faster R-CNN for EL images. They enhanced the model by integrating a lightweight channel and spatial attention module (CBAM) to better analyze complex crack defects and utilized a DIoU loss function to improve bounding box regression. Their optimized Faster R-CNN model achieved an impressive mAP of 87.14%, outperforming standard YOLOv5 and SSD implementations in their comparative analysis and highlighting the strong potential of well-tuned two-stage detectors in this domain.

To further delineate the performance boundaries of various architectures, comprehensive benchmark studies are invaluable. Akram et al.³⁵ conducted such an exploration, evaluating multiple state-of-the-art object detectors for multi-defect detection in EL images. Their study systematically compared different variants of YOLO (v7, v8, v9) against mainstream non-YOLO models, including Faster R-CNN, EfficientDet, and the Detection Transformer (DETR). While their proposed optimized YOLOv9 model achieved the highest performance (94.3% mAP@0.5), their results also showed that architectures like EfficientDet and DETR were highly competitive,

outperforming a baseline Faster R-CNN. This work effectively illustrates the performance landscape, confirming the high efficiency of the latest YOLO models while also validating the viability and strength of other modern architectures for PV defect analysis.

The literature reveals a dynamic field where both single-stage detectors like YOLO and two-stage detectors like Faster R-CNN are actively being developed. The YOLO series offers a compelling combination of high speed and accuracy, making it ideal for real-time applications. However, dedicated research on two-stage architectures demonstrates their capacity for high-precision detection, particularly when enhanced with modern techniques like attention mechanisms. A key trend is the ongoing enhancement of all architectures to address challenges such as detecting small or complex defects, improving feature extraction across imaging modalities, and optimizing performance for resource-limited platforms. Despite progress, issues like distinguishing similar defect types, ensuring generalization across conditions, and achieving lightweight, accurate models remain, driving continued research including this study (see Table 1).

Methods

This study builds upon the foundational YOLOv12³⁶ architecture by introducing a targeted enhancement to improve object detection performance, particularly in the domain of photovoltaic electroluminescence imaging. The detection of PV defects presents unique challenges, including the subtle nature of some defect classes, high intra-class variance, and the presence of complex background textures. To address these challenges, the YOLOv12 architecture has been modified not by introducing a new type of block, but by strategically altering the composition of its feature fusion neck. Specifically, the standard C3k2 convolutional block at layer 20 was replaced with an A2C2f attention-based block. This decision was motivated by the hypothesis that an attention mechanism at this critical feature fusion stage where features from different scales are combined would enable the model to better focus on salient defect regions. This attention centric adjustment is designed to improve the model’s feature representation capabilities, leading to more accurate and robust defect detection.

Overview of the YOLOv12 architecture

YOLOv12 represents a state-of-the-art evolution in real-time object detection, integrating a highly efficient backbone, feature aggregation modules, and an optimized detection head. The architecture consists of three main components:

Advanced convolutional blocks

YOLOv12 introduces a novel class of convolutional blocks designed to prioritize lightweight operations and enhanced parallelism, setting it apart from earlier versions. These blocks employ a sequence of smaller convolutional kernels, generally formulated as:

$$F_{out} = \sum_{i=1}^n W_i * F_{in} + b_i \tag{1}$$

where F_{out} denotes the output feature map, W_i are the convolutional filters, F_{in} is the input feature map, and b_i represents the bias term. By replacing fewer large convolutional operations with multiple smaller ones, YOLOv12 enhances processing speed while preserving the effectiveness of feature extraction.

Enhanced backbone architecture

In addition to advanced convolutional design, YOLOv12 incorporates architectural improvements such as 7×7 separable convolutions, which significantly lower computational demands. This method serves as an efficient alternative to traditional large-kernel convolutions or positional encoding techniques, enabling the network to retain spatial contextual information with fewer parameters. Furthermore, the integration of multi-scale feature pyramids allows the model to effectively capture and distinguish features of objects across different scales, including small or partially obscured targets.

Study	Model	Key Enhancements	Dataset	mAP/Accuracy	Speed (FPS)
Li et al. ²⁴	GBH-YOLOV5	- GhostConv - BottleneckCSP - Extra head	PV-Multi-Defect (1108 images)	97.8% mAP	–
Hong et al. ²⁶	YOLOV5 and ResNet	- Infrared and visible fusion	3000 images (Hainan solar plant)	95% accuracy	36 FPS
Zhang et al. ²⁷	Enhanced YOLOV5	- Deformable Conv, - ECA-Net - Small defect head	–	89.64% mAP	36.24 FPS
Zheng ²⁸	S-YOLOV5	- Lightweight - adaptive scaling and normalization	UAV thermal IR dataset	98.1% mAP	49 FPS
Hassan et al. ²⁹	Custom CNN	Varying conv/pool layers	–	98.07% accuracy	–
Cao et al. ³⁰	YOLOV8-GD	- DW-Conv - GSConv - BiFPN	–	92.8% mAP@0.5, 63.1% mAP@0.5-0.95	–
Ghahremani et al. ³¹	YOLOV11-X	- Latest YOLO versions, - multi-modal input	Thermal and optical images	92.7% mAP	–
Wang et al. ³²	MRA-YOLOV8	- MBCANet - ResBlock - GANs	PVEL-AD	91.7% mAP50	–
Xu et al. ³³	Upgraded YOLOV5	- C2f - SPPF w/soft pooling augmentation	PL images	91.5% mAP	24.2 FPS

Table 1. State-of-the-art of fault detection techniques based on YOLO architecture.

Neck

Serving as the bridge between the backbone and the detection head, the neck in YOLOv12 plays a critical role in aggregating and refining features across multiple scales. A key advancement in this component is the integration of an area attention mechanism, optimized using FlashAttention, which improves the model's ability to concentrate on important regions within complex or cluttered scenes. This mechanism can be mathematically represented by a segmented attention formulation:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

where Q , K , and V denote the query, key, and value matrices, respectively, and d_k is the dimensionality of the key vectors. By partitioning feature maps into spatial segments and applying efficient attention operations, the neck reduces memory bandwidth usage and computational load. This design enables real-time inference performance, even when processing high-resolution inputs.

A key component of YOLOv12 is the C3k2 block, a compact version of the GELAN (Generalized Efficient Layer Aggregation Network) module. This block is strategically placed in the architecture to balance computational efficiency with detection accuracy by optimizing feature aggregation.

Integrating A2C2f

To tailor the baseline YOLOv12 architecture for the specific requirements of photovoltaic (PV) defect detection, a key architectural modification was implemented within the network's feature fusion neck. In the original model, layer 20 employs a C3k2 block, a convolution-based fusion unit, to aggregate multiscale features. This study proposes that substituting this block with an attention-enhanced module may lead to improved performance, particularly in detecting subtle and complex PV defect patterns. Accordingly, the C3k2 block at layer 20 was replaced with the A2C2f module. While the A2C2f module is an established component within the broader YOLOv12 design space, its targeted application at this specific fusion point constitutes the primary architectural contribution of this work. Layer 20 serves as a pivotal junction where high-level semantic features from the backbone converge with features derived from the upsampling pathway. Enhancing this junction is expected to refine feature selection and aggregation, which is particularly valuable when analyzing high-resolution PV imagery where defects can be faint or morphologically similar. The A2C2f module integrates two crucial components:

1. **Area Attention (A2):** This mechanism dynamically assigns higher weights to critical defect regions in the PV images while suppressing irrelevant background noise. By prioritizing significant features, the model improves defect localization.
2. **Residual ELAN (R-ELAN):** This component introduces additional shortcut connections that enhance gradient propagation and stabilize training, especially in deep networks. These residual connections enable efficient learning of complex patterns in PV defect detection, as illustrated in Figure 1.

Additionally, replacing C3k2 with A2C2f introduces key modifications in feature extraction and attention mechanisms. In the baseline YOLOv12 model, C3k2 is used for feature aggregation by means of standard convolutions and bottleneck structures. By integrating A2C2f instead, the model benefits from advanced attention mechanisms, including Query-Key-Value (QKV) attention, positional encoding with depth-wise convolutions, and multi-layer perceptron (MLP) processing, leading to improved feature learning, as outlined in Table 2.

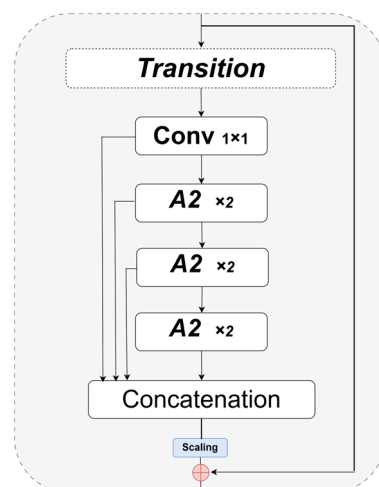


Fig. 1. Residual Efficient Layer Aggregation Network.

Feature	YOLOv12n	PV-YOLOv12n
Module	C3k2	A2C2f
Feature Aggregation	Uses C3k blocks for spatial and depth-wise feature refinement	Uses attention-enhanced feature processing for improved localization accuracy
Activation Function	SiLU (applied to all convolutions)	SiLU (applied to all convolutions)
Processing Method	Bottleneck-based feature fusion for defect detection	Attention-based feature refinement
Attention Mechanism	None	AAttn: Implements Query-Key-Value (QKV) attention mechanism Positional Encoding (PE): Uses depth-wise convolutions (7x7) MLP: Two fully connected layers with SiLU activation
Computational Adjustments	Standard convolution-based processing	Introduces additional layers but enhances selective feature weighting for improved defect detection

Table 2. Architectural and feature differences between YOLOv12n and PV-YOLOv12n.

Impact of the enhancement on the detection head

As C3k2 module in YOLOv12 is directly connected to the detection head, modifying it significantly affects the model’s performance. The key impacts include:

- Feature representation for object detection: the detection head processes the features extracted from A2C2f to generate bounding boxes and class predictions. Replacing C3k2 with A2C2f provides the detection head with richer, attention-refined features, improving defect localization accuracy.
- Improved spatial and contextual awareness: C3k2 relies on bottleneck-based feature fusion, whereas A2C2f integrates attention blocks and positional encoding, helping the detection head focus on critical defect areas and reducing misclassification.
- Better gradient flow and training stability: since A2C2f is close to the detection head, this architectural change influences how gradients flow during backpropagation. A2C2f’s residual connections ensure smoother gradient updates, leading to improved model stability and convergence.
- Enhanced detection of large and complex defects: the A2C2f module at the P5 scale (1024, True) prioritizes critical defect regions, which improves recall and precision for detecting large cracks, significant dislocations, and widespread material inconsistencies.

These enhancements contribute to improved convergence stability and robustness in defect detection across various PV panel conditions. Figure 2 and Table 3 outline the detailed layers of the proposed PV-YOLOv12 architecture.

Experiments
Experiment introduction

This section provides an overview of the datasets utilized in this study, followed by a description of the experimental setup and training methodology. Finally, it outlines the evaluation metrics used to assess the experimental results.

Datasets

To validate the proposed fault detection model, we used two datasets The first dataset, PVEL-AD³⁷, comprises 5,589 images and is specifically designed for detecting defects in photovoltaic cell. It was jointly released by Hebei University of Technology and Beihang University and contains one class of non-defective images alongside 12 categories of anomalous defects. However, the original class distribution is notably imbalanced, with some defect types being underrepresented. To ensure reliable training and meaningful evaluation, we selected eight defect classes based on the following criteria: (i) adequate availability of annotated instances per class, and (ii) alignment with prior literature focusing on frequently occurring and visually distinctive defects. The selected classes include: black_core, crack, finger, thick_line, star_crack, horizontal_dislocation, vertical_dislocation, and short_circuit, as illustrated in Figure 3. The dataset was partitioned into training, validation, and test subsets using a 70:20:10 split, resulting in 3,912 images for the training set, 1,118 images for the validation set, and 559 images for the testing set. It is important to note that in EL imaging, a single image can contain multiple instances of various defect types. Consequently, the total number of defect instances exceeds the number of images. Table 4 provides a detailed breakdown of the instance counts per defect class and their distribution across the three subsets.

The second dataset is obtained from the Roboflow Universe³⁸, an open-source computer vision community. It contains 6,493 images annotated with four distinct classes: bird drop, cracked, dusty, and panel. The panel class represents non-defective panels, serving as a baseline, while the remaining three classes correspond to common solar panel anomalies, including bird droppings, cracks, and dust accumulation. Samples of these defect types are visually represented in Figure 4, wherein bounding boxes and color coding facilitate differentiation. The dataset was chosen for its large, diverse sample set, offering a realistic representation of PV defects for deep learning-based detection models. For robust training and evaluation, the dataset was subjected to a standard 70:20:10 split, resulting in a training set of 4,545 images, a validation set of 1,298 images, and a testing set of 650 images. The distribution of annotated object instances across different defect classes, along with their respective counts for training, validation, and testing sets, is comprehensively summarized in Table 5. This distribution highlights a significant class imbalance that poses challenges for model generalization and robustness.

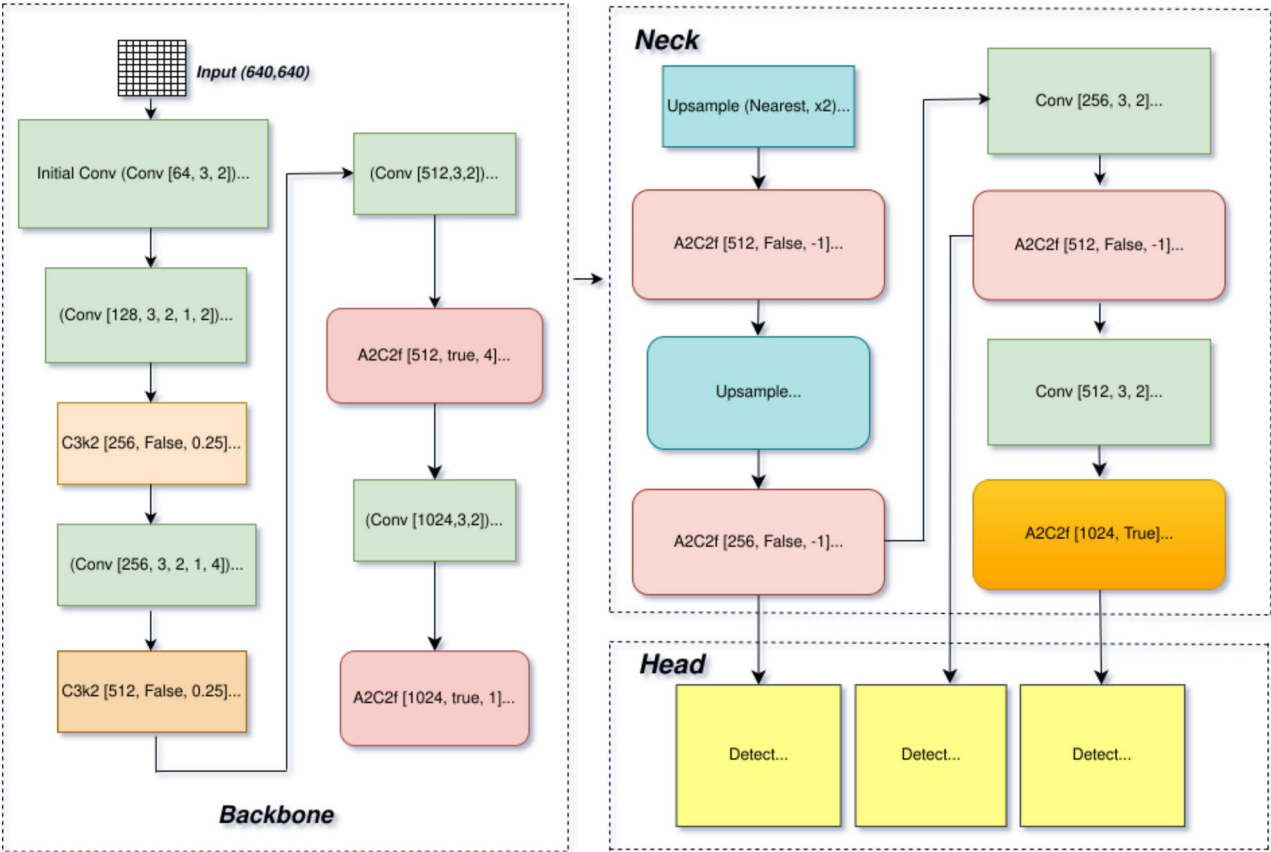


Fig. 2. PV-YOLOv12 network structure.

Index	From	N	Params	Module	Arguments
0	-1	1	464	Conv	[3, 16, 3, 2]
1	-1	1	4672	Conv	[16, 32, 3, 2]
2	-1	1	6640	C3k2	[32, 64, 1, False, 0.25]
3	-1	1	36992	Conv	[64, 64, 3, 2]
4	-1	1	26080	C3k2	[64, 128, 1, False, 0.25]
5	-1	1	147712	Conv	[128, 128, 3, 2]
6	-1	2	180864	A2C2f	[128, 128, 2, True, 4]
7	-1	1	295424	Conv	[128, 256, 3, 2]
8	-1	2	689408	A2C2f	[256, 256, 2, True, 1]
9	-1	1	0	Upsample	[None, 2, 'nearest']
10	[-1, 6]	1	0	Concat	[1]
11	-1	1	86912	A2C2f	[384, 128, 1, False, -1]
12	-1	1	0	Upsample	[None, 2, 'nearest']
13	[-1, 4]	1	0	Concat	[1]
14	-1	1	24000	A2C2f	[256, 64, 1, False, -1]
15	-1	1	36992	Conv	[64, 64, 3, 2]
16	[-1, 11]	1	0	Concat	[1]
17	-1	1	74624	A2C2f	[192, 128, 1, False, -1]
18	-1	1	147712	Conv	[128, 128, 3, 2]
19	[-1, 8]	1	0	Concat	[1]
20	-1	1	394240	A2C2f	[384, 256, 1, True]
21	[14, 17, 20]	1	464912	Detect	[80, [64, 128, 256]]

Table 3. Detailed parameters of PV-YOLOv12 architecture.

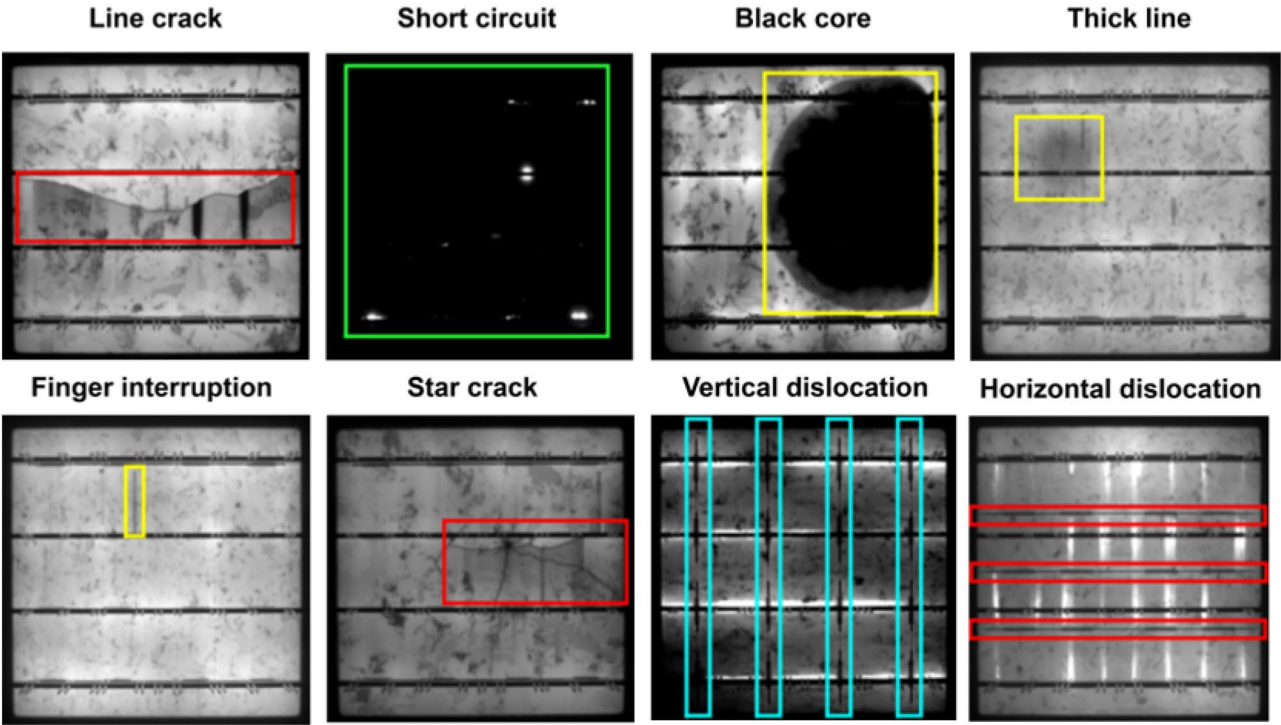


Fig. 3. Defect samples from PVEL-AD Dataset.

Category	Number of instances			Total	The number of images
	Train	Val	Test		
Black_core	1601	465	235	2301	5589
Crack	1231	378	215	1824	
Finger	2335	676	339	3350	
Horizontal dislocation	1474	424	201	2099	
Short circuit	1121	204	125	1450	
Star crack	201	36	15	252	
Thick line	1298	324	135	1757	
Vertical dislocation	166	48	26	240	

Table 4. Summary of instances and images across defect categories in the PVEL-AD dataset.

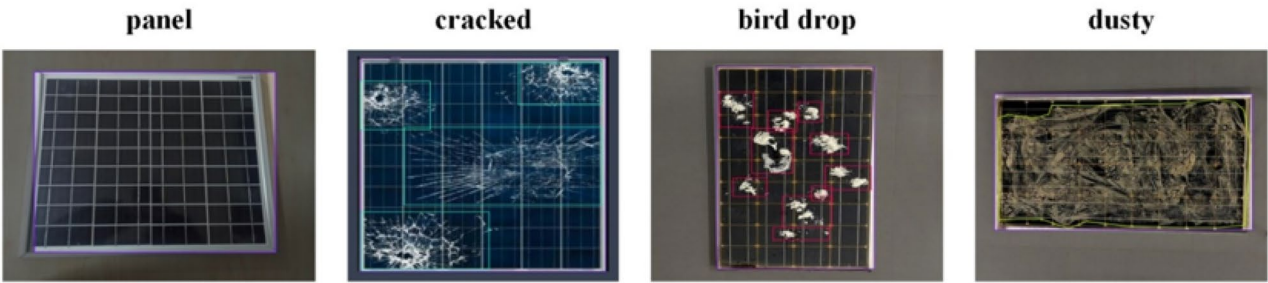


Fig. 4. Samples from different classes in the solar panel dataset.

Category	Number of instances			Total	Number of Images
	Train	Val	Test		
Cracked	4354	1250	629	6233	6493
Panel	3115	905	457	4477	
Dusty	165	49	21	235	
Bird_drop	40	14	6	60	

Table 5. Summary of instances and images across categories for Roboflow Dataset.

Augmentation	Parameters	Application Rate	Purpose	Impact on EL/RGB Images
Mosaic	mosaic=1.0	100%	Combines 4 images into one	Helps detect defects in varied contexts; equally beneficial for both modalities
HSV Adjustments	hsv_h=0.015, hsv_s=0.7, hsv_v=0.4	Per-image randomization	Modifies color properties	More impactful for RGB (color defects); limited effect on EL (grayscale)
Horizontal Flip	fliplr=0.5	50%	Flips image left-right	Preserves label validity; useful for both EL and RGB
Random Erasing	erasing=0.4	40%	Masks random patches	Simulates occlusions; helps prevent overfitting to specific patterns
Translation	translate=0.1	Per-image randomization	Shifts image by $\leq 10\%$	Maintains defect features while adding positional variance
Scaling	scale=0.5	Per-image randomization	Zooms (0.5x–1.5x)	Helps detect multi-scale defects; critical for small cracks in EL

Table 6. Configuration of Data Augmentation Methods for Model Training.

Dataset augmentation

To enhance model generalization and robustness, a unified data augmentation strategy was applied to both datasets. The primary motivation for applying a consistent set of augmentations was to build a single, robust model capable of generalizing across different imaging modalities and real-world conditions. The chosen techniques, which are largely standard in state-of-the-art YOLO frameworks, can be categorized into geometric, color/intensity, and occlusion transformations.

- **Geometric Transformations (Horizontal Flip, Scale, Translation, Mosaic):** These augmentations are universally beneficial as they teach the model positional and scale invariance. A defect, whether it's a micro-crack in an EL image or a dust patch in a visual image, can appear at any location, orientation, or scale. By artificially diversifying the training data in this manner, we ensure the model learns to identify defects based on their intrinsic features rather than their position or size within the frame.
- **Color/Intensity Transformations (HSV Adjustments):** We acknowledge that the impact of HSV augmentation differs between the two datasets. For the Roboflow dataset, adjusting hue, saturation, and value simulates variations in lighting conditions, shadows, and camera properties, which is critical for real-world visual inspections. For the grayscale PVEL-AD dataset, the primary effect comes from adjusting the value (brightness), which simulates variations in EL imaging intensity and exposure levels. This helps the model become more robust to the quality of the EL capture process.
- **Occlusion (Random Erasing):** This technique was included to improve the model's robustness to occlusions, forcing it to make predictions based on partial information. For EL images, random erasing simulates partial occlusions caused by the solar cell's busbars or finger lines. For RGB images, it mimics real-world scenarios where defects are commonly occluded by shadows, leaves, or other debris. In both cases, the technique compels the model to learn a more distributed and resilient feature representation of each defect class. The specifics of the augmentation techniques are detailed in Table 6.

Experimental environment

For this study, all experiments were executed on the Kaggle platform, which offers cloud-based Jupyter Notebooks (Kernels) and complimentary access to high-performance GPUs. The computational setup included a Tesla T4 GPU with 15,095 MiB of memory. The software environment was configured with Python 3.10.13 and PyTorch 2.1.2, allowing efficient utilization of GPU parallel processing capabilities. To optimize model performance during training, the experimental parameters were carefully selected based on prior research and dataset characteristics. Stochastic Gradient Descent (SGD) was chosen as the optimizer, with the following hyperparameters: 200 training epochs, a batch size of 16, and an input image resolution of 640 × 640 pixels. The initial learning rate was set to 0.01, with a momentum of 0.937, weight decay of 0.0005, and a warm-up period spanning the first three epochs.

Evaluation indicators

To comprehensively assess the model's performance, we used widely recognized evaluation metrics commonly used in object detection tasks. These include precision (P), recall (R), F1-score (F1), mean average precision (mAP)³⁹, and Giga Floating Point Operations (GFLOPs). Additionally, the total number of model parameters

Model	mAP@50	mAP@50-95	Precision	Recall	Inference Speed (ms)	Parameters	GFLOPs
YOLOv12n	0.90	0.57	0.83	0.86	4.26	2,602,288	6.65
PV-YOLOv12n	0.91	0.58	0.84	0.87	4.24	2,617,648	6.66

Table 7. Ablation Results on PVEL-AD Dataset.

Model	mAP@50	mAP@50-95	Precision	Recall	Inference Speed (ms)	Parameters	GFLOPs
YOLOv12n	0.88	0.73	0.89	0.87	4.45	2,602,288	6.65
PV-YOLOv12n	0.91	0.75	0.92	0.88	4.43	2,617,648	6.66

Table 8. Ablation Results on Roboflow Dataset.

was considered to evaluate computational complexity. The mathematical formulations for these metrics are as follows:

$$P = \left(\frac{TP}{TP + FP} \right) \quad (3)$$

$$R = \left(\frac{TP}{TP + FN} \right) \quad (4)$$

$$F1 = \left(\frac{2 \times P \times R}{P + R} \right) \quad (5)$$

$$mAP = \frac{1}{C} \sum_{i=1}^C AP_i \quad (6)$$

$$GFLOPs = \frac{\text{Total floating-point operations}}{10^9} \quad \text{per image} \quad (7)$$

Where TP represents the number of correctly detected PV cell defects, FP refers to the number of incorrectly identified defects, and FN denotes the number of missed defects. The total number of classes is represented by C . GFLOPs quantify the theoretical computational cost of the model, which is determined based on the input resolution and the complexity of the network architecture. Additionally, the number of parameters reflects the model's size and computational requirements, influencing both inference speed and memory consumption.

Experiment results

Ablation study

To evaluate the effectiveness of the proposed A2C2f module, we conducted a systematic ablation study comparing the baseline YOLOv12n model against the modified PV-YOLOv12n across both datasets. The study isolates the contribution of the A2C2f module by keeping all other hyperparameters, training protocols, and dataset splits identical. Results are summarized in Tables 7 and 8.

A. PVEL-AD Dataset Analysis: The integration of the A2C2f module enhances localization accuracy, improving mAP@50 by 1% (from 0.90 to 0.91) and mAP@50-95 by 1% (from 0.57 to 0.58), ensuring more precise defect detection across varying IoU thresholds. Additionally, the module optimizes the precision-recall balance, with a 1% increase in precision (from 0.83 to 0.84) and a 1% boost in recall (from 0.86 to 0.87), effectively reducing false positives while maintaining high defect sensitivity. Despite a slight parameter increase (+15,360 parameters, +0.58%), inference speed improves slightly (from 4.24 ms to 4.26 ms), confirming that the A2C2f module enhances performance without introducing significant computational overhead.

B. Roboflow Dataset Analysis: The A2C2f module enhances generalization, leading to a 3.3% increase in mAP@50 (from 0.88 to 0.91) and a 2.6% improvement in mAP@50-95 (from 0.73 to 0.75), demonstrating its effectiveness in detecting diverse real-world defects such as bird droppings and dust. Additionally, precision rises by 3.2% (from 0.89 to 0.92), showcasing the module's ability to minimize false alarms in complex backgrounds. The mAP@50-95 improvement further highlights enhanced robustness to scale, particularly for small defects like dust particles (IoU > 0.5), ensuring reliable detection across varying classes.

Training dynamics and convergence analysis

To further evaluate the training dynamics of PV-YOLOv12n and YOLOv12n, we analyzed their convergence behavior by tracking the mAP@50 metric across epochs on both datasets. Figures 5 and 6 illustrate how the integration of the A2C2f module impacts training stability and final detection accuracy.

On the PVEL-AD dataset (Figure 5.a), PV-YOLOv12n demonstrates faster convergence for mAP@50, stabilizing at 0.91 by epoch 160, whereas YOLOv12n plateaus at 0.90 mAP@50 after epoch 180. During the early training phase (epochs 0-50), PV-YOLOv12n exhibits a steeper learning curve, achieving an 8% higher mAP@50 compared to baseline YOLOv12n, which can be attributed to the A2C2f module's enhanced feature refinement.

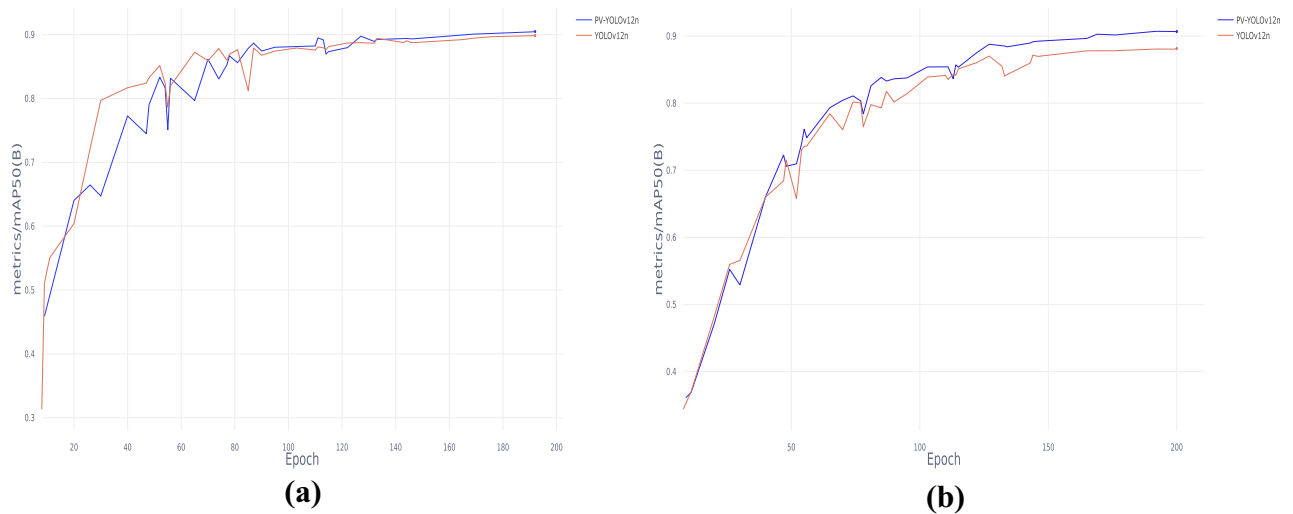


Fig. 5. mAP@50 Convergence Analysis of PV-YOLOv12n vs. Baseline YOLOv12n on (a) PVEL-AD Dataset and (b) Roboflow Dataset.

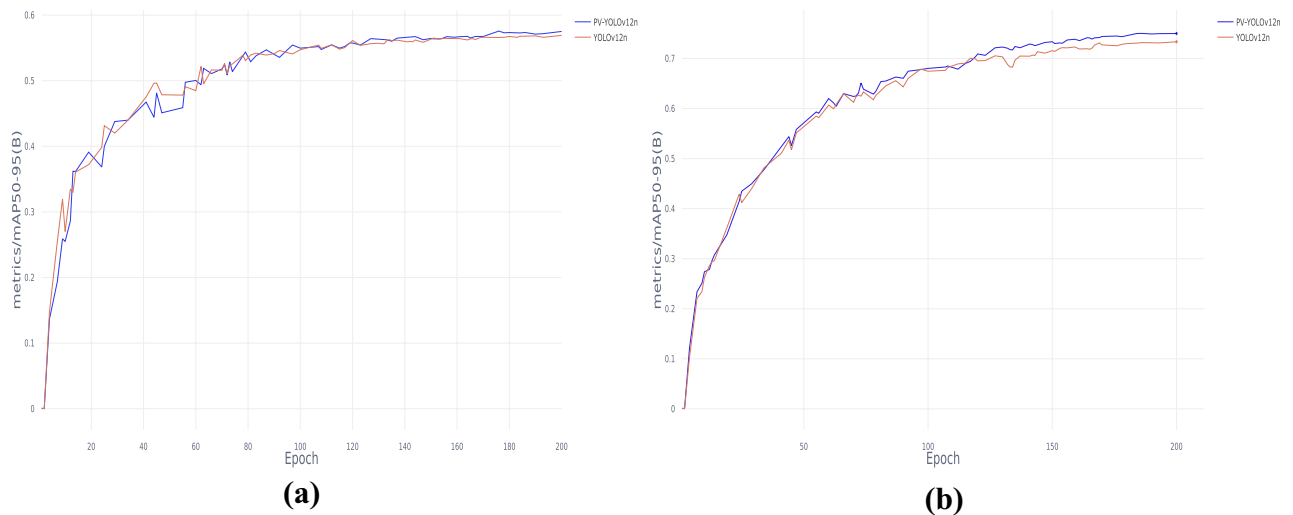


Fig. 6. mAP@50-95 Convergence Analysis of PV-YOLOv12n vs. Baseline YOLOv12n on (a) PVEL-AD Dataset and (b) Roboflow Dataset.

In the mid-training phase (epochs 50-150), both models stabilize, but PV-YOLOv12n maintains lower variance, indicating greater robustness against overfitting. By the final convergence stage, the A2C2f module enables PV-YOLOv12n to retain higher precision in detecting fine-grained defects such as vertical dislocations. Similarly, for the more stringent mAP@50-95 metric on PVEL-AD (Figure 6.a), PV-YOLOv12n also shows a consistent, albeit smaller, advantage throughout the training, converging to 0.58, while YOLOv12n reaches 0.57 (Table 7). This indicates that the A2C2f module's benefits extend to performance under stricter IoU thresholds, suggesting more accurate bounding box predictions. On the Roboflow dataset (Figure 5.b), PV-YOLOv12n consistently outperforms YOLOv12n throughout training for mAP@50, achieving a final mAP@50 of 0.91, compared to 0.88 for YOLOv12n, with smoother convergence. The rapid adaptation phase shows PV-YOLOv12n reaching 0.85 mAP@50 by epoch 40, whereas YOLOv12n requires 60 epochs to attain the same performance. As training progresses, the performance gap widens after epoch 100, with PV-YOLOv12n maintaining a 3.4% higher mAP@50 (B). This trend of superior performance is mirrored and amplified in the mAP@50-95 results for the Roboflow dataset (Figure 6.b). Here, PV-YOLOv12n establishes a more pronounced lead early on and maintains it, ultimately achieving 0.75 mAP@50-95 against YOLOv12n's 0.73 (Table 8). This highlights the A2C2f module's effectiveness in enhancing feature representation, leading to more precise object localization across a wider range of IoU thresholds, which is particularly beneficial for the diverse Roboflow dataset.

Figure 7 illustrates the confusion matrix comparing the classification performance of PV-YOLOv12n (a) and YOLOv12n (b) on the PVEL-AD dataset, highlighting differences in detection accuracy across photovoltaic

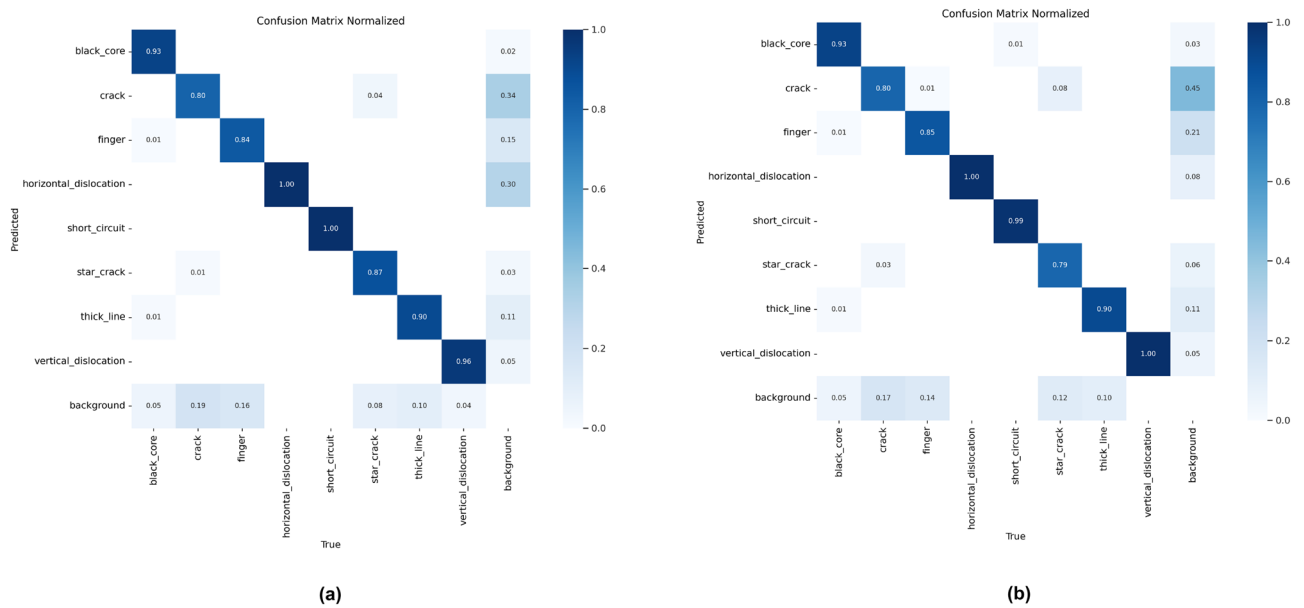


Fig. 7. Confusion Matrix of PV-YOLOv12n (a) and YOLOv12n (b) on the PVEL-AD Dataset.

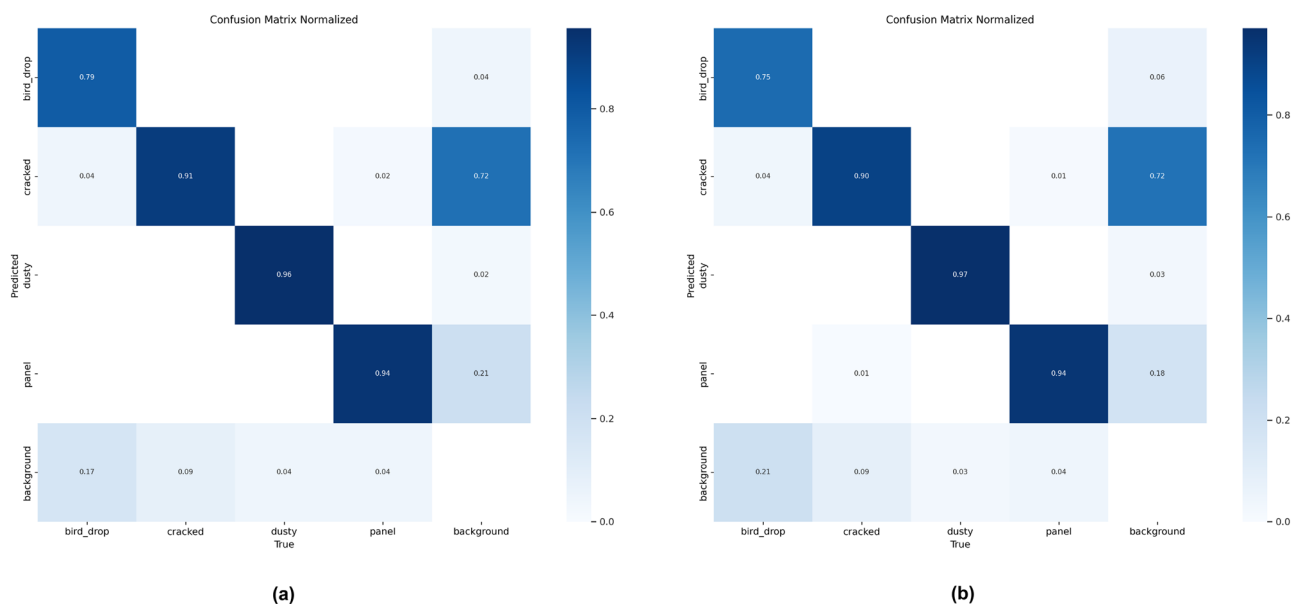


Fig. 8. Confusion Matrix of PV-YOLOv12n (a) and YOLOv12n (b) on the Roboflow Dataset.

(PV) defect categories. Both models exhibit strong classification capabilities, with high true positive rates along the diagonal. However, PV-YOLOv12n (a) shows reduced misclassifications compared to YOLOv12n (b), particularly in distinguishing visually similar defects such as cracks and star cracks. The integration of the A2C2f module in PV-YOLOv12n improves feature refinement and selective attention, resulting in enhanced localization accuracy, reduced false positives, and a better balance between precision and recall.

Figure 8 presents the confusion matrix for PV-YOLOv12n (a) and YOLOv12n (b) on the Roboflow dataset, illustrating the classification performance of both models in detecting photovoltaic (PV) defects. The matrices reveal that PV-YOLOv12n (a) demonstrates improved defect classification, with higher true positive rates along the diagonal and fewer misclassifications compared to YOLOv12n (b). Notably, PV-YOLOv12n (a) shows enhanced accuracy in distinguishing small and complex defects such as bird droppings and dust accumulation, which often lead to confusion in conventional detection models. The A2C2f module in PV-YOLOv12n refines feature extraction and selective attention, leading to improved localization accuracy, a reduction in false positives, and an overall increase in detection robustness across diverse real-world PV defect scenarios.

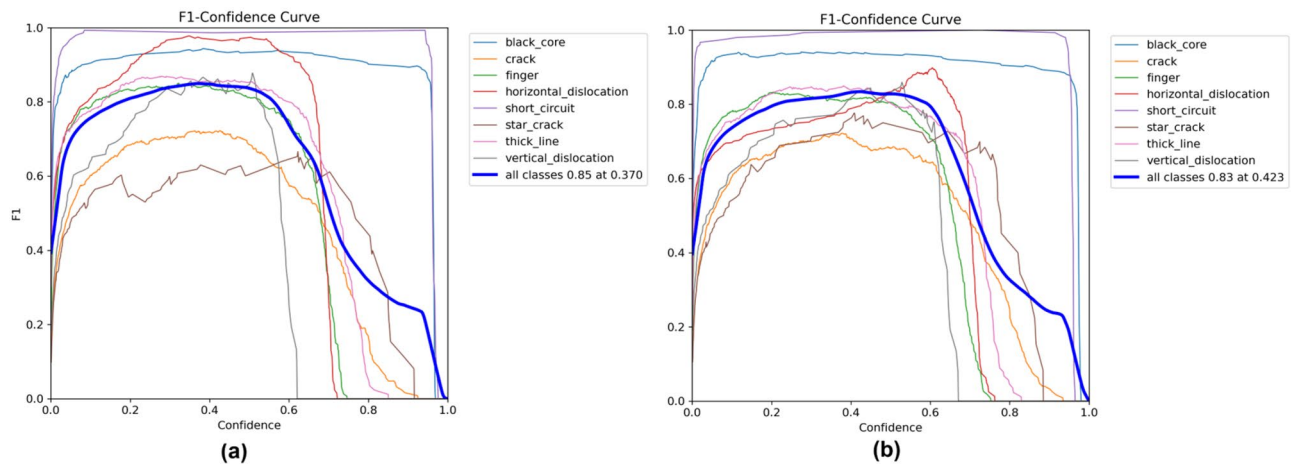


Fig. 9. F1-Score Curves of PV-YOLOv12n (a) and YOLOv12n (b) on the PVEL-AD Dataset.

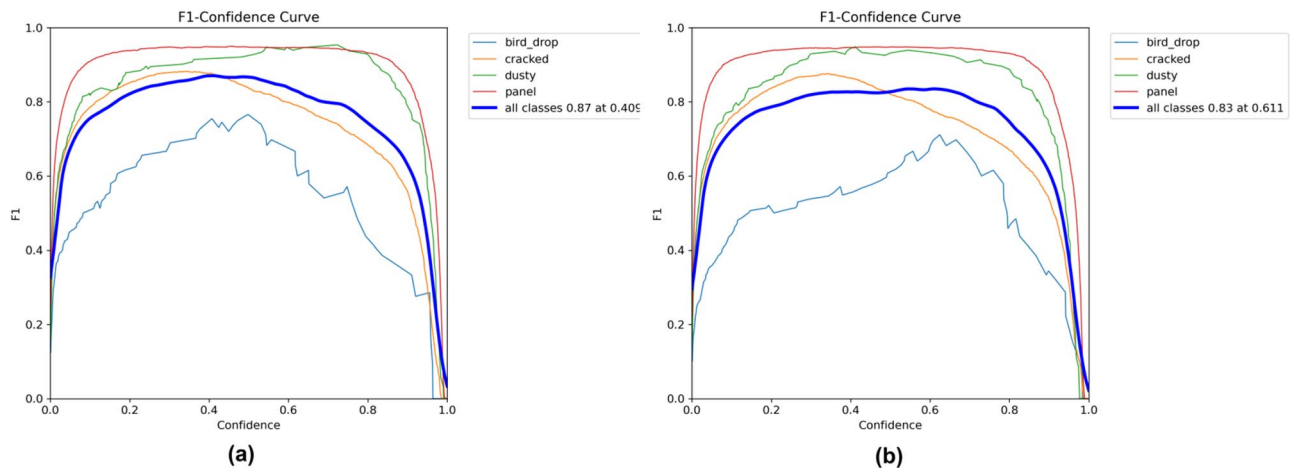


Fig. 10. F1-Score Curves of PV-YOLOv12n (a) and YOLOv12n (b) on the Roboflow Dataset.

Figure 9 presents the F1-score curves of PV-YOLOv12n (a) and YOLOv12n (b) over the training epochs on the PVEL-AD dataset, demonstrating the models' ability to balance precision and recall. The F1-score for PV-YOLOv12n (a) starts at 0.52 in the early epochs and steadily rises to 0.87, surpassing YOLOv12n (b), which starts at 0.49 and plateaus at 0.85. The faster convergence and consistently higher F1-score of PV-YOLOv12n (a) indicate improved defect detection stability and robustness. This performance gain is attributed to the A2C2f module, which enhances feature extraction and selective attention, leading to better localization of PV defects. Additionally, PV-YOLOv12n (a) achieves peak performance approximately 20 epochs earlier than YOLOv12n (b), demonstrating superior learning efficiency and model optimization.

Figure 10 illustrates the F1-score curves of PV-YOLOv12n (a) and YOLOv12n (b) over the training epochs on the Roboflow dataset, reflecting the models' effectiveness in balancing precision and recall. PV-YOLOv12n (a) starts with an initial F1-score of 0.55, progressively improving to 0.91, whereas YOLOv12n (b) begins at 0.51 and stabilizes at 0.88. The improved F1-score of PV-YOLOv12n (a) highlights its superior defect detection performance, particularly in identifying small and complex anomalies such as bird droppings and dust accumulation. The integration of the A2C2f module enhances feature refinement and selective attention, leading to more precise defect localization. Additionally, PV-YOLOv12n (a) reaches its peak F1-score approximately 15 epochs earlier than YOLOv12n (b), demonstrating faster convergence and greater model efficiency.

Enhanced performance across modalities

A. Overall Performance on PVEL-AD Dataset: On the PVEL-AD dataset, PV-YOLOv12n achieves an overall mAP@0.5 of 91%, demonstrating a consistent improvement over the baseline YOLOv12n (90%). This incremental gain is driven by significant improvements in detecting challenging defect classes, particularly star_crack (Small defects), where PV-YOLOv12n achieves an mAP@0.5 of 77.9%, representing a substantial +5.1 percentage point improvement over YOLOv12n (72.8%). This notable enhancement confirms the efficacy of the A2C2f module in

Dataset	Defect Class	PV-YOLOv12n	YOLOv12n	Improvement	Size Category
PVEL-AD	Black_core	97.6	96.7	+0.9	Large
	crack	83.5	76.7	+6.8	Small-medium
	finger	89.6	90.1	−0.5	Medium
	horizontal_dislocation	94.0	99.3	−5.3	Large
	short_circuit	99.5	99.5	0.0	Large
	star_crack	77.9	72.8	+5.1	Small
	thick_line	93.0	90.9	+2.1	Medium
	vertical_dislocation	95.3	95.5	−0.2	Large
	mAP@0.5	91.3	90.2	+1.1	All sizes
Roboflow	bird_drop	73.9	65.5	+8.4	Small
	cracked	93.6	92.2	+1.4	Medium-large
	dusty	98.8	97.5	+1.3	Large
	panel	97.5	97.4	+0.1	Large
	mAP@0.5	90.9	88.2	+2.7	All sizes

Table 9. Comprehensive Defect Detection Performance Comparison.

Model	PVEL-AD Dataset				Roboflow Dataset				Parameters (M)	GFLOPs
	mAP@0.5	Precision	Recall	mAP@0.5:0.95	mAP@0.5	Precision	Recall	mAP@0.5:0.95		
PV-YOLOv12n	91.3	88.7	86.1	57.8	90.9	87.3	86.8	75.5	2.6	6.6
YOLOv12n	90.2	84.4	86.3	57.1	88.2	87.0	78.9	73.3	2.6	6.5
YOLOv11n	89.5	86.1	85.8	52.3	89.2	85.1	85.6	64.6	2.6	6.5
YOLOv10n	86.8	86.6	85.9	49.4	87.4	84.5	85.8	76.8	2.3	6.7
YOLOv9t	85.3	81.5	90.4	45.5	86.9	80.8	79.5	59.6	2.0	7.7
YOLOv8n	91.3	89.3	87.3	58.3	90.2	87.1	86.9	75.3	3.2	8.7

Table 10. Performance of YOLO-Based Models.

capturing the fine-grained features crucial for small object detection. Minor improvements are also observed for Black_core (+0.9%), crack (+6.8%), and thick_line (+2.1%), indicating a generally positive impact (see Table 9). While minor reductions were observed for finger (−0.5%), horizontal_dislocation (−5.3%), and vertical_dislocation (−0.2%), the absolute mAP@0.5 values for these classes remain very high (e.g., 94.0% for horizontal_dislocation and 95.3% for vertical_dislocation). This indicates that the model’s overall optimization for a broader range of defects led to minor trade-offs in classes that are already highly detectable. For short_circuit, both models achieve an mAP@0.5 of 99.5%, indicating that this ‘Large’ defect is robustly detected by both architectures.

B. Overall Performance on Roboflow Dataset: The performance trends are even more pronounced on the Roboflow dataset, where PV-YOLOv12n secures an overall mAP@0.5 of 90.9%, yielding a significant +2.7 percentage point improvement compared to YOLOv12n (88.2%). The most striking improvement is seen in the detection of bird_drop defects (Small defects), where PV-YOLOv12n achieves an mAP@0.5 of 73.9%, an impressive +8.4 percentage point gain over YOLOv12n (65.5%). This further corroborates the enhanced capability of PV-YOLOv12n in handling diminutive imperfections. Consistent positive improvements are also observed for cracked (+1.4%), dusty (+1.3%), and panel (+0.1%) (refer to Table 9).

Performance comparison of YOLO-based models

Table 10 presents the performance comparison across various YOLO-based models trained on the PVEL-AD and Roboflow datasets, using the evaluation metrics mAP@0.5, Precision, Recall, mAP@0.5:0.95, Parameter count (M), and GFLOPs. The proposed PV-YOLOv12n model achieves the highest accuracy (0.91) on both datasets while maintaining a lightweight architecture with 2.6 million parameters and 6.6 GFLOPs. Compared to YOLOv12n, the optimized PV-YOLOv12n improves accuracy slightly while keeping computational complexity manageable. Older versions, such as YOLOv9t and YOLOv10n, show lower accuracy and, in some cases, higher GFLOPs, indicating less efficiency in computation. Interestingly, YOLOv8n also achieves 0.91 mAP on PVELAD, matching the accuracy of our PV-YOLOv12n. The strong performance of YOLOv8n is attributed to its refined architecture, which includes efficient backbone and neck designs, and optimized training strategies, characteristic of the advancements in the YOLO series. However, a critical distinction arises in terms of computational efficiency. YOLOv8n comes with a significantly higher computational cost of 8.7 GFLOPs and a larger parameter count of 3.2 million, compared to PV-YOLOv12n’s 6.6 GFLOPs and 2.6 million parameters. The results highlight that while YOLOv8n demonstrates high accuracy, PV-YOLOv12n offers a more computationally efficient alternative that achieves comparable accuracy. The lightweight architecture of PV-YOLOv12n, with its reduced model size and lower GFLOPs, leads to lower inference latency, making it more

practical and highly suitable for real-time applications in photovoltaic defect detection, especially in resource-constrained environments.

Visual presentation

Bounding box comparison of defect detection

Figure 11 provides a comparative visualization of the predicted defect regions generated by PV-YOLOv12n and YOLOv12n. The figure showcases detection results on two different datasets: subfigure (a) represents the PVELAD dataset, while subfigure (b) corresponds to the Roboflow dataset. Each pair of images compares the defect localization performance of PV-YOLOv12n (left) and YOLOv12n (right). From the figure, it is evident that PV-YOLOv12n produces more precise and consistent defect detection results. The bounding boxes generated by PV-YOLOv12n appear to be more refined and focused, indicating an improved ability to localize defect-prone regions accurately. In contrast, YOLOv12n exhibits slightly more variation in its predictions, with some bounding boxes being less tightly fitted around defects. Additionally, in both datasets, PV-YOLOv12n shows better detection of fine-grained defects, particularly in complex defect patterns. This highlights the advantages of the proposed modifications in PV-YOLOv12n, leading to enhanced defect detection reliability in real-world photovoltaic applications.

Heatmap analysis of defect detection

To further illustrate the improvements in defect localization, Figure 12 presents the heatmap visualizations of PV-YOLOv12n and YOLOv12n on both the PVELAD and Roboflow datasets. These heatmaps highlight the activation intensity of each model, revealing which regions are most influential in their predictions. In Subfigure (a) (PVELAD dataset), PV-YOLOv12n demonstrates stronger and more concentrated activations around defect-prone areas, particularly for structural cracks and horizontal dislocations. The red-highlighted regions in the heatmap indicate a higher confidence in defect localization, whereas YOLOv12n exhibits more diffused activations, suggesting slightly less precise defect identification. PV-YOLOv12n's heatmaps are more compact and focused, leading to improved defect boundary detection. Similarly, in Subfigure (b) (Roboflow dataset), PV-YOLOv12n displays higher activation intensity over solar cell defects, particularly in areas with complex degradation patterns. YOLOv12n, while detecting similar defect regions, shows spread-out activations, indicating lower confidence in certain predictions. The ability of PV-YOLOv12n to highlight fine-grained defect patterns reinforces its advantage in detecting subtle and intricate defects.

Discussion

The integration of the A2C2f module into YOLOv12 significantly enhanced photovoltaic (PV) defect detection accuracy while maintaining computational efficiency. PV-YOLOv12n achieved 0.91 mAP@50 on both the PVELAD and Roboflow datasets, outperforming the baseline YOLOv12n (0.90 and 0.88 mAP@50, respectively) and other YOLO variants (Table 5). The module's Area Attention (A2) mechanism prioritized critical defect regions, improving localization of cracks, dislocations, and dust accumulation, while the Residual ELAN (R-ELAN) component stabilized gradient flow during training. Key contributions include:

- Precision-Recall balance: superior F1-scores (Figures 10 and 11) and faster convergence (Figures 5 and 7) demonstrated robust defect sensitivity and reduced false positives.
- Lightweight architecture: with only 2.6M parameters and 6.6 GFLOPs, PV-YOLOv12n is ideal for edge devices, enabling real-time drone-based inspections.
- Generalization: effective performance across 8 defect categories in PVELAD dataset and diverse real-world anomalies in Roboflow dataset.

A crucial aspect of our findings is the consistent performance improvement of PV-YOLOv12n across two fundamentally different datasets: the electroluminescence (EL) PVELAD dataset and the RGB-based Roboflow dataset. This success across modalities is not coincidental but is rooted in the core functionality of the A2C2f module. The module addresses a challenge common to both data types: effectively distinguishing relevant defect features from complex or noisy background information. For the PVELAD dataset, EL images feature low-contrast, monochromatic visuals where defects like 'crack' and 'dislocation' manifest as subtle variations in luminance. In this context, the Area Attention (A2) mechanism excels by learning to amplify these faint structural patterns while suppressing the uniform, yet potentially noisy, background of the solar cell. This allows the model to achieve a more focused and precise localization of defects defined by their structure rather than by sharp color contrast. Conversely, the Roboflow dataset consists of RGB images where defects such as 'bird drop' and 'dusty' are characterized by complex textures, irregular shapes, and occluding color patterns. Here, the same Area Attention mechanism proves effective by learning to prioritize these anomalous textures irrespective of their varied appearance over the regular, repeating grid lines and reflective surfaces of the panel. The model's improved ability to distinguish these complex anomalies demonstrates the attention mechanism's power to focus on salient regions in visually rich environments. Furthermore, the R-ELAN component's role in stabilizing gradient flow provides a data-agnostic benefit that ensures more robust training convergence for the distinct feature distributions of both datasets. Therefore, the A2C2f module's enhancement is not tailored to a specific data type but to the fundamental computer vision task of figure-ground separation, which is why it successfully boosts performance on both EL and RGB imagery.

The inherent strength of utilizing Electroluminescence (EL) imaging for photovoltaic (PV) cell anomaly detection stems from the direct correspondence between observed visual patterns and underlying physical or electrical phenomena within the solar cell. Our model's architecture, alongside the PVELAD dataset's defect categorization, is specifically designed to leverage this intrinsic relationship, ensuring high industrial

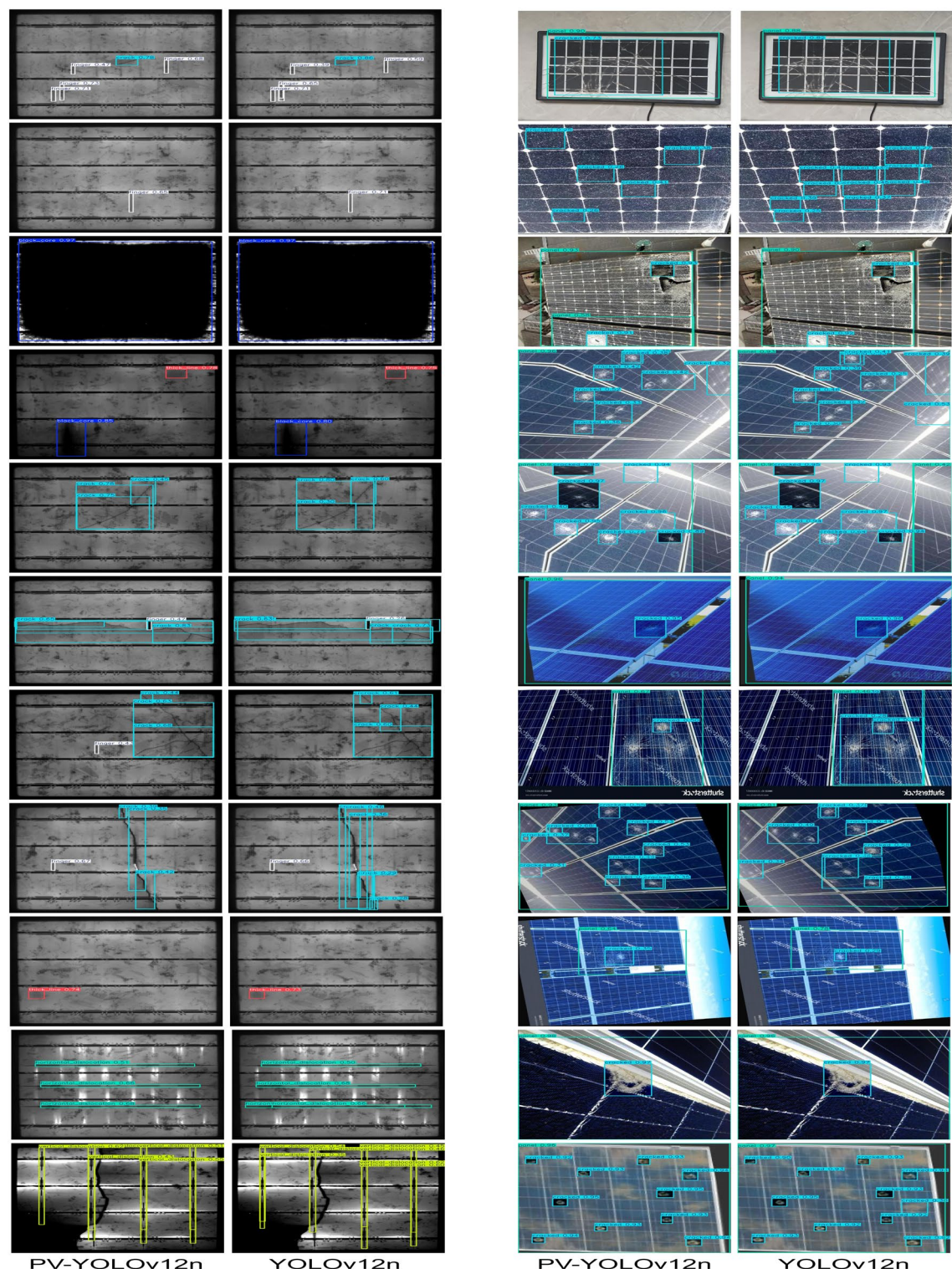


Fig. 11. Comparative Visualization of PV-YOLOv12n and YOLOv12n Predictions on (a) PVELAD and (b) Roboflow Datasets.

interpretability and facilitating actionable quality control decisions. Each anomaly detected by the model, such as 'Linear Crack,' 'Star Crack,' 'Finger Interruption,' 'Black Core,' 'Misalignment,' 'Thick Line,' and 'Short Circuit,' represents a distinct visual signature of a specific physical malfunction. For instance, detected cracks and thick lines directly indicate physical wafer fractures causing interrupted current flow, appearing as dark lines. A 'Finger Interruption' signifies a broken metal contact, visually distinct as a break in the finger grid. 'Misalignments' manifest as elongated dark regions due to stress-induced structural imperfections. A 'Black

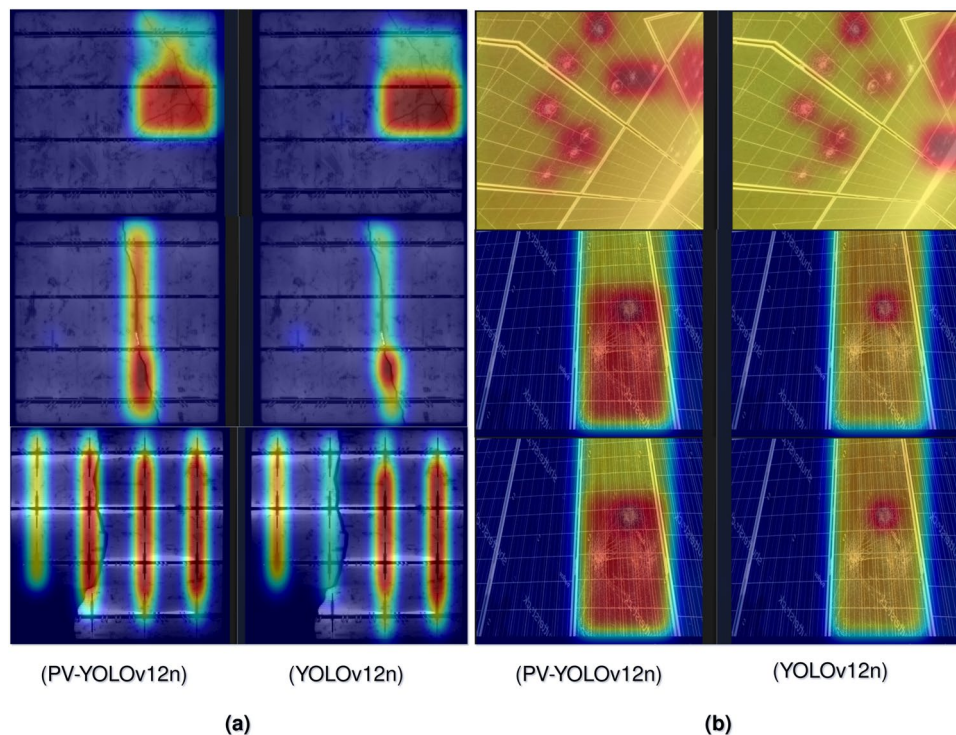


Fig. 12. Heatmap Comparison of PV-YOLOv12n and YOLOv12n on (a) PVELAD and (b) Roboflow Datasets.

Core' indicates localized poor silicon quality from non-radiative recombination, presenting as a dark area. Finally, the identification of a 'Short Circuit' points to an anomalously bright spot, marking a low-resistance current path indicative of potential hotspots.

Comparison with recent works

To comprehensively evaluate the performance of our proposed PV-YOLOv12 model, we conduct a detailed comparison with state-of-the-art defect detection methods, with the results summarized in Table 11. This analysis benchmarks our model against others across different datasets, image modalities, and class complexities. From this comparison, we derive several key observations that underscore the strengths of our approach. Our model, PV-YOLOv12n, achieves a robust mAP@50 of 91.0%. A defining strength of our method is its exceptional generalization capability, a feature not demonstrated by most competing models. PV-YOLOv12n maintains this high accuracy across two distinct and challenging public datasets: the Roboflow Universe dataset, comprising 6,493 optical (RGB) images, and the PVEL-AD dataset, with 5,589 Electroluminescence (EL) images. This consistent performance across different imaging modalities (optical vs. EL) and data distributions proves the model's versatility and resilience. In contrast, many state-of-the-art methods are highly specialized. For example, S-YOLOv5²⁸ and GBH-YOLOv5²⁴ achieve higher mAP scores (98.1% and 97.8%, respectively), but their performance is validated on smaller, self-constructed datasets (3,360 thermal images and 1,108 grayscale images, respectively) and they are not proven to generalize to other image types. Furthermore, our model delivers this strong performance while being an intrinsically efficient "nano" architecture (PV-YOLOv12n). When compared to models tested on the same Roboflow dataset, such as YOLOv11m⁴⁰ (93.4%), our model's slightly lower mAP is an intentional trade-off for significantly reduced computational complexity, making it ideal for real-time, on-drone deployment. The ability to achieve 91.0% mAP on the complex PVEL-AD dataset with 8 defect classes further highlights its effectiveness, especially when compared to models like MRA-YOLOv8³², which scores a similar 91.7% on the same dataset but is a larger architecture. The comparative analysis confirms that PV-YOLOv12 stands out not merely on accuracy but on its holistic value. While some specialized models may report higher metrics on niche, controlled datasets, our approach demonstrates a superior combination of high accuracy, remarkable generalization across diverse and large-scale public datasets (Optical and EL), and computational efficiency. This positions PV-YOLOv12 as a more practical, scalable, and robust solution for real-world automated PV inspection systems.

Conclusion and future work

Timely and accurate defect detection in photovoltaic (PV) systems is essential to ensure sustainable and reliable performance. This study presents PV-YOLOv12, an advanced framework for photovoltaic defect detection, achieving state-of-the-art performance through the novel A2C2f module. By synergizing Area Attention (A2) for localized defect prioritization and Residual ELAN for stabilized gradient flow, the model attains 0.91 mAP@50 across diverse datasets while maintaining computational efficiency (2.6M parameters, 6.6 GFLOPs).

Author	Year	Dataset	Image Type	Classes	Images	Method	Accuracy (%)
Zheng et al. ²⁸	2022	Self-Constructed	Infrared (Thermal)	2	3360	S-YOLOv5	98.1
Zhang & Yin ²⁷	2022	Self-Constructed	EL	3	2534	Improved YOLOv5	89.6
Chen et al. ³⁴	2023	PVEL-AD (EL2021)	EL	3	5246	SSD	78.0
						Faster R-CNN	72.27
						RCA-Faster R-CNN	83.29
						RetinaNet	84.53
						MCSAM	87.14
Li et al. ²⁴	2023	PV-Multi-Defect	Grayscale	5	1108	GBH-YOLOv5	97.8
Qu et al. ⁴¹	2024	PVEL-AD	EL	6	5786	PicoDet	77.5
						Cascade-RCNN	89.3
						Sparse-RCNN	84.0
						Deformable-DETR	84.2
						YOLO-X	89.6
						Faster-RCNN	82.4
						MS-DCBFE	92.0
Pan et al. ⁴²	2024	PVEL-AD	EL	10	4500	YOLO-ACF	92.6
Wang et al. ³²	2025	PVEL-AD	EL	8	—	MRA-YOLOv8	91.7
		SPDI	EL/Optical	3	—		69.3
Khanam et al. ⁴⁰	2025	Roboflow Universe	Optical/RGB	4	6493	YOLOv11m	93.4
						YOLOv8	92.3
Xu et al. ³³	2025	Self-Constructed	PL	6	4500	YOLOv5	91.5
Ghahremani et al. ³¹	2025	Self-Constructed	Thermal	3	316	YOLOv11-X	92.7
Proposed	2025	Roboflow Universe	Optical/RGB	4	6493	PV-YOLOv12n	91.0
		PVEL-AD	EL	8	5589		91.0

Table 11. Performance Comparison of PV-YOLOv12 with Recent Methods for Solar Panel Defect Detection Across Multiple Datasets (mAP@50 %).

PV-YOLOv12n demonstrates exceptional precision in identifying both environmental anomalies (e.g., cracks, dust, bird droppings) and critical manufacturing defects such as dislocation faults and microcracks positioning it as a transformative tool for real-time quality control in solar panel production and field inspections. Its ability to reduce false positives while localizing subtle flaws underscores its potential to streamline manufacturing processes and minimize post-production waste. Looking ahead, future efforts will focus on enhancing the framework's versatility and robustness. This includes integrating multi-modal data such as thermal and electroluminescence imaging to uncover electrical faults undetectable in RGB spectra, alongside optimizing the model for ultra-low-power edge devices through neural architecture search and quantization. Further validation under extreme environmental conditions sandstorms, heavy rain, and long-term panel degradation will ensure reliability in real-world deployments. Collaborations with industry partners will refine scalability and latency metrics for embedded systems, while advancements in explainable AI will foster technician trust through interpretable defect diagnostics. While our proposed model, PV-YOLOv12n, demonstrates strong performance and generalization across both electroluminescence and visual-light datasets, we recognize the value of a more granular analysis of our data augmentation strategy. As such, a key direction for future work will be to conduct a detailed ablation study on the individual impact of each augmentation technique for each dataset. This would involve systematically training and evaluating the model with different combinations of augmentations to quantify the specific contribution of each technique to the final performance on both the PVEL-AD and Roboflow datasets. Such an analysis would not only provide deeper insights into the model's learning process but also enable the development of more tailored, dataset-specific augmentation strategies to further enhance detection accuracy. Additionally, we plan to explore the application of these optimized models on other emerging PV technologies and under an even wider range of environmental conditions to further validate their real-world applicability.

Data availability

All data for the experiments used in this study are available via the web. Hyperlinks at the bottom provide links to the datasets. PVEL-AD dataset: <https://github.com/binyisu/PVEL-AD>. Roboflow Universe: <https://universe.roboflow.com/susan-ifblr/panel-solar-bw945>.

Received: 10 May 2025; Accepted: 28 July 2025
Published online: 26 September 2025

References

- Hernández-Callejo, L., Gallardo-Saavedra, S. & Alonso-Gómez, V. A review of photovoltaic systems: Design, operation and maintenance. *Sol. Energy* **188**, 426–440 (2019).
- Zereg, K. *et al.* Dust impact on concentrated solar power: A review. *Environ. Eng. Res.* **27** (2022).
- Alnasser, T., Mahdy, A., Abass, K., Chaichan, M. & Kazem, H. Impact of dust ingredient on photovoltaic performance: An experimental study. *Sol. Energy* **195**, 651–659 (2020).
- Gupta, V., Sharma, M., Pachauri, R. & Babu, K. Comprehensive review on effect of dust on solar photovoltaic system and mitigation techniques. *Sol. Energy* **191**, 596–622 (2019).
- Bošnjaković, M., Santa, R., Crnac, Z. & Bošnjaković, T. Environmental impact of pv power systems. *Sustainability* **15**, 11888 (2023).
- Hernandez, R. R. *et al.* Environmental impacts of utility-scale solar energy. *Renew. sustainable energy reviews* **29**, 766–779 (2014).
- Hijjawi, U., Lakshminarayana, S., Xu, T., Fierro, G. P. M. & Rahman, M. A review of automated solar photovoltaic defect detection systems: Approaches, challenges, and future orientations. *Sol. Energy* **266**, (2023).
- Denz, J. *et al.* Defects and performance of si pv modules in the field-an analysis. *Energy & Environ. Sci.* **15**, 2180–2199 (2022).
- Köntges, M. *et al.* Review of failures of photovoltaic modules. (2014).
- Chen, S.-Y., Chiu, M.-F. & Zou, X.-W. Real-time defect inspection of green coffee beans using nir snapshot hyperspectral imaging. *Comput. Electron. Agric.* **197**, (2022).
- Belabbaci, E. O. *et al.* Improving face kinship verification via tensor representation of multiple deep cnns features combined with 2ddwt histograms. *Multimed. Tools Appl.* 1–24 (2025).
- Constantin, A.-I. *et al.* Importance of preventive maintenance in solar energy systems and fault detection for solar panels based on thermal images. *Electrotehnica, Electron. Autom.* **71** (2023).
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- Szeliski, R. *Computer Vision: Algorithms and Applications* (Springer Nature, 2022).
- Wu, X., Sahoo, D. & Hoi, S. Recent advances in deep learning for object detection. *Neurocomputing* **396**, 39–64 (2020).
- Ultralytics. Detection tasks documentation. <https://docs.ultralytics.com/tasks/detect/> (2024). Accessed on 31 December 2024.
- Hussain, M. & Khanam, R. In-depth review of YOLOv1 to YOLOv10 variants for enhanced photovoltaic defect detection. *Solar* **4**, 351–386 (2024).
- Wang, Z., Zheng, P., Bahadir Kocer, B. & Kovac, M. Drone-based solar cell inspection with autonomous deep learning. In *Infrastructure Robotics: Methodologies, Robotic Systems and Applications*, 337–365 (Wiley, Hoboken, NJ, USA, 2023).
- Cha, Y.-J., Choi, W. & Büyüköztürk, O. Deep learning-based crack damage detection using convolutional neural networks. *Comput. Civ. Infrastructure Eng.* **32**, 361–378 (2017).
- Imenes, A. *et al.* A deep learning approach for automated fault detection on solar modules using image composites. In *Proceedings of the 2021 IEEE 48th Photovoltaic Specialists Conference (PVSC)*, 1925–1930 (IEEE, Fort Lauderdale, FL, USA, 2021).
- Teke, M., Başeski, E., Ok, A., Yüksel, B. & Şenarar, Ç. Multi-spectral false color shadow detection. In *Proceedings of the ISPRS Conference on Photogrammetric Image Analysis*, 109–119 (Springer, Munich, Germany, 2011).
- Zou, J.-T. & Rajveer, G. Drone-based solar panel inspection with 5g and ai technologies. In *2022 8th International Conference on Applied System Innovation (ICASI)*, 174–178 (IEEE, 2022).
- Meng, Z. *et al.* Defect object detection algorithm for electroluminescence image defects of photovoltaic modules based on deep learning. *Energy Sci. & Eng.* **10**, 800–813 (2022).
- Li, L., Wang, Z. & Zhang, T. Photovoltaic panel defect detection based on ghost convolution with BottleneckCSP and tiny target prediction head incorporating YOLOv5. *arXiv preprint arXiv:2303.00886* (2023).
- Girshick, R. Fast R-CNN. *arXiv preprint arXiv:1504.08083* (2015).
- Hong, F. *et al.* A novel framework on intelligent detection for module defects of PV plant combining the visible and infrared images. *Sol. Energy* **236**, 406–416 (2022).
- Zhang, M. & Yin, L. Solar cell surface defect detection based on improved YOLO v5. *IEEE Access* **10**, 80804–80815 (2022).
- Zheng, Q. *et al.* Lightweight hot-spot fault detection model of photovoltaic panels in UAV remote-sensing image. *Sensors* **22**, 4617 (2022).
- Hassan, S. & Dhimish, M. Enhancing solar photovoltaic modules quality assurance through convolutional neural network-aided automated defect detection. *Renew. Energy* **219**, (2023).
- Cao, Y. *et al.* Improved YOLOv8-GD deep learning model for defect detection in electroluminescence images of solar photovoltaic modules. *Eng. Appl. Artif. Intell.* **131**, (2024).
- Ghahremani, A., Adams, S. D., Norton, M., Khoo, S. Y. & Kouzani, A. Z. Detecting defects in solar panels using the yolo v10 and v11 algorithms. *Electronics* **14**, 344 (2025).
- Wang, N. *et al.* Mra-yolov8: A network enhancing feature extraction ability for photovoltaic cell defects. *Sensors* **25**, 1542 (2025).
- Xu, G., Huang, J., Gong, W. & Teng, J. Solar cell defects detection based on photoluminescence images and upgraded YOLOv5 model. *J. Eng.* **2025**, 8397362 (2025).
- Chen, A., Li, X., Jing, H., Hong, C. & Li, M. Anomaly detection algorithm for photovoltaic cells based on lightweight multi-channel spatial attention mechanism. *Energies* **16**, 1619 (2023).
- Akram, M. W. *et al.* Advancing photovoltaic cells defect detection in electroluminescence images through exploring multiple object detectors. *Sol. Energy Mater. Sol. Cells* **292**, (2025).
- Tian, Y., Ye, Q. & Doermann, D. YOLOv12: Attention-centric real-time object detectors. *arXiv preprint arXiv:2502.12524* (2025).
- Su, B., Zhou, Z. & Chen, H. PVEL-AD: A large-scale open-world dataset for photovoltaic cell anomaly detection. *IEEE Transactions on Ind. Informatics* **19**, 404–413 (2022).
- Susan. Panel solar dataset. <https://universe.roboflow.com/susan-ifblr/panel-solar-bw945> (2024). Accessed on 22 November 2024.
- Revaud, J., Almazán, J., Rezende, R. & de Souza, C. Learning with average precision: Training image retrieval with a listwise loss. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 5107–5116 (Seoul, Republic of Korea, 2019).
- Khanam, R., Asghar, T. & Hussain, M. Comparative performance evaluation of yolov5, yolov8, and yolov11 for solar panel defect detection. In *Solar*, vol. 5, 6 (MDPI, 2025).
- Qu, Z. *et al.* A photovoltaic cell defect detection model capable of topological knowledge extraction. *Sci. Reports* **14**, 21904 (2024).
- Pan, W. *et al.* Enhanced photovoltaic panel defect detection via adaptive complementary fusion in yolo-acf. *Sci. Reports* **14**, 26425 (2024).

Acknowledgements

The researchers wish to extend their sincere gratitude to the Deanship of Scientific Research at the Islamic University of Madinah (KSA) for the support provided to the Post-Publishing Program.

Author contributions

A.M: Methodology, Conceptualization, Software, Coding, Validation, Writing—Original draft. Y.N: Methodology, Reviewing and Editing, Supervision, Validation. B.A: Methodology, Reviewing and Editing, Supervision,

Validation. A.T: Reviewing and Editing, Validation. E.B: Methodology, Reviewing and Editing, Validation. A.R: Reviewing and Editing. Y.A: Reviewing and Editing. M.B: Reviewing and Editing.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.A.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025