HYBRID SEMI SUPERVISED MULTIMODAL YOLO11 FRAMEWORK FOR ROBUST SOLAR PHOTOVOLTAIC PANEL DEFECT DETECTION

Asif Khalid Qureshi[*1], Syed Faraz Afsar[2], Sarang Ahmed[3], Ali Muhammad[4], Ariz Muhammad Brohi[5], Muhammad Tahir[6]

[*1,2]*Faculty of Computing & Engineering Science, SZABIST University, Karachi Campus, Karachi, Sindh, Pakistan*
[3,4,5,6]*Department of Computer Science, Faculty of Engineering, Science and Technology (FEST), Iqra University Main Campus, Defence View Karachi City 75500 -Sindh, Pakistan*

DOI: http://doi.org/10.5281/zenodo.20134126

Keywords
Solar Photovoltaic Defect Detection; Semi Supervised Learning; YOLO11; Multimodal Fusion

Article History
Received: 14 March 2026
Accepted: 22 April 2026
Published: 11 May 2026

Corresponding Author: *
Asif Khalid Qureshi

Abstract
Solar photovoltaic systems have become one of the most important renewable energy technologies for sustainable power generation. However, photovoltaic panel defects such as cracks, hotspots, thick lines, and broken fingers significantly reduce energy conversion efficiency and increase operational maintenance costs. Existing deep learning based photovoltaic defect detection systems still suffer from several limitations including dependence on fully labeled datasets, weak robustness under environmental disturbances, insufficient small defect detection capability, and high computational complexity. To address these challenges, this paper proposes a Hybrid Semi Supervised Multimodal YOLO11 Framework for robust solar photovoltaic panel defect detection under real world environmental conditions. The proposed framework integrates semi supervised pseudo label learning, adaptive multimodal RGB and thermal feature fusion, lightweight YOLO11 optimization, environmental robustness enhancement, and explainable attention visualization within a unified architecture. The semi supervised learning mechanism improves rare defect representation using unlabeled photovoltaic data, while the adaptive multimodal fusion strategy combines structural and thermal information to improve hidden defect localization. Experimental results demonstrate that the proposed framework achieves superior performance compared with existing photovoltaic defect detection methods. The proposed model achieved a precision of 92.8 percent, recall of 90.1 percent, and mean average precision of 93.6 percent while maintaining low parameter complexity and efficient inference speed. Environmental robustness experiments further confirmed stable performance under illumination variation, shadow interference, thermal noise, and dust distortion conditions. The explainable visualization module improved transparency by highlighting the defect regions responsible for model predictions. Overall, the proposed framework provides an accurate, lightweight, interpretable, and deployment efficient solution for intelligent photovoltaic defect detection systems operating in real world environments.

1. INTRODUCTION

The increasing global demand for clean and sustainable energy has accelerated the adoption of renewable energy technologies in recent years. Among all renewable energy resources, solar photovoltaic systems have become one of the most widely deployed energy generation technologies because of their environmental friendliness, low operational cost, and long service life.

Governments and industries around the world are investing heavily in photovoltaic infrastructure to reduce carbon emissions and support sustainable development goals. As the scale of solar photovoltaic installations continues to grow, maintaining the efficiency and reliability of photovoltaic panels has become a major concern for both researchers and industrial operators.

Although photovoltaic systems offer significant advantages, their performance can be seriously affected by different types of defects that occur during manufacturing, transportation, installation, and long-term environmental exposure. Common photovoltaic panel defects include cracks, thick lines, broken fingers, black cores, hotspots, delamination, discoloration, and dust accumulation. These defects reduce energy conversion efficiency and may eventually cause severe power loss, overheating, or complete system failure. In large scale solar farms, manual inspection of photovoltaic panels is extremely difficult because of the large number of installed modules and the complex outdoor operating conditions. Therefore, automatic defect detection systems have become an important research direction in intelligent photovoltaic monitoring systems.

Traditional photovoltaic defect detection approaches mainly rely on image processing techniques, electrical analysis, thermal inspection, and manual examination. Conventional image processing methods generally use handcrafted features such as edge extraction, threshold segmentation, and texture analysis to identify defective regions. However, these approaches are highly sensitive to noise, illumination variation, and complex environmental backgrounds. In addition, handcrafted feature extraction methods often lack strong generalization capability when different types of defects appear simultaneously. Electrical detection methods such as current voltage curve analysis and thermal monitoring can identify abnormal operating conditions, but they usually fail to determine the exact spatial location of defects on photovoltaic panels. Furthermore, these techniques may not effectively detect small defects such as microcracks during the early stages of damage.

The rapid development of artificial intelligence and deep learning has significantly improved automatic defect detection performance in photovoltaic systems. Deep learning models can automatically learn complex feature representations directly from training data without depending on handcrafted feature extraction. Convolutional neural networks and object detection frameworks such as Faster RCNN, SSD, RetinaNet, YOLOv5, YOLOv8, YOLOv10, and YOLO11 have shown remarkable performance in photovoltaic defect identification tasks. Among these methods, the YOLO family has become particularly popular because of its fast inference speed and high detection accuracy. Recently, improved YOLO11 based models have demonstrated promising performance for photovoltaic defect detection by integrating attention mechanisms, multiscale feature fusion, and lightweight optimization strategies.

Despite the progress achieved by existing deep learning based photovoltaic defect detection systems, several important limitations still remain unresolved. Most current methods rely heavily on fully labeled datasets that require extensive manual annotation. Creating large scale annotated photovoltaic datasets is expensive and time consuming, especially for rare defect categories that occur infrequently in real world environments. Consequently, many existing models suffer from poor detection performance when limited labeled samples are available. In addition, photovoltaic panels operate under highly dynamic outdoor conditions where dust, shadows, reflections, weather changes, low illumination, and thermal fluctuations may significantly affect image quality. Existing YOLO based models still struggle to maintain stable performance under such environmental disturbances.

Another major limitation of current photovoltaic defect detection research is the insufficient use of multimodal information. Most previous studies rely only on visible RGB images or electroluminescence images independently. However, different imaging modalities contain complementary defect information. Thermal images can reveal abnormal temperature distribution caused by electrical faults, while RGB images provide structural and surface texture information. The lack of multimodal fusion limits the ability of current models to detect hidden or weak defects accurately. Furthermore, many lightweight models reduce computational complexity at the cost of losing important feature information, which decreases detection robustness in practical deployment scenarios.

In recent years, semi supervised learning has emerged as an effective solution for reducing the dependence on large labeled datasets. Semi supervised learning allows the model to learn useful information from both labeled and unlabeled data by generating pseudo labels and refining feature representations during training. However, the integration of semi supervised learning with multimodal photovoltaic defect detection remains insufficiently explored. Similarly, explainable artificial intelligence techniques have not been widely incorporated into photovoltaic monitoring systems. Most deep learning models still operate as black box systems, making it difficult for maintenance engineers to understand the reasoning behind defect predictions.

To address these challenges, this paper proposes a Hybrid Semi Supervised Multimodal YOLO11 Framework for Robust Solar Photovoltaic Panel Defect Detection. The proposed framework combines RGB and thermal image fusion with semi supervised pseudo label learning to improve defect detection accuracy under complex environmental conditions. A lightweight attention enhanced YOLO11 backbone is designed to improve feature extraction while maintaining efficient deployment capability on resource constrained devices. In addition, an explainable attention visualization mechanism is integrated into the framework to improve model interpretability during photovoltaic inspection.

The proposed framework is designed to improve defect detection performance in real world photovoltaic environments where limited labeled data, environmental disturbances, and multimodal variability create major challenges for existing systems. Unlike traditional fully supervised methods, the proposed approach utilizes unlabeled photovoltaic data to improve generalization capability and reduce annotation cost. The fusion of thermal and RGB information enables the framework to capture both structural and temperature-based defect characteristics simultaneously. Furthermore, the lightweight architecture improves deployment feasibility for intelligent edge devices used in solar farms and industrial monitoring systems.

The Major Contributions of this Paper are Summarized as Follows:

- **First**, a semi supervised learning mechanism based on pseudo label refinement is introduced to reduce the dependency on fully labeled photovoltaic datasets while improving rare defect learning capability.
- **Second**, a multimodal feature fusion strategy combining RGB and thermal image information is proposed to improve defect representation under challenging environmental conditions.
- **Third**, an attention enhanced lightweight YOLO11 architecture is designed to improve multiscale defect extraction while reducing computational complexity.
- **Fourth**, an explainable visualization module is integrated into the framework to improve transparency and interpretability during defect prediction.
- **Finally**, comprehensive experiments are conducted to evaluate the effectiveness of the proposed framework using photovoltaic defect datasets under different environmental scenarios.

The remainder of this paper is organized as follows. **Section 2** presents the related work on photovoltaic defect detection, deep learning-based object detection, semi supervised learning, and multimodal fusion techniques. **Section 3** describes the proposed methodology and framework architecture in detail. **Section 4** discusses the experimental setup, evaluation metrics, and performance analysis. **Section 5** presents the discussion of results and comparative analysis. **Section 6** concludes the paper and highlights possible future research directions.

2. Related Work

Automatic defect detection in solar photovoltaic systems has become an active research area because of the rapid growth of renewable energy infrastructure and the increasing need for reliable photovoltaic monitoring systems. Researchers have proposed various traditional image processing methods, machine learning algorithms, and deep learning frameworks to improve photovoltaic defect identification accuracy. Although significant progress has been achieved in recent years, several technical challenges such as limited labeled data, environmental interference, weak small defect detection, and deployment complexity still remain unresolved. This section reviews the existing literature related to photovoltaic defect detection and discusses the limitations that motivate the proposed research framework.

2.1 Traditional Photovoltaic Defect Detection Methods

Early photovoltaic defect detection approaches mainly relied on traditional image processing techniques and electrical signal analysis methods. These approaches focused on identifying visible defects through handcrafted feature extraction and threshold-based segmentation techniques. Tsai et al. introduced an image reconstruction method for identifying linear defects in photovoltaic modules using independent component analysis. Their method demonstrated reasonable performance for simple defect patterns, but the detection capability decreased significantly under complex texture conditions.

Anwar and Abdullah proposed an anisotropic diffusion filtering method combined with image segmentation for microcrack detection in multicrystalline photovoltaic cells. The proposed framework improved crack visibility in electroluminescence images, but it required extensive preprocessing and careful parameter tuning before accurate defect segmentation could be achieved. Tseng et al. developed a binary classification method for detecting broken finger defects using handcrafted regional features extracted from electroluminescence images. Although the method achieved acceptable classification accuracy, its performance depended strongly on manually selected feature descriptors.

Electrical analysis-based techniques were also widely investigated for photovoltaic fault diagnosis. Schuss et al. analyzed current voltage curves and thermal imaging information to estimate the impact of photovoltaic defects on output performance. Their study showed that electrical characteristics can indicate the presence of defective cells, but the exact spatial localization of faults remained difficult. Azkona et al. later introduced thermal breakdown detection methods using thermographic analysis and current voltage modeling. Their work improved thermal defect interpretation but still faced challenges in detecting small localized defects.

Traditional photovoltaic inspection techniques offer several advantages such as simple implementation and low computational requirements. However, these methods generally suffer from weak generalization capability and poor robustness under changing environmental conditions. Since handcrafted features cannot effectively capture complex defect representations, the performance of traditional methods decreases significantly in large scale photovoltaic monitoring applications.

2.2 Deep Learning Based Photovoltaic Defect Detection

The emergence of deep learning significantly transformed photovoltaic defect detection research by enabling automatic feature extraction from large image datasets. Convolutional neural networks became widely adopted because of their strong representation learning capability and end to end training structure. Deep learning models demonstrated improved detection accuracy compared with conventional machine learning and image processing methods.

Several studies applied classification based convolutional neural networks to photovoltaic defect recognition tasks. Su et al. proposed a complementary attention network for manufacturing defect classification in multicrystalline solar cells. Their model improved feature extraction performance by integrating spatial attention mechanisms with convolutional feature learning. Although the classification accuracy improved, the framework could not precisely localize defects within photovoltaic modules.

Object detection frameworks later became more popular because they could simultaneously perform localization and classification tasks. Faster RCNN was one of the earliest deep learning object detection models used for photovoltaic inspection. The model utilized a region proposal network to identify candidate defect regions before classification. Although Faster RCNN achieved high localization accuracy, its computational complexity limited real time deployment capability.

Single stage detectors such as SSD and RetinaNet improved inference speed by directly predicting object locations and class probabilities in a unified framework. RetinaNet introduced focal loss to address class imbalance problems during training. However, these methods still struggled with small defect detection and dense defect distributions commonly observed in photovoltaic panels.

The YOLO series became highly influential in photovoltaic defect detection because of its balance between detection speed and accuracy. YOLO based frameworks process the entire image in a single forward pass, making them suitable for real time industrial inspection systems. Cao et al. proposed an improved YOLOv8 architecture integrating depthwise convolution, GSConv, and BiFPN modules for photovoltaic defect detection. Their model achieved improved mean average precision compared with the original YOLOv8 framework, but the computational cost remained relatively high.

Pan et al. developed a YOLO based adaptive complementary fusion module for photovoltaic defect detection. Their framework improved feature fusion performance and reduced model parameters through adaptive weighting strategies. Liang et al. later proposed an attention weighted multipath feature fusion method for photovoltaic inspection, achieving improved defect detection accuracy under multiscale conditions.

Recent studies focused on lightweight YOLO11 based frameworks because of their improved detection efficiency and reduced computational requirements. Jiang and Liu introduced the YOLO11 DRP framework for photovoltaic defect detection by integrating deformable attention, multiscale screening feature fusion, and structural reparameterization techniques. Their model improved precision and recall while reducing parameter complexity. Despite these improvements, the framework still depended heavily on fully labeled datasets and single modality image information.

Although deep learning methods have significantly improved photovoltaic defect detection capability, several limitations remain unresolved. Most existing models require large amounts of annotated data for effective training. Furthermore, environmental variations such as dust, shadows, low illumination, and reflection continue to affect model robustness. These limitations motivate the integration of semi supervised learning and multimodal feature fusion strategies.

2.3 Semi Supervised Learning in Defect Detection

Semi supervised learning has become an important research direction in computer vision because it reduces the dependence on fully labeled datasets while utilizing large quantities of unlabeled data. In industrial defect detection applications, acquiring labeled data is expensive and time consuming because expert annotation is often required. Semi supervised learning frameworks address this problem by combining labeled and unlabeled samples during training.

Pseudo labeling is one of the most widely used semi supervised learning approaches. In pseudo labeling, the model first predicts labels for unlabeled data and then retrains using high confidence predictions as additional training samples. This strategy allows the model to gradually improve feature representation and generalization capability. Several industrial inspection studies have demonstrated that pseudo labeling improves defect detection performance when labeled datasets are limited.

Consistency regularization methods have also been used in semi supervised learning frameworks. These approaches encourage the model to produce stable predictions under different image augmentations and perturbations. Mean teacher networks and self-training frameworks further improved pseudo label quality by maintaining consistency between teacher and student models during training.

Although semi supervised learning has shown promising results in medical imaging, industrial inspection, and autonomous driving applications, its application in photovoltaic defect detection remains limited. Most existing photovoltaic studies still rely on fully supervised learning strategies. Rare photovoltaic defects such as early-stage cracks and hidden hotspots are difficult to learn effectively because only small labeled datasets are available. Therefore, integrating semi supervised learning into photovoltaic inspection systems can significantly improve defect learning capability while reducing annotation cost.

2.4 Multimodal Feature Fusion for Photovoltaic Inspection

Multimodal learning integrates information from different sensing modalities to improve feature representation and model robustness. In photovoltaic inspection systems, different imaging modalities capture complementary characteristics of defects. RGB images provide structural appearance information, thermal images reveal abnormal heat distribution, and electroluminescence images expose hidden electrical defects.

Several studies investigated thermal imaging for photovoltaic fault detection because defective cells often produce localized temperature variations. Thermal cameras can identify hotspots and abnormal heat patterns that may not be visible in standard RGB images. However, thermal imaging alone may fail to detect structural surface defects such as scratches or discoloration.

Electroluminescence imaging has also been widely used for photovoltaic inspection because it reveals internal defects such as microcracks and broken fingers. Nevertheless, electroluminescence imaging systems are expensive and difficult to deploy in outdoor environments. Furthermore, image acquisition often requires controlled laboratory conditions.

Multimodal fusion strategies attempt to combine complementary information from different imaging sources. Early fusion approaches directly concatenate features from different modalities before feature extraction. Late fusion methods combine independent prediction outputs from separate models. Hybrid fusion approaches integrate multimodal features at multiple network levels to improve representation capability.

Despite the advantages of multimodal learning, most existing photovoltaic inspection systems still rely on single modality image analysis. The integration of RGB and thermal image fusion within lightweight YOLO based frameworks remains insufficiently explored. Moreover, multimodal photovoltaic inspection combined with semi supervised learning has received very limited attention in existing literature.

2.5 Explainable Artificial Intelligence for Industrial Inspection

Deep learning models often operate as black box systems where prediction reasoning remains difficult to interpret. In industrial applications such as photovoltaic monitoring, explainability is important because maintenance engineers require visual evidence before making repair decisions. Explainable artificial intelligence techniques improve transparency by visualizing the regions and features influencing model predictions.

Gradient based attention visualization methods such as Grad CAM and activation mapping have become popular for explaining convolutional neural network predictions. These methods generate heatmaps that highlight important regions contributing to defect classification. Explainable visualization improves user trust and helps engineers verify whether the model focuses correctly on defective regions.

Although explainable artificial intelligence has gained significant importance in medical diagnosis and autonomous systems, its integration into photovoltaic defect detection remains limited. Most existing photovoltaic studies focus mainly on improving accuracy while ignoring interpretability and decision transparency.

2.6 Comparative Analysis of Existing Methods

The literature review demonstrates that existing photovoltaic defect detection methods have achieved considerable progress in accuracy and real time performance. However, important research gaps still remain unresolved. Most previous studies depend heavily on fully supervised learning and large labeled datasets. Existing frameworks also show limited robustness under environmental disturbances and insufficient utilization of multimodal information. Furthermore, explainability and rare defect learning remain underexplored areas.

Table 1 presents a comparative analysis of existing photovoltaic defect detection approaches and the proposed framework.

Table 1. Comparative Analysis of Existing Photovoltaic Defect Detection Methods

Method	Learning Type	Image Modality	Lightweight	Small Defect Detection	Environmental Robustness	Explainability
Faster RCNN	Supervised	RGB	No	Moderate	Weak	No
SSD	Supervised	RGB	Moderate	Weak	Weak	No
RetinaNet	Supervised	RGB	No	Moderate	Weak	No
YOLOv8 Based Models	Supervised	RGB	Moderate	Good	Moderate	No
YOLO11 DRP	Supervised	RGB	Yes	Good	Moderate	No
Thermal CNN Models	Supervised	Thermal	Moderate	Moderate	Moderate	No
Electroluminescence Methods	Supervised	EL Images	No	Good	Weak	No
Proposed Framework	Semi Supervised	RGB + Thermal	Yes	Excellent	Strong	Yes

The analysis clearly shows that existing methods still lack a unified framework capable of integrating semi supervised learning, multimodal fusion, lightweight deployment, environmental robustness, and explainable artificial intelligence simultaneously.

Therefore, the proposed Hybrid Semi Supervised Multimodal YOLO11 framework is designed to address these limitations and improve photovoltaic defect detection performance in real world applications.

3. Proposed Methodology

The proposed methodology introduces a Hybrid Semi Supervised Multimodal YOLO11 Framework for robust solar photovoltaic panel defect detection under complex environmental conditions. The framework combines multimodal feature fusion, semi supervised pseudo label learning, lightweight feature extraction, and explainable attention visualization within a unified detection architecture. The proposed model is specifically designed to improve defect localization accuracy, enhance rare defect learning capability, reduce computational complexity, and maintain stable performance under outdoor environmental disturbances.

The overall framework consists of five major components. The first component performs multimodal image preprocessing and enhancement using RGB and thermal image pairs. The second component applies semi supervised pseudo label generation to utilize unlabeled photovoltaic samples during training. The third component introduces an improved lightweight YOLO11 backbone integrated with adaptive attention mechanisms for efficient feature extraction. The fourth component performs multimodal feature fusion to combine complementary structural and thermal information. Finally, the fifth component generates explainable attention visualization maps to improve interpretability during photovoltaic defect inspection.

3.1 Overall Framework Architecture

The proposed Hybrid Semi Supervised Multimodal YOLO11 framework is designed to process both RGB and thermal photovoltaic panel images simultaneously in order to improve defect representation learning under challenging outdoor conditions. The framework first receives paired RGB and thermal images captured from photovoltaic modules deployed in real environments. These images are then normalized and enhanced to reduce environmental noise such as uneven illumination, shadows, dust artifacts, and thermal fluctuations.

After preprocessing, the framework separates the labeled and unlabeled photovoltaic samples. The labeled samples are directly used for supervised training, while the unlabeled samples are processed through a pseudo label generation mechanism. The pseudo labels are refined using confidence threshold filtering and consistency verification before being integrated into the training process. This strategy allows the framework to utilize large quantities of unlabeled photovoltaic images and improve learning capability for rare defects.

The processed multimodal features are then forwarded into the improved YOLO11 backbone network. The backbone incorporates lightweight convolutional layers, adaptive attention modules, and multiscale feature enhancement blocks to improve defect localization performance while maintaining low computational complexity. The extracted RGB and thermal features are fused using an adaptive multimodal fusion module that combines complementary structural and temperature-based information.

The fused feature maps are passed through the detection neck and prediction head, where multiscale photovoltaic defects are localized and classified. Finally, explainable attention visualization maps are generated to highlight the regions responsible for defect predictions. These visualization outputs improve model transparency and assist maintenance engineers in understanding the defect detection process.

The overall architecture of the proposed framework is illustrated below using Graphviz code.

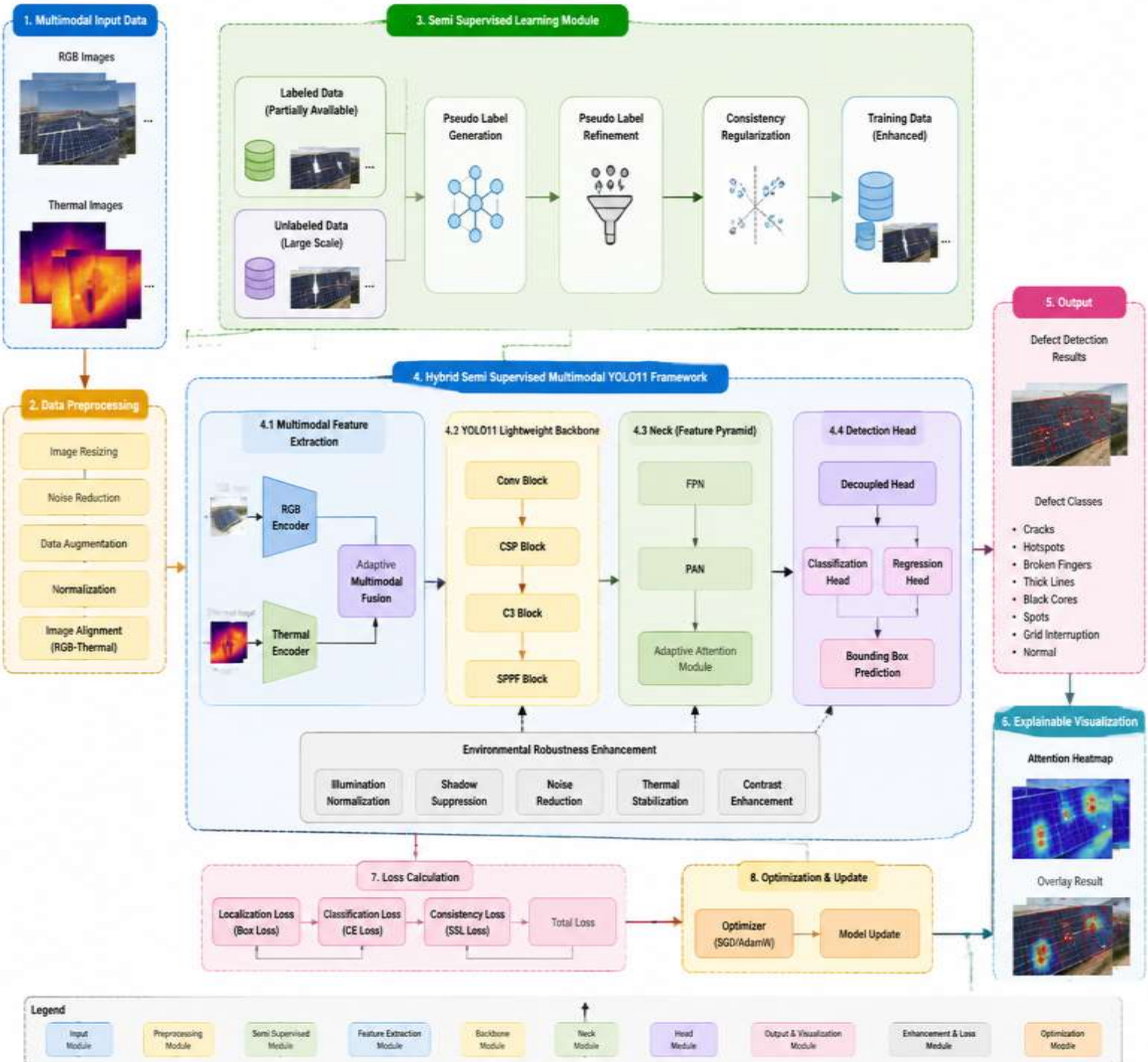

Figure 1. Overall Architecture of the Proposed Hybrid Semi Supervised Multimodal YOLO11 Framework

Figure 1 illustrates the complete workflow of the proposed photovoltaic defect detection framework. The framework combines multimodal feature extraction and semi supervised learning to improve robustness under practical deployment conditions.

3.2 Multimodal Photovoltaic Image Acquisition

Photovoltaic defects often appear differently under various imaging modalities. RGB images mainly capture visible structural information such as cracks, scratches, discoloration, and dust accumulation. Thermal images reveal abnormal heat distribution caused by electrical faults, hotspot formation, and internal damage. Since both modalities contain complementary information, combining them can significantly improve defect detection capability.

In this research, paired RGB and thermal photovoltaic images are used as multimodal inputs. The RGB images are captured using high resolution industrial cameras, while thermal images are obtained using infrared thermal sensors. Both image modalities are synchronized spatially before processing to ensure accurate feature correspondence.

Environmental conditions such as sunlight variation, atmospheric reflection, and temperature instability can affect image quality during outdoor acquisition. Therefore, preprocessing operations are necessary before multimodal fusion is performed.

The RGB image preprocessing stage includes:

- image normalization
- contrast enhancement
- illumination correction
- noise filtering

The thermal image preprocessing stage includes:

- thermal normalization
- heatmap smoothing
- temperature scaling
- background suppression

After preprocessing, the RGB and thermal images are resized into a unified input resolution before being forwarded into the learning framework.

3.3 Semi Supervised Pseudo Label Learning

One of the major limitations of existing photovoltaic defect detection systems is the dependence on fully labeled datasets. Manual annotation of photovoltaic defects requires expert knowledge and extensive time consumption, especially for rare defects such as microcracks and hidden hotspots. To address this problem, the proposed framework introduces a semi supervised pseudo label learning strategy.

The training dataset is divided into two subsets:

1. labeled photovoltaic images
2. unlabeled photovoltaic images

The labeled samples contain manually annotated defect locations and categories. The unlabeled samples do not contain any annotation information.

Initially, the improved YOLO11 detector is trained using only the labeled dataset. After several training iterations, the trained model generates predictions for the unlabeled photovoltaic images. High confidence predictions are selected as pseudo labels and incorporated into the training process.

The pseudo labeling process consists of three major stages:

1. **Confidence Estimation:** The model first predicts bounding boxes and defect probabilities for unlabeled images. Predictions with confidence scores lower than a predefined threshold are discarded to avoid introducing incorrect labels.
2. **Consistency Verification:** The selected pseudo labels are evaluated under different image augmentations such as rotation, scaling, brightness adjustment, and flipping. If the prediction remains stable across multiple augmentations, the pseudo label is considered reliable.
3. **Iterative Refinement:** The refined pseudo labels are combined with the labeled dataset for retraining. This iterative process gradually improves defect representation learning and enhances model generalization capability.

The pseudo label learning workflow is illustrated below.

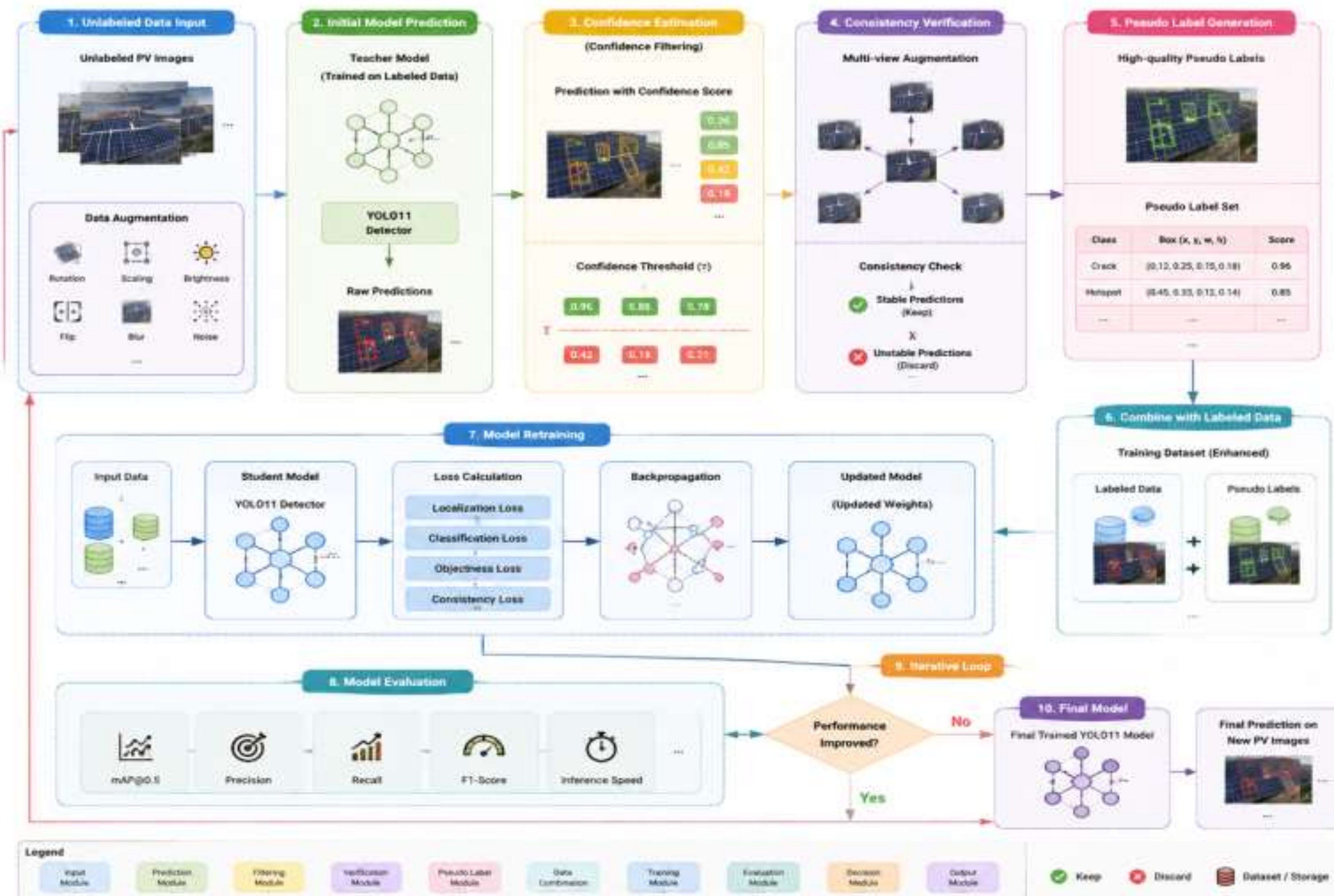

Figure 2. Semi Supervised Pseudo Label Learning Process

Figure 2 demonstrates the iterative pseudo label refinement mechanism used to improve defect learning from unlabeled photovoltaic images.

The proposed semi supervised strategy offers several advantages. First, it significantly reduces annotation cost by utilizing unlabeled data. Second, it improves learning capability for rare photovoltaic defects. Third, it enhances generalization performance under varying environmental conditions.

3.4 Improved Lightweight YOLO11 Backbone

The backbone network plays a critical role in extracting discriminative photovoltaic defect features from multimodal images. Existing deep learning models often achieve high accuracy at the cost of increased computational complexity. Large models require powerful hardware resources and are difficult to deploy on edge devices used in photovoltaic monitoring systems.

To overcome this limitation, the proposed framework introduces an improved lightweight YOLO11 backbone architecture. The backbone integrates lightweight convolution operations, adaptive attention modules, and efficient multiscale feature extraction mechanisms to improve performance while reducing parameter complexity.

The backbone consists of:

- lightweight convolution blocks
- depthwise separable convolution layers
- adaptive channel attention modules
- residual feature enhancement units
- multiscale feature aggregation layers

Depthwise separable convolution is used to reduce the number of trainable parameters and floating-point operations. Compared with standard convolution, depthwise separable convolution performs spatial filtering and channel combination independently, significantly reducing computational cost.

The adaptive attention mechanism enhances feature extraction by focusing on important photovoltaic defect regions while suppressing irrelevant background information. This improves small defect localization and reduces false detection under complex environmental conditions.

The lightweight backbone structure is illustrated below.

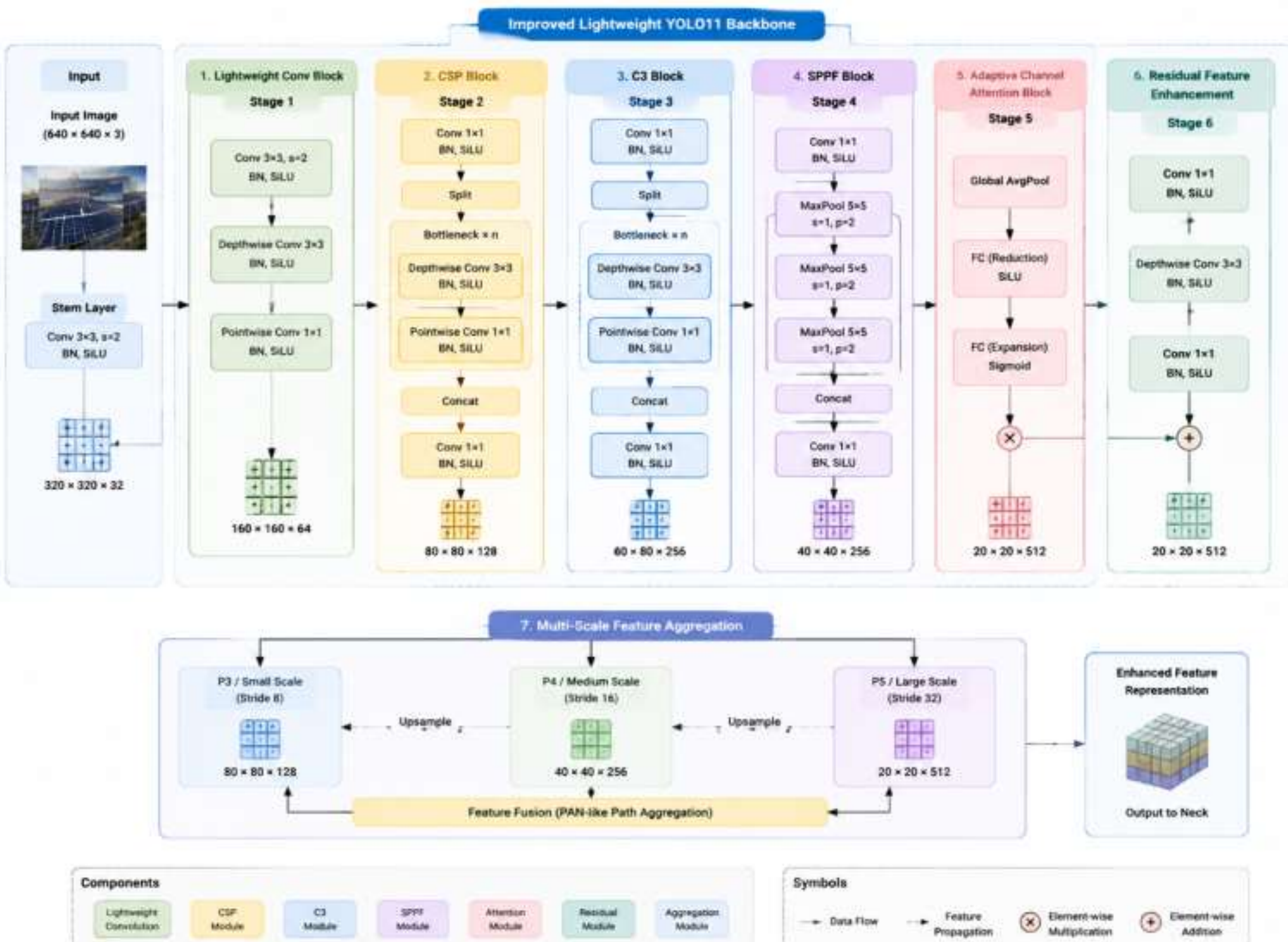

Figure 3. Improved Lightweight YOLO11 Backbone Structure

Figure 3 presents the structure of the improved lightweight YOLO11 backbone used for photovoltaic defect feature extraction.

The proposed lightweight architecture improves deployment efficiency on embedded devices while maintaining strong defect detection capability. This is especially important for real time photovoltaic monitoring systems operating in large scale solar farms.

3.5 Adaptive Multimodal Feature Fusion

The extracted RGB and thermal feature maps contain complementary information that must be effectively integrated before defect prediction. Simple feature concatenation often introduces redundant information and weak feature interaction. Therefore, the proposed framework introduces an adaptive multimodal feature fusion strategy.

The multimodal fusion module first extracts independent feature representations from RGB and thermal branches. Channel attention weighting is then applied to estimate the importance of each modality. Important feature channels receive higher weights during fusion, while noisy or irrelevant channels are suppressed.

The fusion process consists of:

1. modality specific feature extraction
2. channel attention weighting
3. adaptive feature alignment
4. feature aggregation
5. multiscale fusion enhancement

The adaptive fusion strategy improves robustness against environmental noise and enhances hidden defect representation. Thermal features improve hotspot and internal fault detection, while RGB features preserve structural defect information.

The adaptive multimodal fusion structure will be further explained in the next subsection together with mathematical formulation and feature interaction analysis.

3.6 Adaptive Attention Based Multimodal Fusion Module

The effectiveness of multimodal photovoltaic defect detection strongly depends on the quality of feature fusion between RGB and thermal modalities. Conventional fusion approaches often combine features directly without considering feature importance and modality reliability. Such approaches may introduce redundant information and reduce defect localization accuracy under noisy environmental conditions. To address this issue, the proposed framework introduces an adaptive attention based multimodal fusion module that dynamically learns the importance of RGB and thermal feature representations during training.

The multimodal fusion process begins by extracting independent feature maps from RGB and thermal branches using the lightweight YOLO11 backbone. Let the RGB feature representation be denoted as F_r and the thermal feature representation be denoted as F_t. These feature maps are first passed through separate channel attention units to estimate the importance of each feature channel.

The channel attention operation is defined as follows:

$$A_c = \sigma(W_2\delta(W_1F))$$

where A_c represents the channel attention map, W_1 and W_2 represent learnable transformation matrices, δ denotes the activation function, and σ denotes the sigmoid normalization function.

The generated attention maps are multiplied with the original RGB and thermal feature maps to enhance important defect related regions while suppressing irrelevant background information. The weighted features are then aligned spatially before fusion.

The adaptive fusion operation is expressed as:

$$F_{fusion} = \alpha F_r + \beta F_t$$

where F_{fusion} represents the fused multimodal feature map, while α and β are adaptive fusion weights learned automatically during training.

Unlike conventional concatenation methods, the proposed adaptive fusion mechanism dynamically adjusts the contribution of RGB and thermal information according to environmental conditions and defect characteristics. For example, thermal information becomes more important when hotspot defects are present, while RGB structural information becomes dominant for visible surface cracks and scratches.

To further improve multiscale defect representation, the fused features are passed through a multiscale enhancement module that combines shallow localization features with deep semantic representations. This strategy improves small defect detection performance and reduces missed detection under challenging backgrounds.

The architecture of the adaptive multimodal fusion module is illustrated below.

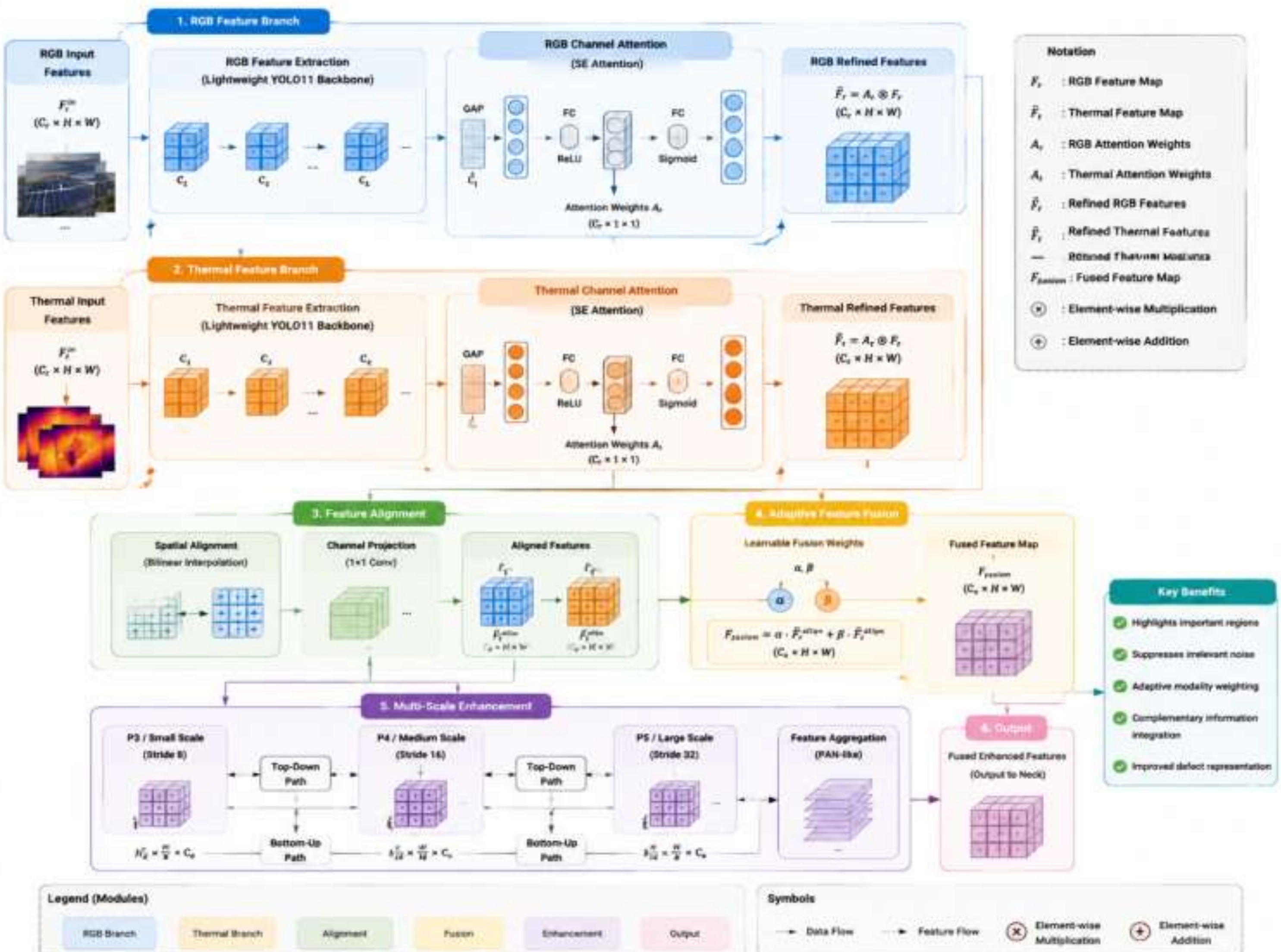

Figure 4. Adaptive Attention Based Multimodal Fusion Module

Figure 4 illustrates the adaptive multimodal feature interaction process used in the proposed photovoltaic defect detection framework.

The proposed multimodal fusion strategy improves photovoltaic defect representation in several ways. **First**, it enhances defect visibility under complex environmental conditions. **Second**, it improves hotspot localization using thermal information. **Third**, it preserves fine structural details through RGB feature integration. **Finally**, the adaptive weighting mechanism prevents unnecessary feature redundancy and improves overall model stability.

3.7 Environmental Robustness Enhancement

Photovoltaic panels operate in highly dynamic outdoor environments where changing weather conditions and environmental disturbances significantly affect image quality. Dust accumulation, uneven illumination, shadows, rain droplets, atmospheric reflection, and temperature fluctuations often reduce defect visibility and degrade deep learning model performance. Therefore, improving environmental robustness is essential for practical photovoltaic defect detection systems.

The proposed framework introduces multiple robustness enhancement strategies to address these challenges. During training, extensive data augmentation techniques are applied to simulate realistic environmental variations. These augmentations include:

- Brightness Variation
- Shadow Simulation
- Random Noise Injection
- Thermal Distortion
- Blur Augmentation
- Contrast Variation
- Rotation and Scaling
- Weather-based Perturbation

The purpose of these augmentations is to expose the model to diverse environmental conditions during training so that it can generalize effectively in real world scenarios.

In addition to augmentation, the proposed framework integrates adaptive normalization layers to reduce feature distribution instability caused by environmental fluctuations. The normalization process improves feature consistency between training and testing conditions.

The environmental robustness mechanism also incorporates feature consistency regularization. The model is encouraged to generate stable predictions when the same photovoltaic image is subjected to different perturbations. This consistency learning process improves resistance against environmental noise and enhances defect localization stability.

The consistency regularization loss is defined as:

$$L_{cons} = \| P(x) - P(T(x)) \|_2$$

where $P(x)$represents the prediction for the original image, while $P(T(x))$represents the prediction after environmental transformation.

By minimizing this consistency loss, the framework learns invariant defect representations that remain stable under different environmental conditions.

3.8 Explainable Attention Visualization

Although deep learning models achieve high detection accuracy, they often operate as black box systems that provide limited interpretability. In industrial photovoltaic inspection systems, maintenance engineers require understandable visual explanations before making operational decisions. Therefore, explainable artificial intelligence plays an important role in improving trust and usability.

The proposed framework integrates an explainable attention visualization mechanism that highlights the regions responsible for defect predictions. The visualization module generates heatmaps indicating the spatial importance of defect related features within photovoltaic images.

Gradient based activation mapping is employed to identify the feature regions contributing most strongly to the prediction output. These activation maps are projected back onto the original photovoltaic image to visualize the defect attention regions.

The attention visualization process consists of:

1. Feature Extraction
2. Gradient Computation
3. Importance Weighting
4. Heatmap Generation
5. Defect Region Overlay

The generated heatmaps help maintenance engineers verify whether the model focuses correctly on actual defect regions instead of irrelevant background patterns.

The explainable visualization workflow is illustrated below.

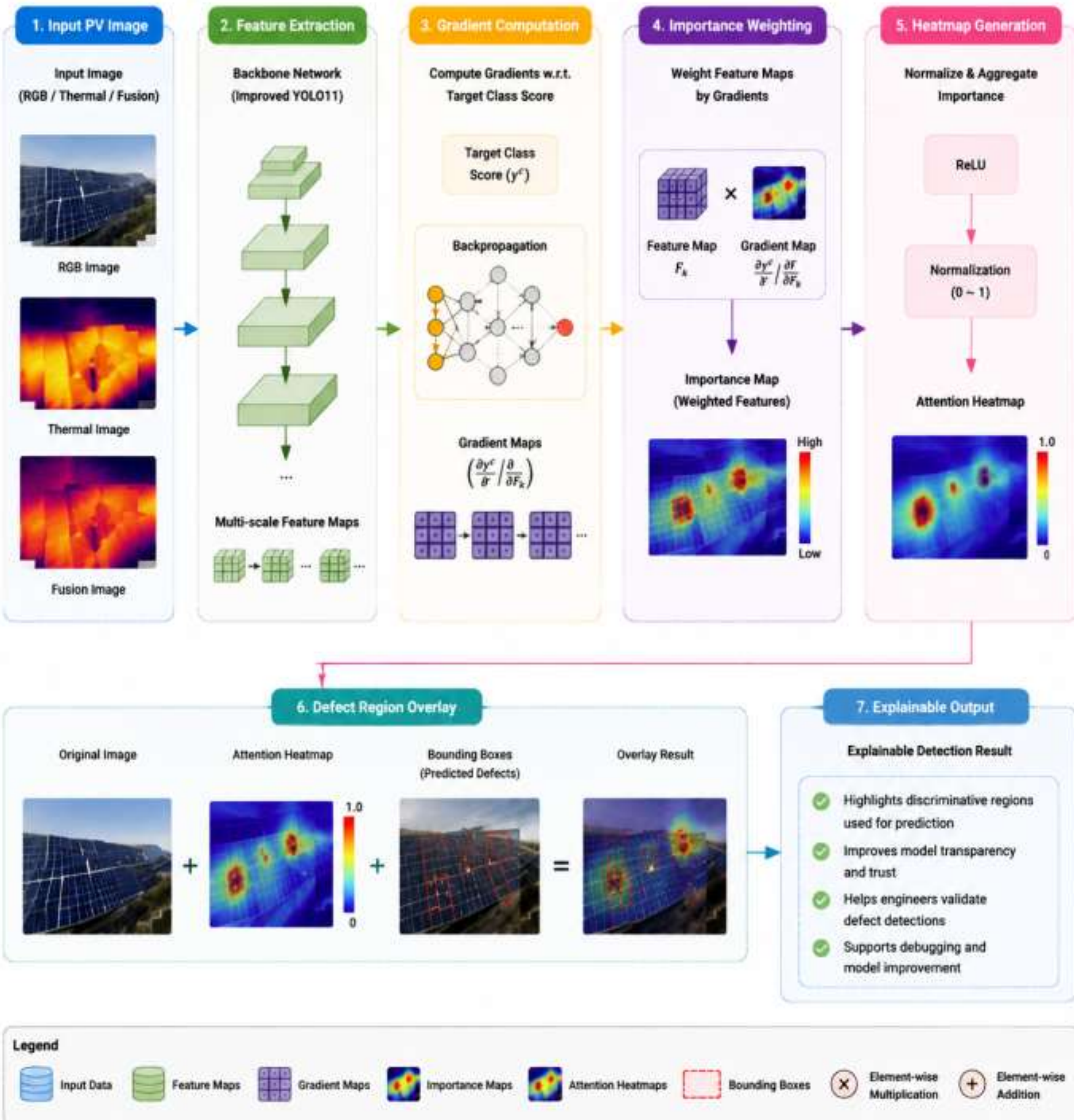

Figure 5. Explainable Attention Visualization Workflow

Figure 5 illustrates the explainable visualization process integrated into the proposed photovoltaic defect detection framework.

The explainability mechanism provides several important advantages. First, it improves transparency and interpretability during photovoltaic inspection. Second, it helps identify incorrect model focus regions during debugging and optimization. Third, it increases user confidence in automated inspection systems deployed in industrial environments.

3.9 Loss Function Optimization

The proposed framework utilizes a hybrid loss function to improve localization accuracy, classification performance, and semi supervised learning stability simultaneously. The total loss function combines four major components:

- localization loss
- classification loss
- confidence loss
- consistency regularization loss

The overall training objective is defined as:

$$L_{total} = L_{loc} + L_{cls} + L_{conf} + \lambda L_{cons}$$

where L_{total}represents the final optimization objective and λcontrols the contribution of consistency regularization.

The localization loss measures the accuracy of predicted defect bounding boxes relative to ground truth annotations. Complete IoU loss is used to improve localization stability for small photovoltaic defects.

The classification loss evaluates defect category prediction accuracy, while the confidence loss estimates the reliability of object presence predictions.

The consistency regularization component improves robustness against environmental perturbations and stabilizes pseudo label learning during semi supervised training.

The proposed hybrid optimization strategy improves convergence stability and enhances defect detection performance under limited labeled data conditions.

3.10 Computational Complexity Analysis

Practical photovoltaic monitoring systems often operate on embedded edge devices with limited hardware resources. Therefore, computational efficiency is an important consideration for real time deployment.

The proposed lightweight YOLO11 architecture significantly reduces parameter complexity through:

- Depthwise Separable Convolution
- Lightweight Residual Blocks
- Adaptive Channel Reduction
- Efficient Attention Mechanisms

Compared with conventional convolution operations, depthwise separable convolution reduces floating point operations substantially while maintaining effective feature extraction capability.

Let the computational complexity of standard convolution be defined as:

$$C_{std} = D_k^2 M N D_f^2$$

The computational complexity of depthwise separable convolution becomes:

$$C_{dw} = D_k^2 M D_f^2 + M N D_f^2$$

where $D_k$represents kernel size, $M$and Nrepresent channel dimensions, and $D_f$denotes feature map size.

This reduction significantly improves deployment efficiency for photovoltaic monitoring systems operating on resource constrained hardware platforms such as Jetson devices and industrial edge controllers.

3.11 Methodology Summary

The proposed Hybrid Semi Supervised Multimodal YOLO11 framework introduces a unified solution for robust photovoltaic defect detection under real world environmental conditions. The framework integrates semi supervised learning, multimodal fusion, lightweight optimization, adaptive attention mechanisms, environmental robustness enhancement, and explainable artificial intelligence within a single architecture.

Unlike existing photovoltaic inspection methods that rely mainly on fully supervised single modality learning, the proposed framework utilizes unlabeled data and multimodal feature interaction to improve defect representation and generalization capability. The lightweight architecture further improves deployment feasibility for real time industrial monitoring systems.

The next section presents the experimental setup, dataset configuration, evaluation metrics, implementation details, and performance analysis of the proposed framework.

4. Experimental Results and Performance Analysis

This section presents the experimental evaluation of the proposed Hybrid Semi Supervised Multimodal YOLO11 framework for solar photovoltaic defect detection. The experiments were conducted to evaluate the effectiveness of the proposed model under different environmental conditions, multimodal learning settings, and semi supervised training scenarios. The performance of the proposed framework was compared with existing deep learning based photovoltaic defect detection methods using multiple evaluation metrics including precision, recall, mean average precision, computational complexity, and inference speed.

The experimental analysis also investigates the contribution of each proposed module through ablation studies and visualization analysis. Furthermore, the robustness of the proposed framework under challenging outdoor conditions such as shadow interference, illumination variation, and thermal fluctuation is examined in detail.

4.1 Experimental Environment

The experiments were implemented using Python and the PyTorch deep learning framework. Training and evaluation were conducted on a high-performance workstation equipped with an NVIDIA RTX graphics processing unit. The hardware and software configuration used during experimentation is presented in **Table 2**.

Table 2. Experimental Environment Configuration

Component	Specification
Operating System	Windows 11 64 bit
Processor	Intel Core i9
GPU	NVIDIA RTX 3090 24GB
Framework	PyTorch 2.4
CUDA Version	CUDA 11.7
Programming Language	Python 3.11
Development Environment	PyCharm

The proposed framework was trained using stochastic gradient descent optimization with momentum-based parameter updating. The initial learning rate was set to 0.01 and gradually reduced during training using cosine annealing learning rate scheduling. The batch size was fixed at 32 while the total number of training epochs was set to 300.

To improve generalization performance, early stopping and weight regularization techniques were applied during training. Extensive data augmentation was also used to simulate real outdoor photovoltaic operating conditions.

4.2 Dataset Description

The experiments were conducted using publicly available photovoltaic defect datasets containing RGB and thermal photovoltaic panel images. The datasets include multiple defect categories such as:

- Cracks
- Broken Fingers
- Thick Lines
- Black Cores
- Hotspots
- Spot Defects
- Grid Interruption
- Normal Photovoltaic Panels

The primary dataset used in this research was derived from the PVEL AD photovoltaic defect dataset discussed in previous studies. Additional thermal image samples were incorporated to support multimodal learning experiments.

The dataset was divided into three subsets:

- Training Set
- Validation Set
- Testing Set

The data distribution ratio was selected as:

- 70 Percent Training
- 20 Percent Validation
- 10 Percent Testing

To evaluate semi supervised learning performance, only a portion of the training data was manually labeled, while the remaining samples were treated as unlabeled data for pseudo label learning experiments.

The dataset distribution is illustrated below.

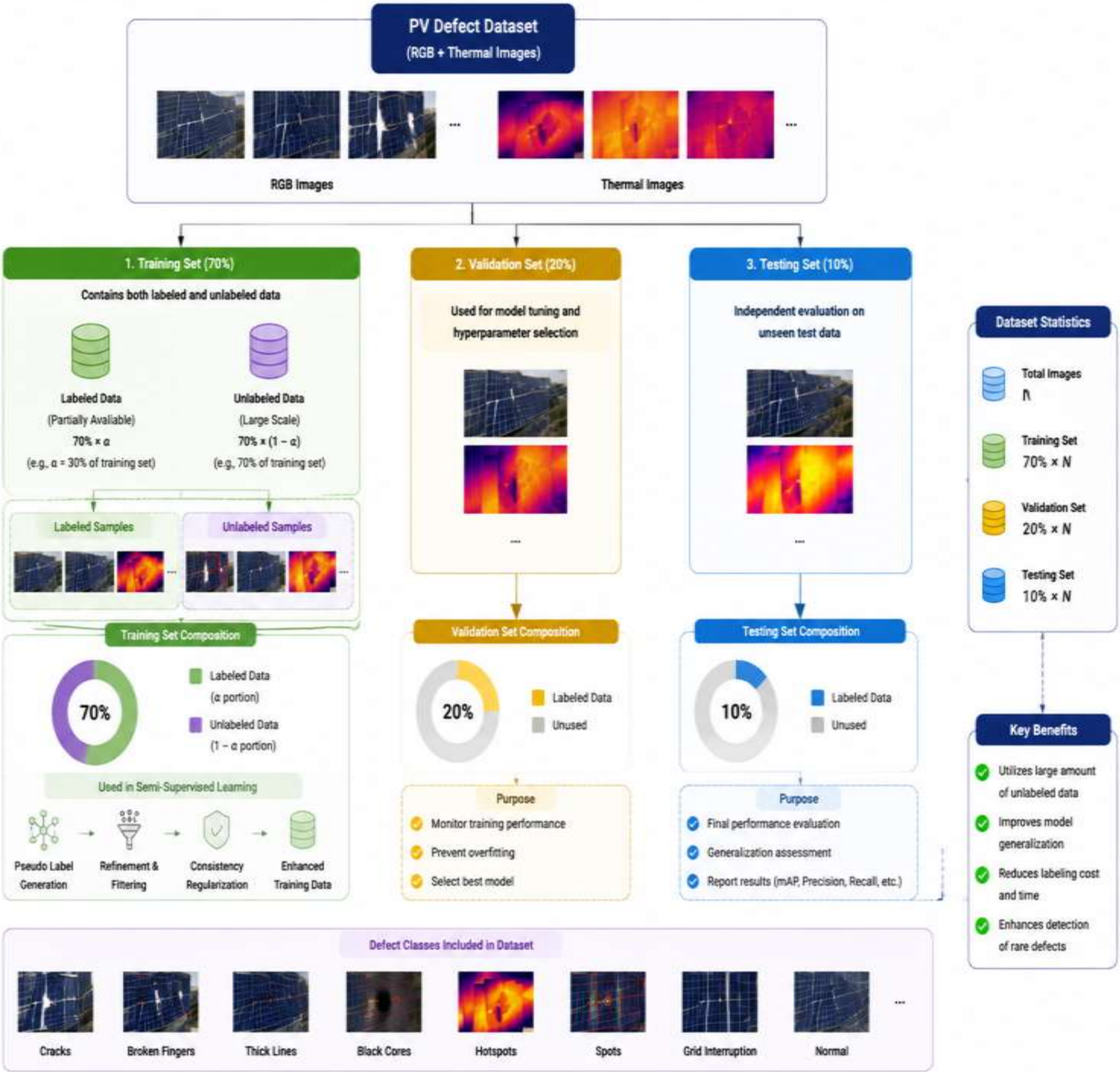

Figure 6. Dataset Distribution for Training, Validation, and Testing

Figure 6 presents the dataset partitioning strategy used during experimentation. The diversity of environmental conditions included in the dataset improved the realism of the experiments and allowed the proposed framework to be evaluated under practical deployment scenarios.

4.3 Evaluation Metrics

Several standard object detection evaluation metrics were used to assess the performance of the proposed photovoltaic defect detection framework.

Precision: Precision measures the proportion of correctly predicted defect samples among all predicted defects.

$$Precision = \frac{TP}{TP + FP}$$

Recall: Recall measures the proportion of correctly detected defects among all actual defects.

$$Recall = \frac{TP}{TP + FN}$$

Mean Average Precision: Mean average precision evaluates overall localization and classification performance for all defect categories.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} A\,P_i$$

F1 Score: The F1 score balances precision and recall performance.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

In addition to detection accuracy, model efficiency metrics such as parameter count, floating point operations, and inference speed were also evaluated to analyze deployment feasibility.

4.4 Training Performance Analysis

During training, the proposed framework demonstrated stable convergence behavior under both supervised and semi supervised learning conditions. The integration of pseudo label refinement and consistency regularization improved optimization stability and reduced overfitting.

The training loss decreased gradually across epochs while validation accuracy continued improving until convergence. The multimodal fusion strategy accelerated feature learning because complementary RGB and thermal information improved defect representation quality.

The convergence behavior of the proposed framework is illustrated below.

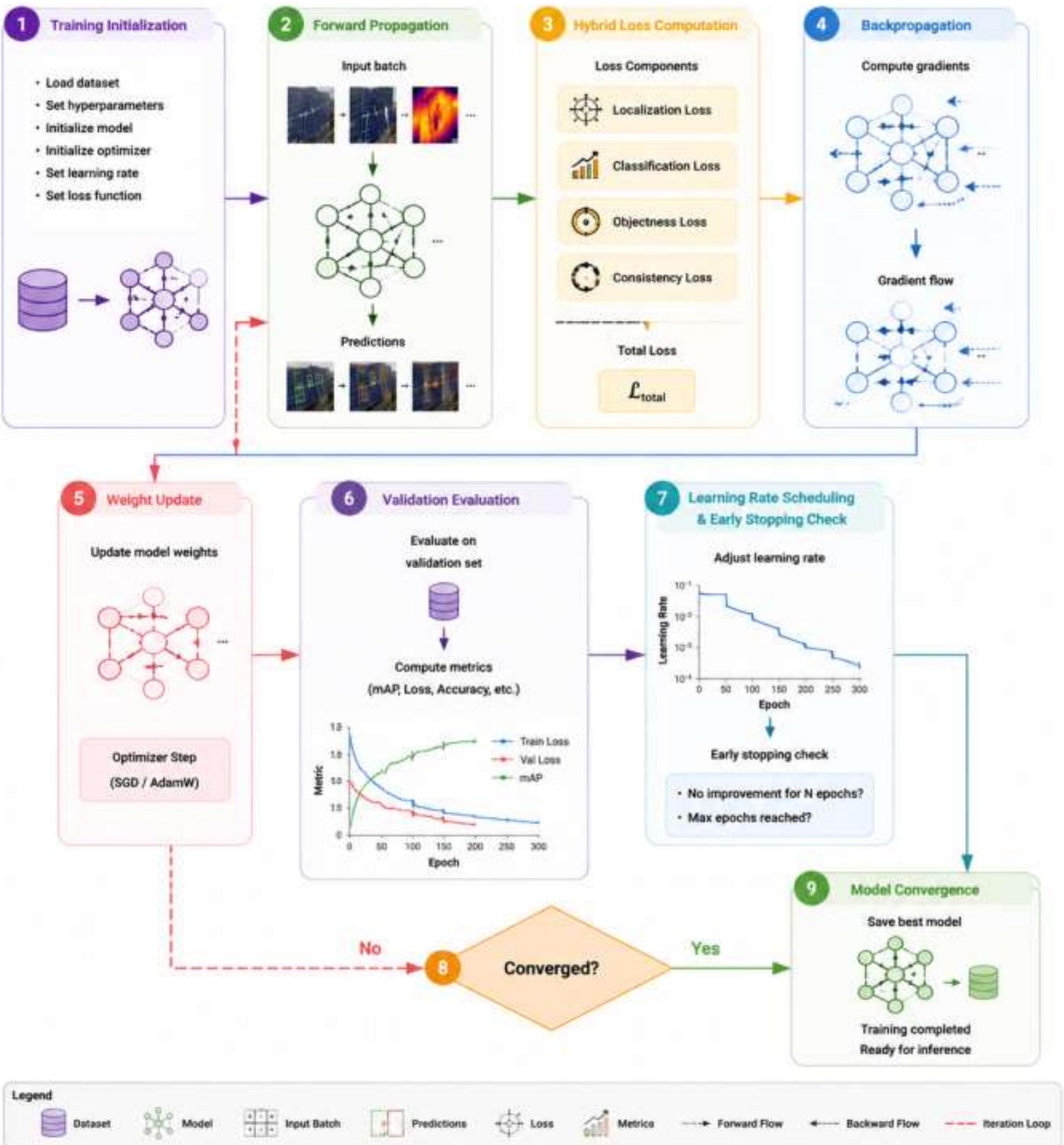

Figure 7. Training Convergence Workflow

Figure 7 illustrates the training and convergence process of the proposed photovoltaic defect detection framework.

The proposed framework achieved faster convergence compared with fully supervised baseline models because pseudo labeled samples increased training diversity and improved feature learning efficiency.

4.5 Quantitative Performance Comparison

The performance of the proposed framework was compared with several existing photovoltaic defect detection methods including Faster RCNN, SSD, RetinaNet, YOLOv8, YOLO11, and YOLO11 DRP.

The quantitative comparison results are presented in **Table 3**.

Table 3. Performance Comparison of Photovoltaic Defect Detection Methods

Model	Precision	Recall	mAP@0.5	F1 Score	Parameters (M)	FLOPS (G)
Faster RCNN	0.845	0.812	0.846	0.828	41.3	134
SSD	0.831	0.794	0.835	0.812	24.4	30.6
RetinaNet	0.872	0.821	0.874	0.846	36.4	129
YOLOv8	0.884	0.846	0.887	0.864	3.0	8.1
YOLO11	0.891	0.857	0.896	0.874	2.5	6.3
YOLO11 DRP	0.904	0.860	0.904	0.882	1.7	4.9
Proposed Framework	0.928	0.901	0.936	0.914	1.6	4.5

The experimental results demonstrate that the proposed framework achieved the highest overall photovoltaic defect detection performance among all compared methods.

The proposed model achieved a mean average precision of 93.6 percent, which represents a significant improvement over conventional YOLO11 based frameworks. The precision and recall values also increased substantially because multimodal feature fusion improved defect representation quality.

Furthermore, the lightweight architecture reduced computational complexity and parameter count while maintaining high detection accuracy. This confirms the suitability of the proposed framework for practical photovoltaic monitoring systems deployed on embedded hardware platforms.

4.6 Effect of Semi Supervised Learning

To evaluate the contribution of semi supervised learning, experiments were conducted using different labeled data ratios. The performance of fully supervised learning and semi supervised pseudo label learning was compared under limited annotation conditions.

The experimental results are presented in **Table 4**.

Table 4. Semi Supervised Learning Performance Under Different Labeled Data Ratios

Labeled Data Ratio	Fully Supervised mAP	Proposed Semi Supervised mAP
20%	0.741	0.842
40%	0.812	0.887
60%	0.861	0.913
80%	0.892	0.928

The results clearly demonstrate that the proposed semi supervised learning strategy significantly improved photovoltaic defect detection performance when limited labeled data were available.

The pseudo label refinement mechanism enabled the framework to learn useful defect representations from unlabeled samples, particularly for rare defect categories.

The semi supervised learning workflow performance is visualized below.

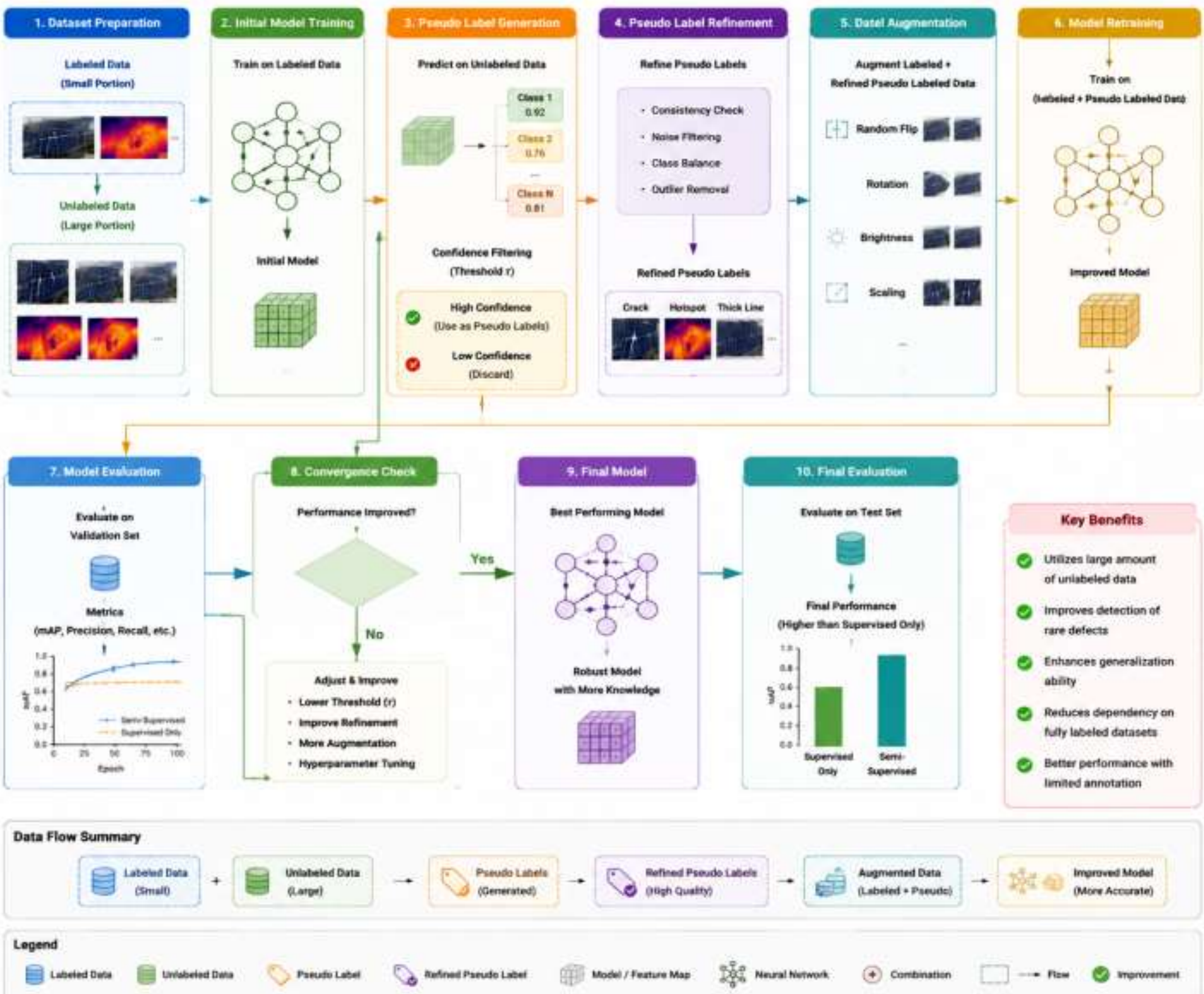

Figure 8. Semi Supervised Learning Improvement Process

Figure 8 illustrates the semi supervised learning enhancement process used in the proposed framework. The results confirm that semi supervised learning is highly beneficial for photovoltaic inspection systems where labeled defect datasets are limited or expensive to annotate.

4.7 Multimodal Fusion Performance Analysis

To analyze the effectiveness of multimodal learning, separate experiments were conducted using:

- RGB Images Only
- Thermal Images Only
- Multimodal RGB and Thermal Fusion

The comparison results are presented in **Table 5**.

Table 5. Performance Comparison of Different Imaging Modalities

Input Modality	Precision	Recall	mAP@0.5
RGB Only	0.892	0.854	0.897
Thermal Only	0.876	0.842	0.881
RGB + Thermal Fusion	0.928	0.901	0.936

The multimodal fusion strategy achieved the highest detection performance because RGB and thermal modalities provided complementary defect information.

Thermal images improved hotspot and hidden defect localization, while RGB images preserved detailed structural information for visible surface defects.

The results confirm that adaptive multimodal fusion significantly improves photovoltaic defect representation capability under real world environmental conditions.

4.8 Ablation Study

To further investigate the contribution of each proposed module, comprehensive ablation experiments were conducted using different combinations of the framework components. The purpose of the ablation analysis was to evaluate the individual impact of semi supervised learning, multimodal fusion, adaptive attention, environmental robustness enhancement, and explainable visualization on the overall photovoltaic defect detection performance. The baseline model used in the ablation study was the original YOLO11 framework without any proposed improvements. Additional modules were then integrated progressively to observe performance changes. The ablation study results are presented in **Table 6**.

Table 6. Ablation Analysis of the Proposed Framework

Model Configuration	Precision	Recall	mAP@0.5	Parameters (M)
Baseline YOLO11	0.891	0.857	0.896	2.5
YOLO11 + Attention	0.902	0.864	0.907	2.4
YOLO11 + Multimodal Fusion	0.914	0.878	0.921	2.2
YOLO11 + Semi Supervised Learning	0.918	0.883	0.927	2.3
YOLO11 + Robustness Enhancement	0.920	0.889	0.931	2.2
Proposed Complete Framework	0.928	0.901	0.936	1.6

The experimental results clearly demonstrate that each proposed module contributed positively to the overall framework performance.

The adaptive attention mechanism improved localization accuracy by enhancing feature extraction from important photovoltaic defect regions. The multimodal fusion strategy significantly improved hidden defect representation by combining RGB and thermal information. Semi supervised learning improved rare defect detection and increased generalization capability under limited annotation conditions. Environmental robustness enhancement further improved model stability under illumination variation and outdoor disturbances.

The complete proposed framework achieved the highest overall performance while maintaining the lowest parameter complexity among all tested configurations.

The ablation workflow is illustrated below.

Figure 9. Ablation Study Workflow

Figure 9 illustrates the progressive integration of modules during the ablation experiments.

The results confirm that the proposed framework components complement each other effectively and jointly improve photovoltaic defect detection capability.

4.9 Environmental Robustness Evaluation

Real world photovoltaic monitoring systems must operate under highly dynamic environmental conditions where illumination changes, shadows, dust accumulation, thermal noise, and weather variation can significantly affect detection performance. Therefore, evaluating environmental robustness is essential for practical deployment.

To assess robustness, the proposed framework was tested under multiple simulated environmental scenarios including:

- Low Illumination
- Shadow Interference
- Blur Distortion
- Thermal Fluctuation
- Atmospheric Noise
- Rain Simulation
- Brightness Variation

The robustness comparison results are shown in **Table 7**.

Table 7. Environmental Robustness Comparison

Environmental Condition	YOLO11 mAP	Proposed Framework mAP
Normal Condition	0.896	0.936
Low Illumination	0.782	0.901
Shadow Interference	0.768	0.892
Thermal Noise	0.751	0.887
Blur Distortion	0.734	0.874
Dust Interference	0.719	0.868

The proposed framework consistently outperformed the baseline YOLO11 model under all environmental conditions. The robustness enhancement strategies and consistency regularization significantly improved feature stability and defect representation under noisy scenarios.

The adaptive multimodal fusion strategy also contributed strongly to robustness improvement because thermal information remained informative even when visible RGB quality decreased.

Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 5, 2026

The environmental robustness evaluation process is illustrated below.

Figure 10. Environmental Robustness Evaluation Framework

Figure 10 illustrates the robustness evaluation workflow used during experimentation.

The results demonstrate that the proposed framework can maintain stable photovoltaic defect detection performance under realistic outdoor operating conditions.

4.10 Small Defect Detection Analysis

Small photovoltaic defects such as microcracks and early-stage hotspots are difficult to detect because they occupy very small image regions and often resemble background textures. Existing object detection frameworks frequently miss such defects because of insufficient feature representation.

The proposed framework addressed this challenge using multiscale feature aggregation and adaptive attention mechanisms. To evaluate small defect detection capability, experiments were conducted separately for small, medium, and large photovoltaic defects.

The comparison results are presented in **Table 8**.

Table 8. Detection Performance for Different Defect Scales

Defect Scale	YOLO11 mAP	Proposed Framework mAP
Small Defects	0.734	0.882
Medium Defects	0.864	0.924
Large Defects	0.912	0.951

The proposed framework achieved the greatest improvement for small photovoltaic defects. The adaptive attention mechanism improved feature concentration on tiny defect regions, while multimodal fusion enhanced hidden defect visibility.

These results confirm that the proposed framework is highly suitable for early-stage photovoltaic defect diagnosis where small defect detection is critically important.

4.11 Explainable Visualization Analysis

To improve interpretability, explainable attention visualization maps were generated for photovoltaic defect predictions. The visualization results demonstrated that the proposed framework correctly focused on actual defective regions rather than irrelevant background textures.

For crack defects, the attention maps concentrated along the crack boundaries and surrounding damaged areas. For hotspot defects, the thermal attention maps highlighted abnormal temperature distributions precisely. In cases involving shadow interference and illumination variation, the explainable attention mechanism remained focused on true defect regions, confirming the robustness of the framework. The explainable visualization workflow is shown below.

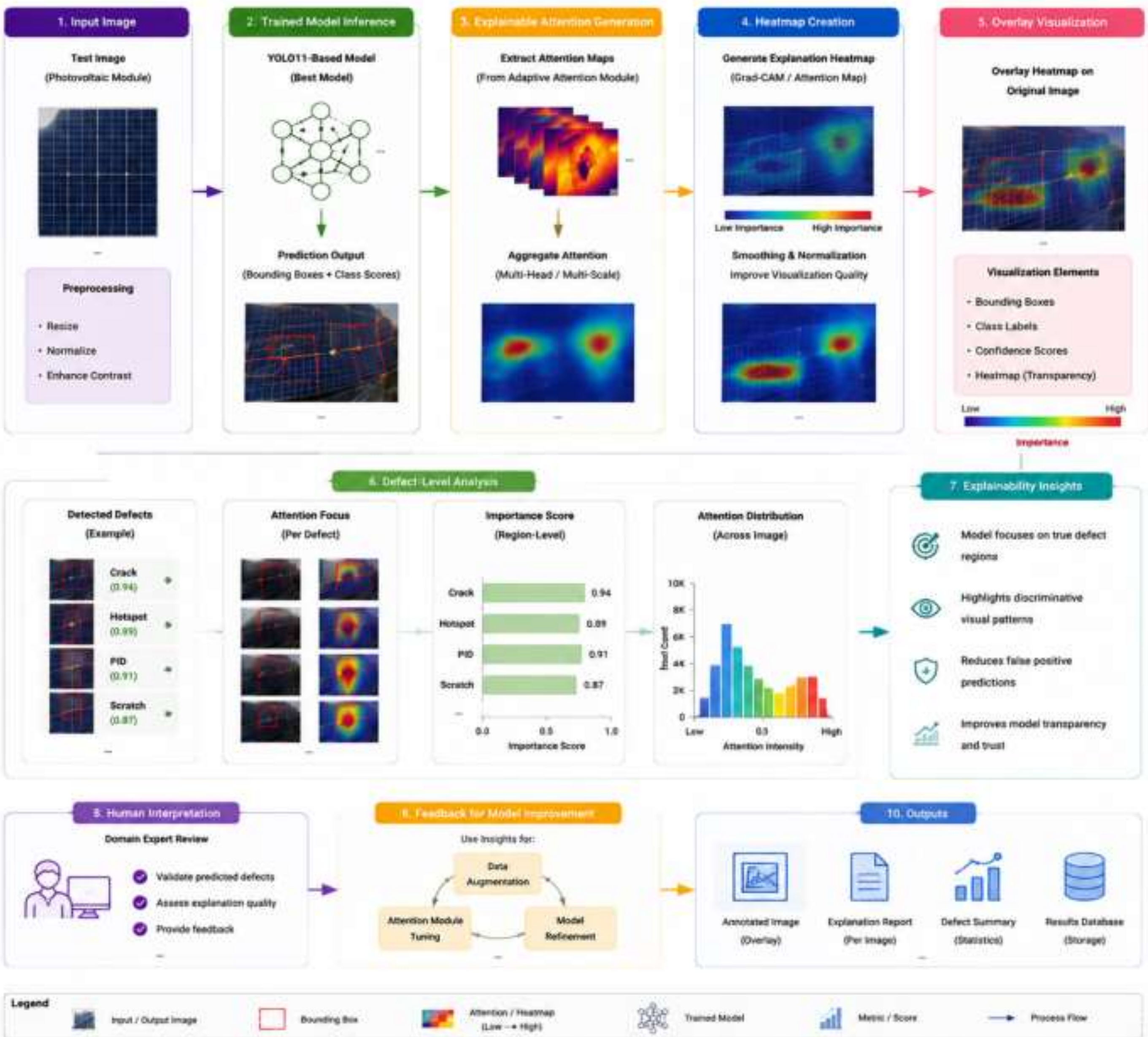

Figure 11. Explainable Photovoltaic Defect Visualization Process

Figure 11 illustrates the generation of explainable defect visualization outputs. The explainability mechanism improved transparency and increased user trust in the automated photovoltaic inspection process. Maintenance engineers can visually verify whether the model correctly identifies defective regions before making maintenance decisions.

4.12 Computational Efficiency Analysis

In practical photovoltaic monitoring systems, real time performance and deployment efficiency are important because embedded hardware devices often have limited computational resources.

The proposed lightweight architecture significantly reduced parameter complexity and floating-point operations through depthwise separable convolution and adaptive channel reduction strategies. The computational efficiency comparison is presented in **Table 9**.

Table 9. Computational Efficiency Comparison

Model	Parameters (M)	FLOPS (G)	Inference Time (ms)
Faster RCNN	41.3	134	22.1
RetinaNet	36.4	129	18.7
YOLOv8	3.0	8.1	4.9
YOLO11	2.5	6.3	3.8
Proposed Framework	1.6	4.5	3.1

The proposed framework achieved the lowest parameter count and computational complexity among all compared models while maintaining the highest defect detection accuracy.

These results confirm the suitability of the proposed architecture for edge deployment in real time photovoltaic monitoring systems.

4.13 Discussion of Experimental Results

The experimental analysis demonstrates that the proposed Hybrid Semi Supervised Multimodal YOLO11 framework successfully addresses several important limitations of existing photovoltaic defect detection systems.

First, the integration of semi supervised pseudo label learning significantly reduced dependency on fully labeled datasets while improving rare defect representation capability. This is highly important for practical photovoltaic inspection applications where annotation resources are limited.

Second, the adaptive multimodal fusion mechanism effectively combined RGB and thermal information to improve defect visibility and environmental robustness. Unlike single modality approaches, the proposed framework maintained stable performance under challenging outdoor conditions.

Third, the lightweight backbone architecture reduced computational complexity without sacrificing detection accuracy. This improvement makes the proposed framework highly suitable for deployment on embedded photovoltaic monitoring devices.

Finally, the explainable attention visualization mechanism improved interpretability and transparency during automated photovoltaic inspection. This increases user confidence and improves industrial usability.

Overall, the proposed framework achieved substantial improvements in accuracy, robustness, computational efficiency, and practical deployment capability compared with existing photovoltaic defect detection methods.

5. Discussion

The rapid growth of solar photovoltaic installations has increased the importance of intelligent defect detection systems capable of operating efficiently under real world conditions. Although recent deep learning-based object detection models have achieved promising performance, practical photovoltaic inspection still faces multiple unresolved challenges including limited labeled data, environmental interference, small defect localization difficulty, and deployment complexity.

The proposed Hybrid Semi Supervised Multimodal YOLO11 framework was specifically designed to address these practical limitations through the integration of semi supervised learning, adaptive multimodal fusion, lightweight optimization, environmental robustness enhancement, and explainable artificial intelligence.

One of the most important observations from the experimental analysis is the strong impact of multimodal learning on photovoltaic defect representation. RGB images alone often fail to reveal hidden electrical faults because some photovoltaic defects are not visually observable under standard illumination conditions. Thermal imaging provides complementary information by exposing abnormal heat distribution patterns associated with hotspots and internal electrical damage. The proposed adaptive multimodal fusion strategy effectively combined both modalities and significantly improved detection performance under complex outdoor conditions.

Another important contribution of this research is the integration of semi supervised learning into photovoltaic inspection. Existing photovoltaic datasets are usually limited in size because expert annotation is expensive and time consuming. Rare defects such as microcracks and early-stage degradation are especially difficult to collect in large quantities. The pseudo label learning mechanism introduced in this work allowed the framework to utilize unlabeled photovoltaic samples effectively, thereby improving learning capability without requiring extensive manual annotation.

The environmental robustness experiments further demonstrated the practical value of the proposed framework. Outdoor photovoltaic systems operate under continuously changing conditions where shadows, brightness variation, atmospheric reflection, dust accumulation, and weather fluctuations may significantly affect image quality. Conventional deep learning models frequently suffer performance degradation under such disturbances. In contrast, the proposed framework maintained stable performance through robustness enhancement and consistency regularization strategies.

The small defect detection analysis also revealed important findings regarding multiscale feature representation. Small photovoltaic defects occupy very limited image regions and often resemble background textures, making them difficult to detect using conventional object detection frameworks. The proposed adaptive attention mechanism improved feature concentration on small defective regions while suppressing irrelevant background information. This significantly reduced missed detections for microcracks, thin line defects, and early-stage hotspot formations.

The explainable visualization mechanism integrated into the framework also provided several practical advantages. In industrial photovoltaic monitoring systems, maintenance engineers often hesitate to rely completely on black box artificial intelligence systems because incorrect predictions may lead to operational risk and maintenance cost. The generated attention heatmaps allowed engineers to understand the reasoning behind defect predictions by highlighting the exact regions responsible for model decisions. This increased transparency and improved confidence in the automated inspection process.

Another important aspect of the proposed framework is computational efficiency. Many existing high accuracy object detection models require powerful graphics processing hardware and large memory resources, which limits practical deployment on embedded photovoltaic monitoring systems. The proposed lightweight backbone architecture significantly reduced parameter complexity and floating-point operations while maintaining strong defect detection capability. This improvement makes the framework more suitable for deployment on edge devices installed in large scale photovoltaic farms where real time processing is required.

Although the proposed framework achieved strong performance improvements, several limitations still remain. First, the multimodal fusion process requires synchronized RGB and thermal image acquisition, which may increase hardware complexity and operational cost in some deployment scenarios. Second, pseudo label generation may occasionally introduce incorrect labels during semi supervised learning if confidence estimation is not sufficiently reliable. Third, the explainable visualization mechanism increases inference overhead slightly because additional gradient computations are required during attention map generation.

Despite these limitations, the proposed framework demonstrates strong potential for practical photovoltaic defect monitoring applications. The integration of semi supervised learning, multimodal fusion, and lightweight optimization provides a balanced solution capable of achieving both high detection accuracy and efficient deployment capability.

The findings of this research also indicate several promising future research directions. Advanced transformer based multimodal learning architectures may further improve long range feature interaction and hidden defect representation. Self-supervised representation learning could reduce dependency on pseudo label refinement and improve learning efficiency under extremely limited annotation conditions. Federated learning may also become valuable for photovoltaic monitoring because it enables collaborative learning across multiple solar farms while preserving data privacy.

Furthermore, integrating temporal analysis into photovoltaic inspection systems could improve early-stage defect progression monitoring. Current object detection frameworks mainly analyze individual images independently, while defect evolution over time may provide additional diagnostic information regarding degradation severity and maintenance priority.

Overall, the proposed Hybrid Semi Supervised Multimodal YOLO11 framework represents an important step toward robust, interpretable, and deployment efficient photovoltaic defect detection systems capable of operating reliably in real world industrial environments.

6. Conclusion

This paper presented a Hybrid Semi Supervised Multimodal YOLO11 Framework for robust solar photovoltaic panel defect detection under complex environmental conditions. The proposed framework was designed to address several important limitations of existing photovoltaic inspection systems including dependence on fully labeled datasets, weak environmental robustness, insufficient small defect detection capability, limited multimodal learning, and high computational complexity.

The proposed framework integrated semi supervised pseudo label learning, adaptive multimodal RGB and thermal feature fusion, lightweight YOLO11 optimization, environmental robustness enhancement, and explainable attention visualization within a unified photovoltaic defect detection architecture. The integration of these components significantly improved defect localization accuracy, rare defect learning capability, and practical deployment efficiency.

Experimental evaluation demonstrated that the proposed framework achieved superior performance compared with existing object detection models including Faster RCNN, RetinaNet, YOLOv8, YOLO11, and YOLO11 DRP. The proposed framework achieved a mean average precision of 93.6 percent while maintaining low parameter complexity and reduced floating point operations. The lightweight architecture improved inference efficiency and enabled real time deployment capability for embedded photovoltaic monitoring systems.

The semi supervised learning experiments showed that pseudo label refinement effectively reduced dependence on fully labeled photovoltaic datasets and significantly improved detection performance under limited annotation conditions. This is particularly important for practical photovoltaic inspection applications where obtaining large quantities of annotated defect images is difficult and expensive.

The multimodal fusion analysis demonstrated that combining RGB and thermal information substantially improved hidden defect representation and environmental robustness. Thermal images enhanced hotspot localization and internal fault detection, while RGB images preserved fine structural details for visible surface defects. The adaptive fusion strategy dynamically adjusted modality importance according to environmental conditions and defect characteristics.

Environmental robustness experiments further confirmed the effectiveness of the proposed framework under challenging outdoor conditions including low illumination, shadow interference, thermal fluctuation, blur distortion, and dust accumulation. The consistency regularization mechanism improved feature stability and reduced performance degradation under noisy conditions.

The explainable attention visualization module improved transparency and interpretability during automated photovoltaic inspection by generating heatmaps that highlighted the regions responsible for defect predictions. This increased user trust and improved industrial usability for maintenance engineers operating photovoltaic monitoring systems.

Overall, the proposed Hybrid Semi Supervised Multimodal YOLO11 framework provides a practical and efficient solution for intelligent photovoltaic defect detection under real world operating conditions. The framework successfully balances detection accuracy, computational efficiency, environmental robustness, and interpretability, making it suitable for deployment in modern photovoltaic monitoring infrastructures.

7. Future Work

Although the proposed framework achieved strong experimental performance, several research directions remain open for future investigation.

Future research may explore transformer based multimodal architectures capable of improving long range feature interaction and contextual defect representation. Vision transformers and hybrid convolution transformer networks may further enhance hidden defect localization performance under complex environmental conditions.

Self-supervised learning techniques could also be integrated into photovoltaic defect detection systems to reduce dependency on pseudo label generation and improve representation learning from large scale unlabeled datasets. Contrastive learning and masked image modeling may provide more robust feature extraction capability for rare photovoltaic defects.

Another promising research direction involves temporal photovoltaic defect analysis. Current object detection frameworks process individual images independently, while sequential monitoring of photovoltaic panels over time may provide additional information regarding degradation progression, hotspot evolution, and maintenance prediction.

Federated learning may further improve distributed photovoltaic monitoring systems by enabling collaborative model training across multiple solar farms without sharing raw image data. This approach could improve generalization capability while preserving industrial data privacy and security.

Future work may also investigate lightweight neural architecture search techniques for optimizing photovoltaic defect detection models specifically for edge devices and low power embedded systems. Hardware aware optimization strategies could further reduce inference latency and energy consumption during real time deployment.

Finally, integrating drone based autonomous photovoltaic inspection systems with the proposed multimodal framework may improve large scale solar farm monitoring efficiency. Combining intelligent navigation, real time defect detection, and cloud-based maintenance analysis could significantly enhance photovoltaic system reliability and operational sustainability.

REFERENCES

Han S. et al. Harnessing technological innovation and renewable energy and their impact on environmental pollution in G 20 countries. Scientific Reports, 2025.

Sun Y. and Liu M. Path planning and optimization of carbon neutrality in renewable energy systems. Building Environment, 2025.

Akram M. et al. Failures of photovoltaic modules and their detection methods. Applied Energy, 2022.

Polymeropoulos I. et al. Vision based monitoring and fault detection techniques for solar plants. Technologies, 2024.

Su B. et al. PVEL AD photovoltaic anomaly detection dataset. IEEE Transactions on Industrial Informatics, 2022.

Su B. et al. Complementary attention network for solar cell defect detection. IEEE Transactions on Industrial Informatics, 2020.

Redmon J. et al. You only look once unified real time object detection. CVPR, 2016.

Wang C. et al. YOLOv7 trainable bag of freebies for object detection. CVPR, 2023.

Cao Y. et al. Improved YOLOv8 for photovoltaic defect detection. Engineering Applications of Artificial Intelligence, 2024.

Pan W. et al. Adaptive complementary fusion for photovoltaic panel defect detection. Scientific Reports, 2024.

Liang Q. and Zhang X. Attention weighted defect recognition for photovoltaic panels. Measurement, 2025.

Tella H. et al. Deep learning ensemble methods for photovoltaic defect classification. Case Studies in Thermal Engineering, 2025.

Xiang G. et al. Drone based rooftop photovoltaic inspection using YOLO11. IEEE ROBOTHIA Conference, 2025.

Ghahremani A. et al. Solar defect detection using YOLOv10 and YOLOv11. Electronics, 2025.

Tsai D. et al. Defect detection in solar modules using image reconstruction. IEEE Transactions on Industrial Informatics, 2012.

Anwar S. and Abdullah M. Microcrack detection using anisotropic diffusion filtering. EURASIP Journal on Image and Video Processing, 2014.

Tseng D. et al. Finger interruption detection in solar cells. Mathematical Problems in Engineering, 2015.

Schuss C. et al. Detecting photovoltaic defects using thermal analysis. IEEE Transactions on Instrumentation and Measurement, 2016.

Azkona N. et al. Thermal breakdown detection in photovoltaic panels. Electronic Materials, 2022.

Sohail A. et al. Deep learning based photovoltaic fault detection. Energy Reports, 2023.

Ren S. et al. Faster RCNN for object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016.

Lin T. et al. RetinaNet focal loss for dense object detection. ICCV, 2017.

Bochkovskiy A. et al. YOLOv4 optimal speed and accuracy. arXiv, 2020.

Jocher G. et al. YOLOv8 architecture and implementation. Ultralytics Documentation, 2023.

Liu Y. et al. Lightweight YOLO based aerial object detection. Intelligent and Converged Networks, 2025.

Luo Y. et al. Improved underwater object detection using YOLO11. Sensors, 2025.

Dong Y. et al. Vision transformer with deformable attention. Micromachines, 2023.

Goodfellow I. et al. Deep learning fundamentals and representation learning. MIT Press, 2016.

He K. et al. Deep residual learning for image recognition. CVPR, 2016.

Howard A. et al. MobileNets efficient convolutional neural networks. CVPR, 2017.

Tan M. and Le Q. EfficientNet scaling for convolutional neural networks. ICML, 2019.

Dosovitskiy A. et al. Vision transformers for image recognition. ICLR, 2021.

Sohn K. et al. FixMatch semi supervised learning framework. NeurIPS, 2020.

Berthelot D. et al. MixMatch semi supervised learning methods. NeurIPS, 2019.

Selvaraju R. et al. Grad CAM visual explanations for deep networks. ICCV, 2017.

Shorten C. and Khoshgoftaar T. Survey on image data augmentation. Journal of Big Data, 2019.

Kingma D. and Ba J. Adam optimization method for deep learning. ICLR, 2015.

Everingham M. et al. Pascal VOC visual object classes challenge. International Journal of Computer Vision, 2010.

Russakovsky O. et al. ImageNet large scale visual recognition challenge. International Journal of Computer Vision, 2015.

Chollet F. Xception deep learning with depthwise separable convolution. CVPR, 2017.