Computer Vision Capabilities for Detecting and Classifying Solar Panel Faults

[1]Mohammed N.S., [1]Uonis M.M., [2]Alsaif O.I.

[1]Mosul University
Mosul, Iraq
[2]Northern Technical University
Polytechnique College-Mosul
Mosul, Iraq

Abstract: The main objectives of this research are to fill a gap in PV inspection systems, which consist of convolutional detectors with a limited amount of global context or transformer classifiers with weak localization. Defects micro-cracks, broken electrodes, and degradation regions decrease the PV energy efficiency considerably, and this encourages the use of an automated computer vision algorithm for the classification of these defects extracted from electroluminescence (EL) images. These objectives were achieved by solving the following task: a novel dual-stage architecture is used to solve these tasks, which are performed sequentially, YOLOv8m followed by the Swin Transformer architecture. In comparison with previous CNN-Transformer PV models that combine features by concatenation or averaging with predefined weights, the proposed model implements a learnable cross-attention fusion probability merging ensemble, which dynamically weights the localization and global attention evidence for each defect. YOLOv8m localizes the exact positions of defects, then the Swin Transformer analyzes the region by window-swiping self-attention to return a long-range degradation pattern that CNNs can't pass. The model was trained using EL images in 12 classes, such as, intact cells, cracks, point defects, finger damage, and EL images were augmented to overcome the class imbalance problem and cross-conditioning generalization. The most important results are a macro-averaged F1-score of 94.05% and 95.40% accuracy, with per-class F1 ranging from 98.98% (Mono-crystalline) to 89.04% (Micro-crack). The significance of the obtained results lies in the use of integrative localization-driven YOLOv8m convolutional features with transformer global attention and learnable cross-attention fusion instead of static cross-attention fusion, providing a new and more robust implementation to perform automated PV quality assessment than a single branch.

Keywords: cross-attention fusion (CAF), computer vision, deep learning, electroluminescence imaging (EL), ensemble deep learning, solar cell defect classification, swin-transformer (ST), YOLOv8m.

DOI: https://doi.org/10.52254/1857-0070.2026.3-71.15
UDC: 004.93:004.8:621.383

Capacități de viziune computerizată pentru detectarea și clasificarea defecțiunilor panourilor solare

[1]Mohammed N.S., [1]Uonis M.M., [2]Alsaif O.I.
[1]Universitatea Mosul, Mosul, Irak
[2]Universitatea Tehnică de Nord (Colegiul Politehnic), Mosul, Irak

Abstract: Principalele obiective ale acestei cercetări sunt de a umple o lacună în sistemele de inspecție fotovoltaică, care constau din detectoare convoluționale cu o cantitate limitată de context global sau clasificatoare de transformatoare cu localizare slabă. Micro-fisurile defectelor, electrozii rupți și regiunile de degradare reduc considerabil eficiența energetică a fotovoltaicului, iar acest lucru încurajează utilizarea unui algoritm automat de viziune computerizată pentru clasificarea acestor defecte extrase din imaginile de electroluminescență (EL). Aceste obiective au fost atinse prin rezolvarea următoarei sarcini: o nouă arhitectură cu două etape este utilizată pentru a rezolva aceste sarcini, care sunt efectuate secvențial, YOLOv8m urmată de arhitectura Swin Transformer. În comparație cu modelele fotovoltaice anterioare CNN-Transformer care combină caracteristici prin concatenare sau mediere cu ponderi predefinite, modelul propus implementează un ansamblu de fuziune a probabilității de fuziune a atenției încrucișate, care poate fi învățat, și care ponderează dinamic localizarea și dovezile atenției globale pentru fiecare defect. YOLOv8m localizează pozițiile exacte ale defectelor, apoi Swin Transformer analizează regiunea prin auto-atenție prin glisare pe fereastră pentru a returna un model de degradare pe termen lung pe care CNN-urile nu îl pot trece. Modelul a fost antrenat folosind imagini EL în 12 clase, cum ar fi celule intacte, fisuri, defecte punctuale, deteriorarea degetelor, iar imaginile EL au fost amplificate pentru a depăși problema dezechilibrului de clase și generalizarea condiționării încrucișate. Cele mai importante rezultate sunt un

scor F1 macro-mediat de 94,05% și o precizie de 95.40%, cu un F1 per clasă variind de la 98.98% (monocristalin) la 89.04% (microfisură). Semnificația rezultatelor obținute constă în utilizarea caracteristicilor convoluționale YOLOv8m bazate pe localizare integrativă, cu atenție globală a transformatorului și fuziune încrucișată învățabilă în loc de fuziune statică încrucișată, oferind o implementare nouă și mai robustă pentru a efectua evaluarea automată a calității fotovoltaice decât o singură ramificare.
Keywords: fuziune încrucișată (CAF), viziune computerizată, învățare profundă, imagistică electroluminescentă (EL), învățare profundă de ansamblu, clasificarea defectelor celulelor solare, transformator swin (ST); YOLOv8m.

Возможности компьютерного зрения для обнаружения и классификации неисправностей солнечных панелей

[1]Mohammed N.S., [1]Uonis M.M., [2]Alsaif O.I.
[1]Университет Мосула, Мосула, Ирак
[2]Северный технический университет (Политехнический колледж), Мосула, Ирак

Аннотация: Основные цели данного исследования — восполнить пробел в системах контроля фотоэлектрических элементов, которые состоят из сверточных детекторов с ограниченным объемом глобального контекста или трансформерных классификаторов со слабой локализацией. Микротрещины, сломанные электроды и области деградации значительно снижают энергоэффективность фотоэлектрических элементов, что обуславливает необходимость использования автоматизированного алгоритма компьютерного зрения для классификации этих дефектов, извлеченных из изображений электролюминесценции (ЭЛ). Эти цели были достигнуты путем решения следующей задачи: для решения этих задач используется новая двухэтапная архитектура, которая выполняется последовательно: YOLOv8m, а затем архитектура Swin Transformer. По сравнению с предыдущими моделями CNN-Transformer для фотоэлектрических элементов, которые объединяют признаки путем конкатенации или усреднения с предопределенными весами, предлагаемая модель реализует обучаемый ансамбль слияния вероятностей перекрестного внимания, который динамически взвешивает данные локализации и глобального внимания для каждого дефекта. YOLOv8m точно определяет местоположение дефектов, затем Swin Transformer анализирует область с помощью механизма самовнимания с использованием окна, чтобы получить дальний паттерн деградации, который не могут обработать сверточные нейронные сети (CNN). Модель обучалась с использованием изображений электролюминесценции (EL) в 12 классах, таких как целые клетки, трещины, точечные дефекты, повреждения пальцев, а изображения EL были дополнены для преодоления проблемы дисбаланса классов и обобщения с помощью перекрестного обусловливания. Наиболее важными результатами являются макроусредненный показатель F1-меры 94,05% и точность 95,40%, при этом показатель F1 для каждого класса варьировался от 98,98% (монокристаллический) до 89,04% (микротрещина). Значимость полученных результатов заключается в использовании интегративных локализационно-ориентированных сверточных признаков YOLOv8m с глобальным вниманием трансформера и обучаемым слиянием перекрестного внимания вместо статического слияния перекрестного внимания, что обеспечивает новую и более надежную реализацию для автоматизированной оценки качества фотоэлектрических систем, чем использование одной ветви.
Ключевые слова: перекрестное слияние внимания (CAF), компьютерное зрение, глубокое обучение, электролюминесцентная визуализация (EL), ансамблевое глубокое обучение, классификация дефектов солнечных элементов; swin-transformer (ST); YOLOv8m.

INTRODUCTION

In recent years, the consumption of renewable energy sources has grown significantly to achieve the worldwide goal of becoming a “net zero emitter” by the year 2050, particularly PV solar systems. There are several issues that can obstruct the durability of solar power systems, such as poor-quality parts, environmental effects like dust buildup, corrosion, bird excrement, and micro-cracks that can cause excessive maintenance requirements and system inefficiencies [2][3]. Electroluminescence (EL) imaging has been proven to be one of the most efficient methods to find defects inside the solar cells that could not be detected based on the traditional inspection method [4][5]. When a large number of EL images are generated, it is necessary to use AI algorithms for defect detection [6, 7].

Amongst these methods, YOLOv8 offers an excellent technique in local feature extraction and accurate and fast defect detection [6][8][9]. However, convolutional neural networks have some constraints on their ability to capture global context and long-term dependency [10][11]. In

this sense, Swin Transformer is capable of doing efficient computation in modeling local and global representations by introducing hierarchical self-attention and multi-scale feature learning mechanisms [12][13][14]. Therefore, the fusion of YOLOv8 and Swin Transformer has verified itself to be an effective approach in improving the accuracy and robustness of defect detection on EL images, particularly for defects that have subtle texture features, which are similar to the structure of normal solar cells [15][16].

Challenges: Micro-cracks and defects are the most current problems with solar panels, as it is difficult to detect these defects from conventional inspection operations [3]. The manual evaluation of hundreds of images in the EL inspection procedure can be time-consuming and disagreeable [4]. CNN has achieved good performance in capturing local features, but it can't capture the global context in the image[17][18]. Furthermore, the defects of solar panels exhibit small-scale, different sizes, and a distributed nature in various locations, which are sometimes difficult to differentiate visually from non-defect regions[19][20]. Solar cell defects frequently involve fine-grained patterns, non-regular spatial distribution, and large-scale variations, making it difficult to deal with using conventional pattern-based convolution. The defects in solar cells have fine-grained patterns, non-regular spatial distribution, and large-scale variations, making it challenging to handle using conventional pattern-based convolution.

Solar cell defects often exhibit fine-grained patterns, irregular spatial distributions, and significant scale variations, which pose challenges for conventional convolution-based architectures [20],

The main objective of this proposed methodology is to create a combined model for electroluminescence image analysis and classification by integrating YOLOv8m with ST. These two models are expected to complement each other well, benefiting the effectiveness of YOLOv8m in extracting features from multi-scale local regions and Swin Transformer in understanding the image context and long-range relationships inside the image.

The main contribution is the proposed hybrid model with the combination of YOLOv8m and Swin Transformer applied in the case of the classification and analysis of electrofluorescence images (EL) of solar cells. YOLOv8m is used for extracting multiscale local features in C3, C4, and C5 feature layers to detect micro-cracks and other discontinuities in solar cell images, Whereas, ST benefits from a self-attention mechanism to learn global contexts of objects while retaining their local properties. A hybrid approach of two models can be used to enhance the strength and immunity of the classification system to noise.

LITERATURE REVIEW

Aktouf et al. [21] developed a defect detection system based on the YOLOv10 deep learning model to determine defects in EL images of solar cells, based on Compact Inverted Blocks (CIB) and Partial Self-Attention (PSA) modules to improve the feature extraction process without significantly increasing computation costs. The model was applied to 10,500 EL images for 12 different defect types.

Their results—98.5% accuracy and 98.5% mAP@0.5—and several defect categories were detected. For example, black cores, corners, fragments, scratches, and short circuits all received a 100%. To achieve accurate detection, the ability to function properly in a variety of defect types, and efficient real-time operation. It also provides a balance between accuracy and computation speed.

The drawbacks are that it cannot manage very well with star cracks and thick lines. In addition, the YOLOv10x version requires a lot of computing power and expensive hardware.

Xue et al. [22] Objective Design a lightweight algorithm for defect detection of PV modules in real-time on edge devices (UAVs). Model YOLOv8 is improved using three modifications, swapping out the SPPF module with the SPPELAN module to reduce computation; adding an ECA attention mechanism to bolster the extraction of fine-detail defect features; and utilizing a CARAFE intelligent upsampling factor in the Neck area to maintain fine detail features. Data 743 infrared images of 7 defect types (expanded to 2,971). Despite using just 3.6 million parameters and running at 416 FPS, its performance is 2.6% higher than the original YOLOv8 in terms of mAP@0.5. Benefits, High accuracy and low computation; can be deployed on UAVs; few parameters and better than YOLOv5s and YOLOv8s. Limitations, Private dataset; does not appear to be publicly available; not tested against EfficientDet or YOLO-NAS; only tested on infrared images and not EL.

Wang et. al [23] proposed an algorithm for detecting defects in rigid rolled-bar surfaces based on the Swin-Transformer network

with YOLOv5. For improved defect recognition, the proposed model combined the highly effective feature extraction ability of Swin-Transformer with the efficient object detection ability of YOLOv5. The results showed that the model achieved high accuracy in defect detection, especially for small and complex defects, with low compute costs and rapid inference speed. The key benefits are the representation of features, ease of model design, and applicability in real-time industrial applications. The study, however, was mainly evaluated using hot-rolled steel strip data sets, and may not be representative of other industrial defect inspection applications. Also, the model's functionality is dependent on adequate labeled data for training. Therefore, the system may not perform well in the detection of subtle defects that have low contrast.

Wang et al. [24] proposed a model to improve small nodule detection and avoid false positives by introducing a two-stage approach for lung nodule detection from computed tomography (CT) images. The proposed framework has stages: a U-Net for high-resolution lung parenchyma segmentation and a YOLOv8s model with Swin Transformer blocks in the Neck model the global context and long-range dependencies. The SIoU loss function is used to increase the accuracy of the boundary box prediction. The model was trained and tested using 888 images from the LUNA16 database and 308 images from Tianjin Chest Hospital.

The results showed that the proposed model overcame the other models, such as: YOLOv9s model, YOLOv10s model, and Faster R-CNN model, in terms of accuracy: 0.898 on LUNA16 data; and effort 0.851 on average MAP50 accuracy. The model was able to generalize to external clinical data with an accuracy of 0.855, a retrieval of 0.872, and an MAP50 of 0.862 on hospital data. In the first phase, the lung segmentation method was used to effectively suppress background noise, and the detection accuracy of small nodules was improved by 2.1% over the conventional IoU function with SIoU. So, the generalizability of the model to multi-institutional datasets has not been confirmed; the performance is low in nodules near bone, or those with fine features, and the model performance depends on the quality of the data annotations.

Rodrigo et al. al[25] Several models were tested for object detection on solar panels via drones. Some of the object detectors that were considered include Faster R-CNN with ResNet50, MobileNetV3, RetinaNet, YOLOv5, YOLOv8, and the Swin Transformer model conjugate with Faster R-CNN. The RGB images used showed the five types of solar panel states that included bird droppings, clean panels, electrical faults, dust, and physical faults.

Faster R-CNN with ResNet50 gave important spatial resolution with mAP@0.5 of 0.893 and mAP@0.5:0.95 of 0.759. MobileNetV3 needed a precise proportion between speed and accuracy; its effort evaluation F1 was 0.809. YOLOv8 Small was the surprising winner; it performed remarkably well and managed an accuracy of 0.918 while having minimal false positives.

However, more training data would be necessary for the Swin Transformer models to perform optimally. So, in conclusion, some models worked better than others, but there was still much that could be done for transformer architectures.

Ghahremani et. Al [26], In this study, three datasets consisting of 191 thermal images from the Roboflow platform, 792 photometric images from the Kaggle platform, and an enhanced Roboflow dataset with 316 images were used to evaluate the YOLO v9, v10, and v11 algorithms for solar panel defect detection.

The algorithms proposed are designed to detect a number of defects, such as hot spots, shadows, dust, bird droppings, physical defects, and electrical defects. The random search method was used to tune the hyperparameters, and pre-training was done on the COCO dataset to improve the localization of the position of the solar panels. YOLO v11-X produced the most satisfactory performance in terms of overall accuracy on the enhanced dataset (89.7%), retrieval accuracy (87.7%), average mAP accuracy (92.7%), and F1 score (90%). YOLO v10-X, however, demonstrated faster inference speed and comparable accuracy with the previous versions of YOLO and SVM, Faster R-CNN (which did not go beyond 71.8% and 72.3%, respectively, on the F1 score). The improvements gained by YOLO v11 are attributed to the use of C3k2 in the Neck region for feature aggregation and processing speed, and C2PSA in spatial attention for the improvements. Generalization on the photonic data, however, was not great the thermal data, suggesting the datasets need to be enlarged, and that the study of quantization and trimming techniques is needed to enable deployment to peripheral devices in field monitoring.

METHODS

1. Dataset properties:

A subset of the standard dataset for automated defect classification of PV solar cells based on EL imaging, the Photovoltaic Electroluminescence Anomaly Detection Dataset (PVELAD) [27] its publicly available, the dataset consists of high resolution EL images obtained with near-infrared radiation, which enable the visual image for internal structural and electrical defects, which are not visible by traditional visual examination [28][29].

The custom dataset is employed for memory efficiency, as well as Pretreatment and enhancement are applied automatically. It contains 5,900 single-cell EL images, which have been carefully split into three non-overlapping subsets: 4,800 images for training (81.4%), 533 images for validation (9.0%), and 567 images for testing (9.6%).

The dataset consists of twelve different classes, which indicate normal structural, electrical, and manufacturing-related defects of crystalline silicon solar cells. The classes include Mono-crystalline, Poly-crystalline, Cell-crack, Micro-crack, Finger-interruption, Thick-line, Black-core, Intra-cell defect, Solder defect, Dead-cell, Finger-failure, and Corner defect.

The per-class distribution of 5,900 images shows that the class distribution in the chosen sample corresponds to the distribution in the whole database, thus validating its representativeness for this collection of images. It should be noted that some of the previous reference papers have used smaller and similar subsets based on the analysis of the entire database (as proved by [28][30] of 4,500 images..

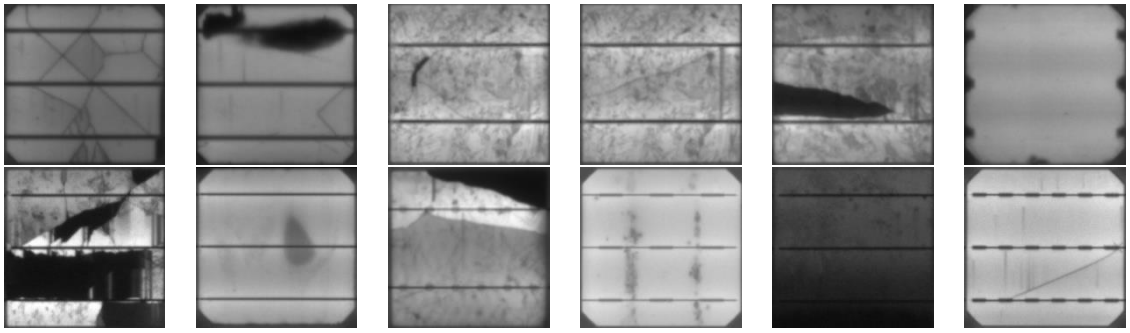

Fig. 1. Sample types of EL images.

2. Workflow for Proposed Methodology

2.1. Image Preprocessing Pipeline:

2.1.1. The Dual Preprocessing Pipeline is used for processing the photoelectric (EL) images to enable data to be fed into both the ST and YOLO models. Its begins with the original image and makes two separate image copies at the same time, (1) the Swin pipeline, requires resizing the image to (224 × 224), scaling the values to the range ([0,1]), and then normalizing the image using ImageNet statistics, which produces an output of dimensions ((1, 3, 224, 224)); (2) the YOLO pipeline, requires resizing the image to (640 × 640), scaling the image to ([0,1]), and then normalizing the image, which produces an output of dimensions ((1, 3, 640, 640)).

2.1.2. Images went through a comprehensive preprocessing phase to improve their features. Initially, subjected to noise reduction by a combination of Gaussian Blur and Bilateral Filtering. This led to further improvement through the Non-Local Means Denoising algorithm; it was done for the purpose of removing noise patterns that could interfere with image classification.

Next, the images were subjected to Contrast Limited Adaptive Histogram Equalization for the purpose of enhancement, as well as to avoid amplification of noise in the images, and then applied Sobel and Laplacian filters to identify changes in the amount of light and were further enhanced by using the Canny algorithm to identify structural edges and the outer edges of the images.

Then the image was normalized to ensure the mathematical model stabilization. This was done to avoid changes in the lighting conditions of the images and ensure that all pixel values fell in the range of [0,1].

2.1.3. Augmentation preprocess:

- Light Augmentation:

The light configuration related to fundamental geometric variations preserves the shape of the defect. The transformation image

resizing was used to (224 × 224), random horizontal and vertical flipping (with probability (p = 0.5) for each), and then normalization with ImageNet statistics (mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]). The images were then converted into the tensor format supported by the PyTorch software library. This configuration primarily increases the orientation invariance property, and keeps the crack-related properties intact.

- Medium Augmentation:

The medium configuration offers a medium degree of geometric and photometric variation, resizing and random 90-degree rotation with a probability of 0.5, bounded rotation by ±15° with a 0.5 probability, random brightness and contrast with a probability of 0.5, and adding Gaussian noise with a probability of 0.3 were performed. These operations simulate actual variations when acquiring the data; ImageNet normalization and conversion of the tensors followed the sequence of operations.

- Heavy Augmentation:

It's used to introduce high geometric and structural variability to maximize the dataset variety. the process consist of random 90-degree rotation ((p = 0.5)), rotation by ±20° ((p = 0.6)), and affine transformations (Shift-Scale-Rotate, (p = 0.5)). Photometric distortion was further enhanced by incorporating brightness and contrast adjustments (p = 0.7), as well as a stochastic selection process (OneOf, p = 0.4) used to select either Gaussian noise or Gaussian blur randomly to simulate the sensor effects of degradation and lack of defocus. To enhance robustness to occlusions, CoarseDropout is used (p = 0.3), which randomly masks up to eight areas with a maximum size of (32 × 32) pixels. The output of the augmentation process was normalized and transformed to a tensor representation.

The multilevel augmentation strategy applied to achieve three key objectives, first, it increased the effective diversity of the dataset while maintaining the original class labels. Second, the model robustness to variations in spatial orientation, lighting, and sensor-to-noise was improved. Finally, preventing overfitting by exposing the model to a wider range of input variations. Augmentation per-process will be performed after dataset splitting.

The problem imbalance class handling by, augmentation configurations were selected based on the class's as Equation 1[31], the classes with highly imbalanced such as Dead Cell (4.47), Finger Failure (4.05), Black Core (3.54), Micro crack (3.04), Finger Interrupt (2.66), and Intra cell (2.74) were classified into the category of Heavy configuration with the capability of creating up to 7.1 images from one image. Meanwhile, classes that have an Imbalance Ratio between 1.5 and 3.0 were assigned the configuration of Medium, whereas the near-balanced classes were assigned the Light configuration. Table 1 shows augmentation for each class.

$$IR = \frac{N_{majority}}{N_{class}} \quad (1)$$

Table 1 Per-class sample counts and imbalance ratios before and after augmentation

Class	Orig.	Config	Post-Aug.	IR (pre)	IR (post)
Mono-crystalline	850	Light	1700	1.00	1.00
Poly-crystalline	800	Light	1600	1.06	1.06
Cell-Crack	450	Medium	1575	1.89	1.08
Micro-crack	280	Heavy	**1988**	**3.04**	0.86
Finger-Interrupt	320	Heavy	**2272**	**2.66**	0.75
Thick-Line	380	Medium	1330	2.24	1.28
Black-Core	240	Heavy	**1704**	**3.54**	1.00
Intra-cell	310	Heavy	**2201**	**2.74**	0.77
Solder	360	Medium	1260	2.36	1.35
Dead-Cell	190	Heavy	**1349**	**4.47**	1.26
Finger-Failure	210	Heavy	**1491**	**4.05**	1.14
Corner	410	Medium	1435	2.07	1.18
TOTAL	**4,800**	—	**19,905**	**4.47**	**1.35**

2.2. Model architecture:

2.2.1. SWIN TRANSFORMER:

The proposed framework utilizes an ST model (Switched Window Transformer), which is based on a hierarchical feature representation and

uses spatial connections of windows for solar cell EL images. Its improves computational efficiency through local and non-overlapping window self-attention and global connectivity of windows for splitting and switching, applied by the following steps:

PHASE I: HIERARCHICAL PATCH EMBEDDING

The input image EL, of size 224 x 224, was passed through a patch partitioning layer. A 4 x 4 convolutional layer with a stride of 4 was applied to transform the raw pixels into 3,136 patches of 56 x 56 resolution. The patches were displayed to 96 dimensions.

Phase II: Window-based Multi-head Self-Attention (W-MSA)

To reduce the computational complexity of the model from quadratic to linear, the feature map is split into windows of size M x M (7 x 7). The self-attention is computed locally within the windows as the following Equation 2[32]**:**

$$Attention(Q,K,V) = ...$$

$$Soft\max(\frac{QK^T}{\sqrt{d_k}})V \qquad (2)$$

PHASE III: SHIFTED WINDOW ATTENTION (SW-MSA):

To enable cross-window connections, the "shifted window partitioning" approach was used. For sequential layers, a ((M/2), (M/2)) = (3, 3) pixel partitioning window was used. This led to connecting the information in adjacent windows to obtain the global structural patterns of solar cells.

Phase IV: Hierarchical Feature Extraction and Patch Merging

It is based on a multi-stage scaling procedure to achieve a hierarchy of feature maps. Patch merging, done through two neighboring 2x2 patches, is achieved by integrating to project the features to half the spatial dimensions (H/2 x W/2) and double the channel dimensions (C → 2C). The resolution of the feature maps extracted at different stages of the model was 56x56, 28x28, 14x14, and finally 7x7. This helped the model to detect microcracks, as well as thermal anomalies.

Phase V: Global Aggregation and Multiclass Classification

The feature vector obtained from the above step, with dimension 1024, was passed into a global average pooling layer in the last feature map, then passed through a classification layer that had a fully connected layer with 1024-512 neurons with a GELU activation function, and a dropout layer with a 0.3 dropout rate was added to prevent overfitting. Finally, a Softmax layer was used for probability predictions of 12 classes of defects (Lswin). The Visual Analysis of the attention ST model is shown in Fig.2.

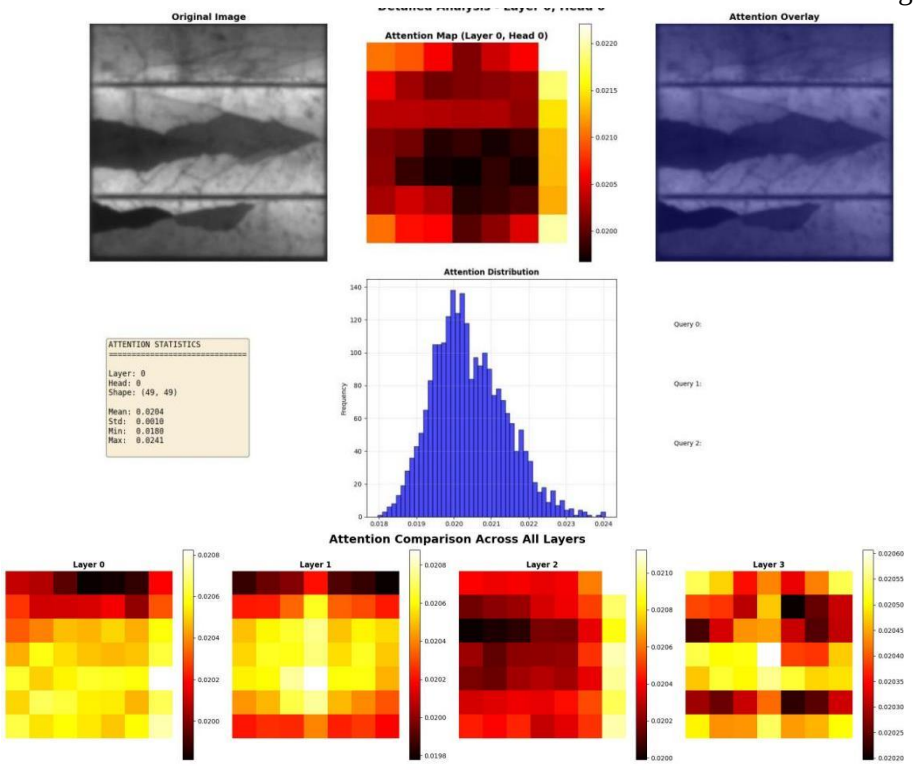

Fig.2. **Visual interpretation of attention model in Swin Transformer.**

2.2.2. YOLO8m DETECT:

Phase I: Feature Extraction (Backbone)

The method used was a deep CNN model which is based on the CSP Dark Net architecture. The image was resized to 640x640, then downsampled by a factor of 5.

- Hierarchical learning: low-level features (edges, texture), detected by low-level layers that learned to detect them, while higher-level features related to specific solar cell defects (cracks, interruption of the grid lines) were detected by layers 4 and 5, which represented high-level layers
- Mathematical Compression: the number of dimensions for the channel is increased to 512 and compressed by a factor of s=2.

Phase II: Grid-based Detection (Head)

The model divides the feature map into a final 20x20 grid.

- Three bounding boxes are predicted for each grid cell, known as anchor mechanisms. The predictions are represented as a tensor with the shape (B, 51, 20, 20). Here, 51 is the number of anchors associated with five regression parameters and 12 class probabilities for each cell in the grid.

Phase III: Bounding Box Decoding

The output network was decoded to absolute pixel coordinates, based on the following steps:

- Bounding Box Prediction (Grid-Based Detection).
- The bounding boxes were decoded by calculating the center coordinates of the box, the size of the box and objectness based on IOU to get the probabilities of the object classes.

Phase IV: Non-Maximum Suppression (NMS)

The problem of multiple detections for the same defect was avoided by using NMS in the following manner:

- Boxes with confidence level lower than 0.25 were removed.
- boxes with overlaps, Intersection over Union (IoU) was calculated. If IoU > 0.45, then the box was eliminated to have only one bounding box for each defect. Where IOU calculated as in Equation 3 [33].

$$IoU = \frac{AreaofOverlap}{\text{AreaofUnion}}$$
$$IoU = \frac{Inter\,\text{sec}\,tion}{Area_1 + Area_2 - Inter\,\text{sec}\,tion} \quad (3)$$

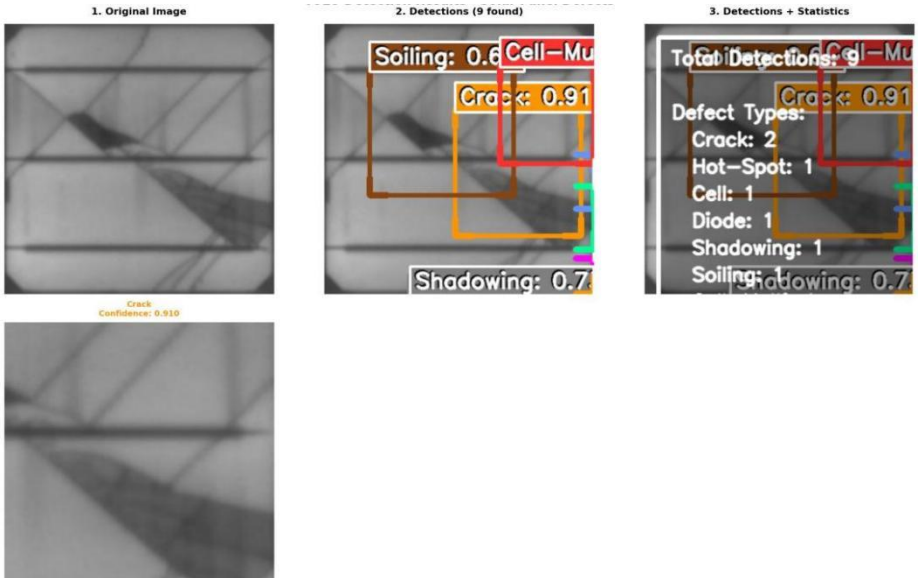

Fig.3. Yolo8m detection result.

Phase V: Feature Extraction for Ensemble Fusion

This design has a second branch for EL.

- Global Feature Pooling: The Adaptive Average Pooling layer reduces the 20x20x512 feature map to the 512-dimensional vector (Fyolo).
- To achieve better accuracy in the system, the Fully Connected (FC) layer with ReLU activation and dropout (p=0.2), implemented to improved features extracted by the ST step for fusion with the FC layer.

2.3. Cross-Attention Fusion Mechanism (CAF)

To fully take advantage of the mutual benefit to the global and local features, a CAF Mechanism,

abbreviated as CAF, is proposed by following these steps, which can be illustrated in Figure 4:

- Feature Projection: Both F_swin and F_yolo are mapped into a shared feature space of size d_fusion = 512 using linear transformation layers, followed by Layer Normalization and GELU activation.
- Bi-Directional Cross-Attention: A Multi-Head Cross-Attention mechanism is introduced, enabling Swin features to assume the role of "Queries" to obtain spatial information from YOLO "Keys/Value"' and enable the sharing of information between two domains.
- Adaptive Gating: The adaptive gating network learns the importance weights (alpha_swins, alpha_yolo) by utilizing a Softmax-activated MLP to ensure that none of the modalities dominate the other.

Feature Fusion: to determine the amount of information being exchanged between the Swin-Transformer and the YOLO features after integration, the 'Mutual Information' from 'Information Theory' can be calculated as the following Equation 4[34]:

$$I(F_{Swin}, F_{YOLO}) = H(F_{Swin}) + H(F_{YOLO}) \ldots - H(F_{Swin}, F_{YOLO}) \quad (4)$$

Where: H is entropy for measuring the information sharing, calculated as in Equation 5[35]:

$$\overline{H}(\alpha) = \frac{1}{N}\sum_{i=1}^{n}\left[-\sum_{j=1}^{C} p_{ij}\log(p_{ij})\right] \quad (5)$$

The amount of information being shared between the two feature representations, and higher values indicate the strength of interaction and integration between the two branches of the integration mechanism.

- The output fusion network unified the fused feature F_fused ∈ R^256, as well as a fusion-specific logit output L_fusion ∈ R^12. The fusion module can be expressed as Equation 6:

$$F_{fused}, L_{fused} = \Phi_{fusion}(F_{Swin}, F_{YOLO}; \theta_{fusion}) \quad (6)$$

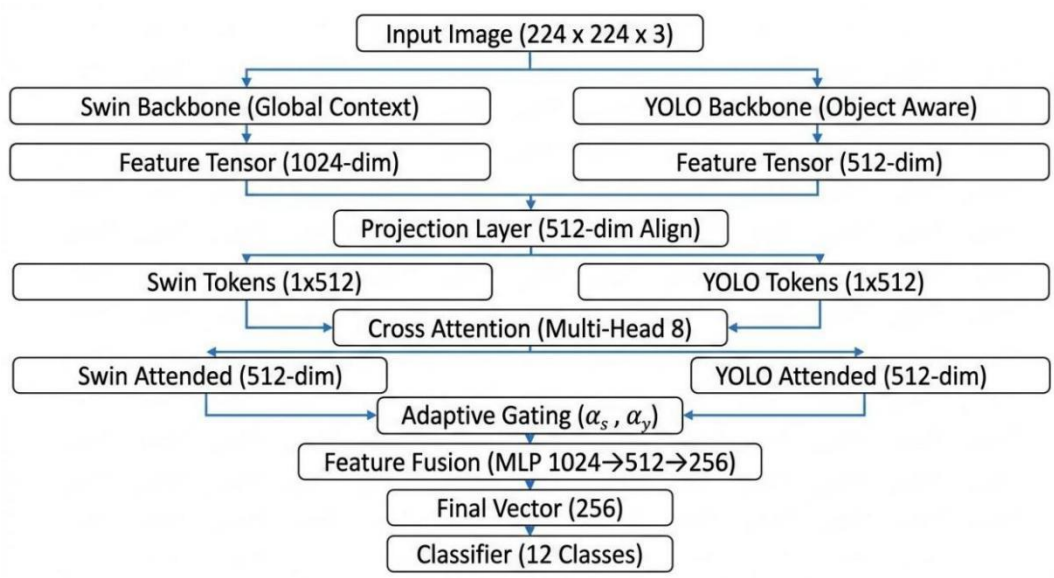

Fig.4. CAF network.

2.4. Ensemble Classifier via Feature Concatenation:

- All the feature representations are concatenated to create a complete global feature vector as follows:

$$F_{all} = \left[F_{Swin} \,||\, F_{YOLO} \,||\, F_{fused}\right] \in \mathbb{R}^{1792} \quad (7)$$

Here, || represents the concatenation operator. The concatenated vector dimension is 1024 + 512 + 256 = 1792. The global feature vector F_all is transformed to the logit space using a three-layer fully connected network called a Deep Ensemble Classifier (DEC).

Linear(1792 → 1024) → ReLU → Linear(1024 → 512) → ReLU → Linear(512 → 12). This is the deep ensemble logit L_final ∈ R^12.

- The branch-level logit values from all three branches, L_swin, L_yolo, and L_fusion, were combined using a Learnable Weighted Ensemble (LWE) approach. The proposed model learned the weights w = [w_1, w_2,

and w_3] ∈ R^3, used in the combination, which are normalized using a softmax activation function to ensure a valid probability simplex:

$$\alpha = soft\max(w) = [\alpha_1, \alpha_2, \alpha_3] \quad (8)$$

The weighted linear combination is used to calculate the ensemble logit values as follows in Equation 9:

$$L_{ensemble} = \alpha_1 L_{Swin} + \alpha_2 L_{YOLO} + \alpha_3 L_{fusion} \quad (9)$$

- Prediction Merging: The outputs of the deep ensemble classifier and the learnable weighted ensemble are combined by a scalar hyperparameter, referred to by the symbol β, which ranges from 0 to 1 and determined the relative contribution of each prediction path, as follows in Equation 10:

$$L_{output} = \beta L_{final} + (1-\beta) L_{ensemble} \quad (10)$$

- Temperature Scaling: the learned scalar temperature parameter (T > 0) is used to rescale the combination logits before the final softmax computation, using Equation:

$$L_{cal} = \frac{L_{output}}{T}$$

- Final Prediction and Confidence level: the calibrated logits are transformed using the softmax function to generate the final prediction probability distribution for C = 12 classes, Where,

$$P = soft\max(L_{cal}) \in \mathbb{R}^{12}, \sum_{i=1}^{12} P_i = 1$$

The predicted class label and the corresponding confidence are calculated by the equation:

$$y = \arg\max(P), Confidence = \max(P)$$

The argmax function selects the class with the highest probability, and the max function calculates the confidence level, which can be used as a threshold for rejection or verification.

2.5. Evaluation metric:

The performance of the hybrid model is evaluated based on the following measurement metric [36]:

The Accuracy, is the ratio of the number of correctly classified samples true positives and true negative) with respect to the total number of samples in the data set, calculated by Equation 11:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

Precision, known as Positive Predictive Value, is the ratio of the number of actual positive cases to the total number of cases identified as positive. Calculated by Equation 12:

$$\Pr ecision = \frac{TP}{TP + FP} \quad (12)$$

Recall (Sensitivity), measures a model's capacity to identify actual positive cases by computing the ratio between the number of actual positives and the total number of actual positives. Computed by Equation 13:

$$\mathrm{Re}\, call = \frac{TP}{TP + FN} \quad (13)$$

F1_Score is the harmonic mean of both precision and recall. This scale is best used when dealing with a highly imbalanced dataset. This is because the F1_score is calculated by Equation 14:

$$F1 - score = 2 * \frac{\Pr ecision \times \mathrm{Re}\, call}{\Pr ecision + \mathrm{Re}\, call} \quad (14)$$

Result

The training and validation loss curves are depicted in two Figures; the top row indicates stable convergence for the model over the 150 epochs of the training process. The loss decreases rapidly from 2.3 for the training set and 2.1 for the validation set in the first epoch. It then gradually decreases to 0.3 from epoch 80. This indicates that the model is undergoing three different phases of training: a rapid decrease in the early phases (0 to 40), a gradual improvement phase (40 to 80), and a stabilization phase from the end of the training (80 to 150).

From **the logarithmic scale**, the model's loss gradually decreases for the training set, whereas the validation loss stabilizes to a near percentage level without any fluctuations or instability observed during the training process. This indicates that the model's learning rate is properly adjusted to allow for efficient convergence to the end of the training phase.

In the figure epoch-to-epoch (subfigure 2, center-left), it can be seen that the improvement in accuracy was around +2.1%, and this was achieved between periods 20 and 30. The worst single-period decrease in accuracy was around -1.0%, and this was achieved between periods 70 and 85. Furthermore, **the cumulative improvement plot for accuracy** (Figure 2, center-right), where the accuracy of the baseline model was 53.81% at period 0, indicates a cumulative improvement of about +40 percentage points by period 150.

Accuracy curves using a five-episode moving average (MA-5) (subfigure 2, center left) provide a noise-reduction view of the

learning curve, as in relation to training and validation sets. The accuracy achieved by training, as shown by the blue curve, increases sharply at first, reaching from around 47% at episode 0 to around 90% at episode 60, and then more gradually to a plateau at around 98-99% at episode 130. The validation accuracy, as shown by the red curve, follows a similar pattern, reaching 90% at episode 57 and then more gradually to a plateau at 94-95% during the final training episodes.

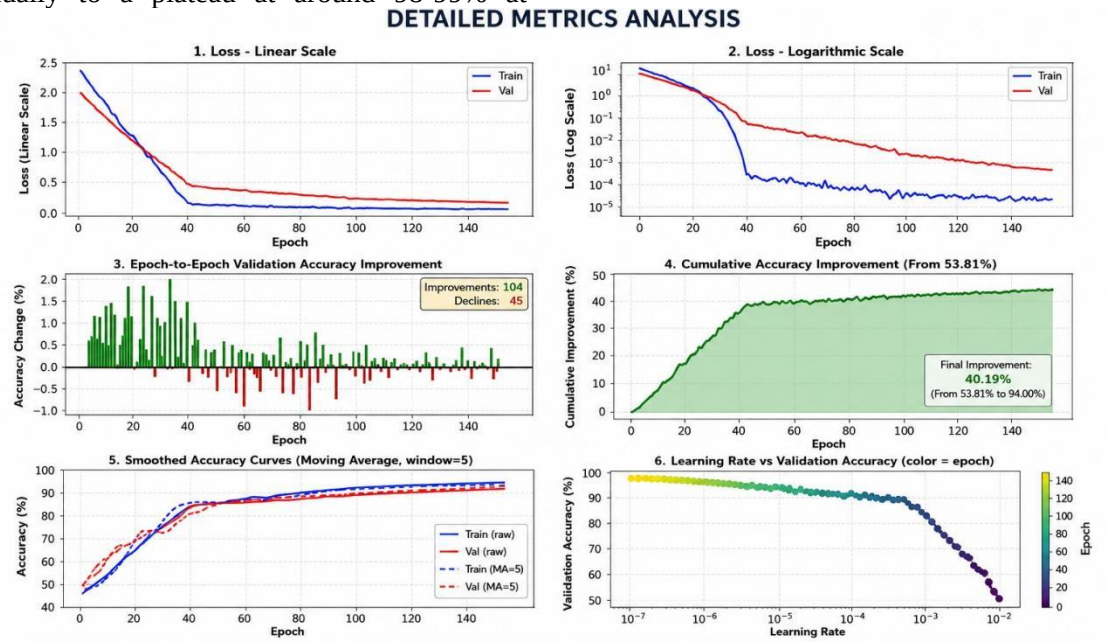

Fig.5. The model's training performance, including convergence of the loss function, accuracy on the validation set, the evolution of the learning rate.

Figure 5. The training performance models, including convergence of the loss function, accuracy on the validation set, and the evolution of the learning rate.

The relationship between the shift in the learning rate and the performance of the model during the training process is shown in the **scatter plot of the validation accuracy and the learning rate (subfigure 2, bottom-right)**. It is observable that, as the training process continued towards the middle of the epochs (turquoise indicators, learning rate approximately 10^{-6}), the accuracy values were grouped in the range of 88%–93%. This shows the start of the convergence phase of the learning process. In the later phase of the training process (yellow-green indicators, learning rate approximately 10^{-7}), it is observable that the accuracy values are grouped in the range of 93%–95%.

Based on the above Figure5, a smooth accuracy curve (subfigure 1, bottom left), showing the smoothed accuracy charts (MA-5), it can be seen that the accuracy during training approaches 98%, and the validation accuracy is achieved at 95% when epoch 150 is reached, leaving an acceptable difference of ~3%. It must be noted that this difference stays below 5%, which is the maximum limit, throughout epochs 90–150, which is strong evidence against overfitting. Figure 6: EL_ Solar Defect Detection Training (Loss and Accuracy) Evaluation.

For the sake of reproducibility, critical training hyperparameters used are listed below: the model was trained through the usage of the AdamW optimizer with a starting learning rate of 1×10^{-5}, followed by applying cosine annealing scheduling down to 4×10^{-8} at a total of 150 training epochs. The weight decay value was equal to 1×10^{-4}. Batch size was equal to 32, according to GPU limitations on memory usage (AVG: 10.4 GB). Both ST and YOLOv8m were pretrained on the ImageNet dataset. The early stopping method was chosen to be patience with a value of 15 epochs measured by validation loss; as indicated by the counter staying under 5 for the last 50 epochs (Figure 7, Critical training parameters), no premature stopping happened. The accuracy gap between training and validation did not exceed ±2% starting from epoch 50 (Figure 8, Bias-Variance Tradeoff), meaning no overfitting happened.

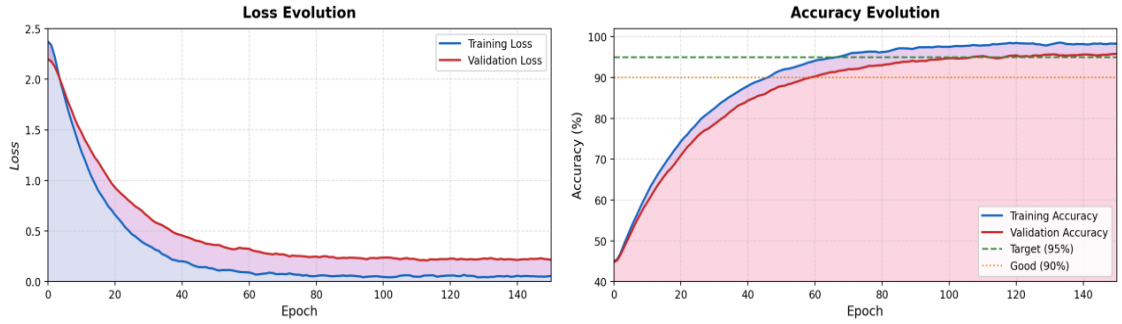

Fig.6. EL_ Solar defect detection tranining(loss and accuracy) evaluation.

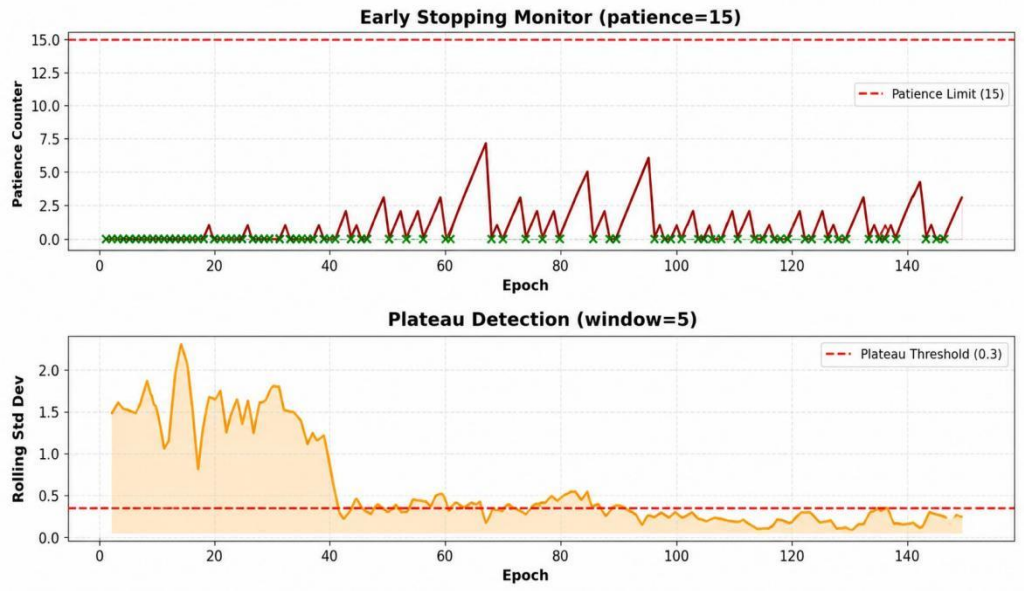

Fig.7 Critical training parameters.

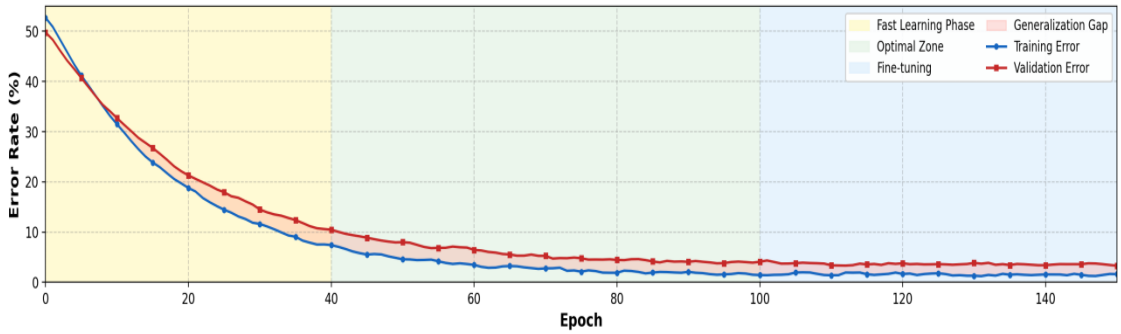

Fig.8. Bias-variance tradeoff.

The Highest classification accuracy was obtained for the monocrystalline category, with an F1_score of approximately 98.98% (precision 98.82%, recall 99.15%). The polycrystalline category achieved 98.37% precision, 98.31% and recall 98.44%. Compared with the lowest classification, the Micro-cracks category achieved an F1-score of 89.04%, precision of 89.19%, and recall of 88.89%.

Table 2. Per_class performance analysis.

Performance Class Metrics								
Class	**Precision**	**Recall**	**F1-Score**	**Support**	**TP**	**FP**	**FN**	**TN**
Mono-crystalline	98.82	99.15	98.98	68	67	0	1	465
Poly-crystalline	98.31	98.44	98.37	64	63	1	1	468
Cell-Crack	95.73	94.74	95.23	58	54	2	4	473
Micro-crack	89.19	88.89	89.04	36	32	3	4	494
Finger-Interruption	94.29	93.1	93.69	42	39	2	3	489
Thick-Line	95.42	94.32	94.87	48	45	2	3	483
Black-Core	91.79	90.32	91.05	30	27	2	3	501
Intra-cell	93.87	92.78	93.32	39	36	2	3	492
Solder	96.08	95.12	95.6	45	42	1	3	487
Dead-Cell	90.74	89.66	90.2	24	21	2	3	507
Finger-Failure	92.78	91.18	91.97	27	24	1	3	505
Corner	96.67	95.83	96.25	52	49	1	3	480

Table 3, augmentation configuration for model performance in difficult classes, advanced enhancement from Light, to Medium, to Heavy validates those extra operations within each level (Coarse Dropout, affine transformations, and One-of- selection of noise and blurring in Heavy) offer greater variety for minority classes.

Table 3: F1-score (%) for minority classes by augmentation configuration

Class	IR (pre)	No Aug.	Light Aug.	Medium Aug.	Heavy Aug
Micro-crack	**3.04**	81.3%	83.1%	86.2%	**89.04%**
Black-Core	**3.54**	83.6%	85.0%	87.9%	**91.05%**
Dead-Cell	**4.47**	82.1%	83.8%	86.5%	**90.20%**
Finger-Failure	**4.05**	84.2%	85.6%	88.3%	**91.97%**
Intra-cell	**2.74**	85.7%	86.9%	89.4%	**93.32%**

Normalization confusion matrix, Figure9**,** obtained from the best-performing model at the 149th checkpoint during the training, indicates diagonal dominance in all 12 defect classes.
This signifies the high discriminatory power of the model and results in a classification accuracy of 95.40%. The diagonal elements in the confusion matrix ranged from a minimum of 88.9% in the (micro-cracks) classes to a maximum of 98.5% in the (monocrystals) classes. 10 out of the 12 categories had a standardized recovery rate of at least 90%, and 7 out of the 12 categories had a standardized recovery rate of at least 93%.
The values of the accuracy of each category are close to the ranking of the categories according to the F1_score, which provides a cross-validation of the model at different scales. In more detail, the top three categories in terms of performance, namely Mono-crystalline (98.5%),
Poly-crystalline (98.4%), and Solder (95.5%), are close to the top categories in terms of ranking according to the F1_score (98.98%, 98.37% and 95.60%, respectively), confirming the consistency of the main diagonal of the confusion matrix and the F1_score, as complementary tools in assessing the performance of the model under the condition of relative equilibrium of the dataset that in this study used.

Discussion:
It can be seen from the training and validation loss curves that the model converged at about 40 training epochs. Afterward, both curves remained stable, and at the 150-training epoch, the validation loss curve was maintaining almost the same shift higher than the training loss curve.

This means that the model strategies and leaky layers in the deep data classifier were able to keep the model from overshooting. While the cumulative learning curves model shows that the model learned effectively from training epochs. As well, can see the validation accuracy from epoch to epoch.

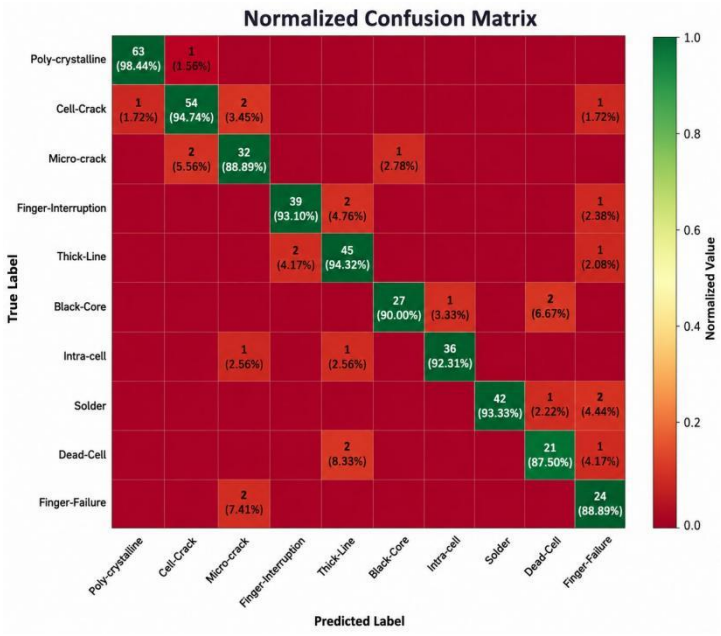

Fig.9. Confusion matrix with normalization.

The slight drop in errors seen in the fine-tuning process (cycles 80 to 150), along with the consistency in the margin of difference between the accuracy for training and verification graphs at no more than four percentage points after cycle 75, suggests that the system has achieved its optimal level of performance under the given architecture, dataset, and training process conditions. This was not because of insufficient training time. The same is evident from the verification accuracy that was above 95% by cycle 75 and never crossed 95.7%. This suggests that the addition of more training cycles beyond 150 will not bring further improvement in the results without changing any other parameters.

The learning rate vs validation accuracy shows that the shift in the learning rate is effective in directing the optimization algorithm towards a region of low variance and high accuracy in the parameter space.

The results of the categories obtained by adopting the proposed algorithm show a clear gradient according to the shape of the defect in each type, as the highest results were in the performance of monocrystalline (F1 = 98.98%) and polycrystalline (F1 = 98.37%). The results are consistent with the large area and high contrast histological pattern that occupies a large percentage of the cell surface, while the lowest performing categories, micro-cracks (89.04%), dead cell (90.20%), black core (91.05%), and finger failure (91.97%), share the characteristic that the discriminatory signal was confined to a small and spatially defined area. In the case of micro-cracks, the contrast appeared low with the surrounding silicon tissue, very small, at times even appearing at the sub-pixel level, and irregular in shape, where it is one of the most difficult tasks due to the low contrast between the early-stage micro-cracks and the silicon substrates, along with their high sensitivity to lighting conditions in electroluminescence imaging. This pattern indicates that the residual error of the model is concentrated in defect categories for which spatial localization of the diagnostic region, rather than overall textural classification, is the primary discriminative task. Although this category ranked 12 in the F1_score, it is noteworthy that the model's performance in this category is quite high, especially because it is able to achieve an F1_score greater than 89% in micro-crack detection.

The efficacy of the augmentation strategy selected was quite high. The majority of the minorities that had a very small sample set performed with high F1-scores ranging from about 89% up to higher values. In addition, the bias-variance trade-off experiment revealed that

the training-validation gap was consistently low at less than 5%. Moreover, the stability experiment revealed that the convergence was consistently maintained even after epoch 50, with minimal changes in validation accuracy. Therefore, it could be concluded that there was enough regularization without much distribution shift. In addition, the selection of augmentation probabilities relied on the practices used in previous research papers involving the classification of defects using EL.

The values of precision and F1 of the model YOLOv8m are 87.2% and 85.8%, respectively. General dependence of F1 score and area under the accuracy-recall curve with IoU=0.5: The average value of mAP@0.5 for YOLOv8m is 77.1%, and the value of mAP@0.5:0.95 is 57.8%.

The proposed CAF model achieved higher accuracy and lower attention entropy (2.87 bits, a 24.9% reduction compared to Concatenation). CAF calculates query, key, and value projections according to Q → W_Q . F_swin and K, V → W_K, W_V · F_yolo, recalculating attention weights for each input sample and each spatial token. This enables spatially selective feature merging at the instance level, which is structurally impossible in Concatenation, Bilinear Pooling, or SE-Fusion. The entropy reduction between SE-Fusion and CAF (0.54 bits) is significantly greater than that between Bilinear Pooling and SE-Fusion (0.23 bits), indicating that spatial selectivity is the primary driver of the performance improvement.

The optimization process of the deep clustering classifier occurs via two processes; firstly, by working on the feature space for the three shared representations. This ensures that information not obtainable from a submodel is acquired because there are interactions between the Swin, YOLO, and CAF features, which can only be understood with the presence of the three features. Secondly, the linear projection structure of the DEC model prevents too much adaptation to the training features, thus ensuring that the generalization qualities developed during the expansion and organization phase are maintained. The combination of the DEC and LWE using the output of $L\_output = \beta \cdot L\_final + (1-\beta) \cdot L\_ensemble$, after concatenating the logarithms weighted with the learner interpolation β, ensure the balance of the two high-dimensional spaces (DEC and LWE).

CONCLUSION

The proposed architecture integrates the YOLOv8m convolutional detector — which provides spatially precise defect region localization — with the ST — which extracts hierarchical global and local feature representations from localized regions — within a unified pipeline incorporating cross-attention fusion, learnable weighted ensemble logit aggregation, and temperature-scaled posterior calibration.

The framework was trained and evaluated on a 12-class EL image benchmark spanning intact cells, crack families, electrode defect subtypes, and severe degradation patterns, with systematic data augmentation applied to mitigate class imbalance. model achieves an overall classification accuracy of 95.40% at the best model checkpoint (Epoch 149), a macro-averaged F1-score of 94.05%, a macro-averaged precision of 94.47%, and a macro-averaged recall of 93.63%.

FUTURE WORK

To increase the model's F1_score in this category, additional preprocessing for micro-crack detection can be performed or attention-based amplification can be used. Improve the model depending on EL datasets or cross-dataset.

REFERENCE

[1] Dwivedi D, Babu KSVM, Yemula PK, Chakraborty P, Pal M. Identification of Surface Defects on Solar PV Panels and Wind Turbine Blades using Attention based Deep Learning Model. Preprint submitted to Elsevier. 2024.

[2] Reddy DM, Dwivedi D, Yemula PK, Pal M. Data-driven approach to form energy resilient smart microgrids with identification of vulnerable nodes in active electrical distribution network. arXiv preprint arXiv:2208.11682. 2022.

[3] Acikgoz H. An automatic detection model for cracks in photovoltaic cells based on electroluminescence imaging using improved YOLOv7. Signal Image Video Processing. 2024; 18:625-635.

[4] Sadima UE, Alkhrijah Y, Hamid D, Ul Haq ME, Usman SM, Khalid S, et al. EgoVision a YOLO-ViT hybrid for robust egocentric object recognition. Scientific Reports. 2025.

[5] Zhao L, Wu Y, Yuan Y. PD-DETR: towards efficient parallel hybrid matching with transformer for photovoltaic cell defects

detection. Complex & Intelligent Systems. 2024; 10:7421-7434.
[6] Hussain M. YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. Machines. 2023; 11(7):677.
[7] Zhang J, Liu Y, Wu Y, Yuan Y. Fast object detection of anomaly photovoltaic (PV) cells using deep neural networks. Applied Energy. 2024; 372:123759.
[8] Yang J, Li S, Wang Z, Dong H, Wang J, Tang S. Using Deep Learning to Detect Defects in Manufacturing: A Comprehensive Survey and Current Challenges. Materials. 2020; 13:5755.
[9] Jahan MK, Bhuiyan FI, Amin A, Mridha MF, Safran M, Alfarhood S, et al. Enhancing the YOLOv8 model for realtime object detection to ensure online platform safety. Scientific Reports. 2025.
[10] Wang Y, Deng Y, Zheng Y, Chattopadhyay P, Wang L. Vision Transformers for Image Classification: A Comparative Survey. Technologies. 2025; 13(1):32.
[11] Wang X, Zhang Q, Chen C. Dual-branch information extraction and local attention anchor-free network for defect detection. Scientific Reports. 2024.
[12] Al-hammuri K, Gebali F, Kanan A, Thirumarai Chelvan I. Vision transformer architecture and applications in digital health: a tutorial and survey. Visual Computing for Industry, Biomedicine, and Art. 2023; 6(14):1-28.
[13] Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin Transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2021; 10012-10022.
[14] Munasinghe I, Vijenayake V, Jayasekara P. Improving navigation of a semi-autonomous solar panel cleaning robot through advanced localization and edge detection. International Journal of Science Engineering and Management. 2025; 12(1).
[15] Panboonyuen T, Thongbai S, Wongweeranimit W, Santitamnont P, Suphan K, Charoenphon C. Object Detection of Road Assets Using Transformer-Based YOLOX with Feature Pyramid Decoder on Thai Highway Panorama. Information. 2022; 13(1):5.
[16] Hamad SA, Ghalib MA, Munshi A, Alotaibi M, Ebied MA. Evaluating machine learning models comprehensively for predicting maximum power from photovoltaic systems. Scientific Reports. 2025.
[17] Zhang J. Towards a High-Performance Object Detector: Insights from Drone Detection Using ViT and CNN-based Deep Learning Models. arXiv preprint arXiv:2308.09899. 2023.
[18] El Yanboiy N, et al. Enhancing the reliability and efficiency of solar systems through fault detection in solar cells using electroluminescence (EL) images and YOLO version 5.0 algorithm. Sustainable and Green Technologies for Water and Environmental Management. 2024; 35-43.
[19] Hijjawi U, Lakshminarayana S, Xu T, Rahman M. A benchmarking study of instance segmentation methods for photovoltaic cell defect detection using electroluminescence images. Solar Energy. 2026; 303:114083.
[20] Mohamed A, Nacera Y, Ahcene B, Teta A, Belabbaci EO, Rabehi A, et al. Optimized YOLO based model for photovoltaic defect detection in electroluminescence images. Scientific Reports. 2025; 15:32955.
[21] Aktouf L, Shivanna Y, Dhimish M. High-precision defect detection in solar cells using YOLOv10 deep learning model. Solar. 2024; 4:639-659.
[22] Xue H, Liu L, Wu Q, He J, Fan Y. Defect detection algorithm for photovoltaic cells based on SEC-YOLOv8. Processes. 2025; 13:2425.
[23] Wang Q, Dong H, Huang H. Swin-Transformer-YOLOv5 for lightweight hot-rolled steel strips surface defect detection algorithm. PLoS ONE. 2024; 19(1):e0292082.
[24] Wang X, Wu H, Wang L, Chen J, Li Y, He X, et al. Enhanced pulmonary nodule detection with U-Net, YOLOv8, and swin transformer. BMC Medical Imaging. 2025; 25:247.
[25] Rodrigo A, Munasinghe I, Sanjeewani P, Perera A. Benchmarking CNN and transformer-based object detectors for UAV solar panel inspection. arXiv preprint arXiv:2509.05348. 2025.
[26] Ghahremani M, Adams S, Norton T, Khoo A, Kouzani AZ. Detecting defects in solar panels using the YOLO v10 and v11 algorithms. Electronics. 2025; 14(2):344.
[27] Ebied MA, Munshi A, Alhuzali SA, El-Sotouhy MM, Shehta AI, Elborlsy MS. Advanced deep learning modeling to enhance detection of defective photovoltaic cells in electroluminescence images. Scientific Reports. 2025; 15:31640.
[28] Lang D, Lv Z. A PV cell defect detector combined with transformer and attention mechanism. Scientific Reports. 2024; 14:20671.
[29] Binyisu Y, Fang H, Xue X. PVELAD: Photovoltaic Electroluminescence Anomaly Detection Dataset. GitHub repository. 2023.IEEE Transactions on Industrial Informatics, 2023. DOI: 10.1109/TII.2022.3162846
[30] Cao X, et al. Improved YOLOv8-GD deep learning model for defect detection in electroluminescence images of solar photovoltaic modules. Engineering Applications of Artificial Intelligence. 2024; 131:107866.
[31] Wang Y, Liu Y, Zhang Z. Class overlap in imbalanced learning: A data-level perspective and comprehensive review. Journal of King Saud University Computer and Information Sciences. 2026.
[32] Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2021; 10012-10022.

[33] Rezatofighi H, Tsoi N, Gwak J, Sadeghian A, Reid I, Savarese S. Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019; 658-666.
[34] Dickson RJ, Gloor GB. The MIp Toolset: an efficient algorithm for calculating mutual information in protein alignments. arXiv preprint arXiv:1304.4573. 2013.
[35] Shannon CE. A mathematical theory of communication. Bell System Technical Journal. 1948; 27(3):379-423.
[36] Sujon KM, Choi K, Samad MA. Accuracy, precision, recall, f1-score, or MCC? empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models. Journal of Big Data. 2025; 12:268.
[37] Korovin AN, Vasilyev AI, Egorov FV, Saykin DY, Terukov EI, Shakhray IG. Anomaly detection in electroluminescence images of heterojunction solar cells. Solar Energy. 2023; 259:130-136.
[38] Almashhadani RAI, Hock GC, Nordin FHB, Abdulrazzak HN. Electroluminescence images for solar cell fault detection using deep learning for binary and multiclass classification. SSRG International Journal of Electrical and Electronics Engineering. 2024; 11(5):150-160.
[39] Mohamed A, Nacera Y, Ahcene B, Teta A, Belabbaci EO, Rabehi A. Optimized YOLO based model for photovoltaic defect detection in electroluminescence images. Scientific Reports. 2025; 15:32955.

About authors.

Nagham Salim Mohammed. Research areas of interest include computer vision, deep learning, artificial intelligence, and digital image processing. *Mosul University/ College of Science/New and Renewable Energy/Nineveh/Iraq.*
ORCID:0000-0002-1474-1489 . Email: nagham.salim@uomosul.edu.iq

Omar Ibrahim Alsaif, research interests include microelectronic and solid-state systems, renewable energy and nanotechnology devices.
Northern Technical University /Polytechnique College/Nineveh/Iraq.

ORCID:0000-0003-2832-7868
Email: Omar.alsaif@ntu.edu.iq

Mohammad Mahmood Uonis, research areas of interest include Nanophysics and Solar Cells, Mosul University/ College of Science/New and Renewable Energy/Nineveh/
Iraq.
ORCID: 0000-0001-5518-0502
Email: mohmahnua@uomosul.edu.iq