IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 24 June 2024, accepted 16 July 2024, date of publication 19 July 2024, date of current version 29 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3431098

RESEARCH ARTICLE

Attention-Based SegNet: Toward Refined Semantic Segmentation of PV Modules Defects

FATMA MAZEN ALI MAZEN[1], YOMNA O. SHAKER[1,2], (Member, IEEE), AND RANIA AHMED ABUL SEOUD[1]

[1]Faculty of Engineering, Electrical Engineering Department, Fayoum University, Fayoum 63514, Egypt
[2]Engineering Department, University of Science and Technology of Fujairah (USTF), Fujairah, United Arab Emirates

Corresponding author: Fatma Mazen Ali Mazen (fma04@fayoum.edu.eg)

ABSTRACT Proper surveillance and maintenance of photovoltaic (PV) systems are crucial to ensure continuous power generation and prevent operational downtimes. However, manual analysis of electroluminescence (EL) images is subjective, time-intensive, and requires significant expertise. To address this issue, a comprehensive deep learning architecture has been developed for the semantic segmentation of 29 different features and defects within EL images of PV panels. The SegNet architecture encoder has been replaced with the VGG16 encoder, which incorporates pre-trained weights to leverage transfer learning during the feature extraction stage. A Convolutional Block Attention Module (CBAM) block has also been introduced to enhance the decoder's ability to generate fine-grained segmentations. Additionally, the suggested architecture has been evaluated through the application of three different loss functions: weighted categorical cross-entropy loss, categorical cross-entropy, and focal loss. The Attention-Based SegNet architecture proposed with a weighted categorical cross-entropy loss exhibits superior performance in terms of accuracy, F1 score, intersection over union (IoU), precision, recall, mean IoU (mIoU), specificity, Jaccard index, and Dice coefficient. It achieves a Dice coefficient of 0.9408 and an mIoU of 0.9101, outperforming the state-of-the-art SEiPV-Net trained on the same dataset by 8.77% and 4.97%, respectively.

INDEX TERMS PV, electroluminescence images, solar cells, semantic segmentation, SegNet, CBAM, SEiPV-Net.

I. INTRODUCTION

In recent years, there has been a widespread transition to renewable energy on a global scale [1]. Compared to conventional energy sources like oil, natural gas, and coal, the solar cell sector had expanded greatly by the end of 2015 [2]. Worldwide, the amount of electricity generated via solar photovoltaic (PV) systems has risen significantly. International PV installations were 623.2 GW (GW) by the end of 2019 [3]. PV systems are expected to become the primary global electricity source by 2050, according to the International Energy Agency [4]. Based on the materials used in their manufacture, solar cells found in PV modules are generally divided into two categories: monocrystalline and polycrystalline silicon. Polycrystalline cells are made up of several silicon sections fused together during the manufacturing process, whereas monocrystalline cells are made up of a single silicon crystal [5]. Companies usually face ongoing trade-offs to provide high-quality PV modules at reduced costs. Over time, the capacity of crystalline silicon cells used in PV modules to generate power decreases due to the susceptibility to cracks, inactive areas, gridlines, and other various defects. Various environmental factors such as snow, wind, thermo-mechanical pressures, as well as human activity during shipping, routine operations, fabrication, maintenance, and installation procedures also have the potential to induce cracks in PV modules [6]. In a competitive market, manufacturers can account for solar panels' quality and reliability by employing two different methods during the production process: electroluminescence

The associate editor coordinating the review of this manuscript and approving it for publication was Shuo Sun.

(EL) testing and flash testing. In the flash test, a short, high-intensity light pulse is used to assess a panel's output performance. The current-voltage (I-V) curve determines the panel's maximum performance. Conversely, PV module defects are detected using EL imaging. During EL testing, photons are produced by solar panels when they come into contact with electricity; the resulting images are examined using infrared or near-infrared cameras to detect and describe different shortcomings. Cracks and other defects appear as dark grey lines or areas in EL images.

Visual inspection of EL images is expensive and time-consuming. Additionally, it is a labor-intensive procedure that requires knowledge in addition to its restricted applicability to a small scale. The application of automated detection techniques is required in order to expand the scope of visual inspection. [7]

The automatic defect detection in EL images of PV modules has been demonstrated to be reliable with computer vision and deep learning approaches. [6]. Deep learning technology has made substantial advances in the last few years in the areas of object detection, image segmentation, and image classification. [8]

Image segmentation is a comprehensive image analysis procedure that involves partitioning a digital image into multiple segments and categorizing the data within each segment. The three primary types of image segmentation tasks encompass semantic segmentation, instance segmentation, and panoptic segmentation. semantic segmentation is a computer vision technique that tries to classify every pixel in an image, assigning a particular label to each pixel based on its content. With this fine-grained comprehension, computers can identify the boundaries and connections between various objects or regions, resulting in a more comprehensive and contextually rich analysis of visual input. [5] The main findings of this research are listed below:

1. The SegNet architecture's encoder was replaced with the VGG16 encoder, incorporating pretrained weights to exploit transfer learning within the feature extraction phase.

2. A Convolutional Block Attention Module (CBAM) block was incorporated to enhance the decoder's ability to produce detailed segmentations.

3. The traditional categorical cross-entropy loss was replaced by the weighted categorical cross-entropy loss. Weighted categorical cross-entropy is beneficial for addressing class imbalances, allowing the model to effectively handle varying class distributions and improve overall performance.

4. The suggested framework is evaluated by applying three distinct loss functions: weighted categorical cross-entropy loss, categorical cross-entropy, and focal loss.

5. The proposed framework performs better than the state-of-the-art SEiPV-Net trained on the same dataset, with a Dice coefficient improvement of 8.77% and an mIoU improvement of 4.97%, respectively.

The structure of the paper is as follows: Section II offers a review of the relevant literature and the contemporary advances in the detection of defects in PV solar cells, encompassing models based on deep learning. Section III delineates the architecture of the suggested model and the key techniques employed in its development. Section IV offers a detailed description of the dataset. It also discusses and analyzes the experimental results. Lastly, Section V encapsulates the study's essential conclusions and outlines potential avenues for future research.

II. RELATED WORK

Numerous scholarly studies have been carried out on the automated detection of defects in solar cells; however, a minority have integrated the use of semantic segmentation in EL imaging. Semantic segmentation involves the process of classification at pixel-level, enabling the identification and categorization of multiple objects within an image. In their work referenced as [6], Pratt et al. pioneered the development of the initial semantic segmentation model utilizing the UNET architecture for the detection and classification of 24 defects in PV modules. The suggested model was trained and tested by the authors using PV modules made up of multicrystalline and monocrystalline silicon cells. The images from the EL dataset and their corresponding ground truth masks were adjusted to a size of 512 $\times$ 512 pixels. The evaluation revealed superior performance of the model on monocrystalline silicon cells compared to multi-crystalline silicon cells. A benchmark dataset was presented by Pratt et al. [9] for the purpose of semantically segmenting 24 distinct features and defects of PV modules. EL images and related ground truth masks with pixel-level annotations for every defect and feature were used to create this dataset. The researchers then used equal, inverse, and custom class weights to train four different deep learning models. They then used a subset of three faults and two features to evaluate the model's performance using the median recall (mRcl) and median intersection over union (mIoU). Ultimately, the DeepLabv3+ model, implemented with custom class weights, demonstrated the most favorable performance in terms of evaluation metrics. The authors concluded that significant performance degradation could result from deviations among the ground truth masks and the EL images. As a result, they provided alternative versions of the dataset to address this challenge.

In a recent study, Eesaar et al. [10] devised a lightweight encoder-decoder architecture termed SEiPV-Net for the semantic segmentation of defects and features within PV modules. The SEiPV-Net model underwent training and evaluation utilizing the dataset introduced in [9]. To tackle the problem of micro-defects occupying a minimal number of image pixels, the researchers employed various class weight assignment strategies. Additionally, they employed three distinct loss functions during the model training process. Notably, the proposed SEiPV-Net performed better than cutting-edge techniques such as U-Net, PSP-Net, and DeepLabv3+ across multiple evaluation metrics. For real-time crack segmentation in complicated scenes, the authors in [11] presented a novel semantic transformer representation

network (STRNet). A focal-Tversky loss function, a multi-head attention-based decoder, coarse upsampling, a squeeze and excitation attention-based encoder, and a learnable swish activation function are among the key components of the STRNet's meticulously constructed architecture. With 91.7%, 92.7%, 92.2%, and 92.6%, respectively, the network notably demonstrates strong performance in precision, recall, F1 score, and mIoU. In order to verify the efficacy of STRNet, the authors carried out a performance comparison study between it and other cutting-edge networks, including Deeplab V3+, FPHBN, Unet++, Attention U-net, and CrackSegNet. A deep learning encoder–decoder framework was developed by the researchers in [12] with the express purpose of semantically segmenting droplets and shunt-type defects in thin-film copper indium gallium selenide (CIGS) solar cells. A collection of 6000 images of identical CIGS modules was used to evaluate their model. Data augmentation approaches were applied to improve the recognition of droplets, specifically to address the lack of annotated images in the droplet class. Various semantic segmentation criteria, including the Jaccard index, Precision, and Recall, were utilized to appraise the model's performance. Fioresi et al. [13] introduced the UCF EL Defect dataset. They proposed the use of a Deeplabv3 model with ResNet-50 backbone for semantic segmentation of five types of defects in monocrystalline and multicrystalline PV cells. The model's training utilized the UCF EL Defect dataset that comprises 17,064 EL images. The model exhibited an mIoU score of 57.3%, a pixel-level accuracy of 95.4%, weighted F1-score of 0.95, and unweighted F1-score of 0.69 and, underscoring its significant potential for industry applications.

In their study [14], Rahman and Chen introduced an enhanced Unet model termed multi-attention U-net (MAU-net) for binary PV defect detection in EL images. For model training, a dataset comprising 828 solar cell images was used to determine whether the cell exhibited defects or was defect-free. Data augmentation methodologies were utilized to enhance the dataset scale and diversity. Furthermore, the model integrated channel attention and spatial attention mechanisms to suppress nonessential features during training and highlight important ones. To address the issue of class imbalance, a combination of focal and dice loss functions was utilized for model training. The suggested model exhibited better performance relative to the baseline Unet model and alternative semantic segmentation methodologies. It achieved an F-measure of 0.799 and an mIoU of 0.699. Han et al. presented a deep learning-based method for defect segmentation in polycrystalline silicon wafers in [15]. The process initially involved using a Region Proposal Network (RPN) to produce defect patches, which were then entered into an improved Unet to perform defect segmentation. A dilated convolution was added to the original Unet structure to handle the variation in the defect sizes. Utilizing dilated convolution can improve the model's ability to rebuild information about small objects. The final findings were then obtained by processing the segmentation outcomes. The model was trained and assessed on a dataset that consisted of 106 1024×1024 images. The experimental findings showed that, in spite of being trained on a relatively small dataset, the suggested model performed better than existing semantic segmentation models. Wang et al. [16] presented a novel algorithm for defect detection in solar cells. This algorithm is an enhanced version of the Faster region-based convolutional neural network. The algorithm leverages a similarity non-maximum suppression mechanism, along with the cosine similarity of the candidate box aspect ratio, and multi-scale feature fusion. The authors conducted comprehensive training, validation, and testing of the proposed approach using the dataset documented in [17], with an augmented image count of 2687 achieved through slicing. The proposed algorithm demonstrated notable superiority over other object detection models, achieving an impressive mean average precision (mAP) of 91.19%.

To automatically detect PV defects in EL images, Mazen et al. [18] have improved YOLOv5. The main improvements are as follows: the network's neck and backbone are enhanced with the Global Attention Module (GAM); the head branch's Adaptive Feature Space Fusion (ASFF) is integrated to increase the effectiveness and accuracy of the model's detection; and the distance intersection over union-non-maximum suppression (DIoU-NMS) is used to create more accurate bounding boxes. The suggested system was trained and tested. Furthermore, after using Test Time Augmentation (TTA), the enhanced YOLOv5 model outperformed the YOLOv8 model, attaining a mAP@0.5 of 77.7%.

El-Rashidy [1] presented a lightweight automated method for PV panel defect detection. First, solar cell images are clustered, and each cluster's identification model is created. A classifier model is developed to categorize solar cell images into clusters, which is then employed to assign a cluster label to each cell image based on its similarity to the images within the cluster. The identification of defective solar cells is subsequently achieved through the utilization of the model developed based on the assigned clusters. A convolutional neural network (CNN)-based classification model was developed by Tang et al. [19] for the identification of four different types of PV module defects: break, defect-free, finger-interrupt, and micro-crack. First, the authors employed Generative Adversarial Networks (GANs) to generate high-resolution EL images. Subsequently, four classifiers, leveraging ResNet50, VGG16, MobileNet, and Inception V3 architectures, were trained utilizing the generated dataset. The experimental findings demonstrated the considerable effect of data augmentation techniques in enhancing the model's accuracy.

Akram et al. [7] presented a classification system that is both time- and power-efficient to recognize PV defects EL images. Owing to the dataset's small size, the authors employed various data augmentation techniques, such as rotation, contrast correction, flipping, and random cropping, to increase its size. To reduce overfitting, they also used

weight regularization and dropout strategies. With an accuracy of 93.02%, the final model proved to be feasible for use in industrial settings for defect detection applications. The ELDDS1400C5 dataset, which includes 1,400 EL images of PV panels for defect detection and segmentation in PV modules, was developed by Rodriguez et al. [20]. There are 1187 training images in the dataset. 1187 training images were used to train a YOLOv5l model. After that, a test set of 190 was used to evaluate the model. The proposed model attained an mAP@0.5 of 0.778. Lastly, the authors segmented the cells using image-gradient-based line localization. A novel dual encoder-decoder architecture called the Polyp Segmentation Network (PSNet) was presented by Lewis et al. [21] for the semantic segmentation of polyps. Multiple deep learning modules were combined to create the PSNet's dual encoder and decoder. The model's performance was assessed using five different polyp datasets, demonstrating higher performance in terms of mean Dice coefficient (mDice) and mIoU when compared to existing state-of-the-art techniques. Zhang and Yin [22]developed an automated system for the identification of PV panel defects in a recent research article. The suggested model is an improved version of YOLOv5, designed especially for identifying three distinct types of defects in PV modules. Notably, the model's increased accuracy in identifying small objects was largely due to the addition of a tiny defect detection head, the use of the ECA-Net attention mechanism, and the inclusion of deformable convolution within the CSP module. To increase the dataset's size, the researchers also used a range of data augmentation strategies. Following these improvements, the model outperformed the original YOLOv5 framework with an mAP of 89.64%.

Another study [23] introduced a deep learning framework, termed the internal damage segmentation network (IDSNet), to semantically segment interior defects in concrete structures. To augment the training dataset for the proposed model, attention-based GAN (AGAN) was employed. The application of AGAN resulted in a notable enhancement of 12% in the mIoU, demonstrating its efficacy in improving the model's performance for semantic segmentation tasks.

The proposed model overcomes the drawbacks found in previous research works using the same dataset: it performs better in semantic segmentation of small, narrow cracks and produces masks that are more accurate than the ground truth masks. Moreover, the model effectively detects defects that the annotators missed, demonstrating its effectiveness in defect identification and segmentation.

III. METHODOLOGY

The dataset, suggested model, and methods used in its creation are summarized in this section.

A. THE DATASET

This work utilizes a 29-class dataset that was made available in [9]. It is further separated into 16 defects and 13 intrinsic

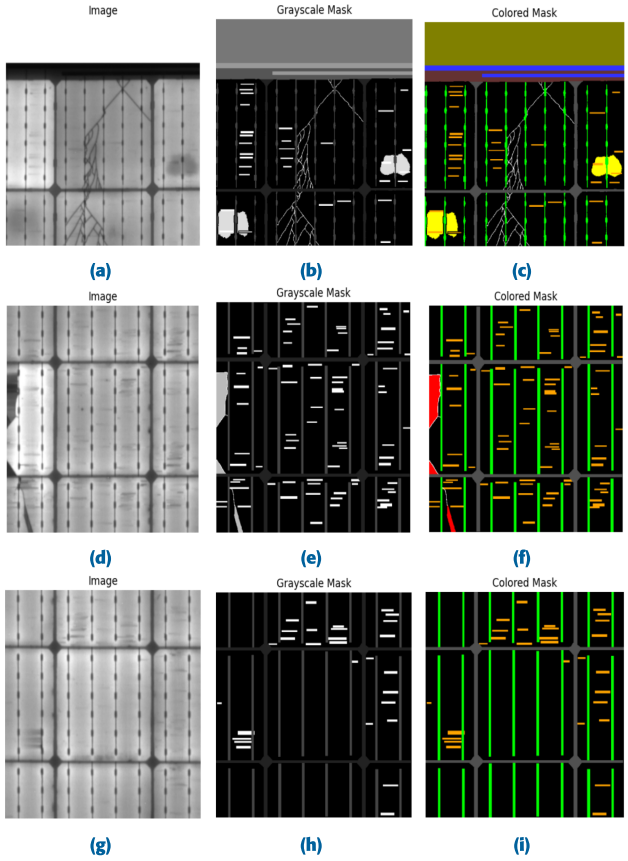

FIGURE 1. (a), (d), (g) EL images (b), (e), (h) Grayscale mask (c), (f), (i) Colored mask.

features of solar PV modules. A PV expert selected its images by visually scanning 80,000 images from the five data sources and selecting samples at random. A PV module's ribbon connector, busbar, or cell spacing are examples of particular elements that are referred to as features in this work. Defects refer to undesirable imperfections that can negatively impact the overall efficiency and lifespan of the PV system like gridlines, inactive areas, and cracks. A solar cell's EL image has a resolution of 512 × 512 pixels. For training, validation, and testing, a total of 2212, 70, and 72 EL images were employed, respectively. For monocrystalline and multicrystalline solar wafers, almost equal amounts of images were used.

the GNU Image Manipulation Program (GIMP) [24] was utilized to generate a ground truth mask for each image so that each class has a distinct color. As verified by Pratt et al. [9] and Eesaar et al. [10] through their benchmark investigation, the initial release of the dataset on November 4, 2021, presented a significant challenge marked by inaccurate labeling, resulting in suboptimal segmentation performance. To address this limitation, this research employs the latest version of the dataset, released on October 8, 2022. Notably, this updated iteration does not provide RGB masks for the training, validation, and test sets.

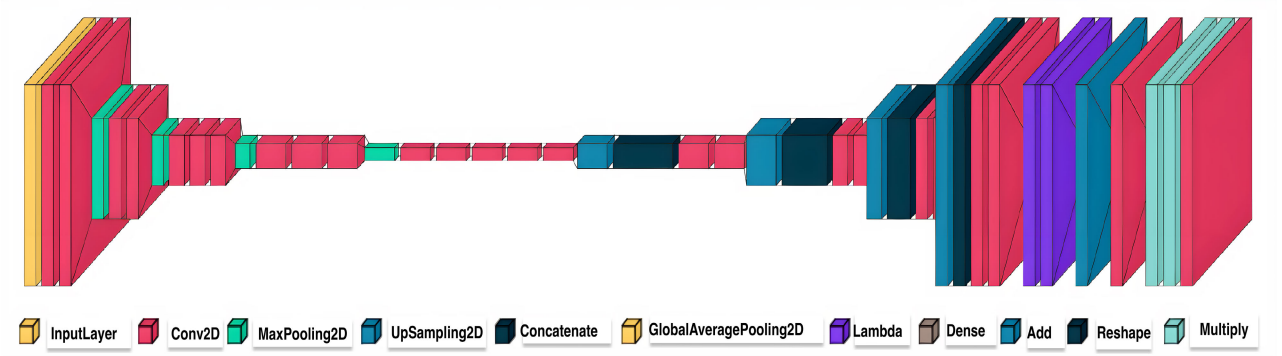

FIGURE 2. The attention-based SegNet model architecture.

Consequently, the color code information accompanying the dataset was used to transform grayscale masks into RGB format, as depicted in FIGURE 1.

B. DETAILS ABOUT THE TRAINING ENVIRONMENT

The study utilized the Kaggle platform with a computational environment consisting of a NVIDIA TESLA P100 GPU, CUDA version 11.1, Python version 3.8.8, keras version 2.13.1, and Tesorflow version 2.13.0. The training process used an Adam optimizer with a learning rate of 0.0001. The number of training epochs was set to 30, while the batch size was set to 8. The input images used in the training process had dimensions of 512×512 pixels. The Attention-Based SegNet model used the VGG16 encoder, which leveraged pre-trained weights obtained from the ImageNet dataset.

C. ATTENTION-BASED SEGNET MODEL

This work presents an improved version of the SegNet architecture presented by Badrinarayanan et al. [25] The SegNet has three main components: an encoder block, a corresponding decoder block, and a pixel-wise classification layer. The authors proposed a 13-convolutional-layer encoder similar to that of the VGG16 [26] architecture to produce a set of feature maps. Each encoder layer has a corresponding decoder layer, resulting in a 13-layer decoder network. On the other hand, the decoder takes feature maps as input, upsamples them, and produces sparse feature maps. The deepest encoder output retains higher-resolution feature maps by discarding fully connected layers. The high-dimensional feature representation emerging from the final decoder output is provided as input to a softmax classifier. The decoder in SegNet employs pooling indices from the max-pooling stage for non-linear upsampling. As a result, SegNet has fewer trainable parameters than other semantic segmentation architectures, making it time- and memory-efficient. The architecture of the proposed Attention-Based SegNet is shown in FIGURE 2 Initially, the VGG16 model is instantiated with pre-trained weights sourced from ImageNet, and the upper layers, specifically the fully connected layers, are omitted. Subsequently, the encoder output is derived by extracting the output of the 'block5_conv3' layer from VGG16. The decoder phase involves the upsampling and concatenation of features originating from the encoder across diverse resolutions. To illustrate, the concatenation with the 'block3_conv3' layer corresponds to a resolution of 128×128, concatenation with the 'block2_conv2' layer corresponds to 256×256, and the concatenation with the 'block1_conv2' layer corresponds to a resolution of 512×512.

Following the resolution adjustment, convolutional layers are applied to fine-tune the feature map. Lastly, the defined CBAM block is employed, followed by a 1×1 convolutional layer with softmax activation, culminating in the generation of the ultimate pixel-wise classification.

D. INTEGRATION OF A CONVOLUTIONAL BLOCK ATTENTION MODULE (CBAM)

The concept of CBAM gained initial prominence by Woo et al. [27]. They demonstrated that using CBAM at every convolutional block can produce enhanced feature maps. The integration of CBAM module into the SegNet network has the potential to greatly improve network performance by enhancing the significance of defect information and reducing the impact of irrelevant data [28]. It overcomes the difficulty of identifying obscured and tiny objects that might be missed otherwise and improves the network's feature extraction capabilities [29].

This CBAM module as shown in FIGURE 3 is divided into two distinct submodules: channel attention module and the spatial attention module. To ensure the concurrent utilization of both channel and spatial features, the attention weight is computed in the spatial and channel dimensions and subsequently applied to the original feature map. This process allows to dynamically modify the features.

1) CHANNEL ATTENTION MODULE

The channel attention mechanism depicted in FIGURE 4 identifies and enhances the most important feature map for learning. Firstly, global average pooling and global max pooling operations are applied to the feature map F, the output

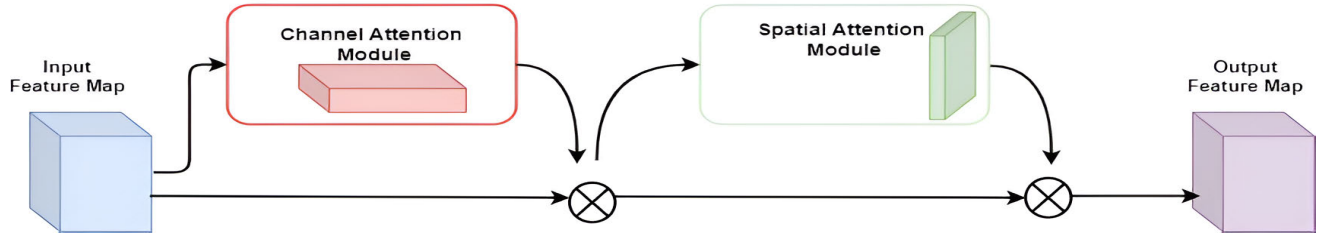

FIGURE 3. CBAM structure diagram.

of the encoder part. The result of this step is two 1 × 1 × C feature maps, F^C_{avg}, and F^C_{max}. The output of this step is then passed through two dense layers to reduce the channel dimension further. The number of neurons in the first and second layer is C/γ, C, respectively, where γ defines the attenuation rate that controls the channel reduction ratio. ReLU activation is then applied for non-linearity, and sigmoid activation is applied to produce attention weights. This process is expressed in Equation 1, where σ represents the Sigmoid activation function and w_0 and w_1 stand for the weights of the first and second dense layers, respectively.

$$\begin{aligned} M_C(F) &= \sigma\left(\text{MLP}\left(F^C_{\text{avg}}\right) + \text{MLP}\left(F^C_{\text{max}}\right)\right) \\ &= \sigma\left(w_1\left(w_0\left(F^C_{\text{avg}}\right)\right) + w_1\left(w_0\left(F^C_{\text{max}}\right)\right)\right) \end{aligned} \quad (1)$$

Finally, the channel attention weights and the input feature map are element-wise multiplied by the channel dimension as depicted in Equation 2 to produce F', the output feature map of the channel attention module that matches the input feature map dimensions.

$$F' = M_C\ (F) \otimes F. \quad (2)$$

2) SPATIAL ATTENTION MODULE

The output of the channel attention module, F' of the size $H \times W \times C$, is inputted into the spatial attention module as illustrated by FIGURE 5. Then, it is subjected to global max pooling and global average pooling of the channel dimensions, respectively. Max and average values are computed along the channel dimension separately to get F^S_{avg} and F^S_{max} of size $H \times W \times 1$. A 7 × 7 convolutional layer with Sigmoid activation function is then used to produce attention weights $M_S(F)$ based on max and average values as illustrated in Equation 3 where σ denotes the Sigmoid activation function, and $f^{7\times7}$ is the 7 × 7 convolution operation.

$$\begin{aligned} M_S(F) &= \sigma\left(f^{7\times7}\left(\left[\text{Avgpool}(F); \text{Maxpool}(F)\right]\right)\right) \\ &= \sigma\left(f^{7\times7}\left(\left[F^S_{\text{avg}}, F^S_{\text{max}}\right]\right)\right) \end{aligned} \quad (3)$$

Finally, the output feature map of the Spatial Attention module F'' is generated through element-wise multiplication of the input feature map F' and $M_S(F)$. The formula of F'' is given by Equation 4:

$$F'' = M_S\ (F) \otimes F'. \quad (4)$$

What is important to learn from the feature map is conveyed through the spatial attention module. The mask created by spatial attention will highlight the features that characterize the object of interest. Through the application of Spatial Attention to refine the feature maps, the input for the subsequent convolutional layers is enhanced, improving the performance of the model..

IV. RESULTS AND DISCUSSION

This section employs a variety of evaluation measures to test and compare the suggested model's performance with that of the most advanced models.

A. EVALUATION METRICS

The quality of the segmentation results is assessed using a variety of measures in order to determine how effective the recommended architecture is. As mentioned in [10], and [14], various metrics can be applied to compare the predicted outcomes with the ground truth masks: F1 Score, accuracy, recall, precision, Dice coefficient and mean Intersection over Union(mIoU). Equations 5, 6, 7, 8, 9, and 10 are used, respectively, to calculate IoU, accuracy, recall, precision, F-measure, and Dice coeficient.

$$IoU = \frac{|J \cap K|}{|J \cup K|} = \frac{TP}{TP + FP + FN} \quad (5)$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (6)$$

$$Recall = \frac{TP}{TP + FP} \quad (7)$$

$$Precision = \frac{TP}{TP + TN} \quad (8)$$

$$F1 - Score = 2 \times \frac{Precision * Recall}{Precision + Recall} \quad (9)$$

$$DiceCoefficient = 2 \times \frac{|J \cap K|}{|J| + |K|} = 2 \times \frac{TP}{2TP + FP + FN} \quad (10)$$

in which FN stands for false negative, FP for false positive, TP for true positive, and TN for true negative. In Equations 5 and 10, J represents the ground truth mask

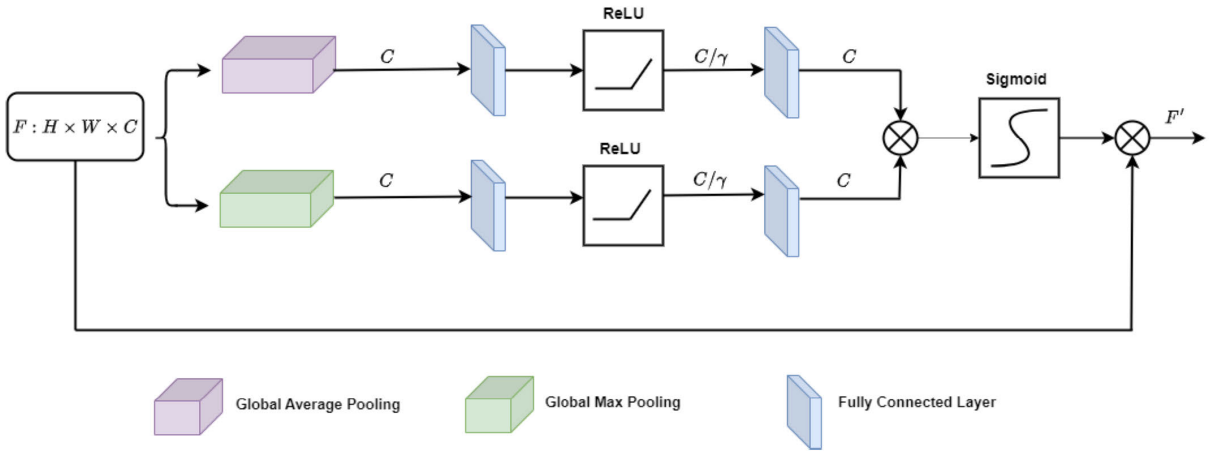

FIGURE 4. Channel attention module.

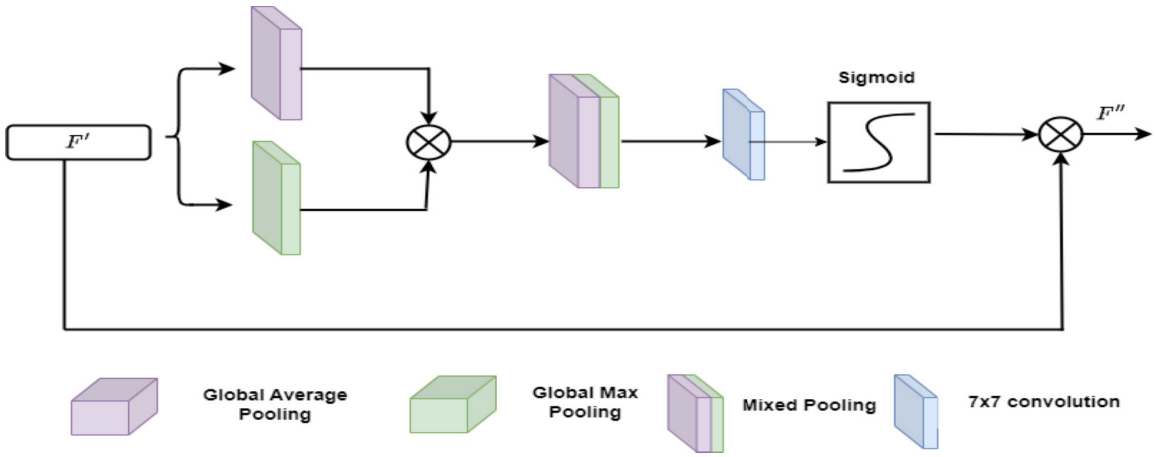

FIGURE 5. Spatial attention module.

for an object, K represents the predicted mask for the same object.

B. DISCUSSION

A comparison of the suggested model's performance against four well-known models for semantic segmentation—SEiPV-Net, DeepLabv3+, U-Net, and PSP-Net—is shown in TABLE 1. In TABLE 1, bold highlighting is used to indicate superior outcomes for emphasis. The experimental findings clearly demonstrate that the proposed model using equal class weights exhibited optimal performance achieving a Dice coefficient of 94.08% and a mIoU of 91.01%, which are 8.77% and 4.97% better, respectively, than the state-of-the-art SEiPV-Net, under equal or custom class conditions which was trained on the same dataset.

Notably, the PSP-Net model incorporates a pre-trained ResNet50 [30] backbone, which was also trained on an ImageNet dataset, but the DeepLabv3+ model is supported by an Xception backbone that was trained on an ImageNet dataset [31]. In conjunction with the analysis of the experimental results based on the evaluation metrics, a visual inspection was also conducted. FIGURE 6 depicts the predicted masks versus the ground truth ones. While certain defects, such as gridlines, inactive areas, and cracks, encompass a relatively limited pixel range within EL images, the proposed model demonstrated the capability to accurately detect them.

As indicated by the authors in [10], the dataset exhibits shortcomings caused by inaccuracies in labeling. A thorough visual examination of FIGURE 6 confirms this observation. The proposed model not only adeptly identifies small-sized defects but also identifies other defects the dataset annotators failed to notice, as illustrated by the pink boxes. For example, in the test image shown in FIGURE 6(a), there are multiple gridline defects that were overlooked by the dataset annotators. They are shown as thin orange horizontal lines bounded by pink boxes in the original image and its corresponding predicted mask. Furthermore, the model is also capable of generating more precise masks, as evidenced in the case of test image depicted in FIGURE 6(d) and

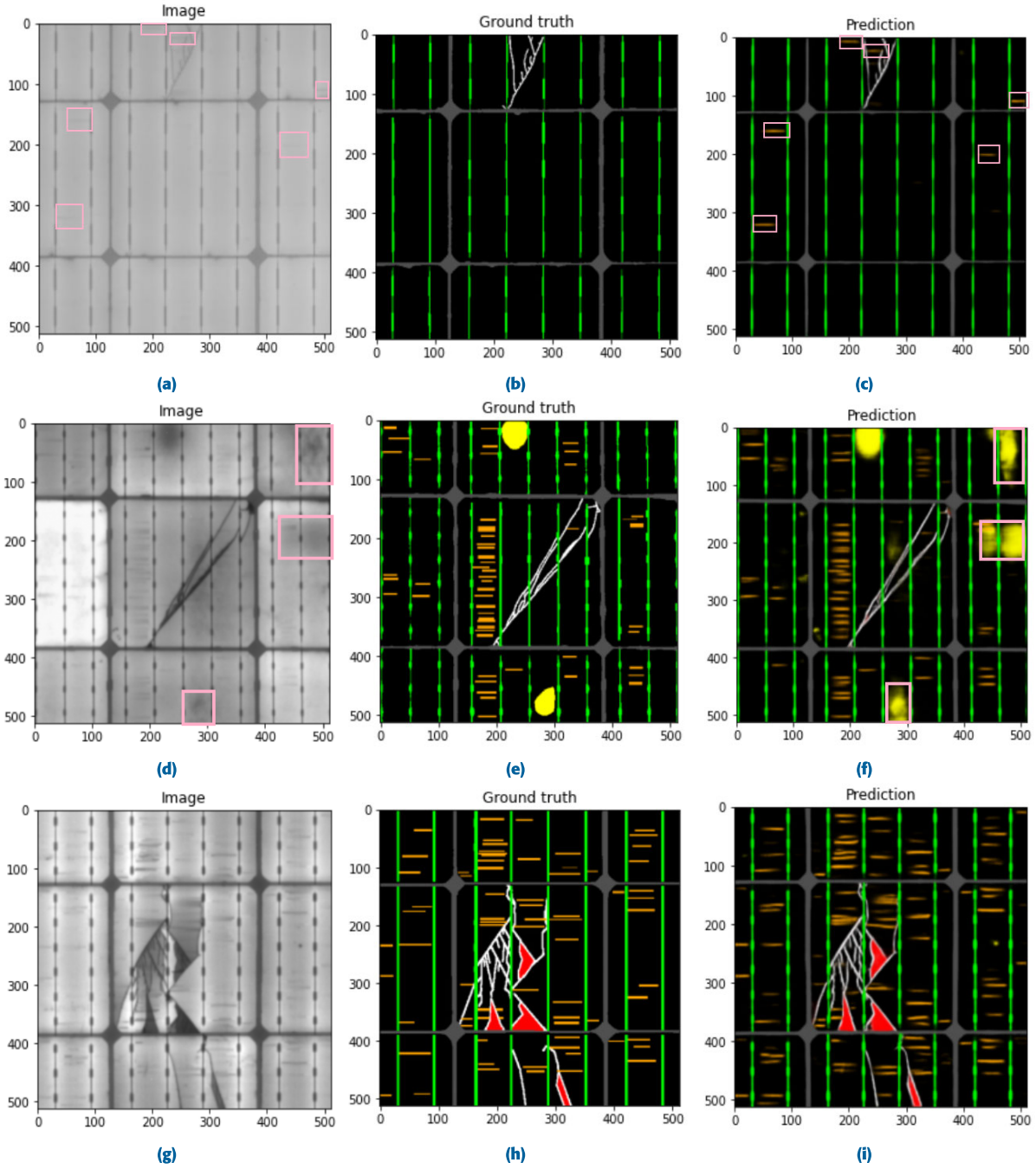

FIGURE 6. (a), (d), (g) EL images (b), (e), (h) Ground truth mask (c), (f), (i) Predicted mask.

its corresponding predicted mask. The two pink boxes on the right illustrate some defects overlooked by the annotators. Additionally, the pink box at the bottom of the image demonstrates that the proposed model can generate a more accurate mask than the provided ground truth. The TABLE 4 presents the Dice coefficient, precision, recall, IoU, F1 score, and mIoU values utilizing the three loss functions, with the bold values emphasizing the optimal results.

The proposed model's superior performance is justified through a comparative analysis with the study carried out by Pratt et al. [9], which used the same dataset. In their study, the authors employed equal, inverse, and custom class weights to train four separate deep-learning models, thereby resulting in twelve different model configurations.

The evaluation used the same subset consisting of three defect types, crack, gridline, and inactive, and two features, ribbons and spacing. The performance assessment used two

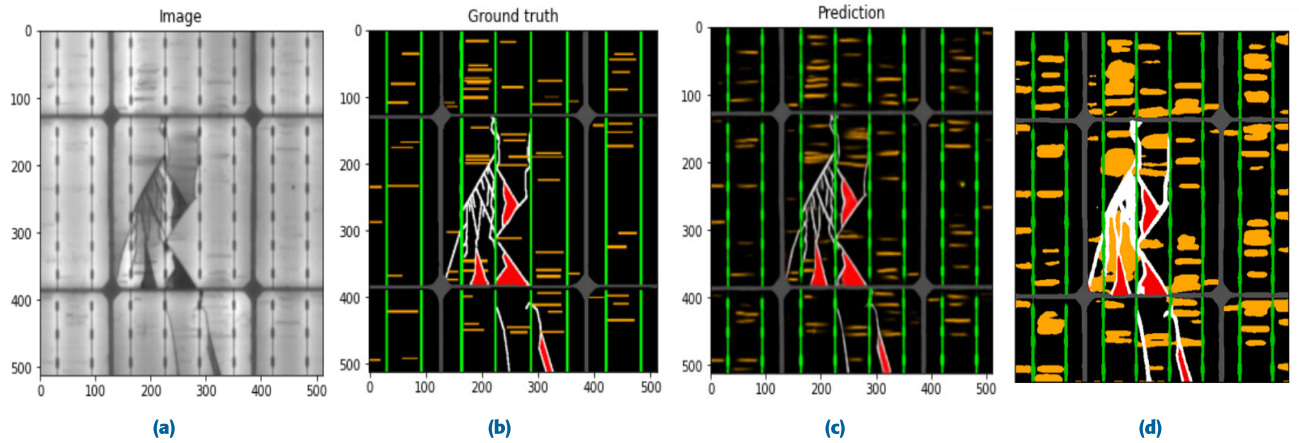

FIGURE 7. (a) EL image (b), Ground truth mask (c), Our predicted mask (d), DeepLabv3+ Pratt et al. [9].

TABLE 1. Comparison of the improved SegNet with state-of-the-art (SOTA) models, highlighting superior results in bold.

Method	Class weights	Dice Coefficient	Precision	IoU	Recall	F1 Score	mIoU
U-NET [10]	Custom	75.03	94.36	60.38	90.28	92.07	80.65
PSP-NET [10]	Custom	80.09	93.85	67.14	90.55	92.10	83.29
DeepLabV3+ [10]	Custom	82.18	93.25	69.98	89.60	91.29	84.57
SEiPV-Net [10]	Custom	83.12	94.63	71.24	92.90	93.75	85.73
U-NET [10]	Equal	79.57	94.64	66.33	91.29	92.90	81.45
PSP-NET [10]	Equal	82.27	93.14	70.43	91.69	92.41	85.16
DeepLabV3+ [10]	Equal	83.63	93.46	71.98	91.71	92.58	83.16
SEiPV-Net [10]	Equal	85.31	94.91	74.47	93.62	94.26	86.04
Attention-Based SegNet [ours]	Equal	**94.08**	**95.46**	**89.02**	**95.31**	**95.38**	**91.01**

metrics: the median recall (mRcl) and the median intersection over union (medIoU).

As the results in TABLE 2 reflect, the DeepLabv3+ model with custom class weights (4c) from Pratt et al.'s study achieved the highest medIoU, with averages of 0.28 for the three defects and 0.70 for the two features. However, the proposed model exceeded these results, achieving an average medIoU of 0.67 for defects and 0.79 for features.

Additionally, the proposed model exhibited exceptional ability in precisely identifying thin, intricate defects like cracks, as shown in FIGURE 7(c). It accurately segmented gridline defects, even though they are small and thin, a task that the DeepLabv3+ model by Pratt et al. could not achieve due to noticeable segmentation overlap and dilation. [9]. These results underscore the advanced precision and overall enhanced performance of the proposed model. TABLE 3 provides a summary of the mRcl and the corresponding averages for a consistent set of five classes. Models utilizing equal class weights (1a, 2a, 3a, 4a) achieved the lowest average mRcl for defects, whereas models employing inverse class weights (1b, 2b, 3b, 4b) exhibited the highest average mRcl for defects, ranging from 0.48 to 0.77. The DeepLabv3+ model with custom class weights (4c) emerged as one of the top performers in terms of the average mRcl for both features and defects. Despite the high recall demonstrated by DeepLabv3+ (4c) across the three defect classes, the medIoU remained relatively low due to low precision. This is visually indicated in FIGURE 7(d) by the dilated predicted masks of defects such as cracks and gridlines relative to the ground truth mask. Notably, the proposed model outperformed the DeepLabv3+ with custom class weights (4c), achieving an average mRcl for defects of 0.79.

Furthermore, a comparative assessment using three distinct loss functions: focal loss [32], categorical cross-entropy loss, and weighted categorical cross-entropy [33] has been conducted.The focal loss represents a refined version of the conventional cross-entropy loss, specifically designed to mitigate the challenges associated with class imbalance [34]. The authors introduced a focusing parameter, γ, to enable a smooth adjustment of the down-weighting rate for simpler examples. Throughout the training process, greater emphasis is placed on challenging examples, while lesser emphasis is placed on well-classified examples. In our experiments, we adhered to the authors' guidance and set the gamma value to 2. Furthermore, for better accuracy, the authors utilized an alpha-balanced variant of the focal loss, with the α value set to 0.25 based on its optimal performance. According to [34], Weighted cross-entropy is widely used for semantic

TABLE 2. Median IoU for U-Net_12 (1a/b/c), U-Net_25 (2a/b/ c), PSPNet (3a/b/c), and DeepLabv3+ (4a/b/c), highlighting superior results in bold.

Method	Weights	crack	gridline	inactive	ribbons	spacing	avg_defects	avg_features
1a [9]	equal	0.13	0.32	0.25	0.68	0.78	0.24	0.73
1b [9]	inverse	0.19	0.15	0.00	0.57	0.70	0.11	0.63
1c [9]	custom	0.26	0.22	0.00	0.58	0.72	0.16	0.65
2a [9]	equal	0.15	0.32	0.09	0.72	0.77	0.19	0.74
2b [9]	inverse	0.19	0.15	0.04	0.54	0.67	0.13	0.61
2c [9]	custom	0.22	0.18	0.00	0.57	0.74	0.13	0.66
3a [9]	equal	0.03	0.23	0.12	0.68	0.80	0.13	0.74
3b [9]	inverse	0.15	0.17	0.00	0.61	0.74	0.10	0.68
3c [9]	custom	0.17	0.14	0.11	0.61	0.77	0.14	0.69
4a [9]	equal	0.00	0.00	0.13	0.70	0.79	0.04	0.75
4b [9]	inverse	0.18	0.14	0.20	0.56	0.66	0.17	0.61
4c [9]	custom	0.25	0.20	0.38	0.63	0.77	0.28	0.70
Attention-Based SegNet[Ours]	equal	**0.42**	0.26	**1.00**	0.68	**0.90**	**0.67**	**0.79**

TABLE 3. Median recall for U-Net_12 (1a/b/c), U-Net_25 (2a/b/ c), PSPNet (3a/b/c), and DeepLabv3+ (4a/b/c).

Method	Weights	crack	gridline	inactive	ribbons	spacing	avg_defects	avg_features
1a [9]	equal	0.14	0.44	0.32	0.74	0.91	0.30	0.83
1b [9]	inverse	0.92	0.87	0.00	0.95	0.97	0.60	0.96
1c [9]	custom	0.90	0.84	0.00	0.97	0.96	0.58	0.96
2a [9]	equal	0.18	0.43	0.20	0.85	0.88	0.27	0.86
2b [9]	inverse	0.92	0.93	0.05	0.97	0.97	0.63	0.97
2c [9]	custom	0.89	0.93	0.00	0.99	0.96	0.61	0.98
3a [9]	equal	0.03	0.30	0.17	0.81	0.90	0.17	0.85
3b [9]	inverse	0.62	0.82	0.00	0.95	0.95	0.48	0.95
3c [9]	custom	0.43	0.68	0.18	0.96	0.96	0.43	0.96
4a [9]	equal	0.00	0.00	0.15	0.83	0.87	0.05	0.85
4b [9]	inverse	0.93	0.92	0.46	0.97	0.99	0.77	0.98
4c [9]	custom	0.86	0.85	0.55	0.98	0.95	0.75	0.96
Attention-Based SegNet[Ours]	equal	0.82	0.56	**1.00**	0.86	**1.00**	**0.79**	0.93

TABLE 4. Comparison of the evaluation metrics using different loss functions.

	focal_loss (alpha=0.25, gamma=2.0)	Categorical cross entropy	Weighted categorical cross entropy
accuracy	95.16	95.07	**95.37**
dice_coef	88.42	92.76	**94.08**
iou	79.55	86.75	**89.02**
jacard	79.55	86.75	**89.02**
precision_m	95.54	95.34	**95.46**
recall_m	94.85	94.85	**95.31**
f1_m	95.19	95.09	**95.38**
specificity	99.84	99.83	**99.84**
mean_iou	90.99	89.93	**91.01**

segmentation with imbalanced datasets. Consequently, this study employed the weighted categorical cross-entropy loss with equal class weight set to a value of 1. Upon examination of the visual performance depicted in FIGURE 8 using these loss functions, it is evident that the focal loss and categorical cross-entropy loss exhibit inadequate performance, as they demonstrate a lack of precision in segmenting the defects. For example, in FIGURE 8(a), the categorical cross-entropy loss failed to identify jbox feature presented in blue color in the ground truth mask in FIGURE 8(b). Additionally, the focal loss masks contains many false positive gridlines depicted in orange color as indicated by white boxes in FIGURE 8(e). Moreover, as demonstrated in FIGURE 8(j) and (i) respectively, in the case of the test image represented in FIGURE 8(f), both the focal loss and categorical cross-entropy loss demonstrated limited capability in detecting the inactive area defect, depicted in red within the ground truth mask. This visual analysis is numerically demonstrated by the results of TABLE 4. Finally, for the last test image FIGURE 8(k), the focal and categorical

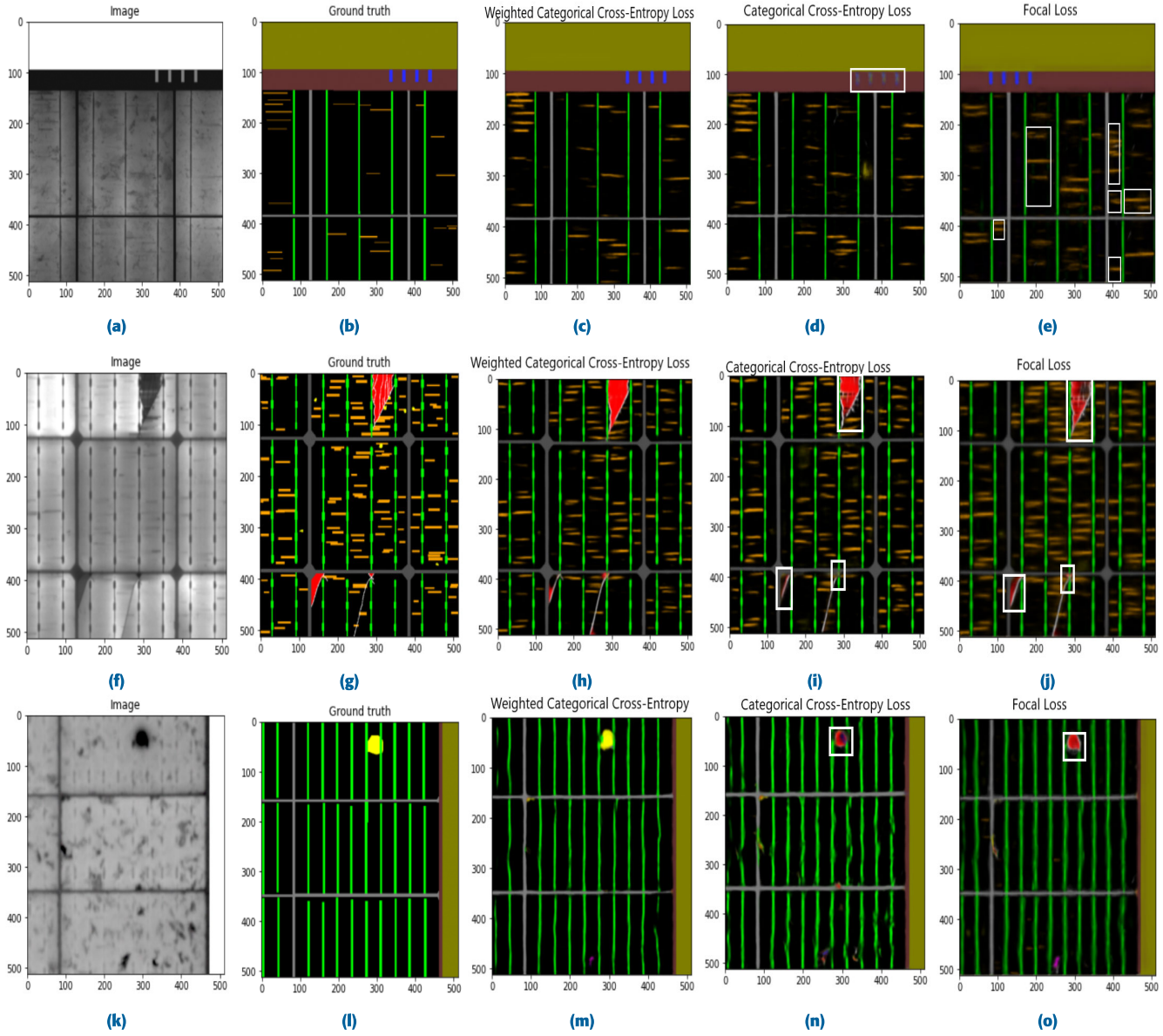

FIGURE 8. (a), (f), (k) EL images (b), (g), (i) GT mask (c), (h), (m) Weighted categorical cross-entropy (d), (i), (n) Categorical cross-entropy (e), (j), (o) Focal loss.

cross-entropy loss functions were unable to detect the scuff defect characterized by yellow color in the ground truth mask FIGURE 8(l) and incorrectly predicted it as inactive area characterized by red colors as shown by white boxes in FIGURE 8(o) and FIGURE 8(n) respectively. In contrast, the weighted categorical cross-entropy loss displays a significantly more accurate segmentation capability for test images (a), (f), and (k). Compared to the groundtruth masks, it has produced the most close and precise segmentation masks depicted in FIGURE 8 (c), (h), and (m).

V. CONCLUSION

This work presents an encoder-decoder framework namely Attention-Based SegNet for semantic segmentation of 29 defects and features of PV modules in EL images. The Attention-Based SegNet replaces the traditional SegNet encoder by the VGG16 encoder with its pretrained weights to make use of transfer learning in feature extraction. Then a CBAM module is added enhance the decoder's ability to generate fine-grained segmentations. While CBAM has demonstrated effectiveness in enhancing the performance of convolutional neural networks (CNNs) by incorporating attention mechanisms, its application in semantic segmentation networks presents challenges. When increasing the image size in the context of the SegNet architecture, there will be a tradeoff between memory efficiency and spatial information preservation during the upsampling process in the decoder network. Furthermore, the default Categorical

Cross-entropy loss function was replaced by Weighted Categorical cross-entropy and focal loss functions and performance is compared.The suggested model is trained, validated and tested using publicly available 29-class dataset. The Attention-Based SegNet model could efficiently detect multiple-scale defects in PV cell EL images, even under complex background conditions. The proposed methodology has demonstrated satisfactory performance in comparison to the SEiPV-Net and the previously employed models in prior investigations like PSP-Net, U-Net, and DeepLabv3+. It is important to highlight that the proposed model, along with other models, attained lower medIoU and mRcl for small, narrow defects like cracks and gridlines, in contrast to larger features such as spacing and ribbons. This research shows that there is a great deal of promise for the suggested model in the PV industry's automated defect semantic segmentation and quality control. According to this study, the suggested model has a lot of potential for use in semantic segmentation and quality control in the photovoltaic industry, offering a lightweight practical and non-intrusive method of extending the lifespan and reliability of PV modules. Future directions may be toward generating new versions of the dataset to address the inaccurate labeling problem. Another important consideration for creating more accurate masks is the incorporation of new loss functions.

REFERENCES

[1] M. A. El-Rashidy, "An efficient and portable solar cell defect detection system," *Neural Comput. Appl.*, vol. 34, no. 21, pp. 18497–18509, Nov. 2022.

[2] A. Shahsavari, F. T. Yazdi, and H. T. Yazdi, "Potential of solar energy in Iran for carbon dioxide mitigation," *Int. J. Environ. Sci. Technol.*, vol. 16, no. 1, pp. 507–524, Jan. 2019.

[3] G. Masson and I. Kaizuka, "IEA PVPS trends in photovoltaic applications 2020," *IEEE Access*, pp. 1–88, 2020. [Online]. Available: https://iea-pvps.org/wp-content/uploads/2020/11/IEAPVPSTrendsReport2020-1.pdf

[4] IEA. (2014). *How Solar Energy Could Be the Largest Source of Electricity By Mid-century—News*. Accessed: Apr. 1, 2023. [Online]. Available: https://www.iea.org/news/how-solar-energy-could-be-the-largest-source-of-electricity-by-mid-century

[5] Y. Jiang and C. Zhao, "Attention classification-and-segmentation network for micro-crack anomaly detection of photovoltaic module cells," *Sol. Energy*, vol. 238, pp. 291–304, May 2022.

[6] L. Pratt, D. Govender, and R. Klein, "Defect detection and quantification in electroluminescence images of solar PV modules using U-net semantic segmentation," *Renew. Energy*, vol. 178, pp. 1211–1222, Nov. 2021.

[7] M. W. Akram, G. Li, Y. Jin, X. Chen, C. Zhu, X. Zhao, A. Khaliq, M. Faheem, and A. Ahmad, "CNN based automatic detection of photovoltaic cell defects in electroluminescence images," *Energy*, vol. 189, Dec. 2019, Art. no. 116319.

[8] Y. Jiang, W. Wang, and C. Zhao, "A machine vision-based realtime anomaly detection method for industrial products using deep learning," in *Proc. Chin. Autom. Congr. (CAC)*, Nov. 2019, pp. 4842–4847.

[9] L. Pratt, J. Mattheus, and R. Klein, "A benchmark dataset for defect detection and classification in electroluminescence images of PV modules using semantic segmentation," *Syst. Soft Comput.*, vol. 5, Dec. 2023, Art. no. 200048.

[10] H. Eesaar, S. Joe, M. U. Rehman, Y. Jang, and K. T. Chong, "SEiPV-net: An efficient deep learning framework for autonomous multi-defect segmentation in electroluminescence images of solar photovoltaic modules," *Energies*, vol. 16, no. 23, p. 7726, Nov. 2023.

[11] D. H. Kang and Y.-J. Cha, "Efficient attention-based deep encoder and decoder for automatic crack segmentation," *Structural Health Monitor.*, vol. 21, no. 5, pp. 2190–2205, Sep. 2022.

[12] E. Sovetkin, E. J. Achterberg, T. Weber, and B. E. Pieters, "Encoder–decoder semantic segmentation models for electroluminescence images of thin-film photovoltaic modules," *IEEE J. Photovolt.*, vol. 11, no. 2, pp. 444–452, Mar. 2021.

[13] J. Fioresi, D. J. Colvin, R. Frota, R. Gupta, M. Li, H. P. Seigneur, S. Vyas, S. Oliveira, M. Shah, and K. O. Davis, "Automated defect detection and localization in photovoltaic cells using semantic segmentation of electroluminescence images," *IEEE J. Photovolt.*, vol. 12, no. 1, pp. 53–61, Jan. 2022.

[14] M. R. U. Rahman and H. Chen, "Defects inspection in polycrystalline solar cells electroluminescence images using deep learning," *IEEE Access*, vol. 8, pp. 40547–40558, 2020.

[15] H. Han, C. Gao, Y. Zhao, S. Liao, L. Tang, and X. Li, "Polycrystalline silicon wafer defect segmentation based on deep convolutional neural networks," *Pattern Recognit. Lett.*, vol. 130, pp. 234–241, Feb. 2020.

[16] Y. Wang, T. Hou, X. Zhang, H. Shangguan, P. Zhang, J. Li, and B. Wei, "Surface defect detection of solar cell based on similarity non-maximum suppression mechanism," *Signal, Image Video Process.*, vol. 17, no. 5, pp. 2583–2593, Jul. 2023.

[17] X. Zhang, Y. Hao, H. Shangguan, P. Zhang, and A. Wang, "Detection of surface defects on solar cells by fusing multi-channel convolution neural networks," *Infr. Phys. Technol.*, vol. 108, Aug. 2020, Art. no. 103334.

[18] F. M. A. Mazen, R. A. A. Seoud, and Y. O. Shaker, "Deep learning for automatic defect detection in PV modules using electroluminescence images," *IEEE Access*, vol. 11, pp. 57783–57795, 2023.

[19] W. Tang, Q. Yang, K. Xiong, and W. Yan, "Deep learning based automatic defect identification of photovoltaic module using electroluminescence images," *Sol. Energy*, vol. 201, pp. 453–460, May 2020.

[20] A. R. Rodriguez, B. Holicza, A. M. Nagy, Z. Vörösházi, G. Bereczky, and L. Czúni, "Segmentation and error detection of PV modules," in *Proc. IEEE 27th Int. Conf. Emerg. Technol. Factory Autom. (ETFA)*, Sep. 2022, pp. 1–4.

[21] J. Lewis, Y.-J. Cha, and J. Kim, "Dual encoder–decoder-based deep polyp segmentation network for colonoscopy images," *Sci. Rep.*, vol. 13, no. 1, p. 1183, Jan. 2023.

[22] M. Zhang and L. Yin, "Solar cell surface defect detection based on improved YOLO v5," *IEEE Access*, vol. 10, pp. 80804–80815, 2022.

[23] R. Ali and Y.-J. Cha, "Attention-based generative adversarial network with internal damage segmentation using thermography," *Autom. Construction*, vol. 141, Sep. 2022, Art. no. 104412.

[24] S. Kimball and P. Mattis, "Gimp 2.10.32," *IEEE Access*, 2022. [Online]. Available: https://www.gimp.org/

[25] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[27] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[28] L. Li, B. Fang, and J. Zhu, "Performance analysis of the YOLOv4 algorithm for pavement damage image detection with different embedding positions of CBAM modules," *Appl. Sci.*, vol. 12, no. 19, p. 10180, Oct. 2022.

[29] Y. Guo, S. E. Aggrey, X. Yang, A. Oladeinde, Y. Qiao, and L. Chai, "Detecting broiler chickens on litter floor with the YOLOv5-CBAM deep learning model," *Artif. Intell. Agricult.*, vol. 9, pp. 36–45, Sep. 2023.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[31] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. CVPR*, Jun. 2009, pp. 248–255.

[32] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007.
[33] Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2020.
[34] S. Jadon, "A survey of loss functions for semantic segmentation," in *Proc. IEEE Conf. Comput. Intell. Bioinf. Comput. Biol. (CIBCB)*, Oct. 2020, pp. 1–7.

FATMA MAZEN ALI MAZEN received the B.Sc. degree (Hons.) in electrical engineering from the Communications and Electronics Department, Faculty of Engineering, Fayoum University, in 2011, and the M.Sc. and Ph.D. degrees from the Faculty of Engineering, Fayoum University, in 2016 and 2021, respectively. She was a Demonstrator with the Faculty of Engineering, Fayoum University, from 2012 to 2016. She was also a Teaching Assistant, from 2017 to 2021, and has been an Assistant Professor, since 2021. Her research interests include artificial neural networks, machine learning, meta-heuristic optimization, deep learning, natural language processing, and computer vision.

YOMNA O. SHAKER (Member, IEEE) received the B.Sc. degree (Hons.) from Fayoum University, Egypt, in 1998, and the M.Sc. and Ph.D. degrees from Cairo University, Cairo, Egypt, in 2003 and 2010, respectively. She was a Visiting Scholar with American University of Sharjah, from 2012 to 2015. She is currently an Assistant Professor with the Department of Electrical Engineering, University of Science and Technology of Fujairah (USTF), United Arab Emirates, and Fayoum University (on leave). Her research and teaching interests include high-voltage equipment, electrical machines, and renewable energy.

RANIA AHMED ABUL SEOUD received the master's and Ph.D. degrees in computer engineering from the Faculty of Engineering, Cairo University (CU). She is currently the Vice Dean of Post Graduates Studies and Research Affairs with the Faculty of Engineering, Fayoum University, Egypt, where she is also a Chief Information Editor. She is also a Full Tenured Professor in computer engineering with FU, where she has also acted on multiple occasions as the Acting Dean of the Faculty of Computers and information. She was a Professor and an Invited Lecturer in different academic institutes around Egypt. She has conducted internationally acclaimed research in artificial intelligence, bioinformatics, computer networks, big data, cloud computing, and electronics. She has authored over 60 scientific publications and supervised 20 doctoral and master's students.