

Received February 5, 2020, accepted February 22, 2020, date of publication February 27, 2020, date of current version March 9, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2976843

Defects Inspection in Polycrystalline Solar Cells Electroluminescence Images Using Deep Learning

MUHAMMAD RAMEEZ UR RAHMAN^{ID} AND HAIYONG CHEN^{ID}

School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin 300130, China

Corresponding author: Haiyong Chen (haiyong.chen@hebut.edu.cn)

This work was supported in part by the National Natural Science Foundation (NNSF) of China under Grant 61873315, in part by the Natural Science Foundation of Hebei Province, under Grant F2018202078, in part by the Technology Project of Hebei Province under Grant 17211804D, and in part by the Outstanding Youth Science Fund of Hebei Province under Grant F2017202062.

ABSTRACT Solar cells defects inspection plays an important role to ensure the efficiency and lifespan of photovoltaic modules. However, it is still an arduous task because of the diverse attributes of electroluminescence images, such as indiscriminative complex background with extremely unbalanced defects and various types of defects. In order to deal with these problems, this paper proposes a new precise and accurate defect inspection method for photovoltaic electroluminescence (EL) images. The proposed algorithm leverages the advantage of multi attention network to efficiently extract the most important features and neglect the nonessential features during training. Firstly, we designed a channel attention to exploit contextual representations and spatial attention to effectively suppress background noise. Secondly, we incorporate both attention networks into modified U-net architecture and named it multi attention U-net (MAU-net) to extract effective multiscale features for defects inspection. Finally, we propose a hybrid loss which combines focal loss and dice loss aiming to solve two problems: a) overcome the class imbalance problem, and b) allowing the network to train with irregular image labels for some complex defects. The proposed multi attention U-net is evaluated on real photovoltaic EL images datasets using 5-fold cross validation technique. Experimental results demonstrate that the proposed network can segment and detect various complex defects correctly. The proposed method achieved the mean intersection over-union (m-IOU) of 0.699 and F-measure of 0.799 which outperforms the previous methods.

INDEX TERMS Multi-attention U-net, cracks segmentation, electroluminescence images, defects inspection.

I. INTRODUCTION

In this era of technology, solar energy provides the most elegant solution for arising energy demand by enabling generation at any scale [1]. Among different solar cell technologies, polycrystalline solar cells dominate the monocrystalline solar cells due to cost. During the production process, solar cells may get damaged due to thermal stress or improper operations. The damage may be due to defects such as finger interruptions, cracks or cell breakages etc. Among these defects, cracks can cause a severe loss in power efficiency of solar cells because they can electrically disconnect certain areas of solar cells [2]. The more severity of crack will lead to greater

electrical power loss of the module [3]. Therefore, timely detection of these cracks is essential to improve the endurance and reliability of solar cells [4]. Thus, this paper presents a method based on deep learning to automatically segment and detect various crack and finger interruption defects in polycrystalline solar cells.

Electroluminescence (EL) imaging is an important non-destructive technology for defects inspection of solar cells with the ability to capture high-resolution solar cell images [5]. The EL imaging highlights the internal defects such as cracks that are difficult to be recognized by the human eye. These cracks are of different sizes, shapes and orientations and appear darker as compared to the background. Figure.1 shows two defective solar cell images with various types of cracks and finger interruption defect. The defects can

The associate editor coordinating the review of this manuscript and approving it for publication was Min Xia^{ID}.

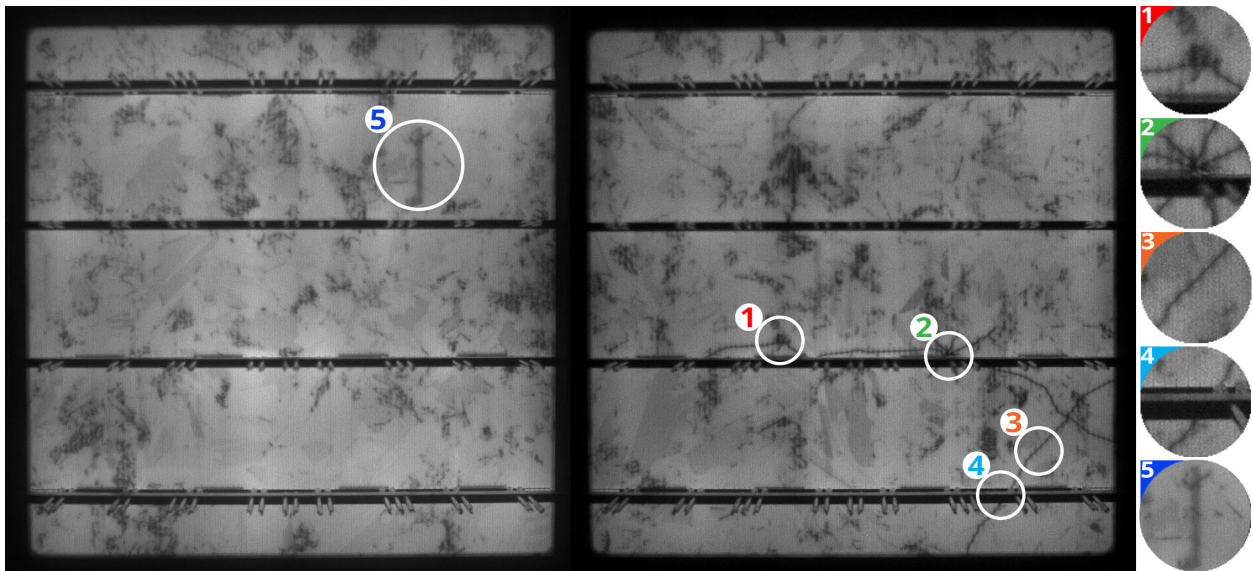


FIGURE 1. Selected polycrystalline solar EL images showing common defect types: (1) Defect submerged into background (2) Complex star crack defect (3) line crack defect (4) Crack separated by bus bar (5) Finger interruption defect.

be categorized into 1) defects submerged into background, 2) complex star-like cracks 3), line crack defects, 4) crack defect separated by bus bar and 5) finger interruption defects. As compared to monocrystalline EL images, polycrystalline EL images have heterogeneous complex background with randomly distributed crystal grains making random patterns. These patterns are unique for every image and may contain the same pixel intensity value as a defect which makes the defect detection task difficult. As shown in Figure.1 the EL imaging makes the defects prominent. However, the visual inspection of polycrystalline solar cell EL images is time consuming and also require expert’s involvement. Several image processing algorithms are proposed to eliminate expert’s involvement. However, they are not robust and are inaccurate. Figure.2(a) shows the defected EL image. As shown in Figure.2(b), Figure.2(c) and Figure.2(d) the segmentation of crack is inaccurate with large amount of noise using traditional image processing algorithms (such as Otsu threshold, Gaussian threshold and global fixed threshold). Thus, we propose an automatic defects inspection method for polycrystalline solar cells which is fast, robust and accurate.

The defects inspection methods are commonly divided into statistical, structural and filter-based approaches. In statistical approaches, the image is separated into distinct regions based on their statistical behavior. Histogram analysis analyzes the image features based on histogram which is a kind of statistical method. Tsanakas and Botsaris [6] proposed a thermographic method for hot-spot detection in defective solar modules. They used non-destructive thermographic approach to view the photovoltaic patterns by using data gained through line profile and image histogram. The proposed method does not filter hot-spots in the image background. Wakaf and Jalab [7] used images with arbitrary gray-level pixel

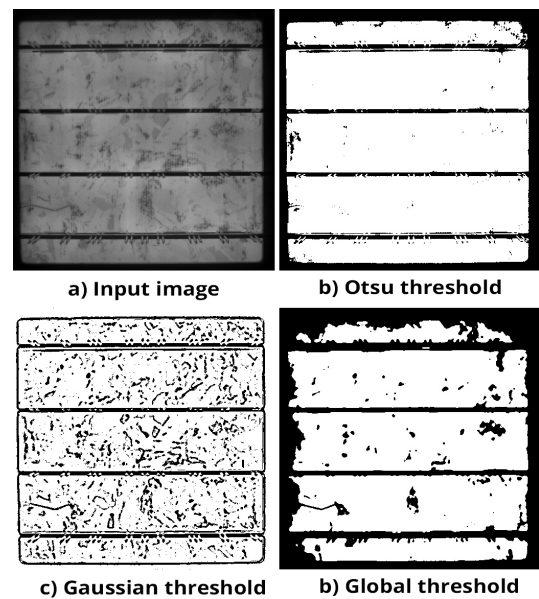


FIGURE 2. Effect of traditional image processing algorithm on defect image: (a) Input image (b) Otsu threshold image (c) Gaussian threshold image (d) Global threshold image.

distribution and propose a method to detect defects from these images. They used histogram matching and separated defective object from foreground from the image background. Gray level co-occurrence matrix (GLCM) is a powerful statistical technique that provides second-order method for producing texture features. GLCM contains frequencies at which two pixels in an image are separated by a certain vector.

The structural analysis methods make up the image textures and spatial arrangements. Qian *et al.* [8] used self-learning features to detect polycrystalline solar cell surface

micro-cracks. In their method, the defect information is obtained from self-learning features then these features are combined with super-pixel segmentation for defect regions localization. The proposed method gives significant results but they used a very small number of defective samples. Furthermore, the method is only evaluated on simple micro-crack defects. Tsai *et al.* [9] used a Haar-like feature extraction techniques and proposed a novel fuzzy c-means clustering technique for cracks and finger interruption defects in polycrystalline solar cell EL images.

Filter based methods are implemented in the spatial domain, frequency domain or joint spatial/spatial-frequency domain. The filter-based methods can carefully select the points of interest and detect irregularities. A particular method has been introduced in [10] using Fourier image reconstruction. The methods eliminate unwanted objects by setting their frequency in the frequency domain to zero and then transforms to a spatial image. Their method takes into account only the straight line-like defects. Chen *et al.* [11] generated an enhanced saliency map using a novel steerable evidence filter then applied morphological operations to accomplish the segmentation of solar cell EL defects. Anwar and Abdullah [12] detects and segment micro cracks in polycrystalline solar cell images using a method based on anisotropic diffusion. The method performs well but didn't consider the cracks submerged with background.

In recent years, deep convolutional neural networks (DCNN) has made convincing progress in the computer vision domain. In comparison to traditional methods, DCNN provides better solutions for arduous problems such as image segmentation [13] and scene recognition [14]. Researcher in the field of industrial defect inspection also adopts the power of DCNN to solve the problems of segmentation and classification of industrial defects. Recently, Han *et al.* [15] added a region proposal network (RPN) into U-net [16] and used dilated convolutions to segment polycrystalline silicon wafer defects. The method gives reasonable results but combining RPN and dilated convolution makes the detection slower. The U-net gives significant results in segmentation tasks with small-scale datasets. Despite having symmetric skip connections to fuse encoder and decoder features, U-net like architecture still faces degradation problem which can be solved by adding residual connections within encoder and decoder [17]. Another problem with U-net is that it fuses low-level features with high-level features based on fixed weights. In order to solve this problem, this paper proposes a multi-attention network to weigh the feature maps according to their importance and suppresses the irrelevant information.

Inspired by human visual system, attention mechanism is a popular trend in deep learning which has proven to offer significant results for image captioning [18], machine translation [19] and classification [20]. Chen *et al.* [21] introduced spatial and channel attention and incorporated into a DCNN for image captioning. Channel attention highlights the most

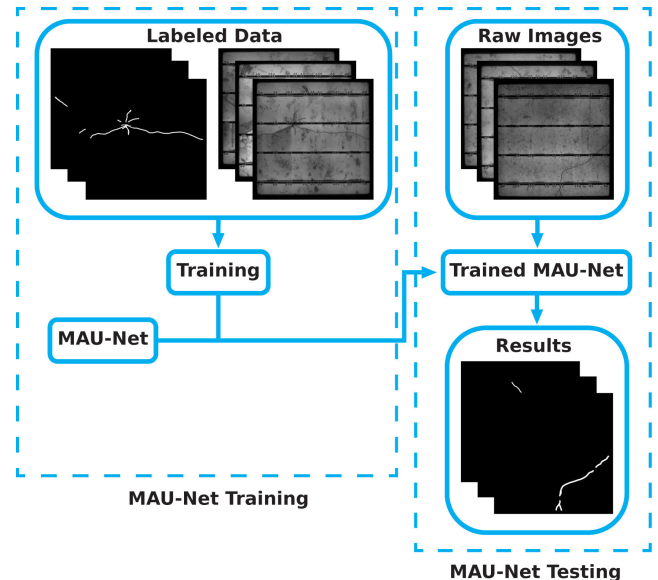


FIGURE 3. Schematic diagram of Multi attention U-net (MAU-net).

important features while spatial attention suppresses the noise information. The architecture outperforms previous methods for image captioning. Xu *et al.* [22] confirm significant improvements in speech recognition task by adding spatial and channel attention into the CNN. Henceforth, we introduce a multi-attention network consisting of spatial and channel attention and incorporate it into modified U-net to solve defects segmentation problem. The schematic diagram of the proposed network is shown in Figure.3. The proposed method effectively improves the performance and speed with a smaller number of parameters. The main contributions of this paper are as follows,

1. A multi attention network is proposed consisting of spatial attention and channel attention which helps to learn and weigh multiscale feature map channels. The channel attention emphasizes on defect regions while spatial attention suppresses the background noise.
2. The proposed multi attention network is added into the modified U-net to accomplish defects inspection in Photovoltaic electroluminescence images. The insertion of multi attention network allows the network to utilize contextual and spatial information effectively.
3. A hybrid loss function is presented to train Multi attention U-net. The hybrid loss combines dice loss and focal loss which solves class imbalance problem and allows the network to learn poorly classified complex cracks. With the proposed loss function and multi attention network, we achieve a reasonable increase in performance.
4. The Multi attention U-net is trained and evaluated on a real industrial dataset. The results are compared with former methods for defects inspection of solar cells. The results verify the significant improvement in defects inspection using Multi attention U-net.

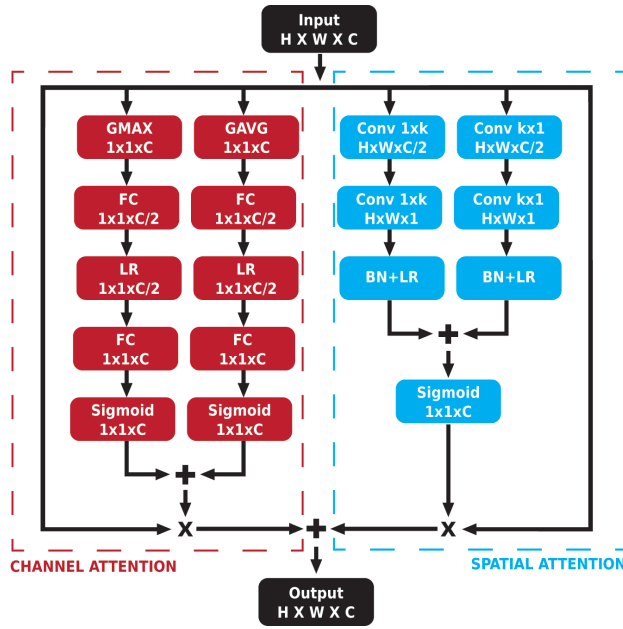


FIGURE 4. Multi attention network.

II. METHODOLOGY

A. MULTI ATTENTION NETWORK

Attention mechanism is inspired by the human visual system. Human visual attention allows to emphasize on a certain area while defocusing the surrounding region. Most convolutional neural networks for defects inspection give equal contribution to all channels, which is ineffective for feature extraction. Generally, each channel of CNNs generate different semantic information from the image. Some features may be valuable for defect localization, while some features may comprise of noises and can result in redundant information which in turn lead to bad segmentation. So, with the aim to extract effective features, a CNN must be capable of highlighting the defects while suppressing the background information. In this paper, we propose a multi attention network, which consists of channel attention and spatial attention to emphasize the defective area and suppress the unwanted details.

As shown in Figure.4, the input features C^h are firstly refined using channel attention and spatial attention, then we will get the channel attention map C^h and spatial attention map S^h . C^h and S^h are used to reweight the input features. Then these features are added together to complete the phenomena of multi attention network. The multi attention network is incorporated into U-net which gives a boost to the performance and will be verified by following experimental results. The phenomenon of multi attention network is given in Equation (1).

$$MAN = S^h + C^h \quad (1)$$

where C^h and S^h are channel attention and spatial attention respectively.

B. CHANNEL ATTENTION

In order to extract effective features, a CNN must be capable of highlighting the defect regions. For, this purpose, we propose a Channel attention mechanism which enables the network to emphasize on the defect regions. The Squeeze and excitation (SE) block in [23] is designed for classification task, and it plays a significant role in improving the performance of the network by placing it in each block of the network. For segmentation task, we improved the original SE block by adding a global max pooling branch together with global average pooling branch. In [24] and [25] the significance of Global max pooling and Global average pooling is shown for distinct objects recognition. After squeezing the inputs using Global max pooling and global average pooling, we apply fully connected layer followed by LeakyRelu layer to add non-linearity and decrease the number of channels to a certain ratio. Then the second fully connected layer followed by a sigmoid activation layer is added to provide each channel a smooth gating function. Finally, we add the outputs of both branches that will be used to re-weight the input feature maps.

As shown in Figure.4 the channel attention block consists of two branches. Global Max pooling (G_{max}) and Global Average pooling (G_{avg}) are used to squeeze the input features in first and second column respectively. In first branch (G_{avg}) is applied to squeeze each input channel $C^h = \{c_0^h, c_1^h, \dots, c_n^h\}$ into a single a single numeric value V^h . This results in $1 \times 1 \times n$ dimensional output $V^h = \{v_0^h, v_1^h, \dots, v_n^h\}$ where n is the number of input channels. Furthermore, V^h is given into a fully connected layer followed by a Leaky Relu activation function which gives output $A^h = \{a_0^h, a_1^h, \dots, a_n^h\}$ of $1 \times 1 \times n/4$ dimension. Finally, subsequent fully connected layer and a sigmoid layer gives the output $M^h = \{m_0^h, m_1^h, \dots, m_n^h\}$. At the end, the outputs of both branches are added together giving $O^h = \{o_0^h, o_1^h, \dots, o_n^h\}$. O^h weighs the input channels based on their importance and gives high weights to the most important features.

$$A^h = \sigma_1(f C_2(\delta(f C_1(V^h, W_1))) \quad (2)$$

$$M^h = \sigma_2(f C_4(\delta(f C_3(V^h, W_2))) \quad (3)$$

$$CA = O^h = M^h + A^h \quad (4)$$

where W represents parameters in channel attention, σ represents the simoid function, $f C_1, f C_2, f C_3$ and $f C_4$ represents the fully connected layers and δ represents the LeakyRelu function. The final output will be obtained by weighting the input features with CA and is given in Equation (5).

$$C^h = CA.C^h \quad (5)$$

C. SPATIAL ATTENTION

Electroluminescence images have heterogeneous complex background containing more background region as compared to the defect region. So, the feature maps may contain a lot

of unnecessary information which may result in bad segmentation. In order to focus more on defects rather than on background, we adopt spatial attention block which suppresses background details while preserving defects details. From [26] large receptive field can increase the representational power of the network leading to an increase in accuracy. Dilated convolutions [27] or Atrous convolutions [28] are used to increase the receptive field size deprived of increasing kernel size but they are slow. Inspired by [29] we use separable large kernel convolutions in spatial attention block in order to increase the receptive field and minimizing memory consumption with a smaller number of parameters. As shown in right side of Figure.4 we separate $k \times k$ convolutional kernel into $1 \times k$ and $k \times 1$ kernels. We tried different values of k such as $k = 5$, $k = 9$, $k = 11$ and $k = 15$. However, $k = 5$ has given best results, so, we set $k = 5$ in each experiment. Each convolution operation is followed by a batch-normalization and a non-linear LeakyRelu layer. Furthermore, we use a sigmoid gating function to map the feature maps to $[0, 1]$. This output feature map is mapped to input features to generate effective features for defects segmentation.

$$S_1 = conv_{1 \times k}(conv_{1 \times k}(S^h, W_1), W_2) \quad (6)$$

$$S_2 = conv_{k \times 1}(conv_{k \times 1}(S^h, W_1), W_2) \quad (7)$$

$$SA = \sigma_3(S_1 + S_2) \quad (8)$$

where W represents parameters in spatial attention, σ represents the simoid function, $conv_{1 \times k}$ and $conv_{k \times 1}$ represents the convolution operations and we set $k = 5$ in the experiments. The final output will be obtained by weighting the input features with SA and is given in Equation (9).

$$S^h = CA.S^h \quad (9)$$

To test the effectiveness of Multi-attention network, we incorporated it into the modified U-net and demonstrated the results in the experiments. The results have shown that by adding multi-attention network into U-net, the defects are detected robustly by suppressing the complex background information.

D. MULTI-ATTENTION U-NET

Based on multi-attention network for effective features extraction, we offer a fast and robust Multi attention U-net architecture. The inputs of the network are grayscale images with size $1 \times 512 \times 512$ and the output is a black and white segmentation map of size $1 \times 512 \times 512$. Each pixel represents a defect or background. In this paper, defects are represented as white pixels while background as black pixels. The structure of MAU-net with definition of each layer is shown in Table.1 in which

- Conv represents the convolution layer.
- LR represents Leaky-Rely layer.
- Max-pooling represents the Max Pooling layer.
- Up-sampling represents the Up-sampling layer.
- Sigmoid represents the sigmoid activation later.

- Concat represents the concatenation layer.

The MAU-net architecture is illustrated in Figure.5. It is composed of 9-level encoder and decoder path. Each level on the encoder path consists of two 3×3 convolutions followed by 2×2 Max-pooling operation for down-sampling the feature maps. As the model grows, the numbers of feature maps are doubled while the size of the feature map is halved. A multi attention network is incorporated at each encoder-decoder path to weigh the feature maps to achieve the attention mechanism. The multi attention network allows the network to focus on the defect features while suppresses the background information. The working principle of the multi attention network is explained in the previous section. On each level in decoder path, we firstly apply convolution operation to halve the number of features using 3×3 convolution, then these features are up-sampled using 2×2 up-sampling layer. We use concatenation operation to fuse these up-sampled features with low-level features passing through the multi-attention network which are at the same level in the encoder path. Then a set of two 3×3 convolutions is applied on these feature maps. It should be noted that each convolution layer is followed by a leaky-Relu layer. In the final layer, 1×1 convolution with sigmoid activation is used to output the segmentation map of the original image. It should be noted that unlike U-net we did not use any cropping. The MAU-net simplifies the U-net architecture and increases its representation power with a smaller number of parameters. In addition, the MAU-net is superior to U-net in training speed, test speed with improved segmentation accuracy.

E. HYBRID LOSS

During the training of the network, it is necessary to estimate the weights to increase robustness and accuracy of the network. This is achieved by using a proper loss function which will be minimized during training. Segmentation of objects exhibiting various sizes and shapes is a challenging task in semantic segmentation. Electroluminescence images has imbalanced pixels distribution i.e. the images contain about 95% of background and crack covers only about 5%. As the contribution of these cracks to the loss is less as compared to the background, so the network may result in low performance

In order to overcome this problem, as suggested in [30], we train our model using a hybrid loss function, which combines both dice loss L_D [31] and the binary focal loss L_{BFL} [32]. L_D helps the network to mitigate the class imbalance problem. On the other hand, L_{BFL} helps the network to learn poorly classified examples in an efficient way.

The dice loss L_D helps to overcome the class imbalance problem in Binary segmentation task by formula given in Equation (10).

$$L_D = -2 \frac{\sum_j p_j g_j}{\sum_j (p_i + g_j)} \quad (10)$$

TABLE 1. Structure of the multi attention U-net (MAU-net).

Layers		Type of operation	Kernel size	Filter number	Step	Output map size
B1	L0	Conv1+LR	3×3	32	1	32×512×512
	L1	Conv2+LR	3×3	32	1	32×512×512
	MAN1	Attention1	-	-	-	32×512×512
B2	L2	Max-pooling	2×2	-	2	32×256×256
	L3	Conv3+LR	3×3	64	1	64×256×256
	L4	Conv4+LR	3×3	64	1	64×256×256
B3	MAN2	Attention2	-	-	-	64×256×256
	L5	Max-pooling	2×2	-	2	64×128×128
	L6	Conv5+LR	3×3	128	1	128×128×128
B4	L7	Conv6+LR	3×3	128	1	128×128×128
	MAN3	Attention1	-	-	-	128×128×128
	L8	Max-pooling	2×2	-	2	128×64×64
B5	L9	Conv7+LR	3×3	256	1	256×64×64
	L10	Conv8+LR	3×3	256	1	256×64×64
	MAN4	Attention1	-	-	-	256×64×64
B6	L11	Max-pooling	2×2	-	2	256×32×32
	L12	Conv9+LR	3×3	512	1	512×32×32
	L13	Conv10+LR	3×3	512	1	512×32×32
B7	L14	Conv11+BN+LR	3×3	256	2	256×32×32
	L15	Up-sampling	2×2	-	2	256×64×64
	L16	Concat (MAN4+L15)	-	-	-	256×64×64
B8	L17	Conv12+LR	3×3	256	1	256×64×64
	L18	Conv13+LR	3×3	256	1	256×64×64
	L19	Conv14+LR	3×3	128	1	128×64×64
B9	L20	Up-sampling	2×2	-	2	128×128×128
	L21	Concat (MAN3+L20)	-	-	-	128×128×128
	L22	Conv15+LR	3×3	128	1	128×128×128
B10	L23	Conv16+LR	3×3	128	1	128×128×128
	L24	Conv17+LR	3×3	64	1	64×128×128
	L25	Up-sampling	2×2	-	2	64×256×256
B11	L26	Concat (MAN2+L25)	-	-	-	64×256×256
	L27	Conv18+LR	3×3	64	1	64×256×256
	L28	Conv19+LR	3×3	64	1	64×256×256
B12	L29	Conv20+LR	3×3	64	1	32×256×256
	L30	Up-sampling	2×2	-	2	32×512×512
	L31	Concat (MAN1+L30)	-	-	-	32×512×512
B13	L32	Conv21+LR	3×3	32	1	32×512×512
	L33	Conv22+LR	3×3	32	1	32×512×512
	L34	Conv23+Sigmoid	1×1	1	1	1×512×512

where $p_j \in [0, 1]$ be the j^{th} output of the final layer of the model passed through sigmoid activation layer and $g_j \in [0, 1]$ is the j^{th} ground truth label.

The Binary focal loss can be given by Equation (11).

$$L_{BFL} = - \sum_{j=1}^{C=2} g_j(1 - p_j)^\gamma \log(p_j) \quad (11)$$

where $(1 - p_i)^\gamma$ is the modulating factor with $\gamma \geq 0$ be the focus parameter to make the loss focus on problematic classes. We experimented different values of γ and $\gamma = 2$ produced the best results.

The final hybrid loss can be calculated by Equation (12).

$$Hybrid\ loss = \beta L_{BFL} + (1 - \beta)L_D \quad (12)$$

where β is a parameter introduced to regulate the stability between dice loss and focal loss. The value of β is set to 0.5.

III. EXPERIMENT RESULTS AND ANALYSIS

A. EVALUATION METRIC

After training, it is compulsory to assess the performance of the proposed architectures. Many metrics are used to determine the performance of segmentation results. In reference to [18], predicted results are compared with ground truth images using five metrics: mean-Intersection-over-Union (m-IOU), accuracy, recall, precision and F-measure. The IOU, accuracy, recall, precision and F-measure are computed by Equations (13), (14), (15), (16) and (17) respectively as follows

$$IOU = \frac{TP}{TP + FP + FN} \quad (13)$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (14)$$

$$Recall = \frac{TP}{TP + FP} \quad (15)$$

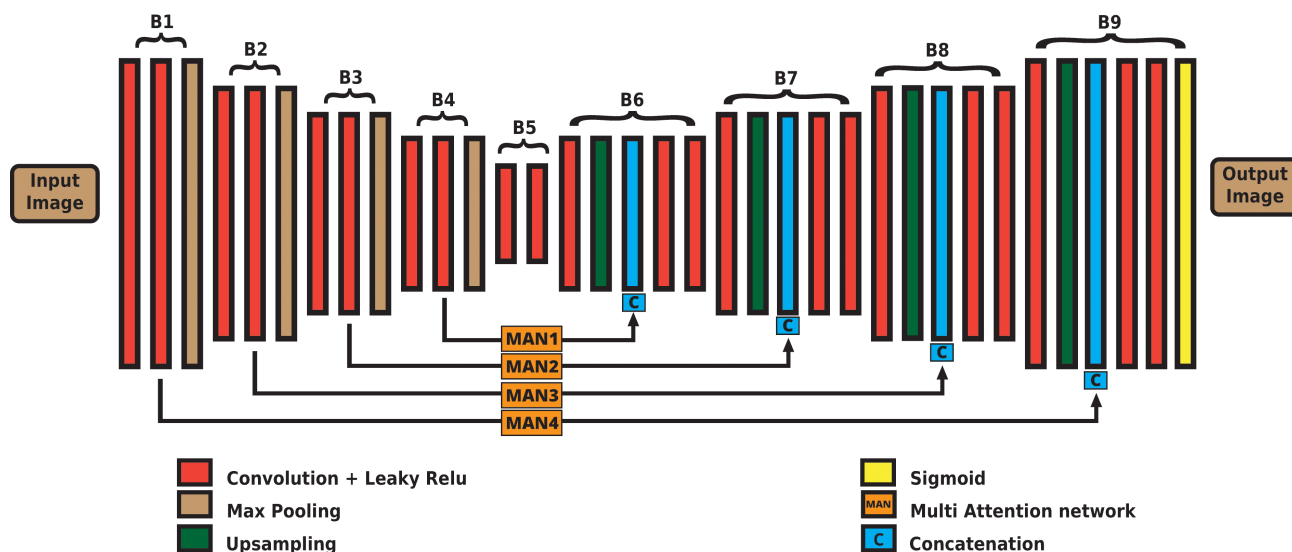


FIGURE 5. Multi attention U-net (MAU-net).

$$Precision = \frac{TP}{TP + TN} \tag{16}$$

$$F - measure = 2x \frac{Precision * Recall}{Precision + Recall} \tag{17}$$

where TP, TN, FP and FN are the number of correctly classified defect pixels, the number of background pixels that are classified correctly, the number of incorrectly classified defect pixels and the number of background pixels that are incorrectly classified respectively. IOU calculates the spatial overlap between predictions and the ground-truth. The value of IOU is between 0 and 1 while 0 means no overlap and 1 means full overlap. Accuracy gives the probability of correctly classified pixels by our model. Recall measures the number of defective pixels in the ground truth image that are also detected as defect by the segmentation model. While precision is the measure of correct positive pixels in the ground truth that are also detected as positive by the segmentation model. F-measure gives the harmonic mean of precision and recall.

B. IMPLEMENTATION DETAILS

The Multi attention U-net and other models for comparison experiment are implemented in Python using Keras 2.2.4 library. All the models are trained for 100 epochs on GTX 1080 Graphics processing unit (GPU) with 12GB of memory. We use RMSprop optimizer with a learning rate of 0.0001 to update the weights of the model. All the parameters are initialized using Xavier uniform initialization. The main CPU parameters are given in Table.2.

C. SOLAR CELL EL IMAGES DATASET

As shown in Table.3, the solar cell EL images dataset consists of total 828 images. There are 406 crack defect images, 359 finger interruption defect images and 63 defect-free images. We only used defective images for training and algorithm

TABLE 2. CPU parameters.

PARAMETER NAME	PARAMETER VALUE
CPU	Intel ® core ™ i7-6700K CPU @ 4.00 GHZ (8CPU's)
RAM	32GB
Hard disk	2TB (Terabyte)

TABLE 3. Solar Cell EL datasets.

DATASET	CRACKS	FINGER INTERRUPTION	DEFECT-FREE	TOTAL
Training	328	292	-	620
Testing	78	67	63	208
Total	406	359	63	828

is tested on both defective and defect free images. All the images are captured in a real industrial environment. Each image has a different contrast, brightness and patterns. All the images are grayscale and are resized to 512 × 512 pixels for faster training. The images are annotated by a specialist using LabelME to provide images for training and testing the segmentation network. Each image has a binary label that means each pixel is either a defect or background: 0 or 1. We marked the defect as white; in which pixel value equals to 1, and the background as black; in which pixel value equals to 0.

Usually, data augmentation is used to enlarge the dataset and overcome the overfitting problem. There are many techniques that are applied to data augmentation for specific problems. In this paper, we use rotation = 40, flipping, height shift = 0.05, width shift = 0.05, shear range = 0.05, horizontal flip, vertical flip and adaptive histogram equalization. In adaptive histogram equalization, several different histograms are computed for each different section of the image and used to adjust lightness values of each section of

TABLE 4. Comparison of segmentation results of proposed method with state of art methods.

METHODS	FOLDS	M-IOU	PRECISION	RECALL	ACCURACY	F-MEASURE
TSAI'S METHOD	1	0.518	0.683	0.562	0.756	0.616
	2	0.462	0.664	0.547	0.744	0.599
	3	0.492	0.675	0.553	0.751	0.607
	4	0.501	0.679	0.548	0.759	0.606
	5	0.486	0.670	0.540	0.749	0.598
	Average	0.491	0.674	0.550	0.751	0.605
SEF METHOD	1	0.503	0.692	0.549	0.783	0.612
	2	0.529	0.712	0.563	0.795	0.628
	3	0.487	0.690	0.540	0.782	0.605
	4	0.512	0.701	0.531	0.785	0.604
	5	0.496	0.691	0.549	0.780	0.611
	Average	0.505	0.697	0.546	0.785	0.612
HAN'S UNET	1	0.631	0.970	0.715	0.971	0.823
	2	0.597	0.958	0.649	0.960	0.773
	3	0.611	0.968	0.659	0.967	0.748
	4	0.629	0.969	0.651	0.965	0.778
	5	0.604	0.967	0.642	0.960	0.771
	Average	0.614	0.966	0.663	0.964	0.778
OURS	1	0.728	0.996	0.710	0.993	0.829
	2	0.672	0.990	0.652	0.992	0.786
	3	0.714	0.996	0.670	0.992	0.801
	4	0.698	0.995	0.667	0.992	0.798
	5	0.685	0.990	0.645	0.990	0.781
	Average	0.699	0.993	0.668	0.991	0.799

image. AHE refines the local contrast of images and improve the quality and definitions of edges in the whole image. Data augmentation increase the generalization abilities and the accuracy of the model.

D. SEGMENTATION RESULTS

Segmentation of cracks in photovoltaic EL images is an arduous task due to the existence of heterogeneous background and various crack shapes. A fast, robust and efficient method for defects detection in polycrystalline Electroluminescence images is proposed in this paper. The proposed method leverages the advantage of multi attention network and hybrid loss. After training and evaluating the network, we obtained promising segmentation results on EL finger interruption and cracks dataset. Thus, we offer an automatic defects inspection algorithm that can segment various defects with high accuracy. It should be noted that the Multi attention U-net offers improved results with only 8.1M parameters.

In this paper, we compare the multi attention u-net with Tsai's method [9], SEF method [11] and Han's u-net [15]. We evaluate the segmentation performance of proposed method and state of the art methods using k-Fold cross validation. We set $k = 5$ for each method. K-fold validation is commonly used to determine the performance of deep learning models on small dataset. We randomly divide the finger interruption and crack defects datasets of 745 images into 5 folds. Thus, each group exhibits 149 images. It should be noted that we don't use defect-free images during testing at this stage. On each fold, four groups are used for training and the fifth group is used for testing. For each fold we calculate the mean intersection over union, precision, recall, accuracy and F-measure of proposed method and state of

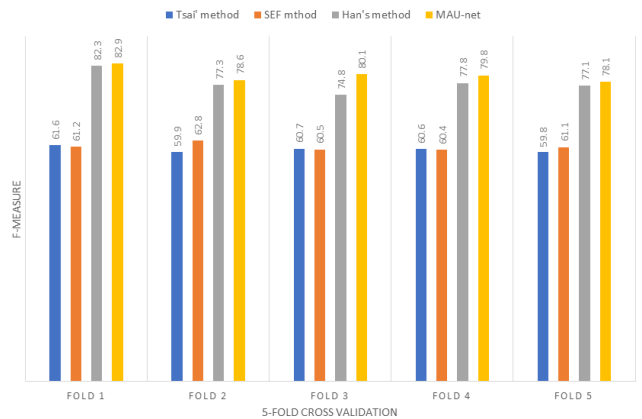


FIGURE 6. F-measure of 5-Fold cross validation of different methods.

the art methods. The results of Tsai's method, SEF method, Han's U-net and proposed method are listed it in Table.4. The last row of each method illustrates the average of 5-folds. From Table.4, it can be seen that Tsai's method has high recall as compared to SEF method. While SEF outperforms Tsai's method in terms of other metrics. Furthermore, in terms of recall proposed method has almost similar results as compared to Han's methods. From the results it can be seen that the proposed method gives the best segmentation results in each metric as compared to other four methods. The comparison of F-measure of 5-fold cross validation is illustrated in Figure.6. Furthermore, we illustrated the segmentation results of proposed method and state of the art methods on crack defects and finger interruption defects in Figure.7. The first and second column indicates the input images and ground truth images while third to sixth column represents corresponding segmentation results.

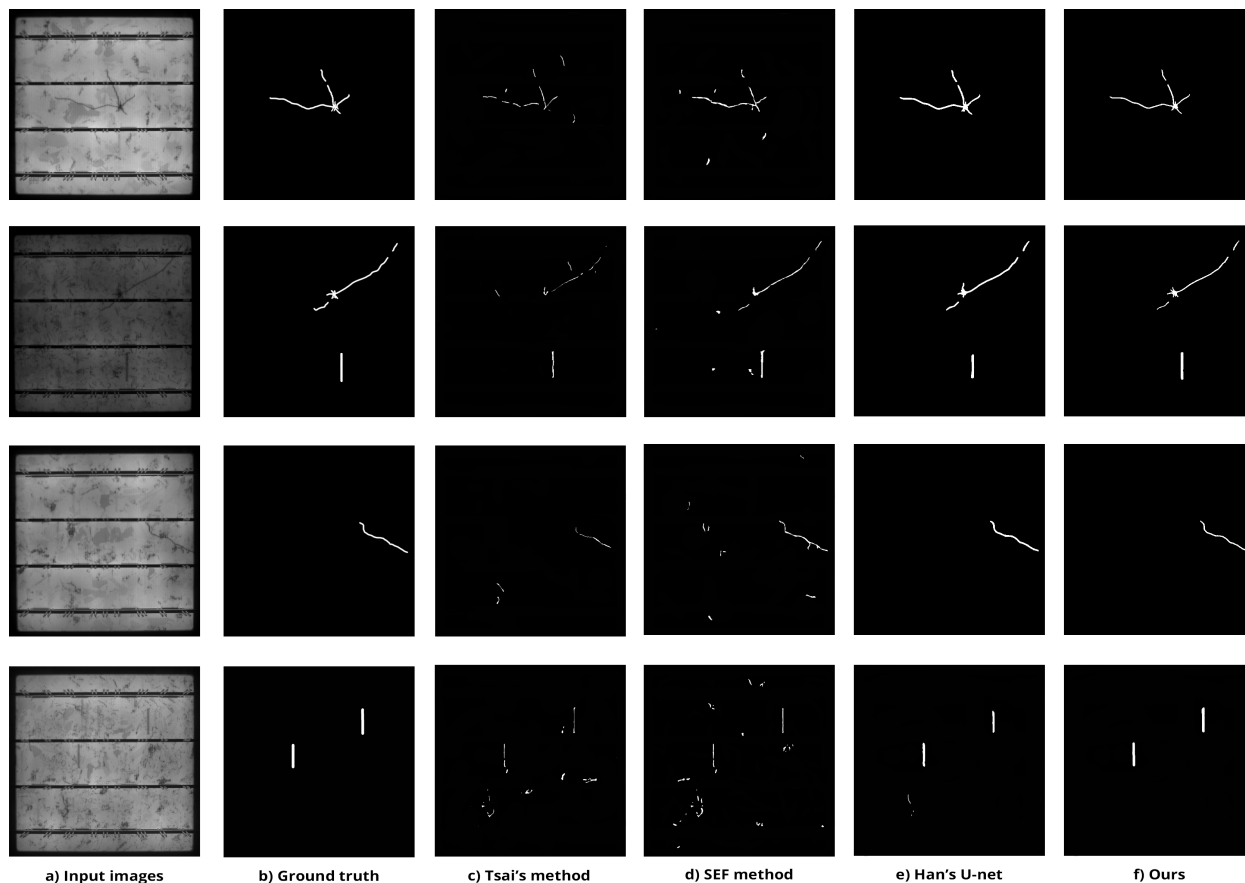


FIGURE 7. Evaluation results on defect images: a) Input image, b) Tsai's method, c) SEF method, d) Han's U-net, e) Ours.

The Tsai's method uses only defect-free images for training. Haar-like features are used to extract the information of background patterns and the fuzzy c -means is applied for clustering. The detection results vary in accordance to the number of clusters c and the control constant t . In this paper we use $c = 30$ and $t = 0.02$. The segmentation results are shown in Figure.7(c). It can be seen that the method easily locates the crack and finger interruption defects but the segmentation is inaccurate and number of background objects similar to defects are also detected as defects. The method shows poor performance when the defects are submerged into background objects.

The SEF method uses a novel steerable evidence filter to create the contrast enhanced saliency map of defects. Then local threshold and minimum spanning tree is applied for defect segmentation. We set all the parameters same as in original SEF paper. The segmentation results are shown in Figure.7(d). SEF method segments both finger interruption and crack defects better than Tsai's method. However, some defects are not segmented accurately and background objects near to defects are also considered as defects by this method.

Hui uses a region proposal network to generate images patches which might contain defects information. Then these patches are fed into modified u-net for defects segmentation.

The modification is done in a sense that the first convolution layers of 4th level and bottleneck are replaced by dilated convolutions. In this paper we use the dilation rate of 2 and 4 in 4th level and bottleneck respectively. The results are illustrated in Figure.7(e). This method segments all finger interruption and crack defects and does better segmentation as compared to the former methods. However, some background objects similar to cracks are also detected as defects.

Context information is very important when analyzing the defects from the background [33]. The proposed architecture considers the contextual information of defects and can robustly differentiate them from the heterogenous background patterns. The insertion of Global attention block between encode-decoder pass highlights the defect region enabling the network to focus more on defects, improving the m-IoU, precision and recall of defects segmentation. The results of proposed method are given in Figure.7(f). It is clear that all the defects are segmented accurately and the background crystal grain patterns are successfully neglected. Furthermore, the multi attention U-net is also prone to occlusions. For-example, at the regions where crack area mixes with the background, the proposed method robustly detects and segments all the defects.

TABLE 5. Evaluation of effectiveness of multi attention network.

METHOD	M-IOU	PRECISION	RECALL	ACCURACY	F-MEASURE
U-Net	0.609	0.952	0.651	0.961	0.773
U-Net + Channel Attention Network	0.684	0.991	0.657	0.993	0.790
U-Net + Spatial Attention Network	0.680	0.980	0.662	0.992	0.790
U-Net + Multi Attention Network	0.699	0.993	0.668	0.993	0.799

TABLE 6. Evaluation of effectiveness of hybrid loss.

Method	M-IOU	PRECISION	RECALL	ACCURACY	F-MEASURE
Multi Attention U-Net + Dice Loss	0.680	0.983	0.660	0.993	0.789
Multi Attention U-Net + Focal Loss	0.691	0.985	0.663	0.993	0.792
Multi Attention U-Net + Hybrid Loss	0.699	0.993	0.668	0.993	0.799

TABLE 7. Qualitative evaluation of defects detection results.

SAMPLE TYPES	NUMBER OF SAMPLES	DETECTION RESULTS (TSAI' METHOD)		DETECTION RESULTS (SEF METHOD)		DETECTION RESULTS (HAN'S U-NET)		DETECTION RESULTS (OUR METHOD)	
		DEFECTIVE	DEFECT-FREE	DEFECTIVE	DEFECT-FREE	DEFECTIVE	DEFECT-FREE	DEFECTIVE	DEFECT-FREE
Cracks	78	76	2	77	1	78	0	78	0
Finger interruption	67	66	1	65	2	66	1	67	0
Defect-free	63	31	32	42	21	12	51	0	63

To test the robustness of the MAU-net against background information, we concluded experiments on defect-free EL images. We use 63 defect-free images for this comparison. The segmentation results of Tsai's method, SEF method, Han's U-net and MAU-net on defect-free EL images are shown in Figure.8. Figure.8(a)-(b) shows the input images and ground truth images respectively. On the other hand, Figure.8(c)-(f) shows the corresponding results of proposed method and state of the art methods. It can be seen that the SEF method shows poor performance on defect free images. While Tsai's method segments the background objects which are similar to cracks and finger interruption defects. Han's U-net has better segmentation performance but still some crystal grain patterns are detected as defects. On the other hand, MAU-net did not segment any background object making it superior to other methods.

E. EVALUATION OF MULTI-ATTENTION NETWORK AND HYBRID LOSS

We modified the original U-net by starting the number of feature maps from 32, however, in the original U-net the features start from 64 in the first-level of U-net. From experiments, we found out that if the features of first-level start from 32, it gives the same result as of 64 features maps in the first-level of U-net. The effect of multi-attention network, channel attention and spatial attention on modified U-net is shown in Table.5. It can be seen that the introduction of channel

attention network into modified U-net increases the precision and m-IOU of the model enabling the model to focus on defects information. While the spatial attention successfully suppresses the background noise and gives the better recall rate. Furthermore, both channel attention network and spatial attention network alone increase the performance while the multi-attention network gives a further increase in segmentation performance. The final m-IOU of modified U-net with multi-attention network is 0.699 which is significantly better than former methods.

Furthermore, we experimented to test the effectiveness of hybrid loss on multi-attention U-net. Table.6 shows the effect of dice loss, focal loss and hybrid loss on the performance of the proposed method. The first, second and third rows shows the results of MAU-net with dice loss, MAU-net with focal loss and MAU-net with Hybrid loss respectively. It can be seen that the multi-attention U-net trained with hybrid loss gives better m-IOU, F-measure and accuracy.

F. DETECTION RESULTS

In order to further evaluate the detection performance of multi attention U-net, we use the scheme that any white object larger than 3×3 pixels is considered as defect. Following this scheme, we calculated the detection results of Tsai's method, SEF method, Han's U-net and MAU-net on 78 crack defects images, 67 finger interruption defect images and 63 defect-free images. All the results are listed in Table.7. It is worth

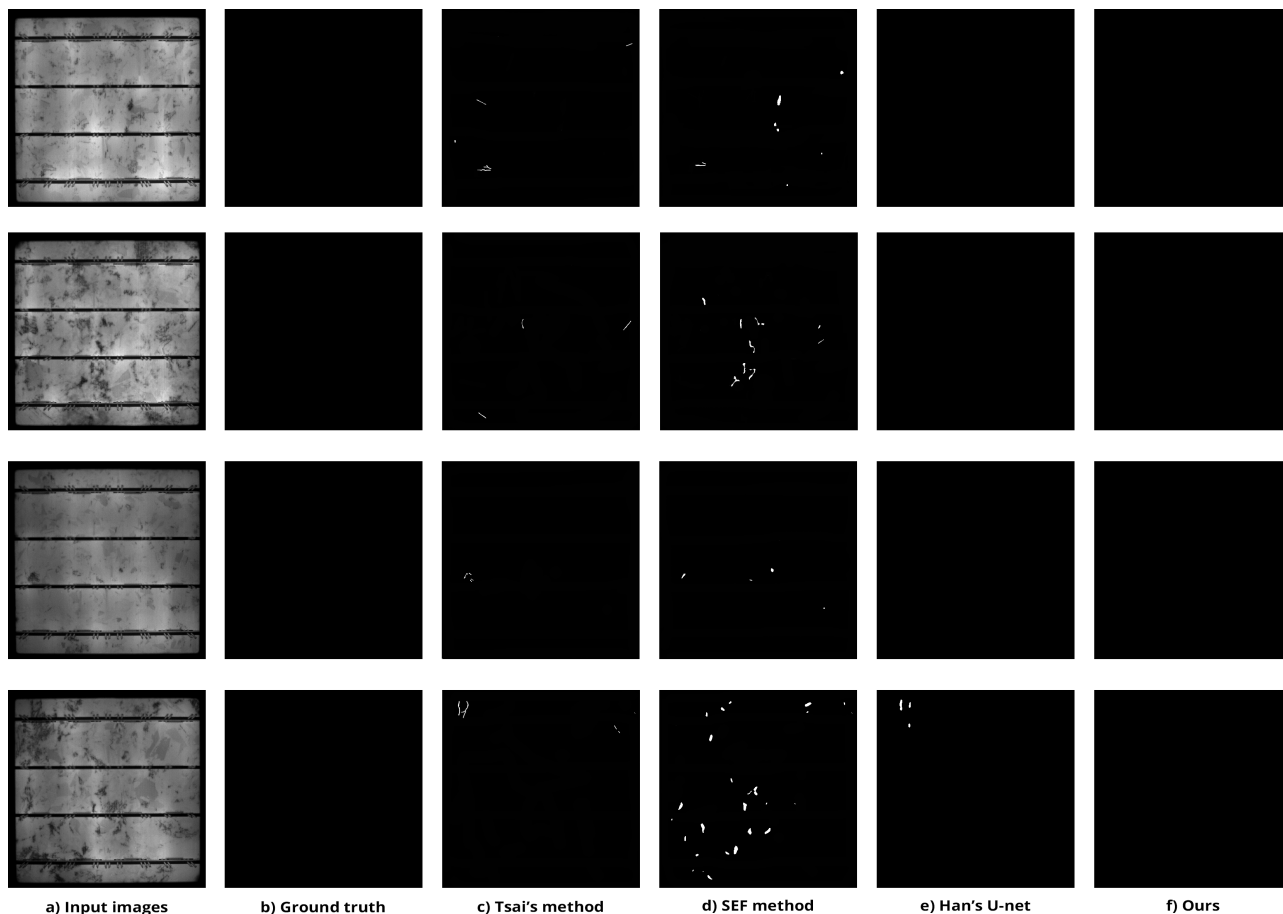


FIGURE 8. Evaluation results on defect free images: a) Input images, b) Ground truths, c) Tsai's method, d) SEF method, e) Han's U-net, f) MAU-net.

TABLE 8. Comparison of test time performances.

METHOD	TEST TIME (Milliseconds)
Tsai's method	98
SEF	250
Han's U-net	183
MAU-net	75

noticing that all the methods gave good defect detection results, however, the multi attention U-net detects all the defected images correctly. Furthermore, the Tsai's method and SEF method show poor detection performance on defect-free images while Han's U-net detects 12 defect free images as defects.

We also made a comparison between Tsai's method, SEF method Han's U-net and MAU-net in terms of test time speed. The test time speed of MAU-net and state of the art methods on 512×512 images are shown in table.8. It can be seen that Tsai's method is faster than SEF method and Han's U-net. While the MAU-net is faster than all methods and processes one image in only 75 milliseconds. Thus, the MAU-net outperforms all state-of-the-art methods in terms of test time

speed, segmentation and detection performance, enabling the network to use in real-time industrial applications.

IV. CONCLUSION

In this paper, we present a fast, robust and efficient method for defects inspection in photovoltaic Electroluminescence images. For evaluation, we used crack and finger interruption defect datasets. The proposed method has two key features: the multi attention network and the hybrid loss. The multi attention network helps the network to focus on defects while suppressing the complex heterogenous background information. Thus, enabling the network to segment complex defects robustly and with higher accuracy. The hybrid loss helps to overcome class imbalance problem and allowing the network to learn poorly classified defects. The dataset for training and testing is collected from real industrial environment. The proposed network is evaluated on defected and defect-free Photovoltaic EL images. Overall results show that the proposed method is superior to the former methods and can contribute to the establishment of more efficient and robust methods for defects inspection in polycrystalline solar cells. Further results show that the proposed method is fast and can be used for real industrial applications.

REFERENCES

- [1] E. Kabir, P. Kumar, S. Kumar, A. A. Adelodun, and K.-H. Kim, "Solar energy: Potential and future prospects," *Renew. Sustain. Energy Rev.*, vol. 82, pp. 894–900, Feb. 2018.
- [2] M. Köntges, I. Kunze, S. Kajari-Schröder, X. Breitenmoser, and B. Björneklett, "The risk of power loss in crystalline silicon based photovoltaic modules due to micro-cracks," *Sol. Energy Mater. Sol. Cells*, vol. 95, no. 4, pp. 1131–1137, Apr. 2011.
- [3] P. Rupnowski and B. Sopori, "Strength of silicon wafers: Fracture mechanics approach," *Int. J. Fract.*, vol. 155, no. 1, pp. 67–74, Mar. 2009.
- [4] B. Du, R. Yang, Y. He, F. Wang, and S. Huang, "Nondestructive inspection, testing and evaluation for si-based, thin film and multi-junction solar cells: An overview," *Renew. Sustain. Energy Rev.*, vol. 78, pp. 1117–1151, Oct. 2017.
- [5] D. B. Sulas, S. Johnston, and D. C. Jordan, "Comparison of photovoltaic module luminescence imaging techniques: Assessing the influence of lateral currents in high-efficiency device structures," *Sol. Energy Mater. Sol. Cells*, vol. 192, pp. 81–87, Apr. 2019.
- [6] A. Tsanakas and P. N. Botsaris, "An infrared thermographic approach as a hot-spot detection tool for photovoltaic modules using image histogram and line profile analysis," *Int. J. Condition Monit.*, vol. 2, no. 1, pp. 22–30, 2012.
- [7] Z. Wakaf and H. A. Jalab, "Defect detection based on extreme edge of defective region histogram," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 30, no. 1, pp. 33–40, Jan. 2018.
- [8] X. Qian, H. Zhang, C. Yang, Y. Wu, Z. He, Q.-E. Wu, and H. Zhang, "Micro-cracks detection of multicrystalline solar cell surface based on self-learning features and low-rank matrix recovery," *Sensor Rev.*, vol. 38, no. 3, pp. 360–368, Jun. 2018.
- [9] D.-M. Tsai, G.-N. Li, W.-C. Li, and W.-Y. Chiu, "Defect detection in multi-crystal solar cells using clustering with uniformity measures," *Adv. Eng. Inform.*, vol. 29, no. 3, pp. 419–430, 2015.
- [10] D.-M. Tsai, S.-C. Wu, and W.-C. Li, "Defect detection of solar cells in electroluminescence images using Fourier image reconstruction," *Sol. Energy Mater. Sol. Cells*, vol. 99, pp. 250–262, Apr. 2012.
- [11] H. Chen, H. Zhao, D. Han, and K. Liu, "Accurate and robust crack detection using steerable evidence filtering in electroluminescence images of solar cells," *Opt. Lasers Eng.*, vol. 118, pp. 22–33, Jul. 2019.
- [12] S. A. Anwar and M. Z. Abdullah, "Micro-crack detection of multicrystalline solar cells featuring an improved anisotropic diffusion filter and image segmentation technique," *EURASIP J. Image Video Process.*, vol. 2014, no. 1, pp. 1–17, Mar. 2014.
- [13] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.
- [14] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in *Proc. NIPS*, 2014, pp. 487–495.
- [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. 2018.
- [16] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.
- [17] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.
- [18] P. Anderson, X. He, C. Buehler, D. Teney, M. Johnson, S. Gould, and L. Zhang, "Bottom-up and top-down attention for image captioning and visual question answering," 2017, *arXiv:1707.07998*. [Online]. Available: <http://arxiv.org/abs/1707.07998>
- [19] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*. [Online]. Available: <http://arxiv.org/abs/1409.0473>
- [20] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3156–3164.
- [21] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, and T.-S. Chua, "SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5659–5667.
- [22] S. Xu and E. Fosler-Lussier, "Spatial and channel attention based convolutional neural networks for modeling noisy speech," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 6625–6629.
- [23] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 7132–7141.
- [24] M. Lin, Q. Chen, and S. Yan, "Network in network," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2014, pp. 1–10.
- [25] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2921–2929.
- [26] K. Smonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, 2015, pp. 1–14.
- [27] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. ICLR*, San Juan, Puerto Rico, 2016, p. 19.
- [28] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," 2018, *arXiv:1802.02611*. [Online]. Available: <http://arxiv.org/abs/1802.02611>
- [29] D. Liu, D. Zhang, Y. Song, C. Zhang, H. Huang, M. Chen, and W. Cai, "Large kernel refine fusion net for neuron membrane segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 2212–2220.
- [30] W. Zhu, Y. Huang, L. Zeng, X. Chen, Y. Liu, Z. Qian, N. Du, W. Fan, and X. Xie, "AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy," *Med. Phys.*, vol. 46, no. 2, pp. 576–589, Feb. 2018.
- [31] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. New York, NY, USA: Springer, 2017, pp. 240–248.
- [32] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.
- [33] H. Zhang, K. Dana, J. Shi, Z. Zhang, X. Wang, A. Tyagi, and A. Agrawal, "Context encoding for semantic segmentation," 2018, *arXiv:1803.08904*. [Online]. Available: <http://arxiv.org/abs/1803.08904>



MUHAMMAD RAMEEZ UR RAHMAN received the B.S. degree from the Federal Urdu University of Arts, Science and Technology, Islamabad, Pakistan, in 2016.

He is currently pursuing the M.S. degree in control science and engineering with the School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin, China. His research interests include robot vision, deep learning, and pattern recognition.



HAIYONG CHEN received the M.S. degree from the Harbin University of Science and Technology, Harbin, China, in 2005, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences (CAS), Beijing, China, in 2008.

He is currently a Professor with the School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin, China. His research interests include image processing, robot vision, and pattern recognition.

• • •