U-net based defects inspection in photovoltaic electroluminecscence images

Muhammad Rameez Ur Rahman
School of Artificial intelligence
Hebei university of technology
Tianjin, China
rameezrehman83@gmail.com

Haiyong Chen
School of Artificial intelligence
Hebei university of technology
Tianjin, China
haiyong.chen@hebut.edu.cn

Wenxi
School of Artificial intelligence
Hebei university of technology
Tianjin, China
15022327520@163.com

***Abstract*—Efficient defects segmentation from photovoltaic (PV) electroluminescence (EL) images is a crucial process due to the random inhomogeneous background and unbalanced crack non-crack pixel distribution. The automatic defect inspection of solar cells greatly influences the quality of photovoltaic cells, so it is necessary to examine defects efficiently and accurately. In this paper we propose a novel end to end deep learning-based architecture for defects segmentation. In the proposed architecture we introduce a novel global attention to extract rich context information. Further, we modified the U-net by adding dilated convolution at both encoder and decoder side with skip connections from early layers to later layers at encoder side. Then the proposed global attention is incorporated into the modified U-net. The model is trained and tested on Photovoltaic electroluminescence 512x512 images dataset and the results are recorded using mean Intersection over union (IOU). In experiments, we reported the results and made comparison between the proposed model and other state of the art methods. The mean IOU of proposed method is 0.6477 with pixel accuracy 0.9738 which is better than the state-of-the-art methods. We demonstrate that the proposed method can give effective results with smaller dataset and is computationally efficient.**

Keywords—U-net, Solar cell defects detection, cracks detection, electroluminescence images

I. INTRODUCTION

During the manufacturing of Solar cells, visual defects inspection is carried out to ensure the product quality. The solar cells can be divided into monocrystalline and polycrystalline. Monocrystalline solar cells have uniform background while the polycrystalline solar cells have randomly distributed complex textural background. Solar cells surface defects can be categorized into cracks, finger interruptions and cell breakages etc. These defects are of various sizes and shapes. The presence of small cracks in high resolution images demands a highly efficient method to segment cracks accurately. Industries often use manual defects inspection systems. Manual defect inspection methods are incapable of detecting defects efficiently and are expensive. Due to the recent developments in computer vision, manual defect inspection is replaced by automatic machine vision methods. Automatic defect inspection methods use image processing and artificial intelligence algorithms which have high speed, better accuracy and low cost. This paper proposes a powerful and effective defect segmentation method based on deep learning which can segment subtle defects in EL images.

In recent years image processing algorithms are successfully applied on photovoltaic EL images to detect defects [1]. Electroluminescence imaging method successfully detect intrinsic defects (e-g dislocations, shunts) and extrinsic defects (e-g cracks, interrupted contacts) in high resolution EL images [2]. In [3] small cracks and finger interruption defects are detected and segmented using Fourier transform. Independent component analysis reconstruction is applied to detect the defects in solar modules [4] [5]. Automatic finger interruption defects are detected by first selecting region of interest and then applying spectral clustering [6]. Solar cell Micro cracks are detected by firstly extracting features via grey level co-occurrence matrix, the support vector machine (SVM) classifier is used to classify these features [7].

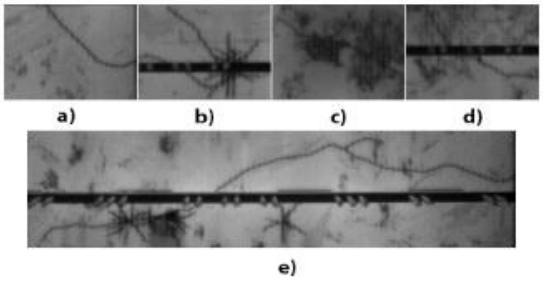

Fig. 1. Defects types: a) line defect, b) star defect, c) Defect inside same light intensity pattern d) defect divided by contact probe, e) defect with all (a) (b) (c) (d) properties

Traditional image processing algorithms require the hand-crafted features extracted by experts which directly affects the detection rate, so they don't have good applicability. Recent advances in deep convolutional neural networks (DCNN) are proved to be efficient and faster for object detection and segmentation task and they don't rely on expert's dependent hand-crafted features. Thus, applying DCNN for segmentation of photovoltaic defects is of significant benefits. However, many state of the art methods are evaluated on scene segmentation in natural images which contains high level semantic information and there is not much difference between ratio of background and object. A significant number of public datasets are available for researchers to evaluate large DCNN models. For new task, researchers use pre-trained models and fine-tune them but this strategy is not very useful for defects inspection of photovoltaic electroluminescence images. Besides this, collecting defective samples of photovoltaic EL images from industrial environment a costly task. Photovoltaic El images are grayscale images with unbalanced pixel distribution of defects and the background. Solar defects se is critical due to the presence of randomly distributed crystal grains on solar cells surface. To analyze solar defects we categorized defects into three types according to crack shapes and texture background. All cracks are of different sizes. The examples of enlarged defects are shown in fig.1. The non-defective area in image contains crystal grains which make up random patterns. These patterns are unique for each image and may have same pixel intensity value as defect which makes them the main reason for false segmentation. Line cracks, which have uniform pixel distribution are easy to

978-1-7281-4607-2/19/$31.00 ©2019 IEEE
DOI 10.1109/ICBK.2019.00036

segment than others. Some line cracks are inside the dark random crystal grain pattern which exhibit same pixel intensity value as defect that makes the defect hard to segment. Star-like cracks exhibit a complicated structure. Cracks are also divided by the dark contact probe which has same pixel intensity value as the crack and can become strong false positive. Some cracks exhibit all above mentioned characteristics. Based on above discussion a new architecture is required to overcome the shortcomings of the earlier methods for defects segmentation in photovoltaic electroluminescence images.

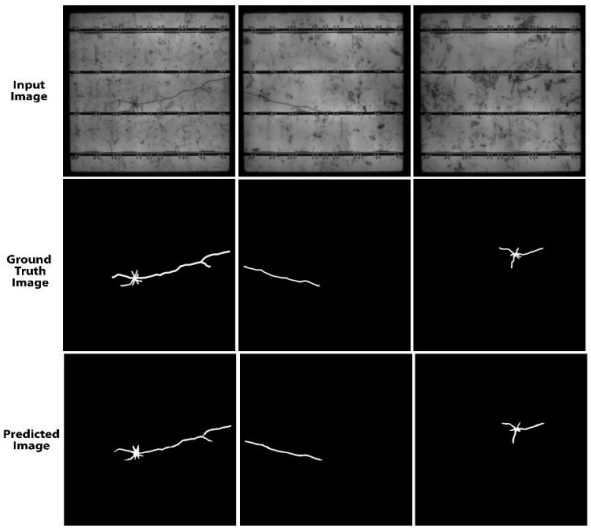

Fig. 2. Defect detection based on proposed method

II. Related work

Recent advances in deep convolutional neural networks (DCNN) are proved to be efficient and faster for the object detection and segmentation tasks and they don't rely on expert's dependent hand-crafted features. Deep belief network (DBN) with transfer learning is used to detect solar and capsule defects but it can only detect the defects on same locations [8]. In [9][10] Convolutional neural network (CNN) combined with sliding window is applied to detect cracks and other six kind of defects with 99.8% accuracy but sliding window technique makes the process slow. A three stage method based on computer vision and machine learning is applied to detect electron microscopy images [11].Firstly defects are detected using cascaded object detector and CNN based screening method is applied to remove false positives, then watershed flood algorithm [12] is applied to extract shapes and sizes of defects. Multiple Steel surface defects are classified [13] using GooLeNet [14] with identity mapping. In [15] fully convolutional network [16] and deeply supervised nets [17] are used to segment the cracks and then guided filtering is applied to refine final results. A U-shaped architecture [18] consisting of contraction and expansion path, highly dependent on data augmentation is successfully applied on various Medical images segmentation tasks [19] [20] [26]. This architecture having symmetric skip connections combines the low-level features with high-level features to improve the performance and object localization. U-shaped architecture may face one problem that it fuses different feature maps based on fixed weights. A region proposal network combined with U-net is used to detect the polycrystalline solar cell defects but the problem with this method is that the bounding boxes are of same size and one defect may contain several bounding boxes [28].

This paper proposes a deep fully convolutional neural network based architecture for segmentation based defects detection in photovoltaic EL images. We choose U-shaped fully convolutional network because it is efficient, can make dense prediction and most importantly accepts inputs of any size. U-net combine low level features with high level features based on fixed weights which may decrease its representational power. Thus, we introduce global attention network to weigh the feature maps according to their relevancy. Then the global attention network is incorporated into U-net to enhance the defect features allowing the U-net to learn defect features better. The examples of defects detection based on proposed method are shown in Fig.2. The main contributions of this paper are

- We introduce the novel attention network and incorporate it into U-net architecture to increase its feature representation power. The proposed model is evaluated on real photovoltaic El images dataset for defects.
- In proposed architecture dilated convolutions are added into U-net to enlarge the receptive field without increasing the number of parameters. Residual Skip connections with average pooling are introduced at encoder side which eases the training process and better features learning.

III. Network architecture

A. Global attention block

During features extraction, some features might be useful while others may be noise or the background object. In order to highlight the most important defect features, the global attention network is proposed. The Squeeze and excitation block in [25] is designed aiming at classification task, and it plays a key role in improving the performance of the network by placing it in each block of the network. To use it in the segmentation of defects we modified the original SE block by squeezing the inputs using global average pooling and global max pooling to get weighted high-quality features. In [29] and [30] the significance of global max pooling and global average pooling is shown for distinct objects recognition.
As shown in fig.3 Global attention block consists of two branches. The only difference between two branches is that, in one branch first global average pooling is applied while in second branch global max pooling is applied. Similar to SE block each input channel $C^h = \{c_0^h, c_1^h, \dots, c_n^h\}$ is squeezed into a single a single numeric value V^h by using global average pooling. This gives $1 \times 1 \times n$ dimensional output $V^h = \{v_0^h, v_1^h, \dots, v_n^h\}$ where n is the number of channels. Later on, V^h is fed into fully connected layer followed by a LeakyRelu activation function resulting an output $A^h = \{a_0^h, a_1^h, \dots, a_n^h\}$ of $1 \times 1 \times n/4$ dimension. The final fully connected layer followed by a sigmoid layer gives the output $A^h = \{a_0^h, a_1^h, \dots, a_n^h\}$. The same process is adopted in global max pooling branch resulting output $M^h = \{m_0^h, m_1^h, \dots, m_n^h\}$. At the end outputs of both branches are added together giving $O^h = \{o_0^h, o_1^h, \dots, o_n^h\}$. O^h Weighs the channels based on their importance and allocate high weights to the most important features.

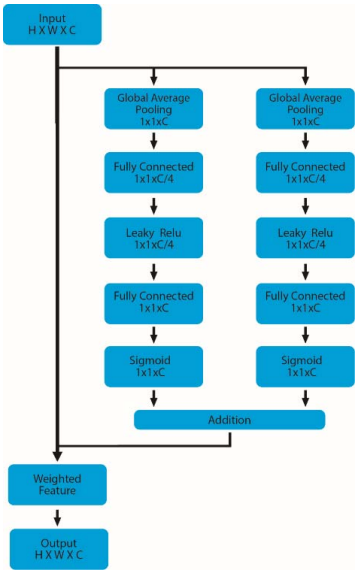

Fig. 3. Global attention block (GAB)

B. *Unet*

The original U-shaped architecture consists of down-sampling path to capture contextual information and up-sampling path for precise object localization. The down-sampling path consists of repeated two consecutive 3x3 unpadded convolution layers for features extraction which are followed by 2x2 max pooling layers for down-sampling. The number of feature channels become doubled after each step which help in extracting advanced features. The size of image decreases as the depth of network increases. The Up-sampling path consist of up-sampling of feature-maps followed by 2x2 up-convolution to recover the size of segmentation map. The up-sampled feature maps are concatenated with the feature maps of down-sampling path which are at the same level. At the end there is 1x1 convolution with 2 feature maps since there are only two classes.

C. *Proposed U-net architecture*

In this section a new architecture is proposed based on Global attention block. Fig.4 shows the architecture of proposed network. Note that the proposed architectures framework is not same as original U-net which is proposed in (Ronneberger, Fischer, and Brox 2015). We modified the original U-net and added new features which are proven to be useful in our task. The proposed architecture consist of 9-level U-net incorporating global attention block, skip connections with average pooling and dilated convolution. The Global attention block is placed at the middle connections between encoder and decoder to improve the representational power of U-net architecture. The global attention block learns to weigh the feature map channels coming from encoder path to accomplish the attention mechanism. The output of attention block is concatenated with upsampled features of decoder path which are at the same level. At encoder and decoder side we use dilation rates of 1, 1, 2, 2, 3 from top to bottom levels. Dilation [23] arbitrarily enlarges the filter's field of view so that no small defect is left undetected. At Encoder side 2x2 Maxpooling layer is used to down-sample feature maps to half. We introduced skip connections with Average pooling at encoder side to down-sample the feature maps. The use of max pooling and average pooling operations together leads to learn discriminative features which results in better detection of defects and edges. The skip connections ease the network training by providing an alternative for gradient to back propagate and help traverse information in network. Different from u-net we added Batch normalization (BN) [21] and LeakyRelu [22] activation layer after each convolution layer. The 1x1 convolution with sigmoid activation is used at the end to output the segmented image. Unlike original U-net, images are not cropped, input and output size is kept to avoid edge loss of image.

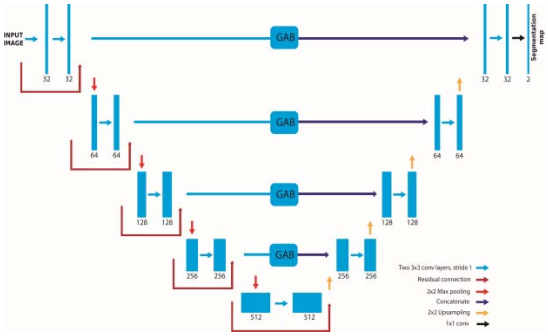

Fig. 4. Proposed U-net architecture

D. *Loss function*

During the training of network, it is necessary to estimate the weights to increase the robustness and accuracy of the network. This is achieved by using a proper loss function which will be minimized during training. As our dataset has unbalanced pixel distribution, around 90% background and 10% defect. So, in this work we use Dice loss [24] which is especially designed for unbalanced data distribution and is defined as

$$Dice\ loss = 1 - 2\,\frac{|X_g| \cap |X_p|}{|X_g| + |X_p|} \quad (1)$$

Where |Xg| is the ground truth image while |Xp| is the predicted image. The training loss and accuracy curve is shown in fig.5.

E. *Dataset*

All the experiments are carried out on 512x512 EL images. The images are taken from real industry environment. The dataset contains 400 defected images, from which 350 images are used for training and 50 images for testing the network. All the images are grayscale 512x512x1. All the training images and ground truths are carefully labelled using LabelME annotation tool. Each pixel in an image is labelled as defect or background. To overcome the problem of small dataset and benefit the training of deep learning model; either speeding up convergence or acting as regularizer, an extensive data augmentation is used. Several parameters are used to augment images which include, rotation =40, width shift= 0.05, height shift=0.05, shear range=0.05, horizontal flips, vertical flips and adaptive histogram equalization (AHE). In adaptive histogram equalization several different histograms are computed for

each different section of image and used to adjust lightness values of each section of image. AHE improves the local contrast and enhance the definitions of edges in each region of image. Data augmentation increase the generalization abilities and the accuracy of model.

IV. Experiment Results and Analysis

We evaluate the performance of the proposed network on photovoltaic electroluminescence images dataset collected from real industrial environment aiming to segment defects in images. Specifically, we compare test set results with two state of the art methods: U-net [18] and R2-Unet [26]. We use all the parameters same for comparison experiments.

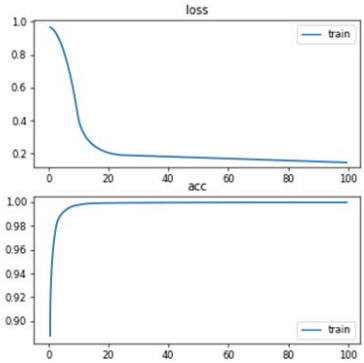

Fig. 5. Training loss and accuracy

A. Implementation *details*

The proposed network and the two baseline networks are implemented in Keras 2.2.4 framework. We use RmsProp optimizer [27] with learning rate of 0.0001 and rho= 0.9 for model optimization. The parameters are initialized using Xavier uniform initializer. We trained the model on GTX 1080 GPU. The batch size is set to 1 and the model is trained for 100 epochs. The main CPU parameters are shown in Table I.

TABLE I. CPU parameters

Parameter name	Parameter value
CPU	Intel ® core ™ i7-6700K CPU @ 4.00 GHZ (8CPU's)
RAM	32GB
Hard disk	2TB

B. *Metrics*

Many evaluation metrics are proposed and are frequently used for semantic segmentation task. To evaluate the performance of our model, predicted results are compared with ground truths using three metrics: Intersection over union (IOU), sensitivity and specificity. These metrics are specifically designed for semantic segmentation tasks.

IOU calculates the spatial overlap between predictions and ground-truth, defined between 0 and 1 while 0 represents no overlap and 1 represents complete overlap. IOU is given as

$$IOU = \frac{TP}{TP+FP+FN} \qquad (2)$$

TP, FP, FN are the true positive, false positive and false negative samples respectively.

Sensitivity is used to find the True positive rate (TPR) of the proposed network. Sensitivity is the measure of positive pixels in the ground truth that are also identified as positive by the segmentation being evaluated. It is given by formula

$$Sensitivity = \frac{TP}{TP+FN} \qquad (3)$$

Specificity is used to find the True negative rate (TNR) of the proposed network. Specificity is the measure of negative pixels in the ground truth that are also identified as negative by the segmentation being evaluated. It is given by formula

$$Sensitivity = \frac{TN}{TN+FP} \qquad (4)$$

Where TP, FP, TN, FN are the true positive, false positive, true negative and false negative respectively. Both specificity and sensitivity are reported between 0 and 1.

C. *Results and discussion*

In this study we present a deep learning-based approach that is able to efficiently segment the defects from photovoltaic electroluminescence images. The proposed method inherent the benefits of dilated convolutions, residual skip connections and Global attention mechanism. After successfully training and evaluating on the EL images, we were able to consistently achieve a good detection result. Thus, we offer a robust defects segmentation framework for photovoltaic EL images. Although the parameters of our network are only 8.3M which is ¼ of original U-net, promising results are achieved on the EL dataset. The comparison of number of parameters is shown in table II.

TABLE II. Comparison of proposed Unet with U-net and R2-Unet in terms of number of parameters

Methods	U-net	R2-Unet	Ours
Number of parameters (millions)	30.8	48.9	**8.3**

TABLE III. Metrics comparison of proposed Unet with U-net and R2-Unet

Methods	U-net	R2-Unet	Ours
Mean IOU	0.6296	0.6384	**0.6477**
Mean Sensitivity	0.8304	0.8313	0.8072
Mean Specificity	0.9787	0.9790	**0.9793**
Pixel Accuracy	0.9736	0.9737	**0.9738**

After training and testing we obtained the results of proposed network using test set consisting of 50 512x512 images. Apart from normalization we did not use any pre-processing or post-processing technique. The output result is

binary segmentation map of defects and are of same size as input. We calculated mean IOU, mean sensitivity and mean specificity to evaluate the segmentation results and presented a comparison between proposed network and state of the art methods. The results of evaluation metrics are presented in Table III. The proposed network achieved the highest Mean IOU of 0.6477, pixel accuracy 0.9738 and specificity 0.8072. In terms of sensitivity the proposed network has lower result than state of the art methods. The final prediction maps of input images and the ground truth images are illustrated in fig.6.

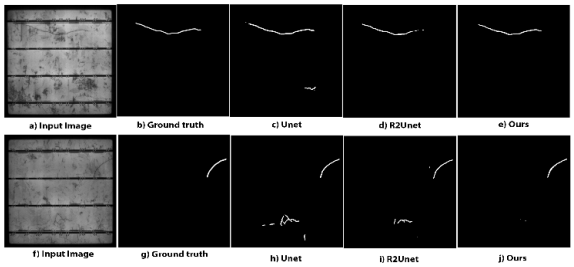

Fig. 6. Comparison of architectures (a)(f) input image, (b)(g) ground truth, (c)(h) with u-net, (d)(i) R2-Unet, (e)(j) ours

It is clear that the proposed method segments the defects accurately and have less false detection than state of the art methods. The insertion of global attention network enhances the defects features and focuses more on defects while improving the precision of cracks detection. Insertion of dilation plays a key role to extract large semantic information. Original U-net detects many background objects as defects resulting in false segmentation. R2-Unet detects defects better than original U-net but it also considers background textures as defects.

V. CONCLUSION

In this paper a novel global attention block is proposed and added into U-net architecture for defects segmentation in EL images taken from real industry environment. The proposed architecture consists of dilated convolutions with residual skip connections containing average pooling aiming to learn discriminative features and achieving meaningful results. The improved U-net model has 8.3M parameters. The only drawback of the proposed network is its less sensitivity than state of the art methods. This maybe be because we did not use any pre-processing or post-processing techniques. We will consider this drawback in our future work and increase the performance of the network with improved sensitivity. The proposed model is compared with state-of-the-art models, i-e U-net and R2-Unet and overall results show that better results can be achieved by adopting the proposed method.

REFERENCES

[1] A. Mansouri, M. Zettl, O. Mayer, M. Lynass, M. Bucher, and O. Stern, "Defect detection in photovoltaic modules using electroluminescence imaging," 27th Eur. Photovolt. Sol. Energy Conf. Exhib., vol. 64617926, pp. 3374–3378, 2012.

[2] Ebner, R., B. Kubicek, and G. Ujvari, Non-destructive techniques for quality control of PV modules: Infrared thermography, electro- and photoluminescence imaging. 2013. 8104-8109.

[3] Tsai, D.-M., S.-C. Wu, and W.-C. Li, Defect detection of solar cells in electroluminescence images using Fourier image reconstruction. Solar Energy Materials and Solar Cells, 2012. 99: p. 250-262.

[4] Tsai, D., S. Wu, and W. Chiu, Defect Detection in Solar Modules Using ICA Basis Images. IEEE Transactions on Industrial Informatics, 2013. 9(1): p. 122-131

[5] Zhang, X., et al., A Novel Method for Surface Defect Detection of Photovoltaic Module Based on Independent Component Analysis %J Mathematical Problems in Engineering. 2013. 2013: p. 8.

[6] Tseng, D.-C., Y.-S. Liu, and C.-M. Chou, Automatic Finger Interruption Detection in Electroluminescence Images of Multicrystalline Solar Cells %J Mathematical Problems in Engineering. 2015. 2015: p. 12.

[7] Teow Wee Teo and Mohd Zaid Abdullah, "Solar Cell Micro-Crack Detection Using Localised Texture Analysis," Journal of Image and Graphics, Vol. 6, No. 1, pp. 54-58, June 2018

[8] Ri-Xian, L., Y. Ming-Hai, and W. Xian-Bao, Defects detection based on deep learning and transfer learning. Metall Min Ind.Vol. 7. 312-321. 2015

[9] Y.J. Cha, W. Choi, O. Büyüköztürk, Deep learning-based crack damage detection using convolutional neural networks, Comput. Aided Civ. Inf. Eng. 32 (5) (2017)

[10] T. Wang, Y. Chen, M. Qiao, and H. Snoussi, ''A fast and robust convolutional neural network-based defect detection model in product quality control,'' Int. J. Adv. Manuf. Technol., vol. 94, pp. 3465–3471, Feb. 2018

[11] Li, W., K.G. Field, and D. Morgan, Automated defect analysis in electron microscopic images. npj Computational Materials, 2018. 4(1): p. 36.

[12] Vincent, L. and P. Soille, Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991. 13(6): p. 583-598.

[13] Liu, Y., Geng, J., Su, Z., Zhang, W., Li, J.: Real-time classification of steel strip surface defects based on deep CNNs. In: Lecture Notes in Electrical Engineering, pp. 257–266. No. 529 (2019)

[14] Szegedy, C., et al. Going deeper with convolutions. in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015.

[15] Zou, Q., et al., DeepCrack: Learning Hierarchical Convolutional Features for Crack Detection. IEEE Transactions on Image Processing, 2019. 28(3): p. 1498-1512.

[16] Long, J., E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015.

[17] Lee, C.-Y., et al., Deeply-Supervised Nets, in Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, L. Guy and S.V.N. Vishwanathan, Editors. 2015, PMLR %J Proceedings of Machine Learning Research: Proceedings of Machine Learning Research. p. 562--570.

[18] Ronneberger, O., P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015. Cham: Springer International Publishing.

[19] Lozej, J., et al., End-to-End Iris Segmentation Using U-Net. 2018. 1-6.

[20] Liu, Z., et al., Liver CT sequence segmentation based with improved U-Net and graph cut. Expert Systems with Applications, 2019. 126: p. 54-63.

[21] Ioffe, S. and C. Szegedy, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, in Proceedings of the 32nd International Conference on Machine Learning, B. Francis and B. David, Editors. 2015, PMLR %J Proceedings of Machine Learning Research: Proceedings of Machine Learning Research. p. 448--456.

[22] A. L. Maas, A. Y. Hannun, and A. Y. Ng. Rectifier nonlinearities improve neural network acoustic models. In International Conference on Machine Learning (ICML).2013

[23] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in ICLR, 2016.

[24] Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J.: 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. pp. 240–248 (2017)

[25] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. arXiv preprint arXiv:1709.01507, 2017.

[26] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, ‘‘Recurrent residual convolutional neural network based on u-net (R2U-Net) for medical image segmentation,’’ Feb. 2018, arXiv:1802.06955

[27] Tieleman, T., Hinton, G.: Lecture 6.5-rmsprop: divide the gradient by a running average of its recent magnitude. In: COURSERA: Neural Networks for Machine Learning (2012)

[28] Han, H., et al., Polycrystalline silicon wafer defect segmentation based on deep convolutional neural networks. Pattern Recognition Letters, 2018

[29] M. Lin, Q. Chen, S. Yan, “Network in network,” in International Conference Learning Representations. (ICLR), 2014, pp. 1–10.

[30] [20] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, “Learning deep features for discriminative localization,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 2921–2929.