ASIA-PACIFIC SOLAR RESEARCH CONFERENCE

PEER REVIEWED
ASIA-PACIFIC
SOLAR RESEARCH CONFERENCE

A Review of Deep Learning for Defects Classification in Silicon Cells and Modules from Luminescence Images

Yoann Buratti,[1] Zubair Abdullah-Vetter,[1] Gaia Maria Javier,[1] Priya Dwivedi,[1] Thorsten Trupke,[1] and Ziv Hameiri[1]

[1] *University of New South Wales, Sydney, Australia*

The silicon (Si) photovoltaic (PV) industry currently cumulates to almost 800 GWp of installed module capacity.[1] To ensure the best possible quality and reliability in this continuously growing industry, efficient and automated methods for monitoring the presence of defects in cells and modules are required. Luminescence imaging, either electroluminescence (EL)[2] or photoluminescence (PL),[3] is a powerful characterisation technique that when combined with machine learning (ML), can meet the challenges of the growing PV industry. **In this paper, we provide a detailed review of recent applications of deep learning algorithms for the classification of defects in Si cells and modules using luminescence images.**

Image analysis using deep learning is based on convolutional neural networks (CNN).[4] Compared to traditional ML, CNNs excel at image analysis due to their ability to learn spatial structure, using convolutional and pooling layers. A competition of image classification, ImageNet,[5] has encouraged researchers to develop different architectures (varying type and number of layers and how to arrange them): AlexNet,[6] ResNet,[7] SqueezeNet,[8] and VGGNet[9] among others.

These architectures and their trained weights are publicly available and can be used as-is (transfer learning), further refined (fine-tuning), or trained from scratch for different classification tasks. All CNNs are composed of two blocks: a feature extractor and a classifier. The feature extractor consists of convolutional and pooling layers while the classifier contains fully connected layers. Removing the classification block enables the CNN to be used to extract a feature vector from the images, which can be passed to more traditional ML algorithms such as random forest (RF)[10] or support vector machine (SVM).[11] Finally, state-of-the-art generative CNN, named generative adversarial network (GAN),[12] can be used as data augmentation for classification problems.

To evaluate a classification task, a confusion matrix is usually created, comparing actual and predicted class labels. For binary classification, the matrix is a 2×2 table comprising of true positive (TP), true negative (TN), false negative (FN), and false positive (FP). Two scoring metrics are defined for this review, the accuracy and F_1-score:[13]

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{F}_1\text{-score} = \frac{\text{TP}}{\text{TP} + 0.5 \cdot (\text{FP} + \text{FN})} \tag{2}$$

The accuracy measures the ratio between correctly predicted class labels and the total number of instances in the dataset. The F_1-score is the harmonic average of the precision (how good is the classifier at finding instances of the class) and recall (did the classifier find all instances of the class). When possible, the F_1-scores are reported in this review for comparison as they inherently consider class imbalance in the dataset, contrary to the accuracy.[14] The definitions are extended to multi-class classification tasks by averaging each class metric.

A dataset containing 2,654 multi-crystalline Si (mc-Si) and mono-crystalline Si (mono-Si) EL images, extracted from 44 modules, was published in 2018 by Buerhop-Lutz *et al*.[15] The dataset, referred to as the ELPV dataset, is available on Github.[15] Each image in the dataset is 300×300 pixels and is annotated as Functional or Defective. The ELPV dataset was first used by Deitsch *et al.*,[16] to apply a CNN (VGG) as a feature extractor with an SVM for the classification task. They achieved an F_1-score of 88.4% and compared their results to a non-deep learning approach involving feature engineering and an SVM classifier achieving an F_1-score of 82.4%. Their study was the first to demonstrate the superiority of automated feature extraction through CNN compared to handcrafted features. Several groups have used the ELPV dataset for the comparison of ML models. Akram *et al.*,[17] introduce data augmentation techniques (random horizontal, vertical flipping, contrast operations) to increase the size of the ELPV dataset. Their architecture was based on VGG with a reduced number of layers and achieved a 92.2% F_1-score. The data augmentation was shown to increase the accuracy by 6.5%.[17] Acharya *et al.*,[18] relabelled the dataset into four classes ('no defect', 'micro defect', 'large defect', and 'low resolution defect'). They then trained a custom CNN architecture achieving 73.8% F_1-score, better than the results achieved by a VGG16 classifier (65.0% F_1-score). Their custom CNN was based on a Siamese CNN where the left and right side of the EL image was passed to separate twin pipelines of the CNN. Finally, Demirci *et al.*,[19] returned to the CNN feature extractor approach, combining the feature of a custom CNN architecture and VGG fed into an SVM classifier for an F_1-score of 95.5% on the binary classification in the ELPV dataset. A summary of the results on the ELPV dataset is shown in Table I. In a span of three years, deep learning-based algorithms are improved at the defect classification task with a learning curve from 88.4% to 95.5% F1-score. These values are higher compared to the performance of the SVM classifier on custom features forr the same task with a 82.4% F_1-score.

Table I. F1-score results on the ELPV dataset for classification tasks

Ref	Year	Model	Classes	F_1-score
Deitsch *et al.*[16]	2019	Custom features + SVM CNN(VGG) + SVM	functional; defective	82.4% 88.4%
Akram *et al.*[17]	2019	CNN(VGG)	functional; defective	92.2%
Acharya *et al.*[18]	2020	CNN(Custom) CNN(VGG)	no defect; micro defect; large defect; low resolution defect	73.8% 65.0%
Demirci *et al.*[19]	2021	CNN(VGG) + SVM CNN(Custom) + SVM CNN(Combined) + SVM	functional; defective	91.9% 93.6% 95.5%

In a different study by the ELPV publishing group, Bartler *et al.*[20] extended the dataset to 1,366 modules resulting in an unbalanced dataset of 98,280 cells with only 3.4% defective cells. As a result, data augmentation techniques (random horizontal and vertical flipping) and resampling have been used to balance the dataset. A CNN based on the VGG network was trained to classify the cells, achieving an F_1-scores of 93.0%. However, the extended dataset was not published by the authors. Larger datasets (not published) have encouraged the training of more complex deep learning algorithms. For example, Ying *et al.*,[21] used 13,835 mc-Si cell EL images to train a custom CNN architecture as a feature extractor and used an RF to classify into six classes: 'no defects', 'open welding', 'broken grid', 'solid black', 'shadow', and 'hidden crack' with an F_1-score of 93.0%. The authors trained a multi-channel CNN where the EL image was resized for each of the channels capturing simple to more complex features in each channel. Karimi *et al.*,[22] applied a CNN technique for defect classification after different times in damp heat exposure, introducing cracks and corroded defects. The authors showed that their custom CNN achieved a 96% accuracy and compared it to a direct classification with RF (96%), and SVM (82%). The F_1-scores were not reported, although as mentioned, F_1-score is more meaningful to unbalanced datasets than accuracy. To increase dataset sizes, it was mentioned that geometric data augmentation could be used.[17] However, GANs can also be trained to generate a diversity of samples from a few examples of a defect class. With GANs, Luo *et al.*,[23] augmented a dataset of 507 defective mono-Si EL images into 40,000 generated images, for each of the four classes ('Grid fingers, 'Micro-crack', 'Material defect', and 'Deep crack'). Multiple CNN architectures were then trained for the classification task and achieved accuracies of 98.42% (ResNet), 74.73% (AlexNet), and 60.6% (SqueezeNet). Even if the F_1-scores were not reported, they should be close to the accuracies due to the balanced nature of a GAN-generated dataset. However, the authors mentioned that due to the unbalanced nature of the original datasets toward the "Grid fingers" class, the trained VGG network achieved lower accuracy for the other classes. This can be explained by the poorer diversity of defect examples generate by the GAN for classes with fewer examples. Tang *et al.*,[24] also used GAN for data augmentation and trained VGG to classify four classes with 82% accuracy. The results for the non-ELPV datasets are summarised in Table II.

Table III. F1-score and accuracy results on non-published dataset for classification tasks

Ref	Year	Model	Datasets	Classes	Performance
Bartler *et al.*[20]	2018	CNN(VGG)	ELPV (extended)	functional; defective	F_1-score: 92.2%
Ying *et al.*[21]	2018	CNN(custom) + RF	13,835 mc-Si cells from module EL	no defect; open welding; broken grid; solid black; shadow; hidden crack	F_1-score: 93.0%
Karimi *et al.*[22]	2019	CNN(custom) RF SVM	3550 mc-Si and mono-Si cells EL	good; cracked; corroded	F_1-score: 98% F_1-score: 96% F_1-score: 82%
Luo *et al.*[23]	2019	CNN(SqueezeNet) CNN(AlexNet) CNN(ResNet)	507 defective monoSi EL into 40,000 GAN generated images	grid fingers; material defect; microcrack; deep cracks	Accuracy: 60.6% Accuracy: 74.7% Accuracy: 98.4%
Tang *et al.*[24]	2020	CNN(VGG) CNN(Custom)	ELPV+Jinko 1800 EL images with GAN augmentation: 7400	defect-free; micro-crack; finger-interruption; break	Accuracy: 82% Accuracy: 82.5%

To summarise, the ELPV dataset provides a clear way to benchmark new approaches or deep learning algorithms as they improve year on year reaching a 95.5% F1-score with deep learning CNNs. Other datasets showed improvements in both F1-score and accuracies for multi-classification tasks throughout the years, with the implementation of data augmentation strategies using GANs. With all these said, here are several recommendations for current and future researchers in this field:

(1) Access to publicly available labelled datasets is a key requirement for benchmarking new approaches and driving the development of novel methods in PV defect classification from luminescence images. This is needed especially for benchmarking the development of custom architecture which has unproven track records compared to published CNN architectures (ResNet, AlexNet, VGG, etc). While the ELPV dataset provides a baseline, larger datasets with multiple labelled classes are needed. We therefore strongly encourage authors to publish their datasets. To circumvent intellectual property restrictions, authors can also publish their trained models. The shared models can be used to benchmark the dataset used in novel approaches in the absence of shared datasets.

(2) Visually inspecting each cell or module for manual labelling is time-consuming and prone to human error or bias, especially for very large datasets. To keep increasing the dataset's size used for deep learning training, automated labelling methods are required, such as unsupervised or semi-supervised learning, to address the challenge of dataset labelling.

(3) A promising direction is the use of GAN to solve the unbalanced dataset issue and to generate large amounts of data for under-represented classes. GANs also solve the issue of manual labelling. With the continued development of deep learning techniques for defect classification, luminescence imaging truly becomes an essential aspect of solar cell and module characterisation.

Acknowledgements

This work was supported by the Australian Government through the Australian Renewable Energy Agency [ARENA; Projects 2017/RND017 and 2020/RND016]. The views expressed herein are not necessarily the views of the Australian Government, and the Australian Government does not accept responsibility for any information or advice contained herein.

References

1. M. Fisher, M. Woodhouse, S. Herritsch, and J. Trube, “International Technology Roadmap for Photovoltaics (12th edition),” VDMA e. V., 2021.
2. T. Fuyuki, H. Kondo, T. Yamazaki, Y. Takahashi, and Y. Uraoka, “Photographic surveying of minority carrier diffusion length in polycrystalline silicon solar cells by electroluminescence,” *Appl. Phys. Lett.*, vol. 86, no. 26, p. 262108, 2005.
3. T. Trupke, R. A. Bardos, M. C. Schubert, and W. Warta, “Photoluminescence imaging of silicon wafers,” *Appl. Phys. Lett.*, vol. 89, no. 4, p. 044107, 2006.
4. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
5. O. Russakovsky *et al.*, “ImageNet large scale visual recognition challenge,” *Int J Comput Vis*, vol. 115, no. 3, pp. 211–252, 2015.
6. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems 25*, pp. 1097–1105, 2012.
7. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
8. F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size,” *arXiv:1602.07360 [cs]*, 2016.
9. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv:1409.1556 [cs]*, 2014.
10. L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
11. H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, “Support vector regression machines,” in *Advances in neural information processing systems*, pp. 155–161, 1997.
12. I. J. Goodfellow *et al.*, “Generative adversarial networks,” *arXiv:1406.2661 [cs, stat]*, 2014.
13. T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.
14. S. W. Glunz, S. Rein, W. Warta, J. Knobloch, and W. Wettling, “Degradation of carrier lifetime in Cz silicon solar cells,” *Solar energy materials and solar cells*, vol. 65, no. 1–4, pp. 219–229, 2001.
15. C. Buerhop-Lutz *et al.*, “A benchmark for visual identification of defective solar cells in electroluminescence imagery,” *35th European Photovoltaic Solar Energy Conference and Exhibition*, pp. 1287–1289, 2018.
16. S. Deitsch *et al.*, “Automatic classification of defective photovoltaic module cells in electroluminescence images,” *Solar Energy*, vol. 185, pp. 455–468, 2019.
17. M. W. Akram *et al.*, “CNN based automatic detection of photovoltaic cell defects in electroluminescence images,” *Energy*, vol. 189, p. 116319, 2019.
18. A. K. Acharya, P. K. Sahu, and S. R. Jena, “Deep neural network based approach for detection of defective solar cell,” *Materials Today: Proceedings*, vol. 39, pp. 2009–2014, 2020.
19. M. Y. Demirci, N. Beşli, and A. Gümüşçü, “Efficient deep feature extraction and classification for identifying defective photovoltaic module cells in electroluminescence images,” *Expert Systems with Applications*, vol. 175, p. 114810, 2021.
20. A. Bartler, L. Mauch, B. Yang, M. Reuter, and L. Stoicescu, “Automated detection of solar cell defects with deep learning,” in *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 2035–2039, 2018.
21. Z. Ying, M. Li, W. Tong, and C. Haiyong, “Automatic detection of photovoltaic module cells using multi-channel convolutional neural network,” in *2018 Chinese Automation Congress (CAC)*, pp. 3571–3576, 2018.
22. A. M. Karimi *et al.*, “Automated pipeline for photovoltaic module electroluminescence image processing and degradation feature classification,” *IEEE Journal of Photovoltaics*, vol. 9, no. 5, pp. 1324–1335, 2019.
23. Z. Luo, S. Y. Cheng, and Q. Y. Zheng, “GAN-based augmentation for improving CNN performance of classification of defective photovoltaic module cells in electroluminescence images,” *IOP Conference Series: Earth and Environmental Science*, vol. 354, p. 012106, 2019.
24. W. Tang, Q. Yang, K. Xiong, and W. Yan, “Deep learning based automatic defect identification of photovoltaic module using electroluminescence images,” *Solar Energy*, vol. 201, pp. 453–460, 2020.