A benchmark dataset for defect detection and classification in electroluminescence images of PV modules using semantic segmentation

Lawrence Pratt [a,b,*], Jana Mattheus [a], Richard Klein [b]

[a] *CSIR, Meiring Naude Road, Pretoria, South Africa*
[b] *School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa*

ARTICLE INFO

ABSTRACT

Electroluminescence (EL) images enable defect detection in solar photovoltaic (PV) modules that are otherwise invisible to the naked eye, much the same way an x-ray enables a doctor to detect cracks and fractures in bones. This paper presents a benchmark dataset and results for automatic detection and classification using deep learning models trained on 24 defects and features in EL images of crystalline silicon solar cells. The dataset consists of 593 cell images with ground truth masks corresponding to the pixel-level labels for each feature and defect. Four deep learning models (U-Net_12, U-Net_25, PSPNet, and DeepLabv3+) were trained using equal class weights, inverse class weights, and custom class weights for a total of twelve sets of predictions for each of 50 test images. The model performance was quantified based on the median intersection over union (mIoU) and median recall (mRcl) for a subset of the most common defects (cracks, inactive areas, and gridline defects) and features (ribbon interconnects and cell spacing) in the dataset. The mIoU measured higher for the two features compared to the three defects across all models which correlates with the size of the large features compared to the small defects that each class occupies in the images. The DeepLabv3+ with custom class weights scores the highest in terms of mIoU for the selected defects in this dataset. While the mIoU for cracks is low (25%) even for the DeepLabv3+, the recall is high (86%), and the resulting prediction masks reliably locate the defects in complex images with both large and small objects. Therefore, the model proves useful in the context of detecting cracks and other defects in EL images. The unique contributions from this work include the benchmark dataset with corresponding ground truth masks for multi-class semantic segmentation in EL images of solar PV cells and the performance metrics from four semantic segmentation models trained using three sets of class weights.

1. Introduction

Electroluminescence (EL) images enable defect detection in solar photovoltaic (PV) modules that are otherwise invisible to the naked eye, much the same way an x-ray enables a doctor to detect cracks and fractures in bones. Millions of EL images are taken every day in factories, labs, and PV plants across the globe. The EL images are essential to identify cracks and other defects in the finished product before the module is sent to market. EL images are also regularly captured during quality testing and troubleshooting over the 25-year typical lifetime of a PV module.

The purpose of this work is to present a unique dataset and evaluate several deep learning models for semantic segmentation applied to the detection of defects and features in EL images of solar cells. Two U-Nets, a PSPNet, and a DeepLabv3+ model were trained to detect 12 features

and 12 defects simultaneously. The objective of the analysis was to identify the most promising model suited to the task which will then be optimized in future work. The unique contributions from this work include the benchmark dataset with corresponding ground truth masks for multi-class defect detection in EL images of solar PV cells. This new dataset provides the first detailed pixel-level labels for EL images of solar cells.

2. Background

Automatic defect detection and classification in solar cells is the subject of many publications since EL imaging of silicon solar cells was first introduced by Fuyuki et al. [1] for detection of deteriorated areas in solar cells in 2005. Since that time, progress has been made towards *in-situ* EL imaging [2–5], improving image quality [6,7], imaging under

---

challenging conditions [8], and correlating defects to power output of PV modules [9–16]. A review of EL camera technology and field applications was published in Ref. [17].

Automated analysis and defect detection of PV module level EL images are critical to derive useful information from batches of PV modules bought and sold throughout the PV value chain. Techniques have been developed to extract and enhance images of solar cells from the PV module-level images [18–21] as a pre-processing step to automate the defect detections and classification typically conducted on cell-level images. Deep learning models have been trained on EL images to classify cells and modules into defect types or varying levels of defect severity [19,20,22–28]. Methods for object detection and localization of multiple defects in EL images have been presented [29–31]. Binary segmentation methods have been used to detect and localize cracks and gridline defects at the pixel level [32–36]. Fioresi et al. [36] presented results using semantic segmentation to detect and localize three classes simultaneously on EL images from a private dataset with 14 labelled classes. However, the labels for cracks and gridline defects in Ref. [36] were assigned to large areas of the cell image, as discussed in Section 3.2.

This paper builds on previous work in which the results of a U-Net model trained on 24 defects and features were presented [37]. In that work, semantic segmentation was introduced as a promising approach for detection and quantification of multiple defect classes in EL images of PV modules. The key contribution from semantic segmentation lies in the pixel-level classification of images [38,39] which enabled the translation from an unstructured image to a structured dataset. Pixel-level classification enabled the model to detect and quantify multiple defect types on a single solar cell rather than simply classify each cell image as having a specific defect or not. The structured dataset was used to quantify the extent of each defect at the module level by summing up the number of pixels associated with each defect observed in the cell-level EL images taken at multiple stages during a sequence of accelerated stress tests. The electrical performance of the PV module can then be correlated to the defects detected in the EL images to better understand the root cause of module performance and degradation over time. In this current work, the labelled dataset is made public and the results from twelve deep learning models are compared and summarized to identify models that might be better suited for defect detection in EL images of PV modules. The comparison includes the same U-Net model presented in Ref. [37] (see Table 1, model '1a') with results from training on the updated dataset.

3. Materials and methods

3.1. Electroluminescence images

EL images of PV modules constructed from crystalline silicon cells are essential for defect detection because micro-cracks and inactive areas impact module performance but often escape detection by the naked eye. An EL image is captured with a camera optimized to record photon emissions near 1100 nm wavelengths, outside the visible light spectrum. The PV module is connected to a power supply and an electrical current is applied to excite the photons during imaging. The camera captures the photon emissions when the electron-hole pairs recombine [17].

Fig. 1 shows ten examples of EL images from the dataset. The top row consists of cells from multi-crystalline silicon wafers with the characteristic square corners and grain boundaries that appear as small, dark regions throughout the cell. The bottom row consists of cells from mono-crystalline silicon wafers with the characteristic rounded corners and noticeable lack of grain boundaries due to the single crystal orientation of the wafer. The images also show examples of some features like cell spacing, ribbon interconnects and text. The defects shown include cracks, inactive areas, gridline defects, and corrosion.

3.2. Dataset

The dataset consists of 593 EL images of solar cells originating from three private and two public sources [40,41] with a roughly equal number of multi-crystalline and mono-crystalline wafers. The original number of labelled images exceeded 600, but some images were excluded following a quality control review by the PV expert. The contributions from the private sources consisted of module-level EL images while the contributions from the public sources consisted of cell-level EL images. The original images were pre-processed following a method described in Ref. [37]. The final 512×512 image contains a full cell in the center surrounded by adjacent cells, the module edge, or padding depending on the source of the image and location of the cell within the module-level image. The two public sources published single-cell images and therefore required padding on all four sides to maintain the full cell at the center of each image.

The images in the benchmark dataset were curated by a PV expert from the $80{,}000 +$ images available from the five data sources combined. The images were chosen by the expert during a visual scan of randomly selected samples from the 80,000 images to form a representative sample of mono- and multi-crystalline cells with common features and defects so that the model would be as generalized as possible given a relatively small sample. The PV expert identified 12 features intrinsic to most PV modules to provide context for the semantic segmentation models. In this paper, a feature refers to a specific component of a PV module such as a busbar, ribbon interconnect, or cell spacing. This should not be confused with a feature designed by a data scientist to improve predictions. The PV expert also identified 12 defects extrinsic to solar cells such as cracks, inactive areas, and gridline defects that can negatively impact module performance. Collectively, the features and defects combine to create 24 classes for the pixel level classification.

A ground truth mask was then created for each image in the dataset using the GNU Image Manipulation Program (GIMP) [42] to assign a unique color code to each class. The selected samples were distributed to a team of labelers who were trained to identify the features and defects.

Table 1
Model ID and descriptions.

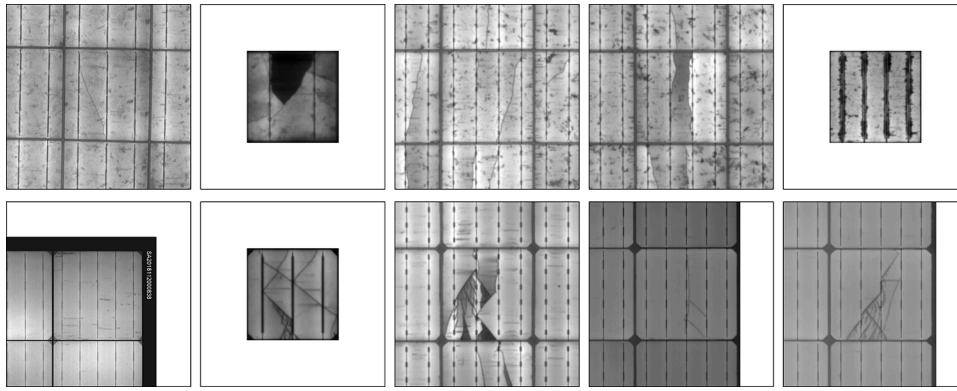| ID	Class Weights	Architecture	Github Authors	Trainable Parameters
1-a	Equal	U-Net	Divam	12,333,720
1-b	Inverse			
1-c	Custom			
2-a	Equal	U-Net	Tomar	25,858,887
2-b	Inverse			
2-c	Custom			
3-a	Equal	PSPNet	Kamikawa	58,038,784
3-b	Inverse			
3-c	Custom			
4-a	Equal	DeepLabv3+	Yakubovskiy and Kawakita	22,443,368
4-b	Inverse			
4-c	Custom			

Fig. 1. EL images from multi-crystalline silicon wafers (top) and mono-crystalline silicon wafers (bottom).

The labelers were also assigned a common set of images so that the process could be assessed for consistency and further training requirements in the early stages. The labelled images were then analyzed to identify and correct any pixels not consistent with the pre-determined RGB color codes. Finally, the PV expert reviewed the ground truth masks and provided instructions to one expert labeler for final revisions.

Fig. 2 shows the EL images and corresponding ground truth masks for the multi-crystalline silicon cells shown in Fig. 1. The labeling of ground truth masks presented challenges when labelling the long, narrow defects such as cracks and gridline defects. A slight misalignment between the image and the ground truth mask can easily lead to false positives and false negatives when the object of interest is only a few pixels wide such as the case for cracks. In Ref. [36], the ground truth masks assign large areas of the image to cracks and gridline defects that extend far beyond the specific pixels associated with either defect as seen in the EL image. While the labelling approach in Ref. [36] minimizes the impact from errors along the edges, the localization of defects is less precise. The dataset presented in this work provides a finer resolution for the size and shape of cracks (white) and gridline defects (orange) as shown in Fig. 2.

The dataset was split into subsets for training, validation, and testing. Fifty images with a prevalence of cracks, gridline defects, and inactive areas were assigned to testing, evenly split for mono- and multi-silicon wafers. Fifty-four images were randomly selected from the remaining images for validation. The remaining images were assigned for training and augmented using a 180° rotation, mirror, and flip yielding 896 mono-si cell images and 1016 multi-si cell images for a total of 1912 images in the training dataset.

3.3. Deep learning models

Table 1 summarizes the 12 fully supervised deep learning models trained and tested to identify the best performing architecture for this benchmark dataset. The model ID is used in the subsequent analysis to identify each of the 12 models. The python code for training each model was adapted from code published on the respective GitHub repositories [43–47]. The U-Net was included because it performs well with a small set of labelled data [48], and it pioneered the symmetric encoder-decoder model with skip connections [38] which forms the basis of many current deep learning models. A second, more complex U-Net was included to determine if the additional trainable parameters would result in better predictions compared to the smaller U-Net. The PSPNet was selected for the depth and complexity of the model which includes a novel pyramid pooling module to improve complex scene understanding [49]. DeepLabv3+ was selected for the promise of improved predictions especially along the object boundaries [50].

Weighted cross entropy was selected for the training loss function as it is widely used for semantic segmentation with imbalanced datasets [51]. Each model was trained with three different sets of class weights. The equal class weights (Set A) assigned a value of one to all classes ($wA_1 = wA_2 = \ldots = wA_c = 1$) where, wA_c is the equal class weight for the cth class. The 'inverse' class weights (Set B) assigned a value for each class equal to the inverse of the median percentage of pixels in each class across all images in the training dataset. (1)

$$wB_c = \frac{1}{median(p_{c1},\ p_{c2},\ \ldots,\ p_{cn})} \tag{1}$$

wB_c = inverse class weight for the cth class p_{cn} = percentage of pixels in the cth class for the nth image n = number of images in the test dataset

The custom class weights (Set C) were pre-defined based on engineering judgement to improve detection of cracks, gridlines, and inactive areas by assigning larger weights to those classes (Table 2).

The ratio of 100:1 for cracks to background was set to enhance crack detection and minimize the impact of noise from the grain boundaries in
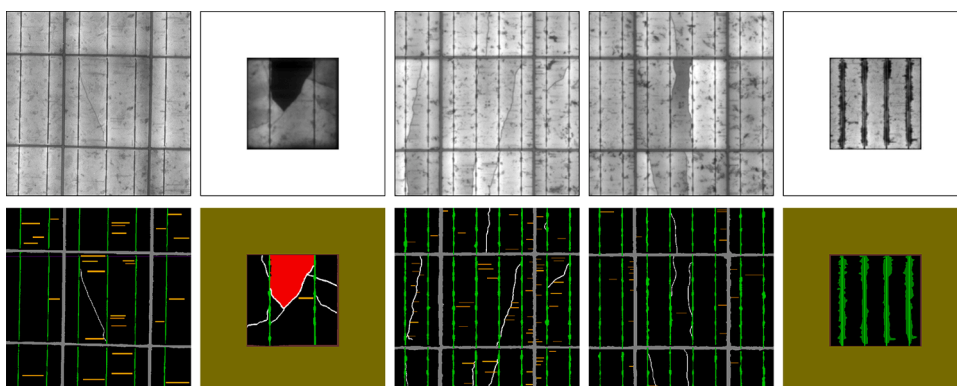

Fig. 2. EL images and ground truth masks for a sample of multi-crystalline silicon cells.

Table 2
Custom class weights.

Description	Value
$wC_{background}$	0.2
$wC_{ribbons}$	3.0
$wC_{gridline}$	10.0
$wC_{inactive}$	10.0
wC_{crack}	20.0
otherwise	1.0

the multi-crystalline solar cells. The class weights for gridline and inactive regions were set to ten to improve detection relative to other defects. The class weight for ribbons was set to 3 to improve detection of this feature and use it to classify the cell type with respect to the number of interconnect ribbons in future work. The choice of the custom weights was validated with a 16-run screening Design of Experiment (DOE) using these custom weights as the center point. The low values were set to one-half the custom weights and the high values were set to one-half the inverse weights. Analysis of the DOE showed the custom weights ranked among the best solutions for detecting background, cracks, gridlines, ribbons, and inactive areas on this benchmark dataset.

Fig. 3 plots the training and validation logs from all twelve models. However, not all loss functions are shown because of the scale chosen for this graphic. The training loss exceeded one for all the DeepLabv3+ models. The validation loss exceeded one for the DeepLabv3+ and the remaining models with inverse class weights.

3.4. Method overview

Fig. 4 summarizes the methodology followed for this work. The EL images and ground truth labels were curated, split into subsets, augmented, and trained using three different class weights. The models were used to generate a prediction mask for each image in the test dataset in which each pixel was labelled from 0 to 23 corresponding to the predicted class. The mask was colored to facilitate the visual analysis of each mask. The median IoU was computed for each image/class combination and plotted for comparison. The EL images and the corresponding ground truth masks were published on github (https://github.com/TheMakiran/BenchmarkELimages).
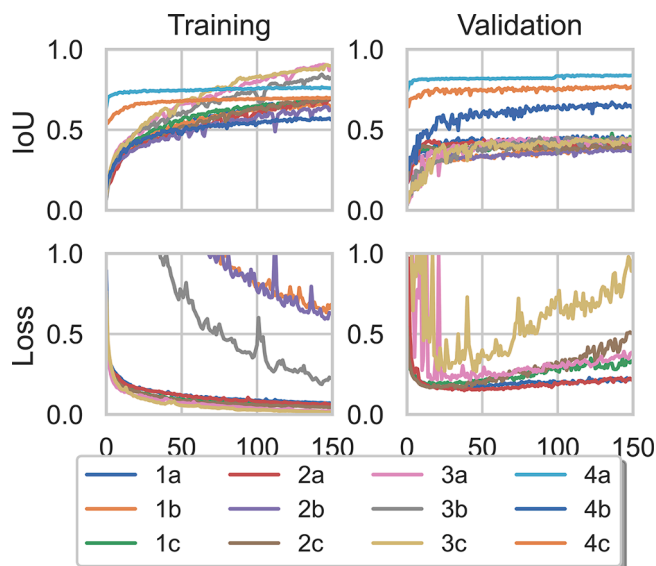

Fig. 3. Median IoU and loss functions for U-Net_12 (1a/b/c), U-Net_25 (2a/b/c), PSPNet (3a/b/c), and DeepLabv3+ (4a/b/c).

4. Results and discussion

Fig. 5 presents the prediction masks for the EL image of one cell (CFVS 00035_r8_c4) from all 12 models. This cell was selected because it contains the three defects of primary concern: cracks, gridline defects, and inactive areas. The rows correspond to the model: U-Net_12 (row 1), U-Net_25 (row 2), PSPNet (row 3), and DeepLabv3+ (row 4). The columns correspond to class weights: equal class weights (a), inverse class weights (b), and custom class weights (c). All four models generate similar prediction masks across the three sets of class weights, i.e., all the rows look similar. They all detect cell spacing (grey) and ribbons (green) well. However, the equal class weights predict fewer/smaller cracks compared to inverse class weights and custom class weights. The prediction masks generated from inverse class weights and custom class weights tend to show dilated cracks and gridline defects.

Fig. 6 shows the IoU for three defects and two features on the predictions from the 12 models for the test dataset. The lines connect the mIoU for each class and the colors correspond to mono- (blue) and multi-crystalline (red) silicon wafers. The grid layout is consistent with Fig. 5 to facilitate a visual comparison of one sample image and the combined results for the test dataset. All models detect the features (spacing and ribbons) better than the defects (cracks, gridlines, and inactive areas). This correlates to the relative size of the features and defects. The features occupy a greater number of pixels in the images than the defects. Based on the median value for number of pixels in the test dataset, the cracks account for 557 pixels, the inactive areas account for 1288 pixels, and the gridlines account for 1750 pixels. The spacing accounts for 12,529 pixels, and the ribbons account for 12,871 pixels. Thus, the features are roughly 10 to 20 times larger than the defects, which may explain the higher mIoU for features compared to defects. The best results for both multi- and mono-silicon wafers was observed on the DeepLabv3+ with custom weights. The mIoU for cracks and gridline defects was generally higher on mono-crystalline silicon compared to multi-crystalline wafers.

Fig. 7 shows a heatmap for the mIoU for the five selected classes across 12 models and the average of the mIoU for defects (crack, gridline, inactive) and features (ribbons, spacing). The darker green cells correspond to high mIoU in each column and red cells correspond to low mIoU. All the models with equal class weights (1a, 2a, 3a, 4a) have relatively high scores for the larger features (0.73–0.75) and relatively low scores for defects (0.04–0.24). The mIoU for the larger features compares reasonably well to the IoU values reported in Ref. [52] for detecting objects such as bikes (0.41), boats (0.68), chairs (0.43), tables (0.65), sofas (0.64), and TVs (0.73). However, the examples in Ref. [52] also illustrate the challenges around detecting long, narrow objects such as bird feet, table legs, chair legs, cat tails, and tires which form part of the larger object. These smaller features were often reduced, dilated, or missing altogether in the predictions presented in Ref. [52]. The mean IoU scores in Ref. [52] ranged from 62.2% to 78% across a range of objects despite the fact that many of the smaller features were missing in the predictions. The mIoU for cracks and gridlines reported in Fig. 7 is understandably low because cracks and gridlines consist exclusively of long, narrow features which are challenging to precisely locate at pixel-level resolution. The highest average mIoU for defects (0.28) was generated by the DeepLabv3+ with custom class weights (4c) making it the best choice for detecting long, narrow defects in this dataset. This result may be due to improved detection along object boundaries by the DeepLabv3+ model as claimed in Ref. [50].

Fig. 8 shows a heatmap for the median recall (mRcl) and the corresponding averages for the same five classes and models shown in Fig. 7. The models using equal class weights (1a, 2a, 3a, 4a) show the lowest average mRcl for defects (0.05–0.3). The models using inverse class weights (1b, 2b, 3b, 4b) show the highest average mRcl for defects (0.48–0.77). These statistics are consistent with the images in Fig. 5 which show dilated cracks and gridline defects in predictions from models when the inverse class weights were applied compared to models
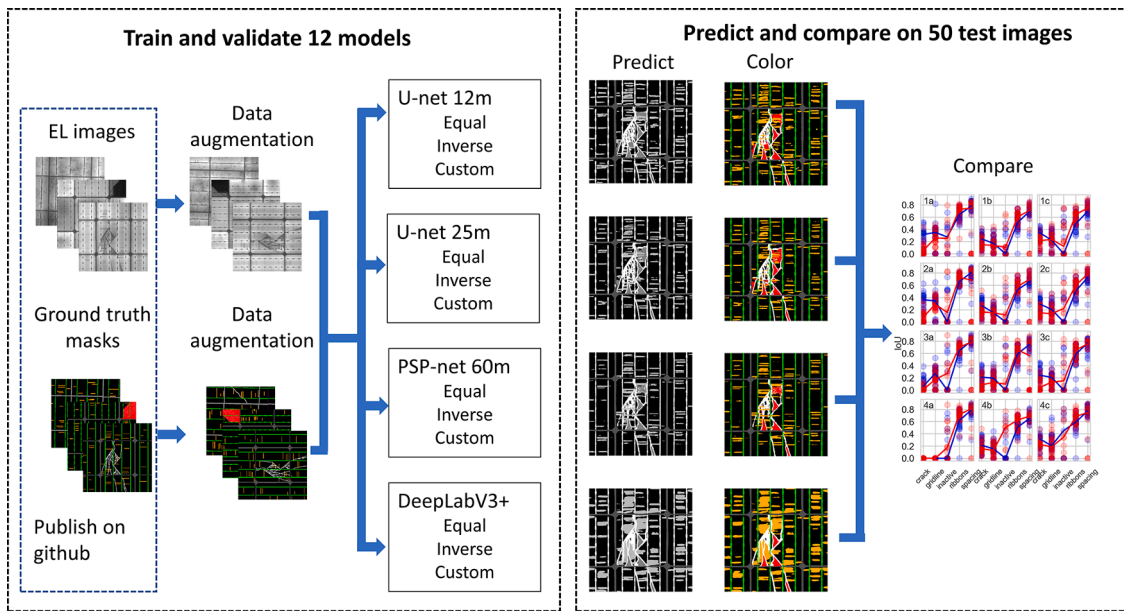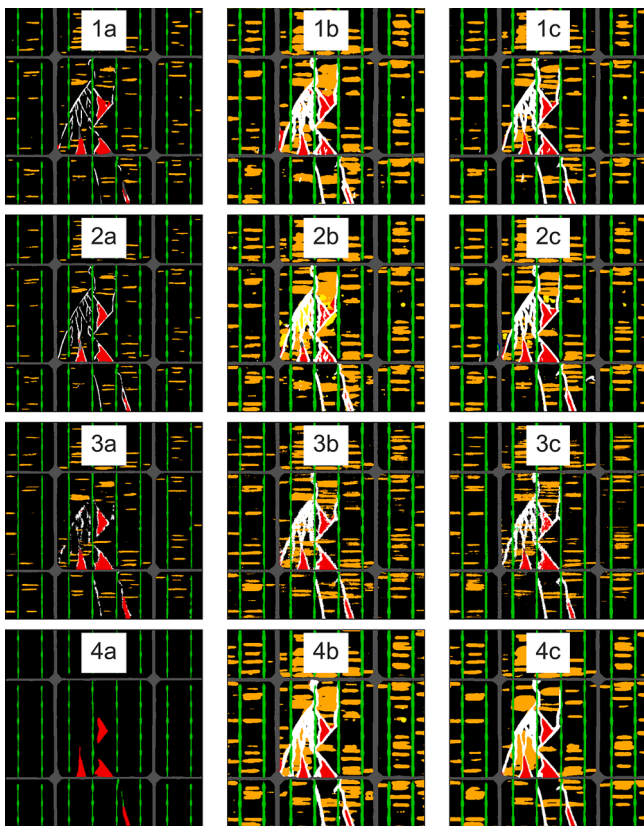
Fig. 4. Graphical summary of the method.

Fig. 5. Prediction masks for cell CFVS 00035_r8_c4: U-Net_12 (row 1), U-Net_25 (row 2), PSPNet (row 3), DeepLabv3+ (row 4), equal class weights (column a), inverse class weights (column b), and custom class weights (column c).

Fig. 6. IoU for mono-crystalline (blue circles), multi-crystalline (red) circles), and mIoU (solid lines) for selected defects and features in 50 test images across 12 models.

with equal class weights. The models using custom class weights (1c, 2c, 3c, 4c) also show relatively high average mRcl for defects and visually dilated defects in Fig. 5. The DeepLabv3+ with custom class weights (4c) is among the top performers for the average of the mRcl for both features and defects. The mRcl for cracks (0.86) and gridline defects (0.85) from
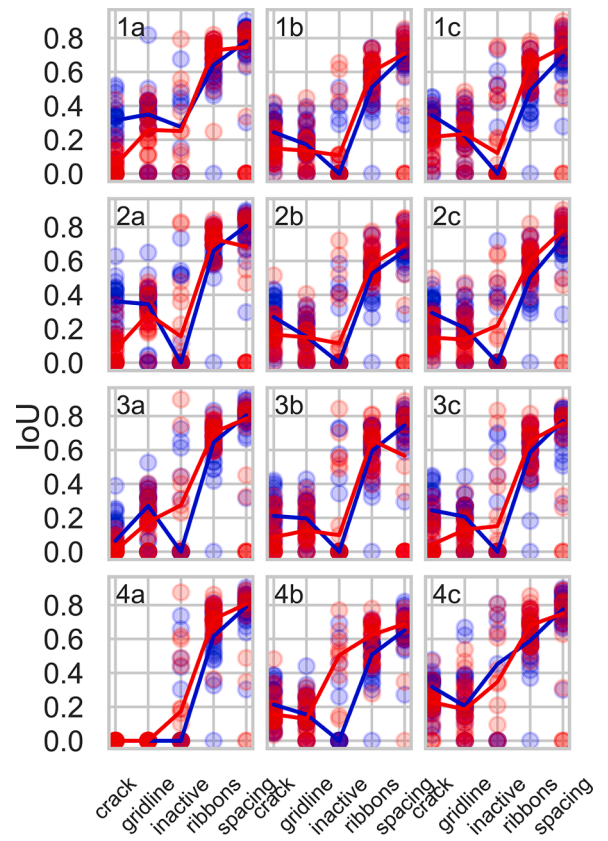
model 4c as shown in Fig. 8 also compares favorably to the recall for cracks (0.82) and gridline/contact defects (0.78) reported in Ref. [36].

While the DeepLabv3+ (4c) shows a high recall averaged across the three defect classes, the mIoU remains relatively low due to the low precision. Fig. 9 provides some insight to explain the low precision. The ground truth mask accurately reflects the cracks and gridline defects in

	crack	gridline	inactive	ribbons	spacing	avg_defects	avg_features
1a	0.13	0.32	0.25	0.68	0.78	0.24	0.73
1b	0.19	0.15	0.00	0.57	0.70	0.11	0.63
1c	0.26	0.22	0.00	0.58	0.72	0.16	0.65
2a	0.15	0.32	0.09	0.72	0.77	0.19	0.74
2b	0.19	0.15	0.04	0.54	0.67	0.13	0.61
2c	0.22	0.18	0.00	0.57	0.74	0.13	0.66
3a	0.03	0.23	0.12	0.68	0.80	0.13	0.74
3b	0.15	0.17	0.00	0.61	0.74	0.10	0.68
3c	0.17	0.14	0.11	0.61	0.77	0.14	0.69
4a	0.00	0.00	0.13	0.70	0.79	0.04	0.75
4b	0.18	0.14	0.20	0.56	0.66	0.17	0.61
4c	0.25	0.20	0.38	0.63	0.77	0.28	0.70

Fig. 7. mIoU for five classes across 12 models and the average of the mIoU for features and defects.

	crack	gridline	inactive	ribbons	spacing	avg_defects	avg_features
1a	0.14	0.44	0.32	0.74	0.91	0.30	0.83
1b	0.92	0.87	0.00	0.95	0.97	0.60	0.96
1c	0.90	0.84	0.00	0.97	0.96	0.58	0.96
2a	0.18	0.43	0.20	0.85	0.88	0.27	0.86
2b	0.92	0.93	0.05	0.97	0.97	0.63	0.97
2c	0.89	0.93	0.00	0.99	0.96	0.61	0.98
3a	0.03	0.30	0.17	0.81	0.90	0.17	0.85
3b	0.62	0.82	0.00	0.95	0.95	0.48	0.95
3c	0.43	0.68	0.18	0.96	0.96	0.43	0.96
4a	0.00	0.00	0.15	0.83	0.87	0.05	0.85
4b	0.93	0.92	0.46	0.97	0.99	0.77	0.98
4c	0.86	0.85	0.55	0.98	0.95	0.75	0.96

Fig. 8. mRcl for five classes across 12 models and the average of the mRcl for features and defects.

the corresponding EL image. The prediction mask identifies and localizes the defects in the EL images, but the defects are dilated relative to the ground truth mask leading to low precision.

The accuracy of the ground truth image can also impact the IoU and recall during testing. Some mismatch between the EL and the ground truth masks is inevitable, despite efforts to mitigate this issue during the labelling process. Fig. 10 shows the bottom left corner of

ARTS_00020_r7_c3 to illustrate the potential for error especially for long, narrow objects. The spacing layer in the prediction mask (gray) covers the corresponding feature in the EL image more completely than the ground truth mask which shows some dark pixels along the edges of the spacing layer unmasked. Similarly, the ribbons layer in the prediction mask (green) may reflect the corresponding feature in the EL image better than the ground truth mask. Errors in labelling along the edges of cracks and gridline defects have a particularly big impact on the mIoU due to the higher perimeter to area aspect ratios of those defects. The prediction mask in Fig. 10 also shows some additional gridline defects that should be added to the ground truth mask.

The low mIoU values for defects are driven by the low precision. The low precision is driven more by dilated predictions of the correct features/defects rather than detection of spurious features/defects. However, the extent of the dilation and spurious detections was deemed acceptable at this stage. The impact of the dilation could be mitigated by adapting the class weights, resulting in a trade-off between recall and precision. As seen in Fig. 5, when the class weights are equal, the dilation of features and defects is minimal.

In addition to the key defects and features analyzed above, the models also learned to detect other features and defects despite relatively small sample sizes. Fig. 11 shows the EL image, ground truth, and prediction mask for an image sourced from [19], one of the public datasets used for training. The authors originally trained models to classify images as 'good', 'cracked', or 'corroded'. In this work, the model detected and localized the corrosion defect in the image despite having only 17 images with the corrosion defect in the training dataset. The corrosion defect (dark green) was detected around the ribbon feature (bright green). The large olive padding feature added around the cell during pre-processing was clearly detected, as was the smaller border feature (brown) surrounding the cell. Some of the gridline defects were detected but the crack was not. The model also learned to correctly ignore the grain boundaries prevalent in multi-crystalline silicon cells.

5. Conclusion

Four deep learning models (U-Net_12, U-Net_25, PSPNet, and DeepLabv3+) were trained using equal class weights, inverse class weights, and custom class weights for a total of twelve sets of predictions for each of 50 test images. The models were trained to simultaneously detect 24 classes in EL images of solar PV cells using semantic segmentation. Twelve classes correspond to intrinsic features of a solar cell, and twelve classes correspond to extrinsic defects. This paper focused on the detection of three critical defects and two common features in crystalline silicon solar cells. The DeepLabv3+ with custom class weights was found to have the highest mIoU averaged across three critical defects identified in this dataset. The mIoU for the features was significantly higher than the mIoU for the defects for all models tested. Notably, the models achieved lower mIoU and mRcl on the small, narrow defects such as cracks and gridlines compared to large features such as spacing and ribbons. This is likely driven more by the dilated features in the predication masks and less due to spurious misclassifications. The models tested are effective in detecting, localizing, and quantifying multiple features and defects in EL images of solar cells. These models can thus be used to not only detect the presence of defects, but to track their evolution over time as modules are re-imaged throughout their lifetime.

The unique contributions from this work include the benchmark dataset published on github and benchmark performance metrics for semantic segmentation on EL images of solar PV cells.

Future work

Future work will focus on semi-supervised learning to incorporate information from the 80,000+ unlabeled EL images available. Image-to-image translation techniques will be investigated for improving the
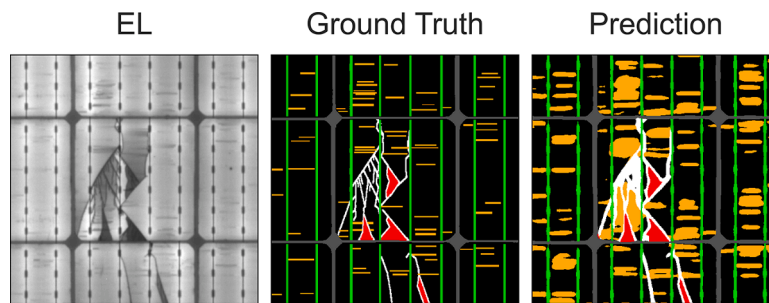
Fig. 9. EL, ground truth mask, and prediction mask from DeepLabv3+ with custom class weights (4c) for CFVS 00035_r8_c4.
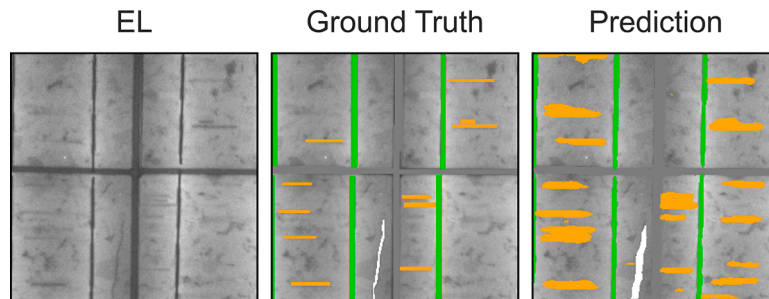

Fig. 10. Bottom left corner of ARTS_00020_r7_c3 showing the labelling errors along the edges of ribbons and cell spacing in the ground truth image leading to errors in precision and recall.
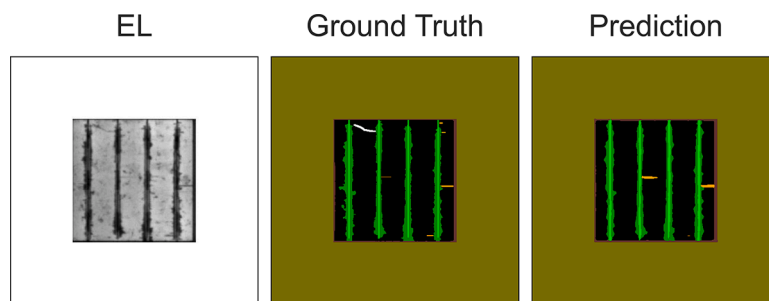

Fig. 11. EL, ground truth, and prediction mask from DeepLabv3+ with custom class weights (4c) for SDLE_00513 showing corrosion of the ribbon interconnects.

accuracy of ground truth images. Finally, the ground truth dataset will be updated and increased to incorporate new features and defects such as half-cut cells, scuffs and edge corrosion.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data will be made public on the github repo referenced in the paper.

Acknowledgments

References

[1] T. Fuyuki, H. Kondo, T. Yamazaki, et al., Photographic surveying of minority carrier diffusion length in polycrystalline silicon solar cells by electroluminescence, Appl. Phys. Lett. 86 (2005), https://doi.org/10.1063/1.1978979.

[2] M. Reuter, L. Stoicescu, J.H. Werner, PV module electroluminescence: enlightening defects. White paper by Solarzentrum Stuttgart GmbH. https://www.solarzentrum-stuttgart.com/uploads/file/platzhalter_vortrag_spezial_hagelschaden_DaySy_April2015_04.pdf.

[3] S. Koch, T. Weber, C. Sobottka, et al., Outdoor electroluminescence imaging of crystalline photovoltaic modules: comparative study between manual ground-level inspections and drone-based aerial surveys, in: Proceedings of the 32nd European Photovoltaic Solar Energy Conference and Exhibition, 2016.

[4] W.B. Hobbs, B. Hamzavy, C.B. Jones, et al., In-field electroluminescence imaging: methods, comparison with indoor imaging, and observed changes in modules over one year, in: Proceedings of the 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC), Waikoloa Village, HI, IEEE, 2018, pp. 3257–3260.

[5] K.G.Bedrich, Y.S. Khoo, Y. Wang, Method, system, and image processing device for capturing and/or processing electroluminescence images, and an aerial vehicle, Patent number WO2021137764, 2021.

[6] K.G. Bedrich, M. Bliss, T.R. Betts, R. Gottschalg, Electroluminescence imaging of PV devices: camera calibration and image correction, in: Proceedings of the 2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC), Portland, OR, USA, IEEE, 2016, pp. 1532–1537.

[7] C. Mantel, S. Spataru, H. Parikh, et al., Correcting for perspective distortion in electroluminescence images of photovoltaic panels, in: Proceedings of the 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC 34th EU PVSEC), 2018, pp. 0433–0437.

[8] M. Owen-Bellini, D.B. Sulas-Kern, G. Perrin, et al., Methods for *in situ* electroluminescence imaging of photovoltaic modules under varying environmental conditions, IEEE J. Photovolt. 10 (2020) 1254–1261, https://doi.org/10.1109/JPHOTOV.2020.3001723.

[9] M. Köntges, I. Kunze, S. Kajari-Schröder, et al., Quantifying the risk of power loss in PV modules due to micro cracks, in: Proceedings of the 25th European Photovoltaic Solar Energy Conference, 2010, p. 9.

[10] M. Abdelhamid, R. Singh, M. Omar, Review of microcrack detection techniques for silicon solar cells, IEEE J. Photovolt. 4 (2014) 514–524, https://doi.org/10.1109/JPHOTOV.2013.2285622.

[11] H. Gopalakrishna, R. Yadav, E. Chan, J.N. Munday, P. Bermel, M.D. Kempe, et al., Durability evaluation of PV modules using image processing tools. New Concepts in Solar and Thermal Radiation Conversion and Reliability, SPIE, San Diego, United States, 2018, p. 36.

[12] K.G. Bedrich, W. Luo, M. Pravettoni, et al., Quantitative electroluminescence imaging analysis for performance estimation of PID-influenced PV modules, IEEE J. Photovolt. 8 (2018) 1281–1288, https://doi.org/10.1109/JPHOTOV.2018.2846665.

[13] M. Bdour, Z. Dalala, M. Al-Addous, et al., A comprehensive evaluation on types of microcracks and possible effects on power degradation in photovoltaic solar panels, Sustainability 12 (2020) 6416, https://doi.org/10.3390/su12166416.

[14] E. Sovetkin, E.J. Achterberg, T. Weber, B.E. Pieters, Encoder–decoder semantic segmentation models for electroluminescence images of thin-film photovoltaic modules, IEEE J. Photovolt. 11 (2021) 444–452, https://doi.org/10.1109/JPHOTOV.2020.3041240.

[15] M. Dhimish, Y. Hu, Rapid testing on the effect of cracks on solar cells output power performance and thermal operation, Scientific Reports 12 (1) (2022) 1–11.

[16] A.M. Karimi, J.S. Fada, N.A. Parrilla, et al., Generalized and mechanistic PV module performance prediction from computer vision and machine learning on electroluminescence images, IEEE J Photovolt. 10 (2020) 878–887, https://doi.org/10.1109/JPHOTOV.2020.2973448.

[17] U. Jahn, M. Herz, M. Köntges, et al (2018), IEA PVPS T13-10 2018 Review on infrared and electroluminescence imaging for PV field applications: International Energy Agency Photovoltaic Power Systems Programme: IEA PVPS Task 13, Subtask 3.3: report IEA-PVPS T13-12:2018. International Energy Agency, Paris, https://iea-pvps.org/wp-content/uploads/2020/01/Review_on_IR_and_EL_Imaging_for_PV_Field_Applications_by_Task_13.pdf.

[18] S. Deitsch, C. Buerhop-Lutz, E. Sovetkin, A. Steland, A. Maier, F. Gallwitz, & C. Riess Segmentation of Photovoltaic Module Cells in Electroluminescence Images. arXiv 2018. *arXiv preprint arXiv:* 1806.06530.

[19] A.M. Karimi, J.S. Fada, M.A. Hossain, et al., Automated pipeline for photovoltaic module electroluminescence image processing and degradation feature classification, IEEE J Photovolt. 9 (5) (2019) 1324–1335.

[20] C. Mantel, F. Villebro, G. Alves dos Reis Benatto, et al., Machine learning prediction of defect types for electroluminescence images of photovoltaic panels. Applications of Machine Learning, SPIE, San Diego, United States, 2019.

[21] E. Sovetkin, A. Steland, Automatic processing and solar cell detection in photovoltaic electroluminescence images, Integrated Computer-Aided Engineering 26 (2) (2019) 123–137.

[22] S. Deitsch, V. Christlein, S. Berger, C. Buerhop-Lutz, A. Maier, F. Gallwitz, C. Riess, Automatic classification of defective photovoltaic module cells in electroluminescence images, Solar Energy 185 (2019) 455–468.

[23] M.W. Akram, G. Li, Y. Jin, et al., CNN based automatic detection of photovoltaic cell defects in electroluminescence images, Energy 189 (2019), 116319, https://doi.org/10.1016/j.energy.2019.116319.

[24] W. Tang, Q. Yang, K. Xiong, W. Yan, Deep learning based automatic defect identification of photovoltaic module using electroluminescence images, Solar Energy 201 (2020) 453–460, https://doi.org/10.1016/j.solener.2020.03.049.

[25] C. Ge, Z. Liu, L. Fang, et al., A hybrid fuzzy convolutional neural network based mechanism for photovoltaic cell defect detection with electroluminescence images, IEEE Trans. Parallel Distrib. Syst. (2020), https://doi.org/10.1109/TPDS.2020.3046018, 1–1.

[26] M.Y. Demirci, N. Beşli, A. Gümüşçü, Efficient deep feature extraction and classification for identifying defective photovoltaic module cells in Electroluminescence images, Expert Systems with Applications 175 (2021), 114810, https://doi.org/10.1016/j.eswa.2021.114810.

[27] W. Tang, Q. Yang, W. Yan, Deep learning-based algorithm for multi-type defects detection in solar cells with aerial EL images for photovoltaic plants, Comput.

Model. Eng. Sci. 130 (2022) 1423–1439, https://doi.org/10.32604/cmes.2022.018313.

[28] C. Huang, Z. Zhang, L. Wang, PSOPruner: PSO-based deep convolutional neural network pruning method for PV module defects classification, IEEE J. Photovolt. (2022) 1–9, https://doi.org/10.1109/JPHOTOV.2022.3195099.

[29] Y. Zhao, K. Zhan, Z. Wang, W. Shen, Deep learning-based automatic detection of multitype defects in photovoltaic modules and application in real production line, Prog. Photovolt. Res. Appl. 29 (2021) 471–484, https://doi.org/10.1002/pip.3395.

[30] X. Zhang, Y. Hao, H. Shangguan, et al., Detection of surface defects on solar cells by fusing multi-channel convolution neural networks, Infrared Phys. Technol. 108 (2020), 103334, https://doi.org/10.1016/j.infrared.2020.103334.

[31] B. Su, H. Chen, P. Chen, et al., Deep learning-based solar-cell manufacturing defect detection with complementary attention network, IEEE Trans Ind Inf 17 (2021) 4084–4095, https://doi.org/10.1109/TII.2020.3008021.

[32] D.M. Tsai, S.C. Wu, W.C. Li, Defect detection of solar cells in electroluminescence images using Fourier image reconstruction, Solar Energy Mater. Solar Cells 99 (2012) 250–262, https://doi.org/10.1016/j.solmat.2011.12.007.

[33] S. Spataru, P. Hacke, D. Sera, Automatic detection and evaluation of solar cell micro-cracks in electroluminescence images using matched filters, in: Proceedings of the 2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC), 2016, pp. 1602–1607.

[34] H. Chen, H. Zhao, D. Han, K. Liu, Accurate and robust crack detection using steerable evidence filtering in electroluminescence images of solar cells, Opt. Lasers Eng. 118 (2019) 22–33, https://doi.org/10.1016/j.optlaseng.2019.01.016.

[35] M.R.U. Rahman, H. Chen, Defects inspection in polycrystalline solar cells electroluminescence images using deep learning, IEEE Access 8 (2020) 40547–40558, https://doi.org/10.1109/ACCESS.2020.2976843.

[36] J. Fioresi, D.J. Colvin, R. Frota, et al., Automated defect detection and localization in photovoltaic cells using semantic segmentation of electroluminescence images, IEEE J. Photovolt. 12 (1) (2021) 53–61.

[37] L. Pratt, D. Govender, R. Klein, Defect detection and quantification in electroluminescence images of solar PV modules using U-net semantic segmentation, Renew. Energy 178 (2021) 1211–1222, https://doi.org/10.1016/j.renene.2021.06.086.

[38] S. Hao, Y. Zhou, Y. Guo, A brief survey on semantic segmentation with deep learning, Neurocomputing 406 (2020) 302–321, https://doi.org/10.1016/j.neucom.2019.11.118.

[39] I. Ulku, E. Akagündüz, A survey on deep learning-based architectures for semantic segmentation on 2D images, Appl. Artif. Intell. (2022) 1–45, https://doi.org/10.1080/08839514.2022.2032924.

[40] R.H. French, A.M. Karimi, J.L. Braid, 4). Electroluminescent (EL) Image Dataset of PV Module Under Step-wise Damp Heat Exposures. https://doi.org/10.17605/OSF.IO/4QRTV.

[41] C. Buerhop, S. Deitsch, A. Maier, et al., A benchmark for visual identification of defective solar cells in electroluminescence imagery, in: Proceedings of the 35th European PV Solar Energy Conference and Exhibition, 2018.

[42] S. Kimball, P. Mattis, and the GIMP Development Team, GIMP 2.10.32, 1995-2022, last checked on 1/25/2023, https://www.gimp.org.

[43] D. Gupta (2021) Image Segmentation Keras : Implementation of Segnet, FCN, UNet, PSPNet and other models in Keras [Source code], Accessed 23 Jan 2022, https://github.com/divamgupta/image-segmentation-keras.

[44] N. Tomar (2022) Semantic-Segmentation-Architecture [Source code], Accessed 23 Jan 2022, https://github.com/nikhilroxtomar/Semantic-Segmentation-Architecture.

[45] Y.Kamikawa (2020) keras-PSPNet [Source code], Accessed 23 Jan 2022, https://github.com/ykamikawa/tf-keras-PSPNet.

[46] S. Kawakita (2021) example_camvid_multiclassB_quita.ipyn [Source code], Accessed 23 Jan 2022, https://github.com/shirokawakita/multiclass-segmentation/blob/main/example_camvid_multiclassB_quita.ipynb.

[47] P. Yakubovskiy (2020) Segmentation Models in Pytorch [Source code], Accessed 10 Apr 2022, https://github.com/qubvel/segmentation_models.pytorch.

[48] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: Proceedings of the International Conference on Medical image computing and computer-assisted intervention, Cham, Springer, 2015.

[49] H. Zhao, J. Shi, X. Qi, et al., Pyramid scene parsing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2017, pp. 2881–2890.

[50] L.C. Chen, Y. Zhu, G. Papandreou, et al., Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European conference on Computer Vision (ECCV), 2018.

[51] S. Jadon, A survey of loss functions for semantic segmentation, in: Proceedings of the 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2020, pp. 1–7.

[52] G. Lin, C. Shen, A. van den Hengel, I. Reid, Efficient piecewise training of deep structured models for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, IEEE, 2016, pp. 3194–3203.