

The Empirical Digital Twin: Representation Learning on Solar Cell Images and Efficient Defect Detection with Human-in-the-Loop

Philipp Kunze¹, Stefan Rein¹, Thomas Mueller¹, Marc Hemsendorf², Klaus Ramspeck³, Matthias Demant¹

¹Fraunhofer Institute for Solar Energy Systems ISE, Heidenhofstr. 2, 79098 Freiburg, Germany

²GP Solar GmbH, Gottlieb-Daimler-Str. 5, 78467 Konstanz, Germany

³h.a.l.m. elektronik GmbH, Friesstr. 20, 60388 Frankfurt am Main, Germany

ABSTRACT: Measurement images of solar cells provide information beyond current-voltage characteristics regarding process and material quality in a spatially resolved manner. However, this information is only partially used because algorithms search for human-defined defects and structures. These labels can be inaccurate and incomplete, a relevance in terms of electrical quality is not necessarily given. Thus, we propose a purely data-based approach to derive a comprehensive representation from the measured images that is meaningful in terms of electrical quality and show how it can be used for efficient defect detection. We call this representation the empirical digital twin. For its calculation, we design a convolutional neural network combining multiple measurement images by correlating them with quality variables. The digital twin is an intermediate representation of the network and summarizes quality-related defect signatures that are visible in the images. We show how this representation can be used to derive sorting criteria for quality inspection within an efficient human-in-the-loop approach detecting defects such as finger interruptions, shunts, etc. The human-in-the-loop method not only needs fewer training samples and thus fewer labels but also improved the F_1 – *Score* detection rate by about 2% on average.

Keywords: Characterization, Deep Learning, Digital Twin, Representation Learning

1 INTRODUCTION

Imaging measurement methods such as electroluminescence (EL), photoluminescence (PL) or infra-red thermography (IR) images are broadly used in production to reveal material and process losses [1–5]. They allow defects such as microcracks, shunts, poor contact formation, etc. to be spatially visualized. Because of this property, they extend the opportunities available through analysis of the current-voltage curve (IV curve) by drawing conclusions about the defect origins and positions.

The interpretation and analysis of the measurement images is challenging even for experts due to overlapping defect structures and high complexity of the data. So far, different approaches have been presented for defect detection:

On the one hand, classical image processing algorithms are developed by experts. These usually combine filtering techniques to extract defect-typical structures and features with machine learning (ML) models such as support vector machines (SVMs) [6] to process those structures and features. For instance, microcracks and finger interruptions in PL images [7,8] or in EL images [9–16] are found by these kinds of techniques.

On the other hand, so-called end-to-end approaches have been pursued in recent years, which are mostly based on convolutional neural networks (CNNs). These are supposed to derive the defects directly from the images. The difference to classical image processing is that here the filters are optimized empirically with a large dataset and not developed by humans. In many works, for example, microcracks, dislocation structures or finger interruptions in EL and IR images are detected using CNNs [17–39].

Both techniques, be it based on image processing or CNNs, have disadvantages in practice limiting their application and making them costly. Especially for CNNs but also for image processing algorithms, annotations of the searched defects are needed before the networks can be trained and evaluated. Firstly, this process of labeling is

both time-consuming and costly, as well as error prone. Greulich et al. have shown that even experts cannot reliably find microcracks in EL images and also vary in their detection rate [40]. Secondly, the transferability to other processes is limited, so that re-labeling for CNNs or algorithmic adaptations may be necessary. Here, purely data-based techniques, as for example proposed by Demant et al. and Kovvali et al. regarding as-cut wafer rating, can help overcome these drawbacks [41–43].

Following these ideas, we propose the empirical digital twin, which is learned from measurement images without label overhead, purely data-driven, and can be used for efficient defect detection.

Our contributions are:

1. We present an approach to learn the empirical digital twin describing the quality status of the solar cell by combining the information from multiple measurement images.
2. We utilize the empirical digital twin for efficiently including expert knowledge into defect detection within an iterative “Human-in-the-Loop”-Approach.

This approach, with additional research and explanation, is described in our journal paper [44].

2 APPROACH

We propose a sequence of algorithms for solar cell inspection. First, we derive a meaningful representation of the solar cell from measurement images by semantically compressing measurement images with respect to quality-describing quantities. The representation, which we call empirical digital twin, and its computation is explained in Section 2.1. Following this, we present an efficient way to use the empirical digital twin to incorporate expert knowledge into defect detection. We describe this iterative human-in-the-loop approach in Section 2.2.

2.1 Learning the empirical digital twin

We compute a comprehensive representation of solar cell measurement images by deriving features from them

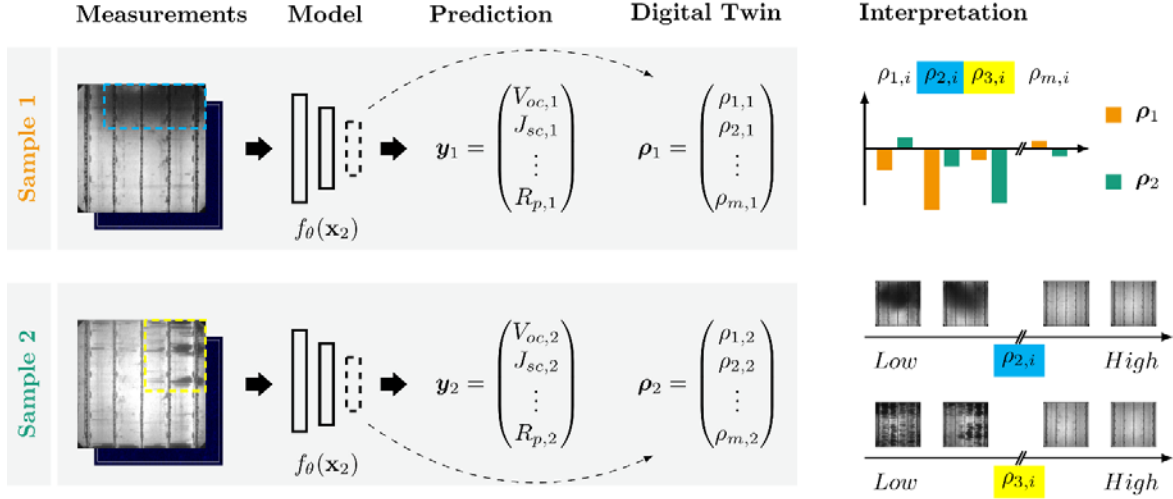


Figure 1: Schematic representation of the calculation of the empirical digital twin for two example cells. The measurements (first column) are passed to a CNN model (second column), which predicts quality variables (third column). The model is used to derive the empirical digital twin (fourth column), whose entries represent quality-describing image features (fifth column).

that are expressive in terms of quality-describing quantities such as IV parameters. The procedure is shown in Figure 1 for two samples (one per row). The measurement images \mathbf{x}_i of the i -th sample are given to a CNN model $f_\theta(\mathbf{x})$ which predicts the quality parameters in \mathbf{y}_i , θ being the model's parameters to be optimized.

The CNN consists of a sequence of some convolutional and pooling layers as well as rectified linear unit (ReLU) activation functions. This gradually reduces the spatial resolution of the measurement images while increasing the semantic expressiveness with respect to the target parameters. The last neurons of the network form a vector $\boldsymbol{\rho}_i$ describing the connection between the images and the quality variables. In the following, we call the vector empirical digital twin (EDT).

The EDT offers the possibility to compare measurement images quantitatively. Column 5 in Figure 1 shows the distribution of $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$ in the form of a bar diagram. Below this, the extreme examples of the dataset studied are shown for features $\rho_{2,i}$ and $\rho_{3,i}$ revealing which image features they represent. Feature 2 (blue) represents poor contact formation during the firing process and feature 3 (yellow) represents finger interruptions. These two process defects occur in the example samples, highlighted by the blue and yellow outlines in the measurement, respectively. Accordingly, these entries in the corresponding EDTs are particularly low as shown in the bar diagram.

In the proposed approach, several different measurement images can be passed to the model and the model can predict various quality parameters. We use EL as well as IR images in conjunction with reflectance values at 390nm and 950nm wavelength, scaled up to match the images' spatial dimensions. In Table 1, columns 1 and 2 show the 18 parameters that the model should predict. These include typical IV parameters such as the open circuit voltage V_{oc} , the short circuit current density J_{sc} , the fill factor FF and the efficiency η , as well as more specific parameters such as the pseudo fill factor pFF , the ideal fill factor FF_0 and their difference $FF_0 - pFF$. We have chosen a large set of parameters so that the EDT can be a comprehensive representation of many properties. However, it is an open question which and how many parameters are best suited for a meaningful EDT.

The EDT vector $\boldsymbol{\rho}$ consists of 720 entries in our implementation. For each of the predicted parameters, 40 entries are provided in $\boldsymbol{\rho}$ within our realization. This adds up to 720 values for 18 parameters. We found this number effective in terms of prediction accuracy, however, it can be optimized in future works.

2.2 Efficient defect detection with Human-in-the-Loop

The digital twin is well suited for quality inspection because it contains the essential measurement image information in a compressed form. Quality-relevant defect structures are already learned indirectly through the regression of the quality parameters. In addition, EDTs have significantly fewer dimensions than their original measurement images and correspondingly less complexity, making it easy to optimize a classifier model based on them.

To use the EDT efficiently and with low label overhead for defect detection, we propose an iterative human-in-the-loop approach, which is schematically outlined in Figure 2. It consists of four steps:

1. The EDTs are calculated for all measurement images of the dataset as described in Section 2.1. From these, a small, representative selection is taken for labeling.
2. This compact dataset is labeled by an expert based on the corresponding measurement images.
3. An ML model based on the labeled compact dataset is trained.
4. The model is applied to the remaining unlabeled dataset and an uncertainty value is calculated for each sample. The most uncertain samples are again given to the expert to be labeled.

Steps 2-4 can be iteratively repeated so that only the most uncertain samples are used for training, since it is assumed that these can give the most information to a classifier model.

The individual steps are described in more detail below. For the representative selection of the of samples for a compact dataset (step 1), we use a k-means [6] with n_{cl} clusters. The n_{cl} samples whose EDTs have the shortest distance to the cluster centroids are selected. These are sorted into defect classes by experts considering the corresponding measurement images (step 2). As a model,

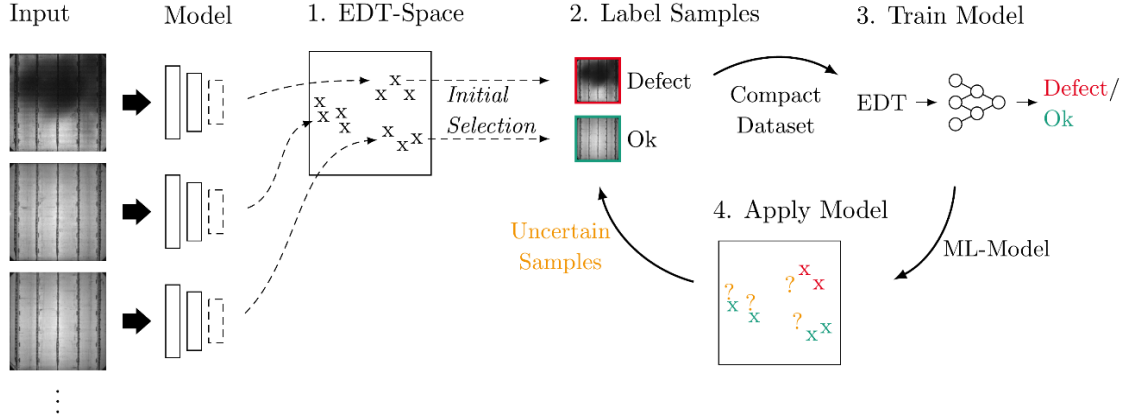


Figure 2: Schematic representation of the human-in-the-loop approach. (1) Measurement images are mapped into the space of empirical digital twins, (2) from which a representative selection based on the measurement images are labeled. (3) Based on the compact dataset, a sorting model is trained. (4) After application to the rest of the dataset, the most uncertain samples are labeled.

we use a small neural classification network (NN) consisting of a few linear layers and ReLU activation functions, which is trained with the n_{cl} EDTs as input and expert labels as targets (step 3). Regarding the uncertainty, we consider the NN as Bayesian approximation by applying dropout [45] multiple times, as described in [46,47]. The NN is described as $g: \boldsymbol{p} \mapsto p_d \in [0,1]$, which maps an EDT \boldsymbol{p} to a defect probability p_d . Based on this, we can calculate the entropy H as described in Equation (1):

$$H(\boldsymbol{p}) = -p_d \log_2 p_d - (1 - p_d) \log_2 (1 - p_d) \quad (1)$$

By applying dropout, i.e. with a certain probability some activations of the NN are randomly set to 0 (are dropped), and multiple entering of the same EDT \boldsymbol{p}_i to the NN, an entropy distribution results. Equation (2) describes that as uncertainty value U the mean value of this distribution is used:

$$U(\boldsymbol{p}_i) = \frac{1}{T} \sum_{t=1}^T H(\boldsymbol{p}_i) \quad (2)$$

T describes the number of inputs.

3 EXPERIMENTS

The dataset consists of 1600 Cz-Si industrially processed passivated emitter and rear cells (PERCs) of size $156 \times 156 \text{ mm}^2$. They were sorted out due to electrical or optical defects. Therefore, they contain various defects such as finger interruptions, microcracks, shunts, edge isolation defects, poor contact formation, etc. Each of the cells was measured by electroluminescence and thermography by a system from h.a.l.m. elektronik GmbH. For the EL measurements, the cells were excited with a current of 20A and the integration time of the Si CCD camera was 50ms. In addition, the reflectance at 390nm and 950nm wavelengths and the parameters from Table 1 were measured.

As CNN we use a variation of the DenseNet [48]. Here, the EL and IR images and the reflectance measurements are first processed in separate CNN branches. Then the results of these are concatenated and processed together. Finally, this path splits to predict the individual quality parameters and compute the EDT. A detailed description of the architecture can be found in [44].

The network is optimized to predict the quality parameters in Table 1, as described in Section 2.1. The input data (EL, IR, and reflectance) are scaled to a size of

Table 1: Predicted parameters, their absolute errors and correlation coefficients with respect to the measured values.

Name	Parameter	Absolute Error	Correlation Coefficient
Open circuit voltage	V_{oc}	0.69 mV	0.93
Short circuit current density	J_{sc}	0.06 mA cm ⁻²	0.62
Fill factor	FF	0.55 %	0.88
Efficiency	η	0.16 %	0.88
Pseudo fill factor	pFF	0.52 %	0.86
Ideal fill factor	FF_0	0.02 %	0.91
Difference: Pseudo & fill factor	$pFF - FF$	0.16 %	0.98
Difference: Ideal & pseudo fill factor	$FF_0 - pFF$	0.53 %	0.86
Suns open circuit voltage	$sunsV_{oc}$	0.65 mV	0.94
Pseudo Efficiency	$p\eta$	0.15 %	0.86
Saturation current density D1	J_{01}	0.01 pA	0.96
Saturation current density D2	J_{02}	2.6 nA	0.94
Reverse current density @ -12V	$J_{rev,1}$	1.05 mA cm ⁻²	0.88
Reverse current density @ -15V	$J_{rev,2}$	1.15 mA cm ⁻²	0.89
Grid Resistance Front	$R_{grid,fr}$	3.56 Ω m ⁻¹	0.93
Grid Resistance Rear	$R_{grid,re}$	1.05 Ω m ⁻¹	0.53
Series Resistance	R_s	0.02 Ω cm ²	0.98
Parallel Resistance	R_p	0.29 Ω cm ²	0.83

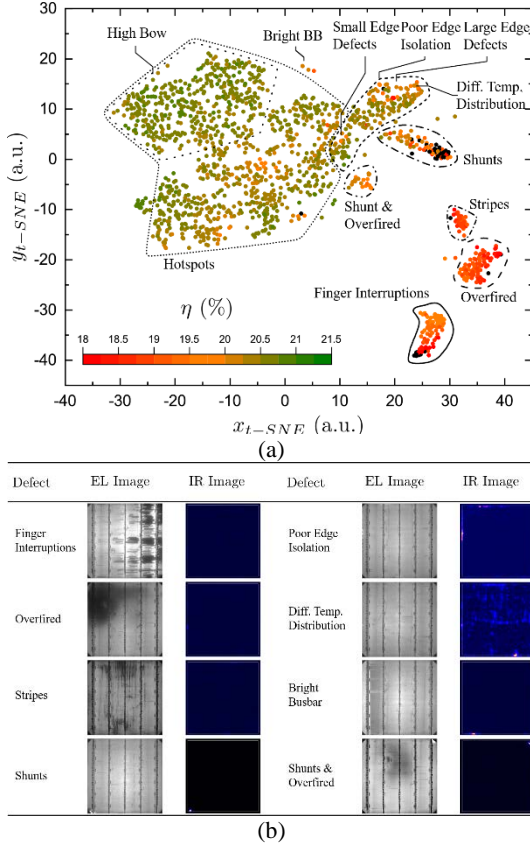


Figure 3: (a) Low-dimensional embedding of the empirical digital twins with cluster annotations and (b) example images for each of the annotated clusters.

$224 \times 224 \text{px}^2$, standardized using mean and standard deviation, and randomly rotated and flipped during training. The model is trained on an Nvidia GeForce RTX 2080 Ti GPU for 250 epochs with a batch size of 40 on the training dataset. The adam-optimizer [49] is used with an L1 loss function. The best model is selected based on the validation dataset and tested on the test dataset.

3.1 Analysis of the empirical digital twins

To investigate the meaningfulness of the EDTs, an exploratory analysis of the representation space is performed. For this purpose, all measurements were examined by an expert, so that the main losses of each cell are known. Since the EDTs have 720 dimensions, they cannot be examined directly. Therefore, the t-SNE [50] algorithm is applied for the analysis, which reduces the high-dimensional EDTs to two dimensions. Here, neighboring points in the high-dimensional space are also close to each other in the two-dimensional space. The resulting clusters are compared with the expert labels.

3.2 Defect detection with Human-in-the-Loop

In this experiment, the possibilities of process defect detection and the human-in-the-loop approach are investigated. All cells were sorted according to the defect classes finger interruptions, poor edge isolation, overfired regions and hotspots and shunts by an expert.

Two approaches to defect detection are compared. First, based on the entire labeled dataset, an NN for defect detection is trained with a 5-fold cross-validation. This NN serves as a supervised reference to the human-in-the-loop approach. Here, the same NN is also trained with a 5-fold cross-validation as described in Section 2.2. Starting with $n_{cl} = 100$ samples, in each iteration, the 100 most uncertain samples are added to the training dataset until 1000 samples are reached. Uncertainty is calculated using Equation (2) with $T = 200$. Results are compared using *Precision*, *Recall* and *F1-Score*. These quantities are described in the appendix.

4 RESULTS

4.1 Results - Analysis of the empirical digital twins

The quality parameters of the cells can be accurately predicted by the trained CNN. In Table 1, columns three and four show the respective absolute error and correlation coefficient between predicted and measured values. In most cases the correlation coefficient is above 0.8 and sometimes even up to 0.98. Exceptions are the grid resistance on the rear side $R_{grid, re}$ and the short circuit current density J_{sc} , which have lower correlation coefficients of 0.53 and 0.62, respectively.

The EDTs contain information regarding quality-relevant process properties and defects. Figure 3 (a) shows the low-dimensional embedding of the EDTs, the points are colored according to their efficiency. Clusters are formed which contain samples with specific defects. Based on our previous expert labels the groups are bordered and annotated. Typical examples for each cluster are shown in Figure 3 (b).

In Figure 3 (a), four smaller clusters with reduced efficiency are observed on the right side. They contain finger interruptions, overfired regions, and stripe structures and shunts. Example EL and IR images are shown in Figure 3 (b) in the left column. Cells with poor edge isolation and diffuse temperature distribution are positioned in the region above. The example images in Figure 3 (b) in the right column show an increased temperature in the IR image at the edges of the cell, and a larger area of increased temperature, respectively. The rest of the large cluster contains only minor losses, but splits into areas of different bows. In addition, there are two smaller clusters which show a combination of shunts and overfiring, or a radiating region around the busbar contact. Example images of this can be found in Figure 3 (b) in the right column in rows three and four.

Table 2: Detection results of the supervised reference and human-in-the-loop approach with respect to the process defects finger interruptions, poor edge isolation, hotspots & shunts and firing defect.

Process Defect	Supervised Reference			Human-in-the-Loop		
	$F_1 - \text{Score}$	<i>Recall</i>	<i>Precision</i>	$F_1 - \text{Score}$	<i>Recall</i>	<i>Precision</i>
Finger Interruptions	0.972	0.946	1	0.991	0.982	1
Poor Edge Isolation	0.861	0.812	0.919	0.897	0.874	0.922
Hotspots & Shunts	0.959	0.970	0.949	0.963	0.965	0.961
Firing Defect	0.762	0.643	0.950	0.782	0.689	0.884

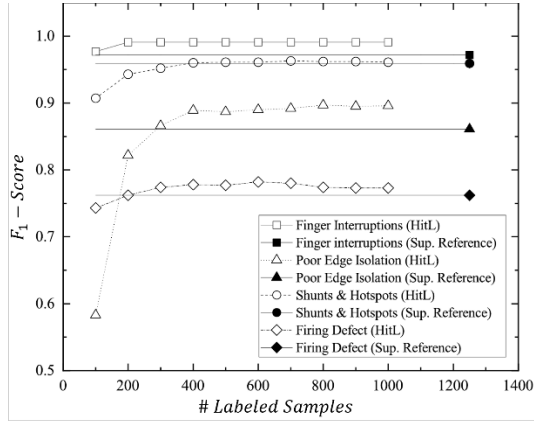


Figure 4: Comparison of F_1 - Scores of the supervised reference (Sup. Reference) and the human-in-the-loop approach with respect to different defect classes.

4.2 Results – Defect detection with Human-in-the-Loop

As a reference to the human-in-the-loop approach, four NNs were trained with respect to the defects finger interruptions, poor edge isolation, hotspots & shunts, and firing defect based on 1250 samples each within a 5-fold cross-validation. They can detect the defects well based on the EDTs. The detection results can be seen in Table 2 in the columns 2-4. In terms of F_1 - Score, the best detection results are finger interruptions (0.972), followed by hotspots & shunts (0.959), poor edge isolation (0.861) and firing defect (0.762).

The human-in-the-loop approach can achieve slightly improved results compared to the supervised reference while only needing few labeled samples. In Figure 4, the F_1 - Score is plotted against the number of samples required for the defect classes mentioned. The filled symbols represent the supervised reference (Sup. Reference) and have a horizontal auxiliary line. The empty symbols represent the human-in-the-loop (HitL) approach. Depending on the defect class, this approach achieves the same or slightly better F_1 - Score already with about 200 – 400 samples.

Figure 4 also shows that the F_1 - Score of the human-in-the-loop approach stabilizes at higher F_1 - Score. The corresponding values can be seen in Table 2: Detection results of the supervised reference and human-in-the-loop approach with respect to the process defects finger interruptions, poor edge isolation, hotspots & shunts and firing defect. in columns 5-7. On average, the F_1 - Score can be increased by about 2%.

5 DISCUSSION

It was shown that quality parameters can be derived from EL and IR images as well as reflectance values. The individual influence of the three inputs remains to be investigated. The empirical digital twins of the solar cells, derivable from the CNN, cluster into visible process defects. This indicates that the EDTs are meaningful in terms of the quality parameters and different defects. Thus, they are suitable to compare measurement images quantitatively. It is desirable to investigate the influence of the input data, the predicted quantities, and the size of the EDT in future studies.

The digital twin is suitable for deriving fast and efficient sorting criteria from process defects. The human-

in-the-loop method is an effective way to derive classification criteria. With little effort, expert knowledge can be incorporated to derive a classification scheme, enabling user-specific defect detection. Slightly better detection results compared to the reference can be achieved in our experiments. In this context, the label effort was reduced by a factor of 4-6. Due to the low effort, the method promises to detect self-defined defects and process properties. Due to the iterative approach, it can also be trained parallel to the production process.

In addition to the mentioned properties, the digital twin promises high transferability between different cell lines and concepts and could also be suitable for process optimization. Since the measurement images are correlated with measured quality variables, no time-consuming and error-prone human labels are necessary. Therefore, they can be quickly adapted to other cell lines. Furthermore, due to the quantitative representation of the sample with regard to process defects, it is also conceivable to utilize the EDTs for process optimization.

6 CONCLUSION

We introduced the empirical digital twin of the solar cell, which contains quality-describing features with respect to the electrical quality from any number of measured images. For this purpose, a deep neural network was trained to correlate high-dimensional image data with IV parameters, allowing the digital twin to represent features in terms of these quantities. As an example, we combine EL and IR images as well as reflectance measurements to predict in total 18 quality variables. The empirical digital twin is derived from the activations of the network in the form of a vector, which allows a quantitative comparison of one or more measurement images.

The digital twin is suitable for quality inspection. Using a human-in-the-loop approach, we show how expert knowledge can be efficiently incorporated into learning of a classification scheme with 4-6 times less label effort. It was demonstrated that the F_1 score of detecting the considered defect types could be increased by about 2% on average. As a result, F_1 scores of 0.99 for finger interruptions, 0.96 for hotspots and shunts, 0.90 for edge isolation problems, and 0.78 for inhomogeneous contact formation due to the firing process could be achieved. A detailed and comprehensive investigation of empirical digital twin with additional applications can be found in [44].

7 ACKNOWLEDGEMENTS

This work was funded by the German Federal Ministry for Economic Affairs and Energy within the project “NextTec” (03EE1001A). Philipp Kunze would like to thank the “Studienstiftung des deutschen Volkes” for support within the scope of his scholarship. Furthermore, many thanks to the PV-TEC Team at Fraunhofer ISE for support during the experiments.

References

- [1] K. Ramspeck, S. Schenk, D. Duphorn, A. Metz, M. Meixner, In-line thermography for reliable hot spot detection and process control, *Energy Procedia* 55 (2014) 133–140.
- [2] P.K. Kuo, T. Ahmed, H. Jin, R.L. Thomas, Phase-locked image acquisition in thermography, in:

- Automated Inspection and High-Speed Vision Architectures II, 1989, pp. 41–47.
- [3] T. Fuyuki, H. Kondo, Y. Kaji, T. Yamazaki, Y. Takahashi, Y. Uraoka, One shot mapping of minority carrier diffusion length in polycrystalline silicon solar cells using electroluminescence, in: Conference Record of the Thirty-first IEEE Photovoltaic Specialists Conference, 2005, 2005, pp. 1343–1345.
 - [4] D. Hinken, C. Schinke, S. Herlufsen, A. Schmidt, K. Bothe, R. Brendel, Experimental setup for camera-based measurements of electrically and optically stimulated luminescence of silicon solar cells and wafers, *Review of Scientific Instruments* 82 (2011) 33706.
 - [5] K. Bothe, P. Pohl, J. Schmidt, T. Weber, P. Altermatt, B. Fischer et al., Electroluminescence imaging as an in-line characterisation tool for solar cell production, in: 21st European Photovoltaic Solae Energy Conference, Dresden, 2006, pp. 597–600.
 - [6] C.M. Bishop, *Pattern recognition and machine learning*, Springer, 2006.
 - [7] M. Demant, T. Welschehold, M. Oswald, S. Bartsch, T. Brox, S. Schoenfelder et al., Microcracks in silicon wafers I: Inline detection and implications of crack morphology on wafer strength, *IEEE Journal of Photovoltaics* 6 (2015) 126–135.
 - [8] I. Zafirovska, M.K. Juhl, J.W. Weber, J. Wong, T. Trupke, Detection of finger interruptions in silicon solar cells using line scan photoluminescence imaging, *IEEE Journal of Photovoltaics* 7 (2017) 1496–1502.
 - [9] D. Stromer, A. Vetter, H.C. Oezkan, C. Probst, A. Maier, Enhanced crack segmentation (eCS): a reference algorithm for segmenting cracks in multicrystalline silicon solar cells, *IEEE Journal of Photovoltaics* 9 (2019) 752–758.
 - [10] S.A. Anwar, M.Z. Abdullah, Micro-crack detection of multicrystalline solar cells featuring shape analysis and support vector machines, in: 2012 IEEE International Conference on Control System, Computing and Engineering, 2012, pp. 143–148.
 - [11] D.-M. Tsai, S.-C. Wu, W.-C. Li, Defect detection of solar cells in electroluminescence images using Fourier image reconstruction, *Solar Energy Materials and Solar Cells* 99 (2012) 250–262.
 - [12] S. Spataru, P. Hacke, D. Sera, Automatic detection and evaluation of solar cell micro-cracks in electroluminescence images using matched filters, in: 2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC), 2016, pp. 1602–1607.
 - [13] B. Su, H. Chen, Y. Zhu, W. Liu, K. Liu, Classification of manufacturing defects in multicrystalline solar cells with novel feature descriptor, *IEEE Transactions on Instrumentation and Measurement* 68 (2019) 4675–4688.
 - [14] D.-C. Tseng, Y.-S. Liu, C.-M. Chou, Automatic finger interruption detection in electroluminescence images of multicrystalline solar cells, *Mathematical Problems in Engineering* 2015 (2015).
 - [15] A.M. Karimi, J.S. Fada, M.A. Hossain, S. Yang, T.J. Peshek, J.L. Braid et al., Automated pipeline for photovoltaic module electroluminescence image processing and degradation feature classification, *IEEE Journal of Photovoltaics* 9 (2019) 1324–1335.
 - [16] A.M. Karimi, J.S. Fada, J. Liu, J.L. Braid, M. Koyutürk, R.H. French, Feature extraction, supervised and unsupervised machine learning classification of PV cell electroluminescence images, in: 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC)(A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC), 2018, pp. 418–424.
 - [17] B. Su, H. Chen, Z. Zhou, BAF-Detector: An Efficient CNN-Based Detector for Photovoltaic Cell Defect Detection, *IEEE Transactions on Industrial Electronics* (2021).
 - [18] Y. Jiang, C. Zhao, W. Ding, L. Hong, Q. Shen, Attention M-net for Automatic Pixel-Level Micro-crack Detection of Photovoltaic Module Cells in Electroluminescence Images, in: 2020 IEEE 9th Data Driven Control and Learning Systems Conference (DDCLS), 2020, pp. 1415–1421.
 - [19] J. Balzategui, L. Eciolaza, N. Arana-Arexolaleiba, J. Altube, J.-P. Aguerre, I. Legarda-Ereño et al., Semi-automatic quality inspection of solar cell based on Convolutional Neural Networks, in: 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), 2019, pp. 529–535.
 - [20] B. Su, H. yong Chen, P. Chen, G.-B. Bian, W. Liu, others, Deep Learning-based Solar-Cell Manufacturing Defect Detection with Complementary Attention Network, *IEEE Transactions on Industrial Informatics* (2020).
 - [21] M. Mayr, M. Hoffmann, A. Maier, V. Christlein, Weakly Supervised Segmentation of Cracks on Solar Cells Using Normalized L p Norm, in: 2019 IEEE International Conference on Image Processing (ICIP), 2019, pp. 1885–1889.
 - [22] 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, *Progress in Photovoltaics: Research and Applications* 29 (2021) 471–484.
 - [23] E. Sovetkin, E.J. Achterberg, T. Weber, B.E. Pieters, Encoder-Decoder Semantic Segmentation Models for Electroluminescence Images of Thin-Film Photovoltaic Modules, *IEEE Journal of Photovoltaics* (2020).
 - [24] J. Balzategui, L. Eciolaza, N. Arana-Arexolaleiba, Defect detection on Polycrystalline solar cells using Electroluminescence and Fully Convolutional Neural Networks, in: 2020 IEEE/SICE International Symposium on System Integration (SII), 2020, pp. 949–953.
 - [25] M.R.U. Rahman, H. Chen, Defects inspection in polycrystalline solar cells electroluminescence images using deep learning, *IEEE Access* 8 (2020) 40547–40558.
 - [26] P. Kunze, J. Greulich, S. Rein, K. Ramspeck, M. Hemsendorf, A. Vetter et al., Efficient Deployment of Deep Neural Networks for Quality Inspection of Solar Cells Using Smart Labeling, in: Presented at the 37th European PV Solar Energy Conference and Exhibition, 2020 % detection but not with cnn, p. 11.
 - [27] S. Deitsch, C. Buerhop-Lutz, E. Sovetkin, A. Steland, A. Maier, F. Gallwitz et al., Segmentation of photovoltaic module cells in electroluminescence images, *arXiv preprint arXiv:1806.06530* (2018).
 - [28] M. Alt, S. Fischer, S. Schenk, S. Zimmermann, K. Ramspeck, M. Meixner, Electroluminescence

- imaging and automatic cell classification in mass production of silicon solar cells, in: 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC)(A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC), 2018, pp. 3298–3304.
- [29] R. Pierdicca, E.S. Malinverni, F. Piccinini, M. Paolanti, A. Felicetti, P. Zingaretti, Deep Convolutional Neural Network for Automatic Detection of Damaged Photovoltaic Cells, *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* 42 (2018).
- [30] A. Bartler, L. Mauch, B. Yang, M. Reuter, L. Stoicescu, Automated detection of solar cell defects with deep learning, in: 2018 26th European signal processing conference (EUSIPCO), 2018, pp. 2035–2039.
- [31] Z. Ying, M. Li, W. Tong, C. Haiyong, Automatic Detection of Photovoltaic Module Cells using Multi-Channel Convolutional Neural Network, in: 2018 Chinese Automation Congress (CAC), 2018, pp. 3571–3576.
- [32] S. Deitsch, V. Christlein, S. Berger, C. Buerhop-Lutz, A. Maier, F. Gallwitz et al., Automatic classification of defective photovoltaic module cells in electroluminescence images, *Solar Energy* 185 (2019) 455–468.
- [33] H. Chen, S. Wang, J. Xing, Detection of Cracks in Electroluminescence Images by Fusing Deep Learning and Structural Decoupling, in: 2019 Chinese Automation Congress (CAC), 2019, pp. 2565–2569.
- [34] C. Dunderdale, W. Brettenny, C. Clohessy, E.E. van Dyk, Photovoltaic defect classification through thermal infrared imaging using a machine learning approach, *Progress in Photovoltaics: Research and Applications* 28 (2020) 177–188.
- [35] A. Ahmad, Y. Jin, C. Zhu, I. Javed, A. Maqsood, M.W. Akram, Photovoltaic cell defect classification using convolutional neural network and support vector machine, *IET Renewable Power Generation* 14 (2020) 2693–2702.
- [36] A. Chindarkkar, S. Priyadarshi, N.S. Shiradkar, A. Kottantharayil, R. Velmurugan, Deep Learning Based Detection of Cracks in Electroluminescence Images of Fielded PV modules, in: 2020 47th IEEE Photovoltaic Specialists Conference (PVSC), 2020, pp. 1612–1616.
- [37] J.S. Fada, M.A. Hossain, J.L. Braid, S. Yang, T.J. Peshek, R.H. French, Electroluminescent image processing and cell degradation type classification via computer vision and statistical learning methodologies, in: 2017 IEEE 44th Photovoltaic Specialist Conference (PVSC), 2017, pp. 3456–3461.
- [38] M.W. Akram, G. Li, Y. Jin, X. Chen, C. Zhu, X. Zhao et al., CNN based automatic detection of photovoltaic cell defects in electroluminescence images, *Energy* 189 (2019) 116319.
- [39] M.Y. Demirci, N. Beşli, A. Gümüşçü, Defective PV Cell Detection Using Deep Transfer Learning and EL Imaging, *Proceedings Book* (2019) 311.
- [40] J.M. Greulich, M. Demant, P. Kunze, G. Dost, K. Ramspeck, A. Vetter et al., Comparison of Inline Crack Detection Systems for Multicrystalline Silicon Solar Cells, *IEEE Journal of Photovoltaics* 10 (2020) 1389–1395.
- [41] M. Demant, P. Virtue, A. Kovvali, X.Y. Stella, S. Rein, Learning quality rating of as-cut mc-si wafers via convolutional regression networks, *IEEE Journal of Photovoltaics* 9 (2019) 1064–1072.
- [42] M. Demant, P. Virtue, A. Kovvali, X.Y. Stella, S. Rein, Visualizing material quality and similarity of mc-Si wafers learned by convolutional regression networks, *IEEE Journal of Photovoltaics* 9 (2019) 1073–1080.
- [43] A.S. Kovvali, M. Demant, S. Rein, Spatially Resolved Material Quality Prediction Via Constrained Deep Learning, in: 2019 IEEE 46th Photovoltaic Specialists Conference (PVSC), 2019, pp. 3059–3062.
- [44] P. Kunze, S. Rein, M. Hemsendorf, K. Ramspeck, M. Demant, Learning an Empirical Digital Twin from Measurement Images for a Comprehensive Quality Inspection of Solar Cells, *Solar RRL*.
- [45] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, *The journal of machine learning research* 15 (2014) 1929–1958.
- [46] Y. Gal, Z. Ghahramani, Dropout as a bayesian approximation: Representing model uncertainty in deep learning, in: international conference on machine learning, 2016, pp. 1050–1059.
- [47] A. Kendall, Y. Gal, What uncertainties do we need in bayesian deep learning for computer vision?, *arXiv preprint arXiv:1703.04977* (2017).
- [48] G. Huang, Z. Liu, L. van der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.
- [49] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).
- [50] L. van der Maaten, G. Hinton, Visualizing data using t-SNE, *Journal of machine learning research* 9 (2008).

APPENDIX

If a defect is correctly predicted, it is called true positive (TP), if a defect is incorrectly predicted, it is called false positive (FP). If a non-defect is correctly predicted, it is called true negative (TN), if a non-defect is incorrectly predicted, it is called false negative (FN). Some quantities can be derived from this. The *Precision* is defined in Equation (3) and indicates how many defect predictions were defects.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

The *Recall*, defined in Equation (4), is a measure of how many defects were found of all defects.

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

The F_1 - *Score*, defined in Equation (5), is the harmonic mean of Precision and Recall.

$$F_1 - Score = 2 \frac{Precision \cdot Recall}{Precision + Recall} \quad (4)$$