Deep Ensemble Inverse Model for Image-Based Estimation of Solar Cell Parameters

M. Battaglia,[1] E. Comi,[1] T. Stadelmann,[1,2] R. Hiestand,[3] B. Ruhstaller,[3,1] and E. Knapp[1]
[1)]*Zürich University of Applied Sciences, Wildbachstrasse 21, 8400 Winterthur, Switzerland*
[2)]*ECLT European Centre for Living Technology (Fellow), Dorsoduro 3911, 30123 Venice, Italy*
[3)]*Fluxim AG, Katharina-Sulzer-Platz 2, 8400 Winterthur, Switzerland*

(*Electronic mail: mattia.battaglia@zhaw.ch)

(Dated: 20 July 2023)

Physical models can help improve solar cell efficiency during the design phase as well as for quality control after the fabrication process. We present a data-driven approach to inverse modeling that can predict the underlying parameters of a finite element method (FEM) solar cell model based on an electroluminescence (EL) image of a solar cell with known cell geometry and laser scribed defects. For training the inverse model, 75,000 synthetic EL images were generated with randomized parameters of the physical cell model. We combine 17 deep convolutional neural networks (CNNs) based on a modified VGG19 architecture into a deep ensemble to add uncertainty estimates. Using the silicon solar cell model, we show that such a novel approach to data-driven statistical inverse modeling can help apply recent developments in deep learning to new engineering applications that require real-time parameterizations of physical models augmented by confidence intervals. The trained network was tested on four different physical solar cell samples and the estimated parameters were used to create the corresponding model representations. Resimulations of the measurements yielded relative deviations of the calculated and the measured junction voltage values of 0.2 % on average with a maximum of 10 %, demonstrating the validity of the approach.

I. INTRODUCTION

The transition of the world energy system to renewable, fossil-free sources relies heavily on photovoltaic electricity[1]. Due to the high energy consumption in the manufacturing process, the energy payback time of monocrystalline solar modules is in the range of 10% of the total lifetime[2]. Therefore, further improvements in cell design and cell production are desirable.

One possible way to shorten the energy payback time of solar cells is to reduce power conversion losses due to electrical resistance in the solar cell and due to cell defects[3,4]. To investigate these losses in detail, various measurement techniques such as electroluminescence (EL), photoluminescence (PL), and infrared (IR) imaging have been established both during the design process in the laboratory and for quality assurance integrated in the production line[5,6].

By combining the imaging techniques with physical modeling, it is possible to gain detailed insight into the impact of specific design decisions and defects on expected module performance. To do so, the defining parameters of the cell have to be extracted from measurements. It has been shown that the characteristics of solar cells can be modelled with Finite Element Method (FEM) models that solve an equivalent circuit representation that is defined by spatially distributed resistances such as the sheet resistance of the electrodes, contact resistivity, parallel shunt resistivity and spatially distributed diode model parameters[7].

The presented work is part of a larger effort to use model calculations of solar cells to characterize defects, quantify their effects on the solar cell and predict the impact of defects and design decision during upscaling to larger cell areas and modules[8]. Standard silicon cells are used in this work because of better reproducibility and ease of use in measurements. However, the method is intended to be transferred to novel cell technologies such as perovskite cells, where the origin of the defects is less well understood and major challenges exist regarding inhomogeneity, upscaling, and stability[9,10].

Parameterization of the numerical model of a physical system can be challenging if its parameters are not directly measurable. Finding an appropriate set of model parameters that will enable the model to reproduce important features of the system must often be done indirectly, by finding a set of parameters that lead to model results which agree with the available measurements. Solving such an inverse problem is often a non-trivial, ill-posed task. In research and laboratory work, model parameters are often manually fitted to match model predictions with measurements. This process can be automated by minimizing the model error with least squares optimization techniques[11]. If statistical analysis is required, the problem can be reformulated in a Bayesian framework and Markov Chain Monte Carlo (MCMC) algorithms can be used to sample from the posterior distribution of the input parameters[12].

For many foreseeable applications of physical models of solar cells, such as quality control in a production line, these traditional approaches are insufficient due to their high computation time. This is particularly true in the use case presented here where two- or three-dimensional models are needed to describe features of the measured solar cell images. Therefore, an explicit form of an inverse model that can replace these steps by directly computing the model parameters based on measured data would be beneficial[13]. Deep neural networks (DNNs) are promising candidates to serve as such inverse models, as they are able to adapt arbitrary nonlinear mappings when trained on a sufficiently large training dataset[14–17].

The use of a neural network as an inverse model has been successfully demonstrated by training an auto encoder surrogate model to recover the material parameters from simulated current-voltage (IV) curves[18]. A hybrid approach in which part of the model parameters are predicted by a CNN while the others are estimated with traditional optimization techniques has showed promising results for determining layer properties based on reflectance spectroscopy[19]. In addition, it has been demonstrated that CNNs can be used to predict the cell efficiency or the location of defects based on electroluminescence images[20–22]. By combining EL, PL and reflectance spectroscopy, CNNs have been used to predict the full IV curve of a cell together with other key characteristics of the cell[23,24]. In a further work, a spatially resolved approach based on a U-net has been used to determine the local dark saturation current[25].

In this work, we propose an approach using a convolutional neural network (CNN) as an inverse model that computes cell parameters from EL measurements of silicon solar cells. The CNN is trained on a training set of synthetic EL images simulated with the numerical model that it aims to invert[26]. The simulated training set has a predefined cell geometry. Shunt defects with high parallel conductivity are added to the cell geometry by placing standardized rectangular subdomains. The numerical model is a 2D+1D FEM model that calculates the 2D potential distribution of the cell's top and bottom electrodes based on domain specific electrode properties and coupling laws. The model is implemented in the simulation software Laoss 4.0 and distributed by Fluxim AG[27,28]. In order to provide an uncertainty estimate of the calculated model parameters, we use a deep ensemble CNN model that is able to predict a Gaussian approximation of the model parameter probability distribution[29]. The CNN model can then be applied to data acquired in a EL measurement setup to estimate the underlying physical model parameters of the solar cells. If successfully applied, this method enables the development and training of a CNN that can almost instantly create a model representation of the presented measurement sample.

The proposed method demonstrates a novel implementation of recent developments in machine learning that could extend existing engineering applications of deep learning for industrial practice[30,31]. Bridging the gap between data-driven and physical models raises new challenges that are less frequently discussed in the deep learning literature: First, CNN network architectures are designed and implemented for classification tasks in the majority of cases. Dealing with physical parameters requires the use of multivariate regression models, which increases the complexity of the training. Second, in science and engineering, there is often a need to provide an uncertainty measure to evaluate the confidence in a result. Common deep neural network architectures are designed to provide point predictions. The use of deep learning in the context of physical models therefore makes it necessary to test and exploit the potential of recent developments in network architectures that provide uncertainty estimates[32]. Our work provides a comprehensive example of a engineering application of a regression CNN that incorporates such an uncertainty estimation. The methodology could be applied to several other regression tasks in the surging area of physics-based deep learning where either sufficient training data or a detailed numerical model is available[33].

The contributions of this work can be summarized as follows. We train deep ensemble CNNs to estimate physical parameters based on PV cell imaging data. The training data is generated by a physical simulation model whose inverse is to be approximated by the CNN. To our knowledge, such a combination of deep ensemble CNNs and inverse modeling is a novel combination of already successful concepts in science and engineering. It is shown that the presented method can be used to estimate physical model parameters without relying on specific symmetries in the layout of the grid lines and bus bars that form the front contact of the PV cell. Therefore, our approach can be applied more generally to different PV cell types. We show that the extracted parameter sets can be used to parameterize the model and reproduce the measured images with high precision. We discuss the changes in the CNN structure and training hyperparameters required to implement and train such a deep ensemble CNN inverse model. In doing so, we provide further evidence that standard implementations of deep neural networks can be adapted for use in a scientific setting with minor modifications.

A schematic of the proposed workflow is shown in Figure 1. The steps of the developed method are described in this paper as follows: The measurement samples are described in section II A. The geometry of this cell sample was extracted from an EL measurement and used as the basis for the inverse model. The Laoss simulation model is explained in section II B. Section II C shows how Laoss is used to model the measured solar cell samples. Additional post-processing steps applied to the Laoss simulation results are explained in section II D. The structure of the inverse model is defined in section II E. The inverse model is trained on a set of simulated EL images. The Laoss parameterization used to generate the training data is discussed in section II F. The network architecture and training of the inverse CNN model are explained in section II G. The results of the inverse model method are evaluated in section III. A discussion and a brief outlook are given in sections IV and V.

II. MATERIAL AND METHODS

A. Measurement Sample

The proposed method was carried out and tested for a monocrystalline solar cell of type XS156B3-200R from Motech Industries[34]. Due to electrical current strength limitations in the available measurement equipment, a laser cutter was used to cut a 2 x 2 cm area from the wafer cell with a busbar at the top. The smaller area of the cell sample also reduced the computation time of the FEM model. An additional laser cut was used to introduce an artificial shunt between two grid lines. Two contact strips were then soldered to the busbar and rear solder pads of the sample cell.

The electroluminescence signal of the sample cell was recorded with a Nikon D800 digital SLR camera. The calibration constant for relating the EL signal with the junction

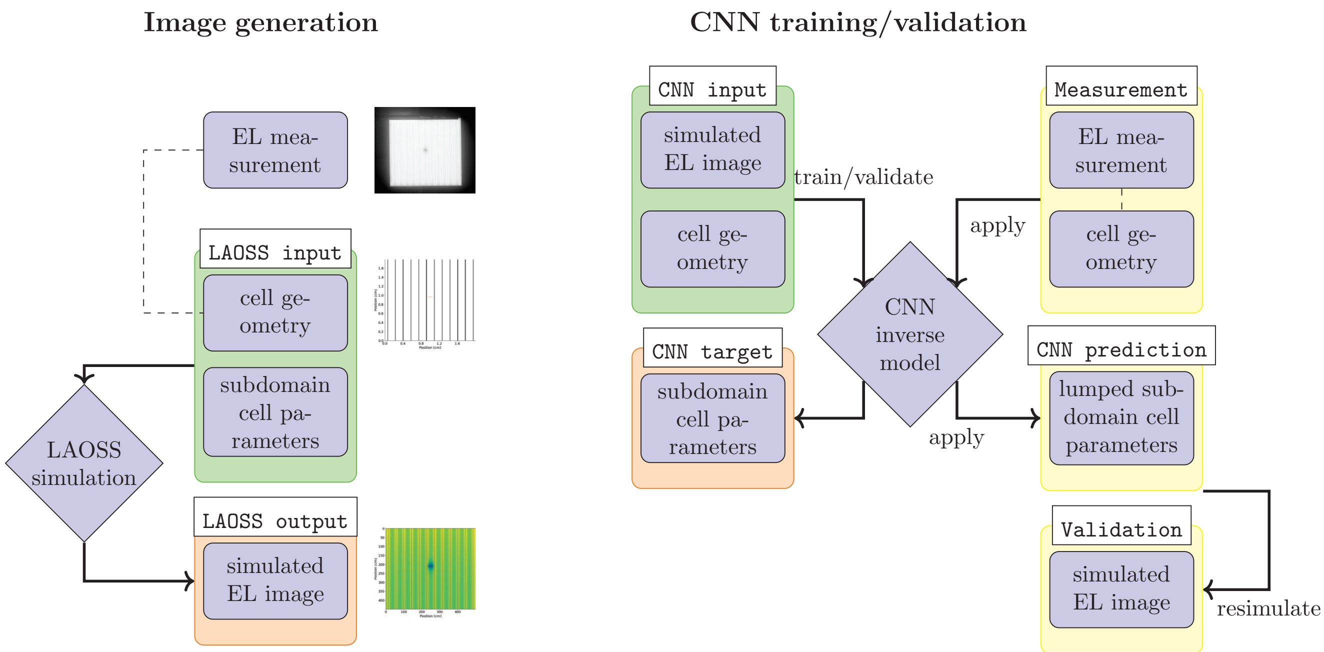

FIG. 1: Workflow of the proposed inverse CNN model training.

voltage was determined with a low forward voltage of 0.55 V assuming negligible current and a therefore a constant voltage across the whole image[35]. The high-voltage EL image was acquired with a forward voltage in the range of 0.60 V-0.64 V. To increase the signal-to-noise ratio of the measurement, the camera exposure time was set to 450 seconds at an ISO level of 200. The measurement image was exported in NEF (RAW) format and converted to grayscale. The resulting measurement image is shown in Figure 2a.

The measured image shows dark areas at the edges of the cell that are not present in the simulation. It is assumed that the edge effects are caused by defects in the junction as a result of the laser cutting process. Since edge effects were not included in the simulation model, this nonsimilarity was removed by cutting an area in the center that is not affected by the edge effects as the input to the inverse model (see Figure 2(a)). Since the calibration constant was calculated based on the averaged luminance signal with the assumption of constant junction voltage in the low-voltage image, the edge effects will decrease the average low-voltage signal and therefore lead to a overestimation of the junction voltage in the area used as the model input.

B. FEM model

The simulation model of Laoss uses a 2D+1D concept, also known as the "interconnected diode model"[28,36–38]. In this modelling approach, the potentials of the top (v^t) and bottom (v^b) electrodes of the solar cell are calculated using the Finite Element Method (FEM) to solve Ohm's law for the given sheet resistance $R_\square$. The current continuity formula for the top electrode is given in Equation 1. The formula for the bottom electrode differs in an additional minus sign on the right hand side.

$$\frac{1}{R^t_\square}\boldsymbol{\Delta} v^t = j(v^t, v^b). \qquad (1)$$

The source current density $j(v^t, v^b)$ in Equation 1 is dependent on the 2D potential distribution and calculated from a user-provided coupling law that represents the diode properties of the solar cell's p-n junction. In this work, the one-diode model is used, which includes both a parallel (ρ_{par}) and an internal series (ρ_{int}) area specific resistivity:

$$j(v^t, v^b) = j(\Delta v) = j_0 \left[exp\left(\frac{\Delta v - j(\Delta v)\rho_{int}}{V_T} \right) - 1 \right] + \frac{\Delta v - j(\Delta v)\rho_{int}}{\rho_{par}}. \qquad (2)$$

The additional parameters used in Equation (2) are the dark saturation current j_0 and the thermal voltage V_T. A schematic representation of the model used in Laoss as well as the input parameters required and the output parameters provided are shown in Fig. 3.

Alternatively, the algebraic diode model can also be replaced by a numerical charge drift-diffusion model considering the entire vertical cell structure (Fluxim's drift-diffusion simulation software Setfos integrates with the 2D+1D model in Laoss). The advantage of the algebraic diode model, however, is the small number of free model parameters to define the coupling law.

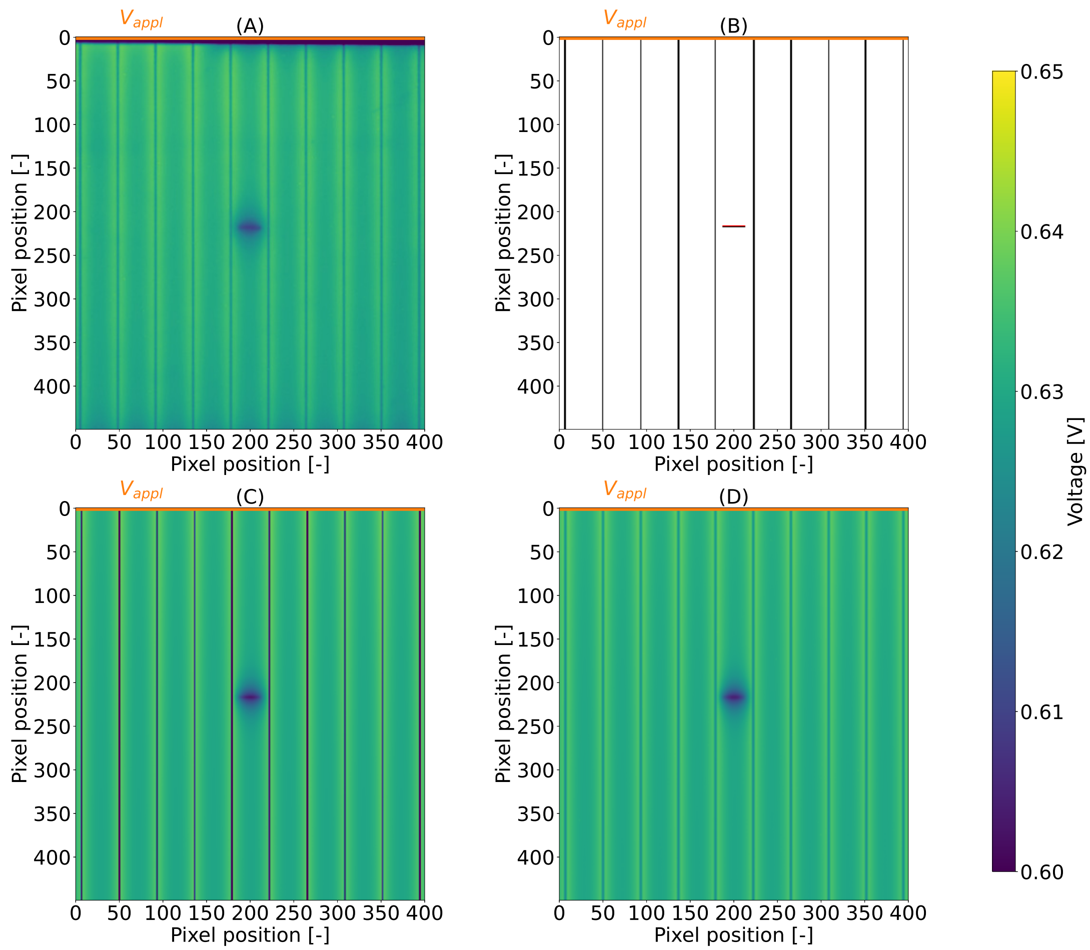

FIG. 2: (A) EL measurement of a sample cell. (B) Cell geometry of the sample cell used for simulation. (withe: active area, black: grid lines, red: shunt) (C) Unprocessed simulated EL image. (D) Post-processed simulated EL image.

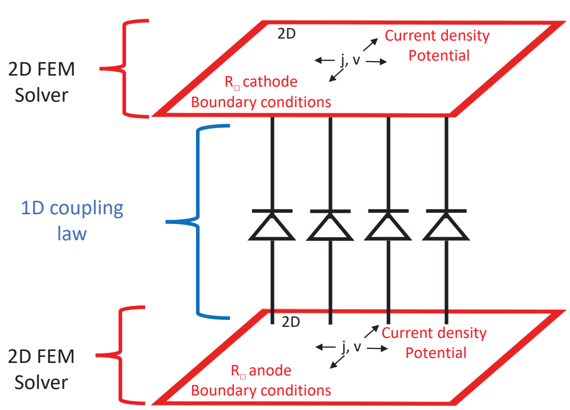

FIG. 3: Structure of the Laoss FEM solver[39].

C. Model representation of the measurement sample

The bias voltage applied to the solar cell during an EL experiment can be represented in Laoss as a fixed potential boundary condition. In the experiment, we applied the forward voltage at the bus bar which is visible as a dark area at the top edge of the measurement image and highlighted by an orange line in Figure 2(A). The corresponding position of the fixed potential boundary condition in the model representation with value V_{appl} is shown analogously in Figure 2(B)-(D). The busbar was not included in the simulation geometry assuming that the fixed potential boundary condition represents the intersection between the busbar and the grid lines as well as the active area on the top electrode. 75’000 parameter/image pairs have been generated to build the training data of the inverse model. For each parameter set, two images were generated at different bias voltages to account for the calibration procedure used in the measurement. The simulated images were generated at random bias voltages. The lower voltage was sampled from a uniform distribution in the range [0.54 V, 0.55 V] and the higher bias voltage was sampled in the range [0.60 V, 0.64 V].

A coupling law with a non-zero value for the internal resistivity ρ_{int} is not yet implemented in Laoss 4. Therefore, Equation 2 was solved externally with the hybrid algorithm of MINPACK and passed to Laoss as discrete values in a text file. The other values of the diode law were ideality factor $n_{id} = 1$, thermal voltage $V_T = 2.38\mathrm{e}{-2}\,\mathrm{V}$ and internal resistivity $\rho_{int} = 2.88\mathrm{e}{-4}\,\Omega\,\mathrm{m}^2$. These values remained unchanged during the simulation.

Laoss exports the potential and current results on a rectangular grid which can then be converted to a raster image. In the following, the spatial position of the simulated and measured values are denoted by $[x,y]$. The simulated luminance value has been calculated according to the following formula[35]:

$$S_{EL}[x,y] = exp\left(\frac{v_{sim}^{junc}[x,y]}{n_{id}V_T}\right). \tag{3}$$

The junction voltage is calculated from the voltage difference between v_{top} and v_{bottom} provided by the Laoss results.

An unprocessed image of $S_{EL}(x,y)$ generated with parameters set manually to reproduce the measured sample is shown in Figure 2(c). The silver grid is assumed to completely block fluorescent light and therefore appears as sharp dark lines.

D. Simulation Post Processing

The simulated image and the camera image of the measurement sample were brought into closer agreement in an additional post-processing step. EL images exhibit inherent blurring due to lateral carrier diffusion in the emitting silicon cell[40], photon scattering in the silicon CCD caused by absorption depths that are longer than the pixel size[41,42], and metal finger scattering[43]. Therefore, the simulated image was convoluted with a Gaussian kernel of size 5 pixels to match the blur level of the EL measurement. In a second step, 148 patches of dark areas were cropped from the camera image. These were then scaled to the range of the pixel values of the simulated image, randomly rotated and flipped, and overlaid with the simulated image to imitate the precise camera noise of the setup that consists of a combination of Gaussian noise, shot noise and salt-and-pepper noise. The post-processed version of the simulation image example is shown in Figure 2(d). A comparison with the unprocessed image in Figure 2(c) shows the effects of blurring and noise, resulting in features similar to the measured image in Figure 2(a).

E. CNN inverse model

The Laoss modelling approach to calculate the simulated junction voltage $v_{sim}^{junc}[x,y]$ described in Section II B can be formalized as following:

$$\begin{aligned} v_{sim}^{junc}[x,y] = \\ \sum_{l=-k}^{k}\sum_{m=-k}^{k} F(V_{appl}, R_{\square}, j_0, 1/\rho_{par})[x-l, y-m]\cdot h[l,m] \\ +\varepsilon[x,y]. \end{aligned} \tag{4}$$

The function F represents the Laoss simulation model, h is the convolution kernel of the artificial camera blur and ε the artifial camera noise added. Given a measurement $v_{meas}[x,y]$, the goal of parametrizing the Laoss model to create the best possible model representation of the measured cell results in the following minimization problem:

$$res = \min_{V_{appl}, R_{\square}, j_0, 1/\rho_{par}} \sum_x\sum_y (v_{sim}^{junc}[x,y] - v_{meas}^{junc}[x,y])^2. \tag{5}$$

In this work, we attempt to avoid this optimization problem by replacing it with an inverse model. The inverse model F^{-1} of the forward calculation

$$F\left(V_{appl}, R_{\square}, j_0, 1/\rho_{par}\right) = v_{sim/meas}^{junc}[x,y] \tag{6}$$

is a function that directly maps measurement outcomes to the model parameter space:

$$F^{-1}(v_{sim/meas}^{junc}[x,y]) = (V_{appl}, R_{\square}, j_0, 1/\rho_{par}). \tag{7}$$

To construct a suitable inverse function F^{-1}, we follow a data-driven approach based on training a regression CNN. The Laoss forward model is used to create training data that converts the model parameter ranges of interest. In the following, the CNN will be trained to recover the model input parameters for each simulated EL image. A conventional way of CNN training is to perform gradient descent steps based on a mean squared error loss function averaged over N training samples:

$$\begin{aligned} loss &= \frac{1}{N}\sum_{i=1}^{N} \sum_{\substack{\alpha\in\{V_{appl}, R_{\square}, \\ j_0, 1/\rho_{par}\}}} (\alpha_i^{pred} - \alpha_i)^2 \\ &= \frac{1}{N}\sum_{i=1}^{N} \sum_{\alpha\in V_{appl}, R_{\square}, j_0, 1/\rho_{par}} (F_{\alpha}^{-1}(v_{sim}^{junc}[x,y]_i) - \alpha_i)^2. \end{aligned} \tag{8}$$

Here, we follow an approach based on deep ensembles that extends the point prediction of neural networks with an uncertainty measure. This is accomplished predicting both a mean μ and a variance σ^2 value for each model parameter which replaces the inverse function as follows:

$$\begin{aligned} F^{-1*}(v_{sim/meas}^{junc}[x,y]) = \\ (\mu_{V_{appl}}, \mu_{R_{\square}}, \mu_{j_0}, \mu_{1/\rho_{par}}, \sigma^2_{V_{appl}}, \sigma^2_{R_{\square}}, \sigma^2_{j_0}, \sigma^2_{1/\rho_{par}}). \end{aligned} \tag{9}$$

The variance prediction of the neural net is then included in the training by using the negative log-likelihood loss function:

$$loss^* = \frac{1}{N}\sum_{i=1}^{N} \sum_{\alpha\in V_{appl}, R_{\square}, j_0, 1/\rho_{par}} \frac{1}{2log(\sigma^2_{\alpha,i}))} \frac{(\mu_{\alpha,i}^{pred} - \alpha_i)^2}{2\sigma^2_{\alpha,i}} \tag{10}$$

The predicted parameter pairs $\{\alpha, var(\alpha)\}$ are assumed to define a Gaussian probability distribution[44]. By minimizing the negative log-likelihood loss function, the network tries to find the predicted distribution in which the probability of the real model parameter α is maximized. Thus, if insufficient correlations are found between simulated data $v_{sim}[x,y]$ and a model parameter α, the network has the freedom to reduce the loss function value by increasing the predicted variance σ_α^2.

The deep ensemble prediction is then completed by training M networks in parallel using different splits for the training and validation data and different initial weights for the CNN. The predictions of the different CNNs are assumed to form a Gaussian mixture. For simplicity, the mean and the variance of this Gaussian mixture are then used to present the results and for further computation:

$$\mu_\alpha^* = \frac{1}{M}\sum_{i=1}^{M}\mu_{\alpha_i} \tag{11}$$

$$\sigma_\alpha^2 = \frac{1}{M}\sum_{i=1}^{M}(\sigma_{\alpha_i}^2 + \mu_\alpha^2) - \mu_\alpha^2 \tag{12}$$

F. Training and Validation Data Generation

The simulations for generating the training data were performed using Laoss 4[39]. The used geometry has three subdomains (active area, grid and shunt) and is shown in Fig. 2(b). The input geometry was manually adjusted to match the dimensions and gridline structure of the sample cell. In the final workflow, this step could be replaced by an algorithm that assists in extracting the cell geometry from the EL image using edge detection and morphological operations. A random number of up to 4 shunts was placed on the artificial images. The position of the shunt was randomly chosen but was constrained by the following conditions:

1. The shunts cannot intersect with a grid line.
2. The shunts cannot intersect each other.
3. All shunts have the same dimensions and orientation (height=0.01 mm, width=1 mm). This corresponds to the assumed shape of the laser shunt scribed into the measurement sample.

A total of 75,000 images were simulated and used for training the inverse CNN model. In each of the three subdomains, four free parameters are chosen to be predicted later by the CNN model and therefore varied during the simulation of the training images. An overview of the parameters can be found in Table I. Some of the parameters are sampled from a probability distribution. The others, such as the parallel resistivity ρ_{par} of the active area, are kept constant at a physically meaningful value. The values of ρ_{par} were independently sampled for shunts on the same simulated cell, resulting in shunts with different intensities that the neural network should learn to distinguish.

G. Network Setup and Training

The network was implemented using the Keras/Tensorflow framework. The network architecture was adapted from the VGG19 implementation of Keras[45]. All modifications follow recommendations given for transferring CNN networks from classification to regression problems[46]. In all hidden layers, ReLu activation functions are used. Dropout layers have been included after the two dense layers at the top of the network and a batch normalization layer was included in front of the last dropout layer. The top layer was removed and replaced by a dense layer consisting of 8 output neurons with a linear activation function to obtain a regression network for the 4 target parameters' (μ,σ^2) pairs.

The input image presented to the CNN was built from the cropped, post-processed simulation images. In order to reduce the memory consumption during training, the images were downsampled to a resolution of 80x40 pixels. Since every subdomain of a simulated EL image has its own distinct model parameters, we designed the CNN such that it predicts the values for one subdomain at the time. The first channel of the input contains the complete voltage image. The second channel encodes the mask in which the pixels defining the subregion for which the parameters should be predicted are set to 1 and the pixels of areas that should be ignored are set to 0. Since a standard implementation of VGG19 was used, a third channel was present but remained empty in all images to keep the original architecture and dependent hyperparameter ranges intact. Figure 4 shows the structure of the input and the output data for the example image of Figure 2(b) which contains a single shunt. The approach results in 3 different two channel images as input to the CNN model with identical values in the first voltage image channel and different masks in the second channel. The parameters of the subdomains can then be collected and used to build a complete simulation model. The CNN model provides all four parameters independently from the subdomain defined in the mask. For shunt subdomains, V_{appl} has no direct physical definition since the fixed voltage boundary condition is only applied to the upper edge of the grid lines and the active area. Therefore, the parameter V_{appl}^{shunt} can be omitted when constructing the simulation model based on the predicted parameters and is put in brackets in Figure 4. The training set consisted of 94 % of all available images, while the validation set contained the remaining 6 %. Target values were scaled to a feature range of 0.01 to 0.99 before training. A batch size of 512 and the Adam solver were used. Training was performed over a total of 200 epochs with an early stopping as soon as the validation loss did not improve for 20 epochs. The negative loss likelihood loss function defined in Equation 10 can lead to numerical instabilities during network training when intermediate predictions of the variance σ^2 are zero or close to zero. To avoid this, the activation of the output neurons for the variance σ^2 was set to a strictly positive *ELU+1* function with $alpha = 1$. In addition, the values for the variance σ^2 have been clipped to the range $[0.0001, 1]$ in the calculation of the loss function and gradient clipping with a value of 0.5 was used in the Adam solver.

TABLE I: Model parameters used for the simulation of the training data

Subdomain	V_{appl} [V]	j_0 [A/m^2]	$1/\rho_{par}$ [S/m^2]	$R_\square$ [$\Omega/\square$]
active area	uniform $[0.6, 0.7]$	log-uniform $[1e{-}10, 1e{-}8]$	constant (50)	uniform $[10, 120]$
grid line	uniform $[0.6, 0.7]$	log-uniform $[1e{-}10, 1e{-}8]$	constant (50)	log-uniform $[1e{-}4, 1e{-}2]$
shunts	not defined	log-uniform $[1e{-}10, 1e{-}8]$	log-uniform $[1e3, 2e6]$	constant (10)

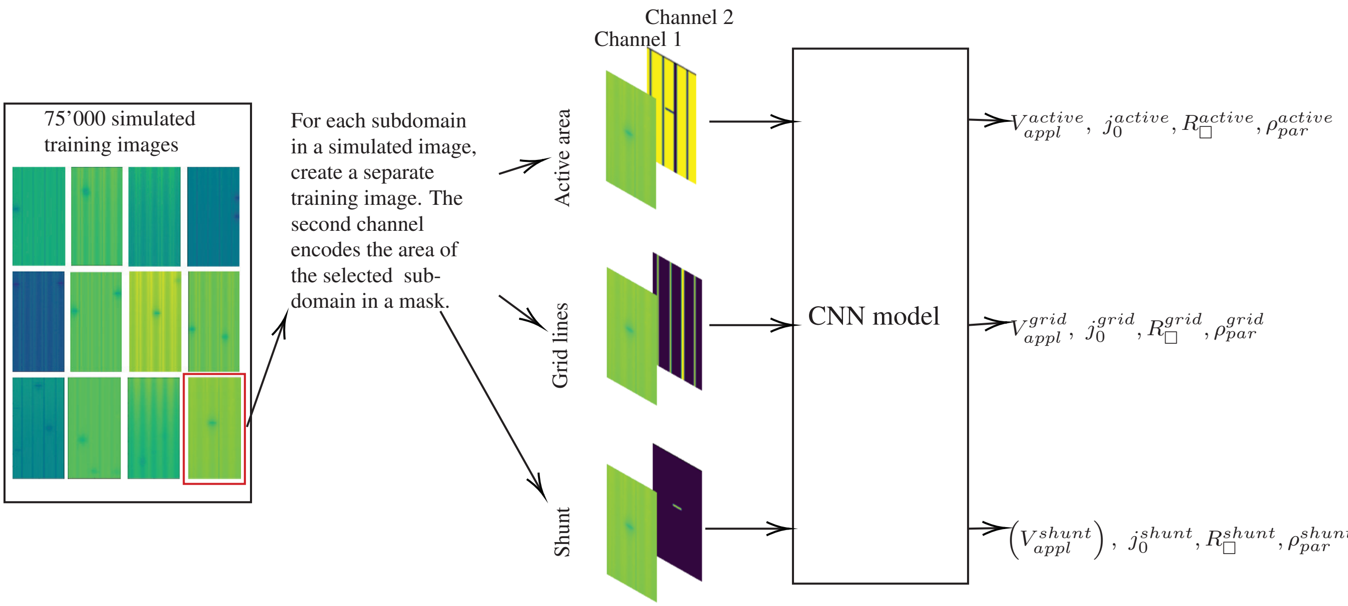

FIG. 4: Structure of the CNN input and output data.

III. RESULTS

Four different measured samples (MO, M1, B1, T2) were used to test the proposed approach. For all samples, the previously trained deep ensemble was used to predict a model parameterization of the FEM simulation model. The quality of this inverse model was then tested by comparing the forward simulation based on these parameters with the original measured data. In addition to the measured test data, a simulated image (MO-sim) included in the validation set during CNN training was used to analyze the extent to which the performance of the approach degrades when real instead of synthetic data is used as input.

The results of the regression output of a single CNN model from the deep ensemble are shown in Figure 5. The x-axis value respresents the value used during the simulation of the EL image. The points on the y-axis show the mean of the predicted Gaussian probability distribution of the parameter. The error bars show the standard deviation calculated from the predicted variance. The results show that the CNN learns to predict V_{appl} from the simulated EL image with very high accuracy, which is correctly represented by corresponding low variance predictions. The predictions of the parameters ρ_{par} and $R_\square$ show higher uncertainty values, which also corresponds to larger offsets between predicted and true values. The largest uncertainties with respect to the defined parameter range are found in the predictions of j_0. Interestingly, the model seems to correctly identify values with large offsets by predicting high variance values in these cases.

The trained deep ensemble consists of 17 CNNs. The ensemble results of the simulated validation image MO-sim are shown in Figure 6. In the case of the simulated validation image, the original parameters used to simulate the image are known and can be compared to the predictions. In general, the mean predictions of each network vary significantly and so the predicted confidence intervals do not necessarily overlap. The true parameter used for the simulation of MO-sim is within the confidence interval of the predicted ensemble distribution in the case of V_{appl}, $R_\square$ of grid lines and ρ_{par} of the laser cur shunt area. The predicted probability distributions of j_0 and $R_\square$ parameters of the active area deviate significantly from the actual simulation parameters, indicating that the EL image does not provide sufficient information to determine the two parameters.

The full parameter prediction results for all four measurement samples and the simulated validation image are shown in Figure 9. The validation images and corresponding resimulations using the mean values of the distribution predicted by

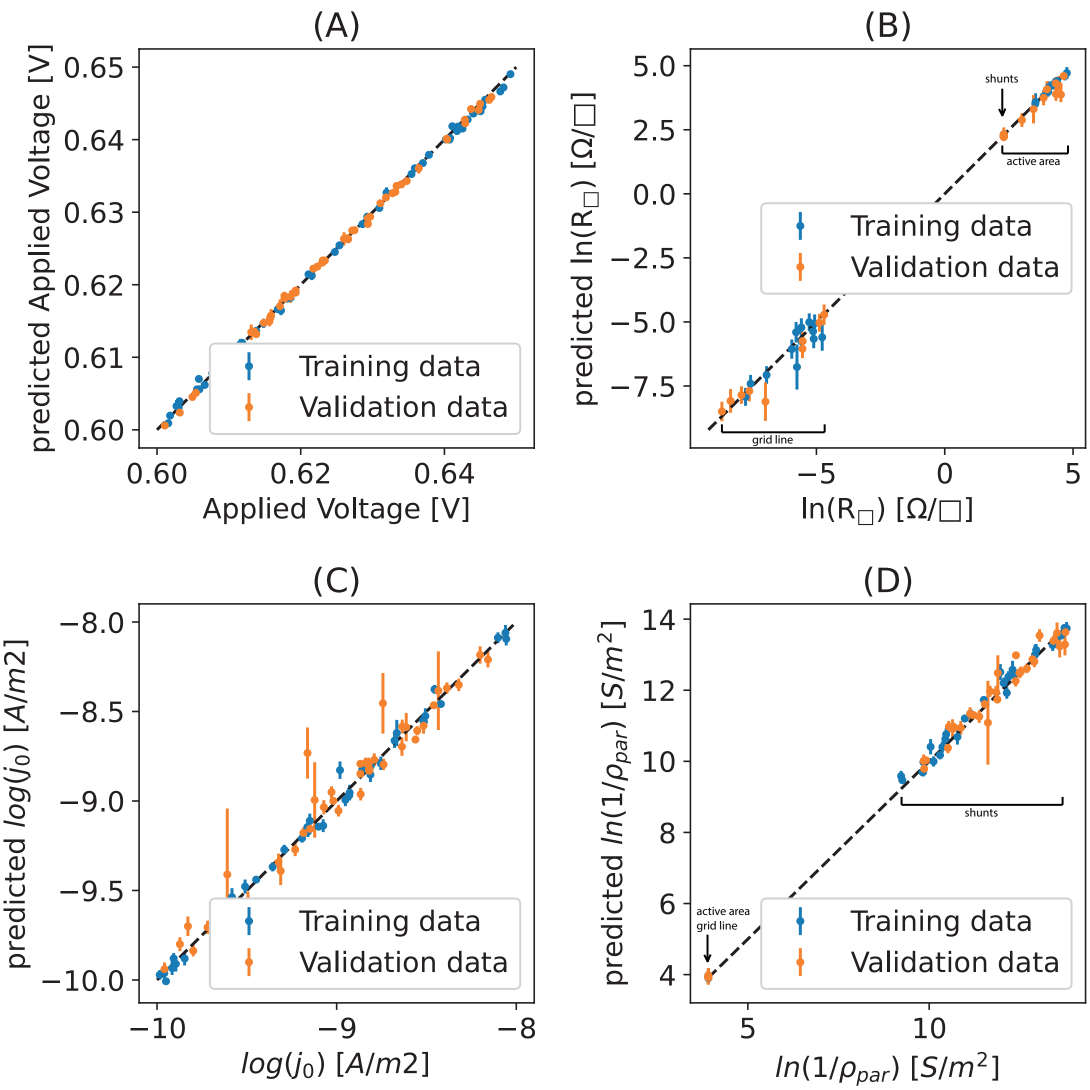

FIG. 5: Regression results of a single instance inverse CNN model's predictions for the applied voltage (A), the sheet resistance (B), the dark saturation current (C), and the parallel resistivity (D) including predicted uncertainties. The plot shows the results for 100 randomly selected images for both the validation set and the training set. The value ranges of the different subdomains are highlighted in (C) and (D). The sheet resistance of the shunts and the parallel resistivity of the active area as well as the grid line were constant in all images. Therefore, all data points of the two sets are superimposed.

the deep ensemble are shown in Figure 7 for (MO-sim, MO and M1). The part of the image that was used as the input of the CNN is highlighted with a red rectangle. A general visual inspection of the images shows good agreement in terms of EL intensity and voltage drops due to the grid line layout. The intensity of the resimulated shunt resembles the measurement closely in the case of the simulated test image. For the measured images, the predicted distribution's mean values for ρ_{par} of the laser cut area result in visually more pronounced voltage drops. The average error in the junction voltage value when comparing the resimulated image with the measured image is 0.2 %. The maximum error that can be found in both images is 10 %.

Horizontal and vertical cross sections of the test cells are shown in Figure 8. For each pixel, the standard deviation has been calculated by simulating 24 images with parameters that have been sampled from the parameter distributions predicted by the deep ensemble. The values of the cross section confirm the impression that the applied voltage range and voltage drops between grid lines are correctly modelled by the mean values of the predicted distributions. The deep ensemble tends to overestimate $1/\rho_{par}$ in the laser cut region. Additionally, the parallel resistance of the cell samples M1, B1, and T2 have been calculated by fitting a lumped-parameter equivalent circuit diode model to the cell's current-voltage curve. The global parallel resistance has then been multiplied with the total shunt area to estimate the area specific resistivity of the shunts. In the case of M1, it is assumed that the two shunts contribute equally to the measured global parallel resistance. The results shown in Figure 9 confirm that the CNN model

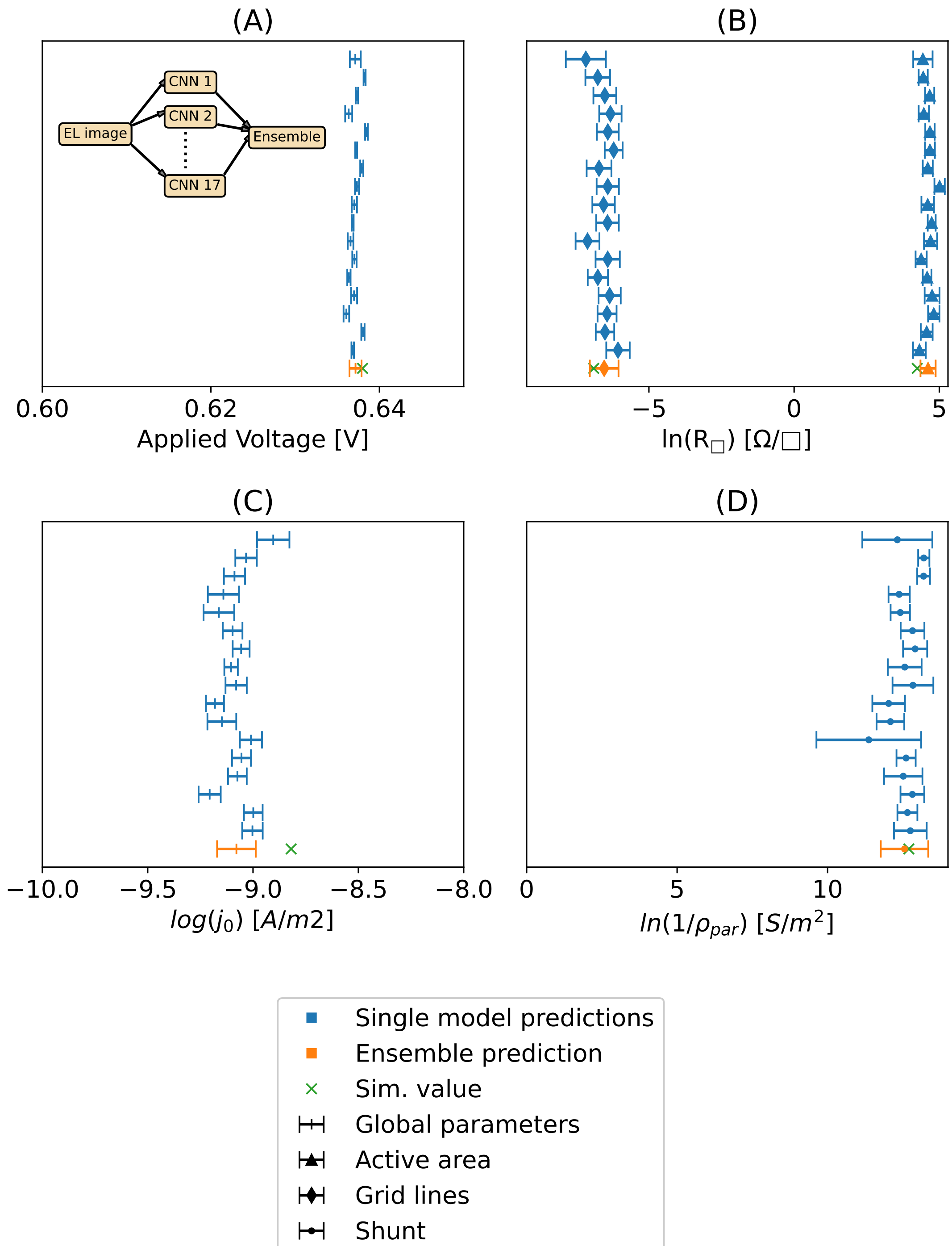

FIG. 6: Model predictions for the applied voltage (A), the sheet resistance (B), the dark saturation current (C), and the parallel resistivity (D) for the validation image MO-sim. The marker shapes indicate the subdomain type for which the parameters were varied and learned during training (see Table I). The are listed in the legend in black and apply to both single model (blue) and ensemble prediction (orange) data points. The color markers and the black shape markers have to be used in combination (e.g. blue diamonds correspond to single model predictions in the active area). The green cross shows the values used during the simulation of MO-sim.

overestimates the value $1/\rho_{par}$ which leads to a stronger voltage drop in the vicinty of the shunt subdomain. However, the calculated confidence ranges for $1/\rho_{par}$ show that the model is able to correctly deliver uncertainty values such that 3 out of 4 measured values lie within the predicted confidence interval.

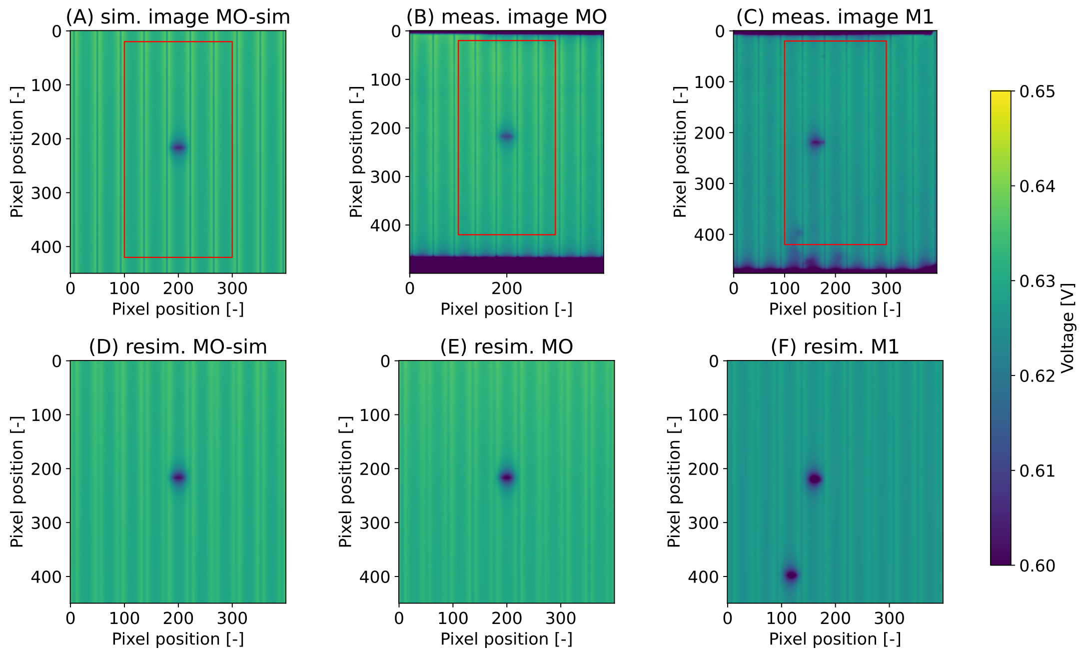

FIG. 7: Comparison of the input images (A)-(C) and the corresponding resimulations (D)-(E) based on the parameters predicted by the CNN inverse model.

IV. DISCUSSION

The results presented in section III show that the deep ensemble CNN used is a promising candidate for an inverse model for a silicon solar cell. It was possible to train the network to predict the used model parameters with high accuracy based on the simulation results. In cases with lower prediction accuracy, the model correctly predicts high error bars thanks to the negative log-likelihood loss function used. Resimulations based on the parameters predicted by the deep ensemble inverse model confirm the overall consistency of the approach by showing good agreement between original data and resimulations where the inverse calculation of the parameters was used for a forward simulation.

High uncertainties exist in the inverse prediction of j_0 and $R_\square$. Nevertheless, the forward calculation confirmed that the predicted parametrizations lead to a valid model representation of the given sample. This is an indication that the inverse problem defined by the equations implemented in the Laoss model is ill-posed. By substituting Equation 2 into Equation 1 one can show that in the regime of low internal resistivity ρ_{int} the derivative of Δv^t depends only on the product of j_0 and $R^t_\square$. This leads to a strong correlation of the two parameters that makes it difficult to resolve them independently from a voltage image alone. The dependency is also confirmed by the resimulation of the ensemble prediction results of MO-sim shown in 6. The CNN model underestimates the value of j_0. Since the value of $R_\square$ is simultaneously overestimated, the resimulation of MO-sim in Figure 8 agrees well with the CNN input image. This behavior is also consistent with the luminance imaging theory, which requires a combination of EL and PL imaging to determine j_0 and $R_\square$ separately. Therefore, if only a model based reconstruction of the EL image is of interest, a possible modification for inverse modeling could be to predict only the product of $j_0 \cdot R_\square$, which would simplify the problem. Similarly, in the discussed regime of low internal resistivity ρ_{int}, the parallel resistivity affects Equation 1 only through the quotient $R_\square/\rho_{par}$. Since there is no domain in which these two parameters have been varied simultaneously, this did not lead to further implications in the present study since one of the two values of the quotient has always been constant. An alternative to avoid the ill-posedness of the problem by guessing well-defined parameter combinations coudl be a physics-informed neural network that includes the knowledge of the governing equations during training of an inverse model neural network. Such an approach could help to force the network to correctly account for interdependent model parameters.

The parallel resistivity of the shunt region were less accurately predicted by the inverse deep ensemble model. Since the deep ensemble performed significantly better on the simulated validation image, it can be assumed that this is partly due to differences in the detailed appearance of the shunt region in the measurement when compared to the simulated data. Although much effort was put into accurately modeling the resulting voltage drop, even the best modeling results showed significant deviations from the measured data. During training, the CNN only learns to recognize the shape of the model-based shunts. Therefore, shunts in the measured data with different appearances are not expected to be handled correctly by the deep ensemble CNN. This is particularly evident for the second shunt in sample *M1*, which has a more triangular shape compared to the ellipsoidal shunts in the other test cells. Due to this unknown geometry, the CNN significantly

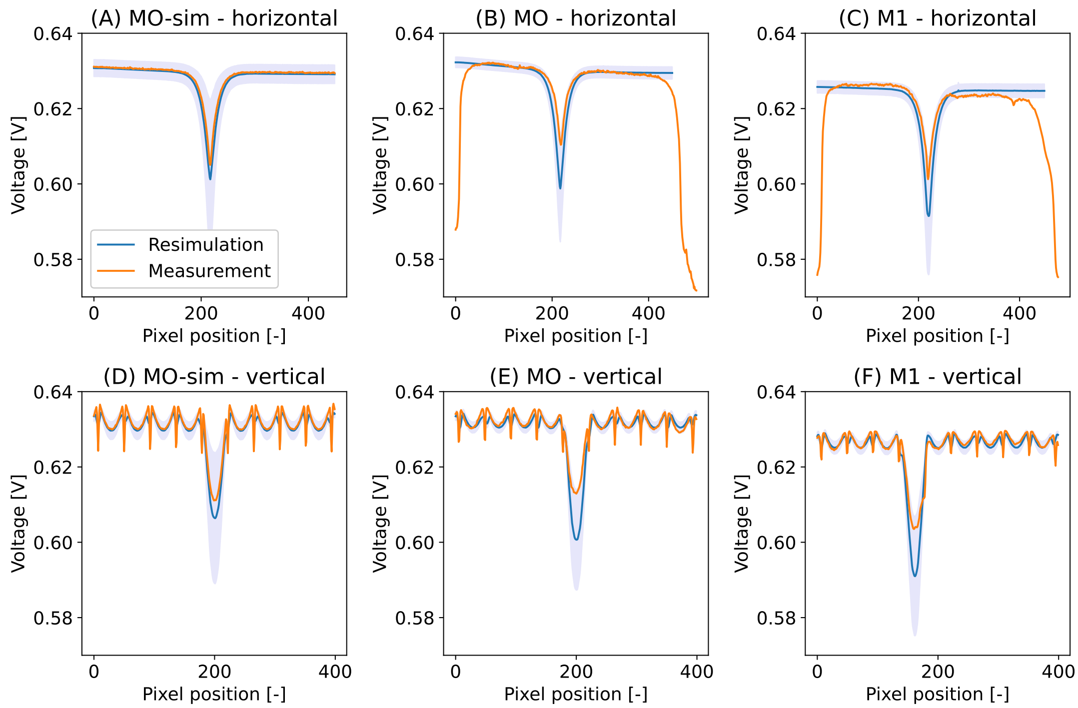

FIG. 8: Horizontal (A)-(C) and vertical (D)-(F) cross section of the input image and the corresponding resimulation based on the parameters predicted by the CNN inverse model including the simulation model uncertainty.

overestimates the conductance of the shunt in this case. Another reason for the overestimation of the shunt conductivities is the overestimation of the measurement's junction voltage in the area that was used for the input to the CNN model due to the inclusion of the areas affected by edge effects in the calibration procedure (see Section II A. The CNN model will interpret this as a higher level of applied voltage in which case the same absolute voltage drop in the shunt area will only be possible with an increased shunt conductivity.

The use of synthetic data for training a neural network can only be successful if the distribution of the training data fully covers all samples of interest to which the model is to be applied. If the network is applied to data that have features that are not present during the training phase, uncontrollable extrapolations are possible. The deep ensemble did not provide reasonable results for two of the four test samples. The results of the failed cases are shown in the Appendix in Figure 10 and in Figure 11. In these cases, the measurement samples were not perfectly aligned due to inaccuracies in the chosen measurement setup. As a result, the grid lines and the mask in channel 2, which define the pixels of the subregion to be predicted, have a small shift that causes the network to calculate values for $R_\square$ that are averaged between the high conductivity of the grid line and the low conductivity of the active region of the cell. This leads to a complete breakdown of the method. The results demonstrate the sensitivity of data-driven methods to the quality and comprehensiveness of the training data set.

V. CONCLUSION AND OUTLOOK

In many engineering applications, finding model parameters of numerical models based on an indirect measurement can be a difficult and time consuming task. In this paper, an inverse modelling approach based on a deep ensemble CNN was demonstrated utilizing a numerical model for the simulation of EL images of silicon solar cells with known cell geometry and known defect areas.

The work confirmed that a CNN is a valuable candidate for a data-driven inverse model. In total 75'000 simulated images have been created with Laoss based on parameters randomly sampled from a predefined range. With 94 % of the images in the training set and 6 % in the validation set, the CNN model has successfully been trained to learn the inverse mapping from the measurement image to the corresponding model parameters. By using a deep ensemble CNN model, an uncertainty prediction for the model parameters was included which is a key component for using the method in a scientific environment.

The model performance was tested with forward simulations based on the predicted model parameter distributions. The tests showed relative deviations of the calculated mean junction voltage from the original measured junction voltage of 0.2 % on average with a maximum of 10 %. The measured junction voltage was within the estimated uncertainties of the model results. The resistivities of the shunt subdomains have been estimated based on the measured current-voltage curve and compared to the values predicted by the CNN model. 3 out of 4 measurements are within the predicted uncertainty range which confirms the consistency of the approach. Failures of the method can be explained by mismatches between the simulation model results and the measurement data which leads to a simulation-reality gap. This critical dependency on accuracy of the synthetic training data is well known in similar methods[47,48].

By essentially pre-calculating an inverse model, the method shows a promising approach to create fast and accurate calculation of the parameters of a physical cell model. Since domain specific information is only present during the training data simulation stage, the approach is highly transferable to other types of solar cell or other engineering applications where a numerical simulator is available. For each cell layout a separate model has to be trained. Additionally, the presented results are restricted to shunts with standardized affected area and orientation. In order to include other defect types, they would have to be included into the training data generation. This would significantly increase the computation time needed for the generation of the images. However this training stage is responsible for the main effort of the method and the model can still provide fast and reliable results during the prediction stage. In particular, we expect that such a model can be used in the near future for efficient extraction of parameters of batches of novel solar cells with identical layout, for quality control, and to enable further improvements in the production and upscaling process. The full potential of such a model would be realized if it is used for quality control in a industry scale production line to characterize cells or modules and improve quality assurance, default classification and defect removal.

In a next step, we plan to apply the method to perovskite cells, where there is a high potential for improvement through detailed defects and cell characterization. Future improvements of the methods could also include integrating physical knowledge into other parts of deep neural network training, such as the loss function or network architecture, to improve data efficiency and out-of-sample predictions[49].

ACKNOWLEDGMENTS

Funding: This work was supported by Innosuisse [grant number 37304.1 IP-ENG]. Franz Baumgartner and Hartmut Nussbaumer from the ZHAW Inst. for Energy Systems and Fluid Engineering are acknowledged for providing the EL imaging camera and laser scribing system used here.

DATA AVAILABILITY STATEMENT

The data that support the findings will be available in Github at https://doi.org/10.5281/zenodo.7528885 following an embargo from the date of publication until 01.01.2028 to allow for commercialization of research findings.

VI. APPENDIX

A. Additional figures

REFERENCES

[1] C. Breyer, D. Bogdanov, A. Gulagi, A. Aghahosseini, L. S. Barbosa, O. Koskinen, M. Barasa, U. Caldera, S. Afanasyeva, M. Child, J. Farfan, and P. Vainikka, "On the role of solar photovoltaics in global energy transition scenarios," Progress in Photovoltaics: Research and Applications **25**, 727–745 (2017).

[2] R. Koppelaar, "Solar-pv energy payback and net energy: Meta-assessment of study quality, reproducibility, and results harmonization," Renewable and Sustainable Energy Reviews **72**, 1241 – 1255 (2017).

[3] E. van Dyk and E. Meyer, "Analysis of the effect of parasitic resistances on the performance of photovoltaic modules," Renewable Energy **29**, 333 – 344 (2004).

[4] M. Kasemann, D. Grote, B. Walter, W. Kwapil, T. Trupke, Y. Augarten, R. A. Bardos, E. Pink, M. D. Abbott, and W. Warta, "Luminescence imaging for the detection of shunts on silicon solar cells," Progress in Photovoltaics: Research and Applications **16**, 297–305 (2008).

[5] A. Helbig, T. Kirchartz, R. Schaeffler, J. H. Werner, and U. Rau, "Quantitative electroluminescence analysis of resistive losses in cu(in, ga)se2 thin-film modules," Solar Energy Materials and Solar Cells **94**, 979 – 984 (2010).

[6] B. Du, R. Yang, Y. He, F. Wang, and S. Huang, "Nondestructive inspection, testing and evaluation for si-based, thin film and multi-junction solar cells: An overview," Renewable and Sustainable Energy Reviews **78**, 1117 – 1151 (2017).

[7] F. Frühauf, J. Wong, J. Bauer, and O. Breitenstein, "Finite element simulation of inhomogeneous solar cells based on lock-in thermography and luminescence imaging," Solar Energy Materials and Solar Cells **162**, 103–113 (2017).

[8] E. L. Comi, E. Knapp, S. Weidmann, C. Kirsch, S. Jenatsch, R. Hiestand, and B. Ruhstaller, "Sinusoidal small-signal (AC) and steady-state (DC) analysis of large-area solar cells," Solar Energy Advances **1**, 100003 (2021).

[9] L. Rakocevic, L. E. Mundt, R. Gehlhaar, T. Merckx, T. Aernouts, M. C. Schubert, S. W. Glunz, and J. Poortmans, "Loss Analysis in Perovskite Photovoltaic Modules," Solar RRL **3**, 1900338 (2019).

[10] E. Comi, E. Knapp, M. Battaglia, S. Weidmann, C. Kirsch, S. Jenatsch, R. Hiestand, M. Bonmarin, and B. Ruhstaller, "Electro-thermal Model for Lock-in Infrared Imaging of Defects in Perovskite Solar Cells," 8th World Conference on Photovoltaic Energy Conversion , 241–246 (2022).

[11] M. L. Johnson and L. M. Faunt, "[1] Parameter estimation by least-squares methods," Methods in Enzymology **210**, 1–37 (1992).

[12] A. Tarantola, "Inverse Problem Theory and Methods for Model Parameter Estimation," Inverse Problem Theory and Methods for Model Parameter Estimation (2005), 10.1137/1.9780898717921.

[13] V. M. Le Corre, T. S. Sherkar, M. Koopmans, and L. J. A. Koster, "Identification of the dominant recombination process for perovskite solar cells based on machine learning," Cell Reports Physical Science **2**, 100346 (2021).

[14] G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, and R. Willett, "Deep Learning Techniques for Inverse Problems in Imaging," IEEE Journal on Selected Areas in Information Theory **1**, 39–56 (2020), arXiv:2005.06001.

[15] V. Das, A. Pollack, U. Wollner, and T. Mukerji, "Convolutional neural network for seismic impedance inversion," https://doi.org/10.1190/geo2018-0838.1 **84**, R869–R880 (2019).

[16] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Networks **61**, 85–117 (2015), arXiv:1404.7828.

[17] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," Nature 2015 521:7553 **521**, 436–444 (2015).

[18] Z. Ren, F. Oviedo, H. Xue, M. Thway, K. Zhang, N. Li, J. D. Perea, M. Layurova, Y. Wang, S. Tian, T. Heumueller, E. Birgersson, F. Lin, A. Aberle, S. Sun, I. M. Peters, R. Stangl, C. J. Brabec, and T. Buonassisi, "Physics-guided characterization and optimization of solar cells using

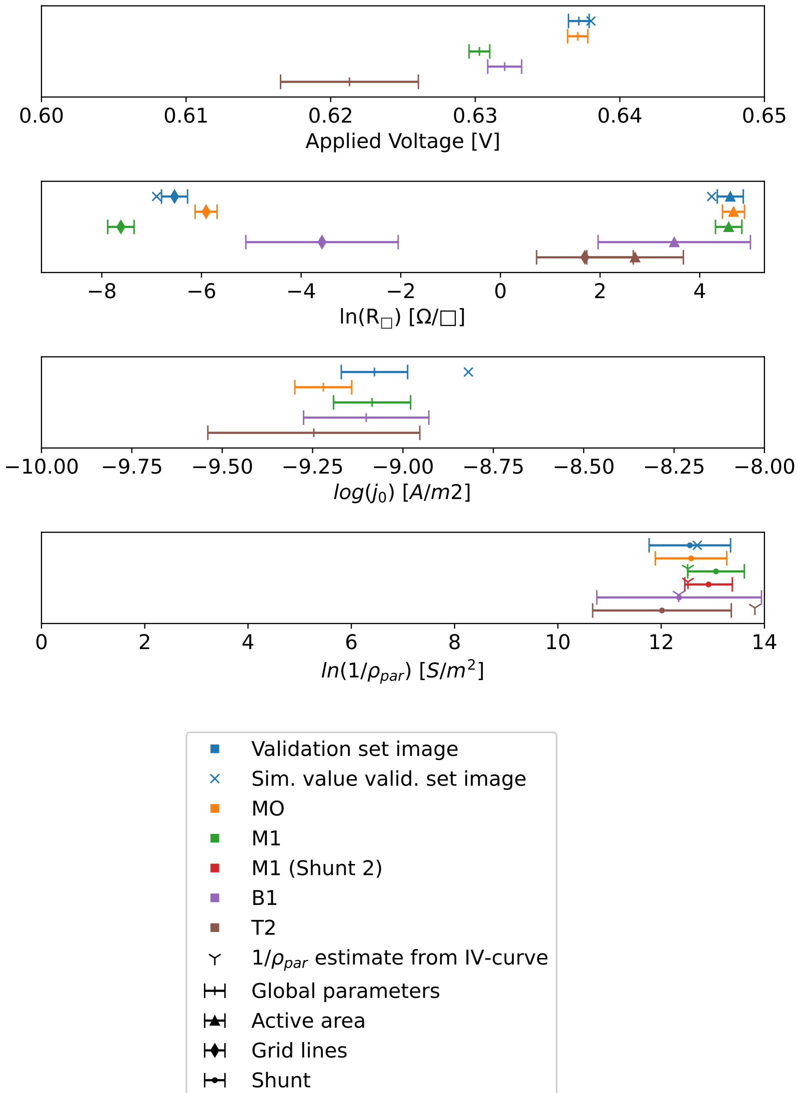

FIG. 9: Detailed ensemble parameter predictions for the validation image and the investigated measurement samples.

surrogate machine learning model," Conference Record of the IEEE Photovoltaic Specialists Conference , 3054–3058 (2019).

[19] A. Wörnhör, M. Demant, H. Vahlman, and S. Rein, "A self-consistent hybrid model connects empirical and optical models for fast, non-destructive inline characterization of thin, porous silicon layers," EPJ Photovoltaics **14**, 4 (2023).

[20] Y. Buratti, A. Sowmya, R. Evans, T. Trupke, and Z. Hameiri, "Half and full solar cell efficiency binning by deep learning on electroluminescence images," Progress in Photovoltaics: Research and Applications **30**, 276–287 (2022).

[21] M. Hoffmann, C. Buerhop-Lutz, L. Reeb, T. Pickel, T. Winkler, B. Doll, T. Würfl, I. Marius Peters, C. Brabec, A. Maier, and V. Christlein, "Deep-learning-based pipeline for module power prediction from electroluminescense measurements," Progress in Photovoltaics: Research and Applications **29**, 920–935 (2021).

[22] X. Zhang, Y. Hao, H. Shangguan, P. Zhang, and A. Wang, "Detection of surface defects on solar cells by fusing Multi-channel convolution neural networks," Infrared Physics Technology **108**, 103334 (2020).

[23] P. Kunze, J. M. Greulich, A. Tummalieh, W. Wirtz, H. Hoeffler, N. Woehrle, S. Glunz, S. Rein, and M. Demant, "Contactless Inline IV Measurement of Solar Cells Using an Empirical Model," Solar RRL **7**, 2200599 (2023).

[24] M. Demant, P. Virtue, A. Kovvali, S. X. Yu, and S. Rein, "Learning Quality Rating of As-Cut mc-Si Wafers via Convolutional Regression Networks," IEEE Journal of Photovoltaics **9**, 1064–1072 (2019).

[25] A. S. Kovvali, M. Demant, and S. Rein, "Spatially Resolved Material Quality Prediction Via Constrained Deep Learning," Conference Record of the IEEE Photovoltaic Specialists Conference , 3059–3062 (2019).

[26] L. Hollenstein, L. Lichtensteiger, T. Stadelmann, M. Amirian, L. Budde, J. Meierhofer, R. M. Füchslin, and T. Friedli, "Unsupervised Learning and Simulation for Complexity Management in Business Operations," Applied Data Science , 313–331 (2019).

[27] M. Diethelm, L. Penninck, M. Regnat, T. Offermans, B. Zimmermann, C. Kirsch, R. Hiestand, S. Altazin, and B. Ruhstaller, "Finite element modeling for analysis of electroluminescence and infrared images of thin-film solar cells," Solar Energy **209**, 186 – 193 (2020).

[28] E. L. Comi, E. Knapp, S. Weidmann, C. Kirsch, S. Jenatsch, R. Hiestand, and B. Ruhstaller, "Sinusoidal small-signal (AC) and steady-state (DC) analysis of large-area solar cells," Solar Energy Advances **1**, 100003 (2021).

[29] B. Lakshminarayanan, A. Pritzel, and C. B. Deepmind, "Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles," Advances in Neural Information Processing Systems **30** (2017).

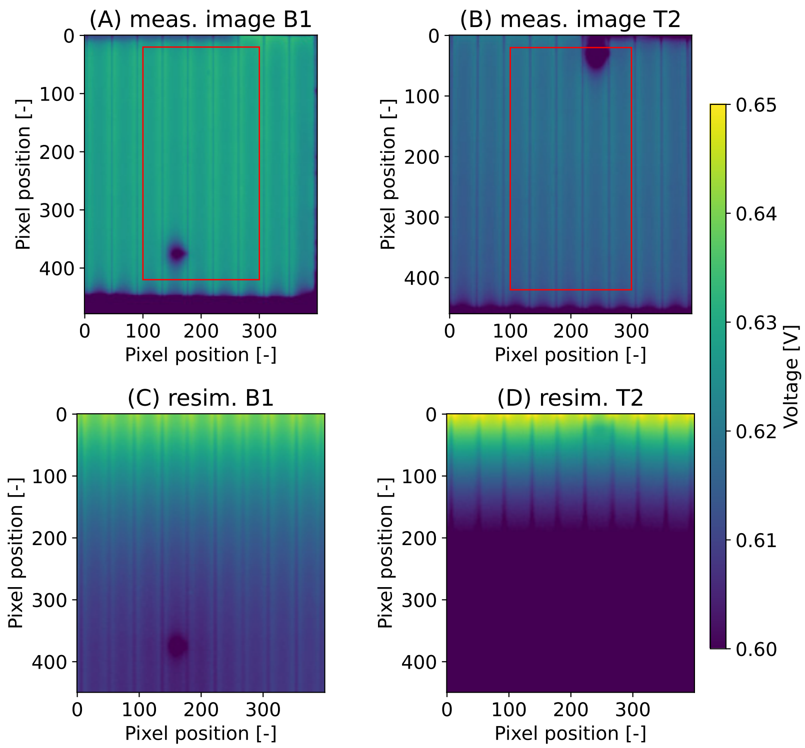

FIG. 10: Comparison of the input images (A) and (B) and the corresponding resimulations (C) and (D) based on the parameters predicted by the CNN inverse model (failed cases).

[30] T. Stadelmann, V. Tolkachev, B. Sick, J. Stampfli, and O. Dürr, "Beyond imagenet: deep learning in industrial practice," in *Applied Data Science* (Springer, 2019) pp. 205–232.

[31] T. Stadelmann, M. Amirian, I. Arabaci, M. Arnold, G. F. Duivesteijn, I. Elezi, M. Geiger, S. Lörwald, B. B. Meier, K. Rombach, and L. Tuggener, "Deep Learning in the Wild," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **11081 LNAI**, 17–38 (2018), arXiv:1807.04950.

[32] J. Gawlikowski, C. R. N. Tassi, M. Ali, J. Lee, M. Humt, J. Feng, A. Kruspe, R. Triebel, P. Jung, R. Roscher, M. Shahzad, W. Yang, R. Bamler, and X. X. Zhu, "A Survey of Uncertainty in Deep Neural Networks," (2021), arXiv:2107.03342.

[33] N. Thuerey, P. Holl, M. Mueller, P. Schnell, F. Trost, and K. Um, "Physics-based Deep Learning," (2021), arXiv:2109.05237.

[34] Motech Industries, "Motech products and services," (2021).

[35] M. Glatthaar, J. Giesecke, M. Kasemann, J. Haunschild, M. The, W. Warta, and S. Rein, "Spatially resolved determination of the dark saturation current of silicon solar cells from electroluminescence images," Journal of Applied Physics **105**, 113110 (2009).

[36] M. Diethelm, L. Penninck, S. Altazin, R. Hiestand, C. Kirsch, and B. Ruhstaller, "Quantitative analysis of pixel crosstalk in amoled displays," Journal of Information Display **19**, 61–69 (2018).

[37] L. Penninck, M. Diethelm, S. Altazin, R. Hiestand, C. Kirsch, and B. Ruhstaller, "Modelling crosstalk through common semiconductor layers in amoled displays," Journal of the Society for Information Display **26**, 546–554 (2018).

[38] T. Lanz, M. Bonmarin, M. Stuckelberger, C. Schlumpf, C. Ballif, and B. Ruhstaller, "Electrothermal finite-element modeling for defect characterization in thin-film silicon solar modules," IEEE Journal on Selected Topics in Quantum Electronics **19** (2013), 10.1109/JSTQE.2013.2250259.

[39] Fluxim AG, "Laoss software to design and optimize large area semiconductor devices," (2021).

[40] S. P. Phang, H. C. Sio, and D. Macdonald, "Carrier de-smearing of photoluminescence images on silicon wafers using the continuity equation," Applied Physics Letters **103**, 192112 (2013), https://doi.org/10.1063/1.4829658.

[41] A. Teal and M. Juhl, "Correcting the inherent distortion in luminescence images of silicon solar cells," in *2015 IEEE 42nd Photovoltaic Specialist Conference (PVSC)* (2015) pp. 1–5.

[42] O. Breitenstein, F. Frühauf, and A. Teal, "An improved method to measure the point spread function of cameras used for electro- and photoluminescence imaging of silicon solar cells," IEEE Journal of Photovoltaics **6**, 522–527 (2016).

[43] Z. Liu, M. Peters, V. Shanmugam, Y. S. Khoo, S. Guo, R. Stangl, A. G. Aberle, and J. Wong, "Luminescence imaging analysis of light harvesting from inactive areas in crystalline silicon pv modules," Solar Energy Materials and Solar Cells **144**, 523–531 (2016).

[44] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles," Advances in Neural Information Processing Systems **2017-December**, 6403–6414 (2016), arXiv:1612.01474.

[45] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," CoRR **abs/1905.11946** (2019), arXiv:1905.11946.

[46] S. Lathuilière, P. Mesejo, X. Alameda-Pineda, and R. Horaud, "A Comprehensive Analysis of Deep Regression," IEEE Transactions on Pattern Analysis and Machine Intelligence **42**, 2065–2081 (2018), arXiv:1803.08450.

[47] T. Inoue, S. Choudhury, G. De Magistris, and S. Dasgupta, "Transfer Learning from Synthetic to Real Images Using Variational Autoencoders for Precise Position Detection," Proceedings - International Conference on Image Processing, ICIP , 2725–2729 (2018).

[48] N. Simmler, P. Sager, P. Andermatt, R. Chavarriaga, F.-P. Schilling, M. Rosenthal, and T. Stadelmann, "A survey of un-, weakly-, and semi-supervised learning methods for noisy, missing and partial labels in industrial vision applications," in *2021 8th Swiss Conference on Data Science (SDS)* (2021) pp. 26–31.

[49] J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar, "Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems," **1**, 1 (2020), arXiv:2003.04919.

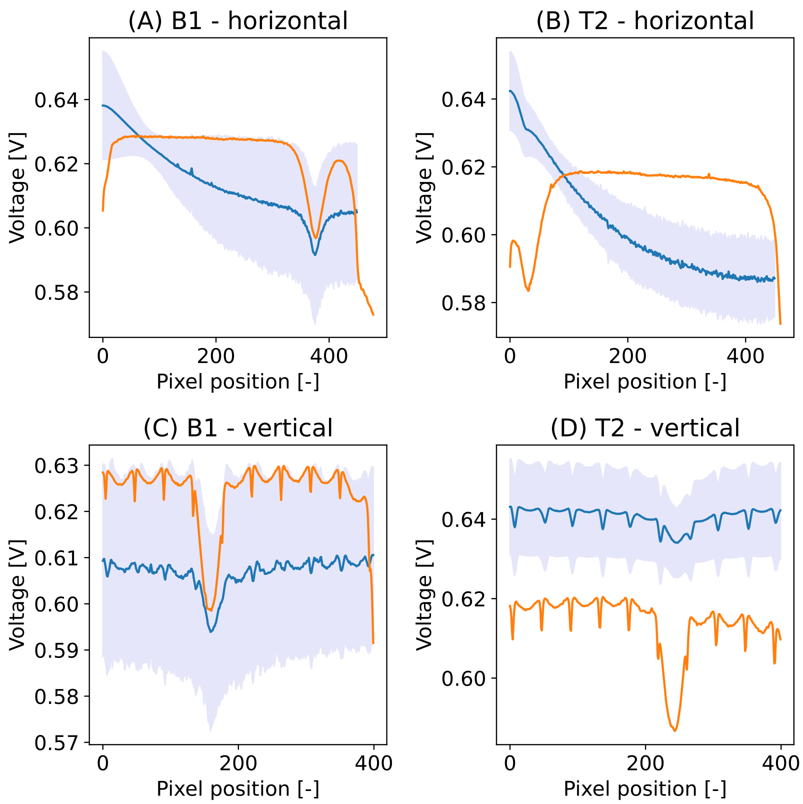

FIG. 11: Horizontal (A) and (B) and vertical (C) and (D) cross section of the input image and the corresponding resimulation based on the parameters predicted by the CNN inverse model including the simulation model uncertainty (failed cases).