Learning Quality Rating of As-Cut mc-Si Wafers via Convolutional Regression Networks

Matthias Demant, Patrick Virtue, Aditya Kovvali, Stella X. Yu, and Stefan Rein

***Abstract*—This paper investigates deep convolutional neural networks (CNNs) for the assessment of defects in multicrystalline silicon (mc-Si) and high-performance mc-Si wafers for solar cell production based on photoluminescence (PL) images. We identify and train a CNN regression model to forecast the I–V parameters of passivated emitter and rear cells from given PL images of the as-cut wafers. The presented end-to-end model directly processes the PL image and does not rely on the human-designed image feature. Domain knowledge is replaced by a model based on a huge variety of empirical data. The comprehensive dataset allows for the evaluation of the generalizability of the model with test wafers from bricks and manufacturers not presented in the training set. We achieve mean absolute prediction errors as low as $0.11\%_{\text{abs}}$ in efficiency for test wafers from "unknown" bricks, which improves handcrafted feature-based methods by $35\%_{\text{rel}}$ at simultaneously lower computational costs for prediction. Samples with high prediction errors are investigated in detail showing an increased iron point defect concentration.**

***Index Terms*—Convolutional neural network (CNN), densely connected neural network (DenseNet), high-performance multicrystalline silicon (HPM-Si), machine learning, material quality, multicrystalline silicon (mc-Si), passivated emitter and rear cell (PERC), photoluminescence (PL), rating, regression, solar cell.**

I. Introduction

THE incoming control for solar cell production of multicrystalline silicon (mc-Si) requires a rapid assessment of the wafer quality in terms of the expected solar cell efficiency, which can be implemented based on photoluminescence (PL) imaging [1] and modern pattern recognition techniques. Since the incoming control for solar cell production takes place directly after wafering, a prediction based on lifetime images of passivated wafers [2] and bricks [3] is not available. By measuring the radiative recombination of the excess charge carriers via PL imaging, crystallization-related defect patterns, which are recombination active, can be observed in as-cut wafers already. The quantity of these defect structures correlates with solar cell performance [4]. Due to the low acquisition time and high spatial resolution of PL imaging, we introduce a rating model based on PL images.

Since the material rating of as-cut samples is challenging for high-quality materials and high-quality solar cell processes, such as the passivated emitter and rear cell (PERC) process [5], the models need to be evaluated on these data. An incoming control requires a rapid investigation of samples without passivating the surface layer. In this early process step, surface recombination limits the measurement of the bulk lifetime of the excess charge carriers. This effect is challenging for the rating of high-quality material and for solar cells with higher efficiency potential. In particular, the development of a prediction model for the PERC has been shown to be more challenging than for solar cells with aluminum back-surface field [6]. We accept this challenge within our investigation: our dataset contains different material types of p-type mc-Si wafers from standard to high performance multicrystalline silicon (HPMC-Si) [7], [8] wafers, which are processed to PERCs.

Traditional rating approaches use human feature engineering to quantify defect features from complex defect patterns. An example for a domain-based sorting metric compares PL intensities at grain boundaries and within grains as criteria for enhanced gettering processes [9]. Other domain-relevant features are regions of reduced lifetime due to crucible contamination, grain boundaries, or dislocation clusters, which are typically detected and quantified based on handcrafted filters and thresholds followed by a regression analysis [6], [10]–[13].

The degree of pattern complexity can be increased by following the bag-of-features paradigm: images are described by a collection of local image descriptors, which represent prototypical structures. In previous investigations, the image of the dark saturation current based on these prototypical PL structures [14] or even the global V_{oc} based on the distribution of these PL structures [15] has been predicted. Although supervised and unsupervised learning steps are applied, the kind of descriptor is still selected manually.

The quantification of defects with handcrafted solutions is difficult.

Manuscript received June 24, 2018; revised November 29, 2018; accepted January 21, 2019. This work was supported by the German Federal Ministry for Economic Affairs and Energy through the project "Q-Crystal" under Contract 0324103A. The work of M. Demant was supported by a fellowship of the German Academic Exchange Service within the program FITweltweit. *(Corresponding author: Matthias Demant.)*

M. Demant is with the Fraunhofer Institute for Solar Energy 79100, Freiburg Germany, and also with the International Computer Science Institute, University of California, Berkeley, CA 94704 USA (e-mail: matthias.demant@ise.fraunhofer.de).

P. Virtue and S. X. Yu are with the International Computer Science Institute University of California, Berkeley, CA 94704 USA (e-mail: virtue@eecs.berkeley.edu; stellayu@berkeley.edu).

A. Kovvali and S. Rein are with the Fraunhofer Institute for Solar Energy 79100 Freiburg, Germany (e-mail: aditya.kovvali@ise.fraunhofer.de; stefan.rein@ise.fraunhofer.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JPHOTOV.2019.2906036

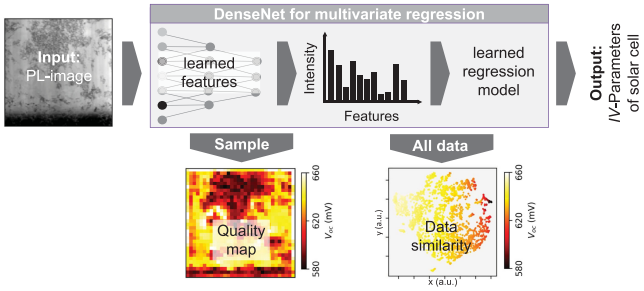


Fig. 1. Rating and visualization based on PL images of as-cut wafers. The two visualization techniques are presented in the second part of our study in [28].

1) Lifetime reducing defects, e.g., contamination from the crucible and structural defects, are not mutually exclusive and reduce PL intensity.
2) They interact during the solar cell process, e.g., when the enhancement of phosphorous diffusion gettering depends on the dislocation density [16]–[21].
3) The spatial distribution of defect patterns plays an important role for V_{oc} prediction [22], [23].

A human expert can hardly oversee the variety of patterns and will lose information using handcrafted methods for feature extraction.

Deep learning allows complex features to be learned in a completely data-driven fashion, which is feasible within our large dataset. Triggered by the successful object classification of the ImageNet database [24], convolutional neural networks (CNNs) have redeemed feature engineering approaches in most computer vision applications. Based on a large amount of empirical data, the network learns to minimize a loss function via gradient-based optimization. The learned representation is capable of considering a large receptive field in the input image and describing complex patterns. The abundance in material variation within our experiment allows efficient and accurate assessment on the performance of material via a data-driven deep learning approach.

In contrast with physical device simulations (e.g., [25]), CNNs are often treated as "blackbox." Yet, a deeper understanding of what has been learned is crucial for the acceptance of CNNs in the photovoltaic community. In addition to high prediction accuracy, CNNs can be used to localize the reasons for changes in performance by analyzing the activation of the network [26]. Furthermore, CNNs can learn a semantic representation of the data during the training process. The network learns to compress the PL images in a meaningful way for the prediction of solar cell performance. This semantic representation can be extracted from the network and used to analyze the similarity of the dataset in terms of defect occurrence and solar cell performance. Visualization techniques [27] can be used to reveal the similarity of the data learned by the network. Both methods are discussed in [28] for the same dataset.

As shown in Fig. 1, we present an end-to-end rating model for material assessment and visualization, where PL images and I–V parameters are used as input and output, respectively. In Section II, we explain how we adapt a densely connected CNN [29] to solve the multivariate regression task to forecast four I–V parameters, namely, efficiency, open-circuit voltage, short-circuit current, and fill factor, in one network. The final regression step is based on learned features only. Results on "unknown" bricks and manufacturers are shown in Sections IV-A and IV-B, respectively. A detailed analysis of the prediction errors is given in Section IV-C. Details and results on quality mapping and data visualization are presented in the second part of this study [28].

II. Our CNN Architecture for Regression

For an introduction to CNNs, we refer the reader to the work of Krizhevsky *et al.* [24] and briefly describe this network structure in the following. In this early CNN called AlexNet, the data are passed sequentially through the network. In subsequent layers, feature information is extracted via convolutional filters, the filter results are normalized, nonlinearly transformed, and downsampled. With increasing depth of the network, the number of filter channels, as well as the receptive field of each feature, increases. Three fully connected layers complete the network. For a regression task, the result of the final fully connected layer returns the predicted value.

With deeper network architectures, training becomes harder. In recent years, network architectures with fully convolutional design [30], [31] and an increasing number of layers became more successful like the visual geometry group (VGG) network [32]. In CNNs, such as AlexNet and VGG, all of the information content relevant to the final output must pass through every single layer. This representational challenge is exacerbated, as the depth of the network increases. This can be addressed by using skip connections, as presented in ResNet [33].

Training such a very deep network is still complicated, due to the vanishing-gradient problem. This problem can appear when the gradient of the error function is backpropagated through the network. Following the chain rule, small values may be multiplied in each layer and, thus, vanish before they reach earlier layers of the network.

Densely connected neural networks (DenseNets) are introduced as easily trainable networks. Huang *et al.* [29] propose to improve the learning process with a DenseNet by adding additional skip connections to the network: for each layer, all previous feature maps are concatenated as input. The resulting network requires less parameters than, e.g., AlexNet or ResNet because of feature reuse. In addition, the dense connections are beneficial for the training process, since the error gradients flow directly from later to earlier layers of the network.

Besides this robust training procedure, the visualization via the activation mapping technique [26] is a good reason for the usage of this network: The DenseNet allows the output of a quality map within few computational steps, as introduced in the second part of this study [28].

Our network structure with three densely connected blocks is visualized in Fig. 2. A dense block is composed of three basic elements:

1) convolutional layers with kernel size $3 \times 3 \times n_{\mathrm{c}}$ ($\mathrm{Conv}_{3\times3}$) to extract image features (n_{c} is the number of input channels);

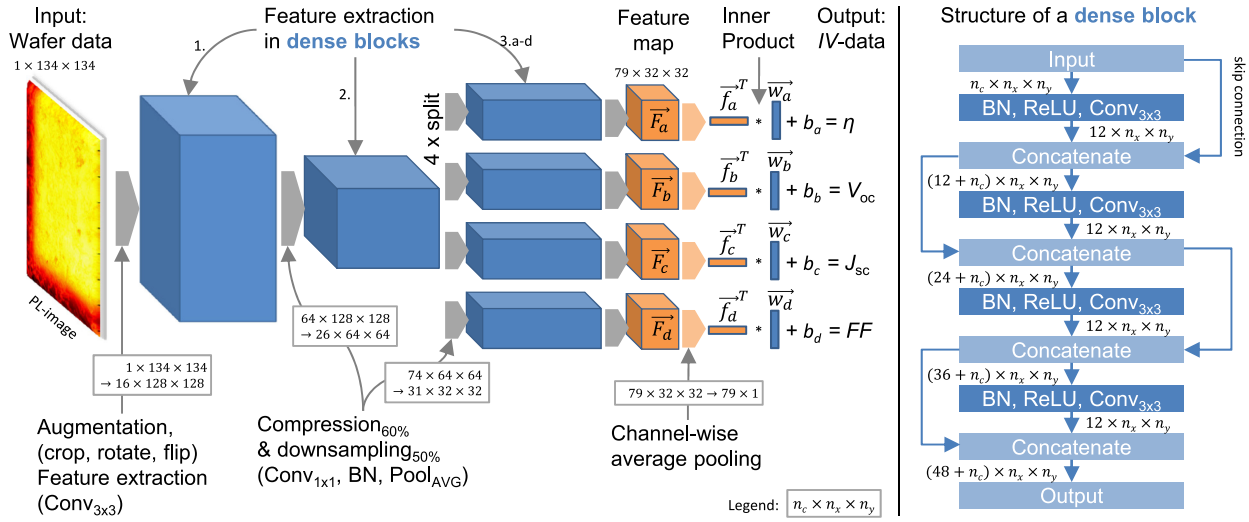


Fig. 2. Scheme of the CNN for multivariate regression of I–V parameters based on PL images of as-cut wafers with a resolution of 134 × 134 pixels.

2) a batch normalization (BN) step;
3) a rectifying linear unit (ReLU) as nonlinear transformation.

As shown in the right scheme of Fig. 2, all previous feature maps are concatenated as input of the current layer. The transition from block to block contains: 1) a convolutional layer ($\text{Conv}_{1\times1}$) for feature compression; and 2) an average pooling layer (Pool_{AVG}) to downsample the data.

Within a walking tour, we describe the forward pass through the network. Our input data have a resolution of 264 × 264 or 134 × 134 pixels due to GPU memory constraints. During the training, an optional data augmentation step can be performed: the images are randomly cropped at both borders, flipped, and rotated. The augmentation step aims for a robust prediction model and avoids overfitting to the training set. Nevertheless, every region of the wafer contributes to the quality of the solar cell. Thus, for the test case, the image is cropped systematically at different positions. Each patch is passed through the network. Finally, the predictions are averaged to consider all image regions for the prediction.

The first convolution step of the first dense block contains 16 channels with a resolution of 128 × 128 for an input image with size 134 × 134 pixels and cropping size of 6 pixels. The feature map grows by 12 feature channels for each of the four subsequent layers in a dense block. After passing the first and second dense blocks, the data are compressed to approximately 40% of the channel size and downsampled by 50% in spatial resolution, respectively. This results in feature maps with a dimension of 26 × 64 × 64 after the first and 31 × 32 × 32 after the second dense block.

Up to now, the different quality prediction values share the same feature maps. At this point, the model will be split: for each of the output values (a) efficiency η, (b) open-circuit voltage V_{oc}, (c) short-circuit current J_{sc}, and (d) fill factor FF, an individual third dense block $k \in \{a, b, c, d\}$ will be applied. The final feature map $F_{k,i}(c, x, y)$ of a sample i contains $n_c = 79$ channels and a spatial resolution of 32 pixel in x- and y-directions. Each of these four dense blocks concludes with a channelwise average pooling step of the final feature map according to

$$\vec{f}_{k,i}(c) = \frac{1}{n_{\text{px}}} \sum_{x,y} \vec{F}_{k,i}(c, x, y) \tag{1}$$

with $\vec{f}_{k,i} \in \mathbb{R}^{79}$ and $n_{\text{px}} = 1024$ elements in each channel c. The final regression step is accomplished by an inner product layer with weights $\vec{w}_k$ and offset b_k for the prediction of one the four I–V parameters $\hat{y}_k$ with

$$\hat{y}_{k,i} = \left\langle \vec{w}_k, \vec{f}_{k,i} \right\rangle + b_k. \tag{2}$$

During the training, loss between predicted quality value $\hat{y}_{k,i}$ and the measured $y_{k,i}$ can be computed as ℓ_1-loss, ℓ_2-loss or weighted mean of both losses.

III. Experimental Approach

A. Materials and Solar Cell Production

1) Material Selection: The dataset was collected within a comprehensive material survey. The presented material selection contains about 7300 wafers from ten different manufacturers taken from different bricks of at least 74 bricks with varying brick positions within 18 blocks. Most of the wafers were sampled from different positions of the brick at regular step sizes, e.g., at each fifth or tenth wafer position from inner, edge, and corner bricks. Material from one of the manufacturers was carefully sampled at two positions from 25 different boxes. The dataset contains HPMC-Si materials from at least three different manufacturers with known brick positions.

2) Measurements and Solar Cell Production: The wafers are laser-marked with a data matrix code to track the samples during the experiment. PL images are measured by a commercially available inline PL system with a line-scan camera with an InGaAs detector, as described by Höffler *et al.* [34]. The wafers have a size of 156 × 156 mm and the PL images a resolution of 1024 × 1024 pixels. The wafers are processed to PERCs within

an industrial production line. The I–V data were measured after solar cell production.

3) Reference Measurement: The concentration of iron point defects is determined by combining PL images at defined states [35]: 1) after storage in the dark, where the defect is present as iron–boron pairs; and (2) after illumination, where the defect is present as interstitial iron. The calculation follows the approach of Zoth and Bergholz [36] and is described in detail in [37].

4) Predicted Parameters: The quality rating aims at predicting the solar cell efficiency, which depends on V_{oc}, J_{sc}, and FF; therefore, the prediction accuracy for all three parameters has to be investigated. V_{oc} depends the strongest on material quality, namely, the recombination current J_0, according to the one-diode equation $V_{\text{oc}} = V_{\text{th}} \cdot \ln J_{\text{sc}}/J_0$ with the thermal voltage V_{th}. J_{sc} depends on recombination, but is sensitive on diffusion length variations in the range of and below wafer thickness. Finally, FF only depends on recombination indirectly via V_{oc}. Therefore, the prediction quality will decrease from V_{oc} via J_{sc} to FF.

B. Definition of Training and Test Sets

To test the generalizability of our approach, the data are split into three disjoint sets for training, validation, and testing of the algorithm. The validation set is used to tune the parameters of the machine learning algorithm. We investigate two scenarios, which are relevant for the application within an incoming control. In the first scenario, we want to predict the quality of materials, which are similar to materials of a crystallization, which has been used for the training of the model. In the second scenario, we analyze material from an "unknown" manufacturer. Therefore, we split the training and test data on brick (box) and manufacturer level, respectively. For each set, no two wafers are from the same brick or manufacturer.

1) Scenario 1, Testing Unknown Bricks: The training set contains about 2900 wafers from 32 bricks and seven boxes and the test set 4325 wafers from 42 bricks and 18 boxes, which are distinct to the training set with regard to the brick origin. A validation set of 424 wafers was randomly selected and removed from the test set for parameter optimization.

2) Scenario 2, Testing Unknown Manufacturers: A prediction model can be used for the performance evaluation of wafers from a new manufacturer with materials never processed within the specific production line, which is addressed in this scenario.

In Scenario 2, a prediction of HPMC-Si wafers is of special interest: HPMC-Si material will displace regular mc-Si material in the future, due to higher solar cell performance. Additionally, a PL-based rating of as-cut wafers gets more challenging for this material class. The measurement of high bulk lifetimes is limited by the surface recombination in as-cut wafers. Therefore, Scenario 2 focuses on the evaluation of HPMC-Si materials.

We define different training and test datasets for each of the three manufacturers (M1, M2, and M3) with HPMC-Si material. Every manufacturer is evaluated individually as "unknown." For each of the three configurations, all wafers from the respective manufacturer are removed from the training set defined in Scenario 1. The training set for the prediction of materials from the respective "unknown" manufacturer contains selected bricks from nine other manufacturers according to Scenario 1 excluding its own material. A DenseNet for V_{oc} prediction is trained based on this training set. The test errors are evaluated for the HPMC-Si bricks of the respective manufacturer Mi. The numbers of ingots, bricks, and wafers in the test set are indicated in Table II.

C. Implementation and Training

The deep learning framework Caffe [38] from the Berkeley Artificial Intelligence Research Lab was used to model the network. The PL data are aligned [39, pp. 58–59], downsampled, and normalized before the training process. The PL images are analyzed with a resolution of 264 × 264 pixels and 134 × 134 pixels before the random cropping step with cropping regions of 8 and 6 pixels, respectively. In addition, a group without random cropping was evaluated at a resolution of 264 × 264 pixels.

The models were trained using stochastic gradient descent for 400k iterations with a batch size of 20 images using Nesterov optimization, which corresponds to 2666 epochs for predicting unknown bricks (training set 1) and slightly more epochs for the prediction of unknown manufacturers (training set 2). In machine learning, the number of epochs indicates how often all training data have been used for training. The learning rate was evaluated on a validation set and initialized to 0.01 and is stepwise lowered by two times after 50 000 iterations, which corresponds to 330 epochs in training set 1. The large number of epochs was selected to compensate the small batch size due to GPU memory constraints. During our evaluations, training runs were performed with NVIDIA Tesla K20 and Titan Xp with 5- and 12-GB memory, respectively. For the latter, each iteration took about 150-ms computation time for image resolutions of 256 × 256 pixels.

D. Reference Method With Designed Features

The deep learning results are compared with our previous feature-based approach [6]. Within this reference method, designed features are calculated from PL images with a resolution of 1024 × 1024 pixels. The features quantify the average PL intensity, recombination active structures, and regions of reduced lifetime due to crucible contamination. In addition, the area fraction of regions with different amount of structural defects and PL intensities is distinguished. All features are used as input to train four distinct regression models to predict one of the four I–V parameters each. The regression models are trained with and without a regularization term to reduce feature activation [40].

IV. Results on Learning Quality Prediction

A. Scenario 1: I–V Prediction for Unknown Bricks

The deep neural network is trained to predict η, V_{oc}, J_{sc}, and FF based on PL images of as-cut wafers within one network. The results for the training and test split according Scenario 1 (see Section III-A) are presented in Table I.

TABLE I
(SCENARIO 1) I–V PREDICTION FOR TEST WAFERS FROM BRICKS NOT IN THE TRAINING SET

Model and resolution	wd*	error	η [%]	V_{oc} [mV]	J_{sc} [mA/cm^2]	FF [%]
Designed features [6] 1024×1024 px^2	□	mae	0.17	2.77	0.16	0.31
		rms	0.30	4.31	0.22	0.53
		corr	0.80	0.87	0.83	0.47
	⊠	mae	0.16	2.73	0.17	0.32
		rms	0.22	3.83	0.22	0.42
		corr	0.88	0.89	0.83	0.53
DenseNet 264×264 px^2 (0px cropped)	□	mae	0.12	2.12	0.14	0.30
		rms	0.17	3.07	0.19	0.39
		corr	0.93	0.93	0.88	0.65
264×264 px^2 (8px cropped)	□	mae	**0.11**	**2.05**	**0.14**	**0.28**
		rms	0.16	3.01	0.18	0.38
		corr	0.94	0.94	0.88	0.67
DenseNet 134×134 px^2 (6px cropped)	□	mae	0.12	2.15	0.15	0.28
		rms	0.17	3.10	0.19	0.37
		corr	0.93	0.93	0.87	0.67
	⊠	mae	0.13	2.33	0.15	0.30
		rms	0.17	3.21	0.20	0.39
		corr	0.93	0.93	0.87	0.62

mae: mean absolute error, rms: root mean squared error,
corr: Pearson correlation coefficient,
wd: weight decay/regularization to reduce feature activation.

The highest prediction quality in terms of correlation coefficient is achieved for the prediction of V_{oc}, followed by η or J_{sc}, and, finally, FF for all test cases.

1) Designed Features: Considering the broad test spectrum, the designed feature approach from Section III-D achieves high prediction quality for V_{oc} and η with mean absolute errors (MAEs) of 2.73 mV and 0.16%$_{abs}$, respectively, for a model with weight decay. Adding weight regularization during the training of the regression model slightly improves the prediction quality, as the model complexity is reduced to avoid an overfitting to the training set. The results for J_{sc} and FF show larger errors.

2) DenseNet: The proposed end-to-end approach improves the traditional feature-engineering approach for the prediction of all four parameters. For "unknown" test bricks (Scenario 1), the MAE for the prediction of the V_{oc} and η is as low as 2.05 mV and 0.11%$_{abs}$ for PL image resolutions of 264×264 pixels and random cropping.

Without random cropping, the prediction quality slightly reduces to 2.12 mV and 0.12%$_{abs}$ for V_{oc} and η prediction, respectively.

Interestingly, a reduction of the image resolution to 134×134 pixels still leads to a high prediction quality. The MAE for V_{oc} prediction slightly increased to 2.15 mV and for η prediction to 0.12%$_{abs}$. Applying weight decay, errors increase to 2.33 mV for V_{oc} and 0.13%$_{abs}$ for η.

The correlation graphs for PL images with a resolution of 134×134 pixels and without weight decay are shown for every other test wafer in Fig. 3.

B. Scenario 2: V_{oc} Prediction for Unknown Manufacturers

The prediction results for each of three HPMC-Si manufacturers (M1, M2, and M3) are heterogeneous, as presented in Table II. During the training of Scenario 2, the models are optimized for V_{oc} prediction only. The PL image resolution was 134×134 pixels with random cropping of 6 pixels.

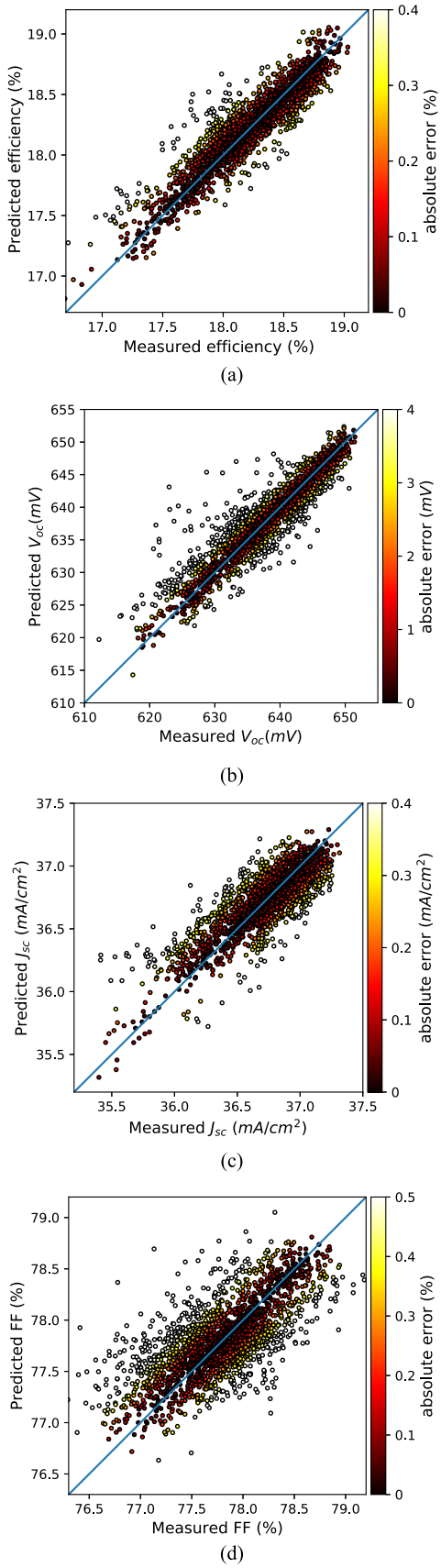
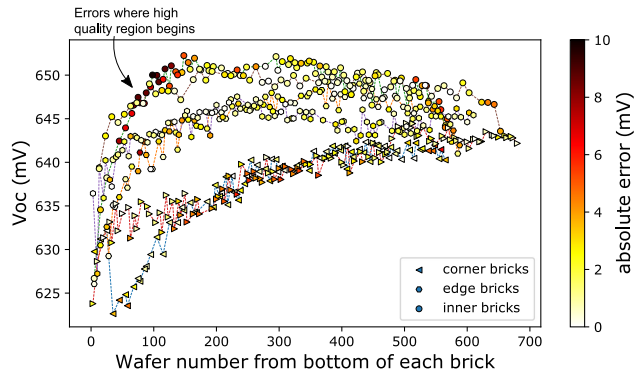


Fig. 3. Measured and predicted solar cell parameters (Scenario 1) based on PL images with a resolution of 134×134 pixels. (a) Efficiency. (b) Open-circuit voltage. (c) Short-circuit current. (d) Fill factor.

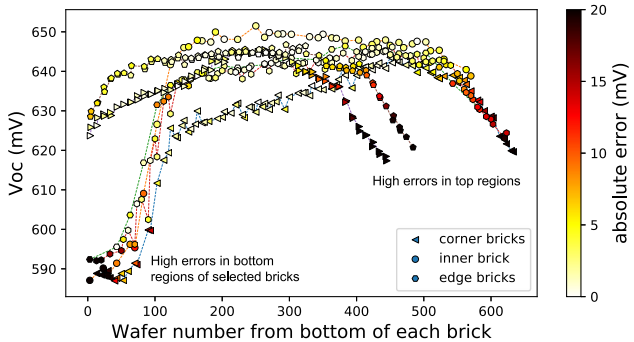
TABLE II
(SCENARIO 2) DENSENET V_{OC} PREDICTION FOR HPMC-Si MANUFACTURERS NOT IN THE TRAINING SET

Manufacturer	Ingots #	Bricks #	Wafers #	mae $[mV]$	rms $[mV]$	cor [-1,1]
M1	2	8	685	1.87	2.5	0.91
M2	1	4	182	1.64	2.19	0.94
M3*	1	6	434	5.09	8.37	0.84

*contains wafers from the very top and bottom brick regions.



(a)



(b)

Fig. 4. Development of measured V_{oc} values along brick height for HPMC-Si wafers from manufacturers M1 and M3. Prediction errors (colors) correspond to models trained without data from the shown manufacturer. The wafer number does not indicate the same absolute height in the ingot for wafers from different bricks, due to variations of the cropping height for each brick. (a) HPMC-Si manufacturer (M1). (b) HPMC-Si manufacturer (M3).

The MAE is as low as 1.87 mV for M1 and 1.64 mV for M2, but very high for M3 with the MAE of 5.09 mV. The very high root-mean-squared (rms) error for M3 indicates a strong variation in prediction accuracy for the samples of this manufacturer.

We analyze the errors for samples from M1 and M3 with respect to wafer brick position. Fig. 4(a) and (b) shows the measured V_{oc} as a function of the wafer number beginning from the bottom of the brick after cropping. The absolute prediction error is given by the color value.

The investigated PERC process is sensitive to contamination from the crucible, as shown in Fig. 4(a). Three different streams with different development in material quality can be observed for corner, edge, and inner bricks. The V_{oc} is low for corner bricks, higher for edge and very high for inner bricks.

1) Errors for M1 [see Fig. 4(a)]: Errors for samples from M1 are located near steep quality changes for few high-quality samples, for samples of corner bricks, and for samples at the top region of the brick. For all bricks of M1, we observe a steep increase in material quality in the first part of the brick. The highest V_{oc} values are achieved for inner bricks after wafer position 100. Especially, within the upper part of this strong slope, the model underestimates the material quality up to 10 mV.

2) Errors for M3 [see Fig. 4(b)]: Errors for samples from M3 can be observed in the very bottom regions of the bricks. Selected bricks from M3 were intentionally cropped near the bottom of the ingot leading to measured V_{oc} values as low as 590 mV. The samples are classified as low-quality material by the model with predicted V_{oc} values around 600 mV. Considering the very low quality in this brick region, the comparable large numerical error can be neglected for these outliers. Assuming a sorting of wafers within an incoming control, samples within this quality range are likely to end up in the lowest quality bin.

High prediction errors can be found in the top region of each brick from M3. In opposite to samples from M1, a steep quality decline can be observed in the top region of each brick from M3. Errors are especially high for wafers from this region due to an overestimation of the material quality by the model, which will be discussed in the following section.

C. Prediction Errors Due to Impurities

As shown in Fig. 4(b), a steep quality decline is observed for materials from the top region of bricks from M3. Interestingly, the PL intensities and PL structures do not change significantly for these samples, as shown in Fig. 5. The PL images in the left column are taken from different brick heights from the top region of the brick and show a slight increase in dislocations from sample position 375 (first row) up to position 445 (third row), which is the topmost sample of the brick.

We have a look at the measured and predicted V_{oc} values, which are annotated above each PL image. In accordance with the additional defect structures in sample 445 compared with sample 375, the predicted V_{oc} decreases by 5.5 mV. In contrast with our prediction, the measured V_{oc} decreases actually by 18.7 mV from 636.1 to 617.4 mV for these samples, leading to high prediction errors.

The deterioration may be referred to back diffusion of impurities during crystallization. For a closer look on these defects, we determine the iron point defect concentration according to Schubert *et al.* [37]. As shown in the right column of Fig. 5, the interstitial iron concentration Fe_i increases with the brick height. The mean value $Fe_{i,mean}$ is annotated above each sample. It increases from $4.1e^{10}$ cm^{-3} for sample at position 375 to $8.8e^{10}$ cm^{-3} for the topmost sample of this brick at position 445. The increase in metastable defects is not observed in these band-to-band PL images only, which can explain the prediction errors. In addition, an increase of the doping concentration in the top region of the brick has been observed. As the PL signal depends on the $n \cdot p$ product, the increased doping concentration can explain a higher PL intensity despite metastable defects.

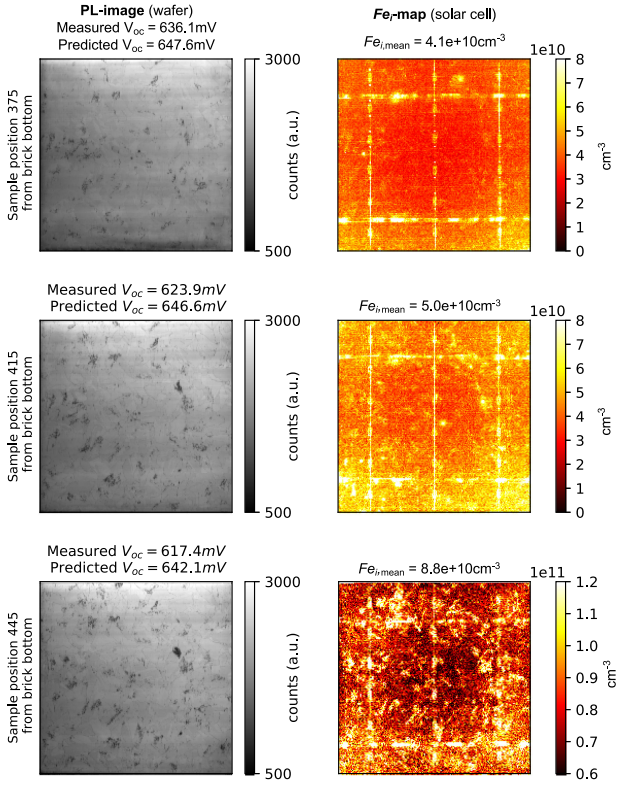


Fig. 5. Pairs of (left) PL image of the as-cut wafer and (right) the image of the iron point defect concentration [37] measured at the finished cell. The samples are selected from different regions near the top of the brick up to the topmost sample in the last row. The measured and predicted V_{oc} values, as well as the mean interstitial iron concentration Fe_i, are annotated. Note the different scales of the Fe_i maps.

V. Discussion and Outlook

Both prediction methods, i.e., the feature-based approach described in Section III-D and the network model from Section II, achieve high prediction quality considering the broad material spectrum.

The network approach improves previous feature-based approaches to predict the quality of silicon wafers for PERC solar cells. The CNN is capable to predict the quality of samples from unknown bricks, as well as samples from an unknown manufacturer with very high accuracy within only a few microseconds.

The model can be extended by adding complementary measurements to improve the prediction quality. So far, an error-free detection cannot be expected, as the PL measurements, used as input data, primarily reveal structural crystallization defects and provide only limited information on the electrical quality in the bulk of the grains. This is because the PL signal in as-cut wafers is strongly limited by surface recombination. Multiple complementary measurements are available for wafer inspection, e.g., the bulk lifetime measured on bricks [3] or spatially resolved grain boundary information [13]. Single parameters or even images can be integrated in the network with few modifications only.

Despite downsampling and cropping of PL images, high-quality predictions can be achieved. The PL images are downsampled before they are feed into the network due to memory constraints of the GPU. The network uses only a part of the information given by most commercially available PL systems. Nevertheless, PL images with very low image resolution show better performance than the feature-engineering approach. In addition, the augmentation step removes information during the training. Yet, models with cropping lead to better prediction results than models without cropping. The random cropping avoids an overfitting of the model to the training data.

The prediction error increases for J_{sc} and FF, as these parameters show a stronger dependence on quality parameters of various steps in the solar cell process, e.g., texturing, contact definition, and geometry. Nevertheless, material properties can influence the quality of process steps, e.g., structural defects have impact on the texturing result [41], and surface roughness can influence contact formation [42]. For a validation of the single-model outputs, we compare the directly predicted efficiency with a calculated efficiency based on the product of predicted V_{oc}, J_{sc}, and FF. The comparison shows an excellent agreement with Pearson's correlation coefficient of 0.99 between both quantities.

For human-designed features, all four regression models are based on the same feature set. The feature coefficients can be compared according to the relevance for the prediction result [13]. Due to the knowledge about the underlying physical defect, these can be interpreted by crystal growers and solar cell manufacturers similar to the analysis in [43]. For a physical interpretation of the deep neural features, further effort is necessary. The presented visualization techniques in [28] analyze the data in more detail. For example, activation maps [26] reveal the spatial distribution of the learned prediction value. They show that regions of reduced V_{oc} correspond to regions with crystallographic defect structures. These techniques are a step toward a transparent learning model for material rating.

VI. Conclusion

In this paper, a machine learning approach for material rating of HPMC-Si and mc-Si wafers for solar cell production has been introduced. A CNN has been trained to predict the I–V parameters based on PL images of as-cut wafers. Our end-to-end model improves existing feature engineering methods for the prediction of the solar cell performance. For materials of "unknown" bricks, the V_{oc} and η were predicted with MAEs as low as 2.05 mV and 0.11%, respectively. The models perform well even for PL images with lowered resolution.

Even a prediction of materials from "unknown" manufacturers showed excellent results for two of three manufacturers with HPM materials. High prediction errors are observed for one manufacturer with high concentration of iron point defects.

In contrast with previous methods, our approach does not rely on human-designed features. Instead, a huge variety of materials formed the database for the successful rating.

Despite the high prediction quality, CNNs are observed as less transparent than human-designed methods. We overcome this perceived "blackbox" behavior: Our network design allows the visualization of what has been learned by the model, which is presented in a second part of this research in [28].

REFERENCES

[1] T. Trupke, R. Bardos, M. Schubert, and W. Warta, "Photoluminescence imaging of silicon wafers," *Appl. Phys. Lett.*, vol. 89, no. 4, 2006, Art. no. 044107.

[2] H. Wagner, M. Müller, G. Fischer, and P. Altermatt, "A simple criterion for predicting multicrystalline Si solar cell performance from lifetime images of wafers prior to cell production," *J. Appl. Phys.*, vol. 114, no. 5, 2013, Art. no. 054501.

[3] B. Mitchell *et al.*, "PERC solar cell performance predictions from multicrystalline silicon ingot metrology data," *IEEE J. Photovolt.*, vol. 7, no. 6, pp. 1619–1626, Nov. 2017.

[4] J. Haunschild *et al.*, "Quality control of as-cut multicrystalline silicon wafers using photoluminescence imaging for solar cell production," *Sol. Energy Mater. Sol. Cells*, vol. 94, no. 12, pp. 2007–2012, 2010.

[5] M. A. Green, "The passivated emitter and rear cell (PERC): From conception to mass production," *Sol. Energy Mater. Sol. Cells*, vol. 143, no. Suppl. C, pp. 190–197, 2015.

[6] M. Demant *et al.*, "Inline quality rating of multi-crystalline wafers based on photoluminescence images," *Prog. Photovolt.: Res. Appl.*, vol. 24, no. 12, pp. 1533–1546, 2015.

[7] C. W. Lan *et al.*, "Grain control in directional solidification of photovoltaic silicon," *J. Cryst. Growth*, vol. 360, no. 0, pp. 68–75, 2012.

[8] Y. M. Yang *et al.*, "Development of high-performance multicrystalline silicon for photovoltaic industry," *Prog. Photovolt.: Res. Appl.*, vol. 23, no. 3, pp. 340–351, 2015.

[9] J. Hofstetter *et al.*, "Sorting metrics for customized phosphorus diffusion gettering," *IEEE J. Photovolt.*, vol. 4, no. 6, pp. 1421–1428, Nov. 2014.

[10] B. Birkmann, A. Hüsler, A. Seidl, K. Ramspeck, and H. Nagel, "Analysis of multicrystalline wafers originating from corner and edge bricks and forecast of cell properties," in *Proc. 26th Eur. Photovolt. Sol. Energy Conf. Exhib.*, 2011, pp. 937–940.

[11] M. Demant, M. Glatthaar, J. Haunschild, and S. Rein, "Analysis of luminescence images applying pattern recognition techniques," in *Proc. 25th Eur. Photovolt. Sol. Energy Conf. Exhib.*, 2010, pp. 1078–1082.

[12] T. Trupke and R. A. Bardos, "Wafer imaging and processing method and apparatus," U.S. Patent 9 546 955, Jan. 17, 2017.

[13] A. Kovvali, M. Demant, T. Trötschler, J. Haunschild, and S. Rein, "About the relevance of features in as-cut multicrystalline silicon wafers on solar cell performance," in *Proc. 8th Int. Conf. Crystalline Silicon Photovolt.*, Lausanne, Switzerland, 2018, Art. no. 130011.

[14] M. Demant, J. Greulich, M. Glatthaar, J. Haunschild, and S. Rein, "Modelling of physically relevant features in photoluminescence images," *Energy Procedia*, vol. 27, pp. 247–252, 2012.

[15] M. Demant *et al.*, "Evaluation and improvement of a feature-based classification framework to rate the quality of multicrytalline silicon wafers," in *Proc. 28th Eur. Photovolt. Sol. Energy Conf. Exhib.*, 2013, pp. 1650–1654.

[16] S. A. McHugo, H. Hieslmair, and E. R. Weber, "Gettering of metallic impurities in photovoltaic silicon," *Appl. Phys. A*, vol. 64, no. 2, pp. 127–137, Jan. 1997.

[17] A. Bentzen and A. Holt, "Overview of phosphorus diffusion and gettering in multicrystalline silicon," *Mater. Sci. Eng.: B*, vol. 159, pp. 228–234, 2009.

[18] D. P. Fenning *et al.*, "Precipitated iron: A limit on gettering efficacy in multicrystalline silicon," *J. Appl. Phys.*, vol. 113, no. 4, 2013, Art. no. 044521. [Online]. Available: https://doi.org/10.1063/1.4788800

[19] S. Castellanos *et al.*, "High-performance and traditional multicrystalline silicon: Comparing gettering responses and lifetime-limiting defects," *IEEE J. Photovolt.*, vol. 6, no. 3, pp. 632–640, May 2016.

[20] S. Bernardini *et al.*, "Nano-XRF analysis of metal impurities distribution at PL active grain boundaries during mc-silicon solar cell processing," *IEEE J. Photovolt.*, vol. 7, no. 1, pp. 244–249, Jan. 2017.

[21] S. Johnston *et al.*, "Correlating multicrystalline silicon defect types using photoluminescence, defect-band emission, and lock-in thermography imaging techniques," *IEEE J. Photovolt.*, vol. 4, no. 1, pp. 348–354, Jan. 2014.

[22] J. Isenberg *et al.*, "Correlation of spatially resolved lifetime measurements with overall solar cell parameters," in *Proc. Conf. Rec. 29th IEEE Photovolt. Spec. Conf.*, May 2002, pp. 198–201.

[23] B. Sopori and W. Chen, "Influence of distributed defects on the photoelectric characteristics of a large-area device," *J. Cryst. Growth*, vol. 210, no. 1, pp. 375–378, 2000.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 26th Annu. Conf. Neural Inf. Process. Syst.*, Lake Tahoe, NV, USA, Dec. 3–6, 2012, pp. 1106–1114.

[25] B. Michl *et al.*, "Efficiency limiting bulk recombination in multicrystalline silicon solar cells," *Sol. Energy Mater. Sol. Cells*, vol. 98, pp. 441–447, 2012.

[26] B. Zhou, A. Khosla, L. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2921–2929.

[27] L. van der Maaten and G. E. Hinton, "Visualizing high-dimensional data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.

[28] M. Demant, P. Virtue, A. Kovvali, S. X. Yu, and S. Rein, "Visualizing material quality and similarity of mc-Si wafers learned by convolutional regression networks," *IEEE J. Photovolt.*, to be published, doi: 10.1109/JPHOTOV.2019.2906037.

[29] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2261–2269.

[30] J. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," in *Proc. Int. Conf. Learn. Represent.*, 2015, arXiv:1412.6806.

[31] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, arXiv:1409.1556.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 770–778.

[34] H. Höffler *et al.*, "Comparison of line-wise PL-imaging and area-wise PL-imaging," *Energy Procedia*, vol. 124, no. Suppl. C, pp. 66–75, 2017.

[35] D. Macdonald, J. Tan, and T. Trupke, "Imaging interstitial iron concentrations in boron-doped crystalline silicon using photoluminescence," *J. Appl. Phys.*, vol. 103, no. 7, 2008, Art. no. 073710.

[36] G. Zoth and W. Bergholz, "A fast, preparation-free method to detect iron in silicon," *J. Appl. Phys.*, vol. 67, no. 11, pp. 6764–6771, 1990.

[37] M. C. Schubert, H. Habenicht, and W. Warta, "Imaging of metastable defects in silicon," *IEEE J. Photovolt.*, vol. 1, no. 2, pp. 168–173, Oct. 2011.

[38] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 675–678.

[39] M. Demant, *Quality Rating of Silicon Wafers—A Pattern Recognition Approach*. Stuttgart, Germany: Fraunhofer-Verlag, 2016.

[40] T. Hastie *et al.*, *Linear Methods for Regression*, vol. 2. New York, NY, USA: Springer, 2009, ch. 3, pp. 43–100.

[41] J. Nievendick *et al.*, "Appearance of rift structure created by acidic texturization and their impact on solar cell efficiency," in *Proc. 35th IEEE Photovolt. Spec. Conf.*, 2010, pp. 447–452.

[42] A. Lorenz *et al.*, "Impact of texture roughness on the front-side metallization of stencil-printed silicon solar cells," *IEEE J. Photovolt.*, vol. 5, no. 4, pp. 1237–1244, Jul. 2015.

[43] T. Strauch, M. Demant, P. Krenckel, S. Riepe, and S. Rein, "Analysis of grain structure evolution based on optical measurements of mc Si wafers," *J. Cryst. Growth*, vol. 454, no. Suppl. C, pp. 147–155, 2016.

Matthias Demant received the Ph.D. degree in engineering sciences from the University of Freiburg, Freiburg, Germany, in 2016, in cooperation with the Fraunhofer Institute for Solar Energy, Freiburg.

As a Postdoctoral Researcher, he joined the Vision Group, International Computer Science Institute, University of California, Berkeley, CA, USA, in 2017. Since 2019, he has been heading a team for the development of machine learning techniques for material characterization and process control in the field of renewable energy with the Fraunhofer Institute for Solar Energy.

Patrick Virtue received the B.S. degree in computer science engineering from the University of Notre Dame, Notre Dame, IN, USA, in 2002. He is working toward the Ph.D. degree in electrical engineering and computer sciences, specializing in deep learning, computer vision, and image processing, with the University of California, Berkeley, CA, USA.

Prior to graduate school, he researched and developed medical image applications as a Software Engineer with GE Healthcare. He has applied his work to a range of applications from medical imaging to military target detection.

Aditya Kovvali is working toward the Ph.D. degree with Albert Ludwig University, Freiburg, Germany.

He is currently with the Department of Quality Assurance, Characterization and Simulation, Fraunhofer Institute for Solar Energy, Freiburg. His work specializes in data-driven approaches for material assessment in photovoltaics.

Stella X. Yu received the Ph.D. degree from Carnegie Mellon University, Pittsburgh, PA, USA, in 2003, where she studied robotics with Robotics Institute, and vision science with the Center for the Neural Basis of Cognition.

She continued her computer vision research as a Postdoctoral Fellow with the University of California, Berkeley, CA, USA, and then studied art and vision as a Clare Boothe Luce Professor with Boston College. She is currently the Director of the Vision Group, International Computer Science Institute and a Senior Fellow with the Berkeley Institute for Data Science, University of California, Berkeley. Her research interests include understanding visual perception from multiple perspectives and using computer vision and machine learning to automate and exceed human expertise in practical applications.

Dr. Yu is a recipient of the National Science Foundation CAREER Award.

Stefan Rein studied physics with the University of Freiburg, Freiburg, Germany. He completed the Diploma thesis in the field of defect characterization in semiconductor materials with the Fraunhofer Institute for Solar Energy (Fraunhofer ISE), Freiburg. He received the Ph.D. degree from the University of Konstanz, Konstanz, Germany, in 2004, for his research work on "Lifetime Spectroscopy" done with Fraunhofer ISE.

In 2005, he joined the Industrial Cell Production Group, Fraunhofer ISE, and accompanied design and ramp-up of the Photovoltaic Technology Evaluation Center. He is the Head of the Department "Quality assurance, characterization and simulation," Fraunhofer ISE. His scientific research interests include defects in crystalline silicon, material and solar cell characterization, development of measurement techniques, and novel approaches for production control.