Deep Learning-Based Multi-Step Solar Forecasting For PV Ramp-Rate Control Using Sky Images

Haoran Wen, Yang Du, Xiaoyang Chen, Eng Gee Lim, Huiqing Wen, Lin Jiang and Wei Xiang

***Abstract*—Solar forecasting is one of the most promising approaches to address the intermittent PV power generation by providing predictions before upcoming ramp events. In this paper, a novel multi-step forecasting (MSF) scheme is proposed for PV power ramp-rate control (PRRC). This method utilizes an ensemble of deep ConvNets without additional time-series models (e.g., RNN or LSTM) and exogenous variables, thus more suitable for industrial applications. The MSF strategy can make multiple predictions in comparison with a single forecasting point produced by a conventional method while maintaining the same high temporal resolution. Besides, stacked sky images that integrate temporal-spatial (ST) information of cloud motions are used to further improve the forecasting performance. The results demonstrate a favorable forecasting accuracy in comparison to the existing forecasting models with the highest skill score of 17.7%. In the PRRC application, the MSF-based PRRC can detect more ramp-rates (RR) violations with a higher control rate of 98.9% compared with the conventional forecasting-based control. Thus, the PV generation can be effectively smoothed with less energy curtailment on both clear and cloudy days using the proposed approach.**

***Index Terms*—solar forecasting, deep learning, stacked sky images, multi-step forecasting (MSF), power ramp-rate control (PRRC).**

I. Introduction

SOLAR energy is one of the most promising renewable energies to replace fossil fuels to mitigate the effects of global warming [1]. However, the increasing integration of photovoltaic (PV) generation continuously challenges the operations of power grids due to its intermittent nature [2]. The PV power generation relies on both intrinsic and extrinsic factors such as cell materials, temperature, and solar irradiance intensity [3]. Among these factors, solar irradiance, which is converted to electricity by PV cells, varies momentarily and substantially depending on different locations and weather patterns [4]. This gives rise to large output power ramp-rates causing grid frequency problems [5], which leads to further concerns to grid stability [6]. To overcome this issue, the utilities have imposed regulations on ramp-rate. Recent works on PV power ramp-rate control seek to effectively smooth its intermittent generation such that the grids suffer less from high RRs. Energy storage system (ESS) is a direct approach to smooth distributed PV power by charging and discharging processes to compensate power variations [7]. Yet the high cost of ESS hinders its prevalent application. Inverter-level active power curtailment (APC) has been introduced to mitigate the required ESS capacities deliberately curtailing PV power [8]. However, it can only be applied to the ramp-up sides, while the ramp-downs are still needed to be compensated by the ESS.

Recently, solar forecasting has been extensively studied to cope with this issue by providing advanced operations for the utilities before upcoming ramp events. The existing solar forecasting techniques can be classified into five categories: regression models, Machine Learning (ML), image-based methods, Numeric Weather Prediction (NWP) and hybrid approaches [9]. The NWP models simulate the irradiance dynamics in the atmosphere [10] and outperform the satellite forecasts at hourly forecasting horizons, but unable to predict the cloud position due to the limited geographic resolution [11]. Regression and machine learning models both utilize historical data of solar irradiance as their inputs to make empirical predictions. The most commonly applied models include artificial neural networks (ANNs) [12], support vector machine (SVM) [13]. A hybrid model [14] combines the merits of different methods to improve forecasting accuracy. However, the forecast horizons of the abovementioned methods are from hours to days, which is far too large for the PRRC application.

In terms of short-term forecasting methods, a ground-based sensor network (GBSN) that aggregates both temporal and spatial information can detect ramp events in very short horizons of several minutes with high temporal resolutions [15]. Depending on how the cloudy information is coupled, the spatial-temporal correlation-based model calculates the maximum cross-correlation of the lagged time (temporal) among adjacent sensors (spatial), while the cloud shadow vector (CSV) tracking-based method maps the power generation to the CSV velocity [16]. However, the GBSN method suffers from varied forecast horizons caused by different cloud speed [17]. Although [18] proposes a fixed-horizon forecasting method, the required sensors, and geographic dispersions increase significantly as the forecast horizon expands.

Manuscript received XX, 2020; revised XX; accepted XX. Date of publication XX; date of current version XX. This work was supported by the National Natural Science Foundation of China (No. 61803315), Jiangsu Science and Technology Program (SBK2018042034), Key Program Special Fund in XJTLU (KSF-A-11), research development fund of XJTLU (RDF-17-01-28) and AI University Research Centre (AI-URC) through XJTLU Key Programme Special Fund (KSF-P-02). Paper no. TII-20-0876. (Corresponding author: Yang Du.)

H. Wen, X. Chen, E. G. Lim and H. Wen are with the Department of Electrical and Electronics Engineering, Xi'an-Jiaotong Liverpool University, Suzhou 215123, China (e-mail: haoran.wen18@student.xjtlu.edu.cn; xiaoyang.chen@xjtlu.edu.cn; enggee.lim@xjtlu.edu.cn; huiqing.wen@xjtlu.edu.cn)

Y. Du and W. Xiang are with College of Science and Engineering, James Cook University, QLD 4870, Australia (e-mail: yang.du@jcu.edu.au; wei.xiang@jcu.edu.au)

L. Jiang is with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3BX, U.K. (e-mail: L.jiang@liverpool.ac.uk)

Alternatively, sky image (SI)-based forecasting strategies provide regional solar forecasts over PV power plants with more flexible forecasting horizons from several minutes to hours by using a single sky camera. The SI-based forecasting methods can be further categorized into physical models and data-driven methods. The physical models include estimation of cloud motion, cloud detection, and cloud classification [19]. [13] utilizes the classified solar irradiance data based on different locations to adaptively adjust the perturbation step of a maximum power point tracking (MPPT) algorithm. A battery-less PV power smoothing method using a sky camera is introduced by [20], where cloud motions are estimated by calculating principal cloud motion vectors.

As for the data-driven models, the feature extraction process is usually involved, where the extracted features are used as predictors for the forecasting models. In [21], an ensemble of five different neural networks with meteorological variables (e.g., wind speed, temperature, and humidity) is used to forecast day-ahead PV output power. [22] uses Red, Green, and Blue (RGB) values, Hue and saturations as the inputs for the multilayer perceptrons (MLPs), which achieve skill scores of 10.1% in global horizontal irradiance (GHI). A recent work [23] deploys multiple convolutional neural networks (ConvNets or CNNs) as feature extractors to transform low-level features of images into high-level latent space. These condensed latent vectors are then forward to several Long-Short Term Memory (LSTM) models that encode temporal information from all past latent vectors. This cascaded structure largely increases the training complexity, number of parameters, and training time.

Comparing with the GBSNs, the SI-based methods require no extra land or other resources, thus more suitable for industrial applications. However, the forecasting accuracy of SI-based models are relatively low since the predictions are empirically made by inferring sky images through a particular model, other than the direct irradiance measurements by sensors in the GBSN. Moreover, the SI-based methods can only produce one single-step forecasting (SSF) point each time, thus the temporal resolution of the forecasts is limited by both forecast horizons and temporal resolution of input images. For example, a model of a 5-min horizon can only produce a single time-step 5 min ahead, and information within the 5 min is not known. For a higher forecast resolution, the model requires the input images with a higher temporal resolution, which results in an increased computational cost and storage demand.

In this paper, instead of using physical models or explicitly extracting features from sky images, we train deep ConvNets to automatically learn the relationship between the historical GHIs and different clouds patterns in the sky images (e.g., clear, moderate cloudy and overcast days). Moreover, an MSF approach is proposed to generate 10 min ahead forecasts at 1-min resolution, with higher accuracy and robustness compared with the conventional approaches. The contributions of this work can be summarized as follows:

1) We propose a multi-step forecasting strategy to forecast GHI values, which is able to provide multiple predictions within the forecast horizon of higher temporal resolution and forecasting accuracy than single-step methods.
2) The stacked sky images that integrate both spatial and temporal information can further improve forecasting accuracy.
3) The proposed MSF strategy can be applied to PV power ramp-rate control, and it results in lower control failure rate and reduced energy curtailment.

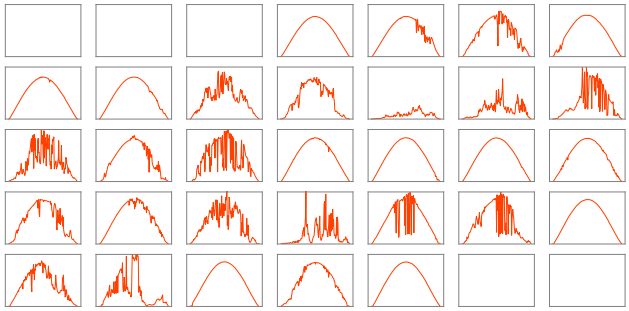


Fig. 1. A GHI calendar of April 2016 at Folsom CA, USA. It contains 23,515 data samples of different weather patterns that reflects the daily GHI variations. The data are used as the test set for our forecasting models.

The remaining of the paper is organized into the following five sections. After the description of the datasets and data pre-processing in the experiments in section II, deep leaning-based solar forecasting and the proposed MSF strategy are introduced in section III. Section IV demonstrates the basic PRRC principal and MSF-based PRRC strategy. After that, the experimental results are presented and discussed in section V and followed by the conclusion in section VI.

II. Data Description

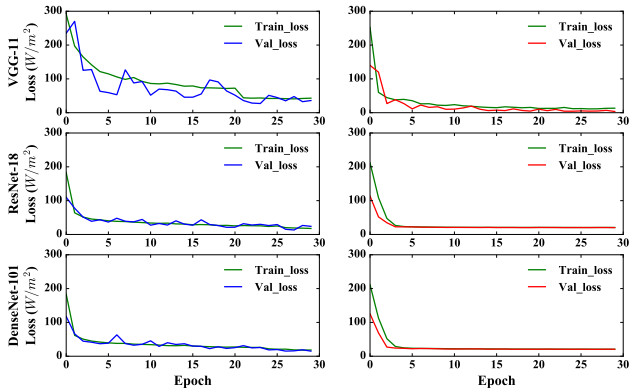
The experiments in the following sections are conducted on two publicly available datasets of sky images obtained at two locations in the United States.

A. Golden, CO Dataset

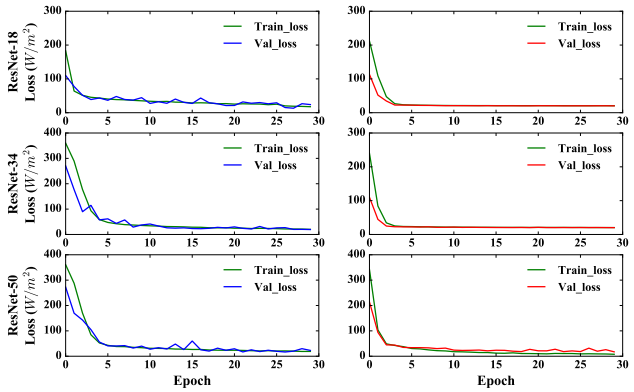
The Solar Radiation Research Laboratory (SRRL) of the National Renewable Energy Laboratory (NREL) has developed a renewable resource climatology for NREL at Golden, Colorado. It provides a variety of measurements of solar radiation and other meteorological information. A Yankee Total Sky Imager (TSI-880) is used to capture all-sky images with the cloud cover values recorded every 10 min. The images available from January 2018 to April 2019 are used for training, validation, and testing [24].

B. Folsom, CA Dataset

The sky images and solar irradiance in the second dataset are obtained by a fish-eye lens camera and a rotating shadow band radiometer (RSR) installed on a roof at Folsom, CA $(38.642°, -121.148°)$ [19]. This dataset provides more than half-million images from 2014 to 2016 with an average interval of 1 min, which makes it possible for a high forecast resolution at a 1-min time step. Considering the seasonal



(a) Loss curves of VGG-11, ResNet-18 and DenseNet-121



(b) Loss curves of ResNet-18, ResNet-34, ResNet-50

Fig. 2. Training and validation losses concerning: (a) different CNN architectures to select forecasting model; (b) ResNet variations of different numbers of hidden layers to investigate the impact of model depth. All the models are trained using both SGD optimizer (blue curves) and Adam Optimizer (red curves).

TABLE I
NUMBERS OF DATA SAMPLES IN TRAIN, VALIDATION AND TEST SET FOR BOTH GOLDEN AND FOLSOM DATASET

Dataset	Task	Train	Validation	Test	Period
Golden	Regression	28,307	3,836	3,409	Jan, 2018 - Apr, 2019
Folsom	Forecast	260,765	57,292	23,515	Jan, 2015 - Apr, 2016

variation of the irradiance, we use more than one year's historical measurements from January 2015 to April 2016 for train, validation, and test sets. April 2016 is selected for testing the model including 12 clear days and 18 cloudy days as shown in Fig. 1. The splits of the datasets for both regression and forecasting tasks are shown in Table I.

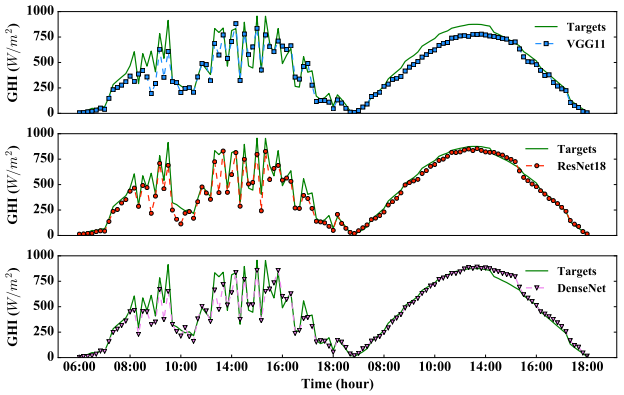


Fig. 3. Regression results of VGG-11, ResNet-18 and DenseNet-121 of two consecutive days on 19 and 20 March at Golden.

C. Data Pre-processing

Given an image set $\mathscr{X} = \{x_{t_0}, x_{t_1}, \ldots, x_{t_i}, \ldots, x_{t_n}\}$, and GHI values $\mathscr{Y} = \{y_{t_0}, y_{t_1}, \ldots, y_{t_i}, \ldots, y_{t_n}\}$, the dataset $\mathscr{D} = \{(x_{t_0}, y_{t_0}), (x_{t_1}, y_{t_1}), \ldots, (x_{t_i}, y_{t_i}) \ldots, (x_{t_n}, y_{t_n})\}$ is constructed by mapping the images with the GHI values according to their time stamps, where $x_{t_i} \in \mathbb{R}^{W \times H}$ and $y_{t_i} \in \mathbb{R}$ are respectively the image and the label in a data sample at time t_i, n is the size of the dataset. The images are in a shape of $W \times H \times C$, which represent for width, height and the number of color channels. The term "size" will only be referred as width and height in the following sections since they are three-channel RGB images. Before loading the images into the model, the original images of sizes 1500×1500 are resized to a uniform size of 256×256 using Python's OpenCV library to accelerate data loading process. The loaded images are converted to tensors of a dimension of $224 \times 224 \times 3$ by the data loader to fit the model input, and each pixel is normalized within range 0 to 1 to make the convergence faster while training the network.

III. DEEP LEARNING-BASED SOLAR FORECASTING

A. Model Selection

Deep learning (DL) is a sub-branch of machine learning and deliberately designed to tackle various difficult vision-related tasks. CNN, as one of the most popular and widely applied DL models, performs mathematical convolution operations with the use of kernels or filters on graphs and produces high-level feature maps.

In this section, we have compared three different DL models, namely VGGNet [25], ResNet [26], DenseNet [27] as well as the impact of different number of hidden layers. The models are pre-trained models available from PyTorch library, and all of the parameters are updated by continuing the backpropagation on the our datasets for the new task. These models all have deep structures in common but differ from the way how their layers are connected. Although it has been mathematically proven that deeper neural networks have more representational ability [28], the shallow versions of each model (VGG-11, ResNet-18, and DensNet-121) are firstly employed for forecasting model selection, and then we compare the number of hidden layers to determine model depths.

All experiments are implemented under the PyTorch deep learning framework with NVIDIA 1080Ti GPUs on a server. The models are trained and tested on the same datasets for 30 epochs, during which the test set are isolated from the training process. The learning rate is initially set to 0.01

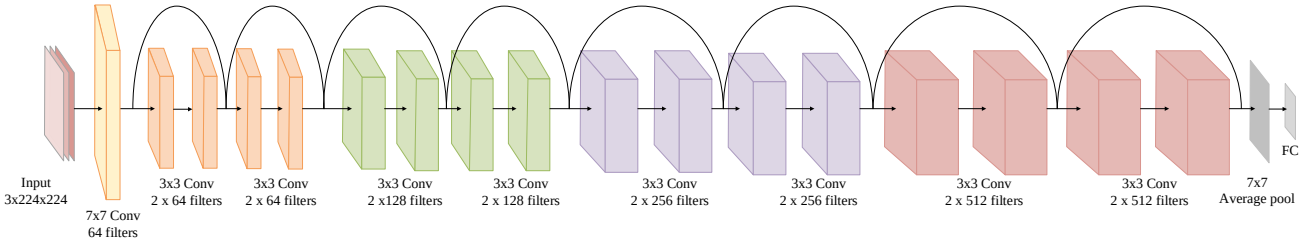


Fig. 4. ResNet-18 architecture for solar forecasting. There are 8 "residual-connected" building blocks, each of them contains 2 filters. The number of 3×3 kernels in each filter increases from 64 to 512 as going deeper. The max-pooling layers, ReLu activations and normalizations of the model are not shown in this Figure.

TABLE II
REGRESSION RESULTS OF DIFFERENT MODELS USING BOTH SGD AND ADAM OPTIMIZERS.

Model	RMSE (W/m^2)		MAE (W/m^2)		Time (s)		Memory (Mb)	
	SGD	Adam	SGD	Adam	SGD	Adam	SGD	Adam
VGG-11	96.3	83.4	65.0	**46.4**	662	654	5743	6433
DenseNet-121	**89.4**	**80.4**	**51.3**	49.1	726	788	6075	6035
ResNet-18	91.4	81.2	53.2	48.3	**384**	**339**	**1855**	**1948**
ResNet-34	90.5	80.9	52.9	49.7	421	438	2489	2085
ResNet-50	92.9	82.8	52.8	49.5	653	612	4285	4065

and automatically tuned by a Plateau learning rate scheduler that reduces the learning rate by a factor of 0.5 when the validation loss has stopped decreasing for over 5 epochs. Both Adam optimizer and Stochastic Gradient Descent (SGD) optimizer that used to find the global minima are compared. The loss curves are demonstrated in Fig. 2. The models are evaluated concerning their testing errors, training speed, and memory usage. The results are summarized in Table II and the forecasting results of two days are shown in Fig. 3.

According to the results, the Adam optimizer that is based on adaptive estimates of lower-order momentum outperforms the SGD optimizer in both testing errors and training time with a little increased memory usage. The loss curves of Adam optimizer shows a faster and more stable convergence process. In terms of models, the VGG-11 model has the worst performance; the DenseNet-121 and ResNet-18 have comparable test errors and convergence processes. However, the ResNet-18 uses much less computational resources and time on training.

To further investigate the impact of model depths, we repeated the above experiments using ResNet variations of different numbers of hidden layers. The performances of ResNet-18 and ResNet-34 are very close in all aspects and their loss curves both converge quickly and stopped decreasing after 4 epochs as shown in Fig. 2(b). As the number of hidden layers increases to 50, the performance of ResNet-50 degrades since slight overfitting occurs, where its validation loss becomes greater the training loss. As a result, the ResNet-18 is used as our base forecasting model.

B. Solar Forecasting Using ResNet-18 Model

Instead of using a cascaded architecture, we deploy stand-alone ResNet-18 models to forecast time-series solar irradiance. The core hypothesis is that the subtle patterns and relationships between cloud motions, sun position, and the irradiance variations are hardly formulated, but can be learned by deep CNNs. Besides, the ResNet architecture features "residual connections" and substantial use of batch normalization and 3×3 convolutional filters. The residual connections can greatly improve gradient flow, thus allowing the training of much deeper models with a large number of layers. An illustration of ResNet-18 architecture is shown in Fig. 4. ReLu activations are applied at the end of each layer with batch normalization. To perform forecasting tasks, Softmax loss is replaced by MSE loss and the fully-connected (FC) layer is reinitialized as a linear layer with 512 inputs and 1 output. The output vector is represented as $\tilde{Y} = [\tilde{y}_{t_0}, \tilde{y}_{t_1}, \ldots, \tilde{y}_{t_i}, \ldots, \tilde{y}_{t_n}]$ indicating the predicted GHI values. Fig. 5 illustrates a visualization of samples of features maps produced by each block of a trained ResNet-18, which gives a more intuitive representation of how a sky image is transformed into high-level abstractions.

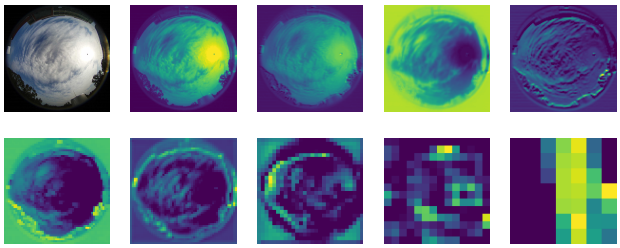


Fig. 5. Visualization of feature maps in hierarchical representations obtained from each block of a pre-trained ResNet-18 network (going deeper from top left to bottom right, row-wise).

The training process is the same as that of the regression tasks except for the use of backward-lagged data. For instance, an image x_{t_i} is labeled with a lagged GHI value $y_{t_{i+\mathscr{H}}}$ by a time interval length $\mathscr{H}$, which is the forecast horizon. However, the model initially suffered from low performance and its accuracy varied largely for different forecast horizons, which happened to the DenseNet and VGG as well. This is because that some images might be mistakenly labeled by lagging the GHI values. For example, at the moments when irradiance varies largely (the sun is to be blocked by clouds or clouds are leaving the sun), some clear sky images are mapped with very low GHI values or overcast images are mapped to high GHI values. These mistakenly labeled data would confuse the model convergence and reduce the forecasting accuracy.

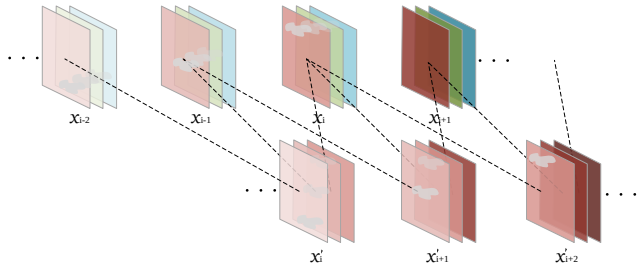


Fig. 6. Illustration of the image stacking process. The upper row is the original sky image set $\mathscr{X}$ in chronological sequence; below is the stacked image set $\mathscr{X}_r$ that consists of the red channels of the present image x_i and two previous images x_{i-1} and x_{i-2}.

Algorithm 1 Image Stacking

Input: Original image set: $\mathscr{X} = \{x_{t_0}, x_{t_1}, \ldots, x_{t_i}, \ldots, x_{t_n}\}$
Output: Stacked image set: $\mathscr{X}_s$

initial $n\_channel$ = 3, layers = $\{\varnothing\}$, t_i = 0
while len($\mathscr{X}$) $- t_i \geq n\_channel$ **do**
  for $i \in$ range($n\_channel$) **do**
    $temp\_x_{t_i}$ = cv2.imread(x_{t_i+i})
    layers.append($temp\_x_{t_i}$[:,:,2])
  end for
  $\tilde{x}_{t_i}$ = np.stack(layers, axis=-1)
  cv2.imwrite('root/to/dir', $\tilde{x}_{t_i}$)
  t_i += 1
end while
Return $\mathscr{X}_s = \{\tilde{x}_{t_0}, \tilde{x}_{t_1}, \ldots, \tilde{x}_{t_i}, \ldots, \tilde{x}_{t_{n-2}}\}$

To alleviate this adverse effect introduced by using the lagged data, the data are down-sampled by a factor of $\frac{1}{\mathscr{H}}$, by which the forecasting accuracy can be effectively improved.

C. Stacked Images

Although the stand-alone deep ConvNets can forecast solar irradiance by historical observations, the forecasting performance highly relies on a variety of factors such as data quality, data processing, model architecture, and training skill. Besides, the end-to-end forecasting process can be less interpretable and hard to be improved without prerequisite knowledge.

To further enhance the forecasting ability, a sequence of stacked images are used as inputs for training and testing the ResNet-18 model. Rather than using different exposures of images to capture surroundings (brightness) of the sun as previously done in [23], we use the stacked sky images that integrate ST information about the cloud motions for a relatively long time interval. This is pivotal in predicting future cloud variation and solar distribution since a stacked image is more capable of capturing the dynamics of passing cloud than a single frame.

To construct stacked images, we extract one certain color channel from each of the three consecutive images $x_{t_i}, x_{t_{i-1}}$ and x_{t_i-2}. The three extracted layers are chronologically stacked to form a new image denoted as $\tilde{x}_{t_i}$. This process is shown in Fig. 6. We then constructed and compared three stacked image sets $\mathscr{X}_r$, $\mathscr{X}_g$ and $\mathscr{X}_b$ using red, blue and green channels respectively. The model trained on $\mathscr{X}_r$ outperforms the others and the model on $\mathscr{X}_b$ has the worst forecasting accuracy. As a result, the red channel of each image is used for the image stacking. Moreover, the number of stacked channels in each image is limited by the model inputs since ResNet takes images of no more than three channels as inputs. Multiple-layer images that contain more information shall be constructed to further improve the conformance together with a DL model of customized architecture.

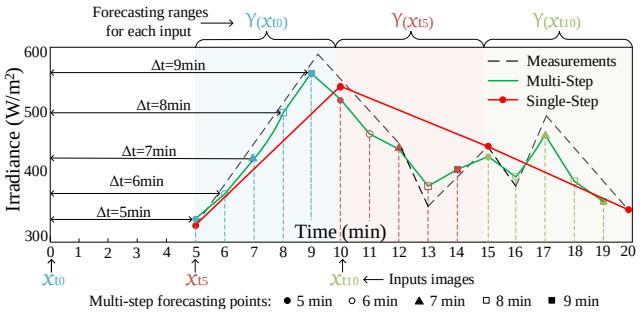


Fig. 7. A schematic representation of the proposed MSF method with three extended forecasting ranges.

D. Multi-Step Forecasting Strategy

The major application of solar forecasting is to assist the operation of power grid. According to [29], there are four parameters describing the temporal nature of solar forecasting, namely (1) forecast horizon ($\mathscr{H}$), (2) forecast resolution ($\mathscr{R}$), (3) forecast lead time ($\mathscr{L}$), and (4) forecast update rate ($\mathscr{U}$). However, a large number of existing algorithms ignore $\mathscr{R}$, $\mathscr{L}$ and $\mathscr{U}$. For instance, hourly forecasts are validated using only hourly data without lead time and intra-hour downscaling. In terms of DL-based solar forecasting, although models trained on images of smaller time intervals can provide smaller forecast resolutions, their forecast horizons are limited because a single model is not capable of generating multiple forecasts within a relatively long horizon.

The proposed MSF strategy provides a remedy to this problem by providing an extended forecast horizon with extra forecasts. The MSF model integrates a linear combination of five ResNet-18 models that are individually trained on datasets of different forecast horizons $\mathscr{H}$ from 5 min to 9 min to provide 1-min forecast resolution ($\mathscr{R}^{1\text{min}}$) for the control criteria. The forecasts are updated every 5 min ($\mathscr{U}^{5\text{min}}$). In practice, the temporal parameters depend on specific operation requirements of systems.

Fig. 7 is a schematic representation of the MSF process that illustrates the forecasts of the MSF model and an SSF model of $\mathscr{H}^{5\text{min}}$ from 5 min to 20 min. Initially, the image x_{t_0} is input to the models at time t_0 and predictions are made from 5 min to 9 min in the blue region and so forth. Although the SSF model (the red line) can ideally predict GHI at every forecasting points, a ramp-down event and a ramp-up event are not detected because of its low forecast resolution. For example, at 16 min the irradiance increases from $380W/m^2$ to $500W/m^2$ within one minute, which is failed to be predicted by the SSF model. Thus, the system is uncontrolled and more

energy (the area between the red line and the dashed line) is curtailed during this period. Therefore, the MSF method has a better overall performance of depicting GHI variation trends than the SSF model. The outputs of the MSF model can be represented by the following equation:

$$\tilde{\mathbb{Y}}_{\text{MSF}} = \begin{matrix} \mathscr{X} \\ \begin{pmatrix} \tilde{Y}_{5\text{min}} \\ \tilde{Y}_{6\text{min}} \\ \vdots \\ \tilde{Y}_{9\text{min}} \end{pmatrix} \end{matrix} = \begin{matrix} x_{t_0} & x_{t_5} & x_{t_{10}} \\ \begin{pmatrix} \tilde{y}_{t_5} & \tilde{y}_{t_{10}} & \tilde{y}_{t_{15}} \\ \tilde{y}_{t_6} & \tilde{y}_{t_{11}} & \tilde{y}_{t_{16}} \\ \vdots & \vdots & \vdots \\ \tilde{y}_{t_9} & \tilde{y}_{t_{14}} & \tilde{y}_{t_{19}} \end{pmatrix} \end{matrix} \in \mathbb{R}^{5\times 3} \quad (1)$$

where $\mathscr{X}$ contains three sample input images x_{t_0}, x_{t_5} and $x_{t_{10}}$; $\tilde{Y}_{5\text{min}}$ to $\tilde{Y}_{9\text{min}}$ are prediction vectors that are produced by each individual model; $\tilde{\mathbb{Y}}_{\text{MSF}}$ is a vectorized representation of the MSF results consisting of all the predictions vectors.

In (1), the right term is a matrix containing all the forecasting points. Each row of the matrix represents for the forecasts generated by a single models, e.g., $\tilde{Y}_{5\text{min}} = [\tilde{y}_{t_5}, \tilde{y}_{t_{10}}, \tilde{y}_{t_{15}}]$. In the meanwhile each column of the matrix represents for the forecasts from the MSF model with one input image x_{t_i}, for example $\tilde{\mathbb{Y}}_{\text{MSF}} = [\tilde{y}_{t_5}, \tilde{y}_{t_6}, \ldots, \tilde{y}_{t_9}]^T$. The final forecasting results of the MSF model are yielded by concatenating all the column vectors of the matrix to a one-dimensional column vector, which is written as:

$$\tilde{\mathbb{Y}}_{\text{MSF}} = [\tilde{y}_{t_5}, \tilde{y}_{t_6}, \ldots, \tilde{y}_{t_{19}}]^T \in \mathbb{R}^{15\times 1} \quad (2)$$

As mentioned in the previous section, because of the mistakenly labeled data introduced by the lagged GHI values, the MSF results look noisy and the forecasting results of each model might be slightly biased (above or below the measurements). To further address this issue, we apply a two-dimensional median filter that reduces random noise to stabilize the output of the MSF model. The function of the median filter can be expressed as:

$$\tilde{\mathbb{Y}}_{\text{med}} = \text{Med}[\tilde{y}_{i-k}, \tilde{y}_{i-k+1}, \ldots, \tilde{y}_i, \ldots, \tilde{y}_{i+k}] \quad (3)$$

where $\tilde{\mathbb{Y}}_{\text{med}}$ is the filtered output and $\tilde{y}_i$ is the prediction sample. The median filter collects a window containing $N = 2k+1$ samples and then performs the median operation on this set where the sample values are replaced with the median value in this window. The window moves one step forward each time, and the only parameter is the filter length $N = 2k+1$, where $N = 5$ in this case.

IV. Forecasting-Based PV Power Ramp-Rate Control Strategy

For the grid-connected PV systems, it is essential to comply with operators' RR requirements, 10% of the rated capacity per minute to stabilize grid frequency and maintain power quality [30]. In this section, we first introduce the basic principles of the conventional PRRC methods based on ESS and inverter-level control. Then, an MSF-based PRRC technique using sky imagery is explained.

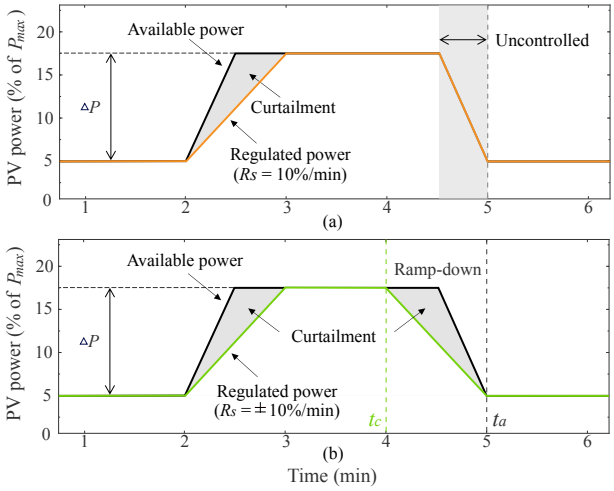


Fig. 8. PV power ramp-rate control curves: (a) PRRC with conventional activate power curtailment; (b) forecasting-based PRRC

A. PV Power Ramp-Rate Control Principle

Ramp rate is defined as how quickly a power plant's generation is changing over time, either ramping up (increasing) or ramping down (decreasing). The PRRC method is used to limit the fluctuations of output power from solar power plants, which can be described by the following equations:

$$R = \frac{P_{pv}(t) - P_{pv}(t - \Delta T)}{\Delta T} \quad (4)$$

$$R_c = \begin{cases} R_s & \text{if } R > R_s \\ -R_s & \text{if } R < -R_s \\ R & \text{otherwise} \end{cases} \quad (5)$$

$$P'_{pv}(t) = P_{pv}(t - \Delta T) + R_c(t - \Delta T) \cdot \Delta T \quad (6)$$

where R, R_c and R_s are the instantaneous RR, regulated RR and the ramp-rate limit respectively. P_{pv} and P'_{pv} are the actual and regulated PV generation at t. ΔT is the observation period. (5) describes how ramp rate is regulated in cases of ramp-up, ramp-down and the ramp within the limit R_s.

PV power ramp rates are commonly regulated by using ESS and APC. The ESS stores energy when PV generation increases faster than R_s and feeds power to the grid when the output power decreases faster than R_s. However, the high cost of ESS is preventing this method from large scale adoption. Additionally, active power curtailment complies with the ramp-rate requirement by curtailing the PV generation, i.e. control the operating point away from the maximum power point. As shown in Fig. 8(a), the main disadvantage of the conventional APC is that it can only be applied to the power ramp-up events because no external energy sources can be used to compensate for the ramp-down fluctuations.

B. MSF-Based PRRC

Fig. 8(b) shows the principle of forecasting-based PRRC. The ramp-up sides are controlled by the conventional active power curtailment method. As for the ramp-down events, the

TABLE III
FORECASTING RESULTS OF RESNET-18 IN DIFFERENT HORIZONS
(∗ INDICATES USING THE STACKED IMAGES)

	Model	MAE (W/m^2)	RMSE (W/m^2)	nMAE (%)	nRMSE (%)	skill (%)
5-min	Persistence	46.5	111.8	10.0	24.5	-
	ResNet-18	58.5	106.6	12.6	23.0	4.7
	ResNet-18*	56.4	105.5	12.2	22.8	**5.6**
6-min	Persistence	49.8	116.9	10.8	25.3	-
	ResNet-18	62.3	105.9	13.4	22.8	9.4
	ResNet-18*	51.9	102.0	11.2	21.2	**12.7**
7-min	Persistence	53.1	120.6	11.5	26.1	-
	ResNet-18	61.2	108.1	13.2	23.3	10.4
	ResNet-18*	60.8	104.8	13.1	22.6	**13.1**
8-min	Persistence	56.2	123.6	12.2	26.7	-
	ResNet-18	53.1	111.2	11.8	23.2	10.0
	ResNet-18*	53.7	104.5	11.6	22.6	**15.4**
9-min	Persistence	58.9	125.6	12.7	27.1	
	ResNet-18	60.7	111.2	13.1	22.8	11.4
	ResNet-18*	54.1	104.3	11.7	22.5	**17.0**
10-min	Persistence	61.2	126.8	13.2	27.4	-
	ResNet-18	61.6	113.6	13.3	24.6	10.4
	ResNet-18*	54.1	104.3	10.8	22.5	**17.7**

ramp-rate violations can be regulated concerning the predictions of the forecasting models. However, the performance of the PRRC highly relies on the forecasting accuracy and forecast resolution. For example, sparse predictions generated by conventional SSF methods will result in less robust controls and unnecessary energy curtailed.

In the contrast, the MSF approach produces multiple predictions within $\mathscr{H}$, associated with a series of corresponding ramp-rate events $R_{t_i}, R_{t_i+1}, \ldots, R_{t_i+\mathscr{H}}$. Instead of using the minimum values as previously done in [20], median values R_{med} in each MSF range is used, which expressed as:

$$R_{\text{med}} = \text{Med}[R_{t_i}, R_{t_i+1}, \ldots, R_{t_i+\mathscr{H}}] \tag{7}$$

Based on (5) and (7), the MSF-based PRRC scheme is:

$$R_c = \begin{cases} R_s & \text{if } R_{\text{med}} > R_s \\ -R_s & \text{if } R_{\text{med}} < -R_s \\ R_{\text{med}} & \text{otherwise} \end{cases} \tag{8}$$

which indicates that if R_{med} in the predictions is larger than the limit R_s, all RRs in the forecast interval will be limit to R_s. Otherwise, the power will be curtailed following the median value of R_{med}. As a result, both ramp-up and ramp-down fluctuations can be regulated by the active power curtailment without ESS.

V. EVALUATION AND ANALYSIS

A. Solar Forecasting Results

The persistence model is selected as the baseline model to evaluate the performance of our model. It is based on the assumption that the future irradiance will remain unchanged over the forecast horizon. The mean absolute error (MAE) and the root mean square error (RMSE) as well as their normalized versions. Additionally, the forecast skill score (s) that was firstly introduced by [31] is used to measure the improvement of the forecasting model over the reference persistence model. It is defined as:

$$\text{s} = 1 - \frac{\text{RMSE}}{\text{RMSE}_{\text{p}}} \tag{9}$$

TABLE IV
PERFORMANCE OF THE PROPOSED MSF ON PRRC

Criteria	MSF		SSF	
	cloudy	clear	cloudy	clear
Ramp events	91	11	91	11
Control failures	**1**	**0**	22	1
Failure rate (%)	**1.1**	**0**	24.2	9.1
Energy curtailment (%)	**11.3**	**3.1**	17.6	7.2

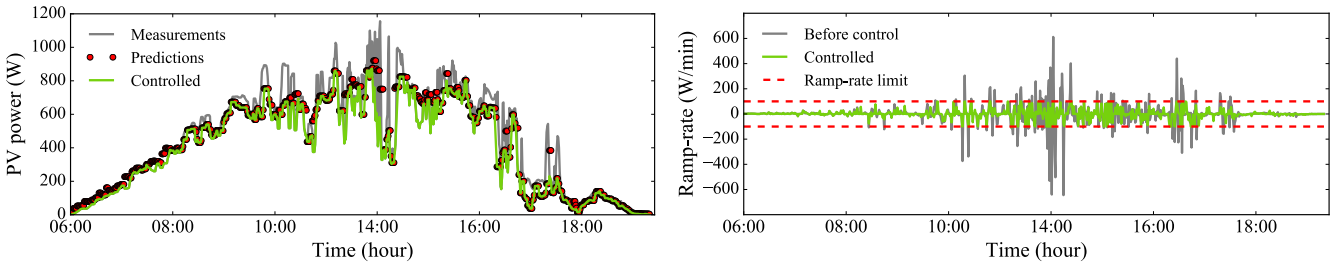
where RMSE_{p} is the error of the persistence model.

In this experiment, we compare the ResNet-18 model to the baseline persistence model on different forecast horizons. Table III reports the experimental results over a long period of 30 consecutive days including various weather conditions and cloud patterns. First, it demonstrates that the ResNet-18 significantly outperforms the baseline model with increasing forecasting skills from 5 min to 10 min, although the MAE of the baseline model is relatively small. The ResNet18* denotes the model trained and tested with the stacked images, by which the accuracy is further improved with the largest improvement from 10.4% to 17.7%. The results imply that the ResNet-18 model is more capable of forecasting GHI within a particular range (e.g., 5 min to 10 min in this experiment), but for the forecasts within 5 min the performance would be less competent. In comparison, the regression models used in [19], achieve average skills from 7.5% to 8.4% for intra-hour GHI from 5 min to 30 min. Meanwhile, these models require extra features extracted from the sky images.

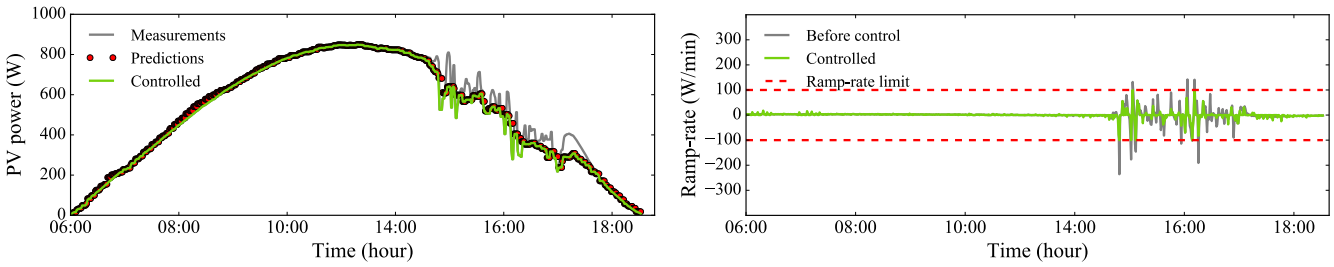
B. PRRC Results

To evaluate the performance of the MSF-based PRRC, the SSF model of the 10-min forecast horizon, which has the highest skill of 17.7% is chosen as the reference model. We use different temporal-resolution images for the MSF and SSF methods respectively to produce forecasts at the same resolution of 1 min required for PRRC. The MSF method uses down-sampled images at 5-min resolution, while the SSF model uses more frequent images at 1-min resolution. Fig. 9(a) and (b) demonstrate the simulation results of the PRRC application on a cloudy day and a clear day. It can be seen that the ramp violations are compliant to the regulation after control and the PV generations are effectively smoothed.

Table IV compares the performance of the proposed control strategy with SSF-based PRRC. There are 91 and 11 ramp violations observed in the original data on two separate days. The MSF-based PRRC successfully regulates 90 ramp violations with only one control failure on a cloudy day. Comparably, 22 RR violations are failed to be controlled by the SSF method associated with a much higher failure rate of 24.2%. In the case of the clear day, their performances are similar except for one control failure in the SSF model. As a result, the SSF-based control curtails 18.1% and 7.2% of the total PV production on each day, which are significantly reduced to 11.3% and 3.1% by using the proposed strategy.



(a) 21 April, 2016



(b) 2 April, 2016

Fig. 9. Simulation results of the MSF-based PRRC on (a) a cloudy day and (b) a clear day with few clouds. The left side shows the MSF predictions (red dots), the original PV generation (gray line) and the regulated PV generation (green line). The right side shows the RR violations compliant to the regulation.

VI. Conclusion

In this paper, the proposed MSF strategy provides accuracy and robust solar forecasts for the PRRC application without using time-series models and additional meteorological information. We can forecast 10 min ahead at a temporal resolution of 1 min utilizing an ensemble of ResNet-18 models with different forecast horizons. The stacked images which integrate spatial-temporal information are used to further enhance the forecasting accuracy. Comparing with the published results, higher forecasting accuracy has been achieved with the skill of 17.7% for a 10-min horizon. The simulation results show that the use of MSF on PRRC outperforms the one using SSF, with less control failures and energy curtailment.

References

[1] M. Woodhouse, R. Jones-Albertus, D. Feldman, R. Fu, K. Horowitz, D. Chung, D. Jordan, and S. Kurtz, "On the path to sunshot: the role of advancements in solar photovoltaic efficiency, reliability, and costs," National Renewable Energy Laboratory (NREL), Golden, CO, USA, Tech. Rep. NREL/TP-6A20-65872, May 2016.

[2] L. Bird, M. Milligan, and D. Lew, "Integrating variable renewable energy: challenges and solutions," National Renewable Energy Laboratory (NREL), Golden, CO, USA, Tech. Rep. NREL/TP-6A20-60451, 2013.

[3] K. A. Kim, P. T. Krein, G. Seo, and B. Cho, "Photovoltaic ac parameter characterization for dynamic partial shading and hot spot detection," in *Applied Power Electronics Conference and Exposition (APEC)*, Long Beach, May 2013, pp. 109–115.

[4] Y. Du, K. Yan, Z. Ren, and W. Xiao, "Designing localized MPPT for PV systems using fuzzy-weighted extreme learning machine," *Energies*, vol. 11, no. 10, pp. 1–10, 2018.

[5] X. Chen, Y. Du, and H. Wen, "Forecasting based power ramp-rate control for PV systems without energy storage," in *2017 IEEE 3rd International Future Energy Electronics Conference and ECCE Asia*, Jul. 2017, pp. 733–738.

[6] N. Mithulananthan, R. Bansal, and V. Ramachandaramurthy, "A review of key power system stability challenges for large-scale PV integration," *Renewable and Sustaniable Energy Reviews*, vol. 41, pp. 1423–1436, 2015.

[7] S. Sukumar, M. Marsadek, K. Agileswari, and H. Mokhlis, "Ramp-rate control smoothing methods to control output power fluctuations from solar photovoltaic (PV) sources - a review," *Journal of Energy Storage*, vol. 20, pp. 218–229, 2018.

[8] A. Sangwongwanich, Y. Yang, and F. Blaabjerg, "A cost-effective power ramp-rate control strategy for single-phase two-stage grid-connected photovoltaic systems," in *Proceedings of the 8th Annual IEEE Energy Conversion Congress and Exposition (ECCE)*, 2016, pp. 1–7.

[9] D. Yang, J. Kleissl, C. A. Gueymardc, H. T. Pedrob, and C. F. Coimbra, "History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining," *Solar Engergy*, vol. 168, pp. 60–101, 2018.

[10] E. Lorenz, J. Kühnert, D. Heinemann, K. P. Nielsen, J. Remund, and S. C. Müller, "Comparison of global horizontal irradiance forecasts based on numerical weather prediction models with different spatio-temporal resolutions," in *31st EU PVSEC*, Hamburg, 2015.

[11] H. Yang, B. Kurtz, D. Nguyen, B. Urquhart, C. W. Chow, M. Ghonima, and J. Kleissl, "Solar irradiance forecasting using a ground-based sky imager developed at uc san diego," *Solar Energy*, vol. 103, pp. 502–524, 2014.

[12] P. Neelamegam and V. A. Amirtham, "Prediction of solar radiation for solar systems by using ann models with different back propagation algorithms," *Journal of Applied Research and Technology*, vol. 14, pp. 206–214, 2016.

[13] K. Yan, Y. Du, and Z. Ren, "MPPT perturbation optimization of photovoltaic power systems based on solar irradiance data classification," *IEEE Trans. Sustain. Energy*, vol. 10, pp. 514–521, 2019.

[14] Y. Wang, Y. X. Shen, S. W. Mao, G. Q. Cao, and R. M. Nelms, "Adaptive learning hybrid model for solar intensity forecasting," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1635–1645, 2018.

[15] X. Chen, Y. Du, H. Wen, J. L, and W. Xiao, "Forecasting based power ramp-rate control strategies for utility-scale PV systems," *IEEE Trans. Ind. Electron.*, vol. 66, pp. 1862–1871, 2019.

[16] X. Chen, Y. Du, E. G. Lim, H. Wen, and J. L, "Sensor network based PV power nowcasting with spatio-temporal preselection for grid-friendly control," *Applied Energy*, vol. 225, pp. 1862–1871, 2019.

[17] M. André, S. Dabo-Niang, T. Soubdhan, and H. Ould-Baba, "Predictive spatio-temperal model for spatially sparse global solar radiation data," *Energy*, vol. 111, no. 15, pp. 599–608, 2016.

[18] B. Elsinga and W. Sark, "Short-term peer-to-peer solar forecasting in a network of photovoltaic systems," *Applied Energy*, vol. 206, pp. 1464–1483, 2017.

[19] H. T. C. Pedro, D. P. Larson, and C. F. M. Coimbra, "A comprehensive dataset for the accelerated development and benchmarking of solar forecasting methods," *Renewable Sustainable Energy*, vol. 11, Jun. 2019.

[20] M. Saleh, L. Meek, M. A. S. Masoum, and M. Abshar, "Battery-less short-term smoothing of photovoltaic generation using sky camera," *IEEE Trans. Ind. Informat.*, vol. 14, pp. 403–414, Feb. 2018.

[21] M. Q. Raze, N. Mithulananthan, J. Li, K. Y. Lee, and H. B. Gooi, "An ensemble framework for day-ahead forecast of PV output power in smart grids," *Trans. Ind. Informat.*, vol. 15, pp. 4624–4634, 2019.

[22] Y. Chu, H. T. Pedro, M. Li, and C. F.Coimbra, "Real-time forecasting of solar irradiance ramps with smart image processing," *Solar Energy*, vol. 114, pp. 91–104, 2015.

[23] J. Zhang, R. Verschae, S. Nobuharac, and J.-F. Lalonde, "Deep photovoltaic nowcasting," *Solar Energy*, vol. 176, pp. 267–276, 2018.

[24] T. Stoffel and A. Andreas, "NREL solar radiation research laboratory (SRRL): Baseline measurement system (BMS); golden, colorado (data)," Jul. 1981.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations (ICLR)*, 2015.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2016, pp. 770–778.

[27] G. Huang, Z. Liu, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, pp. 2261–2269.

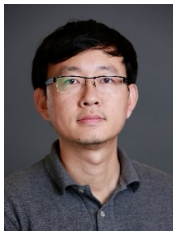
[28] M. Telgarsky, "Benefits of depth in neural networks," *Journal of Machine Learning Reseach*, vol. 49, pp. 1517–1539, Jun. 2016.

[29] D. Yang, "A guideline to solar forecasting research practice: Reproducible, operational, probabilistic or physically-based, ensemble, and skill (ROPES)," *Journal of Renewable and Sustainable Energy*, pp. 11, 022 701, Mar. 2019.

[30] V. Gevorgian and S. Booth, "Review of PREPA technical requirements for inter-connecting wind and solar generation," National Renewable Energy Laboratory (NREL), Golden, CO, USA, Tech. Rep. NREL/TP-5D00-57089, Nov. 2013.

[31] C. W. Chow, B. Urquhart, M. Lave, A. Dominguez, J. Kleissl, J. Shields, and B. Washom, "Intra-hour forecasting with a total sky imager at the uc san diego solar energy testbed," *Journal of Renewable and Sustainable Energy*, pp. 2881–2893, Nov. 2011.

Haoran Wen received his B.S degree in electrical engineering from Xian Jiaotong-Liverpool University, Suzhou, China in 2016; and his M.S. degree in sustainable energy technology from the University of Liverpool, Liverpool, U.K., in 2018. From 2018 to 2019, he was with the GCL New Energy, Co., Ltd., Suzhou, China. He is currently working toward the Ph.D. degree at the University of Liverpool, Liverpool, U.K.. His research interests include solar power forecasting, photovoltaic power systems, machine learning, and image processing.



Yang Du (M'13) received his Ph.D. degree in electrical engineering from The University of Sydney, Australia, in 2013. From 2013 to 2014, he was with Masdar Institute of Science and Technology, Abu Dhabi, UAE, as a post-doctoral research fellow. From 2014 to 2018, he was a lecturer at Xian Jiaotong-Liverpool University, Suzhou, China, where he maintains an honorary position. He joined James Cook University, Cairns, Australia in 2019. He was a visiting scientist at Massachusetts Institute of Technology (MIT) in 2018. He is an associate editor of the IET Renewable Power Generation. His research interest includes solar forecasting, renewable energy integration, power electronics and smart grid.

Xiaoyang Chen (S'17) was born in Suzhou, China, in 1994. He received the B.S. degree in electrical engineering with distinction from Xian Jiaotong-Liverpool University, Suzhou, China, in 2017. He is currently pursuing the Ph.D. degree at the University of Liverpool, Liverpool, U.K.. His research interests include solar power forecasting, photovoltaic power systems reliability, and smart grid.

Eng Gee Lim (SM'98) received the BEng (Hons) and PhD degrees in Electrical and Electronic Engineering from the University of Northumbria, UK in 1998 and 2002 respectively. Prof. Lim worked for Andrew Ltd, a leading communications systems company in the United Kingdom from 2002 to 2007. Since August 2007, Prof. Lim has been at Xian Jiaotong-Liverpool University, where he was formally the head of EEE department and University Dean of Research and Graduate studies. Now, he is School Dean of Advanced Technology, director of AI university research centre and also professor in department of Electrical and Electronic Engineering. He has published over 100 refereed international journals and conference papers. His research interests are Artificial Intelligence, robotics, AI+ Health care, international Standard (ISO/ IEC) in Robotics, antennas, RF/microwave engineering, EM measurements/simulations, energy harvesting, power/energy transfer, smart-grid communication; wireless communication networks for smart and green cities. He is a charter engineer and Fellow of both IET and Engineers Australia. In addition, he is also a senior member of IEEE and Senior Fellow of HEA.

Huiqing Wen (M'13-SM'18) received his B.S. and M.S. degrees in Electrical Engineering from Zhejiang University, Hangzhou, China, in 2002 and 2006, respectively. In 2009, he received his Ph.D. in Electrical Engineering from the Chinese Academy of Sciences, Beijing, China. From 2009 to 2010, he has been an electrical engineer working with the GE (China) Research and Development Center Company, Ltd., Shanghai, China. From 2010 to 2011, he was an engineer at the China Coal Research Institute, Beijing, China. From 2011 to 2012, he was a postdoctoral fellow at the Masdar Institute of Science and Technology, Abu Dhabi, United Arab Emirates. In 2013, he joined the Electrical and Electronic Engineering Department of Xian Jiaotong-Liverpool University (XJTLU), Suzhou, China. Currently, he is a senior associate professor at the XJTLU. He has published more than 100 peer reviewed technical papers in leading journals/conferences and holds over 20 issued/pending patents. His research interests include renewable energy, electric vehicle, power electronics, Microgrid, and power semiconductor devices. He is the associate editor of IEEE ACCESS, International Journal of Photoenergy, and Journal of Power Electronics.

Lin Jiang (M'00) received B.S. and M.S. degrees in electrical engineering from the Huazhong University of Science and Technology China, in 1992 and 1996; and the Ph.D. degree from University of Liverpool U.K in 2001. He is currently a Reader at University of Liverpool. His research interests include optimization and control of smart grid, electrical machine, power electronics and renewable energy.

Wei Xiang (S'00-M'04-SM'10) received the B.Eng. and M.Eng. degrees, both in electronic engineering, from the University of Electronic Science and Technology of China, Chengdu, China, in 1997 and 2000, respectively, and the Ph.D. degree in telecommunications engineering from the University of South Australia, Adelaide, Australia, in 2004.

He is currently the Founding Chair and Head of Discipline of Internet of Things Engineering at James Cook University, Cairns, Australia. He is an elected Fellow of the IET in UK and Engineers Australia. He is the Vice Chair of the IEEE Northern Australia Section. He was an Editor for IEEE Communications Letters (2015-2017), and is an Associate Editor for IEEE Access and Springers Telecommunications Systems. He has published over 250 peer-reviewed papers including 3 academic books and 130 journal articles. His research interest falls under the broad areas of communications and information theory, particularly the Internet of Things, and coding and signal processing for multimedia communications systems.