Sky Image-Based Photovoltaic Output Forecasting

Atindra Jha
Computer Science
Stanford University
atj10@stanford.edu

Dhvaneel Visaria
Mechanical Engineering
Stanford University
dhvaneel@stanford.edu

Abstract

In this study, we evaluate and compare the effectiveness of 2D CNN models like ResNet-18 and ResNet-34 as well as the R(2+1)D-18 video classification model in using sky images to forecast the power output of solar cells in the short term. Our motivation stems from the need to alleviate the unpredictability associated with the short-term power output of solar power generation systems, making it difficult for them to be integrated into larger electric grids. We choose to utilize the aforementioned models in order to analyze the gains in performance, if any, that a computationally expensive video classification architecture can generate over 2D CNNs within the framework of residual learning. We observe that the R(2+1)D-18 model performs yields just marginally higher accuracy while being significantly more computationally expensive that ResNet-34. Moreover, we show that the best results are observed when predicting the average power output over a 5 to 10 minute forecast horizon using images from the previous 2.5 to 5 minutes.

1. Introduction

While energy needs of the world are rising, non-renewable resources conventionally used to produce it are getting exhausted. Coupled with the ongoing climate crisis, this triggers an urgent need for clean and sustainable energy sources to achieve the 2°C goal of the Paris Agreement by phasing out fossil fuels and replacing them with low-carbon sources of energy. Solar energy has emerged as one such promising renewable energy source with a potential of producing 1.74×10^{17} W/year given that the Earth's surface receives approximately 1367 W/m^2 of solar radiation. It is the renewable energy technology with the highest growth rate of more than more than 40% per year over the past decade [19].

Despite the expansion of solar infrastructure and the advancements in engineering, the integration of solar power generation systems into electric grids remains challenging owing to the variability associated with this method of power generation. The amount of electrical power generated from PV systems is dependent on various factors, both internal and external - from solar panels, the semiconductor materials, their energy conversion efficiency to several meteorological variables, including temperature, pressure, relative humidity, aerosol index, wind speed and cloud cover, long-term and short-term weather conditions, tilt of the solar panels.

Measuring solar irradiance may serve as a useful tool for onsite photovoltaic (PV) power forecasting, in turn allowing better predictability with regards to power generation and higher penetration of solar power generation systems into the grid. Solar irradiance is the output of light energy from the entire disk of the Sun, measured at the Earth. It is essentially the brightness of the sunlight at a given spot on the Earth and it is positively correlated with amount of power generated by PV systems. Therefore, it becomes important to study and model the effect of the intensity of solar radiation received by the panels on the output power.

2. Related Work

Given the uncertainty associated with solar energy production owing to its dependence on so many factors, precise forecasting of solar energy availability is highly desirable for the energy industry. Traditionally, empirical models based on geographical and meteorological parameters have been used to estimate the solar irradiance with gradual modifications by incorporating additional metadata such as rainfall, relative humidity, pressure, temperature for better modeling.

With the gradual emergence of capable sensor hardware to capture images/data, the utilization of either satellites or ground-taken sky images has gained popularity. Such high-fidelity data is processed using statistical and signal processing methods including various physical models [2], historical models [11, 20], weather models, cloud trajectory tracking [18, 23, 25]. Therefore, there are two types of

solar forecasting methods - one is weather modeling using numerical forecast models and the other is data-driven approach using either ground-based sky observations or satellite imagery.

State-of-the-art machine learning [14, 22] and deep learning models [12, 15] especially using computer vision have emerged powerful without having to deal with complexity of cloud physical properties, and spatial and temporal dynamics. Unlike empirical models, image-based models capture the cloud information from the sky image dataset, which helps in reasonably accurate solar irradiance forecasting.

Previous works have employed a diverse set of deep learning model architectures depending on the forecast horizon – very short-term, short-term, medium-term and long-term forecasting, from sequence models like long short term memory models (LSTM) [13, 27] and gated recurrent unit (a variant of LSTMs with lesser number of parameters) [24, 26] to other hybrid models – combining image data along with weather metadata using recurrent neural networks (RNN) [9, 16], integrating convolutional neural networks (CNN) with LSTMs [6] and bi-directional gated recurrent unit (Bi-GRU) with autoregressive integrated moving average (ARIMA) model [8].

In this project, we use the ground-based sky images to predict the PV power output on a short-term basis by leveraging the temporal data. Our goal, essentially, is to leverage the time series based image dataset to train a model that predicts the PV output of a solar installation over the next 5, 10 and 15 minutes. Previous work using this dataset, Dissawa, Lasanthika H., et al. [5] has focused on signal processing techniques for cloud motion tracking in order to eventually determine solar irradiance. since cloud cover is the a significant factor causing variation in the solar radiation received. In this project, we implement deep learning techniques using their dataset [4] to compare with their performance.

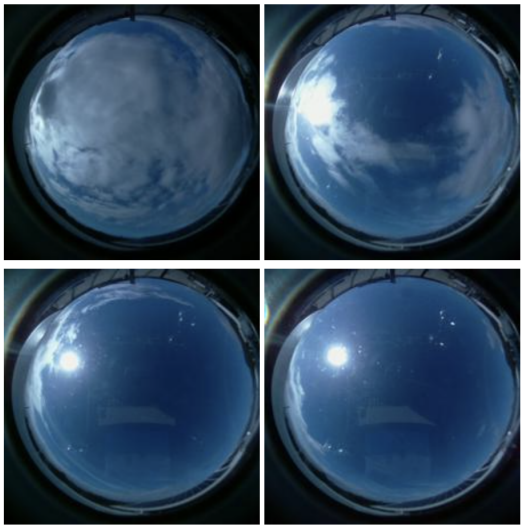

Figure 1. Sky images related to overcast, partially cloudy, and clear skies used for training the deep learning models

3. Dataset

Image data is important to use along with the PV output measurements in modeling the irradiance behaviour since the changes in the power output curve are unpredictable while using the past PV data only. Therefore, we use ground-based sky images (as shown in Fig. 1) related to sunny, partially cloudy, and overcast sky conditions with the corresponding power measurements. These images indicate the solar irradiance by the presence and brightness of the Sun as well as the amount of occlusion created by the clouds.

This data was obtained from two rooftop PV systems at University of Wollongong, Australia using a camera with a large field of view for a high spatial and temporal resolution for the signal processing based study [5]. Specifically, a high-resolution Raspberry Pi camera module with a wide-angle lens was used to grab a vast area of the sky onto the image. The resolution of the captured images was 1024×768 pixels. 3000 images are obtained each day from 10th September 2019 to 12th September 2019, therefore we have 9000 images in total.

The dataset is pre-processed before using them for training and validating the deep learning models for the different experiments. Given the limited computing resources, we crop the sky images to a size of 224×224. The images are normalized according to the ImageNet dataset statistics. Data augmentation is done for certain experiments using random horizontal flip, random vertical flip and random rotation transforms in PyTorch. Since the image capturing and PV power measurement systems are decoupled, the sky images are captured at a time resolution of 10 seconds but the PV output is obtained at a time resolution of 30 seconds. In order to fix the mismatch, we employ simple statistical approaches. Particularly, we use linear interpolation to obtain the PV power outputs corresponding to the image capture times and plot the same in Fig. 2. Evidently, the linear interpolation performs really well despite the rugged nature of the power output curve.

Therefore, we finally have 224×224 sky images at an interval of 10 seconds for three days with the correpsonding PV power output to train the model. The data for the first two days is used for training while the data for the first half

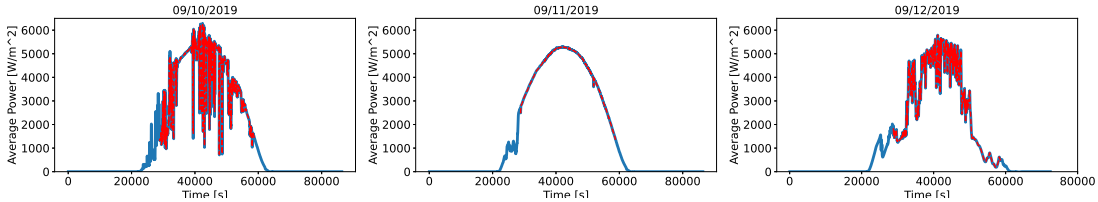

Figure 2. Daily average solar output power (W) from 10th September 2019 to 12th September 2019 measured by the PV measurement system (in blue) and the linearly interpolated PV output values corresponding to sky image capture times (in red)

of the third day is used for validation and that for the second half is used for the final testing to preserve the chronology in favour of randomization.

4. Methods

Given the task at hand of predicting the solar irradiance from the sky images, we first employ transfer learning methods to train simple but effective models. Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task. This optimization helps in rapid progress or improved performance when modeling the second task with lesser computational resources. While dealing with images of the sky, it is straightforward to adopt models trained on ImageNet [3] - a large visual database designed for use in visual object recognition.

Therefore, we start with a minimum viable transfer learning based model in order to test our hypotheses, validate the data pipeline, and establish a baseline performance on the task of predicting PV output based on the data using PyTorch [17]. We use the pretrained AlexNet [10] model with its output layer altered from 1000 neurons to a single neuron in order to suit our regression task as opposed to the original classification task over the ImageNet dataset. The task here was simply to be able to predict the PV output corresponding to the sky image and not forecasting. The input to the AlexNet model are the pre-processed $224 \times 224 \times 3$ RGB images of the sky and the output is the corresponding solar power output in W/m^2.

Having established baseline performance, we transition to deeper, more complex networks for the task of forecasting still using transfer learning. Particularly, we employ pretrained ResNet-18 and ResNet-34 [7] models that are reasonably deep networks for the task at hand. In order to provide the temporal data to our model, we input a sequenced set of the sky images to predict an average PV output power over the forecast horizon. The forecast horizon (fc) is the length of time into the future for which forecasts are to be prepared. Particularly, we input a concatenated tensor of N images of shape $224 \times 224 \times 3N$ at a certain time interval (frequency, f) from time $t-(N-1)f$ to time t to predict the average PV power output over time t to time $t+fc$.

We choose to predict the average power instead of the instantaneous power in order to make the regression task a little easier for the model since we observe that PV output power changes can be very unpredictable as can be seen in Fig. 2. Additionally, a good performance on the average power prediction can still help us against the unreliability of the solar power. Here again the first convolutional layer and the last fully connected layer of the pretrained ResNet models are modified according to the $224 \times 224 \times 3N$ input and singly valued output. For further experiments related to regularization, we also add a dropout layer before the last fully connected layer.

Since the model input consists of the sky images at every ten second interval which is representative of a visual temporal data sequence of a fixed length, this makes it similar to passing a set of frames from a video clip as an input. Therefore, considering the residual network architectures with spatiotemporal convolutions for video classification seemed like the logical next step for improved performance using specialized VideoNet models. Therefore, we employ R(2+1)D [21] that introduces spatiotemporal convolutions, a middle ground between 2D convolutions that consider only two-dimensional spatial information and the computationally expensive 3D convolutions used by R3D18 [1].

R(2+1)D explicitly factors each of the 3D convolutions into multiple 2D spatial convolutions followed by a one-dimensional temporal convolution. The study by Tran et. al. [21] assesses the relative performance of varying spatiotemporal convolutional architectures for video learning, finding that using spatial 2D convolutions followed by a temporal 1D convolution provides increases in accuracy over whole spatiotemporal 3D convolutions. Again, we modify the R(2+1)D18 model to have a single-neuron output layer.

Since R(2+1)D are ResNets with (2+1)D convolutions

used for video classification, we changed our input size from $224 \times 224 \times 3N$ (for 2D residual network architectures) to $3 \times N \times 224 \times 224$ where N is the number of images being fed to the network at once.

In case of all of the aforementioned models, we take the square root of the mean squared error loss. Additionally, we utilize the Adam optimizer because of its versatility with a learning rate of $1e^{-2}$ and weight decay of $1e^{-4}$, hyperparameter values obtained after extensive optimization. We use an accuracy to evaluate the model performance which is defined as the accuracy of the model prediction within 10 % of the ground truth. The root mean squared error and accuracy used for model evaluation can be mathematically expressed as follows:

$$\text{RMSE} = \sqrt{(\frac{1}{n})\sum_{i=1}^{n}(y_i - \hat{y_i})^2}$$

$$\text{Accuracy} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(\frac{y_i - \hat{y_i}}{y_i} \leq 0.1)$$

where y_i is the ground truth while $\hat{y_i}$ is the predicted output. We use a default batch size of 128 for all the models except for R(2+1)D where we use a batch size of 16 due to memory constraints.

5. Experiments and Results

We begin our study by employing the pre-trained AlexNet model for establishing the baseline performance on the simple regression task of predicting the PV power output corresponding to a given sky image. This With resource constraints associated with Google Colab, we train it for approximately 10 epochs to find a reasonable performance on the validation dataset i.e accuracy of 37 %. This is in line with the hypothesis given the use of a relatively simple CNN architecture as a baseline to assess the robustness of the data pipeline.

Next, we move from predicting the PV output associated with a specific sky-image to using a series of sky-images to forecast future PV output. We begin by training 2D residual network architectures on this task. In order to avoid overfitting, we first train ResNet-18, a relatively shallow 18 layer deep convolutional neural network, for 20 epochs. The model was not complex enough to yield RMSE loss and accuracy numbers as low as we had hoped for. The highest validation set accuracy observed was 21.9% and the lowest RMSE loss was 927.29. The Fig. 4 illustrates the poor performance of ResNet-18 using the validation accuracy observed across 20 epochs.

# Images	Frequency (s)	Forecast Horizon (s)	% Accuracy
5	30	300	57.59
5	30	600	68.79
5	30	900	51.76
5	60	300	67.36
5	60	600	62.27
5	60	900	52.83
10	30	300	62.01
10	30	600	68.49
10	30	900	53.55
10	60	300	55.11
10	60	600	54.71
10	60	900	52.07
15	30	300	59.18
15	30	600	57.58
15	30	900	51.14
15	60	300	53.55
15	60	600	56.44
15	60	900	57.42

Table 1. Table containing the performance of ResNet-34 with dropout layer (ratio = 0.3) in terms of best validation accuracy for various discrete values of number of images (N), frequency of the input images (f) and forecast horizon (fc)

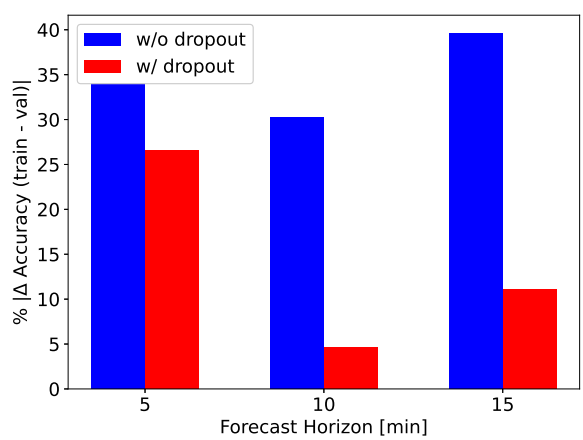

Figure 3. Bar chart depicting the degree of overfitting using the difference between the train and validation accuracies for the best performing ResNet-34 models without and with dropout for $N = 50$, $f = 30$ s, $fc = 5$ min

Having recognized the need for a larger model we employ ResNet-34 in order to achieve better performance on our task. We use a batch size of 128 and optimize the model by tuning the the learning rate and weight decay

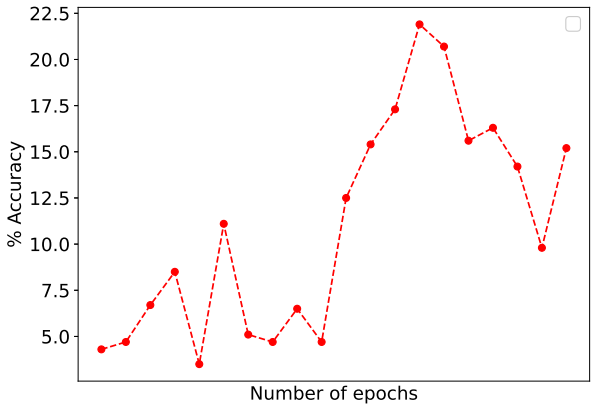

Figure 4. Validation accuracy with number of epochs depicting the poor performance of ResNet-18

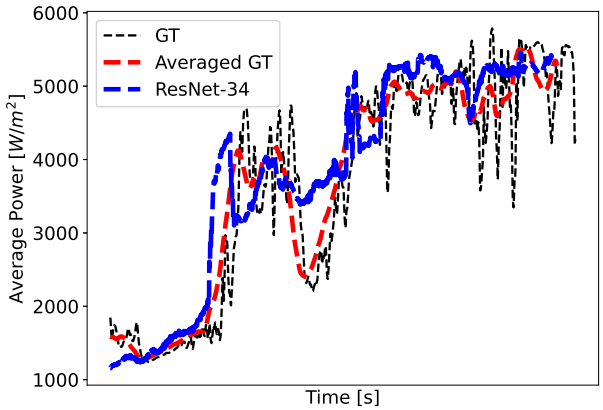

Figure 5. Predictions of the best ResNet-34 model ($N = 5$, $f = 30$ s, $fc = 10$ min) with dropout layer (ratio = 0.5) on the validation set as compared to the ground-truth and the moving average of ground-truth

hyperparameters extensively. The model yielded the best results with a learning rate of $1e-2$ and weight decay of $1e-4$. We further investigate the effects of three important hyperparameters namely, number of images concatenated in the input tensor (N), the frequency or the time interval between each image in the input (f) and the forecast horizon (fc) i.e. the time horizon for which we want our model to predict the average power output.

On training the above model, we observe overfitting i.e. unreasonably high training accuracy as compared to low validation accuracy. Therefore, we add a dropout layer with a random probability, p of 0.3 and 0.5 before the last fully connected layer for higher regularization. While both values of p in the dropout layer help decrease the high variance, having $p = 0.5$ has a better regularizing effect on the model than $p = 0.3$. This helps us obtain a lower degree of overfitting as is evident in the Fig. 3. Here we plot the difference between the training and validation accuracy obtained from models trained with $N = 5$ input images at a frequency interval of $f = 30$ seconds and a forecast horizon, fc of 5, 10 and 15 minutes. It is apparent that the difference between the train and validation accuracy falls across the board with dropout.

Next we study the effect of number of images concatenated in the input (N), the frequency or the time interval (f) and the forecast horizon (fc) on the forecasting model performance. Particularly, we experiment with discrete number of values of $N \in \{5, 10, 15\}$, $f \in \{30, 60\}$ seconds and $fc \in \{5, 10, 15\}$ minutes. It is evident that the model performance decreases with increase in the number of images, particularly at $N = 15$. Also, the performance is much better for $f = 30$ seconds as compared to that for $f = 60$ seconds. Therefore, we infer that in case of short-term solar irradiance forecasting, only very immediate information about the clouds and the sky is representative of the future conditions.

To explain further, the prediction deteriorates if the model receives information from a longer time in the past before time t because the changes in the factors affecting solar irradiance including cloud cover are very fickle i.e. they can change instantly and sometimes dramatically, thus affecting the forecast predictions. We also observe better performance for smaller forecast horizons, as expected, since it is difficult to predict weather changes further apart in time. The forecast prediction of this model on the validation dataset as compared to the groundtruth is plotted in Fig. 5. The best performances is observed for combinations of N, f, and fc that include images from the preceding 150 or 300 seconds and try to forecast the average power output over the following 300 or 600 seconds. Examples of such combinations include $N = 5$, $f = 30$, and $fc = 300$; $N = 10$, $f = 30$, and $fc = 600$; and $N = 5$, $f = 60$, and $fc = 300$.

In order to further enhance performance, we carry out data augmentation by flipping and rotating the images in the dataset. Instead of triggering performance gains, these augmentation techniques result in significantly lower accuracy and higher loss. This is expected given the nature of these sky-images (Fig. 1), which essentially resemble solid blue circles containing a small bright hole representing the Sun

and greyish-white hues representing the clouds. We conjectured that the correct position of the sun and the clouds in each image is vital for the model learning. Unlike images in classification tasks like cat-identification, flipping and rotating sky images i.e. altering the position of the Sun and the clouds, would make learning more difficult because we'd likely be introducing unsound noisy data to train our model.

Our final experiment involved using the R(2+1)D-18 video classification model to examine whether using a 3D ResNet CNN would lead to noteworthy performance gains. As mentioned before, in light of computational constraints, we decreased the batch size to 16. Despite being over twice as computationally expensive as ResNet-34 (in terms of the GPU memory required to train, validate, and test the models), the performance gains were only slightly better than ResNet-34. The graph illustrating the ground truth versus the predicted PV output values can be seen in Fig. 6 As noted in Table 2, the highest test accuracy that we witnessed was 72.65% in the 8th epoch and lowest RMSE loss was 376.84. Compared to the best ResNet-34 test accuracy of 67.81% and lowest RMSE loss of 408, the performance gains were not remarkably high.

6. Discussion and Conclusion

A clear conclusion that can be drawn from the above experiments is that the best results are observed when predicting the average PV output over the next 5 to 10 minutes using images from the previous 2.5 to 5 minutes. Using images more than 5 minutes prior to the point from which PV output must be predicted or increasing the forecast horizon to 15 minutes leads to poor performance likely because of the high volatility associated with PV outputs i.e. their ability to rise or fall dramatically within a very short time frame.

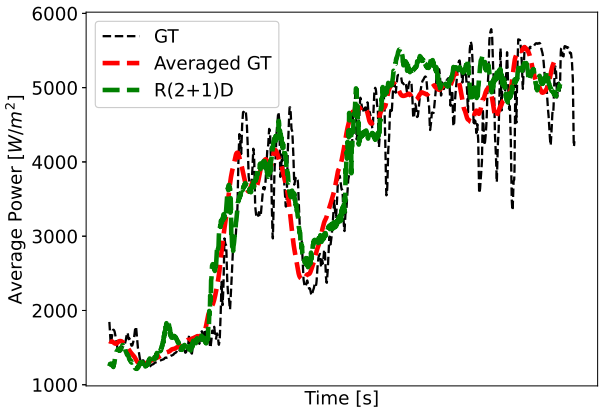

Figure 6. Predictions of the best R(2+1)D model ($N = 5$, $f = 30$s, $fc = 5$min) on the validation set as compared to the ground-truth and the moving average of ground-truth

Model	RMSE [W/m2]	Accuracy
ResNet-34	408	67.81%
R(2+1)D	377	72.65%

Table 2. Table containing the performance of ResNet-34 and R(2+1)D models on the test dataset

It is also noteworthy that general data augmentation techniques like flipping, rotating, and randomly cropping images, used in popular computer vision tasks, may not be helpful, and may even be conterintuitive in case of projects where the relative or exact position of certain features/objects in these influence their corresponding labels as well as the training of the model. The use of these aforementioned data augmentation techniques on our dataset caused model performance to drop significantly.

An unexpected observation is the unremarkable gain in performance that the R(2+1)D-18 model has over ResNet-34. This may be because we reached the peak performance that can be derived out of a residual learning model given the size of our dataset. Increasing model complexity in this case just leads to higher overfitting to the training data, which we observed in case of R(2+1)D-18.

7. Future Work

Given greater computational power and more time, it would be an interesting experiment to collect video data over a longer period of time and train deeper 2D residual CNN models as well as deeper 3D CNNs to compare their respective performances. Augmenting the dataset containing images/videos and PV outputs, with temperature data, azimuth angle, and so on may help improve model performance significantly.

8. Contributions

- Atindra Jha: Initial brainstorming and ideation; PyTorch code, milestone and project report, project poster

- Dhvaneel Vasaria: Initial brainstorming and ideation, PyTorch code, milestone and project report, graphics, tables, and plots

References

[1] Biagio Brattoli, Joseph Tighe, Fedor Zhdanov, Pietro Perona, and Krzysztof Chalupka. Rethinking zero-shot video classification: End-to-end training for realistic applications. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4613–4623, 2020. 3

[2] Chi Wai Chow, Bryan Urquhart, Matthew Lave, Anthony Dominguez, Jan Kleissl, Janet Shields, and Byron Washom. Intra-hour forecasting with a total sky imager at the uc san diego solar energy testbed. *Solar Energy*, 85(11):2881–2893, 2011. 1

[3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 3

[4] Duane; Agalgaonkar Ashish; Godaliyadda Roshan ; Ekanayake Parakrama; Perera Sarath ; Ekanayake Janaka Dissawa, Lasanthika; Robinson. Sky Images and PV Power Measurements for Irradiance Forecasting. https://data.mendeley.com/datasets/cb8t8np9z3/2, 2021. 2

[5] Lasanthika H Dissawa, Roshan I Godaliyadda, Parakrama B Ekanayake, Ashish P Agalgaonkar, Duane Robinson, Janaka B Ekanayake, and Sarath Perera. Sky image-based localized, short-term solar irradiance forecasting for multiple pv sites via cloud motion tracking. *International Journal of Photoenergy*, 2021, 2021. 2

[6] Sujan Ghimire, Ravinesh C Deo, Nawin Raj, and Jianchun Mi. Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. *Applied Energy*, 253:113541, 2019. 2

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3

[8] Mustafa Jaihuni, Jayanta Kumar Basak, Fawad Khan, Frank Gyan Okyere, Elanchezhian Arulmozhi, Anil Bhujel, Jihoon Park, Lee Deog Hyun, and Hyeon Tae Kim. A partially amended hybrid bi-gru—arima model (pahm) for predicting solar irradiance in short and very-short terms. *Energies*, 13(2):435, 2020. 2

[9] Kazım Kaba, Mehmet Sarıgül, Mutlu Avcı, and H Mustafa Kandırmaz. Estimation of daily global solar radiation using deep learning model. *Energy*, 162:126–135, 2018. 2

[10] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012. 3

[11] Dhananjay Kumar, Hitesh Dutt Mathur, Surekha Bhanot, and Ramesh C Bansal. Frequency regulation in islanded microgrid considering stochastic model of wind and pv. *International Transactions on Electrical Energy Systems*, 29(9):e12049, 2019. 1

[12] Pratima Kumari and Durga Toshniwal. Deep learning models for solar irradiance forecasting: A comprehensive review. *Journal of Cleaner Production*, 318:128566, 2021. 2

[13] Aslam Muhammad, Jae Myoung Lee, Sug Won Hong, Seung Jae Lee, and Eui Hyang Lee. Deep learning application in power system with a case study on solar irradiation forecasting. In *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 275–279. IEEE, 2019. 2

[14] Lanre Olatomiwa, Saad Mekhilef, Shahaboddin Shamshirband, Kasra Mohammadi, Dalibor Petković, and Ch Sudheer. A support vector machine–firefly algorithm-based model for global solar radiation prediction. *Solar Energy*, 115:632–644, 2015. 2

[15] Quentin Paletta, Guillaume Arbod, and Joan Lasenby. Benchmarking of deep learning irradiance forecasting models from sky images–an in-depth analysis. *Solar Energy*, 224:855–867, 2021. 2

[16] Zhihong Pang, Fuxin Niu, and Zheng O'Neill. Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons. *Renewable Energy*, 156:279–289, 2020. 2

[17] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017. 3

[18] Zhenzhou Peng, Dantong Yu, Dong Huang, John Heiser, Shinjae Yoo, and Paul Kalb. 3d cloud detection and tracking system for solar forecast using multiple sky imagers. *Solar Energy*, 118:496–519, 2015. 1

[19] Renewables REN21. Global status report, ren21 secretariat, paris, france. In *Tech. Rep.*, pages 91–93. 2017. 1

[20] Walter Richardson, Hariharan Krishnaswami, Rolando Vega, and Michael Cervantes. A low cost, edge computing, all-sky imager for cloud tracking and intra-hour irradiance forecasting. *Sustainability*, 9(4):482, 2017. 1

[21] Du Tran, Heng Wang, Lorenzo Torresani, Jamie Ray, Yann LeCun, and Manohar Paluri. A closer look at spatiotemporal convolutions for action recognition. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 6450–6459, 2018. 3

[22] William VanDeventer, Elmira Jamei, Gokul Sidarth Thirunavukkarasu, Mehdi Seyedmahmoudian, Tey Kok Soon, Ben Horan, Saad Mekhilef, and Alex Stojcevski. Short-term pv power forecasting using hybrid gasvm technique. *Renewable energy*, 140:367–379, 2019. 2

[23] Fei Wang, Zhiming Xuan, Zhao Zhen, Yu Li, Kangping Li, Liqiang Zhao, Miadreza Shafie-khah, and João PS Catalão. A minutely solar irradiance forecasting method based on real-time sky image-irradiance mapping model. *Energy Conversion and Management*, 220:113075, 2020. 1

[24] Jessica Wojtkiewicz, Matin Hosseini, Raju Gottumukkala, and Terrence Lynn Chambers. Hour-ahead solar irradiance forecasting using multivariate gated recurrent units. *Energies*, 12(21):4055, 2019. 2

[25] Phillip Wood-Bradley, José Zapata, John Pye, et al. Cloud tracking with optical flow for short-term solar forecasting. 2012. 1

[26] Ke Yan, Hengle Shen, Lei Wang, Huiming Zhou, Meiling Xu, and Yuchang Mo. Short-term solar irradiance forecast-

ing based on a hybrid deep learning methodology. *Information*, 11(1):32, 2020. 2

[27] Yunjun Yu, Junfei Cao, and Jianyong Zhu. An lstm short-term solar irradiance forecasting under complicated weather conditions. *IEEE Access*, 7:145651–145666, 2019. 2