



Cite this: *Energy Environ. Sci.*,  
2018, 11, 1811

# Solar PV output prediction from video streams using convolutional neural networks

Yuchi Sun,<sup>a</sup> Gergely Szűcs<sup>b</sup> and Adam R. Brandt<sup>a\*</sup>

Solar photovoltaic (PV) installation is growing rapidly across the world, but the variability of solar power hinders its further penetration into the power grid. Part of the short-term variability stems from sudden changes in meteorological conditions, *i.e.*, change in cloud coverage, which can vary PV output significantly over timescales of minutes. Images of the sky provide information on current and future cloud coverage, and are potentially useful in inferring PV generation. This work uses convolutional neural networks (CNN) to correlate PV output to contemporaneous images of the sky (a “now-cast”). The CNN achieves test-set relative-root-mean-square error values (rRMSE) of 26.0% to 30.2% when applied to power outputs from two solar PV systems. We explore the sensitivity of model accuracy to a variety of CNN structures, with different widths, depths, and input image resolutions among other hyper-parameters. This success at “now-cast” prediction points to possible future uses for short-term forecasts.

Received 1st December 2017,  
Accepted 9th April 2018

DOI: 10.1039/c7ee03420b

rsc.li/ees

## Broader context

Despite the steadily increasing adoption of solar power across the world, the intermittency of solar power remains a significant problem. Some of this intermittency comes from the volatility of short-term meteorological conditions. The possibility of experiencing sudden and drastic change in solar panel output has raised concern among grid authorities. Better understanding of these short-term meteorological conditions will allow improved integration of solar power. The development of multi-layer convolutional neural networks (CNNs) is the major driving force behind the recent explosion in “deep learning” application and performance. In this study, we explore the inference of solar panel output solely from concurrent local sky images with a CNN. This research is the first time that CNNs – and by extension deep learning – have been applied to predict solar panel output. We demonstrate that sky images are useful in inferring PV panel output, and CNN is a suitable structure in this application. In the future, this technology could be used to reduce uncertainty in scheduling solar power in the real time market.

## 1 Introduction

Solar photovoltaic (PV) power is playing an increasingly large role in electricity generation. Nameplate PV capacity grew by 28% in the year 2015, reaching 228 GW.<sup>1</sup> Globally, the 50 GW of PV panels installed in 2015 represents over 20% of the net annual addition to power capacity.<sup>2</sup> In terms of electricity generation though, PV is still small; PV currently meets 1.2% of world electricity demand and less than 1% of U.S. demand.<sup>1</sup> Some regions have higher shares; California met 4.3% of electricity demand through PV in 2015 due to its sunny weather and strong policy incentives.<sup>3</sup> As the cost of solar panels continues to decline<sup>4</sup> and awareness of climate change rises, the growth of PV is expected to continue.

Despite the rapid growth of PV, the variability of solar power hinders its large scale deployment, and serves as a challenge in

replacing fossil fuels with solar power. Long-term average power output of PV is well characterized by diurnal and seasonal changes in solar insolation and average cloud cover. However, short-term PV output is much harder to predict due to meteorological volatility, especially cloud coverage. In partly-cloudy conditions, the power output from PV panels can drop to near zero in less than a minute.<sup>5</sup> State of the art short-term predictions of PV output are poor, with a root mean square error (RMSE) of order 200–400 W m<sup>−2</sup> for minute level predictions<sup>6</sup> in partly cloudy conditions. More accurate short-term solar PV forecasts will be useful in a variety of contexts. Generation utilities can use forecasts to develop bidding strategies for real-time markets. Transmission operators and balancing authorities can utilize predictions in determining the need for operating reserves or ancillary services. And large end users with PV installed could use predictions to modulate power demand to avoid peak power penalties (*i.e.* reduce electric car charging in response to cloud cover).

Sky images taken in the vicinity of a PV system can provide information relevant to predicting PV output. Within the image,

<sup>a</sup> Department of Energy Resources Engineering, Stanford University, CA, USA.  
E-mail: [abrandt@stanford.edu](mailto:abrandt@stanford.edu)

<sup>b</sup> Department of Mathematics, Stanford University, CA, USA

the position of the sun directly correlates with theoretical insolation. The distribution, color, and transparency of clouds affects light absorption and reflection. The movement and shape of the clouds even forebodes future sky images and consequently future PV output. Such information is hard to extract and understand with explicit models. With the recent development in deep learning, we explore whether a convolutional neural network (CNN) can learn the correlation between sky images and PV output. In this work, we focus on assessing how images correlate with contemporaneous PV output (a “now-cast”). In the future this could be expanded to using present and past image sequences to predict future output.

## 2 Related work

Prior work on PV output prediction can be divided into two categories: models used for clear sky predictions, and models for general conditions (sunny and cloudy).

Under clear sky conditions, the variation in insolation, and by extension the output of a solar panel, are caused by changing solar zenith angle and air mass. This variability can be accurately characterized with relatively simple equations, termed as the clear sky models. Eqn (1) below is one example from this group of models:<sup>7</sup>

$$\text{GHI} = 1098 \cos(z) e^{\frac{-0.057}{\cos(z)}} \quad (1)$$

In which GHI is the global horizontal insolation ( $\text{W m}^{-2}$ ), and  $z$  is the solar zenith angle (rad). Clear sky models that only utilize the solar zenith angle  $z$ , such as eqn (1), can achieve a relative root mean square error (rRMSE) of  $\approx 10\%$ .<sup>8</sup> Therefore, little uncertainty exists in clear sky prediction.

PV output is harder to predict in general conditions due to the volatility of meteorological systems. This mostly includes changes in cloud coverage of the sun, but also to a lesser extent in water vapor content and aerosol concentration. These effects are largely local and short-term in nature, and are the primary sources of uncertainty in short-term solar panel output.

Researchers have tried to correlate sky images with irradiation. For instance, Kalisch *et al.*<sup>9</sup> used images collected on four transatlantic research cruises to calculate cloud coverage information as parameters for calculating solar downward radiation at the sea surface. Similarly Antón *et al.*<sup>10</sup> quantified one-minute variability of the GHI in Granada, Spain and correlated the variability with the cloud cover obtained from a sky imager.

Artificial neural networks (ANN) have also been applied in predicting insolation. However, these ANN models mostly rely on power generation history rather than images. Sfetsos *et al.*<sup>11</sup> developed a forecast model for mean hourly global horizontal insolation (GHI) based on several different structures of ANNs, whose RMSE revolves around  $40 \text{ W m}^{-2}$  for an hourly forecast. Cao *et al.*<sup>12</sup> combined a recurrent ANN with wavelet analysis for the forecast of daily solar irradiance, which produced a RMSE around  $30 \text{ W m}^{-2}$ . Paoli *et al.* used ANNs for the prediction of daily GHI,<sup>13</sup> and achieved a RMSE of around  $124 \text{ W m}^{-2}$ .

Lauret *et al.*<sup>14</sup> compared the performance of neural network, support vector machine (SVM) and auto-regressive models for hourly GHI prediction. Aside from ANN, SVM is also a popular modeling choice; Sharma *et al.*<sup>15</sup> and Zeng *et al.*<sup>16</sup> both used meteorological input to predict hourly insolation with SVM. Note that these studies focus on hourly or daily insolation, which relatively smoothes out the influence of cloud coverage change.

Very little work has examined combining graphical information and neural networks for PV output forecast. In the cases to date, only traditional ANNs are used. Chu *et al.*<sup>17</sup> fed the cloudiness calculated from an image and lagged output into an ANN and achieved an RMSE between  $77.5$  and  $85.5 \text{ W m}^{-2}$  for minute-scale insolation. Chu trained another ANN which directly takes the mean and variance of color intensity of cloud images as input.<sup>18</sup> This model achieved an RMSE of  $93.1$ – $131.8 \text{ W m}^{-2}$  for 5–20 minute-ahead forecasts. Only utilizing the mean and variance leaves out a wealth of information provided by sky images, which could potentially be better extracted through CNN.

## 3 Data and model

For this problem, the input feature is the sky image at a certain time point  $T_0$ , while the prediction target is the PV panel output at the same time. In the following section, we present the dataset used and the baseline SUNSET model structure.

### 3.1 Sky images

Video recordings of the sky during daytime are taken using a 6 megapixel ( $3072 \times 2048$ ) fish-eye camera with a field of view (FOV) of  $360^\circ$ . The camera is installed on the roof of the Green Earth Sciences Building (GESB) at Stanford University. The wide FOV of a fish-eye camera provides maximal information in a single frame.

Because PV data are available at 1 min resolution (see Section 3.2), frames are snapshotted from the video with the same frequency at corresponding PV output times. Videos were collected from December 2016 to June 2017, with small interruptions due to wiring failure and/or electrical failure of the camera from excessive precipitation. Fig. 1 shows examples of sky images in different weather conditions. Images are down-sampled to a resolution of  $64 \times 64$  in our baseline case to reduce model training time. We explore the effects of input image resolution in Section 4.5.

### 3.2 Solar panel output

PV panel output data are collected from two solar PV arrays close to the camera on the Stanford University campus. The solar panels are located on top of Stanford's Yang and Yamazaki Energy and Environment (henceforth Y2E2) building and Jen-Hsun Huang Engineering Center (henceforth Huang). The two arrays are  $\approx 250 \text{ m}$  and  $125 \text{ m}$  away from the camera, respectively. This is a short distance considering the spatial scale of the clouds. Both Huang and Y2E2 panels report their

12:51:00 Mar.14.2017 10:50:00 Mar.16.2017

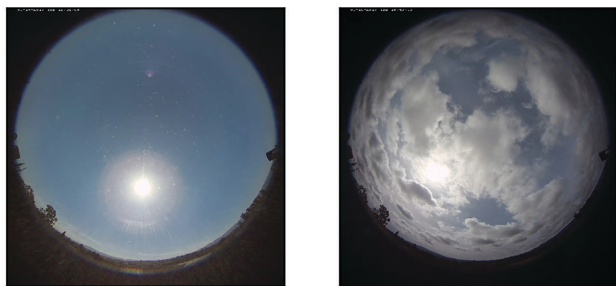


Fig. 1 Sky image examples with time stamp on top: the image on the left shows a completely clear day, while the image on the right presents a cloudy day with obstructed irradiation.

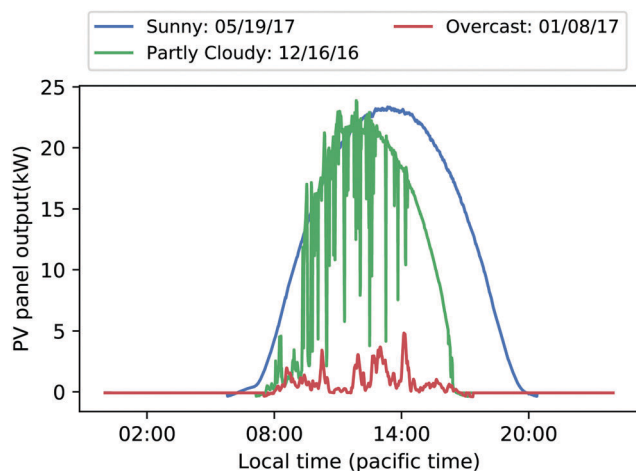


Fig. 2 Minute-by-minute power output from the Huang solar panels under different weather conditions (clear, mixed, and overcast).

output minutely during the daytime. The panel output logs suffer from occasional sensor failure, which results in no data or invalid data for some days of observations. Both the Y2E2 and Huang data streams are filtered to remove erroneous data.

Cloud coverage of the sun obstructs direct normal irradiation (DNI) and leaves only diffuse sunlight to pass through, therefore significantly diminishing electricity generation. Examples of PV panel output profile can be seen in Fig. 2.

### 3.3 Data preprocessing

The total number of valid observations – where synchronous image and PV output during daytime hours are present – totals almost 25 thousand observations for Huang, and over 14 thousand for Y2E2. As per common statistical modeling practice,<sup>19</sup> each dataset is divided into training, validation, and testing sets. The testing set is three complete days of data which represent a sunny day, a partly cloudy day and an overcast day. The testing set is removed completely from the data and never used during model development. The remaining data are split randomly at the observation level into 80% training set and 20% validation set. Training data are used to build the model while validation data are used to diagnose the model fitness, tune the hyper

Table 1 Number of daytime observations in each dataset, where video and PV output are simultaneously valid. And numbers of daytime images rejected

Dataset	Training	Validation	Test	Sum	Rejected
Y2E2	10 338	2585	1479	14 402	824
Huang	15 908	3977	2517	22 402	2495

parameters, and terminate analysis when over-fitting becomes apparent (see Section 3.4). Table 1 shows the number of data points in each set, and also the number of images rejected due to unavailable or invalid PV output data.

### 3.4 Model and CNN architecture

Here we build a specialized deep CNN named SUNSET (Stanford University Neural network for Solar Electricity Trend). Deep CNNs are successful because multi-layer stacking of convolution operations allows better abstraction and comprehension of complex information<sup>20</sup> (like graphical imagery). Widely used CNNs, such as AlexNet,<sup>21</sup> VGG,<sup>22</sup> Inception,<sup>23</sup> and ResNet<sup>24</sup> share a basic common framework; the input image is sequentially passed through several Conv-Pool structures, each of which consists of a set of several convolutional layers and a pooling layer at the end. After the last Conv-Pool structure, the processed input is flattened into a vector and fed into several fully connected layers. The fully connected layers generate a regression result (to predict continuous data) or a classification score (to predict discrete or categorical data).

The SUNSET structure follows the same paradigm as above, and is illustrated in Fig. 3. Input images first go through two Conv-Pool structures. Each Conv-Pool structure includes one convolutional layer, one batch normalization layer, and one pooling layer, in that order. The convolutional layer utilizes a filter size of  $3 \times 3$ , with a step length of one and same-value padding. The activation function used here is Rectified Linear Unit (ReLU).<sup>25</sup> The first Conv-Pool structure contains 12 filters, while the second contains 24 filters. Batch normalization is used after convolution, which makes the model train faster and increases robustness.<sup>26</sup>  $2 \times 2$  max pooling is used in the pooling layer. Max pooling, which aggregates spatially neighboring features, is considered crucial in obtaining translation-invariant features.<sup>27</sup>

After the two Conv-Pool structures, the processed input is passed through two fully connected layers, each containing 1024 neurons, with ReLU as the activation function. A final regression step with linear weights is used to produce the predicted PV panel output.

We train the network using the Adam optimizer,<sup>28</sup> a stochastic gradient descent optimizer popular in the field of machine learning. The loss function to be minimized is defined as the mean square error of the PV output prediction:

$$\text{Loss} = \frac{1}{N} \sum_{i \in S} (P_i - G_i)^2, \quad (2)$$

where  $S$  is the set of samples,  $N = |S|$  is the number of samples,  $P_i$  is the predicted output and  $G_i$  is the true output for sample  $i$ .

Since Adam is a stochastic optimizer, the trained model is slightly different for each run. To mitigate this variability,

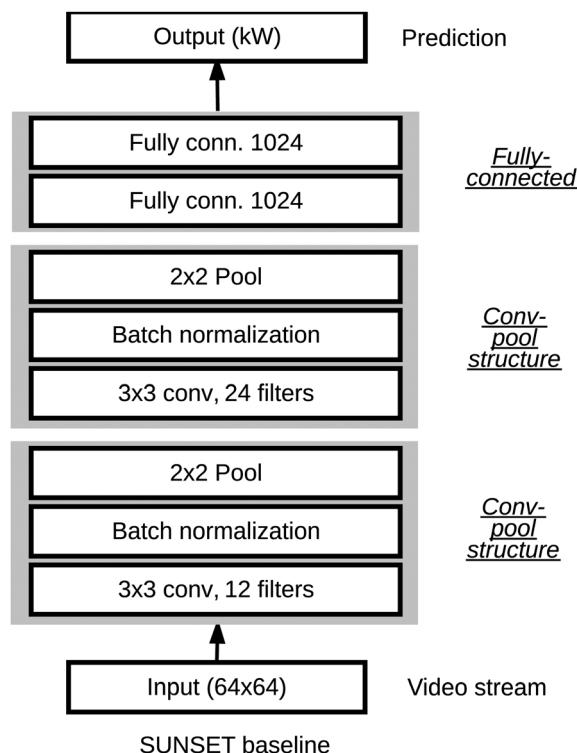


Fig. 3 SUNSET deep CNN network architecture diagram, with input data flow starting from the bottom, going through layers of operation, and forming an output.  $3 \times 3$  Conv., 12 denotes a convolutional layer with 12 filters, each filter with a size of  $3 \times 3$ . Fully conn. 1024 denotes a fully connected layer with 1024 neurons.

Table 2 Summary of sensitivity analysis on the SUNSET baseline model, showing the list of values explored for each architecture choice. Baseline settings in bold font

CNN architecture choice	Sensitivity values
Depth (convolutional layers)	[0, 1, 2, 3, 4, 5, 6, 7, 8]
Depth (fully connected layers)	[0, 1, 2, 3, 4]
Width	[1, 2, 3, 4, 6, 8, <b>12</b> , 16, 24]
Input image side length	[1, 2, 4, 8, 12, 16, 24, 32, 48, <b>64</b> , 96, 128]
Neurons in fully connected layer	[32, 64, 128, 256, 512, 768, <b>1024</b> , 1280, 1536, 2048]

all experiments are repeated 10 times and the loss or model fit for each experiment is reported using box-and-whisker plots. To test the robustness of our structural choice of SUNSET, we experiment with different architecture choices, including the model width, the depth, the number of neurons in fully connected layers, and the resolution of images in Sections 4.2–4.5. Table 2 summarizes all the sensitivity runs we have done on the baseline SUNSET model, which presents the architecture choices we experimented on and the list of values they have taken.

## 4 Results

### 4.1 Results for SUNSET baseline model

The baseline SUNSET model structure is used to train two separate models for the Huang and Y2E2 datasets. Fig. 4 shows

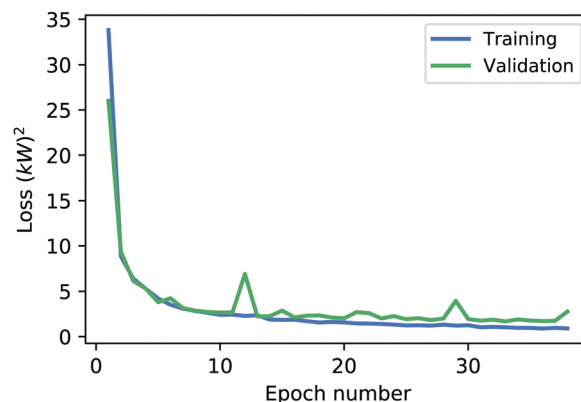


Fig. 4 Loss function over the course of training for the baseline SUNSET model and Huang data.

the typical profile of loss function (mean squared error) in a training cycle. Note that in each epoch every data point in the training set has been used exactly once. The training set and validation set losses initially both decrease. As training continues, the validation set loss stops decreasing while the training set loss continues to decline. This is a sign of over-fitting. We therefore stop training after five epochs of non-decreasing validation set loss. For the baseline SUNSET model this typically occurs after  $\approx 20$  epochs.

The Huang and Y2E2 panels are of different sizes and their outputs are of different magnitudes in kW. In order to compare results across these datasets, we use the relative root mean square error (rRMSE):

$$\text{rRMSE} = \sqrt{\frac{N \times \text{Loss}}{\sum_{i \in S} (G_i - \bar{G})^2}} \quad (3)$$

where  $S$  is the set of samples,  $N = |S|$  is the number of samples, Loss is the calculated mean square error as defined in eqn (2),  $G_i$  is the true output for sample  $i$ , and  $\bar{G}$  is the mean of the true output in  $S$ . Intuitively, rRMSE tell us the fraction of the standard deviation of the true data that is left unexplained by the model. As shown in Fig. 5, the SUNSET baseline model achieves overall a similar level of performance for the Huang and Y2E2 data sets, with the model trained on the Huang dataset doing slightly better in training, validation and the testing set. This might be attributed to the fact the Huang dataset contains more data points and more partly cloudy days, thus making its model more general and robust. We can also see that the training set rRMSE, at 9.9% to 10.9%, is smaller than that of the validation set at 12.8% to 16.3%, which is again smaller than that of the test set at 26.0% to 30.1%. This is expected in most statistical modeling applications.

A more intuitive way to demonstrate the power of the CNN in inferring PV output is through plotting a time series of the test set predictions, which is shown in Fig. 6 and Table 3. We can see that the SUNSET baseline model does an excellent job in sunny condition, and struggles somewhat in partly cloudy and overcast conditions. Across the two datasets, the prediction



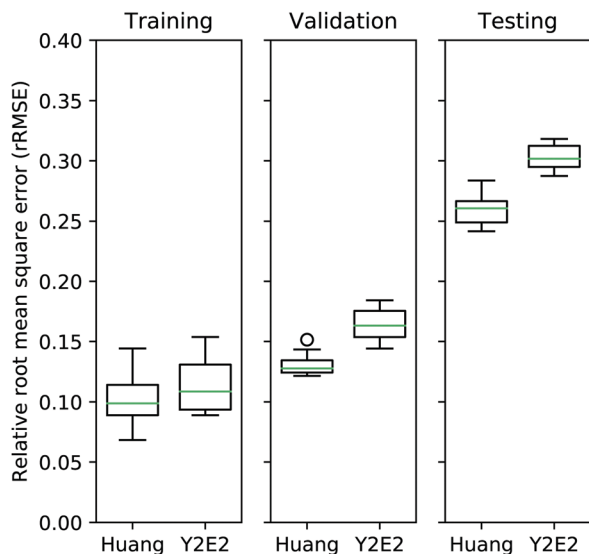


Fig. 5 Comparison of rRMSE between the Huang and Y2E2 datasets, for the training, validation and testing set.

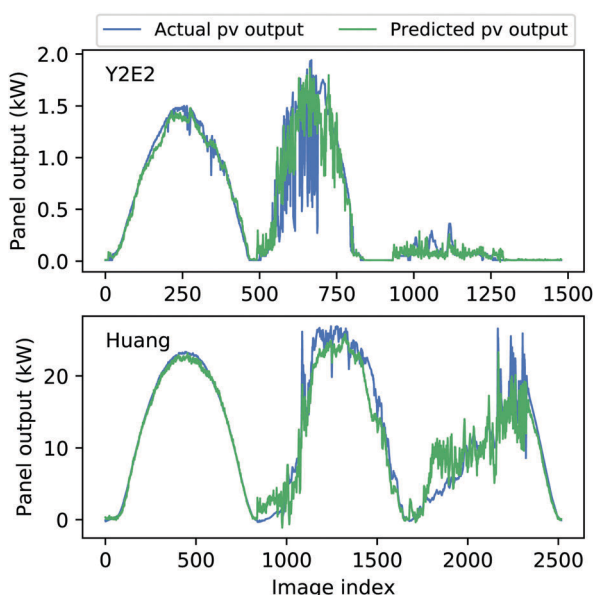


Fig. 6 Time series of the test set prediction for the Y2E2 and Huang datasets, each with a sunny day, a mixed day and an overcast day.

Table 3 A summary of test set prediction errors measured by RMSE, under sunny, partly cloudy, and overcast conditions

Dataset	Sunny RMSE (kW)	Partly cloudy RMSE (kW)	Overcast RMSE (kW)
Y2E2	0.064	0.338	0.051
Huang	0.528	2.483	2.753

error in partly cloudy weather is about 5 times that of the sunny weather. However, inferring solar power in cloudy weather is an inherently harder problem as discussed in Section 1. Nonetheless, the inference result still largely follows the complicated trend of ups and downs in cloudy conditions.

## 4.2 Sensitivity to varying CNN model depth

We perform all sensitivity analysis below on the Y2E2 dataset due to its smaller size and therefore simpler training. The CNN model depth represents the numbers of layers, usually understood as the number of convolutional layers and fully connected layers. We explore the model's sensitivity to zero to eight layers of convolutional operations in total for the two Conv-Pool layers. The Conv-Pool structures are followed by zero to four fully-connected layers. Therefore in total, 45 different model structures are tested, and their results in terms of average validation set loss is presented in Fig. 7.

There is a consistent decrease in median validation loss when we increase the number of convolution layers from 0 to 1, demonstrating the importance of including convolution in our model. However, this benefit is less pronounced when we further increase numbers of convolution layers beyond two. Similarly, more than two fully connected layers does not seem to lead to better performance. These results suggest that 2 convolutional layers and 2 dense layers are sufficient for this application.

One explanation for the relatively small number of convolutional layers needed compared to traditional image classification models is that inference of solar irradiance requires a lower level of abstraction. Image classification models need to extract highly abstract features such as a face or eyes, while a good model for irradiance likely uses low-level features such as brightness and color.

## 4.3 Sensitivity to varying model width

The CNN model width represents the number of filters applied during each convolution layer. We explore how model performance varies with a range of 1 to 24 filters as the model width in the first Conv-Pool structure, while the second Conv-Pool structure is set to twice that amount. The model otherwise has the same structure as the baseline SUNSET model (Section 4.1).

From Fig. 8, we can see that performance steadily improves up to 6 filters. More than 6 filters does not significantly improve

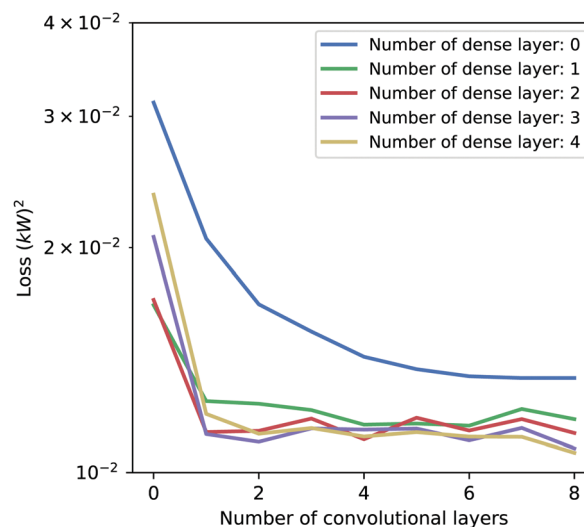


Fig. 7 Median validation set loss of the SUNSET model with varying model depth, both in convolution layer number and fully connected layer number.

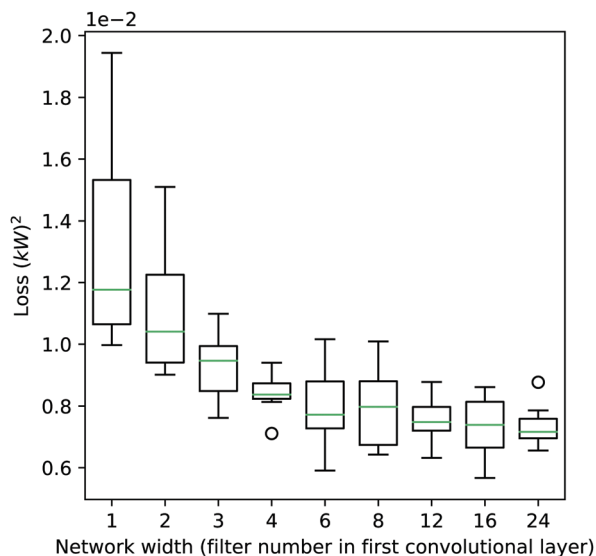


Fig. 8 Box plot of the validation set loss of the SUNSET model with varying model width.

the median validation set loss. The relatively small number of filters needed might be an indication that our model is only looking for a small amount of structural information from the picture, such as average cloud coverage and solar position. Solar PV prediction does not need to recognize a large array of features as traditional image classifiers do, which classifies images into thousands of classes.

#### 4.4 Sensitivity to varying numbers of neurons in the fully-connected layers

As is the case with canonical image classification CNNs like AlexNet<sup>21</sup> and VGGNet,<sup>22</sup> the majority of the trainable parameters, and by extension the size of the model, come from the large number of weights in the fully-connected layers. Here we examine if it is possible to cut down the numbers of neurons in the fully-connected layers.

From Fig. 9, we can see that from 32 to 1280 the validation loss sees a fluctuating decreasing trend, and a neuron number in excess of 1280 is required to achieve optimal performance.

#### 4.5 Sensitivity to varying input image resolution

Theoretically, a higher input image resolution would provide us with more information and a more accurate prediction. However, higher resolution also requires a longer training time and the added information may be redundant. The SUNSET baseline model works with images down-sampled to 64 pixels  $\times$  64 pixels in size. For sensitivity analysis we test a range of different image resolutions from 1 pixel  $\times$  1 pixel to 128  $\times$  128 pixels.

As we can see from Fig. 10, the validation loss decreases with higher resolution up to 48  $\times$  48 pixels, after which the performance plateaus. This is understandable since an image of this size can still show the sun's position and the extent of cloudiness.

One interesting phenomenon to note is that while using a lower input image resolution reduces the training time needed per epoch, the training requires more epochs to converge.

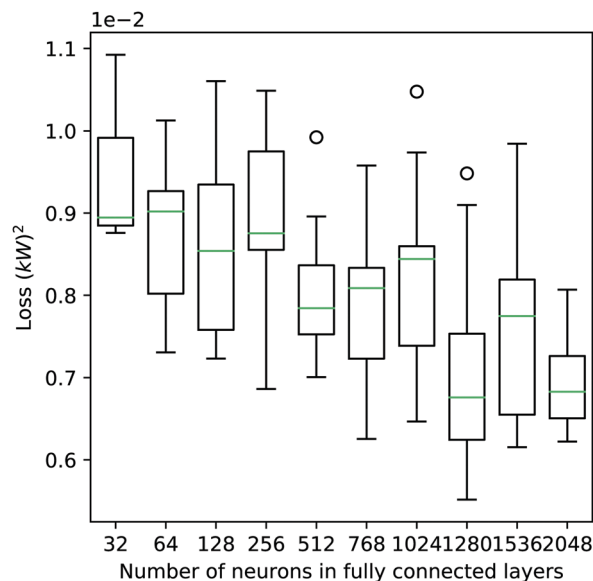


Fig. 9 Box plot of the validation set loss of the SUNSET model with varying number of neurons in the fully connected layers.

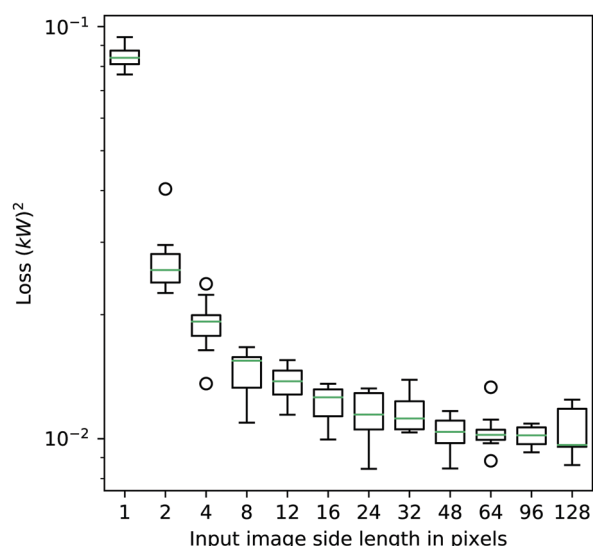


Fig. 10 Box plot of the validation loss of the SUNSET model for various input image resolutions.

In our tests of SUNSET, the largest resolution is 128 times larger than the smallest one (and contains 128<sup>2</sup> times more information). This highest resolution dataset costs  $\approx$  30 times more computational time per epoch, but requires about half the numbers of epochs to reach convergence.

Table 4 summarizes all the sensitivity analysis results in terms of median validation loss, with optimal and close to optimal performance highlighted.

## 5 Future work

In this work, we concentrated solely on how informative images are in indicating PV output. It would be interesting to explore

**Table 4** Summary of model performance for the Y2E2 dataset for all sensitivity analysis. The performance is reported in terms of median validation loss with a unit of  $10^{-2}$  (kW)<sup>2</sup> for different depths, widths, input image resolutions, and number of neurons in the fully connected layer. The best performing model in each sensitivity analysis is marked with [\*], while models with loss of no more than 5% of the optimal model are marked in bold

Conv layers	Loss@FC layer: 0	Loss@FC layer: 1	Loss@FC layer: 2	Loss@FC layer: 3	Loss@FC layer: 4	Width	Loss	Input resolution	Loss	Neurons in FC layers	Loss
0	3.126	1.665	1.704	2.035	2.362	1	1.177	1	8.398	32	0.895
1	2.087	1.220	<b>1.128*</b>	1.129	1.158	2	1.041	2	2.568	64	0.902
2	1.678	1.243	<b>1.138</b>	<b>1.100</b>	<b>1.116</b>	3	0.947	4	1.923	128	0.854
3	1.544	1.222	1.206	1.128	1.157	4	0.837	8	1.548	256	0.875
4	1.416	<b>1.149</b>	<b>1.131</b>	1.132	<b>1.110</b>	6	0.772	12	1.379	512	0.785
5	<b>1.374</b>	<b>1.167</b>	1.200	1.149	1.138	8	0.798	16	1.261	768	0.809
6	<b>1.346</b>	<b>1.139*</b>	<b>1.154</b>	<b>1.101</b>	<b>1.118</b>	12	<b>0.748</b>	24	1.145	1024	0.844
7	<b>1.345</b>	1.251	1.192	1.171	<b>1.116</b>	16	<b>0.739</b>	32	1.120	1280	<b>0.676*</b>
8	<b>1.333*</b>	<b>1.191</b>	<b>1.137</b>	<b>1.069*</b>	<b>1.070*</b>	24	<b>0.716*</b>	48	1.039	1536	0.775
								64	1.022	2048	<b>0.683</b>
								96	1.020		
								128	<b>0.966*</b>		

including other variables as model inputs, such as present and historical meteorological conditions, to achieve an even higher accuracy.

Another future direction is that in practice, predicting the future PV output from a system is of most value. One possible approach to future prediction is to use a stack of images (going back 5 minutes, say), and trying to predict the output of the panels (say 5 minutes ahead). The stack of images would provide necessary information on the previous movement of the clouds, which is vital for predicting future cloud distribution.

## 6 Conclusion

This work explores inferring minute-level solar panel output from sky images, using convolutional neural networks. It validates the intuitive fact that sky images contains abundant information on insolation, and by extension local PV panel generation. Convolutional neural networks are shown to be a suitable structure for interpreting the information contained in the sky images. For this particular problem, two convolutional layers, in which one has 12 filters and the other 24 filters, combined with two fully-connected layers, is the network architecture that is deemed the most suitable. Both convolutional layers and dense layers are important in terms of model performance, while little benefit exists in extending either of their layer numbers beyond two layers. Networks wider than 6 filters do not consistently improve model performance. The input image resolution shows little influence on the model's prediction performance in excess of  $48 \times 48$  pixels. The best architecture ended up with a validation set rRMSE of  $\approx 15\%$ , and a test set rRMSE of  $\approx 28\%$ .

## Conflicts of interest

There are no conflicts to declare.

## References

- 1 IEA-PVPS, *Trends in Photovoltaic Applications – 2016 – 21st edn*, International energy agency technical report, 2016.
- 2 R. 21, *Renewables 2016 global status report*, Renewable energy policy network for the 21st century technical report, 2016.
- 3 C. E. Commission, *California Electrical Energy Generation 2010–2015*, California energy commission technical report, 2016.
- 4 IEA-PVPS, 2014 Snapshot of Global PV Markets, International energy agency technical report, 2014.
- 5 R. Marquez and C. F. Coimbra, *Sol. Energy*, 2013, **92**, 327–336.
- 6 S. Quesada-Ruiz, Y. Chu, J. Tovar-Pescador, H. Pedro and C. Coimbra, *Sol. Energy*, 2014, **102**, 267–275.
- 7 B. Haurwitz, *J. Meteorol.*, 1945, **2**, 154–166.
- 8 M. J. Reno, C. W. Hansen and J. S. Stein, Global horizontal irradiance clear sky models: implementation and analysis, Sandia national laboratories technical report, 2012.
- 9 J. Kalisch and A. Macke, *Meteorol. Z.*, 2008, **17**, 603–611.
- 10 M. Antón, J. Gil, A. Cazorla, J. Fernández-Gálvez, I. Foyo-Moreno, F. Olmo and L. Alados-Arboledas, *Atmos. Environ.*, 2011, **45**, 4815–4821.
- 11 A. Sfetos and A. H. Coonick, *Sol. Energy*, 2000, **69**, 169–178.
- 12 S. Cao and J. Cao, *Appl. Therm. Eng.*, 2005, **25**, 161–172.
- 13 C. Paoli, C. Voyant, M. Muselli and M.-L. Nivet, *Sol. Energy*, 2010, **84**, 2146–2160.
- 14 P. Lauret, C. Voyant, T. Soubdhan, M. David and P. Poggi, *Sol. Energy*, 2015, **112**, 446–457.
- 15 N. Sharma, P. Sharma, D. Irwin and P. Shenoy, Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on, 2011, pp. 528–533.
- 16 J. Zeng and W. Qiao, *Renewable Energy*, 2013, **52**, 118–127.
- 17 Y. Chu, H. T. Pedro, L. Nonnenmacher, R. H. Inman, Z. Liao and C. F. Coimbra, *J. Atmospheric Ocean. Technol.*, 2014, **31**, 1995–2007.
- 18 Y. Chu, M. Li, H. T. Pedro and C. F. Coimbra, *Renewable Energy*, 2015, **83**, 234–244.
- 19 G. James, D. Witten, T. Hastie and R. Tibshirani, *An introduction to statistical learning*, Springer, New York, 2013.
- 20 Y. LeCun, Y. Bengio and G. Hinton, *Nature*, 2015, **521**, 436–444.
- 21 A. Krizhevsky, I. Sutskever and G. E. H. Brandt, *Advances in neural information processing systems*, 2012, pp. 1097–1105.

- 22 K. Simonyan and A. Zisserman, 2014, arXiv preprint arXiv:1409.1556.
- 23 C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
- 24 K. He, X. Zhang, S. Ren and J. Sun, Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
- 25 V. Nair and G. E. Hinton, Proceedings of the 27th international conference on machine learning (ICML-10), 2010, pp. 807–814.
- 26 S. Ioffe and C. Szegedy, *International Conference on Machine Learning*, 2015, pp. 448–456.
- 27 F. J. Huang, Y.-L. Boureau and Y. LeCun, *et al.*, Computer Vision and Pattern Recognition, 2007, CVPR'07, IEEE Conference on, 2007, pp. 1–8.
- 28 D. Kingma and J. Ba, 2014, arXiv preprint arXiv:1412.6980.