Huertas-Tato, J., Galván, I.M., Aler, R. et al. Using a Multi-view Convolutional Neural Network to monitor solar irradiance. *Neural Comput & Applic* (2021).

Using a Multi-view Convolutional Neural Network to monitor solar irradiance

Javier Huertas-Tato · Inés M. Galván · Ricardo Aler · Francisco Javier Rodríguez-Benítez · David Pozo-Vázquez

Abstract In the last years, there is an increasing interest for enhanced method for assessing and monitoring the level of the global horizontal irradiance (GHI) in photovoltaic (PV) systems, fostered by the massive deployment of this energy. Thermopile or photodiode pyranometers provide point measurements, which may not be adequate in cases when areal information is important (as for PV network or large PV plants monitoring). The use of All Sky Imagers paired convolutional neural networks, a powerful technique for estimation, has been proposed as a plausible alternative. In this work, a convolutional neural network architecture is presented to estimate solar irradiance from sets of ground-level Total Sky Images. This neural network is capable of combining images from three cameras. Results show that this approach is more accurate than using only images from a single camera. It has also been shown to improve the performance of two other approaches: a Cloud Fraction model and a Feature Extraction model.

Keywords Deep Learning · Multi-view image · Solar irradiance

Corresponding Author: Javier Huertas-Tato
Universidad Europea de Madrid
E-mail: javier.huertas@universidadeuropea.es

Inés M. Galván
Universidad Carlos III de Madrid
E-mail: igalvan@inf.uc3m.es

Ricardo Aler
Universidad Carlos III de Madrid
E-mail: aler@inf.uc3m.es

Francisco Javier Rodríguez-Benítez
Universidad de Jaen
E-mail: fbenitez@ujaen.es

David Pozo-Vázquez
Universidad de Jaen
E-mail: dpozo@ujaen.es

1 Introduction

The last decade has seen considerable effort to make the solar energy a viable alternative to conventional energy generation systems. Many countries have already reached a notable solar share in their energy mixes and further growth is expected in the near future [1], particularly for photovoltaic (PV) solar technology. For that reason, there is an increasing interest in enhanced methods for assessing and monitoring global horizontal irradiance (GHI) in PV systems. Remarkably, GHI monitoring is important to manage PV distributed networks, micro grids or smart homes [2, 3, 4] and, in general, for PV plants performance assessment [5].

GHI measurements are typically provided by thermopile or photodiode pyranometers, having 95% confidence intervals from ±3% to ±10% for GHI depending on the instruments quality and the sun position [6]. However, pyranometers provide point measurements, which may not be adequate in cases when areal information is important (as for PV network or large PV plants monitoring). In this case, the use of a pyranometer network becomes an expensive choice. Aside from the use of pyranometers, a plausible way for areal solar irradiance monitoring is the use of satellite imagery [7, 8]. Nevertheless, spatial and temporal resolution of satellite irradiance estimates are relatively coarse. This make this approach not suitable for some applications and/or geographical areas. For instance, for relatively small areas or areas with a complex topography or/and showing a high spatial variability of the solar resources related to the cloud characteristics. The use of different parameters measured at the PV system (such as active power flow, cell temperature of direct current and voltage) have been used for GHI estimation [9, 10], but reliability is relatively low.

In the last years, the use of All Sky Imagers (ASI) [11] to estimate solar irradiance has been proposed. Traditional approaches are varied, consisting on extracting features and operating over them. The use the ASIs have some advantages compared to other approaches, particularly regarding the spatial representativeness of the estimates. For instance, [12] showed that, for areas of a few squared kilometers, the ASIs outperform a single pyranometer in representing the spatial irradiance distribution. The work by [13] proposed and evaluated a method for GHI estimation from ASI based on polynomials functions. Authors reported nRMSE values ranging from 6% during clear-sky to 13% for partly-cloudy conditions (9% under all sky conditions) and for 1 min data. [12] evaluated high temporal resolution (15 seconds) GHI estimates derived from an ASI using a network of 99 pyranometers. Results were evaluated on the light of the cloud type. Results proved ASI estimated accuracy to depend on the location of the pyranometers and the cloud type. The highest error were found for Cumulus (RMSE 229 W/m2), Altocumulus (180 W/m2) and Stratocumulus clouds (150 W/m2), while for clear skies the lowest errors were wound (RMSE of 60 W/m2). [14] estimated the GHI using information from a camera system with a charge-coupled device (CCD) sensor. Authors reported nRMSE values for GHI estimates to be in the range

69% for hourly averaged data. [15] proposed a machine learning procedure to estimate the GHI from ASIs. The method uses selected features of the images to train an artificial neural network (ANN) and was tested against pyranometer measurements. Authors reported nRMSE values ranging from 13% to 34%, depending of sky type, for one-minute time resolution estimates.

Machine learning algorithms used in this field have gained popularity over the years. For example, in [16], meteorological variables and geographical information is used with historical data to create an ANN estimator for monthly average irradiance over Turkey. Another approach is presented by [17] where features extracted from ASIs are combined with a neural network for DNI hourly estimations. A relevant use of ANNs is presented by [18] where a principal component analysis (PCA) is performed on the ASI to extract features. Convolutional Neural Networks (CNNs) [19] have seen widespread use in computer vision problems since their introduction. Nowadays, many fields of research are being tackled by means of CNNs, but the estimation problem has not been sufficiently explored in this regard. There are some related applications of CNNs to irradiance estimation, but most do not explore this problem directly. For example, the work of [20] proposes combining video and weather information for forecasting, via intermediate irradiance estimations. Finally, the work of [21] proposes independently using sequences of images to estimate DNI, without further information. Irradiance monitoring works rarely process the raw image. Instead, they rely on either features, cloud estimates, meteorological information, or other human-extracted information; however, when the image is independently treated, some information is lost as 2D representations lack nuance about the sky state, such as cloud volume and altitude.

In this paper the problem of irradiance estimation from raw ASI images using CNNs is addressed, in order to study whether more accurate irradiance estimations can be obtained. The methods proposed in this work aim to remove the use of expert knowledge via automatic image feature extraction using CNNs, letting the learning algorithm discover relevant information on its own. Different approaches based on CNNs are proposed and studied. First, a CNN architecture (the single-view CNN model) is designed to estimate irradiance using a set of images from a single camera. This model is based on a standard CNN but including a dilated convolution after the input channels. This aims to take advantage of spatial information distributed over wide areas of the image without increasing the number of weights. This model receives as inputs the RGB channels from the images and an additional channel containing contextual information about the sun distance. The new additional channel represents the distance from each pixel of the image to the sun. Wherever the sun hits the sensor, the image is saturated in these points, producing severe imperfections [22]. Knowing the distance to the sun center pixel may help the network to take into account this phenomenon as supported by the results from [23].

Second, CNN models are proposed to combine several simultaneous images of the sky from different cameras (the multi-view CNN models), instead of images from a single camera. Although there have been attempts to merge

information of multiple images through CNN's (e.g. [24]), no attempt has been made for irradiance estimation or forecasting as far as we know. The objective is to study whether the joint information from different cameras can improve the estimation of irradiance. These models use as inputs the RGB channels from the images and the additional distance sun channel, similarly to the single-view model. In this work, two approaches are studied: the Feature Average model (the most direct approach to combine simultaneous images from different cameras), where the features extracted by the CNN from each of the three cameras are simply averaged; and a more complex proposal, the Multi-view Pooling model, where the features obtained from the cameras by means of CNN filters are further combined and processed by additional CNN layers.

Different baseline methods have been used for comparison purposes. First, the Cloud Fraction (CF) algorithm [25], one of the most used approaches to estimate irradiance from images. Second, the Random Forest (RF) algorithm [26] using as inputs a set of features extracted from the images, instead of the raw images. Specifically, the features described in [27] have been used in this study. Finally, and with the aim to show the advantages of the CNN architectures proposed in this work, a standard CNN architecture [19] has been also employed as baseline.

All the measurements and ASI images used in this study were collected at the Abengoa Solar Platform of Solúcar (6.25 W, 37.44 N). Each example in the set contains three spatially separated images taken at the same time with different perspectives of the same sky. In order to construct the machine learning models, each picture is associated with an irradiance measurement, obtained with a pyranometer. A total of 1186 instances are available to train, validate and test the proposed models.

In this article, we present relevant contributions to both the irradiance estimation problem, as well as a novel approach to CNN architecture. The first contribution is the use of CNNs to estimate irradiance from ASIs and contextual sky information. The second contribution is a CNN approach to combine different cameras simultaneously to improve solar irradiance estimation, achieving low errors.

This document describes the method used to reach this objective in Section 2. The data that is analysed in this study is described in Section 3. The experimental methodology is defined in Section 4 and results are reported and discussed in Section 5. Finally, Section 6 summarizes the main conclusions of this work.

2 Single and Multi-view models using CNNs

In this section, the CNN models used to estimate the irradiance from a set of ASIs are described. A CNN can be interpreted as a sequence of layers that process an image. In the most common architectures, each layer applies one of three basic operations (convolution, activation, and pooling). Typically,
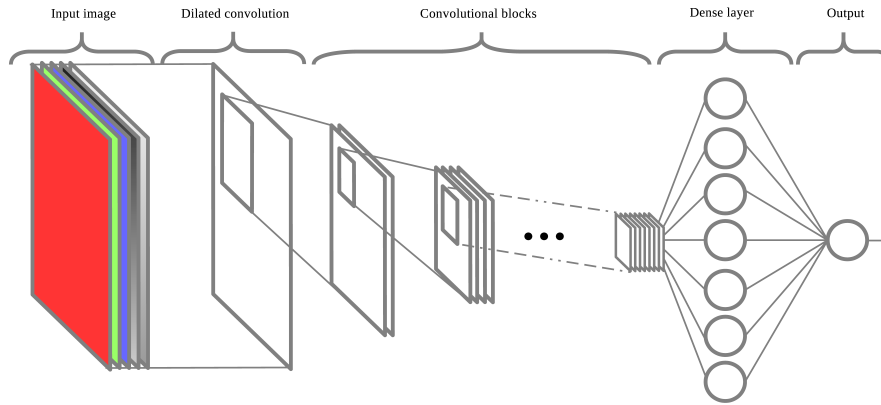
Fig. 1 Representation of the single-view CNN model.

several blocks of convolution-activation-pooling (called blocks from now on) are applied, and after them the output is flattened and processed by a standard multi-layer perceptron or MLP (known as dense layers in the context of CNNs).

Modifications over a standard LeNet [19] CNN architecture are proposed to improve estimations using the information from a single camera ASI (single-view CNN model). Finally, CNN models are proposed to use information from different cameras (multi-view CNN models). Both are described in the following sections.

2.1 Single-view CNN model

The single-view model (Fig. 1) is based on a standard LeNet architecture (see [19]) with some modifications. The model receives ASIs from a single camera, using the information of RGB channels, as usual. However, in this work, additional contextual information is used as an input to the CNN in order to obtain better irradiance estimates. A new channel is added to represent the sun distance, where each pixel contains the distance from the pixel location to the sun radius. It is known that the area around the sun is more saturated [22], and therefore we considered that the network should be aware of distances between each pixel in the image and the pixel sun. The sun centre is calculated with latitude, longitude and solar hour, features that have performed well in past works [16].

A visual representation of the additional information channels along with the original image is shown in Figure 2. In the original image (left image) it can be seen that the pixels near the sun are heavily distorted compared to the rest of the image. A solution would have been covering or filtering the sun through physical means or via image editing. However this would result in the loss of critical information about irradiance. To solve this problem, a fourth channel is created. It contains the Euclidean distance (see Eq. 1) from
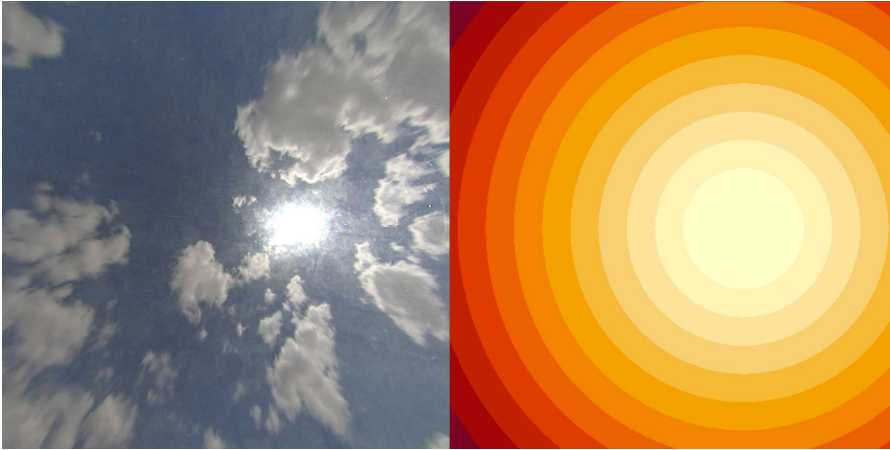
Fig. 2 Example of corrected original image and the additional channel. Left image: Original image of a partially overcast sky, composed of RGB channels. Right image: sun distance channel: pixels closer to the sun are represented as brighter pixels.

any pixel to the sun. This solution aims to include as input the information of which pixels are near the sun. Given the date, latitude, longitude, and camera calibration, the center and radius of the sun can be automatically computed [28].

Eq. 1 displays how the Euclidean distance from each pixel to the sun is computed. (x_{sc}, y_{sc}) are the coordinates of the sun centre and (x_p, y_p) are the coordinates of a pixel. This is computed over every pixel of the image, and finally the pixels inside the sun circle are adjusted to 0.

$$z = \sqrt{(x_p - x_{sc})^2 + (y_p - y_{sc})^2} \tag{1}$$

The single-view CNN model also uses a dilated convolution (see Fig. 1). Instead of directly connecting input channels with the first convolutional-activation-pooling block, the inputs are previously connected to a dilated convolution [29]. Cloud images have rich spatial information that is distributed over wide areas, much bigger areas than for standard computer vision scenarios. The dilated convolution addresses this problem by using a larger window without increasing the number of weights in the network.

Also, the commonly used max pooling operator has been replaced by the average pooling operation on every pooling layer, in order to use the information contained in every pixel. Applying max pooling would lose the irradiance information, because all values in the max pooling sliding window which are below the maximum would be lost.

2.2 Multi-view CNN models

In this work, multi-view models are proposed with the aim of studying whether the use of information from different cameras can improve the estimation of irradiance. In particular, it will be studied whether the use of three concurrent views of the same sky situation is able to improve estimations. The combination of several images from different sources is not a commonly problem addressed in the literature. It is well known that CNNs produce a series of features from the images before being processed by the dense layers. In this work, two approaches will be compared: Feature Average model (CNN-FA) and Multi-view Pooling model (CNN-MvP). Both of them perform the aforementioned generation of features, but differ in the way they are combined, and the processing after the combination. The first one uses a straightforward way to combine simultaneous images, while the second one is a more complex proposal that allows the CNNs to extract more information from images from the different cameras.

For both models proposed there is a step within the network where information is combined. Let $I_{a,b}^{i,(c)}$ be a value with position (a,b) from a channel (c) of the camera i. In the case of several cameras, a set of features would be obtained for each of them, from camera 1: $I_{a,b}^{1,(c)}$ to camera t: $I_{a,b}^{t,(c)}$. In this work, $t = 3$, but any number could be used in principle.

In the case of the **Feature Average model (CNN-FA)** (see Figure 3), a sequence of convolution-activation-pooling layers is applied to every camera. All t such sequences have the same CNN architecture (i.e. the same layers). For every image i the network CNN is the same, sharing the same weights for each image. Those layers generate the $I_{a,b}^{i,(c)}$ mentioned before. Given the CNN's for each camera are identical, every feature found has a counterpart in the feature set from other images. This allows to average the vector of outputs feature-wise, merging the information from the different cameras. The CNN-FA model combines them by averaging the corresponding feature maps, as shown in Eq. 2. In short, for every channel c and pixel (a,b), the average for all t cameras is computed. Once the $I_{a,b}^{\prime(c)}$ have been computed, they are flattened and fed to the final dense layers.

$$I_{a,b}^{\prime(c)} = \text{average}(I_{a,b}^{i,(c)}) \tag{2}$$

The averaging of CNN features obtained from images of different cameras may be too simple an approach to detect further synergies between images. The **Multi-view Pooling model (CNN-MvP)** (see Figure 4) also applies the same CNN architecture to obtain the features from each camera (in this case, these networks will be named CNN_1). But it has two differences with the CNN-FA model. First, max-pooling is used to combine the features (instead of the average), as displayed in Eq. 3. Next, a second convolution-activation-pooling set of layers, CNN_2, is applied to the results. In the case of CNN-MvP, they will be explicitly named first set of convolutional blocks (CNN_1)
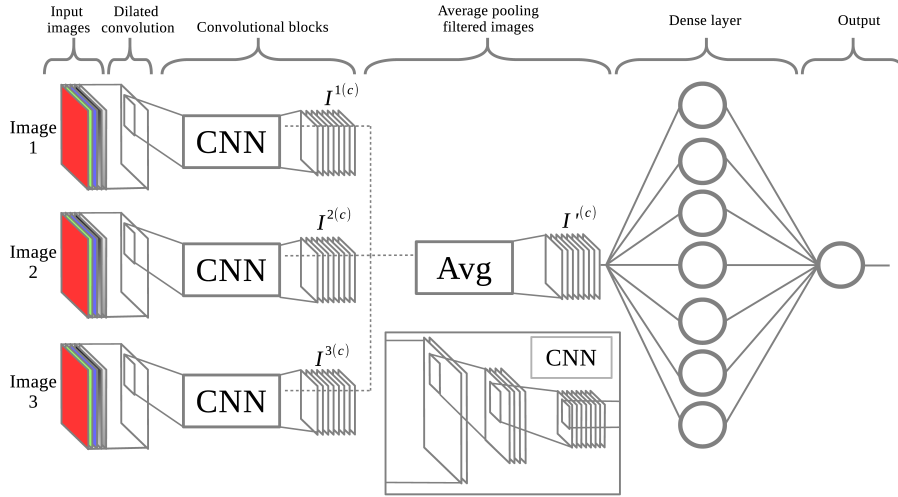
Fig. 3 Representation of the CNN-FA model.

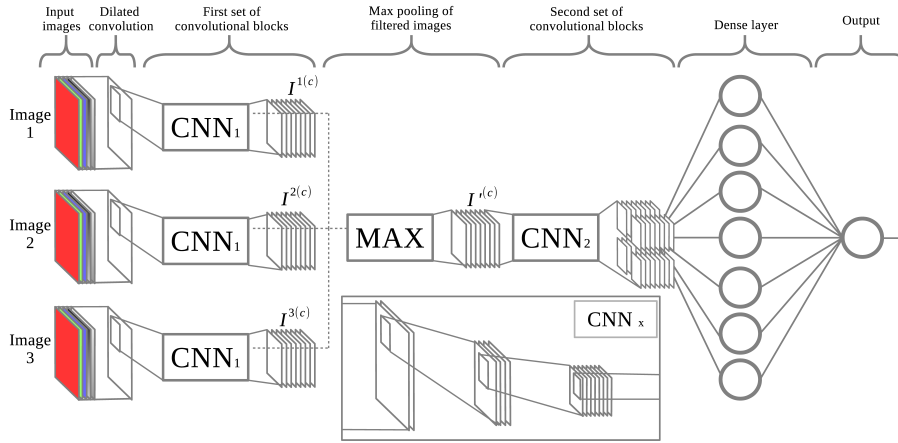**Fig. 4** Representation of the CNN-MvP model

and second set of convolutional blocks (CNN_2). After CNN_2, results are also flattened and fed to the dense layers.

$$I_{a,b}^{\prime(c)} = \max(I_{a,b}^{i,(c)}) \tag{3}$$

3 Data description

All the measurements and ASI images used in this study were collected at the Abengoa Solar Platform of Solúcar (6.25 W, 37.44 N) (Fig. 6). The platform

is located in the southwest of the Iberian Peninsula (near Seville) with an average altitude of 40 meters above sea level and includes several commercial plants (Abengoa Solar operation of 183 MW), pilot plants and research laboratories. Dataset include ASI images collected by three cameras and GHI measurements registered at four radiometric stations, equipped with Kipp&Zonen CMP-6 Pyranometer. Raw data were quality-check analyzed according to the procedures defined by (20), related to physically possible and extremely rare limits. Raw data with solar zenith angle > 75 were discarded. The ASI system is composed by a set of three commercial low-cost video surveillance cameras (model Axis M3007-PV). The distances between ASI cameras are 1, 1.84, and 2.4 km. The cameras provide raw (2592 x 1944) pixels JPEG RGB image files. However, the relevant information is contained in circle of approximate 972 pixels of radio.

The images have been transformed to fit a square image of 500x500, using exclusively the relevant sky information. Every pixel is normalized between 0 and 1. Regard to the irradiance measurements, the average of the values registered at four radiometric stations is considered, because it is more accurate than any individual irradiance measurement.

In this work, instead of directly estimating irradiance, the K_t index is used. This measure is the ratio between irradiance and clear sky irradiance, an analytical calculation based on the date, latitude and longitude. The K_t index removes the solar cycle from the irradiance measurement because it measures the amount of irradiance present relative to the maximum possible for that particular time of the day (and location).

Dataset were registered during a set of days corresponding to the period June to October 2015 and images were taken every 1 minute. There is a set of 1186 examples to perform the experimentation. Each image in the dataset has been labeled with its sky type. Table 1 shows the number of images for each sky type and the average of GHI and the GHI-K_t for each group. There are two special sky types: Multi-cloud and Aerosols. The first happens when different cloud types overlap, resulting in average measurements of irradiance with very high variability (which results in a difficult sky type to estimate). To the best of our knowledge, it is the first time that these two sky types are assessed in the context of ASIs-based solar irradiance estimation.

The state of the sky, and clouds in particular, are the greatest contributors to the intermitency of solar irradiance. Sky type has been used to stratify the training, validation, and test partitions, as it ensures a fair representation of the variety of different sky conditions.

4 Methodology

To ensure the quality of the results and reproducibility of the models, the experimental methodology used in this work is presented in the following paragraphs. Also, the two baseline methods used for comparison are described.

Fig. 5 Spatial location of cameras and sensors.

Table 1 Dataset summary for each sky type: number of images and average GHI. Standard deviation shown in brackets.

Cloud Type	Number of images	Average GHI	Average GHI-K_t
Clear sky	14	737.91 (119.07)	0.938 (0.019)
Cirrus	100	783.82 (91.93)	0.933 (0.094)
Cirrostratus	100	545.54 (113.12)	0.689 (0.134)
Cirrocumulus	30	578.22 (151.80)	0.686 (0.196)
Altocumulus	122	600.15 (184.42)	0.686 (0.2)
Altostratus	41	248.94 (44.69)	0.295 (0.043)
Stratus	163	180.86 (95.44)	0.224 (0.115)
Stratocumulus	220	615.92 (213.45)	0.738 (0.23)
Cumulus	100	757.53 (112.9)	0.834 (0.128)
Multicloud	263	516.43 (235.94)	0.577 (0.258)
Aerosol	33	592.46 (105.1)	0.633 (0.106)

4.1 Baseline methods

Two baseline methods have been compared to the convolutional approaches: a Cloud Fraction model and a Random Forest Feature model.

– **Cloud Fraction:** A common approach to irradiance estimation is Cloud Fraction (CF). This algorithm computes irradiance as the product of total clear-sky irradiance and (1-sky coverage). The coverage is an estimation of the proportion of the sky which is covered by clouds. Coverage has been estimated by means of a HYTA-inspired algorithm [30]. This algorithm first calculates the standard deviation of a normalized index of the image, to decide whether it is unimodal (either clear sky or overcast) or mul-

timodal (partially cloudy skies). If the image is detected as unimodal, a simple thresholding of the pixels is performed. On the second case, the thresholding is dynamic, depending on the mode of each pixel.

For multiple images, this estimation is averaged. This approach is commonplace in irradiance estimation from ASI and despite its simplicity, it offers competitive results [25].

– **Random Forest using Heinle features:** Another approach used for comparison consists on extracting features from the images and use them as inputs to a machine learning method whose output is the estimate of irradiance. Here, the features proposed in [27] have been used. Three types of features are used, spectral, textural and coverage. Spectral features calculate direct statistical measurements from RGB channels and contain: average of the red channel; the average, standard deviation and skewness of the blue channel; and the difference between the averages of red, green and blue channels. Textural features are calculated over the GLCM (Gray-Level Co-Occurrence Matrix) computed over the blue channel. The following features were extracted: homogeneity, energy, entropy and contrast. The last feature is the coverage, calculated using the cloudy pixels detected by a HYTA-based algorithm and calculating the percentage of cloudy pixels in the image.

These features are frequently used for cloud classification, and given that clouds are the main intermitence source for irradiance, it is expected that the irradiance can be estimated from this set of features.

When images from a single camera are used, 12 features are used as inputs to the RF regression method [26]. For experiments using images from multiple cameras (3 in our study), the features extracted from each camera are combined into a vector of features (3x12=36) that will be the inputs to the machine learning algorithm.

The RF optimal hyper-parameters have been found empirically, fixing the number of trees to 500 and searching the number of features per tree between 3 and 11. In this case, the same three stratified train, validation and test partitions as for CNN are used and the validation set is used for selected the hyper-parameters of RF.

4.2 Validation

The complete dataset is split into three stratified partitions: training (70%), validation (15%), and test (15%). The partitions are stratified by cloud type as pointed out in Section 3, with the aim to ensure the fair representation of sky conditions. The training set is used to fit the neural network, validation for optimizing hyper-parameters, and the test set for the final evaluation of the models.

4.3 Metrics

Two different error metrics are used to measure performance: MAE and RMSE. Both metrics can be normalized (relative errors) by the average of the target variable as in Eq.4.

$$rE(Y_{pred}, Y_{obs}) = 100 \times \frac{E(Y_{pred}, Y_{obs})}{mean(Y_{obs})} \qquad (4)$$

where E is MAE or RMSE, and Y_{pred}, Y_{obs} are the sets of predictions and target values, respectively.

In addition to the global metrics, errors have also been broken down by sky type and by the different ranges of irradiance magnitudes (low, medium, high). This is meant to analyse the weaknesses and strengths of the developed models with regard to the different sky types and irradiance levels.

4.4 CNN hyper-parameter tuning

Every network needs to be trained by an optimization algorithm, which has to be carefully selected. In this case the Adam optimizer [31] has been found to be the most suitable algorithm for this problem and dataset.

CNNs have a large number of hyper-parameters. In this article, they have been set via empirical trials. The following hyper-parameters and values have been tested: number of convolutional-activation-pooling blocks {3, 4, 5, 6}, number of filters of the initial dilation {32, 64, 128, 256}, number of filters of the first block {32, 64, 128, 256} (filters always double after the first block), the type of pooling inside a block {Average or Max}, size of the pooling window {2x2 or 3x3}, size of the dilation window {3x3, 5x5, 7x7}, and dilation factor {2, 3, 4}. The search for the optimal values has been incremental, fine-tuning manually the hyper-parameters after each execution, instead of exhaustive, in order to reduce computation time.

After this study the following hyper-parameter values have been selected: the initial dilated convolution has a size of 15x15 dilated by 4, with 128 filters. The single-view model has 5 blocks, with a starting number of 32 filters, doubling the number of filters every block. The convolution filter size is 3x3 (with the standard stride of 1x1). The max-pooling window size is 3x3, with a stride of 2x2. The number of neurons in the dense layer is 512 and the dropout layer has a probability of 75% to drop a neuron (which is a fixed hyper-parameter as a prevention tool against overfitting). The CNN-FA model has the same hyper-parameters as the single-camera model, without further tuning. For the CNN-MvP model, it should be reminded that it has two sets of convolutional blocks. The first one (CNN_1) processes the three images in parallel, while the second one (CNN_2) processes the result of merging the outputs of the first set of blocks by means of max-pooling. CNN_1 has 3 blocks, with a starting number of 32 filters, doubling each block. CNN_2 also has 3 blocks, with a

Table 2 Summary of test metrics (MAE, RMSE as W/m2 and rMAE and rRMSE as percentage) for the single-view camera models (CNNs in the last two rows) and baseline methods (CF, RF, and LeNet, displayed in the first three rows).

	MAE	RMSE	rMAE	rRMSE
CF	110.71	153.47	20.03%	27.76%
RF	83.05	116.22	15.02%	21.02%
LeNet CNN	92.46	125.42	16.73%	22.69%
CNN	78.51	106.29	14.20%	19.23%
CNN + Sun distance	**75.92**	**101.01**	**13.73%**	**18.27%**

starting number of 128 filters, doubling every block. The dense and dropout layers are identical for both models (single-view and multi-view).

Early stopping of training has been applied as follows: networks are trained for a maximum number of 1500 epochs, but for each iteration, training, validation, and test MAE are recorded. Afterwards, the iteration number with minimum validation MAE is selected. The learning rate of the Adam algorithm was set to $1 \cdot 10^{-5}$ while the default momentum value is used.

4.4.1 Technical details

The following experimental results have been obtained on the R environment [32] with the MXNET library [33]. An Nvidia GTX 1080 GPU with 10GB of internal memory has been used.

5 Results and discussion

5.1 Single-view results

Results from the experimentation on a single camera view are presented on Table 2 where all relevant metrics, computed on the test set, are shown. An additional experiment has been conducted, comparing with an standard LeNet architecture [19]. This CNN has no dilation operation. The max pooling operations use a window size and stride of 2x2. Only the RGB channels have been used as inputs. Table 2 contains also the result obtained with baseline methods (CF and Random Forest Feature model).

In Table 2 it is observed that all machine learning methods outperform CF. The baseline method based on feature extraction and RF (second row in Table 2) presents better results than CF and the LeNet CNN architecture, but worse than the CNN architecture proposed in Section 2.1. It is also observed that, for all metrics, the inclusion of the sun distance as an additional input channel to the CNN decreases errors (fourth versus fifth row in Table 2). The use of this information reduces the RMSE by 5.28 W/m2, and the MAE by 2.59 W/m2. If CNN + Sun distance is compared to the standard LeNet, the differences are larger, with a difference of 16.54 W/m2 for MAE and 14.41 for RMSE W/m2. In summary, the CNN architecture proposed and the use of RGB channels and sun distance as inputs to the neural network, achieve the best performance for all metrics.

Table 3 rMAE and rRMSE metrics for the single-view camera models by sky type for each method. The baselines (CF, RF, and LeNet, displayed as the first three columns of each metric) and novel methods (CNN and CNN+Sun distance as the fourth and fifth column of each metric) are compared. These values are expressed in %. The best results for each sky type and experiment are marked in bold.

rMAE	CF	RF	LeNet	CNN	CNN + Sun
Clear-Sky	3.28%	8.57%	5.63%	**3.20%**	4.05%
Cirrus	8.77%	7.20%	7.18%	**6.40%**	8.53%
Cirrostratus	**9.66%**	13.28%	14.33%	10.26%	13.19%
Cirrocumulus	13.40%	**6.85%**	8.37%	8.26%	11.94%
Altocumulus	26.87%	16.75%	24.08%	19.76%	**13.84%**
Altostratus	28.99%	**8.42%**	15.59%	14.80%	15.10%
Stratus	36.10%	34.78%	33.83%	27.87%	**27.20%**
Stratocumulus	31.38%	19.55%	23.02%	21.54%	**18.10%**
Cumulus	9.59%	10.18%	10.16%	**7.58%**	9.70%
Multicloud	24.86%	20.04%	20.64%	17.72%	**16.16%**
Aerosol	34.55%	10.52%	15.61%	9.00%	**8.29%**
rRMSE	CF	RF	LeNet	CNN	CNN + Sun
Clear-Sky	**3.62%**	12.79%	6.84%	4.35%	4.15%
Cirrus	12.62%	**9.05%**	9.84%	9.54%	11.86%
Cirrostratus	**12.32%**	16.51%	16.74%	13.01%	16.01%
Cirrocumulus	15.48%	**9.41%**	10.11%	10.18%	17.30%
Altocumulus	33.85%	20.91%	29.95%	24.01%	**18.59%**
Altostratus	29.69%	**10.21%**	20.58%	16.88%	16.66%
Stratus	45.96%	44.37%	44.03%	**33.91%**	35.11%
Stratocumulus	41.11%	26.22%	30.03%	26.64%	**21.90%**
Cumulus	14.35%	12.28%	13.44%	**9.52%**	12.18%
Multicloud	31.12%	29.06%	26.93%	23.16%	**21.64%**
Aerosol	35.99%	15.53%	17.69%	**12.13%**	12.38

Results have also been broken down by sky type and they can be seen in Table 3. In this case, CF and RF have the best performance on cirrostratus, cirrocumulus and altostratus on both metrics, and cirrus and clear-sky for RMSE. The LeNet architecture always underperforms the other models. The CNNs (RGB and RGB + Sun distance) obtain the best results for the remaining sky types, but no systematic conclusions can be extracted in this case. Another highlight is the performance of CNN methods on the hardest to detect sky types (for example stratocumulus or stratus) improving errors by 2%-5%.

The inclusion of sun distance smooths and decreases most errors for many sky types compared to not using it, although the improvements are not systematically better for all sky types. For instance, there are some cases where errors increase in comparison, for example on cirrostratus where it is worse than CF, or cirrocumulus or altostratus where it is worse than RF. However, for the rest of sky types, errors stay the same or decrease enough for the sun distance to outperform the rest of the models.

An alternative breakdown of results by irradiance magnitude is shown on Table 4. This comparison helps to understand why results are better on the overall metrics but not systematically better on all cloud types. It is observed that while on medium irradiance, the CNN + Sun distance performs worse

Table 4 rMAE and rRMSE metrics (in %) for single-view camera models by irradiance range (low:[0, 441), medium:[441, 699) and high:[699, 1079)). Best results are marked in bold.

	rMAE			rRMSE		
	Low	Medium	High	Low	Medium	High
CF	50.69%	20.48%	9.81%	66.72%	26.18%	16.16%
RF	40.47%	**11.41%**	9.34%	56.00%	**15.06%**	13.26%
LeNet	46.06%	14.20%	9.02%	61.29%	18.26%	12.64%
CNN	32.92%	12.67%	9.14%	44.91%	16.90%	12.46%
CNN + Sun distance	**28.27%**	13.70%	**8.83%**	**39.82%**	17.48%	**11.66%**

than RF, the CNN outperforms the baselines at low and high irradiance. The case of the best performance for low irradiance is important, because the remaining methods performance is poor, which indicates that this case is difficult. For low irradiance, the improvement of CNN+Sun distance over the best of the other methods (RF) is 12% for rMAE and 16% for rRMSE. For high and medium irradiance, differences in the error are smaller: the CNN + Sun distance is best for high irradiance while RF is best for medium, but in both cases differences are smaller than about 2%. This matches the observations on Table 3 where the greatest improvements are seen on high error sky types.

As a conclusion of this analysis, the modifications introduced in the CNN help predict high coverage (hardest type to detect) sky types better than the baseline models. There is a trade-off between medium-high irradiance error and low irradiance error with the CNN + Sun distance method, but the differences on medium irradiances are too low to matter in comparison with the increased capability to correctly predict low irradiance skies.

Convolutional networks are significantly better than Random Forests and Cloud Fraction. While LeNet remains competitive to the baselines, it is worse than our adapted CNN proposal, especially when adding the sun distance feature. The sun distance positively influences results, reducing errors on some cloud types. A feasible explanation for this phenomenon is that the sun distance feature includes otherwise unknown information to the image, revealing new relations within the ASI. It is possible that some defects in the camera are corrected by this method such as image sunburns. Also it is possible that sun distance complements direct irradiance estimations inherent to calculating GHI. There may exist other artificial features for solar irradiance that we did not consider throughout this study, however it is important to include only external information to the image, pieces of information that are not contained within the image.

5.2 Multi-view results

Results from the experimentation on a multiple camera views are presented on Table 5. Here the CNN approaches (CNN-FA and CNN-MvP) described in Section 2.2 are compared to the baseline methods. In this case, the CNNs use as inputs RGB channels and sun distance channel, because it is the best approach when a single-view camera is used (see Section 5.1). As in the case

Table 5 Summary of test metrics (MAE, RMSE, rMAE and rRMSE) for the multi-view models (CNN-FA and CNN-MvP proposals in the last two rows) and baseline methods (CF and RF displayed in the first two rows). Previous results for the single-view CNN (with Sun distance) is shown in the third row.

	MAE	RMSE	rMAE	rRMSE
CF	110.71	153.47	20.03%	27.76%
RF	77.53	106.71	14.02%	19.30%
CNN (single-view)	75.92	101.01	13.73%	18.27%
CNN-FA	84.79	114.47	15.34%	20.71%
CNN-MvP	**70.52**	**98.13**	**12.76%**	**17.75%**

of a single camera, CF performs worse than other techniques for all metrics (MAE and RMSE). RF provides better results than the CNN-FA, but the best performance for all metrics is observed for the CNN-MvP. This approach outperforms the CNN-FA model, with an improvement of 14.27 W/m2 in MAE and 16.34 W/m2 in RMSE. Therefore, information from multiple cameras can be useful when an appropriate combination is performed. With respect to RF (second best approach), the CNN-MvP model improves the errors by 7.01 W/m2 for MAE and 8.58 W/m2 for RMSE. It can also be seen that CNN-MvP improves over the previous single-view CNN in all metrics: it reduces MAE by 5.4 W/m2 (a relative 7.11% improvement) and RMSE by 2.88W/m2 (a relative 2.85% improvement).

As with the single-view results, the metrics are broken down by sky type in Table 6. Here, the CF and CNN-FA models are systematically outperformed by the RF and CNN-MvP models, as the CF and CNN-FA never have the best performance on any cloud type. The CF is still the best algorithm to estimate irradiance in clear skies.

The CNN-MvP model is the best architecture for most sky types on rMAE, and more than half on rRMSE, decreasing errors on some of the most critical types. For example, on stratus (from 11.90% to 5.92% on rRMSE) or cirrostratus (from 12.32% to 8.97%). The CNN-MvP model, in addition to being the best model on the global metrics, it manages to outperform the other models on most sky types.

As with the single camera models, a comparison for different ranges of irradiance magnitudes is presented in Table 7. Here the patterns observed on the single camera comparison are shown again, because the improvements on low irradiance measures are large (10% on rMAE and 11% on rRMSE) with respect to the best baseline method (RF). The error of the CNN-MvP model is worse on medium and high irradiance but the differences with RF are very low (less than about 1%).

When comparing results broken down by sky-type for the single-view and the multi-view methods (Tables 3 and 6), some major improvements of the multi-view approach are observed. Some cloud types have a large reduction in error: altostratus (10%) or cirrostratus (5%). In most cases studied, when multiple cameras are included in the estimation, the results for all sky types improve or remain the same, so including multiple cameras is considered mostly beneficial. If the comparison is done by irradiance magnitudes (Tables 4 and

Table 6 rMAE and rRMSE metrics for the multi-view camera models by sky type for each method. The baselines (CF and RF displayed as the first and second columns of each metric) and novel methods (CNN-FA and CNN-MvP as the third and fourth column of each metric) are compared. These values are expressed in %. The best results for each sky type and experiment are marked in bold.

rMAE	CF	RF	CNN-FA	CNN-MvP
Clear-Sky	**3.28%**	5.75%	4.96%	4.96%
Cirrus	8.77%	**6.20%**	7.58%	9.16%
Cirrostratus	9.66%	11.19%	12.27%	**6.87%**
Cirrocumulus	13.40%	**9.32%**	10.54%	11.78%
Altocumulus	26.87%	14.26%	19.90%	**12.21%**
Altostratus	28.99%	8.96%	19.90%	**5.33%**
Stratus	36.10%	30.70%	29.52%	**22.54%**
Stratocumulus	31.38%	**19.38%**	19.96%	19.98%
Cumulus	9.59%	10.26%	8.48%	**7.40%**
Multicloud	24.86%	18.75%	20.68%	**15.04%**
Aerosol	34.55%	10.25%	8.60%	**7.44%**
rRMSE	CF	RF	CNN-FA	CNN-MvP
Clear-Sky	**3.62%**	7.64%	5.24%	5.45%
Cirrus	12.62%	**8.12%**	11.06%	12.12%
Cirrostratus	12.32%	14.33%	14.83%	**8.97%**
Cirrocumulus	15.48%	**10.89%**	13.15%	15.46%
Altocumulus	33.85%	**17.53%**	24.22%	17.72%
Altostratus	29.69%	11.90%	21.80%	**5.92%**
Stratus	45.96%	38.81%	41.00%	**32.86%**
Stratocumulus	41.11%	**24.96%**	26.85%	25.58%
Cumulus	14.35%	12.52%	9.65%	**8.16%**
Multicloud	31.12%	26.05%	25.80%	**19.43%**
Aerosol	35.99%	16.36%	14.34%	**10.80%**

Table 7 rMAE and rRMSE metrics (in %) for the multi-view camera models and baseline methods by irradiance range (low:$[0, 441)$, medium:$[441, 699)$ and high:$[699, 1079)$). Best results marked in bold.

	rMAE			rRMSE		
	Low	Medium	High	Low	Medium	High
CF	50.69%	20.48%	9.81%	66.72%	26.18%	16.16%
RF	37.32%	**10.62%**	**8.76%**	51.67%	**13.90%**	**11.78%**
CNN-FA	37.6%	13.53%	9.30%	52.00%	17.67%	12.24%
CNN-MvP	**27.44%**	11.21%	8.85%	**39.82%**	15.35%	11.96%

7) there are minor improvements on low and medium irradiance while high irradiance remains similar.

Multi-view results improve upon single camera estimations by adding additional information from different locations. By using spatially separated perspectives in the inference of irradiance, the model is able to discern the best estimation out of the provided inputs. We consider a possibility that this process may approximate a 3D representation of the sky state, not available to the single-view models. Knowing altitude and volume of clouds is very relevant to estimation as overcast skies blocks sun irradiance, and the model being able to better infer low irradiance skies points to this happening as very low, dense cloud types greatly interfere with irradiance. However, whether this approximation is happening is not completely certain, and would make an in-

teresting future research line. To improve the quality of individual and pooled representations the intermediate CNN models (CNN_1 and CNN_2 from the CNN-MvP model) can be enhanced with residual connections, layer normalization or even other state-of-the-art pre-trained networks. Another way to improve the pooled representation would be to use more cameras per sample, or in other words: using more images of the same sky state, which could create more complete representation of the sky type, improving results in turn.

6 Conclusions

In this work, different approaches using CNNs have been proposed and studied to estimate irradiance from a set of ASI images taken at Seville station. Images from three cameras (three different points of view of the same sky situation) are available. The CNN models have been compared with two baseline methods: Cloud Fraction and Random Forest algorithm using as inputs a set of features extracted from the images.

The research started by proposing a single-view CNN model, that use raw images from a single camera. The CNN receives as input the RGB channels of the images, as usual, and an extra input channel containing the distance of every pixel to the sun. The inclusion of this new channel has shown positive results and it has been tested that the removal of sun distance negatively impacts metrics. Moreover, the single-view CNN model outperforms the baseline methods with respect to the global rMAE and rRMSE metrics.

Multi-view CNN approaches have also been proposed, where sky images from three different points of view (three different cameras) are combined. Two approaches (Feature Average model and Multi-view Pooling model) have been compared, that basically differ on the way the features that CNN extracts from each of the three cameras are combined and processed. The first approach simply averages those features, while the second one combines them via max-pooling and performs further processing with additional CNN layers. The use of additional processing layers (Multi-view Pooling model) has proved to be the best alternative to combine information from different cameras. Results have also shown that inclusion of multiple views of the sky reduces the global errors compared to the single-view CNN model and the baseline methods.

The Multi-view Pooling CNN model also reduces the errors across most cloud types, unlike the single-view CNN model, where this behavior is not so systematic. In some critical sky types, such as stratus or cirrostratus, the error decreases considerably for the Multi-view Pooling model. Analysing the results broken down by the level of irradiance, it is concluded that CNN models provide quite smaller errors for low irradiance, while for medium and high irradiance, they perform similar to the Random Forest with feature extraction, which is the best performer among the baseline methods.

In summary, the proposed CNN models show an improvement over other traditional techniques, such as Cloud Fraction or feature-based machine learning methods (Random Forest with feature extraction), specially when joint

information from different cameras is used and when that combination is made through additional CNN layers.

7 Acknowledgements

8 Conflict of interests

There are no conflicts of interests to declare.

References

1. I. Renewables *et al.*, "Analysis and forecast to 2023," *International Energy Agency: Paris, France*, 2018.
2. D. Anagnostos, T. Schmidt, S. Cavadias, D. Soudris, J. Poortmans, and F. Catthoor, "A method for detailed, short-term energy yield forecasting of photovoltaic installations," *Renewable Energy*, vol. 130, pp. 122–129, Jan 2019.
3. G. Merei, J. Moshövel, D. Magnor, and D. U. Sauer, "Optimization of self-consumption and techno-economic analysis of PV-battery systems in commercial applications," *Appl. Energy*, vol. 168, pp. 171–178, Apr 2016.
4. A. Agüera-Pérez, J. C. Palomares-Salas, J. J. González de la Rosa, and O. Florencias-Oliveros, "Weather forecasts for microgrid energy management: Review, discussion and recommendations," *Appl. Energy*, vol. 228, pp. 265–278, Oct 2018.
5. A. Woyte, M. Richter, D. Moser, N. Reich, M. Green, S. Mau, and H. G. Beyer, "Analytical monitoring of grid-connected photovoltaic systems," *IEA-PVPS T13-03*, vol. 2014, 2014.
6. M. Sengupta, A. Habte, S. Kurtz, A. Dobos, S. Wilbert, E. Lorenz, T. Stoffel, D. Renné, C. A. Gueymard, D. Myers, S. Wilcox, P. Blanc, and R. Perez, "Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications," *National Renewable Energy Laboratory*, p. 236p., Feb 2015.
7. C. Rigollier, M. Lefèvre, and L. Wald, "The method Heliosat-2 for deriving shortwave solar radiation from satellite images," *Sol. Energy*, vol. 77, pp. 159–169, Jan 2004.
8. P. Blanc, J. Remund, and L. Vallance, "6 - Short-term solar power forecasting based on satellite images," *Renewable Energy Forecasting*, pp. 179–198, Jan 2017.
9. B. Elsinga and W. G. J. H. M. van Sark, "Short-term peer-to-peer solar forecasting in a network of photovoltaic systems," *Appl. Energy*, vol. 206, pp. 1464–1483, Nov 2017.
10. E. Scolari, F. Sossan, and M. Paolone, "Photovoltaic-Model-Based Solar Irradiance Estimators: Performance Comparison and Application to Maximum Power Forecasting," *IEEE Trans. Sustainable Energy*, vol. 9, pp. 35–44, Jun 2017.

11. A. Kazantzidis, P. Tzoumanikas, P. Blanc, P. Massip, S. Wilbert, and L. Ramirez-Santigosa, "5 - Short-term forecasting based on all-sky cameras," *Renewable Energy Forecasting*, pp. 153–178, Jan 2017.

12. T. Schmidt, J. Kalisch, E. Lorenz, and D. Heinemann, "Evaluating the spatio-temporal performance of sky-imager-based solar irradiance analysis and forecasts," *Atmos. Chem. Phys*, vol. 16, no. 5, pp. 3399–3412, 2016.

13. J. Alonso-Montesinos and F. J. Batlles, "The use of a sky camera for solar radiation estimation based on digital image processing," *Energy*, vol. 90, pp. 377–386, Oct 2015.

14. B. Kurtz and J. Kleissl, "Measuring diffuse, direct, and global irradiance using a sky imager," *Solar Energy*, vol. 141, pp. 311–322, 2017.

15. E. Scolari, F. Sossan, M. Haure-Touzé, and M. Paolone, "Local estimation of the global horizontal irradiance using an all-sky camera," *Sol. Energy*, vol. 173, pp. 1225–1235, Oct 2018.

16. O. Senkal and T. Kuleli, "Estimation of solar radiation over turkey using artificial neural network and satellite data," *Applied Energy*, vol. 86, pp. 1222–1228, Jul 2009.

17. Y. Chu, H. T. Pedro, and C. F. Coimbra, "Hybrid intra-hour DNI forecasts with sky image processing enhanced by stochastic learning," *Solar Energy*, vol. 98, pp. 592–603, 2013.

18. E. Scolari, F. Sossan, M. Haure-Touz, and M. Paolone, "Local estimation of the global horizontal irradiance using an all-sky camera," *Solar Energy*, vol. 173, pp. 1225 – 1235, 2018.

19. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

20. T. A. Siddiqui, S. Bharadwaj, and S. Kalyanaraman, "A deep learning approach to solar-irradiance forecasting in sky-videos," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 2166–2174, IEEE, 2019.

21. X. Zhao, H. Wei, H. Wang, T. Zhu, and K. Zhang, "3d-cnn-based feature extraction of ground-based cloud images for direct normal irradiance prediction," *Solar Energy*, vol. 181, pp. 510 – 518, 2019.

22. J. Alonso, F. Batlles, C. Villarroel, R. Ayala, and J. Burgaleta, "Determination of the sun area in sky camera images using radiometric data," *Energy conversion and management*, vol. 78, pp. 24–31, 2014.

23. J. Alonso-Montesinos and F. J. Batlles, "The use of a sky camera for solar radiation estimation based on digital image processing," *Energy*, vol. 90, no. P1, pp. 377–386, 2015.

24. H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3d shape recognition," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 945–953, IEEE, Dec 2015.

25. L. Chen, G. Yan, T. Wang, H. Ren, J. Calb, J. Zhao, and R. McKenzie, "Estimation of surface shortwave radiation components under all sky conditions: Modeling and sensitivity analysis," *Remote Sensing of Environment*, vol. 123, pp. 457–469, Aug 2012.

26. L. Breiman, "Random Forests," *Machine Learning*, vol. 45, pp. 5–32, Oct. 2001.

27. A. Heinle, A. Macke, and A. Srivastav, "Automatic cloud classification of whole sky images," *Atmospheric Measurement Techniques*, vol. 3, no. 3, pp. 557–567, 2010.

28. I. Reda and A. Andreas, "Solar position algorithm for solar radiation application," *Sol. Energy*, vol. 76, pp. 577–589, May 2004.

29. F. Yu and V. Koltun, "Multi-Scale Context Aggregation by Dilated Convolutions," *arXiv*, Nov 2015.

30. Q. Li, W. Lu, and J. Yang, "A hybrid thresholding algorithm for cloud detection on ground-based color images," *Journal of Atmospheric and Oceanic Technology*, vol. 28, pp. 1286–1296, Oct 2011.

31. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, Dec 2014. arXiv: 1412.6980.

32. R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

33. T. Chen, M. Li, Y. Li, M. Lin, N. Wang, M. Wang, T. Xiao, B. Xu, C. Zhang, and Z. Zhang, "Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems," *arXiv preprint arXiv:1512.01274*, 2015.