SPIN: Simplifying Polar Invariance for Neural networks Application to vision-based irradiance forecasting

Quentin Paletta[1,2] Anthony Hu[1] Guillaume Arbod[2] Philippe Blanc[3] Joan Lasenby[1]

[1]University of Cambridge, UK. [2]ENGIE Lab CRIGEN, France. [3]MINES ParisTech, France.

Abstract

Translational invariance induced by pooling operations is an inherent property of convolutional neural networks, which facilitates numerous computer vision tasks such as classification. Yet to leverage rotational invariant tasks, convolutional architectures require specific rotational invariant layers or extensive data augmentation to learn from diverse rotated versions of a given spatial configuration. Unwrapping the image into its polar coordinates provides a more explicit representation to train a convolutional architecture as the rotational invariance becomes translational, hence the visually distinct but otherwise equivalent rotated versions of a given scene can be learnt from a single image. We show with two common vision-based solar irradiance forecasting challenges (i.e. using ground-taken sky images or satellite images), that this preprocessing step significantly improves prediction results by standardising the scene representation, while decreasing training time by a factor of 4 compared to augmenting data with rotations. In addition, this transformation magnifies the area surrounding the centre of the rotation, leading to more accurate short-term irradiance predictions.

1. Introduction

The presence of clouds in the sky induces a short-term spatiotemporal variability in the production of solar energy. Their complex spatial distribution and temporal dynamics add to the difficulty of accurately predicting cloud displacements, and thus precisely forecasting solar energy yield, e.g. from photovoltaic or concentrated solar thermal electricity converting systems. Consequently, accounting for the high spatiotemporal solar irradiance variability caused by clouds is crucial for the integration of solar energy in the electric grid, or for off-grid hybrid systems, which use a fuel backup generator when the solar output is insufficient.

This short-term variability can be predicted from videos of the cloud cover taken by hemispherical sky cameras on the ground [4, 31] or geostationary satellites [9]. The former provides local predictions with a high temporal resolution up to 20-min ahead, whereas the later gives hours-ahead predictions with a spatial resolution of up to 2-3 km^2 (Figure 2). These vision-based approaches are better suited for prediction than time series from photovoltaic production monitoring or upstream pyranometric in-situ measurement (pyranometer: device used to measure the local solar irradiance), which have no intrinsic ability to predict sun covering by incoming clouds.

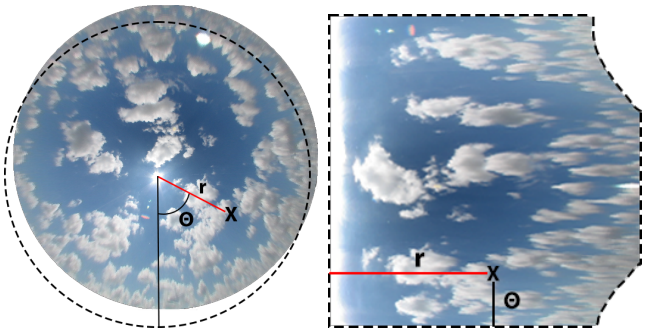


Figure 1. Undistorted image of the sky taken by a fish-eye camera and its polar transformation (raw data are shared by SIRTA laboratory [8]).

Invariances are key in computer vision as they can be integrated in the modelling as inductive biases to facilitate a learning process. For instance, in addition to being translationally equivariant due to weight sharing (a translated input results in similarly translated feature maps), a convolutional neural network using pooling layers is also translationally invariant, i.e. its prediction is unaffected by input translations. Disregarding the position of the object in the image significantly simplifies classification tasks, which share similar translational invariance properties, by avoiding learning to recognise an object for every possible position in the image.

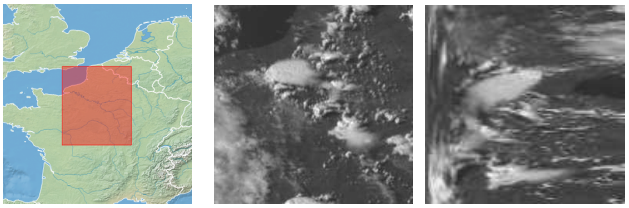


Figure 2. Greyscale satellite image centred on SIRTA's Laboratory (48.713° N, 2.208° E) [5] and corresponding representation in polar coordinates.

Definition A task $\mathcal{T}$ is invariant with respect to an equivalence relation $\sim$ [1] on $\mathcal{X}$ if:

$$\forall x, x' \in \mathcal{X}, x \sim x' \Rightarrow \mathcal{T}(x) = \mathcal{T}(x').$$

However, in some computer vision tasks such as solar energy forecasting, the position of objects is key. A principal component analysis on the learnt spatiotemporal representation of the sequence of past images has shown that the horizontal and vertical positions of the sun account for the 2nd and 4th principal components respectively [26] (see Figure G.1). The position of the sun in the sky and notably its zenith angle directly indicates the level of extra terrestrial irradiance and the amount of air mass crossed by the solar radiation. Hence, it is key to estimate the clear-sky downwelling irradiance (the component of radiation directed toward the earth's surface).

In addition, since the direct component accounts for most of the solar radiation, we expect the zenith angular distance from the centre of the sun to clouds to be of crucial importance to anticipate critical events. The azimuthal position of the clouds with respect the sun's position is, on the contrary, of much less importance. For these reasons, traditional techniques take the position of the sun in the image as a reference to read the spatial configuration and temporal dynamics of the cloud cover [19, 33]. To some extent, the problem of irradiance forecasting from sky images can therefore be considered as rotationally invariant around the sun (polar invariance): the angle from which a cloud is approaching the sun has little impact on the resulting irradiance change. For similar reasons, local irradiance forecasting from satellite images can be approximately seen as invariant to rotation around a point of interest on the globe (a solar facility for instance).

Contributions We introduce the SPIN method (Simplifying Polar Invariance for Neural Networks) to show the benefit for convolutional architectures of representing a rotationally invariant scene with polar coordinates in the context of video analysis. We propose the translation and vertical flip as data augmentation techniques for this polar representation but also temporal flip to leverage the temporal invariance of the problem. We compare it with other transformations and data augmentation approaches never before tried in deep irradiance forecasting from all-sky cameras: sun-centred images, a close-up on the circumsolar area and rotations to leverage the rotational invariance in the image by learning from more diverse cloud-sun spatial configurations. Two irradiance forecasting challenges using ground-taken sky images or satellite images are shown to greatly benefit from representing their corresponding scene with polar coordinates centred on the point of interest in the image, i.e. the sun in the sky image or a solar farm in the satellite image (see Figure 2).

[1] A binary relation $\sim$ on a set $\mathcal{X}$ is said to be an equivalence relation, if and only if it is reflexive ($x \sim x$), symmetric ($x \sim x' \Leftrightarrow x' \sim x$) and transitive ($x \sim x'$ and $x' \sim x'' \Rightarrow x \sim x''$).

2. Prior work

Vision-based solar irradiance forecasting As well as traditional computer vision techniques to model cloud displacement based on cloud detection and tracking in a sequences of sky images [11, 22, 30] or satellite observations [20], recent studies have tried to learn the cloud cover dynamics from training end-to-end various types of neural networks from past sequences of hemispherical sky images [13, 16, 37, 38]. Benefiting from recent advances in deep learning (DL) and an increased availability of large sky images datasets [8, 14, 23, 29] and satellite data, past works have reported high forecasting scores [6, 25, 37]. The common setup is based on supervised learning: a convolutional architecture is trained to extract features from a sequence of past images (sky images or satellite images) and colocated/concomitant time series of pyranometric measurements, to predict future frames and/or future solar irradiance levels [26, 32]. Based on the assumption that the neural network will be able to extract relevant features from raw images, input images are rarely preprocessed before being given to the model. In some studies, images taken by fish-eye cameras are unwrapped on a rectangular regular grid assuming for example, median cloud heights to limit the impact of the distortion on the scene representation. This provides cloud shape and trajectory consistency across the sequence of frames [26]. Recently, some data augmentation techniques (e.g. noise injection, colour space transformations and mixing of images) have been applied to this forecasting task but no significant improvement was observed [21].

Polar invariance Although a rotational invariance around the point of interest is critical, it is not explicitly given to the neural network which has to comprehend a similar event (a cloud hiding the sun or covering the solar facility) multiple times for different positions of the cloud around the sun (or around the power plant in satellite images). A common but

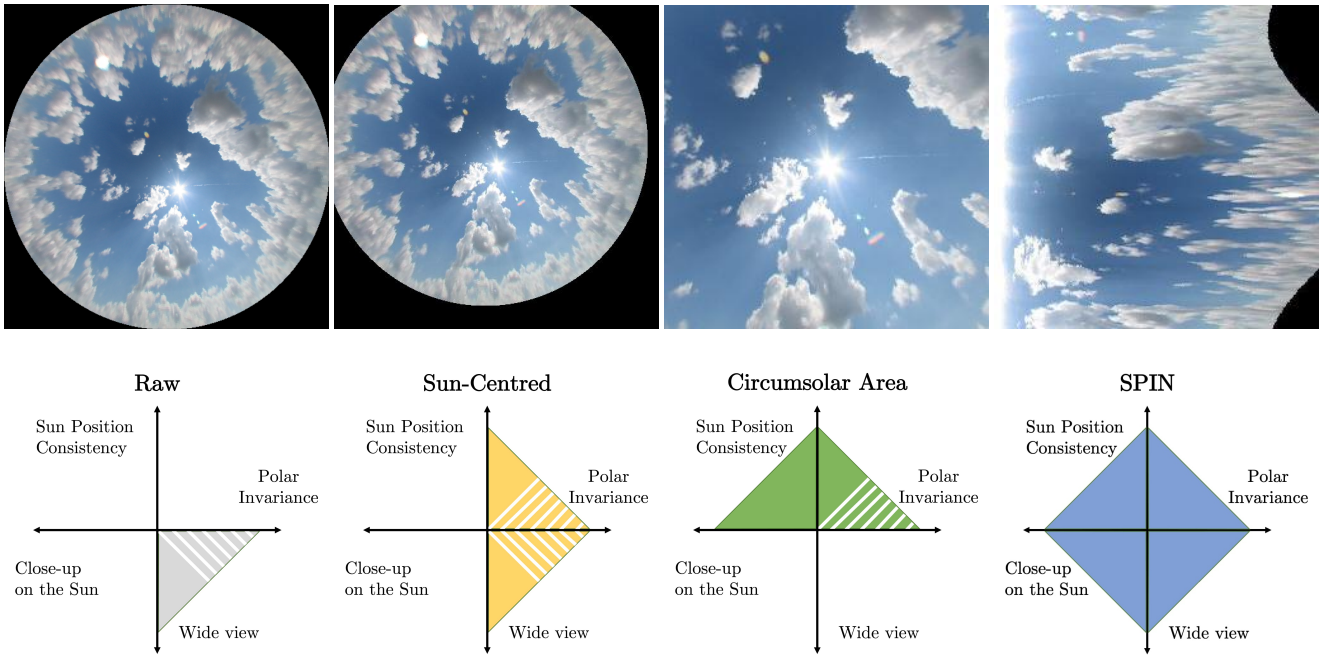


Figure 3. From left to right: 1. Raw undistorted image of the sky taken with a fish-eye camera, 2. Image centred on the sun's position, 3. Close up on the circumsolar area and 4. Polar coordinates. On top the different transformed images and below the corresponding properties diagrams (the white dashed area highlights the gain of leveraging the polar invariance with rotations) taking into account: 1.The polar invariance of the problem, 2. The central position of the sun in the scene, 3. The information close to the sun and 4. A wide view on the more distant part of the image.

expensive way to learn this invariance would be to augment data by training the model on rotated versions of images around the sun (or the facility in the satellite image). Alternatively, one can turn this rotational invariance problem into a one of translational invariance by representing the scene with polar coordinates (see Figure 1) [2]. With this new configuration, the problem becomes invariant by translation along one axis (the argument coordinate), hence is more suited to analysis by convolutional layers. More than a data augmentation technique [34], this transformation applied to rotationally invariant scenes is a more efficient way to learn a task invariance than data augmentation through rotations (which was not shown in common applications [12]) and does not require model tweaks [15]. Additionally, this introduces a distortion in the scene representation resulting in an increased size of the region of interest (the circumsolar area in the sky image or the area surrounding the facility in the satellite image) relative to the rest of the image.

3. Methodology

3.1. Datasets

Irradiance data Irradiance data were generated over a 3 year time window from 2017 to 2019 at SIRTA's lab [8]. It was measured by a pyranometer and reported as a per minute average. Solar irradiance can be predicted from both sky images or satellite images with computer vision.

Sky Images Sky images were generated over the same time window at SIRTA's lab. Images were taken by a hemispherical sky camera (EKO SRF-02) with a 2-min temporal resolution. To tackle the strong distortion induced by the fish-eye lens, images were undistorted [28]. This provides cloud shape and trajectory consistency in consecutive frames as explained in the introduction. The segmentation of images into 5 classes (sky, clouds, sun, saturation and frame) following the method presented in [26] can be briefly summarised as follows: the image is first binary classified into cloud and sky pixels using the Hybrid Thresholding Algorithm (HYTA) [10, 18], then the sun tracker [28] is used to segment the sun if visible. Remaining saturated pixels are classified as saturation. Finally, RGB and segmented images are cropped and downscaled from 768×1024 to a 128×128 pixel resolution.

Satellite Images The set of greyscale satellite images used in this study was obtained from EUMETSAT [24] (Meteosat SEVIRI Rapid Scan image data [2]). These 256×256 pixel images cover a $4.4°$(latitude) $\times$ $4.4°$ (longitude) area centred on SIRTA's laboratory (48.713° N, 2.208° E) (Figure 2). The dataset comprises 148 days in 2017 (training set) and 198 days in 2018 (validation and test sets) with a 5-min temporal resolution.

[2]https://navigator.eumetsat.int/product/EO:EUM:DAT:MSG:MSG15-RSS

3.2. Image transformation and data augmentation

Polar Transformation (SPIN) The transformation of RGB and segmented images into their polar representations is depicted in Figure 1. The centre of the rotation in raw images are respectively the centre of the image in satellite images and the centre of the sun in sky images, which is obtained from a tracking method [28]. The radius r ranges from 0 to half of the original image's width W in both cases and the polar angle θ from 0 to 360°. In practice the original $W \times H \times 3$ image is first mapped to a $W/2 \times r \times 3$ image then downscaled to $128 \times 128 \times 3$ (dimension of the input images in this study). Given the typical "U" shape trajectory of the sun in the sky image (see [28]), the bottom half of the frame is more often visible in the polar transformation than the top half. To keep the resulting dark area in the corners of the resulting transformed image, the direction of the reference angle ($\theta = 0$) is set to the downward vertical.

This transformation introduces several modifications in the representation of the scene, which might impact the performance of the learning (see Figure 3). First, the critical role of the sun position regardless of its visibility (e.g. hidden by a cloud) relative to the rest of the sky is taken into account to standardise the scene. Second, the rotational invariance of the problem around the sun is leveraged by turning it into a translational invariance problem. With this representation, the rotated versions of a given sun-clouds (or solar farm-clouds) spatial configuration can be learnt by a convolutional neural network (CNN) without the need to augment the data. Finally, the transformation distorts the representation of the scene leading to a magnifying effect on the circumsolar area (CSA; area directly surrounding the sun or the solar plant). To disentangle the impact of each aspect of the transformation, we compare the SPIN method with the following intermediate transformations:

Image centred on the Sun (Sky images) To standardise the scene with respect to the position of the sun, we centre sky images on its position using the sun tracking algorithm (second panel in Figure 3).

Close-up on centre of the rotation Building upon the sun-centred image transformation, we disentangle the magnifying effect on the CSA by training models on a close-up of the original image. The new representation depicted in the third panel of Figure 3 corresponds to a fourth of the original image size.

Data augmentation Finally, to leverage the rotational invariance of the problem, we perform random rotations of the input images to generate visually diverse samples from a single spatial configuration. This preprocessing step is applied to sun-centred images, the close-up on the CSA as well as on undistorted images to evaluate the effect of the resulting data augmentation. Although augmenting a polar coordinates image with vertical translations would generate redundant samples for a CNN, it provides some diversity by revealing the clouds split into two at the top and bottom edges of the image.

To compare all transformations keeping the model unchanged, each transformed image is eventually re-scaled to a $128 \times 128 \times 3$ pixel resolution.

3.3. Metrics

Forecast Skill Comparing the performance of forecasting methods evaluated in different conditions is problematic due to confounding effects of multiple variables: weather conditions, localisation, time of the year or of the day, etc. To partially overcome these limiting factors, a standard evaluation method in solar energy forecasting is to estimate the relative performance of a model with a reference mode [36]. The most common baseline is an adaptation of the *persistence model* (PM) called the *smart persistence model* (SPM). Whereas the PM predicts no irradiance changes over a given forecast window ΔT (Equation 1), the SPM takes into account diurnal changes of the extra-terrestrial irradiance given by a clear-sky model [3] (Equation 2).

$$\hat{I}(t + \Delta T) = I(t) \tag{1}$$

$$\hat{I}(t+\Delta T) = k_c(t)\, I_{clr}(t+\Delta T) \text{ with } k_c(t) = \frac{I(t)}{I_{clr}(t)} \tag{2}$$

I represents the measured irradiance and $\hat{I}$ is the predicted irradiance. $I_{clr}(t)$ is the corresponding clear-sky irradiance (without considering cloud) only accounting for the solar zenith angle and the optical transparency state of the atmosphere (or turbidiity). This is a modelled irradiance that is totally exogeneous from this study (McClear model [17]). Given the errors of a model Error_{model} and of the baseline Error_{SPM} based on a given metric (RMSE, MSE, MAE, etc.), the *forecast skill* (FS) is computed as follow:

$$\text{Forecast Skill} = 1 - \frac{\text{Error}_{model}}{\text{Error}_{SPM}} \tag{3}$$

Although better able to generalise a model performance, the FS still has limitations [35]. Particularly with DL approaches, architectures tend to face a frequent time lag correlated with low anticipation skills [25]. This greatly decreases their applicability to real world applications.

Temporal Distortion Index

To quantify this temporal misalignment, [7] introduces the *temporal distortion index* (TDI). Based on dynamic time warping, this metric quantifies temporal distortion by locally warping a times series (e.g. predictions of the model) to resemble another (e.g. ground truth). The resulting TDI

Forecast Horizon	RMSE ↓ [W/m^2] (Forecast Skill ↑ [%]) 2-min	6-min	10-min	95% Quantile ↓ [W/m^2] 2-min	6-min	10-min
Smart Pers.	93.3 (0%)	129.7 (0%)	146.2 (0%)	201.1	304.8	356.9
Raw	85.4 (8.4%)	98.3 (24.2%)	110.4 (24.5%)	184.9	218.2	244.8
Raw + Rotations	78.6 (15.8%)	93.6 (27.8%)	105.5 (27.8%)	168.6	205.9	238.6
CSA	80.4 (13.8%)	97.0 (25.5%)	109.8 (25.1%)	169.9	217.4	244.1
CSA + Rotations	77.1 (17.3%)	93.5 (27.9%)	107.3 (26.6%)	161.2	205.9	240.0
Sun-Centred	81.0 (13.2%)	94.1 (27.4%)	106.6 (27.1%)	174.4	206.8	235.1
Sun-Centred + Rotations	74.9 (19.8%)	89.2 (31.2%)	102.7 (29.8%)	157.8	196.1	227.4
SPIN	74.3 (20.4%)	89.3 (31.1%)	102.9 (29.6%)	156.6	196.6	230.3
SPIN + Translations	**71.8 (23.1%)**	**87.2 (32.8%)**	**101.0 (30.9%)**	**149.1**	**192.2**	**224.0**

Table 1. Performance comparison of the different image transformations based on ECLIPSE 2, 6 and 10-min ahead irradiance predictions from sky images.

is defined as a percentage relative to the maximal distortion. It indicates if the forecasting is, on average, late or in advance, and to what extent.

95% Quantile Anticipating sudden events corresponding to large irradiance shifts is critical to mitigate the effect of the solar production variability. In a hybrid power plant configuration, for instance solar panels with a diesel generator, the fossil fuel backup has to warm up for a few minutes before being able to cover a low solar yield and get synchronized to the local frequency of the isolated grid. To assess the performance of the model on these extreme events, we report the 95% quantile of the sorted list of absolute errors.

3.4. Models

The different transformations presented in this study are evaluated based on the performance of three DL irradiance forecasting models: a CNN [27], a convolutional long short-term memory network (ConvLSTM) [25] network and ECLIPSE [26]. The CNN and ConvLSTM designs are similar: a spatio-temporal representation of the input sequence of past images is learnt through a set of convolutional layers, pooling layers and recurrent modules (for the ConvLSTM). It is then decoded into a single future irradiance value. Instead of predicting a single irradiance value, ECLIPSE is trained end-to-end to recursively predict the sequence of future segmented images and corresponding irradiance levels (Table B.3). This state-of-the-art design was shown to reduce forecast time lag, while improving quantitative performances [26].

3.5. Model Training

Samples with low solar elevation (less than 10° above the horizon) are not used in this study. The remaining samples are split in distinct sets as follows.

Sky Images The model is trained on data from 2017 and 2018 (180,000 samples) and evaluated on even days from 2019 (20,000). The remaining odd days from 2019 are kept for testing (20,000). In addition, the TDM metric is tested on 200 windows randomly sampled from test days. Each window comprises 100 consecutive samples and is 3h20m long.

Satellite Images The model is trained on data from 140 days in 2017 (16,000 samples) and validated on even days from 2018 (5,000). The remaining odd days from 2018 are retained for testing (5,000). In addition, the TDM metric is tested on 100 windows randomly sampled from test days. Each window comprises 50 consecutive samples and is 4h10m long.

4. Results

4.1. Quantitative Performances

The impact of each transformation on model predictions is evaluated through the performance of three models (CNN, ConvLSTM and ECLIPSE) on three forecast horizons (2, 6 and 10-min ahead). The scores of each configuration (Transformation × model × forecast horizon) are reported as their average over two trainings with different random initialisations (Tables E.1 and E.1). The results of ECLIPSE presented in Table 1 are analysed in greater detail as it is the best performing model on most indicators, but all three models display similar trends regarding image transformation and data augmentation.

Although centring the image on the sun provides a considerable performance increase on all forecast horizons compared to the raw image (Figure 4), the polar transformation provides the largest gain. Interestingly, the impact of each transformation on the model predictions depends on the horizon of the prediction. As the forecast horizon

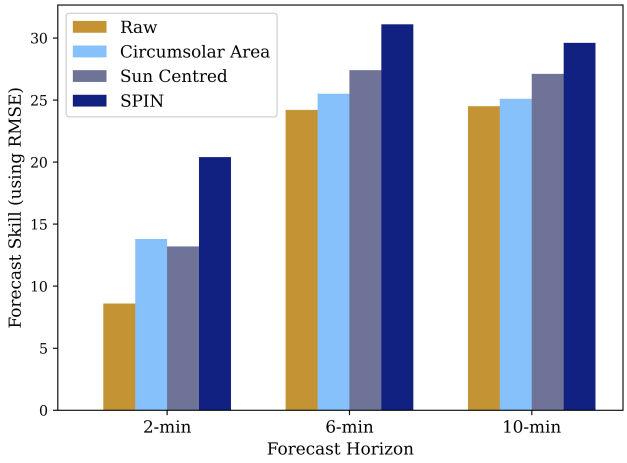


Figure 4. 2, 6 and 10-min ahead FS of ECLIPSE predictions for different sky image transformations.

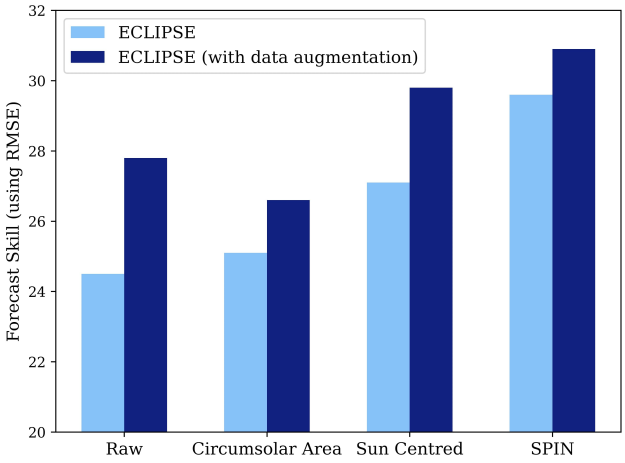


Figure 5. ECLIPSE 10-min ahead FS with and without data augmentation (rotation or translation) applied to the different transformations of sky images.

increases, the range of FS narrows: a difference of 12% at 2-min (from 8.4 to 20.4%), but only 5.1% at 10-min (from 24.5 to 29.6%). Ultimately, we expect all transformations to perform similarly on longer time horizons as current observations provide less information on the longer-term states of the cloud cover. This point is highlighted by the performance of ECLIPSE on the CSA transformation relative to the sun-centred image (Figure 4). Whereas the CSA transformation leads to a better performance on the 2-min forecast (13.8 against 13.2%), it is outperformed on longer time horizons by the sun-centred image (25.1 against 27.1%), which provides more information on distant clouds. Similarly, the distortion observed in the SPIN scene representation seems to account for some of the FS gain on the 2-min ahead predictions: clouds surrounding the sun are more visible given the magnification of the CSA in polar coordinates.

Interestingly, data augmentation (rotation or vertical translation) consistently benefits predictions with an increased FS (Figure 5) and a smaller temporal distortion (Figure 6). Considering that the different rotated versions of a given sun-clouds configuration are learnt by a CNN from a single sample in polar coordinates, the benefit of translating the SPIN representation to reveal clouds split at the edge of the image is smaller than that of rotating the other transformations (Raw, Circumsolar and Sun-centred sky images) to learn from the resulting set of altered spatial configurations. This results in up to four times faster training for SPIN compared to the sun-centred representation with rotations for similar performances (Figure 7).

4.2. Prediction Curves

We compare the different transformations on the 2 and 10-min ahead prediction curves. Regarding the longer horizons, all transformations break the *persistence barrier* (i.e.

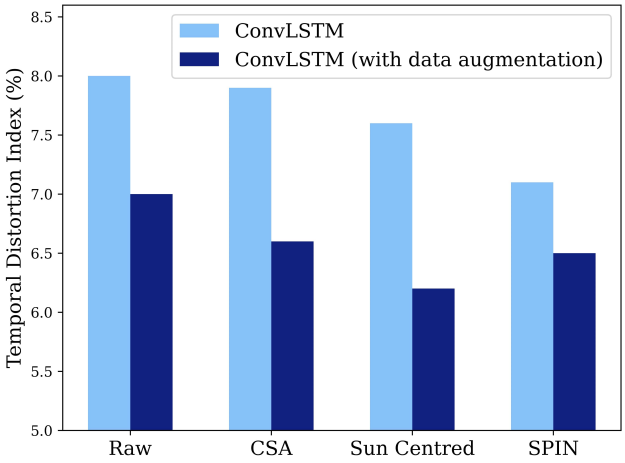


Figure 6. Time distortion index for the different sky image transformations averaged over the 2, 6 and 10-min ahead ConvLSTM's predictions.

are able to foresee events before they happen and decrease time lag below the forecast horizon) except for the CSA (Figure F.2). This is understandable as a close-up on the sun provides little information on distant clouds which are more likely to impact longer-term temporal variability.

In addition, transformations magnifying the circumsolar area (CSA and SPIN) appear to induce a more accurate concomitant irradiance estimate, benefiting shorter term forecasting (2-min). In Figure F.1 for instance, the CSA and SPIN predictions remain close to the ground truth, whereas the raw image and *sun-centred* transformation lead to a high bias (> 100 W/m^2) over large time windows.

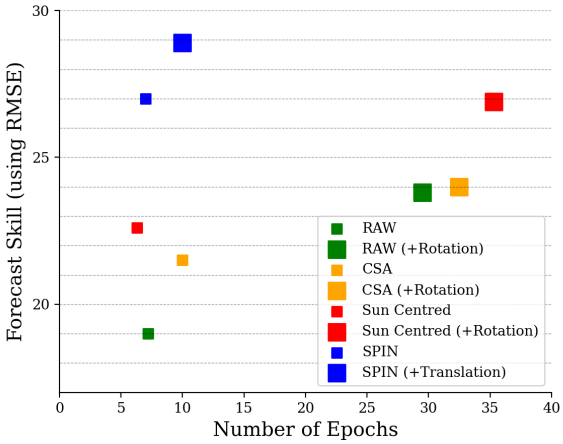


Figure 7. Impact of data augmentation on the training time for the different transformations averaged over the 2, 6 and 10-min ahead ECLIPSE predictions. Data augmentation has limited effect on SPIN's training time contrary to other transformations.

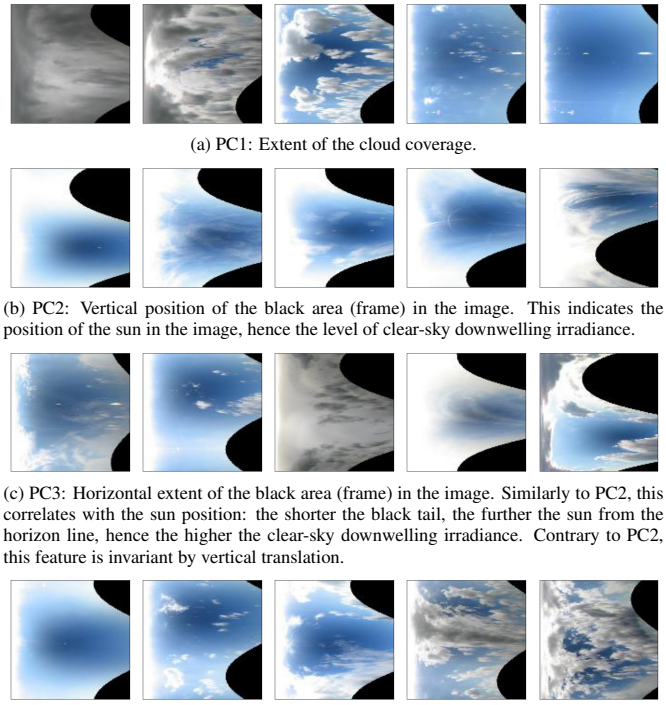
4.3. Visualisation of the Encoded Spatiotemporal Representation

We perform a principal component analysis (PCA) on the spatiotemporal representation of the past sequence of sky images (see Figure 8). The first principal components are similar to those of raw sky images identified by [26] (see Figure G.1). In particular, the position of the sun seems to be extracted from the size and position of the black area (frame) in SPIN images. This indicates that the critical information of the sun position can be retrieved from polar representations, while it would be partly lost in rotated sky images.

4.4. Satellite Images

DL applied to satellite imagery is key for longer-term irradiance forecasting up to 1-2 hours, because of its rapid data processing at inference time compared to heavy numerical weather models, which need more time to reach steady-state simulation [1]. In this study, ECLIPSE is trained on the 10 to 50-min ahead forecasts with a 10-min temporal resolution from greyscale satellite images, the corresponding cloud index map and irradiance measurement (included as a third input channel). Table C.3 shows that models perform significantly better when based on the past irradiance levels, which can be done by inputting past measurements or by predicting the upcoming variation in irradiance levels instead of directly predicting absolute values.

Table 2 shows that both image transformation and data augmentation benefit irradiance forecasting in this new setting. The SPIN transformation leads to FS gains on predictions from both raw and close-up images on all horizons. Contrary to previous results on all-sky cameras, data augmentation on satellite images benefits all settings but the 10-min ahead predictions from polar coordinate representations (SPIN and SPIN close-up). This could be explained by the low level of cloud motion for shorter-term horizons, which could be learned without data augmentation.



(a) PC1: Extent of the cloud coverage.

(b) PC2: Vertical position of the black area (frame) in the image. This indicates the position of the sun in the image, hence the level of clear-sky downwelling irradiance.

(c) PC3: Horizontal extent of the black area (frame) in the image. Similarly to PC2, this correlates with the sun position: the shorter the black tail, the further the sun from the horizon line, hence the higher the clear-sky downwelling irradiance. Contrary to PC2, this feature is invariant by vertical translation.

(d) PC4: Spatial variability of the cloud cover: from fully sunny (or fully cloudy) to partly cloudy.

Figure 8. Four principal components of the spatio-temporal representation encoded by the model (with SPIN method). They respectively account for 8.6%, 6.2%, 4.4% and 2.8% of the variance. The variability of each component is illustrated with images draw from the distribution, from low to high values

	Forecast Skill ↑ [%]		
Forecast Horizon	10-min	30-min	50-min
Raw	8.5	6.5	6.3
Raw (+R)	12.1	12.2	13.7
SPIN	**12.2**	12.5	14.1
SPIN (+T)	12.1	**13.9**	**17.1**
Close-up	10.3	10.6	11.8
Close-up (+R)	10.9	11.1	**18.0**
SPIN Close-up	**13.8**	14.3	14.8
SPIN Close-up (+T)	13.6	**14.4**	**18.0**

Table 2. Performance benchmark of the different satellite image transformations based on ECLIPSE 10, 30 and 50-min ahead irradiance predictions averaged over a 5-min period (R : rotations, T : translations)

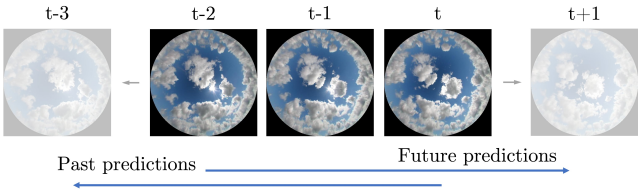


Figure 9. Temporal invariant task: predicting the past from a given sequence is as challenging as predicting the future. Training sequences are flipped with a probability of 0.5 to augment the data.

	Forecast Skill ↑ [%]		
Forecast Horizon	2-min	6-min	10-min
SPIN	21.4	32.2	30.2
SPIN (+TF)	**23.2**	**32.9**	**30.6**
SPIN (+T)	22.2	32.5	30.5
SPIN (+T&TF)	**26.0**	**33.7**	**31.7**

Table 3. Performance benchmark of the different sky image transformations based on ECLIPSE 2, 6 and 10-min ahead irradiance predictions averaged over a 5-min period (T : translations, TF : Temporal flipping). Due to missing data, samples with missing past instances were discarded, which is why the scores of the SPIN baselines are different from other experiments.

4.5. Temporal invariance

Another key invariance in vision-based irradiance forecasting is the temporal invariance: from a given sequence of images, predicting past measurements is as challenging as predicting future measurements (Figure 9). To test that hypothesis, we augment the training data by reversing or not reversing input sequences with a probability of 50% (both directions are seen by the model across multiple epochs). Despite a 20 to 40% increase in training time, results show a benefit on all forecast horizons of using temporal flipping as a data augmentation technique. Notably, the shorter the horizon the higher the FS gain: +1.8% at 2-min, +0.7% at 6-min and +0.4% at 10-min (for SPIN with and without temporal invariance).

4.6. Vertical flip

We experimented vertical flipping as another type of data augmentation (training images were randomly flipped vertically with a probability of 50%) combined with rotation and translation. The results presented in Table 4 show that, although rotations and translations provide a more consistent gain, flipping images improves the baseline by about 1% FS on all horizons. Combining vertical flipping with translations provides mixed results with no FS gain on the 6 and 10-min ahead horizons. Similarly to the temporal flip, the shorter the horizon, the higher the FS increase.

5. Discussion

This study compares different image transformations in the context of rotationally invariant tasks. The characteristics of the different representations and the use of data augmentation leads to variable performance gains regarding forecast skill, clear-sky irradiance level estimation, temporal misalignment or training time. Regarding these aspects of irradiance forecasting, further investigations would focus on the impact of transformations' parameters such as the scaling factor for the CSA transform or the maximum radius for SPIN. In addition, we expect that combining the different representations as multiple input channels would further improve the predictions.

	Forecast Skill ↑ [%]		
Forecast Horizon	2-min	6-min	10-min
SPIN	20.4	31.1	29.6
SPIN (+VF)	**23.8**	**32.3**	**30.4**
SPIN (+T)	23.1	**32.8**	**30.9**
SPIN (+T&VF)	**24.2**	31.9	30.8

Table 4. Performance benchmark of the different sky image data augmentation techniques (T : translations, VF : Vertical flip) based on ECLIPSE 2, 6 and 10-min ahead irradiance predictions

Furthermore, this learning strategy is likely to advance the generalisation properties of a model to observations from other locations. In a transfer learning context, cameras corresponding to a different dataset might not be oriented the same way, hence some spatial features such as the impact of the sun position on the irradiance level would be altered. However, the standardisation of the sun-clouds spatial configuration under the SPIN transformation could facilitate knowledge sharing.

6. Conclusion

By representing a scene with polar coordinates, the rotational invariance of a problem becomes translational. Thus, the visually diverse but equivalent spatial configurations obtained by rotating a given image are learnt by convolutional architectures from a single representation in polar coordinates. Applying this preprocessing step to two polar invariant computer vision-based irradiance forecasting problems, from sky cameras or satellite imagery, provided a significant performance increase compared to the single use of raw data without data transformation or data augmentation with rotations. Models learning from this representation train faster while decreasing temporal misalignment, a key aspect of forecasting tasks. In addition, the SPIN method is likely to facilitate knowledge sharing by standardising the representation of scenes observed with different camera setups.

Acknowledgements The authors acknowledge SIRTA and EUMETSAT for providing the data used in this study. We are grateful to Gisela Lechuga, Aleksandra Marconi and Marcos Gomes-Borges for their technical assistance and valuable comments on the manuscript. This research was supported by ENGIE Lab CRIGEN, EPSRC and the University of Cambridge.

References

[1] Joan Bech and Jorge Luis Chau. *Doppler Radar Observations: Weather Radar, Wind Profiler, Ionospheric Radar, and Other Advanced Applications*. BoD – Books on Demand, Apr. 2012. 7

[2] David Bernecker, Christian Riess, Elli Angelopoulou, and Joachim Hornegger. Continuous short-term irradiance forecasts using sky images. *Solar Energy*, 110:303–315, Dec. 2014. 3

[3] Philippe Blanc, Benoît Gschwind, Mireille Lefèvre, and Lucien Wald. The HelioClim Project: Surface Solar Irradiance Data for Climate Applications. *Remote Sensing*, 3(2):343–361, Feb. 2011. 4, 12

[4] Philippe Blanc, Pierre Massip, Andreas Kazantzidis, Panagiotis Tzoumanikas, Pascal Kuhn, Stefan Wilbert, David Schüler, and Christoph Prahl. Short-term forecasting of high resolution local DNI maps with multiple fish-eye cameras in stereoscopic mode. *AIP Conference Proceedings*, 1850(1):140004, June 2017. 1

[5] EUMETSAT (Organization). *EUMETSAT : The European Organisation for Meteorological Satellites*. EUMETSAT, [1991] ©1991, Darmstadt-Eberstadt, Germany, [1991]. 2

[6] Cong Feng and Jie Zhang. SolarNet: A sky image-based deep convolutional neural network for intra-hour solar forecasting. *Solar Energy*, 204:71–78, July 2020. 2

[7] Laura Frías-Paredes, Fermín Mallor, Teresa León, and Martín Gastón-Romeo. Introducing the Temporal Distortion Index to perform a bidimensional analysis of renewable energy forecast. *Energy*, 94:180–194, Jan. 2016. 4

[8] M. Haeffelin, L. Barthès, O. Bock, C. Boitel, S. Bony, D. Bouniol, H. Chepfer, M. Chiriaco, J. Cuesta, J. Delanoë, P. Drobinski, J.-L. Dufresne, C. Flamant, M. Grall, A. Hodzic, F. Hourdin, F. Lapouge, Y. Lemaître, A. Mathieu, Y. Morille, C. Naud, V. Noël, W. O'Hirok, J. Pelon, C. Pietras, A. Protat, B. Romand, G. Scialom, and R. Vautard. SIRTA, a ground-based atmospheric observatory for cloud and aerosol research. *Annales Geophysicae*, 23(2):253–275, Feb. 2005. 1, 2, 3

[9] A. Hammer, D. Heinemann, E. Lorenz, and B. Lückehe. Short-term forecasting of solar radiation: A statistical approach using satellite data. *Solar Energy*, 67(1):139–150, July 1999. 1

[10] M. Hasenbalg, P. Kuhn, S. Wilbert, B. Nouri, and A. Kazantzidis. Benchmarking of six cloud segmentation algorithms for ground-based all-sky imagers. *Solar Energy*, 201:596–614, May 2020. 3

[11] H. Huang, J. Xu, Z. Peng, S. Yoo, D. Yu, D. Huang, and H. Qin. Cloud motion estimation for short term solar irradiation prediction. In *2013 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 696–701, Oct. 2013. 2

[12] Ruoqiao Jiang and Shaohui Mei. Polar Coordinate Convolutional Neural Network: From Rotation-Invariance to Translation-Invariance. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 355–359, Sept. 2019. 3

[13] Weicong Kong, Youwei Jia, Zhao Yang Dong, Ke Meng, and Songjian Chai. Hybrid approaches based on deep whole-sky-image learning to photovoltaic generation forecasting. *Applied Energy*, 280:115875, Dec. 2020. 2

[14] Benjamin Kurtz, Felipe Mejia, and Jan Kleissl. A virtual sky imager testbed for solar energy forecasting. *Solar Energy*, 158:753–759, Dec. 2017. 2

[15] Dmitry Laptev, Nikolay Savinov, Joachim M. Buhmann, and Marc Pollefeys. TI-Pooling: Transformation-Invariant Pooling for Feature Learning in Convolutional Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 289–297, 2016. 3

[16] Vincent Le Guen and Nicolas Thome. A Deep Physical Model for Solar Irradiance Forecasting With Fisheye Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 630–631, 2020. 2

[17] M. Lefèvre, A. Oumbe, P. Blanc, B. Espinar, B. Gschwind, Z. Qu, L. Wald, M. Schroedter-Homscheidt, C. Hoyer-Klick, A. Arola, A. Benedetti, J. W. Kaiser, and J.-J. Morcrette. McClear: A new model estimating downwelling solar radiation at ground level in clear-sky conditions. *Atmospheric Measurement Techniques*, 6(9):2403–2418, Sept. 2013. 4

[18] Qingyong Li, Weitao Lu, and Jun Yang. A Hybrid Thresholding Algorithm for Cloud Detection on Ground-Based Color Images. *Journal of Atmospheric and Oceanic Technology*, 28(10):1286–1296, Oct. 2011. 3

[19] Ricardo Marquez and Carlos F. M. Coimbra. Intra-hour DNI forecasting based on cloud tracking image analysis. *Solar Energy*, 91:327–336, May 2013. 2

[20] Steven D. Miller, Matthew A. Rogers, John M. Haynes, Manajit Sengupta, and Andrew K. Heidinger. Short-term solar irradiance forecasting via satellite/model coupling. *Solar Energy*, 168:102–117, July 2018. 2

[21] Yuhao Nie, Ahmed S. Zamzam, and Adam Brandt. Resampling and data augmentation for short-term PV output prediction based on an imbalanced sky images dataset using convolutional neural networks. *Solar Energy*, 224:341–354, Aug. 2021. 2

[22] Bijan Nouri, Pascal Kuhn, Stefan Wilbert, Christoph Prahl, Robert Pitz-Paal, Philippe Blanc, Thomas Schmidt, Zeyad Yasser, Lourdes Ramirez Santigosa, and Detlev Heineman. Nowcasting of DNI maps for the solar field based on voxel carving and individual 3D cloud objects from all sky images. *AIP Conference Proceedings*, 2033(1):190011, Nov. 2018. 2

[23] Evangelos Ntavelis, Jan Remund, and Philipp Schmid. Skycam: A dataset of sky images and their irradiance values, 2021. 2

[24] EUMETSAT (Organization). *EUMETSAT : The European Organisation for Meteorological Satellites*. Darmstadt-Eberstadt, Germany, 1991. 3

[25] 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, Aug. 2021. 2, 4, 5

[26] Quentin Paletta, Anthony Hu, Guillaume Arbod, and Joan Lasenby. ECLIPSE : Envisioning Cloud Induced Perturba-

tions in Solar Energy. *arXiv:2104.12419 [cs]*, Apr. 2021. 2, 3, 5, 7, 16

[27] Quentin Paletta and Joan Lasenby. Convolutional Neural Networks Applied to Sky Images for Short-Term Solar Irradiance Forecasting. In *EU PVSEC*, 2020. 5

[28] Quentin Paletta and Joan Lasenby. A temporally consistent image-based sun tracking algorithm for solar energy forecasting applications. In *NeurIPS 2020 Workshop on Tackling Climate Change with Machine Learning*, 2020. 3, 4

[29] Hugo T. C. Pedro, David P. Larson, and Carlos F. M. Coimbra. A comprehensive dataset for the accelerated development and benchmarking of solar forecasting methods. *Journal of Renewable and Sustainable Energy*, 11(3):036102, May 2019. 2

[30] Zhenzhou Peng, Dantong Yu, Dong Huang, John Heiser, and Paul Kalb. A hybrid approach to estimate the complex motions of clouds in sky images. *Solar Energy*, 138:10–25, Nov. 2016. 2

[31] 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, Aug. 2015. 1

[32] Emilio Pérez, Javier Pérez, Jorge Segarra-Tamarit, and Hector Beltran. A deep learning model for intra-day forecasting of solar irradiance using satellite-based estimations in the vicinity of a PV power plant. *Solar Energy*, 218:652–660, Apr. 2021. 2

[33] S. Quesada-Ruiz, Y. Chu, J. Tovar-Pescador, H. T. C. Pedro, and C. F. M. Coimbra. Cloud-tracking methodology for intra-hour DNI forecasting. *Solar Energy*, 102:267–275, Apr. 2014. 2

[34] Hojjat Salehinejad, Shahrokh Valaee, Tim Dowdell, and Joseph Barfett. Image Augmentation Using Radial Transform for Training Deep Neural Networks. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3016–3020, Apr. 2018. 3

[35] Loïc Vallance, Bruno Charbonnier, Nicolas Paul, Stéphanie Dubost, and Philippe Blanc. Towards a standardized procedure to assess solar forecast accuracy: A new ramp and time alignment metric. *Solar Energy*, 150:408–422, July 2017. 4

[36] Dazhi Yang, Stefano Alessandrini, Javier Antonanzas, Fernando Antonanzas-Torres, Viorel Badescu, Hans Georg Beyer, Robert Blaga, John Boland, Jamie M. Bright, Carlos F. M. Coimbra, Mathieu David, Âzeddine Frimane, Christian A. Gueymard, Tao Hong, Merlinde J. Kay, Sven Killinger, Jan Kleissl, Philippe Lauret, Elke Lorenz, Dennis van der Meer, Marius Paulescu, Richard Perez, Oscar Perpiñán-Lamigueiro, Ian Marius Peters, Gordon Reikard, David Renné, Yves-Marie Saint-Drenan, Yong Shuai, Ruben Urraca, Hadrien Verbois, Frank Vignola, Cyril Voyant, and Jie Zhang. Verification of deterministic solar forecasts. *Solar Energy*, 210:20–37, Nov. 2020. 4

[37] Jinsong Zhang, Rodrigo Verschae, Shohei Nobuhara, and Jean-François Lalonde. Deep photovoltaic nowcasting. *Solar Energy*, 176:267–276, Dec. 2018. 2

[38] Xin Zhao, Haikun Wei, Hai Wang, Tingting Zhu, and Kanjian Zhang. 3D-CNN-based feature extraction of ground-based cloud images for direct normal irradiance prediction. *Solar Energy*, 181:510–518, Mar. 2019. 2