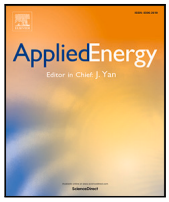
Applied Energy 362 (2024) 122967

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy



Vision transformer models to measure solar irradiance using sky images in temperate climates

Thomas M. Mercier [a,*], Amin Sabet [b], Tasmiat Rahman [c]

[a] Bournemouth University, Poole, BH12 5BB, Dorset, United Kingdom
[b] EscherCloud AI, Laarderhoogtweg 18, Amsterdam, 1101, EA, Netherlands
[c] University of Southampton, University road, Southampton, SO17 1BJ, Hampshire, United Kingdom

ARTICLE INFO

Keywords:
Computer vision
Machine learning
Solar irradiance
Sky imaging

ABSTRACT

Solar Irradiance measurements are critical for a broad range of energy systems, including evaluating performance ratios of photovoltaic systems, as well as forecasting power generation. Using sky images to evaluate solar irradiance, allows for a low-cost, low-maintenance, and easy integration into Internet-of-things network, with minimal data loss. This work demonstrates that a vision transformer-based machine learning model can produce accurate irradiance estimates based on sky-images without any auxiliary data being used. The training data utilizes 17 years of global horizontal, diffuse and direct data, based on a high precision pyranometer and pyrheliometer sun-tracked system; in-conjunction with sky images from a standard lens and a fish-eye camera. The vision transformer-based model learns to attend to relevant features of the sky-images and to produce highly accurate estimates for both global horizontal irradiance (RMSE =52 W/m^2) and diffuse irradiance (RMSE = 31 W/m^2). This work compares the model's performance on wide field of view all-sky images as well as images from a standard camera and shows that the vision transformer model works best for all-sky images. For images from a normal camera both vision transformer and convolutional architectures perform similarly with the convolution-based architecture showing an advantage for direct irradiance with an RMSE of 155 W/m^2.

1. Introduction

Global deployment of photovoltaic systems continues to grow at pace, reaching 1.2 TW capacity by the end of 2022. To sustain such rapid growth, there is an urgent need for highly accurate and reliable measurements of irradiance, which is critical in evaluating performance ratios, as well as power generation forecasting. The accuracy and reliability of these measurements help to reduce the levelized cost of electricity when deploying photovoltaic systems. This is of even further interest for the deployment of bifacial systems, whereby irradiance measurements are used to calculate albedo. Additionally, for sun-tracking systems, which ITRPV predicts 40% market share by 2030 [1], irradiance and subsequent plane-of-array (POA) estimation is needed for real-time tracking algorithms [2]. Currently, the state-of-art and industry standard measurement technique is to utilize Class-A pyranometers for measuring global horizontal irradiance, whilst an additional sun-tracking system and shadow ball is required for diffuse horizontal irradiance. Furthermore, a pyrheliometer as well as a sun-tracked system is required for direct normal irradiance. Therefore, several expensive systems are required to measure GHI, DHI and DNI, which need regular maintenance and costly recalibration [3,4]. In addition, they can be limited in sensitivity and response time, without further investments in expensive datalogging systems. An alternative approach to measuring GHI, DHI and DNI is to utilize a single digital camera, and evaluate the irradiance based on sky imaging through machine learning (ML) models [5]. ML optimizes adaptive models with algorithms like gradient descent, using extensive training data [6]. In a supervised learning approach these models learn by adjusting parameters or "model weights" based on errors calculated against expected outputs for specific inputs. After finding suitable parameters, these models can predict values for new inputs, a process termed inference. Deep learning (DL), a subset of ML, typically features multi-layer neural network based models. It commonly separates the base network responsible for heavy computations from the head network containing fewer layers, the latter of which varies depending on the task. By leveraging transfer learning, the previously learned parameters are retrained in the base model while replacing the head model with one suitable for the new task. Here, pretrained base model parameters from an image classification task are used and fine-tuned to map irradiance values to sky images [7,8]. Two types of sky images are utilized in this study, all-sky images taken using a fish-eye lens camera

* Corresponding author.
E-mail addresses: tmercier@bournemouth.ac.uk (T.M. Mercier), a.sabet@eschercloud.ai (A. Sabet), t.rahman@soton.ac.uk (T. Rahman).

https://doi.org/10.1016/j.apenergy.2024.122967
Received 22 August 2023; Received in revised form 8 February 2024; Accepted 2 March 2024
Available online 13 March 2024

and sky images from a standard camera. Sky images refer to cloud-imaging using cameras situated on the ground, looking onto the sky above. Deploying these models offers cost-effective ways to generate accurate irradiance data both online and retroactively. Sky images are important as they offer a sub-kilometer view of cloud shadows, which will impact the irradiance values. This provides sufficient spatial and temporal resolution to estimate GHI, DHI and DNI effectively [9]. In this paper, it is investigated the use of a vision transformer architecture to create a model that can take in a single sky image without any auxiliary data and produce an irradiance estimate. Additionally, it is shown that images from standard cameras as well as all-sky imagers can be used to accurately model the levels of irradiance by making use of such DL based models. The transformer architecture has revolutionized the field of natural language processing and its recent application to computer vision tasks has shown them to be competitive with deep convolutional neural networks (CNNs) [10]. It utilizes multiple layers of self-attention to focus on the most relevant parts of the image. The high performance and interpretability of this architecture makes it an attractive candidate for mapping of sky images to irradiance.

As the level of solar irradiance seen in a particular location varies based on the cyclical changes of the season, the sun position throughout the day, and weather conditions, it is important to test the DL models from training data in locations where the level of cloud cover changes frequently (i.e. temperate climates). In this work, data from the Chilbolton Facility for Atmospheric and Radio Research (CFARR), Hampshire, UK (51.1445N, 1.4270W) [11–14] is utilized, using over 17 years or sky images and irradiance measurements.

To the best of the authors knowledge this paper is the first to deploy a vision transformer-based model to the task of solar irradiance modeling and is built from training data in temperate climates.

The paper is structured as follows: First an overview of previous work in the field of irradiance modeling is presented. Following this, the proposed modeling framework is introduced detailing the approach to accurately estimate irradiance levels. The fourth section discusses the methods used to evaluate the performance of the proposed model. The dataset used in this work is described in section five. The sixth section details the implementation of the proposed model, including the training. Results and discussion are presented in the seventh section, where the performance of the proposed model is analyzed and compared to previous methods. Finally, a conclusion is provided, summarizing the findings of this study and highlighting the potential benefits and applications of using machine learning to map sky images to irradiance values within the field of solar energy.

The contributions of this paper are as follows:

- Demonstrate that expensive pyranometer equipment can be substituted by all-sky cameras feeding images to DL models
- It is shown that even normal camera images can serve as the basis for usable irradiance estimates
- Vision transformer model is thoroughly compared to a conventional approach and shown to be advantageous for all-sky images
- It is demonstrated that the model learns to attend to relevant features of the sky images

2. Related work

Classically, irradiance modeling has been based on geographic data such as latitude and longitude as well as solar elevation and altitude in addition to meteorological input data such as aerosol content and atmospheric water vapor column [15]. Most of the models that use atmospheric data as input are considered clear-sky models, which estimate the terrestrial surface irradiance for cloudless conditions [16]. These models vary widely in their complexity and in their required input data. Due to high capital and operational costs, frequent calibration and maintenance there is a need for alternative methods for measuring GHI, DHI and DNI [3,17,18]. Due to the general success and the increasingly low barrier of entry of ML, it has seen broad adoption in the physical sciences [19,20]. DL based approaches, in particular, have become increasingly popular for tackling previously intractable or poorly addressed problems. In the field of computer vision DL models utilizing convolutional layers, called deep convolutional networks (CNNs) have been particularly successful [21]. In the field of irradiance modeling such a model has been used to extract relevant features from all-sky images, which were then fed into a multi-layer-perceptron (MLP) to either classify whether the sun was shaded or to directly map images to corresponding irradiance measures [5]. The authors make use of a dataset provided by NREL [22]. To train and validate their model they restrict the dataset to samples collected during summertime between 8 am and 4 pm, totaling 21 600 images. They initialize the model weights for the irradiance mapping task from a model trained to classify sky images as either showing the sun or the sun being covered by clouds. Using this transfer learning approach for the irradiance mapping task they report a root mean square error (RMSE) of 130 W/m^2 on their testing dataset.

Another deep CNN based model has been used for estimating the angular dependence of irradiance based on all-sky images and sun-position [23]. They use the model extract information from the image and concatenate the resulting features with the information from the sun-position fed through an MLP. This serves as the input to the head network consisting of an additional MLP. Using a dataset provided by NREL they show widely varying model performance depending on the tilt angle. ResNet architectures, which integrate residual connections into CNN architectures, have been used in combination with a cloudiness classification to explore modeling of GHI values based on all-sky-images [24,25]. The authors trained sub-models for different sky conditions to output GHI estimates. They present a physics based model to classify images into three sky conditions. Based on a limited dataset of 1200 test images from the SIRTA dataset [26], they report an RMSE of 14.63 W/m^2 for images classified as sunny, 51.63 W/m^2 for partially overcast and 53.38 W/m^2 for overcast. Furthermore they show that increasing the model size only improved performance on sunny images. Another CNN based approach in combination with convolutional block attention modules and local cloud cover values was reported to achieve good results for mapping all-sky images to GHI values [27]. They present a methodology to generate local cloud cover auxiliary data to improve model performance. In their architecture they combine channel attention and spatial attention modules on the output of convolutional blocks. For a dataset of sky images and GHI values provided by NREL, they report an RMSE value normalized by the average ground truth GHI value of 11%. An approach combining classical ML with DL architectures is presented by Henriques de Sá [28]. They extract 17 features such as the fraction of the sun that is covered and the average pixel intensity for every image in a dataset of all-sky images and use an MLP to estimate GHI. On their test set of 9996 images they report an RMSE of 72.3 W/m^2. Chu et al. have demonstrated that a using a network of cameras providing all-sky images in combination with a irradiance mapping algorithm can be utilized to map out irradiance over large areas without the use of radiometers [29]. They mask the images in several different ways and extract a total of 900 features from each image. They report an RMSE of 86.4 W/m^2 for a simple linear regression model and 73.2 W/m^2 for an MLP based model using a subset of 360 features. Using information from several different measurement systems, including an all-sky imager, Blum et al. showed that diffuse irradiance and plane of array irradiance can be estimated [3]. This shows that improving the estimates of irradiance from all-sky imagers can contribute to a variety of solar applications.

As the above discussion shows there have been several attempts to use DL based approaches for the task of irradiance modeling. Such approaches are typically based on CNN architectures while transformer architectures based on the self-attention mechanism are largely unexplored. Furthermore, modeling is largely focused on GHI.

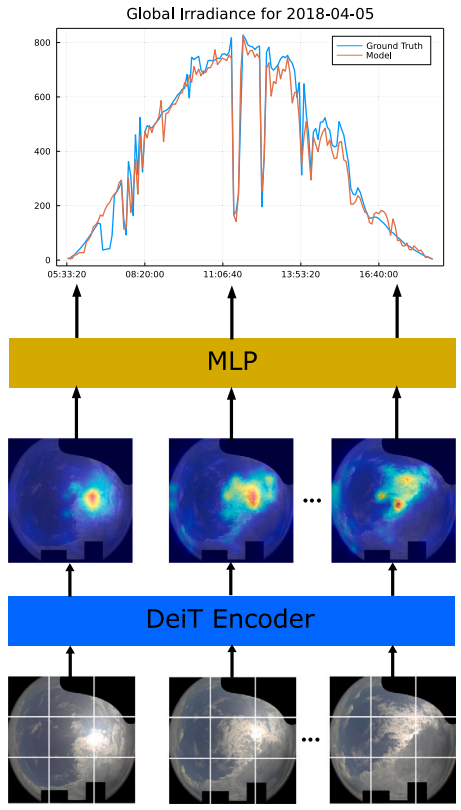


Fig. 1. High level illustration of the model's operation with the number of patches shown being less than the model uses (for illustrative purposes).

3. Proposed framework

Fig. 1 shows a high-level overview of the proposed framework. A data-efficient image transformer (DeiT) with a linear head to map sky images directly to irradiance values is utilized [30,31]. This type of architecture is part of the vision transformer family of models. Unlike convolution based architectures, this type of architecture does work directly with the image arrays. Instead, the input image is first split into a sequence of patches which are then rolled out into vectors, this serves as the input to a trainable linear layer which projects them to the internal encoding dimension of the model. To the sequence of embedded patches, a learnable embedding is prepended, if the model is used for image classification, this embedding represents the class. Positional embeddings are added to give the model positional information. This serves as the input for the transformer encoder, which consists of multiple blocks of multi-headed self-attention and MLP layers. A full discussion of the attention blocks and the underlying mechanism is beyond the scope of this paper and the reader is referred to the original publication [32]. From the output of the transformer encoder, only the prepended class embedding is put through an MLP to produce an irradiance value. The DeiT-based model has a patch size of 16 pixels, an embedding dimension of 192, a depth of 12 layers with 3 attention heads. The choice of model configuration is based on the available pretrained model weights as well as model size considerations.

4. Model evaluation

In order to evaluate the models performance and facilitate comparison to related works the commonly employed metric of RMSE is used. In addition Mean Bias Error (MBE) and t-statistic are reported as evaluation metrics [33,34]. It is a useful metric for regression problems as the resulting numbers can be intuitively evaluated as they are of the same unit and order of magnitude as the target values. The evaluation metrics can be calculated as follows [35]:

Table 1
Models with associated evaluation metrics for both the all-sky image and the normal camera datasets.

Model	Images used	Target	RMSE	*t*-statistic	MBE
DeiT	All-sky	Global	52	130	31
ResNet	All-sky	Global	55	145	36
DeiT	Normal	Global	77	184	44
ResNet	Normal	Global	78	197	47
DeiT	All-sky	Diffuse	31	99	19
ResNet	All-sky	Diffuse	33	108	22
DeiT	Normal	Diffuse	47	97	30
ResNet	Normal	Diffuse	44	103	29
DeiT	All-sky	Direct	94	68	55
ResNet	All-sky	Direct	94	74	58
DeiT	Normal	Direct	172	63	122
ResNet	Normal	Direct	155	62	109

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(y_k - \hat{y}_k\right)^2} \tag{1}$$

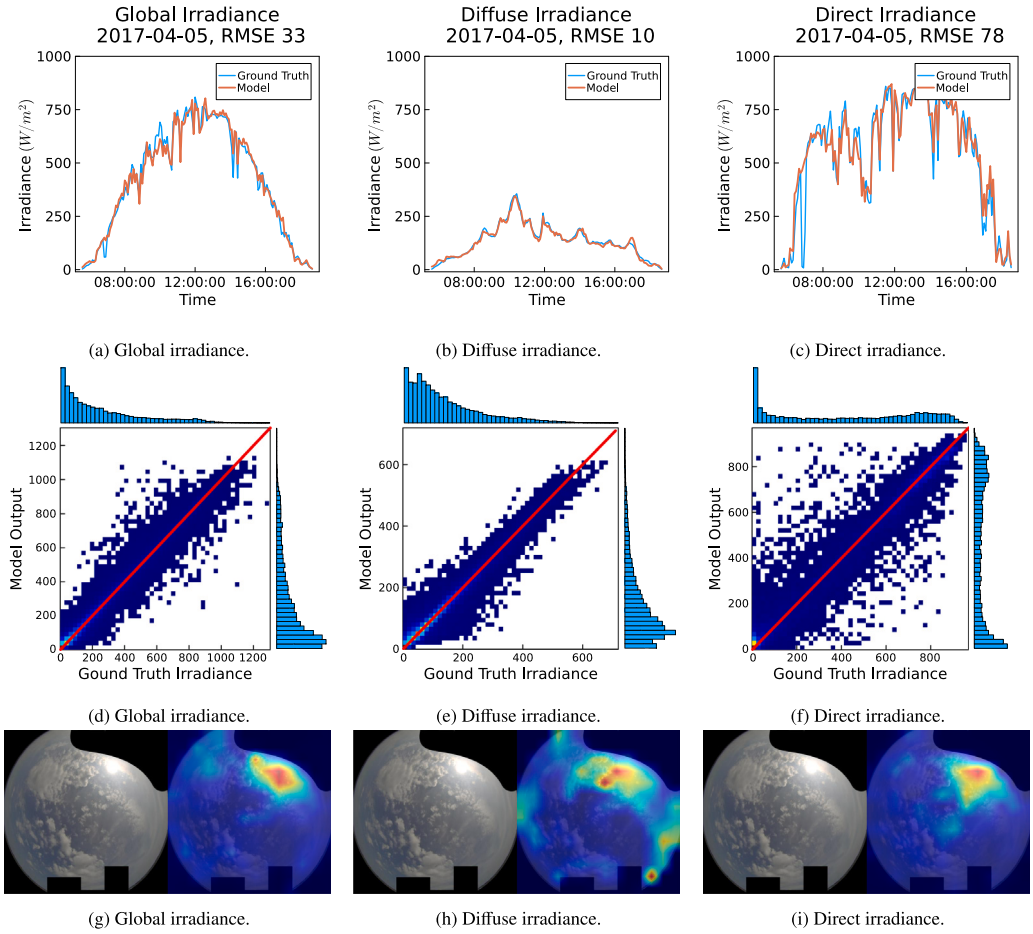
$$MBE = \frac{1}{n}\sum_{k=1}^{n}\left(y_k - \hat{y}_k\right) \tag{2}$$

$$t-statistic = \sqrt{\frac{(n-1)MBE^2}{RMSE^2 - MBE^2}} \tag{3}$$

Here n is the number of samples, y_k is the ground truth irradiance and $\hat{y}_k$ is the model output for an input image. During training of the model, a simple mean squared error (MSE) is used.

5. Dataset

The data was originally gathered at the Chilbolton Facility for Atmospheric and Radio Research (CFARR), Hampshire, UK (51.1445N, 1.4270W) [11–14]. Three types of illumination data were available: two pyranometers collected whole sky radiation and diffuse solar radiation. A pyrheliometer collected direct irradiance. The data for the project consisted of files with the radiometer measurements with each file containing the data for a single day and the cloud image files of two different cameras. The radiometer files contained the measurements in units of W/m^2 with timestamps accurate to 1 s with each file containing about 8000 data points. Two types of cloud images were available, one from a regular camera with a limited field of view, collection of which was discontinued in 2016 and a fisheye (all-sky) camera with a full 180-degree field of view, collection of which started in 2016. The all-sky images were collected by National Center for Atmospheric Science (NCAS). Since the images for both cameras were taken roughly every 5 min, the data was pre-processed such that the radiometer data was averaged over a time window of 30 s and this average was assigned to one image. The data was aligned based on the timestamps so that the time window for averaging the radiometer data always started at the time stamp of the image. This allowed a direct mapping of one image to an irradiance value. Fig. 2 shows the raw measurement data that was available from Chilbolton for a single day for all three targets. As can be seen the illumination data varies smoothly with time of day but shows strong drops in illumination related to change in cloud cover. This raw data was used to assign to each image an average of the measurement data available for a 30 s time window starting from the timestamp of the image. This had the consequence that the target data is somewhat smoother compared to the raw data. The histograms of the target data in Fig. 2 show that the global and diffuse irradiance values have a right skew while the direct irradiance values are dominated by very low values with a noticeable bump near the peak values. The distribution of data here had different threshold values applied below which the data points were removed from the dataset. The threshold values were 10,



(a) Global irradiance. (b) Diffuse irradiance. (c) Direct irradiance.

(d) Global irradiance. (e) Diffuse irradiance. (f) Direct irradiance.

(g) Global irradiance. (h) Diffuse irradiance. (i) Direct irradiance.

Fig. 2. Results of using all-sky images to train the DeiT model with a-c showing comparisons of ground truth irradiance data to the model's estimates for a single day, d-f showing comparison of ground truth to model estimates for all samples in the testset and g-i showing what the model learns to pay attention to illustrated via attention maps.

10 and 2 W/m^2 for global, diffuse and direct irradiance, respectively. To make sure the dataset did not contain any very dark images it was further restricted by removing all datapoints taken between 11 pm and 3 am. Not every image had all three irradiance values available, hence the number of training and testing samples varied between irradiance targets. Table 1 gives an overview of the number of samples used during model training for both cameras and all targets. To reduce the GPU memory requirements and to enable the use of transfer learning from weights pretrained on ImageNet, all images were resized to 224 by 224 pixels, a commonly used image size in DL computer vision tasks. After each image was aligned to its closest window of measurements, the data was split into a training and evaluation dataset as well as a separate testing dataset. This split was done by using days 5 to 9 of each month as the fixed testing dataset, days 15 to 19 as an evaluation dataset while keeping the rest for training. The mean and standard deviation of the training dataset were used to normalize all target data to have a mean of 0 and a standard deviation of 1 during training. Additionally the images were transformed using the mean and standard deviation of the ImageNet dataset [7,8]. For the images taken by the all-sky imager, a mask of black pixels was used to block out objects in the frame to make sure the model learns to extract the relevant information from the image of the sky.

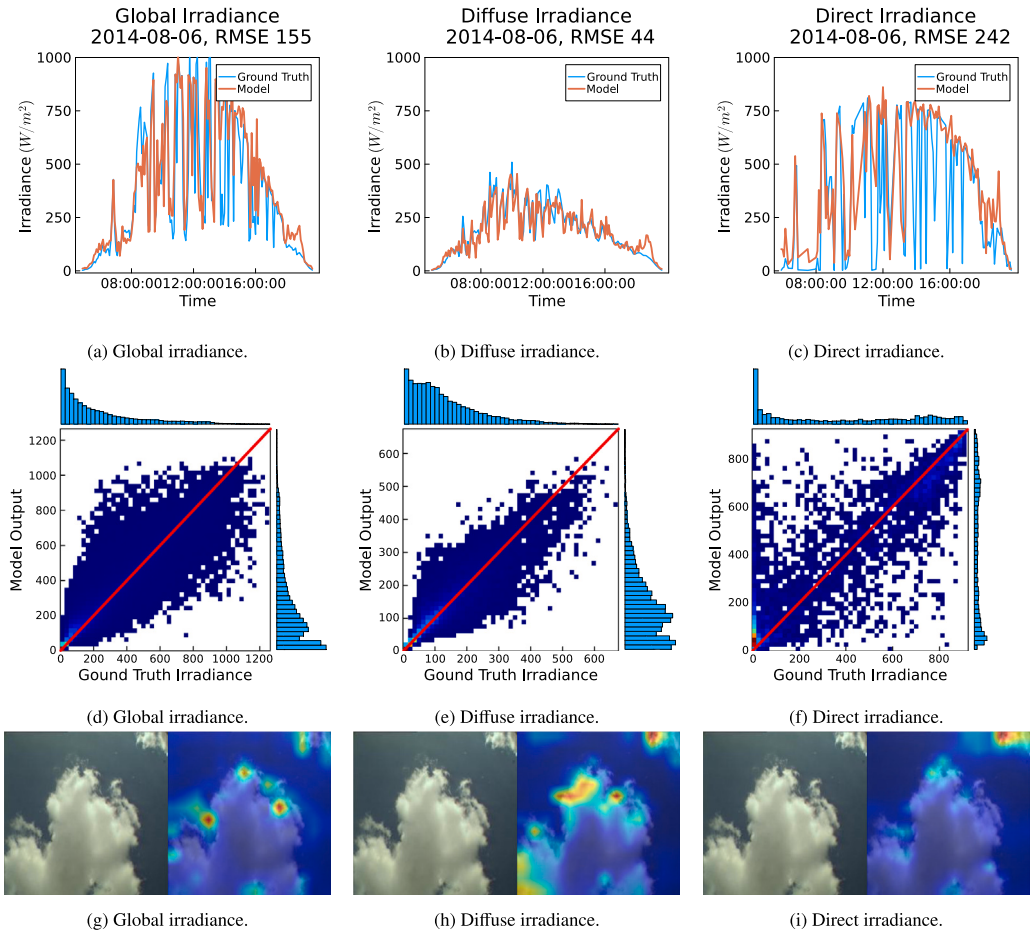
6. Implementation and training setting

Both models were implemented using PyTorch [36] and the code is publicly available.[1] The weights of both models are initialized from models pretrained on the ImageNet dataset [8]. The weights of the MLP head were randomly initialized. The AdamW optimizer with a learning rate of 9e−5 was used for training [37]. An exponential learning rate warmup and decay was used during training. Randomized image augmentation was applied during training by applying random rotations up to 45 degrees with a probability of 10%. The model is trained for 14 epochs with a batch size of 128. The model is trained to produce estimates of global irradiance and then fine tuned to produce estimates for diffuse and direct irradiance. This procedure is carried out for both the all-sky images as well as the images from the ordinary camera.

7. Results and discussion

Table 1 gives an overview of the achieved performance of the DeiT and ResNet models for all types of irradiance. Since most of the competing models found in literature are based on CNNs, a typical ResNet is chosen as a comparison. Both the proposed vision transformer and the conventional ResNet are able to produce accurate estimates for global as well as for diffuse irradiance but struggle with direct irradiance. This is likely related to the distribution of values, which are much more skewed towards very low values with a slight bump in the frequency of values at the higher end for the target of direct irradiance, as illustrated in Fig. 2(f). Learning to estimate direct irradiance from the given dataset is also difficult due to the lower number of training samples available. It is notable that the DeiT model outperforms the ResNet model when all-sky images are used as a basis for the irradiance estimation while the ResNet shows a slight advantage when images from the normal camera are used. Overall, the estimates are much more accurate when all-sky images are used as the input to the model. This is

[1] Model code can be found here: https://github.com/Gittingthehubbing/Solar_Irradiance_ViT.



(a) Global irradiance. (b) Diffuse irradiance. (c) Direct irradiance.

(d) Global irradiance. (e) Diffuse irradiance. (f) Direct irradiance.

(g) Global irradiance. (h) Diffuse irradiance. (i) Direct irradiance.

Fig. 3. Results of using images from normal camera to train the DeiT model with a-c showing comparisons of ground truth irradiance data to the model's estimates for a single day, d-f showing comparison of ground truth to model estimates for all samples in the testset and g-i showing attention maps.

likely related to the amount of information that the model can extract from the larger field of view being much higher. However, if cost is the main concern, even images from a normal camera can produce usable estimates.

The overall performance of the DeiT model is well illustrated in Figs. 2 and 3, which show both the deviation from ground truth for a single day as well as a comparison of the model's estimates to ground truth as density plots for all three types of irradiance targets. Here it can be seen that using all-sky images as the DeiT model's input produces much tighter distribution around the ideal. The density plot also illustrates that the model struggles with direct irradiance, as this presents the most challenging target due to the high volatility and the target's distribution, with this being especially pronounced for the normal images. Overall, Fig. 2 shows that the DeiT based model can produce reliable estimates for all irradiance targets even when there is significant volatility in the data. Furthermore, Fig. 3 shows that the model still manages to produce good estimates for global and diffuse irradiance when images from an ordinary camera are used. However, it struggles to produce useful estimates for direct irradiance. While the model is shown to perform well on data from the Chilbolton facility, it should be mentioned that a limitation of the proposed approach and ML based approaches in general is that the trained model is unlikely to generalize well to datasets far outside its training data distribution, which means a model would have to be fine-tuned on data from the new source if one wishes to use the model on data recorded at a different facility or using different equipment.

To make the inner workings of the DeiT model more interpretable, attention rollout was performed using the same input image for models trained to estimate different targets [38]. As the attention maps in Fig. 2 show, for the all-sky images, the model pays particular attention to relevant features of the sky images for all targets. Particular attention is paid to the part of the image showing the sun and its immediate surrounding area with the pattern of attention being similar when the model is trained to predict global irradiance and direct irradiance but differs significantly for diffuse irradiance predictions. The latter pattern being more distributed across the image and less focused on the position of the sun. Repeating the same procedure using images from a normal camera results in the attention maps shown in Fig. 3. It is notable that the differences in attention maps for the models trained for different targets differs more than for the all-sky images. This further shows that the normal images do not allow for the same level of information extraction and do not enable the model to learn to pay attention to features most relevant for the particular type of irradiance that is to be estimated.

To evaluate how the DeiT model performs under different sky conditions, the testing dataset of the all-sky images is split by clearness index. This index is defined as the ratio of ground level irradiance and extra-terrestrial irradiance [39,40]. Conditions with a clearness index below 0.3 are considered to be overcast, conditions with an index between 0.3 and 0.78 to be intermediate and anything above to be clear. Since Chilboltons rich dataset offers a wide range of sky conditions it can be demonstrated that the model shows good performance in all sky conditions with an RMSE of 41 W/m^2 in overcast, 55 W/m^2 in intermediate and 80 W/m^2 in clear conditions, as can be seen in Fig. 4. A small bias towards overestimating irradiance values under overcast and underestimating irradiance values under clear conditions is present. This further illustrates the robustness of the proposed approach and supports the idea of using sky cameras in various conditions to broaden the acquisition of irradiance measurements in various locations.

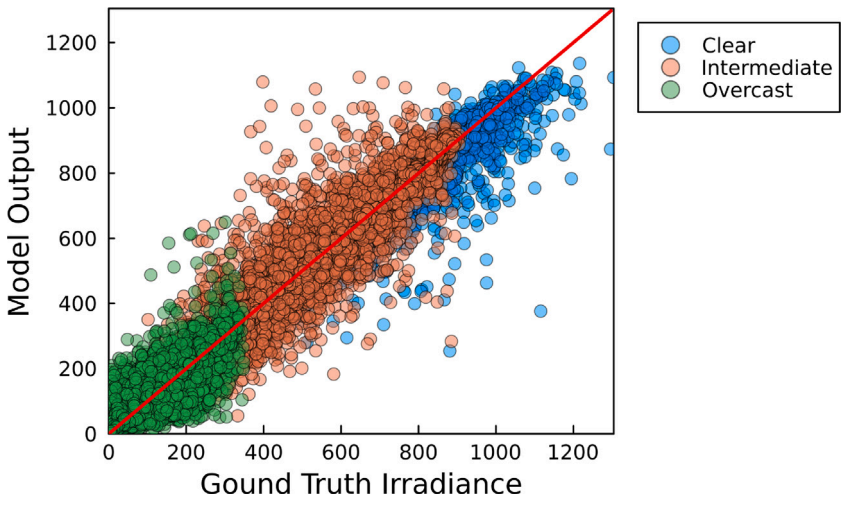


Fig. 4. DeiT model estimates against ground truth for global irradiance using all-sky images as input with the results being split by clearness index.

8. Conclusion

Using field data spanning 17 years from a temperate climate location, it has been demonstrated that a vision transformer-based model can produce accurate irradiance estimates based on sky-images without any auxiliary data being used. It is shown that the model learns to attend to relevant features depending on the type of irradiance that is to be estimated. Comparisons are made for model performance using images from a normal camera as well as images from an all-sky imager with the use of all-sky images resulting in better estimates. Furthermore, to assess the DeiT models performance in the context of previous work, a comparison is made between the proposed attention-based model and a conventional CNN type network, a ResNet, which resulted in the DeiT model outperforming the comparison model when all-sky images are used, while the ResNet showed performance on par with DeiT when normal images are used as input. This work shows that relatively inexpensive cameras in conjunction with DL based models can serve as a reliable replacement for pyranometer and pyrheliometer equipment to assess and monitor site conditions. In the future, this image prediction will be used to link irradiance to POA and optimize tracker position through improved sun-tracking algorithms, as well as improving decomposition and transposition models.

CRediT authorship contribution statement

Thomas M. Mercier: Writing – original draft, Visualization, Software, Methodology, Investigation, Conceptualization. **Amin Sabet:** Software, Methodology. **Tasmiat Rahman:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

A link to the model code is available in footnote 1. The underlying data is available upon request.

Acknowledgments

We would like to thank the Chilbolton observatory for providing the data used in this publication. T Rahman acknowledges funding from EPSRC, United Kingdom EP/X033333/1.

References

[1] Fischer M, Woodhouse M, Herritsch S, Trube J. International technology roadmap for photovoltaics 2022 R&D results. 2022.
[2] Trube J, Herritsch S. International technology roadmap for photovoltaic (ITRPV). Tech. rep., VDMA; 2022, p. 81.
[3] Blum NB, Wilbert S, Nouri B, Lezaca J, Huckebrink D, Kazantzidis A, Heinemann D, Zarzalejo LF, Jiménez MJ, Pitz-Paal R. Measurement of diffuse and plane of array irradiance by a combination of a pyranometer and an all-sky imager. Sol Energy 2022;232:232–47. http://dx.doi.org/10.1016/j.solener.2021.11.064.
[4] Bakouri K, Foqha T, Ahwidi O, Abubaker A, Nassar Y, El-Khozondar H. Learning lessons from Murzuq-Libya Meteorological Station: Evaluation criteria and improvement recommendations. J Sol Energy Sustain Dev 2023;12(1). http://dx.doi.org/10.51646/jsesd.v12i1.149.
[5] Lin Y, Duan D, Hong X, Han X, Cheng X, Yang L, Cui S. Transfer learning on the feature extractions of sky images for solar power production. In: 2019 IEEE power energy society general meeting (PESGM). 2019, p. 1–5. http://dx.doi.org/10.1109/PESGM40551.2019.8973423.
[6] Bishop CM. Pattern recognition and machine learning. In: Information science and statistics, New York: Springer; 2006.
[7] Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. 2009, p. 248–55. http://dx.doi.org/10.1109/CVPR.2009.5206848.
[8] Wightman R. PyTorch image models. GitHub Repos 2019. http://dx.doi.org/10.5281/zenodo.4414861.
[9] Chow CW, Urquhart B, Lave M, Dominguez A, Kleissl J, Shields J, Washom B. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. Sol Energy 2011;85(11):2881–93. http://dx.doi.org/10.1016/j.solener.2011.08.025.
[10] Khan S, Naseer M, Hayat M, Zamir SW, Khan FS, Shah M. Transformers in vision: A survey. ACM Comput Surv 2022;3505244. http://dx.doi.org/10.1145/3505244, arXiv:2101.01169.
[11] Science and Technology Facilities Council, Chilbolton Facility for Atmospheric and Radio Research, Wrench C. Chilbolton facility for atmospheric and radio research (CFARR) visible radiometer data. NCAS British atmospheric data centre, 2023/02/24. 2003, https://catalogue.ceda.ac.uk/uuid/Bf70daf01b6257b2475b057029325869.
[12] Science and Technology Facilities Council, Chilbolton Facility for Atmospheric and Radio Research, Council NER, Ladd D. Chilbolton facility for atmospheric and radio research (CFARR): Cloud camera 2 imagery from chilbolton, hampshire (2016-present). NCAS british atmospheric data centre, 2023/02/24. 2016, https://catalogue.ceda.ac.uk/uuid/F55f5649110b4b98b3d5177d8ff2eac9.
[13] Science and Technology Facilities Council, Chilbolton Facility for Atmospheric and Radio Research, Council NER, Wrench C. Chilbolton facility for atmospheric and radio research (CFARR) meteorological sensor data, chilbolton site. NCAS british atmospheric data centre, 2023/02/24. 2003, https://catalogue.ceda.ac.uk/uuid/45b25a7c531563f4422afcaeea0f07a7.
[14] Science and Technology Facilities Council, Chilbolton Facility for Atmospheric and Radio Research, Wrench C, Agnew J. Chilbolton facility for atmospheric and radio research (CFARR) direct visible radiometer data. NCAS british atmospheric data centre. 2023/02/24. 2003.
[15] Ineichen P. A broadband simplified version of the Solis clear sky model. Sol Energy 2008;82(8):758–62. http://dx.doi.org/10.1016/j.solener.2008.02.009.

[16] Stein JS, Hansen CW, Reno MJ. Global horizontal irradiance clear sky models : Implementation and analysis. Tech. Rep. SAND2012-2389, Albuquerque, NM, and Livermore, CA (United States): Sandia National Laboratories (SNL); 2012, http://dx.doi.org/10.2172/1039404.

[17] Sánchez-Segura CD, Valentín-Coronado L, Peña-Cruz MI, Díaz-Ponce A, Moctezuma D, Flores G, Riveros-Rosas D. Solar irradiance components estimation based on a low-cost sky-imager. Sol Energy 2021;220:269–81. http://dx.doi.org/10.1016/j.solener.2021.02.037.

[18] Rajagukguk RA, Kamil R, Lee H-J. A deep learning model to forecast solar irradiance using a sky camera. Appl Sci 2021;11(11):5049. http://dx.doi.org/10.3390/app11115049.

[19] Dai P, Wang Y, Hu Y, de Groot CH, Muskens O, Duan H, Duan H, Huang R, Huang R. Accurate inverse design of Fabry–Perot-cavity-based color filters far beyond sRGB via a bidirectional artificial neural network. Photon Res PRJ 2021;9(5):B236–46. http://dx.doi.org/10.1364/PRJ.415141.

[20] Liu Z, Zhu D, Rodrigues SP, Lee K-T, Cai W. Generative model for the inverse design of metasurfaces. Nano Lett 2018;18(10):6570–6. http://dx.doi.org/10.1021/acs.nanolett.8b03171.

[21] Rawat W, Wang Z. Deep convolutional neural networks for image classification: A comprehensive review. Neural Comput 2017;29(9):2352–449. http://dx.doi.org/10.1162/neco_a_00990.

[22] Stoffel T, Andreas A. NREL solar radiation research laboratory (SRRL): Baseline measurement system (BMS); golden, colorado (data). Tech. Rep. NREL/DA-5500-56488, Golden, CO (United States): National Renewable Energy Lab. (NREL); 1981, http://dx.doi.org/10.7799/1052221.

[23] Pierce BG, Braid JL, Stein JS, Augustyn J, Riley D. Solar transposition modeling via deep neural networks with sky images. IEEE J Photovolt 2022;12(1):145–51. http://dx.doi.org/10.1109/JPHOTOV.2021.3120508.

[24] Insaf IM, Wickramathilaka HMKD, Upendra MAN, Godaliyadda GMRI, Ekanayake MPB, Herath HMVR, Dissawa DMLH, Ekanayake JB. Global horizontal irradiance modeling from sky images using ResNet architectures. In: 2021 IEEE 16th international conference on industrial and information systems (ICIIS). 2021, p. 239–44. http://dx.doi.org/10.1109/ICIIS53135.2021.9660664.

[25] Zhang R, Ma H, Saha TK, Zhou X. Photovoltaic nowcasting with Bi-level spatio-temporal analysis incorporating sky images. IEEE Trans Sustain Energy 2021;12(3):1766–76. http://dx.doi.org/10.1109/TSTE.2021.3064326.

[26] Haeffelin M, Barthès L, Bock O, Boitel C, Bony S, Bouniol D, Chepfer H, Chiriaco M, Cuesta J, Delanoë J, Drobinski P, Dufresne J-L, Flamant C, Grall M, Hodzic A, Hourdin F, Lapouge F, Lemaître Y, Mathieu A, Morille Y, Naud C, Noël V, O'Hirok W, Pelon J, Pietras C, Protat A, Romand B, Scialom G, Vautard R. SIRTA, a ground-based atmospheric observatory for cloud and aerosol research. Ann Geophys 2005;23(2):253–75. http://dx.doi.org/10.5194/angeo-23-253-2005.

[27] Song S, Yang Z, Goh H, Huang Q, Li G. A novel sky image-based solar irradiance nowcasting model with convolutional block attention mechanism. Energy Rep 2022;8:125–32. http://dx.doi.org/10.1016/j.egyr.2022.02.166.

[28] de Sá Campos MH, Tiba C. Global horizontal irradiance modeling for all sky conditions using an image-pixel approach. Energies 2020;13(24):6719. http://dx.doi.org/10.3390/en13246719.

[29] Chu Y, Li M, Pedro HTC, Coimbra CFM. A network of sky imagers for spatial solar irradiance assessment. Renew Energy 2022;187:1009–19. http://dx.doi.org/10.1016/j.renene.2022.01.032.

[30] Touvron H, Cord M, Douze M, Massa F, Sablayrolles A, Jégou H. Training data-efficient image transformers & distillation through attention. 2021, http://dx.doi.org/10.48550/arXiv.2012.12877, arXiv:2012.12877 [cs].

[31] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N. An image is worth 16x16 words: Transformers for image recognition at scale. 2021, http://dx.doi.org/10.48550/arXiv.2010.11929, arXiv:2010.11929 [cs].

[32] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. 2017, arXiv:1706.03762 [cs].

[33] Alsadi S, Nassar Y. Correction of the ASHRAE clear-sky model parameters based on solar radiation measurements in the Arabic countries. Int J Renew Energy Technol Res 2016;5(4):1–16.

[34] Nassar YF, Hafez AA, Belhaj S, Alsadi SY, Abdunnabi MJ, Belgasim B, Sbeta MN. A generic model for optimum tilt angle of flat-plate solar harvesters for middle east and North Africa region. Appl Sol Energy 2022;58(6):800–12. http://dx.doi.org/10.3103/S0003701X22060135.

[35] Nassar YF, Hafez AA, Alsadi SY. Multi-factorial comparison for 24 distinct transposition models for inclined surface solar irradiance computation in the state of palestine: A case study. Front Energy Res 2020;7. http://dx.doi.org/10.3389/fenrg.2019.00163.

[36] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Köpf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S. PyTorch: An imperative style, high-performance deep learning library. 2019, arXiv:1912.01703.

[37] Loshchilov I, Hutter F. Decoupled weight decay regularization. 2019, http://dx.doi.org/10.48550/arXiv.1711.05101, arXiv:1711.05101.

[38] Abnar S, Zuidema W. Quantifying attention flow in transformers. 2020, arXiv:2005.00928 [cs].

[39] Andrews RW, Stein JS, Hansen C, Riley D. Introduction to the open source PV LIB for python Photovoltaic system modelling package. In: 2014 IEEE 40th photovoltaic specialist conference (PVSC). 2014, p. 0170–4. http://dx.doi.org/10.1109/PVSC.2014.6925501.

[40] Maxwell EL. A quasi-physical model for converting hourly global horizontal to direct normal insolation. Tech. Rep. SERI/TR-215-3087, Golden, CO (USA): Solar Energy Research Inst.; 1987.