

# Short-term Solar Irradiance Prediction from Sky Images with a Clear Sky Model

Huiyu Gao and Miaomiao Liu  
Australian National University, Canberra, Australia  
`{huiyu.gao, miaomiao.liu}@anu.edu.au`

## Abstract

Integrating the solar power into the power grid system while maintaining its stability is essential for utilising such type of clean energy widely. It renders the solar irradiance (determining the solar power) forecasting a critical task. This paper tackles the problem of solar irradiance prediction from a history of sky image sequence. Most existing machine learning methods directly regress the solar irradiance values from a historical image sequence and/or solar irradiance observations. By contrast, we propose a novel deep neural network for short-term solar irradiance forecasting by leveraging a clear sky model. In particular, we build our network structure on the vision transformer to encode the spatial as well as the temporal information in the sky video sequence. We then aim to predict the solar irradiance residual from the learned representation by explicitly using a clear sky model. We evaluated our approach extensively on the existing benchmark datasets, such as TSI880 and ASI16. Results on the nowcasting task, namely estimation of the solar irradiance from the observations, and the forecasting task, which is up to 4-hour ahead-of-time prediction, demonstrate the superior performance of our method compared with existing machine learning algorithms.

## 1. Introduction

Global warming and climate change have become critical issues, which are mainly due to human activities, primarily fossil fuel burning [2]. Using clean energy, especially solar energy, worldwide can reduce the carbon dioxide ( $CO_2$ ) emissions to the atmosphere which is considered as one of the major sources for global warming [1]. While solar power can make the energy system resilient, it is unstable which is largely influenced by real-time weather conditions. Furthermore, the usage of the household roof solar panel may lead to the risk of damaging the power grid infrastructure due to the exporting of excess household solar power to the grid. To maintain the stability of the power grid system, solar irradiance forecasting, namely prediction of the solar irradiance for the future short-term and long-

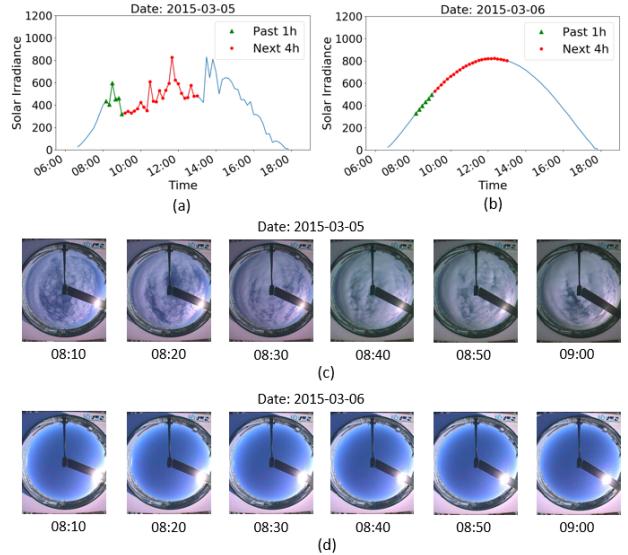


Figure 1. Examples of solar irradiance variations on a cloudy day (a) and a clear day (b). It shows that the shielding of clouds causes solar irradiance to drop compared with that on clear days. Given the observed sky images (c), and (d) for the past 1 hour (see the green triangles in (a), (b) respectively), we aim to forecast the solar irradiance up to 4 hours ahead-of-time (see red dots in (a), (b)).

term horizon from historical observations, is beneficial for solving the problem.

Traditional methods, such as Autoregressive Moving Average [15], and Autoregressive Integrated Moving Average [19], are proven effective in modelling historical solar irradiance sequence only even with limited data. However, solar irradiance is highly influenced by real-time weather conditions. Thus, the most recent common trend in short-term solar irradiance forecasting consists of modelling the sky image video sequences using deep learning models such as long short-term memory (LSTM) for future prediction [21]. Specifically, to reduce the memory requirement in modelling the image sequence, each image in the sequence is encoded as a feature vector by a pretrained nowcasting model, which is a framework to estimate the solar irradiance output at the time instance with image observation.

Then the temporal encoding is achieved via LSTM on the extracted image feature sequences.

Unlike [21], we propose a unified framework based on the encoder of the Vision Transformer (ViT) [10] for feature extraction as well as temporal encoding. Such network structure allows us to encode the long-range spatial information within an image and the temporal information of the video sequence across the long-term horizon.

Similar to [21], we explore the clear sky model which models the solar irradiance values on a clear day. Instead of treating the solar irradiance value from the clear sky model as an extra dimensional feature only, we explicitly predict the residual between the true irradiance value and the one from the clear sky model. Examples of sky images and their corresponding solar irradiance on a cloudy day and a clear day are shown in Fig. 1. As evidenced by our experiments on TSI880 and ASI16, our method shows superior performance.

Our contributions can be summarised as follows.

- We introduce a transformer-based solar irradiance prediction model that can encode the spatial and temporal information in a unified structure.
- Our explicit residual prediction of the solar irradiance using a clear sky model benefits the short-term solar irradiance forecasting.
- It yields the state-of-the-art performance on the now-casting task and shows superior performance on forecasting tasks.

## 2. Related work

**Traditional Methods** Modelling variations in solar irradiance has been studied for decades. In the early stage of the solar energy field, the research is hindered by lacking solar irradiance data collection devices. Therefore, the indirect prediction methods are explored to predict solar irradiance from sunshine duration, diffuse radiation, temperature, and relative humidity. The atmospheric information, such as temperature and relative humidity, has been proven correlated with solar irradiance on clear days. The model proposed by Udagawa [3] achieves the prediction by exploring the relationship between the height of the sun and solar irradiance. Although it adopts limited data and lacks time representation, it still lays a foundation for future research.

Direct prediction models predict the values of future solar irradiance from the former solar irradiance values. More and more facilities are dedicated to collecting solar irradiance data to provide effective support for research purposes [12, 22]. There are several kinds of traditional prediction methods, such as regression model, trend extrapolation, and support vector machine (SVM) [28].

Time series approaches primarily aim at the modelling of long-term solar irradiance forecast, which includes Moving Average, Autoregressive [6], Autoregressive Moving Average [15], and Autoregressive Integrated Moving Average [19] models. These time series models rely on historical solar irradiance only, ignoring the relevant meteorological factors. Besides, it can merely capture linear relationships and require stationary input data. Since these traditional methods have limitations such as manual feature extraction and cannot handle non-linear problems [9], machine learning provides a new approach to this task.

**Machine Learning Methods** Machine learning models, such as Artificial Neural Network (ANNs) [17], have been widely used for solar irradiance forecasting tasks. Wang *et al.* [25] propose a backpropagation neural network, which is a simple fully connected neural network, to perform a short-term solar irradiance forecast. Inputs and outputs of this model are both solar irradiance with an interval of one hour. The simple model can only provide short-term prediction. It could not meet the performance requirements under complex weather conditions. After that, ensemble methods have been introduced, such as Random Forests [7] and Extreme Gradient Boosting [8].

To improve real-time forecasting accuracy and reduce the possible negative effects of photovoltaic systems, Yang *et al.* [27] propose a weather-based hybrid method for 1-day ahead hourly forecasting of solar power output with the application of Self-organizing Map, Learning Vector Quantization, and Support Vector Regression. There is an obvious time series in input data, we can use machine learning algorithms to learn the patterns between these series.

Recent developments in deep learning, particularly in Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM), provide approaches to solve this problem. Qing and Niu [20] use LSTM to catch the dependence between hourly solar irradiance values over a day. Gensler *et al.* [11] employ an auto-encoder architecture to reduce the dimension of historical data, and apply LSTM to forecast the solar irradiance. Nonetheless, these methods just utilise solar irradiance to perform forecast, but do not include sky image features, which can well indicate the current weather conditions and play an important role in forecasting solar irradiance.

**Sky Images** On cloudy days, solar irradiance fluctuate greatly at a small level, which makes the forecast of solar irradiance more challenging. Wolff *et al.* [26] propose a method to track the movement of clouds by using digital image processing technology and calculating the displacement of clouds in the sky images. Alonso-Montesinos *et al.* [4], as well as Kurtz and Kleissl [16] design networks that first predict the cloud distributions base on linear extrapolation, then calculate the solar irradiance according to the predicted distributions of clouds. Zhang *et al.* [29] ex-

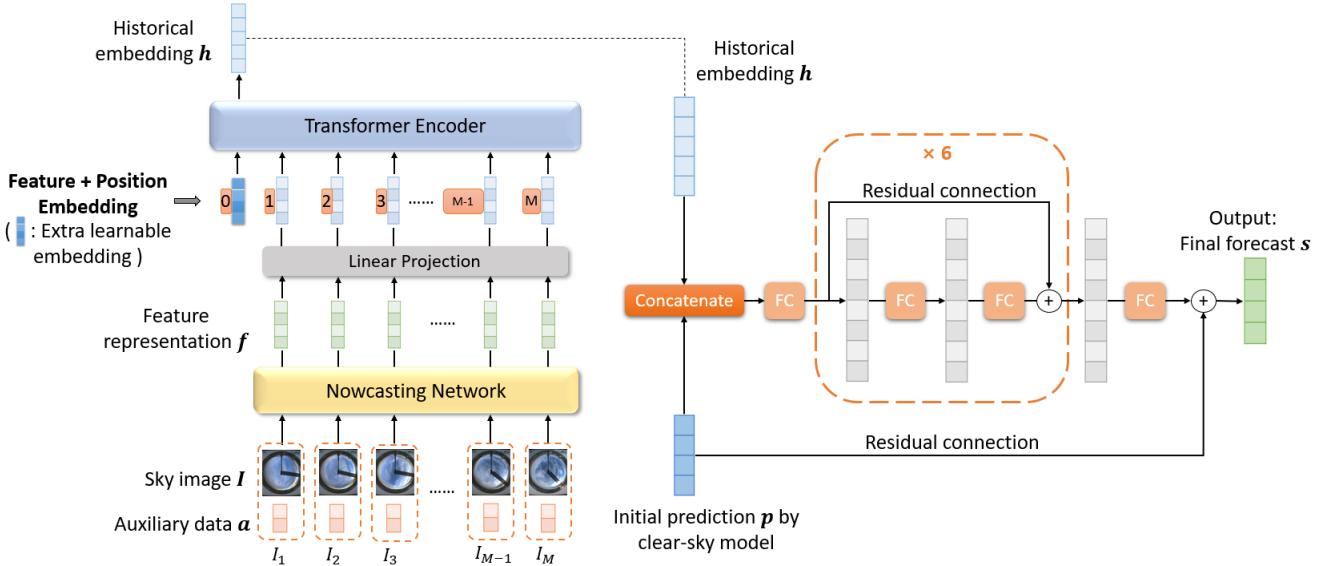


Figure 2. The overview of our transformer-based forecasting network. Given a sequence of historical sky images and their corresponding auxiliary weather data for the past  $M$  frames, this model forecasts solar irradiance up to  $T$  time steps ahead-of-time.

plore different architectures to estimate the variation of solar irradiance from current and previous solar irradiance values. However, their model can only predict one time step and is more suitable for ultra-short-term prediction tasks, such as one minute.

Recently, Siddiqui *et al.* [21] employ a nowcasting model consisting of convolutional layers to extract the features of each sky image followed by using an LSTM to encode the temporal information for forecasting solar irradiance for the next 4 hours. It additionally leverages the clear sky model as an additional feature for solar irradiance forecasting. By contrast, we explicitly explore the clear sky model and predict the residual which is proven effective in the forecasting task.

### 3. Our Approach

Let us now introduce our approach to solar irradiance forecasting. Let  $\mathbf{I}_{1:M}$  be a sequence of  $M$  observed sky images  $\mathbf{I}_{1:M} = \{I_i\}_{i=1}^M$ , where  $I_i$  represents the  $i^{th}$  historical image frame. Assume their corresponding auxiliary data  $\mathbf{A}_{1:M} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_M]$ , where  $\mathbf{a}_m \in R^k$  represents  $k$  weather features such as air temperature, relative humidity, wind speed, and air pressure, etc. Our goal is to forecast a sequence of solar irradiance  $\mathbf{s}_{M+1:M+T} = [s_{M+1}, s_{M+2}, \dots, s_{M+T}]$ , where  $s_t \in R$ ,  $t \in \{M + 1, \dots, M + T\}$ , for the next  $T$  time steps. In our experiments below, we are focusing on short-term solar irradiance prediction by observing sky images in the past 1 hour ( $M = 6$ ) to forecast the solar irradiance up to 4 hours ahead-of-time ( $T = 24$ ).

An overview of our framework is shown in Fig. 2. It mainly consists of three modules, namely the feature extraction module, the temporal encoding module, and the solar irradiance residual prediction module. Below, we discuss these modules in detail.

#### 3.1. Nowcasting Network for Feature Extraction

To obtain  $\mathbf{s}_{M+1:M+T}$  from  $\mathbf{I}_{1:M}$ , the common practice is to build an end-to-end trainable network. While it is likely to yield the best performance due to the end-to-end training, the memory cost will grow significantly with the increase of the image resolution. Therefore, instead of treating the image video sequence as the input to the forecasting network directly, we propose to perform the prediction from the feature of each historical image and auxiliary data extracted from a nowcasting network. Specifically, the nowcasting network is designed to estimate the solar irradiance for each time instance from the observed data. Similar to [21], our nowcasting network takes the observed sky image  $I_t$  as well as the auxiliary information, namely the meteorological information  $\mathbf{a}_t$  which includes wind speed, air temperature, and relative humidity, as input to estimate the corresponding solar irradiance  $s_t$  at time  $t$ .

Our nowcasting network largely follows the most recent transformer-based framework, namely Vision Transformer (ViT) [10], to encode the input information. The network structure is shown in Fig. 3. Given a sky image  $I \in \mathbb{R}^{H \times W \times C}$ , where  $(H, W)$  is the resolution of the input image and  $C$  is the number of channels, we first split it into a sequence of  $N$  patches of size  $(P, P)$ , where

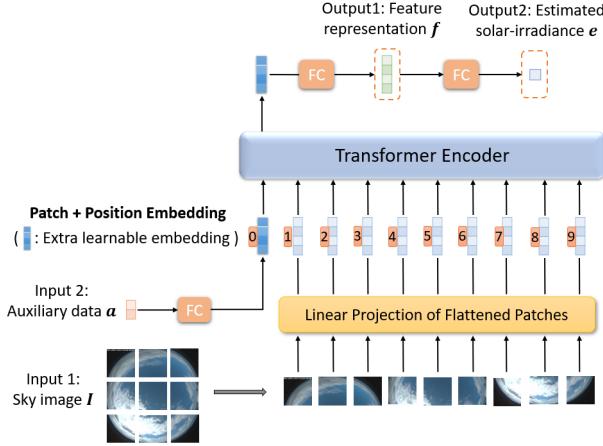


Figure 3. Transformer-based nowcasting network

$N = HW/P^2$ . Then each image patch is flattened into a vector  $\mathbf{x}_n \in R^{P^2 \cdot C}$ ,  $n = \{1, 2, \dots, N\}$ .  $\{\mathbf{x}_n\}_{n=1}^N$  are then further embedded into  $D$  dimensions with a trainable linear projection to obtain the patch embeddings.

Similar to the Vision Transformer (ViT) [10], we put an extra embedding  $\mathbf{x}_0$  at the beginning of the sequence. This extra embedding  $\mathbf{x}_0$  is a learnable mapping from the input auxiliary data  $\mathbf{a}$  which concatenates the information such as air temperature, relative humidity, wind speed, air pressure, azimuth angle, zenith angle, and an estimated solar irradiance value from a clear sky model [13]. Since these auxiliary parameters have different ranges, each component is normalized by subtracting the mean and then dividing by its standard deviation before feeding them into the network.

**Transformer Encoder Module.** Following ViT [10], we encode the image as well as the auxiliary data using the transformer encoder module (see Fig. 4). To be self-contained, we review the transformer encoder module below. This module consists of  $L$  transformer layers, each of which is further composed of a multi-head self-attention (MSA) and a multilayer perceptron (MLP) sub-layer. For each sub-layer, a layernorm (LN) [5] is applied before it and a residual connection [14] is employed around it. The MSA layer is depicted in the right column of Fig. 4, which mainly follows the classical structure proposed by Vaswani *et al.* [24]. An attention mechanism is applied to a query matrix  $Q$  and a set of key-value pairs, which are packed together into a key matrix  $K$  and a value matrix  $V$ . They are essentially the transformed image and auxiliary data embeddings. This can be expressed as

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

where  $d_k$  is the dimension of queries and keys and  $\frac{1}{\sqrt{d_k}}$  is a scaling factor of the dot-product.

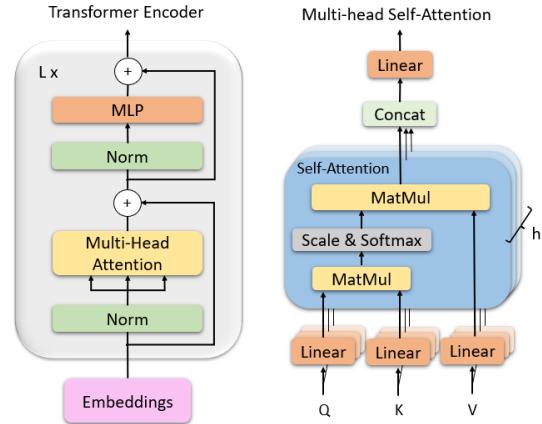


Figure 4. (left) Transformer encoder module [10, 24] (right) Multi-head self-attention structure [24]

In addition, we also apply the multi-head strategy [24] in which the queries  $Q$ , keys  $K$  and values  $V$  are linearly projected  $h$  times with different, learnable matrices  $W_i^Q \in R^{D \times d_k}$ ,  $W_i^K \in R^{D \times d_k}$ ,  $W_i^V \in R^{D \times d_k}$ ,  $i = \{1, 2, \dots, h\}$ . For each projection, we perform the attention mechanism to the projected  $QW_i^Q$ ,  $KW_i^K$ ,  $VW_i^V$  in parallel and regard each of them as an attention head. The outputs from  $h$  self-attention blocks are concatenated and then projected by  $W^O \in R^{hd_k \times D}$  to generate the final representation. That is

$$\text{MSA}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \quad (2)$$

where  $\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$ . Following [10], the formulation of these  $L$  layers of transformation are provided below

$$\begin{aligned} \mathbf{z}_0 &= [\mathbf{x}_0; \mathbf{x}_p^1 \mathbf{E}; \mathbf{x}_p^2 \mathbf{E}; \dots; \mathbf{x}_p^N \mathbf{E}] + \mathbf{E}_{\text{pos}} \\ \mathbf{z}'_\ell &= \text{MSA}(\text{LN}(\mathbf{z}_{\ell-1})) + \mathbf{z}_{\ell-1}, \\ \mathbf{z}_\ell &= \text{MLP}(\text{LN}(\mathbf{z}'_\ell)) + \mathbf{z}'_\ell, \\ \mathbf{f} &= \text{LN}(\mathbf{z}_L^0) \end{aligned} \quad (3)$$

where  $\ell = \{1 \dots L\}$ ,  $\mathbf{E} \in R^{(P^2 \cdot C) \times D}$  and  $\mathbf{E}_{\text{pos}} \in R^{(N+1) \times D}$  represents the position embeddings. We can obtain the image and auxiliary data representation  $\mathbf{f}$  after the encoding process by the transformer.

### 3.2. Transformer-based Temporal Encoding for Forecasting

We extract the representation for the historical observation from the nowcasting network. The observed image sequence as well as the auxiliary data are represented as  $\mathbf{f}_{1:M} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_M]$ , where  $\mathbf{f}_i \in R^d$ ,  $i \in \{1 \dots M\}$ . Note that the nowcasting network can be trained offline. The extracted feature is a low dimensional representation of the image which alleviates the memory constraint on the

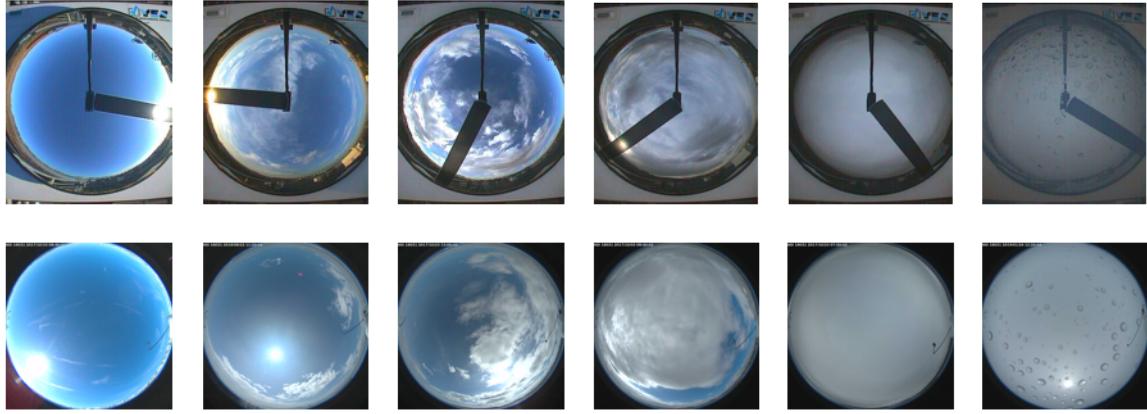


Figure 5. Sky images under different weather conditions from TSI880 dataset (top) and ASI16 dataset (bottom).

forecasting network. Given  $\mathbf{f}_{1:M}$ , we adopt a transformer module for the temporal encoding. Similarly, we introduce an extra embedding  $\mathbf{f}_0$  as a global representation for this observed image sequence. The transformer module explores the long-range dependencies among the observed data and encodes the feature sequence to a sequence embedding  $\mathbf{h}$  for the forecasting task.

### 3.3. Residual Forecasting with a Clear Sky Model

Instead of predicting the solar irradiance for the next  $T$  frames from the sequence embedding  $\mathbf{h}$  directly, we propose to learn the residual solar irradiance relative to an initial prediction  $\mathbf{p} \in \mathbb{R}^T$  that is calculated from a simple clear sky model [13]. This model defines the value of solar irradiance under a cloudless day as

$$\text{GHI} = 1098 \times \cos(z) \times \exp\left(\frac{-0.057}{\cos(z)}\right) \quad (4)$$

where GHI denotes the Global Horizontal Irradiance and  $z$  represents the solar zenith angle. For each time step  $t$  in the next  $T$  frames, we calculate the GHI value using the solar zenith angle  $z_t$  at time  $t$ .

We learn the solar irradiance residual with respect to  $\mathbf{p}$  from the features extracted from historical sequence embedding  $\mathbf{h}$  as well as  $\mathbf{p}$  via a network consisting of a few MLP layers with residual connections.

### 3.4. Training

Let us now introduce the loss function,  $l_n$  and  $l_f$ , we use to train our nowcasting and forecasting model, respectively.

$$l_n = \frac{1}{N} \sum_{i=1}^N |\hat{s}^i - s^i| \quad (5)$$

$$l_f = \frac{1}{N \cdot T} \sum_{i=1}^N \sum_{j=1}^T |\hat{s}_{M+j}^i - s_{M+j}^i| \quad (6)$$

where  $\hat{s}^i$  denotes the estimated solar irradiance by the nowcasting model,  $s^i$  is the ground truth for the  $i$ th sample, and  $\hat{s}_{M+j}^i$  is the predicted solar irradiance for the next  $j^{th}$  time step.

Both our nowcasting network and forecasting network are trained using SGD optimizer with batch size of 32, learning rate as 0.001, and a momentum of 0.9. For the nowcasting network, the input image is of size  $224 \times 224$  and the patch is of size  $(16, 16)$ . We set the number of transformer layers as 12 and the input dimension  $D$  as 768. In each transformer layer, the number of heads  $h$  is chosen as 12, thus the projected vector dimension  $d_k$  in a single attention block becomes  $d_k = D/h = 768/12 = 64$ . The dimension of the feature representation  $\mathbf{f}$  extracted from the nowcasting network for each image is set as 64. In the forecasting network, the number of residual blocks is 6 and the hidden dimension is 512.

## 4. Experimental Results and Analysis

Following the previous work [21], we evaluate our method on TSI880 and ASI16, two publicly available datasets. In the following, we discuss these datasets, evaluation metrics, the baseline approach and present our results.

### 4.1. Datasets

The TSI880 and ASI16 are publicly benchmark datasets built by Solar Radiation Research Laboratory (SRRL) [23] of the National Renewable Energy Laboratory (NREL) [18]. Since 1981, SRRL has been continuously collecting research-quality measurements of solar radiation and surface meteorological parameters using the installed Baseline Measurement System (BMS). BMS has more than 80 meteorological sensors, such as pyranometers, pyrheliometers, pyrgeometers, anemometers, etc.

**TSI880** SRRL has employed a NREL-built All-Sky camera with  $180^\circ$  fish eye lens, named TSI-880, to capture the

Model	TSI880 2015	TSI880 2016	ASI16 2020
Siddiqui <i>et al.</i> (wo) [21]	15.9	17.3	14.0*
Siddiqui <i>et al.</i> [21]	14.6	15.7	13.1*
Ours(wo)	8.2	8.4	8.2
Ours	<b>7.7</b>	<b>7.7</b>	<b>6.9</b>

Table 1. nMAP errors (%) for nowcasting on TSI880 and ASI16 dataset (wo - without the auxiliary input, \* - reproduced results).

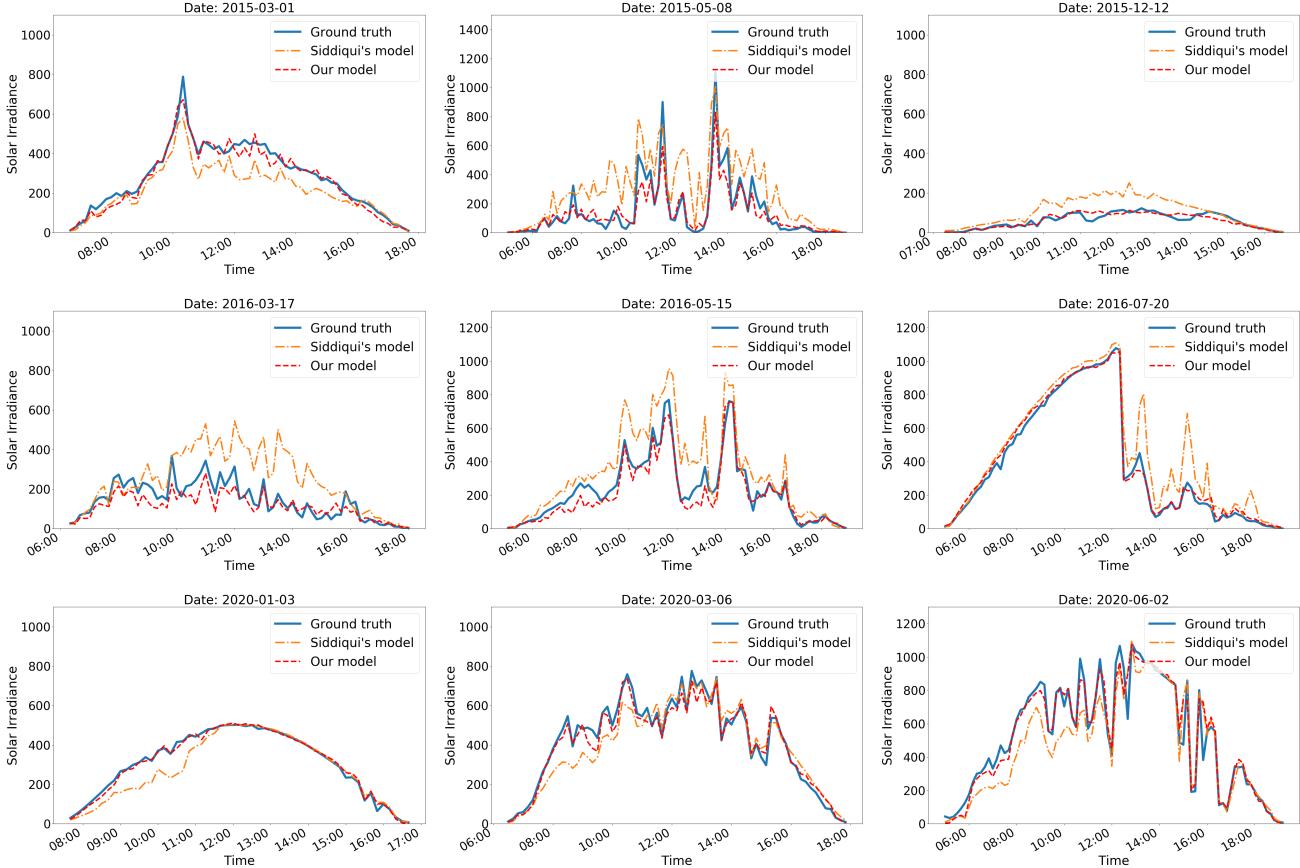


Figure 6. Sampled nowcasting results of solar irradiance on TSI880 2015 dataset (first row), TSI880 2016 dataset (second row), ASI16 2020 dataset (third row) by Siddiqui's model [21] and our model.

sky images every 10 minutes and archive them to disk since July 14, 2004. We show example sky images in the first row of Fig. 5. In TSI880, a mechanical sun tracker is adopted to block the sun and prevent image saturation. Following [21], we train our network on 2004-2014 data and test on 2015-2016 data for both nowcasting and forecasting tasks. For the nowcasting dataset, we match the sky images (taken every 10 minutes) with the auxiliary data in the BMS data stream (refreshed every minute), and the resulting interval of each sample is 10 minutes. It results in 260,754 and 52,707 data pairs for training and testing, respectively. For the forecasting task, the network takes 6 consecutive sky images consisting of one hour observation as input to forecast the

future 24 solar irradiance values which are the 4 hour ahead-of-time prediction.

**ASI16** ASI16 is a dataset of images captured by a high quality camera, ASI-16, every 10 minutes since September 26, 2017 without a sun tracker. As shown in the second row of Fig. 5, the sky images under different weather conditions in the ASI16 dataset have better quality than those in TSI880 due to its wider field of view. For nowcasting, we use sky images captured between 2017 and 2019 (51,091 frames in total) for training and 2020 (25,043 frames) for testing. For forecasting, similar to TSI880, the model is trained by taking the past 6 frames as input to predict the future 24 solar irradiance values.

Model	TSI880 2015				TSI880 2016				ASI16 2020			
	+ 1 h	+ 2 h	+ 3 h	+ 4 h	+ 1 h	+ 2 h	+ 3 h	+ 4 h	+ 1 h	+ 2 h	+ 3 h	+ 4 h
Siddiqui <i>et al.</i> [21]	<b>17.9</b>	<b>25.2</b>	31.6	39.1	<b>16.9</b>	25.0	31.9	39.5	21.6*	27.9*	33.0*	36.9*
Ours	22.6	26.3	<b>30.1</b>	<b>33.7</b>	19.9	<b>23.8</b>	<b>27.2</b>	<b>30.7</b>	<b>17.4</b>	<b>20.9</b>	<b>25.1</b>	<b>29.2</b>

Table 2. nMAP errors (%) for forecasting the solar irradiance for the next 4 hours ( $T = 24$ ) on TSI880 and ASI16 dataset (\* - reproduced result). +1h, +2h, +3h and +4h represent the nMAP error evaluated at the next 6th, 12th, 18th and 24th time step, respectively.

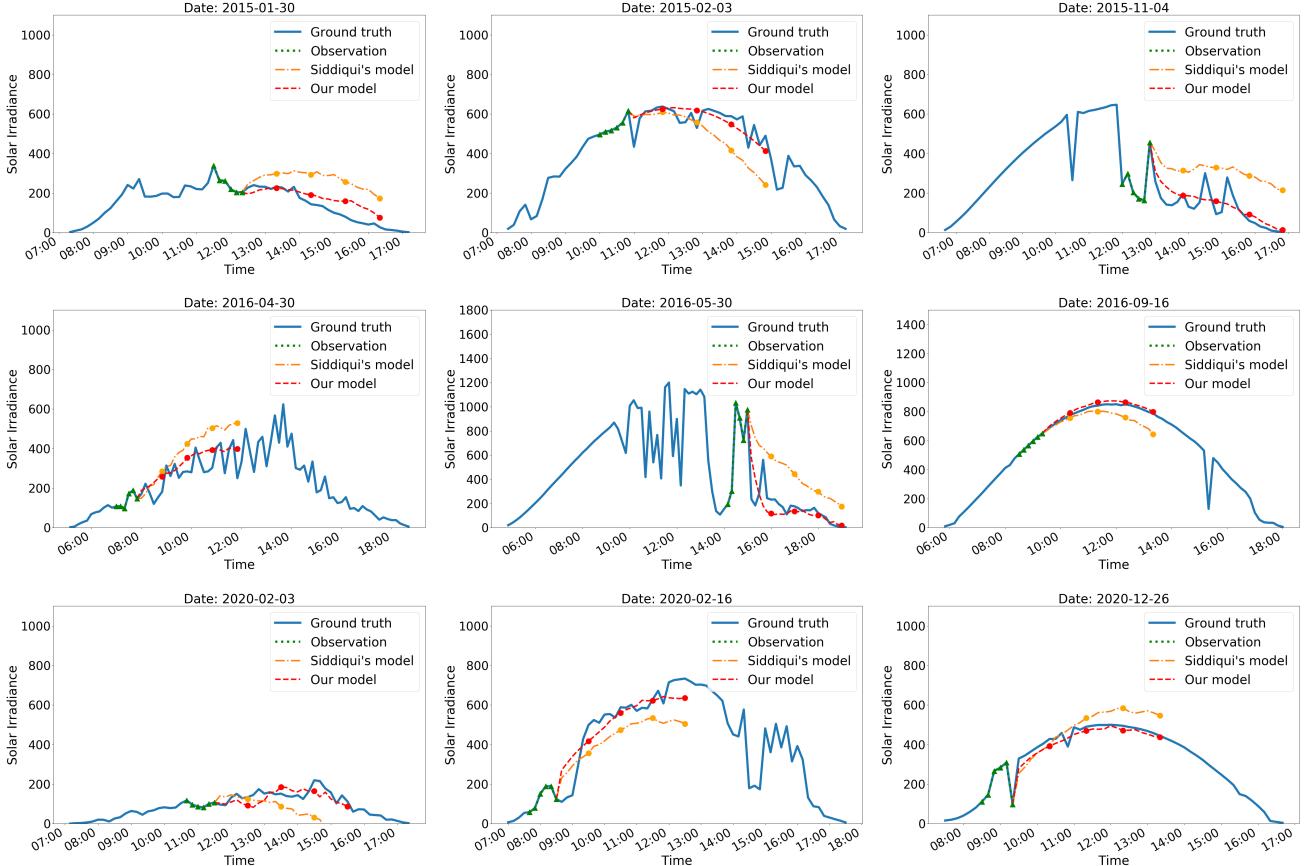


Figure 7. Sampled forecasting results for the next 4 hours on TSI880 2015 dataset (first row), TSI880 2016 dataset (second row), ASI16 2020 dataset (third row) by Siddiqui's model [21] and our method. The four dots shown on each forecasting curve represent the predicted value exactly at +1h, +2h, +3h, +4h.

## 4.2. Evaluation Metric and Baseline

We evaluate the performance of our nowcasting and forecasting models using the normalized mean absolute percentage (nMAP) error defined below,

$$nMAP = \frac{1}{N} \sum_{i=1}^N \frac{|g_i - s_i|}{\frac{1}{N} \sum_{i=1}^N g_i} \times 100 \quad (7)$$

where  $g_i$  is the ground truth and  $s_i$  is the predicted solar irradiance value for the  $i$ th sample,  $i \in \{1, \dots, N\}$ . It is commonly used in solar irradiance forecasting. We compare our approach with [21] which consists of the state of the art.

## 4.3. Results

**TSI880** The nowcasting results on TSI880 dataset are shown in Tab. 1. Compared with [21], our transformer-based nowcasting framework, with auxiliary input, reduces the nMAP error by 6.9% and 8% on the 2015 and 2016 datasets, respectively. The qualitative results shown in Fig. 6 further demonstrate the performance of our approach. In Tab. 2, we report the nMAP errors for the forecasting task on TSI880 dataset. Our proposed method achieves a smaller forecasting error for 3-hour and 4-hour prediction. Specifically, our approach achieves a 5.4% reduction on TSI880 2015 and a 8.8% reduction on TSI880 2016 for 4-hour pre-

Model	TSI880 2015				TSI880 2016				ASI16 2020			
	+ 1 h	+ 2 h	+ 3 h	+ 4 h	+ 1 h	+ 2 h	+ 3 h	+ 4 h	+ 1 h	+ 2 h	+ 3 h	+ 4 h
Clearsky only	39.6	41.3	44.0	46.8	32.6	34.5	36.9	39.6	27.3	29.4	32.8	36.8
Transformer only	23.9	29.3	34.8	40.5	21.3	26.8	31.8	38.4	18.4	22.9	28.2	34.1
Ours	<b>22.6</b>	<b>26.3</b>	<b>30.1</b>	<b>33.7</b>	<b>19.9</b>	<b>23.8</b>	<b>27.2</b>	<b>30.7</b>	<b>17.4</b>	<b>20.9</b>	<b>25.1</b>	<b>29.2</b>

Table 3. nMAP errors (%) for forecasting with and without a clear sky model on TSI880 and ASI16 dataset.

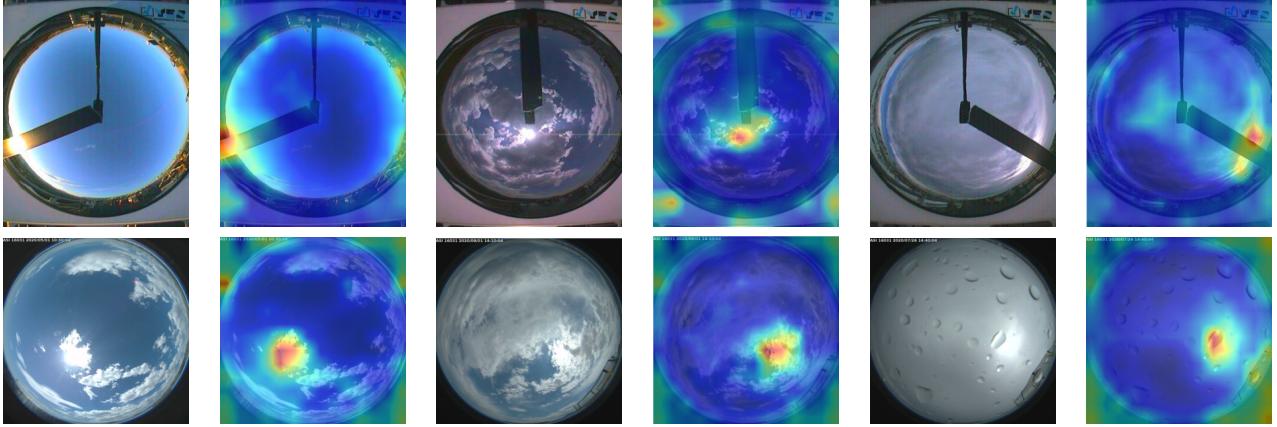


Figure 8. Visualization of sky images as well as their corresponding attention maps by our transformer-based nowcasting network on TSI880 dataset (top) and ASI16 dataset (bottom).

diction. The qualitative comparisons for the forecasting task are shown in Fig. 7. In these figures, the green curves depict the past 1 hour observation, while the orange and red curves represent the forecasting results of solar irradiance for the next four hours by Siddiqui’s method [21] and ours, respectively. It shows that our approach outperforms [21], especially for the next 3-hour and 4-hour prediction, under various weather conditions (reflected by the fluctuation of the solar irradiance curve across the day).

**ASI16** As the source code of [21] is not publicly available, we report the results of [21] on ASI16 based on our own implementation which follows the descriptions of their model in the paper. The results for nowcasting and forecasting tasks are shown in Tab. 1 and Tab. 2, respectively. Our approach achieves a 6.2% error reduction for nowcasting when using sky images to estimate solar irradiance. As illustrated in the third row of Fig. 6, our transformer-based nowcasting network can outperform [21] even under challenging weather conditions. For the forecasting task, our model consistently achieves 4.2%, 7%, 7.9% and 7.7% improvements over [21] for the next 4 hours.

**Ablation Study** To show the influence of the explicit residual prediction with a clear sky model, we train the model with and without the residual learning. Results in Tab. 3 show that learning the residual of the solar irradiance under a clear day indeed benefit the forecasting performance consistently. By explicitly leveraging the clear sky model, the nMAP errors are reduced by 1.3%, 3%, 4.7%,

and 6.8% on TSI880 2015 and 1.4%, 3%, 4.6% and 7.7% on TSI880 2016 for the next four hours forecasting, respectively. For ASI16 2020 dataset, our model also achieves consistent improvement for next four hours solar irradiance forecasting task.

**Visualisation of Attention** In Fig. 8, we visualise sampled attention maps from our transformer-based nowcasting network on TSI880 and ASI16 dataset respectively. The example images cover different weather conditions such as sunny, cloudy, and rainy days. The attention map shows that the attention model can find the most relevant pixels spatially within the image which contributes mostly to the solar irradiance estimation. In particular, the highlighted regions mostly focus on the sun regions in the image. The errors in Fig. 8 are mainly due to the insufficient number of training data which leads to the imperfect attention maps.

## 5. Conclusions

In this paper, we demonstrate that explicitly predicting the residual of the solar irradiance relative to initial values defined by a clear sky model will benefit fore-casting task. In the future, we will explore next frame image prediction and generation for long-term solar irradiance forecasting.

## 6. Acknowledgment

This research was supported by the (ARC) fellowship (DE180100628).

## References

- [1] The cause of climate change. <https://climate.nasa.gov/causes/>.
- [2] Overview: weather, global warming and climate change. <https://climate.nasa.gov/resources/global-warming-vs-climate-change/>.
- [3] Song Aiguo. General situation and analysis of model of solar radiation in japan. *Journal of capital normal university*, page S1, 1997.
- [4] J Alonso-Montesinos, FJ Batlles, and C Portillo. Solar irradiance forecasting at one-minute intervals for different sky conditions using sky camera images. *Energy Conversion and Management*, 105:1166–1177, 2015.
- [5] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.
- [6] Peder Bacher, Henrik Madsen, and Henrik Aalborg Nielsen. Online short-term solar power forecasting. *Solar energy*, 83(10):1772–1783, 2009.
- [7] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [8] Tianqi Chen, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, et al. Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4), 2015.
- [9] Xiao-nan CHEN, Xiao-ke XU, and Ji-dong SUO. Forecasting method of chaotic time series based on rbf. *Journal of Dalian Maritime University*, 1, 2007.
- [10] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [11] André Gensler, Janosch Henze, Bernhard Sick, and Nils Raabe. Deep learning for solar power forecasting—an approach using autoencoder and lstm neural networks. In *2016 IEEE international conference on systems, man, and cybernetics (SMC)*, pages 002858–002865. IEEE, 2016.
- [12] Fang Wei Yu Bingxi Yao Haishun and Li Zhe Gong Chenghu Jin Xifeng. Solar irradiance absolute radiometer and international comparison [j]. *Acta Optica Sinica*, 1, 2003.
- [13] Bernhard Haurwitz. Insolation in relation to cloudiness and cloud density. *Journal of Atmospheric Sciences*, 2(3):154–166, 1945.
- [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [15] Rui Huang, Tiana Huang, Rajit Gadh, and Na Li. Solar generation prediction using the arma model in a laboratory-level micro-grid. In *2012 IEEE third international conference on smart grid communications (SmartGridComm)*, pages 528–533. IEEE, 2012.
- [16] Benjamin Kurtz and Jan Kleissl. Measuring diffuse, direct, and global irradiance using a sky imager. *Solar Energy*, 141:311–322, 2017.
- [17] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.
- [18] NREL. NREL transforming energy. National Renewable Energy Laboratory (NREL). <https://www.nrel.gov/>.
- [19] R Perdomo, E Banguero, and G Gordillo. Statistical modeling for global solar radiation forecasting in bogotá. In *2010 35th IEEE Photovoltaic Specialists Conference*, pages 002374–002379. IEEE, 2010.
- [20] Xiangyun Qing and Yugang Niu. Hourly day-ahead solar irradiance prediction using weather forecasts by lstm. *Energy*, 148:461–468, 2018.
- [21] Talha Ahmad Siddiqui, Samarth Bharadwaj, and Shivkumar Kalyanaraman. A deep learning approach to solar-irradiance forecasting in sky-videos. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 2166–2174. IEEE, 2019.
- [22] Adnan Sözen, Erol Arcaklıoğlu, Mehmet Özalp, and Naci Çağlar. Forecasting based on neural network approach of solar potential in turkey. *Renewable Energy*, 30(7):1075–1090, 2005.
- [23] SRRL. SRRL BMS daily plots and raw data files. NREL Solar Radiation Research Laboratory (SRRL). <https://midcdmz.nrel.gov/apps/sitehome.pl?site=BMS>.
- [24] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017.
- [25] Zhe Wang, Fei Wang, and Shi Su. Solar irradiance short-term prediction model based on bp neural network. *Energy Procedia*, 12:488–494, 2011.
- [26] Björn Wolff, Jan Kühnert, Elke Lorenz, Oliver Kramer, and Detlev Heinemann. Comparing support vector regression for pv power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data. *Solar Energy*, 135:197–208, 2016.
- [27] Hong-Tzer Yang, Chao-Ming Huang, Yann-Chang Huang, and Yi-Shiang Pai. A weather-based hybrid method for 1-day ahead hourly forecasting of pv power output. *IEEE transactions on sustainable energy*, 5(3):917–926, 2014.
- [28] Jianwu Zeng and Wei Qiao. Short-term solar power prediction using a support vector machine. *Renewable Energy*, 52:118–127, 2013.
- [29] Jinsong Zhang, Rodrigo Verschae, Shohei Nobuhara, and Jean-François Lalonde. Deep photovoltaic nowcasting. *Solar Energy*, 176:267–276, 2018.