



Title	Enhancing intra-hour solar irradiation estimation through knowledge distillation and infrared sky images
Authors(s)	Nijhum, Ifran Rahman, Kaloni, Dewansh, Kenny, Paul, Dev, Soumyabrata
Publication date	2024-07-12
Publication information	Nijhum, Ifran Rahman, Dewansh Kaloni, Paul Kenny, and Soumyabrata Dev. "Enhancing Intra-Hour Solar Irradiation Estimation through Knowledge Distillation and Infrared Sky Images." IEEE, July 12, 2024. https://doi.org/10.1109/IGARSS53475.2024.10642288 .
Conference details	The 2024 IEEE International Geoscience and Remote Sensing Symposium (IEEE IEGARSS), Athens, Greece, 7-12 July 2024
Publisher	IEEE
Item record/more information	http://hdl.handle.net/10197/27656
Publisher's statement	© 2024 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Publisher's version (DOI)	10.1109/IGARSS53475.2024.10642288

Downloaded 2026-05-02 10:02:57

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)



© Some rights reserved. For more information

ENHANCING INTRA-HOUR SOLAR IRRADIANCE ESTIMATION THROUGH KNOWLEDGE DISTILLATION AND INFRARED SKY IMAGES

Ifraan Rahman Nijhum^{1,2}, *Dewansh Kalont*³, *Paul Kenny*⁴, *Soumyabrata Dev*^{1,2}

¹School of Computer Science, University College Dublin, Ireland

²The ADAPT SFI Research Centre, Dublin, Ireland

³Birla Institute of Technology And Science, Pilani, India

⁴School of Architecture, Planning and Environmental Policy, University College Dublin, Ireland

ABSTRACT

Recent years have seen increased interest in solar as a popular renewable energy source. The adoption of solar energy is directed towards increasing research interest in the incorporation of solar energy estimation and forecasting in the current system. Solar irradiance is the sun's energy incident on earth, and is dependent on many atmospheric parameters, mainly decided by clouds. The use of a Ground-based sky imaging system is the best for forecasting and estimating for near-real-time temporal spectrum. However, creating efficient solar estimation models for edge devices is still a relatively unexplored field. This paper captures the essence of bringing edge computing on devices for solar estimation on ground-based infrared sky images by proposing a novel knowledge distillation method for enhancing lightweight CNN-regression models. In our study, we leveraged a MobilenetV2-based Teacher model and transferred the knowledge into a simple student model by introducing a sigmoid-based loss in the Knowledge Distillation algorithm. The proposed solution shows an impressive reduction of the Mean Square Error (MSE) of the model from 3015.63 to 2540.67. This research advances the field of solar irradiance estimation by emphasizing the importance of creating efficient and edge device deployable models in this context.

Index Terms— Solar irradiance estimation, Knowledge distillation, CNN regression, Infrared sky images

1. INTRODUCTION

Solar energy is the largest among natural resources in the context of the global energy system [1]. Awareness regarding fossil fuel consumption and climate change is rising around the world, consequently, clean and renewable energy sources such as solar power have been widely accepted [2]. However, solar energy is intermittent and variable in nature, which

introduces significant hurdles to maintaining stable and cost-effective power generation [3]. Thus, precise early forecasting of solar power generation can play a crucial role in enabling grids to prepare in advance. This capability can facilitate the seamless integration of solar power into power grids.

Photovoltaic power generation significantly depends on solar irradiance. The amount of energy received from the sun per unit area is known as solar irradiance [4]. It largely depends on the dynamic changes of cloud coverage over the PV facilities[5]. Thus, instead of relying on historical irradiance data only, integrating solar sky images facilitates better forecasting of solar irradiance. Images captured with infrared technology provide additional information such as temperature variations which are useful for capturing the fluctuations in solar irradiance caused by varying cloud cover [4]. Integrating infrared image data with other sky condition information from multiple sensors proves to be a valuable approach for enhancing forecasting accuracy and reducing power grid instabilities [4, 6, 7].

The advances in deep learning have propelled a transformative evolution in image analysis and computer vision, contributing significantly to different fields and solar estimation is one of them. Recently, many image-based solar irradiance forecasting models have been proposed [8, 9], in addition to open-access sky image datasets [6, 10, 11]. These facilitate the constant development of sky image-based solar power forecasting. However, highly accurate deep learning models are typically complex and require a long time to infer. Thus, while employing neural network models in solar irradiance prediction may help gain highly accurate insights, they may also lead to difficulties in gaining real-time information.

To achieve higher accuracy while utilizing less computation power and resources, there have been many studies on model compression [12, 13, 14, 15]. Knowledge distillation [15] is a popular model compression technique, where a simple model can gain the knowledge of a highly accurate model. However, knowledge distillation is mostly applied to classification problems, whereas solar irradiance estimation is a CNN-regression problem. There have been few works on

This research was conducted with the financial support of Science Foundation Ireland under Grant Agreement No. 13/RC/2106_P2 at the ADAPT SFI Research Centre at University College Dublin. ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology, is funded by Science Foundation Ireland through the SFI Research Centres Programme.

incorporating knowledge distillation in regression problems [16, 17]. Distinct works have been published on estimating and forecasting solar power and irradiance. However, to the best of our knowledge, none of them propose a knowledge-distillation approach to enhance a lightweight neural network model for solar irradiance estimation. Thus, exploring efficient solutions for real-time solar information retrieval from sky images is still a challenge.

In this research¹, we introduce a novel knowledge distillation approach for intra-hour solar irradiance estimation by incorporating infrared sky images to enhance a lightweight CNN model. The key contributions of the study are:

- We proposed a novel sigmoid-based knowledge distillation loss function tailored for regression tasks.
- We improved the accuracy of a simple CNN-regression network by incorporating our distillation approach, minimizing the MSE loss from 3015.63 to 2540.67.
- We evaluated the effectiveness of the proposed method, through experiments, training, and evaluation of our models using infrared sky images

By employing this approach, efficient deep learning models can be built for solar irradiance prediction and deployed into edge devices. In addition, the sigmoid-based loss function can be used to optimize regression models in other domains as well.

2. RELATED WORK

Over past years, as demand and awareness towards renewable energy have drastically increased, so has the research in one of the most prominent forms of renewable energy, i.e. solar. A major persistent issue with current usage is irregularities and abrupt fluctuations in solar energy. This has led researchers to focus on solar irradiance estimation and forecasting.

The Girasol dataset [6] is a sky imaging and global solar irradiance measurement dataset for solar irradiance and provides both visible and infrared images. The sky imaging system consisted of a longwave infrared (IR) camera, and a visible light camera with a fisheye lens for a wider range of sky view to capture. The whole system was then installed in a weatherproof enclosure mounted over a solar tracker. A later more advanced form of the dataset was created which included physical features like divergence, curl, height velocity, and temperature inclusion for specific regions using computation and infrared images.

Solar irradiance, as a problem, has been approached in several ways. One solution used a method for solar power predictions based on sky image and Machine Learning forecasting techniques. The images provided a reliable data source for

¹In the spirit of reproducible research, the code to reproduce the results of this paper can be found in https://github.com/ifran-rahman/Solar_Estimation_KD.

pre-processing steps like cloud segmentation where a UNet-based classifier performed best for the segmentation part. Using the output for DeepLab v3 gave accurate results for solar irradiance estimations. [18]

Another contribution is by presenting a deep learning based CNN-regression model using infrared sky images and MobileNet v2 for solar irradiance estimation. [4] The MobileNetV2 is used as a CNN backbone of the model with a regression head attached to it. The lightweight yet powerful MobileNet model was chosen to boost its efficiency, enabling it to recognize important patterns in data while being suitable for edge devices.

The solution to doing computation quickly and providing consistent output came from research done by [15]. This paper presented a novel approach towards solving the computationally expensive nature of ensemble models, where current-date ensemble models take an approach to use multiple models and provide average predictions. The author presented Distillation of Knowledge as an approach to compress the knowledge and learning of the ensemble into a single model, suitable for fast computation, deployment on edge devices, and less resource extensive.

The main inspiration for this paper, to do Knowledge Distillation in a regression task was published by [16]. The author proposed a Compressing Deep Neural Network model for regression tasks, where an outlier rejection loss, i.e. rejection outliers from the sample set for model prediction is applied for better training practice. The key novelty of Knowledge distillation lies in the loss function in use, the author proposed a new distillation loss specifically for the regression task and multiple-task network which gave better training in feature extraction of student models.

3. METHODOLOGY

3.1. Dataset

In our work, we explored the performance of knowledge distillation on solar irradiance estimation from infrared sky images using a CNN regression model. Infrared images provide valuable insights into the cloud's state as infrared technology can capture temperature variations, in addition to giving information in low-light situations. For our work, we used the Girasol dataset [6]. The dataset consists of irradiance values for particular timestamps, IR, and VI images, and additional data such as temperature, dew point, and atmospheric pressure. However, in this research, we employed infrared images from 8 December 2017 to 7 January 2018 and the following 7 days, as our train and test set respectively. This training and test subset consists of 46,527 and 11,574 images respectively. Each image is of size 60x80 pixels.

3.2. Knowledge Distillation

Knowledge distillation is a process of transferring the knowledge of a large powerful pre-trained model to a smaller, efficient model. The pre-trained powerful model is referred to as a “teacher model” and the smaller model is called a “student model”. The concept of knowledge distillation was proposed by Hinton [15].

For an image I , if we denote the teacher model as “T” and the student model as “S”, then $P_t = T(I)$ and $P_s = S(I)$, where P_t is the logits of the teacher model and P_s is the logits of the student model. Later, P_t and P_s values are divided by a hyper-parameter temperature (τ) to soften the values and passed to a softmax function (σ) to get softened probability distributions 3.

$$P_t = T(I) \quad (1)$$

$$P_s = S(I) \quad (2)$$

$$\sigma(Z_k, \tau) = \frac{e^{\frac{Z_k}{\tau}}}{\sum_{m=1}^n e^{\frac{Z_m}{\tau}}} \text{ for } k = 1, 2 \quad (3)$$

$$L = (1 - \alpha)L_{ce}(\sigma(P_s, Y)) + \alpha L_d(P'_t, P'_s) \quad (4)$$

The softened P_t and P_s can be addressed as P'_t and P'_s . The KL divergence [17] is used to calculate the differences between P'_t and P'_s and the distillation loss (L_d) is obtained. Meanwhile, the hard loss is measured, which is the cross-entropy loss (L_{ce}) [19] between the student model outputs and the target. Finally, the Distillation loss (D_l) is combined with the hard loss (L_{ce}) to obtain the total loss (L) as demonstrated in equation 4. Here, α is another hyperparameter, which can be fine-tuned to adjust the proportion in the model’s reliance on the distillation loss versus the hard loss.

3.3. Sigmoid-based Knowledge Distillation

Knowledge distillation is a popular model compression method in the field of machine learning classification. However, estimating solar irradiance based on sky images is a CNN-regression problem. Thus the traditional approach cannot be used here.

The sigmoid activation function plays an important role in deep learning. The non-linear characteristic of the sigmoid function enables neural networks to capture necessary patterns of complex, non-linear data. The function is defined as stated in the equation 5.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

Here, x is the input value, and $\sigma(x)$ is the output. One of the characteristics of the sigmoid function is, it transforms any

value x into a range of 0 to 1. In our sigmoid-based distillation loss, we utilized this behavior of the function.

The outputs of our CNN-regression models (teacher and student) are continuous values which are the predicted irradiance. Solar irradiance can largely vary depending on the sun and cloud’s formation in the sky.

This is why to get the dissimilarity between the teacher model’s prediction and the student model’s prediction we normalized the outputs to a common scale by applying the sigmoid function. However, as the prediction can be very small or very large, passing it directly through the sigmoid function may saturate and cause the outputs of the function to become too close to 0 or 1 eventually leading to data loss.

Thus, we divide each predicted value by a hyperparameter called temperature (τ) before sending it to the sigmoid function. Equation 6 represents this modified form of the function.

$$\sigma'(x) = \frac{1}{1 + e^{-\frac{x}{\tau}}} \quad (6)$$

$$L_{sd} = L_{mse}(P'_t, P'_s) \quad (7)$$

$$L_H = L_{mse}(P_s, Y) \quad (8)$$

The softened outputs of the sigmoid function for the teacher and student model can be addressed as P'_t and P'_s respectively. After getting the softened predictions, the mean squared error between these two is measured to obtain the distillation/soft loss 7. Meanwhile, the hard loss (L_H) is also calculated to assess the model’s prediction in comparison to the label data using the MSE loss 8.

$$L = (1 - \alpha)L_H + \alpha L_{sd} \quad (9)$$

Finally, the knowledge distillation loss is measured by combining the hard loss (L_H) and the distillation loss (L_{sd}) using the traditional approach demonstrated in the equation 9.

3.4. Model Architectures

The teacher model is inspired by the paper [4] and built based on the pre-trained MobileNetV2 model. We removed the final classification layer of the model. Subsequently, a custom regression layer is attached next to the model. The regression head has a linear layer that produces a single output value as the predicted irradiance value. The output of the MobilenetV2 backbone is passed through an adaptive average pooling and then flattened to create a feature vector. Which is later processed through the linear regression head. Table 1 illustrates a brief description of the teacher model and its layers. The teacher model has 2225153 parameters. In contrast, the student model is a fairly simple convolutional neural network. The proposed model comprises two convolutional layers, two pooling layers, and two fully connected layers. The last fully connected layer outputs a single value which is the predicted

Layer	Description
CNN Backbone	MobilenetV2 (pre-trained) without the classification layer
AdaptiveAvgPool2d	Adaptive average pool
Flatten	Flatten to one-dimensional space
LinearRegressionHead	Fully Connected layer, Output Size: 1

Table 1. Teacher Model Architecture

Layer	Description
Conv1	No. of Filters: 16, Kernel Size: 3X3, Stride: 1, Padding: 1
ReLU1	ReLU activation
MaxPool2d	Kernel Size: 2X2, Stride: 2
Conv2	No. of Filters: 32, Kernel Size: 3X3, Stride: 2, Padding: 0
ReLU2	ReLU activation
MaxPool2d	Kernel Size: 2X2, Stride: 2
FC1	Linear Layer, Output Size: 64
ReLU3	ReLU activation
FC2	Linear Layer, Output Size: 1

Table 2. Student Model Architecture

solar irradiance for that particular image. The student model has 619617 parameters.

4. EXPERIMENTS AND RESULTS

To assess the performance of the proposed distillation process we trained the teacher model and the student model with and without the distillation approach. The chosen training subset consists of 46,527 training and 11,574 test images (as stated in section 3.1). The training process involves optimizing the model’s hyperparameters.

Each model was trained for 50 epochs with a batch size of 50. The learning rate was initially set to 0.01 and a StepLR scheduler was used in conjunction with the Adam optimizer, which reduces the learning rate by a factor of 0.5 after each epoch. During training, the teacher and student model (without kd), mean squared error (MSE) was employed as the evaluation metric. During performing knowledge distillation, the pre-trained teacher model was used to obtain the teacher’s output, and the student model was trained from scratch. For distillation, we set alpha (α) = 8 and temperature (τ)=10 after doing some parameter tuning.

Table 3 demonstrates the performance of the models based on mean squared error (MSE) and root mean squared error (RMSE). The teacher model is based on the MobileNetV2 (pre-trained) backbone and thus it performed well on the data leading to an MSE loss of 1970.75 and RMSE of 43.71. We used this model as our teacher model to evaluate the sigmoid-based distillation process. We also trained the student model from scratch. It achieved the lowest MSE of 3015.63 and RMSE of 54.10. Finally, we trained the student model again from scratch but with the knowledge distillation approach. It can be seen that now, the student model achieved a significantly lower amount of loss in comparison to before

Model	No. of Paramteres	MSE	RMSE
Teacher	2225153	1970.75	43.71
Student	619617	3015.63	54.10
Student with KD	619617	2540.67	49.96

Table 3. Performance comparison of the Teacher and the Student model with and without knowledge distillation approach.

(2540.67 MSE and 49.96 RMSE).

5. CONCLUSION

In our research, we have introduced an innovative way to enhance the efficiency of intra-hour solar irradiance estimations by employing a new knowledge distillation approach specifically designed for optimizing image regression tasks. Integrating infrared sky images captured under varied cloud conditions and utilizing the MobilenetV2 as the backbone of our teacher model has been pivotal to our research.

We validate the effectiveness of the proposed method through extensive experimentation and evaluation. The method illustrates a significant enhancement of the accuracy of the simple student model in predicting solar irradiance from infrared sky images.

The lightweight architecture and fast computation of the student model make it suitable for edge deployment. This study will be a stepping stone for future works on efficient model compression research in the field of solar power prediction and forecasting. Furthermore, the proposed sigmoid-based distillation loss function can be employed in optimizing regression models around different domains.

6. REFERENCES

- [1] M. Aman, K. Solangi, M. Hossain, A. Badarudin, G. Jasmon, H. Mokhlis, A. Bakar, and S. N. Kazi, “A review of safety, health and environmental (she) issues of solar energy system,” *Renewable and Sustainable Energy Reviews*, vol. 41, pp. 1190–1204, 2015.
- [2] G. K. Singh, “Solar power generation by pv (photo-voltaic) technology: A review,” *Energy*, vol. 53, pp. 1–13, 2013.
- [3] N. Tang, Y. Zhang, Y. Niu, and X. Du, “Solar energy curtailment in china: Status quo, reasons and solutions,” *Renewable and Sustainable Energy Reviews*, vol. 97, pp. 509–528, 2018.
- [4] I. R. Nijhum, P. Kenny, and S. Dev, “Intra-hour solar irradiance estimation using infrared sky images and mobilenetv2-based cnn regression,” in *2023 IEEE 7th Conference on Energy Internet and Energy System Integration*, 2024.

- [5] E. A. Papatheofanous, V. Kalekis, G. Venitourakis, F. Tziolos, and D. Reisis, "Deep learning-based image regression for short-term solar irradiance forecasting on the edge," *Electronics*, vol. 11, no. 22, p. 3794, 2022.
- [6] G. Terrén-Serrano, A. Bashir, T. Estrada, and M. Martínez-Ramón, "Girasol, a sky imaging and global solar irradiance dataset," *Data in Brief*, vol. 35, p. 106914, 2021.
- [7] G. Terrén-Serrano and M. Martínez-Ramón, "Deep learning for intra-hour solar forecasting with fusion of features extracted from infrared sky images," *Information Fusion*, vol. 95, pp. 42–61, 2023.
- [8] S. Song, Z. Yang, H. Goh, Q. Huang, and G. Li, "A novel sky image-based solar irradiance nowcasting model with convolutional block attention mechanism," *Energy Reports*, vol. 8, pp. 125–132, 2022.
- [9] Z. Zhen, J. Liu, Z. Zhang, F. Wang, H. Chai, Y. Yu, X. Lu, T. Wang, and Y. Lin, "Deep learning based surface irradiance mapping model for solar pv power forecasting using sky image," *IEEE Transactions on Industry Applications*, vol. 56, no. 4, pp. 3385–3396, 2020.
- [10] D. Buckley, G. Swart, J. Meiring, and L. Stepp, "Ground-based and airborne telescopes," in *Proc. SPIE*, vol. 6267, 2006.
- [11] H. T. Pedro, D. P. Larson, and C. F. Coimbra, "A comprehensive dataset for the accelerated development and benchmarking of solar forecasting methods," *Journal of Renewable and Sustainable Energy*, vol. 11, no. 3, 2019.
- [12] Y. Gao, S. Miyata, and Y. Akashi, "How to improve the application potential of deep learning model in hvac fault diagnosis: Based on pruning and interpretable deep learning method," *Applied Energy*, vol. 348, p. 121591, 2023.
- [13] K. T. Chitty-Venkata, S. Mittal, M. Emani, V. Vishwanath, and A. K. Somani, "A survey of techniques for optimizing transformer inference," *Journal of Systems Architecture*, p. 102990, 2023.
- [14] J. Gou, B. Yu, S. J. Maybank, and D. Tao, "Knowledge distillation: A survey," *International Journal of Computer Vision*, vol. 129, pp. 1789–1819, 2021.
- [15] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.
- [16] M. Takamoto, Y. Morishita, and H. Imaoka, "An efficient method of training small models for regression problems with knowledge distillation," in *2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 2020, pp. 67–72.
- [17] X. Jin, B. Peng, Y. Wu, Y. Liu, J. Liu, D. Liang, J. Yan, and X. Hu, "Knowledge distillation via route constrained optimization," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1345–1354.
- [18] S. Park, Y. Kim, N. J. Ferrier, S. M. Collis, R. Sankaran, and P. H. Beckman, "Prediction of solar irradiance and photovoltaic solar energy product based on cloud coverage estimation using machine learning methods," *Atmosphere*, vol. 12, no. 3, p. 395, 2021.
- [19] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," *Advances in neural information processing systems*, vol. 31, 2018.