# CLOUDUNET: ADAPTING UNET FOR RETRIEVING CLOUD PROPERTIES

*Zahid Hassan Tushar[1], Adeleke Ademakinwa[2], Jianwu Wang[1], Zhibo Zhang[2], Sanjay Purushotham[1]*

[1]Department of Information Systems, University of Maryland, Baltimore County, MD, USA
[2]Physics Department, University of Maryland, Baltimore County, MD, USA
Email: [1]{ztushar1, adeleka1, jianwu, zzbatmos, psanjay}@umbc.edu *

## ABSTRACT

The Earth's radiation budget relies on cloud properties like Cloud Optical Thickness obtained from cloud radiance observations. Traditional physics-based cloud retrieval methods face challenges due to 3D radiative transfer effects. Deep learning approaches have emerged to address this, but their performance are limited by simple deep neural network architectures and vanilla objective functions. To overcome these limitations, we propose CloudUNet, a modified UNet-style architecture that captures spatial context and mitigates 3D radiative transfer effects. We introduce a cloud-sensitive objective function with regularized L2 and SSIM losses to learn thick cloud regions often underrepresented in input radiance data. Experiments using realistic atmospheric and cloud Large-Eddy Simulation data demonstrate that our proposed CloudUNet obtains 5-fold improvement over the existing state-of-the-art deep learning, and physics-based methods.

***Index Terms***— Deep learning, remote sensing, cloud property retrievals

## 1. INTRODUCTION

Clouds play a crucial role in protecting the Earth from cosmic radiation. To estimate this protective capacity, various cloud properties such as cloud optical thickness (COT), cloud effective radius (CER), and cloud top height (CTH) are utilized. Typically, satellites are employed to gather cloud radiance data, enabling the estimation of these cloud properties [1, 2]. If the collected radiance data is one-dimensional, then a simple 1D inversion would result in accurate cloud properties retrievals [3]. However, in reality, clouds exist in three dimensions, necessitating the attainment of 3D radiance observations and the utilization of a 3D inversion method for precise cloud property retrievals. Unfortunately, radiance observations can be at most two dimensions, and attempting to retrieve cloud properties using a 2D inversion technique is hindered by the effects of 3D radiative transfer [4, 5]. The estimation of cloud optical thickness (COT) from radiance data poses a challenging 3D inverse problem, leading researchers to propose approximate solutions. For instance, Nakajima et al. [3] proposed the Independent Pixel Assumption (IPA), which assumes that each pixel in the radiance data is independent and represents individual clouds. However, this assumption is impractical since neighboring pixels are dependent on each other due to the 3D radiative transfer effect. Therefore, the estimated properties can differ significantly from the actual cloud properties. Recently, Deep Learning (DL) solutions have been explored to enhance remote sensing applications [6], particularly for cloud properties retrieval [7, 8]. Okamura et al. [7] developed a deep neural network (DNN) model to jointly retrieve COT and CER. However, their method employed a simple feed-forward neural network with an L2 loss objective for training, which limited its ability to fully capture the spatial context inherent in the radiance data. Another approach by Masuda et al. [8] focused on estimating COT from reflectance data captured by a ground-mounted digital camera using a convolutional neural network (CNN)-based model by mitigating the 3D radiative transfer effects. However, their method is limited by data collection methods and space-time averaging. Nataraja et al. [9] posed COT retrieval as a segmentation problem and retrieved COT using UNet [10, 11, 12, 13]. However, their approach requires high-dimensional radiance observations and is computationally expensive to train.

In this paper, we demonstrate that the above limitations could potentially be overcome with the use of a more sophisticated deep learning model and using refined training objective functions to effectively capture spatial context while handling 3D radiative transfer effects. Our main contributions include proposing CloudUNet, a UNet-style neural network, which employs a novel cloud-sensitive objective (CSO) function with weighted L2 and SSIM [14] loss to retrieve COT from 3D radiance data. Also, empirically we show that our CloudUNet model obtains 5-fold improvement in performance compared to the state-of-the-art deep learning and physics-based methods. Additionally, we conducted ablation studies to demonstrate the significance of the CSO loss function compared to the proposed methods in the literature.

## 2. DATA AND PROBLEM FORMULATION

First, we describe the data used for COT retrievals before introducing our problem formulation.

**Data:** A realistic representation of the atmosphere and

*This research is partially supported by grants IIS-2238743 NSF and 80NSSC21M0027 from NASA.

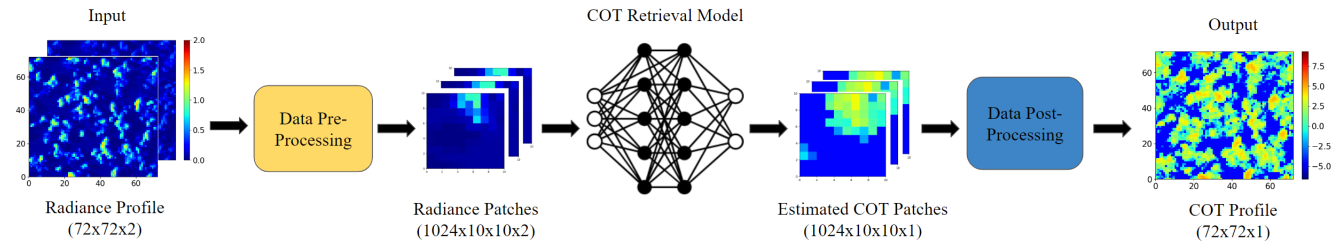


**Fig. 1**: Cloud Optical Thickness Retrieval Framework to estimate COT from the radiance data. Radiance Data are first transformed into small patches and DL/ML model is used to retrieve COT for corresponding patches. Data Post-processing generates the full COT profile from COT patches.

clouds is provided by Large-Eddy Simulation (LES) ARM Symbiotic Simulation and Observations (LASSO) since this representation is consistent with observations from the Atmospheric Radiation Measurement (ARM) program [15]. The LASSO LES cloud field (denoted as cloud profile) has a large domain size of 14km $\times$14km$\times$15km with horizontal resolution of 100m and vertical resolution of 30m below 5km which extends to 300m above 5km with all the cloud properties specified. The large domain size and high-resolution of LASSO LES cloud fields, makes it suitable for our three-dimensional (3-D) effects radiative transfer studies. To conduct radiance computations in 3-D, we utilized the spherical harmonics discrete ordinate method (SHDOM) developed by Evans et al. [16], operating at two specific wavelengths: $0.66\mu m$ and $2.13\mu m$. The angular resolution of the SHDOM used was 12 and 24. The radiative transfer calculation was performed at solar zenith angle (SZA) 60 degrees, solar azimuth angle (SAA) 0 degrees and view zenith angle (VZA) 0 degrees with double periodic horizontal boundary conditions. The surface was treated as a Lambertian surface with surface albedo of $0.05$. Thus, from the simulations at 200m resolution we have the COT as $72 \times 72$ matrix and radiance data as $72 \times 72 \times 2$ tensor, where the 2 represents the two wavelengths, and the remaining dimensions represent height and width of cloud profile, respectively.

**Problem Formulation:** We formulate the problem of COT retrieval as a regression task of predicting continuous COT value matrix provided the input radiance tensor data.

## 3. OUR COT RETRIEVAL FRAMEWORK

Our COT retrieval framework, shown in Figure 1, includes data processing modules and our proposed CloudUNet-based COT retrieval model.

**Data Processing:** The data processing module includes Pre-processing and Post-processing blocks. The Pre-processing block generates patches from the cloud profile's radiance data and maintains a hash table of patch locations. A patch represents a small region in the cloud profile, e.g., a $10 \times 10$ region. Using patches instead of the entire cloud profile allows for efficient data handling and training of smaller COT retrieval models. It's worth noting that simulating a large number of LES cloud profiles is time-consuming and costly, but it's necessary to train larger COT retrieval models. The Post-processing block uses the hash table to combine the COT patches retrieved by the COT retrieval model and generates the complete COT for the full cloud profile.

**COT Retrieval Model:** We propose CloudUNet as the COT retrieval model for accurately estimating COT values given radiance input data. Our CloudUNet, shown in Figure 2, is a UNet model and uses a customized loss function called *Cloud Sensitive Objective* to regress COT values from the radiance data. Additionally, like the popular UNet-style models used in computer vision [17, 18, 19], our CloudUNet model extracts spatial features at multiple scales, and uses skip connections to pass context from preceding layers.

Our CloudUNet model works as follows: given that the radiance patch data of size $10 \times 10$, the model incorporates only two downsampling layers. In each layer, features are extracted using a $3 \times 3$ convolutional operation. To downsample the data before proceeding to the next layer, a $2 \times 2$ max-pooling layer is applied. Since border pixels are lost during this process, the feature maps are cropped before being concatenated with the deconvolved layer. The inclusion of skip connections between opposite layers serves a dual purpose. It addresses the issue of gradient vanishing and provides additional context to the preceding layers. Overall, our CloudUNet model architecture is suited for COT retrieval because the convolutional filters at multiple scales helps to capture the local structure of the radiance data while the skip connections aids in preserving the global context of the entire patch.

**Cloud Sensitive Objective function:** We propose a customized *Cloud Sensitive objective (CSO) function*, shown in equation 3, for training our CloudUNet model. Our CSO function comprises two components: the *L2 loss* and the *SSIM*

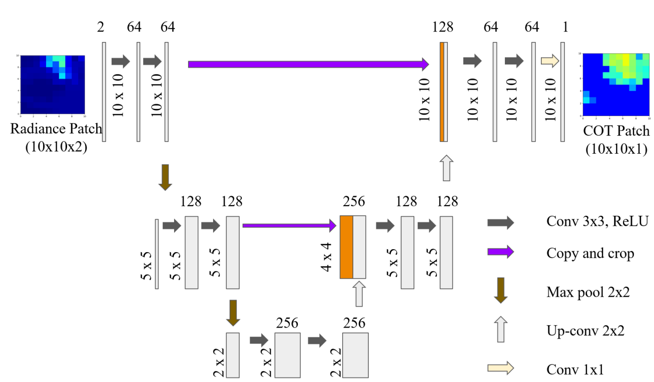


**Fig. 2**: Our Proposed CloudUNet Architecture

*loss*, each of which plays an important role during the model's training, as discussed below.

*L2 loss* is a commonly used objective function for regression task. In COT retrieval task, the rare thick cloud regions which represent the high COT regions, are difficult to learn using L2 loss [20] (Please see Figure 4). To enforce our CloudUNet model to learn the underrepresented high COT values, we propose a simple modified L2 loss function with a weighting parameter as shown in Eq. 1. Here, $\hat{y}$ and $y$ are the estimated and ground truth COT respectively. When the target COT values such as high COT regions are larger than a user-specified threshold $\theta$, then the L2 loss is multiplied by a weighting factor $\alpha$ ($\alpha > 1$). In experiments, we found that the tail distribution corresponds to the high COT regions and used 90th percentile of the COT training data in choosing the threshold.

$$L_P(\hat{y}, y) = \frac{1}{N} \sum_{i=1}^{N} \begin{cases} \alpha \cdot (\hat{y}_i - y_i)^2 & : y_i > \theta \\ (\hat{y}_i - y_i)^2 & : y_i \leq \theta \end{cases} \quad (1)$$

$$L_{SSIM}(\hat{y}, y) = \frac{(2\mu_{\hat{y}}\mu_y + c_1)(2\sigma_{\hat{y}}\sigma_y + c_2)}{(\mu_{\hat{y}}^2 + \mu_y^2 + c_1)(\sigma_{\hat{y}}^2 + \sigma_y^2 + c_2)} \quad (2)$$

$$L_{total}(\hat{y}, y) = \lambda_1 \times L_P(\hat{y}, y) + \lambda_2 \times (1 - L_{SSIM}(\hat{y}, y)) \quad (3)$$

The second loss term, shown in equation 2, corresponds to the *Structural Similarity Index Measure (SSIM)* [14] quality metric. We used the SSIM metric to capture the global structure similarity between the ground-truth COT and the predicted COT. If the ground-truth and predicted COT matrices are identical, then SSIM is +1, otherwise, it is $< 1$ and will be -1 if there are completely incorrect. Here, $c_1$ and $c_2$ are constant terms, and $\mu$ and $\sigma$ are mean and standard deviation respectively.

We trained our CloudUNet model using our proposed CSO loss function with two hyperparameters, $\lambda_1$ and $\lambda_2$, to adjust the weights of the L2 and SSIM losses. The intuition behind $1 - L_{SSIM}$ is that when the estimated COT is close to the ground truth, the $L_{SSIM}$ would be close to 1 and subtracting it from 1 would result in zero loss and when the estimated COT is not the optimal one, it would increase the total loss. Empirically we have found that both loss terms in Eq. 3 have similar ranges and therefore we have chosen $\lambda_1 = \lambda_2 = 1$ in our experimental setup.

## 4. EXPERIMENTS

In this section, we discuss the dataset preparation and the implementation details of our experimental setup.

**Data preparation:** For this study, a total of 102 LASSO LES cloud fields were simulated by employing the cloud Liquid Water Content (LWC) form, and a constant cloud droplet effective radius of 12$\mu$m for every column in all profiles. To account for the limited dataset size, we employed five-fold cross-validation (CV) in our experiments. During training, the Data Pre-processing block generated 1024 patches from each profile using a stride of 2. A thresholding operation was performed to ensure at least 25% cloudy region is present in each patch, specifically for deep learning (DL) training purposes. During inference, we used the Data Processing block to generate non-overlapping patches for COT retrievals, and combined these patches to obtain the complete cloud profile prediction. For example, our CloudUNet model predicted COT with a dimension of $10 \times 10$, so a stride of 10 was used during inference. Each fold of CV had 47,700 training samples, 19,050 validation samples, and 19,050 held-out test samples. The actual range of target COT values was 0 to 360, however, the majority of data is concentrated within a smaller range of 0 to 10. To effectively capture this sparse region, we employed a $log$ scale for estimating COT.

**Table 1**: Comparison of COT Retrieval Methods

| Methods | MSE ($\downarrow$ *Better*) | RRMSE (%) ($\downarrow$ *Better*) |
|---|---|---|
| Physics Based IPA (Nakajima et. al [3]) | $1.966 \pm 0.390$ | $14.49 \pm 6.35$ |
| Random Forests [22] | $2.956 \pm 0.133$ | $16.05 \pm 4.97$ |
| Feed forward DL (Okamura et. al [7]) | $1.891 \pm 0.142$ | $12.95 \pm 3.64$ |
| **CloudUNet (Ours)** | $\mathbf{0.342 \pm 0.038}$ | $\mathbf{5.81 \pm 2.10}$ |

**Implementation:** We conducted experiments using the PyTorch [21], employing various hyperparameters such as optimizers, learning rates, schedulers, and batch sizes. Our CloudUNet model achieved the best results with a learning rate of 0.1, batch size of 128, trained for 250 epochs with early stopping and patience of 50 epochs. The learning rate was adjusted using ReduceLR.

**Model Comparisons:** We compared the following models: (a) IPA retrievals - a physics-based method [3] used by NOAA, (b) Random Forests Regressor [22], (c) state-of-the-art deep learning model [7], (d) Our proposed CloudUNet model. For evaluations of the predicted COT retrieval with respect to the ground truth COT, we employed two metrics: Mean Squared Error (MSE) and Relative Root-Mean-Squared Error (RRMSE). MSE [23] is computed by taking the average of the squared difference between the predicted and the target COT. A lower MSE denotes better retrievals. Relative RMSE (RRMSE) [24] is calculated by dividing RMSE by the average of the squared ground truth where RMSE is the squared root of MSE. When RRMSE is less than 10%, it is considered excellent performance [25].

## 5. RESULTS AND DISCUSSION

**Quantitative Results:** Table 1 illustrates that our proposed CloudUNet model outperforms all the other models by a significant margin. CloudUNet achieves at least 82% lower MSE and a 55% lower Relative RMSE compared to the Okamura et. al [7] and Physics-based IPA retrieval methods [3], and achieves 88% lower MSE and a 64% lower Relative RMSE than Random Forests model. We believe the skip connections in CloudUNet facilitate information reuse and enable the extraction of complex spatial features from different layers.

**Qualitative Results:** We present the visualizations of the estimated COT retrievals of different models for the qualita-

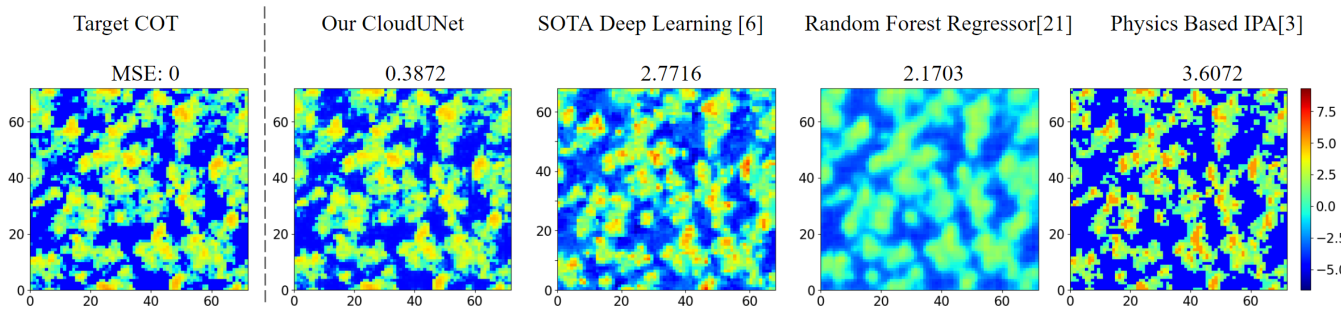


**Fig. 3**: Cloud Optical Thickness (COT) Retrievals using different methods. Colorbar represents the intensity of the COT value in log scale. Blue color indicates thinner or no cloud regions while red color denotes thicker cloud regions.

tive results. Figure 3 shows the COT estimation and MSE errors of an example profile for all the models. The colorbar (in *log* scale) represents the COT value where red indicates the highest possible COT values or thick cloud regions, and the blue indicates the lowest COT values or thin cloud regions. From this Figure 3, we see that the CloudUNet's COT retrieval closely aligns with the ground truth compared to state-of-the-art deep learning-based (Okamura [7]) and physics-based (IPA [3]) retrievals. The IPA retrievals are sharp but tends to overestimate the COT value across different regions, resulting in a high MSE loss. Conversely, both the Okamura and Random Forest models fail to produce sharp retrievals. By referring to the corresponding MSE scores, we can clearly observe the advantage of CloudUNet.

Remarks: Both the qualitative and quantitative results entail the advantages of CloudUNet and its training objective function. The model's architecture preserves global context using skip connections while capturing local features with downsampling and upsampling blocks, and the training objective accounts for the underrepresented thick cloud regions.

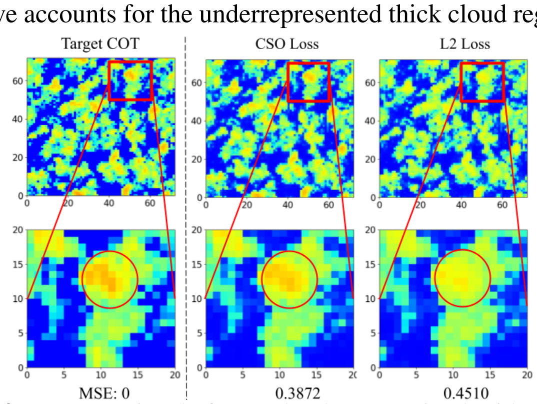


**Fig. 4**: COT Retrievals from CloudUNet trained with different objective functions. Top row represents the COT Profiles and the bottom row is the zoomed version of the specific area from the profiles which are marked by the rectangle.

## 6. ABLATION STUDIES

In this section, we investigate the effectiveness of our objective functions for COT retrieval.

**Impact of the Objective Function:** We compare and contrast the performance of CloudUNet model trained with different objective functions such as Binary Cross-Entropy (BCE) loss [9], L2 loss and proposed CSO function to study the impact of the training objective functions. Note that for the BCE loss we posed the COT retrieval as a 32 class segmentation problem as in [9] for fair comparison. Table 2 shows that the CloudUNet model achieves a lower MSE error with our CSO loss compared to the BCE and L2 loss indicating its effectiveness. The high MSE error for BCE based retrieval is due to the quantization effect i.e., each retrieved pixel always contains error because they are never the exact value rather an approximate one bounded by the bin range.

In Figure 4, we see that when the CloudUNet is trained with L2 loss, the high COT value regions (thick cloud regions) are not estimated accurately. This is because the data inherently has a tail distribution i.e, the thick clouds are rare compared to the clear sky, and an L2 loss provides an average estimation ignoring these regions. Our objective function places more weight on the underrepresented regions so that the model also learns these COT values along with the dominating low COT value regions. The rarity of thick clouds also explains the small improvement in MSE and RRMSE scores.

**Table 2**: Ablation Studies-CloudUNet results with different objective functions

| Objective Function | MSE | RRMSE (%) |
|---|---|---|
| BCE Loss [9] | $9.006 \pm 3.775$ | $25.60 \pm 5.33$ |
| L2 Loss | $0.407 \pm 0.052$ | $6.38 \pm 2.47$ |
| **CSO (Ours)** | $\mathbf{0.342 \pm 0.038}$ | $\mathbf{5.81 \pm 2.10}$ |

## 7. CONCLUSIONS

Estimating cloud properties, such as Cloud Optical Thickness (COT), holds great importance in atmospheric sciences due to its impact on climate and weather. In this paper, we demonstrated that UNet-style deep learning model architectures, such as our proposed CloudUNet model with novel loss function, are well-suited for reducing 3D radiance effects, as supported by both quantitative and qualitative COT retrieval results. In future research, we plan to extend our methods to multi-view data, where both Solar Zenith Angles and View Zenith Angles vary.

## 8. REFERENCES


[1] Han Ma and Shunlin Liang, "Retrieval of multiple land surface and atmospheric parameters from the himawari-8 ahi top-of-atmosphere observations," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 138–141.

[2] Guodong Zhang, Han Ma, and Shunlin Liang, "Estimating 250-m land surface and atmospheric variables from mersi top-of-atmosphere reflectance," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2021.

[3] Teruyuki Nakajima and Michael D King, "Determination of the optical thickness and effective particle radius of clouds from reflected solar radiation measurements. part i: Theory," *Journal of Atmospheric Sciences*, 1990.

[4] Alexander Marshak and Anthony Davis, *3D radiative transfer in cloudy atmospheres*, Springer Science & Business Media, 2005.

[5] A. S. Ademakinwa, Z. H. Tushar, J. Zheng, C. Wang, S. Purushotham, J. Wang, K. G. Meyer, T. Várnai, and Z. Zhang, "Influence of cloud retrieval errors due to three dimensional radiative effects on calculations of broadband cloud radiative effect," *EGUsphere*, vol. 2023, pp. 1–32, 2023.

[6] Seraj Al Mahmud Mostafa et al., "Cnn based ocean eddy detection using cloud services," in *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2023, pp. 4052–4055.

[7] Rintaro Okamura et al., "Feasibility study of multi-pixel retrieval of optical thickness and droplet effective radius of inhomogeneous clouds using deep learning," *Atmospheric Measurement Techniques*, 2017.

[8] Ryosuke Masuda et al., "Retrieval of cloud optical thickness from sky-view camera images using a deep convolutional neural network based on three-dimensional radiative transfer," *Remote Sensing*, 2019.

[9] Vikas Nataraja, Sebastian Schmidt, Hong Chen, Takanobu Yamaguchi, Jan Kazil, Graham Feingold, Kevin Wolf, and Hironobu Iwabuchi, "Segmentation-based multi-pixel cloud optical thickness retrieval using a convolutional neural network," *Atmospheric Measurement Techniques Discussions*, pp. 1–34, 2022.

[10] Olaf Ronneberger et al., "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015.

[11] Lu Li, Chao Wang, Hong Zhang, and Bo Zhang, "Residual unet for urban building change detection with sentinel-1 sar data," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 1498–1501.

[12] Abdul Basit, Muhammad A. Siddique, and M. Saquib Sarfraz, "Deep learning based oil spill classification using unet convolutional neural network," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 2021, pp. 3491–3494.

[13] Zhengquan Chu, Tian Tian, Ruyi Feng, and Lizhe Wang, "Sea-land segmentation with res-unet and fully connected crf," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 3840–3843.

[14] Zhou Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[15] Atmospheric Radiation Measurement (ARM) Program, "Les arm symbiotic simulation and observation activity," 2015.

[16] K Franklin Evans, "The spherical harmonics discrete ordinate method for three-dimensional atmospheric radiative transfer," *JAS*, 1998.

[17] Phillip Isola et al., "Image-to-image translation with conditional adversarial networks," in *CVPR*, 2017, pp. 1125–1134.

[18] Jun-Yan Zhu et al., "Toward multimodal image-to-image translation," *NeurIPS*, vol. 30, 2017.

[19] Dmitrii Torbunov et al., "Uvcgan: Unet vision transformer cycle-consistent gan for unpaired image-to-image translation," in *WACV*, 2023.

[20] Zezhou Cheng et al., "Deep colorization," in *ICCV*, 2015.

[21] Adam Paszke et al., "Pytorch: An imperative style, high-performance deep learning library," *NeurIPS*, 2019.

[22] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *JMLR*, 2011.

[23] Peter J Bickel and Kjell A Doksum, *Mathematical statistics: basic ideas and selected topics, volumes I-II package*, CRC Press, 2015.

[24] Milan Despotovic et al., "Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation," *RSE Reviews*, 2016.

[25] Mao-Fen Li et al., "General models for estimating daily global solar radiation for different solar radiation zones in mainland china," *ECM*, 2013.