CLOUDUNET: A DEEP LEARNING MODEL FOR RETRIEVING CLOUD OPTICAL THICKNESS FROM 3D LES CLOUD FIELDS

Zahid Hassan Tushar, Adeleke Ademakinwa, Jianwu Wang, Zhibo Zhang, Sanjay Purushotham

University of Maryland Baltimore County
Maryland, USA.

ABSTRACT

Earth's radiation budget is determined by using the cloud properties such as Cloud Optical Thickness which are passively obtained from the radiance observations of the clouds. Over the years many researchers have proposed physics-based methods to estimate the cloud properties but were not entirely successful due to the 3D radiative transfer effects. Recently, a couple of deep learning-based solutions are proposed in the literature to reduce the radiative transfer effects in retrieving cloud properties. However, their performance was limited due to the preliminary deep learning model architectures and the usage of vanilla objective functions during training. We proposed CloudUNet, a modified UNet-style architecture with skip connections to comprehend the spatial context in the radiance observations completely and thus alleviating radiative transfer effects that would affect the cloud properties retrievals. CloudUNet uses a modified L2 loss to enforce the model to learn thick cloud regions which are often underrepresented in the reflectance input data. We conducted experiments on the realistic atmospheric and cloud representation simulation data generated from the Large-Eddy Simulation (LES) ARM Symbiotic Simulation and Observation (LASSO) to show the effectiveness of our model and objective function compared to the existing state-of-the-art deep learning and physics-based approaches. Our initial empirical results are promising and we have achieved 5-fold improvement over other existing methods.

***Index Terms*—** COT, deep learning, cloud optical thickness retrieval, remote sensing

1. INTRODUCTION

Atmospheric contents such as cloud help earth to protect itself from cosmic radiations. This protection capacity is estimated by using different cloud properties such as cloud optical thickness (COT), cloud effective radius (CER), cloud top height (CTH). Satellites are used to collect the cloud radiance data from which the cloud properties can be estimated. As shown in 1, if the collected radiance data is one-dimensional, then a simple 1D inversion would result in accurate cloud properties retrievals. However, practically a cloud is 3D and to retrieve accurate cloud properties one would need to achieve the 3D radiance observations and do a 3D inversion. There exists a significant gap between the retrieved cloud properties and the actual cloud properties since the radiance observations can be at most 2D and retrieved cloud properties using a 2D inversion suffers from the 3D radiative transfer effects [1]. Note that 3D radiative transfer denotes the process of computing the transfer of electromagnetic waves through a three-dimensional medium such as cloud.

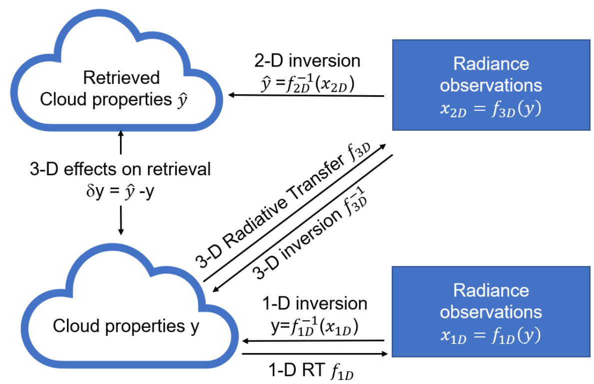

Fig. 1. Conceptual framework to show spatial context between radiance and cloud properties under 3D radiative transfer effect.

Retrieving cloud optical thickness from the radiance data is a 3D inverse problem that forced the physicist to propose approximated solutions with impractical assumptions. For instance, Nakajima et al. [2] assume that each pixel in the radiance data is independent of their neighboring pixels and represents individual clouds. It simplified their calculations but practically the pixels are dependent on each other due to the 3D radiative transfer effect. Therefore the estimated properties quite differ from the actual properties of the cloud.

Retreival of cloud properties such as Cloud Optical Thickness (COT) and Cloud Effective Radius (CER), has been well-studied. Recently, Okamura et al. [3] developed a deep learn-

This work is supported by the NASA award 80NSSC21M0027

ing based neural network model to jointly retrieve COT and CER, and achieved state-of-the-art results compared to the traditional physics-based retrieval methods. However, Okamura et. al proposed a simple feed-forward neural network that uses an L2 loss objective for training, and it was not effective to capture the spatial-temporal context present in the LES cloud data.

The performance of the deep learning models is dependent on the architecture as well as on the objective function that guides the deep learning training. A lot of works in the literature are proposed that introduce different training objectives such as L1 Loss [4, 5], L2 Loss [3, 6, 7], Kullback-Divergence (KD) Loss [8], and least square GAN loss [9] for regression problems. However, the proposed deep learning approaches for COT retrievals only use L2 Loss which might hinder them to learn the underlying data distribution.

Therefore this work aims to retrieve the cloud optical thickness from the radiance data by resolving the existing challenges of appropriate deep learning model and training objectives. We summarise our main contribution as follows:

- To retrieve cloud optical thickness, we propose a modified UNet-style deep neural network and named it CloudUNet which is compact in size and able to model the 3D inverse problem of radiance to COT data mapping.

- To better learn the COT retrieval, we designed an objective function to guide the deep learning model training. This function consists of weighted L2 loss and Structured Similarity Index Measure (SSIM).

- We have conducted experimental analysis to establish the advantage of our design over other state-of-the-art deep learning and physics based approaches.

- We have also designed an end-to-end framework(Fig. 2) to run inference on any size of reflectance profile data. This block provides the option to reduce inference time for a close approximation of the solution.

We expect that this work will inspire researchers to employ deep learning solutions to retrieve different cloud properties. Our belief is that this will work as guide for the researchers to design deep learning models and corresponding objective functions to estimate cloud properties other than COT. The outline of this paper is as follows. We present the literature overview in Section 2. Section 3 contains the description of our proposed method. In Section 4 we provide details of our experimental setup including data specifications. Our findings and results are discussed in Section 5. Finally, the paper is concluded in Section 6 with an overview of this work.

2. RELATED WORKS

Earth's radiation capacity are strongly dependent on the cloud properties such as cloud optical thickness and effective radius [10]. Researchers have proposed complex mathematical models to retrieve the cloud properties [11, 12, 13, 14, 2]. One of the most accurate method was presented in [2] where the authors determined optical thickness and effective particle radius from solar reflectance of the clouds. Primarily they have used reflectance data at two wavelengths for this task. But in case of thin clouds their approach provides ambiguous results which is possible to reduce by using a third wavelength claimed by the authors. However, their methods still produce inaccurate retrievals of cloud properties due to their independent pixel approximation (IPA) even though the pixels in the reflectance data are codependent.

Okamura et al. [3] attempted to resolve 3d radiative transfer effects by using deep learning techniques. They proposed two deep neural network based models: one to correct IPA retrievals and one to retrieve COT and CER. Their later model requires reflectance information using satellites from four wavelengths to correctly estimate the properties. Even though their method was unique and better compared to traditional physics based approach, they could not capture the spatial relation between the pixel accurately which resulted in sub-optimal retrievals. Also another limitation of their approach is that it requires manually setting a threshold level to adjust the estimated COT values. Masuda et al. [6] also approached the COT estimation from reflectance data captured by ground-mounted digital camera using a deep learning model. They proposed a convolutional neural network (CNN) based model to reduce the 3D radiative transfer effects. However, their methods are limited by data collection methods and space-time averaging which could be alleviated with a more sophisticated deep learning architecture and training objective function.

Considering the existing challenges in cloud properties retrievals, we propose a deep learning model which is able to capture the spatial relation in the reflectance data and thereby largely reducing the 3D radiative transfer effects and retrieving COT accurately.

3. METHODOLOGY

Our proposed COT retrieval framework is shown in Fig. 2. There are mainly two blocks in our proposed approach. The Data processing blocks are used to iterate over the radiance data to extract patches and later put the corresponding retrieved COT patches together to generate the COT for the entire profile. The other block is the deep learning model which is employed to estimate COT from radiance data patches. The details of the blocks are included in this section.

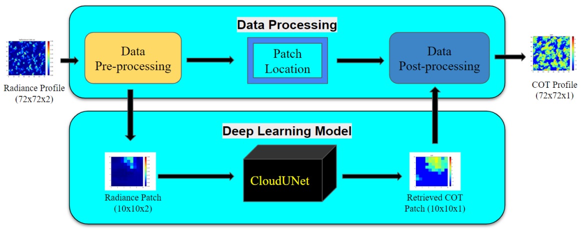

Fig. 2. COT Retrieval Framework is used to estimate cloud optical thickness from the radiance data. Data are first transformed into small patches and a map is maintained with the patch location in the original data. Patches are used by the CloudUNet model to estimate COTs which later are accumulated to retrieve the complete COT. CloudUNet has a modified UNet-style architecture.

3.1. Data Processing Block

Data Processing is done using two blocks: Pre-processing and Post-processing. The purpose of the Data Pre-processing block is to generate patches from the cloud profile's reflectance data and maintain a map that keeps track of the extracted patch locations. Note that, a single patch is a 10×10 small part of the cloud profile. In case of cloud profile's radiance data which has information at two channels, a radiance patch will have a dimension of $10 \times 10 \times 2$. Similarly, a COT patch will have a dimension of $10 \times 10 \times 1$ as COT data has information at single channel. The map is passed to the Data Post-processing block where the retrieved COTs from the patches are put together using the map to generate the COT for the full profile. Overlapping areas are averaged to retrieve the COT. The data processing block has a powerful aspect. By changing the stride, one can control the inference time and the quality of the retrieved COTs. There is a trade-off between the inference time and the retrieved COT quality. When a large stride (max. 10) is used, the number of reflectance patches from a cloud profile would be small, and therefore the deep learning model would be used fewer times to retrieve the COTs from the corresponding reflectance patches. In the case when a smaller stride is used, it results in a large number of reflectance patches, and therefore the usage of deep learning model is increased raising the overall inference time but the retrieval quality of the profile also improves due to overlapping COT patches. Note that the stride number has to be an integer between $1 \sim 10$.

The intuition behind generating patches from cloud profiles instead of using the cloud profile itself is to circumvent the limited size of the data. We know that deep learning model requires a lot of annotated data to train [15]. But generating cloud profile data and annotating them are time-consuming and expensive. Patches generated by the Data Pre-processing block were ample to train our model and achieve generalization over the unseen data.

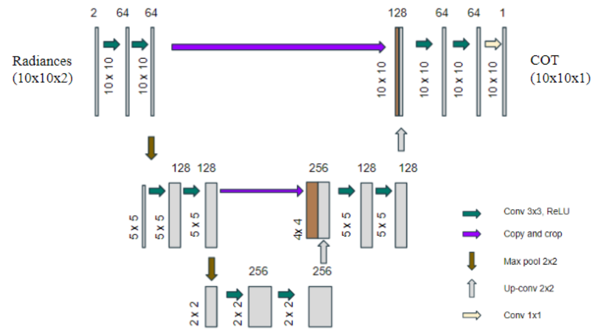

Fig. 3. CloudUNet Architecture: Multiple convolutional layers with skip connections between intermediate layers.

3.2. Model Architecture

Retrieving cloud optical thickness can be considered as image-to-image translation problem. The only differences are that the input will have channels equal to the number of wavelengths present in the radiance data instead of 1 (grayscale) or 3 (RGB), and the output will have only single channel. U-Net style architectures have proven to perform extremely well with image-to-image translation problem [16, 17, 18]. Hence a U-Net style architecture is proposed in this work which we termed as CloudUNet. The benefits of this architecture are the multiple convolutional layers to extract features at different scales and the intermediate skip connections to pass context from preceding layers. The model is shown in Fig. 3.

Since the radiance patch data have a dimension of 10×10, deep architectures cannot be employed. Therefore a shallow network is designed. At each layer, 3×3 convolution is applied to extract features. Before passing to the next layer, it is downsampled through a 2×2 maxplooing layer. The features maps are cropped before concatenating with the deconvolved layer due to loss of border pixels. The skips connections between opposite layers help the model in a two-fold way. It solves the gradient vanishing problem and gives more context to the preceding layers.

3.3. Cloud Sensitive Objective Function

One of the key contributions of this work is the cloud sensitive objective function which helps the CloudUNet to learn the accurate mapping between the radiance data and the COT. The formula for the objective function is given in the following equations.

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

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

$$L_{total}(\hat{y}, y) = L_P(\hat{y}, y) + 1 - L_{SSIM}(\hat{y}, y) \qquad (3)$$

Objective function guides the deep learning training of the CloudUNet model. The COT data has a tail distribution where majority of the data points are distributed in a small range $0 \sim\ 10$. The deep learning model updates its parameters to cover the most of the datapoints. Therefore the loss function in Eq. 1, has a weighting parameter which enforces the model to learn the underrepresented COT ranges as well. $\hat{y}$ and y are the estimated and ground truth COT respectively. When the COT values are higher than a certain threshold θ, the squared difference are multiplied by a weighting factor α ($\alpha > 1$) to place more weights.

As explored in [19], MSE between two images might be the same but the structural quality might not. To account for the structural similarity they have implemented SSIM which stands for Structural Similarity Index Measure. It is implemented in Eq. 2 which gives 1 when the images have same structure and -1 when they are completely different. c_1 and c_2 are constant terms, and μ and σ are mean and standard deviation respectively. For more details please refer to [19].

Eq. 3 is the total loss which combines the loss terms from Eq. 1 and Eq. 2. 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.

4. EXPERIMENTAL SETUP

4.1. Data Preprocessing and Postprocessing

There are 102 LES cloud profiles in the dataset. The profiles are simulated by setting different cloud properties such as cloud effective radius, solar zenith angles, view zenith angles. The input is the reflectance data at two wavelengths $0.66\mu m$ and $2.13\mu m$. For the development purpose $200m$ resolution data are used since the highest resolution of the real satellite is $200m$. Input reflectance data has a dimension of $72 \times 72 \times 2$ where 2 is for two wavelengths and the rest are height and length. The target COT has the same height and length but single channel. Given the small size of the dataset, five-fold cross-validation (CV) is employed. The dataset is divided into five folds. For each simulation run, three folds are used as training data, one as validation, and one as test. After CV separation, a patch generator is used to divide the profiles into 10×10 patches with a stride of 2. It is done at this point so that patch samples from different cloud profiles do not get distributed across train, valid and test sets. Otherwise, this would have led to information leaks which have been avoided. From each profile, we generated 1024 patches during training. A thresholding operation is performed to ensure at least 25% cloudy region on each patch. This is done

Table 1. LES Dataset Summary

LES Cloud Profiles		Patch Dataset
Properties	***No. Profiles***	***No. Patch Samples***
SZA-60 VZA-0 SAA-0 CTH-0 CER-0	102	Train-47,700 Valid-19,050 Test-19,050 Total-85,800

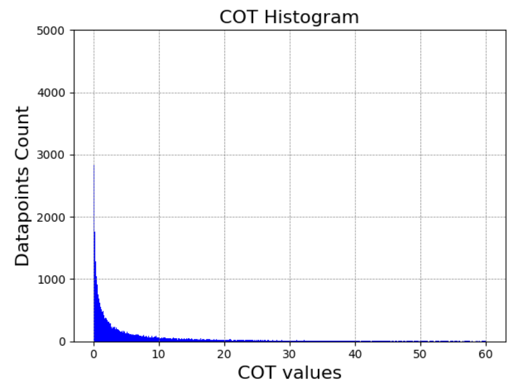

Fig. 4. Histogram of COT distribution. COT bins are in the vertical axis and horizontal axis represents the quantity of datapoints in these ranges. The data has tail distribution reflecting very little thick clouds.

for deep learning training only. For inference, we used the Data Processing block to generate patches with a stride that results in non-overlapping COT retrievals and combine the patches to get the full cloud profile prediction. For instance, the CloudUNet predicts COT with 10×10 dimension, therefore we use a stride of 10 during inference. On the other hand, we use a stride of 6 in the case of Okamura model since its estimated COT has 6×6 dimensions. A summary of the dataset is given in Table.1.

As we can see from the Table.1, there are $47,700$ samples in the training set with $19,050$ samples for validation and equal number of samples for test purpose. The different cloud properties are also provided here. We have generated histograms for the training set to get insight on the data distribution of the COT. The histograms are shown in Fig. 4.

From the Fig. 4, it is noticeable that most of the data are distributed in a small range 0 10. An intuitive way to learn this dense but brief region would be to use log scale when estimating COT. Therefore to emphasize on this region and precise learning of the COT distribution, we have trained the models to predict the log of COT values and reported the MSE errors on this scale as well.

4.2. Implementation

As part of hyperparameter tuning, we have employed different optimizers, learning rates, schedulers and batch size.

Table 2. MSE Loss Comparison Between Different COT Retrieval Methods (IPA retrievals, Okamura and CloudUNet). The scores are based on Test Profiles.

Models		Test MSE Loss
Type	***Name***	
Physics Based baseline Model	IPA Retrievals	1.966 ± 0.390
Deep Learning Based Model	Okamura	1.891 ± 0.142
	CloudUNet (Ours)	$\mathbf{0.372 \pm 0.038}$

As optimizers, we have run simulations using both SGD and Adam optimizers and found Adam to be suitable for our training instead of SGD. Decaying the learning rate is a common practice in deep learning training. Using a smaller learning rate from the start of the training might lead to getting stuck at a local minimum instead of a global minimum. To decay the learning rate, we have used schedulers such as ReduceLR, StepLR and ExponentialLR with different initial learning rates. We also run simulations for a set of batch sizes: 64, 128, 256, and 512. The best result was achieved when the CloudUNet model was trained with a learning rate 0.1, batch size 128, 10000 epochs and an early stopping criterion having the patience of 50 epochs. The learning rate was adjusted using ReduceLR. We compared our model with a physics-based baseline method described in [2]. For comparison with the deep learning based methods, we also implemented the COT retrieval model as deep learning baseline model from the Okamura et al.[3] paper which we refer as the Okamura model.

5. RESULTS AND DISCUSSION

5.1. Quantitative Results

We train the deep learning models in five fold cross validation set up due to the small size of the dataset. The models are trained on the training set of the data. The training is stopped when the model's performance on validation set does not improve with further training. We report the model's performance on the test set to quantify the generalization capability of the model on the unseen data. Mean and standard deviation of the test MSE scores of different models are reported in Table. 2.

The original Okamura model described in [3] requires data in four wavelengths to estimate the COT properly. As we can see from the Table. 2, when data at two wavelengths are only available, the okamura model is not performing well. It's performance is similar to the physics based IPA retrievals. On the other hand, CloudUNet outperforms these two models by a large margin achieving 80% less MSE error compared to the Okamura model. The skip connections between the intermediate layers in the CloudUNet help the model to reuse information from earlier layers and use the other layers to extract more complex features. However, Okamura model do not have this advantage which explains their performance.

5.2. Qualitative Results

In this subsection, we present the visualization of our finding. In Fig. 5, the COT estimation of a specific profile for all the models are shown. Corresponding MSE errors are also reported. Note that the test profiles are unseen to the models and only used during inference. The colorbar represents the COT value where red indicates the highest possible COT values or thick cloud region, and the blue indicates the lowest COT value or thin cloud region. The colorbar is in logarithmic scale. The logic behind this choice is that COTs have a range of 0 360 with most of the values residing in a small range 0 10, and using a regular linear scale would not be very helpful to visualize most of the COT values. A logarithmic scale expanded the prominent small COT ranges while shrunk the sparse large COT value range, thereby increased the visual understanding. 0.01 is added to COT value in order to avoid undefined region of the log scale.

For the deep learning models, the full profile predictions are generated in a modular way unlike the physics based retrievals. For the Okamura and CloudUNet, the predictions are generated in non-overlapping way. It means that we have used a stride of 6 for the Okamura model since it predicts an area of 6×6. Similarly a stride of 10 is used for the CloudUNet model. This also means that CloudUNet is faster to retrieve COT than Okamura and does not lose information at the boundary.

As seen from the Fig. 5, the estimated COT with CloudUNet is quite close to the ground truth compared to the Okamura and IPA retrievals. The IPA retrievals are sharp but tends to overestimate the COT value throughout the regions leading them to large MSE loss. On the other hand, Okamura model failed to generate a sharp retrieval. We can also refer to the corresponding MSE scores to realize the advantage of the CloudUNet. The explanation behind this is that L2 error forces to achieve a global average estimation instead of a sharp one.

For a closer inspection, we also compared between the COT estimation at patch levels. Since Okamura model takes a patch size of $10 \times 10 \times 2$ but predicts a 6×6 region, for fair comparison the IPA retrievals and the CloudUNet retrievals are cropped at the output to match the Okamura model's predictions. An instance of patch level comparison is shown in Fig. 6.

It is noticeable that IPA retrievals have failed to estimate COT for a large region and Okamura's estimation seems somewhat random. On the contrary, CloudUNet accurately predicted most of the region. The MSE scores also reflects the visual results.

6. ABLATION STUDIES

In this section, we aim to find the effectiveness of the modified L2 Loss and SSIM regularization term in the deep learning

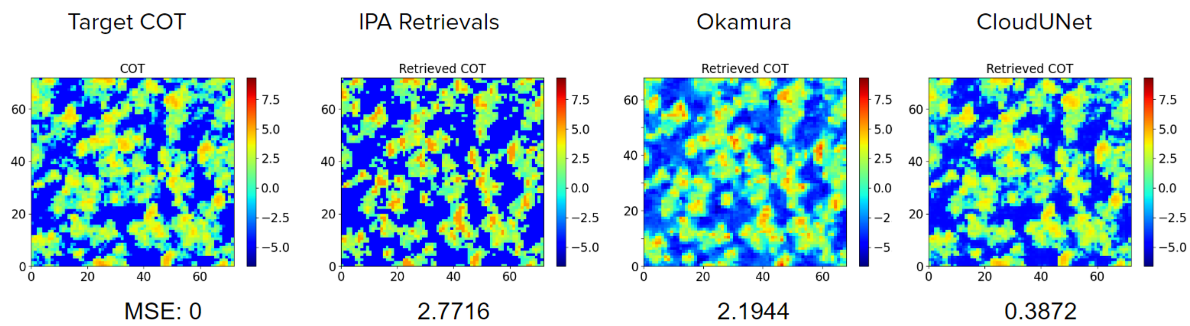

Fig. 5. COT Retrievals using different methods (Profile view). Colorbar represents the intensity of the COT value. Blue indicates no cloud region while red denotes thick clouds. From Left to Right: Ground Truth COT; Retrieved COT from Physics-based Independent Pixel Approx. model (Sharp but over/under estimation, High MSE); Retrieved COT from Okamura model (Blurry); Retrieved COT from CloudUNet model (ours) (Close to Ground Truth).

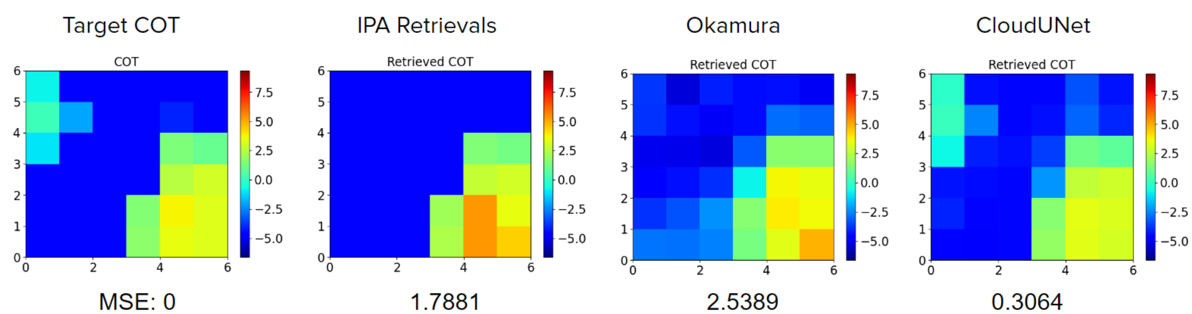

Fig. 6. COT Retrievals using different methods (Patch level view). Colorbar represents the intensity of the COT value. Blue indicates no cloud region while red denotes thick clouds. Top Left: Ground Truth COT; Top right: Retrieved COT from Physics-based Independent Pixel Approx. model (Sharp but missed most regions); Bottom Left: Retrieved COT from Okamura model (Over/ under estimation); Bottom Right: Retrieved COT from CloudUNet model (ours) (Close to Ground Truth).

Table 3. MSE loss Comparison for Experiments on Effectiveness of the Objective Function

Models		**Test MSE Loss**
Name	***Objective Function***	
CloudUNet	L2 Loss	0.440 ± 0.058
	Cloud Sensitive Obj Func (Ours)	$\mathbf{0.372 \pm 0.038}$

model's training as well as the advantage of the CloudUNet architecture.

6.1. Effectiveness of the Objective Function

In this section, we explore the effectiveness of our custom designed objective function. For a fair comparison, the model architecture is kept same which is CloudUNet in this case. We train the CloudUNet with L2 loss and then with our cloud objective function. MSE loss for each simulation is reported in Table. 3.

The CloudUNet model achieves a lower MSE error with the modified L2 loss and SSIM regularization objective function instead of the L2 loss. It is because the data has a tail distribution as we can see from Fig. 4 and our objective function places more weights to the underrepresented regions so that the model learns these COT values too along with the dominating regions.

6.2. Effectiveness of the architecture

Deep learning architecture plays an important role in learning the problem task. We can show this by running a fair comparison. We have trained the okamura and the CloudUNet model using the same objective function which is L2 loss in this case. The test MSE scores are reported in Table. 4. The MSE error is very low for the CloudUNet model compared to the Okamura model. The scores entail that CloudUNet might have learned the reflectance to COT data mapping very well due to its inherent skip connections and encoder-decoder type architecture.

Table 4. MSE Loss Comparison for Experiments on Effectiveness of the Model Architecture (Okamura vs CloudUNet)

Models		**Test MSE Loss**
Objective Function	***Name***	
L2 Loss	Okamura	1.891 ± 0.142
	CloudUNet (Ours)	$\mathbf{0.440 \pm 0.058}$

Table 5. Inference Scores using Radiance from Single wavelength for CloudUNet

Training Specs (Wavelength)		**Test MSE Score**	**Remarks**
$0.66\mu m$	$2.13\mu m$		
Yes	Yes	0.372 ± 0.038	Best Score
Yes	No	49.007 ± 53.335	Performance degrades
No	Yes	98.591 ± 53.747	Performance degrades further

6.3. Relative Comparison of Information Content on Radiance Observing wavelengths

In this ablation studies, we aim to compare the information content on different reflectance wavelengths. We can achieve this by turning off information from either of the channels and evaluate the COT retrievals estimated using only information from a single channel. In the implementation perspective, we entered zeros instead of reflectance data on one of the channels and measured the COT retrievals. The results are reported in Table. 5

As seen in Table. 5, when information from wavelength $2.13\mu m$ is turned off, the CloudUNet's MSE score increased by a large factor. When we turned off information from $0.66\mu m$ and used only $2.13\mu m$'s information, it becomes even worse. This entails that the information from both wavelengths are vital for CloudUNet to retrieve COT accurately. Also it appears that $0.66\mu m$ wavelength relatively carries more weights compared to the other wavelength.

7. CONCLUSIONS

Estimating cloud properties is a very important task for which researchers have introduced different methods. Traditional physics based approaches were hindered by the 3D radiance effect. It is possible to reduce this effect using deep learning based solutions but they require careful design of model architecture and objective function. Our proposed CloudUNet has demonstrated to reduce the effect very well as seen from the quantitative and qualitative results. The inherent skip connections in the architecture and the cloud objective function might have helped to learn the radiance to COT data mapping better than other methods. The synthetic data are from a fixed SZA and VZA making it a single view dataset. In future, we plan to extend our methods to the multi-view data.

8. REFERENCES

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

[2] 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*, vol. 47, no. 15, pp. 1878–1893, 1990.

[3] Rintaro Okamura, Hironobu Iwabuchi, and K Sebastian Schmidt, "Feasibility study of multi-pixel retrieval of optical thickness and droplet effective radius of inhomogeneous clouds using deep learning," *Atmospheric Measurement Techniques*, vol. 10, no. 12, pp. 4747–4759, 2017.

[4] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[5] Soohyun Kim, Jongbeom Baek, Jihye Park, Gyeongnyeon Kim, and Seungryong Kim, "Instaformer: Instance-aware image-to-image translation with transformer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 18321–18331.

[6] Ryosuke Masuda, Hironobu Iwabuchi, Konrad Sebastian Schmidt, Alessandro Damiani, and Rei Kudo, "Retrieval of cloud optical thickness from sky-view camera images using a deep convolutional neural network based on three-dimensional radiative transfer," *Remote Sensing*, vol. 11, no. 17, pp. 1962, 2019.

[7] MM Nabi, Volkan Senyurek, Ali C Gurbuz, and Mehmet Kurum, "Deep learning-based soil moisture retrieval in conus using cygnss delay–doppler maps," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 6867–6881, 2022.

[8] Diederik P Kingma and Max Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[9] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley, "Least squares generative adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2794–2802.

[10] V Ramanathan, "The role of earth radiation budget studies in climate and general circulation research," *Journal of Geophysical Research: Atmospheres*, vol. 92, no. D4, pp. 4075–4095, 1987.

[11] Albert Arking and Jeffrey D Childs, "Retrieval of cloud cover parameters from multispectral satellite images," *Journal of Applied Meteorology and Climatology*, vol. 24, no. 4, pp. 322–333, 1985.

[12] G Molnar and JA Coakley Jr, "Retrieval of cloud cover from satellite imagery data: A statistical approach," *Journal of Geophysical Research: Atmospheres*, vol. 90, no. D7, pp. 12960–12970, 1985.

[13] RJ Curran, HL Kyle, LR Blaine, J Smith, and TD Clem, "Multichannel scanning radiometer for remote sensing cloud physical parameters," *Review of Scientific Instruments*, vol. 52, no. 10, pp. 1546–1555, 1981.

[14] William B Rossow, Leonid C Garder, and Andrew A Lacis, "Global, seasonal cloud variations from satellite radiance measurements. part i: Sensitivity of analysis," *Journal of Climate*, vol. 2, no. 5, pp. 419–458, 1989.

[15] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals, "Understanding deep learning (still) requires rethinking generalization," *Communications of the ACM*, vol. 64, no. 3, pp. 107–115, 2021.

[16] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[17] Jun-Yan Zhu, Richard Zhang, Deepak Pathak, Trevor Darrell, Alexei A Efros, Oliver Wang, and Eli Shechtman, "Toward multimodal image-to-image translation," *Advances in neural information processing systems*, vol. 30, 2017.

[18] Dmitrii Torbunov, Yi Huang, Haiwang Yu, Jin Huang, Shinjae Yoo, Meifeng Lin, Brett Viren, and Yihui Ren, "Uvcgan: Unet vision transformer cycle-consistent gan for unpaired image-to-image translation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 702–712.

[19] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.