Toward Physics-Informed Neural Networks for 3-D Multilayer Cloud Mask Reconstruction

Yiding Wang, *Student Member, IEEE*, Jie Gong, *Member, IEEE*, Dong L. Wu, and Leah Ding, *Member, IEEE*

***Abstract*— Three-dimensional cloud retrievals are critical for understanding their impact on climate and other applications, such as aviation safety, weather prediction, and remote sensing. However, obtaining high-resolution and accurate vertical representation of clouds remains unsolved due to the limitations imposed by satellite instrumentation, viewing conditions, and the complexity of cloud dynamics. Cloud masks are essential for comprehending various cloud vertical properties, but deriving accurate 3-D cloud masks from 2-D satellite imagery data is a challenging task. To tackle these challenges, we introduce a physics-informed loss function for training deep learning models that can extend 2-D cloud images into 3-D cloud masks. The proposed loss, called CloudMask loss, is composed of two domain knowledge-informed loss terms: one for evaluating cloud position and thickness and the other for measuring the number of layers. By combining these loss terms, we improve the trainability of the deep learning models for more accurate and meaningful results. We apply the proposed loss function to different neural networks and demonstrate significant improvements in multilayer cloud mask reconstruction. Utilizing the same neural network architecture, our proposed loss outperforms standard binary cross-entropy (BCE) loss in terms of multilayer cloud classification accuracy, number of layers accuracy, and thickness mean absolute error (MAE). The proposed loss function can be readily integrated into various neural network architectures, resulting in substantial performance gains in 3-D cloud mask generation.**

***Index Terms*— 3-D cloud mask retrievals, multilayer clouds, neural networks, remote sensing, satellite imagery.**

I. Introduction

CLOUDS are ubiquitous. They play critical roles in the energy balance and hydrological cycle. Observational record of clouds could be dated back to the Stone Age [1], but global observations of them were not readily available before the launch of the first weather satellite in 1960. Because of its complicated phase interchange with ambient water vapor (WV), its interaction with aerosols, and its large degree of freedom in its microphysical properties, retrieving cloud characteristics was never an easy task, let alone representing them in models for weather forecast or climate projection.

Manuscript received 9 May 2023; revised 17 September 2023; accepted 2 October 2023. Date of publication 2 November 2023; date of current version 13 November 2023. This work was supported by American University through NASA under Contract 80NSSC22K1763. Efforts at the NASA Goddard Space Flight Center were supported through the NASA CloudSat-CALIPSO Science Program, the U.S. Principal Investigator Program, and the TSIS Science Program. *(Corresponding author: Leah Ding.)*

Yiding Wang and Leah Ding are with the Department of Computer Science, American University, Washington, DC 20016 USA (e-mail: yw2686a@american.edu; ding@american.edu).

Jie Gong and Dong L. Wu are with the NASA Goddard Space Flight Center, Greenbelt, MD 20771 USA (e-mail: jie.gong@nasa.gov; dong.l.wu@nasa.gov).

Digital Object Identifier 10.1109/TGRS.2023.3329649

Among various cloud properties, the most useful one is the cloud mask, which indicates the presence of a cloud or not. However, this is rather a complicated subject than it appears from at least two perspectives. First, cloud detection is "observer-dependent," meaning that it is sensitive to the instrument sensitivity and viewing conditions. A passive visible or infrared sensor can "see" thin clouds that passive microwave sensors are not sensitive to at all, but the former faces challenges in collecting cloud signals from the cloud bottom. For the same sensor, oblique views enable a greater chance of seeing a thin cloud because the signal integration length is longer to stand out from noise. Second, from a satellite bird's-eye view, the definition of "cloud mask" is mostly 2-D, loosing the important vertical dimension of information. For example, for the Advanced Baseline Imager (ABI) that we are going to use in this work, the National Oceanic and Atmospheric Administration (NOAA)'s official cloud mask product only provides cloud top height/temperature and cloud phase information at instrument pixel level (or multipixel averaged level). In addition, the official product assigns "overlapping" type to multilayer cloud without specifying the overlapping structures. Huang et al. [2] compared collocated operational AHI (the equivalent of ABI on Japanese geostationary satellites) cloud top height product with that from the CALIPSO lidar and found poor performance for warm liquid, cirrus, and overlapping clouds. Fmask model [3] uses a Landsat thermal band to estimate cloud height and projects them into shadow candidates for cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 images.

Spaceborne active sensors, such as aforementioned CALIPSO lidar and CloudSat radar launched in the recent two decades, collect an unprecedented detailed record of cloud vertical structures. However, as they are only afforded to scan the nadir, their cloud mask products are essentially still 2-D cross sections. The caveats of the passive and active instruments hence give us a strong motivation for this work to reconstruct the 3-D cloud masks leveraging the merits of the two.

Cloud 3-D masks could be useful for many areas, from aviation safety to atmospheric motion vector retrievals as some possible downstream products. Leinonen et al. [4] developed generative adversarial networks (GANs) to predict the vertical overlapping situation from passive MODIS images by learning from CloudSat radar vertical slices through clouds. This work is motivated to help improve the cloud distribution schemes in global climate models (GCMs). As some other pioneer works, Noh et al. [5], Seaman et al. [6], Haynes et al. [7], and Noh et al. [8] tried to retrieve global cloud 3-D masks

from ABI data using a machine learning approach. Collocated CloudSat observations are used as the "ground truth," and a traditional machine learning random forest was employed. Their output products include eight layers of cloud flags and cloud top and bottom heights if the layer is marked as cloud presence. Using similar training and "truth" dataset, this work takes a rather different approach by designing a cloud-physics tailored loss function. Moreover, we take a step forward to retrieve a cloud mask at 38 levels with a 500-m vertical resolution (covering a range of 0–19 km), allowing us to generate a true 3-D cloud mask.

In this work, we introduce a physics-informed function for training deep learning models that are capable of extending downward 2-D cloud images to 3-D cloud masks. The experimental results show that the physics-informed loss function can modulate the training phase of a deep learning model to explicitly favor convergence toward solutions that adhere to the underlying natural phenomena of cloud distribution.

Specifically, we propose a weighted physics-based loss function that can be expressed in the form of integral and partial differential equations so that it can be efficiently solved by deep neural networks. We present the design of two domain knowledge informed loss terms, where one loss term evaluates the accuracy of cloud position and thickness, and the other loss term measures the accuracy of the number of layers. The proposed loss function is a weighted sum of these two loss terms, with the weights governing the balance between them. These weights can be user-defined or fine-tuned to enhance the trainability and overall performance of the deep learning model.

To summarize, the main contributions of this work are given as follows.

1) We propose a novel loss, called the CloudMask loss, specifically designed to explicitly favor the accurate reconstruction of multilayer cloud structures in terms of their vertical position, number of layers, and thickness for neural networks.
2) We demonstrate that the proposed loss significantly enhanced the quality of multilayer cloud mask reconstructions, outperforming the standard binary cross-entropy (BCE) loss in producing higher accuracy of multilayer classification, number of layers, and lower thickness mean absolute error (MAE).
3) The proposed loss can be easily incorporated into different neural networks, resulting in significant improvements in performance for multilayer cloud mask reconstruction.

The remainder of this article is organized as follows. In Section II, we briefly review related work on 3-D object reconstruction in computer vision and 3-D field retrieval in atmospheric science. The data and data preparation are described in Section III. In Section IV, we introduce the design of CloudMask loss function based on the physical characteristics of clouds. To demonstrate the effectiveness of the proposed CloudMask loss, we evaluate the performance of CloudMask loss using three different deep learning models and present experimental results in Section V, followed by the discussion in Section VI and conclusions in Section VII.

II. Related Work

A. 3-D Object Reconstruction

In computer vision, understanding and reconstructing the 3-D world from 2-D images has wide applications, such as autonomous vehicles where 3-D mapping of the 2-D environment images for obstacle avoidance and navigation, augmented reality where virtual objects are positioned and overlaid in the 3-D space, and medical imaging where 3-D models of anatomical structures from multiple 2-D images are used for diagnosis and surgical planning. A common solution to reconstruct the 3-D object from 2-D images is to leverage multiple images captured from different viewpoints to extract the depth information to create a 3-D representation or structure of the object. For example, Yan et al. [9] proposed a deep learning framework for 3-D object reconstruction from single-view 2-D images without using the ground-truth 3-D volumetric data for training.

The proposed neural network, called perspective transformer nets (PTNs), learns to predict a 3-D transformation matrix that can transform the input 2-D image into a canonical viewpoint, which is then passed to the 3-D decoder module to generate the volumetric 3-D shape. As another example, Kim et al. [10] proposed a GAN to generate 3-D smoke volumes from 2-D sketches.

Light reflections and refraction-based methods have been studied to address the challenge of 3-D reconstruction of dynamic fluid surface [11], [12], [13] and gas flows [14], [15], [16]. These methods typically use a known pattern placed underneath the fluid body, with one or more cameras capturing the reference pattern through the dynamic flow. Single-camera methods often impose additional surface assumptions. Multiple cameras, capturing from various angles, mitigate this by ensuring cross-view consistency. For instance, Qian et al. [11] used this to capture the wavy appearance of a pregenerated random pattern and then estimated correspondences between the captured images and the known background by tracking the pattern.

However, for 3-D cloud mask reconstruction, the data corresponding to the 2-D image are the forever-downward view taken by the Advanced Baseline Imager (ABI) from space with fixed view angle at a given location at the ground. The ground truth of vertical profiles from the CloudSat/CALIPSO is the cross-sectional image that is perpendicular to the downward view, thus sparse. Moreover, cloud 3-D structures have significantly larger variations than that of a rigid object (e.g., the chair shape) applies to most chairs in daily life). Therefore, 3-D cloud mask reconstruction is undoubtedly more challenging than the usual computer vision tasks in 3-D reconstruction [9].

B. 3-D Field Retrieval in Atmospheric Science

The retrieval of the 3-D atmospheric field in geo-coordinates is always an important and challenging topic. Currently, the majority of satellite retrieval algorithms for atmospheric variables only considers radiative transfer along the instrument line-of-sight (LOS) and usually assumes a plane-parallel atmosphere to avoid costly 3-D radiative transfer computations. The

3-D field hence practically comes from the combination of many independent pixel-level retrieved profiles. Tomography aims to recover a 3-D density map of a medium or an object. By taking measurements of the same object (e.g., a cloud) from different view angles that are enabled sometimes during flight campaigns, the tomographic approach can be applied to recover the 3-D object properties unambiguously (e.g., [17]). Tomographic reconstruction is also widely used in other fields, e.g., body mass reconstruction from a CT scan. Forster et al. [18] applied the tomographic retrieval approach to two joint satellite observations, with one scanning at different view angles. Overall, this approach is likely the most accurate and best constrained by physics; however, it is also among the most computationally expensive approaches, which limits its application for operational use.

Machine learning has been trending in the Earth science fields in the past decade. The large volume of daily observations collected by satellites makes it an ideal playground for exploring various machine learning methods. Deep learning for remote sensing has been investigated. Zhu et al. [19] provided a review of recent advances as well as the challenges of using deep learning for remote sensing data analysis. However, by far, the overwhelming majority of deep learning applications in this field employ off-the-shelf loss functions. Unfortunately, these standard off-the-shelf loss functions are designed without considering geophysical parameters. The unique features of geophysical parameters (e.g., often highly imbalanced, with extreme values that require more attention, and continuity) require the tailored design of machine learning methods. In fact, Ding et al. [20] demonstrated the importance of data rebalancing for fair predictions of multilayer clouds using the same dataset in this work. Moreover, traditional machine learning models do not necessarily obey the fundamental governing laws of physical systems nor do they generalize well to scenarios on which they have not been trained. Physics-informed neural networks (PINNs) [21], [22] have recently emerged as a promising approach that combines the strengths of deep learning and physics-based models to overcome the limitations of traditional physics-based models in atmospheric science. Kashinath et al. [23] explored incorporating physics and domain knowledge into machine learning models for weather and climate modeling, showing that such approaches can achieve better physical consistency, reduced training time, improved data efficiency, and better generalization. Several hybrid approaches have combined physical model of clouds and neural networks for cloud detection. For example, Li et al. [24] proposed GANs and a physical model of cloud distortion (CR-GAN-PM) for thin cloud removal.

In this work, we propose a physics-informed loss function as a step toward developing general PINNs specifically designed for multilayer cloud mask generation. By incorporating cloud structures directly into the loss function, neural networks are able to reward predictions that agree with the natural ground truth and penalize those that disagree during the training process. This approach aims to improve the performance of deep learning models for more accurate and reliable cloud mask generation.

TABLE I
ABI CHANNELS AND RESOLUTION

Band	Channel	Function	Resolution
1	0.47 μm	Blue	1 km
2	0.64 μm	Red	0.5 km
3	0.86 μm	Veggie	1 km
4	1.38 μm	Cirrus	2 km
5	1.61 μm	Snow/Ice	1 km
6	2.25 μm	Cloud Particle Size	2 km
7	3.90 μm	Shortwave Window	2 km
8	6.18 μm	Upper-Level Water Vapor	2 km
9	6.95 μm	Mid-Level Water Vapor	2 km
10	7.34 μm	Lower-Level Water Vapor	2 km
11	8.50 μm	Cloud-Top Phase	2 km
12	9.61 μm	Ozone	2 km
13	10.35 μm	Clean IR Longwave Window	2 km
14	11.20 μm	IR Longwave Window	2 km
15	12.30 μm	Dirty Longwave Window	2 km
16	13.30 μm	CO2 Longwave Infrared	2 km

III. DATA

A. Features

In this work, we focus on retrieving 3-D cloud masks using observations from the Advanced Baseline Imager (ABI) in the East Pacific (10°S–10°N, 90°W–150°W), as delineated by the red rectangle in Fig. 1. This region is well known for being the womb of the El Nino/La Nina events. This area is often covered with multilayer clouds and it is very hard to separate the low-, middle-, and high-level clouds from passive satellite imagery data (e.g., ABI). When strong convective systems are absent, clouds in this area predominantly consist of boundary layer low clouds and trade cumulus. Previous efforts on predicting cloud height and layers using machine learning techniques on ABI data have encountered difficulties in this region (see [7]). This area is also chosen because of relatively stable surface emissions, which could otherwise introduce additional noise in cloud signal discrimination.

ABI, onboard the NOAA geostationary GOES-17 satellite, conducts full-disk scans of Earth every 10 min, producing sensing-based images [25]. It has 16 different spectral bands, including two visible channels, four near-infrared channels, and ten infrared channels, as listed in Table I. A complete list of band frequency and fact sheets can be found in [26], [27], and [28]. During nighttime, visible and near-infrared channels are unavailable, leaving only the infrared channels to provide useful information. We retain the original values of these channels as input to the model, and both daytime and nightitime data are used for training. Different channels have varying resolutions (0.5–2 km) and sensitivities to diverse atmosphere or surface features. For this study, all 16 channel features are utilized. For each channel, the footprint size discrepancy is resolved by interpolating to a uniform 1-km resolution, which roughly matches the footprint size of CloudSat.

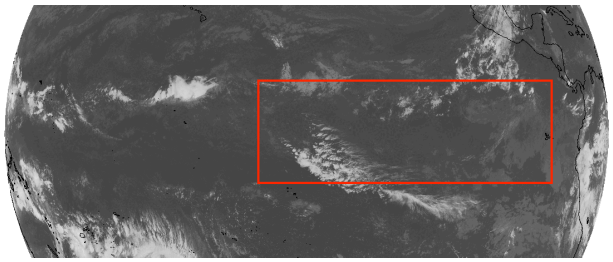

Fig. 1. Sensing-based image example (2022.364 23:00), showing the region of interest—the East Pacific (10°S–10°N, 90°W–150°W), highlighted by the red box.

CloudSat is a spaceborne W-band radar operating at 94 GHz on a polar-orbiting satellite within the A-Train constellation, crossing the Equator at 1:30 A.M. and 1:30 P.M. local solar time. Unfortunately, due to a battery anomaly, CloudSat only functioned during daytime when GOES-17 was launched in March 2018 [29]. Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) is a spaceborne lidar onboard the CALIPSO satellite, which co-flies with CloudSat in the same A-Train orbit. Our dataset was created by collocating data from CloudSat-CALIPSO with ABI, not with other A-Train instruments. By collocating CloudSat-CALIPSO with ABI, we obtain the "ground truth" of the cloud vertical structure, from very thin cirrus to very thick deep convective clouds, except in heavy precipitation regions close to the surface where radar backscatter cannot penetrate further. The synergistic cloud mask product 2B-CLDCLASS-LIDAR is used as the training "ground truth" [30]. This product has a horizontal resolution of 1.4 (across-track) by 2.5 (along-track) kilometers. Fig. 2 shows an example view of the vertical structure of multilayer clouds from the Earth's surface in coordinates of time, latitude, longitude, and height. However, such a high vertical resolution cross section is only available at nadir, which can be compiled to yield cloud statistics (e.g., [31]) but is barely useful for weather applications such as weather system tracking and evolution monitoring. The dataset lacks uncertainty estimation, which is a limitation. However, several studies have evaluated the dataset's quality and acknowledged CloudSat/CALIPSO as the most dependable global cloud vertical structure dataset [31], [32].

Because cloud formation only occurs when WV is saturated due to low temperatures or abundant WV amount, ancillary atmosphere fields from the CloudSat ECMWF-AUX dataset are included as inputs in addition to ABI observations for the deep learning models. These variables are 2-m temperature (T2m), U component of the 10-m wind (U10m), V component of the 10-m wind (V10m), temperature (T), specific humidity (SH), and pressure data in ECMWF-AUX. Based on temperature, SH, and pressure, relative humidity (RH) is calculated [33] as additional information during training. All variables are listed in Table II.

When predicting beyond the collocation tracks, the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) reanalysis is used to extract the ancillary information. While MERRA-2 and ECMWF-AUX data may not be entirely consistent, it turns out that the discrepancy has minimal impact on prediction quality.

TABLE II
LIST OF FEATURES. VARIABLES THAT DEPEND ON HEIGHT ARE VECTORS, REFERRED TO AS "PROFILES"

Variable	Length	Name
Ch	16	ABI 16 channels
T_{2m}	1	two meter temperature
U10	1	U component of the 10 meter wind
V10	1	V component of the 10 meter wind
T	84	temperature of different altitude from 0-20 km
SH	84	specific humidity of different altitude from 0-20 km
RH	84	relative humidity of different altitude from 0-20 km
C	3	Coordinate: Sin/Cos Latitude/Longitude
t	6	Time: Sin/Cos day, Sin/Cos month, Sin/Cos year
Total	280	

When CloudSat radar passes over the region of interest, there is always a corresponding ABI pixel value if ±5-min temporal difference is allowed for the collocation. CloudSat/CALIPSO observations with ABI pixel values, including geographic location and timestamp, from October 2018 to July 2019, are collocated. When combining these two datasets, data samples with a time difference exceeding 500 s or a Euclidean distance larger than $5e^{-05}$ degree (about 788 m in the tropics) are dropped. The aliasing effect is neglected in this study as the selected region is close to nadir.

B. Multilayer Classes

Multilayer clouds refer to the composition of various clouds at different altitudes. Based on cloud top height information from the collocated CloudSat/CALIPSO data, we classify multilayer cloud into eight classes instead of the conventional, more limited three classes (low, mid, and high). The eight classes are clear sky (none), high, middle, low, high + mid, high + low, mid + low, and high + mid + low (all). High clouds have cloud top heights of 9.5 km or higher, while mid clouds have cloud top heights between 5 and 9.5 km. Low clouds are those at or below 5 km.

Fig. 3 shows the statistics for the first six visible and near-infrared channels of the collocated samples across the different classes. Dots within the red rectangles are considered outliers, likely representing measurements under sunglint conditions [34]. These abnormal values are replaced by the clear-sky radiance median, which is a common practice in machine learning field in dealing with outlier data.

The ocean surface's darkness results in significantly smaller median radiance measurements under clear-sky class than in the other seven classes when clouds are present. Later analysis revealed that this outlier substitution approach is effective for the first two visible bands (C01 and C02) but not for the four near-infrared channels (C03–C06). Adjusting outliers for the latter reduces the model's performance, so it is unnecessary to modify them based on the results.

C. Cloud Mask

To retrieve a cloud mask at 38 levels with a 500-m vertical resolution (spanning a range from 0 to 19 km), denoted as

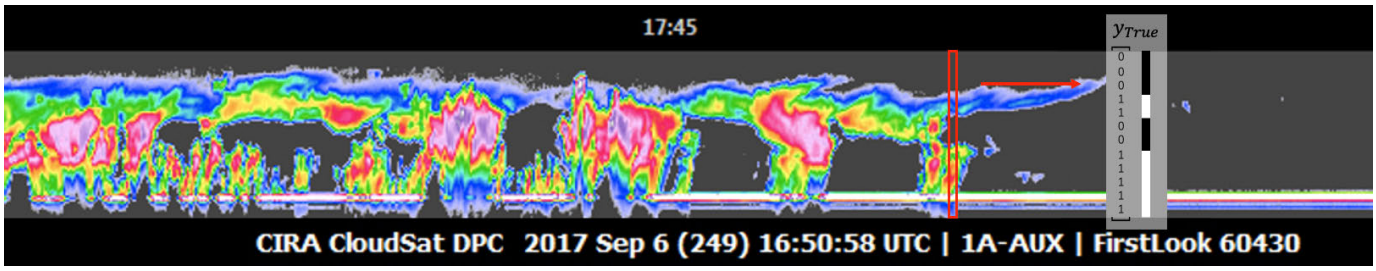

Fig. 2. Example of a cross-sectional scan of CloudSat. The horizontal axis represents geographical coordinates, while the vertical axis represents altitude. The colors represent the intensity of the radar signal (called the "return echo power"). The 2D-CLDCLASS-LIDAR mask product in the red rectangle is plotted to the right, with 0 (black) and 1 (white) corresponding to no-cloud and yes-cloud.

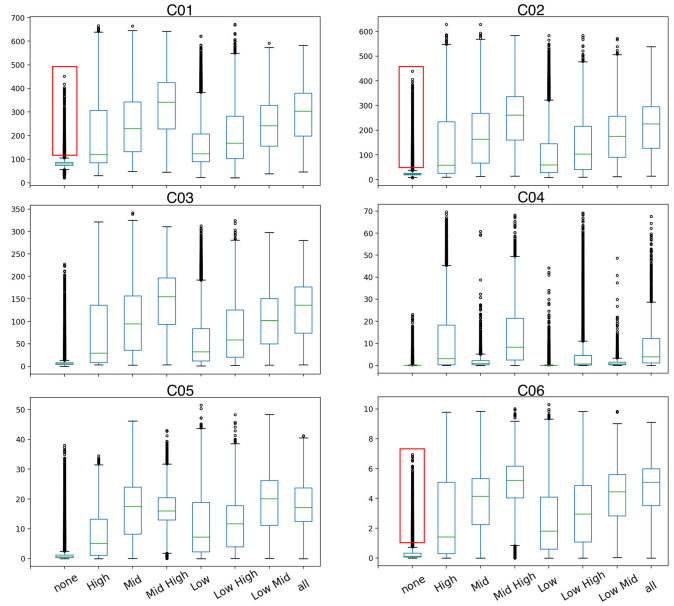

Fig. 3. General statistics of the collocated ABI radiances measurements for a selection of channels (C01–C06 at visible and near-infrared bands) under eight different cloud scenarios. The box plot method is used to show their quartiles. The box extends from the Q1 to Q3 quartile values of the data, with a line at the median (Q2). The whiskers extend from the edges of box to show the range of the data. The extend no more than 1.5 * IQR (IQR = Q3 − Q1) from the edges of the box, ending at the farthest data point within that interval. Outliers are plotted as separate dots. The outliers within the red rectangles will be replaced with the Q2 value.

a binary vector $\mathbf{y}$ of length 38, deep learning models can be trained to approximate the relationship f between the features, denoted as $\boldsymbol{x}$ and $\boldsymbol{y}$

$$\boldsymbol{y} = f(\boldsymbol{x}) \tag{1}$$

where $\boldsymbol{x}$ contains both the ABI and ECMWF-AUX datasets, as detailed in Table I

$$\boldsymbol{x} = \left[\mathbf{Ch}\ \ \mathrm{T}_{2m}\ \ \ \mathrm{U10}\ \ \mathrm{V10}\ \ \mathbf{T}\ \ \mathbf{SH}\ \ \ \mathbf{RH}\ \ \mathbf{C}\ \ \mathbf{t}\right]^{\mathrm{T}}. \tag{2}$$

The 2B-CLDCLASS-LIDAR dataset contains multilayer cloud information with cloud top and bottom height retrieved for each layer. The information is then converted to a binary vector $\boldsymbol{y}$, used as the "ground truth" during training, to indicate the presence of clouds within 500-m intervals (i.e., 1 for cloud present and 0 for cloud absent). Thus, this binary vector $\boldsymbol{y}$ represents cloud presence from sea level up to an altitude of 19 km with a vertical resolution of 0.5 km.

Fig. 2 shows the conversion of a sample (within the red box) into a vector of 38 binary values. As a classification problem, deep learning models predict the probability of cloud presence at 0.5-km vertical intervals. Using a 50% cutoff threshold, probabilities of 50% or higher result in a classification of "cloud presence" at the corresponding altitude, while probabilities lower than 50% result in "no cloud." There is existing work on adjusting threshold to handle imbalanced data such as [35]. In this work, we do not employ a height-dependent probability threshold to predict cloud masks nor analyze uncertainty. While these aspects could be interesting topics for future research, they are not the primary focus and are therefore excluded from the current investigation.

IV. CloudMask Loss

In deep learning, the loss function plays a crucial role in training a model. The loss function measures the discrepancy between the model's predictions and the true values, guiding the model optimization during training where the primary goal is to minimize the loss function value. We have developed a customized loss function, called CloudMask loss, that incorporates cloud physical structure knowledge in deep learning models for more accurate and meaningful results.

First, as clouds are radiatively significant at their tops, deep learning models should favor predictions with accurate cloud top positions. Second, clouds are more likely to be vertically contiguous than broken, so models should be optimized for contiguous geometric given the number of layers. Furthermore, as cloud thickness is largely proportional to the water content it contains and hence impacts the possible precipitation amounts, models should reward accurate predictions of column-integrated geometric thickness. All these physical constraints, originated from real-world cloud structures, are considered in the tailored loss function design, which will be explained in detail next.

To address these requirements, we have designed two terms: **Loss1**, which evaluates the predicted shape and location of clouds, and **Loss2**, which quantifies inaccuracies in the predicted number of cloud layers. By combining these two losses, we tackle the multitasking problem effectively.

To evaluate the predicted shape and location of clouds, we leverage the convolution to transform a binary vector $y(n)$ (i.e., $y(n)_{\text{true}}$ as true masks and $y(n)_{\text{predict}}$ as predicted masks) into a vector $s(n)$ with numerical values beyond the binary range

$$s(n) = (y * g)[n] = \sum_{m=1}^{n} y(m)g(n-m) \tag{3}$$

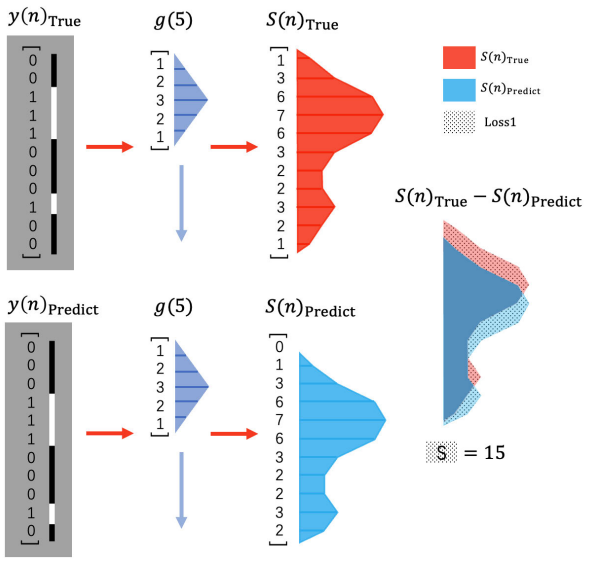

Fig. 4. Loss1: $y(n)_{\text{true}}$ and $y(n)_{\text{predict}}$ are transformed into $S(n)_{\text{true}}$ and $S(n)_{\text{predict}}$, respectively, through a kernel g with a length of 5, which means that it looks at the altitude of 1 km. By subtracting $S(n)_{\text{predict}}$ from $S(n)_{\text{true}}$, we obtain the shaded area S. The area of this shaded region is the loss value. This example shows $y(n)_{\text{predict}}$ that is shifted by 0.5 km toward the sea level compared to $y(n)_{\text{true}}$, resulting in a smaller loss value than in Fig. 5. N is set to 11 in this illustration.

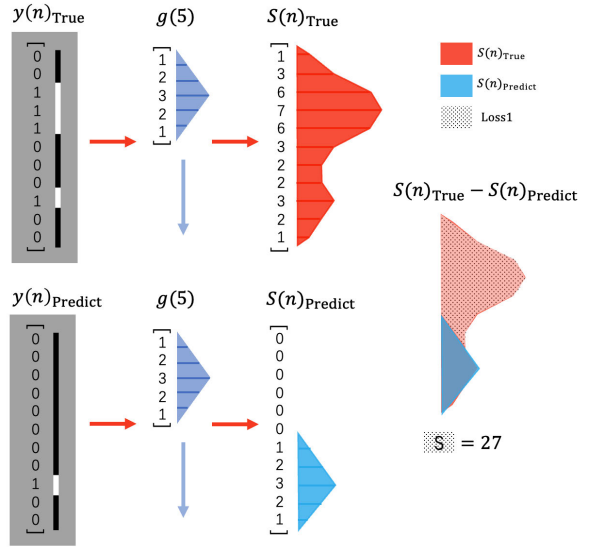

Fig. 5. Loss1: in this example, a thick high cloud is completely missing in $y(n)_{\text{predict}}$ compared to $y(n)_{\text{true}}$. This results in a larger loss value S than in Fig. 4. $N = 11$ in this illustration.

where g is the kernel. The summation runs over all possible positions m where the binary vector $y(m)$ and the kernel $g(n-m)$ overlap, computing the elementwise products and summing the results. This produces the output sequence $s(n)$, which can be used for comparing patterns.

For **Loss1**, as shown in the example in Fig. 4, we transform the binary vector according to the physical presentation of a cloud's shape and position. The length of kernel function $g(n)$ determines the acceptable deviation range of clouds.

In the experiments, for residual network (ResNet), the function kernel is set as $g(n) = [1\ 2\ 3\ 2\ 1]$, where the length is 5, and predicted clouds deviating more 1 km from the true position will have the maximum loss value as a

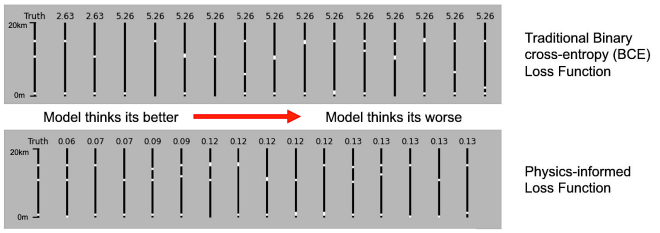

Fig. 6. Loss comparison: loss ranking using BCE and CloudMask loss function. The value above each reconstructed mask represents the loss value, arranged in ascending order. The y-axis represents altitude. BCE compares the number of different values between the predicted and true values, while the physics-informed CloudMask loss tolerates the predicted differences caused by small displacements and penalizes predictions with larger dissimilarities.

penalty for such deviation. Note that the filter is defined by two hyperparameters, the kernel's length and shape that specifies the values of the filter. These two hyperparameters should be tuned based on the dataset and deep learning model. Unlike BCE loss, which cannot be tuned across datasets and models, CloudMask loss has the advantage of being customizable and optimizable based on the specific dataset and model being used.

By setting the kernel function g, the probability of cloud presence at the target location is transformed from binary vector. In other words, if there is a larger chance for a cloud to appear at a given height, after the transformation, the probability at this height will be larger. The value of Loss1 is the sum of squared differences between $s(n)_{\text{true}}$ and $s(n)_{\text{predict}}$, represented by the shaded area in Fig. 4. Here, we use a square function to handle potential negative values in the matrix. The first task thus aligns with our understanding of cloud appearance probability. During the training process, the smaller the shaded area, the lower the value of Loss1.

In the example in Fig. 4, the predicted values are consistently 0.5 km lower than the true values. In Fig. 5, the predicted values are missing the prediction for the middle cloud layer, which leads to a higher loss value

$$\textbf{Loss1} \stackrel{\text{def}}{=} \sum_{n=1}^{N} \left(s(n)_{\text{true}} - s(n)_{\text{predict}} \right)^2 \tag{4}$$

where $g(5) = [1\ 2\ 3\ 2\ 1]$ in (3) and $N = 38$.

It is worth noting that $y(n)_{\text{predict}}$ is a continuous value within $[0, 1]$. However, in Figs. 4 and 5, for illustration purpose, we deliberately set $y(n)_{\text{predict}}$ to 0 and 1 s for a more straightforward comparison with $y(n)_{\text{true}}$.

Loss2 is designed to measure inaccuracies in the predicted number of cloud layers. We first use the convolution kernel $g(2) = [0\ 1]$ as the edge pattern, i.e., a transition from no cloud (0) to cloud presence (1), to transform binary mask vectors, such as $y(n)_{\text{true}}$ and $y(n)_{\text{predict}}$. By subtracting the transformed values from the original binary mask vectors, we can determine the number of cloud layers. We then calculate the difference to find the discrepancies in the number of layers between the true and predicted vectors. Note that with $g(2) = [0\ 1]$, the cloud top results in a positive value, whereas the cloud bottom yields a negative value. By applying the ReLU function, we keep the positive values and discard the negative ones. The ReLU function allows us to identify

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT MODELS WHEN TRAINED USING DISTINCT LOSS FUNCTIONS: BCE LOSS, FOCAL LOSS, AND CLOUDMASK LOSS $\mathcal{L}$

Metric	ResNet BCE	ResNet focal	**ResNet $\mathcal{L}$**	MLP BCE	MLP focal	**MLP $\mathcal{L}$**	CNN BCE	CNN focal	**CNN $\mathcal{L}$**	Transformer BCE	Transformer focal	**Transformer $\mathcal{L}$**
8-class accuracy	62.36%	67.06%	**69.60%**	61.93%	58.24%	68.01%	63.35%	63.68%	65.86%	65.36%	62.62%	68.08%
Number of layers accuracy	61.43%	65.53%	**67.95%**	61.00%	59.96%	66.33%	62.39%	63.55%	64.98%	64.72%	62.25%	66.78%
Thickness MAE	0.9122	1.8601	**0.8346**	0.9803	2.6576	0.8920	0.95429	1.6315	0.9265	0.9683	1.2655	0.9327
CloudMask loss $\mathcal{L}$	0.1275	0.26574	**0.1263**	0.1383	0.3838	0.1355	0.1375	0.2376	0.1354	0.1526	0.1920	0.1486
IoU	0.5528	0.47205	**0.5848**	0.5365	0.3533	0.5333	0.5417	0.4484	0.5453	0.5488	0.4548	0.5451

TABLE IV
CLOUDMASK LOSS $\mathcal{L}$ ABLATION EXPERIMENT

Metric	ResNet-$\mathcal{L}$	ResNet-**Loss1**	ResNet-**Loss2**
8-class accuracy	69.60%	65.23%	66.35%
Number of layers accuracy	67.95%	61.08%	67.21%
Thickness MAE	0.7794	0.8458	2.4840
$\mathcal{L}, w = 0.9$	0.1263	0.1284	0.4127

the cloud top's location and count the number of cloud tops

$$\textbf{Loss2} \overset{\text{def}}{=} \sum_{n=1}^{N} \Big(\max\left(0,\, y(n)_{\text{true}} - s(n)_{\text{true}}\right) - \max\left(0,\, y(n)_{\text{predict}} - s(n)_{\text{predict}}\right) \Big)^2 \quad (5)$$

where $g(2) = [0,\, 1]$ and $N = 38$.

For the CloudMask loss $\mathcal{L}$, we use a hyperparameter w to balance **Loss1** and **Loss2**. According to different requirements, we can set different values w. For example, $w = 1$ focuses on the accuracy of the predicted number of cloud layers. The value of w can be optimized within the range of 0–1, allowing for a balance between different loss components in the model

$$\mathcal{L} \overset{\text{def}}{=} (1 - w) \cdot \textbf{Loss1} + w \cdot \textbf{Loss2}. \quad (6)$$

Fig. 6 shows a comparison of the results obtained using the standard BCE loss (top row) and CloudMask loss (bottom row). Each line represents a reconstructed multilayer cloud vertical mask at the same longitude and latitude. In this example, there are three thin layers of clouds occupying upper, middle, and lower troposphere (the leftmost bar is the truth from CloudSat/CALIPSO). The standard BCE loss values (numbers on the top) cannot differentiate between various wrong predictions, as their loss values are similar. Specifically, there are three thin layers of clouds (high, middle, and low), and the standard BCE loss function is unable to distinguish between wrong predictions of high + low clouds (the second bar from the left in the top row) and high + mid + low clouds (the second bar from the right in the top row), although from radiative perspective, the latter should be favored.

In contrast, the CloudMask loss (bottom row of Fig. 6) effectively ranks predictions based on their similarity to the true mask. The samples with lower loss values (closer to the left) closely reconstruct the truth, while predictions further to the right with higher loss values appear less similar. The CloudMask loss incorporates key physical parameters that can be adjusted to accommodate different objectives. It considers cloud thickness and position and adds penalties for incorrect layer counts. Moreover, hyperparameters in the CloudMask loss function can be adjusted to adapt to different situations when evaluating cloud shape and location.

V. EXPERIMENTAL EVALUATION

The design of CloudMask loss is evaluated using three different deep learning models described in Section V-B. We observe consistent improvements in performance when comparing the CloudMask loss to the BCE loss.

A. Training, Validation, and Test sets

Collocated CloudSat-ABI data from October 2018 to July 2019 is used, which has a total of 876 044 samples. January 2019 is intentionally set aside as an independent test set, which contains 108 337 samples. Data from the second half of December 2018 are used as the validation set. The remaining data serve as the training set. The random split method for training and validation is not employed as the high coherence of adjacent profiles raises concerns about potentially undermining the model's ability to generalize effectively.

B. Deep Learning Models

In this work, we use three different deep learning models—multilayer perceptron (MLP), convolutional neural network (CNN), ResNet, and Transformer—to showcase the robustness of the proposed CloudMask loss function across various model types.

MLP [36] is a fully connected class of feedforward artificial neural networks. It consists of multiple layers of interconnected neurons, where each neuron processes the inputs from the previous layer and transmits the results to the next layer.

CNN is a type of deep learning model that employs convolutional layers to automatically and adaptively learn spatial hierarchies of features, making it highly effective in capturing complex spatial dependencies within input data.

ResNet [37] is a network that uses residual blocks, which allow for the training of very deep neural networks without suffering from the vanishing gradient problem.

The transformer model [38] is a neural network-based attention mechanism. It has had a profound impact on the field of natural language processing (NLP) and has since been applied to various other domains as well.

C. Hyperparameters

We experimented with deep learning models of various sizes to find the optimal balance between performance and model complexity. These models, such as Resnet 34, 50, and 101, exhibited different levels of performance. As model size increased, the performance improved, but the gains diminished beyond a certain point. In our experiments, Resnet-50 is shown to be a suitable choice for this dataset. Resnet-101, compared

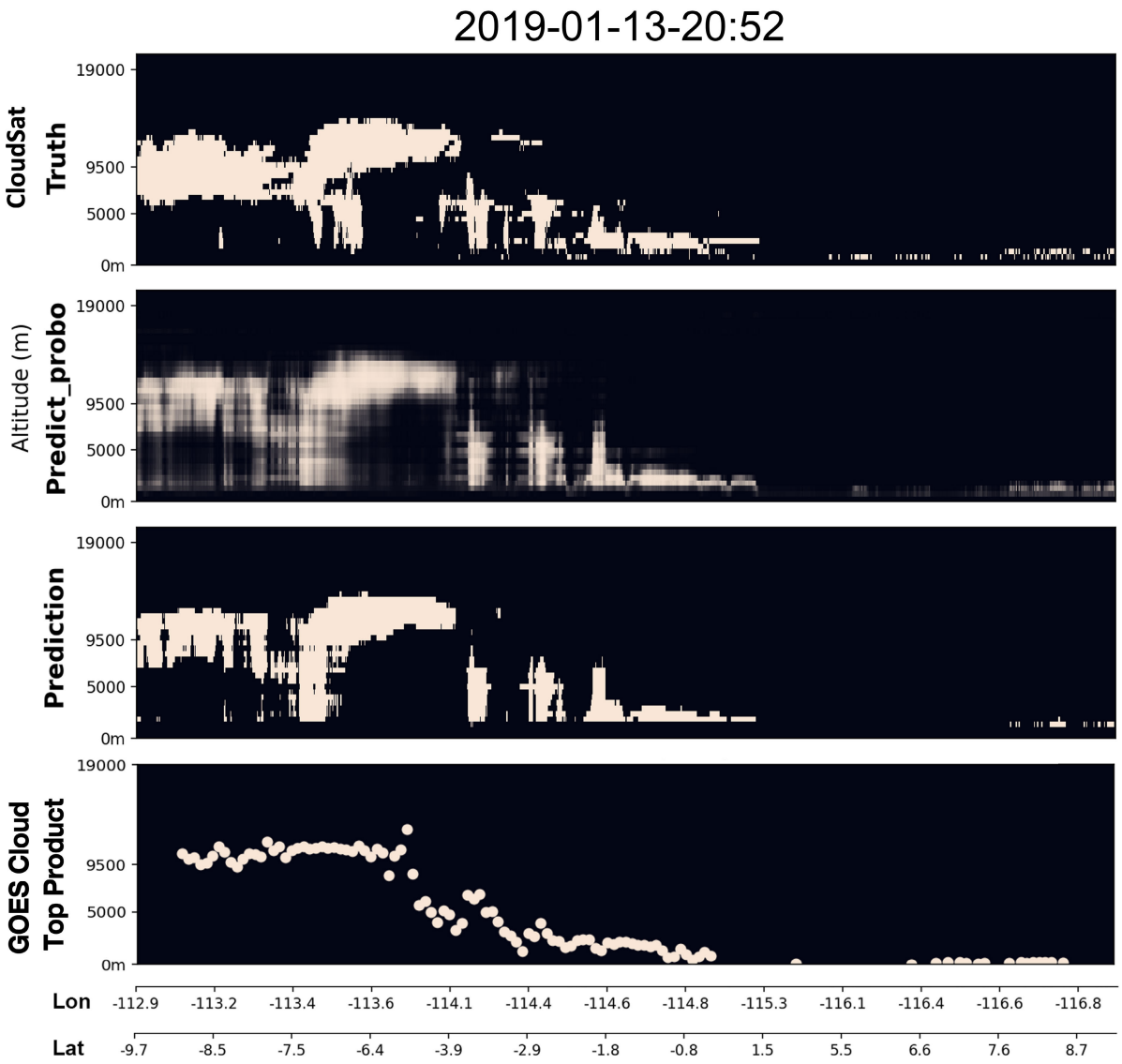

Fig. 7. Curtain of vertical cloud mask along two segments of CloudSat/CALIPSO orbit on January 13, 2019, with the "truth" on the top panel, the predicted probability in the second panel, and the predicted masks in the third panel generated by ResNet-$\mathcal{L}$, and current GOES cloud top height product [40] in the bottom panel.

to Resnet-50, the performance improvement is marginal (less than 0.5%). However, when extending the current design globally, a larger model may be required to handle increased complexity. For example, if we extend the current design globally, we might need a larger Resnet model. Dropout rates ranging from 0 to 0.5 were tested for reducing overfitting and improving the generalization of the model, and 0.15 is chosen. The Nesterov-accelerated adaptive moment estimation is used as the optimizer. Learning rates from $1e^{-}2$ to $1e^{-}5$ were tested, with $3e^{-}4$ being the final selection. To prevent overfitting, early stopping was used during training, which means that training will be stopped when the validation loss of the model stops decreasing. For the MLP model, a hidden layer structure of [196, 128, 128, 128, 38] is used, with ReLU activation function and a dropout rate of 0.1 for each layer. For the CNN model, it includes four 1-D convolutional blocks. Each block consists of a Conv1d layer, a MaxPool1d layer, a BatchNorm1d, and ReLU activation function. Each Conv1d layer has an output channel size of [32, 64, 64, 128] and a kernel size of 3. The MaxPool1d layer has a kernel size of 2. The results are then flattened and passed into an MLP structure of size [256, 256, 38], with a dropout layer using a dropout rate of 0.15. For the transformer, it contains two encoder layers. Each encoder layer has 64 expected features, eight heads in the multihead attention, and 256 dimensions of the feedforward network with a 0.1 dropout rate. Using [1, 1, 1] as CloudMaskLoss kernel can get the best performance for Transformer model. The numbers of model parameters of ResNet, MLP, CNN, and Transformer are 339 640, 181 755, 546 090, and 490 406, respectively.

D. Evaluation Metrics

Following the rationale behind the design of CloudMask loss, the evaluation metric should not be standard, but instead tailored to fit the specific analytical needs.

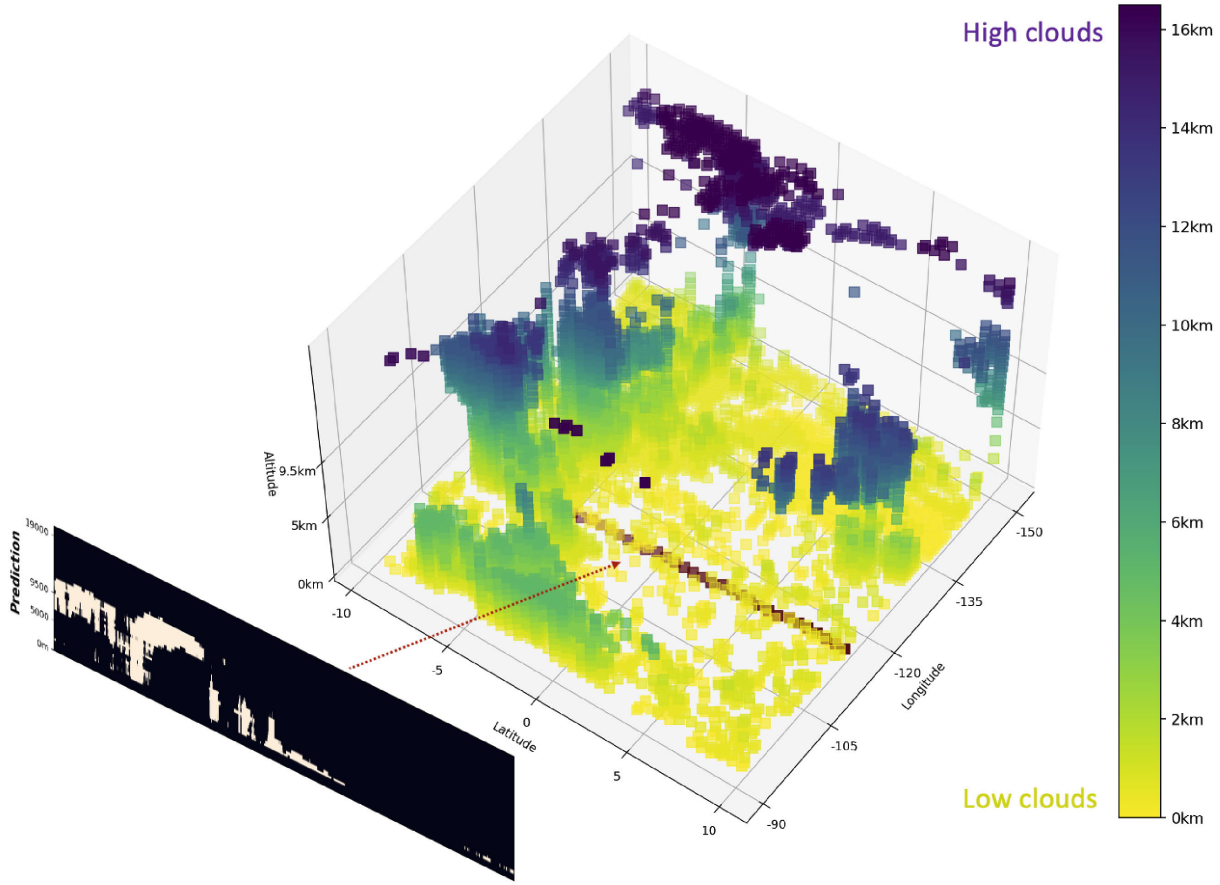

Fig. 8. Predicted the 3-D cloud mask for 21:00 January 13, 2019, longitude from −150° to −90°, latitude from −10° to 10°, using 16 channels of ABI information and temperature and SH data from MERRA-2 passive sensors. The red line represents the position of the scanning line in Fig. 8. The color of the image indicates the height of the clouds, with dark blue representing high-altitude clouds (16 km) and light yellow representing low-altitude clouds (0 km).

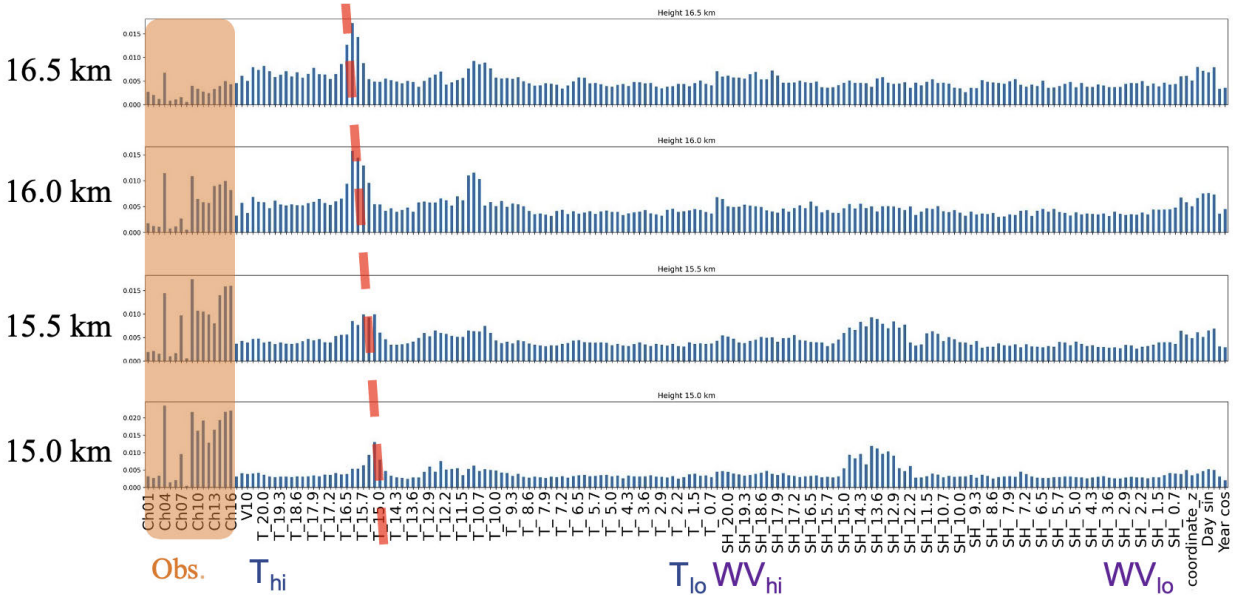

Fig. 9. Importance of features for high cloud prediction. From left to right, the order of features are 16 ABI channels from Ch01 to Ch16, T2m, U10, V10, temperature **T** at 0–20 km, **SH** at 0–20 km, coordinates **C**, and time **t**—month: sin and cos, day: sin and cos, and year: sin and cos. Every third feature is displayed. This plot illustrates the ranking of features used in the model to predict high clouds where the temperature is more important (higher values of T_{hi} marked by the red dashed line). ABI observations from 16 channels are also important (marked by the orange semitransparent rectangles).

We employed five evaluation metrics: eight-class accuracy, number of layers accuracy, thickness MAE, intersection over union (IoU), and the CloudMask loss $\mathcal{L}$. The 38-D classification is downsized to an eight-class classification: 1) clear sky; 2) low cloud (cloud top height $\leq$ 5 km); 3) middle cloud (5 km $<$ cloud top height $<$ 9.5 km); 4) high cloud

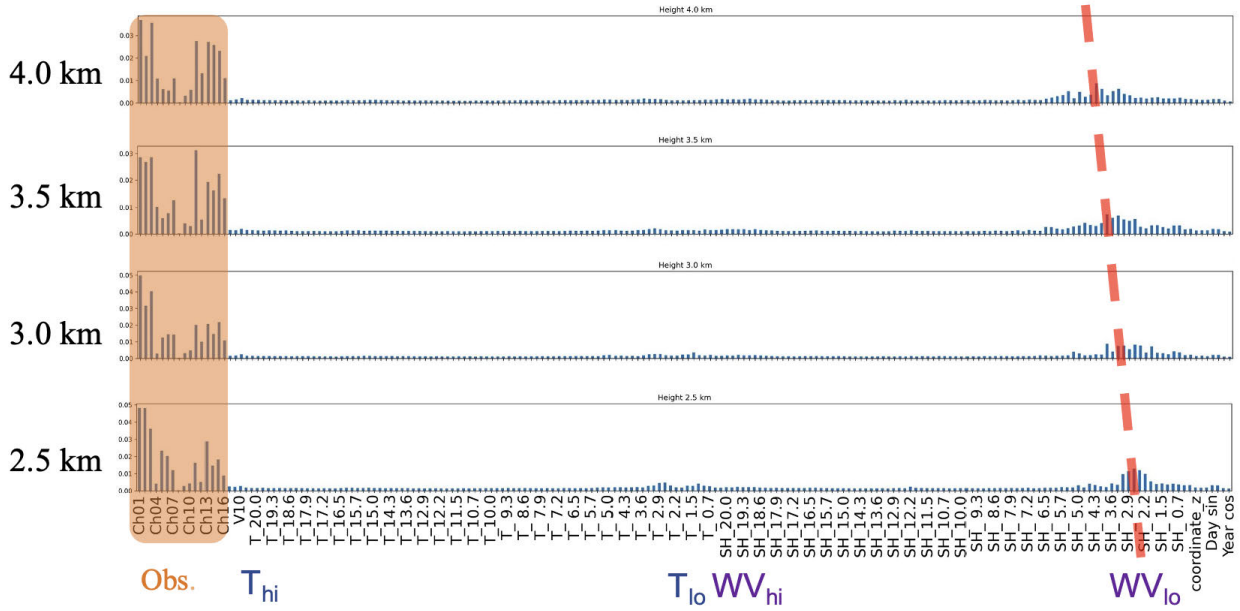

Fig. 10. Importance of features for low cloud prediction. This figure displays the feature importance rankings for low cloud predictions where WV is more important (higher values of WV_{lo} marked by the red dashed line). Similar to high cloud prediction in Fig. 9, ABI observations from 16 channels are important as well (marked by the orange semitransparent rectangles).

(cloud top height $\geq$ 9.5 km); 5) low + middle (overlapping low and middle clouds); 6) low + high (overlapping low and high clouds); 7) mid + high (overlapping middle and high clouds); and 8) low + mid + high (overlapping low, middle, and high clouds). Such a downsize approach is for coarse multilayer classification and for a more straightforward understanding of model performance under the context of traditional cloud diagram.

The number of layers accuracy refers to the accuracy of predicting the number of layers. The thickness MAE is the mean absolute error between the predicted cloud layer's geometric thickness and the ground truth. In the presence of multilayer clouds, the thickness of each layer is summed to obtain the total thickness.

IoU measures the overlap between the prediction and the ground-truth region of an object. IoU ranges between 0 and 1, with 0 indicating no overlap and 1 indicating a perfect match between the prediction and ground truth. Higher IoU values typically indicate better localization accuracy.

The CloudMask loss $\mathcal{L}$ is computed using the same value of $w = 0.9$ as during the training process.

E. Results

We first compared the performance of different models when trained with BCE loss, focal loss [39], and CloudMask loss. The results for different metrics are listed in Table III. The best model is highlighted in bold. As shown in Table III, there is an improvement of about 7% on eight-class accuracy when ResNet and MLP are trained using CloudMask loss compared to BCE loss. For the CNN model, the CloudMask loss improves the eight-class accuracy by about 2.5%. In terms of the accuracy for a number of layers, models trained with CLoudMask loss outperform those trained with BCE loss by 2.5%–6%. For the cloud's thickness, which is physically important for representing an unbiased global hydrological cycle, the CloudMask loss reduces the thickness MAE.

The ResNet-$\mathcal{L}$ model is selected for the following 3-D reconstruction case study due to its overall better performance compared to others.

Adjusting the loss weight w with different values affects the model's optimization. Table IV lists the results for two extreme situations compared to the optimized model performance (left column). The result of using only Loss1 (middle column) is obtained when $w = 0$, and the result of using only Loss2 (right column) is obtained when $w = 1$. As Loss2 is a loss specialized for measuring the number of cloud layers, the prediction accuracy on this particular metric is higher than that of Loss1. ResNet-Loss2 gives an accuracy that is 0.72% lower than that of ResNet-L when combined with Loss1. While Loss2 is specifically tailored to the number of layers, Loss1 helps the model in learning the overall cloud distribution, which can potentially improve the prediction of the number of layers.

The final selection for the value of w should be based on a comprehensive review of all four evaluation metrics, ensuring that the model performance achieves a balanced result among different physics constraints.

VI. Discussion

In this section, we give two examples to further demonstrate the model performance, especially when multilayer clouds or geometrically thick clouds are present. As a passive sensor, ABI is traditionally believed to have a difficult time for these two types of clouds [2]. We demonstrate that with knowledge learned from CloudSat/CALIPSO, we can retrieve unprecedented details of cloud vertical structures from ABI that traditional physics-based methods probably cannot retrieve.

The ResNet model is used for cloud mask reconstruction as it outperforms other models in the evaluation. The vertical

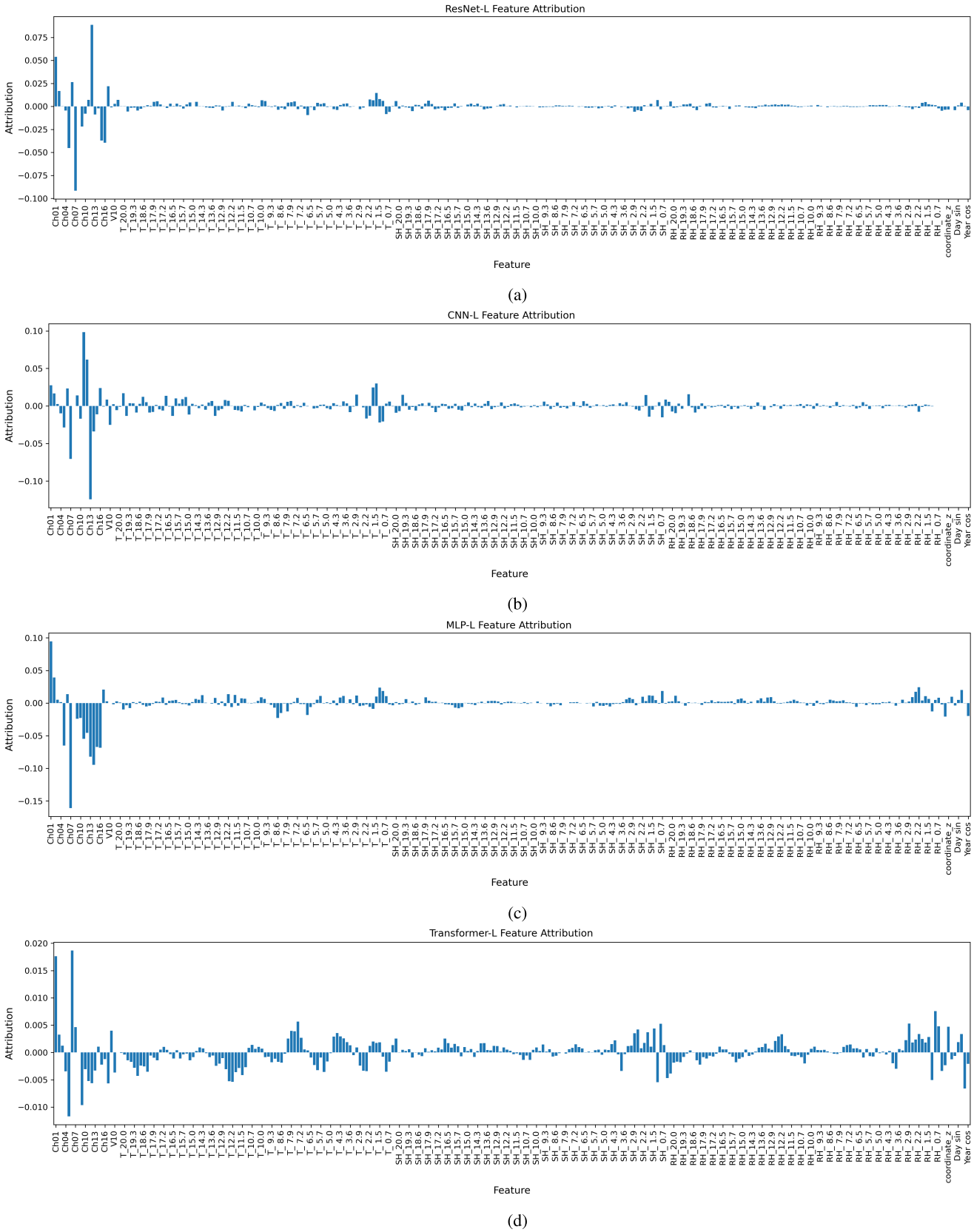

Fig. 11. Feature attribution plots of (a) ResNet-L, (b) CNN-L, (c) MLP-L, and (d) Transformer-L. From left to right, the order of features is 16 ABI channels from Ch01 to Ch16, T2m, U10, V10, temperature **T** at 0–20 km, **SH** at 0–20 km, **RH** at 0–20 km, coordinates **C**, and time **t**—month: sin and cos, day: sin and cos, and year: sin and cos. Every third feature is displayed.

cross section of cloud masks along two segments of CloudSat/CALIPSO orbit on January 13, 2019 is presented in Fig. 7 for the "truth" (top panel), the predicted probability (second panel), and the predicted mask (third panel). In this figure, we use a hard threshold (probability ≥ 0.5) to assign 0/1 as predicted cloud masks, which explains why the middle

and bottom panels look nearly identical. Multilayer clouds' attributes, such height, number of layers, and horizontal connectivity, are well captured in the prediction, even for thick clouds between 7.5∘S and 6.4∘S, and multilayer cloud at around 8.2∘S. The lowest level boundary layer warm clouds (5.5∘N–8.7∘N) are less distinct and broken in the prediction, but they exhibit relatively high probabilities in the middle panel that resembles the "truth" despite being at the marginal value. This is likely due to the misassignment of low cloud and surface. Similarly, the probability at cloud boundaries drops sharply because of ambiguous information between clear-sky and low-level thin clouds.

As a comparison to physics-based products, we show the current GOES cloud top height product [40] in the fourth panel, which lacks the vertical information below the cloud top. In addition, the current operational product produces retrieval for every ten footprints (i.e., 10-km resolution), while ours are at ABI's native resolution (1 km). More results for January 2019 test dataset can be found here.[1]

Furthermore, Fig. 8 shows the 3-D cloud structure for the previous case (the red line on the surface indicates the cross-sectional projection). Although the current approach is only applied to the pixel level, it effectively captures the horizontal continuity of cloud structures, including the high anvil cloud extended from deep convection, the middle-level trade cumuli, and the ubiquitous PBL warm clouds.

VII. Conclusion

To advance machine learning-based cloud 3-D mask retrieval, this article demonstrates the potential of passive sensors (e.g., ABI) for retrieving the cloud 3-D vertical structures that are based on a reliable, high-quality learning database (in this case, CloudSat + CALIPSO). Such 3-D reconstruction capability could not be easily achieved with the traditional radiative transfer-based retrieval framework, given the high dimension of nonlinearities that happened during the cloud formation and satellite signal integration processes. The methodology demonstrated in this study can be applied to other passive plus active sensor combinations.

The highlight of this work is the design of physics-informed customized loss function, which integrates physical priorities for cloud structures into the design of the machine learning model architecture and the evaluation metrics. The current work stops at pixel (or single profile) level. The generality of such a loss function design strategy to different instruments, different spatial/temporal domains, and different machine learning models adds some substantial merits to the current work.

Finally, a sanity check was carried out for the physical consistency using an explainable model (random forest), which emphasizes the importance of having simultaneous all-sky temperature and WV profile retrievals in order to better retrieve cloud vertical information from passive sensors. With the aid of machine learning technique and the advancement of new active remote sensing technologies, it is of high hope that this dream can be realized in the foreseeable future.

[1]https://zenodo.org/record/7865371

As ABI captures high-resolution images at high temporal resolutions, our next step is to extend the loss function to 2-D and subsequently to 3-D. This would allow us to associate the cloud spatial and temporal continuity, considering neighboring pixels across both spatial and temporal domains. In addition, different types and sources of uncertainty have been identified and various approaches have been studied to measure and quantify uncertainty in neural networks, as outlined by Abdar et al. [41]. Moving forward, we will analyze uncertainty to help us further evaluate the prediction uncertainties.

Appendix

Explainability is often a major challenge in applying a deep learning model to a physics problem because of the doubt of physical consistency during the training process. Deep neural networks are intricate nonlinear functions of their inputs. Understanding and interpreting these networks remains a challenging problem and an on-going active topic of research [42], [43].

We trained a random forest model, although its general performance is worse than other deep learning models. For instance, the random forest model has an accuracy of 60.27% for the eight-class classification, 59.29% accuracy for the number of layers prediction, a thickness MAE of 1.183, a CloudMask loss of 0.166, a BEC loss of 3.932, and an IoU of 0.515. However, it provides interpretability, which is enabled from a direct comparison between feature importance rank and our physical understanding of how cloud forms. The rankings are shown in Fig. 9 for the high clouds that top at 16.5, 16, 15.5, and 16 km and Fig. 10 for the clouds that tops at 4, 3.5, 3, and 2.5 km. Note that RH is used in the training, but since it blends the temperature and SH together, this feature is removed in the figures. ABI observations from 16 channels (marked by the orange semitransparent rectangles) undoubtedly dominant the decision processes for both high cloud and low cloud scenarios, with thermal infrared channels being more important for high cloud prediction and visible/near-infrared channels being more important for capturing the low cloud scenes. These are consistent with some previous findings (e.g., [44]). However, ambient atmosphere temperature structures are of equivalent importance to a high cloud prediction, while WV is more important for low cloud prediction. The peak altitudes of most important temperature or WV features coincide with the cloud altitudes (marked by the two red dashed lines). These findings again are consistent with the different cloud formation mechanisms for high clouds and low clouds. In the upper troposphere, cloud ice homogeneous nucleation is extremely sensitive to the in situ temperature threshold. In the oceanic boundary layer when temperature is usually homogeneously warm, low cloud formation is tied to the WV availability. Although temperature and WV profiles are extracted from the ECMWF analysis [45], these findings stress the importance of having simultaneous temperature and WV retrievals in order to accurately retrieve the presence of clouds.

To gain a deeper understanding of feature attributions of the models, we have applied the integrated gradients method [46] to the ResNet, CNN, MLP models, and transformer models.

Fig. 11 plots the attributions of features for each model. As shown in Fig. 11, ABI observations are important compared to other features for all models. Given the page limit and the primary focus of this article, we leave the in-depth study on model explainability with comparison of different models as future work.

REFERENCES

[1] *The Atmosphere and the Ocean*. Accessed: Mar. 14, 2023. [Online]. Available: https://teachersinstitute.yale.edu/curriculum/units/1994/5/94.05.01.x.html

[2] Y. Huang, S. Siems, M. Manton, A. Protat, L. Majewski, and H. Nguyen, "Evaluating Himawari-8 cloud products using shipborne and CALIPSO observations: Cloud-top height and cloud-top temperature," *J. Atmos. Ocean. Technol.*, vol. 36, no. 12, pp. 2327–2347, Dec. 2019.

[3] S. Qiu, Z. Zhu, and B. He, "Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery," *Remote Sens. Environ.*, vol. 231, Sep. 2019, Art. no. 111205.

[4] J. Leinonen, A. Guillaume, and T. Yuan, "Reconstruction of cloud vertical structure with a generative adversarial network," *Geophys. Res. Lett.*, vol. 46, no. 12, pp. 7035–7044, Jun. 2019.

[5] Y.-J. Noh et al., "Cloud-base height estimation from VIIRS. Part II: A statistical algorithm based on A-Train satellite data," *J. Atmos. Ocean. Technol.*, vol. 34, no. 3, pp. 585–598, Mar. 2017.

[6] C. J. Seaman, Y.-J. Noh, S. D. Miller, A. K. Heidinger, and D. T. Lindsey, "Cloud-base height estimation from VIIRS. Part I: Operational algorithm validation against CloudSat," *J. Atmos. Ocean. Technol.*, vol. 34, no. 3, pp. 567–583, Mar. 2017.

[7] J. M. Haynes, Y.-J. Noh, S. D. Miller, K. D. Haynes, I. Ebert-Uphoff, and A. Heidinger, "Low cloud detection in multilayer scenes using satellite imagery with machine learning methods," *J. Atmos. Ocean. Technol.*, vol. 39, no. 3, pp. 319–334, Mar. 2022.

[8] Y.-J. Noh et al., "A framework for satellite-based 3D cloud data: An overview of the VIIRS cloud base height retrieval and user engagement for aviation applications," *Remote Sens.*, vol. 14, no. 21, p. 5524, Nov. 2022.

[9] X. Yan, J. Yang, E. Yumer, Y. Guo, and H. Lee, "Perspective transformer nets: Learning single-view 3D object reconstruction without 3D supervision," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.

[10] B. Kim et al., "Deep reconstruction of 3D smoke densities from artist sketches," *Comput. Graph. Forum*, vol. 41, no. 2, pp. 97–110, May 2022.

[11] Y. Qian, M. Gong, and Y.-H. Yang, "Stereo-based 3D reconstruction of dynamic fluid surfaces by global optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1269–1278.

[12] S. Thapa, N. Li, and J. Ye, "Dynamic fluid surface reconstruction using deep neural network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 21–30.

[13] J. Xiong and W. Heidrich, "In-the-wild single camera 3D reconstruction through moving water surfaces," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 12558–12567.

[14] B. Atcheson et al., "Time-resolved 3D capture of non-stationary gas flows," *ACM Trans. Graph.*, vol. 27, no. 5, pp. 1–9, Dec. 2008.

[15] Y. Ji, J. Ye, and J. Yu, "Reconstructing gas flows using light-path approximation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2507–2514.

[16] T. Xue, M. Rubinstein, N. Wadhwa, A. Levin, F. Durand, and W. T. Freeman, "Refraction wiggles for measuring fluid depth and velocity from video," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 767–782.

[17] A. Levis, Y. Y. Schechner, A. B. Davis, and J. Loveridge, "Multi-view polarimetric scattering cloud tomography and retrieval of droplet size," *Remote Sens.*, vol. 12, no. 17, p. 2831, Sep. 2020.

[18] L. Forster, A. B. Davis, D. J. Diner, and B. Mayer, "Toward cloud tomography from space using MISR and MODIS: Locating the 'veiled core' in opaque convective clouds," *J. Atmos. Sci.*, vol. 78, no. 1, pp. 155–166, 2021.

[19] X. X. Zhu et al., "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.

[20] L. Ding et al., "Imbalanced multi-layer cloud classification with advanced baseline imager (ABI) and CloudSat/CALIPSO data," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2022, pp. 5902–5909.

[21] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, Feb. 2019.

[22] G. E. Karniadakis et al., "Physics-informed machine learning," *Nature Rev. Phys.*, vol. 3, no. 6, pp. 422–440, 2021.

[23] K. Kashinath et al., "Physics-informed machine learning: Case studies for weather and climate modelling," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 379, no. 2194, Apr. 2021, Art. no. 20200093.

[24] J. Li et al., "Thin cloud removal in optical remote sensing images based on generative adversarial networks and physical model of cloud distortion," *ISPRS J. Photogramm. Remote Sens.*, vol. 166, pp. 373–389, Aug. 2020.

[25] *Geostationary Satellite Imargery Dataset*. Accessed: Aug. 26, 2022. [Online]. Available: https://www.ssec.wisc.edu/data/geo/#/animation

[26] *ABI Technical Summary Chart*. Accessed: Mar. 14, 2023. [Online]. Available: https://www.goes-r.gov/spacesegment/ABI-tech-summary.html

[27] *ABI Bands Quick Information*. Accessed: Mar. 14, 2023. [Online]. Available: https://www.goes-r.gov/mission/ABI-bands-quick-info.html

[28] *ABI Performance*. Accessed: Mar. 14, 2023. [Online]. Available: https://www.goes-r.gov/users/GOES-17-ABI-Performance.html

[29] *CloudSat Anomaly Recovery and Operational Lessons Learned*. Accessed: Mar. 14, 2023. [Online]. Available: https://ntrs.nasa.gov/citations/20130009146

[30] *CloudSat/CALIPSO Dataset*. Accessed: Aug. 26, 2022. [Online]. Available: https://www.cloudsat.cira.colostate.edu/data-products/2b-cldclass-lidar

[31] K. Sassen, Z. Wang, and D. Liu, "Global distribution of cirrus clouds from CloudSat/cloud-aerosol LiDAR and infrared pathfinder satellite observations (CALIPSO) measurements," *J. Geophys. Res., Atmos.*, vol. 113, no. D8, Apr. 2008, Art. no. D00A12.

[32] W. B. Rossow and Y. Zhang, "Evaluation of a statistical model of cloud vertical structure using combined CloudSat and CALIPSO cloud layer profiles," *J. Climate*, vol. 23, no. 24, pp. 6641–6653, Dec. 2010.

[33] J. M. Wallace and P. Hobbs, *Atmospheric Science: An Introductory Survey*. 1977, p. 5.

[34] Y. Zhou, Y. Yang, P.-W. Zhai, and M. Gao, "Cloud detection over sunglint regions with observations from the Earth polychromatic imaging camera," *Frontiers Remote Sens.*, vol. 2, Jul. 2021, Art. no. 690010.

[35] C. Esposito, G. A. Landrum, N. Schneider, N. Stiefl, and S. Riniker, "GHOST: Adjusting the decision threshold to handle imbalanced data in machine learning," *J. Chem. Inf. Model.*, vol. 61, no. 6, pp. 2623–2640, Jun. 2021.

[36] M. Kubat, "Neural networks: A comprehensive foundation by Simon Haykin, Macmillan, 1994, ISBN 0–02–352781–7," *Knowl. Eng. Rev.*, vol. 13, no. 4, pp. 409–412, 1999.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[38] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[39] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[40] *Goes Cloud Top Height Product*. Accessed: Sep. 1, 2023. [Online]. Available: https://www.goes-r.gov/products/baseline-cloud-top-height-cloud-layer.html

[41] M. Abdar et al., "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Inf. Fusion*, vol. 76, pp. 243–297, Dec. 2021.

[42] A. B. Arrieta et al., "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020.

[43] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, p. 18, Dec. 2020.

[44] C. Wang, S. Platnick, K. Meyer, Z. Zhang, and Y. Zhou, "A machine-learning-based cloud detection and thermodynamic-phase classification algorithm using passive spectral observations," *Atmos. Meas. Techn.*, vol. 13, no. 5, pp. 2257–2277, May 2020.

[45] *CMWF-AUX Dataset*. Accessed: Mar. 14, 2023. [Online]. Available: https://www.cloudsat.cira.colostate.edu/data-products/ecmwf-aux

[46] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 3319–3328.

Yiding Wang (Student Member, IEEE) received the B.S. degree in computer science from the College of Arts and Sciences, American University, Washington, DC, USA, in 2023, where he is currently pursuing the master's degree in computer science.

Jie Gong (Member, IEEE) received the B.S. degree in atmospheric science from Peking University, Beijing, China, in 2005, and the Ph.D. degree in atmospheric science from Stony Brook University, Stony Brook, NY, USA, in 2009. She did her post-doctoral training at the Jet Propulsion Laboratory (JPL), Pasadena, CA, USA, from 2009 to 2011, where she switched her research area from wave dynamics to satellite remote sensing.

She is currently a Research Physical Scientist at the NASA Goddard Space Flight Center, Greenbelt, MD, USA. She has been a Principal Investigator and a Co-Investigator of many NASA-funded projects. Her research focuses on studying ice and snow microphysical properties, and how they connect to surface precipitation processes from satellite remote sensing observations with a combination use of infrared, sub-millimeter, and microwave techniques.

Dong L. Wu received the B.S. degree in space physics from the University of Science and Technology of China, Hefei, China, in 1985, and the M.S. and Ph.D. degrees in atmospheric science from the University of Michigan, Ann Arbor, MI, USA, in 1993 and 1994, respectively.

He was a Principal Research Scientist and a Supervisor with the Aerosol and Cloud Group, Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena, CA, USA, from 1994 to 2011. He is currently a Project Scientist of NASA's Total and Spectral Solar Irradiance Sensor (TSIS) Mission at the NASA Goddard Space Flight Center, Greenbelt, MD, USA. He was the Principal Investigator (PI) of the Goddards IceCube Project (CubeSat flight demonstration of 883-GHz radiometer for cloud ice measurements). He was a Co-Investigator of Microwave Limb Sounder (MLS) from 1994 to 2008 and CloudSat from 2006 to 2010. He has been a Co-Investigator of the Multi-angle Imaging SpectroRadiometer (MISR) since 2008 and NASA's Global Navigation Satellite Systems (GNSS) since 2007. He has authored or coauthored more than 180 articles in peer-reviewed journals. His research interests include remote sensing of atmospheric clouds and winds.

Dr. Wu received a number of awards, including the NASA Exceptional Achievement Medal in 2001, 2008, and 2022; the JPL Ed Stone Award for Outstanding Research Paper in 2006; and the Robert H. Goddard Award for Science in 2019.

Leah Ding (Member, IEEE) received the Ph.D. degree in electrical engineering from the University at Buffalo, Amherst, NY, USA, in 2013.

She is currently an Associate Professor with the Department of Computer Science, American University, Washington, DC, USA. Her primary research focus centers on trustworthy machine learning and its applications in scientific data analytics.