

# DRLO: Deep Representation Learning for Large Scale Off-track Satellite Remote Sensing Data

1<sup>st</sup> Xin Huang

*Department of Computer and Information Sciences  
Towson University  
Baltimore, MD, USA  
xhuang@towson.edu*

2<sup>nd</sup> Chenxi Wang

*Goddard Earth Sciences Technology and Research (GESTAR) II  
University of Maryland, Baltimore County  
Baltimore, MD, USA  
chenxi@umbc.edu*

3<sup>rd</sup> Wenbin Zhang

*Department of Computer Science  
Michigan Technology University  
Houghton, MI, USA  
wenbinzh@mtu.edu*

4<sup>th</sup> Sanjay Purushotham

*Department of Information Systems  
University of Maryland, Baltimore County  
Baltimore, MD, USA  
psanjay@umbc.edu*

5<sup>th</sup> Jianwu Wang

*Department of Information Systems  
University of Maryland, Baltimore County  
Baltimore, MD, USA  
jianwu@umbc.edu*

**Abstract**—Collocation of measurements from active and passive satellite sensors refers to the combination of data from two sensors that observe the same geographic area at nearly the same time but with differing spatial resolutions and viewing angles. This collocated data, often known as on-track data, comes with precise product labels from the active sensor but comprises only the pixels located directly on the path of an active satellite’s orbit. As a result, its spatial coverage is quite limited, especially when compared to the vast quantities of off-track data. Handling the abundant and information-dense off-track data is crucial for training machine learning models that can effectively integrate the unique features of this data along with on-track data. However, the sheer volume of off-track data presents significant challenges for these models. To address the challenges of large amounts of unlabeled off-track data in remote sensing applications, we introduce a self-supervised representation learning model with VAE and domain adaptation methods to learn a domain invariant classifier for the on-track and off-track data. The model’s performance is enhanced by pre-training off-track data with VAE generative model using off-track data, to learn a good representation that can be transferred to the down-streaming domain adaptation and classification tasks. The classifier is built on these representations to classify different cloud types in passive sensing data, with the goal of achieving higher accuracy in cloud property retrieval. Extensive quantitative and qualitative evaluation demonstrate our method achieves higher accuracy in cloud property retrieval for off-track remote sensing data.

**Index Terms**—representation learning, domain adaptation, multiple satellite data, remote sensing

## I. INTRODUCTION

Constantly covering about two-thirds of Earth’s surface, clouds have a critical role in our climate system, with fundamental influence on the energy, water, and biological cycles. Satellite-based remote sensing plays a crucial role in observing clouds at a global scale. Numerous satellite sensors have been developed to observe and retrieve cloud properties. They can be largely divided into two groups: active sensors such

as CALIPSO and CloudSat, and passive sensors such as MODIS, VIIRS, and ABI. The advantages of active sensors include their capability of resolving the vertical location of the cloud layer and better performance during nighttime and polar regions. On the other hand, passive sensors have a much better spatial sampling rate.

The remote sensing data now grows at an astronomical pace as satellite instruments become more and more powerful, which poses serious challenges to the computational efficiency of physically-based retrieval algorithms. Take the geostationary satellite for example. Until a few years ago, the operational Geostationary Operational Environmental Satellite (GOES) multispectral imager only had a handful of spectral bands and provided a “full-disk” scan of Earth only every 3 hours. The Advanced Baseline Imager (ABI) on the latest GOES-R series (GOES-16+) can provide full-disk scans every 15 minutes in 16 spectral bands with better spatial resolutions. This increased capability leads to a great increase in data, which poses a serious challenge for the physically-based retrieval algorithms. What makes the problem even more challenging is the fact that remote sensing data has also become more heterogeneous. The great success of the A-Train satellite constellation [1] has clearly demonstrated that coordinated and collocated observations from sensors with complementary capabilities, e.g., the combination of passive MODIS and active CALIPSO-CloudSat, can provide a more comprehensive perspective and richer information of clouds that cannot be achieved from the individual instruments alone. However, it is challenging to combine and fuse heterogeneous observations in physically based algorithms because different types of observations often involve dramatically different physics.

Domain adaptation has been thoroughly studied in computer vision [2] [3] and natural language processing (NLP) applications [4] [5]. Recently, the deep learning paradigm has become popular in domain adaptation due to its ability to learn rich, flexible, non-linear domain-invariant representations

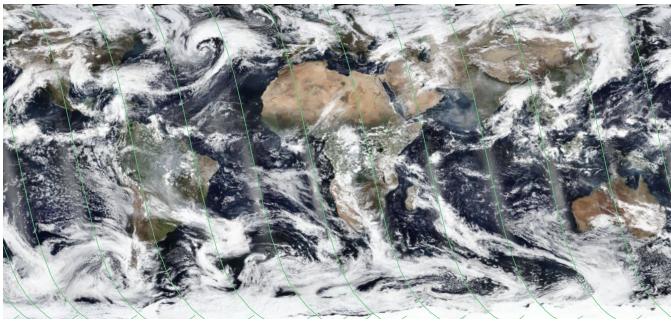


Fig. 1. An example plot of the one-day daytime VIIRS (global coverage) and CALIOP (green lines) orbit tracks (February 8, 2022). Credits: NASA

[6], [7]. However, few of these approaches have been adapted for remote sensing applications. Moreover, domain adaptation techniques using deep neural networks have been mainly used to solve the distribution drifting problem in homogeneous domains [8]. The data in the homogeneous domains usually share similar feature spaces and have the same dimensionalities. Nevertheless, real-world applications often deal with heterogeneous domains that come from completely different feature spaces and different dimensionalities. In our remote sensing application, the two remote sensor datasets collected by active and passive sensors respectively are heterogeneous. In particular, CALIOP actively collects 25 bands of sensing data, it has better sensitivity to aerosol types and cloud phases, and the data are fully labeled with 6 cloud types. VIIRS uses a spectroradiometer sensor to passively collect 20 bands of sensing data with no label or inaccurate labels.

Collocation of measurements from active (e.g., CALIOP) and passive satellite (e.g., VIIRS) sensors involves pairing measurements from two sensors that observe the same location quasi-simultaneously but with different spatial resolutions and at different angles. The collocated data, widely recognized as on-track data, only consists of the pixels on an active satellite orbiting track, thus having very limited spatial coverage compared to large amounts of off-track data. Typically, on-track data are labeled with accurate product type from the active sensor, while off-track data are unlabeled or have inaccurate labels from the passive sensor product. Given the vast amount and valuable information present in off-track data, it's crucial to develop a machine learning model that can integrate the unique features of this data along with on-track data. However, the sheer volume of off-track data presents difficulties in training machine learning models effectively, as this data is often unlabeled and includes considerable noise. Figure 1 shows a coverage difference using a full-day data collection using NASA Earth Data World View website. VIIRS has nearly full coverage of the Earth while CALIOP only covers the green line area which is much smaller than the coverage of VIIRS.

To tackle the issue of handling vast volumes of unlabeled off-track data in remote sensing, we present DRLO, a self-supervised deep representation learning model tailored for large-scale off-track satellite remote sensing data. This model

combines Variational Autoencoders (VAEs) with domain adaptation techniques to develop a domain invariant classifier that efficiently processes both on-track and off-track data. The effectiveness of the model is significantly enhanced by initially pre-training the VAE generative model on off-track data. This pre-training step enables the model to learn a robust representation, which is then applied to domain adaptation and classification tasks. Our comprehensive quantitative and qualitative assessments show that our approach outperforms existing methods in retrieving cloud properties from off-track remote sensing data.

## II. RELATED WORK

Over the past few decades, a variety of aerosol and cloud remote sensing algorithms have been developed based on the physical principles and the radiative transfer of light scattering and absorption within aerosol and cloud fields [9]. These physically-based algorithms are widely used in aerosol and cloud property products for weather and climate studies [10], [11], [12], [13]. Traditionally, many of these algorithms use a lookup table (LUT) approach, in that one must prescribe aerosol and surface properties. The challenge is to ensure the algorithm has the means to select the appropriate model.

Although highly successful, it is challenging to improve these physically-based algorithms. For example, according to [14], there is no absolute separation between “aerosol” and “cloud”. Most, if not all retrieval techniques rely on manually setting thresholds for scene identification, and a test in one region (e.g., tropical ocean) may not work well in another region (e.g., polar land), which necessitates many such “tests”, and painstakingly tuned thresholds. When more than 3 or 4 tests must be combined in a non-sequential way, it is beyond the human ability to interpret. Even with all the effort, dust and cloud scenes are not separated correctly. In addition, even if two sensors are “nearly” the same (e.g., MODIS and VIIRS), spectral bands, resolution, calibration, etc., may be different enough that a threshold applied for one sensor may need revision for another. Thus, physically-based algorithms are expensive.

Machine learning (ML) and artificial intelligence (AI) techniques may overcome the challenges faced by physical-based algorithms. Since ML algorithms are written to autonomously find information (e.g., patterns of spectral, spatial, and/or time series data), they can learn hidden signatures of different types of objects. [15] introduced two Random Forest (RF) machine learning models for cloud mask and cloud thermodynamic phase detection using spectral observations from VIIRS data. [16] developed a deterministic self-organizing map (SOM) approach and applied it to satellite data based cloud type classification. Deep learning [17] is also a promising technique, already having revolutionized many fields such as computer vision [18], natural language processing [19], and is increasingly being used in remote sensing applications [20]. Those approaches can learn representations of multiple variables in a single domain but not multiple domains.

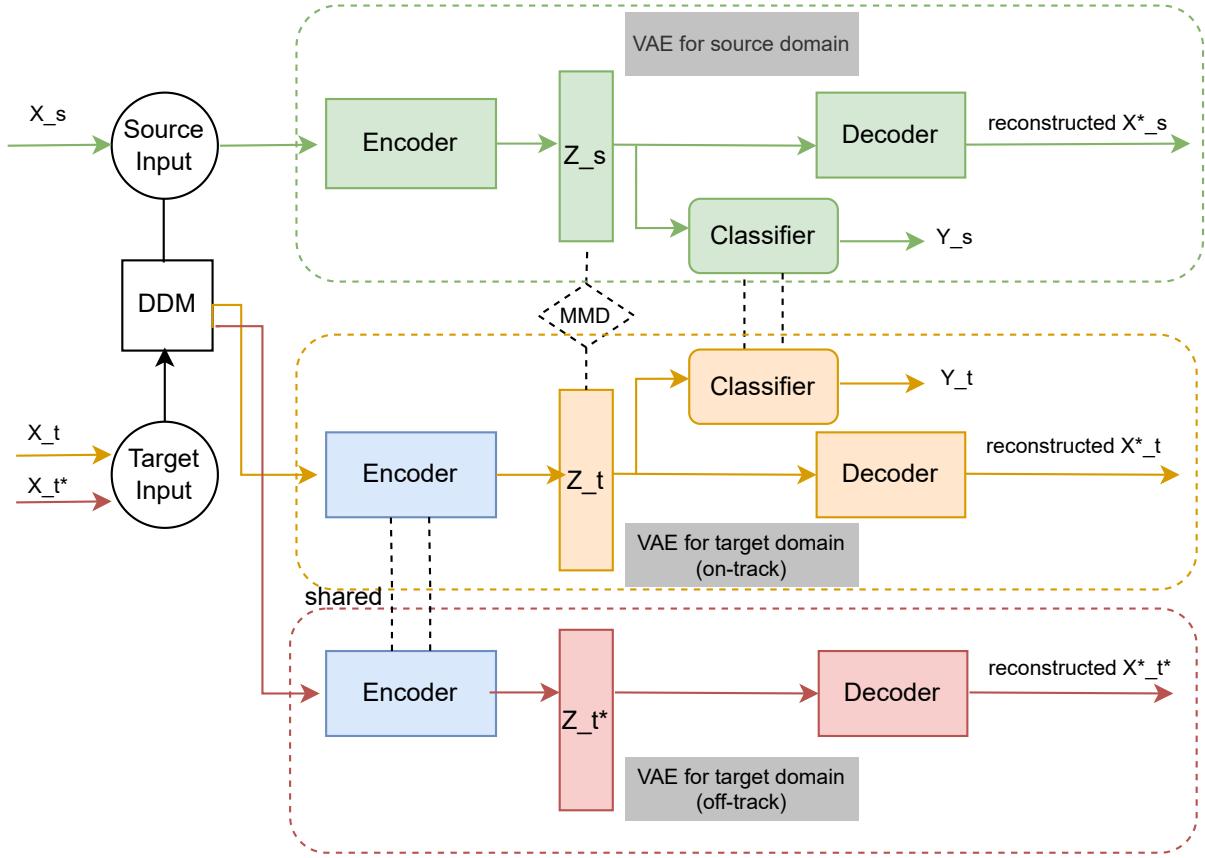


Fig. 2. Framework of DRLO, a deep representation learning model with off-track pre-training and domain adaptation fine-tuning. The model first pre-trains VAE on large scale unlabeled off-track data, it then loads the pre-trained weights and parameters into the VAE-based domain adaptation for on-track source and target domains to build a domain invariant classifier.

Domain adaptation has been widely used in computer vision [2], [3] and natural language processing (NLP) applications [4], [5]. Recently, the deep learning paradigm has become popular in domain adaptation [7], [6], [8] due to its ability to learn rich, flexible, non-linear domain-invariant representations. To the best of our knowledge, those domain adaptation methods can mainly solve the distribution drift problems for homogeneous data collected in different environments and can not be directly adapted to heterogeneous domains such as the active sensor and passive satellite data in remote sensing.

A previous study [21] proposed domain adaptation based cloud type detection methods (DAMA and DAMA-WL) using active and passive satellite data. It develops a deterministic classifier on pixel-level classification but lacks the capacity to capture the spatial correlation among the pixels that are generated by the orbiting satellites and have a strong spatial relationship. [22] further advanced the domain adaptation techniques in multi-satellite remote sensing data by proposing a new Variational Autoencoder (VAE) based domain adaptation method which can capture the spatial correlation and have better generalizability. However, those approaches can only work with on-track collocated data from the paired remote sensors datasets, and they can not directly train a model

utilizing the large-scale off-track unlabeled data, which would have a large distribution drift from the on-track data.

### III. DEEP REPRESENTATION LEARNING WITH VAE AND DOMAIN ADAPTATION METHODS

The passive sensor and active sensor datasets raise more challenges as the two datasets are highly dimensional, globally covered, and heterogeneous. Since off-track data carry rich information and are dominant in the passive remote sensor, it is important to design a model that can incorporate the off-track data in the representation learning, in order to develop a cloud property retrieval method in multi-satellite remote sensing data. Moreover, distribution drift also happens from on-track to off-track data, as some off-track data are very far away from the active sensor's orbiting track, and the environments or surface types change depending on the distance of the off-track data to the on-track data.

To tackle the challenge of distribution drift from on-track to off-track data, we design a deep representation learning framework with VAE and domain adaptation methods that can incorporate off-track feature representation learning to develop a domain invariant classifier for off-track remote sensor data, as shown in Figure 2.

In the training phase, there are three branches of inputs that take source domain data features (CALIOP), target domain data features (off-track VIIRS), and target domain data features (on-track VIIRS). As shown in Figure 2, our model introduces a heterogeneous domain mapping to transform the feature space of the target domain into the feature space of the source domain and uses a feature extraction layer to train the shared representative features between source and target domain. After the domain mapping stage, we pre-train a VAE encoder-decoder using off-track target domain data (red block in Figure 2), and that trained variational encoder is shared with on-track VAE for the target domain (yellow block in Figure 2). It then goes through the VAE-based domain adaptation framework between the on-track source domain (green block) and the on-track target domain (yellow block), in particular, a maximum mean discrepancy (MMD) based domain alignment is applied to the latent spaces generated from source and target domain encoders. By incorporating the domain alignment loss and classification loss in training the domain adaptation network, we find the network can maximize the classification accuracy on the target domain by training this end-to-end deep domain adaptation neural network.

In the testing phase, only target domain data is sent into the deep neural network by going through the deep domain mapping layer and VAE-based feature extraction encoder, which has already captured the hidden representations from both on-track and off-track target data from the training phase. The trained classifier can then be applied to classify the output of the feature extraction layer as the domain invariant feature representation has been generated from the flow.

#### A. Heterogeneous Deep Domain Mapping

Similar to [21], [22], we utilize a deep neural network to perform the deep domain mapping (DDM) between the source and target domain, to adjust for the dissimilarity between the source and target domain by learning a transformation that aligns the target feature space with the source feature space. The goal is to map the VIIRS dataset (target domain) to the CALIOP dataset (source domain) in order to preserve the discriminating power of CALIOP data and transfer it to a downstream machine learning model. This approach ensures that the number of features in both domains is equal and that they are in the same feature space.

For the on-track collocated source and target domain data, the input of the DDM network is the on-track target domain data and the output of the network is the transformed target domain data in the source domain feature space. The source and target domain data in this study are collocated remote sensing data that have the same longitude and latitude coordinates, thus a mean squared error (MSE) loss function is utilized to calculate the error of the DDM network. Specifically, given source domain training examples  $D_s = \{x_i\}, x \in R_s^{d_s}, i = 1, \dots, n_s$  and unlabeled target data set  $D_t = \{u_i\}, u \in R_t^{d_t}, i = 1, \dots, n_t$ , with  $d_s \neq d_t$ ,  $R_s^{d_s} \neq R_t^{d_t}$ , and  $n_s = n_t$ . The DDM model is trained to convert the target domain into the

feature space of the source domain by minimizing the  $L_2$  loss function:

$$L_2 = \frac{1}{n_t} \sum_{i=1}^{n_t} (DDM(u_i) - x_i)^2 \quad (1)$$

#### B. Pre-training VAE for off-track target domain

We first apply the trained DDM from on-track domains to transform off-track target domain data into the feature space of the source domain. Then we perform self-supervised learning by using a variational auto encoder (VAE) to learn a latent space that captures the hidden structures from the off-track target domain data. The off-track target domain VAE is composed of an encoder for feature extraction, a latent feature vector, and a decoder for data reconstruction, respectively. The encoder's purpose is to take input data and create a representation of it in a latent feature space using a parameterized model. The goal is to optimize the parameters of the neural network in order to maximize the variational lower bound by minimizing the KL divergence between the estimated latent vector and the true latent vector, represented as  $L_{kl}$ , and maximizing the expectation of the data points reconstructed from the latent vector, represented as  $L_r$ . We can rewrite the final VAE loss we need to optimize as:

$$L_{vae} = L_{kl} + L_r$$

The learned parameters of the encoder will be shared with encoders used in the domain adaptation module for the on-track data.

#### C. Domain Adaptation on On-track Source and Target Domains

Domain adaptation module for on-track source and target domains consists of four components, i.e., (1) VAE for on-track source domain, (2) VAE for on-track target domain, and (3) domain alignment (4) classifier on the source domain and target domain, respectively.

1) *VAE fine tuning*: We propose a new method to enhance feature representation learning by transferring the knowledge from large-scale off-track data to on-track data. The structure and relationship among the off-track data can be learned via self-supervised VAE and transferred to on-track data by exploiting the idea of transfer learning. To implement the transfer learning based feature representation learning, we use a pre-training and fine-tuning approach. First, the encoder layer of the VAE is pre-trained on off-track target domain data as introduced in Section III-B, and then the pre-trained model with updated parameters is used as an input into our model to give the model some 'knowledge' of training. In particular, transfer learning with the help of off-track target domain data is performed for the external learning of backbone structure and mapping. The weights and parameters of encoders are optimized during pre-training, and the outputs of pre-training are used as starting parameters to train on a similar network that requires on-track target domain data.

Our VAE-based domain adaptation model utilizes two networks: one for the source domain (CALIOP) and another for the target domain (VIIRS). The model optimizes VAE losses for both networks:

$$L_{vae}^s = L_{kl}^s + L_r^s \quad \text{and} \quad L_{vae}^t = L_{kl}^t + L_r^t \quad (2)$$

2) *Domain alignment*: The use of VAE can uncover hidden features in input data, but it can be difficult to extract domain-invariant features due to a significant difference between the source and target domains. To address this issue, we propose a domain alignment module that reduces the discrepancy between the source and target domains and improves the robustness of the classifier in the target domain. Specifically, we add a feature adaptation layer to the VAE for both the source and target domains and use the maximum mean discrepancy (MMD) [23], [22] as a metric for measuring the difference between the domains. In particular, it maps the features of both domains to a common reproducing kernel Hilbert space (RKHS) so that the distance between distributions can be represented as the distance between their kernel embeddings.

VAE's encoder layers  $E_s$  and  $E_t$  are applied to source domain data and target domain data, respectively, resulting in hidden features represented as  $E(X_s)$  and  $E(X_t)$ . The domain alignment loss that needs to be optimized is defined as:

$$L_{mmd} = MMD(E_s(X_s), E_t(X_t)) \quad (3)$$

3) *Weakly supervised classifiers with focal loss*: In our remote sensing application, the classifier is predicting three cloud properties, which are Clear Sky, Liquid Cloud, and Ice Cloud. The labels for the source domain (CALIOP) data are considered accurate and serve as the ground truth due to the active remote sensing method used. However, the labels for the target domain (VIIRS) data are generated by a physical-based retrieval algorithm and are therefore less accurate, referred to as "weak labels". These labels contain only three cloud types (Clear Sky, Liquid Cloud, Ice Cloud) and are only 86% accurate when compared to the ground truth (CALIOP) labels. In addition, the cloud types studied in our remote sensing application are also imbalanced.

To encourage the classifier to learn about the difficulty of classifying samples and confront the challenge of class imbalance, we introduce a weighted focal loss to down-weight the loss for well-classified examples and up-weight the loss for misclassified examples, as well as add the weight of class imbalance. This is achieved by introducing a modulating factor (referred to as  $\gamma$ ) in the cross-entropy loss, which makes the loss function more focused on the hard examples. The weighted focal loss is defined as:

$$FL(p_t) = w * -(1 - p_t)^\gamma * \log(p_t)$$

Where,  $p_t = p$  if true class, otherwise  $1 - p$  and  $p$  is the predicted probability of the true class.  $w$  is the rescaling weight given to each class.

TABLE I  
ACCURACY AND AUC ON PREDICTING THE CLOUD TYPES ON OFF-TRACK CATS (TARGET) DATASET. TRAINING DATA ARE COLLOCATED CALIOP AND VIIRS SATELLITE DATA, AND TESTING DATA ARE CATS SATELLITE DATA.

Models	Accuracy Metric			AUC Metric		
	2015	2016	2017	2015	2016	2017
1) Module 1	0.544	0.556	0.546	0.763	0.782	0.779
2) Module 2	0.703	0.719	0.722	0.845	0.853	0.851
3) DANN	0.418	0.432	0.423	0.667	0.688	0.678
4) DSAN	0.495	0.501	0.515	0.724	0.727	0.733
5) DAMA-WL	0.644	0.661	0.654	0.820	0.832	0.829
6) VDAM	0.652	0.671	0.674	0.822	0.835	0.831
7) DRLO	<b>0.714</b>	<b>0.730</b>	<b>0.730</b>	<b>0.847</b>	<b>0.854</b>	<b>0.853</b>

Our model utilizes the latent feature vectors obtained from the VAEs to construct a source classifier  $C_s$  and a target classifier  $C_t$  by adding fully connected feature layers and ReLU activation functions to the encoders of the source and target domains, respectively. The weighted focal loss  $FL_c^s$  is used for source labels and the weighted focal loss  $FL_c^t$  for target labels. Additionally, the fully connected feature layers for the source and target domains share the same weights, resulting in a classifier that is invariant to the domain.

#### IV. EXPERIMENTS AND EVALUATIONS

We conduct experiments on real-world remote sensing datasets to compare the performance of our proposed model with state-of-the-art ML models.

##### A. Dataset

1) *Training data*: Our experiments are performed using remote satellite sensing datasets from CALIOP active sensor (source) and VIIRS passive sensor (target). The source domain data from CALIOP includes 25 attributes, collected by the CALIOP active spaceborne Lidar sensor. The target domain data from VIIRS includes 20 attributes, collected by the VIIRS passive spectroradiometer sensor. The details of attributes can be found in [24].

The ground truth labels for the source domain are obtained from the aerosol-free pixels of CALIOP, which are divided into three categories: Clear Sky, Pure Liquid Cloud, and Pure Ice Cloud. The weak labels for the target domain are obtained from the aerosol-free pixels of the VIIRS Cloud Top and Optical Properties Product [25]. It should be noted that 86% of the VIIRS data points are found to match the labels of the corresponding CALIOP data points.

Our training datasets consists of off-track target domain (VIIRS) dataset and on-track collocated source domain (CALIOP) and target domain (VIIRS) datasets. The Off-track VIIRS dataset has 600k samples randomly selected from the year 2016 VIIRS data. On-track collocated dataset is collocated for January 2016 of CALIOP and VIIRS datasets with 1,197,536 data points.

2) *Testing data*: The CATS is a lidar remote-sensing instrument that measures atmospheric components from the International Space Station (ISS), which has a non-sun-synchronous orbit [26], [27]. Launched on January 10th, 2015, CATS

provides more than 2-year's continuous observations of clouds and aerosols. Similar to CALIPSO, backscattered signals in the 1064nm channel are used to identify many types of clouds and aerosols and derive vertical structures along the ISS orbit [27].

In this study, we collocate testing data from VIIRS and CATS for the entire CATS operation period (March 2015 – October 2017). Level-1 observations from VIIRS at native 750-m resolution in 16 VIIRS M-bands [28] and vertically resolved cloud-aerosol types from CATS Level-2 Layer product at 5-km resolution [27] are collected. Since the two instruments have different platforms and orbit types, observations outside a 15-minute window and a 5-km distance are discarded. In total, 420k, 740k, and 605k pixels are collected to evaluate our models on the 2015, 2016, and 2017 datasets, respectively.

For model testing purposes, we create pixel labels using collocated CATS Level-2 Layer products. Similar to the labels used for model training, we also use clear sky, pure liquid cloud, and pure ice cloud to represent all kinds of pixel types. In particular, clear means no cloud, liquid (ice) means only liquid (ice)-phase clouds are found in the whole column. It is important to emphasize that, as CALIPSO has dual 532 and 1064nm channels which offer better aerosol detection capabilities [29], the generated labels from CALIPSO and CATS may have some differences.

3) *Evaluation metrics*: The first evaluation metric we use to compare all the models is accuracy, which is defined as:

$$\text{Accuracy} = \frac{\text{Total number of correct predictions}}{\text{Total number of data points}} \quad (4)$$

We also leverage receiver operating characteristic (ROC) curves to assess and compare the efficacy of various models. A ROC curve graphically illustrates the relationship between the true positive rate (TPR), commonly known as sensitivity, and the false positive rate (FPR), referred to as 1-specificity, across different classification score thresholds. By integrating the curve (TPR values) with respect to the FPR values from zero to one, we calculate the area under the ROC curve (AUC). This AUC metric serves as a comprehensive performance indicator, encapsulating the model's effectiveness across all conceivable thresholds. The AUC values fall within the range of [0, 1], with larger AUC values indicating superior performance.

#### B. Quantitative Evaluation

For domain adaptation model comparisons, we conducted experiments on five models that use both source and target data variables (features). These comparison models include: 1) Module 1: The proposed model without domain mapping, it removes the domain mapping module from the proposed model, 2) Module 2: The proposed model without focal loss, using traditional cross-entropy loss, 3) DANN [30]: domain-adversarial training of neural networks, 4)DSAN [31]: Deep Subdomain Adaptation Network, 5) DAMA-WL [21], 6) VDAM [22], 7) DRLO (Our method). Comparing our proposed model with these baseline and state-of-the-art domain adaptation models can help us understand the importance of each module in our model as well as performance improvement to other deep domain adaptation methods.

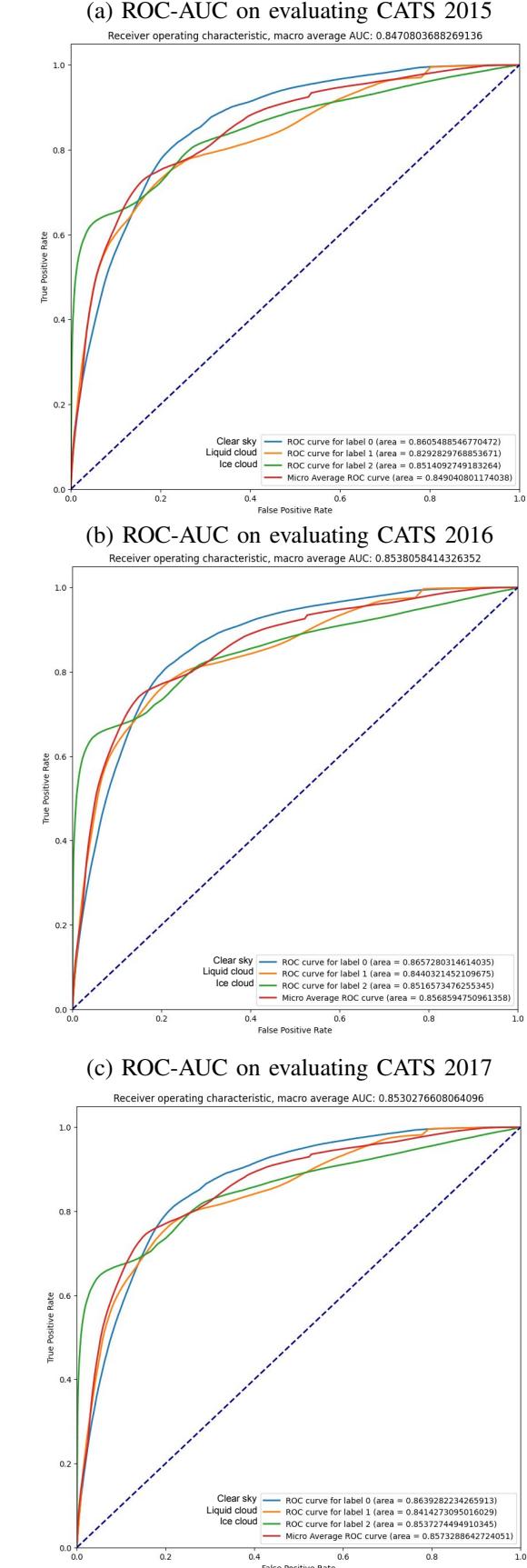


Fig. 3. ROC-AUC performance on evaluating CATS 2015, CATS 2016, and CATS 2017 dataset, respectively.

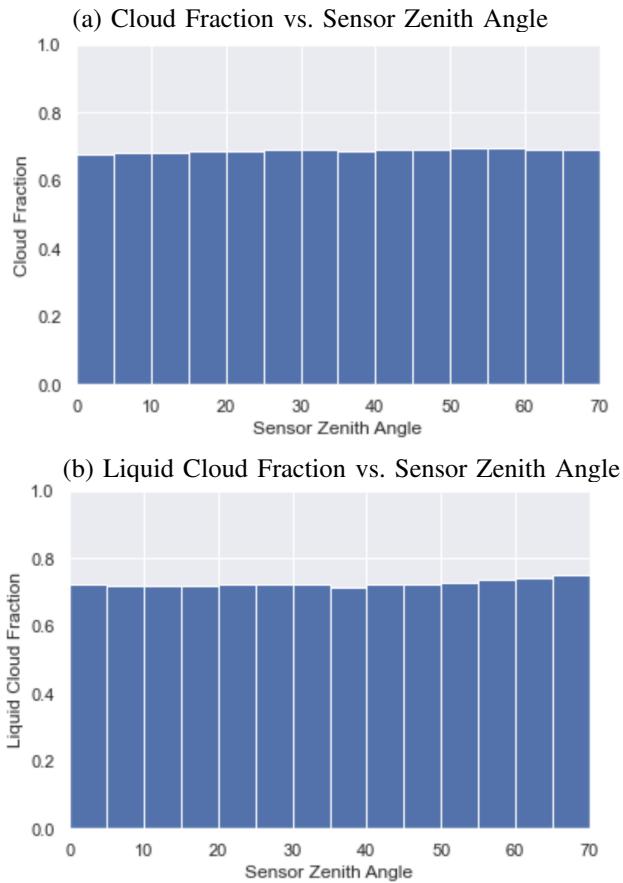


Fig. 4. Cloud fraction and liquid cloud fraction of DRLO's prediction versus sensor zenith angle. The more level the bar is, the more accurate the forecast is, as it tells the independence of the prediction to the sensor's zenith angle.

As shown in Table I, our proposed model outperforms the other domain adaptation baselines significantly. Firstly, our method improves the accuracy by 33% and AUC 9% on average of all the predictions over three years when compared to the model without the DDM module. The low accuracy (around 54%) in predicting without domain mapping illustrates the difficulties in representing heterogeneous data and the challenges of directly using existing domain adaptation methods in such domains. Our proposed deep domain mapping method can bridge the gap between heterogeneous source and target domains and extract a domain-invariant representation by combining it with a domain adaptation technique.

Secondly, we see the weighted focal loss improves our method's accuracy by about 1.1% for all three years used in the evaluation. Thirdly, DANN [30] and DSAN [31], the state-of-the-art domain adaptation techniques, only achieve about 42% and 50% accuracy, respectively, about 40% lower than our DRLO method. This demonstrates that existing domain adaptation techniques, which are primarily designed to address distribution shifts in homogeneous domains, would have inferior performance when applied to heterogeneous domains. Finally, we also observe the proposed model outperforms the other domain adaptation method DAMA-WL model developed

for satellite remote sensing data in [21] and [22], showing around 7% and 6% accuracy improvement for all three testing datasets.

We further illustrate the classification performance for each cloud label using ROC-AUC on evaluating CATS 2015, 2016, and 2017 datasets as shown in Figure 3, in which the clear sky classification has superior performance over liquid and ice clouds. Moreover, similar ROC-AUC performance on datasets of different years also demonstrates that DRLO is stable and robust enough to classify cloud types over long periods by training previous years' data.

### C. Qualitative Evaluation

We conduct extensive climatology evaluations of the proposed approach for satellite remote sensing cloud properties. We collaborate with climatology domain experts in our team to conduct scientific evaluation and compare the proposed datasets to be generated with physically-based algorithms and current EOSDIS level-2 products through multiple aspects including statistics, climatology, ground observation, and ad-hoc case studies. We compare the climatology of a variety of cloud bulk properties, such as cloud fraction and cloud phase (Liquid/Ice cloud fraction) for the VIIRS off-track data.

Figure 4 shows the cloud and liquid cloud fraction of our model's prediction versus the sensor zenith angle. A good prediction model that captures the inherent characteristic of the data should be independent of the sensor's zenith angle. As the bar is very flat across the zenith angles in Figure 4 on both cloud fraction and liquid cloud fraction, it tells that our model's prediction is independent of the sensor's zenith angle and captures good representations of data.

We also conduct a qualitative evaluation with the help of our NASA collaborators from climate science to compare DRLO and DAMA-WL on visualizing the prediction on a granule of VIIRS data, as shown in Figure 5. The clear sky is shown in blue, the ice cloud is shown in red, and the liquid cloud is shown in green. Compared to DAMA-WL, DRLO's prediction is more balanced, by revealing some more ice clouds (red) and clear sky pixels (blue), due to better representation learned from our pre-training/fine-tuning technique and the weighted focal loss used in dealing with the imbalanced data. Moreover, in the regions that are dominant with clear sky (blue), the proposed method is able to identify some ice clouds, which the DAMA-WL method fails to detect.

## V. CONCLUSION

This paper introduces a novel representation learning method using pre-training/fine-tuning VAE and domain adaptation to predict cloud types for off-track remote sensing data. The method pre-trains a VAE-based generative model using large scale off-track data to capture structures of unlabeled data. It then uses a fine-tuning strategy to load the pre-trained VAE model to a domain adaptation network to learn domain-invariant representations from multiple satellite remote sensing data. The method then uses these representations to classify different cloud types in passive sensing data, with the goal

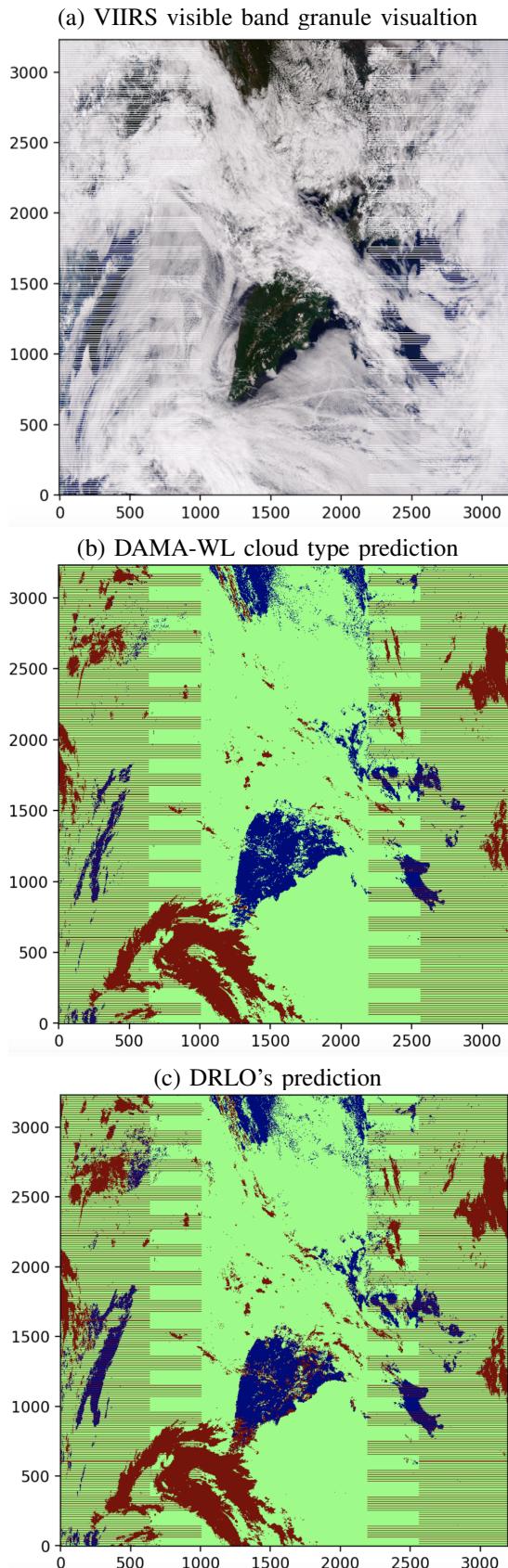


Fig. 5. Granule visualization comparison between DAMA-WL and DRLO on VIIRS cloud type prediction: clear sky (blue), ice cloud(red), and liquid cloud (green).

of achieving higher accuracy in cloud property retrieval. Extensive quantitative and qualitative evaluations show that this method outperforms other state-of-the-art machine learning methods. We also want to highlight our work's impact as better cloud retrieval from the proposed model could help understand climate change.

#### ACKNOWLEDGMENT

We thank our UMBC collaborator Zhibo Zhang and NASA collaborators Kerry Meyer and Benjamin Marchant for insightful discussions. This work is supported by grants OAC-1942714, IIS-1948399 from the National Science Foundation (NSF), and grant 80NSSC21M0027 from the National Aeronautics and Space Administration (NASA).

#### REFERENCES

- [1] D. Vane and G. L. Stephens, "The cloudsat mission and the a-train: A revolutionary approach to observing earth's atmosphere," in *2008 IEEE Aerospace Conference*, 2008, pp. 1–5.
- [2] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 2066–2073.
- [3] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *2013 IEEE international conference on computer vision*, 2013, pp. 2960–2967.
- [4] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification," in *Proceedings of the 45th annual meeting of the association of computational linguistics*, 2007, pp. 440–447.
- [5] G. Foster, C. Goutte, and R. Kuhn, "Discriminative instance weighting for domain adaptation in statistical machine translation," in *2010 conference on empirical methods in natural language processing*. Association for Computational Linguistics, 2010, pp. 451–459.
- [6] E. Tzeng, J. Hoffman, T. Darrell, and K. Saenko, "Simultaneous deep transfer across domains and tasks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 4068–4076.
- [7] S. Purushotham, W. Carvalho, T. Nilanon, and Y. Liu, "Variational recurrent adversarial deep domain adaptation," in *International Conference on Learning Representations*, 2017.
- [8] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *European conference on computer vision*. Springer, 2016, pp. 443–450.
- [9] S. Ackerman, S. Platnick, P. Bhartia, B. Duncan, T. L'Ecuyer, A. Heidinger, G. Skofronick-Jackson, N. Loeb, T. Schmit, and N. Smith, "Satellites see the world's atmosphere," *Meteorological Monographs*, vol. 59, pp. 4–1, 2018.
- [10] S. Ackerman, R. Holz, R. Frey, E. Eloranta, B. Maddux, and M. McGill, "Cloud detection with modis. part ii: validation," *Journal of Atmospheric and Oceanic Technology*, vol. 25, no. 7, pp. 1073–1086, 2008.
- [11] R. Levy, S. Mattoo, L. Munchak, L. Remer, A. Sayer, and N. Hsu, "The collection 6 modis aerosol products over land and," *Atmospheric Measurement Techniques Discussions*, vol. 6, pp. 159–259, 11 2013.
- [12] N. Hsu, M.-J. Jeong, C. Bettenhausen, A. Sayer, R. Hansell, C. Seftor, J. Huang, and S.-C. Tsay, "Enhanced deep blue aerosol retrieval algorithm: The second generation," *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 16, pp. 9296–9315, 2013.
- [13] S. Platnick, K. Meyer, M. King, G. Wind, N. Amarasinghe, B. Marchant, G. Arnold, Z. Zhang, P. Hubanks, R. Holz, P. Yang, W. Ridgway, and J. Riedi, "The modis cloud optical and microphysical products: Collection 6 updates and examples from terra and aqua," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 1, pp. 502–525, 2017.
- [14] J. V. Martins, D. Tanré, L. Remer, Y. Kaufman, S. Mattoo, and R. Levy, "Modis cloud screening for remote sensing of aerosols over oceans using spatial variability," *Geophysical Research Letters*, vol. 29, no. 12, pp. MOD4–1, 2002.
- [15] 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," *Atmospheric Measurement Techniques*, vol. 13, no. 5, pp. 2257–2277, 2020.

- [16] W. Zhang, J. Wang, D. Jin, L. Oreopoulos, and Z. Zhang, “A deterministic self-organizing map approach and its application on satellite data based cloud type classification,” in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 2027–2034.
- [17] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [18] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [19] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” in *Advances in neural information processing systems*, 2013, pp. 3111–3119.
- [20] L. Zhang, L. Zhang, and B. Du, “Deep learning for remote sensing data: A technical tutorial on the state of the art,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 22–40, 2016.
- [21] X. Huang, S. Ali, C. Wang, Z. Ning, S. Purushotham, J. Wang, and Z. Zhang, “Deep Domain Adaptation based Cloud Type Detection using Active and Passive Satellite Data,” in *In: 2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 2020, pp. 1330–1337.
- [22] X. Huang, C. Wang, S. Purushotham, and J. Wang, “Vdam: Vae based domain adaptation for cloud property retrieval from multi-satellite data,” in *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*, ser. SIGSPATIAL ’22. Association for Computing Machinery, 2022.
- [23] K. Borgwardt, A. Gretton, M. Rasch, H. Kriegel, B. Schölkopf, and A. Smola, “Integrating structured biological data by kernel maximum mean discrepancy,” *Bioinformatics (Oxford, England)*, vol. 22, pp. e49–57, 08 2006.
- [24] X. Huang, S. Ali, S. Purushotham, J. Wang, C. Wang, and Z. Zhang, “Deep multi-sensor domain adaptation on active and passive satellite remote sensing data,” in *1st KDD Workshop on Deep Learning for Spatiotemporal Data, Applications, and Systems (DeepSpatial 2020)*. ACM, 2020.
- [25] S. Ackerman, R. Frey, A. Heidinger, Y. Li, A. Walther, S. Platnick, K. Meyer, G. Wind, N. Amarasinghe, C. Wang *et al.*, “Eos modis and snpp viirs cloud properties: User guide for climate data record continuity level-2 cloud top and optical properties product (cldprop), version 1,” *NASA MODIS Adaptive Processing System, Goddard Space Flight Center, USA*, 2019.
- [26] J. E. Yorks, M. J. McGill, S. P. Palm, D. L. Hlavka, P. A. Selmer, E. P. Nowottnick, M. A. Vaughan, S. D. Rodier, and W. D. Hart, “An overview of the CATS level 1 processing algorithms and data products,” *Geophys. Res. Lett.*, vol. 43, no. 9, pp. 4632–4639, 2016.
- [27] J. E. Yorks, S. P. Palm, M. J. McGill, D. L. Hlavka, W. D. Hart, P. A. Selmer, and E. P. Nowottnick, “CATS Algorithm Theoretical Basis Document, Level 1 and Level 2 Data Products, release 1.2.” *Technical Report*, 2016.
- [28] C. Cao, F. J. De Luccia, X. Xiong, R. Wolfe, and F. Weng, “Early on-orbit performance of the visible infrared imaging radiometer suite onboard the suomi national polar-orbiting partnership (s-npp) satellite,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 2, pp. 1142–1156, 2014.
- [29] Z. Liu, M. Vaughan, D. Winker, C. Kittaka, B. Getzewich, R. Kuehn, A. Omar, C. Trepte, and C. Hostetler, “The calipso lidar cloud and aerosol discrimination: Version 2 algorithm and initial assessment of performance,” *Journal of Atmospheric and Oceanic Technology*, vol. 26, pp. 1198–1213, 01 2009.
- [30] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, “Domain-adversarial training of neural networks,” in *Journal of Machine Learning Research*, vol. 17, 2016, pp. 1–35.
- [31] Y. Zhu, F. Zhuang, J. Wang, G. Ke, J. Chen, J. Bian, H. Xiong, and Q. He, “Deep subdomain adaptation network for image classification,” *IEEE Transactions on Neural Networks and Learning Systems*, 2020.