DOI: https://doi.org/10.1145/3557915.3561044

VDAM: VAE based Domain Adaptation for Cloud Property Retrieval from Multi-Satellite Data

Xin Huang
Department of Information Systems, University of Maryland, Baltimore County
Baltimore, MD, USA
xinh1@umbc.edu

Chenxi Wang
Goddard Earth Sciences Technology and Research (GESTAR) II, University of Maryland Baltimore County
Goddard Space Flight Center – NASA
Greenbelt, MD, USA
chenxi.wang@nasa.gov

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

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

ABSTRACT

Domain adaptation techniques using deep neural networks have been mainly used to solve the distribution shift problem in homogeneous domains where data 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 with different dimensionalities. In our remote sensing application, two remote sensing datasets collected by an active sensor and a passive one are heterogeneous. In particular, CALIOP actively measures each atmospheric column. In this study, 25 measured variables/features that are sensitive to cloud phase are used and they are fully labeled. VIIRS is an imaging radiometer, which collects radiometric measurements of the surface and atmosphere in the visible and infrared bands. Recent studies have shown that passive sensors may have difficulties in prediction cloud/aerosol types in complicated atmospheres (e.g., overlapping cloud and aerosol layers, cloud over snow/ice surface, etc.). To overcome the challenge of the cloud property retrieval in passive sensor, we develop a novel VAE based approach to learn domain invariant representation that capture the spatial pattern from multiple satellite remote sensing data (VDAM), to build a domain invariant cloud property retrieval method to accurately classify different cloud types (labels) in the passive sensing dataset. We further exploit the weight based alignment method on the label space to learn a powerful domain adaptation technique that is pertinent to the remote sensing application. Experiments demonstrate our method outperforms other state-of-the-art machine learning methods and achieves higher accuracy in cloud property retrieval in the passive satellite dataset.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

CCS CONCEPTS

• **Applied computing** → **Physical sciences and engineering**; • **Computing methodologies** → *Machine learning*.

KEYWORDS

Domain Adaptation, VAE, Remote Sensing, Deep Learning

ACM Reference Format:
Xin Huang, Chenxi Wang, Sanjay Purushotham, and Jianwu Wang. 2023. VDAM: VAE based Domain Adaptation for Cloud Property Retrieval from Multi-Satellite Data. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Clouds are a critical component of the Earth's climate, having significant impacts on Earth's energy, hydrological and biological cycles [6]. In particular, clouds constantly cover about two-thirds of Earth's surface and they have a key role in modulating global radiative energy distribution. Yet, how clouds will respond to climate change is still uncertain and debated. To understand the radiative effects of clouds in the climate system and how they interact and evolve with the environments, we need to know a variety of cloud properties. Some can be classified as bulk cloud properties, including cloud mask (e.g., a satellite pixel is cloudy or clear), cloud vertical distribution (e.g., cloud top height), and cloud thermodynamic phase (e.g., liquid or ice). Others are often classified as microphysical and optical properties, such as cloud droplet (or cloud ice crystal for ice cloud) effective radius (CER), and cloud optical thickness (COT). Our work focuses on the retrieval of cloud mask and cloud thermodynamic phases from the satellite remote sensing data.

Satellite-based remote sensing is the only way to monitor the above mentioned cloud properties on a global scale. Thus, improvements in cloud observations are a major focus of NASA's Earth science endeavor. Numerous satellite sensors have been developed to observe and retrieve cloud properties. They can be largely divided into two groups according to the origin of the signal: 1) active sensors such as spaceborne lidar (e.g., CALIOP) and radar (e.g., CloudSat) that measure signals transmitted by the sensor that were reflected, refracted or scattered by the Earth's surface or the atmospheric components; and 2) passive radiometers such as MODIS,

VIIRS and ABI that measure reflected solar radiation and/or thermal emission from the Earth-atmosphere system. For cloud remote sensing, the advantages of active sensors include their capability of resolving the vertical distribution of cloud layer and better performance during nighttime and polar region in comparison with passive sensors. On the other hand, passive sensors generally have order-of-magnitude wider swaths and thereby substantially better spatial coverage.

Over the past few decades, a variety of cloud remote sensing algorithms have been developed based on the physical principles and the radiative transfer of light scattering and absorption within cloud fields (see review by Ackerman et al. [3]). Although the underlying physics is largely known, developing a flexible and efficient physics-based cloud property retrieval algorithm is still highly challenging for a number of reasons. First, the algorithms must account for the dramatic variability of the clouds (i.e., low vs. high clouds, liquid- vs. ice-clouds, overcast vs. broken clouds). Secondly, they must take into account the vast different environments, from tropical ocean to Sahara Desert to Arctic sea ice, in which the clouds are observed. Finally, an operational algorithm must be computationally efficient to satisfy users' needs (e.g., hurricane tracking). Traditionally, many physically-based algorithms relied heavily on complex decision trees to deal with the variability of clouds and their environments, and greatly simplify the radiative transfer (RT) process to reduce computation cost. Thus, such algorithms suffer from various retrieval artifacts. For example, threshold-based, complex decision trees are hard to control and often lead to unphysical abrupt changes of cloud properties within small-scale.

Machine learning algorithms such as Random Forests (RF) have shown potential improvement over physically-based algorithms, however, they are often ill-suited to learn from multiple active and passive sensors. For example, RF can be either developed for CALIOP or VIIRS data, but it cannot jointly learn from both these sensors since there is a mismatch of features (variables) among the sensors. Further, RF and other ML algorithms often do not generalize to new combinations of the learned features beyond those seen during the training phase. Moreover, many ML algorithms generally cannot do joint label predictions if the labels are missing for one of these sensors during training phase. To address these issues, we will develop deep learning (DL) models which can automatically learn feature representations from multiple sensors with different features/variables in an end-to-end fashion. Our DL models will be able to predict labels for all sensors even when the labels are absent for some sensors during training phase by transferring knowledge from one set of sensors (e.g. CALIOP) to another set of sensors (e.g. VIIRS).

Active and passive satellite sensors generally measure different variables due to different instruments or applications they were designed for. We can use all these variables from multiple sensors to learn a common representation and can capture any existing common predictive patterns that might aid in inferring knowledge from multiple sensors or transferring knowledge from one sensor to another sensor. The transfer of knowledge or patterns across sensors (here, sensors correspond to domains) is known as domain adaptation. Domain adaptation has been thoroughly studied in computer vision [9, 12] and natural language processing (NLP) applications [4, 10]. Recently, the deep learning paradigm has become popular in domain adaptation [29–31] due to its ability to learn rich, flexible, non-linear domain-invariant representations. However, very few of these approaches have been adapted for remote sensing applications. Our previous study [17] proposed domain adaptation based cloud type detection methods (DAMA and DAMA-WL) using active and passive satellite data. It develops a deterministic classifier that assumes the input satellite pixels are independent and identically distributed (IID). It lacks the capacity to capture the spatial correlation among the pixels that are generated by the orbiting satellites and have strong spatial relationships.

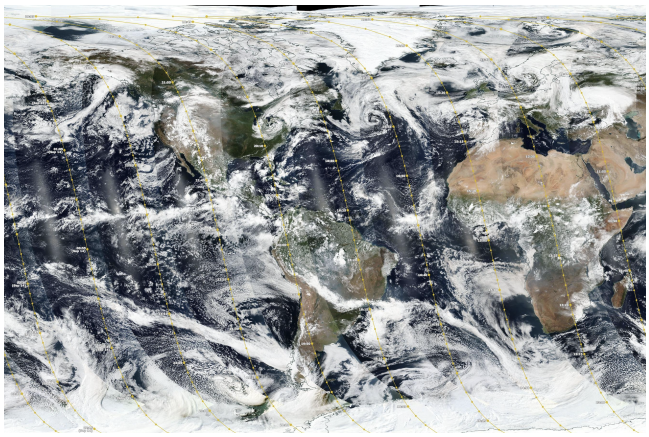

Figure 1: An example plot of the one-day daytime VIIRS (global coverage) and CALIOP (yellow lines) orbit tracks (September 13 2022). Credits: NASA

In this work, we further advance 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. In particular, VAE is generative deep learning model that utilizes neural layers and developed for generating a low-dimension latent space that captures a good representation of the input data. VAE is able to work with highly complex data and learn a latent distribution to represent the internal structure and store it in the hidden nodes. In our remote sensing data (as shown in Figure 1), the collocated data (shown in yellow lines) generated from active and passive satellites are sparser compared to the original data that has global coverage in passive satellite VIIRS, therefore, using a generative model such as VAE can learn the distribution in the latent space that can capture hidden structures and can be used to generate more representative samples than the original data. Moreover, the CNN based encoder network learns the spatial correlations since the collocated satellite data inherently contain the spatial information of the orbiting track, that is, neighboring pixels share more similar data distributions or patterns if they are closer than pixels that are further away on the collocated orbiting track.

Our contributions can be summarized as follows. We open sourced VDAM implementation in PyTorch[1]. 1) We develop a novel VAE based domain adaptation method for multi-satellite Data (**VDAM**)

[1]VDAM source codes are available at: https://github.com/AI-4-atmosphere-remote-sensing/cloud-phase-prediction

to learn domain invariant representation that captures the spatial pattern from multiple satellite remote sensing data, and further build a domain invariant cloud property retrieval method to accurately classify different cloud types (labels) in the passive sensing dataset. 2) We exploit the domain alignment methods on the feature domain and label domain respectively to learn a powerful domain adaptation technique that is pertinent to the remote sensing application. 3) Experiments demonstrate our method outperforms other state-of-the-art ML methods and achieves higher accuracy in cloud property retrieval in the passive satellite dataset.

2 RELATED WORK

In the last 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 [3]. These physically-based algorithms are widely used in aerosol and cloud property products for weather and climate studies [2, 15, 23, 28]. 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 that the algorithm has the means to select the appropriate model. For example, recent efforts[21, 38] implement a non-spherical model to improve dust retrievals. However, this also requires a priori knowledge (dust identification) because uniformly assuming a dust model biases the results [7, 8]. Similarly, cloud optical and radiative properties strongly depend on thermodynamic phase [28]. Therefore, errors in detecting cloud/aerosol, and identifying cloud thermodynamic phase and/or aerosol types can significantly impact downstream retrieval products and scientific analyses.

Machine Learning (ML) and Deep Learning (DL) techniques may overcome the challenges faced by physically-based algorithms. Since ML algorithms are written to autonomously find information (e.g., spectral, spatial, and/or time series patterns), they can learn hidden signatures of different types of objects. ML algorithms are portable and can be easily applied to active and/or passive sensor measurements. Among all the ML algorithms, deep learning (DL) [20] is a promising technique, already having revolutionized many fields such as computer vision [13], natural language processing [27], and is increasingly being used in remote sensing applications [22, 37]. DL models can be automatically trained "end-to-end" and are conducive to learning representations from multiple input variables coming from different sensors.

Deep Learning models (DL) (also called Deep Neural Networks or Deep Models) have become a successful approach for automated extraction of complex data representations for end-to-end training. DL models consist of a layered, hierarchical architecture of neurons for learning and representing data. The main advantage of DL approach is its ability to automatically learn good feature representations from raw data, and thus significantly reducing efforts of hand-crafted feature engineering. In addition, DL models learn distributed representations of data which enables generalization to new combinations of the values of learned features beyond those seen during the training phase. Empirical studies have demonstrated that DL models often yield better machine learning results, e.g., improved classification modeling, and the invariant property of data representations [20]. Deep Learning models have yielded outstanding results in several applications, including computer vision [13] and natural language processing [27]. Recently, machine learning researchers [29] have shown that deep learning approaches can achieve state-of-art performance in analyzing time-series datasets.

Unlike physically-based algorithms or other ML algorithms, DL models can be adapted to new domains by using knowledge transfer technique known as domain adaptation. Domain adaptation has been widely used in solving distribution shift problems in computer vision [9, 12, 25] and natural language processing (NLP) applications [4, 10]. Recently, the deep learning paradigm has become popular in domain adaptation [11, 24, 29, 31, 32] 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 shift problems for homogeneous data collected at different environments and can not be directly adapted to heterogeneous domains such as the active sensor and passive satellite data in remote sensing.

Autoencoders are a specific type of feedforward neural network used to learn efficient representation for a set of data in an unsupervised manner [19]. It consists of three components: encoder, code and decoder, and the input to the neural network is the same as the output. Autoencoders compress the input into a lower-dimensional code through an encoder, and then reconstruct the output from this representation through a decoder. The code is a compact summary of the input, also called the latent-space representation. Variational Autoencoder (VAE) is a variant of Autoencoders as generative models. VAE learns the parameters of the probability distribution modeling the input data, instead of learning an arbitrary function in the case of vanilla autoencoders [18]. By sampling points from this distribution we can also use the VAE as a generative model.

Recently VAE based domain adaptation methods have been applied in image to image translation [24], medical image segmentation [35] and satellite image classification [26] and segmentation [36]. The existing VAE based domain adaptation techniques mainly work with 2D image data which are usually homogeneous (share the same features space and the same dimension), while our approach solves a more challenging problem in heterogeneous domain adaptation with different feature spaces and different dimensions, for active and passive sensors which are very typical settings in satellite remote sensors and have widespread applicability.

In this paper, we develop deep domain adaptation models based on variational auto encoder (VAE) and Deep Convolutional Neural Networks (CNN) [20], to address the challenges of jointly modeling multi-sensor data, and to generate highly accurate labels for passive satellite remote sensing of cloud properties by transferring good representations from active satellite sensing data. Compared to our previous study [17], this paper exploits the characteristics of satellite orbiting track and applies 1D-CNN to capture spatial correlations among neighboring data points to achieve better feature representation, while previous study [17] assumes pixels are IID and have no dependence. Additionally, this paper develops a VAE based generative domain adaptation method to learn the latent representations for active (source domain) and passive sensing data (target domain), while previous study is a deterministic method. Finally, this paper introduces domain alignment methods on both

feature space and label space to develop a powerful domain adaptation technique that is pertinent to the remote sensing application, while previous study only conducts second order statistics based correlation alignment on feature space.

3 SPATIAL DATA COLLOCATION FOR PASSIVE AND ACTIVE SENSORS

Collocation of measurements from two satellite sensors involves pairing measurements from two sensors that observe the same location quasi-simultaneously but with different spatial resolutions and at different angles. Most of the current collocation schemes are carried out on the Earth surface and thus face three challenges. First, it is hard to deal with off-nadir field of view distortion. Second, the collocation involves a time consuming search process. Third, parallax cloud displacement causes large errors for off-nadir pixels.

3.1 Data Collocation based on Spatial Coordinates

In this paper, we adapted the proposed method [14] into our remote sensing dataset to generate the collocation data for the passive and active sensor data. The method is developed to rapidly collocate data from passive radiometers (e.g., MODIS and VIIRS) and/or active sensors at moderate or fine spatial resolutions (< 5km). This algorithm has three notable features. First, by selecting a master pixel from a passive radiometer and subordinate pixels from an active sensor (or another passive sensor), the algorithm derives how the master pixel is transected by subordinate pixels. This algorithm then automatically provides the indices of the subordinate pixels that are located within the master pixel. Second, for collocations of two instruments that have different orbits (e.g., VIIRS and CALIOP), the observation time intervals and pixel distances will be reported for all collocated pixels. Users can easily define a threshold in terms of time interval and/or pixel distance to balance the number of samples and data quality. Third, although active sensors have quasi-nadir viewing geometries, passive radiometers have a wider range of view angles. To deal with the well-known parallax effect, in this algorithm, a parallax effect removal module is designed to mitigate this effect by using lidar detected cloud-top height.

3.2 Active and Passive Satellite Data

The Visible Infrared Imaging Radiometer Suite (VIIRS) [1] is a passive instrument onboard polar orbiting satellites Suomi-NPP and JPSS-1, and will be JPSS-2 after 2022. VIIRS collects visible and infrared imagery and provides less than 1-km spatial resolution observations (native 750m for VIIRS moderate spatial resolutions bands) and wide spatial coverage. Passive sensors observe column-integrated radiation. Accuracy of cloud and/or other atmospheric particles detections could significantly decrease if the whole column is highly heterogeneous (e.g., multi-layered clouds with different thermodynamic phases). In contrast to passive satellite sensors that observe column-integrated, active sensors are more reliable in recognizing objects in different layers because of the high vertical resolutions. For example, the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) onboard CALIPSO satellite [34] operates at wavelengths of 532nm and 1064nm, measuring lidar backscattering profiles at a 30m vertical and 333m along-track resolution. CALIOP also measures the perpendicular and parallel signals at 532nm, along with the depolarization ratio at 532nm that is frequently used in cloud phase discrimination algorithms because of its strong particle shape dependence [34]. Although active sensors are very sensitive to cloud and aerosol layers, they have limited spatial coverage. By taking into account these strengths and weaknesses of both CALIOP and VIIRS, we intend to generate reliable label datasets based on CALIOP Level-2 (version 4) product. The VIIRS Level-1B observations and solar/satellite geometries, and the CALIOP and VIIRS Level-2 cloud mask and thermodynamic phase products will be used for training, validation, testing, and comparison.

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 yellow line area which is much smaller than the coverage of VIIRS.

4 VAE BASED DOMAIN ADAPTATION FOR CLOUD PROPERTY RETRIEVAL

In heterogeneous domain adaptation, the feature spaces between source and target domain are nonequivalent and the dimensions may also generally differ [5]. In our satellite remote sensing application, the source (active) domain data, contains sensing data with 25 attributes collected by the CALIOP active spaceborne lidar sensor while the target (passive) domain data contains another group of sensing data with 20 attributes collected by the VIIRS passive spectroradiometer sensor. The two remote sensing datasets have completely different feature spaces due to the nature of the data they collect. CALIOP data has better representations for the cloud types with its capability of resolving the vertical location of aerosol/cloud layer, better sensitivity to aerosol type and cloud phase, and better performance during night-time and polar region.

In this paper, we propose a VAE based deep domain adaptation technique that captures the spatial patterns and performs heterogeneous domain alignment, and apply it to the cloud property retrieval in our multi-satellite remote sensing data. The passive sensor and active sensor datasets raise more challenges as the two datasets are high dimensional, globally covered and heterogeneous. In the training phase, there are two branches of inputs that are source domain data features (CALIOP) and target domain data features (VIIRS) that have different dimensionalities and heterogeneous feature spaces. As shown in Figure 2, our model introduces a heterogeneous domain mapping to transform the feature space of target domain into the feature space of source domain, and uses feature extraction layer to train the shared representative features between source and target domain. After the domain mapping stage, 1D-CNN is applied to the pixel of interest and its neighboring pixels (1x5) to extract the features with spatial information from its neighbors. It then goes through a VAE encoder-decoder on source branch and target branch, respectively, to generate the latent space that captures hidden structure. Then, a maximum mean discrepancy (MMD) based domain alignment is applied to the latent spaces from source and target domain. 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

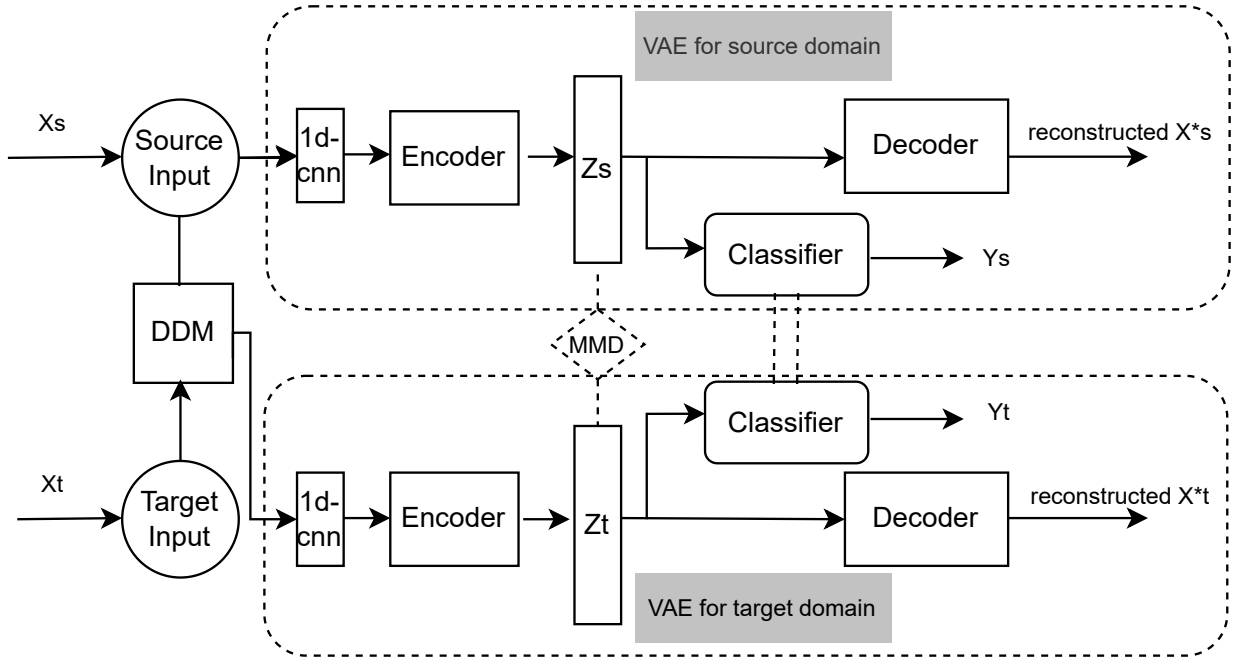

Figure 2: Network architecture of VDAM. For each domain, we construct a customized VAE model, which contains an encoder to extract latent features, a distribution in latent space, a decoder for data reconstruction, and a multi-class classifier for cloud property retrieval for source domain and target domain, respectively. The domain discrepancy between the source domain and target domain is minimized by a domain alignment technique called MMD.

neural network. In the testing phase, only VIIRS (target domain) data is fed into the deep neural network by going through the deep domain mapping layer and VAE based feature extraction layer. 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. Figure 2 demonstrates our VAE based deep domain adaptation which will be further explained in detail in the rest of the section.

4.1 Deep Domain Mapping (DDM)

To adapt to the completely different feature spaces, i.e., the heterogeneity of source and target domain, we introduce a deep learning based approach to learn a transformation to map the target feature space into the source feature space. It equalizes the number of features in source and target domains, and also transforms both domains into the same feature space.

In our remote sensing datasets, the target domain (VIIRS) has wider spatial coverage but with no label information. The source domain (CALIOP) has better representation for cloud types and is fully labeled, so mapping the target domain to source domain can preserve the discriminating power of the source domain and can also transfer it into the down-streaming learner.

We design a deep neural network to perform the deep domain mapping (DDM) between source and target domain. Input of the DDM network is the target domain data and output of the network is the transformed target domain data in the source domain feature space. Because the source domain data and target domain data are collocated remote sensing data with the same longitude and latitude coordinates, mean squared error (MSE) loss function is used to measure the error of the DDM network. Specifically, given source domain training examples $D_s = \{x_i\}, x \in R_s^{d_s}, i = 1, ..., n_s$ and unlabeled target data set $D_t = \{u_i\}, u \in R_t^{d_t}, i = 1, ..., n_t$, with $d_s \neq d_t$ and $R_s^{d_s} \neq R_t^{d_t}$. Because the source domain and target domain are collocated data so we have $n_s = n_t$. The DDM is learned to transform the target domain into source domain feature space by minimizing L_2 loss function:

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

By minimizing the L_2 error, we aim to map the features of target domain into the feature space of source domain that has better feature representation. The L_2 loss is co-trained with domain alignment loss and classifier loss in an end-to-end fashion.

Our multiple domain experiments in Table 1 show DDM can significantly improve the classification accuracy, demonstrate that the proposed domain adaptation and alignment (to be introduced in next section) method works well on the multiple domain data from the same feature space. The proposed heterogeneous deep domain mapping network is also generic and flexible. It can be plugged into other domain adaptation methods and used in areas other than climate data analytics.

4.2 Spatial Feature Representation Learning

The active and passive datasets (CALIOP and VIIRS) used in our application are collocated on the CALIOP orbiting track as shown in Figure 3, which implies the spatial dependency among the pixels that are collected in the dataset. To learn the feature representation with spatial correlation, we use 1D-CNNs on the source branch and target branch after the data mapping subnetwork. CNN is a variant of artificial neural network (ANN) that has been successfully applied to visual imagery analysis and computer vision. By using smaller and simpler patterns embossed in the filters, CNN exploits the hierarchical pattern in data and assemble patterns of increasing complexity. In particular, the input to the 1D-CNN is a tensor with a shape (N, C_{in}, T_{in}). After passing through a convolutional layer, the input sequence becomes abstracted to a feature map, also called an activation map, with a shape (N, C_{out}, T_{out}), N denotes a batch size, C denotes a number of channels, T denotes the length of data sequence. Convolution layer extracts the significant features from the sequence with a spatial kernel of a smaller size k. Pooling layer is also supplied to extracts the dominant features from the convoluted feature map. In this study, T_{in} is set to 5 with a small filer size $k = 3$, and max pooling is used to return the maximum value from the data sequence covered by the filter.

Figure 3 shows the collocated VIIRS (green) and CALIOP (red) pixels sequences that will be fed into the input of the two separate 1D-CNN layers to extract the features with spatial dependencies.

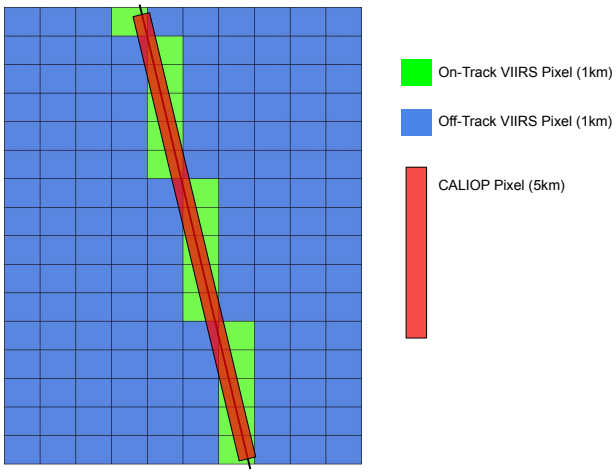

Figure 3: Illustration of on-track and off-track VIIRS pixels as well as the collocated CALIOP track and pixels. In our study, two 1D-CNN layers are applied on the overlapped VIIRS (green) and CALIOP (red) pixel sequences, respectively.

4.3 VAE based Domain Adaptation

Figure 2 shows the main modules of VDAM, which consists of three components, i.e., (1) the customized VAE for source domain, (2) the customized VAE for target domain, and (3) the module for domain alignment (4) the classifier on source domain and target domain, respectively. The source domain VAE and target domain VAE are composed of an encoder for feature extraction, a latent feature vector, a decoder for data reconstruction, respectively.

4.3.1 ***Latent Space Learning*** . The encoder maps the input data into a latent feature space, and approximates the posterior probability by a parameterized model. In VAE, the sample $P_\theta(z)$ can be generated from the latent space z. Then it can get the reconstruction given z as $P_\theta(\mathbf{x}|\mathbf{z})$, θ is the learned parameter. The goal is to maximize the log-likelihood of the data, which is composed of a sum over the marginal likelihoods of individual data points, $\log p_\theta(\mathbf{x}^{(i)}, ..., \mathbf{x}^{(N)}) = \sum_{i=1}^{N} \log p_\theta(\mathbf{x}^{(i)})$, which can be rewritten [18] as:

$\log p_\theta(\mathbf{x}^{(i)}) = D_{KL}(q_\phi(\mathbf{z}|\mathbf{x}^{(i)})||p_\theta(\mathbf{z}|\mathbf{x}^{(i)})) + L(\theta, \phi; \mathbf{x}^{(i)})$.

The first RHS term is the KL divergence of the approximate from the true posterior. KL-divergence is non-negative, and the second RHS term $L(\theta, \phi; \mathbf{x}^{(i)})$ is the (variational) lower bound on the marginal likelihood of datapoint i (ϕ is variational approximation parameter) that can be written as:

$$\log p_\theta(\mathbf{x}^{(i)}) \geq L(\theta, \phi; \mathbf{x}^{(i)}) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[-\log q_\phi(\mathbf{z}|\mathbf{x}) + \log p_\theta(\mathbf{x}, \mathbf{z})]$$

And it can also be written as:

$$L(\theta, \phi; \mathbf{x}^{(i)}) = -D_{KL}(q_\phi(\mathbf{z}|\mathbf{x}^{(i)})||p_\theta(\mathbf{z})) + \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x}^{(i)})}[\log p_\theta(\mathbf{x}^{(i)}|\mathbf{z})]$$

The goal is to maximize the variational lower bound by optimizing the parameters θ and ϕ of the neural network. In particular, we need to minimize the KL divergence between the estimated latent vector and the true latent vector, which can be simplified as L_{kl} ; and maximize the expectation of the reconstructed data points sampled from the latent vector, which can be represented as L_r. We can rewrite the final VAE loss we need to optimize as:

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

At the training stage, we sample the latent features from the posterior distribution, and feed them into the subsequent networks of the classifier and decoder. The classifier is composed of several fully connected layers that outputs the probability of cloud property category (clear, liquid cloud, ice cloud, mixed cloud) for each pixel, and the decoder is used to reconstruct the input data. Our VAE based domain adaptation model consists of two networks, one is the VAE network for the source domain (CALIOP) and the other is also a VAE network but for target domain (VIIRS). So the VAE losses that our model optimizes is:

$$L^s_{vae} = L^s_{kl} + L^s_r \quad \text{and} \quad L^t_{vae} = L^t_{kl} + L^t_r \tag{2}$$

4.3.2 ***MMD based Domain Alignment***. Although using VAE can discover hidden features of the given input data, there is significant distribution discrepancy between the source and the target domain that makes it difficult to extract domain-invariant features. We introduce a domain alignment module to minimize the discrepancy between source and target domains and help develop a domain invariant classifier that can be robust in the target domain. In particular, we add a feature adaptation layer to the auto encoder pairs of source and target domain to measure the domain discrepancy loss. We used the maximum mean discrepancy (MMD) [5], an effective non-parametric empirical estimation method, as the

domain distance metric between the source domain and target domain adaptation layer. The idea is to convert two sets of source and target domain features to a common reproducing kernel Hilbert space (RKHS), so that representing distances between distributions as distances between kernel embedding of distributions.

Specifically, let $X = \{x_1, \cdots, x_{n_1}\}$ and $Y = \{y_1, \cdots, y_{n_2}\}$ be random variable sets from two distributions $\mathcal{P}$ and $\mathcal{Q}$. The empirical estimate of the distance MMD between $\mathcal{P}$ and $\mathcal{Q}$ is

$$\mathrm{MMD}(\mathbf{X}, \mathbf{Y}) = ||\frac{1}{n_1}\sum_{i=1}^{n_1}\psi(x_i) - \frac{1}{n_2}\sum_{i=1}^{n_2}\psi(y_i)||_{\mathcal{H}},$$

where $\mathcal{H}$ is a universal RKHS, $||*||_{\mathcal{H}}$ is RKHS norm, and $\psi : X \rightarrow \mathcal{H}$. Then the distance between distributions of two variables can be estimated by the distances between the means of the samples mapped into a RKHS.

Source domain data and target domain data go through the encoder layers E_s and E_t of VAE and generate the hidden features represented as $E(X_s)$ and $E(X_t)$. The domain alignment loss to be optimized can be defined as:

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

4.3.3 ***Weakly Supervised Classifiers.*** With learned latent feature vectors from the VAEs, we add two sets of fully connected feature layers and ReLU activation functions after the source domain and target domain encoders to build a source classifier C_s (with cross entropy loss L_c^s for source labels) and a target classifier C_t (with an instance weighted cross entropy loss L_c^t for target label), respectively. The fully connected feature layers from the source domain share the weights with the fully connected feature layers of the target domain, so that the learned classifier can be domain invariant.

In our remote sensing data, the cloud properties the classifier is predicting are Clear Sky, Liquid Cloud, Ice Cloud and Mixed Cloud. The labels of source domain (CALIOP) data are derived from active remote sensing, so they are considered as accurate and ground truth of the cloud labels. For target domain (VIIRS) data, inaccurate VIIRS labels are also generated from the physical-based retrieval algorithm because it could provide supplementary information for the globally-covered pixels, although the labels of the target domain (VIIRS) are noisy/inaccurate. We consider the physical-based VIIRS label set as weak labels, as it contains only three cloud types (Clear Sky, Liquid Cloud, Ice Cloud), which is also only 80% accurate when compared to the ground truth (CALIOP) labels.

To incorporate the interplay between weak label and accurate label, we introduce a label alignment approach to help learn the classifier. That is, we assign a weight w_i to each collocated data pair used in calculating the target classifier loss L_c^t. The idea is to assign a lower weight to a data pair when the weak label from target domain equals to that of source domain, while assigning a higher weight when weak label from target domain differs to that of source domain. This can facilitate the model to focus on learning toward the more challenging area that the classifier is uncertain about. The weights we used for different label values alignment are:

$$w_{s_i,t_i} = \begin{cases} 1.5 & \text{if label of } s_i \text{ differs to label of } t_i \\ 1.25 & \text{if label of } s_i \text{ is mixed cloud} \\ 1 & \text{if label of } s_i \text{ equals to label of } t_i \end{cases} \tag{4}$$

, in which s_i, t_i are collocated data pair with s_i represents the source domain (CALIOP) data point and t_i represents the target domain (VIIRS) data point, respectively.

The loss functions of the source domain classifier and target domain classifier are:

$$L_c^s = \sum_{i=1}^{n_s} CE(Y_{s_i}, Y_i) \tag{5}$$

$$L_c^t = \sum_{i=1}^{n_t} CE(Y_{t_i}, Y_i) * w_{s_i,t_i} \tag{6}$$

in which CE represents the standard cross entropy between the classifier's prediction (Y_s or Y_t) and ground truth (Y), and w_{s_i,t_i} is the label weight from formula 4.

4.4 End-to-End Joint Training

The domain mapping and VAE based feature extraction and classification modules are trained jointly in an end-to-end fashion in order to align the heterogeneous source and target domains and build the domain invariant classifier. In each training epoch, the parameters of the VAE with domain alignment and domain mapping module are updated using back-propagation algorithm. The joint loss is composed of the loss of deep domain mapping, the losses of VAE losses for source domain and target domain, the loss of MMD based domain alignment, the loss of source classifier and the loss of target classifier with weak label:

$$Loss = L_2 + L_{vae}^s + L_{vae}^t + L_{mmd} + L_c^s + L_c^t \tag{7}$$

5 EXPERIMENTS

We evaluate our method with several experiments on real world remote sensing datasets to compare the performance of the proposed model with the state-of-the-art ML models. In this section, we first give a description of our datasets, then we summarize experiment results and analysis.

5.1 Datasets

We conduct experiments on CALIOP active sensor (source) and VIIRS passive sensor (target) remote satellite sensing datasets. The source domain CALIOP data, contains sensing data with 25 attributes (features) collected by the CALIOP active spaceborne Lidar sensor. The detailed description of the attributes can be found at the previous study [16]. The target domain VIIRS data contains another group of sensing data with 20 attributes collected by the VIIRS passive spectroradiometer sensor. Four auxiliary attributes shared in both CALIOP and VIIRS datasets are surface temperatures, surface emissivity, surface type and snow ice index. The latitude and longitude of the pixel are also provided in both CALIOP and VIIRS datasets. In total, there are 6 auxiliary features supplemented to both source and target domains to train the domain adaptation model. Following our previous work at [33], we filter nighttime

data records and choose the daytime records with 0 < Solar Zenith angle (SZA) < 80.

We consider the aerosol-free pixels from CALIOP as ground truth labels for source domain, with the following four categories: 1) Clear Sky (no cloud), 2) Pure Liquid Cloud, and 3) Pure Ice Cloud, 4) Mixed Liquid and Ice Cloud. We also consider aerosol-free pixels from VIIRS Cloud Top and Optical Properties Product [1] as weak labels in the target domain. The weak labels correspond to the following three categories: 1) Clear Sky (no cloud), 2) Pure Liquid Cloud, and 3) Pure Ice Cloud. As mentioned earlier, 80% of the VIIRS data points match the labels of the collocated CALIOP data points.

In this study, we create a new dataset with 4 aforementioned cloud type labels that exclude the aerosols (compared to 7-label Dataset 1 used in our previous work [17]) and add a new mixed liquid/ice cloud (compared to 3-label Dataset 2 in [17]). We work with domain experts (our NASA collaborators) to identify the 4-label cloud types (used in this study) that are most interesting and important in the cloud retrieval domain. With this new dataset, the overall accuracy in this paper (as shown in Table 1 is around 13% higher than 7-label Dataset 1 (DAMA model) in [17] and 8% lower than the 3-label Dataset 2 (DAMA-WL model) in [17], which is reasonable and consistent in terms of the complexity of the label space. Our training dataset is collocated for January 2013 of CALIOP and VIIRS datasets with 1,270,583 data points. Each built model is evaluated by predicting the labels for the month of four years, i.e., January 2014, January 2015, January 2016 and January 2017. Figure 4 shows the class distribution against each cloud type (class) for CALIOP labels (4 cloud types). Similarly, Figure 5 shows class distribution against each cloud type (class) for the training and test VIIRS datasets with weak VIIRS labels (3 cloud types). Analyzing the class distribution in the training dataset for CALIOP, as illustrated in Figure 4, we can see some class imbalance with highest class label data available for "Pure Liquid" and lowest class label data available for "Mixed Cloud". Similarly, for VIIRS weak label of the second label setting, we can see highest class label data available for "Pure Liquid" and lowest class label data available for "Clear Sky", as illustrated in Figure 5.

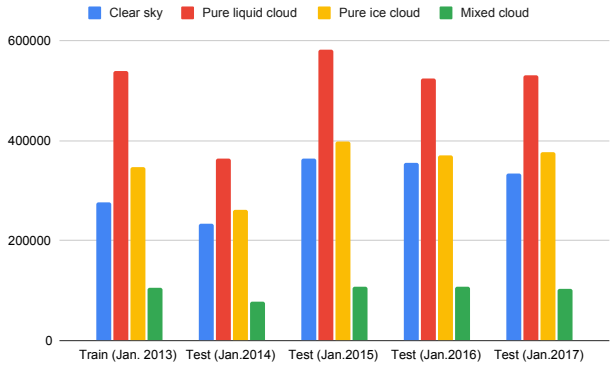

Figure 4: Source domain (CALIOP) data distribution (data point count for each of 4 cloud types) for training and testing.

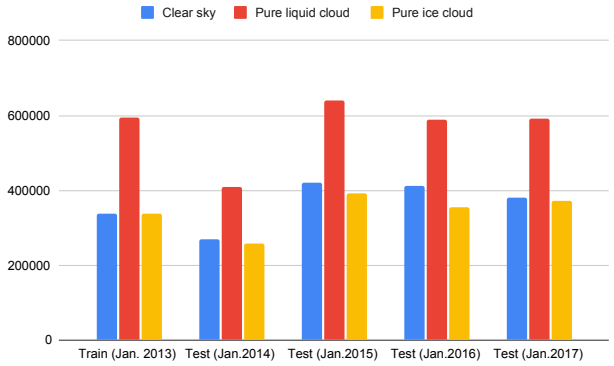

Figure 5: Target domain (VIIRS) data distribution (data point count for each of 3 cloud types) for training and testing.

We evaluate our model with accuracy metric to compare all the models:

$$Accuracy = \frac{\text{Total number of correct predictions}}{\text{Total number of data points}} \tag{8}$$

5.2 Performance Comparison using Data from Single Domain

For non-domain adaptation model comparison, we conducted experiments on three baseline models which were trained on data from a single domain. These baseline models include 1) RF model: Random Forest trained on VIIRS data, 2) MLP-VIIRS: A deep learning based MLP model trained on VIIRS data, 3) MLP-CALIOP: A deep learning based MLP model trained on CALIOP data.

In order to make a fair comparison to our proposed model, we apply the same neural network used in the shared layer of our VDAM network to build the neural network for baseline models (MLP-CALIOP and MLP-VIIRS), with the same type and number of layers. In our experiments, the MLP (shared) layers are followed with a ReLU activation function and Dropout (0.25). To train the RF model, we specify 100 as the number of trees and 15 as the maximum depth of the trees in the forest, chosen after hyperparameter tuning.

As shown in Table 1, as an ML-based baseline result, RF achieves around 82% test accuracy (for January 2017). For the single domain experiments, we can see MLP-CALIOP achieves 100% accuracy in predicting the active sensing dataset, which is expected as the data distribution of each cloud type is very discriminative in CALIOP dataset. Our ultimate goal is to transfer the discriminative representation from this active sensor CALIOP to passive sensor VIIRS in order to accurately classify the cloud types in the passive dataset. In comparison, MLP-VIIRS model has lower accuracy around 81%, as VIIRS is a passive dataset collected by detecting the reflection of natural radiation and their feature discrimination power is weak. This observation highlights the importance of using multiple sensor data to better understand and classify the unlabeled passive sensing data that has wider spatial coverage. Our proposed VAE based deep domain adaptation model aims to achieve higher accuracy than using single domain data by transferring the discriminating power from the source domain to target domain.

Table 1: Accuracy on predicting the cloud types on VIIRS (target) dataset.

Models - Single Domain	Label	Source	Target	Jan. 2014	Jan. 2015	Jan. 2016	Jan. 2017
Random Forest	CALIOP	VIIRS	VIIRS	0.815	0.823	0.821	0.828
Random Forest-WL	VIIRS	VIIRS	VIIRS	0.775	0.783	0.781	0.790
MLP-VIIRS	CALIOP	VIIRS	VIIRS	0.805	0.811	0.810	0.815
MLP-CALIOP	CALIOP	CALIOP	CALIOP	1.000	1.000	1.000	1.000
Models - Multiple Domains							
Auto Encoder model	CALIOP + VIIRS	CALIOP	VIIRS	0.821	0.833	0.830	0.836
Model without domain mapping	CALIOP + VIIRS	CALIOP	VIIRS	0.512	0.533	0.530	0.539
Model without 1D-CNN	CALIOP + VIIRS	CALIOP	VIIRS	0.855	0.863	0.861	0.866
DANN [11]	CALIOP + VIIRS	CALIOP	VIIRS	0.563	0.578	0.570	0.582
DAMA-WL [17]	CALIOP + VIIRS	CALIOP	VIIRS	0.842	0.848	0.846	0.851
VDAM (Our method)	CALIOP + VIIRS	CALIOP	VIIRS	**0.868**	**0.872**	**0.871**	**0.878**

By comparing VDAM's prediction to traditional ML methods as shown in Table 1, our proposed model's prediction accuracy outperforms (5% higher) Random Forest Model that is widely used in climate data. Supervised learning models such as Random Forest in the single domain assume the label information on the target domain is available, in comparison, our VAE based domain adaptation is weakly unsupervised domain adaptation that does not require accurate label information in the target domain, and mainly rely on the label information of source domain and domain alignment between source and target domain to build the model and make prediction.

5.3 Performance Comparison of using Data from Multiple Domains

For domain adaptation model comparisons, we conducted experiments on five baseline models that use both source and target data variables (features). These baseline models include the following: 1) Auto Encoder model: uses auto encoder instead of VAE in the proposed model, 2) Model without domain mapping: removes the domain mapping module from the proposed model, 3) Model without 1D-CNN: removes the 1D-CNN layers from the proposed model, 4) DANN [11]: domain-adversarial training of neural networks, 5) DAMA-WL [17], 6) VDAM (Our method). Comparing these baseline models with our proposed model can help understand the importance of each module in our model as well as performance improvement to other deep domain adaptation methods.

From the result of multiple sources based models in Table 1, our proposed model outperforms the other domain adaptation baselines significantly. Firstly, by comparing the results of VAE module to using Auto Encoder method, we can see there are about 4% improvement in using VAE, which tells the learned variational distribution help train a better model. Secondly, Our method improves the accuracy by 34% on average of all the predictions from January 2013 to January 2017 when compared to the model without Domain Mapping module. The very low accuracy (around 53%) in predicting cloud satellite data without domain mapping exemplifies inherent complexities in heterogeneous data representation and the challenge of directly applying existing domain adaptation methods in heterogeneous domains. Our proposed deep domain mapping can mitigate the gap between the heterogeneous source and target domains and extract the domain invariant representation by integrating with the domain adaptation technique. Thirdly, we see the 1D-CNN layers improves our method's accuracy by about 1% for all four months used in testing, that tells the spatial information is captured in our model and helps the model performance. Additionally, DANN [11], the widely used adversarial learning based domain adaptation method, only achieves about 58% accuracy, 30% lower than our VDAM method. It shows that existing domain adaptation techniques, proposed to mainly solve distribution shift in homogeneous domains, would have degraded performance when applying to heterogeneous domains. Finally, we also observe the proposed model outperforms the other domain adaptation method DAMA-WL model introduced in [17], showing around 2.5% accuracy improvement for all the 4 testing datasets.

Our work is a pilot study in using VAE based generative models in domain adaptation to transfer the knowledge from satellite active sensor to passive sensor. Recent studies show GAN is being used in computer vision / imaging which handles high dimensional space of 2D/3D imaging data, with the aim of generating visually plausible images. The satellite remote sensing data in our study has lower dimensions with the goal of cloud property classification. We chose VAE as the pilot method, instead of GAN, due to its simplicity as a powerful generative model. The experiment results show our VAE based domain adaptation approach outperforms other single domain or multiple domain approaches. Because of the promising results of our approach, we will further explore other generative models (such as GAN, Probabilistic/Bayesian model, etc) and compare their performance in our future work.

6 CONCLUSION

In this paper, we present a novel VAE based deep domain adaptation method (VDAM) to employ both active and passive sensing data in cloud type prediction. We develop a VAE based generative method to learn domain invariant representation that capture the spatial pattern from multiple satellite remote sensing data, and further build a domain invariant cloud property retrieval method to accurately classify different cloud types (labels) in the passive sensing data, by transferring good cloud representations from active satellite sensing data. We further exploit the weight based alignment method on label space to learn a powerful domain adaptation

technique that is pertinent to the remote sensing application. Experiments demonstrate our method outperforms other state-of-the-art ML methods and achieves higher accuracy in cloud property retrieval in the passive satellite data. For future work, we plan to investigate taking into account the information of off track pixels and incorporating deep learning models that can capture temporal information (e.g., LSTM, Transformer) to further improve cloud property retrieval.

ACKNOWLEDGMENTS

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−1730250, OAC−1942714, IIS−1948399 from the National Science Foundation (NSF) and grant 80NSSC21M0027 from the National Aeronautics and Space Administration (NASA).

REFERENCES

[1] SA. Ackerman, R. Frey, A. Heidinger, Y. Li, A. Walther, S. Platnick, K. Meyer, G. Wind, N. Amarasinghe, C. Wang, et al. 2019. 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).

[2] SA. Ackerman, RE. Holz, R. Frey, EW. Eloranta, BC. Maddux, and M. McGill. 2008. Cloud detection with MODIS. Part II: validation. *Journal of Atmospheric and Oceanic Technology* 25, 7 (2008), 1073−1086.

[3] SA. Ackerman, S. Platnick, PK. Bhartia, B. Duncan, T. L'Ecuyer, A. Heidinger, G. Skofronick-Jackson, N. Loeb, T. Schmit, and N. Smith. 2018. Satellites see the world's atmosphere. *Meteorological Monographs* 59 (2018), 4−1.

[4] J. Blitzer, M. Dredze, and F. Pereira. 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of the 45th annual meeting of the association of computational linguistics*. 440−447.

[5] K. Borgwardt, A. Gretton, M. Rasch, HP. Kriegel, B. Schölkopf, and A. Smola. 2006. Integrating structured biological data by Kernel Maximum Mean Discrepancy. *Bioinformatics (Oxford, England)* 22 (08 2006), e49−57. https://doi.org/10.1093/bioinformatics/btl242

[6] O. Boucher, D. Randall, P. Artaxo, C. Bretherton, G. Feingold, P. Forster, V.-M. Kerminen, Y. Kondo, H. Liao, U. Lohmann, P. Rasch, S. K. Satheesh, S. Sherwood, B. Stevens, and X. Y. Zhang. 2013. Clouds and Aerosols. In *Climate Change 2013*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

[7] HM. Cho, S. Nasiri, P. Yang, I. Laszlo, and T. Zhao. 2013. Detection of Optically Thin Mineral Dust Aerosol Layers over the Ocean Using MODIS. *Journal of Atmospheric and Oceanic Technology* 30 (05 2013), 896−916. https://doi.org/10.1175/JTECH-D-12-00079.1

[8] P. Ciren and S. Kondragunta. 2014. Dust aerosol index (DAI) algorithm for MODIS. *Journal of Geophysical Research: Atmospheres* 119 (04 2014). https://doi.org/10.1002/2013JD020855

[9] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars. 2013. Unsupervised visual domain adaptation using subspace alignment. In *2013 IEEE international conference on computer vision*. 2960−2967.

[10] G. Foster, C. Goutte, and R. Kuhn. 2010. Discriminative instance weighting for domain adaptation in statistical machine translation. In *2010 conference on empirical methods in natural language processing*. Association for Computational Linguistics, 451−459.

[11] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and VS. Lempitsky. 2016. Domain-Adversarial Training of Neural Networks. In *Journal of Machine Learning Research*, Vol. 17. 1−35.

[12] B. Gong, Y. Shi, F. Sha, and K. Grauman. 2012. Geodesic flow kernel for unsupervised domain adaptation. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2066−2073.

[13] K. He, X. Zhang, S. Ren, and J. Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770−778.

[14] RE. Holz, SA. Ackerman, FW. Nagle, R. Frey, S. Dutcher, RE. Kuehn, MA. Vaughan, and B. Baum. 2008. Global Moderate Resolution Imaging Spectroradiometer (MODIS) cloud detection and height evaluation using CALIOP. *Journal of Geophysical Research: Atmospheres* 113, D8 (2008).

[15] NC. Hsu, M-J. Jeong, C. Bettenhausen, AM. Sayer, R. Hansell, CS. Seftor, J. Huang, and S-C. Tsay. 2013. Enhanced Deep Blue aerosol retrieval algorithm: The second generation. *Journal of Geophysical Research: Atmospheres* 118, 16 (2013), 9296−9315.

[16] X. Huang, S. Ali, S. Purushotham, J. Wang, C. Wang, and Z. Zhang. 2020. 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.

[17] X. Huang, S. Ali, C. Wang, Z. Ning, S. Purushotham, J. Wang, and Z. Zhang. 2020. 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, 1330−1337.

[18] DP. Kingma and M. Welling. 2014. Auto-Encoding Variational Bayes. *CoRR* abs/1312.6114 (2014).

[19] MA. Kramer. 1991. Nonlinear principal component analysis using autoassociative neural networks. *Aiche Journal* 37 (1991), 233−243.

[20] Y. LeCun, Y. Bengio, and G. Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436−444.

[21] J. Lee, N. Hsu, A. Sayer, C. Bettenhausen, and P. Yang. 2017. AERONET-based nonspherical dust optical models and effects on the VIIRS Deep Blue/SOAR over-water aerosol product. *Journal of Geophysical Research: Atmospheres* 122 (09 2017). https://doi.org/10.1002/2017jd027258

[22] J. Lee, Y. Shi, C Cai, P. Ciren, J. Wang, A. Gangopadhyay, and Z. Zhang. 2021. Machine Learning Based Algorithms for Global Dust Aerosol Detection from Satellite Images: Inter-Comparisons and Evaluation. *Remote Sensing* 13, 3 (2021). https://doi.org/10.3390/rs13030456

[23] R. Levy, S. Mattoo, L. Munchak, L. Remer, A. Sayer, and N. Hsu. 2013. The Collection 6 MODIS aerosol products over land and. *Atmospheric Measurement Techniques Discussions* 6 (11 2013), 159−259. https://doi.org/10.5194/amtd-6-159-2013

[24] MY. Liu, T. Breuel, and J. Kautz. 2017. Unsupervised Image-to-Image Translation Networks *(NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 700−708.

[25] M. Long, Y. Cao, J. Wang, and MI. Jordan. 2015. Learning transferable features with deep adaptation networks. *arXiv preprint arXiv:1502.02791* (2015).

[26] D. Lunga, H. Yang, A. Reith, J. Weaver, J. Yuan, and B. Bhaduri. 2018. Domain-Adapted Convolutional Networks for Satellite Image Classification: A Large-Scale Interactive Learning Workflow. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11, 3 (2018), 962−977. https://doi.org/10.1109/JSTARS.2018.2795753

[27] T. Mikolov, I. Sutskever, K. Chen, GS. Corrado, and J. Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111−3119.

[28] S. Platnick, KG. Meyer, MD. King, G. Wind, N. Amarasinghe, B. Marchant, GT. Arnold, Z. Zhang, PA.. Hubanks, RE. Holz, P. Yang, WL. Ridgway, and J. Riedi. 2017. The MODIS Cloud Optical and Microphysical Products: Collection 6 Updates and Examples From Terra and Aqua. *IEEE Transactions on Geoscience and Remote Sensing* 55, 1 (2017), 502−525. https://doi.org/10.1109/TGRS.2016.2610522

[29] S. Purushotham, W. Carvalho, T. Nilanon, and Y. Liu. 2017. Variational recurrent adversarial deep domain adaptation. In *International Conference on Learning Representations*.

[30] B. Sun and K. Saenko. 2016. Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision*. Springer, 443−450.

[31] E. Tzeng, J. Hoffman, T. Darrell, and K. Saenko. 2015. Simultaneous deep transfer across domains and tasks. In *Proceedings of the IEEE International Conference on Computer Vision*. 4068−4076.

[32] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell. 2014. Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474* (2014).

[33] C. Wang, S. Platnick, K. Meyer, Z. Zhang, and Y. Zhou. 2020. A machine-learning-based cloud detection and thermodynamic-phase classification algorithm using passive spectral observations. *Atmospheric Measurement Techniques* 13, 5 (2020), 2257−2277.

[34] DM. Winker, JL. Tackett, BJ. Getzewich, Z. Liu, MA. Vaughan, and RR. Rogers. 2013. The global 3-D distribution of tropospheric aerosols as characterized by CALIOP. *Atmospheric Chemistry & Physics* 13, 6 (2013).

[35] F. Wu and X. Zhuang. 2021. Unsupervised Domain Adaptation With Variational Approximation for Cardiac Segmentation. *IEEE Transactions on Medical Imaging* 40, 12 (2021), 3555−3567. https://doi.org/10.1109/TMI.2021.3090412

[36] L. Zhang, M. Lan, J. Zhang, and D. Tao. 2022. Stagewise Unsupervised Domain Adaptation With Adversarial Self-Training for Road Segmentation of Remote-Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), 1−13.

[37] L. Zhang, L. Zhang, and B. Du. 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine* 4, 2 (2016), 22−40.

[38] Y. Zhou, R. Levy, L. Remer, S. Mattoo, Y. Shi, and C. Wang. 2020. Dust Aerosol Retrieval Over the Oceans With the MODIS/VIIRS Dark-Target Algorithm: 1. Dust Detection. *Earth and Space Science* 7 (10 2020). https://doi.org/10.1029/2020EA001221