

Article

Discriminative Feature Learning Constrained Unsupervised Network for Cloud Detection in Remote Sensing Imagery

Weiyang Xie , Jian Yang, Yunsong Li *, Jie Lei , Jiaping Zhong  and Jiaojiao Li 

State Key Laboratory of Integrated Service Networks, School of Telecommunications Engineering, Xidian University, Xi'an 710071, China; wyxie@xidian.edu.cn (W.X.); jianyang_d@stu.xidian.edu.cn (J.Y.); jielei@mail.xidian.edu.cn (J.L.); jpzhong@stu.xidian.edu.cn (J.Z.); jjli@xidian.edu.cn (J.L.)

* Correspondence: ysl@mail.xidian.edu.cn; Tel.: +86-029-8820-3116

Received: 11 December 2019; Accepted: 20 January 2020; Published: 1 February 2020



Abstract: Cloud detection is a significant preprocessing step for increasing the exploitability of remote sensing imagery that faces various levels of difficulty due to the complexity of underlying surfaces, insufficient training data, and redundant information in high-dimensional data. To solve these problems, we propose an unsupervised network for cloud detection (UNCD) on multispectral (MS) and hyperspectral (HS) remote sensing images. The UNCD method enforces discriminative feature learning to obtain the residual error between the original input and the background in deep latent space, which is based on the observation that clouds are sparse and modeled as sparse outliers in remote sensing imagery. The UNCD enforces discriminative feature learning to obtain the residual error between the original input and the background in deep latent space, which is based on the observation that clouds are sparse and modeled as sparse outliers in remote sensing imagery. First, a compact representation of the original imagery is obtained by a latent adversarial learning constrained encoder. Meanwhile, the majority class with sufficient samples (i.e., background pixels) is more accurately reconstructed than the clouds with limited samples by the decoder. An image discriminator is used to prevent the generalization of out-of-class features caused by latent adversarial learning. To further highlight the background information in the deep latent space, a multivariate Gaussian distribution is introduced. In particular, the residual error with clouds highlighted and background samples suppressed is applied in the cloud detection in deep latent space. To evaluate the performance of the proposed UNCD method, experiments were conducted on both MS and HS datasets that were captured by various sensors over various scenes, and the results demonstrate its state-of-the-art performance. The sensors that captured the datasets include Landsat 8, GaoFen-1 (GF-1), and GaoFen-5 (GF-5). Landsat 8 was launched at Vandenberg Air Force Base in California on 11 February 2013, in a mission that was initially known as the Landsat Data Continuity Mission (LDCM). China launched the GF-1 satellite. The GF-5 satellite captures hyperspectral observations in the Chinese Key Projects of High-Resolution Earth Observation System. The overall accuracy (OA) values for Images I and II from the Landsat 8 dataset were 0.9526 and 0.9536, respectively, and the OA values for Images III and IV from the GF-1 wide field of view (WFV) dataset were 0.9957 and 0.9934, respectively. Hence, the proposed method outperformed the other considered methods.

Keywords: cloud detection; unsupervised network; adversarial learning; residual error; multivariate Gaussian distribution; remote sensing

1. Introduction

Remote sensing imaging technology such as multispectral (MS) imaging and hyperspectral (HS) imaging can perceive targets or natural phenomena remotely [1–3]. Depending on the wide-scale

monitoring capability, remote sensing images have been successfully applied to target detection [4], anomaly detection [5], and classification [6]. In [4], the current scenario and challenges of hyperspectral target detection are reviewed. Yuan et al. [5] introduced a method of hyperspectral anomaly detection that uses image pixel selection. Based on deep learning, the method in [6] increases the classification accuracy.

However, due to the significant impact of the atmospheric density and the cloud layer on the image acquisition process, most remote sensing images are inevitably polluted by clouds to different degrees [7], thereby resulting in inaccurate spectral characteristics of targets or natural phenomena, which reduces the exploitability of the image [8]. Consequently, cloud detection is an important preprocessing step for the promotion of the subsequent application and the improvement of the utilization rate of such cloud-contaminated remote sensing images.

Single-image cloud detection methods include threshold-based, machine-learning-based, and deep-neural-network-based methods. Threshold-based methods in traditional modeling are widely used because they are efficient and straightforward. Many cloud detection methods of this type have been proposed. Fisher [9] integrated the morphological features with the threshold-based method to realize cloud and cloud shadow detection in SPOT5 high-resolution geometric (HRG) imagery. A progressive refinement scheme (PRS) was introduced for the detection of cloud regions [10], which firstly utilized a threshold to obtain a coarse cloud detection map. An automatic multi-feature combined (MFC) method [11] was introduced for cloud and cloud shadow detection in GaoFen-1 (GF-1) wide field of view (WFV) imagery, which first implemented threshold segmentation to produce a preliminary cloud mask. Recently, Zhong et al. [12] proposed an improved threshold method that used strict and loose thresholds to produce a cloud detection map. The determination of a suitable threshold for complex surface areas of various cloud types remains challenging.

Most machine learning methods are supervised, such as support vector machine (SVM) [13] and scene learning (SL) [14]. These methods typically require hand-crafted features such as texture features [15] and morphological features [9] as the input of classifiers. It is difficult to capture the characteristics of clouds in complex scenes accurately. While these methods and techniques have made substantial advances [9,13–15], they are still far from being automatic and practical. The K-means method is a typical unsupervised cloud detection method [16]—the cloud detection performance of which should be further improved.

More recently, deep neural networks (DNNs) have exhibited strong performances in modeling and generalization from complex datasets [17–20], which can more accurately capture the characteristics of cloud-contaminated data than traditional methods. Shi et al. [21] and Goff et al. [22] exploited the convolutional neural network (CNN) to define clouds in the super-pixels regions. Ozkan et al. [23] used deep pyramid networks to realize cloud detection from RGB color remote sensing images. On MS images with nine bands from Landsat 8, Zi et al. [24] combined a double-branch principal component analysis network (PCANet) with SVM to achieve accurate cloud detection based on the coarse result that was obtained using a spectral threshold function. Shao et al. [25] extracted multiscale features via a method that was based on CNN for cloud detection. Yang et al. [26] proposed a cloud detection neural network (CDnet) based on CNN by simultaneously exploiting multiscale and global contextual information, preserving score map resolution, and refining object boundaries. Li et al. [27] introduced a multiscale convolutional feature fusion (MSCFF) method for remote sensing images that were captured by various sensors. Most methods using CNN are intended to extract deep characteristics. Experimental results have demonstrated that CNN-based methods can outperform traditional methods in cloud detection. However, these supervised methods depend on manually generated large datasets. It is challenging to produce a reliable training dataset for wide-scale monitoring with millions of pixels. Moreover, these recent CNN-based cloud detection methods were applied on datasets with few spectral bands. When they are applied to HS images with hundreds of bands, they are impractical due to the heavy computational burden and the large memory requirement. Therefore, the development of an unsupervised cloud detection method that can overcome the problem of insufficient training data with strong generalization performance is significant.

In view of this, we propose a discriminative feature learning constrained unsupervised network for cloud detection (UNCD) in MS and HS images. The proposed UNCD method depends on an important observation that clouds are sparse and modeled as sparse outliers [28–30]. Adversarial feature learning is conducted on an unsupervised neural network, namely, an autoencoder (AE) [31,32], to extract a compact representation of the original input image in the deep latent space. An image discriminator is introduced to correct image features in order to avoid the generalization of out-of-class features that is caused by the adversarial feature learning. With sufficient training samples, the background can be more distinctively reconstructed than the clouds. Besides, a multivariate Gaussian distribution is adopted to extract a discriminative feature matrix of the background in the latent space; hence, if the dataset contains clouds, the encoder will encourage the learning of the background distribution of the dataset. As a consequence, the residual error between these two lower-dimensional representations is beneficial to cloud detection. We conducted experiments on both MS and HS images and evaluated the performance of the proposed UNCD framework in terms of detection accuracy and generalization.

The contributions of this paper are fourfold:

1. A novel UNCD method is proposed to address the issue of insufficient training data in remote sensing images, especially hyperspectral data, in the field of cloud detection. To the best of our knowledge, in this paper, such an unsupervised adversarial feature learning model is utilized for the first time for MS and HS cloud detection.
2. Latent adversarial learning is introduced such that the AE focuses on extracting a compact representation of the input image in the latent space.
3. An image discriminator is used to prevent the generalization of out-of-class features.
4. A multivariate Gaussian distribution is adopted to extract a discriminative feature matrix of the background in the latent space, and the residual error between the low-dimensional representations of the original and background pixels is beneficial to cloud detection.

The remainder of this paper is organized as follows. In Section 2, AE and adversarial learning models are briefly described. Section 3 introduces the proposed UNCD framework. The experimental results and discussion are presented in Section 4. The discussion of the experimental results is introduced in Section 5. Finally, Section 6 presents the conclusions of this study.

2. Related Work

2.1. Generative Adversarial Network

A generative adversarial network (GAN) is a highly efficient generation model proposed by Goodfellow et al. [33] in 2014. GANs consist of a generator G and a discriminator D , which are trained against each other. The generator G inputs a random data \mathbf{z} whose probability distribution is $p_{\mathbf{z}}(\mathbf{z})$ and outputs a data $\mathbf{x}_{fake} = G(\mathbf{z})$. The discriminator D receives two inputs, real data \mathbf{x} with a probability distribution $p_{data}(\mathbf{x})$ and generated data $G(\mathbf{z})$. The discriminator is trained to identify the real data. Via successive adversarial learning, G generates increasingly many real data. The objective of this adversarial learning process can be expressed as [33]:

$$\min_G \max_D \mathbb{E}_{\mathbf{x} \sim p_{data}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} [\log (1 - D(G(\mathbf{z})))] \quad (1)$$

Due to its advantages in terms of generation and adversarial learning performance, we constructed a GAN model that is based on an encoder and a decoder for the improved production of discriminative features.

2.2. Variational Autoencoders

An AE updates the parameters by minimizing the distance between the input data and the reconstructed data, which is an unsupervised neural network. Many extensions that are based on

AEs have been proposed and broadly used, one of which is the variational AE (VAE) [32]. The VAE conducts variational inference to match the distribution of the hidden code vector of the AE with a predefined prior distribution. From the perspective of the probabilistic graph model, the encoder and decoder of the VAE can be interpreted as a probabilistic encoder and a probabilistic decoder, which are denoted as $p_{\theta}(\mathbf{x}|\mathbf{z})$ and $q_{\phi}(\mathbf{z}|\mathbf{x})$, respectively [32]. To match the distribution of the hidden code vector \mathbf{z} with a predefined prior distribution, variational inference is performed to optimize the variational lower bound on the marginal log-likelihood of each observation. Thus, the objective function of VAE can be expressed as:

$$\ell_{vae} = -KL\left(q_{\phi}\left(\mathbf{z}|\mathbf{x}^i\right)\|p_{\theta}\left(\mathbf{z}\right)\right) + \mathbb{E}_{q_{\phi}\left(\mathbf{z}|\mathbf{x}^i\right)}\left[\log p_{\theta}\left(\mathbf{x}^i|\mathbf{z}\right)\right], \tag{2}$$

where \mathbf{x}^i refers to a sample from the training dataset. The first term is the KL-divergence term, which represents the difference between the distribution of the extracted latent feature samples and the Gaussian distribution. The smaller the difference, the closer the distribution of the extracted latent features is to the Gaussian distribution. The second term is the AE’s reconstruction error, which represents the expectation of the reconstruction error, namely, the gap between the input and the output. The smaller the expected value of the reconstruction error, the closer the output can be to the input. The hidden code vector \mathbf{z} is sampled from $\mathbf{z}^{i,l} = \mu^i + \sigma^i \odot \varepsilon^l$, where $\varepsilon^l \sim N(0, I)$. Both the encoder and the decoder originally leverage the sigmoid function as the nonlinear activation function. Recently, the rectified linear unit (ReLU) [34] has been widely used as the nonlinear activation function.

3. Proposed Method

Figure 1 illustrates the overall framework of our proposed UNCD method. Let $\mathbf{Y} = \{\mathbf{y}^i\}_{i=1}^{M \times N}$, $\mathbf{y}^i \in \mathbb{R}^{L \times 1}$ denote an input remote sensing image with $M \times N$ spectral vectors, and each spectral vector contains L dimensions. Firstly, the underlying characteristics are extracted by an encoder E and a decoder De trained via an adversarial approach, thereby yielding a compact representation $\mathbf{Z} = \{\mathbf{z}^i\}_{i=1}^{M \times N}$, $\mathbf{z}^i \in \mathbb{R}^{l \times 1}$, $l < L$ of the redundant remote sensing images. Based on the observation that clouds are sparse and modeled as sparse outliers, some powerful constraints are imposed on the unsupervised neural network for the extraction of a discriminative feature matrix $E(De(\mathbf{Z}))$ of the background, which has the same dimension as \mathbf{Z} . Due to the reconstruction ability of AEs and the generation capability of GANs, it is possible to obtain features of the full image and the background in deep latent space, respectively. As a consequence, \mathbf{Z} and $E(De(\mathbf{Z}))$ are compact representations of the original image and background, respectively. Specifically, the residual error between \mathbf{Z} and $E(De(\mathbf{Z}))$ is beneficial to cloud detection. More details of each step are described as follows.

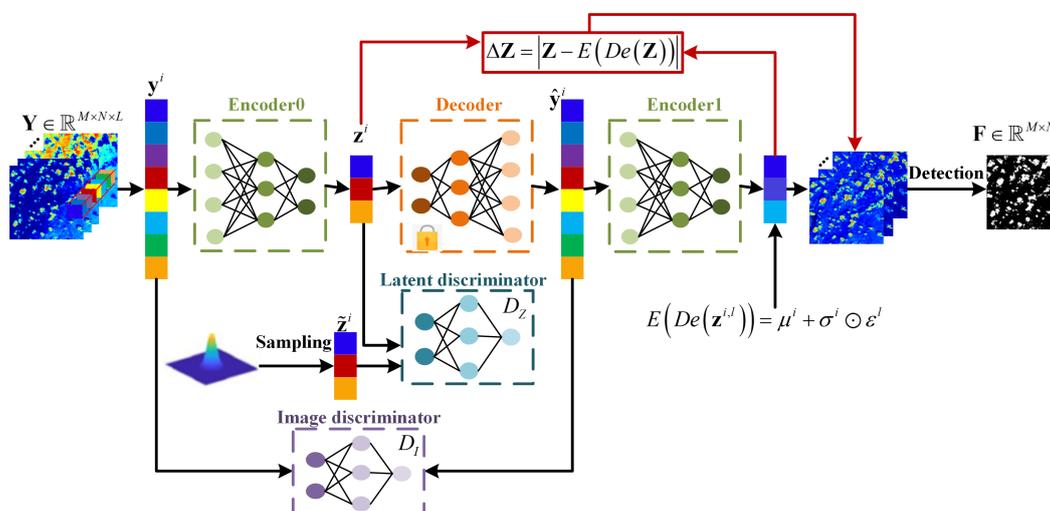


Figure 1. The detailed description of the proposed unsupervised network for cloud detection (UNCD) in remote sensing imagery.

3.1. Constructing the Residual Error in the Latent Space

The proposed UNCD method relies on two assumptions for cloud detection. One is that the background can be distinguished from the clouds in the latent feature space and most of a remote sensing scene is background, while clouds are sparse. When a remote sensing image \mathbf{Y} that contains clouds is input into the network, the network is expected to generate a compact representation of \mathbf{Y} and to reconstruct a corresponding background via an unsupervised approach. Consequently, the residual error refers to the deep latent feature space in which the clouds are enhanced, and the background is suppressed:

$$\Delta\mathbf{Z} = |\mathbf{Z} - E(De(\mathbf{Z}))|. \quad (3)$$

The encoder learns a mapping from input \mathbf{Y} into the deep latent space, which produces a feature matrix \mathbf{Z} that preserves the essential information including the clouds and background. Then, \mathbf{Z} is input into the decoder De to reconstruct \mathbf{Y} . Since sufficient training samples are available for most of the background and clouds are sparse with limited training samples in a remote sensing scene, the decoder produces small reconstruction error for the background region but relatively large reconstruction error for the clouds. The adversarial learning terms and physical constraints regarding the characteristics of the cloud-contaminated images are imposed on the network to enhance the capability of discriminative feature learning performance.

3.2. Adversarial Feature Learning Term

Adversarial feature learning is used in the latent space to extract distinctive features. As shown in Figure 1, the encoder E^0 acts as a generator to produce latent feature variables $\mathbf{z}^i \sim q(E^0(\mathbf{y}^i))$, which strives to fool the latent discriminator D_z with $\mathbf{z}^i \sim p(\mathbf{z}^i)$ sampled from the prior distribution. The latent discriminator D_z aims to distinguish between input that comes from the encoder $\mathbf{z}^i \sim q(E^0(\mathbf{y}^i))$ versus that from prior distribution $\mathbf{z}^i \sim p(\mathbf{z}^i)$. As a consequence, the generator (encoder) E^0 attempts to minimize this objective against the latent discriminator D_z that attempts to solve $\min_{E^0} \max_{D_z} Loss_{adv_z}$, where $Loss_{adv_z}$ is denoted as:

$$Loss_{adv_z} = \mathbb{E}_{\mathbf{z}^i \sim N(0, I)} [\log D_z(\mathbf{z}^i)] + \mathbb{E}_{\mathbf{z}^i \sim q(E^0(\mathbf{y}^i))} [\log(1 - D_z(E^0(\mathbf{y}^i)))] \quad (4)$$

Here, we set the prior distribution $\mathbf{z}^i \sim p(\mathbf{z}^i)$ as multivariate Gaussian distribution to extract feature variables that are more beneficial to cloud detection considering that clouds are sparse and the majority is background.

3.3. Adversarial Image Learning Term

The adversarial feature learning in the previous part may generate out-of-class features due to the capability of GAN to generate new variants. To generate features of the original remote sensing images, we constrain the network by the image discriminator D_I to avoid the generation of out-of-class features, as illustrated in Figure 1. Via this approach, the decoder De and the image discriminator D_I are combined into a GAN again, in the training phase of which the decoder De is changed to a generator to ensure that the generated image can fool the image discriminator D_I . The image discriminator D_I aims to distinguish whether its input comes from the generator De or the real input. The optimization problem is to achieve $\min_{De} \max_{D_I} Loss_{adv_I}$, where $Loss_{adv_I}$ can be expressed as:

$$Loss_{adv_I} = \mathbb{E}_{\mathbf{y}^i \sim p_{data}} [\log D_I(\mathbf{y}^i)] + \mathbb{E}_{\mathbf{z}^i \sim N(0, I)} [\log(1 - D_I(De(\mathbf{z}^i)))] \quad (5)$$

The first term indicates that the discriminator attempts to maximize the output of the real sample to make it closer to 1. The second term indicates that by optimizing the discriminator, the output to

the generated sample becomes closer to zero. At the same time, the generator is optimized, thereby making the output of the discriminator on the generated samples closer to 1.

3.4. Latent Representation of the Background

The Gaussian distribution that is imposed on the network focuses on reconstructing the majority class, namely, the background, with relatively sufficient samples. Consequently, the generated image $De(\mathbf{Z})$ with the same size as the input is closer to the background than to the clouds. Due to the strong representational capability of AE, it can still reconstruct the sparse pixels (i.e., the clouds). To further reduce the proportion of clouds in the reconstructed image $De(\mathbf{Z})$, we impose the representation consistency constraint in the deep latent feature space, that is, we enforce the satisfaction of a multivariate Gaussian distribution by $E(De(\mathbf{Z}))$ via an unsupervised approach that better accords with the background characteristics:

$$Loss_{Gauss} = KL(q(E(De(\mathbf{z}^i)) | De(\mathbf{z}^i))) \| p(E(De(\mathbf{z}^i))), \quad (6)$$

where $E(De(\mathbf{z}^{i,l})) = \mu^i + \sigma^i \odot \varepsilon^l$ and $\varepsilon^l \sim N(0, I)$. The encoded output of the second encoder is represented by $E(De(\mathbf{z}^i))$, and the output of the first decoder is denoted by $De(\mathbf{z}^i)$. During the training of encoder1, only the parameters of the encoder are optimized, while the parameters of the decoder De are kept fixed as, illustrated in Figure 1. $E(De(\mathbf{z}^{i,l}))$ is the projection of the reconstructed spectral vector $De(\mathbf{z}^i)$ in the low-dimensional latent space, which is the enhanced background representation obtained by imposing the representation consistency constraint. Thus, the distinctive residual error for distinguishing between the background and the clouds in the low-dimensional latent space can be calculated via Equation (3), and an example is represented in Figure 2.

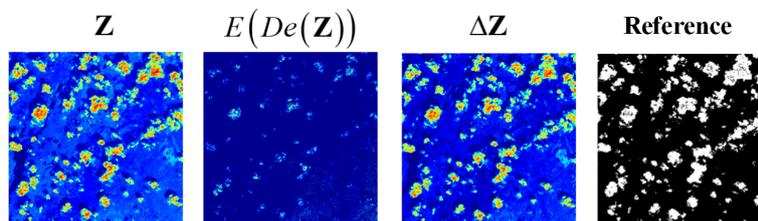


Figure 2. An example of a residual error map.

3.5. Reconstruction Loss

In a traditional AE network, the input data are encoded by the hidden layers to be decoded correctly in the output. The parameters are selected such that the objective of minimizing the following cost function is realized:

$$Loss_r = \|\mathbf{y}^i - \hat{\mathbf{y}}^i\|_2, \quad (7)$$

where \mathbf{y}^i denotes an input spectral vector with L dimensions and $\hat{\mathbf{y}}^i$ denotes the corresponding reconstructed spectral vector. Finally, the combination of the aforementioned losses yields

$$Loss = Loss_{adv_z} + Loss_{adv_I} + Loss_{Gauss} + Loss_r. \quad (8)$$

The model is trained and updated via the stochastic gradient descent (SGD) algorithm. When the model loss converges, the parameters, including the weight matrix and the bias, are obtained. According to Figure 2, the compact representation of the original image, \mathbf{Z} , contains features of both clouds and background regions, whereas $E(De(\mathbf{Z}))$ can well represent the background information with the clouds restricted. As a consequence, the residual error $\Delta\mathbf{Z}$ obtained from the pixel-wise difference enhances the clouds while suppressing the background in the lower-dimensional feature space, which facilitates the detection of clouds as shown in Figure 3. Subsequently, an adaptive weighting method that was

proposed in our previous work [35,36] is applied on each dimension of the residual error $\Delta\mathbf{Z}$ to discard redundant information and to construct a comprehensive map, where the structure tensor (ST) is utilized. The ST of the i th dimension in the residual error $\Delta\mathbf{Z}$ can be defined as [35,36]:

$$\mathbf{S}^i = \begin{bmatrix} (\Delta\mathbf{Z}^i)_x^2 & (\Delta\mathbf{Z}^i)_x(\Delta\mathbf{Z}^i)_y \\ (\Delta\mathbf{Z}^i)_x(\Delta\mathbf{Z}^i)_y & (\Delta\mathbf{Z}^i)_y^2 \end{bmatrix}, \tag{9}$$

where $(\Delta\mathbf{Z}^i)_x = \frac{\partial\Delta\mathbf{Z}^i}{\partial x}$ and $(\Delta\mathbf{Z}^i)_y = \frac{\partial\Delta\mathbf{Z}^i}{\partial y}$ represent the derivatives of the i th dimension of the residual error $\Delta\mathbf{Z}$ along the x and y directions, respectively. Since this structure tensor \mathbf{S}^i of the i th dimension is a semi-determined matrix, it can be decomposed into [35,36]:

$$\mathbf{S}^i = \begin{bmatrix} \eta_1^i & \eta_2^i \end{bmatrix} \begin{bmatrix} \lambda_1^i & 0 \\ 0 & \lambda_2^i \end{bmatrix} \begin{bmatrix} \eta_1^i \\ \eta_2^i \end{bmatrix}, \tag{10}$$

where λ_1^i and λ_2^i are the non-negative eigenvalues of the i th dimension and η_1^i, η_2^i are the corresponding eigenvectors. The larger eigenvalue is represented by λ_1^i . As discussed in our previous work [35,36], the relative larger eigenvalue can represent the response intensity of each pixel in the corresponding dimension of $\Delta\mathbf{Z}$. Therefore, the weight vector for each dimension of $\Delta\mathbf{Z}$ is calculated from the relative larger eigenvalue. The larger the edge intensity of the i th dimension, the more structural information the i th dimension of $\Delta\mathbf{Z}$ contains. Consequently, the i th dimension should occupy a larger proportion. The $(i - 1)$ th dimension should also occupy a larger proportion. If the larger eigenvalues of all the dimensions are same, all the dimensions will have the same proportion, which is the same as the result obtained with the averaging function. In summary, the weighted residual error $\Delta\hat{\mathbf{Z}}$ is calculated as [35,36]:

$$\Delta\hat{\mathbf{Z}} = \sum_{i=1}^l \frac{\lambda_1^i}{\sum_{i=1}^l \lambda_1^i} \Delta\mathbf{Z}^i, \tag{11}$$

where l is the dimension of $\Delta\mathbf{Z}$. To further consider that the adjacent pixels have high correlations with each other, we utilize a guided filter [37] to move the neighboring pixels to the same object for both the background and clouds. Finally, we use an iterative optimization step to further increase the detection accuracy, namely, we multiply the initial detection map by the residual error $\Delta\mathbf{Z}$, and repeat the detection procedure until the following stopping rule is satisfied:

$$\left| \mathbf{C}_D^{f-1} - \mathbf{C}_D^f \right| \leq 10^{-4}, \tag{12}$$

where \mathbf{C}_D^{f-1} and \mathbf{C}_D^f represent the $(f - 1)$ th and f th detection maps, respectively. According to this stopping rule, if the $(f - 1)$ th and f th detection maps are highly similar, the iterative optimization will terminate.

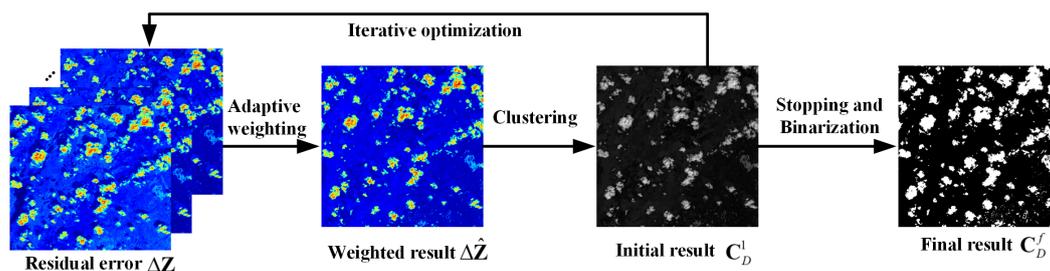


Figure 3. Detailed description of the proposed iterative optimization method.

3.6. Data Description

3.6.1. Landsat 8 Dataset

The Landsat 8 dataset was made available online <https://www.usgs.gov/land-resources/nli/landsat/spatial-procedures-automated-removal-cloud-and-shadow-sparcs-validation>, and is widely used to evaluate the cloud detection methods due to its wide coverage and high resolution. This dataset consists of 80 MS remote sensing images, with details described in [38]. The Landsat 8 dataset was collected globally using two instruments: OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor). We use the data from both of these. All the channels are used for cloud detection. The spatial resolution of each Landsat image is 30 m—that is, each pixel represents information on an area of 900 m². The location of the dataset is global. Each image contains 1000 × 1000 pixels in the spatial domain and ten bands in the spectral domain. The corresponding reference maps are shown in Figure 4.

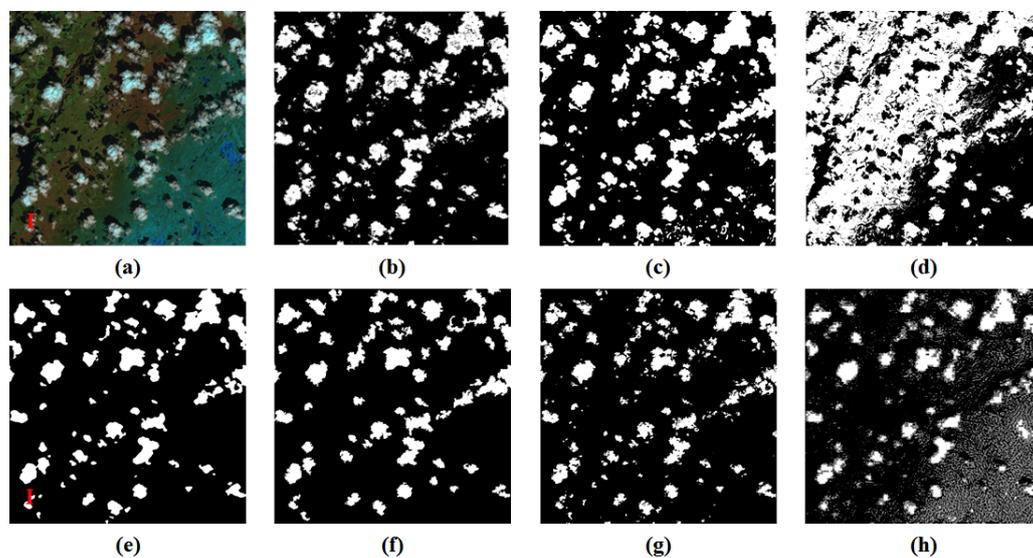


Figure 4. Cloud detection maps of the compared methods: (a) Input images labeled I from the Landsat 8 dataset, (b) reference detection map, (c) proposed, (d) K-means, (e) progressive refinement scheme (PRS), (f) principal component analysis network (PCANet), (g) support vector machine (SVM), and (h) scene learning (SL).

3.6.2. GF-1 WFV Dataset

The GaoFen-1 (GF-1) satellite was launched by China. GaoFen means “high resolution” in Chinese. Detailed information on the GF-1 satellite and wide field of view (WFV) images are available online <http://sendimage.whu.edu.cn/en/resources/mfc-validation-data/>. The GF-1 WFV image considered here is a Class 2A product produced via relative radiation correction and system geometry correction. The Class 1A data are the original digital product for regular radiation calibration, while the Class 2A data were generated after system geometry correction, where all pixels are re-sampled to 16-m resolution with 10-bit data. The false-color images and the corresponding reference maps are shown in Figure 5.

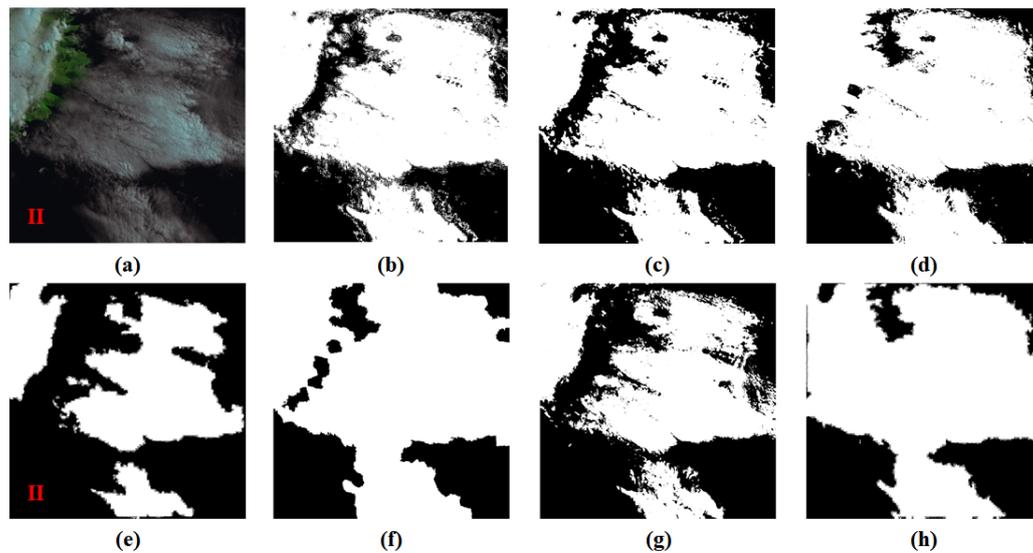


Figure 5. Cloud detection maps of the compared methods: (a) Input images labeled II from the Landsat 8 dataset, (b) reference detection map, (c) proposed, (d) K-means, (e) PRS, (f) PCANet, (g) SVM, and (h) SL.

3.6.3. GF-5 Hyperspectral Dataset

The GaoFen-5 (GF-5) satellite can capture hyperspectral observations in the Chinese Key Projects of the High-Resolution Earth Observation System. Six payloads were carried on it: an Advanced Hyperspectral Imager (AHSI), a Visual and Infrared Multispectral Imager (VIMI), an Atmospheric Infrared Ultra-Spectral Sounder (AIUS), a Greenhouse Gases Monitoring Instrument (GMI), an Environmental Monitoring Instrument (EMI), and a Directional Polarization Camera (DPC). The spectral coverage of these sensors ranges from ultraviolet to long-wave infrared bands. Two HS images captured by the GF-5 AHSI sensor were used to evaluate the performance of the proposed method. Each image has a size of $430 \times 430 \times 180$ pixels. The false-color images are shown in Figure 6. The reference maps for this dataset have not been published yet.

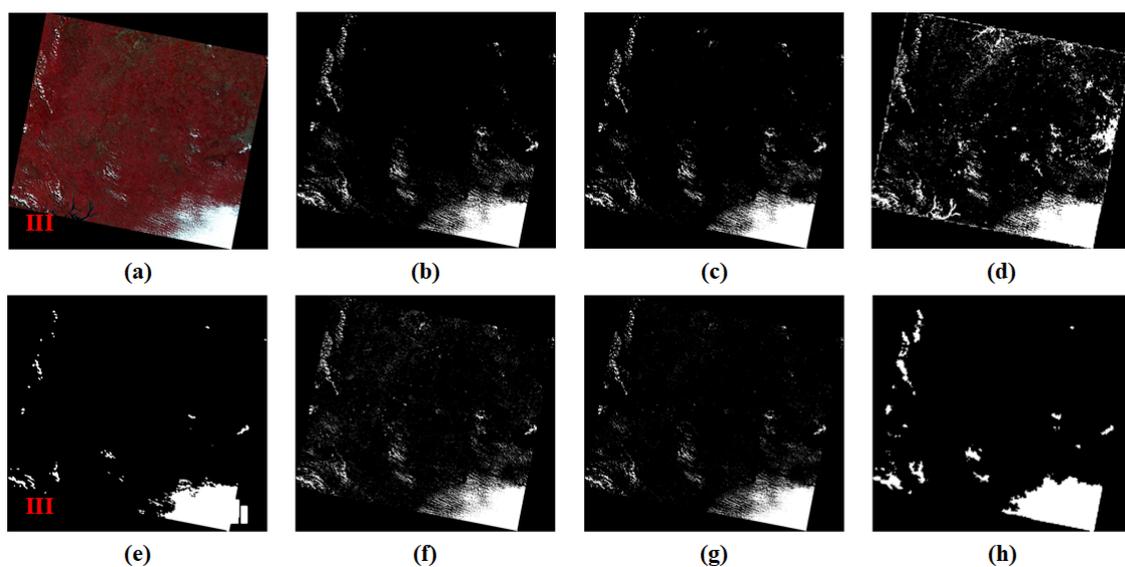


Figure 6. Cloud detection maps of the compared methods: (a) Input images labeled III from the GaoFen-1 (GF-1) wide field of view (WFV) dataset, (b) reference detection map, (c) proposed, (d) K-means, (e) PRS, (f) PCANet, (g) SVM, and (h) SL.

4. Experimental Results

We comprehensively evaluated the proposed method on three real datasets, which included MS and HS images captured by various imaging sensors over various scenes. First, we describe the dataset. Then, we introduce the compared methods and evaluation criterion. Third, we further investigate the performance of the proposed UNCD method, both qualitatively and quantitatively. Finally, we analyze the impact of the network structure on the detection performance.

4.1. Experimental Setting

Due to insufficient samples in the remote sensing imagery, we fixed the depth of our UNCD to 2 to avoid overfitting. Inspired by [39], the number of hidden nodes was fixed to $\sqrt{L} + 1$, where L is the number of bands of the input remote sensing image. The leaky ReLU (LReLU) was used as the nonlinear activation function with a slope of 0.2 to compress the negative input and retain the negative part of the information to a certain extent. As the number of epochs increases, both the detection performance and the computational complexity increase. Thus, we set the number of epochs to 1000 as a trade-off. The batch size was fixed to the number of pixels in the spatial domain, namely, $M \times N$ for each remote sensing image. The learning rate was set to 0.01. The parameters introduced above were set to the default values in our experiments, and they can be tuned by the user for the optimal results.

We conducted experiments on 8 NVIDIA Tesla k80 graphics cards based on a system running Python 3.6.0 and TensorFlow 1.10.0. All compared methods were implemented in MATLAB R2017a.

4.2. Compared Methods and Evaluation Criterion

To evaluate the performance of the proposed UNCD model, several representative cloud detection methods that are frequently cited in the literature, namely, K-means, PRS, SVM, PCANet, and SL, were employed for comparison in terms of both visual effects and quantitative evaluations. The K-means method is a typical unsupervised method. The PRS method yields satisfactory results on several remote sensing images, and is also an unsupervised method. Since the PRS method is only applicable to RGB images, we combined band 4-red, band 3-green, and band 2-blue into RGB images. The SVM, PCANet, and SL methods are supervised learning methods for cloud detection.

To comprehensively evaluate the cloud detection results, the three commonly used evaluation criteria, namely, the area under the curve (AUC) [40] of the receiver operating characteristic (ROC), the overall accuracy (OA), and the kappa coefficient (Kappa), were employed.

The area under the curve (AUC) [40] is widely used in objective evaluation indices. The AUC can identify general trends in the detection performance. The larger the AUC value of the ROC curve, the better the performance. The OA is defined as [41]:

$$OA = \frac{TP + TN}{TP + TN + FN + FP}, \quad (13)$$

where true positives (TPs), true negatives (TNs), false positives (FPs) and false negatives (FNs) represent the number of correctly detected cloud pixels, the number of correctly detected non-cloud pixels, the number of false-alarm pixels, and the number of missed cloud pixels, respectively. The kappa coefficient reflects the agreement between a final cloud detection image and the ground-truth map. Compared with the overall accuracy (OA), the kappa coefficients can more objectively reflect the accuracy of the results. The larger the value of the kappa coefficient, the higher the accuracy of the result will be. The kappa coefficient is calculated as [41]:

$$Kappa = \frac{OA - P}{1 - P}, \quad (14)$$

where

$$P = \frac{(TP + FP)(TP + FN)}{(TP + FP + FN + TN)^2} + \frac{(FN + TN)(FP + TN)}{(TP + FP + FN + TN)^2}. \quad (15)$$

4.3. Cloud Detection Results

4.3.1. Landsat 8 Dataset Results

The reference maps and the visual cloud detection results that were obtained by the competing methods for two images from the Landsat 8 dataset are shown in Figures 4 and 5. The reference map is published at the same time as the dataset and used together with the data. Moreover, the provider of the dataset marks the annotation. Image I is an example of a case with thin clouds, and Image II is an example of a case with thick clouds. With these two images, we compared our method with K-means, PRS, SVM, PCANet, and SL. We implemented these compared methods via the publicly released codes. According to Figures 4 and 5, the proposed UNCD method caused the smallest visual difference between the reference map and the detection map compared to K-means, PRS, SVM, PCANet, and SL. As illustrated in Figures 4 and 5, the K-means method yielded many noise-like detection results. The PRS method realized minimal improvement for Image I with respect to the K-means method, and thin clouds were undetected in large areas in Image I. The SL method generated some detection mistakes in the image due to the thinness and thickness of the clouds. The PCANet and SVM methods outperformed the K-means, PRS, and SL methods. In particular, the proposed UNCD method could accurately distinguish clouds and complex background scenes. The reason is that the proposed detection method extracts the spectral features of the image while utilizing the spatial background information between adjacent pixels. In comparison to these methods, the advantage of the proposed UNCD method is its robust detection performances for different clouds with different scales in different scenes. The objective evaluations, including AUC, OA, and Kappa obtained by all the considered algorithms, are reported in Table 1, which complies with the visual observation. Concretely, the AUCs obtained by the proposed UNCD method were 0.9543 and 0.9637 for Images I and II, respectively, which are much higher than those obtained by the second-best approach in each case, 0.8485 (PRS method) and 0.8848 (K-means method). The OA and Kappa obtained by the proposed UNCD method were also the highest, and much higher than those obtained by the second-best method.

Table 1. The detection performance of different methods on Images I and II from the Landsat 8 dataset. AUC: area under the curve; Kappa: kappa coefficient; OA: overall accuracy. The bold fonts indicate the best results.

Image I	Proposed	K-means	PRS	PCANet	SVM	SL
AUC	0.9543	0.7979	0.8485	0.8468	0.8286	0.7401
OA	0.9526	0.6745	0.9391	0.9359	0.9343	0.8448
Kappa	0.8719	0.3545	0.7747	0.7646	0.7503	0.4814
Image II	Proposed	K-means	PRS	PCANet	SVM	SL
AUC	0.9637	0.8848	0.8184	0.8701	0.8762	0.8593
OA	0.9536	0.9062	0.7668	0.8962	0.8409	0.8821
Kappa	0.9016	0.7899	0.5560	0.7659	0.6845	0.7365

4.3.2. GF-1 WFV Dataset

The reference maps and detection maps obtained by the compared methods are shown in Figures 6 and 7. The detection results obtained by the proposed method were similar to the reference maps, indicating that the proposed method could obtain a satisfactory detection performance. It is apparent that the UNCD method achieved better detection results than the K-means, PRS, PCANet, SVM, and SL methods. For the GF-1 WFV dataset, Table 2 lists the corresponding quantitative metrics. In addition, the competing methods are also reported in Table 2. From Table 2, the AUC values, OA, and Kappa obtained by the proposed UNCD method were the best among the compared methods. The reported OA scores on Images III and IV (0.9957 and 0.9934) obtained by the proposed UNCD

method were higher than that obtained by the second-best methods (0.9835 for SVM on Image III and 0.9690 for SL on Image IV).

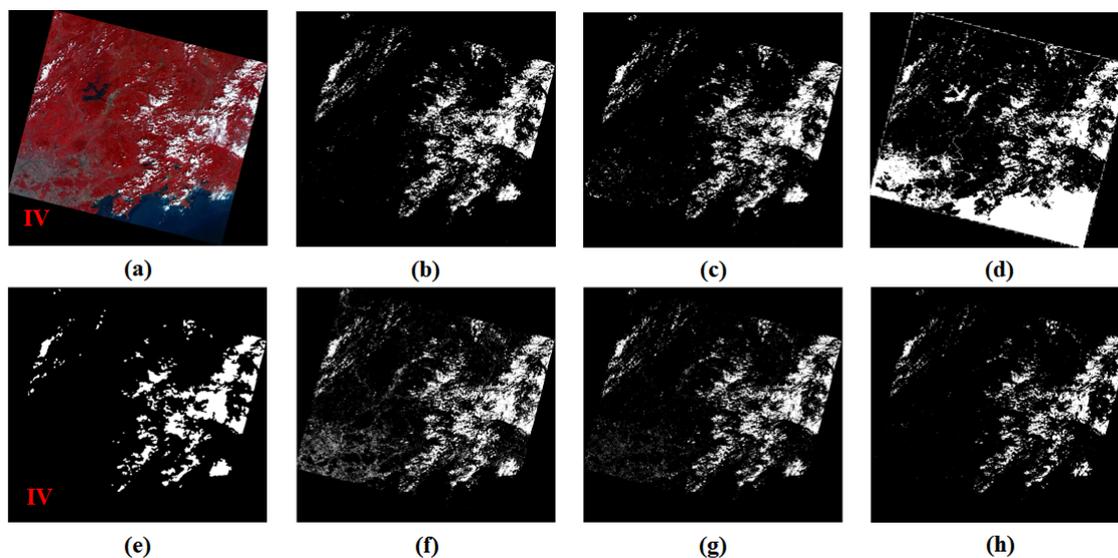


Figure 7. Cloud detection maps of the compared methods: (a) Input images labeled IV from the GF-1 WFV dataset, (b) reference detection map, (c) proposed, (d) K-means, (e) PRS, (f) PCANet, (g) SVM, and (h) SL.

Table 2. The detection performance of different methods on Images III and IV from the GaoFen-1 (GF-1) wide field of view (WFV) dataset. The bold fonts indicate the best results.

Image III	Proposed	K-means	PRS	PCANet	SVM	SL
AUC	0.9676	0.9310	0.8387	0.9168	0.8873	0.8840
OA	0.9957	0.9491	0.9726	0.9815	0.9835	0.9661
Kappa	0.9636	0.6646	0.7434	0.8403	0.8458	0.7261
Image IV	Proposed	K-means	PRS	PCANet	SVM	SL
AUC	0.9860	0.8753	0.8586	0.9304	0.8980	0.8280
OA	0.9934	0.8585	0.9623	0.9591	0.9743	0.9690
Kappa	0.9630	0.4621	0.7551	0.7732	0.8338	0.7743

4.3.3. GF-5 Hyperspectral Dataset

The GF-5 dataset was used to evaluate the feasibility of our proposed UNCD method on real hyperspectral data. The cloud detection results are displayed in Figures 8 and 9. Image V in Figure 8 contains large clouds, while Image VI in Figure 9 contains many small clouds. It can be observed that our method was capable of discriminating clouds of different sizes from the background pixels. The K-means method introduced many false positives. PRS and PCANet failed to detect all the clouds. By contrast, the SVM and SL methods had better performance while also generating some detection mistakes. For example, the buildings in Image V were still detected as clouds by the SVM and SL methods.

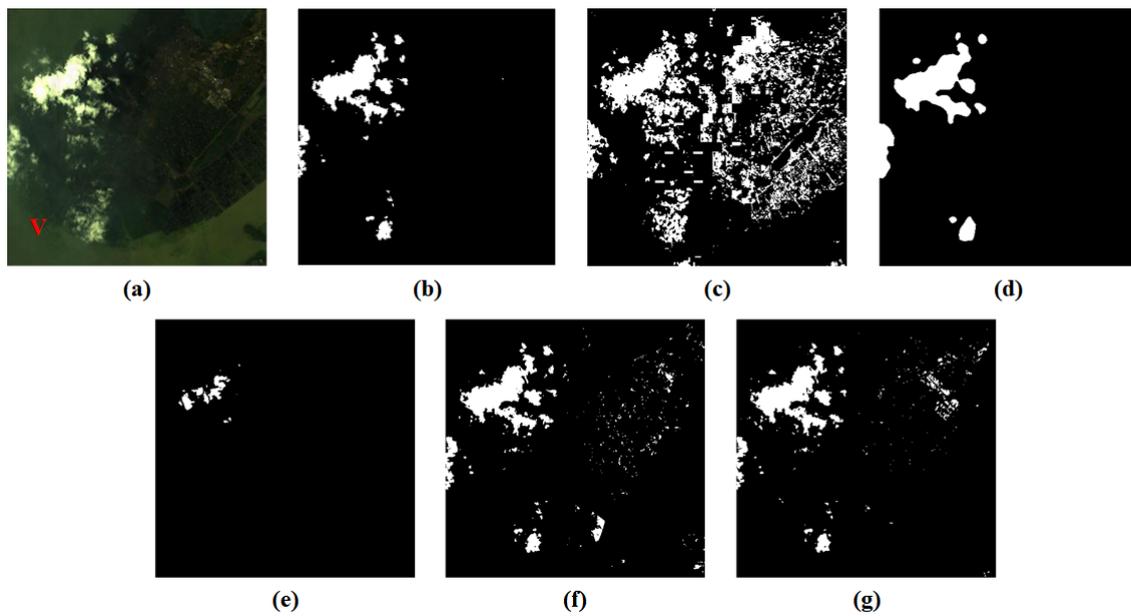


Figure 8. Cloud detection maps of the compared methods: (a) Input images labeled V from the GF-5 dataset, (b) proposed, (c) K-means, (d) PRS, (e) PCANet, (f) SVM, and (g) SL.

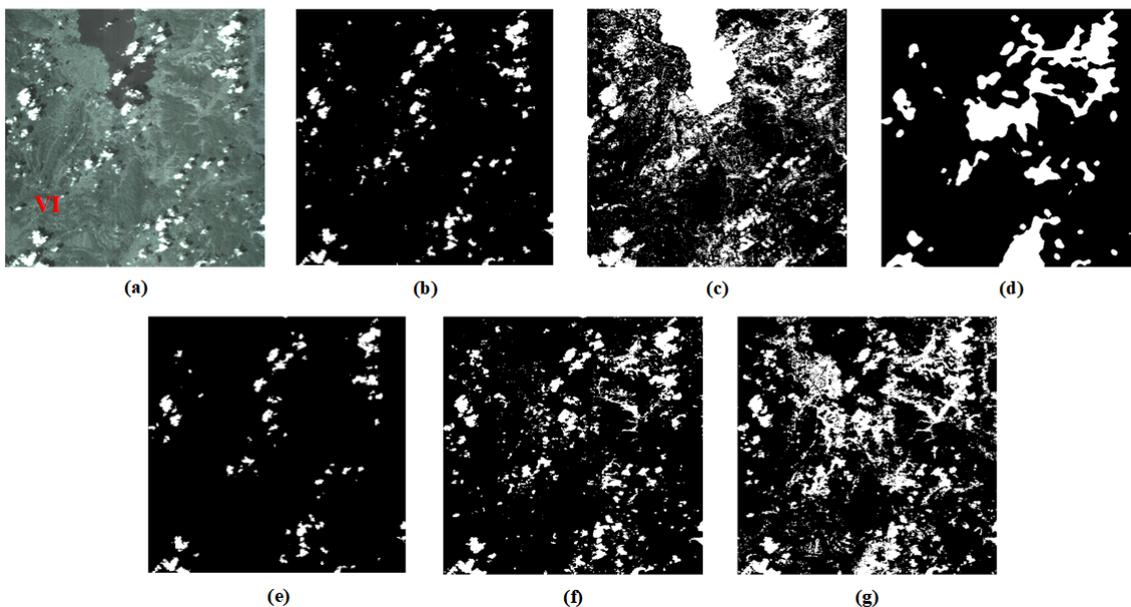


Figure 9. Cloud detection maps of the compared methods: (a) Input images labeled VI from the GF-5 dataset, (b) proposed, (c) K-means, (d) PRS, (e) PCANet, (f) SVM, and (g) SL.

4.4. Component Analysis

This section analyzes the effects of the significant processing components on the detection performance on each dataset. Since the reference maps of the GF-5 hyperspectral dataset are not publicly available, the objective evaluations could not be obtained for this dataset. Therefore, four remote sensing images coming from the Landsat 8 dataset and the GF-1 WFV dataset were used to evaluate the effect of each component objectively. Three comparison experiments were conducted. In the first experiment, only the AE was considered, which is a basic model for joint encoder–decoder training. The second considered AE with additional adversarial training (a latent feature discriminator). The third utilized the proposed method. The AUC, OA, and Kappa values were calculated as reported in Table 3. The better the detection performance, the higher the AUC, OA, and Kappa values. The AUC

values were 0.9254, 0.9025, 0.9436, and 0.9462 for four images when only using AE. When applying the adversarial feature learning, the AUC values were improved to 0.9477, 0.9379, 0.9506, and 0.9689, respectively. When the multivariate Gaussian distribution was introduced, the method yielded the best AUC values, which reached 0.9543, 0.9637, 0.9676, and 0.9860, respectively. Similarly, the other two indicators (i.e., OA and Kappa) also increased. These results demonstrate that each component of the proposed UNCD method has a positive influence on the cloud detection performance.

Table 3. The area under the curve (AUC), overall accuracy (OA) and Kappa for component analysis of UNCD for different datasets. AE: autoencoder. The bold fonts indicate the best results.

AUC				
Component	Image I	Image II	Image III	Image IV
Only AE	0.9254	0.9025	0.9436	0.9462
AE with D_z	0.9477	0.9379	0.9506	0.9689
Proposed	0.9543	0.9637	0.9676	0.9860
OA				
Component	Image I	Image II	Image III	Image IV
Only AE	0.9211	0.8749	0.9917	0.9897
AE with D_z	0.9248	0.9310	0.9922	0.9923
Proposed	0.9526	0.9536	0.9959	0.9934
Kappa				
Component	Image I	Image II	Image III	Image IV
Only AE	0.8678	0.7469	0.9258	0.9336
AE with D_z	0.8540	0.8535	0.9309	0.9515
Proposed	0.8719	0.9016	0.9636	0.9630

5. Discussion

According to the values of AUC, OA, and Kappa in Tables 1 and 2 and the visual observations in Figures 4–9 for various types of datasets, compared to several state-of-the-art methods, the proposed UNCD method performed the best in detection according to both the objective evaluation results and the visual observations. The superior performance of the proposed method is due to the latent adversarial learning constrained encoder, the image discriminator, and the multivariate Gaussian distribution in the network architecture. It can be concluded from the component analysis experiments that each part had a positive impact on the detection results. While the proposed UNCD method yielded promising results in cloud detection, several areas were identified during experiments for improvement in future work. It is worthwhile to further utilize the data characteristics of remote sensing images in order to optimize unsupervised networks and to improve the cloud detection performance. In addition, the network architecture can be enhanced by adding some loss functions and constraints.

The OA values of Images I–IV that were obtained by the proposed UNCD method were 0.9526, 0.9536, 0.9957, and 0.9934, respectively. Our method outperformed the second-best method by 1.44%, 5.23%, 1.24%, and 1.96%, respectively. While the performance of our method was the best among the considered methods, there is still room for improvement. Moreover, the accuracy of cloud detection was not high, and there were still many missed detections. However, other comparison methods also have this problem. Meanwhile, the proposed UNCD method is devoted to cloud detection. In the future, we will further realize cloud shadow detection. At the same time, to make the method universal, we will conduct additional experiments on datasets acquired by additional sensors.

6. Conclusions

In this paper, we proposed a discriminative feature learning constrained unsupervised network for cloud detection (UNCD) for remote sensing imagery. The induced latent discriminator, image

discriminator, and multivariate Gaussian distribution consider the fact that clouds are sparse and modeled as outliers and realize a discriminative residual map between the original input and the background. Based on the analysis of the strong correlation between adjacent pixels, a guided filter is employed on the residual map to obtain an initial detection map. To further improve the detection performance, an iterative optimization algorithm is introduced which terminates automatically if the stopping condition is satisfied. Extensive experimental results on several datasets demonstrate that the proposed UNCD not only realizes a more favorable detection performance but also generalizes better to different datasets compared with other state-of-the-art methods. Moreover, the OA values for Images III and IV from the GF-1WV dataset were 0.9957 and 0.9934, respectively. This signifies that our algorithm performs better than other known algorithms. In future work, we will expand the datasets used for the experiments to improve the performance of the algorithm.

Author Contributions: W.X. conceived and designed the study; J.Y. and J.L. (Jie Lei) performed the experiments and analyzed the result data; W.X. and Y.L. investigated related work; J.Z. and J.L. (Jiaojiao Li) reviewed and edited the manuscript; W.X. wrote the paper. All authors read and approved the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 61801359, Grant 61571345, Grant 91538101, Grant 61501346, Grant 61502367, and Grant 61701360, in part by the Young Talent fund of University Association for Science and Technology in Shaanxi of China under Grant 20190103, in part by the Special Financial Grant from the China Postdoctoral Science Foundation under Grant 2019T120878, in part by the 111 project under Grant B08038, in part by the joint fund project of NSFC under Grant U1704130, in part by the Fundamental Research Funds for the Central Universities under Grant JB180104, in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2019JQ153, Grant 2016JQ6023, and Grant 2016JQ6018, in part by the General Financial Grant from the China Postdoctoral Science Foundation under Grant 2017M620440, in part by the Yangtze Rive Scholar Bonus Schemes under Grant CJT160102, and in part by the Ten Thousand Talent Program.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AHSI	Advanced Hyperspectral Imager
AIUS	Atmospheric Infrared Ultra-spectral Sounder
AUC	Area under the curve
CDnet	Cloud detection neural network
CNN	Convolutional neural network
DNN	Deep neural network
DPC	Directional Polarization Camera
EMI	Environmental Monitoring Instrument
FN	False negative
FP	False positive
GAN	Generative adversarial network
GF-1	GaoFen-1
GF-5	GaoFen-5
GMI	Greenhouse Gases Monitoring Instrument
HRG	High-resolution geometric
HS	Hyperspectral
Kappa	Kappa coefficient
MS	Multispectral
MSCFF	Multiscale convolutional feature fusion
OA	Overall accuracy
OLI	Operational Land Imager
PCANet	Principal component analysis network
PRS	Progressive refinement scheme
ROC	Receiver operating characteristic
SGD	Stochastic gradient descent

SL	Scene learning
ST	Structure tensor
SVM	Support vector machine
TP	True positive
TN	True negative
TIRS	Thermal Infrared Sensor
UNCD	unsupervised network for cloud detection
VAE	Variational autoencoder
VIMI	Visual and Infrared Multispectral Imager
WFV	Wide field of view

References

1. Lei, J.; Xie, W.; Yang, J.; Li, Y.; Chang, C.I. Spectral-spatial feature extraction for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8131–8143. [[CrossRef](#)]
2. Xie, W.; Shi, Y.; Li, Y.; Jia, X.; Lei, J. High-quality spectral-spatial reconstruction using saliency detection and deep feature enhancement. *Pattern Recognit.* **2019**, *88*, 139–152. [[CrossRef](#)]
3. Jiang, T.; Li, Y.; Xie, W.; Du, Q. Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2020**. to be published. [[CrossRef](#)]
4. Nasrabadi, N.M. Hyperspectral target detection: An overview of current and future challenges. *IEEE Signal Process. Mag.* **2014**, *31*, 34–44. [[CrossRef](#)]
5. Yuan, Y.; Ma, D.; Wang, Q. Hyperspectral anomaly detection by graph pixel selection. *IEEE Trans. Cybern.* **2016**, *46*, 3123–3134. [[CrossRef](#)]
6. Li, Y.; Xie, W.; Li, H. Hyperspectral image reconstruction by deep convolutional neural network for classification. *Pattern Recognit.* **2017**, *63*, 371–383. [[CrossRef](#)]
7. Zhang, Y.; Rossow, W.B.; Lacis, A.A.; Oinas, V.; Mishchenko, M.I. Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data. *J. Geophys. Res. Atmos.* **2004**, *109*, D19105. [[CrossRef](#)]
8. Xie, W.; Li, Y.; Zhou, W.; Zheng, Y. Efficient coarse-to-fine spectral rectification for hyperspectral image. *Neurocomputing* **2018**, *275*, 2490–2504. [[CrossRef](#)]
9. Fisher, A. Cloud and cloud-shadow detection in SPOT5 HRG imagery with automated morphological feature extraction. *Remote Sens.* **2014**, *6*, 776–800. [[CrossRef](#)]
10. Zhang, Q.; Xiao, C. Cloud detection of RGB color aerial photographs by progressive refinement scheme. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7264–7275. [[CrossRef](#)]
11. Li, Z.; Shen, H.; Li, H.; Xia, G.; Gamba, P.; Zhang, L. Multi-feature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery. *Remote Sens. Environ.* **2017**, *191*, 342–358. [[CrossRef](#)]
12. Zhong, B.; Chen, W.; Wu, S.; Hu, L.; Luo, X.; Liu, Q. A cloud detection method based on relationship between objects of cloud and cloud-shadow for Chinese moderate to high resolution satellite imagery. *IEEE J. Sel. Top. Appl. Earth Observat. Remote Sens.* **2017**, *10*, 4898–4908. [[CrossRef](#)]
13. Ishida, H.; Oishi, Y.; Morite, K.; Moriwaki, K.; Nakajima, T.Y. Development of a support vector machine based cloud detection method for MODIS with the adjustability to various conditions. *Remote Sens. Environ.* **2018**, *205*, 309–407. [[CrossRef](#)]
14. An, Z.; Shi, Z. Scene learning for cloud detection on remote-sensing images. *IEEE J. Sel. Top. Appl. Earth Observat. Remote Sens.* **2015**, *8*, 4206–4222. [[CrossRef](#)]
15. Li, P.; Dong, L.; Xiao, H.; Xu, M. A cloud image detection method based on SVM vector machine. *Neurocomputing* **2015**, *169*, 34–42. [[CrossRef](#)]
16. Kanungo, T.; Mount, D.M.; Netanyahu, N.S.; Piatko, C.D.; Silverman, R.; Wu, A.Y. An efficient K-means clustering algorithm: Analysis and implementation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 881–892. [[CrossRef](#)]
17. Xie, W.; Jia, X.; Li, Y.; Lei, J. Hyperspectral image super-resolution using deep feature matrix factorization. *IEEE Trans. Geosci. Remote Sens.* **2019**. to be published. [[CrossRef](#)]
18. Wu, H.; Prasad, S. Semi-supervised deep learning using pseudo labels for hyperspectral image classification. *IEEE Trans. Image Process.* **2018**, *27*, 1259–1270. [[CrossRef](#)]

19. Xie, W.; Li, L.; Hu, J.; Chen, D.Y. Trainable spectral difference learning with spatial starting for hyperspectral image denoising. *IEEE Trans. Image Process.* **2018**, *108*, 272–286. [[CrossRef](#)]
20. Ienco, D.; Pensa, R.G.; Meo, R. A semisupervised approach to the detection and characterization of outliers in categorical data. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 1017–1029. [[CrossRef](#)]
21. Shi, M.; Xie, F.; Zi, Y.; Yin, J. Cloud detection of remote sensing images by deep learning. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 701–704.
22. Goff, M.L.; Tourneret, J.Y.; Wendt, H.; Ortner, M.; Spigai, M. Deep learning for cloud detection. In Proceedings of the International Conference of Pattern Recognition Systems(ICPRS), Fort Worth, TX, USA, 23–28 July 2017; pp. 1–6.
23. Ozkan, S.; Efendioglu, M.; Demirpolat, C. Cloud detection from rgb color remote sensing images with deep pyramid networks. *arXiv* **2018**, arXiv:cs.CV/1801.08706.
24. Zi, Y.; Xie, F.; Jiang, Z. A cloud detection method for Landsat 8 images based on PCANet. *Remote Sens.* **2018**, *10*, 877. [[CrossRef](#)]
25. Shao, Z.; Pan, Y.; Diao, C.; Cai, J. Cloud detection in remote sensing images based on multiscale features-convolutional neural network. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4062–4076. [[CrossRef](#)]
26. Yang, J.; Guo, J.; Yue, H.; Liu, Z.; Hu, H.; Li, K. CDnet: CNN-based cloud detection for remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**. to be published. [[CrossRef](#)]
27. Li, Z.; Shen, H.; Cheng, Q.; Liu, Y.; You, S.; He, Z. Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors. *ISPRS J. Photogramm. Remote. Sens.* **2019**, *150*, 197–212. [[CrossRef](#)]
28. Zhang, Y.; Wen, F.; Gao, Z.; Ling, X. A coarse-to-fine framework for cloud removal in remote sensing image sequence. *IEEE Trans. Geosci. Remote Sens.* **2019**, to be published. [[CrossRef](#)]
29. Wen, F.; Zhang, Y.; Gao, Z.; Ling, X. Two-pass robust component analysis for cloud removal in satellite image sequence. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1090–1094. [[CrossRef](#)]
30. Lorenzi, L.; Melgani, F.; Mercier, G. Missing-area reconstruction in multispectral images under a compressive sensing perspective. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 3998–4008. [[CrossRef](#)]
31. Makhzani, A.; Shlens, J.; Jaitly, N.; Goodfellow, I.; Frey, B. Adversarial autoencoders. *arXiv* **2015**, arXiv:cs.LG/1511.05644.
32. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2013**, arXiv: cs.LG/1312.6114.
33. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
34. Nair, V.; Hinton, G.E. Rectified linear units improve restricted Boltzmann machines. In Proceedings of the International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010.
35. Xie, W.; Jiang, T.; Li, Y.; Jia, X.; Lei, J. Structure tensor and guided filtering-based algorithm for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4218–4230. [[CrossRef](#)]
36. Xie, W.; Lei, J.; Cui, Y.; Li, Y.; Du, Q. Hyperspectral pansharpening with deep priors. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, to be published. [[CrossRef](#)] [[PubMed](#)]
37. He, K.; Sun, J.; Tang, X. Guided image filtering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1397–1409. [[CrossRef](#)] [[PubMed](#)]
38. Hughes, M.J.; Hayes, D.J. Automated detection of cloud and cloud shadow in single-date Landsat imagery using neural networks and spatial post-processing. *Remote Sens.* **2014**, *6*, 4907–4926. [[CrossRef](#)]
39. Cao, V.L.; Nicolau, M.; McDermott, J. A hybrid autoencoder and density estimation model for anomaly detection. In Proceedings of the IEEE International Parallel Problem Solving from Nature, Edinburgh, Scotland, 17–21 September 2016; pp. 717–726.

40. Ferri, C.; Hernández-Orallo, J.; Flach, P. A coherent interpretation of AUC as a measure of aggregated classification performance. In Proceedings of the International Conference on Machine Learning, Fort Lauderdale, FL, USA, 11–13 April 2011; pp. 657–664.
41. Wang, Q.; Yuan, Z.; Du, Q.; Li, X. GETNET: A General End-to-End 2-D CNN Framework for Hyperspectral Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3–13. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).