



Article

Fusion Information Multi-View Classification Method for Remote Sensing Cloud Detection

Qi Hao ^{1,2}, Wenguang Zheng ^{1,2,*}  and Yingyuan Xiao ^{1,2,*} 

¹ Engineering Research Center of Learning-Based Intelligent System, Ministry of Education, Tianjin 300384, China; haoqi_77@outlook.com

² Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin University of Technology, Tianjin 300384, China

* Correspondence: wenguangz@tjut.edu.cn (W.Z.); yyxiao@tjut.edu.cn (Y.X.)

Abstract: In recent years, many studies have been carried out to detect clouds on remote sensing images. Due to the complex terrain, the variety of clouds, the density, and content of clouds are various, and the current model has difficulty accurately detecting the cloud in the image. In our strategy, a multi-view data training set based on super pixel is constructed. View A uses multi-level network to extract the boundary, texture, and deep abstract feature of super pixels. View B is the statistical feature of the three channels of the image. Privilege information View P contains the cloud content of super pixels and the tag status of adjacent super pixels. Finally, we propose a cloud detection method for remote sensing image classification based on multi-view support vector machine (SVM). The proposed method is tested on images of different terrain and cloud distribution in GF-1_WHU and Cloud-38 remote sensing datasets. Visual performance and quantitative analysis show that the method has excellent cloud detection performance.

Keywords: cloud detection; feature extraction network; fusion information; multi-view learning



Citation: Hao, Q.; Zheng, W.; Xiao, Y. Fusion Information Multi-View Classification Method for Remote Sensing Cloud Detection. *Appl. Sci.* **2022**, *12*, 7295. <https://doi.org/10.3390/app12147295>

Academic Editor: Byung-Gyu Kim

Received: 29 May 2022

Accepted: 18 July 2022

Published: 20 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Remote sensing images are widely used in land resource utilization, meteorological monitoring, and geology [1,2]; however, the monitoring data of sensors are often affected by clouds. Many remote sensing studies have been plagued by cloud occlusion, resulting in inaccurate observation results [3]. Therefore, it is necessary to accurately detect the cloud in remote sensing images.

Most cloud detection methods are highly dependent on available spectral bands, and use specific physical constraints to separate different categories according to spectral bands [4]. In particular, when relying on handmade features or experiences and searching for specific thresholds, these methods will have some problems in the segmentation of categories, and there may be some situations that cannot be segmented from the spectral threshold, such as deserts and super-high-brightness pixel areas [5]. The key of the threshold-based method is how to choose the best threshold to distinguish foreground cloud and background surface. Early fixed threshold methods failed to meet the increasing accuracy requirements. Therefore, more and more dynamic adaptive threshold methods are proposed for the difference between cloud features and surface features. Jedlovec et al. [6] used two images with different channels to incorporate the spatiotemporal change threshold into the cloud detection process. Zhang et al. [7] proposed an automatic cloud detection algorithm for remote sensing image observation statistics. They improved the global threshold method and gradually improved the detection results.

With the development of machine learning, the method based on Markov random field [8] and the widely used SVM [9–12] are becoming more and more popular in cloud detection. There are also some deep networks used as tools for cloud detection tasks. For example, Mohajerani et al. [13] proposed a hybrid full convolution network (FCN) and gradi-

ent recognition algorithm and applied FCN to the field of cloud detection. Manzo et al. [14] proposed a framework that combines convolutional neural networks, adapted to the cloud recognition task through a transfer learning approach, using voting rules. The cloud-net algorithm has better detection effect by redesigning the convolution block based on FCN [15], and usually uses the cloud network as the baseline of the deep learning cloud detection network. The multilevel feature fused segmentation network (MFFSNet) algorithm uses a pyramid pooling module to aggregate feature information at different scales to improve the utilization of local and global features of clouds in images [16]. In the detection of cloud area in cloud-containing remote sensing images, judging from the pixel level or dividing the image into rectangular blocks, super-pixel-level judgment is more effective than pixel-level judgment [17,18]. Liu et al. [19] proposed a cloud judgment method combined with several statistical characteristics of super pixels and conducted experiments. The above studies show that judging cloud or non-cloud with super pixel as the basic unit can obtain excellent cloud and non-cloud segmentation results of remote sensing images.

Many methods determine labels according to the number of cloud-containing pixels or the proportion of cloud-containing pixels in the super pixel, and cannot make good use of the label data of adjacent pixels and their own cloud proportion data. The information between super pixels cannot be well combined. The current methods are difficult to define the super pixel as a certain category, or the super pixel cannot fully contain the cloud, and the corresponding cloud-containing state information cannot be reasonably used. At the same time, most remote sensing images now use multiple sensors to collect data, but there will be only three RGB visible light channels of data. For this image, a segmentation method with high accuracy is also needed.

In order to solve the above problems, a multi-layer network structure is used to extract the texture, boundary, and high-level abstract information of super pixels, and the statistical features of the basic three-channel color data are utilized. The data of different extracted views are essentially relevant because they provide complementary information for the same data in semantics. Many methods show that the combination of multiple views learning is superior to the simple method of using a connected view or learning from each view alone [20,21]. The structural relationship between pixels and the cloud state of the pixel itself belong to a privilege information. In order to unify the use of each view, a multi-view classification method based on information fusion is proposed. In the training stage, three feature views are used. The model introducing the privileged information mechanism can accurately determine the super pixel category by using the extracted two feature views (except the privilege information features). Overall, our main contributions are as follows:

1. A feature extraction network at the super pixel level is constructed to extract the fast features of super pixels in cloud-containing remote sensing images at multi-scales. The cloud content in super pixels and the cloud-containing marker state information of adjacent super pixels are effectively utilized.
2. A multi-view support vector machine cloud detection classifier based on fusion information is constructed and the solving algorithm based on quadratic convex optimization is given.
3. We provide a multi-view classification dataset based on remote sensing cloud super pixels. The new model is used to classify the super pixels and synthesize the cloud mask bipartite graph. Experiments are carried out in images with different cloud contents to verify the effectiveness of the new method.

The rest of this article is organized as follows. Section 2 introduces the research status of multi-view learning and an advanced multi-view classification method. In Section 3, we introduce the framework of the proposed method in detail, including the multi-view feature extraction method and fusion information multi-view classification model. Section 4 shows the experimental results on two remote sensing datasets and discusses the results of the experiment. Finally, Section 5 summarizes the research work.

2. Related Work

2.1. Multi-View Learning

Multi-view learning algorithms can be divided into co-training, multi-kernel learning, and subspace learning [21,22]. The co-training algorithm iteratively maximizes mutual conventions on two distinct views to ensure consistency on the same validation data, such as multi-view collaborative clustering algorithm and research [23] using multiple collaborative training for document classification [24]. The multi-kernel learning (MKL) algorithm uses the kernels corresponding to different views and combines them linearly or nonlinearly to improve performance. For example, the support kernel machine (SKM) model introduced in [25], and the sequential minimum optimization (SMO) algorithm is developed to solve it. Multi-kernel framework with nonparallel support vector machine (MKNPSVM) works by integrating non-parallel support vector machines into the MKL framework to learn the optimal kernel combination [26]. The subspace learning algorithm assumes that the input view comes from a potential subspace and aims to realize the potential subspace shared by multiple views, such as SVM-2K combined with two support vector machines and kernel canonical correlation analysis (KCCA) [27], SVM classification method with coupling privileged kernel method [28], etc.

2.2. Coupling Privileged Kernel Method for Multi-View Learning

Tang et al. [28] proposed a simple and effective multi-view learning coupling privileged kernel method (MCPK). MCPK integrates consensus and complementarity principles into a unified framework. In particular, consistency is captured by coupling terms between two views. Because the multi-view data collected from different domains can complement each other, a different feature view can receive explicit privilege information from its view. MCPK can be built as Equation (1).

$$\begin{aligned} \min_{w_A, w_B, \zeta_A, \zeta_B} \quad & \frac{1}{2} (\|w_A\|^2 + \gamma \|w_B\|^2) + C_A \sum_{i=1}^l \zeta_i^A + C_B \sum_{i=1}^l \zeta_i^B + C \sum_{i=1}^l \zeta_i^A \zeta_i^B, \\ \text{s.t.} \quad & y_i(w_A \cdot \phi_A(x_i^A)) \geq 1 - \zeta_i^A, \\ & y_i(w_B \cdot \phi_B(x_i^B)) \geq 1 - \zeta_i^B, \\ & \zeta_i^A \geq y_i(w_B \cdot \phi_B(x_i^B)), \\ & \zeta_i^B \geq y_i(w_A \cdot \phi_A(x_i^A)), \\ & \zeta_i^A \geq 0, \zeta_i^B \geq 0, i = 1, \dots, l. \end{aligned} \quad (1)$$

where w_A and w_B are the weight vectors of view A and view B, respectively, and the two views are weighed by the non-negative trade-off parameter γ . As slack variables, ζ_i^A and ζ_i^B are constrained by the correction functions determined by the two views. The coupling term $C \sum_{i=1}^l \zeta_i^A \zeta_i^B$ makes the product of error variables of the two views as small as possible. When classifiers constructed from different views are more consistent, errors from both views are small, resulting in smaller couplings. Therefore, its consistency can be fully ensured. C is a non-negative coupling parameter that controls the influence of the coupling term. C_A and C_B are non-negative penalty parameters.

3. Proposed Method

3.1. Multi-View Feature Extraction

View A: Texture, boundary, and other features are extracted from the super pixel circumscribed rectangular region through a multi-layer joint convolutional neural network (CNN). Cloud area has similar shape to super pixel boundary, which is quite different from grassland area and water area boundary. In this paper, we use the inherent multi-scale pyramid level of a deep convolution network to develop a top architecture with horizontal connection, which is used to construct high-level feature maps at all scales. Features can be easily extracted by multi-scale feature extraction structure in the super

pixel circumscribed rectangle, and texture, boundary, gradient, and other information can be obtained from multiple horizons. The feature extraction layer Conv3 in the deepest part will extract more abstract features. View A feature extraction is realized by cloud super pixel feature extraction network (CSPFE-Net); CSPFE-Net structure is shown in Figure 1. In the network structure, the super pixel external rectangular region has the input of $3 \times 16 \times 16$ images. Conv refers to the convolution layer. Relu increases the nonlinear ability. Pooling downsampling is used to further extract the image. LRN is normalized to constrain the data within a certain range. Concat refers to stitching vectors. The features extracted from each feature extraction layer are weighted splicing, and are jointly output as the feature vector through the full connection layer. The feature vector with the super pixel output structure of 1×64 under the fine scale is sufficient to complete the feature representation.

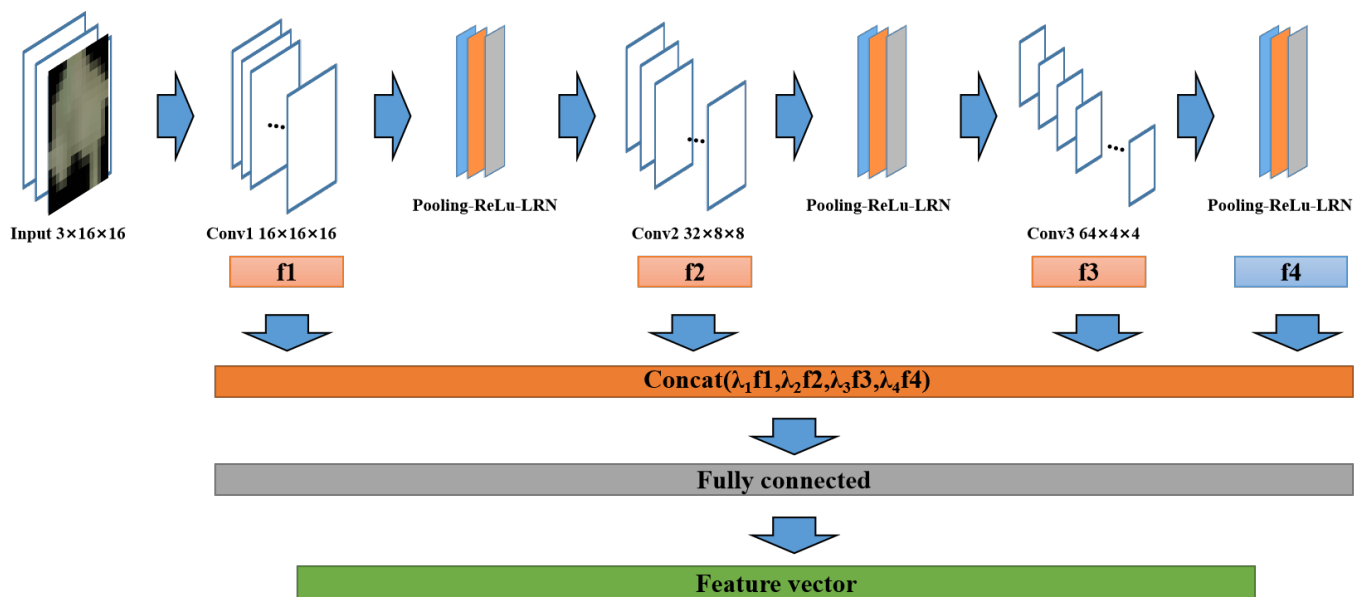


Figure 1. CSPFE-NET structure.

View B: The color statistical characteristics extracted by the super pixel itself, excluding the black edge region of the circumscribed rectangle. Specifically, we calculate the mean, variance, maximum, minimum, and median of the data describing the color in each channel. SP_i represents three-channel data for the i -th super pixel, and RGB color statistics can be calculated by Equation (3). The feature vector format of view b is 1×15 .

$$getCF(x) = [mean(x), std(x), max(x), min(x), median(x)] \tag{2}$$

$$RGBfeature = [getCF(SP_i(R)), getCF(SP_i(G)), getCF(SP_i(B))] \tag{3}$$

where $getCF(x)$ extracts and splices the statistical features of the data of one channel x ; $SP_i(R)$, $SP_i(G)$, and $SP_i(B)$ represent three visible light channels: red, green, and blue.

View P: View P (privileged information view) is the feature of privileged information. The correction space guided by privilege information can correct the classification plane of multi-view, which can make the performance of multi-view classifier better. The feature of cloud-containing super pixel privileged information includes two parts. The first part is the specific cloud content in the super pixel. The second part is the label of the fast adjacent block of the super pixel. Generally, two super pixel blocks with the nearest center distance are selected, and this part is the feature vector of 1×3 .

3.2. Fusion Information Multi-View SVM Classification Method

Fusion information multi-view SVM classification method (FIMV-SVM) is based on MCPK, using feature extraction network and three-channel statistical features as view A and view B data, using privileged information view P to correct the separation hyperplane of view A and view B. The optimized structure constructed is as per Equation (4).

$$\begin{aligned}
 \min_{w_A, w_B, w_P, \zeta_A, \zeta_B, \zeta_P} & \frac{1}{2} (\|w_A\|^2 + \gamma \|w_B\|^2 + \gamma_P \|w_P\|^2) \\
 & + C_A \sum_{i=1}^l \zeta_i^A + C_B \sum_{i=1}^l \zeta_i^B + C_P \sum_{i=1}^l \zeta_i^P + C \sum_{i=1}^l \zeta_i^A \zeta_i^B, \\
 \text{s.t.} & \quad y_i(w_A \cdot \phi_A(x_i^A)) \geq 1 - \zeta_i^A, \\
 & \quad y_i(w_B \cdot \phi_B(x_i^B)) \geq 1 - \zeta_i^B, \\
 & \quad y_i(w_P \cdot \phi_P(x_i^P)) \geq 1 - \zeta_i^P, \\
 & \quad \zeta_i^A \geq y_i(w_B \cdot \phi_B(x_i^B)), \\
 & \quad \zeta_i^B \geq y_i(w_A \cdot \phi_A(x_i^A)), \\
 & \quad \zeta_i^A \geq y_i(w_P \cdot \phi_P(x_i^P)), \\
 & \quad \zeta_i^B \geq y_i(w_P \cdot \phi_P(x_i^P)), \\
 & \quad \zeta_i^A \geq 0, \zeta_i^B \geq 0, \zeta_i^P \geq 0, i = 1, \dots, l.
 \end{aligned} \tag{4}$$

In optimization problem Equation (4), $\|w_A\|^2$ and $\|w_B\|^2$ are regularization terms of view A and view B, $\|w_P\|^2$ is a regularization term for privileged information view P, respectively. C_A , C_B , and C_P are non-negative penalty parameters, $\zeta_i^A = [\zeta_1^A, \dots, \zeta_l^A]$, $\zeta_i^B = [\zeta_1^B, \dots, \zeta_l^B]$, and $\zeta_i^P = [\zeta_1^P, \dots, \zeta_l^P]$ are non-negative slack parameters. γ is the balance parameter to balance the weight of view A and view B. γ_P is used to weigh the influence of privileged information view P; $\phi_A(x_i^A)$, $\phi_B(x_i^B)$, and $\phi_P(x_i^P)$ represent the mapping of views data. In the constraint, $y_i(w_P \cdot \phi_P(x_i^P)) \geq 1 - \zeta_i^P$ denote that the slack variables are constrained by the view P, $\zeta_i^A \geq y_i(w_P \cdot \phi_P(x_i^P))$, and $\zeta_i^B \geq y_i(w_P \cdot \phi_P(x_i^P))$ correcting constraints on the classification hyperplane through the correction hyperplane formed by privileged information. For the solution of the above optimization problem, it can be transformed into a Lagrangian dual problem and solved by solving quadratic convex optimization. The Lagrangian function is Equation (5).

$$\begin{aligned}
 L = & \frac{1}{2} (\|w_A\|^2 + \gamma \|w_B\|^2 + \gamma_P \|w_P\|^2) + C_A \sum_{i=1}^l \zeta_i^A + C_B \sum_{i=1}^l \zeta_i^B + C_P \sum_{i=1}^l \zeta_i^P + C \sum_{i=1}^l \zeta_i^A \zeta_i^B \\
 & + \sum_{i=1}^l \alpha_i^A (1 - \zeta_i^A - y_i(w_A \cdot \phi_A(x_i^A))) + \sum_{i=1}^l \alpha_i^B (1 - \zeta_i^B - y_i(w_B \cdot \phi_B(x_i^B))) \\
 & + \sum_{i=1}^l \alpha_i^P (1 - \zeta_i^P - y_i(w_P \cdot \phi_P(x_i^P))) \\
 & + \sum_{i=1}^l \lambda_i^A (y_i(w_B \cdot \phi_B(x_i^B)) - \zeta_i^A) + \sum_{i=1}^l \lambda_i^B (y_i(w_A \cdot \phi_A(x_i^A)) - \zeta_i^B) \\
 & + \sum_{i=1}^l \mu_i^A (y_i(w_A \cdot \phi_A(x_i^A)) - \zeta_i^P) + \sum_{i=1}^l \mu_i^B (y_i(w_B \cdot \phi_B(x_i^B)) - \zeta_i^P) \\
 & - \sum_{i=1}^l \beta_i^A \zeta_i^A - \sum_{i=1}^l \beta_i^B \zeta_i^B.
 \end{aligned} \tag{5}$$

Therefore, the dual programming of Equation (4) can be obtained by finding the partial derivatives of the optimization parameters.

$$\begin{aligned}
 \min \quad & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l ((\alpha_i^A - \lambda_i^B) y_i K_A(x_i^A, x_j^A) (\alpha_j^A - \lambda_j^B) y_j \\
 & + \frac{1}{\gamma} (\alpha_i^B - \lambda_i^A) y_i K_B(x_i^B, x_j^B) (\alpha_j^B - \lambda_j^A) y_j \\
 & + \frac{1}{\gamma_P} (\alpha_i^P y_i K_P(x_i^P, x_j^P) \alpha_j^P y_j)) - \sum_{i=1}^l (\alpha_i^A + \alpha_i^B + \alpha_i^P) \\
 & + \frac{1}{C} \sum_{i=1}^l (\alpha_i^A + \lambda_i^A + \mu_i^A + \beta_i^A - C_A) (\alpha_i^B + \lambda_i^B + \mu_i^B + \beta_i^B - C_B), \\
 \text{s.t.} \quad & \alpha_i^A, \alpha_i^B, \lambda_i^A, \lambda_i^B, \mu_i^A, \mu_i^B, \beta_i^A, \beta_i^B \geq 0.
 \end{aligned} \tag{6}$$

where $K_A(x_i^A, x_j^A), K_B(x_i^B, x_j^B), K_P(x_i^P, x_j^P)$ represents the kernel mapping mode of view A, view B, and view P feature data, respectively. The optimization problem Equation (6) is a quadratic convex programming problem, which can be solved by the quadratic convex programming method. Solving the optimal parameters $\alpha_i^{A*}, \alpha_i^{B*}, \beta_i^{A*}, \beta_i^{B*}, \lambda_i^{A*}, \lambda_i^{B*}, \mu_i^{A*}, \mu_i^{B*}$, we use the Karush–Kuhn–Tucker(KKT) [29] condition to obtain the optimal result w_A^* and w_B^* . The calculation results are shown in Equations (7) and (8).

$$w_A^* = \sum_{i=1}^l (\alpha_i^{A*} y_i - \lambda_i^{B*} y_i) \phi_A(x_i^A), \tag{7}$$

$$w_B^* = \sum_{i=1}^l (\alpha_i^{B*} y_i - \lambda_i^{A*} y_i) \phi_B(x_i^B). \tag{8}$$

After obtaining the optimal w_A^* and w_B^* , we use the following formula to predict the labels of the new samples (x^A, x^B) from view A and view B. The final predictor of multi-views can be constructed as the average prediction factor of each view and is shown in (9).

$$f = \text{sign}(\frac{1}{2} f_A(x^A) + \frac{1}{2} f_B(x^B)) = \text{sign}(\frac{1}{2} w_A^{*\top} \phi_A(x^A) + \frac{1}{2} w_B^{*\top} \phi_B(x^B)). \tag{9}$$

where f_A represents the decision function of view A and f_B represents the decision function of view B. The FIMV-SVM solution process is clearly represented in Algorithm 1.

3.3. Cloud Detection Model Training and Application Process

Combined with the above contents, the cloud detection model method can be summarized. The specific process is shown in Figure 2. Notably, privileged information does not appear in the application phase.

In the model training phase, the following steps are taken: (1) The super pixels come from simple linear iterative cluster (SLIC) method [30], which is used for super pixel division in the original image dataset to construct the dataset with super pixel blocks as the classification objects; (2) three different feature extraction methods are used to extract the feature of super pixels one by one, and the feature vectors of view A, view B, and view P are composed; (3) FIMV-SVM classifier is trained by using the extracted numerical feature dataset and labels.

Algorithm 1 QP Algorithm for FIMV-SVM

Require: $S = \{x_i^A, x_i^B, x_i^P, y_i\}_{i=1}^l = \{(x_i^A; 1), (x_i^B; 1), (x_i^P, 1)\}_{i=1}^l, y_i \in \{+1, -1\}$

Ensure: Decision functions: $f = \text{sign}(\frac{1}{2}w_A^{*\top} \phi_A(x^A) + \frac{1}{2}w_B^{*\top} \phi_B(x^B))$

- 1: Grid method generates parameter sets: $\text{paraSet} = \{C_A, C_B, C_P, C, \gamma, \gamma_P\}_{i=1}^n$.
- 2: **for** each $i \in [1, n]$ **do**
- 3: Set parameters $C_A, C_B, C_P, C, \gamma, \gamma_P = \text{paraSet}[i, :]$.
- 4: Set kernels function of view A, view B and view P: $K_A(x_i^A, x_j^A), K_B(x_i^B, x_j^B), K_P(x_i^P, x_j^P)$.
- 5: Create and solve quadratic programming problem and Solving quadratic programming and retaining optimal parameters $\alpha_i^{A*}, \alpha_i^{B*}, \beta_i^{A*}, \beta_i^{B*}, \lambda_i^{A*}, \lambda_i^{B*}, \mu_i^{A*}, \mu_i^{B*}$.
- 6: Get the optimal weight w_A^* and w_B^* by substituting formula:

$$w_A^* = \sum_{i=1}^l (\alpha_i^{A*} y_i - \lambda_i^{B*} y_i) \phi_A(x_i^A), w_B^* = \sum_{i=1}^l (\alpha_i^{B*} y_i - \lambda_i^{A*} y_i) \phi_B(x_i^B).$$

- 7: Decision function is solved by parameters w_A and w_B .
- 8: The accuracy of statistical validation set was calculated as $\text{ACC}[i], W_A[i] = w_A, W_B[i] = w_B$.
- 9: **end for**
- 10: The final w_A^*, w_B^* :

$$w_A^* = W_A[\text{find}(\text{acc} == \max(\text{ACC}))], w_B^* = W_B[\text{find}(\text{acc} == \max(\text{ACC}))].$$

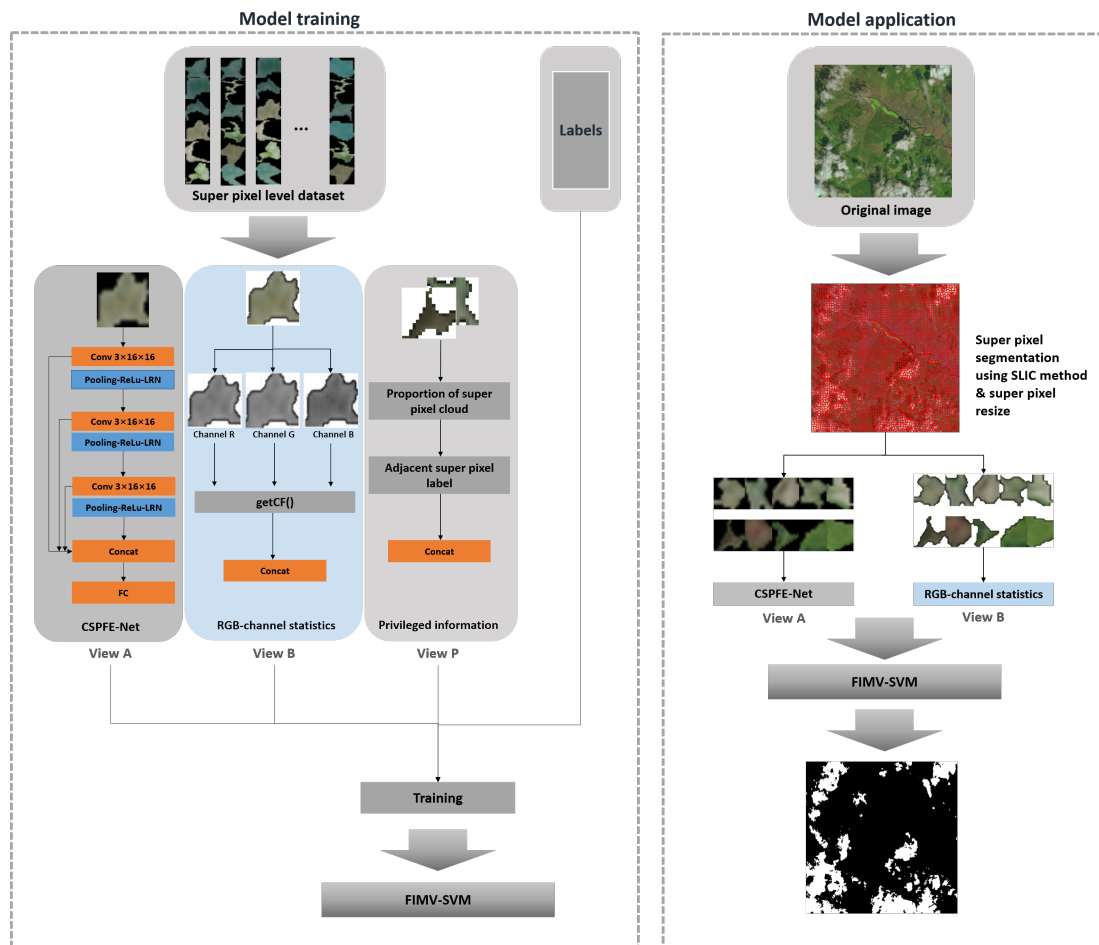


Figure 2. FIMV-SVM cloud detection model training and application process framework.

In the model application phase, the following steps are taken: (1) The image to be processed is divided by SLIC super pixel segmentation method, and then the super pixel block is resized; (2) the features of view A and view B are extracted from the super pixel set one by one to form their feature vectors; (3) super pixel classification object view A, view B, two view feature vectors, are input through the FIMV-SVM decision function to obtain the corresponding classification results; (4) the final segmentation cloud mask result is formed by combining the super pixel classification results.

4. Experiment

Our experiment was carried out on a personal computer with i7-6500 CPU and 16 GB RAM. The environment used in the experiment is Python 3.7 combined with pyTorch framework (version 1.11.0), and the cvxopt tool [31] on MATLAB 2016b is used to solve the convex optimization problem.

For the experiment in GF-1_WHU remote sensing images [32] and datasets of Cloud-38 [33], the detailed description is shown in Table 1. The original high-pixel image is divided into sub-images of $400 \times 400 \times 3$, and only their visible light channels are used.

Table 1. Public dataset description.

Name	Number of Images	Resource Acquisition	Remarks
GF-1_WHU	108	http://sendimage.whu.edu.cn/en/mfc-validation-data/ (accessed on 15 July 2022)	GF-1_WHU includes 108 GF-1 wide field of view (WFV) level-2A scenes and its reference cloud and cloud shadow masks.
Cloud-38	38	https://www.kaggle.com/datasets/sorour/38cloud-cloud-segmentation-in-satellite-images/download (accessed on 15 July 2022)	There are four spectral channels, namely, red (band 4), green (band 3), blue (band 2), and near-infrared (band 5).

In the process of setting the level of segmentation parameters of super pixel segmentation, considering that the super pixel object needs to have certain information inclusion ability, in the image of $400 \times 400 \times 3$, the classification levels are 1000, 1600, 2000, 2400, and 3000, which can be divided into about 160, 100, 80, 67, and 54 pixels in the super pixel block. In order to explore the optimal parameters of the super pixel partition level, we construct a parameter optimization dataset from the original datasets with the size of 1000. A total of 80% of the parameter optimization dataset is used for training, and the rest is used for testing. The average accuracy in the test process is used as the evaluation index. The experimental results are shown in Figure 3. The optimal result is obtained when the super pixel level is 2000, which shows that the number of super pixels in the super pixel is about 80, which is suitable for the smallest unit of the cloud recognition task. Therefore, for the super pixel segmentation method using convenient and efficient SLIC method, the segmentation level is 2000, that is, an image contains about 2000 super pixels, and each super pixel contains about 80 pixels. By resize operation, the number of rectangular pixels is about 256 pixels (16×16). The resize process uses the `imresize` function of MATLAB software, and `imresize` uses bicubic interpolation by default. The super pixel blocks with cloud content exceeding 45% are automatically labeled as cloud super pixels, and the others are labeled as non-cloud super pixels.

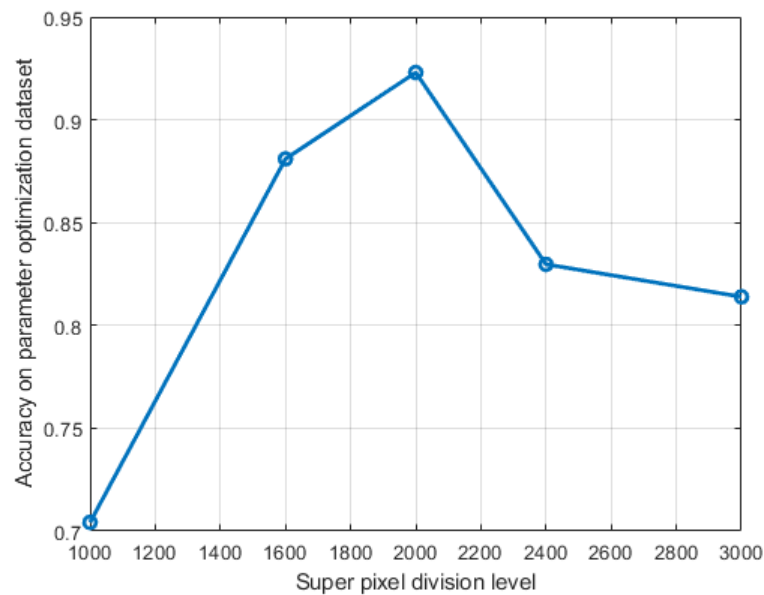


Figure 3. Performance of different super pixel classification levels on parameter optimization dataset.

4.1. Parameter Setting

In the experiment, we set the following parameters. The values of λ_1 , λ_2 , λ_3 , and λ_4 in CSPFE-Net of view A are 0.3, 0.3, 0.2, and 0.2. FIMV-SVM parameters are determined by grid search method and five-fold cross validation during training. The determination of parameters is based on the highest accuracy. C , C_A , C_B , and C_P select their values from the set $[10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3]$. The value set of γ and γ_P is $[0.2, 0.4, 0.6, 0.8, 1]$. Gaussian radial basis function (RBF) kernel function $K(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{2\sigma})$ is the kernel function of SVM method, σ for the Gaussian RBF kernel function is selected from $[10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3]$. For the association between parameter tuning and training, see Algorithm 1 above.

4.2. Visual Performance

We compare the proposed method with several advanced cloud detection methods. Cloud-Net is a cloud detection method based on deep learning and is widely used as the baseline of cloud detection experiments. Hierarchical fusion convolutional neural network (HFCNN) [19] is a network detection method with multi-level feature extraction, and HFCNN is a super-pixel-level judgment method. Furthermore, we added the SVM method and the MCPK method. In the SVM method, the extracted view A data and view B data are used as one-dimensional features for training. In MCPK, the data of view A and view B are used. Through these two experiments, we can explore the effectiveness of multi-view learning and privileged information addition.

Typical cloud-containing image blocks are selected from GF-1_WHU and Cloud-38 to display the detection results. These images have a variety of cloud coverage and backgrounds. The comparison results with the proposed method are shown in Figures 4 and 5. In the figures, from left to right, are original image, ground truth, Cloud-net, HFCNN, SVM, MCPK, and our method. From the visual results, on the whole, the proposed method is closer to ground truth. There are several aspects worth noting. In Figure 4, the test object in the third line is an image with highlighted features. Cloud-Net has a certain misjudgment, and SVM method judges a large area into a cloud. Compared with HFCNN, the proposed method is more rigorous for thin cloud detection. SVM methods often lack large-scale cloud areas, and MCPK will misjudge to a certain extent in continuous clouds.

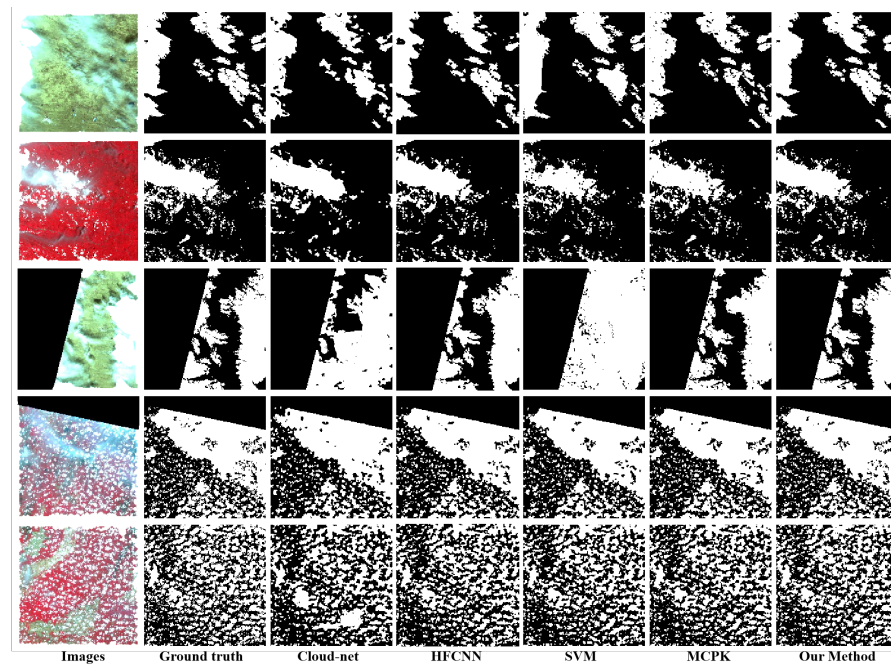


Figure 4. The performance of several methods on GF-1_WHU dataset [32] with different cloud distribution styles.

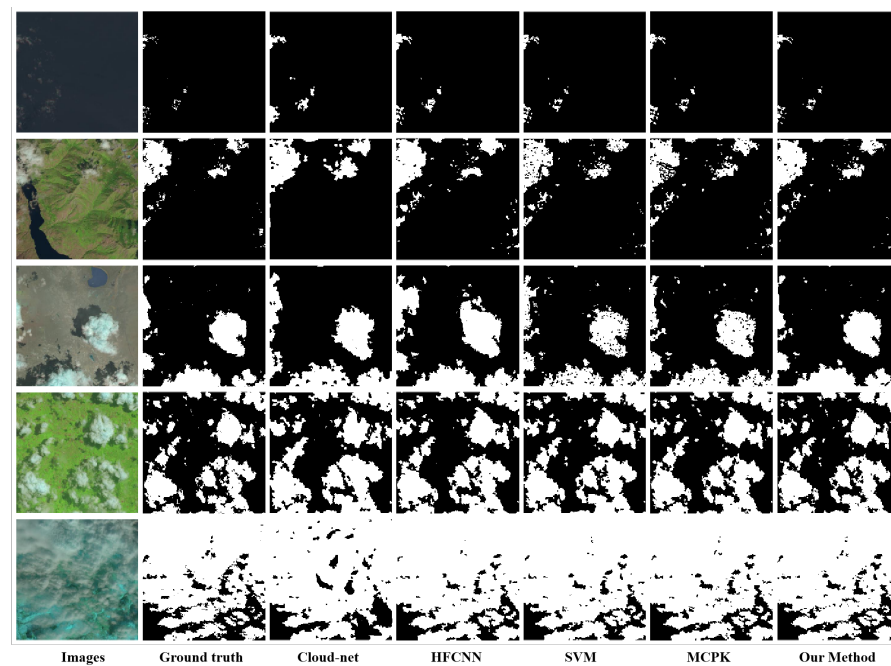


Figure 5. The performance of several methods on Cloud-38 dataset [33] with different cloud distribution styles.

4.3. Quantitative Analysis

In the description of the overall results, the Jaccard index is used to describe the similarity between the predicted mask and the real mask, which is widely used in the performance evaluation of cloud detection tasks. Precision is the ratio that predicts the number of true values to the number of cloud tags in cloud data. Recall represents how many clouds can be predicted in all marked cloud data. Specificity index is used to measure the integrity of error prediction, and the overall accuracy index is used to represent the accuracy of the cloud/non-cloud binary classification. F1-score considers the relationship

between precision and recall. The calculation method of each evaluation index is shown as Equations (10)–(15).

We divide each data image in GF-1_WHU and Cloud-38 into $400 \times 400 \times 3$ specifications, and randomly select 80% of the dataset as the training set image, which is the test set image. Specifically, in the GF-1_WHU dataset, there are 4246 images, including 3369 training images and 850 test images. In the cloud-38 dataset, there are 15,200 images, including 12,160 training images and 3040 test images. Each index data is shown as (mean \pm variance).

$$JaccardIndex = \frac{TP}{TP + FN + FP} \tag{10}$$

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

$$Specificity = \frac{TN}{TN + FP} \tag{13}$$

$$OverallAccuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{15}$$

where *TP* represents the number of positive samples with positive judgment results, *TN* represents the number of negative samples with negative judgment results, *FP* represents the number of negative samples with positive judgment results, and *FN* represents the number of positive samples with negative judgment results.

As shown in Tables 2 and 3, our method achieves high scores in both groups of test results. The Cloud-net [15] method needs a certain number of training sets to annotate data, and cannot use the location information of pixel blocks. HFCNN also uses super pixels as the basic unit of cloud detection, but it uses only one feature extraction method, and cannot use the privilege information marked by super pixels, so it performs poorly in some confusing areas. The basis of its use is $32 \times 32 \times 3$ and the segmentation effect may be rougher. By comparing the SVM, MCPK, and our methods, it is proved that the multi-view structure has better performance than the method of directly stitching combined features. At the same time, thanks to the feature extraction of super pixel multi-view and the utilization of privilege information, our method is more accurate for the recognition of the cloud area. The experimental results show that no matter whether the type of substrate in the background is more or less, our method can obtain excellent results, and the results of each index prove the feasibility and effectiveness of the proposed remote cloud sensing detection method.

Table 2. Performance on GF-1_WHU test dataset (mean \pm variance (%)).

Method	Jaccard Index	Pression	Recall	Specificity	Overall Accuracy	F1-Score
Cloud-Net [15]	81.93 \pm 7.22	87.03 \pm 5.48	93.49 \pm 4.36	82.86 \pm 4.34	88.46 \pm 3.68	91.21 \pm 4.46
HFCNN [19]	87.94 \pm 5.12	91.87 \pm 4.05	96.02 \pm 5.13	91.28 \pm 3.74	92.79 \pm 5.04	94.36 \pm 4.76
SVM	67.12 \pm 3.77	71.36 \pm 5.14	92.13 \pm 5.49	72.56 \pm 4.87	78.85 \pm 3.24	79.62 \pm 3.64
MCPK	85.32 \pm 4.19	87.93 \pm 5.33	96.07 \pm 3.96	85.77 \pm 5.08	90.76 \pm 4.39	92.43 \pm 3.13
Our method	91.69 \pm 1.34	94.67 \pm 2.01	97.38 \pm 1.65	93.04 \pm 2.12	96.31 \pm 1.46	96.03 \pm 1.72

Table 3. Performance on Cloud-38 test dataset (mean \pm variance (%)).

Method	Jaccard Index	Precision	Recall	Specificity	Overall Accuracy	F1-Score
Cloud-Net [15]	87.13 \pm 5.14	91.22 \pm 4.37	95.94 \pm 3.89	89.03 \pm 4.34	93.14 \pm 5.36	92.87 \pm 4.68
HFCNN [19]	92.04 \pm 3.69	92.95 \pm 3.98	98.25 \pm 5.01	92.24 \pm 4.36	95.86 \pm 5.23	94.21 \pm 3.67
SVM	79.02 \pm 4.37	82.83 \pm 5.04	92.87 \pm 4.76	82.78 \pm 3.85	86.74 \pm 4.19	87.86 \pm 3.92
MCPK	87.45 \pm 3.94	92.32 \pm 4.18	96.83 \pm 5.03	88.34 \pm 4.76	92.89 \pm 5.14	94.03 \pm 3.87
Our method	94.87 \pm 1.24	95.28 \pm 1.39	98.74 \pm 2.07	93.48 \pm 2.11	97.28 \pm 1.47	97.37 \pm 1.86

5. Conclusions

This paper mainly studies the three-channel remote sensing cloud detection method based on the fusion of multi-view information at the super pixel level. Firstly, the segmented super pixels are used to establish a super pixel remote sensing image database. Secondly, a variety of feature extraction mechanisms are used to extract three view features of super pixel blocks containing privileged information views. Finally, an SVM classifier that can utilize privileged information features is constructed, and a solution strategy based on quadratic convex optimization is proposed. The classifier is used to judge the super pixel category organization one by one to generate the cloud mask. Experiments are carried out on GF1_WHU and Cloud-38 datasets with different cloud content data. From the results of qualitative and quantitative analysis, we can see that the proposed method has good performance, and it also has good detection effect in scenarios with large differences in cloud distribution and cloud content. In the future, we consider improving the model by using transfer learning technology to make the model quickly adapt to the cloud recognition of multi-style remote sensing images. The algorithm proposed in this paper is based on the improvement of SVM binary classifiers. It is necessary to study a new strategy for multi-classification tasks such as accurate cloud classification, which is also a direction worthy of study in the future.

Author Contributions: Conceptualization, Q.H. and Y.X.; methodology, Q.H. and W.Z.; software, Q.H.; validation, Q.H., Y.X. and W.Z.; writing—original draft preparation, Q.H.; writing—review and editing, Q.H., Y.X. and W.Z.; visualization, Q.H.; supervision, Y.X.; project administration, W.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Tianjin “Project + Team” Key Training Project grant number XC202022 and Tianjin Research Innovation Project for Postgraduate Students grant number 2021YJSS095.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Abbreviations

The following is a description of the abbreviations used in this paper.

SVM	Support vector machine
MKL	Multi-kernel learning
SMO	Sequential minimum optimization
MKNPSVM	Multi-kernel framework with nonparallel support vector machine
KCCA	Kernel canonical correlation analysis
FCN	Full convolution networks
MCPK	Multi-view learning coupling privileged kernel method
CNN	Convolutional neural network
CSPFE-Net	Cloud super pixel feature extraction network

FIMV-SVM	Fusion information multi-view SVM
KKT	Karush–Kuhn–Tucke
SLIC	Simple linear iterative cluster
HFCNN	Hierarchical fusion convolutional neural network
RBF	Radial basis function

References

- Sola, I.; Álvarez-Mozos, J.; González-Audicana, M. Inter-Comparison of Atmospheric Correction Methods on Sentinel-2 Images Applied to Croplands. In Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2018, Valencia, Spain, 22–27 July 2018; IEEE: Piscataway Township, NJ, USA, 2018; pp. 5940–5943.
- Vermote, E.F.; Saleous, N.; Justice, C.O. Atmospheric correction of MODIS data in the visible to middle infrared: first results. *Remote Sens. Environ.* **2002**, *83*, 97–111. [[CrossRef](#)]
- Irish, R.R.; Barker, J.L.; Goward, S.N.; Arvidson, T. Characterization of the Landsat-7 ETM+ automated cloud-cover assessment (ACCA) algorithm. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 1179–1188. [[CrossRef](#)]
- Zhu, Z.; Woodcock, C.E. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* **2012**, *118*, 83–94. [[CrossRef](#)]
- Bley, S.; Deneke, H. A threshold-based cloud mask for the high-resolution visible channel of Meteosat Second Generation SEVIRI. *Atmos. Meas. Tech.* **2013**, *6*, 2713–2723. [[CrossRef](#)]
- Jedlovec, G.J.; Haines, S.L.; LaFontaine, F.J. Spatial and temporal varying thresholds for cloud detection in GOES imagery. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1705–1717. [[CrossRef](#)]
- 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](#)]
- Li, P.; Dong, L.; Xiao, H.; Xu, M. A cloud image detection method based on SVM vector machine. *Neurocomputing* **2015**, *169*, 34–42. [[CrossRef](#)]
- Le Hégarat-Masclé, S.; André, C. Use of Markov random fields for automatic cloud/shadow detection on high resolution optical images. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 351–366. [[CrossRef](#)]
- Shao, Z.; Deng, J.; Wang, L.; Fan, Y.; Sumari, N.S.; Cheng, Q. Fuzzy autoencode based cloud detection for remote sensing imagery. *Remote Sens.* **2017**, *9*, 311. [[CrossRef](#)]
- Ishida, H.; Oishi, Y.; Morita, 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*, 390–407. [[CrossRef](#)]
- Yuan, Y.; Hu, X. Bag-of-words and object-based classification for cloud extraction from satellite imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 4197–4205. [[CrossRef](#)]
- Mohajerani, S.; Krammer, T.A.; Saeedi, P. Cloud detection algorithm for remote sensing images using fully convolutional neural networks. *arXiv* **2018**, arXiv:1810.05782.
- Manzo, M.; Pellino, S. Voting in transfer learning system for ground-based cloud classification. *Mach. Learn. Knowl. Extr.* **2021**, *3*, 542–553. [[CrossRef](#)]
- Mohajerani, S.; Saeedi, P. Cloud-Net: An end-to-end cloud detection algorithm for Landsat 8 imagery. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; IEEE: Piscataway Township, NJ, USA, 2019, pp. 1029–1032.
- Yan, Z.; Yan, M.; Sun, H.; Fu, K.; Hong, J.; Sun, J.; Zhang, Y.; Sun, X. Cloud and cloud shadow detection using multilevel feature fused segmentation network. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1600–1604. [[CrossRef](#)]
- Tan, K.; Zhang, Y.; Tong, X. Cloud extraction from Chinese high resolution satellite imagery by probabilistic latent semantic analysis and object-based machine learning. *Remote Sens.* **2016**, *8*, 963. [[CrossRef](#)]
- Xie, F.; Shi, M.; Shi, Z.; Yin, J.; Zhao, D. Multilevel cloud detection in remote sensing images based on deep learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3631–3640. [[CrossRef](#)]
- Liu, H.; Du, H.; Zeng, D.; Tian, Q. Cloud detection using super pixel classification and semantic segmentation. *J. Comput. Sci. Technol.* **2019**, *34*, 622–633. [[CrossRef](#)]
- Gao, X.; Fan, L.; Xu, H. Multiple rank multi-linear kernel support vector machine for matrix data classification. *Int. J. Mach. Learn. Cybern.* **2018**, *9*, 251–261. [[CrossRef](#)]
- Xu, C.; Tao, D.; Xu, C. A survey on multi-view learning. *arXiv* **2013**, arXiv:1304.5634.
- Tang, J.; He, Y.; Tian, Y.; Liu, D.; Kou, G.; Alsaadi, F.E. Coupling loss and self-used privileged information guided multi-view transfer learning. *Inf. Sci.* **2021**, *551*, 245–269. [[CrossRef](#)]
- Appice, A.; Malerba, D. A co-training strategy for multiple view clustering in process mining. *IEEE Trans. Serv. Comput.* **2015**, *9*, 832–845. [[CrossRef](#)]
- Kim, D.; Seo, D.; Cho, S.; Kang, P. Multi-co-training for document classification using various document representations: TF-IDF, LDA, and Doc2Vec. *Inf. Sci.* **2019**, *477*, 15–29. [[CrossRef](#)]
- Cao, L.-L.; Huang, W.B.; Sun, F.-C. Optimization-based extreme learning machine with multi-kernel learning approach for classification. In Proceedings of the 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014; IEEE: Piscataway Township, NJ, USA, 2014; pp. 3564–3569.

26. Tang, J.; Tian, Y. A multi-kernel framework with nonparallel support vector machine. *Neurocomputing* **2017**, *266*, 226–238. [[CrossRef](#)]
27. Chao, G.; Sun, S. Consensus and complementarity based maximum entropy discrimination for multi-view classification. *Inf. Sci.* **2016**, *367*, 296–310. [[CrossRef](#)]
28. Tang, J.; Tian, Y.; Liu, D.; Kou, G. Coupling privileged kernel method for multi-view learning. *Inf. Sci.* **2019**, *481*, 110–127. [[CrossRef](#)]
29. Deng, N.; Tian, Y.; Zhang, C. *Support Vector Machines: Optimization Based Theory, Algorithms, and Extensions*; CRC Press: Boca Raton, FL, USA, 2012.
30. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC super pixels compared to state-of-the-art super pixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [[CrossRef](#)]
31. Grant, M.; Boyd, S.; Ye, Y. CVX: Matlab Software for Disciplined Convex Programming. 2008. Available online: <http://cvxr.com/cvx/> (accessed on 15 July 2022).
32. 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](#)]
33. Roy, D.P.; Wulder, M.A.; Loveland, T.R.; Woodcock, C.E.; Allen, R.G.; Anderson, M.C.; Helder, D.; Irons, J.R.; Johnson, D.M.; Kennedy, R.; et al. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sens. Environ.* **2014**, *145*, 154–172. [[CrossRef](#)]