Cloud Image Retrieval for Sea Fog Recognition (CIR-SFR) Using Double Branch Residual Neural Network

Tianjiao Hu, Zhuzhang Jin, Wanxin Yao, Jiezhi Lv, and Wei Jin

***Abstract*—Sea fog is a common weather phenomenon at sea, which reduces visibility and causes tremendous hazards to marine transportation, marine fishing, and other maritime operations. Traditional sea fog monitoring methods have enormous difficulties in characterizing the diversity of sea fog and distinguishing sea fog from low-level clouds. Thus, we propose a cloud image retrieval method for sea fog recognition (CIR-SFR) in a deep learning (DL) framework by combining the advantages of metric learning. CIR-SFR includes the feature extraction module and the retrieval-based SFR module. The feature extraction module adopts the double branch residual neural network (DBRNN) to comprehensively extract the global and local features of cloud images. By introducing local branches and using activation masks, DBRNN can focus on regions of interest in cloud images. Moreover, cloud image features are projected into the semantic space by introducing multisimilarity loss, which effectively improves the discrimination ability of sea fog and low-level clouds. For the retrieval-based SFR module, similar cloud images are retrieved from the cloud image dataset according to the distance in the feature space, and accurate SFR results are obtained by counting the percentage of various cloud image types in the retrieval results. To evaluate the SFR system, we establish a dataset of 2544 cloud images including clear sky, low-level cloud, medium high cloud, and sea fog. Experimental results show that the proposed method outperforms the traditional methods in SFR, which provides a new way for SFR.**

***Index Terms*—Cloud image retrieval (CIR), double branch residual neural network (DBRNN), metric learning, multisimilarity loss, sea fog recognition (SFR).**

I. Introduction

SEA fog is a phenomenon of low-level water vapor condensation that occurs at sea or in coastal areas, where the accumulation of large amounts of water droplets or ice crystals will reduce horizontal visibility to less than 1 km. Offshore activities, such as shipping, fishing, and other maritime operations can suffer significantly from abysmal visibility. Approximately 70% of ship collisions, groundings, and other events are caused by sea fog, making it one of the most serious disaster. Most traditional sea fog monitoring relies on observation located at sea or in coastal areas. Due to the shortage of stations and the high susceptibility of observation facilities to corrosion by seawater, large-scale continuous monitoring of sea fog is challenging. Therefore, the development of the marine economy as well as disaster prevention and mitigation depends on the exploration of alternative and effective methods for sea fog monitoring.

With the development of satellite remote sensing technology, the application of high-precision and multichannel satellite remote sensing images in sea fog monitoring is getting more and more attention. Eyre et al. [1] discussed the potential of bright temperature difference in nighttime sea fog recognition (SFR) by analyzing mid-infrared and far-infrared channels data from the polar-orbiting satellite NOAA; Husi et al. [2] used the brightness temperature difference between the mid-infrared and long-wave infrared channels of the Himawari-8 satellite, and combined it with the snow coverage index to establish a daytime SFR model; Deng et al. [3] proposed a daytime sea fog monitoring method on a dynamic threshold using FY2E satellite data. These researchers implemented sea fog monitoring by seeking a suitable threshold to exploit the difference in cloud radiation between different imaging channels of remote sensing. However, there are some problems in these methods, such as the difficulty in determining the threshold, the difficulty in using the spatial relationship between remote sensing image pixels, and the sea fog monitoring model established that cannot make full use of the satellite remote sensing data of different channels. In recent years, with the development of research on the human brain's visual perception mechanism and computer technology, DL has attracted significant attention from academics across the globe. In particular, deep convolutional neural networks (deep CNNs) have made significant progress in applying remote sensing image processing. In the field of satellite cloud image analysis, a large number of research results have been achieved in Cloud detection, tropical cyclone classification, and cloud cover calculation based on DL, and the application potential of DL in satellite cloud image analysis has been revealed.

Content-based image retrieval (CBIR) is a method that aims to use the image's content to find the same or similar samples from an image database as the query image [4]. In general, different types of clouds correspond to different weather information. If the cloud image similar to the current cloud information can be found in the historical cloud library, by analyzing the weather

Manuscript received 6 December 2022; revised 8 February 2023; accepted 17 February 2023. Date of publication 14 March 2023; date of current version 5 April 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 42071323, and in part by the Public Welfare Science and Technology Project of Ningbo under Grant 202002N3104. *(Corresponding author: Wei Jin.)*

The authors are with the Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China (e-mail: k592483107@outlook.com; fatejzz@foxmail.com; xincase@foxmail.com; 1261416840@qq.com; jinwei@nbu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3257042

conditions and their development trends at a certain moment in history, it is possible to provide supporting information for current weather forecasts or warnings of catastrophic weather. As a promising technology for monitoring severe weather, satellite cloud image retrieval (CIR) is expected to play an important role in sea fog monitoring. However, the characteristics of satellite cloud images differ from those of natural images. For example, satellite cloud images contain a wealth of spectral data, and the properties of various cloud systems in terms of cloud type, extent, boundary shape, and texture are intricate. Furthermore, when people judge the similarity of cloud images, it is based on the understanding of the meteorological semantic information reflected in them, but not based on the similarity of the visual content of the images. Therefore, how effectively understanding and describing cloud images have become the key to satellite CIR.

Before formally conducting the research on the CIR method for SFR, we still need to solve the problems of sea fog diversity and the distinction between sea fog and low-level clouds. Due to the influence of various factors, such as geographic location and season, the sea fog cloud patterns often show multiple manifestations. Moreover, there is no essential physical property difference between low-level clouds and sea fog, and their spectral characteristics are also highly similar. The accuracy of SFR will be hampered by intraclass diversity and interclass similarity that often cause semantic discrimination errors in sea fog recognition. To solve the above problems, deep metric learning (DML) [5] provides a feasible idea.

In this study, in order to solve the problems encountered by the traditional threshold method, using Himawari-8 satellite cloud images as a basis, we propose the cloud image retrieval for sea fog recognition (CIR-SFR). CIR-SFR base on the characteristics of sea fog itself, under the framework of DL, along with the benefits of metric learning in displaying the meteorological semantic features of satellite cloud images. CIR-SFR mainly includes the feature extraction module and the retrieval-based sea fog recognition module. In CIR-SFR, the feature extraction module uses a double branch residual neural network (DBRNN). The backbone structure of DBRNN is a double-branch network, where the global branch contains the feature extractor (FE) and generates the activation mask (AM), which acts on the original input cloud image, and the local branch contains the FE to extract the features of the main cloud regions in the cloud image. Due to the high similarity between low-level cloud and sea fog in satellite cloud image representation, the generated cloud image features will show high intraclass differences and interclass similarities in the embedding space. Therefore, in the process of network training, DBRNN will be co-trained through multisimilarity loss (MS Loss) and dual branches under the basic requirement of CIR. The trained model embeds the cloud images into a metric space in which the distance between different cloud images increases and the distance between the same cloud images decreases. The retrieval-based sea fog recognition module uses the cloud image features extracted by the trained DBRNN to retrieve the historical cloud images, which are similar to the query in the database according to the distances between different cloud images in the feature space. By calculating the weight share of similar historical cloud images in various types of cloud images, the category of the query is inferred, and accurate SFR results are obtained. The main contributions of this article can be summarized as follows:

1) Addressing the difficult to determine thresholds of traditional SFR methods encountered, we propose a CIR-SFR.
2) We construct DBRNN with DML to solve the problem that the traditional deep features are high intraclass differences and interclass similarities of different categories of satellite cloud images, so as to improve the discrimination ability of sea fog and low-level cloud.
3) We establish a dataset containing the most significant cloud categories, such as clear sky, medium high cloud, low-level cloud, and sea fog. This lays the foundation for further research on satellite CIR, cloud image identification, and sea fog monitoring.

II. Related Work

In this section, we present some work related to CIR-SFR, including deep CNNs, DML, and satellite CIR.

A. Deep CNNs

In order to avoid the complicated image feature extraction process, classical seep CNNs algorithms, such as AlexNet [6], GoogleNet [7], VGGNet [8], and ResNet [9], which can better map ordinary images to the feature space and portray the semantic features of images in the form of space vectors. In recent years, deep CNNs have also received extensive attention from scholars in the field of satellite remote sensing [10]. However, due to the rich spectral information contained in satellite cloud images, and the cloud characteristics reflected in satellite cloud images, such as shapes and textures are very complicated, ordinary classical networks cannot effectively describe the meteorological information contained in cloud images. Therefore, various improved versions of deep networks have emerged and been applied in cloud image segmentation, cloud image recognition, and cloud image classification. Kaur Buttar et al. [11] proposed a segmentation method for clouds and certain terrians in satellite images. The method combines the U-Net++ with a light weight channel attention mechanism to create crisp cloud boundaries. Shao et al. [12] proposed a CNN based on multiscale features. The network combines high-level semantic information and low-level spatial information generated during feature learning to achieve the simultaneous classification of thin, thick, and noncloudy pixels. However, the application of deep CNNs in satellite cloud image is still few, the development time is short, and the algorithm is not mature enough. Deep CNNs has also been applied and developed in the field of sea fog detection and prediction. Ran et al. [13] developed an algorithm for sea fog detection during morning and evening hours in the framework of DL combined with terrain constraints. Zhu et al. [14] used U-Net DL model combined with PCA to accomplish effective sea fog detection. In the field of SFR, the threshold method still occupies the main position, and DL is still less. The algorithm cannot reasonably use multichannel data, and the ability of distinguishing sea fog from low-level cloud is limited. Therefore, we use the multichannel satellite cloud image data, combined with the advantages of DL, to further explore the SFR method.

B. DML

Traditional feature extraction methods based on DL have a pair of contradictions: one contradiction is that the spatial position relationship between samples will not be considered by the deep network during training; the other is that when the extracted features are used for image classification and image retrieval, it is necessary to assume that there is a meaningful distance measure in the input space [15]. The emergence of DML provides a feasible idea to solve this contradiction. The purpose of DML is to learn a low-dimensional image embedding function, through which the image is embedded into a metric space. In this metric space, the distance between images of the same class is smaller than that between images of different classes, so as to minimize the intraclass distance and maximize the interclass distance. At present, DML is widely used in image retrieval [16], recognition [17], verification [18], and feature matching [19]. Triplet training [5] is a common way to implement DML. The triplet $t=\{g^a, g^p, g^n\}$ consists of anchor samples, positive samples, and negative samples, where anchor samples g^a and positive samples g^p are in the same category, the negative samples g^n belong to different categories. During training, g^a and g^p form a set of positive sample pairs, while g^a and g^n form a set of negative sample pairs, which makes the distance between positive sample pairs reduced and the distance between negative sample pairs expanded through network training. In addition, contrastive training based on the Siamese network and quadruplet training based on triple adding a negative sample can also be practical for DML. However, in the process of model training, multivariate methods need to construct multivariate sample groups, which will lead to the redundant selection of training samples, further causing difficulties in model training convergence and model performance degradation. In order to solve this problem, we introduce multisimilarity loss (MS loss) [20] in CIR-SFR for double-branch co-training and optimization. MS Loss combines self-similarity and relative similarity, which can excavate difficult sample pairs suitable for training to reduce the total number of sample pairs during training, which can not only effectively improve the training speed but also obtain more discriminative image features in the embedding space.

C. Satellite Cloud Image Retrieval

Satellite CIR belongs to the category of remote sensing image retrieval. Compared with other common images, the retrieval accuracy of satellite cloud images depends more on the model's understanding and description of the cloud image itself, which is the representativeness and comprehensiveness of the features extracted by the model. Traditional satellite CIR relies on the accuracy of manually extracted features, which are divided into three main visual features: 1) hue, 2) structure, and 3) texture. Acqua et al. [21] used point diffusion technique to compare the shape similarity between cloud images, thus realizing CIR. Gurve et al. [22] extracted morphological, color and texture features from satellite cloud images, respectively, and developed a content-based image retrieval system. With the continuous progress of DL, its advantages in feature extraction have gradually emerged, and DL is gradually applied to remote sensing image retrieval. S Roy et al. [23] proposed a DML based on hash network, which integrated transfer learning and a triple training scheme to optimize the target retrieval task and obtain binary hash codes for fast search. Y Liu et al. [24] adopted the similarity-based conjoined CNN (SBS-CNN) to generate compact image features, and proposed an unsupervised deep transfer learning method based on similarity for remote sensing image retrieval. Although DL has made much progress in remote sensing image retrieval, little research has been applied to CIR. In order to solve the problem that traditional SFR method based on the threshold is difficult to depict cloud image spatial information, and the traditional SFR method based on classification model cannot effectively use cloud image of history information. Inspired by the success of DL in the field of remote sensing image retrieval, in this article, we study in DL framework and carry out the work on CIR-SFR.

III. CIR-SFR

In this section, we propose the motivation for the study of CIR-SFR, then present the general framework of CIR-SFR, and finally show the implementation and optimization of the components of CIR-SFR.

A. Problem Formulation

Different types of clouds correspond to different weather information. If the features of two temporal clouds are similar, the weather development processes corresponding to them are likely likewise identical. Therefore, if the cloud image that is similar to the current cloud information can be found in the historical cloud library, by analyzing the weather conditions and its development trend at a certain moment in history, it is possible to provide auxiliary information for current weather forecasts or warnings of catastrophic weather. Sea fog has specific spectral and textural characteristics on satellite cloud images as a type of catastrophic weather. Therefore, the problem of SFR can be solved by cloud class recognition of satellite cloud images. However, traditional cloud image classification based on DL can only give the discriminant result, but not fully express cloud meteorological semantics, which results in poor interpretability of the method. Therefore, to increase the accuracy of sea fog detection and simultaneously improve the recognition results, this work presents a SFR method based on CIR with the support of similar historical cloud images. Considering the rich spectral information and complex texture characteristics of satellite cloud images, the features extracted by a single network structure are difficult to comprehensively and pertinently describe the meteorological information contained in cloud images. In this article, a double branch residual neural network (DBRNN) is used as the backbone structure of the model, MS loss is introduced to conduct CIR, and finally realize SFR based on the retrieval results. In order to improve the clarity of the text, we have installed the following Table I to illustrate some important symbols and definitions in the article.

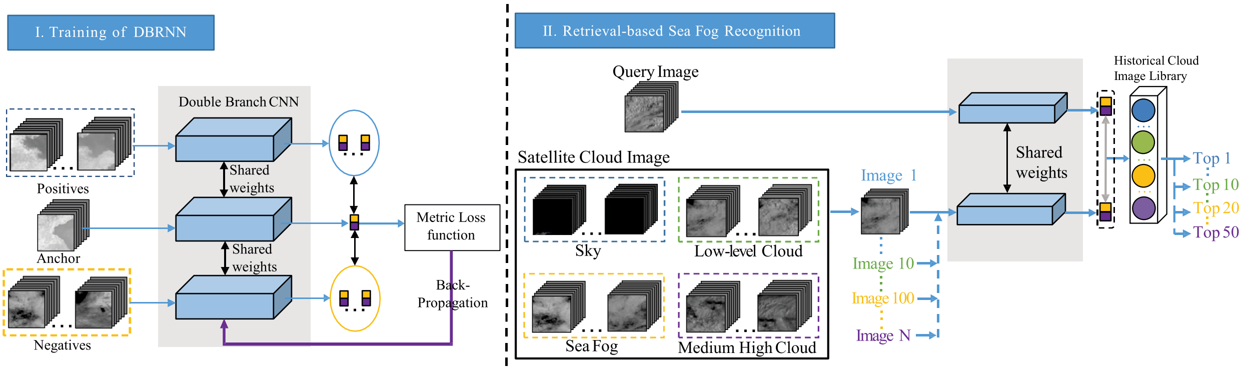


Fig. 1. Structure of the proposed CIR-SFR.

TABLE I
NOTATIONS AND DEFINITIONS

Notation	Definitions
X	Raw images.
C	Category of cloud image.
F	Feature map.
$\max()$	Maximum value function.
$\min()$	Minimum value function.
H	Aggregated feature map.
$\hat{H}$	Activation map.
α, β, γ	Hyperparameters.

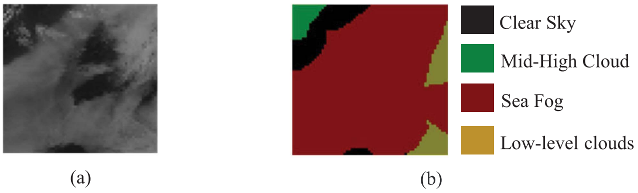


Fig. 2. Cloud image labeled as sea fog with its semantic segmentation mask. (a) Sea fog cloud image. (b) Cloud image corresponds to the semantic segmentation mask image.

B. CIR-SFR

The structure of CIR-SFR proposed in this article is shown in Fig. 1. The backbone network of CIR-SFR includes the training of DBRNN module (the feature extraction module) and the retrieval-based SFR module. In the training of DBRNN module, first, we choose one cloud image as the anchor sample, and its similar and nonsimilar cloud images form the positive and negative sample sets, respectively. Second, in order to construct the historical cloud image library, we introduce metric learning to train DBRNN, using anchor samples and samples from the positive (negative) sample sets. The trained DBRNN is used to extract cloud image features and construct a historical cloud image library. Then, in the retrieval-based SFR module, we extract the features of query by the trained DBRNN. The Top-50 similar cloud images are retrieved from the historical cloud image library based on the similarity among the features. Finally, based on the feature distance between similar cloud images and query, we calculate the weighted scores of different categories of cloud images in the retrieval results by “weighted voting,” so as to determine the category of query and realize SFR.

1) DBRNN: In this article, in order to recognize sea fog, we try to use image retrieval to determine the query category, which needs to solve two problems. The first problem is that satellite cloud images are different from natural images. It is so difficult to label the satellite cloud images that the number of labeled samples are small, which often leads to overfitting and damages the performance of CIR model during the deep network training. In order to solve this problem, metric learning is introduced to train the network. On the one hand, in order to alleviate the problem of having few training samples, metric learning can reuse samples by constructing sample groups. On the other hand, an embedding space is obtained by metric learning, in which the embedding vectors of similar samples are pulled closer while the embedding vectors of different samples are pushed away in order to enhance the network’s ability to express the features of samples. The second problem is that in the process of cloud image labeling, if the proportion of a certain category of cloud or fog regions in the cloud image is more than 50%, the label of the cloud image is labeled with this category. However, the cloud image often contains other types of clouds or fog regions, and the ability of network to describe the essential characteristics of the cloud image will be affected by these other types of cloud or fog regions. A cloud image labeled as sea fog and its semantic segmentation mask in the training set are shown in Fig. 2.

Fig. 2 shows that although most of this cloud image are the sea fog region, which is the main semantic object of this cloud image, there are still some medium high clouds, low-level clouds, and clear skies mixed in. In order to reduce the interference of nondominant semantic objects in the cloud image as much as possible, this article constructs the DBRNN combined with local features to extract features based on the traditional CNN. The framework of DBRNN is shown in Fig. 3.

Fig. 3 shows that DBRNN includes the global branch and the local branch, where the global branch extracts the global features of the cloud image, and the local branch extracts the local features of the main semantic objects in the cloud image by the effect of the activation mask (AM). Both branches use ResNet50 as the backbone network, and its feature extraction is mainly divided into four stages, each stage is composed of a number of residual blocks. Stage 1 and stage 4 have three residual blocks each, stage 2 has four residual blocks, and stage 3 has six

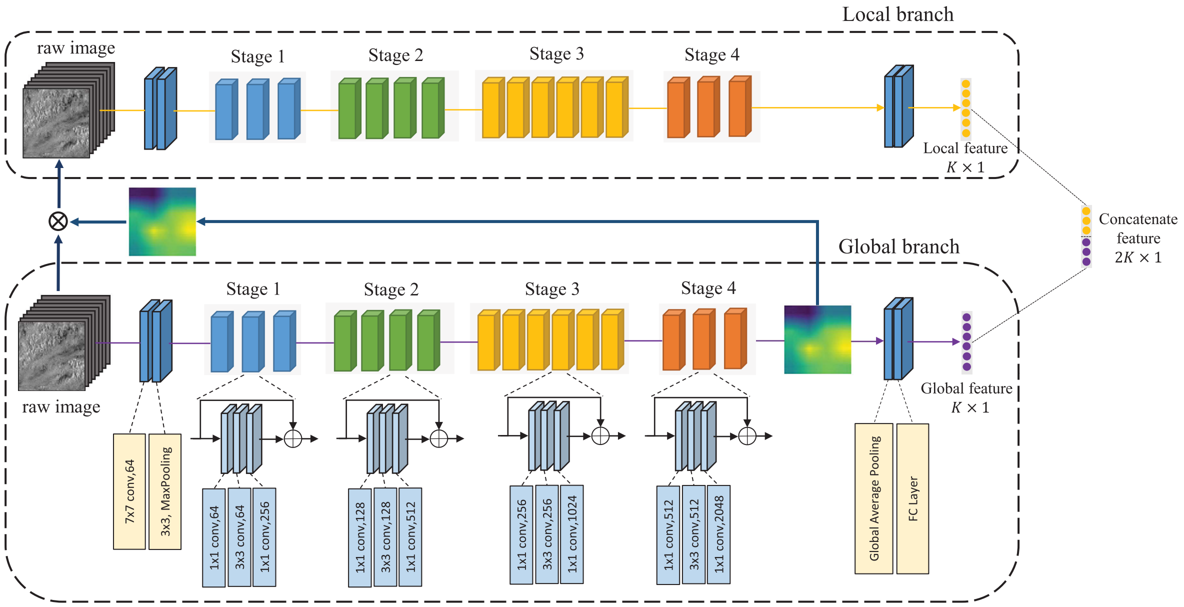


Fig. 3. Framework of the DBRNN.

residual blocks. After the fourth stage, the feature map will pass through the global max pooling layer, and the fully connected layer with the ReLu function as the activation function, each channel's spatial information is aggregated and mapped to the low-dimensional embedding space to generate a feature vector of length K. Finally, for representing the cloud image, the feature vectors of the double branches are concatenate to form a feature vector of length $2\,K$.

2) AM: In order to extract features from the main semantic objects of the cloud image, the AM is generated by performing a series of processing on the feature map of the global branch stage 4, and then the AM is multiplied with the original cloud image as the input of the local branch, as follows.

We assume that the map of global branching stage 4 is $F \in R^{C\times L\times W}$, where C, L, and W represent the number of channels, height, and width of the feature map, respectively. According to the characteristics of the CNN, different semantic objects in the cloud image are represented in different positions of the high-level feature map. Therefore, when the network detects the existence of a certain semantic, the corresponding region in the feature map will be activated. If the feature maps of multichannels are activated in the same region, it often means that the corresponding region in the original cloud image is the main semantic object. For this purpose, we superimpose the feature map F along the channel dimension to obtain the "aggregated feature map" H, it is defined as

$$H(i,j) = \sum_{c=1}^{C} F(c,i,j) \tag{1}$$

where $H(i,j)$ represents the amount of information at position (i,j) in the H. The greater the amount of information, the more likely the corresponding region contains the main semantic object. To measure the strength of the information at different locations and to establish a one-to-one mapping between the information and each pixel of the original cloud, we normalize H and upsample it to the original cloud size to generate AM $\hat{H}$, which can be defined as follows:

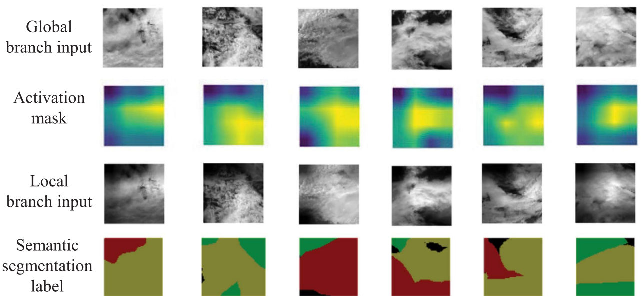


Fig. 4. Cloud image and its corresponding semantic segmentation labels, activation masks, and local branching input.

$$\hat{H}(i,j) = \text{upsample}\left(\frac{H(i,j) - \min(H)}{\max(H) - \min(H)}\right) \tag{2}$$

where max() and min() return the maximum and minimum pixel values of H, respectively, and upsample() indicates the upsampling operation. In order to emphasize the main semantic objects in the cloud image and suppress the interference of other nonmain semantic objects, we multiply the original cloud image with the AM. Several cloud images with their semantic segmentation labels, activation masks, and local branching input are shown in Fig. 4.

Fig. 4 shows that the AM well characterizes the main semantic regions of the cloud image. Compared with the original cloud image, the input of the local branch pays more attention to the dominant cloud region in the cloud image to make the extracted features better reflect the category information.

3) Design of Loss Function and Hard Sample Mining: The loss function plays a significant role in DML, and many loss functions for DML are built on top of sample pairs or sample triples. In the SFR task, there is no essential difference in physical features between sea fog and low-level cloud, and the spectral characteristics are also very similar, so the recognition problem of nonidentical but extremely similar samples needs to be solved. In order to solve this problem, we use the strategy of sample pairs to train the network and introduce multisimilarity loss (MS Loss) [20]. MS loss captures better information from sample pairs by comprehensively considering three aspects of the similarity of self-similarity of sample pairs, the similarity of negative sample pairs, and similarity of positive sample pairs. Since the proposed DBRNN has both global and local branches, in order to combine the information of the sample pairs in the double branches, we designed the following loss function based on the cosine similarity of the sample pairs, which is defined as follows:

$$\begin{aligned} L_p = \frac{1}{D\alpha}\sum_{i=1}^{D}\Bigg(\log\Bigg[1+\sum_{K\in P_i} e^{-\alpha(S_{ik}^{g}-\lambda)}\Bigg] \\ \times\Bigg[1+\sum_{K\in P_i} e^{-\alpha(S_{ik}^{l}-\lambda)}\Bigg]\Bigg) \\ +\frac{1}{D\beta}\sum_{i=1}^{D}\Bigg(\log\Bigg[1+\sum_{K\in N_i} e^{\beta\left(S_{ik}^{g}-\lambda\right)}\Bigg] \\ \times\Bigg[1+\sum_{K\in N_i} e^{\beta(S_{ik}^{l}-\lambda)}\Bigg]\Bigg) \end{aligned} \tag{3}$$

where D is the batch-size, S_{ik}^{g} and S_{ik}^{l} are the cosine similarity between the features extracted from the anchor sample cloud X_i and the paired sample X_k after global branching and local branching; P_i and N_i are the set of positive and negative samples of the anchor sample cloud map X_i; α, β are hyperparameters as in Binomial deviance loss [20], where α controls the weight of positive pairs and β controls the weight of negative pairs; λ represents the margin of similarity. The processing of positive samples in batch is shown in the top half of (3). In this part of the loss, λ controls the closeness of positive sample pairs and penalizes those positive sample pairs with similarity < λ. The processing of negative samples in the batch is shown in the bottom half of (3). This partial loss ensures that the similarity of negative samples to the anchor is as low as possible, which means that negative samples close to the anchor (i.e., with high similarity) should be penalized more than negative samples far from the anchor (i.e., with lower similarity). Therefore, through (3), the network can learn an embedding space that brings similar samples closer and pushes different samples away, so that the extracted features are more conducive to completing the SFR task.

When randomly selecting anchor samples and positive (negative) samples to form training sample pairs, a large number of redundant samples are generated, which will reduce the training speed and make little contribution to the improvement of model performance [20]. Therefore, we introduce a difficult sample selection strategy to retain only those sample pairs with more valuable information. Taking global branching as an example, we assume that X_i represents the cloud image anchor sample, y_i represents its label, and X_j represents the cloud image chosen randomly from the dataset such that it belongs to a class y_j, where $y_j \neq y_i$. Only when the similarity S_{ij} between X_i and X_j in the corresponding negative sample satisfied (4), the two form a valid negative sample pair. The process of negative sample selection can be expressed as follows:

$$S_{ij}^{-} > \min_{y_i=y_k} S_{ik} - \varepsilon \tag{4}$$

where ε represents a preset threshold, which is set to 0.1 in this article. Equation (4) indicates that only negative samples whose similarity to the anchor samples are larger than the minimum of their positive sample's similarity will be included in the training. Similar to the negative sample selection rule, we assume that X_l represents the samples of the same category as X_i. Only when the similarity S_{il} between X_i and X_l satisfied (5) they will form a valid positive sample pair. The process of positive sample selection can be expressed as follows:

$$S_{il}^{+} < \max_{y_i\neq y_k} S_{ik} + \varepsilon. \tag{5}$$

Equation (5) shows that only positive samples whose similarity to the anchor samples are smaller than the maximum of their negative sample's similarity will be included in the training.

C. Cloud Retrieval and Sea Fog Recognition

The satellite cloud image pairs will be embedded into the feature space by the trained DBRNN, and the retrieval is based on the distance of the cloud image pairs in the feature space: the same class is closer, while the different classes are far away. The labeled training set data are used as the gallery set, and the test set is used as the query set. Cloud class recognition can be achieved by CBIR and majority voting. The details can be defined as follows.

First, the query X is input to DBRNN to generate its global features and local features, and they are concatenated to form the feature vector characterizing the cloud image

$$x = [x^g, x^l] \tag{6}$$

where x^g and x^l, respectively, represent the outputs of the global branch and local branch. Then, x will be compared with the features of other clouds in the gallery set, and the retrieval results are sorted by the distance between features from smallest to largest. In order to avoid the problem that the proportion of the two types of clouds in the returned results is the same, which makes it impossible to complete the recognition task according to the majority voting. We motivated by the weighted K-nearest neighbor (KNN) algorithm, combined with the Gaussian function, and finally, defined the weight of the returned cloud image based on the distance between the returned cloud image features and the query features. The weight can be described as follows:

$$W(X,Z) = ae^{-\frac{(\mathrm{dist}(X,Z)-b)^2}{2c^2}}. \tag{7}$$

In the equation, $\mathrm{dist}(X, Z)$ is the Euclidean distance between the query X and the returned cloud image Z features; a represents the peak of the Gaussian curve; b is the horizontal coordinate corresponding to the peak; c is the standard deviation, which is experimentally determined to be 1, 0, and 0.3, respectively, to get better results. To get the score that the query belongs to different categories, we summarize the weight sum of the returned TOP N cloud images belonging to different categories by category. The score can be expressed as

$$\mathrm{Score}(X, C_i) = \sum_{Z_j \in C_i} W(X, Z_j), j = 0, 1, \ldots, N. \quad (8)$$

In the equation, C_i represents different cloud image categories, and we determine the category of the query by the highest score, which is defines as follows:

$$c(x) = \arg\max_{C_i}(\mathrm{Score}(X, C_i)) \quad (9)$$

where $c(x)$ represents the category to which the query X is interpreted by the model, and if this category is sea fog, then SFR is achieved at the same time.

IV. Construction of the Dataset

A. Analysis of Spectral Characterization of Satellite Cloud Images

In general, different types of cloud images often have different spectral characteristics, but there is no essential difference between low-level cloud and sea fog in physical properties and spectral characteristics. Therefore, it is still a challenge to distinguish low-level clouds from sea fog by remote sensing detection. In order to solve this problem, first, we analyze the remote sensing spectral characteristics of four types cloud images, such as sea fog, medium high cloud, low-level cloud, and clear sky. The purpose is to screen out the imaging channels that can effectively reflect the characteristics of different types of cloud images.

In the visible to near-infrared channel, the signal received by satellite consists of a cloud layer, the solar radiation reflected by the underlying surface, and scattered radiation of solar radiation in the atmosphere, Earth's atmospheric radiation due to the latter than the former proportion is very small, negligible. Therefore, the cloud image of this channel mainly reflects the reflective nature of the cloud and the underlying surface to solar radiation. Based on Mie's scattering theory [25], clouds and fog have obvious scattering effects, and their reflectance is significantly greater than the information on the sea surface, land, and other features, so the visible-NIR channel cloud images can effectively distinguish between clouds and clear sky. Moreover, since the sea fog is closer to the surface, its diffuse reflection from the ground or other directions is less, which causes that if the low-level cloud and the sea fog are of the same thickness, the reflectivity of the sea fog will be smaller than that of the low-level cloud. In addition, studies by Zeng-Zhou Hao et al. [26] also pointed out that the particle size differences between cloud and mist can cause them in visible to near-infrared wave channels to have different reflection and scattering effects, which not only caused the cloud in near-infrared wave channel imaging characteristic difference, also makes the low-level clouds and the sea fog in the visible channel have different texture features. Therefore, the visible-NIR channel can identify cloudy areas and clear sky, low-level clouds, and sea fog.

In the far-infrared channel, the radiation received by satellite mainly comes from the thermal radiation emitted by the target feature itself, so the cloud image in the far-infrared channel mainly represents the temperature and specific emissivity of the target feature itself. The lower the target temperature, the lower the specific emissivity, and the lower the radiation value received by satellite. Medium high clouds have high altitudes and low temperatures, so their bright radiation temperature is significantly lower than other underlying surfaces, while low-level clouds are close to the sea surface, and their bright temperature is close to the clear sky. Therefore, the far-infrared channel can be used as an auxiliary channel to distinguish medium high clouds from low-level clouds.

Combined with the above analysis, in order to comprehensively utilize the information from the different imaging channels, we select cloud image data in three visible channels, three near-infrared channels and one far-infrared channel (11.2 μm) from the Himawari-8 satellite for the study of CIR for SFR.

B. Cloud Images Acquisition and Annotation

We select the Bohai Sea and Yellow Sea waters which are located at longitudes $116.5°$–$129.25°$E and latitudes $30°$–$42.5°$N as the main study area. This region is affected by the encounter between the warm current entering the Yellow Sea along the northwest direction south of Jeju Island and the coastal current of the Yellow Sea, so the sea fog occurs more frequently. According to incomplete statistics, more than 80 offshore foggy weather events occurred in this sea area during 2018–2020. In order to obtain the appropriate satellite cloud image of sea fog, first, we obtain the sea fog occurrence date according to the fog monitoring report published by China Ecological Remote Sensing Information Service Network. Then, we collect the Himawari-8 full-disk cloud image data of the sea fog occurrence date from 2017 to 2020. Finally, according to the longitude and latitude of the study area, the cloud image data are intercepted to generate the corresponding satellite cloud image of sea fog.

Referring to the cloud classification products of Himawari-8 and combining with the fog monitoring report, we collected different categories of cloud images to form a cloud image dataset, and the specific methods are as follows.

1) *Cloud Image Segmentation:* Himawari-8's cloud classification product divides clouds into nine categories: 1) Cirrus, 2) Cirrostratus, 3) deep convection, 4) Altocumulus, 5) Altostratus, 6) Nimbostratus, 7) Cumulus, 8) Stratocumulus, and 9) Stratus. Cirrus, Cirrostratus, deep convective, high cumulus, high stratus, and rain stratus clouds belong to medium high clouds; cumulus, stratocumulus, and stratus clouds belong to low-level clouds. Therefore, based on cloud classification products, through pixel segmentation, we can classify satellite cloud images

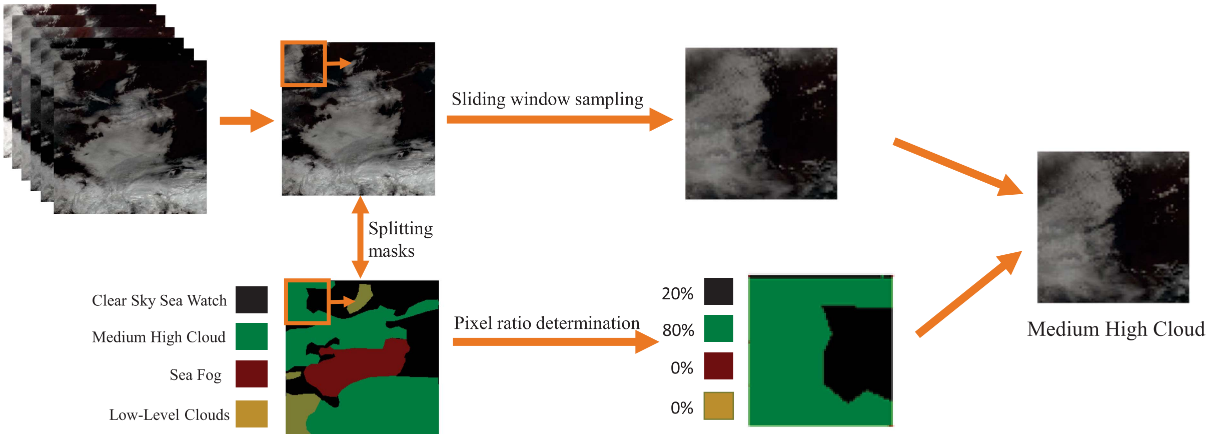


Fig. 5. Cloud image collection and annotation process.

into three categories: 1) medium high cloud, 2) low-level cloud, and 3) clear sky.

2) *Generate Semantic Segmentation Mask Maps*. Combined with the fog monitoring reports, the low-level cloud and fog regions in the cloud image are further segmented into low-level cloud and sea fog, so as to generate a semantic segmentation mask map with four objects types: 1) medium high cloud, 2) low-level cloud, 3) sea fog, and 4) clear sky.

3) *Cloud Image Collection to Generate Datasets*. We sample the original cloud image using a sliding window with a window width of 128 pixels and a step size of 64 pixels. During the sampling process, with reference to the segmentation mask map, when the pixel share of a certain category of clouds in the window is more than 50%, the cloud image covered by the window is intercepted as a sample and labeled with the corresponding category. This process is shown in Fig. 5.

Fig. 5 shows an example of the acquisition and annotation process for a sample of medium- high cloud images. According to this method, the cloud image dataset we constructed contains 566 clear sky cloud images, 579 sea fog cloud images, 726 medium high cloud images, and 662 low-level cloud images, for which we conduct the CIR-SFR studies.

V. Experiment

The experimental system was configured with a Windows 10 operating system, a 4.1 GHz Intel Core i5-10600KF CPU, and a computer with 32 GB of running memory, the programming language Python 3.7. The deep network model builds on the Keras and Tensorflow frameworks, with all convolutional operations performed on the graphics card NVIDIA GeForce RTX3060. We divide the Yellow Bohai Sea sea fog dataset into a training set and a test set in the ratio of 8:2, and add 11 sea fog cloud images from 2021 to 2022 to the test set. In order to mitigate the overfitting problem of the network, each training cloud image will be rotated by 90°, 180°, and 270° to augment the training set. We randomly select 16 cloud images from various types of cloud images to form a batch and used Adam optimization [27] for network training, with the initial learning rate set to 10^{-5}. By referring to the analysis in literature [20], the hyperparameters α, β, and λ of the loss function set to 2, 20, and 1, respectively. The number of neurons K, in the fully connected layers of the global branch and the local branch in the DBRNN model, is set to 64. The goal of this article is to solve the SFR problem using image retrieval, so the following experiments are conducted to evaluate the performance of the proposed model in CIR-SFR.

For the SFR task, we divide all samples of the dataset into two categories: 1) positive and 2) negative classes where all cloud images belonging to sea fog are positive classes and all cloud images of other categories are negative classes. After the test dataset is discriminated by the model, there are four main cases: 1) True positive (TP) represents the sample recognized by the model as sea fog, which is actually also sea fog; 2) false positive (FP) represents the sample recognized by the model as sea fog, which is actually not sea fog; 3) false negative (FN) represents the sample recognized by the model as nonsea fog, which is actually sea fog; 4) true negative (TN) represents the sample recognized by the model as nonsea fog, which is actually also nonsea fog. We use precision (PRE), recall, and F1-score ($F1$) to evaluate the performance of the model for SFR. The PRE, recall, and $F1$ can be formulated as follows:

$$\text{PRE} = \frac{TP}{TP + FP} \tag{10}$$

$$\text{recall} = \frac{TP}{TP + FN} \tag{11}$$

$$F1 = \frac{2 \times \text{PRE} \times \text{recall}}{\text{PRE} + \text{recall}}. \tag{12}$$

In the equation, PRE represents the fraction of sea fog correctly recognized by the model as sea fog over all the clouds predicted to be sea fog, recall represents the proportion of sea fog clouds correctly recognized by the model to all the sea fog clouds, and $F1$ can combine the PRE and recall of the model, which is drawn within the range of [0, 1]. The higher values of PRE, recall, and $F1$, the better the model performance.

We also use the PRE, recall, and $F1$ of the DBRNN model for recognizing other types of cloud images and calculate them as an average metric to measure the performance of the model in cloud graph classification, which can be defined as follows.

$$\mathrm{PRE}_{\mathrm{avg}} = \frac{1}{4}\sum_{i=1}^{4}\mathrm{PRE}_i \tag{13}$$

$$\mathrm{recall}_{\mathrm{avg}} = \frac{1}{4}\sum_{i=1}^{4}\mathrm{recall}_i \tag{14}$$

$$F1_{\mathrm{avg}} = \frac{1}{4}\sum_{i=1}^{4}F1_i. \tag{15}$$

For the cloud image retrieval task, we evaluated the performance of the model using Precision@L ($P@L$) and Mean Average Precision (mAP) [28].

$$P@L = \frac{1}{Q}\sum_{i=1}^{Q}\frac{m(x_i)}{L} \tag{16}$$

$$mAP = \frac{\sum_{q=1}^{Q} AP(x_i)}{Q} \tag{17}$$

where L is the number of returned clouds given at the time of retrieval, $m(x_i)$ represents the number of clouds images returned in the same category as the cloud image x_i to be retrieved, and Q is the total number of images to be retrieved. For clarity of presentation, we abbreviate Precision@L as $P@L$ in the following. In (17), AP is the average precision of retrieval for each cloud image, which is calculated as follows:

$$AP(x_i) = \frac{1}{M(x_i)}\sum_{j=1}^{M(x_i)}\frac{j}{L_j} \tag{18}$$

where $M(x_i)$ represents the number of cloud images in the same category as x_i in the cloud image dataset. At the jth retrieval, the system retrieves exactly j cloud images of the same category when the number of returned cloud images is given as L_j.

A. Performance Analysis of DBRNN Model for Cloud Graph Retrieval

In order to show the advantages of the DBRNN model compared with state-of-the-art method, this article conducts a comparison experiment with other retrieval algorithms on the Yellow Sea and Bohai Sea sea fog dataset, and the comparison methods include: DLBHS [29], Milan [23], and DSH [30]. The length of cloud features of all methods is 64, DLBHS, Milan, and DSH only use visible clouds to achieve retrieval. Experiments were conducted to evaluate the retrieval performance of different methods in terms of $P@L$ and mAP when returning 10, 20, 30, and 50 clouds, and the results are shown in Table II.

TABLE II
COMPARISON OF RETRIEVAL PERFORMANCE BETWEEN DIFFERENT RETRIEVAL METHODS

Methods	$mAP(\%)$	$P@10(\%)$	$P@20(\%)$	$P@30(\%)$	$P@50(\%)$
DLBHS	82.07	83.13	83.42	83.57	83.34
MiLan	68.86	72.84	72.72	72.35	71.94
DSH	67.55	53.02	62.66	67.57	69.48
Proposed	**89.90**	**90.58**	**90.57**	**90.53**	**90.48**

The performance of our model is highlighted in bold.

As can be seen from Table II, compared with the traditional methods that only use visible bands to extract cloud image features, the DBRNN model extracts more discriminative features from multichannels cloud images. The DBRNN has a mAP value of 89.90%. It is 7.83%, 21.04%, 22.35% higher that DLBHS, MilLan, DSH, respectively. The reason for this is that through the operation of AM, the global feature and local feature information of the cloud image are fused, so that the extracted features can better describe the main meteorological semantic objects in the cloud image. Moreover, during the training of the network, the DBRNN model uses the MS Loss function to optimize the distribution of samples in the embedding space, which further improves the retrieval performance.

To visually show the retrieval performance of the DBRNN model, we present the retrieval results of the same unlabeled sea fog cloud images using the DLBHS and DBRNN models, respectively. The results are shown in Fig. 6.

The top 1 row shows the retrieval result for the DBRNN model, while the bottom 1 row shows the retrieval result for DLBHS. As can be seen from the TOP 7 returned clouds, the retrieval results of the DBRNN model are all sea fog clouds, and the visual similarity between the cloud images and those to be retrieved is well characterized by the distance. However, the retrieval results of DLBHS contain several images of low-level cloud, which indicate that the retrieval accuracy of DLBHS is unsatisfactory, and the physical properties of low-level clouds and sea fog are too similar to be effectively identified by traditional method. This experiment shows that the proposed DBRNN model can better solve distinguish the low-level cloud and sea fog in sea fog monitoring. In order to further demonstrate the cloud retrieval performance of DBRNN and DLBHS, we use the T-distributed stochastic neighbor embedding (T-SNE) [31] model.

From Fig. 7, it can be seen that in the embedding space, the DBRNN model has better cloud image aggregation and better differentiation ability than the DLBHS method for various types of clouds and fog. As an example, in the case of sea fog images, many sea fog samples of DLBHS method are scattered in the aggregation region of low-level clouds, while the DBRNN model can well achieve the separation of low-level cloud samples from sea fog samples. Moreover, the DLBHS method suffers from serious confusion between low-level cloud and medium-high cloud samples, while the DBRNN model is advantageous for low-level cloud and medium-high cloud identification.

B. Performance Analysis of DBRNN Model for Cloud Image Classification

The CIR system based on the DBRNN model can infer the category of the query by retrieving the historical cloud images that are similar to the query in the cloud images library, so as to achieve cloud image classification. In order to verify the performance of the DBRNN model on cloud classification tasks, the classical machine learning (ML) methods and conventional

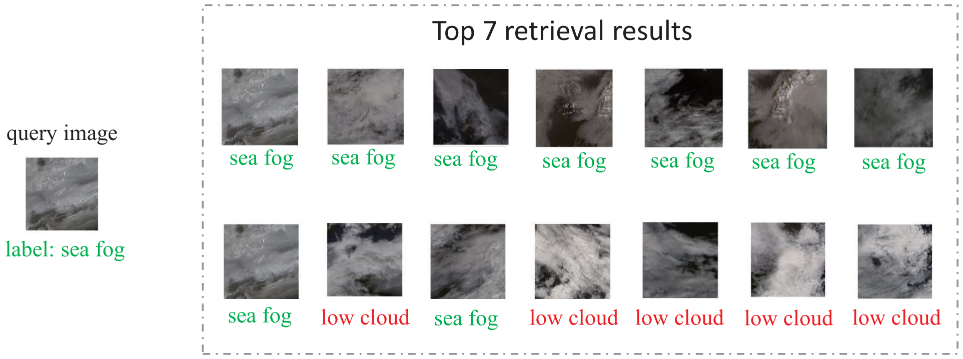
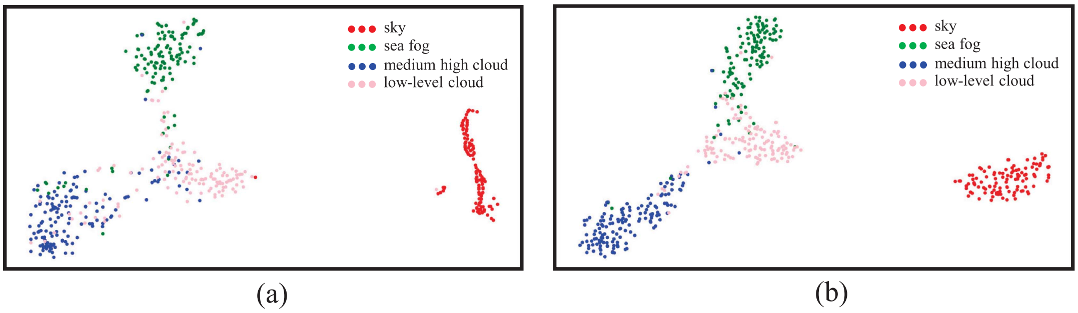


Fig. 6. Retrieval results of the same query.



(a)

(b)

Fig. 7. Visualization of Embedding Spaces for Different Methods. (a) DLBHS. (b) Model we proposed.

DL methods are chosen to conduct comparison experiments with the DBRNN model-based methods. The classical ML methods include random forest classifier (RF) [32], logistic regression (LR) [33], and support vector machines (SVM) [34]. In the feature extraction stage, the model first generates gray level cooccurrence matrix (GLCM) according to the visible channel cloud images, and then extracts statistics, such as angular second moment (ASM), contrast, and entropy as the texture features of the cloud images using GLCM. Finally, the model combines the spectral features in different channels for cloud images to carry out cloud classification experiments. For the DL methods, we carry out two groups of experiments: the first group adopts the traditional DL method, uses ResNet50 as the basic network, train the network with visible channel cloud image data (vis), and multichannel cloud image data (mc), respectively; the second group introduces the generative adversarial model (GAN) [35] based on the original DBRNN's model, and uses GAN to generate pseudo-cloud images to increase the number of training samples. They are compared with the DBRNN based cloud classification model, and the experimental results are shown in Table III.

TABLE III
Cloud Classification Performance of Different Models

	Method	PRE$_{\text{avg}}(\%)$	$recall_{\text{avg}}(\%)$	$F1_{\text{avg}}(\%)$
ML	RF	85.01	83.99	84.25
	LR	76.36	75.83	75.97
	SVM	70.98	68.91	69.36
DL	Resnet50(Vis)	86.91	86.06	86.16
	Resnet50(mc)	91.03	90.53	90.61
	DBRNN(GAN)	92.65	92.44	92.49
	Proposed	**94.95**	**94.71**	**94.76**

The performance of our model is highlighted in bold.

As can be seen from Table III that compared with other classical ML methods, RF has the best performance, the PRE$_{\text{avg}}$, recall$_{\text{avg}}$, $F1_{\text{avg}}$ value of RF are 85.01%, 83.99% and 84.25%, but its performance is still lower than that of the cloud classification model based on DL. In the traditional DL approach, the PRE, recall, and $F1$ values of the ResNet50 model trained using visible wavelengths are 86.91%, 86.06%, and 86.16%, respectively. However, the network trained using multiwavelength cloud image data shows further improvement in the cloud classification task, indicating that the spectral information reflected by multichannel cloud images is more helpful to improve the performance of the model in the cloud classification task. Cloud classification model based on DBRNN not only alleviates the problem of uneven distribution and small number of training samples through metric learning but also builds on the global features of the cloud images extracted by the backbone network, introducing local branch to extract the cloud semantic information, so as to effectively improve the performance of the model cloud classification. The PRE, recall, and $F1$ values reached 94.95%, 94.71%, and 94.76%, respectively. However, the introduction of GAN makes the model take longer to run. The base GAN model runs in nearly 15 min on the 3060 RTX GPU, while it cannot be computed on the CPU. And since there is no mask product as an aid and the image representations of low clouds and sea fog are inherently closer, the pseudo-cloud maps generated by the GAN model will

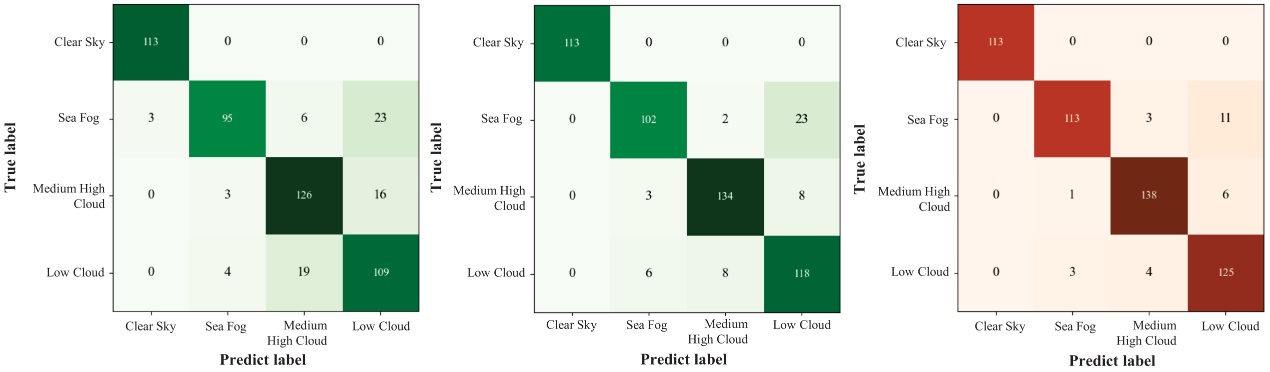


Fig. 8. Confusion matrices for deep learning models. (a) Resnet50(vis). (b) Resnet50(mc). (c) Model we proposed.

TABLE IV
SEA FOG RECOGNITION PERFORMANCE OF DIFFERENT MODEL CLOUD RETRIEVAL METHODS

Methods	PRE(%)	$recall(\%)$	$F1(\%)$
Resnet50(vis)	93.14	74.80	82.97
Resnet50(mc)	91.89	80.31	85.71
Proposed	**96.58**	**88.98**	**92.62**

The performance of our model is highlighted in bold.

TABLE V
MODEL PERFORMANCE OF DIFFERENT NETWORK FRAMEWORKS

Network Architecture	$\mathrm{PRE}_{\mathrm{avg}}(\%)$	$recall_{\mathrm{avg}}(\%)$	$F1_{\mathrm{avg}}(\%)$
Single branch	92.55	92.21	92.27
Double branch	**94.95**	**94.71**	**94.76**
Res-CBAM	92.14	91.73	91.80

The performance of our model is highlighted in bold.

produce greater ambiguity in the label definitions. This will lead to some degradation in model performance.

To visualize the performance of the proposed cloud classification model, the confusion matrix of cloud classification results from the traditional DL and the method we proposed based on the test dataset is shown in Fig. 8.

As can be seen from Fig. 8, for clear sky cloud images, all the three models can be correctly classified; for 127 sea fog samples, the Resnet50(vis) model, misclassified six cases of them as medium-high clouds, 23 cases as low-level clouds, and even misclassified three cases as clear-sky, Resnet50(mc) has improved cloud classification performance due to the use of multichannel information, and the samples misclassified as medium-high clouds were effectively reduced. However, 23 cases are misclassified as low-level clouds, which further expresses the difficulty in distinguishing sea fog from low-level clouds. Not only does the cloud classification model we proposed has much better discrimination ability than the Resnet50(vis) and Resnet50(mc) models between medium-high cloud and low-level cloud samples but also it has an excellent capacity to discriminate for sea fog.

C. Performance Analysis of DBRNN Model for Sea Fog Identification

In order to evaluate the effectiveness of the CIR method we proposed on SFR, we compare it with the CIR methods based on traditional DL models, and the results are shown in Table IV.

As can be seen from the Table IV, compared with the methods based on traditional DL models, the performance of the method we proposed on sea fog recognition is greatly improved, with PRE, recall, and $F1$ values of SFR have reached 96.58%, 88.98%, and 92.62%, respectively. This is mainly due to the fact that the DBRNN model not only utilizes the multichannel information of satellite cloud images but also effectively fuses the global and local features of satellite cloud images through the global branching and local branching structures. Moreover, the experimental results also show that it is feasible to achieve SFR through CIR.

D. Validity Analysis of DBRNN Model

In this article, we construct DBRNN to extract the features of the cloud image. The global branch of the DBRNN extracts the global features, and the local branch extracts the local features from the main semantic region of the cloud image through the AM, which makes use of the feature map of the global branch to enhance the network's perception of the main semantic objects. This mechanism is similar to the convolutional block attention module (CBAM). To analyze the effectiveness of the double branch structure and show the performance comparison between the AM mechanism and CBAM, in this article, we conduct a comparative experiment on cloud image classification with Resnet50 as the backbone network and use different network frameworks. The results are shown in Table V, where the Res-CBAM represents the single branch combined with CBAM structure.

As can be seen from the Table V, the model performance of the proposed double branch structure is substantially improved when applied to cloud class recognition, with an increase of 2.49% in F1 score, compared with the single branch network. Moreover, although CBAM increases the weight of the spatial or channel with critical information in the cloud image, the performance of the model compared to the basic single-branch network is not effectively improved. Therefore, the proposed AM mechanism performs better than CBAM in SFR.

In general, the feature maps obtained at each stage of the global branch of the DBRNN can be used for AM generation.

TABLE VI
MODEL PERFORMANCE OF DIFFERENT INTERMEDIATE FEATURE MAPS

Feature map stage	$\mathrm{PRE}_{\mathrm{avg}}(\%)$	$recall_{\mathrm{avg}}(\%)$	$F1_{\mathrm{avg}}(\%)$
stage 1	93.20	93.04	93.09
stage 2	93.76	93.55	93.59
stage 3	93.87	93.78	93.80
stage 4	**94.95**	**94.71**	**94.76**

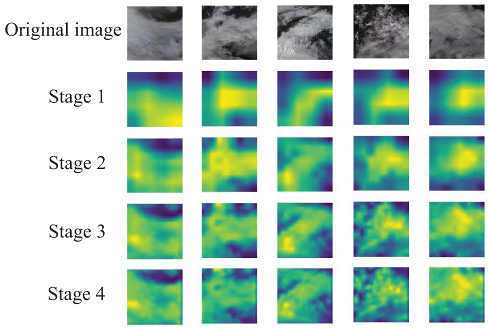


Fig. 9. Original image and its activation map in different stages.

Since the feature maps generated at different stages contain different information, in order to analyze the influence between the AM generated at different stages and the model performance, in this article, we use feature maps generated at different stages as AM to conduct a comparative experiment for cloud classifications, and the results are shown in Table VI.

As can be seen from the Table VI, the AM generated by the bottom feature map has the worst performance of the trained model. As the network deepens, the performance of the trained model also improves. The model trained on the AM generated by the top-level feature maps achieve the best performance with an F1 score of 94.76%. Therefore, the proposed model in this article uses the feature maps from stage 4 to generate the AM. To visualize the semantic information of the cloud image reflected by feature maps at different levels, the AM generated by feature maps at different stage is visualized in Fig. 9.

As can be seen in Fig. 9, the AM generated by the low-level feature maps is relatively clutter and mainly reflects the texture and edge features of the cloud image. As the network deepens, the activation regions reflected by the AM generated by the high-level network gradually gather and can represent the main semantic objects in the cloud image. Therefore, the high-level AM applied to the original cloud images will improve the model's ability to perceive the main cloud and fog regions.

VI. DISCUSSION

This article proposes a CIR-SFR. The method fully investigates the physical significance of each channel in the Himawari-8 satellite, combines the advantages of DL in feature extraction, and uses a novel CIR-SFR. This work has important implications for further work on image retrieval, image classification, and SFR.

One limitation of this article is that the model we created relies on historical satellite cloud image data for loss function optimization, and the limited number of historical satellite cloud image data can have an impact on the training of the model. This article introduces deep metric learning to mitigate this aspect and enable the model to present better recognition results.

Another limitation encountered in this article is that in the construction of the dataset. Since the labeling of the dataset depends on the cloud classification products of the Himawari-8 satellite, this will prevent the dataset from being accurately subclassified due to the problem of defining the cloud classification products. In future work, we will consider introducing other derivative products to further explore the precise classification of cloud products.

Furthermore, in order to achieve more accurate SFR, we will consider introducing other data that are important for SFR in our future work: SST data, regional latitude, and longitude data, etc.

Despite a few limitations, the article mitigates the impact of the limitation by using methods, such as deep metric learning, and achieves excellent results in the experiments, which is impossible in traditional methods. This approach can be applied in global sea fog image recognition and provides a new way of thinking for SFR.

VII. CONCLUSION

In this work, we construct the DBRNN to combine local features and global features by using three visible channels, three near-infrared channels and one far-infrared channel of the Himawari-8 satellite cloud images. Based on the backbone network, the proposed network introduces local branches and uses the AM to enhance the main cloud regions in the cloud image, which enables comprehensive extraction of global and local features of the cloud image. In addition, during network training, MS loss is introduced to map cloud features into semantic space, which can not only effectively alleviate the problem of insufficient number and uneven distribution of training samples but also improves the discriminative ability of the model for sea fog and low-level cloud. CIR-SFR not only deepens researchers' perceptual understanding of various types of clouds but also facilitates researchers to carry out further analysis of cloud images, intuitively enhancing the interpretability of the sea fog identification method. The experimental results show that the PRE, recall, and $F1$ value of the proposed SFR method reached 94.95%, 94.71% and 94.76% on the Yellow Sea and Bohai Sea sea fog dataset, respectively, which is better than the traditional SFR methods and provides a new idea to realize sea fog recognition.

REFERENCES

[1] J. Eyre, J. Brownscombe, and R. Allan, "Detection of fog at night using advanced high resolution radiometer (AVHRR) imagery," Meteorological Mag., vol. 113, pp. 66–271, 1984.
[2] Husi et al., "Development of a daytime cloud and haze detection algorithm for himawari-8 satellite measurements over central and eastern China," *J. Geophys. Res., D. Atmos.: JGR*, vol. 122, no. 6, pp. 3528–3543, 2017.
[3] Y. Deng, Y. Tian, and J. Wang, "Dynamic detection of daytime sea fog using geostationary meteorological satellite data," *Scientia Geographica Sinica*, vol. 36, pp. 1581–1587, 2016.
[4] S. Ceri, A. Bozzon, M. Brambilla, E. Della Valle, P. Fraternali, and S. Quarteroni, *An Introduction to Information Retrieval*. Berlin, Heidelberg, Germany: Springer, 2013, pp. 3–11.

[5] J. Lu, J. Hu, and J. Zhou, "Deep metric learning for visual understanding: An overview of recent advances," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 76–84, Nov. 2017.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25. Red Hook, NY, USA: Curran Assoc., Inc., 2012.

[7] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.

[8] K.-H. Chan, S.-K. Im, and W. Ke, "Vggrenet: A light-weight vggnet with reused convolutional set," in Proc. *IEEE/ACM 13th Int. Conf. Utility Cloud Comput.*, 2020, pp. 434–439.

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

[10] B. Zhang et al., "Progress and challenges in intelligent remote sensing satellite systems," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1814–1822, 2022.

[11] P. K. Buttar and M. K. Sachan, "Semantic segmentation of clouds in satellite images based on U-Net architecture and attention mechanism," *Expert Syst. Appl.*, vol. 209, 2022, Art. no. 118380.

[12] Z. Shao, Y. Pan, C. Diao, and J. Cai, "Cloud detection in remote sensing images based on multiscale features-convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 4062–4076, Jun. 2019.

[13] R. Yinze, H. Ma, Z. Liu, X. Wu, Y. Li, and H. Feng, "Satellite fog detection at dawn and dusk based on the deep learning algorithm under terrain-restriction," *Remote Sens.*, vol. 14, 2022, Art. no. 4328.

[14] Z. Chunyang, W. Jianhua, L. Shanwei, S. Hui, and X. yanfang, "Sea fog detection using U-net deep learning model based on MODIS data," in *Proc. IEEE 10th Workshop Hyperspectral Imag. Signal Process.: Evol. Remote Sens.*, 2019, pp. 1–5.

[15] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2006, vol. 2, pp. 1735–1742.

[16] M. Zhang, Q. Cheng, F. Luo, and L. Ye, "A triplet nonlocal neural network with dual-anchor triplet loss for high-resolution remote sensing image retrieval," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2711–2723, 2021.

[17] K. Sohn, "Improved deep metric learning with multi-class N-pair loss objective," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, Red Hook, NY, USA, 2016, pp. 1857–1865.

[18] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a "siamese" time delay neural network," in *Proc. 6th Int. Conf. Neural Inf. Process. Syst.*, San Francisco, CA, USA, 1993, pp. 737–744.

[19] C. B. Choy, J. Gwak, S. Savarese, and M. Chandraker, "Universal correspondence network," in *Proc. Int. Conf. Ad. Neural Inf. Process. Syst.*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29. Red Hook, NY, USA: Curran Assoc., Inc., 2016.

[20] X. Wang, X. Han, W. Huang, D. Dong, and M. R. Scott, "Multi-similarity loss with general pair weighting for deep metric learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5017–5025.

[21] F. Dell'Acqua and P. Gamba, "Query-by-shape in meteorological image archives using the point diffusion technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 9, pp. 1834–1843, Sep. 2001.

[22] M. K. Gurve and J. Sarup, "Satellite cloud image processing and information retrieval system," in *Proc. IEEE World Congr. Inf. Commun. Technol.*, 2012, pp. 292–296.

[23] S. Roy, E. Sangineto, B. Demir, and N. Sebe, "Metric-learning-based deep hashing network for content-based retrieval of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 2, pp. 226–230, Feb. 2021.

[24] Y. Liu, L. Ding, C. Chen, and Y. Liu, "Similarity-based unsupervised deep transfer learning for remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 7872–7889, Nov. 2020.

[25] G. Mie, "Beiträge zur optik trüber medien, speziell kolloidaler metallösungen," *Annalen der Physik*, vol. 330, no. 3, pp. 377–445, 1908.

[26] Z. Hao, D. Pan, F. Hong, and Q. Zhu, "Optical radiance characteristics of sea fog based on remote sensing," *Acta Optica Sinica*, vol. 28, pp. 2420–2426, 2008.

[27] R. Yedida, S. Saha, and T. Prashanth, "Lipschitzlr: Using theoretically computed adaptive learning rates for fast convergence," *Appl. Intell.*, vol. 51, no. 3, pp. 1460–1478, Mar. 2021.

[28] L. Liao, Z. Li, and S. Zhang, "Image retrieval method based on deep residual network and iterative quantization hashing," *J. Comput. Appl.*, vol. 42, pp. 2845–2852, 2022.

[29] K. Lin, H.-F. Yang, J.-H. Hsiao, and C.-S. Chen, "Deep learning of binary hash codes for fast image retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2015, pp. 27–35.

[30] H. Liu, R. Wang, S. Shan, and X. Chen, "Deep supervised hashing for fast image retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2064–2072.

[31] L. van der Maaten and G. Hinton, "Visualizing high-dimensional data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.

[32] T. K. Ho, "Random decision forests," in *Proc. 3rd Int. Conf. Document Anal. Recognit.*, 1995, vol. 1, pp. 278–282.

[33] J. Cramer, "The origins of logistic regression," Tinbergen Inst., Amsterdam, Netherlands, Tinbergen Institute Discussion Papers 02-119/4, Dec. 2002. [Online]. Available: https://ideas.repec.org/p/tin/wpaper/20020119.html

[34] S. Suthaharan, *Support Vector Mach.* Boston, MA, USA: Springer US, 2016, pp. 207–235.

[35] I. J. Goodfellow et al., "Generative adversarial nets," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, Cambridge, MA, USA, 2014, vol. 2, pp. 2672–2680.

Tianjiao Hu received the bachelor's degree in management from Zhejiang Agriculture and Forestry University, Zhejiang, China, in 2021. She is currently working toward the master's degree in computer technology with the Faculty of Electrical Engineering and Computer Science, Ningbo University.

Her research interests include deep learning, saliency detection and its application in remote sensing image recognition.

Zhuzhang Jin received the master's degree in computer science and technology from Ningbo University, Ningbo, China, in 2022.

He is currently working in computer engineering. His research interests include deep learning, saliency detection and its application in remote sensing target detection.

Wanxin Yao is currently working toward the B.S. degree in computer science and technology with a focus on network technology and engineering from Ningbo University, Ningbo, China.

Her research interests include image classification, image recognition, and deep learning.

Jiezhi Lv received the bachelor's degree in computer science and technology from Jinan University, Jinan, China, in 2022. He is currently working toward the master's degree in computer science and technology with the Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo, China.

His research interests include deep learning, image retrieval and its application in remote sensing image retrieval.

Wei Jin received the Ph.D. degree in optical engineering from Chongqing University, Chongqing, China, in 2006.

He is currently a Professor with the Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo, China. His research interests include sparse representation, deep learning, computer vision, and image processing.