A Multiscale Dual Attention Network for the Automatic Classification of Polar Sea Ice and Open Water Based on Sentinel-1 SAR Images

Zheng Zhang, Guangbo Deng, Chuyao Luo, Xutao Li, Yunming Ye, and Di Xian

Abstract—**Automatic classification of sea ice and open water plays a vital role in climate change research, polar shipping, and other applications. Many deep-learning-based methods are proposed to automatically classify sea ice and open water to address this issue. Even though these methods have achieved remarkable success, the noise phenomenon in synthetic aperture radar (SAR) images still causes considerable limitations in the model performance. Meanwhile, these existing methods ignore multiscale global information from large-scale SAR images, which tends to produce misclassification. In this article, we propose a novel multiscale dual attention network (MSDA-Net) for the task. To tackle the first drawback, we introduce the information of relative position and high-pass filtering as two extra channels to reduce the noisy effects. Moreover, we propose a patch dual attention mechanism and embed it into the ConvNeXt blocks to capture the multichannel and spatial features. To address the second problem, we propose a multiscale spatial attention module to capture multiscale global spatial information. The experiments show that the proposed method significantly outperforms state-of-the-art methods. In addition, comprehensive case studies are conducted, which verify the effectiveness of MSDA-Net in different SAR scenes.**

Index Terms—**Deep learning, sea ice classification, synthetic aperture radar (SAR).**

I. Introduction

Automatic classification of sea ice and open water is vital in climate change research, polar shipping, navigation in polar regions, and marine operations. The global warming causes the reduction of the amount of polar sea ice. As an essential indicator of global climate change [1], researchers paid more attention to sea ice change. Besides, large areas of sea ice melting, breaking up and drifting in summer raise unpredictable risks for Arctic shipping and transport safety [2]. A growing interest in polar sea ice classification has been shown to provide valuable information for sea ice change and safe navigation in the Arctic.

To monitor the constantly moving and changing sea ice conditions in time, satellite remote sensing technologies are used to collect data. With the launch of the Nimbus-5 Electrically Scanning Microwave Radiometer (EMSR), humanity began to use passive microwave remote sensing to obtain complete observations of sea ice coverage at both poles. Since then, the increasing launch of sensors (such as Nimbus-7 scanning multichannel microwave radiometer [3], special Sensor microwave imager [4], advanced microwave scanning radiometer for the Earth observing system [5], microwave radiation imager [6], and advanced microwave scanning radiometer 2 [7]), led to the emergence of various sea ice classification algorithms. With the improvement of observational capabilities of the North and South polar sea ice coverage, on-board high spatial resolution optical-infrared sensors [8], [9] and synthetic aperture radar (SAR) [10] provide effective means of acquiring regional sea ice information. Unlike optical-infrared sensors, SAR, with high spatial resolution and comprehensive coverage, is unaffected by polar nights and can penetrate the clouds. Benefit from the advantages of SAR, it has been extensively used for sea ice monitoring and classification.

Based on SAR images [11], researchers devote themselves to studying algorithms of sea ice classification. Classical methods are based on machine learning (ML) methods, including threshold-based methods [12] and expert systems [13], neural network (NN) [14], [15], support vector machine (SVM) [16], [17], random forest [18], and Markov random fields [19]. By learning from massive datasets, they can build mappings from sea ice features to sea ice categories. However, the performances of these methods depend on prior expert knowledge and sophisticated manual engineering to build model features. In other words, these methods take extensive labor cost while having poor generalization capability [20], which hinders their applications in real world.

Recently, deep learning techniques shed some light on overcoming the limitation, which directly constructs the mapping from raw data to classification results. In this kind of method, sea ice classification on SAR images can be defined as a pixel-level segmentation problem. Initially, some fully connected neural network (FC NN)-based methods are introduced to classify sea ice and open water [21], [22], [23], [24], [25], [26]. However,

Manuscript received 12 October 2023; revised 25 December 2023; accepted 11 January 2024. Date of publication 18 January 2024; date of current version 4 March 2024. This work was supported in part by Shenzhen Science and Technology Program, China, under Grant JCYJ20210324120208022 and Grant KCXFZ20211020163403005. *(Zheng Zhang and Guangbo Deng contributed equally to this work.) (Corresponding author: Di Xian.)*

Zheng Zhang, Guangbo Deng, Chuyao Luo, Xutao Li, and Yunming Ye are with the Department of Computer Science and Shenzhen Key Laboratory of Internet Information Collaboration, Harbin Institute of Technology, Shenzhen 518055, China, and also with the Shenzhen Key Laboratory of Internet Information Collaboration, Harbin Institute of Technology, Shenzhen 518055, China (e-mail: 37zhangz@gmail.com; dengguangbo05@gmail.com; luochuyao.dalian@gmail.com; lixutao@hit.edu.cn; yym@hit.edu.cn).

Di Xian is with the National Satellite Meteorological Center, China Meteorological Administration, Beijing 100081, China (e-mail: xiandi@cma.gov.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3354912

the FC NN-based method is poor in extracting the local spatial features. To solve this challenge, Wang et al. [27] proposed a U-Net-based methods to exploit spatial representations on different scales. They provide a more precise classification between sea ice and open water compared to FC NN. To improve the spatial representations ability further, DAU-Net [28] combines the advantages of residual block [29] and the self-attention mechanism to achieve excellent performance.

With the development of deep learning, some of the latest semantic segmentation methods achieve better performance in pixel-level segmentation. DeepLab is a fully convolutional network (FCN) developed and constantly updated by Google that has shown good performance in pixel-level segmentation [30], [31]. Due to the locality of convolution operation, convolutional neural network (CNN)-based methods are usually challenging to learn global and long-range semantic correlations. To leverage such correlations, some transformer-based [32] methods are introduced to the computer vision field [33], [34]. For example, Swin-UNet has achieved great success in medical image segmentation by hierarchical Swin transformer with shifted windows [35], while Segformer [36] uses fewer parameters to get better performance. In addition, some methods combine the advantages of Transformer and the CNN network. ConvNeXt [37] designed the CNN network according to the architecture of Swin-Transformer and achieved better classification performance. Visual attention network (VAN) [38] uses the large kernel attention mechanism and achieves adaptivity not only in the spatial dimension but also in the channel dimension.

However, the existing methods cannot effectively solve the problem of sea ice segmentation due to the difference between the remote sensing images and regular images. This task faces several challenging issues.

1) Sentinel-1 SAR images are often affected by noise interference, which deteriorates image clarity and contrast, posing challenges to the segmentation of sea ice and limiting model performance. Specifically, noise speckles, appearing as fine-grained vertical lines or point-like noise, are prominent in Sentinel-1 SAR images. Although several preprocessing methods [39], [40] have been proposed to eliminate vertical lines in the HV channel, these approaches rely on noise area identification, introducing significant time complexity and hindering automatic and real-time detection.
2) Mainstream segmentation methods encounter difficulties in effectively handling sea ice objects of different scales. Sea ice exhibits a wide range of sizes, from small blocks to large formations, posing challenges for segmentation tasks. Conventional semantic segmentation methods primarily focus on segmenting larger objects, often labeling smaller sea ice objects as noise or background, resulting in inaccurate segmentation outcomes. Conversely, techniques specialized in small object segmentation may not adequately capture the intricate characteristics of large-scale sea ice. Consequently, existing segmentation methods may fail to comprehensively express and represent the complex features of sea ice with different scales, leading to less precise segmentation results.

In this article, we propose a multiscale dual attention network (MSDA-Net) for automatic polar sea ice classification to address these two problems. First, we introduce the information of relative position and high-pass filtering as two extra channels to reduce the influence of noise. The former can help models classify those regions with noise. The latter can preserve the appearance of sea ice and reduce the disturbance by noise. Moreover, we propose a patch dual attention mechanism (PDAM) and embed it into the ConvNeXt blocks to capture the multichannel and spatial features. PDAM-ConvNeXt blocks can enrich the representation of relative position and high-pass filtering channels to focus on the noisy position and spatial correlation. Second, we propose a multiscale spatial attention (MSSA) module to extract global features from SAR images. It utilizes the advantage of the transformer [32] to capture multiscale global spatial information.

In brief, we summarize our main contributions as follows.

1) We propose a method called MSDA-Net, which demonstrates good robustness and the ability to capture multiscale spatial information in the classification of sea ice and open water areas, achieving excellent classification results. The MSDA-Net excels in robust edge detail detection, reduces misclassification, and demonstrates excellent performance in detecting fragmented ice in the region.
2) Our research introduces two key modules, MSSA and PDAM, which significantly improve the classification of sea ice and open water, particularly in fragmented ice detection and noisy conditions, respectively.
3) We conducted extensive experiments and ablation studies to verify the effectiveness of our proposed method. The experimental results clearly show that our method is significantly superior to the state-of-the-art methods. These experimental results further confirm the feasibility and superiority of the proposed key modules.
4) In addition, we created a novel dataset consisting of high-quality samples with precise labels for sea ice and open water classification. We employed a method to prevent duplicate samples and appropriately divided the sample proportions to ensure the robustness of all models. The dataset is accessible at https://doi.org/10.5281/zenodo.7260842.

The rest of this article is organized as follows. The related work about sea ice classification with SAR satellite images is introduced in Section II. The proposed method is presented in Section III. The dataset preparation, comprehensive experiments, and analysis are presented in Section IV. Finally, Section V concludes this article.

II. Related Work

The automatic sea ice classification methods can be divided into classical methods, namely ML methods and deep learning methods.

A. Classical Method

The earliest sea ice automatic segmentation methods are based on threshold, probabilistic, and statistical methods. These methods can be traced back to 1986 [11]. They mainly focus on extracting texture from the image, and then, detecting sea ice by lookup tables [41] or Bayesian classification algorithm [42], [43], [44]. Moreover, a sea ice classification framework based on a projection matrix is proposed to preserve spatial localities of features from multisource images such as SAR and multispectral images [45]. Based on these classical methods, the introduction of incidence-angle dependent intensity decay rates achieves better segmentation results [46]. However, the adjacent pixels of SAR scenes change randomly in radar echo signals due to coherence. So, these threshold based methods are not applicable to sea ice classification.

B. ML Method

ML models usually require labeled datasets for supervised training. However, there was no representative dataset, and only several scenes were used in the early stage. Some ML methods such as wavelet transforms [47], [48], Bayes classifier [19], and Maximum likelihood [49] are introduced to classify sea ice and open water. With the completeness of dataset, methods based on NN and SVM further improve the accuracy of classification. As important features of sea ice, texture, incidence angle, and polarization features are used extensively in NN [50], [51] and SVM [52], [53]. However, prior expert knowledge and complex manual engineering limitations are not suitable for processing polar sea ice data in large volumes.

C. Deep Learning Method

To address the aforementioned problem, some deep-learning-based methods are introduced to classify sea ice and open water. Sea ice classification can be regarded as a semantic segmentation task, which means pixel-level classification. The FCN [54] is the first work to solve the semantic segmentation task. U-Net [55] is the most representative of the FCN-based segmentation model. U-Net consists of a shrinkage path and a symmetric expansion path. The shrinkage path is used to obtain contextual information, and the symmetric expansion path is used to precisely locate the segmentation boundary. Based on U-Net, the symmetric U-shaped structure is also widely used in sea ice classification benefitted from its applicability [24], [27].

The FCN-based semantic segmentation model is coarse and ignores the spatial consistency relationship between pixels. To overcome the drawback, Google proposes the series that introduces Atrous spatial pyramid pooling (ASPP), which extracts features by using multiple dilated convolutions with different sampling rates to capture contextual information with different scales [30], [31]. By introducing the ASPP, the model further improves the accuracy of sea ice classification [56].

Since the classification of sea ice and open water can be regarded as a semantic segmentation problem, here we review some segmentation methods. Recently, the vision transformer (ViT) [33] attracted much attention due to its superior performance. Different from CNN-based methods, this model with the self-attention mechanism has a larger receptive field. Therefore, it can capture more spatial information and generate better results. However, the high computational cost hinders their application in semantic segmentation on a large scale. To address this problem, researchers proposed Swin-Transformer [57] and Swin-UNet [35], which achieved great success in the field of image segmentation by using a U-shaped architecture and convolutional downsampling. Besides, Segformer [36] has designed a more lightweight network structure. An efficient transformer block and a simple multilayer perceptron (MLP) segmentation reduce computational cost significantly.

Different from existing methods, the proposed MSDA-Net combines the advantages of the CNN and transformer networks to extract local and multiscale global spatial features of Sentinel 1 SAR images. On the one hand, to improve the local representation ability of the model, we proposed a PDAM based on ConvNeXt [37] blocks. It preserves local spatial information by splitting the original feature map into independent parts, and then, extracts their spatial representation by the dual attention mechanism. On the other hand, to extract multiscale global spatial information, we proposed an MSSA. It captures the multiscale global spatial semantic information by the transformer with different window sizes. Benefiting from the PDAM and MSSA, the model achieves effective improvement and significantly outperforms the state-of-the-art methods.

III. Proposed Methods

A. Problem Definition

The classification of sea ice and open water can be defined as a semantic segmentation problem. It can be described as follows. Given an image X with N channels, it aims to predict a matrix $\hat{X}$ with two channels, where these two channels denote the probability of each pixel being detected as sea ice and open water, respectively.

B. Overall Architecture

In this article, we propose an MSDA-Net to classify sea ice and open water. The overall architecture of MSDA-Net is presented in Fig. 1. The MSDA-Net consists of encoder, bottleneck, and decoder. The encoder has a convolution block for feature exaction, three downsamples, and three PDAM-ConvNeXt layers where each PDAM-ConvNeXt layer containstwo PDAM-ConvNeXt blocks. The shape of inputs is $B \times N \times 512 \times 512$, where B denotes the batch size. As the output of encoder, the spatial representation with the shape $B \times 384 \times 64\times$ 64 is extracted. The bottleneck is compose of two PDAM-ConvNeXt layers and an MSSA module. In the bottleneck part, the shape of the feature map is invariable, which is also $B \times 384 \times 64 \times 64$. The decoder comprises of two DUpsample layers, a PDAM-ConvNeXt layer, and a classification layer, which is a convolution layer to convert the channel to the categories counts. The shape of the output is $B \times 2 \times 512 \times 512$.

Besides, MSDA-Net also contains a skip connection between the encoder and the decoder. The extracted context features are fused with multiscale features from the encoder via skip connections to complement the loss of spatial information caused by downsampling. It contributes to reducing the loss of spatial information caused by downsampling because the skip connections can fuse the extracted features in shallow layer where the consumption of spatial representation is less.

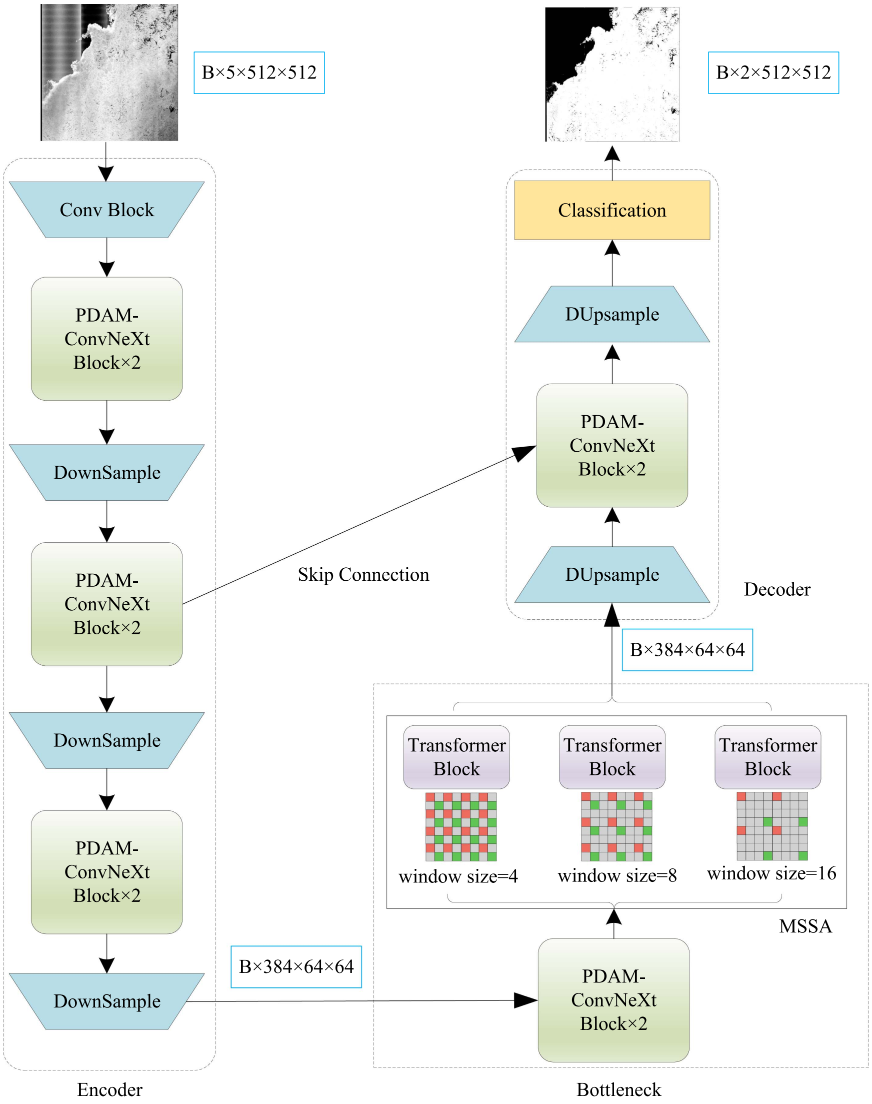


Fig. 1. Structure of MSDA-Net.

C. Patch Dual Attention Mechanism

To extract channel and spatial feature of feature map, we propose a PDAM. The structure of the PDAM is shown in Fig. 2. Given the input feature map $F_i \in R^{B\times C\times H\times W}$, where $B, C, H,$ and W denote the batch size, channel, height, and width of F_i. According to Fig. 2(a), the original input F_i is split into 4×4 overlapping patches, and then, they stack together to generate a new input $F_s \in R^{B\times 16C\times \frac{H}{4}\times \frac{W}{4}}$. The output F' of the PDAM can be obtained according to the following formula:

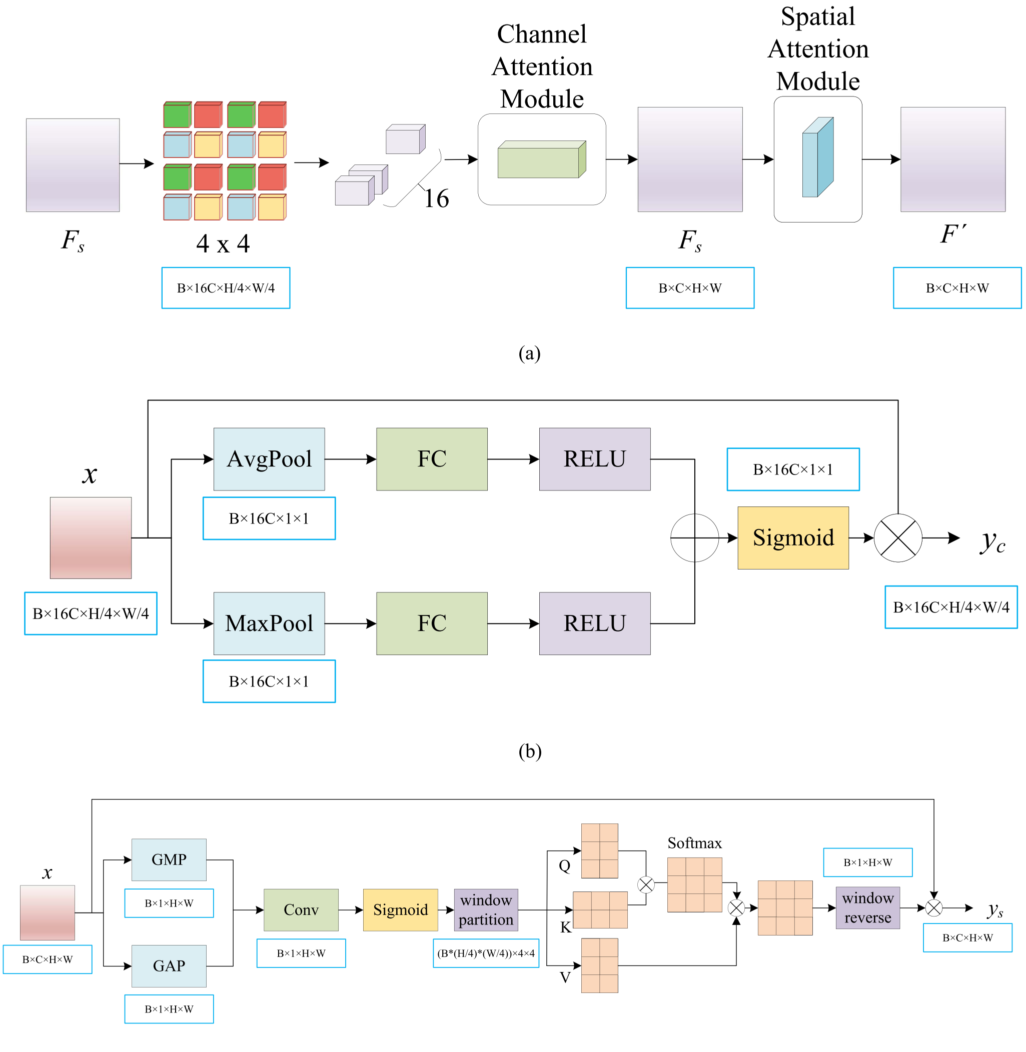
$$F' = \text{SAM}(\text{CAM}(F_s)) \tag{1}$$

where the CAM, SAM, and $F' \in R^{B\times C\times H\times W}$ denote the channel attention module, spatial attention module, and the output, respectively. Here, the CAM and SAM extract channel and spatial features of F effectively, respectively.

As Fig. 2(b) shows, the introduction of CAM is conducted to capture channel information of the feature map. Given the input feature map $x \in R^{B\times 16C\times \frac{H}{4}\times \frac{W}{4}}$, the output y_c of CAM is shown as

$$y_c = x \otimes \text{Sigmoid}(\text{AAP}(x) + \text{AMP}(x)) \tag{2}$$

where the AAP denotes the adaptive average pooling with FC and RELU. The AMP represents the adaptive max pooling connected to the FC and RELU, while the Sigmoid is the activation function. The specific structure is shown in Fig. 2(b). $\otimes$, FC, and RELU are the multiply operation, fully connected layers, and the activation function, respectively. In CAM, we partition the feature map into 4×4 small blocks to obtain enhanced local information, facilitating a more detailed analysis of image details and enabling the extraction of additional features specifically relevant to detecting small fragmented ice. the kernel size of pooling is set to $[\frac{H}{4}, \frac{W}{4}]$ (the shape is same to feature map F_s) to only preserve the channel information. Therefore, by adding the outputs of AAP and AMP, the shape of result after activation function is converted to $B \times 16C \times 1 \times 1$. By multiplying to original input x, the important channel of x are enhanced further,



(a)

(b)

(c)

Fig. 2. Details of the PDAM. (a) Structure of the PDAM. (b) Structure of the channel attention mechanism. (c) Structure of the spatial attention mechanism.

while the output of CAM with the shape $B \times 16C \times \frac{H}{4} \times \frac{W}{4}$ is generated.

The SAM is introduced to extract the spatial information of the feature map in Fig. 2(c). Given the input feature map $x \in R^{B\times C\times H\times W}$, the output $y_s \in R^{B\times C\times H\times W}$ of SAM can be formulated as

$$y_s = x \otimes \mathrm{SA}(\mathrm{Sigmoid}(\mathrm{Conv}(\mathrm{Concat}[\mathrm{GAP}(\mathrm{x}), \mathrm{GMP}(\mathrm{x})]))) \tag{3}$$

where the SA, GMP, GAP, Concat, and Conv denote the self-attention module, global average pooling, global max pooling, concatenate operation, and the convolution, respectively. In SAM, the GMP and GAP are introduced to preserve the spatial information by calculating the mean and the max value of all pixels in each channel. Then, we obtain two feature maps with the shape $B \times 1 \times H \times W$. Then, these two feature maps are concatenated and the shape is convoluted to $B \times 1 \times H \times W$. After activation function, the shape of the feature map is converted to $B \times 1 \times H \times W$. Here, the SA is introduced to enhance additional important spatial information. $y_s' \in R^{B\times 1\times H\times W}$ is split into $\frac{H}{4} \times \frac{W}{4}$ overlapping patches by window partition, and then, they stack together to generate a new input $f \in R^{(B*\frac{H}{4}*\frac{W}{4})\times 4\times 4}$. Here, the window partition is based on matrix transformation, which can adapt the tensors to any shapes we need and the window reverse is the inverse operation of window partition. f is used as query, key, and value, and the SA can be calculated as follows:

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d}}\right) \cdot V \tag{4}$$

where $Q, K, V \in R^{(B*\frac{H}{4}*\frac{W}{4})\times 4\times 4}$ represent the query, key, and value matrices. In SA, the shapes of the input and output are invariable. The window reverse module transforms the shape $(B * \frac{H}{4} * \frac{W}{4}) \times 4 \times 4$ to $B \times 1 \times H \times W$ after SA. By multiplying to original input x, important pixels of x are enhanced further, while the output of SAM y_s with the shape $B \times C \times H \times W$ is generated in (3).

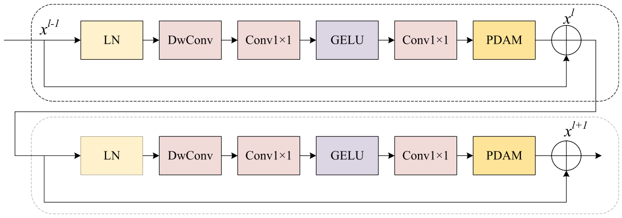


Fig. 3. Structure of two consecutive PDAM-ConvNeXt Blocks.

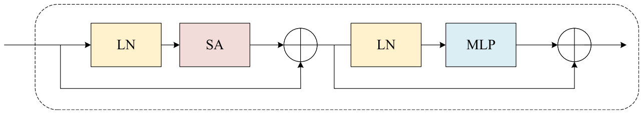


Fig. 4. Structure of the transformer block.

As a kind of attention mechanism, the PDAM can be introduced to any CNN backbone. Due to the promising performance of ConvNeXt [37] for feature extraction, we embed PDAM into the end of ConvNeXt block and name the new block to PDAM-ConvNeXt in this article. Fig. 3 shows the structure of two consecutive PDAM-ConvNeXt blocks, which is denoted by green block in Fig. 1. Each PDAM-ConvNeXt block includes a LayerNorm (LN), a depthwise convolution module (DwConv) [58], two convolution modules with 1×1 kernel size, and a PDAM module. Besides, it applies the GELU activation function and residual connection [29]. Formally, given input x^{l-1}, the output of two continuous PDAM-ConvNeXt blocks can be formulated as

$$\begin{aligned} x^{l} &= x^{l-1} + \text{PDAM}(\text{Conv}(\text{Conv}(\text{DwConv}(x^{l-1})))) \\ x^{l+1} &= x^{l} + \text{PDAM}(\text{Conv}(\text{Conv}(\text{DwConv}(x^{l})))) \end{aligned} \tag{5}$$

where x^l represents the output of the lth block. By embedding the PDAM, the ability to capture channel and spatial information is significantly improved.

D. Multiscale Spatial Attention

To capture more multiscale global and long-range semantic information interaction of SAR images, we propose an MSSA module. It combines the advantages of the transformer [32] and ASPP [30], [31] to capture the global spatial information and extract multiscale context representation, respectively. MSSA is composed of three transformer blocks that extract global spatial features with different scales. According to window partition with different window sizes, the feature map is split into patches with different shapes in Fig. 1. Then, the patched feature map is fed into transformer blocks with the same structure, which is shown in Fig. 4. Here, LN, SA, and MLP denote the LayerNorm, self-attention module, and MLP module, respectively.

E. Downsample and Upsample Operators

To reduce the computational cost and enlarge the receptive field, a 2×2 convolution is applied. Besides, upsample layers apply DUpsampling [59] module whose structure is shown in Fig. 5. Different from bilinear interpolation and transposed convolution upsampling methods, it not only considers the correlation between each pixel but also has high computational efficiency. Specifically, given a scale factor s, an output channel C' and a feature map $F \in R^{C\times H\times W}$, where C, H, and W denote the channel, height, and width of F, respectively. First, convolution converts the shape of F to $H \times W \times N$, where N is equal to $s \times s \times C'$. Then, the shape of every new tensor is converted to $s \times s \times \frac{N}{s\times s}$. After that, the upsampled feature map with the shape $(s \times H) \times (s \times W) \times \frac{N}{s\times s}$ is generated.

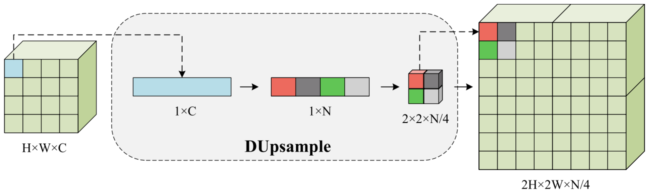


Fig. 5. Structure of DUpsample.

IV. Experiment

A. Dataset

1) Sentinel-1 Images: The data[1] used in this article are derived from SAR scenes from the Sentinel-1 satellite. This satellite works in the C-Band, which has been verified to perform well for discriminating between sea ice and open water due to its high spatial resolution [60], [61]. Here, medium-resolution level 1 extra-wide (EW) ground range detected scenes (GRDM) have been frequently chosen because scenes with this type are the most favorable combination of extensive coverage and high spatial resolution in marine polar regions. Each scene contains up to 10000×10000 pixels. The spatial resolution of each scene is 40 m × 40 m.

We collected 101 representative SAR scenes from Sentinel-1 A and B sensor data. These scenes covered the period from August 10 to August 15 in 2021. In Sentinel-1 SAR, HH- (horizontal copolarization) and HV-polarized (cross-polarization) data mean the electromagnetic waves are received from horizontal and vertical directions, respectively. Due to the sensitivity of ice-induced volume scattering in the cross-polarization channel, the radar backscatter of sea ice is generally higher than that of open water. Hence, the HV-polarization channel was usually used to generate Arctic sea ice cover products. Moreover, the difference between copolarization and cross-polarization data has proven to be an optimal combination for distinguishing between sea ice and open water [62], [63]. Therefore, in addition to incorporating HV-polarized data directly, we employ polarization difference (HH − HV) and polarization ratio (HH/HV) data to conduct our data labeling and training.

2) Data Preprocessing: Due to the persistent speckle noise and thermal noise-induced subswath transitions in Sentinel SAR

[1] The original Sentinel-1 SAR images can be downloaded from the website of the Alaska Satellite Facility (https://search.asf.alaska.edu/).

scenes, some long-grained vertical lines greatly influence data labeling and model inference.

Park et al. [39] developed an effective method for reducing noise in Sentinel-1 GRD products by calculating the average values of factors from five segmented azimuthal blocks and tuning the noise vectors provided by the European Space Agency (ESA) using empirically determined correction coefficients. Sun and Li [40] further improved the denoising method by segmenting the image into a greater number of azimuthal blocks and introducing a variance factor to distinguish between homogeneous and inhomogeneous blocks. This resulted in more precise scaling and balancing factors based on the homogeneity of local regions within each subswath. As denoising efficacy increases, the computational resources and time required for preprocessing increase exponentially. However, even with extensive preprocessing and denoising methods, it is impossible to completely eliminate noise interference.

First, Sentinel Application Platform (SNAP) 3.0 software[2] [64] is usually used to preprocess SAR images, which is also used here. Specifically, we leverage SNAP 3.0 to remove GRD border noise, reduce thermal noise, filter out speckle noise, and perform radiometric calibration on the aforementioned SAR images. Second, the processed images of HV, HH − HV, and HH/HV channels can be described by the backscattering coefficient, and the formula is shown as

$$\sigma_0(\mathrm{dB}) = 10 \times \log_{10}\sigma_0 \tag{6}$$

where the σ_0 and $\sigma_0(\mathrm{dB})$ denote the backscattering coefficient and its log-transformed result, respectively. Third, we normalize them to the range of 0–255 while all scenes are resized to 5120 × 5120. Finally, we use a land and sea mask to filter out the land in each scene so that only the open water and sea ice regions are preserved.

Since the pixel values in regions of sea ice are different in various channels, sea ice can be distinguished from open water by the threshold constraint of the three channels. Based on this property, we created an annotation tool for labeling sea ice and open water. Sea ice and open water are denoted as 0 and 1, respectively.

Moreover, since the SAR images with 5120 × 5120 pixels are also too costly for computation, it is necessary to crop them into smaller patches. In order to prevent image cutting loss, the edge part of the image is filled with zero. Each image is divided into 121 patches with 512 × 512 pixels chronologically in the same cutting order. This operation avoids overlapping images in the dataset so that each part of the original image appears only once in each training iteration.

Finally, we divide the training set, validation set, and test set in the ratio of 6:1:3 in chronological and crop order. The number of samples in test dataset is large because more test sets can better test the robustness of the model. Finally, we divide 7381 images as the training set, 1210 images as the validation set, and 3630 images as the test set.

3) Data Denoising: The HV, HH − HV, and HH/HV images are chosen to be the inputs. The scanning technique TOPSAR of Sentinel-1 SAR usually causes persistent speckle noise and thermal noise-induced subswath transitions. These noises are easily to be misclassified as sea ice due to their high pixel values. To reduce the impact of these noises, we introduce two additional channels, high-pass filter (HPF) and position encoding (PE) channels.

The HPF channel is generated from the HV channel in Fig. 6(c). Intuitively, the regions that change drastically belong to the high-pass component, while the regions that change slowly belong to the low-pass component. We can see that most of noise areas are in the SAR scene and belong to low-pass areas. We remove the low-frequency information and keep the high-frequency information by using high-pass filtering. Based on the aforementioned concepts, the next step is to describe the process of obtaining the HPF channel, which involves the following steps.

1) *Fourier transform:* Convert the input image to the frequency domain using fast Fourier transform for frequency analysis and processing.
2) *Spectrum centering:* Shift the zero-frequency component (dc component) to the center of the spectrum for better visualization and analysis.
3) *Amplitude spectrum calculation:* Compute the amplitude spectrum of the frequency-domain image by taking the absolute value and applying a logarithmic function to map the values to the range of 0–255 for enhanced visualization.
4) *HPF matrix generation:* Generate a transformation matrix for the HPF based on a specified threshold to selectively attenuate low-frequency components.
5) *Application of transformation matrix:* Multiply the frequency-domain image by the HPF matrix to obtain an enhanced image with emphasized high-frequency information.
6) *Inverse Fourier transform:* Perform inverse Fourier Transform to convert the frequency-domain image back to the spatial domain.

In this article, the threshold of the HPF is 30. Fig. 6(a), (c), and (e) shows the original image, the result after high-pass filtering, and the label. From these figures, we find that most of the long-grained vertical lines are filtered out. However, even if the high-pass filtering filters out most of the noisy areas, the image becomes rough and part of the ices areas are also filtered out. We use this result as a new extra input.

Moreover, the PE channel is added to preserve the relative positions before cropping. To record the position of each pixel, the original image is encoded as follows:

$$x_p = i/n \tag{7}$$

where x_p, i, and n denote the value of each pixel in the PE channel, the order of image crops, and the total number of crops, respectively. As the noise position in original image (5120 × 5120) is fixed, cropped images preserve the location of the noise. Once positions of noise are obtained, the model

[2] This software can be downloaded from the website (http://step.esa.int/main/download/snapdownload/)

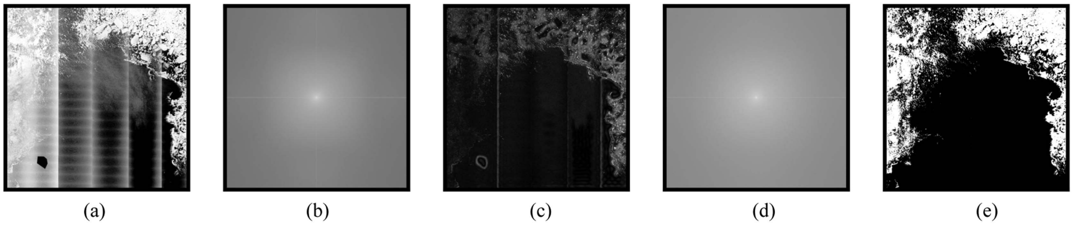


Fig. 6. (a) HV channel. (b) Spectrogram of the HV channel. (c) HPF channel. (d) Spectrogram of the HPF channel. (e) Label (the white part denotes sea ice).

can adaptively learn how to segment the sea ice in those regions with noise.

B. Evaluation Metrics

In this article, to evaluate the experimental results of the sea ice classification, several common measurements based on the confusion matrix are used. According to the segmentation result and ground truth, the true positives (TP, the number of pixels that segmentation and ground truth are both sea ice), true negatives (TN, the number of pixels that segmentation and ground truth are both open water), false positives (FP, the number of pixels that segmentation is sea ice and the ground truth is open water), and false negatives (FN, the number of pixels that segmentation is open water and the ground truth is sea ice) can be counted. Based on these indices (TP, TN, FP, and FN), the mean intersection over union (MIoU), accuracy (Acc), Jaccard index (Jaccard), frequency weighted intersection over union (FWIoU), precision, recall, and F1-score can be calculated by the following formula:

$$\begin{aligned}
\mathrm{MIoU} &= \frac{1}{k}\sum_{i=1}^{k}\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}\\
\mathrm{Acc} &= \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}\\
\mathrm{Jaccard} &= \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}\\
\mathrm{FWIoU} &= \frac{\mathrm{TP}+\mathrm{FN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}\times\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}\\
\mathrm{Precision} &= \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}\\
\mathrm{Recall} &= \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}\\
F1-\mathrm{Score} &= \frac{2\times\mathrm{Precision}\times\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}} \qquad (8)
\end{aligned}$$

where the k denotes the kth category. MIoU denotes the average value of each class after calculating their ratio of the intersection and union. The Acc is the proportion of correctly classified pixels in all pixels. The Jaccard is used to compare the similarities and differences between two samples. FWIoU is the ratio of intersection and concatenation weighted by the frequency of occurrence of each class. The F1-score is a powerful evaluation metric used to determine the harmonic mean of the precision and recall metrics. The higher values of these evaluation metrics indicate the better segmentation performance.

C. Parameter Setting

The MSDA-Net is implemented based on Python 3.8 and Pytorch 1.7.0. For all training cases, data augmentations such as flips and rotations are used to increase data diversity. The input image size is set as 512×512. We train all the models on two NVIDIA 3090 GPUs. During the training period, the batch size is 8, and the learning rate is 0.001. The seed is set to 42, and an AdamW [68] optimizer with a momentum of 0.9 is used to optimize all models for backpropagation.

D. Baseline Methods

We compare the proposed method MSDA-Net with the following existing methods.

1) *UNet [55]:* UNet is a U-shape structure NN for semantic segmentation.
2) *ResUNet [29]:* ResUnet is a U-shaped deep residual network, where the basic block is ResNet-34. Here, ResNet is a deep learning model for classification.
3) *DenseUNet [65]:* DenseUNet is a U-shaped densely connected convolutional network, which the basic block is dense block.
4) *DeeplabV3 [30]:* DeeplabV3 is a deep NN for semantic segmentation, which is built with ResNet-34 and ASPP.
5) *DeeplabV3+ [31]:* DeeplabV3+ is an enhanced version of DeeplabV3, in which the basic block is ResNet-34.
6) *TransUNet [34]:* TransUNet is a transformer-based U-shaped network for semantic segmentation.
7) *Swin-UNet [35]:* Swin-UNet is a Unet-like pure transformer, in which the basic block is swin-transformer block. Here, Swin-UNet initially uses four times downsampling.
8) *DAU-Net [28]:* DAU-Net is a dual-attention U-Net model based on ResUNet, where the basic block is ResNet-34.
9) *Segformer [36]:* Segformer is a transformer-based network for semantic segmentation.
10) *ConvUNeXt [37]:* ConvUNeXt is a U-shaped network built upon on the ConvNeXt block.
11) *VAN [38]:* VAN is a visual attention network.
12) *CBAM [66]:* CBAM is a simple yet effective attention module for feed-forward CNNs.
13) *GeM [67]:* GeM is a pooling operation based on generalized mean pooling, which adapts its shape to accommodate features of different scales and shapes.

TABLE I
COMPARISON RESULTS

Method	MIoU ↑	FWIoU ↑	Acc ↑	Jaccard ↑	Precision ↑		Recall ↑		F1-Score ↑	
					Sea ice	Open water	Sea ice	Open water	Sea ice	Open water
UNet [55] (2015)	80.1	81.9	88.7	85.4	89.0	92.1	95.4	81.9	92.2	87.0
ResUNet [29](2015)	81.6	82.4	89.3	85.6	90.1	90.9	94.5	84.0	92.3	87.5
DenseUNet [65] (2016)	82.8	83.6	89.8	86.7	90.1	92.9	95.9	83.8	93.0	88.4
Deeplabv3 [30] (2017)	89.9	90.4	94.3	92.2	94.4	95.9	97.5	91.1	96.0	93.5
Deeplabv3+ [31] (2018)	85.3	86.0	91.6	88.5	92.1	93.1	95.8	87.4	94.0	90.3
TransUNet [34] (2021)	85.8	86.5	91.6	89.2	91.2	95.9	97.6	85.5	94.3	90.7
Swin-UNet [35] (2021)	86.2	86.8	91.9	89.3	91.9	94.9	97.0	86.8	94.5	90.9
DAU-Net [28] (2021)	89.2	89.7	94.1	91.4	94.8	94.1	96.2	92.0	95.5	93.1
Segformer [36] (2021)	89.6	90.1	94.0	91.9	94.1	96.0	97.5	90.7	95.8	93.4
ConvUNeXt [37] (2022)	88.1	88.6	93.6	90.5	93.9	92.5	95.1	92.1	95.0	92.3
VAN [38] (2022)	89.7	90.2	94.3	91.9	94.7	95.1	96.9	91.7	95.8	93.4
MSDA-Net+CBAM[66] (2018)	92.1	92.4	95.6	93.8	96.3	95.9	93.6	**97.7**	94.9	**96.8**
MSDA-Net+GeM[67] (2018)	90.8	91.2	94.9	92.8	95.6	95.3	92.7	97.2	94.1	96.2
MSDA-Net (Ours)	**93.0**	**93.4**	**96.1**	**94.5**	**96.3**	**97.0**	**98.1**	94.2	**97.2**	95.6

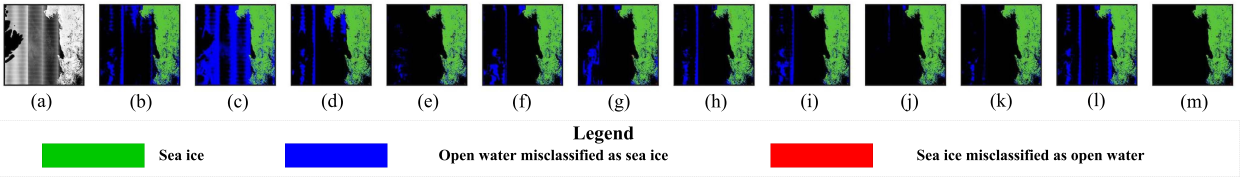

(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
(m)
Legend
Sea ice
Open water misclassified as sea ice
Sea ice misclassified as open water

Fig. 7. SAR scene (HV channel), the classification results of all models. (a) SAR scene (HV channel). (b) UNet. (c) ResUNet. (d) DenseUNet. (e) DeeplabV3. (f) DeeplabV3+. (g) TransUNet. (h) Swin-UNet. (i) DAU-Net. (j) Segformer. (k) ConvUNeXt. (l) VAN. (m) MSDA-Net (Ours).

E. Result and Analysis

We compare the proposed MSDA-Net with other prevailing models. Table I shows the evaluated metrics of all models. The best results are indicated in bold, and the second results are labeled by underlining. It can be obviously observed that MSDA-Net achieves the best performance. Besides, both CNN-based methods (DeeplabV3) and transformer-based (Segformer) methods perform well. The results indicate that the combination of VAN and DAU-Net can improve the performance. Moreover, we can see that the improvements of existing models are extremely limited. The improvement of the second (DeeplabV3) and third (VAN) highest models in terms of MIoU is 0.2%. However, compared with VAN in terms of MIoU, the improvement of MSDA-Net is 3.2%, which is 15 times more efficient than DeeplabV3. Similarly, the performances in terms of FWIoU, Acc, and Jaccard all illustrate this view. Fig. 7 shows the visualization of an SAR scene. It can be observed that our model with superior performance tend to generate fewer blue regions on the left of the results, which means the model misclassifies less open water as sea ice. In Fig. 7, we observe that compared with other models, MSDA-Net barely misclassifies open water as sea ice. Besides, the highest performance of MSDA-Net in terms of precision, recall, and F1-Score shows the robustness of this model, which produces the fewest red and blue regions in the figure.

TABLE II
COMPLEXITY ANALYSIS: PARAMETERS AND FLOPS COMPARISON

Method	Input size	Params	FLOPs
UNet [55]	5 × 512 × 512	17.3M	161.1G
ResUNet [29]	5 × 512 × 512	13.0M	20.4G
DenseUNet [65]	5 × 512 × 512	1.9M	152.2G
Deeplabv3 [30]	5 × 512 × 512	25.4M	79.5G
Deeplabv3+ [31]	5 × 512 × 512	26.7M	164.5G
TransUNet [34]	5 × 512 × 512	66.8M	131.2G
Swin-UNet [35]	5 × 512 × 512	27.1M	31.0G
DAU-Net [28]	5 × 512 × 512	18.6M	21.8G
Segformer [36]	5 × 512 × 512	82.0M	64.4G
ConvUNeXt [37]	5 × 512 × 512	46.2M	124.3G
VAN [38]	5 × 512 × 512	12.9M	51.7G
MSDA-Net+CBAM [66]	5 × 512 × 512	10.6M	153.5G
MSDA-Net+GeM [67]	5 × 512 × 512	10.6M	153.4G
MSDA-Net (Ours)	5 × 512 × 512	10.6M	153.5G

We conducted two separate experiments on MSDA-Net. In the first experiment, we replaced the PDAM module with the CBAM module and embedded CBAM into the ConvNeXt blocks, forming CBAM-ConvNeXt blocks. In the second experiment, we replaced the maxpooling and avgpooling operations in the PDAM module with the more advanced gempooling operation. We compared the performance of these two modules, and the results are shown in Table I. The comparison results showed that, compared to CBAM, our PDAM module demonstrated improvements in MIoU, FWIoU, Acc, Jaccard, and precision metrics. However, for Recall and F1-Score metrics specifically related to open water areas, the PDAM module did not perform as well as the CBAM module. This is because the PDAM module focuses more on local fragmented ice features, while the CBAM module emphasizes global information. In addition, replacing the maxpooling and avgpooling operations with gempooling did not result in performance improvement.

In Table II, we conducted a detailed comparison of the complexities of various models, with a particular focus on two key metrics: parameter count (Params) and floating-point operations (FLOPs). Our proposed MSDA network exhibits a relatively small parameter count, standing at 10.6 M. When compared to other models, it is only larger than the parameter count of

TABLE III
COMPARISON RESULTS OF DIFFERENT REPRESENTATIVE SCENES

Method	Time	MIoU ↑	FWIoU ↑	Acc ↑	Jaccard ↑	Precision ↑		Recall ↑		F1-Score ↑	
						Sea ice	Open water	Sea ice	Open water	Sea ice	Open water
DeeplabV3	20210814T071803	84.0	93.9	89.4	96.4	97.8	**86.8**	**98.6**	80.1	98.2	83.5
	20210814T085813	87.3	87.5	93.3	88.5	94.4	92.0	93.5	93.1	94.0	92.6
	20210815T062219	84.7	88.8	92.3	77.2	85.0	96.7	89.4	95.2	87.2	96.0
	20210815T093835	82.6	93.9	89.5	96.4	98.1	82.5	98.3	80.7	98.2	81.6
Segformer	20210814T071803	85.2	94.1	94.1	96.4	98.9	80.1	97.4	80.8	98.2	85.4
	20210814T085813	89.6	89.7	94.8	90.4	**97.8**	91.0	92.2	97.4	95.0	94.2
	20210815T062219	85.0	88.7	94.4	78.7	80.9	**98.6**	95.7	93.2	88.3	95.9
	20210815T093835	85.0	94.5	95.8	96.6	99.4	76.8	97.1	94.4	98.3	95.6
VAN	20210814T071803	88.4	95.5	95.5	97.3	99.2	84.7	98.1	92.9	98.6	88.8
	20210814T085813	77.6	77.6	88.2	77.8	96.5	79.6	80.0	96.4	88.3	88.0
	20210815T062219	83.4	87.4	94.1	76.0	78.3	98.8	96.2	91.9	87.3	95.3
	20210815T093835	79.4	91.9	94.5	94.7	99.3	66.9	95.4	93.6	97.3	80.2
DAU-Net	20210814T071803	88.2	95.4	95.4	97.3	99.2	84.4	98.1	92.7	98.6	88.6
	20210814T085813	38.3	37.0	61.8	25.8	91.7	51.6	26.5	97.0	59.1	74.3
	20210815T062219	66.6	72.2	87.7	56.2	56.9	99.2	**97.8**	77.5	77.4	88.3
	20210815T093835	55.1	74.5	87.6	78.9	99.5	31.7	79.2	**95.9**	89.3	63.8
MSDA-Net (Ours)	20210814T071803	**89.3**	**95.9**	**95.8**	**97.5**	**99.2**	86.1	98.3	**93.3**	**98.8**	**89.7**
	20210814T085813	**90.8**	**90.9**	**95.4**	**91.5**	97.3	**92.8**	**93.9**	**96.8**	**95.6**	**94.8**
	20210815T062219	**90.7**	**93.2**	**95.7**	**86.0**	**90.6**	98.3	94.4	**97.0**	**92.5**	**97.6**
	20210815T093835	**89.0**	**96.1**	**96.1**	**97.7**	**99.4**	**84.7**	**98.3**	93.8	**98.8**	**89.3**

the DenseUNet model (1.9 M), while being smaller than the parameter counts of other mainstream models. This suggests that the model demands a smaller storage footprint. However, the MSDA network demonstrates a higher FLOPs value, reaching 153.5 G. This is attributed to the ability of the MSDA-Net's MSSA module and PDAM module to extract multiscale and local information, which requires a higher computational cost due to the increased number of calculations involved in capturing these intricate features.

F. Case Study

Due to thermal noise, the performance of the model is by varying factors. Therefore, to evaluate the performances of MSDA-Net in different scenes (such as an entire block of sea ice, the complex sea ice boundary, floes, and geographic location), we conduct a series of case studies to validate the robustness and accuracy of our model. The experimental results show the advantages of MSDA-Net with robust edge detail detection, minor misclassification, and superior regional crushed ice detection.

Here, four different representative scenes are chosen to analyze the performances of the models. Table III shows the results of evaluation metrics in these four scenes and Fig. 8 shows the visualizations of these four scenes. In Fig. 8, the first column is the HV channel of these scenes and the reminder columns represent the segmentation results from various models. Here, the first scene is an entire block of sea ice, while the second scene includes the junction of ocean, land, and ice. The third scene contains many small floes that could be challenging for the classification of open water and sea ice, and the fourth scene has a complex sea ice boundary. In Fig. 8, other columns show the visualization of predictions from several models, which achieve top five MIoU in Table III. Moreover, the green, blue, and red parts denote the regions of sea ice, open water misclassified as sea ice, and sea ice misclassified as open water, respectively, while the whole black area of the HV channel is the land. Next, we will analyze the aforementioned four scenes. respectively.

The first scene contains an entire block of sea ice, which is the most explicit scene for model classifying. The results in Table III show MSDA-Net achieves the best performance in this type of scene. Moreover, we observe that other models contain broad red or blue parts in the lower right of the scenes in the first row of Fig. 8.

For the second scene, not only does the influence of thermal noise play a role, but the presence of ocean, land, and ice junction also causes enormous difficulty in accurately classifying. DAU-Net achieves the worst results in Table III. According to the precision and recall, we find that DAU-Net has misclassified a lot of sea ice as open water. The largest red parts of DAU-Net in Fig. 8(k) also verified this view. In the middle right of Fig. 8(j), the red parts indicate VAN cannot identify the sea ice in the region of ocean, land, and ice junction. DeeplabV3 and Segformer perform better than the aforementioned two models. However, it is difficult for DeeplabV3 and Segformer to accurately classify the edge of the junction in Fig. 8(h) and (i). Compared with these models, our model achieves the best performance on the whole and edges.

In the third scene, many small floes and thermal noise are challenging for the classification of sea ice and open water. It is obvious that MSDA-Net achieves the best performance in almost all indexes in Table III. For other models, the low precision of sea ice shows that there are many water pixels misclassified as sea ice pixels. Moreover, according to the first row of Fig. 8, we observe that the blue regions on the left, which denote pixels of high value, cause the misclassification. In most SAR sea ice scenes, this phenomenon with a lot of thermal noise is common. The better performance of DeeplabV3 and MSDA-Net indicates the robustness of these two models. Besides, MSDA-Net not only misclassifies less water as sea ice, but it also classifies the sea ice of edges more accurately than DeeplabV3.

There is a complex sea ice boundary in the fourth scene, which causes a huge challenge for the identification of the details. DAU-Net achieves the worst results in Table III and the largest

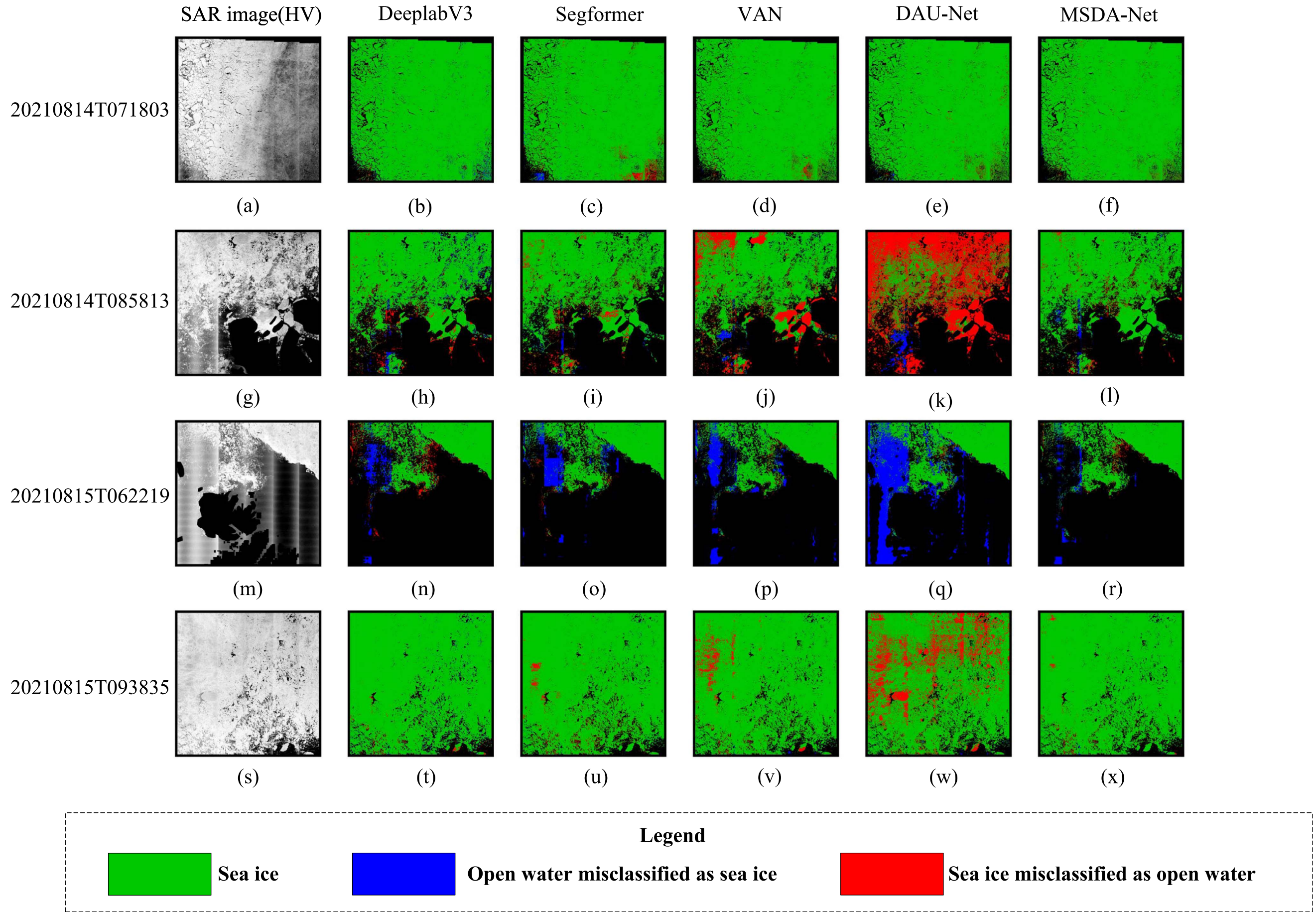


Fig. 8. SAR scene (HV channel), the classification results of DeeplabV3, Segformer, VAN, DAU-Net, and MSDA-Net (Ours). (a)–(f) 20210814T071803. (g)–(l) 20210814T085813. (m)–(r) 20210815T062219. (s)–(x) 20210815T093835.

red area in Fig. 8(w). Besides, the crushed ice in the lower right corner of the scene is difficult to identify. Only MSDA-Net and Segformer can modify the region. Moreover, according to the results in Table III, it is obvious that MSDA-Net is superior to Segformer.

To highlight the details of fine ice fragments captured by our model, we divided the original image into four equal parts vertically and horizontally, focusing on the regions where small sea ice clusters occur. In Fig. 9, we selected four densely fragmented ice scenes to demonstrate our model's ability to capture multiscale sea ice objects. The first column here contains the HV channel images of these scenes. The left half displays the data in their original size, while the right half shows the data from the zoomed-in region. The remaining columns demonstrate the segmentation results obtained using different models within this zoomed-in region.

In these four densely fragmented ice scenes, we observed two primary classification errors. First, there is the misclassification of blue vertical stripes, which can be attributed to thermal noise interference in SAR images. This interference may lead to the inaccurate classification of certain water areas as sea ice areas, resulting in the appearance of blue regions. Second, we identified the misclassification of small, dot-like ice fragments, which presents as red dot-shaped errors in the images. This misclassification could be attributed to the inherent challenge faced by the model in accurately capturing the structure of fine ice fragments. Consequently, these smaller ice fragment areas may be incorrectly identified as open-water regions. Compared to the other four models, MSDA-Net effectively avoids the interference of thermal noise and performs well in minimizing misclassifications of blue stripes. In addition, MSDA-Net utilizes multiscale module feature extraction, enabling better extraction and capturing of detailed information on dot-like ice fragments, resulting in excellent performance in classifying dot-like ice fragments. In summary, MSDA-Net demonstrates advantages over the other four models in dealing with challenges such as thermal noise interference and dot-like ice fragment classification, showing good performance and robustness.

G. Ablation Study

In order to investigate the effects of different modules and channels, we conducted ablation experiments and summarized their outcomes. The ablation experiments were divided into three parts: the first part involved further analysis of the main

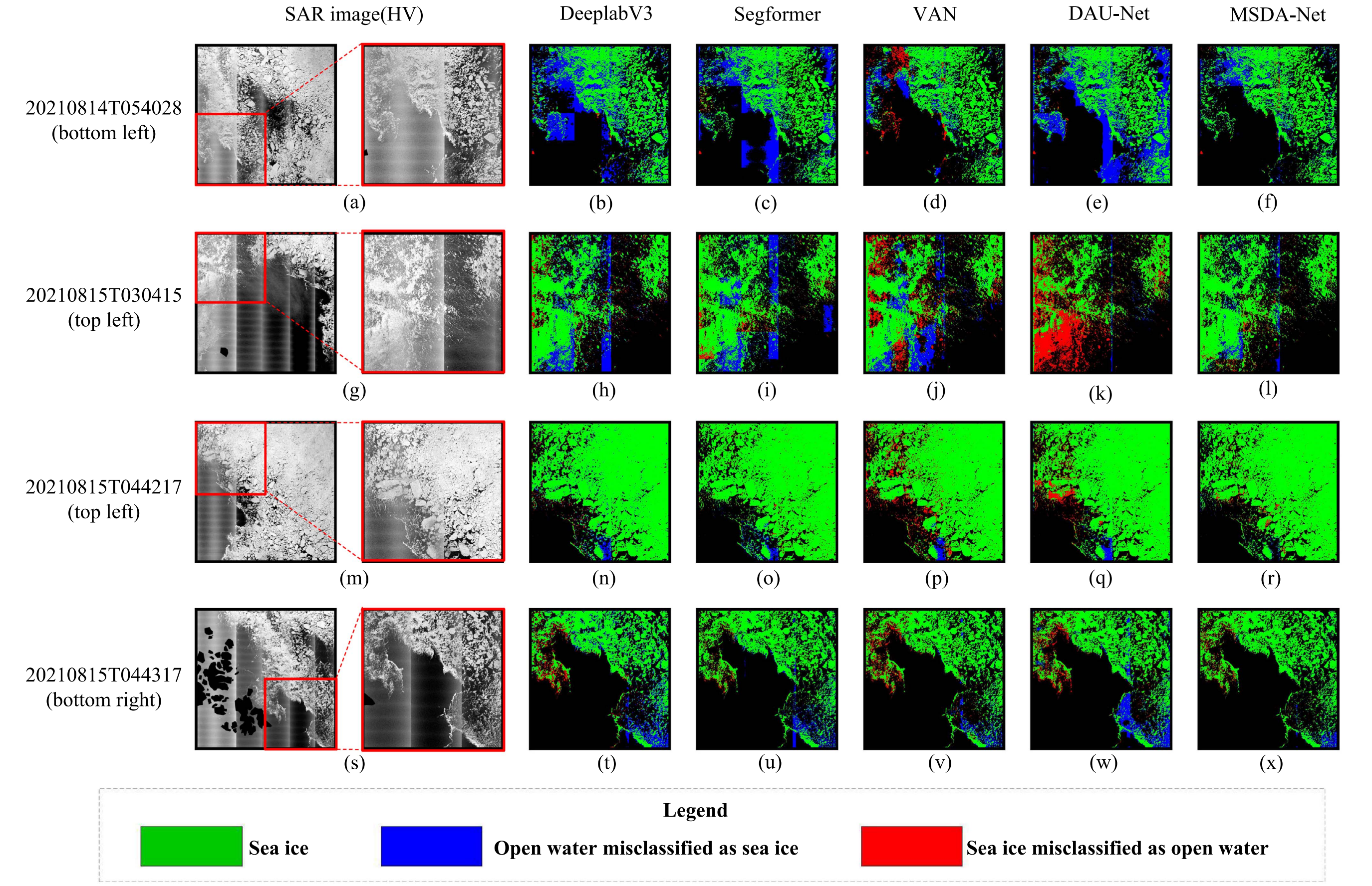

SAR image(HV)
DeeplabV3
Segformer
VAN
DAU-Net
MSDA-Net
20210814T054028 (bottom left)
20210815T030415 (top left)
20210815T044217 (top left)
20210815T044317 (bottom right)
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
(m)
(n)
(o)
(p)
(q)
(r)
(s)
(t)
(u)
(v)
(w)
(x)
Legend
🟩 Sea ice
🟦 Open water misclassified as sea ice
🟥 Sea ice misclassified as open water

Fig. 9. SAR scene (HV channel), the classification results of DeeplabV3, Segformer, VAN, DAU-Net, and MSDA-Net (Ours). (a)–(f) 20210814T054028 (bottom left). (g)–(l) 20210815T030415 (top left). (m)–(r) 20210815T044217 (top left). (s)–(x) 20210815T044317 (bottom right).

TABLE IV
Comparison Results of Module Ablation Study

Method	MSSA	PDAM		MIoU ↑	FWIoU ↑	Acc ↑	Jaccard ↑	Precision ↑		Recall ↑		F1-Score ↑	
		CAM	SAM					Sea ice	Open water	Sea ice	Open water	Sea ice	Open water
Baseline				88.8	89.2	94.1	91.0	95.7	92.2	94.9	93.4	95.3	92.8
MSDA-Net$_M$	✓			90.3	90.8	94.3	92.5	94.0	97.2	98.3	90.4	96.2	93.8
MSDA-Net$_P$		✓	✓	91.4	91.9	95.0	93.4	94.6	**97.8**	**98.7**	91.3	96.7	94.6
MSDA-Net$_C$		✓		91.0	91.4	95.0	92.9	95.9	95.3	92.6	97.4	94.2	**96.3**
MSDA-Net$_S$			✓	89.8	90.3	94.2	92.1	96.3	94.1	90.6	**97.7**	93.4	95.8
MSDA-Net$_{MC}$	✓	✓		91.9	92.3	95.5	93.6	96.3	95.8	93.4	**97.7**	94.8	96.7
MSDA-Net$_{MS}$	✓		✓	90.5	90.9	94.6	92.6	96.2	96.5	91.6	97.6	93.8	96.1
MSDA-Net	✓	✓	✓	**93.0**	**93.4**	**96.1**	**94.5**	**96.3**	97.0	98.1	94.2	**97.2**	95.6

modules, PDAM, and MSSA; the second part compared the two submodules, SAM and CAM, under PDAM; and the third part conducted ablation experiments based on the HPF and PE channels.

1) Analysis of PDAM and MSSA Modules: The evaluation metrics are presented in Table IV. The baseline refers to MSDA-Net without PDAM and MSSA, MSDA-Net$_M$ represents MSDA-Net with MSSA, and MSDA-Net$_P$ represents MSDA-Net with PDAM. To ensure the similarity of structures and fairness in the experimental results, we substitute MSSA with a convolution layer for both baseline and MSDA-Net$_P$. The convolution layer has a kernel size, stride, and padding of 3, 1, and 1, respectively.

From Table IV, the results of MSDA-Net$_M$ and MSDA-Net$_P$ are higher than baseline, which shows the advantage of introducing MSSA and PDAM, respectively. The lower prediction and higher recall of sea ice indicate that MSDA-Net$_M$ and MSDA-Net$_P$ tend to identify more accurate sea ice, although accompanied by some degree of misclassification. However, MSDA-Net combines the advantage of PDAM and MSSA and achieves huge improvement on any evaluation metrics.

Besides, Table V and Fig. 10 show the influence of PDAM and MSSA under different types of scenes. Here, Table V presents the results of evaluation metrics in different representative scenes, and Fig. 10 shows the visualization of different models.

TABLE V
COMPARISON MODULE ABLATION RESULTS OF REPRESENTATIVE SCENES

Method	Time	MIoU ↑	FWIoU ↑	Acc ↑	Jaccard ↑	Precision ↑		Recall ↑		F1-Score ↑	
						Sea ice	Open water	Sea ice	Open water	Sea ice	Open water
Baseline	20210814T071803	83.5	93.2	95.6	95.7	**99.4**	74.1	96.2	**95.0**	97.8	84.6
	20210814T085813	53.0	52.3	72.8	46.3	**98.3**	60.0	46.6	**99.0**	72.5	79.5
	20210815T062219	82.7	86.8	93.7	75.0	77.4	98.7	95.9	91.5	86.7	95.1
	20210815T093835	74.9	89.6	94.2	92.8	**99.5**	58.8	93.3	**95.2**	96.4	77.0
MSDA-Net$_M$	20210814T071803	88.6	95.5	95.7	97.3	99.2	84.6	98.1	93.4	98.6	89.0
	20210814T085813	88.0	88.1	94.0	88.8	97.5	89.5	90.8	97.1	94.2	93.3
	20210815T062219	75.3	80.3	91.6	65.9	66.8	**99.3**	**97.9**	85.2	82.4	92.3
	20210815T093835	85.8	94.8	96.1	96.8	99.5	77.8	97.3	95.0	98.4	86.4
MSDA-Net$_P$	20210814T071803	**89.5**	**96.0**	93.8	**97.7**	98.7	**90.7**	**99.0**	88.7	98.8	89.7
	20210814T085813	**91.3**	**91.4**	**95.5**	**92.2**	95.9	**95.0**	**95.9**	95.0	**95.9**	**95.0**
	20210815T062219	83.3	87.3	94.3	75.9	77.7	99.0	97.1	91.5	87.4	95.3
	20210815T093835	87.0	95.4	94.5	97.2	99.1	83.1	98.1	90.9	98.6	87.0
MSDA-Net (Ours)	20210814T071803	89.3	95.8	**95.8**	97.5	99.2	86.1	98.3	93.3	**98.8**	**89.7**
	20210814T085813	90.8	90.9	95.4	91.5	97.3	92.8	93.9	96.8	95.6	94.8
	20210815T062219	**90.7**	**93.2**	**95.7**	**86.0**	**90.6**	98.3	**94.4**	**97.0**	**92.5**	**97.6**
	20210815T093835	**89.0**	**96.1**	**96.1**	**97.7**	99.4	**84.7**	**98.3**	93.8	**98.8**	**89.3**

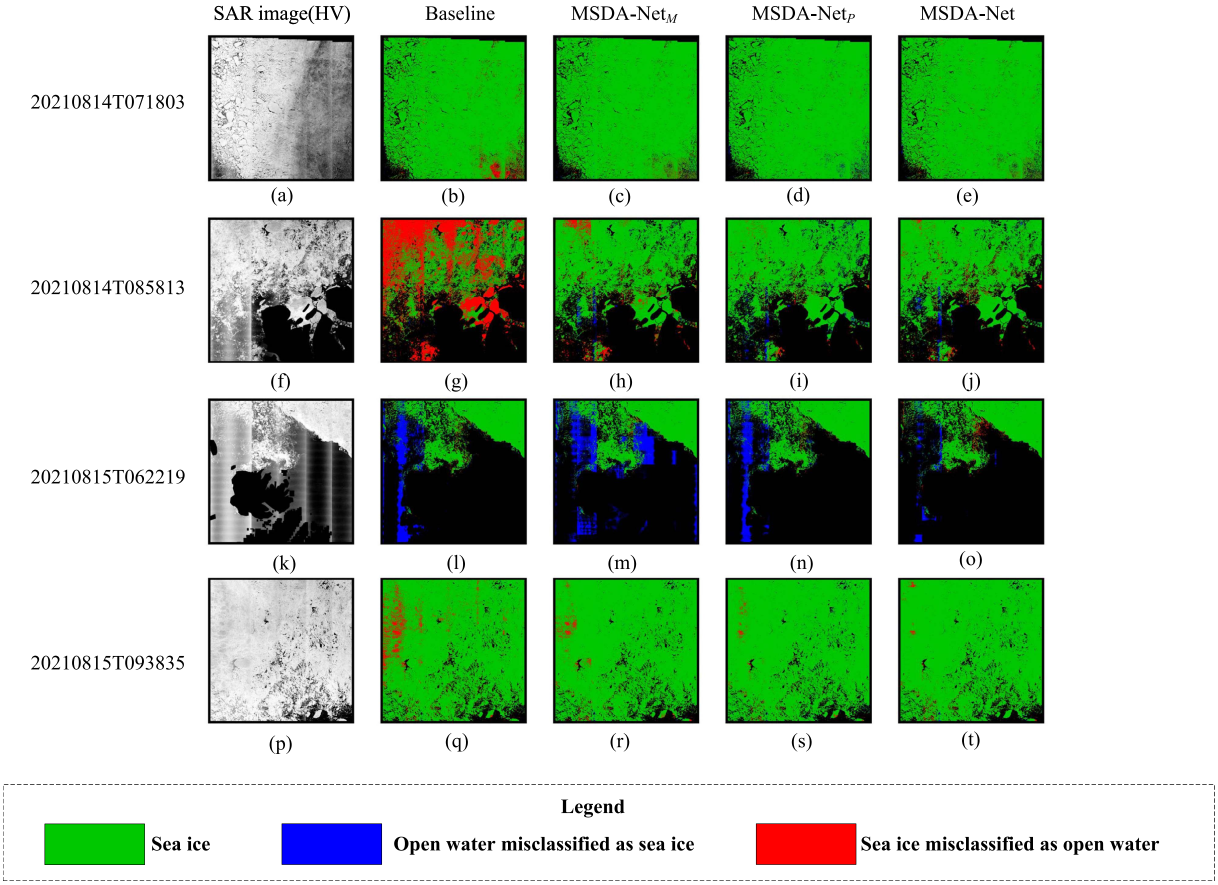


Fig. 10. SAR scene (HV channel), the classification results of baseline, MSDA-Net$_M$, MSDA-Net$_P$, and MSDA-Net (Ours). (a)–(e) 20210814T071803. (f)–(j) 20210814T085813. (k)–(o) 20210815T062219. (p)–(t) 20210815T093835.

TABLE VI
COMPARISON RESULTS OF CHANNEL ABLATION STUDY

Method	HPF	PE	MIoU ↑	FWIoU ↑	Acc ↑	Jaccard ↑	Precision ↑		Recall ↑		F1-Score ↑	
							Sea ice	Open water	Sea ice	Open water	Sea ice	Open water
Baseline			90.3	90.8	94.6	92.4	94.7	-	97.4	-	96.1	-
MSDA-Net$_{\mathrm{HPF}}$	✓		91.3	91.7	95.2	93.1	95.6	-	97.3	-	96.5	-
MSDA-Net$_{\mathrm{PE}}$		✓	92.0	92.4	95.6	93.7	95.8	-	97.7	-	96.8	-
MSDA-Net	✓	✓	**93.0**	**93.4**	**96.1**	**94.5**	**96.3**	**97.0**	**98.1**	**94.2**	**97.2**	**95.6**

For the first scene, both MSDA-Net$_P$ and MSDA-Net$_M$ achieve superior performance than the baseline model in Table V. According to the first row in Fig. 10, we can see that PDAM and MSSA are helpful in improving the identity of the lower right of the scenes.

The baseline model achieves the worst performance in the second scene. Compared with the baseline model, it is obvious that the introduction of PDAM and MSSA can improve the performance of the classification in the ocean, land, and ice junction. Moreover, MSDA-Net$_P$ performs better than MSDA-Net, which implies PDAM can enhance edge identification.

In the third scene, MSDA-Net$_M$ has misclassified more open water pixels with thermal noise as sea ice than the baseline model, while MSDA-Net$_P$ achieves great performance in this condition. The superior performance of MSDA-Net$_P$ can contribute to PDAM, which extracts an important feature of PE and HDF channels. Moreover, the best result of MSDA-Net shows that the combination of PDAM and MSSA can capture more valuable channel and multiscale spatial information.

For the fourth scene, there is a broad red region on the left of Fig. 10(q), which implies the baseline model cannot identify complex textures of sea ice. Both MSDA-Net$_M$ and MSDA-Net$_P$ can enhance the identification of textures in Fig. 10(r) and (s). Moreover, the combination of PDAM and MSSA can generate the best performance.

2) Analysis of SAM and CAM Submodules Under PDAM: The PDAM module effectively enhances the representation of relative positions and high-pass filtering channels, leading to significant improvements in all evaluation metrics when incorporated into the model. Within the PDAM module, there are two submodules, CAM and SAM. To investigate the internal functions of CAM and SAM and understand their contributions, we conducted ablation experiments. Specifically, we denoted the models as MSDA-Net$_C$, MSDA-Net$_S$, MSDA-Net$_{\text{MC}}$, and MSDA-Net$_{\text{MS}}$, which correspond to MSDA-Net with CAM, MSDA-Net with SAM, MSDA-Net with MSSA and CAM, and MSDA-Net with MSSA and SAM, respectively.

Table IV presents the comparison of four sets of experiments: MSDA-Net$_C$ versus the baseline, MSDA-Net$_S$ versus the baseline, MSDA-Net$_{\text{MC}}$ versus MSDA-Net$_M$, and MSDA-Net$_{\text{MS}}$ versus MSDA-Net$_M$. The results indicate a significant improvement in the model's performance in terms of MIoU, FWIoU, Acc, and other indicators upon integrating the CAM and SAM submodules, thus confirming their effectiveness. Specifically, the CAM submodule has a more prominent role than the SAM submodule in enhancing the model performance.

In the context of sea ice images, the primary issue lies in noise interference. The CAM submodule effectively utilizes additional HPF and PE channels to explore channel relationships and focus on vertical regions of noise, contributing to noise removal. Consequently, the CAM submodule exhibits more substantial performance enhancement in sea ice images.

However, sea ice exhibits a multiscale nature, with small ice floes potentially overlapping with noise, making accurate localization of sea ice regions challenging when relying solely on the SAM submodule. The SAM submodule emphasizes global features instead of local areas, potentially underutilizing semantic information embedded in sea ice images. Nevertheless, synergistically combining the CAM and SAM submodules in PDAM fully leverages their advantages. The CAM submodule provides accurate semantic information, while the SAM submodule focuses on global features, enabling the PDAM to achieve enhanced sea ice region localization and extraction.

Consequently, the combined employment of the CAM and SAM submodules within the PDAM module effectively utilizes rich semantic information to analyze and process sea ice images across multiple scales, leading to improved results. The integration of these CAM and SAM submodules proves more effective in enhancing model performance.

3) Analysis of HPF and PE Channels: To examine the impact of incorporating additional channels into the model, we conducted a series of ablation experiments focusing on the PE and high-pass filtering channels. The effects were summarized and are presented in Table VI. The baseline model refers to the utilization of only the HV, HH − HV, and HH/HV channels within the multiscale attention network. MSDA-Net $_{\text{HPF}}$ represents the multiscale attention network that includes the HV, HH− HV, HH/HV, and high-pass filtering channels. MSDA-Net $_{\text{PE}}$ corresponds to the multiscale attention network that incorporates the HV, HH− HV, HH/HV, and PE channels. MSDA-Net denotes the multiscale attention network that combines the HV, HH − HV, HH/HV, high-pass filtering, and PE channels.

The experimental findings demonstrate that both MSDA-Net$_{\text{HPF}}$ and MSDA-Net$_{\text{PE}}$ outperform the baseline model, showcasing the enhanced capability of the model in sea ice monitoring. From Table VI, it is evident that the high-pass filtering channel primarily improves the model's accuracy by reducing the misclassification of sea ice. On the other hand, the PE channel not only contributes to accuracy improvement but also performs well in recall. By leveraging the advantages of the PE channel and the high-pass filtering channel, MSDA-Net achieves significant improvements across all evaluation metrics.

V. Conclusion

In this article, we introduce the MSDA-Net for open water and sea ice classification. We identify two challenges and propose corresponding solutions to address them. The first challenge involves the common noise interference in Sentinel-1 SAR images, while the second challenge relates to handling sea ice objects of multiple scales.

To address the first challenge, we introduce specific measures to mitigate noise interference, including the incorporation of additional positional embeddings and HPF channels. These components aid in noise filtering and localization of vertical line noise. In addition, we introduce the PDAM to capture both channel and spatial features, effectively leveraging the extra channel information to reduce noise interference in Sentinel-1 SAR images. To tackle the second challenge, we develop the MSSA module within the network's architecture. This module enhances the extraction of multiscale spatial features, enabling the network to effectively handle sea ice objects at various scales, particularly dense small ice fragments and punctate ice fragments.

Our comprehensive experimental results demonstrate the effectiveness and superiority of the proposed model, achieving state-of-the-art performance among all models. We conduct extensive ablation experiments, including independent ablations on each module, investigating the impact of the PDAM and MSSA modules. Furthermore, we also specifically evaluate the contributions of the CAM and SAM submodules within PDAM. Besides, we perform channel ablation experiments to explore the effects of additional position embedding and HPF channels on the model's performance. The results indicate that these enhancements have resulted in significant performance improvements.

For future work, our primary focus will be on finding more effective methods to mitigate the influence of thermal noise. Although we have verified the effectiveness of incorporating extra position embedding and HPF channels, we aim to explore additional techniques and strategies to further reduce the impact of thermal noise on our model's performance.

References

[1] D. Olonscheck, T. Mauritsen, and D. Notz, "Arctic sea-ice variability is primarily driven by atmospheric temperature fluctuations," *Nature Geosci.*, vol. 12, no. 6, pp. 430–434, 2019.

[2] M. Dabboor and T. Geldsetzer, "Towards sea ice classification using simulated radarsat constellation mission compact polarimetric SAR imagery," *Remote Sens. Environ.*, vol. 140, pp. 189–195, 2014.

[3] D. J. Cavalieri, P. Gloersen, and W. J. Campbell, "Determination of sea ice parameters with the Nimbus-7 SMMR," *J. Geophys. Res., Atmos.*, vol. 89, no. D4, pp. 5355–5369, 1984.

[4] D. Cavalieri et al., "Aircraft active and passive microwave validation of sea ice concentration from the defense meteorological satellite program special sensor microwave imager," *J. Geophys. Res., Oceans*, vol. 96, no. C12, pp. 21989–22008, 1991.

[5] T. Kawanishi et al., "The advanced microwave scanning radiometer for the Earth observing system (AMSR-E), NASDA's contribution to the EOS for global energy and water cycle studies," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 2, pp. 184–194, Feb. 2003.

[6] X. Zhao et al., "Sea ice concentration derived from FY-3D MWRI and its accuracy assessment," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4300418.

[7] L. Kilic, R. T. Tonboe, C. Prigent, and G. Heygster, "Estimating the snow depth, the snow–ice interface temperature, and the effective temperature of arctic sea ice using advanced microwave scanning Radiometer 2 and ice mass balance buoy data," *Cryosphere*, vol. 13, no. 4, pp. 1283–1296, 2019.

[8] G. A. Riggs, D. K. Hall, and V. V. Salomonson, "Modis sea ice products user guide to collection 5," *NASA Goddard Space Flight Center*, vol. 49, pp. 1–50, 2006.

[9] D. Xian, P. Zhang, L. Gao, R. Sun, H. Zhang, and X. Jia, "Fengyun meteorological satellite products for Earth system science applications," *Adv. Atmospheric Sci.*, vol. 38, no. 8, pp. 1267–1284, 2021.

[10] P. Maillard, D. A. Clausi, and H. Deng, "Operational map-guided classification of SAR sea ice imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 12, pp. 2940–2951, Dec. 2005.

[11] M. Fily and D. A. Rothrock, "Extracting sea ice data from satellite SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. GE-24, no. 6, pp. 849–854, Nov. 1986.

[12] F. Fetterer, C. Bertoia, and J. P. Ye, "Multi-year ice concentration from RADARSAT," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. Proc. Remote Sens., Sci. Vis. Sustain. Develop.*, vol. 1, 1997, pp. 402–404.

[13] L.-K. Soh, C. Tsatsoulis, D. Gineris, and C. Bertoia, "Arktos: An intelligent system for SAR sea ice image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 1, pp. 229–248, Jan. 2004.

[14] J. A. Karvonen, "Baltic sea ice SAR segmentation and classification using modified pulse-coupled neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1566–1574, Jul. 2004.

[15] S. Chen et al., "MYI floes identification based on the texture and shape feature from dual-polarized sentinel-1 imagery," *Remote Sens.*, vol. 12, no. 19, 2020, Art. no. 3221.

[16] S. Leigh, Z. Wang, and D. A. Clausi, "Automated ice–water classification using dual polarization SAR satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5529–5539, Sep. 2014.

[17] X.-M. Li, Y. Sun, and Q. Zhang, "Extraction of sea ice cover by sentinel-1 SAR based on support vector machine with unsupervised generation of training data," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3040–3053, Apr. 2021.

[18] J.-W. Park, A. A. Korosov, M. Babiker, J.-S. Won, M. W. Hansen, and H.-C. Kim, "Classification of sea ice types in Sentinel-1 synthetic aperture radar images," *Cryosphere*, vol. 14, no. 8, pp. 2629–2645, 2020.

[19] H. Deng and D. A. Clausi, "Unsupervised segmentation of synthetic aperture radar sea ice imagery using a novel Markov random field model," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 528–538, Mar. 2005.

[20] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 364–368, Mar. 2016.

[21] Y. Xu and K. A. Scott, "Sea ice and open water classification of SAR imagery using CNN-based transfer learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Fort Worth, TX, USA, 2017, pp. 3262–3265.

[22] C. Wang, H. Zhang, Y. Wang, and B. Zhang, "Sea ice classification with convolutional neural networks using Sentinel-l scanSAR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Valencia, Spain, 2018, pp. 7125–7128.

[23] J. Li, C. Wang, S. Wang, H. Zhang, Q. Fu, and Y. Wang, "Gaofen-3 sea ice detection based on deep learning," in *Proc. Prog. Electromagn. Res. Symp.-Fall*, 2017, pp. 933–939.

[24] Y. Gao, F. Gao, J. Dong, and S. Wang, "Transferred deep learning for sea ice change detection from synthetic-aperture radar images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 10, pp. 1655–1659, Oct. 2019.

[25] H. Boulze, A. Korosov, and J. Brajard, "Classification of sea ice types in sentinel-1 SAR data using convolutional neural networks," *Remote Sens.*, vol. 12, no. 13, 2020, Art. no. 2165.

[26] S. Khaleghian, H. Ullah, T. Kræmer, N. Hughes, T. Eltoft, and A. Marinoni, "Sea ice classification of SAR imagery based on convolution neural networks," *Remote Sens.*, vol. 13, no. 9, 2021, Art. no. 1734.

[27] Y.-R. Wang and X.-M. Li, "Arctic sea ice cover data from spaceborne SAR by deep learning," *Earth Syst. Sci. Data Discuss*, vol. 2020, pp. 1–30, 2020.

[28] Y. Ren, X. Li, X. Yang, and H. Xu, "Development of a dual-attention u-net model for sea ice and open water classification on SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Feb. 2021, Art. no. 4010205.

[29] 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.

[30] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[31] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 801–818.

[32] A. Vaswani et al., "Attention is all you need," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 5998–6008.

[33] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Representations*, 2021, pp. 1–22.

[34] J. Chen et al., "TransuNet: Transformers make strong encoders for medical image segmentation," 2021, *arXiv:2102.04306*.

[35] H. Cao et al., "Swin-UNet: UNet-like pure transformer for medical image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 205–218.

[36] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 12077–12090.

[37] Z. Liu, H. Mao, C. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, New Orleans, LA, USA, 2022, pp. 11966–11976.

[38] M.-H. Guo, C.-Z. Lu, Z.-N. Liu, M.-M. Cheng, and S.-M. Hu, "Visual attention network," *Comput. Vis. Media*, vol. 9, no. 4, pp. 733–752, 2023.

[39] J.-W. Park, A. A. Korosov, M. Babiker, S. Sandven, and J.-S. Won, "Efficient thermal noise removal for Sentinel-1 topSAR cross-polarization channel," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1555–1565, Mar. 2018.

[40] Y. Sun and X.-M. Li, "Denoising Sentinel-1 extra-wide mode cross-polarization images over sea ice," *IEEE Trans. Geosci. Remote. Sens.*, vol. 59, no. 3, pp. 2116–2131, Mar. 2021.

[41] R. Kwok, E. Rignot, B. Holt, and R. Onstott, "Identification of sea ice types in spaceborne synthetic aperture radar data," *J. Geophys. Res., Oceans*, vol. 97, no. C2, pp. 2391–2402, 1992.

[42] L.-K. Soh and C. Tsatsoulis, "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 2, pp. 780–795, Mar. 1999.

[43] M.-A. Moen et al., "Comparison of feature based segmentation of full polarimetric SAR satellite sea ice images with manually drawn ice charts," *Cryosphere*, vol. 7, no. 6, pp. 1693–1705, 2013.

[44] A. S. Fors, C. Brekke, A. P. Doulgeris, T. Eltoft, A. H. Renner, and S. Gerland, "Late-summer sea ice segmentation with multi-polarisation SAR features in C and X band," *Cryosphere*, vol. 10, no. 1, pp. 401–415, 2016.

[45] Z. Yu, T. Wang, X. Zhang, J. Zhang, and P. Ren, "Locality preserving fusion of multi-source images for sea-ice classification," *Acta Oceanologica Sinica*, vol. 38, no. 7, pp. 129–136, 2019.

[46] A. Cristea, J. Van Houtte, and A. P. Doulgeris, "Integrating incidence angle dependencies into the clustering-based segmentation of SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2925–2939, May 2020.

[47] A. K. Liu, S. Martin, and R. Kwok, "Tracking of ice edges and ice floes by wavelet analysis of SAR images," *J. Atmospheric Ocean. Technol.*, vol. 14, no. 5, pp. 1187–1198, 1997.

[48] Q. Yu, C. Moloney, and F. M. Williams, "SAR sea-ice texture classification using discrete wavelet transform based methods," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Toronto, ON, Canada, 2002, pp. 3041–3043.

[49] J. Haarpaintner and S. Solbø, "Automatic ice-ocean discrimination in SAR imagery," *Norut IT- Rep.*, vol. 6, 2007, Art. no. 28.

[50] N. Y. Zakhvatkina, V. Y. Alexandrov, O. M. Johannessen, S. Sandven, and I. Y. Frolov, "Classification of sea ice types in ENVISAT synthetic aperture radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2587–2600, May 2013.

[51] W. Aldenhoff, C. Heuzé, and L. E. Eriksson, "Comparison of ice/water classification in Fram Strait from C-and L-band SAR imagery," *Ann. Glaciol.*, vol. 59, no. 76, pp. 112–123, 2018.

[52] H. Liu, H. Guo, and L. Zhang, "SVM-based sea ice classification using textural features and concentration from RADARSAT-2 dual-pol scanSAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 4, pp. 1601–1613, Apr. 2015.

[53] D.-B. Hong and C.-S. Yang, "Automatic discrimination approach of sea ice in the arctic ocean using sentinel-1 extra wide swath dual-polarized SAR data," *Int. J. Remote Sens.*, vol. 39, no. 13, pp. 4469–4483, 2018.

[54] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 3431–3440.

[55] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, 2015, pp. 234–241.

[56] D. Malmgren-Hansen et al., "A convolutional neural network architecture for Sentinel-1 and AMSR2 data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 1890–1902, Mar. 2021.

[57] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Montreal, QC, Canada, 2021, pp. 9992–10002.

[58] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[59] Z. Tian, T. He, C. Shen, and Y. Yan, "Decoders matter for semantic segmentation: Data-dependent decoding enables flexible feature aggregation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Long Beach, CA, USA, 2019, pp. 3126–3135.

[60] W. Dierking, "Mapping of different sea ice regimes using images from sentinel-1 and ALOS synthetic aperture radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1045–1058, Mar. 2010.

[61] W. Dierking, "Sea ice monitoring by synthetic aperture radar," *Oceanography*, vol. 26, no. 2, pp. 100–111, 2013.

[62] W. Tan, J. Li, L. Xu, and M. A. Chapman, "Semiautomated segmentation of Sentinel-1 SAR imagery for mapping sea ice in labrador coast," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1419–1432, May 2018.

[63] J. Karvonen, "Baltic sea ice concentration estimation based on c-band dual-polarized SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5558–5566, Sep. 2014.

[64] ESA, "SNAP." Accessed on: Jul. 2023. [Online]. Available: https://step.esa.int/main/toolboxes/snap/

[65] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[66] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

[67] F. Radenović, G. Tolias, and O. Chum, "Fine-tuning CNN image retrieval with no human annotation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1655–1668, Jul. 2019.

[68] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. Int. Conf. Learn. Representations*, 2019, pp. 1–18.

Zheng Zhang received the bachelor's degree from Jilin University, Changchun, China, in 2020, and the master's degree from the Harbin Institute of Technology, Shenzhen, China, in 2023, both in computer science. He is currently working toward the Ph.D. degree with Centre for Vision Speech and Signal Processing (CVSSP), University of Surrey, Guildford, U.K.

His research includes computer vision, medical image analysis, deep learning, and precipitation nowcasting.

Guangbo Deng received the B.S. degree in computer science, in 2022, from the Harbin Institute of Technology, Shenzhen, China, where he is currently working toward the Ph.D. degree in computer science.

His research interests include precipitation nowcasting, remote sensing image processing, and spatiotemporal predictive learning.

Chuyao Luo received the bachelor's degree in Internet of Things engineering from Dalian Maritime University, Dalian, China, in 2017, and the doctoral degree in computer science from the Harbin Institute of Technology, Shenzhen, China, in 2022.

He is currently a Postdoctoral Fellow with the Harbin Institute of Technology. His research includes data mining, computer vision, time-series data prediction, and precipitation nowcasting.

Xutao Li received the bachelor's degree from the Lanzhou University of Technology, Lanzhou, China, in 2007, and the master's and Ph.D. degrees from the Harbin Institute of Technology, Shenzhen, China, in 2009 and 2013, respectively, all in computer science.

He is currently a Professor with the School of Computer Science and Technology, Harbin Institute of Technology. His research interests include data mining, machine learning, graph mining, and social network analysis, especially tensor-based learning, and mining algorithms.

Yunming Ye received the Ph.D. degree in computer science from Shanghai Jiao Tong University, Shanghai, China, in 2004.

He is currently a Professor with the School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China. His research interests include data mining, text mining, and ensemble learning algorithms.

Di Xian received the Bachelor of Science degree in meteorology from Nanjing University, Nanjing, China, in 2002, and the master's degree in atmospheric physics and atmospheric environment from Peking University, Beijing, China, in 2009.

He is currently a Research Professor and the Director of S&T and International Cooperation Office, National Satellite Meteorological Center, China Meteorological Administration. His research interests including remote sensing, satellite data application, and Big Data system architecture .