

CNN and Transformer Fusion Network for Sea Ice Classification Using GaoFen-3 Polarimetric SAR Images

Jiande Zhang , Wenyi Zhang , Xiao Zhou, Qingwei Chu , Xiaoyi Yin, Guangzuo Li, Xiangyu Dai , Shuo Hu, and Fukun Jin

Abstract—This article investigates the safety risks associated with sea ice along the Arctic Route by using polar sea ice images obtained by Gaofen-3 (GF3) Synthetic Aperture Radar (SAR) satellites. We collected three SAR datasets, representative of GF3 satellites' operational modes, and constructed semantic segmentation datasets through meticulous annotation using concurrent optical satellite imagery obtained by Landsat satellites to ensure precision. We propose SI-CTFNet, an innovative sea ice semantic segmentation model that integrates convolutional neural networks (CNNs) and vision transformers (ViTs) for enhanced multiscale feature extraction. SI-CTFNet features a dual-pathway architecture designed to optimize feature extraction, complemented by the BiA-Fusion module, which effectively merges local and global features to improve decoding accuracy. In addition, we introduce the MSDAM module to facilitate a comprehensive multiscale contextual analysis, addressing the diverse distribution of ice types in the imagery. The model incorporates advanced decoding techniques, including a progressive upsampling approach for the CNN-fusion branch and an efficient All-MLP module for the Transformer branch. Performance evaluations across three distinct datasets reveal that SI-CTFNet significantly outperforms existing methods in key metrics and maintains efficacy with supplementary Sentinel1A C-band satellite data. Furthermore, we present a streamlined variant of SI-CTFNet, which achieves a threefold increase in inference speed with minimal reduction in classification accuracy. The ultimate objective of this work is to advance a precise sea ice forecasting and navigation system for polar regions, aimed at automating sea ice classification within a smart polar shipping framework.

Index Terms—CNN and transformer, sea ice classification, semantic segmentation, synthetic aperture radar.

I. INTRODUCTION

ACCORDING to data released by the US National Snow and Ice Data Center (NSIDC), in September 2022, the average extent of Arctic sea ice was 4.87 million square kilometers, which is 1.54 million square kilometers less than the

Received 11 June 2024; revised 26 August 2024; accepted 7 September 2024. Date of publication 23 September 2024; date of current version 23 October 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62331206. (Corresponding author: Wenyi Zhang.)

Jiande Zhang, Xiao Zhou, and Shuo Hu are with the QiLu Aerospace Information Research Institute, Jinan 250132, China (e-mail: zhangjd@aircas.ac.cn; zhouxiao@aircas.ac.cn; hushuo@aircas.ac.cn).

Wenyi Zhang, Qingwei Chu, Xiaoyi Yin, Guangzuo Li, Xiangyu Dai, and Fukun Jin are with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: wyzhang@aircas.ac.cn; chuqw@aircas.ac.cn; yinx@aircas.ac.cn; ligz@aircas.ac.cn; daixianguy22@mails.ucas.ac.cn; fkj15504196531@163.com).

Digital Object Identifier 10.1109/JSTARS.2024.3464691

median value of sea ice extent from 1981 to 2010. The decreasing extent of Arctic sea ice has led to a year-on-year extension of the navigable window, especially in the Northeast Route, increasing the commercial value of the Arctic Route [1]. Compared to traditional shipping routes in China, the opening of the Arctic Route can save 30% to 50% of sailing distance [2], greatly enhancing the potential for maritime trade [3]. However, the harsh navigational conditions in the Arctic pose significant challenges to vessels traversing the Arctic Routes. From the current research on safety hazards and maritime accidents in the Arctic navigation, sea ice stands out as one of the major safety hazards for maritime navigation [4], [5], [6]. Therefore, in the process of monitoring Arctic sea ice, the classification and assessment of sea ice conditions, as well as ice forecasting, are of crucial importance and assistance for polar navigation.

With the development of science and technology, the use of remote sensing imagery for sea ice classification research has become increasingly common. Synthetic Aperture Radar (SAR) is not affected by cloud cover or sunlight, allowing for observations to be conducted under all-day and all-weather conditions, and has been proven suitable for sea ice observations in polar regions [7], [8]. SAR images are formed by receiving the echoes reflected from sea ice, and different types of sea ice are distinguished in the images based on their unique reflection characteristics. From existing research results, C-band multipolarimetric SAR data have been demonstrated to be suitable for sea ice classification research [9], [10], [11], [12], [13], [14], [15].

Currently, one of the simplest and most direct methods for sea ice classification involves using calibrated backscatter coefficients for differentiation. However, since different types of sea ice can exhibit similar backscatter coefficients, this approach may result in unreliable classification outcomes [16]. More advanced classification methods require the incorporation of additional sea ice features, such as texture features [17], [18], scattering intensity [19], [20], and polarimetric characteristics [13], [21], to construct comprehensive feature vectors. These vectors are then used to train various classifiers, including neural networks [21], support vector machines [22], decision trees, random forests [23], and other sophisticated image segmentation methods [24], [25]. Despite these advancements, the manual extraction of features is largely based on expertise and intuition, making it impractical to exhaustively search for the optimal

feature combination. Therefore, there remains a significant need for the continued improvement of sea ice classification methods.

In recent years, CNNs have achieved remarkable success in image recognition and image classification, and have subsequently gained increasing favor in the field of sea ice classification. Some simple classification methods involve constructing a shallow CNN classifier and distinguishing different types of sea ice based on the varying backscatter coefficient intensities in SAR sea ice images. [26], [27], [28]. Subsequently, further improvements to sea ice classification models based on some publicly available deep learning models (such as ResNet [14], [29], [30], [31], VGG-16 [32], LSTM [30], [33], MobileNet [34], AlexNet [35], etc.) have been widely proposed, with their classification performance often surpassing traditional machine learning methods. Nevertheless, most image-based classification models require training samples to be cropped from the original image through a sliding window, a process heavily influenced by subjective preferences. To achieve pixel-level classification, sea ice classification models based on image segmentation have been gradually developed [36], [37], [38], [39]. These models use manually annotating labels that correspond pixel-by-pixel with the training samples, thereby eliminating the cumbersome step of generating training samples via a sliding window approach. In addition, using multisource sea ice data in the sea ice classification process can also improve classification accuracy. Studies by [20], [40], [41] choose to fuse SAR sea ice images and multispectral sea ice image information, while studies by [42], [43] introduces sea ice concentration data into SAR sea ice images.

The problems existing in the SAR sea ice classification field based on deep learning methods are as follows:

- 1) SAR sea ice images are difficult to interpret. Although Arctic coastal countries such as Canada, Russia, Norway, and others regularly release sea ice classification products, these products often have low resolution, making them difficult to match with conventional SAR images.
- 2) The classification model needs improvement. Hindered by the difficulty in acquiring ground truth sea ice data, it is challenging to establish authoritative and reliable datasets. Deep learning methods driven by data often struggle to find utility. Existing models typically categorize small sea ice slices based on image recognition, which fragments the contextual information in the images, leading to subjectivity and lack of persuasiveness.
- 3) The sources of data used are relatively limited. Most existing sea ice classification studies based on deep learning utilize imagery data from a single satellite.

Increasing data diversity aids in validating whether classification models are equally applicable to other satellite data sources.

The research objective of this article is to combine state-of-the-art deep learning algorithms to propose a highly reliable Arctic Route sea ice classification algorithm based on deep learning models and validate its classification performance using different SAR satellites sea ice image data. The research steps of this article are as follows:

- 1) 1) Select GF3 SAR sea ice images and Landsat-8 multispectral sea ice images from the same location and close

timeframes. Crop the overlapping regions based on latitude and longitude, and label the SAR sea ice images using the optical sea ice images to obtain reliable high-resolution sea ice ground truth data, thereby creating the sea ice dataset.

- 2) Choose an image segmentation-based deep learning backbone and integrate it with the currently state-of-the-art deep learning algorithms. Improve the existing models and propose a highly reliable model named Sea Ice CNN and Transformer Fusion Network (SI-CTFNet) for polar sea ice classification. Design the proposed model to be lightweight based on practical application scenarios to enhance its feasibility and efficiency in real-world applications.
- 3) Test the model's classification performance using SAR sea ice images from different modes of the GF3 satellites.
- 4) Evaluate the proposed method's applicability to other C-band satellite SAR sea ice images using Sentinel1A satellite images.

As is well known, CNNs are good at capturing local features of images but are insensitive to location information. Transformers [44] were initially applied in the NLP domain and later migrated to the field of image recognition, resulting in the ViT [45] model. ViT specializes in modeling the global information of the entire image, thereby improving the understanding of its structure and semantics, although it is insensitive to local features. To enjoy the benefit of both, considerable efforts have been made to integrate CNNs with ViTs. Inspired by the TransFuse [46] model, we borrowed and improved upon its overall structure to propose a novel image segmentation network for sea ice classification called SI-CTFNet. This model largely extends the framework of CNN and transformer dual-branch parallel feature extraction in the transfuse model, and the specific improvements can be summarized as follows:

- 1) 1) Aiming at the characteristics of poor decipherability of SAR sea ice images, the downsampling scale is reduced in the feature extraction framework of the dual-branch to retain more original information in the image, and the transformer branch is modified to output multiscale feature maps like the CNN branch;
- 2) Aiming at the characteristics of the different categories of sea ice accounting for different proportions of the SAR sea ice images, the MSDAM module is proposed to adaptively capture different sizes of information inside the image by means of the atrous convolution and multiscale dynamic convolution;
- 3) BiAFusion module is designed to realize the effective combination of feature maps acquired by CNNs and ViTs at the same resolution.

Our ultimate goal is to develop an automated sea ice forecasting and route planning system tailored for polar vessels, utilizing satellite-borne SAR sea ice imagery data and deep learning algorithms, in order to reduce maritime accidents caused by sea ice and ensure safe navigation in polar regions. The research findings of this article will be applied to implement automatic sea ice classification functionality within a polar intelligent shipping system. Utilizing SAR satellites such as GF3, LT-1A/B, HJ-2E/F

for polar shipping route sea ice monitoring, real-time processing, and generation of sea ice classification products will provide timely ice condition information for polar navigation, assisting in route planning or modification.

II. DATASETS

A. Satellites Introduction

1) *GF3 Satellites*: The GF3 satellite 01 (GF3-01/GF3A) was launched in 2016, marking China's first high-resolution C-band Synthetic Aperture Radar (SAR) satellite. It features multiple imaging modes and is capable of providing SAR image products with resolutions ranging from 1 to 500 m and swath widths from 10 to 650 km [47], enabling all-weather, all-day monitoring of global marine and terrestrial information. Subsequently, China launched the GF3 satellite 02 (GF3-02/GF3B) and GF3 satellite 03 (GF3-03/GF3C) on 23 November 2021, and 7 April 2022, respectively, achieving networked operations and significantly enhancing China's radar satellite Earth observation capabilities. Among the various operational modes of the GF3 satellites, both Quad-Polarimetric Stripmap I (QPSI) and Quad-Polarimetric Stripmap II (QPSII) provide quad-polarimetric SAR images, while Standard Stripmap (SS), Fine Stripmap I (FSI), and Fine Stripmap II (FSII) offer dual-polarimetric SAR images. In this study, Quad-Polarimetric Stripmap data from the GF3A satellite and Fine Stripmap data from the GF3B satellite were utilized.

2) *Landsat Satellites*: This article utilizes image data from Landsat8 and Landsat9 satellites. The Landsat8 and Landsat9 satellites were launched on 11 February 2013, and 27 September 2021, respectively, as successive satellites in the U.S. NASA's Landsat program. Landsat8/9 is equipped with two primary sensors: the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). OLI has higher spectral resolution and a wider spectral range, allowing it to provide more detailed surface information. The OLI sensor on the Landsat8/9 satellites provides image data for nine bands, including a panchromatic band with a spatial resolution of 15 m for Band 8, and bands 2/3/4 with spatial resolutions of 30 m for the blue, green, and red bands, respectively, which can be used for synthesizing natural-color images. This study utilizes multispectral image data from three groups of Landsat8/9 satellites.

3) *Sentinel1A Satellite*: The Sentinel1 satellite, part of the European Space Agency's Copernicus program (GMES), consists of two satellites, Sentinel1A (S1A) and Sentinel1B (S1B). Sentinel1A was launched on 3 April 2014, equipped with a C-SAR sensor capable of providing dual-polarimetric scan mode image products. This study utilized Sentinel1A dual-polarimetric scan mode image data.

B. Remote Sensing Image Datasets

The study plans to use visible light sea ice images to assist in interpreting SAR sea ice images, hence minimizing the time interval between the GF3 satellites images and Landsat8 satellite images. We searched for GF3 satellites image products and Landsat8 satellite image products from 2020 to present, collecting two sets of sea ice image products suitable for dataset

TABLE I
DETAILED INFORMATION OF THE REMOTE SENSING IMAGES USED IN THE OKHOTSK SEA DATASET

| Satellite | Id | Time | near incidence angle | far incidence angle |
|-----------|-----|----------|----------------------|---------------------|
| GF3A | g1 | 21:05:35 | 35.3336 | 37.1574 |
| | g2 | 21:05:40 | 35.3331 | 37.1586 |
| | g3 | 21:05:45 | 35.3326 | 37.1598 |
| | g4 | 21:05:50 | 35.3321 | 37.1592 |
| | g5 | 21:05:55 | 35.3318 | 37.1588 |
| | g6 | 21:06:00 | 35.3312 | 37.1584 |
| | g7 | 21:06:04 | 35.3308 | 37.1579 |
| | g8 | 21:06:09 | 35.3303 | 37.1573 |
| | g9 | 21:06:14 | 35.3298 | 37.1568 |
| | g10 | 21:06:19 | 35.3294 | 37.1563 |
| Landsat8 | L1 | 01:41:08 | - | - |
| | L2 | 01:41:32 | - | - |

creation. These correspond to the typical strip modes of GF3 satellites, named Okhotsk Sea and Sakhalin1, respectively. In addition, we separately collected a set of sea ice image products containing GF3, Landsat, and Sentinel satellite data, named Sakhalin2, for subsequent experiments.

1) *Okhotsk Sea*: This dataset is from the same pass as the data used in [41], and we supplemented it with four additional scenes of GF3A satellite image products to enrich our dataset. All data in this group were acquired on 28 February 2020, over the Okhotsk Sea area, as shown in Fig. 1(a). The green boxes in Fig. 1(a) represent the coverage of GF3A satellite Quad-Polarimetric Stripmap I (QPSI) SAR sea ice images, numbered g1 to g10, while the red boxes represent the coverage of Landsat8 satellite multispectral sea ice images, numbered L1 and L2. The GF3A satellite SAR sea ice images were acquired between 21:05 and 21:06 (UTC) with a swath width of 30 km and a nominal resolution of 8 m. The Landsat8 multispectral sea ice images were acquired at 01:41 (UTC) with resolutions of 30 and 15 m. Detailed information about this dataset is provided in Table I.

2) *Sakhalin1*: The data in this group were all acquired on 19–20 February 2024, covering the approximate latitude and longitude range of the southern waters of Kuye Island, as shown in Fig. 1(b). The green boxes in Fig. 1(b) represent the coverage of GF3B satellite dual-polarimetric Fine Stripmap I (FSI) SAR sea ice images, numbered g13-g20, while the red boxes represent the Landsat8 multispectral sea ice images, numbered L3-L5. The SAR sea ice images from the GF3B satellite were acquired between 20:37 and 20:38, operating in Fine Stripmap mode with a swath width of 50 km and a nominal resolution of 5 m. The Landsat8 multispectral sea ice images were acquired between 01:11 and 01:12 with resolutions of 30 and 15 m. Detailed information about this dataset is provided in Table II.

3) *Sakhalin2*: The data in this group were all captured on 10 February 2024, in the northern waters of the Sea of Okhotsk, as illustrated in Fig. 1(c). The green box in Fig. 1(c) represents the SAR sea ice image captured by the GF3B satellite in dual-polarimetric Fine Stripmap II Mode (FSII) at 8:08, with a swath width of 100 km and a nominal resolution of 10 m. The red box in Fig. 1(c) indicates the Landsat8 multispectral sea ice image

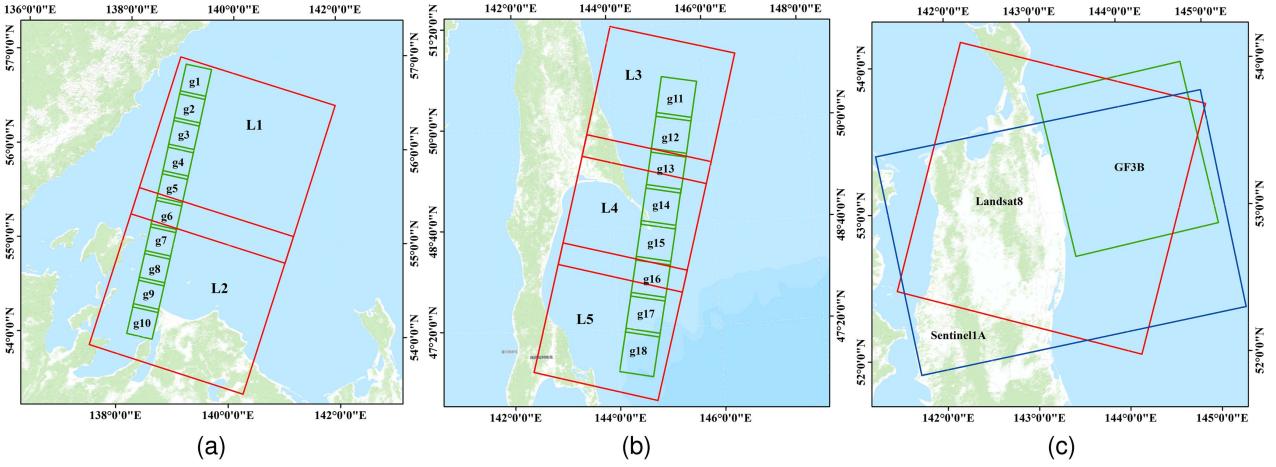


Fig. 1. Coverage of remote sensing images used in this article, The red boxes represent Landsat satellites images, the green boxes represent GF3 satellites images, and the blue box represents Sentinel1A satellite images. (a) Coverage of remote sensing images used in Okhotsk Sea Dataset. (b) Coverage of remote sensing images used in Sakhalin1 Dataset.(c) Coverage of remote sensing images used in Sakhalin2 Dataset.

TABLE II
DETAILED INFORMATION OF THE REMOTE SENSING IMAGES USED IN THE
SAKHALIN1 DATASET

| Satellite | Id | Time | near incidence angle | far incidence angle |
|-----------|-----|----------|----------------------|---------------------|
| GF3B | g11 | 20:37:08 | 31.4137 | 34.1761 |
| | g12 | 20:37:16 | 31.4131 | 34.1888 |
| | g13 | 20:37:24 | 31.4124 | 34.2013 |
| | g14 | 20:37:32 | 31.4117 | 34.2139 |
| | g15 | 20:37:40 | 31.4110 | 34.2263 |
| | g16 | 20:37:48 | 31.4103 | 34.2388 |
| | g17 | 20:37:56 | 31.4098 | 34.2513 |
| | g18 | 20:38:04 | 31.4092 | 34.2643 |
| | L3 | 01:11:54 | - | - |
| Landsat8 | L4 | 01:12:18 | - | - |
| | L5 | 01:12:42 | - | - |

TABLE III
DETAILED INFORMATION OF THE REMOTE SENSING IMAGES USED IN THE
SAKHALIN2 DATASET

| Satellite | Id | Time | near incidence angle | far incidence angle |
|------------|-----|----------|----------------------|---------------------|
| GF3B | g19 | 08:08:21 | 31.4129 | 38.1988 |
| Landsat9 | L6 | 01:23:43 | - | - |
| Sentinel1A | S1 | 08:14:54 | 30.3756 | 45.7353 |

captured at 1:23, while the blue box represents the Sentinel1A dual-polarimetric scan mode SAR sea ice image captured at 8:14, operating in the Interferometric Wide Swath (IW) mode with a swath width of 250 km and a resolution of 20 m. Detailed information for this dataset is provided in Table III. The purpose of using Sentinel-1 satellite data was to provide a comparison experiment with data from different C-band satellites to validate the applicability of this method across various satellite datasets.

C. Data Preprocessing

The Landsat8 multispectral sea ice images were processed for image fusion using ENVI 5.3 software. The high-resolution

image was the panchromatic image from the eighth band, while the low-resolution image was a natural color image composed of bands 4/3/2. After fusion, a natural color sea ice image with a spatial resolution of 15 m was obtained. The GF3 satellites SAR image products were single-look complex (SLC) images. Initially, each pixel position in the image was taken modulus to obtain amplitude image data. Subsequently, the amplitude images, lacking geographic coordinate information, were uniformly projected onto the UTM coordinate system using the provided RPC files from the GF3 satellites products. Strictly speaking, since the backscatter intensity in SAR images is influenced by the incidence angle [48], [49], [50], [51], many studies include an additional step for incidence angle correction during data processing [17], [22]. Since the range of incidence angles in the data used in this study is relatively small, the incidence angle correction step was omitted during the preprocessing of SAR images.

Thus, we obtained both low-resolution optical sea ice images and high-resolution SAR sea ice images in the same coordinate system. As shown in Fig. 1, there is a considerable overlap between the visible light sea ice images and SAR sea ice images, which can be cropped using the GDAL library. Initially, the four corner points' latitude and longitude of the two images to be cropped are obtained. The latitudes and longitudes of the four corner points of the overlapping quadrilateral are determined using the two sets of coordinates. The cropping function provided by the GDAL library is then used, employing the quadrilateral's corner points' latitude and longitude to crop both images simultaneously, while also resampling both images to the same spatial resolution.

D. Data Labeling

According to the classification standards established by the World Meteorological Organization (WMO) for sea ice conditions in the polar and high-latitude regions, different classifications of sea ice are determined, with those based on ice thickness

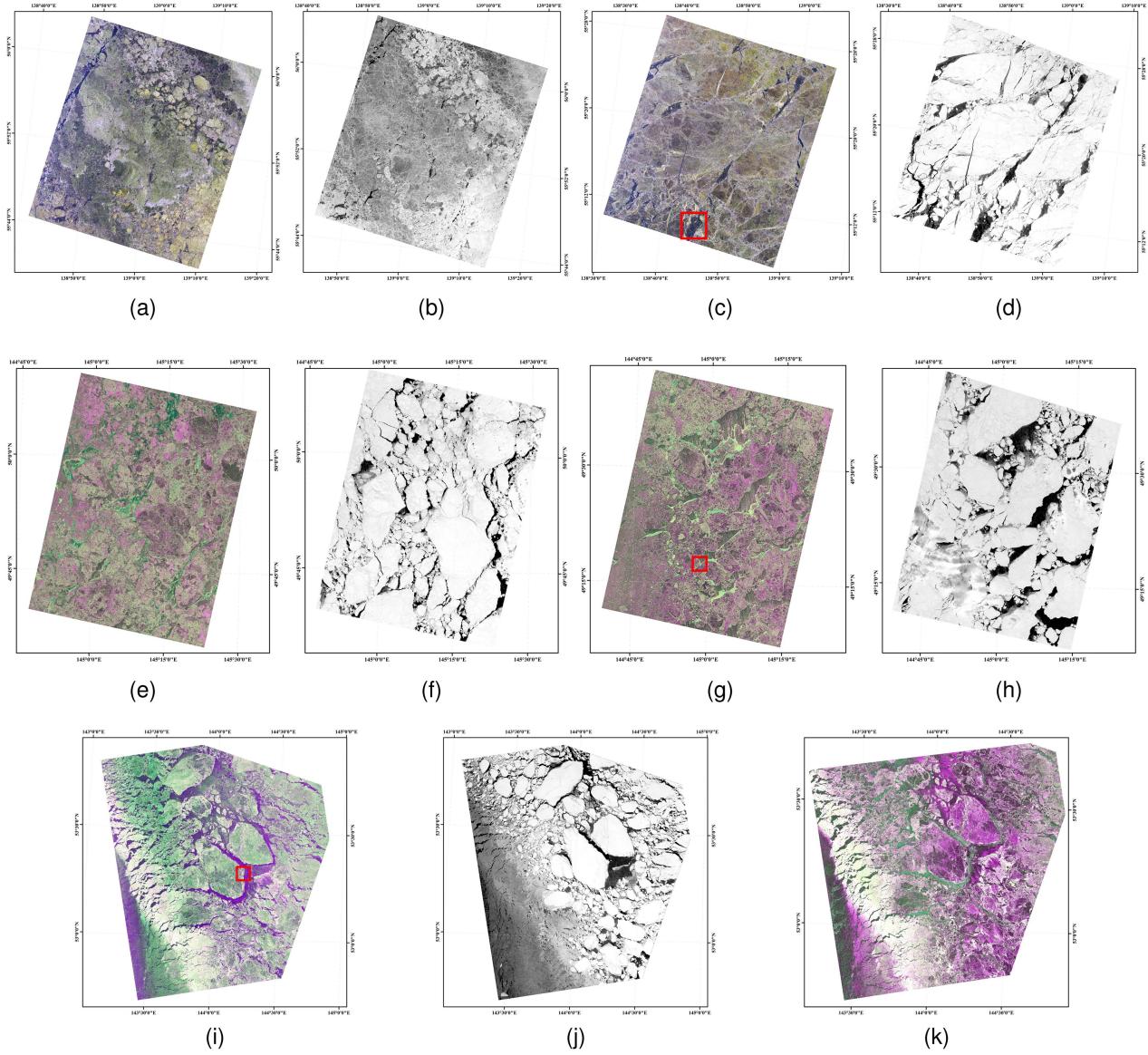


Fig. 2. Example scenes used in this article. (a) and (c) correspond to g6 and g8 ($R = HH, G = HV, B = VV$), (b) and (d) are cropped area corresponding to (a) and (c), respectively. (e) and (g) correspond to g12 and g13 ($R = HH, G = HV, B = \sqrt{|R^2 - G^2|}$), (g) and (h) are cropped area corresponding to (e) and (g), respectively. (i) and (k) correspond to g19 and S1 ($R = HH, G = HV, B = \sqrt{|R^2 - G^2|}$), respectively. (j) is cropped area corresponding to (i).

widely adopted. This article primarily adopts the sea ice category names defined by the WMO. By referencing the distinct characteristics of different sea ice categories as depicted in Landsat8 optical sea ice images, the sea ice is classified into four categories across three datasets. In the Okhotsk Sea dataset and Sakhalin2 dataset, the four categories of sea ice are classified as Open Water (OW), New Ice (NI), Young Ice (YI), and First-Year Ice (FYI), as shown in Fig. 3(a) and Fig. 3(c). In the Sakhalin1 dataset, due to the presence of numerous small pieces of sea ice resembling first-year ice and a lack of clear differentiation between sea water and new ice, during the annotation process, sea water and new ice were grouped into one category, and an additional category of floe ice was added. The four categories of sea ice are New Ice (NI), Young Ice (YI), First-Year Ice (FYI), and Floe Ice (FI), as demonstrated in the annotation example in Fig. 3(b).

From Landsat optical images, it can be observed that areas of OW and NI are mostly darker, while YI can be further classified into Gray Ice and Gray-White Ice, corresponding to gray and gray-white regions in the optical images, respectively. FYI, also known as White Ice, appears as white regions in optical images, and FI areas often consist of various types of small ice blocks intermixed. In SAR images, the characteristics of OW and NI are similar to those in optical images, appearing as relatively dark or smooth areas due to their smooth surfaces, resulting in lower backscatter coefficients. YI exhibits the highest backscatter intensity among these four types of sea ice and often appears as the brightest areas in SAR images. The structures of FYI and FI are more complex, typically containing multiple scattering types, resulting in a more intricate pattern in SAR images. FI generally exhibits slightly higher backscatter intensity compared to FYI.

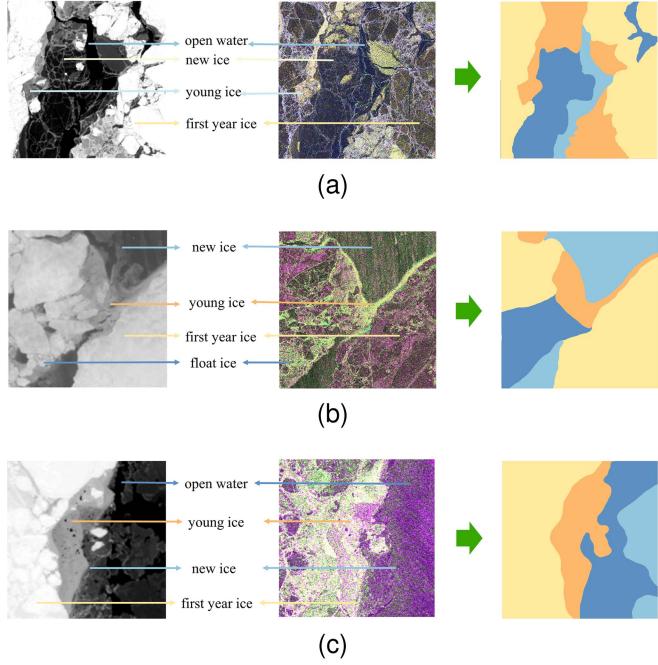


Fig. 3. Different types of sea ice in coincident images. The area is an enlargement of the red rectangular areas in (a) Fig. 2(c), (b) Fig. 2(g) and (c) Fig. 2(i). In (a), (b) and (c), the left images are Landsat satellites optical images, the middle images are GF3 satellites SAR images, and the right images are manually annotating labels.

Fig. 3 illustrates three representative image samples, along with the corresponding manually annotating results at sea ice junctions, which align with the locations marked by red rectangles in Fig. 2. The images in the left column are a portion of the sea ice images cropped from Landsat8 satellite images based on the latitude and longitude region, while the images in the middle column are false color images obtained from the combination of GF3 multipolarimetric SAR images.

III. METHODOLOGY

A. Overall Structure

SI-CTFNet consists of three branches, as shown in Fig. 4: the CNN branch is derived from deeplabV3+ [52], the transformer branch is derived from Segformer [53], and the fusion branch consists of BiAFusion modules, which inputs the feature maps of CNN and ViT corresponding layers into the fusion network to obtain the fused feature map. At the output end, we set three learnable weights α , β , and γ . The three branches independently complete the segmentation of the input image, and then multiply and add them with these three weights. Finally, the segmentation result is obtained through a convolutional block. The above process can be expressed by

$$F_{out} = Conv(\alpha F_{cnn} + \beta F_{fuse} + \gamma F_{vit}) \quad (1)$$

$$\alpha + \beta + \gamma = 1 \quad (2)$$

where F_{out} represents the final output prediction result, F_{cnn} , F_{fuse} , and F_{vit} , respectively, denote the prediction results of the

three branches. α , β , and γ are three learnable weight parameters, and Conv represents the convolutional block operation.

The input of CNN branch is an image of size $H \times W \times 3$, with the backbone network being a modified ResNet18. ResNet18 consists mainly of four residual blocks, as illustrated by the orange dashed box in Fig. 4. Each residual block has a convolutional kernel size of 3, with a stride of 2 for the first convolutional layer and a stride of 1 for the second convolutional layer. Consequently, after each residual block operation, the dimensions of the input feature map are halved, resulting in feature map dimensions of $H/2 \times W/2$, $H/4 \times W/4$, $H/8 \times W/8$, and $H/16 \times W/16$ for each residual block's output. The feature map outputted by the last residual block is inputted into the MSDAM, which is an improvement upon the ASPP module in the original DeeplabV3+. This module incorporates channel attention mechanism and multiscale feature fusion module, as detailed in Section III-B. The feature map outputted by the MSDAM, along with the shallow feature maps outputted by the other three residual blocks, serves as the input to the decoder. The decoder structure is not specifically designed and directly adopts progressive upsampling (PUP) method, as explained in Section III-D. The output of the decoder is adjusted to 32 channels through a 1×1 convolutional layer, and multiplied by the learnable parameter α to obtain the output of this branch.

Unlike ViT that can only generate a single-resolution feature map, the goal of this transformer branch is, given an input image, to generate CNN-like multilevel features. In transformer branch, we use cascading ViTs similar to that in the Segformer model to extract both high-resolution fine features and low-resolution coarse features. Initially, the input raw image is divided into fixed-size image patches. Each patch contains local information from the image. Each image patch is linearly transformed into a low-dimensional feature space, along with the addition of positional encoding information, resulting in $X = \{x_1, x_2, \dots, x_n\}$, where n is the number of divided image patches, and x_i is the vector corresponding to the i th patch after encoding. Subsequently, X is multiplied by three different transformation matrices to obtain three vectors Q , K , and V , as shown in the following:

$$(Q, K, V) = (XW_q, XW_k, XW_v) \quad (3)$$

where W_q , W_k , and W_v represent the weight matrix of linear transformation used to generate query, key, and value tensors, respectively. Then, Q , K , and V are fed into the multihead self-attention module to compute the self-attention feature, which is formulated as follows:

$$Attention = \text{Softmax} \left(\frac{Q \times K^T}{\sqrt{d_k}} \right) \times V \quad (4)$$

where d_k is the channel dimension of K , and division by $\sqrt{d_k}$ can be considered as an approximate normalization. The Softmax function is calculated for each row in the matrix. Finally, we adapt a backbone similar to that in Segformer to implement the feature extraction network of this branch, which comprises four cascaded transformer blocks. The input to each transformer block is the feature map after patching and positional encoding. For the decoding part, we use a lightweight All-MLP module,

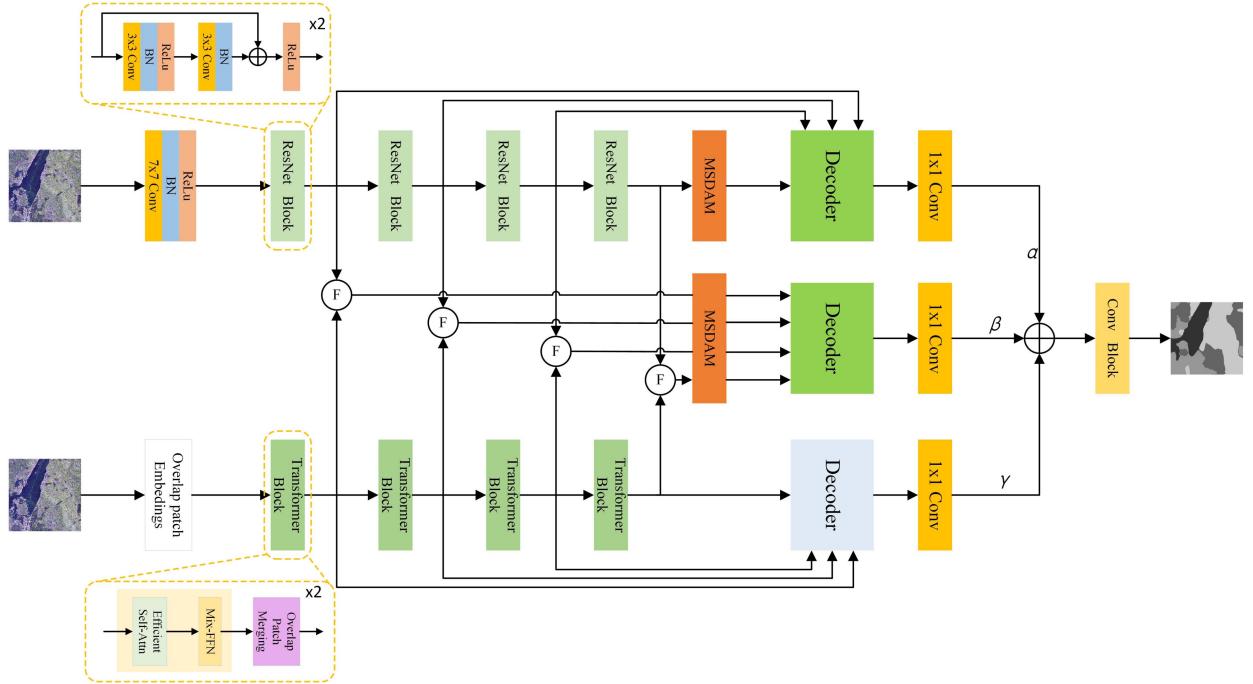


Fig. 4. Overall network structure of the SI-CTFNet.

which integrates feature maps from multiple scales to obtain the prediction results.

The fusion branch comprises four BiAFusion modules (the circled F in Fig. 4), each receiving the feature maps outputted by the corresponding layers in the CNN branch and transformer branch, to accomplish the fusion of the two types of feature maps. Following the four BiAFusion modules, the feature maps undergo further extraction of deep features through the MSDAM module, and the results are simultaneously inputted into the decoder. Subsequently, the structure of this branch is the same as that of the CNN branch, and the output result is multiplied by the learnable parameter β to obtain the output of this branch. The specific structure of the BiAFusion module is detailed in Section III-C.

B. MSDAM Module

Global features play a crucial role in semantic segmentation. To further extract global features, we designed the Multi-Scale Dynamic Attention Module (MSDAM), the detailed structure of which is shown in Fig. 5(a).

In the original DeeplabV3+ network, this position corresponds to the Atrous Spatial Pyramid Pooling (ASPP) module, which enlarges the receptive field by setting different dilation rates, capturing more contextual semantic information without losing resolution. However, excessively large dilation rates may lead to information loss in small targets, resulting in grid effects and boundary effects. Therefore, based on the original ASPP module, we added a multiscale module to dynamically segment the semantics using filters of different scales and capture multi-scale semantic information. The module takes an input feature map of size $H \times W \times C$ and consists of two branches. One

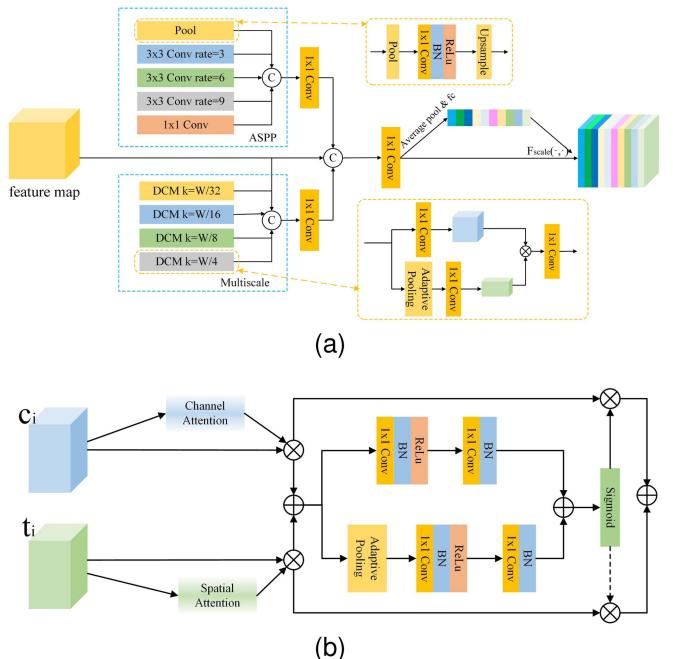


Fig. 5. (a) Overall network structure of MSDAM Module. (b) Overall network structure of BiAFusion Module.

branch is the ASPP module from the original DeeplabV3+, where we reduced the original dilation rates to $\{3, 6, 9\}$ to mitigate information loss issues while keeping the other structures unchanged. The other branch is the multiscale module, comprising four Dynamic Convolutional Modules (DCM) with variable k values. Each DCM splits into two subbranches: one subbranch reduces the feature map dimension to $H \times W \times C$ through a

1×1 convolution, while the other subbranch transforms the feature map to $k \times k \times C$ through adaptive pooling and a 1×1 convolution, followed by fusion through depthwise conv operation. The outputs of the ASPP module and the multiscale module are concatenated with the original input feature map, resulting in a feature map of size $H \times W \times (C \times 3)$, which is then dimensionally reduced through a 1×1 convolutional module. Finally, a channel attention module is applied, obtaining weights for each channel through global pooling and fully connected layers, which are then multiplied with the original feature map channel-wise to obtain the final output feature map.

C. BiAFusion Module

The structure of the BiAFusion module is illustrated in Fig. 5(b). This module takes two inputs, f_c and f_t , where f_c represents the feature map from the convolutional neural network (CNN), and f_t represents the feature map from the vision transformer. We apply channel attention mechanism to f_c and spatial attention mechanism to f_t , and then combine them to obtain a feature map with both local features and global semantic information. Since the importance of the two types of feature maps may not be entirely equal, we introduce a trainable weight calculation module. This module takes the fused feature map as input and computes global features and local features separately through two branches. Subsequently, the outputs are summed and passed through a sigmoid function to compress the result between 0 and 1. The weighted f_c with channel attention mechanism is then multiplied by this weight, and the result is subtracted from 1 and multiplied by f_t with spatial attention mechanism. Finally, the products of the two paths are added together to obtain the final output. This process can be described as

$$f_{out} = wCA(f_c) + (1 - w)SA(f_t) \quad (5)$$

$$w = \sigma(\mathbb{L}(CA(f_c) + SA(f_t)) + \mathbb{G}(CA(f_c) + SA(f_t))) \quad (6)$$

where CA represents channel attention operation, SA represents spatial attention operation, σ denotes sigmoid operation, \mathbb{L} and \mathbb{G} represent obtaining local features and global features respectively, and f_c and f_t are the input feature maps. It should be noted that the extraction of global features is obtained by the 1×1 convolution module, while the extraction of local features is obtained by the adaptive pooling module and then the 1×1 convolution module.

D. Decoder

As is well known, shallow features contain rich spatial detail information, such as color, texture, and edges, but lack deep semantic information, while deep features contain rich semantic information but have lower spatial resolution. Therefore, organically combining features from different scales often effectively improves segmentation accuracy. In this article, no further improvements were made to the decoder. Instead, features from different scales were simply inputted into the decoder simultaneously, and multiscale feature fusion was achieved through

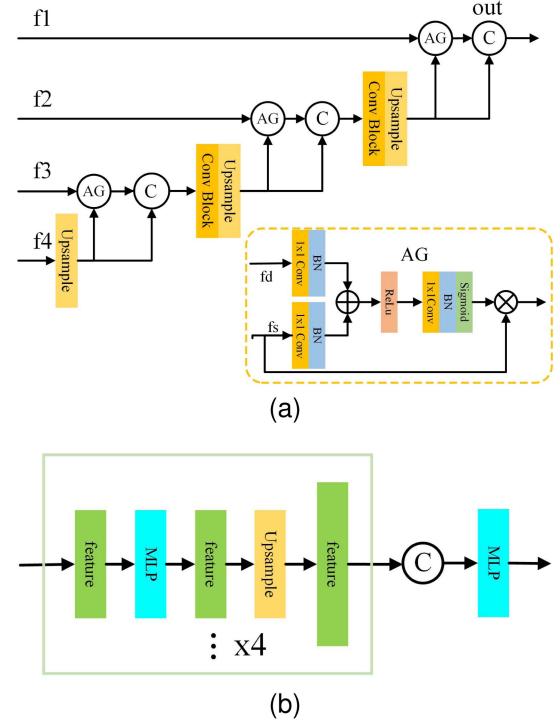


Fig. 6. Overall network structure of decoder of (a) CNN branch and fusion branch (b) transformer branch.

attention gates and upsampling, as shown in Fig. 6(a). Assuming input deep features f_d and shallow features f_s , f_d , and f_s are firstly input into the attention gate to obtain the fused feature map. The fused feature map is then concatenated with the deep feature map, and the channel dimension is adjusted through a 1×1 convolution module, which is repeated with the next shallow feature. The above process can be represented by

$$f_{out} = \text{Conv}(\text{cat}(f_d, \text{AG}(f_d, f_s))) \quad (7)$$

where f_d and f_s represent the input deep features and shallow features, respectively, cat denotes concatenation along the channel dimension, and AG stands for attention gate operation.

E. Lightweight Designed

The primary motivation for the lightweight design is to enhance the model's inference speed to meet the demands of time-sensitive application scenarios. This study does not implement targeted lightweight design for each module within SI-CTFNet. Instead, the inference speed is improved by reducing the number of parameters and decreasing the number of model branches. The structure of the lightweight modified model is shown in Fig. 7.

The lightweight SI-CTFNet largely retains the overall structure of the original SI-CTFNet, with the following main modifications: 1) An additional downsampling step is introduced in the feature extraction network, reducing the dimensions of the feature maps at the corresponding scales in the original network by half. 2) The number of feature map branches input to the decoder is reduced, using only two scales of feature maps. One branch, directly input to the MSDAM or MLP module, represents feature maps with dimensions of $H/16 \times W/16 \times C$. The other branch

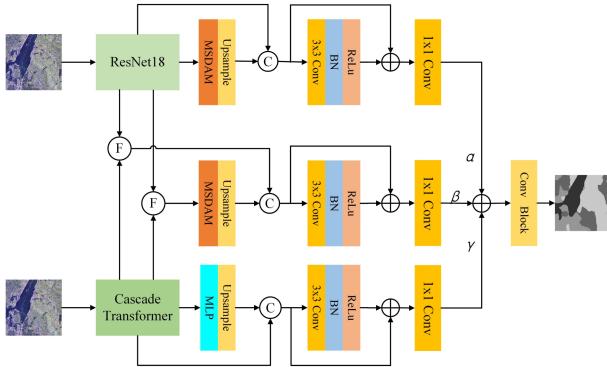


Fig. 7. Overall network structure of the lightweight SI-CTFNet.

represents feature maps with dimensions of $H/4 \times W/4 \times C$. 3) In the decoder, the two scales of feature maps are directly concatenated through upsampling, followed by a convolutional residual block.

IV. EVALUATION METHOD AND EXPERIMENTAL SETTING

A. Evaluation Metrics

We used five metrics to evaluate the classification performance of the model: PA(Pixel Accuracy), CPA(Class Pixel Accuracy), mPA(mean Pixel Accuracy), IoU(Intersection over Union), and mIoU(mean Intersection over Union). PA represents the proportion of correctly classified pixels in all categories to all pixels. CPA represents the proportion of correctly classified pixels in a single category to all pixels. mPA is the average of CPA. IoU represents the overlap ratio between the predicted region and the ground truth region for a single category, i.e., the ratio of their intersection to their union. mIoU is the average of IoU. The formulas for calculating these metrics are given by

$$PA = \frac{\sum_{k=1}^N TP_k + TN_k}{\sum_{k=1}^N TP_k + TN_k + FP_k + FN_k} \quad (8)$$

$$CPA_k = \frac{TP_k}{TP_k + FP_k}, k \in \{1, 2, \dots, N\} \quad (9)$$

$$mPA = \frac{1}{N} \sum_{k=1}^N CPA_k \quad (10)$$

$$IoU_k = \frac{TP_k}{TP_k + FP_k + FN_k}, k \in \{1, 2, \dots, N\} \quad (11)$$

$$mIoU = \frac{1}{N} \sum_{k=1}^N IoU_k \quad (12)$$

where TP_k , FP_k , TN_k , and FN_k denote true positives, false positives, true negatives, and false negatives, respectively, for object indexed as class k . N is the number of classes.

B. Experiments Setups

We employed a sliding window approach with a size of 512×512 to select training samples on each of the three datasets,

and manually annotating them with reference to optical sea ice images at corresponding locations. Due to the significant workload involved in manual annotation, we ultimately selected 600 images of size 512×512 from each dataset for the final experimental data, with 500 images used for training and 100 images for testing.

During the process of data sample collection, to mitigate the impact of sample imbalance, we employed two strategies. First, due to the intricate and overlapping boundaries of different ice categories, we computed edge intensity using the Sobel operator for each selected sample, aiming to choose regions with higher intensity as training samples to ensure they encompass multiple ice categories. Second, we manually selected a small number of samples for training and utilized a random forest algorithm to train a simple classifier. This classifier was then used to roughly estimate the proportion of each ice category in each selected sample, striving to maintain balance in the sample quantities within the dataset. In addition, we augmented the dataset using data augmentation techniques. All experiments were trained and validated on a single NVIDIA RTX 3090Ti GPU using the PyTorch framework to ensure a fair comparison.

V. EXPERIMENT RESULTS AND ANALYSIS

A. Comparison With State-of-The-Art Methods on Okhotsk Sea Dataset

Table IV displays the comparative performance of various classification methods tested on the Okhotsk Sea dataset. The data demonstrate that our proposed method outperforms existing approaches in three key metrics—PA, mPA, and mIoU, with scores of 90.413%, 90.437%, and 82.772%, respectively, which are 2.431%, 2.384%, and 1.777% higher than the second-best results. In addition, our method achieved the highest classification accuracy for NI and YI and recorded the best IoU across all categories, but it did not perform as well as ResUnet in the classification accuracy of OW and FYI. However, ResUnet's frequent misclassifications in two classes reduced its IoU scores for OW and FYI.

Fig. 8 illustrates the comparative segmentation results of our method against other widely used methods, highlighting the differences. To visualize the differences between our method and other popular methods, we show the visualization results for each method in Fig. 8, which demonstrates that SI-CTFNet produces superior segmentation results compared to alternative methods. From the segmentation results, it can be observed that SI-CTFNet delineates boundaries more similarly to our manually labeled ground truth. When encountering smaller annotated regions, it achieves more complete delineation of their contours. Moreover, when dealing with larger annotated regions, it minimizes the occurrence of misclassifying other categories within them.

B. Comparison With State-of-The-Art Methods on Sakhalin1 Dataset

We retrained the model on another dataset and conducted identical tests; the results are presented in Table V. From the

TABLE IV
QUANTITATIVE COMPARISON WITH STATE-OF-THE-ART MODELS ON THE OKHOTSK SEA DATASET

| Model | CPA | | | | IoU | | | | PA | mPA | mIoU |
|-----------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | OW | FYI | NI | YI | OW | FYI | NI | YI | | | |
| Unet [54] | 94.651 | 78.387 | 79.273 | 86.200 | 84.307 | 66.291 | 67.370 | 75.522 | 84.453 | 84.628 | 73.373 |
| TransFuse [46] | 93.249 | 81.146 | 80.262 | 88.380 | 85.729 | 68.600 | 68.866 | 78.255 | 85.798 | 85.759 | 75.363 |
| Transunet [55] | 94.404 | 75.006 | 71.843 | 84.926 | 83.094 | 60.390 | 60.272 | 72.698 | 81.507 | 81.545 | 69.114 |
| Attunet [56] | 94.264 | 76.154 | 76.940 | 86.087 | 83.730 | 64.035 | 63.906 | 74.791 | 83.283 | 83.361 | 71.615 |
| Resunet [57] | 97.463 | 88.039 | 57.428 | 67.433 | 78.552 | 57.438 | 48.819 | 63.134 | 76.134 | 77.591 | 61.986 |
| DeeplabV3+ [52] | 93.164 | 83.024 | 86.684 | 89.341 | 88.944 | 71.286 | 75.224 | 80.526 | 87.982 | 88.053 | 78.995 |
| Segformer [53] | 93.309 | 81.299 | 86.096 | 90.501 | 87.219 | 71.585 | 74.867 | 80.617 | 87.858 | 87.801 | 78.572 |
| Ours | 94.242 | 86.946 | 88.958 | 91.601 | 90.291 | 76.963 | 79.750 | 84.082 | 90.413 | 90.437 | 82.772 |

The best values in the columns are in bold. All scores are expressed as percentages (%).

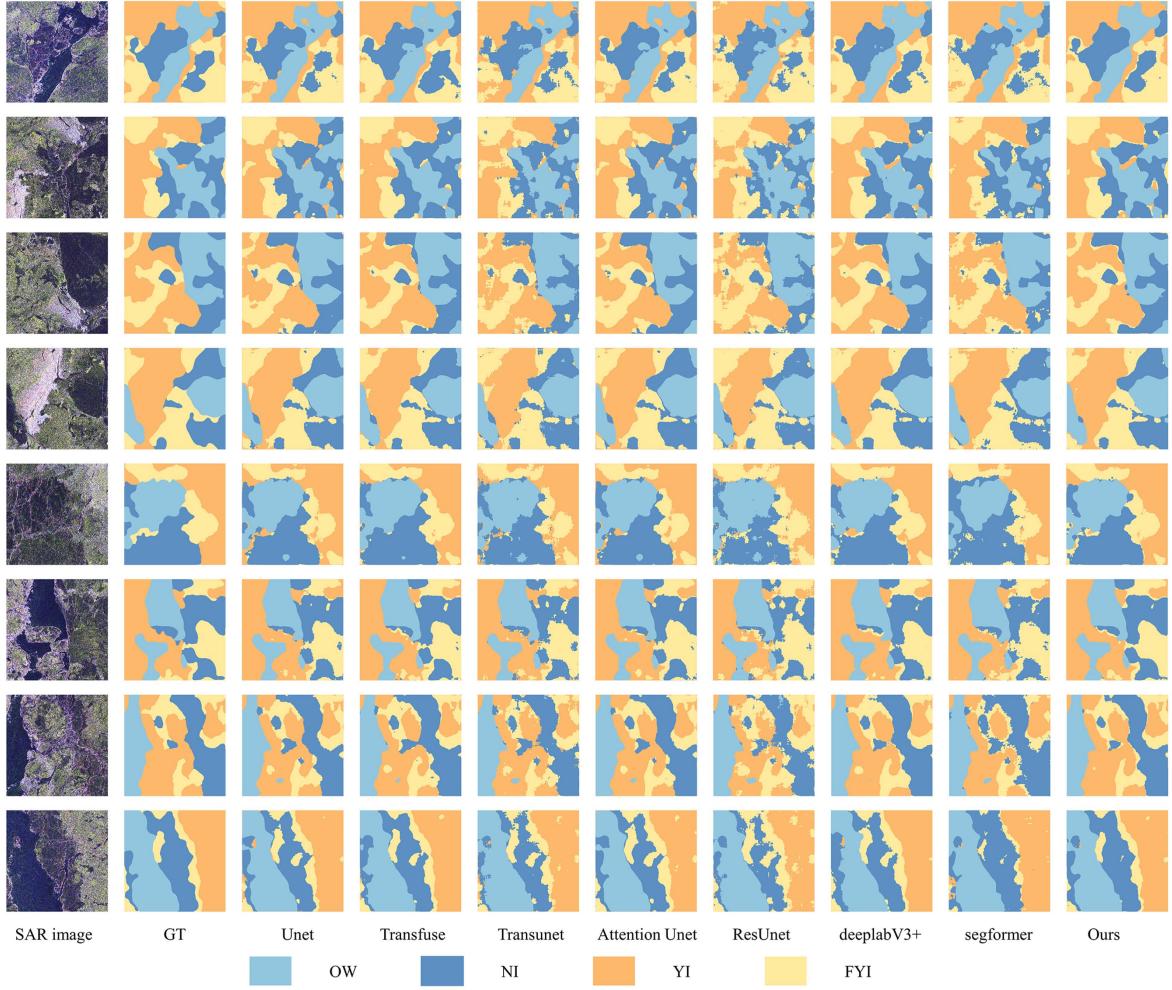


Fig. 8. Qualitative comparison of the visualization results of our method with other methods on the Okhotsk Sea dataset.

experimental results, it can be observed that our proposed method achieves the best performance in terms of PA, mPA, and mIoU, with scores of 88.575%, 89.102%, and 80.300%, respectively. These scores are 1.650%, 1.485%, and 2.519% higher than the second-best results. In this test, our proposed method outperformed all other methods in all metrics except for

OW classification accuracy, where it lagged behind ResUnet. However, the test results on the Sakhalin1 dataset were generally lower than those on the Okhotsk Sea dataset.

Also, to visualize the differences between our method and other popular methods, we show the visualization results for each method in Fig. 9. Similar to the test results on the Okhotsk

TABLE V
QUANTITATIVE COMPARISON WITH STATE-OF-THE-ART MODELS ON THE SAKHALIN1 DATASET

| Model | CPA | | | | IoU | | | | PA | mPA | mIoU |
|-----------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | OW | FYI | FI | YI | OW | FYI | FI | YI | | | |
| Unet [54] | 90.456 | 84.491 | 83.401 | 90.175 | 82.605 | 71.448 | 73.69 | 81.142 | 86.57 | 87.130 | 77.221 |
| TransFuse [46] | 89.518 | 85.029 | 83.230 | 89.933 | 82.963 | 71.001 | 73.188 | 81.379 | 86.434 | 86.927 | 77.133 |
| Transunet [55] | 89.716 | 76.887 | 76.882 | 86.322 | 78.468 | 61.704 | 64.731 | 75.433 | 81.437 | 82.452 | 70.084 |
| Attunet [56] | 89.386 | 84.513 | 83.833 | 89.553 | 81.945 | 71.245 | 73.558 | 80.722 | 86.357 | 86.821 | 76.867 |
| Resunet [57] | 92.515 | 83.969 | 73.807 | 80.333 | 79.565 | 64.079 | 63.465 | 74.450 | 81.579 | 82.656 | 70.390 |
| DeeplabV3+ [52] | 91.272 | 83.758 | 83.373 | 90.090 | 83.380 | 70.970 | 72.656 | 82.064 | 86.484 | 87.123 | 77.268 |
| Segformer [53] | 91.747 | 82.298 | 84.032 | 92.392 | 82.171 | 70.081 | 74.760 | 84.113 | 86.925 | 87.617 | 77.781 |
| Ours | 91.895 | 85.522 | 86.104 | 92.889 | 84.774 | 74.370 | 77.162 | 84.892 | 88.575 | 89.102 | 80.300 |

The best values in the columns are in bold. All scores are expressed as percentages (%).

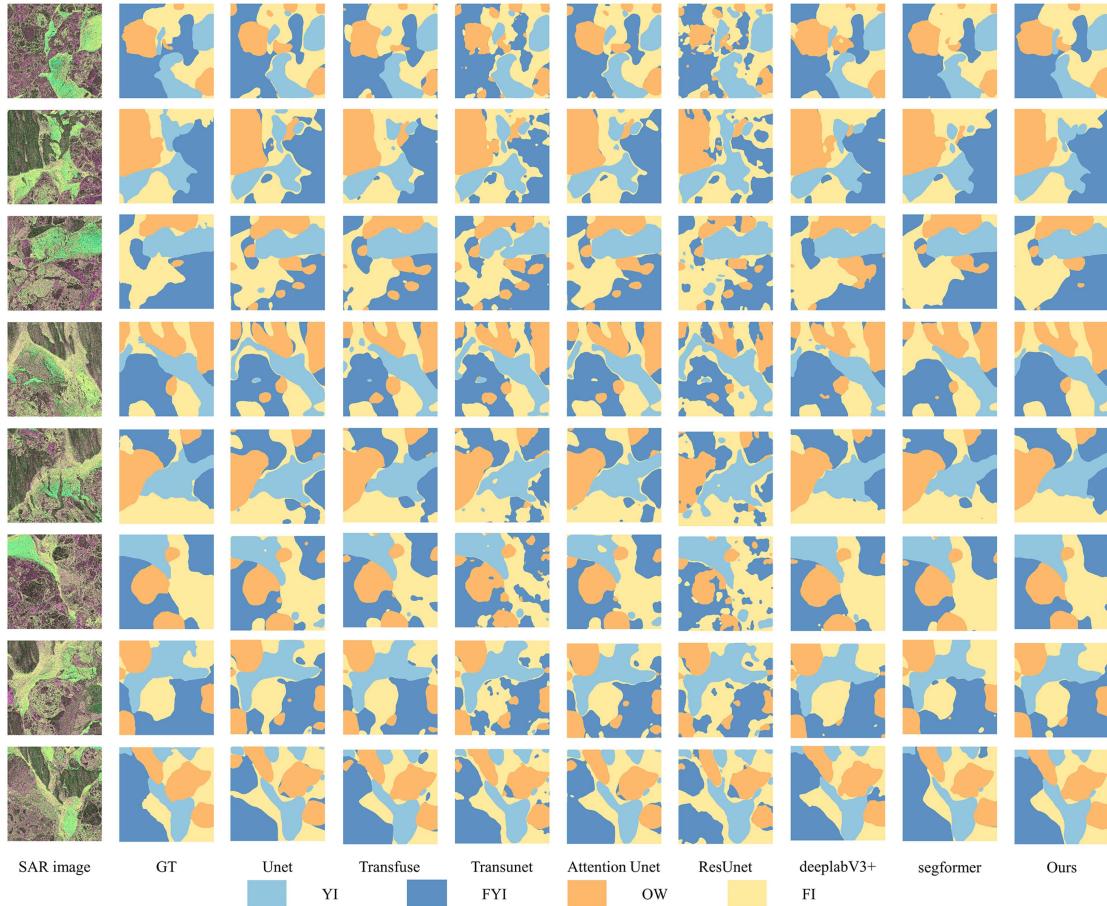


Fig. 9. Qualitative comparison of the visualization results of our method with other methods on the Sakhalin1 dataset.

Sea dataset, our method captures more details on the boundaries between different categories and achieves segmentation results closest to manual annotations.

C. Ablation Experiments on Okhotsk Sea and Sakhalin1

To evaluate the performance of each module of the proposed SI-CTFNet, ablation experiments were carried out on the Okhotsk Sea dataset and Sakhalin1 dataset. The effectiveness

of each module was tested by overlaying it in turn. The experimental results are shown in Tables VI and VII, respectively. Fig 10 shows the visualization results of two sets of ablation experiments.

1) *Baseline*: The Baseline is formed by removing the MSDAM module, BiAFusion module, and Transformer branches from SI-CTFNet, utilizing the remaining components. It employs ResNet18 as the encoder and the module depicted in Fig. 6(a) as the decoder. As evident from Tables VI and VII,

TABLE VI
ABLATION STUDY OF EACH COMPONENT OF THE SI-CTFNET ON THE OKHOTSK SEA DATASET

| Model | CPA | | | | IoU | | | | PA | mPA | mIoU |
|------------------------------|---------------|--------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | OW | FYI | NI | YI | OW | FYI | NI | YI | | | |
| Baseline | 94.766 | 82.222 | 77.769 | 89.611 | 85.101 | 69.791 | 68.812 | 79.436 | 86.221 | 86.092 | 75.785 |
| Baseline +MSDAM | 93.885 | 86.998 | 87.018 | 92.514 | 90.164 | 76.900 | 78.579 | 84.082 | 90.232 | 90.104 | 82.431 |
| Baseline +MSDAM +Transformer | 94.259 | 87.29 | 89.677 | 90.465 | 90.014 | 76.859 | 79.651 | 83.73 | 90.275 | 90.423 | 82.564 |
| SI-CTFNet | 94.242 | 86.946 | 88.958 | 91.601 | 90.291 | 76.963 | 79.750 | 84.082 | 90.413 | 90.437 | 82.772 |

The best values in the columns are in bold. All scores are expressed as percentages (%).

TABLE VII
ABLATION STUDY OF EACH COMPONENT OF THE SI-CTFNET ON THE SAKHALIN1 DATASET

| Model | CPA | | | | IoU | | | | PA | mPA | mIoU |
|------------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | OW | FYI | NI | YI | OW | FYI | NI | YI | | | |
| Baseline | 90.765 | 80.347 | 82.848 | 87.391 | 80.131 | 68.156 | 70.456 | 78.599 | 84.646 | 85.338 | 74.335 |
| Baseline +MSDAM | 91.713 | 85.955 | 85.733 | 92.072 | 85.322 | 74.712 | 76.204 | 83.867 | 88.359 | 88.868 | 80.026 |
| Baseline +MSDAM +Transformer | 92.281 | 86.317 | 84.950 | 92.303 | 85.681 | 74.909 | 75.958 | 83.785 | 88.381 | 88.963 | 80.083 |
| SI-CTFNet | 91.895 | 85.522 | 86.104 | 92.889 | 84.774 | 74.370 | 77.162 | 84.892 | 88.575 | 89.102 | 80.300 |

The best values in the columns are in bold. All scores are expressed as percentages (%).

using only the Baseline yields PA, mPA, and mIoU of 86.221%, 86.092%, and 75.785%, respectively, on the Okhotsk Sea dataset, and PA, mPA, and mIoU of 84.646%, 85.338%, and 74.335%, respectively, on the Sakhalin1 dataset.

2) *Baseline+MSDAM*: Based on the baseline, we incorporate the MSDAM module, which consists of dilated convolution and multiscale dynamic convolution modules. This addition enables the model to enlarge the receptive field of convolutional kernels, thereby capturing broader contextual information, while also allowing the model to capture feature information at different scales, thus enhancing feature representation and generalization capability. As shown in Tables VI and VII, the inclusion of the MSDAM module leads to an improvement in PA, mPA, and mIoU by 4.011%, 4.012%, and 6.646%, respectively, on the Okhotsk Sea dataset. Particularly noteworthy is the increase in classification accuracy of NI by 9.249%. On the Sakhalin1 dataset, PA, mPA, and mIoU are enhanced by 3.713%, 3.530%, and 5.691%, respectively.

3) *Baseline+Msdam+Transformer*: Based on the baseline, the MSDAM module and transformer branch are added, allowing the model to simultaneously receive inputs from both branches during the final decision-making process, with adaptive weights attached to each branch. From Tables VI and VII, it can be observed that with the continued addition of the transformer branch, on the Okhotsk Sea dataset, PA, mPA, and mIoU are respectively increased by 0.043%, 0.319%, and 0.133%. However, there is a decrease in YI's CPA metric and IoU metric other than NI. On the Sakhalin1 dataset, PA, mPA, and mIoU are respectively increased by 0.022%, 0.095%, and 0.057%, while there is a certain degree of decline in NI's CPA and IoU metrics.

4) *SI-CTFNet*: On the basis of baseline, the MSDAM module, transformer branch, and BiAFusion module are added,

so that the model receives the input of three branches at the same time in the final decision, and adds adaptive weights to the three branches, respectively. As can be seen from Tables VI and VII, the PA, mPA, and mIoU of SI-CTFNet on the Okhotsk Sea dataset increased by 0.138%, 0.014%, and 0.208%, respectively, reaching the highest 90.413%, 90.437%, and 82.772%. The PA, mPA, and mIoU on the Sakhalin1 dataset increased by 0.194%, 0.139%, and 0.217%, respectively, reaching the highest of 88.575%, 89.102%, and 80.300%.

D. Comparison With Sentinel1A Satellite on Sakhalin2 Dataset

To compare with other SAR satellite data, we specifically selected a scan mode image from the Sentinel1A satellite, which operates in the C-band, to create the Sakhalin2 dataset. The samples in the Sakhalin2 dataset are all cropped from SAR images of GF3B satellite and Sentinel1A satellite at the same location. We used training samples from data acquired by both satellites to train the model and tested the performance of our proposed method on SAR sea ice images from both satellites.

From the experimental results shown in Table VIII, it can be observed that our proposed method achieves slightly higher performance on the GF3B satellite dataset compared to the Sentinel1A satellite dataset in terms of PA, mPA, and mIoU. Specifically, the values are 92.100%, 92.097%, and 85.564%, respectively, which are 0.601%, 0.606%, and 0.955% higher than those obtained on the Sentinel satellite dataset. In terms of individual class accuracy, the GF3B satellite dataset performs slightly better on OW and YI categories compared to the Sentinel satellite dataset, while the opposite is observed for the other two categories. Regarding IoU, the results show that the GF3B

TABLE VIII

COMPARISON EXPERIMENT RESULTS OF CLASSIFICATION PERFORMANCE BETWEEN SENTINEL1A SATELLITE IMAGES AND GF3B SATELLITE IMAGES

| Satellite | CPA | | | | IoU | | | | PA | mPA | mIoU |
|------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | OW | FYI | NI | YI | OW | FYI | NI | YI | | | |
| Sentinel1A | 93.737 | 90.166 | 87.664 | 94.397 | 90.190 | 79.348 | 77.410 | 91.487 | 91.499 | 91.491 | 84.609 |
| GF3B | 95.268 | 90.061 | 87.373 | 95.685 | 91.370 | 80.188 | 78.712 | 91.986 | 92.100 | 92.097 | 85.564 |

The best values in the columns are in bold. All scores are expressed as percentages (%).

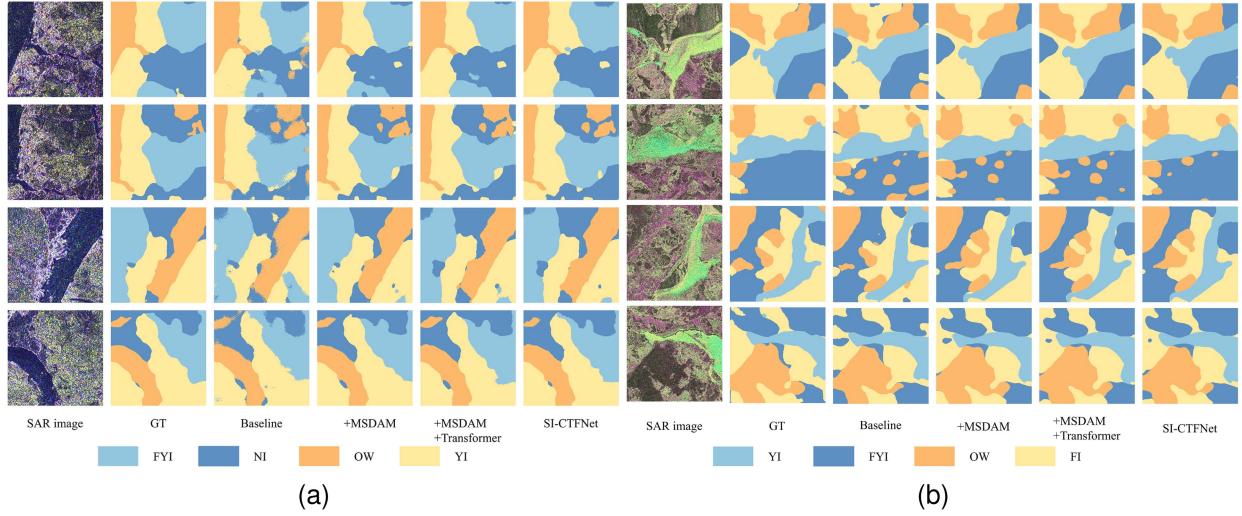


Fig. 10. Visualization of ablation experiments. (a) Okhotsk Sea Dataset. (b) Sakhalin1 Dataset.

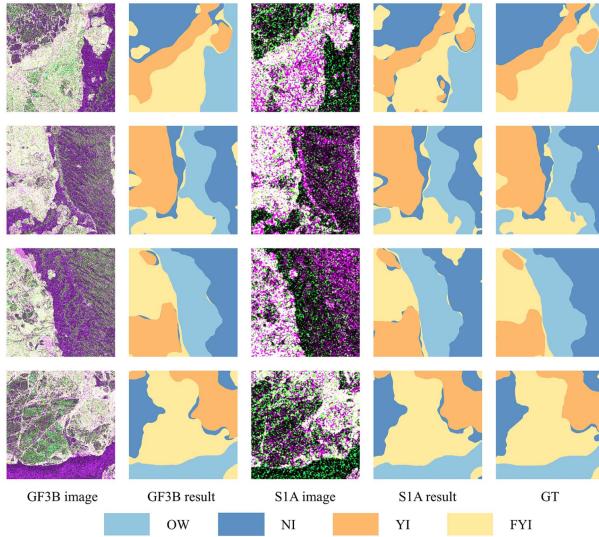


Fig. 11. Comparison of classification results between GF3B satellite image and Sentinel1A satellite image.

satellite dataset performs slightly better than the Sentinel1A satellite dataset. Overall, the performance differences between the two satellite datasets are minor across all evaluated metrics.

To demonstrate the segmentation results of the two satellite datasets, we visualize them in Fig. 11. As shown in Fig. 11, due to the small time interval between the passages of the two satellites, the similarity of sea ice at the same location is relatively high. However, the imaging parameters and working modes of

the two satellites are different, resulting in significant differences in the features of the same sea ice in the corresponding images. In addition, Sentinel1A satellite images exhibit slightly more speckle noise, reflected in the presence of more green spots in the color images. In terms of segmentation results, the segmentation results for larger annotated areas are generally similar between the two, while in smaller annotated areas, the segmentation performance of Sentinel1A satellite images is slightly inferior to that of GF3B satellite images.

E. Comparison With Lightweight Model on Inference Speed

We tested the inference speed of SI-CTFNet and Lightweight SI-CTFNet on the Okhotsk Sea dataset. To provide a more intuitive comparison of the test results, in addition to the previously used metrics PA, mPA, and mIoU, we introduced three additional evaluation metrics: FLOPs (Floating-point Operations Per Second), Params (number of model parameters), and FPS (Frames Per Second). The final test results are shown in Table IX.

As shown in Table IX, the classification performance of the lightweight model decreased by 1–2% in terms of PA, mPA, and mIoU. The FLOPs metric decreased from over 500 to 92.51, indicating a significant reduction in model complexity. Due to downsampling and branch reduction, the number of model parameters also decreased by approximately 6M. The inference speed improved from the original 24FPS to 73FPS, achieving a threefold increase in inference speed.

TABLE IX
COMPARISON EXPERIMENT RESULTS OF INFERENCE SPEED BETWEEN
SI-CTFNET AND LIGHTWEIGHT SI-CTFNET

| Model | PA (%) | mPA (%) | mIoU (%) | FLOPs (Gbps) | Params (Mb) | FPS |
|-----------------------|---------------|---------------|---------------|--------------|--------------|--------------|
| SI-CTFNet | 90.413 | 90.437 | 82.772 | 502.23 | 25.84 | 24.03 |
| Lightweight SI-CTFNet | 89.165 | 89.206 | 80.675 | 92.51 | 19.78 | 73.49 |

The best values in the columns are in bold.

VI. DISCUSSION

A. Sample Labeling Method

This study utilizes semantic segmentation models to address the sea ice classification issue. Such supervised learning models rely on precise pixel-level labeling to achieve accurate classification accuracy. However, the ice charts issued by major national ice and snow departments only provide rough regional labels rather than pixel-level ones, resulting in very limited ground truth information available. Therefore, to obtain reliable pixel-level sample annotations, this study employs contemporaneous optical satellite imagery to assist in manually annotating SAR sea ice images. Despite this, the method still encounters some difficulties. Ideally, sample labels could be directly obtained by annotating optical sea ice images. However, during the process of cropping overlapping regions, it is found that there are certain pixel errors in the corresponding points of images captured by different satellites. These errors arise partly from geometric correction and partly from the movement of sea ice between the acquisition times of the two images. Consequently, resorting to manually annotating SAR images to obtain ground truth labels inevitably increases the workload significantly.

B. Error Analysis for Classification

Due to the intricate interweaving of various types of sea ice at the boundaries in SAR images, manually annotating SAR sea ice images is exceedingly challenging. In addition, constrained by computational resources, we are only able to select fixed-size patches from the original large images as samples. Furthermore, when selecting samples, we tend to prioritize those from regions with high confidence (such as the boundaries between different types of sea ice) for both training and testing purposes.

However, due to the very similar appearance of different types of ice in SAR images, especially when they are in continuous developmental stages, such as the YI and NI within the red box in Fig. 12(a). Inside the red box, it can be observed from Landsat 8 optical images that the area below open water should be NI. However, in SAR images, this portion of NI exhibits scattering characteristics more similar to YI. As a result, in the segmentation results, this portion is classified as YI. Furthermore, there are significant differences between the FYI on the left and right sides of the SAR image. The scattering intensity of the FYI on the left is noticeably weaker than that on the right, even exhibiting more similar characteristics to the NI within the red box. This also leads to a portion of the NI within the red box being classified as FYI.

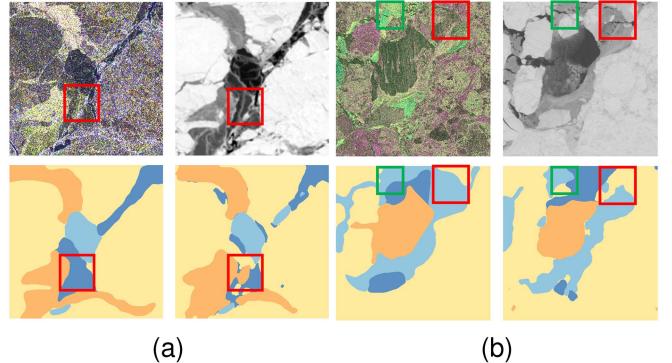


Fig. 12. Two samples for visualizing classification error analysis. (a) Selected from Okhotsk Sea Dataset. (b) Selected from Sakhalin1 Dataset.

On the other hand, ice of the same category in SAR images may exhibit different appearances. For instance, YI within the red and green boxes in Fig. 12(a) demonstrates markedly different scattering characteristics. The YI in the green box exhibits higher scattering intensity, similar to the characteristics of young ice in other images, while the YI in the red box exhibits features similar to FYI. This results in better segmentation of the YI in the green box during the segmentation process, whereas the YI in the red box is misclassified as FYI. In addition, in SAR sea ice images from the Sakhalin1 dataset, apart from open water, the boundaries between other categories of sea ice are more ambiguous. For example, FI itself consists of small patches of FYI mixed with some small fragments of other categories of sea ice, making it challenging to label in SAR images and leading to difficulties in segmentation.

C. Significance of Lightweight Model in Practical Applications

The primary motivation for lightweighting SI-CTFNet is to integrate sea ice classification processing into real-time satellite data reception and processing systems deployed on the ground or aboard ships. This allows for rapid, pixel-level, high-precision sea ice classification to be performed concurrently with satellite data reception. To illustrate, we discuss using a $20k \times 20k$ dual-polarimetric FSII image from GF3 satellites. Based on operational experience with GF3 satellites imaging programs, generating a $20k \times 20k$ FSII SAR image product from raw data typically requires 10 s, resulting in a total of 20 s for a dual-polarimetric SAR image product. Given that the input size for SI-CTFNet is 512, predicting a $20k \times 20k$ image with the original SI-CTFNet requires 66 s, whereas the lightweight SI-CTFNet accomplishes this in just 22 s. Referring to the 2019 collision of the Xue Long with an iceberg in Antarctica, the vessel's speed in the ice floe region was 3 knots (1.72 m/s). According to the crew's experience, the visibility of icebergs was approximately several tens of meters. In the aforementioned situation, using the lightweight model for prediction takes 42 s, providing a warning for icebergs 72 m away, which is sufficient to guide the vessel to alter its route before manual observation. In contrast, using the original model requires 86 s for prediction, providing a warning for icebergs only 148 m away, which does

not allow for route adjustment before manual observation. In summary, the lightweight model enables rapid real-time ice condition prediction in complex sea states without relying on manual observation (if the interpretation time is shorter, it enables the detection of closer icebergs, allowing for more timely routes adjustments).

VII. CONCLUSION

This study addresses the safety hazards posed by sea ice in Arctic Route by proposing and training a novel sea ice semantic segmentation network model, SI-CTFNet, which integrates CNNs with ViTs, rigorously validating the model's classification effectiveness and thoroughly discussing its practical applicability. In data preprocessing stage, we used simultaneous optical images from the Landsat satellites, in conjunction with manual annotation, to enhance the interpretation of SAR sea ice images captured by the GF3 satellites, thereby achieving pixel-level labeling with increased reliability. SI-CTFNet adopts a dual-branch parallel feature extraction approach, independently acquiring feature maps of diverse scales from the CNN and transformer branches, and subsequently adaptively fusing them at the same scale through the BiAFusion module. In addition, the MSDAM module is employed to enable the model to accommodate sea ice of varying sizes within the images. We demonstrate the effectiveness of our proposed method through extensive comparative testing and ablation experiments on the Okhotsk Sea dataset and the Sakhalin1 dataset. Comparative segmentation experiments on SAR images from two satellites were conducted on the Sakhalin2 dataset by introducing Sentinel1A SAR image to increase experimental data diversity, thereby validating the applicability of our proposed method across different satellite data. In view of practical application requirements, a lightweight model was proposed, achieving a threefold increase in inference speed on the Okhotsk Sea dataset without significantly compromising accuracy. Currently, this method is limited to the segmentation of small SAR sea ice images. Future research will first address the issue of insufficient inference speed by further optimizing the lightweight design, particularly reducing the complexity of the ViT component, aiming to achieve real-time reception and processing with the quick view system. In addition, the limited and manually selected data used for training results in a model that may not generalize well to full sea ice images. The study should focus on increasing sample data to enhance model versatility and effectively address issues related to stitching small sample segments.

ACKNOWLEDGMENT

The authors would like to thank Lijian Shi (National Satellite Ocean Application Service, China) for providing GF3A quad-polarimetric SAR sea ice images.

REFERENCES

- [1] Y. Cao et al., "Trans-arctic shipping routes expanding faster than the model projections," *Glob. Environ. Change*, vol. 73, 2022, Art. no. 102488. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0959378022000267>
- [2] M. Cai, W. Cao, and B. Hu, "Analysis of the economy and future development trend of the northeast arctic passage," *Traffic Inform. Saf.*, vol. 38, pp. 105–111, 2020.
- [3] S. T. Koçak and F. Yercan, "Comparative cost-effectiveness analysis of arctic and international shipping routes: A fuzzy analytic hierarchy process," *Transp. Policy*, vol. 114, pp. 147–164, 2021.
- [4] B. Lin, M. Zheng, X. Chu, W. Mao, D. Zhang, and M. Zhang, "An overview of scholarly literature on navigation hazards in arctic shipping routes," *Environ. Sci. Pollut. Res.*, vol. 31, no. 28, pp. 40419–40435, 2024. [Online]. Available: <https://api.semanticscholar.org/CorpusID:261527772>
- [5] Y. Wang, K. Liu, R. Zhang, L. Qian, and Y. Shan, "Feasibility of the northeast passage: The role of vessel speed, route planning, and ice-breaking assistance determined by sea-ice conditions for the container shipping market during 2020–2030," *Transp. Res. Part E, Logistics Transp. Rev.*, vol. 149, 2021, Art. no. 102235. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1366554521000132>
- [6] X. Yang, J. Zhi, W. Zhang, S. Xu, and X. Meng, "A novel data-driven prediction framework for ship navigation accidents in the arctic region," *J. Mar. Sci. Eng.*, vol. 11, no. 12, 2023, Art. no. 2300. [Online]. Available: <https://www.mdpi.com/2077-1312/11/12/2300>
- [7] W. J. Campbell et al., "Microwave remote sensing of sea ice in the aidjex main experiment," *Boundary-Layer Meteorol.*, vol. 13, pp. 309–337, 1978.
- [8] L. Fu and B. Holt, "Seasat views oceans and sea ice with synthetic aperture radar," 1982. [Online]. Available: <https://api.semanticscholar.org/CorpusID:126840428>
- [9] A. M. Johansson, C. Brekke, G. Spreen, and J. A. King, "X-, C-, and L-band SAR signatures of newly formed sea ice in arctic leads during winter and spring," *Remote Sens. Environ.*, vol. 204, pp. 162–180, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425717304960>
- [10] S. Singha, M. Johansson, N. Hughes, S. Hvidegaard, and H. Skourup, "Arctic sea ice characterization using spaceborne fully polarimetric L-, C-, and X-band SAR with validation by airborne measurements," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3715–3734, 2018, doi: [10.1109/TGRS.2018.2809504](https://doi.org/10.1109/TGRS.2018.2809504).
- [11] J. A. Casey, S. E. Howell, A. Tivy, and C. Haas, "Separability of sea ice types from wide swath C- and L-band synthetic aperture radar imagery acquired during the melt season," *Remote Sens. Environ.*, vol. 174, pp. 314–328, 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S003442571530242X>
- [12] J. P. S. Gill and J. J. Yackel, "Evaluation of C-band SAR polarimetric parameters for discrimination of first-year sea ice types," *Can. J. Remote Sens.*, vol. 38, pp. 306–323, 2012. [Online]. Available: <https://api.semanticscholar.org/CorpusID:129882972>
- [13] L. He, X. He, F. Hui, Y. Ye, T. Zhang, and X. Cheng, "Investigation of polarimetric decomposition for arctic summer sea ice classification using Gaofen-3 fully polarimetric SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3904–3915, 2022.
- [14] T. Zhang et al., "Deep learning based sea ice classification with Gaofen-3 fully polarimetric SAR data," *Remote. Sens.*, vol. 13, no. 8, 2021, Art. no. 1452, doi: [10.3390/rs13081452](https://doi.org/10.3390/rs13081452).
- [15] M. Shokr and M. Dabboor, "Polarimetric SAR applications of sea ice: A review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 6627–6641, 2023.
- [16] 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.
- [17] 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.
- [18] M. Jiang, D. A. Clausi, and L. Xu, "Sea-ice mapping of RADARSAT-2 imagery by integrating spatial contexture with textural features," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 7964–7977, 2022.
- [19] J.-W. Park, A. A. Korosov, M. Babiker, J.-S. Won, M. W. Hansen, and H. Kim, "Classification of sea ice types in Sentinel-1 synthetic aperture radar images," *The Cryosphere*, vol. 14, no. 8, pp. 2629–2645, 2020. [Online]. Available: <https://tc.copernicus.org/articles/14/2629/2020/>
- [20] W. Guo, P. Itkin, J. Lohse, M. Johansson, and A. P. Doulgeris, "Cross-platform classification of level and deformed sea ice considering per-class incident angle dependency of backscatter intensity," *Cryosphere*, vol. 16, no. 1, pp. 237–257, 2022. [Online]. Available: <https://tc.copernicus.org/articles/16/237/2022/>
- [21] R. Ressel, S. Singha, S. Lehner, A. Rösel, and G. Spreen, "Investigation into different polarimetric features for sea ice classification using x-band

- synthetic aperture radar,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, pp. 3131–3143, Jul. 2016. [Online]. Available: <https://api.semanticscholar.org/CorpusID:41865726>
- [22] 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.
- [23] J. Lohse, A. P. Doulgeris, and W. Dierking, “An optimal decision-tree design strategy and its application to sea ice classification from SAR imagery,” *Remote. Sens.*, vol. 11, no. 13, 2019, Art. no. 1574. [Online]. Available: <https://api.semanticscholar.org/CorpusID:198425418>
- [24] S. Ochilov and D. A. Clausi, “Operational SAR sea-ice image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4397–4408, Nov. 2012.
- [25] Q. Yu and D. A. Clausi, “SAR sea-ice image analysis based on iterative region growing using semantics,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, pp. 3919–3931, Dec. 2007. [Online]. Available: <https://api.semanticscholar.org/CorpusID:6259949>
- [26] H. Boulze, A. 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. [Online]. Available: <https://www.mdpi.com/2072-4292/12/13/2165>
- [27] B. Montpetit, B. Deschamps, J. King, and J. Duffe, “Assessing the parameterization of RADARSAT-2 dual-polarized ScanSAR scenes on the accuracy of a convolutional neural network for sea ice classification: Case study over coronation gulf, Canada,” *Can. J. Remote Sens.*, vol. 49, no. 1, 2023, Art. no. 2247091, doi: [10.1080/07038992.2023.2247091](https://doi.org/10.1080/07038992.2023.2247091)
- [28] M. Jiang, X. Chen, L. Xu, and D. A. Clausi, “Semi-supervised sea ice classification of SAR imagery based on graph convolutional network,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 1031–1034, [Online]. Available: <https://api.semanticscholar.org/CorpusID:252588697>
- [29] W. Song, M. Li, Q. He, D. Huang, C. Perra, and A. Liotta, “A residual convolution neural network for sea ice classification with sentinel-1 SAR imagery,” in *Proc. IEEE Int. Conf. Data Mining Workshops*, 2018, pp. 795–802.
- [30] W. Song et al., “Automatic sea-ice classification of SAR images based on spatial and temporal features learning,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 9887–9901, Dec. 2021.
- [31] H. Lyu, W. Huang, and M. Mahdianpari, “Eastern arctic sea ice sensing: First results from the RADARSAT constellation mission data,” *Remote Sens.*, vol. 14, no. 5, 2022, Art. no. 1165, doi: [10.3390/rs14051165](https://doi.org/10.3390/rs14051165).
- [32] S. Khaleghian, H. Ullah, T. Krämer, 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. [Online]. Available: <https://www.mdpi.com/2072-4292/13/9/1734>
- [33] W. Song, W. Gao, Q. He, A. Liotta, and W. Guo, “SI-STSAR-7: A large SAR images dataset with spatial and temporal information for classification of winter sea ice in hudson bay,” *Remote. Sens.*, vol. 14, no. 1, 2021, Art. no. 168, [Online]. Available: <https://www.mdpi.com/2072-4292/14/1/168>
- [34] J. Zhang, W. Zhang, Y. Hu, Q. Chu, and L. Liu, “An improved sea ice classification algorithm with gaofen-3 dual-polarization SAR data based on deep convolutional neural networks,” *Remote. Sens.*, vol. 14, no. 4, 2022, Art. no. 906. [Online]. Available: <https://www.mdpi.com/2072-4292/14/4/906>
- [35] 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.*, 2017, pp. 3262–3265.
- [36] 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, 2022, Art. no. 4010205.
- [37] J. Zhao, L. Chen, J. Li, and Y. Zhao, “Semantic segmentation of sea ice based on u-net network modification,” in *Proc. IEEE Int. Conf. Robot. Biomimetics*, 2022, pp. 1151–1156.
- [38] R. Huang, C. Wang, J. Li, and Y. Sui, “DF-UHRNet: A modified CNN-based deep learning method for automatic sea ice classification from sentinel-1A/B SAR images,” *Remote. Sens.*, vol. 15, 2023, Art. no. 2448. [Online]. Available: <https://api.semanticscholar.org/CorpusID:258580420>
- [39] H. Wan et al., “Multi-featured sea ice classification with SAR image based on convolutional neural network,” *Remote. Sens.*, vol. 15, no. 16, 2023, Art. no. 4014. [Online]. Available: <https://www.mdpi.com/2072-4292/15/16/4014>
- [40] Y. Han, Y. Liu, Z. Hong, Y. Zhang, S. Yang, and J. Wang, “Sea ice image classification based on heterogeneous data fusion and deep learning,” *Remote. Sens.*, vol. 13, no. 4, 2021, Art. no. 592. [Online]. Available: <https://www.mdpi.com/2072-4292/13/4/592>
- [41] K. Yang et al., “Fine resolution classification of new ice, young ice, and first-year ice based on feature selection from Gaofen-3 quad-polarization SAR,” *Remote Sens.*, vol. 15, no. 9, 2023, Art. no. 2399. [Online]. Available: <https://www.mdpi.com/2072-4292/15/9/2399>
- [42] 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.
- [43] L. Zhao, T. Xie, W. Perrie, and J. Yang, “Deep-learning-based sea ice classification with Sentinel-1 and AMSR-2 data,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 5514–5525, 2023.
- [44] A. Vaswani et al., “Attention is all you need,” *Neural Inf. Process. Syst.*, 2017. [Online]. Available: <https://arxiv.org/abs/1706.03762>
- [45] A. Dosovitskiy et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” *arxiv* 2010.11929, 2020. [Online]. Available: <https://arxiv.org/abs/2010.11929>
- [46] Y. Zhang, H. Liu, and Q. Hu, “Transfuse: Fusing transformers and CNNs for medical image segmentation,” *Med. Image Computing Comput. Assisted Intervention – MICCAI 2021*, *arxiv* 2102.08005, 2021, pp. 14–24.
- [47] Q. Zhang, “System design and key technologies of the GF-3 satellite,” *Acta Geodaetica et Cartographica Sinica*, vol. 46, no. 3, 2017, Art. no. 269, doi: [10.11947/j.AGCS.2017.20170049](https://doi.org/10.11947/j.AGCS.2017.20170049).
- [48] M. Mäkinen and J. Karvonen, “Incidence angle dependence of first-year sea ice backscattering coefficient in Sentinel-1 SAR imagery over the kara sea,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6170–6181, Nov. 2017.
- [49] J. Lohse, A. P. Doulgeris, and W. Dierking, “Incident angle dependence of Sentinel-1 texture features for sea ice classification,” *Remote. Sens.*, vol. 13, no. 4, 2021, Art. no. 552. [Online]. Available: <https://www.mdpi.com/2072-4292/13/4/552>
- [50] S. Singha, A. M. Johansson, and A. P. Doulgeris, “Robustness of SAR sea ice type classification across incidence angles and seasons at L-band,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, pp. 9941–9952, Dec. 2021. [Online]. Available: <https://api.semanticscholar.org/CorpusID:229253398>
- [51] J. Lohse, A. P. Doulgeris, and W. Dierking, “Mapping sea-ice types from Sentinel-1 considering the surface-type dependent effect of incidence angle,” *Ann. Glaciol.*, vol. 61, pp. 260–270, 2020. [Online]. Available: <https://api.semanticscholar.org/CorpusID:221709547>
- [52] 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, Art. no. 3638670. [Online]. Available: <https://api.semanticscholar.org/CorpusID>
- [53] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Álvarez, and P. Luo, “SegFormer: Simple and efficient design for semantic segmentation with transformers,” *Neural Inf. Process. Syst.*, 2021. [Online]. Available: <https://arxiv.org/abs/2105.15203>
- [54] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” *arxiv* 1505.04597, 2015. [Online]. Available: <https://arxiv.org/abs/1505.04597>
- [55] J. Chen et al., “TransuNet: Transformers make strong encoders for medical image segmentation,” *arxiv* 2102.04306, 2021. [Online]. Available: <https://arxiv.org/abs/2102.04306>
- [56] O. Oktay et al., “Attention U-Net: Learning where to look for the pancreas,” *arxiv* 1804.03999, 2018. [Online]. Available: <https://arxiv.org/abs/1804.03999>
- [57] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, “ResuNet-A: A deep learning framework for semantic segmentation of remotely sensed data,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 162, pp. 94–114, Apr. 2020, doi: [10.1016/j.isprsjprs.2020.01.013](https://doi.org/10.1016/j.isprsjprs.2020.01.013).

Jiande Zhang received the B.S. degree in communication engineering from ShanDong University, ShanDong, China, in 2020, and the M.S. degree in information and communication engineering from the University of Chinese Academy of Sciences, Beijing, China, in 2023.

He is currently an Assistant Engineer with the Qilu Aerospace Information Research Institute, Jinan. His research interests include remote sensing satellite ground systems, SAR image processing and interpretation.





Wenyi Zhang received the B.S. degree in electronic information engineering from Beihang University, Beijing, China, in 2006, and the Ph.D. degree in signal and information processing from the University of Chinese Academy of Sciences, Beijing, in 2012.

He is currently an Associate Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing. His research interests include remote sensing satellite ground systems, remote sensing information processing, artificial intelligence, and Big Data based on remote sensing data.



Guangzuo Li received the bachelor's degree in electronic information engineering, and the master's degree in information and communication engineering from Tsinghua University, Beijing, China, in 2008, 2013, respectively, and the Ph.D. degree in information and communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2020.

He is currently an Associate Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing. His research interests include synthetic aperture radar, radar imaging processing technology, sparse signal processing, and remote sensing data processing.



Xiao Zhou received the B.S. degree from China University of Mining and Technology, Beijing, China, in 2009, and the Ph.D. degree in photogrammetry and remote sensing from Peking University, Beijing, in 2014.

He is a Distinguished Researcher specially appointed by the Chinese Academy of Sciences, a member of the Youth Innovation Promotion Association, the Director of a research office at the Aerospace Information Research Institute of the Chinese Academy of Sciences, and the Head of the field responsible for space situational awareness and space manipulation.



Xiangyu Dai received the B.Eng. degree in electronic information engineering from the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, China, in 2022. He is currently working toward the M.S. degree in electronic information engineering with the Aerospace Information Research Institute, Chinese Academy of Sciences.

His research interests include remote sensing image processing and object detection.



Qingwei Chu received the master's degree in information and communication engineering from the University of Chinese Academy of Sciences, Beijing, China, in 2013.

He is currently an Associate Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, working in the field of remote sensing satellite ground data receiving and processing systems.



Shuo Hu received the master's degree in information and communication engineering from the Jilin University, Changchun, China, in 2020.

He is currently an Engineer with the Qilu Aerospace Information Research Institute, Jinan, working in the field of remote sensing satellite ground data receiving and processing systems.



Xiaoyi Yin received the B.S. degree in electronic information engineering from China Agricultural University, Beijing, China, in 2013. He received the Ph.D. degree in computer vision from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, in 2021.

He is currently an Assistant Researcher with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing. His research interests include remote sensing satellite ground systems, remote sensing information processing, artificial intelligence, Big Data based on remote sensing data, machine learning, and applied mathematics.



Fukun Jin received the B.Eng. degree in 2023 from Xidian University, Xi'an, China, majoring in communication engineering. He is currently working toward the M.S. degree in information and communication engineering with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China.

His research interests include remote sensing image processing.