

# Deep-Learning-Based Sea Ice Classification With Sentinel-1 and AMSR-2 Data

Li Zhao , Tao Xie , William Perrie , and Jingsong Yang 

**Abstract**—In the era of big data, how to utilize synthetic aperture radar (SAR) and passive microwave radiometer data for better sea ice monitoring by deep-learning technology has recently attracted wide attention. In this article, we first propose a universal and lightweight multiscale cascade network (MCNet) for Sentinel-1 SAR-based sea ice classification. In comparison with the previous local inference methods that split SAR images to small patches, our proposed global inference method MCNet is able to segment whole SAR images directly. Then, taking MCNet as a basis, we investigate four different fusion methods for Sentinel-1 SAR and the advanced microwave scanning radiometer-2 data. These are the early fusion, deep fusion, late fusion, and the hybrid method, which fuse data at the input level, feature level, decision level, as well as both feature and decision levels, respectively. Experiments demonstrate that MCNet performs better than the commonly used U-Net in terms of accuracy, memory usage, inference speed, and in capturing small-scale local details. As for data fusion, compared with MCNet, significant improvements have been achieved for all data fusion methods, except the early fusion method. Both deep fusion and late fusion methods have their own advantages in classifying certain sea ice types. By combining them together, the proposed hybrid method achieves optimal performance. Finally, with regard to the class imbalance problem, we recommend the application of self-supervised learning to mine the value of massively unlabeled SAR images.

**Index Terms**—Advanced microwave scanning radiometer-2 (AMSR-2), data fusion, deep learning (DL), sea ice classification, Sentinel-1, synthetic aperture radar (SAR).

Manuscript received 2 March 2023; revised 2 May 2023; accepted 10 June 2023. Date of publication 13 June 2023; date of current version 26 June 2023. This work was supported in part by the National Natural Science Foundation of China under Project 42176180, in part by the National Key R&D Program of China under Grant 2021YFC2803302 and Grant 2022YFC3104900, in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant KYCX20\_0930, in part by the China Scholarship Council for 2 year's study at the Bedford Institute of Oceanography under Grant 202008320523, and in part by Canada's SWOT (Surface Water and Ocean Topography) programs and MAXSS, the ESA initiative for Maximum Atmospheric eXtreme Satellite Synergy. (Corresponding author: Tao Xie.)

Li Zhao is with the School of Marine Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: nuist-zhaoli@foxmail.com).

Tao Xie is with the Laboratory for Regional Oceanography and Numerical Modeling, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China, and also with the School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: xietao@nuist.edu.cn).

William Perrie is with the Department of Fisheries and Oceans Canada, Bedford Institute of Oceanography, Dartmouth, NS B2Y 4A2, Canada (e-mail: william.perrie@dfo-mpo.gc.ca).

Jingsong Yang is with the State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China (e-mail: jsyang@sio.org.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3285857

## I. INTRODUCTION

ARCTIC sea ice plays an important role in maintaining the energy balance of the Earth's climate system and is a key indicator of global climate change. Therefore, it has become a hot topic to develop automatic algorithms to obtain accurate, high-resolution sea ice-type maps from satellite remote sensing data, especially given the present and anticipated future climate, with the accelerating decline in Arctic sea ice cover [1], [2], [3].

Nowadays, the synthetic aperture radar (SAR) and passive microwave radiometer (MWR) have become the most powerful tools to monitor sea ice due to their powerful ability to work day and night, almost without regard to weather. SAR images have the advantage of high resolution (<100 m) and can, therefore, catch the fine spatial and texture details of sea ice, especially in the marginal ice zone (MIZ). However, the backscattering signatures of open water (OW) and different sea ice conditions are quite complex in SAR images; especially at high wind conditions, where the backscattering signatures turn out to be more easily ambiguous, compared with low wind conditions. Specifically, in high winds, various sea ice types and OW can have very close backscatter signatures. This is a great challenge for making high-accuracy sea ice maps. Compared with SAR images, MWR data, such as advanced microwave scanning radiometer-2 (AMSR-2) brightness temperature (BT), exhibit very distinct differences between OW and sea ice, but at poor spatial resolution (~10 km). Therefore, the potential benefit is large for sea ice classification, by combining SAR with AMSR-2.

In this work, a new lightweight deep-learning (DL) architecture is proposed for Sentinel-1 SAR-based sea ice classification. This architecture is different from previous methods. It adopts multiscale SAR images as input instead of small patches cropped from SAR images. This architecture has significant advantages over the commonly used U-Net model in terms of prediction accuracy, memory usage, and inference speed. Based on this, we investigate four different DL-based data fusion methods, including early fusion, deep fusion, late fusion, and the hybrid method, for Sentinel-1 SAR and AMSR-2 fusion. Experiments demonstrate the advantages that the hybrid method has compared with the three other methods. We also investigate the effect of label errors on model performance and find that a small number of wrong labels does not harm model performance significantly. However, we also find that the class imbalance problem harms model performance severely. A solution is proposed to deal with this problem.

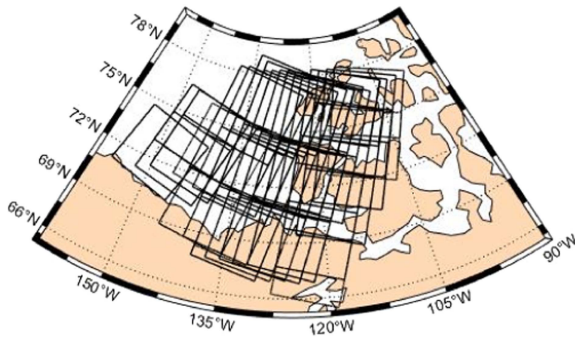


Fig. 1. Study area and footprints of Sentinel-1 SAR images.

The rest of this article is organized as follows. The study area and data are described in Section II. Related work is introduced in Section III. Our method is presented in Section IV. Results and discussion are given in Sections V and VI, respectively. Finally, Section VII concludes this article.

## II. STUDY AREA AND DATA

To capture a variety of sea ice conditions in this work, the Western Canadian Arctic (see Fig. 1) is chosen as the region of interest. This is one of the five regions where the Canadian Ice Service (CIS) carries out long-term sea ice monitoring and provides digital operational sea ice charts. At present, sea ice charts produce the most precise sea ice-type estimates and can, therefore, be used as ground truth for model training and validation, although some subjective errors may exist. This area is monitored by Sentinel-1 SAR and AMSR-2 satellites, day and night, which are also widely used for Arctic sea ice monitoring. More details about these data are introduced in the following sections.

### A. Sentinel-1 SAR Imagery

The Sentinel-1 mission consists of two polar-orbiting SAR satellites: Sentinel-1A and Sentinel-1B, operated by the European Space Agency. The revisit cycle for each Sentinel-1 satellite is 12 days. The two-satellite constellation shortens the revisit period to six days. Unfortunately, the Sentinel-1B mission ended on December 23, 2021, due to an anomaly related to the instrument electronics power supply provided by the satellite platform.

We use the medium-resolution Sentinel-1 ground range detected products in HH/HV polarization in this study. The pixel spacing is 40 m and an extra wide swath of over 400 km is used. The processing includes internal calibration, Doppler centroid estimation, as well as range and azimuth processing. The Sentinel-1 SAR images were downloaded from the Alaska Satellite Facility. We collected 574 Sentinel-1 SAR images for the period from October 2019 to September 2020. The footprints of all SAR images are shown in Fig. 1.

### B. Sea Ice Charts

Weekly regional CIS ice charts are available from the National Snow and Ice Data Center. These charts are made by ice forecast

TABLE I  
SEA ICE-TYPE DEFINITION AND CODE

Stage of development	Thickness (in cm)	WMO code
OW	NA	55
NI	< 10	81
GI	10 – 15	84
GWI	15 – 30	85
ThinFYI	30 – 70	87
MFYI	70 – 120	91
ThickFYI	≥ 120	93
OI	NA	95
SYI	NA	96
MYI	NA	97

experts based on the manual interpretation of satellite data, visual observations from ships and aircraft, and weather and oceanographic information. Satellite data are collected over several days in order to have complete coverage of any given area. The charts are provided as digital shapefiles encoded in SIGRID-3 format.

The charts provide the ice concentration estimates in increments of 10% (0, 10%, 20%, ..., 100%) and ice-type estimates. Ice information is coded using the World Meteorological Organization (WMO) standards. Based on the ice thickness, the sea ice is divided into OW, new ice (NI), gray ice (GI), gray–white ice (GWI), thin first-year ice (ThinFYI), medium first-year ice (MFYI), thick first-year ice (ThickFYI), old ice (OI), second-year ice (SYI), and multiyear ice (MYI). Their thickness and corresponding WMO codes are shown in Table I. More detailed definitions can be found at Environment and Climate Change Canada.

Although many studies have reported high classification accuracy, the number of classified sea ice types is quite different among these many studies. Some studies focus on discriminating sea ice from OW [4], [5]. Most studies can discriminate about 3–5 sea ice types [6], [7], [8], [9]. Clearly, without doubt, the difficulty in solving the sea ice-type classification task increases as the number of sea ice types to be estimated increases. This is because sea ice types are associated closely with stages of sea ice development; these can be close to one another and have very similar visual appearances in SAR images.

In this study, we focus on identifying more refined sea ice types according to the sea ice development stage: OW, NI, GI, GWI, ThinFYI, MFYI, ThickFYI, OI, SYI, and MYI. Since the ice charts also provide a land mask, we also estimate the land locations.

### C. AMSR-2 Data

The AMSR-2 is a dual-polarized (vertical and horizontal) passive MWR onboard the Japan Aerospace Exploration Agency's (JAXA's) GCOM-W1 spacecraft. AMSR2 takes the measurements of BT from the Earth's surface and the atmosphere at multiple frequency bands (6/7/10/18/23/36/89 GHz). The AMSR-2 data have been used to generate routine sea ice products, such as the ASI sea ice concentration products by Bremen University and the global sea ice-type product by OSISAF.

We use the AMSR-2 L3 BT product at 10 km resolution, which can be freely downloaded from the JAXA G-Portal. These

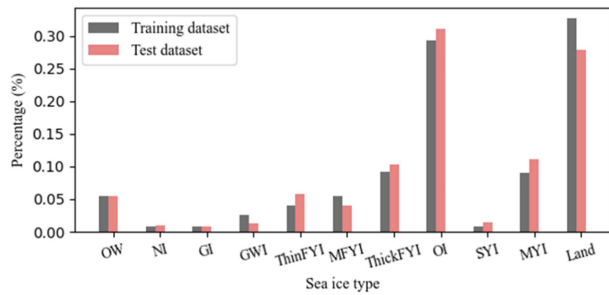


Fig. 2. Sea ice-type distributions for training and test datasets.

data include daily ascending and descending BT fields for all channels. For convenience, we only use the ascending BT data. In general, TB at low-frequency bands, such as 18 and 36 GHz, is more often used for sea ice classification [10], [11]. TB at 89 GHz is also used for sea ice classification [12], but it is more sensitive to atmospheric effects. In this study, BT at all bands of the Sentinel-1 SAR images is used to conduct the data fusion.

#### D. Dataset Construction

To construct the dataset for model training and validation, we first follow a generic workflow to process SAR images. Radiometric corrections and thermal noise removal are applied to both HH- and HV-polarized SAR images to obtain the normalized radar cross section (NRCS) in dB. The size of the original SAR images is about  $10\,000 \times 10\,000$  pixels. For convenience, all SAR images are cropped to a uniform size of  $9000 \times 9000$  pixels. To reduce the speckle noise and computational load, we downsample all SAR images to different spatial scales of 200, 400, and 800 m by boxcar averaging, corresponding to an input size of  $1800 \times 1800$ ,  $900 \times 900$ , and  $450 \times 450$  pixels. Then, the sea ice-type labels are mapped onto the SAR image grids by nearest neighbor interpolation. The AMSR-2 BT data are resampled to 800 m. We normalize the NRCS and BT values by their mean and standard deviation (SD), respectively, using the Z-score method. Then, the whole dataset is randomly split into training and test datasets at a ratio of 7:3. The sea ice distributions for training and test datasets are illustrated in Fig. 2.

### III. RELATED WORK

#### A. SAR-Based Sea Ice Classification

Nowadays, popular methods for SAR-based sea ice classification can be summarized as two types: statistical machine learning (SML) based methods and DL-based methods.

Typically, SML-based methods can be considered as having two steps. First, polarimetric and texture features are extracted from the SAR images and then an SML model is applied to make sea ice classification. The polarimetric features can reveal the scattering mechanisms or statistical properties of sea ice well. The latter properties can be extracted from SAR images based on the transformation of polarimetric channels [13], [14], [15], [16] or eigendecomposition methods [13], [15], [17]. For texture feature extraction, gray-level co-occurrence matrices [15], [18],

[19], [20], [21] are the most popular texture analysis method. Based on these extracted features, various SML models can be employed to classify sea ice, such as supporting vector machine (SVM) [18], [19], [20], [21], random forest (RF) [21], k-means [13], [15], maximum likelihood [13], decision tree [14], and Bayesian methods [22].

The combined use of different methods is often the most effective. Leigh et al. [23] proposed the map-guided ice classification system that combines the iterative region growth using semantics algorithm and a pixel-based SVM method using a nonlinear radial basis function. By integrating the SVM results into a conditional random field (CRF), Zhu et al. [24] developed the SVM-CRF algorithm for the classification of five different ice types.

Although many achievements have been obtained for SML-based methods in sea ice classification, there are some notable limitations. *First*, these methods need a high level of professional knowledge in order to implement their rather complicated feature engineering methodologies and they can, therefore, easily experience robustness problems. *Second*, their computation efficiency cannot meet the requirements of processing immense amounts of high-resolution SAR images in the era of big data. *Last* but not the least, the performance of the traditional SML methods becomes steady when the amount of data increases beyond a certain number, whereas the performance of DL methods keeps increasing with respect to the growing amounts of data [25].

To deal with these problems, DL has attracted wide attention recently due to its powerful ability to learn low- and high-level semantic features from SAR images automatically, without the need to perform complex hand-designed feature extraction. At present, the most commonly used DL architecture in SAR-based sea ice classification is CNN [26]. Typically, CNN consists of a series of convolutional layers, pooling layers, fully connected layers, and activation functions. To reduce the computation load and prevent overfitting, the convolutional layers consist of a set of filters, which share the same learnable weights. Usually, they are followed by an activation function, which is usually rectified linear units (ReLU), which help introduce strong nonlinearity into the model, thereby enabling the model to learn complex semantic features. After this, the extracted semantic features are downsampled by the pooling operation, e.g., maximum pooling, to further decrease the calculation requirements. Finally, the fully connected layers are applied to make classifications based on the extracted semantic features.

In recent years, many studies have developed SAR-based sea ice classification methodologies using CNN. Boulze et al. [9] trained LeNet with Sentinel-1 SAR data and sea ice charts for four ice-type classifications and achieved accuracy that exceeded that of the RF algorithm based on the texture features. Khaleghian et al. [8] utilized data argumentation technology to handle the class imbalance problem and significantly improved the classification performance. Zhang et al. [6] used MobileNetV3 as a backbone network and combined a multiscale feature fusion method to establish the Multiscale MobileNet model for sea ice classification based on Gaofen-3 SAR data. Lyu et al. [7] combined RADARSAT Constellation Mission data



and the normalizer-free residual network (ResNet) for sea ice detection and classification. Thus, the superior capacity of the DL model approach over the traditional SML models was again confirmed.

Recently, a more general way has been proposed to design a DL model for sea ice classification, which follows two steps. *First*, classical CNNs are used as the backbone for feature extraction. *Second*, a specifically designed neural network is applied to make predictions based on the extracted representations. For example, Song et al. [27] extracted spatial features of SAR imagery based on the ResNet model and then fed them into the so-called “long short-term memory” network to learn complementary temporal features for final prediction.

Although CNN-based methods have shown good performance, they have some inherent disadvantages in dealing with high-resolution SAR images. On the one hand, in order to make sea ice classification at the pixel level, SAR images should be first split into small patches based on a sliding window strategy. Then, the patches are fed to the model in order to predict the labels. Therefore, the time complexity is with respect to the size of the SAR image. According to Song et al. [27], it takes about 15–20 min to process one SAR image, which is quite time-consuming. On the other hand, the CNNs can only “see” the small patches instead of the whole SAR image, ignoring the relationships among adjacent patches. As a result, this significantly restricts the performance of CNN methodologies in sea ice classification due to the lack of global information.

To overcome these shortcomings, fully convolutional networks (FCNs) [28] have been applied to SAR-based sea ice classification [5] and sea ice concentration estimation [29], [30]. By replacing fully connected layers in CNNs with  $1 \times 1$  convolutional layers, FCNs are able to make predictions at the pixel level. Benefiting from this, FCNs can take inputs of arbitrary size and then make predictions with the corresponding size efficiently. Unfortunately, due to GPU memory limitations, it is difficult for the existing FCNs, such as U-Net [31] and DeepLabV3 [32], to directly deal with the entire high-resolution SAR images because they are not specially designed for high-resolution images. To reduce the computational load, the common practice is to reduce the resolution of SAR images by, e.g., 200 m [29], [30], and then divide the SAR images into small patches to further reduce the input size. As mentioned above, this also confines the ability of FCNs to learn global information and, therefore, degrades model performance.

To solve this problem, we develop a lightweight FCN model based on the idea of a multiscale cascade [33], [34]. The proposed model consists of three network branches with different depths to learn multiscale representations from multiresolution SAR images. In this way, our model achieves better performance in accuracy, memory usage, and inference speed.

### B. DL-Based Data Fusion Techniques

The overall objective of data fusion is to combine the advantages of multiple data sources in order to improve the derived data products, compared with only using a single data source. With a large amount of Earth observation satellites present in

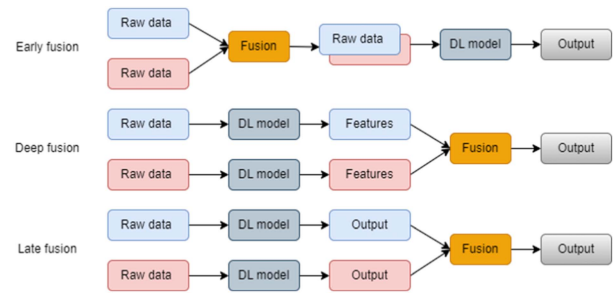


Fig. 3. Generic DL architectures for early fusion, deep fusion, and late fusion.

orbit, it is of great value to develop DL-based data fusion techniques in this era of big data. In methodologies for DL-based data fusion, one can distinguish three common types: early fusion, deep fusion, and late fusion.

Fig. 3 illustrates the generic DL architectures for *early fusion*, *deep fusion*, and *late fusion*. In the *early fusion* strategy, raw data are fused at the input or data level. By directly concatenating these raw data inputs in the original input space, they are fused channel-by-channel as the multichannel inputs that are used to learn a fused semantic feature representation. In the *deep fusion* strategy, each raw data input is used as a single input to train the individual DL model, and then these learned semantic features are fused for the final task. In the *late fusion* strategy, similar to the deep fusion strategy, each raw data input is inputted into an individual DL model. The single DL model can better dig for the unique information of the corresponding data. The outputs of each DL model will then be integrated to generate the final output. Therefore, it has been concluded that DL-based data fusion methods can be put into three categories: input/data level, feature level, and decision level. One can refer to recent review articles for an overview of these activities [35], [36], [37].

Based on the FCN model for SAR-based sea ice classification, we explore four different fusion methods: early fusion, deep fusion, late fusion, and a hybrid method (combining deep fusion and late fusion), for Sentinel-1 SAR and AMSR-2 data fusion, for the first time.

## IV. METHOD

### A. Multiscale Cascade Network (MCNet) for SAR-Based Sea Ice Classification

In general, deeper networks usually lead to better performance. However, their huge memory usage makes it difficult for them to deal with full-scale SAR images. To reduce the memory burden, previous methods have usually cropped SAR images into small patches, which harms model accuracy due to the lack of global information. In contrast, we try to segment entire SAR images directly. Taking into account the computation efficiency and prediction accuracy, MCNet adopts a multiscale cascade architecture [33], [34] to extract low- and high-level semantic features from multiscale SAR images. As shown in Fig. 4, this approach consists of two shallow network branches and one relatively deep network branch. The shallow network branches are used to extract low-level spatial features from high- and medium-resolution SAR images, while the deep network branches are

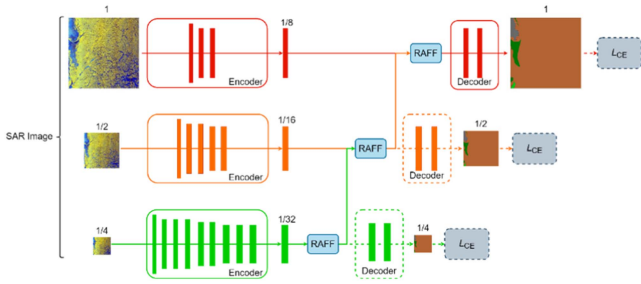


Fig. 4. Overall architecture of MCNet. Numbers are the ratios of the image or feature map size, compared with the highest resolution of the input data. Dotted lines and boxes represent modules that are only used in the training stage.

designed to extract complementary high-level semantic features from relatively low-resolution SAR images. Thus, the RAFF module [34] is employed to fuse semantic features deeply from adjacent network branches. Finally, lightweight decoders are utilized to make predictions based on the fused features. The essential structure of the encoder, RAFF module, and decoder are described in detail in the following sections.

1) *Encoder*: For high- and medium-resolution branches, we only use the first three and four stages of the original short-term dense concatenate network (STDCNet) [38], respectively. Each stage of STDCNet consists of several blocks, including one convolutional layer, one batch normalization layer, and the ReLU activation layer in each block. The convolutional kernel size of the first stage is 1, and the kernel sizes of the others are 3. The resolution of the feature map in each stage is reduced by half by using a stride of 2. As a result, the output resolutions of the encoder for high- and medium-resolution branches are 1/8 and 1/16 of the SAR image resolution, respectively. For stages 1–4, the number of output channels is 32, 64, 256, and 512, respectively.

The encoder for the deep network branch is DeepLabv3 [32], including four ResNet18 [39] blocks and one atrous spatial pyramid pooling (ASPP) module. ASPP is composed of one  $1 \times 1$  convolution and three  $3 \times 3$  convolutions, with rates of 12, 24, and 36. The output resolution of DeepLabv3 is 1/32 of the SAR image resolution.

2) *RAFF*: How to fuse these multiscale semantic features, extracted from SAR images, is one of the key problems in this study. Common methods employ simple addition or concatenation approaches to perform feature fusion. However, this ignores the complex relationships between features from adjacent branches [34]. Therefore, we introduce the RAFF module proposed by Guo et al. [34] to learn the relationship automatically.

The overall structure of the RAFF module is shown in Fig. 5. Let  $F_1^{C \times H_1 \times W_1}$  and  $F_2^{C \times H_2 \times W_2}$  denote the feature maps from two different branches, where  $C$ ,  $H_1/H_2$ , and  $W_1/W_2$  denote the channel size, height, and width of feature maps, respectively. Channelwise attention  $\text{att}$  is calculated as follows:

$$\text{att} = \text{Conv}(\text{GAP}(F)) \quad (1)$$

where  $\text{Conv}$  denotes the convolution operation with kernel size and stride assumed to be 1, and  $\text{GAP}$  represents the global

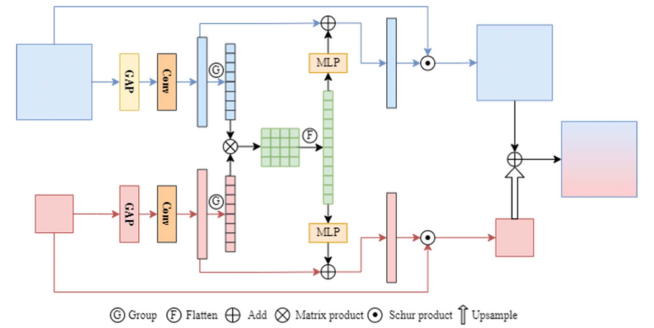


Fig. 5. Overall structure of the RAFF module. Light blue and light red indicate feature maps from two adjacent branches.

average pooling. Thus, attention vectors  $\text{att}_1^C$  and  $\text{att}_2^C$  can be computed by (1), and  $\text{att}_1^C$  and  $\text{att}_2^C$  are divided into  $K$  groups with length  $L$ , i.e.,  $G_1^{K \times L}$  and  $G_2^{K \times L}$ .

The relationship matrix  $R$  between  $G_1$  and  $G_2$  can be defined by an inner product for each of the group pairs

$$R = G_1 G_2^T \quad (2)$$

where  $T$  denotes the matrix transpose. After that,  $R$  is flattened to a 1-D vector  $R'$ , which is then imputed to the multilayer perceptron (MLP). The output of MLP is defined as follows:

$$M = \text{LN}(\text{ReLU}(\text{LN}(R'))) \quad (3)$$

where  $\text{LN}$  and  $\text{ReLU}$  denote the linear layer and activation function, respectively.

Finally, the fused feature  $F_f$  is calculated as follows:

$$F_f = \beta_1 \cdot F_1 + \text{Upsample}(\beta_2 \cdot F_2) \quad (4)$$

where  $\beta^C$  is the modulation factor, defined as follows:

$$\beta = \sigma(\text{att} + \alpha M). \quad (5)$$

3) *Decoder*: An identical lightweight decoder is employed for each branch to make final segmentations. It sequentially consists of one  $3 \times 3$  convolution, one batch normalization layer, one ReLU activation layer, and one  $1 \times 1$  convolution. During the test stage, the decoder of the last two branches can be discarded.

## B. Data Fusion Methods

On the basis of MCNet, we propose four different methods for fusing Sentinel-1 SAR and AMSR-2 data: early fusion, deep fusion, late fusion, and a hybrid method. We describe these methods in detail in the following sections.

1) *Early Fusion*: Typically, the early fusion method fuses data at the input level. However, it will increase both the computation and memory load seriously, making it difficult to carry out large-scale training. Therefore, we concatenate AMSR-2 data and the multiscale features extracted from SAR images for early fusion, following the work of Malmgren-Hansen et al. [40]. Although they declare that this is a feature-level or deep fusion method, we consider it an early fusion approach because no features are extracted from AMSR-2 data. Based on MCNet,

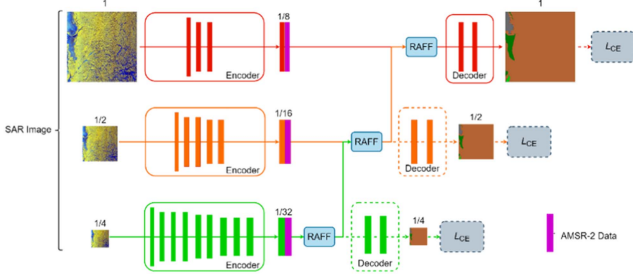


Fig. 6. Overall architecture of MCNet-E.

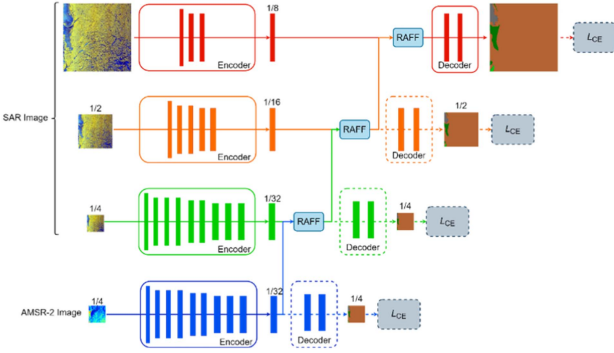


Fig. 7. Overall architecture of MCNet-D.

we just concatenate AMSR-2 data and the features extracted from SAR images for early fusion. The overall architecture of MCNet-E is shown in Fig. 6.

2) *Deep Fusion*: The deep fusion method fuses data at the feature level. Inspired by the basic idea of MCNet, an additional network branch is utilized to extract semantic features from AMSR-2 data. Then, these semantic features are fused with the multiscale features extracted from Sentinel-1 SAR images by the RAFF module. The overall architecture of MCNet-D is illustrated in Fig. 7. DeepLabV3 is used as the encoder to extract features from AMSR-2 data and this decoder is the same as that of MCNet.

3) *Late Fusion*: The late fusion method fuses data at the decision level. To be specific, the classifiers MCNet and DeepLabV3 are used to obtain sea ice maps from Sentinel-1 SAR images and AMSR-2 data, respectively. Then, we fuse their results for the final classification. Many fusion methods have been proposed for late fusion, such as averaging the confidence of the individual networks [41] or the naive Bayes (NB) method [42]. We find that the performance of the NB method is more stable and robust than that of the averaging-based approach. Therefore, we adopt the NB approach for late fusion.

The NB method assumes that the classifiers are mutually independent, given a certain class label. This is called conditional independence, which allows for the following formula:

$$p(s|w_k) = p(s_1, s_2, \dots, s_L|w_k) = \prod_{i=1}^L p(s_i|w_k), \quad (6)$$

$$k = 1, \dots, c$$

where  $s_i$  represents the output label of the  $i$ th classifier and  $w_k$  is the ground truth label.  $L$  and  $c$  are the numbers of classifiers of classes, respectively.  $p(s_i|w_k)$  denotes the probability that the  $i$ th classifier labels the sample in class  $s_i$ .

According to the Bayesian theory, the posterior probability needed to label a certain sample can be calculated as follows:

$$p(w_k|s) = \frac{p(w_k)p(s|w_k)}{p(s)} = \frac{p(w_k) \prod_{i=1}^L p(s_i|w_k)}{p(s)}. \quad (7)$$

It is obvious that the denominator is independent of  $w_k$  and, therefore, can be ignored. Then, the support for class  $w_k$  can be computed as follows:

$$\mu_k \propto p(w_k) \prod_{i=1}^L p(s_i|w_k). \quad (8)$$

The final class is determined by the maximum value of  $\mu_k$ . Based on the confusion matrix calculated for each classifier on the training dataset, (8) can be rewritten as

$$\mu_k \propto \frac{1}{N_k^{L-1}} \prod_{i=1}^L CM_{k,s_i}^i \quad (9)$$

where  $CM_{k,s_i}^i$  denotes the confusion matrix with regard to the  $i$ th classifier.  $N_k$  is the number of elements of the dataset from class  $w_k$ .

4) *Hybrid Method*: We find that the combination of deep fusion and late fusion is more effective, compared with the other possible approaches. Specifically, the classification results obtained from MCNet (SAR), DeepLabV3 (AMSR-2), and MCNet-D (SAR + AMSR-2) are fused for final predictions based on the NB method. In essence, this is also a late fusion method. To discriminate this method from the late fusion method, above, we give an explanation in this separate section.

### C. Evaluation Metrics

We evaluate model performance for three aspects: prediction accuracy, GPU memory usage, and inference time on GPU/CPU. We follow the common practice to use **mean intersection over union** (mIoU) as the accuracy metric of the semantic segmentation task. To compute mIoU, we first calculate **intersection over union** (IoU) for each class as follows:

$$\text{IoU} = \frac{\text{TP}}{\text{FP} + \text{FN} + \text{TP}} \quad (10)$$

where TP, FP, and FN represent the true positive, false positive, and false negative, respectively, which can be derived from the confusion matrix.

Then, mIoU can be calculated by averaging the classwise IoU. To evaluate the accuracy of models more objectively, we adopt a fivefold cross-validation method. The training dataset is randomly divided into five equal-sized subsets. Of the five subsets, a single subset is retained as the validation data for validating the model, and the remaining four subsets are used as the training data. The cross-validation process is then repeated five times, with each of the five subsets used exactly once as the validation data. Thus, we get five trained models. We evaluate

TABLE II  
PERFORMANCE COMPARISON BETWEEN GLOBAL INFERENCE AND LOCAL INFERENCE METHODS

Model	mIoU (Mean/SD)	Memory (Mb)	Inference time (s) (GPU/CPU)
Local inference			
U-Net-100	50.24/1.13	3316	3.06/10.73
U-Net-400	54.39/0.95	4770	1.76/6.57
U-Net-800	56.75/0.80	7754	1.92/10.46
Global inference			
MCNet	62.93/1.04		
MCNet-M	59.14/1.17	2314	1.51/5.47
MCNet-L	57.45/1.23		

their performance on the test dataset. Their mean and SD values are used to measure the model performance.

Memory and GPU inference time are measured on a GPU with a batch size of 1. We also provide the CPU inference time, in case the GPU is not available.

#### D. Implementation Details

All experiments are performed on a Ubuntu workstation with four RTX 3090 GPUs (24 GB memory) and two Intel Xeon Gold 6248R CPUs. We employ the AdamW [43] optimizer with a weight decay of 0.01 and a cosine learning rate schedule, gradually decaying from  $5 \times 10^{-4}$  to  $10^{-6}$ . The batch size is set to 4 per GPU. Therefore, the total batch size is 16. To improve generalization ability and reduce the risk of overfitting, random horizontal flip, random vertical flip, and random rotation by  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$  are applied to the training dataset, as the data augmentation. We train all experiments for 100 epochs and use warm up [44] for the first two epochs to improve training stability and reduce early overfitting.

## V. RESULTS

#### A. Global Inference Versus Local Inference

We first evaluate the performance of the proposed MCNet method for SAR-based sea ice classification. This is a global inference method since MCNet can segment whole SAR images. In comparison, the previous work on this topic usually splits SAR images into small patches as model input due to the limitations of GPU memory and computational power. We refer to this method as a local inference method.

We compare MCNet with the generic local inference methods, which train and test their models on cropped patches. We select U-Net [31] as the baseline model because it has been widely used in sea ice classification and sea ice concentration estimation [5], [30], and has achieved good performance. The evaluation results are shown in Table II. The number in the ‘‘Model name’’ column represents the patch size of the model input. As expected, the performance of U-Net improves with increasing patch size. But at the same time, the GPU memory increases greatly with patch size, which makes training and prediction difficult for large-scale SAR images. The inference time on GPU/CPU is determined by both patch size and patch number. This can explain why U-Net-400 achieves faster inference speed than

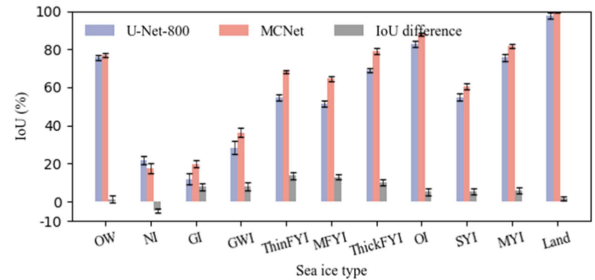


Fig. 8. Classwise IoU of U-Net-800 and MCNet, and corresponding performance differences between them. Error bars represent SD.

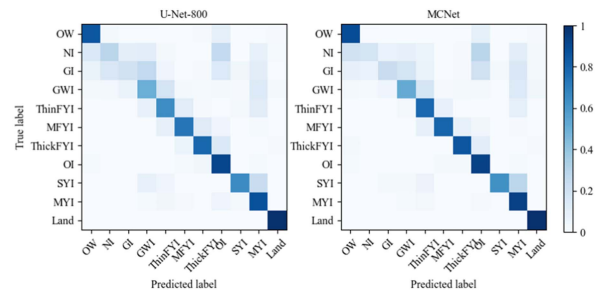


Fig. 9. Confusion matrices of U-Net-800 and MCNet.

U-Net-100 and U-Net-800. Our proposed MCNet achieves the best performance indicators. At the cost of very small memory requirements, MCNet achieves the highest mIoU and the fastest inference speed. To evaluate the effectiveness of our proposed model architecture further, we also assess the performance of the medium-resolution branch (MCNet-M) and low-resolution branch (MCNet-L). It can be seen that both medium- and high-resolution branches can benefit from the fusion of semantic features extracted from multiscale SAR images. This can demonstrate that our proposed architecture is effective.

Fig. 8 shows the classwise IoU of U-Net-800 and MCNet, and the corresponding performance differences between them. It can be seen that MCNet outperforms U-Net-800 in detecting all sea ice types, except NI. In general, it is much harder to discriminate sea ice types with close development stages because they present very similar visual features in SAR images. By increasing the patch size, MCNet is able to utilize more information to identify sea ice types with similar visual representations. Especially, for the identification of first-year ice with different thicknesses, significant improvements are observed for MCNet.

The confusion matrices of U-Net-800 and MCNet are shown in Fig. 9. Obviously, it is more difficult for both U-Net-800 and MCNet to discriminate sea ice types with close development stages, as mentioned above. Initially, it seems strange that for NI, GI, and GWI, a large proportion ( $>10\%$ ) is misclassified into OI and MYI. We, subsequently, check all data carefully and find that these misclassified sea ice types usually appear near OI and MYI. Because they suffer from the serious class imbalance problem, it is difficult for U-Net-800 and MCNet to learn sufficiently good feature representations from inadequate samples. We will discuss this problem in detail in Section VI.



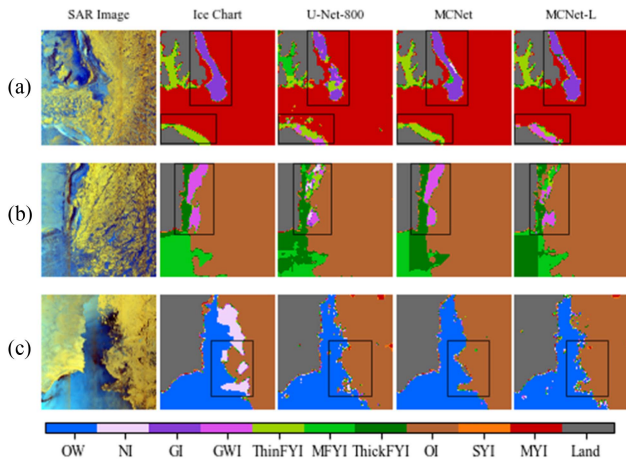


Fig. 10. Cases of segmentation results. The columns from left to right represent RGB (R: HV, G: HH, B: HH/HV) SAR imagerys, ice charts, and classification results from U-Net-800, MCNet, and MCNet-L. In this article, the colors indicate differing sea ice types following WMO standards.

Qualitative comparison results are given in Fig. 10. In case (a), the classification result from MCNet has good agreement with ice chart data, except that very few parts of GI are misclassified as MFYI. By contrast, U-Net-800 misclassifies some ThinFYI as GWI and MFYI. Moreover, some GIs are misclassified as GWI, ThinFYI, and MYI. In case (b), U-Net-800 fails to recognize GWI, while MCNet identifies GWI correctly. A large part of MFYI at the bottom left corner is misclassified as ThickFYI by both U-Net-800 and MCNet. Objectively speaking, it is quite difficult to distinguish FYI with different thicknesses because they are presented by very similar visual representations in SAR images.

In the MIZ corresponding to case (c), MCNet can better characterize the shape and structure of the MIZ than U-Net-800. Due to the class imbalance problem, both fail to distinguish NI. Compared with MCNet, MCNet-L performs worse in identifying small-scale sea ice characteristics. This is because some detailed information is discarded during downsampling.

In summary, compared with the local inference methods, our proposed global inference approach, MCNet, achieves more accurate and refined classification results, as well as higher inference speed with fewer requirements for memory usage. Due to its strong ability for learning multiscale semantic features, MCNet can better capture the small-scale local details, which are very important for sea ice applications, such as ice navigation and high-resolution and high-precision climate model developments.

### B. Evaluation of Different Data Fusion Methods

Based on MCNet, we implement four different data fusion methods. Their performances are evaluated comprehensively and given in Table III. Compared with MCNet, the accuracies of all fusion methods show obvious improvements, except for the early fusion method, which is even slightly poorer than MCNet. The hybrid method achieves the highest overall accuracy. Due to the need to deal with additional AMSR-2 data with up to

TABLE III  
PERFORMANCE COMPARISON AMONG DIFFERENT DATA FUSION METHODS

Method	mIoU (Mean/SD)	Memory (Mb)	Inference time (s) (GPU/CPU)
Early fusion	61.83/0.84	10642	2.09/5.52
Deep fusion	65.27/0.90	10726	2.08/5.72
Late fusion	64.89/0.65	2314~10788	7.84/12.28
Hybrid method	66.71/0.74	2314~10788	7.85/15.93

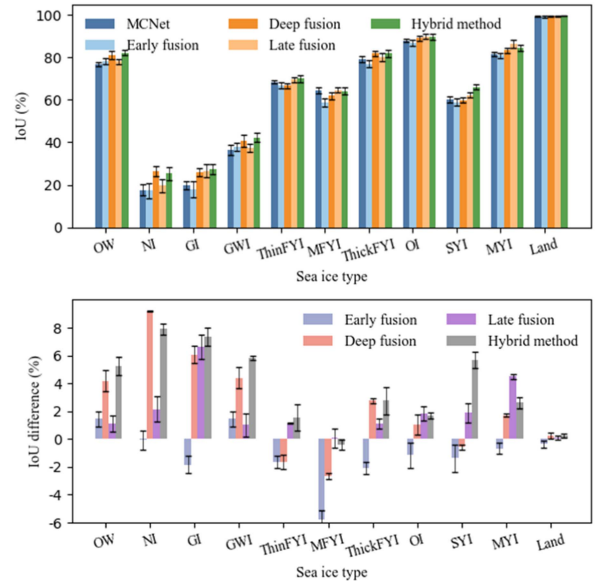


Fig. 11. Top: classwise IoU; bottom: classwise IoU difference between four data fusion methods and MCNet. Error bars represent SD.

14 channels, there is a significant increase in the GPU memory usage for these four data fusion methods. But the difference in GPU memory usage among them is quite small. Moreover, the inference time of the early fusion and deep fusion methods is comparable with that of MCNet, while the inference time of the late fusion and hybrid methods increases significantly by four times for GPU and one–two times for CPU, respectively. This is because both the late fusion and hybrid methods need to do a fusion of classification results from multiple models, resulting in the increase of computational requirements.

To further analyze their classification performance in each class in detail, the classwise IoU (Top) results for all fusion methods and MCNet are shown in Fig. 11 (Top). Combined with Fig. 2, we can see that there is a high correlation (Pearson correlation  $> 0.7$ ) between sea ice-type distribution on the training dataset and classwise IoU distribution on the test dataset. The high correlation indicates that the data imbalance has a significant negative effect on model performance over the minority classes.

Fig. 11 (Bottom) illustrates the classwise IoU difference between different fusion methods and MCNet. The early fusion method slightly improves the classification accuracy of OW, NI, GWI, MYI, and land, whereas the classification accuracy of other sea ice types is reduced. Notably, the classification accuracy of MFYI decreases by up to 4.97%. This demonstrates that the early fusion method is not an ideal approach for fusing



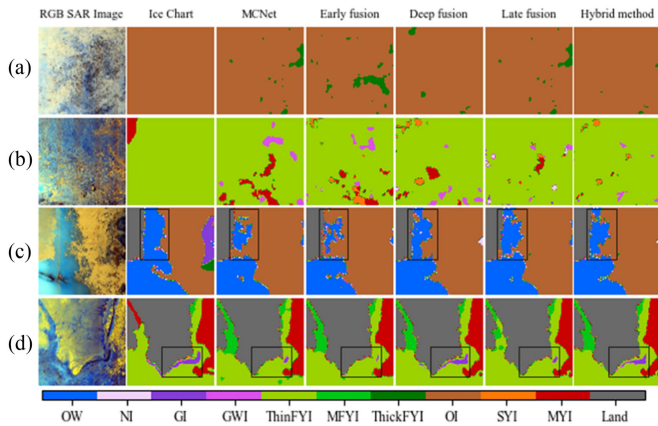


Fig. 12. Cases of segmentation results. The columns from left to right represent RGB (R: HV, G: HH, B: HH/HV) SAR imageries, ice charts, and classification results from MCNet, early fusion, deep fusion, late fusion, and the hybrid method.

Sentinel-1 SAR and AMSR-2 data for sea ice classification. By contrast, the deep fusion method gains a significant improvement ( $>4\%$ ) in classifying OW, NI, GI, and GWI. Meanwhile, the classification accuracy of ThickFYI, OI, MYI, and land also increases slightly. But the classification accuracy of ThinFYI, MFYI, and SYI drops slightly. Overall, the deep fusion method is obviously superior to the early fusion approach. This indicates that it is more effective to extract semantic features from AMSR-2 data for feature fusion, instead of simply concatenating AMSR-2 raw data and features extracted from SAR.

In contrast with the early fusion and deep fusion methods, the late fusion approach is more stable because no decrease in classification accuracy is observed for all sea ice types. By comparing the IoU difference between the deep fusion and late fusion methods, we can conclude that the deep fusion method performs better in the classification of OW, NI, GWI, ThickFYI, and land, while the late fusion approach performs better in the classification of other sea ice types. Each has its own advantages. Therefore, it is possible to make full use of their particular superior characteristics in order to get better data fusion. Following this idea, we propose a hybrid method for Sentinel-1 SAR and AMSR-2 data fusion by combining the deep fusion method and the late fusion method together. As shown in Fig. 11 (bottom), the classwise classification accuracy of the hybrid method achieves a good balance between that of the deep fusion method and the late fusion method.

To better illustrate the difference among different fusion methods, four cases are given in Fig. 12. The early fusion method can hardly help improve the classification results and can even make the results [case (a)] worse. In cases (a) and (d), the deep fusion method performs better than the late fusion method, while in cases (b) and (c), the results are reversed. By absorbing the advantages of the deep fusion and late fusion methods, the hybrid method achieves better results, overall. Both cases (a) and (b) demonstrate that we can obtain more confident results at large spatial scales by data fusion. Moreover, it is proven by case (c) that data fusion can help improve the results in the MIZ. We also observe that data fusion is able to help identify

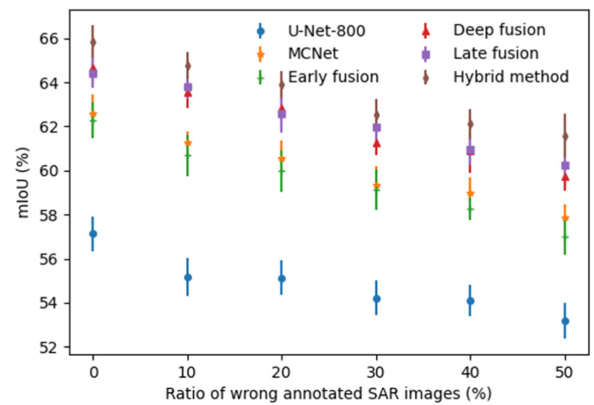


Fig. 13. Effect of label errors on model performance. Error bars represent SD.

small-scale local details in case (d). This is surprising because the spatial resolution of AMSR-2 data is coarse, on the order of 10 km. A reasonable explanation is that additional AMSR-2 data can enhance the generalization ability and robustness of the proposed DL models.

## VI. DISCUSSION

### A. Effect of Label Errors on Model Performance

Currently, ice charts are produced by experienced ice experts and are based on the manual interpretation of SAR images. In SAR images, a large spatial region with homogeneous ice characteristics is assigned with a single label. The particular ice type at a specific location may be different from the label provided by ice charts. Moreover, ice charts have been known to have biases due to the subjectivity of the ice experts [45]. Therefore, it is necessary to study the effect of label errors on model performance. However, it is quite difficult to quantitatively evaluate the label errors of ice charts. To deal with this problem, we manually introduce some errors to the training labels. Specifically, although one SAR image may contain several ice types, we randomly select only one ice type and replace its label with ice with similar development stages. This is reasonable because it is hard to distinguish sea ice types with close development stages. In this way, we evaluate the effect of label errors on model performance indirectly. It can be seen from Fig. 13 that the performance of all models decreases with the ratio of wrongly annotated SAR images. When 20% of SAR images have label errors, the mIoU of all models decreases by less than 2. Therefore, although ice charts may have some label errors, they are not generally expected to have a significant negative effect on model performance.

### B. Possible Solution to the Class Imbalance Problem

In the computer vision domain, DL models are usually developed using artificially balanced datasets. However, the datasets in the real world are usually class imbalanced, such as the sea ice datasets used in this work (see Fig. 2). As mentioned above, this problem has a significant negative effect on model performance,

over the minority classes (see Fig. 11). Therefore, it is necessary to explore possible solutions to this problem.

A lot of effort has been expended in order to deal with the imbalance problem in the field of supervised learning (SL), such as data resampling [46], [47], loss reweighting [48], [49], and representation and classifier decoupling [50], [51]. We find that the above-mentioned methods improve model performance on the so-called “instance-rare classes,” at the cost of the “instance-rich classes.” To some extent, the majority classes may be more important since they occupy the largest proportions of data in the natural world.

From a data-driven perspective, this problem constitutes the inherent issue in the DL domain. Data have never been a problem for DL, but the lack of labeled data is a problem. However, making high-quality ice charts is time-consuming and requires professional ice forecasting expertise. In fact, the labeled SAR images only occupy a small proportion, compared with the whole SAR image world. In order to solve this problem and to build more robust and more generalized DL models for SAR-based sea ice classification, we suggest exploiting the value of huge amounts of unlabeled SAR images by self-supervised learning (SSL), without significant limitations related to specific human-designed strong data augmentation in semi-SL [52].

Moreover, some studies have reported that SSL is more robust in handling the issue of data imbalance [53], [54] compared with SL. In particular, SSL has established a breakthrough with the birth of masked autoencoders (MAE) [55]. However, the encoder for MAE is a vision transformer [56], which is not suitable for processing high-resolution SAR images due to its huge memory usage. Besides, MAE will experience a serious overfitting problem if it is trained on small-scale datasets; the performance of MAE is 10% lower than that of general CNNs. In practice, previous studies have demonstrated that MAE performs better than CNNs for large enough datasets [55], [56]. More recently, Li et al. [57] proposed an “architecture-agnostic masked image modeling” (A<sup>2</sup>MIM) framework, which is compatible with both CNNs and transformers, in a universal way. Based on A<sup>2</sup>MIM, our proposed DL models can also be used for SSL. We will investigate this in the future.

## VII. CONCLUSION

In this article, we present a DL-based automatic sea ice classification algorithm with Sentinel-1 SAR images. The proposed MCNet combines three network branches with different depths to learn low- and high-level features of sea ice from multiresolution SAR images. Benefitting from an elegant design, it achieves better performance in overall accuracy, memory usage, and inference speed, compared with the common U-Net method. It is able to classify 11 sea ice types with different development stages and has the advantage of identifying small-scale sea ice conditions, with the help of multiscale feature fusion.

Based on MCNet, we adopt four different fusion methods: early fusion, deep fusion, late fusion, and the hybrid method, to fuse SAR and AMSR-2 data. Their performances are fully evaluated. The early fusion method, which concatenates AMSR-2 data and SAR image features for input-level data fusion,

provides little improvement to the classification accuracy. The deep fusion and late fusion methods fuse data at the feature level and at the decision level, respectively. Their performance is significantly improved compared with MCNet. Although each has its own strengths, their overall accuracy is comparable. The late fusion method is able to improve the classification accuracy of almost all sea ice types, whereas the deep fusion method achieves remarkable advantages in classifying certain sea ice types. By combining their advantages, the hybrid method achieves a good balance between these advantages and disadvantages. Moreover, among all fusion methods, it obtains the highest accuracy.

However, there are still some limitations in this work. In particular, the class imbalance problem has not been solved. Real-world classification problems usually present an imbalanced distribution where most classes only have a few samples. Owing to the lack of samples, performance in such classes is challenging. In this case, the overall accuracy of NI, GI, and GWI is quite poor. By investigating various studies in the literature, we think SSL is a feasible solution through mining the value of numerous unlabeled SAR images. We will further study this problem in the future.

The architecture of all DL models proposed in this study has the advantage of universality. Namely, our methods can be further improved by adopting more advanced encoder, decoder, and feature fusion modules, as the development of DL technology continues to advance.

## REFERENCES

- [1] Q. Cai, J. Wang, D. Beletsky, J. Overland, M. Ikeda, and L. Wan, “Accelerated decline of summer Arctic sea ice during 1850–2017 and the amplified Arctic warming during the recent decades,” *Environ. Res. Lett.*, vol. 16, no. 3, 2021, Art. no. 034015.
- [2] Z. Liu et al., “Acceleration of western Arctic sea ice loss linked to the Pacific North American pattern,” *Nature Commun.*, vol. 12, no. 1, 2021, Art. no. 1519.
- [3] Z. Wang et al., “Spatial and temporal variations of Arctic sea ice from 2002 to 2017,” *Earth Space Sci.*, vol. 7, no. 9, 2020, Art. no. e2020EA001278.
- [4] Y.-R. Wang and X.-M. Li, “Arctic sea ice cover data from spaceborne synthetic aperture radar by deep learning,” *Earth Syst. Sci. Data*, vol. 13, pp. 2723–2742, 2021.
- [5] 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, doi: [10.1109/LGRS.2021.3058049](https://doi.org/10.1109/LGRS.2021.3058049).
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] R. Xu, C. Zhao, X. Zhai, and G. Chen, “Arctic sea ice type classification by combining CFOSCAT and AMSR-2 data,” *Earth Space Sci.*, vol. 9, no. 2, 2022, Art. no. e2021EA002052.
- [11] Z. Zhang, Y. Yu, X. Li, F. Hui, X. Cheng, and Z. Chen, “Arctic sea ice classification using microwave scatterometer and radiometer data during 2002–2017,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5319–5328, Aug. 2019.

- [12] T. Markus and D. J. Cavalieri, "The AMSR-E NT2 sea ice concentration algorithm: Its basis and implementation," *J. Remote Sens. Soc. Jpn.*, vol. 29, no. 1, pp. 216–225, 2009.
- [13] 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, no. 3, pp. 306–323, 2012.
- [14] T. Geldsetzer and J. Yackel, "Sea ice type and open water discrimination using dual co-polarized C-band SAR," *Can. J. Remote Sens.*, vol. 35, no. 1, pp. 73–84, 2009.
- [15] J. P. S. Gill, J. J. Yackel, and T. Geldsetzer, "Analysis of consistency in first-year sea ice classification potential of C-band SAR polarimetric parameters," *Can. J. Remote Sens.*, vol. 39, no. 2, pp. 101–117, 2013.
- [16] M. Dabboor, B. Montpetit, and S. Howell, "Assessment of the high resolution SAR mode of the RADARSAT constellation mission for first year ice and multiyear ice characterization," *Remote Sens.*, vol. 10, no. 4, 2018, Art. no. 594.
- [17] M. Dabboor, B. Montpetit, S. Howell, and C. Haas, "Improving sea ice characterization in dry ice winter conditions using polarimetric parameters from C- and L-band SAR data," *Remote Sens.*, vol. 9, no. 12, 2017, Art. no. 1270.
- [18] 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.
- [19] N. Zakhvatkina, A. Korosov, S. Muckenhuber, S. Sandven, and M. Babiker, "Operational algorithm for ice–water classification on dual-polarized RADARSAT-2 images," *Cryosphere*, vol. 11, no. 1, pp. 33–46, 2017.
- [20] L. Zhang, H. Liu, X. Gu, H. Guo, J. Chen, and G. Liu, "Sea ice classification using TerraSAR-X ScanSAR data with removal of scalloping and interscan banding," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 2, pp. 589–598, Feb. 2019.
- [21] 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, pp. 2629–2645, 2020.
- [22] 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.
- [23] 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.
- [24] T. Zhu, F. Li, G. Heygster, and S. Zhang, "Antarctic sea-ice classification based on conditional random fields from RADARSAT-2 dual-polarization satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2451–2467, Jun. 2016.
- [25] M. Z. Alom et al., "A state-of-the-art survey on deep learning theory and architectures," *Electronics*, vol. 8, no. 3, 2019, Art. no. 292.
- [26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017, doi: [0.1145/3065386](https://doi.org/10.1145/3065386).
- [27] 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.
- [28] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.
- [29] I. de Gelis, A. Colin, and N. Longepe, "Prediction of categorized sea ice concentration from Sentinel-1 SAR images based on a fully convolutional network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 5831–5841, Apr. 2021, doi: [10.1109/JSTARS.2021.3074068](https://doi.org/10.1109/JSTARS.2021.3074068).
- [30] K. Radhakrishnan, K. A. Scott, and D. A. Clausi, "Sea ice concentration estimation: Using passive microwave and SAR data with a U-Net and curriculum learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 5339–5351, Apr. 2021, doi: [10.1109/JSTARS.2021.3076109](https://doi.org/10.1109/JSTARS.2021.3076109).
- [31] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2015, pp. 234–241.
- [32] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.
- [33] H. Zhao, X. Qi, X. Shen, J. Shi, and J. Jia, "ICNet for real-time semantic segmentation on high-resolution images," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 418–434.
- [34] S. Guo et al., "ISDNet: Integrating shallow and deep networks for efficient ultra-high resolution segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 4351–4360.
- [35] T. Zhou, S. Ruan, and S. Canu, "A review: Deep learning for medical image segmentation using multi-modality fusion," *Array*, vol. 3–4, 2019, Art. no. 100004.
- [36] K. Huang, B. Shi, X. Li, X. Li, S. Huang, and Y. Li, "Multi-modal sensor fusion for auto driving perception: A survey," 2022, *arXiv:2202.02703*.
- [37] J. Liu, T. Li, P. Xie, S. Du, F. Teng, and X. Yang, "Urban big data fusion based on deep learning: An overview," *Inf. Fusion*, vol. 53, pp. 123–133, 2020.
- [38] M. Fan et al., "Rethinking BiSeNet for real-time semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 9711–9720.
- [39] 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.
- [40] 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.
- [41] K. Kamnitsas et al., "Ensembles of multiple models and architectures for robust brain tumour segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*, 2018, pp. 450–462.
- [42] B. Bigdeli, F. Samadzadegan, and P. Reinartz, "A decision fusion method based on multiple support vector machine system for fusion of hyperspectral and LIDAR data," *Int. J. Image Data Fusion*, vol. 5, no. 3, pp. 196–209, 2014.
- [43] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. Int. Conf. Learn. Representations*, 2019.
- [44] P. Goyal et al., "Accurate, large minibatch sgd: Training imagenet in 1 hour," 2017, *arXiv:1706.02677*.
- [45] M.-A. N. 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.
- [46] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
- [47] S. Pouyanfar et al., "Dynamic sampling in convolutional neural networks for imbalanced data classification," in *Proc. IEEE Conf. Multimedia Inf. Process. Retrieval*, 2018, pp. 112–117.
- [48] C. Huang, Y. Li, C. C. Loy, and X. Tang, "Deep imbalanced learning for face recognition and attribute prediction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 11, pp. 2781–2794, Nov. 2020.
- [49] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 9260–9269.
- [50] B. Zhou, Q. Cui, X.-S. Wei, and Z.-M. Chen, "BBN: Bilateral-branch network with cumulative learning for long-tailed visual recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 9716–9725.
- [51] B. Kang et al., "Decoupling representation and classifier for long-tailed recognition," in *Proc. Int. Conf. Learn. Representations*, 2020.
- [52] K. Sohn et al., "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," in *Proc. 34th Int. Conf. Neural Inf. Process. Syst.*, 2020, pp. 596–608.
- [53] H. Liu, J. Z. HaoChen, A. Gaidon, and T. Ma, "Self-supervised learning is more robust to dataset imbalance," in *Proc. NeurIPS 2021 Workshop Distrib. Shifts: Connect. Methods Appl.*, 2021.
- [54] Y. Yang and Z. Xu, "Rethinking the value of labels for improving class-imbalanced learning," in *Proc. 34th Int. Conf. Neural Inf. Process. Syst.*, 2020, pp. 19290–19301.
- [55] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022.
- [56] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Representations*, 2021.
- [57] S. Li et al., "Architecture-agnostic masked image modeling—from ViT back to CNN," 2022, *arXiv:2205.13943*.



**Li Zhao** received the M.S. degree in marine meteorology in 2019 from the Nanjing University of Information Science and Technology, Nanjing, China, where he is currently working toward the Ph.D. degree.

He is currently a visiting student with the Bedford Institute of Oceanography, Dartmouth, NS, Canada. His research interests include sea ice monitoring based on synthetic aperture radar and application of deep learning in ocean remote sensing field.





**Tao Xie** received the M.S. degree in materials from the Guizhou University of Technology, Guiyang, China, in 2002, and the Ph.D. degree in electromagnetic fields and microwave technology from Shanghai Jiaotong University, Shanghai, China, in 2005.

He was a Postdoctoral Fellow in oceanography and ocean remote sensing with the Bedford Institute of Oceanography, Dartmouth, NS, Canada. He is currently a Professor with the Nanjing University of Information Science and Technology, Nanjing, China. His research interests are modeling of electromagnetic backscattering from ocean surface, ocean surface parameters retrieval from synthetic aperture radar (SAR) imageries, oil spilling detection by SAR, and Arctic sea ice remote sensing.



**William Perrie** received the B.S. degree in physics and mathematics from the University of Toronto, Toronto, ON, Canada, in 1973, and the Ph.D. degree in meteorology and oceanography from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1979.

He was a Postdoctoral Fellow with the University of British Columbia, Canada, and the National Center for Atmospheric Research, Boulder, CO, USA. He is currently a Senior Research Scientist with the Bedford Institute of Oceanography, Dartmouth, NS, Canada, and an Adjunct Professor with Dalhousie University, Halifax, NS, Canada. His research interests include ocean waves, air–sea fluxes, coupled atmosphere–ocean interactions, and impacts of climate change and satellite remote sensing.



**Jingsong Yang** received the B.Sc. degree in physics from Zhejiang University, Hangzhou, China, in 1990, and the Ph.D. degree in physical oceanography from the Ocean University of Qingdao, Qingdao, China, in 2001.

He has been with the Second Institute of Oceanography, State Oceanic Administration, Hangzhou, China, since 1996. He is currently the Head of the microwave marine remote sensing with the State Key Laboratory of Satellite Ocean Environment Dynamics, Hangzhou, China. He has been a Doctoral Supervisor with Zhejiang University, since 2011, and a Supervisor of graduate students of the Second Institute of Oceanography since 2002. He has more than 19 years of experience in microwave marine remote sensing. He has been a PI and a participant in more than 20 research projects and published more than 100 scientific papers in peer-reviewed journals and international conference proceedings. His research interests include microwave marine remote sensing, data fusion, image processing, and satellite oceanography.