Hierarchical Information-Sharing Convolutional Neural Network for the Prediction of Arctic Sea Ice Concentration and Velocity

Younghyun Koo, *Member, IEEE*, and Maryam Rahnemoonfar, *Member, IEEE*

Abstract—Forecasting sea ice concentration (SIC) and sea ice velocity (SIV) in the Arctic Ocean is of great significance as the Arctic environment has been changed by the recent warming climate. Given that physical sea ice models require high computational costs with complex parameterization, deep learning techniques can effectively replace the physical model and improve the performance of sea ice prediction. This study proposes a novel multitask fully conventional network architecture named hierarchical information-sharing U-net (HIS-Unet) to predict daily SIC and SIV. Instead of learning SIC and SIV separately at each branch, we allow the SIC and SIV layers to share their information and assist each other's prediction through the weighting attention modules (WAMs). Consequently, our HIS-Unet outperforms other statistical approaches, sea ice physical models, and neural networks without such information-sharing units. The improvement of HIS-Unet is more significant to when and where SIC changes seasonally, which implies that the information sharing between SIC and SIV through WAMs helps learn the dynamic changes of SIC and SIV. The weight values of the WAMs imply that SIC information plays a more critical role in SIV prediction, compared to that of SIV information in SIC prediction, and information sharing is more active in marginal ice zones [e.g., East Greenland (EG) and Hudson/Baffin Bays (HBB)] than in the central Arctic (CA).

Index Terms—Cryosphere, information sharing, machine learning (ML), sea ice forecast, sea ice motion, U-net, weighting attention module (WAM).

I. INTRODUCTION

THE Arctic sea ice extent (SIE) and thickness have changed dramatically over the last few decades. SIE has been reduced by more than 50 000 km^2/year (4%/decade) since the era of satellite observation in the 1970s [1], [2], which is mainly attributed to anthropogenic CO_2 emission and resultant global warming [3]. Such a loss of the Arctic sea ice cover has been observed for all seasons, both winter and summer, and almost all regions in the Arctic Ocean [4]. Meanwhile, the Arctic sea ice thickness (SIT) has decreased by more than 2 m (>60%), and more than half of multiyear ice (MYI) has disappeared for the last few decades [5], [6].

Such dramatic changes in SIE and SIT could have affected the thermodynamic and dynamic conditions of the Arctic sea ice [7], [8], [9]. In particular, since the dynamic movement of sea ice has significant impacts on sea ice area and sea ice volume, it is important to understand both sea ice dynamics and area fraction together as a clue in the future Arctic and global climate [10].

To predict the Arctic sea ice motions, a number of numerical sea ice models have been developed based on a physical understanding of sea ice and its interaction with the atmosphere and ocean. However, such physics models require too many complicated parameterizations [11], [12] accompanying high computational costs to run the models. Moreover, considering that numerical models are highly sensitive to initial conditions and physical assumptions, they can produce prediction results inconsistent with real observations [13].

Recently, in addition to these physical numerical models, machine learning (ML) techniques, particularly deep learning, have emerged as another efficient way to forecast sea ice conditions. The most common sea ice variable predicted by the deep learning model is sea ice concentration (SIC). Myriad deep learning models, including convolutional neural network (CNN) and recurrent neural network (RNN), have been proposed to predict SIC [14], [15], [16], [17], [18] with lead times varying from daily [19], [20], [21] to weekly [22] and monthly scale [23], [24]. While most previous deep learning studies have focused on predicting SIC, deep learning has rarely been used for predicting sea ice velocity (SIV). Nevertheless, several studies demonstrated that CNN outperformed other statistical methods in daily SIV prediction [25], [26].

However, those previous ML studies have only focused on the prediction of a single variable, such as either SIC or SIV alone. Considering that SIC and SIV affect each other, these two variables can potentially complement each other to improve their own prediction. Thus, this study develops a fully convolutional deep learning model performing short-term forecasting of SIC and SIV simultaneously. We design a fully convolutional network (FCN) architecture consisting of two branches specialized for SIC and SIV prediction, which share their intermediate prediction layers through information-sharing blocks. We aim to improve both SIC and SIV prediction performances by allowing the SIC and SIV intermediate layers to share their information during the model training.

The main contributions of this research work consist of the following.

1) We design a novel FCN architecture called hierarchical information-sharing U-net (HIS-Unet) to predict SIC and SIV simultaneously.
2) To allow sharing and emphasizing important information between SIC and SIV, we insert weighting attention modules (WAMs) between SIC and SIV branches, consisting of linear weights to SIC and SIV features and channel and spatial attention modules.
3) We conduct extensive experiments to check if sharing and highlighting information between SIC and SIV branches via WAMs improves SIC and SIV predictability and when and where these improvements are significant.
4) By interpreting the spatial patterns of information sharing at WAMs, we exhibit how the SIC and SIV information is shared for different regions.
5) By comparing the climatological conditions of the test year with training years, we check whether sharing SIC and SIV information can assist each other's prediction even in climatologically anomalous years.

The remainder of the article is organized as follows. Section II reviews several literature regarding the physical models for sea ice prediction, ML for sea ice prediction, and multitask ML. Section III explains details of remote sensing and meteorological data used in this study, and Section IV presents the detailed architecture of our information-sharing network and the baseline models. The accuracy and implication of the model are given in Sections V and VI, respectively.

II. RELATED WORK

In Section II-A, we first discuss the prediction of sea ice dynamics with physical models. Next, in Section II-B, we discuss the recent development of neural network techniques for sea ice prediction. Finally, in Section II-C, we discuss the recent development of multitask neural networks in remote sensing.

A. Dynamic Sea Ice Models

In general, changes in SIC are driven by two components: dynamic and thermodynamic processes. The evolution of SIC (A) is governed by the following equation [27]:

$$\frac{\partial A}{\partial t} + \nabla \cdot (\boldsymbol{u} A) = f_c - r \qquad (1)$$

where $\boldsymbol{u}$ is ice motion, f_c is SIC change from freezing or melting (thermodynamic process), and r is SIC change from mass-conserving mechanical ice redistribution processes (e.g., ridging or rafting) that convert ice area to ice thickness. Based on this relationship, the SIC changes in the Arctic and Southern Oceans have been explored as the combination of sea ice motion, thermodynamic growth, and mechanical ice deformation [27], [28].

In addition, numerous physical sea ice models have been proposed to explain and predict the dynamic behavior of the Arctic sea ice. Assuming the elastic–viscous–plastic (EVP)

condition of sea ice, these models are governed by the following momentum equation [29]:

$$m\frac{D\boldsymbol{u}}{Dt} = -mf\boldsymbol{k} \times \boldsymbol{u} + \tau_{ai} + \tau_{wi} + \boldsymbol{F} - mg\nabla H \qquad (2)$$

where $D/Dt = \partial/\partial t + \boldsymbol{u} \cdot \nabla$ is the substantial time derivative, m is the ice mass per unite area, $\boldsymbol{k}$ is a unit vector normal to the surface, $\boldsymbol{u}$ is the ice velocity, f is the Coriolis parameter, τ_{ai} and τ_{wi} are the forces due to air and water stresses, H is the elevation of the sea surface, g is the gravity acceleration, and $\boldsymbol{F}$ is the force due to variations in internal ice stress. As many previous studies have already suggested, wind and ocean forcings have primary impacts on SIV. Particularly, wind velocity has been treated as a major variable in SIV, which can contribute to up to 70% of the SIV variances [30] depending on season or region [31]. Nevertheless, predicting sea ice dynamics based on physical models is still challenging due to its intrinsic complexity and dependency on numerous atmospheric and oceanic parameterizations.

B. Deep Learning for Sea Ice Prediction

Nowadays, various ML and deep learning techniques have been extensively used in many applications of remote sensing of polar oceans [32], [33], [34], [35], [36]. In particular, deep learning techniques have been effective in predicting the high complexity of sea ice conditions (i.e., SIC and SIV) [37] (Table I). First, in terms of SIC, Andersson et al. [14] proposed a deep-learning sea ice forecasting system named IceNet, which is designed to forecast monthly SIC for the next six months. Kim et al. [15] also used CNN to predict after-one-month SIC from satellite-based SIC observations and weather data for the previous months, and their model showed a mean absolute error (MAE) of 2.28% and anomaly correlation coefficient of 0.98. Similarly, a CNN model proposed by Ren and Li [16] showed <1% of MAE in daily SIC prediction for the melting season, and U-Net by Ren et al. [17] showed 2%–3% of MAE in daily SIC prediction. In addition to CNN, long short-term memory (LSTM), an advanced RNN architecture, has been commonly used to predict SIC. LSTM models generally have shown >0.98 correlation coefficient and <10% of root mean square error (RMSE) for a daily-scale SIC prediction [19], [21]. The monthly SIC predictions performed by LSTM models have shown <10%–12% of RMSE [18], [23], [24].

CNN and LSTM have also been used for SIV prediction in several studies. Zhai and Bitz [26] used CNN to predict SIV, and their model showed a correlation coefficient >0.8, outperforming other statistical and physical models. Similarly, the CNN model proposed by Hoffman et al. [25] showed approximately 0.8 correlation coefficient in SIV prediction. The convolutional LSTM presented by Petrou and Tian [38] showed 2.5–3.1 km/day of MAE and 3.7–4.2 km/day of RMSE in SIV prediction. However, to our knowledge, most previous ML studies only focused on deriving either SIC or SIV, not attempting to integrate SIC and SIV information. This study is the first to integrate SIC and SIV information in a single multitask neural network architecture, aiming to improve the prediction of both variables.

TABLE I
EXAMPLES OF DEEP LEARNING FOR SEA ICE PREDICTION

Sea ice variable	Method	References
Sea ice concentration	CNN	[14]–[17]
	LSTM	[18], [20], [23], [24]
	Conv-LSTM	[19], [21], [22]
Sea ice velocity	CNN	[25], [26]
	LSTM	[38]

C. Multitask Neural Network

In the existing multitask network studies, the most common architecture is the "shared-trunk": all tasks share certain hidden layers to learn common features for all tasks but keep several task-specific branched layers [39]. Conventionally, multitask networks can be categorized into two formulations depending on how to choose the branched level and share the information between tasks: 1) early-branched network and 2) late-branched network. The early-branched or late-branched approaches have been adopted in various fields for the classification and segmentation of remote sensing imagery data [40], [41], [42], [43]. Alhichri [40] used a multitasking architecture similar to a late-branch network for multiple land-use classification tasks. Papadomanolaki et al. [41] proposed a UNet-like multitask architecture to perform segmentation and change detection from multispectral image sets. Ilteralp et al. [42] used a multitask convolutional network for Chlorophyll-a estimation and month classification with satellite images. Li et al. [43] proposed a multitask network to perform mask prediction, edge prediction, and distance map estimation.

However, this shared-trunk architecture makes a model inflexible, so the model performance highly depends on where different branches split up. A late-branch network can lose significant task-specific features because nearly all network parameters are shared across all training processes [44]. Excessive sharing can also hinder the performance of another task that has different needs (also known as negative transfer). On the contrary, an early-branched network can fail to leverage information effectively between tasks because of too little sharing [39]. Hence, instead of the shared trunk architecture, another *crosstalk* approach has been proposed: each task has a separate network, but information flow is added between parallel layers of the task networks. For example, Misra et al. [45] proposed cross-stitch units between separate CNNs for different tasks, which provide the optimal linear combinations for a given set of tasks. Similarly, a sluice meta-network introduced by Ruder et al. [46] consists of a shared input layer, two task-specific output layers, and three hidden layers that have task-specific and shared subspaces. The input of each layer is determined as a linear combination of task-specific and shared outputs of the previous layers. Instead of a linear combination of the additional parameters, Gao et al. [47] used 1×1 convolution to perform neural discriminative dimensionality reduction (NDDR) to retain the discriminative information from task-specific and shared features. Similarly, He et al. [44] introduced task consistency learning (TCL) blocks in their hierarchically fused U-Net. However, to the best of our knowledge, no studies have conducted such a multitask

network for sea ice applications. This article proposes the first multitask network designed for SIC and SIV prediction.

III. DATA

We collect SIV and SIC satellite observation data as the input and output of the models. We also collect wind and air temperature data from the ERA5 climate reanalysis product as additional input variables. Sea ice physics model data is also collected to be used for a baseline comparison. We collect the data from 2016 to 2022 and filter out any data with null values or physically invalid values. In Sections III-A–III-D provide details of the datasets we use (see Table II).

A. Sea Ice Velocity

For the SIV data, we use the NSIDC polar pathfinder daily 25 km EASE-grid sea ice motion vectors version 4 [48], [49]. This product derives daily sea ice drift from three primary types of sources: 1) gridded satellite imagery (e.g., AVHRR, SSMI, SSMI/S, and AMSR); 2) NCEP/NCAR wind reanalysis data; and 3) buoy position data from the International Arctic Buoy Program (IABP). The u component (along-x) and v component (along-y) of sea ice motions are independently derived from each of these sources and optimally interpolated onto a 25 km equal-area scalable Earth (EASE) grid by combining all sources. When sea ice drift is derived from satellite image data, a correlation coefficient is calculated between a small target area in one image and a searching area in the second image. Then, the location in the second image where the correlation coefficient is the highest is determined as the displacement of ice [48]. The mean difference between the interpolated u components and the buoy vectors was approximately 0.1 km/day with a root mean square (rms) error of 2.9 km/day, and 0.3 km/day with an rms error of 2.9 km/day for v components [48]. The SIV extraction schemes of this dataset are valid at significant distances from the ice edges, in areas where ice conditions are relatively stable, stationary, homogenous, and isotropic day to day. Therefore, surface melting in the summer season can deteriorate drift accuracy because it affects the passive-microwave identification of ice parcels [49]. In this study, we use SIV values at least 50 km (or 2 pixels) away from the coastline.

B. Sea Ice Concentration

For the SIC data, we use NOAA/NSIDC climate data record (CDR) of passive microwave (PMW) SIC version 4 data [50]. This dataset provides a CDR of SIC (i.e., the areal fraction of ice within a grid cell) from PMW data, including SMMR, SSMI, and SSMI/S. The CDR algorithm output is the combination of SIC estimations from two algorithms: the NASA Team (NT) algorithm [51] and the NASA Bootstrap (BT) algorithm [52]. These empirical algorithms estimate SIC from the PMW brightness temperatures at different frequencies and polarizations (i.e., vertical and horizontal polarizations at 19, 22, and 37 GHz). Then, the CDR product adjusts algorithm coefficients for each sensor to optimize the consistency of daily and monthly SIC time series. Several assessments showed that the error of this SIC estimation is approximately 5% within

TABLE II
INPUT DATASETS FOR ML MODELS

Dataset	Name	Spatial resolution	Temporal resolution
Sea ice velocity (u and v)	NSIDC Polar Pathfinder Daily 25 km EASE-Grid Sea Ice Motion Vectors [49]	25 km	Daily
Sea Ice Concentration	NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration [50]	25 km	Daily
Wind velocity (u and v)	ECMWF Reanalysis v5 (ERA5) hourly data on single levels [57]	0.25°	Hourly
Air temperature	ECMWF Reanalysis v5 (ERA5) hourly data on single levels [57]	0.25°	Hourly
X Y coordinates	25 km Equal-Area Scalable Earth (EASE) Grid	25 km	-

the consolidated ice pack during cold winter conditions [53], [54], [55]. However, in the summer season, the error can rise to more than 20% due to surface melt and the presence of melt ponds [56]. Due to the data quality issue near coastal areas, we use the SIC data more than 50 km from the coastline. Since the projection of this SIC data (NSIDC Sea Ice Polar Stereographic North) is different from that of SIV data (EASE grid), we reproject SIC into the EASE grid using bilinear interpolation.

C. ERA5 Climate Reanalysis

ERA5 is the fifth generation ECMWF (European Center for Medium-Range Weather Forecasts) atmospheric reanalysis of the global climate covering the period from January 1940 to the present [57]. ERA5 provides hourly estimates of atmospheric, land, and oceanic climate variables [57]. We acquire the daily average wind velocity (u and v components) at 10 m height and 2 m air temperature from this hourly data. In addition, since the raw ERA5 data is gridded to a regular latitude-longitude grid of 0.25°, we reproject this data onto the 25 km EASE grid using bilinear interpolation to co-locate with the SIV data.

D. Sea Ice Physics Model

As a baseline to compare the performance of our ML model, we use the Arctic Ocean physics analysis and forecast data (level 4) provided by Copernicus Marine Services. This data uses the operational TOPAZ4 Arctic Ocean system [58] with a hybrid coordinate ocean model (HyCOM) [59]. It is run daily to provide ten days of forecast (average of ten members) of the 3-D physical ocean, including sea ice albedo, sea ice area fraction, SIT, and SIV. The original gridded data (12.5 km resolution at the North Pole on a polar stereographic projection) is also converted to the 25 km EASE grid.

IV. METHOD

Given that the interaction between SIC and SIV is extremely complicated by various thermodynamic and dynamic mechanisms, the multitask model for SIC and SIV prediction should be flexible and interactive. Hence, in order to share SIC and SIV information efficiently, we adopt the crosstalk multitask architecture [39] instead of the shared-trunk approach. In addition, based on the assumption that sharing SIC and SIV information throughout the learning process improves their prediction, we add WAMs between the SIC and SIV task-specific networks to allow sharing and highlighting of essential information. The details of the proposed multitask

network architecture and comparison with baseline models are described in this section.

A. Hierarchical Information-Sharing U-net

Fig. 1(a) describes the architecture of the multitask network we propose in this study: HIS-Unet. In this architecture, two separate SIV and SIC prediction tasks have their own branches after a common convolutional layer of 32 filters. Each branch has a U-net structure [60] consisting of the encoder (contracting path) and decoder (expansive path). Each encoder consists of repeated two 3 × 3 convolutions and a 2 × 2 max pooling operation with stride 2 for downsampling. The number of feature channels is doubled after each downsampling step. The decoder consists of several upsampling steps implemented by a 2 × 2 up-convolution that halves the number of feature channels, a concatenation with the cropped feature map, and two 3 × 3 convolutions [60]. The hyperbolic tangent (tanh) activation function is applied after each convolutional layer. Although each branch has a separate U-net construction that produces SIV or SIC, they share and transfer their information through six WAMs. By adopting linear weighting parameters and channel/spatial attention modules in WAMs, we enable these WAMs to 1) determine how much information from SIC and SIV is shared with each other and 2) highlight important channel and spatial information. We insert six WAMs right after the max pooling or up-convolution operations: three blocks in the encoder steps and three blocks in the decoder steps.

As shown in Fig. 2(a), each WAM first receives information from SIV and SIC branches and calculates the weighted sum of them. Letting a WAM receive the SIV feature map ($\xi_{in,SIV}$; image height H, image width W, and channels C) and SIC feature map ($\xi_{in,SIC}$; $H \times W \times C$) from the previous max-pooling or up-convolution operations, the input shared information ($\xi_{in,share}$) is determined by following:

$$\xi_{in,share} = A_{in,SIV}\xi_{in,SIV} + A_{in,SIC}\xi_{in,SIC} \qquad (3)$$

where $A_{in,SIV}$ and $A_{in,SIC}$ denote the weights for the SIV and SIC information, respectively. We initially set both $A_{in,SIV}$ and $A_{in,SIC}$ as 0.5, which indicates both SIV and SIC information are equally shared before training the model. As the model is trained with data, $A_{in,SIV}$ and $A_{in,SIC}$ are fine-tuned for the best accuracy. If $A_{in,SIV}$ is high, more SIV information is shared during the information sharing; if $A_{in,SIC}$ is high, more SIC information is shared.

After the input shared information $\xi_{in,share}$ is determined by the linear weighting combination of $\xi_{in,SIV}$ and $\xi_{in,SIC}$, this shared information passes through channel attention module
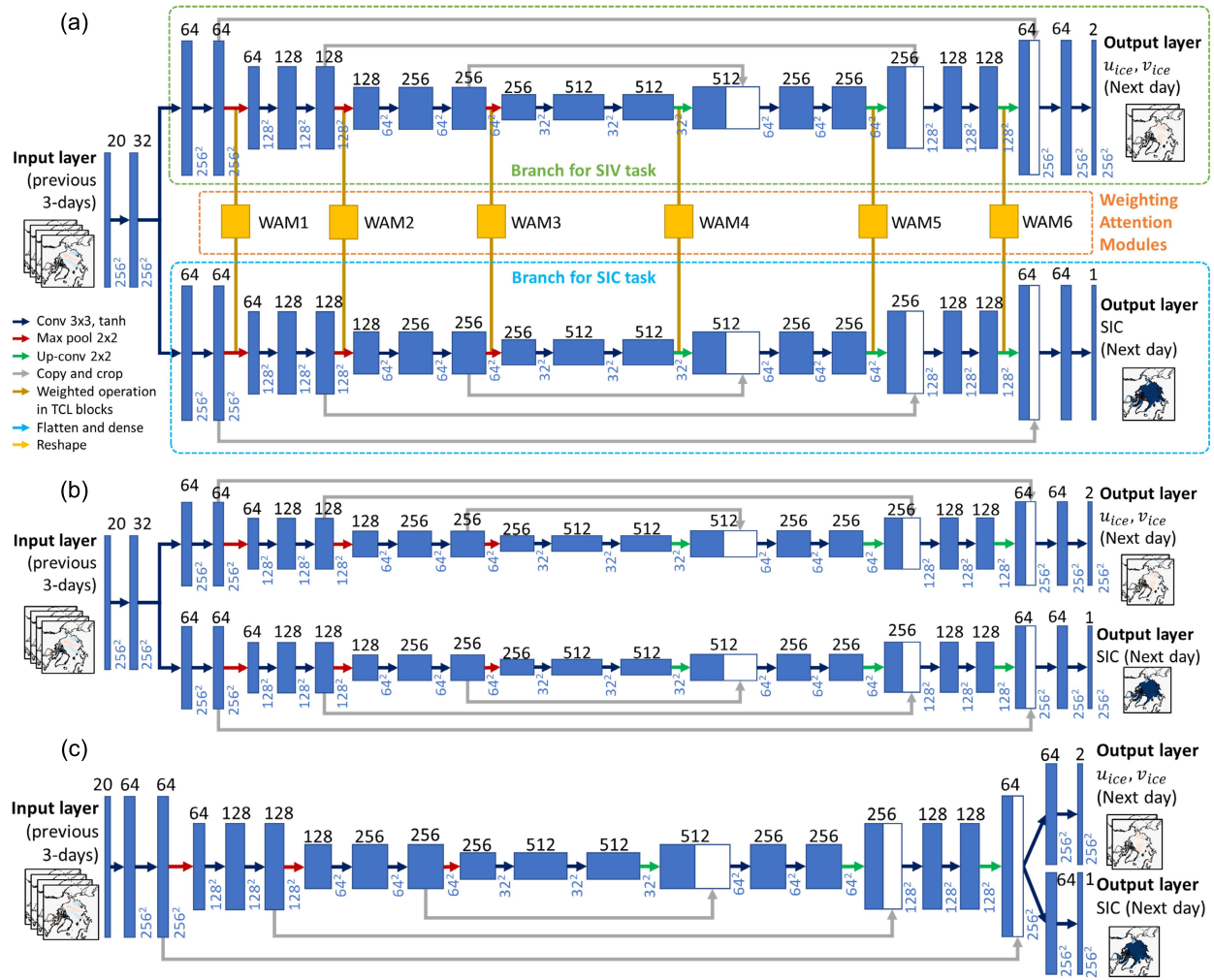
Fig. 1. Architecture of multitask CNN models tested in this study. (a) HIS-Unet. (b) EB-Unet. (c) LB-Unet.

[see Fig. 2(b)] and spatial attention module [see Fig. 2(c)]. The channel and spatial attention module were proposed by Woo et al. [61] and have been widely used to improve the representation power of CNNs. The channel attention highlights what channel is meaningful in a given input image, and the spatial attention highlights where an informative part is located [61]. In the channel attention module, the spatial information of an input feature map is aggregated using max pooling and average pooling operations, and these two descriptors are forwarded to a multilayer perceptron (MLP) with one hidden layer. Consequently, the channel attention map of $1 \times 1 \times C$ dimension is generated, and this channel attention map is multiplied with the input feature of $H \times W \times C$ dimension. After the channel attention is applied, the spatial attention is applied. In the spatial attention module, average pooling and max pooling are applied along the channel axis, and these two feature descriptors generate a spatial attention map ($H \times W \times C$ dimension) through a convolutional layer. As a result, these channel and spatial attention modules convert the input shared information $\xi_{\text{in,share}}$ into the output shared information ($\xi_{\text{out,share}}$)

$$\xi_{\text{out,share}} = M_s(\xi_{\text{in,share}}) \otimes (M_c(\xi_{\text{in,share}}) \otimes \xi_{\text{in,share}}) \quad (4)$$

where M_s and M_c denote spatial attention map and channel attention map derived from spatial and channel attention modules, respectively, and $\otimes$ denotes element-wise multiplication. If a certain channel is important in model training, this channel should have a higher channel attention value than the others. Similarly, if a specific spatial coordinate is important, this location is highlighted as higher value in the spatial attention. By arranging the channel attention and spatial attention sequentially, we can highlight what and where shared information is important. These channel and spatial attention modules are also applied to the input SIV and SIC feature maps ($\xi_{\text{in,SIV}}$ and $\xi_{\text{in,SIC}}$).

Then, the attention shared information is sent to the SIV and SIC branches after multiplying output weights ($A_{\text{out,SIV}}$ and $A_{\text{out,SIC}}$) and adding attention SIV and SIC information, respectively,

$$\xi_{\text{out,SIV}} = A_{\text{out,SIV}}\xi_{\text{out,share}}$$
$$+ M_s(\xi_{\text{in,SIV}}) \otimes (M_c(\xi_{\text{in,SIV}}) \otimes \xi_{\text{in,SIV}}) \quad (5)$$
$$\xi_{\text{out,SIC}} = A_{\text{out,SIC}}\xi_{\text{out,share}}$$
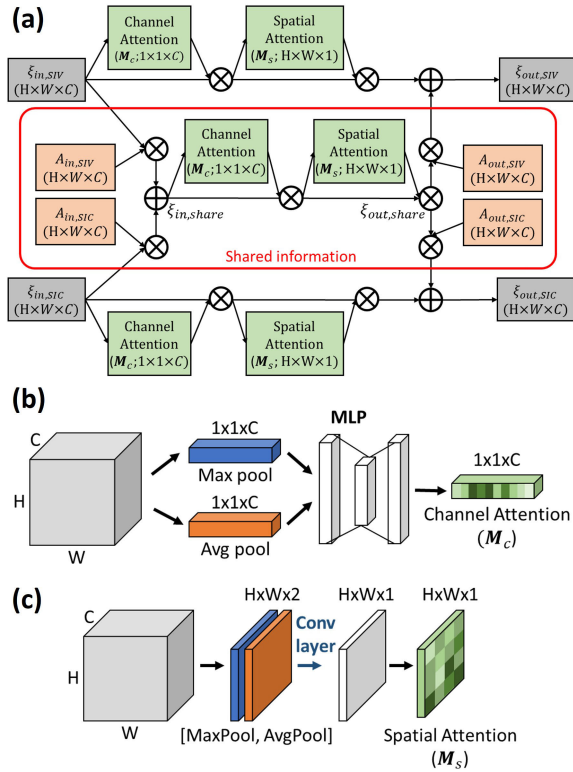$$+ M_s(\xi_{\text{in,SIC}}) \otimes (M_c(\xi_{\text{in,SIC}}) \otimes \xi_{\text{in,SIC}}). \quad (6)$$

Fig. 2. Schematic of (a) WAM, (b) channel attention module, and (c) spatial attention module.

Here, $A_{\text{out,SIV}}$ and $A_{\text{out,SIC}}$ determine the relative importance of shared information to each branch. A higher $A_{\text{out,SIV}}$ value means that more shared information is added to the SIV branch, implying that SIC information is helpful for SIV prediction. Likewise, a higher $A_{\text{out,SIC}}$ value corresponds to a greater contribution of SIV information to the SIC prediction. In Section VI-B, we will discuss details about how those weights in the WAMs affect the SIC and SIV predictions.

The objective loss function we use for learning is the sum of the mean square error (MSE) for u-component SIV (u), v-component SIV (v), and SIC (A)

$$\text{Loss} = \sum \left(|u_{\text{prd}} - u_{\text{obs}}|^2 + |v_{\text{prd}} - v_{\text{obs}}|^2 + \beta |A_{\text{prd}} - A_{\text{obs}}|^2 \right) \tag{7}$$

where the subscript obs means observation and prd means prediction, and β is the weight for SIC mse. When we experimented with β values of 0.1, 0.2, 0.5, and 1.0, $\beta = 0.5$ produced the best accuracy, balancing both SIV and SIC predictability. Thus, we set β to 0.5.

B. Baseline Models

To compare with the HIS-Unet, we evaluate five neural network models and two simple statistical models. Five neural network models are 1) early-branched U-net (EB-Unet), 2) late-branched U-net (LB-Unet), 3) U-net, 4) FCN only with convolutional layers, and 5) CNN proposed by Hoffman et al. [25] for SIV; two statistical models are 1) persistence and 2) linear regression (LR) models.

Based on the U-net architecture [60], we test two multitask networks: EB-Unet [see Fig. 1(b)] and LB-Unet [see Fig. 1(c)]. The EB-Unet only shares the first convolutional layer, but the rest of the convolutional layers are separated into SIV and SIC branches of independent U-net architecture. On the other hand, the LB-Unet has the same architecture as the original U-net, but the last convolutional layer and output layer are branched for SIV and SIC prediction separately. In addition to these multitask U-nets, we also test the original U-net architecture [60] with three channels of u-component SIV, v-component SIC, and SIC at the output layer.

Another neural network model we evaluate is a simple FCN consisting of seven convolutional layers of 64 filters for each and an output layer with three channels of u-component SIV, v-component SIV, and SIC. We refer to this network as FCN7. The final neural network we evaluate is the CNN architecture proposed by Hoffman et al. [25]. This model consists of five sequential layers of convolutional and max pooling, followed by the flattening, dense, and reshaping layers. Since the original CNN model [25] only predicts u- and v-component SIV, we add an SIC output channel at the output layer.

As a statistical model, the persistence model simply assumes that the sea ice condition remains the same as that of the previous day. The LR model, another statistical model, predicts u, v, and A at pixel (i, j) from the linear combination of the input variables as follows:

$$u(i, j) = \sum_{k=1}^{n} a_{u,k}(i, j) x_k(i, j) \tag{8}$$

$$v(i, j) = \sum_{k=1}^{n} a_{v,k}(i, j) x_k(i, j) \tag{9}$$

$$A(i, j) = \sum_{k=1}^{n} a_{A,k}(i, j) x_k(i, j) \tag{10}$$

where n is the number of input variables and x_k is the kth input variable. $a_{u,k}$, $a_{v,k}$, and $a_{A,k}$ are the linear coefficients corresponding to x_k for the prediction of u, v, and A, respectively.

C. Training and Testing Models

To predict the daily SIV and SIC, we use the previous three-days of SIV (u- and v-components), SIC, air temperature, wind velocity (u- and v-components), and geographic location (i.e., x and y coordinates) as the inputs of the prediction models. Consequently, the input layer has 20 channels (three-day data of six sea ice or meteorological variables and two geographical variables) of 256×256 grid size. All input values are normalized from -1 to 1 based on the nominal maximum and minimum values that each variable can have. Then, this model is trained to predict the SIV and SIC for the next day. We collect the data from 2016 to 2022; 2016–2021 datasets are used to train/validate the model, and 2022 datasets are used to test the model. The 2016–2021 datasets are randomly divided into 80% of train datasets and 20% of validation datasets. The number of train, validation, and test datasets is 1742, 438, and 363, respectively. The HIS-Unet is trained to minimize the loss function in (7) and

TABLE III
ACCURACY OF SIC AND SIV PREDICTION FOR EACH MODEL

Model	Sea ice concentration (SIC)			Sea ice velocity (SIV)		
	MAE (%)	RMSE (%)	R	MAE (km/d)	RMSE (km/d)	R
HIS-Unet	**2.934**	**6.122**	**0.978**	**1.812**	**2.667**	**0.834**
EB-Unet	3.139	6.238	0.974	1.926	2.773	0.818
LB-Unet	3.122	6.575	0.970	1.960	2.814	0.813
Unet	3.400	6.601	0.970	1.981	2.834	0.815
FCN7	4.750	7.896	0.958	2.331	3.213	0.770
CNN [25]	10.071	16.963	0.839	2.223	3.217	0.742
LR	4.232	8.178	0.965	3.340	5.142	0.780
Persistence	3.135	7.435	0.924	2.871	4.181	0.629
HYCOM [59]	8.874	15.878	0.865	3.782	5.506	0.508

optimized by Adam stochastic gradient descent algorithm [62] with 100 epochs and 0.001 learning rate. Such hyperparameter settings are the same for the other neural network models. We implement this model in Python using the PyTorch library. All scripts are executed on the GPU nodes of the Frontera supercomputing system from the Texas Advanced Computing Center (TACC), each node equipping four NVIDIA Quadro RTX 5000 GPUs of 16-GB memory.

D. Evaluation Metrics

The performances of the neural network models, statistical baseline models, and sea ice physics model are evaluated by the following three metrics: 1) correlation coefficient (R); 2) RMSE; and 3) MAE:

$$R = \frac{\sum (x - \overline{x})(y - \overline{y})}{\sqrt{\sum (x - \overline{x})^2 \sum (y - \overline{y})^2}} \tag{11}$$

$$RMSE = \sqrt{\frac{(x - y)^2}{N}} \tag{12}$$

$$MAE = \sum \frac{|x - y|}{N} \tag{13}$$

where x denotes the prediction values, y denotes the observation values, and N is the number of samples. In calculating these metrics for SIV, we calculate the average of the u- and v-components of SIV. In general, R can be useful to assess the overall spatiotemporal pattern of SIC and SIV variations, and RMSE or MAE can be useful to assess the magnitude of prediction errors.

V. RESULTS

Table III shows the test accuracy of SIC and SIV in 2022 for different models. The HIS-Unet model generally shows the best SIC and SIV prediction results, with the highest R (0.978 for SIC and 0.834 for SIV) and lowest RMSE (6.122% for SIC and 2.677 km/day for SIV), followed by EB-Unet, LB-Unet, Unet, and FCN7. These five fully CNN models (HIS-Unet, EB-Unet, LB-Unet, Unet, and FCN7) generally show better performance than the persistence model for both SIC and SIV, suggesting that these models can significantly predict the daily variations of sea ice conditions. In addition, these ML models outperform the LR model and HYCOM physical model. It is also worth mentioning that the CNN proposed by Hoffman et al. [25] has better fidelity than the persistence model in SIV but the worst performance in SIC

prediction. Since this CNN model was originally designed for only SIV prediction, this architecture does not produce robust predictions for SIC.

The EB-Unet shows considerable improvement in both SIV and SIC prediction compared to LB-Unet and U-net. This implies that separating SIC and SIV information at the early-branch stage is more efficient in predicting both variables than using the late-branch architecture or adding channels at the output layer. However, considering that such an early-branch approach prevents information sharing at the intermediate learning stages, the information sharing between SIC and SIV through WAMs in the middle of HIS-Unet architecture allows a better prediction for both SIC and SIV than the EB-Unet.

Fig. 3 depicts the monthly accuracy for each model. It is notable that the HIS-Unet has remarkably better performance during all months in both SIC and SIV prediction. In particular, the correlation coefficient derived by HIS-Unet outperforms the other ML models in both SIC and SIV prediction, which supports that the HIS-Uent produces the spatiotemporal variations of SIC and SIV more successfully [see Fig. 3(a) and (b)]. In terms of SIC prediction, the HIS-Unet shows a higher R by 0.002–0.006 and a lower RMSE by 0.05%–0.25% compared to the EB-Unet. In particular, the largest improvement occurs in the summer months from July to October. While R varies significantly from July to October by other models (<0.97), the HIS-Unet keeps $R > 0.97$ in these months, maintaining a similar level of R to winter months. October shows the most conspicuous improvement in SIC, reducing RMSE by 0.26% and increasing R by 0.006. In general, the Arctic SIE decreases rapidly in July and reaches the minimum in September. Then, sea ice starts to grow from September, recording the fastest ice growth rate in October [63]. As a result, the thermodynamic and dynamic sea ice condition changes more abruptly in the summer months. Integrating SIV information into SIC prediction via HIS-Unet might contribute to predicting SIC in this fast-melting and fast-growing season. Furthermore, as shown in Fig. 4(a), the improvement of R is observed near marginal sea ice zones, rather than the central Arctic (CA). This implies that SIV information can help predict spatiotemporal trends of SIC for where SIC changes dynamically in melting and freezing seasons. On the other hand, regarding the RMSE value of the SIC prediction, the HIS-Unet also contributes to reducing SIC errors slightly in the CA [see Fig. 4(b)].
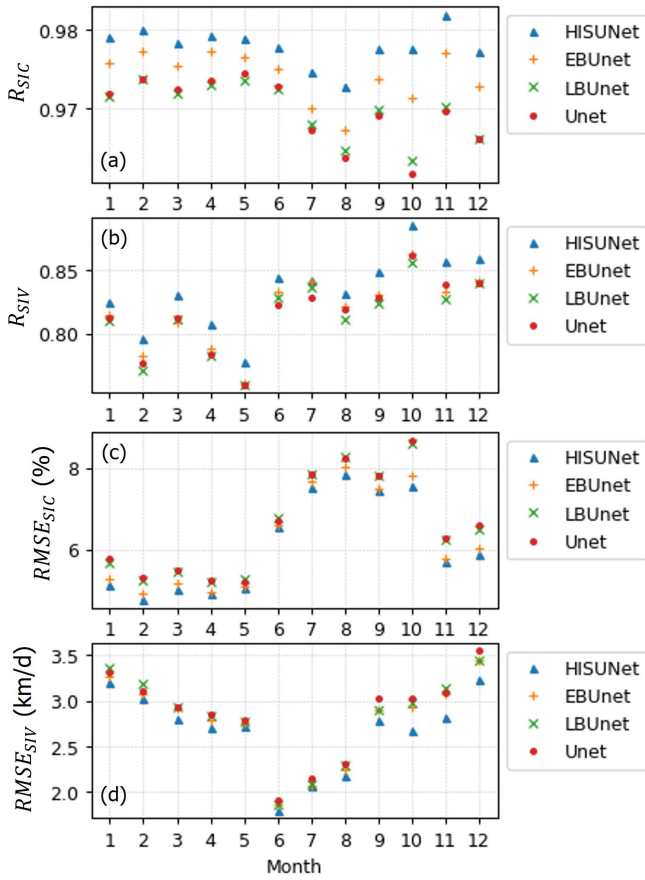
Fig. 3. Comparison of monthly accuracy of the models in 2022: (a) R of SIC, (b) R of SIV, (c) RMSE of SIC, and (d) RMSE of SIV.

Fig. 4. Difference of R and RMSE between HIS-Unet and EB-Unet: (a) R of SIC, (b) RMSE of SIC, (c) R of SIV, and (d) RMSE of SIV. All pixel values are the average of all days of 2022 when the SIC and SIV values of the pixels are valid.

accuracy changes for different regions of the Arctic Ocean. Furthermore, in Section VI-B, we examine the characteristics of information sharing between SIV and SIC layers, which occurs in WAMs. Finally, in Section VI-C, we discuss the climatological representativity of 2022 with respect to its impact on the training and testing of ML models.

A. Model Accuracy for Different Arctic Subregions

We assess the accuracy of HIS-Unet for six different sub-regions in the Arctic Ocean (see Fig. 5): CA, Chukchi and Beaufort Seas (CBS), Laptev and East Siberian Seas (LESS), Kara and Barents Seas (KBS), East Greenland (EG), and Hudson and Baffin Bays (HBB). The division of these Arctic subregions has been conducted in many previous studies on the basis of their unique atmospheric and sea ice characteristics [65], [66], [67], [68], [69]. Fig. 6 shows the RMSE and R of SIC and SIV prediction for these six subregions every month in 2022.

Regarding the SIC prediction, CA shows the most accurate prediction, with the highest R and lowest RMSE for all months. This should be because the CA region is almost covered by high-concentration sea ice all months [see Fig. 6(c)]. Since SIC remains stable at almost 100% for all seasons, the prediction of SIC in this region is relatively easier than in other regions. All other subregions also show R values greater than 0.95 and RMSE less than 10% for all seasons. The CA, CBS, and LESS regions show a seasonal trend in the SIC predictability: the RMSE increases in the summer months from June to October. The lower SIC prediction performance in the summer months might be because SIC changes significantly in these regions during the summer months, accompanied by fast melting and fast-growing [see Fig. 6(c)]. When these three regions are entirely covered by sea ice from January to April, the RMSE of SIC is less than 2% [see Fig. 6(b)]. KBS shows a low error of SIC from August to September because of its almost ice-free condition during these months,

As for the SIV prediction, the R of HIS-Unet is higher than the EB-Unet by 0.01–0.02 and the RMSE is lower by 0.05–0.25 km/day on average. In the case of SIV, the largest improvement by HIS-Unet occurs in October–December: RMSE decreases more than 0.2 km/day, and R increases by 0.02. During these months, the Arctic SIE starts to increase and sea ice becomes compacted. Such a compacted sea ice pack can either regulate or accelerate sea ice movement. Consequently, the integration of SIC information via HIS-Unet improves the SIV prediction when sea ice compactness increases in October–December. As shown in Fig. 4(c) and (d), the improvement of SIV prediction occurs in marginal sea ice zones out of the CA. Since the north of Greenland and the Canadian Archipelago are mainly covered by MYI that survives even in summer [64], the SIC in the CA does not change much over a year. Although the integration of SIC information via the HIS-Unet does not efficiently improve the SIV predictability in such a static region of the CA, it improves the SIV prediction broadly for the marginal ice zones where SIC changes seasonally.

VI. DISCUSSION

Considering that the interaction between SIV and SIC can vary by sea ice conditions (e.g., marginal sea ice zones versus CA), it is worth comparing the model accuracy for different regions. Thus, in Section VI-A, we discuss how the model
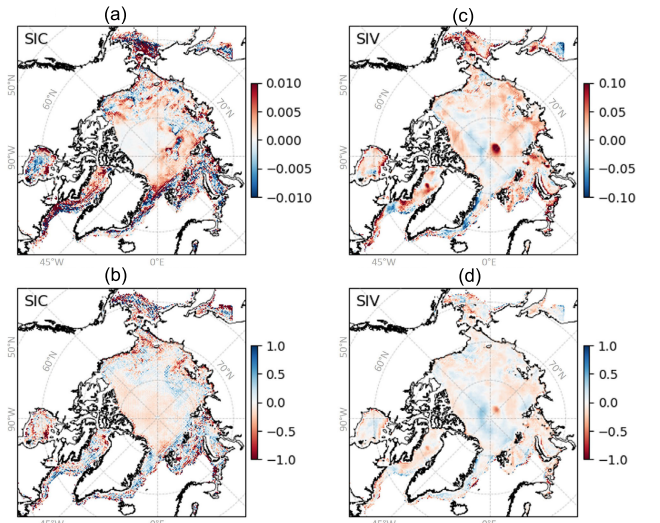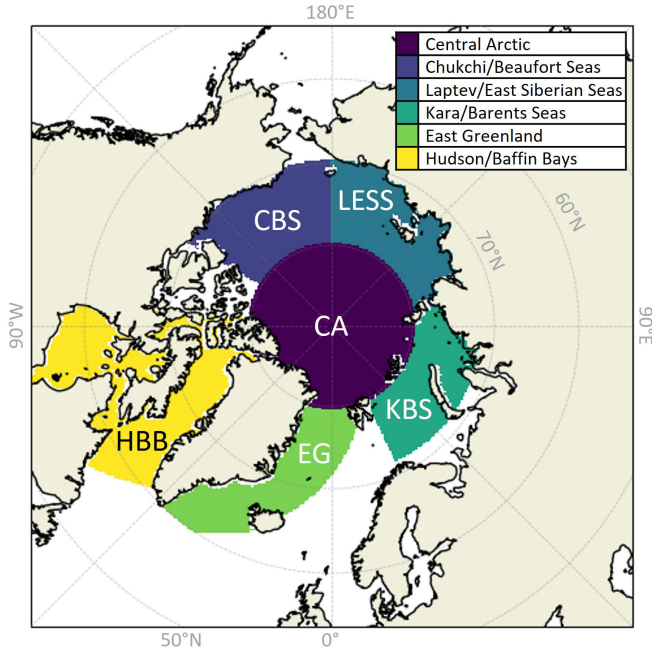
Fig. 5. Six Arctic subregions defined in this study: CA, CBS, LESS, KBS, EG, and HBB.

Fig. 6. SIC prediction accuracy by subregions and months: (a) R of SIC and (b) RMSE of SIC. Bright white indicates better accuracy and dark red indicates worse accuracy. The monthly mean and uncertainty of SIC are displayed in (c) and (d), respectively. SIV prediction accuracy by subregions and months: (e) R of SIV and (f) RMSE of SIV. The monthly mean and uncertainty of SIV are displayed in (g) and (h), respectively.

but accuracy deteriorates in the fast ice-growing season in October–November and melting season in May–July. HBB also shows a similar accuracy variation with higher errors in melting seasons from May to July. EG shows relatively stable accuracy in SIC with $R > 0.97$ and RMSE $< 5\%$, along with a consistent SIC all over the seasons.

Next, in terms of the SIV prediction, it is interesting that the prediction performance is relatively stable in the CA region. While the other regions experience large fluctuations in R and RMSE values, the fluctuation of R and RMSE in CA is not as large as the other regions. The R-value remains 0.8–0.9 and RMSE 3.0–5.0 km/day for most of the months in CA. In CBS and LESS, the RMSE of SIV does not exceed 4.5 km/day before October, but the RMSE increases up to 6 km/day in October–December. The KBS and HBB regions show abnormally high R and low RMSE values from July to October, but these values can be caused by very low SIC [see Fig. 6(c)] and slow sea ice drift [see Fig. 6(g)]. Compared to the other subregions, EG shows extremely fast SIV [see Fig. 6(c)] as a gateway to a magnificent amount of sea ice exports through the Fram Strait [70]. Despite such fast SIV, this region shows relatively better accuracy than the other subregions, with < 4 km/day of RMSE except for September.

Although our HIS-Unet model shows significant performance in both SIC and SIV prediction, here we note that the uncertainties in the target SIC and SIV observations should be taken into account. We calculate the monthly uncertainties in SIC and SIV observations for each subregion [see Fig. 6(d) and (h)], which can be obtained from the SIC and SIV datasets [48], [49]. The uncertainties of SIC are relatively low in CA, CBS, and LESS compared to those in KBS, EG, and HBB, and the SIC uncertainties increase substantially in the summer months in all subregions. On the other hand, the
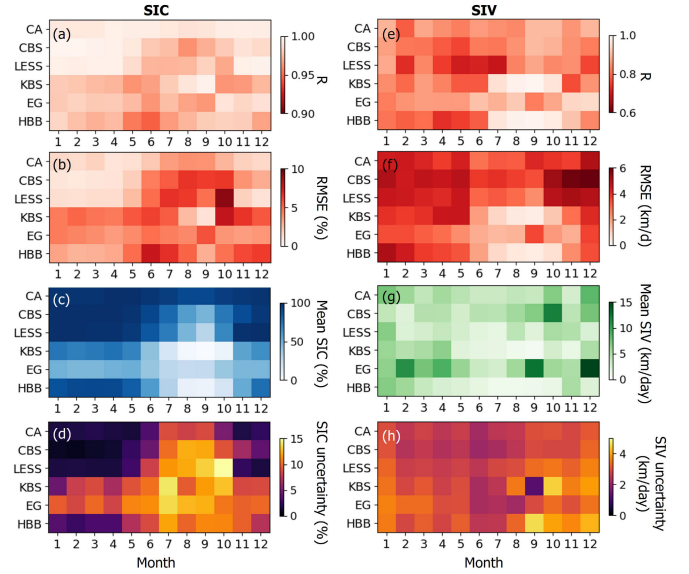
regional and seasonal variations in SIV uncertainties are not as significant as those of SIC. However, KBS and HBB show slightly larger SIC uncertainties compared to the other regions from August to October, when sea ice melts and grows fast.

B. Characterization of Information Sharing Pattern in WAMs

In the HIS-Unet architecture, the WAMs make it possible for the SIC and SIV prediction layers to share their information with each other. In the WAM, the information sharing is implemented by linear weight combination between SIC and SIV layers followed by channel and spatial attention modules. Moreover, the output feature from this WAM is passed through the multiplication of the output weights before being transferred to each branch. These input and output weights determine how much SIC and SIV information is to be mixed with each other and how significant this shared information is to be for each branch. Thus, we extract and examine the weight metrics of each WAM.

The input and output weights for WAM levels 1–6 are displayed in Fig. 7(a) and (b), respectively. Since each weight has the same dimension as the input layer ($H \times W \times C$), here we calculate the mean weights along all the channels ($H \times W$). In this section, we discuss the weighting values and spatial patterns exhibited only in the WAM1, WAM2, WAM5, and WAM6 because their $H \times W$ dimensions are closer to the original input and output dimension (256×256). There is a possibility that other intermediate WAMs represent fine-tuning the weights and biases in the model rather than weighted information sharing.

First, when the SIV and SIC information are first blended through WAM1, SIC information is more weighted than SIV with higher input weights. In particular, interesting is that a higher $A_{\text{in,SIC}}$ is observed near the EG and HBB regions. These
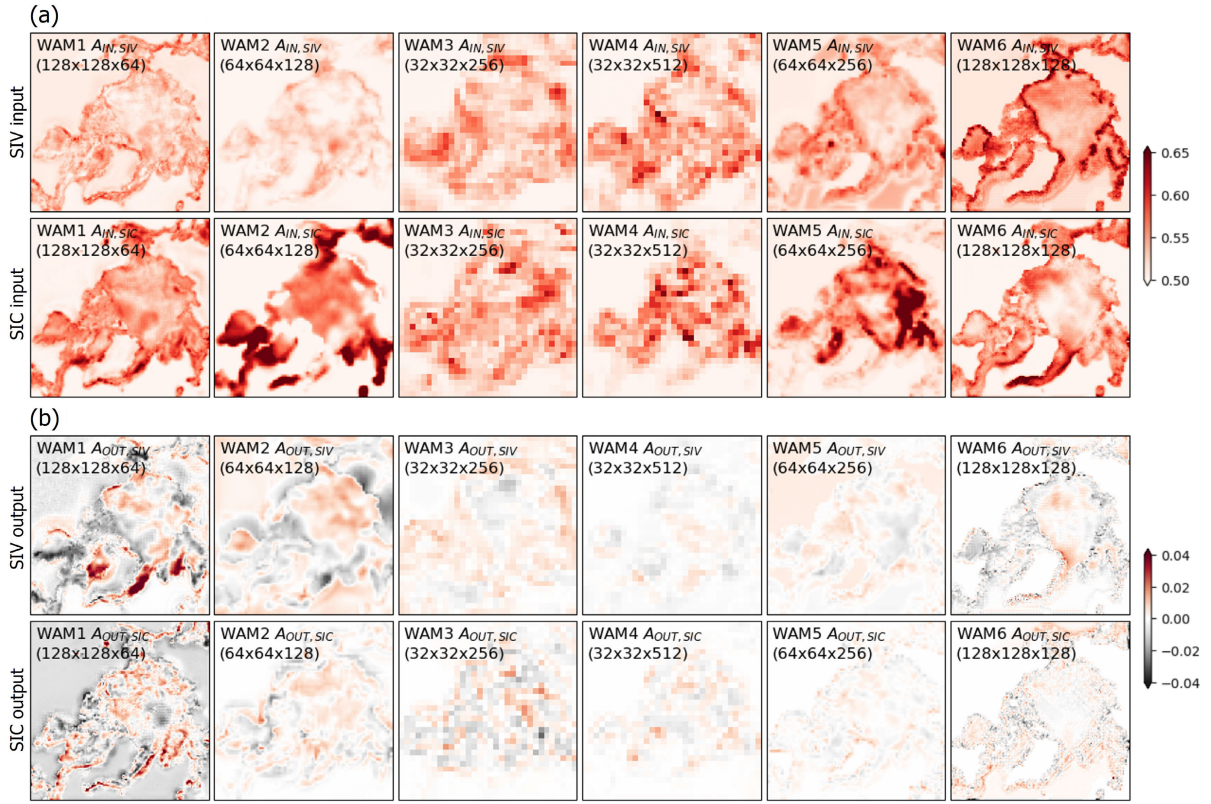
Fig. 7. (a) WAM input weights [$A_{\text{in,SIV}}$ and $A_{\text{in,SIC}}$; refer to Fig. 2 and (3)] and (b) WAM output weights [$A_{\text{out,SIV}}$ and $A_{\text{out,SIC}}$; refer to Fig. 2 and (5) and (6)] for each level of HIS-Unet in Fig. 1(a).

regions are known to have somehow different characteristics of sea ice dynamics from the other regions; HBB is relatively isolated from global oceanic currents in the CA (e.g., transpolar drift and Beaufort Gyre), and EG is a highly advective region [see the fastest ice velocity in Fig. 6(f)] because of the fast sea ice near the Fram Strait. A higher $A_{\text{in,SIC}}$ in these regions implies that SIC information is useful to predict such distinctive SIV characteristics. When this shared information is transferred to each branch, the shared information in the HBB, EG, and KBS regions is weighted for the SIV branch (i.e., shown as higher $A_{\text{out,SIV}}$), and the shared information in the EG and KBS regions is weighted for the SIC branch (i.e., shown as higher $A_{\text{out,SIC}}$). We emphasize that the shared information is highly weighted in EG for both SIV and SIC predictions, which suggests that the information sharing between SIV and SIC is beneficial for predicting SIV and SIC in extremely dynamic regions like EG. In general, the magnitudes of output weights are higher in the SIV branch compared to the SIC branch.

At the next level of information sharing at WAM2, SIC information is also more weighted in sharing information, and the SIV branch receives more shared information in transferring the shared information. When the WAM2 receives information from SIV and SIC branches, SIC information is significantly weighted near marginal sea ice zones out of the CA, particularly in KBS and HBB (i.e., higher $A_{\text{in,SIC}}$). This similar pattern is also observed in WAM5: more SIC contribution to shared information, particularly

near KBS. Given that the variation of SIC in KBS is significantly influenced by the inflow and outflow of sea ice from the CA [70], the collaboration of SIC and SIV information can improve the model predictability in this region.

At the last information sharing state (WAM6), it is noted that the SIV information is weighted along the coastline in the information sharing (i.e., high $A_{\text{in,SIV}}$ values). The SIC information is also weighted more in the coastal regions (i.e., high $A_{\text{in,SIC}}$ values), especially near the EG region, than in the CA. This indicates that the SIC information significantly contributes to predicting SIV in the EG region, in agreement with the results from WAM1. When the shared information transfers to each branch, the CBS and EG regions have higher $A_{\text{out,SIV}}$, while the magnitude of $A_{\text{out,SIV}}$ in the CA is close to zero. The spatial variability of $A_{\text{out,SIC}}$ in WAM6 appears negligible.

In summary, we find the following noticeable characteristics in information sharing through WAMs of HIS-Unet: 1) information sharing between SIC and SIV is relatively active in sea ice marginal zones compared to the CA; 2) when WAMs receive information from each branch and share them, SIC information is more weighted; 3) when WAMs pass the sharing information to each branch, the shared information is more weighted for the SIV branch than the SIC branch; and 4) the information sharing between SIC and SIV is helpful to predict SIV in the HBB, EG, and KBS regions that have distinctive sea ice dynamics characteristics.
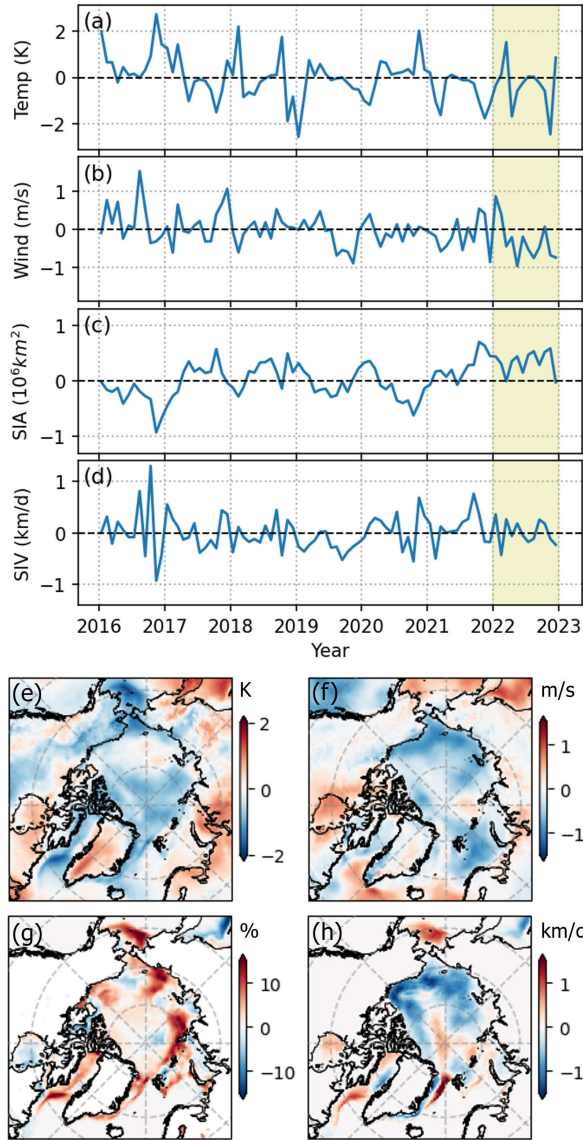
Fig. 8. Monthly anomaly of (a) air temperature, (b) wind speed, (c) sea ice area, and (d) SIV to the 2016–2021 monthly average in the Arctic Ocean. Sea ice area is calculated as the area where SIC >15%. Anomaly maps of (e) air temperature, (f) wind speed, (g) SIC, and (h) SIV in 2022 to the 2016–2021 average.

C. Climatological Representability in 2022

In this study, we use 2016–2021 data as training datasets and 2022 data as test datasets. If the climatological condition of 2022 is different from the other years, it can possibly make the model performance biased to the training datasets and reduce the representability of the model. Hence, in this section, we check if any climatological difference exists between training and testing years and how it affects the model performance. We compare the monthly anomalies of air temperature, wind speed, SIC, and SIV in 2022 to the average in 2016–2021. As shown in Fig. 8, 2022 showed lower temperature [see Fig. 8(a)], lower wind speed [see Fig. 8(b)], greater sea ice area [see Fig. 8(c)], and slower SIV than 2016–2021. The lower anomalies of air temperature and wind speed are observed across the entire Arctic Ocean [see Fig. 8(e) and (f)]. SIC

anomaly also occurs in most Arctic regions; in particular, LESS and KBS regions have extremely positive anomalies with >10% higher than the 2016–2021 average [see Fig. 8(g)]. Such SIC anomalies in LESS and KBS regions could bring relatively higher RMSE of SIC in these regions [see Fig. 6(b)]. The slow SIV in CBS and LESS [see Fig. 8(h)] appears somehow related to the negative anomaly of wind speed in these regions [see Fig. 8(f)].

Herein, we emphasize that our HIS-Unet model shows significant performance in such abnormal climate and sea ice conditions in 2022 (as shown in Table III and Figs. 3 and 6), even though the model is trained only with 2016–2021 data. Interesting is that the spatial distribution of SIC anomaly [see Fig. 8(g)] is somehow overlapped with where HIS-Unet shows higher R than the EB-Unet [see Fig. 4(a)]. Introducing SIV information into SIC prediction could be beneficial to predicting the SIC anomaly condition in 2022. Regarding SIV, it is worth mentioning that the negative SIV anomaly regions in CBS and LESS show a slight improvement of SIV prediction by HIS-Unet [see Fig. 4(c) and (d)].

VII. CONCLUSION

In this study, we propose a multitask FCN to predict Arctic SIC and SIV simultaneously. We accomplish the information sharing between SIC and SIV layers by adding six WAMs between separate SIC and SIV U-net branches. The important SIC and SIV information is shared and highlighted in WAMs by following procedures: 1) intermediate SIC and SIV features are mixed with multiplying linear input weights; 2) the shared information is highlighted through channel and spatial attention modules; and 3) the highlighted shared information is transferred to SIC and SIV branches with multiplying linear output weights, respectively. Since SIC and SIV affect each other in thermodynamic and dynamic ways, the information sharing and attention through WAMs can indirectly allow the network to learn their complicated interactions.

Our HIS-Unet shows better performance than other network architectures, including convolutional network, U-net, LB-Unet, and EB-Unet. In particular, our HIS-Unet outperforms these models in the melting or early freezing seasons and shows a stable performance in the CA, where the SIC is relatively consistent throughout the year. For the other regions, although the model performance varies by the sea ice conditions, the fidelity of HIS-Unet is close to or better than the previous SIC and SIV prediction studies. When the input and output weights of WAMs are examined, we find that the information-sharing scheme of HIS-Unet plays an important role in predicting SIC and SIV in the marginal sea ice zones where sea ice conditions change dynamically, such as EG, HBB, and Krara/Barents Seas. In these regions, WAMs explicitly contribute to SIC and SIV predictions by 1) assigning higher weights to SIC information in these regions when mixing SIC and SIV information and 2) assigning higher weights to shared information in these regions when transferring shared information to the SIC and SIV branches. Furthermore, the information sharing between SIC and SIV

also helps predict the anomalous sea ice conditions in 2022, even though this model is trained without this year's data.

REFERENCES

[1] D. J. Cavalieri and C. L. Parkinson, "Arctic sea ice variability and trends, 1979–2010," *Cryosphere*, vol. 6, no. 4, pp. 881–889, Aug. 2012.

[2] J. C. Stroeve et al., "Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations," *Geophys. Res. Lett.*, vol. 39, no. 16, pp. 1–7, Aug. 2012.

[3] D. Notz and J. Stroeve, "Observed Arctic sea-ice loss directly follows anthropogenic CO_2 emission," *Science*, vol. 354, no. 6313, pp. 747–750, Nov. 2016.

[4] I. H. Onarheim, T. Eldevik, L. H. Smedsrud, and J. C. Stroeve, "Seasonal and regional manifestation of Arctic sea ice loss," *J. Climate*, vol. 31, no. 12, pp. 4917–4932, Jun. 2018.

[5] R. Lindsay and A. Schweiger, "Arctic sea ice thickness loss determined using subsurface, aircraft, and satellite observations," *Cryosphere*, vol. 9, no. 1, pp. 269–283, Feb. 2015.

[6] R. Kwok, "Arctic sea ice thickness, volume, and multiyear ice coverage: Losses and coupled variability (1958–2018)," *Environ. Res. Lett.*, vol. 13, no. 10, Oct. 2018, Art. no. 105005.

[7] X. Yu et al., "Evaluation of Arctic sea ice drift and its dependency on near-surface wind and sea ice conditions in the coupled regional climate model HIRHAM–NAOSIM," *Cryosphere*, vol. 14, no. 5, pp. 1727–1746, May 2020.

[8] R. Kwok, G. Spreen, and S. Pang, "Arctic sea ice circulation and drift speed: Decadal trends and ocean currents," *J. Geophys. Res., Oceans*, vol. 118, no. 5, pp. 2408–2425, May 2013.

[9] J. Anheuser, Y. Liu, and J. R. Key, "A climatology of thermodynamic vs. dynamic Arctic wintertime sea ice thickness effects during the CryoSat-2 era," *Cryosphere*, vol. 17, no. 7, pp. 2871–2889, Jul. 2023.

[10] T. J. W. Wagner, I. Eisenman, and H. C. Mason, "How sea ice drift influences sea ice area and volume," *Geophys. Res. Lett.*, vol. 48, no. 19, Oct. 2021, Art. no. e2021GL093069.

[11] E. Blockley et al., "The future of sea ice modeling: Where do we go from here?" *Bull. Amer. Meteorolog. Soc.*, vol. 101, no. 8, pp. E1304–E1311, 2020.

[12] E. C. Hunke, W. H. Lipscomb, and A. K. Turner, "Sea-ice models for climate study: Retrospective and new directions," *J. Glaciol.*, vol. 56, no. 200, pp. 1162–1172, 2010.

[13] E. Blanchard-Wrigglesworth, R. I. Cullather, W. Wang, J. Zhang, and C. M. Bitz, "Model forecast skill and sensitivity to initial conditions in the seasonal sea ice outlook," *Geophys. Res. Lett.*, vol. 42, no. 19, pp. 8042–8048, Oct. 2015.

[14] T. R. Andersson et al., "Seasonal Arctic sea ice forecasting with probabilistic deep learning," *Nature Commun.*, vol. 12, no. 1, p. 5124, Aug. 2021.

[15] Y. J. Kim, H.-C. Kim, D. Han, S. Lee, and J. Im, "Prediction of monthly Arctic sea ice concentrations using satellite and reanalysis data based on convolutional neural networks," *Cryosphere*, vol. 14, no. 3, pp. 1083–1104, p. 5124, Mar. 2020.

[16] Y. Ren and X. Li, "Predicting daily Arctic sea ice concentration in the melt season based on a deep fully convolution network model," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2021, pp. 5540–5543.

[17] Y. Ren, X. Li, and W. Zhang, "A data-driven deep learning model for weekly sea ice concentration prediction of the pan-Arctic during the melting season," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4304819.

[18] J. Wei, R. Hang, and J.-J. Luo, "Prediction of pan-Arctic sea ice using attention-based LSTM neural networks," *Frontiers Mar. Sci.*, vol. 9, pp. 1–14, Jun. 2022.

[19] Q. Liu, R. Zhang, Y. Wang, H. Yan, and M. Hong, "Daily prediction of the Arctic sea ice concentration using reanalysis data based on a convolutional LSTM network," *J. Mar. Sci. Eng.*, vol. 9, no. 3, p. 330, Mar. 2021.

[20] Q. Zheng, W. Li, Q. Shao, G. Han, and X. Wang, "A Mid- and long-term Arctic sea ice concentration prediction model based on deep learning technology," *Remote Sens.*, vol. 14, no. 12, p. 2889, Jun. 2022.

[21] Q. Liu, R. Zhang, Y. Wang, H. Yan, and M. Hong, "Short-term daily prediction of sea ice concentration based on deep learning of gradient loss function," *Frontiers Mar. Sci.*, vol. 8, pp. 1–13, Sep. 2021.

[22] Y. Liu, L. Bogaardt, J. Attema, and W. Hazeleger, "Extended-range Arctic sea ice forecast with convolutional long short-term memory networks," *Monthly Weather Rev.*, vol. 149, no. 6, pp. 1673–1693, 2021.

[23] J. Chi and H.-C. Kim, "Prediction of Arctic sea ice concentration using a fully data driven deep neural network," *Remote Sens.*, vol. 9, no. 12, p. 1305, Dec. 2017.

[24] B. Mu, X. Luo, S. Yuan, and X. Liang, "IceTFT v1.0.0: Interpretable long-term prediction of Arctic sea ice extent with deep learning," *Geosci. Model Develop.*, vol. 16, no. 16, pp. 4677–4697, Aug. 2023.

[25] L. Hoffman et al., "Machine learning for daily forecasts of Arctic sea ice motion: An attribution assessment of model predictive skill," *Artif. Intell. Earth Syst.*, vol. 2, no. 4, pp. 1–45, Oct. 2023.

[26] J. Zhai and C. M. Bitz, "A machine learning model of Arctic sea ice motions," 2021, *arXiv:2108.10925*.

[27] P. R. Holland and R. Kwok, "Wind-driven trends in Antarctic sea-ice drift," *Nature Geosci.*, vol. 5, no. 12, pp. 872–875, Dec. 2012.

[28] P. R. Holland and N. Kimura, "Observed concentration budgets of Arctic and Antarctic sea ice," *J. Climate*, vol. 29, no. 14, pp. 5241–5249, Jul. 2016.

[29] W. D. Hibler, "A dynamic thermodynamic sea ice model," *J. Phys. Oceanogr.*, vol. 9, no. 4, pp. 815–846, Jul. 1979.

[30] A. S. Thorndike and R. Colony, "Sea ice motion in response to geostrophic winds," *J. Geophys. Res., Oceans*, vol. 87, no. C8, pp. 5845–5852, Jul. 1982.

[31] K. Maeda, N. Kimura, and H. Yamaguchi, "Temporal and spatial change in the relationship between sea-ice motion and wind in the Arctic," *Polar Res.*, vol. 39, pp. 1–10, Nov. 2020.

[32] S.-H. Baek, E.-J. Kang, B.-J. Sohn, S.-W. Kim, and H. Shi, "Development of ANN-based algorithm to estimate wintertime sea ice temperature profile over the Arctic ocean," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4302117.

[33] J. Huang, J. Chen, Y. Mu, C. Cao, and H. Shen, "Remote-sensing monitoring of colored dissolved organic matter in the Arctic ocean," *Mar. Pollut. Bull.*, vol. 204, Jul. 2024, Art. no. 116529.

[34] A. Stokholm, T. Wulf, A. Kucik, R. Saldo, J. Buus-Hinkler, and S. M. Hvidegaard, "AI4SeaIce: Toward solving ambiguous SAR textures in convolutional neural networks for automatic sea ice concentration charting," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4304013.

[35] X. Li et al., "Deep-learning-based information mining from ocean remote-sensing imagery," *Nat. Sci. Rev.*, vol. 7, no. 10, pp. 1584–1605, Oct. 2020.

[36] G. Gao, L. Yao, W. Li, L. Zhang, and M. Zhang, "Onboard information fusion for multisatellite collaborative observation: Summary, challenges, and perspectives," *IEEE Geosci. Remote Sens. Mag.*, vol. 11, no. 2, pp. 40–59, Jun. 2023.

[37] W. Li, C.-Y. Hsu, and M. Tedesco, "Advancing Arctic sea ice remote sensing with AI and deep learning: Opportunities and challenges," *Remote Sens.*, vol. 16, no. 20, p. 3764, Oct. 2024.

[38] Z. I. Petrou and Y. Tian, "Prediction of sea ice motion with convolutional long short-term memory networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6865–6876, Sep. 2019.

[39] M. Crawshaw, "Multi-task learning with deep neural networks: A survey," 2020, *arXiv:2009.09796*.

[40] H. Alhichri, "Multitask classification of remote sensing scenes using deep neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 1195–1198.

[41] M. Papadomanolaki, M. Vakalopoulou, and K. Karantzalos, "A deep multitask learning framework coupling semantic segmentation and fully convolutional LSTM networks for urban change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7651–7668, Sep. 2021.

[42] M. Ilteralp, S. Ariman, and E. Aptoula, "A deep multitask semisupervised learning approach for Chlorophyll–A retrieval from remote sensing images," *Remote Sens.*, vol. 14, no. 1, p. 18, Dec. 2021.

[43] M. Li, J. Long, A. Stein, and X. Wang, "Using a semantic edge-aware multi-task neural network to delineate agricultural parcels from remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 200, pp. 24–40, Jun. 2023.

[44] K. He et al., "HF-UNet: Learning hierarchically inter-task relevance in multi-task U-Net for accurate prostate segmentation in CT images," *IEEE Trans. Med. Imag.*, vol. 40, no. 8, pp. 2118–2128, Aug. 2021.

[45] I. Misra, A. Shrivastava, A. Gupta, and M. Hebert, "Cross-stitch networks for multi-task learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3994–4003.

[46] S. Ruder, J. Bingel, I. Augenstein, and A. Søgaard, "Latent multi-task architecture learning," in *Proc. 33rd AAAI Conf. Artif. Intell. 31st Innov. Appl. Artif. Intell. Conf. 9th AAAI Symp. Educ. Adv. Artif. Intell.*, 2019, pp. 4822–4829.

[47] Y. Gao, J. Ma, M. Zhao, W. Liu, and A. L. Yuille, "NDDR-CNN: Layerwise feature fusing in multi-task CNNs by neural discriminative dimensionality reduction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3200–3209.

[48] M. Tschudi, J. S. Meier, W. N. Stewart, C. Fowler, and J. Maslanik, "Polar pathfinder daily 25 km ease-grid sea ice motion vectors, version 4," Nat. Snow Ice Data Center, Boulder, CO, USA, 2019.

[49] M. A. Tschudi, W. N. Meier, and J. S. Stewart, "An enhancement to sea ice motion and age products at the national snow and ice data center (NSIDC)," *Cryosphere*, vol. 14, no. 5, pp. 1519–1536, May 2020.

[50] W. N. Meier, F. Fetterer, A. K. Windnagel, and J. S. Stewart, "NOAA/NSIDC climate data record of passive microwave sea ice concentration, version 4," Nat. Snow Ice Data Center, Boulder, CO, USA, 2021.

[51] D. J. Cavalieri, P. Gloersen, and W. J. Campbell, "Determination of sea ice parameters with the NIMBUS 7 SMMR," *J. Geophys. Res., Atmos.*, vol. 89, no. 4, pp. 5355–5369, 1984.

[52] J. C. Comiso, "Characteristics of Arctic winter sea ice from satellite multispectral microwave observations," *J. Geophys. Res., Oceans*, vol. 91, no. 1, pp. 975–994, Jan. 1986.

[53] W. N. Meier, "Comparison of passive microwave ice concentration algorithm retrievals with AVHRR imagery in Arctic peripheral seas," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1324–1337, Jun. 2005.

[54] J. C. Comiso, D. J. Cavalieri, C. L. Parkinson, and P. Gloersen, "Passive microwave algorithms for sea ice concentration: A comparison of two techniques," *Remote Sens. Environ.*, vol. 60, no. 3, pp. 357–384, Jun. 1997.

[55] N. Ivanova et al., "Inter-comparison and evaluation of sea ice algorithms: Towards further identification of challenges and optimal approach using passive microwave observations," *Cryosphere*, vol. 9, no. 5, pp. 1797–1817, Sep. 2015.

[56] S. Kern, T. Lavergne, D. Notz, L. T. Pedersen, and R. Tonboe, "Satellite passive microwave sea-ice concentration data set inter-comparison for Arctic summer conditions," *Cryosphere*, vol. 14, no. 7, pp. 2469–2493, Jul. 2020.

[57] H. Hersbach et al., "The ERA5 global reanalysis," *Quart. J. Roy. Meteorol. Soc.*, vol. 146, no. 730, pp. 1999–2049, Jul. 2020.

[58] P. Sakov, F. Counillon, L. Bertino, K. A. Lisæter, P. R. Oke, and A. Korablev, "TOPAZ4: An ocean-sea ice data assimilation system for the North Atlantic and Arctic," *Ocean Sci.*, vol. 8, no. 4, pp. 633–656, Aug. 2012.

[59] E. P. Chassignet et al., "The HYCOM (HYbrid Coordinate Ocean Model) data assimilative system," *J. Mar. Syst.*, vol. 65, nos. 1–4, pp. 60–83, Mar. 2007.

[60] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, vol. 9349, no. 18, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Cham, Switzerland: Springer, Oct. 2015, pp. 234–241.

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

[62] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015, pp. 1–13.

[63] F. Fetterer, K. Knowles, and W. N. Meier, "Sea ice index, version 3 [data set]," Nat. Snow Ice Data Center, Boulder, CO, USA, 2017.

[64] W. Meier and J. Stroeve, "An updated assessment of the changing Arctic sea ice cover," *Oceanography*, vol. 35, nos. 3–4, pp. 10–19, 2022.

[65] C. L. Parkinson, D. J. Cavalieri, P. Gloersen, H. J. Zwally, and J. C. Comiso, "Arctic sea ice extents, areas, and trends, 1978–1996," *J. Geophys. Res., Oceans*, vol. 104, no. C9, pp. 20837–20856, Sep. 1999.

[66] J. E. Overland and M. Wang, "Future regional Arctic sea ice declines," *Geophys. Res. Lett.*, vol. 34, no. 17, pp. 1–7, Sep. 2007.

[67] I. S. Cabral, I. R. Young, and A. Toffoli, "Long-term and seasonal variability of wind and wave extremes in the Arctic ocean," *Frontiers Mar. Sci.*, vol. 9, pp. 1–13, May 2022.

[68] J. E. Walsh, H. Eicken, K. Redilla, and M. Johnson, "Sea ice breakup and freeze-up indicators for users of the Arctic coastal environment," *Cryosphere*, vol. 16, no. 11, pp. 4617–4635, Nov. 2022.

[69] S. Chen, K. Li, H. Fu, Y. C. Wu, and Y. Huang, "Sea ice extent prediction with machine learning methods and subregional analysis in the Arctic," *Atmosphere*, vol. 14, no. 6, p. 1023, Jun. 2023.

[70] R. Kwok, "Outflow of Arctic ocean sea ice into the Greenland and Barents seas: 1979–2007," *J. Climate*, vol. 22, no. 9, pp. 2438–2457, May 2009.
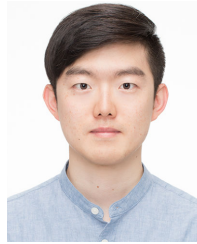
Younghyun Koo (Member, IEEE) received the B.S. degree in energy resources engineering and the M.S. degree in energy systems engineering from Seoul National University, Seoul, Republic of Korea, in 2017 and 2019, respectively, and the Ph.D. degree in environmental science and engineering from the University of Texas at San Antonio, San Antonio, TX, USA, in 2023.

He is currently a Post-Doctoral Research Associate with the Computer Vision and Remote Sensing Laboratory (Bina Laboratory), Lehigh University, Bethlehem, PA, USA. His research interests include the application of machine learning and remote sensing techniques for environmental monitoring, particularly in polar regions.
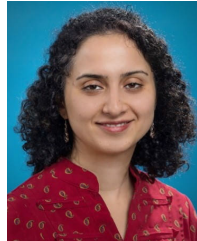
Maryam Rahnemoonfar (Member, IEEE) received the Ph.D. degree in computer science from the University of Salford, Manchester, U.K., in 2010.

She is currently an Associate Professor and the Director with the Computer Vision and Remote Sensing Laboratory (Bina Laboratory), Lehigh University, Bethlehem, PA, USA. Her research interests include deep learning, computer vision, data science, AI for social good, remote sensing, and document image analysis. Her research specifically focuses on developing novel machine learning and computer vision algorithms for heterogeneous sensors such as radar, sonar, multispectral, and optical.

Dr. Rahnemoonfar's research has been funded by several awards including the NSF HDR Institute Award-iHARP, NSF BIGDATA Award, Amazon Academic Research Award, Amazon Machine Learning Award, Microsoft, and IBM.