Sea Ice Concentration Estimation: Using Passive Microwave and SAR Data With a U-Net and Curriculum Learning

Keerthijan Radhakrishnan, *Student Member, IEEE*, K. Andrea Scott, *Member, IEEE*, and David A. Clausi, *Senior Member, IEEE*

***Abstract*—Ice concentration estimates are typically acquired from algorithms using passive microwave satellite data, and from image analysis charts, but these have limitations. Estimates acquired from passive microwave data have coarse spatial resolution, may have errors due to atmospheric contamination, and often perform poorly in marginal ice zones. Image analysis charts are not as precise, subject to analyst interpretation, and only available over specific geographic areas. We have implemented a U-net with synthetic aperture radar images as inputs and use ice concentration estimates retrieved from passive microwave data as training labels. The U-net, due to not being sensitive to patch size, is shown to be an improvement over previous work with convolutional neural networks that use fully connected layers at the output. Data augmentation and an L1 loss function were applied along with a novel training scheme that leverages curriculum learning. In this training scheme, the model is first trained with samples from open water and consolidated ice regions before incorporating samples from marginal ice regions. In a tenfold cross validation experiment, we achieve 3–4% mean absolute error comparing to estimates using passive microwave data and observe curriculum learning models having more stable training. Predictions on four with-held SAR scenes with difficult ice conditions were evaluated with image analysis charts. A mean absolute error of 7.18% is achieved, which is lower than errors associated with passive microwave data alone. Qualitative improvements in marginal ice zone estimates are achieved, while still preserving smooth consolidated ice regions, and openings in ice cover.**

***Index Terms*—Convolutional neural network (CNN), curriculum learning, fully convolutional network (FCN), ice concentration.**

I. Introduction

SEA ice concentration is calculated as a numeric value between zero and one, defined as the total area of ice in a specified region divided by the total area of that region. Navigators are interested in sea ice concentration because routes with significant ice cover can be dangerous and time consuming. Vessels often get stuck in ice which delays delivery time and expends fuel, with negative economic and environmental consequences. Captains of ice-class vessels have indicated that better knowledge of openings in the ice cover (leads) is important for safety [1]. For climate scientists, more accurate ice concentration information enables better models and ability to predict climate change occurrences [2].

Synthetic aperture radar (SAR) sensors and passive microwave sensors are more often used to monitor sea ice in arctic regions than optical sensors because they are less affected by cloud cover and do not depend on solar illumination. Data provided from these sensors can also be used to obtain ice concentration estimates. Images acquired from passive microwave sensors have a lower spatial resolution than SAR images and they are affected by atmospheric moisture. There are numerous algorithms using passive microwave data for ice concentration retrieval [3]. Algorithms using lower frequencies have a lower spatial resolution, while higher frequency algorithms have higher spatial resolution, but they are more affected by atmospheric moisture. For example, the smallest instantaneous field of view (IFOV) available from present passive microwave sensors is 3–5 km at 89 GHz, for which the atmospheric contamination in the marginal ice zone is significant. Fig. 1 shows a region, where two different algorithms using passive microwave data provide very different sea ice concentration estimates in a marginal ice zone.

The present study demonstrates improved ice concentration retrievals in marginal ice zones through the use of SAR data. Although there are benefits to using SAR data to retrieve ice concentration estimates, it is a difficult process because of the complexity of the interaction of the SAR signal with water and ice. The backscatter of the SAR signal is dependent on the imaging geometry, surface conditions, and ice types. Open water that is calm generally has a lower intensity than ice. Rough water conditions from wind and ocean currents can increase the backscatter. Previous studies have shown that deep learning models are a suitable choice for estimating ice concentration from SAR because they can learn backscatter patterns of ice and water [4]–[6]. The present study builds upon the previous work of using deep learning models for ice concentration estimation.

In this study, we utilize SAR imagery, which contains details of the marginal ice zones, in tandem with sea ice concentration

Manuscript received December 29, 2020; revised March 21, 2021; accepted April 21, 2021. Date of publication April 27, 2021; date of current version June 8, 2021. This work was supported in part by Marine Environmental Observation, Prediction and Response (MEOPAR) and in part by Environment and Climate Change Canada (ECCC). *(Corresponding author: Keerthijan Radhakrishnan.)*

Keerthijan Radhakrishnan is with the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: kradhakr@uwaterloo.ca).

K. Andrea Scott and David A. Clausi are with the University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: ka3scott@uwaterloo.ca; dclausi@uwaterloo.ca).

Digital Object Identifier 10.1109/JSTARS.2021.3076109

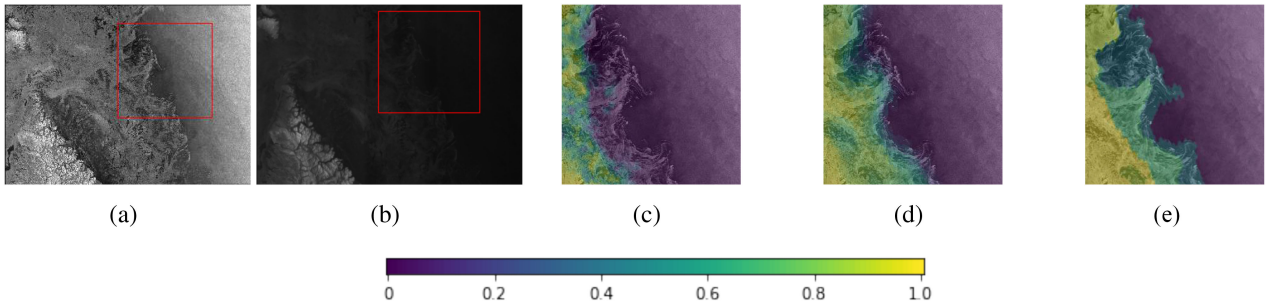

Fig. 1. Estimates of sea ice concentration on marginal ice zone region from algorithms using passive microwave data and from image analysis charts. (a) HH Polarized RADARSAT-2 SAR image and (b) HV Polarized RADARSAT-2 SAR image acquired 02/27/2018. Marginal ice zone region is outlined by a red box. (c) ASI estimates, (d) NT2 estimates, (e) image analysis chart estimates overlaid on HH Polarized RADARSAT-2 SAR image for region outlined by red box. The estimate acquired from the ASI algorithm underpredicts the ice concentration. Although some of this difference could be due to difference between the acquisition times of the passive microwave and SAR data, a bias of this nature has been observed in an earlier study [4]. The estimate acquired from the NT2 algorithm does not underpredict the marginal ice zone region, but details in this region are being missed due to the coarse resolution of the sensors used. Image analysis charts are estimates given to large polygons of homogeneous regions and do not capture fine details.

from passive microwave sensors, to obtain an improved estimate of sea ice concentration. Previous studies have used convolutional neural networks (CNNs) with fully connected layers that make predictions for each pixel on SAR imagery [4], [5]. We propose a deep learning model with a U-net architecture, which is in essence a fully convolutional network (FCN) with skip connections and lacks fully connected layers. This architecture has significant benefits over the traditionally used models that make predictions for each pixel as it is not dependent on patch size and is able to make predictions with finer details and less noise. Parallel studies have used U-nets for sea ice cover information and have seen good results [7], [8]. We also introduce a novel training scheme based on curriculum learning to improve estimates in marginal ice zones. The authors do not know of any papers employing curriculum learning in this manner.

The contributions of the curriculum learning training scheme can be summarized as follows.

1) Improved estimation in marginal ice zones where small scale details are captured while not contaminating the open water and consolidated ice regions using a relatively small dataset.
2) Increased stability during model training.

II. Background

With a greater number of SAR images and ice concentration labels being available, deep learning methods have garnered popularity for their ability to learn suitable features from SAR images. Deep learning for ice concentration estimation has been studied as early as 2014, where multilayer perceptron neural networks were trained using HH/HV SAR channels along with incidence angle information to obtain ice concentration estimates [9], [10]. Since then, many advancements in deep learning have been made and deeper models have shown great success in ice concentration estimation. In this section, an overview of deep learning architectures and their research relevant to this study is given.

A. Convolutional Neural Networks

Recently, CNNs have shown significant potential for sea ice concentration estimation from SAR [4]–[6]. A CNN is a deep learning model that consists of both convolutional and fully connected layers. The output of a convolutional layer is the convolution operation between the input and a number of filters. This provides the model with spatial information from the images. Before the fully connected layers, the output from the convolutional layers is flattened to a 1D vector. Between convolutional layers, there are often pooling operations to reduce the overall size of the input to the next layer.

CNNs have been successful in estimating ice concentration when using image analysis chart estimates as training labels [4], [6]. Passive microwave data has also been used as training labels for a DenseNet model, a CNN with direct connections between all layers [11], to estimate ice concentration on the Gulf of St. Lawrence and Arctic Archipelago regions [5]. The limitation of the aforementioned CNNs for ice concentration estimation is that they make predictions for a single pixel each iteration. A parallel study utilized atrous convolutions with spatial pyramid pooling for ice concentration estimation, which allowed multidimensional predictions to be made in one pass and eliminated the need for downscaling [6]. In the present study, we follow a different approach to make multidimensional predictions that utilizes a U-net.

B. U-Net

In the domain of remote sensing, U-nets are primarily used for segmentation problems [12], [13], but have also been used for regression problems such as pansharpening, and have achieved good results [14]. U-nets have also been used successfully for sea ice cover segmentation problems [7], [8]. Therefore, in this study we investigate the ability of a U-net to estimate sea ice concentration.

A U-net is an FCN, a CNN that lacks fully connected layers. FCNs are advantageous as their input is not restricted to a specified patch size and they are able to make predictions for full SAR images in one pass. A U-net is a unique FCN which contains encoding and decoding stages along with skip connections [15]. Skip connections, which connect the downsampling and upsampling branches of the U-net, reintroduce features that may have been lost in the downsampling stages. This allows the model to capture fine details in ice cover.

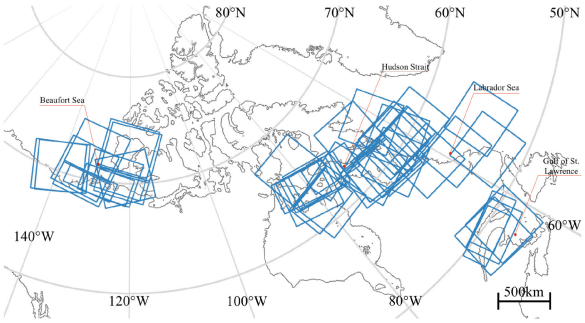

Fig. 2. Scenes taken from Hudson Strait, Gulf of St. Lawrence, Labrador Sea, and Beaufort Sea were used in this study. Outlines of scenes used are shown in blue.

C. Curriculum Learning

Studies have found that models struggle to learn when presented with full datasets at once [16]. Curriculum learning is the structured training of a machine learning model by progressively increasing the difficulty of the training data. This has been proven to increase model performance, speed up convergence, and improve generalizability [16], [17]. Curriculum learning is most commonly used for classification tasks where there is a disparity in the difficulty of the classes or in the difficulty within classes [16], [18], [19]. It has also been explored for classification tasks with class imbalance and shown success [20]. Applying curriculum learning for sea ice concentration estimation is desirable because marginal ice zones are more difficult to estimate and not as abundant in datasets as consolidated ice and open water regions.

D. Data and Study Region

The study area, Hudson Strait, which connects the North Atlantic to Hudson Bay, was chosen for its current usage and anticipated usage by shipping vessels as well as its ice conditions [21]. The study period is January–March 2018. In this time period, a variety of ice types can be seen, such as new ice, ice floes, ice eddies and filaments, in addition to grey, grey-white, and thicker first year ice.

In this study, SAR images, estimates acquired from passive microwave data, and estimates acquired from image analysis charts were used.

The SAR images used were RADARSAT-2 ScanSAR Wide images of HH and HV polarization. Uncalibrated SAR data was used because a significant difference was not observed when using calibrated SAR data in a previous study [5]. The grid size of this data is 50 m. Outside the study area, additional scenes from the Gulf of St. Lawrence (February 2011), Beaufort Sea (April 2015/2016), and Labrador Sea (February 2011) were also included in the training as a form of data augmentation to increase the number and variety of consolidated ice and open water conditions. These regions consist of different ice types than those from the Hudson Strait, mainly thicker consolidated ice regions and some open water regions. The location of the SAR images used for this study are shown in Fig. 2.

Two sea ice concentration estimates obtained from passive microwave data were used in this study. The first is from the Institute of Environmental Physics at the University of Bremen [22]. These estimates were obtained by applying the ARTIST sea ice (ASI) algorithm [22] to microwave radiometer data of the advanced microwave scanning radiometer 2 (AMSR-2) sensor on the JAXA satellite GCOM-W1 [22]. The 89 GHz channels were used because they provide the finest spatial resolution among all channels from AMSR-2. The grid size of ice concentration estimates from this data source is 3.125 km [22]. The higher frequency channels used by the ASI algorithm may have fine resolution, but they also have some shortcomings. The 89-GHz channels are heavily influenced by atmospheric water vapor and cloud liquid water. As is the case with most sea ice concentration algorithms, they use a weather filter which is a bulk correction for atmospheric opacity that sets the ice concentration to zero [22]. The weather filter is problematic because it is difficult to distinguish an atmospheric signature from intermediate ice concentration in the marginal ice zone and sometimes the weather filters remove thin or diffuse ice instead of erroneous retrievals due to weather. These filters also use lower frequency channels of the passive microwave sensors, which have coarser spatial resolution. The second sea ice concentration estimates used were obtained from the NASA team 2 (NT2) algorithm [23]. The NT2 algorithm is an enhanced version of the NASA team (NT) algorithm [23] that estimates ice concentration from passive microwave data of frequencies 18.7–89 Ghz [24]. The lower frequencies are affected less by atmospheric conditions and cloud cover. The grid size for ice concentration estimates from this data source is 12.5 km, which is coarser than that for ASI estimates [23]. NT2 estimates have also been shown to struggle in diffuse ice conditions and new ice (16% bias) [25]. Therefore, ice concentration estimates obtained from this source are not as precise as the ASI estimates. Estimates acquired from passive microwave data has been shown to be sensitive to the presence of melt water on the ice, atmospheric water vapor, and cloud liquid water [3], [26], [27].

Our last data source is image analysis charts. In this visual interpretation, an ice analyst assigns ice concentration values to regions that are considered homogeneous in appearance, but this also has some shortcomings. First, with increasing volumes of SAR images being available, automated methods for ice concentration estimation from SAR images would be ideal. Moreover, the precision of the image analysis charts is no more than 10% because ice concentration estimates are given in increments of 10%. Furthermore, image analysis charts provide a single label for a large spatial region with homogeneous ice characteristics. The ice concentration at a specific point may be different than the label given to the region, for example these charts do not capture small-scale details such as floes and openings in the ice cover. Image analysis charts have also been known to have bias due to subjectivity of the ice experts [28]. Image analysis charts were acquired for the evaluation stage.

III. Methodology

In this study, we implement a U-net model employing a novel training method to estimate sea ice concentration and we perform experimentation to validate the model design choices. We develop our model using SAR images as input and use

TABLE I
SCENES USED FOR TRAINING AND EVALUATION

Set	Location	Dates	No. Scenes
Basic Train	Hudson Strait	January-March 2018	22
Enhancement Set	Gulf of St. Lawrence	February 2011	6
	Labrador Sea	February 2011	4
	Beaufort Sea	April 2015	8
Hudson Strait Evaluation	Hudson Strait	January-March 2018	4

passive microwave data, specifically ASI estimates, as training labels. Estimates from passive microwave data were preferred as training labels over image analysis charts because they covered a larger number of images from our dataset and they are not subject to analyst interpretation. Furthermore, if this model is to be extended to a larger dataset, it would be more difficult to obtain image analysis charts. ASI estimates were chosen over NT2 estimates because they capture finer details, such as ice cracks [29]. NT2 estimates and image analysis chart estimates were reserved for the test set as a form of cross validation. Our methodology to develop such a model is described in this section. Data processing was completed in Python and models were built using the PyTorch framework on Python [30].

A. Data Processing

Before developing a model, some processing of the raw data is required. This consists of processing the SAR data, splitting the dataset to train/test sets, annotation of the images, and extracting patches for training.

For SAR image processing, an 8×8 downsample was performed, where each 8×8 block was replaced with the average value within the block. This put the grid size as 400 m. The downsampling was performed to make the images more manageable for training and to reduce speckle noise. Next, we create the model input by stacking the HH and HV SAR images. We will refer to the stacked images of shape $X \times Y \times 2$ as simply the SAR images.

The full dataset was split into training and test sets. The data from the Hudson Strait was split into a training set, *basic train*, and an evaluation set, *Hudson Strait Evaluation*. Only images where there was a reasonable visual agreement between SAR and ASI estimates were used for basic train. Another training set, *enhancement set*, was also made using the data outside the Hudson Strait. The enhancement set was kept separate from the basic train to study the effects of incorporating a region with different ice types into the training. The SAR datasets are described in more detail in Table I. ASI estimates were acquired for the full dataset while NT2 and image analysis chart estimates were acquired for the Hudson Strait Evaluation set only.

Next, we annotate the SAR images with the ASI, NT2, and image analysis chart estimates. To annotate the images, nearest neighbor lookup process with the estimates was performed to obtain an ice concentration value at each pixel on the SAR image [31]. Using these annotations, we generate new images with the same length and width as the SAR images. We use these annotated images as training labels and as labels to evaluate our model predictions.

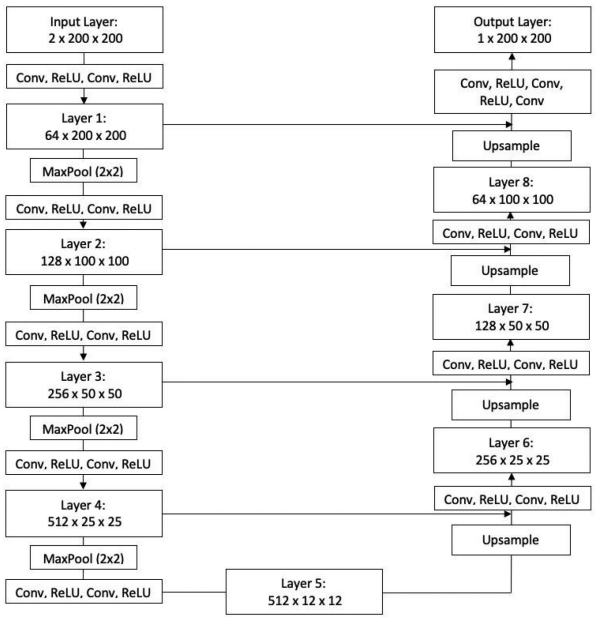

Fig. 3. U-net structure.

The U-net requires patches due to memory limitations. Therefore, the last step is extracting patches from the SAR images and the annotated images for training. Patches of size 250×250 were extracted away from land with a stride of 50, which resulted in 2860 patches from basic train and 3709 patches from enhancement set. Only the center 200×200 was used to calculate the loss and update the weights. This is done to prevent boundary effects. When making predictions, patches are extracted with a stride of 200 and predictions are made for the center 200×200 region.

B. U-Net Architecture

For this problem, a U-net consisting of 19 convolutional layers and 4 downsampling operations was used. The U-net architecture was chosen based on a cross validation experiment with U-nets of varying depths. The chosen U-net architecture is shown in Fig. 3. The convolution operation is performed with 3×3 filters for all convolutional layers. Downsampling was performed by a 2×2 max pooling operation which replaces each 2×2 block with the maximum value in that block. The U-net has a rectified linear unit (ReLU) activation function to introduce nonlinearity between each layer [32]. Transpose convolution operation with a stride of 2 is used to upsample the image [33]. The optimizer used for the U-net was stochastic gradient descent

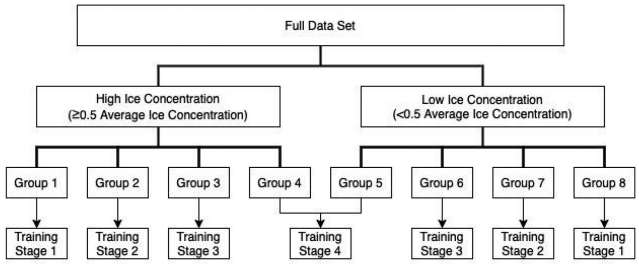

Fig. 4. Training stages for training scheme based on curriculum learning. Four equal-sized groups were created from the high-ice concentration and low-ice concentration set with lower group numbers having scenes with a higher average ice concentration.

with an initial learning rate of 1×10^1. The learning rate was decayed every 40 epochs by a factor of 10 for a total of 160 epochs. The decay rate was chosen to be every 40 epochs as the validation and training loss plateaus at this point.

C. Training Method

To benefit both from the wider range of consolidated ice and open water in enhancement set and the marginal ice zone samples in basic train, a revised training schedule based on curriculum learning was developed. To do this, we split the the training dataset into two groups, high-ice concentration ($\geq$0.5 ice concentration) and low-ice concentration set ($<$0.5 ice concentration), based on average ice concentration in the patch. Next, within each set, the patches were arranged by increasing average ice concentration and split into four groups of equal size. Finally, opposing groups from the high-ice concentration set and low-ice concentration set were joined to form the four training stages. We number these training stages from one to four by increasing levels of difficulty (increasing marginal ice). This process is outlined in Fig. 4.

The learning method is different when implementing curriculum learning. Initially, the model is trained only with training stage 1. The other training stages are progressively added every 20 epochs. After training with the full set for 20 epochs, learning rate decay begins with a factor of 10 every 20 epochs to reach a final learning rate of 1×10^{-4}.

IV. Experiments

In this section, we perform experiments to validate our proposed methods and we establish design choices that are most suitable for sea ice concentration estimation. To choose the optimal model, the overall accuracy and the ability to generalize to many scenes was considered. The model must also overcome the low bias and coarse resolution issue, shown in Fig. 1. Hudson Strait evaluation was used for a qualitative comparison of different models.

Images included in Hudson Strait evaluation are shown in Fig. 5. Three of these images (a), (b), (d) capture the dynamic marginal ice zone along the Labrador coast. Furthermore, in Fig. 5(b), the ASI algorithm underpredicts the ice concentration in the vicinity of the ice edge. Fig. 5(c) was chosen because it contains smooth ice, ice that lacks cracks, which can be problematic for machine learning models.

A. Traditional CNN Versus U-Net

For our study, a CNN consisting of three convolutional layers and two fully connected layers was used as the baseline model, referred to as *baseline CNN* hereafter. Baseline CNN consisted of 2×2 maxpooling between convolutional layers, ReLU activation functions between each layer, and a final linear layer. Increasing model depth or using models designed for other applications, such as that from [5], did not have improvements and did not solve the shortcomings of traditional CNNs, such as patch dependency, discussed in Section I. Therefore, a deeper CNN architecture was not selected as the baseline. The input to the baseline CNN is a patch and the output is an ice concentration estimate for the center pixel of the patch. The optimizer used for the CNN was Adam with an initial learning rate of 1×10^3 [34]. The learning rate was reduced by a factor of 10 every 20 epochs for a total of 100 epochs. This learning structure was used because it took roughly 20 epochs for the validation loss to plateau at each learning rate and reduction in loss was minimal after 80 epochs.

The drawback to the baseline CNN is that the model's FOV is the size of the patch. Therefore, the spatial information received by the model is that which is contained in the patch. FCNs, such as the proposed U-net, do not have this problem because they are not restricted to a patch size. To see the effects of different patch sizes for a CNN, three models were trained that used patches of size 25×25, 45×45, and 65×65. These models are referred to as CNN-25, CNN-45, and CNN-65, hereafter. For training, patches were extracted away from land from basic train using a stride equal to half the patch size. This resulted in training sets of size 280 836, 64 700, and 29 955 patches for CNN-25, CNN-45, and CNN-65, respectively. To keep the baseline CNN model consistent, the 25×25 and 45×45 patches were upsampled to 65×65 using bilinear interpolation. The predictions made by the baseline CNNs are shown in Fig. 6.

CNN-25 captures the fine details in the SAR image, but produces a noisy output due to the lack of spatial information. Predictions from CNN-45 and CNN-65 retain fewer details, but are less noisy because larger patches are less discriminative. For example, when the patch size is large, the patches of two neighboring pixels will have more overlap. Therefore, it would be difficult for a CNN to differentiate patches of neighboring pixels, resulting in a blurrier output. Furthermore, a large patch provides the CNN with more spatial information but may also include features that are irrelevant to the ice concentration of the center pixel.

Note that patch size is not the only difference between CNN-25, CNN-45, and CNN-65. The FOV of the kernels used in the convolution operations are also different. To confirm that patch size is the reason predictions made by CNN-65 and CNN-45 are blurrier than CNN-25, another experiment was conducted. The kernel size was increased to 8×8 for CNN-25 and increased to 4×4 for CNN-45. This was done so that the kernel's FOV would cover approximately the same physical dimensions for each patch size. In this experiment CNN-45 and CNN-65 were again blurrier than CNN-25.

Sea ice concentration estimates with a U-net [Fig. 7(e)–(h)] retain more details than those from the baseline CNN, consistent

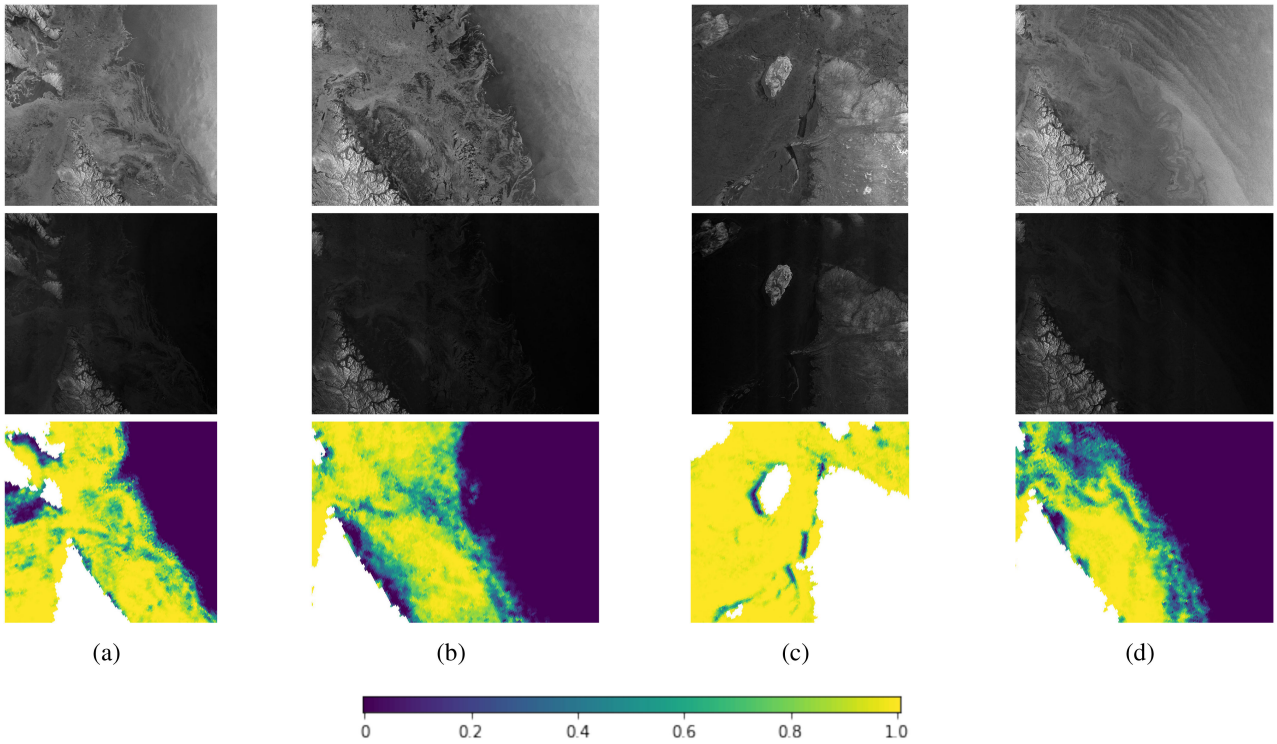

Fig. 5. SAR HH/HV polarized images and ASI estimates taken from Hudson Strait. Dates and center locations. (a) 01/20/2018 (60.83°N, 62.85°W). (b) 02/27/2018 (60.14°N, 61.06°W). (c) 03/04/2018 (61.52°N, 78.64°W). (d) 02/03/2018 (59.832°N, 61.18°W). All images show details in the SAR images being missed in ASI estimates.

with other studies [15]. Another benefit of the U-net is that it requires less time to make predictions than the traditional CNN models. For example, the baseline CNN required approximately 1.5 h to make predictions for the 4 test images, while the U-net requires less than 7 s. This is due to the time required for baseline CNN to iterate through every pixel on the image. Note that the training time for the U-net was higher than for the CNN because the U-net had more parameters to train. The U-net required 2.5 h to train, while the CNN only required 30 min.

Another observation from this experiment is that all models also predict water free from noise. This was a property of all models in our study and it is partially due to the truncation of values to a range between zero and one. The models generally perform well in open water conditions, but they suffer with open water that is very rough, as shown in Fig. 5(d). Rough water conditions were limited in the training set. More exposure to such conditions may improve the performance of the model.

B. Loss Function

In this section, we build on earlier studies, such as [4], [5], and explore the effects of training with L1 and L2 loss functions. The two loss functions are shown in (1) and (2) and they are commonly used for regression problems

$$\text{Loss}_{L1} = \frac{1}{n}\sum_{i=1}^{n} |f(x^{(i)}, \theta) - y^{(i)}| \tag{1}$$

$$\text{Loss}_{L2} = \frac{1}{n}\sum_{i=1}^{n} (f(x^{(i)}, \theta) - y^{(i)})^2. \tag{2}$$

Outputs to the model trained with L2 loss function and L1 loss function is shown in Fig. 7(e)–(h) and (i)–(l) respectively. The model trained with L2 loss function has a bias toward lower ice concentration in all images and suffers from banding effect to a greater extent. The output from the model trained with the L2 loss function retained more of the shortcomings of the ASI estimates because L2 loss takes the square of the difference between the predicted value and the label field. Therefore, larger differences affect the loss terms more than small ones. Larger error values could be due to the ASI algorithm labelling ice regions as water and the L2 loss function forces the model to fit these situations to obtain a lower loss in the training set. Consequently, attempting to fit consolidated ice regions as water will make the model acquire a bias toward lower ice concentration values. Models trained with an L1 loss function are not affected to the same degree because there is no amplification of large errors.

C. Dataset Augmentation

To provide the models with a greater variety of ice and open water conditions, we use additional images (enhancement set) as a form of data augmentation. Outputs from the U-net trained with basic train + enhancement set and L1 loss function are shown in Fig. 7(m)–(p).

Comparing with the models trained with basic train, we can see that the bias is improved. The underpredicting of ice near land in Fig. 7(n) does not occur and the banding effect issue is eliminated. The model also has better predictions for smooth ice, as seen in Fig. 7(o). These benefits are attributed to the

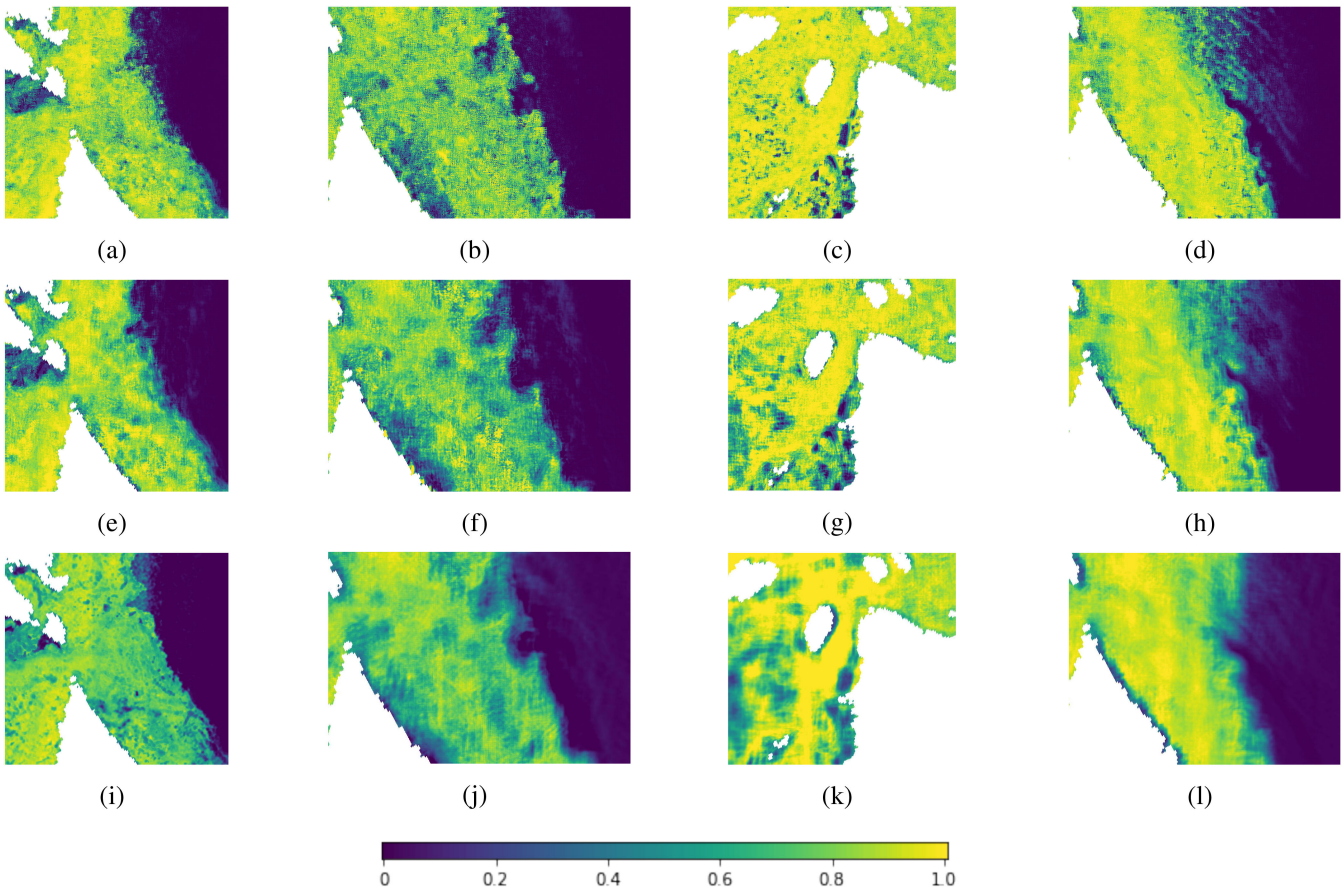

Fig. 6. Comparison of baseline CNN predictions using different patch sizes as input. (a)–(d) CNN trained with 25×25 patches output noisier images but also capture more details. (e)–(h) CNN trained with 45×45 and (i)–(l) 65×65 patches have less noise but capture fewer details. Smaller patches were upsampled using bilinear interpolation to 65×65 to ensure that the same baseline CNN model is used.

model being exposed to different types and tones of ice and water regions. Although increasing the data size improves the smooth ice, bias, and banding effect issues, the amount of detail the model captures is also reduced. The model does not capture many details present in the ice edge that the previous models captured. This could be due to oversaturation of consolidated ice regions and the lack of marginal ice zones in enhancement set.

D. Proposed Curriculum Learning Training Method

Results from applying the training method discussed in Section III are shown in Fig. 7(q)–(t). When comparing with the U-net trained with enhancement set with standard training, it can be seen that the model using the proposed training method captures more details that were originally missed, while maintaining the benefits of enhancing the dataset. This method, based on curriculum learning, benefits the model because it is able to learn features of consolidated ice and open water before learning features in marginal ice zones. We note that predictions are not as good for the image dominated by ice eddies [Fig. 7(r), likely because of a lack of appropriate samples in the training dataset.

In Fig. 8, we overlay a marginal ice zone region from three models with the corresponding SAR image. We can see that the U-net trained with only basic train captures the ice edge relatively well, but it visually appears to underpredict the ice concentration. The U-net leveraging enhancement set does not have the bias issue, but it can be seen that the ice edge is no longer defined as well. From the last model, the U-net leveraging enhancement set and employing curriculum learning, it can be seen that curriculum learning solves the bias issue while preserving some of the details in the ice edge. Furthermore, when applying curriculum learning, fewer epochs are required and early epochs are shorter because there is less data to train. Therefore, from this experiment, we show that curriculum learning achieves better performance in less time.

V. Model Evaluation

In this section, we evaluate the performance of our proposed model. First, we perform tenfold cross validation using basic train with our proposed model and other models. Next, we evaluate the proposed model with unseen images from Hudson Strait evaluation and compare it with passive microwave data and image analysis charts. The motivation for comparing our model with multiple data sources stems from the significant bias between passive microwave estimates and image analysis charts that previous studies reported for this area [35]. Therefore,

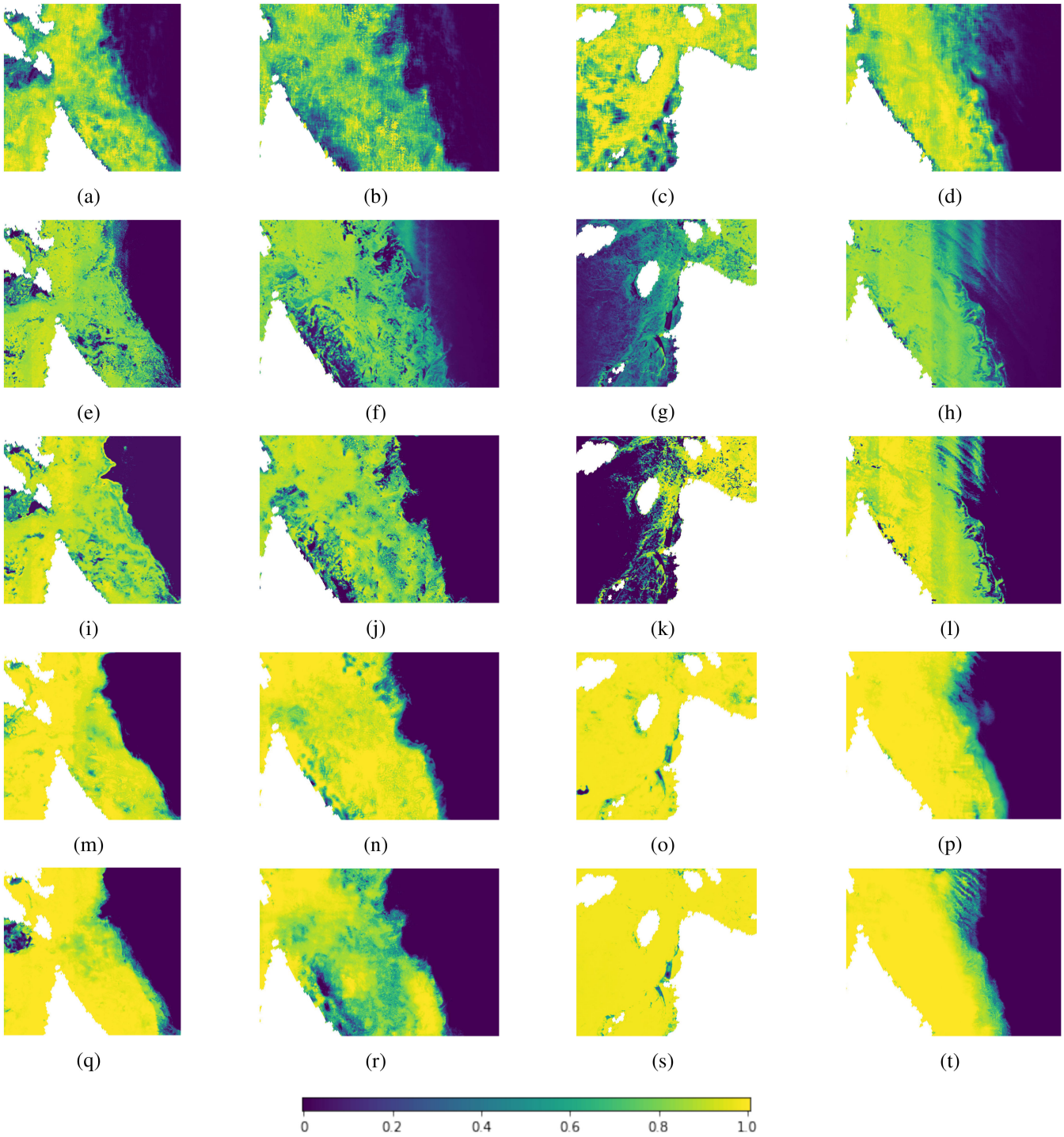

Fig. 7. Comparison of model predictions. (a)–(d) CNN trained with patch size of 45×45, (e)–(h) U-net trained with L2 loss function, (i)–(l) U-net trained with L1 loss function, (m)–(p) U-net trained with L1 loss function with an enhanced data set, (q)–(t) U-net trained with L1 loss function with an enhanced dataset using curriculum learning. Unlike CNNs, U-nets do not have patch size dependency. L1 loss function reduces the bias toward lower values. Enhancing the dataset reduces the bias further, but details from the SAR imagery are being missed. The addition of curriculum learning retains more details while preserving the bias reduction benefit from the enhanced set.

we are interested in understanding the agreement of our model with the different ice concentration estimation methods as well as comparing their agreement with one another. Histograms breaking down the samples used for evaluation are shown in Fig. 9.

A. *Tenfold Cross Validation*

A tenfold cross validation experiment was conducted with five models to quantitatively evaluate design choices. In this experiment, basic train was split into 10 sets randomly and 1 set was chosen to be a test set. The remaining sets were used

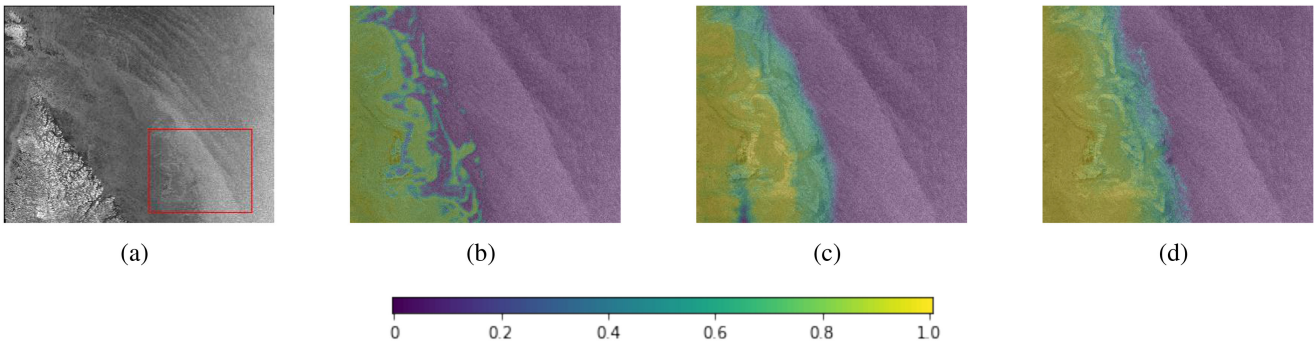

Fig. 8. Estimates for marginal ice zone region from 02/03/2018 overlaid on HH SAR image. (a) HH SAR image with marginal ice region of interest outlined by red box. Sea ice concentration estimates for region in red box from (b) U-net using L1 loss function, (c) U-net using L1 loss function and data augmentation, (d) U-net using L1 loss function, data augmentation, and curriculum learning.

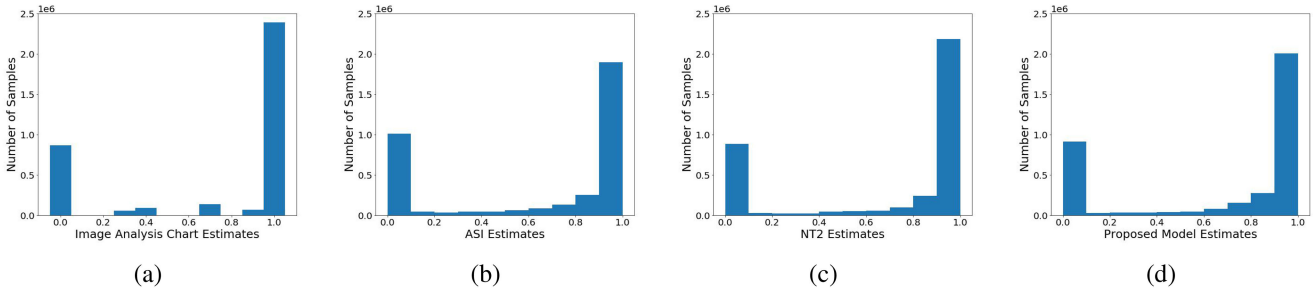

Fig. 9. Histogram of ice concentration estimates. (a) Image analysis charts. (b) ASI. (c) NT2. (d) Proposed model.

TABLE II
MEAN AND STANDARD DEVIATION OF MEAN ABSOLUTE ERROR WHEN PERFORMING TENFOLD CROSS VALIDATION USING BASIC TRAIN AND ASI ESTIMATES

Model	Train Mean (Std. Dev)	Test Mean (Std. Dev)
CNN-45	8.37% (0.731%)	10.10% (0.280%)
U-net (HS/L2)	13.62% (1.163%)	13.81% (0.996%)
U-net (HS/L1)	10.43% (0.171%)	10.82% (0.161%)
U-net (HS+E/L1)	5.04% (1.536%)	5.14% (1.589%)
U-net (HS+E/L1/Curr. Learning)	3.85% (0.397%)	3.87% (0.387%)

for training. This was repeated ten times for each model and the mean and standard deviation of the errors are reported in Table II. The average error of the baseline CNN is lower than that of the U-nets trained with basic train only. This is due to the poor performance of the U-nets in smooth ice conditions. As expected, the U-net trained with L1 loss function has a lower average mean absolute error than the U-net trained with L2 loss function and incorporating enhancement set further reduces the average error. Employing curriculum learning reduces the average error further and it also reduces the standard deviation of the errors. This suggests employing curriculum learning also results in more stable training.

B. Passive Microwave Evaluation

In this section, we compare our model estimates with estimates obtained from passive microwave data. Estimates from passive microwave data were grouped into bins of size 0.1. The proposed model predictions from every pixel from each bin group was extracted and the mean and standard deviation were calculated. The mean for each bin was plotted along with its standard deviation in Fig. 10. We do this comparison for our model using NT2 and ASI estimates. From this figure, it can be seen that there is a high bias from the proposed model when compared with ASI estimates. This is appropriate because we selected images for basic train that are of good visual agreement with the SAR images, whereas images for Hudson Strait evaluation were selected as images we wanted to improve on (visual bias between ASI and SAR). Therefore, this result, Figure 10(b), is expected. A positive correlation can also be seen between the ASI and proposed model estimates. This shows that the proposed model and ASI estimates have similar relative predictions for ice concentrations. The proposed model has better agreement with the NT2 estimates. The ASI estimates show a similar bias when compared with NT2. Note that although there is good agreement between the model and NT2, ice concentration predictions from NT2 lack small-scale details.

Upon qualitative observations, we can see that the proposed model predicts fine details that are missed by NT2 and ASI estimates. An example of the model successfully predicting

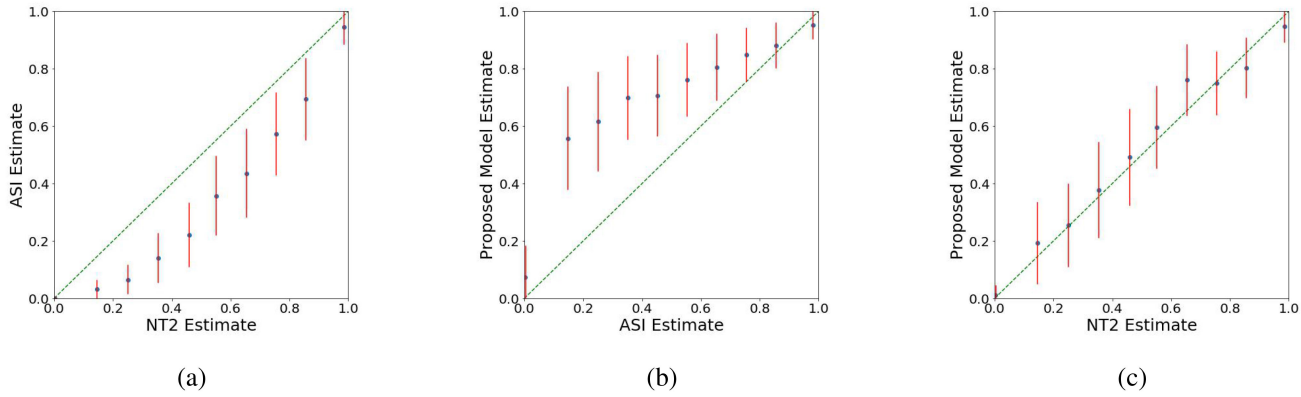

Fig. 10. Comparison of estimates between (a) NT2 and ASI, (b) proposed model and ASI, (c) proposed model and NT2. Proposed model is high biased when compared with ASI, which is expected given that the evaluation dataset corresponds to images for which the ASI ice concentration was visibly underestimated. NT2 estimates show a similar bias when compared with ASI estimates. Proposed model does not show significant bias with NT2 estimates. The blue dots represent the mean of the estimates on the *X*/*Y* axis. The red line represents the standard deviation of the values for the *Y* axis. The dashed green line is $y = x$ line.

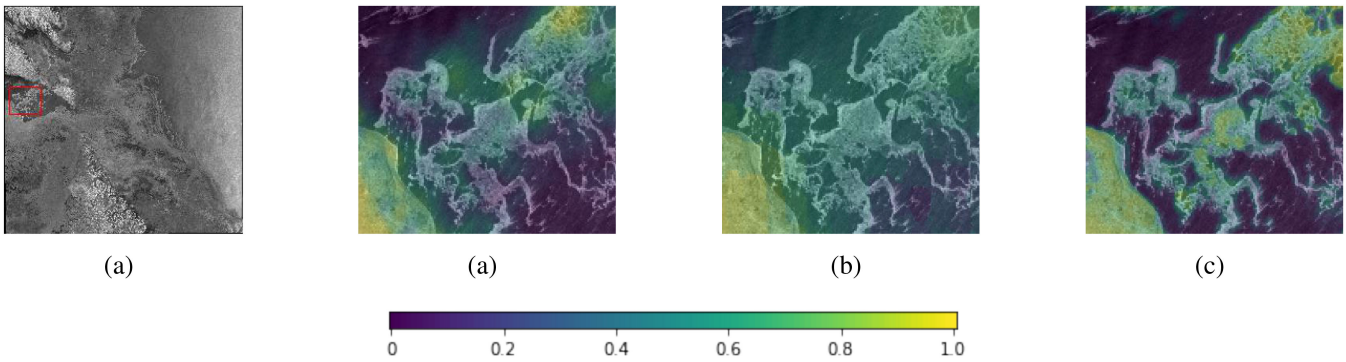

Fig. 11. (b) ASI, (c) NT2, and (d) proposed model ice concentration estimates overlaid on marginal ice zone region (outlined by red box) of (a) HH SAR image acquired 01/20/2018. Proposed model captures more precise details that are missed by ASI and NT2 estimates due to the coarse resolution of the passive microwave sensors.

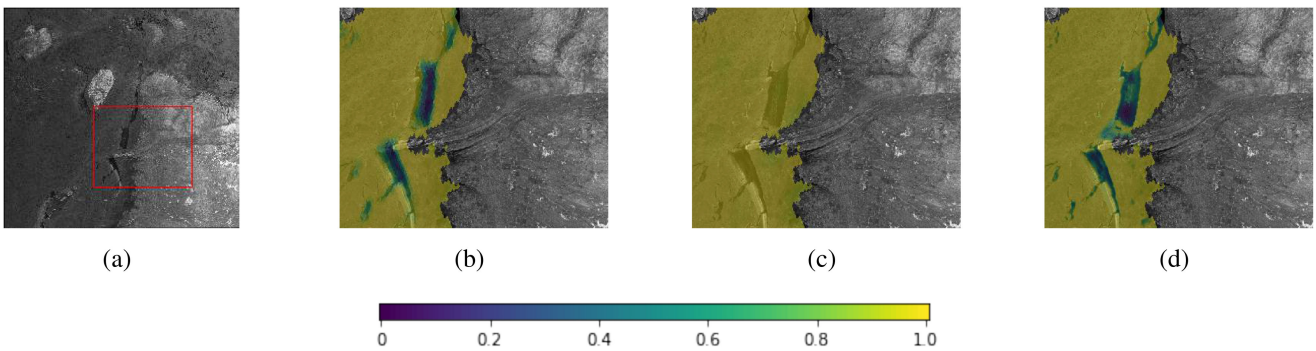

Fig. 12. (b) ASI, (c) NT2, and (d) proposed model ice concentration estimates overlaid on marginal ice zone region (outlined by red box) of HH SAR image (a) acquired 03/04/2018. Proposed model captures ice cracks more accurately than ASI estimates. NT2 estimates do not identify the ice cracks.

boundaries in the marginal ice zone that NT2 and ASI estimates are incapable of doing is shown in Fig. 11. The coarse resolution of passive microwave data makes NT2 and ASI estimates incapable of precisely identifying features in the ice edge and cracks in the ice cover. This is shown in Fig. 12. The success of the proposed model in these instances are due to predictions made on SAR images which have higher spatial resolution than passive microwave data. The proposed model also does not underpredict the ice near land to the same extent as the ASI estimates.

C. *Image Analysis Chart Evaluation*

Image analysis charts were also used to evaluate the model performance. The image analysis charts are not as precise as estimates retrieved from passive microwave data or the proposed model as large homogeneous regions (polygons) are given ice concentration values in 0.1 intervals. From Hudson Strait evaluation, 21 polygons were used in this study. They are summarized in Table III.

TABLE III
NUMBER OF POLYGONS FOR EACH ICE CONCENTRATION VALUE FROM IMAGE ANALYSIS CHARTS USED FOR EVALUATION

Ice Concentration Label	Number of Polygons
0.0	3
0.3	4
0.4	3
0.7	5
0.9	3
1.0	3

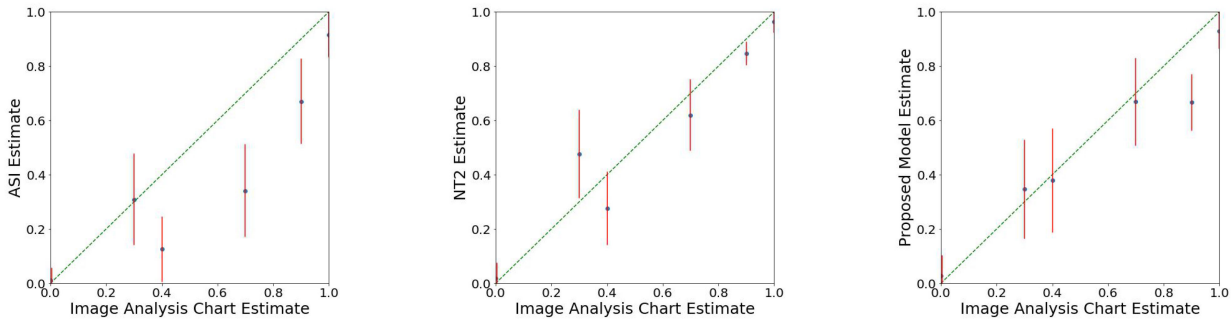

Fig. 13. Comparison of estimates between image analysis charts and (a) ASI, (b) NT2, and (c) proposed model. ASI estimates have a bias toward lower predictions when compared with image analysis charts. Proposed model and NT2 show better agreement with the image analysis chart estimates. The blue dots represent the mean of the estimates on the *X*/*Y* axis. The red line represents the standard deviation of the values for the *Y* axis. The dashed green line is $y = x$ line. Breakdown of polygons used for these plots are described in Table III.

TABLE IV
MEAN ABSOLUTE ERROR AND PEARSON CORRELATION COEFFICIENT BETWEEN ASI/NT2/PROPOSED MODEL AND IMAGE ANALYSIS CHART ESTIMATES OBTAINED FROM PLOTS FROM FIG. 13

	ASI	NT2	Proposed Model
Mean Absolute Error	16.10%	8.12%	7.18%
Pearson Correlation Coefficient	0.914	0.961	0.977

Note that the error for the proposed model is lower than that for NT2 and it is able to capture the small scale details better (see Figs. 11 and 12).

To compare the results, the mean and standard deviation of the estimates from the pixels of each image analysis chart label value was obtained. The plots comparing the passive microwave estimates and proposed model estimates are shown in Fig. 13. The greater underprediction seen from the ASI estimates on Fig. 13(a) further demonstrates that the ASI estimates have bias. The proposed model and the NT2 estimates show better agreement with the image analysis charts. We show statistics obtained from these plots in Table IV. From this table, it can be seen that the proposed model and NT2 estimates have a lower mean absolute error and higher Pearson correlation coefficient than the ASI estimates when compared with image analysis chart estimates. The proposed model also has slightly better agreement with image analysis charts than NT2.

VI. CONCLUSION

In this study, we show a U-net, which accepts images of any size, can be used in tandem with curriculum learning to obtain sea ice concentration estimates from HH and HV polarized SAR images that contain enhanced small-scale details in the marginal ice zone in comparison to the coarser data that are used for training. Hence, we can say the U-net is effectively able to upscale the passive microwave data using the image features learned from SAR. Through a tenfold cross validation with the train set, we select a U-net using an L1 cost function, data augmentation, and the curriculum learning training method as our proposed model. Our model has a mean absolute error of 3.87% with a standard deviation of 0.387% on the test data. We also found curriculum learning led to more stable and efficient training. The curriculum learning first trains the model with only consolidated ice and open water conditions before being exposed to marginal ice zones. This approach is advantageous in this study because consolidated ice and open water regions are more abundant and easier to identify. Allowing the model to only train with these regions early on gave it the opportunity to identify typical features of ice and water faster. Better performance was also observed in marginal ice zones which suggests the model is able to focus on these regions in later epochs.

In this study, we also show that using a U-net is advantageous to traditionally used CNNs as it eliminates shortcomings seen in previous models for ice concentration estimation. First, the U-net is significantly faster when making predictions as it makes predictions over 1000× faster in our experiments. Furthermore, unlike traditional CNNs, the U-net does not require a user defined patch size to make predictions.

When comparing predictions from our proposed model with that of algorithms using passive microwave data (ASI and NT2 algorithms), we found some advantages. The proposed model captures more details in the marginal ice zone that are missed by ASI and NT2 estimates. From our comparison with image analysis charts, our proposed model performs better than ASI and NT2 estimates as it achieves a mean absolute error of 7.18% with the image analysis charts, whereas ASI and NT2 estimates achieved a mean absolute error of 16.10% and 8.12%, respectively.

Lastly, aside from a U-net model, there are other deep learning models that have not been explored for sea ice concentration estimation, such as standard FCN models. A U-net was preferred in this study because it was observed in our experiments to capture finer details in the ice, such as ice cracks and leads, than standard FCNs. Furthermore, previous studies in segmentation on remote sensing images have found similar observations to ours that FCNs that employ gradual upsampling and skip connections preserve finer details than those that do not [36]. For this reason, other FCN architectures were not explored further in this study.

To develop an operational model, it should be trained with a wider variety of images. For example, the train set had a limited number of images with rough water conditions, large regions of smooth new ice and ice eddies. In the future, the model will be evaluated over a wider range of conditions, such as different seasons and locations, to obtain a more accurate measure of its performance. With a larger dataset, more complex models, such as a deeper U-net model, can also be explored. A deeper U-net model will be able to learn more spatial features. A curriculum learning method with more stages can also be implemented. This may help the model be more precise in the marginal ice zone ice concentration estimation. Lastly, it would be beneficial to have an operational model that can estimate ice concentration on full SAR images that have not been downsampled. Using full images would allow models to capture finer details. This task is more challenging as the computation requirements are greater.

Ultimately, in this study, not only do we show a new method to obtain ice concentration estimates with passive microwave data and SAR data, we also retain good estimates in consolidated ice and improve estimates in marginal ice zones.

Acknowledgment

The authors would like to thank ECCC for providing the SAR data and the image analysis charts, University of Bremen for providing the ASI data [22], and NSIDC for providing the NT2 data [23].

References

[1] C. T. Canada, G. Timco, B. Gorman, J. Falkingham, and B. O'Connell, "Scoping study: Ice information requirements for marine transportation of natural gas from the high Arctic," Nat. Res. Council Canada, Ottawa, ON, Canada, CHC-TR-029, 2005.

[2] A. S. Gagnon and W. A. Gough, "Trends in the dates of ice freeze-up and breakup over Hudson Bay, Canada," *Arctic*, vol. 58, no. 4, pp. 370–382, 2005.

[3] 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, pp. 1797–1817, 2015.

[4] L. Wang, K. Scott, and D. Clausi, "Sea ice concentration estimation during freeze-up from SAR imagery using a convolutional neural network," *Remote Sens.*, vol. 9, no. 5, 2017, Art. no. 408.

[5] C. L. Cooke and K. A. Scott, "Estimating sea ice concentration from SAR: Training convolutional neural networks with passive microwave data," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4735–47477, Jul. 2019.

[6] 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.

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

[8] R. Kruk, M. C. Fuller, A. S. Komarov, D. Isleifson, and I. Jeffrey, "Proof of concept for sea ice stage of development classification using deep learning," *Remote Sens.*, vol. 12, no. 15, 2020, Art. no. 2486.

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

[10] J. Karvonen, "A sea ice concentration estimation algorithm utilizing radiometer and SAR data," *The Cryosphere*, vol. 8, no. 5, pp. 1639–1650, 2014.

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

[12] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.

[13] R. Li *et al.*, "Deepunet: A deep fully convolutional network for pixel-level sea-land segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 3954–3962, Nov. 2018.

[14] W. Yao, Z. Zeng, C. Lian, and H. Tang, "Pixel-wise regression using U-net and its application on pansharpening," *Neurocomputing*, vol. 312, pp. 364–371, 2018.

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

[16] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 41–48.

[17] D. Weinshall, G. Cohen, and D. Amir, "Curriculum learning by transfer learning: Theory and experiments with deep networks," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 5238–5246.

[18] W. Lotter, G. Sorensen, and D. Cox, "A multi-scale CNN and curriculum learning strategy for mammogram classification," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2017, pp. 169–177.

[19] N. Sarafianos, T. Giannakopoulos, C. Nikou, and I. A. Kakadiaris, "Curriculum learning for multi-task classification of visual attributes," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2017, pp. 2608–2615.

[20] Y. Wang, W. Gan, J. Yang, W. Wu, and J. Yan, "Dynamic curriculum learning for imbalanced data classification," in *Proc. IEEE Int. Conf. Computer Vis.*, 2019, pp. 5017–5026.

[21] L. Brigham, "Marine protection in the arctic cannot wait," *Nature*, vol. 478, no. 7368, pp. 157–157, 2011.

[22] G. Spreen, L. Kaleschke, and G. Heygster, "Sea ice remote sensing using AMSR-E 89-GHz channels," *J. Geophysical Res.: Oceans*, vol. 113, no. C2, 2008, Art. no. C02S03.

[23] T. M. Meier, W. N. and J. C. Comiso, "AMSR-E/AMSR2 unified l3 daily 12.5 km brightness temperatures, sea ice concentration, motion & snow depth polar grids, version 1," [Hudson Strait, Labrador Sea, Gulf of St Lawerence, Beaufort Sea]. *NASA National Snow and Ice Data Center Distributed Active Archive Center*, Boulder, CO, USA, Aug. 2019.

[24] 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.

[25] D. J. Cavalieri, T. Markus, D. K. Hall, A. J. Gasiewski, M. Klein, and A. Ivanoff, "Assessment of EOS aqua AMSR-E arctic sea ice concentrations using Landsat-7 and airborne microwave imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 2057–3069, Nov. 2006.

[26] S. Andersen, R. Tonboe, S. Kern, and H. Schyberg, "Improved retrieval of sea ice total concentration from spaceborne passive microwave observations using numerical weather prediction model fields: An intercomparison of nine algorithms," *Remote Sens. Environ.*, vol. 104, no. 4, pp. 374–392, 2006.

[27] J. Lu, G. Heygster, and G. Spreen, "Atmospheric correction of sea ice concentration retrieval for 89 GHz AMSR-E observations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1442–1457, May 2018.

[28] M.-A. Moen *et al.*, "Comparison of feature based segmentation of full polarimetric SAR satellite sea ice images with manually drawn ice charts," *Cryosphere*, vol. 7, pp. 1693–1705, 2013.
[29] K. Radhakrishnan, "Sea ice concentration estimation: Using passive microwave and SAR data with fully convolutional networks," Master's thesis, University of Waterloo, 2020. [Online]. Available: http://hdl.handle.net/10012/16213
[30] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Sys. 32*, 2019, pp. 8024–8035.
[31] S. Maneewongvatana and D. M. Mount, "Analysis of approximate nearest neighbor searching with clustered point sets," in *Data Structures, Near Neighbor Searches, and Methodology: Fifth and Sixth DIMACS Implementation Challenges*, Providence, RI, USA: Amer. Math. Soc., vol. 59, pp. 105–123, 2002.
[32] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statistics, ser. Proc. Mach. Learn. Res.*, G. Gordon, D. Dunson, and M. Dudík, Eds., vol. 15, Apr. 2011, pp. 315–323.
[33] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 2528–2535.
[34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *ICLR*, 2015.
[35] K. A. Scott, M. Buehner, A. Caya, and T. Carrieres, "Direct assimilation of AMSR-E brightness temperatures for estimating sea ice concentration," *Monthly Weather Rev.*, vol. 140, no. 3, pp. 997–1013, 2012.
[36] S. Piramanayagam, E. Saber, W. Schwartzkopf, and F. W. Koehler, "Supervised classification of multisensor remotely sensed images using a deep learning framework," *Remote Sens.*, vol. 10, no. 9, 2018, Art. no.. 1429.

Keerthijan Radhakrishnan (Student Member, IEEE) received the B.A.Sc. degree in mechanical engineering and the M.A.Sc. degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 2018 and 2020, respectively.

His research interests include applying machine learning, computer vision, and image processing toward remote sensing applications.

K. Andrea Scott (Member, IEEE) received the B.A.Sc. from the University of Waterloo, Waterloo, ON, Canada, in 1999, the M.A.Sc. degree from McMaster University, Hamilton, ON, Canada, in 2001, and the Ph.D. degree from the University of Waterloo, Waterloo, ON, Canada, in 2008, all in mechanical engineering.

She was a Postdoctoral Researcher with the Data Assimilation and Satellite Meteorology Research Section, Environment and Climate Change Canada, Toronto, ON, Canada, in 2008, where she was part of a team involved in the development of a sea ice data assimilation system. In 2012, she joined the Department of Systems Design Engineering, University of Waterloo, as a Faculty Member with a specialization in sea ice remote sensing and data assimilation.

David A. Clausi (Senior Member, IEEE) received the Ph.D. degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 1996.

He worked in medical imaging with Mitra Imaging, Waterloo, ON, Canada, in 1996. He started his academic career in 1997 as an Assistant Professor in Geomatics Engineering with the University of Calgary, Calgary, AB, Canada. In 1999, he returned to the University of Waterloo and is currently a Professor specializing in the field of Intelligent Systems and has been an Associate Dean, Research & External Partnerships with the Faculty of Engineering, since 2018. He is an active interdisciplinary academic who has an extensive research record, and has authored or coauthored refereed journal and conference papers. His research interests include remote sensing, computer vision, image processing, and algorithm design.

Prof. Clausi was the recipient of numerous scholarships, paper awards, research awards, teaching excellence awards and his efforts have led to successful commercial implementations, culminating in the creation and sale of a high-tech company. He was the Co-chair of IAPR Technical Committee 7–Remote Sensing from 2004 to 2006.