# Sea Ice Concentration Estimation During Melt From Dual-Pol SAR Scenes Using Deep Convolutional Neural Networks: A Case Study

Lei Wang, *Graduate Student Member, IEEE*, K. Andrea Scott, *Member, IEEE*, Linlin Xu, *Member, IEEE*, and David A. Clausi, *Senior Member, IEEE*

*Abstract*—**High-resolution ice concentration maps are of great interest for ship navigation and ice hazard forecasting. In this case study, a convolutional neural network (CNN) has been used to estimate ice concentration using synthetic aperture radar (SAR) scenes captured during the melt season. These dual-pol RADARSAT-2 satellite images are used as input, and the ice concentration is the direct output from the CNN. With no feature extraction or segmentation postprocessing, the absolute mean errors of the generated ice concentration maps are less than 10% on average when compared with manual interpretation of the ice state by ice experts. The CNN is demonstrated to produce ice concentration maps with more detail than produced operationally. Reasonable ice concentration estimations are made in melt regions and in regions of low ice concentration.**

*Index Terms*—**Convolutional neural network (CNN), ice concentration, synthetic aperture radar (SAR).**

## I. INTRODUCTION

WITH recent reductions in Arctic ice extent, there has been growing economic interest in shipping and natural resource extraction in the Arctic [1]. To support safe Arctic operations and navigation in ice-infested waters, timely high-resolution information of the ice coverage is crucial [2]. In this paper, we focus on ice concentration, which is defined as the percentage of ice coverage over a given spatial area.

Satellite sensors are used as the main data source for ice monitoring in the Arctic region. Due to the frequent cloud coverage and long periods with little sunlight, optical remote sensing technology cannot generate continuous observations in the Arctic. Instead, passive microwave and synthetic aperture radar (SAR) sensors are commonly used [3], both of which can generate observations regardless of the weather conditions (cloud, rain) and sunlight. Passive microwave sensors detect radiation emitted by the target. A large footprint is normally required to have sufficient energy for a measurement. This constrains the spatial resolution of the passive microwave data, which is normally above 4 km. In contrast, SAR transmits radiation to the Earth and measures the portion of energy backscattered to the sensor. The spatial resolution of SAR imagery used for ice mapping is normally less than 100 m, which is much finer than the resolution of passive microwave imagery.

Due to the lack of robust automatic SAR image interpretation algorithms, SAR images are still manually interpreted for ice mapping by national ice agencies such as the Canadian Ice Service (CIS) [4]. Manual interpretation of SAR images is time consuming, generating low-resolution ice maps, and subjective [5]. Moreover, several new SAR satellite missions are under development, such as Canada's RADARSAT Constellation mission and the European Space Agency Sentinel-1B mission. With the rapidly growing volume of data, automatic ice concentration estimation, and even data mining from the large quantity of SAR images, is strongly desired.

The automatic interpretation of the high-resolution SAR sea ice images is a difficult problem due to the complex interaction between the ice or water surface and the SAR signal. The magnitude of the reflected SAR signal is largely determined by the radar frequency, the radar incidence angle, and the target conditions such as surface roughness, moisture, snow cover, and salinity [6]. For example, calm water generally has low backscatter compared with wind-roughened water, most noticeably in horizontal transmit and horizontal receive (HH) pol and vertical transmit and vertical receive (VV) pol SAR images compared with horizontal transmit and vertical receive (HV) pol [6]. This is not always the case due the effect of the incidence angle. The sensitivity of scattered SAR signals to wind speed varies with incidence angle. The appearance of ice in SAR images also changes through the ice season due to melting, ice dynamics, and floe distribution. These factors make it challenging to design image features to capture the diverse patterns of ice and water in SAR imagery.

However, there has been a constant effort toward automatic ice concentration estimation from SAR imagery [7]–[13]. A two-step approach is normally used [7]–[13]: SAR image features are first designed and calculated to describe the differences between ice and water; then, a regression model is built upon those features to estimate the ice concentration. Empirical linear regression and multiple-layer perceptron are examples of such regression models [9], [10], [12], [13].

Both single-polarized [8], [10], [14] and dual-polarized (HH and HV) [11]–[13] SAR images have been used to extract ice and water features. HH and HV pol complement each other. HH-pol images contain more details but are strongly affected by incidence angle. HV-pol images contain different information of the ice and are visually less affected by incidence angle. The

use of dual-polarized SAR images has been shown to benefit ice information retrieval [7], [12], [13], [15], particularly in water regions and melt conditions [12].

A number of features have been proposed to estimate ice concentration from SAR images, such as autocorrelation and gray-level cooccurrence matrix texture features [10], [11], in addition to local normalized intensity and intensity difference of different polarizations [10], [12]. These features were reported to be affected by varying wind conditions [10]. In a recent study [11], a complex model between SAR image features and ice concentration that takes into account wind speed, snow depth, and surface air temperature was developed. No single feature was found to correlate well with the ice concentration, whereas using many features together led to improved correlation with ice concentration [11]. This study demonstrated that the image features chosen *a priori* are not able to capture ice concentration from SAR imagery well. In [9], the densities of edges and corners are used as features to detect ice in SAR images. These carefully designed features were found to improve ice concentration estimation. However, there was still a degradation of performance in the melt season, as well as overestimation of ice concentration in open water regions, as corner and edge features are unstable to melt and wind-roughened water. To get more robust and representative features, backscatter calibration [12], [13] and incidence angle correction [16] are often applied to SAR images before feature extraction. The effect of such preprocessing is conditioned on the SAR imagery and surface conditions. Segmentation is commonly used to improve the numerical quality and visual appearance of the estimated ice concentration for most of the aforementioned methods. The drawback is that the quality of the results is affected by the quality of the segmentation, which leads to a less robust algorithm.

These observations indicate that higher level image features (such as compositions of basic image structures such as edges and corners) need to be used in order to capture ice appearances in SAR imageries under variable and changing surface conditions. The design of such features is difficult and nonintuitive. Redesign of the algorithm or features may be necessary for different regions or times of the year. Therefore, it is of interest to investigate the use of feature learning instead of feature design for ice concentration estimation from SAR images. If features can be learned from training data, the most effective features can be generated for specific data and conditions without human intervention. This is highly desirable for large-scale automatic ice concentration estimation tasks.

In this paper, the estimation of ice concentration under the framework of deep learning is investigated. Deep learning is a feature learning method that uses multiple layers of neural networks [17], which has recently demonstrated great potential in many different recognition tasks such as speech recognition and image object classification [18]–[21]. A deep convolutional neural network (CNN) has been used to estimate ice concentration directly from dual-band SAR images (HH and HV) in the melt season. CNN is a neural network model that enforces weight sharing and local connections between adjacent layers of neurons. This method has a demonstrated ability to achieve high performance for image-related recognition tasks [20], [22]–[24]. In this paper, its capability to generate high-quality ice concentration estimation in the melt season is demonstrated
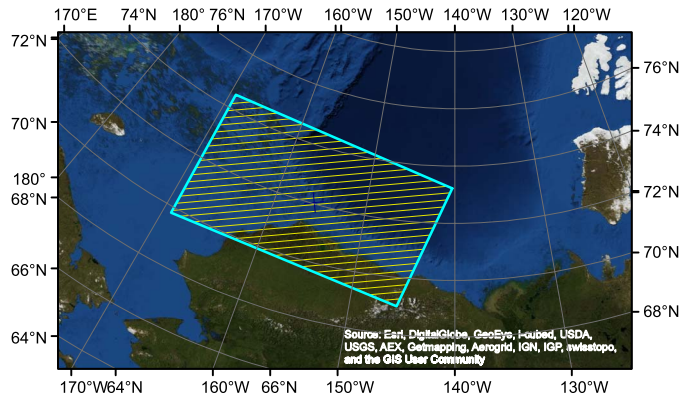


Fig. 1. Study area in the Beaufort Sea used in this study.

by a case study in the Beaufort Sea, under a variety of surface conditions, without incidence angle correction or postprocessing. A state-of-the-art ice concentration estimate from SAR imageries is achieved. To the best of our knowledge, this is the first study in which a CNN is used to extract ice concentration from SAR images.

The outline of this paper is as follows. Section II describes the study area and data used. CNN is introduced in Section III. The method for ice concentration estimation using CNN is described in Section IV. Results and analysis are described in Sections V and VI, followed by conclusions in Section VII.

## II. DATA SET USED IN THIS STUDY

The study area is located in the Beaufort Sea, north of Alaska. This region is a particularly relevant area for high-resolution ice concentration mapping research due to oil and gas exploration in this region. The period of study is July to September. Ice at this time of year in the Beaufort Sea is mainly a mixture of multiyear ice, first-year ice, and some new ice. Based on CIS ice charts, the coverage of first-year ice is between 0% and 50%, with under 30% of coverage for most images. The coverage of multiyear ice varies from around 30% to 90%. Packed multiyear ice often breaks into smaller ice floes, leaving large areas of open water between the floes, which may refreeze to form new ice. Melting ice surfaces can be observed in this region until October [25].

The data set used in this study is a set of 11 RADARSAT-2 dual-polarized (HH and HV) ScanSAR wide-beam images of the Beaufort Sea (see Fig. 1) provided by MDA. These images are acquired between July and September in 2010 and 2011. Each image covers a 500 km × 500 km area with a 50 m × 50 m nominal pixel spacing. The incidence angle ranges from 20° to 49°. Exact dates of the images used are given in Table I.

Image analyses were used to train and verify the neural network in this study. The image analyses are ice maps produced by ice experts from the CIS through visual interpretation of SAR images [4]. An image analysis is composed of contiguous polygons that are generated by the ice analyst. Each polygon is assigned an ice concentration value and fractions of different ice types manually. The ice concentration values are quantized by intervals of 0.1 (e.g., ice concentration only has 11 possible values: 0, 0.1, 0.2, ..., 0.9, 1). The precision of the image analysis ice concentration is normally considered to be approximately 10% [4]. Examples of HH- and HV-pol SAR images

TABLE I
ICE CONCENTRATION ESTIMATION RESULTS VERIFIED AGAINST IMAGE ANALYSES FOR THE LEAVE-ONE-OUT EXPERIMENT. THE FIRST 11 ROWS OF THE DATA CORRESPOND TO THE 11 ROUNDS OF THE LEAVE-ONE-OUT EXPERIMENT. THE ROW "AVERAGE" IS THE MEAN ERRORS OF THE LEAVE-ONE-OUT EXPERIMENT WEIGHTED BY THEIR CORRESPONDING NUMBER OF SAMPLES USED. THE LAST ROW "SI" IS THE ERRORS OF THE AMSR-E ICE CONCENTRATION DATA WHEN VERIFIED AGAINST IMAGE ANALYSIS. "NUM." IS THE NUMBER OF SAMPLES USED IN EACH STAGE. "SCENE" REPRESENTS THE SCENE NUMBER LABELED IN THE FORMAT OF YYMMDD

| Training | | | | Validation | | | | | Testing | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Num.$ | $E_{sgn}$ | $E_{L1}$ | $E_{std}$ | $Scene$ | $Num.$ | $E_{sgn}$ | $E_{L1}$ | $E_{std}$ | $Scene$ | $Num.$ | $E_{sgn}$ | $E_{L1}$ | $E_{std}$ |
| 47905 | 0.03 | 0.06 | 0.10 | 100806 | 5759 | -0.02 | 0.12 | 0.18 | 100730 | 613 | -0.11 | 0.12 | 0.13 |
| 41150 | -0.01 | 0.05 | 0.10 | 110817 | 7368 | 0.01 | 0.06 | 0.11 | 100806 | 5759 | -0.08 | 0.13 | 0.18 |
| 37198 | -0.03 | 0.07 | 0.12 | 110817 | 7368 | 0.01 | 0.06 | 0.11 | 100822 | 9711 | -0.02 | 0.06 | 0.11 |
| 44984 | -0.03 | 0.07 | 0.11 | 110817 | 7368 | 0.01 | 0.06 | 0.11 | 100829 | 1925 | -0.07 | 0.09 | 0.12 |
| 43175 | -0.01 | 0.07 | 0.11 | 110817 | 7368 | 0.01 | 0.06 | 0.11 | 100909 | 3734 | -0.03 | 0.05 | 0.09 |
| 48096 | 0.03 | 0.06 | 0.10 | 100806 | 5759 | -0.02 | 0.13 | 0.18 | 110709 | 422 | -0.03 | 0.12 | 0.15 |
| 42260 | -0.02 | 0.06 | 0.11 | 110811 | 8974 | 0.02 | 0.06 | 0.11 | 110710 | 3043 | -0.03 | 0.11 | 0.15 |
| 43521 | -0.03 | 0.06 | 0.11 | 110811 | 8974 | 0.02 | 0.06 | 0.11 | 110720 | 1782 | -0.09 | 0.13 | 0.19 |
| 37572 | 0.02 | 0.06 | 0.10 | 100806 | 5759 | -0.01 | 0.12 | 0.19 | 110725 | 10946 | 0.02 | 0.08 | 0.12 |
| 34357 | -0.01 | 0.06 | 0.11 | 110725 | 10946 | 0.02 | 0.07 | 0.10 | 110811 | 8974 | 0.04 | 0.07 | 0.11 |
| 37935 | -0.03 | 0.07 | 0.12 | 110811 | 8974 | 0.04 | 0.07 | 0.11 | 110817 | 7368 | 0.02 | 0.06 | 0.11 |
| Average | -0.01 | 0.06 | 0.11 | | 7692 | 0.01 | 0.08 | 0.13 | | 4934 | **-0.01** | **0.08** | **0.13** |
| ASI | | | | | | | | | | | -0.12 | 0.12 | 0.18 |
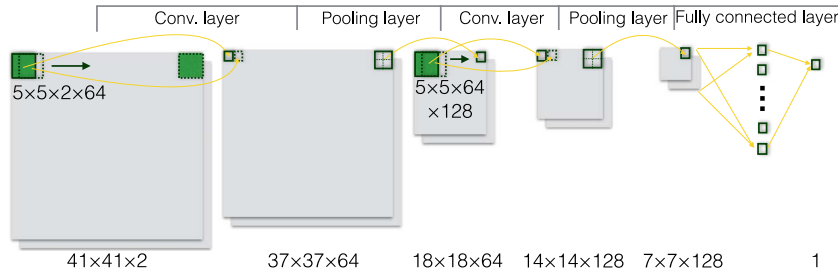


Fig. 2. Structure of the CNN model used in this study. The outer rectangles are inputs and outputs of different layers. The size of the input of each layer (width, height, and number of bands) is shown at the bottom. The smaller filled rectangles are convolution filters the sizes of which are described beneath the filters. For example, the first convolution layer contains 64 filters of size $5 \times 5 \times 2$. The input of this layer is an image patch with two bands, and each band is $41 \times 41$ pixels.

and image analyses are shown in Figs. 6 and 7(a), (d), and (g). While image analyses may contain errors, they are used as the ground truth in this study because image analysis is the most reliable ice concentration data available for this study. Without confusion, we also refer to the image analysis as ground truth.

The results of this study are also compared with the AMSR-E daily ice concentration maps acquired from http://www.iup.uni-bremen.de/seaice/amsr/. This ice concentration data set features a pixel spacing of 6.25 km. It was generated from the AMSR-E 89-GHz channel data using the ARTIST sea ice (ASI) algorithm [26]. This algorithm is known to underestimate ice concentration for thin ice or when ice concentration is low [26]. This data set is referred to as ASI ice concentration in the rest of this paper.

## III. CNN

A CNN is an artificial neural network introduced by LeCun *et al.* [27]. A CNN consists of a stack of alternating convolution layers and pooling layers with a number of fully connected layers on top. Fig. 2 shows the structure of a CNN with two pairs of convolution layers and pooling layers with one fully connected layer on top. The filter parameters in the convolution layers and the weight matrix in the fully connected layers are the parameters of the CNN, which are learned through training.

### A. Convolution Layer

A convolution layer is mainly characterized by its number of convolution filters and the size of these filters. Each filter moves within the image (the filter has to be within the image completely) to generate a convolved image. The output of a convolution layer $h_{ij}$, which is also called "feature map," is a pixelwise transformation of the convolved images by an activation function. With multiple filters, multiple feature maps are generated and stacked to form a multilayer feature map. The convolution process for each filter is described by

$$\boldsymbol{h}_{ij} = \boldsymbol{\sigma}\left((\boldsymbol{W} * \boldsymbol{X})_{ij} + b\right), \text{ in which} \quad (1a)$$
$$i = 1, \dots, M_x, \quad j = 1, \dots, M_y \quad (1b)$$
$$M_x = \frac{S_x - K_x}{P} + 1 \quad (1c)$$
$$M_y = \frac{S_y - K_y}{P} + 1. \quad (1d)$$

In (1), $X$ is the input image of width and height denoted by $S_x$ and $S_y$. A filter $W$ of width $K_x$ and $K_y$ moves along the input space to convolve with the input image $X$, and the operation of convolution is denoted by $*$. The bias term for this filter is denoted by $b$; the activation function is denoted by $\sigma$. The generated feature map $h$ has width and height $M_x$ and $M_y$ and depends on the step size of the filters, which is denoted by $P$. When $P$ is 1, every possible location in the input image is filtered. When $P$ is 2, every other pixel of the input image is filtered.

### B. Max Pooling Layer

A max pooling layer subsamples the feature map generated by the previous convolution layer by mapping each nonoverlapped rectangle image block to the maximum value of this block [28], [29]. If the input image of a pooling layer has multiple bands, max pooling will be performed on each band, and the results will be stacked. Pooling reduces information redundancy and resolution of the input to the pooling layer. This redundancy arises because the convolution windows of neighboring locations are highly overlapped. Upper convolution layers always have larger receptive fields than lower convolution layers, which means that upper convolution layers model larger scale information in the image. This enables the CNN to model spatial context information at multiple scales.

### C. Fully Connected Layer

There may be several pairs of convolution and pooling layers in a CNN. The feature maps generated by the last pooling layer are lexicographically expanded to form a vector. One or more fully connected layers are built on this vector. Each fully connected layer maps the input vector to an output vector by (1a).

## IV. Ice Concentration Estimation Method

In this paper, a supervised CNN model is used as a feature extractor and a regression model to estimate ice concentration from SAR images. The processing scheme is shown in Fig. 3. It is composed of three major steps: preprocessing of the SAR images, training of the CNN model, and prediction using the trained CNN model. Performance of the trained CNN model is evaluated by testing on independent images to the images used in training. Prediction follows the same procedure as testing, except that an ice concentration is estimated for every pixel location in the input SAR image; therefore, only the first two steps will be described.

### A. Preprocessing of SAR Images

All the SAR images are subsampled by averaging nonoverlapping $8 \times 8$ blocks of the image to reduce image noise and data volume. The subsampled images have 400-m pixel spacing with pixel values between 0 and 255. Standard CNNs, which use sigmoid activation functions, take input values normalized to $[-1, 1]$. In this paper, rectified linear units (ReLUs) (2a) are used as the activation function for all layers, except the output layer. When an ReLU activation function is used, input values below 0 will not generate activation; therefore, the gradient of
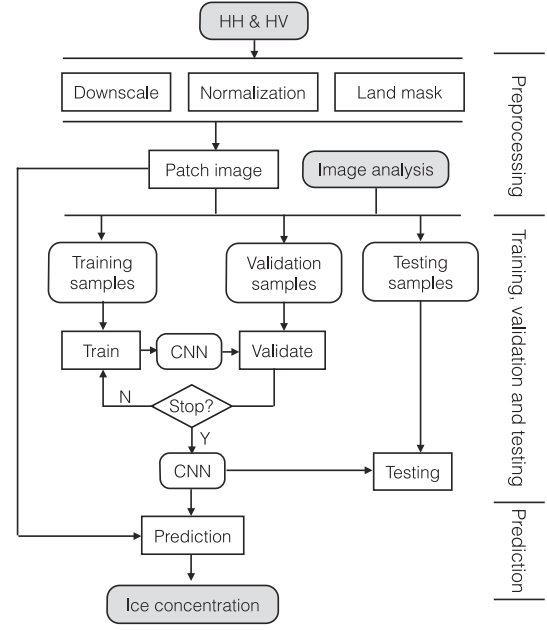


Fig. 3. Flowchart of the major steps of the algorithm.

units with negative input will be 0. No learning is happening when trained using stochastic gradient descent (SGD) at those units. For this reason, instead of normalizing the input to $[-1, 1]$, pixel values of the dual-polarized SAR images are normalized to [0, 1] by dividing each pixel value by 255. A land mask is applied to remove the pixels located on land.

The image analyses obtained from CIS in this study are sampled points of the original polygon-based image analyses. The sampling interval is about 8 km in the north–south direction by 5 km in the east–west direction. Note that, since each polygon of the image analysis is labeled with one ice concentration value, the actual ice concentration at the sampled locations may be different from the ice concentration of the polygon. This introduces representation errors when using the sample points from the image analysis for training. The image analyses only cover a small portion of the whole image for some scenes [see Fig. 7(a)]. The number of image analysis sample points for each SAR image is given in Table I.

Image patches of $41 \times 41$ pixels (16.4 km $\times$ 16.4 km) centered at the image analysis sample locations are extracted from the normalized SAR images. If an image patch overlaps with land or lies outside the SAR image, its corresponding image analysis sample point is not used. Because both HH and HV pol are considered, each image patch is a 3-D matrix of size $41 \times 41 \times 2$. Each SAR image patch and the ice concentration located at the patch center from the image analysis are one sample used in this experiment. The patch size is tuned by evaluating the performance of the algorithm using patch sizes from 28 to 60, and the patch size with the minimal testing error is selected.

### B. Structure of CNN

The structure of the CNN used in this study is illustrated in Fig. 2. This CNN contains two convolution layers and one fully connected layer on top. The first convolution layer has 64 filters of width and height set to 5. Each filter is a 3-D array (height, height, bands). Inputs to this filter are image patches of

size $41 \times 41 \times 2$; thus, the size of each filter is $(5, 5, 2)$. The filter outputs convolved images of size $(41-5+1, 41-5+1, 64)$, that is, $(37, 37, 64)$ based on (1). The first convolution layer is followed by a max pooling layer of pooling size $2 \times 2$ and stride 2 [20], [23]. The max pooling layer outputs the maximum activation of every unoverlapping $2 \times 2$ block of the convolved image to form a feature map of size $(37/2, 37/2, 64)$, which is $(18, 18, 64)$ after the last column and row of the feature map are removed. The second convolution layer contains 128 filters of size $(5, 5, 64)$. It outputs convolved image of size $(18-5+1, 18-5+1, 128)$. The following pooling layer is the same as the first pooling layer. Its output feature map is $(7, 7, 128)$. The fully connected layer takes input of dimension $7 \times 7 \times 128 = 6272$ and outputs only one single value, which is the ice concentration estimate.

ReLUs (2a) were chosen for both convolution layers because they have been demonstrated to lead to faster learning and better features compared with sigmoid activation units due to the fact that they do not saturate as sigmoid activation units [20], [30]. The top linear layer uses a linear activation function (2b) with one output unit, which is the ice concentration. By modifying the activation function of the top layer, the CNN model is casted from a classification model to a regression model. Equations (2a) and (2b) are as follows:

$$f(x) = \max(0, x) \tag{2a}$$

$$f(x) = x. \tag{2b}$$

### C. Training and Testing

Backpropagation and minibatch stochastic gradient descendant (SGD) [31] are used for training. The training samples are randomly separated to $N/128$ minibatches, where $N$ is the total number of training samples. Each minibatch contains 128 training samples. In our experiment, the mean square error over the minibatch is used as the cost function for training, i.e.,

$$L\left(F(\boldsymbol{X}; \boldsymbol{W}), \boldsymbol{y}\right) = \frac{1}{128} \sum_{i=1}^{128} \left(F(\boldsymbol{X}_i; \boldsymbol{W}) - \boldsymbol{y}_i\right)^2. \tag{3}$$

In (3), the forward propagation of the CNN is noted by $F$. The $i$th image patch and corresponding image analysis of the minibatch are denoted by $\boldsymbol{X}_i$ and $\boldsymbol{y}_i$. The CNN takes $\boldsymbol{X}$ as input and generates an estimate of ice concentration based on its current weights $\boldsymbol{W}$. The parameters of the CNN are updated according to the derivative of the parameters to the cost function over one minibatch, which is described by

$$\boldsymbol{V}_{i+1} = m \cdot \boldsymbol{V}_i - r \cdot \epsilon \cdot \boldsymbol{W}_i - \epsilon \frac{\partial \boldsymbol{L}}{\partial \boldsymbol{W}} |_{\boldsymbol{w}_i} \tag{4a}$$

$$\boldsymbol{W}_{i+1} = \boldsymbol{W}_i + \boldsymbol{V}_{i+1}. \tag{4b}$$

The gradients of the cost function with respect to the weights $(\partial \boldsymbol{L}/\partial \boldsymbol{W})$ are calculated and averaged over the minibatch. The weights $\boldsymbol{W}$ are updated by $V_{i+1}$ at iteration $i + 1$ with learning rate $\epsilon = 10^{-3}$ and a weight decay of $r = 2 \times 10^{-5}$ with momentum $m = 0.9$. The setting of the training parameters for SGD is similar to the setting in [20]. Some adjustments are made by tuning. Similar to [20], the parameters of the CNN are initialized by uniform random sampling between $-0.05$ and $0.05$.

An epoch training scheme [31] is adopted. For each epoch, all the minibatches of the training samples are iterated once by the training algorithm. To accelerate the training process, the training is set to stop after 500 epochs or when the score of cost function is changing less than 0.001 for 20 consecutive epochs, in case the training converges early (which is typical [32]).

Usually, there is only a training and testing data set for image classification tasks. Due to the fact that CNN tends to overfit due to its large parameter space, it is a common practice to use a validation data set when training a CNN model to relieve the overfitting problem [20]. The derived CNN model is validated after each training epoch by calculating the cost function on the validation data using the current model. The CNN with the smallest validation error will be selected as the trained CNN. Note that validation is used for model selection, and it is therefore part of the training scheme.

The 11 scenes (see Table I) are divided to training and testing sets. A leave-one-out experiment is conducted. One scene is used for testing, another scene is used for validation, and the rest are used for training for each round of the experiment. There are 11 rounds in the leave-one-out experiment. Validation is essential for the successful training of the CNN model. Thus, the validation image needs to be selected carefully. The four scenes with large image analysis spatial coverage ratio and complete ice concentration levels in their image analysis are selected as validation candidate images. One of the candidate image is randomly chosen as the validation image for each round of the leave-one-out experiment.

## V. Ice Concentration Estimation Results

The ice concentration estimates are evaluated against image analyses in the SAR image space. In other words, each image analysis sample point is compared with the ice concentration of its nearest SAR image pixel. The mean error $E_{\text{sgn}}$, the mean absolute error $E_{L1}$, and the error standard deviation $E_{\text{std}}$ are calculated for evaluation purposes using

$$E_{\text{sgn}} = \text{mean}(\text{IC}_{\text{CNN}} - \text{IA}) \tag{5a}$$

$$E_{L1} = \text{mean}\left(|\text{IC}_{\text{CNN}} - \text{IA}|\right) \tag{5b}$$

$$E_{\text{std}} = \text{std}(\text{IC}_{\text{CNN}} - \text{IA}) \tag{5c}$$

where $\text{IC}_{\text{CNN}}$ denotes the estimated ice concentration, and IA denotes the image analysis. The evaluation results for all training, validation, and testing for all rounds of the leave-one-out experiment are shown in Table I.

Based on Table I, the mean error $E_{\text{sgn}}$ is 0.01 on average. The overall bias of the estimated ice concentration is 0.08. The largest mean error $(E_{\text{sgn}})$ and the mean absolute error $(E_{L1})$ for the testing image are $-0.11$ and 0.13, respectively. The error standard deviation $(E_{\text{std}})$ for testing is at the same level as for training and validation. The performance of testing is not showing a large decrease to the performance of training and validation, which indicates a low level of overfitting of the trained CNN model. This suggests that the method can generalize to unseen data well. Compared with ASI ice concentration, the CNN result is showing a much better performance in all three statistics.
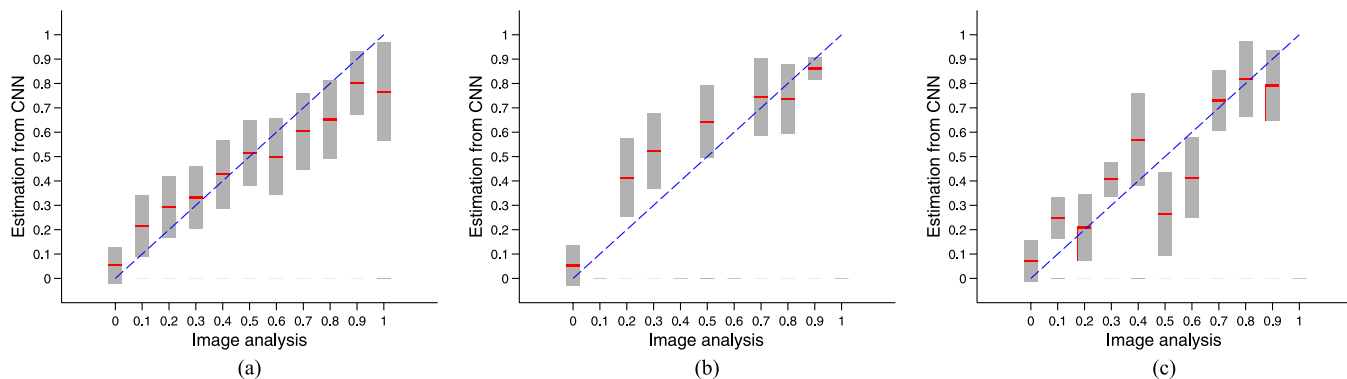
Fig. 4. Errors at different ice concentration levels for (a) training, (b) validation, and (c) testing when using scene 20110725 for testing, scene 20110811 for validation, and the remaining scenes for training.
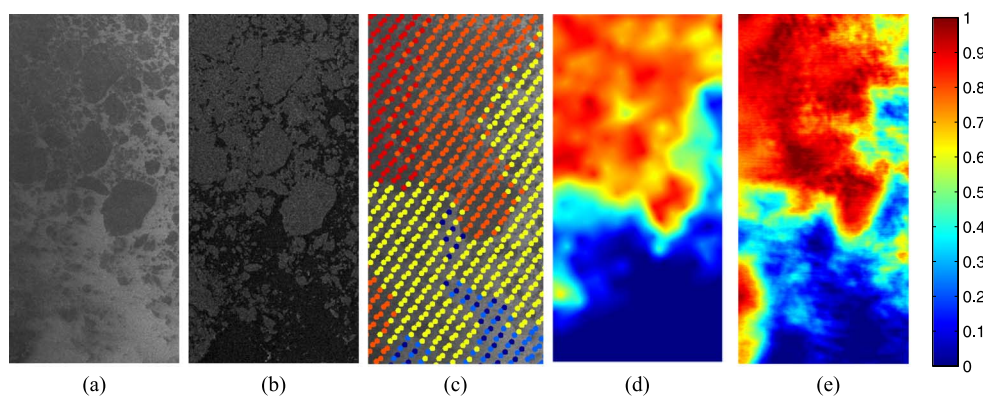


Fig. 5. Ice concentration estimation details shown for a subset of scene 20110725. (a) HH. (b) HV (enhanced for visualization). (c) Image analysis plotted over the HH band image using the same color scheme as (d). (d) AMSR-E ice concentration generated using the ASI algorithm. (e) Estimated ice concentration when using scene 20110725 for testing.

While the image analysis is quantized, the ice concentration estimation using CNN is continuous. This difference may introduce errors into the evaluation statistics. Therefore, the ice concentration estimates are also quantized to 11 levels between 0 and 1 and reevaluated against the image analyses. The evaluation results are similar but slightly improved after quantizing the ice concentration estimation and are therefore not shown here.

Fig. 4 shows the mean and standard deviations of ice concentration estimation errors for different ice concentration bins when using scene 110725 as testing. There is a clear trend between image analyses and ice concentration estimates from the CNN for all three sets in general. Because of the abundant water samples in the training data set, the estimation of pure water is better than that of the ice concentration for both testing and validation. It is reasonable to infer that the estimation precision can be improved by using more training samples of intermediate ice concentration levels. A strong underestimation of ice concentration at 0.5 and 0.6 for test data is observed. This may be caused by the representation error in the image analysis. For example, Fig. 5 shows an area that is labeled as having an ice concentration of 0.6 in the image analysis (see the area covered by yellow dots in Fig. 5). By inspecting the SAR images (see Fig. 6), it was found that this region is a mixture of ice floes and water. This is captured correctly by the CNN. The estimated ice concentration shows large water regions with

low ice concentration, as well as ice floes having larger ice concentration. The estimated ice concentration contains more details than the image analysis in this region. This is very desirable for navigation and operation in the Arctic.

Scenes 100730, 110725, and 110811 from training, testing, and validation, separately, are shown in Fig. 6 in sequential rows. These scenes are selected due to their complicated ice features. The corresponding image analysis, the ASI concentration, and the ice concentration estimation by CNN are shown in Fig. 7. Note that, for scene 100730, only about 1/8 of the total image area has a corresponding image analysis.

For all scenes, there is an obvious incidence angle effect that can be seen in the HH images. The tone is brighter at the near-range portion of the scene. There is no obvious effect of incidence angle visible in the ice concentration estimation. This indicates that the CNN model is able to model the incidence angle effect very well. Banding caused by a SAR defect in the HV image causes a slight overestimation of the ice concentration in the middle left of the water region for 110725 and 110811. Regions with less strong banding are successfully modeled, as the right part of the water regions in all three images has clean estimation of 0% ice concentration. The banding might be the cause of the overestimation at low-ice-concentration areas in Fig. 4. Melting causes the appearance of ice in the SAR image to become similar to water. At this time of year, the darker regions in the top left and middle left parts of scene 110811
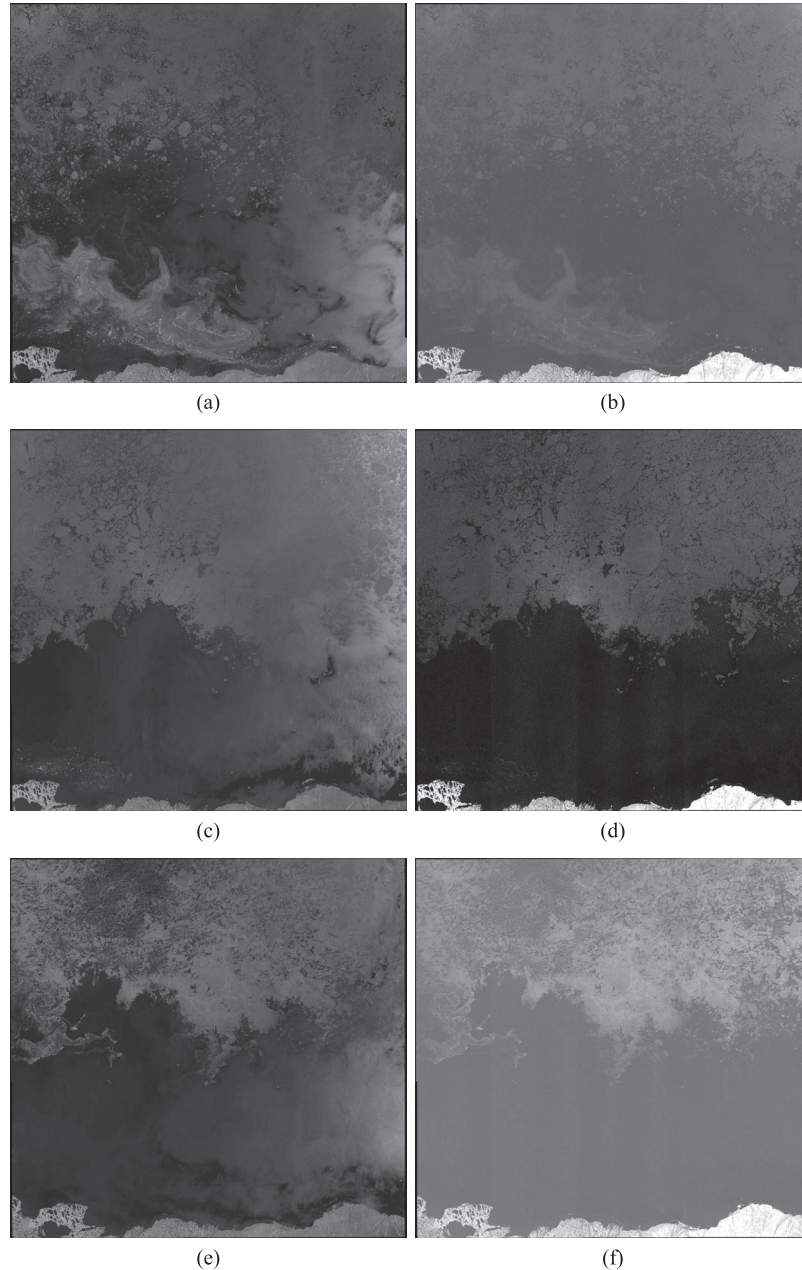
Fig. 6. Selected scenes of the RADARSAT-2 images used. The first to third rows are scenes 100730, 110725, and 110811, respectively. The first and second columns are the HH and HV bands, respectively. The HV images are enhanced for visualization.

are very likely to have water on the surface. The uniform dark regions in the HH image can be easily misinterpreted as water in automatic SAR image analysis methods and are similarly usually mistaken for ice of low concentration by passive microwave ice concentration retrieval algorithms [26], [33]. Those regions are labeled by CNN with intermediate ice concentration levels. In comparison, the ASI data indicate ice concentration that is very low, whereas the image analysis is between the two.

The ice concentration estimation by CNN also contains more details than the image analysis and the ASI ice concentration, particularly in low-ice-concentration regions (such as marginal ice zones). However, CNN is also generating noisy ice concentration estimates in some open water regions. It has been observed that this occurs when the scale of the feature on

the wind-roughened open water surface is on the order of the CNN window size. This effect could be reduced by using more training samples at those regions.

## VI. DISCUSSION

Experiments have been also carried out using HV pol or HH pol only (results not shown here). The use of dual-pol SAR image is able to produce a better ice concentration estimation than using HH or HV only, as has been demonstrated in previous research [12], [34]. When using HH pol only, the results are strongly affected by the incidence angle, whereas using only HV pol shows banding in the resulting ice concentration estimate.
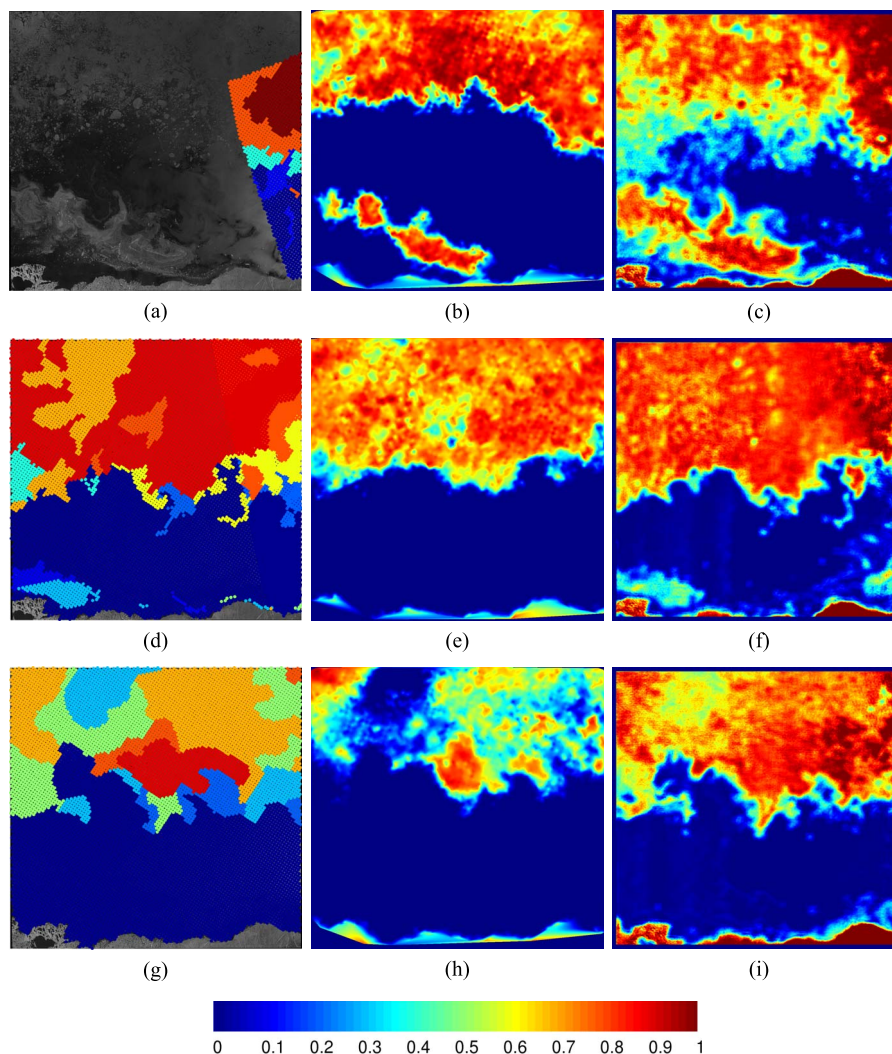
Fig. 7. (Column 1) Ice concentration from image analysis. (Column 2) ASI ice concentration. (Column 3) Estimation of CNN. The first to third rows correspond to scenes 100730, 110725, and 110811, respectively.

CNNs have a large number of hyperparameters, such as input image patch size, number of layers, number of filters, size of filters, and parameters for training. The tuning of these hyperparameters is essential to obtain good performance for a specific task. The patch size of the input for CNN, the filter size and pooling stride of the first convolution layer, and the size of the filters have been tuned to maximize performance in this study. Exhaustive searching of the best combination of hyperparameters is very time consuming. In this experiment, the hyperparameters are tuned in two steps. First, the input image patch size and filter sizes are tuned using only three images, each of which is used for training, validation, and training. Then, the numbers of filters and layers are tuned on the full data set using the size of input image patch and the size of filters from the last step.

The patch size of the input image and the filter size are related with the intrinsic scale and complexity of the problem. In the tuning experiment, different patch sizes were tested. The patch size was found to have an impact on the results with most of the differences coming from the banding effect and melt surfaces. A patch size of 41 showed the least banding effect and ice concentration underestimation, which might be caused by melt conditions, and was therefore adopted. CNN models with smaller patch size tend to underestimate ice concentration. Intuitively, the strength and the small-scale texture of the backscatter of a melt pond are very similar to calm water [33]; thus, the correct identification of melt ice needs more information from its neighborhood in the image. A small patch size causes confusion between melt and water due to this lack of enough supporting neighborhood information. Using patch size over 41 does not lead to improved performance. Similarly, larger patch size also benefits the recognition of wind-roughened water. If the banding effect were to be totally removed from the image, the optimal patch size might be different. Our tuning experiments suggest that the model is not very sensitive to the selection of other parameters as long as the model is large enough (sufficient number of filters and layers).

The image analyses are subsampled because their spatial resolution is much coarser than the SAR images, which introduces representation errors. It would be beneficial to model the errors explicitly [35], although the CNN is relatively robust to training sample errors [17]. Another benefit of the CNN is that it is
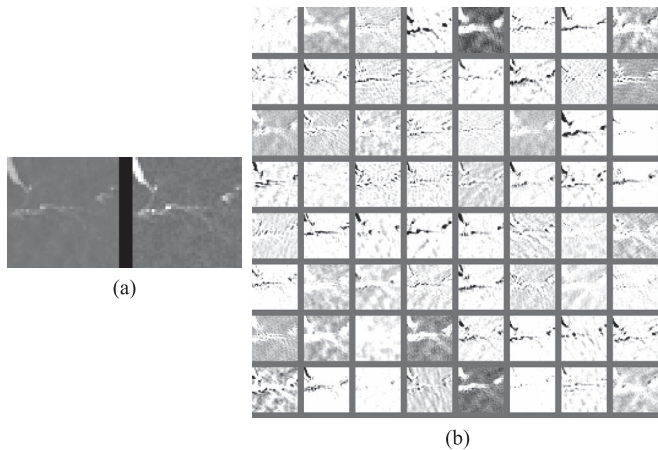
Fig. 8. CNN identifies SAR image structures. (a) (left) HH and (right) HV image patches of size 41 × 41 with ice concentration 0.2 indicated by image analysis (color enhanced). (b) Activation (output) of the 64 filters of the first convolutional layer (the color is reversed and the contrast is enhanced for better viewing).

a flexible computational structure. More observations can be easily incorporated into the current CNN as additional features of the input.

The training process ran for 500 epochs. The training and validation error started to converge at around 300 epochs, which is relatively slow. This may be caused by imbalance in the training samples used. Imbalanced training samples means there is a dominant class in the samples. In our case, the number of water (ice concentration is 0) samples is about eight times the second most common ice concentration level in the training samples. Intuitively, this pushes the model to a "dangerous local minimum" quickly [36]. When the model is at this local minimum, it detects most of the input as water and can still achieve a low cost. This causes underestimation of the ice concentration in general if the training is stopped early. It may take many epochs to get out of this local minimum, which leads to long training period. There are several approaches that may be investigated to resolve this issue, including undersampling the majority [37], [38], oversampling the minority [37], [38], or using a Bayesian cross-entropy cost function [36]. Our experiments (not shown here) show that learning normally converges within 50 epochs when using those methods, but none of the aforementioned methods converges to a model better than training directly on all the training samples for a long time. In this paper, we choose to prioritize precision and bear with the long training time. A better method to reduce the effect of imbalanced data is needed for better performance in terms of computational time and result accuracy.

Preliminary investigation of the connection between CNN filters and sea surface conditions is conducted by visualizing the activation maps (output) of the trained convolution layers (see Fig. 8). It shows that the network identifies SAR image structures. Some convolutional filters of the first convolutional layer produce feature maps with similar structures to the input SAR image patches.

The implementation of this algorithm is based on Theano [39] and Pylearn2 [40], which are two deep learning packages running in python. The code used for this study is shared on GitHub (https://github.com/leinxx/pylearn2_cnn).

## VII. CONCLUSION

In this paper, a CNN has been applied to dual-polarized SAR (HH and HV) images to generate ice concentration estimates. The CNN used takes the image patches of the intensity-scaled dual-pol SAR images as input and outputs ice concentration directly. Image analysis charts are used for training. State-of-the-art pixel-level result is acquired during the melt season in the Beaufort Sea. The ice concentration from the CNN contains abundant details of the ice compared with the image analyses. The results suggest that CNN is a robust method that can model the effect of incidence angle, SAR image noise, and the effect of wind on water and melt. Low-ice-concentration regions are also captured by the CNN model used.

The training on 11 images takes about 9 h, the prediction of ice concentration for one image takes around 10 min using an Nvidia GTX 780 graphic card with around 2000 processing cores. With more powerful graphic cards and appropriate optimization, the processing time can be largely reduced. Once the model is trained, the ice concentration estimation using the model can run in parallel on multiple graphics processing units easily. This is very promising for operational applications which require timely and robust ice concentration estimates over large regions.

This is a case study due to the limited amount of data available. For operational use, more comprehensive study and evaluation of this method in different regions and times of the year are required, which also requires the understanding of the connection between the CNN features and the sea surface conditions, which can be considered in future studies.

## REFERENCES

[1] O. M. Johannessen, V. Y. Alexandrov, V. Alexandrov, I. Y. Frolov, and L. P. Bobylev, *Remote Sensing of Sea Ice in the Northern Sea Route*. Chichester, U.K.: Praxis Publishing Limited, 2007.

[2] D. N. Thomas and G. S. Dieckmann, *Sea Ice*. Hoboken, NJ, USA: Wiley-Blackwell, 2009.

[3] J. L. Awange and J. B. Kiema, "Microwave remote sensing," in *Environmental Geoinformatics*. Berlin, Germany: Springer-Verlag, 2013, pp. 133–144.

[4] M. Buehner, A. Caya, L. Pogson, T. Carrieres, and P. Pestieau, "A new Environment Canada regional ice analysis system," *Atmos.-Ocean*, vol. 51, no. 1, pp. 18–34, Feb. 2013.

[5] M. A. N. Moen *et al.*, "Comparison of automatic segmentation of full polarimetric SAR sea ice images with manually drawn ice charts," *Cryosphere*, vol. 7, no. 6, pp. 1693–1705, Jun. 2013.

[6] F. D. Carsey, *Microwave Remote Sensing of Sea Ice*. Washington, DC, USA: AGU, vol. 68, 1992.

[7] B. Scheuchl, D. Flett, R. Caves, and I. Cumming, "Potential of RADARSAT-2 data for operational sea ice monitoring," *Can. J. Remote Sens.*, vol. 30, no. 3, pp. 448–461, Jan. 2004.

[8] T. Bøvith and S. Andersen, "Sea ice concentration from single-polarized SAR data using second-order grey level statistics and learning vector quantization," Danish Meteorol. Inst., Copenhagen, Denmark, Tech. Rep. 05-04, 2005.

[9] J. Karvonen, B. Cheng, T. Vihma, M. Arkett, and T. Carrieres, "A method for sea ice thickness and concentration analysis based on SAR data and a thermodynamic model," *Cryosphere*, vol. 6, no. 6, pp. 1507–1526, Dec. 2012.

[10] A. Berg and L. E. B. Eriksson, "SAR algorithm for sea ice concentration—Evaluation for the Baltic sea," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 5, pp. 938–942, Sep. 2012.

[11] N. Kasapoglu, "Sea ice concentration retrieval using composite ScanSAR features in a SAR data assimilation process," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2085–2089, Dec. 2014.

[12] 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.

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

[14] J. Karvonen, "Baltic sea ice concentration estimation based on C-band HH-polarized SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 6, pp. 1874–1884, Dec. 2012.

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

[16] M. P. Mäkynen *et al.*, "Incidence angle dependence of the statistical properties of C-band HH-polarization backscattering signatures of the Baltic Sea ice," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 12, pp. 2593–2605, Dec. 2002.

[17] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, Jan. 2009.

[18] G. Hinton, "Learning multiple layers of representation," *Trends Cogn. Sci.*, vol. 11, no. 10, pp. 428–434, Oct. 2007.

[19] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, Inc., 2012, pp. 1097–1105.

[21] Y. Bengio, "Deep learning of representations: Looking," in *Statistical Language and Speech Processing*, ser. Lecture Notes in Artifitial Intelligence. Berlin, Germany: Springer-Verlag, 2013, vol. 7978, pp. 1–37.

[22] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 609–616.

[23] D. Ciresan and U. Meier, "Flexible, high performance convolutional neural networks for image classification," in *Proc. 22nd Int. Joint Conf. Artif. Intell.*, 2011, pp. 1237–1242.

[24] A. Karpathy *et al.*, "Large-scale video classification with convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 1725–1732, 2014.

[25] J. E. Overland, "Meteorology of the Beaufort Sea," *J. Geophys. Res., Oceans*, vol. 114, no. C1, Jan. 2009, Art. no. C00A07.

[26] G. Spreen, L. Kaleschke, and G. Heygster, "Sea ice remote sensing using AMSR-E 89-Ghz channels," *J. Geophys. Res.*, vol. 113, no. C2, Feb. 2008, Art. no. C02S03.

[27] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[28] Y.-L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in *Proc. 27th ICML*, 2010, pp. 111–118.

[29] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *Artificial Neural Networks—ICANN*. Berlin, Germany: Springer-Verlag, 2010, pp. 92–101.

[30] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th ICML*, 2010, pp. 807–814.

[31] Y. LeCun, L. Bottou, G. Orr, and K. Müller, "Efficient backprop," in *Neural Networks: Tricks of the Trade*. Berlin, Germany: Springer-Verlag, 2012, pp. 9–48.

[32] L. Prechelt, "Early stopping—But when?" in *Neural Networks: Tricks of the Trade*. Berlin, Germany: Springer-Verlag, 2012, pp. 53–67.

[33] J. M. Hanesiak, D. G. Barber, R. A. De Abreu, and J. J. Yackel, "Local and regional albedo observations of Arctic first-year sea ice during melt ponding," *J. Geophys. Res., Oceans*, vol. 106, no. C1, pp. 1005–1016, Jan. 2001.

[34] L. Wang, K. Scott, and D. A. Clausi, "Automatic feature learning of SAR images for sea ice concentration estimation using feed-forward neural networks," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 3969–3971.

[35] V. Mnih and G. E. Hinton, "Learning to label aerial images from noisy data," in *Proc. 29th ICML*, 2012, pp. 567–574.

[36] A. Dalyac, M. Shanahan, and J. Kelly, "Tackling class imbalance with deep convolutional neural networks," Dept. Comput. Sci., Imperial College London, London, U.K., Tech. Rep., 2014.

[37] M. A. Maloof, "Learning when data sets are imbalanced and when costs are unequal and unknown," in *Proc. Int. Conf. Mach. Learn. Workshop Learn. Imbalanced Data Sets II*, 2003, pp. 1–8.

[38] Z.-H. Zhou and X.-Y. Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 1, pp. 63–77, Jan. 2006.

[39] F. Bastien *et al.*, "Theano: New features and speed improvements," in *Proc. Deep Learn. Unsupervised Feature Learn. NIPS Workshop*, 2012, pp. 1–10.

[40] I. J. Goodfellow *et al.*, "Pylearn2: A machine learning research library," unpublished paper, 2013. [Online]. Available: http://arxiv.org/abs/1308.4214
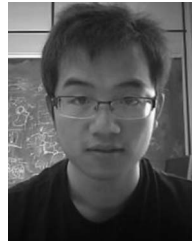
**Lei Wang** (S'14) received the B.S. and M.S. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2009 and 2012, separately. He is currently working toward the Ph.D. degree in systems design engineering at the University of Waterloo, Waterloo, Canada.

His research interests are in the field of large-scale machine learning for remote-sensed images.

**K. Andrea Scott** (M'12) received the B.A.Sc. degree from the University of Waterloo, Waterloo, ON, Canada, in 1999, the M.A.Sc. degree from McMaster University, Hamilton, ON, in 2001, and the Ph.D. degree from the University of Waterloo in 2008.

After completing her doctorate studies, she worked as a Postdoctoral Researcher in the Data Assimilation and Satellite Meteorology Research Section of Environment Canada. At Environment Canada, she was part of a team working on the development of a sea-ice data assimilation system. Since 2012, she has held a faculty position in the Department of Systems Design Engineering at the University of Waterloo, specializing in sea-ice remote sensing and data assimilation.
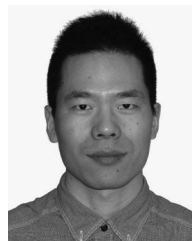
**Linlin Xu** (M'14) received the B.Eng. and M.Sc. degrees in geomatics engineering from the China University of Geosciences, Beijing, China, in 2007 and 2010, respectively, and the Ph.D. degree in geography from the University of Waterloo, Waterloo, ON, Canada, in 2014.

He is currently a Postdoctoral Fellow with the Vision and Image Processing Laboratory, Department of Systems Design Engineering, University of Waterloo. His current research interests are in the areas of hyperspectral and synthetic aperture radar image processing.

**David A. Clausi** (S'93–M'96–SM'03) received the B.A.Sc., M.A.Sc., and Ph.D. degrees from the University of Waterloo, Waterloo, ON, Canada, in 1990, 1992, and 1996, respectively, all in systems design engineering.

He was in the medical imaging field at Agfa, Waterloo. In 1997, he was an Assistant Professor of geomatics engineering with the University of Calgary, Calgary, Canada. In 1999, he returned to the University of Waterloo, where he is currently a Professor, specializing in the fields of intelligent and environmental systems. He is an active interdisciplinary and multidisciplinary researcher. He has authored or coauthored various papers published in refereed journals and conference proceedings in the diverse fields of remote sensing, computer vision, algorithm design, and biomechanics. His research efforts have led to successful commercial implementations.

Prof. Clausi was the Cochair of the International Association for Pattern Recognition Technical Committee 7 on Remote Sensing, during 2004–2006. He was a recipient of numerous scholarships, paper awards, and two Teaching Excellence Awards. He was also a recipient of the award for Research Excellence and Service to the Research Community by the Canadian Image Processing and Pattern Recognition Society in 2010.