

Ice concentration estimation from dual-polarized SAR images using deep convolutional neural networks

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

Abstract

High resolution ice concentration maps are of great interest for shipping and operations in the Arctic. A convolutional neural network (CNN) has been used to estimate ice concentration from C-band dual-polarized RADARSAT-2 satellite images in the melt season. SAR images are used as input and the ice concentration is the direct output from the CNN. With no feature extraction or segmentation-based post processing, the errors of the generated ice concentration maps compared to image analyses are less than 10% on average. The CNN is demonstrated to produce ice concentration maps with more details than image analyses. Reasonable ice concentration estimations are made in the melt season, and regions of low ice concentration ice are also well captured.

Index Terms

Sea ice concentration, synthetic aperture radar, convolutional neural network

I. INTRODUCTION

With the recent reductions in Arctic ice extent, there has been a growing economic interest in shipping and mining in the Arctic [1]. To support safe Arctic operations and navigation in ice infested waters, high resolution ice concentration information is crucial [2]. Ice concentration is the percentage of sea ice coverage over a given spatial area. Sea ice concentration is operationally monitored using satellite sensors. Due to the frequent cloud coverage and long periods with little solar illumination, 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 the existence of sun illumination. Passive microwave sensors gather and measure 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 sends signals to the earth and measures the reflected signals. 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. Therefore, SAR images contain more details than passive microwave images.

Due to the lack of robust automatic SAR ice concentration estimation algorithms, SAR images are still manually interpreted for operational purposes, such as sea ice mapping and prediction, 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 the Canada's RADARSAT Constellation mission and the European Space Agency Sentinel mission. With the rapidly growing data volume, automatic ice concentration estimation and even data mining from the large amount of SAR images are strongly desired.

The main difficulty in automatic ice concentration estimation from SAR images is caused by 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 instrument incidence angle and surface conditions such as surface roughness, moisture, snow cover and salinity [6]. For example, calm water generally has low backscatter compared to wind-roughened water most noticeably in HH band SAR images. The sensitivity of scattered SAR signals to wind speed varies with incidence angle. The appearance of ice in SAR images also changes through the year due to melting, ice dynamics and floe distribution. These factors make it a challenging task to design image features to capture all the diverse patterns seen in SAR sea ice imagery.

However, there has been a constant effort towards automatic ice concentration estimation from SAR imagery [7–13]. A two step approach is normally used [7–13]: first, image features are 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 (MLP) are examples of such regression models

[9, 10, 12, 13].

A number of features have been proposed to estimate ice concentration from SAR images, such as auto correlation and GLCM [10, 11], local normalized intensity and intensity difference of different polarizations [10, 12]. In Karvonen’s work [9], the density of edges and corners are used as features to detect ice from SAR images. To calculate more robust and representative features, preprocessing of SAR images, such as backscatter calibration [12, 13] and incidence angle correction [9, 12], are normally applied. 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, while using many features together led to improved correlation with ice concentration [11]. This study demonstrated the difficulty in designing features to capture ice concentration from SAR imagery.

Both single polarized [8, 10, 14] and dual polarized [11–13] SAR images have been used to estimate ice concentration. The use of dual polarized SAR images has been shown to be beneficial for ice information retrieval from SAR images [7, 12, 13, 15], because the backscatter of the HV band is not strongly affected by the incidence angle. The use of dual-polarized SAR images has been found to reduce the over estimation of ice concentration in open water regions and improve the estimation of ice concentration in melt conditions [12].

Most of the aforementioned methods use segmentation to improve the numerical quality and visual appearance of their results, which will decrease the resolution of the estimated ice concentration maps. To generate high resolution ice concentration estimation, it is more appropriate to use pixel-based ice concentration estimation methods (without segmentation). To generate “smooth” ice concentration estimations without segmentation based refinement, the estimation algorithm is required to be very robust. As the ice and water are changing through time and location, it is challenging to design robust SAR image features that can work for all different locations and seasons. Redesign of the algorithm or features may be necessary for different situations. Based on these observations, it is of interest to investigate the use of feature learning instead of feature design for ice concentration estimation from SAR images. If the right features can be learned from training data using feature learning, the right features can be generated for specific data and conditions without human intervention. This is highly desirable for large scale automatic ice concentration estimation tasks. Deep learning is a feature learning method that uses multiple layers of neural networks [16]. Deep learning has demonstrated great potential in many different recognition tasks such as speech

recognition and image object classification [17–20].

For this purpose, in this study, the estimation of ice concentration under the framework of deep learning is investigated. A deep convolutional neural network (CNN) has been used to estimate ice concentration directly from dual band SAR images (HH and HV). CNN is a deep learning 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 [19, 21–23]. This algorithm is demonstrated to be able to automatically generate high quality ice concentration estimation in the melt season, under a variable surface conditions, without incidence angle correction.

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 Section V and VI, which is followed by conclusions in Section VII.

II. DATASET USED IN THIS STUDY

The study area is located in the Beaufort Sea, north of Alaska. This region is important for oil and gas exploration, and it is a particularly relevant area for high resolution ice concentration mapping research. The Beaufort Sea is frozen throughout the year except in the summer (between July and October). Ice at this time of year is a mixture of multi-year ice, first year ice and new ice. Packed multi-year ice often breaks into smaller ice floes, leaving large areas of open water between the floes. Melting ice surfaces can be observed in this region until October [24].

The dataset used in this study is a set of 15 RADARSAT-2 dual polarized (HH and HV) ScanSAR wide beam images of the Beaufort Sea (Fig. 1) provided by MDA[®]. It is a subset of the dataset described in [15]. Only the scenes with corresponding image analyses are used. Each scene covers a 500 km by 500 km area with 50 m by 50 m nominal pixel spacing. The incidence angle for each scene ranges from 20° to 49°. Exact dates of the images used are given in Table I. All scenes are acquired between July and October, with 9 scenes from 2010 and 6 scenes from 2011.

Image analyses are used to train and verify the neural network in this study. The image analyses used are ice maps produced by ice experts from the Canadian Ice Service (CIS) by visually interpreting SAR images [4]. An image analysis is composed of contiguous polygons that are generated by the ice analyst as regions with certain ice concentration. Each polygon

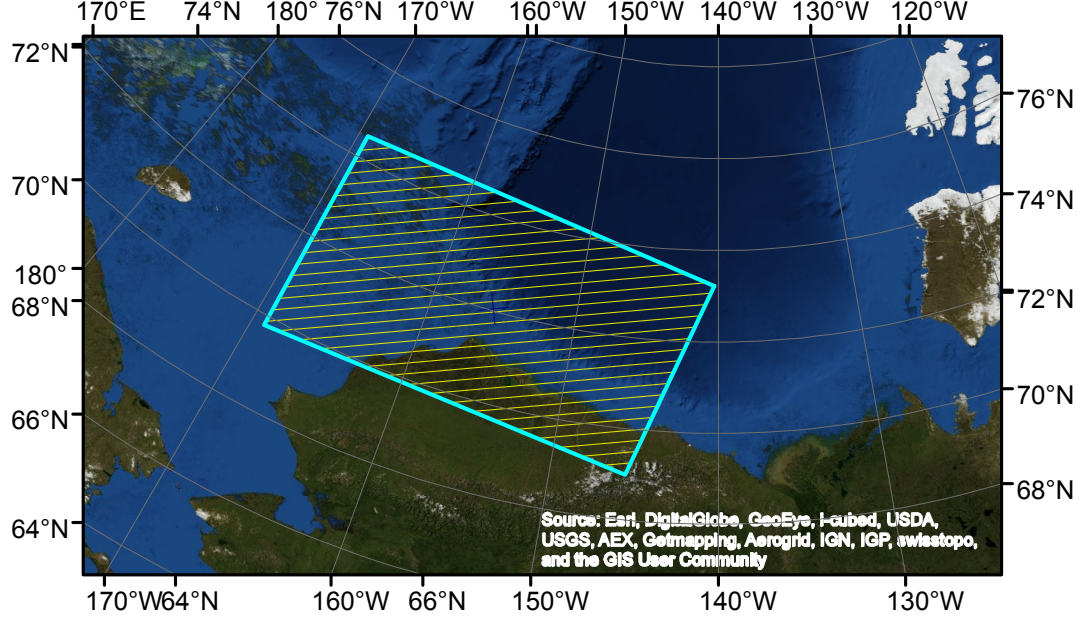


Fig. 1: The study area in the Beaufort Sea used in this study.

is assigned an ice concentration value and portions 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 which represent the portion of ice in that polygon). The precision (error standard deviation) of the image analysis ice concentration is approximately 10% [4]. The image analyses obtained from CIS in this study consist of sampled points of the original image analysis charts using a sampling interval of about 8 km by 5 km. Since the details within one polygon in the image analysis are not considered, the true ice concentration at the sample locations may be very different from the ice concentration of the polygon. This introduces representation errors when using the sample points from the image analysis for training. For some scenes, the image analyses only cover a small portion of the whole image (Fig. 7(a)). The number of image analysis sample points for each SAR image are given in Table I. Illustrations of SAR images and image analyses are shown in Fig. 6 and Fig. 7(a, d, g).

The AMSR-E daily ice concentration maps are used for comparison purposes. The AMSR-E ice concentration data were acquired from <http://www.iup.uni-bremen.de/seaice/amsr/>. The ice concentration used was generated from the AMSR-E 89GHz channel data using the ARTIST sea ice (ASI) algorithm [25]. This algorithm is known to underestimate ice concentration when ice concentration is low, and was chosen due to its relatively high spatial

| Year | Date (month.days) | number of samples |
|------|-------------------|-------------------|
| 2010 | 7.30 | 1254 |
| | 8.06 | 6125 |
| | 8.22 | 10516 |
| | 8.29 | 2145 |
| | 9.09 | 10236 |
| | 9.29 | 7473 |
| | 10.03 | 11560 |
| | 10.06 | 7981 |
| | 10.08 | 1752 |
| 2011 | 7.09 | 561 |
| | 7.10 | 3570 |
| | 7.20 | 1943 |
| | 7.25 | 12005 |
| | 8.11 | 9736 |
| | 8.17 | 7927 |

TABLE I: Data acquisition times and the number of sample points in their corresponding image analyses. Note that, because of the dense uniform sampling pace within the SAR image, one thousand samples can only cover a small portion (around 8-12%) of the image.

resolution [25]. The pixel spacing of this dataset is 6.25 km. This dataset is referred to as ASI ice concentration in the remaining of this paper.

III. CONVOLUTIONAL NEURAL NETWORK (CNN)

A CNN is an artificial neural network introduced by LeCun et al. [26]. A CNN consists of a stack of alternating convolution layers and pooling layers with a number of fully connected layers on top. Fig. 3 shows the structure of a CNN with two pairs of convolution layers and pooling layers and 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 pixel-wise transformation of the convolved images by an activation function. With multiple filters, multiple feature maps are generated and stacked to a multilayer feature map. The convolution process for each filter is described by (1).

$$\mathbf{h}_{ij} = \sigma((\mathbf{W} * \mathbf{X})_{ij} + b), \text{ in which,} \quad (1a)$$

$$i = 1, \dots, M_x, j = 1, \dots, M_y \quad (1b)$$

$$M_x = \frac{S_x - K_x}{P} + 1 \quad (1c)$$

$$M_y = \frac{S_y - K_y}{P} + 1. \quad (1d)$$

In (1), \mathbf{X} is the input image of width and height denoted by S_x and S_y . A filter \mathbf{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 σ . The generated feature map \mathbf{h} has width and height M_x and M_y and depends on the step size of the filters, denoted by P . When P is one, every possible location in the input image is filtered. When P is two, every other pixel of the input image is filtered.

B. Max pooling layer

A max pooling layer sub-samples the feature map generated by its previous convolution layer by mapping each non-overlapped rectangle image block to the maximum value of this block [27, 28]. 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 of the pooling layer. This redundancy arises because the convolution window of neighboring locations are highly overlapped. Upper convolution layers always have larger receptive fields than their previous convolution layers, which means 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 study, a supervised CNN model is used as a regression model to generate ice concentration from SAR images. The processing scheme is shown in Fig. 2. It is composed of three major steps: preprocessing of the SAR images, training (with validation and testing) and prediction using the trained CNN. Prediction is the same as testing except that an ice concentration is estimated for every location in the input SAR image, therefore only the first two steps will be described.

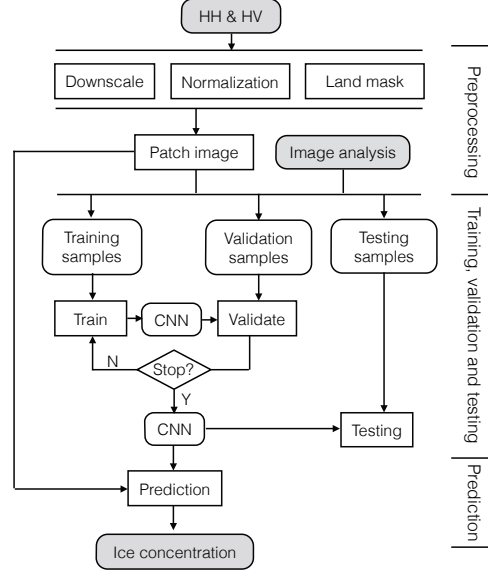


Fig. 2: Flow chart of the major steps of the algorithm.

A. Preprocessing of SAR images

All the SAR images are sub-sampled by averaging non-overlapped 8 by 8 blocks of the image to reduce image noise and data volume. The sub-sampled images have 400 m spatial resolution with pixel values between 0 and 255. Standard CNNs using sigmoid activation functions take input values normalized to $[-1, 1]$. In this study, rectified linear units (ReLU) units (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 units with negative input will be 0. No learning is happening when using stochastic gradient descent (SGD) on those units. For this reason, instead of normalizing the input to $[-1, 1]$, pixel values of the dual-band SAR images are normalized to $[0, 1]$ by dividing each pixel value by 255. A land mask is applied to remove the samples located on land.

Image patches of 40 by 40 pixels centered at the image analysis sample locations are extracted. 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 band are considered, each image patch is a three dimensional matrix of size 40 by 40 by 2. Each SAR image patch and the ice concentration located at the patch center from the image analysis is one sample used in this experiment. The patch size has been tuned by testing the patch sizes from 28 to 60 at 4 pixel interval and selecting that with the minimal estimation error.

B. Structure of CNN

The structure of the CNN used in this study is illustrated in Fig. 3. 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. Inputs to this filter are image patches of size 40 by 40 by 2 (height, width, number of bands), so the size of each filter is (5, 5, 2). The filter outputs convolved images of size (40-5+1, 40-5+1, 64), that is (36, 36, 64) based on (1). The first convolution layer is followed by a max pooling layer of pooling size 2 by 2 [19, 22]. The max pooling layer outputs the max activation of every unoverlapping 2 by 2 block of the convolved image to form a feature map of size (36/2, 36/2, 64), which is (18, 18, 64). 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 $7*7*128 = 6272$ and outputs only one single value, which is the ice concentration estimation.

ReLU (2a) were chosen for both convolution layers because they have been demonstrated to lead to faster learning and better features compared to sigmoid activation units due to the fact that they do not saturate as sigmoid activation units [19, 29]. The top linear layer uses a linear activation function (2b) with one output unit, which is the ice concentration.

$$f(x) = \max(0, x) \quad (2a)$$

$$f(x) = x. \quad (2b)$$

C. Training, validation and testing

Back-propagation and mini-batch stochastic gradient descendant (SGD) [30] are used for training. The training samples are randomly separated to $N/128$ mini-batches, in which N

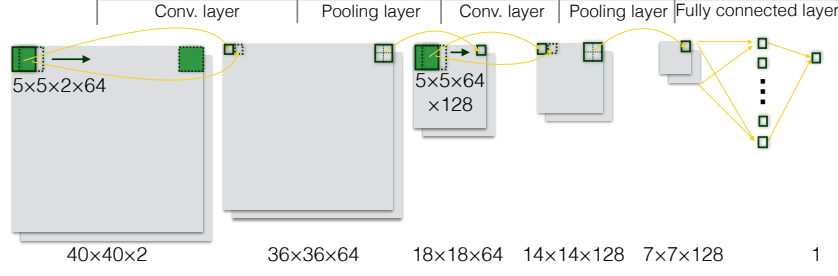


Fig. 3: The structure of the CNN model used in this study. The outer rectangles are inputs and outputs of different layers. The size of the input of each layer (width, height and number of bands) is shown at the bottom. The smaller filled rectangles are convolution filters the sizes of which are described beneath the filters. For example, the first convolution layer contains 64 filters of size 5 by 5 by 2. The input of this layer is a image patch with 2 bands and each band is 50 by 50 pixels.

is the total number of training samples. Each mini-batch contains 128 training samples. In our experiment, the mean squared-error over the mini-batch is used (3) as the cost function for training.

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

In (3), the forward propagation of the CNN is noted by F . The i th image patch and corresponding image analysis of the mini-batch is noted by \mathbf{X}_i and \mathbf{y}_i . The CNN takes \mathbf{X} as input and generate a estimation of ice concentration based on its current weights \mathbf{W} . The parameters of the CNN are updated according to the derivative of the parameters to the cost function over one mini-batch, which is described in (4).

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

$$\mathbf{W}_{i+1} = \mathbf{W}_i + \mathbf{V}_{i+1}. \quad (4b)$$

The gradients of the weights with respect to the cost function ($\frac{\partial L}{\partial \mathbf{W}}$) are calculated and averaged over the mini-batch. The weights \mathbf{W} are updated with learning rate $\epsilon = 10^{-3}$ and weight decay of $r = 2 \times 10^{-5}$ with momentum $m = 0.9$. The setting of the training parameters for SGD are similar to the setting in [19]. Some adjustments are made by tuning. Similar to [19], the parameters of the CNN are initialized by uniform random sampling between -0.05 and 0.05.

An epoch training scheme [30] is adopted. For each epoch, all the mini-batches of the training samples are iterated once by the training algorithm. To accelerate the training process in case the training converges early (which is typical [31]), 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.

To avoid over-fitting, it is a common practice to use validation dataset to select the model of best performance during the training process [19]. The 15 scenes are divided to training, testing and validation sets. 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. The trained CNN model is then evaluated on the test data.

A leave-one-out experiment is conducted. In this study, one scene is used for testing, another scene is used for validation and the remaining are used for training because a large number of training samples is required to train the CNN. The scene with the most complete samples of ice concentration levels, which is 20110811, is selected for validation for all rounds of the leave-one-out experiment. Of the remaining 14 scenes (without scene 20110811), there are 7 for which samples from image analysis charts cover over half the image area, these are used as testing candidate images, which means the leave-one-out experiment consists of seven rounds. In each round, a test scene is selected from the test candidates and the remaining 13 scenes are selected for training. The testing is not done on every image because of two reasons: the incomplete image analyses contains a small amount of samples in a few ice concentration levels, testing on those images is not thorough due to the lack of ice concentration levels; secondly, each round of the experiment requires over one day to run, it requires over two weeks to run all the testing cases.

V. ICE CONCENTRATION ESTIMATION RESULTS

The estimated ice concentration is evaluated against image analyses obtained from CIS. The mean difference between the estimated ice concentration and image analysis (E_{L1}), the mean of the absolute difference between the two (E_{sgn}) and the standard deviation of the error (E_{std}) are used to evaluate the performance of the algorithm. The evaluation results for all training, validation and testing for all rounds of the leave-one-out experiment are shown in Table II.

Based on Table II, the performance of this algorithm is quite stable. The mean error E_{sgn}

| | Training | | | Validation | | | Testing | | |
|------------|-----------|----------|-----------|------------|----------|-----------|-----------|----------|-----------|
| Test image | E_{sgn} | E_{L1} | E_{std} | E_{sgn} | E_{L1} | E_{std} | E_{sgn} | E_{L1} | E_{std} |
| 20100822 | -0.06 | 0.10 | 0.16 | 0.05 | 0.08 | 0.12 | 0.00 | 0.08 | 0.14 |
| 20100909 | -0.05 | 0.09 | 0.16 | 0.05 | 0.08 | 0.12 | -0.00 | 0.04 | 0.09 |
| 20100929 | -0.05 | 0.09 | 0.15 | 0.04 | 0.07 | 0.12 | -0.02 | 0.10 | 0.19 |
| 20101003 | -0.03 | 0.09 | 0.14 | 0.03 | 0.07 | 0.12 | -0.08 | 0.14 | 0.20 |
| 20101006 | -0.05 | 0.09 | 0.15 | 0.03 | 0.07 | 0.12 | -0.06 | 0.10 | 0.16 |
| 20110817 | -0.06 | 0.10 | 0.16 | 0.04 | 0.07 | 0.12 | 0.03 | 0.07 | 0.12 |
| 20110725 | 0.02 | 0.08 | 0.13 | 0.09 | 0.10 | 0.14 | 0.01 | 0.09 | 0.13 |
| Average | -0.04 | 0.09 | 0.15 | 0.05 | 0.08 | 0.12 | -0.02 | 0.09 | 0.15 |
| ASI | -0.16 | 0.16 | 0.24 | -0.13 | 0.13 | 0.17 | -0.11 | 0.11 | 0.16 |

TABLE II: The ice concentration estimation results verified against image analyses. The validate image is 20110811 for all rounds. The testing image used is given in the first column. Row one to seven are the error statistics for the seven rounds of the leave-one-out experiment. Row “Average” represents the average statistics of the seven rounds. Row “ASI” is the ASI ice concentration verified against image analysis. The different sets (training, validation and testing) for ASI means the ASI ice concentration at days of the corresponding SAR images in that set (scene 20110725 for testing, 20110811 for validation and the rest for training).

is less than 5% on average. So the overall bias of the estimated ice concentration is less than 5%. The largest mean error for training, validation and testing are -0.06, 0.09 and -0.08 respectively. The mean error is less than 0.1 (10%) in all cases. The average absolute difference (E_{L1}) for all three sets are less than 0.1 (10%). The mean standard deviation (E_{std}) for testing is slightly more diverse for testing than training and validation. This is reasonable because the validation data are not changed between rounds, and only a portion of the training sets is different for each round, but the testing image is completely changed for the different rounds. Compared to ASI ice concentration, ice concentration generated by the CNN is closer to the image analysis. The performance of the CNN on the testing and validation dataset is very close to its performance on the training dataset. This indicates that the method can generalize to unseen data well. While the image analysis is quantized, the ice concentration estimation using CNN is continuous. This difference may potentially introduce errors into the evaluation statistics. Therefore, the ice concentration estimates are also evaluated after been quantized to 11 levels between 0 and 1 similar to the image analysis. The evaluation results are slightly improved after quantizing the ice concentration estimation (results not

shown here).

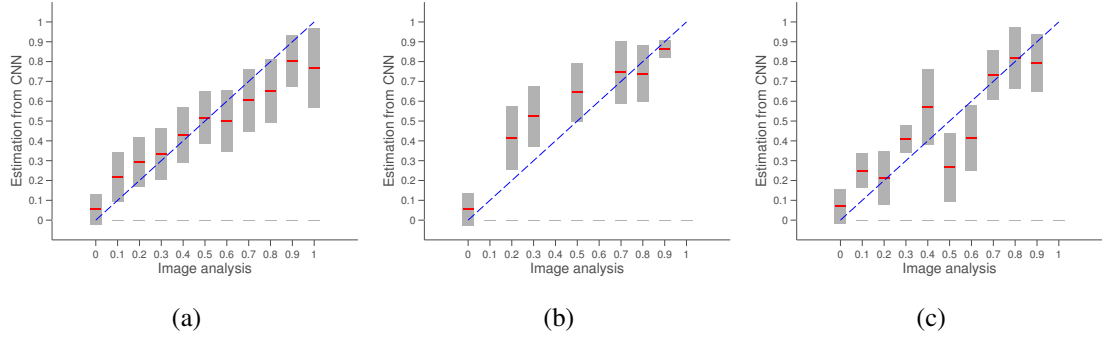


Fig. 4: Errors at different ice concentration levels for training (a), validation (b) and testing (c) when using scene 20110725 for testing, scene 20110811 for validation and the remaining scenes for training.

Fig. 4 shows the means and standard deviations of ice concentration estimation errors for different ice concentration bins when using scene 20110725 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 dataset, the classification of pure water is better than the rest 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. 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 (the area covered by yellow dots in Fig. 5). By inspecting the SAR images (Fig. 6), we 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, and 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.

Selected scenes 20100730, 20110725 and 20110811 are shown in Fig. 6 in sequential rows. These three images are used for training, testing and validation separately. The corresponding image analysis, ASI ice concentration and ice concentration estimation by CNN are shown in Fig. 7. Note that for scene 20100730, only about 1/8 of the total image area has a corresponding image analysis.

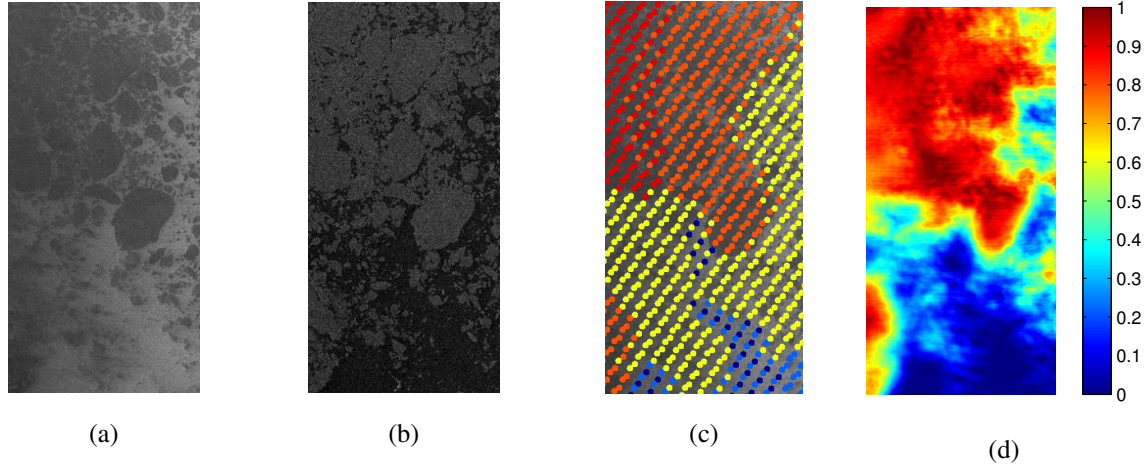


Fig. 5: Ice concentration estimation details shown for a subset of the scene 20110725. (a) HH; (b) HV (enhanced for visualization); (c) image analysis plotted over the HH band image using the same color scheme as (d); (d) estimated ice concentration when using scene 20110725 for testing.

For all scenes, there is an obvious incidence angle effect that can be seen in the HH images. The tone is brighter at the right end of the images. 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 over estimation of the ice concentration in the middle left of the water region for 20110725 and 20110811. Regions with less strong banding are successfully modeled, as the right part of the water regions in all three images have clean estimation of 0% ice concentration. The banding might be the cause of the over estimation at low ice concentration 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 the middle left parts of scene 20110811 are very likely to have water on the surface. The uniform dark regions in 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. Those regions are captured by CNN as being regions of intermediate ice concentration. In comparison, the ASI data indicate ice concentration that is very low, while the image analysis is between the two.

The ice concentration estimation by CNN also contains more details than the image analysis and ASI ice concentration especially at low ice concentration regions (such as ice water

boundaries). However, CNN is also generating noisy ice concentration estimates in some open water regions. This effect could be reduced by selecting more samples at those regions or through post processing.

VI. DISCUSSION

Experiments have also been carried out using HV band or HH band only (results not shown here). The use of dual band 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, 32]. When using HH band only, the results are strongly affected by the incidence angle, while using only HV band shows banding in the resulting ice concentration estimate. In our previous work, multi-layer perceptron was used to estimate ice concentration under similar settings [32].

CNNs have a large number of hyper parameters, such as input image patch size, the number of layers and the number of filters, the size of filters and parameters for training. The tuning of these hyper-parameters 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 hyper-parameters is very time consuming. In this experiment, the hyper-parameters 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 number of filters and layers are tuned on the full dataset using the size of input image patch and size of filters from the last step.

Patch size of the input image and 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 a small impact on the results with most of the differences coming from the banding effect. A patch size of 40 showed the least banding effect in the result and was therefore adopted. If the banding effect were to be 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 filters and layers).

The image analyses are sub-sampled because their spatial resolution is coarser than the SAR images, which introduces representation errors. It would be beneficial to model the errors explicitly [33], even though the CNN is relatively robust to training sample errors[16].

Another benefit of the CNN is that it is a flexible computational structure. More observations can be easily incorporated into the current CNN as additional bands 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 8 times the second most common ice concentration level in the training samples. Intuitively, this pushes the model to a “dangerous local minimum” quickly [34]. 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 under estimation of the ice concentration in general. It may take many epochs to get out of this local minimum. There are several approaches that may be investigated to resolve this issue including under sampling the majority [35, 36], over sampling the minority [35, 36] or using a Bayesian cross entropy cost function [34]. 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 the imbalanced data for a long time. A better method to reduce the effect of imbalanced data is needed for better performance.

VII. CONCLUSION

In this paper, a convolution neural network is applied to dual polarized SAR images to generate ice concentration maps. The CNN used takes the image patches of the intensity-scaled dual band SAR images as input and outputs ice concentration directly. Image analysis charts are used for training. The state-of-the-art pixel based result is acquired during the melting season in the Beaufort sea. The ice concentration from the CNN contains abundant details of sea ice compared to the image analysis. The results suggest that the CNN is a simple yet robust model that can model the incidence angle, banding effect, SAR image noise, the effect of wind on water and melting. Low ice concentration regions are also captured by the CNN.

The training on 13 images takes about 10 hours, the prediction of ice concentration for one image takes 0.5 hours using Nvidia GTX 780 graphic card with around 2000 processing cores. With more powerful graphic cards and parallel computing, 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 GPUs easily. This is very promising for operational

applications which requires timely ice concentration estimations.

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*. Praxis Publishing Limited, 2007.
- [2] D. N. Thomas and G. S. Dieckmann, *Sea ice*. Wiley-Blackwell, 2009.
- [3] J. L. Awange and J. B. Kiema, “Microwave Remote Sensing,” in *Environmental Geoinformatics*, Springer, 2013, pp. 133–144.
- [4] M. Buehner, A. Caya, L. Pogson, T. Carrieres, and P. Pestieau, “A new Environment Canada regional ice analysis system,” *Atmosphere-Ocean*, vol. 51, no. 1, pp. 18–34, 2013.
- [5] M. A. N. Moen, A. P. Doulgeris, S. N. Anfinson, A. H. H. Renner, N. Hughes, S. Gerland, and T. Eltoft, “Comparison of automatic segmentation of full polarimetric SAR sea ice images with manually drawn ice charts,” *The Cryosphere*, no. 7, 2013.
- [6] F. D. Carsey, *Microwave remote sensing of sea ice*. American Geophysical Union, 1992, vol. 68.
- [7] B. Scheuchl, D. Flett, R. Caves, and I. Cumming, “Potential of RADARSAT-2 data for operational sea ice monitoring,” *Canadian Journal of Remote Sensing*, vol. 30, no. 3, pp. 448–461, 2004.
- [8] T. Bvith and S. Andersen, “Sea ice concentration from single-polarized SAR data using second-order grey level statistics and learning vector quantization,” *DMI Scientific Report*, no. 05-04,
- [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,” *The Cryosphere*, vol. 6, pp. 1507–1526, 2012.
- [10] A. Berg and L. E. B. Eriksson, “SAR algorithm for sea ice concentration-evaluation for the Baltic sea,” *Geoscience and Remote Sensing Letters, IEEE*, vol. 9, no. 5, pp. 938–942, 2012.
- [11] N. Kasapoglu, “Sea ice concentration retrieval using composite scansar features in a sar data assimilation process,” *Geoscience and Remote Sensing Letters, IEEE*, vol. 11, no. 12, pp. 2085–2089, 2014.

- [12] J Karvonen, “Baltic sea ice concentration estimation based on C-band dual-polarized SAR data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 9, pp. 5558–5566, 2014.
- [13] J. A. Karvonen, “A sea ice concentration estimation algorithm utilizing radiometer and SAR data,” *The Cryosphere Discussions*, vol. 8, no. 2, pp. 2213–2241, 2014.
- [14] J Karvonen, “Baltic sea ice concentration estimation based on C-band HH-polarized SAR data,” *Selected Topics in Applied Earth Observations and remote sensing*, vol. 5, no. 6, pp. 1874–1884, 2012.
- [15] S. Leigh and D. A. Clausi, “Automated ice-water classification using dual polarization SAR satellite imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 9, pp. 5529–5539, Sep. 2014.
- [16] Y. Bengio, “Learning deep architectures for AI,” *Foundations and trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.
- [17] G. E. Hinton, “Learning multiple layers of representation,” *Trends in Cognitive Sciences*, vol. 11, no. 10, pp. 428–434, Oct. 2007.
- [18] G. Hinton, L. Deng, D. Yu, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. S. G. Dahl, and B. Kingsbury, “Deep neural networks for acoustic modeling in speech recognition,” *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, Nov. 2012.
- [19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [20] Y. Bengio, “Deep learning of representations: looking forward,” May 2013. arXiv:1305.0445.
- [21] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, “Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations,” in *Proceedings of the 26th Annual International Conference on Machine Learning*, ACM, 2009, pp. 609–616.
- [22] D. Ciresan and U. Meier, “Flexible, high performance convolutional neural networks for image classification,” *Proceedings Of The Twenty-Second International Joint Conference On Artificial Intelligence*, pp. 1237–1242, 2011.

- [23] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, “Large-scale video classification with convolutional neural networks,” *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1725–1732, Jun. 2014.
- [24] J. E. Overland, “Meteorology of the Beaufort Sea,” *Journal of Geophysical Research: Oceans (1978–2012)*, vol. 114, no. C1, 2009.
- [25] G Spreen, L Kaleschke, and G Heygster, “Sea ice remote sensing using AMSR-E 89-GHz channels,” *Journal of Geophysical Research*, vol. 113, no. C2, C02S03+, Feb. 2008.
- [26] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [27] Y.-L. Boureau, J. Ponce, and Y. LeCun, “A theoretical analysis of feature pooling in visual recognition,” in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 111–118.
- [28] D. Scherer, A. Müller, and S. Behnke, “Evaluation of pooling operations in convolutional architectures for object recognition,” in *Artificial Neural Networks–ICANN 2010*, Springer, 2010, pp. 92–101.
- [29] V. Nair and G. E. Hinton, “Rectified linear units improve restricted Boltzmann machines,” in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 807–814.
- [30] Y. LeCun, L Bottou, G. Orr, and K. Müller, “Efficient backprop,” in *Neural Networks: Tricks of the Trade*, 2012.
- [31] L. Prechelt, “Early stopping-but when?,” in *Neural Networks: Tricks of the Trade*, Springer, 2012, pp. 53–67.
- [32] 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 *2014 IEEE Geoscience and Remote Sensing Symposium*, IEEE, Jul. 2014, pp. 3969–3971.
- [33] V. Mnih and G. E. Hinton, “Learning to label aerial images from noisy data,” in *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, 2012, pp. 567–574.
- [34] A. Dalryac, M. Shanahan, and J. Kelly, “Tackling class imbalance with deep convolutional neural networks,” Tech. Rep., 2014.
- [35] M. A. Maloof, “Learning when data sets are imbalanced and when costs are unequal and unknown,” in *Workshop on Learning from Imbalanced Data Sets II, ICML 2003*.

- [36] Z.-H. Zhou and X.-Y. Liu, “Training cost-sensitive neural networks with methods addressing the class imbalance problem,” *Knowledge and Data Engineering, IEEE Transactions on*, vol. 18, no. 1, pp. 63–77, 2006.

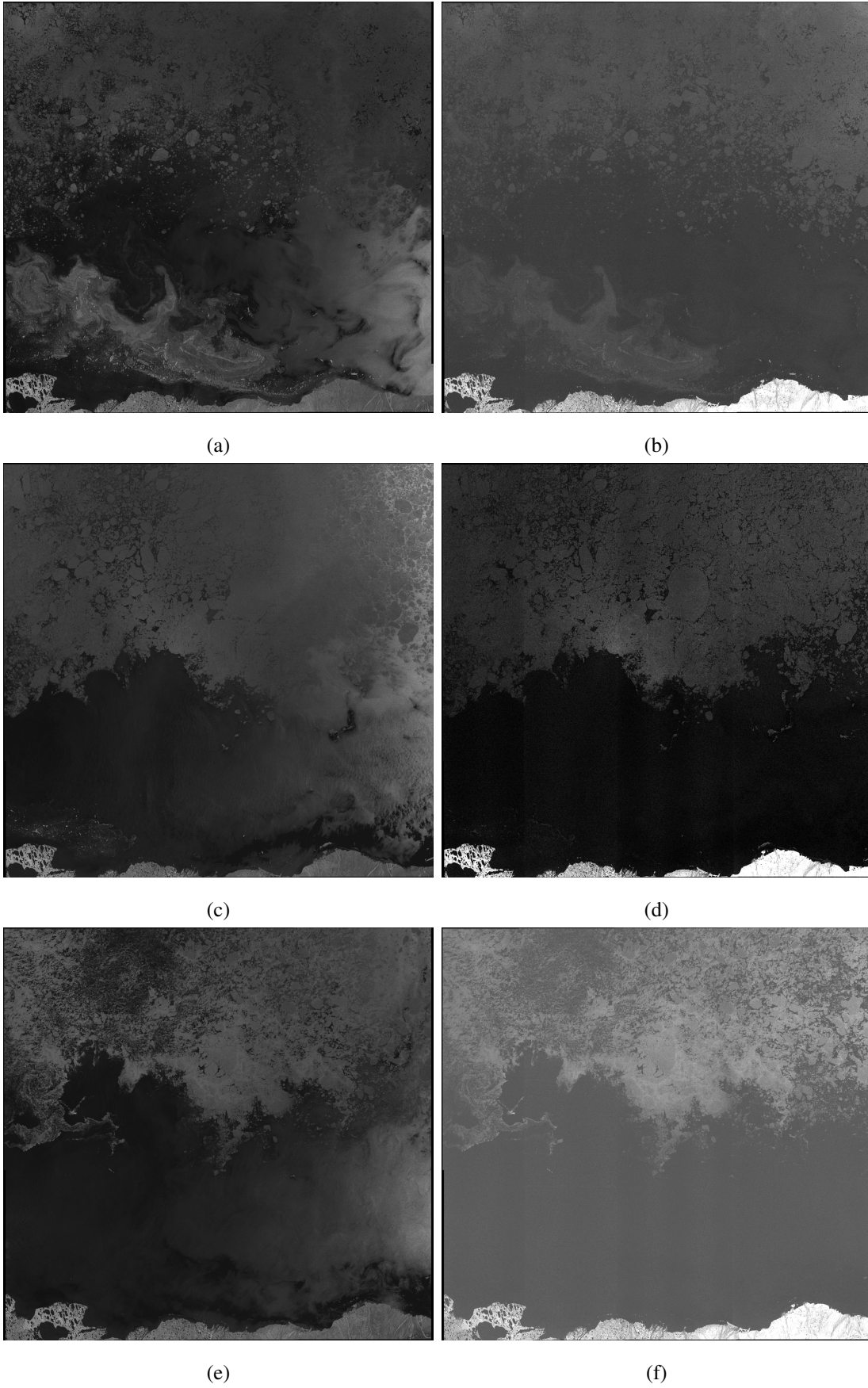


Fig. 6: Selected scenes of the RADARSAT II images used. First row to third row are scenes 20100730, 20110725, 20110811 respectively. The first and second column are the HH and HV bands respectively. The HV images are enhanced for visualization.

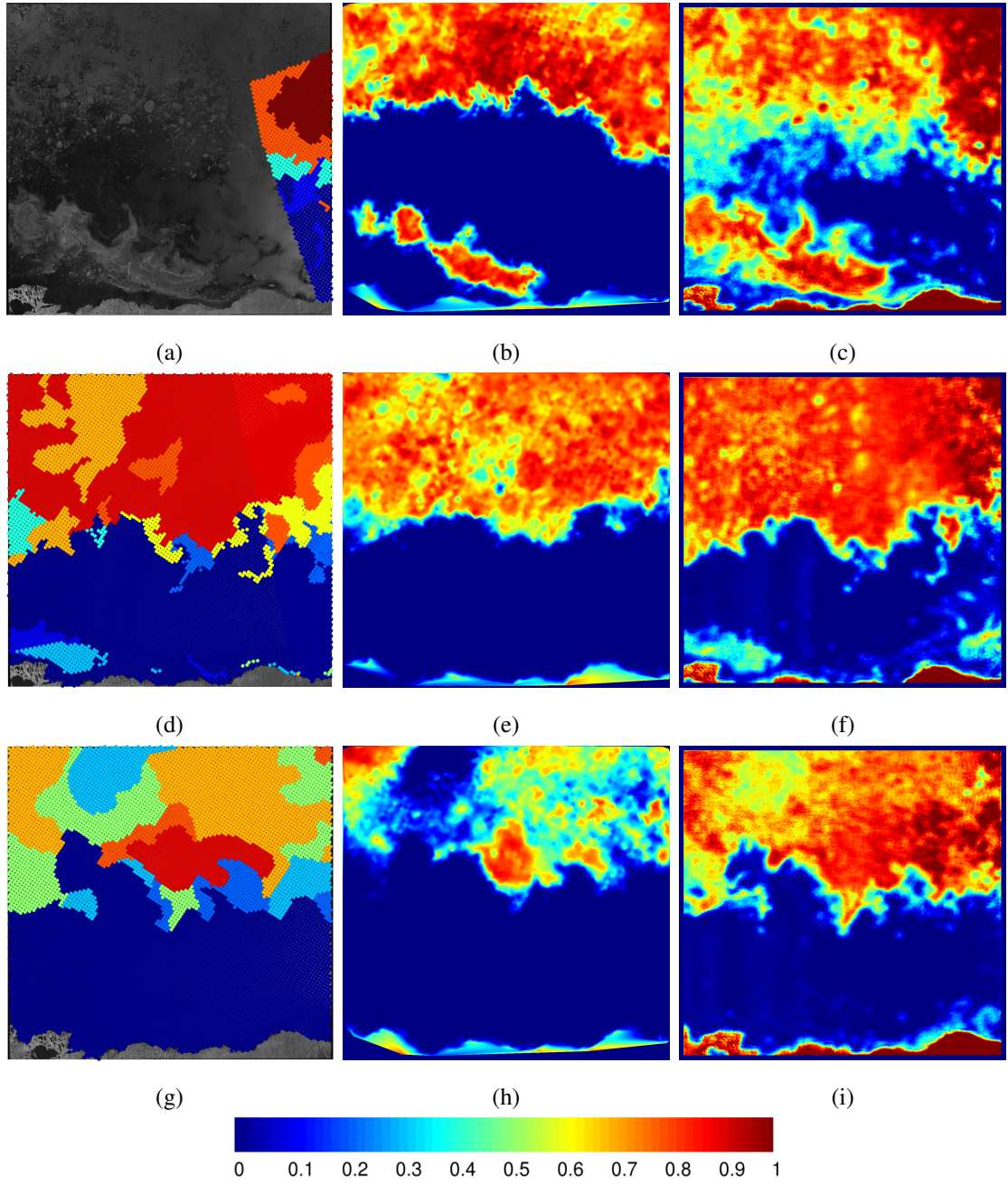


Fig. 7: The ice concentration from image analysis (column 1), ASI ice concentration (column 2) and the estimation of CNN (column 3). The first to third row are scene 20100730, 20110725, 20110811 separately.