Sea Ice SAR Imagery Classification and Regression Based On Convolutional Neural Networks

by

Yan Xu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2018

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Due to the global warming, there have been signficant reductions in the ice extent and ice thickness in the Arctic and marginal seas [57]. Monitoring these changes in sea ice is very important for human activities including weather forecasting, natural-resource extraction, and ship navigation [52].

Of the various sea ice monitoring activities, and sea ice and open water classification, sea ice concentration estimation has attracted significant attention due to the importance of this type of information. Satellite imagery is widely used for monitoring the ice cover. In this regard, images from synthetic aperture radar (SAR) are of interest due to their high spatial resolution. However, automated SAR imagery interpretation is a complex recognition task that requires algorithms with strong ability to learn complex features.

Convolutional neural networks (CNNs) are the state-of-the-art in the image recognition field and CNNs have demonstrated an excellent ability to learn complicated image features. In this thesis, we first used a CNN-based transfer learning method to address sea ice and water classification challenge, which achieves an impressive classification accuracy (92.36%). Then sea ice concentration estimation from SAR image using CNNs is developed. The CNN models are trained from scratch using image analysis charts as ground truth.

Based on the designed CNN, several studies are conducted. We first demonstrate the importance of including samples of intermediate ice concentration in our training data. Then experiments are carried out to increase the number of these samples in our dataset. The results from experiments indicate that model performance can be improved by adding more intermediate ice concentration samples from new datasets, regardless of the location, time, and sea ice features of new datasets. Another benefit of balancing the dataset is that the estimation results of intermediate ice concentrations from the CNN become more accurate. In addition, the CNN model we adopted is found to outperform other algorithms on distinguishing the marginal ice zone.

Acknowledgements

During the last two years, I would like to express my gratitude to my supervisor, Prof. Andrea Scott, who supports me and gives me help whenever I encounter problems. I am also grateful to my supervisor for her patience and support in overcoming numerous obstacles I have been facing through my research.

I am sincerely grateful to my readers: Prof. Clausi and Prof. Orchard for sharing their truthful and illuminating views and suggestions on a number of issues related to my thesis.

I would like to thank my lab mates for helping me solve many problems. In addition, I am going to express my gratitude to everyone who supported me throughout my master program. I am thankful for their aspiring guidance, invaluably constructive criticism and friendly advice during the last two years.

Last but not the least, I would like to thank my family. Thanks for their support and eternal love.

Thank you.

Yan Xu

Dedication

This is dedicated to my family.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Obtaining knowledge of sea ice is important to understand climate change, as well as for weather forecasting and marine operations in ice covered regions. According to recent studies, sea ice cover is currently undergoing significant changes with an observed decline not only in the extent of multiyear sea ice [7], but also an overall reduction in thickness [39]. The rate of these changes has being increasing in recent years [39]. In order to monitor the condition of sea ice and support safe Arctic operations and navigations in ice-infested waters, effective and high resolution information of the ice coverage is desire [66, 57].

Satellite observations such as synthetic aperture radar (SAR) images are widely used for sea ice information retrieval. These images have high spatial resolution and have the ability to see through cloud cover, and are ideal for sea ice mapping. It is challenging to interpret SAR images since the interactions between the SAR signal and sea ice are affected by various factors, including the imaging frequency and incidence angle, and the surface conditions of sea ice [43, 42].

Apart from several physical algorithms that can be applied to retrieve ice information from SAR imagery, machine learning algorithms such as support vector machine (SVM), random forest, and k-nearest neighbour (KNN), combined with hand-crafted features (e.g., geometrical feature [62], Fourier descriptors) also can be adopted to obtain impressive performance on sea ice information retrival. For example, a SVM model trained using SAR image texture features was proposed to identify multiple sea ice types[45].

One approach of machine learning called convolutional neural networks (CNNs) is attracting increasing attention in the remote sensing domain because CNNs are now the state-of-the-art for image related tasks [2]. Due to the availability of large datasets and the use of GPUs, which can alleviate the training inefficiency, and some novel approaches

to reduce over-fitting [75, 38, 67], CNNs have been shown to significantly outperform traditional methods for image classification tasks [2]. Due to the success of CNNs on image classification tasks, CNNs are expected to perform well on SAR imagery classification problem as well [23]. In addition, with continuous advancements and developments in satellite technology, the volume of available synthetic aperture radar (SAR) images is increasing, which increases the need for automated algorithms to analyse these images. The objective of this thesis is to apply CNN-based learning approaches for the analysis of sea ice SAR imagery, providing more effective and timely information about sea ice so as to facilitate human activities.

One challenging task in sea ice remote sensing field is to classify sea ice and open water in satellite imagery accurately [84]. Distinguishing sea ice and open water from SAR imagery is critical for many activities, such as ship navigation, search, and rescue operations. Nowadays, automated methods to distinguish sea ice and water in SAR imagery are in demand. CNN-based approach can be used to solve this problem. In this thesis, a CNN-based transfer learning approach is adopted to classify sea ice and open water from SAR images, image features are extracted from a pre-trained CNN and softmax classifier is adopted to achieve final classification of sea ice and water. This approach achieved 92.36% classification accuracy on average after applying cross-validation.

The other topic of this thesis is sea ice concentration estimation. Currently, SAR images are manually interpreted by human experts working in government agencies to produce operational image analysis charts, which display sea ice concentration information over certain regions [75, 54]. However, preparation of these charts is very time-consuming and this human visual interpretation tends to be biased. It is of interest to find an intelligent and robust method to facilitate the estimation of sea ice concentration from SAR imagery. Although there are numerous methods available, this thesis mainly focuses on sea ice concentration estimation from SAR images by training a CNN model from scratch. However, the ice concentration from this CNN has the same problem as that from Wang [75], which is that the ice concentration in the range of 10% to 80%, is not estimated very accurately.

By looking into the available dataset, the imbalanced dataset problem was discovered and several possible methods to address this problem are proposed. The experimental results show that the efforts of balancing the dataset by introducing minority categories from new datasets do make a difference, even though the sea ice features and SAR image acquisition periods between datasets differ. With the size limitation of available new datasets, the original dataset could not be totally balanced by the new samples introduced by new datasets. However, the results do show an impressive improvement according to many experiments which aim at balancing the original dataset. Moreover, it is found that

the CNN model outperform other methods on identifying marginal ice zone. This finding is supported by the experiments in this study.

The outline of this thesis is as follows: We first give a brief background introduction in chapter 2 of some essential concepts that will be discussed in this thesis. In chapter 3, various data preprocessing methods for most of our experiments are illustrated in detail, and the method of classification of sea ice and water from SAR images is also presented. In chapter 4, sea ice concentration estimation by using CNN is discussed and the imbalanced problem is studied.

Chapter 2

Background

This chapter presents background knowledge related to this thesis. It includes relevant fundamental concepts of remote sensing, such as properties of synthetic aperture radar imagery and sea ice monitoring. In addition, background of convolutional neural networks, and recent research of applying CNN approaches in remote sensing are presented in this chapter. A CNN-based method called transfer learning is introduced as well. Finally, the imbalanced dataset issue is briefly discussed at the end.

2.1 Sea Ice and SAR Introduction

2.1.1 Sea Ice Basics

Monitoring sea ice is crucial for many high latitude countries [63]. Sea ice is a part of the cryosphere, interacting constantly with the underlying ocean and the overlaying atmosphere [9, 63]. Sea ice floats on the ocean's surface because sea ice is less dense than water [9]. Due to the effects of winds, currents and temperature fluctuations, sea ice is very dynamic in the ocean [9]. Therefore, there are numerous sea ice types, for example, deformed and undeformed sea ice, and various sea ice features, such as pressure ridges or cracks in the sea ice cover, are generated [9].

The percentage of sea ice coverage in a given sea area is defined as sea ice concentration, or ice concentration (IC), ranging from 0% to 100%. The definition of 0% ice concentration means there is no ice but water in the given sea area, while 100% ice concentration means ice is covering all the given area [35].

The information of sea ice is important for numerous applications. First of all, sea ice can impede ship traffic and increase the risk of traffic accidents in the Arctic [63, 9]. Sea ice is highly related to climate, it can be considered as a climate indicator. For example, reductions in sea ice extent and thickness are believed to be due to warming climate [63, 9]. A consolidated ice cover can be regarded as an insulator that hinders the exchange of mass, momentum and chemical constituent between the ocean and the atmosphere [63]. Finally, fluxes at the ice surface provide boundary conditions for weather forecasting models.

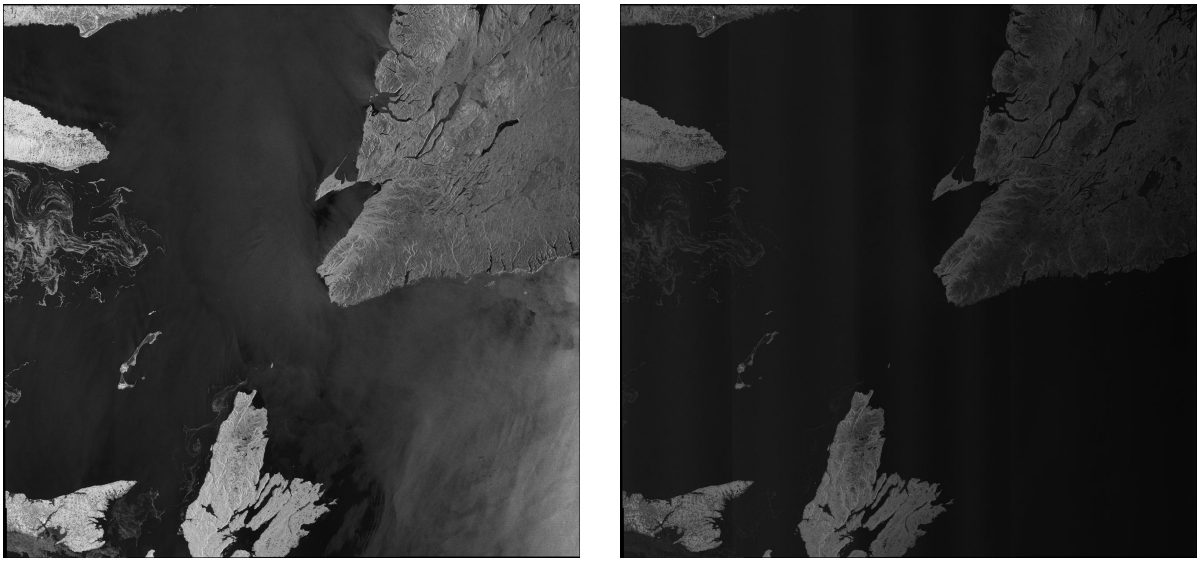
2.1.2 SAR Imagery

Synthetic aperture radar (SAR) is widely used for sea ice monitoring. It is well known for its all-day and all-weather sensing capability [30], because it can capture images of the earth surface even through clouds and in total darkness (without sunlight) [17]. Generally speaking, SAR is a form of radar, which is used to create 2D or 3D images of an object or earth's surface [10]. SAR images are able to represent more details of the ice cover as compared to passive microwave images because of their high spatial resolution [23]. In comparison to the resolution of operational ice concentration products based on passive radiometer data, the resolution of SAR data is significantly higher [35].

Nowadays, the main data source that is used for sea ice monitoring in Canada is RADARSAT-2. RADARSAT-2 has four different polarization configurations for SAR systems:

1. HH: Horizontal polarization on transmit, Horizontal polarization on receive
2. HV: Horizontal polarization on transmit, Vertical polarization on receive
3. VH: Vertical polarization on transmit, Horizontal polarization on receive
4. VV: Vertical polarization on transmit, Vertical polarization on receive

Each polarization configuration corresponds to one SAR image. A dual-pol SAR scene contains two images, one HH and one HV, and covers a large spatial region (approximately 500km x 500km), and it is commonly used for ice charting. In contrast, a quad-pol scene contains four images (HH,HV,VH,VV) and covers a smaller geographic region. Experiments in this thesis are based on dual-pol SAR scenes. For example, a dual-pol SAR scene is shown in Figure 2.1. The image 2.1a is HH polarization image, while the image 2.1b is HV polarization image.



(a) HH (b) HV

Figure 2.1: An example of a dual-pol SAR scene acquired on 2014-01-18 in the Gulf of St. Lawrence.

The ice analysis charts produced by Canadian Ice Service have relied heavily on RADARSAT-2 since it was launched in 2007 [37, 49]. The ongoing satellite misson RADASAT-2 in Canada is presently considered as the most reliable source of information on the condition of sea ice [37].

2.1.3 Sea Ice Monitoring

Many approaches have been developed to monitor sea ice conditions using satellite data. The three most popular data sources are optical sensors, passive microwave sensors, and SAR [27].

Optical sensors have the benefit of relatively fine spatial resolution (1km) but also some shortcomings, image capture conditions are limited by cloud coverage and sun illumination over the Arctic during winter [37]. Thus, the data from optical sensors cannot satisfy the requirements of ice operations because optical sensors only provide data under cloud-free conditions. Other data sources used to monitor sea ice, passive microwave and scatterometer observations, are not significantly restricted by clouds and sunlight [37]. The strong contrast of the emissivity of sea ice and water makes water and different sea ice types easily

distinguished in passive microwave images [3, 75]. However, the low resolution of passive microwave data makes them barely applicable in the vicinity of a shoreline and within narrow channels [37].

Compared to the two data sources discussed above, high-resolution synthetic aperture radar images (nominal pixel spacing less than or equal to 50m) [37] from RADARSAT have been considered as the main data source for ice monitoring in several countries, particularly in Canada [68], who uses large amounts of wide-swath SAR images to monitor Canadian ice regions [63]. This thesis mainly focuses on using SAR imagery to retrieve sea ice information.

2.1.4 Sea Ice Features of The Gulf of St. Lawrence

The main study area of this thesis located in the Gulf of St. Lawrence (GSL), which is a semienclosed sea along the east coast of Canada [74]. A picture of this area is shown in Figure 2.2, most of SAR images used for the experiments of this thesis were captured from Gulf of St. Lawrence in the period of January 16, 2014 to February 10, 2014. This time period corresponds to the freeze up stage of sea ice in the Gulf of St. Lawrence.

The sea ice in GSL has strong seasonality and dynamic motion [74]. A typical timeline of sea ice formation in the Gulf of St. Lawrence is as follows: new ice begins to form in the upper of GSL near the end of December, and the ice cover augments in concentration at the beginning of January. Sea ice occurs first in shallow areas at the very begining of the ice season and then spreads eastward [74]. Ice covers almost the entire Gulf of St.Lawrence by the third week of February [16]. Sea ice in the GSL reaches its maximum stage of formation by the third week of March, after that, melting and dispersal procedure begin [74].

Land-fast ice has limited extent in the Gulf of St. Lawrence, and ice moves from shore into the main part of the Gulf of St. Lawrence due to wind and water motion [16]. As a result, floes and some new ice are generated in the central region of the gulf. Wind affects the motion of sea ice significantly in the northeastern of the Gulf of St. Lawrence [16]. Sometimes, very large floes will occur in this area [16]. Sea ice is very mobile in the Gulf of St Lawrence, the size of floes is commomly small [16]. Ridging can be rather extensive in some areas, but not developed to any great height.

Figure 2.2: East cost of Canada with the Gulf of Saint Lawrence shown. (Source: [8])

2.2 Image Analysis Charts

To provide ice information to customers, the Canadian Ice Service (CIS) relies heavily on SAR imagery [15]. These images will be analysed to create image analysis charts, which are ice maps produced by ice experts from CIS through visual interpretation of SAR imagery [43, 42]. The zone analysized by the image analysis concurs with an area relating to the satellite path. Therefore, the satellite's path varies each day thus the chart area shifts from day to day [15]. According to the official website of Environment and Climate Change Canada (ECCC), around 11,000 SAR images from RADARSAT-2 are received per year [15]. Sea ice experts take approximately 4 hours to analyse a SAR scene [15]. In most cases, analysis results are issued in the form of image analysis charts [15].

To produce image analysis charts, expert analysts visually identify regions in the image that have homogeneous ice conditions and assign an ice concentration label to that region. These labels are assigned in tenths (e.g., 10%, 20% ...), and therefore the precision of this ice concentration is normally considered to be approximately 10%.

When analysing the satellite imagery, expert analysts are not only using their abundant

experience of visually looking at SAR images, but also their experience of the ice conditions in the region. Analyzing images is particularly challenging if seas are rough, or sometimes if water is on the ice, accumulated either from melt or rainfall [15]. Although image analysis charts are normally considered reliable, there are several source of uncertainty and errors in the image analysis charts because they are based on human interpretation [19]. In addition, the process of generating image analysis charts is time-consuming, that is the reason why automating the process of ice information retrieval is significant [75].

2.3 Convolutional Neural Networks

Convolutional neural networks (CNNs) models have recently been applied to large-scale visual recognition tasks. These networks are trained via backpropagation through layers of convolutional filters [41, 14, 67]. With large quantities of training data, CNN models outperform other models, and had early success in digit classification tasks. Nowadays, many technologies related to CNNs are used in our daily life, for instance, face recognition [51] and machine translation [28]. CNNs are an active area of research and have achieved appealing progresses in many fields, such as medical imaging and automobile industry. However, CNNs are still not widely used in remote sensing, there are still substantial problems that are worth exploring [22]. In this section, we will provide a brief introduction about CNNs, and the algorithm of training a CNN.

2.3.1 Convolutional Neural Networks Basics

CNNs are fundamentally the same as ordinary neural networks, which are comprised of neurons that have weights and biases to be learned [32]. A CNN consists of multiple layers, which are typically convolutional layers, pooling layers, fully-connected layers, and non-linear layers. The arrangement of these layers determines the architecture of a CNN. Normally, each neuron receives some inputs, performs a dot product between the inputs and the filter weights, followed by a non-linearity, which is called the activation function [31]. The most popular activations functions are Sigmoid, Tanh, and ReLU [31]. The convolutional layers are typically followed by fully-connected layers. For a classification problem, these layers provide inputs to a softmax function that computes a probability, and finally a loss function, such as cross-entropy, is used to compare the network outputs to the target outputs [32].

CNN feature learning algorithms have achieved superior performance over hand-designed features in many benchmark data sets [38]. The most salient characteristics of a CNN are

local connections and parameter sharing. These two characteritics lead to the main advantage of CNNs, which means that the computational cost to train a CNN is reasonable as there are fewer parameters than fully connected networks with the identical number of hidden layers [31].

In recent years, more and more complex CNN models have been developed, for example, AlexNet [38], VGG-16-Net [67], GoogLeNet [71], and the latest one, Res-Net [21], all of them are the champions of ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) respectively in the passing years. Generally speaking, the deeper and wider of a CNN structure, the better the performance it will achieve. However, memory and the computational resources are a challenge when training these CNNs. In addition, to avoid overfitting problem, a considerably large dataset is in demand for deeper and wider CNNs.

2.3.2 The Architecture of Convolutional Neural Networks

A typical CNN consists of a number of convolutional layers, pooling layers optionally followed by several fully-connected layers [58]. As the convolution operation is in charge of the most heavy computation part of a CNN, convolutional layers are regarded as the most important components [31]. The function of convolutional layer is to compute the output of neurons that are connected to local regions in the input, each computing a dot product between their weights and a small region they are connected to in the input volume [58]. The pooling layer basically performs a down sampling operation, while the category scores are computed after the fully-connected layer, where fully-conected means that each neuron in this layer and in the previous layer is connected to each neural in the previous layer. A typical CNN structure is shown in Figure 2.3.

2.3.3 Optimization For Convolutional Neural Networks

Neural network training is one of the most difficult optimization problems [18]. Gradient descent is the most commonly used modern optimization algorithm, which follows the gradient of the entire training set downhill. Gradient descent may be accelerated significantly by using stochastic gradient descent with fixed learning rate [18]. Instead of using fixed learning rate, numerous algorithms have been proposed that can adapt the leaning rate recently during training [18]. For example, AdaGrad [36], RMSProp [73].

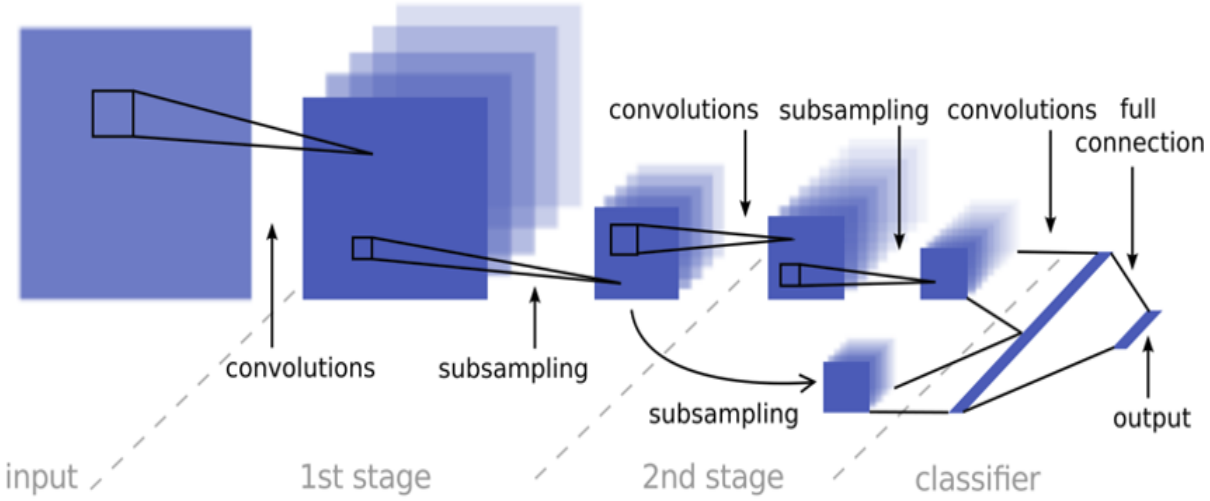


Figure 2.3: The architecture of a typical CNN. Source:[40]

2.3.4 Model Training

The algorithm that is commonly used to train CNN models is stochastic gradient descent (SGD), which uses backpropagation to compute its gradients [61]. Backpropagation is the most popular algorithm for multi-layer networks, the core of this algorithm is to apply the chain rule to compute the effect of each weight in the network with respect to error function [61]. This algorithm can be applied to effectively compute the gradients for SGD algorithm.

The input samples first propogate through CNN, and then the derivatives of the loss with respect to the parameters throughout the network are computed by backpropagation algorithm. Regarding the optimization method, the weights are updated through mini-batch stochastic gradient descent (SGD) [85]. The objective of SGD algorithm is to minimize the loss function so as to find the optimal parameters in the convolutional network model with a good generalization ability for the target problem. First of all, the general form of cost function $J(\theta)$ is defined as follows [18],

$$J(\theta) = L(f(x;\theta), y) = \frac{1}{m}\sum_{i=1}^{m} L(f(x^{(i)};\theta), y^{(i)}))).$$

L is the loss function, and $f(x;\theta)$ is the output from a model with input x and parameter

set θ, where m is the number of samples in the mini-batch, y denotes the desired output or label for each sample.

L2 loss (Euclidean loss) is typically used for CNN-based image regression tasks [77]. In this study, an Euclidean loss layer is adopted at the last layer of the network. Euclidean loss with L2 regularization term has the form [18],

$$\mathbb{L}_{f(x;\theta),y} = \frac{1}{2m}\sum_{i=0}^{m}(f(x^{(i)};\theta) - y^{(i)})^2 + \lambda\sum_{i=1}^{m}\theta_i^2.$$

This function is otherwise called the "squared error function", or "mean squared error". The mean is halved for the sake of a convenience for the computation of the gradient descent, as the derivative term of the square function will cancel out the 1/2 term. In order to update weights, the gradient of the loss with respect to the weights is needed [18].

$$\hat{g} = \frac{\partial \mathbb{L}_{f(x;\theta),y}}{\partial\theta} = \frac{1}{m}\sum_{i=0}^{m}(f(x^{(i)};\theta) - y^{(i)})\frac{\partial f(x^{(i)};\theta)}{\partial\theta} + 2\lambda\theta_i.$$

where $\hat{g}$ denotes the gradient calculated based on m samples randomly sampled from the training data. The calculated gradient will be used for mini-batch SGD algorithm, updating θ in the direction of $\hat{g}$, a mini-batch SGD algorithm is based on the following steps [18].
Step 1: Randomly initiate weights θ and set values for hyperparameters.
Step 2: Iterate the followings until stop criteria is met.
Sampling a minibatch m sample $\{x^1, x^2, ...x^m\}$ from training set and the corresponding labels $y^{(i)}$

$$\text{calculate the gradient:} \quad \hat{g} := \frac{1}{m}\bigtriangledown_\theta\sum_i L(f(x^{(i)};\theta), y^{(i)}))$$

$$\text{calculate velocity update:} \quad v := \alpha v - \epsilon\hat{g}$$

$$\text{update each weight:} \quad \theta := \theta + v.$$

where ϵ denotes learning rate, we use v denotes velocity and α denotes momentum parameter. The symbol := means to update the parameter on the left side of the symbol by the computed value of the equation on the right side of the symbol. The symbol $\bigtriangledown_\theta$ denotes the gradient with respect to parameter θ

2.4 Transfer Learning

Training an entire deep CNN from scratch, which means initializing the weights at random, is often impractical in practice, because it is not feasible to obtain a large number of training samples. People normally collect limited datasets with far less than the total parameters in a deep CNN [33], which has millions of parameters that need be determined [23]. However, in particular, there is no large-scale labelled SAR imagery dataset that is widely used to test deep learning algorithms. To balance this trade-off, transfer learning is expected to play a significant role in the case of deep CNNs for tasks using SAR imagery. Transfer learning is an effective way of training a large CNN without overfitting when the dataset is insufficient [23].

The main concept of transfer learning is as follows. First, we need to train a base network on a relatively huge dataset, for example, ImageNet, which contains about 22,000 classes and nearly 15 million labeled images. Second, we reuse the learned weights or features of the pretrained network in a new network that will be used for our dataset (target dataset) [83]. This process will be successful as long as general features are learned well from the original huge base dataset, and the new target database also contains these kind of general features [83]. In practice, numerous studies have implemented this transfer learning method for various tasks, with results that are comparable to those from conventional methods.

There is an interesting common phenomenon that has been observed from many deep CNNs trained on natural images. Typically, simple features, such as Gabor features, edges, and color blobs are learned at the first few layers [33, 83]. Such features are not specific to a particular dataset, they are generic features that will be useful for a variety of image classification and recognition tasks. While this fact is heuristic, it does provide a fundamental theoretical support for applying transfer learning method. However, features detected from the later layers of the neural network are more specific to the base dataset [33]. Therefore, we only need to modify the last few layers to fit target tasks. For most CNNs trained for image classification, we know that the features computed by the network's last layer must depend greatly on the chosen specific dataset, for example, if a network that has been trained with softmax layer as the last layer, then each output unit will be specific to a particular category [33, 83]. This means that if you have 10 classes, then there will be 10 corresponding outputs.

In summary, transfer learning can be summaried as follows: First, to pretrain a CNN on a large dataset, and then use the weights of pretrained CNN either as an initialization of weights or regarding the entire CNN as a fixed feature extractor for new datasets or tasks

[33]. There are two commonly used transfer learning scenarios which will be presented in the following subsections in detail [33].

2.4.1 Scenario I: Feature Extractor

The basic idea of the first transfer learning scenario is to treat a pre-trained CNN as a fixed feature extractor and retain the classifier [33]. To be more specific, a CNN pretrained on ImageNet is taken, and then removing the last fully-connected layer. After that, we can treat the rest of the CNN as a fixed feature extractor for the target dataset [33]. If we take an AlexNet as an example, this operation will return a 4096-dimension feature vector for every image that is fed through the AlexNet [33]. This feature vector is then used to train a linear classifier, for instance, Linear SVM, KNN, or softmax classifier for the target dataset [33].

2.4.2 Scenario II: Fine-tuning

In addition to the method discussed above, there is another transfer learning method that is popular, which is copying weights from the first several layers of a trained base network to the first several layers of a new network [83]. The remaining layers of the target network are trained using the new dataset with random weights initialization [83]. During training process, there are two options: one options is to backpropagate the errors from the new task into all the layers including the first several layers to fine-tune the copied weights. The other option is keeping copied weights of the first several layers fixed, which means the weights for these layer will remain unchanged during training [83]. As for the remaining several last layers, they will be retrained to be more specific to the new task. The decision whether or not to retrain the first layers not only depends on the size of the new dataset, but also relies on the number of parameters need to be trained in the first several layers [83].

Keeping a portion of the prior layers frozen and fine-tuning the rest of the network is appealing due to overfitting concerns [33]. Overfitting can occur when the new dataset is small and the number of parameters need to be trained is relatively large. In this case, it is better to keep the weights of the majority of the layers unchanged [83]. However, overfitting will not be a problem when the new dataset is sufficient, and in this case tweaking weights of a few layers of the network can boost the performance on the new dataset [83]. In an extreme case, there is no need to do transfer learning if you are confident that the size of the target dataset is sufficient, i.e, datasets that weights are being transferred to is

big enough. That is because learning from scratch on the large target dataset will give satisfying results without overfitting concerns [83]. However, The computing time of this method is much longer than that of transfer learning.

2.5 Previous Work Relevant To This Thesis

2.5.1 Research on CNNs In Remote Sensing Applications

In recent years, there is a large and growing volume of published studies describing the role of CNNs in different fields. Implementing CNNs algorithms is becoming more popular in the remote sensing community. Studies have shown that CNNs approaches outperform conventional methods in a wide range of tasks, for instance, SAR image depeckling, remote sensing image target recognition, classification, edge detection and so on. In this section, I give a literature review of previous research of applying CNNs approaches in the domain of remote sensing.

The studies of Otávio AB et al. [59] demonstrate that CNNs not only can be used to recognize natural objects, but trained CNN models generalizes well on the task of classifying remote sensing scenes. A CNN-based approach for detecting edges in remote sensing images is proposed [81], and found that their approach can detect the continuous edges of remote sensing images and has strong robustness compared with classic algorithms. Jun et al. [13]explicitly explored the ability of a deep CNN with data augmentation on SAR target recognition.

Kampffmeyer et al. [29] proposed a deep CNN for semantic segmantation of urban remote sensing images. The results show that they obtain an overall classification accuracy of 87%, this score is higher than the state-of-the-art for the dataset they used. Maggiori et al [48] proposed a four stacked convolutional layers for extracting features and a deconvolutional layer as the last layer. This convolutional network is designed for pixelwise labeling of remote sensing imagery. Their experimental results surpass previous approaches not only in terms of efficiency, but also in terms of accuracy [48].

CNNs are found to be able to deal with SAR image depeckling problems as well. An eight convolutional layer network with a method that similar to residual learning is designed in [6] for SAR image depeckling. The CNN model they proposed was found to outperform three other image despeckling algorithms. In addition, CNNs are also found to be able to tackle the problem of pansharpening remote sensing images. An example can be found in [50].

One approach that is particularly related to the studies in chapter 3 is the study described in [44]. They have shown a CNN model can be successfully used on sea ice and water classification task. They directly use polygon-wise ice concentration of image analysis chart, classifying sea ice and open water from SAR imagery based an Alter-CNN algorithm [44]. They obtain impressive results in comparison with other algorithms.

Another study that is similar to our studies presented in chapter 4 is the study proposed by Wang [75]. A CNN structure and a fully-convolutional neural network (FCNN) structure taking SAR images and incidence angle as input are developed to estimate sea ice concentration for SAR imagery. Both the FCNN and the CNN models are able to obtain good results even for very thin new ice, which is difficult to distinguish visually in the imagery [75]. The root mean squared errors are 0.22 and 0.21 for CNN and FCNN respectively. The CNN models are found to be able to generate reasonable ice concentration in the melt season when evaluating the performance on a Beaufort Sea dataset, and achieved a state-of-the-art ice concentration estimation [75]. Wang also evaluate the transferability of CNN models and found CNN models are transferable between different SAR image datasets [75]. Wang's works in [75] is the first study in which a CNN is applied to estimate sea ice concentration from SAR images.

2.5.2 Research On Transfer Learning In Remote Sensing

The main approach presently used for transfer learning in remote sensing is using pretrained CNNs as feature extractors. Morgan et al. [55] explored how a deep CNN approach performs on the classification of SAR imagery, and they conclude that a CNN method can be adapted quickly to new SAR imagery classification tasks [55]. One interesting transfer learning based method was proposed by Huang et al. [23]. They transferred from a tacked convolutional auto-encoders model trained on unlabelled sufficient SAR images to a labelled new dataset with limited number of training SAR images [23]. They report that their method is competitive with the state-of-the-art method on their dataset, and their method performs even better if the size of training sample is small compared to other methods [23].

Kang et al [30] combined an algorithm for SAR ship detection based on transfer learning method. They extracted features of SAR images from VGG [67], they discover that combining pixels and deep feature can boost the performance of multiscale ship detection in SAR images [30].

The method proposed by Castelluccio et al. [4] is similar to the method presented in chapter 3, which is the classification of sea ice and open water from SAR imagery,

Castelluccio et al. adopted CaffeNet and GoogLeNet, finetuning these pre-trained CNN for semantic classification of remote sensing scenes. In chapter 3, transfer learning approach will be illustrated more specifically.

2.6 Research on Imbalanced Datasets

A dataset is regarded as imbalanced when there are significant differences among the number of samples available for each category [5]. An example of an imbalanced dataset is sea ice SAR imagery datasets, which often contain more samples of open water and ice at high concentrations than samples at intermediate ice concentrations.

Obtaining a balanced dataset at the very beginning is not always possible because real-world datasets are more or less imbalanced. If a dataset is imbalanced, the minority categories are not well represented and models will focus more on majority categories during model training. Thus, the prediction ability for these minority categories is weaker compared to majority categories [47].

Coping with imbalanced dataset problem is important in many applications, especially when the primary interest is on the minority categories, such as detection of fraud calls, and filtering fraud transaction in banking operations [60]. Examples for sea ice would be detecting an isolated ice floe, or an opening in a consolidated ice cover.

There are three commonly used methods that can deal with problems caused by imbalanced datasets. The first two methods are based on sampling method. Intuitively, as our goal is to balance a dataset, sampling methods can help to tackle this kind of problem. One way is oversampling samples in the minority categories, the intent is to increase the number of samples in the minority categories until the number of samples in each category is almost equal before applying machine learning algorithms. However, when applying oversampling algorithm, we cannot just repeatedly get samples from minority categories, because it can lead to overfitting. One classical oversampling method for imbalanced dataset is called SMOTE proposed in [5], the minority categories are oversampled by creating "synthetic" examples rather than by over-sampling with replacement [5]. A"synthetic" example is generated according to the formula,

$$x_{i1} = x_i + \zeta(x_{nn} - x_i).$$

where x_{i1} is the newly generated"synthetic" sample, x_i is the original sample, x_{nn} the nearest neighbour of x_i, ζ is a number between 0 and 1.

The second method is undersampling samples in the majority categories [87], i.e, randomly removing several samples from majority categories until the number of majority categories is almost as small as the number of samples in the minority categories before applying learning algorithms [87]. Chawla et al. [5] conclude that undersampling method enables better classifier than oversampling method. On the other hand, if we abandon samples from majority categories at random, then important information can be ommited [87, 47].

The third method that can be adopted for addressing imbalanced dataset problem is based on modifing the learning algorithms. For example, support vector machine combined with boosting or ensemble method can perform better than learning from the original imbalanced dataset [47]. Another widely used approach is called cost-sensitive learning [46], which addresses the imbalanced dataset problem through incurring different costs for different classes. One typical approach of cost-sensitive learning is to move the decision thresholds according to their corresponding misclassification costs. To explain this approach in a simple way, assume the imbalanced dataset consists of two classes, a negative category and a positive class. p denotes the probability estimation results is positive category from a binary classifier, the decision is based on a threshold,

$$\text{if } \frac{p}{1-p} > 1; \text{positive class} \tag{1}$$

$$\text{else; negative class.}$$

However, if the number of positive samples is different from the number of negative samples in the training dataset, then the threshold should be as follows:

$$\text{if } \frac{p}{1-p} > \frac{m^+}{m^-}; \text{positive class} \tag{2}$$

$$\text{else; negative class}$$

where m^+ denotes the number of positive samples in the training set and m^- is the number of negative samples in the training set.

However, if our classifier is based on (1), then we need to adjust the predicted probability p when using(1) to make decision, the formula is,

$$\text{if } \frac{p}{1-p} \times \frac{m^-}{m^+} > 1; \text{positive class} \tag{3}$$

$$\text{else; negative class.}$$

This is just a basic idea, we can sort of adjust the predicted probabilities when making decisions. However, it is way more complicated in reality; We need to embed something similar to (3) to decision process [87].

Chapter 3

Sea Ice And Water Classification Of SAR Imagery Using Transfer Learning: Scenario I

Accurate and robust classification methods of sea ice and open water are in demand by ice services worldwide [75, 82]. There are many methods that can be used to indentify ice and water from SAR imagery. CNNs are becoming increasingly popular for image related tasks in many research communities due to availability of large image datasets and high-performance computing systems [38, 67, 82]. As CNNs have achieved great success on many image classification tasks, it is reasonable to investigate if CNNs can be used for the classification of image patches from synthetic aperture radar (SAR) imagery into ice and water [82]. In this chapter, several data preprocessing methods, for instance, image analysis charts interpolation, the generation of patches, are discussed in detail. Furthermore, Scenario I of transfer learning has been implemented by extracting features of the patches from AlexNet and training a softmax classifier. This method achieves an overall classification accuracy of 96.69% based on the held-out test data, and obtained 92.36% classification accuracy after applying leave-one-out evaluation. The objectives of the study in this chapter are as follows:

1. To preprocess orginal data into a form suitable for application of CNN.
2. To develop an automated algorithm for ice and water classification from RADARSAT-2 ScanSAR dual-polarization images by using transfer learning scenario I: feature extractor.

3. To demonstrate the exact procedure of implementing transfer learning to SAR imagery classification tasks.

4. To verify and discuss the developed CNN-based transfer learning approach against the ground truth, i.e., image analysis charts from Canadian ice service.

5. To evaluate the CNN model by using holdout method and leave-one-out method.

3.1 Dataset

3.1.1 SAR Scenes

The SAR image dataset used for this study consists of 25 SAR scenes, all of them are dual-polarization SAR images [82]. Each scene consists of HH (horizontal transmit polarization, vertical receive polarization), and an HV (horizontal transmit polarization, horizontal receive polarization) image.

The SAR images were acquired in ScanSAR wide mode of RADARSAT-2, with swath width of 500km and nominal pixel spacing of 50m [82]. The available SAR images were captured from Gulf of St. Lawrence in the period of January 16, 2014 to February 10, 2014, which corresponds to freeze-up. Ice concentration and thickness are increasing from January to February [82]. From the image analysis it is known that the ice cover consists mainly of thin, new, gray and gray-white ice, with thickness less than 30cm [16, 82], and becomes looser approximately 20-50 km from the ice edge [76, 82]. The details of the 25 SAR scenes are shown in Table 3.1. Note that some SAR scenes contain mostly land.

Table 3.1: Detailed information of the GSL2014 dataset

ID	Scene ID	Date Acquired	The number of Image Patches(Stride=20)
1	20140116_223042	2014-01-16	66
2	20140117_103914	2014-01-17	579
3	20140118_101002	2014-01-18	2320
4	20140120_105149	2014-01-20	120
5	20140121_214420	2014-01-21	1704
6	20140122_095247	2014-01-22	1332
7	20140123_222627	2014-01-23	146
8	20140124_103501	2014-01-24	946
9	20140124_215646	2014-01-24	2375
10	20140125_100500	2014-01-25	1130
11	20140126_223850	2014-01-26	0
12	20140127_104734	2014-01-27	227
13	20140127_221027	2014-01-27	79
14	20140128_101751	2014-01-28	1572
15	20140130_110029	2014-01-30	7
16	20140130_222234	2014-01-30	257
17	20140131_103053	2014-01-31	1327
18	20140131_215240	2014-01-31	2428
19	20140206_221744	2014-02-06	598
20	20140207_102631	2014-02-07	1705
21	20140207_214938	2014-02-07	86
22	20140208_095758	2014-02-08	2204
23	20140209_223030	2014-02-09	60
24	20140210_103911	2014-02-10	380
25	20140210_220111	2014-02-10	1893

3.2 Data Preprocessing

3.2.1 Incidence Angle Data Processing

It is known that the backscatter in a SAR image is sensitive to the incidence angle of the acquisition [1]. The sensitivity to incidence angle is dependent on wind and associated

ocean roughness conditions [1]. At steep incidence angles during wind roughened conditions, the bright signature seen in the HH channel often makes ice versus open water discrimination difficult [1]. Overall, steep incidence angles are preferred to maximize new and thin ice separability in the HH channel and act as a complement to ice versus open water separation in the HVchannel [1].

The incidence angle data interpolation method in this study is the same as that described in [75]. The incidence angle data are stored as incidence angle images [75]. These images can be used as the third channel of a image patch (image patches will be explained in the next subsection). Finally, these incidence angle images are rescaled to be in the same data range of SAR imagery.

3.2.2 SAR imagery Processing

SAR images are corrupted with significant multiplicative speckle noise due to the coherent nature of the imaging process [82]. In order to eliminate the influence of noise and reduce the data volume, all the SAR images are sub-sampled by 8 x 8 block averaging [82]. The input of the CNN model used requires 3 dimensional data, such as image data with three color channels (RGB). For this study, the HH polarization image, HV polarization image, and incidence angle information for each pixel are combined together to generate the 3D images. An example is shown in Figure 3.1.

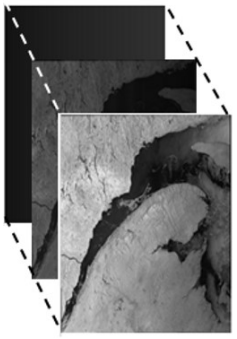


Figure 3.1: Example of a 3D image used for extracting patches. This image was acquired on January-17 2014. Source[82]

3.2.3 Extracting Image Patches

Image patches, with size of 45 pixels × 45 pixels × 3, are extracted from the downsampled SAR image scenes and incidence angle images [75, 82]. A patch size of 45 pixels × 45 pixels is chosen because experiments conducted in [75, 82] found the patch size of 45 pixels × 45 pixels is the best option of achieving better performance for the task of ice concentration estimation using a CNN.

The method for extracting patches is to extract one patch every 20 pixels (e.g., with a stride of 20) from HH, HV, and incidence angle image respectively and record the central point ice concentration value of the corresponding image analysis chart [82]. The label of a patch is then defined by thresholding the ice concentration. Points with ice concentration greater than 30% are considered to be ice, while points with ice concentration less than and equal to 30% are considered to be water [82]. 23541 patches in total were obtained after discarding land contaminated patches, which are patches containing land pixels [82]. The number of patches extracted from different SAR scene is shown in Table 3.1. Several image patches are shown in Figure 3.2. Note that the images shown in Figure 3.2 are not the original patches. These patches are processed in order to observe features. The image patches in the top row are water patches, while the image patches in the bottom row are ice patches. It can be seen that the water patches are visually smoother than the ice patches, and that in some cases it can be difficult to distinguish between the ice and water patches visually [82]. This motivates the use of an intelligent model.

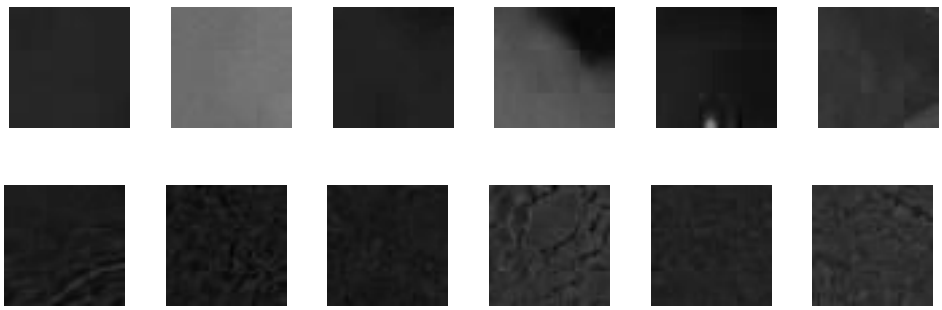


Figure 3.2: Samples of patches with size 45 × 45 × 3: Each patch has an associated label from the image analysis chart. The top row are water patches, while the bottom row are ice patches.

3.2.4 Image Analysis Charts Processing

Image analysis charts from CIS is used as the ground truth. Background materials pertaining to image analysis charts have been presented in Section 2.2. Since image analysis charts contain information about sea ice concentration, while the experiments in this chapter aim to classify sea ice and open water, the image analysis charts are modified by thresholding ice concentration at 30% [82].

Note that there will not be a corresponding ice concentration value (ground truth) from the image analysis chart for every ocean pixel of a specific HH image, because the spatial resolution of the SAR image is much finer than the image analysis chart. Therefore, interpolation of the image analysis ice concentration to the pixel locations in the SAR imagery is necessary. To deal with this problem, nearest neighbour interpolation method is adopted, which interpolates ice concentration from image analysis charts for every ocean pixels based on closest latitude and longitude of original ice analysis charts. However, this also introduces the assumption that the image analysis can be interpreted as valid at this finer spatial resolution.

3.3 Feature Extraction and Classification Method

In this study, instead of building a CNN from scratch, a CNN that has been constructed in an earlier study (AlexNet [38, 82]) is used in a transfer learning approach [83]to classify ice and water in SAR imagery. AlexNet was designed and trained for the task of labelling images from the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) database. These are images of everyday entities, such as a dog, cat or car, and are significantly different from SAR sea ice imagery [82]. Nevertheless, the experimental results support the use of such an approach for this task [82].

The experiments in this chapter start with the bundled CaffeNet model, which is based on the network architecture of Krizhevsky et al. [38] for ImageNet called AlexNet. From this, the generalization ability of the network is explained when retraining the softmax classifier on the top [82]. To do this, a transfer learning method is adopted. The labelled image patches are fed through AlexNet, which is used as a fixed feature extractor [82]. Features from the sixth fully-connected layer (FC6) of AlexNet are used in a softmax classifier that was retrained to estimate the probability of ice for the center pixel of the patches [82]. In this way, sea ice and open water can be classified [82]. The classification accuracy can be determined by comparing the estimation results to the labels of the testing

set [82]. The main method is shown in a flowchart in Figure 3.3. In short, this method basically consists of three steps: preprocessing, feature extraction, and prediction [82].

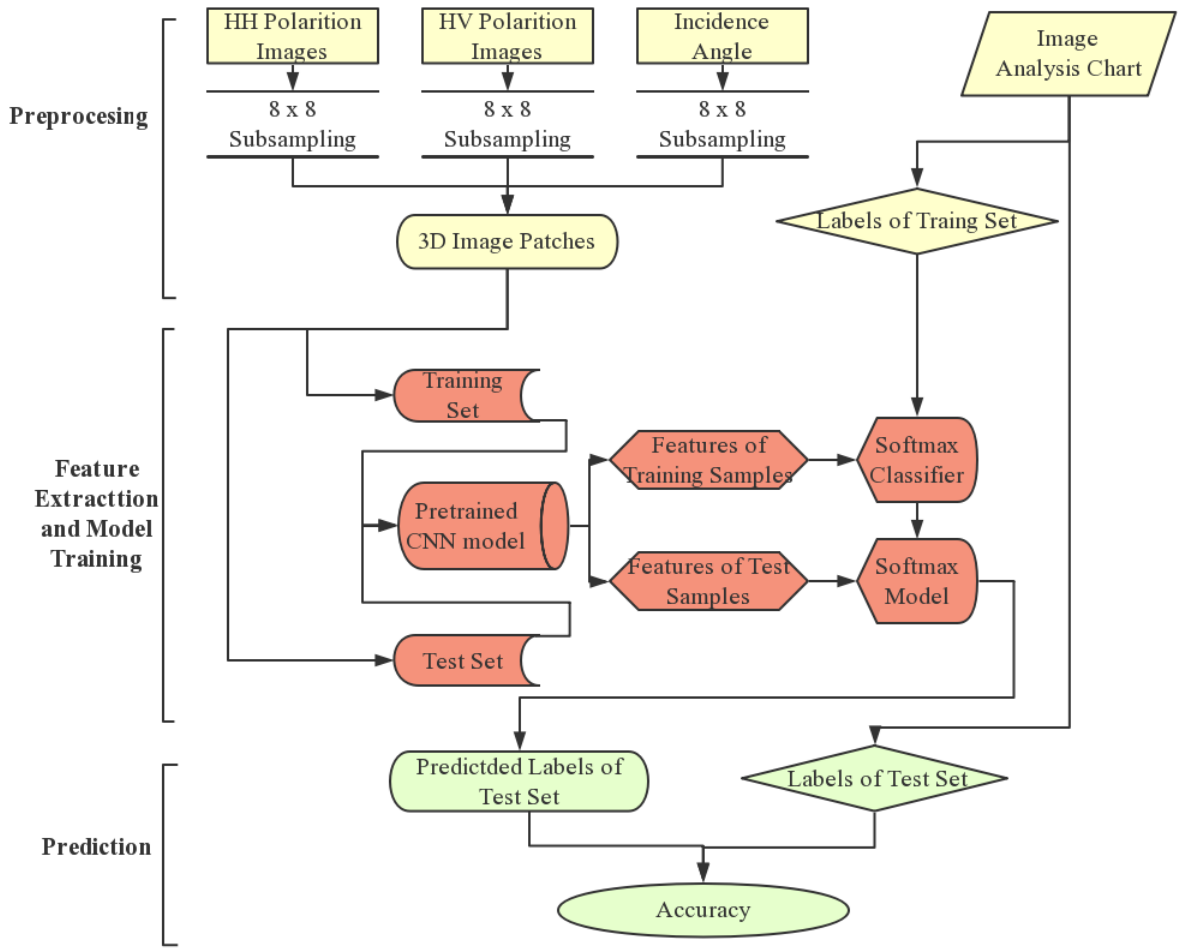


Figure 3.3: Flowchart of method for sea ice and water classification.

3.4 Model Evaluation Methods

In order to evaluate the performance of the model in this study, two methods are adopted for estimating the generalization error depending on different training and testing splits.

One is called holdout method, the other is called leave-one-out, which is a special variant of cross-validation method.

3.4.1 Holdout Method

Holdout method is relatively easy to understand, the target dataset with labeled samples is divided into two independent sets, namely, the training set and the test set [72]. After obtaining a classification model only using the training set, we evaluate the performance of the model on the test set to obtain an independent estimation of classification accuracy [72]. The percentage of samples in the dataset for training and for testing is typically 80 percent and 20 percent respectively [72].

The holdout method has several shortcomings. First of all, when 20% of the samples are reserved for test set, the number of training samples might be insufficient if the model we want to train is very complex [72]. Thus, the model trained by 80% of the dataset may not be as good as a model trained by all the samples. In addition, the model may be highly dependent on the way the training set and test set are structured [72]. On the other hand, it is a kind of trade-off, when we decide to use a large training set, then the test set must be small for a limited dataset. As a result, the computed classification accuracy from the test set is no longer reliable [72].

3.4.2 Leave-one-out Method

Leave-one-out method is an variant of cross-validation method. It involves using one observation as the test set and the rest of observations as the training set. In terms of this study, one SAR scene was chosen as test set each time, and the remaining SAR scenes are used for training. This is repeated for each image in the set, or 25 times. Leave-one-out approach has the advantage of utilizing as much data as possible for training. In addition, the test sets are mutually exclusive and they effectively cover the entire data set. The drawback of this approach is that it is computationally expensive.

3.5 Experimental Setup

In these experiments, stride of 20 was used between image patches when extracting patches from SAR images, which means spacing between patches is 20x400m or 8km [82]. The

spacing of image analysis pixels is 8km x 5km [82]. Therefore, this model will not provide prediction for every pixel in a SAR image [82]. In addition, pixels near image borders are unclassified as it is not possible to extract 45 x 45 patches around image border. Moreover, pixels near land boundaries are not classified to avoid contamination of the dataset with unphysical values.

In terms of the holdout method, the first 20 SAR scenes in Table 3.1 were chosen as training data (18918 patches), and the last five scenes were used as the test data (4623 patches). In terms of the leave-one-out method, each scene will be regarded as test set once. As there are no patches extracted from scene 20140126_223850 due to landcover, 24-fold cross validation is executed. The softmax classifier was trained by using a large set of labelled images patches $\{x_i, y_i\}$, where x_i represents inputs image patches, y_i represents labels of image patches. A cross-entropy loss function, a typical loss function for image classification, is used for the comparison between the true distribution p and the estimated distribution q, the mathematical expression of cross-entropy has this form, [34]

$$H(p, q) = -\sum_x p(x) \log \; q(x).$$

Softmax classifier produces the probabilty distribution of each class. As the estimated category probability q of the i-th category from the softmax classifier is defined as [34],

$$q_i = \frac{e^{f_{y_i}}}{\sum_j e^{f_j}}.$$

and f_j mean the j-th element of the vector of category scores [34]. Softmax function is a standard way to turn numbers to probability distributions.

3.6 Results

Measuring the performance of the model on an independent test set is very useful because such a measure provides an unbiased estimate of its generalization error [72]. On the same domain, the accuracy or error rate computed from the test set can also be used to compare the relative performance of classifiers [72]. In this study, we use classification accuracy to measure model performance. Classification accuracy has the following definition:

$$\text{Classification Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

In order to evaluate the performance of the model in this study, the two model evaluation methods discussed in Section 3.4 are adopted for estimating the generalization error depending on different training and testing splits. Results of both methods will be shown in the following two subsections. At the last part of this section, several conclusions are presented by comparing classification results with image analysis charts.

3.6.1 Holdout Results

In the experiments carried out here by calculating the number of patches extracted from each SAR scenes, the dataset is split such that around 80% is used for training, and 20% is used for testing, corresponding to five SAR scenes used for testing, and the remaining SAR scenes used for training. Table 3.2 below displays the classification results on the five test scenes. It can be clearly observed from table that the overall classification accuracy is around 96.690%. Some SAR scenes have even more accurate classification results. It can be seen in Table 3.2 that when a SAR scene has fewer patches, the classification accuracy is relatively lower. One reason for this may be that the classification results on small test set may not represent the real performance of the trained model, if the data in the test set are not well represented in the training data.

Table 3.2: Holdout results for ice-water classification for SAR scenes in the test set using transfer learning. Number of patches means the total number of patches extracted from each SAR scene in the test set.

Scene ID	**Number of Patches**	**Classification Accuracy**
20140207_214938	86	86.047%
20140208_095758	2204	97.096%
20140209_223030	60	88.333%
20140210_103911	380	98.684%
20140210_220111	1893	96.355%
Overall	**4623**	**96.690%**

3.6.2 Leave-one-out Results

Table 3.3: Leave-one-out results for ice-water classification for the GSL2014 dataset using transfer learning. Number of patches means the total number of patches extracted from each SAR scene

Scene ID	Number of patches	Classification Accuracy
20140116_223042	66	62.121%
20140117_103914	579	71.848%
20140118_101002	2320	88.836%
20140120_105149	120	82.500%
20140121_214420	1704	92.019%
20140122_095247	1332	91.967%
20140123_222627	146	93.836%
20140124_103501	946	97.463%
20140124_215646	2375	92.547%
20140125_100500	1130	93.009%
20140127_104734	227	79.295%
20140127_221027	79	89.873%
20140128_101751	1572	91.667%
20140130_110029	7	84.436%
20140130_222234	257	84.825%
20140131_103053	1327	96.458%
20140131_215240	2428	89.086%
20140206_221744	598	92.642%
20140207_102631	1705	96.716%
20140207_214938	86	86.047%
20140208_095758	2204	97.232%
20140209_223030	60	90.000%
20140210_103911	380	97.632%
20140210_220111	1893	96.144%
Overall	**23541**	**92.360%**

Results for the leave-one-out experiments are shown in Table 3.3. By averaging the classification results of every experiment weighted by the number of patches in the test image, around 92.3% classification accuracy is achieved. Although this overall classification accu-

racy is lower than the classification computed by hold-out method, it is still very impressive. According to these results, we can conclude that the method proposed in this chapter performs well on our dataset and it also has a good generalization ability. In contrast, Leigh [42] et al. developed a ice-water classification system named MAp-Guided Ice Classification (MAGIC) with an impressive leave-one-out classification accuracy with 96.42%. A paper [84] investigated sea ice and water classification and proposed a SVM based method by using hand-crafted texture features, their method achieved average $91\% \pm 4\%$ classification accuracy on validation SAR scenes.

3.6.3 Results Visualization

In order to present the classification results for visualization, the image patches are mapped back to the HH pol SAR image, and are compared with the image analysis charts. Examples are shown in Figure 3.4 and Figure 3.5.

In Figure 3.4, two sub-images of two SAR scenes with high classification results (97.232% and 96.144% respectively) are exhibited. The two pictures at the first row are SAR images in HH pol. The two images in the middle are corresponding image analysis charts after thresholding ice concentration. Pixels near land and image boundary are unclassified because the method has not been developed to extract sea ice or water patches from these region. According to Figure 3.4, the classification results are fairly good compared with image analysis charts. Many of the misclassified points are close to the ice edge, i.e., the ice-water boundary. There are only a few misclassified pixels in the open water and ice cover regions respectively, indicating that the method can distinguish both sea ice and open water equally well.

Apart from looking at SAR scenes with high classification accuracies, it is also of interest to look into a SAR scene with relatively low classification accuracy to help us understand cases where the model does not perform well. Figure 3.5 presents a sub-image of a SAR scene with 89% classification accuracy, highlighting the region with classification errors. We can clearly find that the main errors made by model are near ice water boundary, some regions identified as water by the image analysis chart are classified as ice. One interesting finding is that the region highlighted by a yellow circle. This region contains low ice concentration (mainly consists of water) according to ice analysis chart and HH pol image, while the predicted results show that that area mainly consists of sea ice. It seems that the model classification ability for low ice concentration region is not that good. Adding more training samples with low ice concentration might help in this case.

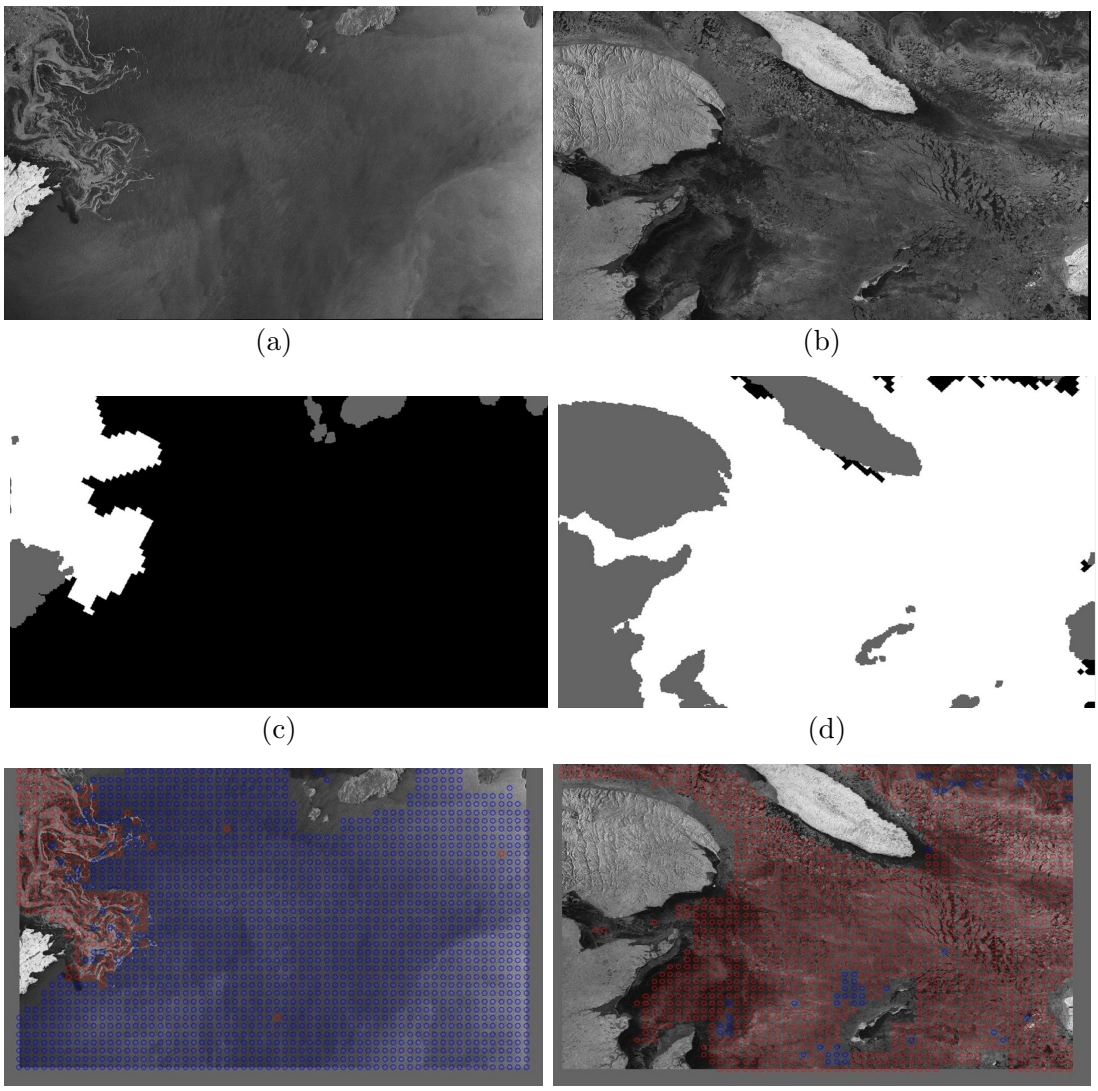
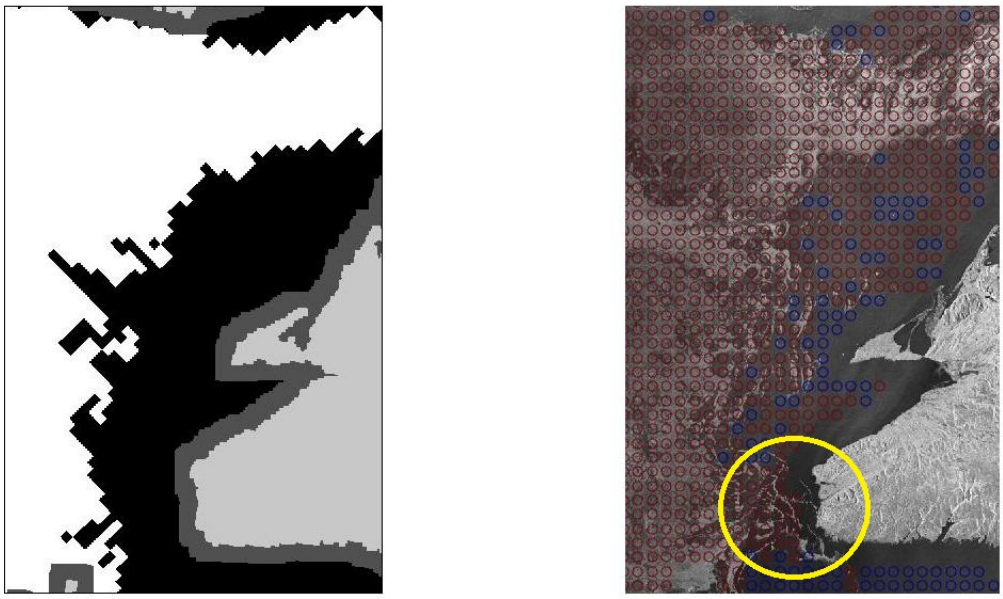


Figure 3.4: Water and sea ice classification results visualization of two SAR scenes. Source[82], (a), (c), (e) are HH pol image, ground truth, and prediction results of SAR scene (ID: 20140208_095758) respectively; (b), (d), (f) are HH pol image, ground truth, and prediction results of SAR scene (ID: 20140210_220111) respectively. For plots (c) and (d), white represents sea ice, black stands for open water, grey areas are unclassified areas because these are land. The two images at bottom are classification result plots, red circle and blue circle represent sea ice and open water respectively.



(a) (b)

Figure 3.5: An example of a SAR scene with relatively low classification accuracy. This is a sub-image of Scene ID: 20140131_215240. (a) The corresponding image analysis chart, white represents sea ice, black represents open water, grey represents land and unclassified area. (b) Classification results mapped on SAR image in HH pol, red circle: sea ice. blue circle: water, the yellow circle highlights a region of interest.

3.7 Discussion

By comparing the classification results with image analysis charts, it can be found that regions near sea ice and water boundary tend to be misclassified. The weak ability to classify pixels near sea ice and water boundary might caused by the way we assign labels for each patch. As the models are trained to estimate the label of the central pixel of a patch, when giving a label for one patch of size $45 \times 45 \times 3$, we just assign the label of the center pixel as the label for the whole patch. This label assignment method potentially introduced errors because it might assign wrong labels for some patches. For example, if most of the pixels in one patch are water, but the center point is regarded as sea ice according to a image analysis chart, then the whole patch will be labelled as sea ice, while the true label for this patch should be water.

Note that some of these misclassifications could be due to errors in the image analyses, or due differences in how information is represented in image analyses (polygons) as compared to our method (image patches).

Some points near land and near image borders are unclassified as well, which might be a concern if information in these regions is desired. One method to address this problem will be discussed in the next Chapter.

Chapter 4

Sea Ice Concentration Estimation Using CNN-based Regression

The previous chapter presented a method to classify pixels in SAR sea ice imagery as ice and water. In this chapter, the problem of sea ice concentration estimation from SAR imagery is addressed.

Accurate sea ice concentration information is important for the human population. Sea ice has an impact on weather and climate. For example, the amount of heat transfered through the ice is directly dependent on the ice concentration [78]. The information of marginal ice zone is also of interest because it is an indicator of Arctic warming and it provides a nice living environment for phytoplankton and creatures such as seabirds and zooplankton [24]. Sea ice concentration also plays an important role in ship navigation. Navigators depend on sea ice concentration charts to find safe routes for icebreakers and ships [78]. For example, regions with low sea ice concentration are passable for many ships while regions with high ice concentration could be very dangerous. Sea ice concentration estimation from SAR imagery is currently done manually by experts working at the Canadian Ice Service. However, these ice charts often lack the small scale details of the ice cover, such as individual floes or openings. It is of interest to develop a method to automatically estimate ice concentration from SAR imagery that contains the relevant details.

Different from hand-crafted feature extraction based methods, CNN-based methods learn features from a dataset automatically [23], which can be a good fit for this sea ice concentration estimation problem. In this Chapter, a CNN-based regression model is developed for estimating sea ice concentration based on dual-polarization images (HH and HV), combined with incidence angle image. The model is similar to that used in an earlier

study [75], although the testing method used here is more thorough. In addition, various experiments are carried out to investigate the role of the training data on the estimation of intermediate ice concentrations (those between 10% and 80%). This will be referred to addressing the imbalanced data problem. Before applying CNN regression models, several data preprocessing techniques, for examples, land mirroring, are discussed in this Chapter.

4.1 Dataset

In this study, the base SAR image dataset is the same dataset used for the experiments in chapter 3, i.e., the SAR image dataset from the Gulf of St. Lawrence captured in 2014. The description of that dataset can be found in Section 3.1.

Apart from this dataset, three complementary datasets are introduced in order to address the imbalanced dataset problem. These SAR images are also acquired in ScanSAR wide mode from RADARSAT-2 but from different regions and years. One dataset is very similar to the aforementioned dataset, and contains SAR images captured from the Gulf of St. Lawrence during the ice-covered season within the period of January 21st, 2012 to February 11th, 2012. Six available SAR scenes with reliable image analysis charts from that dataset are used. The other dataset consists of SAR images captured from Labrador Sea in the period of February 14th to February 20th, 2011. The Labrador Sea is located northeast of the Gulf of St. Lawrence. In this region, sea ice occurs near the coast of Labrador in December and sea ice becomes thicker from January to February [64]. The sea ice cover generally consists of a marginal ice zone next to the open water [64], which is quite similar to the ice features in the Gulf of St. Lawrence. There are 4 SAR scenes with image analysis charts available in this dataset. The third dataset contains SAR images captured from Canadian Arctic Archipelago in the period of July 24th 2013 to September 13th 2013. and differs from the other datasets in that some of the SAR scenes contains regions populated with multi-year and first-year ice floes. This dataset consists of 6 SAR scenes. More details about these datasets are given in Table 4.1

Table 4.1: Image acquisition dates for all images used in the ice concentration estimation study

The Gulf of St. Lawrence 2014 Dataset

Number	*Scene Number*
1	20140116_223042
2	20140117_103914
3	20140118_101002
4	20140120_105149
5	20140121_214420
6	20140122_095247
7	20140123_222627
8	20140124_103501
9	20140124_215646
10	20140125_100500
11	20140126_223850
12	20140127_104734
13	20140127_221027
14	20140128_101751
15	20140130_110029
16	20140130_222234
17	20140131_103053
18	20140131_215240
19	20140206_221744
20	20140207_102631
21	20140207_214938
22	20140208_095758
23	20140209_223030
24	20140210_103911
25	20140210_220111

The Gulf of St. Lawrence 2012 Dataset

Number	*Scene Number*
1	20120121_104328
2	20120125_102632
3	20120128_220047
4	20120204_103503
5	20120208_101753
6	20120211_103046

Labrador Sea 2011 Dataset

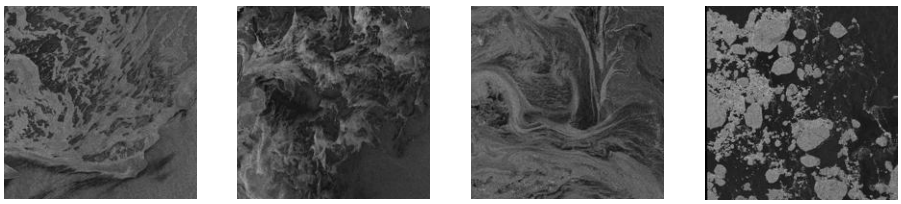
Number	*Scene Number*
1	20110214_094659
2	20110216_102816
3	20110217_100003
4	20110220_101204

Canadian Arctic Archipelago 2013 Dataset

Number	*Scene Number*
1	20130724_145706
2	20130731_131133
3	20130801_223532
4	20130908_021928
5	20130910_131613
6	20130913_150901

In order to describe different datasets easily and clearly, dataset of the Gulf of St. Lawrence acquired in 2014 will be referred to as GSL2014; The dataset of the Gulf of St. Lawrence acquired in 2012 and the dataset of Labrador Sea acquired in 2011 will be together referred to as GSL2012LS2011. The reason why these two datasets are combined is because the sea ice features in those two datasets are very similar. The Canadian Arctic

Archipelago 2013 dataset is referred as CAA2013, ice features in this dataset are not very similar to that of the GSL2014 dataset. For example,the region contains both multi-year and first-year-ice with distinct ice floes. Figure 4.1 displays a sample of the HH SAR scene from each dataset.



(a) (b) (c) (d)

Figure 4.1: Sea ice features in different datasets.(a) GSL2014; (b) GSL2012; (c) LS2011; (d) CAA2013

The geographical location of each SAR scene in the datasets is shown in Figure 4.2. For each SAR scene, the mean value of the latitude and longitude of the image data locations are computed and used to represent the image location. Each color in Figure 4.2 represents a different dataset.

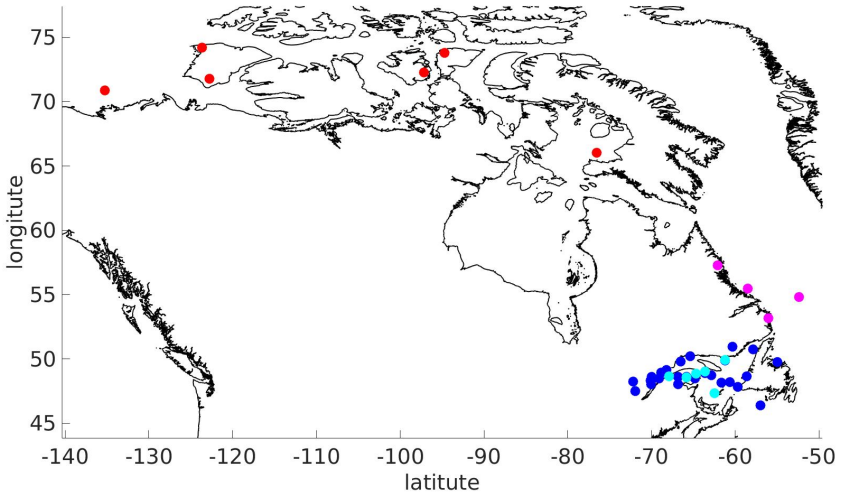


Figure 4.2: Geographical location of the datasets used for sea ice concentration estimation study. Blue: GSL2014, Cyan: GSL2012, Purple: LS2011, Red: CAA2013

4.2 Data Preprocessing

For all the experiments in this Chapter, the data preprocessing steps include image analysis charts interpolation, SAR imagery downsampling, incidence angle data processing, and are similar to what have been presented in Section 3.2. One difference is that a smaller stride (stride=4) as compared to stride of 20 from previous chapter is used to extract patches, because we want ice concentration at a relatively fine spatial resolution. Another difference is that a land mirroring method is adopted to enable estimation of sea ice concentration near coast lines.

4.2.1 Land Mirroring

Most of SAR images acquired from the Gulf of St.Lawrence contain land, which creates a problem when estimating sea ice concentration using square patches. As the size of a patch is 45 pixels x 45 pixels (18 km x 18 km), patches extracted near land have high probability of containing land pixels. In order to satisfy the needs of obtaining sea ice concentration information near land, a method called land mirroring [75] is used to deal with land pixels in image patches. Figure 4.3 shows an example of land mirroring. Note that if the label of the center pixel of a patch is land, then that patch will not be used. If not, the land mirroring method will be applied. The land mirroring method is as follows: For a 45 pixel x 45 pixel image patch that contains land pixels, the coast line is regarded as a mirror. Each land pixel (e.g., pixel B) in a patch will be replaced by the mirrored sea ice or water pixel (e.g., pixel A), if there are some land pixels that cannot find their mirrored pixel, then these pixel values will be replaced by the center pixel value. This does make sense because the models are trained to estimate the sea ice concentration of the central pixel of a patch. Land mirroring is a very helpful way to remove land contamination effects and provides sea ice concentration information near coast lines and islands.

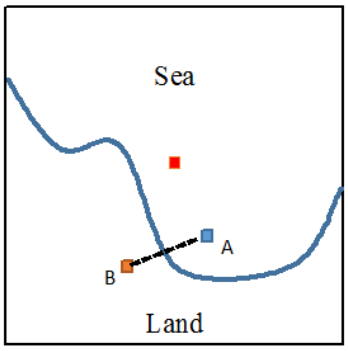


Figure 4.3: A example of land mirroring

4.3 Convolutional Neural Network Model

4.3.1 CNN structure

The CNN-based regression model structure used for studies in this chapter is a classical CNN structure. It mainly consists of three convolutional layers, three pooling layers, and two fully-connected layers. The last layer is sum of squares or Euclidean loss layer. A dropout [70] layer is used near the top layer in order to reduce overfitting. Dropout sets a neuron active with probability p. Here, the dropout rate is set as 0.5. The rectified linear unit (Relu) is used as activation function which has the form, $f(x) = \max(0, x)$. Relu is chosen here rather than Tanh because it is mentioned in [38] that there is more than six times improvement in convergence speed with Relu comparing to Tanh. Local response normalization (LRN) is the normalization algorithm which performs a kind of 'lateral inhibition' by normalizing over local input regions [86]. The detailed structure of the CNN is shown in Table 4.2 below.

Table 4.2: The structure of CNN-based regression model

Layer	Method	Settings	Output	The # of Weights
Input layer		$45 \times 45 \times 3$	$45 \times 45 \times 3$	0
Conv1	Relu	Kernel_size: 5 Stride : 1; Padding:2 Num_output: 64	$45 \times 45 \times 64$	$5 \times 5 \times 3 \times 64$ +64= 4864
Pool1	Maxpooling	Kernel_size: 3 Stride: 2; Padding:1	$23 \times 23 \times 64$	0
Norm1	LRN		$23 \times 23 \times 64$	0
Conv2	Relu	Kernel_size: 5 Stride : 1; Padding: 2 Num_output: 128	$23 \times 23 \times 128$	$5 \times 5 \times 64 \times 128$ +128=204928
Pool2	Maxpooling	Kernel_size: 3 Stride : 2	$11 \times 11 \times 128$	0
Norm2	LRN		$11 \times 11 \times 128$	0
Conv3	Relu	Kernel_size: 5 Stride : 1; Padding: 2 Num_output: 128	$11 \times 11 \times 128$	$5 \times 5 \times 128 \times 128$ +128= 409728
Pool3	Maxpooling	Kernel_size: 3 Stride : 2	$4 \times 4 \times 128$	0
Fc1	InnerProduct	Num_output: 1024	1024	$4 \times 4 \times 128 \times 1024$ +1024 = 2098176
Dropout1		Dropout rate: 0.5	1024	0
Fc2	Euclidian Loss	Num_output: 1	1	$1 \times 1024 + 1 = 1025$
Number of weights that need to be learned: 2718721				

The convolutional layers each contain 64 to 128 filters with kernels of size 5×5. Max-pooling kernels of size 3×3 are used. Finally, Euclidean loss are used as the loss function of the CNN. There are around 2.7 million weights in this convolutional network that need to be learned. There are no weights in pooling layers, dropout layer, and normalization layers.

To understand the number of weights associated with the convolutional and fully-connected layers. Let us use K, S, P, F to denote the number of filters, the stride, the amount of zero padding, and filter size respectively. $W_i \times W_i \times D_i$ denotes the size of output at the ith layer, so W_{i-1} is the input size of the ith layer, and the relationships

between them are presented as follows:
$W_i = (W_{i-1} - F + 2P)/S + 1$ [31]
$D_i = K$ [31]
According to the structure of convolutional layer, it introduces $F \times F \times D_{i-1}$ weights per filter, so the total of $F \times F \times D_{i-1} \times K + K$ weights in a convolutional layer need to be determined [31], where the $F \times F \times D_{i-1} \times K$ weights are from the convolution operations and the K weights are from the bias terms. In terms of fully-connected layer, the number of weights equals the size of output from last layer times the number of neurons in current fully-connected layer add biases.

Hyper-parameter tuning is also an important aspect that can affect the CNN model performance. Table 4.3 shows several important parameters that we used in the experiments in this chapter.

Table 4.3: Hyper-parameters of the CNN

Learning Rate	Momentum	Weight Decay	Dropout Rate	Batch Size
0.001	0.9	0.05	0.5	256

4.3.2 Implementation

Most data preprocessing methods, SAR imagery preprocessing, image analysis charts interpolation and result visualization et al., are implemented using Matlab and Python. Statistical testing is implemented by R studio, CNN Model training and prediction are implemented based on a deep learning framework called Caffe [26].

4.4 Studies of The Impacts of Removing Intermediate Ice Concentration Training Samples

4.4.1 Experiment Design

In this study, only the GSL2014 dataset is used. The CNN regression model explained in the section 4.3 is adopted for all the experiments in this section. There are two main types of training data included in this study, one consists of all values of ice concentrations (AICs) from the image analysis chart, the other is using only high and low ice concentrations

(HLICs) as training data, which satisfy the following requirement:

$$IC <= 20\% \ or \ IC >= 80\%$$

The reason why this idea is proposed is because it is believed among sea ice community that ice concentration between 20%- 80% might not be accurate in image analysis charts. Therefore, it was desired to investigate if it would be beneficial to remove these samples when training a CNN because training samples with more accurate labels are used for training. However, one drawback is that the number of training samples is decreased. In order to investigate if the effect is from the number of training samples, another experiment of using the same number of AICs training samples as the number of HLICs samples is carried out.

After the training samples are generated, the CNN regression model is trained, followed by validation, which is carried out using five-fold cross validation. Finally, five-fold cross validation is adopted. To do the cross-validation, the dataset is first split into five equal subsets, where each subset contains the same number of SAR scenes [80]. Of these subsets, four subsets are used to train the model and the remaining subset is used to test model performance [80]. This process is repeated until all subsets have participated in the test dataset.

In order to compare the performance of the CNN regression model with other algorithms, sea ice concentration retrieved using the Artist Sea Ice (ASI) is used. This method was proposed in [69], which is an algorithm of estimating sea ice concentration by using the the brightness temperature difference of the vertically and horizontally polarized channels at 89GHz [69, 75]. The advantage of this algorithm is that it has high spatial resolution and less bias of thin ice in comparison with other passive microwave retrieval algorithms [56].

4.4.2 Results and Discussion

The 25 SAR scenes available are partitioned into five folds as shown in Table 4.4. Thus, 20 SAR scenes are used for training and the five SAR scenes are used for testing for each fold. In this way each of the 25 SAR scenes participates in the testing. In order to clearly evaluate the performance of models, root mean squared error (E_{rms}) is used for the cross-validation experiments.

$$E_{rms} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f(x^{(i)};\theta) - y^{(i)})^2}$$

The root mean squared error (E_{rms}) are computed for each SAR scene and results are presented in Table 4.4. The # of patches denotes the number of patches in the test data.

Table 4.4: Experimental results from using AICs as compared to HLICs in the training data. Results shown here are for all five folds. The number of patches means the total number of patches extracted from each SAR scene

Fold	**Scene ID**	**The # of Patches**	E_{rms} **for AICs**	E_{rms} **for HLICs**
1	20140116_223042	4373	0.2339	0.2670
1	20140125_100500	36050	0.1720	0.1830
1	20140118_101002	64955	0.1637	0.2066
1	20140131_103053	43989	0.1526	0.1847
1	20140210_103911	13728	0.1360	0.1530
2	20140122_095247	40470	0.2002	0.2500
2	20140130_110029	1669	0.2553	0.3092
2	20140131_215240	71188	0.2042	0.2232
2	20140208_095758	59263	0.1349	0.1972
2	20140117_103914	23842	0.2285	0.2506
3	20140126_223850	1411	0.4777	0.5623
3	20140207_102631	51965	0.1456	0.1536
3	20140210_220111	57086	0.1655	0.1559
3	20140124_103501	33947	0.1500	0.1597
3	20140120_105149	6691	0.2869	0.3169
4	20140130_222234	10088	0.2837	0.2596
4	20140209_223030	4424	0.2722	0.2793
4	20140130_222234	10088	0.2837	0.2596
4	20140128_101751	47642	0.1746	0.2059
4	20140121_214420	53220	0.2472	0.2200
5	20140207_214938	3546	0.2211	0.2238
5	20140206_221744	24277	0.2470	0.2588
5	20140127_221027	3833	0.0494	0.2223
5	20140124_215646	72053	0.1977	0.1825
5	20140123_222627	7693	0.2557	0.2506
Average E_{rms}		**N/A**	**0.1917**	**0.2085**

It can be seen in Table 4.4, using all ice concentration to train the CNN performs better than only using HLICs to train the CNN. For some specific SAR scenes, for in-

stance, 20140121_214420 and 20140124_215646, using HLICs to train CNN models improved model performance. The average root mean squared error of using all ice concentration is around 0.1917, while that of using HLICs is around 0.2085. The results for scene ID 20140126_22385 have the highest error. This is because most patches in that SAR scene contain land, so there are not too many patches can be extracted in that SAR scene. A small number of test samples cannot fully describe underlying error. In addition, areas near land are a challenge for the CNN because of the approximation brought in through the land mirroring.

To look at the statistical significance of the differences between the results from AICs and HLICs, a Wilcoxon signed-rank test [79] is conducted. The Wilcoxon signed-rank test is a non-parametric method appropriate for examining the median difference between paired observations [11]. This statistical test has been used in previous studies comparing ice concentration datasets [20]. Here, significance level $\alpha = 5\%$ is used, the null hypothesis and alternative hypothesis are:

$$H0 : \text{the median difference} = 0$$

$$H1 : \text{the median difference} > 0$$

After running the Wilcoxon signed-rank test by using the statistics package in R studio, the P_value = 0.0003575, which means that the null hypothesis is rejected. Therefore, we have statistically significant evidence at $\alpha = 0.05$ to show that the root mean squared error when using AICs to train CNN models is lower than that from using HLICs to train our CNN for the dataset investigated here.

To investigate whether the effect is due to the number of training samples or the lack of HLICs training samples. The three experiments in Table 4.5 are carried out. The evaluation method for these experiments are based on the holdout method, the 1st fold in table 4.4 is used as the testing set and the rest of the data are used as the training set. Apart from E_{rms} defined above, the following statistics are used for the analysis of errors for the experiments in Table 4.5.

$$E_{mean} = \frac{1}{N}\sum_{i=1}^{N}(f(x^{(i)};\theta) - y^{(i)})$$

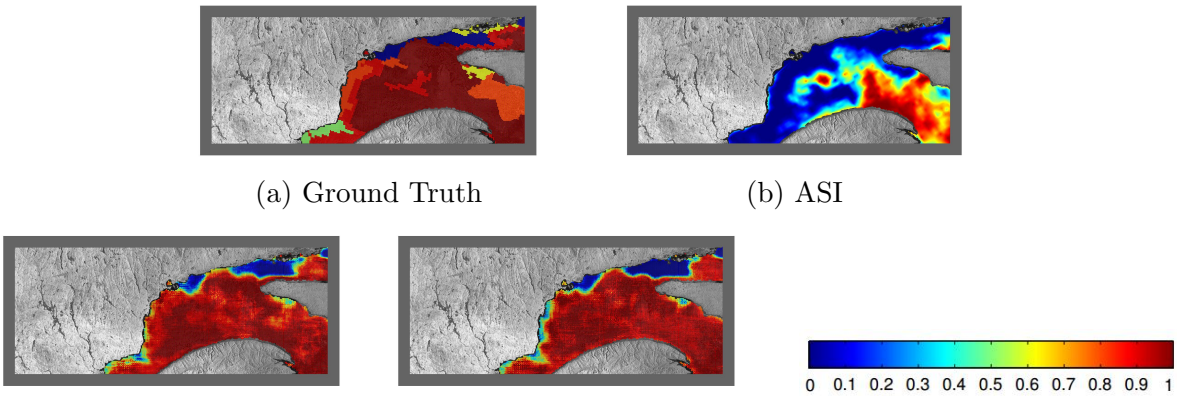
$$E_{L1} = \frac{1}{N}\sum_{i=1}^{N}\left|f(x^{(i)};\theta) - y^{(i)}\right|$$

where E_{mean} is the mean value of the difference between predicted value and ground truth. N denotes the total number of samples in the test set E_{L1} denotes L1 error.

Table 4.5: The comparison of three experiments. Results shown here are for fold 1 test data only, where training is done using fold 2-5.

ID	Dataset	The # of Training Samples	The # of Testing Samples	E_{mean}	E_{L1}	E_{rms}
1	AICs of GSL2014	584415	163095	-0.0012	0.0909	0.1630
2	HLICs of GSL2014	514561	163095	0.0022	0.1182	0.1936
3	A subset of AICs of GSL2014	514561	163095	-0.0063	0.0947	0.1668

It can be seen from Table 4.5 that the effect of using HLICs to train a CNN model is not from the number of training samples. Although Exp.2 and Exp.3 have the same number of training samples, the E_{rms} of Exp.2 is significantly higher than that of Exp.3. We can conclude that intermediate ice concentration samples are very important for training a CNN model to estimate sea ice concentration.

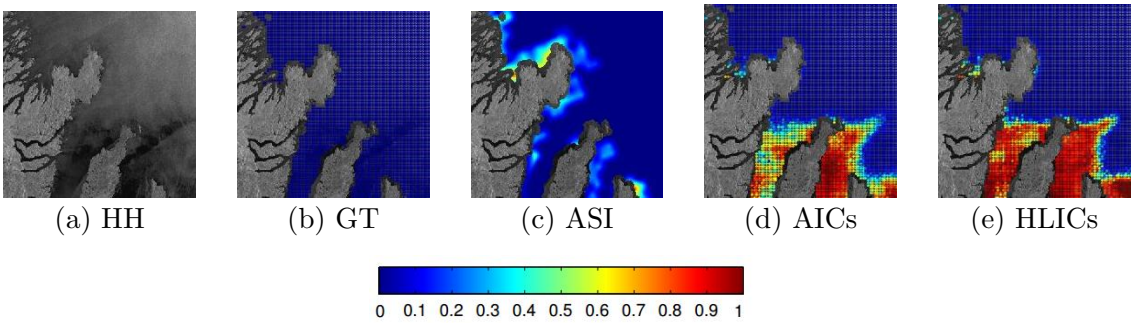


(a) Ground Truth (b) ASI

(c) CNN with AICs (d) CNN with HLICs

Figure 4.4: Results visualization for 20140210_103911.

Visual comparison of sea ice concentration from the CNN with that from ASI and the image analysis are shown in Figure 4.4. It can be seen that sea ice concentration estimated



(a) HH (b) GT (c) ASI (d) AICs (e) HLICs

Figure 4.5: Sub-scene from 20140122_095247 showing where CNN regression model yield incorrect results. (a) SAR imagery in HH polarization. (b) GT refers to ground truth, which is the image analysis chart. (c) Estimation results from ASI. (d) AICs. (e) HLICs.

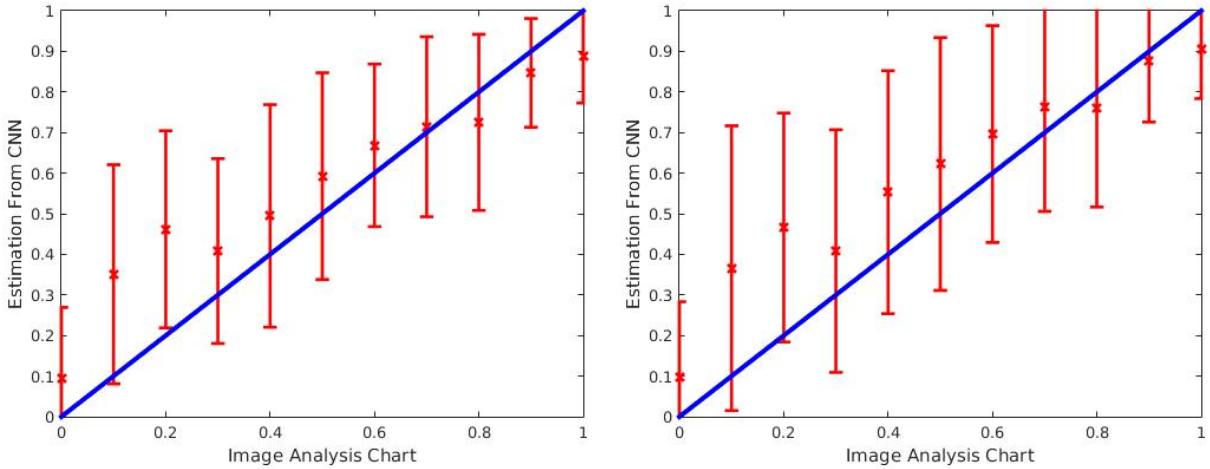
from ASI are relatively inaccurate compared to the image analysis chart, while sea ice concentration estimated from the CNNs are much better. Results from other scenes can be seen in Figure A.1-A.6 in Appendix A. From Figure A.2 and Figure A.6, the difference between using all ice concentrations for training and using HLICs for training can be observed. Sea ice concentration in image analysis charts with values over 60% are typically overestimated in the experiment using HLICs, and sea ice concentration in image analysis charts with values under 40% are typically underestimated in the experiment using HLICs. This means that the model trained by HLICs tends to identify marginal ice zone either as consolidated ice zone or water.

It is worthwhile to understand in what kind of situation CNN models perform poorly on SAR imagery, so that we can improve model design in the future. Some intuition can be gained by visually looking at the results. For instance, it can be seen in Figure 4.5 that there are dark regions near the land boundary in the bottom left of the SAR image in HH pol. These dark regions are due to reduced windspeed in this region because of local topography. By comparing Figure 4.5d with Figure 4.5e, it can be found that using all ice concentrations performs slightly better than using HLICs, as CNN models trained by HLICs tend to estimate higher ice concentrations in this case.

4.4.3 Error Analysis

To compare the estimation result and the ground truth for each ice concentration category, two pictures showing the means and standard deviations of estimation results are generated in Figure 4.6. In this figure, the blue line represents the desired relationship between sea

ice concentration from image analysis charts and sea ice concentration estimated from CNNs, the red cross and the red bar represent the estimated mean value and standard deviation for each ice concentration category respectively. It can be seen that means are close to the true means but the standard deviations are high. By comparing Figure 4.6a and Figure 4.6b, we can conclude that the CNN trained on all ice concentration performs better than the CNN trained on HLICs, especially for those marginal ice zone (ice concentration greater than 20% or less than 80%). As the HLICs is mainly consists of low sea ice concentration categories and high sea ice concentration categories, the CNN was trained to be more sensitive to either low ice concentrations or high ice concentrations rather than intermediate ice concentrations, its performance in marginal ice zone is weak due to the lack of intermediate ice concentrations training samples.



(a) Training on AICs (b) Training on HLICs

Figure 4.6: Ice concentrations prediction distribution calculated based on cross validation results of the entire GSL2014 dataset. The blue line represents the desired relationship between sea ice concentration from image analysis charts and sea ice concentration estimated from CNNs. The red cross and the red bar represent the estimated mean value and standard deviation for each ice concentration category respectively

Figure 4.7 presents sea ice concentration prediction distribution for each ice concentration category versus the image analysis charts. These are the prediction results from the whole GSL2014 dataset using five-fold cross-validation, and this is the prediction results by using AICs to train the CNN models. It can be seen that the CNNs perform well on distinguishing high ice concentration categories and low ice concentration categories,

while the performance on intermediate ice concentrations categories is not that good. This problem will be discussed more in the next section.

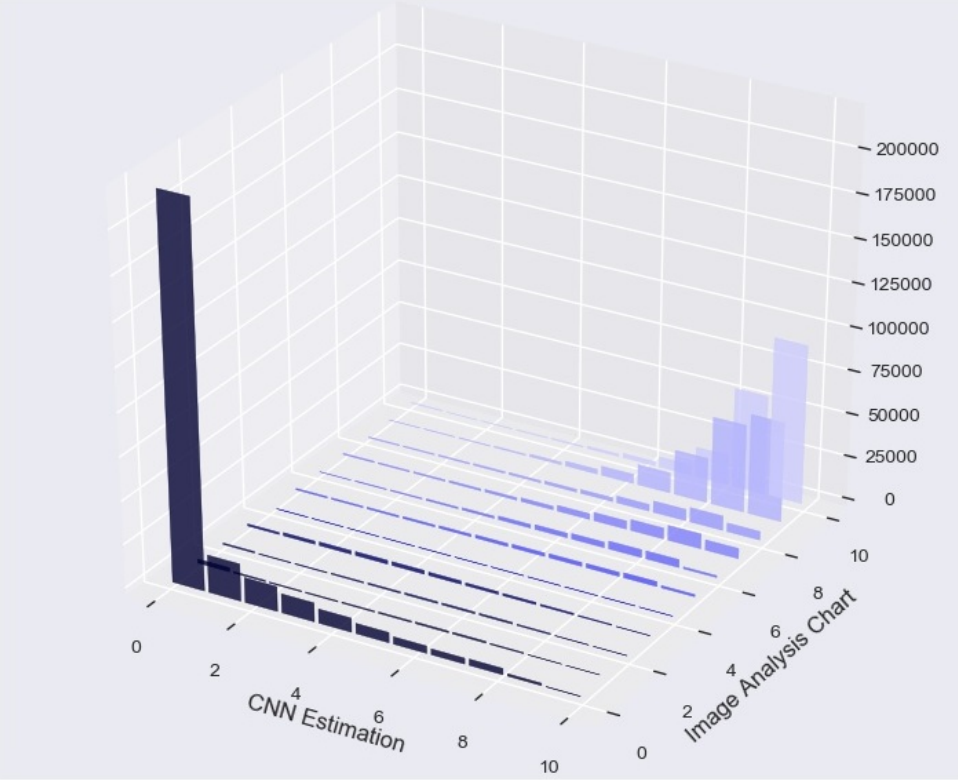


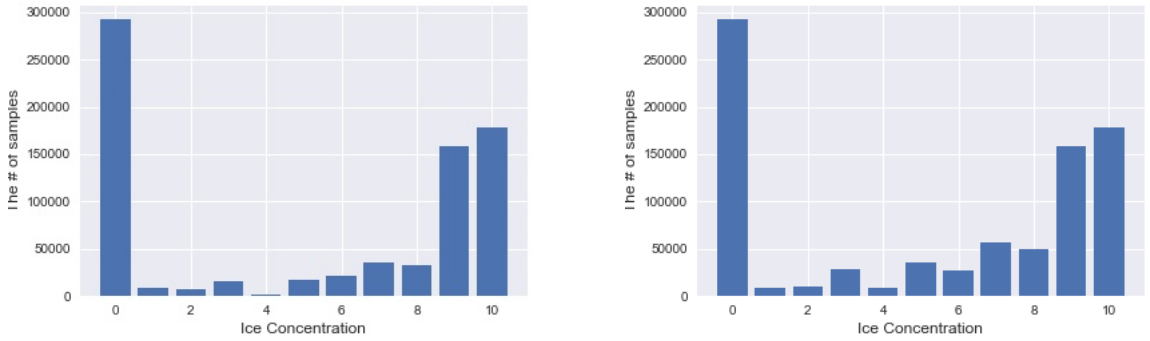
Figure 4.7: Estimation results from the CNN versus image analysis chart. This figure is generated based on the cross-validation results using all ice concentrations to train the CNNs. The vertical axis represents the number of patches

4.5 Studies of The Impacts of Increasing Intermediate Ice Concentration Training Samples

4.5.1 Problem Description

Various datasets available in real life are imbalanced due to the restriction of obtaining data for some specific categories or due to the nature of the dataset. For example, dataset

of patients who have or do not have cancer may be highly imbalanced, because few people may have cancer given a randomly sampled set of patient records. For GSL2014 dataset, marginal ice zones in the Gulf of St. Lawrence are less common than that open water and consolidated ice. Figure 4.8a displays the distribution of GSL2014 dataset over different ice concentration categories. It is very clear that this dataset is a very typical example of imbalanced dataset. Figure 4.8a reveals that water (0% ice concentration) and high ice concentration categories (90% and 100% ice concentration) consist of the major part of the dataset, while there are far fewer ice samples with ice concentration in the range of 10% to 80%.



(a) (b)

Figure 4.8: Distribution of ice concentration categories. The two figures show the distribution of datasets over ice concentration categories; (a) The GSL2014 dataset; (b) The GSL2014 plus the intermediate ice concentration samples of the GSL2012LS2011 dataset.

4.5.2 Experiment Design

One set of methods for addressing imbalanced dataset problem consists of various ways to balance the dataset. Undersampling the majority classes is one method that can be used to solve the problem of an imbalanced dataset. However, this method is not applicable for the present study. Recall that the 2014 dataset has only 25 SAR scenes. If we undersample the majority category to have almost the same number of samples as the minority classes, we will not have enough samples to train a relatively good CNN model due to the high number of parameters in the model.

Another way is to apply data augmentation based on the original GSL2014 dataset by rotating and flipping image patches with intermediate ice concentrations in the GSL2014 dataset. As an alternative approach, we can introduce more intermediate ice concentration samples from other available SAR image datasets captured from different regions and years.

To investigate this, several experiments are conducted as described in Table 4.6. One experiment uses the GSL2014 dataset plus intermediate ice samples of the GSL2012LS2011 dataset. This experiment aims at balancing the dataset by adding more intermediate ice concentration samples with similar sea ice features as the GSL2014 dataset. Figure 4.8b shows the distribution of ice concentration categories after adding more samples. As the number of available SAR scenes is small, the dataset is still imbalanced, but it is less imbalanced compared to Figure 4.8a. The second experiment uses the GSL2014 dataset plus the whole GSL2012LS2011 dataset. The objective of this experiment is to compare with the first experiment. The third experiment uses the GSL2014 dataset and intermediate ice concentration samples of the CAA2013 dataset. This experiment aims at balancing the GSL2014 dataset by increasing more marginal ice samples for the case where sea ice features are unlike the GSL2014 dataset.

The evaluation method for this study is based on the holdout method, the 1st fold in table 4.4 is used as the testing set and the rest of the data are used as the training set. The following statistics are used for the analysis of errors for the experiments,

$$E_{mean} = \frac{1}{N}\sum_{i=1}^{N}(f(x^{(i)};\theta) - y^{(i)})$$

$$E_{L1} = \frac{1}{N}\sum_{i=1}^{N}\left|f(x^{(i)};\theta) - y^{(i)}\right|$$

$$E_{rms} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f(x^{(i)};\theta) - y^{(i)})^2}$$

where E_{mean} is the mean value of the difference between predicted value and ground truth. N denotes the total number of samples in the test set E_{L1} denotes L1 error and E_{rms} denotes the root mean squared error.

4.5.3 Results and Discussion

Results from experiments using different training datasets, evaluated for the case where fold 1 is the test data, are shown in Table 4.6 below.

Table 4.6: Results from experiments using different training datasets. Results shown here are for fold 1 test data only, where training is done using fold 2-5.

Exp_id	Dataset	The # of Training Samples	The # of Testing Samples	E_{mean}	E_{L1}	E_{rms}
1	GSL2014	584415	163095	-0.0012	0.0909	0.1630
2	HLICs of GSL2014	514561	163095	0.0022	0.1182	0.1936
3	GSL2014 +data augmentation of intermediate ICs	799981	163095	-0.0065	0.1093	0.1782
4	GSL2014 + GSL2012LS2011	1074391	163095	-0.0132	0.0899	0.1563
5	GSL2014 + intermediate ICs of GSL2012LS2011	667611	163095	-0.0035	0.0898	0.1562
6	GSL2014+ intermediate ICs of CAA2013	681549	163096	-0.0104	0.0974	0.1543

Comparing the results from Exp.1 and Exp.3, it can be seen that implementing data augmentation by adding more marginal ice samples by flipping and rotating existing samples does not help to improve model performance. The E_{rms} of Exp.3 is even higher than that of Exp.1, In contrast to Exp.5 and Exp.6 with Exp.1, we can conclude that balancing the GSL2014 dataset by using more intermediate ice concentration samples from other datasets does improve the model performance even though the GSL2014 dataset is not fully balanced. One interesting finding is that extra intermediate ice concentration samples from other datasets do not necessarily have to be similar to the sea ice features in the GSL2014 dataset. This finding is supported by Exp.6, which obtained the lowest root mean squared error in this study. In addition, the number of training samples used for Exp.4 almost doubled the number of training samples used in Exp.1, while the number of training samples used for Exp.5 is just around 1/6 more than the number of training samples used in Exp.1, but the E_{rms} of Exp.4 and Exp.5 are almost the same (0.1563 and 0.1562 respectively). This means that the decreased percentage of E_{rms} in Exp.4 compared to Exp.1 is not because of the rise of the number of training samples, it is because the GSL2014 dataset is balanced to some extent.

According to the experimental results in Table 4.6, we can see that addressing imbalance in the dataset does make a difference on the CNN model performance. The root mean squared error of the sea ice concentration from the CNN is reduced when the balance in the dataset is improved using samples from other dataset. From Table 4.6, it can be seen that using data augmentation of minority classes did not show clear advantage of decreasing the root mean squared error.

After running two Wilcoxon signed-rank tests, one test for Exp.1 and Exp.5, the other test for Exp.1 and Exp.6, the P-values are less than the significance level alpha = 0.05. We can conclude that the median of Exp.1 is significantly different from the medians of Exp.5 and Exp.6.

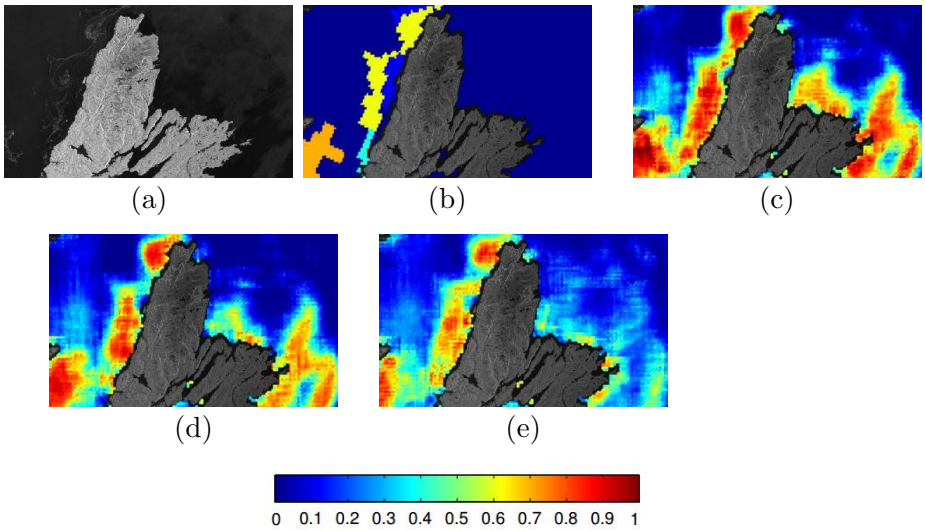
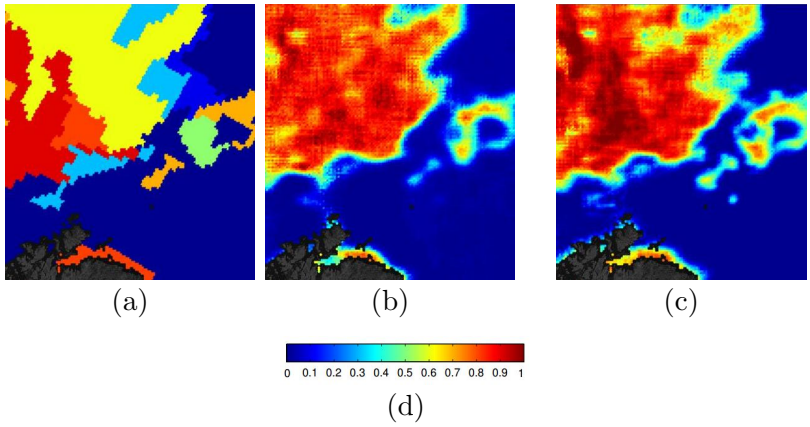


Figure 4.9: Visulization of the improvement by using different training datasets. These are sub-images of scene ID: 20140118_101002. (a) SAR imagery in HH pol; (b) Image analysis chart;(c) Prediction results when using only the GSL2014 dataset; (d) Prediction results when using the GSL2014 dataset and intermediate ice concentration samples in the GSL2012LS2011 dataset; (e) Prediction results when using the GSL2014 and intermediate ice concentration samples in the CAA2013 dataset.

Figure 4.9 shows how the CNN model performance differs with respect to the experiments with different training datasets. It can be observed that some regions identified as water in the image analysis are predicted as high ice concentration in Figure 4.9c. After adding more marginal ice samples (results are shown in Figure 4.9d and Figure 4.9e),



(a) (b) (c)

(d)

Figure 4.10: An example of results visulization of different experiments using different training datasets. (a) Image analysis chart (ground truth). (b) Prediction result when using only GSL2014. (c) Prediction results when using GSL2014 and intermediate ice samples in GSL2012LS2011.

improvements can be observed as those regions are no longer predicted as high ice concentration, but predicted as low ice concentrations, which reduces the estimation error.

Figure 4.10 shows a region with multiple sea ice concentration categories. This is a sub-image from scene ID: 20140122_095247. It can be seen that different sea ice concentration categories can be distinguished more clearly in Figure 4.10c than that in Figure 4.10b. In addition, it can be seen from the Figure 4.10 that the estimation results near the boundary of water and ice tend to be inaccurate. This ambiguity is due to the way of extracting patches and assigning labels. Results from other SAR scenes in test sets can be seen in Figure B.1 –B.3.

As can be seen in Figure 4.11, another benefit of using relatively balanced dataset to train CNN model is that the standard deviation of the errors for each ice concentration category are lower. Comparing Figure 4.11a and Figure 4.11b, we can conclude that if we just solely increase the number of training samples by using the whole GSL2014 dataset and the whole GSL2012LS2011 dataset instead of focusing on balancing the imbalanced GSL2014 dataset, then the standard deviation and mean of each ice concentration category remain almost unchanged. There is a significant improvement of the estimated mean value of the errors from the case where ice concentration equals 40%, the estimated mean is closer to the true mean when the GSL2014 dataset is more balanced.

To further investigate the impact of introducing more intermediate ice concentration

samples, we evaluate the prediction of marginal ice zone (MIZ). We threshold ice concentrations into three classes:
category I: Water ($IC <= 0.1$)
category II: MIZ($0.1 < IC <= 0.8$)
category III: Ice($IC > 0.8$)

After thresholding ground truth and ice concentration estimation from the CNNs, Confusion matrices for each experiments in this study are calculated as shown in Table 4.7. Note that these confusion matrices are computed from five test SAR scenes in fold 1 as shown in Table 4.4. The first row is the predicted labels for the three categories, while the first column is the true labels for the three categories. It can be seen that increasing the number of intermediate ice concentration samples in the training data can increase the classification accuracy of MIZ. In addition, It can be found that most of the misclassified MIZ samples are classified as ice from the experiemts in this study, just a few of them are misclassified as water. The classification accuracy of ice and water are equally good in most of the experiments in this study.

In order to compare with other people's results, we also calculated confusion matrices for the results from ASI. In contrast to the confusion matrix of ASI, it can be found that our CNN models have better ability to distinguish marginal ice zone.

Table 4.7: Confusion matrices for the experiments carried out in this study, in addition to those from ASI. Results shown here are for fold 1 test data only, where training is done using fold 2-5.

	Water	MIZ	ICE
	Exp.1: GSL2014		
Water	0.82	0.17	0.01
MIZ	0.02	0.70	0.28
ICE	0.00	0.13	0.87
	Exp.2: HLICs of GSL2014		
Water	0.72	0.27	0.01
MIZ	0.09	0.61	0.30
ICE	0.01	0.15	0.84
	Exp.3: GSL2014+data augmentation of intermediate ICs		
Water	0.76	0.23	0.01
MIZ	0.01	0.78	0.21
ICE	0.00	0.24	0.76
	Exp.4: GSL2014 + GSL2012LS2011		
Water	0.85	0.14	0.01
MIZ	0.03	0.73	0.24
ICE	0.00	0.16	0.84
	Exp.5: GSL2014 + intermediate ICs of GSL2012LS2011		
Water	0.81	0.18	0.01
MIZ	0.02	0.74	0.24
ICE	0.00	0.14	0.86
	Exp.6: GSL2014+ intermediate ICs of CAA2013		
Water	0.75	0.25	0.00
MIZ	0.03	0.71	0.27
ICE	0.00	0.15	0.85
	ASI		
Water	0.95	0.05	0.00
MIZ	0.67	0.31	0.02
ICE	0.19	0.57	0.24

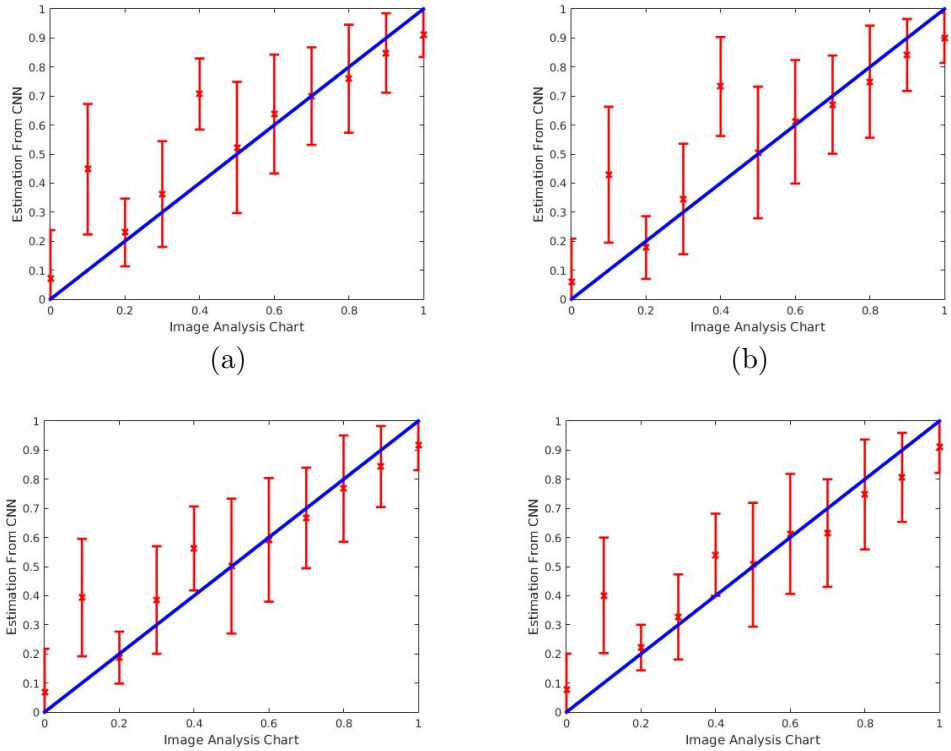


Figure 4.11: These figures show the estimation results from CNN models verse ice analysis charts from experiments using different datasets. Results shown here are for fold 1 test data only, where training is done using fold 2-5.(a) using only GSL2014; (b) using GSL2014 and GSL2012LS2011; (c) using GSL2014 and intermediate ice concentration samples in GSL2012LS2011; (d) using GSL2014 and intermediate ice concentration samples in CAA2013.

Chapter 5

Conclusion

5.1 Summary

The two main topics of this thesis are sea ice and open water classification, and sea ice concentration estimation, from dual-pol SAR scenes. To carry out these tasks, we adopted CNN-based methods and explored how the training data could affect CNN performance.

Transfer learning from natural image datasets (ImageNet) to SAR image datasets has shown desirable outcomes according to the sea ice and open water classification studies in chapter 3. The studies also show that the model trained on ImageNet has good generalization ability to remote sensing images. The studies in chapter 3 adopted AlexNet as a feature extractor and utilized softmax as a classifier to conduct sea ice and water classification of SAR imagery. The method that we adopted performs quite well with an overall classification accuracy of 92.36% through applying leave-one-out method, which enables evaluation of the model against unseen data.

It is known that CNNs can solve image classification problems very well. The studies in chapter 4 presents another functionality of CNNs; CNNs can also be used to solve regression problems [25, 76, 53]. A seven-layer CNN was developed for the problem of sea ice concentration estimation from SAR imagery. The method obtained an overall root mean squared error of 0.1917 in a five fold cross validation experiment. To compare with the results from Wang [75], the root mean squared error from the cross-validation experiment of using AICs to train the CNN in 4.4, calculated using the same test set as that used by Wang, is 0.1882 while root mean squared error from the CNN that Wang implemented is 0.2214. Note that the specific SAR scenes used as test set can be found

in Wang's paper[75]. Based on this CNN approach, experiments were conducted in order to better understand how the different training data set can be used to improve estimates of intermediate ice concentrations (or the marginal ice zone). Our original dataset is imbalanced because it lacks intermediate sea ice concentration categories. To reduce the imbalance, training samples with intermediate ice concentration from four datasets are introduced, and the results show this is very helpful in reducing the estimation errors of intermediate ice concentration samples. The results also reveal that the newly introduced training samples do not necessarily have similar sea ice features with the original dataset, which is meaningful as we may use more sea ice samples regardless of where SAR images were captured from. In addition, the experiments in this study prove that the CNN model have a better ability to identify marginal ice zone compared to other methods.

5.2 Future Work

There are several possible aspects that could be investigated to further develop to approaches discussed in this thesis. In terms of sea ice and open water classification studies, there are more advanced and well-trained CNN models available online than AlexNet. Implementing one of these models in a transfer learning approach might achieve a better performance than using AlexNet. In addition, ensemble learning is also a good option. Bagging and boosting techniques can be adopted to boost the model performance as well, but this requires higher GPU computational power.

As explained before, there are some errors in image analysis charts due to human visual interpretation of SAR imagery, which means our ground truth may not be that accurate. If these errors can be identified before training the CNNs, then the training data being supplied to the model would be of higher accuracy. However, validation of intermediate ice concentration data is challenging.

Deep CNNs have shown their strong power on image classification tasks according to recent research. It is of interest to train a deep CNN from scratch when a considerably large SAR imagery dataset and its corresponding image analysis charts are available. When the depth of commonly used CNNs getting deeper, more weights need to be learned. Thus, more training samples are required. Furthermore, we should consider the trade-off between the deeper CNNs and the number of training samples when designing a CNN structure. Theoretically, the more training data the better, as limited dataset may result in overfitting [65].

Intermediate ice concentrations categories are always the minority categories in sea ice studies due to the nature of sea ice cover. These imbalanced datasets are a challenge

when automated methods are applied that are expecting a balanced dataset. Although some experiments in chapter 4 have demonstrated that even a small improvement in the dataset balance can boost model performance, we still do not have sufficient intermediate ice concentration samples.

The HV polarization images used for this study can be refined if one can figure out a method to remove the vertical stripes in HV images. These vertical stripes occur in almost every HV image used for this study, and the CNN tends to consider this as a feature even though these stripes do not represent useful information for the sea ice concentration estimation problem. These stripes come from sensor noise and therefore it is difficult to eliminate them. Hybrid-polarity SAR architecture will be included in future SAR missions such as the Canadian RADARSAT Constellation Mission (RCM) [12]. These sensors will enable the use of compact polarimetry (CP) data in wide swath imagery [12]. Therefore, dual-pol SAR imagery may not be widely used in the future.

Bibliography

[1] Matt Arkett, Dean Flett, and Roger De Abreu. C-band multiple polarization SAR for ice monitoring–What can it do for the Canadian ice service. *Envisat Symposium 2007*, SP(636), 2007.

[2] Nicolas Audebert, Bertrand Le Saux, and Sébastien Lefèvre. Semantic segmentation of earth observation data using multimodal and multi-scale deep networks. In *Asian Conference on Computer Vision*, pages 180–196. Springer, 2016.

[3] Frank D Carsey. *Microwave remote sensing of sea ice*. American Geophysical Union, 1992.

[4] Marco Castelluccio, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva. Land use classification in remote sensing images by convolutional neural networks. *arXiv preprint arXiv:1508.00092*, 2015.

[5] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[6] Giovanni Chierchia, Davide Cozzolino, Giovanni Poggi, and Luisa Verdoliva. SAR image despeckling through convolutional neural networks. *arXiv preprint arXiv:1704.00275*, 2017.

[7] Josefino C Comiso. Large decadal decline of the Arctic multiyear ice cover. *Journal of Climate*, 25(4):1176–1193, 2012.

[8] Wikipedia contributors. Gulf of saint lawrence — wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Gulf_of_Saint_Lawrence&oldid=812484449.

[9] Wikipedia contributors. Sea ice — wikipedia, the free encyclopedia, 2017. https://en.wikipedia.org/w/index.php?title=Sea_ice&oldid=815289206.

[10] Wikipedia contributors. Synthetic-aperture radar — wikipedia, the free encyclopedia, [Online; accessed 12-December-2017]. https://en.wikipedia.org/w/index.php?title=Synthetic-aperture_radar&oldid=812953600.

[11] Wikipedia contributors. Wilcoxon signed-rank test — wikipedia, the free encyclopedia, [Online; accessed 29-March-2018]. https://en.wikipedia.org/w/index.php?title=Wilcoxon_signed-rank_test&oldid=827681159.

[12] M Dabboor and T Geldsetzer. Towards sea ice classification using simulated RADARSAT Constellation Mission compact polarimetric SAR imagery. *Remote sensing of environment*, 140:189–195, 2014.

[13] Jun Ding, Bo Chen, Hongwei Liu, and Mengyuan Huang. Convolutional neural network with data augmentation for SAR target recognition. *IEEE Geoscience and Remote Sensing Letters*, 13(3):364–368, 2016.

[14] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In *International conference on machine learning*, pages 647–655, 2014.

[15] Environment and Climate Change Canada. Ice chart descriptions. https://www.canada.ca/en/environment-climate-change/services/ice-forecasts-observations/latest-conditions/products-guides/chart-descriptions.html#iac, 2016.

[16] Environment and Climate Change Canada. Sea ice climatic atlas for the east coast 1981-2010. https://www.canada.ca/en/environment-climate-change/services/ice-forecasts-observations/publications/sea-climatic-atlas-east-coast-1981-2010/chapter-2.html, 2016.

[17] Environment and Climate Change Canada. Synthetic aperture radar. http://www.nrcan.gc.ca/earth-sciences/geomatics/satellite-imagery-air-photos/sensors-methods/synthetic-aperture-radar/10968, 2016.

[18] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[19] Hyangsun Han, Sang-Hoon Hong, Hyun-cheol Kim, Tae-Byeong Chae, and Hae-Jin Choi. A study of the feasibility of using KOMPSAT-5 SAR data to map sea ice in the Chukchi Sea in late summer. *Remote Sensing Letters*, 8(5):468–477, 2017.

[20] Hyangsun Han and Hyun-cheol Kim. Evaluation of summer passive microwave sea ice concentrations in the Chukchi Sea based on KOMPSAT-5 SAR and numerical weather prediction data. *Remote Sensing of Environment*, 209:343–362, 2018.

[21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[22] Fan Hu, Gui-Song Xia, Jingwen Hu, and Liangpei Zhang. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing*, 7(11):14680–14707, 2015.

[23] Zhongling Huang, Zongxu Pan, and Bin Lei. Transfer Learning with Deep Convolutional Neural Network for SAR Target Classification with Limited Labeled Data. *Remote Sensing*, 9(9):907, 2017.

[24] Norwegian Polar Institute. The marginal ice zone, [Online; accessed 02-May-2017]. http://www.npolar.no/en/facts/the-marginal-ice-zone.html.

[25] Aaron S Jackson, Adrian Bulat, Vasileios Argyriou, and Georgios Tzimiropoulos. Large pose 3D face reconstruction from a single image via direct volumetric CNN regression. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 1031–1039. IEEE, 2017.

[26] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional Architecture for Fast Feature Embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[27] Rishi Jobanputra and David A Clausi. Preserving boundaries for image texture segmentation using grey level co-occurring probabilities. *Pattern Recognition*, 39(2):234–245, 2006.

[28] Nal Kalchbrenner, Lasse Espeholt, Karen Simonyan, Aaron van den Oord, Alex Graves, and Koray Kavukcuoglu. Neural machine translation in linear time. *arXiv preprint arXiv:1610.10099*, 2016.

[29] Michael Kampffmeyer, Arnt-Børre Salberg, and Robert Jenssen. Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2016 IEEE Conference on*, pages 680–688. IEEE, 2016.

[30] Miao Kang, Xiangguang Leng, Zhao Lin, and Kefeng Ji. A modified faster R-CNN based on CFAR algorithm for SAR ship detection. In *Remote Sensing with Intelligent Processing (RSIP), 2017 International Workshop on*, pages 1–4. IEEE, 2017.

[31] Andrej Karpathy. Cs231n lecture notes: Neural networks part 1: Setting up the architecture. http://cs231n.github.io/convolutional-networks/.

[32] Andrej Karpathy. Cs231n lecture notes: Quick intro without brain analogies. http://cs231n.github.io/neural-networks-1/.

[33] Andrej Karpathy. Cs231n lecture notes: Transfer learning. http://cs231n.github.io/transfer-learning/.

[34] Andrej Karpathy. Cs231n lecture notes:networks for visual recognition. http://cs231n.github.io/linear-classify/.

[35] Juha Karvonen. Baltic sea ice concentration estimation based on C-band HH-polarized SAR data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(6):1874–1884, 2012.

[36] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[37] Alexander S Komarov and Mark Buehner. Automated Detection of Ice and Open Water From Dual-Polarization RADARSAT-2 Images for Data Assimilation. *IEEE Transactions on Geoscience and Remote Sensing*, 55(10):5755–5769, 2017.

[38] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[39] Seymour W Laxon, Katharine A Giles, Andy L Ridout, Duncan J Wingham, Rosemary Willatt, Robert Cullen, Ron Kwok, Axel Schweiger, Jinlun Zhang, Christian Haas, et al. CryoSat-2 estimates of Arctic sea ice thickness and volume. *Geophysical Research Letters*, 40(4):732–737, 2013.

[40] Yann LeCun. Deep Learning Tutorial: ICML, Atlanta, 2013-06-16, 2013[Online]. https://cs.nyu.edu/~yann/talks/lecun-ranzato-icml2013.pdf.

[41] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

[42] Steven Leigh. Automated Ice-Water Classification using Dual Polarization SAR Imagery. Master's thesis, 2013.

[43] Steven Leigh, Zhijie Wang, and David A Clausi. Automated ice–water classification using dual polarization SAR satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 52(9):5529–5539, 2014.

[44] Fan Li and Graham Taylor. Alter-cnn: An approach to learning from label proportions with application to ice-water classification. In *Neural Information Processing Systems Workshops (NIPSW) on Learning and privacy with incomplete data and weak supervision*, 2015.

[45] Huiying Liu, Huadong Guo, and Lu Zhang. SVM-based sea ice classification using textural features and concentration from RADARSAT-2 Dual-Pol ScanSAR data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(4):1601–1613, 2015.

[46] Xu-Ying Liu and Zhi-Hua Zhou. The influence of class imbalance on cost-sensitive learning: An empirical study. In *Data Mining, 2006. ICDM'06. Sixth International Conference on*, pages 970–974. IEEE, 2006.

[47] Rushi Longadge and Snehalata Dongre. Class imbalance problem in data mining review. *arXiv preprint arXiv:1305.1707*, 2013.

[48] Emmanuel Maggiori, Yuliya Tarabalka, Guillaume Charpiat, and Pierre Alliez. Fully convolutional neural networks for remote sensing image classification. In *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*, pages 5071–5074. IEEE, 2016.

[49] MANICE. Manual of standard procedures for observing and reporting ice conditions. 1994.

[50] Giuseppe Masi, Davide Cozzolino, Luisa Verdoliva, and Giuseppe Scarpa. Cnn-based pansharpening of multi-resolution remote-sensing images. In *Urban Remote Sensing Event (JURSE), 2017 Joint*, pages 1–4. IEEE, 2017.

[51] Masakazu Matsugu, Katsuhiko Mori, Yusuke Mitari, and Yuji Kaneda. Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks*, 16(5-6):555–559, 2003.

[52] Walter N Meier, Greta K Hovelsrud, Bob EH Oort, Jeffrey R Key, Kit M Kovacs, Christine Michel, Christian Haas, Mats A Granskog, Sebastian Gerland, Donald K Perovich, et al. Arctic sea ice in transformation: A review of recent observed changes and impacts on biology and human activity. *Reviews of Geophysics*, 52(3):185–217, 2014.

[53] Shun Miao, Z Jane Wang, and Rui Liao. A CNN regression approach for real-time 2D/3D registration. *IEEE transactions on medical imaging*, 35(5):1352–1363, 2016.

[54] M-AN Moen, Anthony Paul Doulgeris, Stian Normann Anfinsen, Angelika HH Renner, Nick Hughes, Sebastian Gerland, and Torbjørn Eltoft. Comparison of feature based segmentation of full polarimetric sar satellite sea ice images with manually drawn ice charts. *The Cryosphere*, 7(6):1693, 2013.

[55] David AE Morgan. Deep convolutional neural networks for ATR from SAR imagery. *Proceedings of the Algorithms for Synthetic Aperture Radar Imagery XXII, Baltimore, MD, USA*, 23:94750F, 2015.

[56] Leif T Pedersen Natalia Ivanova and Rasmus Tonboe. *D2.5 Product Validation and Algorithm Selection Report (PVASR) Sea Ice Concentration.* SICCI-PVASR, 2013. http://ocean.dmi.dk/projects/SICCI/documents/SICCI_PVASR_D2.5_Issue1.1_20130828.pdf.

[57] Thomas Newman, Sinead L Farrell, Jacqueline Richter-Menge, Laurence N Connor, Nathan T Kurtz, Bruce C Elder, and David McAdoo. Assessment of radar-derived snow depth over Arctic sea ice. *Journal of Geophysical Research: Oceans*, 119(12):8578–8602, 2014.

[58] Andrew Ng. Unsupervised feature learning and deep learning tutorials: convolutional neural networks. http://ufldl.stanford.edu/tutorial/supervised/ConvolutionalNeuralNetwork/.

[59] Otávio AB Penatti, Keiller Nogueira, and Jefersson A dos Santos. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 44–51, 2015.

[60] D Ramyachitra and P Manikandan. Imbalanced dataset classification and solutions: a review. *International Journal of Computing and Business Research (IJCBR)*, 5(4), 2014.

[61] Martin Riedmiller and Heinrich Braun. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In *Neural Networks, 1993., IEEE International Conference on*, pages 586–591. IEEE, 1993.

[62] Steven K Rogers, John M Colombi, Curtis E Martin, James C Gainey, Ken H Fielding, Tom J Burns, Dennis W Ruck, Matthew Kabrisky, and Mark Oxley. Neural networks for automatic target recognition. *Neural networks*, 8(7):1153–1184, 1995.

[63] Stein Sandven, Ola M Johannessen, and Kjell Kloster. Sea ice monitoring by remote sensing. *Encyclopedia of Analytical Chemistry*, 2006.

[64] K Andrea Scott, Mark Buehner, and Tom Carrieres. An assessment of sea-ice thickness along the Labrador coast from AMSR-E and MODIS data for operational data assimilation. *IEEE Transactions on Geoscience and Remote Sensing*, 52(5):2726–2737, 2014.

[65] Hoo-Chang Shin, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, 35(5):1285–1298, 2016.

[66] Mohammed Shokr and Nirmal Sinha. *Sea ice: physics and remote sensing*. John Wiley & Sons, 2015.

[67] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[68] L-K Soh, Costas Tsatsoulis, Denise Gineris, and Cheryl Bertoia. ARKTOS: An intelligent system for SAR sea ice image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 42(1):229–248, 2004.

[69] Gunnar Spreen, Lars Kaleschke, and Georg Heygster. Sea ice remote sensing using AMSR-E 89-GHz channels. *Journal of Geophysical Research: Oceans*, 113(C2), 2008.

[70] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[71] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[72] Pang-Ning Tan et al. *Introduction To Data Mining*. Pearson Education India, 2006.

[73] Tijmen Tieleman and Geoffrey Hinton. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26–31, 2012.

[74] Jorge Urrego-Blanco and Jinyu Sheng. Formation and distribution of sea ice in the Gulf of St. Lawrence: A process-oriented study using a coupled ocean-ice model. *Journal of Geophysical Research: Oceans*, 119(10):7099–7122, 2014.

[75] Lei Wang. *Learning to Estimate Sea Ice Concentration from SAR Imagery*. PhD thesis, University of Waterloo, Systems Design Engineering Department, 2016. http://hdl.handle.net/10012/10954.

[76] Lei Wang, K Andrea Scott, Linlin Xu, and David A Clausi. Sea ice concentration estimation during melt from dual-pol SAR scenes using deep convolutional neural networks: A case study. *IEEE Transactions on Geoscience and Remote Sensing*, 54(8):4524–4533, 2016.

[77] Puyang Wang, He Zhang, and Vishal M Patel. SAR Image Despeckling Using a Convolutional Neural Network. *IEEE Signal Processing Letters*, 24(12):1763–1767, 2017.

[78] Wikipedia contributors. Sea ice concentration — Wikipedia, the free encyclopedia, 2017. https://en.wikipedia.org/w/index.php?title=Sea_ice_concentration&oldid=812786266.

[79] Wikipedia contributors. Wilcoxon signed-rank test — Wikipedia, the free encyclopedia, [Online; accessed 10-April-2018]. https://en.wikipedia.org/w/index.php?title=Wilcoxon_signed-rank_test&oldid=827681159.

[80] Wikipedia contributors. Cross-validation (statistics) — Wikipedia, the free encyclopedia, [Online; accessed 2-May-2018]. https://en.wikipedia.org/w/index.php?title=Cross-validation_(statistics)&oldid=834387091.

[81] Guo-bao Xu, Gui-yan Zhao, Yi-xin Yin, et al. A CNN-based edge detection algorithm for remote sensing image. In *Control and Decision Conference, 2008. CCDC 2008. Chinese*, pages 2558–2561. IEEE, 2008.

[82] Yan Xu and K. A Scott. Sea ice and open water classification of SAR imagery using CNN-based transfer learning. *International Geoscience and Remote Sensing Symposium*, 2017.

[83] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In *Advances in neural information processing systems*, pages 3320–3328, 2014.

[84] Natalia Zakhvatkina, Anton Korosov, Stefan Muckenhuber, Stein Sandven, and Mohamed Babiker. Operational algorithm for ice–water classification on dual-polarized radarsat-2 images. *The Cryosphere*, 11(1):33, 2017.

[85] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *Computer Vision – ECCV 2014*, pages 818–833. Springer, 2014.

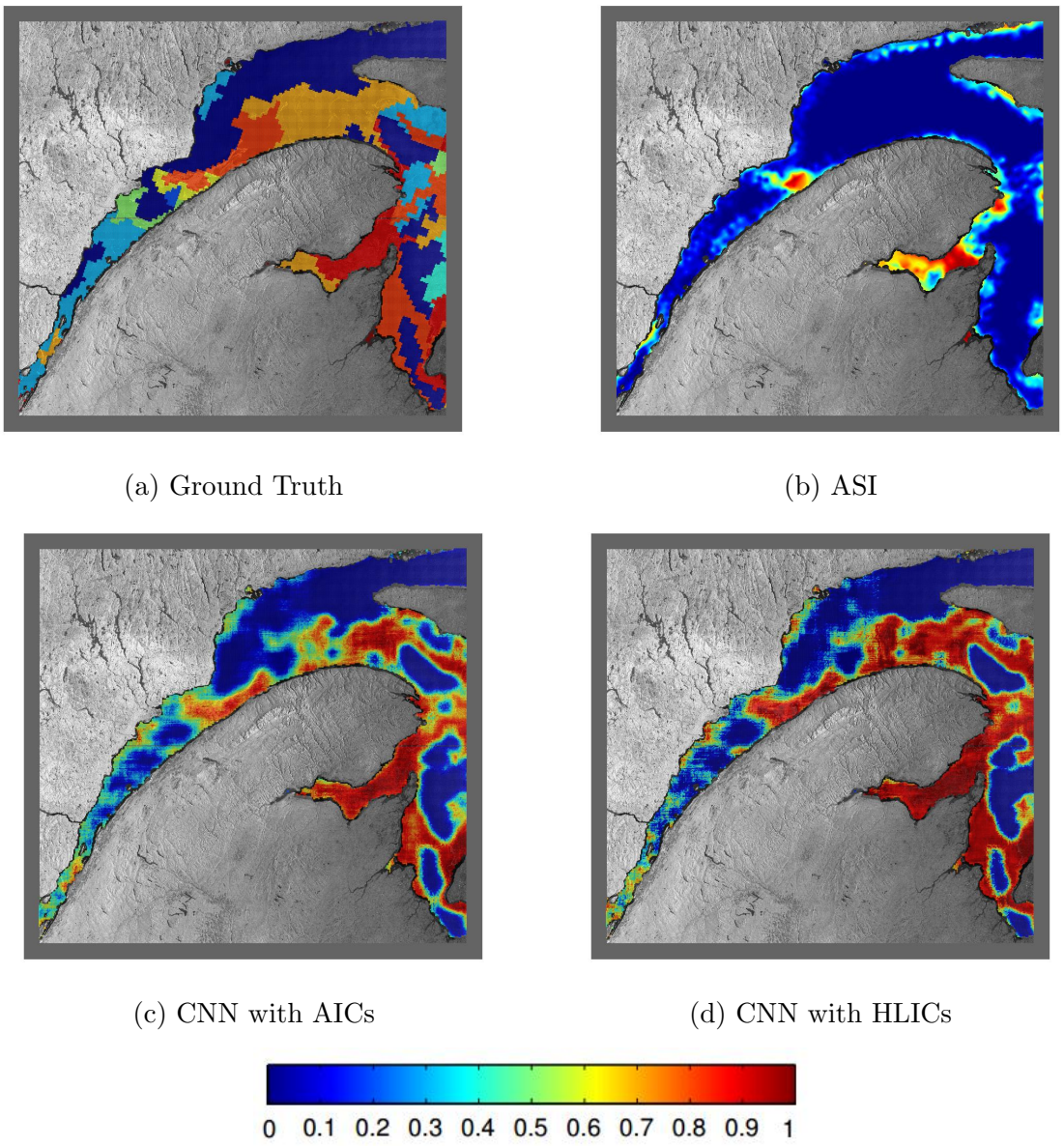
[86] Fan Zhang, Fabio Duarte, Ruixian Ma, Dimitrios Milioris, Hui Lin, and Carlo Ratti. Indoor space recognition using deep convolutional neural network: a case study at MIT campus. *arXiv preprint arXiv:1610.02414*, 2016.

[87] Zhihua Zhou. *Machine Learning*, volume 4. Tsinghua University Press, 2016.

Appendices

Appendix A

Results visualization from experiments using AICs and HLICs as training data



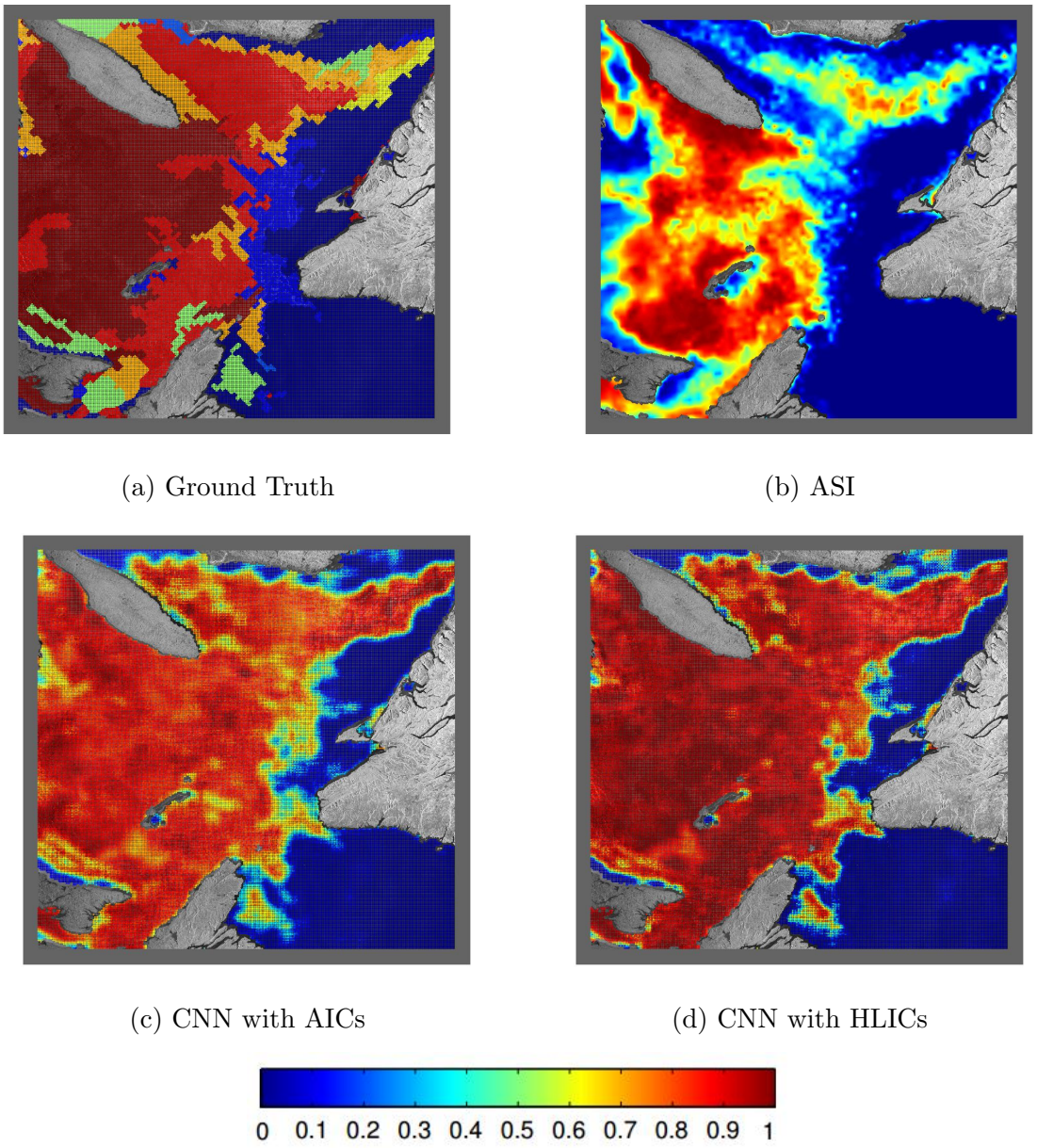
(a) Ground Truth

(b) ASI

(c) CNN with AICs

(d) CNN with HLICs

Figure A.1: Results visualization for 20140117_103914 from experiments using AICs and HLICs.



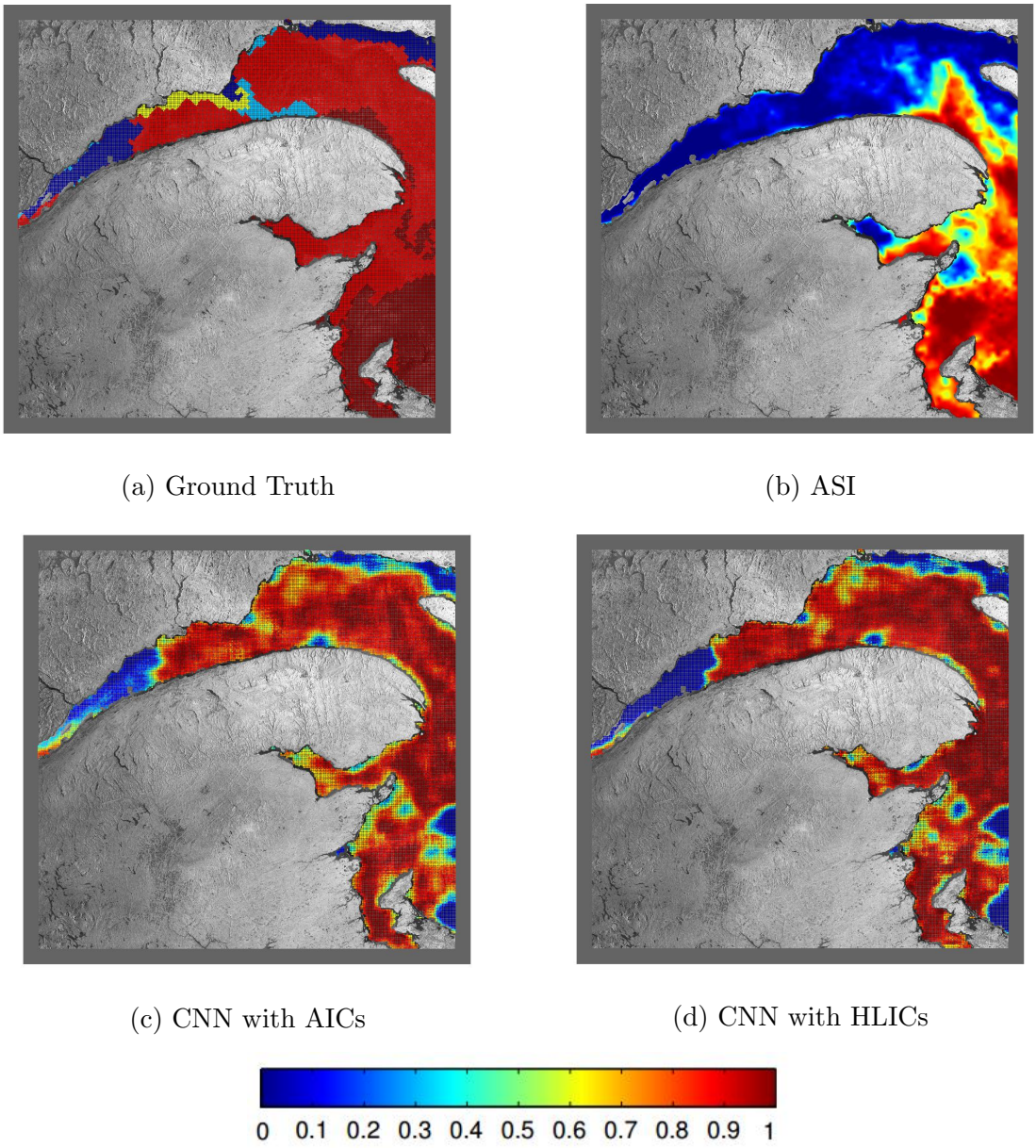
(a) Ground Truth

(b) ASI

(c) CNN with AICs

(d) CNN with HLICs

Figure A.2: Results visualization for 20140131_215240 from experiments using AICs and HLICs



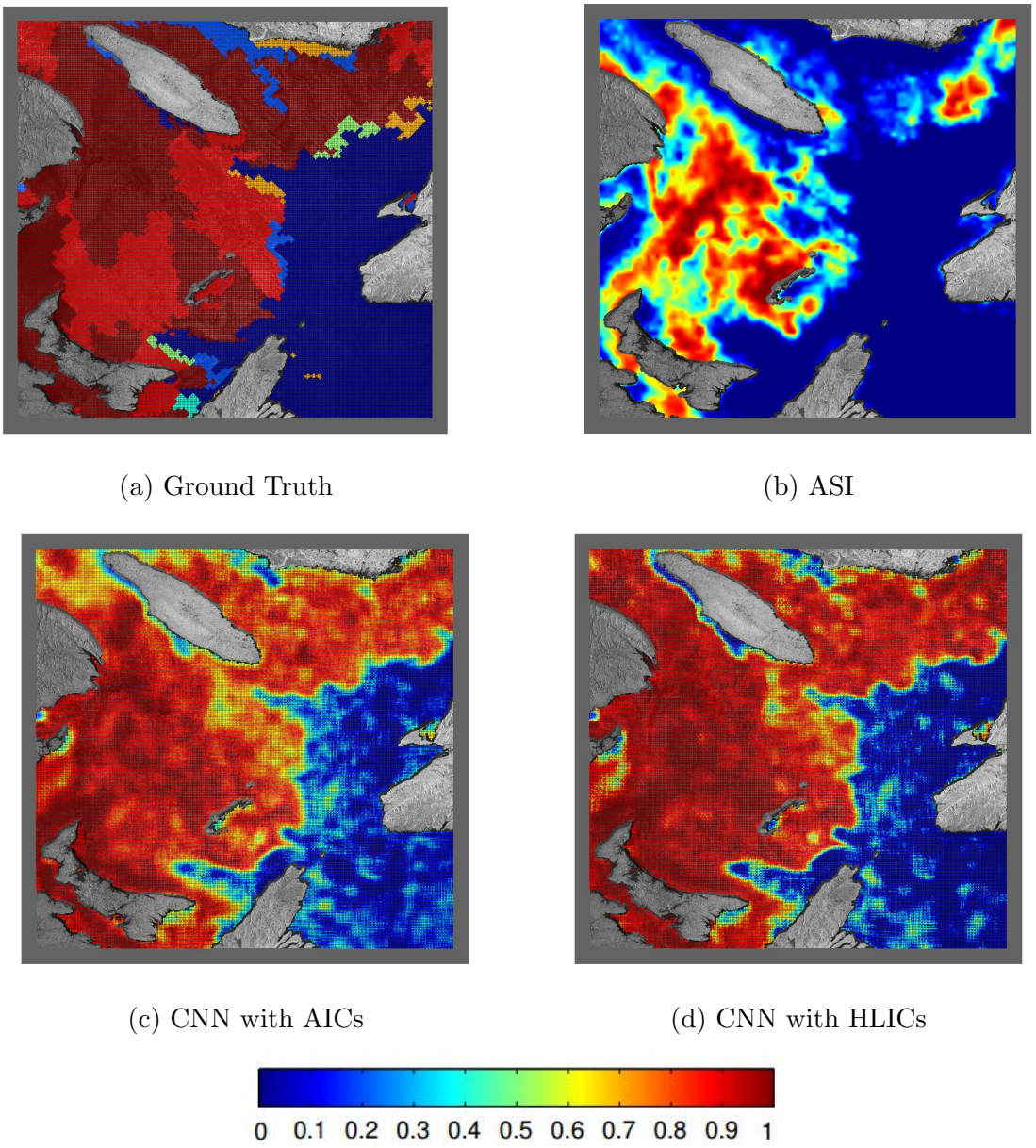
(a) Ground Truth

(b) ASI

(c) CNN with AICs

(d) CNN with HLICs

Figure A.3: Results visualization for 20140206_221744 from experiments using AICs and HLICs



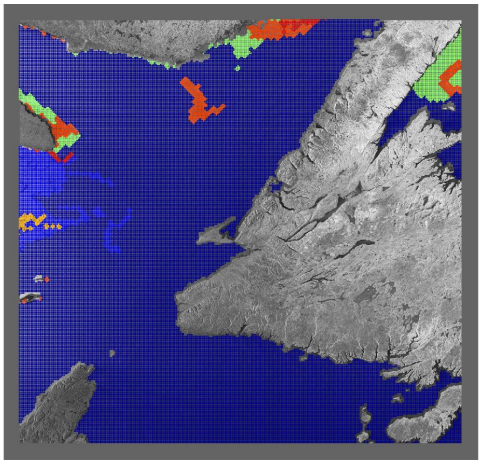
(a) Ground Truth

(b) ASI

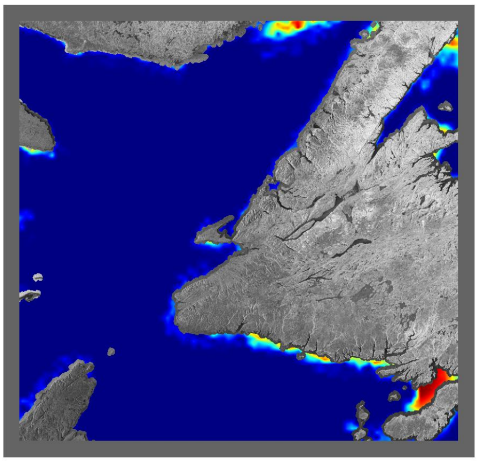
(c) CNN with AICs

(d) CNN with HLICs

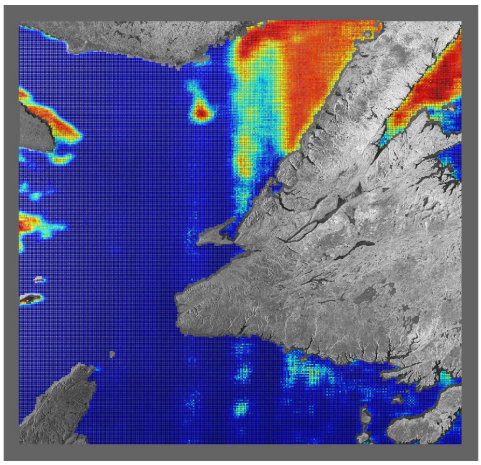
Figure A.4: Results visualization for 20140124_215646 from experiments using AICs and HLICs.



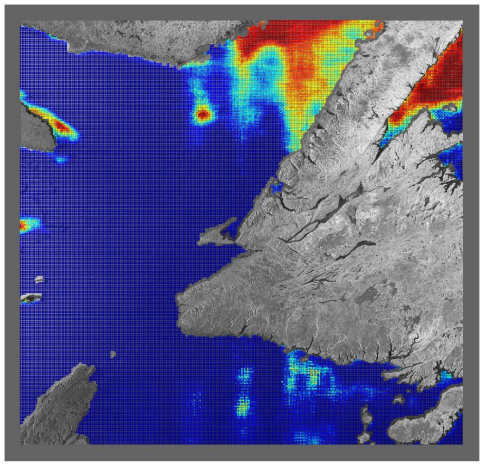
(a) Ground Truth



(b) ASI



(c) CNN with AICs



(d) CNN with HLICs

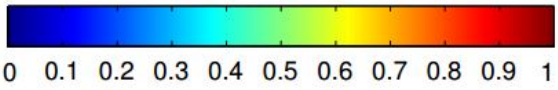
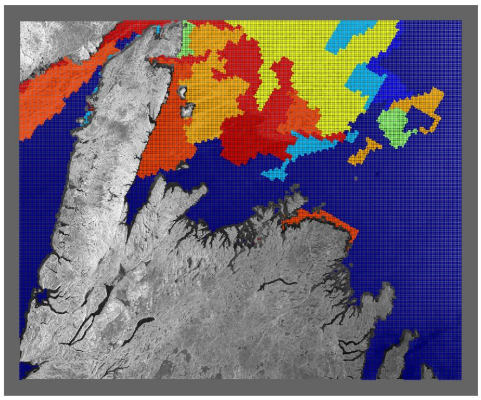
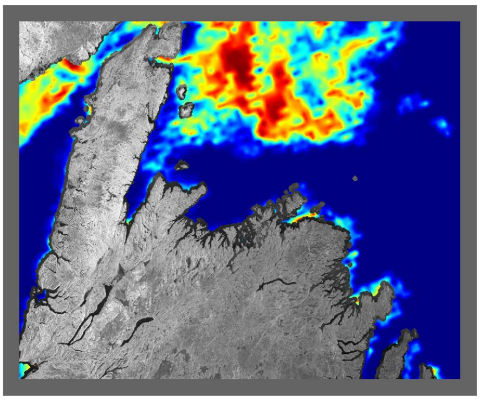


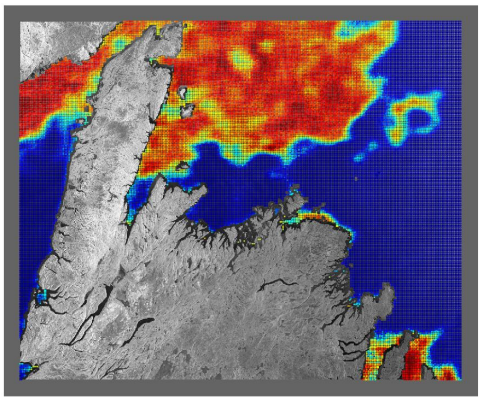
Figure A.5: Results visualization for 20140121_214420 from experiments using AICs and HLICs



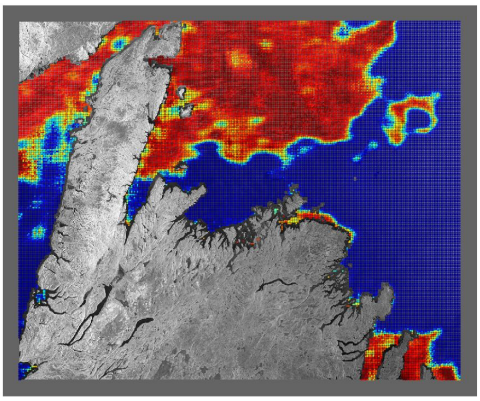
(a) Ground Truth



(b) ASI



(c) CNN with AICs



(d) CNN with HLICs

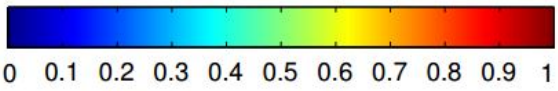


Figure A.6: Results visualization for 20140122_095247 from experiments using AICs and HLICs

Appendix B

Results visualization from experiments using different training datasets for fold 1

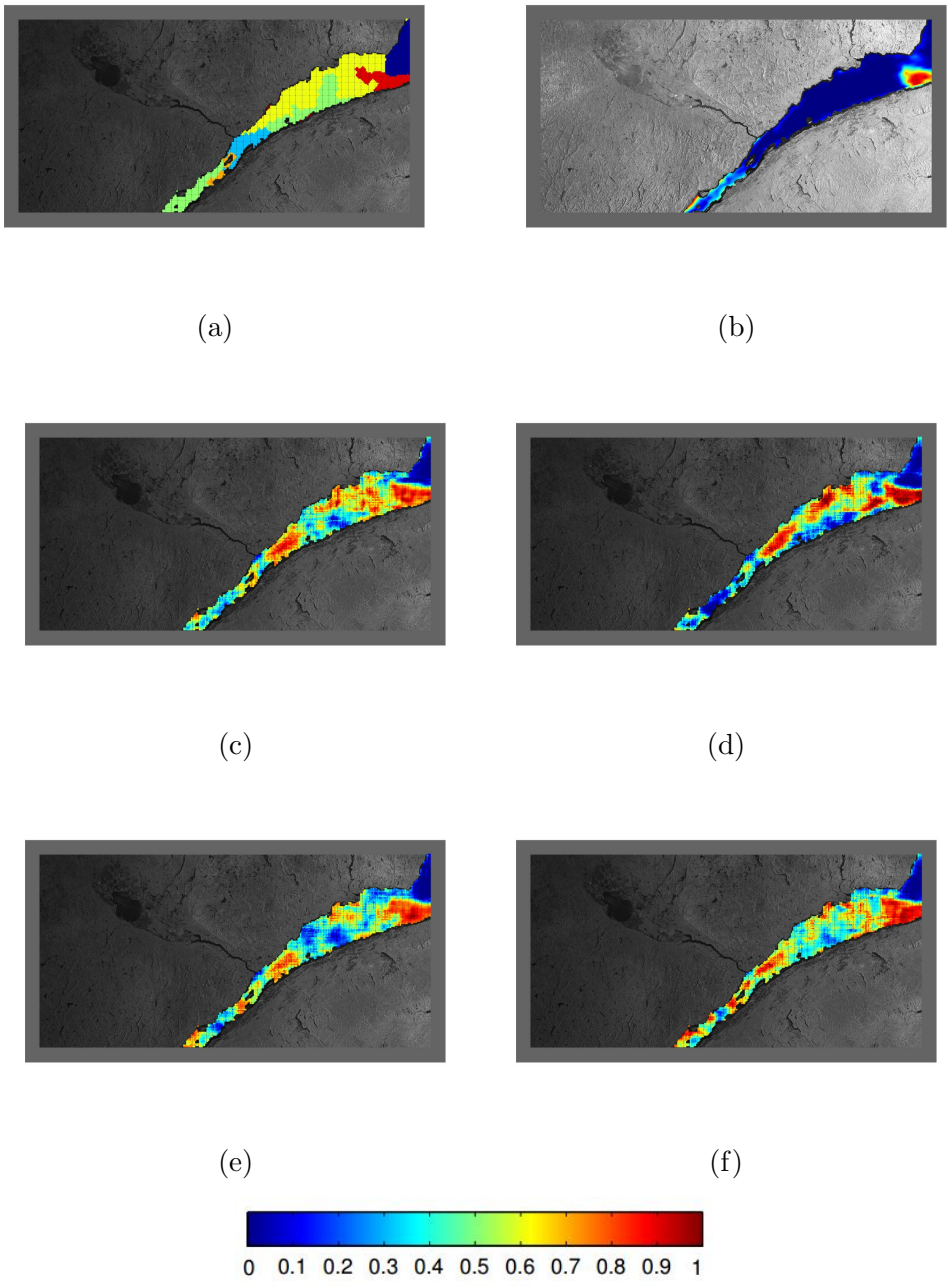


Figure B.1: Results visualization from experiments using different training datasets. Scene ID 20140116_223042. (a) ground truth, (b) ASI, (c) GSL2014, (d) HLICs of GSL2014, (e) GSL2014 + intermediate ICs of GSL2012LS2011, (f) GSL2014+ intermediate ICs of CAA2013.

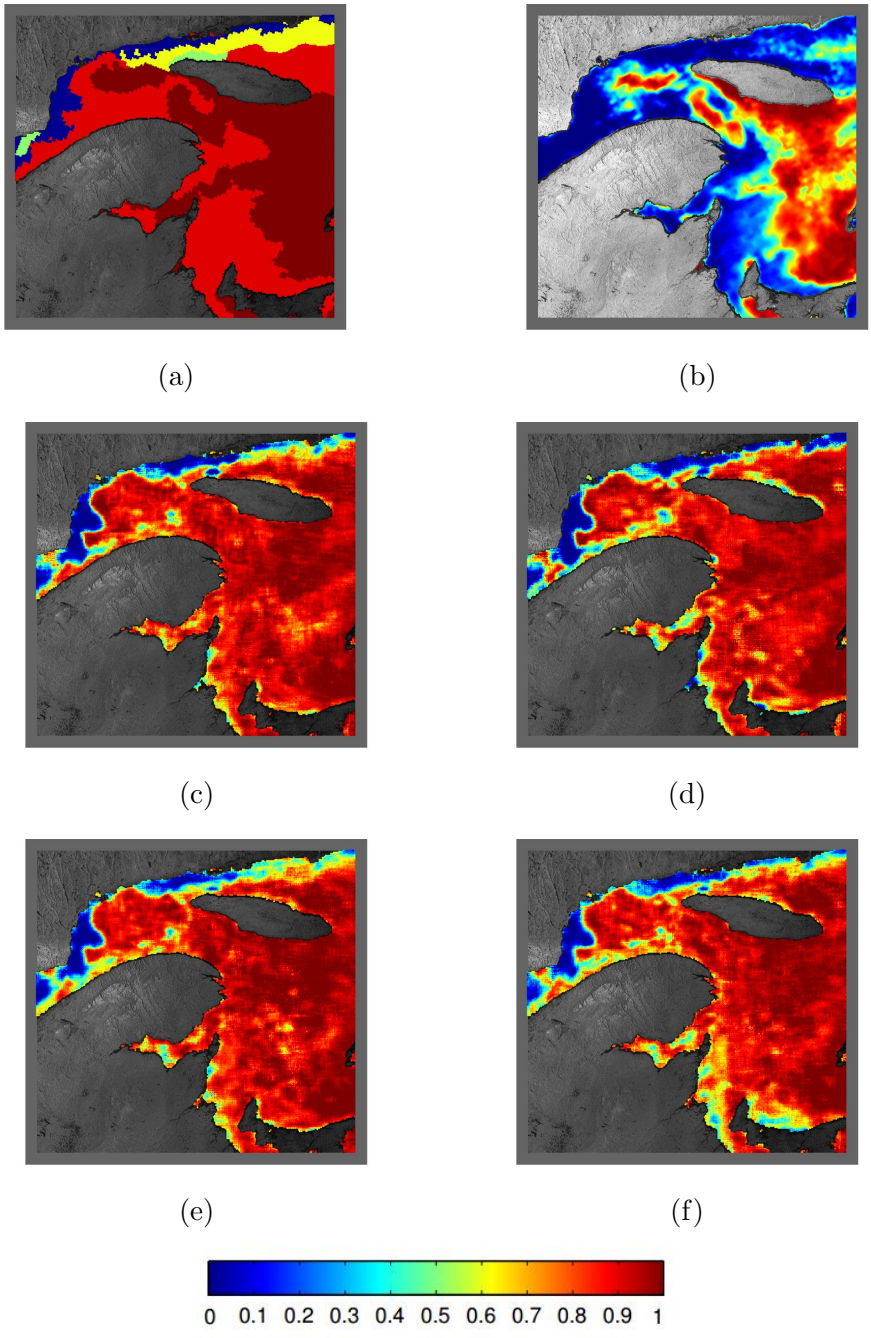


Figure B.2: Results visualization from experiments using different training datasets. Scene ID 20140131_103053. (a) ground truth, (b) ASI, (c) GSL2014, (d) HLICs of GSL2014, (e) GSL2014 + intermediate ICs of GSL2012LS2011, (f) GSL2014+ intermediate ICs of CAA2013.

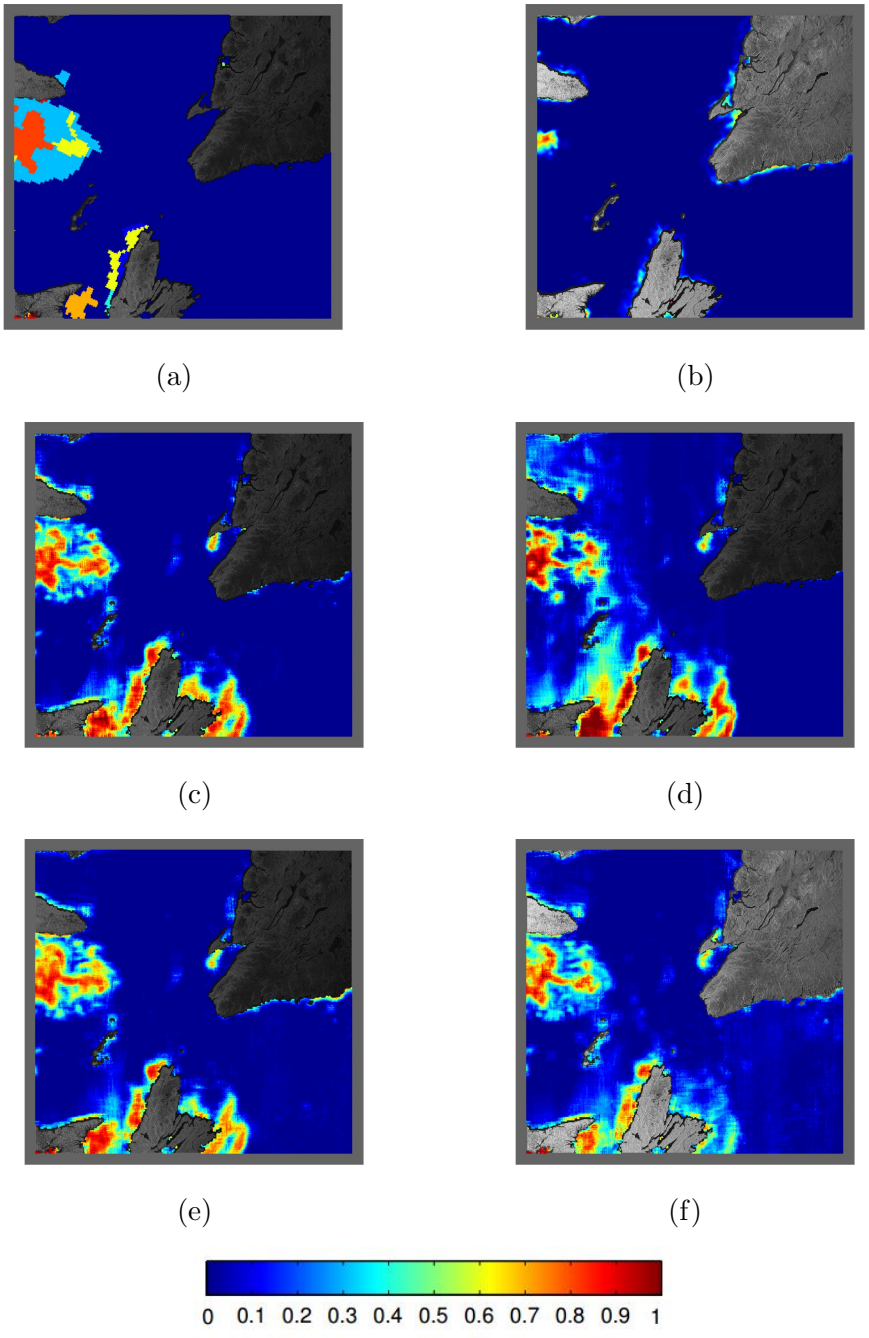


Figure B.3: Results visualization from experiments using different training datasets. Scene ID 20140118_101002. (a) ground truth, (b) ASI, (c) GSL2014, (d) HLICs of GSL2014, (e) GSL2014 + intermediate ICs of GSL2012LS2011, (f) GSL2014+ intermediate ICs of CAA2013.