Learning to Estimate Sea Ice Concentration from SAR Imagery

by

Lei Wang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2016

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

Through the growing interest in the Arctic for shipping, mining and climate research, large-scale high quality ice concentration is of great interest. Due to the unavailability of suitable ice concentration estimation algorithms, ice concentration maps are interpreted from synthetic aperture radar (SAR) images manually by ice experts for operational uses. An automatic ice concentration estimation algorithm is required for accurate large-scale ice mapping. In this thesis, a set of algorithms are developed aiming at operational ice concentration estimation from SAR images.

The major difficulty in designing a robust algorithm for ice concentration estimation from SAR images is the constantly changing SAR image features of ice and water in time and location. This difficulty is addressed by learning features instead of designing features from SAR images. A set of convolutional neural network based ice concentration estimation algorithms are developed to learn multi-scale SAR image features and simultaneously regress ice concentration from the learned image features. We first demonstrated the capability of CNNs in ice concentration estimation from SAR images when trained using image analysis charts as ground truth. Then the model is further improved by taking into account the errors in the image analysis charts. Ice concentration estimates with improved robustness to training samples errors, accuracy and scale of details are obtained. The robustness of the developed methods are further demonstrated in the melt season of the Beaufort Sea, where reasonable ice concentration estimates are acquired. In order to reduce the model training time, it is desired to reuse existing models. The model transferability is evaluated and suggestions on using existing models to accelerate the training process are given, which is shown to reduce the training time by over 10 times in our case.

Acknowledgments

I would like to express my gratitude to my co-supervisors, professor David Clausi and professor Andrea Scott. Thank you for been extremely patient and helpful through out the last four years. Your have taught me how to grow as a rigorous researcher.

I would sincerely like to thank my Ph.D. committee members Dr. David Clausi, Dr. Andrea Scott, Dr. Paul Fieguth and Dr. Kumaraswamy Ponnambalam from the department of System Design Engineering at the University of Waterloo, Dr. Pascal Poupart from the department of Computer Science at the University of Waterloo and Dr. Leif Toudal Pedersen from Danmarks Meteorologiske Institut. Your insight and acknowledgment are very helpful and encouraging.

Many thanks to my colleagues in the Vision and Image Processing lab at the University of Waterloo. People here are knowledgeable and friendly. I have learned so much from them. Many thanks to Alex Wong, Linlin Xu and Fan Li. They have contributed many valuable ideas and discussions to the thesis.

Lastly, and most importantly, I would like to thank my family for all of their support and endless love. Special thanks to Zhenyi Gao. I am lucky to have your company in the last two years.

Thank you.

Dedication

This is dedicated to my family.

Table of Contents

List of Tables

List of Figures

Nomenclature

α Momentum

$\bar{\boldsymbol{y}}$ Mean of $\boldsymbol{y}$

$\boldsymbol{\theta}'$ Neural network weights of the previous EM iteration

$\boldsymbol{\theta}_m$ All weights of a neural network at iteration m

$\boldsymbol{C^k}$ kth convolutional filter in a convolutional layer

$\boldsymbol{h^k}_{ij}$ Output of convolutional layer for the kth convolutional filter at pixel location i, j

$\boldsymbol{V}_t$ Update of $\boldsymbol{\theta}$ at iteration t in SGD

$\boldsymbol{W}$ Weight matrix of a fully connected layer

$\boldsymbol{x}$ Matrix

$\boldsymbol{z}$ Image analysis

ϵ Learning rate

$\mathbb{L}$ Loss

ρ Scaling factor in EM based approach

σ Standard deviation

b Bias term

C_x Width of $\boldsymbol{C^k}$

C_y Height of $\boldsymbol{C^k}$

E_{L1} Mean absolute error

E_{rmse} Root Mean squared error

E_{sgn} Mean error

E_{std} Error standard deviation

F Function of the network

f Activation function

K Number of convolutional filters in a convolutional layer

L Distance function

M Number of pixels in $\boldsymbol{z}$

M_i Number of pixels with image analysis i

M_x Width of layer output

M_y Height of layer output

P Step size of convolutional operations

Q Log-likelihood of data

r Weight decay

S_x Width of $\boldsymbol{x}$

S_y Height of $\boldsymbol{x}$

S_z Number of $\boldsymbol{x}$

Chapter 1

Introduction

Sea ice covers about 7% of the Earth's surface [84]. It has a significant influence on the environment and global climate [88]. Sea ice acts like a blanket on the ocean surface, reducing the absorbed solar energy, evaporation and heat exchange between the ocean and atmosphere, which makes it an important component in climate feedback mechanisms. Sea ice also affects the movement of ocean water by releasing salt into water. Water with more salt content is denser and sinks to the bottom of the ocean, which contributes to ocean water circulation [1]. Sea ice also has an influence on the wildlife and the society of northern communities by affecting the fresh water supply, hunting and shipping [36]. Therefore, accurate ice information is critical for the understanding of the Arctic environment and global climate change.

There is increasing evidence showing that sea ice extent in the Arctic is decreasing [87]. The decline in ice extent is creating opportunities for new sea routes and the exploration of natural resources in the Arctic. The Arctic shipping routes provide alternative seasonal shipping lanes that connect the Atlantic and Pacific Oceans, which are gaining attraction in recent years. The Arctic region is estimated to have 22% of the world's oil and natural gas reserves [32], and a large portion of the natural resources in the Arctic are offshore and unexplored [61]. More industrial activities are expected in this region in the future. The need for timely and accurate ice information is increasing to support the increasing human activities in the Arctic [88].

Satellite sensors are used as the main tools for ice monitoring in the Arctic region. Due to the frequent cloud coverage and long periods of darkness, optical remote sensing technology cannot generate continuous observations in the Arctic. Instead, microwave sensors, which include passive and active (e.g., synthetic aperture radar) microwave sensors, are

commonly used for ice monitoring due to their ability to generate observations regardless of the weather conditions (cloud, rain) and with or without sunlight [4]. Passive microwave sensors are mainly constrained by their coarse spatial resolution, which is normally between 4 km and 100 km. SARs are the main data source used for high resolution ice monitoring and operational purposes [27]. With a spatial resolution typically less than 100 meters, SARs provide observations with rich details of sea ice. However, the interpretation of SAR images is a difficult task due to the extremely complex interactions between SAR signals and sea ice, which depends on the SAR signals (frequency, polarization, incidence angle, noise level and spatial resolution), surface conditions of sea ice (moisture, surface roughness, grain size, thickness and water content of snow cover) and the property of sea ice (temperature, salinity profile, air bubbles, thickness) [23, 27].

The focus of this thesis is ice concentration, the percentage of ice coverage in an area, which is a fundamental ice information required for operational ice monitoring and climate research [27]. Several empirical algorithms have been developed to extract ice concentration from passive microwave brightness temperatures, based on the differences or ratios between signals of ice and open water at different microwave frequencies and polarizations [62]. Global passive microwave based ice concentration maps are generated automatically using these algorithms on a daily basis for applications that do not require fine detailed ice structures, such as large scale climate modeling. However, passive microwave based ice concentration maps have known issues of underestimation for regions of low ice concentration levels, due to signal contamination from the atmosphere and mixing of different surface types [62].

Due to the lack of suitable algorithms for automatic ice concentration estimation from SAR images, SAR images with ancillary data are manually interpreted by human experts working in government agencies to produce operational ice charts. Several types of ice charts are regularly produced to provide general ice information at different time spans for selected regions [28]. In an ice chart, the target region is divided into contiguous regions (polygons), and each polygon is labeled with general information about the ice distribution (ice types with corresponding concentrations and floe sizes). Ice charts are the most accurate ice maps available and have been widely used for climate change research and ship navigation, offshore drilling and other marine operations, such as fishery, search and rescue [27]. However, the manual interpretation of SAR images is time consuming and subjective, based on the skills of the ice experts [27]. Small-scale details of the SAR images are also not fully interpreted in ice charts, due to the use of large polygons. With the growing research and human activities in the Arctic, more SAR satellite missions are planned, such as the Sentinel-1B (European Space Agency) and RADARSAT Constellation Mission (Canada), to provide improved spatial and tempo observations of the Arctic.

With the increased data volumes from such satellites, the use of these satellite data is constrained by the inability to interpret the data efficiently. The ability to produce timely ice concentration maps with improved spatial resolution and spatial coverage is increasingly desired.

There has been a constant effort in automating the ice charting process, but little success has been achieved [27]. With the most recent advances in machine learning, especially the rapidly growing interests in deep neural networks, complex recognition tasks such as classification of natural scenes and language translation have been largely improved due to the enhanced learning capacity of deep neural networks [82, 34, 48, 15]. The automation of ice concentration estimation from SAR images may also benefit from the using of deep neural networks, because

1. SAR image interpretation is a complex recognition task that requires machine learning algorithms with the ability to learn complex models.

2. The large volume of ice chart archives and SAR images provide a large number of training samples required for the training of complex deep neural networks.

The objective of this thesis is to find a method for large scale ice concentration estimation, with high accuracy and fine spatial resolution compared to existing operational ice concentration estimation techniques, e.g., passive microwave ice concentration mapping and SAR image based manual ice charting, to support marine activities and climate change research in the Arctic. Methods for automatic ice concentration estimation from SAR images based on convolutional neural networks (CNNs), which is a type of deep neural networks designed for image recognition tasks, are developed and studied. In order to apply convolutional neural networks for ice concentration estimation from SAR images for operational use, these research questions need to be answered, including:

1. Is CNN a useful method for ice concentration estimation from SAR images?

2. How to design a reasonable CNN model for the ice concentration estimation problem?

3. How to effectively train CNN models using existing ice maps, e.g., ice charts, to generate accurate and high resolution ice concentration estimates?

4. How robust are the developed methods in different sea surface conditions and ice types?

5. How to effectively deploy CNN models for operational uses, especially on how to reuse trained CNN models in regions other than their training region?

These research problems are addressed sequentially in this thesis. We first provide the background of the ice concentration estimation problem with discussions of our motivation and contributions in Chapter 2. In Chapter 3, the usefulness of CNNs and the design of CNN models for ice concentration estimation from SAR images are discussed. In Chapter 4, methods are developed to train CNNs using ice charts with training sample errors modeled. Accurate ice concentration estimates with finer details than existing operational ice charts are produced. In Chapter 5, the robustness of the developed algorithms in previous chapters are demonstrated. In Chapter 6, model re-usability (transferability) is studied and suggestions for operational deployment of the developed methods are provided.

Chapter 2

Background and Contributions

This chapter presents the fundamentals of sea ice monitoring using remotely sensed data. First, necessary information about sea ice development and how it is used to develop qualitative sea ice classes are introduced. Then, the principles of passive microwave and SAR remote sensing of sea ice are explained. Last, the methods for ice concentration estimation from the remotely sensed data in both operational centers and research are reviewed with discussions that lead to the proposed methodology.

2.1   Ice Classes

Ice is usually classified based on its thickness and stage of development. Ice charts use a detailed ice classification scheme with over 10 types of ice [28, 11], which can be summarized to new ice, nilas, young ice, first-year ice and multi-year ice [39]. New ice is the newly formed ice that appears at the beginning of frozen season or when open water refreezes. In this stage, ice crystals start to accumulate under the effect of wind and waves. The accumulated new ice is called nilas when it is less than 10 cm thick. When ice keeps on growing, it enters the stage of young ice with thickness varying between 10 cm and 30 cm. This is a transition stage between nilas and first-year ice. There are two sub-types of young ice: gray ice with thickness between 10 cm to 15 cm, and gray-white ice with a thickness between 15 to 30 cm. First-year ice is the ice that is still growing after the stage of young ice but has not grown for more than one year. The ice thickness at this stage varies between 30 cm to 2 m. Ice that survives one or more melting season is called multi-year ice. Multi-year ice is typically thicker than new ice, young ice and first-year

ice. The thickness usually ranges from 2 m to 4 m. Also, multi-year ice is much lower in salinity than first-year ice because there is brine drainage from the ice during melt.

2.2 Remote Sensing Technology for Ice Monitoring

There are three types of satellite sensors that are used for ice monitoring: optical sensors, passive microwave sensors and SARs (synthetic aperture radar). The application of optical satellites in sea ice monitoring is largely constrained by cloud cover and the lack of sun illumination in the Arctic region in the winter. In contrast, remote sensing in the microwave region of the electromagnetic spectrum does not require sunlight and is much less sensitive to cloud cover or atmospheric conditions, which makes microwave remote sensing, passive microwave and SAR, widely used for sea ice monitoring [13]. Therefore, only passive microwave and SAR remote sensing of sea ice will be discussed.

2.2.1 Passive Microwave

Passive microwave data are the most widely used data for global sea ice monitoring [67]. Satellite passive microwave radiometers collect and measure the natural radiation of the Earth's surface in multiple frequencies and multiple polarizations (horizontal and vertical). Satellite passive microwave sensors are normally designed to have a very large footprint (from 4 km to 100 km). For the low-frequency channels (e.g., 6.9GHz) on a passive microwave sensor there is no atmospheric influence on the observed brightness temperature. However, at these frequencies the spatial footprint of the measurement is large ($\approx 50km$) which is mainly related to the limited diameter of the antenna reflector. As the frequency increases, the size of the spatial footprint decreases (to as low as 3 km), at the expense of increased atmospheric contamination of the surface signal [95, 4].

The principle of passive microwave remote sensing is stated by the Raleigh-Jeans law, which is an approximation of the Plancks Law in microwave frequency [13, 3]. The Raleigh-Jeans law describes the spectral radiance of a black body at a given temperature and wavelength [95]. Based on the Raleigh-Jeans law, the observed radiance, which is called brightness temperature, is the multiplication of the actual temperature and emissivity of the target. The microwave emissivity of sea ice is a function of the dielectric constant, which is mostly affected by the sea ice body structure, surface roughness, moisture content and salt content [13, 3]. The emissivity of seawater is very low and stable under wind of middle strength. The emissivity of young ice and first-year ice are often very strong.

The emissivity of multi-year ice is slightly lower than that of the first-year. The difference in emissivity makes water, young ice and multi-year ice easily distinguishable in passive microwave images [13].

In addition to the variability of surface conditions, the observed radiance is also affected by atmospheric absorption and emission of microwave energy and wind roughening over open water, which are reported to reduce the effectiveness of ice concentration estimation algorithms [86].

SSM/I (Special Sensor Microwave/Imager) and AMSR-E (Advanced Microwave Scanning Radiometer - Earth Observing System) are two passive microwave satellite sensors that have been widely used for the estimation of sea ice concentration. SSM/I has a spatial resolution of 25 km on average, with 12.5 km at higher frequency bands [60]. Data from the AMSR-E sensor have a spatial resolution from 75 km to 5 km which is the finest among its kind [94]. The AMSR-E sensor failed in Oct. 2011. Its successor, the AMSR2, operates in the same frequencies with slightly finer spatial resolution (62 km to 3 km) [78].

2.2.2   Synthetic Aperture Radar

Synthetic aperture radar (SAR) is an active microwave remote sensing technology that operates at low frequencies (e.g., 5GHz) of the microwave range of the electromagnetic spectrum. Similar to passive microwave radiometers, SARs are also not significantly attenuated by weather conditions and sun illumination. Unlike passive microwave radiometers, which measure the emission of the targets, SAR sends electromagnetic radiation (EM radiation) pulses to the ground and collects reflected signals as the measurements. By using signal processing to "synthesize" antenna size along the flight direction, the spatial resolution is largely improved to beyond the limitation of the physical antenna size [19]. The unique characteristics of SAR make it widely used in environmental monitoring. In regions such as the Arctic and the tropics, where a clear sky is rare [70], SAR is the only method to collect relatively continuous (e.g., RADARSAT-2 can acquire images in the Gulf of St. Lawrence everyday [85]) large-scale information of the Earth's surface with fine resolution. There are also other types of active microwave sensors used for imaging the Earth's surface, such as scatterometers and altimeters, which measure the backscatter response of the transmited signals. However, they normally operate in much coarser spatial resolution (e.g., over 10 km [95]).

Often SAR imaging systems are designed to work at different polarization configurations. Polarization is the orientation of the EM waves sent or received by an antenna.

Because scatter from the imaging surface can change the polarization of the EM waves, there are four different polarization configurations for SAR systems:

1. HH. Send and receive horizontal linear polarized EM waves.

2. VV. Send and receive vertical linear polarized EM waves.

3. HV. Send horizontal linear polarized EM waves and receive vertical linear polarized EM waves.

4. VH. Send vertical linear polarized EM waves and receive linear horizontal polarized EM waves.

Some SAR systems work in several different polarization configurations simultaneously. Different polarizations provide different but correlated observations of the Earth's surface. The combination of different polarizations generally provides more information than a single polarized SAR image.

There have been a steadily growing number of SAR systems. Most of the currently planned SAR missions are satellite constellations, i.e., a group of satellites working in a coordinated way, such as the Sentinel-1, SAOCOM, METEOR-M and RADARSAT Constellation mission (RCM) [72]. Most of the planned SAR satellites will have the capability to work in dual or quad polarizations simultaneously to provide more information of the Earth's surface. Sentinel-1a is already producing data, which are free to download. The data acquired by the RCM will also be freely available to the public. Tools to process and extract information from these large volumes of data will likely be in great demand in the future.

An important characteristic of SAR images that must be considered when processing SAR images is the heavy "contamination" of speckle noise due to the coherent nature of SAR imaging systems. There is little information in a single pixel (Fig. 2.1). Spatial texture and structure (the relationship between different parts of an image in different scales) of the backscatter (signals that returns to the sensor) image are often required to interpret SAR images in addition to backscatters of pixels [70]. Due to the processing of the data to achieve ScanSAR mode, the HV pol RADARSAT-2 images used in this study contain vertical banding that leads to distortion that can be seen in the images. (see Fig. 2.2 for an example). Due to the fact that the HV pol is close to the noise floor, the effect of this banding noise can be substantial, which make it more challenging to extract information automatically from SAR images. SAR images shown in this thesis are

8
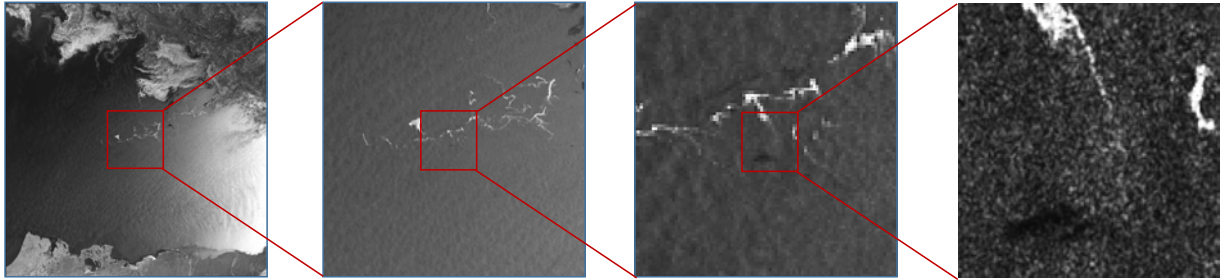
Figure 2.1: An example of HH pol SAR image. Each image corresponds to the rectangle region in its left image. In the first image, the backscatter of water (the area in the middle of the scene) increases from left to right due to the incidence angle effect. In the middle two images, sparse ice floes (white) on large water body (dark). Water exhibits a non-smooth pattern due to wind roughening. Wind roughening is less obvious when incidence angle increases and water becomes darker and smoother as shown in the first image. In the last image, the speckle noise of SAR image is visible at this scale. Images are enhanced for better viewing.

enhanced for visualization using the *imadjust.m* function in Matlab®. The contrast of some images has also been adjusted manually to clarify the image features under discussion.

The SAR backscatter of water depends on the incidence angle of SAR signal, wind speed and direction. Calm water generally has lower backscatter compared to wind-roughened water (Fig. 2.3), most noticeably in HH polarized and VV polarized SAR images as compared to HV polarized SAR images [13]. This is not always the case due the effect of incidence angle, which is the angle between the emitted SAR signal and the normal to the local surface. The backscatter of water is stronger with lower incidence angle for HH pol, which can be even stronger than the backscatter of ice (Fig. 2.1). The sensitivity of scattered SAR signals to wind speed also varies with incidence angle [57].

SAR backscatter in ice-covered regions is complex. It depends on many physical parameters, such as surface roughness, salinity, ice structure, ice thickness, melt, snow cover, temperature and floe distribution [13]. The appearance of ice changes through seasons and locations because these parameters are constantly changing due to the ice-ocean and ice-atmosphere dynamics [37]. The backscatter of sea ice is composed of surface scattering, which occurs near the surface of snow or ice, and volume scattering, which occurs inside the snow body or ice body. Which mechanism is dominant depends on the penetration depth of the radar pulses, which is largely determined by the dielectric constant of the surface. The absolute value of the dielectric constant determines the reflection coefficient, and the imaginary part of the dielectric constant determines the penetration depth [47].

9
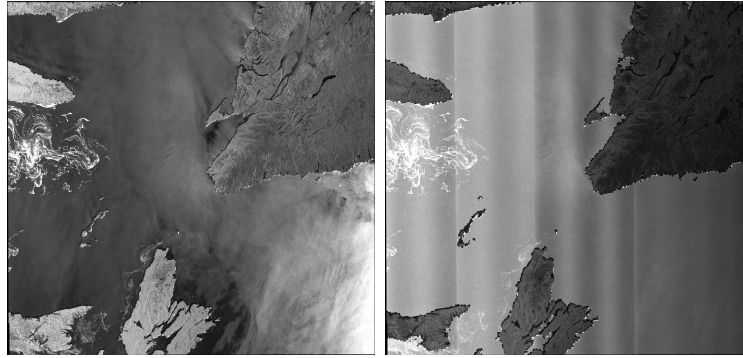
(a) HH (b) HV

Figure 2.2: The banding effect in HV pol RADARSAT-2 images. The vertical strips in (b) are caused by the banding effect in HV pol. The banding in this case is obvious and causing inhomogeneous water observations and sharp pixel intensity changes in water regions. Images are enhanced for better viewing.

In general, wet materials and materials with high salinity have high dielectric constant such as seawater, wet snow and young sea ice. There is little penetration of radar waves into these surface types, i.e., there is almost no volume scattering. In such cases, surface scattering dominates the backscattering, and the strength of surface backscatter is largely determined by the surface roughness, where rougher surface causes stronger backscatters. Dry materials with low salinity have a low dielectric constant such as dry snow and multi-year ice. Volume scattering contributes a large portion of the backscatter for such surface types, which is mainly affected by the salinity, crystal structure and air pockets in ice.

The backscatter of most new ice types is low due to the smooth and wet surface. New ice may be confused with water when wind speed is low. Two examples are shown in Fig. 2.4. The first row of Fig. 2.4 is an example of typical calm water with high incidence angle. The second row of Fig. 2.4 has new ice of intermediate ice concentration in the dark region (left of the image). The third row of Fig. 2.4 has new ice for the middle upper dark region with ice concentration 1. The appearances of the new ice in both samples are barely distinguishable from calm water at low incidence angles in both polarizations.

The backscatter will increase significantly for young ice. Young ice often appears bright in SAR images due to the growth of frost flowers (highly saline ice crystals that grow on the ice surface) or deformed ice surface. Compared with nilas, young ice breaks as a result of the sea swell, so distinct blocks are often evident in SAR images.

(a) HH   (b) HV   (c) Land mask
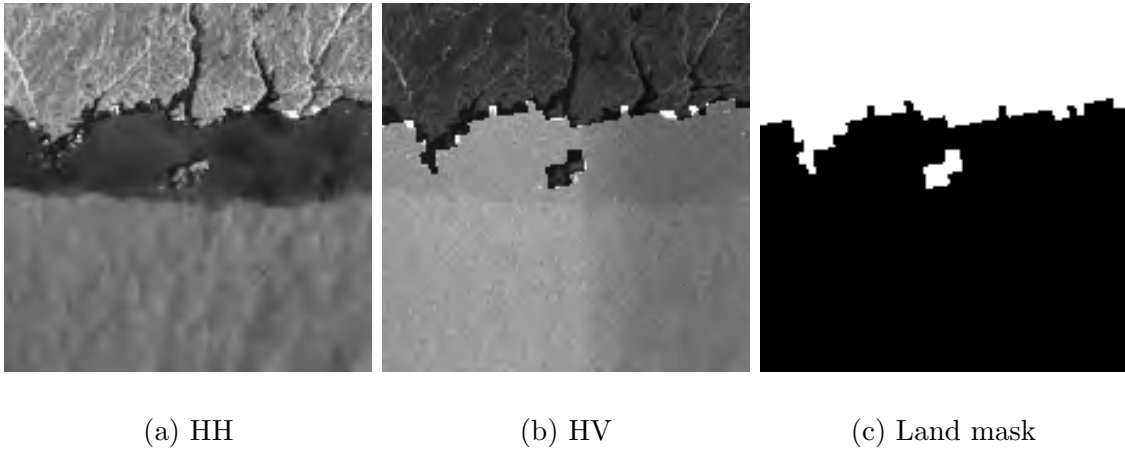
Figure 2.3: Water in SAR images have different level of backscatters for different regions in both HH and HV polarizations in (a) and (b). (c) is the land mask for (a) and (b) with land in white and sea in black. Images are enhanced for better viewing.

First-year ice appears over the entire ice season on the Canadian east coast and in Hudson Bay. The backscatter of first-year ice is generally lower than the backscatter of young ice. Surface scattering is dominant for first-year ice because the salt in the ice makes the dielectric constant high. The backscatter of first-year ice is higher when deformation exists, because the ice ridges and edges caused by deformation will cause corner reflection. Very large first-year ice floes are often formed when several large ice floes are frozen together. Therefore, first-year ice tends to have SAR signatures that are inhomogeneous in both tone and texture.

Second-year ice and multi-year ice produces the strongest backscatter of level ice due to the strong volume backscattering. This is the main difference between the backscatter of first-year ice and second-year ice or multi-year ice. Second-year ice and multi-year ice normally show a rougher texture and brighter tone in SAR images during the winter compared to first-year ice. In multi-year ice, one can often see strips of leads, which are not that common in first-year ice.

2.3 Operational Ice Concentration Estimation

2.3.1 Automatic Passive Microwave Ice Concentration Estimation

Ice concentration estimation from passive microwave data has reached a mature state in general. It has been widely used in operational ice monitoring systems [27]. The AMSR-E / AMSR2 ice daily ice concentration map is a widely used ice concentration map generated from passive microwave data.

Daily AMSR-E/AMSR2 global ice concentration maps are produced by the PHAROS group at the University of Bremen. This ice concentration dataset features a pixel spacing of 6.25 km for both AMSR-E and AMSR2, which is the finest passive microwave ice concentration product. It was generated from the AMSR-E/AMSR2 89 GHz channel data using the ARTIST sea ice (ASI) algorithm [86]. The ASI algorithm uses the brightness temperature difference of the vertically and horizontally polarized channels at 89GHz to determine ice concentration. It assumes a cubic relationship between ice concentration and the brightness temperature difference. Ice concentrations are determined by interpolating the brightness differences of tie points, which are the typical brightness difference of ice and water in an image. The variation of brightness temperatures for different surface types is the main source of errors for the ASI algorithm. This is often caused by the variations in surface salinity, moisture and snow cover. Melting will also strongly change the emissivity in microwave bands. Atmospheric emission and absorption may also contaminate the passive microwave observations. For the AMSR-E/AMSR2 data, the ASI algorithm is reported to produce ice concentration with errors less than 10% for middle and high ice concentrations (above 65%) [86]. The errors for low ice concentrations may have substantial variations depending on the atmospheric conditions [86], which is mostly related to the using of weather filtering to reduce the effect of atmospheric contamination. The use of weather filtering is also related with the underestimation of ice concentration for thin ice or very low ice concentration [86]. Note that AMSR-E was used before the instrument stopped producing data in October 2011.

2.3.2 Manual SAR Ice Concentration Estimation

SAR images are used as the main data source for operational ice charting. Ice charts are manually interpreted ice maps prepared by ice experts using mainly SAR images and other ancillary data such as ship reports and in situ observations [10]. Ice charts are considered

to be the most reliable ice concentration product available. An example of ice chart is demonstrated in Fig. 2.5. Each ice chart is composed of contiguous polygons with assigned ice concentration in tenth (e.g., ice concentration only has 11 possible values: 0, 0.1, 0.2, ... 0.9, 1). The accuracy of ice charts is normally considered to be approximately 10% [10]. Each polygon has a separate letter indicating the "egg code" of the polygon, which is the total concentration, partial concentration, stage of development and form of ice in an oval form [28]. Compared to the daily AMSR-E / AMSR2 ice concentration maps, ice charts are more accurate, especially for regions of low ice concentration, and have more accurate ice-water boundaries for regions where ice-water boundaries show simple structures and are easy to interpret (Fig. 2.6).

In Canada, the Canadian Ice Service (CIS) produces three different types of ice charts: image analysis charts, which are valid at SAR image acquisition time; daily ice charts, which incorporate SAR images and other sources of data, valid at 1800 UTC each day, produced during shipping seasons in various regions; and regional ice charts which are produced weekly all year round over all Canadian waters.

2.4 Automatic SAR ice concentration estimation

Published algorithms for ice concentration estimation using SAR imagery normally consists of three major steps: feature extraction, regression, and post-processing [81, 12, 42, 8, 46, 44, 45]. Each of these steps will be discussed in order.

A number of image features have been used by algorithms to estimate ice concentration from SAR images, such as auto correlation and GLCM (gray-level co-occurrence matrix) texture features [8, 46], locally normalized backscatter intensity and intensity differences between SAR images of different polarizations [8, 44]. These features were reported to be affected by the variation of wind conditions [8] and incidence angles [46]. In a recent study [46], a set of GLCM (gray-level co-occurrence matrix) features, intensity, filtered intensity of the SAR images are used together to estimate ice concentration from SAR images. No single feature was found to be well correlated with the ice concentration, while using many features together led to improved correlation with ice concentration [46]. Some image features are also designed to detect ice in SAR images, such as the density of edges and corners [42]. These carefully designed features were found to improve ice concentration estimation. However, there was still a degradation of performance in the melt season, and overestimation of ice concentration in some open water regions, as corner and edge features are unstable to melt and wind roughened water. To obtain more robust and representative features, backscatter calibration and incidence angle correction may be applied to SAR

13

images before feature extraction [44, 45], the effect of which depends on the SAR imagery and surface conditions.

A few types of regression models have been used to estimate ice concentration from SAR imagery, such as empirical linear regression [42], Gaussian mixture models [43], and multiple-layer perceptron (MLP) [8, 44]. The effectiveness of those models depends strongly on the features used.

Segmentation based smoothing is a commonly used post-processing technique to improve the numerical quality and visual appearance of the estimated ice concentration [43, 45, 44]. It separates the target region into visually homogeneous smaller regions. Proper segmentation of SAR images should have segmentation boundaries consistent with the natural separation of different sea surface types, which also requires the chosen features are able to separate difference sea surface types. All the pixels in a region are assigned with the same ice concentration, which is typically the mean of the estimated ice concentration in that region. Segmentation based smoothing is not able to correct the local bias in the results. Any local systematic errors, such as overestimation of wind roughened open water, will be passed to the results after segmentation based smoothing. Additionally, small ice structures may be considered as noise and smoothed out by the segmentation algorithms.

Both single pol [12, 8, 43] and dual pol (HH and HV) [44, 45, 46] SAR images have been used to extract ice and water features. As mentioned earlier, they are highly correlated, but also contain complementary information. HH pol images have more detail but are strongly affected by incidence angle. HV pol images are less affected by incidence angle, but have lower signal-to-noise ratio (SNR) and fewer details. The use of dual polarized SAR images has been shown to benefit ice information retrieval [81, 45, 44, 56], especially in open water regions and in melt conditions [44]. This improvement can be contributed by the fact that HV pol is insensitivity to wind roughening and has different backscatters to water [79].

2.5   Motivations and Contributions

The difficulty in estimating sea ice concentration from SAR images is essentially the difficulty in finding representative SAR image features for different surface types. The use of handcrafted image features, such as polarization differences and texture measurements, may be unable to distinguish the complex SAR image signatures of ice and water as has been demonstrated in previous studies [8, 42, 46, 44]. Representative features for SAR images should be more complex due to the complexity of the underlying physics. However,

14

designing complex and representative features to hard code knowledge into the machine learning system is difficult and non-intuitive [6]. Moreover, different features might need to be designed for different ice forms at different locations or time of the year, due to the variation of sea ice and water features in time and space, which makes it difficult to implement in large-scale operational applications. Therefore, it is of interest to investigate the use of feature learning, instead of feature design, to estimate sea ice concentration from SAR images.

Feature learning is of interest to identify and disentangle the underlying explanatory factors from raw data [24, 6], which has become an increasingly active area of research in the past several years [66]. It can reduce the dependency of feature engineering and prior knowledge for machine learning systems [6], which makes it appealing for analyzing complex and large volumes of data [68, 25, 66].

Deep learning is a type of feature learning method that can automatically extract complex data representations at high levels of abstraction [33, 5, 66]. It has recently demonstrated success in many complex recognition tasks such as speech recognition and image object classification [33, 34, 48, 7]. Such algorithms use multilayer (deep) neural networks to extract representative features in a hierarchical architecture, where higher-level features are defined in terms of lower-level features [66]. This hierarchical architecture simulates the process of multiple levels of abstraction with higher-level features in the hierarchical architecture correspond to more abstract features. For example, a dog may be represented by its body parts, which are then represented by texture and contour features. Those texture and contour features are then represented by lines and corners. Neural networks can simulate this abstraction process by representing lines and corners at lower-level layers, texture and contour features at middle layers, body parts at higher layers and the concept of dog at the top layer.

Based on the nature of the application, deep neural networks can be structured with different layer configurations. For image recognition tasks, deep convolutional neural networks (CNN) are the most widely used deep neural network structures due to their ability to model local image structures at multiple scales efficiently, which is achieved by weight sharing within a layer and local connectivity between adjacent layers. CNNs have a demonstrated ability to achieve high performance for image related recognition tasks [55, 17, 48, 41]. This has motivated the use of CNNs for ice concentration estimation from SAR images. More details of CNNs are given in Chapter 3.

There has been limited research in learning features from satellite images. Related studies include using convolutional neural networks for road classification from aerial images [63] and vehicle detection from high resolution satellite images [16]. Recently, Penatti

et al. (2015) argue that the features generated from everyday objects can be used to improve aerial image classification [76]. No published research in applying CNNs on SAR images or other sea ice imagery is known to the author. The effectiveness of CNNs and what is a proper CNN model structure for ice concentration estimation need to be explored first.

The training of CNN models requires a large quantity of high quality training samples. For many remote sensing problems, gathering high quality ground truth is expensive and sometimes not feasible, due to the vast study area and diversity of surface conditions. This is also the case for ice concentration mapping, because it is difficult to send people to sample many locations simultaneously while the scene is being captured, and visual analysis of satellite images contains uncertainty. Ice concentration from passive microwave sensors and ice charts are the two potential sources of training samples. The bias in passive microwave ice concentration in low ice concentration levels [86] will lead to CNN models with similar biases. Ice charts are a better source of training samples than ice concentration from passive microwave data. Making use of the large number of ice charts available through the archives at CIS could greatly benefit the building of a robust CNN model. However, in doing this, the errors in ice charts need to be taken into account. Even though the CNN is relatively robust to training sample errors, modeling of training sample errors explicitly is still beneficial [5, 64]. One type of error is human error, which is caused by the uncertainty in manual interpretation and misalignment of the manually interpreted ice-water boundaries and the actual ice-water boundaries. A more important type of error is the representation error. The actual ice concentration of a pixel in a polygon in an ice chart may vary from the labeled ice concentration of that polygon. Using the ice concentration of a polygon in an ice chart as the ice concentration for every pixel in that polygon causes representation errors, for regions containing mixtures of ice and water. The effect of these training sample errors needs to be addressed to better train the CNN models with ice charts.

The main contribution of this work is an advanced SAR image ice concentration estimation method. More specifically,

1. CNN is evaluated for ice concentration estimation from SAR images for the first time to the best of our knowledge. A base CNN model and guidelines for hyper-parameter selection for the problem of SAR ice concentration estimation is given which provides a baseline for future research. State-of-the-art ice concentration estimates with finer ice structures than image analysis charts are acquired when trained with image analysis charts for the Gulf of St. Lawrence.

2. Two models that consider both human errors and representation errors in the training

16

samples are developed and analyzed. Human errors and representation errors are two types of errors in the image analyses. Neural network models trained without these two types of errors taken into account are found to produce estimates with erroneous water regions and blurry ice-water boundaries. Both models generate ice concentration estimates with further improved accuracy and spatial resolution.

3. The effectiveness of CNNs and the developed methods in the thesis for ice concentration estimation in the melt season is evaluated. The ability to correctly estimate ice concentration in the melt season is important for operational ice charting. Melt normally causes underestimation of ice concentration as reported in previous methods. The developed models in this thesis are demonstrated to generate reasonable ice concentration estimates in the melt season.

4. The transferability, which is the effectiveness of a model on related tasks, of CNN models is evaluated to guide the application of the developed models for large scale ice concentration estimation tasks operationally. For large-scale ice concentration mapping, it is desirable to reuse a model trained for one location to reduce the training time for another location. By transferring (reusing) the layers trained on another dataset, the training time of CNN models, which is the bottleneck of the computational cost, are reduced by over 10 times in our case with no observed decrease in accuracy.

2.6 Outline of Thesis

The performance of CNNs on SAR image ice concentration estimation is evaluated and discussed in Chapter 3. The search for the optimal CNN model structure (hyper-parameter optimization) is time consuming due to the large number of hyper-parameters and the high computational complexity of each training and testing cycle. In this work, the CNN model is built progressively by steadily improving the best available model. Since no reference CNN model is available for SAR images, a base CNN model structure is first developed with extensive hyper-parameter testing. A variation of the CNN model, fully connected convolutional neural network (FCNN), is also tested using the same structure as the CNN model. The resulted FCNN model is chosen over CNN for further development for later chapters due to improved accuracy and efficiency.

Methods to account for both human errors and representation errors are discussed in Chapter 4 based on the FCNN model developed in Chapter 3. L_1 loss is used to reduce the effect of human errors in training samples, and is able to produce improved results as

17

compared to the FCNN model trained with L_2 loss. Two independent methods, an EM based learning scheme and a novel loss function that models the representation error, are developed to improve the training of CNNs with the existence of representation errors in the training data.

The effectiveness of the algorithms developed in Chapter 3 and Chapter 4 are further tested on another dataset in the Beaufort Sea, which covers the melt season and has different ice features from that of the Gulf of St. Lawrence, in Chapter 5.

The model transferability is evaluated in Chapter 6 to further guide the use of the developed methods in operational situations. The CNN model trained on the data acquired in the winter of Gulf of St. Lawrence is evaluated on the data acquired during the melt season of the Beaufort Sea with a number of top layers (layers close to the network output layer) fine tuned.

Chapter 7 provides a summary of the thesis with limitations and future perspectives.

HH   HV   ImA

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
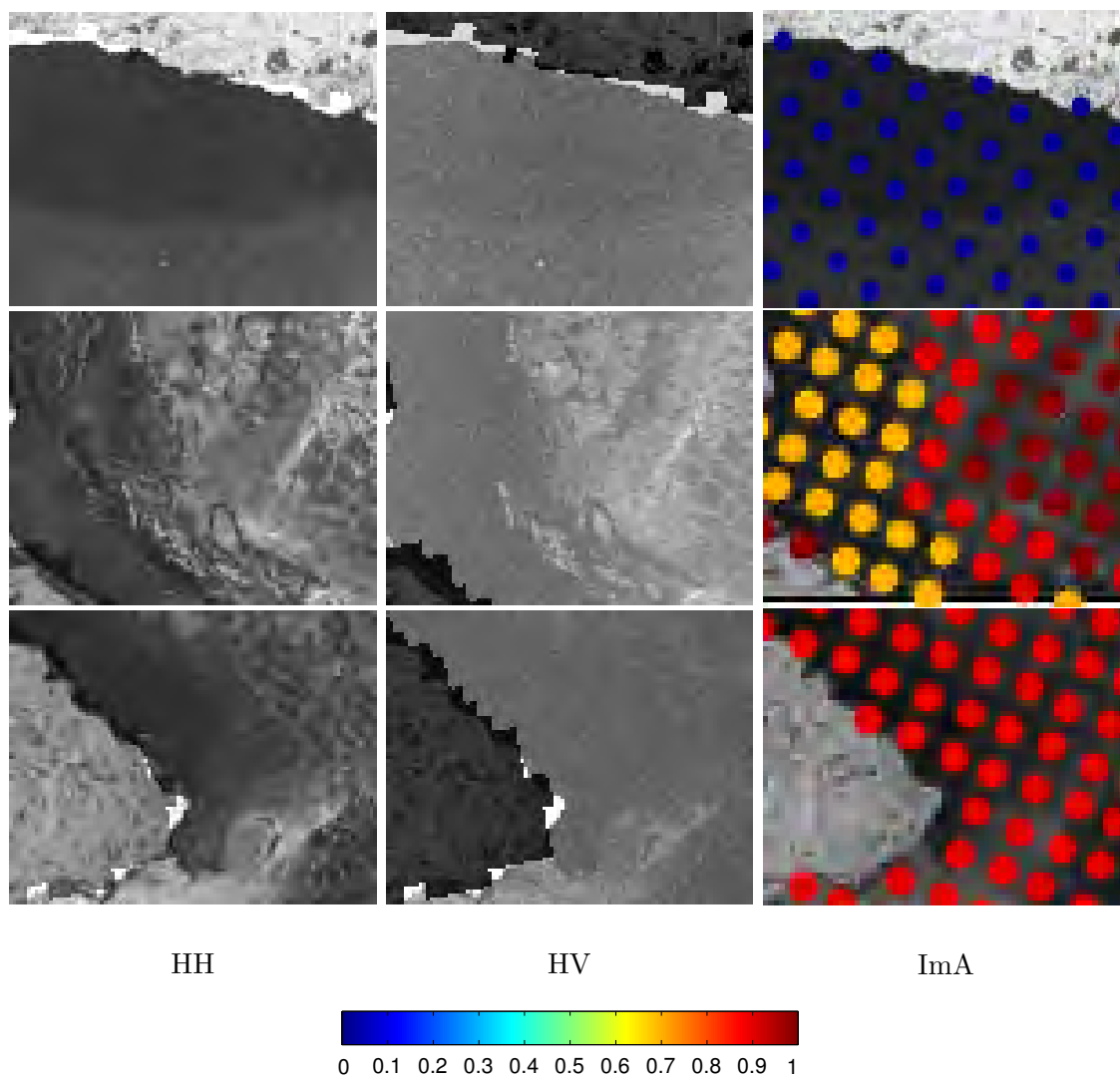
Figure 2.4: Examples of similar image appearances of ice and water when ice has low backscatter and smooth texture. Each row shows the dual-pol SAR images and the corresponding image analyses (ImA) of an example. The first row is an example of typical calm water at high incidence angle in SAR images: low backscatter in both HH and HV pols. The middle row has new ice of ice concentration 0.7 in the left of the image. The last row shows another region of new ice with ice concentration 0.9 in the middle of the image. New ice in both the second and the third row has image textures similar to calm water. The legend at the bottom is a color coding for the ice concentration. Images are enhanced for better viewing.

19
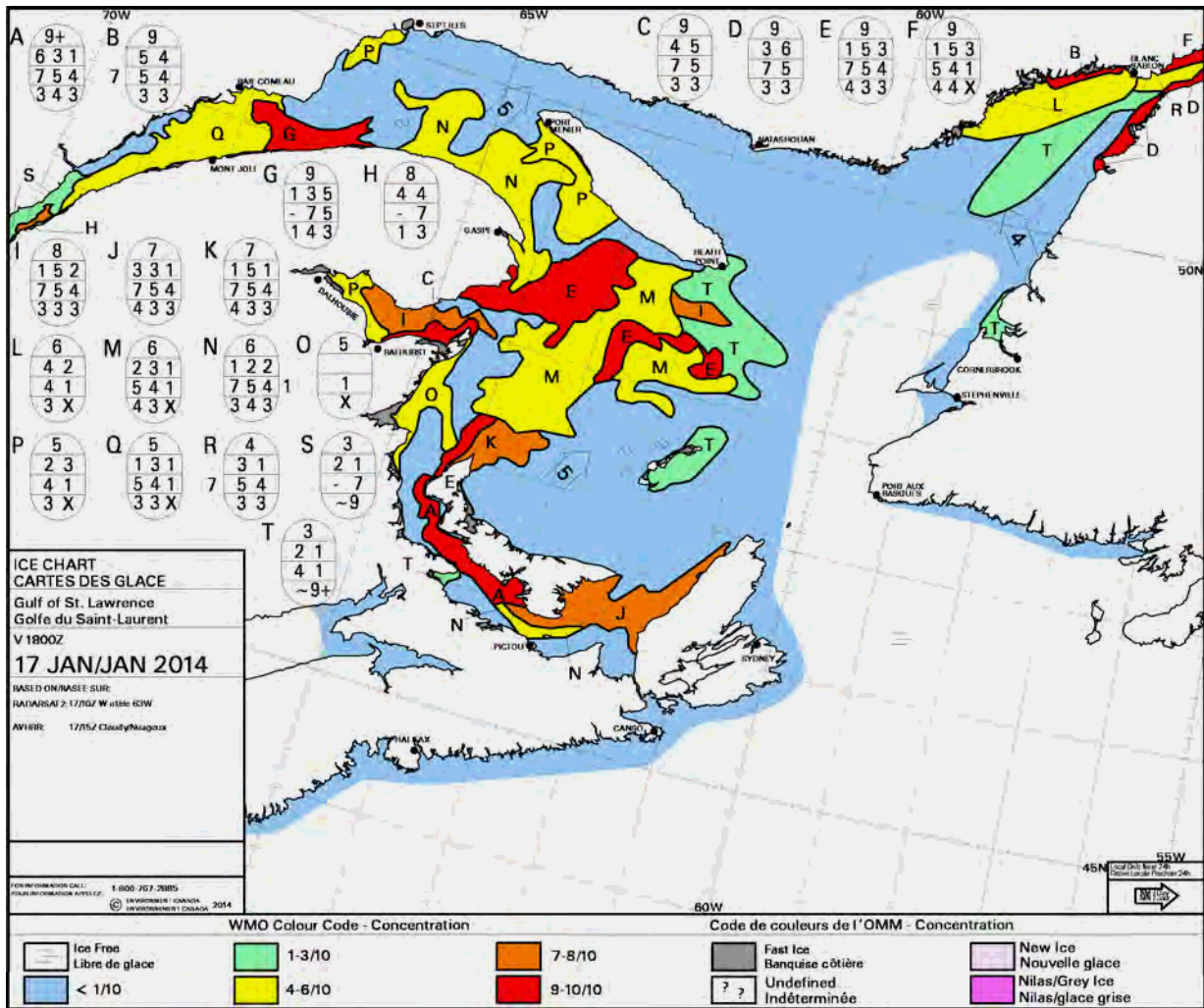
Figure 2.5: Daily ice chart of the Gulf of Saint Lawrence, January 17, 2011.
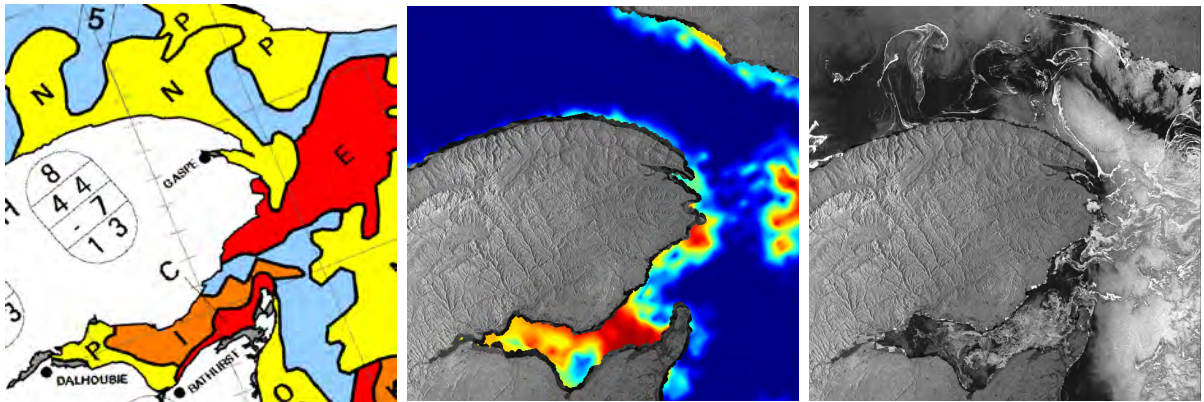
(a) Daily ice chart (b) ASI (c) HH

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
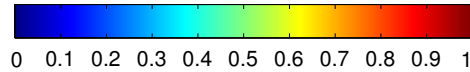
Figure 2.6: Details of two ice concentration products and corresponding SAR images at the region near the center of the daily ice chart shown in Fig. 2.5. ASI refers to the ice concentration generated from AMSR-E or AMSR2 data using the ASI algorithm [86]. Ice charts have more accurate ice water boundaries for some regions and is more accurate compared to ASI ice concentration. Small scale details of the ice structure in the SAR images are not presented in the ice charts or the ASI ice concentration. At the bottom is the color scheme for (b). The color scheme for ice chart (a) is shown in Fig. 2.5.

Chapter 3

Estimating Ice Concentration Using Convolutional Neural Networks

In this chapter, convolutional neural networks (CNNs) are applied to SAR images for ice concentration estimation. Both CNN and its fully convolutional variation FCNN are evaluated. The developed CNN and FCNN structures are given and the details of the experiment design are explained. The is followed by sensitivity analysis of the hyper-parameters, which are the variables that control the network structures and layer configurations, to justify the design of the network structures. Both CNN and FCNN are compared with the state-of-the-art ice concentration estimation algorithm from the literature to demonstrate the improved performance. To show how CNN is detecting ice and water internally, a visualization of the CNN intermediate features is provided at the end of this chapter.

3.1 Convolutional Neural Network and Its Variations

3.1.1 Convolutional Neural Network

CNN is a trainable architecture composed of multiple stages [51, 53, 54]. The input and output of each stage are multidimensional matrices, which are called "feature maps". Each stage is composed of three consecutive operations (layers): convolutional filtering, non-linear transformation and sub-sampling (pooling). A CNN model normally contains multiple stages followed by a stack of fully connected layers on top [54]. An example of CNN structure is illustrated in Fig. 3.1.

In a convolutional layer, the layer input matrix $\boldsymbol{x}$ (width S_x, height S_y and number of channels S_z) is convolved with K convolution filters of size (C_x, C_y, S_z), which are denoted by $\boldsymbol{C^k}, k = 1, \ldots, K$, at step size (stride) P (convolution is calculated for locations that are P pixels apart). A total of K feature maps of dimension M_x and M_y will be generated as the output of this convolutional layer (3.1). Each convolutional layer is mainly characterized by the size and number of filters. The values of these filters are learned through training.

$$\boldsymbol{h^k}_{ij} = (\boldsymbol{C^k} * \boldsymbol{x})_{ij} + b, \text{ in which,} \tag{3.1a}$$

$$i = 1, \ldots, M_x, j = 1, \ldots, M_y, k = 1, \ldots, K \tag{3.1b}$$

$$M_x = \frac{S_x - C_x}{P} + 1 \tag{3.1c}$$

$$M_y = \frac{S_y - C_y}{P} + 1, \tag{3.1d}$$

Where the operation of convolution is denoted by $*$. Notation b is the bias term used to stabilize the training process, which is also trained from data [31].

A convolutional layer is followed by a nonlinear transformation layer, which applies a nonlinear function to the feature maps element-wisely. This nonlinear function is also referred as the activation function. ReLU (3.3a) is the most commonly used activation functions for CNNs and FCNNs. ReLU activation has been demonstrated to lead to faster learning and better features than traditionally used sigmoid activation function, because ReLU activation does not saturate compared to sigmoid activation which does saturate [65, 48].

The sub-sampling layer is also known as the pooling layer, which is used after the nonlinear transformation layer. Pooling provides CNN the ability to extract more abstract features in higher layers and invariance to spatial translations over the pooling window. Max pooling is the most commonly used type of pooling layer due to its simplicity and effectiveness [18, 80, 54, 48]. It outputs the maximum value over each pooling window. A pooling layer is mainly characterized by its pooling method (max or mean), size of pooling window and step size of pooling window. For example, when pooling window size and step size are both set to 2, a max-pooling layer outputs the maximum value of every two by two non-overlapping window of its input.

The top fully connected layers are used as classification modules using the features extracted by the previous multiple stages. Every neuron in a fully connected layer is connected to all the neurons of its input layer. The input matrix $\boldsymbol{x}$ to a fully connected layer is flattened to a vector and transformed to the output space by a weight matrix $\boldsymbol{W}$ and bias b, then an activation function f is applied element-wise to generate its output $\boldsymbol{h}$.

That is,

$$\boldsymbol{h} = f((\boldsymbol{W} * \boldsymbol{x}) + b). \tag{3.2}$$

For the case of CNNs, the first fully connected layer takes a stack of feature maps as input. The feature maps are flattened to a vector first and (3.2) is applied. A fully connected layer can be seen as a special case of convolutional layer when the convolutional filters are of the same size as the layer input. For example, given input feature maps of dimension $S_x \times S_y \times S_z$, the weight matrix w for a fully connected layer of C_z units has dimension $C_z \times (S_x \times S_y \times S_z)$. This weight matrix w can be interpreted as C_z convolutional filters of dimension $S_x \times S_y \times S_z$, where each convolutional filter calculates the inner product of the input and the filter weights as fully connected layers do.

The design of CNN is based on the assumptions of statistical stationarity and locality of pixel dependencies of natural images [52]. These two assumptions are enforced by the convolutional filtering process where weights are shared among different locations and only locally connected. Compared to standard fully connected neural networks with similar number of units, CNNs are able to model local spatial information more efficiently with fewer trainable parameters, which also makes them easier to train [52, 48].
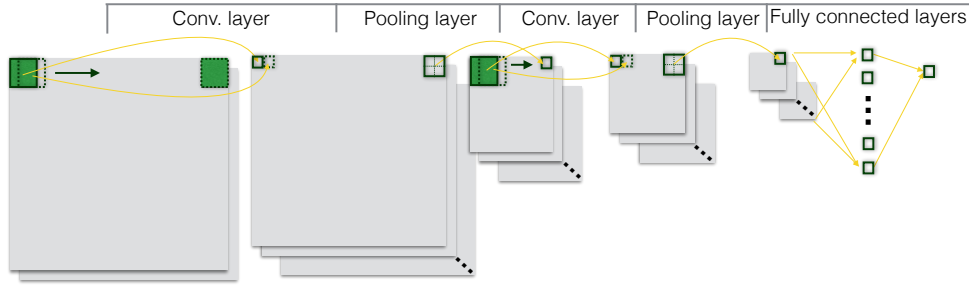
Figure 3.1: An illustration of an example CNN model with two convolutional layers and two fully connected layers. The gray squares are inputs and outputs of different layers. The green squares are convolutional filters. In convolutional layers, the convolutional filters are applied to everywhere of the layer input and output a stack of feature maps. These feature maps are sub-sampled by the subsequent pooling layer. The output of the second pooling layer is inputed to a stack of two fully connected layers where the last fully connected layer has only one output unit.

The computational complexity of a convolutional layer is in the order of $\mathcal{O}(S_z * K * M_x * M_y * C_x * C_y)$. The number of operations for a pooling layer is in the order of $\mathcal{O}(K * M_x * M_y)$, which is much smaller than convolutional layers. The computational

complexity for a fully connected layer is in the order of $\mathcal{O}(W_x * W_y)$, in which W_x and W_y are the number of rows and columns of weight matrix W. W can be very large, especially for the first fully connected layer of a CNN, because W_x is the number of elements of the input feature map ($K * M_x * M_y$), and W_y is normally set to a large number, e.g., 1024. For each unit, an activation function is applied, which is simple to compute in general. The computation required for an activation function can be considered constant. So the computational complexity for all the activation functions in the order of the number of parameters in a neural network. The computational complexity of a neural network is therefore mainly determined by Convolutional layers and the fully connected layers.

3.1.2   Fully Convolutional Neural Network

CNNs were originally designed for image labeling tasks, which require one label for one target image [51, 53, 48]. Feature sub-sampling and spatial translation invariance are helpful for image labeling, but hinder pixel level labeling tasks by leading to inaccurate object boundary locations. The prediction for pixel level labeling is inefficient when each forward-propagation only produces the prediction for a single pixel. The computationally expensive CNN model needs to be applied to all pixel locations densely on an image to generate pixel level predictions. Fully convolutional neural network (FCNN) is a CNN variant designed with these two problems addressed for pixel-level labeling tasks [58].

The basic idea of FCNN is to replace the fully connected layers of CNNs by convolutional layers. Because convolutional layers convolve the input feature maps with a set of filters, convolutional layers can be applied to arbitrary input size, while fully connected layers can only take input of predefined sizes, which has constrained the input size of CNNs. By replacing the fully connected layers of a CNN model with convolutional layers, FCNN can be applied to an input of arbitrary size. However, by doing this, the size of the output of the network (the output of the top layer of a neural network) is smaller than the input size of the network due to the use of sub-sampling (pooling) layers. For example, if a FCNN contains three max-pooling layers, each of which reduces the feature maps by half, the FCNN output is roughly $1/2^3$ of its input size, i.e., the output of the network corresponds to input image locations of every 8 pixels away.

Several methods have been proposed to produce denser outputs from FCNNs [30, 83, 75, 15]. The FCNN used in this study [75, 15] is chosen based on its simplicity relative to other methods. It first removes the sub-sampling step after the max-pooling so that the dimension of feature maps is not reduced. However, this will reduce the receptive field of the neurons (the size of the local region in the input, which affects the responses of a filter)

in upper layers of these feature maps. To keep the receptive field of the neurons unchanged, zeros are inserted to the convolutional filters to increase their length. Applying this zero-inserted filter on the pooling results without sub-sampling is equivalent to applying the original compact filter on the sub-sampled pooling results in the sense that they have similar sizes of receptive field, but the output size is not reduced. This is called the "hole algorithm" [15]. For example, in a CNN model, a feature map $f0$ of size 10×10 input to a pooling layer with stride 2 will be reduced to size 5×5, which is noted by $f1$. Applying a 3×3 filter to the $f1$ outputs a feature map $f2$ of size 5×5 when 1 pixel wide padding is applied around $f2$. In an FCNN, the pooling layer uses pooling stride 1 to calculate local maximum values of a 3 by 3 window for every pixel location in $f0$, and outputs feature map $f1'$ of size 10 by 10 when 1 pixel wide padding is applied. By inserting one zero between elements of a 3×3 filter, the filter becomes 5×5. Applying the 5×5 filter on $f1'$ with padding 2 leads to a feature map $f2'$ of size 10×10, which is twice the size of $f2$. $f2$ corresponds to the even rows and cols of $f2$.

3.2   Data and Study Area

The study area is located in the Gulf of Saint Lawrence area, which covers the St. Lawrence Estuary eastward from Quebec and the entire Gulf of St. Lawrence (Fig. 2.5). The period of study extends from January 17th, 2014 to February 10th, 2014. This time of year corresponds to freeze-up in the Gulf of Saint Lawrence, with both ice concentration and thickness increasing from January into February. For the duration of the study, most of the ice is new, gray and gray-white ice. Ice is compact in the middle of the Gulf of Saint Lawrence and becomes looser approximately 20-50 km from the ice edge [26].

A total of 25 RADARSAT-2 dual-pol (HH and HV) ScanSAR Wide [85] images acquired in January and February of 2014 are used for the present study. The full list of the SAR images used are provided in Table 3.1. The pixel spacing of the acquired SAR images is 50 m by 50 m, and the incidence angle ranges from 20° to 49°. The outlines of all the SAR images in the dataset are shown in 3.2.

Image analysis charts are used for training. Image analysis charts are more accurate compared to passive microwave based ice concentration products [2]. Compared to other types of ice charts (daily ice chart and regional ice chart), image analysis charts provide a more detailed interpretation of SAR images and are valid at the SAR image acquisition time [28].

The image analysis training data obtained from CIS in this study are grid-point data from the image analysis charts. The sampling interval is about 8 km in the north-south
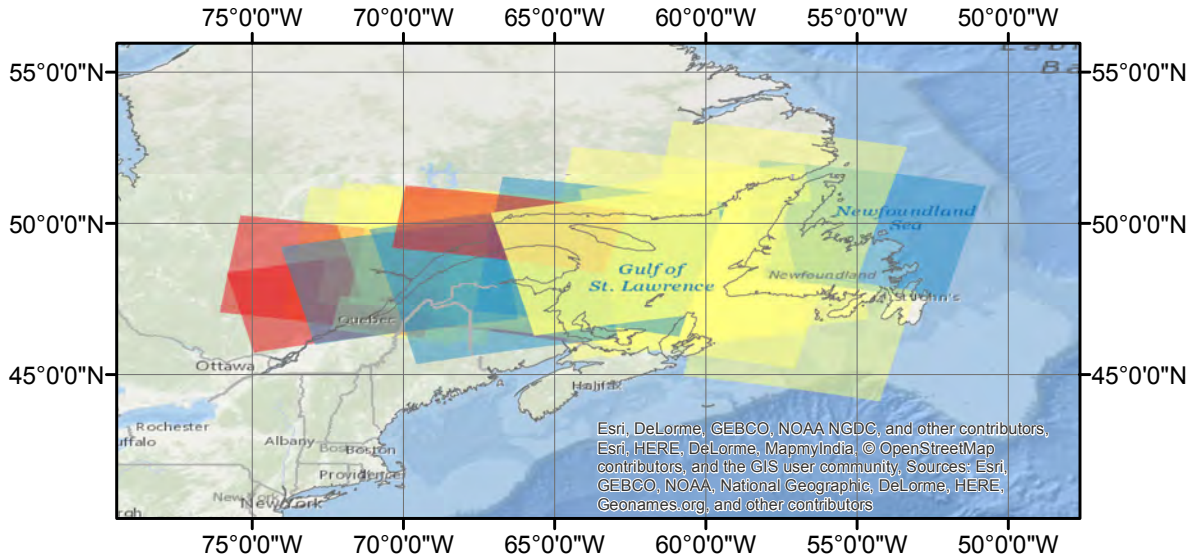
Figure 3.2: Study area and the dataset for the Gulf Saint Lawrence. There are 25 scenes of dual-pol SAR images acquired between Jan. 16 2014 and Feb. 10 2014 in this area. The coverage for each scene is marked in a translucent polygon with different colors. Yellow scenes are used for training, red are used for validation and blue for testing.

direction and 5 km in the east-west direction. Note that since each polygon of the image analysis is labeled with a single ice concentration value for the entire polygon, the actual ice concentration at the grid-point locations may be different from the ice concentration of the polygon, depending on the spatial distribution of ice within the polygon. The number of image analysis grid points for each SAR image varies from a few hundred to several thousand, which depends on the area of sea surface in that scene. The numbers of image analysis grid points for the SAR images used are provided in Table 3.1.

Corresponding daily AMSR2 ice concentration maps for each SAR scene are downloaded from the website of PHAROS group at the University of Bremen [86]. These AMSR2 ice concentration maps are reprojected to their corresponding SAR image pixel grids with cubic interpolation. The reprojected AMSR2 ice concentration is referred to as ASI ice concentration in the rest of this thesis.

Table 3.1: Details of the Gulf of Saint Lawrence dataset

Set	Scene ID	Date acquired	Number of image analysis points
Training	20140131_103053	31 January 2014	8231
	20140127_221027	27 January 2014	1319
	20140203_104323	03 February 2014	3019
	20140116_223042	16 January 2014	530
	20140208_095758	08 February 2014	13872
	20140210_220111	10 February 2014	8358
	20140207_214938	07 February 2014	612
	20140125_100500	25 January 2014	5200
	20140131_215240	31 January 2014	11111
	20140124_103501	24 January 2014	6900
	20140120_105149	20 January 2014	829
	20140118_101002	18 January 2014	7492
	20140128_101751	28 January 2014	12791
	20140130_222234	30 January 2014	1407
	20140123_222627	23 January 2014	950
	20140127_104734	27 January 2014	3427
	20140124_215646	24 January 2014	10964
	20140121_214420	21 January 2014	15897
Validation	20140122_095247	22 January 2014	5014
	20140206_221744	06 February 2014	3395
	20140209_223030	09 February 2014	545
	20140207_102631	07 February 2014	9228
Testing	20140210_103911	10 February 2014	2918
	20140130_110029	30 January 2014	425
	20140126_223850	26 January 2014	165
	20140117_103914	17 January 2014	2922

3.3 Ice concentration estimation work flow

3.3.1 Overview

The overall processing scheme is shown in Fig. 3.3. There are three major steps: preprocessing of the SAR images, training the model and predicting using the trained model. In
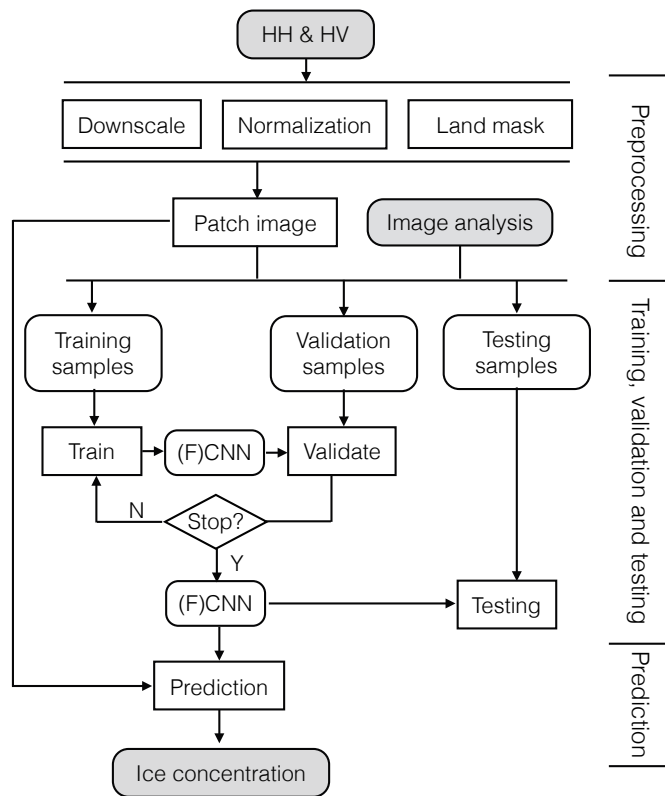
Figure 3.3: Flow chart of using CNN for ice concentration estimation.

the preprocessing step, described in Section 3.3.2, SAR images are downscaled and normalized and a mask is applied around the land boundary. SAR image patches of a predefined size are cropped at the grid point locations of the image analysis from the preprocessed images. The SAR image patches and the corresponding ice concentration from the image analysis are used to train CNN or FCNN models. Here the best CNN/FCNN model we found are used. The rational for the chosen CNN/FCNN model configuration and the impact of the hyper-parameters on performance are provided later in section 3.6. The performance of the trained CNN model is evaluated by testing on images that are independent of those used in training and validation. Prediction follows the same procedure as testing except that an ice concentration is estimated for every pixel location of the input SAR image.

3.3.2 Preprocessing of SAR images

All the SAR images are sub-sampled by 8×8 block averaging to reduce image noise and reduce data volume by 1/64. Learning at this reduced scale requires smaller spatial context window and therefore smaller neural networks. This is desired because of the limited number of training samples available (0.152 million image analysis sample points) for our study compared to model size (≈ 3.9 million parameters). The sub-sampled images have 400 m pixel spacing with pixel values between 0 and 255. Input normalization is a common practice to improve the performances of CNNs [48, 21]. In this study, the pixel values of the dual-polarized SAR images are normalized by first calculating the mean and standard deviation of pixel values over the entire dataset for each channel, then subtracting from each pixel value this mean, and dividing by the standard deviation. If training sample patches are selected near land, the land pixels may be recognized as ice. This will lead to overestimation (contamination) of ice concentrations near land. The size of land contaminated regions depends on the size of training sample patches. In our case, image patch size 45 by 45 is used, which corresponds to 18 km $\times$ 18 km ground distance. The land contamination can potentially affect water regions within 18 km distance to the coast. This is a concern for the Gulf of Saint Lawrence due to the long coastline. Direct masking out land pixels to 0 is not used because the masked pixels may be confused with dark new ice. Instead, land masks are applied to the SAR images and land pixels are replaced by their corresponding mirrored water pixels to reduce land contamination as illustrated in Fig. 3.4. By doing this, the estimated ice concentration only depends on local ice or water pixels. The actual ice concentration may be changed by the land mirroring process, depending on the shape of the coastline. However, in our testing, land mirroring can largely reduce the effect of land on ice concentration estimation. Therefore, no further investigation on better methods to mask land pixels is performed.

The incidence angle information for each SAR image pixel is calculated from the image meta data using linear interpolation and stored as incidence angle images. These incidence angle images are also normalized to have similar value ranges as the normalized SAR images. Image patches of 45×45 pixels (18 km $\times$ 18 km) centered at grid-point locations of the image analysis charts are extracted from the preprocessed SAR images and the corresponding incidence angle images. Each image patch is a three dimensional matrix of size $3 \times 45 \times 45$ (the last dimension corresponds to 3 values: HH, HV and incidence angle). Each extracted patch and the ice concentration located at the patch center from the image analysis is one sample used to train the CNN. For FCNN, the ice concentration for every SAR image pixel is interpolated from the image analysis charts using nearest neighbor interpolation to increase the number of training samples. The image patches and
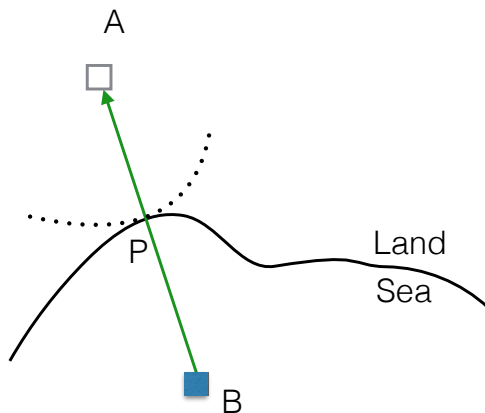
Figure 3.4: An illustration showing how land pixels are replaced with mirrored water pixels to remove land contamination. For land pixel A, its' nearest coast pixel P is found. The value of the water pixel B, located along the same line as AP with the length BP = length AP, is used to replace the value of pixel A.

the corresponding interpolated image analysis are used for training.

Training sample augmentation, which artificially enlarges the dataset by label-preserving transformations, such as rotation and flipping, is a common practice to reduce overfitting and improve the performance of CNN models [48, 14]. In our experiment, training samples are augmented on-the-fly by random rotating and flipping. The random transformed SAR image patches are used for forward-propagation, which corresponds to increasing the training set by a factor of several hundred times.

3.3.3 Structure of CNN and FCNN

CNN and FCNN structures used in this study are illustrated in Table 3.2. The CNN contains three convolutional layers and two fully connected layers on top. ReLU is used as the activation function for all except the last layers for both CNN and FCNN (3.3a). A linear activation function is used in the top layer with one output unit to generate real valued ice concentration (3.3b). This is the best CNN/FCNN configuration we have found though hyper-parameter tuning. Details about the chosen of CNN/FCNN structures are discussed in Section 3.6.

Table 3.2: Structure and configuration of the used CNN and FCNN model. Model CNN/FCNN represents the common structures between the CNN and FCNN model used. For configurations that are different for CNN and FCNN, prefixes "C" and "F" are used for CNN and FCNN separately. The filter size for convolutional layers are in the form of number of filters * filter channels * filter height * filter width.

Model	Layer	Filter/Pooling size	Stride	Pad	Activation	Output)
CNN/FCNN	Data					3 * 45 * 45
	Conv1	64 * 3 * 5 * 5	1	C0 F2	ReLU	C64 * 41 * 41 F64 * 45 * 45
	Pool1	2 * 2	2	C0 F1	Max	C64 * 21 * 21 F64 * 23 * 23
	Conv2	128 * 64 * 5 * 5	1	C0 F2	ReLU	C128 * 17 * 17 F128 * 23 * 23
	Pool2	2 * 2	2	C0 F2	Max	C128 * 9 * 9 F128 * 12 * 12
	Conv3	128 * 128 * 5 * 5	1	C0 F2	ReLU	C128 * 5 * 5 F128 * 12 * 12
CNN	FC4	1024 * (128 * 5 * 5)		0	ReLU	1024 * 1
	Dropout	Drop rate: 0.5				1024 * 1
	FC5	1 * 1024		0	Linear	1
FCNN	Conv4	1024 * 128 * 6 * 6	1, hole 2	5	ReLU	1024 * 12 * 12
	Dropout	Drop rate: 0.5				1024 * 12 * 12
	Conv5	1 * 1024 * 1 * 1	1	0	Linear	1 * 12 * 12

$$f(x) = max(0, x) \tag{3.3a}$$
$$f(x) = x. \tag{3.3b}$$

FCNN shares the same structure with the CNN model for the first three convolutional layers. The two fully connected layers *FC4* and *FC5* in CNN are replaced by two convolutional layers *Conv4* and *Conv5* in the FCNN. The output of the FCNN model has spatial dimension 12×12 which is about 4 times smaller than the input image patch of FCNN. Zero padding is used for the convolutional layers in FCNN to keep the convolutional layer output the same rows and columns of that layer input.

There are a total of 3.9 million weights and 5.3 million weights for CNN and FCNN separately. The increased number of weights for FCNN comes from the *Conv4* layer.

The computation required for a forward pass for this FCNN is roughly 56 times of the computation required for this CNN. However, FCNN estimates 12×12 ice concentrations per forward pass, which is more efficient for prediction than CNN. Due to the parallelized GPU computing, the actual difference between the computational time required for CNN and FCNN are much smaller, which is reported in Section 3.5.2.

Besides data augmentation, dropout is another technique used to reduce overfitting. A dropout layer randomly sets the outputs of neurons (also referred as units) in a layer to zero with predefined probability [35]. Those dropped neurons are not contributing to the forward pass and therefore are not updated in the back-propagation. The use of dropout can reduce the co-adaptations between neurons because a neuron cannot rely on the presence of other neurons [48, 35]. The network is therefore forced to learn more representative features. A dropout layer is used before the last fully connected layer for CNN and before the last convolutional layer for FCNN. The drop rate is set to 0.5, i.e., half of the neurons are randomly chosen and their outputs are set to zero.

3.3.4 Training and Testing

Our network is trained to output the ice concentration from SAR image patches. Instead of using softmax loss [48], which is commonly used in classification CNNs, the L_2 loss is used (3.4) for this regression problem to penalize the discrepancy between the CNN output and the ice concentration provided by the image analysis charts:

$$\mathbb{L}^{L_2}(F(\boldsymbol{x}; \boldsymbol{\theta}), \boldsymbol{z}) = \frac{1}{M} \sum_{m=1}^{M} (F(\boldsymbol{x}; \boldsymbol{\theta})_m - \boldsymbol{z}_m)^2. \tag{3.4}$$

Where $F(\boldsymbol{x}; \boldsymbol{\theta})$ is the network output given input $\boldsymbol{x}$ and parameterization $\boldsymbol{\theta}$. $\boldsymbol{z}_m$ is the ice concentration for the mth sample from image analyses. M is the number of samples used in each training sample batch. For batch sizes of larger than 1, the overall loss of this mini-batch is the average loss of all samples in that mini-batch.

Back-propagation and mini-batch stochastic gradient descent (SGD) [50] are used as the training algorithm. SGD iteratively updates the model weights using the gradient of loss with respect to the model parameters calculated using a subset of the training samples (mini-batch).

The derivatives of loss (3.4) to the network parameters are

$$\frac{\partial \mathbb{L}^{L_2}}{\partial \boldsymbol{\theta}} = \frac{2}{M} \sum (F(\boldsymbol{x}; \boldsymbol{\theta})_m - \boldsymbol{z}_m) \frac{\partial F(\boldsymbol{x}; \boldsymbol{\theta})_m}{\partial \boldsymbol{\theta}}, m = 1, ..., M. \tag{3.5}$$

The derivatives are back-propagated through each pixel in the predictions. The network parameters are updated according to the derivative of the loss to the parameters over each mini-batch, which is described by (3.6).

$$\boldsymbol{V}_{t+1} = \alpha \cdot \boldsymbol{V}_t - r \cdot \epsilon \cdot \boldsymbol{\theta}_t - \epsilon \frac{\partial \mathbb{L}}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}_t} \tag{3.6a}$$

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \boldsymbol{V}_{t+1}. \tag{3.6b}$$

The gradients of the loss with respect to the network parameters ($\partial \mathbb{L}/\partial \boldsymbol{\theta}$) are calculated and averaged over the mini-batch. The weights $\boldsymbol{\theta}$ are updated by V_{i+t} at iteration $t+1$ with learning rate $\epsilon = 10^{-3}$ and weight decay of $r = 2 \times 10^{-5}$ with momentum $m = 0.9$. The setting of the training parameters for SGD is similar to the published setting by Krizhevsky et al. [48]. Adjustments are made by tuning the training parameters sequentially. ϵ is first tuned due to its significant effect on the training results. Then r and α are tuned. Similar to Krizhevsky et al. [48], the parameters of the CNN are initialized by uniform random sampling between -0.05 and 0.05.

An epoch training scheme [50] is adopted. For each epoch, all the training samples are iterated once by the training algorithm. The learning rate is reduced by a factor of 10 for every 20 thousand mini-batches (about 17 epochs). To accelerate the training process, the training is set to stop when the score of the loss function is changing less than 0.001 for 20 consecutive epochs, in case the training converges early (which is typical [77]).

Usually, there are only training and testing datasets for image classification tasks. Due to the large parameter space for CNNs, CNNs tend to overfit easily [48]. A common practice is to use a validation dataset to validate the CNN model during training time [48]. The derived CNN model is evaluated after each training epoch by calculating the loss function on the validation dataset using the current model. The CNN model with the smallest validation error will be selected as the trained CNN. Note that validation is used for model selection and it is therefore part of the training scheme. In this case, the 25 scenes are randomly divided to 17 training images, 4 testing images and 4 validation images.

Once the CNN and FCNN models are trained, ice concentration for each pixel location is estimated by applying the trained model on the target SAR images. Since CNN can only predict a single location in one forward-propagation, the CNN model is used on input images with stride 1, i.e., the input window moves one pixel every time. The output size of FCNN is 12×12, which is roughly $1/4$ times of its input size for the network structure used here. Bilinear interpolation is applied to rescale the FCNN output to the size of its input patch, to simplify the comparison and visualization of the results. To speed up the prediction, the input image window moves 5 pixels each time, and 7 pixels are overlapped

34

between neighboring predicted patches. Only the central 5 by 5 pixels of the interpolated FCNN output are used.

3.3.5 Implementation

Caffe [38], a popular C++ open-source deep learning package, is used in this study. It provides a ready to use implementation of CNN and FCNN. SAR image preprocessing and patching are implemented in Python. A data layer is implemented using C++ under Caffe to read the patched images and their corresponding image analyses and output pairs of inputs and targets for the CNN or FCNN model. In-situ training sample augmentation is also implemented in the data layer.

3.4 Existing SAR Ice Concentration Algorithm Implementation

For the purpose of evaluation, a variation of the recent ice concentration estimation algorithm developed by Karvonen [44] is implemented in Matlab. Karvonen's work is chosen because of his extensive publication record of developing algorithms using SAR data for automated ice monitoring. In Karvonen's paper, a preliminary ice concentration is first estimated from the autocorrelation of HH pol SAR images by a segmentation based approach [43]. The preliminary ice concentration is then used with four other SAR image features (HV, HV/HH, (HH-HV)/HH, and incidence angle) in a fully connected neural network (also named multi-layer perceptron) with one hidden layer of 10 units.

In our implementation, the ice concentration is estimated directly from a fully connected neural network with one hidden layer of size 40. Ten GLCM features are used in addition to the five features used by Karvonen. These ten GLCM features are identified as the most important ten SAR image features from a pool of 172 SAR image features for distinguishing ice and water [56]. These features should also benefit the ice concentration estimation tasks. This implementation is referred to as MLP40 (multi-layer perceptron network with one hidden layer of 40 units). The full list of image features used is listed in Table 3.3.

There are four major differences between our implementation of Karvonen's paper and the original implementation [44]:

1. The original implementation used an initial ice concentration estimate (generated

Table 3.3: Image features used for method MLP40

#	Pol	Feature
1	HV	GLCM mean 25 by 25 step 5
2	HH	GLCM correlation 51 by 51 step 5
3	HH	GLCM mean 25 by 25 step 1
4	HH	GLCM dissimilarity 51 by 51 step 20
5	HH	GLCM applied second moment 101 by 101 step 5
6	HH	Intensity
7	HV	Average 25 by 25 window
8	HH	Average 5 by 5 window
9	HH	GLCM dissimilarity 51 by 51 step 5
10	HH	GLCM mean 101 by 101 step 20
11	HV	Intensity
12	HH,HV	HV/HH
13	HH,HV	(HH-HV)/HH
14	HH	Intensity autocorrelation
15		Incidence angle

from HH pol autocorrelation) as an input into the MLP. Our implementation just uses the HH pol autocorrelation [43] as an input to the MLP.

2. SAR image segmentation results are used to produce smoother estimates in the original implementation which are not included in our implementation.

3. A set of GLCM features are used as input to the MLP in our implementation.

4. The original implementation used an MLP with 10 hidden neurons while the MLP used here has 40 hidden neurons.

The HH pol ice concentration is generated from the distribution of autocorrelations in a segmentation [43]. A linear relationship between the ice concentration and the HH pol autocorrelation is assumed [43]. When estimating ice concentration of pixels, each pixel can be considered to be a small segment. Therefore, the use of HH pol ice concentration in an MLP is equivalent to the use of HH pol autocorrelation in an MLP for pixel level ice concentration estimation tasks. Segmentation is also removed from our implementation for fair comparison with the CNN and FCNN, which do not use segmentation. To represent the additional input image features in the hidden layer, the number of hidden neurons

needs to be increased. The resulting MLP has higher ratio of hidden neurons to input features (40/15) than Karvonen's implementation (10/6). As a result, our implementation of Karvonen's method is an augmented version of the original implementation in [43, 44] for pixel level concentration estimation.

In Leigh's paper, image features are extracted from 4 by 4 block averaged SAR images [56]. The 8 by 8 block averaged SAR images are used in this thesis. For consistency, the image features are first calculated from 4 by 4 block averaged SAR images as done in Leigh's paper. Then the calculated image features are averaged for every 2 by 2 block. The same training scheme used by Karvonen is used to train the MLP [44].

3.5 Results

3.5.1 Evaluation

The ice concentration estimated from the SAR images using the CNN and FCNN described in Section 3.3.3, as well as ice concentration from ASI and MLP40 are evaluated against image analyses in the SAR image space. In other words, each image analysis sample point is compared to the ice concentration of its nearest pixel in the associated SAR scene, which means the image analysis samples are used at a finer spatial resolution than what the analyst intended. The mean error (E_{sgn}), mean absolute error (E_{L1}), error standard deviation (E_{std}) and root mean squared error (E_{rmse}) are calculated for evaluation purposes using (3.7)

$$E_{sgn} = mean(IC - ImA) \tag{3.7a}$$

$$E_{L1} = mean(|IC - ImA|) \tag{3.7b}$$

$$E_{std} = std(IC - ImA) \tag{3.7c}$$

$$E_{rmse} = sqrt(mean[(IC - ImA)^2]). \tag{3.7d}$$

The term IC denotes the estimated ice concentration using CNN or FCNN and ImA denotes the ice concentration from the image analysis for the rest of the thesis. Due to the overestimation in image analysis [10], the actual error of the estimates are asymmetric, i.e., for the same discrepancy to the image analysis, underestimated ice concentration contains less error than overestimated ice concentration statistically. When there is a lack of better ice concentration data, these error measurements provide a reasonable estimation of the quality of the ice concentration estimates.

While the ice concentration derived from the image analysis is a discrete number (0-10) scaled between 0 and 1 (0, 0.1, ..., 1.0), the ice concentration from FCNN and CNN is determined as a real number between 0 and 1, this difference may introduce errors into the evaluation statistics. Therefore, the ice concentration estimates are also quantized by rounding to 11 levels between 0 and 1 and re-evaluated against the image analyses. The evaluation results are similar with slight improvement after quantization, and are therefore not shown in this thesis.

The evaluation results for training, testing and validation datasets are given in Table 3.4. Both CNN and FCNN produce more accurate ice concentration estimates than ASI and MLP40. ASI underestimates ice concentration by around 25% when compared with image analyses (Table 3.4). Since the CNN and FCNN are trained using image analysis charts, these methods are expected to have lower errors than ASI since (a) ASI is not trained with image analysis charts and (b) results are based on comparison to image analysis charts. Previous studies reported that the ASI ice concentration normally has errors less than 10% for middle and high ice concentrations [86]. The large underestimation of ice concentration for middle and high ice concentrations by ASI observed in this study is mainly caused by the underestimation of ice concentration for new ice and gray ice (i.e., thin ice). The underestimation of ice concentration is improved by CNN and FCNN compared to MLP40.

The error standard deviation (E_{std}) for testing is at the same level as training and validation for CNN and FCNN. The performance of testing does not show a large decrease relative to the performance of training and validation, which indicates a low level of over-fitting for the trained CNN and FCNN models. Note that the validation errors are larger than testing errors for MLP40. This might be caused by the insufficient testing samples used, which could lead to different distributions of image surface types for validation and testing images.

Statistical significance of the improvement in E_{rmse} of FCNN to CNN is evaluated using the test dataset, because E_{rmse} is used as the main criterion for performance evaluation. As the empirical distributions of the prediction errors for CNN and FCNN are similar in shape to a Gaussian, the Gaussian distribution is used to model the prediction errors of the test dataset. This means that the statistical significance of the E_{rmse} can be assessed using a z-test, with the E_{rmse} assumed to follow a chi-squared distribution [93]. The null hypothesis is that the E_{rmse} of FCNN and CNN has the same distribution. The calculated p-value is 0.007 from a z-test [93]. Therefore, the improvement in FCNN to CNN is statistically significant for significance level 0.01. Similarly, the improvement of CNN compared to MLP is also shown to be statistically significant (p-value $<< 0.001$). The error for ASI is strongly biased and cannot be modeled using a Gaussian distribution. Also, the error of ASI is considerably higher than that of the other methods. Therefore, the statistical

significance test of MLP40 against ASI is not conducted.

Table 3.4: Average error statistics across different methods for Gulf of Saint Lawrence dataset

Method	Set	E_{sgn}	E_{L1}	E_{std}	E_{rmse}
ASI	Training	-0.2423	0.2605	0.3207	0.4020
	Validation	-0.3416	0.3768	0.3693	0.5031
	Testing	-0.2717	0.2877	0.3097	0.4121
MLP40	Training	0.0002	0.1460	0.2050	0.2049
	Validation	-0.0410	0.2381	0.2986	0.3015
	Testing	-0.0819	0.1727	0.2325	0.2466
CNN	Training	-0.0039	0.0845	0.1506	**0.1507**
	Validation	-0.0123	0.1253	0.2056	0.2059
	Testing	-0.0274	0.1295	0.2197	0.2214
FCNN	Training	0.0146	0.0810	0.1559	0.1565
	Validation	-0.0082	0.1257	0.2011	**0.2012**
	Testing	-0.0055	0.1239	0.2141	**0.2142**

Fig. 3.5 shows the means and standard deviations of ice concentration estimation errors for different ice concentration bins for training, validation and testing datasets. There is a clear trend between image analyses and ice concentration estimates generated from SAR images for all three sets in general. ASI shows underestimation for almost all ice concentration levels, with larger underestimation for higher ice concentration values. MLP40 overestimates ice concentration for water regions by about 15% for all three datasets, and underestimates ice concentration by 20% to 40% for training, testing and validation. Both CNN and FCNN show relatively less overestimation for water regions and less underestimation for ice regions compared to MLP40. For water, CNN and FCNN overestimate on average by approximately 5% for training and 10% for testing and validation. For ice (ice concentration is 1), CNN and FCNN underestimate ice concentration by 10% on average for all three sets. The estimation of pure water or ice generally has smaller error standard deviation than the estimate for intermediate ice concentration levels. This might be caused by the abundant water samples and ice samples in the training dataset (Fig. 3.6), or the better quality (less errors) of ice/water samples than samples of intermediate ice concentration levels. It is reasonable to assume that the estimation precision can be improved by using more training samples of intermediate ice concentration levels.

The ice concentration estimates for scene 20140117_103914 in the test set are illustrated
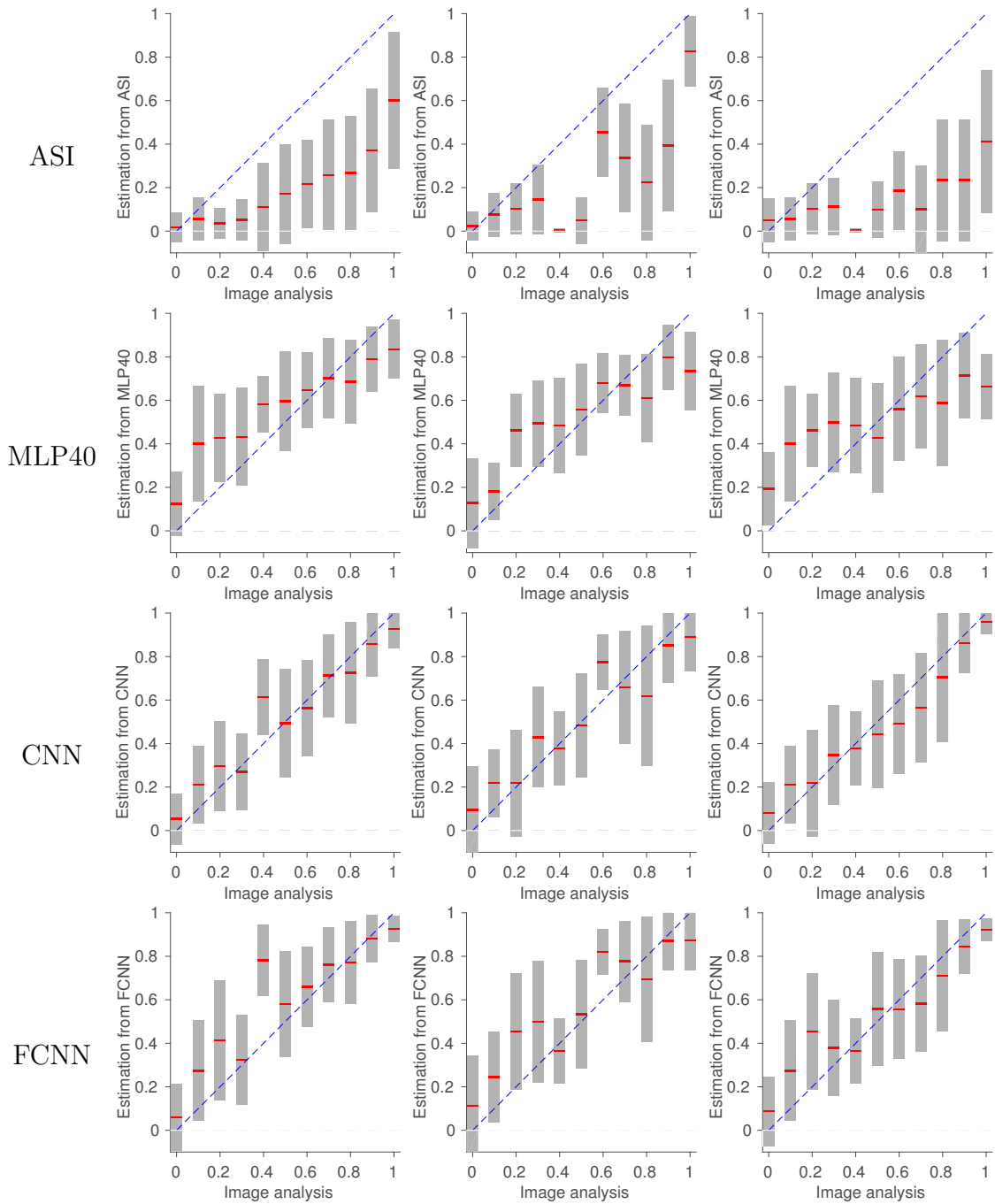
Figure 3.5: Errors at different ice concentration levels for ASI (1st row), MLP40 (2nd row), CNN (3rd row) and FCNN (4th row) for training (1st column), validation (2nd column) and testing (3rd column) sets. The red lines represent mean error, and half length of a bar represents error standard deviation.

(a) Training   (b) Validation   (c) Testing
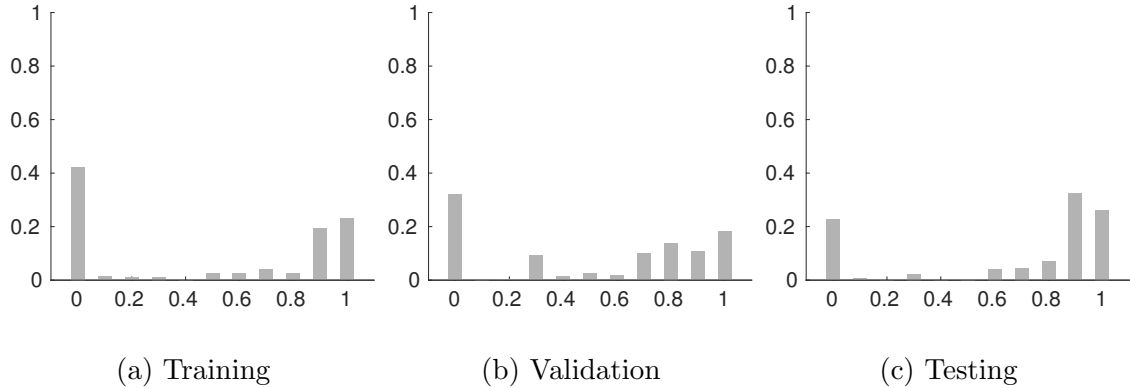
Figure 3.6: Histogram of image analysis ice concentration for training, validation and testing dataset of the Gulf of Saint Lawrence. The training samples are strongly biased since the majority of the training samples are water and ice.

in Fig. 3.7. This scene is selected due to its diverse ice features and ice concentration values. Due to the fact that the CNN and FCNN estimates are visually similar for this scene, only the result of CNN is shown. There is an obvious incidence angle effect that can be seen in the HH image (Fig. 3.7(a)). The tone is brighter at the near-range portion (right) of the scene. There is no obvious effect of incidence angle visible in the ice concentration estimation in Fig. 3.7(e-f). This indicates that both the CNN and MLP40 are able to model the incidence angle effect well through learning. Most of the ice is underestimated as pure water by ASI (Fig. 3.7(d)). MLP40 generally overestimates the ice concentration for water regions (Fig. 3.7(e)). CNN is able to produce less noisy estimates over both ice and water regions (Fig. 3.7(f)) than MLP40.

All SAR based algorithms produce ice concentration estimates with more details and clearer ice-water boundaries than the ASI data (Fig. 3.7(d, e, f)), which is partly due to the higher resolution of SAR images. Due to the direct use of SAR image pixel values, MLP40 is more responsive to backscatter changes in SAR images than CNN and FCNN. Therefore, MLP40 is able to capture the finest details in SAR images in all the methods tested, which also introduced large errors into the estimated ice concentration for both ice and water regions (Fig. 3.8(e)). The ice concentration estimates by CNN/FCNN contain less noise than the result of MLP40, but still contain more details than the image analysis charts, especially in low ice concentration regions and marginal ice zones (Fig. 3.8(f), (g)). This may be caused by the difficulty to manually identify accurate boundaries of low ice concentration regions by ice analysts or the limited number of polygons they can use for each image. As the image analyses contain ice concentration values at coarser scales

(a) HH   (b) HV   (c) ImA

(d) ASI   (e) MLP40   (f) CNN

0   0.1   0.2   0.3   0.4   0.5   0.6   0.7   0.8   0.9   1
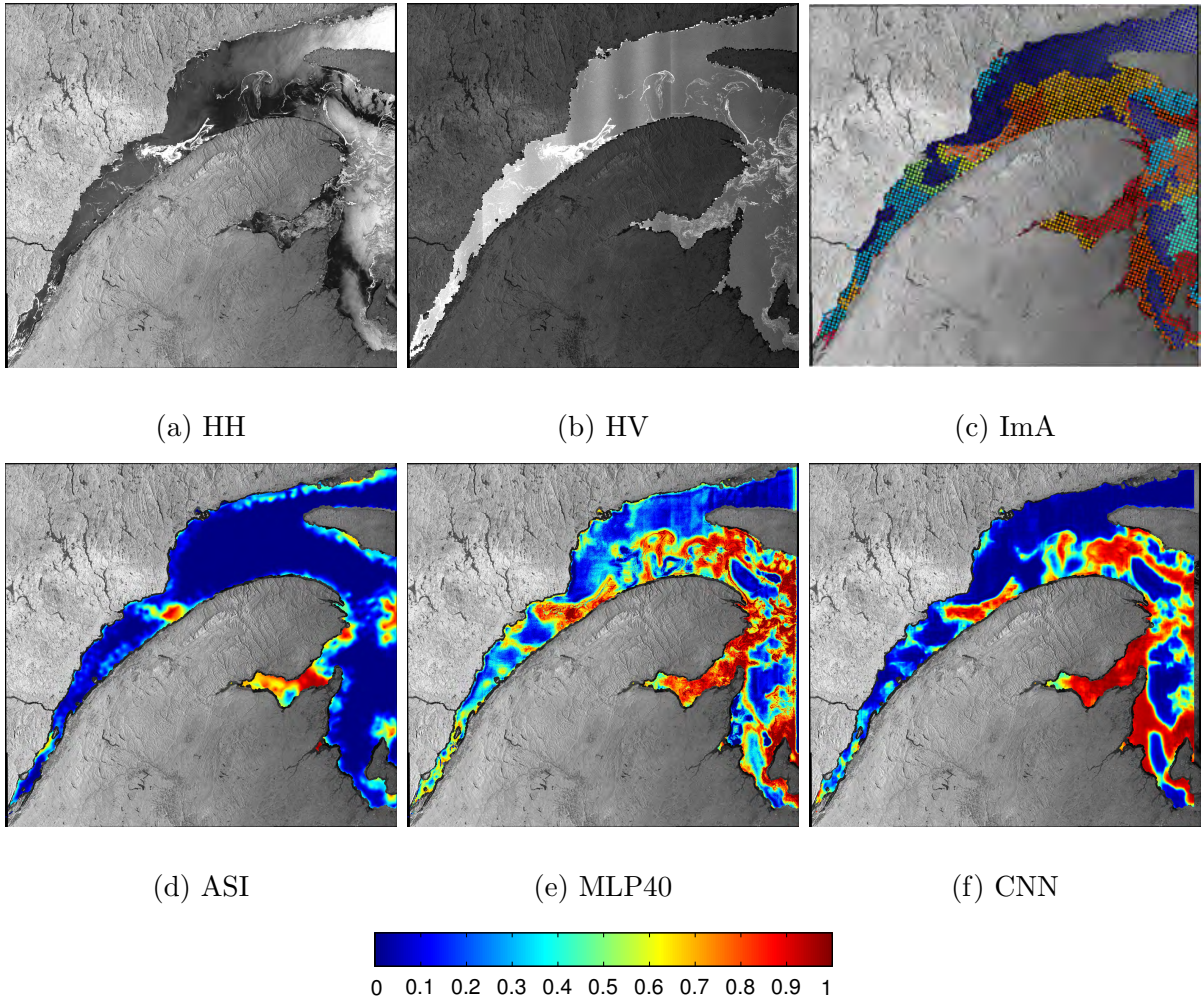
Figure 3.7: The ice concentration estimation result by CNN compared to ASI and MLP40 ice concentration for scene 20140117_103914, which is used for testing. (c) ImA refers to image analysis.

than SAR image pixels, CNN/FCNN models trained using the image analyses produce ice concentration estimates at a similar spatial scale to the image analyses, which leads to results with blurrier ice-water boundaries and less detail of ice structures compared to SAR images.

New ice can be easily misinterpreted as water in automatic SAR image analysis methods and also in passive microwave ice concentration retrieval algorithms [86]. The dark regions

(a) HH   (b) HV   (c) ImA   (d) ASI

(e) MLP40   (f) CNN   (g) FCNN

0   0.1   0.2   0.3   0.4   0.5   0.6   0.7   0.8   0.9   1
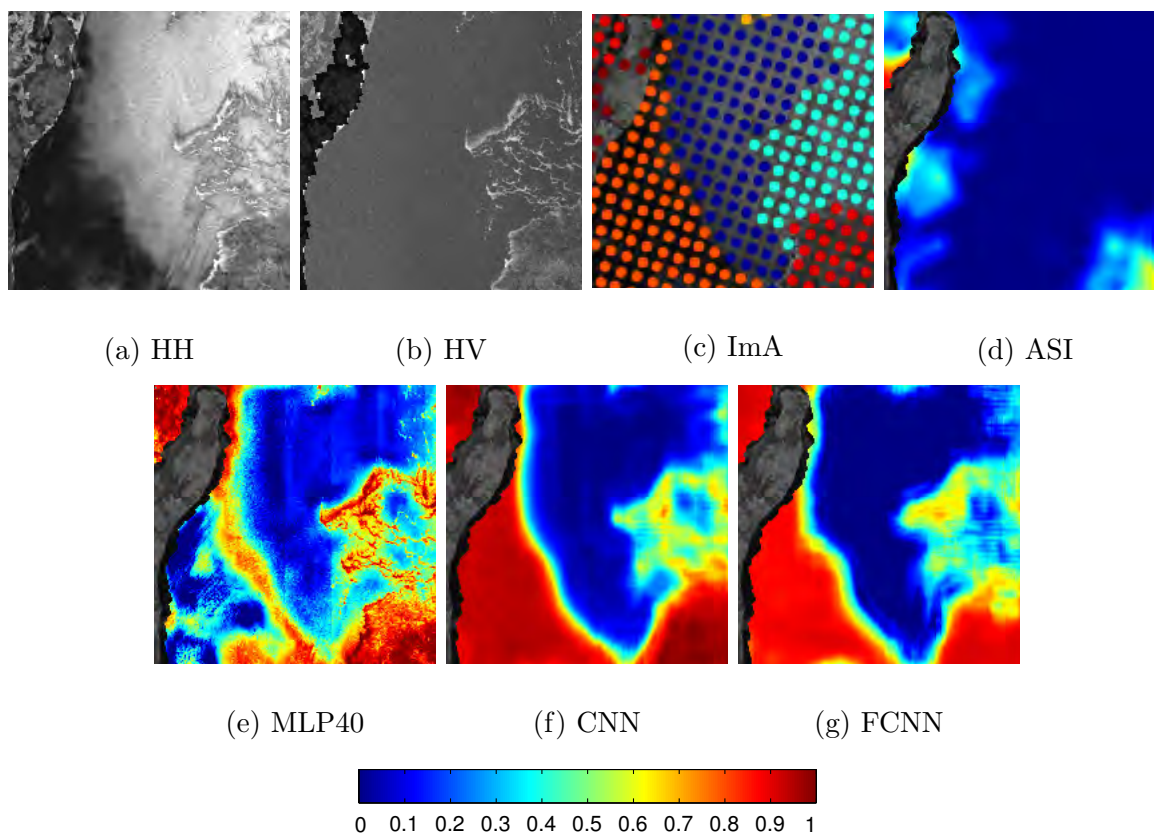
Figure 3.8: An example shows the details of the results for new ice and water. (d) ASI result is mainly water for this region. (e) MLP40 produces noisy ice concentration estimates with new ice in the bottom left been identified as water and low ice concentration levels. Both (f) CNN and (g) FCNN are able to correctly identify new ice and water with less noise. (g) FCNN generates improved water estimates than CNN near ice. The color bar in the bottom is the color scheme used for (c) to (g). Subscene of dimension 6 km × 6 km from 20140117_103914.

in middle top and right bottom of Fig. 3.7(a) are new ice according to the ice chart, which have similar appearances with calm water at high incidence angles. The details of the results for the right bottom of Fig. 3.7(a) are compared in Fig. 3.8. CNN and FCNN are found to correctly identify this region, as corresponding to one with high ice concentration, while ASI and MLP40 misidentified this region as low ice concentration. However, some new ice in the middle top of Fig. 3.7(a) is still misidentified as water. Some regions with similar SAR image appearances and incidence angles are found to be identified as

water by the image analysis in the training set (Fig. 3.9). This might have caused the misidentification of new ice as water.

(a) HH   (b) HV   (c) ImA

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
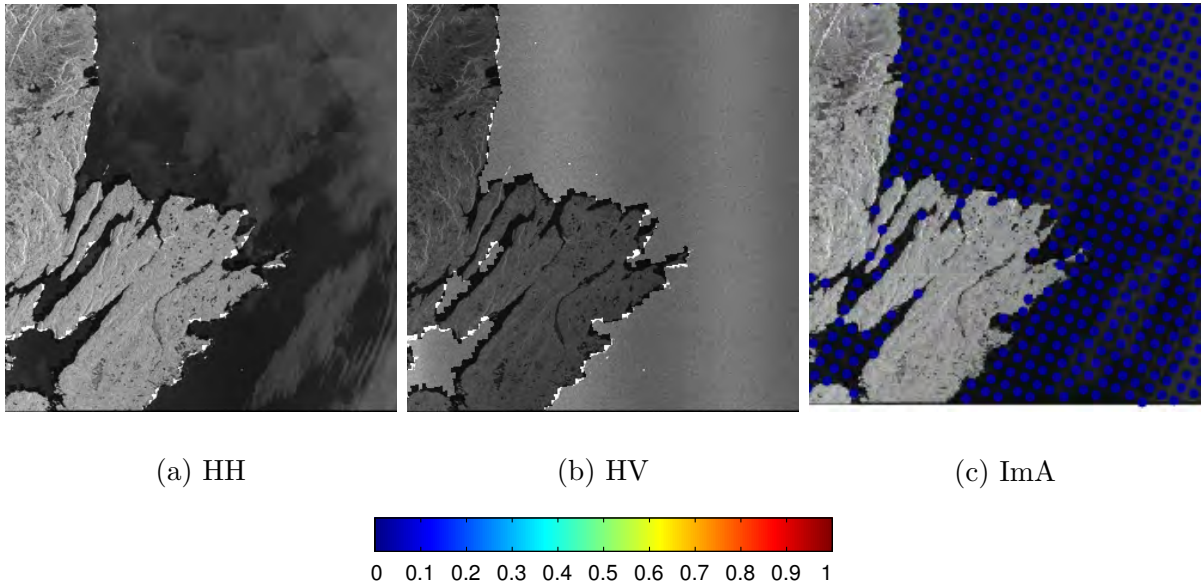
Figure 3.9: Image analysis used for training identifies regions as water, which have similar appearance and incidence angle to the dark new ice in the middle top of Fig. 3.7(a). (c) ImA refers to image analysis. Subscene of dimension 100 km × 100 km from 20140118_101002 in the training set.

Fig. 3.7(b) shows a moderate banding effect in HV pol. The banding effect does not appear to have an obvious impact on the CNN result for this scene. CNN and FCNN are found to be more robust to banding than MLP40. However, strong banding may cause overestimation of ice concentration (spurious ice) for water regions. Such an example is given in Fig. 3.10, where both CNN and FCNN overestimate ice concentration for water regions with strong banding in the HV pol.

Calm water can generally be correctly identified when the banding effect is not strong. Strong wind roughening can lead to noisy ice concentration estimates in some open water regions for both CNN and FCNN. An example is shown in Fig. 3.11, where CNN and FCNN overestimate ice concentration for some of the water regions. The estimate of MLP40 has a similar pattern to the roughening of water in the HH pol image, e.g., stronger wind roughening leads to stronger overestimation. The estimates of CNN and FCNN are not correlated with the strength of roughening but appears more randomly in this region. Later we will demonstrate in Chapter 4 that this is caused by training sample errors.

(a) HH   (b) HV   (c) ImA   (d) ASI

(e) MLP40   (f) CNN   (g) FCNN

0   0.1   0.2   0.3   0.4   0.5   0.6   0.7   0.8   0.9   1
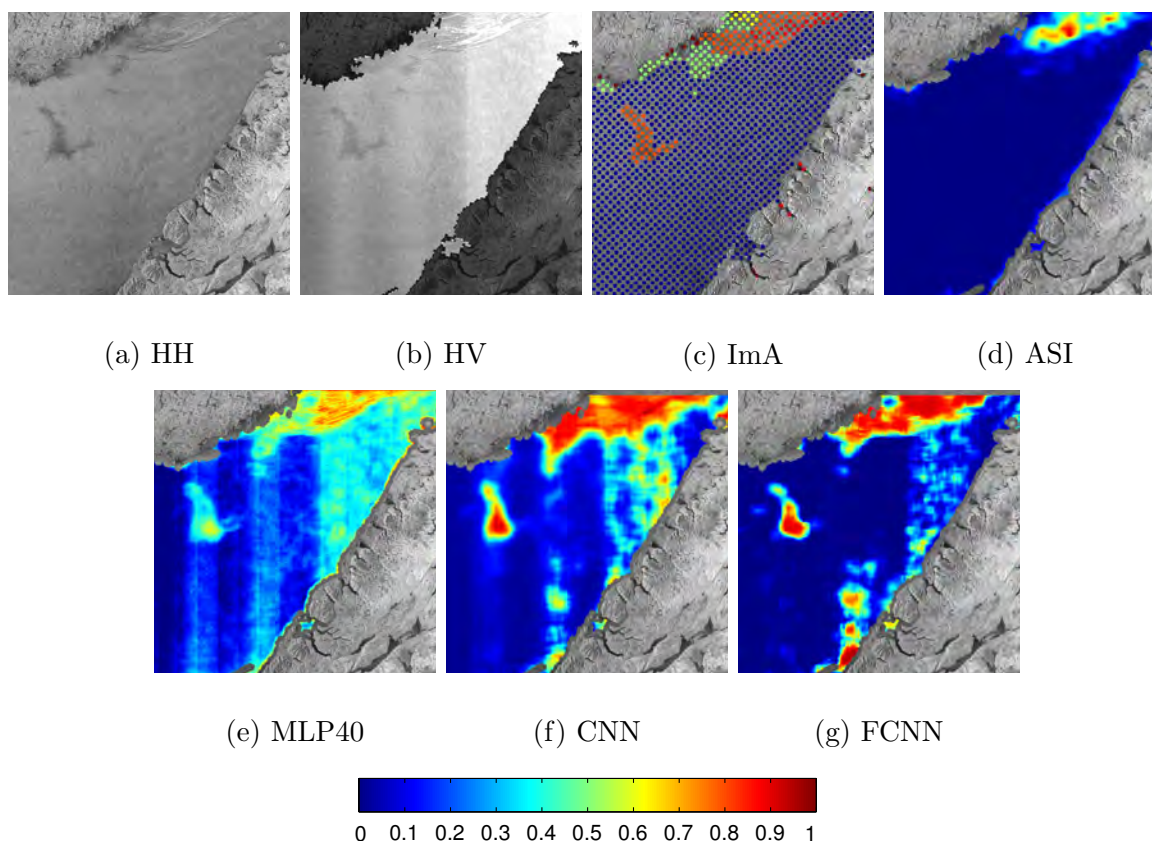
Figure 3.10: Example of water misidentified as ice for both MLP40, CNN and FCNN due to the banding effect in HV pol. Water in the right part of (b) HV pol is obviously brighter than water in the left. The overestimate in water regions for (e) MLP40, (f) CNN and (g) FCNN coincide with the location and shape of the banding in (b) HV. Subscene of dimension 20 km × 20 km from 20140121_214420.

3.5.2   Comparison of CNN and FCNN

The details of CNN and FCNN estimates for a region of scene 20140117_103914 are shown in Fig. 3.8. CNN and FCNN generate similar ice concentration estimates in general. There are several small differences. CNN mis-identifies water regions near ice with non-zero low ice concentration, while FCNN identifies more water regions with 0 ice concentration. FCNN also generates better water predictions than CNN when the banding effect is strong (Fig. 3.10). As FCNN uses interpolated image analyses as training samples, the number of effective water and ice training samples are increased for FCNN, which may partially
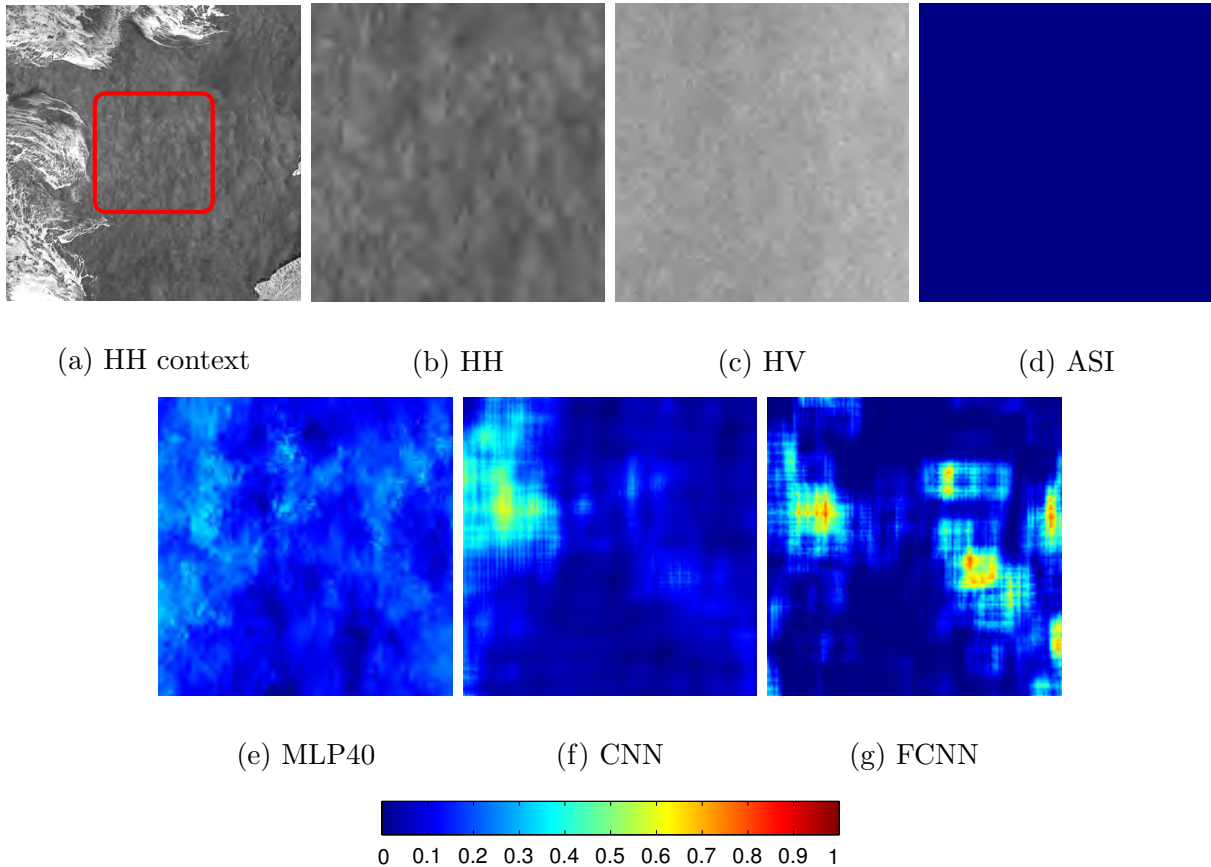
Figure 3.11: Overestimation for wind roughened water by (e) MLP40, (f) CNN and (g) FCNN. (b-g) correspond to the region in the red window in (a). Subscene of dimension 6 km × 6 km from 20140124_215646.

account for FCNN's improved water predictions.

The main advantage of FCNN over CNN is its significantly faster prediction, which is essential for large-scale applications. For each forward pass, CNN predicts one location/pixel. There are redundant convolutional computations when predicting neighboring locations due to the large overlap of input patches. FCNN predicts multiple locations in each forward pass. The convolutional operations are shared between locations, which reduces the computational redundancy. In our experiments using one Nvidia GTX 780, one four-core 3.6GHz CPU and 32Gb memory, the typical computational time required to predict a 500 km by 500 km sea surface using the developed FCNN model is less than 3 minutes, while CNN normally takes 13 minutes. However, FCNN takes longer time to train

Table 3.5: The average error statistics for networks trained with or without incidence angle data using CNN on the Gulf of Saint Lawrence data

	Set	E_{sgn}	E_{L1}	E_{std}	E_{rmse}
w/ incidence angle	Training	-0.0039	0.0845	0.1506	0.1507
	Validation	-0.0123	0.1253	0.2056	0.2059
	Testing	-0.0274	0.1295	0.2197	0.2214
w/o incidence angle	Training	0.0052	0.0817	0.1434	**0.1435**
	Validation	0.0035	0.1183	0.1837	**0.1836**
	Testing	-0.0119	0.1220	0.2031	**0.2035**

compared to CNN due to the increased computations by zero padding and the removal of the pooling layer. In our experiment, CNN and FCNN converge at similar iterations, but the training time for one iteration of FCNN is 1.5 times of the time required for CNN.

3.6 Architecture Evaluation

Patch size and network depth are found to have substantial impact on the estimated ice concentration. Therefore, a sensitivity analysis for these two hyper-parameters is carried out. The effect of using incidence angle data is also evaluated. Our tuning experiments suggest that the model is not very sensitive to the selection of other parameters, such as the size and number of convolutional filters, pooling size, number of fully connected neurons. Therefore, these hyper-parameters are chosen based on experience from previous studies [48] without further tuning.

3.6.1 Use of Incidence Angle Data

The brightness of water in HH pol is strongly affected by incidence angle. Water at low incidence angles can have similar backscatter as dark new ice. Therefore, incidence angle may provide useful information for the separation of ice and water. Two CNNs are trained with and without incidence angle data as the third input channel using the network structure and training parameters introduced earlier this chapter.

The errors evaluated against image analyses are illustrated in Table 3.5. The two models have similar E_{rmse}.

The use of incidence angle is found to reduce the banding effect on the estimated ice concentration, but also leads to results that are more sensitive to wind roughening, which may have caused larger E_{rmse} for the network trained with incidence angle data, as illustrated in Fig. 3.12. The results without incidence angle (Fig. 3.12(d)) are more affected by the banding in the HV pol. The errors at water regions for results with incidence angle (Fig. 3.12(c)) are large but do not coincide with the location and shape of banding. The result without incidence angle also has cluttered errors, but the magnitude of the errors is small. The use of incidence angle can help the information extraction from HH pol and therefore may reduce the dependency of the model on HV pol, which could have led to the reduced banding effect. The cluttered errors in water are not found to be related to any obvious patterns in the image. They may be caused by training sample errors, especially representation errors, which may label some water regions with high ice concentration. When incidence angle data are used, the model is larger, which is more likely to overfit and preserve the errors in the training samples. More new ice may be correctly identified when incidence angle is used (Fig. 3.13). The use of incidence angle can also help to distinguish new ice from water, especially when the incidence angle is low and water is dark in general. Because the cluttered random errors in water regions may be removed if the training sample errors are modeled properly, the incidence angle data are used in our model for the robustness to banding effect and the improved detection of new ice.

(a) HH   (b) HV   (c) w/ IA   (d) w/o IA

0   0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
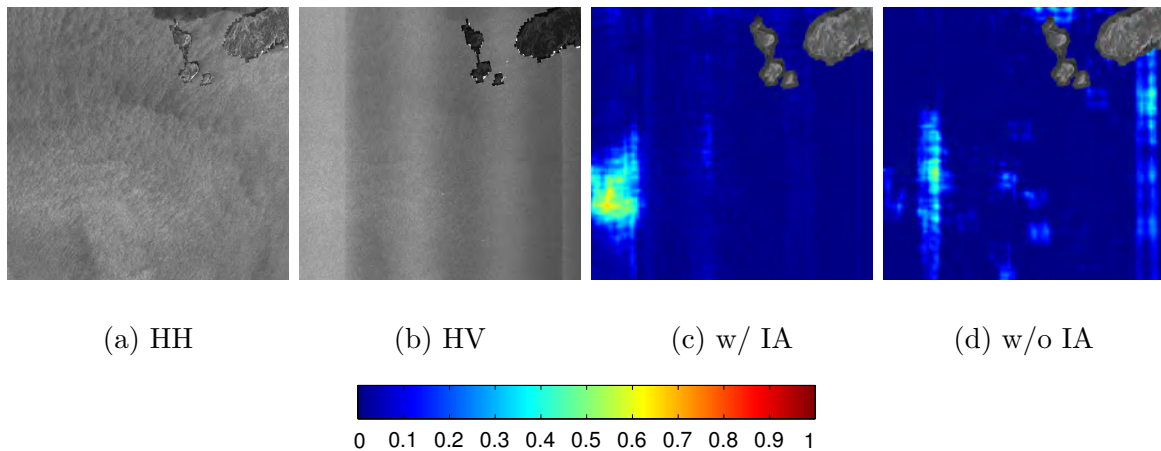
Figure 3.12: The effect of incidence angle data on water predictions. (c) using incidence angle information reduces the impact of banding effect, but causes blocky errors in water. (d) result of not using incidence angle data has errors coincide with the banding in HV pol for water regions. Subscene of dimension 16 km × 16 km from 20140208_095758.

(a) HH	(b) ImA	(c) L2 w/ IA	(d) L2 w/o IA

0   0.1   0.2   0.3   0.4   0.5   0.6   0.7   0.8   0.9   1
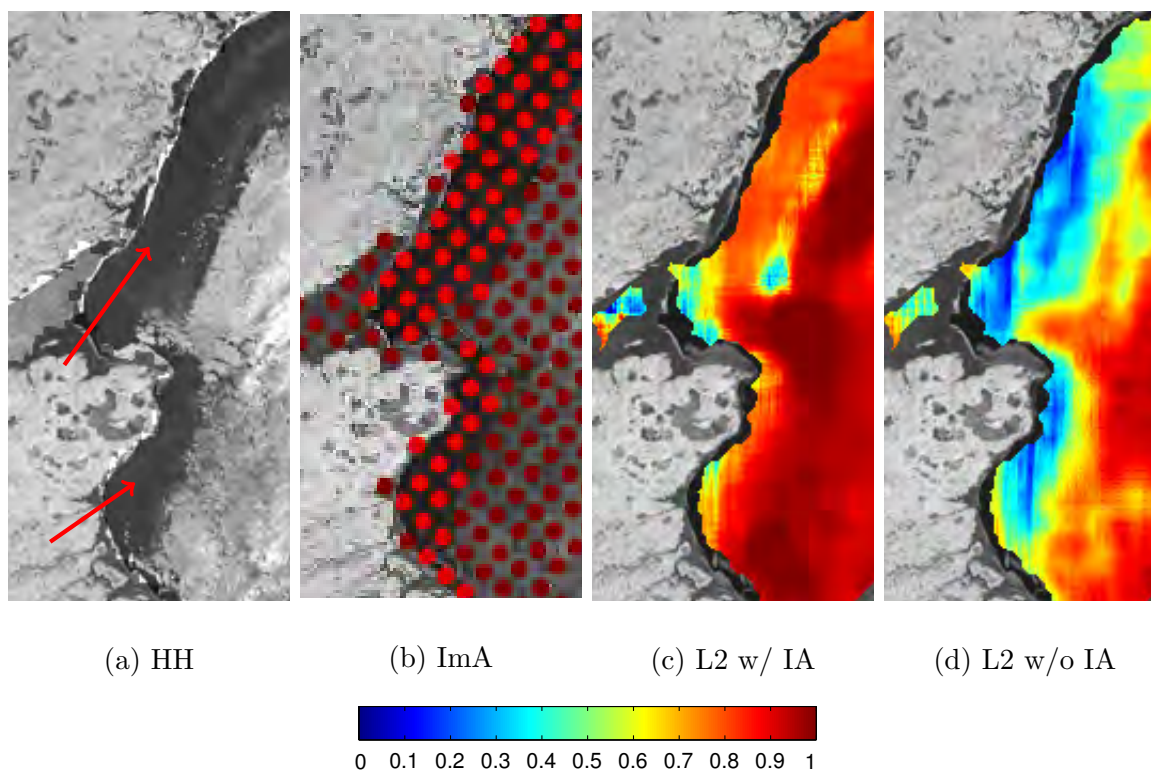
Figure 3.13: New ice (the arrow pointed dark regions along the coast) been correctly identified as ice when incidence angle data are used. Subscene of dimension 15 km × 6 km from 20140206_221744.

3.6.2   Patch Size

The patch size of the input image and size of convolutional filters are related to the intrinsic scale and complexity of the problem. In this tuning experiment, patch sizes of 25, 35, 45 and 55 are tested, which correspond to ground distance of 10 km (25 * 400 m), 14 km (35 * 400 m), 18 km (45 * 400 m) and 22 km (55 * 400 m). Fig. 3.14 shows the E_{rmse} of the output of CNNs using different input patch sizes. With larger patch size, the model fits the training data better and the E_{rmse} of the training data decreases. The E_{rmse} for test and validation data decreases when the patch size increases from 25 to 45. However, when the patch size increased from 45 to 55, the E_{rmse} for test and validation slightly increased, which could be an indication of slight overfitting for the dataset used. Therefore, 45 (18 km ground distance) is the patch size used in this study. Note that this is only the case for the dataset used in this study. When more training samples are available, a larger patch

49

size might produce better results.
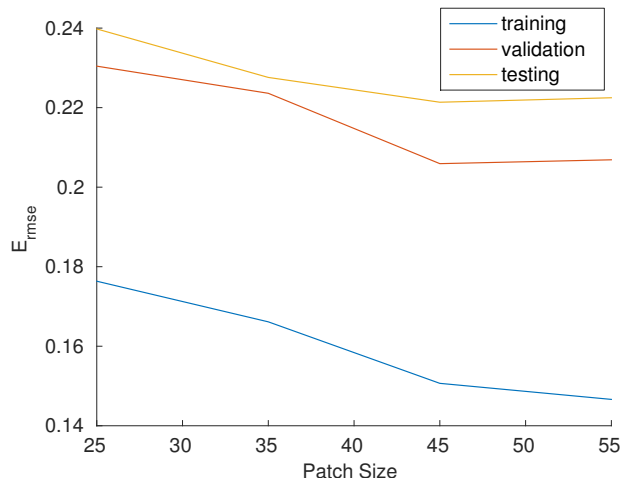
Figure 3.14: Comparison of the E_{mse} for different patch sizes.

Patch size is found to have an impact on the results with most of the difference coming from the banding effect contaminated regions and wind roughed water regions. A visual comparison of the results for different patch sizes is illustrated in Fig. 3.15. A patch size of 45 (Fig. 3.15(e)) was found to generate the best prediction of water under conditions of wind-roughening, while smaller patch sizes (Fig. 3.15(c-d)) strongly overestimate ice concentration in water regions due to wind and banding effect. These results suggest that the separation of wind roughened water and ice requires spatial context information over a larger region. In contrast, ice is generally well identified for all tested patch sizes. Using small patch sizes tends to slightly underestimate ice concentration, leads to ice cover that is less homogeneous, as compared to larger patch sizes. However, larger patch size means larger model and requires more training samples and higher computational complexity.

3.6.3   Network Depth

Network depth is an important parameter of the network which determines the level of abstractions used for classification or regression. Here, CNN models with two and three convolutional layers are trained and evaluated. The error statistics against image analyses are illustrated in Table 3.6. The use of two or three layers of convolutional neural networks generate similar error statistics.

(a) HH (b) HV (c) ImA

(d) Patch size 25 (e) Patch size 35 (f) Patch size 45 (g) Patch size 55
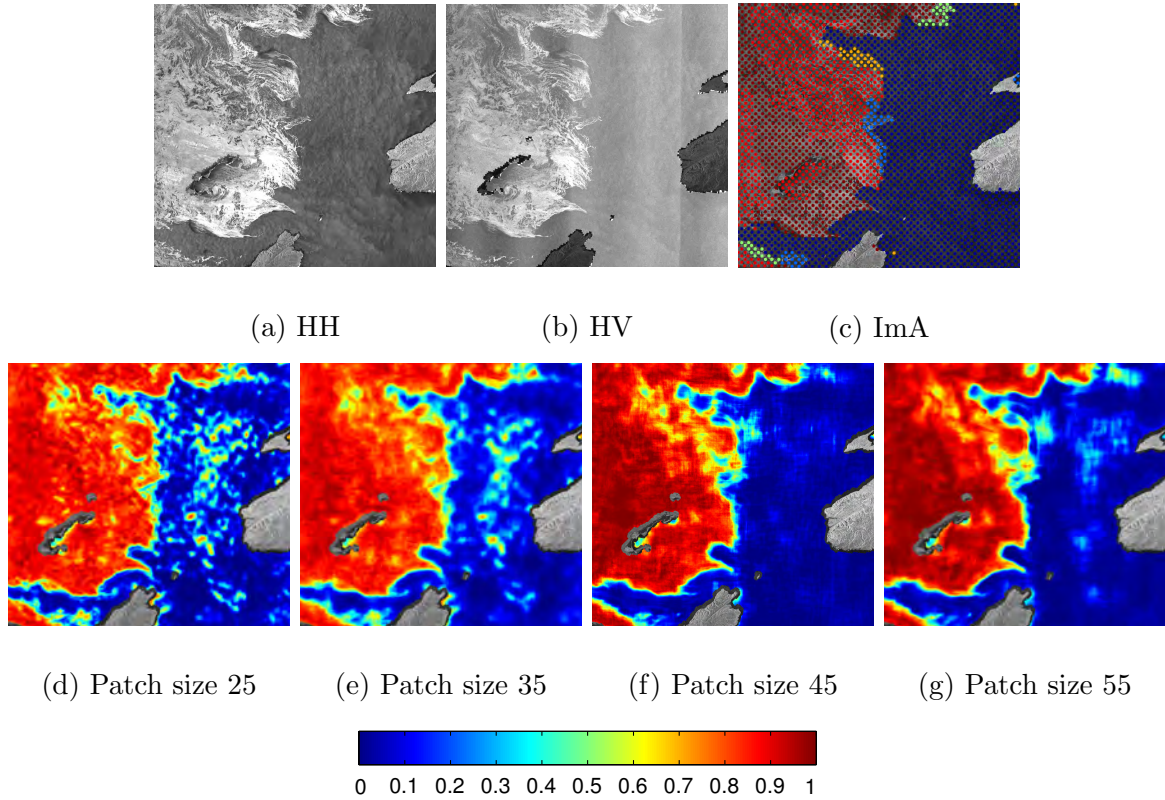
Figure 3.15: Visual comparison of different patch sizes. Estimate for ice is improved when patch size increases. Patch size 45 which corresponding to ground distance of 18 km has cleaner water estimates than the rest. Subscene of dimension 26 km × 26 km from 20140124_215646.

Table 3.6: Average error statistics for networks with two convolutional layers and three convolutional layers on the Gulf of Saint Lawrence dataset

Set	Two Convolutional Layers				Three Convolutional Layers			
	E_{sgn}	E_{L1}	E_{std}	E_{rmse}	E_{sgn}	E_{L1}	E_{std}	E_{rmse}
Training	-0.0055	0.0874	0.1266	**0.1269**	-0.0039	0.0845	0.1506	0.1507
Validation	-0.0028	0.1229	0.1933	**0.1934**	-0.0123	0.1253	0.2056	0.2059
Testing	0.0054	0.1556	0.2300	0.2302	-0.0274	0.1295	0.2197	**0.2214**

Visually, the network with three convolutional layers produces smoother ice concentration estimates as shown in Fig. 3.16. This makes sense as deeper networks extract more

51

abstract features so that the results are less sensitive to raw pixel values. The network with three convolutional layers also produces more reasonable ice concentration estimation. The water regions in Fig. 3.16(a) that are incorrectly identified by the network with two convolutional layers (Fig. 3.16(c)) are correctly identified by the network with three convolutional layers (Fig. 3.16(d)). Ice concentration for new ice is also improved by using deeper structures as shown in Fig. 3.17.

(a) HH   (b) ImA   (c) Two conv. layers   (d) Three conv. layers

0   0.1   0.2   0.3   0.4   0.5   0.6   0.7   0.8   0.9   1
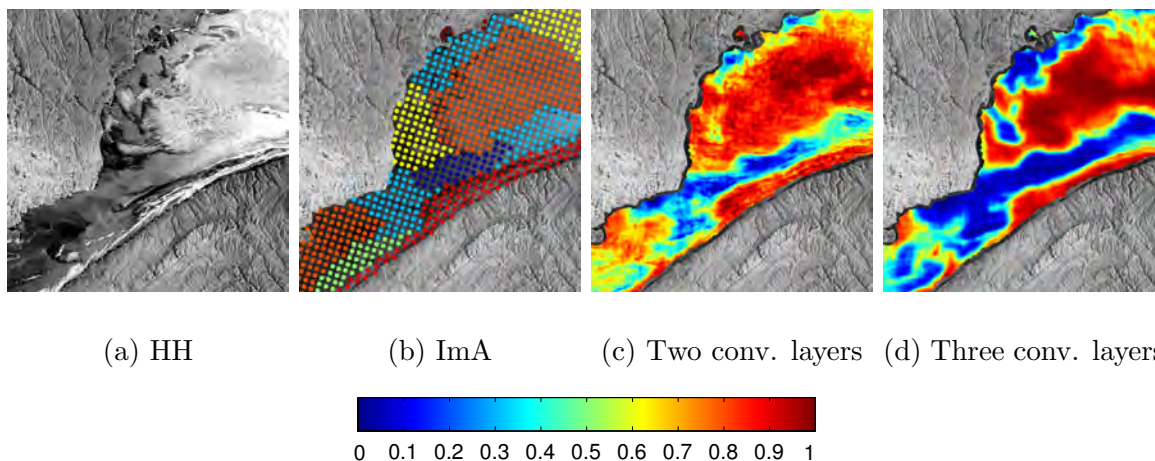
Figure 3.16: Comparison of results produced by networks with two-convolutional-layer and three-convolutional-layer structures for a sample location of size 200 km × 200 km in scene 20140127_104734 in Gulf of Saint Lawrence. Estimate by the two-convolutional-layer network is noisier. The three-convolutional-layer network produces smoother and more reasonable results.

The results are similar when more convolutional layers are used. As adding more layers would lead to increased computational complexity, the three-convolutional-layer structure is chosen as the base model structure.

3.7   Visualization of CNN Features

An investigation of the connection between CNN filters and sea surface conditions is conducted by visualizing the activation maps of the first and last convolutional layers (nonlinear transformed h in (3.1)).

Fig. 3.18 shows the visualization of CNN activations (3.1) for two types of ice structures in two rows. The two ice structures (a) and (d) are labeled with ice concentration 1 and

(a) HH	(b) ImA	(c) Two conv. layer	(d) Three conv. layer

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
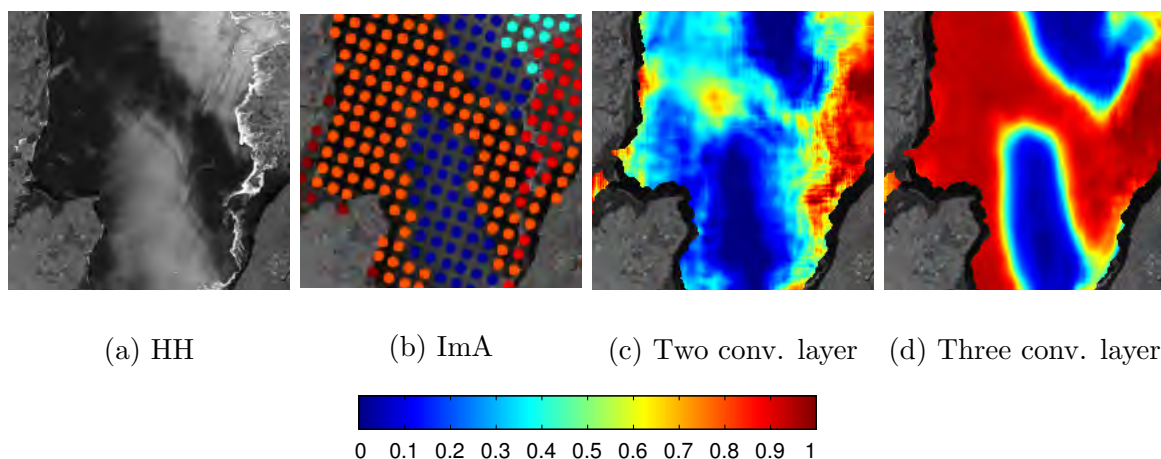
Figure 3.17: The network with three convolutional layers improves the estimation for new ice compared to network with two convolutional layers. The color bar in the bottom is the color scheme used for (b) to (d). Subscene of dimension 8 km × 8 km from 20140117_103914.

0.2 separately in image analysis. For Fig. 3.18(a), the activations in the first convolutional layer are reacting to the brightness changes of SAR images. Similar convolutional filters of the first convolutional layer are activated for ice in both patches (Fig. 3.18(b, e)) for network input (a) and (d). The activated filters for the last convolutional layer are less similar in Fig. 3.18(e) and (f). Fig. 3.18(d) contains ice (white) on water body (gray). Ice and water have different activations for most filters in the first convolutional layer in Fig. 3.18(e). This suggests that CNN is generating features to distinguish ice and water internally.

Water at different incidence angles also has different activation patterns, which is demonstrated in Fig. 3.19. The first and second row of Fig. 3.19 are the dual-pol SAR image and their activations of the first and last convolutional layers for SAR water at incidence angle 30° and 45° (they look less diverse due to the visual enhancement). Different filters are activated for water at different incidence angles. The filters that are activated in common in both cases also exhibit different textures in the activation map. Thus, different filters are learned to estimate water at different incidence angles in CNNs.

Calm water and wind roughened water active some filters in common in the first and third convolutional layer, as demonstrated in Fig. 3.20. The activation patterns of the same filter for calm water and wind roughened water are quite different. In general, wind roughened water has rougher textures in the activation maps of the first convolutional

layer in Fig. 3.20(b) than calm water does in Fig. 3.20(e). The activation textures of the first convolutional layer for wind roughened water regions (Fig. 3.20(b)) is similar to the activation textures of ice in Fig. 3.18(b). The similarity in activations to some types of ice might have caused the overestimation for wind roughened water regions. However, the overestimation for both wind roughened and calm water in Fig. 3.20(a) and Fig. 3.20(d) is less than 5%, because of the difference in activated filters for ice and water.

3.8   Discussion

In this chapter, CNN and FCNN are applied to estimate ice concentration from SAR images in the Gulf of St. Lawrence. A convolutional neural network structure is developed through the sensitivity analysis of hyper-parameters. Experiments using HV pol or HH pol only have also been carried out (results not shown here). Using dual-pol SAR image is able to produce a better ice concentration estimate than using HH pol or HV pol only. When using HH pol only, the results are strongly affected by the incidence angle effect, which causes overestimations for water regions of low incidence angles. Using only HV pol shows banding in the estimated ice concentration. Similar results have been demonstrated in previous research [44, 92].

Two convolutional neural network models, CNN and FCNN, are tuned and applied to dual polarized SAR (HH and HV) images. State-of-the-art ice concentration estimates with finer details than the image analysis chart are acquired. Both CNN and FCNN are demonstrated to generate ice concentration estimates with improved details and accuracy as compared to the currently available passive microwave ice concentration products and our implementation of a recent SAR ice concentration estimation method MLP40 [44], which uses an MLP to regress ice concentration from a set of predefined SAR image features. Because of the shallow network structure, MLP40 is more sensitive to the SAR image backscatter values than the CNN and FCNN used, which causes noisy ice concentration estimates. The small model used by MLP40 does not have the large learning capacity as the used CNN/FCNN model. Some complex cases, such as dark new ice, are not recognized correctly, which will cause systematic errors in the results. Systematic errors cannot be corrected by segmentation based post-processing. Therefore, the deeper and larger CNNs used here can generate more accurate ice concentration estimates than MLP40.

A linear activation function has been chosen as the activation function of the last fully connected layer (convolutional layer in the case of FCNN). Sigmoid activation is a more intuitive choice here as it naturally bounds the output of the CNN or FCNN to 0 and 1. In our experiment, sigmoid activation is found to produce saturated ice concentration

predictions close to 0 or 1, which causes large errors for intermediate ice concentration levels.

The results of CNN and FCNN also contain more details of the ice cover than the image analysis charts, which are used to train the neural networks. However, the ice-water boundaries in the ice concentration estimate by CNN and FCNN are not as crisp as can be seen in the SAR images. Both CNN and FCNN are robust to the incidence angle effect even without the use of incidence angle data. However, the use of incidence angle data as input to the neural networks improves water estimation especially at ice-water boundaries. In general, FCNN produces more visually appealing results than CNN due to improved water predictions and reduced effect of banding in HV pol. Still, predictions for water regions contain errors. Both blurry boundaries and noisy water estimates may be caused by the errors in the image analysis charts, which will be addressed in Chapter 4.

Ice with low reflectance and smooth textures may be misidentified as low ice concentration due to the confusion of SAR image appearances of ice and water. Image structure information is required to solve this problem. Model with larger patch size and larger training samples may also benefit the detection of this kind of ice.

(a)

(b)

(c)

(d)

(e)

(f)
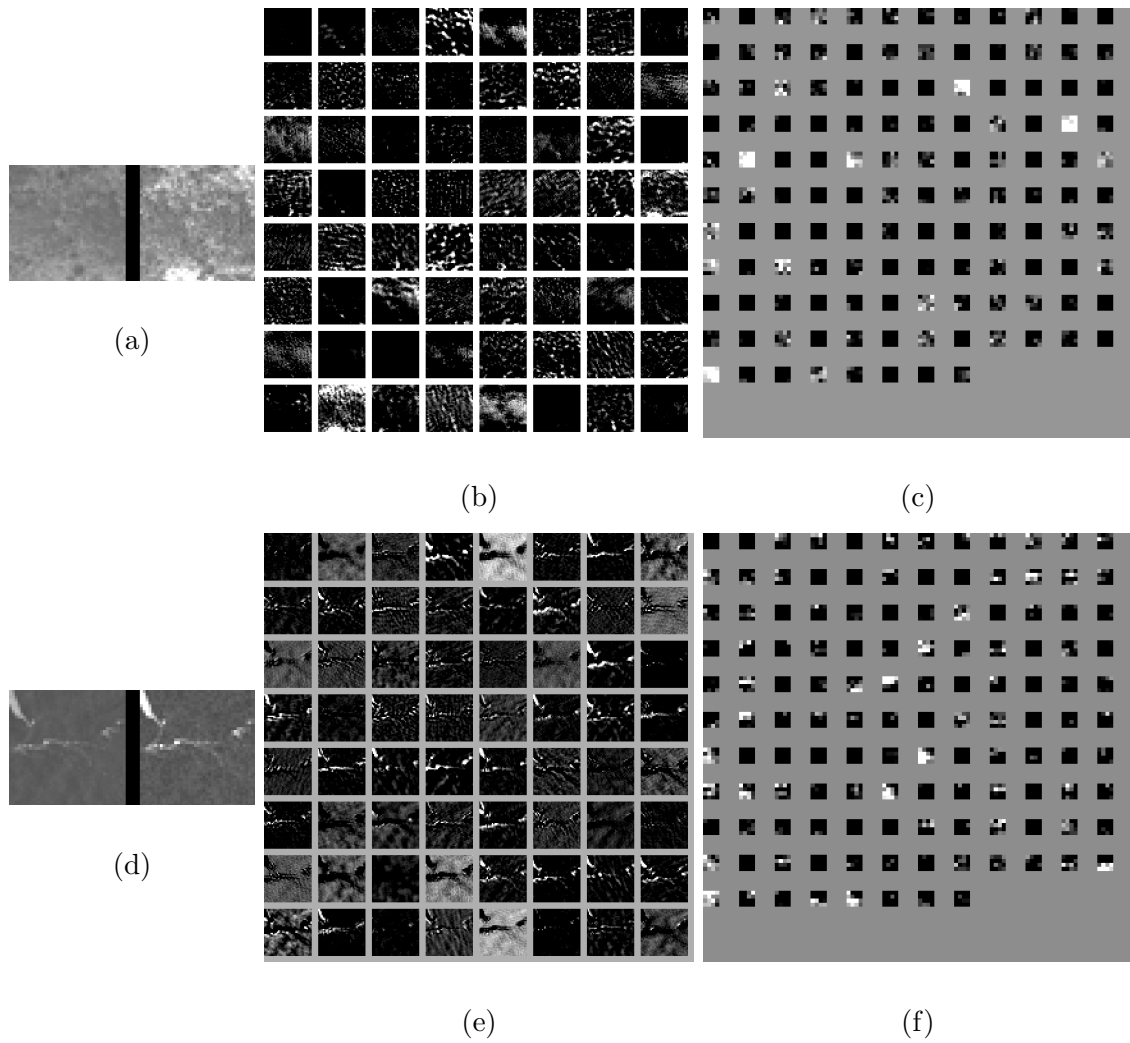
Figure 3.18: CNN identifies SAR image structures. Two rows correspond to two ice samples (ice concentration 1 and 0.2). Column one: HH (left) and HV (right) image patches of size 45 by 45; column two: activations (output) of the first convolutional layer; column three: activations (output) of the last convolutional layer. Images are enhanced for better viewing.

56

(a)

(b)

(c)

(d)

(e)

(f)
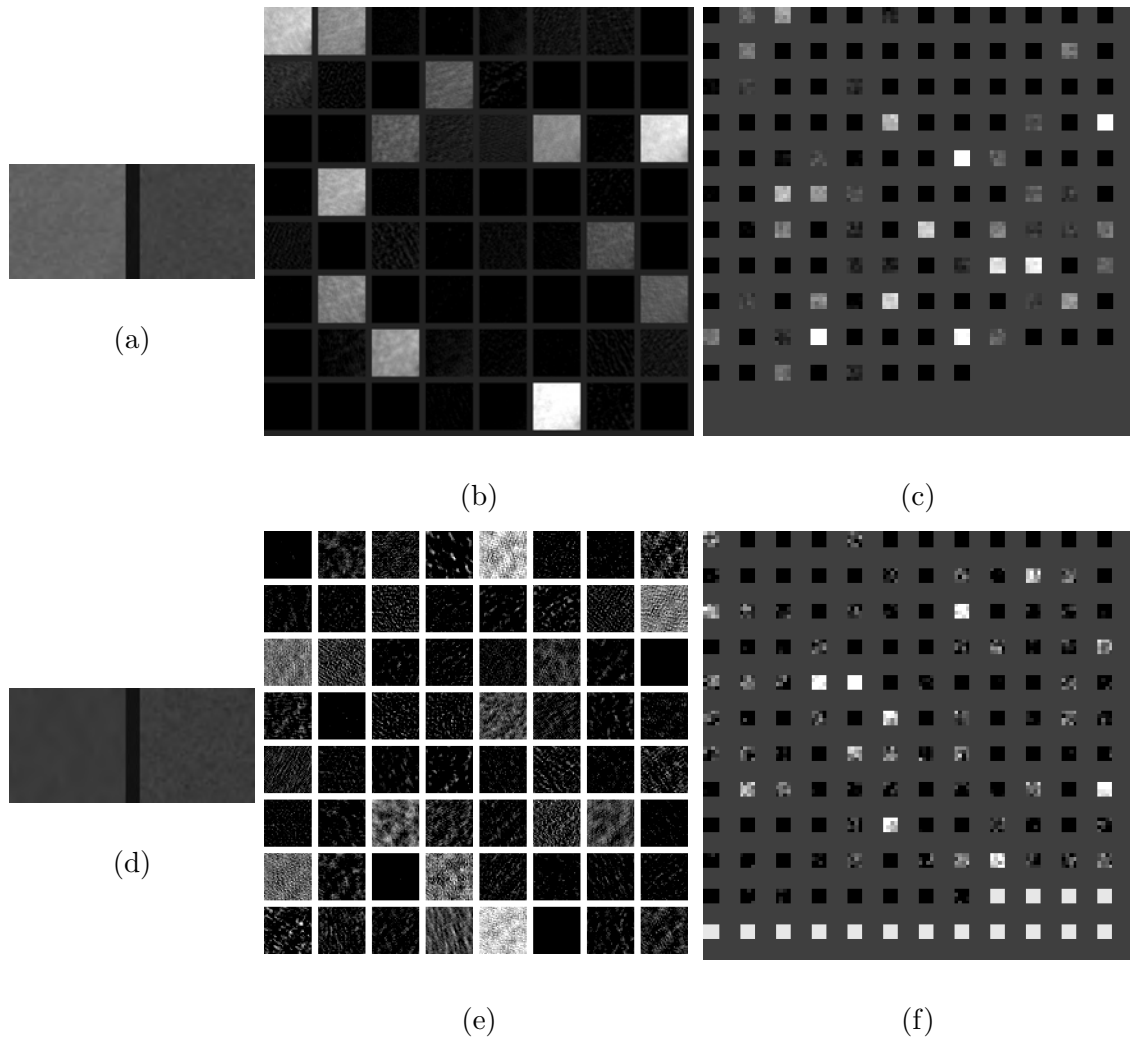
Figure 3.19: CNN features for water of different incidence angles. Row one: incidence angle of 30°; row two: incidence angle 45°. Column one: HH and HV image patches; column two: activation of first convolutional layer; column three: activation of the last convolutional layer. Images are enhanced for better viewing.

(a)

(b)

(c)

(d)

(e)

(f)
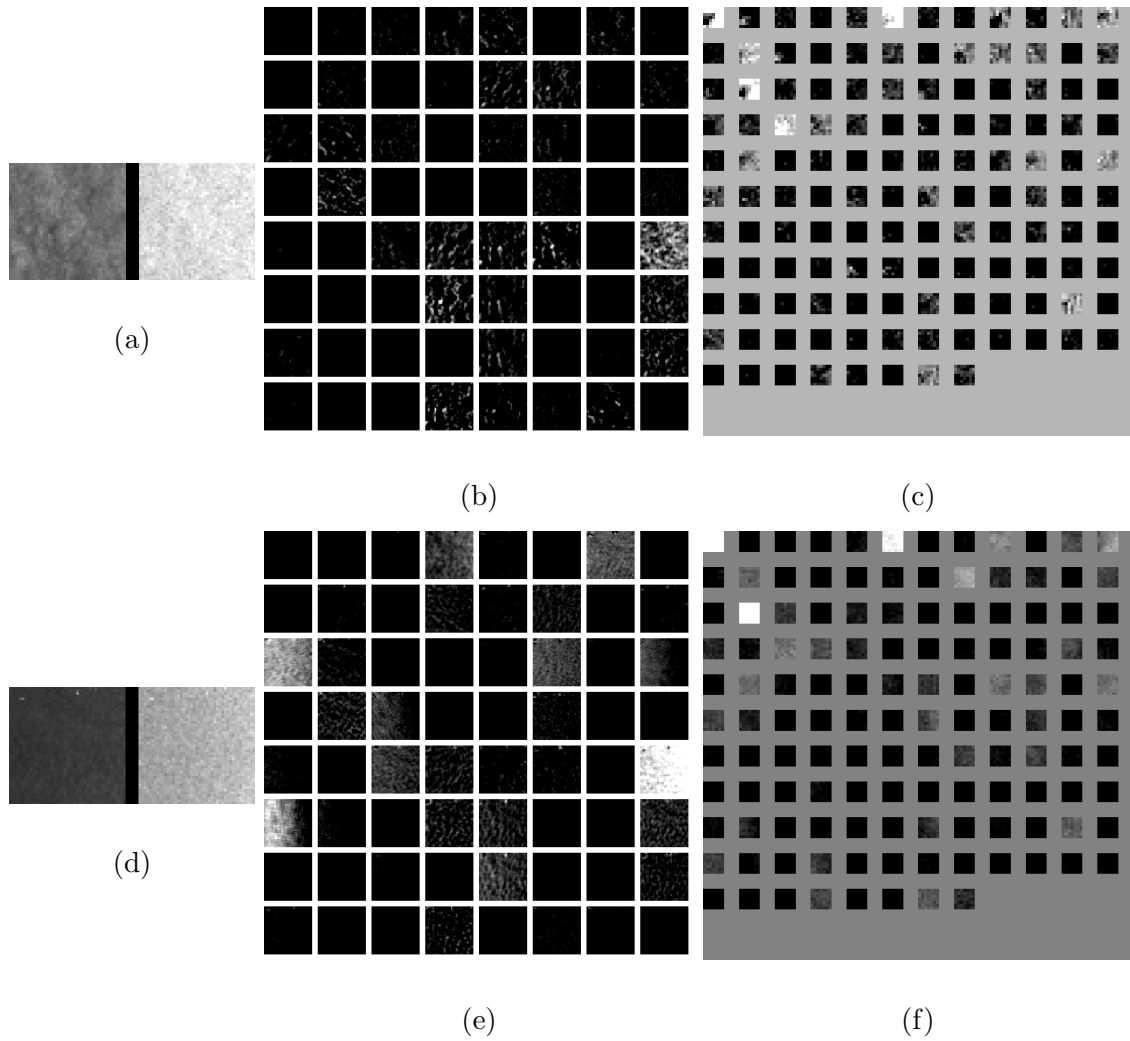
Figure 3.20: CNN features for calm water and wind roughened water. Row one: wind roughened water; Row two: calm water. Column one: HH and HV image patches; column two: activation of first convolutional layer; column three: activation of the last convolutional layer. Images are enhanced for better viewing.

Chapter 4

Accounting for Training Sample Errors

There are two types of errors identified in the image analysis charts: human error and representation error. Human error refers to the uncertainty in manual interpretation, such as inaccurate user-provided ice concentration estimates and misalignment of manually drawn polygons to the natural boundaries. Representation error refers to the spatial variation of ice concentration within a polygon that is not captured by image analysis charts. For example, as an extreme case, a polygon may be identified as 50% concentration, but half of the polygon's area could be 100% ice and half could be open water.

The effect of each of the two types of training sample errors are evaluated in this chapter. L_1 loss is used instead of L_2 loss to reduce the effect of human errors in the training samples. An EM (expectation maximization) based approach and mean-split loss are developed to deal with the representation errors, which lead to results with improved detail and accuracy, especially in water regions.

4.1 The Effect of Training Sample Errors

4.1.1 Human Error

Examples of human errors are shown in Fig. 4.1, where visually homogeneous regions are labeled with different ice concentrations and the image analysis ice concentration does not align well with the ice-water boundaries in SAR images. CNNs trained with these errors are
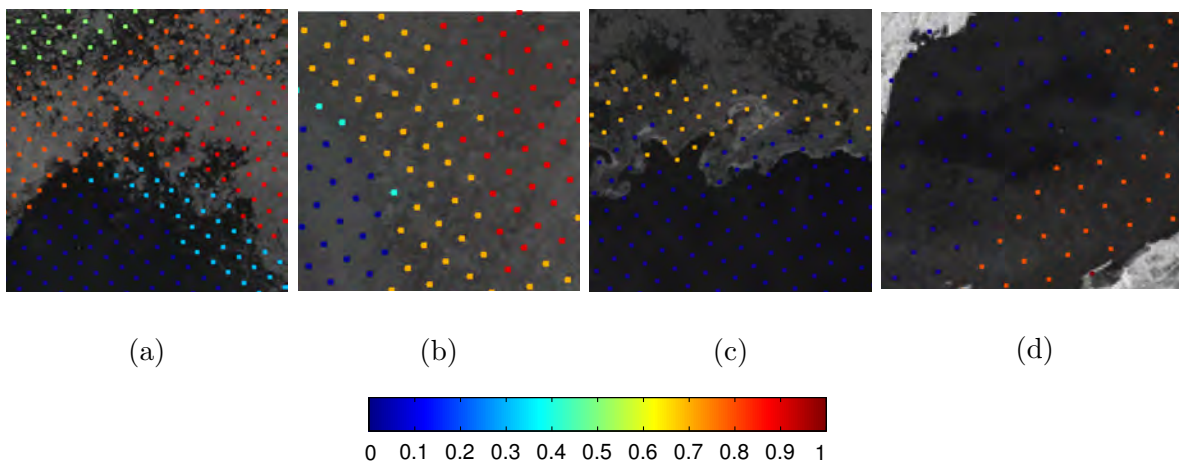
Figure 4.1: Example of random errors in image analysis. (a) to (d) are HH pol SAR images overlaid with image analysis charts which use the color scheme in the bottom. (a) Ice in the middle left of the image is labeled with ice concentration 0, and two regions with similar ice features are identified with ice concentration 0.7 (orange) and 0.9 (red) separately; (b) Similar ice features are identified with ice concentration 0.6 (yellow) and 0.9 (red); (c) Misalignment of ice-water boundaries in image analysis and actual ice-water boundaries; (d) Water region labeled with ice concentration 0.9 (red).

expected to have larger uncertainty around ice-boundaries, or erroneous ice concentration estimates.

The human error is modeled as a Gaussian noise added to the true ice concentration. The effect of human errors in the training samples is evaluated by training and validating CNNs with noise-added training data and validation data. The ice concentration (range from 0 to 1) of every image analysis sample has random, zero mean, Gaussian noise (with predefined parameter standard deviation) added to it. The CNN model used in Chapter 3 is trained with the noise-added samples using the same training scheme as described in Section 3.3.4. The noise added ice concentration values, which are larger than 1 or smaller than 0, are clipped to 1 and 0. In this experiment, random noise with a standard deviation of 0 (no noise), 0.02, 0.05, 0.1, 0.15 and 0.2 are tested. As ice charts are both quantized, and typically, biased (ice charts tend to overestimate ice concentration to reduce the risk for operational uses of the ice charts), the using of zero mean Gaussian noise model to represent the training sample errors is not ideal. The experiments carried out here can be considered as the first step toward evaluating the impact of human errors.

The E_{rmse} of the ice concentration in the test dataset, given different noise standard

deviations, are shown in Fig. 4.2. The E_{rmse} increases quickly by a factor of two times when the standard deviation of noise increases from 0 to 0.02. The increase of E_{rmse} slows down when the standard deviation of noise is larger than 0.02. Therefore, the performance of CNN trained with L_2 loss is sensitive to human errors in the training samples especially when the errors are small. The accuracy (in terms of standard deviation) of ice concentration in ice charts is considered to be around 0.1 in practice [10]. However, due to the uncertainty in manual interpretation, the level of human errors in image analysis charts may be affected by the skills of the ice expert, season and ice form. The performance of neural networks trained with image analysis charts using L_2 loss will depend on these factors. Therefore, it is necessary to take into account the human errors for more consistent and stable performance. This will be discussed further in Section 4.2.
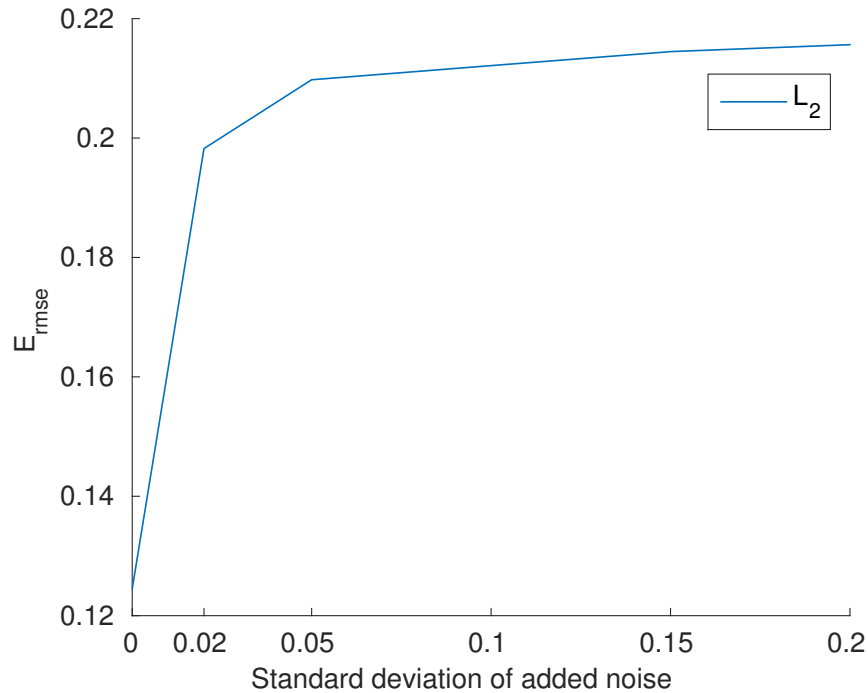
Figure 4.2: Effect of adding random error to the training samples for CNNs trained with L_2 loss.

4.1.2 Representation Error

The impact of representation error is examined by comparing the results from CNNs trained with and without the representation errors. In this experiment, a few scenes of the Gulf of Saint Lawrence dataset with obvious representation errors are selected. A CNN model is trained and tested on all the selected scenes. As the training and testing are using identical and small dataset, the test result should be almost perfect in this setting unless the training samples contain errors. The data used in this experiment are illustrated in Fig. 4.3. Fig. 4.3(a, b, c) are the dual-pol SAR image and image analysis of a region for which the image analysis contains representation errors. This region shows sparse ice cover surrounded by open water. The sparse ice cover is assigned an ice concentration 0.6 in the image analysis (orange in Fig. 4.3(c)), while it appears that the ice concentration should be either 0 or 1 on a pixel level from the SAR imagery. The CNN model trained with this part of the image analysis produces erroneous estimates for some water regions, as can be seen in Fig. 4.3(f). This is because pixels of ice concentration near 0 and 1 in the SAR image are assigned with ice concentration 0.6 in the training process. The errors in water in the ice concentration estimate are removed when the image analyses with representation errors are excluded from the training process (orange in Fig. 4.3(g)). Therefore, representation errors can degrade the performance of the CNN model, leading to over prediction of ice concentration in water and low ice concentration regions.

4.2 Accounting for Human Error Using L_1 Loss

The least squared error (L_2 norm) is known to be sensitive to data noise [96], which is also demonstrated in Fig. 4.2. In using a mean-squared difference, higher weight is given to larger residuals in the L_2 norm, which makes the regression model easily biased by outliers. The least absolute deviation (L_1 norm) is more robust to outliers as all residuals contribute the same to the derivative of loss to model parameters. Therefore, the use of L_1 loss can also partly reduce the effect of representation error.

While L_2 loss (3.4) calculates the mean-squared difference, the L_1 loss calculates the absolute mean difference between the neural network output ($F(\boldsymbol{x};\boldsymbol{\theta})$) and the ice concentration from the image analysis ($\boldsymbol{z}$):

$$\mathbb{L}^{L_1}(F(\boldsymbol{x};\boldsymbol{\theta}),\boldsymbol{z}) = \frac{1}{M}\sum_{m=1}^{M}|F(\boldsymbol{x};\boldsymbol{\theta})_m - \boldsymbol{z}_m|, \tag{4.1}$$

(a) HH   (b) HV   (c) ImA

(d) Target HH   (e) Target HV   (f) w/ error   (g) w/o error
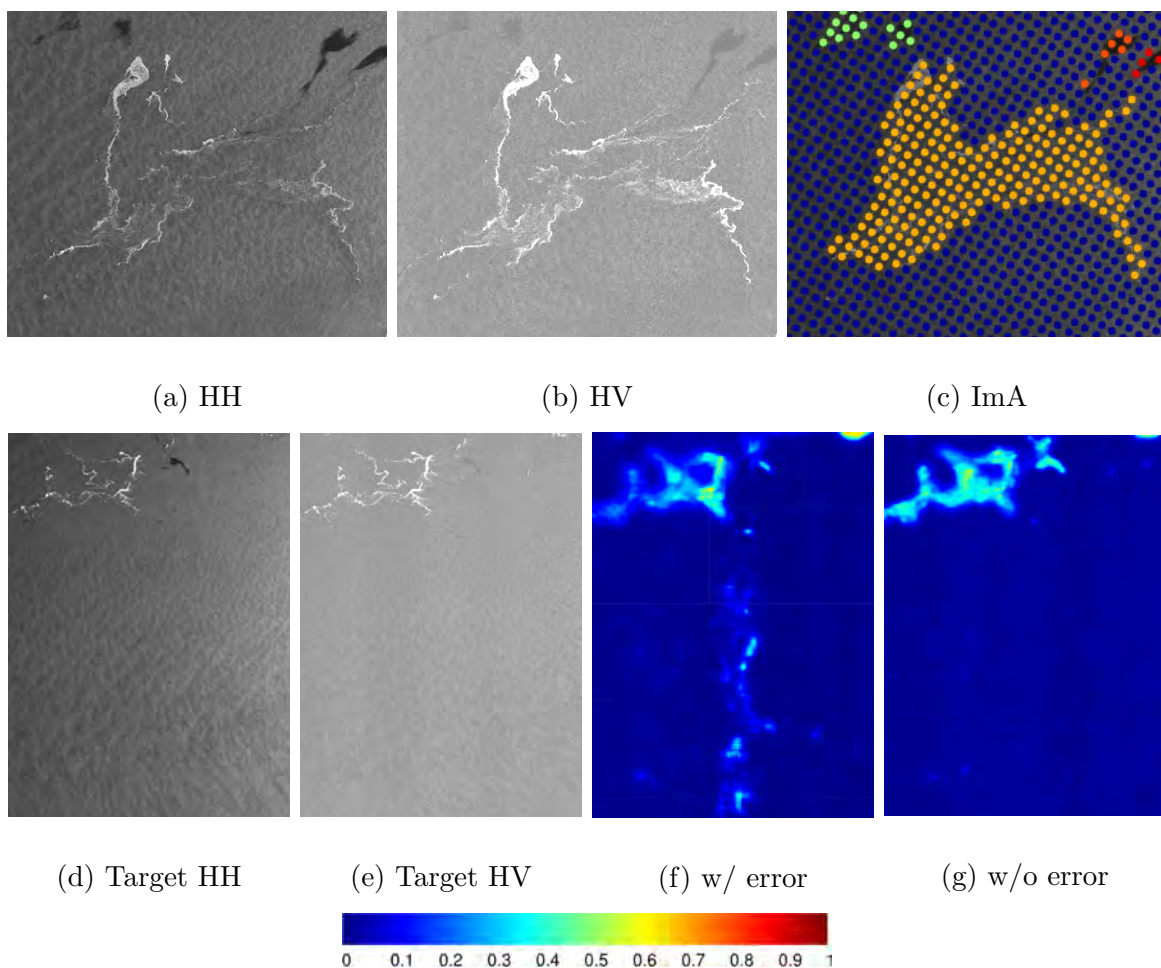
Figure 4.3: Representation errors in image analysis causes errors in water regions. (a) and (b): region of sparse ice cover identified as 60% ice concentration in the image analysis (c), while the actual ice concentration in a finer scale is mostly water and ice, which leads to large representation errors. (d) HH pol and (e) HV pol with mostly open water and some ice floes at the top. (f) Noisy water prediction (0.2-0.3 in the middle of the region) of the target region when the all samples in (c) are used for training. (g) Clean water prediction of the target region when the erroneous training samples in (c) are omitted from the training process. Subplots (a), (b), (d) and (e) are visually enhanced for better viewing. (c), (f) and (g) follow the provided legend to show concentration across the whole image.

where $\boldsymbol{x}$ and $\boldsymbol{\theta}$ are the input patch matrix and weights of the CNN/FCNN respectively. Notation M is the number of pixels in the patch $F(\boldsymbol{x}; \boldsymbol{\theta})$, and m is the index of elements

63

in a matrix. The derivative when the argument of the absolute value function is zero is not defined. For the purpose of back-propagation, the derivative of L_1 is set to 0 when $F(\boldsymbol{x}; \boldsymbol{\theta})_m = \boldsymbol{z}_m$. That is

$$\frac{\partial \mathbb{L}^{L_1}}{\partial F(\boldsymbol{x}; \boldsymbol{\theta})_m} = \begin{cases} -1/M & \text{if } F(\boldsymbol{x}; \boldsymbol{\theta})_m < \boldsymbol{z}_m, \\ 0 & \text{if } F(\boldsymbol{x}; \boldsymbol{\theta})_m = \boldsymbol{z}_m, \\ 1/M & \text{if } F(\boldsymbol{x}; \boldsymbol{\theta})_m > \boldsymbol{z}_m. \end{cases} \tag{4.2}$$

For batch size larger than 1, the calculated loss is the mean loss of all the patches in that batch.

4.2.1 Experiment Setup

The FCNN model described in Section 3.3.3 is used here. The same training, validation and testing dataset and training scheme as in Section 3.3.4 are used, except the learning rate. Because the derivative of loss to the neural network output is larger for L_1 loss compared to L_2 loss, which results in larger updates for network parameters during training, the learning rate is reduced to 5×10^{-4} for L_1 loss. The resulting models are noted by FCNN-L_1 and FCNN-L_2 separately.

4.2.2 Results

The average error statistics are shown in Table 4.1. The E_{rmse} for FCNN-L_1 and FCNN-L_2 are similar for all sets. In terms of these statistics, the two methods generate similar results.

Table 4.1: The error statistics for networks trained using L_1 and L_2 loss on the Gulf of Saint Lawrence dataset

	Set	E_{sgn}	E_{L1}	E_{std}	E_{rmse}
	Training	0.0146	0.0810	0.1559	0.1565
FCNN-L_2	Validation	-0.0055	0.1239	0.2141	**0.2142**
	Testing	-0.0082	0.1257	0.2011	**0.2012**
	Training	-0.0027	0.0656	0.1406	**0.1407**
FCNN-L_1	Validation	-0.0114	0.1206	0.2209	0.2211
	Testing	0.0031	0.1291	0.2234	0.2234

Visually, the results of FCNN-L_1 have sharper boundaries and improved water predictions as illustrated in Fig. 4.4. Water regions near ice-water boundaries are cleaner when L_1 is used. These improvements in visual appearance do not always mean better consistency with the image analyses. Because larger errors have a higher penalty in L_2 loss than in L_1 loss, the results of FCNN-L_2 are closer to the image analysis in this case than the results of FCNN-L_1 as shown in Table 4.1.

(a) HH   (b) ImA   (c) FCNN-L_2 (d) FCNN-L_1

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
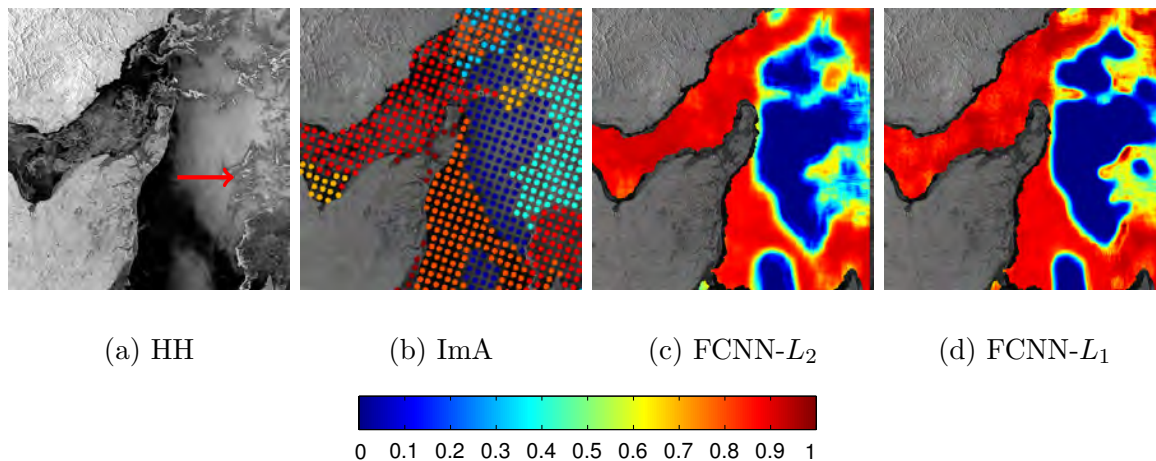
Figure 4.4: FCNN trained with L_1 loss has improved ability to capture ice boundaries as compared to FCNN trained with L_2 loss. The image analysis does not provide the necessary details to resolve ice concentration in detailed regions and an example of this is provided in the region identified by the arrow in (a). Subscene of dimension 10 km × 10 km from 20140117_103914 in testing set.

The overestimation of water caused by the banding effect in the HV pol is also improved when L_1 loss is used. An example is shown in Fig. 4.5, where most of the errors in water predictions are removed when L_1 loss is used (Fig. 4.5(e)).

To test the effect of human errors of training samples on the performance of CNNs trained with L_1 loss, different levels of random noise are added to the image analysis ice concentration to simulate the overall effect of human errors, i.e., a combined effect of random errors and gross errors). The clipping of noise added image analysis to 0 and 1 can break the symmetry of the distribution of the added noise, which can partly simulate the gross errors of human errors. The performance of CNNs trained with these noise-added training samples are tested, as done in the beginning of this chapter (Section 4.1.1). The E_{rmse} for the test dataset with respect to the standard deviation of the added Gaussian noise is illustrated in Fig. 4.6. CNNs trained with L_1 loss are relatively stable to human

(a) HV	(b) ImA	(c) FCNN-L_2	(d) FCNN-L_1
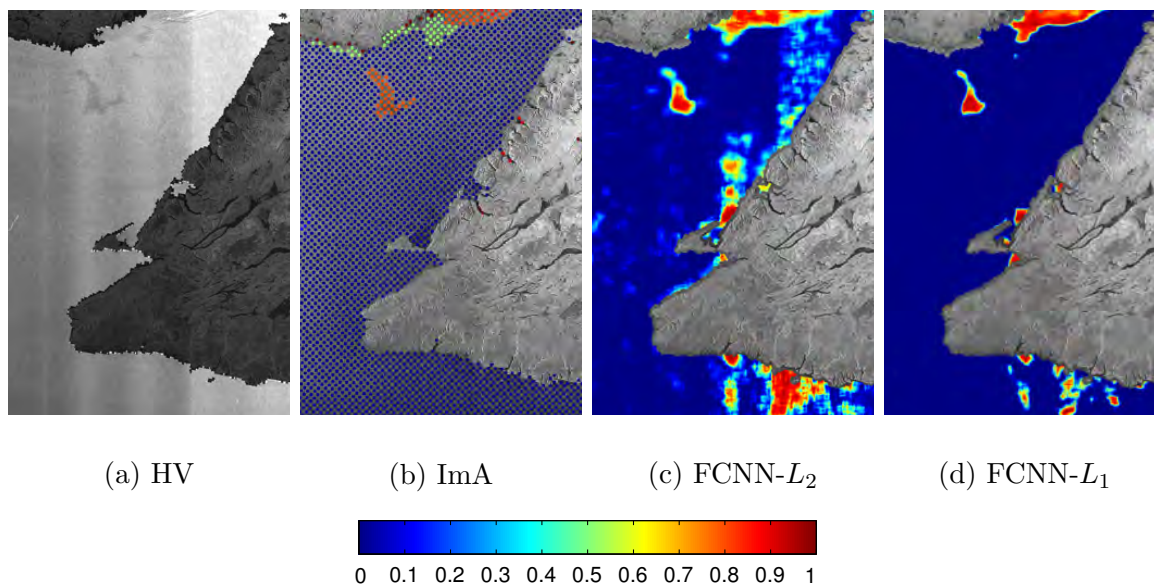
Figure 4.5: An example showing improved open water estimates of FCNN-L1 compared to FCNN-L2. (a) HV pol contaminated by banding effect (visually enhanced); (b) corresponding image analysis; (c) the result of FCNN-L_2 with erroneous water estimates that coincide with the location of banding; (d) the result of FCNN-L_1 that has more accurate open water estimates. Subscene of dimension 900×600 from 20140121_214420.

errors in the training samples as compared to CNNs trained with L_2 loss. The results of CNNs trained with added random noise of standard deviation 0.2 using L_1 and L_2 are compared in Fig. 4.7. Using L_2 heavily overestimates ice concentration for water regions when the added noise has standard deviation is 0.2 (Fig. 4.7(e)), while using L_1 can produce similar water predictions for both the case of no noise, and noise with a standard deviation of 0.2.

4.3 Accounting for Representation Error

The representation error is caused by using the image analysis samples as pixel level ice concentration instead of the average ice concentration of a polygon. To reduce the effect of representation error, the image analysis ice concentration should be used as the mean ice concentration in a polygon region. However, due to the lack of the polygon shape information in our image analysis dataset, a rectangle of predefined size instead of the
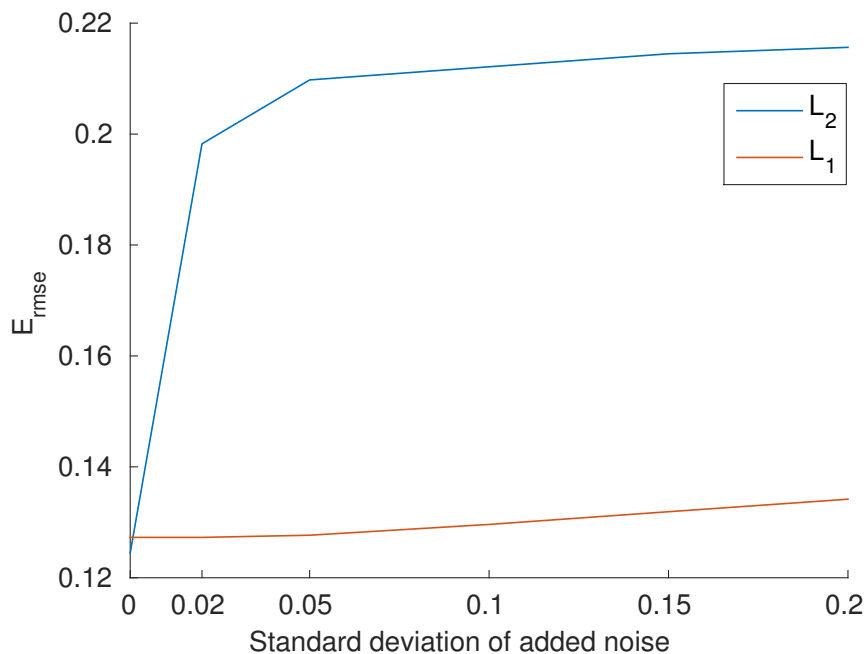
Figure 4.6: Effect of adding random error to the training samples for CNNs trained with L_1 and L_2 loss.

image analysis polygon is used to calculate the mean ice concentration, i.e., the mean ice concentration from image analysis over a rectangle of predefined size is considered to be an estimate of the mean of the ice concentration in that rectangle. For convenience, the size of the rectangle is defined to be the input patch size of the neural network. In this section, two methods are developed to use image analysis under this semantic meaning: an EM algorithm which considers the pixel level ice concentration as a latent variable, and a new loss function which calculates the loss from the difference of the mean of image analysis and the mean of the network output. We first introduce these two methods. Then the testing and comparison of the two methods are followed.

4.3.1 Expectation Maximization

An expectation-maximization (EM) algorithm is developed to train FCNN with the representation errors of the image analyses modeled. EM algorithm is an iterative method to find the maximum likelihood estimation of parameters in statistical models with latent variables [22, 9]. When the representation errors of image analyses are considered, image

67
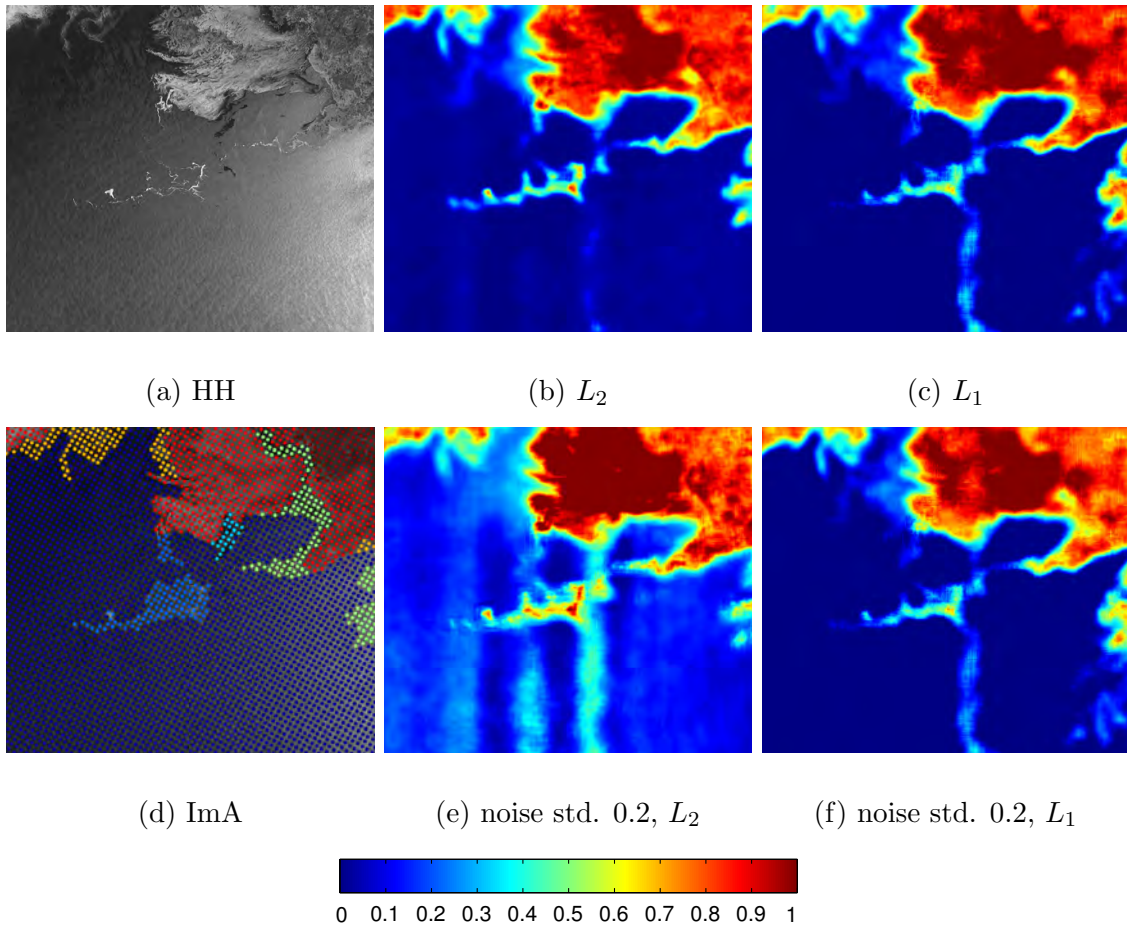
Figure 4.7: Comparing the results of CNNs trained with L_1 and L_2 loss with random training samples errors. L_1 loss is more robust to random training sample errors as compared to L_2 loss.

analyses are interpreted as manually estimated mean of pixel level ice concentration values for given regions. The ice concentration at each pixel is the unknown latent variable in this system. The CNN model parameters and pixel level ice concentration depend on each other. So EM algorithm is a proper algorithm to train this system.

An EM algorithm iteratively runs E step and M step. In the E step, the maximum a posterior (MAP) estimation of the latent variable is calculated. In the M step, parameters of the statistical model are adjusted to maximize the data-likelihood using the estimated latent variable from the last E step. This algorithm is implemented as an EM layer which

takes the image patch and the ground truth label patch as input and outputs the modified label patch.

The formulation of the EM algorithm used in this study is based on Andre Ng's course notes [69] and inspired by Papandreou's work [74]. The input SAR image patch is denoted by $\boldsymbol{x}$. The pixel level ice concentration is denoted by $\boldsymbol{y}$. The ice concentration from the image analysis is denoted by $\boldsymbol{z}$. Here the image analysis $\boldsymbol{z}$ is interpreted as the mean of pixel level ice concentration $\boldsymbol{y}$. So $\boldsymbol{z}$ is conditionally independent from the input $\boldsymbol{x}$ given $\boldsymbol{y}$. With this assumption, we have the following decomposition of $p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\theta})$:

$$p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\theta}) = p(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta})p(\boldsymbol{z}|\boldsymbol{y}), \tag{4.3}$$

in which $\boldsymbol{\theta}$ is the vector containing neural network weights.

The expected log-likelihood of the data given the previous parameter estimate $\boldsymbol{\theta}'$ is [69]:

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}') = \sum_{\boldsymbol{y}} p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta}') log \frac{p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\theta})}{p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta}')}. \tag{4.4}$$

Decompose $p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\theta})$ using (4.3), we get

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}') = \sum_{\boldsymbol{y}} p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta}')(log[p(\boldsymbol{x})] + log[p(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta})] + log[p(\boldsymbol{z}|\boldsymbol{y})] - log[p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta}')]).$$
$$\tag{4.5}$$

Ignoring the terms that do not depend on $\boldsymbol{\theta}$, Q can be simplified to

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}') = \sum_{\boldsymbol{y}} p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta}') log[p(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta})]. \tag{4.6}$$

In E step, the latent pixel ice concentration is estimated by maximizing $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta}')$, which can be decomposed by (4.3). The optimal $\boldsymbol{y}$ can be written as:

$$\hat{\boldsymbol{y}} = argmax_{\boldsymbol{y}}[p(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta}')p(\boldsymbol{z}|\boldsymbol{y})] \tag{4.7a}$$
$$= argmax_{\boldsymbol{y}}[log[p(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta}')] + log[p(\boldsymbol{z}|\boldsymbol{y})]]. \tag{4.7b}$$

In M step, $\boldsymbol{y}$ is set to the estimated $\hat{\boldsymbol{y}}$ and $\boldsymbol{\theta}$ is adjusted to maximize Q in (4.6). By assigning $\boldsymbol{y}$ to $\hat{\boldsymbol{y}}$, $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta}')$ is approximated to a distribution with probability 1 at $\hat{\boldsymbol{y}}$ and 0 otherwise [74]. So (4.6) becomes

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}') \approx log[p(\hat{\boldsymbol{y}}|\boldsymbol{x}; \boldsymbol{\theta})]. \tag{4.8}$$

Standard back-propagation from the loss calculated with $\hat{\boldsymbol{y}}$ as the ground truth is used to update $\boldsymbol{\theta}$ to maximize Q.

Both $p(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta}')$ and $p(\boldsymbol{z}|\boldsymbol{y})$ need to be specified to complete the E step. We first consider the case when the whole input SAR image patch lies inside a polygon in image analysis. Therefore, $\boldsymbol{z}$ becomes a single real value between 0 and 1, which is denoted by z. Later, the developed method will be extended to the case when input SAR image patches overlap with multiple image analysis polygons.

Assuming that image analysis ice concentration is sampled from a Gaussian distribution centered at the mean of the pixel level ice concentration, which is denoted by $\bar{\boldsymbol{y}}$. The probability of the image analysis ice concentration ($\boldsymbol{z}$) given the pixel level ice concentration ($\boldsymbol{y}$) can be written as

$$p(\boldsymbol{z}|\boldsymbol{y}) \propto exp(-\frac{(\bar{\boldsymbol{y}} - z)^2}{\sigma_z^2}) \tag{4.9a}$$

$$log[p(\boldsymbol{z}|\boldsymbol{y})] = -\frac{(\bar{\boldsymbol{y}} - z)^2}{\sigma_z^2} + (const), \tag{4.9b}$$

where M is the number of pixels in $\boldsymbol{y}$, and σ_z is an unknown parameter.

Due to the bias and quantization of ice concentration values in image analyses, the probability of the image analysis ice concentration given the pixel level ice concentration $p(\boldsymbol{z}|\boldsymbol{y})$ is asymmetric and discrete. However, the amount of bias is unknown, and may vary for different ice conditions. Modeling the actual $p(\boldsymbol{z}|\boldsymbol{y})$ relationship would be difficult. The Gaussian model used here is an approximation of $p(\boldsymbol{z}|\boldsymbol{y})$. It will be shown later that the estimated $\boldsymbol{y}$ does not depend on the shape of $p(\boldsymbol{z}|\boldsymbol{y})$ as long as it peaks when the mean of $\boldsymbol{y}$ is equal to $\boldsymbol{z}$. So the Gaussian model used here for $p(\boldsymbol{z}|\boldsymbol{y})$ is a reasonable approximation, that is demonstrated to work well in practice.

The FCNN output is an estimate of $\boldsymbol{y}$ using the information in SAR image $\boldsymbol{x}$. Since the FCNN does not output the conditional distribution $p(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta}')$ directly, it needs to be assigned. We assume that the shape of the ice concentration distribution in a patch can be correctly predicted by the FCNN, meaning, for example, that pixels with a higher ice concentration have a higher predicted ice concentration by the FCNN. However, the value range (difference between maximum and minimum values in a patch) of the FCNN output $F(\boldsymbol{x}; \boldsymbol{\theta})$ should be smaller than the true value range of $\boldsymbol{y}$, because the training samples used have a smaller value range than their true value range. For example, given a region labeled with ice concentration 0.5, the actual ice concentration might be 0 and 1 for some of the pixel locations. Neural network models trained with samples labeled with ice concentration 0.5 in this polygon are expected to generate ice concentration predictions around 0.5 with

value range smaller than [0, 1]. This is generally consistent with our observation of the FCNN results. That is to say, the relative ice concentration between different pixels in the FCNN output $F(\boldsymbol{x}; \boldsymbol{\theta})$ is considered to be an approximation of the actual relative ice concentration.

A linear model with Gaussian noise is used to model the relationship between the value range of the predicted ice concentration and the actual value range, i.e., $\boldsymbol{y}$ is estimated by shifting $F(\boldsymbol{x}; \boldsymbol{\theta})$ and rescaling from its mean:

$$\boldsymbol{y} = \bar{\boldsymbol{y}} + [F(\boldsymbol{x}; \boldsymbol{\theta}) - \overline{F(\boldsymbol{x}; \boldsymbol{\theta})}](1 + \rho) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma_y^2) \tag{4.10a}$$

$$log[p(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta}')] = -\sum_{m=1}^{M} \frac{((y_m - \bar{\boldsymbol{y}}) - [(F(\boldsymbol{x}; \boldsymbol{\theta})_m - \overline{F(\boldsymbol{x}; \boldsymbol{\theta})})(1 + \rho)])^2}{\sigma_y^2} + (const). \tag{4.10b}$$

In which, the $F(\boldsymbol{x}; \boldsymbol{\theta})$ is first shifted to zero mean by subtracting the mean over a patch, which is denoted by $\overline{F(\boldsymbol{x}; \boldsymbol{\theta})}$, and then scaled to have larger value range by a factor of $(1 + \rho)$. ε is a random error term with zero mean and standard deviation σ_y. The scaled values are shifted to the mean of the pixel level ice concentration $\bar{\boldsymbol{y}}$. By using (4.10), the pixels in $\boldsymbol{y}$ are considered independent given $\boldsymbol{x}$.

With $p(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta}')$ and $p(\boldsymbol{z}|\boldsymbol{y})$ specified in (4.9) and (4.10), the solution for $\hat{\boldsymbol{y}}$ in (4.7) is,

$$\hat{\boldsymbol{y}} = z + (F(\boldsymbol{x}; \boldsymbol{\theta}) - \overline{F(\boldsymbol{x}; \boldsymbol{\theta})})(1 + \rho), \tag{4.11}$$

where both $p(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta}')$ and $p(\boldsymbol{z}|\boldsymbol{y})$ can reach their own maximum. So $\hat{\boldsymbol{y}}$ is not affected by the value of σ_z and σ_y. These two values do not need to be specified.

The resulting pixel level ice concentration estimate, $\hat{\boldsymbol{y}}$, retains the relative relationship between FCNN output pixels and the mean ice concentration estimate over the patch is $\boldsymbol{z}$. The parameter ρ needs to satisfy several constraints:

1. Parameter ρ must be non-negative to make sure the value range of $\hat{\boldsymbol{y}}$ is not decreased after scaling because less diverse training samples ($\boldsymbol{z}$ is smoother than the actual situation) are used to generate the $\boldsymbol{y}$, which will result $F(\boldsymbol{x}; \boldsymbol{\theta})$ with smaller range value than the real situation.

2. Parameter ρ should be small enough to make sure the scaled values still lie between 0 and 1.

3. An image patch with a more diverse appearance and intermediate ice concentration levels in the image analysis charts often should have larger ρ due to a larger possibility

of diverse ice concentration levels within the patch. This is a major assumption used in this EM algorithm. The heuristics is that training samples of ice concentration 0 or 1 generally contain less representation errors because they are all ice or water, while training samples with intermediate ice concentration values are more likely to be a mixture of ice pixels and water pixels when they contain pixels of very different appearances within their SAR image patches.

Based on these constraints, the parameter ρ is calculated by

$$\boldsymbol{F}' = F(\boldsymbol{x}; \boldsymbol{\theta}) - \overline{F(\boldsymbol{x}; \boldsymbol{\theta})} + z \tag{4.12a}$$

$$\rho_{max} = min\left(\frac{z}{z - min_m \boldsymbol{F}'_m}, \frac{1 - z}{max_m \boldsymbol{F}'_m - z}\right) \tag{4.12b}$$

$$\rho = min(\alpha/\sigma_{\boldsymbol{x}}, \rho_{max}), \alpha > 0, \sigma_{\boldsymbol{x}} > 0, \tag{4.12c}$$

where $\boldsymbol{F}'$ is the FCNN prediction shifted to mean z in (4.12a). (4.12b) calculates the maximum possible ρ by constraining the scaled $\boldsymbol{F}'$ to interval $[0, 1]$. Notation m is the index of the elements in $\boldsymbol{F}'$. Notation $\sigma_{\boldsymbol{x}}$ and α are a measure of the homogeneity of $\boldsymbol{x}$ and a preset coefficient that maps the homogeneity measurement to ρ. ρ is then selected from the smaller value between $\alpha\sigma_{\boldsymbol{x}}$ and ρ_{max} in (4.12c). Any homogeneity measurement can be used as $\sigma_{\boldsymbol{x}}$. Difference homogeneity measurements can be found in the literature [89, 90, 40]. Here a simple standard deviation is used:

$$\sigma_{\boldsymbol{x}} = 1/std(\boldsymbol{x}). \tag{4.13}$$

Less homogeneous SAR images would be scaled to larger ice concentration ranges. Since the scaled ice concentration is constrained to between 0 and 1, ρ would equal to ρ_{max} which is close to 0 when shifted ice concentration estimates ($\boldsymbol{F}'$) are close to 0 or 1. The appearance of the input image patch does not affect the results in such cases. The appearance of the input image will affect the training samples when they are identified as intermediate ice concentration and the backscatters appear to indicate a mixture of ice and water.

For the patches with multiple ice concentration levels in the image analyses, their predictions are grouped according to their image analysis ice concentration, and $\hat{\boldsymbol{y}}$ is calculated for the given patch for every group separately in the E step using the above equations.

4.3.2 Mean-Split Loss

Instead of using the EM algorithm to deal with representation errors in the training samples, another model is developed in which the representation error is modeled directly. As

the image analysis indicates the average ice concentration in a specified region, a direct way to use the image analysis data is to compare the mean of the predicted ice concentration and the mean of the image analysis over a region for loss calculation. This loss function is referred to as mean-split loss in the rest of the thesis.

In order to calculate the mean-split loss, pixels in a patch are grouped according to the ice concentration levels $I = 0, 0.1..., 1$ in the corresponding area in the image analysis. The loss for each group is calculated from the image analysis ice concentration $\boldsymbol{z}$ and the mean of their predicted ice concentration $\bar{\boldsymbol{y}}_i$ using distance function L. The total loss is the weighted sum of the group losses weighted by their portion in the patch $\frac{M_i}{M}$. That is

$$\mathbb{L}^{MS}(\boldsymbol{z}, \boldsymbol{y}) = \sum_{i \in I} \frac{M_i}{M} L(i, \bar{\boldsymbol{y}}_i), \tag{4.14}$$

in which, M_i represents the number of pixels in a patch with image analysis ice concentration i. All M_i present in a patch add up to M which is the total number of values in FCNN output $\boldsymbol{y}$ in a patch. The distance function L is either L_1 or L_2 norm in this study.

The back-propagation for all the pixels in a group are the same. When using L_1 loss for L, for pixel $\boldsymbol{y}_m$ in group i, the back-propagation is

$$\frac{\partial \mathbb{L}^{MS}}{\partial \boldsymbol{y}_m} = \frac{M_i}{M} \frac{\partial L}{\partial \bar{\boldsymbol{y}}_i} \frac{\partial \bar{\boldsymbol{y}}_i}{\partial \boldsymbol{y}_m} \tag{4.15a}$$

$$= \frac{M_i}{M} \frac{\partial L}{\partial \bar{\boldsymbol{y}}_i} \frac{1}{M_i} \tag{4.15b}$$

$$= \frac{1}{M} \frac{\partial L}{\partial \bar{\boldsymbol{y}}_i}. \tag{4.15c}$$

The derivative of L to $\bar{\boldsymbol{y}}_i$ ($\frac{\partial L}{\partial \bar{\boldsymbol{y}}_i}$) can be calculated numerically from (4.2) or (3.5) when L is L_1 or L_2 norm.

The mean-split loss is implemented as a mean-split layer within the framework of neural networks. The mean-split layer takes the CNN/FCNN output and the image analysis as input, and outputs the loss using (4.14). The errors are back-propagated to every pixel in the CNN output separately using (4.15).

4.3.3 Experiments

The FCNN model developed in Chapter 3 is used. The EM layer is inserted after the last FCNN layer and L_1 loss is calculated from the output of EM layer and the image analyses.

This model is referred to as FCNN-EM-L_1 in the rest of the thesis. There is one parameter that needs to be manually set in the proposed EM method, which is the coefficient α for σ_x. The coefficient α is set to 0.5. The same training scheme and parameter setting used to train the FCNN-L_1 model in Chapter 4.2 is used here. Because the EM algorithm uses the prediction of the network to refine the image analyses during training, a good initial prediction would benefit the convergence of the training process. Therefore, the model FCNN-EM-L_1 is fine tuned from a trained FCNN-L_1 model, which shows better performance than one directly trained from randomly initialized model weights.

The same FCNN model is used to test the mean-split loss. L_1 is used for the distance function L in (4.14). This model is referred to as FCNN-MS-L_1. The same training scheme and parameter setting used to train the FCNN-L_1 model in Chapter 4.2 is used here.

4.3.4 Results

The error statistics are given in Table 4.2. All three methods generate ice concentration with small biases (E_{sgn}) when evaluated against image analysis charts. The E_{std} of FCNN-EM-L_1 for testing and validation are increased by 0.04 and 0.01 than that of the FCNN-L_1 model, which is relatively small. The increase in E_{std} is mainly due to the increased level of detail in the FCNN ice concentration, which are not captured by the image analysis.

Table 4.2: The average error statistics for FCNN-L_1, FCNN-EM-L_1 and FCNN-MS-L_1 using the Gulf of Saint Lawrence data

	Set	E_{sgn}	E_{L1}	E_{std}	E_{rmse}
	Training	-0.0027	0.0656	0.1406	0.1407
FCNN-L_1	Validation	-0.0114	0.1206	0.2209	**0.2211**
	Testing	0.0031	0.1291	0.2234	**0.2234**
	Training	-0.0064	0.0546	0.1313	**0.1315**
FCNN-EM-L_1	Validation	-0.0118	0.1271	0.2314	0.2317
	Testing	0.0015	0.1508	0.2601	0.26
	Training	-0.0066	0.0617	0.1379	0.1382
FCNN-MS-L_1	Validation	-0.0084	0.1364	0.2380	0.2381
	Testing	0.0222	0.1638	0.2686	0.2694

Overview of the results for a testing scene is illustrated in Fig. 4.8. All three methods can generate reasonable ice concentration in general. FCNN-EM-L_1 and FCNN-MS-L_1

show less underestimation of ice concentration for ice regions as compared to FCNN-L_1. FCNN-EM-L_1 and FCNN-MS-L_1 have different ice concentration estimates for ice floes. The results for the left-bottom sub-scene of Fig. 4.8 is extracted and shown in Fig. 4.9(a)-(f). These small ice floes are identified with higher ice concentration by FCNN-MS-L_1.

(a) HH   (b) HV   (c) ImA

(d) FCNN-L_1 (e) FCNN-EM-L_1 (f) FCNN-MS-L_1

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

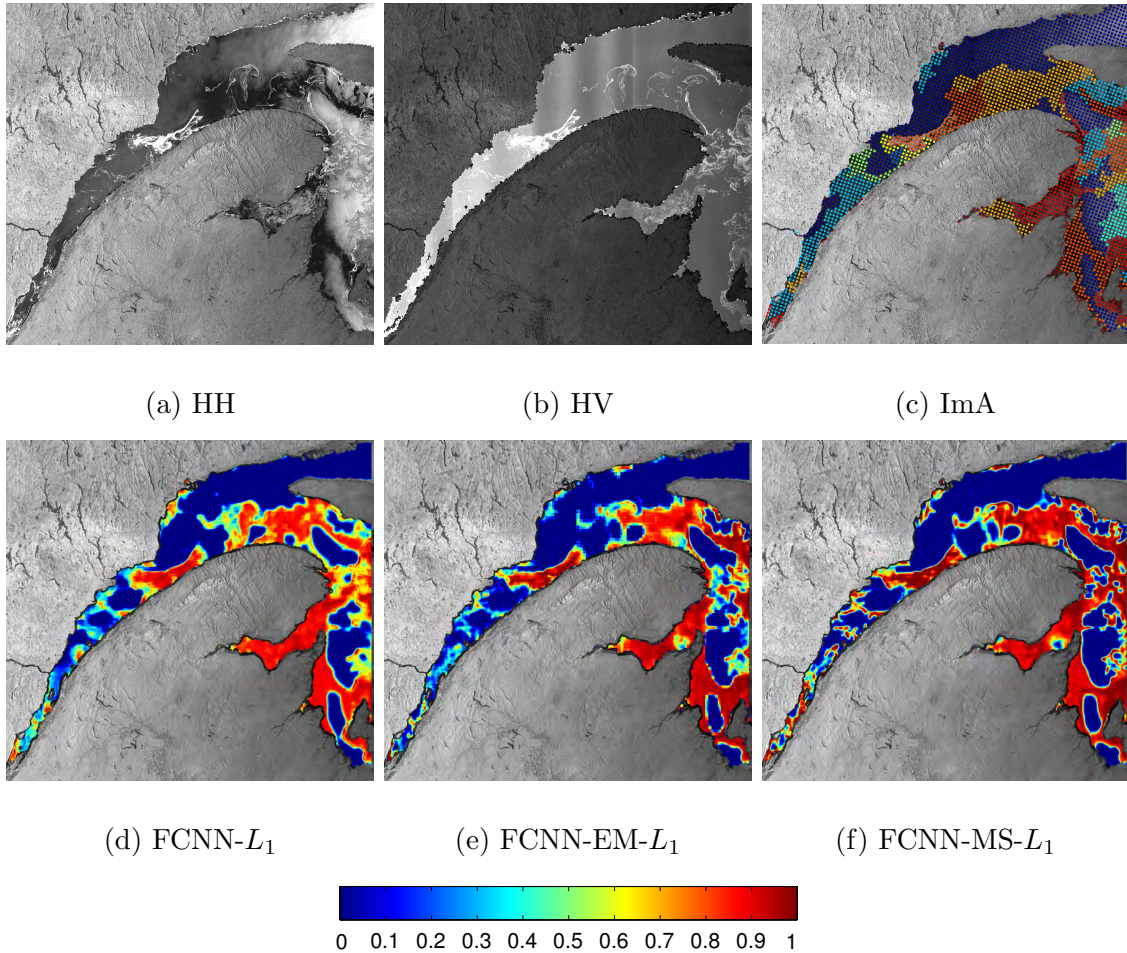Figure 4.8: Overview of the results of FCNN-L_1, FCNN-EM-L_1 and FCNN-MS-L_1. Scene 20140117_103914 in the testing set.

An example of the results for a region of intermediate ice concentration is illustrated in Fig. 4.10. FCNN-L_1 produces a result that is more consistent with the image analysis than FCNN-EM-L_1 and FCNN-MS-L_1, but the use of the EM algorithm and the mean-split loss

75

| (a) HH | (b) HV | (c) ImA | (d) FCNN-L_1 | (e) EM-L_1 | (f) MS-L_1 |

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
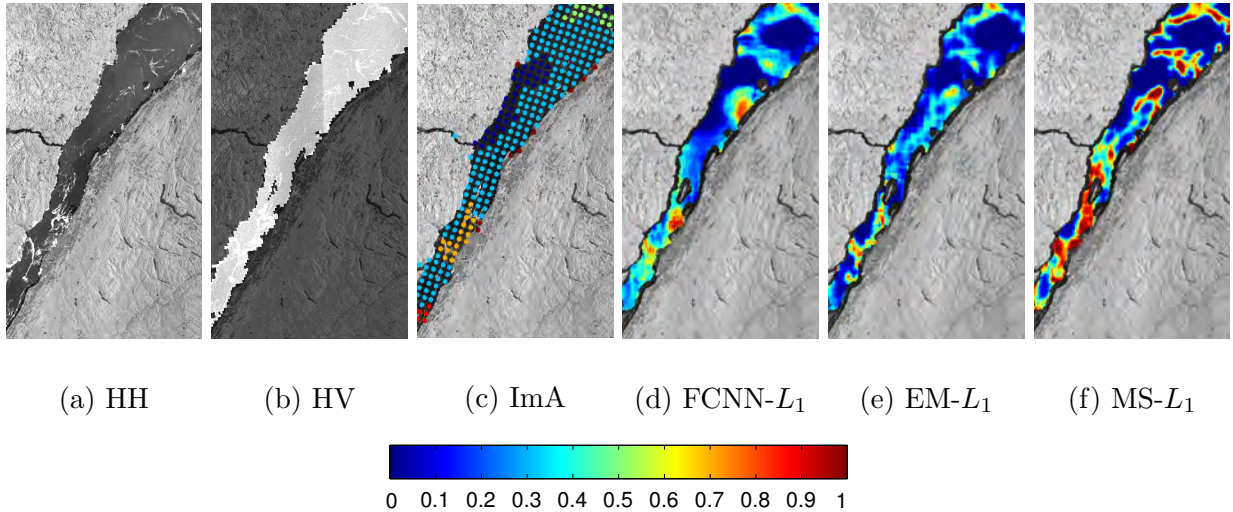
Figure 4.9: Comparision of FCNN-L_1, FCNN-EM-L_1 and FCNN-MS-L_1. FCNN-EM-L_1 and FCNN-MS-L_1 generate results with differences in details mainly for middle and low ice concentration regions. The first row is a subscene of 18 km × 16 km from 20140117_103914 in the testing set. FCNN-EM-L_1 identifies sparse ice floes with lower concentration than FCNN-MS-L_1.

can capture details (Fig. 4.10(e, f)) which are not captured by the image analysis (Fig. 4.10(c)) or the FCNN-L_1 (Fig. 4.10(d)). The water blobs in the FCNN-EM-L_1 result and FCNN-MS-L_1 result coincide with the dark regions in the SAR image (Fig. 4.10(a) and (b)), which could be water as this region is labeled with ice concentration 0.5 and 0.6. The modeling of representation errors using EM based algorithm or mean-split loss does not introduce extra trainable weights to the FCNN model used. Therefore, the improvement in predicting training images suggests that the modeling of representation errors is able to improve the learning of neural networks. Between FCNN-MS-L_1 and FCNN-EM-L_1, the level of details that can be seen in the ice cover are different. FCNN-MS-L_1 estimates the dark blobs in the HH pol image (Fig. 4.10(a)) with intermediate ice concentration, while the result of FCNN-EM-L_1 has lower ice concentration for the dark blobs. The EM algorithm is generating an ice concentration estimate with sharper features than the other approaches, which is expected. The EM based approach and the mean-split loss both use the image analysis as the average ice concentration in a patch, but the EM based approach used a linear model to increase the value range of the predictions. This increased value range makes the modified target value closer to 0 and 1, therefore the FCNN trained on these modified target values generates ice concentration values closer to 0 and 1 than the

(a) HH   (b) HV   (c) ImA   (d) FCNN-L_1 (e) EM-L_1 (f) MS-L_1
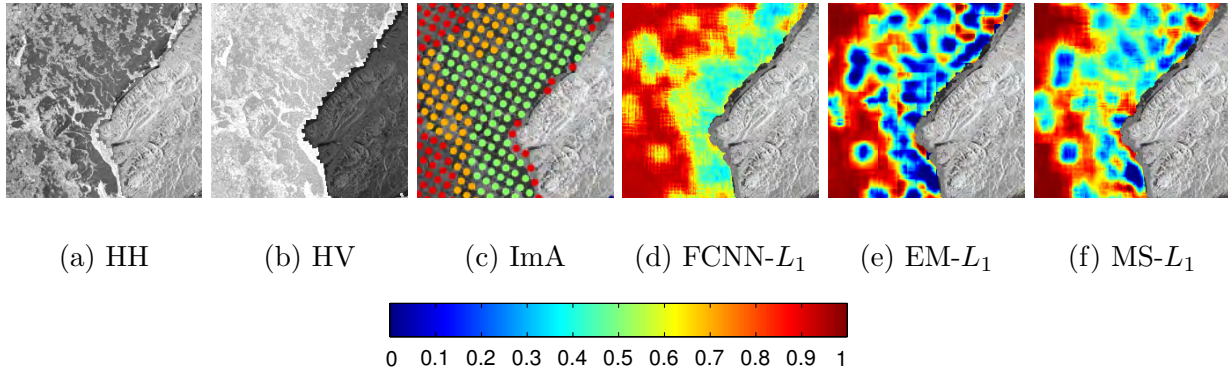
Figure 4.10: Comparison of FCNN-MS-L_1, FCNN-EM-L_1 and FCNN-L_1 for regions with intermediate ice concentration levels. (a) and (b) are visually enhanced for better viewing. Subscene of dimension 8 km × 8 km from 20140208_095758 in the training set.

FCNN trained on the mean-split loss.

Estimates for water regions are also improved as shown in Fig. 4.11. The result of FCNN-L_1 contains small chunks of spurious ice over water bodies. The spurious ice is reduced by the EM algorithm. Because training samples of ice concentration 0 or 1 are used directly for FCNN-L_1, FCNN-EM-L_1 and FCNN-MS-L_1, the improvement in water regions is due to the improved utilization of intermediate ice concentration samples that often contain a certain level of representation errors.

(a) HH   (b) HV   (c) ImA   (d) FCNN-L_1 (e) EM-L_1 (f) MS-L_1

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
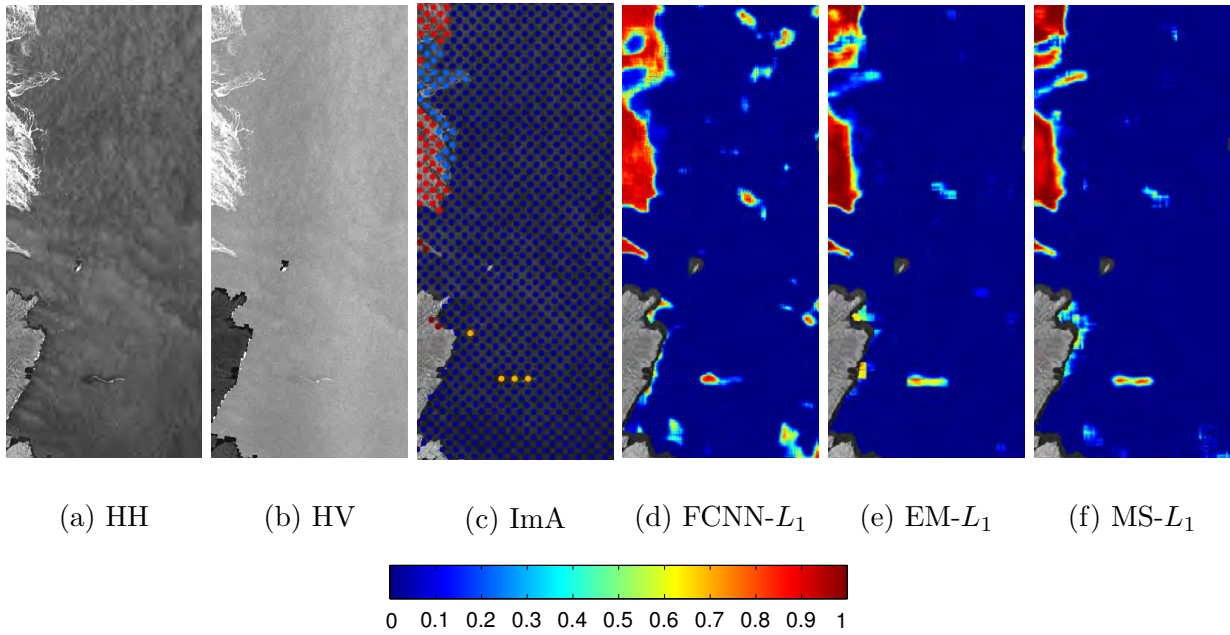
Figure 4.11: Comparision of water predictions for FCNN-MS-L_1 and FCNN-L_1. (d) FCNN-MS-L_1 is less affected by wind roughening than FCNN-L_1. Subscene of dimension 28 km × 16 km from 20140124_215646 in the training set.
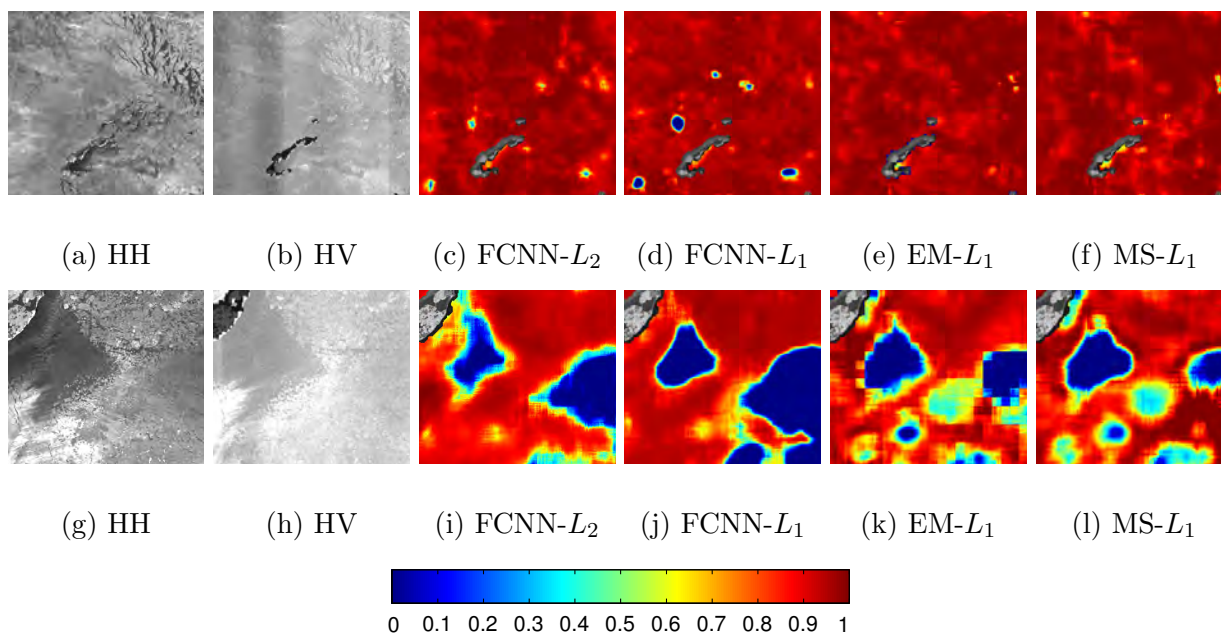
Figure 4.12: Comparison of FCNN-L_2, FCNN-L_1, FCNN-EM-L_1 and FCNN-MS-L_1 for ice estimates. Each row correspond to one example. Misidentification of ice with smooth textures and low backscatter is improved by FCNN-EM-L_1 and FCNN-MS-L_1 in (e), (f), (k) and (l) as compared to FCNN-L_2 and FCNN-L_1 in (c), (d), (i) and (j). Due to the limited space, EM-L_1 and MS-L_1 are used to represent FCNN-EM-L_1 and FCNN-MS-L_1. Both HH and HV are visually enhanced for better viewing. Subscene of dimension 10 km $\times$ 10 km from 20140206_221744 in valid dataset.

Ice with smooth textures and low backscatters is sometimes misidentified as water by the FCNN-L_1 model (Fig. 4.12). Misidentification of this type of ice is partially corrected by FCNN-EM-L_1 and FCNN-MS-L_1 in Fig. 4.12.

4.3.5 Discussion

Through the modeling of the two types of training sample errors, the ice concentration estimates are significantly improved. This improvement is mainly in water regions and the details. Due to the limitation in the resolution of the ground truth data, the actual improvement is not able to be quantified in this study, so the evaluation of the developed methods depends on visual inspection.

The EM algorithm and the mean-split loss can be used to refine existing ice charts. These algorithms use existing ice charts as training samples and learn a model to refine these ice charts in both accuracy and details. If applied for ice chart refinement, all the data can be used for training. Since the performance of the algorithms on training samples is generally good and stable, the refinement of existing ice charts is expected to have high potential for operational use.

The EM algorithm and mean-split loss can also be used together with L_2 loss. However the results show similar overestimation of water regions to FCNN-L_2, because L_1 loss is important for accurate water estimates. Therefore, only the results using L_1 loss are reported.

Limitations of the developed algorithms exist for regions of ice with smooth texture and low backscatter that can lead to confusion between ice and water, depending on the incidence angle. Since ice with a dark, smooth appearance is not common in the training data used, incorrect relationships between incidence angle and ice texture could be learned so that a change of incidence angle may lead to misidentification of ice to water under this circumstance. Additional training samples that include ice of this type under different incidence angles should benefit ice concentration estimation for these regions. Segmentation based post-processing may not be able to help in this situation because the features of smooth dark ice are very different from the features of its surrounding ice. Using a larger patch size to increase the spatial context information could potentially help to distinguish between dark, smooth ice and water, but it will also tend to blur the ice-water boundaries at the same time.

Chapter 5

Evaluation Using Beaufort Sea Data

The Gulf of Saint Lawrence data are acquired in the winter. In this semi-closed region, ice deformation is significantly affected by the coast and current. For large scale ice monitoring purposes, the performance of the methods developed in Chapter 3 and Chapter 4 need to be evaluated on other types of ice conditions. Ice with ponding or wet snow is a particularly interesting condition because it is common in the summer, at which time human activities are increased compared to other seasons, and it is difficult to separate water covered ice and open water using SAR or passive microwave data [42, 86]. In this chapter, the ice concentration estimation methods in previous chapters are evaluated on a dataset acquired during the melt season of the Beaufort Sea to further evaluate their robustness.

5.1   Dataset and Study Area

The study area is located in the Beaufort Sea, north of Alaska (Fig. 5.1). This region is particularly relevant for high resolution ice concentration mapping research due to oil and gas exploration in the region. The period of study is July to October in 2010 and 2011. Ice at this time of year in the Beaufort Sea is mainly a mixture of multi-year ice, and first year ice, with smaller quantities of new ice. Based on CIS (Canadian Ice Service) ice charts, the coverage of first-year ice is between 0 to 50% with under 30% of coverage for most images. The coverage of multi-year ice varies from around 30% to 90%. Packed multi-year ice often breaks into smaller ice floes, leaving large areas of open water between the floes which may refreeze to form new ice. Melting ice surfaces can be observed in this region until October [73]. Wet snow and melt ponds can be easily confused with open
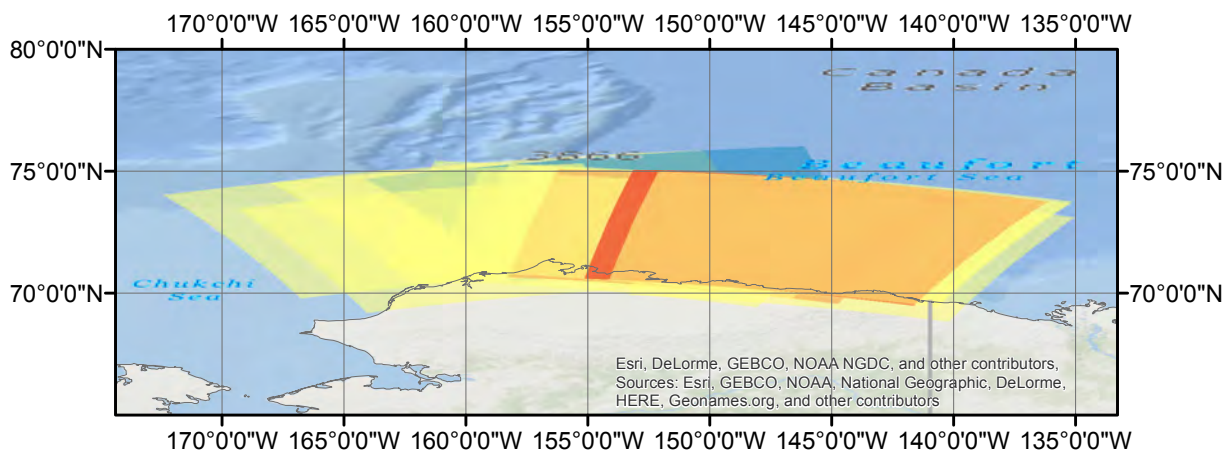
Figure 5.1: Study area and dataset for the Beaufort Sea. The coverage for each scene is marked in a translucent polygon with different colors. Yellow scenes are used for training, red are used for validation and blue for testing.

water in both SAR and passive microwave imagery, which can lead to difficulties in ice concentration estimation.

A set of 15 RADARSAT-2 dual pol (HH and HV) ScanSAR Wide beam images of the Beaufort Sea has been provided by MDA®. These images were acquired between July and October in 2010 and 2011. Pixel spacing and incidence angle range are the same as the Gulf of Saint Lawrence data which is 50 m by 50 m and 20° to 49° respectively. Each image covers roughly a 500 km by 500 km area. Corresponding image analyses are obtained and used as training data. The image analysis used for some scenes only cover a small portion of the SAR scene area. Different from the Gulf of Saint Lawrence dataset, banding in the HV pol for these Beaufort Sea images are already partly removed by MDA®.

5.2 Experiment Setup

The 15 scenes are divided into 12 scenes for training, 1 scene for validation and 2 scenes for testing (Table 5.1). The three scenes are selected for testing and validation because these images contain the almost all ice concentration levels and a variety of sea surface types, which can make validation and testing more informative. All the CNN based models are trained from random initialized weights. Dual polarized SAR images and incidence angle maps are input to the models together as three channels. The training parameters

described in Chapter 3 and Chapter 4 are used here. Since this region is less dominated by land boundaries than the Gulf of Saint Lawrence, land pixels of SAR images are masked to 0 instead of replaced by their nearest pixels of sea ice or water. The following methods are tested: CNN-L_2, FCNN-L_2, FCNN-L_1, FCNN-EM-L_1, FCNN-MS-L_1, MLP40.

Table 5.1: Beaufort dataset

Set	Scene ID	Date acquired	Number of image analysis points
Train	20100730	30 July 2010	1117
	20100806	6 August 2010	5934
	20100822	22 August 2010	8131
	20100829	29 August 2010	1915
	20100909	9 September 2010	8631
	20100929	29 September 2010	7162
	20101003	3 October 2010	9142
	20101008	8 October 2010	1569
	20110709	9 July 2010	488
	20110710	10 July 2010	2725
	20110720	20 July 2010	1851
	20110725	25 July 2010	9372
Valid	20110817	17 August 2010	6056
Test	20110811	11 August 2010	9246
	20101006	6 October 2010	7658

5.3   Results

The error statistics for the CNN based models, ASI and MLP40 are shown in Table 5.2. The image analysis charts are again used as the validation data. The ASI ice concentration used here is the AMSR-E daily ice concentration data [86]. The error statistics have a similar pattern to that of the Gulf of Saint Lawrence data. In general, the ice concentration estimated using the CNN models developed in this thesis have significantly better error statistics than the results from ASI and MLP40. ASI underestimates ice concentration by an average of about 15% for this dataset. The statistics of MLP40 are closer to our results. The CNN and FCNN generate results with better error statistics compared to FCNN-EM-L_1 and FCNN-MS-L_1 as the model is trained to fit the image analysis instead of the

modified image analysis. The biases for the four CNN-based methods developed in this thesis are around 5% when compared with the image analyses, the standard deviation of error is less than 13% for CNN and FCNN and less than 19% for FCNN-EM-L_1 and FCNN-MS-L_1 algorithm. Considering that the accuracy of the image analysis ice concentration is 10%, the CNN and FCNN results are in reasonable agreement with the image analysis. The error statistics are in general at the same level for training, testing and validation, which indicates that the impact of any overfitting is minimal.

An overview of the results of all the methods tested for the scene acquired in August 8, 2010 is illustrated in Fig. 5.2. This scene is used for training. It contains ice at the top and bottom and water in the middle. There is a transition area from ice to water that contains many ice floes of various sizes. Individual ice floes are not well captured by the image analysis as shown in Fig. 5.2(c). The ASI ice concentration largely underestimates ice concentration in areas with ice floes and other middle to low ice concentration regions. The methods developed in this thesis are able to produce more reasonable ice concentration than ASI (Fig. 5.2(e-i)) and MLP40. MLP40 produces an ice concentration estimate that is noisy with large overestimation in water regions. The four CNN based approaches have much cleaner and smoother ice concentration estimates compared to MLP40. The FCNN-L_2 result is slightly better than CNN-L_2 due to reduced underestimation in ice and reduced overestimation in water (Fig. 5.2(e) has mild overestimation in water regions). The use of L_1 or L_2 with FCNN (Fig. 5.2(g)) produce similar results, which are close to the image analysis. The FCNN-EM-L_1 and the FCNN-MS-L_1 algorithm produce results with greatly improved details of ice structures and ice-water boundaries. Even very small ice floes are captured and represented as regions with high ice concentration. Similar results are acquired for testing image. Such an example is shown in Fig. 5.3.

The details of a subregion in Fig. 5.2 are shown in Fig. 5.4. This patch is an area of mixed ice and water. The image analysis chart (Fig. 5.4(c)) indicates the upper region of this patch has intermediate ice concentration. MLP40 is able to capture small details of ice structures in SAR images (Fig. 5.2(e)). CNN-L_2 underestimates ice concentration slightly (Fig. 5.2(f)). Compared to CNN-L_2, FCNN-L_2 produces results closer to the image analysis. The details and ice structures are much improved by EM and MS as compared to FCNN-$L1$, although the errors evaluated against the image analysis could be even higher than the CNN or FCNN results (Fig. 5.4(e-g)). Compared to other methods, FCNN-EM-L_1 and FCNN-MS-L_1 estimate the ice floes in the center of the image with substantially higher ice concentration, which is more reasonable when ice concentration is estimated in finer scale than the image analyses. FCNN-EM-L_1 and FCNN-MS-L_1 have similar results in this area, with slight difference in the left top of the image. Quantitative comparison of FCNN-EM-L_1 and FCNN-MS-L_1 is currently impossible due to the lack of

(a) HH   (b) HV   (c) ImA

(d) ASI   (e) MLP40   (f) CNN-L_2

(g) FCNN-L_2 (h) FCNN-L_1 (i) FCNN-EM-L_1
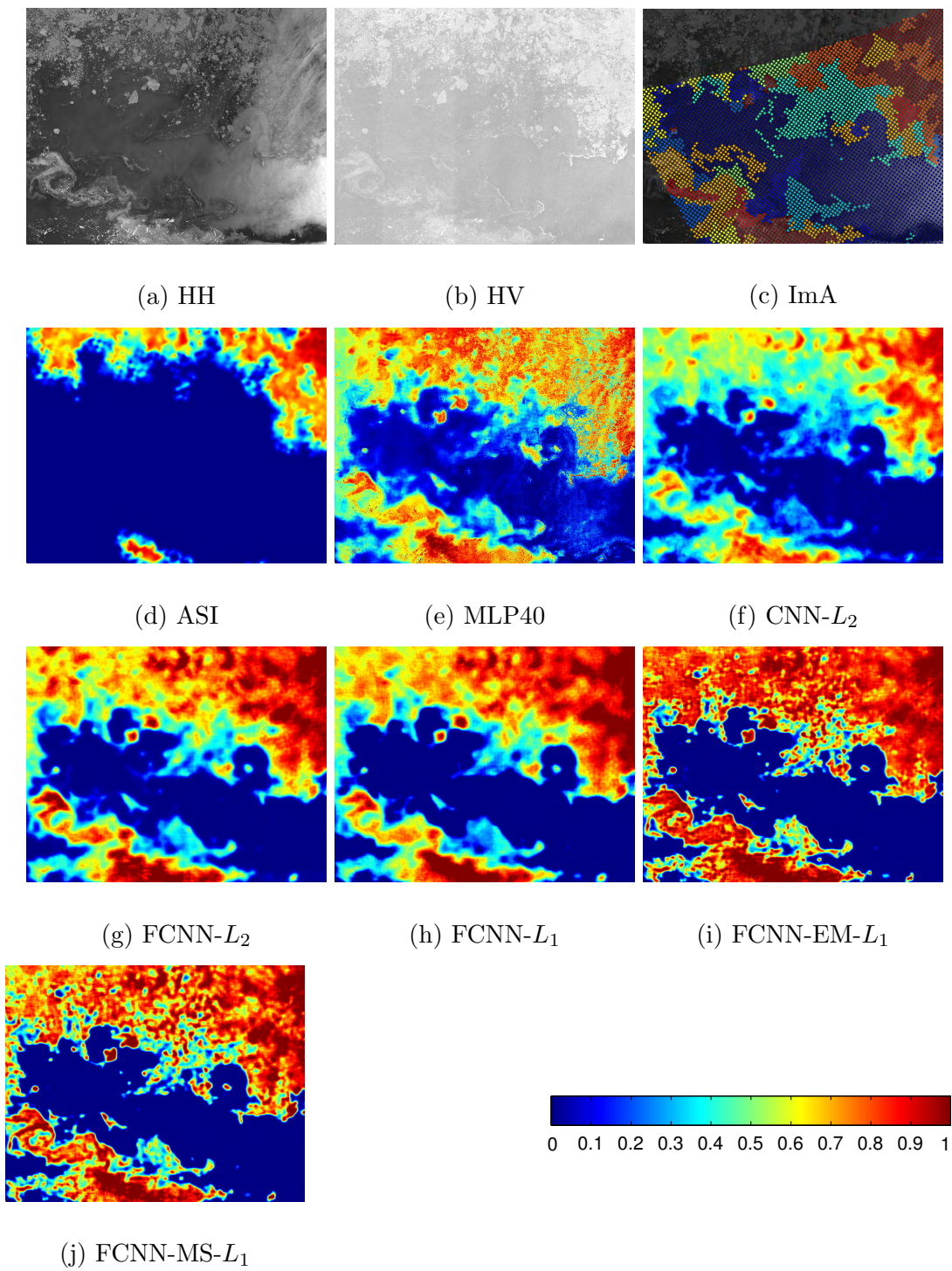
(j) FCNN-MS-L_1

Figure 5.2: The results of all the methods developed in the Beaufort Sea for a scene acquired in August 6, 2010 which is used for training.

Table 5.2: The average error statistics of different algorithms on the Beaufort Sea data

Set	E_{sgn}	E_{L1}	E_{std}	E_{rmse}
ASI				
Training	-0.1571	0.1649	0.2523	0.2973
Validation	-0.1493	0.1610	0.2315	0.2755
Testing	-0.0578	0.0718	0.1399	0.1513
MLP40				
Training	-0.0093	0.0963	0.1549	0.1552
Validation	0.0724	0.1674	0.2117	0.2236
Testing	0.1211	0.1550	0.1746	0.2124
CNN-L_2				
Training	-0.0423	0.0911	0.1546	0.1603
Validation	-0.0112	0.0817	0.1385	0.1389
Testing	-0.0063	0.0562	0.1122	**0.1122**
FCNN-L_2				
Training	−0.0129	0.0633	0.1289	0.1296
Validation	0.0196	0.0701	0.1266	**0.1281**
Testing	0.0319	0.0625	0.1177	0.1221
FCNN-L_1				
Training	-0.0065	0.0611	0.1283	**0.1285**
Validation	0.0249	0.0740	0.1336	0.1360
Testing	0.0464	0.0723	0.1346	0.1425
FCNN-EM-L_1				
Training	0.0103	0.0734	0.1520	0.1523
Validation	0.0346	0.0825	0.1547	0.1584
Testing	0.0868	0.1050	0.1912	0.2100
FCNN-MS-L_1				
Training	-0.0037	0.0721	0.1498	0.1497
Validation	0.0243	0.0844	0.1584	0.1603
Testing	0.0581	0.0886	0.1726	0.1822

ice concentration ground truth at this scale.

The predictions over open water regions are significantly improved when training sample errors are taken into account, as shown in Fig. 5.5. This is a particularly difficult scene due

to the complex appearance of water regions, which have rough textures and low backscatter in both pols in the upper left and smooth texture and high backscatter in the lower right. The intersection of the two water regions has a sharp change in backscatter, which can be easily misidentified as ice. The direct use of CNN or FCNN with L_2 loss overestimate ice concentration in the water regions. The models that consider the training sample errors in Fig. 5.5(h, i, j) are able to produce accurate ice concentration estimate for open water regions under the complex surface conditions. Therefore, more representative features can be learned by dealing with training sample errors.

5.4 Discussion

The impact of patch size has also been tested on another model developed earlier on the Beaufort Sea data [91]. CNN models with smaller patch size tend to underestimate ice concentration in this dataset. Intuitively, the strength and small scale texture of the backscatter of a melt pond are very similar to calm water, so the correct identification of melt ice needs more information from its neighborhood in the image. A small patch size leads to confusion between melt pond and open water due to insufficient spatial context information. Using small patch size also shows more overestimation caused by banding in the HV pol. This is different from the Gulf of Saint Lawrence data, where the impact of patch size comes mostly from the banding effect and wind roughening. A patch size of 45 pixels (18 km ground distance) is found to generate good results in practice.

Due to the limited available training samples for the Beaufort Sea dataset, a small number of test scenes are carefully selected to cover as many types of ice conditions and concentrations as possible in our experiment, which may still not represent the distribution of the training samples very well. In our case, the selected test data contains a lower percentage of melt surfaces than in the training data. Because melted surfaces are generally more likely to be confused with water than other types of ice, the estimated ice concentration generally have larger errors for melted surfaces than other types of surface conditions. This may be the cause of the unusual situation in which the test errors are lower than the training errors for CNN-L_2 and FCNN-L_2. Further experimentation is needed to determine this.

The models are also trained with randomly divided training, validation and testing samples. The number of training samples is relatively small for the Beaufort Sea dataset as compared to the Gulf Saint Lawrence dataset. Smooth ice at high incidence angles can be misidentified as water sometimes due to insufficient relevant training samples. Due to the melt and smooth surface, i.e., first year ice can appear smooth in backscatter, it

is common for ice during the melt season to have smooth image textures. Therefore, the correct identification of such ice surfaces is important for ice mapping during melt conditions. When the incidence angle data are not used, the smooth ice can be correctly identified in general, but ice concentration estimates in water regions are noisy [91]. The model may tend to find a simple relationship between the image backscatter and the incidence angle to determine ice concentration when certain types of ice samples are small and happen to occur in narrow incidence angle ranges, which is the case for the Beaufort Sea data. To confirm our speculation, incidence angle data and the SAR image in the CNN or FCNN model are decoupled (experiments are not included in this thesis), i.e., two streams of convolutional layers that take incidence angle and SAR image as input separately are used, and the two streams are combined in upper layers. By decoupling SAR image and incidence angle in the lower layers, the CNN feature extraction is less affected by the incidence angles. The decoupling of incidence angle from SAR images in the model shows reduced misidentification of ice as water when training samples are not sufficient (Fig. 5.6). When enough data are available in operational situations, the decoupling of incidence angle from image is expected to have little effect on the results. Therefore, the incidence angle data are used with HH and HV pol SAR images as input to the first convolutional layer.

The training process may run for 5×10^4 mini-batches (128 samples/patches in each mini-batch). In the experiment, we observe that the training error curve has a long tail. The training and validation error started to converge at around 10^4 mini-batches. Then the training error kept on decreasing slowly for 4×10^4 more mini-batches until convergence. This long and inefficient training process may be caused by imbalanced training samples used. Imbalanced training samples means there is a dominant class in the samples. In our case, the number of water (ice concentration is 0) samples is about 8 times the second most common ice concentration level in the training samples. Intuitively, this pushes the model to a "dangerous local minimum" quickly [20]. When the model is at this local minimum, it detects most of the input as water and can still achieve a low cost. This causes underestimation of the ice concentration in general, if the training is stopped early. Many epochs are expected to be required to move away from this local minimum. There are several approaches that may be investigated to resolve this issue, including under sampling the majority [59, 98], over sampling the minority [59, 98] or using a Bayesian cross entropy loss function [20]. Our experiments show that learning normally converges quickly when using those methods, but none of the aforementioned methods converge to a model better than training directly on all the training samples for a long time. In this study, we choose to prioritize precision and accept the long training time. The network is trained for about 10 hours to ensure convergence.

(a) HH

(b) HV

(c) ImA

(d) ASI

(e) MLP40

(f) CNN-L_2

(g) FCNN-L_2

(h) FCNN-L_1

(i) FCNN-EM-L_1

(j) FCNN-MS-L_1
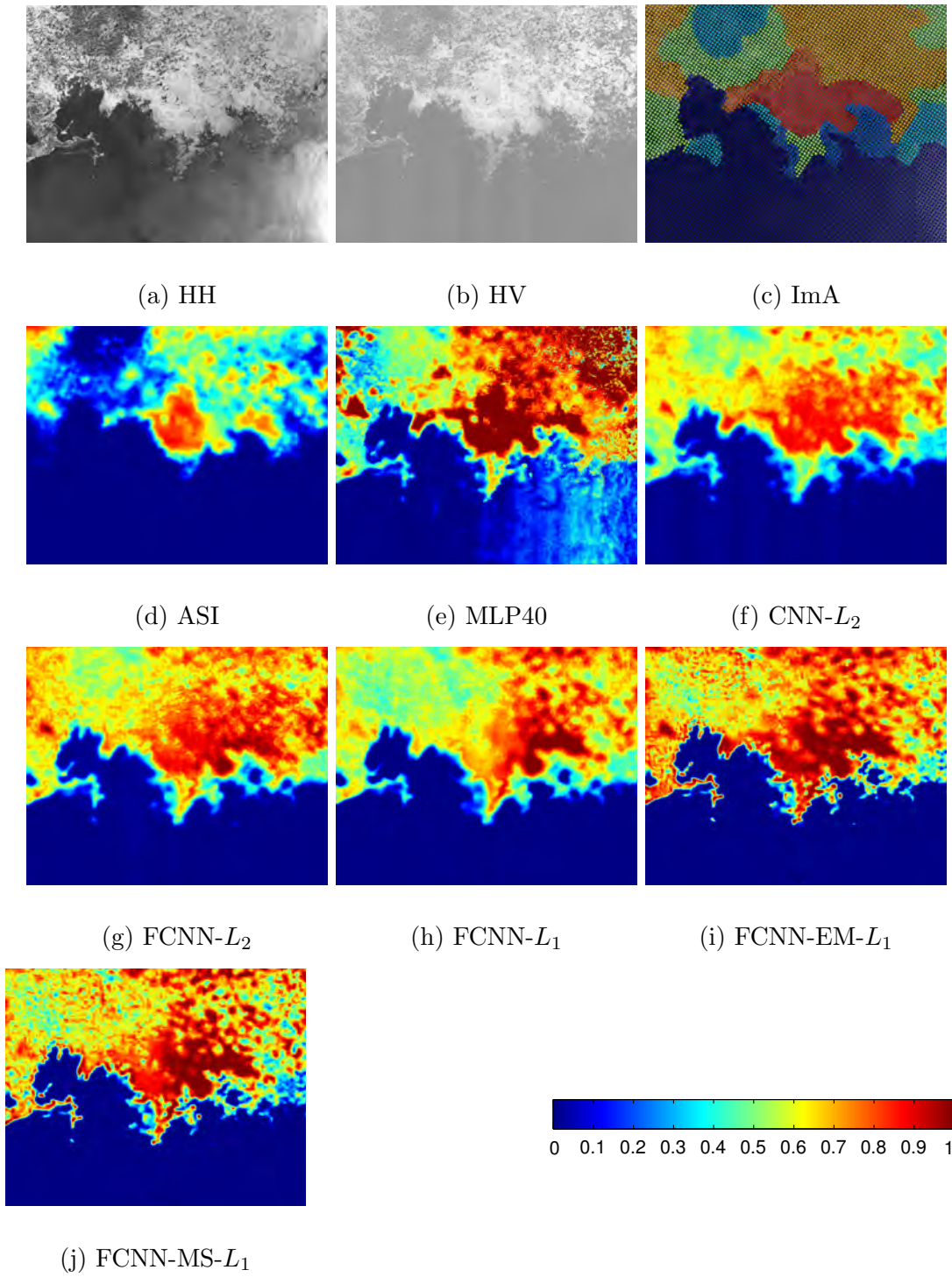
Figure 5.3: The results of all the methods developed in the Beaufort Sea for a scene acquired in August 11, 2011 which is used for testing.

(a) Location   (b) HH   (c) HV   (d) ImA

(e) ASI   (f) MLP40   (g) CNN-L_2 (h) FCNN-L_2

(i) FCNN-L_1 (j) FCNN-EM-L_1 (k) FCNN-MS-L_1
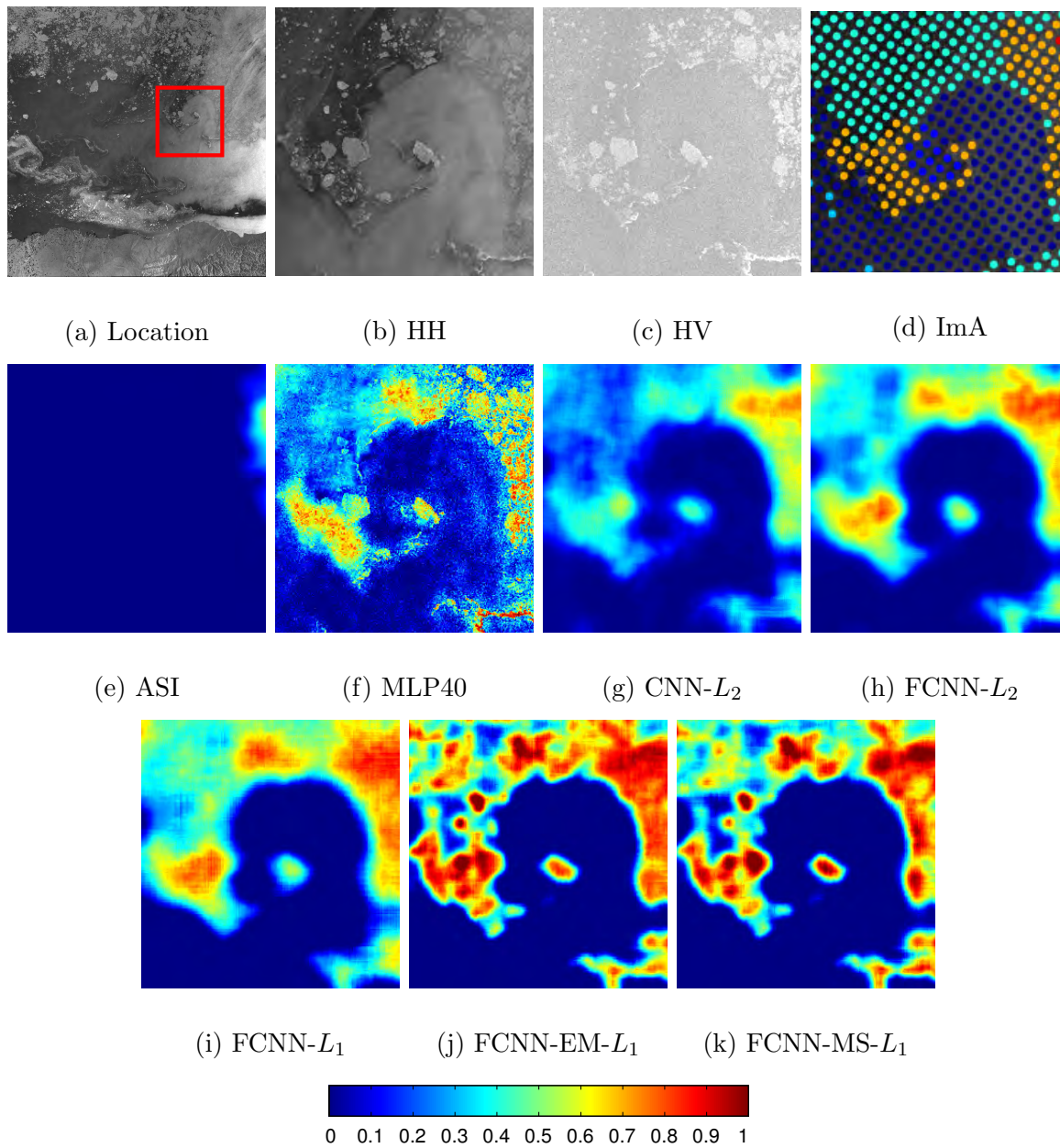
Figure 5.4: The details of the results of all the methods developed in the Beaufort Sea for a scene acquired in August 6, 2010. (a) marks the location of (b)-(k) in the scene by a red rectangle. (d)-(k) use the color scheme in the bottom.

(a) HH (b) HV (c) ImA

(d) ASI (e) MLP40 (f) CNN-L_2

(g) FCNN-L_2 (h) FCNN-L_1 (i) FCNN-EM-L_1

(j) FCNN-MS-L_1
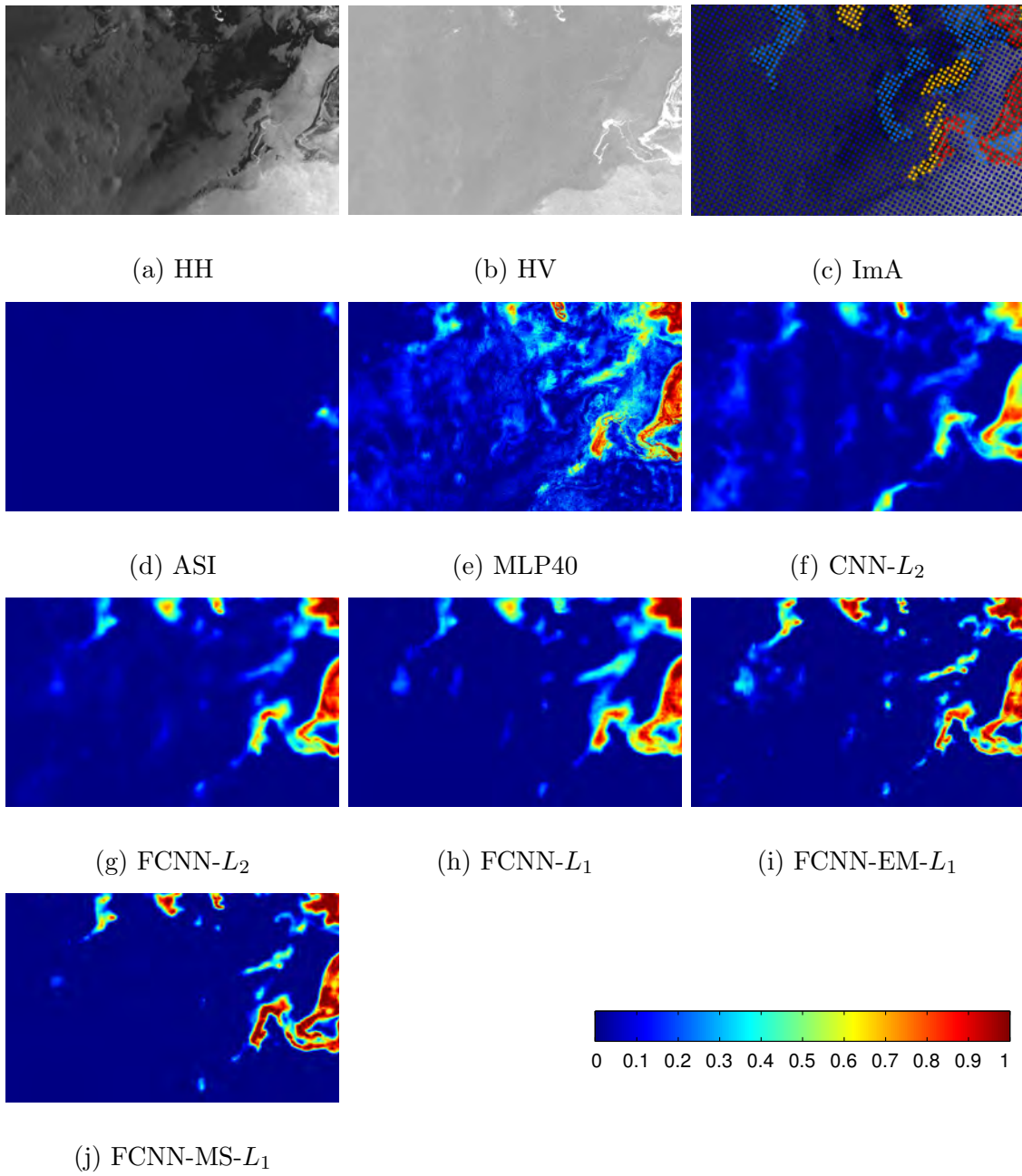
Figure 5.5: Results for all methods applied to Beaufort Sea scene acquired September 29 2010. The modeling of human errors and representation errors in training samples are able to improve predictions in water regions.

(a) HH   (b) HV   (c) coupled   (d) decoupled

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
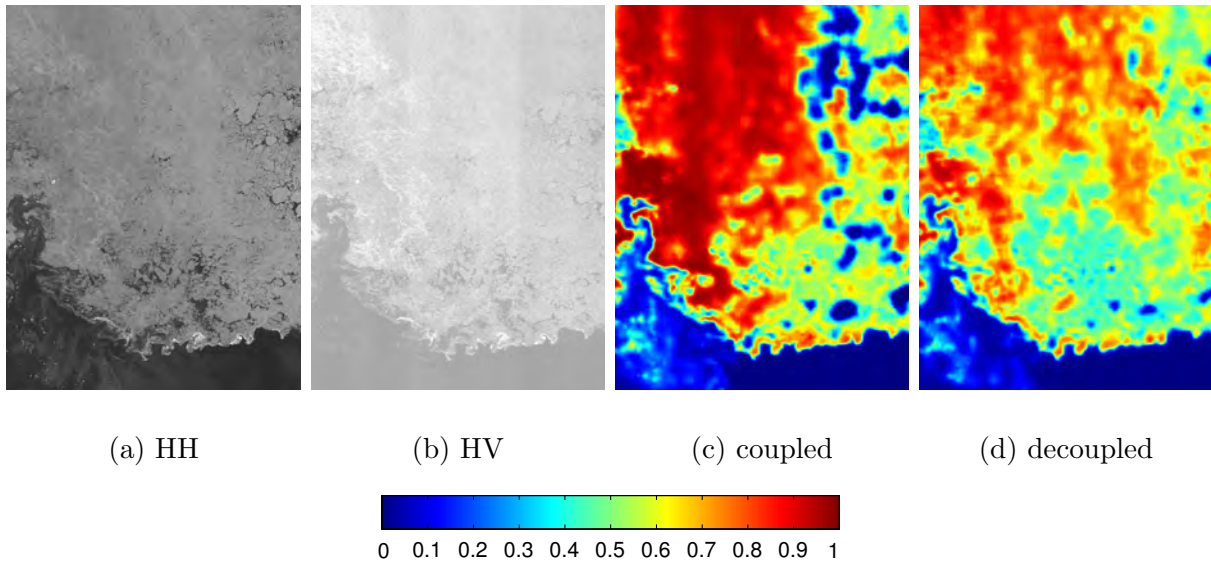
Figure 5.6: Decoupling of incidence angle data from SAR images reduces misidentification of smooth ice as water when training samples are insufficient. The image analysis chart for this scene does not cover this selected region. The dark regions in the bottom are open water with some ice floes in the left. The rest of the image is ice with smooth textures. (c) smooth ice been misidentified as water in the top right of the image due to the lack of relative training samples. (d) misidentification of smooth ice is reduced by decoupling incidence angle and SAR image in the network. Image acquired in July 20, 2011.

Chapter 6

Transferability Analysis

Many deep convolutional neural networks generate features that are similar to Gabor filters in their bottom layers [97], even for CNNs trained for irrelevant tasks. The learned features in lower layers of the networks appear to be less specific to certain data or task and can be applied to many datasets and tasks. How well the neural network trained on one task or dataset performs on other tasks or dataset is the transferability of the model. Ice concentration estimation for the Beaufort Sea and Gulf of Saint Lawrence are two closely related tasks. How the model trained in one region performs in another region and how a trained model can be optimally used for other regions are two questions related to the deployment of machine learning models for operational use. A case study pertaining to these two questions is presented in this chapter by analyzing the transferability of the CNN model trained on the Gulf of Saint Lawrence to the Beaufort Sea data.

6.1 Experiment Setup

In CNNs and FCNNs, lower layers are more transferable than upper layers because they are less specific to the task. So the transferability analysis is normally done by layer [97, 71]. However, upper layer features in a CNN depend on their lower layer features, so using a layer without all of its' lower layers in a different task will not generate any useful result. So, a number of upper layers of a neural network model trained on one dataset are retrained on another dataset to test the transferability of these bottom layers [97, 71]. This retraining process is called "fine tuning" [71]. Because of the fragile dependencies (also called co-adaptation) between features extracted in neighboring layers of a network [97], i.e., the co-adaption cannot be fully recovered when some layers are transferred from other models,

neural network parameters for all layers are initialized from the model trained on the Gulf of Saint Lawrence data and fine tuned on the Beaufort Sea data in our experiment. There are five trainable layers for our neural network structure. Experiments of fine tuning with freezing 0 (allow all layers to be fine-tuned) to 5 (model not fine tuned) bottom layers are performed.

Compared to the other models developed in this thesis, FCNN-L_1 generates results that are more similar to the image analyses in both scale and accuracy, because it is trained to mimic the scale of image analysis with robustness for human errors in the training samples. The evaluation against image analyses is more meaningful for FCNN-L_1. Therefore, the transferability of FCNN-L_1 model from the Gulf of Saint Lawrence to Beaufort Sea is tested.

Since fine tuning is trying to adapt an existing model to new data, small changes of the model weights are expected, so the learning rate for the fine tuning experiment is decreased from 10^{-3} to 10^{-4}. All other training parameters are not changed from the experiment described in Chapter 3.

6.2   Results and Discussion

The main results of the transferability analysis are provided in Fig. 6.1. Fine tuning all layers of the network trained on the Gulf of Saint Lawrence data does not show any improvement in performance, as compared to training from random initialized weights in this case. When more bottom layers are frozen in the training process, fewer layers are allowed to adapt to the new data which results in increased errors. The increase of errors is small when the first or the first two layers are frozen. The fact that the first two layers trained in the Gulf of Saint Lawrence can be reused without modification in the Beaufort Sea data suggests that CNN models can find generic low-level representations for ice concentration estimation. This is unexpected as the ice features are quite different in the Beaufort Sea from the ice features in the Gulf of Saint Lawrence. Transferring the bottom two layers from models trained on one dataset to another dataset can reduce the required number of training samples, which could be useful when the training samples are insufficient for a new region. When more layers are frozen, the increase of errors also increases, because the features from higher layers are more specific to the task and are less transferable than lower level features. When using the CNN model trained on the Gulf of Saint Lawrence dataset with all layers frozen, the E_{rmse} is 0.4 when tested on Beaufort Sea data, which is not usable.
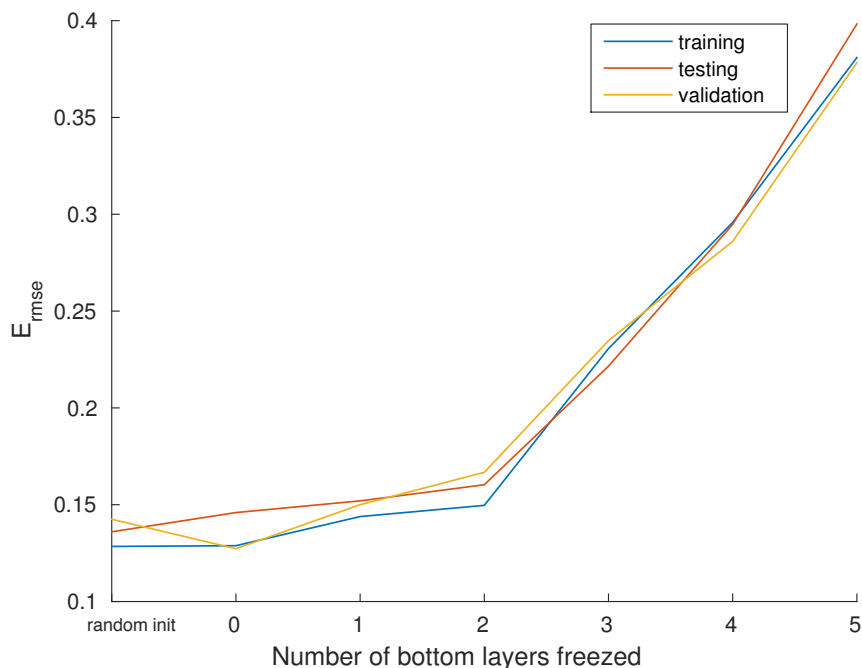
Figure 6.1: The mean squared errors for the fine tuning experiment. Model trained in the Gulf of Saint Lawrence is fine tuned on Beaufort Sea data. Label 0 to 5 in the horizontal axis are the number of bottom layers frozen in fine tuning. "random init." in the left end of the horizontal axis represent the model trained on the Beaufort Sea data from random initialized weights as has been done in previous chapters. The vertical axis is the mean squared error evaluated on the Beaufort Sea data.

The fine tuning experiment often converges faster than training from random initialized weights. The number of iterations required for convergence is illustrated in Fig. 6.2. The training loss converges in 20000 iterations in general for all experiments. Fine tuning with all the layers open converges in 1000 iterations, which is 19 times less than the neural network trained from random initialized weights (Fig. 6.2(a)). When more convolutional layers are frozen in training, the number of iterations required for convergence first increases. This may be caused by the fragile dependencies (co-adaptation) between layers. The recovering of the correct upper level features based on existing low level features is difficult and may lead to increased training iterations. The number of iterations required for convergence reaches the highest when the first three convolutional layers are frozen. With more layers frozen, the iterations required for convergence decreases, due to the significantly reduced number of trainable parameters.
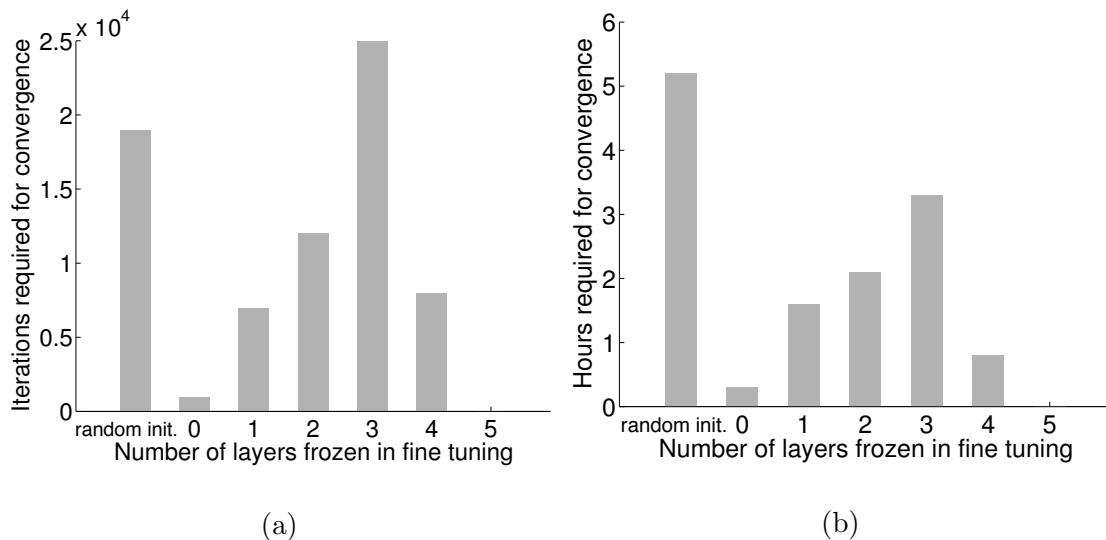
95

Figure 6.2: Iterations and time required for convergence change with the number of layers frozen during fine tuning. (a) Iterations required for convergence to the number of bottom layers frozen. (b) Computational time required for convergence to the number of bottom layers frozen.

The training time is the number of iterations multiplied by the time required per iteration. In each training iteration, there is one forward pass and one backward pass. With more layers frozen in training, the computation required for one forward pass is not affected, but the computation required for backward pass is reduced because back-propagation is not calculated for the frozen layers. So the training time required for the fine tuning experiments are also plotted separately in Fig. 6.2(b). All fine tuning experiments require less training time than training from random initialized weights. Except for no fine tuning (when the number of frozen layers is five), fine tuning all layers takes the least time to train in our case, which is 17 times faster than trained from random initialized weights.

6.3 Operational deployment

The transferability analysis suggests that fine tuning an existing model on the target dataset can significantly improve the performance. The benefit of fine tuning has also been demonstrated in many other studies [49, 29, 71]. We conclude that in operational deployment, it is always recommended to fine tune an existing model on the target area,

instead of training a new model from random initialized weights or using an existing model without modification.

The benefit of fine tuning depends on how the base model is trained. If the base model is trained on a large training dataset that covers a large variety of sea surface conditions, the fine tuning might not be able to improve the accuracy because the large training sample base has a large possibility to contain samples with similar surface conditions to the target location.

This concern naturally leads to the question of how to train a good base model for ice concentration estimation. A big base model can be generated for all locations and seasons, or different models can be generated for different locations or seasons. One base model is simple to use and maintain. However, this model needs to have more weights than location or season specific ice concentration estimation models to deal with the increased complexity in larger scale ice concentration estimation applications, which will require more time for using the model for ice concentration estimation. It is also not straightforward to train one big model for all locations and seasons with the location and season information included. Location and season provide prior knowledge of the data which can help the separation of ice and water from SAR images. Therefore, generating small models that correspond to different locations or seasons is more reasonable and can potentially achieve better performance than the one model approach. These small models can share the same low-level layers to reduce the requirement for the number of training samples.

Chapter 7

Conclusion

The need for high resolution ice concentration mapping in the Arctic is continuously increasing with the increasing human activities in the Arctic. Automation of the current ice charting method, which depends on manual interpretation of SAR images, is desired. There have been several attempts in automatic ice concentration estimation from SAR images in the past decade. The difficulty in recognizing ice and water for constantly changing surface conditions in space and time has prevented these algorithms from being deployed for operational use. This work is set out to design an algorithm to estimate ice concentration from dual-polarized SAR images that is easy to deploy and can adapt to the variations of surface conditions in space and time. This goal has been accomplished by building a complex neural network model to model the complexity of ice and water from the large and still growing volume of archive ice charts and SAR images.

In Chapter 3 two convolutional neural network designs are presented and tested on data acquired in the winter of the Gulf of Saint Lawrence. The two models are conventional convolutional neural network and its fully convolutional variation. Through sensitivity analysis of the hyper-parameters, a base network structure that takes SAR images and incidence angle as input is developed. Input patch size, number of convolutional layers and incidence angle data are found to have substantial influence on the performance. In general, larger patch size, deeper network and the using of incidence angle are demonstrated to improve the performance. The fully connected convolutional neural network is used for further development mainly due to its efficiency in prediction and slightly improved accuracy compared to the classic convolutional neural networks.

In Chapter 4 training sample errors are modeled and improvements are found in ice-water boundaries and water regions. L_1 loss is used to reduce the effect of human errors in

the training samples, which is demonstrated to be more robust to training sample errors compared to L_2 loss. Two independent methods, an EM algorithm and mean-split loss are devised to model the representation errors. The EM and mean-split algorithm have one and zero parameters separately, which makes them simple to use and relatively robust to parameter setting. However, due to the lack of ground truth data at higher resolution than the image analyses, the improvement in ice structures and ice-water boundaries cannot be evaluated quantitatively. The evaluation of the results therefore heavily depends on visual analysis and comparison.

In Chapter 5 the methods developed in Chapter 3 and Chapter 4 are used on a Beaufort Sea dataset that was acquired in the summer of 2010 and 2011 to evaluate their performance on summer ice with different spatial structures than the ice in the Gulf of Saint Lawrence. Sate-of-the-art ice concentration estimates are achieved in the Beaufort Sea. The CNN models are found to be able to generate reasonable ice concentration in the melt season, which has demonstrated its wide applicability for ice concentration estimation tasks.

Lastly, the model transferability is evaluated by a case study in Chapter 6. The model trained on the Gulf of Saint Lawrence dataset is used and fine tuned on the Beaufort Sea dataset. The first two convolutional layers are found to be transferable between the two study areas. The transferability analysis suggests that models trained in a different dataset for a different time or region can be fine tuned to adapt to new regions without losing performance. The major benefit of fine tuning from a previously trained model is its faster convergence, which would be desired for operational uses. In this case, the model convergence can be accelerated by over 10 times.

7.1   List of Contributions

The unique and major contributions of this work are as follows:

1. Application of convolutional neural networks to the problem of ice concentration estimation from SAR images. Convolutional neural networks are found to be promising for the purpose of operational ice charting.

2. An optimal convolutional neural network structure that is suitable for ice concentration estimation from SAR images.

3. The effect of human errors and representation errors in the image analyses are analyzed. A set of methods to train the convolutional neural network that considers

the human errors and representation errors in image analyses, which can generate ice concentration estimates with abundant details is proposed.

4. Analyzed the transferability of CNN models for ice concentration estimation in the Beaufort Sea and Gulf of Saint Lawrence, which resulted a training scheme with greatly improved efficiency.

7.2 Limitations

There are several limitations of the methods developed in this thesis:

1. The main limitation is the misidentification of dark ice with smooth image textures, such as those from calm water, which is caused by the confusion of water and ice features in SAR images. The correct identification of this type of ice may require the use of prior knowledge or other types of data, or spatial context analysis.

2. Due to the limited data available, the CNN or FCNN structure developed in this study may not be the optimal structure for SAR images in other geographic regions or seasons.

3. Even though the ice-water boundaries are refined when the representation errors in training samples are considered, the ice-water boundaries in the results are still not as sharp as the ice-water boundaries in SAR images. Small ice structures are missing sometimes when the ice concentration is low.

7.3 Extensions and Future Work

Based on the limitations of this study, there are several possible extensions to improve the current algorithms for future studies.

1. Jointly train a conditional random field (CRF) model and network parameters with the representation errors modeled. Using CRF to explicitly model the local image structure can help the identification of dark, smooth ice and can also sharpen the ice-water boundaries in the results.

2. Reducing the co-adaptation between SAR image and incidence angle, which may help the detection of ice that is often confused with water and reducing the required number of training samples. This could be helpful when the available training samples are insufficient. Some investigation in this direction by separating the SAR images and incidence angle data to different layers shows some level of success. Another possible method to reduce the co-adaptation is to perturb the incidence angle data for regions of ice, as the appearance of ice is less affected by incidence angle when there is no melt condition in the ice surface.

3. Polygon based training. Currently, ice concentration from the image analysis is compared to the average ice concentration in a patch. Comparing ice concentration from the image analysis with the average ice concentration from that polygon is a more natural and accurate way to use the image analyses, which could potentially improve the training of FCNN models.

4. Assimilation of the high resolution ice concentration estimates. The forecasting of ice concentration could potentially benefit from the improved ice concentration estimates. Consider every pixel as an observation in the data assimilation system, the assimilation of such dense observations is a challenging task.

References

[1] Ko Aagaard and E. C. Carmack. The role of sea ice and other fresh water in the Arctic circulation. *Journal of Geophysical Research: Oceans*, 94(C10):14485–14498, 1989.

[2] Tom Agnew and Stephen Howell. The use of operational ice charts for evaluating passive microwave ice concentration data. *Atmosphere-ocean*, 41(4):317–331, 2003.

[3] San Antonio. Microwave Remote Sensing of Sea Ice Basics - I. 2010.

[4] Joseph L. Awange and John B. Kiema. Microwave Remote Sensing. In *Environmental Geoinformatics*, pages 133–144. Springer, 2013.

[5] Yoshua Bengio. Learning deep architectures for AI. *Foundations and trends in Machine Learning*, 2(1):1–127, 2009.

[6] Yoshua Bengio. Representation Learning. Technical report, 2012.

[7] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(8):1798–1828, August 2013.

[8] A. Berg and L. E. B. Eriksson. SAR algorithm for sea ice concentration-evaluation for the Baltic Sea. *Geoscience and Remote Sensing Letters, IEEE*, 9(5):938–942, 2012.

[9] Sean Borman. The expectation maximization algorithm – a short tutorial. *Unpublished paper*, 2004.

[10] Mark Buehner, Alain Caya, Lynn Pogson, Tom Carrieres, and Paul Pestieau. A new Environment Canada regional ice analysis system. *Atmosphere-Ocean*, 51(1):18–34, 2013.

[11] A.V. Bushuyev. Sea ice nomenclature. Technical Report WMO-No.259 Suppl.No.5, The Joint WMO-IOC Technical Commission for Oceanography and Marine Meteorology, 2009.

[12] Thomas Bvith and Sren Andersen. Sea ice concentration from single-polarized SAR data using second-order grey level statistics and learning vector quantization. *DMI Scientific Report*, (05-04), 2005.

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

[14] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*, 2014.

[15] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Semantic image segmentation with deep convolutional nets and fully connected CRFs. *arXiv preprint arXiv:1412.7062*, 2014.

[16] Xueyun Chen, Shiming Xiang, Cheng-Lin Liu, and Chun-Hong Pan. Vehicle detection in satellite images by hybrid deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 11(10):1797–1801, 2014.

[17] D. C. Ciresan and Ueli Meier. Flexible, high performance convolutional neural networks for image classification. *Proceedings Of The Twenty-Second International Joint Conference On Artificial Intelligence*, pages 1237–1242, 2011.

[18] Dan C Cireşan, Ueli Meier, Jonathan Masci, Luca M Gambardella, and Jürgen Schmidhuber. Flexible, high performance convolutional neural networks for image classification. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Two*, pages 1237–1242. AAAI Press, 2011.

[19] John C. Curlander and Robert N. McDonough. *Synthetic aperture radar*, volume 199. Wiley New York, 1991.

[20] Alexandre Dalyac, Murray Shanahan, and Jack Kelly. Tackling class imbalance with deep convolutional neural networks. Technical report, Department of Computer Science, Imperial College London, 2014.

[21] Anderson de Andrade. Best practices for convolutional neural networks applied to object recognition in images. Technical report, University of Toronto, 2014.

[22] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

[23] Wolfgang Dierking. Sea ice monitoring by synthetic aperture radar. *Oceanography*, 26, 2013.

[24] Pedro Domingos. A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87, October 2012.

[25] E. Dumbill. *Strata 2012: Making Data Work*. O'Reilly, Santa Clara, CA O'Reilly, 2012.

[26] Environment and Climate Change Canada. Sea Ice Climatic Atlas for the East Coast 1981-2010. https://www.ec.gc.ca/glaces-ice/default.asp?lang=En&n=AE4A459A-1&offset=2&toc=show, 2015. [Online; accessed 20-March-2016].

[27] John C. Falkingham. Global satellite observation requirements for floating ice: focusing on synthetic aperture radar. Technical report, 2014.

[28] Dan Fequest. *MANICE: manual of standard procedures for observing and reporting ice conditions*. Environment Canada, 2002.

[29] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. November 2013.

[30] Alessandro Giusti, Dan C Cireşan, Jonathan Masci, Luca M Gambardella, and Jürgen Schmidhuber. Fast image scanning with deep max-pooling convolutional neural networks. *arXiv preprint arXiv:1302.1700*, 2013.

[31] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning*, volume 2. Springer, 2009.

[32] US Congressional Hearing. Strategic importance of the arctic in u.s. policy. Technical report, 2009.

[33] Geoffrey E Hinton. Learning multiple layers of representation. *Trends in Cognitive Sciences*, 11(10):428–434, 2007.

[34] Geoffrey E. Hinton, Li Deng, Dong Yu, et al. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine*, 29(6):82–97, November 2012.

[35] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, July 2012.

[36] Larry D. Hinzman, Neil D. Bettez, et al. Evidence and implications of recent climate change in northern Alaska and other Arctic regions. *Climatic Change*, 72(3):251–298, 2005.

[37] Christopher R Jackson, John R Apel, et al. *Synthetic aperture radar marine user's manual*. US Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite, Data, and Information Serve, Office of Research and Applications, 2004.

[38] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the ACM International Conference on Multimedia*, pages 675–678. ACM, 2014.

[39] Ola M. Johannessen, Vitaly Y. Alexandrov, Vitali Alexandrov, Ivan Y. Frolov, and Leonid P. Bobylev. *Remote sensing of sea ice in the Northern Sea Route*. Praxis Publishing Limited, 2007.

[40] Aranzazu Jurio, Humberto Bustince, Miguel Pagola, Pedro Couto, and Witold Pedrycz. New measures of homogeneity for image processing: an application to fingerprint segmentation. *Soft Computing*, 18(6):1055–1066, 2014.

[41] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1725–1732, June 2014.

[42] J. Karvonen, B. Cheng, T. Vihma, M. Arkett, and T. Carrieres. A method for sea ice thickness and concentration analysis based on SAR data and a thermodynamic model. *The Cryosphere*, 6:1507–1526, 2012.

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

[44] Juha Karvonen. Baltic sea ice concentration estimation based on C-band dual-polarized SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 52(9):5558–5566, 2014.

[45] Juha Karvonen. A sea ice concentration estimation algorithm utilizing radiometer and SAR data. *The Cryosphere*, 8(5):1639–1650, 2014.

[46] N.G. Kasapoglu. Sea ice concentration retrieval using composite scansar features in a sar data assimilation process. *Geoscience and Remote Sensing Letters, IEEE*, 11(12):2085–2089, 2014.

[47] V Komarov, S Wang, and J Tang. Permittivity and measurements. *Encyclopedia of RF and microwave engineering*.

[48] 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.

[49] H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin. Exploring strategies for training deep neural networks. *J Mach Learn Res*, 10, 2009.

[50] Yann LeCun, L. Bottou, GB Orr, and KR Müller. Efficient backprop. In *Neural Networks: Tricks of the Trade*. 2012.

[51] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[52] Yann LeCun et al. Generalization and network design strategies. *Connections in Perspective. North-Holland, Amsterdam*, pages 143–55, 1989.

[53] Yann LeCun, Fu Jie Huang, and Leon Bottou. Learning methods for generic object recognition with invariance to pose and lighting. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–97. IEEE, 2004.

[54] Yann LeCun, Koray Kavukcuoglu, Clément Farabet, et al. Convolutional networks and applications in vision. In *ISCAS*, pages 253–256, 2010.

[55] Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations.

In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 609–616. ACM, 2009.

[56] 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, September 2014.

[57] Hui Lin, Qing Xu, and Quanan Zheng. An overview on SAR measurements of sea surface wind. *Progress in Natural Science*, 18(8):913 – 919, 2008.

[58] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

[59] Marcus A. Maloof. Learning when data sets are imbalanced and when costs are unequal and unknown. In *Workshop on Learning from Imbalanced Data Sets II, ICML 2003*.

[60] J. Maslanik and J. Stroeve. DMSP SSM/I-SSMIS daily polar gridded brightness temperatures, version 4. http://nsidc.org/data/docs/daac/nsidc0001_ssmi_tbs.gd.html, 2012.

[61] Jonathan Masters. The Thawing Arctic: Risks and Opportunities . http://www.cfr.org/arctic/thawing-arctic-risks-opportunities/p32082, 2013. [Online; accessed 20-March-2016].

[62] Walter N Meier, Florence Fetterer, J Scott Stewart, and Sean Helfrich. How do sea-ice concentrations from operational data compare with passive microwave estimates? Implications for improved model evaluations and forecasting. *Annals of Glaciology*, 56(69):332–340, 2015.

[63] Volodymyr Mnih and Geoffrey E Hinton. Learning to detect roads in high-resolution aerial images. In *Computer Vision-ECCV 2010*, pages 210–223. Springer, 2010.

[64] Volodymyr Mnih and Geoffrey E Hinton. Learning to label aerial images from noisy data. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pages 567–574, 2012.

[65] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 807–814, 2010.

[66] Maryam M. Najafabadi, Flavio Villanustre, et al. Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1):1–21, 2015.

[67] National Center for Atmospheric Research Staff (Eds). The climate data guide: Sea ice concentration data: Overview, comparison table and graphs. https://climatedataguide.ucar.edu/climate-data/sea-ice-concentration-data-overview-comparison-table-and-graphs, 2013. [Online; accessed 20-March-2016].

[68] National Research Council. *Frontiers in Massive Data Analysis*. The National Academies Press, Washington, DC, 2013.

[69] Andre Ng. CS229 lecture notes: The EM algorithm. http://cs229.stanford.edu/notes/cs229-notes8.pdf, 2016. [Online; accessed 20-March-2016].

[70] Chris Oliver and Shaun Quegan. *Understanding synthetic aperture radar images*. SciTech Publishing, NC, USOverland, 2004.

[71] Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic. Learning and transferring mid-level image representations using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1717–1724, 2014.

[72] Kazuo Ouchi. Recent trend and advance of synthetic aperture radar with selected topics. *Remote Sensing*, 5(2), 2013.

[73] James E. Overland. Meteorology of the Beaufort Sea. *Journal of Geophysical Research: Oceans*, 114(C1):C00A07, 2009.

[74] George Papandreou, Liang-Chieh Chen, Kevin Murphy, and Alan L Yuille. Weakly- and semi-supervised learning of a dcnn for semantic image segmentation. *arXiv preprint arXiv:1502.02734*, 2015.

[75] George Papandreou, Iasonas Kokkinos, and Pierre-André Savalle. Untangling local and global deformations in deep convolutional networks for image classification and sliding window detection. *arXiv preprint arXiv:1412.0296*, 2014.

[76] Otavio Penatti, Keiller Nogueira, and Jefersson 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.

[77] Lutz Prechelt. Early stopping-but when? In *Neural Networks: Tricks of the Trade*, pages 53–67. Springer, 2012.

[78] Remote Sensing Systems, Santa Rosa, CA. AMSR2 / AMSRE. http://www.remss.com/missions/amsre. [Online; accessed 20-March-2016].

[79] R. K. Scharien, J. J. Yackel, D. G. Barber, M. Asplin, M. Gupta, and D. Isleifson. Geophysical controls on C band polarimetric backscatter from melt pond covered Arctic first-year sea ice: Assessment using high-resolution scatterometry. *Journal of Geophysical Research: Oceans*, 117(C9), 2012.

[80] Dominik Scherer, Andreas Müller, and Sven Behnke. Evaluation of pooling operations in convolutional architectures for object recognition. In *Artificial Neural Networks–ICANN 2010*, pages 92–101. Springer, 2010.

[81] Bernd Scheuchl, Dean Flett, Ron Caves, and Ian Cumming. Potential of RADARSAT-2 data for operational sea ice monitoring. *Canadian Journal of Remote Sensing*, 30(3):448–461, 2004.

[82] Frank Seide, Gang Li, and Dong Yu. Conversational speech transcription using context-dependent deep neural networks. In *Interspeech 2011*. International Speech Communication Association, August 2011.

[83] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*, 2013.

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

[85] Bob Slade. *RADARSAT-2 Product Description*, November 2009.

[86] G. Spreen, L. Kaleschke, and G. Heygster. Sea ice remote sensing using AMSR-E 89-GHz channels. *Journal of Geophysical Research*, 113(C2):C02S03, February 2008.

[87] Julienne Stroeve and Dirk Notz. Insights on past and future sea-ice evolution from combining observations and models. *Global and Planetary Change*, 135:119 – 132, 2015.

[88] David N. Thomas and Gerhard S. Dieckmann. *Sea ice.* Wiley-Blackwell, 2009.

[89] Dimitri Van De Ville, Mike Nachtegael, Dietrich Van der Weken, Etienne E Kerre, Wilfried Philips, and Ignace Lemahieu. Noise reduction by fuzzy image filtering. *Fuzzy Systems, IEEE Transactions on*, 11(4):429–436, 2003.

[90] Dietrich Van der Weken, Mike Nachtegael, and Etienne E Kerre. Using similarity measures and homogeneity for the comparison of images. *Image and Vision Computing*, 22(9):695–702, 2004.

[91] L. Wang, K. A. Scott, L. Xu, and D. 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*, PP(99):1–10, 2016.

[92] Lei Wang, K. Scott, and David A. Clausi. Automatic feature learning of SAR images for sea ice concentration estimation using feed-forward neural networks. In *2014 IEEE Geoscience and Remote Sensing Symposium*, pages 3969–3971. IEEE, July 2014.

[93] Joseph C. Watkins. Probability theory. *Course note for Probability Theory, University of Arizona*, 2006.

[94] F. J. Wentz and T. Meissner. AMSR-E ocean algorithms; supplement 1. Technical Report 051707, Remote Sensing Systems, Santa Rosa, CA., 2007.

[95] Iain H Woodhouse. *Introduction to microwave remote sensing*. CRC press, 2005.

[96] Huan Xu, Constantine Caramanis, and Shie Mannor. Robust regression and lasso. In *Advances in Neural Information Processing Systems*, pages 1801–1808, 2009.

[97] 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.

[98] Zhi-Hua Zhou and Xu-Ying Liu. Training cost-sensitive neural networks with methods addressing the class imbalance problem. *Knowledge and Data Engineering, IEEE Transactions on*, 18(1):63–77, 2006.