



## Article

# RUF: Effective Sea Ice Floe Segmentation Using End-to-End RES-UNET-CRF with Dual Loss

Anmol Sharan Nagi <sup>1,\*</sup>, Devinder Kumar <sup>2</sup>, Daniel Sola <sup>1</sup> and K. Andrea Scott <sup>1</sup>

<sup>1</sup> Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L G31, Canada; dsola@uwaterloo.ca (D.S.); ka3scott@uwaterloo.ca (K.A.S.)

<sup>2</sup> School of Medicine, Stanford University, Stanford, CA 94305, USA; devkumar@stanford.edu

\* Correspondence: asnagi@uwaterloo.ca

**Abstract:** Sea ice observations through satellite imaging have led to advancements in environmental research, ship navigation, and ice hazard forecasting in cold regions. Machine learning and, recently, deep learning techniques are being explored by various researchers to process vast amounts of Synthetic Aperture Radar (SAR) data for detecting potential hazards in navigational routes. Detection of hazards such as sea ice floes in Marginal Ice Zones (MIZs) is quite challenging as the floes are often embedded in a multiscale ice cover composed of ice filaments and eddies in addition to floes. This study proposes a segmentation model tailored for detecting ice floes in SAR images. The model exploits the advantages of both convolutional neural networks and convolutional conditional random field (Conv-CRF) in a combined manner. The residual UNET (RES-UNET) computes expressive features to generate coarse segmentation maps while the Conv-CRF exploits the spatial co-occurrence pairwise potentials along with the RES-UNET unary/segmentation maps to generate final predictions. The whole pipeline is trained end-to-end using a dual loss function. This dual loss function is composed of a weighted average of binary cross entropy and soft dice loss. The comparison of experimental results with the conventional segmentation networks such as UNET, DeepLabV3, and FCN-8 demonstrates the effectiveness of the proposed architecture.

**Keywords:** sea ice; ice floe; SAR; deep learning; conditional random fields; segmentation



**Citation:** Nagi, A.S.; Kumar, D.; Sola, D.; Scott, K.A. RUF: Effective Sea Ice Floe Segmentation Using End-to-End RES-UNET-CRF with Dual Loss. *Remote Sens.* **2021**, *13*, 2460. <https://doi.org/10.3390/rs13132460>

Academic Editor: Juha Karvonen

Received: 1 May 2021  
Accepted: 8 June 2021  
Published: 24 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Sea ice is one of the greatest physical constraints for shipping activities in the Arctic. Due to the lengthening of the open-water season, maritime traffic in the Arctic has increased three-fold over the past few years [1]. The story is similar in the Canadian Arctic as well, with Hudson Strait being the most traffic prone area [2]. However, although the ice extent and thickness are reducing across the Arctic, the risks and hazards involved in sailing through the Arctic are even more significant than in the past.

As the ice melts, the ice pack becomes more mobile, allowing hazards such as ice floes to break away. These ice floes move at high speeds in a dynamic fashion and can cause damage to vessels and man-made structures [3]. Ice charting is performed by the Canadian Ice Services (CIS) for estimating sea ice concentration and stage of development (which includes floes size information) to make management decisions for ensuring safety and efficient maritime activities in the Canadian Arctic [4]. CIS uses data from various data sources for the production of ice charts and the development of guidelines for mariners. One of the prominent data sources used for the task of sea ice monitoring is Synthetic Aperture Radar (SAR), which provides high spatial resolution images irrespective of the daylight conditions [5]. This data is manually interpreted, for which highly skilled personnel are required for the job.

To process the vast volumes of available data in real-time, there is a need for automated methods for ice floe detection. Previous studies [6–11] used image processing techniques

coupled with traditional machine learning on SAR images for the task of ice–water segmentation and ice floe separation. SAR images usually have intrinsic speckle noise due to the coherent nature of the imaging process [12]. The presence of this speckle-noise has been identified as a limitation in classification accuracy [6,11]. To circumvent this issue, instead of using SAR images, some studies [13–17] used data from vessel mounted cameras for identifying ice floes. This approach solves the issue of speckle-noise contamination. However, geometric error compensation is required to tackle the problem of underestimation/overestimation of ice cover caused due to oblique sensor placement [14]. Information from shipboard sensors is also biased in the sense that the ships prefer to transit through regions with lower ice concentration.

In this paper, we present an end-to-end ResUnet-CRF (RUF)-architecture-based model with a dual loss for ice floe segmentation in SAR images. From the perspective of deep learning, the proposed RUF architecture integrates three main modules: encoder–decoder framework [18], deep residual connections [19], and a probabilistic graphical model [20]. The encoder–decoder framework allows the network to learn the latent space representation of the data. Such networks have been shown to work with image noise for tasks such as image deblurring [21] and super-resolution [22]. Hence, such a network was chosen for the present study to aid in dealing with speckle noise typically present in SAR images. Residual connections ease the network training and the conditional random field aids in the refinement of segmentation boundaries. We train our proposed network architecture in an end-to-end manner using a dual loss function. The dual loss [23] is a combination of BCE and Dice loss.

Our main contributions are as follows:

1. We propose a novel encoder–decoder-based deep residual network embedded with a dense probabilistic graphical model for sea ice floe segmentation. To the best of our knowledge, this is the first time such a network has been successfully implemented in the domain of sea ice segmentation.
2. Passive microwave data does not provide precise information about sea ice concentration (SIC) in low SIC areas with small ice floes due to a coarse resolution and low instrument sensitivity. Our method successfully detects ice floes in SAR images, especially in the regions with less than 20% SIC, which could be important for marine hazard monitoring and wildlife management.
3. The proposed approach, RUF, is able to achieve higher metric scores along with visually superior results when compared with standard state of the art segmentation backbone models such as FCN-8 and DeepLabV3. These results have been achieved with fewer weights than other leading approaches, as our approach uses 26 M parameters, while FCN-8 and DeepLabV3 use 54 M and 60 M parameters, respectively.

This paper is organized as follows: after a brief literature review (Section 2), the description of the study area, image database, and data annotation are provided (Section 3). Next, detailed information regarding the various components of our proposed network architecture is presented in Section 4, followed by the description of evaluation metrics (Section 5) used in this paper. Section 6 provides information regarding the experimental setup, conducted experiments, and obtained results. Finally, the paper ends with the conclusion and future improvements in Section 7.

## 2. Background

Sea ice charting is typically conducted by national ice services to identify the boundaries between ice and open water, and to identify the dominant ice types and ice concentration for a given region. In the past years, due to the improvement in both aerial and remote sensing sensors, numerous studies using sea ice data have emerged [24]. These studies cover ice–water segmentation [25–27], ice concentration estimation [28–30], ice thickness estimation [31,32], ice type classification [33,34], and sea ice feature detection [35,36]. Methods using superpixel segmentation [37,38], watershed segmentation [39,40], and active contours [41,42] have been actively employed in SAR image segmentation. There are

several related studies in the area of ship detection in ice-covered waters [43–45]. The study in [43] used a novel approach, combining depthwise convolution and pointwise convolution to enable a lightweight and efficient network. This may be interesting to explore in future work. Many studies on ship detection use the SAR detection dataset (SSDD), which consists of quad-pol SAR imagery with spatial resolution of 1–15 m. The ships appear in these images as small bright regions. The study [45] also looked at a wide-swath Sentinel-1 image, which is comparable to the data source used here.

For the task of ice floe detection, various researchers have used different data platforms. Studies conducted by Hall et al. [13], Lu et al. [14], Heyn et al. [15,16], and Wang et al. [17] used vessel mounted camera sensors to obtain photographic data to identify ice floes. However, due to the oblique sensor placement, accurate measurements of sensor height, tilt, and focal length are required to calculate the geometric distortion. Moreover, compensation for ship sway is required for the success of these methods.

Images obtained from SAR provide a continuous stream of high spatial resolution data irrespective of the weather conditions and natural illumination. Earlier studies [6–11] aimed to solve the problem of ice floe identification in two steps. The first step involved the ice–water segmentation while the second step involved delineating different floes. Studies by Steer et al. and Toyota et al. [6,7] involved different thresholding methods for sea ice segmentation followed by morphological dilation/erosion operations to split different floes. Holt et al. [8] used local dynamic thresholding [46] and shrinking/growing algorithm [47] for floe segmentation. Hwang et al. [9] proposed a segmentation technique using Kernel Graph Cuts (KGC) [48] for ice–water segmentation and a combination of distance transformation, watershed [49], and a rule-based boundary revalidation processing for floe splitting [50]. Graphical models such as Markov Random Field and Conditional Random Field have also been used for the task of sea ice segmentation [10,11]. Due to the presence of speckle noise in SAR images, it can be difficult to segment sea ice floes using traditional machine learning techniques.

Recently, Convolutional Neural Networks (CNNs) have been proven to be good at learning the low- and high-level abstract features from raw images. Hence, they have been extensively used in tasks such as image classification [19,51–53], semantic segmentation [18,54], and object detection [55,56]. Long et al. [54] introduced a fully convolutional network for the task of image segmentation, while Chen et al. [57] proposed a combination of CNNs and CRFs to tackle poor localization of CNNs. Ronnerberger et al. [18] proposed an encoder–decoder-based network for the task of medical image segmentation. Later, Chen et al. proposed the DeepLab family networks [57–59] with dilated convolutions to reduce the computational complexity while maintaining the same receptive field.

Recently, Singh [60] et al. compared various segmentation models (e.g., DeepLab [57], UNet [18], SegNet [61], DenseNet [9]) for the task of river ice floe segmentation. Zhang et al. [62] introduced a convolutional network with dual attention streams for ice segmentation in rivers. Both of these studies use optical image datasets. To improve on pixel representations of CNNs and take advantage of residual learning, we propose RUF, an encoder–decoder network with residual blocks integrated with a convolutional CRF and trained in an end-to-end manner with dual loss function for the task of ice floe segmentation.

### 3. Dataset

The geographical area of interest for the dataset used in this paper spans over the Hudson Strait located in Eastern Canada, and its outflow into the Labrador Sea in the North Atlantic. The dataset is composed of 9 RADARSAT-2 C-band ScanSAR wide-beam mode images acquired in HH (horizontal transmit and horizontal receive) polarization. The images were acquired at a center frequency of 5.405 GHz with a 500-Km swath width and provide a nominal pixel spacing of 50 m. The SAR images were captured from the area as shown in Figure 1, with the red polygon describing the extent of one SAR

image. Information regarding the image acquisition dates, central latitude and longitude, and instances of annotated floes is given in Table 1.



**Figure 1.** The geographical area of interest of captured SAR scenes. Red polygon represents the geographic extent of a SAR image used in the study.

**Table 1.** Description of RADARSAT-2 scenes dataset that was used to train, validate, and test various model architectures.

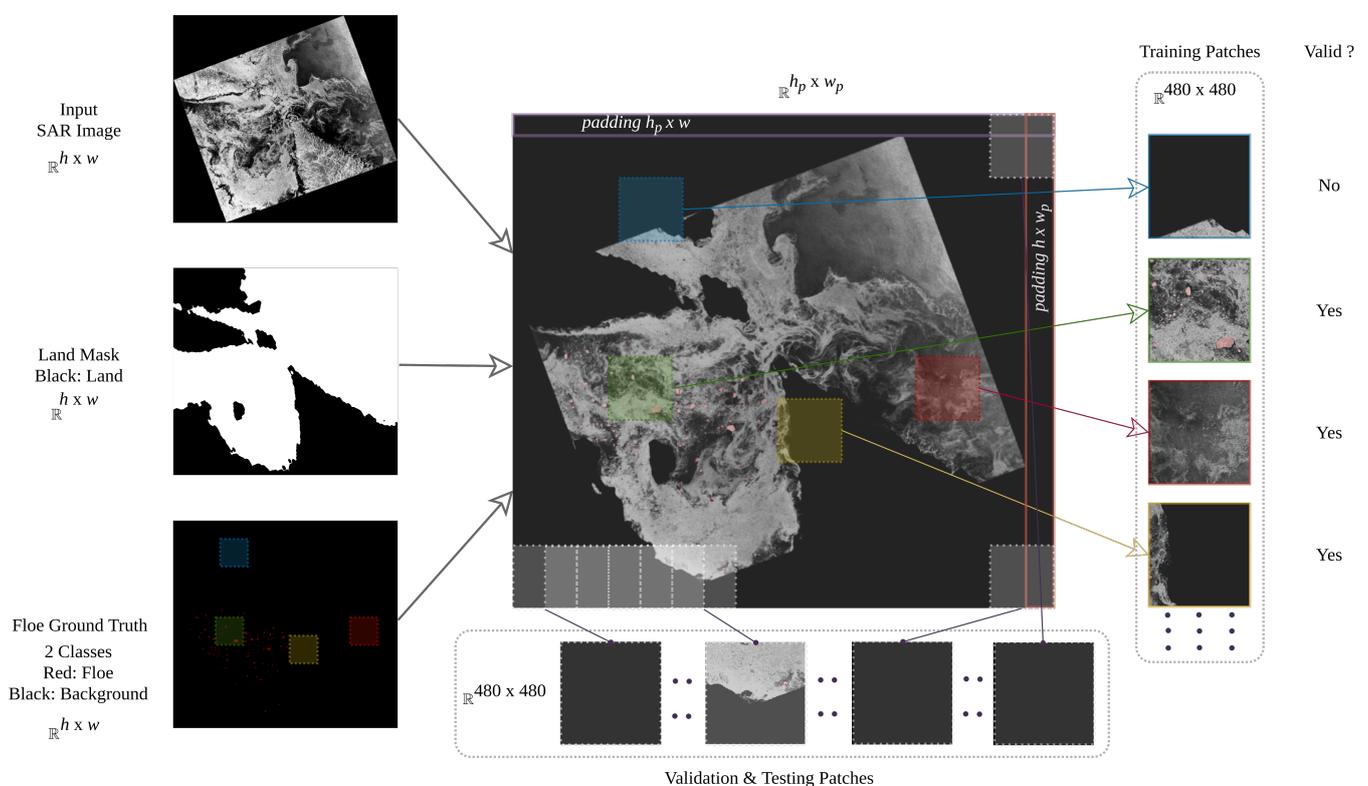
Dataset	Scene Acquisition		Number of Floes	Central Coordinate
	Date	Time		
Train	2008-11-13	11:29:00	204	63°32'40"N, 75°22'25"W
	2011-06-12	22:12:53	208	60°28'50"N, 68°47'22"W
	2012-06-05	22:42:59	177	62°38'24"N, 77°2'1"W
	2013-06-03	11:25:54	348	62°47'46"N, 72°38'1"W
	2013-06-14	22:34:35	203	62°38'14"N, 74°56'25"W
Total			1140	
Validation	2012-05-08	11:29:49	212	63°32'40"N, 75°22'25"W
	2012-06-06	10:44:18	128	60°42'8"N, 65°1'W
Total			340	
Test	2011-06-15	10:56:38	59	62°20'26"N, 67°29'54"W
	2014-06-15	11:29:40	288	63°30'41"N, 75°22'W
Total			347	

### 3.1. Data Preprocessing

SAR images have a grainy ‘salt and pepper’ appearance, also called speckle noise, which is caused due to random interference between coherent returns. To reduce this intrinsic contamination of speckle noise, the SAR images were downsampled four-fold. The downsampling operation was carried out by averaging over  $4 \times 4$  pixel nonoverlapping

blocks. Downsampling operation changed the nominal pixel spacing to 200 m with a reduction in data volume to 1/16 of the original. The local average filtering operation helps to reduce the speckle noise [63,64]. Downsampling is the result of this average filtering. The local average filtering also helps with reducing the data volume and makes it more manageable for training the neural network models.

Note that the Hudson Strait and the neighboring geographical areas of interest have a long coastline, and the images have significant land cover. As ice floes are a water body phenomenon, pixels representing land in the images were masked as black, as shown in Figure 2. To generate the land masks for the images, we applied a threshold of 0 m-elevation to the elevation masks and later a Gaussian blur with (5,5) kernel size to remove the rough edges along the sea-land boundaries. The elevation masks were generated using *ACE2\_5Min* digital elevation model with bilinear interpolation using the European Space Agency (ESA) toolbox, Sentinel Application Platform.



**Figure 2.** Patch generation for training, validation, and testing. Training: input SAR image ( $h \times w$ ) is land masked and patches of size  $480 \times 480$  are selected randomly from the image. For a patch to be considered as a valid training sample, it should not contain more than 50% black pixels. Validation and testing: land masked SAR image is padded using black pixels. Patches are drawn serially from the land masked SAR image with a patch overlap of 50%. Total training/validation patches extracted from an image are  $(h_p \times w_p) \div (480 \times 480)$ . Image acquired on date 2012-06-06 at time 10:44:18 along the Canadian east coast with central coordinate  $60^\circ 42' 8'' \text{N}$ ,  $65^\circ 1' \text{W}$ .

### 3.2. Data Annotation

The dataset contains 9 images with 1627 manual annotations of ice floes. For our experiments, a single class was defined for annotation, namely, 'floe', and all the remaining pixels were categorized as background. In the pixel-level annotations, we used the following criteria to annotate a closed contour as a floe:

- The contour contains consolidated ice as determined via visual inspection.
- At least 30% of the contour boundary is in contact with seawater.
- The contour contains at least 60 pixels.

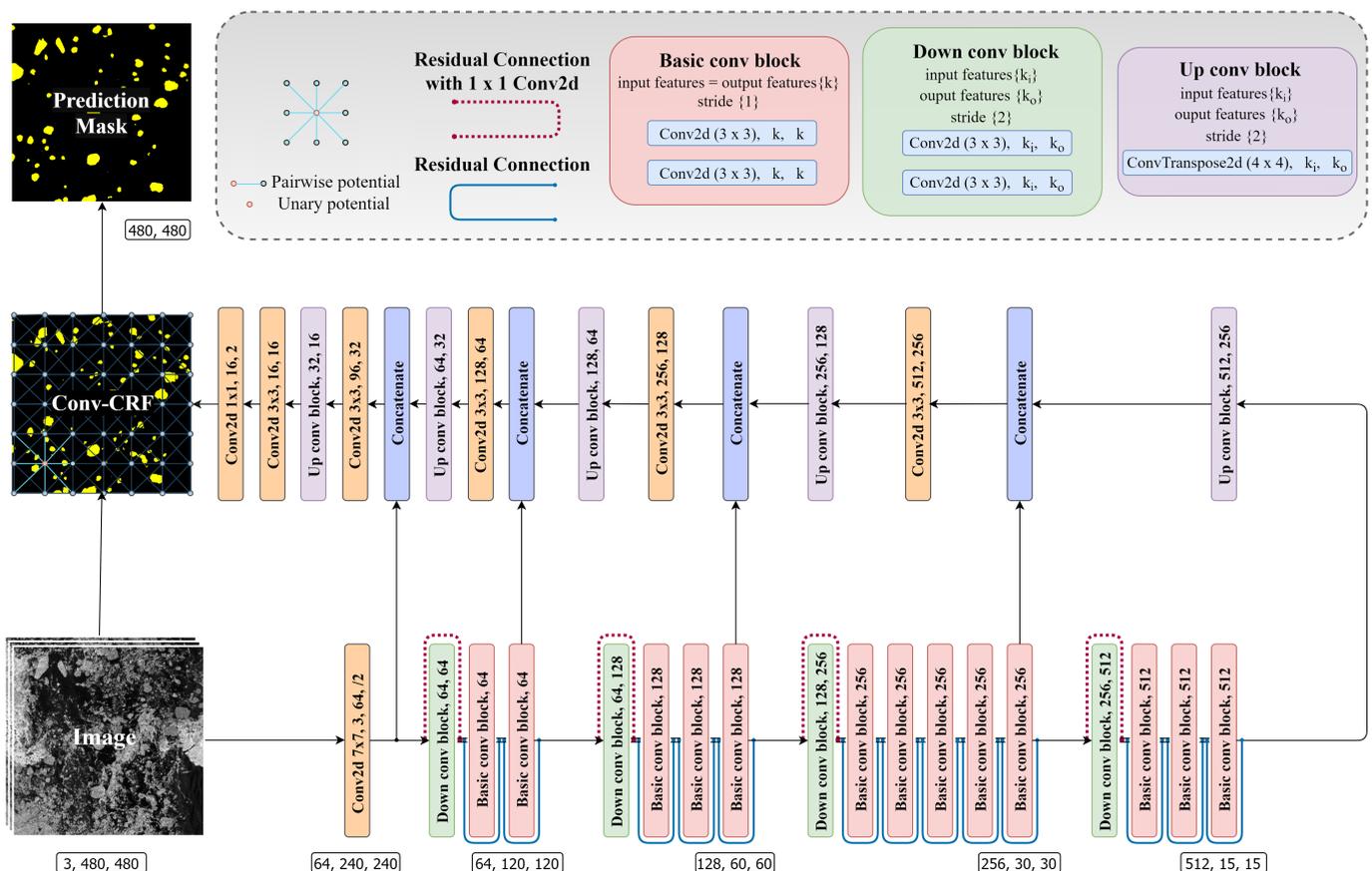
These criteria were chosen to eliminate the closed contours in ice covers (frozen in flocs) and to reduce noise artifacts. The flocs were annotated through visual inspection of the imagery using the Computer Vision Annotation Tool (CVAT).

### 3.3. Dataset Split

We split the dataset into the train, validation, and test sets with a rough 60:20:20 ratio split, such that the training set contains 60% of annotated flocs while validation and test sets each contain 20% of floc annotations. Full SAR images were placed in each set in order to have truly independent samples across different sets, as shown in Table 1. Splitting the dataset as such allows us to ascertain better generalization of the model.

## 4. Methodology

This section introduces the proposed RUF network architecture. The proposed model leverages the advantages of residual CNNs and Convolutional Conditional Random Fields (Conv-CRFs) [20] alongside a dual loss function. Residual blocks allow a network to train easily while the UNET skip connections help with easy information propagation between different layers of the network. To facilitate learning of the network weights, we jointly train the RES-UNET and Conv-CRF parts with the dual loss function. The overall architecture of the proposed network is illustrated in Figure 3. Information about various components of the network is given in the subsequent subsections.



**Figure 3.** Figure of the proposed network architecture. The lower branch of the architecture consists of the downsampling/encoding path. The upper branch of the architecture consists of the upsampling/decoding path. Skip connections (interconnections between the two parts) help with integrating the location information in the downsampling paths to the contextual information in the upsampling path. The output of the final layer (loosely labeled mask) in the upsampling path is fed to the Conv-CRF as unary. Conv-CRF processes the loosely labeled mask along with input image to generate the Prediction Mask.

#### 4.1. RES-UNET

##### 4.1.1. Residual Block

Neural network architectures with multiple layers, commonly known as deep neural networks, can learn richer features than their shallower counterparts [65,66]. Even though these deep neural network architectures are effective, they do struggle with a degradation problem where, upon adding more layers to an already deep neural network, the training accuracy of the model decreases. This is counter-intuitive since a deeper model should be able to fit the training data as well as a shallower model. To overcome this degradation problem, He et al. [19] proposed residual neural networks with stacked residual blocks. Given an input  $\mathbf{x}_l$  and output of the  $l$ -th residual unit  $\mathbf{x}_{l+1}$ , a residual block can be illustrated as

$$\begin{aligned} \mathbf{y}_l &= h(\mathbf{x}_l) + R(\mathbf{x}_l, w_l), \\ \mathbf{x}_{l+1} &= f(\mathbf{y}_l), \end{aligned} \quad (1)$$

where  $R(\cdot)$  is the residual function,  $f(\mathbf{y}_l)$  is an activation function, and  $h(\mathbf{x}_l)$  is an identity mapping function.

##### 4.1.2. UNET

The UNET [18] is a fully convolutional image segmentation architecture with symmetric downsampling and upsampling paths. To help with projecting the discriminatory features learned at different levels of the downsampling path, UNET uses skip connections. Skip connections help in integrating the location information in the downsampling paths to the contextual information in the upsampling paths. Rather than adding the input to the output in the case of residual blocks, a skip connection concatenates the input from the downsampling path to the output of the upsampling path.

#### 4.2. Convolutional Conditional Random Field: Conv-CRF

In a semantic segmentation task, the pixelwise predictions of the CNN models are prone to having inaccurate boundaries. To reduce inaccuracies in the boundary, global and contextual information models such as CRFs can be used in conjunction with the CNNs [57].

In the case of semantic segmentation, the label of each pixel is  $x_i \in \{1, \dots, X\}$ , where  $i$  is a pixel in image  $I$  with  $N$  pixels. In modern approaches, a fully connected CRF (FC-CRF) takes the CNN's output to compute the unary potentials [57,67]:  $\Psi_u(x_i) \in \mathbb{R}^X$ . Further, the pairwise potentials that account for the joint distribution of pixel pairs  $i, j$  as  $\Psi_p(x_i, x_j) = \mu(x_i, x_j)K(f_i, f_j)$ , where  $\mu(x_i, x_j)$  is a compatibility transformation such as in Potts model  $\mu(x_i, x_j) = [x_i \neq x_j]$ , and  $K$  is a kernel function such as the Gaussian kernel function.

Conv-CRFs [20] have an add-on assumption of conditional independence over the FC-CRFs. In the Conv-CRF model, two pixels  $i, j$  are considered conditionally independent when their L1 norm is greater than a threshold:  $d(i, j) > k$ , where  $d(\cdot)$  is the L1 norm and  $k$  is the distance threshold or the filter size. This means that the pairwise potential is zero for all the pixels at a distance greater than the threshold  $k$ . The Gibbs energy for a label sequence  $x$  can then be written as

$$E(x) = \sum_i^N \Psi_u(x_i) + \sum_{i=1}^N \sum_{j=1}^{k \times k} \Psi_p(x_i, x_j). \quad (2)$$

This greatly reduces the computational complexity. Teichmann et al. [20] also introduced a new message passing kernel that is similar to 2d convolutions of CNNs and can be efficiently implemented using convolutional libraries.

Efficient computations and exact message-passing lead to better run-time and performance when compared to FC-CRFs, which makes Conv-CRFs a better candidate for our architecture.

#### 4.3. End-To-End Training

For training the proposed architecture, we first feed the SAR images to the base RES-UNET network, where pixelwise segmentation maps from the base network are fed to the Conv-CRF network as the unary. The Conv-CRF network cleans up spurious predictions and enhances object boundary predictions. Training these two parts in an end-to-end manner allows the gradients to flow through the whole pipeline and enables both networks to learn simultaneously. Hence, with this approach, we optimize both models with respect to each other to provide optimum results.

#### 4.4. Dual Loss Function

For the task of multiclass classification, network weights are generally trained using the categorical cross entropy loss. In case of segmentation, losses involving ground truth and prediction overlap are generally employed. To train our network, we optimize the proposed RUF architecture using a weighted dual loss function including Binary Cross Entropy (BCE) loss and Soft Dice (SD) loss:

$$\mathcal{L}_{total} = \alpha \cdot \mathcal{L}_{BCE} + (1 - \alpha) \cdot \mathcal{L}_{Dice_s}, \quad (3)$$

where  $\alpha \in [0, 0.5, 1]$  is the weight parameter. BCE loss measures the classification accuracy of the model prediction and it increases as the prediction diverges from the ground truth [68]. SD loss, which is derived from Dice Coefficient, measures the similarity between two sets [69]. BCE loss and SD loss can be defined as Equations (4) and (5), respectively:

$$\mathcal{L}_{BCE} = - \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)], \quad (4)$$

$$\mathcal{L}_{Dice_s} = 1 - \frac{2 \sum_{i=1}^N y_i \hat{y}_i}{\sum_{i=1}^N \hat{y}_i + \sum_{i=1}^N y_i}, \quad (5)$$

where  $y_i$  is the label or ground truth and  $\hat{y}_i$  is the prediction for the  $i$ th pixel.

#### 4.5. RUF Architecture

RUF is a five-level deep convolutional network with symmetric downsampling and upsampling paths, as shown in Figure 3. The downsampling path encodes the image into a condensed representation while the upsampling path decodes this information into pixel-wise categorization. The downsampling path has four residual blocks. Each residual block contains multiple residual units built with two  $3 \times 3$  convolutional layers and a residual connection. The convolutional layers are accompanied by a BatchNorm2d layer with a ReLu activation function. Rather than employing a max-pooling operation to downsample the feature maps [18], we use down-convolutional blocks with strided convolutions. Max-pooling downsamples the feature maps by taking the maximum value in pooling window to represent the pixels in that window, whereas strided convolutions allow the network to summarize the pixels in that receptive field. Strided convolutions allow the network to learn the spatial relationships without losing localization accuracy, as in max-pooling when downsampling is performed multiple times. A bottleneck in the network forces the model to compress the information to learn useful features from the previous layers. The upsampling path has a similar structure to the downsampling path. Feature maps are upsampled at each level using up conv blocks employing transposed convolutions. Skip connections concatenate the output of each level in the downsampling path to the upsampling path and help in combining coarse information with finer information. At the final level, a  $1 \times 1$  convolution projects the multichannel feature maps to our intermediate

segmentation mask. This mask is then processed in conjunction with the input image to calculate the unary and pairwise potentials of the Conv-CRF for further refinement. A softmax operation is applied to the output of Conv-CRF, which is later thresholded at 50 percent confidence to obtain the prediction mask.

## 5. Metrics

The following metrics were used to evaluate the proposed approach and perform a comparison with other standard segmentation approaches.

1. Mean Intersection over Union (mIoU): mIoU is a popular metric that is often used for the task of semantic segmentation. mIoU is calculated by averaging the Jaccard Score ( $J$ ) over all the given classes in a segmentation task. Jaccard Score represents the ratio of the area of intersection between the ground truth ( $G$ ) and predicted segmentation ( $P$ ) maps to the union between ground truth and predicted segmentation maps:

$$J(G, P) = \frac{|G \cap P|}{|G \cup P|}, \quad (6)$$

where  $G$  and  $P$  are the ground truth and predicted segmentation maps, respectively.

2. Mean Pixel Accuracy (mPA): mPA is the ratio of correctly classified pixels per class, then averaged over the total number of classes. For  $k + 1$  classes ( $k$  foreground classes and one background):

$$mPA = \frac{1}{k + 1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}}. \quad (7)$$

3. F1 Score: F1 Score is the harmonic mean of Precision and Recall. Precision and Recall:

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN}, \quad (8)$$

where  $TP$  is the number true positives,  $FP$  is the number of false positives, and  $FN$  is the number of false negatives.

4. Dice Score: Dice Coefficient is the ratio of twice the area of intersection between ground truth and predicted segmentation maps, with the total number of pixels in the maps:

$$Dice = \frac{2|G \cap P|}{|G| + |P|}. \quad (9)$$

Both Dice Coefficient and IoU are positively correlated. In the case of binary segmentation, where foreground class is considered the positive class, F1 Score and Dice Coefficient are identical:

$$F1 = \frac{2|G \cap P|}{|G| + |P|} = Dice. \quad (10)$$

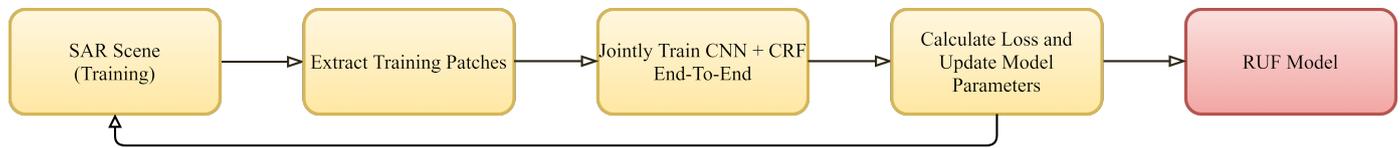
Hence, for our binary segmentation problem of ice floes, in this approach, we use Dice Coefficient as F1 Score hereafter.

## 6. Experiments and Results

### 6.1. Training Procedure

There are a total of 1140 annotated floes in the training set, which comprise approximately 60% of the overall annotated floes. Due to limited training images, we employ a random patch draw policy to train the model. An image is first randomly selected from the training set. We then randomly draw a patch of the given patch size from this selected image. Figure 2 illustrates the training patch selection process. A patch is considered a valid training sample if it does not contain more than 50% black area (due to image boundaries or land masking) or contains at least one floe either fully or partially. For example, a patch with 70% land and containing one floe is a valid sample. The above process is repeated until we find enough training samples for the training batch. We randomly rotate the eligible

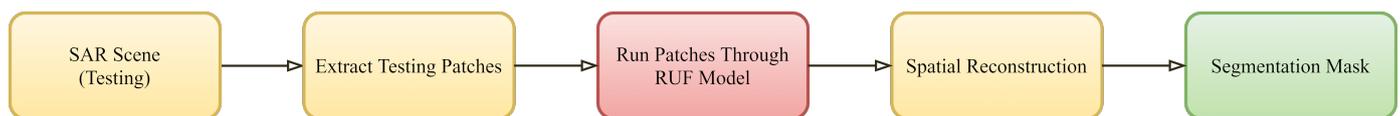
training samples by 0, 90, 180, or 270 degrees for data augmentation to increase the overall data available to train the models. This data augmentation provides a regularization effect that helps the models to generalize better on the overall dataset by reducing overfitting. The model training pipeline is illustrated in Figure 4.



**Figure 4.** Model training pipeline. Input SAR image is processed as illustrated in Figure 2 to extract patches for the training batch. These patches are fed to the model and the model parameters are updated according to the loss.

### 6.2. Validation and Testing Procedure

The validation and test sets contain 2 images each, which account for approximately 20% of the overall annotated floes. To validate and test the model, patches are extracted serially from the images with an overlap of 50%. Figure 2 provides details about selecting validation and testing patches. The model validation and testing pipeline is illustrated in Figure 5. The validation dataset is used to check if the model is overfitting while the test dataset is used to compare different models.



**Figure 5.** Model testing pipeline. Input SAR image is processed as illustrated in Figure 2 to feed patches serially to the trained RUF model for inference. These patches are then reconstructed to yield the segmentation mask for the whole image.

### 6.3. Setup

The code was implemented using Pytorch 1.3.1 and Torchvision 0.4.2 open source frameworks. For all experiments, model weights were initialized using kaiming uniform initialization [70]. To optimize the model parameters further, we used the ADAM optimizer at an initial  $1 \times 10^{-4}$  learning rate. The Conv-CRF was initialized using default parameters from [20,71] with an exception, We removed the Gaussian blur as the SAR images in our dataset are already downsampled with 4-fold averaging.

### 6.4. UNET Backbone-Selection

To select the primary segmentation network for our pipeline, we first compared the UNET architecture with different backbone UNETs. The key difference between a UNET and a backbone-UNET is that we replace the two convolutional layers, and the  $2 \times 2$  max-pooling operations for each level in the downsampling path are replaced with the different convolutional blocks of the backbone architecture, while the skip connections and the convolutional layers in the upsampling path remain the same. Information regarding the number of parameters and comparison between different UNET architectures is given in Table 2. We use VGG19, Inception V3, and ResNet34 to construct different UNET architectures. We observe that UNET architecture with a ResNet34-based encoder achieves the best scores on the validation set and is selected for further improvement. Refer to Table 2 for more information.

**Table 2.** Comparison of different UNET backbones on validation set. Best results are in bold.

Model	Parameters	mIOU	F1	mPA
UNET	7,788,545	70.50	58.77	74.83
VGG-UNET	20,267,281	71.52	60.71	73.62
Inception-UNET	27,161,264	71.59	60.89	74.68
RES-UNET	25,658,402	<b>72.57</b>	<b>62.88</b>	<b>80.40</b>

**Table 3.** Comparison of different RUF joint training approaches on validation set. Best results are in bold.

Method	mIoU	F1	mPA
Decoupled	68.25	54.23	72.90
Stepwise	69.73	57.27	74.87
End-to-end	<b>75.07</b>	<b>67.28</b>	<b>81.83</b>

### 6.5. Joint Training with Conv-CRF

With the UNET backbone selected, we examine different configurations of joint training. We experiment with decoupled learning, stepwise learning, and end-to-end learning approaches. The results of which are illustrated in Table 3. Decoupled learning is based on the assumption that CRF needs an accurate prediction of the unary [71] to learn efficiently. In this approach, initially, only the CNN model is training till loss convergence, i.e., when the loss begins to saturate. After this step, the CRF is trained as a standalone model with the CNN output as unary to the CRF. The gradients never flow through the whole model in an iteration. Stepwise learning is similar to decoupled learning but in the second step, both the CNN and CRF are trained jointly such that weights of the whole model are updated in the second step. End-to-end learning involves training the whole architecture jointly from the very beginning such that the gradients flow through the whole network from the first epoch.

**Table 4.** Comparison of different loss functions for various model architectures using end-to-end training approach on validation set. BCE—Binary cross entropy loss; Dice—Dice loss; BCE + Dice—Equally weighted BCE and Dice Loss. Best results are in bold.

Model	Loss Function								
	BCE			Dice			Dice + BCE		
	mIoU	F1	mPA	mIoU	F1	mPA	mIoU	F1	mPA
UNET	65.76	48.66	66.39	65.64	48.61	71.96	70.51	58.78	74.83
RES-UNET	70.83	59.40	74.41	70.32	58.44	75.41	72.57	62.88	80.40
RUF ( <i>Ours</i> )	70.91	59.59	75.36	71.39	60.48	74.61	<b>75.07</b>	<b>67.28</b>	<b>81.83</b>

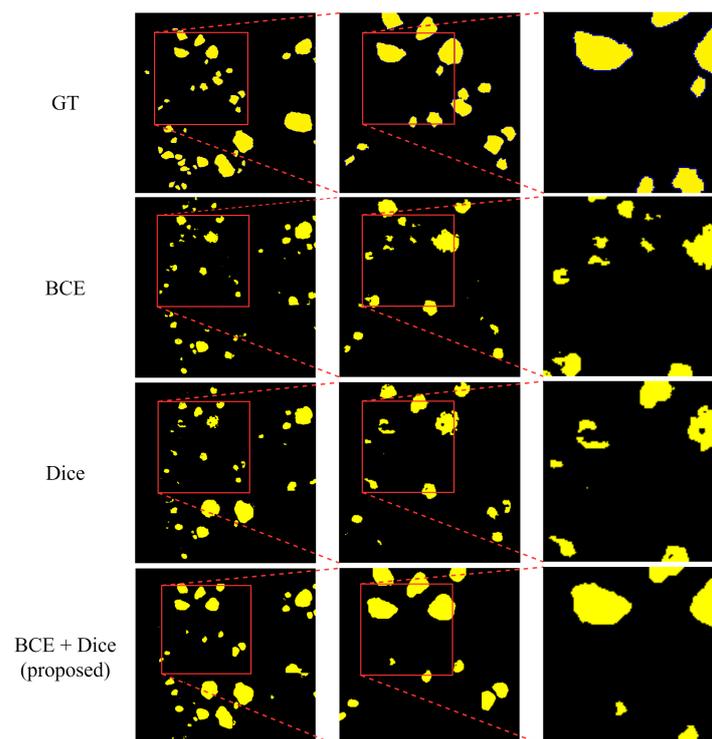
We observed that for our task, the end-to-end learning approach was more stable and yielded preminent results. The main reason for the success of this approach is that when the CNN and CRF are trained together, they are able to coadapt easily with respect to each other. Thus, the end-to-end learning approach was used for further experiments.

### 6.6. Dual Loss Function Selection

To select the optimum weights of our dual loss function, we trained different RUF models using various  $\alpha$  values for  $\mathcal{L}_{total}$  in Equation (3). We evaluated three different configurations with  $\alpha \in [0, 0.5, 1]$ , which yield only BCE loss, equally weighted BCE and Dice loss, and only Dice loss, respectively. The results of our experiments are given in Table 4. We observe that using a dual loss comprising of equally weighted BCE and Dice loss helps the network to generate better predictions.

We observed that training with either only BCE or Dice loss did help the model loss to converge quickly but the predictions from such models lack discrete boundaries,

as illustrated in Figure 6. BCE is a local loss that accounts for the correctness of individual pixels and is biased towards the majority class (background). Dice loss deals with the problem of unbalanced datasets by design, such that the network is disincentivized to ignore the minority class. But Dice loss can have high variance over batches as all object instances are given the same weight irrespective of the object size. As Dice loss focuses on the similarity of two sets and BCE loss is the sum of the distance between the corresponding pixels in ground truth and prediction mask, combining them allows the network to optimize the loss in both image-level and pixel-level domains. We observe that models trained using dual loss yield better metrics and visual predictions than those without, irrespective of using the CRF component, as in Figure 6.



**Figure 6.** Comparison between the predictions of RUF model trained using BCE, Dice, and BCE + Dice loss. It can be observed that the predictions obtained using BCE + Dice loss have discrete continuous boundaries similar to the ground truth (GT).

### 6.7. Patch Size Selection

As the overall segmentation network configuration is selected, we investigate further to choose the optimum patch size to train our model. We trained various RUF models beginning with a patch size of  $96 \times 96$  pixels and a batch size of 64 patches. We observed that this model was able to identify small floes, while it struggled to detect floes that appear partially in the patch. To overcome this issue, we decided to increase the patch size until we obtained optimum results. To train the subsequent models with bigger patch sizes, we decreased the batch size to obtain optimum GPU memory utilization. Results of our experiment are given in Table 5.

**Table 5.** Comparison of different models for various patch sizes on validation set. Best results are in bold.

Patch Size	Batch Size	mIoU	F1	mPA
96	64	70.53	58.81	74.09
192	32	70.77	59.38	77.65
288	16	70.75	59.24	74.16
384	12	71.99	59.77	74.40
480	8	<b>75.07</b>	<b>67.28</b>	<b>81.83</b>
576	6	69.85	57.44	72.13

We observed a direct correlation between a bigger patch size and the segmentation accuracy, as evident from increasing metrics in Table 5—with an exception to the patch size of  $576 \times 576$  pixels, where a considerable dip in performance was observed. To investigate this issue, we trained models with patch sizes  $432 \times 432$  and  $528 \times 528$  with various batch sizes. The results of this experiment are presented in Table 6. We observed that when the training patch size increases from 384 at 12 batch size to 432 at 10 batch size, the model's performance increases. However, when the 432-sized patch's batch size is decreased to 8, a dip in performance is observed. When we changed the training patch size from 480 at 8 batch size to 528 at 7 batch size, we observed that the model's performance decreases, although the patch size has increased. A similar trend has also been discussed in [72], such that increasing the training batch size helps the model to learn and converge quickly while yielding better results. Due to the GPU memory constraint, a trade-off between patch size and batch size had to be made in order to obtain an optimum patch and batch size.

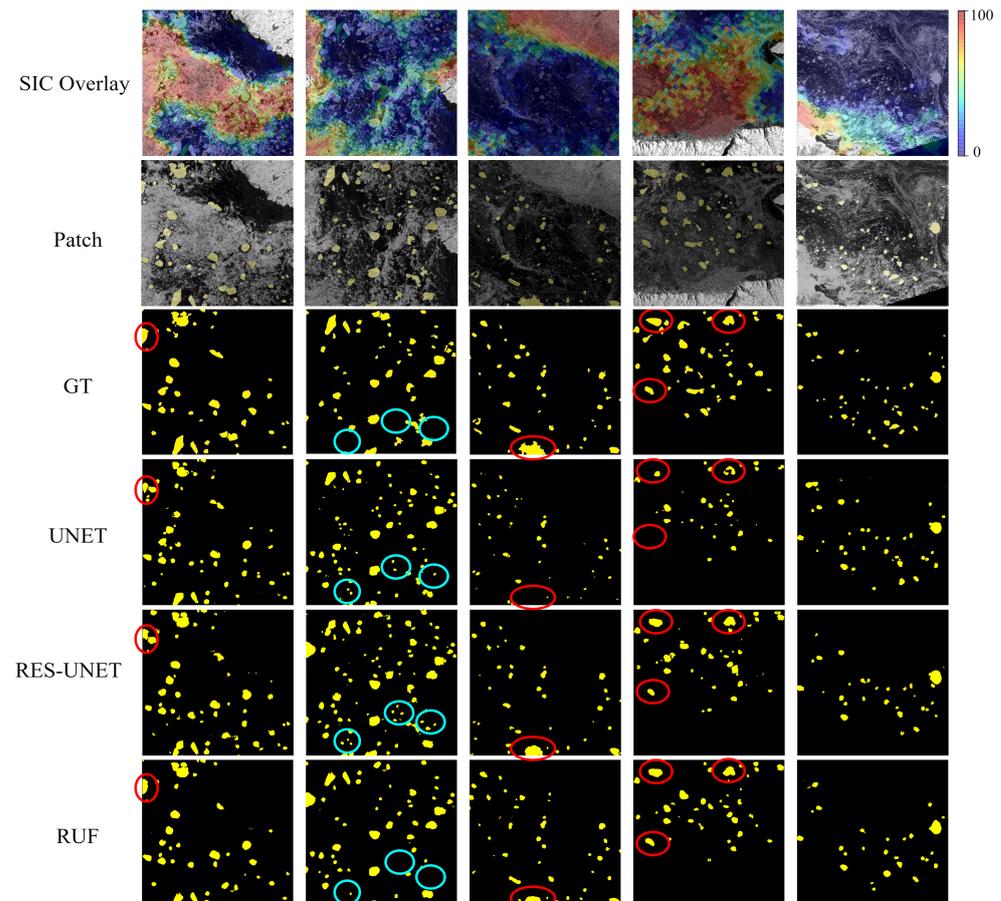
**Table 6.** Comparison of patch size and batch size on validation set. Best results are in bold.

Patch Size	Batch Size	mIoU	F1	mPA
384	12	71.99	59.77	74.40
432	10	73.77	63.98	78.38
432	8	71.87	60.18	74.82
480	8	<b>75.07</b>	<b>67.28</b>	<b>81.83</b>
528	7	73.65	63.74	77.87
528	6	69.12	56.97	71.65
576	6	69.85	57.44	72.13

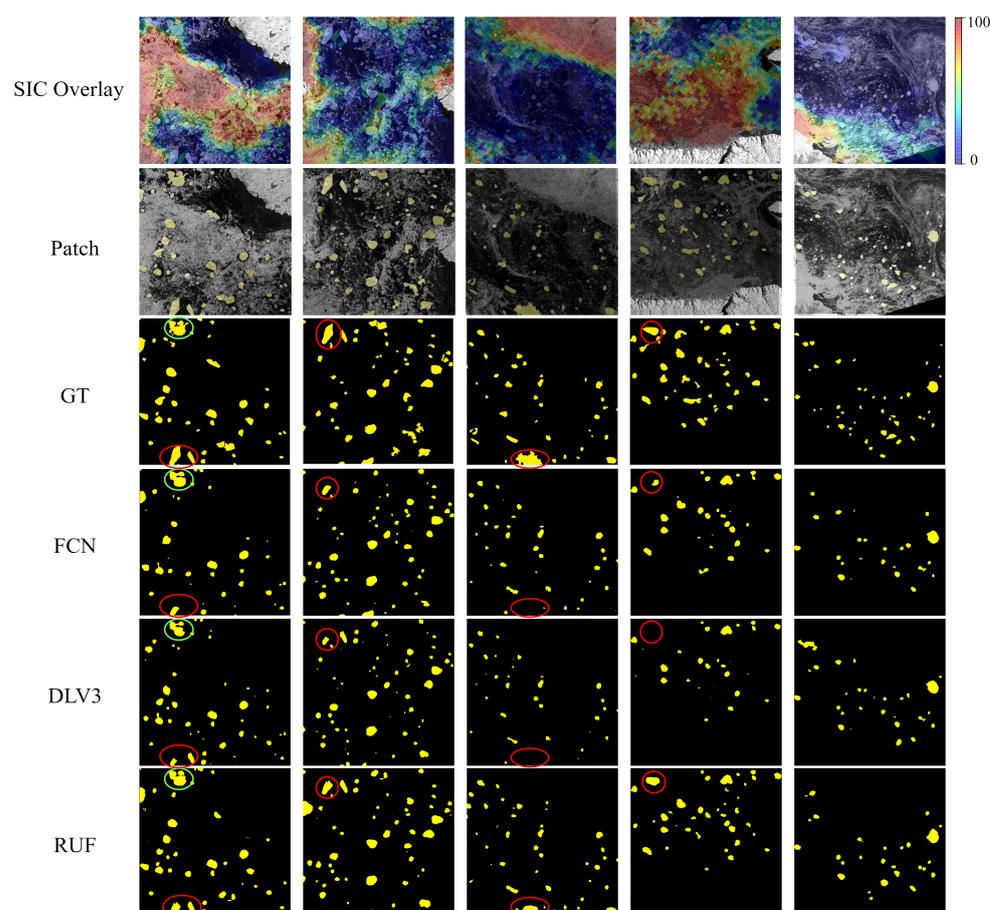
### 6.8. Comparison with Other Models

Using the optimal input size and other network parameters, we obtained the best results using our proposed method. We then compared our method with several other standard segmentation models that are typically used in SAR image segmentation. Test set images are used in this comparison. To gauge the iterative improvement component of our method, we plot the predictions of UNET, RES-UNET, and RUF together in Figure 7. Upon initial observation, we notice that both UNET and RES-UNET generate a greater number of false-positive predictions compared to RUF. We observe that RUF is able to detect floes of various shapes and sizes, while UNET struggles to detect bigger floes. UNET is also unable to delineate the floes in close proximity to each other. The main reason for this issue would be the absence of residual connections in the UNET, as the model is unable to learn finer features compared to other more complex models. It can also be observed that both UNET and RES-UNET generate multiple broken boundary predictions where a single floe is segmented as multiple floes.

We also compare RUF with other standard segmentation models such as FCN-8 (FCN) and DeepLabV3 (DLV3), the predictions of which are illustrated in Figure 8. We can observe that the predictions from both FCN and DeepLab do not suffer from false-positives; however, they do struggle to detect bigger floes and delineate floes in close proximity. The absence of a dense component such as a conditional random field to enhance finer segmentation details would explain this issue. Apart from these qualitative observations, the superiority of RUF for the task of ice floe segmentation is also evident from the quantitative analysis, as given in Table 7. Zoomed-in segmentation results of a patch are given in Figure 9.



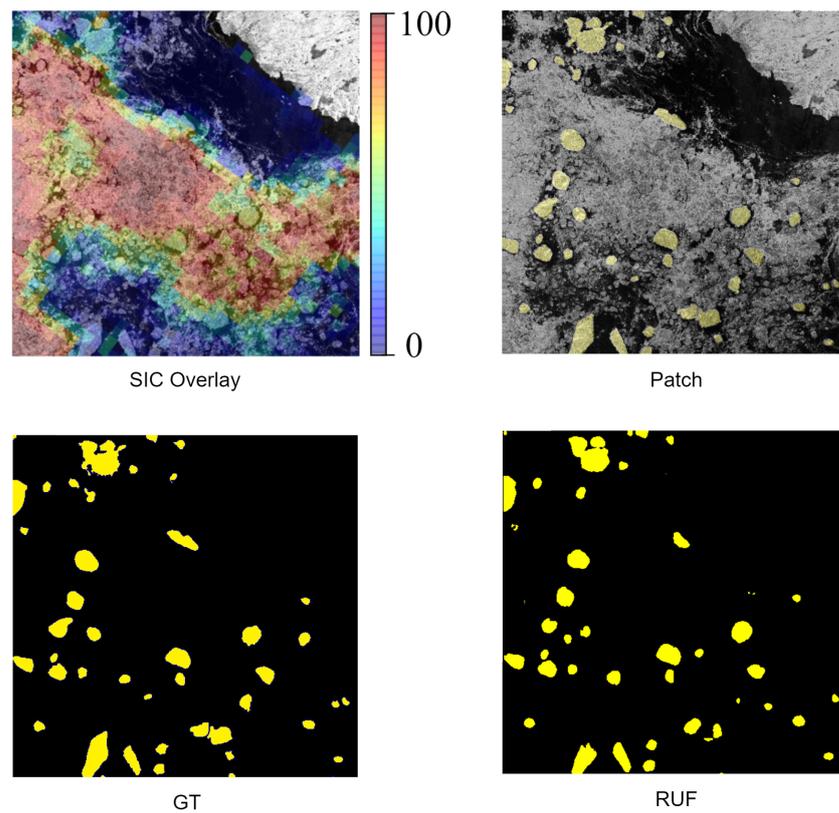
**Figure 7.** Comparison between the predictions of UNET, RES-UNET, and RUF for the task of ice floe segmentation on patches from the test set. Red ovals highlight segmentation results where both UNET and RES-UNET are unable to properly segment a fully or partially visible ice floe. Blue ovals highlight segmentation results where both UNET and RES-UNET produced many false-positive predictions. It can be observed that the proposed RUF architecture is able to produce finer segmentation results compared to other backbone architectures on our dataset. Here, Patch and GT denote original image patch input and ground truth, respectively. The sea ice concentration (SIC) masks are overlaid on the respective patches for comparison between sea ice as continuous (passive microwave data) vs. discrete (processed ice floe masks from SAR) information.



**Figure 8.** Comparison between the predictions of FCN-8, DeepLabV3, and RUF for the task of ice floe segmentation on patches from the test set. Red ovals highlight segmentation results where both FCN and DLV3 are unable to properly segment a fully or partially visible ice floe. Green ovals highlight segmentation results where both FCN and DLV3 are unable to delineate floes in close proximity. It can be observed that the proposed RUF architecture is able to produce finer segmentation results compared to the other frequently used segmentation methods in satellite imaging on our dataset. Here, Patch and GT denote original image patch input and ground truth, respectively. The sea ice concentration (SIC) masks are overlaid on the respective patches for comparison between sea ice as continuous (passive microwave data) vs. discrete (processed ice floe masks from SAR) information.

**Table 7.** Comparison between various segmentation architectures on test set. Best results are in bold.

Model	Parameters	mIoU	F1	mPA
UNET	7,788,545	69.20	56.54	71.87
RES-UNET	25,658,402	71.76	61.24	74.66
FCN8	54,304,086	69.88	57.48	72.24
DLV3	60,991,062	67.47	52.43	68.94
RUF (ours)	26,119,208	<b>74.52</b>	<b>66.26</b>	<b>77.88</b>



**Figure 9.** Zoom-in on the segmentation result of a patch. Here, Patch and GT denote original image patch input and ground truth, respectively, where GT are the floes that were annotated manually according to our selection criteria. The sea ice concentration (SIC) mask is overlaid on the patch for comparison between sea ice as a continuous (passive microwave data) vs. discrete (processed ice floe masks from SAR) information.

## 7. Conclusions

In this paper, we presented a sea ice floe segmentation network for SAR images. RUF is introduced as a Conv-CRF embedded residual UNET, end-to-end trained using a dual loss function. With our model, we demonstrate that we can fully exploit the usefulness of Conv-CRF for the task of segmentation by integrating it inside a deep learning algorithm and training it in an end-to-end fashion. We performed extensive analysis for selecting model parameters to ensure the reliability of the proposed model. The experimental results confirm that our proposed model can achieve higher scores on all metrics and visually superior results that show a reduced number of false alarms and better representation of floe shape than other approaches on the limited dataset used in the study. It is noteworthy that this has been achieved with fewer weights than other leading approaches. We observed that in some instances, RUF was unable to detect floes of very large sizes. These floes were comparable to the prediction window/patch size. Upon investigation, we noticed that in our images, there were a greater number of mid-sized and small-sized floes compared to large-sized floes. In the future, we aim to use a much larger dataset over a wider geographic region to include a greater variety of floes. For example, including marginal ice zones, such as those of Antarctica, Baffin Bay, and the Beaufort Sea, would allow a combination of ice floes and larger icebergs in addition to multiyear ice floes. We would like to remove the 4x4 downsample and use full-resolution images. This may require more lightweight architectures, such as EfficientNet [73]. Our end goal is to provide floes from SAR that can be used for various purposes, such as the validation of sea ice concentration from passive microwave data, to provide information for wildlife habitat studies, and for studies of wave–ice interactions.

**Author Contributions:** A.S.N., D.K. and K.A.S. conceived and designed the framework of the study. A.S.N. and D.S. completed the data collection and processing. A.S.N. and D.K. completed the algorithm design and the data analysis. A.S.N. completed the experiments. A.S.N. was the lead author of the manuscript with contributions by D.K., D.S. and K.A.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Canadian ArcticNet program and MEOPAR the Marine Environmental Observation Prediction and Response Network.

**Data Availability Statement:** The data is not publicly available due to 3rd party data policy.

**Acknowledgments:** The RADARSAT-2 data used in this paper was provided by the Canadian Ice Service (CIS). All SAR images are copyrighted to MacDONALD, Dettwiler and Associates Ltd. (2010)-All Rights Reserved. RADARSAT is an official mark of the Canadian Space Agency.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Barnhart, K.R.; Miller, C.R.; Overeem, I.; Kay, J.E. Mapping the future expansion of Arctic open water. *Nat. Clim. Chang.* **2016**, *6*, 280–285. [[CrossRef](#)]
2. Andrews, J.; Babb, D.; Barber, D. Climate change and sea ice: Shipping in Hudson Bay, Hudson Strait, and Foxe Basin (1980–2016). *Elem. Sci. Anth.* **2018**, *6*, 19. [[CrossRef](#)]
3. Barber, D.; Babb, D.; Ehn, J.; Chan, W.; Matthes, L.; Dalman, L.; Campbell, Y.; Harasyn, M.; Firoozy, N.; Theriault, N.; et al. Increasing mobility of high Arctic sea ice increases marine hazards off the east coast of Newfoundland. *Geophys. Res. Lett.* **2018**, *45*, 2370–2379. [[CrossRef](#)]
4. Scheuchl, B.; Flett, D.; Caves, R.; Cumming, I. Potential of RADARSAT-2 data for operational sea ice monitoring. *Can. J. Remote Sens.* **2004**, *30*, 448–461. [[CrossRef](#)]
5. Smirnov, V.; Bychkova, I. Satellite monitoring of ice features to ensure safety of offshore operations in the Arctic seas. *Izv. Atmos. Ocean. Phys.* **2015**, *51*, 935–942. [[CrossRef](#)]
6. Steer, A.; Worby, A.; Heil, P. Observed changes in sea-ice floe size distribution during early summer in the western Weddell Sea. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **2008**, *55*, 933–942. [[CrossRef](#)]
7. Toyota, T.; Haas, C.; Tamura, T. Size distribution and shape properties of relatively small sea-ice floes in the Antarctic marginal ice zone in late winter. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **2011**, *58*, 1182–1193. [[CrossRef](#)]
8. Holt, B.; Martin, S. The effect of a storm on the 1992 summer sea ice cover of the Beaufort, Chukchi, and East Siberian Seas. *J. Geophys. Res. Ocean.* **2001**, *106*, 1017–1032. [[CrossRef](#)]
9. Hwang, B.; Ren, J.; McCormack, S.; Berry, C.; Ayed, I.B.; Graber, H.C.; Aptoula, E. A practical algorithm for the retrieval of floe size distribution of Arctic sea ice from high-resolution satellite Synthetic Aperture Radar imagery. *Elem. Sci. Anth.* **2017**, *5*, 38. [[CrossRef](#)]
10. Clausi, D.A.; Yue, B. Comparing cooccurrence probabilities and Markov random fields for texture analysis of SAR sea ice imagery. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 215–228. [[CrossRef](#)]
11. Zhu, T.; Li, F.; Heygster, G.; Zhang, S. Antarctic Sea-Ice Classification Based on Conditional Random Fields From RADARSAT-2 Dual-Polarization Satellite Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 2451–2467. [[CrossRef](#)]
12. Cooke, C.L.V.; Scott, K.A. Estimating Sea Ice Concentration From SAR: Training Convolutional Neural Networks With Passive Microwave Data. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4735–4747. [[CrossRef](#)]
13. Hall, R.J.; Hughes, N.; Wadhams, P. A systematic method of obtaining ice concentration measurements from ship-based observations. *Cold Reg. Sci. Technol.* **2002**, *34*, 97–102. [[CrossRef](#)]
14. Lu, P.; Li, Z. A method of obtaining ice concentration and floe size from shipboard oblique sea ice images. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2771–2780.
15. Heyn, H.M.; Blanke, M.; Skjetne, R. Ice condition assessment using onboard accelerometers and statistical change detection. *IEEE J. Ocean. Eng.* **2019**, *45*, 898–914. [[CrossRef](#)]
16. Heyn, H.M.; Knoche, M.; Zhang, Q.; Skjetne, R. A system for automated vision-based sea-ice concentration detection and floe-size distribution indication from an icebreaker. In Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering, Trondheim, Norway, 25–30 June 2017; American Society of Mechanical Engineers: New York, NY, USA, 2017; Volume 57762, p. V008T07A012.
17. Wang, Q.; Li, Z.; Lu, P.; Lei, R.; Cheng, B. 2014 summer Arctic sea ice thickness and concentration from shipborne observations. *Int. J. Digit. Earth* **2019**, *12*, 931–947. [[CrossRef](#)]
18. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
19. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

20. Teichmann, M.T.T.; Cipolla, R. Convolutional CRFs for Semantic Segmentation. *arXiv* **2018**, arXiv:cs.CV/1805.04777.
21. Chen, G.; Gao, Z.; Wang, Q.; Luo, Q. U-net like deep autoencoders for deblurring atmospheric turbulence. *J. Electron. Imaging* **2019**, *28*, 053024. [[CrossRef](#)]
22. Hu, X.; Naiel, M.A.; Wong, A.; Lamm, M.; Fieguth, P. RUNet: A Robust UNet Architecture for Image Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019.
23. Lei, Y.; Tian, S.; He, X.; Wang, T.; Wang, B.; Patel, P.; Jani, A.B.; Mao, H.; Curran, W.J.; Liu, T.; et al. Ultrasound prostate segmentation based on multidirectional deeply supervised V-Net. *Med. Phys.* **2019**, *46*, 3194–3206. [[CrossRef](#)]
24. Haugen, J.; Imsland, L.; Løset, S.; Skjetne, R. Ice observer system for ice management operations. In Proceedings of the Twenty-First International Offshore and Polar Engineering Conference, Maui, HI, USA, 19–24 June 2011.
25. Leigh, S.; Wang, Z.; Clausi, D.A. Automated ice–water classification using dual polarization SAR satellite imagery. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 5529–5539. [[CrossRef](#)]
26. Hoekstra, M.; Jiang, M.; Clausi, D.A.; Duguay, C. Lake Ice-Water Classification of RADARSAT-2 Images by Integrating IRGS Segmentation with Pixel-Based Random Forest Labeling. *Remote Sens.* **2020**, *12*, 1425. [[CrossRef](#)]
27. Xie, T.; Perrie, W.; Wei, C.; Zhao, L. Discrimination of open water from sea ice in the Labrador Sea using quad-polarized synthetic aperture radar. *Remote Sens. Environ.* **2020**, *247*, 111948. [[CrossRef](#)]
28. Wang, L.; Scott, K.A.; Xu, L.; Clausi, D.A. Sea ice concentration estimation during melt from dual-pol SAR scenes using deep convolutional neural networks: A case study. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4524–4533. [[CrossRef](#)]
29. Cheng, A.; Casati, B.; Tivy, A.; Zagon, T.; Lemieux, J.F.; Tremblay, L.B. Accuracy and inter-analyst agreement of visually estimated sea ice concentrations in Canadian Ice Service ice charts using single-polarization RADARSAT-2. *Cryosphere* **2020**, *14*, 1289–1310. [[CrossRef](#)]
30. Karvonen, J. Baltic sea ice concentration estimation using SENTINEL-1 SAR and AMSR2 microwave radiometer data. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2871–2883. [[CrossRef](#)]
31. Shi, L.; Karvonen, J.; Cheng, B.; Vihma, T.; Lin, M.; Liu, Y.; Wang, Q.; Jia, Y. Sea ice thickness retrieval from SAR imagery over Bohai sea. In Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014; pp. 4864–4867.
32. Karvonen, J.; Shi, L.; Cheng, B.; Similä, M.; Mäkynen, M.; Vihma, T. Bohai Sea ice parameter estimation based on thermodynamic ice model and Earth observation data. *Remote Sens.* **2017**, *9*, 234. [[CrossRef](#)]
33. Zakhvatkina, N.Y.; Alexandrov, V.Y.; Johannessen, O.M.; Sandven, S.; Frolov, I.Y. Classification of sea ice types in ENVISAT synthetic aperture radar images. *IEEE Trans. Geosci. Remote Sens.* **2012**, *51*, 2587–2600. [[CrossRef](#)]
34. Park, J.W.; Korosov, A.A.; Babiker, M.; Won, J.S.; Hansen, M.W.; Kim, H.C. Classification of Sea Ice Types in Sentinel-1 SAR images. *Cryosphere Discuss.* **2019**, *2019*, 1–23.
35. Nagi, A.S.; Minhas, M.S.; Xu, L.; Scott, K.A. A Multi-Scale Technique to Detect Marginal Ice Zones Using Convolutional Neural Networks. In Proceedings of the IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 3035–3038.
36. Sola, D.; Nagi, A.S.; Scott, K.A. Identifying Sea Ice Ridging in SAR Imagery Using Convolutional Neural Networks. In Proceedings of the IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 6930–6933.
37. Lang, F.; Yang, J.; Yan, S.; Qin, F. Superpixel segmentation of polarimetric synthetic aperture radar (sar) images based on generalized mean shift. *Remote Sens.* **2018**, *10*, 1592. [[CrossRef](#)]
38. Zhang, W.; Xiang, D.; Su, Y. Fast Multiscale Superpixel Segmentation for SAR Imagery. *IEEE Geosci. Remote. Sens. Lett.* **2020**, 1–5. [[CrossRef](#)]
39. Ciecholewski, M. River channel segmentation in polarimetric SAR images: Watershed transform combined with average contrast maximisation. *Expert Syst. Appl.* **2017**, *82*, 196–215. [[CrossRef](#)]
40. Ijtona, T.B.; Ren, J.; Hwang, P.B. SAR sea ice image segmentation using watershed with intensity-based region merging. In Proceedings of the 2014 IEEE International Conference on Computer and Information Technology, Xi’an, China, 11–13 September 2014; pp. 168–172.
41. Braga, A.M.; Marques, R.C.; Rodrigues, F.A.; Medeiros, F.N. A median regularized level set for hierarchical segmentation of SAR images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1171–1175. [[CrossRef](#)]
42. Jin, R.; Yin, J.; Zhou, W.; Yang, J. Level set segmentation algorithm for high-resolution polarimetric SAR images based on a heterogeneous clutter model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4565–4579. [[CrossRef](#)]
43. Zhang, T.; Zhang, X.; Shi, J.; Wei, S. Depthwise separable convolution neural network for high-speed SAR ship detection. *Remote Sens.* **2019**, *11*, 2483. [[CrossRef](#)]
44. Zhang, T.; Zhang, X. High-speed ship detection in SAR images based on a grid convolutional neural network. *Remote Sens.* **2019**, *11*, 1206. [[CrossRef](#)]
45. Zhang, T.; Zhang, X.; Ke, X.; Zhan, X.; Shi, J.; Wei, S.; Pan, D.; Li, J.; Su, H.; Zhou, Y.; et al. LS-SSDD-v1.0: A Deep Learning Dataset Dedicated to Small Ship Detection from Large-Scale Sentinel-1 SAR Images. *Remote Sens.* **2020**, *12*, 2997. [[CrossRef](#)]
46. Haverkamp, D.; Soh, L.K.; Tsatsoulis, C. A comprehensive, automated approach to determining sea ice thickness from SAR data. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 46–57. [[CrossRef](#)]

47. Tsatsoulis, C.; Kwok, R. *Analysis of SAR Data of the Polar Oceans: Recent Advances*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
48. Salah, M.B.; Mitiche, A.; Ayed, I.B. Multiregion image segmentation by parametric kernel graph cuts. *IEEE Trans. Image Process.* **2010**, *20*, 545–557. [[CrossRef](#)]
49. Kato, Z.; Berthod, M.; Zerubia, J. A hierarchical Markov random field model and multitemperature annealing for parallel image classification. *Graph. Model. Image Process.* **1996**, *58*, 18–37. [[CrossRef](#)]
50. Ren, J.; Hwang, B.; Murray, P.; Sakhalkar, S.; McCormack, S. Effective SAR sea ice image segmentation and touch floe separation using a combined multi-stage approach. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 1040–1043.
51. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
52. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
53. Iandola, F.; Moskewicz, M.; Karayev, S.; Girshick, R.; Darrell, T.; Keutzer, K. Densenet: Implementing efficient convnet descriptor pyramids. *arXiv* **2014**, arXiv:1404.1869.
54. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
55. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 142–158. [[CrossRef](#)]
56. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
57. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [[CrossRef](#)] [[PubMed](#)]
58. Wang, P.; Chen, P.; Yuan, Y.; Liu, D.; Huang, Z.; Hou, X.; Cottrell, G. Understanding convolution for semantic segmentation. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 1451–1460.
59. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
60. Singh, A.; Kalke, H.; Loewen, M.; Ray, N. River ice segmentation with deep learning. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7570–7579. [[CrossRef](#)]
61. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)]
62. Zhang, X.; Jin, J.; Lan, Z.; Li, C.; Fan, M.; Wang, Y.; Yu, X.; Zhang, Y. ICENET: A Semantic Segmentation Deep Network for River Ice by Fusing Positional and Channel-Wise Attentive Features. *Remote Sens.* **2020**, *12*, 221. [[CrossRef](#)]
63. Pogson, L.; Geldsetzer, T.; Buehner, M.; Carrieres, T.; Ross, M.; Scott, K.A. Collecting empirically derived SAR characteristic values over one year of sea ice environments for use in data assimilation. *Mon. Weather Rev.* **2017**, *145*, 323–334. [[CrossRef](#)]
64. Clausi, D.; Qin, A.; Chowdhury, M.; Yu, P.; Maillard, P. MAGIC: Map-guided ice classification system for operational analysis. In Proceedings of the 2008 IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS 2008), Tampa, FL, USA, 7 December 2008; pp. 1–4.
65. Rolnick, D.; Tegmark, M. The power of deeper networks for expressing natural functions. *arXiv* **2017**, arXiv:1705.05502.
66. Mhaskar, H.; Liao, Q.; Poggio, T. Learning functions: When is deep better than shallow. *arXiv* **2016**, arXiv:1603.00988.
67. Zheng, S.; Jayasumana, S.; Romera-Paredes, B.; Vineet, V.; Su, Z.; Du, D.; Huang, C.; Torr, P.H. Conditional random fields as recurrent neural networks. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 11–18 December 2015; pp. 1529–1537.
68. Ma, Y.-D.; Liu, Q.; Qian, A.B. Automated image segmentation using improved PCNN model based on cross-entropy. In Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, Hong Kong, China, 20–22 October 2004; pp. 743–746.
69. Sudre, C.H.; Li, W.; Vercauteren, T.; Ourselin, S.; Cardoso, M.J. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 240–248.
70. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 11–18 December 2015; pp. 1026–1034.
71. Krähenbühl, P.; Koltun, V. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. *arXiv* **2012**, arXiv:cs.CV/1210.5644.
72. Smith, S.L.; Kindermans, P.J.; Ying, C.; Le, Q.V. Don't decay the learning rate, increase the batch size. *arXiv* **2017**, arXiv:1711.00489.
73. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 10–15 June 2019; pp. 6105–6114.