

# End-to-End Classification Network for Ice Sheet Subsurface Targets in Radar Imagery

Yiheng Cai , Shaobin Hu , Shinan Lang \*, Yajun Guo and Jiaqi Liu

Department of Information, Beijing University of Technology, Beijing 100124, China; caiyiheng@bjut.edu.cn (Y.C.); 13027520936@163.com (S.H.); 13366380576@163.com (Y.G.); liu0321@emails.bjut.edu.cn (J.L.)

\* Correspondence: langshinan@bjut.edu.cn

Received: 25 January 2020; Accepted: 1 April 2020; Published: 5 April 2020



**Abstract:** Sea level rise, caused by the accelerated melting of glaciers in Greenland and Antarctica in recent decades, has become a major concern in the scientific, environmental, and political arenas. A comprehensive study of the properties of the ice subsurface targets is particularly important for a reliable analysis of their future evolution. Newer deep learning techniques greatly outperform the traditional techniques based on hand-crafted feature engineering. Therefore, we propose an efficient end-to-end network for the automatic classification of ice sheet subsurface targets in radar imagery. Our network uses bilateral filtering to reduce noise and consists of ResNet module, improved Atrous Spatial Pyramid Pooling (ASPP) module, and decoder module. With radar images provided by the Center of Remote Sensing of Ice Sheets (CReSIS) from 2009 to 2011 as our training and testing data, experimental results confirm the robustness and effectiveness of the proposed network in radargram.

**Keywords:** end-to-end network; classification; ice sheet subsurface targets; radar imagery

## 1. Introduction

In recent years, global warming has exacerbated the melting of glaciers in Greenland and Antarctica, which has considerable influence on sea level rise and, subsequently, the safety of those living in coastal and seaside locations. In order to gather data about glacier melting, it is vital to identify and quantify changes in the ice and bedrock layers of these glaciers. Historically, glaciologists probed the subsurface structure of ice sheets in polar regions by drilling ice cores, but newer methods—such as ground-penetrating radar (GPR) technology—enable scientists to gather robust data sets quickly and efficiently.

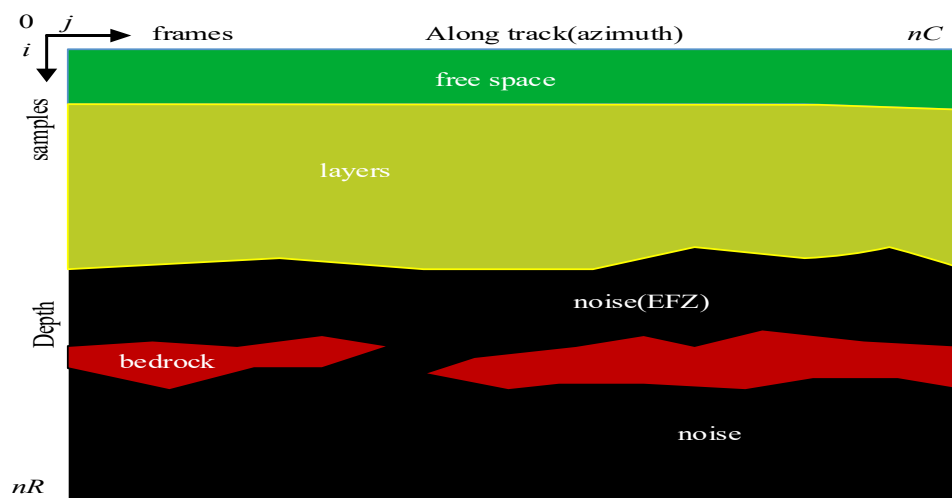
Prior work on this topic involves analyzing data from radar sounder instruments (known as radargrams or echograms) to draw inferences about the properties of the ice sampled. Radar sounders, which are usually operated on airborne or satellite platforms, are active instruments that can perform non-intrusive depth measurements of the subsurface structure of the ice sheets on a large spatial scale. While the advent of ground-penetrating radar made the data collection process more efficient, the data analysis process is still intensely time-consuming because it is typically done by hand.

Newer work in this field utilizes image processing, computer vision, and deep learning techniques to automatically or semi-automatically determine ice surface and bottom boundaries from echograms [1–7]. Gifford et al. [1] employs both the edge-based and active contour methodologies to automate the task of locating polar ice and bedrock layers from airborne radar data acquired over Greenland and Antarctica. In the edge-based approach, edge detection, thresholding, and edge following are utilized to identify the layers of interest for the ice thickness estimation. The active contour approach involves fitting a contour to the boundary using image and contour costs, as well as a gradient force that pushes the contour upward from the bottom of the image. Both methods have their

pros and cons; the edge-based method is more efficient, but the active contour method produces more robust results [1]. A third technique, the level-set model, is better at identifying curved boundaries by embedding the evolving curve into a higher dimensional surface. Mitchell et al. [2] uses a level-set technique for estimating bedrock and surface layers, but find it problematic because of the need to reinitialize the curve manually for each radar image. Therefore, Rahnemoonfar et al. [3] introduces the distance regularization term, which essentially maintains the regularity of the level-set and leads to a stable numerical procedure without the need for reinitialization. Recently, Rahnemoonfar et al. [4] proposes a novel approach that automatically detects the complex topology of the ice surface and bottom boundaries based on the charged particle concept. This method detects the contours in the image based on Coulomb's Law and the assumption that each pixel is an electrically charged particle. Another approach is to use graphical models to detect ice layers in radar echograms. Lee et al. [5] frame boundary tracking as an inference problem and employs a Markov Chain Monte Carlo (MCMC) technique to sample from the joint distribution over all possible layers that exist on a given image. Xu et al. [6] revisits this issue and uses a tree-reweighted message passing (TRW) technique, which first generates a seed surface subject with a set of constraints, and then incorporates additional sources of evidence to refine it via discrete energy minimization. Berger et al. [7] further improves upon the method of Xu et al. [6] by incorporating additional domain-specific knowledge into the cost functions and modeling algorithms.

There are also some efforts to achieve the automatic classification of ice subsurface targets in echograms [8–11]. Ilisei et al. [8,9] exploits the statistical properties of the radar signal to pre-process the radar image data in order to generate a statistical map of the subsurface, and then deploys a segmentation algorithm that had been attuned to that specific study area. Bruzzone et al. [10,11] develops an automatic classification system for subsurface targets, which includes extraction of relevant features based on both the statistical properties of the radar sounder signals and the spatial distribution of the ice subsurface targets. Once specific features are identified, they are then categorized with a support vector machine (SVM) classifier. The classification of each category can be used to focus on the study of specific areas. The classification of the whole bedrock area can be used for geological studies, to understand the thickness of bedrock, to evaluate the type of bedrock material and to derive the absorption characteristics of bedrock [12]. Meanwhile, the thickness of the ice column can be calculated by the classification of layers and bedrock.

At present, the Center of Remote Sensing of Ice Sheets (CReSIS) provide a wide variety of radar data from various radar equipments. A radargram provided by CReSIS is a 2-D matrix with  $nR$  rows (samples  $i$ ) and  $nC$  columns (frames  $j$ ). The typical radargram model includes free space, layers, echo-free zone (EFZ), bedrock, and noise, as shown in Figure 1 (water and freeze-on ice are not considered here). Layers and bedrock are two physical components of the ice sheet subsurface. The ice column consists of a sequence of ice layers. Bedrock is the deepest reflection area of radar wave, and completely attenuates the radar wave. Therefore, the radar equipment under the bedrock can only receive noise. Especially EFZ, which is not a physical component, is the result of the lack of coherent reflector due to the layer disturbance caused by the ice flow at the base interface [13]. In the EFZ, the reflected wave is buried in the thermal noise, so it has a similar distribution with the noise [10]. For this reason, in the automatic classification of ice subsurface targets, the EFZ and noise classes are merged within a single no backscattering target class [10]. Therefore, free space, layers, bedrock, and noise (including EFZ region) are regarded as the classification targets in this paper.



**Figure 1.** The typical radargram model includes free space, layers, echo-free zone (EFZ), bedrock, and noise.

The traditional classification methods need to design features manually, which are high complexity and slow speed, and are not suitable for large, complex datasets. In a more recent development, deep learning technology is employed to better estimate the ice and bedrock boundaries in glacier echograms. Kamangir et al. [14] combines holistically-nested edge detection (HED) [15] with the undecimated wavelet transform technique to develop an end-to-end ice boundary detection network. Xu et al. [16] proposes a multi-task spatiotemporal neural network that combines 3D ConvNets and recurrent neural network (RNN) to estimate ice surface boundaries from sequences of tomographic radar images. However, the identification of the ice subsurface targets using deep learning technology has not been fully explored. Deep learning algorithms are efficient in many public data sets (e.g., Cityscapes dataset, PASCAL-VOC dataset) because they can automatically learn the features of the different data scales without manual intervention. Chen et al. [17–19] propose a series of DeepLab networks for image pixel classification, where the latest network, DeepLabv3+, provides the accurate and high resolution results in PASCAL-VOC2012. Yuan et al. [20] introduces an object context pooling (OCP) scheme and focus on the context aggregation strategy for robust classification. Fu et al. [21] propose a dual attention network (DANet) to adaptively integrate local features with their global dependencies. Therefore, it is worth exploring strategies for applying these deep learning methods to the classification of the ice subsurface targets.

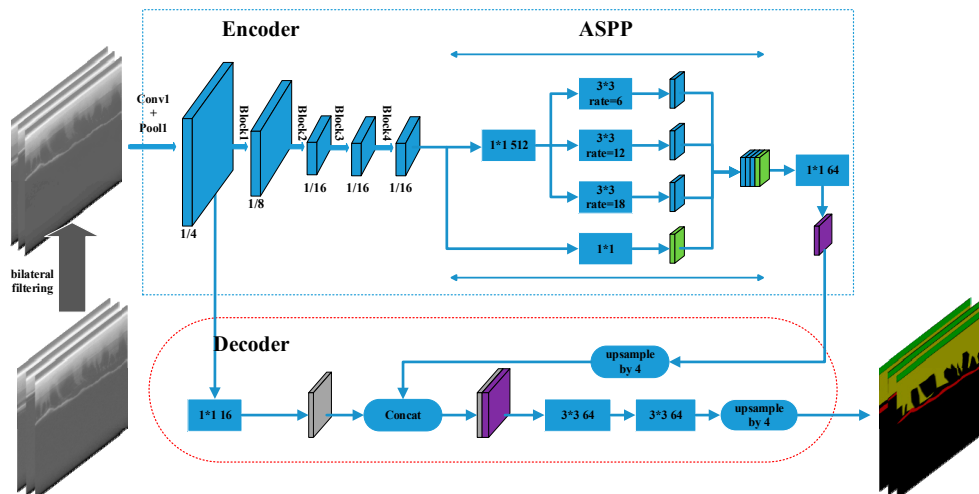
In this paper, we propose a deep convolution classification network to achieve pixel-level classification of ice subsurface targets. This network is composed of filter processing, encoder and decoder. Our network has been validated on the radargram data set provided by the Center of Remote Sensing of Ice Sheets (CReSIS) from 2009 to 2011, and the results show that our network can outperform an automatic classification system for the ice sheet subsurface targets based on support vector machine [10], Deeplab networks [17–19], object context network [20], and dual attention network for scene segmentation [21].

The main contributions of this paper are listed as follows: (1) for the first time in the literature, we introduce the deep convolution network to realize the classification of the ice subsurface targets, and the network realizes the end-to-end processing, which avoids the complex feature engineering for the ice subsurface targets; (2) in the encoder, the modified ASPP structure is used to obtain multi-scale features and improve classification accuracy by removing image level features and changing feature dimensions; (3) in the decoder, a reasonable method of feature fusion is used to solve the problem of long network training and testing times and further improve the accuracy; (4) we use a bilateral filtering algorithm to reduce the speckle noise of radar image to provide the high level of information from radar images.

## 2. Materials and Methods

To decipher and categorize these radargrams, we first reduce the noise in the radar images. Then, we apply the proposed network to robustly classify these ice subsurface targets.

The proposed network structure is shown in Figure 2. Images are pre-processed with bilateral filtering [22], which reduces the amount of noise interference. The network is comprised of an encoder and a decoder. The encoder, which includes the network backbone and the ASPP module, extracts radar image features. The ResNet network, which includes atrous convolution, forms the network backbone, and captures long-range information without changing the image resolution. The decoder combines the features extracted by the encoder with the low-level features. We then apply two  $3 \times 3$  convolutions to refine the features, follow by a simple bilinear upsampling by a factor of 4. We adopt an improved ASPP structure. By removing image-level features, we change the five branches structure into a four branches structure. Additional input feature dimension compression during convolution removes redundant information. To reduce the training time, we modify the channel combination of the decoder by changing the original channel combination from (256,48) to (64,16). The improved network achieves end-to-end processing, and results in faster, more accurate classification of the ice sheet subsurface targets in radar images.



**Figure 2.** Our network architecture consists of ResNet module, improved atrous spatial pyramid pooling (ASPP) module and decoder module.

### 2.1. Noise Removal during Image Pre-Processing

Ground-penetrating radar is capable of retrieving sample data across vast areas of ice, but the resulting images are often plagued by speckles and other types of interference. Speckle noise reduces the quality of radar images, and seriously hinders the interpretation and further interpretation of images (i.e., image feature extraction, image segmentation, etc.). Radar image quality can be improved with a number of noise-reduction techniques. Rahnmounfar et al. [4] successfully denoised images using the anisotropic diffusion method, which requires regional filter smoothing. Kamangir et al. [3] applied an undecimated wavelet transform to decompose the ice radar image into wavelet sub-bands, and then improved overall image quality using threshold processing. Radar image noise reduction must not only reduce noise, but also maintain relatively complete edge information. Inspired by the work proposed by Tomasi and Manuchi [22], in which a bilateral filtering method reduces speckle noise and maintains image edges, we employ bilateral filtering to reduce the noise in the radar images in our proposed classification framework.

Bilateral filtering [22] is a filtering method that considers both the spatial and range information of an image. The method is non-iterative, local, and simple. Bilateral filtering model can be described as

$$h(x) = k^{-1}(x) \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\xi) c(\xi, x) s(f(\xi), f(x)) d\xi \quad (1)$$

with the normalization

$$k(x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} c(\xi, x) s(f(\xi), f(x)) d\xi \quad (2)$$

where  $f(\cdot)$  and  $h(\cdot)$  are the input and output image, respectively.  $x$  denotes the current pixel, and  $\xi$  references the pixels adjacent to pixel  $x$ .  $c(\xi, x)$  signifies the geometric distance between  $x$  and  $\xi$ .  $s(f(\xi), f(x))$  approximates the grayscale value between  $x$  and  $\xi$ .  $c(\xi, x)$  and  $s(f(\xi), f(x))$  are expressed as

$$c(\xi, x) = \exp\left(-\frac{1}{2}\left(\frac{\|\xi - x\|}{\sigma_d}\right)^2\right) \quad (3)$$

$$s(f(\xi), f(x)) = \exp\left(-\frac{1}{2}\left(\frac{\|f(\xi) - f(x)\|}{\sigma_r}\right)^2\right) \quad (4)$$

where  $\sigma_d$  denotes geometric distance variance and  $\sigma_r$  denotes approximate variance of the grayscale value.

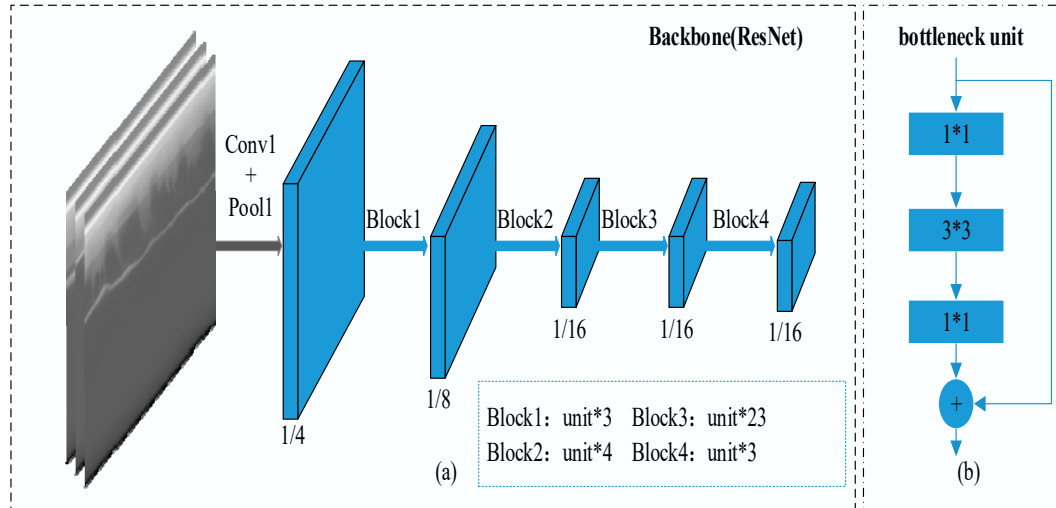
From (3), we can see that the points far from the center  $x$  in space have little influence on the final filtering result of  $x$ . Spatial filtering weights the adjacent points in space, and the weighting coefficient diminishes with the increased distance. From (4), we can see that the range filtering weights the adjacent points with similar grayscale values, and the weighting coefficient decreases with the large variation in adjacent pixel grayscale values. Because bilateral filtering includes both spatial and range information of the image, it can significantly reduce speckle noise in radar images. We use bilateral filtering [22] in training and test images, which allows for noise reduction of images and improves classification performance in the radar images.

## 2.2. Encoder Structure

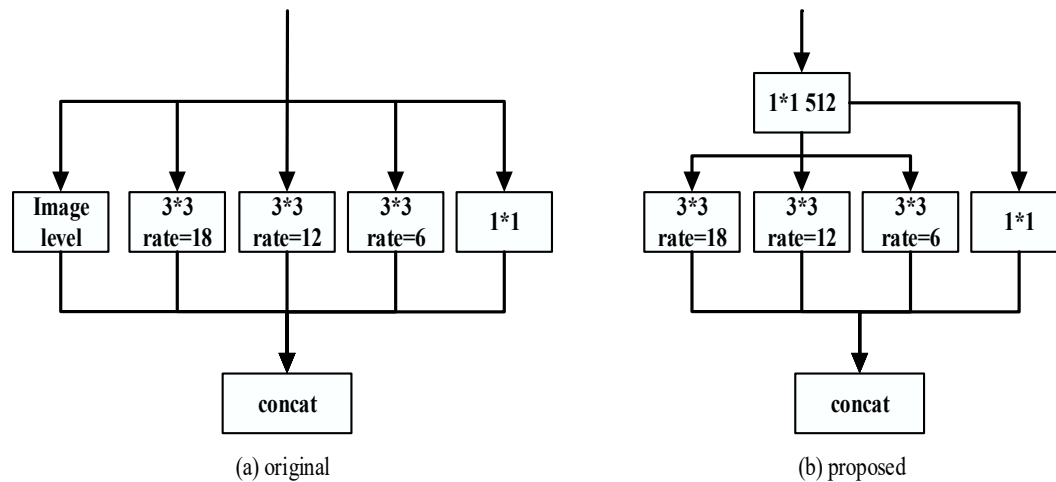
We use ResNet-101 [23] as the backbone network of the encoder structure. ResNet is a residual learning framework used to ease the training of networks that are substantially deeper than those used previously. The backbone network (ResNet)—which consists of conv1, a pool1, and four block modules—extracts high-level features from ice radar images (Figure 3a). Each block consists of several bottleneck units (Figure 3b). We use output stride to represent the ratio of input image spatial resolution to final output resolution, and set the output stride to 16. After the filtered image is input into the backbone network and passed through conv1 and pool1, the output resolution is 1/4 of the input image resolution. Then, after passing through block1, block2, block3, and block4 in turn, the output resolution becomes 1/8, 1/16, 1/16, and 1/16 of the input image, respectively. Here, the stride of the last pooling or convolutional layer in block3 is set to 1 to avoid signal decimation. Then, all subsequent convolutional layers are replaced with atrous convolutional layers with rate  $r = 2$ . Atrous convolution adaptively modifies the filter's field-of-view by changing the rate value, which allows us to extract dense feature responses without changing the image spatial resolution or learning any extra parameters [17,18]. The backbone network finally outputs 2048 dimensional features as the input of the atrous spatial pyramid pooling module.

Atrous spatial pyramid pooling (ASPP) is inspired by the efficacy of spatial pyramid pooling [24], which demonstrates that it is effective to resample features at different scales to accurately and efficiently classify regions of an arbitrary scale. We review the ASPP module proposed in [19], as shown in Figure 4a, where three atrous convolutions with different atrous rates, image-level features, and a

$1 \times 1$  convolution are applied behind the network backbone. The features of all branch results are concatenated and then fed through a  $1 \times 1$  convolution. Based on the success of the original ASPP module, we propose a new ASPP structure to further improve the classification performance and reduce training parameters, as illustrated in Figure 4b.



**Figure 3.** (a) Resnet-101 as the backbone network. (b) Several bottleneck units reside within each block.



**Figure 4.** (a) The original ASPP structure. (b) The improved ASPP structure. We changed the input feature dimension of the atrous convolution and removed the image-level features.

Considering the ordered distribution of ice subsurface targets, the features captured by the atrous convolution with different rates in the ASPP module represent the global contextual information captured by image-level features. Therefore, we remove image-level features from consideration and use a four branches structure to replace the five branches structure of traditional ASPP. In addition, we use a  $1 \times 1$  convolution with 512 dimensions to change the input feature dimension of atrous convolution. The original ASPP module is designed to evaluate data with respect to the PASCAL VOC 2012 semantic segmentation benchmark [25], which contains 20 foreground object classes and one background class, and has an input feature dimension of 2048. However, ice radar images can only be divided into four categories. In the original ASPP module, the input feature dimension of 2048 is high, which will not only cause feature redundancy, but also increase the training time. Therefore, a  $1 \times 1$  convolution is employed to reduce both the input feature dimension (to 512) and the time needed to reconstruct



features. Similarly, we modify the encoder output feature from 256 to 64, which is beneficial to further refine features used in the decoder.

### 2.3. Decoder Structure

We use a decoder structure in the whole network framework as shown in Figure 5. In order to use the edge information in low-level features to enrich the classified edge, the decoder concatenates the high-level and the low-level features in proportion. Then two  $3 \times 3$  convolutions are used to refine the features, and the image features are bilinearly up-sampled by a factor of 4 to achieve end-to-end processing.

Because the number of encoder features channels is modified from original 256 to 64, we also reduce the channel number of the low-level features in the decoder from 48 to 16. The two types of features are then combined and processed using two  $3 \times 3$  convolutions with 64 channels. In the decoder structure, the (64,16) channel combination can reduce redundant information in feature fusion. We find that (64,16) channel combination improves the accuracy of radar image classification and significantly reduces the training time.

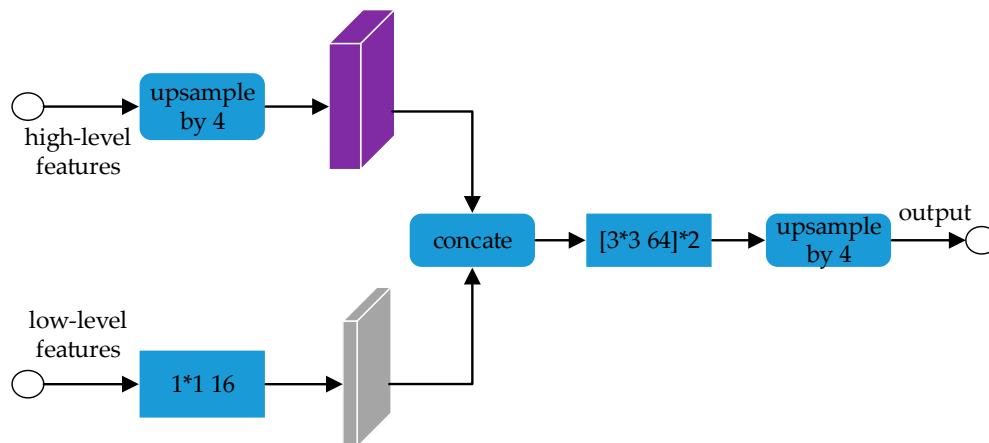


Figure 5. The decoder concatenates the high-level and the low-level features.

### 3. Results

Our sample data consist of radar images of Antarctica acquired between 2009 and 2011 (i.e., 2009\_Antarctica DC8, 2010 Antarctica DC8 and 2011 Antarctica TO). The first and second data sets (i.e., 2009 and 2010) were obtained by MCoRDS with bandwidth of 9.5 MHz and transmission power of 550 W. The third data set (i.e., 2011) was acquired by MCoRDS2 with bandwidth of 30 MHz and transmission power of 1050 W. For all data sets, pulse compression and windowing algorithms are used to improve the range resolution of the image, and synthetic aperture radar (SAR) processing is used to improve the azimuth resolution. The range resolution of radar image obtained by MCoRDS and MCoRDS2 is 13.6 m and 4.3 m respectively. The azimuth resolution of all images is 25 m. Moreover, a minimum variance distortionless response technique is employed to suppress the clutter from the cross-track direction [26]. In this work, 360 images are used for training and 100 images are used for testing. All images of training and testing contain all target classes, i.e., free space, layers, bedrock, and noise (including EFZ). Although the radar image dataset is small, the network can accelerate the convergence speed of gradient descent and obtain better classification performance, due to the ResNet model which is pretrained by the ILSVRC-2012-CLS image classification dataset in the training phase. Our implementation is built on TensorFlow. The batch normalization parameters are trained with decay = 0.9997. Momentum and weight decay coefficients are set to 0.9 and 0.0002 respectively. Batchsize is set to 4 with 10,240 iterations. We employ a ‘poly’ learning rate policy where the initial

learning rate is 0.007. During training, we set the image crop size to 513, randomly flip the images from right to left, and scale the input images from 0.5 to 2.0.

Like other classification methods of the ice subsurface targets [10,11], this paper employs overall accuracy (OA) to evaluate classification performance of our network. Instead of taking OA as the evaluation of window sample classification (i.e., each window contains multiple pixels) as in [10] and [11], we take OA to measure the image pixel classification results. OA represents the percentage of predicted correct classification to total samples. At the same time, we also use kappa coefficient as the evaluation metric, which integrates the diagonal and non-diagonal terms of the confusion matrix.

### 3.1. ASPP Design Choices

To evaluate the effect of different branches in ASPP, we compare the performance of different structures after removing different branches in ASPP, as shown in Table 1. After removing image-level features, the classification performance of the structure is better than that of the original network and other variant structures, and an observation that validates our choice to remove image-level features from ASPP structure to improve network performance with respect to radar image classification.

**Table 1.** Comparison of the classification performance of the different ASPP branches.

Conv 1*1	Conv 3*3 Rate = 6	Conv 3*3 Rate = 12	Conv 3*3 Rate = 18	Image Level	OA	Kappa
✓	✓	✓	✓	✓	97.55	95.18
✓	✓	✓	✓		97.68	95.44
✓	✓	✓		✓	97.61	95.34
✓	✓		✓	✓	97.58	95.26
✓		✓	✓	✓	97.54	94.59
	✓	✓	✓	✓	97.51	94.57

Using different input feature dimension values, we further demonstrate the positive effect of reducing the ASPP input feature dimension from 2048 to 512 (Table 2). The OA and Kappa values are optimal with 512 dimensions; other dimension values resulted in lower classification performance because of feature redundancy (higher dimension), or loss of image details (lower dimension). With this improved ASPP structure, we have improved the classification performance over that of the original ASPP structure.

**Table 2.** Comparison of different ASPP input feature dimension values to evaluate the classification performance of the corresponding network.

Input Channel	64	256	512	1024	2048	OA	Kappa
atrous conv (6,12,18)	✓					97.45	94.36
		✓				97.57	94.68
			✓			97.68	95.44
				✓		97.59	95.29
					✓	97.44	95.00

### 3.2. Decoder Design Choices

We change the channel combination of both the encoder features and the low-level features in the decoder structure. In order to evaluate the effectiveness of our proposed channel combination method, we compare the classification performances of four different channel combinations and the number of the network model parameters generated during their training. As shown in Table 3, the channel combination (64,16) has the highest OA and Kappa values and the best classification performance. This channel combination also requires the second lowest number of parameters, surpassed only by channel combination (32,8). Compared to the original structure, the number of parameters in combinations



(32,8) and (64,16) are reduced by 25% and 22%, respectively. Channel combination (64,16) represents the optimal combination of strong classification performance and a reduced number of parameters in the network model.

**Table 3.** Comparison of network classification performance and parameters with different channel combinations.

Channel	OA	Kappa	Parameters (M)
(256,48)	97.66	95.44	474.8
(128,48)	97.67	95.45	404.8
(64,16)	97.71	95.50	371.8
(32,8)	97.66	95.39	356.0

We also design different convolution structures for the decoder module, and report the findings in Table 4. The feature dimension of our convolution structure is 64; after concatenating the feature maps, we find that it is best to employ two 3\*3 convolutions with 64 channels to refine the feature map.

**Table 4.** Comparison of different convolution structures after combining features.

Conv Structure	OA	Kappa
[3*3 64]	97.44	94.36
[3*3 64]*2	97.71	95.50
[3*3 64]*3	97.54	94.52
[1*1 64]+[3*3 64]	97.63	94.80

### 3.3. Image Filtering

We compare the noise suppression effects of three different noise reduction methods on radar images: Lee filtering, anisotropic diffusion, and bilateral filtering. To quantify the noise reduction of each method, we use the equivalent number of looks (ENL) and edge preserving index (EPI) metrics as the primary evaluation criteria. ENL represents the image smoothing effect, and EPI represents the ability of the network to preserve image edges. ENL and EPI are expressed as

$$ENL = \frac{\bar{\mu}^2}{\sigma^2} \quad (5)$$

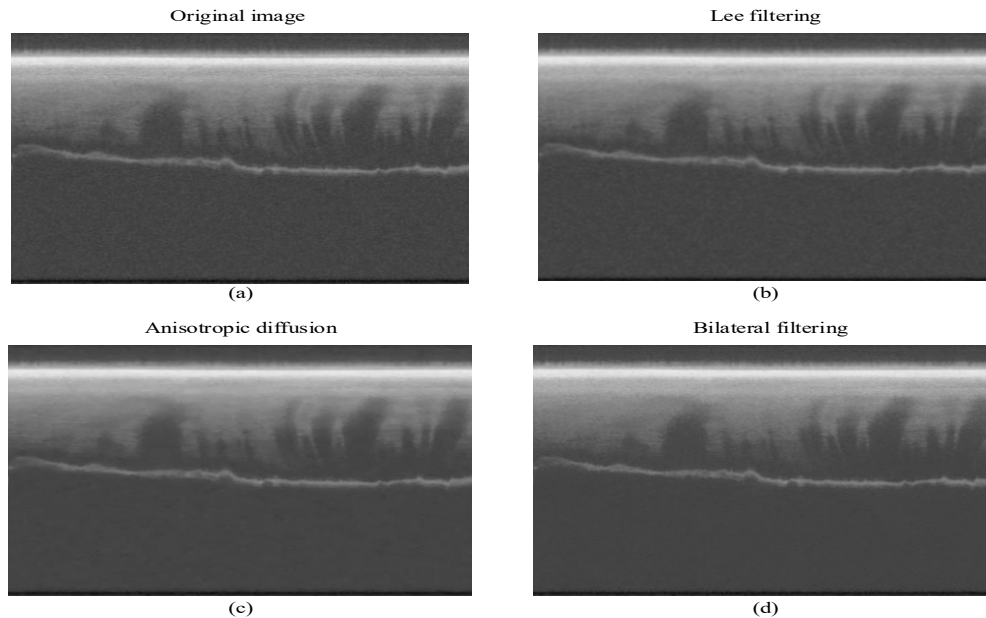
$$EPI = \frac{\sum(|p_s(i, j) - p_s(i + 1, j)| + |p_s(i, j) - p_s(i, j + 1)|)}{\sum(|p_o(i, j) - p_o(i + 1, j)| + |p_o(i, j) - p_o(i, j + 1)|)} \quad (6)$$

where  $\bar{\mu}$  denotes the mean value of the image,  $\sigma^2$  denotes the variance of the image,  $p_s(i, j)$  denotes the grayscale value of the output image at the point  $(i, j)$ , and  $p_o(i, j)$  denotes the grayscale value of the input image at the point  $(i, j)$ . A higher value of ENL corresponds to a smoother image; a higher EPI value corresponds to better preserved image edges.

Using our experimental dataset, we calculate the EPI and ENL values for Lee filtering, anisotropic diffusion, and bilateral filtering [22] (Table 5). Bilateral filtering yields the best image smoothing and edge preservation of the radar images. Figure 6 visualizes the images after different filtering methods: original image (Figure 6a), Lee filtering (Figure 6b), anisotropic diffusion (Figure 6c), and bilateral filtering (Figure 6d). To demonstrate that filtering can improve the classification accuracy of the network, we calculate the OA and Kappa values for network models that had been trained and tested on the original image and the filtered image (Table 6); the network trained and tested with the filtered images results removes more noise from the images, resulting in improved ice subsurface target classification.

**Table 5.** Comparison of different noise reduction methods.

Filter Type	Equivalent Number of Looks (ENL)	Edge Preserving Index (EPI)
Original images	8.0418	1
Lee filtering	8.1804	0.4740
Anisotropic diffusion	8.1196	0.4204
Bilateral filtering	8.2257	0.4965

**Figure 6.** (a) Original image. (b) Lee filtering image. (c) Anisotropic diffusion image. (d) Bilateral filtering image.**Table 6.** Classification performance using the original image and the filtered image as the network input.

Input Type	OA	Kappa
Original image	97.56	95.23
Filtered image	97.73	95.53

### 3.4. Comparison with Deep Learning Methods

To assess the classification performance of our network, we use our data set to calculate the OA and Kappa values for the other methods (Table 7). The first three are different DeepLab network versions proposed recently by Chen et al., all of which use ResNet-101 as the network backbone, and are generally considered to be methods with high accuracy and robustness. The next two methods are two kinds of OCNet, ResNet-OC, and ASP-OC, which introduce object context pooling (OCP) on the basis of ResNet and ASPP modules respectively. DANet proposes to append two types of attention modules on top of dilated FCN. OCNet and DANet all achieve new state-of-the-art performance in challenging scene pixel level classification.

As shown in Table 7, the proposed approach shows a consistent performance improvement over the other classification methods. OCNet and DANet do not achieve as good a performance in the classification of ice subsurface targets as they do in scene pixel level classification. From this comparison, note that Deeplabv3+ obtains better classification results, which indicates that ASPP and decoder modules are necessary for the classification of ice subsurface targets. Although Deeplabv3+ and the proposed method both have ASPP and decoder modules in the structure, as compared with Deeplabv3+, the proposed method shows improvements of 0.13% OA on a radar image dataset; that is, the proposed method modified using bilateral filtering, ASPP of removing image level features,

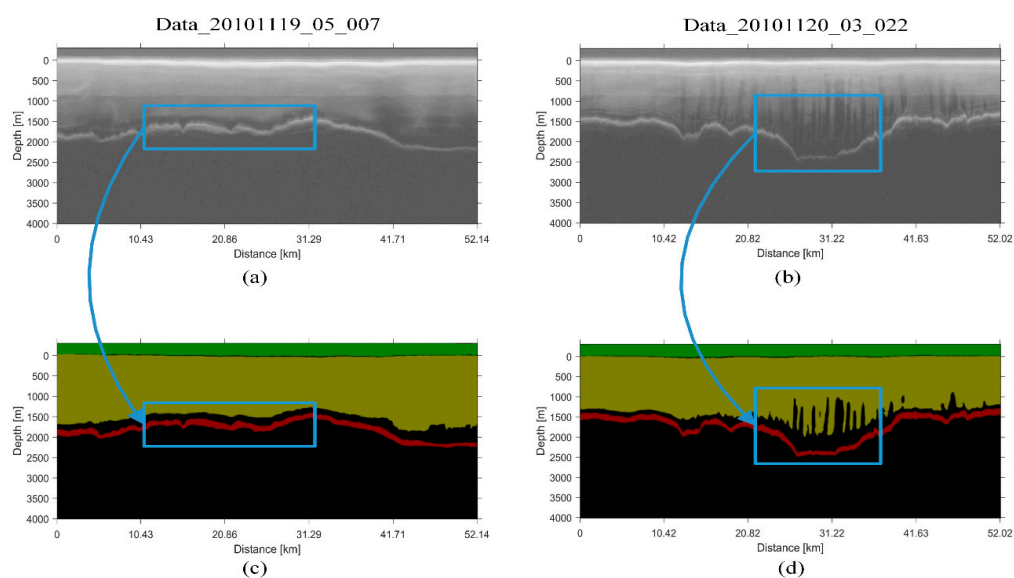
changing feature dimensions, and effective decoding of feature fusion can learn more information from radar images.

**Table 7.** Comparison of classification performance of different deep learning methods.

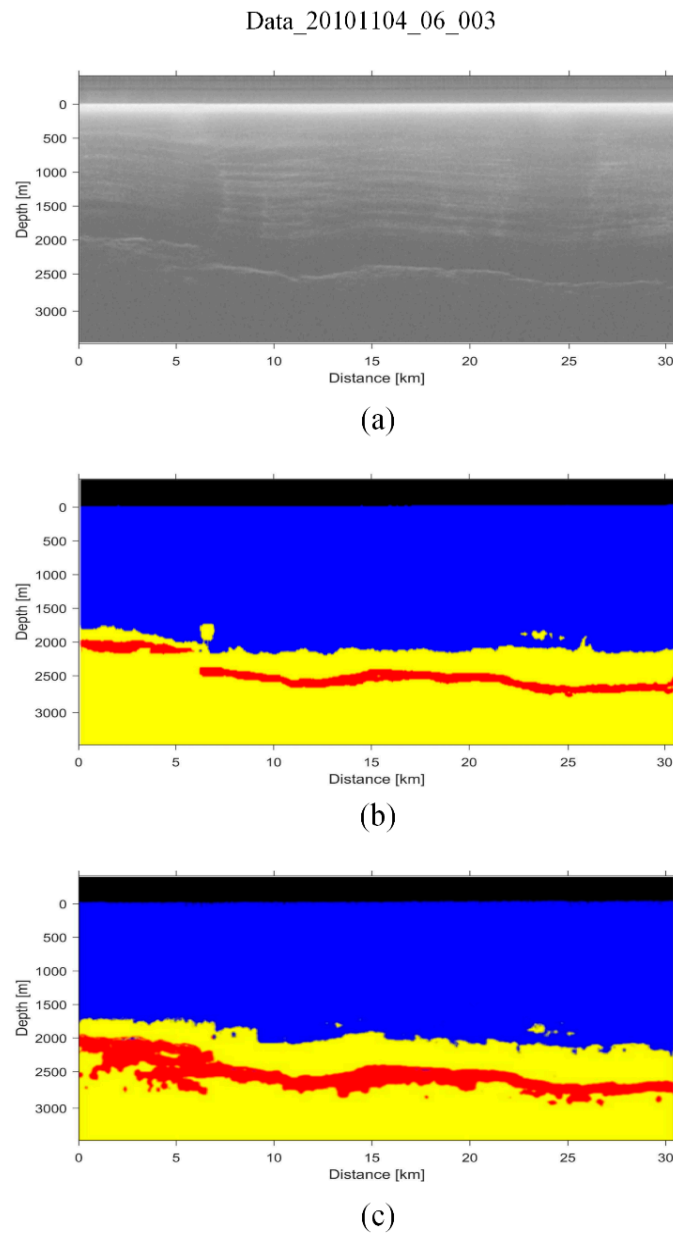
Method	OA	Kappa
Deeplabv2 (ResNet-101) [17]	97.60	95.24
Deeplabv3 (ResNet-101) [18]	95.48	90.91
Deeplabv3+ (ResNet-101) [19]	97.60	95.47
OCNet(ResNet-OC) [20]	96.39	92.19
OCNet(ASP-OC) [20]	96.80	93.13
DANet [21]	96.44	92.59
Our network	97.73	95.53

### 3.5. Visualization of Results

To evaluate the classification performance of our method more intuitively, we visualize the classification maps in Figure 7. The results show that our method has a good effect on the details of classification targets, especially the highlighted areas marked by rectangles in the figures where the bedrock exhibits slight changes (Figure 7a,c) and the layers change continuously (Figure 7b,d). We further compare our method with other classification method of ice subsurface targets (i.e., reference [10]). Considering that [10] takes OA as the evaluation of window sample classification, which is not suitable for comparison with the OA used to measure the more detailed image pixel classification in this paper, we only qualitatively visualize the classification results of the two methods, as shown in Figure 8. Compared with the bedrock area in the radargram, the classification results of the method reported in [10] in the bedrock area are noticeably wider (Figure 8a,c), which can be caused by fixed window classification. However, our method gets more accurate results in bedrock areas (Figure 8b). Moreover, our method is very fast, taking only an average of 2 s to infer each image on a computer with Intel Core i7-7700 @ 3.6GHz and NVIDIA GTX1080Ti GPU card. The method of [10] needs about 45 s to infer each image by using a cluster of 192 CPUs (at 2.05 GHz).



**Figure 7.** Examples of (a,b) radargrams and (c,d) corresponding classification maps generated with the presented network. In the classification maps, each class corresponds to a different color: free space (green); layers (yellow); bedrock (red); noise and EFZ (black).



**Figure 8.** Examples of (a) radargram, (b) corresponding classification maps generated with the presented network and (c) corresponding classification maps generated with the [10] method. In the classification maps, each color represents a different target class: free space (black); layers (blue); bedrock (red); noise and EFZ (yellow).

### 3.6. Experimental Results on Image Boundaries

In order to evaluate the performance of our classification results on the boundary, we obtain the boundary results of our method (i.e., ice surface and ice bottom), which directly affect the accuracy of the ice thickness calculation. We use the widely used balanced *F-measure* for evaluation. The *F-measure* equation is

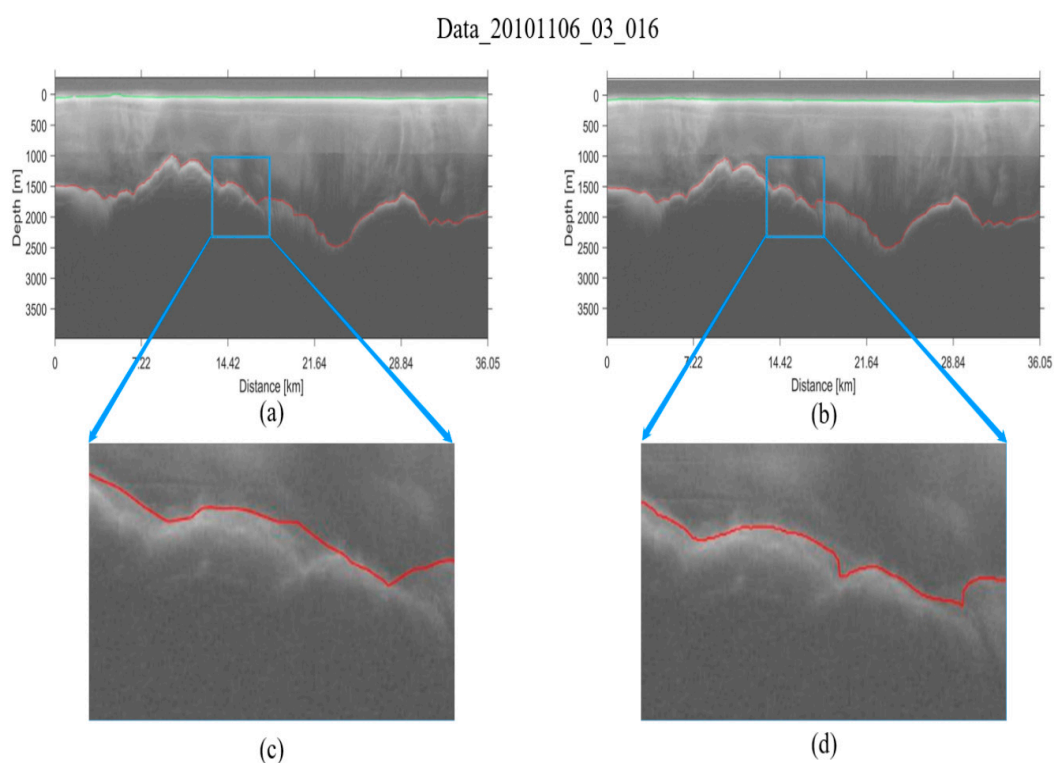
$$F\text{-measure} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (7)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad (8)$$

$$recall = \frac{TP}{TP + FN} \quad (9)$$

where  $TP$  is a true positive,  $FP$  is a false positive, and  $FN$  is a false negative. The F-measure is the weighted harmonic mean of precision and recall.

We calculate the precision, recall, and F-measure in the test set. Our method obtains 77% F-measure for the entire test set, which is same accuracy as the existing state-of-the-art boundary detection method [14]. Figure 9 visualizes the boundary result maps. Comparing Figure 9b, the result of the proposed network, with Figure 9a, the manually picked interfaces, it can be concluded that the proposed network gets the similar result as the manually picked interfaces; furthermore, our result appears to be even more accurate in some parts as shown in Figure 9c,d. In conclusion, our method not only gets high precision classification results, but also shows promising results with respect to manually picked data on the boundary.



**Figure 9.** (a) Manually picked interfaces. (b) The boundary map generated with the presented method. (c,d) Magnified section of (a,b). Each color represents a different boundary: ice surface (green); ice bottom (red).

#### 4. Conclusions

In this paper, we have presented a novel method for the automatic classification of ice sheet subsurface targets, which can automatically divide the radar images into various categories for analysis of ice sheet characteristics and therefore solve the time-consuming problem of ice radar image data analysis. Our methodology, which uses ResNet and improved ASPP modules to extract multi-scale features, and utilizes the decoder module to fuse high-level semantic features and low-level features, can accurately classify ice radar images at the pixel level. Improvements we make to the ASPP and decoder modules, as well as the employment of bilateral filtering, resulted in less noisy radar images and, subsequently, better classification performance with CReSIS radar image data from 2009 to 2011.

**Author Contributions:** Conceptualization, Y.C., S.L., and S.H.; Methodology, Y.C. and S.H.; Validation, S.H.; Formal analysis, S.H., Y.G., and J.L.; Writing—original draft preparation, Y.C. and S.H.; Writing—review and editing, Y.C., S.L., S.H., Y.G., and J.L.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key Research and Development Program of China (no. 2017YFC1703302) and the National Natural Science Foundation of China (No. 41776186).

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

## References

- Gifford, C.M.; Finyom, G.; Jefferson, M.; Reid, M.; Akers, E.L.; Agah, A. Automated polar ice thickness estimation from radar imagery. *IEEE Trans. Image Process.* **2010**, *19*, 2456–2469. [CrossRef] [PubMed]
- Mitchell, J.E.; Crandall, D.J.; Fox, G.C.; Rahnemoonfar, M.; Paden, J.D. A semi-automatic approach for estimating bedrock and surface layers from multichannel coherent radar depth sounder imagery. *Proc. SPIE* **2013**, 8892, 88921E–88926E.
- Rahnemoonfar, M.; Fox, G.C.; Yari, M.; Paden, J.D. Automatic ice surface and bottom boundaries estimation in radar imagery based on Level-Set approach. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 5115–5122. [CrossRef]
- Rahnemoonfar, M.; Habashi, A.A.; Paden, J.D.; Fox, G.C. Automatic ice thickness estimation in radar imagery based on charged particles concept. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium, Fort Worth, TX, USA, 23–28 July 2017; pp. 3743–3746.
- Lee, S.; Mitchell, J.; Crandall, D.J.; Fox, G.C. Estimation bedrock and surface layer boundaries and confidence intervals in ice sheet radar imagery using MCMC. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 111–115.
- Xu, M.; Crandall, D.J.; Fox, G.C.; Paden, J.D. Automatic estimation of ice bottom surfaces from radar imagery. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 340–344.
- Berger, V.; Xu, M.; Chu, S.; Crandall, D.J.; Paden, J.D.; Fox, G.C. Automated tracking of 2D and 3D ice radar imagery using VITERBI and TRW-S. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Valencia, Spain, 22–27 July 2018; pp. 4162–4165.
- Ilisei, A.M.; Ferro, A.; Bruzzone, L. A technique for the automatic estimation of ice thickness and bedrock properties from radar sounder data acquired at antarctica. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany, 22–27 July 2012; pp. 4457–4460.
- Ilisei, A.M.; Bruzzone, L. A model-based technique for the automatic detection of earth continental ice subsurface targets in radar sounder data. *IEEE Geosci. Remote Sens. Lett.* **2014**, *28*, 1911–1915. [CrossRef]
- Ilisei, A.M.; Bruzzone, L. A system for the automatic classification of ice sheet subsurface targets in radar sounder data. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3260–3277. [CrossRef]
- Khodadadzadeh, M.; Ilisei, A.M.; Bruzzone, L. A technique based on adaptive windows for the classification of radar sounder data. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 3739–3742.
- Daniels, D. *Ground Penetrating Radar*; Institution of Engineering and Technology: London, UK, 2004.
- Drews, R. Layer disturbances and the radio-echo free zone in ice sheets. *Cryosphere* **2009**, *3*, 195–203. [CrossRef]
- Kamangir, H.; Rahnemoonfar, M.; Dobbs, D.; Paden, J.D.; Fox, G.C. Deep hybrid wavelet network for ice boundary detection in radar imagery. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Valencia, Spain, 22–27 July 2018; pp. 3449–3452.
- Xie, S.; Tu, Z. Holistically-nested edge detection. In Proceedings of the IEEE International Conference Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1395–1403.
- Xu, M.; Fan, C.; Paden, J.D.; Fox, G.C.; Crandall, D.J. Multi-Task Spatiotemporal Neural Networks for Structured Surface Reconstruction. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 1273–1282.
- 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.* **2018**, *40*, 834–848. [CrossRef] [PubMed]
- Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. Available online: <https://arxiv.org/abs/1706.05587> (accessed on 5 December 2017).



19. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 833–851.
20. Yuan, Y.; Wang, J.D. OCNet: Object Context Network for Scene Parsing. Presented at Arxiv. Available online: <https://arxiv.org/pdf/1809.00916.pdf> (accessed on 4 September 2018).
21. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.J.; Fang, Z.W.; Lu, H.Q. Dual Attention Network for Scene Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019.
22. Tomas, C.; Manduchi, R. Bilateral filtering for gray and color images. In Proceedings of the International Conference on Computer Vision (ICCV), Bombay, India, 4–7 January 1998; pp. 839–846.
23. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June 2016; pp. 770–778.
24. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. In Proceedings of the European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014; pp. 346–361.
25. Everingham, M.; Eslami, S.M.A.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The pascal visual object classes challenge a retrospective. *Int. J. Comput. Vision (IJCV)* **2014**, *111*, 98–136. [[CrossRef](#)]
26. Li, J.; Paden, J.; Leuschen, C.; Rodriguez-Morales, F.; Hale, R.D.; Arnold, E.J.; Crowe, R.; Gomez-Garcia, D.; Gogineni, P. High-Altitude Radar Measurements of Ice Thickness Over the Antarctic and Greenland Ice Sheets as a Part of Operation IceBridge. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 742–754. [[CrossRef](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).