

© 2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

A Deep Learning Architecture for Semantic Segmentation of Radar Sounder Data

This paper appears in: IEEE Transactions on Geoscience and Remote Sensing

Date of Publication: 13 November 2021

Author(s): Elena Donini; Francesca Bovolo; Lorenzo Bruzzone

Volume: 60

Page(s): 1 – 14

DOI: <https://doi.org/10.1109/TGRS.2021.3125773>

A Deep Learning Architecture for Semantic Segmentation of Radar Sounder Data

Elena Donini¹, Member, IEEE, Francesca Bovolo², Senior Member, IEEE,
and Lorenzo Bruzzone³, Fellow, IEEE

Abstract—During the last decades, radar sounders provided direct measurements (radargrams) of the Earth’s polar caps’ subsurface. Radargrams are of critical importance for a better understanding of glaciologic structures and processes of the ice sheet in the framework of climate change. This article aims to automatically extract information on basal boundary conditions given their substantial relevance for modeling the ice-sheet processes, such as the sliding. We introduce a novel automatic method based on deep learning to detect the basal layer and basal units in radargrams acquired in the inland of icy areas. Radargrams are segmented into englacial layers, bedrock, basal units, and noise-limited regions; the latter includes the echo-free zone (EFZ), thermal noise, and signal perturbation. The network is a U-Net with attention gates and the Atrous Spatial Pyramid Pooling (ASPP) module that automatically extract semantically meaningful features at different scales. Experimental results on two datasets acquired in north Greenland and west Antarctica by the Multichannel Coherent Radar Depth Sounder (MCoRDS3) indicate a high overall segmentation accuracy. The accuracy of basal ice and signal perturbation detection is high, and that of the other classes is comparable with the literature techniques based on handcrafted features. The results show the effectiveness of the proposed method in automatically extracting semantically meaningful features to segment radargrams and map the basal layer and basal units.

Index Terms—Basal boundary conditions, basal units, convolutional neural network, cryosphere, deep learning, radar sounder (RS), remote sensing.

I. INTRODUCTION

ICE sheets feed ice shelves that calve and melt into the sea, representing one of the leading causes of the global sea-level rise. Ice sheet seaward flow depends on the basal boundary conditions, related to several thermal and mechanical factors at the basal interface [1]–[3]. The basal layer, i.e., the deepest part of the ice column, is critical for understanding ice sheet processes, including sliding. Therefore, accurate modeling of the basal conditions requires direct measurements

Manuscript received March 23, 2021; revised August 6, 2021 and November 2, 2021; accepted November 2, 2021. Date of publication November 8, 2021; date of current version February 15, 2022. This work was supported by the Italian Space Agency through the “Attività Scientifiche per JUICE fase C-D” under Contract Agenzia Spaziale Italian—Istituto Nazionale di Astrofisica (ASI-INAF) and Contract 2018-25-HH.0. (Corresponding author: Francesca Bovolo.)

Elena Donini is with the Center for Digital Society, Fondazione Bruno Kessler, 38123 Trento, Italy, and also with the Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy (e-mail: edonini@fbk.eu).

Francesca Bovolo is with the Center for Digital Society, Fondazione Bruno Kessler, 38123 Trento, Italy (e-mail: bovolo@fbk.eu).

Lorenzo Bruzzone is with the Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy.

Digital Object Identifier 10.1109/TGRS.2021.3125773

of the ice sheet down to the bedrock. Traditionally, information on the ice sheet is extracted by analyzing either ice core samples obtained by drilling or radar sounder (RS) data. Ice core samples have been providing a large amount of information [4], [5], but extracting them is intrusive and expensive. To deal with these issues, RSs have been widely used during the last decades to probe the inside of the ice sheets in Greenland and Antarctica [6]. RSs are active sensors that transmit electromagnetic (EM) waves toward nadir with frequencies usually below 1 GHz and with a wide enough bandwidth to resolve ice features in the vertical dimension. The EM waves propagate into the subsurface and interact with the buried geologic structures. Part of the wave power is reflected toward the antenna by each dielectric discontinuity. The coherent sum of these reflections is collected in radargrams. A large amount of airborne RS data is available, and the data volume is expected to grow further in the future with planned missions [7]–[9]. Recently, the basal layers have been visually analyzed to identify geologic structures, such as basal units [10], [11] and Units of Disrupted Radio-stratigraphy (UDR) [12], [13]. They elongate for several kilometers and extend up to half of the ice sheet thickness, causing the ice layering to fold toward the surface. Although visual inspection has been supporting the characterization of several targets, it is subjected to several limitations. Visual inspection is time-demanding and not suitable for analyzing a large number of radargrams. Moreover, it is subjective and can lead to inconsistencies and misinterpretations. Hence, automatic methods are now emerging to address these limitations and automatically extract information on the subsurface [14]–[17].

In the literature, automatic methods for analyzing the basal boundary conditions in RS data fall into three categories. The first uses numerical modeling for ice sheet geologic processes and locally infers the basal boundary condition type. Numerical models [18], [19] consider data from different sensors (e.g., altimeter and optical images) and analyses (e.g., subsurface temperature profile) as input, extracting only the ice sheet thickness and the topography of the basal interface from RS data. However, other studies [11], [20], [21] indicate that RS data contain more information, including basal units and water at the interfaces that can further enhance the modeling of the basal conditions.

The second strategy automatically analyzes the layer stratigraphy of radargrams to identify disruptions in the basal layer without considering the basal interface. The englacial stratigraphy (i.e., the layer) is extracted with a line detection algorithm, and the line slopes are used to model

the disruption [12], [22], [23]. However, these techniques focus on analyzing the ice stratigraphy disruption without mapping the basal layer, e.g., basal ice and basal units. The third strategy extracts the main geologic targets by semantically segmenting radargrams. The study in [16] segmented radargrams into three targets: the stratigraphy of the ice layering, the bedrock, and a combination of the echo-free zone (EFZ) and thermal noise. Donini *et al.* [15] expanded the work in [16] by considering also basal ice. Both methods extract handcrafted features based on *a priori* knowledge on the target class spatial and statistical properties to perform a pixel-based classification with support vector machines (SVMs). Handcrafted features strongly depend on human design and model-specific classes. However, increasing the number of classes, this approach is subjected to performance degradation in tackling the problem complexity. Thus, there is a need to develop methods capable of extracting robust features automatically. Recently, deep learning has provided powerful means for image segmentation in various applications [24]–[29]. Moreover, deep-learning-based methods handle the feature extraction automatically and provide robust results in other applications.

In the semantic deep learning literature, several works exist performing a supervised pixel-based learning task, given reliable training samples [24]–[29]. The standard approach is based on dense or fully convolutional deep neural networks (FCNNs). Dense networks are made of fully connected layers, where each neuron in one layer is connected to each neuron in the previous and next ones [24]. FCNNs consist of 1×1 convolutions that perform the task of the fully connected layers in dense networks. In recent years, many variants of FCNNs have been proposed [25]–[29] mostly to analyze passive data (e.g., optical data). For instance, Ronneberger *et al.* [28] proposed a U-Net architecture with a contracting path to capture context, and a symmetric expanding path enabling segmentation was proposed. Attention U-Net is enriched with attention gate (AG) modules that force the network to focus on relevant regions of the input data in [30]. The model was further improved in [31] by adding the Atrous Spatial Pyramid Pooling (ASPP) module in the bottleneck to control the feature resolution. The ASPP module robustly segments images at multiple scales by enlarging the receptive field to incorporate a more extensive context while saving on the number of the network parameters compared to the regular convolution [29]. Recently, deep learning has been applied to radargrams obtaining good performance for: 1) detecting the ice layers [32], [33]; 2) simulating RS images with generative adversarial network (GAN) [34]; 3) target detection [35]; and 4) segmentation [36]–[38]. The study in [38] applied a ResNet with an ASPP module to segment radargrams in ice layers, thermal noise and EFZ, and bedrock with an overall accuracy (OA) slightly lower than other literature methods based on the SVM [15], [16]. In addition, the study did not consider the basal layer and the targets related to basal boundary conditions.

This article proposes a novel automatic method based on deep learning to segment icy inland radargrams and locate meaningful geologic targets for better understanding

the ice sheet and the basal boundary condition processes. Radargrams are segmented into classes, including englacial layers, bedrock, noise-limited regions already considered in the literature, as well as a new class: basal ice. The classes are characterized by a specific spatial distribution in the range and azimuth directions. Moreover, the dimension and scale of the targets significantly vary between them. Inland radargrams are processed with an attention U-Net with the ASPP module that automatically extracts relevant features for the semantic segmentation problem. The encoder blocks are connected to the decoder block at the same level with skip connections. AGs filter skip connection signals to remove the irrelevant information. At the bottleneck of the network, the ASPP module extracts features at multiscale levels to improve the identification of targets having different sizes. An argmax operation analyzes the output features to predict labels. The segmentation map is refined using morphological filters. The method effectiveness was tested on two datasets of radargrams by the Multichannel Coherent Radar Depth Sounder (MCoRDS3) RS in the inland of Greenland and Antarctica with basal units in the basal layers.

This article is structured in five sections. Section II defines the segmentation problem. Section III proposes the deep learning method to segment radargrams and describes the related architecture. Section IV illustrates the datasets and the preprocessing. Section V reports the network setup, the evaluation criteria, and the experimental results to prove the effectiveness of the proposed method. Finally, Section VI provides concluding remarks and insights for future works.

II. FORMULATION OF THE PROBLEM

Let us consider radargrams acquired over the central part of ice sheets and caps. We aim at locating meaningful geologic targets for the analysis of the ice sheet and basal boundary conditions, such as the basal unit. Let \mathbf{R} be a 2-D radargram of N_T traces and N_S samples

$$\mathbf{R} = \{P_{dB}(a, r) | a \in [1, \dots, N_T], r \in [1, \dots, N_S]\} \quad (1)$$

where P_{dB} is the power of the reflections stored in the radargrams in decibels and (a, r) are the azimuth and range coordinates, respectively. We aim at segmenting inland icy radargrams into N_C classes representing meaningful geologic targets related to the basal boundary conditions. The classes are defined as $\{\omega_c, c \in [1, \dots, N_C = 4]\}$. Moving down from the surface along the range dimension (see Fig. 1), ω_1 is englacial layers, ω_2 is basal ice, ω_3 is bedrock, and ω_4 represents the noise-limited regions (i.e., thermal noise, EFZ, and signal perturbations, collectively). The EFZ and the thermal noise represent the regions that reflect waves having a lower power than the antenna sensitivity.

Basal ice occurs at the base of the ice sheet, and there may be melt-water at the basal interface, i.e., the interface between the ice and the media below it (e.g., bedrock or till) [20], [39] (see Fig. 1). Close to the core of the basal ice, the layers fold toward the surface, generating large structures with steep edges, called disrupted layers that can extend up to half of the layer pack [11], [12]. However, the layer upper part, close to the surface, is usually not affected by the folding and,

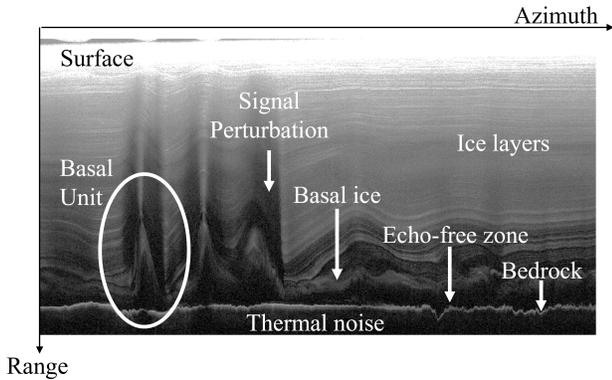


Fig. 1. Part of a radargram from MCoRDS [6] campaign 20170413 in Greenland imaging an ice sheet with three basal units. The labels indicate the geographic targets in the basal layer and the classes for the proposed method.

thus, remains flat. Increasing the steepness of the disrupted layer, the intensity of the signal perturbation increases and masks part of the ice strata (see Fig. 1). It is worth noting that the target dimension largely varies between them. Ice layers cover a larger area than the other classes. The bedrock is typically thinner than the ice pack, the thermal noise, EFZ, and signal perturbation. The basal layer and the signal perturbation dimensions depend on the deformation of the ice sheet bottom. Thus, classes are modeled best at different scales.

Basal units are characterized by a strong reflection, having four main radiometric properties. First, the signal of the folded layers above the basal units is weaker due to the steep slope scattering energy away from the radar rather than back toward it [39]. Second, the basal ice is characterized by diffuse scattering that is often of greater amplitude than the amplitude of the echoes surrounding it [20], [21]. Third, the basal layer shows a loss in the reflection power, given by deformation of the layers, impurities (e.g., soil or stones), and the higher ice temperature [21]. Indeed, basal ice has a higher temperature than the rest of the ice sheet due to geothermal heat flux and hydrological transport of heat (liquid water from the surface). Finally, the last radiometric characteristic regards the bedrock under freeze-on ice that may appear as a specular reflector, i.e., a smooth and bright reflection. This is because, with the basal unit, water is generally present at the basal interface as a thin film or even a lake [10], [11], [20], [21]. The EM interactions with one target also affect the representation of the other targets in the radargram. Therefore, this property should be considered in the data analysis.

III. PROPOSED DEEP-LEARNING-BASED METHOD

This section presents the proposed method for the segmentation of the inland radargrams. The method is based on a supervised CNN that segments the radargram into four classes (see Fig. 2). Here, we consider a U-Net enriched with the ASPP module, and AGs as the architecture previously showed good performance for semantic segmentation [28], [30], [31]. The network learns the characteristics of the classes and automatically extracts relevant features at different scales for the radargram segmentation. Extracted features are semantically meaningful given the generalization capability of the network [40]. The AGs help to focus on the relevant

regions of the radargram by filtering the features input to the decoder from the skip connections [30]. The AGs suppress irrelevant low-level features to better focus on the salient signal features in contrast with the noise-limited regions [41]. The ASPP module extracts multiscale features to better segment objects and classes at different scales [29], such as the ice layers, basal ice, and bedrock that show by a different scale in radargrams. The multiscale features avoid the loss of information in the network compression that generates poor-resolution segmentation maps, i.e., fuzzy boundaries. In addition, ASPP helps the processing of inland radargram to detect classes with different dimensions and scales, such as the thin bedrock and thick layers. The training of the network consists of two stages. In the first step, training minimizes the reconstruction loss between the input and the output to initialize the network parameters and extract features at multiple scales. In the second step, the network is tuned in a supervised fashion to segment radargrams into the classes of interest. Segmentation labels are obtained by the argmax classification of the features of the last layer of the network. Then, the map is refined in the spatial domain using morphological filters.

A. Deep Network Architecture

The proposed network has an autoencoder shape: 1) the encoder compresses the input into a lower dimension and 2) the decoder decompresses it to obtain an output with the exact dimensions as the encoder input. Here, we use a U-Net architecture [28] that showed outstanding performance in the segmentation task coupled with efficient use of the GPU memory [41]. U-Net has skip connections linking the encoder and the decoder (see Fig. 3). Input radargrams \mathbf{R} are split into nonoverlapping patches (h, w) . The patches have the same size as the radargram depth in the range direction, i.e., $h = N_s$. This leads to samples that capture the global context and are fully representative of the class characteristics (see Section II) in terms of the variability, pattern, and vertical spatial distribution. In the azimuth direction, the patch size w is proportional to the average dimension of basal units [42], [43]. After the processing, the output patches are concatenated to have the same shape as the original radargram. In the following, the structure of the network is described in detail.

Let us define the input and output at each point in the network as $H_i^{l,j}$, where $l \in [1, \dots, 4]$ indicates the encoder and decoder block number, j can be either e or f with e indicating blocks of the encoder, and f can be those of the decoder. $H_i^{l,j}$ is the feature set extracted by the l th encoder or decoder block and has dimensions $(M^l \times h^l \times w^l \times M_0)$, where M^l indicates the number of features, w^l and h^l are the dimensions of the features, and M_0 is the input patch channel size and is fixed to $M_0 = 1$. The index i indicates the stage in the block: c stands for the convolutional layer output, b the batch normalization output, r the ReLU activation function output, m the max-pooling output, d the deconvolutional layer output, and u the upsampling output. Finally, H_{ASPP} is the output of the ASPP module and has shape $(M^{ASPP} \times h^{ASPP} \times w^{ASPP})$, where M^{ASPP} indicates the number of features,

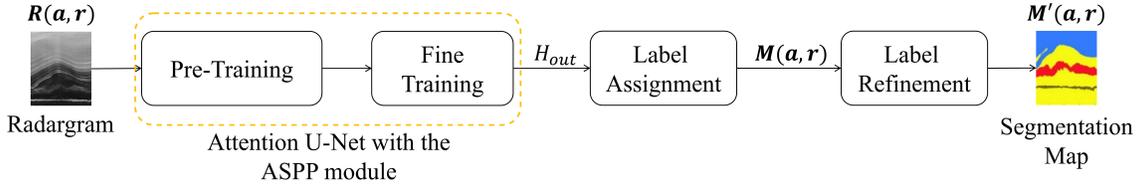


Fig. 2. Flowchart of the proposed deep-learning method for the supervised segmentation of radargrams.

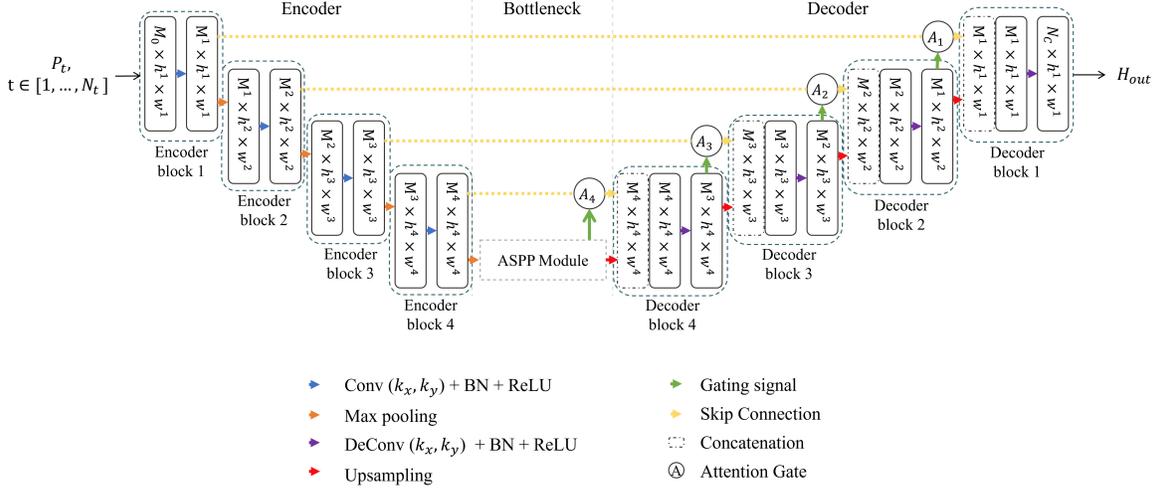


Fig. 3. Architecture of the attention U-Net embedded with the ASPP module.

and h^{ASPP} and w^{ASPP} indicate the height and the width of the features, respectively. H_{AG}^l is the output of the l th AG (i.e., the input of the l th decoder block) and has shape of $(M^{l+1} \times h^{l+1} \times w^{l+1})$.

1) *Encoder*: The encoder is made of N_L downsampling blocks, consisting of two sets each made of a convolution layer, followed by a batch normalization layer and a rectified linear unit (ReLU). Each convolutional layer l learns high-level semantic features from the spatial and semantic information in the data [40]. The first block ($l = 1$) input is a single radargram patch P_t , $t \in [1, \dots, N_t]$. Each block l outputs a set of M^l features for each input pixel followed by downsampling by the max-pooling operation.

To extract the features, the convolutional layer applies a convolution across all features with a kernel of size (k_x, k_y) such that the learnable weight set W^l for block l has size $M^l \times k_x \times k_y \times M_0$. M_0 is the channel size of the input and is fixed to $M_0 = 1$ since the radar images are single-channel log power detected images. The kernel size (k_x, k_y) is the same for all layers. Each channel of the output of the convolutional layer $H_c^{l,e}$ is normalized with a batch normalization layer according to the mean and standard deviation of each batch to obtain $H_b^{l,e}$. An ReLU activation function introduces nonlinearity to obtain $H_r^{l,e}$. After the activation, a max-pooling operation is applied to reduce the features' dimensions and, thus, the number of parameters to learn and the computational cost. The size and stride are both given by (m_x, m_y) , where m_x is the value in azimuth and m_y is the value in range.

The number of output features after max pooling is $M^l \times h^{l+1} \times w^{l+1}$.

2) *Atrous Spatial Pyramid Pooling*: In the bottleneck of the network, the ASPP module, based on the atrous (also called dilated) convolution [29], [44], applies parallel atrous convolution to vary the receptive field of the network filters (see Fig. 3). This helps the network to integrate the global view and focus on the radargram details. Atrous convolution is an operation that contracts or expands the receptive field of the convolutional filters by varying the dilation rate parameter d . d corresponds to the input signal sampling stride, and standard convolution is a particular case with $d = 1$. The dilated rate d indicates that the convolutional filter are enlarged by adding $d - 1$ zeros between the filter coefficients to change the filter resolution (see Fig. 4). The kernel dimension is enlarged from $(k_x \times k_y)$ to $(k'_x \times k'_y)$, where

$$\begin{aligned} k'_x &= k_x + (k_x - 1)(d - 1) \\ k'_y &= k_y + (k_y - 1)(d - 1). \end{aligned} \quad (2)$$

The ASPP module applies a set of $N_{\text{ASPP}} = 4$ parallel atrous convolutions with varying dilation rates (see Fig. 4) and a global average pooling (GAP) (see Fig. 5). Different dilation rate filters extract features at different scales and are concatenated with the GAP map, which includes the global context information. Each atrous block contains: 1) a dilated convolution layer with dilated rate d_i , kernel size (k_x, k_y) , and unitary stride extracting M_{ASPP} features;

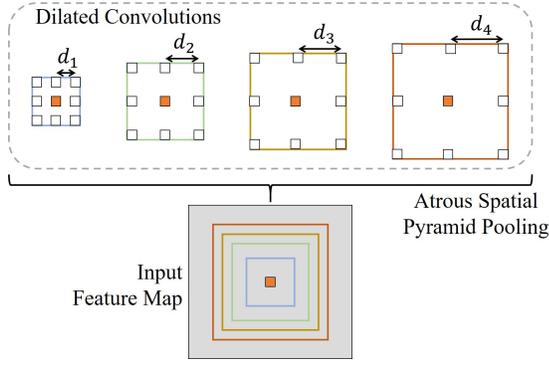


Fig. 4. Representation of the dilated convolution filters with different dilation rate coefficients $[d_1, \dots, d_4]$ applied to the orange pixel. The ASPP module applies parallel dilated convolutions to the input map to extract multiscale features. The different receptive fields of the filters are shown in different colors.

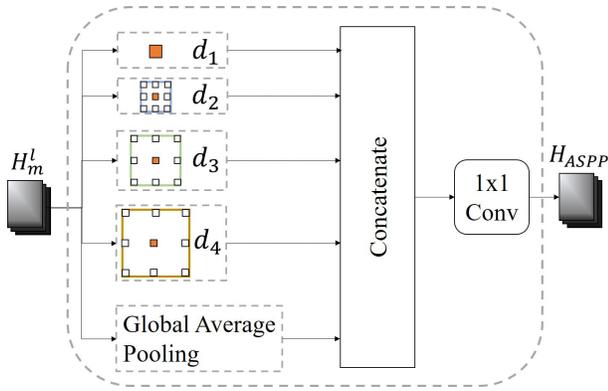


Fig. 5. Structure of the ASPP module. The output of the encoder H_m^l is processed by parallel dilated convolutions with different dilation rate coefficients $[d_1, \dots, d_4]$ and a GAP. All the convolutions are followed by a batch normalization layer and an ReLU. The output of the branches are concatenated and processed by a (1×1) convolution to obtain the output of the ASPP module H_{ASPP} .

2) a batch normalization layer; and 3) an ReLU activation function. The $(N_{ASPP} \times M_{ASPP})$ features from the atrous blocks and the GAP feature are concatenated together and fed to a (1×1) convolution operation extracting M_{ASPP} features.

3) *Decoder and the Attention Gates*: The decoder consists of L expanding blocks having embedded the attention mechanism from [30] to suppress the activation of irrelevant radargram regions by filtering the features from the skip connections. The l th AG takes as input the features $H_m^{l,e}$ from the skip connections and the features from the next lowest layer $(l+1)$ as gating signal g^l . The assumption is that low-level features from the encoder are redundant and poorly represent the data. Features from the decoder better represent the data as they are extracted at a higher level in the network. Hence, decoder features can be used to filter the features $H_m^{l,e}$ from the skip connections and activate the most informative regions [30]. To highlight radargram salient regions, for each pixel, the l th AG module extracts a scalar attention value $\alpha \in [0, 1]$ from $H_m^{l,e}$ considering the gating signal g^l [30]. In multiclass semantic segmentation tasks, AGs extract multidimensional attention coefficients [45].

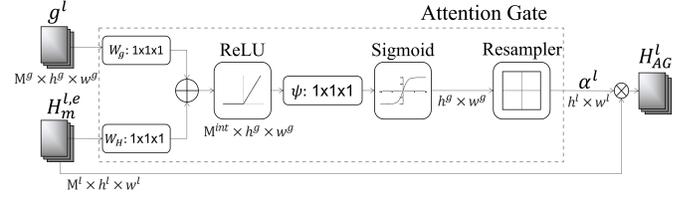


Fig. 6. Structure of the AG module [30] to filter the signal from the skip connection $H_m^{l,e}$. The gate signal g^l highlights the salient spatial regions of the feature set $H_m^{l,e}$ and determines the attention coefficients α^l . The attention coefficients are then elementwise multiplied with the low-level feature set to obtain the output of the l th AG module H_{AG}^l .

As shown in Fig. 6, each AG takes as input the set of lower level feature $H_m^{l,e}$ from the skip connections and the gating vector $g^l \in \mathbb{R}^{M_g}$, where M_g is the number of features. For each level l , the gating vector is the features extracted from the previous layer $(l+1)$. For the AG with $l = N_L$, the gating signal consists of the output features of the ASPP module H_{ASPP} (see Fig. 3), i.e., $g^l = H_{ASPP}$ and $M_g = M_{ASPP}$. For the other AGs ($l \in [1, \dots, N_L - 1]$), the gating signal is the output of the previous layer $(l+1)$, i.e., $g^l = H_r^{l+1,f}$ and $M_g = M^{l+1}$. Each AG applies three linear transformations (W_g , W_H , and Ψ) to the gating signal and the lower level features by using channelwise 1-D convolutions. W_g and W_H map the input features g^l and $H_m^{l,e}$ in a \mathbb{R}^{int} space, while Ψ reduces the feature number from M_{int} to 1. The resulting features are summed to obtain the gating coefficients (additive attention) (see Fig. 6). The gating coefficients are processed by an ReLU and a channelwise 1-D convolution that outputs one feature per pixel, followed by a sigmoid activation function. Finally, the sigmoid output signal is resampled with the bilinear interpolation by an upsampling layer that doubles the dimensions to match the shape of the following decoder block input. This generates a grid signal conditioned to the data spatial information so that the attention coefficient α^l of the l th AG has a shape $(h^l \times w^l)$. Finally, the output of the l th AG H_{AG}^l is the elementwise multiplication $[\cdot]$ between the each of the features in $H_m^{l,e}$ and the attention coefficients α^l , i.e., $H_{AG}^l = H_m^{l,e} \cdot \alpha^l$. The output of the l th AG H_{AG}^l is concatenated with the upsampled version of the gating signal $H_u^{l+1,f}$. For the first decoder block ($l = N_L$), the AG output is concatenated with the upsampled output of the ASPP H_{ASPP} . For the other decoders block ($l \in [1, \dots, N_L - 1]$), the AG output is concatenated with the upsampled output of the previous network layer $H_u^{l+1,f}$.

Decoder blocks $l \in [1, \dots, L]$ consist of two sets each made of a (k_x, k_y) deconvolutional layer and a batch normalization layer, followed by an ReLU (see Fig. 3). Each block takes as input the skip connection signal filtered by the AG and the output upsampled features from the next lowest block. Each block l gives, in output, a set of M^l features $H_d^{l,f}$ of shape $(M^l \times h^l \times w^l)$. In the case of the last block, the output $H_d^{l,f}$ is further processed with a (k_x, k_y) deconvolutional layer and a batch normalization layer, followed by an ReLU. This generates the output of the network H_{out} that has a shape $(M^{\text{out}} \times h \times w)$. The value of M^{out} depends on the task of interest, as described in the following.

B. Pretraining of the Network

All the network weights can be learned by training the network according to the optimization of a loss function. Using randomized parameters is less efficient than using a pretraining phase [46]. Hence, this step aims at initializing the network parameters by training the network to learn the characteristics of the radar data. To this end, we aim at training the network to learn how to reconstruct a given input radargram patch. This is a regression task as the network should reconstruct the values of each pixel of the input radargram at the output. Therefore, the number of features expected in the network output is fixed to $M^{\text{out}} = 1$. For training, we use a regression loss. Here, we use the mean square error (mse), defined in (3), but any regression loss from the literature can be used

$$(\text{mse}) = \frac{1}{N_S N_T} \sum_{a=1}^{N_T} \sum_{r=1}^{N_S} [\mathbf{R}_{\text{pred}}(a, r) - \mathbf{R}(\mathbf{a}, \mathbf{r})]^2. \quad (3)$$

The mse estimates the average squared difference between the predicted values $\mathbf{R}_{\text{pred}}(\mathbf{a}, \mathbf{r})$ and the actual values $\mathbf{R}(\mathbf{a}, \mathbf{r})$ of the log power radargrams. This metric gives an estimation of the goodness of the reconstruction.

C. Generation of the Segmentation Map

After initializing the network parameters, we train the network in a supervised manner to segment the radargram into ω_1 englacial layers, ω_2 basal ice, ω_3 bedrock, ω_4 thermal noise, EFZ, and signal perturbations. The last deconvolutional layer of the network extracts $M^{\text{out}} = N_c$ features, where N_c is equal to the number of classes of the segmentation task. This article considers four classes as we investigate inland icy radargrams and characterize the basal layer. However, the network is flexible enough to analyze more classes [40] if a sufficient labeled dataset for training exists.

1) *Loss Function*: For the supervised segmentation, we use the cross-entropy loss function between the radargram and reference labels. For each pixel (a, r) in the patch, the cross-entropy loss function is applied to the predicted label $\omega_c(a, r)$ and the reference label $\omega_{r_l}(a, r)$ as follows:

$$\ell_t(a, r) = - \sum_{c=1}^{N_c} \omega_{r_l}(a, r) \log[\omega_c(a, r)] \quad (4)$$

where N_c is number of classes of the segmentation task. The loss term \mathcal{L}_t for all the pixels of the radargram is computed as follows:

$$\mathcal{L}_t = \frac{1}{N_S N_T} \sum_{a=1}^{N_T} \sum_{r=1}^{N_S} \ell_t(a, r). \quad (5)$$

After processing the radargram \mathbf{R} , we obtain a matrix of M^{out} deep features $H_{\text{out}}(a, r, k)$, where $k \in [1, \dots, M_{\text{out}}]$.

2) *Label Assignment*: Analyzing a radargram \mathbf{R} with such a network, semantically similar pixels produce high values in the same deep features. Hence, for each pixel (a, r) , the label ω_c is defined by choosing the output feature with the maximum value [47]

$$\omega_c(a, r) = \underset{k \in [1, \dots, M_{\text{out}}]}{\text{argmax}} \{H_{\text{out}}(a, r, k)\}. \quad (6)$$

The segmentation map \mathbf{M} is constructed from the individual patch outputs and is defined as

$$\mathbf{M} = \{M(a, r) | a \in [1, \dots, N_T], r \in [1, \dots, N_S]\} \quad (7)$$

where the labels for $M(a, r)$ are assigned according to (6).

3) *Label Refinement*: In semantic segmentation, pixels are expected to be spatially correlated. Even if this property is partially ensured by the network convolutional nature, we refine the segmentation map to obtain \mathbf{M}' with morphological attribute operators [48]. To enhance the spatial consistency of the segmentation map, we applied the morphological opening operator followed by the morphological closing operator that is both based on a sequence of dilation and erosion [48]. Opening removes objects and closing fills holes that are smaller than the structuring element (SE), while structures that are larger than the SE are not modified [48]. The SE defines the neighborhood used to process each pixel, i.e., the moving window used in the transformations and can have different shapes, such as circular. The morphological opening Γ^T consists of two operations that: 1) preserve the connected region containing the pixel (a, r) and 2) preserve or remove the connected region based on the evaluation of a criterion T . The opening attribute $\gamma^T(\mathbf{M})$ can be formulated for each pixel (a, r) and each level $x \in [1, \dots, N_c]$ of the gray image as in [48]

$$\gamma^T(\mathbf{M})(a, r) = \max \left\{ x : (a, r) \in \bigcup_{(a,r) \in T_x(\mathbf{M})} \Gamma^T[\text{Th}_x] \right\} \quad (8)$$

where Th_x represents the binary image obtained by thresholding the segmentation map \mathbf{M} at each level x , with x ranging on the gray levels of \mathbf{M} . Similarly, the morphological attribute closing $\phi^T(\mathbf{M})$ can be defined as

$$\phi^T(\mathbf{M})(a, r) = \max \left\{ x : (a, r) \in \bigcup_{(a,r) \in T_x(\mathbf{M})} \Phi^T[\text{Th}_x] \right\}. \quad (9)$$

The morphological profile, i.e., the refined segmentation map, is calculated by applying sequential filtering, where the criterion T is evaluated at each iteration.

Given that \mathbf{M} is a digital gray level image of N_c levels (one for each class) defined in the \mathbb{Z}^2 domain, we can apply an opening operator to \mathbf{M} to obtain the morphological opening attribute profile $\Pi_{\gamma^T}(\mathbf{M})$ defined as

$$\Pi_{\gamma^T}(\mathbf{M}) = \{ \Pi_{\gamma^T} : \Pi_{\gamma^T} = \gamma^T(\mathbf{M}) \}. \quad (10)$$

$\Pi_{\gamma^T}(\mathbf{M})$ is further processed by a closing filter to obtain the morphological closing attribute profile $\Pi_{\psi^T}(\mathbf{M})$, i.e., the refined segmentation map \mathbf{M}' , as

$$\begin{aligned} \mathbf{M}' &= \Pi_{\psi^T}(\Pi_{\gamma^T}(\mathbf{M})) \\ &= \{ \Pi_{\psi^T} : \Pi_{\psi^T} = \psi^T(\Pi_{\gamma^T}(\mathbf{M})) \}. \end{aligned} \quad (11)$$

IV. DATASET DESCRIPTION

This section describes the datasets used to assess the effectiveness of the proposed method. We consider two datasets acquired over the polar areas of the Earth by the MCoRDS3

TABLE I

PARAMETERS OF THE RS INSTRUMENT MCoRDS [6] FOR THE CAMPAIGNS ACQUIRED IN 2017, 2014, AND 2018 AND THE GEOMETRICAL RESOLUTIONS OF THE RADARGRAMS

Parameters	2017 campaigns	2014 and 2018 campaigns
Central frequency f_c	195 MHz	190 MHz
Wavelength λ	1.54 m	1.58 m
Bandwidth BW	30 Mhz	50 MHz
Sampling Frequency	30 MHz	50 MHz
Transmitted Power P_{tr}	1050 W	6 kW
Aircraft Altitude AGL h	~500 m	~500 m
Range Resolution in Ice	4.3 m	2.6 m
Azimuth Resolution	27.5 m	27.5 m
Azimuth sampling rate	15 m	15 m

RS [6] mounted on an aircraft (see Table I for the radar parameter details). The first dataset consists of radargrams acquired in Northern Greenland. The second dataset consists of radargrams acquired in Antarctica. As reference data, samples representing the four classes were manually picked from the radargrams. These regions are selected in a way that there is no ambiguity regarding the target class. To this end, we perform the selection based on examples of radargrams visually inspected and available in the literature [10], [11], [23], [42], [43].

A. Preprocessing of the Radargrams

The RS data [6] are distributed by the Center for Remote Sensing of Ice Sheets (CREGIS) for NASA Operation Ice-Bridge. The radar data are already range compressed, azimuth focused via synthetic aperture radar (SAR) techniques, and the clutter is partially compensated to get the resolutions shown in Table I. To prepare the radar data, we concatenated adjacent acquired radargrams with attention to removing overlapping regions. Moreover, we corrected the data for the fluctuations due to the movements of the platform. While acquiring the data, an aircraft has three undesired movements, i.e., pitch, roll, and yaw, which influences and degrades the quality of the radargram signal. To mitigate the signal degradation, we normalized the power of the radargram concerning the surface power for each column. This helps reduce the fading of the power due to the rolling movement of the aircraft on the curves. Considering the preprocessing, the vector of the range coordinate of the surface $s(a)$ was identified as the maximum power for each azimuth column a . Hence, the power of the surface $P_s(a)$ of each azimuth trace was defined as $P_s(a) = \mathbf{R}(\mathbf{a}, s(a))$. The region above the surface is the free space and was removed by masking it as we aim to study the subsurface. Radar data are affected by speckle, which can be approximated as multiplicative noise. Hence, by applying a log-transformation to the power radargram, the noise can be approximated as an additive, reducing the complexity of handling the data and enhancing the spatial information.

B. Generation of the Datasets

Adjacent radargrams are concatenated and preprocessed with the steps described above. Next, each radargram is

TABLE II

LIST OF THE CAMPAIGNS AND THE RADARGRAM IDS USED IN THE SUPERVISED TRAINING FOR EACH DATASET AND THE NUMBER OF TRACES N_T FOR EACH CAMPAIGN

	Campaign	Radargram ID	N_T
Dataset 1	2017_0403_01	6–16	35551
	North Greenland 2017_0413_01	36–56	70026
Dataset 2	2014_1114_02	12–13, 21–24	19992
	West Antarctica 2018_1104_01	4–5	6668
	2018_1020_01	4, 8–10, 15–18, 21–22	30009

divided into patches that are normalized by removing the mean of the dataset and dividing by the standard deviation. For both the North Greenland and Antarctica data, we generated two datasets \mathcal{J}^1 and \mathcal{J}^2 . \mathcal{J}^1 is used for the pretraining and consists of the preprocessed patches. \mathcal{J}^2 is used for the supervised training and consists of pairs of preprocessed and manually labeled patches. Table II lists the IDs of the campaign and radargrams used in \mathcal{J}^2 for both Northern Greenland and west Antarctica datasets. Both datasets \mathcal{J}^a , $a = [1, 2]$ are further divided into training \mathcal{J}_{tr}^a and test \mathcal{J}_{te}^a patches so that 90% of the patches are used for the training and the remaining 10% for the test phase.

1) *Dataset 1: North Greenland*: Fig. 7(a) shows the ground track of the campaigns used for both datasets \mathcal{J}^1 and \mathcal{J}^2 : the radargrams selected for \mathcal{J}^1 are in red, and those for \mathcal{J}^2 are in blue and green; they belong to the campaigns 20170403_01 and 20170413_01, respectively. Dataset \mathcal{J}^1 for Northern Greenland consists of 282 radargrams imaging grounded ice acquired in several campaigns over Greenland in 2017. The radargrams are preprocessed to generate concatenated radargram with $N_T^1 = 930600$ traces. The radargrams are divided into patches of size $w = 64$ and $h = 1280$ in the range and azimuth directions, respectively. h is set so that the neural network has a complete view of the vertical distribution of the classes. Dataset \mathcal{J}^1 consists of 14541 patches, which are divided into the dataset \mathcal{J}_{tr}^1 of 13087 patches for the training phase, and \mathcal{J}_{te}^1 of 1454 patches for the test phase. Dataset \mathcal{J}^2 consists of 30 radargrams acquired in the north of Greenland in 2017 (see Table II). We consider three campaigns imaging a region where the basal layer is strongly visible, and there is evidence of basal units [12], [42]. The radargrams are preprocessed to generate a concatenated radargram with $N_T^2 = 99000$ traces, which is divided into patches. Dataset \mathcal{J}^2 consists of 1547 patches before data augmentation. We perform the data augmentation to increase by three times the number of labeled patches. The patches in \mathcal{J}^2 are divided into the set \mathcal{J}_{tr}^2 of 4641 patches and \mathcal{J}_{te}^2 of 465 patches.

2) *Dataset 2: West Antarctica*: Dataset 2 consists of radargrams acquired in 12 campaigns by MCoRDS [6] imaging grounded ice in the west of Antarctica in 2014 and 2018. Fig. 7(b) shows the ground track of the campaigns used for both datasets \mathcal{J}^1 and \mathcal{J}^2 : the radargrams selected for \mathcal{J}^1 are in red, and those for \mathcal{J}^2 are in blue, green, and black and belong to the campaigns 20141114_02, 20181020_01, and 20181104_01, respectively. Dataset \mathcal{J}^1 consists of 207 radargrams that are preprocessed to generate concatenated

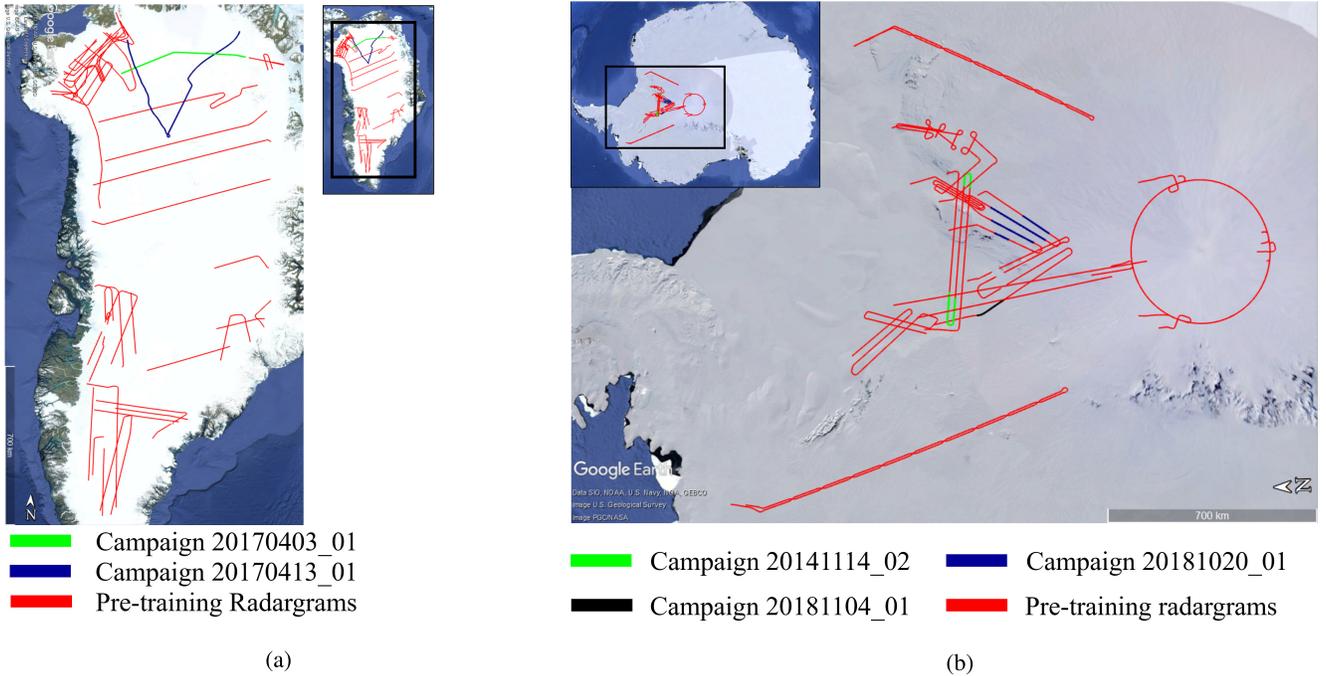


Fig. 7. Ground tracks of the campaigns used for (a) dataset 1 and (b) dataset 2. The ground track of the radargrams selected for the pretraining is in red for both datasets. Radargrams for the supervised training of dataset 1 are from campaign 20170403_01 [in blue in (a)] and campaign 20170413_01 [in green in (a)] in North Greenland. Radargrams for the supervised training of dataset 2 are from campaign 20141114_02 [in green in (b)], campaign 20181020_01 [in blue in (b)], and campaign 20181104_01 [in black in (b)] in West Antarctica.

radargrams with $N_T^1 = 724\,500$ traces. The concatenated radargram is divided into patches of size $w = 64$ and $h = 1680$ in the range and azimuth directions, respectively. Note that the h -dimension of the patches from the Antarctica dataset is greater than for the other dataset as, on average, the Antarctica ice sheet is thicker. Dataset \mathcal{J}^1 consists of 11 320 patches that are divided into the dataset \mathcal{J}_{tr}^1 of 10 188 patches for the training phase and \mathcal{J}_{te}^1 of 1132 patches for the test phase. Dataset \mathcal{J}^2 consists of 15 radargrams acquired in West Antarctica in 2018 and 2014 (see Table II). We consider three campaigns imaging a region where the basal layer and basal units are strongly visible [10], [23]. The radargrams are preprocessed to generate a concatenated radargram with a number of traces $N_T^2 = 56\,669$, which is divided into patches. Dataset \mathcal{J}^2 consists of 885 patches before data augmentation. We perform the data augmentation to increase the number of labeled patches by five times. The patches in \mathcal{J}^2 are divided into the set \mathcal{J}_{tr}^2 of 3983 patches and \mathcal{J}_{te}^2 of 442 patches.

C. Data Augmentation

To increase the network invariance and robustness when the number of available training samples is small, data augmentation is essential to generate a larger labeled dataset [28]. Accordingly, here, as the number of labeled samples for the training is limited (see Table II), we use standard data augmentation techniques (i.e., affine transformations and horizontal flips) to increase the dimension of the labeled dataset. The transformations are designed so that the geometrical properties of the geologic subsurface structures are preserved, and the

augmented data are realistic. Here, we apply random horizontal flips, rotation, and elastic deformations to the original and labeled patches to generate new pairs. Random horizontal flip transformation is applied with a probability $p_f = 0.5$. Random rotation transformation is applied with a probability $p_r = 0.8$ with the rotation degrees in the range $[-\theta_{\max}, +\theta_{\max}]$, where θ_{\max} is the maximum steepness angle of the surface. The random elastic deformation transformation is applied with a probability $p_e = 0.9$ to the vertical Δ_r (range) and horizontal Δ_a (azimuth) directions. We generate smooth deformations applying a random displacement field based on coarse grids of dimensions 8×8 and 16×16 cells. For each cell and directions Δ_r and Δ_a , the displacement vectors are sampled from a Gaussian distribution with standard deviation σ_e . σ_e is sampled from another Gaussian distribution defined as $\sigma_e \sim \mathcal{N}(\mu'_e = 8; \sigma'_e = 0.6)$. The stress field is applied to the patches by moving each pixel to a new position and using spline interpolation of order one [49]. We apply a mirroring resampling filter for mitigating the low pass effect at the border pixels. The parameters of the Gaussian distribution are experimentally chosen so that the deformations in the augmented patches are realistic. Fig. 8 shows two random augmentation outputs from the patch in Fig. 8(a).

V. EXPERIMENTAL SETUP AND RESULTS

A. Experimental Setup

1) *Network Setup*: Table III shows the parameters of the network. We used four downsampling blocks in the encoder and four upsampling blocks in the decoder. The convolutional kernels are of size $(k_x = 3, k_y = 3)$, and the upsampling

TABLE III
PARAMETERS OF THE PROPOSED METHOD AND THE NEURAL NETWORK

Parameter	Description	Dataset 1	Dataset 2
(h, w)	Patch size in the range and the azimuth	(64, 1280)	(64, 1680)
M_0	Number of channels of the input data	1	1
N_L	Number of downsampling and upsampling blocks, and AGs	4	4
$M_l, l = [1, \dots, 4]$	Features extracted by the l -th block of the encoder and the decoder	[64, 128, 256, 512]	[64, 128, 256, 512]
(k_x, k_y)	Convolutional kernel size in the range and azimuth	(3, 3)	(3, 3)
(m_x^l, m_y^l)	Maxpooling and upsampling size in range and azimuth	(2, 2)	(2, 2)
M_{ASPP}	Features extracted by each atrous block and the ASPP module	512	
$d_i, i = [1, \dots, 4]$	Dilate rate parameters of each atrous block	[1, 6, 12, 18]	
ℓ_r	Learning rate	0.0001	0.0001
\mathcal{E}	Maxium number of epochs for training phase	100	100
N_B	Mini-batch size	16	16

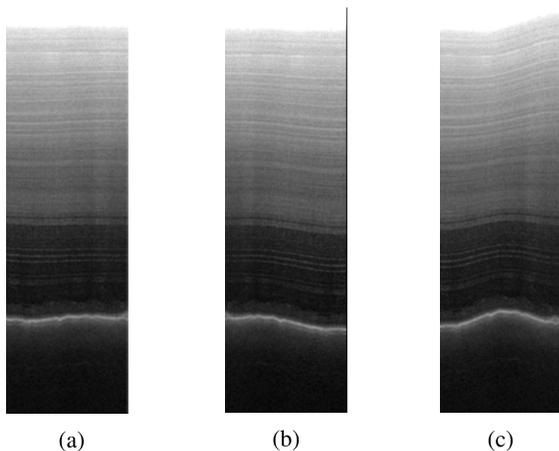


Fig. 8. Example of the patches generated by data augmentation step: (a) original patch from radargram 36 of campaign 20170413_01, (b) random outcome 1 of augmenting patch (a), and (c) random outcome 2 of augmenting patch (a).

and downsampling are with $m_x^l = 2, m_y^l = 2$. The ASPP consists of $N_{ASPP} = 4$ dilated convolutions with the dilation rate parameters equal to $[d_1 = 1, d_2 = 6, d_3 = 12, d_4 = 18]$. At each epoch of the training, a small batch of N_B patches is given input to the network to accelerate the training phase. For this reason and memory constraints, the minibatch size is set equal to $N_B = 16$ for both datasets. For the optimization, we used Adam optimization as it shows good results in the literature [50]. The learning rate is set to $\ell_r = 0.0001$ after running several experiments varying the values from 0.1 to 10^{-6} . To reduce the overfitting and increase the generalization capability of the network, the training stop conditions are based on: 1) the maximum number of epochs, which is set to $\epsilon = 100$ and 2) the training and validation losses at each epoch. A validation loss higher than the training loss indicates that the model is overfitting. Hence, the training is stopped when the validation loss is higher than the training loss, and the model trained in the previous epoch is considered for the next steps.

2) *Comparison Methods*: To assess the effectiveness of each element of the proposed method, we compare the results with those of the baseline methods: the U-Net [28], the U-Net with

AGs [30], and the U-Net with AGs and ASPP without data augmentation. We also compare the results of the proposed method with the literature methods for the segmentation of inland icy radargrams based on handcrafted features processed by the SVM classifier in [15] and [16].

B. Evaluation Criteria

To evaluate the performance of the proposed method, we considered two sets of metrics: one for the reconstruction of the radargrams and the other for the supervised segmentation task. For the reconstruction task, we consider the mse between the input and the output radargrams, as defined in (3). For the semantic segmentation task, we consider how well the pixels of each class are classified. In the inference phase, N_{tot} pixels are analyzed with the proposed semantic segmentation method. For each pixel, the method predicts a label that can be compared with the reference data. TP^c is defined as the number of true positive pixels, i.e., the pixels correctly classified for the c th class and TN^c the true negative, i.e., the number of pixels correctly classified as belonging to one of the other $N_C - 1$ classes. FN^c is defined as the false negative, i.e., the number of pixels belonging to the class ω_c but wrongly labeled, and the FP^c as the false positive, i.e., the number of pixels not belonging to the class ω_c but labeled with ω_c . The sensitivity indicates the probability that a pixel is classified as the c th class, given that it actually belongs to that class

$$\text{Sensitivity} = \frac{TP^c}{TP^c + FN^c}, \quad c \in [1, \dots, N_C]. \quad (12)$$

The specificity indicates the probability that a pixels is not labeled as ω_c , given that it does not belong to that class

$$\text{Specificity} = \frac{TN^c}{TN^c + FP^c}, \quad c \in [1, \dots, N_C]. \quad (13)$$

Finally, the OA is defined as the number of correctly classified pixels TP over the number of processed pixels, i.e., $OA = TP/N_{tot}$.

C. Dataset 1: North Greenland

We first pretrain the network with the dataset \mathcal{J}^1 to initialize the network parameters to extract relevant features for the

TABLE IV
SEGMENTATION PERFORMANCE OF THE PROPOSED METHOD, DEEP BASELINE METHOD, AND LITERATURE METHODS FOR RADARGRAM SEGMENTATION APPLIED TO THE GREENLAND DATASET, WHERE ω_1 IS ENGLACIAL LAYERS, ω_2 IS BASAL ICE, ω_3 IS BEDROCK, ω_4 IS THERMAL NOISE, EFZ, AND SIGNAL PERTURBATIONS, AND ω_4^R IS THE REDUCED VERSION OF ω_4 THAT CONSIDERS THERMAL NOISE AND THE EFZ AS IN [15] AND [16]

Method	Metric	Overall	ω_1	ω_2	ω_3	ω_4	ω_4^R
Proposed Method	Accuracy	0.9837	0.9872	0.9811	0.9801	0.9863	0.9875
	Sensitivity	0.9882	0.9957	0.9861	0.9817	0.9895	0.9881
	Specificity	0.9871	0.9952	0.9824	0.9831	0.9879	0.9899
Proposed method without data augmentation	Accuracy	0.8375	0.8516	0.8306	0.8223	0.8453	0.8568
U-Net with AGs [30]	Accuracy	0.9584	0.9629	0.9589	0.9497	0.9621	0.9660
U-Net [28]	Accuracy	0.9182	0.9288	0.9150	0.9036	0.9254	0.9288
Handcrafted features and SVM [15, 16]	Accuracy [16]	0.9821	0.9868	–	0.9751	–	0.9845
	Accuracy [15]	0.9817	0.9898	0.9723	0.9832	–	0.9815
Extended SVM [15]	Accuracy	0.9524	0.9942	0.9821	0.9986	0.8347	0.9872

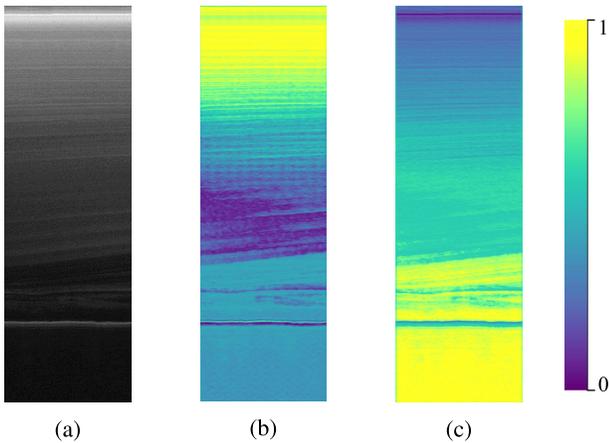


Fig. 9. Example of the features extracted from different convolutional layers for the patch in (a) with the network pretrained only to reconstruct the input radargram. The patch in (a) is part of radargram 20170413_01, feature (b) is extracted from decoder block 3, and (c) from decoder block 4.

radar data. Fig. 9 shows examples of features extracted from different convolutional layers of the network decoder. Note that the features have a spatial pattern that is similar to the original patch [see Fig. 9(a)], even if the values of the features (represented by the color map) strongly vary. Note that the top skip connection does not thwart the pretraining of the rest of the network since the features extracted by the lower blocks $N_l, l \in [2, 4]$ are meaningful [see Fig. 9(b)]. This is because the top AG filters the features from the top skip connection (i.e., the features extracted by the first encoder block) by focusing only on relevant regions. At the end of pretraining, the reconstruction error is equal to $\text{mse} = 0.0110$. Then, we train the neural network with the labeled dataset \mathcal{J}^2 to extract N_c semantically meaningful features automatically. The labels are assigned to each pixel (a, r) according to (6) and refined by applying a morphological disk-shaped SE of radius 3. For an example, we show radargrams 43–46 of campaign 20170413_01 of dataset 1 in Fig. 10(a) and the related segmentation map in Fig. 10(b). Fig. 11 shows a detail of dataset 1 (part of radargram 20170413_01-44) that contains

two basal units. From qualitative analysis, basal ice and signal perturbation classes are better and finer segmented with the proposed method than with that in [15], which is based on handcrafted features processed by the SVM classifier. The handcrafted features were manually designed for modeling the properties of englacial layers, basal ice, bedrock, thermal noise, and EFZ. Note that the handcrafted features are not designed for detecting signal perturbation. Table IV shows the accuracy of the proposed method for dataset 1 calculated considering as reference the pixels with manually assigned unambiguous labels (e.g., by leaving out the pixels along borders between the targets/classes) to avoid including possible errors in the quantitative evaluation. The accuracy for the basal ice is 98.11%, while that for the signal perturbation is 98.63%. The accuracy of the other classes (i.e., bedrock and ice layers) is comparable or slightly lower than those of the state-of-the-art methods in [15] and [16] (see Table IV). To better understand the effectiveness of the proposed method in extracting semantically meaningful features for the segmentation problem, we further compared the classification results of the proposed method with those of the method presented in [15] for segmenting the noise-limited regions. The last row of Table IV reports the results of the analysis of dataset 1 with the method in [15], including the new subclass, i.e., signal perturbation. The method in [15] achieves high accuracy for the classes considered in feature design. However, as expected, the accuracy is low for class ω_4 , including pixels of the signal perturbation, EFZ, and thermal noise. Furthermore, the segmentation accuracy ω_4^R , i.e., the reduced version of ω_4 that considers the thermal noise and the EFZ classes as in [15] and [16], points out that, for the handcrafted feature method, the accuracy of the nonmodeled class is poor. On the contrary, unlike other literature methods, the proposed method can automatically extract semantically meaningful features that accurately discriminate all the classes without the need for manual design, which is time demanding and requires specific expertise. This means that, if we add or remove target classes in the problem, the method can adaptively and automatically extract new optimal features without the need for a complex

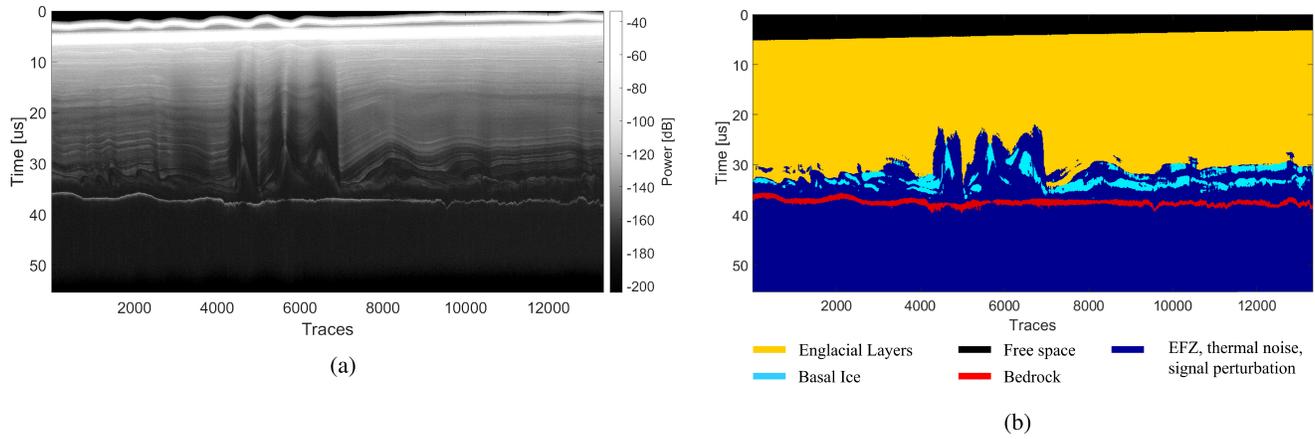


Fig. 10. Example of the method applied to part of dataset 1 (radargrams 43–46) of the campaign 20170413_01: (a) input radargram and (b) segmented radargram. Engacial layers are in yellow; the basal ice is in light blue; the bedrock is in red; and the signal perturbation, EFZ, and thermal noise are in dark blue.

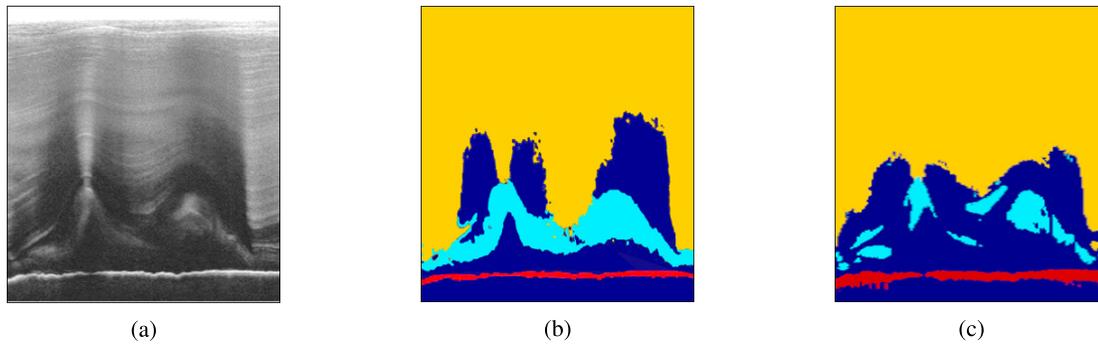


Fig. 11. Part of radargram 44 of the campaign 20170413_01: (a) input radargram, (b) segmented radargram with [15] considering the signal perturbation class, and (c) segmented radargram with the proposed method. Engacial layers are in yellow; the basal ice is in light blue; the bedrock is in red; and the signal perturbation, EFZ, and thermal noise are in dark blue.

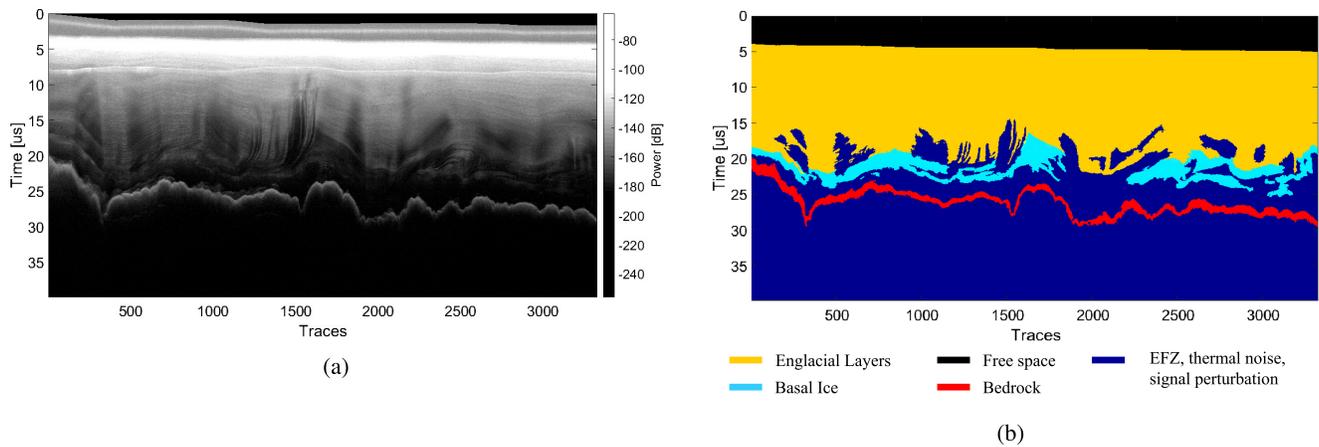


Fig. 12. Example of the method applied to part of dataset 2 (radargram 23) of the campaign 20141114_02: (a) input radargram, (b) segmented radargram. Engacial layers are in yellow; the basal ice is in light blue; the bedrock is in red; and the signal perturbation, EFZ, and thermal noise are in dark blue.

and lengthy process of design, extraction, and selection of handcrafted features. Table IV presents the accuracy of the proposed method and the U-Net, as well as the U-Net with AGs to understand better the effectiveness of the ASPP, the AGs, and the data augmentation elements. The table shows that the proposed method outperforms the three baseline methods

for every class. The accuracy of the proposed method without data augmentation is extremely low, confirming the importance of data augmentation with small labeled datasets. AGs increase the U-Net performance for all the classes, even if the accuracy of the bedrock and basal ice is lower than those of the layers and the noise-limited regions due to the different prior

TABLE V

SEGMENTATION PERFORMANCE OF THE PROPOSED METHOD, DEEP BASELINE METHOD, AND LITERATURE METHODS FOR RADARGRAM SEGMENTATION APPLIED TO THE ANTARCTICA DATASET, WHERE ω_1 IS ENGLACIAL LAYERS, ω_2 IS BASAL ICE, ω_3 IS BEDROCK, ω_4 IS THERMAL NOISE, EFZ, AND SIGNAL PERTURBATIONS, AND ω_4^R IS THE REDUCED VERSION OF ω_4 AS IT CONSIDER THE THERMAL NOISE AND THE EFZ AS IN [15] AND [16]

Method	Metric	Overall	ω_1	ω_2	ω_3	ω_4	ω_4^R
Proposed Method	Accuracy	0.9843	0.9881	0.9855	0.9789	0.9848	0.9819
	Sensitivity	0.9885	0.9921	0.9859	0.9860	0.9901	0.9889
	Specificity	0.9853	0.9902	0.9862	0.9817	0.9831	0.9874
Proposed method without data augmentation	Accuracy	0.7847	0.7913	0.7865	0.7598	0.8012	0.8001
U-Net with AGs [30]	Accuracy	0.9569	0.9631	0.9577	0.9469	0.9600	0.9551
U-Net [28]	Accuracy	0.9169	0.9205	0.9122	0.9087	0.9263	0.9301
Handcrafted features and SVM [15, 16]	Accuracy [16]	0.9837	0.9966	–	0.9751	–	0.9793
	Accuracy [15]	0.9846	0.9858	0.9891	0.9774	–	0.9859
Extended SVM [15]	Accuracy	0.9566	0.9868	0.9771	0.9885	0.8739	0.9846

probabilities of the classes. The ASPP increases the accuracy of the bedrock and basal ice, extracting multiscale features and mitigating the problem of imbalanced classes.

D. Dataset 2: West Antarctica

For the Antarctica dataset, we perform the same steps as for the Greenland dataset. The pretraining with \mathcal{J}^1 gives a reconstruction error equal to $\text{mse} = 0.00938$. For an example, we show radargram 23 of campaign 20141114_02 in Fig. 12(a) and the related segmentation map in Fig. 12(b). Table V reports the performance of each method for dataset 2. The accuracy for the basal ice is 98.55%, while that for the signal perturbation is 98.48%. The accuracy of the other classes (i.e., bedrock and ice layers) are comparable with the state of the art in [15] and [16] (see Table V). Also, for this dataset, the accuracy of the handcrafted features and the SVM classifier is comparable with the proposed method for the classes for which the features are designed including the layers (ω_1) and the EFZ and thermal noise (ω_4^R) (see Table V). As for dataset 1, class ω_3 has a slightly lower accuracy for the proposed method than for the literature method in [15] as it is the class with the lowest number of labeled samples. For the classes not considered in the feature extraction design step (such as the signal perturbation), the proposed method is performing better: the accuracy of class ω_4 (signal perturbation, thermal noise, and EFZ) for the proposed method is equal to 98.77%, and that for the SVM method in [15] is 87.39%. Considering that the accuracy for the thermal noise and EFZ (class ω_4^R) is 98.46% with the SVM method in [15], the handcrafted features may be limited in their ability to model targets not considered in the design step. On the contrary, the proposed method can automatically extract semantically meaningful features in the learning phase. Table V shows that, also for dataset 2, all the techniques had the expected improvements. In addition, the table shows that the accuracy of the proposed method significantly improves when data augmentation is applied. The performance also improves with respect to the baseline U-Net by adding the AGs and the ASPP module: the network can

extract semantically more meaningful features and mitigate the problem of the strongly unbalanced classes.

VI. CONCLUSION

We proposed a fully automatic method for detecting the basal layer and the basal units in airborne inland radargrams that segments radargrams into four classes: 1) englacial layering; 2) basal ice; 3) bedrock; and 4) noise-limited regions, including the EFZ, thermal noise, and signal perturbation. We proposed to use a U-Net with embedded AGs and the ASPP module to extract relevant features for segmenting radargrams. The AGs filter irrelevant features to focus the network on critical areas of the radargrams. The ASPP module extracts features with different receptive fields, improving the identification of targets of different dimensions and scales, such as the basal ice, the ice layers, and the bedrock. The network weights are initialized with pretraining for better handling the properties of the radar signal. Next, the attention U-Net is trained to extract features for labeling. Finally, morphological filters are refined the predicted labels. The method is tested on two datasets of radargrams acquired in Greenland and Antarctica, where the basal layer and basal units are visible. We applied the method to vertical strides of the radargram (patches) to capture the global context and the classes' vertical spatial distribution. The proposed method requires the availability of a labeled dataset training set that was manually generated for this work. The results prove that the method can assign semantically accurate labels, and each method element had the expected improvements. The proposed method demonstrated the ability to identify the defined basal ice and signal perturbation classes with high accuracy. The other classes, i.e., the englacial layers, the bedrock, the EFZ, and the thermal noise, are segmented with a high accuracy that is comparable to the literature methods without requiring the feature extraction design procedure. As the technique does not use handcrafted features, it is adaptable to analyze data acquired in different scenarios, such as coastal areas that are characterized by floating ice, marine and meteoric ice, and the grounding areas.

In future works, we also plan to test the proposed method on other icy areas to detect different targets, such as ice shelves and crevasses in coastal regions, which have highly different geometric and radiometric properties than those in the inland areas. We plan to investigate the possibility to adapt the model trained on the Antarctica dataset to perform segmentation on the Greenland dataset, and vice versa, by means of domain adaptation techniques that consider the different geologic properties (e.g., the thickness of the ice and depth of the bedrock) of the two domains. Finally, we plan to investigate other approaches for semisupervised training to improve the network performance while reducing the number of labeled samples required for supervised training.

ACKNOWLEDGMENT

The authors acknowledge the use of data and data products from the Center for Remote Sensing of Ice Sheets (CRE SIS) generated with support from the University of Kansas, NASA Operation IceBridge Grant NNX16AH54G, NSF Grants ACI-1443054, OPP-1739003, and IIS-1838230, Lilly Endowment Inc., and Indiana METACyt Initiative.

REFERENCES

- [1] M. M. Bennett and N. F. Glasser, *Glacial Geology: Ice Sheets and Landforms*. Hoboken, NJ, USA: Wiley, 2011.
- [2] R. E. Bell, M. Studinger, C. A. Shuman, M. A. Fahnestock, and I. Joughin, "Large subglacial lakes in East Antarctica at the onset of fast-flowing ice streams," *Nature*, vol. 445, no. 7130, p. 904, 2007.
- [3] D. M. Schroeder, D. D. Blankenship, and D. A. Young, "Evidence for a water system transition beneath Thwaites Glacier, West Antarctica," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 30, pp. 12225–12228, 2013.
- [4] O. Alemany *et al.*, "The SUBGLACIOR drilling probe: Concept and design," *Ann. Glaciol.*, vol. 55, no. 68, pp. 233–242, 2014.
- [5] Q. Schiermeier, "Speedy Antarctic drills start hunt for Earth's oldest ice," *Nature News*, vol. 540, no. 7631, p. 18, 2016.
- [6] F. Rodriguez-Morales *et al.*, "Advanced multifrequency radar instrumentation for polar research," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2824–2842, May 2014.
- [7] S. Thakur, E. Donini, F. Bovolo, and L. Bruzzone, "An approach to the assessment of detectability of subsurface targets in polar ice from satellite radar sounders," *IEEE Trans. Geosci. Remote Sens.*, early access, Oct. 20, 2021, doi: [10.1109/TGRS.2021.3119047](https://doi.org/10.1109/TGRS.2021.3119047).
- [8] L. Bruzzone *et al.*, "Envision mission to venus: Subsurface radar sounding," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2020, pp. 5960–5963.
- [9] L. Bruzzone *et al.*, "RIME: Radar for icy moon exploration," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jan. 2013, pp. 3907–3910.
- [10] R. E. Bell *et al.*, "Widespread persistent thickening of the east Antarctic ice sheet by freezing from the base," *Science*, vol. 331, no. 6024, pp. 1592–1595, Mar. 2011.
- [11] R. E. Bell *et al.*, "Deformation, warming and softening of Greenland's ice by refreezing meltwater," *Nature Geosci.*, vol. 7, no. 7, p. 497, 2014.
- [12] C. Panton and N. B. Karlsson, "Automated mapping of near bed radio-echo layer disruptions in the Greenland ice sheet," *Earth Planet. Sci. Lett.*, vol. 432, pp. 323–331, Dec. 2015.
- [13] C. F. Dow, N. B. Karlsson, and M. A. Werder, "Limited impact of subglacial supercooling freeze on the Greenland ice sheet stratigraphy," *Geophys. Res. Lett.*, vol. 45, no. 3, pp. 1481–1489, Feb. 2018.
- [14] E. Donini, L. Carrer, C. Gerekos, L. Bruzzone, and F. Bovolo, "An unsupervised fuzzy system for the automatic detection of candidate lava tubes in radar sounder data," *IEEE Trans. Geosci. Remote Sens.*, early access, Mar. 12, 2021, doi: [10.1109/TGRS.2021.3062753](https://doi.org/10.1109/TGRS.2021.3062753).
- [15] E. Donini, S. Thakur, F. Bovolo, and L. Bruzzone, "An automatic approach to map refreezing ice in radar sounder data," *Proc. SPIE Image Signal Process. Remote Sens.*, vol. 11155, Dec. 2019, Art. no. 111551B.
- [16] A. M. Ilisei and L. Bruzzone, "A system for the automatic classification of ice sheet subsurface targets in radar sounder data," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3260–3277, Jun. 2015.
- [17] A.-M. Ilisei, M. Khodadadzadeh, A. Ferro, and L. Bruzzone, "An automatic method for subglacial lake detection in ice sheet radar sounder data," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3252–3270, Jun. 2019.
- [18] M. Schäfer *et al.*, "Sensitivity of basal conditions in an inverse model: Vestfonna ice cap, nordaustlandet/svalbard," *Cryosphere*, vol. 6, no. 4, pp. 771–783, Jul. 2012.
- [19] K. Christianson *et al.*, "Basal conditions at the grounding zone of Whillans Ice Stream, West Antarctica, from ice-penetrating radar," *J. Geophys. Res., Earth Surf.*, vol. 121, no. 11, pp. 1954–1983, Nov. 2016.
- [20] S. P. Carter, H. A. Fricker, and M. R. Siegfried, "Antarctic subglacial lakes drain through sediment-floored canals: Theory and model testing on real and idealized domains," *Cryosphere*, vol. 11, no. 1, p. 381, 2017.
- [21] D. A. Young, D. M. Schroeder, D. D. Blankenship, S. D. Kempf, and E. Quartini, "The distribution of basal water between Antarctic subglacial lakes from radar sounding," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 374, no. 2059, Jan. 2016, Art. no. 20140297.
- [22] L. C. Sime, R. C. A. Hindmarsh, and H. Corr, "Automated processing to derive dip angles of englacial radar reflectors in ice sheets," *J. Glaciol.*, vol. 57, no. 202, pp. 260–266, 2011.
- [23] M. L. Goldberg, D. M. Schroeder, D. Castelletti, E. Mantelli, N. Ross, and J. M. Siegert, "Automated detection and characterization of Antarctic basal units using radar sounding data: Demonstration in Institute Ice Stream, West Antarctica," *Ann. Glaciol.*, vol. 61, no. 8, pp. 242–248, 2020.
- [24] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.
- [25] O. Tasar, Y. Tarabalka, and P. Alliez, "Incremental learning for semantic segmentation of large-scale remote sensing data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3524–3537, Sep. 2019.
- [26] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [27] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jan. 2014, pp. 580–587.
- [28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI (Lecture Notes in Computer Science)*, vol. 9351, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds. Cham, Switzerland: Springer, 2015, doi: [10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28).
- [29] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2017.
- [30] O. Oktay *et al.*, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.
- [31] P. Guo, X. Su, H. Zhang, M. Wang, and F. Bao, "A multi-scaled receptive field learning approach for medical image segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1414–1418.
- [32] H. Kamangir, M. Rahmehoonfar, D. Dobbs, J. Paden, and G. Fox, "Deep hybrid wavelet network for ice boundary detection in radar imagery," in *Proc. Int. Geosci. Remote Sens. Symp.*, Jan. 2018, pp. 3449–3452.
- [33] M. Yari, M. Rahmehoonfar, J. Paden, I. Oluwanisola, L. Koenig, and L. Montgomery, "Smart tracking of internal layers of ice in radar data via multi-scale learning," in *Proc. IEEE Int. Conf. Big Data*, Dec. 2019, pp. 5462–5468.
- [34] M. Rahmehoonfar, J. Johnson, and J. Paden, "AI radar sensor: Creating radar depth sounder images based on generative adversarial network," *Sensors*, vol. 19, no. 24, p. 5479, Dec. 2019.
- [35] E. Donini, F. Bovolo, and L. Bruzzone, "An unsupervised deep learning method for subsurface target detection in radar sounder data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2021, pp. 2955–2958.
- [36] M. H. García, E. Donini, and F. Bovolo, "Automatic segmentation of ice shelves with deep learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. IGARSS*, Jul. 2021, pp. 4833–4836.
- [37] M. H. García, E. Donini, and F. Bovolo, "Transfer learning for the semantic segmentation of cryosphere radargrams," *Proc. SPIE Image Signal Process. Remote Sens.*, vol. 11862, pp. 223–233, Dec. 2021.

- [38] Y. Cai, S. Hu, S. Lang, Y. Guo, and J. Liu, "End-to-end classification network for ice sheet subsurface targets in radar imagery," *Appl. Sci.*, vol. 10, no. 7, p. 2501, Apr. 2020.
- [39] G. K. A. Oswald and S. Gogineni, "Recovery of subglacial water extent from Greenland radar survey data," *J. Glaciol.*, vol. 54, no. 184, pp. 94–106, 2008.
- [40] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [41] J. Schlemper *et al.*, "Attention gated networks: Learning to leverage salient regions in medical images," *Med. Image Anal.*, vol. 53, pp. 197–207, Apr. 2019.
- [42] N. Ross, H. Corr, and M. Siegert, "Large-scale englacial folding and deep-ice stratigraphy within the west Antarctic ice sheet," *Cryosphere*, vol. 14, no. 6, pp. 2103–2114, Jun. 2020.
- [43] C. Panton, "Automated mapping of local layer slope and tracing of internal layers in radio echograms," *Ann. Glaciol.*, vol. 55, no. 67, pp. 71–77, 2014.
- [44] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*.
- [45] T. Shen, T. Zhou, G. Long, J. Jiang, S. Pan, and C. Zhang, "DiSAN: Directional self-attention network for RNN/CNN-free language understanding," 2017, *arXiv:1709.04696*.
- [46] D. Erhan, A. Courville, Y. Bengio, and P. Vincent, "Why does unsupervised pre-training help deep learning?" in *Proc. 13rd Int. Conf. Artif. Intell. Statist.*, 2010, pp. 201–208.
- [47] A. Kanazaki, "Unsupervised image segmentation by backpropagation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 1543–1547.
- [48] P. Soille, *Morphological Image Analysis: Principles and Applications*. Cham, Switzerland: Springer, 2013.
- [49] P. Y. Simard *et al.*, "Best practices for convolutional neural networks applied to visual document analysis," in *Proc. ICDAR*, vol. 3, 2003, pp. 1–4.
- [50] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.



Elena Donini (Member, IEEE) received the B.Sc. degree in electronics and telecommunication engineering, the M.Sc. degree (*summa cum laude*) in telecommunication engineering, and the Ph.D. degree (*cum laude*) in information and communication technologies from the University of Trento, Trento, Italy, in 2015, 2017, and 2021, respectively.

She is currently a Post-Doctoral Researcher with the Remote Sensing for Digital Earth Unit, Fondazione Bruno Kessler, Trento, where she is working on the design of automatic methods for the analysis of planetary and terrestrial radar sounder data. She is also a member of the Remote Sensing Laboratory, Department of Information and Communication Technologies, University of Trento. She is also a member of the Scientific Team working on the Radar for Icy Moon Exploration (RIME) Instrument that will be onboard the upcoming ESA mission JUPITER ICY moons Explorer (JUICE) to the Jupiter system and a member of the Science Study Team of the EnVision mission to Venus. Her research interests are related to the automatic analysis of terrestrial and planetary radar sounder data with machine learning and deep learning techniques.

Dr. Donini was a recipient of the Prize for the 2017 Best Italian Master Thesis in remote sensing awarded by the Italian Chapter of the IEEE Geoscience and Remote Sensing Society.



Francesca Bovolo (Senior Member, IEEE) received the Laurea (B.S.) degree, the Laurea Specialistica (M.S.) degree (*summa cum laude*) in telecommunication engineering, and the Ph.D. degree in communication and information technologies from the University of Trento, Trento, Italy, in 2001, 2003, and 2006, respectively.

She was a Research Fellow with the University of Trento until 2013. She is currently the Founder and the Head of Remote Sensing for Digital Earth Unit, Fondazione Bruno Kessler, Trento, and a member of the Remote Sensing Laboratory, Trento. She is one of the co-investigators of the Radar for Icy Moon Exploration Instrument of the European Space Agency Jupiter Icy Moons Explorer and a member of the Science Study Team of the EnVision mission to Venus. Her research interests include remote-sensing image processing, multitemporal remote sensing image analysis, change detection in multispectral, hyperspectral, and synthetic aperture radar images,

and very high-resolution images, time-series analysis, content-based time-series retrieval, domain adaptation, and light detection and ranging (LiDAR) and radar sounders. She conducts research on these topics within the context of several national and international projects.

Dr. Bovolo is also a member of the program and scientific committees of several international conferences and workshops. She was a recipient of the First Place in the Student Prize Paper Competition of the 2006 IEEE International Geoscience and Remote Sensing Symposium (Denver, 2006). She was the Technical Chair of the Sixth International Workshop on the Analysis of Multitemporal Remote-Sensing Images (MultiTemp 2011 and 2019). She has been the Co-Chair of the SPIE International Conference on Signal and Image Processing for Remote Sensing since 2014. She was the Publication Chair of the International Geoscience and Remote Sensing Symposium in 2015. She has been an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING since 2011 and a Guest Editor of the Special Issue on Analysis of Multitemporal Remote Sensing Data of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. She is also a referee for several international journals.



Lorenzo Bruzzone (Fellow, IEEE) received the Laurea (M.S.) degree (*summa cum laude*) in electronic engineering and the Ph.D. degree in telecommunications from the University of Genoa, Genoa, Italy, in 1993 and 1998, respectively.

He is currently a Full Professor of telecommunications with the University of Trento, Trento, Italy, where he teaches remote sensing, radar, and digital communications. He is also the Founder and the Director of the Remote Sensing Laboratory, Department of Information Engineering and Computer Science, University of Trento. He is also a principal investigator of many research projects. Among the others, he is currently a Principal Investigator of the Radar for icy Moon exploration (RIME) instrument in the framework of the JUPITER ICY moons Explorer (JUICE) mission of the European Space Agency (ESA) and the Science Lead for the High Resolution Land Cover project in the framework of the Climate Change Initiative of ESA. He is the author (or a coauthor) of 236 scientific publications in refereed international journals (172 in IEEE journals), more than 310 papers in conference proceedings, and 21 book chapters. He is an editor/co-editor of 18 books/conference proceedings and one scientific book. His papers are highly cited, as proven from the total number of citations (more than 27 000) and the value of the H-index (74) (source: Google Scholar). His research interests are in the areas of remote sensing, radar, and synthetic aperture radar (SAR), signal processing, machine learning, and pattern recognition. He promotes and supervises research on these topics within the frameworks of many national and international projects.

Dr. Bruzzone has been a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS) since 2009, where he has been the Vice-President for Professional Activities since 2019. He ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Seattle, July 1998. He was a recipient of many international and national honors and awards, including the recent IEEE GRSS 2015 Outstanding Service Award, the 2017 IEEE IGARSS Symposium Prize Paper Award, and the 2018 IEEE IGARSS Symposium Prize Paper Award. Since 2003, he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. He was invited as a keynote speaker in more than 32 international conferences and workshops. He was a Guest Co-Editor of many special issues of international journals. He is also the Co-Founder of the IEEE International Workshop on the Analysis of Multi-Temporal Remote-Sensing Images (MultiTemp) series and a member of the Permanent Steering Committee of this series of workshops. He has been the Founder of the *IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE* for which he was the Editor-in-Chief from 2013 to 2017. He is also an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was a Distinguished Speaker of the IEEE Geoscience and Remote Sensing Society from 2012 to 2016.