Transfer learning for the semantic segmentation of cryosphere radargrams

Miguel Hoyo García[1,2], Elena Donini[1], and Francesca Bovolo[1]

[1]Digital Society Center, Fondazione Bruno Kessler, Via Sommarive, 18 I-38123, Trento, Italy
[2]Department of Information Engineering and Computer Science, University of Trento, Via Sommarive, 5 I-38123, Trento, Italy

ABSTRACT

Radar sounders (RSs) mounted on airborne platforms are active sensors widely employed to acquire subsurface data of the cryosphere for Earth observation. RS data, also called radargrams, provide information on the buried geology by identifying dielectric discontinuities in the subsurface. Recently, a strong effort can be observed in designing automatic techniques to identify the main targets of the cryosphere. However, most of the methods are based on target-specific handcrafted features. Newly convolutional neural networks (CNNs) automatically extract meaningful features from data. However, supervised training requires numerous labeled data that are hard to retrieve in the RS domain. In this work, we adopt a CNN pre-trained in domains other than RS for automatically segmenting cryosphere radargrams. To adapt to the radargram characteristics, we introduce convolutional layers at the beginning of the pre-trained network. We modify the top layers of the network to a U-fashion autoencoder to extract relevant features for the target task. The new layers are fine-tuned with few labeled radargrams to identify and segment five targets: free space, continental ice layering, floating ice, bedrock, and EFZ and thermal noise. The pre-trained weights are not updated during fine-tuning. We applied the proposed approach to radargrams from Antarctica acquired by MCoRDS3, obtaining high overall accuracy. These results demonstrate the effectiveness of the method in segmenting radargrams and discriminating continental and coastal ice structures.

Keywords: Radar sounder, remote sensing, cryosphere, deep learning, transfer learning, domain adaptation.

1. INTRODUCTION

Understanding the dynamics of the ice sheets and ice shelves helps to predict the impact of climate change on the cryosphere[1] and the evolution of the ice at the Poles. This can be done by analyzing the inner structure of ice sheets and ice shelves to estimate climate change indicators for example related to the ice shelf melting and the ice sheet mass balance decrease. The analysis of the mass balance of ice sheets and ice shelves requires the direct measurement of the ice up to the basal interface to extract important information on the subsurface geologic structures and processes. These measurements can be provided by radar sounders (RSs), acquiring radargrams.

RSs transmit an electromagnetic wave in the range of High Frequency (HF) or Very High Frequency (VHF). The wave is transmitted in the nadir direction and propagates through the subsurface. The interaction of the EM wave with the dielectric interfaces in the subsurface results in backscattered echoes that are captured by the sensor. The radargram is generated by adding the echoes and concatenating them in the along-track direction of the radar. Hence, radargrams image the targets in the cryosphere, like continental ice layering, bedrock, floating ice and crevasses, and echo-free zone (EFZ) and noise (see Fig 1).

RSs for Earth observation are mainly mounted on an airborne platform and generally probe the subsurface of Greenland and Antarctica. Motivated by the expected increasing amount of radar sounders with planned airborne and satellite-borne missions,[2,3] there exist automatic approaches for radargram analysis based on statistical methods and machine learning techniques to identify targets in the subsurface.[4,5] Ilisei et al.[5] proposed a

Further author information: (Send correspondence to Miguel Hoyo García or Francesca Bovolo)
Miguel Hoyo García: E-mail: mhoyogarcia@fbk.eu
Francesca Bovolo: E-mail: bovolo@fbk.eu

method based on the Support Vector Machine (SVM) to automatically segment grounded ice sheet radargrams. This approach employed several hand-crafted features, e.g. the entropy, which are target-specific. SVM-based methods have been also employed to detect and segment specific targets in cryosphere radargrams, such as Donini et al.[4] that presented a method to detect refreezing ice. Similar to 5, this approach uses several class-specific handcrafted features. Due to the use of handcrafted features, the main disadvantage of these approaches is that every new target calls for a complete redesign of the features.

In recent years, advanced deep learning (DL) techniques, such as convolutional neural networks (CCNs), have become prominent to analyze and segment several data types. Deep CNNs automatically learn meaningful features from the data during training at the cost of a large amount of labeled data[6]. When a large labeled dataset is not available, as in the RS domain, there are possible mitigation techniques as randomly initializing the network parameters leads to overfitting or suboptimal results.[6] For instance, Ronneberber et al.[7] propose the U-Net that employs data augmentation techniques. However, when the training samples are not enough, data augmentation could still lead to overfitting, since data augmentation does not introduce a great variability in the training set.[8] An alternative is a lightweight CNN, such as the MobileNet v2,[9] that employs residual connections and obtains good performance in several tasks. Other techniques to cope with the lack of training data are transfer learning and domain adaptation. These techniques are based on pre-training the network on a large training set to avoid overfitting and learn features with large generalization capabilities. Subsequently, the network is adapted and fine-tuned with a small amount of target data. The literature has demonstrated the effectiveness of domain adaptation and transfer learning to analyze different domains (including remote sensing) and perform different tasks.[10,11] Transfer learning aims to reuse solidly pre-trained architectures on a source task ($\mathcal{T}_S$) to adapt a network to perform a target task ($\mathcal{T}_T$).[12] To analyze data of different domains, domain adaptation techniques are employed to minimize the distance across domains and re-use in the target domain ($\mathcal{D}_T$) features already learned in the source domain ($\mathcal{D}_S$). It is assumed that the domains $\mathcal{D}_T$ and $\mathcal{D}_S$ have different yet related distributions.[13]

DL for RSs faces several specific challenges caused by the early stage of development of DL for RS data. The main challenge is the lack of reliable labeled RS data, which makes it difficult to properly train from scratch, i.e, with random initialization of the network parameters, a DL architecture without overfitting. However, techniques like data augmentation are not suitable since they would not introduce enough variability in the RS dataset to avoid overfitting. Although domain adaptation and transfer learning are widely used for computer vision and other remote sensing domains, these techniques have never been used for radar sounder data. Using these techniques for RS requires considering the radar sounder data properties, e.g., signal and noise distribution. Moreover, DL for RSs has other challenges, such as the unbalanced prior probabilities of the subsurface targets in the radargrams.

In this work, we propose a transfer learning and domain adaptation approach that reuses a CNN pre-trained in domains other than RS data to segment radargrams (i.e., assign a label to each pixel of the radargram). The method uses domain adaptation and transfer learning to adapt a CNN for image classification (i.e., assign a label for each image) to automatically segment cryosphere radargrams. The network is pre-trained in the source domain with a multimedia reliable labeled dataset on a classification task. We apply domain adaptation by adding a convolutional layer at the beginning of the CNN architecture to make the network able to handle radargram properties, e.g., having one channel instead of three. In addition, we employ transfer learning to modify the pre-trained architecture to segment radargrams. To this end, we modify the ending layers to have a U-Net fashion autoencoder able to extract meaningful features for all the radargram pixels. Then, the new elements in the CNN are fine-tuned on the target domain, i.e., radargrams, with a segmentation task. Note that the parameters of the pre-trained CNN are not modified during fine-tuning. The method is tested on MCoRDS radargrams acquired in Antarctica imaging both ice sheets and ice shelves to segment radargrams into five classes (free space or air, floating ice, ice layering, bedrock, and thermal noise and echo-free zone). Here, as pre-trained CNN, we use the MobileNetV2 that performs well in many segmentation and classification tasks [9]. The goodness of the results demonstrates that it is possible to import CNN pre-trained weights from other domains to analyze radargrams. The rest of this paper is organized as follows: Section 2 formulates the problem and describes the radargram characteristics, and Section 3 describes the proposed methodology. The experimental results are in Section 3. Finally, Section 4 concludes the paper and presents the directions for future work.

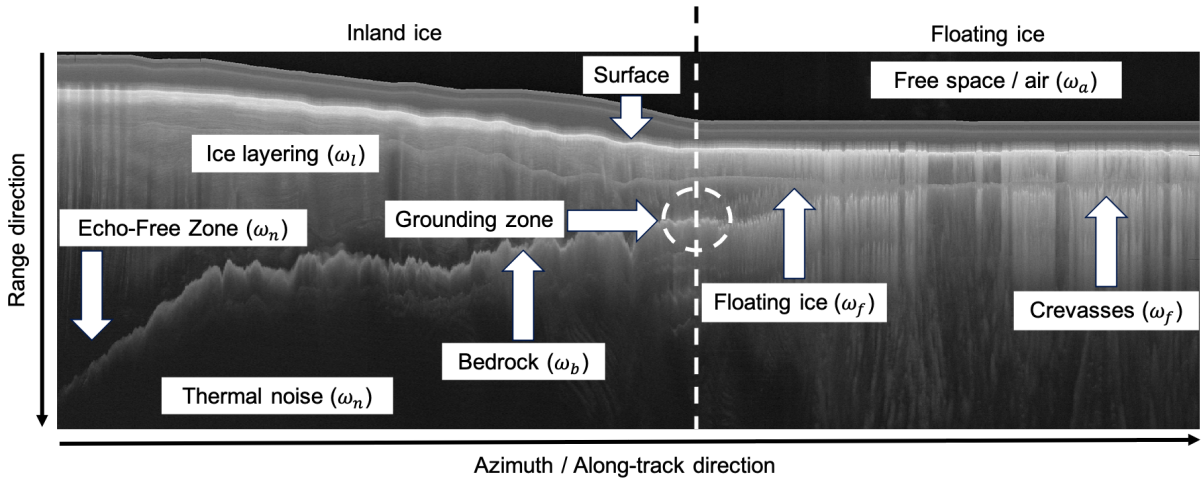


Figure 1: Detail of a radargram acquired in Antarctica by MCoRDS3 imaging the main subsurface targets.

2. PROBLEM FORMULATION

Let us consider a radargram $\mathbf{R}$ as a 2-dimensional matrix with n_T traces along track or azimuth direction (columns) and n_S samples in the depth or range direction (rows):

$$\mathbf{R} = \{R(x,y)|x \in X = [1,...,n_T], y \in Y = [1,...,n_s]\}. \quad (1)$$

Cryosphere radargrams image the ice sheets in the continental areas and the ice shelves floating on the ocean. In both, the first interface is the ice surface that generates the strongest reflections in terms of power in the radargram. This interface delimits the ice and the free space above the surface. Below the surface in radargrams of inland ice, the ice layering generates bright lines as a result of variations of the dielectric coefficient. The deepest reflection in continental radargrams is the bedrock, which is generated by the rocky interface under the ice that reflects all the remaining incident signal by generating a peak in the backscattered signal. The noisy contributions in continental ice radargrams are above the bedrock, called EFZ,[14] and below the bedrock, the thermal noise. Both noisy contributions have statistical properties similar to thermal noise added by the receiver. Coastal areas of radargrams mostly image the floating ice and crevasses. Crevasses are vertical fractures caused by the movement of the ice shelf over the water as a result of the ice melting.[15] Crevasses are identified in radargrams as high scattered vertical lines caused by the penetration of the transmitted signal through the fractures (see Fig. 1). In coastal radargrams, the target that causes the deepest reflection is the interface between the floating ice and the ocean or the crevasses in absence of floating ice. Finally, the noisy area in floating ice radargrams is placed below the floating ice and water interface. A cryosphere radargram of Antarctica is shown in Fig. 1. Note that the targets in radargrams are mainly distributed along the range direction and that the targets cover areas of different sizes in radargrams and show different prior probabilities. For example, the area covered by the layers is 20 times greater than that of the bedrock.

To summarize, the main challenges faced in the RS domain when using DL are: i) the complexity and variability of jointly processing inland and coastal icy areas, ii) the extremely unbalanced prior probabilities of the targets, iii) the limited amount of reliably labeled data available for training, and iv) the limited number of available labeled samples that leads to overfitting with the random initialization of the CNN weights when training from scratch.

3. PROPOSED METHOD

This section describes the problem and introduces the proposed method for radargram segmentation. We will explain in detail the proposed methodology for radargram segmentation based on transfer learning and domain adaptation.

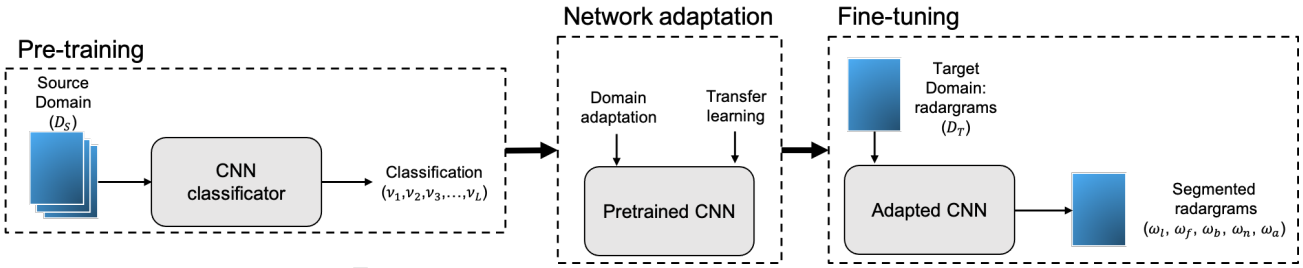


Figure 2: Block scheme of the proposed methodology.

Here, we propose a method for the automatic segmentation of radar sounder data acquired in the cryosphere. The aim is to segment a radargram $\mathbf{R}$ into N classes $\Omega = \{\omega_1, \omega_2, ..., \omega_N\}$, using a CNN able to automatically extract semantically meaningful features for each pixel.[8] To extract features in a supervised way, a CNN needs a large amount of labeled data that are not usually available in the RS data domain, the target domain $\mathcal{D}_T$. To mitigate this problem, the CNN is pre-trained in another domain, called source domain $\mathcal{D}_S$, to perform a source task $\mathcal{T}_S$, so that the network weights are not set to a random initial value decreasing the performance of the network.[6] The source domain $\mathcal{D}_S$, usually presents different characteristics, such as a higher number of channels, and the different statistical distribution P_S. So, to performs the adaptation from $\mathcal{D}_S$ to $\mathcal{D}_T$, we design domain adaptation by adding a convolutional layer at the beginning of the architecture. We introduce transfer learning techniques to reuse the pre-trained CNN weights and adapt the CNN architecture to perform the target task $\mathcal{T}_T$, i.e., radargram segmentation. The last layers of the pre-trained CNN are deleted since they identify the most specific features from $\mathcal{D}_S$,[8] resulting in a reduced pre-trained CNN. The reduced CNN is modified to a U-fashion autoencoder, where the reduced CNN is the encoder and the decoder consists of several up-convolutional layers to match the input size. The up-convolutional layers identify the target classes Ω in the target domain $\mathcal{D}_T$.

The method is divided in two steps as shown in Fig. 2:

1. The network is pre-trained in the source domain $\mathcal{D}_S$ to perform the source task $\mathcal{T}_S$, which consists of image classification into M classes $\mathcal{V} = \{\nu_1, \nu_2, ..., \nu_M\}$ (different than Ω) with a multimedia labeled dataset.

2. We adapt the network for the radargrams characteristics and to perform segmentation by employing domain adaptation and transfer learning techniques to analyze data of the target domain $\mathcal{D}_T$, i.e, radargrams, and perform the target task $\mathcal{T}_T$, i.e., radargram segmentation.

3. Finally, the network is fine-tuned in $\mathcal{D}_T$ to extract semantically meaningful features for each pixel of the radargrams.

3.1 Pre-training in $\mathcal{D}_S$

In the pre-training step, the network is trained in the source domain $\mathcal{D}_S$ to perform a source task $\mathcal{T}_S$. Choosing the source domain $\mathcal{D}_S$ and the task $\mathcal{T}_S$ depend on i) the availability of reliable labeled data in the source domain, and ii) the similarity of the source domain to the target domain in terms of data probability density distribution. The source domain $\mathcal{D}_S$ is formed by a feature space $\mathcal{X}_S$ and a related probability distribution $P(X_S)$, where X_S is the set of labeled training samples $X_S = \{x_{S_1}, x_{S_2}, ..., x_{S_r}\} \in \mathcal{X}_S$. Thus, given a source domain $\mathcal{D}_S = \{\mathcal{X}_S, P(X_S)\}$, the source task $\mathcal{T}_S$ associated to $\mathcal{D}_S$ can be defined by: the label space $\mathcal{V}$ and the predictive function $f_S()$. The label space $\mathcal{V} = \{\nu_1, \nu_2, ..., \nu_M\}$ consists of the labels of the training samples X_S. The predictive function f_S is an unobserved attribute that predicts a label $y_{S_i} \in \mathcal{V}$ corresponding to a given sample $x_{S_i} \in X_S$. Therefore, the source task can be defined as:

$$\mathcal{T}_S = \{\mathcal{V}, f_S(\cdot)\}. \tag{2}$$

From a probabilistic viewpoint, $f_S(x)$ can be expressed as $P_S(y_S|x_S)$. Hence, the source domain data D_S can be defined as:

$$D_S = \{(x_{S_1}, y_{S_1}), ..., (x_{S_r}, y_{S_r})\}. \tag{3}$$

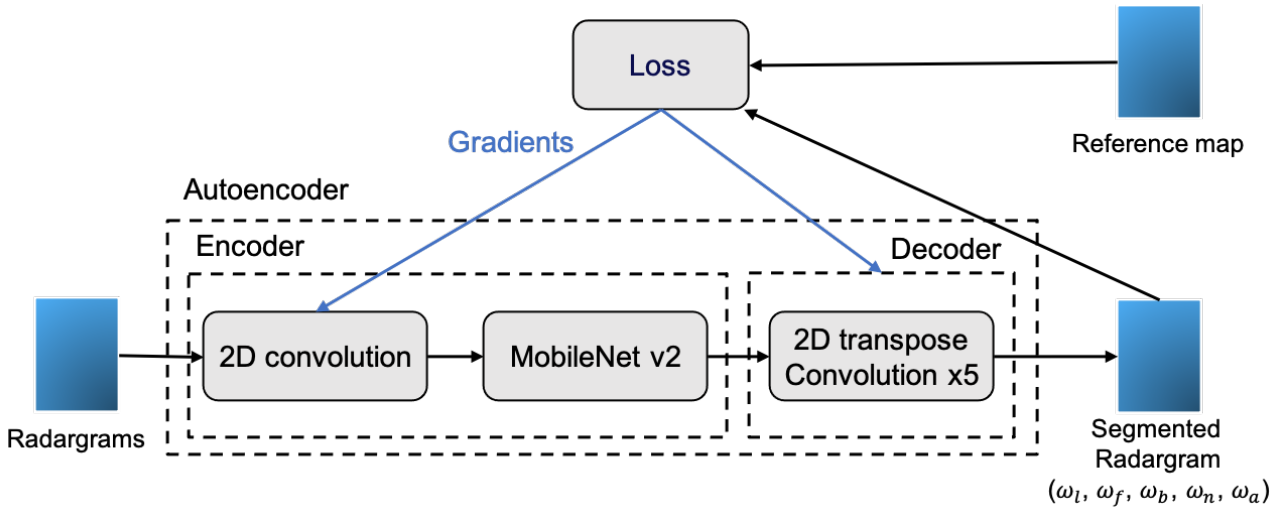


Figure 3: Block scheme of the proposed new architecture.

Deep learning for visual recognition is currently in a very advanced stage of development. So, this led us to choose an extremely optimized approach with enough training data to be robustly trained. Here, robust training is critical since the better the extracted features from data are in D_S, the better results will be reached in the target domain.[6] We choose the MobileNet V2 that performs well in many segmentation and classification tasks.[9] Moreover, the number of parameters is very low compared to other DL solutions, which is optimal for the RS domain to avoid overfitting. The CNN employs depthwise separable convolutions, a low computational cost convolution with a significant small trade-off in accuracy reduction to improve the performance.[16] Also, the network takes advantage of inverted residual blocks, a variation of the traditional residual blocks that insert shortcuts between the bottlenecks to avoid transmitting non-linear transformations. As Mobilenet is developed in the computer vision domain, available Mobilenet V2 is pre-trained with 3-channel RGB images to perform a classification task.[9]

3.2 Network adaptation and fine-tuning in $\mathcal{D}_T$

Here, we apply domain adaptation techniques to make the network analyze data in the target domain $\mathcal{D}_T$, i.e, radargrams, and transfer learning techniques to perform the target task ($\mathcal{T}_T$), i.e., radargram segmentation, reusing the CNN pre-trained weights. Let us define the target domain ($\mathcal{D}_T$) similarly to $\mathcal{D}_S$: $\mathcal{D}_T = \{\mathcal{X}_T, P(X_T)\}$. $\mathcal{D}_T$ consists of the feature space $\mathcal{X}_T$, and the related probability distribution $P(X_T)$. Analogously, the set of training samples X_T is defined as $X_T = \{x_{T_1}, x_{T_2}, ..., x_{T_r}\} \in \mathcal{X}_T$. The target task ($\mathcal{T}_S$) can be defined as:

$$\mathcal{T}_T = \{\Omega, f_T(\cdot)\}, \tag{4}$$

where Ω represents the segmentation classes and f_T is the function that segments the radargram in the classes Ω. Thus, f_T can be expressed as $P(y_T|x_T), y_T \in \Omega$ and $Y_T = \{y_{T_1}, y_{T_2}, ..., y_{T_r}\} \in \Omega$, therefore the target domain data is:

$$D_T = \{(x_{T_1}, y_{T_1}), ..., (x_{T_r}, y_{T_r})\}. \tag{5}$$

We adapt the MobileNet V2 to consider the radargram characteristics by proposing a new architecture with a U-fashion shape, as shown in Fig. 3. We divide the implementation into two steps: adaptation to the radargram characteristics, and adaptation to the radargram segmentation task.

Adaptation to the radargram characteristics. This step aims to perform the domain adaptation to handle the target domain that has i) one channel instead of three, and ii) different marginal probability of the radar data $P(X_T) \neq P(X_S)$ than the multimedia data, which means different distribution of the features between the domains. Radar data are affected by a noise that can be approximated as multiplicative, which has different properties than the additive noise of optical or multimedia data.[17] We assume that $\mathcal{X}_T \in \mathcal{X}_S$, since a radargram

is a 1 channel image representation. To learn and compensate for the difference between the domains, we include a 2D convolution layer at the top of the MobileNet V2. This layer performs the domain adaptation function f_{da} to fit the target domain in input to the system to the source domain data by doing the 1-to-3 channel conversion as:

$$D_S = \{(f_{da}(x_{T_1}), y_{S_1}), ..., (f_{da}(x_{T_q}), y_{S_q})\}. \tag{6}$$

Adaptation to the radargram segmentation. Here, we apply transfer learning to make the MobileNet V2 to segment the input radargrams into the Ω classes reusing the pre-trained CNN weights. So we adapt the network to extract relevant features for the target task $\mathcal{T}_S$. The last three layers of the pre-trained network (two 2D convolutional layers and an average pool layer) are discarded since they extract features specific to the source domain and task. We introduce a new reduced target task $\mathcal{T}_{S_{red}} = \{\mathcal{V}_{red}, f_{S_{red}}(\cdot)\}$, a new label space ($\mathcal{V}_{red}$), and a new predictive function ($f_{S_{red}}$), which is $P_{S_{red}}(y_{S_{red}}|f_{da}(x_T)), y_{S_{red}} \in \mathcal{V}_{red}$. This can be fitted in the source domain data:

$$D_S = \{(f_{da}(x_{T_1}), y_{S_{red_1}}), ..., (f_{da}(x_{T_q}), y_{S_{red_q}})\}. \tag{7}$$

Due to the MobileNet V2 architecture, the new output features are significantly smaller in size compared to the input radargram. Since we aim to segment the input radargrams, we add five up-convolutional layers acting as a decoder to increase the size of the output features. Additionally, each up-convolutional layer in the decoder is shortcutted with a matching size layer in the reduced MobileNet V2 to facilitate gradient propagation. The decoder classifies each pixel of the radargram in one of the N classes in Ω, acting as f_T. The target task is:

$$\mathcal{T}_T = \{\Omega, f_T(f_{S_{red}}(\cdot))\}. \tag{8}$$

Finally, the network layers added in this step are fine-tuned with a small amount of labeled data to set the network weights to extract semantically meaningful features from the pixels for the segmentation. Note that the pre-trained weights of the MobileNet V2 are not updated in this step. The loss optimized in fine-tuning is the sparse categorical cross-entropy loss. The error during fine-tuning is used to update the weights of the new convolutional layers. The pre-trained weights remain unchanged to avoid over-fitting due to the small amount of labeled data in the target domain.

Due to the extremely unbalanced appearance of classes in $\mathcal{D}_T$, we introduce weight maps to increase the importance of the less frequent classes otherwise under-considered in the training. For instance, the noise class appears over twenty times more frequently than the bedrock in cryosphere radargrams. Hence, the less frequent classes are not well differentiated since their contribution to the overall error value is low. The weight of each class w_{ω_i} depends on the class prior probability in X_T:

$$w_{\omega_i} = \frac{N_p}{N_{\omega_i} N} \tag{9}$$

where N_{ω_i} is the total number of pixels of ω_i in Y_T, and N_p identifies the total number of pixels in X_T. The value of w_{ω_i} weights the error calculated for each class during the fine-tuning.

4. EXPERIMENTAL RESULTS

4.1 Dataset and experimental setup

In this subsection, we describe the datasets for the pre-train and fine-tuning of the network and the experimental set-ups. The radargram segmentation aims at labeling each pixel of the input radargrams into the five considered classes: air, layers of the ice sheet, floating ice, bedrock, and EFZ and thermal noise ($\Omega = \{\omega_{free\ space}, \omega_{layers}, \omega_{bedrock}, \omega_{floating\ ice}, \omega_{noise}\}$).

As source domain dataset D_S, we choose a multimedia dataset that can be easily collected. In this case we use the ImageNet dataset that consists of 14 million RGB images of 3 channels divided into 1000 classes[18] (See Table 1). The source task $\mathcal{T}_S$ is to classify the images. Hence, the network is pre-trained with this dataset.

The dataset for fine-tuning (D_T) includes radargrams acquired by MCoRDS-3[19] in four campaigns in inland and coastal areas of Antarctica to have all the target classes ($\Omega = \{\omega_{free\ space}, \omega_{layers}, \omega_{bedrock}, \omega_{floating\ ice}, \omega_{noise}\}$)

Table 1: Pre-training dataset and experimental set-up.

Parameter	Value
Dataset	ImageNet
Image Size (height×width×channels)	224×224×3
Images	14000000
Classes	1000

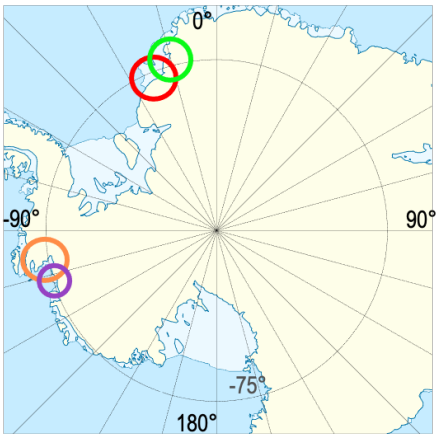


Figure 4: Location of the campaigns: 20161109_03 (orange), 20181015_01 (red), 20181115_01 (purple), 20181012_01 (green) in Antarctica.

and are provided by the Center of Remote Sensing of Ice Sheets (CReSIS): 20161109_03, 20181012_01, 20181015_01, and 20181115_01 with a total of 116288 traces. The location of the campaigns is shown in Fig. 4. The MCoRDS-3 is mounted in a Douglas DC-8 aircraft, and the radar parameters are shown in Table 2.

The data are range compressed and azimuth focused via synthetic aperture radar (SAR) processing. The fluctuations caused by the movements of the aircraft were corrected, and the power information of each radargram was log scaled and normalized to approximate the noise as additive. The radargrams were divided into 1817 patches of size 1536×64 pixels to meet the pre-trained network requirements. The vertical dimension is chosen to maximize the variability in terms of the considered classes in the range direction. Finally, D_T is divided into training for fine-tuning, validation, and test sets by making sure to use different campaigns for each set. The network is fine-tuned with a learning rate of 0.001 using the Adam optimizer, and minimizing the sparse categorical cross-entropy loss. The fine-tuning parameters are shown in Table 3.

4.2 Evaluation Metrics

To evaluate the performance of the proposed methodology, we consider i) the average producer accuracy (APA), which is computed by dividing the number of correctly classified pixels on each class C_{ω_i} by the number of reference pixels of that class N_{ω_i} (see Eq. 10), ii) the average user accuracy (AUA) that results from dividing the number of correctly classified pixels in each class C_{ω_i} by the total number of pixels that were classified in that class U_{ω_i} (see Eq. 11), and iii) the overall accuracy (OA) that is calculated by dividing the correctly classified pixels by the network C_{ω_i} by the total number of pixels N_p (see Eq. 12):

$$APA = \frac{C_{\omega_i}}{N_{\omega_i}} \cdot 100\% \tag{10}$$

$$AUA = \frac{C_{\omega_i}}{U_{\omega_i}} \cdot 100\% \tag{11}$$

$$OA = \frac{C_{\omega_i}}{N_p} \cdot 100\% \tag{12}$$

Table 2: Parameters and characteristics of MCoRDS-3 mounted on DC-8 aircraft.[19]

Parameter	Value
Central Frequency (f_c)	190 MHz
Bandwidth (BW)	50 MHz
Transmitted Power (P_{tr})	6000 W (1000W/channel)
Aircraft Altitude (h)	1500 m
Range Resolution in Ice (R_r)	4.3 m
Along-track Resolution (R_a)	27.5 m

Table 3: Fine-tuning dataset and experimental set-up parameters.

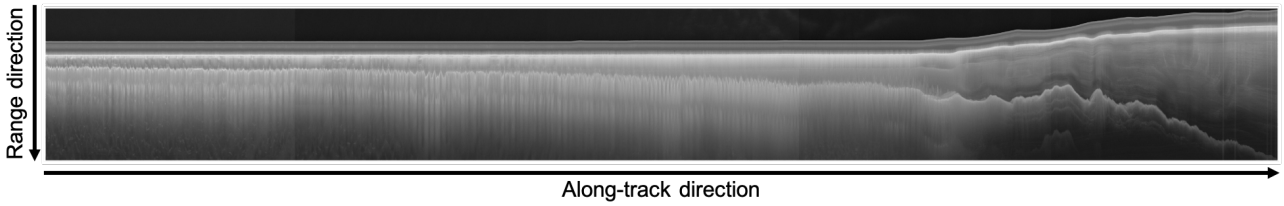
Parameter	Value
Campaign ID (Radargram ID)	20161109_03 (045-047), 20181015_01 (004-009), 20181115_01 (001-002, 012-014), 20181012_01 (009-013, 017-026, 028-030)
Patch size (height×width×channels)	1536×64×1
Number of traces	116288
Dimension of training set	939
Dimension of validation set	234
Dimension of test set	410
Number of classes	5
Learning rate	0.001
Optimizer	Adam
Loss	Sparse categorical cross entropy loss
Epochs	200

4.3 Segmentation results

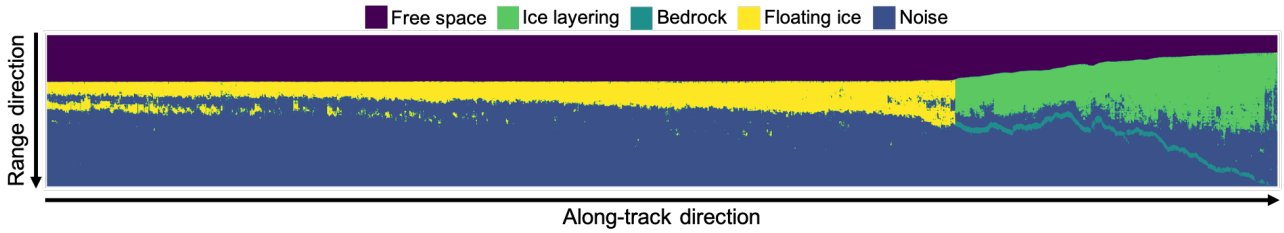
Fig. 5b shows the segmentation map of the radargram presented in Fig. 5a. The radargram was acquired in the Pine Island Glacier. The floating ice is on the left part of the radargram and the inland ice is on the right part of the radargram. The segmentation map shows that classes are well discriminated. The network manages to properly delimit the floating ice by identifying the interface between free space and ice, and ice and ocean. We can also see that the CNN efficiently differentiates the floating ice, the inland ice, and the bedrock, which could help to identify the grounding lines. The segmentation map proves the robustness of the network since no data acquired from this campaign were used for training.

Fig. 6b presents the segmentation map of the radargram in Fig. 6a. Note that the network can identify and segment the five classes in this radargram with good accuracy. Again, the network differentiates between the continental ice and the floating ice. We can also see how the network properly segments other areas as the bedrock. However, we can see that the segmentation of the floating ice is less accurate in this classification map. This is probably caused by the artifacts that can be observed in the original radargram. In any case, the network manages to delimit accurately the floating ice in most of the traces.

Table 4 shows the error matrix in pixel classification obtained by the network on the test set. The matrix shows the accuracy results obtained for each class and the overall accuracy, which is over 94%. These results prove that the learned features from the source domain D_S are extendable and effective to analyze radar sounder data and the learned features in the added layers are meaningful for segmenting radargrams. We can notice that the latest layers of the network manage to extract meaningful features from the main targets in the radargram, see Fig. 7. In addition, the overall accuracy reached by the network demonstrates that the method is robust and that the network has not overfitted during training. This could be critical due to the small number of training samples. We compared these results with the same autoencoder with random initialization of the weights of the added layers instead of fine-tuning. However, the accuracy results obtained without fine-tuning were not satisfactory, below 50% OA.

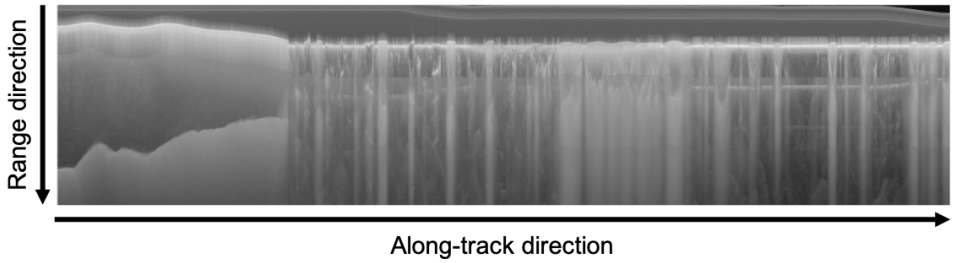


(a)

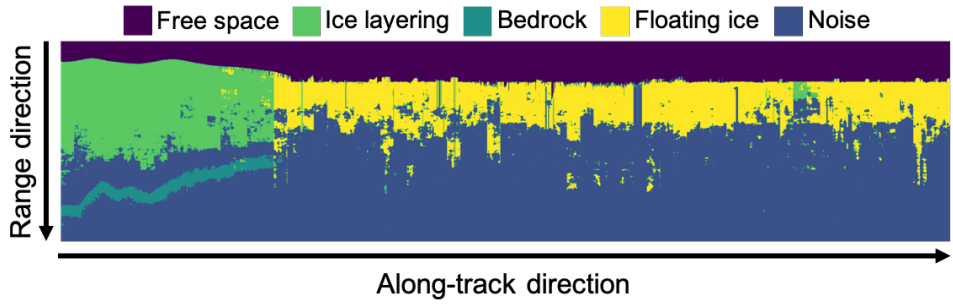


(b)

Figure 5: (a) Radargram 045-047 acquired in the campaign 20161109_03 and (b) segmentation map by the proposed network. The free space is represented in purple, the floating ice is mapped in yellow, the layers of the continental ice sheet in green, the bedrock is represented in light blue, and dark blue is the noise and the EFZ.



(a)



(b)

Figure 6: (a) Radargram 012-013 acquired in the campaign 20181115_01 and (b) segmentation map by the proposed network. The free space is represented in purple, the floating ice is mapped in yellow, the layers of the continental ice sheet in green, the bedrock is represented in light blue, and dark blue is the noise and the EFZ.

Table 4: Error matrix of the proposed methodology on the test samples.

		Reference Pixels						
	Class	**Free Space**	**Noise**	**Bedrock**	**Layers**	**Floating Ice**	**Total**	**AUA**
Pred. Value	**Free Space**	960624	8453	0	7485	9544	9631723	99.74%
	Noise	0	21459213	124559	173601	590948	22348321	96.02%
	Bedrock	0	76554	284195	334	417	361500	78.62%
	Layers	2039	784520	1235	3500774	27690	4316258	81.11%
	Floating ice	13903	228275	4952	60356	3385595	3693081	91.67%
	Total	9622183	22557015	414941	3742550	4014194		
	APA	99.83%	95.13%	68.49%	93.54%	84.34%	**OA = 94.76%**	

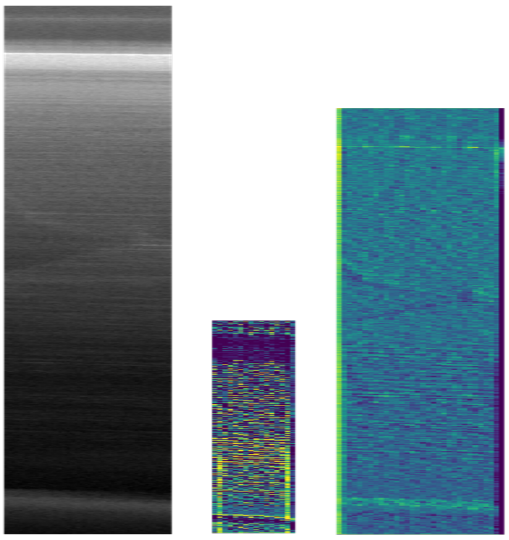


Figure 7: From left to right: original patch, features extracted from the third 2D transpose convolution in the decoder, and features extracted from the fourth 2D transpose convolution in the decoder.

5. CONCLUSION

In this work, we proposed a method based on transfer learning and domain adaptation to reuse a pre-trained MobileNet V2 to automatically segment radargrams. The radargrams are acquired in the cryosphere and radargrams are segmented with high accuracy in the five target classes.

We have employed domain adaptation techniques to be able to employ a pre-trained network with RGB images to analyze data from a very different domain by adding a convolutional layer at the beginning of the architecture. In addition, we employed transfer learning techniques to reuse the pre-trained weights and to redesign the MobileNet V2 architecture, which was originally designed to classify images, into a U-fashion autoencoder to segment radargrams in the target RS domain classes. Finally, we have fine-tuned the elements added to the redesigned architecture with a small labeled dataset in the RS domain so the new layers learn semantically meaningful features for each pixel.

We tested our approach on radargrams acquired in Antarctica by MCoRDS-3 over a complex scenario including both inland and coastal areas. The results that we obtained show that the network can identify the target classes. Also, we proved the robustness of the approach by reaching over 94% of segmentation accuracy on radargrams whose campaigns were not used for fine-tuning. With these results, we also proved that it is possible to reuse pre-trained features from other domains in the RS domain. We also demonstrated the effectiveness of the fine-tuning since the fine-tuned autoencoder outperformed the randomly initialized one.

As future work, we aim to compare this approach with other pre-trained architectures and with other approaches, such as unsupervised learning. We also aim to expand this approach to data acquired in other regions as Greenland or other planets of the Solar System.

REFERENCES

[1] Smith, B., Fricker, H. A., Gardner, A. S., Medley, B., Nilsson, J., Paolo, F. S., Holschuh, N., Adusumilli, S., Brunt, K., Csatho, B., et al., "Pervasive ice sheet mass loss reflects competing ocean and atmosphere processes," *Science* **368**(6496), 1239–1242 (2020).

[2] Donini, E., Thakur, S., Bovolo, F., and Bruzzone, L., "Assessing the detection performance on icy targets acquired by an orbiting radar sounder," in [*IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*], 997–1000, IEEE (2019).

[3] Bruzzone, L., Bovolo, F., Thakur, S., Carrer, L., Donini, E., Gerekos, C., Paterna, S., Santoni, M., and Sbalchiero, E., "Envision mission to venus: Subsurface radar sounding," in [*IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*], 5960–5963, IEEE (2020).

[4] Donini, E., Thakur, S., Bovolo, F., and Bruzzone, L., "An automatic approach to map refreezing ice in radar sounder data," in [*Image and Signal Processing for Remote Sensing XXV*], **11155**, 111551B, International Society for Optics and Photonics (2019).

[5] Ilisei, A. M. and Bruzzone, L., "A system for the automatic classification of ice sheet subsurface targets in radar sounder data," *IEEE Transactions on Geoscience and Remote Sensing* **53**(6), 3260–3277 (2015).

[6] Erhan, D., Courville, A., Bengio, Y., and Vincent, P., "Why does unsupervised pre-training help deep learning?," in [*Proceedings of the thirteenth international conference on artificial intelligence and statistics*], 201–208, JMLR Workshop and Conference Proceedings (2010).

[7] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in [*International Conference on Medical image computing and computer-assisted intervention*], 234–241, Springer (2015).

[8] Goodfellow, I., Bengio, Y., and Courville, A., [*Deep learning*], MIT press (2016).

[9] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C., "Mobilenetv2: Inverted residuals and linear bottlenecks," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 4510–4520 (2018).

[10] Yosinski, J., Clune, J., Bengio, Y., and Lipson, H., "How transferable are features in deep neural networks?," *arXiv preprint arXiv:1411.1792* (2014).

[11] Xie, M., Jean, N., Burke, M., Lobell, D., and Ermon, S., "Transfer learning from deep features for remote sensing and poverty mapping," in [*Thirtieth AAAI Conference on Artificial Intelligence*], (2016).

[12] Pan, S. J. and Yang, Q., "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering* **22**(10), 1345–1359 (2009).

[13] Bruzzone, L. and Marconcini, M., "Domain adaptation problems: A dasvm classification technique and a circular validation strategy," *IEEE transactions on pattern analysis and machine intelligence* **32**(5), 770–787 (2009).

[14] Robin, G. d. Q. and Millar, D., "Flow of ice sheets in the vicinity of subglacial peaks," *Annals of Glaciology* **3**, 290–294 (1982).

[15] Lindzey, L., "A brief introduction to ice-penetrating radar [online]," *lindzey.github.io* (2015).

[16] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861* (2017).

[17] Ferro, A. and Bruzzone, L., "Analysis of radar sounder signals for the automatic detection and characterization of subsurface features," *IEEE Transactions on Geoscience and Remote Sensing* **50**(11), 4333–4348 (2012).

[18] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al., "Imagenet large scale visual recognition challenge," *International journal of computer vision* **115**(3), 211–252 (2015).

[19] Rodriguez-Morales, F., Gogineni, S., Leuschen, C. J., Paden, J. D., Li, J., Lewis, C. C., Panzer, B., Alvestegui, D. G.-G., Patel, A., Byers, K., et al., "Advanced multifrequency radar instrumentation for polar research," *IEEE Transactions on Geoscience and Remote Sensing* **52**(5), 2824–2842 (2013).