

APPLICATION PAPER

Skip-WaveNet: a wavelet based multi-scale architecture to trace snow layers in radar echograms

Debvrat Varshney¹ , Masoud Yari², Oluwanisola Ibikunle³, Jilu Li³, John Paden³, Aryya Gangopadhyay¹ and Maryam Rahmehoonfar^{2,4}

¹Department of Information Systems, University of Maryland Baltimore County, Baltimore, MD, USA

²Department of Computer Science and Engineering, Lehigh University, Bethlehem, PA, USA

³Center for Remote Sensing and Integrated Systems, University of Kansas, Lawrence, KS, USA

⁴Department of Civil and Environmental Engineering, Lehigh University, Bethlehem, PA, USA

Corresponding author: Maryam Rahmehoonfar, Email: maryam@lehigh.edu

Received: 28 February 2023; **Revised:** 02 July 2024; **Accepted:** 09 August 2024

Keywords: multi-scale networks; radar; snow layer tracking; snow thickness; wavelet transforms

Abstract

Airborne radar sensors capture the profile of snow layers present on top of an ice sheet. Accurate tracking of these layers is essential to calculate their thicknesses, which are required to investigate the contribution of polar ice cap melt to sea-level rise. However, automatically processing the radar echograms to detect the underlying snow layers is a challenging problem. In our work, we develop wavelet-based multi-scale deep learning architectures for these radar echograms to improve snow layer detection. These architectures estimate the layer depths with a mean absolute error of 3.31 pixels and 94.3% average precision, achieving higher generalizability as compared to state-of-the-art snow layer detection networks. These depth estimates also agree well with physically drilled stake measurements. Such robust architectures can be used on echograms from future missions to efficiently trace snow layers, estimate their individual thicknesses, and thus support sea-level rise projection models.

Impact Statement

Global warming is a reality that is harshly affecting the polar ice caps. The increase in their annual melt rates and their contribution to global sea-level rise can potentially cause drastic socio-economic damage. Radar devices are the most popular sensors to monitor ice caps, but processing and analyzing their data is not straightforward. In this work, we propose a deep-learning algorithm that can effectively process polar radar echograms and calculate the thickness of snow accumulated on top of an ice sheet. The estimated thicknesses can be used by glaciological models to project global sea-level rise, eventually helping us prepare for any future calamities.

1. Introduction

The increase of global mean annual temperature every year is having a drastic effect on the polar ice-caps. Intergovernmental Panel on Climate Change (IPCC) 2021 report (Masson-Delmotte et al., 2021) states that the Arctic has warmed at more than twice the global rate over the past 50 years, and the melting of the Greenland Ice Sheet (GrIS) will cause sea levels to rise by 2 m by 2,100, potentially flooding regional coastal areas (Kirezci et al., 2020). Modern climate models that simulate and project sea-level rise rely on the thickness of the snow accumulated on top of the ice caps (Koenig et al., 2016). Continuous monitoring

and evaluation of this thickness is hence imperative to support climate models make accurate projections and prepare us for future natural disasters.

Polar ice sheets are thick masses of ice, on top of which snow accumulates every year to form snow accumulation layers. The change in volume of these snow layers changes the surface mass balance (SMB) of the ice sheet. If the SMB of an ice sheet turns negative, it results in a sea-level increase. Airborne radar systems such as the Snow Radar (Gogineni et al., 2013) monitor and capture the internal state of these snow layers annually. Significant variations in snow permittivity reflect the electromagnetic signals transmitted by Snow Radar. This reflection, which can be seen as the layers in radar echograms, often occurs at the boundaries of snow stratigraphy, where changes in the physical properties of the snow, such as density, hardness, grain size, and shape, result in significant discontinuities in the snow's dielectric properties. However, the visibility and clarity of the snow layers in Snow Radar echograms degrade when the magnitude of the discontinuity in the snow's dielectric properties decreases. Routine data processing has applied deconvolution, filtering, and coherent and incoherent integration techniques to remove non-ideal system characteristics and improve the signal-to-noise ratio; however, off-nadir surface backscattering, multipath scattering, and signal loss in the medium may still exacerbate the degradation of Snow Radar echogram's quality (see the examples in the leftmost columns of Figures 1-3).

Due to the degraded quality of the echograms, tracking snow accumulation layers is challenging for experienced glaciologists as well as conventional vision algorithms (Koenig et al., 2016; Rahnemoonfar

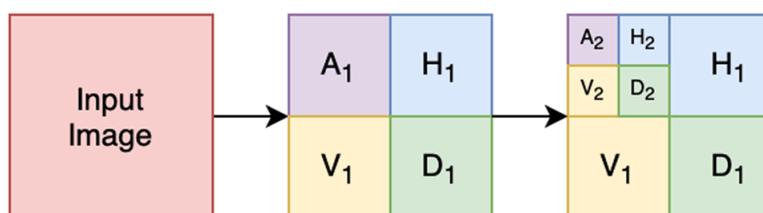


Figure 1. Qualitative comparison of model outputs. From left to right, are the Snow Radar echogram, the ground truth labels, output by MS-CNN (Rahnemoonfar et al., 2021), and our proposed Skip-WaveNet-dmey model. The red boxes highlight some of the regions where Skip-WaveNet gives a cleaner prediction compared to MS-CNN. The two networks achieved an MAE of 3.74 and 4.22 pixels, respectively, across all the layers in this echogram.

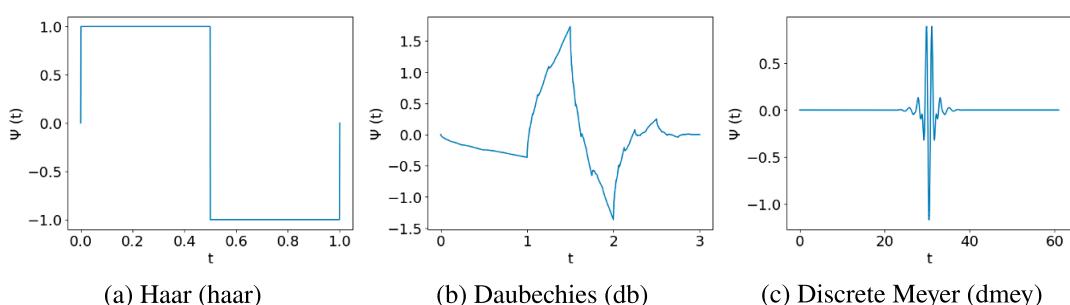


Figure 2. Qualitative comparison of model outputs. From left to right, are the Snow Radar echogram, the ground truth labels, output by MS-CNN (Rahnemoonfar et al., 2021), and our proposed Skip-WaveNet-dmey model. The red boxes highlight some of the regions where Skip-WaveNet gives a cleaner prediction compared to MS-CNN. The two networks achieved an MAE of 3.17 and 4.22 pixels, respectively, across all the layers in this echogram.

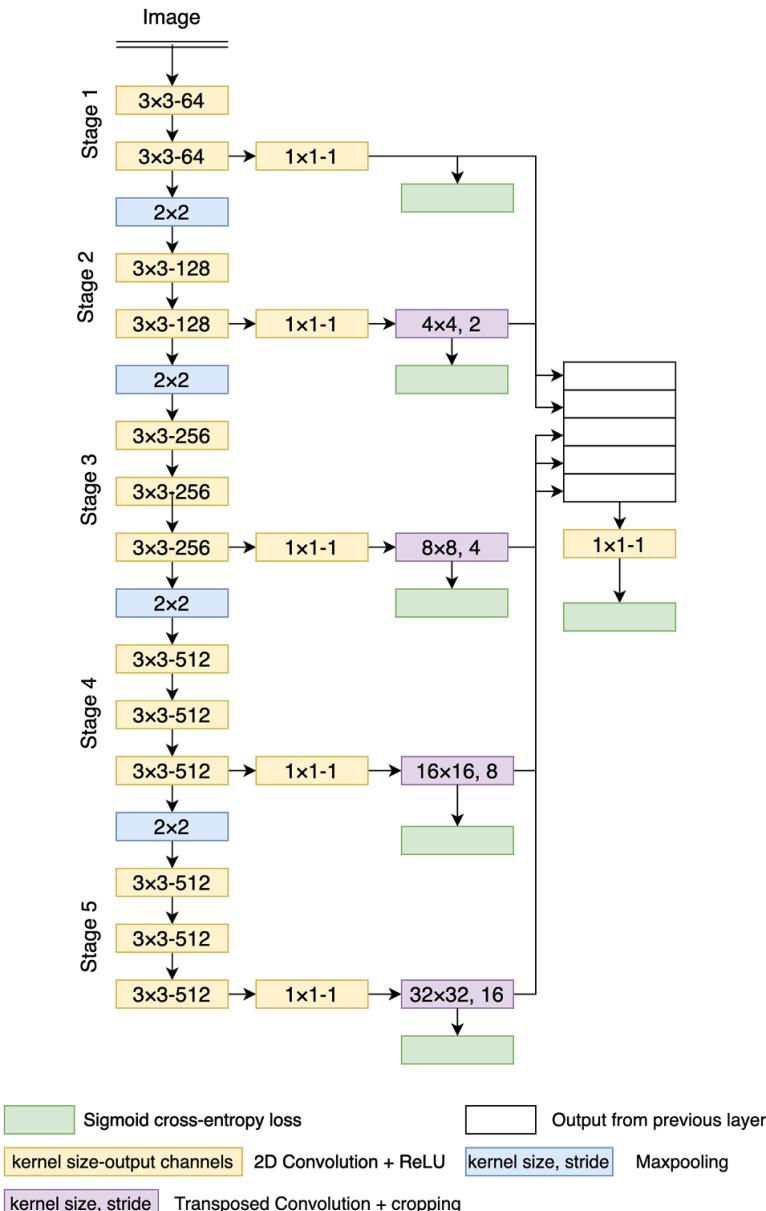


Figure 3. Qualitative comparison of model outputs. From left to right, are the Snow Radar echogram, the ground truth labels, output by MS-CNN (Rahnemoonfar et al., 2021), and our proposed Skip-WaveNet-dmey model. The red boxes highlight some of the regions where Skip-WaveNet gives a cleaner prediction compared to MS-CNN. The two networks achieved an MAE of 5.93 and 8.09 pixels, respectively, across all the layers in this echogram.

et al., 2021). Even convolutional neural networks (CNNs; LeCun et al., 2015), which have become the standard machine learning algorithms for image processing, have issues in making accurate predictions from these images (Rahnemoonfar et al., 2021; Varshney, Rahnemoonfar, Yari, Paden, et al., 2021; Varshney et al., 2020) since any interruptions to their input can drastically affect their prediction capability (Q. Li et al., 2021). To improve the effectiveness of CNNs in processing these radar images, we use

wavelet transforms (Mallat, 1989), which are signal processing techniques that can represent an image in a multi-resolution format. They have localization properties, in both the spatial and frequency domains, and can depict the contextual as well as textural information in the image at different scales through their detail coefficients (Huang et al., 2017; Mallat, 1989). We use these detail coefficients to guide the CNN toward an enhanced snow layer representation.

In this work, we exploit the multi-scale nature of wavelet transforms and fuse them in a multi-scale convolutional neural network (Rahnemoonfar et al., 2021) to improve the accuracy of snow layer detection. We also show that taking wavelet transforms of intermediate scales of the neural network helps in learning and extracting features, as compared to taking multi-level wavelet transforms of the input image. This work is an extension of Varshney, Yari, et al. (2021) with a stronger backbone network, more types of wavelet transform, analysis of “static” wavelets vs “dynamic” wavelets, comparison with state-of-the-art networks, calculation of individual layer depths (in terms of number of pixels), comparison with in-situ stake measurements, along with the use of a more practical Snow Radar dataset.

The rest of the paper is organized as follows—we highlight the past work on snow layer tracking and wavelet combined neural networks in Section 2; give a brief background of wavelet transforms, explain the proposed wavelet-based multi-scale architecture(s), and showcase the evaluation techniques in Section 3; describe the Snow Radar data and the in-situ observations that we use for our experiments in Section 4; discuss the quantitative and qualitative results in Section 5, and finally conclude the paper in Section 6 highlighting some future work.

2. Related works

Recently, there have been extensive studies on using deep learning for tracking snow layers through radar echograms. At the same time, wavelet transformations are being exhaustively used to improve the feature extraction capabilities of deep learning networks. Here, we provide a quick overview of the research being done in these two areas.

2.1. Snow layer tracking through deep learning

Feature extraction capabilities and good generalizability of deep learning networks have been used on Snow Radar images to track snow layers in Yari et al. (2019), Yari et al. (2020), Rahnemoonfar et al. (2021), Varshney, Rahnemoonfar, Yari, Paden, et al. (2021), Varshney, Yari, et al. (2021), and Wang et al. (2021). Deep learning has also been used to segment the radar depth sounder (RDS) images in Donini et al. (2022), Ghosh and Bovolo (2022a), and Ghosh and Bovolo (2022b) into ice and bedrock through various modifications of UNet (Ronneberger et al., 2015), with or without transformers (Vaswani et al., 2017). In this paper, we focus on the shallow sensor, the Snow Radar (Gogineni et al., 2013), which tracks the snow layers.

Wang et al. (2021) used a CNN-RNN (recurrent neural network) architecture for tiered segmentation of radar images to identify the top 4–6 snow layers. Similarly, Ibikunle et al. (2020) built an iterative neural network architecture to track snow layers, but on simulated radar images, as real radar images were too complex for feature extraction. Varshney, Rahnemoonfar, Yari, Paden, et al. (2021) used pyramid pooling modules to learn local-to-global spatio-contextual information of snow layer pixels and perform semantic segmentation. In the line of multi-scale architectures, Yari et al. (2019) first explored multi-scale contour detection CNNs for the purpose of snow layer extraction. The authors noted that using popular pre-training strategies would not work on Snow Radar images due to the inherent noise present in these images. Further extensions of this work by Rahnemoonfar et al. (2021) and Yari et al. (2020) showed the usefulness of multi-scale architectures on synthetic radar images and temporal transfer learning, respectively. Subsequently, wavelet-based multi-scale architectures for snow layer tracking were first developed by Varshney, Yari, et al. (2021) which showed preliminary results in this domain. This prior work used a VGG-13 architecture (Simonyan & Zisserman, 2015) on a small Snow Radar dataset, which achieved an F-score of a little over 0.7. We expand upon this work by

training a VGG-16 architecture, which is known to give robust representations (Rahnemoonfar et al., 2021), on a new dataset specifically catered towards dry zone areas, where the snow layers can be tracked. Furthermore, and most importantly, we also calculate the individual layer depths in terms of the number of pixels. We also show that using wavelet transforms of each scale of a multi-scale network helps in learning inherent image features, as compared to using multi-level wavelet transforms of the input image.

2.2. CNNs with wavelet transformations

Wavelet transforms have had immense applications in image denoising (Bnou et al., 2020; P. Liu et al., 2019; Mallat, 1989), image compression (Naveen Kumar et al., 2019; Xiong & Ramchandran, 2009), and image restoration (Bae et al., 2017; Huang et al., 2017; P. Liu et al., 2019). Many recent works such as P. Liu et al. (2019), Williams and Li (2018), Huang et al. (2017), Han and Ye (2018), and Bae et al. (2017) have used the downsampling properties of wavelets and replaced the convolutional or pooling layers in a CNN, since incorporating the wavelet coefficients typically helps reduce information loss (P. Liu et al., 2019). Williams and Li (2018) showed that using wavelet transform for downsampling would create much cleaner and sharper images, as compared to using pooling layers, and improve generalizability. Similarly, P. Liu et al. (2019) found that wavelet transforms not only enlarge a kernel's receptive field but also prevent information leakage which generally takes place during a pooling operation. By using multi-level wavelet transforms, Han and Ye (2018) developed deep convolutional framelets to reconstruct sparse-view computed tomography images and Huang et al. (2017) developed a wavelet-based loss function to super-resolve facial images. Furthermore, Bae et al. (2017) showed that residual learning improves by training on wavelet subbands. The variety of these works shows that by incorporating wavelet transforms in a CNN, the feature extraction capability of the latter improves due to sharper feature maps, increased receptive field, and enhanced residual learning.

3. Methodology

In this section, we briefly explain the background of wavelet transforms, describe our base architecture without wavelets called Multi-Scale CNN, and then explain the wavelet combined neural networks for snow layer tracking.

3.1. Wavelet transform

We use the discrete wavelet transform (DWT) for our computations. For an image $f(x,y)$ having dimensions $X \times Y$, DWT is given as Equations 3.1 and 3.2 for $j \geq j_0$:

$$W_\phi(j_0, m, n) = \frac{1}{\sqrt{X \cdot Y}} \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} f(x, y) \phi_{j_0, m, n}(x, y) \quad (3.1)$$

$$W_\psi^i(j, m, n) = \frac{1}{\sqrt{X \cdot Y}} \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} f(x, y) \psi_{j, m, n}^i(x, y) \quad (3.2)$$

In these equations, W_ϕ is the approximation coefficient (A), and W_ψ^i are the detail coefficients for each level of wavelet transform j , where $i \in \{H, V, D\}$. H, V, and D are the horizontal, vertical, and diagonal details, respectively, m, n are the subband dimensions (Williams, Li, et al., 2018), and j_0 is an arbitrary starting level. Each subsequent level of wavelet transform is computed on the previous level's approximation coefficient. In the beginning, the input raw image is treated as an approximation coefficient. Furthermore, ϕ is a scaling function, and ψ is the wavelet function, both of which downsample an input image by a factor of 2 in both X and Y dimensions. Readers are encouraged to go through Daubechies

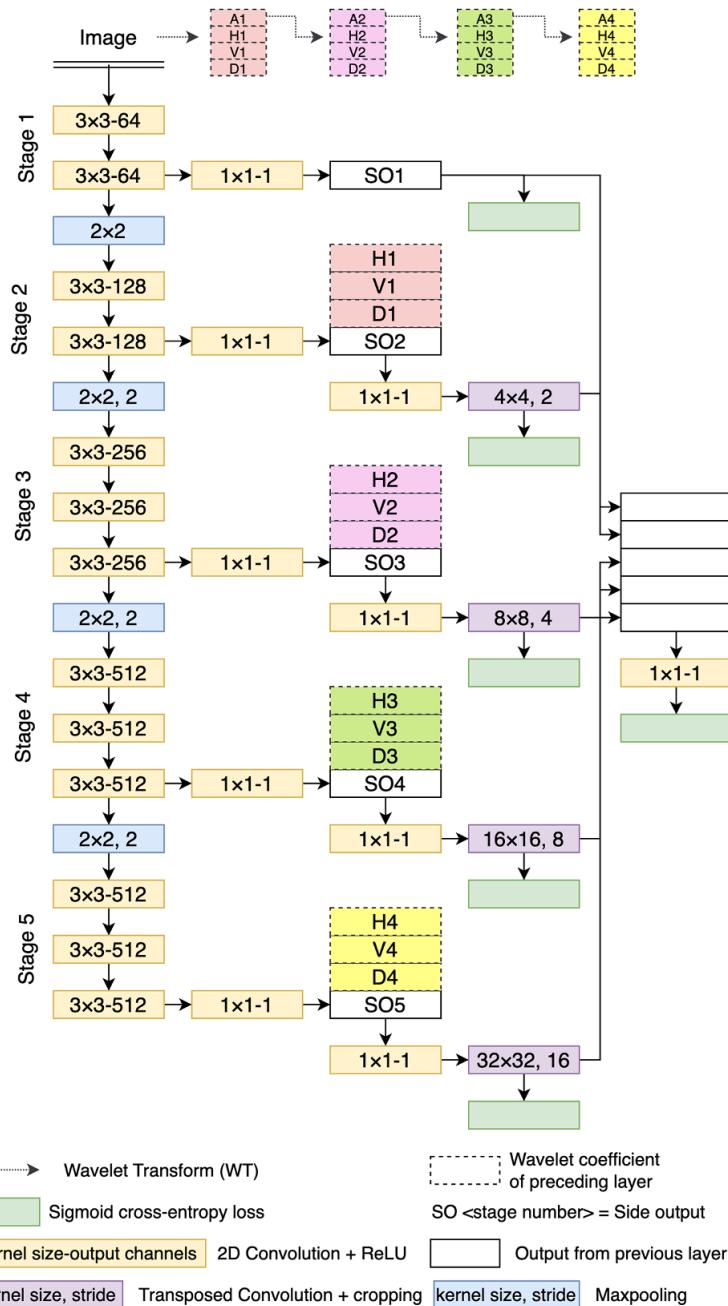


Figure 4. A level 2 wavelet transform of a given input image. The subscript denotes the level number.

(1992) to get a detailed understanding of ϕ , ψ , and the wavelet basis functions. We use the downsampling property of wavelets and combine their transforms with the side outputs of multi-scale networks. Figure 4 shows a level 2 (i.e. $j = 2$) transform of an input image. In our work, we experiment with three popular discrete wavelets, i.e. Haar, Daubechies, and “Discrete” Meyer. We will be abbreviating them as “haar,” “db,” and “dmey” in subsequent sections for ease of reading. The wavelet functions (ψ) for the three wavelet types have been plotted in Figure 5.

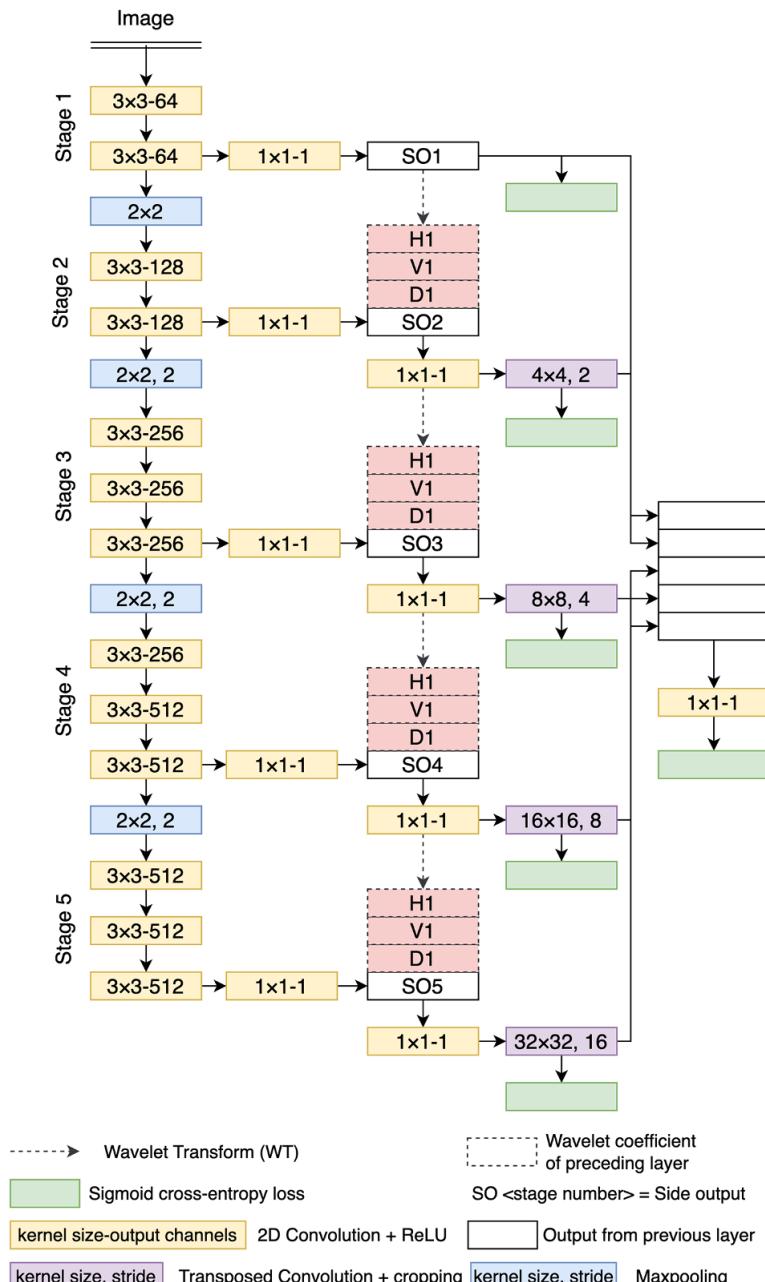


Figure 5. The three types of wavelets that we use for our experiments, and their shorthand representation in parenthesis.

3.2. Multi-scale CNN (MS-CNN)

In this work, we use a multi-scale network built for contour detection (Xie & Tu, 2015), which was proven to be useful for snow layer tracking (Rahnemoonfar et al., 2021; Yari et al., 2019). This is a VGG-16 (Simonyan & Zisserman, 2015) network with the terminal fully connected layers and the last pooling layer removed. The final convolutional layers of every stage, right before the max pooling layer of every stage, are convolved with a 1×1 filter to generate what we call “side output(s)”. Each side

output is at a downsampled resolution of the previous stage's side output. All side outputs are finally upsampled through transposed convolutions, followed by cropping, to match the resolution of the input image. The five side outputs, from the five stages, are concatenated together to form a “fuse” layer. All five side outputs along with the final “fuse” layer are trained together in a deeply supervised manner (Lee et al., 2015) with a cumulative loss function, Equation 3.5, as explained in the following paragraph.

To train the network, we calculate the binary cross entropy loss (l) for every pixel i as follows:

$$l(x_i; W) = \begin{cases} \alpha \cdot \log(1 - x_i) & \text{if } y_i = 0 \\ \beta \cdot \log(x_i) & \text{if } y_i = 1 \end{cases} \quad (3.3)$$

In Equation 3.3, x_i is the sigmoid activation map obtained from a network with weights W , y_i is the ground truth label of the corresponding pixel in the input image I having a total of $|I|$ pixels, and α, β are defined as Equation 3.4.

$$\begin{aligned} \alpha &= \lambda \cdot \frac{|Y^+|}{|Y^+| + |Y^-|} \\ \beta &= \frac{|Y^-|}{|Y^+| + |Y^-|} \end{aligned} \quad (3.4)$$

In Equation 3.4, $|Y^+|$ denotes the count of all positive labels, i.e. those pixels representing the top of a layer ($y_i = 1$) and $|Y^-|$ denotes the count of all negative labels i.e. all other pixels which are background ($y_i = 0$). λ is a hyperparameter used to balance these positive and negative labels.

The total loss is computed as Equation 5 where k depicts each of the side outputs or stages, i.e. $K = 5$. This network, called Multi-Scale CNN, is the current state-of-the-art model for snow layer tracking (Rahnemoonfar et al., 2021) and forms our baseline architecture to compare with our wavelet-based architectures. Multi-Scale CNN is shown in Figure 6 and most of the time abbreviated as MS-CNN for the rest of the paper.

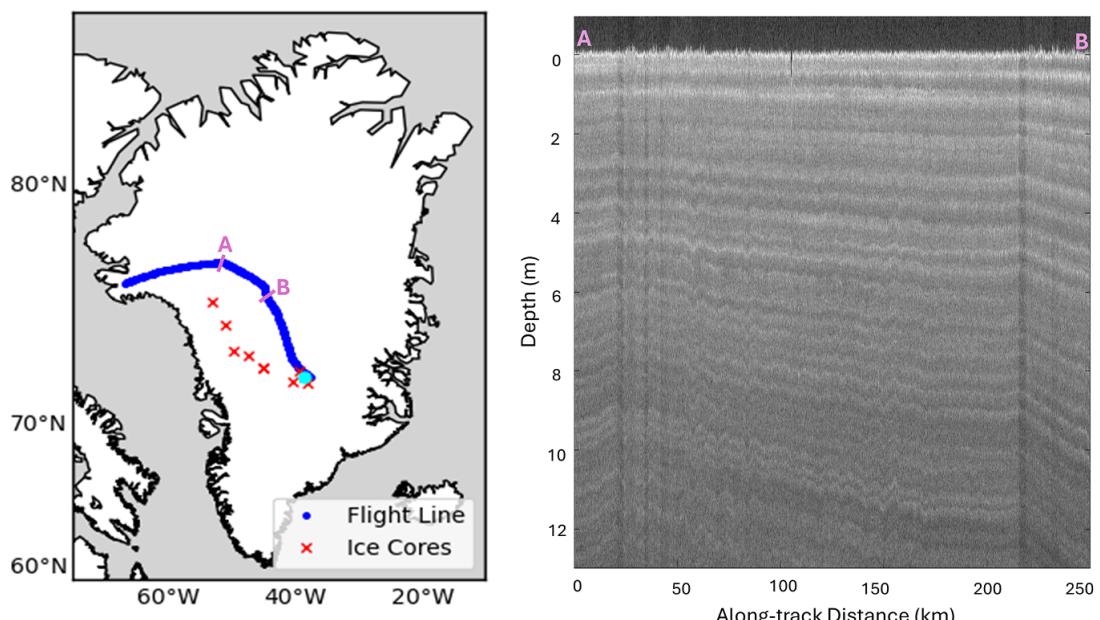


Figure 6. MS-CNN—The multi-scale architecture of Rahnemoonfar et al. (2021) which forms our base model.

$$L(W) = \sum_{i=1}^{|I|} \left(\sum_{k=1}^K l(x_i^k; W) + l(x_i^{fuse}; W) \right) \quad (3.5)$$

3.3. Wavelet combined neural networks

We set up two wavelet-based architectures, and exploit the multi-level nature of wavelet transforms to embed them into the base multi-scale architecture, MS-CNN. The two wavelet-based network modifications are termed as “WaveNet” and “Skip-WaveNet,” and are described below:

3.3.1. WaveNet

In Figure 3, from stages 2 to 5 of the base architecture, the feature maps get downsampled to a scale of $2 \times$ to $16 \times$ (with respect to the input image) to form the side outputs. Multi-level wavelet transforms of the input image also give us downsampled features at similar scales, which are sharp, denoised, and contain the horizontal, vertical, and diagonal detail coefficients (see subsection 3.1). By fusing wavelet information of the image to every scale, local as well as global, of the multi-scale network, the side outputs can be enriched for feature extraction. Hence, in this architecture, we take a level 4 wavelet transform (since there are four stages which have downsampled resolution as compared to the input image—stages 2 to 5) of the input image and fuse the detail coefficients in the following manner: we concatenate all three detail coefficients i.e. H, V, D of a wavelet transform of level l of the input radar echogram to the side output $l+1$ of the base architecture. This means that the detail coefficients from level 1 of the wavelet transform fuse with side output 2 of the architecture, detail coefficients from level 2 of the wavelet transform fuse with side output 3 of the architecture, and so on. The WaveNet architecture is shown in Figure 7.

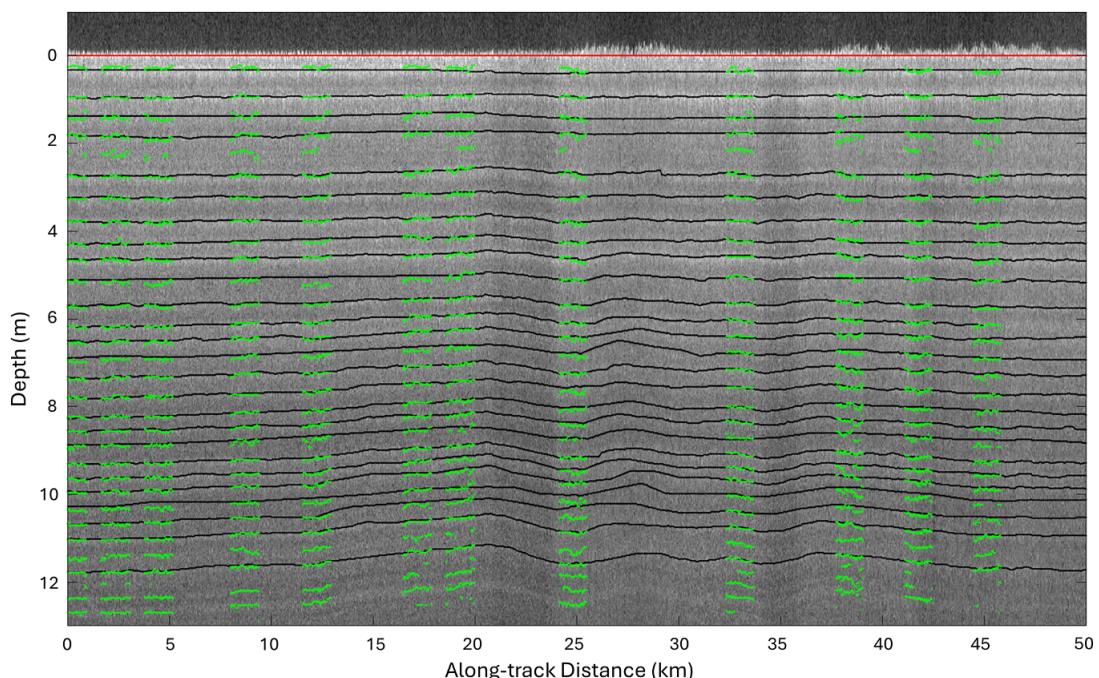


Figure 7. WaveNet—A wavelet-based architecture. Here, the input image goes through a multi-level wavelet transform, where each level is shown with a different color, ranging from light pink to yellow.

3.3.2. Skip-WaveNet

The side outputs of MS-CNN are feature maps containing image patterns extracted at different scales, from local to global. These feature maps contain an improved representation of the features, i.e. are “more learned,” after every iteration of training and also contain noise in the form of unuseful network weights. Hence, they have the potential to be denoised further which can add value to overall network training. So for the Skip-WaveNet architecture, we take a wavelet transform of each side output and add it as an extra layer to the successive side output to supplement network learning. This will help form skip connections, support residual learning (Bae et al., 2017; He et al., 2016) and also propagate information between scales. Since the wavelet transform is at a downsampled resolution to its input signal, the dimensions of a transform of side output s will be the same as the dimensions of side output $s+1$. Furthermore, by fusing the wavelet transform of a scale to its successive scale, we are also propagating information between scales which can support network learning. With this intuition, we form the Skip-WaveNet architecture as follows: we take a level 1 wavelet transform of a side output s ($s \in [1, 4]$) and fuse its detail coefficients to side output $s+1$. In this case, for a given input image, the wavelet transforms will be renewed after each epoch of training. This can be compared against the WaveNet architecture where the transforms generated for an input image will be the same during all epochs. Hence, such multi-scale “dynamic” wavelets should improve the denoising and feature extraction capabilities of the network. The architecture of Skip-WaveNet is shown in Figure 8.

Apart from the above-defined modifications, WaveNet and Skip-WaveNet are trained in the same way as the base architecture MS-CNN, i.e. they are trained with deep supervision through binary cross-entropy loss, and have the final fusion of five side outputs, etc.

3.4. Experimental setup

All the architectures, including the base architecture, and wavelet combined neural networks share the same hyperparameters as those used in Xie and Tu (2015) and Rahnemoonfar et al. (2021). For our experiments, we augment the training dataset with scaling factors $\in [0.25, 0.5, 0.75]$ and left-to-right flipping. The augmentation helps in expanding the training dataset to 6,430 images, which we train for 15 epochs. All networks took approximately 10 hours, on average, when trained on NVIDIA GeForce RTX 2080 Ti GPU with an Intel Core i9 processor.

3.5. Post-processing and evaluation

We perform non-maximum suppression (NMS) on the network outputs to make finer predictions of the layers. NMS gives grayscale predictions, which we compare against our ground truth by evaluating the ODS (optimal dataset scale) and OIS (optimal image scale) F-scores (Y. Liu et al., 2019). The ODS F-score looks for a single threshold across the entire dataset which can binarize the predicted image, and give the most optimum F-score with respect to the ground truth. On the other hand, the OIS F-score searches for a similar optimum threshold for each image. The optimum F-scores thus found from each image are then averaged over the entire dataset to obtain the OIS F-score. Here, the F-score refers to the standard F1-score used in computer vision, defined as Equation 3.6. In this equation, TP, FP, and FN are true positives, false positives, and false negatives, respectively. We also use average precision (AP), which is the area under a precision-recall curve (Su et al., 2015), as an evaluation metric of network performance.

$$F = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (3.6)$$

Based on the optimum threshold we get, we binarize the network outputs to get predictions similar to the ground truth. We then calculate the row indices (or the “range bin” indices of an echogram) of each layer in the predicted image matrix and compare them with corresponding row indices of the manually labeled ground truth to evaluate the mean absolute error (MAE) of layer depth using Equation 3.7. The

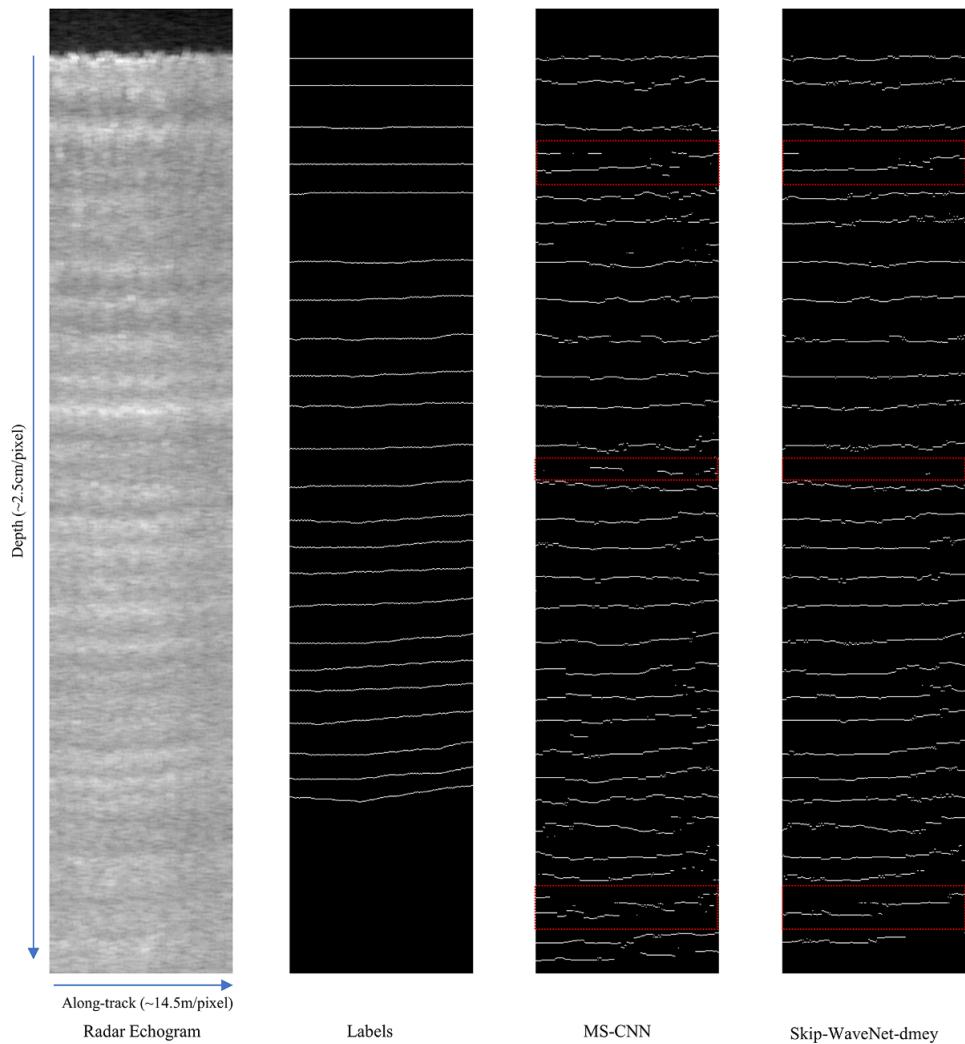


Figure 8. Skip-WaveNet—A wavelet-based architecture with skip connections. Here, we take a level 1 wavelet transform of each side output. This is in contrast to WaveNet where the wavelet transform was of the input image.

overall MAE was then obtained by estimating the MAE for each layer in every echogram for the entire test set using Equation 3.8.

$$MAE_j^k = \frac{\sum_{i=1}^{i=W_k} |GTD_i^{kj} - PD_i^{kj}|}{W_k} \quad (3.7)$$

$$MAE = \frac{1}{T} \sum_{k=1}^{k=T} \frac{\sum_{j=1}^{j=N_k} MAE_j^k}{N_k} \quad (3.8)$$

In Equations 3.7 and 3.8, GTD , and PD are the ground truth and predicted layer depths, respectively, in terms of the number of pixels. For example, PD_i^{kj} represents the the row index of the j^{th} predicted layer in the i^{th} column of the echogram k . W_k and N_k , respectively, represent the total number of columns (width),

and the total number of traced layers, in an echogram k . T represents the total number of echograms in the test set.

3.6. Comparison with *in-situ* stake measurements

The snow thickness and annual accumulation rates can be computed from traced annual layers. To assess the effects of tracing imperfections on the annual accumulation rate estimates, we compare the radar-derived accumulation rates with the ground measurements. Of the available *in-situ* stake measurements compiled by the SUMup group (see Section 4), the closest measurement to the flight line of the radar data set used in this paper is chosen as the reference for comparison. The stake location is only ~ 16 km away and is thus suitable for reliable comparison. Given the scale of the ice-sheet, we do not expect the accumulation to vary at such a short distance. We estimate the radar-derived snow accumulation rates similar to J. Li et al. (2023) which uses Benson's density-depth model (Benson, 1960) and an interpretation model from Clarke et al. (1989). The *in-situ* stake measurements provide the required initial values of snow density and accumulation at the surface for the interpretation model (Clarke et al., 1989). The monthly stake-measured snow accumulations between September and August in two consecutive years are added up as the annual accumulation rate. The MAE effects are assessed by comparing the radar-derived accumulation rates with and without the MAE of each layer, which is calculated by Equation 3.8.

4. Dataset

We use Snow Radar images publicly hosted by the Center for Remote Sensing and Integrated Systems (CReSIS; CReSIS, 2012). This dataset was captured over different regions of the GrIS and used to analyze the ice sheet's annual accumulation rates (Koenig et al., 2016). The blue line on the map in the left panel of Figure 9 shows one of the transects. The right panel of the figure presents the radar echogram of a 250 km section along the transect showing the snow layers to a depth of ~ 12 m. This transect contains 1,286 training images and 321 test images where each image has a vertical resolution of approximately 2.5 cm and an along-track resolution of 14.5 m (Table 1). We use binary labels for training and evaluating our model, where the top of each snow layer is labeled as "1," and all other pixels are labeled as "0." These labels are prepared manually by scientific experts tracing out the echograms through visual inspection.

To validate our model predictions with *in-situ* observations, we use the SUMup dataset (Dibb & Fahnestock, 2004; Thompson-Munson et al., 2022). This dataset contains surface mass-balance field observations of snow density, accumulation on land ice, and its errors over ~ 60 -year time period across the Arctic and Antarctic. This data set contains ice cores/snow pits, radar isochorones, and stake measurements. In the stake measurements, midpoint depth, snow layer densities, monthly snow accumulation, and error in accumulation are available. Figure 9 highlights the ice cores from the

Table 1. Key parameters of the Snow Radar sensor used for data collection

Parameter	Value
Bandwidth	2–8 GHz
Pulse duration	$250 \mu\text{s}$
PRF	2 kHz
Transmit power	100 mW
Intermediate frequency range	62.5–125 MHz
Sampling frequency	125 MHz
Range resolution	~ 4 cm
Along-track resolution	14.5 m

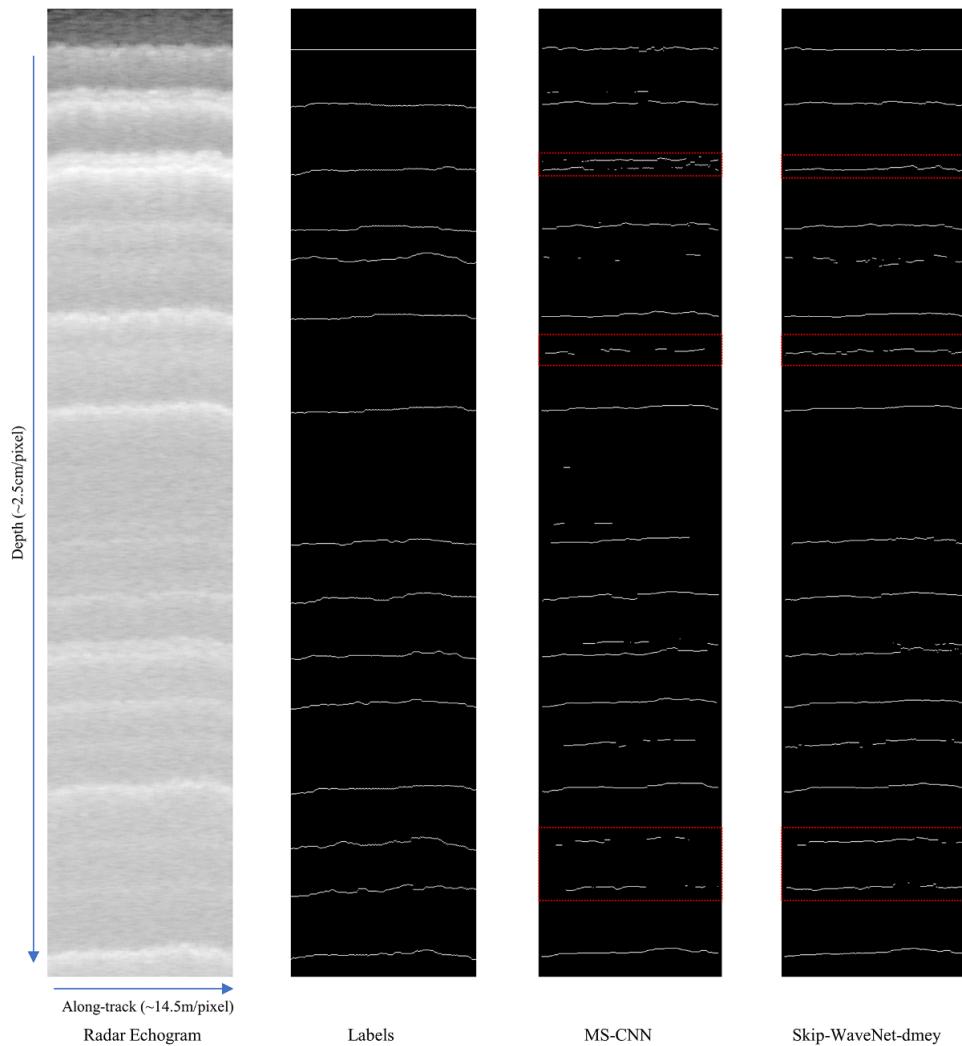


Figure 9. Left: Flightline of NASA Operation IceBridge 2012 in blue, nearest ice cores to the flight line in red, and the site which has a temporal overlap with IceBridge data in cyan. Right: A radar echogram spans 250 km along this flight line, marked by points “A” and “B”.

SUMup dataset which are closest to our flight line in 2012. For a fair comparison, we chose a stake (Dibb & Fahnestock, 2004) whose measurements have a temporal overlap with the radar data collection time span.

5. Results and discussion

We note that training with all five side outputs of our base MS-CNN model on the new dry-zone dataset by CReSIS is detrimental to performance since there are hardly any global features or contours that this network (or humans in general) can detect (see for example, the radar echograms in Figures 1 and 2). Hence, we use only the first four side outputs and the fuse layer to train the baseline MS-CNN model. This is not the case for wavelet combined neural networks which are able to detect layer contours across all scales and can generate the fifth side output as well, capturing global context. In this section, we showcase our results on wavelet combined neural networks vs the baseline model and also discuss the wavelet

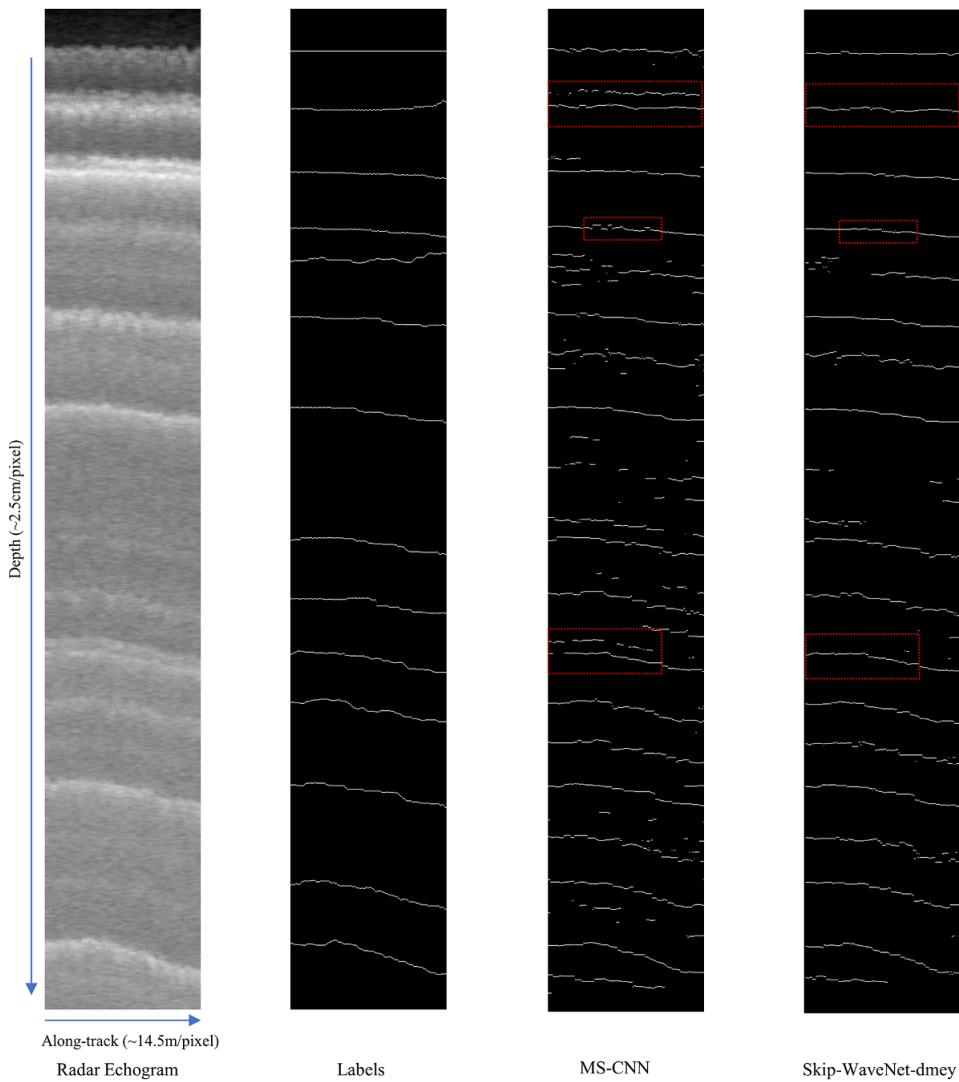


Figure 10. Radar echogram of a 50-km transect with the ground truth marked in black. Layers predicted by Skip-WaveNet on the test regions are shown in green. The network is able to trace faint layers deeper than 12 m, whose ground truth is difficult to prepare.

architectures WaveNet vs Skip-WaveNet. We also compare our performance with the only other work on Snow Radar layer tracing (Wang et al., 2021). Furthermore, in Figure 10, we showcase Skip-WaveNet’s ability to trace snow layers along a 50-km transect.

5.1. Wavelet combined neural networks vs MS-CNN

We tabulate the ODS and OIS F-scores, the average precision (AP), as well as the mean absolute error (MAE) of depth estimates in Table 2. These estimates are calculated on the final fuse layer of our experiments. From Table 2, we see that all wavelet combined neural networks, except for WaveNet with a dmey wavelet, perform better than the baseline MS-CNN model. The wavelet-based networks also give higher AP and MAE scores (3.31 pixels) as compared to CNN3B+RNN (Wang et al., 2021). Wang et al. (2021) were able to predict the top four to six layers from an echogram, while the multi-scale architecture,

Table 2. Evaluation metrics were obtained across different model architectures on the test set. Highest scores are highlighted in bold

Network	Wavelet	ODS	OIS	AP	MAE
CNN3B+RNN	None	NA	NA	0.853	8.730
MS-CNN	None	0.852	0.866	0.918	9.492
WaveNet	haar	0.876	0.888	0.936	3.517
	db	0.870	0.883	0.931	3.541
	dmey	0.835	0.851	0.905	3.967
Skip-WaveNet	haar	0.880	0.892	0.938	3.451
	db	0.879	0.890	0.937	3.438
	dmey	0.886	0.898	0.943	3.309

whether MSCNN or wavelet-based, can trace out the deeper, twenty to thirty, layers from the echogram (Figures 1-3).

In the wavelet-based networks, Skip-WaveNet with a dmey wavelet achieves the highest F-scores, with all Skip-WaveNet models performing better than WaveNet. In the rest of the manuscript, we use the words Skip-WaveNet and Skip-WaveNet-dmey interchangeably. The wavelet transforms help in improving the ODS F-score by 2.11%–3.99% and the OIS F-score by 1.96%–3.7%, over the baseline architecture, for different experiments. Thus, the wavelet transforms not only help in denoising feature maps but also extract features at the global scale (fifth stage), which the comparative MS-CNN does not. Since images are extremely noisy with different images having varying levels of intensity and contrast (see the different sample radar echograms in Figures 1-3), the layers can be almost impossible to see by human eyes. Hence, extracting features at a global fifth stage is a useful property that the wavelet architectures have achieved. In Figures 1-3, it can be seen that the layers detected by the wavelet network give a consistent and useful tracking result than the baseline MS-CNN model. The baseline model is misguided by a lot of degraded features in the input radar echograms, and falsely classifies redundant backscatterings as a layer, as can be seen by its intermediary predictions between two layers, highlighted in red in the figures. The layers predicted by MS-CNN mostly have missing values and are discontinuous in nature (i.e. have gaps in between) giving a reduced tracking accuracy (Table 2). MSCNN's MAE is worse by ~6 pixels which will give inaccurate thickness estimates eventually adding to uncertainties in accumulation rates (Kahle et al., 2021). Further, it should be noted that all multiscale architectures (MS-CNN, WaveNet, and Skip-WaveNet) have the same baseline CNN architecture, and hence the same number of trainable parameters. In WaveNet and Skip-WaveNet, the wavelet coefficients are additional filters computed from either the convolutional layers (side outputs) or the input images, and hence neither act as parameters which require training nor are updated by backpropagation. In this regard, both Skip-WaveNet and WaveNet are able to achieve a higher tracking accuracy even though their complexity, in terms of number of trainable parameters, is the same as that of MS-CNN.

5.2. Skip-WaveNet vs WaveNet

Our experiments confirm that wavelet-based architectures, by supplying additional information from the wavelet coefficients, enhance feature extraction. From Table 2 we see that Skip-WaveNet performs better than others, where the wavelet transform obtained in each epoch is from a learned and improved side output (so the wavelets are “dynamic” and change after each epoch), whereas in WaveNet, the multi-level wavelet transforms are always fixed for a given input image for every epoch (the wavelets are in a way “static” across the entire training period). We further see that the performance of the WaveNet experiment deteriorates with the wavelet complexity, whereas the performance of Skip-WaveNet improves with wavelet complexity (ODS F-score of dmey > db > haar). What would be interesting for future studies is to investigate how the different types of wavelets are affecting network learning in different ways.

5.3. Validation with snow stake measurements

In addition to the geographical proximity, the measurements at the selected snow stake site (Dibb & Fahnestock, 2004; the cyan dot in the left panel of Figure 9) span the temporal range of 2003 to 2016, thereby exhibiting a temporal overlap with the radar data collection period. This temporal alignment renders it particularly well-suited for conducting comparative analyses with our model predictions applied to the 2012 IceBridge dataset. We use this snow stake site and find the test echogram closest to it, to compare Skip-WaveNet's predictions over this echogram and the measurements at the selected stake site. Figure 11 shows the annual accumulation rate for different years (2004–2012), as well as the associated error in these measurements at the stake measurement site. Similarly, the plot also shows the radar-derived accumulation rate and its error due to the mean absolute error achieved by SkipWaveNet-dmey for the first eight annual snow layers between 2004 and 2011. The depth of the 2004 layer is ~4.5 m below the surface. As can be seen from the plot, the estimates from SkipWaveNet are within the error bounds of the stake accumulation measurements. Moreover, the radar-derived accumulation rates from training labels (i.e. without MAE, and marked by *) agree well with the accumulation rate measured at

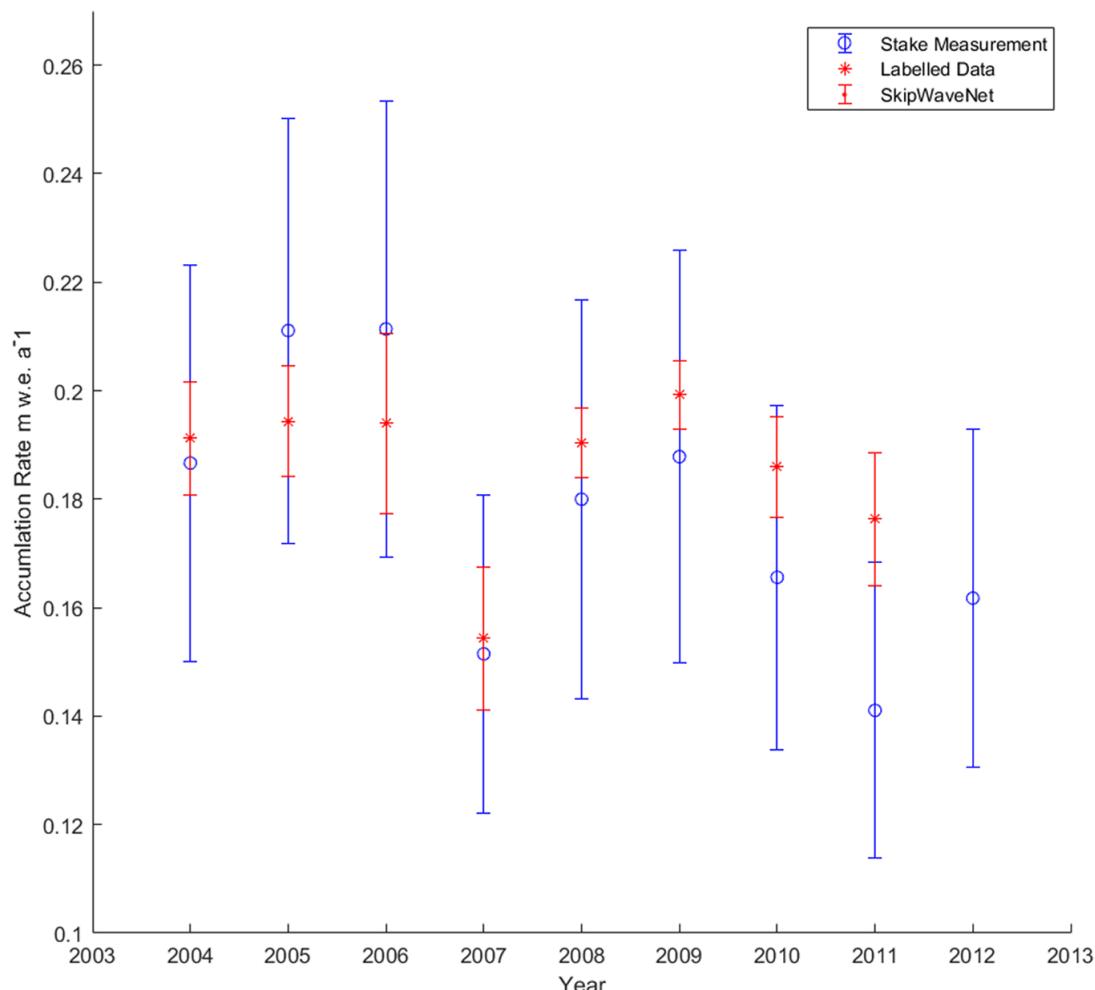


Figure 11. Comparison of the accumulation rate and its error at the site of stake measurements (blue) with the radar-derived accumulation rate from the traced annual layers marked by the training labels (*), and with the mean absolute error achieved by SkipWaveNet (red).

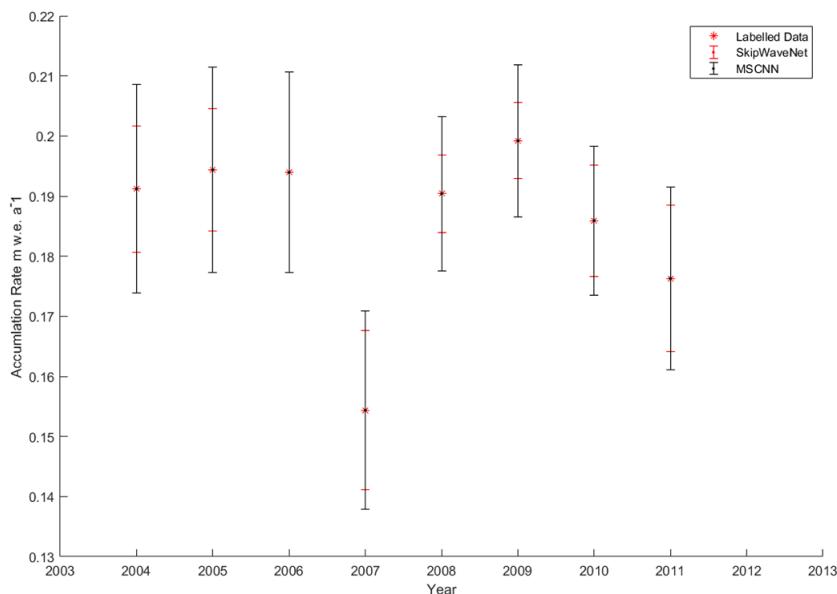


Figure 12. Comparison of the radar-derived accumulation rate from the mean absolute error of SkipWaveNet (red) and MSCNN (black), both centered at the radar-derived accumulation rate from the traced annual layers (*). Clearly, SkipWaveNet gives a lower error.

the stake site. The MAE of the first eight layers in SkipWaveNet's predictions is 2.2 pixels corresponding to an error in accumulation rate estimation of ~ 0.011 m w.e. a-1. Therefore the tracing improvement of 6.2 pixels by Skip-WaveNet dmey over MS-CNN (see Table 2) results in reducing the uncertainty by 0.031 m w.e. a-1 in accumulation rate estimation. We also compare the radar-derived accumulation rates from the mean absolute error of SkipWaveNet and MSCNN for the same echogram in Figure 12 which shows that Skip-WaveNet was able to give a smaller error in accumulation rates estimates as compared to MS-CNN.

At this stake site, however, the available density measurements were made between 1997 and 2002 in different months. In Figure 13 we plot these density measurements, made from the surface to a depth of 3 m, in the form of blue circles. The red line shows the density-depth profile obtained from linear fitting with the least mean square error. The orange line shows the density-depth profile used by the interpretation model (Clarke et al., 1989) for the accumulation rate estimation from tracked snow layers in a radar echogram. Figure 14 presents the corresponding profiles of snow deposition age and dielectric constant versus depth, respectively.

6. Conclusion

Radar echograms are used for capturing the profile of multiple snow layers accumulated on top of an ice sheet. These echograms can have varying characteristics that can make snow layers difficult to detect visually or through automated algorithms. In this work, we use wavelet transforms to extract multi-scale textural information from radar echograms to supplement neural network learning. Our novel architecture helps in improving layer detection from noisy echograms. We also find that using wavelet transforms of intermediate side outputs and forming skip connections with them help in feature extraction and network learning, as compared to using 'static' wavelets of the input radar echogram. We show that our algorithm can be used for layer tracking and calculating snow thickness accurately, thus reducing accumulation rate error as compared to in-situ stake measurements. Since our models are data-driven in nature they are

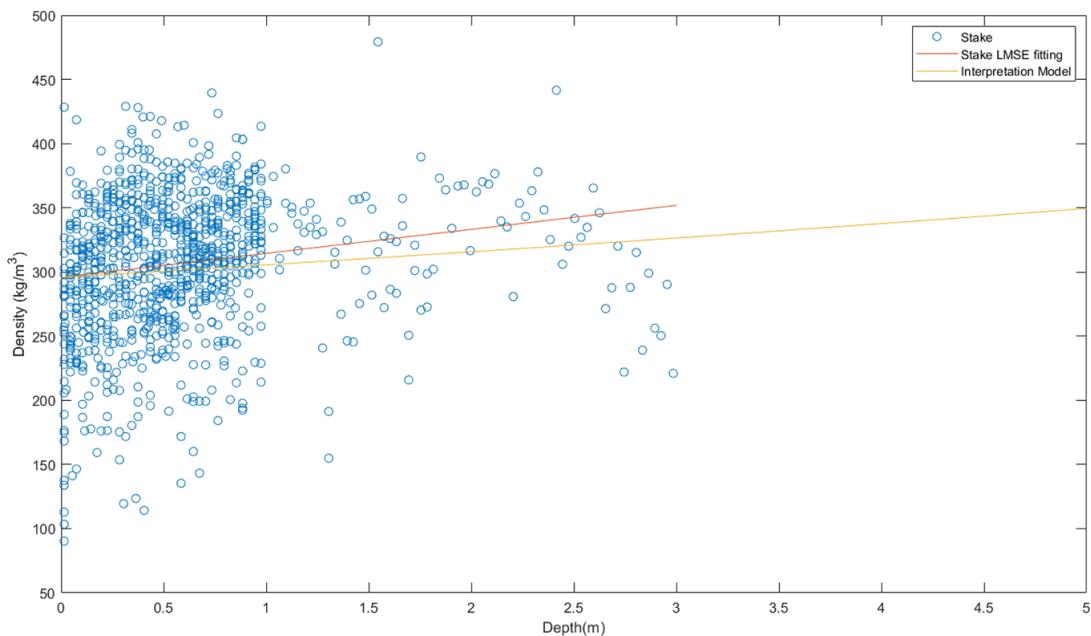


Figure 13. Snow density measurements at the stake site (Dibb & Fahnestock, 2004) in blue, their linear fit in red, and the density-depth profile (through the interpretation model of Clarke et al. (1989)) of snow layers traced from the closest radar echogram in orange.

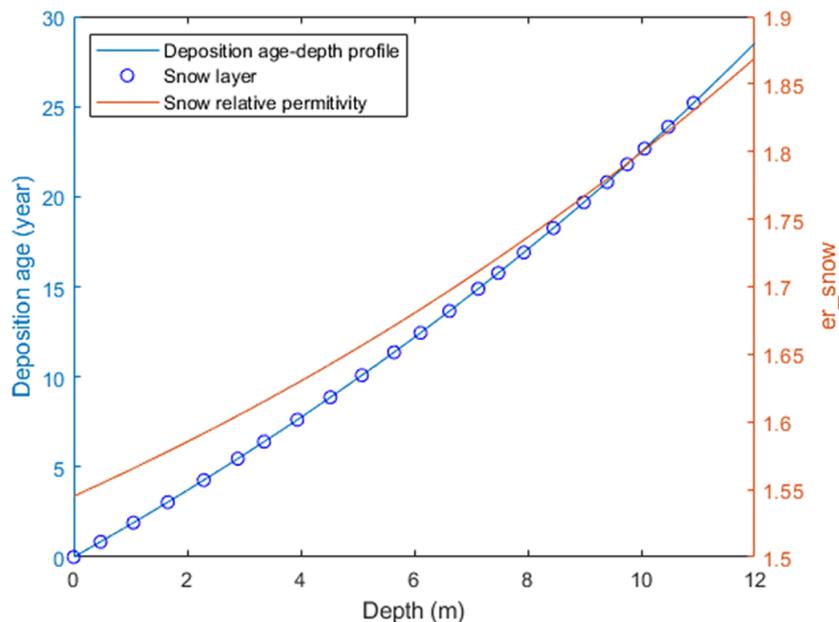


Figure 14. Profiles of snow layer deposition age (in blue) and dielectric constant (in red) versus depth estimated through the interpretation model of Clark et al. (1989) at the location of the radar echogram closest to the stake site (Dibb & Fahnestock, 2004). Layers traced by Skip-WaveNet on this radar echogram are marked by circles.

mainly effective for echograms captured in similar dry snow zone regions and can provide thickness estimates for glaciological models to use. Our future works are looking at tracing the layers for ablation zones (where the layers are much fainter, as seen through radar), leveraging larger datasets to incorporate regional variation and improve generalizability, as well as using multiple sensors to introduce complementary information.

Author contribution. Conceptualization: M.R. Resources: M.R; J.P; A.G. Software & Methodology: D.V; M.Y.; M. R.; O.I.; J. L. Data curation: O.I. Data visualization: D.V. Formal Analysis: D.V; M.R. Snow accumulation rate comparison and validation: J. L; Supervision: M.R. Funding acquisition: M.R; Investigation: D.V; M.R. Writing original draft: D.V. All authors reviewed and approved the final submitted draft.

Data availability statement. The dataset is publicly hosted by the Center for Remote Sensing and Integrated Systems (CReSIS) at CReSIS (2012).

Funding statement. This study was funded by NSF BIGDATA awards (IIS-1838230, IIS-1838024), the U.S. Army Grant No. W911NF21-20076, IBM, and Amazon.

Competing interest. None

References

Bae, W, Yoo, J, and Ye, JC (2017). Beyond Deep Residual Learning for Image Restoration: Persistent Homology-Guided Manifold Simplification. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1141–1149. <https://doi.org/10.1109/CVPRW.2017.152>

Benson, CS (1960). *Stratigraphic studies in the snow and firn of the greenland ice sheet* [Doctoral dissertation, California Institute of Technology].

Bnou, K, Raghay, S, and Hakim, A (2020). A wavelet denoising approach based on unsupervised learning model. *EURASIP Journal on Advances in Signal Processing*, 2020(1), 1–26.

Clarke, GKC., Cross, GM, and Benson, CS (1989). Radar imaging of glaciovolcanic stratigraphy, mount wrangell caldera, alaska: Interpretation model and results. *Journal of Geophysical Research: Solid Earth*, 94(B6), 7237–7249. <https://doi.org/10.1029/JB094iB06p07237>

CReSIS. (2012). NASA OIB Greenland. https://data.cresis.ku.edu/data/temp/internal_layers/NASA_OIB_test_files/image_files/Dataset/snow/temp_data_share/OLD_2012_Greenland_P3/2012_Greenland_P3/frames_001_243_20120330_04/ Accessed 08/10/2024.

Daubechies, I (1992). *Ten Lectures on Wavelets*. Society for Industrial; Applied Mathematics. <https://doi.org/10.1137/1.9781611970104>

Dibb, JE, and Fahnestock, M (2004). Snow accumulation, surface height change, and firn densification at Summit, Greenland: Insights from 2 years of in situ observation. *Journal of Geophysical Research: Atmospheres*, 109(D24), D24113. <https://doi.org/10.1029/2003JD004300>

Donini, E, Bovolo, F, and Bruzzone, L (2022). A Deep Learning Architecture for Semantic Segmentation of Radar Sounder Data. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–14. <https://doi.org/10.1109/TGRS.2021.3125773>

Ghosh, R, and Bovolo, F (2022a). A Hybrid CNN-Transformer Architecture for Semantic Segmentation of Radar Sounder data. *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 1320–1323. <https://doi.org/10.1109/IGARSS46834.2022.9883124>

Ghosh, R, and Bovolo, F (2022b). TransSounder: A Hybrid TransUNet-TransFuse Architectural Framework for Semantic Segmentation of Radar Sounder Data. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–13. <https://doi.org/10.1109/TGRS.2022.3180761>

Gogineni, S, Yan, JB, Gomez, D, Rodriguez-Morales, F, Paden, J, and Leuschen, C (2013). Ultra-wideband radars for remote sensing of snow and ice. *IEEE MTT-S International Microwave and RF Conference*, 1–4. <https://doi.org/10.1109/IMaRC.2013.6777743>

Han, Y, and Ye, JC (2018). Framing U-Net via Deep Convolutional Framelets: Application to Sparse-ViewCT. *IEEE Transactions on Medical Imaging*, 37(6), 1418–1429.

He, K, Zhang, X, Ren, S, and Sun, J (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Huang, H, He, R, Sun, Z, and Tan, T (2017). Wavelet-SRNet: A Wavelet-based CNN for Multi-scale Face Super Resolution. *Proceedings of the IEEE International Conference on Computer Vision*, 1689–1697.

Ibikunle, O, Paden, J, Rahnemoonfar, M, Crandall, D, and Yari, M (2020). Snow Radar Layer Tracking Using Iterative Neural Network Approach. *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 2960–2963. <https://doi.org/10.1109/IGARSS39084.2020.9323957>

Kahle, EC, Steig, EJ, Jones, TR, Fudge, T, Koutnik, MR, Morris, VA, Vaughn, BH, Schauer, AJ, Stevens, CM, Conway, H, et al. (2021). Reconstruction of temperature, accumulation rate, and layer thinning from an ice core at South Pole, using a statistical inverse method. *Journal of Geophysical Research: Atmospheres*, 126(13), e2020JD033300.

Kirezci, E, Young, IR, Ranasinghe, R, Muis, S, Nicholls, RJ, Lincke, D, and Hinkel, J (2020). Projections of global-scale extreme sea levels and resulting episodic coastal flooding over the 21st Century. *Scientific reports*, 10(1), 1–12.

Koenig, LS, Ivanoff, A, Alexander, PM, MacGregor, JA, Fettweis, X, Panzer, B, Paden, JD, Forster, RR, Das, I, McConnell, JR, et al. (2016). Annual Greenland accumulation rates (2009–2012) from airborne snow radar. *The Cryosphere*, 10(4), 1739–1752.

LeCun, Y, Bengio, Y, and Hinton, G (2015). Deep learning. *nature*, 521(7553), 436–444.

Lee C-Y, Xie S, Gallagher P, Zhang Z and Tu Z (2015) Deeply-supervised nets. In Lebanon G and Vishwanathan SVN (eds.), *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, Vol. 38. San Diego, CA: Proceedings of Machine Learning Research, pp. 562–570. Available at <https://proceedings.mlr.press/v38/lee15a.html>.

Li, J, Rodriguez-Morales, F, Fettweis, X, Ibikunle, O, Leuschen, C, Paden, J, Gomez-Garcia, D, and Arnold, E. (2023). Snow stratigraphy observations from operation icebridge surveys in alaska using s and c band airborne ultra-wideband fmaw (frequency-modulated continuous wave) radar. *The Cryosphere*, 17(1), 175–193. <https://doi.org/10.5194/tc-17-175-2023>

Li, Q, Shen, L, Guo, S, and Lai, Z (2021). WaveCNet: Wavelet Integrated CNNs to Suppress Aliasing Effect for Noise-Robust Image Classification. *IEEE Transactions on Image Processing*, 30, 7074–7089. <https://doi.org/10.1109/TIP.2021.3101395>

Liu, P, Zhang, H, Lian, W, and Zuo, W (2019). Multi-Level Wavelet Convolutional Neural Networks. *IEEE Access*, 7, 74973–74985.

Liu, Y, Cheng, M-M, Hu, X, Bian, J-W, Zhang, L, Bai, X, and Tang, J (2019). Richer Convolutional Features for Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(8), 1939–1946. <https://doi.org/10.1109/TPAMI.2018.2878849>

Mallat, SG (1989). A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), 674–693.

Masson-Delmotte, V, Zhai, P, Pirani, A, Connors, S, Péan, C, Berger, S, Caud, N, Chen, Y, Goldfarb, L, Gomis, M, Huang, M, Leitzell, K, Lonnoy, E, Matthews, J, Maycock, T, Waterfield, T, Yelekçi, O, Yu, R, and (eds.), B. Z. (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (tech. rep.). Intergovernmental Panel on Climate Change (IPCC). Cambridge University Press.

Naveen Kumar, R, Jagadale, B, and Bhat, J (2019). A lossless image compression algorithm using wavelets and fractional Fourier transform. *SN Applied Sciences*, 1(3), 1–8.

Rahnemoonfar, M, Yari, M, Paden, J, Koenig, L, and Ibikunle, O (2021). Deep Multi-Scale Learning for Automatic Tracking of Internal Layers of Ice in Radar Data. *Journal of Glaciology*, 67(261), 39–48.

Ronneberger, O, Fischer, P, and Brox, T (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241.

Simonyan, K, and Zisserman, A (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *International Conference on Learning Representations*, 1–14.

Su, W, Yuan, Y, and Zhu, M (2015). A Relationship between the Average Precision and the Area under the ROC Curve, 349–352. <https://doi.org/10.1145/2808194.2809481>

Thompson-Munson, M, Montgomery, L, Lenaerts, J, and Koenig, L (2022). Surface Mass Balance and Snow Depth on Sea Ice Working Group (SUMup) snow density, accumulation on land ice, and snow depth on sea ice datasets 1952–2019. *Arctic Data Center*. <https://doi.org/10.18739/A24Q7QR58>

Varshney, D, Rahnemoonfar, M, Yari, M, and Paden, J (2020). Deep Ice layer Tracking and Thickness Estimation using Fully Convolutional Networks. *2020 IEEE International Conference on Big Data (Big Data)*, 3943–3952.

Varshney, D, Rahnemoonfar, M, Yari, M, Paden, J, Ibikunle, O, and Li, J (2021). Deep Learning on Airborne Radar Echograms for Tracing Snow Accumulation Layers of the Greenland Ice Sheet. *Remote Sensing*, 13(14). <https://doi.org/10.3390/rs13142707>

Varshney, D, Yari, M, Chowdhury, T, and Rahnemoonfar, M (2021). Refining Ice Layer Tracking through Wavelet combined Neural Networks. *ICML 2021 Workshop on Tackling Climate Change with Machine Learning*. <https://www.climatechange.ai/papers/icml2021/49>

Vaswani, A, Shazeer, N, Parmar, N, Uszkoreit, J, Jones, L, Gomez, AN, Kaiser, L, and Polosukhin, I (2017). Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), *Advances in neural information processing systems* (NeurIPS) (Vol. 30). Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fb0d053c1c4a845aa-Paper.pdf>

Wang, Y, Xu, M, Paden, JD, Koenig, LS, Fox, GC, and Crandall, DJ (2021). Deep Tiered Image Segmentation for Detecting Internal Ice Layers in Radar Imagery. *2021 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6. <https://doi.org/10.1109/ICME51207.2021.9428365>

Williams, T, Li, R, et al. (2018). An Ensemble of Convolutional Neural Networks Using Wavelets for Image Classification. *Journal of Software Engineering and Applications*, 11(02), 69.

Williams, T, and Li, R (2018). Wavelet Pooling for Convolutional Neural Networks. *International Conference on Learning Representations*.

Xie, S, and Tu, Z (2015). Holistically-Nested Edge Detection. *2015 IEEE International Conference on Computer Vision (ICCV)*, 1395–1403. <https://doi.org/10.1109/ICCV.2015.164>

Xiong, Z, and Ramchandran, K (2009). Chapter 18 - wavelet image compression. In A. Bovik (Ed.), *The essential guide to image processing* (pp. 463–493). Academic Press. <https://doi.org/10.1016/B978-0-12-374457-9.00018-4>

Yari, M, Rahmimoonfar, M, and Paden, J (2020). Multi-Scale and Temporal Transfer Learning for Automatic Tracking of Internal Ice Layers. *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 6934–6937. <https://doi.org/10.1109/IGARSS39084.2020.9323758>

Yari, M, Rahmimoonfar, M, Paden, J, Oluwanisola, I, Koenig, L, and Montgomery, L (2019). Smart Tracking of Internal Layers of Ice in Radar Data via Multi-Scale Learning. *2019 IEEE International Conference on Big Data (Big Data)*, 5462–5468.

Cite this article: Varshney D, Yari M, Ibikunle O, Li J, Paden J, Gangopadhyay A and Rahmimoonfar M (2024). Skip-WaveNet: a wavelet based multi-scale architecture to trace snow layers in radar echograms. *Environmental Data Science*, 3: e39. doi:10.1017/eds.2024.25