How to Get the Most Out of U-Net for Glacier Calving Front Segmentation

Maniraman Periyasamy, Amirabbas Davari, Thorsten Seehaus, Matthias Braun, Andreas Maier, *Senior Member, IEEE*, and Vincent Christlein

***Abstract*—The melting of ice sheets and glaciers is one of the main contributors to global sea-level rise. Hence, continuous monitoring of glacier changes and in particular the mapping of positional changes of their calving front is of significant importance. This delineation process, in general, has been carried out manually, which is time-consuming and not feasible for the abundance of available data within the past decade. Automatic delineation of the glacier fronts in synthetic aperture radar (SAR) images can be performed using deep learning-based U-Net models. This article aims to study and survey the components of a U-Net model and optimize the model to get the most out of U-Net for glacier (calving front) segmentation. We trained the U-Net to segment the SAR images of Sjogren-Inlet and Dinsmoore–Bombardier–Edgworth glacier systems on the Antarctica Peninsula region taken by ERS-1/2, Envisat, RadarSAT-1, ALOS, TerraSAR-X, and TanDEM-X missions. The U-Net model was optimized in six aspects. The first two aspects, namely data preprocessing and data augmentation, enhanced the representation of information in the image. The remaining four aspects optimized the feature extraction of U-Net by finding the best-suited loss function, bottleneck, normalization technique, and dropouts for the glacier segmentation task. The optimized U-Net model achieves a dice coefficient score of 0.9378 with a 20% improvement over the baseline U-Net model, which achieved a score of 0.7377. This segmentation result is further postprocessed to delineate the calving front. The optimized U-Net model shows 23% improvement in the glacier front delineation compared to the baseline model.**

***Index Terms*—Glacier calving front segmentation, optimized U-Net, semantic segmentation.**

I. Introduction

THE ice masses of the earth cryosphere are strongly affected by climate change. Data, such as the position of the calving front of marine or lake terminating glaciers, the location where the glacier terminates, or where icebergs calve off, can serve as an indicator for changes in the climatic setting or the ice dynamics of a glacier system. Calving front recession can destabilize the ice flow and enhance further recession, by reducing the buttressing forces at the glacier terminus or unpinning from so-called pinning points (bedrock undulations) of the floating glacier tongue [1]. Glaciers, ice caps, and ice sheets are mainly located in remote places, like high mountain or polar regions, and extend over large regions. This makes remote sensing data ideal for studying changes in ice masses.

Various types of remote-sensing data are used to infer calving front positions. Commonly, multispectral or synthetic aperture radar (SAR) imagery is used. SAR data have the advantage of not being limited by cloud cover or polar night. However, the acquisition geometry and topography of the monitored area can cause effects such as SAR layover and shadowing. Most analyses of calving front positions relied on manual mapping in the past. These manual delineation methods are labor-intensive and error-prone when working with rapidly growing satellite image archives [2]. The water surface in front of the calving front is often covered by icebergs and sea/lake ice, forming the ice-mélange. The surface texture of glacier tongue and ice mélange can be quite similar, making it difficult to separate both, even for an experienced mapper. Hence, automating the glacier (front) segmentation method is a challenge that has been actively studied in the past decades [3]–[8]. Various (semi)automatic routines were developed based on image classification, edge detection, and edge enhancement. The performance of these methods is mostly good, but a major problem arises when the water region at the calving front is covered by ice mélange, which is especially the case in many polar regions [2].

One of the earlier works on automatic segmentation of glaciers was by Sohn and Jezek [3]. They used Roberts cross operator to detect the edges of the Jakobshavn glacier in Greenland. Seale *et al*. [5] proposed a method to delineate the glacier calving fronts automatically and delineated the MODIS data for ten years along the Greenlandic coast. They performed a multitemporal analysis on the image and used a Sobel operator and the brightness gradient to detect the edges. Klinger *et al*. [4], on the other hand, formulated a fully automated approach for delineating the calving fronts based on the active contours (also referred to as "snakes") and nearest neighbor (NN) classifier [9]. This method takes the initial coastline position and uses the classification parameters and the NN classifier to calculate the extent to which the new front position has been warped. All

Manuscript received August 18, 2021; revised November 16, 2021 and January 6, 2022; accepted January 22, 2022. Date of publication February 1, 2022; date of current version February 16, 2022. This work was supported in part by Friedrich-Alexander-Universität Erlangen-Nürnberg and in part by Stadtler-Stiftung under the Emerging Field Initiative TAPE: Tapping the Potential of Earth Observation. *(M. Periyasamy and A. Davari are co-first authors.) (Corresponding author: Amirabbas Davari.)*

Maniraman Periyasamy, Amirabbas Davari, Andreas Maier, and Vincent Christlein are with the Computer Science Department, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany (e-mail: maniraman.periyasamy@fau.de; amir.davari@fau.de; andreas.maier@fau.de; vincent.christlein@fau.de).

Thorsten Seehaus and Matthias Braun are with the Geography and Geosciences Department, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany (e-mail: thorsten.seehaus@fau.de; matthias.h.braun@fau.de).

Digital Object Identifier 10.1109/JSTARS.2022.3148033

these methods are analytical, which have various advantages, such as reduced computation time and easier analysis of detection errors. However, they require either prior knowledge of the data or labor-intensive manual work for pre- or post-processing.

Deep learning-based algorithms, mainly based on convolutional neural networks (CNN) have shown better performance than the analytical methods for image processing, object detection, and segmentation tasks [10]. Deep learning has also been explored in fields where remote sensing imagery is used and high accuracies have been achieved for applications like hyperspectral image analysis, interpretation of SAR images, interpretation of high-resolution satellite images, multimodal data fusion, and 3-D reconstruction [11]. These algorithms learn the information from a given input by transforming it to higher levels of abstract representation using a gradient-based optimization scheme [12]. Pixel-level classification of semantic regions is known as semantic segmentation. Various authors have proposed numerous CNN-based deep learning architectures for image semantic segmentation tasks [13]. Among those, fully convolutional neural network (FCNN) proposed by Long *et al*. [14] and FCNN-based U-Net introduced by Ronneberger *et al*. [15] has gained lots of attention due to their high performance even with low contrast medical images.

U-Nets have also been explored for glacier calving front segmentation. One of the first studies regarding glacier calving front segmentation using U-Net was conducted by Mohajerani *et al*. [7] on Greenland glaciers. They trained a U-Net to segment the input image into the glacier calving front region and nonfront region. They demonstrated a significant performance increase in calving front delineation compared to the analytical method like the Sobel operator. Here, however, the authors resize the large images to a smaller size and feed them into the network, which will lead to potential information loss. Furthermore, all the input images have to be rotated such that the glacier flow is in the same direction. Zhang *et al*. [6] segmented the glacier by tiling the large input image into smaller tiles and trained the U-Net with these patches. In the next step, they postprocessed the results to extract the glacier calving front. Here the authors showcased a small test error suggesting that the accuracy of a well-trained network can be close to the human level [6]. But, tiling the image into smaller tiles leads to a class imbalance in the patches and makes the training process challenging. To tackle the class imbalance problem, the authors [6] have dropped 80% of the tiles with only one class in them. This class imbalance issue is handled via reformulating the segmentation problem into a pixel-wise regression task by Davari *et al*. [16] in their study on Antarctic Peninsula glaciers. The authors construct a distance map representing each pixel's distance from the calving front and train a U-Net to approximate this distance map given a SAR image. The calving front is further extracted from this distance map during postprocessing. The same class imbalance issue is tackled by tuning the learning hyperparameter in another study [17]. This is done in two steps; first, the Mathews correlation coefficient metric is used as a metric for the validation set to decide early stopping and to avoid overfitting; second, they proposed a distance map-based BCE loss function for optimizing the U-Net. In both studies, the authors demonstrate the impact of fine-tuning the individual aspects of a deep learning model for calving front delineation and its overall performance gain. Another study, which uses U-Nets for glacier calving front segmentation, has been conducted by Baumhoer *et al*. [18]. The authors explore the performance of U-Net trained on different glaciers. Finally, a comprehensive study on glacier front detection for 66 Greenland glaciers using over 1500 optical images with manual delineations and a convolutional neural network was conducted by Cheng *et al*. [8]. They constructed their architecture by adapting the DeepLabV3+ Xception model with modifications. This modified model provides a probability mask output from which the calving front is extracted via substantial postprocessing. Though Cheng *et al*. performed their study on large glacier ensembles, the entirety of the data is too large to train in a reasonable time. To accommodate the large dataset, the authors have greatly down-scaled large images so that the training can be performed with limited resources and time. This resizing may lead to potential information loss.

While the U-Net architecture in its original form shows a high affinity to detect glacier calving fronts in SAR images compared to the analytical methods, various issues and imperfections negatively impact its performance. Some of the major issues in training a U-Net on SAR image are explained below.

A clear and separable representation of information in the data plays a vital role in the performance of a deep learning model [19]. However, SAR images are prone to speckle noises [20], [21] and exhibit low contrast between the glacier and nonglacier regions [18]. As the nature of the noise and low contrast issues are well known, preprocessing steps with various filters were suggested [22], [23]. Besides noisy data, the lack of sufficiently well-annotated glacier image data is a major obstacle to successfully training a U-Net. Lack of data becomes a problem because deep learning models rely on a diverse training dataset, which represents the problem domain completely to avoid over-fitting and learn better features [24]. Data augmentation increases the amount of annotated training data, which in a way serves as regularization to prevent the model from over-fitting [25], [26].

Another challenging factor, while training a U-Net for glacier calving front segmentation, is the class imbalance between glacier calving front and nonfront regions in the SAR image. The class imbalance in the training dataset causes the U-Net to misclassify most of the regions in the glacier image as a nonglacier calving front region because of its higher prior probability [27]. The negative effects of class imbalance while training a model can be reduced via appropriate data-level manipulation or algorithm-level constraints [27]. Data-level manipulation such as data augmentation and dynamic sampling [28]–[32] are effective in handling class imbalance issue, but increases the computational cost. Simultaneously, the algorithm-level constraints such as loss functions are case-specific and not a generic method [27]. Further, the U-Net also suffers in extracting features that capture non-local information at the lower layers of the encoder part [33]. This happens due to the small receptive fields at the bottleneck layers. The final issue in training a U-Net for glacier (front) segmentation is its slow convergence rate. Activation functions such as rectified linear unit (ReLU), which

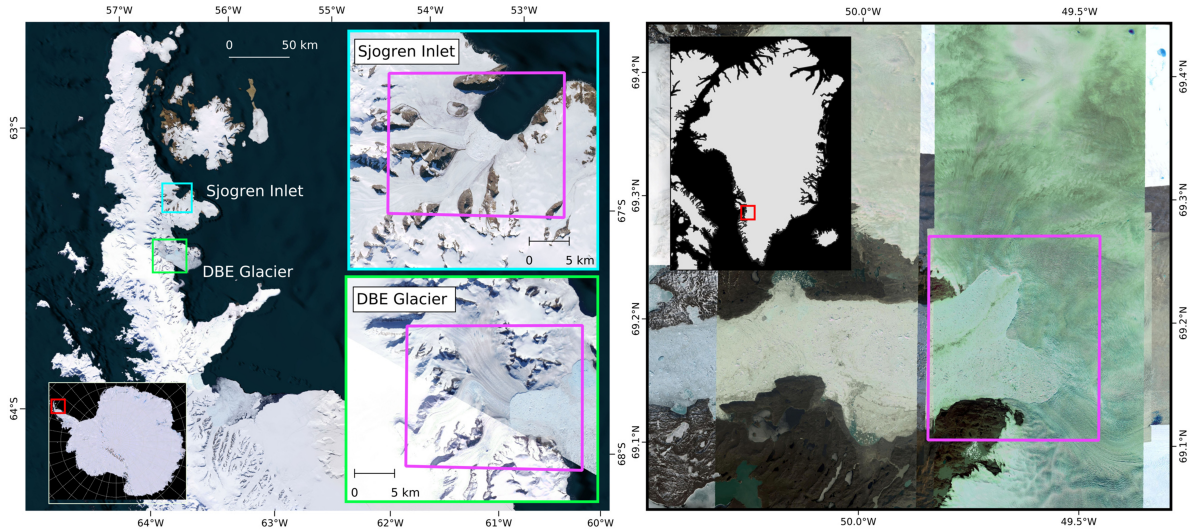


Fig. 1. Overview of study area at the northern Antarctic Peninsula. Subsets show analyses glacier systems. Pink polygons indicate areas of interest used in the analysis. Background: ESRI Satellite © ESRI; Map of Antarctic and Antarctic Peninsula: Landsat LIMA Mosaic © USGS, NASA, BAS, NFS.

are not zero-centered, cause internal covariate shift, thereby leading to slower convergence [34]. The U-Net architecture uses the ReLU activation function after every convolution layer. Hence, it is prone to internal covariate shift, which can be addressed using normalization and regularization techniques.

This study addresses all the abovementioned issues, both on data-level and algorithm-level that could hinder the performance of a U-Net for glacier calving front segmentation, by comparing, analyzing, and optimizing the U-Net components in the following six stages:

1) data preprocessing;
2) data augmentation;
3) optimal loss function;
4) optimal bottleneck;
5) optimal normalization; and
6) optimal dropout ratios.

The rest of this article is structured as follows. In Section II, we explain the data used in this study. Afterward, the base segmentation model is introduced in Section III. In Section IV, we propose the modifications and evaluate their effects. We evaluate our optimized model against two strong state-of-the-art models in Section VI and show its superiority. In Section V, we conduct an ablation study to evaluate the effect of different transfer learning strategies. Finally, we compute the delineation of the calving front in Section VII. Section IX concludes this article.

II. Data

A. Study Sites

Along the Antarctic Peninsula and Greenland's margins, marine-terminating outlet glaciers are draining the respective ice sheets. These outlet glaciers are typically located in narrow fjords. A pronounced retreat of the glacier calving fronts was reported along the coastline of the Antarctic Peninsula for the last decades [35], [36]. In particular, the retreat and disintegration of ice shelves along the Antarctic Peninsula has strongly affected the ice dynamics of the tributary glaciers, leading to glacier thinning, further recession, and high ice losses [37] and [38]. In this study, we selected outlet glaciers on the Antarctic Peninsula and Greenland to test and analyze the capabilities of CNN for calving front detection on SAR intensity images. The Sjögren-Inlet (SI) and Dinsmoore–Bombardier–Edgworth (DBE) glacier systems, which are major tributaries to the Prince–Gustav-Channel and Larsen-A ice shelves, respectively, were selected as study sites on the Antarctic Peninsula. At both sites, a comprehensive analysis of the reaction of the glaciers to the break-up of the ice shelves in 1995 was carried out by [38] and [39]. Strong acceleration of the ice flow and a subsequent long-term slow down was observed at both glacier systems. Additionally, a pronounced and rapid recession of the calving front shortly after the ice shelf disintegration, followed by variations in the calving front positions, was mapped at both sites. In Greenland, we selected Jakobshavn Isbræ for our investigations. It is one of the major outlet glaciers of the Greenland ice sheet and drains into the ocean on the central west coast of Greenland. Within the last decades, the strong retreat of the calving front and pronounced variations in the ice flow speed were observed [40]–[42].

B. SAR Imagery Processing

We used satellite SAR imagery from the ERS-1/2, Envisat, RadarSAT-1, ALOS, TerraSAR-X (TSX), and TanDEM-X (TDX) missions at both study sites at the Antarctic Peninsula, acquired between 1995 and 2014. SAR speckle was reduced by multilooking the acquisitions. The multilooked intensity images were geo-referenced and ortho-rectified by means of the ASTER digital elevation model from [43]. The processing of the SAR imagery was done using the GAMMA RS Software. At our study site in Greenland, and Jakobshavn Isbrea, we used the same acquisitions as [6]. The authors kindly provided us with the

TABLE I
OVERVIEW OF THE SAR SENSORS AND SPECIFICATIONS USED TO ACQUIRE THE SAR IMAGES OF THE ANTARCTIC PENINSULA AND JAKOBSHAVN ISBREA DATASETS

Platform	Primary data format	Sensor	Mode	SAR band	Repetition cycle [d]	Time interval	Multi looking factor	Ground range resolution [m]	Number of SAR images
Antarctic Peninsula									
ERS-1/2	SLC	SAR	IM	C band	35/1	13. November 1995 - 26. February 2010	1x5	20	27
RADARSAT	SLC/PRI	SAR	ST	C band	24	10. September 2000 - 20. January 2008	1x4	20 (SLC) 12.5 (PRI)	51
Envisat	SLC	ASAR	IM	C band	35	05. December 2003 - 03. July 2010	1x5	20	55
ALOS	SLC	PALSAR	FBS	L band	46	18. May 2006 - 03. March 2011	2x5	16,7	26
TerraSAR-X TanDEM-X	SLC	SAR	SM	X band	11	13. October 2008 - 22. December 2014	3x3	6.7	85
Jakobshavn Isbrea									
TerraSAR-X	EEC	SAR	SM	X band	11	16. April 2009 - 23. December 2015	5x5	6	149

routines to carry out the same processing on the TSX enhanced-ellipsoid-corrected (EEC) imagery. A detailed description of the processing can be found in [6]. The specifications and parameters of the SAR sensors and imagery are provided in Table I.

C. Labeled Image Generation

The manually mapped calving front locations at DBE and SI glacier from [38] and [39], and at Jakobshavn Isbrea from [6] were used to generate the training labels. The preprocessed SAR acquisitions were cropped to the areas of interest as shown in Fig. 1 and converted to 16-b single-channel images. Classified 8-b single-channel images showing the defined classes (glacier and nonglacier regions) were generated using the respective glacier calving front position (spatial lines) in combination with the coastline and rock-outcrop polygons (Antarctic Peninsula: from the Antarctic Digital Database, Jakobshavn Isbrea: manually defined).

D. Data Statistics

The Antarctic Peninsula dataset consists of 244 images with various image sizes ranging from 458×377 to 4581×3370 pixels. Apart from varying image sizes, the images in the dataset also have different spatial resolutions, ranging from 6.7 to 20 m per pixel. From these 244 images, 50 were chosen at random for the test set. From the remaining images, 50 were chosen for the validation set, and the rest were used as the training set. The Jakobshavn Isbræ dataset consists of 149 images with a constant spatial resolution of $6\,\mathrm{m}$ per pixel and image size being either 3644×1642 or 3563×1650. From these 149 images, 30 were chosen at random for the test set. From the remaining images, 30 were chosen for the validation set, and the rest became the training set.

III. BASE SEGMENTATION MODEL

The base segmentation model is derived from the U-Net architecture for glacier calving front segmentation proposed by Zhang *et al*. [6]. Our work focuses on optimizing the U-Net architecture from a deep learning point of view. Zhang *et al*. [6] adapt a common U-Net architecture for glacier calving front segmentation in SAR imagery. Hence, it has been chosen as the baseline method in this work.

A. U-Net

U-Net is a deep learning-based image segmentation model introduced by Ronneberger *et al*. [15] for medical images. U-Net is intended to address one of the significant issues in the field of medical image segmentation: the lack of large datasets. In general, the U-Net architecture consists of two parts: an encoder and a decoder. The encoder extracts the features from the input image via multiple consecutive convolution operations. A pooling operation follows each convolution operation to reduce the overall computation cost without losing vital information. The decoder decodes the data from the encoder via multiple upsampling operations and convolution operations. A series of skip connections between the encoder and decoder are implemented to provide the spatial information of the input to the decoder. The decoder, with the decoded information, provides a pixel-wise classification result for the input image. The last layer of the encoder that transmits the encoded features and the first layer of the decoder that receives the encoder's information surround the bottleneck of the U-Net architecture. These layers decide the amount of information and the quality of the information transmitted from the encoder to the decoder.

B. Hyperparameters

Zhang *et al*. [6] trained this model with input images of size 960×720 as all the images in his study were larger than this. In contrast, the dataset used for this study consists of images with varying image sizes. Thus, the input size for the

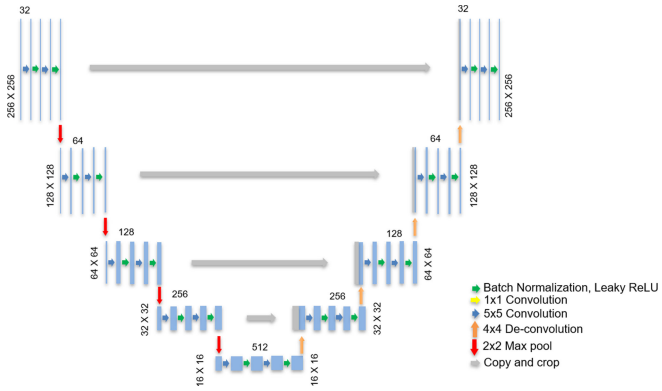


Fig. 2. Base segmentation model used in this study.

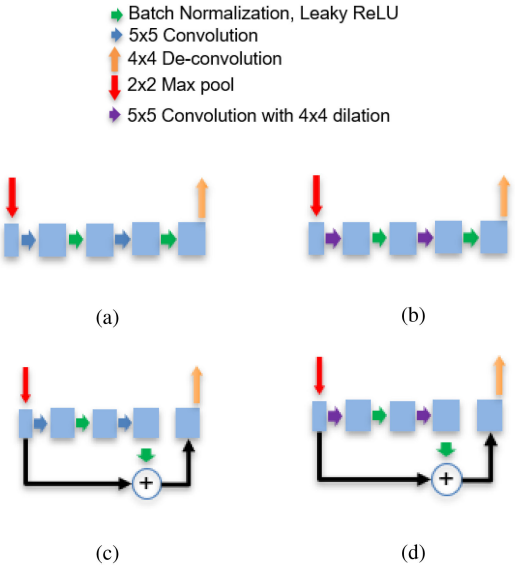


Fig. 3. Various bottlenecks of U-Net. (a) Bottleneck in the baseline segmentation model, (b) U-Net with dilated bottleneck, (c) U-Net with residual bottleneck, and (d) U-Net with dilated bottleneck and residual connection.

TABLE II
COMMON HYPER-PARAMETERS USED IN THIS STUDY

Parameter	Value
Input size	256×256
Max epochs	256
Batch size	16
Optimizer	Adam
Learning rate	1. Exp. III-D - constant $[1e-3]$, 2. Exp. IV-A to IV-F, V - Cyclic $[1e-7, 1e-3]$.
Weight Initializer	He Initialization
Stopping criteria	1. End of max epochs, 2. Early stopping with the patience of 30 epochs.

base segmentation model of this study is set to 256×256. The architecture of the base segmentation model with the number of feature-maps extracted after every block is shown in Fig. 2. The parameters given in Table II are constant for all the experiments conducted in this study. The batch size is fixed to 16 due to hardware limitations. The maximum number of epochs was fixed to 256 with a high patience of 30 epochs on validation loss as stopping criteria in the early stopping regime for finding the

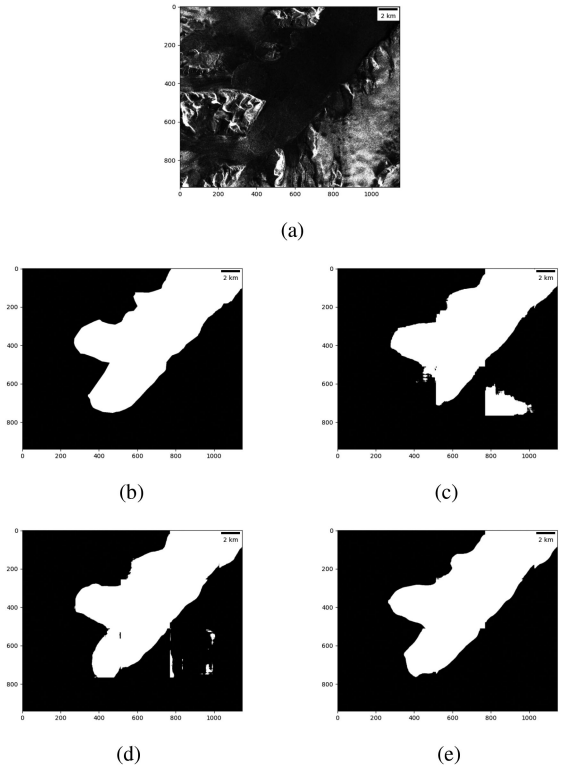


Fig. 4. Qualitative results of different bottleneck architectures. The SAR image is acquired using RADARSAT 1 (SAR), on March 12, 2007. (a) Original SAR image. (b) Ground truth. (c) Residual connection. (d) Dilated convolution. (e) Residual + Dilated.

optimal convergence. The optimizer and weight initializers were chosen based on the reference work [6]. The train and validation losses were used to validate the optimal convergence to a solution without the model getting under- or overfitted.

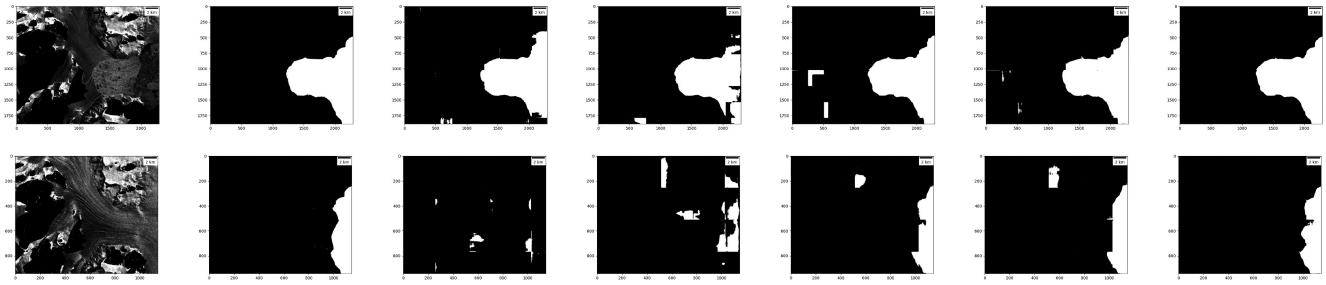
C. Evaluation Metrics

To evaluate and quantize the performance of the models, we use three commonly used metrics in the image segmentation community [44], namely, intersection over union (IOU), pixel-wise accuracy, and Dice coefficient. These metrics are mathematically expressed in (1)–(3), respectively. In these equations, TP denotes the number of true positive predictions, TN is the number of true negative predictions, FP is the number of false positive predictions, and FN is the number of false negative predictions. Out of the three, pixel-wise accuracy is the most relaxed performance metric. Pixel-wise accuracy metric is the ratio between correct predictions over the total number of predictions and does not take the wrong predictions into consideration. However, IOU and Dice coefficient on the other hand penalizes the false positives and false negatives more stringently

$$\text{IOU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{Pixel-wise accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{2}$$

$$\text{Dice coefficient} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}}. \tag{3}$$



(a) (b) (c) (d) (e) (f) (g)

Fig. 5. Qualitative results on segmenting (a) SAR image for each experiment in comparison with the (b) ground truth: (c) results from baseline segmentation model, (d) after data preprocessing, (e) data augmentation, (f) optimal loss, and (g) optimal bottleneck. The SAR image in the first row is acquired using TerraSAR-X (SAR) on June 10, 2012 and the SAR image in the second row is acquired using ERS-1/2 (SAR) on August 28, 1993.

D. Baseline Segmentation Results

The baseline segmentation model was trained and tested with the Antarctic Peninsula dataset. This model took 142 epochs to converge using the Adam optimization method with a learning rate of 1×10^{-3} and binary cross-entropy (BCE) loss. The model results in an IOU score of 0.6224, pixel-wise accuracy of 0.9219, and a dice coefficient of 0.7377. These results are the baseline against which our proposed modifications are compared.

IV. Optimizing U-Net Model for Glacier Segmentation

The performance of a U-Net depends on the following five aspects:

1) how clear and separable the key pieces of information in the data are [19];
2) how good the constraints for the optimization are [45];
3) how generalizable the model is [46];
4) the prior knowledge of the model [47]; and
5) how good the model can extract features from the input [48].

All these five aspects of a U-Net model are analyzed and optimized in the following seven aspects:

1) data preprocessing;
2) data augmentation;
3) optimal loss function;
4) normalization layer;
5) dropout;
6) optimal bottleneck of U-Net; and
7) transfer learning.

All suggested improvements are evaluated directly and sequentially processed. Changes to the baseline segmentation model or the data that yields the highest performance upgrade in the current experiment are used in the subsequent experiments. The order of the experiments is chosen randomly as the individual experiments are independent of each other and can be re-ordered or removed if needed.

A. Data Preprocessing

Low contrast regions and speckle noise present in SAR images make the detection of edges and boundaries between glacier and ocean very hard. A bilateral filter with a 10-pixel neighborhood, color sigma 95, coordinate sigma 75, and a contrast limited adaptive histogram equalization (CLAHE) filter with a clip limit of 5.0 and tile grid size of 5×5 were applied to further reduce the noise and increase the contrast within the different image regions in the Antarctic Peninsula glacier images. Baseline segmentation model, trained and tested on the preprocessed dataset, obtained an IOU score of 0.6805, pixel-wise accuracy of 0.9293, and dice coefficient of 0.7787, which represents an 4.1% improvement compared to the initial results. Fig. 5(d) presents the qualitative results of the data preprocessing experiment.

B. Data Augmentation

Next, we studied the effect of data augmentation on the baseline segmentation model for glacier image segmentation. The geometric transformations to be applied to the preprocessed Antarctic Peninsula dataset for augmenting the train set are 0°, 90°, 180°, 270° rotations and a vertical flip. This data augmentation resulted in an eightfold increase in the amount of data, from 144 images to 1152 images for training. Though this eight-fold increase might bring better generalization to the model, this also drastically increases the training time. Hence, we also trained, tested, and evaluated twofold data augmentation performance by applying only the vertical flip transformation to the dataset. Both the results of eight-fold and twofold data augmentation for all the metrics are given in Table III. Data augmentation has improved the dice coefficient drastically to 0.9026. The qualitative results in Fig. 5(e) show that the data augmentation has increased the true-positive predictions by the model.

C. Optimal Loss Function

In a deep neural network (DNN), the loss function is the objective function that has to be minimized. For an accurate and faster convergence to the optimal solution, the loss function should cover all possible input distribution. The author of the article [49] has categorized and defined the loss function into four types. They are, distribution-based such as BCE loss, region-based such as Dice loss [50], boundary-based loss, and compound loss. Equations (4) and (5) define the BCE loss and Dice loss, respectively, where $\vec{y}$ is the ground truth, i.e., the manual annotations generated by experts, $\hat{\vec{y}}$ is the prediction, M

TABLE III
QUANTITATIVE RESULTS OF THE BASELINE SEGMENTATION MODEL AND THE SUGGESTED IMPROVEMENTS

	Variables	Metrics			Improvement
		IOU	pixel-wise acc.	Dice coefficient	
Baseline model	-	0.6224	0.9219	0.7377	-
Pre-processing	-	0.6805	0.9293	**0.7787**	4.1 %
Data Augmentation	Flip	0.7498	0.9471	0.8268	7.30 %
	Flip + Rotations	0.8340	0.9599	**0.9026**	
Optimal Loss	1.0 · BCE + 0.0 · Dice	0.8340	0.9599	0.9026	0.90 %
	0.8 · BCE + 0.2 · Dice	0.8400	0.9608	0.9086	
	0.5 · BCE + 0.5 · Dice	0.8464	0.9646	**0.9117**	
	0.2 · BCE + 0.8 · Dice	0.8234	0.9591	0.8960	
	0.0 · BCE + 1.0 · Dice	0.7258	0.9327	0.8234	
Optimal Bottleneck	Residual Connection	0.8538	0.9675	0.9122	2.61 %
	Dilated Convolution	0.8879	0.9740	**0.9381**	
	Residual + Dilated	0.8870	0.9740	**0.9378**	
Normalization Layer	**BatchNorm**	0.8870	0.9740	**0.9378**	0.00
	LayerNorm	0.7863	0.9482	0.8696	
	InstanceNorm	0.7415	0.9373	0.8349	
	GroupNorm	0.7716	0.9453	0.8621	
	WeightNorm	0.8331	0.9574	0.9018	
Dropout Layer	0.1	0.8846	0.9704	0.9321	0.00
	0.2	0.8037	0.9494	0.8765	
	0.3	0.6510	0.9012	0.7725	
	0.4	0.5130	0.8253	0.6605	
	0.5	0.3508	0.6885	0.4993	

The improvements are calculated using the dice-coefficient of the best-performing model highlighted in bold.

is the number of inputs, and K is the number of classes

$$L_{\text{BCE}}(\vec{W}) = -\sum_{m=1}^{M}\sum_{k=0}^{K} y_{k,m}\ln(\hat{y}_{k,m}) \tag{4}$$

$$L_{\text{Dice}}(\vec{W}) = 1 - \frac{2\sum_{m=1}^{M}\sum_{k=0}^{K} y_{k,m}\hat{y}_{k,m}}{\sum_{m=1}^{M}\sum_{k=0}^{K} y_{k,m}^2 + \sum_{m=1}^{M}\sum_{k=0}^{K} \hat{y}_{k,m}^2}. \tag{5}$$

Different loss functions can tackle different issues such as class imbalance, label noise robustness, etc. [51]. For segmentation problems with an issue in a dataset like class imbalance and regions with no information, a weighted combination of BCE and DICE loss have shown high-performance [52]. In this stage of the experiment, we test different combinations of BCE and DICE loss, i.e.,

$$L_{\text{Total}} = \alpha \cdot L_{\text{BCE}} + \beta \cdot L_{\text{Dice}} \tag{6}$$

where α, β are weights for the losses L_{BCE} and L_{Dice}, respectively. L_{BCE} and L_{Dice} are given in (4) and (5). The model gave the best performance for an equally weighted compound loss combining L_{BCE} and L_{Dice}, with the weight α and β being 0.5. Fig. 5(f) shows the qualitative results of the baseline segmentation model with the optimal loss function. The use of this optimal loss function increased the dice coefficient of the model from 0.9026 to 0.9117 and has reduced the false positives cases while increasing the true positives along the boundary. The metric results for various combinations of loss functions tested during this experiment are given in Table III.

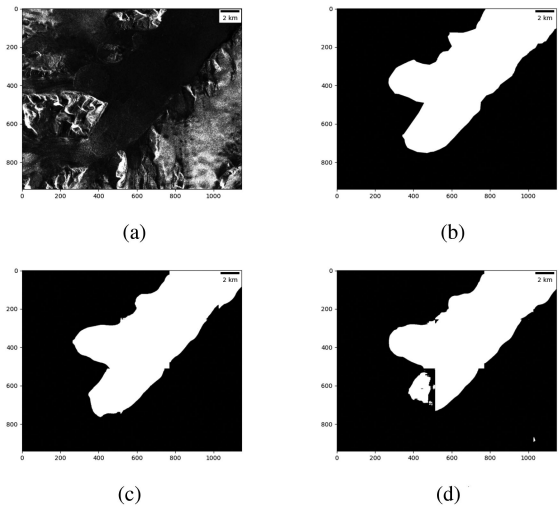


Fig. 6. Effect of adding dropout layer to the optimal model. The SAR image is acquired using RADARSAT 1 (SAR), on March 12, 2007. (a) Original SAR image. (b) Ground truth. (c) Without dropout. (d) With drouput.

D. Optimal U-Net Bottleneck

The bottleneck acts as the lowest layer of the encoder block, which extracts the last set of features with which the decoder decodes and reconstructs the pixel-wise classification output. Due to the nature of U-Net's architecture, the lower layers have a small receptive field that prevents the network from extracting features that capture nonlocal information [33]. Using dilated convolutions with a residual connection as a bottleneck has been shown to increase the receptive field and to enhance the gradient

flow during the backward pass [33]. This experiment is divided into two parts. The first part of the experiment is conducted to find the optimal dilation rate for convolution layers in the bottleneck. Here, we replaced all the convolution layers in the bottleneck with dilated convolution layers, and the dilation rates of (2, 2), (4, 4), (8, 8), (2, 2) + (4, 4), and (2, 2) + (4, 4) + (8, 8) were tested to find the optimal dilation rate. The baseline segmentation model's architecture with dilated bottleneck is shown in Fig. 3(b). The dilation rate of (4, 4) produced the highest dice coefficient of 0.9381. Hence, this model was further compared with other models in the second part of this experiment to find the optimal bottleneck.

The second part compares the model with dilated bottleneck against the model with the residual connected bottleneck [see Fig. 3(c)] and the residual connected dilated bottleneck [see Fig. 3(d)]. The numeric results are given in Table III.

The dice coefficient results of these three models were very close to each other; therefore, the qualitative results were compared. From the qualitative results, as shown in Fig. 4, we observe that the U-Net with combined dilated bottleneck and residual connection has near-perfect predictions along the boundary with very few false positives and false negatives. Hence, dilated bottleneck with residual connection is chosen as the optimal bottleneck for the U-Net for glacier segmentation. The prediction results of the baseline segmentation model with optimal bottleneck and all optimizations from previous experiments (from now on referred to as the optimized model) are depicted in Fig. 5.

E. Normalization Layer

Normalization layers play a key role in DNNs, such as achieving generalization and faster convergence [34]. In this experiment, we intend to analyze the effect of different normalization techniques on the U-Net for glacier segmentation. Note that in the optimized model, batch normalization is applied after every convolution operation. We replaced the batch normalization with layer normalization, instance normalization, group normalization, and weight normalization one after the other. The quantitative results of various normalization layers are shown in Table III. We can conclude that batch normalization as proposed in the baseline segmentation model performed the best for glacier segmentation.

F. Dropout

In this experiment, we analyze the effect of dropout on the U-Net for glacier segmentation. All the normalization layers are replaced with a dropout layer. Dropout rates in the range of [0.1, 0.5] with an incremental step size of 0.1 were tested. The optimal model did not converge with dropout replacing the normalization layers. On the other hand, it has been observed that adding a dropout layer after every normalization layer increased the overall performance of the model [53]. Hence, the case where a dropout layer is added after every normalization layer was also tested. Unlike replacing the normalization layers with dropout, adding a dropout after every normalization layer lets the network converge. Table III shows that the model with a dropout of 0.1 performed similarly to the optimized model from the previous stage. Nevertheless, as shown in Fig. 6, the qualitative results depict that the dropout layer increases the false-negative cases in the prediction. Hence, we conclude that the use of dropout layers does not improve the model performance.

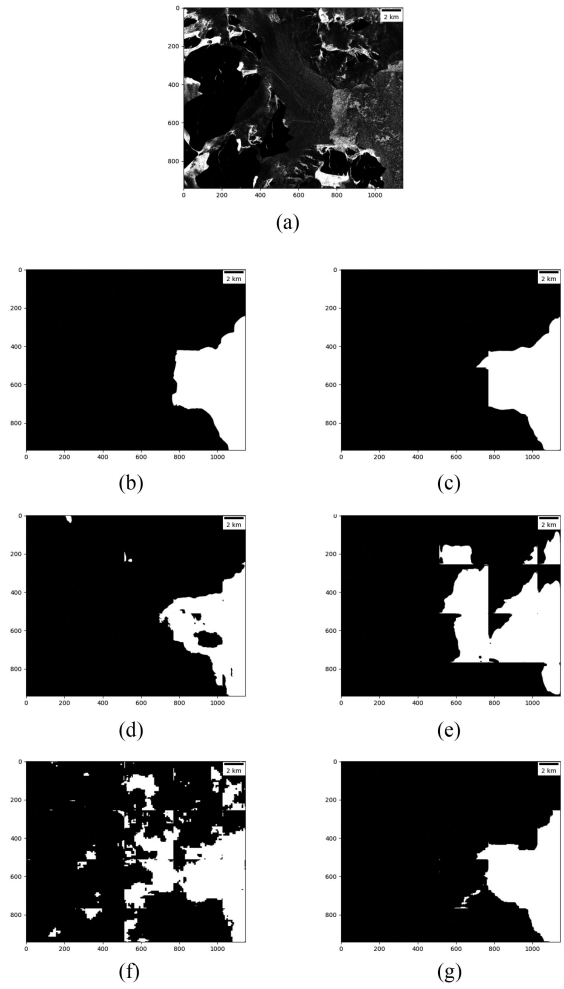


Fig. 7. Qualitative results of benchmarked segmentation models. The SAR image is acquired using ERS-1/2 (SAR) on January 12, 2005. (a) Original SAR image. (b) Ground truth. (c) Optimized U-Net. (d) Nested U-Net with SGD. (e) Nested U-Net with Adam. (f) FCNN with Adam. (g) FCNN with SGD.

V. Transfer Learning

The knowledge gained by training our optimal model from the previous experiment on DBE and SI glaciers is transferred and used for training the model to segment Jakobshavn Isbræ glacier. To analyze the extent of further training required after knowledge transfer, we tested the following three different protocols.

1) Full transfer learning and full training, where the knowledge from both encoder and decoder blocks of the U-Net is transferred, and both the blocks are trained further.
2) Full transfer learning and encoder training: The knowledge from both encoder and decoder blocks of the U-Net are transferred, but only the encoder block is trained further while the decoder block is not trained any further and is kept frozen.

TABLE IV
PERFORMANCE OF THE MODEL TRAINED FROM SCRATCH AND MODELS TRAINED USING TRANSFER LEARNING

Training type	IOU	pixel-wise acc.	Dice coeff.
Train from scratch	0.9674	0.9908	0.9833
Transfer + Full Train	0.9188	0.9764	0.9572
Transfer + Encoder Train	0.9193	0.9765	0.9575
Transfer + Decoder Train	0.9574	0.9878	0.9781

3) Full transfer learning and decoder training: The knowledge from both encoder and decoder blocks of the U-Net are transferred, but only the decoder block is trained further while the encoder block is not trained any further and kept frozen.
4) All these transfer learning methods were tested and compared against a model trained from scratch on the Jakobshavn Isbrae glacier.

The quantitative results given in Table IV show that the model trained from scratch (protocol 4) and the model with the full transfer learning and decoder training (protocol 1) exhibit a comparable numerical performance. However note that the transfer learning model (protocol 3) converged already after 5 epochs, whereas the model trained from scratch needed 64 epochs. This shows that transfer learning is computationally very efficient with a minimal loss in accuracy compared to training from scratch.

VI. COMPARISON WITH STATE OF THE ART

Finally, we intend to compare the performance of the optimized model from this study against the state-of-the-art technologies for image segmentation. Two of the most widely used image segmentation architectures next to U-Nets are FCNN [14] and a variation of the U-Net denoted as Nested U-Net [54].

A. Fully Convolutional Neural Network

The FCNN architecture is called a fully convoluted neural network as the fully connected layers, which were traditionally used as the final set of layers in the model, are replaced with 1×1 convolution layers. The multichannel 1×1 outputs from the last layer will be upsampled to produce the pixel-wise classification results. One major drawback with this upsampling procedure is that the multichannel 1×1 outputs from the network will contain the class information of the pixel, whereas its spatial information is lost. Hence, the authors used skip connections to the network's final prediction and previous layers to provide the spatial information during the upsampling process. FCNN is one of the first high-impact CNN-based segmentation models [55].

B. Nested U-Net

Nested U-Net is an advancement of the U-Net architecture. It was introduced by Zhou *et al*. [54] for medical image segmentation. This deeply supervised architecture resembles the U-Net architecture but with more layers and a series of nested skip connections. Nested U-Net implements multiple convolutions in the encoder and transposed convolutions (*aka.* deconvolutions) in the decoder, which are interconnected and let the optimizer choose the most optimal path for the segmentation task. Due to the higher number of convolution and deconvolution layers, the number of trainable parameters in the network is much higher compared to the standard U-Net architecture. Thus, the computational cost for the training of a Nested U-Net also increases compared to the U-Net model.

Note, the original FCNN has very few trainable parameters compared to our optimized model. Hence, to make a fair comparison, we implemented a larger version of the FCNN with the number of parameters being comparable to the Nested U-Net. Both architectures were trained from scratch on the preprocessed and augmented DBE and SI glacier dataset using Adam and stochastic gradient descent (SGD) optimizers. Hence, the optimized model from this study was benchmarked against four different models, namely, Nested U-Net with Adam optimizer, Nested U-Net with SGD optimizer, FCNN with Adam optimizer, and FCNN with SGD optimizer.

TABLE V
EVALUATION RESULTS OF BENCHMARKED SEGMENTATION MODELS

Model	Parameters	IOU	pixel-wise acc.	Dice coeff.
FCN + Adam	∼30 million	0.5143	0.8448	0.6671
FCN + SGD	∼30 million	0.8150	0.9559	0.8892
Nested U-Net + Adam	∼28 million	0.5636	0.8485	0.7106
Nested U-Net + SGD	∼28 million	0.7586	0.9440	0.8517
Optimized U-Net model	∼24 million	0.8724	0.9672	0.9248

C. Comparison

The evaluation metric results and the number of trainable parameters in each model are given in Table V. FCNN with Adam converged within 25 epochs, whereas Nested U-Net with Adam and FCNN with SGD took 40 epochs and 53 epochs, respectively. Nested U-Net with SGD completed the maximum number of epochs without converging to a solution. Fig. 7 shows qualitative results of all the benchmarked models.

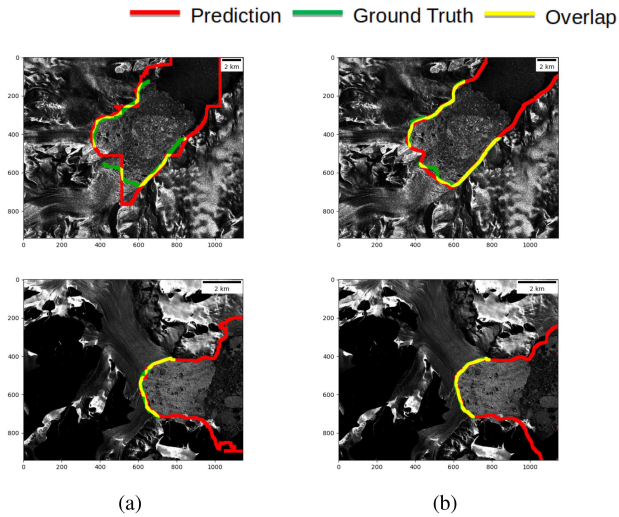
From the evaluation metrics, qualitative results, and the number of epochs needed for convergence, we can conclude that all the other state-of-the-art models have either obtained a bad local optimal solution or did not obtain a solution at all. Our optimized U-Net with fewer trainable parameters is thus preferable to the other two state-of-the-art models.

VII. DELINEATION OF CALVING FRONT

The delineation of the calving front from the predicted glacier region is performed in two steps. First, the false positives are removed by dropping all the contours except for the contour with the largest connected component. Then, the edge of this area is delineated using a canny edge detector. This detected edge is evaluated using dice coefficient. The dice coefficient was chosen as the metric over others to evaluate the performance of the model on delineating the calving front as it penalizes the false positives and false negatives more stringently. However, the dice coefficient metric on the calving front is tough to evaluate as the metric will consider the prediction a miss even if it shifted by one pixel. Hence, to make the metric more compendious,

TABLE VI
DICE COEFFICIENT SCORE ON CALVING FRONT DELINEATION WITH VARIOUS TOLERANCES

Glacier	Tolerance	Baseline Model	Optimized Model
Antarctic Peninsula	50 m	0.1660	0.3425
	100 m	0.2433	0.4538
	150 m	0.3239	0.5718
Jakobshavn Isbræ	50 m	0.4985	0.6181
	100 m	0.6183	0.7096
	150 m	0.7138	0.7838



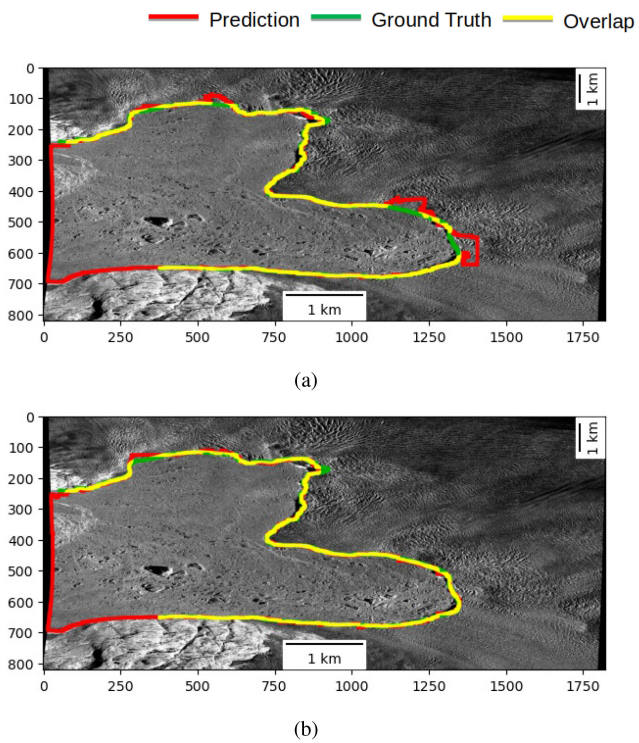
(a) (b)

Fig. 8. Calving front delineation by (a) baseline segmentation model, and (b) optimized U-Net model on Antarctic Peninsula glacier. The SAR image in the first row is acquired using ERS-1/2 (SAR) on 15.10.2004 and the SAR image in the second row is acquired using TerraSAR-X (SAR) on 10.6.2012.

we dilate the prediction and ground truth such that it has a tolerance on the nearby predictions. The dice coefficient scores for calving front predictions with a tolerance of 50 m, 100 m, 150 m are given in Table VI. The tolerance of 50 m, 100 m, 150 m is translated to the number of pixels using the ground range resolution of the corresponding satellite mission given in Table I. Figs. 8, and 9 depict the calving front delineation via the U-Net model optimized in this study.

VIII. OPTIMIZATION RESULTS

To obtain the best results using a U-Net as base architecture for glacier segmentation, we studied five main aspects of the U-Net model in seven different experiments. Our baseline model started with a dice coefficient of 0.7377. We first optimized the data aspect of the model by analyzing the effects of data preprocessing and data augmentation on the Antarctic Peninsula dataset. Optimizing the data via data preprocessing and data augmentation resulted in an 4.1% and 7.3% increase in performance from the baseline segmentation model, respectively. We then enhanced the feature extraction capability of the model by finding the optimal loss constraint resulting in a 0.9% increase in performance, finding the optimal bottleneck resulting in a 2.1% increase in performance, and finding the optimal normalization for the glacier segmentation task. The optimized U-Net attained a dice coefficient score of 0.9378, which is a 20% increase



(a) (b)

Fig. 9. Calving front delineation by (a) baseline segmentation model and (b) optimized U-Net model on Greenland glacier. The SAR image is acquired using TerraSAR-X (SAR) on 6.10.2014.

compared to the dice coefficient score of 0.7377 obtained by the baseline segmentation model [6].

IX. CONCLUSION

This article optimizes the U-Net architecture for glacier segmentation and postprocess the segmentation results to delineate the calving front. Postprocessing the resulting segmented glaciers from optimized U-Net model to delineate the calving front led to a dice coefficient of 0.5718 with a tolerance of 150 m. This is a 23% improvement compared to the baseline segmentation model, which obtained a dice coefficient of 0.3239. The optimized U-Net model attained similar performance improvement on the Greenland glacier when trained from scratch. Hence, it can be concluded that the presented optimization scheme is highly effective for glacier calving segmentation in SAR imagery and future works can benefit from it in their segmentation pipelines. Additionally, complementary works aiming to improve the loss function and hyper-parameter optimization could be explored to improve the accuracy further [16], [17].

ACKNOWLEDGMENT

The authors would like to thank the German Aerospace Center (DLR), the European Space Agency (ESA), and the Alaska Satellite Facility (ASF) for providing the SAR data for the analysis. The authors would also like to thank financial support by Deutsche Forschungsgemeinschaft and Friedrich-Alexander-Universität Erlangen-Nürnberg within the funding programme "Open Access Publication Funding".

REFERENCES

[1] J. Fürst *et al.*, "The safety band of antarctic ice shelves," *Nature Climate Change*, vol. 6, pp. 479–482, 1993. [Online]. Available: https://doi.org/10.1038/nclimate2912

[2] C. Baumhoer *et al.*, "Remote sensing of antarctic glacier and ice-shelf front dynamics-a review," *Remote Sens.*, vol. 10, 2018, Art. no. 1445. [Online]. Available: https://doi.org/10.1038/nclimate2912

[3] H.-G. Sohn and K. C. Jezek, "Mapping ice sheet margins from ERS-1 SAR and SPOT imagery," *Int. J. Remote Sens.*, vol. 20, no. 15/16, pp. 3201–3216, 1999.

[4] T. Klinger, M. Ziems, C. Heipke, H. W. Schenke, and N. Ott, "Antarctic coastline detection using snakes küstenliniendetektion in der antarktis mit hilfe von snakes," *Photogrammetrie—Fernerkundung—Geoinformation*, vol. 2011, pp. 421–434, Dec. 2011.

[5] A. Seale, P. Christoffersen, R. I. Mugford, and M. O'Leary, "Ocean forcing of the Greenland ice sheet: Calving fronts and patterns of retreat identified by automatic satellite monitoring of eastern outlet glaciers," *J. Geophysical Res., Earth Surf.*, vol. 116, no. F3, 2011.

[6] E. Zhang, L. Liu, and L. Huang, "Automatically delineating the calving front of jakobshavn Isbræfrom multitemporal terrasar-x images: A deep learning approach," *Cryosphere*, vol. 13, no. 6, pp. 1729–1741, 2019.

[7] Y. Mohajerani, M. Wood, I. Velicogna, and E. Rignot, "Detection of glacier calving margins with convolutional neural networks: A case study," *Remote Sens.*, vol. 11, no. 1, 2019.

[8] D. Cheng *et al.*, "Calving front machine (CALFIN): Glacial termini dataset and automated deep learning extraction method for Greenland, 1972-2019," *Cryosphere*, vol. 15, no. 3, pp. 1663–1675, 2021.

[9] P. Cunningham and S. J. Delany, "k-nearest neighbour classifiers: A Tutorial," *ACM Comput. Surveys*, vol. 54, no. 6, pp. 1–25, Jul. 2022.

[10] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019.

[11] X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.

[12] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," in *Proc. 2nd Int. Conf. Comput. Sci. Technol.*, 2021, pp. 124–133.

[13] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2021.3059968.

[14] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, 2015, pp. 234–241.

[16] A. Davari, C. Baller, T. Seehaus, M. Braun, A. Maier, and V. Christlein, "Pixel-wise distance regression for glacier calving front detection and segmentation," 2021, *arXiv:2103.05715*.

[17] A. Davari *et al.*, "On Mathews correlation coefficient and improved distance map loss for automatic glacier calving front segmentation in SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, Oct. 2021.

[18] C. A. Baumhoer, A. J. Dietz, C. Kneisel, and C. Kuenzer, "Automated extraction of Antarctic glacier and ice shelf fronts from sentinel-1 imagery using deep learning," *Remote Sens.*, vol. 11, no. 21, 2019, Art. no. 2529.

[19] Y. Zhou and Y. Wu, "Analyses on influence of training data set to neural network supervised learning performance," in *Advances in Computer Science, Intelligent System and Environment*, D. Jin and S. Lin, Eds. Berlin, Germany: Springer, 2011, pp. 19–25.

[20] P. Singh and R. Shree, "Analysis and effects of speckle noise in SAR images," in *Proc. 2nd Int. Conf. Adv. Comput., Commun., Autom.*, 2016, pp. 1–5.

[21] Z. Yu, W. Wang, C. Li, W. Liu, and J. Yang, "Speckle noise suppression in SAR images using a three-step algorithm," *Sensors*, vol. 18, no. 11, 2018, Art. no. 3643.

[22] M. Mansourpour, M. Rajabi, and J. Blais, "Effects and performance of speckle noise reduction filters on active radar and SAR images," in *Proc. ISPRS*, 2006, vol. 36, p. W41.

[23] W. Zhang, F. Liu, and L. Jiao, "SAR image despeckling via bilateral filtering," *Electron. Lett.*, vol. 45, pp. 781–783, Aug. 2009.

[24] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, Jul. 2019, Art. no. 60.

[25] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, "Understanding data augmentation for classification: When to warp," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl.*, 2016, pp. 1–6.

[26] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proc. 7th Int. Conf. Document Anal. Recognit.*, 2003, pp. 958–963.

[27] J. Johnson and T. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, Mar. 2019, Art. no. 27.

[28] H. Lee, M. Park, and J. Kim, "Plankton classification on imbalanced large scale database via convolutional neural networks with transfer learning," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 3713–3717.

[29] S. Pouyanfar *et al.*, "Dynamic sampling in convolutional neural networks for imbalanced data classification," in *Proc. IEEE Conf. Multimedia Inf. Process. Retrieval*, 2018, pp. 112–117.

[30] A. Davari, H. C. Özkan, A. Maier, and C. Riess, "Fast and efficient limited data hyperspectral remote sensing image classification via GMM-Based synthetic samples," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2107–2120, Jul. 2019.

[31] A. Davari, V. Christlein, S. Vesal, A. Maier, and C. Riess, "GMM supervectors for limited training data in hyperspectral remote sensing image classification," in *Proc. Int. Conf. Comput. Anal. Images Patterns*, 2017, pp. 296–306.

[32] F. N. Hatamian, N. Ravikumar, S. Vesal, F. P. Kemeth, M. Struck, and A. Maier, "The effect of data augmentation on classification of atrial fibrillation in short single-lead ECG signals using deep neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2020, pp. 1264–1268.

[33] S. Vesal, N. Ravikumar, and A. Maier, "A 2D dilated residual U-net for multi-organ segmentation in thoracic CT," 2019, *arXiv:1905.07710*.

[34] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[35] A. J. Cook, P. R. Holland, M. P. Meredith, T. Murray, A. Luckman, and D. G. Vaughan, "Ocean forcing of glacier retreat in the western antarctic peninsula," *Science*, vol. 353, no. 6296, pp. 283–286, 2016.

[36] A. Cook, D. Vaughan, A. Luckman, and T. Murray, "A new antarctic peninsula glacier basin inventory and observed area changes since the 1940 s," *Antarctic Sci.*, vol. 26, no. 6, pp. 614–624, 2014.

[37] P. Friedl, T. C. Seehaus, A. Wendt, M. H. Braun, and K. Höppner, "Recent dynamic changes on fleming glacier after the disintegration of Wordie Ice shelf, Antarctic Peninsula," *Cryosphere*, vol. 12, no. 4, pp. 1347–1365, 2018.

[38] T. C. Seehaus *et al.*, "Dynamic response of Sjögren glaciers, antarctic peninsula, to ice shelf breakup derived from multi-mission remote sensing time series," *Front. Earth Sci.*, vol. 4, 2016, Art. no. 66.

[39] T. Seehaus, S. Marinsek, V. Helm, P. Skvarca, and M. Braun, "Changes in ice dynamics, elevation and mass discharge of Dinsmoor-Bombardier-Edgeworth glacier system, Antarctic Peninsula," *Earth Planet. Sci. Lett.*, vol. 427, pp. 125–135, Sep. 2015.

[40] I. Joughin *et al.*, "Continued evolution of Jakobshavn Isbrae following its rapid speedup," *J. Geophysical Res. (Earth Surf.)*, vol. 113, Dec. 2008, Art. no. F0 4006.

[41] I. Joughin *et al.*, "Seasonal to decadal scale variations in the surface velocity of Jakobshavn Isbræ, Greenland: Observation and model-based analysis," *J. Geophys. Res.*, vol. 117, May 2012.

[42] R. Rosenau, E. Schwalbe, H.-G. Maas, M. Baessler, and R. Dietrich, "Grounding line migration and high-resolution calving dynamics of Jakobshavn Isbræ, West Greenland," *J. Geophysical Res., Earth Surf.*, vol. 118, pp. 382–395, Jun. 2013.

[43] A. Cook, T. Murray, A. Luckman, D. Vaughan, and N. Barrand, "A new 100-m digital elevation model of the antarctic peninsula derived from aster global DEM: Methods and accuracy assessment," *Earth Syst. Sci. Data Discuss.*, vol. 5, pp. 129–142, May 2012.

[44] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Med. Imag.*, vol. 15, Aug. 2015, Art. no. 29.

[45] H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein, "Visualizing the loss landscape of neural nets," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 6391–6401.

[46] N. Ismoilov and S.-B. Jang, "A comparison of regularization techniques in deep neural networks," *Symmetry*, vol. 10, Nov. 2018, Art. no. 648.

[47] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *Proc. Int. Conf. Artif. Neural Netw. Mach. Learn.*, 2018, pp. 270–279.
[48] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *Neural Inf. Process. Syst.*, vol. 25, pp. 1097–1105, Jan. 2012.
[49] J. Ma *et al.*, "Loss odyssey in medical image segmentation," *Med. Image Anal.*, vol. 71, Jul. 2021, Art. no. 102035.
[50] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis.*, 2016, pp. 565–571.
[51] K. Janocha and W. M. Czarnecki, "On loss functions for deep neural networks in classification," 2017, *arXiv:1702.05659*.
[52] M. Leipert, G. Vogeler, M. Seuret, A. Maier, and V. Christlein, "The notary in the haystack - Countering class imbalance in document processing with CNNs," in *Proc. Int. Workshop Document Anal. Syst.*, 2020, pp. 246–261.
[53] X. Li, S. Chen, X. Hu, and J. Yang, "Understanding the disharmony between dropout and batch normalization by variance shift," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2677–2685.
[54] Z. Zhou, R. S. Md Mahfuzur, T. Nima, and L. Jianming, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support*, Cham, Switzerland: Springer, 2018, pp. 3–11.
[55] S. A. Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad, and G. Hamarneh, "Deep semantic segmentation of natural and medical images: A review," *Artif. Intell. Rev.*, vol. 54, pp. 137–178, Jan. 2021.

Maniraman Periyasamy received the B.E. degree in mechatronics from the Aarupadai Veedu Institute of Technology, Chennai, India, in 2016, and the M.Sc. degree in computational engineering from Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, in 2021.

He is currently working as a Research Assistant with Fraunhofer-IIS, Nürnberg, Germany, specializing in the field of quantum reinforcement learning. His research interests include machine learning, deep learning, reinforcement learning, and quantum computing.

Amirabbas Davari received the B.Sc. degree in electrical engineering from the University of Tehran, Tehran, Iran, in 2011, the M.Sc. degree in electrical engineering from Özyeğin Uuniversity, Istanbul, Turkey, in 2013, and the Ph.D. degree in computer science from the Pattern Recognition Laboratory, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany, in 2022.

His research interests include image processing, pattern recognition, and deep learning with a focus on remote sensing, SAR, and hyperspecral image analysis.

Thorsten Seehaus received the Diploma in physics from the University Würzburg, Wrzburg, Germany, in 2011 and the Ph.D. degree in geography from the Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, in 2016.

He finished an apprenticeship as a Mechatronics Technician with Jopp GmbH, Bad Neustadt an der Saale, Germany in 2003. In 2012, he joined the working group of GIS and remote sensing with the Institute of Geography, Friedrich-Alexander-Universität Erlangen-Nürnberg, where he is currently a Research Assistant. He uses mainly multimission SAR imagery to assess glacier variables like mass balances and area changes. His research interests include developing and applying remote sensing techniques for monitoring glacier changes on various scales and in various regions worldwide.

Matthias Braun received the Diploma in hydrology (M.S.) and the Dr. rer.nat. (Ph.D.) degree (with honors) from the University of Freiburg, Breisgau, Germany, in 1997 and 2001, respectively.

From 1997 to 2001, he was with the University of Freiburg, where he worked on modeling surface energy balance of an Antarctic ice cap supported by synthetic aperture satellite image analysis. From 2001 to 2010, he was the Scientific Coordinator of the Interdisciplinary Center for Remote Sensing of Land Surfaces, Bonn University, Bonn, Germany. He was appointed as Associate Professor of Geophysics with the University of Alaska Fairbanks, Fairbanks, AK, USA, in 2010 and as Professor with Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, in 2011. He was leading various field campaigns in Antarctica, Greenland, Svalbard, Patagonia, High Mountain Asia, and the Alps. His research interests include the analysis of earth observation time series for glaciological variables.

Andreas Maier (Senior Member, IEEE) was born in Erlangen, Germany, on November 26, 1980. He received the graduation degree in computer science and the Ph.D. degree in medical signal processing in speech data from the Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, in 2005 and 2009, respectively.

From 2005 to 2009, he was with the Pattern Recognition Lab, Computer Science Department, University of Erlangen-Nuremberg. During this period, he developed the first online speech intelligibility assessment tool—PEAKS—that has been used to analyze over 4000 patient and control subjects so far. From 2009 to 2010, he started working on flat-panel C-arm CT as Postdoctoral Fellow with the Radiological Sciences Laboratory, Department of Radiology, Stanford University, Stanford, CA, USA. From 2011 to 2012, he joined Siemens Healthcare, Erlangen, Germany, as Innovation Project Manager and was responsible for reconstruction topics in the angiography and X-ray business unit. In 2012, he returned to the University of Erlangen-Nuremberg as the Head of the Medical Reconstruction Group, Pattern Recognition Lab. In 2015, he became a Professor and the Head of the Pattern Recognition Lab. His research interests include medical imaging, image and audio processing, digital humanities, and interpretable machine learning and the use of known operators.

Dr. Maier has been a Member of the Steering Committee of the European Time Machine Consortium since 2016. He was awarded an ERC Synergy Grant "4D Nanoscope" in 2018.

Vincent Christlein received the graduation degree in computer science 2012 and the Ph.D. (Dr.-Ing.) degree in 2018 from the Friedrich-Alexander-Universität of Erlangen-Nürnberg (FAU), Erlangen, Germany.

During his studies, he worked on automatic handwriting analysis with focus on writer identification and writer retrieval. Since 2018, he has been working as a Research Associate with the Pattern Recognition Lab, FAU, where he was promoted to Academic Councilor in 2020. He heads the Computer Vision Group, which covers a wide variance of topics, e.g., environmental projects such as glacier segmentation or solar cell crack recognition, but also computational humanities topics, such as document and art analysis.