NABLA-N for meltpond detection

Aqsa Sultana[a], Vijayan K. Asari*[a], Theus Aspiras[a], Ruixu Liu[a], Ivan Sudakow[b], Lee W. Cooper[c]
[a]University of Dayton, Vision Lab, Dayton, Ohio, USA 45469
[b]The Open University, Milton Keynes, England, MK7 6AA, United Kingdom
[c]University of Maryland, Center for Environmental Science, Maryland, USA 21613

ABSTRACT

With the increase in global temperatures due to anthropogenic climate change, sea ice in the Arctic has experienced rapid melting, resulting in increasing numbers of meltponds. As meltponds have a much lower albedo than sea ice or snow, more solar radiation will be absorbed by the water, further accelerating the melting rate of the sea ice. The dynamic nature of the meltponds exhibit complex shapes and boundaries, which makes manual analysis tedious and taxing. Several classical image processing approaches have been extensively used for the detection of meltpond regions in the Arctic area. We propose a CNN based multiclass segmentation model known as NABLA-N for automated detection and segmentation of meltponds. The architectural framework of NABLA-N consists of an encoding unit and multiple decoding units that decode from several latent spaces. The fusion of multiple feature spaces in the decoding units enables better representation of features due to the combination of low and high-level feature maps. The proposed model is evaluated on high-resolution aerial photographs of Arctic region obtained during the Healy-Oden Trans Arctic Expedition (HO-TRAX) in 2005 and NASA's Operation IceBridge DMS L1B Geolocated and Orthorectified data in 2016. These images are classified into three classes: meltpond, open water and sea ice. In this paper, NABLA-N demonstrates superior performance on segmentation of meltpond data compared to other state-of-the-art networks such as UNet and Recurrent Residual UNet (R2UNet).

Keywords: Segmentation, meltponds, arctic region, ensembled models, multiple latent spaces, HO-TRAX, Operation IceBridge, encoding unit, decoding unit

1. INTRODUCTION

The earth's average temperature has rapidly increased over the past 150 years in large part due to human-made greenhouse gas emissions. This increase has had an outsized effect on the Arctic, where higher temperatures have resulted in above-average rates of melting for sea ice. Such melting forms pools of water known as meltponds, which have a much lower albedo than snow or ice [1]. The decreased albedo causes greater absorption of solar radiation in areas covered by meltponds, leading to more melting of the sea ice. This results in a positive feedback loop accelerating the rate of melting of sea ice. Since meltponds play a significant role in increasing the loss of sea ice, they could be a useful metric in quantifying how Arctic Sea ice is responding to global warming [2].

Due to their localized nature and rapid development, an autonomous method for meltpond detection would prove extremely useful for environmental monitoring. Currently, tracking the formation and growth of meltponds requires manual annotation and evaluation. This makes data collection tedious, and calculating automated bounding boxes around meltponds would allow for more timely calculations of important metrics such as the melting rate of said meltponds. Additionally, general metrics of open water and sea ice allow for ratio calculations which will allow researchers to better track the overall activities of the Arctic region [1]. Figure 1 shows the meltpond in southwestern Greenland's glacial ice field. The image is natural-color and was acquired by the Advanced Land Imager on NASA's Earth Observing-1 (EO-1) satellite [3].

This study focuses exclusively on creating automated bounding boxes for the regions (snow/ice, open water and meltponds) in Arctic. We used state of the art neural networks to compare and create a benchmark for our network.

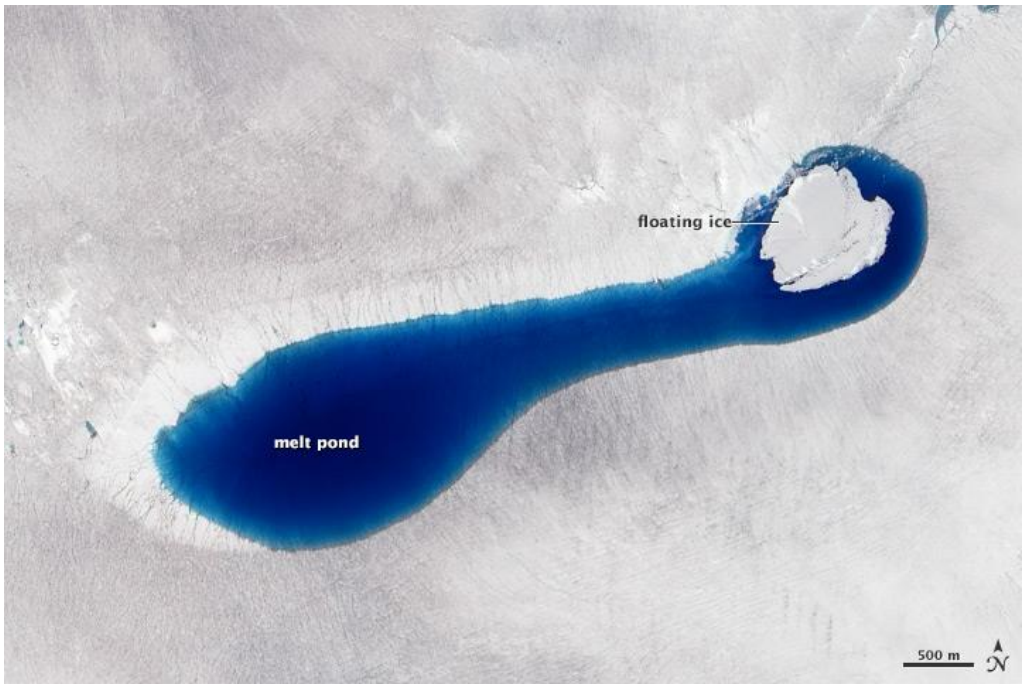


Figure 1. Image acquired from NASA's Earth Observing-1 satellite by the Advanced Land Imager

UNet

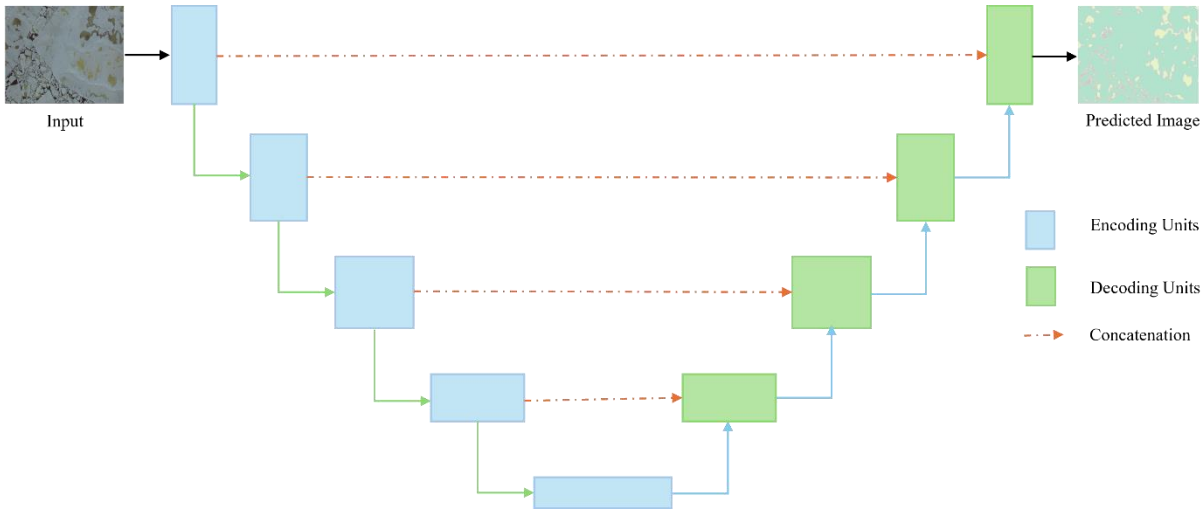


Figure 2. Architectural structure of UNet

1.1 UNet

Figure 2 displays the architectural structure of UNet [4]. UNet consists of four encoders and four decoders which are built in the shape of the letter 'U'. Each encoder block consists of 3x3 convolutions followed by a Rectified Linear Unit (ReLU) activation function. Each stepping block in the contracting path downsamples input data by doubling the feature channels while decreasing the spatial dimension in half by using a 2x2 max-pooling operation. For the smooth flow of information from encoder to decoder, they are connected by a bridge consisting of a 3x3 convolution followed by a ReLU activation function.

In the expansive path, the decoder uses 2x2 transpose convolution to up-sample the sizes of the images. Each decoder is concatenated with its corresponding encoder to retain the low-level information from earlier layers. Finally, two convolutions of 3x3 with ReLU activation function are used. The output layer of the decoder uses a 1x1 convolution layer with either sigmoid or softmax activation depending on the number of classes. This activation function provides the segmentation mask, which represents the pixel-level classification.

1.2 R2UNet

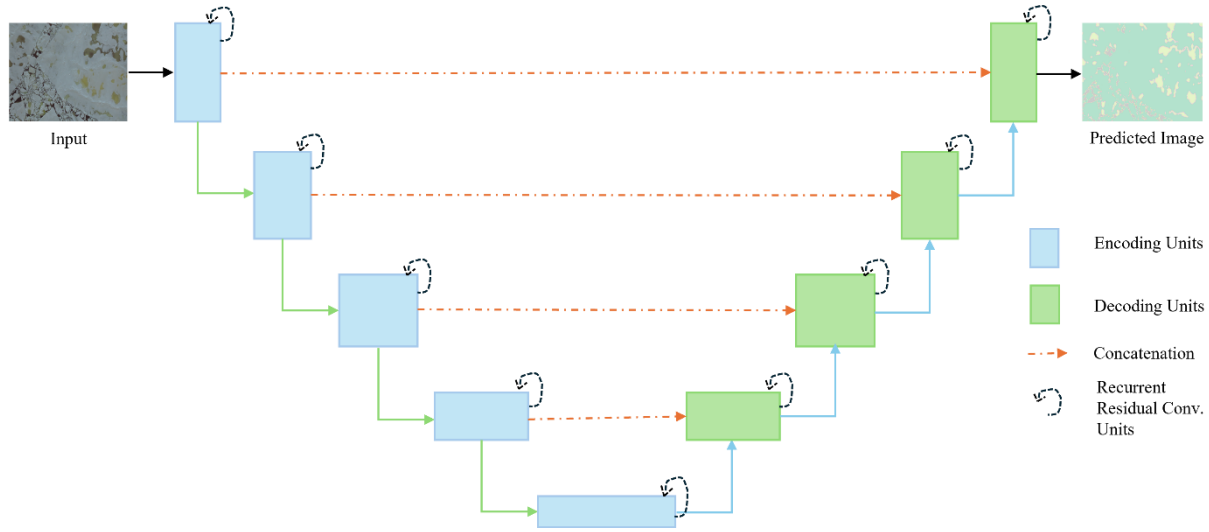


Figure 3. Architectural structure of R2UNet

Figure 3 shows the architectural details of R2UNet [5]. The structure of R2UNet is based on UNet consisting of encoder and decoder. Each encoder has a recurrent convolutional block following the residual connection before downsampling. Figure 4 (a) shows the recurrent residual convolutional unit. In each convolutional layer, subsequent recurrent convolutional layers are used. As the recurrent operation is based on the number of time steps, we used time step t=3. Thus, one convolutional layer has three recurrent layers as shown in Fig. 4 (b) [1,5,6,7]. The encoder is connected to the decoder via a bridge.

The decoder performs up-sampling of feature maps similar to UNet, except that each decoder block consists of a recurrent convolutional block following the residual connection before a transpose operation. The feature maps on the encoder are concatenated with feature maps of the decoder to retain relevant low-level information. The addition of a feedback loop in each convolutional layer yields accumulation of features required for bettering of semantic representation of extracted features. Meanwhile, the addition of a residual connection improves the learning efficacy and prevents exploding and vanishing gradients [1, 5].

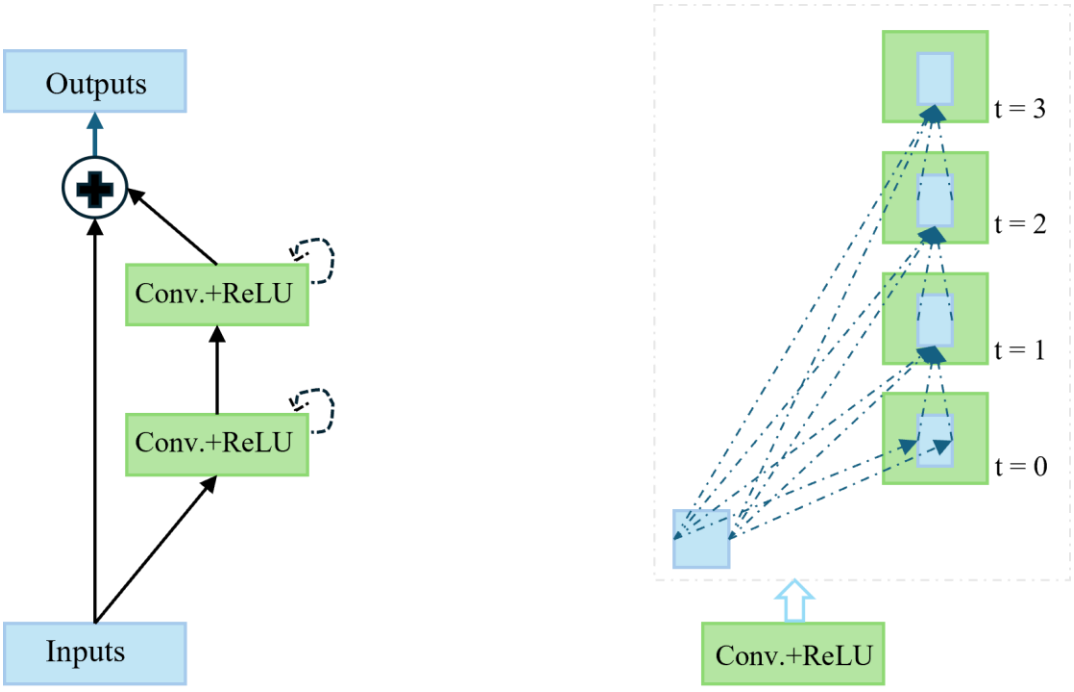


Figure 4. (a) Unfolded recurrent residual convolutional unit (b) Unfolded recurrent convolutional unit t=3

In this work, we introduce NABLA-N (∇^N -Net) for creating bounding boxes for the regions in the Arctic Area. ∇^N -N network embodies fusion and ensembling of multiple decoders and learns from many latent spaces for segmentation tasks. Our study shows that learning from multiple latent spaces enables better representation of features due to the combination of low and high-level feature maps. The model is evaluated on HO-TRAX and Operation IceBridge database acquired in year 2005 and 2016 respectively. Both databases consist of three annotated regions: open water, melt pond and snow/ice.

2. RELATED WORKS

The goal of this work is to employ NABLA-N, a segmentation model, and compare the qualitative and quantitative results against other state-of-the-art architectures. In the last few years, several semantic segmentation models have been proposed and have proven very successful in segmentation tasks in many different fields. In 2015, UNet was introduced for biomedical image segmentation [4]. The UNet model was efficiently applied on different modalities based on segmentation problems. In 2018, an improved version of UNet with residual and recurrent operations was introduced, named Recurrent Residual UNet [5]. In the same year, another architecture, LadderNet, which was a chain of multiple UNets, was introduced [9]. In 2019, the Fusion Net architecture, which was made up of multiple UNets in parallel, was proposed [10]. Additionally, in 2017, NABLA-Net was proposed. This architecture consisted of FCN based encoding and decoding units [11]. In 2019, NABLA-N network was introduced for the segmentation of skin cancer [8]. In this work, we applied NABLA-N network (∇^N -Net) for the evaluation and creation of bounding boxes of the regions in Arctic area. The network is evaluated on HO-TRAX and Operation IceBridge dataset [12]. Additionally, MeltpondNet, which was based on Swin Transformer UNet [13] and R2UNet [1] for detection of meltponds on Arctic Sea ice, was also evaluated on the HO-TRAX dataset.

3. MODEL ARCHITECTURE

The NABLA-N network derives its name from the symbol known as “NABLA,” which is an upside-down version of the Greek letter Delta, as the network’s shape resembles the symbol. The ‘N’ in the network is derived based on the number

of feature spaces used while employing NABLA. We used three latent spaces, hence the title ∇^3 -Net. Figure 5 shows the ∇^3 -Net architecture with three feature spaces. ∇^3 -Net is a combination of the UNet, R2UNet, LadderNet and FusionNet architectures, consisting of encoding and decoding unit. In the encoder, similar to UNet and its variants, the input image is fed through several forward convolutional techniques and is subsampled using max-pooling operations of 2x2. Here, the depth of the image is doubled by increasing the number of feature maps and size of the image is reduced half its spatial dimension.

The decoder has several convolutional transpose operations for upsampling the image. Here, the image size is increased, and feature maps are reduced. The encoded features from the inputs are decoded through bottleneck layer. According to representation strategy of features, the deeper layers with a greater number of feature maps represent high level features in representing feature to object space. As the bottleneck has high feature representation, the flow of information from encoders to decoders is prone to noise sensitivity and decoders are very crucial in producing accurate segmentation results. To combat this tendency, our ∇^3 -Net model consists of three decoders utilizing three feature spaces in the deeper layers of the encoder to produce enhanced and precise segmentation masks. The encoding unit encodes the input samples and multiple decoding units decode the encoded features from different latent spaces as shown in Fig. 5. The encoding unit is concatenated with its multiple corresponding decoding units to retain relevant low-level features, and feature fusion operations are applied between decoding units using addition. A 1x1 convolution is performed in the output layer after concatenation. As the ∇^3 -Net model is deeper, we used recurrent residual convolutions for convolutional operations in both encoding and decoding units to prevent exploding and vanishing gradients. These operations ensure efficient learning and better feature accumulation, which is required for segmentation and detection tasks. The robustness of our model is evaluated on HO-TRAX and Operation IceBridge dataset and is discussed in further sessions [8].

We used:
3→16(2)→32(2)→64(2)→128(2)→256(2)→128 (2)→64(2)→32(2)→16(2)→3, where numbers within the parentheses indicate filter size of the receptive field, and numbers outside the parentheses indicate filter size of the filter maps. The number of total, trainable, and non-trainable parameters of the model are 3,037,923, 3,036,931 and 992 respectively.

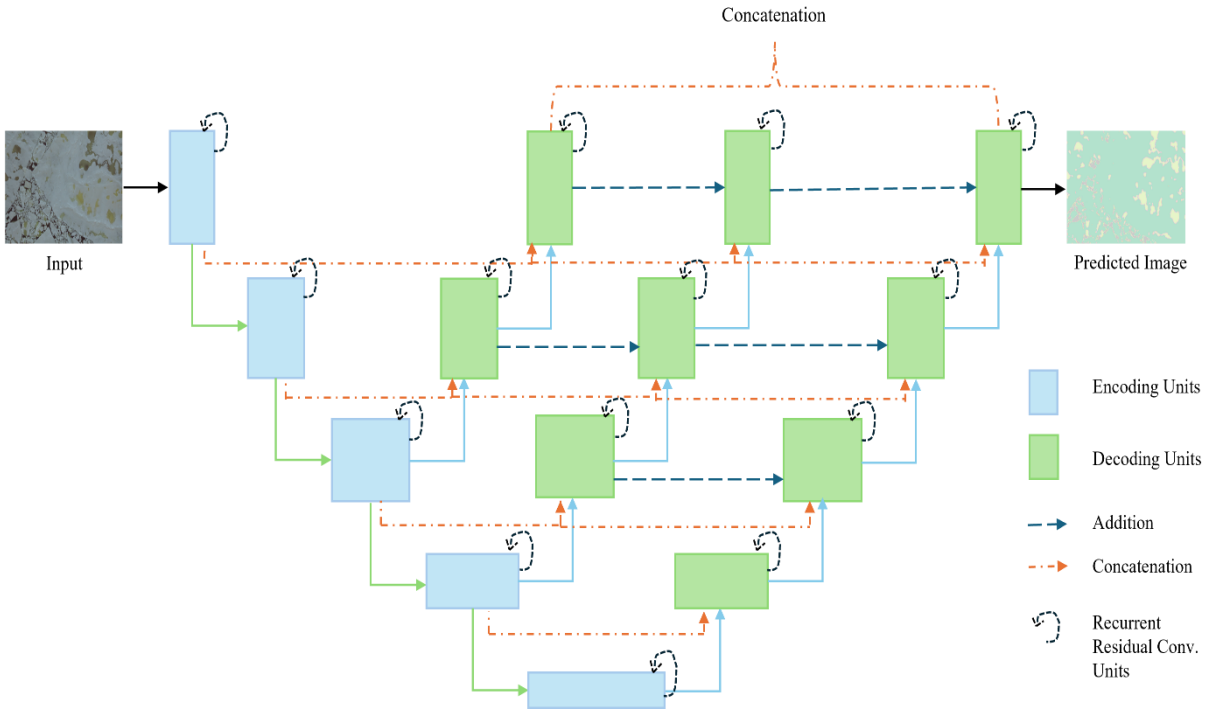


Figure 5. The ∇^3 -Net architecture

4. EXPERIMENTAL SETUP AND RESULTS

The experiments were conducted on two NVIDIA GEFORCE RTX Titan GPUs with 24 GB of VRAM each (for a total of 48 GB VRAM) and 128 GB RAM. We used the TensorFlow framework in Python with a Keras backend. The experimental setup was consistent for all the experiments conducted for this work.

4.1 Dataset

4.1.1 HO-TRAX

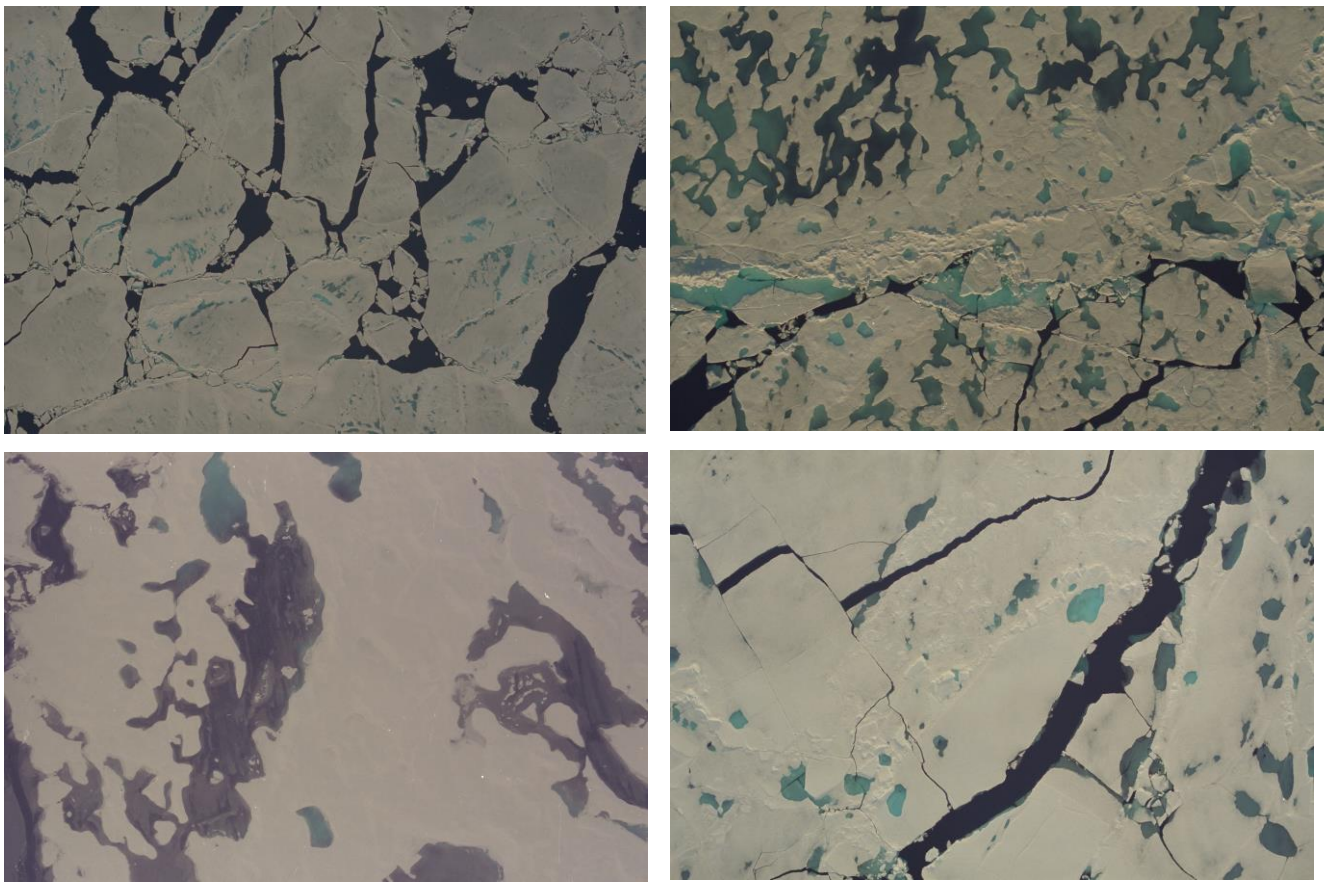


Figure 6. Healy Oden Trans Arctic Expedition (HO-TRAX) image dataset

The aerial imagery was captured from a helicopter with a Nikon D70 digital camera between 5 August and 30 September 2005 during the Healy Oden Trans Arctic Expedition (HO-TRAX). Flights were conducted at relatively low altitudes between 150–700 m to avoid low clouds. The captured images had an average size of 3042×2048 pixels. The highly detailed images were visually analyzed, and the zones were divided into three classes: open water, meltponds and sea ice [1,12]. The sample images are shown in Fig. 6.

4.1.1 Operation IceBridge

The IceBridge dataset consists of Level 1B imagery acquired from the Digital Mapping System (DMS) over Greenland and Alaskan waters in 2016, as part of NASA's Operation IceBridge. The aircraft was flown over the Chukchi Sea in July when the sea ice would have been melting. The collected imagery had a 10 cm ground sample distance with varying temporal resolution [14]. The sample images are shown in Fig. 7.

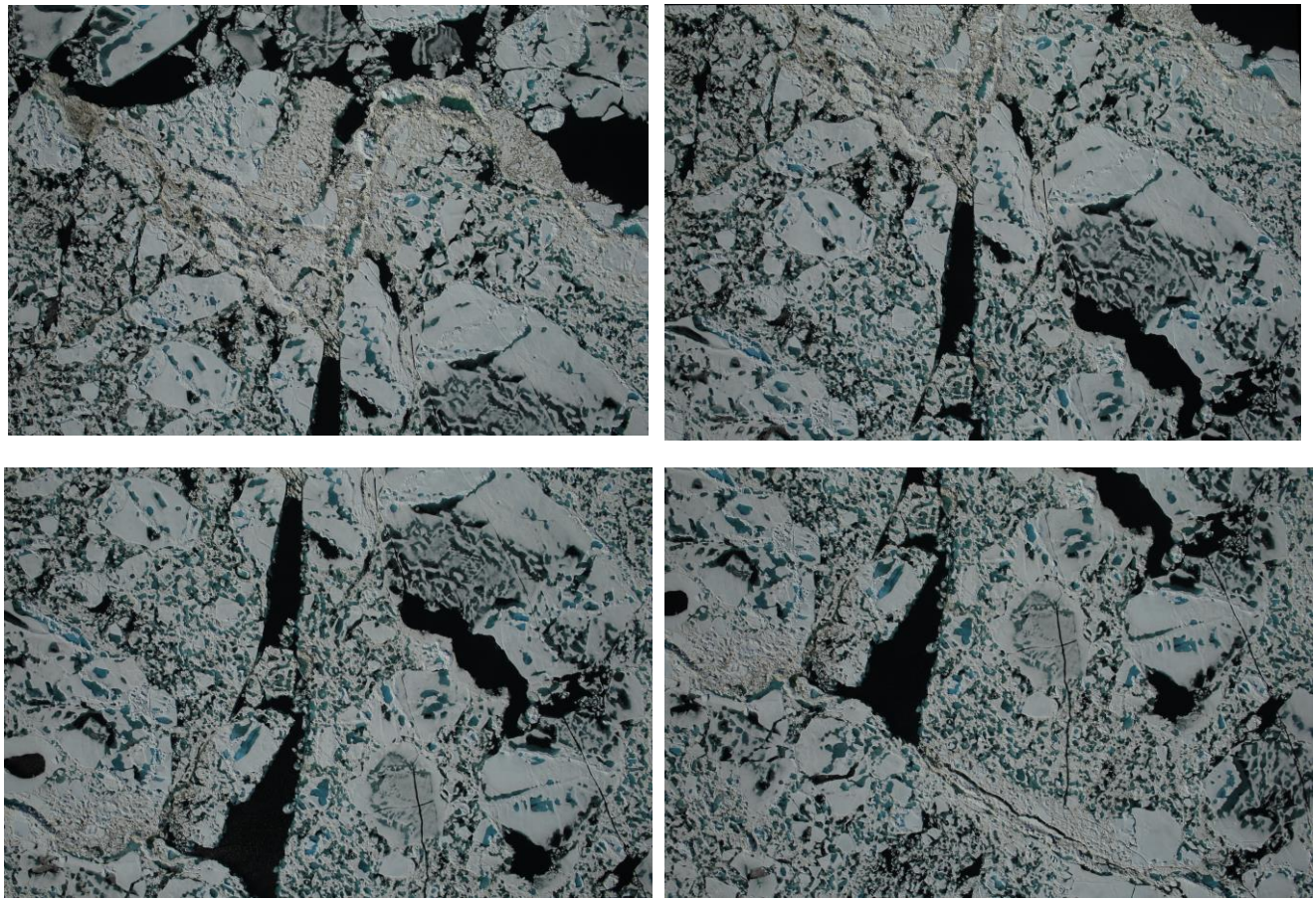


Figure 7. NASA's Operation IceBridge image dataset

4.2 Quantitative analysis approaches

For quantitative analysis of the experimental results of our model, we employed F1-score, accuracy, precision, recall, Jaccard similarity and mean IoU. We utilized the following evaluation metrics to evaluate the performance of our model against UNet and R2UNet.

F1 – Score:

$$F1 - Score = \frac{2TP}{2TP + FP + FN} = \frac{2 \times (Precision \times Recall)}{Precision + Recall}$$

Accuracy:

$$AC = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision:

$$PR = \frac{TP}{TP + FP}$$

Recall:

$$R = \frac{TP}{TP + FN}$$

Jaccard Similarity:

$$JS = \frac{|GT \cap SR|}{|GT \cup SR|}$$

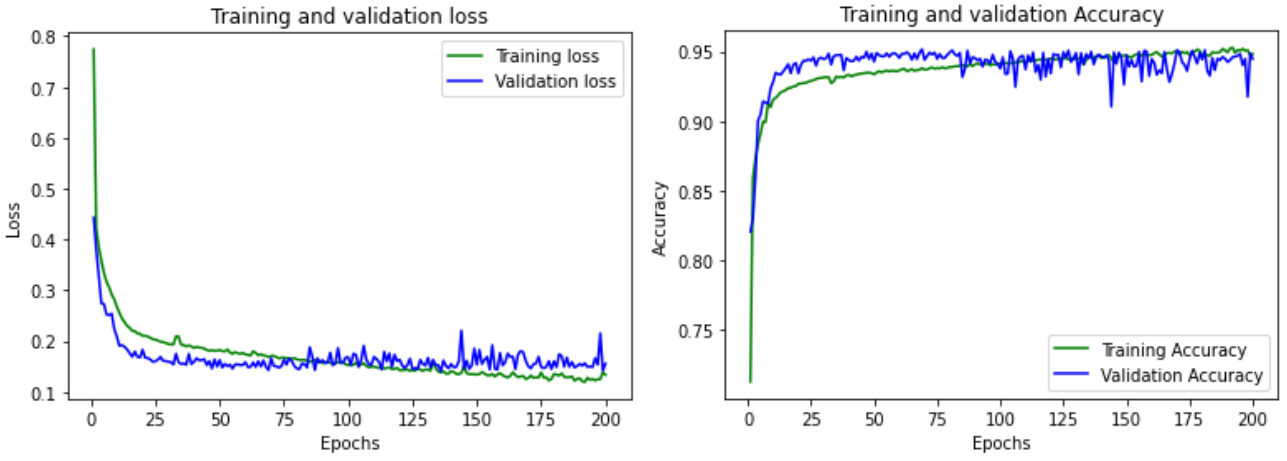
Mean IoU:

$$MeanIoU = \frac{TP}{TP + FP + FN}$$

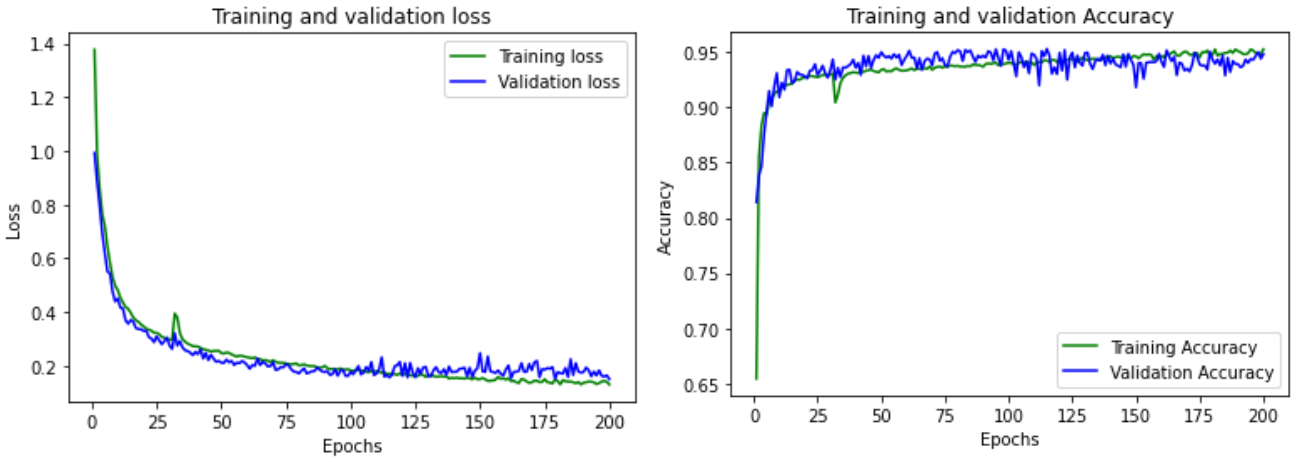
Here, GT is ground truth, SR is segmentation result, TP is True Positive, TN is True Negative, FP is False Positive and FN is False Negative.

4.3 Training method

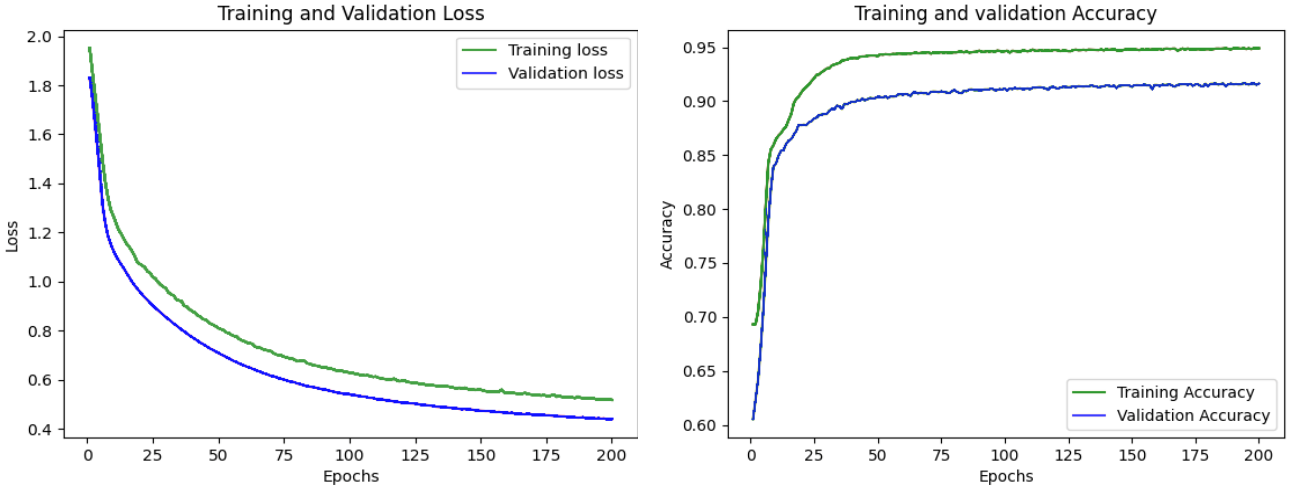
All the three models viz. UNet, R2UNet and ∇^N -Net are trained for 200 epochs. For the optimizer, we used ADAMW with a learning rate and weight decay of 1×10^{-4}. As meltpond detection is a multi class segmentation problem, we converted our labels to a binary class matrix to return either 1 or 0 from a class vector. We used categorical cross entropy for the loss function. For HO-TRAX, we split the data set into 60, 25, and 15 images for training, validation, and testing, respectively [1]. Figures 8 (a), (b) (c) show the training and validation loss, training accuracy, and validation accuracy of UNet, R2UNet and ∇^3 -Net. For the IceBridge dataset, we cropped the images to a size of 640 x 640 pixels since the original images are very large. After cropping, we had 210, 70, and 70 images for training, validation, and testing, respectively. Figures 9(a), (b) (c) show the training and validation loss, training, and validation accuracy of UNet, R2UNet and ∇^N -Net for the Operation IceBridge dataset.



(a)

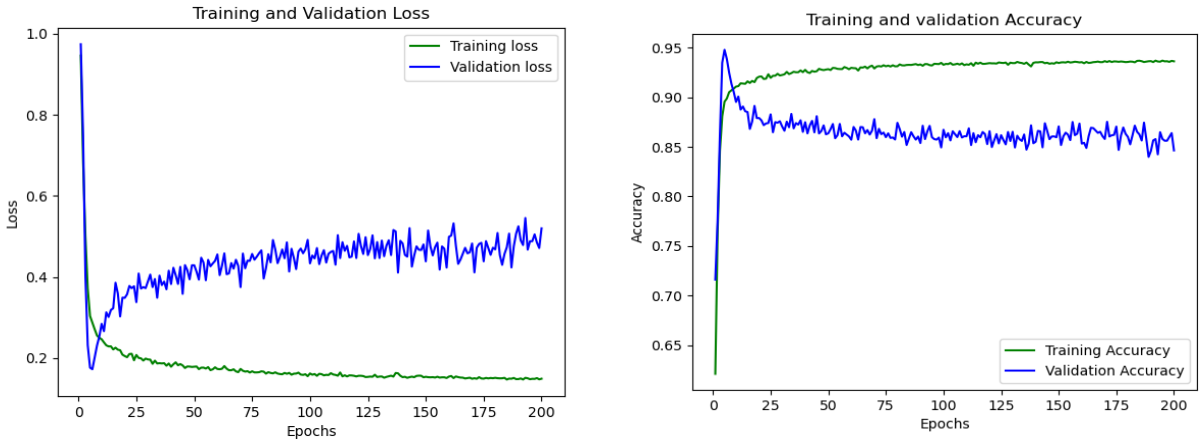


(b)

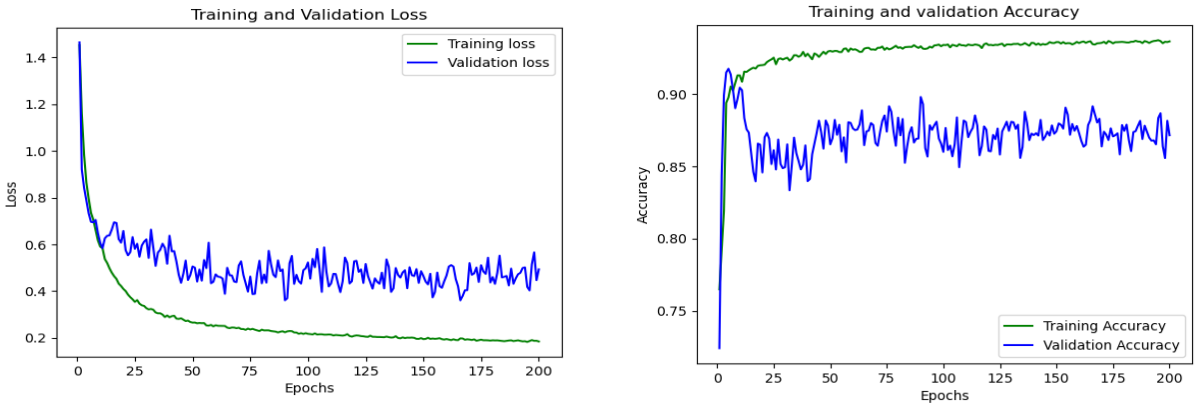


(c)

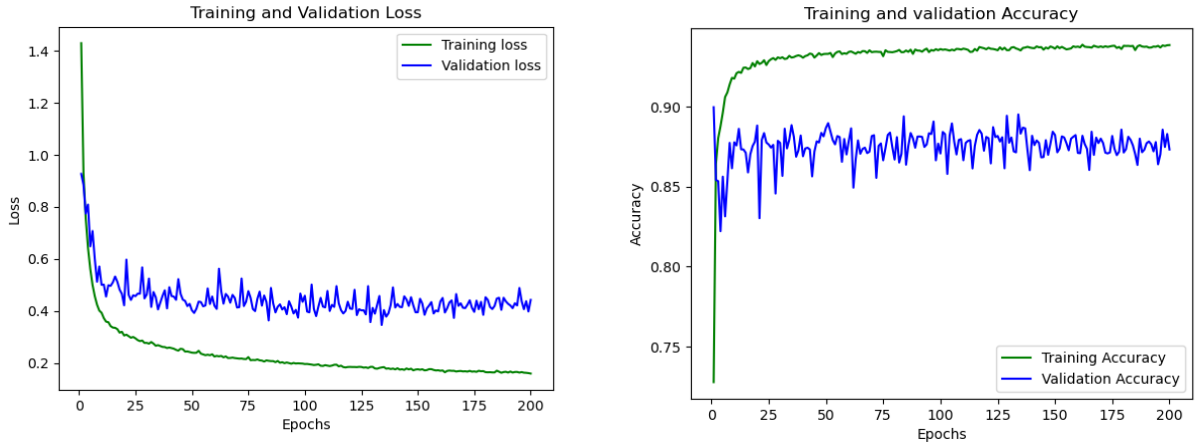
Figure 8. HO-TRAX training and validation loss and accuracy (a) UNet (b) R2UNet (c) ∇^N -Net



(a)



(b)



(c)

Figure 9. IceBridge training and validation loss and accuracy (a) UNet (b) R2UNet (c) ∇^{N} -Net

4.4 Results

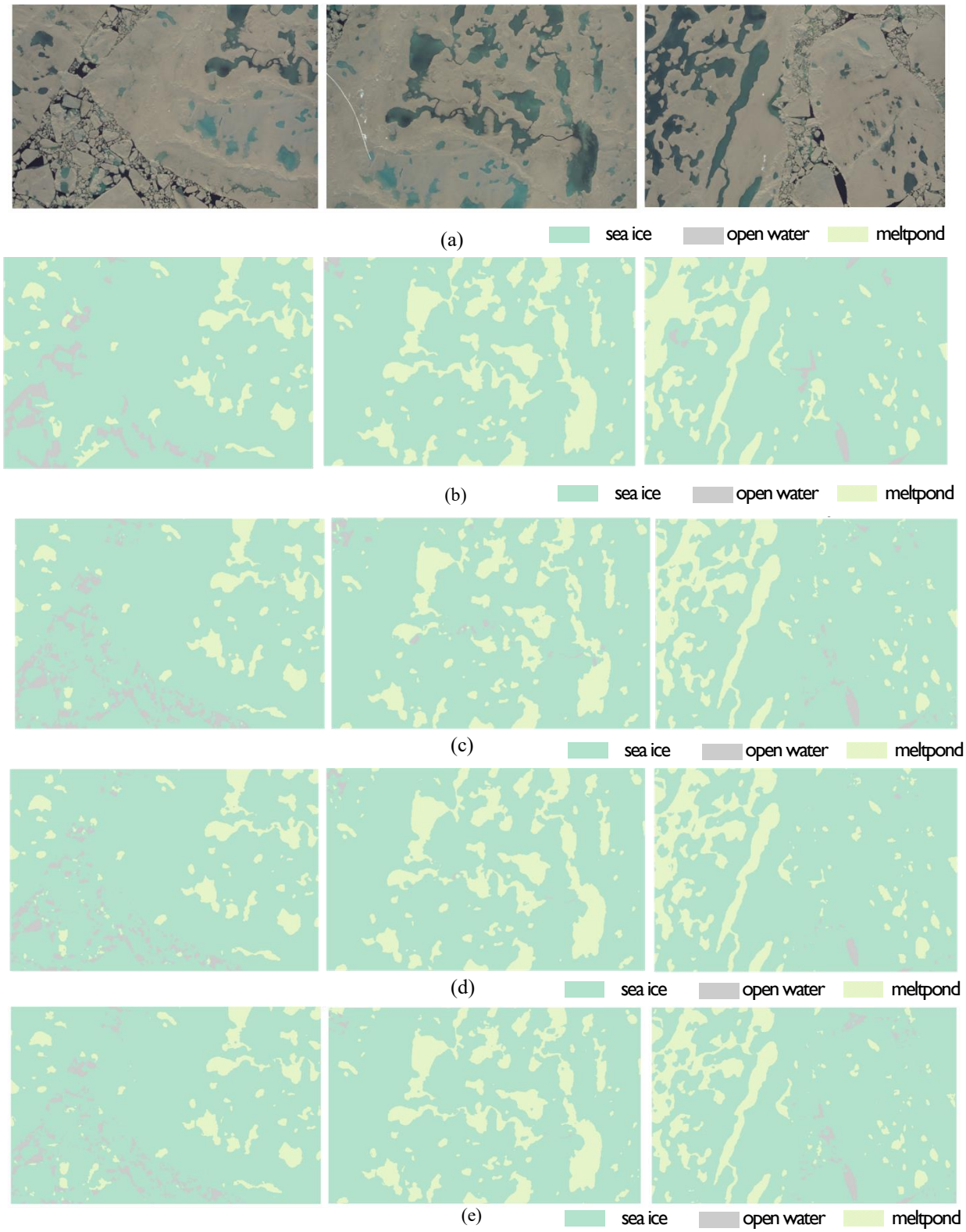


Figure 10. HO-TRAX database: (a) Original image (b) Ground truth (c) UNet (d) R2UNet (e) ∇^3 -Net

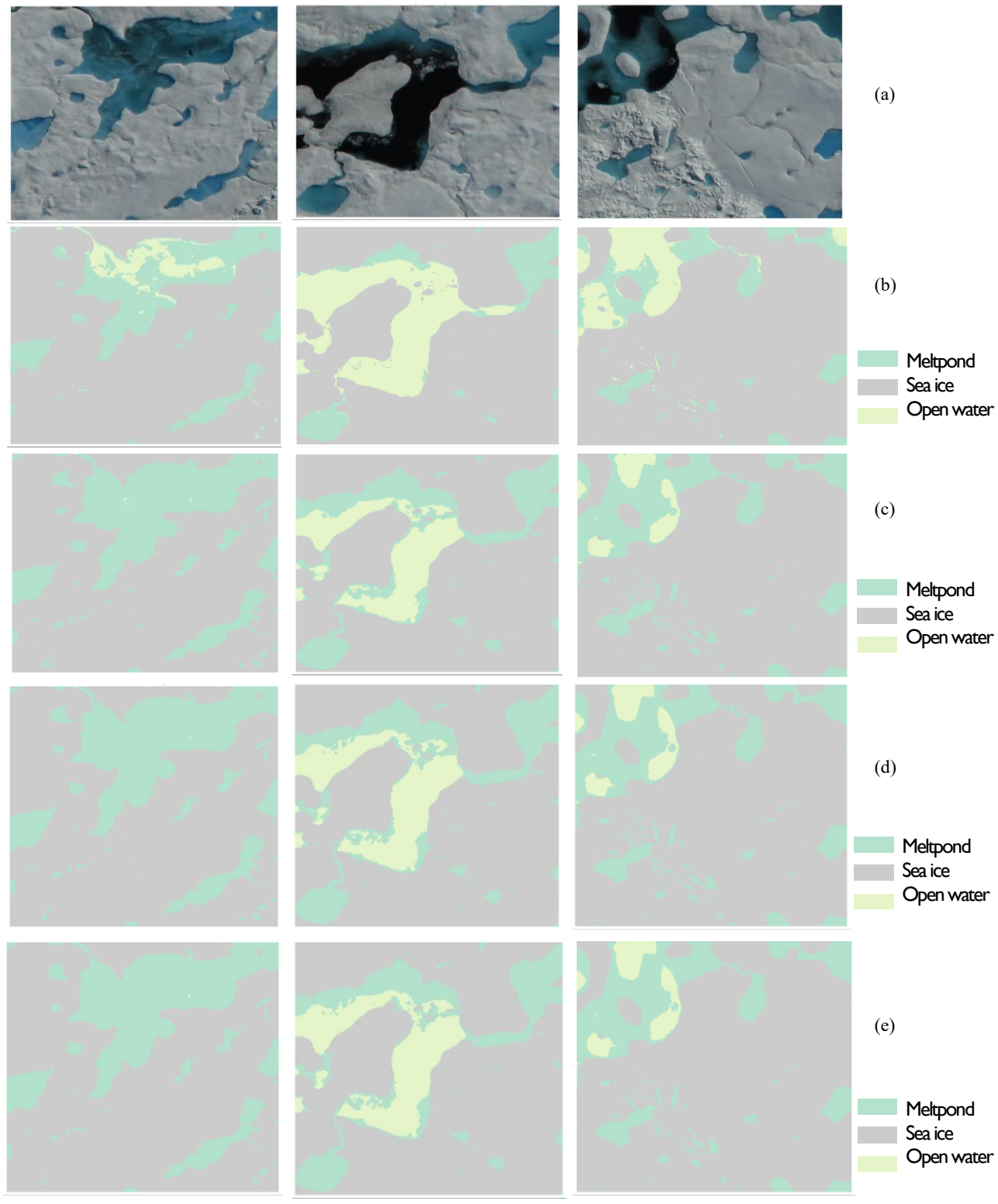


Figure 11. Operation IceBridge database: (a) Original image (b) Ground truth (c) UNet (d) R2UNet (e) ∇^3 -Net

All three trained models were evaluated on the same test images to compare their performance and robustness for segmentation. Under the same experimental and architectural settings for HO-TRAX data, UNet's accuracy was the lowest at 94.65%. The second lowest, R2UNet, showed an accuracy of 94.96%. Finally, ∇^3 -Net had the highest accuracy at 95.44%. The original images, ground truths and predicted images are shown in Fig. 10. Table 1 displays the performance of all the models and their comparison with different parameters such as F1-Score, Accuracy, Precision, Recall, Jaccard similarity and Mean IoU.

For Operation IceBridge data, ∇^N -Net (N=3) had the highest accuracy at 88.01%, The second highest, R2UNet, showed an accuracy of 86.39% and UNet had the lowest accuracy at 85.94%. The original images, ground truths and predicted images are shown in Fig. 11. Table 2 displays the performance of all the models and their comparison with different parameters such as F1-Score, Accuracy, Precision, Recall, Jaccard similarity and Mean IoU. Due to their ability to distinguish positive values from negative values, the classes that were incorrectly labeled by human error were generalized and predicted correctly by all of the models [1].

Table 1. Experimental results of UNet, R2UNet and ∇^3 -Net for the segmentation of three regions in HO-TRAX database

	F1-Score	**Accuracy**	**Precision**	**Recall**	**Jaccard Similarity**	**Mean IoU**
UNet	0.8859	0.9465	0.8847	0.8883	0.8016	0.7629
R2UNet	0.8933	0.9496	0.8992	0.8917	0.8129	0.7754
∇^N -Net (N=3)	**0.8997**	**0.9544**	**0.9058**	**0.8941**	**0.8227**	**0.7871**

Table 2. Experimental results of UNet, R2UNet and ∇^3 -Net for the segmentation of three regions in Operation IceBridge database

	F1-Score	**Accuracy**	**Precision**	**Recall**	**Jaccard Similarity**	**Mean IoU**
UNet	0.7497	0.8594	0.8363	0.7860	0.6234	0.6204
R2UNet	0.7767	0.8639	0.8390	0.8122	0.6522	0.6493
∇^N -Net (N=3)	**0.7905**	**0.8814**	**0.8504**	**0.8198**	**0.6714**	**0.6684**

5. CONCLUSION

In this work, we proposed and implemented a new architecture, ∇^N -Net, for pixel-level multiclass segmentation. Our network derives the name from its architectural structure "NABLA" which is an inverted Greek Delta. Our feature extraction is based on ∇^3 -Net (N=3), which is composed of three latent spaces. It is a CNN based framework consisting of encoding unit and several decoding units with multiple latent space enabling enhanced performance and better feature representation. As there are multiple decoding units with multiple latent spaces, the flow of high-level information is more efficient. Our framework engenders UNet, R2UNet, LadderNet and FusionNet. The model is evaluated on Operation IceBridge and HO-TRAX databases for segmentation. The quantitative and qualitative results demonstrate enhanced and robust performance when compared against UNet and R2UNet architectures. ∇^3 -Net showed superior accuracy of 95.44% on HO-TRAX dataset and 88.14% on Operation IceBridge dataset. Our further investigation will include relabeling the Operation IceBridge dataset using the results of ∇^3-Net and retrain the model for superior results and performance. We would also like to make internal tweaks in the model for better feature accumulation and representation.

ACKNOWLEDGEMENTS

This work was supported by the Division of Physics at the National Science Foundation (NSF), Grant No. PHY 2102906.

REFERENCES

[1] Aqsa Sultana, Vijayan K. Asari, Ivan Sudakow, Theus Aspiras, Ruixu Liu, and Denis Demchev "R2UNet for meltpond detection", Proc. SPIE 12527, Pattern Recognition and Tracking XXXIV, 125270R (13 June 2023); http://doi.org/10.1117/12.2663982

[2] Julienne Stroeve and Dirk Notz 2018 Environ. Res. Lett. 13 103001, doi: 10.1088/1748-9326/aade56

[3] NASA Earth Observatory. (n.d.). *Greenland melt ponds*. https://www.earthobservatory.nasa.gov/images/80677/greenland-melt-ponds

[4] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015

[5] Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M. Taha, and Vijayan K. Asari, "Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation," arXiv.org, Computer Vision and Pattern Recognition, arXiv:1802.06955, pp. 1-12, May 2018. (arXiv)

[6] Aqsa Sultana, Vijayan K. Asari and Theus Aspiras "Residues in succession recurrent U-Net for segmentation of retinal blood vessels", Proc. SPIE 12527, Pattern Recognition and Tracking XXXIV, 1252706 (13 June 2023); https://doi.org/10.1117/12.2664876

[7] Sultana, Aqsa. *Residues in Succession U-Net for Fast and Efficient Segmentation.* 2022. University of Dayton, Master's thesis. *OhioLINK Electronic Theses and Dissertations Center*, http://rave.ohiolink.edu/etdc/view?acc_num=dayton1659016279233472.

[8] Alom, M. Z., Aspiras, T., Taha, T. M., & Asari, V. K. (2019). Skin Cancer Segmentation and Classification with NABLA-N and Inception Recurrent Residual Convolutional Networks. *ArXiv*. /abs/1904.11126

[9] Zhuang, Juntang. "LadderNet: Multi-path networks based on U-Net for medical image segmentation." ArXiv abs/1810.07810 (2018): n. pag.

[10] Quan, T. M., Hildebrand, D. G. C., & Jeong, W. (2021). FusionNet: a deep fully residual convolutional neural network for image segmentation in connectomics. *Frontiers in Computer Science*, *3*. https://doi.org/10.3389/fcomp.2021.613981

[11] McKinley, R. *et al.* (2016). Nabla-net: A Deep Dag-Like Convolutional Architecture for Biomedical Image Segmentation. In: Crimi, A., Menze, B., Maier, O., Reyes, M., Winzeck, S., Handels, H. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2016. Lecture Notes in Computer Science(), vol 10154. Springer, Cham. https://doi.org/10.1007/978-3-319-55524-9_12

[12] Ivan Sudakow, Vijayan Asari, Ruixu Liu, & Denis Demchev. (2022). Melt pond from aerial photographs of the Healy–Oden Trans Arctic Expedition (HOTRAX) (1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.6602409

[13] I. Sudakow, V. K. Asari, R. Liu and D. Demchev, "MeltPondNet: A Swin Transformer U-Net for Detection of Melt Ponds on Arctic Sea Ice," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 8776-8784, 2022, doi: 10.1109/JSTARS.2022.3213192.

[14] Dominguez, R. (2010). IceBridge DMS L1B Geolocated and Orthorectified Images, Version 1 [Data Set]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/OZ6VNOPMPRJ0. Date Accessed 03-15-2024.