Automated Recognition of Permafrost Disturbances using High-spatial Resolution Satellite Imagery and Deep Learning Models

Mahendra Rajitha Udawalpola[1], Chandi Witharana[1], Amit Hasan[1], Anna Liljedahl[2], Melissa Ward Jones[3], Benjamin Jones[3]

[1] Dept. of Natural Resources and the Environment, University of Connecticut, Storrs, Connecticut, USA - (mahendra.udawalpola, chandi.witharana, amit.hasan)@uconn.edu

[2] Woodwell Climate Research Center, Falmouth, Massachusetts, USA - aliljedahl@whrc.org

[3]University of Alaska Fairbanks, AK, USA (mkwardjones, bmjones3)@alaska.edu

KEYWORDS: Arctic, Permafrost, Retrogressive Thaw Slumps, Deep learning, Satellite imagery, Mapping

ABSTRACT:
The accelerated warming conditions of the high Arctic have intensified the extensive thawing of permafrost. Retrogressive thaw slumps (RTSs) are considered as the most active landforms in the Arctic permafrost. An increase in RTSs has been observed in the Arctic in recent decades. Continuous monitoring of RTSs is important to understand climate change-driven disturbances in the region. Manual detection of these landforms is extremely difficult as they occur over exceptionally large areas. Only very few studies have explored the utility of very high spatial resolution (VHSR) commercial satellite imagery in the automated mapping of RTSs. We have developed deep learning (DL) convolution neural net (CNN) based workflow to automatically detect RTSs from VHRS satellite imagery. This study systematically compared the performance of different DLCNN model architectures and varying backbones. Our candidate CNN models include: DeepLabV3+, UNet, UNet++, Multi-scale Attention Net (MA-Net), and Pyramid Attention Network (PAN) with ResNet50, ResNet101 and ResNet152 backbones. The RTS modeling experiment was conducted on Banks Island and Ellesmere Island in Canada. The UNet++ model demonstrated the highest accuracy (F1 score of 87%) with the ResNet50 backbone at the expense of training and inferencing time. PAN, DeepLabV3, MaNet, and UNet, models reported mediocre F1 scores of 72%, 75%, 80%, and 81% respectively. Our findings unravel the performances of different DLCNNs in imagery-enabled RTS mapping and provide useful insights on operationalizing the mapping application across the Arctic.

1. INTRODUCTION

The Arctic is going through rapid changes in recent years. The temperatures in the region are rising at two to fourfold the global average (Screen, 2010). Due to the warming Arctic, the occurrence of permafrost disturbances, such as retrogressive thaw slumps (RTSs) has increased (Lants 2008). It is important to perform continuous monitoring of these disturbances to evaluate the impact on the Arctic environment. However, monitoring these disturbances is difficult in the Arctic compared to other parts of the world due to extreme weather, remoteness, and logistical challenges.

RTSs are thermokarst features created by the rapid thaw of ice-rich permafrost on slopes of permafrost. An active thaw slump consists of an exposed headwall that defines the upslope boundary of the RTS. Below the headwall, there is a scar zone consisting of muddy exposed soil. The materials in the scar zone can move downslope by creating a tongue-like shape at the other end of the RTS (Figure 1).

Figure 1. Retrogressive thaw slump headwall, scar zone, and debris tongue.

RTSs impact infrastructure, and aquatic and terrestrial ecosystems (Kokelj et al. 2013). Sediment and solutes released by RTS alter the properties of soils and surface waters. A mass movement of sediments and runoff can change the turbidity of adjacent rivers, lakes, and coastal environments. (Segal, 2015)

There are many attempts have been made to map RTSs in the Arctic region. Most of the mapping has been done using remote sensing images with manual techniques. There are only a few attempts have been made to automatically map RTSs using remote sensing images. Huang et al 2020 used Planet CubeSate images of 3m resolution to map RTSs in Tibetian Platue with DeepLabV3+. Recently Nitze et al. 2022 utilized PanetScope satellite imagery of 3.15m resolution to map RTS using UNet and UNet++. Witharana et. al 2022 employ high-resolution satellite images of 0.5m resolution to detect RTS using UNet. In that, they analyze the effect of different image tile sizes and spatial resolutions on the deep learning model prediction performances.

The morphometric features of RTSs (headwall, scar zone, and debris tongue) are well suited to be exploited with machine/deep learning algorithms. We use Deep Learning Convolutional Neural Networks (DL-CNN) to automatically detect RTSs. The main objective of this study is to investigate how different DL-CNN networks perform on RTSs detection using very high-resolution satellite imagery. Based on five candidate DL-CNN architectures, we systematically compared their training and detection performances.

2. METHODS

2.1 Mapping of RTS using satellite images

We used a transfer learning strategy to train the candidate DL-CNNs. In transfer learning, we have two stages. In the first stage, we use backbone CNN and in the second stage, we use classifier network. Figure 2 shows a schematic diagram for this approach. The backbone CNN is used to extract features from the images. The backbones of the networks have been pre-trained on ImageNet

datasets. Therefore, we can use a small number of samples to train the CNN. The extracted features are used to segment the RTSs in satellite images. We use different CNN networks for the segmentation of RTSs.

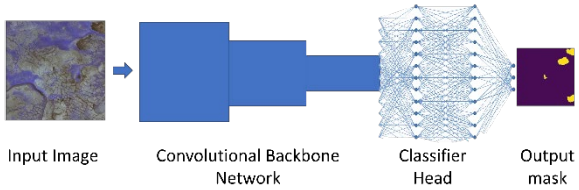


Figure 2. Simplified schematic diagram of transfer learning in convolutional neural networks. (Imagery © 2016 DigitalGlobe, Inc).

We tasked three convolutional backbone networks, 1) ResNet50, 2) ResNet101, and 3) ResNet152 (He, 2016) in this study. We used the pre-trained weights on the ImageNet dataset and froze weight values while training on our custom RTS dataset. Our comparative analysis entailed five semantic segmentation algorithms: UNet (Ronneberger, 2015), Pyramid Attention Network (PAN) (Li, 2018), Multi-scale Attention Net (MANet) (Fan, 2020), and UNet++ (Zhou, 2018). Table 1 shows the number of total parameters and the number of trainable parameters in each network.

Model	Backbone	Number of parameters (millions)	Number of trainable parameters (millions)
UNet	Resnet50	32M	9M
	Resnet101	51M	9M
	Resnet152	67M	9M
PAN	Resnet50	24M	1M
	Resnet101	43M	1M
	Resnet152	58M	1M
MANet	Resnet50	147M	123M
	Resnet101	166M	123M
	Resnet152	182M	123M
DeeplabV3	Resnet50	26M	3M
	Resnet101	45M	3M
	Resnet152	61M	3M
UNet++	Resnet50	48M	25M
	Resnet101	67M	25M
	Resnet152	83M	25M

Table 1. Comparison of the size variation of candidate DL-CNN models

2.2 Model Training

The RTS modeling was conducted based on the high res satellite imagery from Banks Island and Ellesmere Island in north Arctic Canada (Figure 3). We selected 12 WorldView-2 satellite images from Banks Island and 14 WorldView-2 satellite images from Ellesmere Island to generate hand-annotated RTS training data. Image scenes were acquired during July - Aug at 0.5m spatial resolution with 4 spectral channels (red, green blue, and near infra-red). Pansharpened and orthorectified imagery were provided by the Polar Geospatial Center, University of Minnesota.

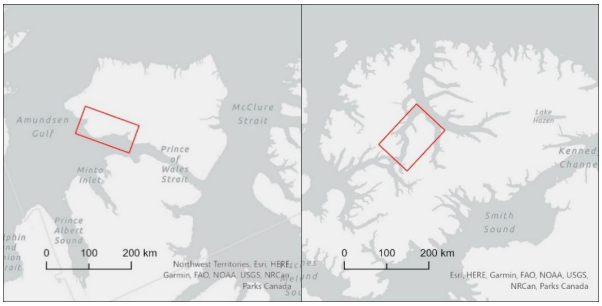


Figure 3. Selected study areas from Banks Island (left) and Ellesmere Island (right) in Canada.

For the model training, 475 image tiles (2048 x 2048 pixels or ~ 1km x 1km on the ground) were selected from each of the study sites shown in Figure 3. The dataset was split into 80%, 10%, and 10% for training, validations, and testing, respectively.

We utilized Adam optimization algorithms with a learning rate of 10^{-4} for the first 25 epochs and 10^{-5} for the rest of the epochs. We used dice loss for calculating training and value loss while training. All models were trained across 100 epochs. We employed 3 augmentations (horizontal flip, vertical flip, and random 90-degree rotation) to the datasets with 50% probability in each epoch.

Figures 4-8 show the training F1 scores for different CNN architectures coupled with three backbone networks ResNet50(blue), ResNet101(orange), and ResNet152(green). Figure 4 shows the F1 scores for MANet. All backbone networks achieved 97% accuracy at the end of epoch 50. Figure 5 shows the F1 scores for the DeepLanV3 network. Here all three backbones reported 96% accuracy at the end of the training. Training accuracy for the UNet model is shown in Figure 6. At the end of the training, all three backbones achieved 97% accuracy. Figure 7 shows the training F1 scores for the PAN network. All three backbone networks scored 96% accuracy. As seen in Figure 8, UNet++ with Resnet50 showed elevated F1 scores (at epoch 50 it's around 98%) compared to the other two backbones.

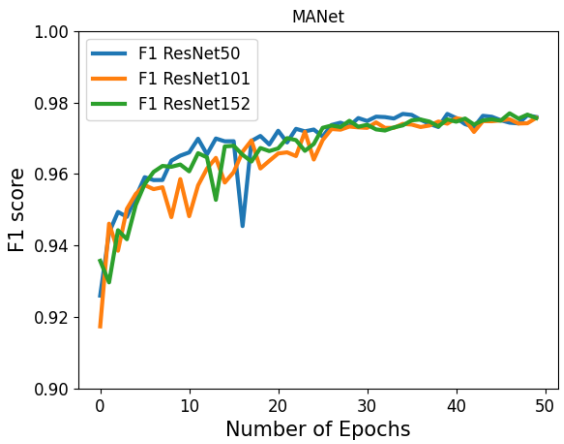


Figure 4. F1 score for training with ResNet50(blue), ResNet101(orange), and ResNet152(green) for MANet network.

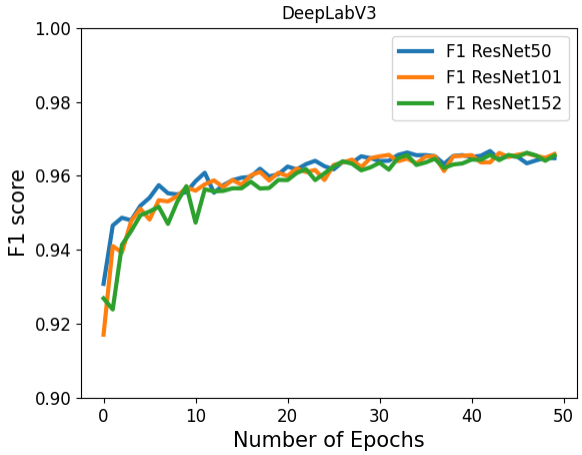


Figure 5. F1 score for training with ResNet50(blue), ResNet101(orange) and ResNet152(green) for DeepLabV3 network.

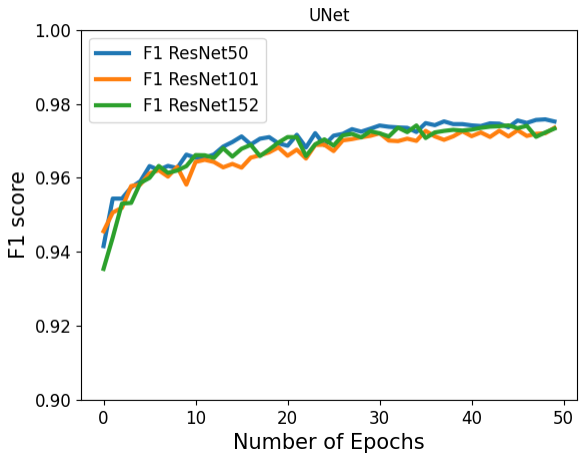


Figure 6. F1 score for training with ResNet50(blue), ResNet101(orange) and ResNet152(green) for UNet network.

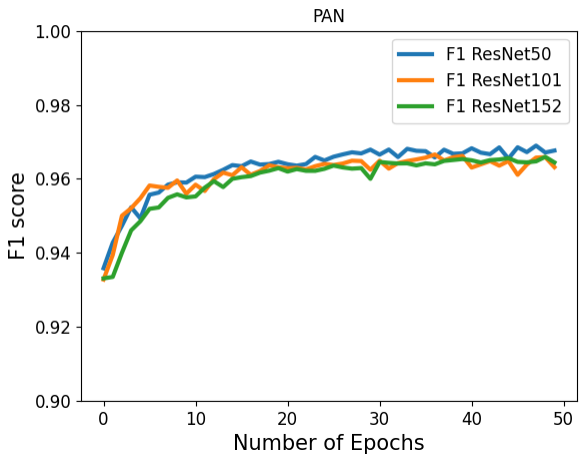


Figure 7. F1 score for training with ResNet50(blue), ResNet101(orange) and ResNet152(green) for PAN network.

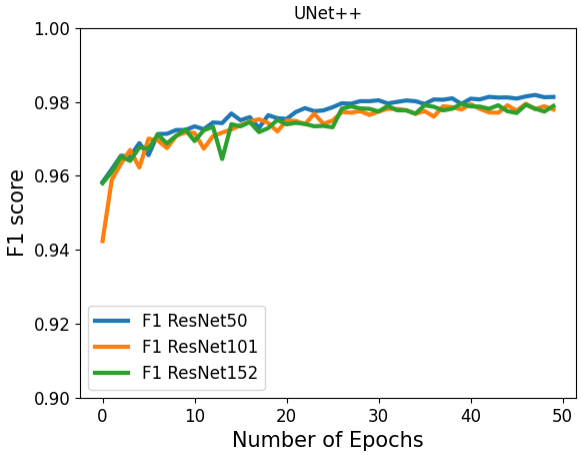


Figure 8. F1 score for training with ResNet50(blue), ResNet101(orange) and ResNet152(green) for UNet++ network.

Based on the training accuracy budget (Figures 4-8), we selected the UNet++ model with the ResNet50 backbone as our best-performing model to detect RTSs in the study area. Automated detection of RTSs using high-resolution imagery is a challenging task. A typical 0.5m resolution image scene is about 20 km x 20 km in size and contains about 1.6 billion pixels. An image scene as it is does not fit the GPU memory, therefore we need to split the image scene into small tiles. As shown in Figure 9, we first partitioned the image into 2000 x 2000 pixel tiles. Then we feed these tiles into the trained DL-CNN model for predictions.

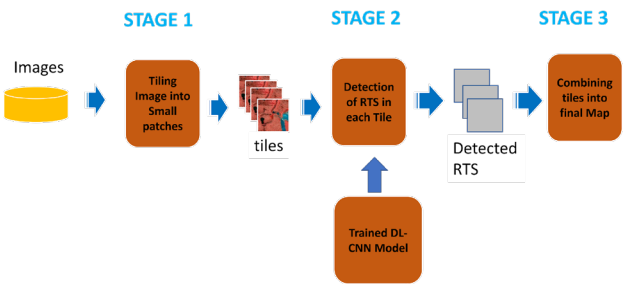


Figure 9. Semantic diagram of high-resolution satellite imagery workflow.

We used NVIDIA A100 GPU with 40Gb memory to run our DL-CNN models. The different models were executed using the PyTorch Segmentation Models library (Yakubovskiy 2019). We further utilized other libraries such as OpenCV for image processing, GDAL for accessing satellite images, and Albumentations for image augmentation.

3. RESULTS

3.1 Model Comparison

ResNet50 backbone network consistently performs better in the training stage according to Figures 4-8. Figure 10 exhibits CNN model performance with respect to the test dataset. Here we have chosen ResNet50 which was the best performing network for proceeding CNN model comparison. Accuracy scores from the comparative model analysis (Figure 10) elected the UNet++ model as the best contender The lower F1 scores were reported

by the DeepLab V3. The MANet and the UNet demonstrated the second and third best performances, respectively.

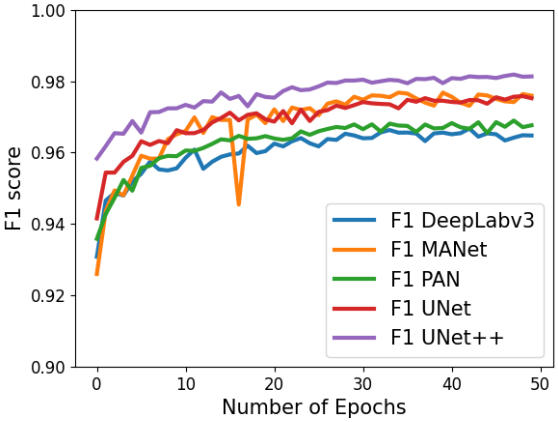


Figure 10. F1 score for training with ResNet50 with different DL-CNNs. DeepLabV3(blue), MANet(orange), PAN(green), UNet(red), UNet++(purple)

Figure 11 shows the training times for each model combination. The UNet++ model is slower compared to the other models. DeepLabV3 was the fastest among the candidate networks. The use of a lighter Resnet50 is faster in training than a larger backbone of Resnet152. Both PAN and UNet exhibited similar training time to that of DeepLab V3.

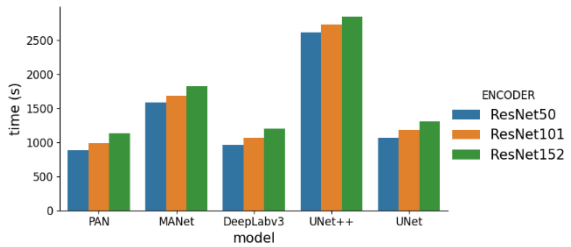


Figure 11. Time taken for training for different models with different backbones combinations.

Figure 12 depicts the F1 scores pertaining to the test data. The UNet++ outperforms the other CNN models on the test dataset. UNet++ with ResNet50 showed the highest F1 and PAN network with ResNet101 showed the lowest F1 score. ResNet50 backbone network consistently showed better F1 scores in all combinations.

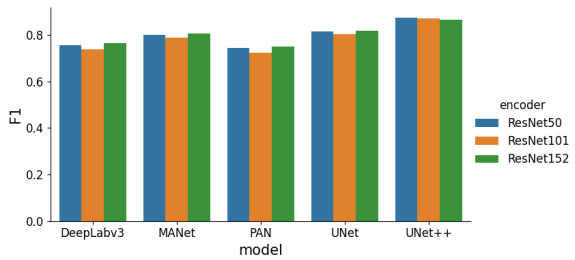


Figure 12. Reported F1 scores on test data for models trained with ResNet50(blue), ResNet101(orange), and ResNet152(green) for different DL-CNNs.

Figure 13 shows three examples of detected RTSs and ground truth annotations. Each row shows the image tile (left), ground truth (middle), and predicted RTSs (right), respectively. Visual inspections revealed that the UNet++ DL-CNN was able to accurately detect and delineate RTSs. Some miss detections were observed when the RTSs are smaller in size (see Figure13, last row).

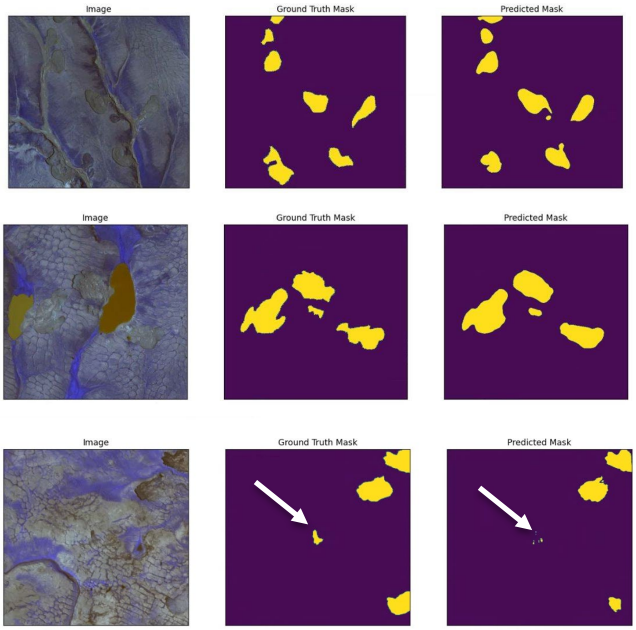


Figure 13. Three image sample tiles of the test dataset and the ground truths and predicted masks of those images. (Imagery © 2016 DigitalGlobe, Inc).

Among different CNN model-encoder combinations, the UNet++ model with the Rsetnet50 backbone demonstrated the highest accuracy (F1 score of 87%) at the expense of training and inferencing time. The PAN, DeepLabV3, MaNet, and UNet, models reported mediocre F1 scores of 72%, 75%, 80%, and 81% respectively.

3.2 RTS Prediction

We have applied the trained UNet++ model with the Reset50 backbone on satellite imagery from Banks Island and Ellesmere Island Figure 14(a) shows example detection in Banks Island. Over 90% of the RTS were correctly detected by the UNet++ with the ResNet50 backbone. Figures 14(b) show the zoomed-inviews of example areas. The trained model was able to detect the RTSs in Banks Island accurately. Figures 15(a) show the example detection in Ellesmere Island. Similar to Banks Island, over 90% of the RTS were correctly detected by the UNet++ with ResNet50 backbone. Figures 15(b) shows the zoomed-in view of example detections.

In all the cases the RTS headwall was correctly detected. In some cases, the RTS only in the scar zone (refer to an anatomy of RTS shown in Figure 1) was detected. In other instances the debris tongue was also included, however, it was not consistent across all predictions.

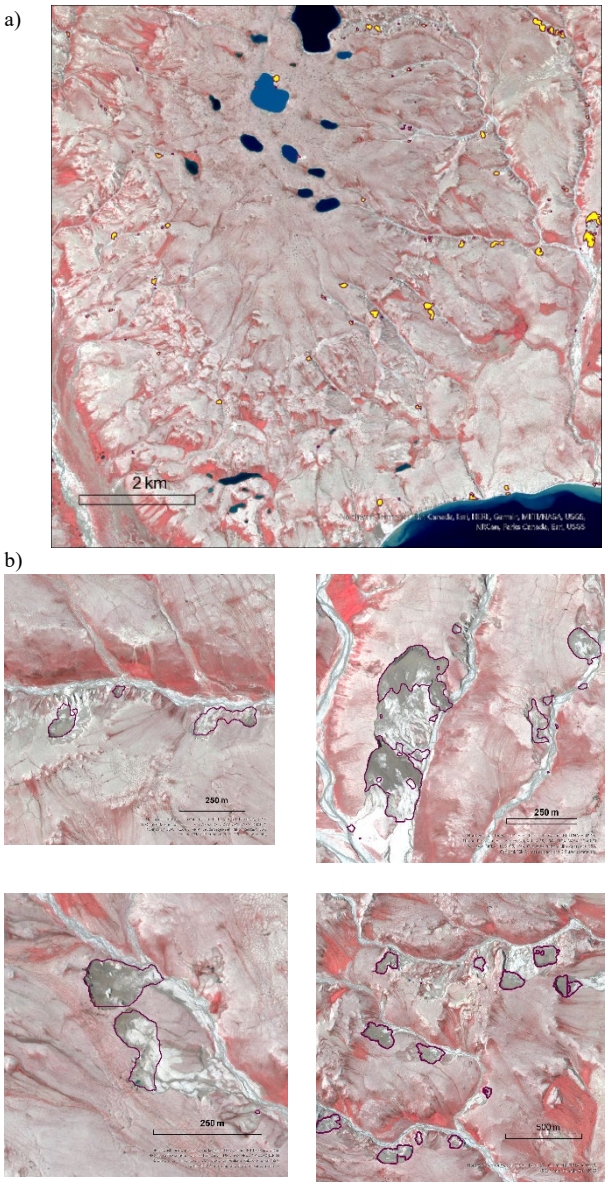


Figure 14. (a) Model application in Banks Island. (b) Zoomed-in views of the detections. (Imagery © 2016 DigitalGlobe, Inc).

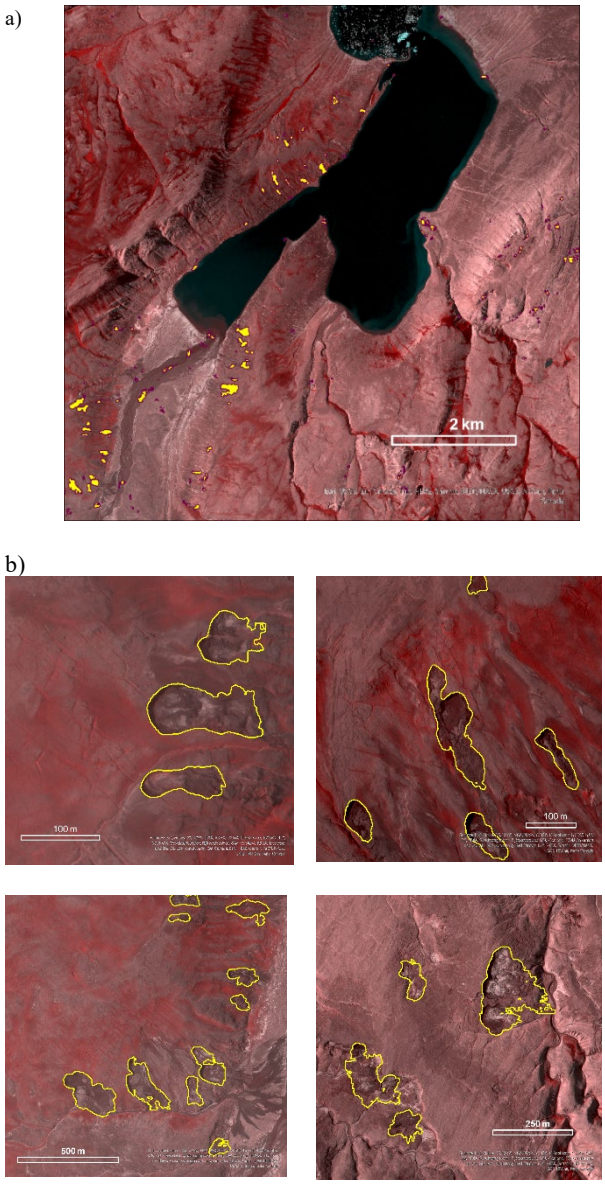


Figure 15. (a) The model application is Ellesmere Island. (b) Zoomed vies of the detections. (Imagery © 2016 DigitalGlobe, Inc).

Figure 16 demonstrates the potential of multi-temporal RTS detection. Figure 16(a) and (b) correspond to images acquired in 2015 and 2019, respectively. The green outline represents prediction based on the 2015 image and the yellow outline represents the RTS detection based on the 2019 image. As shown in the figures, we can clearly see the upward movement of the headwall in 2019. The 2015 scar zone had been stabilized by 2019. This example elucidates the potential usage of the DLCNN approaches for monitoring RTS activity using high-resolution satellite imagery. Because of the sub-meter scale spatial resolution, it is possible to differentiate RTS' morphometric variations.

4. CONCLUSION

The central goal of this study was to understand the performances of different deep learning CNN algorithms pertaining to automated recognition of retrogressive thaw slumps from very high spatial resolution commercial satellite imagery. Our comparative analysis entailed five DL-CNN with three encoders (backbone) types.

Our findings unravel the performances of different DLCNNs in imagery-enabled RTS mapping and provided useful insights on operationalizing the mapping application over large areas. We also demonstrated that our method can be used to find temporal changes in RTS accurately.

The headwalls of RTS have been detected in all the predictions. But the detection of scar zone and debris tongue boundaries were not consistent throughout the region. One reason for this can be that there is no clear definition to annotate debris tongue and scar zone. When we closely inspected RTS annotations from other studies, it was evident that the annotation process lacks formality.

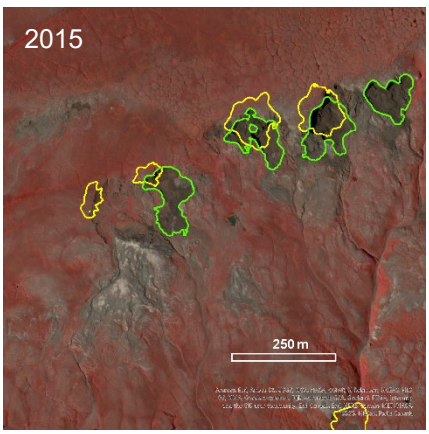
Among many, some of the important questions that arise in the annotation process include, should annotation include debris flow? deposition area? if those should be included how far away from the headwall?. In some instances, debris flow is way more extensive than the RTS itself. So consistent agreement should be prepared for consistent detection of RTS using deep learning models.

The UNet++ model performs well in our study candidate study sites. But to employ RTS detection in a circumpolar mapping context, one has to test the selected model in other areas of the Arctic. This requires a systematic model transferability analysis. Our study area is one of the more challenging to be used in DL-CNN models as there is no visible vegetation. With vegetation cover, the RTS stands out. Thus, we think that the inclusion of a substantial amount of training data representing the heterogeneity of multiple permafrost landscapes would elevate the interoperability of the UNet++ model.

5. ACKNOWLEDGEMENTS

This research was supported by the U.S. National Science Foundation grants #1927872, 1927723, 1927729, 1927720 & 1927920. The authors would like to thank the Polar Geospatial Center, the University of Minnesota for its imagery support.

a)



b)

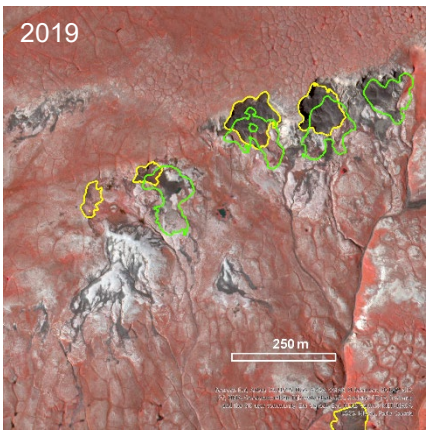


Figure 16. Multi-temporal RTS detection. (a) and (b) represent satellite images acquired in 2015 and 2019, respectively. The green outline corresponds to the RTS detection based on the 2015 image whereas the yellow outline corresponds to the RTS detection based on the 2019 image. (Imagery © 2016 DigitalGlobe, Inc).

6. REFERENCES

Chen, L.C., Papandreou, G., Schroff, F. and Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587.

Fan, T. Wang, G. Li, Y. and Wang, H. "MA-Net: A Multi-Scale Attention Network for Liver and Tumor Segmentation," in IEEE Access, vol. 8, pp. 179656-179665, 2020

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Huang, L. Luo, J. Lin, Z. Niu, F. and Liu, L. Using deep learning to map retrogressive thaw slumps in the Beiluhe region (Tibetan Plateau) from CubeSat images, Remote Sensing of Environment, Volume 237, 2020

Kokelj, S.V., and Jorgenson, M.T. 2013. Advances in thermokarst research. Permafrost and Periglacial Processes, 24: 108–119. doi:10.1002/ppp.1779

Lantz, T.C. and Kokelj, S.V., 2008. Increasing rates of retrogressive thaw slump activity in the Mackenzie Delta region, NWT, Canada. Geophysical Research Letters, 35(6).

Li, H., Xiong, P., An, J. and Wang, L., 2018. Pyramid attention network for semantic segmentation. arXiv preprint arXiv:1805.10180.

Ronneberger, O., Fischer, P. and Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

Screen, J., Simmonds, I. The central role of diminishing sea ice in recent Arctic temperature amplification. Nature 464, 1334–1337 (2010). https://doi.org/10.1038/nature09051

Segal, R.A. Lantz, T.C Kokelj S.V. Inventory of active retrogressive thaw slumps in the Peel Plateur, Northwest Territories, NWT Open Report 2015-020.

Witharana, C. Udawalpola, M.R. Liljedhl, A.K. Ward-Jones, M. Jones, B., Hasan, A Joshi D. Manos, E. Automated recognition of retrogressive thaw slumps in the Arctic permafrost tundra using high spatial resolution commercial satellite imagery, Remote Sensing, 2022 (In Reiveiw)

Yakubovskiy, P. Segmentation Model Pytorch, Github repository, 2019. https://github.com/qubvel/segmentation_models.pytorch

Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N. and Liang, J., 2018. Unet++: A nested u-net architecture for medical image segmentation. In Deep learning in medical image analysis and multimodal learning for clinical decision support (pp. 3-11). Springer, Cham.