Recurrent Graph Convolutional Networks for Spatiotemporal Prediction of Snow Accumulation Using Airborne Radar

Benjamin Zalatan
Department of Computer Science and Engineering
Lehigh University
Bethlehem, PA, USA
bjz222@lehigh.edu

Maryam Rahnemoonfar
Department of Computer Science and Engineering
Department of Civil and Environmental Engineering
Lehigh University
Bethlehem, PA, USA
maryam@lehigh.edu

Abstract—The accurate prediction and estimation of annual snow accumulation has grown in importance as we deal with the effects of climate change and the increase of global atmospheric temperatures. Airborne radar sensors, such as the Snow Radar, are able to measure accumulation rate patterns at a large-scale and monitor the effects of ongoing climate change on Greenland's precipitation and run-off. The Snow Radar's use of an ultra-wide bandwidth enables a fine vertical resolution that helps in capturing internal ice layers. Given the amount of snow accumulation in previous years using the radar data, in this paper, we propose a machine learning model based on recurrent graph convolutional networks to predict the snow accumulation in recent consecutive years at a certain location. We found that the model performs better and with more consistency than equivalent nongeometric and nontemporal models.

Index Terms—deep learning, graph neural networks, ice thickness, remote sensing

I. INTRODUCTION

As global atmospheric temperatures increase over time, the accurate prediction of annual snow accumulation and internal layers of ice sheets grows in importance. Tracking and forecasting these internal ice sheet layers is important for calculating snow mass balance, extrapolating ice age from direct measurements of the subsurface, and inferring otherwise difficult to observe ice dynamic processes. Precise understanding of the spatiotemporal variability of snow accumulation in the Greenland ice sheet is important to reducing the uncertainties in current climate model predictions and future sea level rise.

Measurements about the mass balance of ice sheets are traditionally collected by drilling ice cores and shallow pits. However, it is challenging to capture catchment-wide accumulation rates due to inherent sparsity, access difficulty, and high cost. Attempts to interpolate in-situ measurements further introduce uncertainties due to the high variability in local accumulation rate.

Airborne measurements collected over Greenland using nadir-looking radar sensors is one of the complementing methods of mapping topography and monitoring accumulation rates with a broad spatial coverage that has the advantage of revealing isochronous ice layers beneath the ice surface. Sub-surface ice layers that appear in snow radar echograms unveil historic annual and multi-annual snow accumulation. Snow Radar [1] provides a spatiotemporal accumulation map that can help in quantifying the impact of atmospheric warming on polar ice caps and project its contribution to future sea level rise.

A large amount of airborne radar data from polar regions has been collected over the past few decades. However, the complexity, scale, and heterogeneity of the data has necessitated advanced automatic solutions. In recent years, several deep learning approaches [2]–[10] based on convolutional neural networks (CNNs) have been proposed to track the internal layers of ice sheets and calculate historic snow accumulation. Snow Radar data, even after post-processing, contains a significant amount of noise. In addition to being very sensitive to noise, CNNs cannot be used for spatiotemporal tasks, such as predicting snow accumulation of the future years based on data from the past years.

Recent studies involving graph convolutional networks (GCNs) [11] have shown promise in spatiotemporal tasks such as traffic forecasting [12] [13] [14], wind speed forecasting [15], and power outage prediction [16]. In this paper, we propose a geometric deep learning model that uses a supervised multi-target long short-term memory graph convolutional network (GCN-LSTM) [17] to predict the thickness of multiple shallow ice layers at specific coordinates given the thicknesses of multiple deep ice layers.

We focus on the Snow Radar [1] dataset collected by the Center for Remote Sensing of Ice Sheets (CReSIS) as part of NASA's Operation IceBridge. The Snow Radar operates from 2-8 GHz and is able to track deep layers of ice with a high resolution over wide areas of an ice sheet. The sensor produces a two-dimensional grayscale profile of historic snow accumulation over consecutive years (see the left image in Figure 1), where the horizontal axis represents the along-track direction, and the vertical axis represents layer depth. Pixel brightness is directly proportional to the strength of the returning signal. Pixels representing surface layers are generally brighter and more well-defined due to high reflectance and snow density
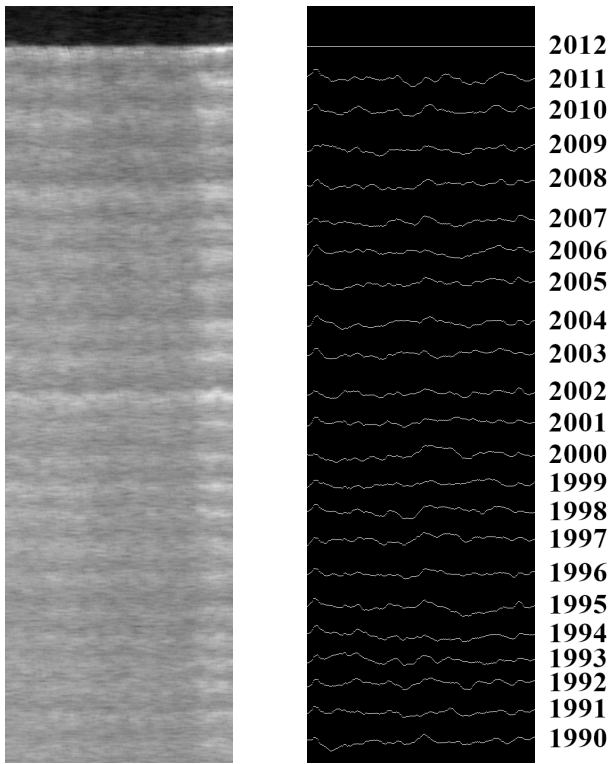
Fig. 1. Snow radar echogram (left) and its corresponding binary ground-truth labelling (right) showing layers of the Greenland ice sheet and their respective years.

variation, while pixels representing deeper layers are generally darker and noisier due to increased density and a lower return-signal strength. In our experiments, we use radar data from selected Snow Radar flights over Greenland in the year 2012. In many areas, each ice layer represents an annual isochrone [18]. As such, we may refer to specific ice layers by their corresponding year (see Figure 1).

In our experiments, we convert internal ice layer thickness information in radar images into sequences of temporal graphs to be used as input to our model. More specifically, we convert ten deep ice layers (corresponding to annual snowfall from 1997 to 2006, respectively) into ten spatiotemporal graphs. Our model then performs multi-target regression to predict the thicknesses of the five shallow ice layers immediately below the surface, corresponding to annual snow accumulation from 2007 to 2011, respectively.

Our contribution in this work includes the following:

1) A method of converting Snow Radar ice layer data into sets of spatiotemporal graphs.
2) A generalizable machine learning model based on GCN-LSTM that predicts the thicknesses of shallow ice layers.

Our proposed model was shown to perform better with more consistency than equivalent non-geometric and non-temporal models.

II. RELATED WORK

A. Deep Learning for the Calculation of Ice Layer Thickness and Snow Accumulation

Automated techniques, such as [19]–[21], have been developed for detecting the surface and bottom layers of echograms using radar depth sounder sensors. Tracking the internal layers of an ice sheet is inherently more challenging due to the close proximity of each layer. Due to its exceptional performance in feature extraction, deep learning has been applied extensively on remote sensing images, such as Snow Radar data, for tracking the internal layers of ice sheets [2] [3] [7] [8]. [7] used a multi-scale contour-detection CNN to segment the different internal ice layers found in Snow Radar images. The authors of this work also experimented with pretraining on the Berkeley Segmentation Dataset and Benchmark (BSDS) dataset [22] and found that this was not very effective due to the large amount of noise in Snow Radar images. In [2], the authors trained a multi-scale neural network on synthetic Snow Radar images for more robust training. A multi-scale network was also used in [8], where the authors trained a model on images from the year 2012 and fine tuned it by training on a small number of images from other years. [3] found that using pyramid pooling modules, a type of multi-scale architecture, helps in learning the spatio-contextual distribution of pixels for a certain ice layer. The authors also found that denoising the images prior to the CNN layer improved both the accuracy and F-score. These works have all used multi-scale networks in order to segment the ice layers within Snow Radar images, but have commonly noted that the high amount of noise present in Snow Radar images is an issue that must be addressed in the future.

While these works have attempted to track and segment the annual ice layers present within Snow Radar images, none have yet attempted to use the spatiotemporal patterns of the ice layers to predict the thicknesses of shallow ice layers given the thicknesses of deeper ice layers.

B. Graph Convolutional Networks

Graph convolutional networks have had numerous real-world applications in a vast array of different fields. In the field of computer vision, recurrent GCNs have been used to generate and refine "scene graphs", where each node corresponds to the bounding box of an object in an image and the edges between nodes are weighted by a learned "relatedness" factor that prunes unrelated edges [23] [24]. GCNs have also been used to segment and classify point clouds generated from LiDAR scans [25] [26]. Recurrent GCNs have been used in traffic forecasting, such as in [12], where graph nodes represented sensors installed on different roads, edges between nodes were weighted by the distance between sensors, and node features consisted of average detected traffic speed over a certain period of time.

Some existing recurrent graph-based weather prediction models, such as [15] and [27], have experimented with defining the edge weights in graph adjacency matrices as learnable

TABLE I
KEY PARAMETERS OF THE SNOW RADAR USED DURING DATA
COLLECTION.

	Snow Radar Parameters
Bandwidth	2-8 GHz
Pulse duration	250 μs
PRF	2 kHz
Transmit power	100 mW
Intermediate frequency range	62.5-125 MHz
Sampling frequency	125 MHz
Range resolution	$\sim$ 4 cm
Along-track footprint	14.5 m

parameters rather than static values. While this did show slightly improved performance by allowing a model to learn relationships between nodes more complex than simple geographic distance, it exponentially increases the computational complexity of the model, making it considerably more difficult to train and increasing its susceptibility to overfitting. While it is possible to reduce the number of learnable parameters if the adjacency matrix is low-rank, this is not always guaranteed and the number of learnable parameters may remain very high. This number may also be decreased by using a more sparse graph structure, such as restricting a node's neighborhood to its k-nearest neighbors. This strategy can be applied for some problems, but may not work for spatial tasks where most or all nodes are in close proximity.

III. DATASET

In this study, we used Snow Radar data captured during the NASA's Operation Ice Bridge mission [1]. CReSIS has made this data publicly available on their website (https://data.cresis.ku.edu/). The data gathered during each flight was processed and separated into a series of grayscale radar echogram images, each with a width of 256 pixels and a height ranging between 1200 and 1700 pixels. Each pixel in a column corresponds to approximately 4 cm of ice, and each radar image has an along-track footprint of 14.5m (see Table I). The value of each pixel is proportional to the relative received power of the returning radar signal at that depth. Accompanying each image was a 256×2 matrix specifying the geographic latitude and longitude associated with each vertical column of pixels. In order to gather ground-truth thickness data, the images were manually labelled in a binary format where white pixels represented the tops of each firn layer, and all other pixels were black (see the right image in Figure 1).

In order to capture a sufficient amount of data, only radar images containing a minimum of 16 ice layers were used: one unused surface layer, five shallow layers for ground truths, and ten deeper layers for use as node features (the surface sheet is unused as it remains flat with constant height throughout all images in the dataset). Ten deep layers and five shallow layers were chosen in order to maximize the number of usable images while also maintaining a reasonable amount of data per image. This limitation reduced the total number of usable images from 816 down to 568. Five different training and testing sets were generated by taking five random permutations of

all usable images and splitting them at a ratio of 4:1. The five training sets contained 454 images each, and the five testing sets contained 114 images each.

IV. METHODS

A. Graph Convolutional Networks

Traditional CNNs use a matrix of learnable weights, often referred to as a kernel or filter, as a sliding window across pixels in an input image. This allows a model to automatically extract image features that would otherwise need to be identified and inputted manually. Graph convolutional networks apply similar logic to graphs, but rather than using a sliding window of weights across an image, GCN performs weighted-average convolution on each node's neighborhood to extract features that reflect the structure of a graph. The size of the neighborhood on which convolution takes place is dictated by the number of GCN layers present in the model (i.e. K GCN layers result in K-hop convolution). In a sense, graph convolutional networks are a generalized form of traditional CNN's that enable variable degree and weighted adjacency.

B. Recurrent Neural Networks

Recurrent neural networks (RNNs) are able to process a sequence of data points as input, rather than a single static data point, and learn the long-term relationships between them. Many traditional RNN structures had issues with vanishing and exploding gradients on long input sequences. Long short-term memory (LSTM) attempts to mitigate those issues by implementing gated memory cells that guarantee constant error flow [28]. Applying LSTM to GCN using GCN-LSTM allows for a model to learn not only the relationships between nodes in a graph, but also how those relationships change (or persist) over time.

C. Model Architecture

Our model uses ten consecutive GCN-LSTM layers (one for each of the ten deep ice layers from 1997 to 2006), each with 64 output channels, that lead into two fully-connected layers. The first fully-connected layer has 32 output channels, and the second fully-connected layer has five output channels. The output of the model is the predicted thickness (in pixels) of the five shallow ice layers (2007 to 2011) immediately beneath the surface. Between each layer is the Hardswish activation function [29], chosen due to its superior performance when compared to ReLU and its derivatives [30]. Between the two fully-connected layers is Dropout [31] with $p = 0.2$. We use the Adam optimizer [32] with a constant learning rate of 0.01, and we train for 150 epochs using mean-squared error loss. The structure of the model is shown visually in Figure 2.

D. Graph Generation

1) Model Input: Each radar image is converted into ten graphs, each consisting of 256 nodes. Each graph corresponds to a single ice layer for each year from 1997 to 2006. Each node represents a vertical column of pixels in the radar echogram image and has three features: two for the latitude
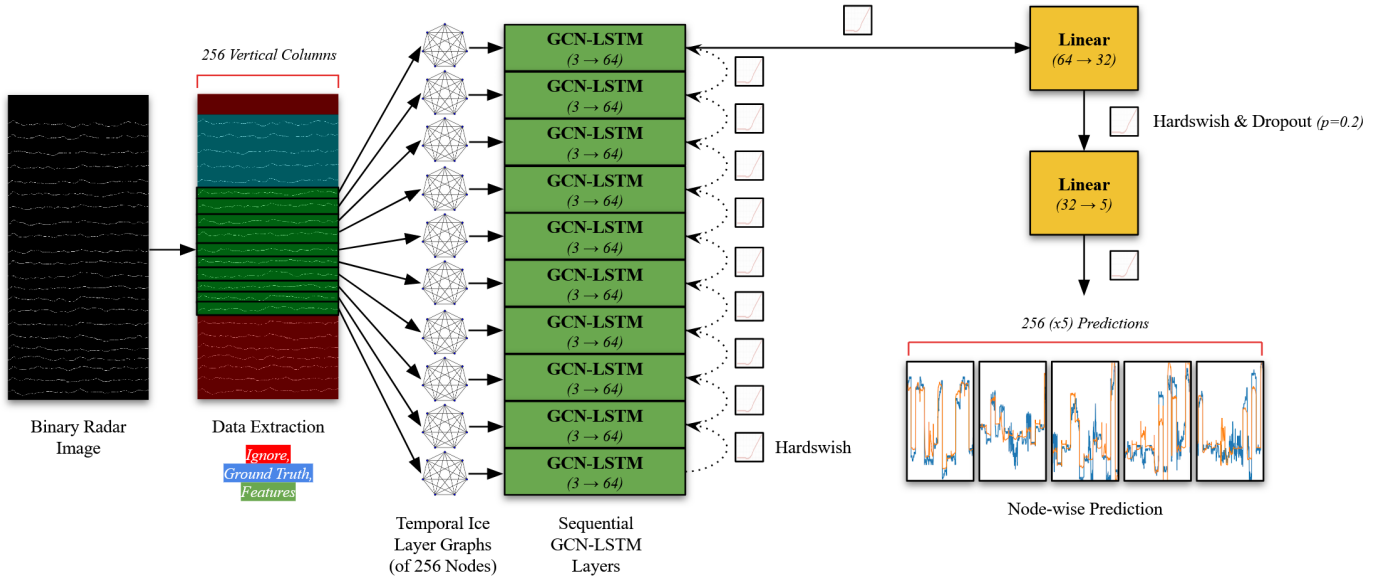
Fig. 2. Architecture of the proposed model.

and longitude that correspond to that column, and one for the thickness of the corresponding year's ice layer within that column.

2) Graph Adjacency: All graphs are fully connected and undirected. All edges are inversely weighted by the geographic distance between node locations using the haversine formula. For a weighted adjacency matrix A:

$$A_{i,j} = \frac{1}{2\arcsin\left(\text{hav}(\phi_j - \phi_i) + \cos(\phi_i)\cos(\phi_j)\ \text{hav}(\lambda_j - \lambda_i)\right)}$$

where

$$\text{hav}(\theta) = \sin^2\left(\frac{\theta}{2}\right)$$

$A_{i,j}$ represents the weight of the edge between nodes i and j. ϕ and λ represent the latitude and longitude features of a node, respectively. Node features of all graphs are collectively normalized using z-score normalization. Weights in the adjacency matrices of all graphs are collectively normalized using min-max normalization.

V. RESULTS

A. Baselines

To highlight the performance of the GCN-LSTM model, as well as verify that the temporal and geometric aspects of the model serve to its benefit, we compared its performance with equivalent models that utilize a traditional GCN as well as a traditional, non-geometric LSTM.

For the GCN model, all hyperparameters remain the same, but the GCN-LSTM layer sequence is replaced by a single GCN layer. Rather than generating ten graphs for each of the ten deep ice layers, we generate a single graph with each node having 12 features: two for the latitude and longitude, and ten for the thicknesses of the 1997 to 2006 deep ice layers. The

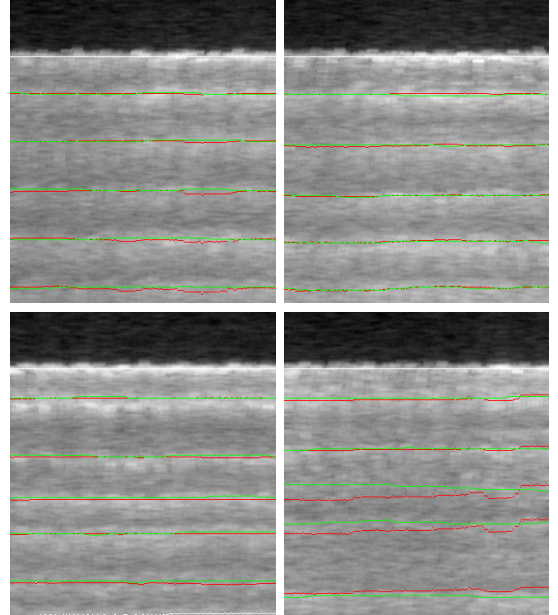rest of the model, including the adjacency matrix generation, is identical to the proposed model.

Fig. 3. Example radar images overlayed with predicted thicknesses from our model. Green pixels are the ground truth, red pixels are the predicted value. A poor-fitting example is shown in the bottom-right.

For the LSTM model, all hyperparameters remain the same, but ten sequential LSTM layers are used, rather than ten sequential GCN-LSTM layers. Since this model is non-geometric, nodes are simply converted into rows in a feature vector, and no adjacency information is supplied. The rest of the model is identical to the proposed model.

TABLE II
RESULTS FROM THE LSTM, GCN, AND GCN-LSTM MODELS ON THE FIVE PREDICTED ANNUAL ICE LAYER THICKNESSES FROM 2007 TO 2011.
RESULTS ARE SHOWN AS THE MEAN $\pm$ STANDARD DEVIATION OF THE RMSE OVER FIVE TRIALS (IN PIXELS).

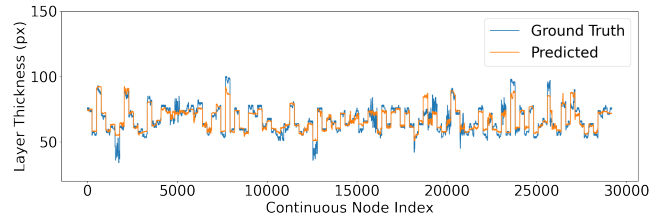|	2007	2008	2009	2010	2011	Total
LSTM	7.834 ± 1.050	8.620 ± 3.392	7.243 ± 1.213	5.653 ± 0.370	5.995 ± 0.764	7.268 ± 1.162
GCN	6.165 ± 0.431	6.458 ± 0.866	6.376 ± 1.381	5.267 ± 0.184	5.184 ± 0.768	5.956 ± 0.468
GCN-LSTM	$\mathbf{4.614 \pm 0.504}$	$\mathbf{4.733 \pm 1.111}$	$\mathbf{5.278 \pm 1.074}$	$\mathbf{4.449 \pm 0.336}$	$\mathbf{3.855 \pm 0.272}$	$\mathbf{4.651 \pm 0.414}$
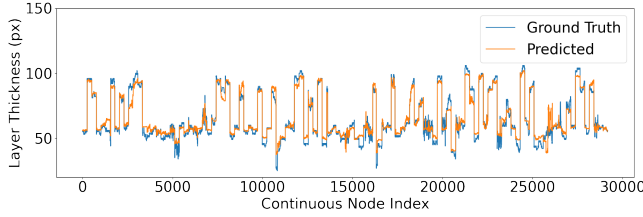
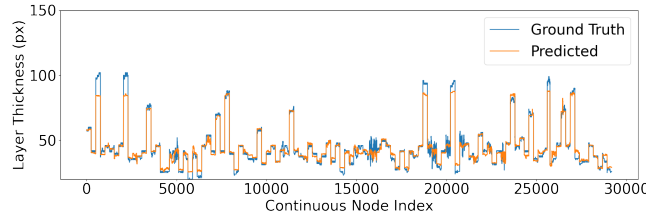(a) Continuous predicted thicknesses for 2007.
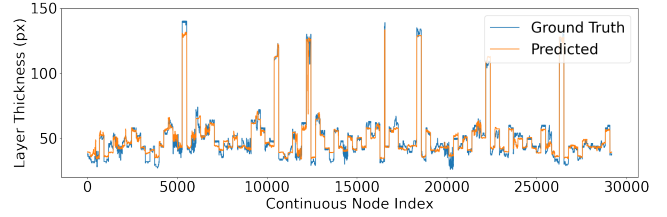
(b) Continuous predicted thicknesses for 2008.
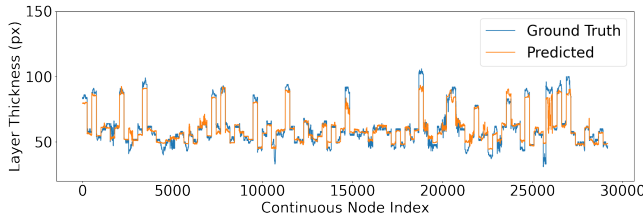
(c) Continuous predicted thicknesses for 2009.

(d) Continuous predicted thicknesses for 2010.

(e) Continuous predicted thicknesses for 2011.

Fig. 4. Continuous real vs. predicted thicknesses for years 2007-2011. Ground truths are labelled in blue, predicted values are labelled in orange. The y-axis represents ice layer thickness in pixels, and the x-axis represents the unsorted, randomized node index in the testing set.

B. Experimental Results

In each trial, the root mean squared error (RMSE) was taken between the predicted and ground truth thicknesses for each of the 2007-2011 ice layers among the entire testing set. The mean RMSE and standard deviations were found for each year over all five trials. These results are showcased in Table II. Example well-fitting and poor-fitting results are overlaid on their respective radar echogram images in Figure 3. Graphed continuous results over all (unordered) testing samples for years 2007-2011 are shown in Figure 4. The proposed GCN-LSTM model performed significantly better than the baseline GCN and LSTM models both in terms of mean RMSE and consistency (shown by a lower standard deviation). The lack of adjacency data in the LSTM model and complex temporal learning in the GCN model may contribute to these results.

VI. CONCLUSION

In this work, we proposed a temporal, geometric, multi-target machine learning model based on GCN-LSTM that predicts the annual snow accumulation of Greenland from 2007 to 2011 given the annual snow accumulation from 1997 to 2006. Our proposed model was shown to perform better and with more consistency than equivalent non-geometric and non-temporal models. While our model succeeds at predicting shallow layer thicknesses with reasonable accuracy, there are opportunities for improvement and generalization, some of which are outlined in the next section.

A. Generalizability and Future Work

1) Additional Ice Layers: While our model was restricted to learning on only ten ice layers and predicting only five ice layers, it is easily expandable to a higher number of both feature and predicted layers. Future work may take advantage of this expandability and test a similar model on a much larger number of layers.

2) Additional Node Features: In our model, minimal node features are present; only latitude, longitude, and layer thickness are used as node features. It may be possible to improve

performance by integrating additional node features, such as physical properties of the ice at each layer.

3) Learnable Graph Adjacency: In our model, edge weights between nodes are static values inversely proportional to their geographic distance. It may be possible to improve performance by defining edge weights as learnable parameters, allowing the model to learn relationships between nodes more complex than simple geographic distance.

4) Generalizing to Future Snow Accumulation: In these experiments, our model is restricted to learning and predicting on data gathered about existing ice layers. If radar image data gathered over multiple years were used, our model could be generalized to forecast future snow accumulation.

5) Generalizing to Deep Ice Layers: Since Snow Radar echogram images inherently become noisier and more difficult to track as depth and firn density increases, it may be useful to attempt to reverse this model and use the thicknesses of shallow ice layers to predict the thicknesses of deeper ice layers.

VII. Acknowledgements

References

[1] S. Gogineni, J. B. Yan, D. Gomez, F. Rodriguez-Morales, J. Paden, and C. Leuschen. Ultra-wideband radars for remote sensing of snow and ice. In *IEEE MTT-S International Microwave and RF Conference*, pages 1–4, 2013.

[2] Maryam Rahnemoonfar, Masoud Yari, John Paden, Lora Koenig, and Oluwanisola Ibikunle. Deep Multi-Scale Learning for Automatic Tracking of Internal Layers of Ice in Radar Data. *Journal of Glaciology*, 67(261):39–48, 2021.

[3] Debvrat Varshney, Maryam Rahnemoonfar, Masoud Yari, and John Paden. Deep Ice layer Tracking and Thickness Estimation using Fully Convolutional Networks. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 3943–3952. IEEE, 2020.

[4] Debvrat Varshney, Maryam Rahnemoonfar, Masoud Yari, John Paden, Oluwanisola Ibikunle, and Jilu Li. Deep Learning on Airborne Radar Echograms for Tracing Snow Accumulation Layers of the Greenland Ice Sheet. *Remote Sensing*, 13(14), 2021.

[5] Debvrat Varshney, Masoud Yari, Tashnim Chowdhury, and Maryam Rahnemoonfar. Refining Ice Layer Tracking through Wavelet combined Neural Networks. 2021.

[6] Masoud Yari, Oluwanisola Ibikunle, Debvrat Varshney, Tashnim Chowdhury, Argho Sarkar, John Paden, Jilu Li, and Maryam Rahnemoonfar. Airborne snow radar data simulation with deep learning and physics-driven methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:12035–12047, 2021.

[7] Masoud Yari, Maryam Rahnemoonfar, John Paden, Ibikunle Oluwanisola, Lora Koenig, and Lynn Montgomery. Smart Tracking of Internal Layers of Ice in Radar Data via Multi-Scale Learning. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 5462–5468. IEEE, 2019.

[8] Masoud Yari, Maryam Rahnemoonfar, and John Paden. Multi-Scale and Temporal Transfer Learning for Automatic Tracking of Internal Ice Layers. In *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, pages 6934–6937, 2020.

[9] Masoud Yari, Oluwanisola Ibikunle, Debvrat Varshney, Tashnim Chowdhury, Argho Sarkar, John Paden, Jilu Li, and Maryam Rahnemoonfar. Airborne snow radar data simulation with deep learning and physics-driven methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:12035–12047, 2021.

[10] Debvrat Varshney, Oluwanisola Ibikunle, John Paden, and Maryam Rahnemoonfar. Learning snow layer thickness through physics defined labels. In *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, pages 1233–1236, 2022.

[11] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *Proceedings of ICLR 2017*, 2017.

[12] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting, 2017.

[13] Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018.

[14] Shengnan Guo, Youfang Lin, Ning Feng, Chao Song, and Huaiyu Wan. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 922–929, Jul. 2019.

[15] Dogan Aykas and Siamak Mehrkanoon. Multistream graph attention networks for wind speed forecasting. *CoRR*, abs/2108.07063, 2021.

[16] Damian Owerko, Fernando Gama, and Alejandro Ribeiro. Predicting power outages using graph neural networks. pages 743–747, 11 2018.

[17] Youngjoo Seo, Michaël Defferrard, Pierre Vandergheynst, and Xavier Bresson. Structured sequence modeling with graph convolutional recurrent networks. In *Proceedings of ICLR 2017*, 2017.

[18] Lora S Koenig, Alvaro Ivanoff, Patrick M Alexander, Joseph A MacGregor, Xavier Fettweis, Ben Panzer, John D Paden, Richard R Forster, Indrani Das, Joesph R McConnell, et al. Annual Greenland accumulation rates (2009–2012) from airborne snow radar. *The Cryosphere*, 10(4):1739–1752, 2016.

[19] M. Rahnemoonfar, G. C. Fox, M. Yari, and J. Paden. Automatic ice surface and bottom boundaries estimation in radar imagery based on level-set approach. *IEEE Transactions on Geoscience and Remote Sensing*, 55(9):5115–5122, 2017.

[20] M. Rahnemoonfar, A. A. Habashi, J. Paden, and G. C. Fox. Automatic ice thickness estimation in radar imagery based on charged particles concept. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 3743–3746, 2017.

[21] Maryam Rahnemoonfar, Jimmy Johnson, and John Paden. Ai radar sensor: Creating radar depth sounder images based on generative adversarial network. *Sensors*, 19(24):5479, Dec 2019.

[22] Pablo Arbeláez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour Detection and Hierarchical Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):898–916, 2011.

[23] Danfei Xu, Yuke Zhu, Christopher B. Choy, and Li Fei-Fei. Scene graph generation by iterative message passing, 2017.

[24] Jianwei Yang, Jiasen Lu, Stefan Lee, Dhruv Batra, and Devi Parikh. Graph r-cnn for scene graph generation, 2018.

[25] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. Dynamic graph cnn for learning on point clouds, 2018.

[26] Loic Landrieu and Martin Simonovsky. Large-scale point cloud semantic segmentation with superpoint graphs, 2017.

[27] Tyler Wilson, Pang-Ning Tan, and Lifeng Luo. A low rank weighted graph convolutional approach to weather prediction. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 627–636, 2018.

[28] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9:1735–80, 12 1997.

[29] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, and Hartwig Adam. Searching for mobilenetv3. *CoRR*, abs/1905.02244, 2019.

[30] Prajit Ramachandran, Barret Zoph, and Quoc V. Le. Searching for activation functions, 2017.

[31] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors, 2012.

[32] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.