IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium | 978-1-7281-6374-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/IGARSS39084.2020.9323758

MULTI-SCALE AND TEMPORAL TRANSFER LEARNING FOR AUTOMATIC TRACKING OF INTERNAL ICE LAYERS

Masoud Yari [1], Maryam Rahnemoonfar [1], John Paden [2]

1. Computer Vision and Remote Sensing Laboratory, University of Maryland, Baltimore County, USA
2. Center for Remote Sensing of Ice Sheets, University of Kansas, Kansas, USA.

ABSTRACT

Pragmatic Deep Learning techniques in recent years have greatly influenced our approaches to data analysis. However, in many real-world problems, even when a large dataset is available, Deep Learning methods have shown less success, for the lack of large labeled dataset, presence of noise, or missing data. In this work, our goal is to track internal ice layers in radar images gathered with various sensors in different years. We will show that transfer learning will not generally work well. However, if the Deep Learning model gets trained on noisy images, there would be a significant improvement. Unlike spatial Transfer Learning, our experiments show that temporal Transfer Learning can provide considerably better results.

Index Terms—

neural network, radar images, automatic tracking, machine learning, multi-class classification

1. INTRODUCTION

Deep learning methods have displayed impressive success in many areas of practical interest such in optical imagery, as classification [1], object recognition [2], counting [3, 4, 5], and semantic segmentation [6, 7]. However, it has been less successful in problems related to non-optical sensors, such as radar imagery, due to coherent noise in the data. The goal of this work is to examine deep neural network for tracking the internal layers of ice.

Surface and bottom tracking in radar images has been extensively studied in several works, [8, 9, 10, 11, 12]. But tracking internal layers is a significantly more difficult task because of the large number of layers in close proximity [13, 14]. MacGregor et al. [13] developed a semi-automated layer tracker for the CReSIS radar depth sounder and applied it to several seasons of data in the first large scale effort to do internal layer tracking. Even using the semi-automated method, the task took several years to complete. Koenig et al. [14] have tracked these interfaces and used the tracked layers to measure annual snow fall over a large area. Since the techniques are not able to detect all internal layers and are not fully automatic, they cannot be easily scaled up to big dataset for routine application.

In the present work, we aim to analyze snow Radar [15] images produced by the Center of Remote Sensing of Ice Sheets for NASA Operation IceBridge. The snow radar is a profiling instrument which produces vertical sounding images of snow layers over ice sheets and ice caps. The radar signal is sensitive to annual density changes that occur due to the seasonal transitions from summer to winter; this density change interface scatters the radar signal which is measured by the radar's receiver. For training our network we used the output of semi-supervised layer tracking by [14], which were corrected by human.

The best labeled data we have belongs to year 2012. This is also our largest dataset from year 2008 to 2017. We used this dataset for training. It must be noted that the labeled data is produced semi-automatically. Our goal is to develop a deep learning approach that can track ice layer boundaries in the presence of noise and missing information.

2. METHODOLOGY

We denote our training dataset by $\mathbb{X}$. Our dataset consists of pair of vectors (X, Y) where X is the original image and Y is the corresponding binary layer data. In a multi-scale learning process, we define an edge detection function at each scale s, and we denote it by E_s . We pull out several side-outputs at different scales in our network. Suppose we denote each output at scale s by $O_s(X)$. Then the edge map prediction at scale s is $\widetilde{Y}_s = D_s(O_s(X))$, where $D_s(\cdot)$ is an upsampling function to re-scale the output to match the size of the layers vector. The final prediction is defined as a linear combination of edge map predictions across all scales. In other words,

$$\widetilde{Y} = \sum_{s \in S} w_s \widetilde{Y}_s, \tag{1}$$

where S is the set of all scales. We also define loss functions at all scales, denoted by ℓ_s, and one loss function for the final prediction $\widetilde{Y}$, denoted by ℓ_f. The learning process can be described as

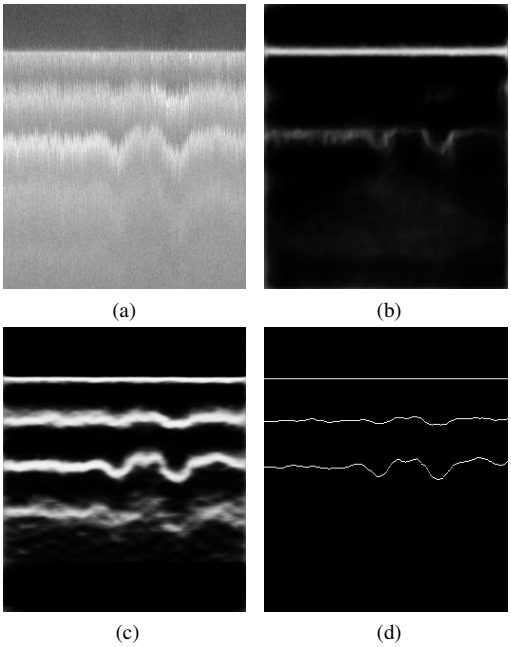

Fig. 1. (a) The original image. (b) The result of training on augmented BSDS500. (c) The result of training from scratch on ICE2012. (d) Human labeled edges.

$$(\theta, w)^* = \underset{\theta, w}{argmin} \left(\ell_f + \sum_{s \in S} \ell_s \right) \tag{2}$$

with $\theta = (\theta_0, \theta_s)_{s \in S}$, where θ_0 represents parameters of the network, θ_s output parameters at scale s, and $w = (w_s)_{s \in S}$ the fusion weights.

3. EXPERIMENTAL RESULTS AND EVALUATION

We took two main approaches for the learning method. First we examined transfer learning. We carried out several experiments in transfer learning; the best result came from training the model on an augmented BSDS500 data. In the second approach, we trained the model on our 2012 dataset with a normal distribution initialization. The test results in the second approach was considerably better than the first approach, that is transfer learning. In the training process, we have the learning rate 10^{-6} and the momentum 0.9. We have also used weight decay rate of 2×10^{-4}. Fig.1 shows the outputs of both experiments for one test sample.

As discussed in some recent works [16], deep learning models are unstable in the presence of noise. Therefore, one can expect that the first method do poorly in tracking the edges, but, training from scratch provides much better results even with respect to the human labeling (See Table 1).

Our dataset contains data from 2009 to 20017. However, the quality of images in all years are lower comparing to 2012 data; also we have less labeled data in other years. Therefore we were not able to use them in training process. Our 2012 dataset consists of 2360 train and 260 test images. We have used our multi-scale learning approach, described in the methodology section, on 2012 data. In Section sec.method, it was mentioned that the model produces side-outputs in different scales. In our model, there are five side-outputs. Finally, there is the fusion of the side-outputs. Figure 3 provides a sample of all six outputs of one input data. The fusion coefficients are also learned during the training procedure. Each time we apply the model on a test dataset, the model is going to produce five side-outputs and a fusion result.

The test results for year 2012, is provided in Table 1. For other years, we used transfer learning. Here we provide a few examples of our results for different years. Fig. 2(a) for instance is an image taken from our images in 2009. (b) shows

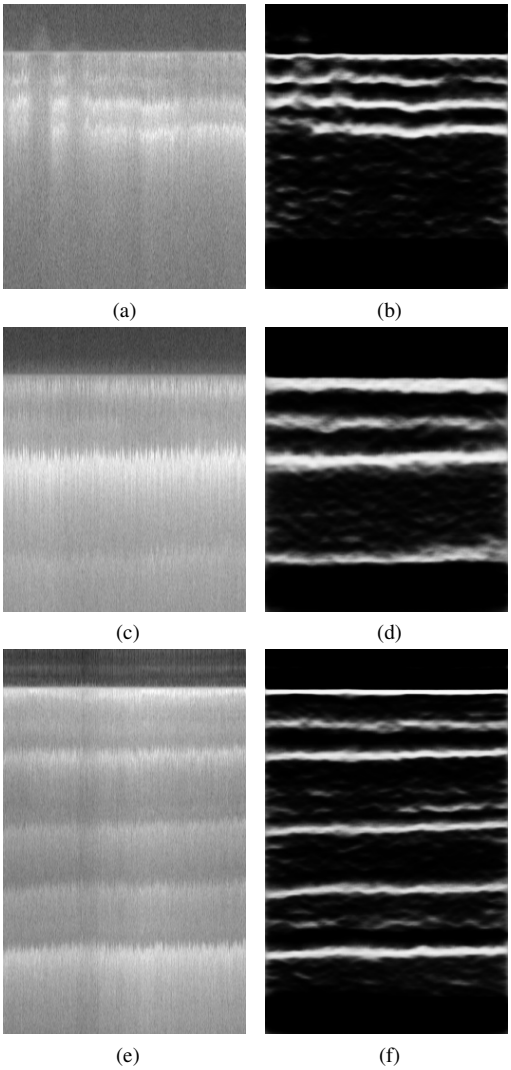

Fig. 2. Left: original images from 2009 (a), 2011(c), and 2016(e). Right: corresponding prediction results.

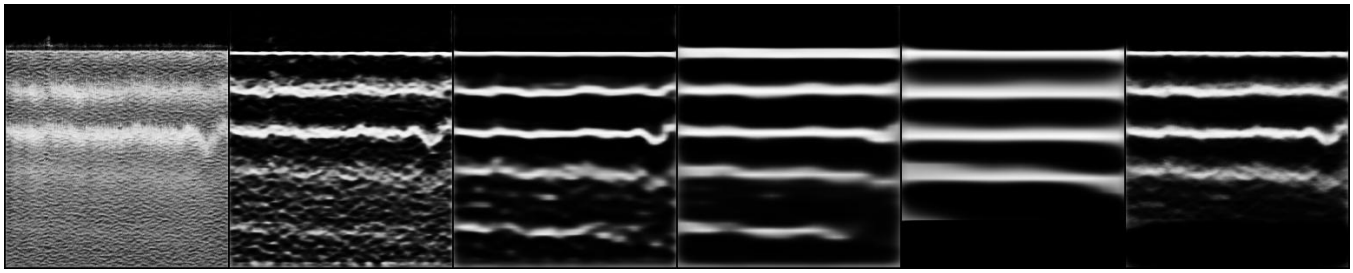

Fig. 3. From left to right: the original images, side-outputs 1-5, and the fusion of side-outputs.

Table 1. Evaluation results for different time intervals

	Test on 2011			Test on 2012			Test on 2016			Test all, no 2012			Test all		
	ODS	OIS	AP	ODS	OIS	AP	ODS	OIS	AP	ODS	OIS	AP	ODS	OIS	AP
Side 1	0.164	0.131	0.087	0.320	0.465	0.261	0.325	0.436	0.270	0.249	0.325	0.182	0.218	0.398	0.167
Side 2	0.703	0.565	0.672	0.763	0.779	0.760	0.710	0.732	0.657	0.703	0.742	0.696	0.722	0.771	0.722
Side 3	0.731	0.659	0.680	0.796	0.824	0.786	0.729	0.746	0.709	0.759	0.799	0.740	0.778	0.819	0.774
Side 4	0.595	0.553	0.511	0.732	0.769	0.645	0.700	0.724	0.594	0.727	0.765	0.637	0.740	0.784	0.647
Side 5	0.549	0.566	0.405	0.512	0.572	0.399	0.574	0.609	0.468	0.547	0.582	0.401	0.507	0.565	0.371
Fuse	0.738	0.662	0.733	**0.815**	**0.854**	**0.815**	0.751	0.766	0.742	0.753	0.803	0.768	0.770	0.833	0.784

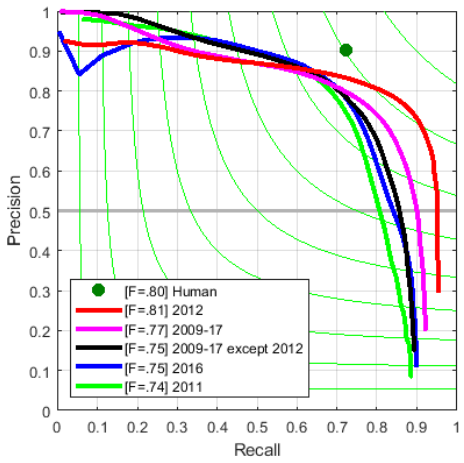

Fig. 4. Precision-recall curves for fusion results of different temporal samples.

our prediction result. The prediction result is the fuse of all side-outputs as described in Section 2. In all three examples in Fig. 2, some fainted layers are also detected, which are not normally detected by humans.

We report three different quantities for evaluating our results; the Optimal Dataset Scale (ODS) or best F-measure on the dataset for a fixed scale, the Optimal Image Scale (OIS) or aggregate F-measure on the dataset for the best scale in each image, and the Average Precision (AP) on the full recall range (equivalently, the area under the precision-recall curve), see [17].

We evaluate the result on datasets of each year. Table 1 contains information for years 2011, 2012, and 2016. The last column in the table shows evaluation results for a test set that contained images from all different years, while the forth column contains the same evaluation metrics without data from 2012. As expected, the best evaluation results pertains to year 2012, then to the test set that contained 2012 data (approximately 10% of the data is from 2012). Table 1 also contains the evaluation metrics for all side-outputs and their fusion. As we should expect, the fusion results are the best results in each column. The Precision-recall curves for fusion results are shown in Fig .4.

4. CONCLUSIONS

There are two main obstacles in applying deep learning methods on complex data, such as radar data. The first and most important obstacle is the lack of labeled data. In our case, we have a semi-automated system to generate labeled data; however, it can generate a limited amount of labeled data, and will not work well when the image is too noisy. The second obstacle is the high frequency of noise in our images. In this case, transfer learning works very poorly. Normally, deep learning models get trained on optical images with low background noise; therefore, they are unstable in the presence of too much noise.

To address the first issue, we used a multi-scale deep learning architecture. That is a more robust approach, as shown in our experiments. To deal with the second issue, we trained the model on the largest available dataset in one year only. We showed that a transfer learning technique works quite well for other years.

5. REFERENCES

[1] C. Sheppard and M. Rahnemoonfar, "Real-time scene understanding for UAV imagery based on deep convolutional neural networks," in *Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International*. IEEE, 2017, pp. 2243–2246.

[2] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik, *Simultaneous Detection and Segmentation*. Cham: Springer International Publishing, 2014, pp. 297–312.

[3] M. Rahnemoonfar and C. Sheppard, "Deep count: fruit counting based on deep simulated learning," *Sensors*, vol. 17, no. 4, p. 905, 2017.

[4] ——, "Real-time yield estimation based on deep learning," in *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping II*, vol. 10218. International Society for Optics and Photonics, 2017, p. 1021809.

[5] M. Rahnemoonfar, D. Dobbs, M. Yari, and M. J. Starek, "Discountnet: Discriminating and counting network for real-time counting and localization of sparse objects in high-resolution UAV imagery," *Remote Sensing*, vol. 11, no. 9, p. 1128, 2019.

[6] M. Rahnemoonfar, M. Robin, M. V. Miguel, D. Dobbs, and A. Adams, "Flooded area detection from UAV images based on densely connected recurrent neural networks," in *Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International*. IEEE, 2018, pp. 3743–3746.

[7] M. Rahnemoonfar and D. Dobbs, "Semantic segmentation of underwater sonar imagery with deep learning," in *Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International*. IEEE, 2019, pp. 9455–9458.

[8] S. Lee, J. Mitchell, D. J. Crandall, and G. C. Fox, "Estimating bedrock and surface layer boundaries and confidence intervals in ice sheet radar imagery using mcmc," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 111–115.

[9] J. E. Mitchell, D. J. Crandall, G. C. Fox, M. Rahnemoonfar, and J. D. Paden, "A semi-automatic approach for estimating bedrock and surface layers from multichannel coherent radar depth sounder imagery," in *SPIE Remote Sensing*. International Society for Optics and Photonics, 2013, Conference Proceedings, pp. 88 921–88 926.

[10] M. Rahnemoonfar, M. Yari, and G. C. Fox, "Automatic polar ice thickness estimation from sar imagery," in *SPIE Defense+ Security*. International Society for Optics and Photonics, 2016, pp. 982 902–982 902.

[11] 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*, 2017.

[12] M. Rahnemoonfar, , A. Abbassi, and F. G. C. Paden, John, "Automatic ice thickness estimation in radar imagery based on charged particle concept," *IEEE International Geoscience and Remote Sensing Symposium*, 2017.

[13] J. A. MacGregor, M. A. Fahnestock, G. A. Catania, J. D. Paden, S. P. Gogineni, S. K. Young, S. C. Rybarski, A. N. Mabrey, B. M. Wagman, and M. Morlighem, "Radiostratigraphy and age structure of the greenland ice sheet," *Journal of Geophysical Research: Earth Surface*, vol. 120, no. 2, pp. 212–241, 2015.

[14] L. S. Koenig, A. Ivanoff, P. M. Alexander, J. A. MacGregor, X. Fettweis, B. Panzer, R. R. Forster, I. Das, J. R. McConnell, M. Tedesco *et al.*, "Annual greenland accumulation rates (2009–2012) from airborne snow radar," *The Cryosphere*, vol. 10, no. 4, 2016.

[15] F. Rodriguez-Morales, D. G.-G. Alvestegui, E. J. Arnold, R. D. Hale, S. Keshmiri, C. J. Leuschen, J. Li, J. D. Paden, and C. Cardenas, "Radar systems for ice and snow measurements onboard manned and unmanned aircraft," *IEEE Latin America Transactions*, vol. 16, no. 9, pp. 2473–2480, 2018.

[16] D. Heaven, "Why deep-learning ais are so easy to fool," pp. 163–166, 2019.

[17] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 5, pp. 898–916, 2011.