PREDICTION OF SEA ICE MOTION WITH RECURRENT NEURAL NETWORKS

Zisis I. Petrou and YingLi Tian

Department of Electrical Engineering, The City College of New York,
The City University of New York, New York, NY 10031, USA
e-mails: {zpetrou,ytian}@ccny.cuny.edu

ABSTRACT

Prediction of sea ice motion is important for ocean-atmosphere interaction modeling and safe naval operations in polar regions. In this study, we investigate the potential of Recurrent Neural Networks (RNNs) in predictions of motion for several days in the future based only on previously observed satellite image data. We collect a large dataset of daily Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E) images that cover the entire Arctic. Optical flow is employed to calculate dense sea ice motion between images of each consecutive-day pair. The optical flow images are then used to train an encoder-decoder Long Short-Term Memory (LSTM) RNN and estimate motion for several days in the future. Experiments demonstrate that the proposed method is successful in predicting short-term sea ice motion with accuracy close to motion calculated from the original images and buoys, and proves promising for further applications and research.

Index Terms— Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E), deep learning, Long Short-Term Memory (LSTM), motion prediction, optical flow

1. INTRODUCTION

Knowledge of sea ice motion in polar and near-polar regions is essential for safe naval operations and modeling of the energy and mass exchange between the atmosphere and the ocean [1, 2]. Thus, timely prediction of future sea ice motion is important for planning safe ship navigation, oil and gas exploration, and fisheries, and may serve as feedback mechanism to improve ocean-atmosphere models.

Despite its importance, the challenging task of future motion prediction is a largely unexplored area. The few presented studies [3, 4] rely mainly on numerical models that require a large amount of data from multiple sources, including surface winds, ice thickness, water currents, ice collision rheology data, etc., with the respective uncertainties of each source. On the contrary, in this study we aim at predicting sea ice motion based only on a sequence of previously observed satellite images. To our knowledge, this is the first attempt to predict sea ice motion solely based on satellite imagery.

Motion prediction remains a challenging task in computer vision. Recent deep learning techniques have been applied to predict future instances based on large datasets of observed past instances. Among them, RNNs have provided promising results in predicting a small number of future natural [5] or ground-based radar echo images [6], based on an observed image sequence. In this study, we propose a methodology for future sea ice motion prediction based on LSTMs and optical flow image sequences. We calculate a sequence of optical flow images from the satellite image data. The optical flow images are encoded in a fixed-size vector representation by an LSTM network, and decoded by a second LSTM network for short-term predictions of optical flow instances for several days in the future. To our best knowledge, it is the first use of deep learning, and in particular, LSTM networks in sea ice motion prediction.

2. DATA

Horizontal polarization brightness temperature images from the Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E) sensor on NASA's Aqua satellite are used as the principal data sources [7]. In particular, the entire set of 277 daily averaging images of 36.5 GHz from 2011 is collected, ranging from January 1st to October 4th, the last day AMSR-E data are provided for. The images are in a polar stereographic grid, tangent to the Earth's surface at 70 degrees northern latitude, and have a spatial resolution of 12.5 km. Each image covers the entire Arctic region, expanding within the following corner coordinates: top-left 30.98°N, 168.35°E; bottom-left 33.92°N, 80.74°W; top-right 31.37°N, 102.34°E; bottom-right 34.35°N, 9.97°W. The original 16-bit images are transformed to 8-bit images for further processing, scaled between the minimum and maximum brightness temperature values observed over the entire dataset. In addition, the sea ice concentration daily average product included within the image dataset is used to generate a daily sea ice mask, by filtering out the indicated open water and land mask pixels.

This work was supported in part by ONR grant N000141310450.

3. METHODOLOGY

3.1. Optical flow

Sea ice motion is calculated for each pair of consecutive-day AMSR-E images. We employ an optical flow approach that has recently proven to outperform a state-of-the-art pattern matching approach to calculate sea ice motion i) densely, i.e., for every pixel of the original image, instead of the sparse calculation of pattern matching, ii) more accurately, and iii) requiring less processing time [8].

First, the sea ice mask image for each day is used to generate AMSR-E images with non-sea-ice pixels filtered out. These new images are used to calculate the optical flow. A structure learning-based edge detection methodology [9] is then applied to each image. In parallel, sparse pixel correspondences are found for each pair of images, using a bottom-up multi-stage non-rigid matching approach [10]. The calculated edges and pixel correspondences are then used to apply sparse-to-dense interpolation and variational energy minimization [11] to calculate optical flow for each pixel between the paired images.

3.2. LSTM encoder-decoder

Our LSTM network consists of an encoder LSTM and two decoder LSTM networks. The composite network follows the structure proposed by Srivastava *et al.* [5]. The encoder network consists of a number of LSTM units that sequentially receive as input optical flow patches from continuous dates and encode them to a fixed-length vector representation. Then, the first decoder network receives this representation as input and reconstructs the original patches, from the most recent to the oldest. Apart from the first reconstructed patch in the time series, where only the encoder representation is used as input, for each reconstruction the output from the previous time step is used as input together with the encoder representation. The second decoder network works in a similar fashion, but instead, it tries to predict a number of future optical flow patches. Again, the previously predicted patch is used as input for the next prediction together with the encoder representation. The LSTM encoder-decoder structure allows an end-to-end training of the entire network in an unsupervised manner, with no requirement for image annotation.

Two-layer encoder and decoder architectures are used, where the output of the first LSTM layer serves as the input to the second layer. Each LSTM unit has a *memory cell* $\boldsymbol{c}_t$ at time t, connected with an *input*, $\boldsymbol{i}_t$, *forget*, $\boldsymbol{f}_t$, and *output*, $\boldsymbol{o}_t$ logistic sigmoidal gate. The input sources at step t are the current optical flow patch—reordering the 2D optical flow patch as 1D vector—or its encoded representation vector, $\boldsymbol{x}_t$, and the output, or *hidden state*, from the previous step, $\boldsymbol{h}_{t-1}$. Together with the previous cell state, these input sources pass through the input and forget gates and update the cell state. The input sources and the updated cell state are then passed through the output gate and update the hidden state of the unit through a tanh non-linearity. The overall formulas describing the calculation of the current hidden state output, $\boldsymbol{h}_t$, are [12]:

$$\boldsymbol{i}_t = \sigma(W_{xi}\boldsymbol{x}_t + W_{hi}\boldsymbol{h}_{t-1} + W_{ci}\boldsymbol{c}_{t-1} + \boldsymbol{b}_i), \quad (1)$$

$$\boldsymbol{f}_t = \sigma(W_{xf}\boldsymbol{x}_t + W_{hf}\boldsymbol{h}_{t-1} + W_{cf}\boldsymbol{c}_{t-1} + \boldsymbol{b}_f), \quad (2)$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t\boldsymbol{c}_{t-1} + \boldsymbol{i}_t \tanh(W_{xc}\boldsymbol{x}_t + W_{hc}\boldsymbol{h}_{t-1} + \boldsymbol{b}_c), \quad (3)$$

$$\boldsymbol{o}_t = \sigma(W_{xo}\boldsymbol{x}_t + W_{ho}\boldsymbol{h}_{t-1} + W_{co}\boldsymbol{c}_{t-1} + \boldsymbol{b}_o), \quad (4)$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \tanh(\boldsymbol{c}_t), \quad (5)$$

where σ represents the sigmoid function, W weight matrices, and $\boldsymbol{b}$ biases. Multiplications between vectors are element-wise. The weight matrices W_{ci}, W_{cf}, and W_{co}, representing the so-called *peephole* connections between the cell and the gates, are diagonal, i.e., each element of a gate vector receives input only from the respective element of the cell vector. The rest of the weight matrices are dense. The cell state of the LSTM unit is capable of summing activities over time. This allows the LSTM network to capture and accurately describe long-term dependencies and avoid the exploding or vanishing gradient problems that traditional RNNs suffer from.

4. RESULTS AND DISCUSSION

4.1. Network parameters

In our settings, since the desired output is optical flow real values, we use linear output units in the LSTM network. The squared loss function is used to evaluate the predictions during training. The weight matrices are initialized by sampling from a uniform distribution and biases are set to zero. We use sequences of 20 optical flow images for training, out of which the 10 first serve as the observed images and the next 10 as the ones to be predicted.

Contrary to [5] that used a training and test set, we additionally employ a validation set to fine-tune the network and prevent overfitting. From the total 276 optical flow images, we randomly select 31 consecutive images for validation and another 31 for testing. This results in 11 overlapping sequences of 20 images each for validation and another 11 sequences for prediction. The original images of width-height-depth dimensions $608\times896\times2$ pixels are split into 532 patches of $32\times32\times2$ pixels. From the validation set, patches with no pixels falling in the sea ice mask area are excluded. In total, 1202 sequences are used for validation and other 5852 sequences for testing, each with patches from 20 sequential images. The remaining images form the training set. To keep the training set totally independent, we exclude all images belonging to either the validation or test set. After removing all patches falling entirely outside the sea ice area, we end up with a training set of 32184 sequences, each with patches from 20 sequential images.

The first and second layers of the encoder-decoder network has 4096 and 2048 hidden units, respectively. This

means that, for a 20-patch sequence, the first layer of the encoder transforms the first 10 patches into a 4096-dimensional vector, the second layer into a 2048-dimensional vector, and the decoder networks follow the reverse order to reconstruct the first 10 patches and predict the next 10 *future* patches. Minibatch training of 100 sequences with rotation is used, i.e., the last minibatch fills the missing sequences from the beginning of the set and the next minibatch continues from there. Validation during training is applied every 500 batch iterations on the validation set. Instead of a fixed number of training iterations as in [5], we apply early-stopping after 10 validation steps without change in the best validation performance. Then, the model with the best validation performance is selected.

4.2. Experimental results

In our experiments the first 10 optical flow patches of each test set sequence are used as input to predict the next 10 patches. First, we compare the predicted optical flow with the reference optical flow patches on a per pixel basis, for both x and y motion axes. Pixels outside the sea ice area are masked out. Fig. 1 draws error statistics for each predicted step in the future, e.g., if patches from days 1–10 are used as inputs, step 1 represents the predicted patches of day 11, step 2 of day 12, etc. After masking out non-sea ice pixels, around 700,000 pixels remain at each prediction step. As expected, the minimum error values appear in the first prediction step, for both motion axes. The errors then, generally, gradually increase as prediction moves further in the future. Although in the last steps the errors slightly decrease for only the x axis, the overall increasing trend remains. However, overall, the performance degrades in a slow pace and even after 10 prediction steps the error values remain as low as double the errors in the first prediction step.

To further evaluate the accuracy of our proposed approach, the predicted motion is compared against buoys from the International Arctic Buoy Programme (IABP) [13]. The 12:00 GMT reported daily buoy positions are selected and the data are reprojected to the polar stereographic grid of the input data. Table 1 reports the corresponding errors for each prediction step in the future. Each step involves between 280 and 287 buoys, i.e., around 26 buoys for each daily image of the test set. It is noted that the corresponding root mean-squared error (RMSE) and mean-absolute error (MAE) of the optical flow data we use as input for the prediction lie around 2.5–3 km and 1.6–2.3 km, respectively. This demonstrates the close proximity of the error of the predicted patches to the theoretically minimum possible error of the input data. In addition, predictions of up to several steps in the future are more accurate even from vectors estimated from observed, instead of predicted, similar AMSR-E 36.5 GHz data (RMSE 4.5 km on the x and 4.83 km on the y axis in [14]).

Fig. 2 draws an example of the step 1 prediction vectors

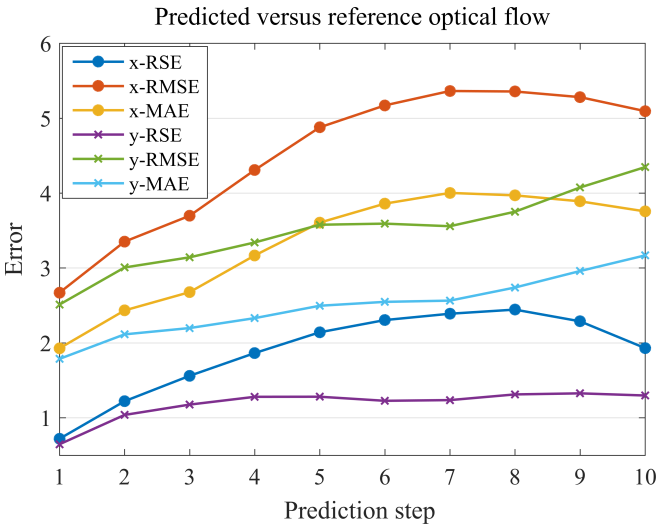

Fig. 1. Accuracy evaluation of the predicted versus the reference optical flow for all sea ice pixels, for each prediction step in the future. The relative squared error (RSE), root mean-squared error (RMSE) in km, and mean-absolute error (MAE) in km are calculated for both x and y axes.

Table 1. Accuracy evaluation of the predicted motion against buoys in the entire Arctic, for each prediction step in the future (fut). RMSE and MAE are in km.

	x axis			y axis		
fut	RSE	RMSE	MAE	RSE	RMSE	MAE
1	0.456	3.607	2.609	0.572	3.167	2.306
2	0.589	3.793	2.777	0.836	3.521	2.589
3	0.589	3.806	2.698	1.202	4.252	2.950
4	0.650	3.944	2.912	1.652	5.096	3.591
5	0.591	3.662	2.662	1.937	5.893	4.305
6	0.595	3.875	2.781	1.995	6.521	4.887
7	0.696	4.211	3.164	2.085	6.747	5.072
8	0.823	4.437	3.351	2.117	6.756	5.110
9	0.808	4.398	3.325	2.012	6.681	5.130
10	0.822	4.632	3.473	1.834	6.611	4.970

compared with the reference optical flow and the IABP buoys. For visualization, a close-up look of the original image extent is shown, vectors are drawn every 56 and 38 pixels in the vertical and horizontal dimension, respectively, and they are magnified by 20 times to be clearly visible (see scale on the bottom right corner of the figure). Although in several cases the predicted motion vectors underestimate the true magnitude of the buoys or the orientation of the reference optical flow vectors, they correlate overall well with both. In total, as the aforementioned quantitative results also suggest, the proposed LSTM and optical flow approach provides satisfactory

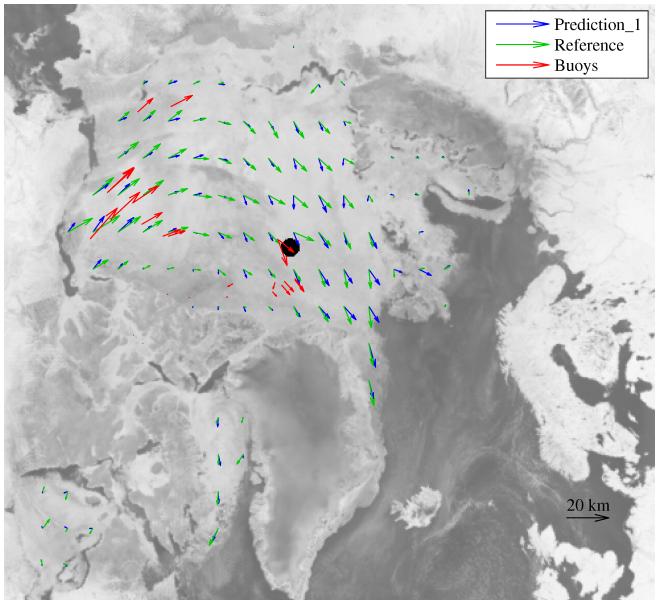

Fig. 2. Close-up look of one-step predicted motion vectors, reference optical flow and buoy data, for the day pair of May 20–21, 2011. All vectors are magnified by 20 times for visualization. Background: AMSR-E image from May 20, 2011.

first predictions of future sea ice motion and a promising direction for future research.

5. CONCLUSION

We have proposed an approach to predict sea ice motion with a LSTM deep learning network and optical flow input, using only satellite AMSR-E source data. The approach provided satisfactory results for predictions as far as 10 steps in the future. It is the first attempt to use LSTMs for sea ice motion prediction and the results are promising for future research and extension to further motion prediction tasks.

6. REFERENCES

[1] T. Kræmer, H. Johnsen, and C. Brekke, "Emulating Sentinel-1 Doppler radial ice drift measurements using Envisat ASAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6407–6418, Dec. 2015.

[2] F. Girard-Ardhuin and R. Ezraty, "Enhanced arctic sea ice drift estimation merging radiometer and scatterometer data," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2639–2648, July 2012.

[3] L. W. A. De Silva, H. Yamaguchi, and J. Ono, "Ice–ocean coupled computations for sea-ice prediction to support ice navigation in Arctic sea routes," *Polar Res.*, vol. 34, article no. 25008, Nov. 2015.

[4] A. J. Schweiger and J. Zhang, "Accuracy of short-term sea ice drift forecasts using a coupled ice-ocean model," *J. Geophys. Res. Ocean.*, vol. 120, no. 12, pp. 7827–7841, Dec. 2015.

[5] N. Srivastava, E. Mansimov, and R. Salakhudinov, "Unsupervised learning of video representations using LSTMs," in *Proc. 32nd Int. Conf. Machine Learning*, 2015, pp. 843–852.

[6] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W. Wong, and W. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems 28*, 2015, pp. 802–810.

[7] D. J. Cavalieri, T. Markus, and J. C. Comiso, "AMSR-E/Aqua Daily L3 12.5 km Brightness Temperature, Sea Ice Concentration, & Snow Depth Polar Grids, Version 3," Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center, 2014. doi: http://dx.doi.org/10.5067/AMSR-E/AE_SI12.003. Accessed: Dec. 16, 2016.

[8] Z. I. Petrou and Y. Tian, "High-resolution sea ice motion estimation with optical flow using satellite spectroradiometer data," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1339–1350, Mar. 2017.

[9] P. Dollár and C. L. Zitnick, "Fast edge detection using structured forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1558–1570, Aug. 2015.

[10] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "DeepFlow: Large displacement optical flow with deep matching," in *IEEE Int. Conf. Computer Vision*, 2013, pp. 1385–1392.

[11] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, "EpicFlow: Edge-preserving interpolation of correspondences for optical flow," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2015, pp. 1164–1172.

[12] A. Graves, "Generating sequences with recurrent neural networks," 2013, [Online]. Available: arXiv:1308.0850.

[13] M. Tschudi, C. Fowler, J. Maslanik, J. S. Stewart, and W. Meier, "Polar Pathfinder daily 25 km EASE-Grid sea ice motion vectors, version 3, Buoys," Boulder, Colorado USA: National Snow and Ice Data Center, 2016. doi: http://dx.doi.org/10.5067/O57VAIT2AYYY. Accessed: Dec. 20, 2016.

[14] W. N. Meier and M. Dai, "High-resolution sea-ice motions from AMSR-E imagery," *Ann. Glaciol.*, vol. 44, pp. 352–356, Nov. 2006.