

Prediction of Sea Ice Motion With Convolutional Long Short-Term Memory Networks

Zisis I. Petrou[✉], Member, IEEE, and Yingli Tian[✉], Fellow, IEEE

Abstract—Prediction of sea ice motion is important for safeguarding human activities in polar regions, such as ship navigation, fisheries, and oil and gas exploration, as well as for climate and ocean-atmosphere interaction models. Numerical prediction models used for sea ice motion prediction often require a large number of data from diverse sources with varying uncertainties. In this paper, a deep learning approach is proposed to predict sea ice motion for several days in the future, given only a series of past motion observations. The proposed approach consists of an encoder-decoder network with convolutional long short-term memory (LSTM) units. Optical flow is calculated from satellite passive microwave and scatterometer daily images covering the entire Arctic and used in the network. The network proves able to learn long-time dependencies within the motion time series, whereas its convolutional structure effectively captures spatial correlations among neighboring motion vectors. The approach is unsupervised and end-to-end trainable, requiring no manual annotation. Experiments demonstrate that the proposed approach is effective in predicting sea ice motion of up to 10 days in the future, outperforming previous deep learning networks and being a promising alternative or complementary approach to resource-demanding numerical prediction methods.

Index Terms—Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E), Advanced Scatterometer (ASCAT), AMSR2, Arctic sea ice, convLSTM, deep neural networks, drift prediction, optical flow, recurrent neural networks (RNNs)

I. INTRODUCTION

SEA ice motion significantly affects the thickness distribution of ice in polar regions, causing leads or ridging. It has a critical role in the advection of sea ice out of the Arctic region, whereas it overall influences the ice mass balance and fluxes between the ocean and the atmosphere [1]–[4]. It has been used for the initialization, assimilation, and validation of climate models simulating sea ice dynamics [5]–[9]. On a local level, sea ice motion affects human activities in the Arctic region. Following the continuous decline of sea ice extent

Manuscript received August 5, 2018; revised December 18, 2018 and March 5, 2019; accepted March 21, 2019. Date of publication April 29, 2019; date of current version August 27, 2019. This work was supported in part by the Office of Naval Research (ONR) under Grant N000141310450 and in part by NSF under Grant IIS-1400802. (*Corresponding author: Yingli Tian*.)

Z. I. Petrou is with the Department of Electrical Engineering, The City College, The City University of New York, New York, NY 10031 USA (e-mail: zpetrou@ccny.cuny.edu).

Y. Tian is with the Department of Electrical Engineering, The City College, The City University of New York, New York, NY 10031 USA, and also with the Department of Computer Science, The Graduate Center, The City University of New York, New York, NY 10016 USA (e-mail: ytian@ccny.cuny.edu).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2019.2909057

during the last decades and the projections for ice-free periods in the future [10]–[12], the potential for more extensive usage of the region for shipping, fisheries, oil and gas exploration, and tourism has been raised [13]. In order to safeguard human activities and enhance climate modeling, accurate short-term predictions of sea ice motion for several days in the future are increasingly important.

Predictions of sea ice motion, together with sea ice extent and related variables, have almost exclusively been attempted using numerical prediction models, including fully coupled atmosphere-ocean global climate models, statistical regression-based approaches, and heuristic methods [14]–[19]. A number of models have been designed to address seasonal, interannual, and decadal predictions [13], [20], whereas others focus on short-term forecasting [21], with the latter being crucial for safe maritime operations. To generate the desired outputs, these models usually rely on a large number of sources of information, including surface winds, ice thickness, water currents, ice collision rheology data, or previous predictions. For a specific time frame or area of interest, the data sources can be limited [22], thus, reducing the applicability of the models. In addition, the model predictions are associated with several sources of uncertainty, including model initial conditions and uncertainties in the input data, e.g., ice thickness data [23], [24].

More specifically, one of the main drivers of sea ice motion, especially in central Arctic, rather than the coastlines, is wind [25], [26]. This is why weather prediction models have been extensively used to predict sea ice motion. Thus, wind prediction limitations of these models consequently limit the accuracy of sea ice motion prediction. In addition, given that the average sea ice drift speed has increased during the past decades to a much larger degree than wind speed [25], [27], such predictions become more challenging. Additional factors that affect sea ice motion are ocean currents, forcing due to the Coriolis force and density-driven gravitational forcing, and ice cover and internal ice strength, with the latter ones acting as resisting motion factors.

In this paper, we propose a short-term prediction approach based on single-source data time series. In particular, we employ sea ice motion data calculated from sequences of daily passive microwave and scatterometer satellite images and predict motion for several days in the future. Our proposed approach is based on deep neural networks, specifically convolutional long short-term memory (LSTM) networks [28], and offers several advantages compared with the aforementioned

numerical model prediction paradigm: 1) it relies on a single source of data, and thus, it is prone to single-source uncertainties; 2) the data are available on a daily basis and cover the entire Arctic; 3) the proposed approach is completely unsupervised and automated, i.e., the proposed network is trained end-to-end without requiring any human annotation of the data; and 4) although the methodology explicitly ignores the aforementioned factors that affect sea ice motion prediction (wind, ocean currents, sea ice cover), these factors can be implicitly modeled through the calculations of spatial and temporal correlations of sea ice motion. In addition, contrary to the only previous deep learning approach for sea ice motion prediction [29], our proposed convolutional LSTM approach can explicitly capture spatial relationships among neighboring motion vectors.

Our study is organized as follows. Section II discusses the previous work on motion prediction. Section III presents the data used in the study and Section IV details our proposed method. Experimental results are drawn and discussed in Section V. Conclusions are given in Section VI.

II. RELATED WORK

Recurrent neural networks (RNNs) have been used in a variety of remote sensing applications, including image classification [30], [31]; automatic clustering of multispectral images [32]; spatiotemporal prediction of snow depth and distribution driven by meteorological data [33]; retrieval of areal extent of snow cover [34]; rainfall nowcasting from multisatellite passive-sensor images [35]; and estimation of soil moisture dynamics [36]. In particular, LSTM networks have been additionally employed in image classification [37]; image captioning [38]; land cover change detection [39]; semantic understanding of high-resolution images [40]; and supervised texture segmentation [41]. Contrary to the aforementioned applications, LSTM networks are employed in this study to encode spatiotemporal information, more specifically sea ice motion, from 2-D time series and predict future time-series in the 2-D space for several days in the future.

Prediction of motion has been an increasingly studied field. A number of approaches have focused in predicting motion trajectories of image pixels or objects in the 2-D space [42]–[44]. Despite the similarities, the objective of this paper is different, as we do not aim at predicting the trajectory of specific sea ice floes, but instead at predicting the motion in the dense image field, i.e., for every pixel, for several time instances. Prediction of traffic or crowd flow is another relevant field to our study where deep learning techniques have been proposed recently [45]–[49]. Ma *et al.* [46] proposed a 2-D image representation of spatiotemporal traffic and a convolutional neural network (CNN) approach for short-term traffic prediction. Similar to [49], space is considered as one dimension, with spatial correlations captured in 1-D space, and, thus, different from the 2-D desired motion in this study. Closer to the problem formulation in this study, Yu *et al.* [45] used CNN to generate 2-D spatial traffic features and fully connected LSTM for prediction. Zhang *et al.* [47], [48] leveraged the existence of motion temporal patterns and proposed the fusion

of temporal closeness, period, and seasonal trend to capture temporal dependencies for short-term crowd flow prediction. The method could be promising for sea ice motion prediction under the assumption of seasonality patterns of the motion, which have not been observed on our available single-year data. Furthermore, several approaches have employed static images to predict future image frames [50]–[52]. Such approaches are not appropriate in our study, since single-time images, or even single-time sea ice motion, convey limited information on future sea ice motion and any prediction would be mostly arbitrary.

A number of approaches based on image time series as input for future prediction have been proposed recently, mainly in the field of unsupervised video prediction [28], [29], [53]–[61]. Srivastava *et al.* [59] proposed an encoder–decoder architecture based on LSTM to predict a number of future images from an input sequence of images. A variation of the network using a predicted frame as additional input for the next prediction provided the best results. Shi *et al.* [28] introduced the idea of convolutional LSTM (convLSTM) networks that, besides the ability to encode and decode temporal information, were able to capture spatial relationships between neighboring pixels for precipitation nowcasting. The idea of convolutional LSTM models has further been employed in prediction studies [53], [55], [56], [58]. A recent study used optical flow, i.e., dense 2-D sea ice motion, directly as input to an encoder–decoder LSTM network to predict sea ice motion for several days in the future [29]. However, the input in this approach was converted to a 1-D vector to be handled by the fully connected LSTM networks, thus ignoring spatial relationships between neighboring pixels.

In this study, we propose a deep convolutional LSTM approach for sea ice motion prediction. A major difference with aforementioned frame prediction approaches is that we do not try to predict future images and use these images to estimate motion. Instead, we use optical flow time series as input to the convolutional LSTM network and directly predict motion for several days in the future. Contrary to [29], the proposed convolutional structure is able to capture spatial relationships among neighboring pixels, whose motion is highly correlated.

III. DATA

The principal data used to train the prediction model are a large number of Level-3 brightness temperature images from the Advanced Microwave Scanning Radiometer–Earth Observing System (AMSR-E) sensor on NASA’s Aqua satellite [62]. The spatial resolution of the images is 12.5 km. In total, a number of 276 daily images, from January 1 to October 3, 2011, are collected, most of them used for the model training. The images are of horizontal polarization at 36.5 GHz and cover the entire Arctic (Fig. 1), with the coordinates of their corners being: top left 30.98°N, 168.35°E; bottom left 33.92°N, 80.74°W; top right 31.37°N, 102.34°E; and bottom right 34.35°N, 9.97°W.

A variety of data are used to evaluate the prediction model, coming from different sensors, properties, and peri-

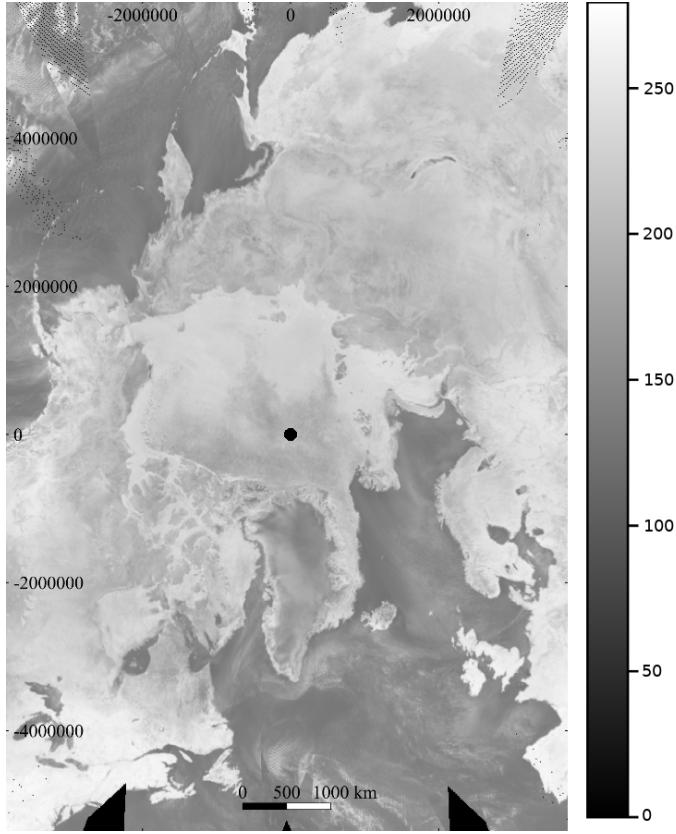


Fig. 1. AMSR-E image from January 1, 2011, with brightness temperature colorbar (in K). The image is projected on the North Polar Stereographic Projection [65].

ods of the year. First, a random sample from the aforementioned AMSR-E images is used for testing. In addition, images from the AMSR2 onboard the Global Change Observation Mission–Water (GCOM-W) satellite are used. AMSR2 data from both 36- and 89-GHz bands are collected from the winter (February 21, 2018–March 24, 2018) and summer (August 21, 2018–September 30, 2018) seasons [63]. In particular, Level-3 horizontal polarization brightness temperature images of 10-km spatial resolution are collected. The images are downsampled to 12.5-km resolution to have the same extent and size to the AMSR-E images. In order to evaluate the performance of the approach beyond passive microwave data, images from the Advanced Scatterometer (ASCAT) onboard the EUMETSAT METOP satellite are also collected for testing, both from the winter (February 19, 2017–March 30, 2017) and the summer (August 28, 2017–September 27, 2017) seasons [64]. In particular, daily multiorbit incidence angle normalized sigma-0 (sigma-0 at 40° incidence angle) Arctic images processed by the multivariate Scatterometer Image Reconstruction algorithm with filtering (SIRF) are used. The images are downsampled from 4.45 to 11.20 km, so their width matches the width of the AMSR-E and AMSR2 images.

All images are projected in the National Snow and Ice Data Center (NSIDC) Polar Stereographic North grid, with 70°N latitude of true scale and the origin longitude at 45°E [65]. The images are scaled between the overall minimum and

maximum values of the entire set and converted to 8-bit images for memory and processing efficiency. The open water and land are masked out from each image, based on daily masks generated by or distributed with the above products.

As a note, a main reason for the selection of the data sources above for this study, besides the previous successful use in sea ice motion estimation studies and the wide spatial coverage, is the availability of daily data. This temporal resolution is necessary to allow daily predictions. Thus, such large swath width radiometer and scatterometer data are more appropriate for the requirements of this study than smaller swath width data sources, such as Synthetic Aperture Data (SAR). Although SAR has proven accurate in providing high-resolution sea ice motion estimation, the daily availability of images covering the entire Arctic is currently limited [66].

Note that the images and their corresponding sea ice masks used in the experiments in this paper, for the AMSR-E, AMSR2, and ASCAT sensors are available at DOI: 10.21227/f87y-by42 with the title of “Data for Prediction of Sea Ice Motion with Convolutional Long Short-Term Memory Networks.”

IV. METHODS

A. Optical Flow Estimation

Optical flow is calculated for each pair of masked images of consecutive days for AMSR-E, AMSR2, and ASCAT data, respectively. Given a pair of images, the output of the calculation is a 3-D tensor with the two axes representing the axes of the original 2-D images in pixels, and the third axis including the latitudinal and longitudinal motion of sea ice for each pixel from the first to the second image. As a note, the terms “latitudinal” and “longitudinal” motion are used in the paper to indicate motions in the vertical and horizontal axis of the reprojected images, respectively. Despite the fact that, due to the selected map projection, the terms do not align with the geographic latitude and longitude coordinates (e.g., positive longitudinal motion in the upper part of the projected image describes motion toward decreasing geographic longitudes), the terms are selected to highlight the georeferenced nature of the analyzed motion. The optical flow method proposed in [67] is adopted, proven more accurate than traditionally employed pattern matching in sea ice motion estimation and able to provide dense calculation in the continuous (subpixel) space [66].

The calculated optical flow data are grouped in overlapping sequences including $T + P$ consecutive-day flow instances, where the first T flow instances represent the input to the prediction model and the last P ones the desired output.

B. Conditioned Convolution LSTM Network

1) *Convolutional LSTM Unit*: The proposed network consists of connected convolutional LSTM units that have the typical structure of LSTM with peephole connections [28]. The structure of an LSTM unit is shown in Fig. 2.

Each LSTM unit consists of a *memory cell*, C , that is connected with an *input*, i , *forget*, f , and *output*, o , logistic sigmoidal gates. This structure allows the memory cell to

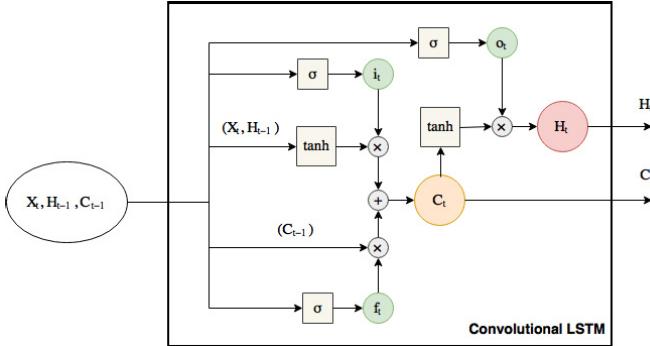


Fig. 2. Representation of an LSTM unit. The inputs of the unit at time t are shown in the ellipse outside the box. Arrows with input variables in parenthesis denote that only the specific input(s) is(are) passed. The outputs of the LSTM unit are the updated cell, C_t , and hidden, H_t , states.

sum activities over time and capture long-term properties, counteracting exploding and vanishing gradient problems that limit vanilla RNN networks. At time t , the unit gets as input the t th optical flow of a sequence, X_t , and the output, or *hidden state*, from the previous step, H_{t-1} . In the case of the encoder, X_t comes from the input optical flow sequence, whereas in the case of the decoder from the predicted sequence. The two input sources, X_t and H_{t-1} , pass through the input and forget gates and update the cell state. The updated cell state, C_t , passes together with X_t and H_{t-1} through the output gate and update the hidden state of the unit through a tanh nonlinearity. All H_t , i_t , f_t , and o_t are 3-D arrays whose last two dimensions match the width and height dimensions of X_t . The following formulas describe the overall LSTM connections:

$$\begin{aligned} i_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \odot C_{t-1} + b_i) \\ f_t &= \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \odot C_{t-1} + b_f) \\ C_t &= f_t \odot C_{t-1} + i_t \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \\ o_t &= \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \odot C_{t-1} + b_o) \\ H_t &= o_t \odot \tanh(C_t), \end{aligned}$$

where σ represents the sigmoid function, W the weight matrices, and b the biases. The peephole connection weight matrices W_{ci} , W_{cf} , and W_{co} are diagonal, whereas the rest are dense. Convolutions are denoted by $*$, whereas \odot stands for the Hadamard product.

2) *LSTM Network*: The proposed prediction network has an encoder-decoder structure with convolutional LSTM units (Fig. 3). For a specific sequence, the input daily flows, X , are given sequentially as input to the LSTM encoder and transformed into a compact representation. The encoded representation is then fed to the decoder network that predicts the next daily optical flows. The output is fed to a forward convolutional layer that reconstructs it to the same feature dimensions as the input. During the training of the network, the predicted flows, \hat{Y}_t , $j = T+1, \dots, T+P$, are compared with the corresponding calculated flows for the same days, X_t , $t = T+1, \dots, T+P$. Thus, the entire network is end-to-end trainable in an unsupervised fashion requiring no manual annotation.

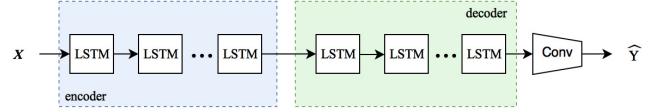


Fig. 3. Overview of the structure of the proposed convolutional LSTM prediction network. X represents the optical flow input sequence, “LSTM” and “Conv” stand for convolutional LSTM and forward convolutional layers, respectively, and \hat{Y} represents the predicted flow array.

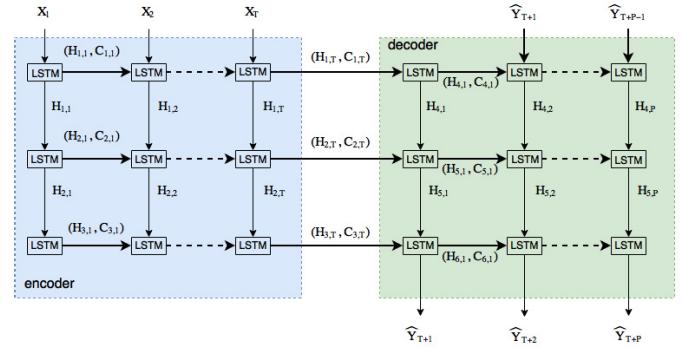


Fig. 4. Breakdown of the encoder-decoder structure of the proposed prediction network. X_i represents the i th optical flow array of the input sequence X , \hat{Y}_j represents the j th predicted flow array; $C_{l,t}$ and $H_{l,t}$ stand for the cell and hidden states of a convolutional LSTM unit in layer l and time t .

Fig. 4 draws in more detail the interactions among the network components. As depicted, the last states and cell outputs of the encoder are copied to the decoder network. The encoder and decoder can consist of multiple layers of LSTM units stacked together, with the output, H , of one layer being the input of the next one. Contrary to the original convolutional LSTM model [28], our decoder network is conditioned on the previous output [59]. This structure allows the decoder network to use as additional input for each predicted flow, \hat{Y}_t , $t = T+1, \dots, T+P$, the previous prediction, \hat{Y}_{t-1} , with the exception of the first prediction, \hat{Y}_{T+1} , where the encoded information is the only input to the decoder (Fig. 4). During training, the observed optical flow in time $t-1$ is used as input, whereas during validation and testing, the predicted flow is used instead. This modification is supported by the notion that sea ice motion between consecutive days usually does not change dramatically as far as magnitude and direction are concerned, at least at the 12.5-km spatial and the daily temporal scale of the imagery. There can be regular direction and speed changes on subkilometer and subdaily scales due to tides, inertial effects, and ice leads and ridging [68], [69], but their effects are averaged out on the scales of analysis of this study. Thus, prior knowledge of the predicted motion at time $t-1$ is expected to constrain and enhance the prediction at time t . Furthermore, in [28], the outputs from all decoder LSTM layers are concatenated and fed to the 1×1 convolutional layer to produce the final prediction. On the contrary, in our network, only the output of the last LSTM layer is fed to the final 1×1 convolutional layer. The notion of this simpler connection network is to keep the role of each layer distinct in the propagation of information between layers, so network optimization is more focused and targeted

during training. In addition, the number of network parameters is also decreased in this final layer, preventing overfitting.

3) Optimization and Loss: Backpropagation through time (BPTT) and Root Mean Square Propagation (RMSProp) [70] are used for minibatch training of our network. During training, having an optical flow sequence $\{X_1, \dots, X_T\}$ as input, the predicted sequence $\{\hat{Y}_{T+1}, \dots, \hat{Y}_{T+P}\}$ is compared with the actual sequence $\{Y_{T+1}, \dots, Y_{T+P}\}$. Instead of the commonly used squared loss [28], [29], [59], we additionally employ l_2 gradient difference loss [57] to penalize changes in differences between neighboring pixels beyond the expected ones. Our final loss function is

$$\begin{aligned} L(\hat{Y}, Y) &= a \sum_{i,j,k} |Y_{i,j,k} - \hat{Y}_{i,j,k}|^2 \\ &\quad + b \sum_{i,j,k} ||Y_{i,j,k} - Y_{i-1,j,k}| - |\hat{Y}_{i,j,k} - \hat{Y}_{i-1,j,k}||^2 \\ &\quad + b \sum_{i,j,k} ||Y_{i,j,k} - Y_{i,j-1,k}| - |\hat{Y}_{i,j,k} - \hat{Y}_{i,j-1,k}||^2, \end{aligned} \quad (1)$$

where i , j , and k are width, height, and depth, pixel indices, respectively. The first summation term represents the squared loss, whereas the last two terms the gradient loss. The weights $a \in \mathbb{R}$ and $b \in \mathbb{R}$ define the relative contribution of the two terms in the overall loss.

V. RESULTS AND DISCUSSION

A. Experimental Settings

The daily optical flow arrays form sequences of 20 consecutive-day arrays. In each sequence, the first ten arrays are used as input ($T = 10$) and the last ten as the ground-truth prediction output ($P = 10$). Data from AMSR-E are used to train the prediction model. A number of temporally overlapping sequences are randomly selected to form the validation set during the network training. Another set of temporally overlapping sequences are randomly selected as testing set, under the condition of having no common dates with the validation set. The remaining sequences are used as the training set, with no common dates with the validation and testing sets. The sequences are split in 32×32 spatially nonoverlapping patches over the longitude and latitude image dimensions, forming sequences of 20 such patches. For training efficiency, the sequences with no sea ice pixels are removed from the training and validation sets. However, since we aim at simulating real-life conditions and make the algorithm agnostic of the underlying surface, we keep training and validation sequences that include nonmoving (e.g., land, land-fast ice) pixels together with moving (sea ice) pixels, so that the algorithm can learn zero-motion patterns as well. The testing sets consist of sequences covering the entire Arctic region, including both moving and nonmoving regions.

Following the above process, the training and validation sets include 32 087 and 1208 sequences of $20 \times 32 \times 32$ optical flow patches from AMSR-E data, respectively. The testing data include: 1) 5852 sequences of patches from AMSR-E; 2) 13 718 and 14 440 sequences from ASCAT winter and

summer data, respectively; and 3) 20 216 sequences from each of AMSR2 36- and 89-GHz winter data type and 21 280 sequences from each summer data type.

We use a three-layer encoder and a three-layer decoder network. The number of hidden states of the convolutional LSTM layers are 128, 64, and 32, respectively, whereas the state-to-kernel size is 5×5 . We use minibatch training of 16 sequences, a learning rate of 0.001, and a decay rate of 0.9 for RMSProp optimization. The size of LSTM layers and the learning and decay rates are in line with related studies [28], [55], [58]; we experiment with varying values without noticing significant changes in performance. The batch size is selected based on experimentation and capabilities of the available graphics processing units (GPUs). As widely employed, the convolutional LSTM layer weights, W , were initialized from a uniform distribution in $[-1, 1]$, whereas the initial biases, b , were set to zero. We select the l_2 gradient loss function described in (1), which outperformed l_2 , l_1 , and l_1 gradient loss functions in our experiments. The weights of the loss function in (1), a and b , are set to 1.

We compare our conditioned convolutional LSTM network with the fully connected conditioned LSTM network by Petrou and Tian [29] and the convLSTM network by Shi *et al.* [28]. For the former, we employ a two-layer encoder–autoencoder–decoder architecture with 4096 and 2048 hidden states, uniform initialization of the weight matrices, and biases set to 1, based on recommendations in [71]. The two convLSTM versions have the same parameters as our network, one with the originally proposed 3×3 state-to-kernel size and one with 5×5 . Based on the performance in the validation set, the same number of 50 epochs is selected to train all networks.

B. Comparison With Reference Optical Flow

The predicted optical flow from our proposed network for 10 days in the future is compared with the reference optical flow for the same days. Thus, the average difference of the predicted 2-D motion vector at each pixel with the reference optical flow at the same pixels is calculated.

1) AMSR-E Data: Table I shows the mean absolute error (MAE) and root-mean-squared error (RMSE) between the vectors for the latitude (x) and longitude (y) axes for 10 prediction steps, as well as the overall average performance.

As expected, the best predictions—highlighted in bold—are made for the first prediction step, i.e., the first day after the observed optical flow. The performance of the prediction starts degrading as predictions move further into the future. In the y -axis, the worst performance is observed at the tenth prediction step, for the x -axis at the eighth step, whereas, overall, the errors are larger the later the predictions are into the future. The MAE errors follow a quasi-linear increase as the predictions move into the future. It is noteworthy that the worst error predictions are less than by 50% worse than the best early predictions. This indicates that prediction holds a similar, not quickly decreasing performance after several prediction steps in the future.

2) Winter ASCAT and AMSR2 Data: Fig. 5 draws the MAE errors for the ASCAT and AMSR2 data during the dry winter

TABLE I

COMPARISON OF THE PREDICTED OPTICAL FLOW WITH THE REFERENCE OPTICAL FLOW VECTORS FOR THE AMSR-E DATA, FOR EACH PREDICTION STEP IN THE FUTURE (“FUT”). COLUMN “VECTORS” REPORTS THE TOTAL NUMBER OF VECTORS COMPARED. MAE AND RMSE ARE REPORTED FOR THE LATITUDINAL (x) AND LONGITUDINAL (y) MOTION AXES, BOTH IN KILOMETERS. THE AVERAGE ERRORS FROM ALL STEPS ARE ALSO REPORTED

fut	vectors	x-MAE	x-RMSE	y-MAE	y-RMSE
1	807626	2.34	3.25	2.11	3.07
2	803467	2.65	3.69	2.32	3.44
3	799080	2.78	3.84	2.18	3.23
4	794545	3.00	4.10	2.30	3.35
5	790216	3.28	4.46	2.50	3.62
6	786609	3.37	4.60	2.68	3.84
7	782872	3.40	4.62	2.76	3.93
8	779041	3.44	4.63	2.86	4.05
9	774893	3.39	4.56	3.03	4.24
10	770026	3.28	4.40	3.09	4.33
Avg.	788837.5	3.09	4.21	2.58	3.71

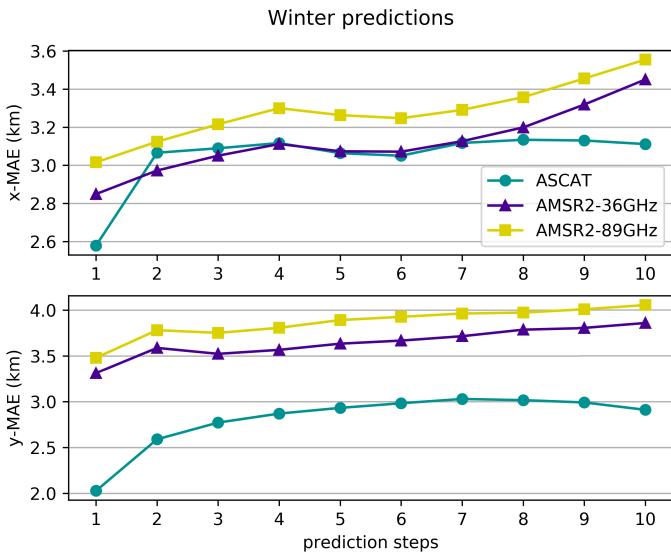


Fig. 5. Comparison of the predicted optical flow with the reference optical flow vectors, for each prediction step in the future. MAE errors of the ASCAT and AMSR2 winter data are reported for the latitudinal (x) and longitudinal (y) motion axes.

season for the ten prediction steps. Masking out nonsea ice pixels, about 1.5 million and 1.1 million motion vectors are included per prediction step. The predictions with ASCAT data follow more closely the observed optical flow vectors than the predictions from AMSR2 data for all prediction steps, as far as the combined x - and y -axis errors are concerned. The predictions from 36-GHz AMSR2 data are consistently better than the 89-GHz AMSR2 data, but following a highly correlated pattern per prediction step. For all sources of data, the predictions get less accurate as they move further in the future. However, similar to the AMSR-E data, the prediction drop from the first to the tenth precision step is only about 20% or lower, for any of the ASCAT and AMSR2 data, demonstrating that the performance of the model remains stable after several days in the future. It is also noteworthy

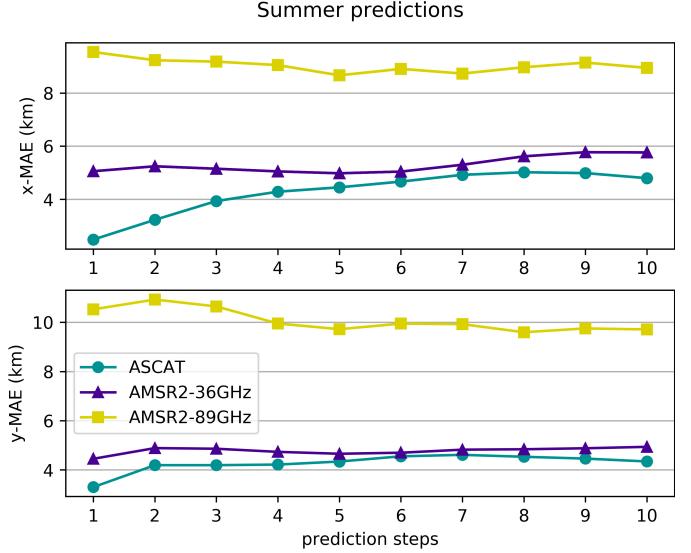


Fig. 6. Comparison of the predicted optical flow with the reference optical flow vectors, for each prediction step in the future. MAE errors of the ASCAT and AMSR2 summer data are reported for the latitudinal (x) and longitudinal (y) motion axes.

that predictions, mostly from the ASCAT data, have similar accuracy with the predictions from AMSR-E data reported in Table I. This demonstrates that our model is transferable to data sources different than the ones used to train it, and highlights the importance of using optical flow data rather than sensor-specific raw image values, such as brightness temperature.

3) *Summer ASCAT and AMSR2 Data:* The MAE errors for the ASCAT and AMSR2 data during the melt summer season are drawn in Fig. 6 for the ten prediction steps. After masking out nonsea ice pixels, each prediction step considers about 320 000 and 565 000 vectors from ASCAT and AMSR2 data, respectively. As in the dry season, the predictions with ASCAT data are consistently more accurate than AMSR2 data. For most prediction steps, the predictions with 36-GHz AMSR2 data are not significantly worse than with ASCAT data. On the contrary, 89-GHz AMSR2 errors are almost double as high as 36-GHz AMSR2 for most prediction steps. As reported in previous studies [72], [73], 89-GHz microwave data are more affected by atmospheric and surface properties than 36 GHz and lead to less accurate estimation of sea ice motion. Such atmospheric conditions, more intense during the summer melt season, have affected the accuracy of the calculated optical flow, used as input and reference data. Thus, the input data from 89 GHz can include inconsistent and erroneous motion patterns that the model has not been trained to identify, and thus, lead to predictions less consistent with the calculated reference flow.

4) *Marginal Ice Zone and Canadian Archipelago:* Besides the regions of thick ice, mainly in the center of Arctic and around the North Pole, the area of study includes several regions of thinner ice, where sea ice motion is usually more intense and may be more challenging to predict. In addition to the evaluation for the entire Arctic presented in Sections V-B2 and V-B3, we measure the performance of the algorithm

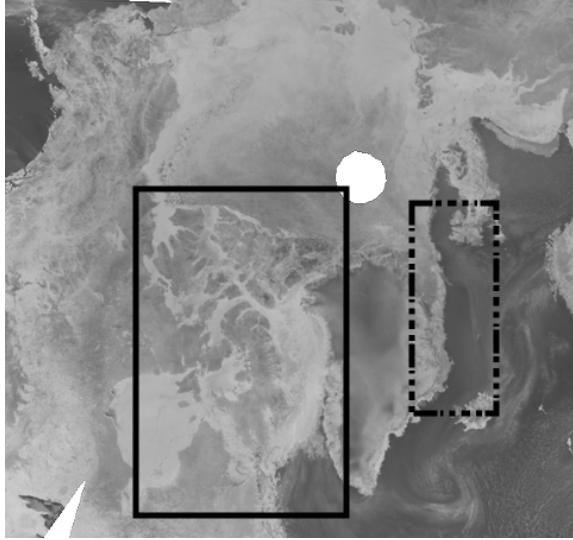


Fig. 7. Canadian Archipelago (solid line) and marginal ice zone (dashed line) regions selected for study, on zoomed-in view of the 36-GHz AMSR2 image from March 1, 2018.

TABLE II

COMPARISON OF THE PREDICTED OPTICAL FLOW WITH THE REFERENCE OPTICAL FLOW VECTORS, IN THE CANADIAN ARCHIPELAGO FOR ONE AND 10 PREDICTION STEPS IN THE FUTURE (“FUT”). COLUMN “VECTORS” REPORTS THE AVERAGE NUMBER OF VECTORS COMPARED PER PREDICTION STEP. MAE IS REPORTED FOR THE LATITUDINAL (x) AND LONGITUDINAL (y) MOTION AXES. THE AVERAGE ERRORS FROM ALL STEPS ARE ALSO REPORTED. “W,” “S,” “36,” AND “89” INDICATE WINTER, SUMMER, 36-, AND 89-GHz DATA, RESPECTIVELY

source	vectors	x-MAE (km)		y-MAE (km)	
		fut 1	fut 10	fut 1	fut 10
Canadian Archipelago					
ASCAT-w	327709	1.31	1.89	1.50	2.07
AMSR2-36-w	317075	1.75	2.33	1.83	2.56
AMSR2-89-w	317075	1.88	2.51	1.88	2.96
ASCAT-s	63021	1.21	3.73	1.23	1.80
AMSR2-36-s	187304	4.12	3.99	3.26	3.27
AMSR2-89-s	187304	9.65	7.52	10.46	8.04
Marginal ice zone					
ASCAT-w	68911	2.01	3.16	2.76	4.38
AMSR2-36-w	54819	2.49	4.15	2.64	4.29
AMSR2-89-w	54819	2.60	4.27	2.76	4.18
ASCAT-s	6449	1.91	3.09	2.79	1.30
AMSR2-36-s	29459	5.40	5.70	6.93	7.37
AMSR2-89-s	29459	5.31	5.19	6.20	6.23

separately in two such areas of high operational interest, the Canadian Archipelago and a marginal ice zone east of Greenland (Fig. 7).

Table II reports the MAE errors of the predicted versus the observed reference flow with the ASCAT and AMSR2 data, for both the winter and summer seasons. For space economy, only the first and tenth predictions are reported, in most cases representing the lower and upper error limits among the ten predictions. The errors for the two regions are very similar with the errors reported for the entire Arctic (Figs. 5 and 6). Our proposed model is able to predict the motion in these challenging regions with similar accuracy as the entire study

TABLE III

COMPARISON OF THE PROPOSED METHOD WITH THE CONDITIONED LSTM APPROACH BY [29] AND THE CONVOLUTIONAL LSTM APPROACH BY [28], THE LATTER WITH 3×3 AND 5×5 STATE-TO-KERNEL SIZES, ON THE AMSR-E DATA. THE ACCURACY OF COPYING THE LAST INPUT FLOW IS ALSO REPORTED AS “BASELINE.” THE AVERAGE MAE AND RMSEs, BOTH IN KILOMETERS, FOR TEN PREDICTION STEPS ARE SHOWN, FOR THE LATITUDINAL (x) AND LONGITUDINAL (y) MOTION AXES

Method	x-MAE	x-RMSE	y-MAE	y-RMSE
baseline	3.61	4.94	3.10	4.43
[29]	3.36	4.53	2.68	3.71
[28] [3×3]	3.19	4.35	2.81	4.01
[28] [5×5]	3.12	4.24	2.63	3.72
Ours	3.09	4.21	2.58	3.71

area. As observed in the previous experiments, the ASCAT data provide more accurate predictions than the AMSR2 data. Similarly, the predictions are more close to the reference data during the dry winter season than the melt summer season.

C. Comparison With Previous Methods

We compare our proposed network with previous multistep prediction approaches, as detailed in Section V-A, on the AMSR-E data. In addition, the performance of simply copying the last observed input optical flow is considered as baseline. Table III presents the average errors for 10 prediction steps in the future. The conditioned LSTM network proposed by Petrou and Tian [29] performs similarly with the convLSTM network with 3×3 state-to-kernel size by Shi *et al.* [28]. Increasing the state-to-kernel size to 5×5 captures motion correlations in a larger area and proves beneficial for the performance of the network. Our proposed approach achieves lower errors than both methods, as highlighted in bold. It proves able to capture spatial correlations of motion with the convolutional LSTM layers, whereas conditioning each prediction to the previous one, acts as a priori—unsupervised—knowledge that is beneficial for the next predictions.

D. Comparison With Buoys

Following an unsupervised end-to-end training, our model is trained to minimize the loss against the reference optical flow values. These reference optical flow values, calculated from the corresponding observed AMSR-E image pairs, are considered as a ground-truth during the model training. However, the optical flow calculations with observed data entail errors as evidenced when compared with more accurate validation sources [67]. To get a further estimate on the accuracy of the model predictions, the predicted optical flow vectors are additionally evaluated against buoys from the International Arctic Buoy Program (IABP) [74]. The buoys have estimated position accuracy around 0.5 km/day [72], [74], providing position accuracy at a much finer scale than the reference optical flow estimated from the coarse 12.5 km resolution AMSR-E images. The daily buoy positions at 12:00 GMT are collected to calculate their motion within consecutive

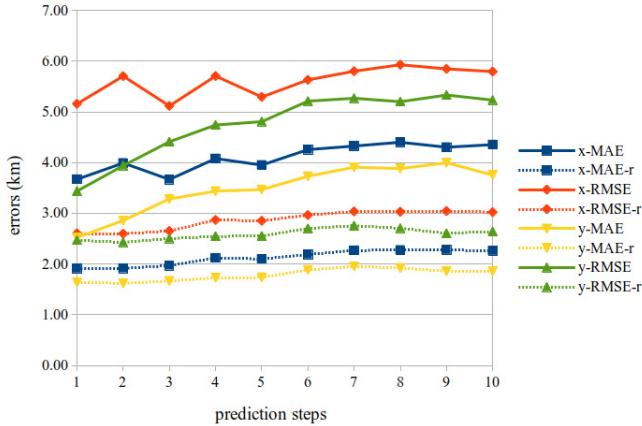


Fig. 8. Comparison of the predicted optical flow and the reference optical flow vectors from the AMSR-E data against buoys, for each prediction step in the future. MAE and RMSE are reported for the latitudinal (x) and longitudinal (y) motion axes. The reference data errors are indicated by the prefix “-r.”

days, and projected to the NSIDC Polar Stereographic North grid [65].

Fig. 8 shows the performance of the proposed approach against the buoy motion, on the AMSR-E data, for every prediction step. The corresponding errors of the reference optical flows are provided as well, i.e., the errors calculating the optical flow from observed images on the same days, rather than predicting it. As expected, the reference flow errors remain almost stable as the prediction steps change, since they are independent from the prediction process. On the contrary, the errors of the predictions, overall, increase as the prediction moves further to the future. Some individual high error values in the first prediction steps on the x -axis may be considered as statistical outliers, justified by the fact that the model is not trained to minimize the errors against the buoys, but against the reference flow data.

However, the plots can still offer useful indications on the predictive power of the proposed network. We calculate the relative difference between the corresponding predicted and reference types of error (e.g., the MAE on the x -axis for the predicted and reference data), as $(e_{a,p} - er_{a,p})/e_{a,p}$, where $e_{a,p}$ is the average prediction error for axis a , $a = \{x, y\}$ and type p , $p = \{\text{MAE}, \text{RMSE}\}$, for the 10 prediction steps, and $er_{a,p,t}$ the corresponding reference error. This relative error lies between 0.84 and 0.95 for the different error measures and axes, indicating that the prediction errors are less than double the reference optical flow error. In addition, it is worth noting that several predicted motions are as accurate as estimations of motion from observed images from previous studies. In particular, pattern matching on similar AMSR-E 36.5-GHz data estimated motion with RMSE 4.5 km on the x -axis and 4.83 km on the y -axis [72]. As observed on Fig. 8, predictions of up to five steps on the y axis have lower error than the motion estimated with observed data from the previous approach. This demonstrates, on the one hand, that the selected optical flow approach offers a solid basis for the training of our network compared with pattern matching approaches, and on the other hand, that our prediction method

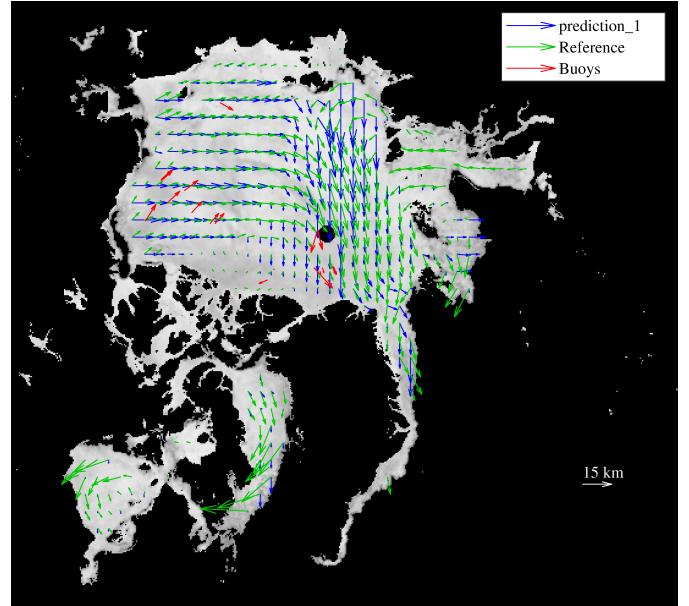


Fig. 9. Focus on predicted sea ice motion for one day in the future with the AMSR-E data, for the dates between May 24 and 25, 2011. The reference optical flow calculated from observed images and buoy motion are also drawn.

offers promising results for motion prediction, even without using any buoy information for training. Assimilation of such data in the training process is a candidate future pathway to further increase our method’s prediction accuracy.

Fig. 9 focuses on the sea ice motions within the Arctic Ocean. The predicted motion for one step in the future is drawn, together with the reference optical flow calculated from the corresponding observed AMSR-E images and the buoy motion. Although the optical flow is predicted for each pixel of the image, only one-fourteenth and one-ninth of the vectors are drawn for clearer illustration for the latitude and longitude axis, respectively. The scale of the motion is also magnified. As observed, the predicted motion follows closely the reference one. For the largest part of the image, the predicted and reference motion vectors are highly correlated both in magnitude and direction. It is noteworthy that in certain areas such as the Beaufort Sea and Chukchi Sea on the left part of the depicted area, the predicted motion does not agree well with the ground truth buoy motion, since no buoy information was used during the network training. Instead, the predictions follow closely the reference flow, which demonstrates the success of our method to extrapolate observations into the future, but also some of its limitations associated with its limited training sources.

E. Image Prediction

We additionally evaluate our network ability to predict future AMSR-E images, given a sequence of past images as input. The depth dimension of the network is changed to 1, since single-band greyscale images are the input and desired output in this case. The rest of the network parameters remain the same. Fig. 10 shows examples of AMSR-E image prediction for one, six, and 10 steps in the future. In particular,

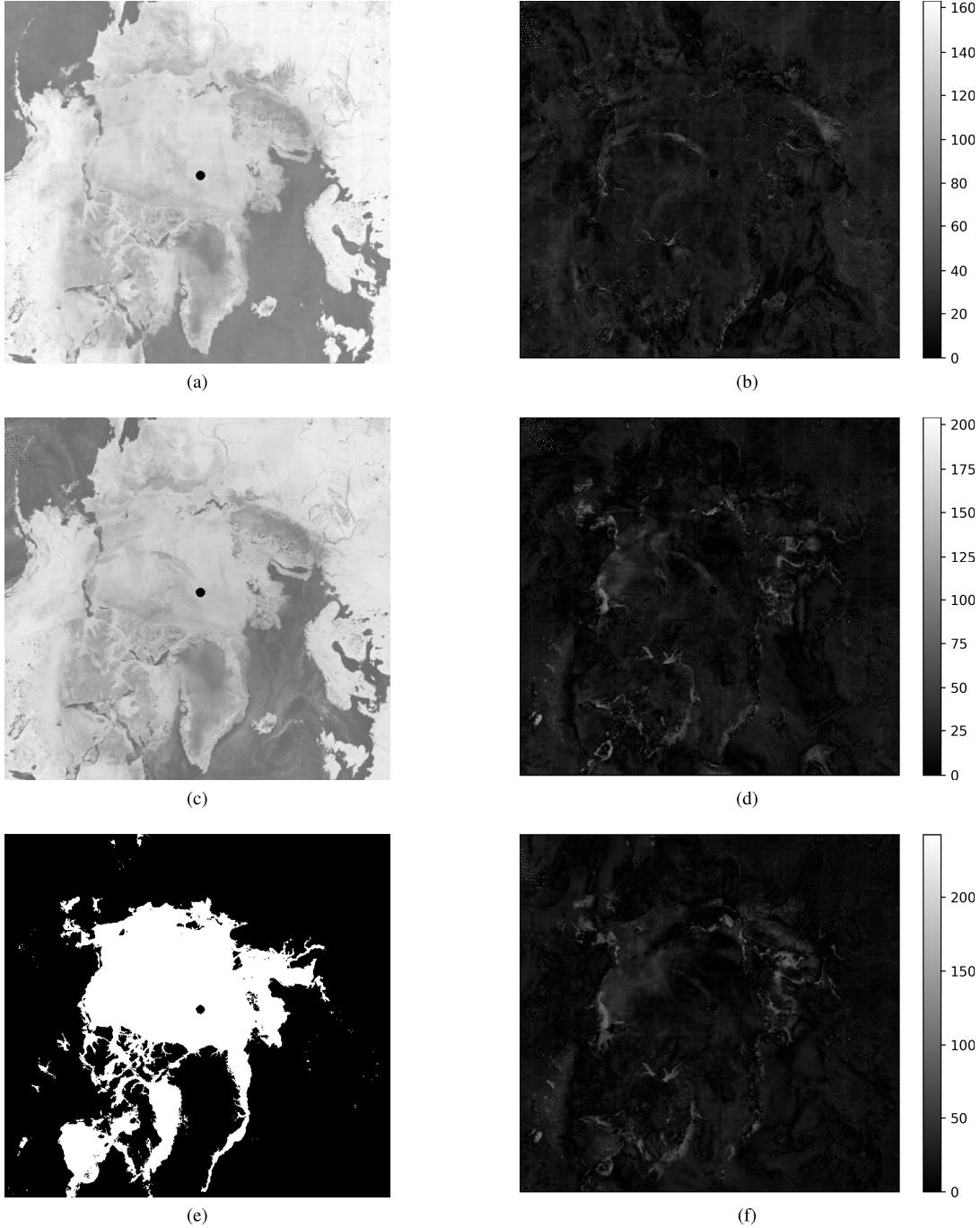


Fig. 10. Focus on (a) predicted AMSR-E image for one step in the future and (b) original one from May 20, 2011, together with (e) corresponding sea ice mask. Images on the right show the absolute difference between the original and predicted AMSR-E images for (b) 1, (d) 6, and (f) 10 days in the future, for the dates May 20, 25, and 29, 2011.

a focused view of the predicted image for one step in the future together with the original image are shown in Fig. 10(a) and (c), respectively. The sea ice mask image is also drawn to highlight the areas mostly used during the model training. It is observed that the proposed network is able to predict most of the structure elements of the image, in the sea ice, open ocean, and land areas. Some patch edge effects can be observed in the predicted images, coming from the 32×32 pixel predicted patches that were mosaicked to form

the final images. In addition, the predicted image appears more blur than the original one, with several sharp details useful to extract sea ice motion being smoothed out. However, the network proves able to predict the overall structure of the area, even in the parts of no sea ice that were masked out during the training process.

Fig. 10(b), (d), and (f) shows the absolute differences between predicted and observed images for one, six, and ten steps in the future, for the dates of May 20, 25, and 29, 2011.

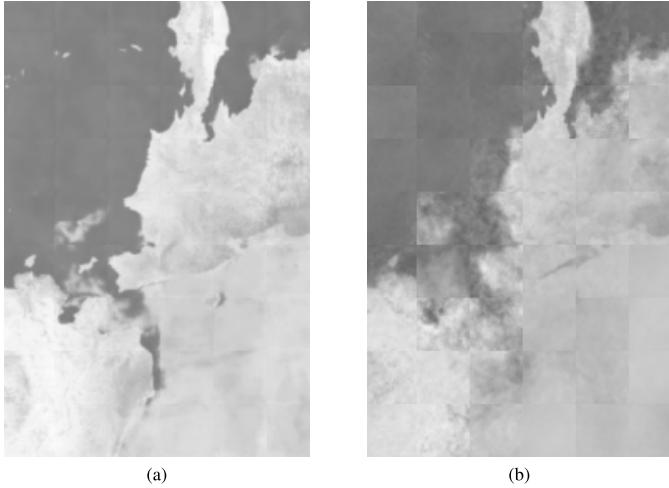


Fig. 11. Detail of predicted AMSR-E image by the proposed convolutional LSTM network (left) and the fully connected LSTM network by [29] (right) for 1 day in the future. The depicted detail is from the top-left part of the original image, including the Chukchi Sea and the Bering Strait, from May 20, 2011.

The overall predictions for all steps in the future are relatively small, as indicated by the dark pixels in the images. As expected, the first-step predictions are the closest to the observed image. As predictions move further in the future, the differences with the observed images slightly increase. However, most differences appear in areas with nonsea ice pixels, which were masked out during model training. Areas in the central Arctic region, where the main sea ice layer exists as drawn in Fig. 10(e), show small differences even for ten-step predictions.

Fig. 11 shows a detail of the predicted image of May 20, 2011, for 1 day in the future by the proposed convolutional LSTM network and the fully connected network in [29]. The improvement in accuracy and the ability to preserve edges brought by the convolutional LSTM network is observed. Besides these advantages, though, the proposed network appears also more robust to patch effects created by splitting the image (and optical flow) data into sequences of 32×32 pixels during the training and prediction processes. These effects clearly affect the quality of the image predicted by the fully connected LSTM network, where the edges around each patch are obvious, creating a tile effect. This effect can hardly be observed in the predicted image by the proposed network. It is noteworthy that the patch sequences are fed to the algorithm during training in a random order, thus spatial relationships among adjacent patches are not explicitly defined. However, the ability of the proposed network to predict images with indistinct patch edges is one more qualitative piece of evidence that the network not only preserves spatial relationships among neighboring pixels, but it provides accurate predictions in absolute terms that allow smooth transitions among adjacent patches.

VI. CONCLUSION

The convolutional LSTM approach proposed in this study proved able to predict sea ice motion for several days in the future. Optical flow calculated from pairs of AMSR-E

images was the only input to the unsupervised end-to-end trainable system requiring no manual annotation. Conditioning each prediction to the previous one, i.e., using as input at each prediction step the previous prediction, took advantage of the temporal correlations that exist in sea ice motion between consecutive days. Besides AMSR-E data, the model was able to provide accurate predictions on data from other sources, AMSR2 and ASCAT, which had not been used for training. Our proposed approach outperformed previous LSTM-based multistep prediction methods, and although being trained to minimize only reference flow data of sparse resolution, it showed promising results when evaluated against with precision buoy vector data. The approach is, thus, a promising single-source data-driven alternative or complementary approach to complex numerical methods requiring multisource data.

REFERENCES

- [1] S. E. L. Howell *et al.*, “Comparing L- and C-band synthetic aperture radar estimates of sea ice motion over different ice regimes,” *Remote Sens. Environ.*, vol. 204, pp. 380–391, Jan. 2018.
- [2] W. N. Meier, “Losing Arctic sea ice: Observations of the recent decline and the long-term context,” in *Sea Ice*, D. N. Thomas, Ed., 3rd ed. Chichester, U.K.: Wiley, 2017, ch. 11, pp. 290–303.
- [3] C. Haas, “Sea ice thickness distribution,” in *Sea Ice*, D. N. Thomas, Ed., 3rd ed. Chichester, U.K.: Wiley, 2017, ch. 2, pp. 42–64.
- [4] A. Gettelman and R. B. Rood, *Demystifying Climate Models: A Users Guide to Earth System Models*. Berlin, Germany: Springer, 2016, ch. 6, pp. 87–108.
- [5] 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.
- [6] A. Berg, L. Axell, and L. E. B. Eriksson, “Comparison between SAR derived sea ice displacement and hindcasts by the operational ocean model HIROMB,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2013, pp. 3630–3633.
- [7] N. Kimura, A. Nishimura, Y. Tanaka, and H. Yamaguchi, “Influence of winter sea-ice motion on summer ice cover in the Arctic,” *Polar Res.*, vol. 32, no. 1, Nov. 2013, Art. no. 20193.
- [8] M. L. Van Woert, C. Zou, W. N. Meier, P. D. Hovey, R. H. Preller, and P. G. Posey, “Forecast verification of the polar ice prediction system (PIPS) sea ice concentration fields,” *J. Atmos. Ocean. Technol.*, vol. 21, no. 6, pp. 944–957, Jun. 2004.
- [9] W. N. Meier, J. A. Maslanik, and C. W. Fowler, “Error analysis and assimilation of remotely sensed ice motion within an Arctic sea ice model,” *J. Geophys. Res.*, vol. 105, pp. 3339–3356, Feb. 2000.
- [10] M. C. Serreze and J. Stroeve, “Arctic sea ice trends, variability and implications for seasonal ice forecasting,” *Philos. Trans. Roy. Soc. London A, Math. Phys. Sci.*, vol. 373, Jul. 2015, Art. no. 20140159.
- [11] J. C. Stroeve *et al.*, “Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations,” *Geophys. Res. Lett.*, vol. 39, no. 16, Aug. 2012, Art. no. L16502.
- [12] M. Wang and J. E. Overland, “A sea ice free summer Arctic within 30 years: An update from CMIP5 models,” *Geophys. Res. Lett.*, vol. 39, no. 18, 2012, Art. no. L18501.
- [13] E. Hawkins, S. Tietsche, J. J. Day, N. Melia, K. Haines, and S. Keeley, “Aspects of designing and evaluating seasonal-to-interannual Arctic sea-ice prediction systems,” *Quat. J. Roy. Meteorol. Soc.*, vol. 142, no. 695, pp. 672–683, 2016.
- [14] A. Dirkson, W. J. Merryfield, and A. Monahan, “Impacts of sea ice thickness initialization on seasonal Arctic sea ice predictions,” *J. Climate*, vol. 30, no. 3, pp. 1001–1017, 2017.
- [15] G. C. Smith *et al.*, “Sea ice forecast verification in the Canadian Global Ice Ocean Prediction System,” *Quant. J. Roy. Meteorol. Soc.*, vol. 142, no. 695, pp. 659–671, 2016.
- [16] J.-F. Lemieux *et al.*, “The Regional Ice Prediction System (RIPS): Verification of forecast sea ice concentration,” *Quant. J. Roy. Meteorol. Soc.*, vol. 142, pp. 632–643, Jan. 2016.
- [17] 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.

- [18] P. Sakov, F. Counillon, L. Bertino, K. A. Lisæter, P. R. Oke, and A. Koralev, "TOPAZ4: An ocean-sea ice data assimilation system for the North Atlantic and Arctic," *Ocean Sci.*, vol. 8, no. 4, pp. 633–656, Aug. 2012.
- [19] R. H. Preller, P. G. Posey, W. Maslowski, D. Stark, and T. T. C. Pham, "Navy sea ice prediction systems," *Oceanography*, vol. 15, no. 1, pp. 44–56, 2002.
- [20] V. Guemas *et al.*, "A review on Arctic sea-ice predictability and prediction on seasonal to decadal time-scales," *Quant. J. Roy. Meteor. Soc.*, vol. 142, no. 695, pp. 546–561, Jan. 2016.
- [21] D. A. Hebert *et al.*, "Short-term sea ice forecasting: An assessment of ice concentration and ice drift forecasts using the U.S. Navy's Arctic Cap Nowcast/Forecast System," *J. Geophys. Res. Ocean.*, vol. 120, no. 12, pp. 8327–8345, Dec. 2015.
- [22] F. Kauker *et al.*, "Seasonal sea ice predictions for the Arctic based on assimilation of remotely sensed observations," *Cryosphere Discuss.*, vol. 9, pp. 5521–5554, Oct. 2015.
- [23] 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, Nov. 2015, Art. no. 25008.
- [24] R. Msadek, G. A. Vecchi, M. Winton, and R. G. Gudgel, "Importance of initial conditions in seasonal predictions of Arctic sea ice extent," *Geophys. Res. Lett.*, vol. 41, no. 14, pp. 5208–5215, Jul. 2014.
- [25] G. Spreen, R. Kwok, and D. Menemenlis, "Trends in Arctic sea ice drift and role of wind forcing: 1992–2009," *Geophys. Res. Lett.*, vol. 38, no. 19, Oct. 2011, Art. no. L19501.
- [26] M. Steele, J. Zhang, D. Rothrock, and H. Stern, "The force balance of sea ice in a numerical model of the Arctic Ocean," *J. Geophys. Res. Oceans*, vol. 102, no. C9, pp. 21061–21079, Sep. 1997.
- [27] P. Rampal, J. Weiss, and D. Marsan, "Positive trend in the mean speed and deformation rate of Arctic sea ice, 1979–2007," *J. Geophys. Res.*, vol. 114, no. C5, May 2009, Art. no. C05013.
- [28] 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 *Proc. 28th Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.
- [29] Z. I. Petrou and Y. Tian, "Prediction of sea ice motion with recurrent neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Fort Worth, TX, USA, Jul. 2017, pp. 5422–5425.
- [30] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017.
- [31] E. Maggiori, G. Charpiat, Y. Tarabalka, and P. Alliez, "Recurrent neural networks to correct satellite image classification maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 4962–4971, Sep. 2017.
- [32] P. Koprinkova-Hristova, K. Alexiev, D. Borisova, G. Jelev, and V. Atanassov, "Recurrent neural networks for automatic clustering of multispectral satellite images," *Proc. SPIE*, vol. 8892, pp. 88920X-1–88920X-8, Oct. 2013.
- [33] T. Sauter, B. Weitzenkamp, and C. Schneider, "Spatio-temporal prediction of snow cover in the Black Forest mountain range using remote sensing and a recurrent neural network," *Int. J. Climatol.*, vol. 30, no. 15, pp. 2330–2341, Dec. 2010.
- [34] J. J. Simpson and T. J. McIntire, "A recurrent neural network classifier for improved retrievals of areal extent of snow cover," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 10, pp. 2135–2147, Oct. 2001.
- [35] F. S. Marzano, G. Rivolta, E. Coppola, B. Tomassetti, and M. Verdecchia, "Rainfall nowcasting from multisatellite passive-sensor images using a recurrent neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3800–3812, Nov. 2007.
- [36] Y. C. Tzeng, K. T. Fan, C. Y. Lin, Y. J. Lee, and K. S. Chen, "Estimation of soil moisture dynamics using a recurrent dynamic learning neural network," in *Proc. IEEE Int. Geosci. Remote Sens. Congr.*, Munich, Germany, Jul. 2012, pp. 1251–1253.
- [37] J. Geng, H. Wang, J. Fan, and X. Ma, "SAR image classification via deep recurrent encoding neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2255–2269, Apr. 2018.
- [38] X. Lu, B. Wang, X. Zheng, and X. Li, "Exploring models and data for remote sensing image caption generation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2183–2195, Apr. 2018.
- [39] H. Lyu, H. Lu, and L. Mou, "Learning a transferable change rule from a recurrent neural network for land cover change detection," *Remote Sens.*, vol. 8, no. 6, Jun. 2016, Art. no. 506.
- [40] B. Qu, X. Li, D. Tao, and X. Lu, "Deep semantic understanding of high resolution remote sensing image," in *Proc. Int. Conf. Comput. Inf. Telecommun. Syst.*, Kunming, China, Jul. 2016, pp. 1–5.
- [41] W. Byeon and T. M. Breuel, "Supervised texture segmentation using 2D LSTM networks," in *Proc. IEEE Int. Conf. Image Process.*, Paris, France, Oct. 2014, pp. 4373–4377.
- [42] J. Jaipuria, G. Habibi, and J. P. How, (2018). "A transferable pedestrian motion prediction model for intersections with different geometries." [Online]. Available: <https://arxiv.org/abs/1806.09444>
- [43] L. Ballan, F. Castaldo, A. Alahi, F. Palmieri, and S. Savarese, "Knowledge transfer for scene-specific motion prediction," in *Proc. Eur. Conf. Comput. Vis.*, vol. 1, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Amsterdam, The Netherlands: Springer, 2016, pp. 697–713.
- [44] M. Monfort, A. Liu, and B. D. Ziebart, "Intent prediction and trajectory forecasting via predictive inverse linear-quadratic regulation," in *Proc. AAAI Conf. Artif. Intell.*, 2015, pp. 3672–3678.
- [45] H. Yu, Z. Wu, S. Wang, Y. Wang, and X. Ma, "Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks," *Sensors*, vol. 17, no. 7, 2017, Art. no. 1501.
- [46] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, 2017, Art. no. 818.
- [47] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 1655–1661.
- [48] J. Zhang, Y. Zheng, D. Qi, R. Li, and X. Yi, "DNN-based prediction model for spatio-temporal data," in *Proc. 24th ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, 2016, Art. no. 92.
- [49] Y. Wu and H. Tan, (2016). "Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework." [Online]. Available: <https://arxiv.org/abs/1612.01022>
- [50] J. Walker, C. Doersch, A. Gupta, and M. Hebert, "An uncertain future: Forecasting from static images using variational autoencoders," in *Proc. Eur. Conf. Comput. Vis.*, vol. 7, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Amsterdam, The Netherlands: Springer, 2016, pp. 835–851.
- [51] J. Walker, A. Gupta, and M. Hebert, "Dense optical flow prediction from a static image," in *Proc. IEEE Int. Conf. Comput. Vis.*, Santiago, Chile, Jun. 2015, pp. 2443–2451.
- [52] S. L. Pintea, J. C. van Gemert, and A. W. M. Smeulders, "Déjà Vu: Motion prediction in static images," in *Proc. Eur. Conf. Comput. Vis.*, vol. 3, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Zürich, Switzerland: Springer, 2014, pp. 172–187.
- [53] A. X. Lee, R. Zhang, F. Ebert, P. Abbeel, C. Finn, and S. Levine, (2018). "Stochastic adversarial video prediction." [Online]. Available: <https://arxiv.org/abs/1804.01523>
- [54] M. Babaeizadeh, C. Finn, D. Erhan, R. H. Campbell, and S. Levine, "Stochastic variational video prediction," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–15.
- [55] W. Lotter, G. Kreiman, and D. Cox, "Deep predictive coding networks for video prediction and unsupervised learning," in *Proc. Int. Conf. Learn. Represent.*, Toulon, France, 2017, pp. 1–18.
- [56] R. Villegas, J. Yang, S. Hong, X. Lin, and H. Lee, "Decomposing motion and content for natural video sequence prediction," in *Proc. Int. Conf. Learn. Represent.*, Toulon, France, 2017, pp. 1–22.
- [57] M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," in *Proc. Int. Conf. Learn. Represent.*, San Juan, Puerto Rico, 2016, pp. 1–14.
- [58] V. Patraucean, A. Handa, and R. Cipolla, "Spatio-temporal video autoencoder with differentiable memory," in *Proc. Int. Conf. Learn. Represent.*, San Juan, Puerto Rico, 2016, pp. 1–13.
- [59] N. Srivastava, E. Mansimov, and R. Salakhudinov, "Unsupervised learning of video representations using LSTMs," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 843–852.
- [60] M. Ranzato, A. Szlam, J. Bruna, M. Mathieu, R. Collobert, and S. Chopra, (2014). "Video (language) modeling: A baseline for generative models of natural videos." [Online]. Available: <https://arxiv.org/abs/1412.6604>
- [61] V. Michalski, R. Memisevic, and K. Konda, "Modeling deep temporal dependencies with recurrent ‘grammar cells’," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, Inc., 2014, pp. 1925–1933.
- [62] 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 V003," NASA Nat. Snow Ice Data Center Distrib. Active Arch. Center, Boulder, CO, USA,

- Tech. Rep., 2014. Accessed: Mar. 4, 2019. [Online]. Available: https://nsidc.org/data/ae_si12/versions/3, doi: [10.5067/AMSR-E/AE_SI12.003](https://doi.org/10.5067/AMSR-E/AE_SI12.003).
- [63] JAXA. (2017). *Globe Portal System*. Accessed: Mar. 4, 2019. [Online]. Available: <https://gportal.jaxa.jp/gpr/index/index>
- [64] Brigham Young University. (2018). *ASCAT Enhanced Resolution Image Products (Beta Version)*. Accessed: Mar. 4, 2019. [Online]. Available: http://www.scp.byu.edu/data/Ascat/SIR/Ascat_sir.html
- [65] National Snow and Ice Data Center. (2016). *Documentation: Polar Stereographic Projection and Grid*. Accessed: Mar. 4, 2019. [Online]. Available: http://nsidc.org/data/polar-stereo/ps_grids.html
- [66] Z. I. Petrou, Y. Xian, and Y. Tian, "Towards breaking the spatial resolution barriers: An optical flow and super-resolution approach for sea ice motion estimation," *ISPRS J. Photogramm.*, vol. 138, pp. 164–175, Apr. 2018.
- [67] 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, 2017.
- [68] J. K. Hutchings, P. Heil, A. Steer, and W. D. Hibler, "Subsynoptic scale spatial variability of sea ice deformation in the western Weddell Sea during early summer," *J. Geophys. Res. Oceans*, vol. 117, no. C1, Jan. 2012, Art. no. C01002.
- [69] J. K. Hutchings, A. Roberts, C. A. Geiger, and J. Richter-Menge, "Spatial and temporal characterization of sea-ice deformation," *Ann. Glaciol.*, vol. 52, no. 57, pp. 360–368, 2011.
- [70] T. Tielemans and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *Coursera, Neural Netw. Mach. Learn.*, vol. 4, no. 2, pp. 26–31, 2012.
- [71] R. Jozefowicz, W. Zaremba, and I. Sutskever, "An empirical exploration of recurrent network architectures," *J. Mach. Learn. Res.*, vol. 37, pp. 2342–2350, Jul. 2015.
- [72] W. N. Meier and M. Dai, "High-resolution sea-ice motions from AMSR-E imagery," *Ann. Glaciol.*, vol. 44, pp. 352–356, Nov. 2006.
- [73] T. Lavergne, S. Eastwood, Z. Teffah, H. Schyberg, and L. A. Breivik, "Sea ice motion from low-resolution satellite sensors: An alternative method and its validation in the Arctic," *J. Geophys. Res.*, vol. 115, no. C10, Oct. 2010, Art. no. C10032.
- [74] 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," Nat. Snow Ice Data Center, Boulder, CO, USA, Tech. Rep., 2016. Accessed: Mar. 4, 2019. [Online]. Available: <https://nsidc.org/data/nsidc-0116/versions/3>, doi: [10.5067/O57VAIT2AYYY](https://doi.org/10.5067/O57VAIT2AYYY).



Zisis I. Petrou (S'05–M'15) received the Diploma degree in electrical and computer engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2007, the M.Sc. degree in space studies from the International Space University, Strasbourg, France, in 2009, and the Ph.D. degree in remote sensing from the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K., in 2015.

From 2009 to 2016, he was a Research Assistant with the Center for Research and Technology Hellas, Thessaloniki, Greece. From 2016 to 2017, he was a Post-Doctoral Research Scholar with The City College of New York, The City University of New York, New York, NY, USA. He currently works as a Software Engineer with Foursquare Labs, New York, NY, USA. His research interests include classification, motion detection and prediction analysis on remote sensing images using advanced machine learning, image processing, and computer vision techniques.

Dr. Petrou was the recipient of the Best Paper Award for Young Professionals under 35 at the 2nd International Conference on Space Technology, Athens, in 2011, sponsored by IEEE. He has been a Scholar of the European Space Agency (ESA) and the Onassis Foundation, Greece.



Yingli Tian (M'99–SM'01–F'18) received the B.S. and M.S. degrees from Tianjin University, Tianjin, China, in 1987 and 1990, respectively, and the Ph.D. degree from Chinese University of Hong Kong, Hong Kong, in 1996.

She held a faculty position with the National Laboratory of Pattern Recognition, Chinese Academy of Sciences, Beijing, China. In 1998, she joined Carnegie Mellon University Pittsburgh, PA, USA, where she was a Post-Doctoral Fellow with the Robotics Institute. From 2001 to 2008, she was a Research Staff Member with the IBM T. J. Watson Research Center, Yorktown Heights, NY, USA. She is one of the inventors of the IBM Smart Surveillance Solutions. Since 2008, she has been a Professor with the Department of Electrical Engineering, The City College, and also with the Department of Computer Science, The Graduate Center, The City University of New York, New York, NY, USA. Her research interests include remote sensing and a wide range of computer vision problems from object recognition, scene understanding, to human behavior analysis, facial expression recognition, and assistive technology.