A CFCC-LSTM Model for Sea Surface Temperature Prediction

Yuting Yang, Junyu Dong, *Member, IEEE*, Xin Sun, *Member, IEEE*, Estanislau Lima, Quanquan Mu, and Xinhua Wang

***Abstract*— Sea surface temperature (SST) prediction is not only theoretically important but also has a number of practical applications across a variety of ocean-related fields. Although a large amount of SST data obtained via remote sensor are available, previous work rarely attempted to predict future SST values from history data in spatiotemporal perspective. This letter regards SST prediction as a sequence prediction problem and builds an end-to-end trainable long short term memory (LSTM) neural network model. LSTM naturally has the ability to learn the temporal relationship of time series data. Besides temporal information, spatial information is also included in our LSTM model. The local correlation and global coherence of each pixel can be expressed and retained by patches with fixed dimensions. The proposed model essentially combines the temporal and spatial information to predict future SST values. Its structure includes one fully connected LSTM layer and one convolution layer. Experimental results on two data sets, i.e., one Advanced Very High Resolution Radiometer SST data set covering China Coastal waters and one National Oceanic and Atmospheric Administration High-Resolution SST data set covering the Bohai Sea, confirmed the effectiveness of the proposed model.**

***Index Terms*— Long short term memory (LSTM), sea surface temperature (SST), spatiotemporal sequence prediction.**

I. Introduction

IN RECENT years, the prediction of sea surface temperature (SST) attracted increasing attention in various ocean-related fields [1] such as fisheries [2], global warming [3], and oceanic environmental protection. Many different methods have been proposed in attempting to predict SST, with varying degrees of success. These methods can be classified into two general categories [4]. One category is known as the numerical model, based on physics [5]. It makes use of a series of complex physics equations to describe SST variation rules. These equations are usually very complicated and demand massive computational efforts. The other category is related to data-driven models, based on data analysis. It tries to automatically learn SST variation rules from SST data. With more and more remote sensing SST data collected, stored, processed, and disseminated by organizations, machine learning methods such as support vector machines (SVMs) were used to solve the prediction problem in the past [6]. However, the training of SVM is time consuming. In case of enough data available, deep learning outperforms SVM under most circumstances. And neural network models, such as recurrent neural network (RNN), are introduced to handle the sequence prediction problem in many areas. Nevertheless, such RNN-based sequence prediction models have hardly been used for SST prediction.

Recent advances in RNN models offer some useful ideas on sequence prediction problems. Long short term memory (LSTM) architecture, which contributes with an improvement to the hidden layer of RNN, has been successfully used to perform various supervised sequence learning tasks, such as machine translation, caption generation for images, and speech recognition. The LSTM network uses memory blocks to store and retrieve information over short or long time periods. The memory block in turn uses recurrently connected cells to learn the dependencies between two time frames, and then transfer the probabilistic inference to the next frame. These techniques have since been improved, and later approaches provide more frameworks for temporal learning problems. For example, Srivastava *et al.* [7] proposed an encoder–decoder LSTM that has successfully realized video sequence reconstruction and prediction. These advances raise some interesting possibilities. Zhang *et al.* [8] attempt at using LSTM models to solve SST prediction problems. To the best of our knowledge, their work is the first one of using the state-of-the-art sequence prediction method. In their experiments, a fully connected LSTM (FC-LSTM) architecture is used to model the sequence relationship and predict SST data. However, SST prediction is actually a spatiotemporal sequence prediction problem using historical SST data as input and future SST data as output, whereas the FC-LSTM model [8] takes only temporal information into account. This drawback of the FC-LSTM model limits the improvement of prediction accuracy. Thus, there is a large amount of information loss in the prediction process and the prediction accuracy is hard to be improved.

SST prediction is still a challenging task. Existing LSTM models [7], [8] generally ignore the spatial information of the image sequences, resulting in low prediction accuracy. Thus, a prediction model that could fully exploit the spatiotemporal information of the image sequences is greatly desired [9], [10]. In our work, we propose a prediction model that combines the temporal and spatial information, named CFCC-LSTM

Manuscript received June 25, 2017; revised September 28, 2017 and November 26, 2017; accepted December 1, 2017. Date of publication December 28, 2017; date of current version January 23, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 41741007, Grant 41576011, and Grant 41706010, in part by the Key Research and Development Program of Shandong Province under Grant GG201703140154, and in part by the Open Funding of State Key Laboratory of Applied Optics. *(Corresponding authors: Junyu Dong; Xin Sun.)*

Y. Yang, J. Dong, X. Sun, and E. Lima are with the Department of Computer Science and Technology, Ocean University of China, Qingdao 266100, China (e-mail: dongjunyu@ouc.edu.cn; sunxin@ouc.edu.cn).

Q. Mu and X. Wang are with the State Key Laboratory of Applied Optics, Chinese Academy of Sciences, Changchun 130033, China.

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

Digital Object Identifier 10.1109/LGRS.2017.2780843

(combined FC-LSTM and convolution neural network) model. This model is composed of one FC-LSTM layer and one convolution layer. Our results show that the performance of the CFCC-LSTM model for solving SST prediction problems outperforms that of the original FC-LSTM model, and those of traditional models such as SVM.

The rest of this letter is organized as follows. Section II presents the innovation of the CFCC-LSTM model and its workflow. Section III describes the experimental data and analyzes the results. Finally, our conclusion is drawn in Section IV.

II. Proposed Model

Two most important requirements for a machine learning method are sufficient data and a suitable end-to-end model. First, we propose a CFCC-LSTM model, aiming at solving sequence prediction problems, especially for complex SST images to satisfy the model requirement. Second, we introduce two typical data sets, the China Ocean data set [11] and the Bohai Sea data set, to satisfy the data requirement. The following section describes the innovation of the CFCC-LSTM model.

A. CFCC-LSTM Model

The innovation of the CFCC-LSTM model is to combine both the temporal and spatial information. The following section explains step by step how we accomplish this goal.

1) How to Deal With the Spatial Information: In this letter, we propose a 3-D grid to handle the spatial information. As shown in Fig. 1(a), the 3-D grids are composed of a sequence of patches. Each patch is a matrix of K dimensions, which consists of one center pixel and $K-1$ adjacent pixels. The adjacent pixels include the nearest neighbor pixels that are the four pixels surrounding the center pixel in the four cardinal directions. The nearest neighbor pixels contain local context and the whole adjacent pixels contain global correlations [12]. Thus, both the local context and global correlation of each pixel can be expressed and retained by patches with fixed dimensions. Here, K is fixed as 5×5. The reason is explained as follows.

The dimension of each patch should be set corresponding to the prior of the data set. Since the largest dimension of the images used in our experiments is 600×450, the patch can hardly cover oceanic large-scale features that generally span thousands of kilometers in horizontal scale. Similarly, since the highest resolution of the images used in our experiments is $1/20°$ latitude by $1/20°$ longitude, the patch can hardly cover oceanic small-scale features that have a horizontal scale not more than tens of kilometers. As a consequence, the patch could only cover oceanic mesoscale features. For the two SST data sets used in our experiments, the minimum region that each pixel covers is approximately 3 km in latitude and 5 km in longitude, respectively. Besides, mesoscale often refers to a span of hundreds of kilometers in horizontal scale. Therefore, patches with the dimension of 5×5 are the smallest regions to keep mesoscale spatial information. Thus, our model, which uses 3-D grids of $5 \times 5 \times L$ dimensions as input, is able to combine the spatial information of the image sequence. And L represents the sequence length.

2) How to Deal With the Temporal Information: To combine the temporal information, we use FC-LSTM as the first layer of our model. LSTM is first proposed in [13]. It is a specific type of RNN architecture that is explicitly designed to learn long-range dependencies among sequences.

Thus, it is reasonable and effective to use FC-LSTM as a method together with 3-D grids to combine the temporal and spatial information. The FC-LSTM layer is composed of a series of memory blocks, which in turn consists of a series of cells. The whole computation of each cell can be defined by a series of equations as follows:

$$i_t = f_1(W_i \cdot x_t + H_i \cdot h_{t-1} + C_i \cdot c_{t-1}) + b_i \quad (1)$$

$$f_t = f_1(W_f \cdot x_t + H_f \cdot h_{t-1} + C_f \cdot c_{t-1}) + b_f \quad (2)$$

$$o_t = f_1(W_o \cdot x_t + H_o \cdot h_{t-1} + C_o \cdot c_{t-1}) + b_o \quad (3)$$

$$g_t = f_2(W_g \cdot x_t + H_g \cdot h_{t-1} + C_g \cdot c_{t-1}) \quad (4)$$

$$c_t = i_t \cdot g_t + c_{t-1} \cdot f_t \quad (5)$$

$$h_{t-1} = f_2(c_t) \cdot o_t \cdot W_h \quad (6)$$

where W, H, and C are weight matrices from input to gates, recurrent connections, and cell to gates with its bias terms b_i, b_f, and b_o, respectively. There are three kinds of gates: the input gate i_t, the forget gate f_t, and the output gate o_t; each is activated by a logistic sigmoid function $f_1(\cdot)$ or a hyperbolic tangent $f_2(\cdot)$. The input gate can decide how much input information enters the current cell. The forget gate can decide how much information should be forgotten for the previous memory vector c_{t-1}, and the output gate can decide what information will be given from the current cell. The cell state c_t is computed using the gated previous state and the gated input, and it can be overwritten, kept, or retrieved by the three gates. g_t is used to transform the input and previous state to be taken into account into the current state. Finally, h_t is the output of the cell and it is scaled by a tangent transformation of the current state. t and $t-1$ are the present time and the previous time, respectively. The following section gives us a clear track of how our prediction method works.

B. Workflow of the Proposed SST Prediction Method

The workflow of the proposed SST prediction method is shown in Fig. 1. It is composed of two parts, which are the preprocessing method as shown in Fig. 1(a) and the CFCC-LSTM model as shown in Fig. 1(b). The input sequence is preprocessed to combine the temporal and spatial information into a set of 3-D grids. Then, these 3-D grids are sent into CFCC-LSTM to generate predictions.

The CFCC-LSTM model contains two parts: the FC-LSTM layer and the convolution layer. The FC-LSTM layer takes a 3-D grid as input and outputs a 3-D grid as prediction. And the dimensions of the predicted grid are the same as the input grid, as shown in Fig. 1(b). The convolution operation comprises two processes: the first process is recombination and the second process is convolution. The goal of the recombination process is to adjust the organization of the grid. In this process, the grid is converted into a 25-D vector sequence. Convolution

解决空间信息的问题，使用的是3维网格

mesoscale 中尺度

上面是对这个模型的一个大体模块介绍，主要包括空间和时间问题是各自怎么解决的，时间问题的话是用的全连接LSTM，下面这一块就是这个模型究竟是如何工作的。

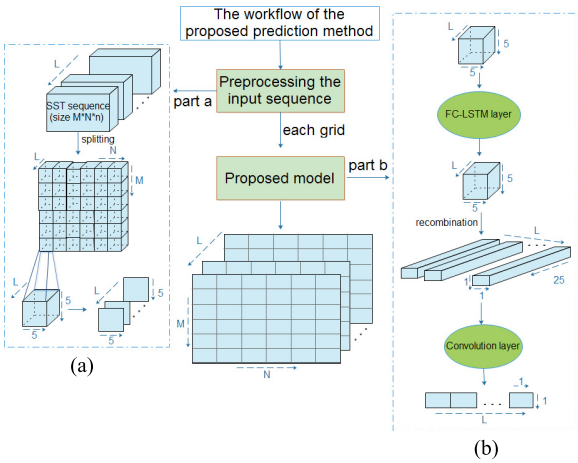


Fig. 1. Process of the proposed SST prediction method and its inner structure. (a) Preprocessing method. (b) Inner structure of the CFCC-LSTM model.

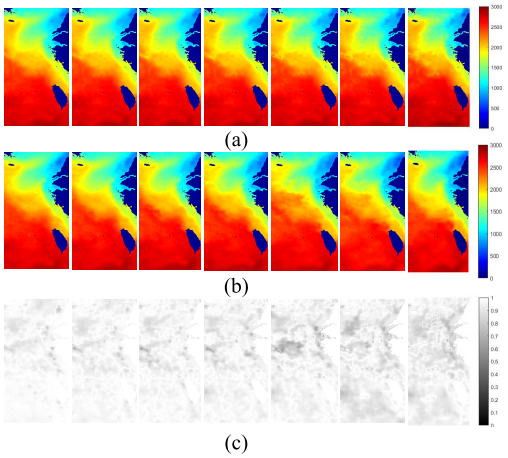


Fig. 2. Comparison between the predicted output and the ground truth for seven days on China Ocean data set. (a) Predicted SST data for seven days. (b) Ground truth SST data for seven days. (c) Difference between the ground truth and the predicted output.

is then applied to the results of recombination. And the 25-D sequence of dimensions $25 \times M \times N \times L$ is transformed into a 1-D sequence with dimensions $M \times N \times L$.

In our scenario, both the dimensions of the prediction image and the input image are $M \times N$. Therefore, CFCC-LSTM is an end-to-end model.

III. Experiments

A. Data Sets

We create two SST data sets covering the China Ocean and the Bohai Sea, respectively. The first data set is a subset of the Advanced Very High Resolution Radiometer data set, which is labeled the China Ocean data set in this letter. The second data set is a subset of the National Oceanic and Atmospheric Administration High-Resolution SST data set, namely, the Bohai Sea data set. The China Ocean data set contains 600 daily SST images from 2007 to 2012 covering China coastal waters. Each image has dimensions of 300×150 pixels. The Bohai Sea data set used in this letter contains a total of 12 868 daily patches, and each patch has dimensions of 16×15 pixels. The Bohai Sea is the innermost gulf of the Yellow Sea on the coast of Northeastern China. The differences between the two data sets mainly lie in their resolution and coverage. The resolution of the China Ocean data set is 1/20° latitude by 1/20° longitude, while that of the Bohai Sea data set is 1/4° latitude by 1/4° longitude. The China Ocean data set covers the area from $20N$ to $30N$ in latitude and $120E$ to $131.25E$ in longitude, while the Bohai Sea data set covers the area from $37.07N$ to $41N$ in latitude and $117.35E$ to $121.10E$ in longitude.

这里的两个数据集可以借鉴一下

B. Results and Analysis

In this letter, we evaluate the effectiveness of different prediction methods using the root of mean squared error (*RMSE*) and the prediction accuracy (*ACC*). The error formulation and the performance metric are listed as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(X_{\mathrm{obs},i} - X_{\mathrm{model},i})^2}{n}} \tag{7}$$

$$\mathrm{ACC} = 1 - \frac{\sum_{i=1}^{n}\left(\frac{|X_{\mathrm{model},i} - X_{\mathrm{obs},i}|}{X_{\mathrm{obs},i}}\right)}{n} \tag{8}$$

where $X_{\mathrm{obs},i}$ is observed values and $X_{\mathrm{model},i}$ is modeled values at location i. i is numbered in the order of row column. n is the total number of the pixels to be predicted. Here, $n = M \times N$.

Here, *RMSE* can be regarded as absolute error and *ACC* as relative accuracy. The smaller the *RMSE* the better performance we get, while *ACC* is the opposite. For area predictions, we use the area average *RMSE* and area average *ACC*. The boldface items in Tables I–VI represent the best performance, i.e., the smallest RMSE and the largest ACC.

标注黑体的要进行说明是啥意思

First, we choose the learning rates (lr) between 0.02 and 0.5, which have the advantage of better convergence. When lr is smaller than 0.02, there exists the problem of slow convergence. When lr is larger than 0.5, too early convergence and no convergence might appear. Second, the reason we make predictions for 1 day, 7 days, and 30 days in the future is to verify the proposed model in the case of short-term prediction, midlong-term prediction, and long-term prediction, respectively. Besides, it is clear that these time spans of one day, one week, and one month are closely related to daily life and industry requirement. Thus, the proposed model could be more practical. Third, the reason we set the iteration times (it) as 50 is intend to shorten the training time, while 100 is intend to obtain a better convergence. When it is smaller than 50, it is difficult to converge. However, when it is larger than 100, the increasement in it can hardly improve the accuracy.

At the beginning, we randomly initialize the weights. We set $\mathrm{lr} = 0.5$, $\mathrm{it} = 100$, and use four days' historical SST data for one-day prediction and 20 for seven-day prediction, respectively. Then, we carry out experiments using different it and different convolution strategies, i.e., using only the FC-LSTM layer, using the FC-LSTM layer combined with average convolution, and using the FC-LSTM layer combined with weighted convolution, respectively.

As shown in Table I, the FC-LSTM layer combined with weighted convolution achieves the best performance with

TABLE I

PREDICTION RESULTS (AREA AVERAGE *RMSE* AND *ACC*) USING DIFFERENT CONVOLUTION STRATEGIES AND ITERATION TIMES ON THE BOHAI SEA DATA SET

CFCC-LSTM Model	Metrics	Iterations (prediction days = 1,7)			
		50	100	50	100
Only FC-LSTM Layer	$RMSE$	0.1826	0.2281	0.5119	0.5351
	ACC	99.21	99.21	98.23	98.23
FC-LSTM Layer + Average Convolution	$RMSE$	**0.1420**	0.1716	0.4699	0.4725
	ACC	99.27	99.27	98.38	98.38
FC-LSTM Layer + Weighted Convolution	$RMSE$	0.1583	**0.1583**	**0.3242**	**0.3242**
	ACC	**99.32**	**99.32**	**98.41**	**98.41**

TABLE II

PREDICTION RESULTS (AREA AVERAGE *RMSE* AND *ACC*) FOR FIVE DIFFERENT LOCATIONS USING DIFFERENT LEARNING RATES AND INITIALIZATION CONDITIONS

Initialize	Metrics	p_1	p_2	p_3	p_4	p_5
Random lr=0.02	$RMSE$	0.5480	0.7790	0.7160	0.2260	0.5360
	ACC	98.12	97.60	97.82	98.78	98.12
Fixed lr=0.02	$RMSE$	0.3913	0.3883	0.4401	0.3022	0.5331
	ACC	98.48	98.57	98.22	98.73	98.13
Random lr=0.5	$RMSE$	0.5023	0.4042	0.9022	1.020	0.4022
	ACC	98.19	98.43	97.50	97.22	98.44
Fixed lr=0.5	$RMSE$	**0.2501**	**0.2200**	**0.1370**	**0.1330**	**0.1870**
	ACC	**98.71**	**98.79**	**99.04**	**99.07**	**98.87**

it $= 50$. It shows that smaller it obtains equivalent or even better results. Besides, the FC-LSTM layer combined with average convolution achieves higher accuracy and lower *RMSE* than with only an FC-LSTM layer. Nevertheless, the FC-LSTM layer combined with weighted convolution achieves even higher accuracy than with average convolution. Thus, we decide to use the FC-LSTM layer combined with weighted convolution as our model and set it $= 50$ for our following experiments.

In order to obtain the best performance, we carry out experiments with different initializations and lr. We randomly choose five locations from the Bohai Sea data set, denoted by $p_1, p_2, \ldots, p_5$ to predict SST values for the next seven days. First, we carry out experiments on these five locations to choose a better initialization method from random to fixed initialization. The top part of Table II shows the results on five different locations with different initialization conditions. It can be seen from the results that the best performance is achieved with the fixed initialization scheme. Then, we carry out experiments on the same five locations to choose a better lr from values between 0.02 and 0.5. There is a general trend that CFCC-LSTM performs better as lr increases. The bottom part of Table II shows the results using different lr, and it can be seen from the results that the best performance is achieved with lr $= 0.5$. Thus, in the following experiments, we set lr $= 0.5$ and fix the initialization weights.

Then, we perform experiments using k days' historical SST data to predict the future SST data, i.e., the future one day and seven days. Here, k represents the number of days used for prediction. As shown in Tables III and IV, the best k values are seven for one-day prediction and 20 for seven-day prediction.

TABLE III

PREDICTION RESULTS (AREA AVERAGE *RMSE* AND *ACC*) USING DIFFERENT k ON DIFFERENT DATA SETS

CFCC-LSTM Model	Metrics	k		
		4	7	20
Bohai Sea Dataset	$RMSE$	**0.1399**	0.1466	0.1445
	ACC	99.49	**99.58**	99.53
China Ocean Dataset	$RMSE$	0.4584	**0.4078**	0.4109
	ACC	98.32	**98.52**	98.50

TABLE IV

PREDICTION RESULTS (AREA AVERAGE *RMSE* AND *ACC*) USING DIFFERENT k ON DIFFERENT DATA SETS

CFCC-LSTM Model	Metrics	k		
		7	20	50
Bohai Sea Dataset	$RMSE$	0.2911	**0.2772**	0.2882
	ACC	98.69	98.54	**98.89**
China Ocean Dataset	$RMSE$	0.6884	**0.6307**	0.6311
	ACC	97.52	**97.61**	97.60

TABLE V

PREDICTION RESULTS (AREA AVERAGE *RMSE* AND *ACC*) ON THE BOHAI SEA DATA SET (37.07N TO 41N, 117.35E TO 121.10E)

Methods	Metrics	Prediction Days		
		1	7	30
SVM	$RMSE$	0.3998	0.8388	1.2477
	ACC	98.72	97.28	95.93
SVR	$RMSE$	0.3847	0.6305	1.3052
	ACC	98.49	97.82	96.03
FC-LSTM	$RMSE$	**0.0767**	0.6540	1.1363
	ACC	99.23	97.55	96.90
CFCC-LSTM Model	$RMSE$	0.1466	**0.2772**	**0.7260**
	ACC	**99.58**	**98.54**	**97.62**

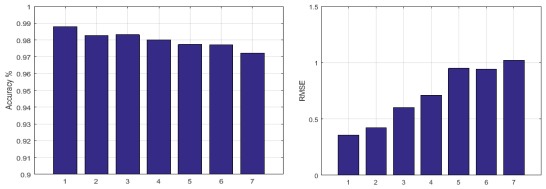
In order to compare CFCC-LSTM with FC-LSTM, SVM, and SVR, all of the three methods are tested with our two data sets. For the FC-LSTM network, we set lr $= 1$, $k = 10$, 30, and 120 for one-day prediction, seven-day prediction, and 30-day prediction, respectively. For SVM, we use libSVM for prediction and perform experiments with MATLAB. For SVR, we use the RBF kernel for prediction and perform experiments using the TensorFlow Toolkit [14]. Moreover, the kernel width for RBF is set as $\sigma = 1.6$, which is chosen by cross validation. For both SVM and SVR, we set $k = 7$, 20, and 50 for one-day prediction, seven-day prediction, and 30-day prediction, respectively. The performances of these methods are shown in Tables V and VI, respectively.

Possibly the most notable trend may originate from the bottom rows of Tables V and VI, which shows that CFCC-LSTM with lr $= 0.5$ and it $= 50$ outperforms the existing models on the two data sets. Its accuracies, i.e., 98.52%, 97.61%, and 96.59% on the China Ocean data set and 99.58%, 98.54%, and 97.62% on the Bohai Sea data set, are significantly higher than that of SVM, SVR, and FC-LSTM, respectively. Also, the long-term prediction stability of the CFCC-LSTM model outperforms that of the FC-LSTM model and the SVM method. Its RMSE is close to 50% lower than

TABLE VI

PREDICTION RESULTS (AREA AVERAGE *RMSE* AND *ACC*) ON THE CHINA OCEAN DATA SET (20N TO 30N, 120E TO 131.25E)

Methods	Metrics	Prediction Days		
		1	7	30
SVM	*RMSE*	0.6590	1.0088	1.2159
	ACC	98.14	96.84	95.95
SVR	*RMSE*	0.5555	0.9239	1.3672
	ACC	97.71	97.28	95.78
FC-LSTM	*RMSE*	0.9342	1.1652	1.4670
	ACC	97.68	96.42	95.98
CFCC-LSTM Model	*RMSE*	**0.4078**	**0.6307**	**1.0701**
	ACC	**98.52**	**97.61**	**96.59**



(a) (b)

Fig. 3. (a) Accuracy achieved on the China Ocean data set for seven days. (b) Corresponding *RMSE* achieved on the China Ocean data set for seven days.

that of FC-LSTM and SVM for seven-day prediction and 30-day prediction.

Besides, the accuracy achieved on the China Ocean data set is lower than that achieved on the Bohai Sea data set. The reason lies in the differences of resolution and area coverage. It is obvious that the higher the resolution is, the harder is to generate an accurate prediction. The resolution of the China Ocean data set is 25 times higher than that of the Bohai Sea data set, and the regional coverage of the China Ocean data set is seven times larger than that of the Bohai Sea data set. Therefore, the accuracy on both the China Ocean data set and the Bohai Sea data set could verify the robustness of the CFCC-LSTM model.

To further prove the effectiveness of CFCC-LSTM, we visualize the seven-day prediction and the ground truth in Fig. 2. To put things into places in the overall view, there is high similarity between the prediction that is shown in Fig. 2(a) and the ground truth that is shown in Fig. 2(b). The image differences between the ground truth and the prediction are shown in Fig. 2(c). We invert the black and white for better visual display. However, when we check the details, most of the differences are located in the areas where the SST variation over time is the local maximum. This is because the SST in these areas changes quite rapidly and is affected by many factors. Thus, it is hard to get an accurate prediction according to historical data in these areas.

Fig. 3(a) and (b) gives us a more direct exhibition of the *ACC* changing trend and that of *RMSE*, respectively. Through the two histograms, we can see the prediction accuracy shows a downward trend, while the prediction error presents a rising trend. This is because the prediction difficulty increases over time, and the prediction error rises correspondingly. More importantly, the *ACC* remains above 0.97, and the *RMSE* remains below 1.5. Thus, the stability of *ACC* and *RMSE* could verify the effectiveness of our model.

IV. CONCLUSION

In this letter, we improved the SST prediction accuracy by combining both the spatial and temporal information. We achieved this goal using a two-part strategy: one 3-D grid, and the combination of one FC-LSTM layer and one convolution layer. The 3-D grid constrains the local correlation and the global coherence of the center pixel. The convolution layer further encodes the spatial and temporal information for SST prediction. Our experimental results conducted on two different SST data sets show that our model achieved the best performance for SST prediction compared with other methods.

For further work, the spatial information can be encoded in a more complex way. And the proposed method should be employed into practical applications, such as ocean front detection [15], [16] and eddy recognition.

REFERENCES

[1] Z. Z. Hu, A. Kumar, B. Huang, W. Wang, J. Zhu, and C. Wen, "Prediction skill of monthly SST in the North Atlantic Ocean in NCEP Climate Forecast system version 2," *Climate Dyn.*, vol. 40, nos. 11–12, pp. 2745–2759, 2013.

[2] H. U. Solanki, D. Bhatpuria, and P. Chauhan, "Integrative analysis of AltiKa-SSHa, MODIS-SST, and OCM-chlorophyll signatures for fisheries applications," *Marine Geodesy*, vol. 38, pp. 672–683, Mar. 2015.

[3] C. C. Funk and A. Hoell, "The leading mode of observed and CMIP5 ENSO-residual sea surface temperatures and associated changes in indo-Pacific climate," *J. Climate*, vol. 28, no. 11, p. 150202132719008, 2015.

[4] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1279–1289, Jun. 2016.

[5] T. N. Stockdale, M. A. Balmaseda, and A. Vidard, "Tropical Atlantic SST prediction with coupled ocean–atmosphere GCMs," *J. Climate*, vol. 19, no. 19, pp. 6047–6061, 2010.

[6] I. D. Lins *et al.*, "Sea surface temperature prediction via support vector machines combined with particle swarm optimization," in *Proc. Int. Probabilistic Safety Assessment Manage. Conf.*, 2010, pp. 16–29.

[7] N. Srivastava, E. Mansimov, and R. Salakhutdinov, "Unsupervised learning of video representations using LSTMs," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 843–852.

[8] Q. Zhang, H. Wang, J. Dong, G. Zhong, and X. Sun, "Prediction of sea surface temperature using long short-term memory," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1745–1749, Oct. 2017.

[9] Q. Wang, Z. Meng, and X. Li, "Locality adaptive discriminant analysis for spectral–spatial classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2077–2081, Nov. 2017.

[10] X. Sun, F. Zhou, J. Dong, F. Gao, Q. Mu, and X. Wang, "Encoding spectral and spatial context information for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2250–2254, Dec. 2017.

[11] Y. Yang, J. Dong, X. Sun, R. Lguensat, M. Jian, and X. Wang, "Ocean front detection from instant remote sensing SST images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1960–1964, Dec. 2016.

[12] Y. Yuan, J. Lin, and Q. Wang, "Hyperspectral image classification via multitask joint sparse representation and stepwise MRF optimization," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 2966–2977, Dec. 2015.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[14] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. Usenix Conf. Oper. Syst. Design Implement.*, 2016, pp. 265–283.

[15] E. Lima, X. Sun, Y. Yang, and J. Dong, "Application of deep convolutional neural networks for ocean front recognition," *J. Appl. Remote Sens.*, vol. 11, no. 4, p. 042610, 2017.

[16] E. Lima, X. Sun, J. Dong, H. Wang, Y. Yang, and L. Liu, "Learning and transferring convolutional neural network knowledge to ocean front recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 354–358, Mar. 2017.