An Attention-based Context Fusion Network for Spatiotemporal Prediction of Sea Surface Temperature

Benyun Shi, Yingjian Hao, Liu Feng, Conghui Ge, Yue Peng, and Hailun He

***Abstract*—Sea surface temperature (SST) is a fundamental parameter in the field of oceanography as it significantly influences various physical, chemical, and biological processes within the marine environment. In this study, we propose an Attention-based Context Fusion Network (ACFN) model for short-term prediction of SST based on the Operational SST and Sea Ice Analysis (OSTIA) data. The ACFN model combines an attention-based context fusion block with the Convolutional Long Short-Term Memory (ConvLSTM) model, enabling the exploration of intricate spatiotemporal correlations between the previous context state and the current input state in ConvLSTM. To assess the performance of the ACFN model, we apply it to predict SST in the Bohai Sea over lead times spanning from 1 to 10 days. The results demonstrate that our proposed model outperforms several state-of-the-art models, i.e., ConvLSTM, PredRNN, and MoDeRNN, in terms of mean absolute error and coefficient of determination. In particular, our analysis reveals that the prediction errors near the coastal areas exhibit relatively higher values compared to those in the central Bohai Sea.**

***Index Terms*—Sea surface temperature, Artificial neural networks, ConvLSTM, Attention-based context fusion network, Spatiotemporal prediction.**

I. Introduction

SHORT-TERM prediction of sea surface temperature (SST) is of great significance for operational weather and ocean forecasting, as it provides valuable insights into the dynamic behavior of the marine environment [1]–[3]. However, accurately forecasting SST over a wide geographical expanse poses challenges due to the complex nature of the SST anomalies, as well as the intrinsic uncertainties on the conditions of sea or ocean systems [4]–[6]. Taking into consideration the SSTs observed in the Bohai Sea, evidence has shown that the spatiotemporal variability of SST in this region is notably pronounced [7].

Traditionally, extensive methods have been proposed in the domain of SST prediction, such as the use of Ocean General Circulation Models (OGCMs) and various physics-based numerical models [8]. These methods, firmly grounded in established principles of physical oceanography and statistical modeling, have significantly contributed to the understanding of SST dynamics. Nevertheless, they grapple with inherent limitations, such as their incapacity to effectively capture intricate nonlinear relationships and manage the inherent high-dimensional complexity inherent in SST data. Furthermore, these approaches encounter challenges in adapting to dynamic shifts in data distributions or the emergence of new patterns, thereby impeding their capacity to remain adaptable in tandem with evolving environmental conditions.

In recent years, the advent of artificial neural networks (ANNs) has revolutionized the domain of SST prediction, offering distinct advantages over traditional methods [9]–[12]. The key advantage of ANNs lies in their capacity to model nonlinear relationships, adapt to changing data distributions, and capture subtle dependencies that may elude conventional techniques [13]. Neural network-based predictions for SST involve taking factors that affect SST as inputs and corresponding SST values as targets. These inputs and targets are connected by neurons, and a learning algorithm adjusts the weights of these neurons to establish a specific relationship between inputs and targets. The set of weights obtained stores the nonlinear relationship between factors and SST.

Existing studies have predominantly utilized SST or SST anomaly (SSTA) time series as predictors while emphasizing SST prediction at fixed spatial coordinates organized into regular grids [14]. The primary challenge lies in how to effectively capture the underlying spatiotemporal correlations present in SST data. Recent advances in recurrent neural networks (RNNs), such as Long Short-Term Memory (LSTM) networks [14], [15] or Gated Recurrent Units (GRU), have shown better capability in capturing temporal dependencies of SST time series [16].

In addition to pre-processing the SST series, convolutional neural networks (CNNs) can directly extract the spatial features from the original data. Therefore, in order to better exploit the spatiotemporal correlations within SST data, researchers have proposed research architectures that combine CNNs with RNNs. For example, Xiao et al. used the convolutional LSTM model to predict SSTs in the East China Sea [17]. Wei and Guan proposed a 3DConv-LSTM model for 7-day SST prediction [12]. However, most models in use tend to extract spatial correlations and temporal correlations sequentially. There remains a significant scope for enhancement in the simultaneous extraction of spatiotemporal correlations.

Convolutional LSTM (ConvLSTM) extends the LSTM model by extending the fully connected LSTM (FC-LSTM) to have convolutional structures in both the input-to-state and state-to-state transitions [18]. In doing so, the spatiotemporal relationships can be captured simultaneously through an end-

Benyun Shi, Yingjian Hao, Liu Feng, Conghui Ge, and Yue Peng are with the College of Computer and Information Engineering, Nanjing Tech University, Nanjing, China; Hailun He is with the State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, China. (Corresponding author: Hailun He, e-mail: hehailun@sio.org.cn)

to-end learning structure. Due to the outstanding performance of ConvLSTM in spatiotemporal prediction, an increasing number of variant models have been introduced. For example, predictive RNN (PredRNN) enhances ConvLSTM by adding a spatiotemporal memory unit [19]. Motion Details RNN (MoDeRNN) improves ConvLSTM by introducing a detail context block to increase the correlation between input states and hidden states [20].

Based on the aforementioned work, this paper introduces an attention-based context fusion network (ACFN) model. The main focus lies in the investigation of the spatiotemporal correlations between the previous hidden state and the current inputs, within the framework of ConvLSTM. To achieve this, we incorporate the convolutional block attention module (CBAM) within the detail context block for adaptive feature refinement [20], [21]. Meanwhile, we implement the ACFN model SST prediction in the Bohai Sea area across lead times ranging from 1 to 10 days. A comprehensive evaluation of its performance is conducted by comparing it against several state-of-the-art models.

The remainder of this letter is organized as follows: In Section II, we formally define the SST prediction problem. In Section III, we introduce the ACFN model in detail. In Section IV, we evaluate the performance of the ACFN model in predicting the SST in the Bohai Sea. Finally, we conclude this work in Section V.

II. Problem Statement

Given a dataset $\mathcal{D} = \{(X_1, X_2, \ldots, X_\tau)_i\}_{i=1}^N$, consisting of N samples of SST data. Each sample in $\mathcal{D}$ represents a sequence of τ SST fields. Each SST field is represented as a grid matrix with dimensions $W \times H$, where the SST values are recorded in a single channel. Therefore, the shape of each sample in $\mathcal{D}$ is $(\tau, 1, W, H)$, where 1 indicates the channel representing SST. The goal of the multiple-input multiple-output (MIMO) SST prediction problem is to predict the SST values for the next k time steps by utilizing the observed SST values from the preceding τ time steps.

Mathematically, the prediction task can be formulated as the learning a mapping function $f\colon X^{\tau\times 1\times W\times H} \to Y^{k\times 1\times W\times H}$. Here, X signifies the input tensor encompassing historical data related to the SST field with dimensions $W \times H$, while Y represents the corresponding observed output tensor within the same field. Given the training set $\mathcal{D}$, our primary objective is to minimize the error between the predicted values and the actual observations. This minimization objective can be expressed through the following equation:

$$\arg\min_f \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{k} \textit{LOSS}(f(X_i)_j - Y_{i,j}).$$

Here, $f(X_i)_j$ is the predicted value of the SST on the j-th day based on the i-th sample, and $Y_{i,j}$ is the actual observation of the SST on the corresponding day.

III. Proposed Model

A. Model Architecture

In this section, we present an ACFN model for short-term SST prediction. The architecture of our proposed model, designed for MIMO SST prediction, is illustrated in Figure 1(a). The given inputs $[X_{1:\tau}] = (X_1, X_2, \ldots, X_\tau)$ are first encoded by a weight-shared 2D CNN encoder, denoted as $[F_{1:\tau}] = \mathcal{E}([X_{1:\tau}])$, which then serves as the input of N-layer ACFNs to generate high-order spatiotemporal representations of $[X_{1:\tau}]$ and output states $[\hat{X}_{\tau+1:\tau+k}] = (\hat{X}_{\tau+1}, \ldots, \hat{X}_{\tau+k})$. Finally, the output states are decoded by 2D CNN decoder iteratively to forecast the next k SST fields, that is, $[\hat{Y}_{\tau+1:\tau+k}] = \mathcal{D}([\hat{X}_{\tau+1:\tau+k}])$. In this study, the encoder and decoder both use 1×1 kernel CNN layers are kept consistent with the approach presented in the previous work of PredRNN [19].

B. Attention-based Detail Context Block

For the previous hidden state H_{t-1} and the intermediate feature representation $\hat{F}_t$ in ADCB illustrated in Figure 1(b), we adopt a CBAM to explore important features and suppress unnecessary ones [21]. Take $\hat{F}_t$ as an example, the CBAM sequentially infers a 1-D channel attention map $\mathbf{M_c} \in \mathbb{R}^{C\times 1\times 1}$ and a 2-D spatial attention map $\mathbf{M_s} \in \mathbb{R}^{1\times W\times H}$. The entire CBAM process is summarized as follows:

$$\hat{F}'_t = \mathbf{M_c}(\hat{F}_t) \otimes \hat{F}_t$$
$$\hat{F}''_t = \mathbf{M_s}(\hat{F}'_t) \otimes \hat{F}'_t$$

where $\otimes$ denotes element-wise multiplication. In this paper, we use the same equations for channel and spatial attention as outlined in [21]:

$$\mathbf{M_c}(\hat{F}_t) = \sigma(MLP(AvgPool(\hat{F}_t)) + MLP(MaxPool(\hat{F}_t))),$$

and

$$\mathbf{M_s}(\hat{F}'_t) = \sigma(f([AvgPool(\hat{F}'_t); MaxPool(\hat{F}'_t)])),$$

where σ denotes the sigmoid function, MLP is a multi-layer perceptron with one hidden layer, and f is a 2-D convolutional operation. By doing so, CBAM enables the model to focus on important spatial or channel features.

We further adopt the process of detail context block to capture fine-grained intricate spatiotemporal features and update upper hidden state H_{t-1} and the current input state F_t [20]. Denote Att_H and Att_F the outputs of CBAM for H_{t-1} and $\hat{F}_t$ respectively, the remaining procedure of ADCB can be summarized as follows:

$$Att_H = CBAM(H_{t-1})$$
$$\hat{F}_t = s \times Att_H \times F_t$$
$$Att_F = CBAM(\hat{F}_t)$$
$$\hat{H}_{t-1} = s \times Att_F \times H_{t-1}.$$

In this context, the input feature F_t undergoes a reweighting process by taking the Hadamard product with Att_H and multiplying it by a scale factor s. Then, the updated feature $\hat{F}_t$ is used to generate Att_F, which is further used to update H_{t-1} in the same way. In doing so, the updated states $\hat{H}_{t-1}$ and $\hat{F}_t$ will capture rich spatiotemporal features.

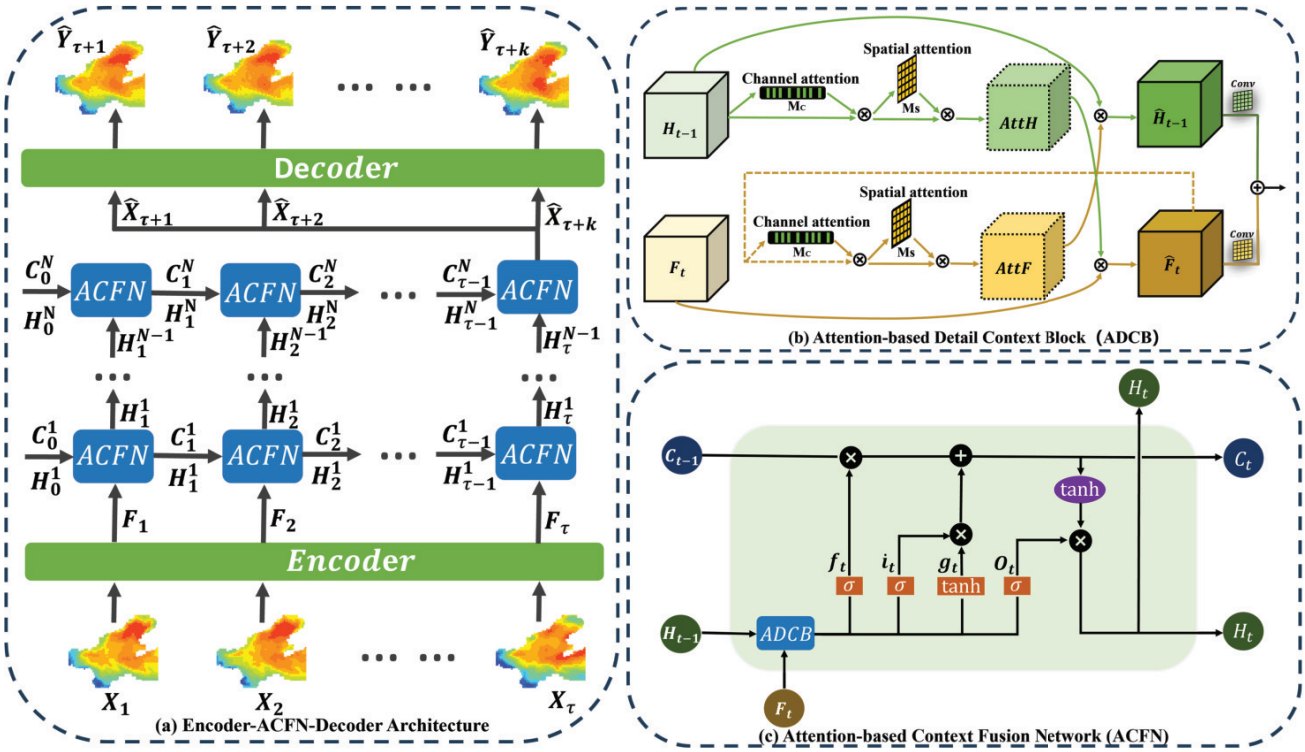

Fig. 1. Overview of the proposed attention-based context fusion network (ACFN) model. (a) The Encoder-ACFN-Decoder architecture for multiple-input multiple-output SST prediction; (b) The attention-based detail context block (ADCB), which is embedded in the standard Convolutional Long Short-term Memory (ConvLSTM) model; (c) The attention-based context fusion network module, which incorporates the ADCB block with the ConvLSTM model.

C. Attention-based Context Fusion Network

The proposed ACFN model incorporates an attention-based detail context block (ADCB) with the standard ConvLSTM model, as illustrated in Figure 1(c). This fusion enables a comprehensive exploration of intricate spatiotemporal relationships between the previous hidden state and the current input state in ConvLSTM. Specifically, ACFN utilizes the updated states $\hat{H}_{t-1}$ and $\hat{F}_t$ to derive the detailed output state H_t and memory state C_t. In practice, it employs the ConvLSTM model and accepts the updated states $\hat{H}_{t-1}$ and $\hat{F}_t$ as inputs. The process can be formulated as follows:

$$
\begin{aligned}
i_t &= \sigma(W_{xi} * \hat{F}_t + W_{hi} * \hat{H}_{t-1} + b_i) \\
f_t &= \sigma(W_{xf} * \hat{F}_t + W_{hf} * \hat{H}_{t-1} + b_f) \\
o_t &= \sigma(W_{xo} * \hat{F}_t + W_{h}o * \hat{H}_{t-1} + b_o) \\
g_t &= \tanh\left(W_{xg} * \hat{F}_t + W_{hg} * \hat{H}_{t-1} + b_g\right) \\
C_t &= f_t \otimes C_{t-1} + i_t \odot g_t \\
H_t &= o_t \otimes \tanh(C_t)
\end{aligned}
$$

where i_t, f_t, o_t, g_t, C_t, and H_t represent the input gate, forget gate, output gate, control gate, cell state, and hidden state, respectively. W_{xg}, W_{hg}, W_{xf}, W_{hf}, W_{xo}, W_{ho}, W_{xi} and W_{hi} are all 5×5 kernels for gate operation, and b represents the bias vector. Moreover, $*$ is the convolution operator. the obtained H_t and C_t are utilized as inputs for the ACFN module in the subsequent time step $t+1$. Moreover, H_t is also fed into the ACFN module in the upper layer.

IV. Experiments

In this section, we evaluate the performance of the ACFN model for SST prediction in the Bohai Sea over lead times spanning from 1 to 10 days. This evaluation involves a comparative analysis with three state-of-the-art baseline models: ConvLSTM, PredRNN, and MoDeRNN.

A. Study area and Dataset

We conducted an experimental study for SST prediction within the Bohai Sea region, which spans from 37.5°N to 40.7°N in latitude and 118.81°E to 122.0°E in longitude. This region is characterized as a semi-enclosed shelf sea, connecting to the Yellow Sea via the Bohai Strait. For this study, authentic SST data were sourced from the Operational SST and Sea Ice Analysis (OSTIA) system [22]. The spatial resolution is $1/20^{\circ}\times1/20^{\circ}$ in longitude and latitude, with a time range from October 1, 1981 to the present [22].

B. Experimental Settings

In this investigation, a dataset comprising a total of 14,600 temporal data points spanning from January 1982 to December 2021 was assembled. To prepare the data for model training, all SST records underwent a normalization process to scale them within the range of 0 to 1. Following normalization, a sliding backward window technique was applied to organize the data into a 5-dimensional tensor with dimensions $(14581, 10, 1, 64, 64)$. To create datasets for model training, validation, and testing, the 5-dimensional tensor data was

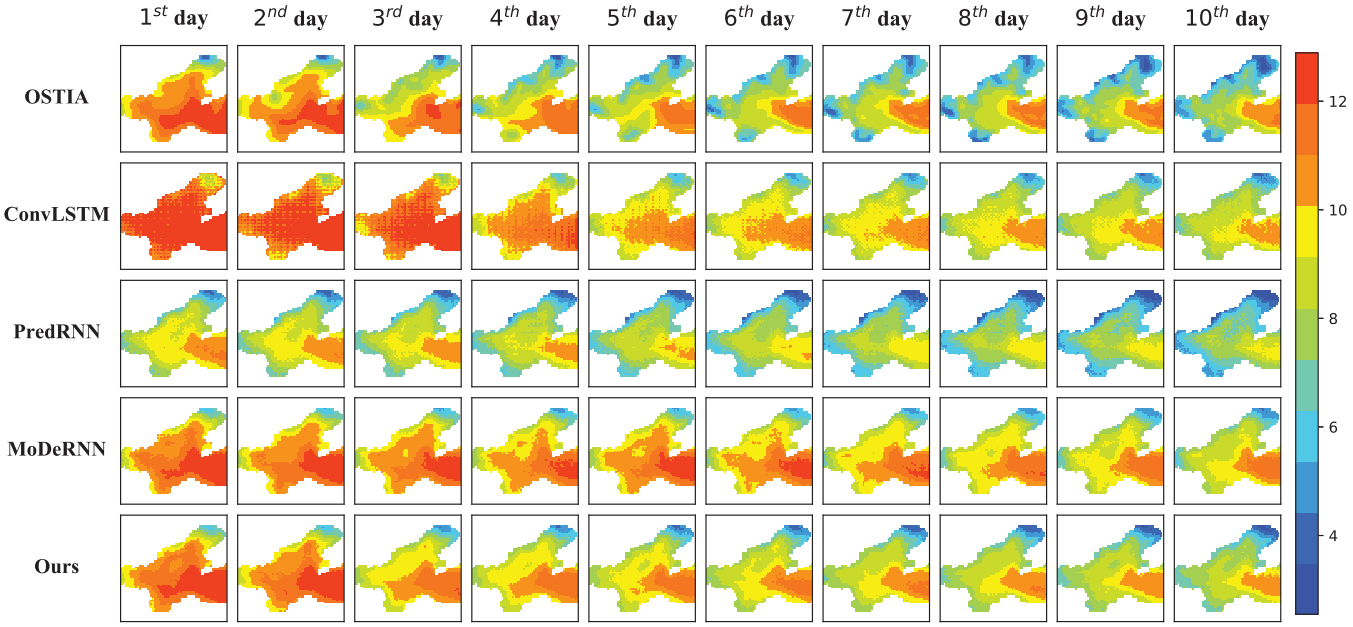

Fig. 2. Observed OSTIA and predicted SST with the reference time being November 24, 2015 using different neural network models, namely, ConvLSTM, PredRNN, MoDeRNN, and our proposed ACFN.

partitioned in a 7:1:2 ratio. In addition, two evaluation metrics, namely Mean Absolute Error (MAE) and Coefficient of Determination (R^2), were employed to gauge the predictive performance of the model.

C. Experimental Results

Figure 2 illustrates predicted SST snapshots, with the reference time being November 24, 2015. For the 1-day lead time prediction (corresponding to November 25, 2015), the true OSTIA SST data indicate the expansion of warm water from the Yellow Sea into the Bohai Sea, with relatively cooler temperatures observed in the northeastern Bohai Sea compared to other areas. In the case of lead times ranging from 2 to 4 days, the true SST exhibits a cooling trend along the near-coastal and central regions of the Bohai Sea. ConvLSTM, on the other hand, suggests a warming trend for the 1-day lead time, which contrasts with the corresponding OSTIA observations. However, for lead times of 2 to 4 days, ConvLSTM fails to capture the observed SST cooling in the near-coastal and central areas of the Bohai Sea. PredRNN, in general, indicates lower SST values compared to actual observations in both the Bohai Sea and the adjacent Yellow Sea. MoDeRNN and ACFN demonstrate superior performance, particularly for lead times of 1 and 2 days. ACFN exhibits predictions that are closer to the observed values than MoDeRNN. For instance, at a 3-day lead time, ACFN aligns with OSTIA observations, depicting the disappearance of the 10°C warm water in the northern Bohai Sea. Conversely, MoDeRNN suggests the persistence of 10°C warm water across the main Bohai Sea. Finally, for a 10-day lead time, ACFN forecasts significant SST cooling in the northeastern part of the Bohai Sea, consistent with observational data. However, MoDeRNN predicts a relatively weaker cooling signal in this region.

Figure 3 presents the spatial distribution of MAE within the studied area. For the 1-day lead time prediction, ConvLSTM exhibits MAE values exceeding 1.2°C, with relatively lower MAE values observed around the Bohai Strait, approximately 1°C. PredRNN showcases improved performance compared to ConvLSTM in the central and northeastern parts of the Bohai Sea, but performs less favorably around the Bohai Strait, where MAE values are higher. MoDeRNN and ACFN consistently demonstrate notably low MAE values across the studied area. Even in the vicinity of the coast, the maximum MAE remains below 0.6°C for a 1-day lead time prediction. As lead times extend, MAEs for PredRNN, MoDeRNN, and ACFN generally exhibit an increase. Notably, ACFN displays a discernible improvement compared to MoDeRNN, particularly evident in the 8-day lead time prediction. In the western Bohai Sea, MoDeRNN exhibits MAE values exceeding 0.8°C, whereas ACFN consistently maintains lower MAE values, primarily around 0.6°C.

Figure 4 presents the geographical distribution of the R^2 for various predictive models. In the case of ConvLSTM, R^2 values are notably higher in the central area of the Bohai Sea compared to the Bohai Strait area. Furthermore, the R^2 for a 1-day lead time is lower than that predicted for a 7-day lead time. PredRNN, conversely, exhibits higher R^2 values in the western and northeastern parts of the Bohai Sea, in contrast to lower values in the vicinity of the Bohai Strait and the adjacent Yellow Sea. Moreover, R^2 values in PredRNN tend to decrease as lead time extends. In the case of both MoDeRNN and ACFN, R^2 values approach near-perfection for a 1-day lead time, with R^2 exceeding 0.99 in most areas of the Bohai Sea. Concerning the comparative advantages of ACFN over MoDeRNN, our ACFN model demonstrates superior R^2 values in the central area of the Bohai Sea, particularly for lead times spanning from 6 to 10 days. Notably, the R^2 values

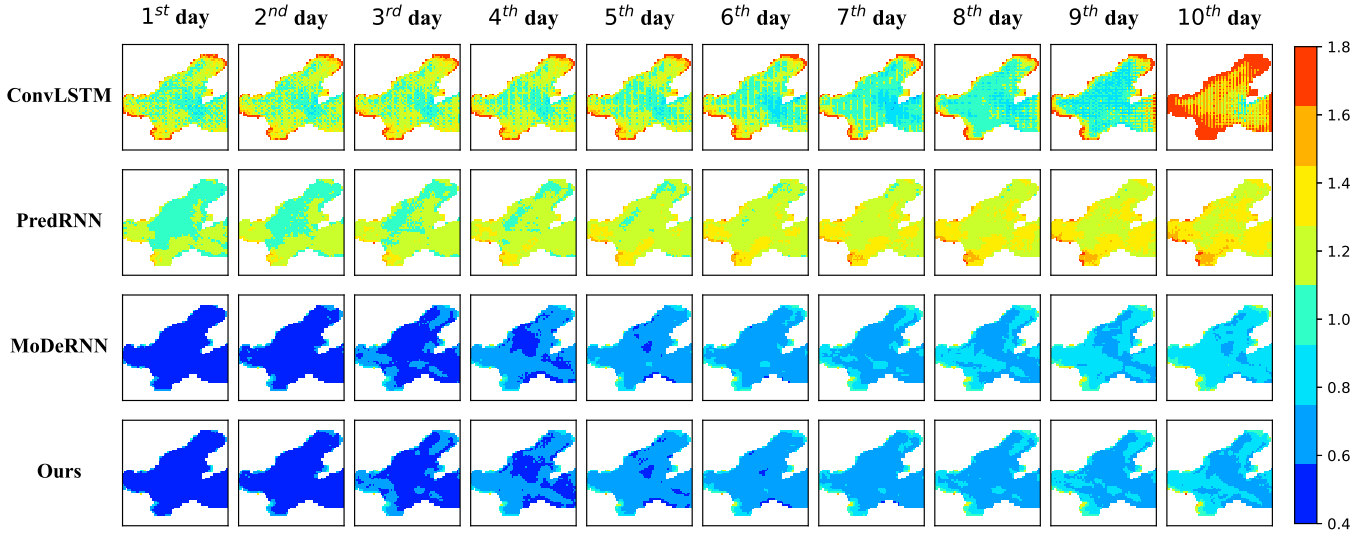

Fig. 3. Spatial distribution of MAE on all testing samples for different lead times with respect to four neural network models, namely, ConvLSTM, PredRNN, MoDeRNN and our proposed ACFN.

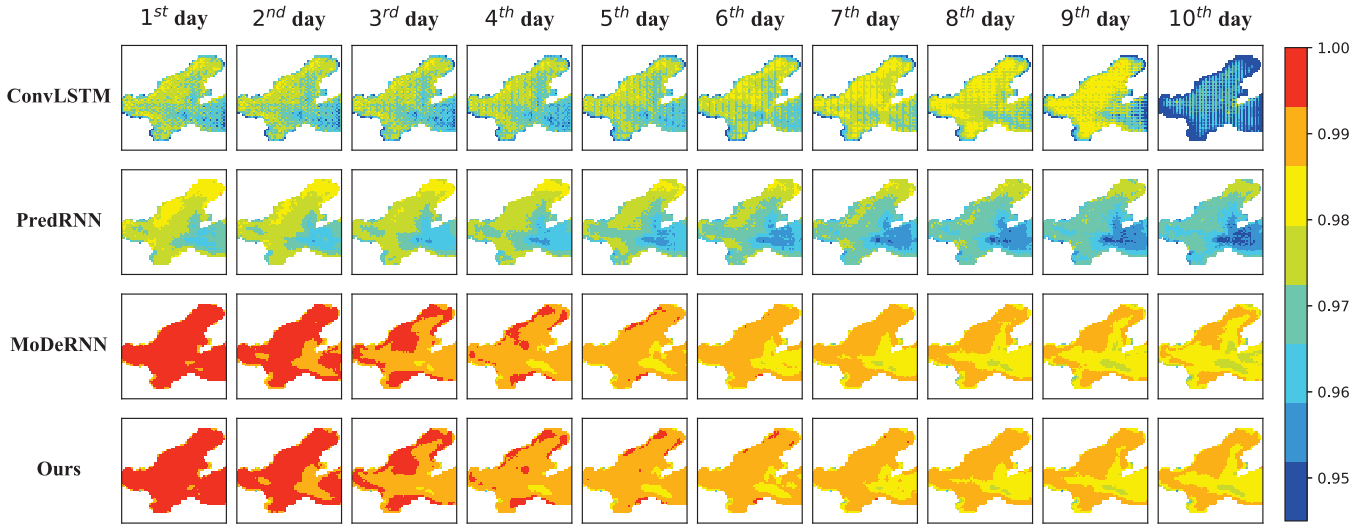

Fig. 4. Spatial distribution of R^2 on all testing samples for different lead times with respect to four neural network models, namely, ConvLSTM, PredRNN, MoDeRNN and our proposed ACFN.

for ACFN remain consistently high, surpassing 0.99, within the central Bohai Sea even for the extended 10-day lead time predictions.

Table I presents the Bohai Sea averaged MAE values for lead times ranging from 1 to 10 days. It is evident that the ACFN model demonstrates a substantial performance advantage over the ConvLSTM and PredRNN models. Specifically, the ACFN model achieves a remarkably low MAE of 0.41°C for a 1-day lead time and a commendable 0.79°C for a 10-day lead time. In contrast, the ConvLSTM and PredRNN models exhibit relatively higher MAE values, with ConvLSTM scoring 1.19°C (1-day lead time) and 1.71°C (10-day lead time), while PredRNN records 1.09°C (1-day lead time) and 1.32°C (10-day lead time). For lead times spanning from 3 to 10 days, ACFN consistently outperforms MoDeRNN, with the extent of improvement in MAE gradually increasing with the extension of lead time.

TABLE I
MAE FOR SST PREDICTION OVER 1- TO 10-DAY LEAD TIMES USING DIFFERENT MODELS.

Models	**day1**	**day2**	**day3**	**day4**	**day5**	**day6**	**day7**	**day6**	**day9**	**day10**
ConvLSTM	1.19	1.20	1.21	1.21	1.20	1.17	1.11	1.10	1.10	1.71
PredRNN	1.09	1.12	1.15	1.18	1.21	1.23	1.25	1.28	1.30	1.32
MoDeRNN	0.41	0.51	0.58	0.63	0.68	0.72	0.75	0.78	0.81	0.84
ACFN	**0.41**	**0.51**	**0.57**	**0.62**	**0.66**	**0.69**	**0.72**	**0.74**	**0.77**	**0.79**

TABLE II
R^2 FOR SST PREDICTION OVER 1- TO 10-DAY LEAD TIMES USING DIFFERENT MODELS.

Models	**day1**	**day2**	**day3**	**day4**	**day5**	**day6**	**day7**	**day6**	**day9**	**day10**
ConvLSTM	0.973	0.973	0.972	0.972	0.972	0.974	0.976	0.976	0.976	0.949
PredRNN	0.977	0.976	0.975	0.973	0.972	0.971	0.970	0.969	0.968	0.967
MoDeRNN	0.996	0.994	0.992	0.991	0.990	0.989	0.988	0.987	0.986	0.985
ACFN	**0.996**	**0.994**	**0.993**	**0.991**	**0.990**	**0.989**	**0.989**	**0.988**	**0.987**	**0.986**

Table II presents a comparison of the Bohai Sea averaged

R^2 among four forecasting models. For a 1-day lead time, ConvLSTM achieves an averaged R^2 of 0.973, while PredRNN attains a slightly higher R^2 of 0.977. Notably, MoDeRNN and ACFN further enhance the R^2 to an impressive 0.996. Across lead times from 1 to 5 days, both MoDeRNN and ACFN consistently maintain R^2 values surpassing 0.99, indicating high predictive accuracy. In the case of an extended 10-day lead time, MoDeRNN and ACFN achieve R^2 values of 0.985 and 0.986, respectively. Although the increment in R^2 when comparing ACFN to MoDeRNN is relatively small at 0.001, it is worth noting that ACFN demonstrates a positive improvement in R^2. Hence, ACFN showcases superior performance compared to ConvLSTM, PredRNN, and MoDeRNN in terms of both averaged MAE and R^2 in the studied area of the Bohai Sea.

V. Conclusion

In this paper, we have introduced an attention-based context fusion network (ACFN) for SST prediction in the Bohai Sea. By incorporating the convolutional block attention module within the detail context block, the proposed model empowers the exploration of intricate spatiotemporal correlations between the previous context state and the current input state in the ConvLSTM framework. Experimental results have shown that ACFN outperforms three state-of-the-art models, namely ConvLSTM, PredRNN and MoDeRNN, in short-term forecast of SST in the studied area in terms of MAE and R^2. Our study underscores the practical utility of ACFN for short-term SST prediction. Additionally, in the forecast experiments, it is worth noting that SST prediction errors tend to be relatively higher in coastal areas compared to deeper shelf regions.

Acknowledgments

This study is supported partially by the NSFC/RGC Joint Research Scheme (Grant no. 62261160387), the Oceanic Interdisciplinary Program of Shanghai Jiao Tong University (Grant no. SL2020MS030), and the Open Fund of State Key Laboratory of Satellite Ocean Environment Dynamics (Grant no. SOEDZZ2206).

References

[1] M. D. Sumner, K. J. Michael, C. J. Bradshaw, and M. A. Hindell, "Remote sensing of southern ocean sea surface temperature: implications for marine biophysical models," *Remote Sensing of Environment*, vol. 84, no. 2, pp. 161–173, 2003.

[2] F. J. Wentz, C. Gentemann, D. Smith, and D. Chelton, "Satellite measurements of sea surface temperature through clouds," *Science*, vol. 288, no. 5467, pp. 847–850, 2000.

[3] A. O. Tarakanov and A. V. Borisova, "Galapagos indicator of El Niño using monthly SST from NASA Giovanni system," *Environmental Modelling & Software*, vol. 50, pp. 12–15, 2013.

[4] J. F. De Paz, J. Bajo, A. González, S. Rodríguez, and J. M. Corchado, "Combining case-based reasoning systems and support vector regression to evaluate the atmosphere–ocean interaction," *Knowledge and Information Systems*, vol. 30, pp. 155–177, 2012.

[5] K. Patil, M. Deo, and M. Ravichandran, "Prediction of sea surface temperature by combining numerical and neural techniques," *Journal of Atmospheric and Oceanic Technology*, vol. 33, no. 8, pp. 1715–1726, 2016.

[6] J. Xie, J. Zhang, J. Yu, and L. Xu, "An adaptive scale sea surface temperature predicting method based on deep learning with attention mechanism," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 5, pp. 740–744, 2019.

[7] Y. Yan, Y. Zhou, Y. Xu, and W. Gu, "Assessment of the spatiotemporal variability of seawater temperature and salinity in the yellow and bohai seas from multiple high-resolution reanalysis datasets," *Ocean Dynamics*, pp. 1–17, 2023.

[8] M. Haghbin, A. Sharafati, D. Motta, N. Al-Ansari, and M. H. M. Noghani, "Applications of soft computing models for predicting sea surface temperature: a comprehensive review and assessment," *Progress in Earth and Planetary Science*, vol. 8, no. 1, pp. 1–19, 2021.

[9] E. Pisoni, F. Pastor, and M. Volta, "Artificial neural networks to reconstruct incomplete satellite data: application to the Mediterranean sea surface temperature," *Nonlinear Processes in Geophysics*, vol. 15, no. 1, pp. 61–70, 2008.

[10] M. Jahanbakht, W. Xiang, and M. R. Azghadi, "Sea surface temperature forecasting with ensemble of stacked deep neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.

[11] K. Zhang, X. Geng, and X.-H. Yan, "Prediction of 3-D ocean temperature by multilayer convolutional LSTM," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 8, pp. 1303–1307, 2020.

[12] L. Wei and L. Guan, "Seven-day sea surface temperature prediction using a 3DConv-LSTM model," *Frontiers in Marine Science*, vol. 9, p. 905848, 2022.

[13] L. Chaudhary, S. Sharma, and M. Sajwan, "Systematic literature review of various neural network techniques for sea surface temperature prediction using remote sensing data," *Archives of Computational Methods in Engineering*, pp. 1–33, 2023.

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

[15] T. Song, J. Jiang, W. Li, and D. Xu, "A deep learning method with merged LSTM neural networks for SSHA prediction," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 2853–2860, 2020.

[16] Z. Zhang, X. Pan, T. Jiang, B. Sui, C. Liu, and W. Sun, "Monthly and quarterly sea surface temperature prediction based on gated recurrent unit neural network," *Journal of Marine Science and Engineering*, vol. 8, no. 4, p. 249, 2020.

[17] C. Xiao, N. Chen, C. Hu, K. Wang, Z. Xu, Y. Cai, L. Xu, Z. Chen, and J. Gong, "A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data," *Environmental Modelling & Software*, vol. 120, p. 104502, 2019.

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

[19] Y. Wang, M. Long, J. Wang, Z. Gao, and P. S. Yu, "Predrnn: Recurrent neural networks for predictive learning using spatiotemporal LSTMs," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[20] Z. Chai, Z. Xu, and C. Yuan, "Modernn: Towards fine-grained motion details for spatiotemporal predictive learning," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 4658–4662.

[21] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.

[22] C. J. Donlon, M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer, "The operational sea surface temperature and sea ice analysis (OSTIA) system," *Remote Sensing of Environment*, vol. 116, pp. 140–158, 2012.