Physics-Guided Generative Adversarial Networks for Sea Subsurface Temperature Prediction

Yuxin Meng, *Student Member, IEEE*, Eric Rigall, *Student Member, IEEE*, Xueen Chen, *Member, IEEE*, Feng Gao, *Member, IEEE*, Junyu Dong, *Member, IEEE*, and Sheng Chen, *Fellow, IEEE*

Abstract—Sea subsurface temperature, an essential component of aquatic wildlife, underwater dynamics, and heat transfer with the sea surface, is affected by global warming in climate change. Existing research is commonly based on either physics-based numerical models or data-based models. Physical modeling and machine learning are traditionally considered as two unrelated fields for the sea subsurface temperature prediction task, with very different scientific paradigms (physics-driven and data-driven). However, we believe that both methods are complementary to each other. Physical modeling methods can offer the potential for extrapolation beyond observational conditions, while data-driven methods are flexible in adapting to data and are capable of detecting unexpected patterns. The combination of both approaches is very attractive and offers potential performance improvement. In this article, we propose a novel framework based on a generative adversarial network (GAN) combined with a numerical model to predict sea subsurface temperature. First, a GAN-based model is used to learn the simplified physics between the surface temperature and the target subsurface temperature in the numerical model. Then, observation data are used to calibrate the GAN-based model parameters to obtain a better prediction. We evaluate the proposed framework by predicting daily sea subsurface temperature in the South China Sea. Extensive experiments demonstrate the effectiveness of the proposed framework compared to existing state-of-the-art methods.

Index Terms—Generative adversarial network (GAN), numerical ocean model, ocean physical laws, sea subsurface temperature, sea surface temperature (SST).

I. INTRODUCTION

SEA subsurface is the part of the ocean below the sea surface. Its temperature plays an important role in ocean science research [1]. The sea subsurface temperature is

important information for understanding the global ocean ecosystem and Earth climate system. The study of the spatial and temporal distribution of sea temperature and its variation law is not only a critical issue in marine geography but also of considerable significance to the fishery, navigation, and underwater acoustics. Diverse sources of external factors, such as radiation and diurnal wind, affect the sea subsurface temperature, and the prediction of the sea subsurface information is very challenging [2]. Existing studies on sea subsurface temperature rely on numerical modeling and observational data [3]–[6].

Numerical modeling is a widely used technique to tackle complex ocean problems by data simulation based on the equations of ocean physical laws. Currently, the Princeton ocean model (POM) [7], the HYbrid coordinate ocean model (HYCOM) [8], and the finite-volume coastal ocean model (FVCOM) [9] are commonly used in oceanography. POM is a classic traditional ocean model with a clear structure, concise model specifications, and thorough model physical interpretation. The flexible vertical hierarchical structure of HYCOM makes it more suitable for the significant expansion of the stratification effect. The FVCOM model includes the momentum equation, the continuity equation, the thermosalt conservation equation, and the state equation. The numerical solution of FVCOM adopts the finite volume method (FVM), which has the advantages of accurate and fast calculation and good fitting of coastline boundary and seabed topography based on the unstructured mesh. This is because FVM can better guarantee the conservation of each physical quantity not only in each unit but also in the whole calculation area. All these numerical models are constructed based on our knowledge of ocean physics, and they are often applied to simulate ocean dynamics and predict sea subsurface temperature. However, their prediction accuracy can hardly be guaranteed since there exists a large range of environmental factors that affect marine environments.

In order to improve the prediction accuracy of the numerical models, assimilation methods are commonly used. Traditional assimilation methods can improve the model prediction performance by fusing new observational data in the dynamic running process of a numerical model. Smedstad and O'Brien [10] summarized the data assimilation methods developed before 1991 and classified them into polynomial interpolation methods, optimal interpolation methods, and variational analysis methods. Anderson *et al.* [11] also surveyed the data assimilation methods in Physical Oceanography. Although the prediction accuracy of the traditional

assimilation methods is much higher than that of the numerical models, there are ample rooms that these methods can be further improved.

In contrast to the physics-based numerical models, data-driven models, such as neural networks, rely purely on observational data to learn the underlying data distribution. However, it is unclear how these models produce specific decisions, and interpreting these data-driven models physically is very difficult. Since these methods only rely on training data, their generalization ability on unseen data is often limited, whereas most physics-based models do not utilize training data and, therefore, may perform well on unseen data, provided that the physical laws employed to build these models accurately represent the underlying data distribution. Nevertheless, the physical rules are often incomplete, and these numerical models need to be improved and supplemented.

A fundamental principle in data modeling is to incorporate available *a priori* information regarding the underlying data generating mechanism into the modeling process. Data-physics hybrid models that are capable of incorporating prior knowledge typically outperform data-driven modeling [12], [13]. Motivated by this fundamental principle for data modeling, in this article, we focus on developing a physics-guided framework for training a neural network to predict sea subsurface temperature, which combines numerical modeling and observational data modeling. We demonstrate that this data-physics hybrid modeling approach can not only take advantage of our prior knowledge of ocean physical laws but also improve the overall prediction accuracy.

In recent years, deep learning in computer vision [14]–[16] and natural language processing [17]–[19] have achieved breakthrough progress. Its underlying motivation is to simulate the human brain neural connection structures [20]–[23]. When handling high-dimensional data, high-level features are extracted through multiple layers progressively to identify the concepts relevant to humans [24]–[26]. Deep learning models can be roughly divided into two categories: discriminant models and generative models [27], [28]. Discriminant models are trained to distinguish the correct output among possible output choices [29], [30]. On the other hand, generative models are trained to obtain better understandings of the data samples. Specifically, a generative model learns a distribution from the input samples and then generates similar samples based on this distribution to enhance the model. Goodfellow *et al.* [31] proposed the generative adversarial network (GAN), which uses adversarial training to train a generative network and a discriminative network jointly. The generative network captures the potential distribution of the real data, while the discriminative network is commonly a binary classifier, which judges whether the input samples are real or not. Many GAN-based models have been proposed to solve the problem of high-quality image generation. Isola *et al.* [32] proposed Pix2pix for image translation. In Pix2pix, a pair of image datasets from different domains is fed into the model during training, and an image can be transformed from one domain to the other. Zhu *et al.* [33] proposed CycleGAN to learn mappings between an input image and an output image when
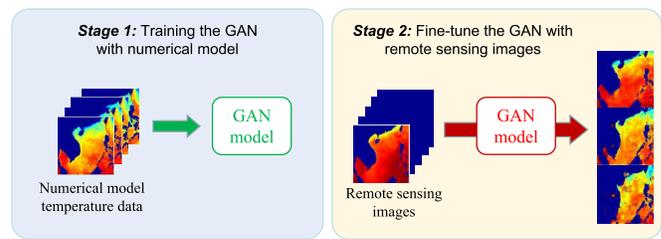

Fig. 1.   Two-stage sea subsurface temperature prediction framework. In the first stage, generative adversarial training is performed on the model with the data from the physics-based numerical model. In the second stage, the model is fine-tuned with observational data.

paired training data are unavailable. A cycle consistency loss is introduced to achieve this goal.

The deep neural network has strong predictive power, but it does not follow the laws of physics. By contrast, a numerical model simulates the ocean dynamics based on the knowledge of ocean physics. Karpatne *et al.* [34] blended the numerical model with multilayer perceptron to correct lake temperature. In this work, the authors applied all the variables related to the lake temperature and the output of the numerical model for the lake temperature as the inputs to the neural network. If the numerical model accurately simulates the motion of the lake temperature, the output of their model is generated by the numerical model; otherwise, the result is generated by the neural network. This approach basically chooses the result from either the physics-based numerical model or the neural network trained by observation data. Ideally, we would like to design a prediction method by combining both the physics-based numerical model and the data-driven model. Jia *et al.* [35] combined a recurrent neural network (RNN) model with the numerical model to predict the lake temperature. Their model was trained over the numerical model data and then fine-tuned on the limited observation data. However, their model was applied for each depth separately, and the data from the same depth are used to predict the lake temperature of the same depth. In addition, they only predicted the temperature value at one subsurface point, not over an entire area. We also note that most existing studies concentrate on sea surface prediction, while there is a paucity of contributions on the daily sea subsurface temperature prediction. This will be further discussed in the related work section.

In this article, to tackle the abovementioned limitations in the existing sea temperature analysis literature, we propose a new framework to predict the sea subsurface temperature by combining the physics-based numerical model with deep neural networks. In our method, we apply the physics-based numerical model to train the neural network model in the first phase, and then, observational data are used to calibrate the model parameters in the second phase. More specifically, we design two neural networks in the proposed framework, as illustrated in Fig. 1. The first network learns the simplified physics laws from the numerical model. The weights of this first network are shared by the second network. This effectively encodes the knowledge of ocean physics into this second network model, and its weights are then fine-tuned by observational data. It can be seen that the merits of

both physics-based numerical modeling and observational data modeling approaches are combined, and consequently, the prediction accuracy is further enhanced. The main contributions of this article are summarized as follows.

1) A novel GAN-based framework is proposed, which predicts the daily sea subsurface temperature by learning the relationship between sea surface temperature (SST) and subsurface temperature.
2) We explore the use of GAN combined with the physics-based numerical model for building a hybrid prediction model incorporating more effectively the known ocean physics with the observational data information.
3) We propose a physics-based loss with a mask as prior knowledge. The mask filters out land locations, and this loss automatically encodes the knowledge of ocean physics into the modeling process, leading to prediction performance improvement.

The rest of the article is organized as follows. Section II presents the background of GAN models and sea temperature prediction. Section III details the proposed framework for sea subsurface temperature prediction. The experimental results are reported in Section IV. We draw concluding remarks and discuss the future work in Section V.

II. RELATED WORKS

A. Generative Adversarial Networks

Inspired by the binary zero-sum game, Goodfellow *et al.* [31] proposed GAN, in which two neural networks contest each other in a game. More specifically, GAN is composed of two networks: a generative network G and a discriminative network D. The generator G iteratively learns the distribution of the real input samples, and it generates samples following the learned distribution. The generated fake samples are then fed into the discriminator D, and D is trained to judge whether the input samples are real or fake.

In the training process, the generator G learns the input data distribution. During this learning process, fake samples can be identified by the discriminator D from the real data distribution. In such adversarial learning, the generator G tries to "fool" the discriminator D by producing samples as similar as possible to the real samples. With this mutual competitive reinforcement, the performances of both G and D are jointly enhanced.

The conditional generative adversarial network (CGAN) [36] is an extension of GAN in which a conditional setting is applied. In CGAN, both the generator G and discriminator D are conditioned on class labels. As a result, the model can learn mappings from inputs to outputs by feeding it with contextual information.

Yang *et al.* [37] solved the stochastic differential equations by encoding the known physical laws into the GAN. Lütjens *et al.* [38] used GAN to learn the latent features of the numerical model data in order to generate more realistic coastal floor data. Zheng *et al.* [39] reconstructed the image based on its known pixels by employing a GAN model. These works used the GAN model to learn the latent features

from the numerical model. Then, they applied the pretrained GAN model to do the corresponding tasks. In other words, these works used the GAN models to replace part or the entire numerical model. The works [37]–[39] highlight the potential application of the GAN model in physical-relevant tasks. However, the difference between these works with our hybrid physics-data-based GAN is huge. Not only we pretrain the GAN with the physics-based numerical model but also we adopt the observational data to calibrate the pretrained GAN model. In other words, our GAN model not only learns the physical laws from the numerical model but also adapts itself using observational data.

B. Sea Subsurface Temperature Prediction

Temperature is an important factor in marine hydrology and climate change [40]. Existing studies based on satellite remote sensing data mainly focus on SST and assessment. Yang *et al.* [41] consider the task of sea temperature prediction as a sequence prediction problem and build an end-to-end trainable long short-term memory (LSTM) neural network model. Then, the temporal and spatial features are combined to predict sea temperature. Wei *et al.* [42] used Ice Analysis (OSTIA) data to train a neural network for South China Sea temperature prediction. Deep learning-based methods have also been utilized to predict the SST in the Bohai Sea and Indian Ocean [43]–[45].

The abovementioned studies mainly focus on temperature prediction of the sea surface. However, the sea subsurface temperature prediction research is scarce. Han *et al.* [46] applied the convolutional neural network (CNN) to predict the subsurface temperature from sets of the remote sensing data. Lu *et al.* [47] adopted the preclustered neural network method to estimate the subsurface temperature, and the results are better than those obtained without clustering. Wu *et al.* [48] used the self-organizing map neural network to predict the subsurface temperature anomaly in the North Atlantic. These methods can reliably predict the monthly subsurface temperature using a neural network due to the fact that sufficient monthly observational data of the subsurface temperature are available for training neural network models. However, due to the very limited daily observation data, the prediction of the daily subsurface temperature cannot be carried out efficiently and accurately only using neural networks. Zhang *et al.* [49] used monthly Argo data to predict the sea subsurface temperature, but no physics-based numerical model was utilized in this monthly sea subsurface temperature prediction model. These works indicate the lack of research on daily subsurface temperature prediction. In this article, we combine deep neural networks and a physics-based numerical model into a unified framework, which is capable of predicting the daily sea subsurface temperature.

III. PROPOSED FRAMEWORK

The proposed framework, as depicted in Fig. 2, composes of two stages: 1) generative adversarial pretraining on numerical model data and 2) fine-tuning of the GAN model with observational data. In the first stage, the generator learns the
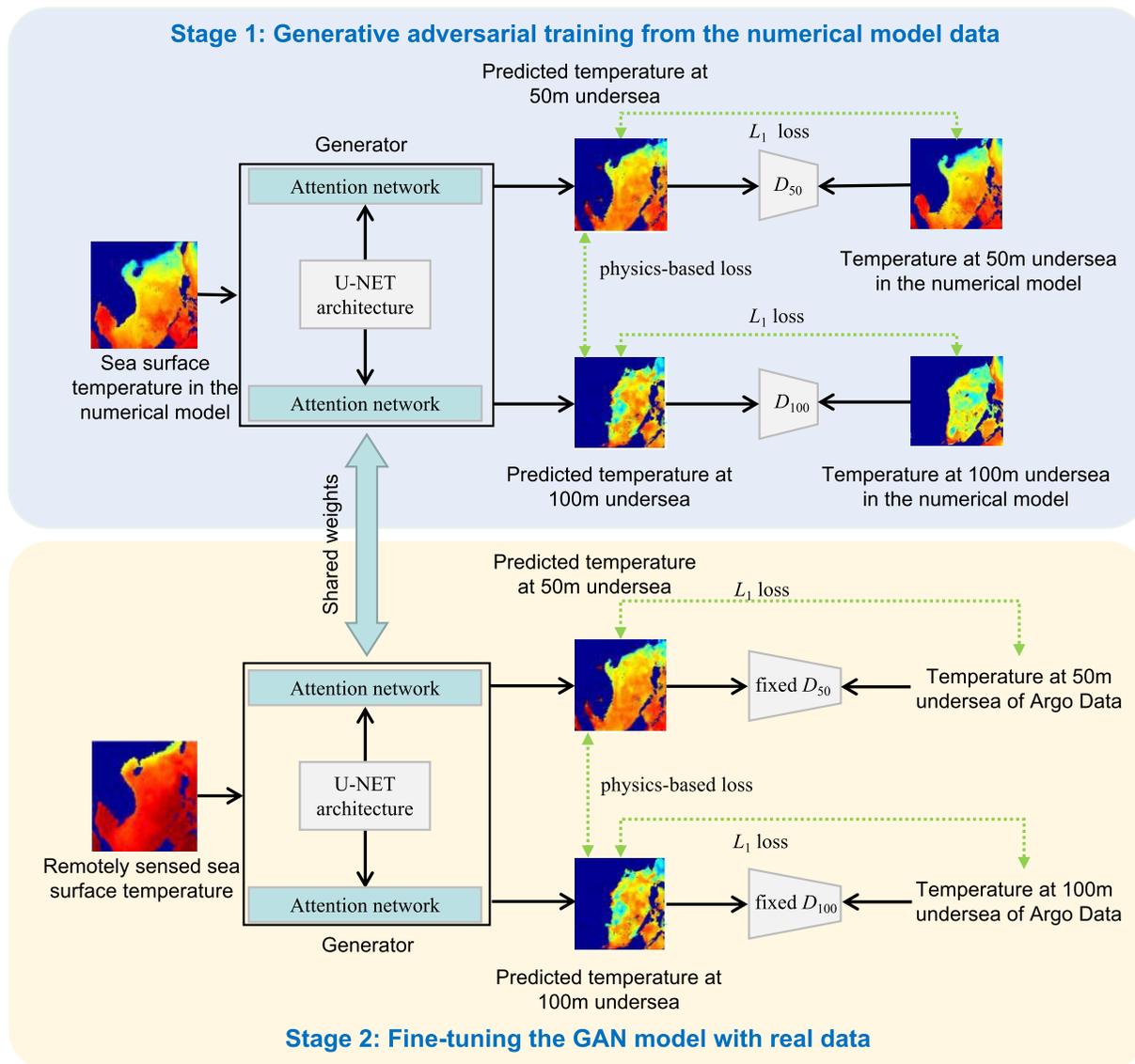
Fig. 2.   Proposed GAN-based sea subsurface temperature prediction framework. Stage 1: generator learns the mapping from the sea surface temperature (SST) to the target depth temperature in the numerical model. The generator is composed of two components: one single shared network and several task-specific attention networks. The shared network learns the mapping from SST and random noise to the numerical model data. The task-specific attention networks capture the mapping between the SST and the sea subsurface temperature. Stage 2: fine-tuning the GAN model with observational data. The weights of the generator are shared with Stage 1, and the weights of the discriminator are fixed.

mapping from the SST to the target depth temperature using numerical model data. This effectively encodes the knowledge of ocean physics into the neural network model. In the second stage, real-world observation data are used to fine-tune the weights of the neural network model. This enables the model to learn the real-data distribution and compensate for the physics knowledge missing in the numeral model. We now detail these two stages.

A. Stage 1: Generative Adversarial Training on Numerical Model Data

Numerical models play an important role in understanding the ocean's influence on global climate. They simulate the ocean properties and circulation based on the equations of ocean physics laws. Since numerical models approximate the

physical correlations among different depths of the ocean, we use a GAN model in the proposed framework to acquire these relationships from the data generated by a numerical model.

Without loss of generality, we consider predicting the subsurface temperatures at 50-, 100-, and 150-m underwater simultaneously. The prediction tasks of different depth temperatures can be achieved jointly by multitask learning. In order to obtain good performance for each task, attention modules are used to enable both the task-shared and task-specific feature learning in an end-to-end manner [50]. The generator architecture is depicted in Fig. 3, which is comprised of multiple sets of attention modules and the U-NET architecture. Each set of attention modules can learn the features for individual tasks. Specifically, each attention module learns
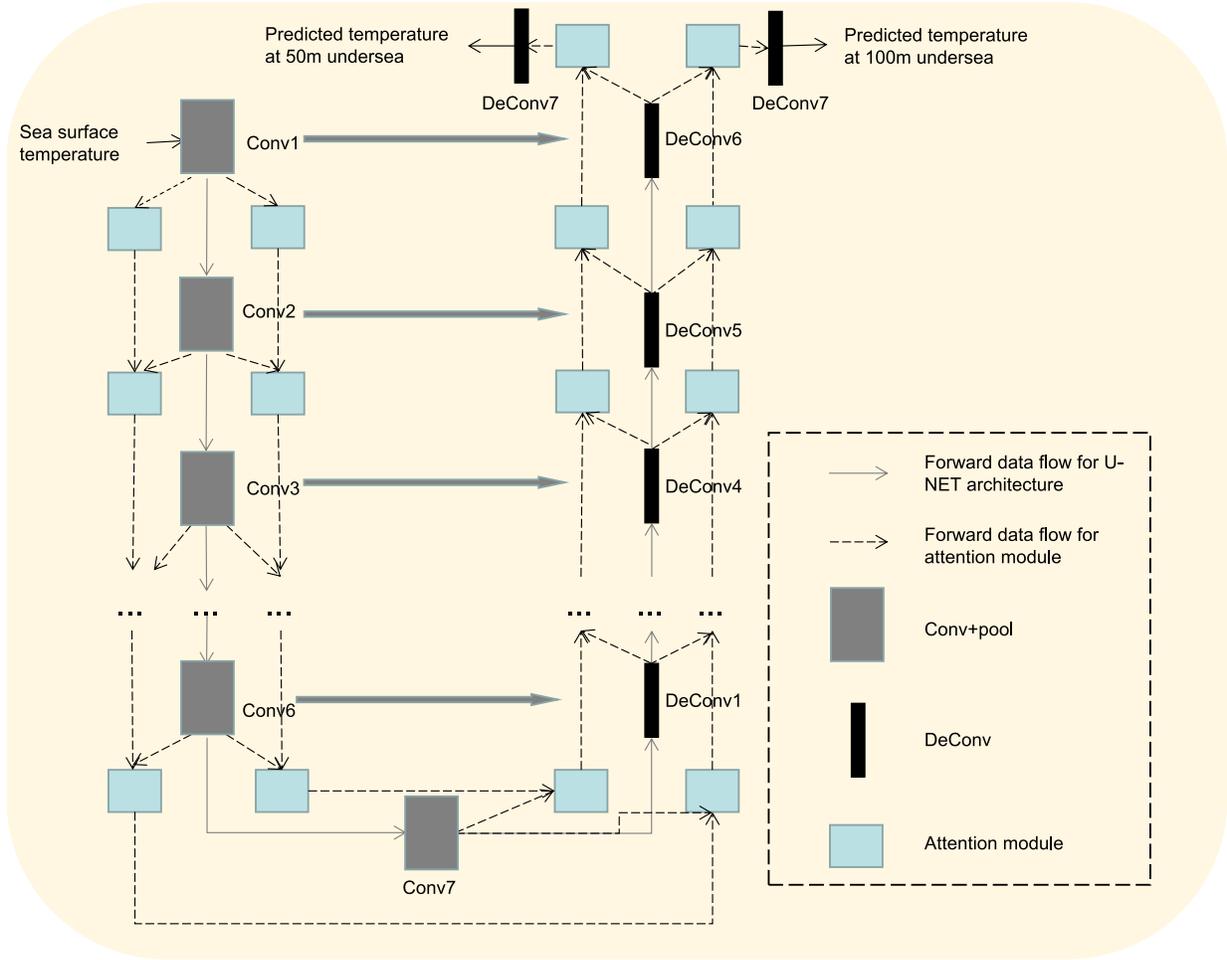
Fig. 3. Illustration of the generator architecture. The generator comprises the U-NET architecture and the two sets of attention modules. The attention module is connected with the output of the convolution block and the attention module from the last layer, which is passed one by one.

a soft attention mask, which is dependent on the features in the shared network. The features in the shared network and the soft attention masks can be trained jointly to optimize the generalization of the features across different tasks.

As shown in Fig. 4, the shared features after pooling are denoted as p, and the learned attention mask in the layer for task i is denoted as a_i. The task-specific features $\hat{a}_i$ are computed by elementwise multiplication of the attention mask with the shared features as $\hat{a}_i = a_i \odot p$, where $\odot$ denotes the elementwise multiplication operator. The attention module has strong capabilities of emphasizing nontrivial features and weakening unimportant ones. Moreover, as the seawater temperature generally decreases with the increase of depth, we exploit this fact and set it as prior knowledge. If the seawater temperature in a lower layer is estimated higher than the one in an upper layer, the model is penalized. Hence, we apply this physics-based loss to guide the fitting ability of the model between different depths.[1]

As mentioned in Section II-A, a GAN model is composed of two networks: the generative network G and the discriminative

[1]In some high latitude oceanic regions, seawater temperature at 50 m can actually be higher than SST. In this case, the physics-based loss should not be applied to this first underwater layer.
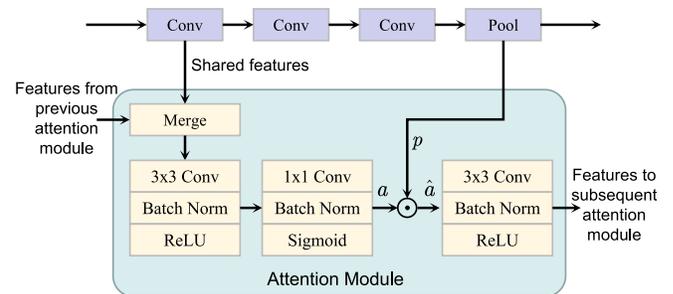
Fig. 4. Illustration of the attention module.

network D. In our model, the generator contains two parts: one single shared network and three task-specific attention networks. The shared network uses a conditional GAN model that learns a mapping from the observed image x and random noise z to real image y. The network objective is defined as follows:

$$L_{S_1}(G, D) = E_{x,y} \log D(x, y)$$
$$+ E_{x,z} \log(1 - D(x, G(x, z))) \quad (1)$$

where $E_{x,y}$ denotes the expectation operator with respect to x and y, $D(x, y)$ distinguishes whether x and y are the

true paired data, and $G(x, z)$ learns the mapping from the data x and random vector z to the real data y. As can be observed in (1), in the shared network, the generator G tries to minimize the objective while D tries to maximize it. In the generator model, the input noise z and conditional information x jointly constitute the joint hidden layer representation in order to model the same distribution with domain y. To further improve the prediction performance, we mix the conditional GAN objective with an L_1 distance, which is defined as

$$L_{L_1}(G) = E_{x,y,z} \|G(x, z) - y\|_1. \qquad (2)$$

Besides the shared network, we build three task-specific attention networks, $G_{0 \to 50}$, $G_{0 \to 100}$, and $G_{0 \to 150}$, to capture the mappings between the SST and the undersea temperatures at 50, 100, and 150 m, respectively. Correspondingly, the discriminative network can be decomposed into three subdiscriminative networks, namely, $D = \{D_{50}, D_{100}, D_{150}\}$.

In our implementation, the SST x_0 is obtained from the HYCOM data [8]. Besides x_0, we generate three masks, M_0, M_{50}, and M_{100}. Starting from M_0, its value at a given location is set to 1 if the SST is available from the numerical model at this location, and the value is set to 0 if the temperature is not exploitable, e.g., the location is on the land. This mask can filter out noise regions, such as the land. We further set the margin to 0.1. If the temperature of the deep layer is $0.1°$ higher than that of the shallow layer, the model is penalized. Specifically, we define an objective function $L_{0 \sim 50}(G)$ as follows:

$$L_{0 \sim 50}(G)$$
$$= E_{x,z} \|\max\{(G_{0 \to 50}(x_0, z) \odot M_0 - x_0 \odot M_0), 0.1\}\|_1. \quad (3)$$

The purpose of the mask M_0 can be seen clearly from the objective function $L_{0 \sim 50}(G)$. Only when the temperature at the 0-m depth is exploitable, i.e., this location is not on land, the comparison between the temperature at the -50-m depth and the temperature at the 0-m depth is meaningful. Similarly, we have the mask M_{50}, whose value at a location is set to 0 if the numerical model data indicates that this 50-m depth location is on the land; otherwise, the temperature at this location is exploitable, and we set $M_{50} = 1$. Likewise, we can calculate M_{100}. Hence, we can define the objectives $L_{50 \sim 100}(G)$ and $L_{100 \sim 150}(G)$, respectively, as

$$L_{50 \sim 100}(G) = E_{x,z} \|\max\{(G_{0 \to 100}(x_0, z) \odot M_{50}$$
$$- G_{0 \to 50}(x_0, z) \odot M_{50}), 0.1\}\|_1 \quad (4)$$

$$L_{100 \sim 150}(G) = E_{x,z} \|\max\{(G_{0 \to 150}(x_0, z) \odot M_{100}$$
$$- G_{0 \to 100}(x_0, z) \odot M_{100}), 0.1\}\|_1. \quad (5)$$

Based on the above three objective functions, we propose the physics-based loss by using the three masks as prior knowledge, which leads to improved prediction performance. Hence, the physics-based loss in Stage 1 is defined as

$$L_{P_1}(G) = L_{0 \sim 50}(G) + L_{50 \sim 100}(G) + L_{100 \sim 150}(G). \quad (6)$$

It can be seen that this physics-based loss applies "shallow sea temperature" as masks (SL masks). Specifically, when comparing the temperature difference of a deeper layer and the shallower layer, the mask is referencing the shallower-layer sea temperature. It is natural to ask whether we can adopt "deeper sea temperature" as masks (DP masks). That is, when comparing the temperature difference of a deeper layer and the shallower layer, the mask is referencing the deeper layer sea temperature. Adopting DP masks in the proposed physics-based loss corresponds to replacing M_0 in the loss (3) with M_{50} and replacing M_{50} in the loss (4) with M_{100}, as well as calculating the mask M_{150} and using it to replace M_{100} in the loss (5). This DP mask approach, however, is less effective than the SL mask approach. This is because the land area at the 0-m depth is smaller than that at the -50-m depth, and the land area at the -50-m depth is smaller than that at the -100-m depth, and so on. Therefore, the exploitable values of M_0 ($M_0 = 1$) or the size of M_0 is much larger than that of M_{50}, the size of M_{50} is much larger than that of M_{100}, and the size of M_{100} is much large than that of M_{150}. Hence, adopting the SL mask approach enables the model to exploit larger sea subsurface areas. In the ablation study of Section IV-C, we will demonstrate that better performance can be obtained by adopting the SL mask approach than the DP mask approach.

By employing a physics-based loss, the generator can learn the mapping from the SST to the temperature at 50-, 100-, and 150-m undersea from the numerical model simultaneously. According to the prior knowledge, the SST should be higher than the one at 50-m undersea, which should be higher than the temperature at 100-m underwater, and so on. If there is some irregular data, the penalty term will be added to the training process.

According to the above discussion, the full objective function in the first stage of generative adversarial training using the physics-based numerical model data is expressed as

$$L(G, D) = L_{S_1}(G, D) + L_{L_1}(G) + L_{P_1}(G). \quad (7)$$

Algorithm 1 implements the first stage of the training process in our proposed method. The weights of the discriminators and the generator are updated based on the costs (1) and (7) separately. In our implementation, the first $n_1 = 100$ epochs maintain a constant learning rate of $l_1 = 0.0002$, followed by another $n_2 = 100$ epochs with a linearly decaying learning rate whose decaying factor η satisfies $0 < \eta < (l_1/n_2)$. This setting is the same as the original Pix2Pix method [32].

B. Stage 2: Fine-Tuning GAN Model With Observation Data

Since numerical models rely heavily on simplified physics law, their results sometimes exhibit discrepancies from the observed data. Therefore, we utilized remotely sensing data, Argo data [51], to correct numerical data errors.

As illustrated in Fig. 2, AVHRR SST data [52] are fed as the input of the model, while Argo data are employed as the real data. The generator shares the weights with the model from the first stage, while the weights of the discriminators are fixed. The generator in the second stage is composed of one single shared network and two task-specific attention networks. The objective function of the shared network is given

Algorithm 1 Stage I Training Procedure

Require: HYCOM model training data x, y, random noise vector z, sea temperature masks M_0, M_{50} and M_{100}, initial learning rate l_1, learning rate decaying factor η, numbers of critic iterations n_1, n_2

Require: Initial generator parameters θ_g, initial discriminator parameters $\theta_d = \{\theta_d^i\}_{i \in \{50,100,150\}}$

Ensure: Generator G and discriminator $D = \{D_i\}_{i \in \{50,100,150\}}$

1: **while** not converged **do**
2:     Set learning rate to $l = l_1$;
3:     **for** $t = 0, \ldots, n_1$ **do**
4:         Sample image pair $\{x_0^i\}_{i=1}^N$ and $\{y_{50}^i\}_{i=1}^N$, $\{x_0^i\}_{i=1}^N$ and $\{y_{100}^i\}_{i=1}^N$, $\{x_0^i\}_{i=1}^N$ and $\{y_{150}^i\}_{i=1}^N$;
5:         Update D by gradient descent based on cost (1);
6:         Update G by gradient descent based on cost (7);
7:     **end for**
8:     **for** $t = n_1 + 1, \ldots, n_1 + n_2$ **do**
9:         Sample image pair $\{x_0^i\}_{i=1}^N$ and $\{y_{50}^i\}_{i=1}^N$, $\{x_0^i\}_{i=1}^N$ and $\{y_{100}^i\}_{i=1}^N$, $\{x_0^i\}_{i=1}^N$ and $\{y_{150}^i\}_{i=1}^N$;
10:        Update D by gradient descent based on cost (1);
11:        Update G by gradient descent based on cost (7);
12:        Update learning rate $l = l_1 - \eta(t - n_1)$;
13:     **end for**
14: **end while**

Algorithm 2 Stage II Training Procedure

Require: Remote sensing satellite training data x, Argo training data Argo_{50}, Argo_{100} and Argo_{150}, random noise vector z, sea temperature masks M_{50} and M_{100}, initial learning rate l_1, learning rate decaying factor η, numbers of critic iterations n_1, n_2

Require: Generator parameters θ_g and discriminator parameters $\theta_d = \{\theta_d^i\}_{i \in \{50,100,150\}}$

Ensure: Generator G

1: **while** not converged **do**
2:     Set learning rate to $l = l_1$;
3:     **for** $t = 0, \ldots, n_1$ **do**
4:         Sample image pair $\{x_0^i\}_{i=1}^N$ and $\{\text{Argo}_{50}^i\}_{i=1}^N$, $\{x_0^i\}_{i=1}^N$ and $\{\mathbf{Argo}_{100}^i\}_{i=1}^N$, $\{x_0^i\}_{i=1}^N$ and $\{\text{Argo}_{150}^i\}_{i=1}^N$;
5:         Update G by gradient descent based on cost (11);
6:     **end for**
7:     **for** $t = n_1 + 1, \ldots, n_1 + n_2$ **do**
8:         Sample image pair $\{x_0^i\}_{i=1}^N$ and $\{\text{Argo}_{50}^i\}_{i=1}^N$, $\{x_0^i\}_{i=1}^N$ and $\{\text{Argo}_{100}^i\}_{i=1}^N$, $\{x_0^i\}_{i=1}^N$ and $\{\text{Argo}_{150}^i\}_{i=1}^N$;
9:         Update G by gradient descent based on cost (11);
10:        Update learning rate $l = l_1 - \eta(t - n_1)$;
11:     **end for**
12: **end while**

as follows:

$$L_{S_2}(G) = E_{x,z} \log(1 - D(x, G(x, z))) \qquad (8)$$

where the discriminator D does not update its weights, and only the generator updates its parameters through backpropagation. In this stage, the real data are Argo data. As Argo data contain the temperature information at a single location, we cannot use Argo data to train the discriminator. Instead, we have to fix the discriminator in order to predict the temperatures on the entire area, not at a point location like Argo data. Since daily Argo data are point data, to predict values from the point to the entire plane, the discriminator pretrained in the first stage is used to measure the differences between the generated samples and the real data according to

$$L_{dot}(G) = E_{x,z} \| G(x_{i,j}, z) - \text{Argo}_{i,j} \|_1 \qquad (9)$$

where the index pairs i and j denote the locations of the temperature values from Argo data. Since daily Argo data only contain one temperature value, we employ L_1 distance to measure the temperature error between Argo data and the corresponding generated sample in (9). By doing this, the adjustment from point to plane can be achieved.

It should be noted that, in the second stage, two task-specific attention networks are employed. Due to the imprecision of AVHRR SST data, the temperature difference between the sea surface and 50-m undersea is not taken into account. In the experiment section, this will be fully explained. Therefore, the physics-based loss in Stage 2 is defined as

$$L_{P_2}(G) = L_{50\sim100}(G) + L_{100\sim150}(G) \qquad (10)$$

where $L_{50\sim100}(G)$ and $L_{100\sim150}(G)$ use the same configurations as the corresponding objective functions in the first stage.

The full objective function employed in the second stage is, therefore, given by

$$L(G) = L_{S_2}(G) + L_{dot}(G) + L_{P_2}(G). \qquad (11)$$

Algorithm 2 implements the second stage of the training process in our method.

IV. EXPERIMENTS

A. Study Area and Data

The study was conducted on the South China Sea, a marginal sea in the western Pacific Ocean, located in the south of Mainland China. The sea has an area of about 3.5 million square kilometers, with an average depth of 1212 m and a maximum depth of 5559 m. A typical study area of (3.99 °N–24.78 °N, 98.4 °E–124.4 °E) was selected.

The numerical model data, satellite remote sensing data, and Argo data from May 2007 to November 2017 were used for training. The remote sensing data from January 2004 to April 2007 were employed as the test input data. The Argo data from January 2004 to April 2007 were used as the true values for the comparison with the predictions, i.e., in the testing, the predicted results are compared with the Argo data.

The numerical model data used in our experiments are HYCOM from [54]. The HYCOM data format is NetCDF, and its spatial resolution is $1°/12° \times 1°/12°$. The data are configured with 32 layers in the vertical direction.

The National Oceanic and Atmospheric Administration (NOAA) optimum interpolation SST (OISST) data from [52] is used in this article. The spatial resolution of the SST data is $0.25° \times 0.25°$, and daily mean data are employed in our study.

The Argo data employed in our study are collected from [51]. The Argo data are composed of the data collected from different buoys placed at different locations in the South China Sea. The daily Argo data are sea subsurface temperature data acquired at only one point in the whole sea area. As the Argo data are point data, we randomly choose one point from the predicted temperature results at the target locations to compare with the true value of the Argo data at the same locations.

More specifically, the sea temperatures of the numerical model data at the depths of 0, -50, -100, and -150 m are used for the first training stage. In the second training stage, we train the model over the satellite remote sensing data and the Argo data at the depths of -50, -100, and -150 m. The input data for the first stage are constructed in the format: $[3856, 128, 128, 1]$, where the first number is the size of the training dataset, the next two numbers are the height and the width of the input data, respectively, and the last number represents the gray-scale map with one color channel. Similarly, the format for the input data is $[2020, 128, 128, 1]$ in the second training stage, after removing the bad quality Argo data. In the test stage, the formats of the input data and the output data are $[180, 128, 128, 1]$ and $[540, 128, 128, 1]$, respectively, where the output data include equal numbers of data samples for the sea subsurface temperature at 50, 100, and 150 m.

B. Baseline Models and Evaluation Metrics

To the best of our knowledge, this article is the first to predict daily sea subsurface temperature by using methods other than numerical models. Due to the sparsity of the observational sea subsurface temperature data for training, it is not feasible to predict the temperature of a whole ocean area by solely relying on a neural network. Therefore, in our experimental evaluation, we combine the neural network methods with the numerical model and the traditional data assimilation approach to perform the study. Since there are only a few Argo devices in the entire China South Sea, the daily sea temperature can only be obtained in a small set of data points in the entire region. Thus, our method adopts numerical model data to do predictions first due to limited observational data, as the numerical model can simulate ocean dynamics and obtain sea temperature in the entire region. Then, we use the set of observational data to fine-tune the model. In the experiments, when we need to compare with other state-of-the-art methods, we also train the model with numerical model data using those methods, and then, observational data are applied for fine-tuning. The data generated by the numerical model assimilation method are obtained from [54]. This HYCOM assimilation data have a spatial resolution of $1°/12° \times 1°/12°$, a temporal resolution of one day, and a vertical resolution from the sea surface to 5000-m undersea. It is much closer to the observational data compared to the HYCOM model data. We compare these data with our method in the following experiment part.

Furthermore, the following neural network methods are selected as the baselines to compare with our model: Pix2pix [32], CycleGAN [33], and PGNN [34]. For Pix2pix

TABLE I
STUDY ON THE MULTITASK LEARNING

Model	RMSE (°C)		
	50m	100m	150m
Model without TANs	0.9532	1.3265	1.2475
Model with TANs	0.9435	1.3067	1.2439
Model	R^2		
	50m	100m	150m
Model without TANs	0.5431	0.3129	0.5410
Model with TANs	0.5437	0.3374	0.5514

and CycleGAN, we use the publicly available source codes provided by the authors, with the same default parameters. Specifically, for Pix2pix, $\lambda = 100$ and 70×70 PatchGAN are employed, as mentioned in [32]. For CycleGAN, an Adam solver [53] is with a learning rate of 0.0002. For PGNN, its output comes from either the neural network or the numerical model. However, having an output solely relying on a pure neural network is unsuitable for daily sea subsurface prediction over the whole sea area. Therefore, we cannot directly compare the PGNN with our method. Since PGNN uses a physics-based loss to guide the training of its neural network, in our experiments, we compare the physics-based loss obtained by PGNN with the physics-based loss obtained by our method. In addition, we also compare our method with the methods of [35] and [49].

The two evaluation criteria, the root mean square error (RMSE) and the coefficient of determination (R^2) [55], are used to assess the performance of the compared methods.

C. Experiment Design and Ablation Study

All our experiments are implemented on an NVIDIA GeForce 2080Ti GPU. Training iterations and learning rates are the same for both phases of our approach. We train our model for $n_1 + n_2 = 200$ epochs. The first $n_1 = 100$ epochs maintain a constant learning rate of 0.0002, followed by another $n_2 = 100$ epochs with a linearly decaying learning rate. The main network of the generator adopts a U-NET architecture [56], and each convolution is followed by an attention module. The discriminator applies the same six-layer convolutional network as in pix2pix [32]. We construct the data as 128×128 square-shaped heatmaps. Due to the inconsistency of Argo data underwater position, the 1-D interpolation method was applied to obtain the data of 50-, 100-, and 150-m underwater. We use the Z-score standardization method to preprocess the data.

We perform an extensive ablation study to demonstrate the effectiveness of multitask learning and physics-based loss. The influence of different margin values in physics-based loss is also studied. Moreover, the temperature difference between the sea surface and 50-m undersea is analyzed in detail.

1) Effectiveness of Multitask Learning: Multitask learning exploits the correlation among different tasks to promote each other, and consequently, the performance of the whole model is enhanced. We add multiple task-specific attention networks (TANs) to learn the mappings from the SST to 50-, 100-, and 150-m undersea simultaneously. Table I illustrates

TABLE II
STUDY ON THE MASK IN PHYSICS-BASED LOSS

Method	RMSE (°C)		
	50m	100m	150m
NO mask	0.9480	1.3663	1.1977
DP mask	0.9647	1.3114	1.2048
SL mask	0.9333	1.2931	1.1969
Method	R^2		
	50m	100m	150m
NO mask	0.5276	0.3020	0.5865
DP mask	0.5217	0.3277	0.5742
SL mask	0.5457	0.3577	0.5885

TABLE III
STUDY ON DIFFERENT MARGINS

Margin	RMSE (°C)		
	50m	100m	150m
zero	0.9976	1.3668	1.2345
0.001	0.9401	1.3077	1.2152
0.010	0.9731	1.3631	1.2418
0.100	0.9333	1.2931	1.1969
max	0.9403	1.3297	1.2063
Margin	R^2		
	50m	100m	150m
zero	0.4768	0.2795	0.5596
0.001	0.5397	0.3489	0.5597
0.010	0.5236	0.2813	0.5596
0.100	0.5457	0.3577	0.5885
max	0.5300	0.3166	0.5831

TABLE IV
STUDY ON THE PHYSICS-BASED LOSS IN STAGE 2

Method	RMSE(°C)		
	50m	100m	150m
Model with $L_{0\to50}(G)$ in stage 2	0.9465	1.3386	1.2031
Model without $L_{0\to50}(G)$ in stage 2	0.9333	1.2931	1.1969
Method	R^2		
	50m	100m	150m
Model with $L_{0\to50}(G)$ in stage 2	0.5218	0.3210	0.5826
Model without $L_{0\to50}(G)$ in stage 2	0.5457	0.3577	0.5885

TABLE V
STUDY ON THE NETWORK ARCHITECTURE

Method	RMSE (°C)		
	50m	100m	150m
One attention module	0.9562	1.3286	1.2003
One discriminator	1.0075	1.2965	1.2010
Our method	0.9333	1.2931	1.1969
Method	R^2		
	50m	100m	150m
One attention module	0.5429	0.2799	0.5741
One discriminator	0.4765	0.3511	0.5736
Our method	0.5457	0.3577	0.5885

the RMSE and R^2 results on the usefulness of TANs. By using TANs, the RSME values improve by 0.0097, 0.0198, and 0.0036, respectively, for predicting the sea subsurface temperatures 50-, 100-, and 150-m undersea. Using TANs also improves the R^2 values. The results of Table I, therefore, demonstrate that multitask learning is effective to improve the prediction performance.

2) Effectiveness of the Mask: When we employ the physics-based loss to guide the network training, the temperature between the upper and lower layers is compared by using a mask, and the margin is set to 0.1 here. Here we compare several schemes: no use of mask (NO mask), deeper-layer sea temperature as a mask (DP mask), and shallower-layer sea temperature as a mask (SL mask). Table II summarizes the RMSE and R^2 results obtained with these mask schemes. It can be seen that the method with the SL mask achieves the best RMSE and R^2 values. Therefore, we adopt the SL mask in the physics-based loss for our approach [see (3)–(5) and the discussions after (6)].

3) Analysis of the Margin: Likewise, in order to obtain better fitting ability, we add a margin in the physics-based loss. First, we calculated the maximum temperature difference between the samples of two depths. Then, the margin of the physics-based loss is scaled from 0 to Max. Table III shows the prediction results of different margins. The best RMSE and R^2 values are obtained when the margin is set to 0.100. Therefore, in our approach, we set the margin to 0.100 [see (3)–(5)].

4) Analysis of the Physics-Based Loss in Stage 2: In the second phase of the proposed method, we apply remote sensing data and Argo data to fine-tune the model. We estimate

the contribution of $L_{0\to50}(G)$ in the physics-based loss in Table IV. It can be observed that the model without $L_{0\to50}(G)$ performs better. The reason is due to the imprecision of remote sensing AVHRR SST data, which degrades the performance of the model with $L_{0\to50}(G)$. Therefore, in our proposed method, we do not take $L_{0\to50}(G)$ into account in the physics-based loss in the second stage [see (10)].

5) Network Architecture Design: We use three attention modules and three discriminators for the three specific tasks, respectively. Considering the similarity in these tasks, the network architecture designs that exploit one attention module or one discriminator to learn different tasks are also experimented, and the results obtained are compared with our design in Table V. The experimental results show that using more attention modules and discriminators can achieve better performance. Although our model performs better in this study than the model with a single attention module and single discriminator, it has a higher computational complexity than the latter. In the case of predicting the subsurface temperatures at more than three depths, a single attention module with a single discriminator may become a better choice.

D. Experimental Results and Analysis

For the Argo data from January 2004 to April 2007, after removing the invalid data, we obtain 180 daily temperature observation values. We compare the predicted results with these 180 remaining Argo observational data. Fig. 5 compares the predicted temperature at 50-m undersea with the corresponding Argo data. It can be observed that the predicted results of the proposed method fit well the Argo data. Similarly, the Argo data and the corresponding predicted temperatures at 100-m undersea and 150-m undersea are illustrated in Figs. 6 and 7, respectively. These results demonstrate
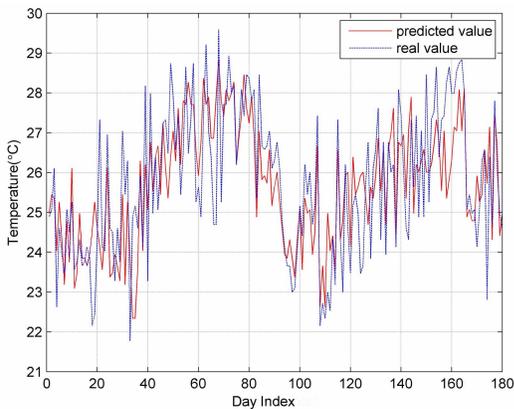
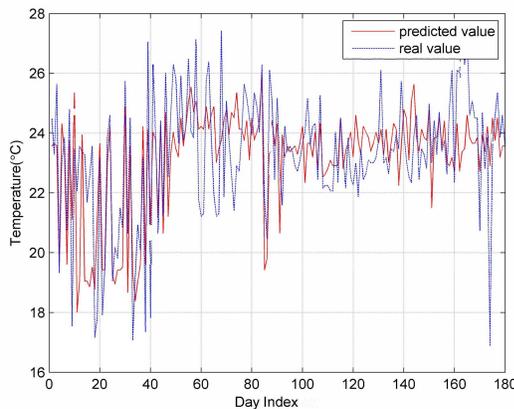Fig. 5.   Predicted temperature at 50-m undersea versus Argo data.

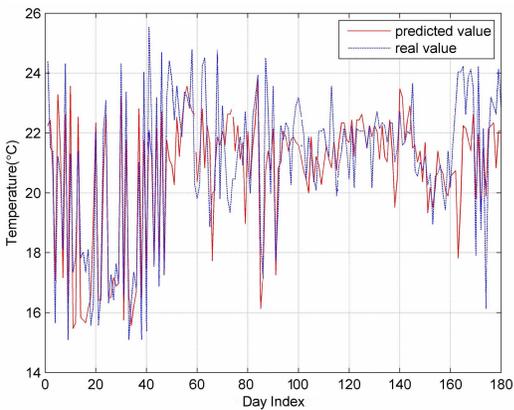Fig. 6.   Predicted temperature at 100-m undersea versus Argo data.

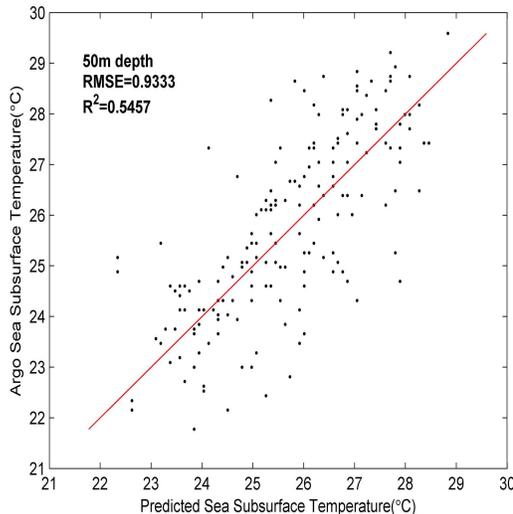Fig. 7.   Predicted temperature at 150-m undersea versus Argo data.

Fig. 8.   Predicted temperature at 50-m undersea and corresponding Argo data scatter plot.

Fig. 9.   Predicted temperature at 100-m undersea and corresponding Argo data scatter plot.

that the proposed method can generate reliable and accurate temperature predictions at different depths of the sea.

A correlation scatter plot between the predicted temperature at 50-m undersea and the Argo data is depicted in Fig. 8. If the data points are more evenly and densely distributed near the diagonal red line, the prediction result is better. Similar scatter plots of the prediction results at 100-m undersea and 150-m undersea are shown in Figs. 9 and 10, respectively. As can be observed, the prediction results at 50-m undersea are

better than the results at 100- and 150-m undersea. Evidently, as depth increases, the prediction accuracy decreases.

Fig. 11 displays the temperatures predicted by the proposed method at different depths (50, 100, and 150 m) together with the corresponding Argo observations on November 9, 2006. The visual results show that the predicted results by the proposed method are very close to the ground-truth Argo data. This demonstrates that the proposed method is reliable and accurate.

The temperature prediction experiment for each model is repeated ten independent runs with different random initializations. We summarize the temperature prediction results, presented as average $\pm$ standard deviation (STD), of different methods in Table VI, where PGpix2pix and PGcycleGAN are the pix2pix method with the numerical model and the CycleGAN method with the numerical model, respectively. Note that applying neural networks, such as pix2pix and CycleGAN, without considering the numerical model is

TABLE VI
SEA SUBSURFACE TEMPERATURE PREDICTION RESULTS (AVERAGE ± STD) OF DIFFERENT METHODS AVERAGED OVER TEN RANDOM RUNS

Model	RMSE(°C)		
	50m	100m	150m
Assimilation method	1.4520	1.8201	1.6774
PGpix2pix	0.9528±0.0114	1.3301±0.024	1.2890±0.0415
PGcycleGAN	2.6155±0.0812	2.5345±0.042	2.7954±0.180
PGNN	0.9482±0.0070	1.3691±0.0259	1.2837±0.039
PGConvLSTM	1.9213±0.223	1.6928±0.021	1.9974±0.127
PGsim	1.1132±0.033	1.4659±0.083	1.3281±0.004
Our method without PLoss	0.9517±0.0082	1.3312±0.0251	1.2648±0.0313
Our method with PLoss	**0.9402±0.0069**	**1.2894±0.0038**	**1.2330±0.0361**

| Model	R^2		
	50m	100m	150m
Assimilation method	-0.4393	-0.2661	0.1938
PGpix2pix	0.5447±0.0119	0.3581±0.051	0.4992±0.043
PGcycleGAN	-2.8694±0.44	-1.9791±0.2798	-1.0427±0.4027
PGNN	0.5381±0.0112	0.2621±0.2089	0.2583±0.1958
PGConvLSTM	-0.8927±0.2110	-0.4445±0.3627	-0.0655±0.1212
PGsim	0.3555±0.0191	0.0179±0.2842	0.4514±0.042
Our method without PLoss	0.5512±0.0175	0.2934±0.045	0.5515±0.0212
Our method with PLoss	**0.5610±0.0153**	**0.3957±0.0392**	**0.5665±0.024**
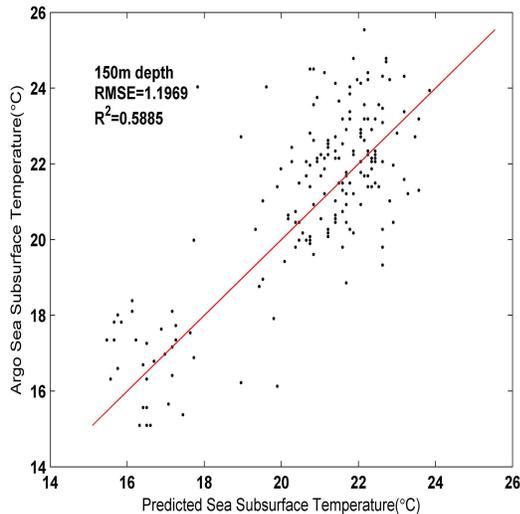
Fig. 10. Predicted temperature at 150-m undersea and corresponding Argo data scatter plot.

incapable of predicting daily sea subsurface temperature effectively due to very limited observational data. Consequently, we have to adapt our idea of physics-guided (PG) enhancement by combining neural networks with a numerical model. To compare with the methods in [35] and [49], we adopt the ConvLSTM model to replace the RNN model for acquiring the sea subsurface temperature prediction in the whole area of China South Sea, which we refer to as PGConvLSTM. Specifically, we train the ConvLSTM model by using the same training mode as ours and removing the mapping from the surface temperature to the subsurface temperature learned by the GAN model as the works [35], [49] did. In our proposed model, we pretrain the GAN on the numerical model data and then fine-tune the GAN model with the observational data. To compare with this two-stage training, we also simply concatenate the numerical simulation data onto the observational data together to train the GAN, which we refer to as PGsim.

Our framework uses the physics loss to automatically encode the knowledge of ocean physics into the modeling process. In addition to our method with physics loss (our method with PLoss), the results of our method without physics loss (our method without PLoss) are also shown in Table VI.

The results of the PGConvLSTM are poor, as this approach does not exploit the mapping from the surface temperature to the subsurface temperature learned by the GAN model [35], [49]. This demonstrates that this mapping is essential in the prediction of the daily subsurface temperature. Our proposed GAN-based framework effectively exploits the merits of both the numerical model and a neural network, and can learn the map from the surface to the subsurface well through the proposed two-stage training. By contrast, simply concatenating the numerical data and the observational data together to train the model (PGsim) is less accurate than our approach. It can be seen from Table VI that our proposed method with the physics loss attains the best performance. In terms of RMSE, it outperforms PGpix2pix by 0.0126 °C, 0.0407 °C, and 0.056 °C for predicting the sea temperatures 50-, 100-, and 150-m undersea, respectively. In terms of R^2 statistic, our method outperforms PGpix2pix by 0.0163, 0.0376, and 0.0673 for predicting the sea subsurface temperatures at these depths, respectively. In addition, our method and PGpix2pix have similar STDs for both performance metrics. Also, observe that our method with physics loss outperforms the one without physics loss. Hence, the experimental results clearly demonstrate that the proposed method is capable of enhancing the daily sea subsurface temperature prediction over the existing state-of-the-art methods.

Currently, only the traditional assimilation method can predict the daily sea subsurface temperature. Our proposed method is the first, which can significantly improve the accuracy of the daily sea subsurface temperature prediction compared with the assimilation method. We believe that exploiting the numerical model data and the two-stage training mode that we propose are essential to perform the daily sea subsurface
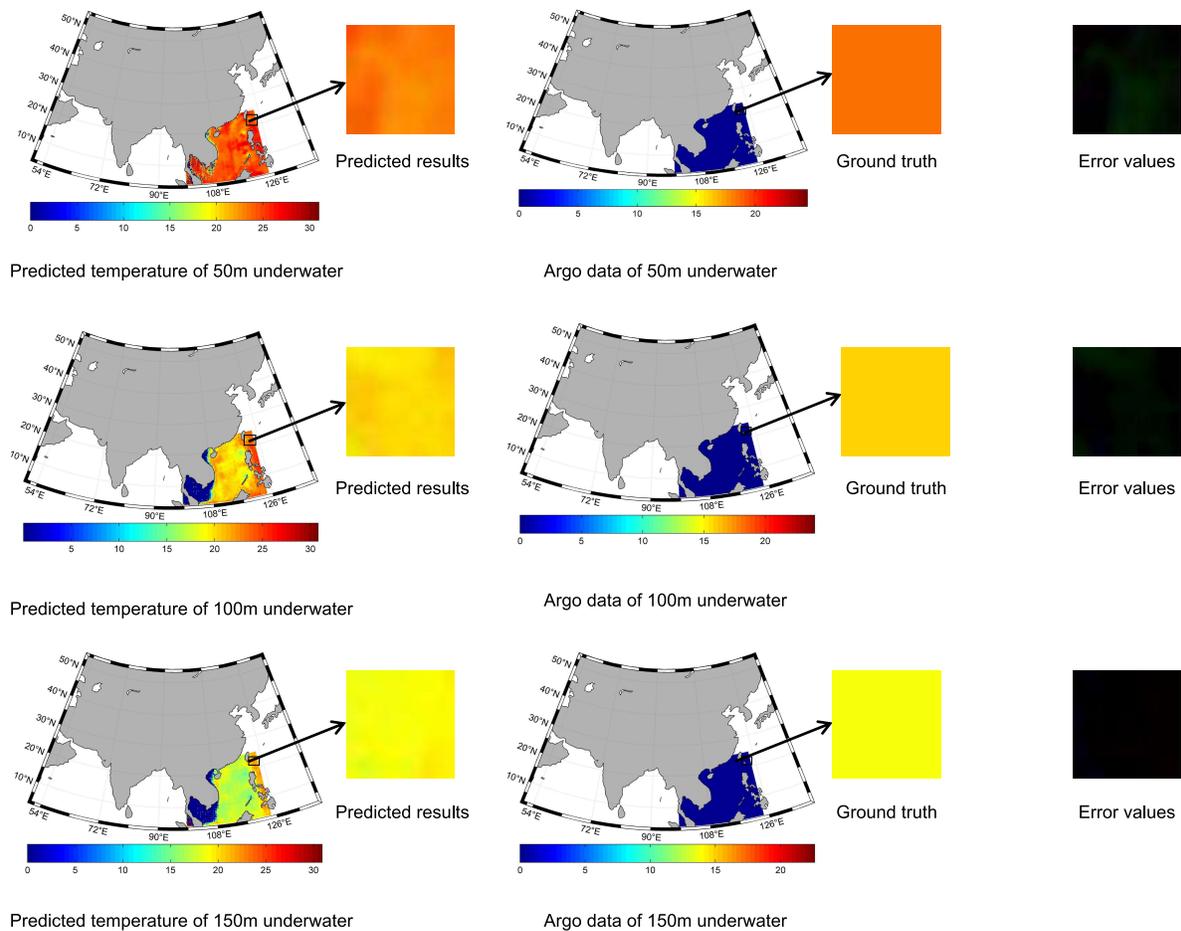
Fig. 11. Display of predicted temperature values and measurements at different depths on November 9, 2006.

temperature prediction task. Multitask learning is integrated into the proposed method to enable the prediction of the temperatures at 50-, 100-, and 150-m underwater simultaneously. A physics-based loss is also added to our model to further improve the accuracy of the daily sea subsurface temperature prediction. The experimental results have verified the effectiveness of our proposed method. Furthermore, our proposed framework also benefits other existing neural network-based methods. Although only applying the pix2pix framework or other neural network is incapable of predicting the daily sea subsurface temperature effectively due to the very limited observational data, by adopting our idea of combining neural network and the numerical model, PGpix2pix becomes capable of significantly outperforming the assimilation method. The experimental results confirm that our proposed method outperforms the existing state-of-the-art methods. Compared with PGpix2pix, although the performance gain is small, our method can predict the temperatures for all the target depths simultaneously, while PGpix2pix needs multiple models to predict the temperatures of different depths.

In our experiments, we have to discard a lot of data since not every daily Argo data are valid. It is clear that, by using more usable datasets to provide sufficient training and testing data, the accuracy of prediction can further be improved. In addition,

our experimental results will also be enhanced by looking for better quality remote sensed images.

V. CONCLUSION AND FUTURE WORK

In this article, we have proposed a novel GAN-based framework for challenging daily sea subsurface temperature prediction. In our method, a physics-based numerical model is employed in a GAN to acquire the simplified physical laws at different ocean depths, and observational data are used for fine-tuning the model parameters to obtain better prediction results. Our method has effectively exploited the complementary merit of the physics-based numerical model and observational data-based neural network. Moreover, a physics-based loss based on a mask has been employed, which leads to improved prediction performance. The experimental results have demonstrated that the proposed method can achieve better performance in daily sea subsurface temperature prediction compared with the state-of-the-art baselines.

In the future, we plan to extend our work to temporal dimensions with better quality and large scales traits, which will provide more information to further improve the prediction accuracy. In addition, we also plan to investigate the use of several self-attention networks to enhance the overall performance of our model.

REFERENCES

[1] C. Chen, K. Yang, Y. Ma, and Y. Wang, "Reconstructing the subsurface temperature field by using sea surface data through self-organizing map method," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 12, pp. 1812–1816, Dec. 2018.

[2] M. J. McPhaden *et al.*, "The tropical ocean-global atmosphere observing system: A decade of progress," *J. Geophys. Res.: Oceans*, vol. 103, no. C7, pp. 14169–14240, 1998.

[3] P. Tandeo *et al.*, "A multivariate regression approach to adjust AATSR sea surface temperature to *in situ* measurements," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 1, pp. 8–12, Jan. 2009.

[4] X. Meng and J. Cheng, "Estimating land and sea surface temperature from cross-calibrated Chinese Gaofen-5 thermal infrared data using split-window algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 3, pp. 509–513, Mar. 2020.

[5] R. Wu, X. Cao, and W. Chen, "Surface wind speed-SST relationship during the passage of typhoons over the South China sea," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 5, pp. 933–937, Sep. 2012.

[6] K. Hosoda, H. Kawamura, K.-W. Lan, T. Shimada, and F. Sakaida, "Temporal scale of sea surface temperature fronts revealed by microwave observations," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 1, pp. 3–7, Jan. 2012.

[7] L. Umlauf and H. Burchard, "A generic length-scale equation for geophysical turbulence models," *J. Mar. Res.*, vol. 61, no. 2, pp. 235–265, Mar. 2003.

[8] E. P. Chassignet *et al.*, "The HYCOM (hybrid coordinate ocean model) data assimilative system," *J. Mar. Syst.*, vol. 65, nos. 1–4, pp. 60–83, Mar. 2007.

[9] C. Chen, R. Beardsley, and G. Cowles, "An unstructured grid, finite-volume coastal ocean model (FVCOM) system," *Oceanography*, vol. 19, no. 1, pp. 78–89, Mar. 2006.

[10] O. M. Smedstad and J. J. O'Brien, "Variational data assimilation and parameter estimation in an equatorial Pacific Ocean model," *Prog. Oceanogr.*, vol. 26, no. 2, pp. 179–241, 1991.

[11] L. T. Anderson, D. J. Sheinbaum, and K. Haines, "Data assimilation in ocean models," *Rep. Prog. Phys.*, vol. 59, no. 10, pp. 1209–1266, 1996.

[12] X. Hong and S. Chen, "A new RBF neural network with boundary value constraints," *IEEE Trans. Syst., Man, Cybern., B (Cybern.)*, vol. 39, no. 1, pp. 298–303, Feb. 2009.

[13] S. Chen, X. Hong, and C. J. Harris, "Grey-box radial basis function modelling," *Neurocomputing*, vol. 74, no. 10, pp. 1564–1571, May 2011.

[14] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, May 2018.

[15] 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.

[16] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, "Missing data reconstruction in remote sensing image with a unified spatial-temporal-spectral deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4274–4288, Mar. 2018.

[17] X. Li, T. Qin, J. Yang, and T.-Y. Liu, "LightRNN: Memory and computation-efficient recurrent neural networks," in *Proc. NIPS*, Barcelona, Spain, Dec. 2016, pp. 4385–4393.

[18] R. Sarikaya, G. E. Hinton, and A. Deoras, "Application of deep belief networks for natural language understanding," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, no. 4, pp. 778–784, Apr. 2014.

[19] M. Korpusik and J. Glass, "Deep learning for database mapping and asking clarification questions in dialogue systems," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 8, pp. 1321–1334, Aug. 2019.

[20] Q. Liu, R. Hang, H. Song, and Z. Li, "Learning multiscale deep features for high-resolution satellite image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 117–126, Jan. 2018.

[21] Y. Sun, Q. Zeng, B. Geng, X. Lin, B. Sude, and L. Chen, "Deep learning architecture for estimating hourly ground-level PM2.5 using satellite remote sensing," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 9, pp. 1343–1347, Sep. 2019.

[22] C. Henry, S. M. Azimi, and N. Merkle, "Road segmentation in SAR satellite images with deep fully convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 12, pp. 1867–1871, Dec. 2018.

[23] M. Ogut, X. Bosch-Lluis, and S. Reising, "A deep learning approach for microwave and millimeter-wave radiometer calibration," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5344–5355, Aug. 2019.

[24] S.-Z. Chen, C.-C. Guo, and J.-H. Lai, "Deep ranking for person re-identification via joint representation learning," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2353–2367, May 2016.

[25] F. Wang, H. Huang, and J. Liu, "Variational-based mixed noise removal with CNN deep learning regularization," *IEEE Trans. Image Process.*, vol. 29, pp. 1246–1258, 2019.

[26] P. Perera and V. M. Patel, "Learning deep features for one-class classification," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5450–5463, Nov. 2019.

[27] Q. Wang, S. Liu, J. Chanussot, and X. Li, "Scene classification with recurrent attention of VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1155–1167, Feb. 2019.

[28] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3–13, Jan. 2019.

[29] C. Yang, "CNN-based polarimetric decomposition feature selection for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8796–8812, Nov. 2019.

[30] H. Sun, X. Zheng, X. Lu, and S. Wu, "Spectral-spatial attention network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3232–3245, May 2020.

[31] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. NIPS*, Montreal, QC, Canada, Dec. 2014, pp. 2672–2680.

[32] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1125–1134.

[33] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2223–2231.

[34] A. Daw, A. Karpatne, W. Watkins, J. Read, and V. Kumar, "Physics-guided neural networks (PGNN): An application in lake temperature modeling," 2017, *arXiv:1710.11431*.

[35] X. Jia *et al.*, "Physics guided RNNs for modeling dynamical systems: A case study in simulating lake temperature profiles," in *Proc. SIAM Int. Conf. Data Mining*, Calgary, AB, Canada, May 2019, pp. 558–566.

[36] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.

[37] L. Yang, D. Zhang, and G. E. Karniadakis, "Physics-informed generative adversarial networks for stochastic differential equations," *SIAM J. Sci. Comput.*, vol. 42, no. 1, pp. A292–A317, Jan. 2020.

[38] B. Lütjens *et al.*, "Physics-informed GANs for coastal flood visualization," 2020, *arXiv:2010.08103*.

[39] Q. Zheng, L. Zeng, and G. E. Karniadakis, "Physics-informed semantic inpainting: Application to geostatistical modeling," *J. Comput. Phys.*, vol. 419, Oct. 2020, Art. no. 109676.

[40] M. Liu, L. Guan, W. Zhao, and G. Chen, "Evaluation of sea surface temperature from the HY-2 scanning microwave radiometer," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1372–1380, Mar. 2017.

[41] Y. Yang, J. Dong, X. Sun, E. Lima, Q. Mu, and X. Wang, "A CFCC-LSTM model for sea surface temperature prediction," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 207–211, Feb. 2018.

[42] L. Wei, L. Guan, and L.-Q. Qu, "Prediction of sea surface temperature in the South China sea by artificial neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 558–562, Apr. 2020.

[43] 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.

[44] K. Patil and M. C. Deo, "Prediction of daily sea surface temperature using efficient neural networks," *Ocean Dyn.*, vol. 67, nos. 3–7, pp. 357–368, Apr. 2017.

[45] K. Patil and M. C. Deo, "Basin-scale prediction of sea surface temperature with artificial neural networks," *J. Atmos. Ocean Technol.*, vol. 35, no. 7, pp. 1441–1455, Jul. 2018.

[46] M. Han, Y. Feng, X. Zhao, C. Sun, F. Hong, and C. Liu, "A convolutional neural network using surface data to predict subsurface temperatures in the Pacific ocean," *IEEE Access*, vol. 7, pp. 172816–172829, 2019.

[47] W. Lu, H. Su, X. Yang, and X.-H. Yan, "Subsurface temperature estimation from remote sensing data using a clustering-neural network method," *Remote Sens. Environ.*, vol. 229, pp. 213–222, Aug. 2019.

[48] X. Wu and X.-H. Yan, "Estimation of subsurface temperature anomaly in the north Atlantic using a self-organizing map neural network," *J. Atmos. Ocean. Technol.*, vol. 29, no. 11, pp. 1675–1688, 2012.

[49] K. Zhang, X. Geng, and X.-H. Yan, "Prediction of 3-D ocean temperature by multilayer convolutional LSTM," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1303–1307, Aug. 2020.

[50] S. Liu, E. Johns, and A. J. Davison, "End-to-end multi-task learning with attention," in *Proc. CVPR*, Long Beach, CA, USA, Jun. 2019, pp. 1871–1880.

[51] *Argo Data Management*. Accessed: May 10, 2021. [Online]. Available: http://www.argodatamgt.org

[52] *NOAA Physical Sciences Lab*. Accessed: Jun. 1, 2021. [Online]. Available: https://psl.noaa.gov

[53] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, San Diego, CA, USA, May 2015, pp. 1–15.

[54] *HYCOM*. Accessed: May 10, 2021. [Online]. Available: https://www.hycom.org

[55] N. R. Draper and H. Smith, *Applied Regression Analysis*. Hoboken, NJ, USA: Wiley, 1998.

[56] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015, *arXiv:1505.04597*. [Online]. Available: http://arxiv.org/abs/1505.04597

Feng Gao (Member, IEEE) received the B.Sc. degree in software engineering from Chongqing University, Chongqing, China, in 2008, and the Ph.D. degree in computer science and technology from Beihang University, Beijing, China, in 2015.

He is currently an Associate Professor with the School of Computer Science and Technology, Ocean University of China, Qingdao, China. His research interests include remote sensing image analysis, pattern recognition, and machine learning.

Yuxin Meng (Student Member, IEEE) received the B.Eng. degree in computer science and technology from the Anhui University of Science and Technology, Huainan, China, in 2010. She is currently pursuing the Ph.D. degree with the Vision Lab, Ocean University of China, Qingdao, China, supervised by Prof. J. Dong.

Her research interests include image processing and computer vision.

Junyu Dong (Member, IEEE) received the B.Sc. and M.Sc. degrees from the Department of Applied Mathematics, Ocean University of China, Qingdao, China, in 1993 and 1999, respectively, and the Ph.D. degree in image processing from the Department of Computer Science, Heriot-Watt University, Edinburgh, U.K., in 2003.

He is currently a Professor and the Dean of the School of Computer Science and Technology, Ocean University of China. His research interests include visual information analysis and understanding, and underwater image processing.

Eric Rigall (Student Member, IEEE) received the engineering degree from the Graduate School of Engineering, University of Nantes, Nantes, France, in 2018. He is currently pursuing the Ph.D. degree with the Vision Lab, Ocean University of China, Qingdao, China, supervised by Prof. J. Dong.

His research interests include signal and image processing, machine learning, and computer vision.

Sheng Chen (Fellow, IEEE) received the B.Eng. degree in control engineering from the East China Petroleum Institute, Dongying, China, in 1982, the Ph.D. degree in control engineering from the City, University of London, London, U.K., in 1986, and the D.Sc. degree from the University of Southampton, Southampton, U.K., in 2005.

From 1986 to 1999, he held research and academic appointments at The University of Sheffield, Sheffield, U.K., the University of Edinburgh, Edinburgh, U.K., and the University of Portsmouth, Portsmouth, U.K. Since 1999, he has been with the School of Electronics and Computer Science, University of Southampton, where he holds the post of Professor of intelligent systems and signal processing. He has published over 600 research articles. He has more than 16 800 Web of Science citations with H-index 57 and more than 33 400 Google Scholar citations with H-index 78. His research interests include adaptive signal processing, wireless communications, modeling and identification of nonlinear systems, neural networks and machine learning, intelligent control system design, and evolutionary computation methods and optimization.

Dr. Chen is also a fellow of the U.K. Royal Academy of Engineering, the Asia-Pacific Artificial Intelligence Association, and IET. He is one of the original ISI highly cited researchers in engineering (March 2004).

Xueen Chen (Member, IEEE) received the B.Sc. degree from the Department of Applied Mathematics, Ocean University of China, Qingdao, China, in 1994, and the Ph.D. degree in marine science from the University of Hamburg, Hamburg, Germany, in 2004.

He is currently a Professor with the School of Oceanic and Atmospheric Science, Ocean University of China. His research interests include ocean dynmaics, regional climate analysis, and ocean data assimilation.