

## A dual-attention embedded CNN model for estimating mixed layer depths in the Bay of Bengal\*

Wentao JIA<sup>1,5</sup>, Xun GONG<sup>2,3,4</sup>, Shanliang ZHU<sup>1,5,\*\*</sup>, Jifeng QI<sup>6,7,\*\*</sup>, Xianmei ZHOU<sup>1,5</sup>, Hengkai YAO<sup>1,5</sup>, Xiang GONG<sup>1,5</sup>, Wenwu WANG<sup>1,5</sup>

<sup>1</sup>School of Mathematics and Physics, Qingdao University of Science and Technology, Qingdao 266061, China

<sup>2</sup>Institute for Advanced Marine Research, China University of Geosciences, Guangzhou 511462, China

<sup>3</sup>State Key Laboratory of Biogeology and Environmental Geology, Hubei Key Laboratory of Marine Geological Resources, China University of Geosciences, Wuhan 430074, China

<sup>4</sup>Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan 250014, China

<sup>5</sup>Qingdao Innovation Center of Artificial Intelligence Ocean Technology, Qingdao 266061, China

<sup>6</sup>Key Laboratory of Ocean Observation and Forecasting and Key Laboratory of Ocean Circulation and Waves, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China

<sup>7</sup>University of Chinese Academy of Sciences, Beijing 100049, China

Received May 9, 2024; accepted in principle Jul. 2, 2024; accepted for publication Aug. 27, 2024

© Chinese Society for Oceanology and Limnology, Science Press and Springer-Verlag GmbH Germany, part of Springer Nature 2025

**Abstract** Variations in ocean mixed layer depth (MLD) show a significant impact on energy balance in the global climate systems and marine ecosystems. At present, the accuracy of modeling MLD, especially in the region with complex ocean dynamics, remains a challenge, thus calling for an emergency using artificial intelligence approach to improve the assessment of the MLD. A novel convolutional neural network model was developed based on a dual-attention module (DA-CNN) to estimate the MLD in the Bay of Bengal (BoB) by integrating multi-source remote sensing data and Argo gridded data. Compared with the original CNN model, the DA-CNN model exhibits superior performance with notable improvements in the annual average root mean square error (RMSE) and  $R^2$  values by 13.0% and 8.4%, respectively, while more accurately capturing the seasonal variations in MLD. Moreover, the results using the DA-CNN model show minimum RMSE and maximum  $R^2$  values, in comparison to the calculation by the random forest, artificial neural network model, and the hybrid coordinate ocean model. Accordingly, our findings suggest that the newly developed DA-CNN model provides an effective advantage in studying the MLD and the associated ocean processes.

**Keyword:** mixed layer depth (MLD); remote sensing observation; dual-attention module (DA-CNN); Bay of Bengal

### 1 INTRODUCTION

Ocean mixed layer refers to surface depths of the ocean with quasi-homogeneous temperature, salinity, and density. It is critical for marine primary production, the exchange of heat, and momentum in ocean-atmosphere interactions (Lorbacher et al., 2006). In particular, mixed layer depth (MLD) is a vital factor and plays an important role in regulating energy balance in the global climate system and carbon cycles (Kara et al., 2003; Keerthi et al.,

2013). For instance, it has been shown in heat budget analysis that the MLD has important implications for determining the location and seasonal evolution of warm blobs and temperature diagnosis over the New Pacific region (Shi et al., 2022). More studies have also indicated that the

\* Supported by the Ministry of Science and Technology of the People's Republic of China (No. 2019YFE0125000) and the National Natural Science Foundation of China (No. 42376032)

\*\* Corresponding authors: zhushanliang@qust.edu.cn; jfqi@qdio.ac.cn

MLD variability can affect the rate of heat flux exchange between the ocean and atmosphere, the ocean's ability to store and transport heat as well as carbon (Gadgil et al., 1984; Yamamoto et al., 2015; Dall'Olmo et al., 2016; Jang et al., 2017; Yu et al., 2019), and the availability of light and nutrients to support the growth of phytoplankton (Dickey et al., 1993; Polovina et al., 1995; de Fommervault et al., 2017; Diaz et al., 2021; Xue et al., 2022).

According to in-situ observations, the MLD is predominantly reliant on vertical ocean temperature, salinity, and thus density profiles (Paillet et al., 1999; Kara et al., 2000a; Thomson and Fine, 2003; Holte and Talley, 2009; Helber et al., 2012). The integration of multi-source observational data and threshold calculation methods has established a fundamental groundwork for the investigation of the MLD and upper-ocean dynamic processes. However, the spatiotemporal data from in-situ observations are often discontinuous, and their discontinuity and incompleteness can affect the estimation of the MLD. In technique, Holte et al. (2017) suggested the calculation of MLD using superior quality indicators better than threshold methods and significantly mitigates the tendency of these methods to overestimate the MLD in some regions with the deep mixed layer in winter. Li et al. (2017) enhanced the classic Barnes method by utilizing optimal parameters and response functions to reduce the error caused by the uneven spatial distribution of Argo observation data and established a MLD dataset that retains a more comprehensive set of mesoscale features. Although the spatiotemporal resolution of in-situ observation data and the calculation methods of MLD have been improved, classic methods still face challenges such as limited spatial coverage, low spatiotemporal resolution, and low accuracy (Hosoda et al., 2010; Holte et al., 2017; Li et al., 2017).

Over the past few decades, various methods have been widely applied to estimate the MLD, thanks to the rapid accumulation of in-situ, remote sensing, and aerial survey data, as well as the rapid development of ocean information detection techniques. Previous studies suggested that many oceanic subsurface phenomena can be characterized by the relevant surface parameter data (Fiedler, 1988; Vernieres et al., 2014). For example, Rintoul and Trull (2001) explored the seasonal variations in MLD and nutrient concentrations by comprehensively analyzing aerial survey data of many years in the sub-Antarctic region, revealing

the seasonal characteristics of the MLD in this region, which are shallow in summer and deep in winter. Li et al. (2000) estimated the MLD by matching the measured internal wave group velocities with those calculated by the model. Although observational data has achieved continuous and extensive sampling in both time and space, previous methods such as data assimilation and numerical simulation of subsurface ocean variables have generally been complex and computationally expensive, and their estimation accuracy cannot be guaranteed (Courtois et al., 2017; Dwivedi et al., 2018; Wei et al., 2023).

Data-driven artificial intelligence (AI) models have received quite extensive attention in the field of oceanography in recent years, demonstrating superior performance in estimating internal ocean variables from different observation data (Meng et al., 2022; Yue et al., 2024). For example, Su et al. (2021b) proposed a bi-directional long short-term memory neural networks (Bi-LSTM) method to predict the global ocean sea surface temperature anomaly (STA) and sea surface salinity anomaly (SSA) in combination with surface remote sensing observations and subsurface Argo gridded data. Pauthenet et al. (2022) proposed an estimation method for the MLD in the Gulf of Mexico based on a multilayer perceptron, demonstrating the potential of machine learning methods in MLD estimation. Jeong et al. (2019) utilized high spatiotemporal resolution satellite sea surface data to reconstruct a 3D thermohaline field in the ocean subsurface layer, and further estimated and analyzed the decadal variations in the global MLD. Foster et al. (2021) tested a variety of traditional and probabilistic machine learning techniques for the southern Indian and eastern equatorial Pacific regions and found that machine learning models combined with sea surface data can effectively improve the estimation accuracy of MLD compared with the optimal interpolation of Argo observation data. Su et al. (2024) proposed a Residual Convolutional Gate Recurrent Unit neural network to estimate the global MLD and the model can efficiently extract spatio-temporal features from ocean observations. These research methods based on AI techniques provide a new way of developing MLD estimation.

Although AI models have exhibited capability in estimating the MLD, the limitations of observational data and models have led to some remaining issues such as few input parameters, relatively simple models (Foster et al., 2021; Gu et al., 2022;

Pauthenet et al., 2022) and high estimation errors (Su et al., 2024). For example, Gu et al. (2022) used a pre-clustering-based artificial neural network (ANN) model and estimated the MLD in the Indian Ocean but with an averaged estimation error of about 3.79 m, and the estimation errors increased to 7.02 m and 9.15 m in the Arabian Sea and the Bay of Bengal (BoB), respectively. This could be attributed to the insufficient ability of the model to extract and learn spatiotemporal characteristics with complex nonlinear relationships. Therefore, both the model itself and its accuracy have considerable potential improvement.

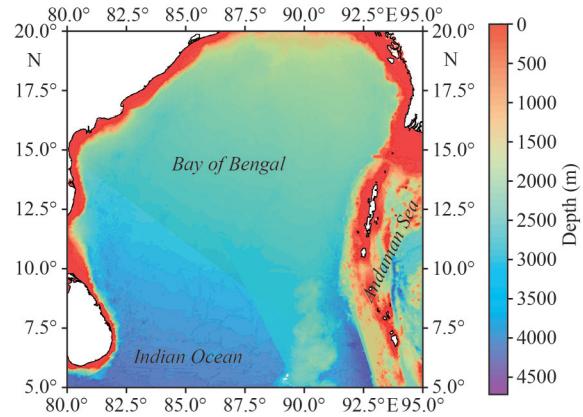
Furthermore, deep learning models based on attention mechanisms have attracted the attention of oceanographers (Li et al., 2022). Qi et al. (2023) developed a CNN model based on the attention mechanism to reconstruct the 3D thermohaline field in the Indian Ocean and achieved excellent results. Ren et al. (2022) developed a U-net model based on an attention module to classify the sea ice and open water from SAR images, and the results showed that the proposed method significantly improved the classification accuracy compared with the original U-net model.

Motivated by the aforementioned discussions, the primary objective of this study is to investigate a new way to estimate the MLD in some typical regions with complex dynamic processes by developing a novel AI model based on multi-source remote sensing data. Here, we proposed a CNN model based on a dual-attention module (DA-CNN) to estimate the MLD using multi-source satellite observation data in the BoB, as a case study. In addition, we evaluated the proposed model by comparing its performance with the data-driven CNN model, random forest (RF) model, and ANN model, as well as the physics-driven hybrid coordinate ocean model (HYCOM).

## 2 STUDY AREA AND DATA

### 2.1 Study area

The BoB ( $5^{\circ}\text{N}$ – $20^{\circ}\text{N}$ ,  $80^{\circ}\text{E}$ – $95^{\circ}\text{E}$ ), adjoining Asia and occupying the eastern part of the tropical Indian Ocean, is an important part of the Indo-Pacific warm pool (Fig.1). The salinity in the BoB shows significant spatial and temporal variations due to the freshwater runoff from the hinterland river and substantial precipitation associated with summer monsoons (Howden and Murtugudde, 2001; Vinayachandran et al., 2002; Akhil et al.,



**Fig.1** Bathymetry in the Bay of Bengal ( $5^{\circ}\text{N}$ – $20^{\circ}\text{N}$ ,  $80^{\circ}\text{E}$ – $95^{\circ}\text{E}$ )

2020). As a result, these abnormal fluctuations lead to large variability in the MLD of this region, which has a critical impact on some regional oceanic phenomena such as tropical cyclones, El Niño events, Indian Ocean dipoles, and monsoon variations (Masson et al., 2005; Yang et al., 2007; Balaguru et al., 2012; Kumari et al., 2018; Goswami et al., 2022). Therefore, the accurate estimation of MLD helps comprehend the variability of the ocean-atmosphere heat flux and analyze the dynamic mechanism of these oceanic phenomena in the region.

### 2.2 Data source and preprocessing

The datasets in this study involve a series of sea surface remote sensing data and Argo gridded data in the BoB from January 2010 to December 2019, as summarized in Table 1. The sea surface temperature (SST) data of a horizontal resolution of  $0.25^{\circ}\times 0.25^{\circ}$  is generated using a daily optimum interpolation method based on radiometer satellite from the National Oceanic and Atmospheric Administration (NOAA) and ship observations (Banzon et al., 2016). In addition, the sea surface salinity (SSS) data is sourced from the Soil Moisture and Ocean Salinity (SMOS) Level-3 salinity product at a spatial resolution of  $0.25^{\circ}\times 0.25^{\circ}$  (Boutin et al.,

**Table 1** Summary of the data used in this study

Variable	Data source	Time range	Resolution
SSS	SMOS		Monthly/ $0.25^{\circ}\times 0.25^{\circ}$
SST	NOAA		Monthly/ $0.25^{\circ}\times 0.25^{\circ}$
SSH	AVISO	2010–2019	Monthly/ $0.25^{\circ}\times 0.25^{\circ}$
SSW	CCMP		Monthly/ $0.25^{\circ}\times 0.25^{\circ}$
MLD	Argo		Monthly/ $1^{\circ}\times 1^{\circ}$

2018). The sea surface height (SSH) data is obtained from the Archiving, Validation, and Interpretation of Satellite Oceanographic Data (AVISO) project also at a spatial resolution of  $0.25^\circ \times 0.25^\circ$  (Hauser et al., 2021). The eastward component sea surface wind (USSW) and northward component sea surface wind (VSSW) values are from the Cross-Calibrated Multi-Platform (CCMP) product of a spatial resolution of  $0.25^\circ \times 0.25^\circ$  (Atlas et al., 2011). Moreover, Argo gridded data are obtained from the Asia Pacific Data Research Center (APDRC) with a spatial resolution of  $1^\circ \times 1^\circ$  (Wong et al., 2020). Additionally, the MLD used for comparison is derived from the HYCOM reanalysis data to validate the estimation performance of the proposed model.

In our calculation, the SST, SSS, SSH, VSSW, and USSW are independent input variables for the proposed model. Additionally, geographic information such as longitude (LON) and latitude (LAT) can affect the performance of the estimation model (Gueye et al., 2014; Su et al., 2021a). Consequently, LON and LAT parameters with the same resolutions as other input parameters are selected as input variables for the model. The Argo-derived MLD gridded data are used as the training and validation labels for the proposed model in this study. At the data preprocessing stage, all the data are monthly and interpolated onto a grid with a resolution of  $0.5^\circ \times 0.5^\circ$ , in line with the temporal and spatial coverage of the BoB to ensure consistency and accuracy in the modeling calculation and evaluation. Any data point with missing parameters within the BoB is excluded. After that, a total of 120 monthly valid datasets from January 2010 to December 2019 are obtained, with 653 valid data points for each variable per month. Finally, all data are normalized by utilizing the mean and standard deviation of the data to expedite model convergence.

### 3 METHOD

The CNN model is widely employed across various domains in deep learning (Lecun et al., 1998; Qi et al., 2023). It functions on local connections and weight sharing, enabling efficient extraction and learning of features from high-dimensional geographical spatial data. On the other hand, the CNN model still confronts challenges, for instance, information overload and complex network structures, leading to slow parameter updates and suboptimal expressive capabilities (Liu

et al., 2018). In recent years, various attention mechanisms have been successfully applied to reduce computational complexity and enhance the network's ability to process information (Li et al., 2022; Ren et al., 2022; Qi et al., 2023). In this study, an improved DA-CNN model for complex multidimensional ocean data is proposed by integrating the dual attention (DA) module into the CNN architecture and is used to estimate the MLD in BoB, as an application case. In Section 3.1, the working principles and advantages of the DA module are presented, while Section 3.2 provides a detailed description of the specific architecture and estimation modeling process of the DA-CNN model.

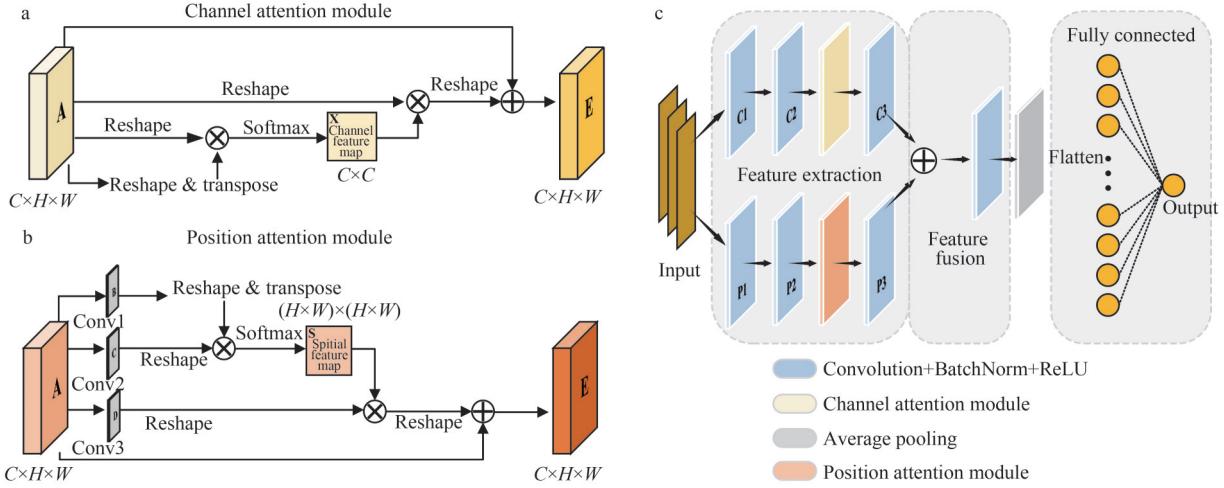
#### 3.1 DA module

The DA module is an algorithm that utilizes self-attention mechanisms to adaptively integrate local semantic features, thereby enhancing the expressive capabilities of AI models in computer vision tasks (Fu et al., 2019). The DA module consists of two submodules, including the channel attention module (CAM) and the position attention module (PAM). The CAM captures the channel dependencies between any two channels and aids the model in more efficiently combining and selecting features, ultimately improving its expressiveness and generalization abilities. The PAM focuses on the spatial information of input data to capture critical information at various spatial positions. It helps the model better understand the significance of different positions when processing data like images, thereby enhancing the model's localization and discrimination capabilities.

Figure 2 shows the structure of the DA module and DA-CNN model, where the structure of CAM is shown in Fig.2a.  $\mathbf{A} \in R^{C \times H \times W}$  is the local feature map output by the previous processing.  $\mathbf{A}$  is reshaped into  $R^{C \times N}$  and performed matrix multiplication with its transpose  $\mathbf{A}'$ . The softmax layer is applied to obtain the channel attention map  $\mathbf{X} \in R^{C \times C}$ , where  $x_{ji}$  measures the  $i^{\text{th}}$  channel's impact on the  $j^{\text{th}}$  channel.  $\mathbf{A}'$  is multiplied with  $\mathbf{X}$  and reshaped into  $R^{C \times H \times W}$ .  $R$  is multiplied by a scale parameter  $\alpha$  and added to the input  $\mathbf{A}$  to obtain the output  $\mathbf{E} \in R^{C \times H \times W}$  element by element.

$$x_{ji} = \frac{e^{A_i A_j}}{\sum_{i=1}^C e^{A_i A_j}}, \quad (1)$$

$$E_j = \alpha \sum_{i=1}^C (x_{ji} A_i) + A_j, \quad (2)$$



**Fig.2 Structure of CAM (a), PAM (b), and DA-CNN model (c)**

where  $\alpha$  is initialized to 0 and gradually learns to assign more weights by the backpropagation. It can be inferred from Eq.2 that the resulting feature  $\mathbf{E}$  at each channel is the weighted sum of the global features  $R$  and local features  $\mathbf{A}$ , it models the long-term semantic dependencies between feature maps.

The structure of the PAM is shown in Fig.2b.  $\mathbf{A} \in R^{C \times H \times W}$  is the local feature map extracted by the previous processing.  $H$ ,  $W$ , and  $C$  are the rows, columns, and channels, respectively.  $\mathbf{A}$  is transformed and reshaped to new feature maps  $\mathbf{B}$ ,  $\mathbf{C}$ , and  $\mathbf{D}$  by passing through three convolutional layers, respectively, where  $\{\mathbf{B}, \mathbf{C}, \mathbf{D}\} \in R^{C \times N}$  ( $N = H \times W$ ).  $\mathbf{A}$  matrix multiplication and softmax activation are performed on  $\mathbf{B}$  and  $\mathbf{C}$  to obtain the spatial position map  $\mathbf{S} \in R^{N \times N}$ . The more similar feature representations of the two positions contribute to a greater correlation between them, where  $s_{ji}$  measures the  $i^{\text{th}}$  position's impact on  $j^{\text{th}}$  the position.  $\mathbf{S}$  is multiplied with  $\mathbf{D}$  and reshaped to  $R^{C \times H \times W}$ . For each channel of  $R$ , the element of a position is the weighted sum of elements across all positions in  $\mathbf{D}$  based on the weights in  $\mathbf{S}$ .  $R$  is multiplied by a scale parameter  $\beta$  and added to the input  $\mathbf{A}$  in element-wise to obtain the output  $\mathbf{E} \in R^{C \times H \times W}$ .

$$s_{ji} = \frac{e^{B_i C_j}}{\sum_{i=1}^N e^{B_i C_j}}, \quad (3)$$

$$E_j = \beta \sum_{i=1}^N (s_{ji} D_i) + A_j, \quad (4)$$

where  $\beta$  is initialized to 0 and gradually learns to assign more weights by the backpropagation. It can be inferred from Eq.4 that the resulting feature  $\mathbf{E}$  integrates the local features  $\mathbf{A}$  and the global

features  $R$ . Thus, it has a global context view and selectively aggregates context based on a position attention map.

### 3.2 DA-CNN model

The DA-CNN model is an improved model of the CNN model, specifically designed to enhance the accuracy of estimating MLD from multi-source remote sensing data. The PAM can help the model capture spatial relationships among features in the input data, enabling it to focus on important spatial locations. The CAM can emphasize the importance of different feature channels, highlighting significant features and suppressing irrelevant ones. The model is capable of focusing its attention on key variables and informative features by incorporating both the PAM and CAM, capturing the local and global dependencies of satellite data in the spatial and channel dimensions, thereby improving the generalization ability of the model.

The DA-CNN model is composed of convolutional layers, batch normalization (BN) layers, Rectified Linear Unit (ReLU) activation layers, DA modules, global average pooling (GAP) layers, and fully connected (FC) layers. BN layers and ReLU activation layers are applied after each convolutional layer to prevent overfitting as well as gradient explosion or vanishing. Figure 2c shows the overall architecture of the DA-CNN model. The preprocessed satellite observation data are fed into two parallel attention branches, where local feature maps are generated through successive convolutional operations. Each map undergoes individual attention mechanism processing and passes to deeper convolutional layers. Subsequently,

the features from both branches are merged and further processed through a convolutional layer to facilitate feature fusion. This convolutional layer stacks features along the channel dimension, effectively combining complementary information from two branches and enhancing feature discrimination. It reduces the dimensionality while preserving the information, decreasing the model complexity and computational resources. Additionally, a GAP layer is introduced to reduce computational complexity, followed by FC layers to output the MLD at a specific coordinate point. This streamlined process helps the network better understand and capture complex nonlinear relationships in the input data, while also speeding up the model training.

In this study, a DA-CNN model for estimating the MLD is developed using multi-source satellite observation data in the BoB. As shown in Fig.3, the study flow is divided into three stages. The collection and processing of raw data is the first stage. The raw datasets of the five sea-surface parameters (SST, SSS, SSH, USSW, VSSW), as well as LON and LAT for each point, are collected and preprocessed from multi-source databases. The training and testing datasets covering 120 months and 653 data points per month are established and the Argo-derived MLD data are used as the training and testing labels. In the second stage, the monthly average data from January 2010 through December 2018 are taken as the training datasets to optimize the model parameters. Specifically, 108-month data is divided into 12 separate training datasets according to the same month. Each training dataset covering 9 months and 653 data points for each variable per month is fed into a DA-CNN estimation model for training respectively, leading to 12 estimation models. The experiments determine

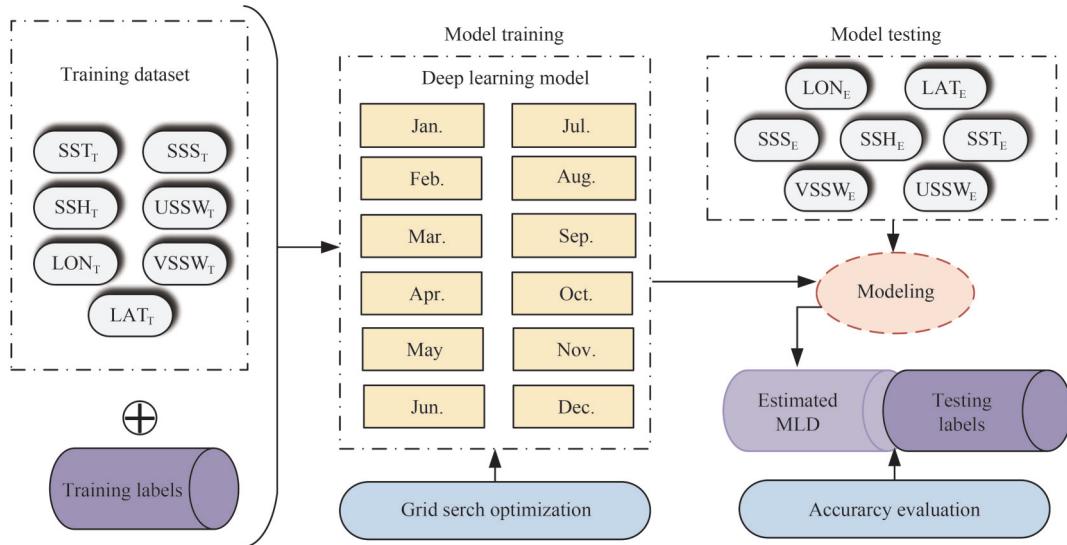
several key hyperparameters that have a significant impact on the model performance, including the learning rate and the number of neurons per layer. The grid search method is employed to systematically explore combinations of these hyperparameters over a specified range. The search range for the learning rate is set between 0.01 and 0.1, and the number of neurons ranges from 16 to 256. The best combination of parameters for the 12 DA-CNN models is presented in Table 2. In the third stage, the testing data for each month in 2019 is input into the model for the corresponding month to obtain the results for this month. Meanwhile, the root mean square error (RMSE) and determination coefficient ( $R^2$ ) are chosen to evaluate the performance of the 12 DA-CNN models. The DA module code is based on open-source code, and the DA-CNN models are implemented and tested on an RTX 3090 (24GB) graphics card using a Python program based on PyTorch.

## 4 RESULT

#### 4.1 Validation of satellite-derived SSS and SST

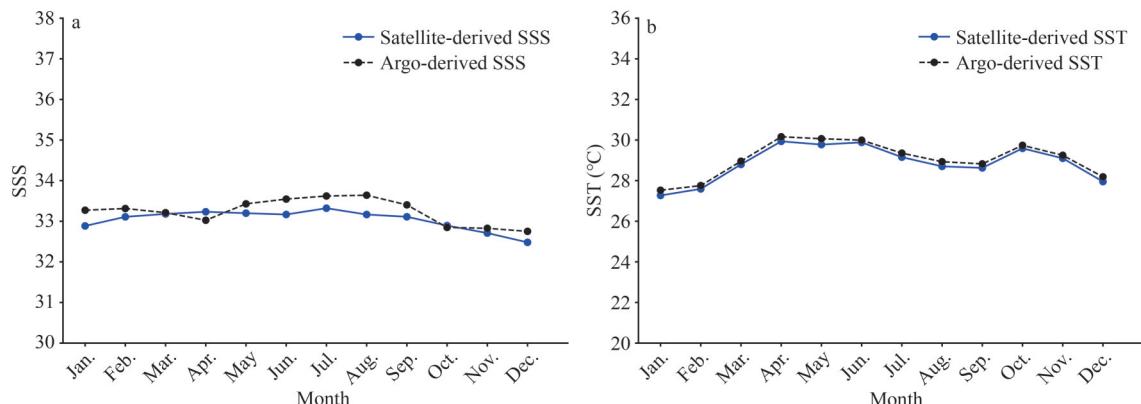
AI models are data-driven models, and the quality of the input dataset directly affects their performance (Jiang et al., 2021). This study uses high-resolution satellite observation data in the training process of the DA-CNN model and derives label data from Argo gridded data. Since the MLD in the Argo dataset is calculated based on vertical temperature and salinity data, it is necessary to validate the reliability of the satellite-derived SSS and SST data by comparing them with the Argo gridded data when they are used as input variables of the model. As shown in Fig.4a, the monthly average of satellite-derived SSS and Argo-derived SSS show good agreement and similar seasonal

**Table 2** Optimal parameter values for DA-CNN models in different months



**Fig.3** The workflow of the methodologies used in this study

variation features. For instance, the SSS values from remote sensing and Argo observations are 33.2 and 32.9 in March and October, respectively, and are close to equal. It can be also observed that the relatively large difference between the two datasets occurs in summer and is approximately 0.4, which may be due to their different measurement errors and sources (Zhao et al., 2023). Although there exists a certain level of discrepancy, they are insignificant. Similarly, the satellite-derived SST data averaged over the BoB is also in agreement with the Argo SST data on a seasonal scale (Fig.4b). For example, both of them show that the maximum SST value ( $>29.8^{\circ}\text{C}$ ) occurs in April, while the SST minimal values ( $<27.5^{\circ}\text{C}$ ) in January, and the differences between the two datasets are less than  $0.2^{\circ}\text{C}$ . These comparative results demonstrate the reliability of the satellite-derived SSS and SST data used in this study.



**Fig.4** Comparison of the Argo (dashed black line) and satellite (solid blue line) for the monthly average SSS (a) and the monthly average SST (b) in the BoB from January 2010 to December 2019

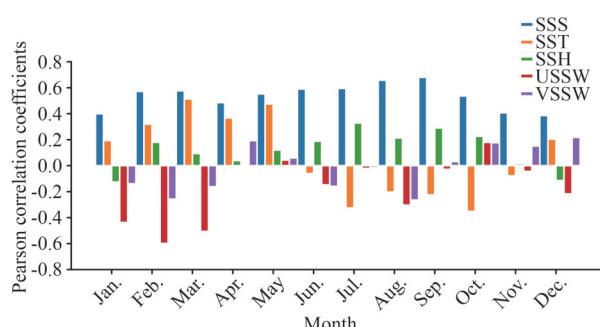
#### 4.2 Identification of input variables

To determine the optimal combination of input variables for the DA-CNN model, the Pearson correlation coefficient is applied to quantitatively analyze the correlation between the MLD and each input variable. Figure 5 shows the monthly average Pearson correlation coefficients of the MLD and SSS/SST/SSH/USSW/VSSW (individually) from January 2010 to December 2019. Generally, an increase in SST decreases the density, making the mixed layer shallower, while an increase in SSS raises it, making the layer deeper (Kara et al., 2000b, 2003; Mignot et al., 2007). As seen in Fig.5, the SSS exhibits a positive correlation with the MLD throughout the year. Their monthly average correlation coefficient values are consistently around 0.5 or higher, up to 0.7. Moreover, these

values exceed the corresponding values between the MLD and the other four variables in most months. The findings indicate that the SSS is the most significant factor influencing the MLD in the BoB, consistent with previous studies (Pailler et al., 1999). The SST is positively correlated with the MLD in winter and spring, suggesting that simultaneous increases in the SST and SSS promote the MLD deepening. On the contrary, the negative correlation between the SST and MLD in summer and autumn indicates that the SSS dominates the MLD variations during this period, while the SST plays a minor role in its variations. The correlation coefficient between the MLD and SSH is close to 0.4 in summer, while their correlation was relatively smaller in winter at approximately 0.2. This indicates that SSH shows a greater effect on the MLD of the BoB in summer than in winter. In parallel, the MLD is better correlated to the USSW compared to the VSSW (Fig.5). The negative correlation coefficient between the MLD and USSW in spring reaches 0.6. This demonstrates a greater linkage between the USSW and the MLD. The correlation analysis between the MLD and sea surface parameters elucidates the effect of each sea surface parameter on the MLD and provides the reasons for the involvement of the selected input variables for the model.

#### 4.3 Input variable comparison experiment

Studies have shown that sea surface parameters, e.g., SST, SSH, SSS, USSW, VSSW, and geographic information, help improve the accuracy of MLD estimation (Su et al., 2021a; Wang et al., 2021; Gu et al., 2022; Qi et al., 2023). However, there are few quantitative analyses to assess the importance of each input variable in individually affecting the



**Fig.5** Pearson correlation coefficients between the sea surface parameters (SSS, SST, SSH, USSW, and VSSW) and the Argo-derived MLD from January 2010 to December 2019

MLD estimation. To investigate the quantitative impact of each input variable on the estimation performance of the proposed DA-CNN model, we designed four cases of experiments with different combinations of input variables, as shown in Table 3. Since the MLD is derived from the density threshold method in this study, the combination of the SSS, SST, and geographical information is set as the baseline experiment, named Case 1. For comparison, Case 2 and Case 3 add the SSH, and then USSW and VSSW, to Case 1, respectively. Case 4 refers to an experiment where all parameters are used as the input variables of the model. As listed in Table 3, the annual average RMSE and  $R^2$  of the DA-CNN model in Case 4 are 2.71 m and 0.85, respectively, and the model achieves the best estimation results among the four experiments. Compared to Case 1, the annual average RMSE and  $R^2$  are improved by 10.1% and 8.6% for Case 2 with the SSH, 6.8% and 8.8% for Case 3 with the USSW and VSSW, and 16.4% and 13.1% for Case 4, respectively. The comparison results show that the model in Case 4 has the best estimation performance, while the SSH in Case 2 displays superior improvements on the model results in comparison to the USSW and VSSW in Case 3.

The estimation performance of the model in the different cases is further analyzed on a seasonal scale. Figure 6 illustrates the variations in the monthly average RMSE and  $R^2$  for the DA-CNN model for Case 1 to Case 4 in 2019. According to the comprehensive comparison of all cases, it can be noted that the model performance in the four cases shows similar seasonality. Input variables play different roles in the estimation effects of the model in different months, generally consistent with the results of the previous correlation analysis between the variables. For example, the correlation between the MLD and SSW is greater than that of the SSH from January to March (Fig.5). Correspondingly, the DA-CNN model in Case 3 has smaller RMSEs and larger  $R^2$  values compared to Case 2. Case 4 shows

**Table 3 Comparison experiment quantitative results**

Experiment	Annual average RMSE (m)	Annual average $R^2$	RMSE improvement (%)	$R^2$ improvement (%)
Case 1	3.25	0.75	—	—
Case 2	2.92	0.81	10.1	8.6
Case 3	3.03	0.81	6.8	8.8
Case 4	2.71	0.85	16.4	13.1

— means no data.

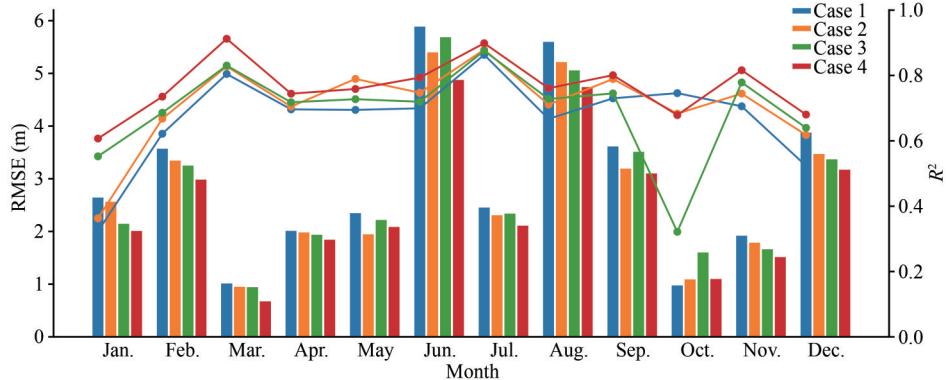


Fig.6 Variations of monthly average RMSE (m) and  $R^2$  for the variable comparison experiment in 2019

The bars and lines represent RMSE and  $R^2$ , respectively.

the best estimation of the annual cycle in the MLD except for May and October. About the estimation for May, Case 4 presents RMSE and  $R^2$  of 2.10 m and 0.76, slightly worse than the values of 1.96 m, and 0.79 for the experiment Case 2. Regarding the estimation for October, Case 2 and Case 4 are insignificantly different from Case 1. In comparison, the performance of Case 3 shows the worst estimation of the MLD. This can be attributed to the fact that the input variables of the model in both months are characteristically entangled, thus limiting the model's capability to capture the nonlinear relationship among the input variables (Karras et al., 2021). In addition, the strongest positive Indian Ocean Dipole (IOD) event in the Indian Ocean in 2019 may also make these variables anomalous (Du et al., 2020), which may have an impact on the performance of the estimation model.

#### 4.4 Evaluation of the DA-CNN model in Case 4

In this section, the performance of the DA-CNN model in Case 4 is comprehensively evaluated from multiple perspectives. In Section 4.4.1, three experiments on model ablation are conducted to demonstrate the superiority of the DA-CNN model. Two experiments are performed in Section 4.4.2 to verify the generalization ability of the DA-CNN model. Moreover, the monthly average MLD in the BoB in 2019 is estimated by the DA-CNN model, and its spatial distribution characteristics are analyzed in Section 4.4.3. The effect of the DA module in the CNN model on the estimation performance is quantitatively analyzed from the aspect of the season in Section 4.4.4.

##### 4.4.1 Experiment on model ablation

Three experiments on model ablation were

designed to better understand the estimation performance of the DA-CNN model. The ablation models are the original CNN model without attentional mechanism, the CNN model with only the CAM (CAM-CNN), and the CNN model with only the PAM (PAM-CNN). These models were applied to estimate the MLD in the representative months (February, May, August, and November) of the four seasons in 2019. As shown in Table 4, any attention mechanism can improve the estimation accuracy of the model compared with the original CNN model. For example, The RMSE values of the CAM-CNN model, PAM-CNN model, and DA-CNN model are reduced by 0.43, 0.36, and 0.87 m compared to that of the CNN model in August, respectively. Moreover, the DA-CNN model achieves the best estimation performance in each season, especially in the summer when the MLD is the deepest. These experimental results suggest that the introduction of the DA module significantly enhances the fitting ability of the CNN model and improves the estimation accuracy of the model.

##### 4.4.2 Experiment on model generalization ability

The DA-CNN models for February, May, August, and November, as representative for the four seasons in 2019, are used to estimate the MLD for these months to verify whether the model for specified months can be generalized to estimate the

Table 4 Experiment results on model ablation

RMSE (m)	Feb.	May	Aug.	Nov.
CNN	3.07	2.48	5.64	1.76
CAM-CNN	3.19	2.32	5.21	1.69
PAM-CNN	3.07	2.43	5.28	1.66
DA-CNN	3.03	2.1	4.77	1.53

MLDs in other months. The RMSE values for each month are listed in Table 5 with the best RMSE values highlighted in red. As shown in Table 5, the models trained for a particular month each show the most efficient estimation of the MLD for the corresponding month. This suggests that for oceans with significant seasonality, it is necessary to build models for different months, i.e., generic parameter sets for different seasons will reduce the accuracy of MLD estimates.

The trained DA-CNN models from January to April are applied to the MLD estimation for the corresponding months in 2019 and 2020 to further validate the estimation performance of these models in different years, respectively. The monthly average RMSE values of the models for January to April in 2019 and 2020 are shown in Table 6. It can be seen that the RMSE for each month in 2020 increases in different degrees compared to that in 2019. This may be due to the fact that the trained model adequately learns the continuous spatio-temporal features from 2010 to 2018 and thus the estimation performance of the model performs better in 2019, whereas its performance is poor in 2020 due to the lack of extraction of features for 2019. The above experiments indicate that the DA-CNN model has the generalization ability to a certain extent, but there is room for further improvement.

#### 4.4.3 Spatiotemporal distribution of the estimated MLD

Spatiotemporal characteristics of the MLD in 2019 are estimated by the DA-CNN model, and then evaluated to analyze the modelling accuracy. Figure 7 shows a comparison of average monthly estimated and Argo-derived MLDs with seasonal

variations in 2019. As shown, the estimated MLD in the BoB presents an asymmetric bimodal peak in the summer and winter of the Northern Hemisphere, with the maximum of about 40 m in summer being significantly larger than the maximum of about 25 m in winter. This is generally in line with the Argo-derived values. Moreover, Fig.8 displays the spatial distribution of the seasonality between the Argo-derived MLD and the estimated MLD, also in general consistent with that of the Argo-derived MLD in a typical month (February, May, August, and November) of each season, with the differences ranging from -4 to 4 m in most regions. There is a noticeable gradient trend in the Argo-derived MLD from northeast to southwest in May 2019, with depths deepening from 20 m in the northeast to 35 m in the southwest, which trend is also revealed by the proposed model while keeping the error within 4 m in most regions. Therefore, the model effectively captures the geographic variations of the MLD in the BoB. However, there are still some differences between the estimated MLD and the Argo-derived MLD in certain regions with complex ocean dynamics processes. For example, the MLD is overestimated in coastal regions in the northern BoB during February and August, with a difference of less than 6 m. This may be caused by the complex change in salinity due to a large number of runoff inputs and substantial precipitation in the region (Akhil et al., 2020). These differences indicate that the DA-CNN model made better estimations in open oceans compared to ocean areas of nonlinear signals due to complex dynamic processes.

#### 4.4.4 Performance verification of DA module

In this section, the accuracy of the MLDs that are estimated by the DA-CNN model in Case 4 and the original CNN model are compared. Figure 9 shows the variations of the monthly average RMSE and  $R^2$  for the two models in different months during 2019.

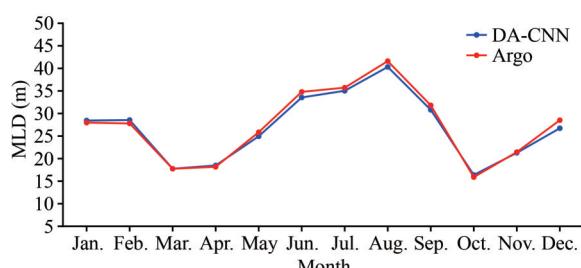


Fig.7 Monthly average MLD from Argo-derived and DA-CNN estimation at different months in 2019

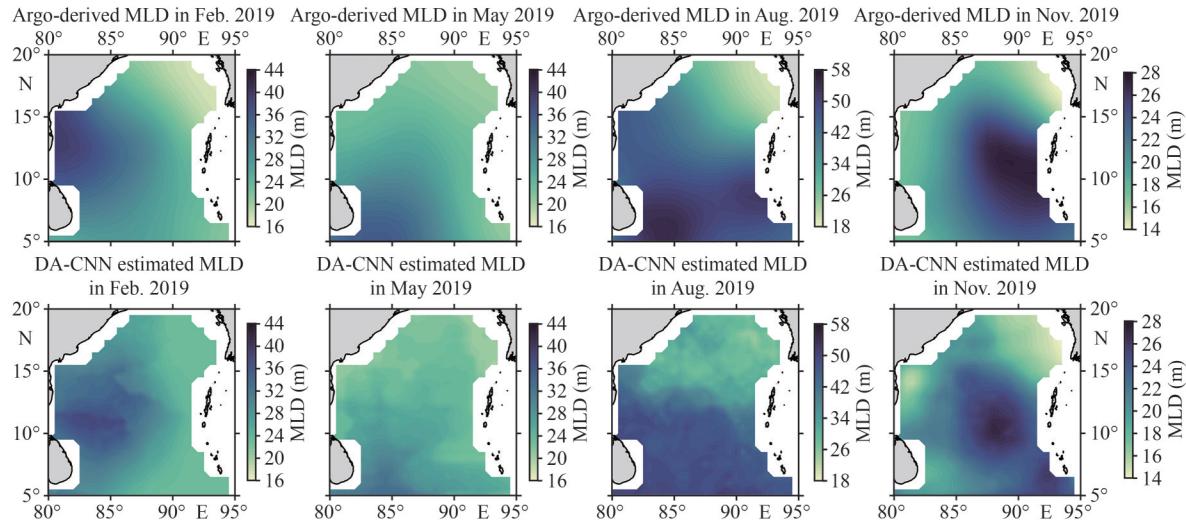
Table 5 Experimental results on universal parameters of the model

RMSE (m)	Feb-Model	May-Model	Aug-Model	Nov-Model
Feb.	<b>3.03</b>	4.92	6.33	7.3
May	4.95	<b>2.10</b>	3.78	5.74
Aug.	9.27	8.52	<b>4.77</b>	13.31
Nov.	4.70	4.12	4.95	<b>1.53</b>

The red value indicates the optimal RMSE for different models to estimate the current month.

Table 6 Results of the MLD estimation for the first four months of 2019 and 2020

RMSE (m)	Jan.	Feb.	Mar.	Apr.
2019	2.04	3.03	0.69	1.87
2020	8.58	4.45	2.30	2.13

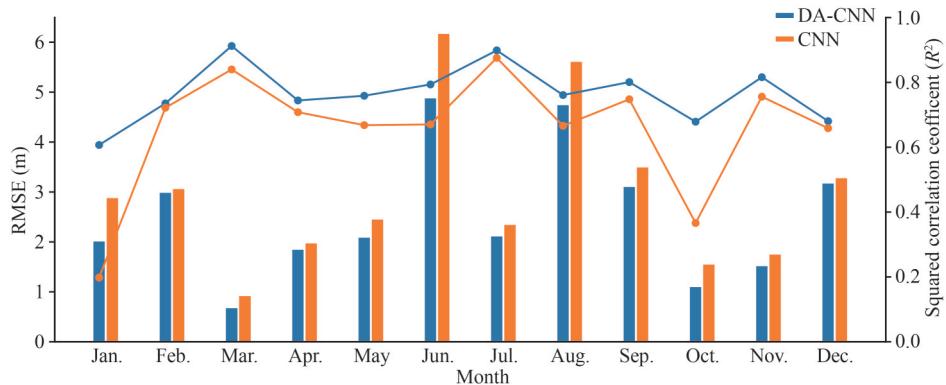


**Fig.8** Spatial distribution of the MLD from Argo-derived (top panel) and DA-CNN model estimation (bottom panel) in four typical months (February, May, August, and November) in 2019

It is indicated that the DA-CNN model and the original CNN model have good consistency in estimation accuracy during most months. For example, both models effectively capture the annual variations of MLD in February and December. Meanwhile, the differences in the RMSE as well as  $R^2$  values of both models are not obvious in the two months, with discrepancies of approximately 0.05 m and 0.01, 0.1 m and 0.03, respectively. Through a year, the DA-CNN model exhibits superior estimation compared to the CNN model. In particular, the estimation accuracy of the DA-CNN model significantly improved in January and October, with the RMSE reduced by 0.9 and 0.45 m, and the  $R^2$  increased by 0.5 and 0.3, respectively. This suggests that the DA module effectively enhances the nonlinearly fitting ability of the CNN model, although the Argo-derived MLD in January

and October of 2019 shows extreme phenomena compared to the same months in the rest years (Du et al., 2020). Notably, the estimation accuracy of both the CNN and D-CNN models is not good enough in June and August of 2019. This may be related to the fact that these two months underwent the onset and retreat processes of the strongest southwest monsoon in the BoB over the past 25 years (Greaser et al., 2020; Ratna et al., 2021). Overall, our results indicate that the CNN model, with its outstanding feature extraction capabilities on high-dimensional data, achieves favorable estimation results. Furthermore, the involvement of the DA mechanism significantly improves the estimation of MLD using the CNN model.

Moreover, the impact of the DA added onto the CNN model is validated by evaluating annual



**Fig.9** The monthly average RMSE (m) and  $R^2$  for the DA-CNN model and CNN model at different months in 2019

The bars indicate RMSE (m) and the lines indicate  $R^2$ .

mean estimation. Figure 10 displays the spatial distribution of the RMSE and  $R^2$  in the annual average estimated by the DA-CNN model and the original CNN model. As shown, the DA-CNN model outperforms that of the original CNN model along certain coasts in the northern and western BoB, although the estimation accuracy of both models is not good enough due to the fact that these regions are affected by many negative factors such as the large fluctuations in salinity and the complexity in ocean currents (Rao et al., 2010; Ray et al., 2022). For example, the annual average RMSE and  $R^2$  values for the DA-CNN model are improved by 2.0 m and 0.4 in northern regions ( $15^\circ\text{N}$ – $20^\circ\text{N}$ ,  $90^\circ\text{E}$ – $95^\circ\text{E}$ ), and 1.5 m and 0.3 in western regions ( $10^\circ\text{N}$ – $15^\circ\text{N}$ ,  $80^\circ\text{E}$ – $85^\circ\text{E}$ ) with respect to those of the CNN model, respectively. This indicates that the DA-CNN model may effectively estimate the MLD in regions with complex ocean dynamics. Quantitatively, the RMSE and  $R^2$  values for the DA-CNN model in terms of annual averages are significantly improved by 13.0% and 8.4% compared to those for the CNN model of 3.12 m and 0.78, respectively. Therefore, by integrating with the attention mechanism, the estimation accuracy of the DA-CNN model is effectively improved compared to the original CNN model.

#### 4.5 Estimation performance comparison with other models

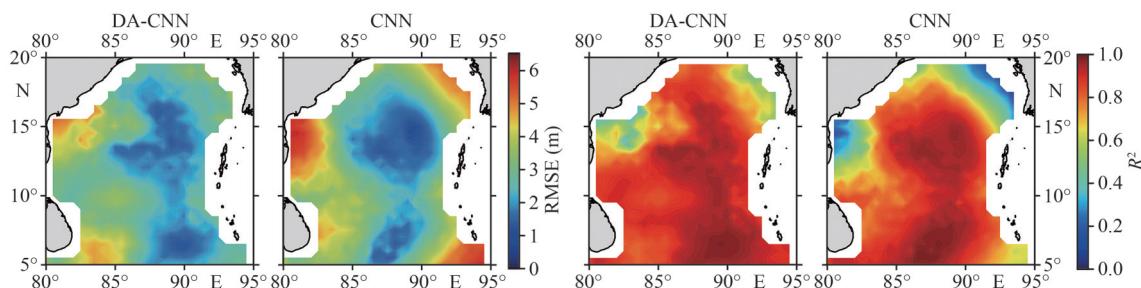
In this section, the results using the DA-CNN model in Case 4 are compared with the data-driven CNN model, RF model, and ANN model, as well as the physics-driven HYCOM model. The parameters of the CNN model use the same parameters as in the DA-CNN model except for the removal of the DA module from the model architecture, as listed in Table 2. The ANN model consists of three linear layers, with 64 and 128 neurons in each layer, respectively. The RF model constructs decision trees

by randomly extracting features and combining them from the dataset. In addition, the RF model depends on 4 key parameters of the number of decision trees in the model (n\_estimators), the minimum number of samples to split an internal node (min\_samples\_split), the minimum number of samples to split a leaf node (min\_samples\_leaf) and the maximum depth of the tree (max\_depth), using values of 100, 2, 4, and 10, respectively.

##### 4.5.1 Annual average performance among models

The annual average RMSE and  $R^2$  values for five estimation models are shown in Table 7. The DA-CNN model has the best estimation performance among models, as demonstrated by minimum RMSE and maximum  $R^2$  values. Figure 11a–f display the spatial distributions of the annual average MLDs obtained by the Argo and these models in 2019. Meanwhile, the differences between the Argo-derived MLD and that estimated by these models are also depicted in Fig.11g–k. As shown, the MLDs estimated by these models are overestimated in the northern BoB meanwhile underestimated in the western BoB. Nevertheless, the spatial distribution differences of the DA-CNN model are still the smallest. In comparison, the HYCOM model shows the worst performance among the five models. Here, in spite of the highest spatiotemporal resolution, the HYCOM encounters cases where the initial or boundary conditions are not suitable for local regions, resulting in the poor estimation of the regional MLD (Duerr et al., 2012). This further highlights the advantage of using AI models compared to the traditional physics-driven models in the estimation of the MLD.

The correlation densities of the annual average MLD in 2019 estimated by each model and the Argo-derived are presented in 2019 (Fig.12). It is shown that most data points of the DA-CNN model are closer to the equal value lines than those of the other models, while this model also has the



**Fig.10** Spatial distribution of the annual average RMSE (m) and  $R^2$  in 2019

**Table 7 Annual average RMSE for these models**

Model	Annual average RMSE (m)	Annual average $R^2$
DA-CNN	2.71	0.85
CNN	3.12	0.78
RF	3.24	0.76
ANN	4.64	0.63
HYCOM	7.11	0.29

minimum RMSE and maximum  $R^2$ . This indicates that the results estimated by the DA-CNN model are the most reliable. The reason may be that the introduced DA module helps the CNN model to learn the latent nonlinear relationships within the data more efficiently. Additionally, the results of the CNN and RF models for estimating MLD are almost equivalent but commonly show less accuracy compared to that of the DA-CNN model. Whereas, the density distributions of the RF model are relatively more scattered compared to those of the CNN model. Possibly, this is due to the low complexity of the RF model, making it difficult to capture complex local characteristics or reveal potential quantitative relationships. The ANN model estimates and HYCOM reanalysis data contain a large number of outliers, indicating a limited ability to estimate the MLD accurately in the BoB. Overall, the DA-CNN model exhibits the best accuracy in estimating the MLD in the BoB.

#### 4.5.2 Comparison of seasonal average performance of models

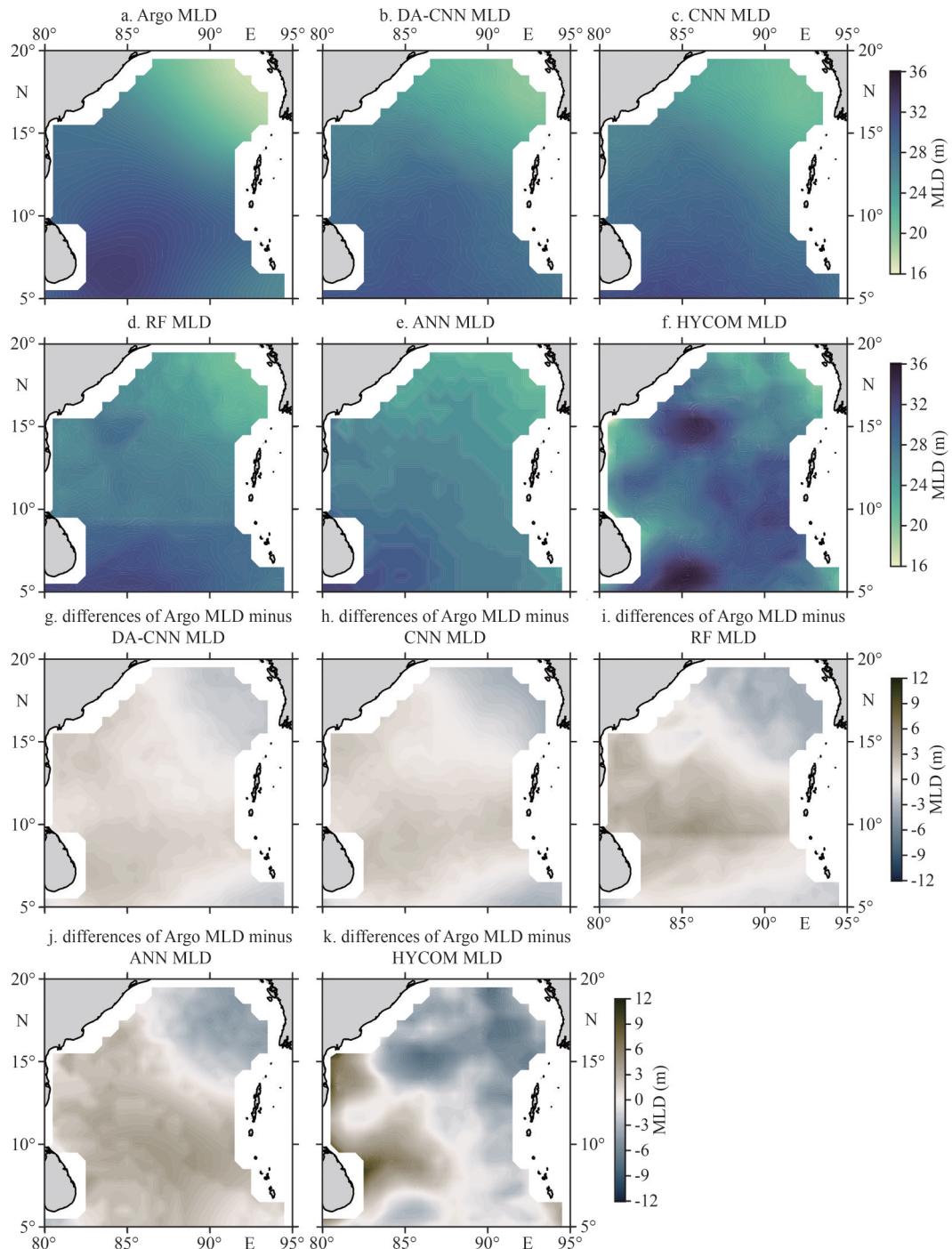
In this section, the estimation performance among the five models with seasonal variations is compared on the basis of averaged RMSEs in different seasons. Figure 13 shows the boxplot distribution of the seasonal and annual average RMSEs for the five models in 2019. The seasonal average RMSE values for these models are shown in Table 8. As shown in Fig.13, the DA-CNN and CNN models based on the convolutional architecture have fewer outliers and fewer offsets compared to the other models, reaffirming the excellent learning capability of convolutional layers for high-dimensional feature data. Moreover, the DA-CNN model shows the best estimation accuracy, as demonstrated by the minimum RMSE values of 1.52, 3.65, 1.87, and 2.66 m in the four seasons, respectively (Table 8). In addition, the summer MLD reaches peak values in summer (Fig.7), leading to a general decrease in the estimation accuracy of all models in this season, and the

estimation accuracy of the DA-CNN model is most substantially improved in the summer with respect to the other seasons (Fig.13). Therefore, the proposed DA-CNN model outperforms the other data-driven and physics-driven models, which effectively captures the spatiotemporal characteristics of the MLD in the BoB and accurately simulates seasonal variations, demonstrating robust and effective estimation capabilities.

## 5 CONCLUSION

This study aims to provide an estimation approach capable of assessing and analyzing the MLD in some typical ocean regions with complex dynamic processes, using the BoB as a case study. We develop a new DA-CNN model to estimate the MLD by integrating multi-source remote sensing data and Argo data. The multi-source datasets spanning 120 months from 2010 to 2019, with 653 data points for each variable per month, are collected from satellite observations and Argo gridded data in the BoB. The seven parameters (SST, SSS, SSH, USSW, VSSW, LON, LAT) and Argo-derived MLD are utilized as input and output variables of the model, respectively. Four groups of comparison experiments denoted as Case 1 to Case 4 are designed to verify the performance of the DA-CNN model with variable inputs in estimating the MLD in the BoB. The comparative results demonstrate that the model in Case 4 shows the best capability in capturing the complex features of the monthly MLD in the region. The DA-CNN model demonstrates improved accuracy in all months of 2019 compared with the original CNN model, with annual average RMSE and  $R^2$  values are 2.71 m and 0.85, respectively. This superior performance can be attributed to the better ability of the model incorporating the DA module to capture spatial features and characterize complex ocean processes. The performance of the DA-CNN model in Case 4 is also evaluated in multiple perspectives of spatial distributions in different months, comparison with other models, and seasonal variations. Overall, the developed DA-CNN model exhibits superior performance and can effectively estimate the MLD in the BoB regions.

The MLD estimated by the DA-CNN model in Case 4 not only shows good agreement with the Argo-derived MLD from a spatiotemporal distribution perspective but also simulates the two peaks of the MLD occurring in both the summer and

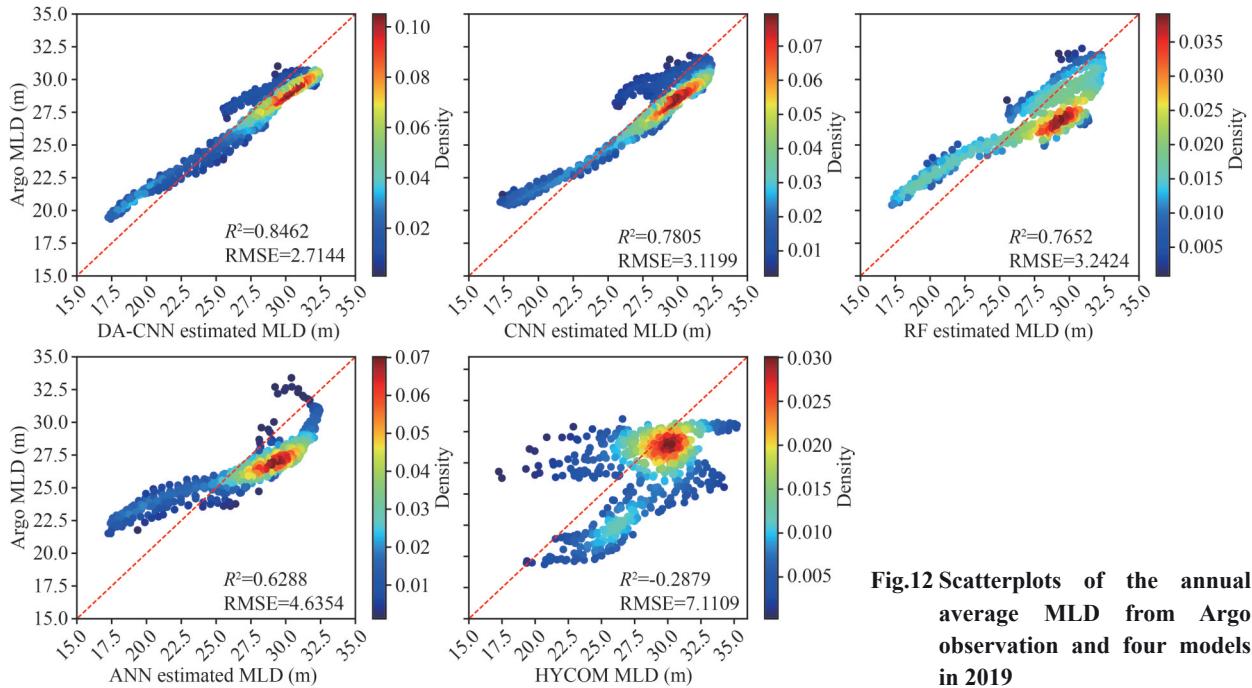


**Fig.11** Spatial distribution of annual average MLD in the BoB obtained from Argo (a), five models estimated MLDs (b–f) and the difference between the Argo-derived MLD in 2019 (g–k)

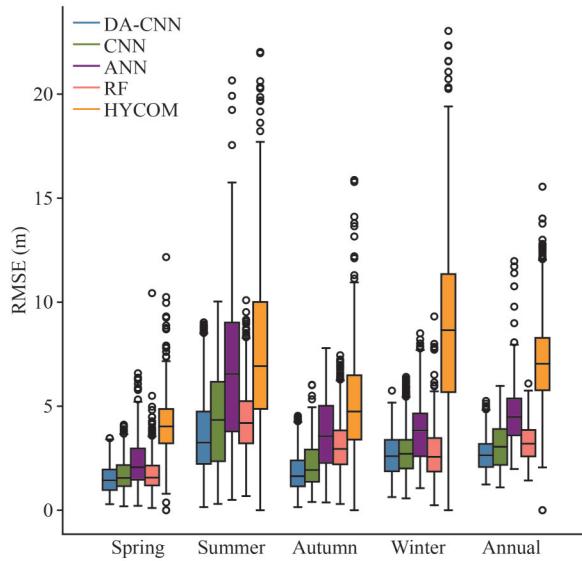
The brown region indicates that the difference is more significant than zero, meaning that the models underestimate the MLD values. The gray region means the models overestimate the MLD values.

winter in the BoB. Three experiments on model ablation show that the introduction of the DA module significantly improves the estimation accuracy of the DA-CNN model, surpassing the original CNN model, PAM-CNN model, and CAM-CNN model. Meanwhile, the performance of the

proposed model integrated with the attention mechanism has been effectively improved with respect to that of the original CNN model in some regions with complex ocean-atmosphere interactions. For example, the annual average RMSE and  $R^2$  values for the DA-CNN model are



**Fig.12** Scatterplots of the annual average MLD from Argo observation and four models in 2019



**Fig.13** Boxplot distribution of the seasonal and annual average RMSE (m) of five models in 2019

Boxes capture 25%–75% of the monthly RMSE values; the middle black line represents the median RMSE values, and the dots outside the box are considered outliers, whose values are 1.5×lower/upper quantile.

improved by 2.0 m and 0.4 in northern BoB, and 1.5 m and 0.3 in western BoB compared with those of the CNN model, respectively. The DA-CNN model in Case 4 is further compared with the RF, ANN model, and HYCOM model. The comparison results show that the DA-CNN model has the best performance among the five models, which can improve the RMSE and  $R^2$  values by 13.0% and

**Table 8** Seasonal average RMSEs for these models

Model	Spring RMSE (m)	Summer RMSE (m)	Autumn RMSE (m)	Winter RMSE (m)
DA-CNN	1.52	3.65	1.87	2.66
CNN	1.72	4.44	2.20	2.86
RF	1.73	4.29	3.11	2.72
ANN	2.30	6.64	3.64	3.74
HYCOM	4.12	7.81	5.06	8.77

8.4%, 16.3% and 10.6% compared to the original CNN model and RF, respectively. Meanwhile, the ANN model and HYCOM model show a large number of offsets compared to the Argo observations. Finally, the performance of these models with seasonal variations is further quantitatively evaluated. The results suggest that the DA-CNN model is also able to demonstrate the seasonality of the MLD, surpassing other models of the RF model, ANN model, and HYCOM model.

Our findings aid in detecting and monitoring the seasonal variation of the mixed layer in the BoB, offering valuable insights for further scientific understanding of oceanographic processes in this region. On the other hand, the DA-CNN model has limitations in some aspects such as estimating extreme anomaly events and interpreting the physical mechanisms of the results. Moreover, there is still potential improvement in the accuracy and generalization capability of the model due to the impact of the Argo gridded data and the inherent





