This is the Preprint Version.

The following publication H. Yang et al., "UniOcean: A Unified Framework for Predicting Multiple Ocean Factors of Varying Temporal Scales," in IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1-16, 2024, Art no. 4213116 is available at https://doi.org/10.1109/TGRS.2024.3506058.

UniOcean: A Unified Framework for Predicting Multiple Ocean Factors of Varying Temporal Scales

Hanchen Yang, Yuwei Chen, Jiannong Cao, *Fellow, IEEE*, Wengen Li*,
Yu Yang, Shuyu Wang, Jihong Guan, Rufu Qin, and Shuigeng Zhou, *Senior Member, IEEE*

***Abstract*—Accurate prediction of ocean factors (e.g., temperature and salinity) is crucial for plenty of applications, including weather forecasting, storm tracking, and ecosystem protection. Meanwhile, it is well-known that the ocean is a unified system and various ocean factors usually influence each other. For example, the changes in temperature would affect the distribution of salinity in ocean. However, existing studies for ocean factor prediction mainly focus on designing individual models for predicting specific factors and ignore the correlations between different factors, thus having potentials to be further improved. Therefore, we propose a unified framework UniOcean to predict multiple ocean factors simultaneously, and capture the correlations between them to improve the prediction accuracy. First, considering that ocean factors are usually collected with different temporal scales, we develop the fine-grained multi-scale data fusion module to integrate multiple ocean factors with different temporal scales, and effectively learn their hierarchical patterns at different levels. Then, since the correlations between ocean factors may vary across different time periods, the multi-factor correlation learning module is constructed to adaptively learn the dynamic correlations between different factors. Finally, we utilize the factor-specific towers to predict multiple ocean factors simultaneously. Experimental results on five real-world remote-sensing datasets demonstrate that UniOcean significantly improves the prediction accuracy by 11%-53% in terms of MSE for different ocean factors.**

***Index Terms*—Multiple ocean factors, different temporal scales, spatial-temporal prediction, unified model.**

I. Introduction

OCEAN covers more than two-thirds of the Earth, and plays a crucial role in regulating the global climate. Various ocean factors, e.g., sea surface temperature (SST), sea surface salinity (SSS), and ocean heat content (OHC), have been studied for decades to identify extreme climate events (e.g., El Nin˜o phenomena and typhoons), understand ocean circulation, and protect the ocean ecosystem [1]. With the development of sensing technology, more and more remote sensors have been deployed to monitor the changes of ocean factors [2], [3]. Various ocean datasets are thus collected and published by different organizations (e.g., National Aeronautics and Space Administration (NASA), European Centre for Medium-Range Weather Forecasts (ECMWF), and the Chinese Academy of Sciences (CAS)), which lays the solid foundation for analyzing the changing patterns and regularity of the ocean. For example, since 1979, NASA has launched a series of satellites [4], [5] to continually monitor the SST and SSS globally. Over the past few decades, there have been numerous studies utilizing the collected data for predicting various ocean factors, which hold great significance for a lot of applications, e.g., weather forecasting [6], fishing detection [7], and storm tracking [8].

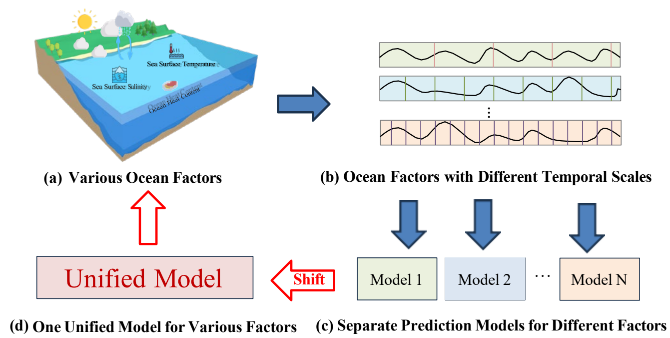

Fig. 1: The ocean system has various factors (a), and different factors are usually collected with different temporal scales (b). The existing methods predict the specific ocean factors with separate models (c). In contrast, our work (d) explores a unified model to predict multiple ocean factors simultaneously.

Recently, with the development of deep learning techniques, more and more data-driven methods [1], [9], [10] have drawn increasing attention in ocean factor prediction and shown significant improvement compared with traditional physical methods [11]–[13]. However, a major limitation of existing data-driven methods is that they mainly focus on designing separate models for predicting specific ocean factors. As shown in Fig. 1, existing methods often focus on predicting specific ocean factors, e.g., SST, SSS, and OHC, independently rather than predicting multiple factors simultaneously. In fact, the ocean is a unified system and different ocean factors record different aspects of the ocean system. These factors thus could provide complementary information for each other [14]. For example, SSS and SST are closely interconnected by solubility, precipitation, and evaporation. With the increase

Hanchen Yang is with the Department of Computer Science and Technology, Tongji University, Shanghai, China, and also the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China (email: neoyang@tongji.edu.cn).

Jiannong Cao and Yu Yang are with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China (email: {jiannong.cao, cs-yu.yang, 2203}@polyu.edu.hk).

Yuwei Chen, Wengen Li, Shuyu Wang, and Jihong Guan are with the Department of Computer Science and Technology, Tongji University, Shanghai, China (email: {2051505, lwengen, tj_sywang, jhguan}@tongji.edu.cn).

Rufu Qin is with the School of Ocean and Earth Science, Tongji University, Shanghai, China (email: qinrufu@tongji.edu.cn).

Shuigeng Zhou is with the School of Computer Science, Fudan University, Shanghai, China (email: sgzhou@fudan.edu.cn).

* Wengen Li is the corresponding author.

of SST, the movement of water molecules is accelerating, resulting in the separation of salt molecules and the reduction of SSS. Meanwhile, different ocean factors are recorded in the same spatial-temporal context and usually have similar temporal dependencies, e.g., tendencies, periodicity, and seasonality. Capturing such dependencies is essential for accurate prediction of ocean factors. Therefore, designing a unified framework to jointly model the changes of different ocean factors and their correlations is meaningful and promising to further improve the prediction accuracy.

Incorporating different ocean factor prediction tasks (e.g., SST prediction [6], SSS prediction [15], OHC prediction [16]) can be regarded as a typical multi-task learning (MTL) problem. MTL is a popular machine learning approach that trains a single model to perform multiple tasks simultaneously [17]. It has demonstrated superior performance in various fields, e.g., computer vision (CV) and natural language processing (NLP). MTL can learn robust and universal data representations to capture the shared knowledge among multiple tasks to improve their performance. Currently, there are already some studies on MTL methods in the ocean [18]–[21]. For example, MTL method was introduced to jointly predict the sea ice concentration and sea ice extent in the Arctic ocean [19]. Liu et al. [22] proposed a multi-task neural network model to simultaneously recognize abnormal signals and locate abnormal regions in the East China Sea. Despite the improvement, developing the aforementioned unified ocean factors prediction model remains technically non-trivial due to the following two challenges.

On the one hand, most existing MTL methods often operate on the data with the same temporal scales but are unable to model multi-scale remote sensing data [18], [19]. Due to the variations in sensing technology and data processing standards, the data of different ocean factors are usually of varying temporal scales, e.g., daily scale, weekly scale, and monthly scale. As shown in Fig. 2, three ocean factors are collected in the same period but with different temporal scales. When the MTL methods receive the multi-scale data as input, the temporal dependencies of different scales may exhibit diverse seasonality and tendencies, introducing significant complexity for the model to learn the consistent temporal dependencies. The existing MTL methods typically utilize the manually selected data with the same temporal scale, or employ specific techniques such as data interpolation, data resampling, and feature extraction to calibrate the data into the same temporal scale [23]. For instance, Eliot et al. [19] converted the daily sea ice concentration data into monthly averages to align it with the monthly sea ice extent data. However, such transformations may introduce additional noise and result in the loss of data details. As a result, the model may mistakenly learn the temporal dependencies across different temporal scales and exhibit poor prediction performance. Therefore, it is crucial to develop an end-to-end architecture that can directly capture fine-grained patterns in the multi-scale data.

On the other hand, the dynamic correlations between different ocean factors have not been well-captured. Current MTL methods basically employ a static module (e.g., linear layer and attention layer) or a combined training loss

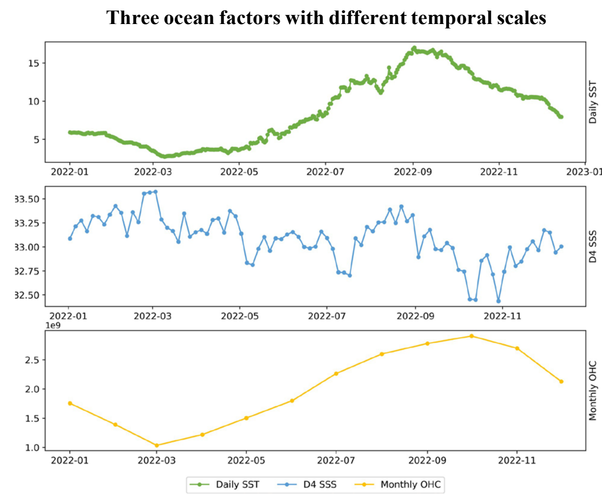

Fig. 2: The SST, SSS, and OHC data with different temporal scales, i.e., daily scale, scale of four days, and monthly scale, are collected by NOAA in 2022.

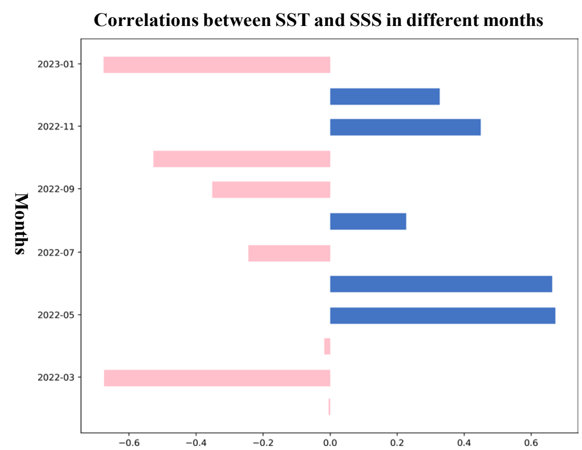

Fig. 3: The Pearson correlations between SST and SSS over different months, where blue and pink colors represent positive and negative correlations, respectively.

to integrate different ocean data analytic tasks [20]. However, the correlations between various ocean factors are complicated and may change over time. As shown in Fig. 3, we utilize the Pearson correlation coefficient [24] to measure the correlations between SST and SSS. It is clear that the correlations are dynamically changing over different months. Therefore, it is necessary to effectively capture the spatial-temporal variability in the correlations between ocean factors.

To solve the above challenging issues, we proposed a unified prediction framework, UniOcean, to predict multiple ocean factors of varying temporal scales simultaneously. First, we develop the *Fine-grained Multi-scale Data Fusion* module to fuse multiple ocean factors of different temporal scales through the hierarchical residual network. Then, the *Multi-factor Correlation Learning* module is constructed to adaptively learn the dynamic correlations between different ocean factors over time. Finally, we design the *Factor-specific Tower* module to produce the prediction results for each ocean factor. Specifically, our main contributions include:

- We propose a novel unified framework *UniOcean* to achieve multiple ocean factors prediction simultaneously. UniOcean can efficiently align the different temporal scales of multiple ocean factors, capture the unified patterns, and learn the dynamic correlations between ocean factors to boost the prediction performance.
- We develop a *Fine-grained Multi-scale Data Fusion* module to capture the temporal dependencies of ocean factors at different scales and then introduce the hierarchical residual network to fuse these dependencies.
- We design a *Multi-factor Correlation Learning* module to capture the dynamic correlations between different ocean factors, in which the correlations are fully self-adaptive.
- Experimental results on five remote-sensing datasets demonstrate that UniOcean is a universal framework to enhance various models to achieve significant prediction performance on multiple ocean factors.

II. Data

A. Datasets

In this study, we utilize five real-world datasets, i.e., Daily SST, D4 SSS, Weekly SST, Monthly SSS, and Monthly OHC, to evaluate the performance of our method. They are all reanalysis remote sensing datasets of high quality and have been processed to fill the missing values by their publishers. We summarize these datasets in Table I.

TABLE I: Summary of datasets.

Dataset	#Scale	#Period	#Spatial Coverage
Daily SST	Daily	1981-now	Global
D4 SSS	4 Day	2011-now	Global
Weekly SST	Weekly	1981-now	Global
Monthly SSS	Monthly	1960-now	Global
Monthly OHC	Monthly	1940-now	Global

- **Daily SST.** A high-resolution sea surface temperature dataset from Physical Sciences Laboratory (PSL)[1]. The data is processed using Optimum Interpolation to achieve daily scale. The temporal coverage is from 1981 to the present.
- **D4 SSS.** A public dataset that is produced by the Earth and Space Research (ESR) in collaboration with the Remote Sensing Systems (RSS)[2], and is derived from satellite L-band radiometer measurements from 2011 to now. The data is collected every four days.
- **Weekly SST.** A dataset that collects the weekly global sea surface temperature from NOAA[3]. The temporal coverage is from 1981 to now.
- **Monthly SSS.** A monthly ocean salinity dataset that is provided by the Institute of Atmospheric Physics, Chinese Academy of Sciences[4]. The temporal coverage is from 1960 to now.
- **Monthly OHC.** A monthly ocean heat content dataset that is also provided by the Institute of Atmospheric Physics, Chinese Academy of Sciences. The temporal coverage is from 1940 to now.

[1]Dataset available at: https://www.ncei.noaa.gov/products
[2]Dataset available at: https://podaac.jpl.nasa.gov
[3]Dataset available at: https://www.esrl.noaa.gov/psd

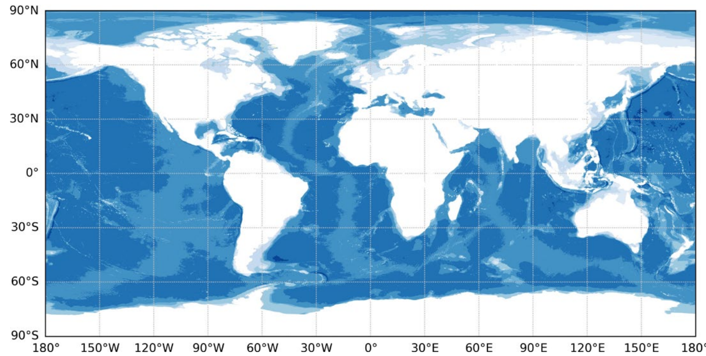

Fig. 4: Study area (blue): global ocean data (180° E–180° W, 90° S–90° N) with the spatial resolution of 1°×1°.

B. Study Area

In this work, as shown in Fig. 4, we conduct the study on the global ocean (180° E–180° W, 90° S–90° N) with the spatial resolution of $1° \times 1°$. Thus, the total number of grid regions is $180 \times 360 = 64800$. The values on the land a, i.e., the white areas in Fig. 4, are set to 0.. Thus, the total number of grid regions for the ocean is about 46000.

III. Related Works

In this section, we overview the existing methods for ocean factor prediction and multi-task learning, and give a discussion about the advantages and limitations of these methods.

A. Ocean Factor Prediction

Prediction of ocean factors (e.g., SST, SSS, and OHC) is an important research topic in ocean science. It aims to understand the regularity of past observations and accurately predict future observations, benefiting a lot of applications such as weather forecasting, disaster warning, and ocean environment protection [2], [3]. Various methods have been developed to predict ocean factors, and these methods can be roughly divided into two categories, i.e., physical models and data-driven models.

Physical models. The basic idea of physical models is to combine the laws of physics, e.g., Newton's laws of motion, the law of conservation of energy, and the seawater equation of state, to predict ocean factors. For instance, the Global Forecast System (GFS) [11] is a weather forecast model that carefully selects parameters to combine multiple physic laws (e.g., the Navier-Stokes equation and solar radiation) to predict the change of SST. Another representative method is the CMCC (Centro Euro-Mediterraneo sui Cambiamenti

[4]Dataset available at: http://www.ocean.iap.ac.cn

Climatici) model [12], [13] which is an ensemble model that integrates multiple physical models (e.g., C-GLORS, OceanVar, and NEMO) [12] to achieve accurate SST prediction. Although physical models have been widely used in many applications, they require a good understanding of the underlying changing mechanisms of ocean factors to choose the critical parameters for physical models. However, such mechanisms are usually complicated, which makes it challenging to use only the manually settled parameters to effectively learn and monitor the patterns of ocean factors.

Data-driven models. The latest data-driven models for ocean factor prediction can be further divided into traditional machine learning-based models and deep learning-based models. *Traditional machine learning-based methods*, e.g., vector autoregressive models [25], autoregressive integrated moving average (ARIMA) [26], hidden Markov models (HMM) [27], and support vector machines (SVM) [28], have been widely used to predict ocean factors. For example, Xue et al. [27] proposed a seasonally varying Markov model constructed in a multivariate space to predict SST. A feed forward neural network [29] was used to predict the Chl-a concentration in Lake Kasumigaura of Japan and achieved better performance than the physical methods. Li et al. [30] combined kernel Granger Causality analysis (KGC) and SVM to predict the daily sea ice concentration. Although these methods can predict the trend of ocean factors to a certain extent, their overall prediction accuracy is still low. When the prediction length is over one month, the model's performance deteriorates dramatically due to the gradient vanishing issue.

(2) *Deep learning-based models*, e.g., RNN, CNN, and transformers, have been widely used for predicting ocean factors because of their superior ability to model the complex dependencies among data [31]–[35]. For example, Yang et al. [36] combined the Markov random field with LSTM to predict the SST, achieving better results than SVM and HMM models. Wang et al. [37] used the fully convolutional neural networks (FCNN) model to predict sea ice concentration along the east coast of Canada and achieved high prediction accuracy. Self-attention-based methods such as Transformer [38] and its variants [39] achieve high-quality SST prediction by utilizing the non-autoregressive mechanism. In addition, some recent studies find that it is promising to combine physical models with data-driven methods [40]. Taking the SST prediction as an example, Arka et al. [41] integrated traditional physical laws (e.g., temperature density and energy conservation) and recurrent graph networks (RGN) to predict SST, which have achieved better results than RGN model. In sum, various deep learning-based methods dominate the prediction of ocean factors [1].

B. Multi-task learning

Multi-task learning (MTL) is a fundamental learning paradigm in machine learning. MTL leverages the correlations among multiple related tasks to improve the performance of all tasks [17]. Over past decades, MTL has attracted much attention in various fields, e.g., computer vision (CV) [42], natural language processing (NLP) [43], [44] and spatial-temporal data mining (STDM) [45], [46]. Specifically, in the STDM field, there are numerous MTL-based studies for prediction problems. For example, Zhang et al. [47] proposed the multi-task deep learning model (MDL) to utilize a convolutional neural network to capture the spatial dependencies and employ the linear fusion layer to jointly predict the traffic flow on the road and region level. Wang et al. [48] developed a multi-task adversarial spatial-temporal network (MT-ASTN) to jointly capture the shared patterns of human crowd inflow and outflow for prediction. Zhang et al. [49] proposed an LSTM-based MTL method to jointly consider the taxi pick-up and drop-off demand together for prediction. During training, the average loss across the two prediction tasks (pick-up and drop-off demand) is calculated, encouraging the model to balance the performance on both tasks and achieve good performance.

Specifically, there are several multi-task learning models for predicting ocean factors. Ling et al. [18] designed a multi-task machine learning framework to jointly predict short-term and long-term abnormal SST prediction in the Indian Ocean. Han et al. [50] designed a MasterGNN model utilizing the recurrent graph neural networks to predict the air PM2.5, weather temperature, pressure, humidity, wind speed, and wind direction together. Liu et al. [22] proposed a multi-task convolutional neural network (MTCNN) model to simultaneously recognize submarine cable magnetic anomaly (SCMA) signals and locate abnormal regions in the East China Sea. However, these models are restricted to tasks with the same temporal scale data, ignoring how to deal with multi-scale data, which is more common in remote sensing data.

C. Discussion

According to the above overview, the main limitation of existing studies on ocean factor prediction is that they cannot sufficiently capture and utilize the correlations between different factors to enhance the prediction accuracy. Although some MTL methods utilize the linear fusion layer and combined loss to consider the correlations, they still fall short in capturing the dynamic correlations over time. Moreover, due to the varying temporal scales of different ocean factors, it is challenging to directly apply existing methods to learn the unified patterns of multiple ocean factors. To solve these issues, we proposed a novel approach, UniOcean, to align the multi-scale ocean factors and incorporate fine-grained patterns together to boost the prediction performance.

IV. Problem Definition

Definition 1: Ocean Factor Prediction. Given the historical observations $X = \{X_1, X_2, \cdots, X_t, \cdots, X_T\} \in \mathbb{R}^{N\times T}$ of one ocean factor, e.g., SST, SSS or OHC, each $X_t = \{x_{1,t}, x_{2,t}, \cdots, x_{N,t}\}^T \in \mathbb{R}^{N\times 1}$ records the ocean factor values at N different spatial locations at the time step t. Ocean factor prediction aims to predict the future value $\hat{Y} = \{\hat{Y}_{T+1}, \hat{Y}_{T+2}, \cdots, \hat{Y}_{T+T_\tau}\} \in \mathbb{R}^{N\times T_\tau}$ for upcoming T_τ time steps based on the historical data. Thus, prediction problem is to seek a model F as follows:

$$\hat{Y} = F_\theta\{X\} \tag{1}$$

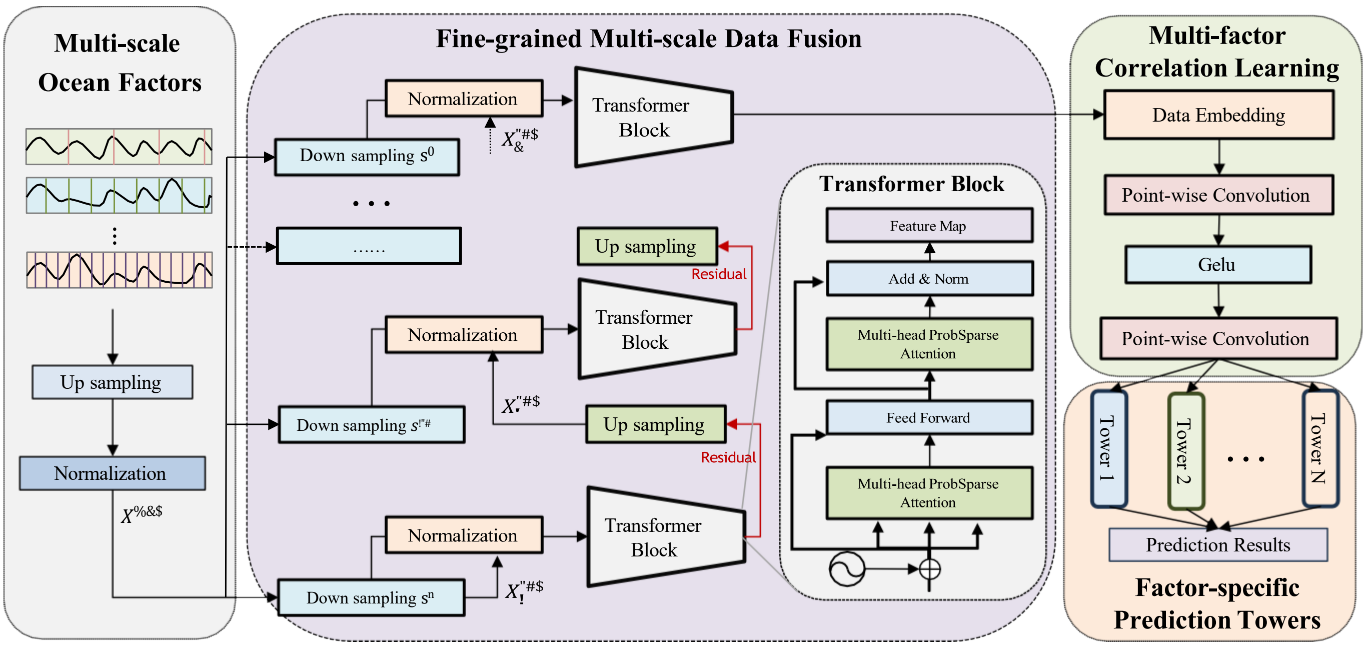

Fig. 5: The proposed unified prediction framework (UniOcean). The multi-scale ocean factors data is input into the *Fine-grained Multi-scale Data Fusion* module to learn patterns at different scales effectively and fuse them through the hierarchical residual network. Next, the *Multi-factor Correlation Learning* module is constructed to adaptively learn the dynamic correlations between different factors. Finally, we utilize the *Factor-specific Prediction Towers* to predict all ocean factors simultaneously.

where θ denotes all the learnable parameters in the model.

Definition 2: Multiple Ocean Factor Prediction. Given the historical observations of M ocean factors $X^{MF} = \{X^1, X^2, \cdots, X^i, \cdots, X^M\}$, each X^i represents the observations of factor i. Multiple ocean factor prediction aims to seek the model F to learn a unified prediction framework for all factors to improve prediction performance, i.e.,

$$\hat{Y}^{MF} = F_\theta\{X^1, X^2, \cdots, X^i, \cdots, X^M\} \tag{2}$$

where θ denotes all the learnable parameters in the model, $\hat{Y}^{MF}$ represents the prediction result of all ocean factors.

In general, each ocean factor X^i may have different scales s. It is necessary and important to consider the multi-scale characteristics in predicting multiple ocean factors. Thus, we denote this problem as:

$$\{\hat{Y}^1_{s1}, \hat{Y}^2_{s2}, \cdots, \hat{Y}^M_{sm}\} = F_\theta\{X^1_{s1}, X^2_{s2}, \cdots, X^M_{sm}\} \tag{3}$$

where θ denotes all the learnable parameters in the model, $\hat{Y}^i_s$ represents the prediction result of each ocean factor, and si denotes the scale of each factor X^i.

V. Method

A. Framework of UniOcean

Fig. 5 illustrates the framework of UniOcean. Specifically, we first design the *Fine-grained Multi-scale Data Fusion* module to integrate multi-scale ocean factors to capture the temporal dependencies at different levels and fuse them via the hierarchical residual network. Then, we propose the *Multi-factor Correlation Learning* module to aggregate the data embedding of different ocean factors using point-wise convolution and adaptively capture the dynamic correlations between the factors. Finally, we develop the *Factor-specific Prediction Towers* to predict the results for all ocean factors at the same time. More details are elaborated as follows.

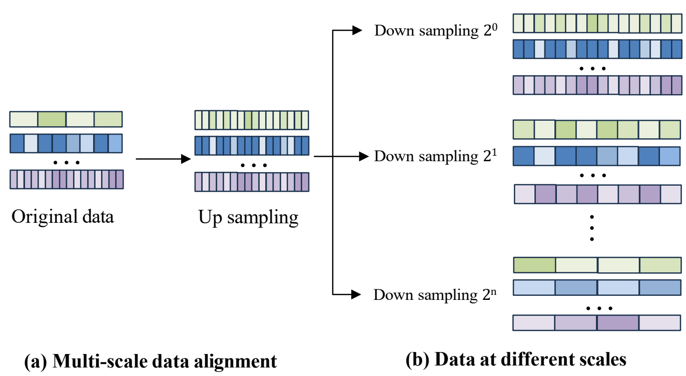

Fig. 6: Illustration of the data scales at different layers in the fine-grained multi-scale data fusion module.

B. Fine-grained Multi-scale Data Fusion

Due to the difference in temporal scales, the temporal dependencies of ocean factors may exhibit diverse seasonality and tendency, which introduces significant complexity for the model to learn such dependencies. As a result, the prediction model may mistakenly learn patterns across different scales of ocean factors and exhibit poor performance. To solve this problem, we design the Fine-grained Multi-scale Pattern Fusion module, which iteratively applies multiple transformer blocks to learn the temporal dependencies at different scales and utilizes hierarchical residual networks to fuse the learned dependencies. Concretely, the Fine-grained Multi-scale Pattern Fusion module consists of three components, i.e., multi-scale data alignment, fine-grained pattern learning, and hierarchical residual network.

1) **Multi-scale data alignment**: Given the historical data of different ocean factors $X^{MF} = \{X^1, X^1, \cdots, X^M\}$, as shown in Fig. 6(a), we first select the highest scale among all M factors as S_h and initially utilize a simple upsampling layer to align all the data to the highest scale by repeating the values, i.e.,

$$X_{up}^{MF} = \text{Upsamping}(X^{MF}, S_h) \tag{4}$$

For example, we use the upsampling layer to align monthly SSS and weekly SST data into weekly scale by linear data interpolation. After that, we can get the same-scale data of all ocean factors. Although this approach provides a simple way to align data to the same scale, the temporal dependencies from different scales remain hidden and are not learned adequately.

2) **Fine-grained pattern learning**: To capture the temporal dependencies from different scales, we first introduce a set of temporal scales $S = \{s^n, s^{n-1}, \cdots, s^1, s^0\}$ to downsample the original data to different scales, where s is the scale factor. For example, as shown in Fig. 6(b), if we have the scale factor $s = 2$, the scale set S will consist of consecutive powers of 2, i.e., $S = 2^2, 2^1, 2^0$. Concretely, we downsample the X^{MF} by a scale factor of s^n via an average pooling operation to get the input data X^{Enc} as:

$$X^{Enc} = \text{AveragePooling}(X^{MF}, s^n) = \frac{1}{s^n}\sum_{t}^{t+s^n} X \tag{5}$$

where AveragePooling() calculates the mean value of the data $\{X_i, X_{i+1}, \cdots, X_{i+s^n}\}$ for each factor.

Then, we utilize the transformer blocks to learn the dependencies at different temporal scales. Transformer is a fundamental model widely used in natural language processing (NLP) and Computer Vision (CV) and plays a crucial role in capturing the contextual relationships and dependencies within a sequence of input data. UniOcean is an universal framework which could easily combine with different variation of transformer block (e.g., Informer, Autoformer, and Fedformer). Here, we introduce basic transformer block [51]. In each transformer block, we encode the positional information of data as:

$$\begin{aligned} PE_{(pos,2i)} &= sin(\frac{pos}{10000^{2i/d}}) \\ PE_{(pos,2i+1)} &= cos(\frac{pos}{10000^{2i/d}}) \end{aligned} \tag{6}$$

where pos is the temporal position of a certain item of the input data, and d is the embedding dimension. After calculating the positional embedding, we capture the temporal patterns at different scales. The transformer blocks calculate self-attention using different input data, i.e.,

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}}) \tag{7}$$

$$\begin{aligned} MultiHead(Q, K, V) &= Concat(head_1, \cdots, head_h) \\ head_i &= Attention(QW_i^q, KW_i^k, VW_i^v) \end{aligned} \tag{8}$$

where W represents the learnable parameters of the transformer block, and Q, K, V represent the position encoding of the input data.

3) **Hierarchical residual network**: To fuse the patterns of different scales inspired by [52], we construct the hierarchical residual network to model the patterns from bottom to top. Specifically, the input of transformer block i consists two parts, i.e., the encoder X_i^{Enc}, and the decoder X_i^{Dec}. The encoder X_i^{Enc} is the down sample data of X_{up}^{MF} with scales i. The decoder is obtained using linear interpolation to upsample the output of transformer block $i-1$. Thus, it can align the encoder and decoder data to the same scale to feed into the transformer block.

$$\begin{aligned} X_i^{Out} &= \text{TransformerBlock}(X_i^{Enc}, X_i^{Dec}) \\ where: X_i^{Enc} &= \text{LinearInterpolation}(X_{i-1}^{Out}, s) \end{aligned} \tag{9}$$

where LinearInterpolation() imputes the missing values by assuming a constant rate of change between the known values, and the X_i^{Out} and X_{i-1}^{Out} respectively represent the outputs of transformer blocks i and $i-1$.

Algorithm 1 Fine-grained Multi-scale Data Fusion

Data: The data for different ocean factors $X^{MF} = \{X^1, X^2, \cdots, X^M\}$, scale factor s, and a set of temporal scales $S = \{s^n, s^{n-1}, \cdots, s^1, s^0\}$.
X_{up}^{MF} = Upsampling(X^{MF})
for *layer i in* 0 **to** n **do**
  X_i^{Enc} = AveragePooling(X_{up}^{MT}, s^i)
  if *i equal to 0* **then**
    X_i^{Dec} = X_i^{Enc}
  else
    X_i^{Dec} = LinearInterpolation(X_{i-1}^{Out}, s);
  end
  X_i^{Out} = TransformerBlock(X_i^{Enc}, X_i^{Dec})
end
Result: The fused multi-scale pattern embedding.

Algorithm 1 presents the pseudocode for the whole process of fine-grained multi-scale data fusion module. Finally, we obtain the aligned embedding X^{Emb} of multi-scale data as the input for the multi-factor correlation learning module.

C. Multi-factor Correlation Learning

The multi-factor correlation learning module aims to leverage the shared information and dynamic correlations between factors to improve the overall prediction performance. To this end, we construct two point-wise convolution layers to capture the correlations between ocean factors over time.

Point-wise convolution, also known as 1x1 convolution, is a type of convolutional operation commonly used in deep learning architectures. Point-wise convolution can operate on the channel dimension of the input to learn the correlations of ocean factors. In detail, as shown in Fig. 7, the input data embedding $X^{Emb} \in \mathbb{R}^{D \times F}$ is generated by the multi-scale data fusion module, and we set the two kernel dimensions of point-wise convolutions as $F \times d$ and $d \times F$, respectively, where F represents the number of ocean factors, and d

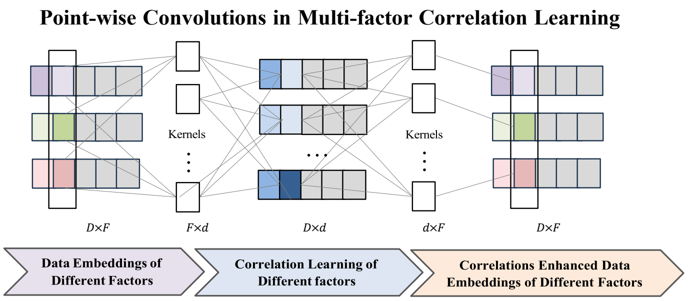

Fig. 7: Illustration of point-wise convolution to capture the correlations between factors.

denotes the data embedding dimension. For the first point-wise convolution layer, by setting the kernel dimensions as $F \times d$, we ensure that only the data embedding of different ocean factors can interact with each other. We utilize a sliding window to calculate the correlations of different factors at the specific window (period). This allows the model to capture the dynamic correlations among different ocean factors and map such correlations to a hidden space of dimensions $D \times d$, i.e.,

$$X^{Hid} = \sum_{F} X^{Inp} \times \text{Kernel}(F, d) \tag{10}$$

where X^{Inp} represents the data input of the correlation learning module, and X^{Hid} is the hidden embedding of the correlations between different ocean factors.

Then, we use the Gelu (Gaussian Error Linear Unit) function to activate the important correlations in X^{Hid}, i.e.,

$$X^{Hid} = \text{Gelu}(X^{Hid}) \tag{11}$$

$$\text{Gelu}(x) = \frac{1}{2}\left(1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right) \cdot x \tag{12}$$

where the erf() is an error function defined as $\frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt$, which provides smoother weights for effectively training.

After that, we employ another point-wise convolution layer to transform the learned correlation of different factors into data embedding, thereby obtaining data embedding that are enhanced by factors' correlations.

$$X^{Out} = \sum_{T} X^{Hid} \times \text{Kernel}(d, F) \tag{13}$$

where X^{Hid} represents the learned dynamic factor correlations, T denotes the total number of kernels, and X^{Out} is the correlation-enhanced data embedding of different ocean factors.

D. Factor-specific Prediction Towers

Finally, we utilize the factor-specific prediction towers to generate the prediction results for all ocean factors. For each factor, the corresponding tower is one multi-layer perception (MLPs), i.e.,

$$\hat{Y}^{MF} = WX^{Out} + b \tag{14}$$

where $W \in \mathbb{R}^{N \times d \times T}$ represents the learnable parameters, b is the bias, $X^{Out} \in \mathbb{R}^{N \times d}$ is the data representation in latent space, and d is the hidden dimension. For different ocean factors, the output length corresponds to their original scales. We choose L1 loss as the training objective and calculate the loss at each time step to optimize the prediction model. As a result, the loss function for multiple ocean factors prediction in UniOcean is formulated as follows:

$$L_{pred} = L(F_\theta) = \sum_{i=t}^{t+\tau} \left\| Y_i^{MF} - \hat{Y}_i^{MF} \right\|_1 \tag{15}$$

where F_θ denotes all the learnable parameters, Y_i^{MF} and $\hat{Y}_i^{MF}$ are the ground truth and the prediction results, respectively, at time step i.

VI. Experiments

We conduct experiments on five datasets to demonstrate the superiority of the proposed UniOcean method against multiple strong baseline methods. We also evaluate the impacts of the hyper-parameters and the effectiveness of model components. Moreover, we present visualizations to showcase the outcomes of our method as well as the baseline approaches, providing clear and concise comparisons. The codes are public available in the online repository[5].

A. Datasets of Ocean Factors

To unify the time range of all datasets, we conduct two sets of ocean factor prediction experiments, focusing on fine-grained and coarse-grained prediction, respectively. First, we test the performance of coarse-grained ocean factor prediction (30 years), i.e., SST prediction, SSS prediction, and OHC prediction. It includes weekly SST, monthly SSS, and monthly OHC datasets, and we chose 30 years of data from January 1993 to December 2022. The second set of experiments is to test the fine-grained ocean factor prediction (10 years), consisting of daily SST prediction and D4 SSS (collected per 4 days) prediction spanning from 2011 to 2022. In both datasets, 70% data is allocated for training, 10% for validation, and 20% for testing.

B. Baselines

We compare our method with the following seven baselines, including two typical Recurrent Neural Network (RNN) based models and five state-of-the-art Transformer models.

- **GRU** (Cho et al. 2014 [32]): a variant of recurrent neural networks that overcomes the vanishing gradient problem in traditional RNNs by employing gating mechanisms.
- **ConvLSTM** (Shi et al. 2015 [53]): A CNN-based model that uses convolutional neural networks to extract the spatial correlations and uses LSTM to model the temporal dependencies.
- **Informer** (Zhou et al. 2020 [39]): An advanced transformer-based model designed to address the challenges of long sequence time-series forecasting with improved efficiency and accuracy.

[5]https://github.com/Neoyanghc/Uniocean

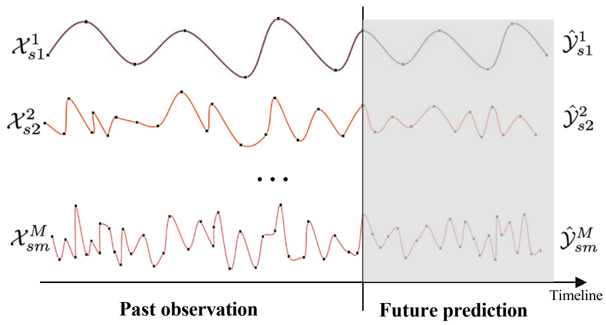

Fig. 8: Illustration of the multi-scale data prediction scheme.

- **Autoformer** (Wu et al. 2021 [54]): A transformer-based model that employs decomposition transformer with auto-correlation mechanisms specifically tailored for long sequence time-series forecasting.
- **FEDformer** (Zhou et al. 2022 [55]): A time series forecasting model that learns a graph structure among multiple time series and forecasts them simultaneously with a GNN.
- **MTformer** (Wu et al. 2021 [54]): A variation of Informer that predict all tasks together.
- **Scaleformer** (Mohammad et al. 2023 [52]): A time series forecasting model that utilizes a scale-aware structure to capture multi-scale information in one model.

C. Experimental Setups

Here, we briefly introduce the prediction scheme, experimental environment, parameter setting, and evaluation metrics.

Prediction scheme. All methods predict the results (i.e., SST, SSS, and OHC) of the following $\tau = 32, 64, 96$ time steps from the present time, based on the historical data of the last $T = 32, 64, 96$ time steps. For example, for the daily SST dataset, the $T = 32$ means that we utilize past values of 32 days to predict the next 32 days. Fig. 8 illustrates the prediction scheme to utilize the data with different temporal scales for prediction, where black squares mean the data samples and different factors within the same period may have different numbers of data samples. In our prediction task, we use the same historical temporal period for all factors to predict future values.

Experimental environment. All the deep-learning based models are implemented in Python with Pytorch 1.13.1. We use the source code and hyperparameters used in the original papers to evaluate the baseline methods on all datasets. All the models are run on a server with four NVIDIA 3090 GPUs. We optimize all the deep learning-based models by the Adam optimizer with a maximum of 50 epochs, and the early stopping is employed in validation loss to avoid over-fittings.

Parameter setting. We repeat each experiment ten times, and the best parameters for all deep learning-based models are chosen through a careful parameter-tuning process. The hyperparameters for our models are chosen through experiments, which are introduced in detail in Section VI-D. The batch size is set to 32, 24, and 16 for the experiments with prediction lengths of 32, 64, and 96, respectively. The original

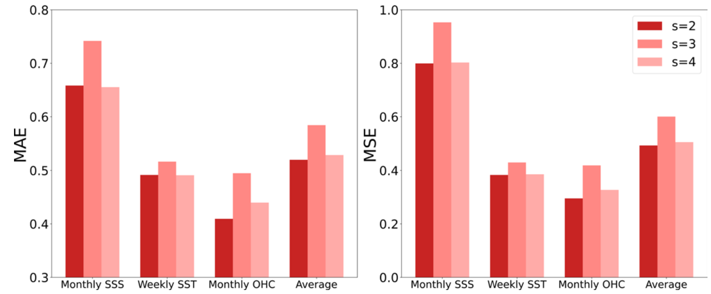

Fig. 9: The comparisons of the MSE and MAE results of different scales in UniOcean framework, i.e., 2, 3, 4.

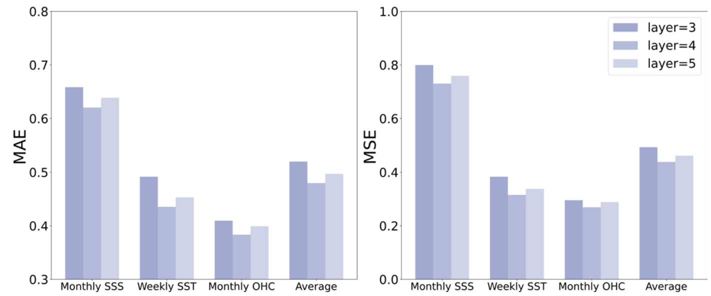

Fig. 10: The comparisons of the MSE and MAE results of the different number of layers in the hierarchical residual network in UniOcean framework, i.e., 3, 4, 5.

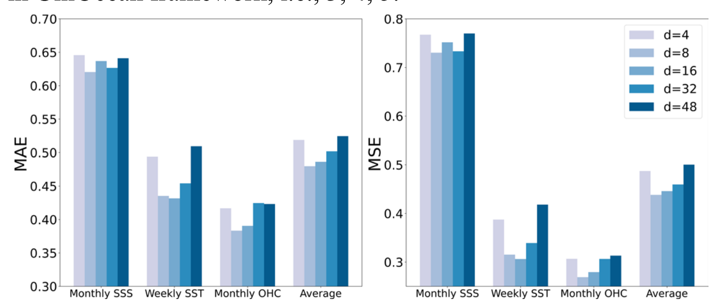

Fig. 11: The comparisons of the MSE and MAE results of different hidden dimensions of multi-factor correlation learning in UniOcean framework, i.e., 4, 8, 16, 32, 48.

learning rate is set to 0.0001 and halves every four epochs. The optimization problem is solved via back-propagation.

Evaluation metrics. We use two widely used metrics, i.e., Mean Absolute Error (MAE) and Mean Square Error (MSE), to measure the performance of prediction models, i.e.,

$$\begin{aligned} \mathrm{MAE}(Y, \hat{Y}) &= \frac{1}{|\tau|} \sum_{i \in \tau} |Y_i - \hat{Y}_i| \\ \mathrm{MSE}(Y, \hat{Y}) &= \frac{1}{|\tau|} \sum_{i \in \tau} (Y_i - \hat{Y}_i)^2 \end{aligned} \tag{16}$$

where $Y = Y_1, \ldots, Y_\tau$ denotes the ground truth, $\hat{Y} = \hat{Y}_1, \cdots, \hat{Y}_\tau$ represents the predicted values, and τ denotes the time steps to be predicted. In our experiments, τ is set to 32, 64, and 96, respectively.

D. Parameter Study

To study the effects of hyperparameters, we conduct a parameter study on the three core hyper-parameters of our framework, i.e., the scale factor s in the multi-scale fusion

TABLE II: The result comparison of 7 baseline models on coarse-grained ocean factor prediction (30 years), i.e., SST prediction, SSS prediction, and OHC prediction. Our UniOcean enhanced methods significantly improve compared with corresponding methods, and UniOcean-Fedformer achieves the best overall performance.

Methods	Prediction length=32						Prediction length=64						Prediction length=96						Avg
	Monthly SSS		Weekly SST		Monthly OHC		Monthly SSS		Weekly SST		Monthly OHC		Monthly SSS		Weekly SST		Monthly OHC		
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	
GRU	1.499	0.888	1.312	0.902	0.978	0.800	1.539	0.945	1.595	0.977	0.987	0.803	1.996	1.089	2.077	1.096	1.113	0.854	1.371
ConvLSTM	1.244	0.839	1.739	1.015	1.241	0.829	1.165	0.788	1.648	0.953	0.917	0.706	1.082	0.771	1.465	0.898	0.821	0.674	0.952
MTformer	1.354	0.887	0.745	0.677	0.487	0.557	1.241	0.805	0.813	0.703	0.509	0.568	1.286	0.836	0.795	0.692	0.546	0.591	0.791
Scaleformer	0.933	0.741	0.327	0.435	0.264	0.376	0.883	0.693	0.499	0.551	0.378	0.469	0.997	0.744	0.359	0.458	0.322	0.421	0.550
Informer	1.542	0.955	0.876	0.613	0.602	0.350	1.571	0.953	0.989	0.799	0.768	0.700	1.923	1.087	1.294	0.923	1.225	0.892	1.224
UniOcean-Informer	**0.646**	**0.593**	**0.283**	**0.404**	**0.219**	**0.340**	**0.733**	**0.626**	**0.339**	**0.454**	**0.306**	**0.424**	**0.873**	**0.694**	**0.388**	**0.485**	**0.328**	**0.431**	**0.533**
Autoformer	0.776	0.655	0.270	0.368	0.261	0.365	0.872	0.682	0.530	0.569	0.460	0.533	0.884	0.692	0.448	0.519	0.446	0.521	0.585
UniOcean-Autoformer	**0.628**	**0.581**	**0.256**	**0.377**	**0.200**	**0.323**	**0.790**	**0.661**	**0.385**	**0.495**	**0.313**	**0.424**	**0.864**	**0.685**	**0.458**	**0.538**	**0.373**	**0.472**	**0.565**
Fedformer	0.799	0.662	**0.191**	**0.293**	0.238	0.352	0.897	0.698	**0.263**	**0.384**	0.311	0.411	0.947	0.735	**0.342**	**0.448**	0.380	0.468	0.553
UniOcean-Fedformer	**0.608**	**0.571**	0.270	0.394	**0.202**	**0.329**	**0.741**	**0.624**	0.309	0.430	**0.261**	**0.380**	**0.804**	**0.670**	0.361	0.473	**0.283**	**0.399**	**0.498**

module, the number n of layers in the resid module, and the number of hidden dimensions d in multi-factor correlation learning module. We repeat each experiment 5 times and report the average of MAE on the test set. In each experiment, we only change one parameter while fixing all other parameters.

Scale factor. One key parameter in the UniOcean method is the scale factor s for pattern fusion, which learns the patterns of multi-scale data. Fig. 9 shows the results of different scale factors. With the scale factor $s = 2$, UniOcean-Informer obtains the best performance. The scale factors $s = 3$ and $s = 4$ may cause the loss of detailed information thus can not effectively capture the fine-grained patterns in the data, leading to low performance.

The number of hierarchical layers. Another crucial parameter is the number of layers n in the hierarchical residual network, which fuses the patterns of multi-scale data. As shown in Fig. 10, increasing the number of layers from 3 to 4 can improve the representation ability and reduce MSE. However, the number of 5 layers causes the over-fitting problem, significantly degrading performance. Therefore we set 4 layers in the hierarchical residual network.

The number of hidden dimensions. an essential parameter in the UniOcean method is the number of hidden dimensions $\mathbf{d}$ in the point-wise convolution layer to capture the correlation of different ocean factors. The number of hidden dimensions is determined by the hidden dimension d. Fig. 11 shows the results of UniOcean-Informer while setting different hidden dimension ratios. Our method obtains the best performance with the d number of 8. But for a bigger ratio, it may cause an over-fitting problem.

E. Experimental Results

Three experiments are conducted to examine UniOcean's performance and generalization ability. First, we evaluate our framework on the coarse-grained ocean factor prediction (30 years) datasets collected from 1993 to 2022, as shown in Table II. Second, we compare the performance of our framework by feeding it with different combinations of ocean factors to assess the method's generalization ability, as shown in Table III. Third, we utilize a more recent fine-grained prediction (10 years) dataset collected from 2011-2022 to evaluate the performance of our framework, as shown in Table IV. The results demonstrate that our models achieve superior performance among all the state-of-the-art methods.

Firstly, Table II shows the performance results of adding our proposed UniOcean framework with Transformer-based models and different baselines on coarse-grained ocean factor prediction (SST prediction, SSS prediction, and OHC prediction). We can observe that:

1) By combining our UniOcean framework with baseline methods, the average performance significantly improved. UniOcean-Fedformer achieved the best average performance on all baselines. Compared with the Informer model, the UniOcean-Informer achieved remarkable improvements ranging from 53% to 73% on all ocean factors in MSE results. Similarly, when comparing the Autoformer models, the UniOcean-Autoformer demonstrated substantial performance improvements. These results highlight that the proposed UniOcean is an universal and effective framework to enhance various models. 2) Using multi-scale pattern fusion, our UniOcean-Informer model achieved better performance than the simple multi-task model MTformer. By incorporating multi-scale patterns into the fusion process, our framework can capture and leverage information from different scales, enhancing performance. 4) Compared to the multi-scale method, Scaleformer, our approach has significantly improved due to the correlation modeling between different ocean factors. By modeling the correlation between ocean factors, our method can effectively capture and utilize the shared information to achieve better performance. 5) Traditional RNN-based models (GRU, ConvLSTM) perform much worse than most Transformer-based methods, indicating the potential of Transformer architecture in predicting ocean factors. 6) as the forecasting sequence length increases, UniOcean-Fedformer demonstrates superior performance compared to Fedformer. While it slightly lags in SST, UniOcean-Fedformer outperforms Fedformer in SSS and OHC. This can be attributed to the task balance issue in multi-task learning. Nonetheless, UniOcean-Fedformer maintains a solid overall performance.

Secondly, Table III demonstrates the result comparison of different combinations of the ocean factors; We investigated the impact of training our methods on various combinations

TABLE III: The result of training our methods on different combinations of ocean factor prediction (three, two, and single factors). The result shows that incorporating more factors help improve overall prediction results more significantly.

Datasets		Trained on three factors		Trained on two factors						Trained on single factors					
		SSS+SST+OHC		SSS+SST		SSS+OHC		SST+OHC		SSS		SST		OHC	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Monthly SSS	32	**0.646**	**0.593**	0.781	0.662	0.835	0.700	-	-	1.542	0.955	-	-	-	-
Weekly SST	32	**0.283**	**0.404**	0.437	0.514	-	-	0.338	0.450	-	-	0.876	0.613	-	-
Monthly OHC	32	**0.219**	**0.340**	-	-	0.294	0.411	0.291	0.396	-	-	-	-	0.602	0.350
Monthly SSS	64	**0.733**	**0.626**	0.779	0.646	0.940	0.721	-	-	1.571	0.953	-	-	-	-
Weekly SST	64	**0.339**	**0.454**	0.357	0.467	-	-	0.438	0.520	-	-	0.989	0.799	-	-
Monthly OHC	64	**0.306**	**0.424**	-	-	0.493	0.546	0.334	0.437	-	-	-	-	0.768	0.700
Monthly SSS	96	**0.873**	**0.694**	0.918	0.728	1.123	0.816	-	-	1.923	1.087	-	-	-	-
Weekly SST	96	**0.388**	**0.485**	0.461	0.538	-	-	0.427	0.511	-	-	1.294	0.923	-	-
Monthly OHC	96	**0.328**	**0.431**	-	-	0.731	0.675	0.355	0.448	-	-	-	-	1.225	0.892

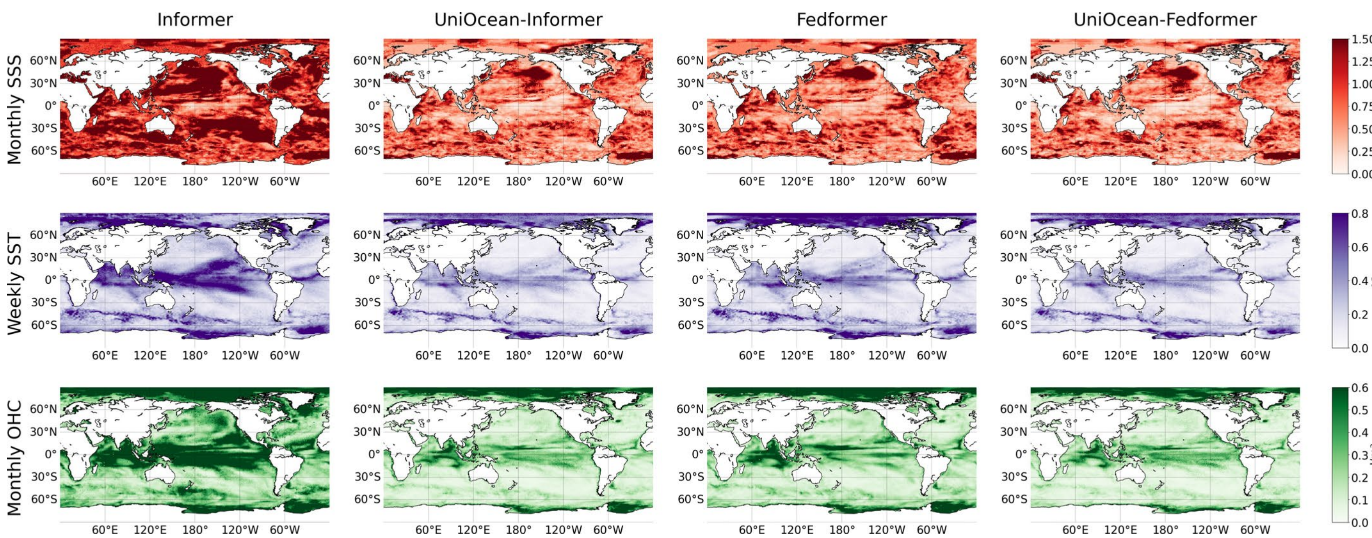

Fig. 12: The comparisons of the overall MSE results between our methods Informer, UniOcean-Informer, Fedformer, and UniOcean-Fedformer on the test sets of three datasets, SST, SSS, and OHC, using experimental data with a prediction sequence length of 32. And the darker the color, the larger the error.

TABLE IV: The comparisons of the result of UniOcean-Informer and Informer on fine-grained ocean factor prediction (Daily SST and D4 SSS), our method can significantly enhance the prediction performance.

Datasets		Informer		UniOcean+Informer		**Improvement**
		MSE	MAE	MSE	MAE	
D4 SSS	32	0.819	0.526	**0.383**	**0.367**	**44.18%**
Daily SST	32	0.227	0.327	**0.095**	**0.202**	**46.31%**
D4 SSS	64	0.792	0.530	**0.520**	**0.278**	**39.64%**
Daily SST	64	0.258	0.368	**0.153**	**0.278**	**31.09%**
D4 SSS	96	0.760	0.516	**0.583**	**0.482**	**16.50%**
Daily SST	96	0.257	0.373	**0.171**	**0.296**	**25.91%**

of multiple ocean factors, including three ocean factors, two ocean factors, and a single ocean factor. We can see that the performance improves with an increasing number of combined ocean factors. When the UniOcean-Informer model is trained on multiple ocean factors simultaneously, its performance improves compared to training on a single ocean factor. The results revealed that incorporating multiple ocean factors leads to a more substantial improvement in overall prediction performance.

Thirdly, as shown in Table IV, we can see that our methods continue to demonstrate superior performance on fine-grained ocean factor prediction, with 40% improvement on MSE and 28% improvement on MAE averagely.

In a nutshell, our framework achieves promising prediction performance by fully fusing the multi-scale data and capturing the correlations between different ocean factors.

F. Visualization Comparison.

To give a clear view of the advantages of our proposed methods, we provide the visualization comparison from two perspectives, i.e., overall comparison and location comparison.

Firstly, we compare the overall MSE performance of our method and baseline methods, Informer and Fedformer, and

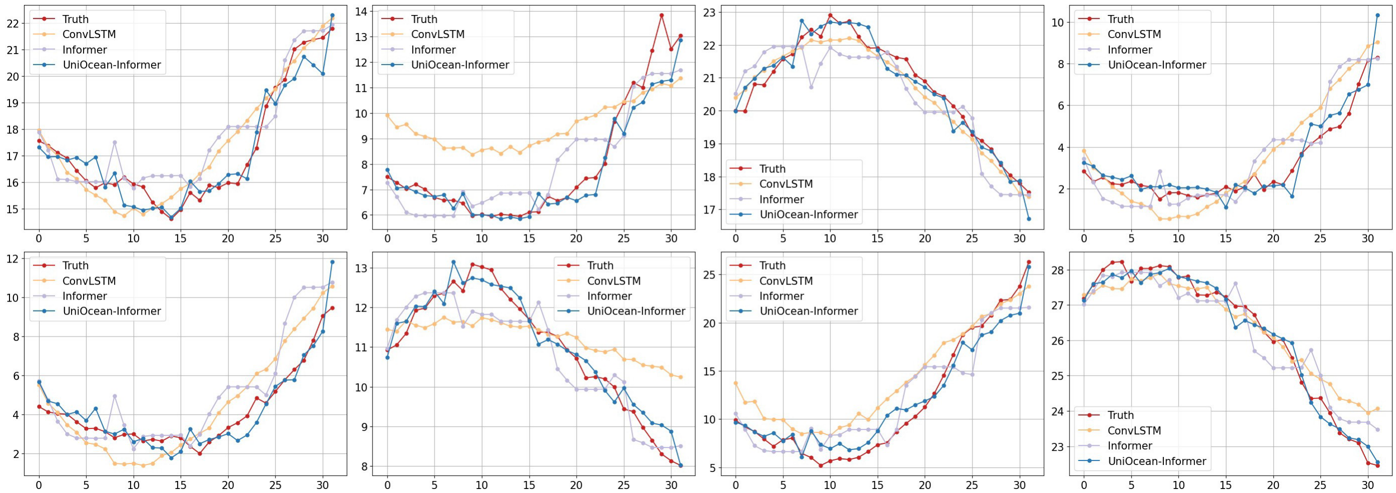

Fig. 13: The prediction visualization of Convlstm, Informer, and UniOcean-Informer of 8 selected locations on the Weekly SST dataset. The ground truth is in red. Our method (Blue line) significantly outperforms others.

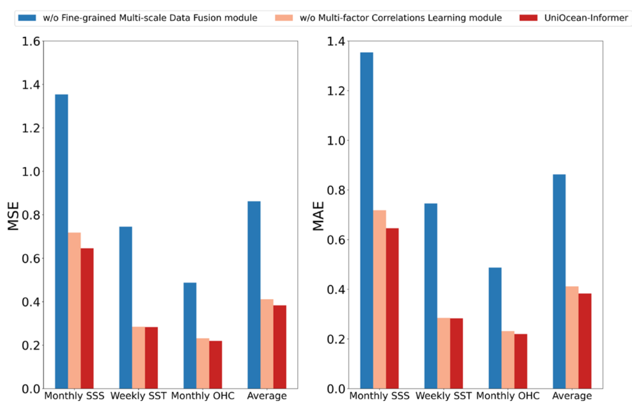

Fig. 14: The ablation study on different variants of UniOcean on MSE and MAE metrics.

corresponding UniOcean-Informer and UniOcean-Fedformer. We visualize the MSE heatmaps in the spatial regions on the test sets with the prediction length 32. As shown in Fig. 12, each subplot independently plots the MSE for the global ocean. The lighter color indicates a small absolute error, whereas the deeper color means a large error. According to the visualization, UniOcean helps the baseline models achieve significant improvements compared to the original model.

Secondly, we provide clear temporal comparisons of results in 8 specific locations. As shown in Fig. 13, each sub-plot visualizes the SST value of one single location in the test set of the weekly SST dataset. Our UniOcean-Informer model (blue line) shows the smallest gap with the ground-truth values (red line), and the errors of ConvLSTM and Informer are considerably significant. Compared with baseline models, the predicted results of our method are most similar to the ground truth values.

In a word, all the results indicate that our model achieves superior performance on all ocean factors.

G. *Ablation Study*

To evaluate the effectiveness of the key components in our method, we conduct a detailed ablation study on the UniOcean framework with Informer by removing different components to get different variants, i.e.,

- **w/o Fine-grained Multi-scale Data Fusion module**: removing the Hierarchical residual network designed for fusing the patterns of multi-scale data.
- **w/o Multi-factor Correlation Learning module**: deleting the point-wise convolution layers, which captures the dynamic correlations for all factors.

As shown in Fig. 14, w/o Fine-grained Multi-scale data Fusion module, the model has the worst performance among all variants, which indicates that multi-scale data fusion improves the performance significantly by utilizing the hierarchical structures to fuse the spatial-temporal dependencies at different scale levels. As for w/o Multi-factor Correlation Learning module, we can observe that this mechanism largely improves average performance, especially in SSS and OHC prediction, with a slight decrease in SST prediction performance. This observation aligns with the common trade-off encountered in MTL models. The MTL model often involves a trade-off between individual ocean factor prediction performance and overall prediction performance across all ocean factors. Unsurprisingly, the model may sacrifice some task-specific performance to achieve the best overall performance.

In summary, both multi-scale data fusion and correlation learning modules are essential for our method to get better performance.

VII. Conclusion

In this study, we introduce a unified ocean factors prediction framework (UniOcean) that leverages multi-scale remote sensing data to simultaneously predict multiple ocean factors. Our approach incorporates the Fine-grained Multi-scale data Fusion module, which effectively captures patterns at different scales and fuses these patterns through hierarchical structures. Additionally, we employ the Multi-factor Correlations Learning module, utilizing point-wise convolution, to adaptively learn the correlations between different factors. Finally, we utilize the factor-specific towers module to predict ocean factors simultaneously. Experimental results on five real-world remote-sensing datasets validate the superior prediction performance achieved by UniOcean, highlighting its potential for enhancing predictions across all ocean factors. Our study draws the following conclusions:

1) Our work UniOcean explores a new perspective on modeling the multiple ocean factors with different temporal scales and provides good performance.
2) Multi-factor learning is an effective way to enhance prediction performance in remote sensing data. By capturing the correlation between different factors, it can be found that overall prediction performance is effectively improved.
3) Fine-grained multi-scale data fusion is highly important to capture the temporal patterns in all ocean factors, when encountering the data with different scales.
4) UniOcean can be easily implemented into other various transform-based methods to enhance their prediction performance without modifying their structures.

In the future, we will further consider and align the different spatial scales across multiple ocean factors to enhance the prediction performance.

References

[1] K. Bi, L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, "Accurate medium-range global weather forecasting with 3d neural networks," *Nature*, vol. 619, no. 7970, pp. 533–538, 2023.

[2] H. Yang, W. Li, S. Wang, H. Li, J. Guan, S. Zhou, and J. Cao, "Spatial-temporal data mining for ocean science: Data, methodologies, and opportunities," *arXiv preprint arXiv:2307.10803*, 2023.

[3] C. Justice, J. Townshend, E. Vermote, E. Masuoka, R. Wolfe, N. Saleous, D. Roy, and J. Morisette, "An overview of modis land data processing and product status," *Remote sensing of Environment*, vol. 83, no. 1-2, pp. 3–15, 2002.

[4] X. Tian, L. Gao, J. Li, L. Chen, J. Ren, and C. Li, "Retrieval of atmospheric aerosol optical depth from avhrr over land with global coverage using machine learning method," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2021.

[5] A. K. Heidinger, C. Cao, and J. T. Sullivan, "Using moderate resolution imaging spectrometer (modis) to calibrate advanced very high resolution radiometer reflectance channels," *Journal of Geophysical Research: Atmospheres*, vol. 107, no. D23, pp. AAC 11–1–AAC 11–10, 2002.

[6] H. Yang, W. Li, S. Hou, J. Guan, and S. Zhou, "Higrn: A hierarchical graph recurrent network for global sea surface temperature prediction," *ACM Transactions on Intelligent Systems and Technology*, vol. 14, no. 4, pp. 1–19, 2023.

[7] J. P. Eveson, A. J. Hobday, J. R. Hartog, C. M. Spillman, and K. M. Rough, "Seasonal forecasting of tuna habitat in the great australian bight," *Fisheries Research*, vol. 170, pp. 39–49, 2015.

[8] A. Kuwano-Yoshida and S. Minobe, "Storm-track response to sst fronts in the northwestern pacific region in an agcm," *Journal of Climate*, vol. 30, no. 3, pp. 1081–1102, feb 2017.

[9] K.-H. Han and S. Hong, "Ocean salinity retrieval and prediction for soil moisture active passive satellite using data-to-data translation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023.

[10] Y. Ren, X. Li, and W. Zhang, "A data-driven deep learning model for weekly sea ice concentration prediction of the pan-arctic during the melting season," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–19, 2022.

[11] S.-J. Lin, W. Putman, and L. Harris, "THE GFDL FINITE-VOLUME CUBED-SPHERE DYNAMICAL CORE," p. 58.

[12] A. Storto, S. Masina, and A. Navarra, "Evaluation of the CMCC eddy-permitting global ocean physical reanalysis system (C-GLORS, 1982–2012) and its assimilation components," *Quarterly Journal of the Royal Meteorological Society*, vol. 142, no. 695, pp. 738–758, 2016.

[13] A. Storto and S. Masina, "The CMCC Global Ocean Physical Reanalysis System (C-GLORS) version 3.1: Configuration and basic validation," p. 12.

[14] B. S. Halpern, C. Longo, D. Hardy, K. L. McLeod, J. F. Samhouri, S. K. Katona, K. Kleisner, S. E. Lester, J. O'Leary, M. Ranelletti *et al.*, "An index to assess the health and benefits of the global ocean," *Nature*, vol. 488, no. 7413, pp. 615–620, 2012.

[15] M. Montero, N. Reul, C. de Boyer Monte´gut, J. Vialard, S. Brachet, S. Guimbard, D. Vandemark, and J. Tournadre, "Sss estimates from amsr-e radiometer in the bay of bengal: Algorithm principles and limits," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–21, 2023.

[16] Z. M. Labe and E. A. Barnes, "Predicting slowdowns in decadal climate warming trends with explainable neural networks," *Geophysical Research Letters*, vol. 49, no. 9, p. e2022GL098173, 2022.

[17] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 12, pp. 5586–5609, 2021.

[18] F. Ling, J.-J. Luo, Y. Li, T. Tang, L. Bai, W. Ouyang, and T. Yamagata, "Multi-task machine learning improves multi-seasonal prediction of the indian ocean dipole," *Nature Communications*, vol. 13, no. 1, p. 7681, 2022.

[19] E. Kim, P. Kruse, S. Lama, J. Bourne, M. Hu, S. Ali, Y. Huang, and J. Wang, "Multi-task deep learning based spatiotemporal arctic sea ice forecasting," in *2021 IEEE International Conference on Big Data (Big Data)*, 2021, pp. 1847–1857.

[20] S. Wang, X. Zhang, Y. Qin, W. Song, and B. Li, "Marine target magnetic anomaly detection based on multitask deep transfer learning," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023.

[21] Y. Liu, H. Niu, and Z. Li, "A multi-task learning convolutional neural network for source localization in deep ocean," *The Journal of the Acoustical Society of America*, vol. 148, no. 2, pp. 873–883, 2020.

[22] Y. Liu, Y. Wu, L. Yang, P. Zhou, J. Kuang, W. Yu, J. Wang, Z. Xu, and G. Li, "A multi-task learning for submarine cable magnetic anomaly recognition," *Journal of Marine Science and Engineering*, vol. 11, no. 5, p. 900, 2023.

[23] Q. Zheng, R. Wang, G. Han, W. Li, X. Wang, Q. Shao, X. Wu, L. Cao, G. Zhou, and S. Hu, "A spatio-temporal multiscale deep learning model for subseasonal prediction of arctic sea ice," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[24] I. Cohen, Y. Huang, J. Chen, J. Benesty, J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," *Noise reduction in speech processing*, pp. 1–4, 2009.

[25] E. Zivot and J. Wang, *Vector Autoregressive Models for Multivariate Time Series*. New York, NY: Springer New York, 2003, pp. 369–413.

[26] G. Zhang, "Time series forecasting using a hybrid arima and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[27] Y. Xue and A. Leetmaa, "Forecasts of tropical pacific sst and sea level using a markov model," *Geophysical Res. Lett.*, pp. 2701–2704, 2000.

[28] S. Aguilar-Martinez and W. W. Hsieh, "Forecasts of Tropical Pacific Sea Surface Temperatures by Neural Networks and Support Vector Regression," *International Journal of Oceanography*, pp. 1–13, 2009.

[29] F. Recknagel, J. Bobbin, P. Whigham, and H. Wilson, "Comparative application of artificial neural networks and genetic algorithms for multivariate time-series modelling of algal blooms in freshwater lakes," *Journal of Hydroinformatics*, vol. 4, no. 2, pp. 125–133, 03 2002.

[30] M. Li, R. Zhang, and K. Liu, "Machine learning incorporated with causal analysis for short-term prediction of sea ice," *Frontiers in Marine Science*, vol. 8, p. 649378, 2021.

[31] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, pp. 1735–1780, 1997.

[32] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1724–1734.

[33] W. Cao, D. Wang, J. Li, H. Zhou, L. Li, and Y. Li, "Brits: Bidirectional recurrent imputation for time series," in *Advances in Neural Information Processing Systems*, 2018, pp. 1–11.

[34] G. Lai, W.-C. Chang, Y. Yang, and H. Liu, "Modeling long- and short-term temporal patterns with deep neural networks," in *The 41st International ACM SIGIR Conference on Research amp; Development in Information Retrieval*, 2018, pp. 95–104.

[35] S. S. Rangapuram, M. W. Seeger, J. Gasthaus, L. Stella, Y. Wang, and T. Januschowski, “Deep state space models for time series forecasting,” in *Advances in Neural Information Processing Systems*, 2018, pp. 1–10.

[36] Y. Yang, J. Dong, X. Sun, E. Lima, Q. Mu, and X. Wang, “A CFCC-LSTM Model for Sea Surface Temperature Prediction,” *IEEE Geoscience and Remote Sensing Letters*, no. 2, pp. 207–211, 2018.

[37] L. Wang, K. A. Scott, D. A. Clausi, and Y. Xu, “Ice concentration estimation in the gulf of st. lawrence using fully convolutional neural network,” in *2017 IGARSS*. IEEE, 2017, pp. 4991–4994.

[38] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in neural information processing systems*, vol. 30, 2017, pp. 1–11.

[39] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, “Informer: Beyond efficient transformer for long sequence time-series forecasting,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 12, 2021, pp. 11 106–11 115.

[40] L. Von Rueden, S. Mayer, K. Beckh, B. Georgiev, S. Giesselbach, R. Heese, B. Kirsch, J. Pfrommer, A. Pick, R. Ramamurthy *et al.*, “Informed machine learning–a taxonomy and survey of integrating prior knowledge into learning systems,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 1, pp. 614–633, 2021.

[41] X. Jia, J. Zwart, J. Sadler, A. Appling, S. Oliver, S. Markstrom, J. Willard, S. Xu, M. Steinbach, J. Read *et al.*, “Physics-guided recurrent graph model for predicting flow and temperature in river networks,” in *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)*. SIAM, 2021, pp. 612–620.

[42] L. Qu, S. Liu, M. Wang, and Z. Song, “Transmef: A transformer-based multi-exposure image fusion framework using self-supervised multi-task learning,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, no. 2, 2022, pp. 2126–2134.

[43] Y. Kittenplon, I. Lavi, S. Fogel, Y. Bar, R. Manmatha, and P. Perona, “Towards weakly-supervised text spotting using a multi-task transformer,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 4604–4613.

[44] Y. Fu, W. Ou, Z. Yu, and Y. Lin, “Miga: a unified multi-task generation framework for conversational text-to-sql,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 11, 2023, pp. 12 790–12 798.

[45] Z. Zhang, X. Zhao, H. Miao, C. Zhang, H. Zhao, and J. Zhang, “Autostl: Automated spatio-temporal multi-task learning,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 4, pp. 4902–4910, Jun. 2023.

[46] X. Zhao, W. Fan, H. Liu, and J. Tang, “Multi-type urban crime prediction,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 4, 2022, pp. 4388–4396.

[47] J. Zhang, Y. Zheng, J. Sun, and D. Qi, “Flow prediction in spatio-temporal networks based on multitask deep learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 3, pp. 468–478, 2019.

[48] S. Wang, H. Miao, H. Chen, and Z. Huang, “Multi-task adversarial spatial-temporal networks for crowd flow prediction,” in *Proceedings of the 29th ACM international conference on information & knowledge management*, 2020, pp. 1555–1564.

[49] C. Zhang, F. Zhu, X. Wang, L. Sun, H. Tang, and Y. Lv, “Taxi demand prediction using parallel multi-task learning model,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 794–803, 2020.

[50] J. Han, H. Liu, H. Zhu, H. Xiong, and D. Dou, “Joint air quality and weather prediction based on multi-adversarial spatiotemporal networks,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 5, 2021, pp. 4081–4089.

[51] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.

[52] M. A. Shabani, A. H. Abdi, L. Meng, and T. Sylvain, “Scaleformer: Iterative multi-scale refining transformers for time series forecasting,” in *The Eleventh International Conference on Learning Representations*, 2022, pp. 1–13.

[53] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo, “Convolutional lstm network: A machine learning approach for precipitation nowcasting,” in *Proceedings of the 28th International Conference on Neural Information Processing Systems*, 2015, pp. 802–810.

[54] H. Wu, J. Xu, J. Wang, and M. Long, “Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting,” *Advances in neural information processing systems*, vol. 34, pp. 22 419–22 430, 2021.

[55] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, “Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting,” in *International conference on machine learning*. PMLR, 2022, pp. 27 268–27 286.

VIII. Biography Section

Hanchen Yang received the Bachelor’s degree from Beijing jiaotong University, Beijing, China, in 2020. He is working toward a Ph.D. at the Department of Computer Science and Technology, Tongji University. He is a dual Ph.D. student in the Department of Computing at the Hong Kong Polytechnic University, Hong Kong, China. His research interests include spatial-temporal data mining, graph neural networks, and deep learning.

Yuwei Chen Yuwei Chen currently works for bachelor’s degree at Tongji University, Shanghai, China, majoring in Data Science. She will work towards the master’s degree in computer science with the Department of Computer Science and Technology, Tongji University, Shanghai, China, in 2024. Her research interests include Spatio-temporal data analysis and deep learning.

Jiannong Cao (M’93-SM’05-F’15) received the B.Sc. degree in Computer Science from Nanjing University, China, in 1982, and the M.Sc. and Ph.D. degrees in Computer Science from Washington State University, USA, in 1986 and 1990, respectively. He is currently the Otto Poon Charitable Foundation Professor in Data Science and the Chair Professor of Distributed and Mobile Computing in the Department of Computing at The Hong Kong Polytechnic University (PolyU), Hong Kong. He is also the Dean of Graduate School, the director of Research Institute for Artificial Intelligence of Things (RIAIoT) in PolyU, the director of the Internet and Mobile Computing Lab (IMCL). He was the founding director and now the associate director of PolyU’s University’s Research Facility in Big Data Analytics (UBDA). He served the department head from 2011 to 2017. Prof. Cao is a member of Academia Europaea, a fellow of IEEE, and an ACM distinguished member. He is also a distinguished member of China Computer Federation (CCF). His current research interests include big data analytics, edge intelligence, and mobile computing.

Wengen Li received the B.Eng. degree and Ph.D. degree in Computer Science from Tongji University, Shanghai, China, in 2011 and 2017, respectively. In addition, he received a dual Ph.D. degree in Computer Science from the Hong Kong Polytechnic University in 2018. He is currently an Associate Professor of the Department of Computer Science and Technology at Tongji University. His research interests include multi-modal data analytics, and spatio-temporal data analytics for urban computing and ocean computing. He is a member of China Computer Federation (CCF), a member of IEEE, and a member of ACM.

Yu Yang is currently a Research Assistant Professor with the Department of Computing, The Hong Kong Polytechnic University. He received the M.Eng. degree in Pattern Recognition and Intelligence System from Shenzhen University in 2015 and the Ph.D. degree in Computer Science from The Hong Kong Polytechnic University in 2021. His research interests include spatiotemporal data analysis, representation learning, urban computing, and learning analytics.

Shuyu Wang received the master's degree in Computer Science and Technology from China University of Petroleum, Shandong, China, in 2021. She is currently working toward the Ph.D. degree at the Department of Computer Science and Technology, Tongji University. Her research interests include remote sensing data processing, spatial-temporal data imputation, and deep learning.

Jihong Guan received the bachelor's degree from Huazhong Normal University in 1991, the master's degree from Wuhan Technical University of Surveying and Mapping (merged into Wuhan University since 2000) in 1991, and the PhD degree from Wuhan University in 2002. She is currently a professor in the Department of Computer Science and Technology, Tongji University, Shanghai, China. Before joining Tongji University, she served in the Department of Computer, Wuhan Technical University of Surveying and Mapping from 1991 to 1997, as an assistant professor and an associate professor (since August 2000), respectively. She was an associate professor (2000-2003) and a professor (Since 2003) in the School of Computer, Wuhan University. Her research interests include databases, data mining, distributed computing, bioinformatics, and geographic information systems (GIS).

Rufu Qin received the bachelor's degree in applied geophysice from Jilin University, Changchun, China, in 2001, and the Ph.D. degree in geoexploration and information technology from Jilin University, Changchun, China, in 2006. He is currently an Associate Professor with the School of Ocean and Earth Science, Tongji University, Shanghai, China. His research interests include geographic information systems, remote sensing, and data mining.

Shuigeng Zhou is a professor of School of Computer Science, Fudan University, Shanghai, China. He received his Bachelor degree from Huazhong University of Science and Technology (HUST) in 1988, his Master degree from University of Electronic Science and Technology of China (UESTC) in 1991, and his PhD of Computer Science from Fudan University in 2000. He served in Shanghai Academy of Spaceflight Technology from 1991 to 1997, as an engineer and a senior engineer (since August 1995) respectively. He was a post-doctoral researcher in State Key Lab of Software Engineering, Wuhan University from 2000 to 2002. His research interests include big data management and analysis, artificial intelligence, and bioinformatics. He has published more than 200 papers in domestic and international journals (including ACM TITS, IEEE TKDE, IEEE TPDS, VLDB Journal, IEEE TCBB, Bioinformatics *etc.*) and conferences (including SIGMOD, VLDB, ICDE, SIGKDD, SIGIR, AAAI, IJCAI, ICCV, CVPR, SODA, ISMB and RECOMB *etc.*). Currently, he is a senior member of IEEE and a member of ACM.