Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2023.29646

297 (2023) 138–159
June

Remote sensing image analysis and cyanobacterial bloom prediction method based on ACL3D-Pix2Pix

Li Wang[a], Qianhui Tang[a], Wenhao Li[a], Xiaoyi Wang[b,*], Haiyan Zhang[a,*], Jiping Xu[a], Zhiyao Zhao[a], Jiabin Yu[a], Huiyan Zhang[a], Qian Sun[a], Yuting Bai[a]

[a]*Beijing Laboratory for Intelligent Environmental Protection, School of Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China, emails: haiyanzhang@btbu.edu.cn (H.y. Zhang), wangli@th.btbu.edu.cn (L. Wang), 3499789140@qq.com (Q.h. Tang), 1437192518@qq.com (W.h. Li), xujiping@139.com (J.p. Xu), zzy_btbu@163.com (Z.y. Zhao), yujiabin_1984@hotmail.com (J.b. Yu), zhanghuiyan369@126.com (H.y. Zhang), sungian0403@hotmail.com (Q. Sun), byting@foxmail.com (Y.t. Bai)*
[b]*Beijing Institute of Fashion Technology, Beijing 100029, China, email: sdwangxy@163.com (X.y. Wang)*

Received 24 September 2022; Accepted 22 April 2023

ABSTRACT

Currently, the prediction of cyanobacterial blooms in rivers and lakes is mainly based on data obtained from on-site water quality monitoring, which cannot analyze the overall water area. Remote sensing images can reflect spatiotemporal and spatial information of the overall water area. However, the existing methods cannot effectively predict the spatial and temporal distribution of overall water bloom by pixel-level prediction of remote sensing images, and the raw remote sensing images often have many problems that cannot be directly used for modeling. Therefore, the method of remote sensing image time series pre-processing is firstly proposed in this paper. Secondly, attentional convolution long short-term memory network embedding to 3D U-Net's Pix2Pix (ACL3D-Pix2Pix) is proposed in this paper to achieve pixel-level prediction of remote sensing images. The model generator is a convolutional long short-term memory network (ConvLSTM) embedded into the 3D U-Net network. The attention mechanism is added to the ConvLSTM, the residual structure is added to the 3D U-Net network, and then the loss function of the overall model is updated to achieve pixel-level prediction of remote sensing images. On this basis, the spatial and temporal distribution prediction of cyanobacterial blooms based on remote sensing image analysis is realized by adjusting the existing eutrophication grading criteria of cyanobacterial bloom water bodies. Finally, the experimental results show that the method is effective in predicting cyanobacterial blooms.

Keywords: Remote sensing image; Cyanobacterial bloom; Water eutrophication; Prediction; Pix2Pix model

1. Introduction

With the development of technology and society, the phenomenon of water eutrophication caused by water pollution is becoming more and more serious and has become a global environmental problem. Cyanobacterial bloom is a kind of water pollution phenomenon in the eutrophic state of water body, where algae proliferate and gather and make the water body appear blue-green after reaching a certain concentration [1]. Outbreaks of cyanobacterial blooms caused by eutrophication in water bodies can be extremely harmful to people's daily lives [2,3], not only causing enormous environmental pollution, resulting in water shortages [4], but also affecting local economic

* Corresponding authors.

development, and it has been shown that economic development and environmental level are closely related [5,6]. As cyanobacterial blooms are unexpected events, it is difficult to control and cost a great deal of human and material resources once they arise. As a result, if it is possible to predict the occurrence of cyanobacterial blooms in advance, the relevant departments can take appropriate emergency measures and make early decisions to achieve double the results with half the effort as well as prevent harm caused by blooms of cyanobacterial algae.

Currently, there are two main categories of methods for predicting cyanobacterial blooms: mechanism-driven methods and data-driven methods. Mechanism-driven models use differential or partial differential equations to describe the interactions of influencing factors, considering physical, chemical, and biological processes in aquatic ecosystems [7–10]. In spite of this, the mechanism-driven model does not generalize well due to the different algal growth processes in different water bodies.

Based on the type of data used for modeling, data-driven prediction models can be classified into prediction models based on sensor numerical data and prediction models based on remote sensing image data. Sensor numerical data is obtained by setting a certain number of underwater monitoring points evenly in a certain range of water, taking samples by hand, and bringing them back to the laboratory for testing. It is difficult to make an accurate judgment about the cyanobacteria bloom situation of the entire lake due to the cumbersome nature of this monitoring procedure, which is easily affected by weather conditions. Remote sensing image data is obtained through satellite images, which have the characteristics of strong monitoring timeliness and a wide monitoring range. As hyperspectral data, remote sensing images provide rich spatial information, as well as multidimensional, correlation, non-linearity, and large data volumes. Therefore, deep learning can be used to extract deep-seated space-time information from remote sensing images and better predict upcoming remote sensing images.

The prediction models based on sensor numerical data are mainly divided into 2 categories: mathematical-statistical methods and artificial intelligence methods. Mathematical statistical models are more generalizable than mechanistic models but do not apply to systems with significant nonlinearity; artificial intelligence models include both traditional machine learning and deep learning methods. Traditional machine learning displays strong self-learning and self-adaptive ability when dealing with nonlinear problems such as cyanobacterial blooms [11–13], and is well suited to complex nonlinear systems, however, it requires the use of a manual feature set and is not suitable for learning large quantities of data [14,15]; while deep learning is a type of representation learning, it is capable of learning a higher level of abstract representation of data and automatically extracting deep features from it [16], and the model's capability will increase exponentially with the level of abstraction [17]. There are several deep learning networks commonly used, including convolutional neural networks [18], recurrent neural networks [19], graph neural networks [20], and their improved and combined network models [21,22], however, the deep learning methods based on sensor data do not take into account the spatial and temporal characteristics of water bodies in an integrated manner.

Prediction models based on remote sensing image data are mostly deep learning-based methods because they involve image processing, and can be divided into 2 categories, feature-level prediction, and pixel-level prediction, according to the output types of the models. Feature-level prediction refers to extracting spatial and temporal features from remote sensing images and then performing numerical prediction based on the features. Relevant studies include image prediction methods based on convolutional neural networks (CNN) combined with 2-dimensional Gabor filtering [23]; image time series prediction based on long short-term memory (LSTM) [24]; prediction based on 2DCNN networks superimposed on channels and 3DCNN networks based on remote sensing image prediction methods [25]. While feature-level predictions utilize remote sensing images as input and water body characteristics as output, the prediction results are still numerical data, which cannot reflect the spatial and temporal distribution of cyanobacteria blooms. Pixel-level prediction refers to the prediction of the pixels of the remote sensing image at the future time. Relevant research includes time series prediction of the remote sensing image based on the convolutional long short-term memory network (ConvLSTM) [26], which uses zero padding in the convolution process, as a result, image edge information prediction is not very accurate; remote sensing image prediction based on generative adversarial networks (GAN) and its variants such as Pix2Pix [27–29]. Pixel-level prediction takes remote sensing images as input and output, and the prediction results are still images, which can reflect the future time and space distribution. Nevertheless, gradient explosions and gradient disappearances can occur easily, and it is often difficult to extract both temporal and spatial features accurately at the same time, resulting in poor image prediction.

Various cyanobacterial bloom prediction methods have been analyzed, and it has been concluded that pixel-level prediction methods based on remote sensing image data can better reflect the spatial and temporal distribution of cyanobacterial blooms. As a variant of GAN, the Pix2Pix model [30] can handle the pixel transformation of images in space better than other pixel-level prediction methods and can theoretically achieve Nash equilibrium, allowing more accurate pixel-level prediction. As a result of the generator's use of a U-Net network, the original Pix2Pix model is limited to predicting a single image based on a single image, and the model cannot extract time series features of images, as well as the model, cannot converge easily during training, increasing the training difficulty.

Chlorophyll-a concentration increases significantly when cyanobacterial blooms occur, and therefore chlorophyll-a concentration is the most direct indicator to characterize cyanobacterial blooms. Thus, for the remote sensing image based on chlorophyll-a concentration, the existing remote sensing image prediction methods are difficult to effectively predict the remote sensing image at the pixel level, which makes it impossible to predict the spatial and temporal distribution of cyanobacterial blooms in the whole water area.

On the basis of ACL3D-Pix2Pix, the preprocessing of remote sensing image time series and pixel level prediction

methods are proposed in this paper, and thus the prediction method of the spatial and temporal distribution of cyanobacterial blooms in the whole water area is proposed. Based on the predicted chlorophyll-a concentration, it is possible to determine the cyanobacterial bloom outbreak caused by eutrophication in the water body, and thus the degree of eutrophication in the water body can be judged.

The prediction results of cyanobacterial bloom spatial and temporal distribution allow the future eutrophication degree and distribution of cyanobacterial blooms to be evaluated effectively. Thus, the relevant departments will be able to focus on preventing and controlling areas with a high degree of eutrophication in order to reduce environmental pollution and control costs.

2. Preliminaries information

The overall research route of this paper, the base model for remote sensing image prediction, and the remote sensing image dataset used are presented in this section.

2.1. Research route

An analysis of remote sensing images and a method for predicting cyanobacterial blooms based on the ACL3D-Pix2Pix are presented in this paper. There are three main parts, as shown in Fig. 1.

As can be seen from Fig. 1, the main research routes of this paper are:

1) Study of pre-processing methods for time series of remote sensing images. This paper addresses the problems that existing remote sensing image raw data are not perfect, and that images have a non-uniform scale, missing local information, and unequal sampling time intervals. In pre-processing, data scale unification based on pixel substitution method, remote sensing image repair based on attention mechanism Pix2Pix model and spatial weights, and remote sensing image time series filling based on linear interpolation method are used.
2) Study of remote sensing image prediction method based on ACL3D-Pix2Pix. First, the remote sensing images of chlorophyll-a concentration at four consecutive moments were superimposed as video frames, after which the ConvLSTM network is fused with the 3D U-Net network, a layer of ConvLSTM network is added after each downsampling and upsampling layer in the 3D U-Net network, and the attention mechanism is introduced into the ConvLSTM to fully extract the temporal and spatial features of remote sensing images, and the residual structure is incorporated into the upsampling and downsampling to avoid the phenomenon of overfitting in the network. Finally, the L_1 loss function of the original Pix2Pix model is replaced by the L_2 loss

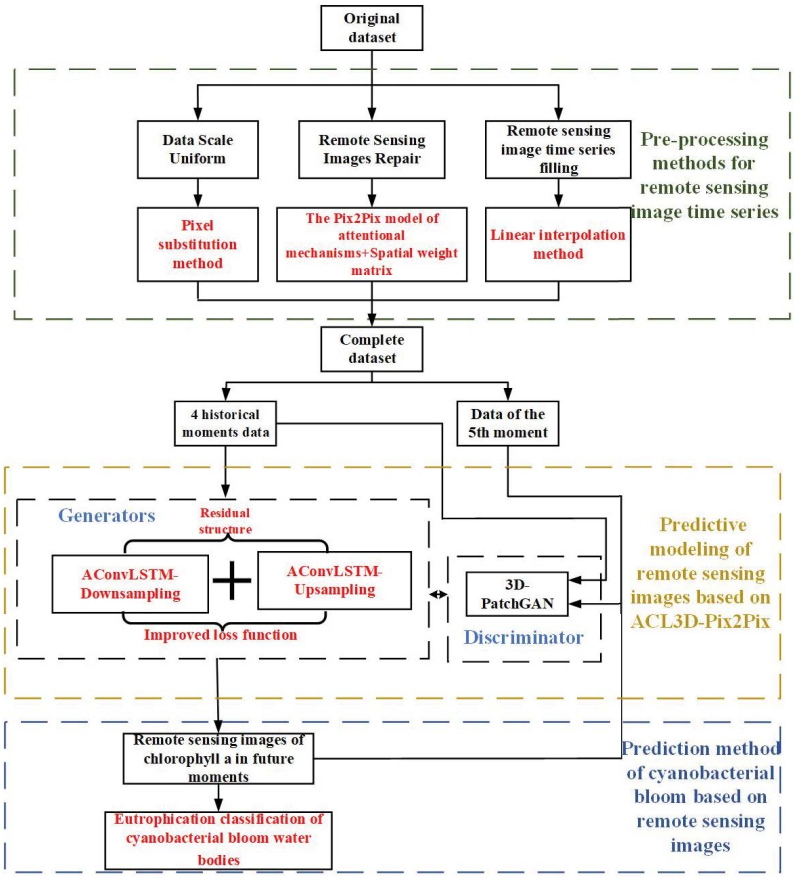


Fig. 1. Research route of remote sensing image analysis and cyanobacteria bloom prediction based on ACL3D-Pix2Pix.

function to speed up the convergence of the network and improve the accuracy of the pixel-level remote sensing image prediction.

3) Study of spatial and temporal distribution prediction method of cyanobacterial blooms based on remote sensing images. To solve the problem that the existing prediction methods of cyanobacterial blooms cannot realize the prediction of spatial-temporal distribution, utilizing the unified data scale of remote sensing images, the existing eutrophication classification standard of cyanobacterial blooms expressed by chlorophyll-a concentration is adjusted in order to predict spatial and temporal distributions of the cyanobacterial bloom eutrophication grade in the whole water area.

2.2. Basic model

During this section, the basic models related to the ACL3D-Pix2Pix presented in this paper are discussed.

2.2.1. Original Pix2Pix model

The original Pix2Pix model is based on conditional general adverse nets (cGAN). The input image is equivalent to the conditions input into cGAN, which is used to guide the generator to generate images. Initially, Pix2Pix was used to convert image styles. Fig. 2 is a structural diagram of the original Pix2Pix model.

As shown in Fig. 2, the original Pix2Pix model is composed of a generator and a discriminator. The generator is composed of a traditional U-Net network. The upsampling process involves the addition of a dropout layer, through which noise is added to diversify the output of the network. The discriminator is composed of PatchGAN, which divides the input image into $N \times N$, and then calculates the probability of each small block. Finally, the average value of these probabilities is taken as the output of the whole. As a result, the amount of computation is reduced and the training speed and convergence speed can be increased. An input image is received by the generator and converted into a predicted image by encoding and decoding, while the discriminator optimizes the parameters of the generator by discriminating between the predicted image and the real image. For the generator, the training process is to constantly "cheat" the discriminator with the generated new data. For the discriminator, continuous learning is needed to prevent being "cheated".

2.2.2. ConvLSTM model

A limitation of the original Pix2Pix model is its inability to extract time series features from images, whereas ConvLSTM is suitable for extracting time series features from images. In ConvLSTM, the full connection part of LSTM is turned into a convolution operation, which preserves the ability of LSTM to extract time series as well as the ability to process image data and extract its spatial-temporal characteristics. So ConvLSTM is applied to video detection [31], gesture recognition [32] and other fields [33].

As shown in Fig. 3, ConvLSTM is comprised of 3 gate control units and 1 central node, which are the input gate, forgetting gate, output gate, and memory cell. The biggest difference from LSTM is that single-layer convolution calculation is performed after the current time input and short-term memory are combined. This difference is the key to extracting spatial structure information. ConvLSTM can be described as [34]:

$$I_t = \sigma\left(W_{\text{xi}} * X_t + W_{\text{hi}} * H_{t-1} + b_i\right) \tag{1}$$

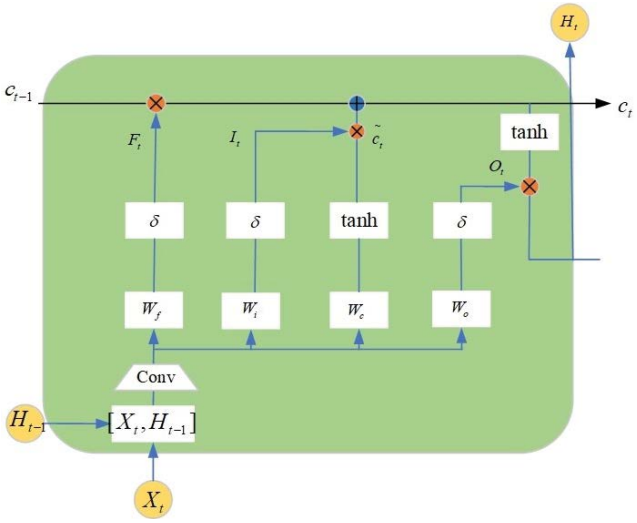


Fig. 3. ConvLSTM structure diagram.

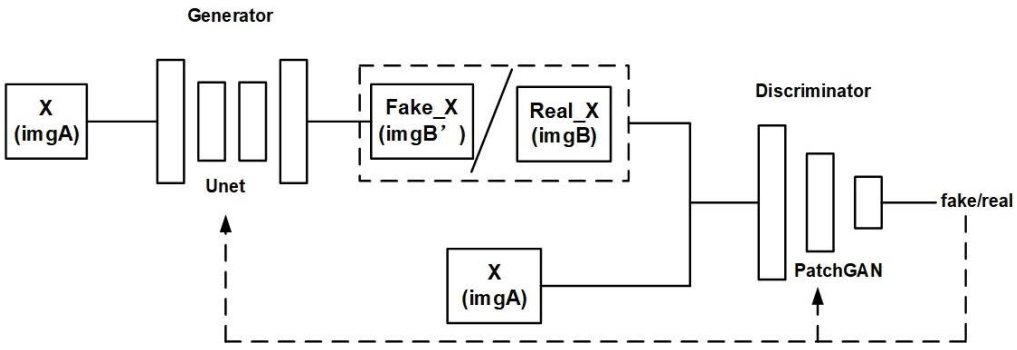


Fig. 2. Structure diagram of original Pix2Pix model.

$$F_t = \sigma\left(W_{\text{xf}} * X_t + W_{\text{hf}} * H_{t-1} + b_f\right) \quad (2)$$

$$\tilde{c}_t = \tanh\left(W_{\text{xc}} * X_t + W_{\text{hc}} * H_{t-1} + b_c\right) \quad (3)$$

$$O_t = \sigma\left(W_{\text{xo}} * X_t + W_{\text{ho}} * H_{t-1} + b_o\right) \quad (4)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \quad (5)$$

$$H_t = O_t \circ \tanh\left(c_t\right) \quad (6)$$

where * denotes convolution operation, ∘ denotes Hadamard product, σ denotes sigmoid function, $W_{x\sim}$ and $W_{h\sim}$ denotes 2-dimensional convolution kernel, X_t denotes input image at the current time, c_t denotes long-term memory at the current time, $\tilde{c}_t$ denotes newly generated information, H_t denotes short-term memory at the current time, I_t denotes input gate output, F_t denotes forgetting gate output and O_t output gate output.

2.2.3. U-Net model

The generator of the original Pix2Pix model is the U-Net network, which was first used to solve the problem of medical image segmentation [35]. Fig. 4 is the structure diagram of the U-Net network.

As shown in Fig. 4, U-Net is a U-shaped network structure to obtain context information and location information. The left side of the U-Net is the downsampling network for feature extraction, and the right side is the upsampling network for feature fusion. In feature extraction, image information will be lost and image resolution will be reduced. In feature fusion, the feature map with a larger size obtained by upsampling through deconvolution lacks edge information. Therefore, edge features can be completed through feature stitching to obtain a complete feature map. It is for these reasons that U-Net networks are widely used in the segmentation of images [36–38] and generation of images [39].

2.3. Remote sensing image data set

The data were obtained from Lake-Watershed Science SubCenter, National Earth System Science Data Center, National Science & Technology Infrastructure of China, which is the 250 m resolution remote sensing image of Taihu Lake region acquired by MODIS satellite, and the raw data of this remote sensing image is in .hdf format. On this basis, based on the relative humidity, wind speed, wind direction, and barometric pressure meteorological data of the same moment in the surrounding meteorological stations of Taihu Lake, the Rayleigh scattering correction of the original data was performed by using python software for the atmosphere. After that, the chlorophyll-a concentration data measured by the water sensor were used to construct an empirical model and a semi-analytical model using python software to invert the atmospheric corrected data to obtain the raster data of the spatial distribution of chlorophyll-a concentration in the format of .img grayscale map. Finally, the raster data is converted into colorful remote sensing image data in .jpg format using ArcGis software, which is the remote sensing image data set used in this study. After data screening, 100 raw data of chlorophyll-a concentration collected from March 9, 2010 to December 29, 2010 were selected as training data, and 28 raw images of chlorophyll-a concentration collected from October 6, 2009 to December 28, 2009 were selected as test data.

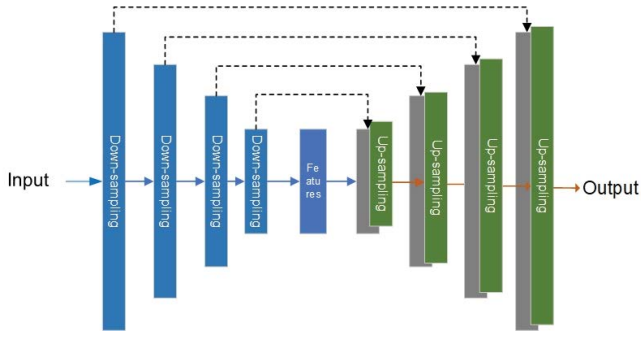


Fig. 4. U-Net model.

3. Method

First of all, the original data of the actual remote sensing image is not perfect due to differences in inversion, weather changes, poor remote sensing communication, and other factors, and as a result, cannot be directly used for prediction modeling. Therefore, a method for preprocessing remote sensing image time series is proposed. Secondly, to realize pixel-level prediction of remote sensing images, a prediction method of remote sensing image time series is proposed by ACL3D-Pix2Pix. Finally, in order to better judge the spatial-temporal distribution of cyanobacterial blooms in the future by adjusting the water eutrophication classification standard, a method of predicting the spatial-temporal distribution of cyanobacterial blooms based on remote sensing images is proposed [40]. All the above methods are implemented in python.

3.1. Preprocessing method of remote sensing image time series

In this paper, corresponding preprocessing methods are proposed to solve the problems of imperfect original time series data of remote sensing images, that is, inconsistent scales, missing local information of images, and unequal sampling time intervals, including:

1) To solve the problem that all image data scales cannot be unified due to the difference of image data scales in the inversion process, the pixel replacement method is adopted to unify the data scales;
2) To solve the problems of missing information and data anomalies caused by weather changes during the sampling of remote sensing images, the Pix2Pix model based on the attention mechanism combined with the spatial weight matrix is used to repair remote sensing images;
3) To solve the problem of missing remote sensing images at some sampling times due to poor remote sensing communication, that is, unequal sampling time

intervals, the linear interpolation method is used to supplement the time series data.

3.1.1. Data scale of remote sensing images unification based on pixel substitution method

In the actual remote sensing image, the scale of each image data is not uniform due to the inversion difference between different images, as shown in Fig. 5.

The red dotted line part in Fig. 5 represents the data scale of remote sensing images. The data scale of each remote sensing image uses the same 9 color grades to represent the 9 concentration ranges of chlorophyll-a, but the concentration ranges represented by the same color grade on different images are not the same. Therefore, it is necessary to develop a unified data scale for all remote sensing images and replace the pixel values of all remote sensing images according to the unified data scale.

First, according to Eqs. (7) and (8), the average concentration range represented by each color level on all remote sensing images is calculated as a unified data scale. After that, calculate the Euclidean distance between each color level in the original remote sensing image, and each color level in the data scale, as shown in Eq. (9), and then each pixel value in the original image is replaced with the pixel value of the color level in the data scale with the smallest Euclidean distance, thus completing the unification of the data scale.

$$\mathrm{LN}_i(\max)=\mathrm{avg}\left(\sum_{n=1}^{k}L_i^n(\max)\right) \tag{7}$$

$$\mathrm{LN}_i(\min)=\mathrm{avg}\left(\sum_{n=1}^{k}L_i^n(\min)\right) \tag{8}$$

$$\left(\mathrm{LN}_i(\max),\mathrm{LN}_i(\min)\right)=\arg\min\left(\sqrt{\left(L(\max)-\mathrm{LN}_i(\max)\right)^2+\left(L(\min)-\mathrm{LN}_i(\min)\right)^2}\right) \tag{9}$$

where, $\mathrm{LN}_i(\max)$, $\mathrm{LN}_i(\max)$ are the maximum and minimum values of chlorophyll-a at grade i after data scale unification, LN_i denotes the concentration of chlorophyll-a at grade i, $L_i^n(\max)$ and $L_i^n(\min)$ are the maximum and minimum values of chlorophyll-a at grade i in the nth sheet in the original data set, k is the total number of samples in the data set, avg denotes the mean value operation, LN(min), LN(max) refers to the maximum and minimum values of each grade on each remote sensing image.

3.1.2. Remote sensing image repair based on attention mechanism Pix2Pix model and spatial weight matrix

During the sampling of remote sensing images, cloud and anomaly data may cover some areas, resulting in damage to some remote sensing images, as shown in Fig. 6. Therefore, it is necessary to fix the missing area data after unifying the data scale.

The gray area in Fig. 6a represents the remote sensing image damage caused by cloud cover, the black area in Fig. 6b represents the remote sensing image damage caused by cloud cover, and the white area in Fig. 6c represents the remote sensing image damage caused by abnormal data. For the remote sensing images with data corruption, the remote sensing images are repaired by using the Pix2Pix model based on the attention mechanism and the spatial weight matrix.

When using attention mechanism Pix2Pix model for remote sensing image repair, firstly, the undamaged remote sensing images in the dataset are selected and the 'damaged remote sensing images' are constructed by manually adding a mask. In order to consider the correlation between images, the 'damaged remote sensing image' and its previous and next undamaged remote sensing images are superimposed on the channel as the input of the Pix2Pix model based on the attention mechanism, and the undamaged remote sensing image before adding the mask is used as the target output. The comparison of remote sensing images before and after adding the mask is shown in Fig. 7.

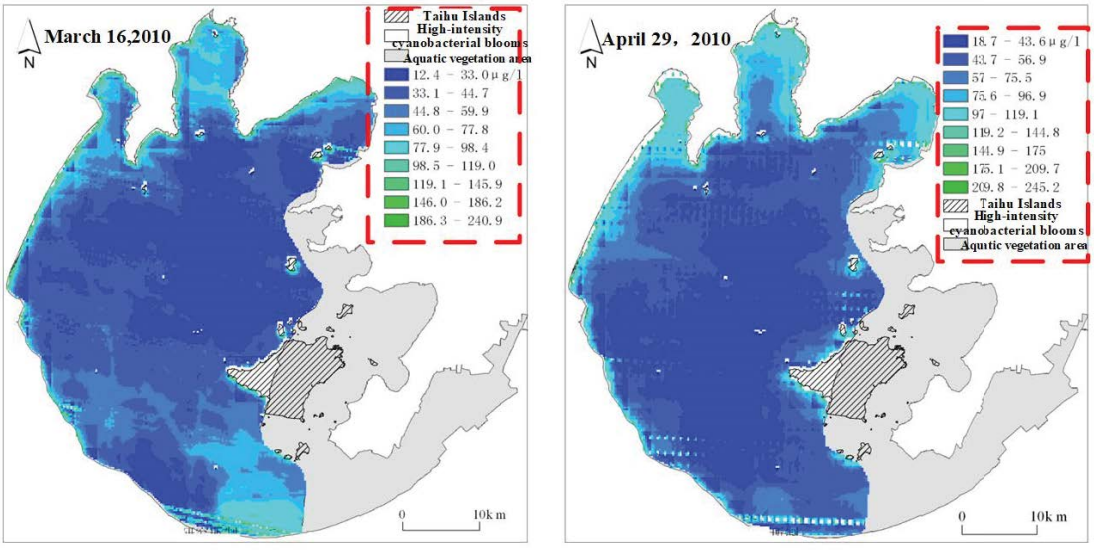


Fig. 5. Comparison chart of remote sensing image data scales.

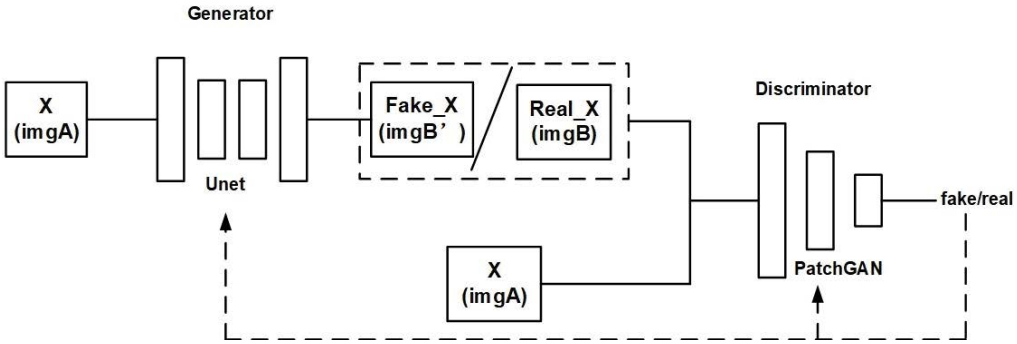


Fig. 6. Schematic diagram of damage in some areas of remote sensing image.

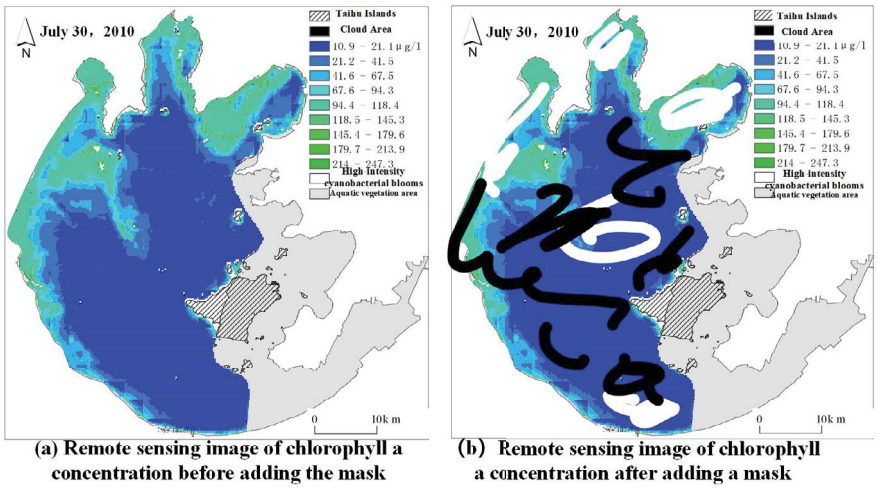


Fig. 7. Comparison of remote sensing images before and after adding masks.

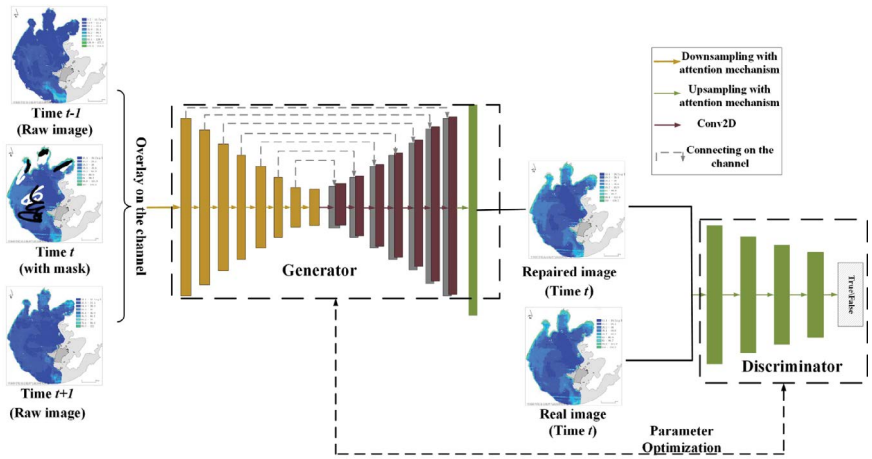


Fig. 8. Pix2Pix model based on attention mechanism.

As shown in Fig. 7 the original remote sensing image is on the left and the remote sensing image after adding the mask is on the right. The structure of the Pix2Pix model based on the attention mechanism is shown in Fig. 8.

The attention mechanism in this model consists of the channel attention mechanism and the spatial attention mechanism, respectively, where the structure of the channel attention mechanism is shown in Fig. 9.

Channel attention can be specifically described as:

$$a_1 = \text{Global Average Pooling}\left(x^{\text{in}}_{i(H,W,C)}\right) \tag{10}$$

$$a_2 = f\left(W_c * a_{1(c,1)}\right) \tag{11}$$

$$f_{\text{channel}} = a_{2(1,1,C)} \circ x^{\text{in}}_{i(H,W,C)} \tag{12}$$

where $x^{\text{in}}_{i(H,W,C)}$ represents the remote sensing image on day i, (H,W,C) represents the data format of the remote sensing image as a 3D data, GlobalAveragePooling represents the global level pooling, a_1 is the result of the global average pooling to obtain a feature of length C, W_c is a weight matrix, $a_{1(C,1)}$ represents the reconstruction of the data of a_1 into a matrix of C×1, the order of the data in a_1 does not change, * represents the convolution, f is the sigmoid activation function, a_2 is the output of a two-dimensional matrix, ∘ is the Hadamard product f_{channel} is the output of the channel attention mechanism, and $a_{2(1,1,C)}$ is the reconstructed data after a_2 reconstruction.

The structure of the spatial attention mechanism is shown in Fig. 10:

The spatial attention mechanism can be described as follows:

$$b_1 = \text{Global Average Pooling}\left(f_{\text{channel}}\right) \tag{13}$$

$$b_2 = \text{Global max Pooling}\left(f_{\text{channel}}\right) \tag{14}$$

$$b_3 = f\left(\left[b_1, b_2\right] * w_c\right) \tag{15}$$

$$f_{\text{sp}} = b_3 \circ f_{\text{channel}} \tag{16}$$

where f_{channel} is the output of the channel attention mechanism, GlobalAveragePooling and GlobalmaxPooling are the global average pooling and global maximum pooling, b_1 and b_2 corresponding to their outputs, f is the sigmoid activation function, w_c is the weight matrix, * denotes convolution, b_3 is the intermediate variable, and f_{sp} is the output of the spatial attention mechanism.

By using the model shown in Fig. 8 for training, the trained model is used for remote sensing image repair. However, the amount of existing remote sensing image data is small, so some areas cannot be repaired based on deep learning repair, and for the areas that cannot be repaired, the spatial weight matrix is used for secondary repair. The spatial weight schematic is shown in Fig. 11, after repairing the missing region contour, the center part is filled using

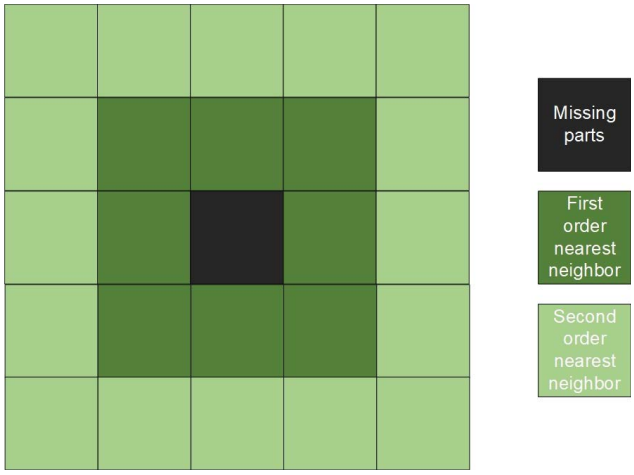


Fig. 11. Spatial weight matrix.

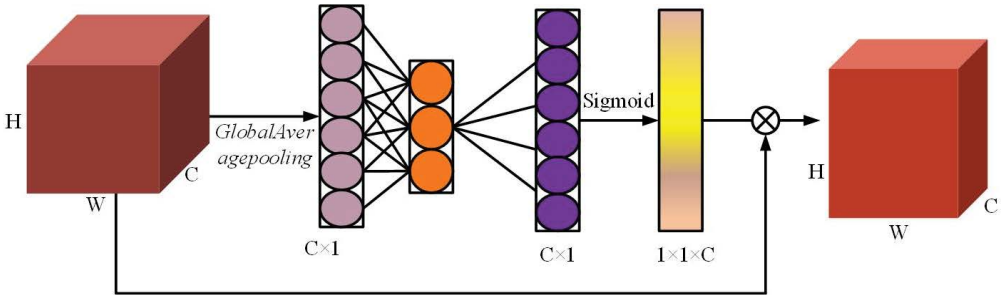


Fig. 9. Channel attention mechanism structure diagram.

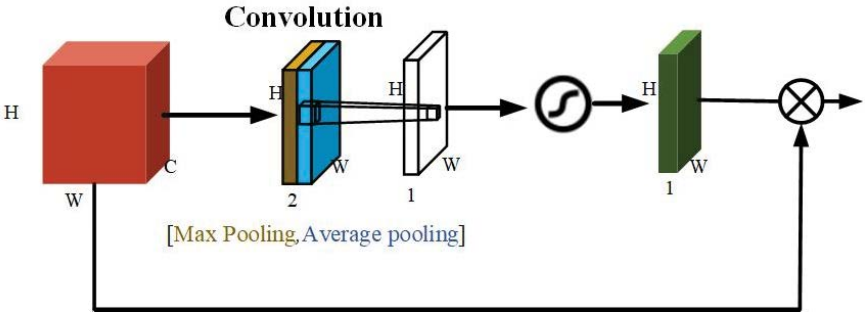


Fig. 10. Structure diagram of spatial attention mechanism.

the nearest neighboring contour parts. The missing part in the center is filled with the contour part of the nearest neighbor.

The weight of the first-order nearest neighbor is the square of the inverse of the Euclidean distance, and the weight of the second-order nearest neighbor is the in-verse of the Euclidean distance, as shown in Eq. (17) for the specific repair method.

$$P(i,j)=\frac{\sum_{a=1}^{n} w_1 \times P_1(i_a,j_a)+\sum_{b=1}^{m} w_2 \times P_2(i_b,j_b)}{n * w_1 + m * w_2} \tag{17}$$

where w_1 is the weight of the first-order nearest neighbor, w_2 is the weight of the second-order nearest neighbor matrix, $P(i,j)$ is the pixel value of the missing site, $P_1(i_a,j_a)$ is the pixel point of the first-order nearest neighbor, $P_2(i_b,j_b)$ is the pixel point of the second-order nearest neighbor, n is the number of first-order neighboring pixels, and m is the number of second-order nearest neighboring pixels.

3.1.3. Time series data filling based on linear interpolation

Based on the repair of a single remote sensing image, the remote sensing images could not be acquired at some moments due to poor remote sensing communication and weather, so the data were sampled at varying intervals in the data set, as shown in Fig. 12.

As shown in Fig. 12, the sampling intervals are 2 d and 1 d, respectively. To supplement the missing images and create a complete dataset with equal sampling intervals, linear interpolation is employed in this paper.

As the original data set has a minimum sampling interval of 1 d, the data set is supplemented according to the sampling interval of 1 d. The formula is as follows:

$$P(m+\lambda)=P(m)+\frac{\lambda(P(n)-P(m))}{(n-m)} \tag{18}$$

where, $P(m + \lambda)$ represents the pixel value of the day $m + \lambda$, λ is the number of days from m,$m < \lambda < n$, m, n are the dates, and $P(m)$, $P(n)$ are the pixel values of the 2 dates already in the original data. The schematic diagram of linear interpolation is shown in Fig. 13.

As can be seen from Fig. 13, the pixel values of any one of the 2 d can be obtained by the pixel values of the 2 d that are not adjacent to each other.

3.2. Remote sensing image prediction method based on ACL3D-Pix2Pix

In order to solve the problems that the existing pixel-level prediction methods are difficult to fully extract time series information from images, the model is prone to gradient explosion and the model training speed is slow, this paper makes the following improvements to the original Pix2Pix model:

1) First, the ConvLSTM network is embedded into the 3D U-Net network for image time series prediction. After each downsampling and upsampling layer in the 3D U-Net network, another layer of ConvLSTM network is added, and on this basis, the attention machine is added to the ConvLSTM to build the attention convolution long short-term memory network (AConvLSTM), so as to improve the ability of the network to extract the image time series features; and the residual structure is added to the upsampling and downsampling of the 3D U-Net, so as to avoid the network overfitting phenomenon.
2) In the discriminator, the original Pix2Pix model inputs the input image into the discriminator as a label, without considering the time correlation. In this paper, the image of the historical moment does the data on the channel to do the overlay as the label, and the time series

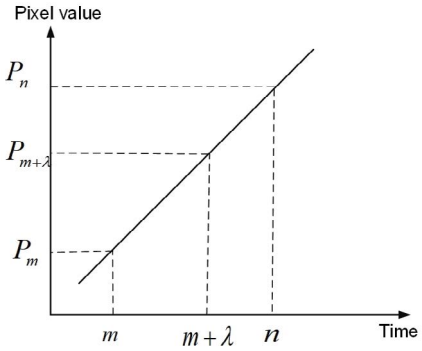
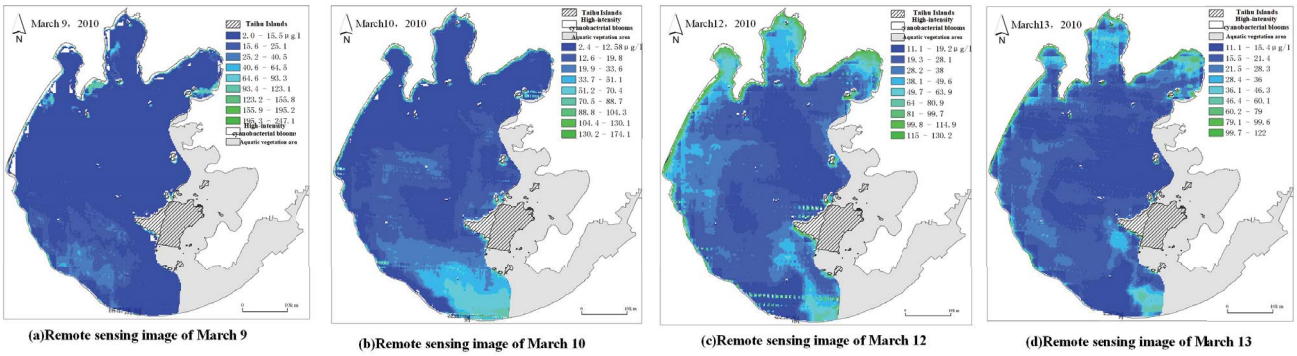


Fig. 13. Schematic diagram of linear interpolation.



(a)Remote sensing image of March 9 (b)Remote sensing image of March 10 (c)Remote sensing image of March 12 (d)Remote sensing image of March 13

Fig. 12. Remote sensing images sampled at varying intervals.

information is incorporated in the criteria to provide the distinction between genuine and fake images.

3) In terms of the loss function, Since the original Pix2Pix model adopts the L_1 loss function, its gradient is fixed and nondifferentiable at the extreme point, resulting in a slow convergence of the network. In this paper, the L_2 loss function is used instead of the L_1 loss function. The gradient of the L_2 loss function decreases as the error decreases, and this contributes to the network's convergence and speeds up the prediction process. Fig. 14 is a structural diagram of ACL3D-Pix2Pix.

As shown in Fig. 14, the remotely sensed images of chlorophyll-a concentration at four consecutive moments are input to the generator, which implements the pixel-level prediction of the remotely sensed image of chlorophyll-a concentration at the fifth moment by embedding the ConvLSTM into the 3D U-Net network. And the remote sensing images of these four moments are superimposed on the channels as the labels for discrimination, and the network performance is optimized by adversarial training of the generator and the discriminator. The reason for using four consecutive moments to predict the fifth moment is to fully consider the change rule of remote sensing image with time and the required hardware resources.

3.2.1. Generator improvements

The generator mainly consists of AConvLSTM, downsampling with residual structure, and upsampling with residual structure.

3.2.1.1. Construction of AConvLSTM

Attention mechanisms can not only enable the network to learn the allocation of attention autonomously but also help the network to fuse various important information. The AConvLSTM model is constructed by adding an attention mechanism before the gate of ConvLSTM, which allows the network to autonomously fuse information in an image as well as improve its ability to extract information from images. The AConvLSTM structure is shown in Fig. 15.

Here, the red dashed box part represents the attention mechanism added in this paper. The schematic diagram of the attention mechanism is shown in Fig. 16.

As shown in Fig. 16, the attention mechanism in this paper is composed of channel attention and spatial attention in series. The channel attention mechanism can model the dependency between each feature map and adaptively adjust the characteristic response value of each channel. The operation process of the channel attention mechanism can be described as follows:

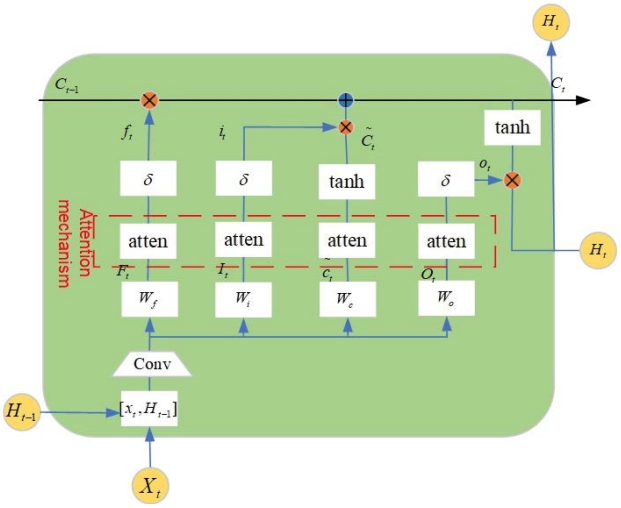


Fig. 15. Structure diagram of attention convolution long short-term memory network.

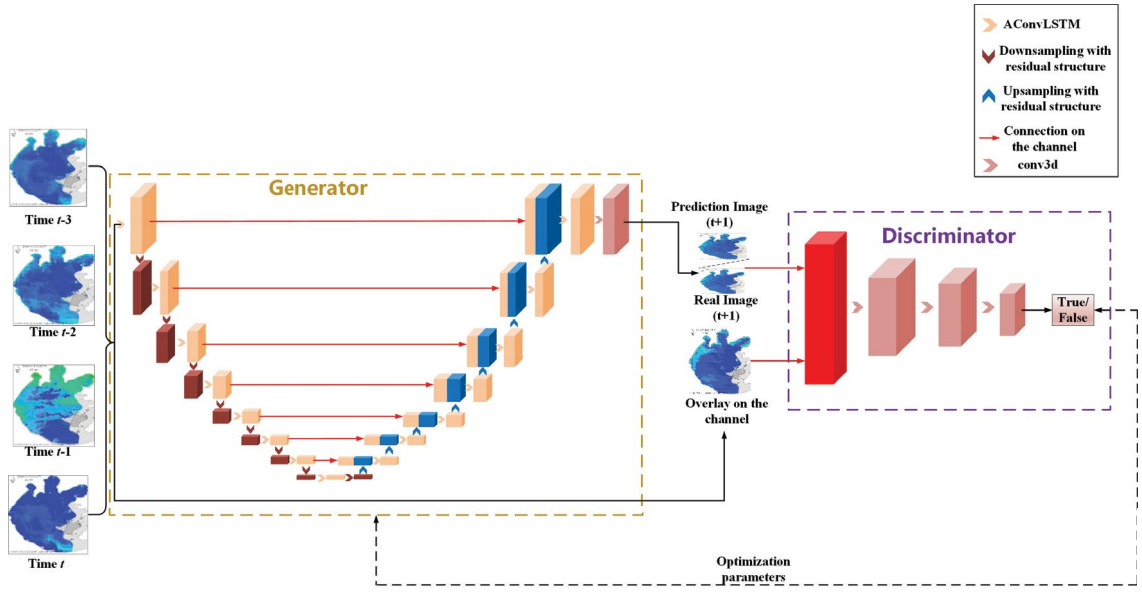


Fig. 14. Structure diagram of ACL3D-Pix2Pix.

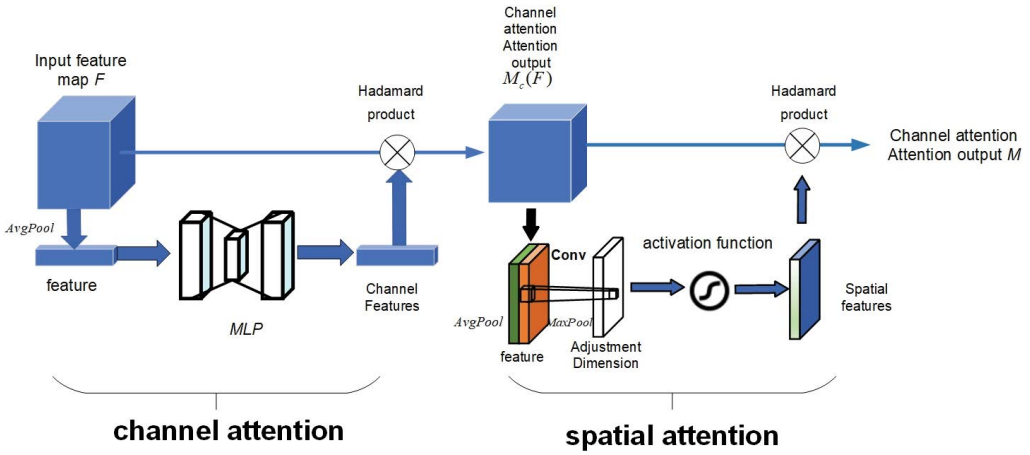


Fig. 16. Schematic diagram of attention mechanism.

$$M_c(F)=\sigma\left\{\mathrm{MLP}\left[\mathrm{Avg\ Pool}(F)\right]\right\}\circ F \tag{19}$$

The input feature map F is globally average pooling to get the features, and the features are input to a multiply-layer perceptron (MLP), and the output of the MLP is multiplied by the activation function to get the weight of each channel, which completes the channel attention construction, where $M_c(F)$ is the feature map output by the channel attention mechanism, MLP is the multiply-layer perceptron, σ is the activation function, o is the Hadamard product, AvgPool is the average pooling, and F is the input feature map.

The spatial attention mechanism can automatically extract the focused areas in the image, which increases the computational consumption but greatly improves the network performance. The operation process of the spatial attention mechanism can be described as follows:

$$M=\sigma\left\{f^{3\times3\times3}\left[\begin{array}{l}\mathrm{Avg\ Pool}\left(M_c(F)\right);\\ \mathrm{Max\ Pool}\left(M_c(F)\right)\end{array}\right]\right\}\circ M_c(F) \tag{20}$$

In which the maximum value and the average value are taken on the channel of each feature point in the output feature map obtained from the channel attention, the 2 results are stacked and the number of channels is adjusted by using the convolution with the number of convolution kernels of 1. Finally, the weight of each feature is obtained by the activation function, and the weight is multiplied with the input to complete the spatial attention construction. M is the final output.

As a result of introducing the above attention mechanism before ConvLSTM enters the activation function of the gating unit, AConvLSTM can be described as follows: the AConvLSTM proposed in this paper first accepts the input image X_t of this moment, and the short-term memory output H_{t-1} of the previous moment, and enters the attention mechanism shown in Fig. 15 after convolving to obtain the input F_t of the forgetting gate attention mechanism and finally obtain the forgetting gate output f_t after the activation function.

$$F_t=W_{\mathrm{xf}}*X_t+W_{\mathrm{hf}}*H_{t-1}+b_f \tag{21}$$

$$f_t=\delta\left\{M\left[M_c(F_t)\right]\right\} \tag{22}$$

The input gate output is similar to the forgetting gate output process, where the input I_t of the input gate attention mechanism and the input gate output i_t can be expressed as:

$$I_t=W_{\mathrm{xi}}*X_t+W_{\mathrm{hi}}*H_{t-1}+b_i \tag{23}$$

$$i_t=\delta\left\{M\left[M_c(I_t)\right]\right\} \tag{24}$$

The attention mechanism input $\tilde{c}_t$ of the nascent information is passed through the attention mechanism after the tanh activation function to obtain the nascent information $\tilde{C}_t$, $\tilde{C}_t$ and the input gate output i_t does Hadamard product after and the previous moment long-term memory C_{t-1} and forgetting gate output f_t do Hadamard product result summed to obtain the long-term memory C_t of this moment, which can be expressed as follows.

$$\tilde{c}_t=W_{\mathrm{xc}}*X_t+W_{\mathrm{hc}}*H_{t-1}+b_c \tag{25}$$

$$\tilde{C}_t=\tanh\left\{M\left(M_c(\tilde{c}_t)\right)\right\} \tag{26}$$

$$C_t=f_t\circ C_{t-1}+i_t\circ\tilde{C}_t \tag{27}$$

The input O_t of the output gate attention mechanism passes through the attention mechanism and then passes through the tanh activation function to obtain the output o_t of the output gate at this moment, and the output o_t of the output gate makes the Hadamard product with the long-term memory C_t at this moment activated by the tanh activation function to obtain the short-term memory output H_t at this moment, which can be specifically expressed as follows:

$$O_t=W_{\mathrm{xo}}*X_t+W_{\mathrm{ho}}*H_{t-1}+b_o \tag{28}$$

$$o_t=\delta\left\{M\left(M_c(O_t)\right)\right\} \tag{29}$$

$$H_t = o_t \circ \tanh C_t \tag{30}$$

where * denotes the convolution operation, ∘ is the Hadamard product, δ denotes the Sigmoid function, $W_{x\sim}$ and $W_{h\sim}$ are 2-dimensional convolution kernels, X_t is the input image at the current moment, I_t, O_t, $\tilde{c}_t$, F_t are the inputs of the attention mechanism, respectively, C_t is the long-term memory at the current moment, $\tilde{C}_t$ is the nascent information, H_t is the short-term memory at the current moment, i_t is the input gate output, f_t is the forgetting gate output and o_t output gate output, all are 3D tensor. The spatial tensor of $W \times H \times C_{in}$, where C_{in} is the number of channels.

3.2.1.2. Construction of downsampling with residual structure

The downsampling with residual structure consists of the conv3d module as well as the residual module, and the structure is shown in Fig. 17.

The red dashed box in Fig. 17 indicates the residual structure added in the downsampling layer. The conv3d module is composed of a 3D convolutional layers, a batch normalization layer, and an activation function layer. The calculation of the conv3d module is shown in Eq. (31):

$$x_{out} = f\left(B\left(x_{in} * \omega + b\right)\right) \tag{31}$$

where x_{in} is the input, ω and b represents the weights and bias values of the convolution kernel, * represents the convolution operation, B represents the batch normalization process, f represents the activation function, and the LeakyReLu function, x_{out} represents the output after the conv3d module.

The residual module consists of 2 conv3d modules and 1 connection layer. In the residual module, the number of convolution kernels of the second conv3d module is twice that of the first conv3d module, and the size of the convolution kernels are 1×1×1 and 3×3×3, respectively, with a step size of 1. The network output size is guaranteed to be the same as the input size after the conv3d module, and the output is superimposed with the input of the residual module after the 2 conv3d modules, The residual module formula is shown below:

$$y_1 = f\left(B\left(x_{out} * \omega_1 + b_1\right)\right) \tag{32}$$

$$y_2 = f\left(B\left(y_1 * \omega_2 + b_2\right)\right) \tag{33}$$

$$y_{out} = x_{out} + y_2 \tag{34}$$

where, y_{out} is the residual module output, x_{out} is the output of the previous conv3d, ω_1 and ω_2 are the weights, b_1 and b_2 are the bias values, and y_1 and y_2 are both intermediate variables.

3.2.1.3. Construction of upsampling with residual structure

Fig. 18 for the structure diagram of upsampling with residual structure.

The red dashed box in Fig. 18 represents the residual structure added in the upsampling layer. The upsampling module is composed of an 3D deconvolution layer, a batch normalization layer, an activation function layer, a dropout layer, and a residuals module. Dropout layers provide diversity to the network and prevent it from overfitting. Because the structure of the residual network is introduced in the process of upsampling and downsampling, the network can generate a clearer image, and avoid the network degradation caused by the deepening of network layers.

3.2.2. Improvement of discriminator

In the discriminator network, 3 conv3d modules comprise PatchGAN. In the original Pix2Pix model, the input image of the generator is used as the output of the label guidance discriminator. The structure of the discriminator in this paper is shown in Fig. 19.

As shown in Fig. 19 the discriminator uses the superimposition of the remote images of 4 historical times input to the generator in the channel dimension as the output of the label guidance discriminator, which not only ensures the consistency of the label but also provides a standard for distinguishing the true and false images on the premise of considering the time characteristics.

3.2.3. Improvement of loss function

The loss function of the original Pix2Pix model is composed of generator loss and discriminator loss.

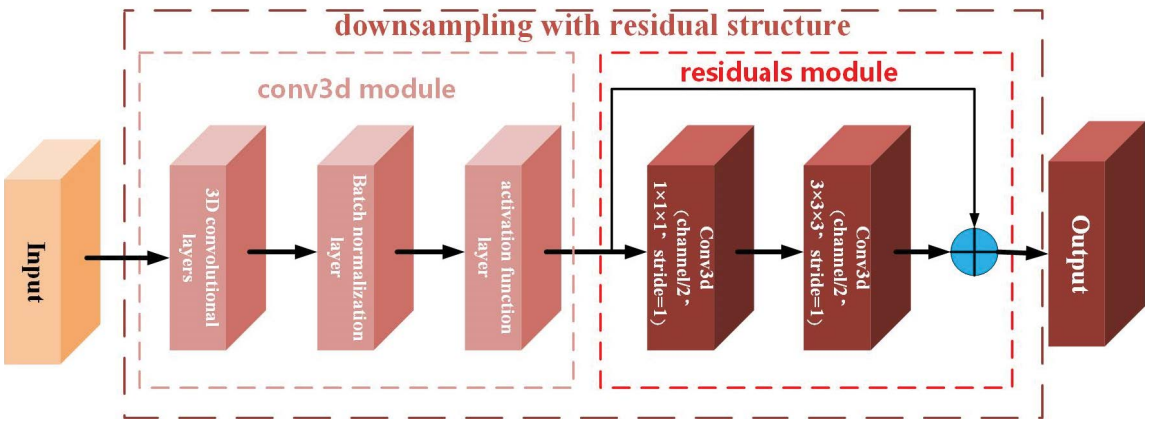


Fig. 17. Structure diagram of 3D U-Net downsampling module with residual structure.

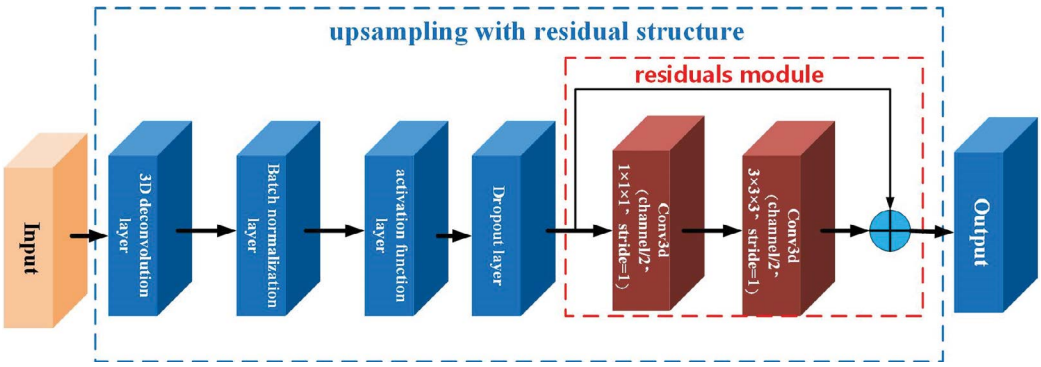


Fig. 18. Structure diagram of 3D U-Net upsampling module with residual structure.

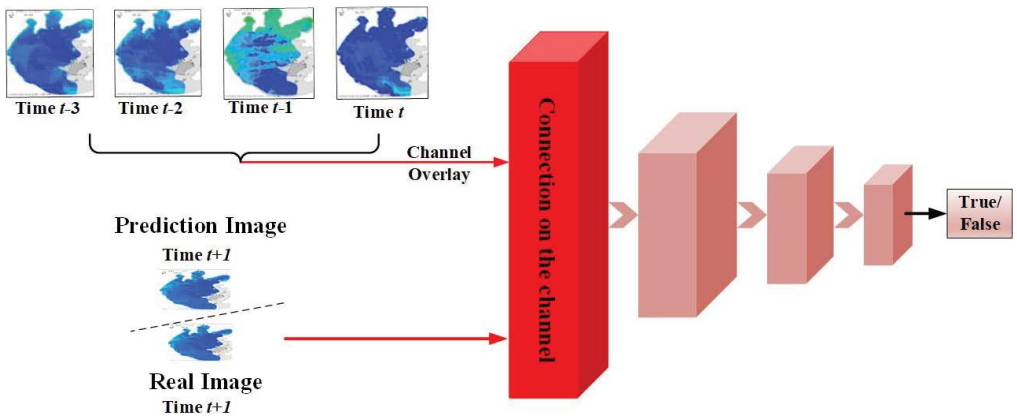


Fig. 19. Schematic diagram of the discriminator.

The discriminator uses binary cross entropy as the loss function. The generator loss is composed of L_1 loss and binary cross entropy function. The difference between the generated image and the real image is constrained by L_1 distance.

The overall loss function of the original Pix2Pix model is [30]:

$$L_{\text{Pix2Pix}} = \arg\min_G \max_D L_{\text{cGAN}}(G,D) + \lambda L_{L1}(G) \tag{35}$$

where:

$$L_{\text{cGAN}} = \arg\min_G \max_D E_{x,y\sim\text{Pdata}(x,y)}\left[\log D(x,y)\right] + E_{x\sim\text{Pdata}}\left[\log\left(1-D\left(x,G(x)\right)\right)\right] \tag{36}$$

$$L_{L1}(G) = E_{x,y\sim\text{Pdata}}\left[\left\|y-G(x)\right\|_1\right] \tag{37}$$

where L_{Pix2Pix} represents the original Pix2Pix overall loss function, G represents the generator, D represents the discriminator, L_{cGAN} represents the cGAN loss function, x represents the input image, y represents the real image, $E_{x,y\sim\text{Pdata}(x,y)}$ represents the expected value after selecting pairs of data from the dataset, $L_{L1(G)}$ is the L_1 loss of the generator. E represents the expectation, λ is a positive integer.

The L_1 loss function has a stable gradient, so the L_1 loss function has better robustness. When there are more outliers in the training data set, the L_1 loss function will be more effective. However, since the gradient is a fixed value when the loss value is small, the loss function will fluctuate around the stable value, and it is difficult to converge to higher accuracy. In the task of image prediction, since the difference between the predicted value and the real value is not large, the L_1 loss function is rarely used. In practical application, L_2 loss is fast and easy to solve. Therefore, the L_1 loss is replaced by L_2 loss in this paper, and the L_2 loss and the overall optimization objective after the change are as follows:

$$L_{L2}(G) = E_{x,y,z}\left[\left\|y-G(x,z)\right\|_2\right] \tag{38}$$

$$G_F = \arg\min_G \max_D \begin{pmatrix} E_{x,y}\left[\log D(x,y)\right] \\ +E_{x,z}\left[\log\left(1-D\left(x,G(x,z)\right)\right)\right] \\ +\lambda L_{L2}(G) \end{pmatrix} \tag{39}$$

where L_{L2} represents the L_2 loss of the generator, G_F represents the final optimization function, E represents the expectation, y represents the real image, $E_{x,y}$ represents the expectation value taken after the operation on the input image and the real image, $E_{x,z}$ represents the expectation value taken after the operation on the input image and the noise, x represents the input to the generator, G represents the generator, D represents the discriminator,

and λ is a positive integer with a value of 100. When training the generative adversarial network, the discriminator is trained first and the generator is trained after fixing the discriminator.

3.3. Prediction of the spatial and temporal distribution of cyanobacterial blooms

To resolve the problem that the existing eutrophication classification standard for cyanobacterial blooms has a too small numerical range, resulting in the inability to cover the numerical range of remote sensing images. After the unified data scale of remote sensing images, the numerical range of the eutrophication classification standard for waterbodies is adjusted and expanded in this study to predict future spatial and temporal distributions of cyanobacterial blooms.

3.3.1. Adjustment of eutrophication classification standard of cyanobacteria bloom

To better evaluate the outbreak of cyanobacteria bloom, this paper first refers to the eutrophication classification standard of water bodies specified by the Organization for Economic Co-operation and Development (OECD), as shown in Table 1.

In Table 1 the nutritional grades of the lake are divided into poor nutrition (<3 μg/L), medium nutrition (3~11 μg/L), eutrophication (11–78 μg/L), heavy eutrophication (≥78 μg/L). However, Taihu Lake is eutrophic throughout the year, and the fluctuations in chlorophyll-a concentrations cannot cover the entire classification range. Therefore, some scholars [25] further subdivided the eutrophic grade in Table 1 into 4 grades eutrophic I, II, III, and IV according to the actual situation, as shown in Table 2.

Table 2 classifies the nutritional grades into 7 grades. However, the grades in Tables 1 and 2 are based on the average chlorophyll-a concentration of the whole lake. However, for the remote sensing image of Taihu Lake, the chlorophyll-a concentration of each pixel ranges from 2.3 to 247 μg/L, and the concentration range of chlorophyll-a in Tables 1 and 2 is too small (3–78 μg/L) to cover the chlorophyll-a concentration range of each pixel in the remote sensing image.

Therefore, it is necessary to expand the concentration range of water eutrophication classification standard according to the chlorophyll-a concentration range of each pixel of remote sensing image. In this paper, the unified data scale of remote sensing images is calculated according to the method in section 3.1.1. According to the concentration range of chlorophyll-a represented by the unified 9 color grades, and referring to the eutrophication classification standards in Tables 1 and 2, the eutrophication grade of the water body is redivided into 9 grades, that is, the poor nutrition grade and the medium nutritional grade are merged into 1 grade, and the eutrophication grade is re divided into 3 grades, The heavy eutrophication grade was reclassified into 5 grades.

The unified data scale and adjusted water eutrophication classification standard are shown in Fig. 20.

It can be seen from Fig. 20 that in the unified data scale, the 9 color grades of the remote sensing image correspond to the 9 concentration ranges of chlorophyll-a, and the adjusted 9 water eutrophication grades also correspond to them.

Table 1
Water eutrophication classification standard (OECD)

Nutritional grade	Whole-lake average chlorophyll-a concentration (μg/L)
Poor nutrition	<3
Medium nutrition	3~11
Eutrophication	11~78
Heavy eutrophication	≥78

3.3.2. Prediction of spatial and temporal distribution of cyanobacterial blooms

In order to determine the nutrient level of each pixel, the remote sensing images predicted in section 3.2 can be compared to the adjusted water eutrophication grading criteria shown in Fig. 19. By analyzing the spatial and temporal distribution of cyanobacterial blooms at the pixel level, the relevant departments can focus on areas that are more eutrophicated in order to prevent outbreaks of cyanobacterial blooms in the waters in the future.

Table 2
Water eutrophication classification standard (Taihu Lake)

Nutritional grade	Whole-lake average chlorophyll-a concentration (μg/L)
Poor nutrition	<3
Medium nutrition	3~11
Eutrophication I	11~21
Eutrophication II	21~24
Eutrophication III	24~26
Eutrophication IV	26~78
Heavy eutrophication	≥78

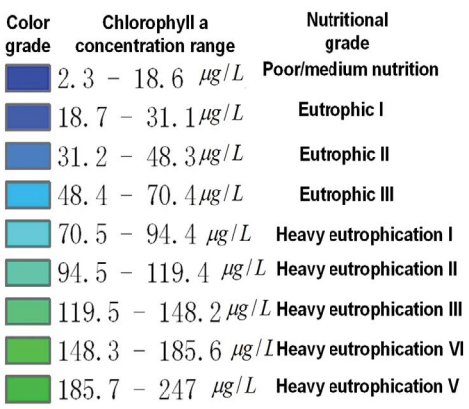


Fig. 20. Unified data scale and adjusted water eutrophication classification standard (remote sensing image).

4. Results

4.1. Results of preprocessing remote sensing image time series

4.1.1. Results of unified data scale based on pixel substitution method

According to the unified data scale (Fig. 20), the pixel value for each color grade in the original remote sensing image is replaced with the pixel value for the corresponding color grade in the unified data scale according to Eq. (9), thus completing the pixel replacement of the unified data scale is completed. A comparison of remote sensing images before and after unifying the data scale is shown in Fig. 21.

As shown in Fig. 20, the left figure is the original remote sensing image, and the right figure is the remote sensing image after the unified data scale. It can be seen that the color changes of pixels in some areas of the remote sensing image before and after the unification. By unifying the data scale, the same color grade in all remote sensing images represents the same chlorophyll-a concentration range.

4.1.2. Results of remote sensing image repair based on attention mechanism Pix2Pix model and spatial weight matrix

The downsampling parameters in the Pix2Pix model of attention mechanism for remote sensing image repair are shown in Table 3.

The upsampling parameters are shown in Table 4.

The results of remote sensing image repair are shown in Fig. 22, which shows a comparison of remote sensing images before and after repair.

First, the damaged remote sensing images are repaired using the attention mechanism Pix2Pix model, and for the parts with poor repair results, a secondary repair is performed using the spatial weight matrix on the basis of the attention mechanism Pix2Pix model repair.

Table 3
Downsampling parameters

Network layer	Number of convolution kernels	Convolution kernel size	Step length
Downsampling 1	32		
Downsampling 2	64		
Downsampling 3	128		
Downsampling 4	128		
Downsampling 5	128	3 × 3	2
Downsampling 6	256		
Downsampling 7	256		
Downsampling 8	512		
Downsampling 9	512		

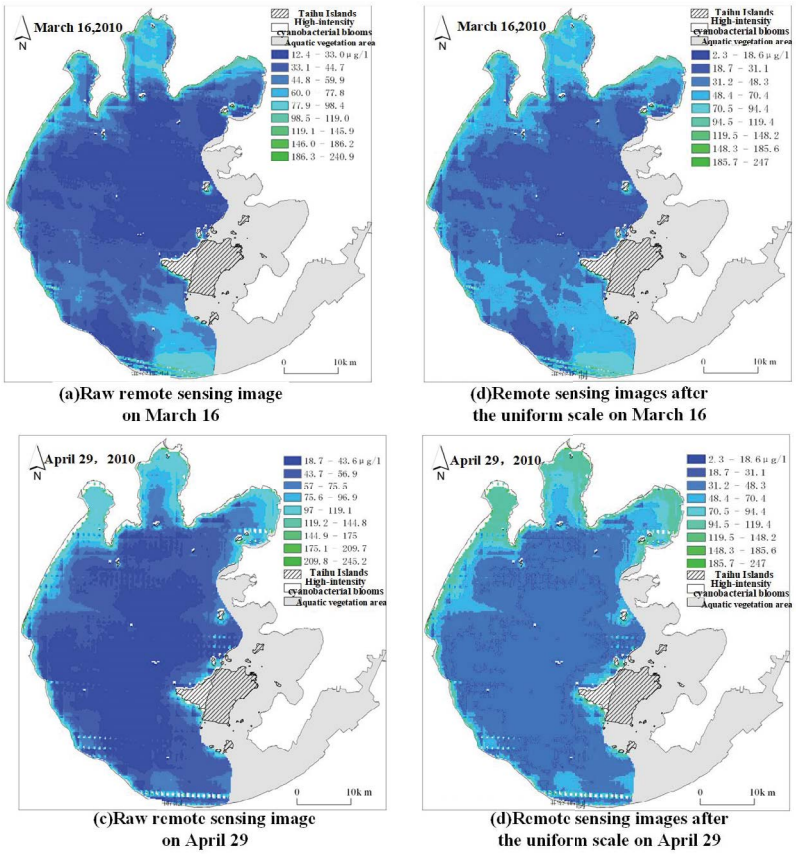


Fig. 21. Comparison of remote sensing images before and after scale unification.

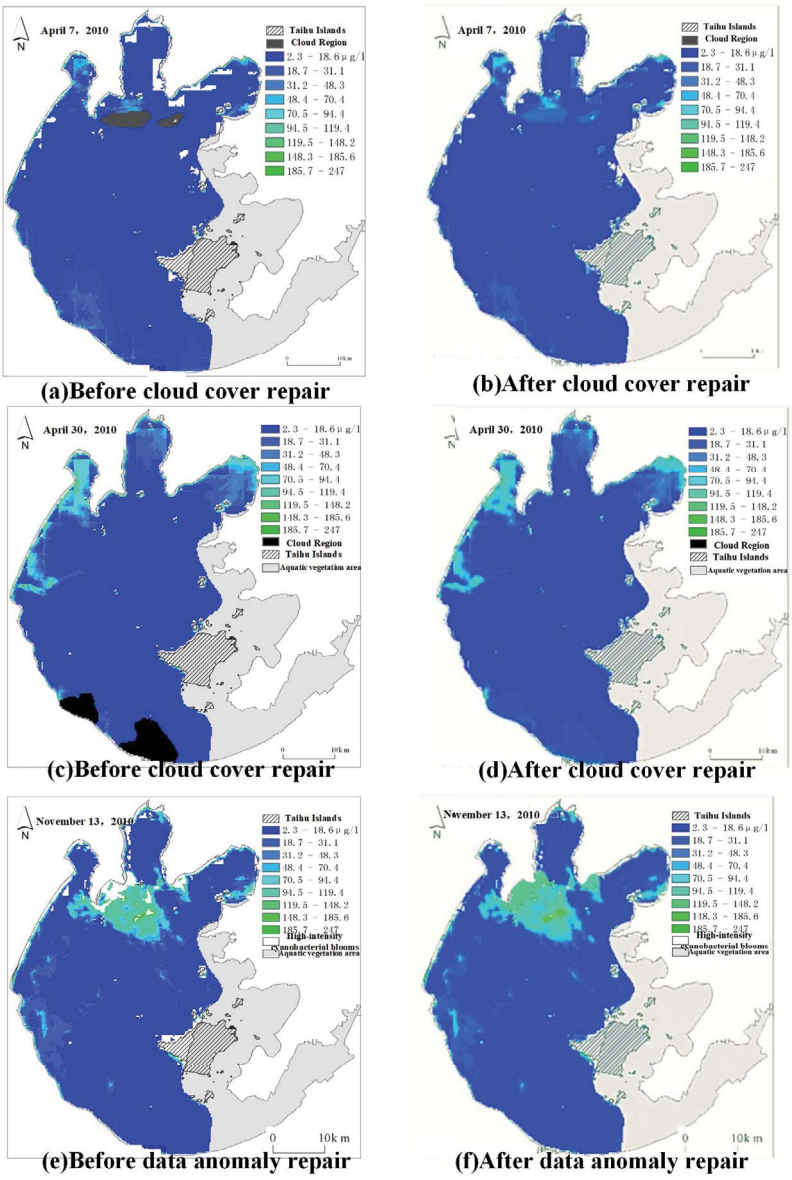


Fig. 22. Comparison of remote sensing images before and after repair.

4.1.3. Results of time series data filling based on linear interpolation

Fig. 23 shows the time series of remote sensing images after linear interpolation obtained using formula (18).

Fig. 23 depicts the real data in the original data set as March 9, March 10, March 12, and March 13, and the filling data obtained by linear interpolation as March 11. By linear interpolation, the sampling time interval in the data set is set to 1 d.

4.2. Remote sensing image prediction results based on ACL3D-Pix2Pix

4.2.1. Parameter setting

A discussion of ACL3D-Pix2Pix parameters is presented in this section.

4.2.1.1. Generator parameter settings

The generator of ACL3D-Pix2Pix is the AConvLSTM embedded in the 3D U-Net network. It mainly consists of AConvLSTM, downsampling with residuals, and upsampling with residuals. The parameter settings of AConvLSTM are shown in Table 5.

Parameters in Table 5 were derived based on the empirical method and the data format required. The parameters of the downsampling module with residuals are shown in Table 6.

An empirical method was used to determine the number and size of convolution kernels in Table 6. There are 8 layers of upsampling with residuals, and the parameters are shown in Table 7.

As shown in Table 7, the number of deconvolution kernels and the size of the deconvolution sum are also

empirically derived. Step sizes for upsampling and downsampling should be the same.

4.2.1.2. Discriminator parameter setting

The discriminator constructed in this paper is the PatchGAN model, and the network parameters are shown in Table 8.

As shown in Table 8, PatchGAN has 3 network layers. In this study, the remote sensing images of chlorophyll-a concentration, temperature, and phycocyanin concentration at the same moment were first superimposed on the channels to construct multi-factor data of one moment, and the multi-factor data of 4 historical moments were used to predict the remote sensing images of chlorophyll-a concentration at the 5th moment. First, 4 historical multi-factor data are superimposed as video frames, and an Adam optimizer is used to optimize the network with a learning rate of 0.001.

4.2.2. Prediction results of remote sensing images

Fig. 24 shows the prediction images for ACL3D-Pix2Pix.

Fig. 24a and e represent the real image and the predicted image based on the ACL3D-Pix2Pix. As can be seen intuitively, the predicted image using the method described in this paper is quite similar to the real image.

It is proposed in this paper to evaluate the prediction quality of the model more objectively by evaluating structural similarity (SSIM), peak signal-to-noise ratio (PSNR), cosine similarity (cosine), and mutual information. SSIM measures image similarity based on brightness, contrast, and structure; PSNR can reflect the mean square error between 2 images; cosine can determine the similarity between 2 images by calculating the cosine distance between the vectors; mutual information describes similarity by calculating the mutual information between 2 images. In order to compare the effectiveness of the ACL3D-Pix2Pix proposed in this paper with the existing newer pixel-level prediction methods for remote sensing image prediction, the original Pix2Pix model [30], the Channel-Pix2Pix [25] model with superimposed historical moment image channels, and the 3D U-Net as a generator were used to predict remote sensing images of 3D-Pix2Pix model for remote sensing image prediction [29].

Table 9 shows the average and standard deviation of the similarity between the predicted and real images for each prediction model.

Table 4
Upsampling parameters

Network layer	Number of deconvolution kernels	Deconvolution kernel size	Step length
Upsampling 1	512		
Upsampling 2	256		
Upsampling 3	256		
Upsampling 4	124		
Upsampling 5	128	3*3	2
Upsampling 6	128		
Upsampling 7	64		
Upsampling 8	32		
Upsampling 9	16		
Output layer	3	1*1	1

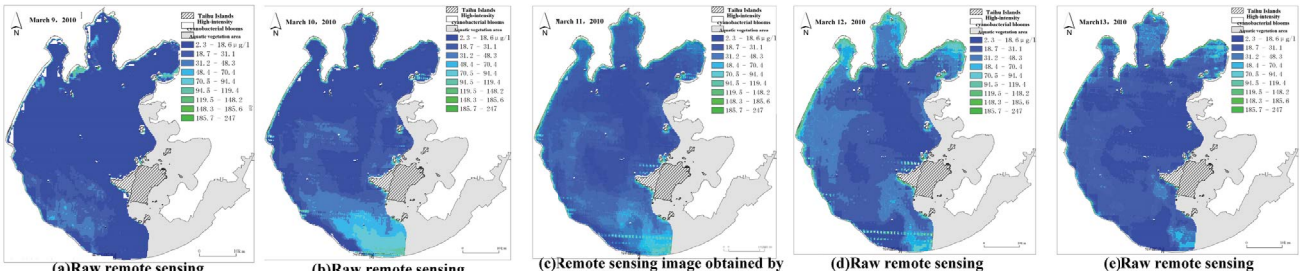


Fig. 23. Time series of remote sensing images after linear interpolation.

Table 5
AConvLSTM parameter setting

Network layer	Parameter setting	Activation function
Average pooling layer of channel attention	Output size = 1	NA
Channel attention full connection layer 1	Out features = 8	ReLu
Channel attention full connection layer 2	Out features = 32	Sigmoid
Spatial attention global average pool layer	Output size = 1	NA
Spatial attention maximum pooling layer	Output size = 1	NA
Spatial attention convolution layer	Filters = 1, kernel size = 7	Sigmoid
ConvLSTM layer	Kernel size = 3	LeakyReLu

Table 6
Downsampling parameters with residuals

Network layer	Number of convolution kernels	Convolution kernel size	Step length	Padding
Downsampling 1	32	(1,3,3)		'same'
Downsampling 2	64	(2,1,1)		'valid'
Downsampling 3	128	(1,3,3)	(1,2,2)	'same'
Downsampling 4	256	(2,1,1)		'valid'
Downsampling 5	256	(1,3,3)		'same'
Downsampling 6	512	(2,1,1)		'valid'
Downsampling 7	512	(1,3,3)	(2,2,2)	'same'

Table 7
Upsampling parameters with residuals

Network layer	Number of deconvolution kernels	Dconvolution kernel size	Step length	Padding
Upsampling 1	512	(1,3,3)		'same'
Upsampling 2	256	(2,1,1)		'valid'
Upsampling 3	256	(1,3,3)		'same'
Upsampling 4	128	(2,1,1)	(1,2,2)	'valid'
Upsampling 5	64	(1,3,3)		'same'
Upsampling 6	32	(2,1,1)		'valid'
Upsampling 7	3	(1,3,3)		'same'
conv3d	3	(4,1,1)	(1,1,1)	'valid'

Table 8
PatchGAN network parameters

Network layer	Number of convolution kernels	Convolution kernel size	Step length	Padding
conv3d-1	64	(1,3,3)	(1,2,2)	'same'
conv3d-2	128			
conv3d-3	1		(1,1,1)	'valid'

As can be seen from the last row of Table 9, the prediction results of ACL3D-Pix2Pix proposed in this paper maintain good consistency with the real images in terms of brightness, contrast, structure, mean square error, cosine distance, and mutual information. The possible reasons for this result are shown:

1) In the original Pix2Pix model, pixel-to-pixel prediction is possible, but it cannot take into account time correlation; therefore, the prediction effect of image time series is general;
2) In Channel-Pix2Pix, time correlation is taken into account, so the prediction effect is better than that of the original Pix2Pix model, but since each image is superimposed on the channel in accordance with time, it is still operating on 2-dimensional data in essence, so the prediction effect is not as good as 3D-Pix2Pix;
3) As 3D-Pix2Pix uses 3D convolution to extract spatiotemporal features, the prediction effect is superior to Channel-Pix2Pix. However, because 3D convolution is not suitable for extracting image time series features, and the method is not easy to converge during training, the prediction effect is inferior to that of The ACL3D-Pix2Pix;
4) ACL3D-Pix2Pix introduced in this paper can better extract spatiotemporal features and prevent overfitting and gradient explosion than the 3D-Pix2Pix model by introducing an attention mechanism;
5) The ACL3D-Pix2Pix proposed by this study showed better performance in the evaluation of both average and standard deviation, indicating that the model not only outperforms other models in terms of prediction accuracy, but also outperforms other models in terms of stability of prediction results.
6) ACL3D-Pix2Pix proposed in this paper is used to predict the changes in pixel values of remote sensing images over time. Therefore, this model is not only effective for pixel-level prediction of chlorophyll-a concentration remote sensing images but also applicable to the pixel-level prediction of various remote sensing images theoretically.

4.3. Results of the predicted spatial and temporal distribution of cyanobacterial blooms

Due to space constraints, Fig. 25 shows only actual and predicted images of remote sensing images obtained on October 27 and December 27, along with their corresponding nutrient grades.

As shown in Fig. 25, in October 27, the upper left side and part of the lower right side of Taihu Lake are at the level of heavy eutrophication IV, which means that the possibility of cyanobacterial bloom outbreak in this part is higher, while the rest of the areas are at the level of poor or

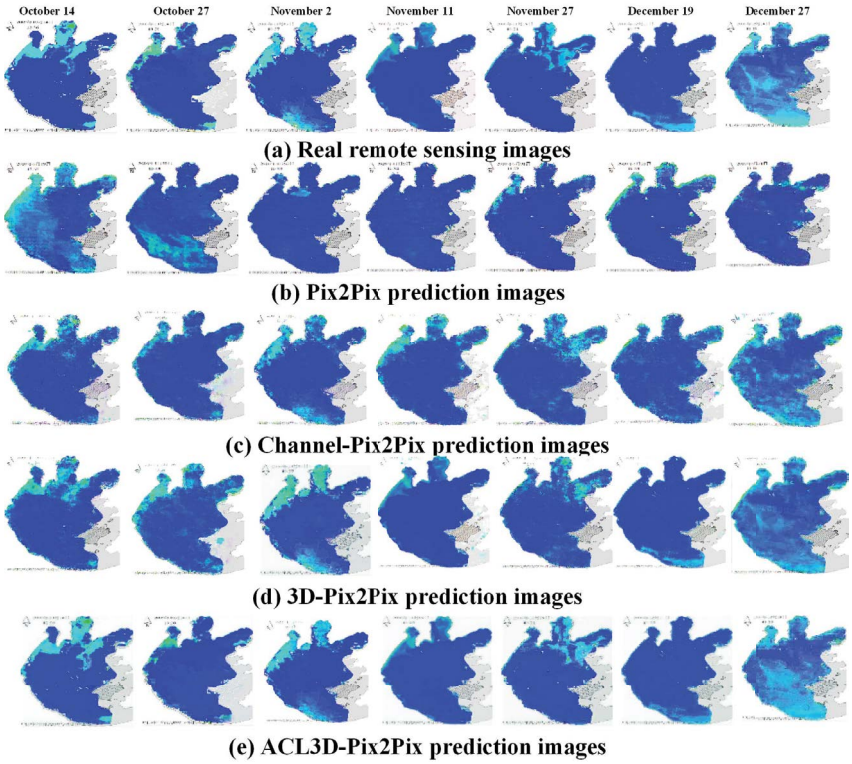


Fig. 24. Comparison of real image and the predicted image.

Table 9
Similarity comparison of model prediction results (average/standard deviation)

Models	SSIM	PSNR	Cosine	Mutual information
Pix2Pix	0.70319/0.06255	15.58665/1.69133	0.99219/0.00161	1.44716/0.23410
Channel-Pix2Pix	0.81031/0.06065	20.70479/1.80155	0.99244/0.00182	1.45557/0.23963
3D-Pix2Pix	0.83656/0.05164	27.23929/1.6650	0.99488/0.00109	1.71282/0.24161
ACL3D-Pix2Pix	0.91312/0.04080	29.79848/1.48564	0.99849/0.00104	1.985731/0.23027

medium nutrition; in the predicted image of December 27, the edge and part of the lower left side of Taihu Lake are at the level of heavy eutrophication I, while the rest of the areas are at the level of medium nutrition to eutrophication II. The treatment of areas with high eutrophication levels can effectively prevent the outbreak and reduce the impact of cyanobacterial blooms.

5. Discussion

Firstly, a series of methods for remote sensing image preprocessing were proposed in this study, based on which the ACL3D-Pix2Pix model was constructed to realize the pixel-level prediction of remote sensing images, and finally the prediction method of spatial and temporal distribution of cyanobacterial blooms based on remote sensing images was proposed.

- Compared with the traditional mechanism-driven model prediction methods [7–10], this study does not need to master the complex growth mechanism of cyanobacterial blooms and a priori knowledge, and fits the algal biomass changes by establishing the optimal mathematical expression relationship between the input and output of remote sensing images of blooms, so this study is more applicable and the model is more anti-interference;
- Compared with data-driven model prediction methods based on numerical data from underwater sensors [11–15], this study is firstly more convenient in terms of data acquisition, and underwater sensors cover a small area and can only represent the water quality indicators at the location of the sensors, while this study comprehensively considers the water quality information of the overall waters, and the prediction results are more reasonable;
- Compared with the feature-level prediction method based on remote sensing image data [23–25], although the feature-level prediction based on remote sensing image considers the overall water quality information of the water, the method only extracts and predicts the global features of the remote sensing image, and can only obtain the global average water quality change of

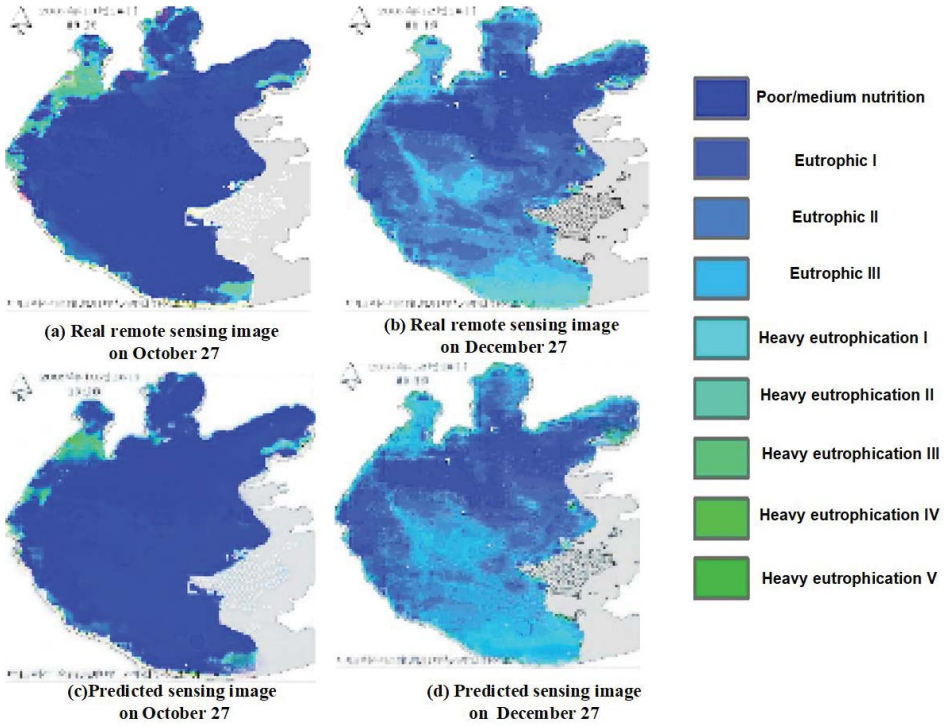


Fig. 25. Spatial and temporal distribution of cyanobacteria bloom on October 27 and December 27.

the water, while this study is to predict each pixel in the remote sensing image at the future moment, and can obtain the specific value of water quality change at any pixel location of the water, which is not possible in the previous cyanobacterial bloom prediction studies;

- Among the existing pixel-level prediction methods based on remote sensing image data, compared with the pixel-level prediction methods based on the original Pix2Pix model [30], the original Pix2Pix model is unable to consider the time dimension, and thus has a low prediction accuracy for remote sensing images of future moments, which leads to a low reference significance of the predicted spatio-temporal distribution results of cyanobacterial blooms, whereas this study is able to consider the time dimension. By embedding the ConvLSTM network in the original Pix2Pix model generator, it can fully extract the time series information of the historical images and predict the images in the future, which greatly improves the accuracy of the pixel-level prediction of the remote sensing images in the future, so the predicted spatial and temporal distribution of cyanobacterial blooms has great reference significance; Compared with the pixel-level prediction method based on a single ConvLSTM model [26], the prediction results of the single ConvLSTM model are not clear at the edges, thus leading to a lower validity of the spatial and temporal distribution of cyanobacterial blooms at the edges of the water, while this study improves the validity of the prediction results of the spatial and temporal distribution of cyanobacterial blooms by embedding the ConvLSTM into the Unet network and using the feature fusion structure in the U-Net network, the splicing of features can be used to complete the edge features and thus make the predicted image edges clearer; Compared with the pixel-level prediction method based on the existing 3D-Pix2Pix model [29], the model is difficult to train because the existing 3D-Pix2Pix model does not provide a targeted treatment for the case of gradient explosion in the network, which leads to lower stability and efficiency in predicting the spatial and temporal distribution of cyanobacterial blooms. This study avoids the gradient dispersion or explosion caused by cumulative multiplication when solving the gradient by adding the residual structure to 3D U-Net, improves the robustness of the network, and improves the loss function to speed up the convergence of the network, which improves the efficiency and stability of the spatial and temporal distribution prediction for cyanobacterial water blooms.
- Fig. 24 and Table 9 illustrate that the model proposed in this study improves the accuracy and confidence of the prediction results.
- In this study, we predict the spatial and temporal distribution of cyanobacterial blooms by predicting the eutrophication level of any location in the lake at the future time through pixel-level prediction of remote sensing images of water bodies at the future time. By referring to the prediction results of this study, we can prevent cyanobacterial blooms by only treating the areas with high eutrophication levels in the whole lake in the future. Since the whole lake does not need to be treated, not only the efficiency of treatment is improved, but also the cost of treatment is reduced. For example, for areas with high eutrophication levels predicted for future moments, water exchange can be enhanced, local deep water discharge can be carried out to remove

nutrients in the area, or mechanical methods can be used to carry out local aeration and promote water flow, it is also possible to use porous adsorption materials or specific adsorbents to adsorb pollutants such as nitrogen and phosphorus to achieve the transfer of nutrients in the region, thereby preventing the occurrence of cyanobacteria blooms.

6. Conclusion

- In this paper, based on the remote sensing image data of chlorophyll-a concentration, a series of pre-processing methods for remote sensing image time series data are proposed to improve the existing dataset, and for the improved dataset, a new remote sensing image pixel-level prediction method is proposed to achieve the prediction of chlorophyll-a concentration at the future time, and then the eutrophication level of any location in the overall water is evaluated according to the results of the obtained remote sensing image pixel-level prediction of chlorophyll-a concentration, and the spatial and temporal distribution of cyanobacterial blooms is finally predicted according to the eutrophication level.
- There are three innovations in this paper: firstly, for the problems of non-uniform pixel scale, damaged remote sensing images and unequal sampling interval in the pre-processing of remote sensing image time series, the pixel substitution method for data scale uniformity, the attention mechanism Pix2Pix model method for image repair, and the linear interpolation method for time series filling are proposed respectively to solve the above problems; secondly, the pixel-level prediction method of remote sensing images based on ACL3D-Pix2Pix model is proposed in this study, which improves the problems of low prediction accuracy and easy gradient explosion or gradient disappearance during training in the existing pixel-level prediction; Finally, the existing eutrophication classification standard for water bodies is not applicable to remote sensing image data, and based on the data scale in remote sensing images, the eutrophication classification standard for water bodies based on remote sensing images is proposed in this study to achieve the prediction of spatial and temporal distribution of cyanobacterial blooms.
- The experimental results show that the model proposed in this study can obtain the prediction results of spatial and temporal distribution of cyanobacterial blooms, so the model can provide relatively accurate early warning for the key areas of cyanobacterial bloom outbreak, and carry out targeted treatment for the key areas according to the prediction results, which not only reduces the waste of resources but also improves the treatment efficiency, promotes the progress of water environment resource protection and water environment prediction technology, and has a guiding effect on the problem of cyanobacterial bloom prediction.
- In addition, it is worth noting that the attention mechanisms constructed in this study are all spatial attention mechanisms and channel attention mechanisms, which are relatively basic attention models. In subsequent studies, the above attention mechanisms can be considered to be replaced by more complex multi-headed attention mechanisms, which can be used to enhance the model's ability to extract features; the outbreak of cyanobacterial blooms is also affected by total nitrogen and total phosphorus, as well as wind speed and rainfall. However, there is a lack of remote sensing image data mentioned above. Therefore, in future research, multimodal data fusion between remote sensing image data and numerical data can be considered to make full use of the existing data for predictive modeling of cyanobacterial bloom and water eutrophication level.

Acknowledgment

Thanks to the National Social Science Foundation of China and the Beijing Outstanding Talents Cultivation Fund for the Young Talented Team Project.

Funding statement

This study was supported by the National Social Science Foundation of China (19BGL184) and the Beijing outstanding talent training and supporting youth top-notch team project (2018000026833TD01).

Conflicts of interest

The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] I. Ogashawara, Determination of phycocyanin from space—a bibliometric analysis, Remote Sens., 12 (2022) 567, doi: 10.3390/rs12030567.
[2] E.M. Isenstein, D. Kim, M.-H. Park, Modeling for multi-temporal cyanobacterial bloom dominance and distributions using landsat imagery, Ecol. Inf., 59 (2020) 101119, doi: 10.1016/j.ecoinf.2020.101119.
[3] B. Nowruzi, N. Bouaïcha, J.S. Metcalf, S.J. Porzani, O. Konur, Plant-cyanobacteria interactions: beneficial and harmful effects of cyanobacterial bioactive compounds on soil-plant systems and subsequent risk to animal and human health, Phytochemistry, 192 (2021) 112959, doi: 10.1016/j.phytochem.2021.112959.
[4] L. Wang, T.R. Zhang, X.B. Jin, J.P. Xu, X.Y. Wang, H.Y. Zhang, J.B. Yu, Q. Sun, Z.Y. Zhao, Y.X. Xie, An approach of recursive timing deep belief network for algal bloom forecasting, Neural Comput. Appl., 32 (2020) 163–171.
[5] H.M. Li, Z.M. Jiang, G.H. Dong, L.Y. Wang, X. Huang, X. Gu, Y.J. Guo, Spatiotemporal coupling coordination analysis of social economy and resource environment of central cities in the Yellow River basin, Discrete Dyn. Nat. Soc., 2021 (2021) 6637631, doi: 10.1155/2021/6637631.
[6] U.A. Bhatti, Z. Zeeshan, M.M. Nizamani, S. Bazai, Z.Y. Yu, L.W. Yuan, Assessing the change of ambient air quality patterns in Jiangsu Province of China pre-to post-COVID-19, Chemosphere, 288 (2022) 132569, doi: 10.1016/j.chemosphere.2021.132569.
[7] W.M. Woelmer, R. Quinn Thomas, M.E. Lofton, R.P. McClure, H.L. Wander, C.C. Carey, Near-term phytoplankton forecasts reveal the effects of model time step and forecast horizon on predictability, Ecol. Appl., 32 (2022) e2642, doi: 10.1002/eap.2642.
[8] L. Wang, J.P. Kang, X.Y. Wang, J.P. Xua, X.B. Jin, H.Y. Zhang, J.B. Yu, Q. Sun, Z.Y. Zhao, L. Zheng, Formation mechanism time series modelling and expert system prediction of algal bloom, J. Environ. Prot. Ecol., 19 (2018) 1561–1572.

[9] X.Y. Wang, J. Jia, T.L. Su, Z.Y. Zhao, J.P. Xu, L. Wang, A fusion water quality soft-sensing method based on wasp model and its application in water eutrophication evaluation, J. Chem., 2018 (2018) 9616841, doi: 10.1155/2018/9616841.
[10] X.M. Fan, L.X. Song, D.B. Ji, J.K. Shen, D.F. Liu, Research on mechanism of algal blooms based on the critical depth theory, Environ. Sci. Technol., 40 (2017) 89–94.
[11] L. Wang, T.R. Zhang, J.P. Xu, J.B. Yu, X.Y. Wang, H.Y. Zhang, Z.Y. Zhao, An approach of improved dynamic deep belief nets modeling for algae bloom prediction, Cluster Comput., 22 (2019) 11713–11721.
[12] H. Yajima, J. Derot, Application of the random forest model for chlorophyll-a forecasts in fresh and brackish water bodies in Japan, using multivariate long-term databases, J. Hydroinf., 20 (2018) 206–220.
[13] W. Feki-Sahnoun, H. Njah, A. Hamza, N. Barraj, M. Mahfoudi, A. Rebai, M.B. Hassen, Using a naive Bayes classifier to explore the factors driving the harmful dinoflagellate *Karenia selliformis* blooms in a southeastern Mediterranean lagoon, Ocean Dyn., 70 (2020) 897–911.
[14] W. Ying, Gated recurrent unit based on feature attention mechanism for physical behavior recognition analysis, Comput. Sci. Inf. Eng., 26 (2023) 357–365.
[15] A.A. Kashyap, S. Raviraj, A. Devarakonda, S.R. Nayak K, K.V. Santhosh, S.J. Bhat, F. Galatioto, Traffic flow prediction models – a review of deep learning techniques, Cogent Eng., 9 (2022) 2010510, doi: 10.1080/23311916.2021.2010510.
[16] Y.L. Cun, Y. Bengio, G. Hinton, Deep learning, Nature, 521 (2015) 436–444.
[17] S. Dong, P. Wang, K. Abbas, A survey on deep learning and its applications, Comput. Sci. Rev., 40 (2021) 100379, doi: 10.1016/j.cosrev.2021.100379.
[18] J. Jin, Y.N. Zhang, Z. Hao, R.L. Xia, W.S. Yang, H.L. Yin, X.W. Zhang, Benchmarking data-driven rainfall-runoff modeling across 54 catchments in the Yellow River Basin: overfitting, calibration length, dry frequency, J. Hydrol.: Reg. Stud., 42 (2022) 101119, doi: 10.1016/j.ejrh.2022.101119.
[19] Y.R. Li, Z.F. Zhu, D.Q. Kong, H. Han, Y. Zhao, EA-LSTM: Evolutionary attention-based LSTM for time series prediction, Knowledge-Based Syst., 181 (2019) 104785, doi: 10.1016/j.knosys.2019.05.028.
[20] S. Liu, M. Li, Z. Zhang, B.H. Xiao, X.Z. Cao, Multimodal ground-based cloud classification using joint fusion convolutional neural network, Remote Sens., 10 (2018) 822, doi: 10.3390/rs10060822.
[21] L. Wang, X.Y. Wang, Z.Y. Zhao, Y.X. Wu, J.P. Xu, H.Y. Zhang, J.B. Yu, Q. Sun, Y.T. Bai, Multi-factor status prediction by 4D fractal CNN based on remote sensing images, Fractals, 30 (2022) 2240101, doi: 10.1142/S0218348X22401016.
[22] S.S. Hwang, G.W. Jeon, J.P. Jeong, J.Y. Lee, A novel time series based Seq2Seq model for temperature prediction in firing furnace process, Procedia Comput. Sci., 155 (2019) 9–26.
[23] U.A. Bhatti, Z.Y. Yu, J. Chanussot, Z. Zeeshan, L.W. Yuan, W. Luo, S.A. Nawa, Local similarity-based spatial–spectral fusion hyperspectral image classification with deep CNN and gabor filtering, IEEE Trans. Geosci. Remote Sens., 60 (2021) 1–15.
[24] C.H. Qi, S. Huang, X.F. Wang, Monitoring water quality parameters of Taihu Lake based on remote sensing images and LSTM-RNN, IEEE Access, 8 (2020) 188068–188081.
[25] L. Wang, Y.X. Wu, J.P. Xu, H.Y. Zhang, X.Y. Wang, J.B. Yu, Q. Sun, Z.Y. Zhao, Status prediction by 3D fractal net CNN based on remote sensing images, Fractals, 28 (2020) 2040018, doi: 10.1142/S0218348X20400186.
[26] Y.B. Ma, J.B. Wei, W.C. Tang, R.X. Tang, Explicit and stepwise models for spatiotemporal fusion of remote sensing images with deep neural networks, Int. J. Appl. Earth Obs. Geoinf., 105 (2021) 102611, doi: 10.1016/j.jag.2021.102611.
[27] H. Li, S. Gao, G.Y. Liu, D.L. Guo, C. Grecos, P. Ren, Visual prediction of typhoon clouds with hierarchical generative adversarial networks, IEEE Geosci. Remote Sens. Lett., 17 (2019) 1478–1482.
[28] M. Rüttgers, S.S. Lee, S.W. Jeon, D.Y. You, Prediction of a typhoon track using a generative adversarial network and satellite images, Sci. Rep.-UK, 9 (2019) 6057, doi: 10.1038/s41598-019-42339-y.
[29] A. Bihlo, A generative adversarial network approach to (ensemble) weather prediction, Neural Networks, 139 (2021) 1–16.
[30] P. Isola, J.-Y. Zhu, T.H. Zhou, A.A. Efros, Image-to-Image Translation with Conditional Adversarial Networks, Computer Vision and Pattern Recognition, Honolulu, USA, 2017, pp. 1125–1134.
[31] C. Zhang, J.H. Kim, Video object detection with two-path convolutional LSTM pyramid, IEEE Access, 8 (2020) 151681–151691.
[32] G.M. Zhu, L. Zhang, L. Yang, L. Mei, S.A.A. Shah, M. Bennamoun, P.Y. Shen, Redundancy and attention in convolutional LSTM for gesture recognition, IEEE Trans. Neural Networks Learn. Syst., 3 (2019) 1323–1335.
[33] Z.H. Hu, J.B. Zhou, K.J. Huang, E.Y. Zhang, A data-driven approach for traffic crash prediction: a case study in Ningbo, China, Int. J. Intell. Transp. Syst. Res., 20 (2022) 508–518.
[34] X.J. Shi, Z.R. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, W.-C. Woo, Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting, NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems, Vol. 1, 2015, pp. 802–810.
[35] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, Computer Vision and Pattern Recognition, Munich, Germany, 2015, pp. 234–241.
[36] X.J. Li, J.Q. Ding, J.J. Tang, F. Guo, Res2Unet: a multi-scale channel attention network for retinal vessel segmentation, Neural Comput. Appl., 34 (2022) 12001–12015.
[37] X.M. Li, H. Chen, X.J. Qi, Q. Dou, C.-W. Fu, P.A. Heng, H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes, IEEE Trans. Med. Imaging, 37 (2018) 2663–2674.
[38] J.Y. Yao, S.G. Jin, Multi-category segmentation of Sentinel-2 images based on the Swin UNet method, Remote Sens., 14 (2022) 3382, doi: 10.3390/rs14143382.
[39] H.-K. Kim, K.-Y. Yoo, H.-Y. Jung, Color image generation from LiDAR reflection data by using selected connection UNET, Remote Sens., 20 (2020) 3387, doi: 10.3390/s20123387.
[40] L. Wang, W. Li, X. Wang, J. Xu, Remote sensing image analysis and prediction based on improved Pix2Pix model for water environment protection of smart cities, PeerJ Comput. Sci., 9 (2023) 341–345.