

# Global chlorophyll-*a* concentration estimation from moderate resolution imaging spectroradiometer using convolutional neural networks

Bowen Yu,<sup>a</sup> Linlin Xu,<sup>b,c,\*</sup> Junhuan Peng,<sup>b,d</sup> Zhongzheng Hu,<sup>e</sup>  
and Alexander Wong<sup>c,f</sup>

<sup>a</sup>China Academy of Electronics and Information Technology, Beijing, China

<sup>b</sup>China University of Geosciences (Beijing), School of Land Science and Technology, Beijing, China

<sup>c</sup>University of Waterloo, Department of Systems Design Engineering, Waterloo, Canada

<sup>d</sup>Shanxi Key Laboratory of Resources, Environment, and Disaster Monitoring, Jinzhong, China

<sup>e</sup>China Center for Resources Satellite Data and Application, Beijing, China

<sup>f</sup>University of Waterloo, Waterloo AI Institute, Waterloo, Canada

**Abstract.** The accurate estimation of global chlorophyll-*a* (Chla) concentration from the large remote sensing data in a timely manner is crucial for supporting various applications. Moderate resolution imaging spectroradiometer (MODIS) is one of the most widely used earth observation data sources, which has the characteristics of global coverage, high spectral resolution, and short revisit period. So the estimation of global Chla concentration from MODIS imagery in a fast and accurate manner is significant. Nevertheless, the estimation of Chla concentration from MODIS using traditional machine learning approaches is challenging due to their limited modeling capability to capture the complex relationship between MODIS spatial-spectral observations and the Chla concentration, and also their low computational efficiency to address large MODIS data in a timely manner. We, therefore, explore the potential of deep convolutional neural networks (CNNs) for Chla concentration estimation from MODIS imagery. The Ocean Color Climate Change Initiative (OC-CCI) Chla concentration image is used as ground truth because it is a well-recognized Chla concentration product that is produced by assimilating different satellite data through a complex data processing steps. A total of 12 monthly OC-CCI global Chla concentration maps and the associated MODIS images are used to investigate the CNN approach using a cross-validation approach. The classical machine learning approach, i.e., the supported vector regression (SVR), is used to compare with the proposed CNN approach. Comparing with the SVR, the CNN performs better with the mean log root-mean-square error and  $R^2$  of being 0.129 and 0.901, respectively, indicating that using the MODIS images alone, the CNN approach can achieve results that is close to the OC-CCI Chla concentration images. These results demonstrate that CNNs may provide Chla concentration images that are reliable, stable and timely, and as such CNN constitutes a useful technique for operational Chla concentration estimation from large MODIS data. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JRS.14.034520](https://doi.org/10.1117/1.JRS.14.034520)]

**Keywords:** convolutional neural network; chlorophyll-*a*; moderate resolution imaging spectroradiometer; Ocean Color Climate Change Initiative; ocean color; deep learning.

Paper 200282 received May 7, 2020; accepted for publication Aug. 7, 2020; published online Sep. 8, 2020.

## 1 Introduction

Chlorophyll-*a* (Chla) concentration is a key indicator of the biophysical status of water bodies. Numerous satellites programs operated by the National Aeronautics and Space Administration (NASA), the National Oceanic and Atmospheric Administration (NOAA), and the European

---

\*Address all correspondence to Linlin Xu, E-mail: [xulinlin@cugb.edu.cn](mailto:xulinlin@cugb.edu.cn)

**Table 1** Summary of primary ocean color satellites.

Acronym	Full name	Operating institution	Service period
CZCS	Coastal zone color scanner	NASA	1978 to 1986
SeaWiFS	Sea-viewing wide field-of-view sensor	NASA	1997 to 2010
MERIS	Medium resolution imaging spectrometer	ESA	2002 to 2012
MODIS	Moderate resolution imaging spectroradiometer	NASA	1999 to present
VIIRS	Visible and infrared imager radiometer suite	NASA and NOAA	2011 to present
OLCI	Ocean and land color instrument	ESA	2016 to present

Space Agency (ESA) have global ocean monitoring capability (Table 1). Accurate and fast estimation of global Chla concentration from these satellites' images is crucial for various purposes, such as the net primary production derivation,<sup>1,2</sup> harmful algal blooms detection,<sup>3-6</sup> physical and biological interactions,<sup>7,8</sup> and ecosystem models validation.<sup>9,10</sup>

The twin-moderate resolution imaging spectroradiometer (MODIS), terra-MODIS and aqua-MODIS, have high temporal resolution to view the entire Earth's surface every 1 to 2 days which have provided big remote sensing data. The big data characteristics of satellite remote sensing calls for an efficient Chla concentration estimation algorithm that can handle the big remote sensing data in an efficient and effective manner. Nevertheless, the relationship between the Chla concentration and the remote sensing observations is a very complex nonlinear function, as a consequence, design of an end-to-end machine learning approach with strong learning capability that can capture the complex nonlinear relationship between the Chla concentration and the remote sensing observations is an important research issue. Support vector machine, a classical traditional machine learning approaches, has limitations due to their relative weak modeling capability compared to neural networks to handle the complexity and uncertainty in the inverse problem and also the low computational efficiency.<sup>11</sup> Deep learning, especially convolutional neural networks (CNNs), is well known for its using both spatial and spectral information, strong modeling capability and leveraging GPU for high computational efficiency, by greatly outperforming the other machine learning techniques in a lot of computer vision tasks.<sup>12-16</sup> Remote sensing scientists have seen the value of CNNs when they face the tasks of classification, segmentation, object detection, change detection, and super resolution.<sup>17-19</sup> However, very limited studies focus on using CNNs to solve regression problems. Although CNNs have huge potential for addressing the problems in Chla estimation, to the best of our knowledge, there is no study to explore the capability of CNNs for the estimation of global Chla concentration. Therefore, this paper explores deep CNN for the estimation of global Chla concentration. Nevertheless, deep CNN requires a lot of training samples, but the ground-truth field data are very limited, and as such we adopt the Ocean Color Climate Change Initiative (OC-CCI) images as ground truth.

The OC-CCI project is one of the 13 projects in the ESA CCI program that aim at addressing a particular essential climate variable from satellite observations. The objective of this project is to produce a long-term, multisensor time-series of satellite ocean color data, e.g., Chla concentration, with specific information on errors and uncertainties.<sup>20</sup> The OC-CCI Chla concentration dataset is produced by combining the ocean chlorophyll 3 (OC3) algorithm, the ocean chlorophyll 5 (OC5) algorithm, and the OCI algorithm [ocean chlorophyll 4 (OC4) algorithm + color index (CI) algorithm]<sup>21-23</sup> on the merged data of SeaWiFS, MERIS, MODIS, and VIIRS. Implemented by a sophisticated processing chain, OC-CCI intends to provide a global Chla concentration product of the highest quality, particularly on Chla retrievals from case 2 waters in global scale, although these may not be the latest.<sup>24</sup> Although inevitable biases exist in the OC-CCI products when being matched against the *in situ* observations, they are the most accessible and reliable Chla concentration product.

This paper explores the use of the CNN regression method for the end-to-end estimation of the global Chla concentration from the imagery recorded by the MODIS sensor on board

the Aqua satellite. Unlike approaches that build the relationship between reflectance and Chla concentration obtained from the SeaBASS dataset or cruise measurements,<sup>25,26</sup> the CNN method here derives the relationship between reflectance and Chla concentration acquired from the OC-CCI Chla image, which can provide abundant training samples for effective training of the CNN model. MODIS remote sensing reflectance (Rrs) images and OC-CCI monthly Chla concentration images are used as input and ground truth, respectively. Nearly 15,000,000 samples are selected for training the patch-based CNN regression model. A total of 12 monthly global Chla concentration images are produced by the CNN and support vector regression (SVR), respectively. Qualitative and quantitative analyses are used for evaluation and comparison analysis, and the results imply that the CNN model can be successfully used to estimate the global Chla concentration. This effort represents an initial attempt to estimate Chla concentration using CNN.

## 2 CNN

A CNN usually takes an image as input and outputs its corresponding class label. CNN architectures comprise typically several layers: convolutional layers, pooling layers, and fully connected layers. A slightly different version from the classical LeNet<sup>12</sup> architecture is shown in Fig. 1 to explain CNN.

The convolution layer 1 takes an RGB image  $X$  and kernels  $W^{(1)}$  which consist of weights as inputs and outputs feature maps  $a^{(1)}$ . The size of  $X$  is channel  $\times$  height  $\times$  width =  $3 \times 28 \times 28$  and the size of  $W^{(1)}$  is number  $\times$  channel  $\times$  height  $\times$  width =  $20 \times 3 \times 5 \times 5$  which means that there are 20 three-dimensional (3-D) kernels. Kernels should have the same channels as the corresponding input image. In general, there will be multiple filters and multiple feature maps are generated and stacked to form a multilayer feature map.<sup>27,28</sup> Each feature map is obtained by dot product between receptive field and kernel in specific rules of scanning the input image. The weights in the kernel will be learned by training. The process of convolutional layer 1 can be expressed as

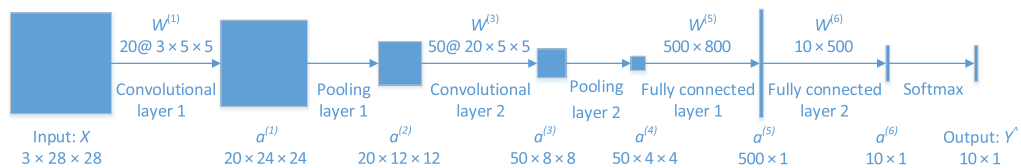
$$a^{(1)} = f[W^{(1)} * X + B^{(1)}], \tag{1}$$

where the operation of convolution is denoted by  $*$  and  $f(\cdot)$  means the activation function.  $B^{(1)}$  is the bias of first convolution layer with the same size as  $a^{(1)}$ . The height and width of feature maps  $a^{(1)}$  are decided by the height and width of kernels  $W^{(1)}$  and the step size of the kernels named stride. They will be calculated as

$$a_{\text{height}}^{(1)} = \frac{X_{\text{height}} - W_{\text{height}}^{(1)}}{\text{stride}} + 1, \tag{2}$$

$$a_{\text{width}}^{(1)} = \frac{X_{\text{width}} - W_{\text{width}}^{(1)}}{\text{stride}} + 1. \tag{3}$$

The stride of  $W^{(1)}$  is 1, so the height and width of feature maps  $a^{(1)}$  is  $28 - 5 + 1 = 24$ . The size of  $a^{(1)}$  is channel  $\times$  height  $\times$  width =  $20 \times 24 \times 24$ , and the channel is same as the number in



**Fig. 1** Structure of a CNN which is slightly different from the classical LeNet. The order of 3-D  $X$  and  $a$  are channel  $\times$  height  $\times$  width. Kernels  $W$  in each convolution layer are four-dimensional and the number before at stands for number of the 3-D kernels, so the order of the four-dimensional kernels  $W$  is number  $\times$  channel  $\times$  height  $\times$  width.

kernels  $W^{(1)}$ . It can be explained that the 20 feature maps  $a^{(1)}$  could be treated as a 3-D images as  $X$ . The activation function employs nonlinear transformation [e.g., sigmoids, tanh, rectified linear unit (ReLU), etc.] to each element in the outputs of convolutional process.

Conventionally, each convolutional layer is followed by a pooling layer in order to reduce the number of parameters to learn and reduce the variance of features,<sup>29-31</sup> which can be accomplished by operations like maximum, average, etc. No weights need to be learned in this layer. The process of pooling layer 1 can be expressed as

$$a^{(2)} = \text{down}[a^{(1)}], \tag{4}$$

where  $\text{down}[\cdot]$  represents down sampling. In general, the output of a pooling layer will be reduced by half of input, so the size of  $a^{(2)}$  is  $\text{channel} \times \text{height} \times \text{width} = 20 \times 12 \times 12$ . The process of convolutional layer 2 and pooling layer 2 are same as convolutional layer 1 and pooling layer 1 which can be expressed as

$$a^{(3)} = f[W^{(3)} * a^{(2)} + B^{(3)}], \tag{5}$$

$$a^{(4)} = \text{down}[a^{(3)}]. \tag{6}$$

A fully connected layer takes all neurons in the input layer, no matter a convolutional layer or fully connected layer, and connects them to each neuron in its layer. It is regularly used as the last few layers in a CNN architecture and yields as many outputs as labels to a classifier layer. The process of two fully connected layers can be expressed as

$$a^{(5)} = f[W^{(5)} \cdot a^{(4)}(:) + B^{(5)}], \tag{7}$$

$$a^{(6)} = W^{(6)} \cdot a^{(5)} + B^{(6)}. \tag{8}$$

In the fully connected layer 1, the 3-D  $a^{(4)}$  will be flatted to a vector with the size changing from  $\text{channel} \times \text{height} \times \text{width} = 50 \times 4 \times 4$  to  $(50 \times 4 \times 4) = 800$ . An inner product performs between  $W^{(5)}$  and  $a^{(4)}$ , and the size of  $b^{(5)}$  is same as  $a^{(5)}$ . The fully connected layer 2 is similar but without an activation function, and the output is set as 10 neurons according to the number of class, which is the 10 numbers of handwritten and machine-printed character recognition in LeNet. In the last layer, a softmax function  $\sigma(\cdot)$  will be used as classifier to generate the label  $\hat{Y}$ . Based on the equation above, the output of the CNN is formulated as

$$\hat{Y} = \sigma\{W^{(6)} \cdot f[W^{(5)} \cdot \text{down}\{f[W^{(3)} * \text{down}\{f[W^{(1)} * X + B^{(1)}]\} + B^{(3)}]\}(:) + B^{(5)}] + B^{(6)}\}. \tag{9}$$

### 3 Methodology

In this paper, a patch-based CNN regression method is used to estimate global Chla concentration from the MODIS imagery. The patch-based regression CNN model scans through images and crops a patch from the images at each location of the scanning. For each patch, it uses the regression CNN to predict the Chla concentration value of the pixel at the patch center. Then the model assembles the whole predicted values to form the global Chla concentration map. The complete procedure is shown in Fig. 2. It consists of three parts: the preprocessing of the input data, the training of the CNN model, and the data prediction employing the trained CNN, which has the same procedure as testing. Performance of the CNN method is evaluated on the whole dataset.

#### 3.1 Data

In order to train and evaluate the CNN model in the task of the Chla concentration estimation, Aqua MODIS level 3 monthly standard mapped image Rrs and OC-CCI monthly Chla concentration in version 3.1 are used as input and ground truth, respectively. Monthly data of January

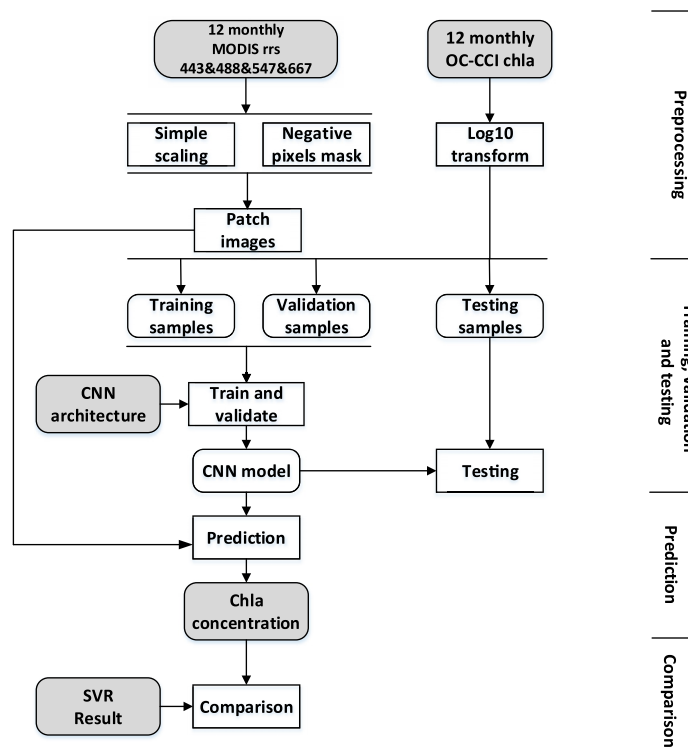


Fig. 2 Flowchart of the proposed algorithm.

Table 2 Details of the dataset used for training, validating, and testing.

Set	Input	Ground truth
Training	January 2016 MODIS Rrs 443, 488, 547, and 667	January 2016 OC-CCI
Validation	January 2016 MODIS Rrs 443, 488, 547, and 667	January 2016 OC-CCI
Testing	January 2016 MODIS Rrs 443, 488, 547, and 667	January 2016 OC-CCI
	February 2016 MODIS Rrs 443, 488, 547, and 667	February 2016 OC-CCI
	March 2016 MODIS Rrs 443, 488, 547, and 667	March 2016 OC-CCI
	April 2016 MODIS Rrs 443, 488, 547, and 667	April 2016 OC-CCI
	May 2016 MODIS Rrs 443, 488, 547, and 667	May 2016 OC-CCI
	June 2016 MODIS Rrs 443, 488, 547, and 667	June 2016 OC-CCI
	July 2016 MODIS Rrs 443, 488, 547, and 667	July 2016 OC-CCI
	August 2016 MODIS Rrs 443, 488, 547, and 667	August 2016 OC-CCI
	September 2016 MODIS Rrs 443, 488, 547, and 667	September 2016 OC-CCI
	October 2016 MODIS Rrs 443, 488, 547, and 667	October 2016 OC-CCI
	November 2016 MODIS Rrs 443, 488, 547, and 667	November 2016 OC-CCI
	December 2016 MODIS Rrs 443, 488, 547, and 667	December 2016 OC-CCI

2016 will be used for training and validating, and monthly data of the whole year (include January) will be used for testing. Details are shown in Table 2. Usually, there is only a training and testing data set for inverse problem. Due to the fact that CNNs own large parameters, it is a common practice to use a validation data set to help tune the parameters. The derived CNN

model is validated after some iterations by calculating the cost function on the validation data using the current model. The CNN with the smallest validation error will be selected as the trained CNN.

The Rrs products are obtained from the NASA ocean color website.<sup>32</sup> These data have a monthly temporal resolution and 4 km (at the equator) spatial resolution with global coverage in equirectangular projection. The atmospheric correction algorithm employed by NASA has been illustrated in detail in the previous studies.<sup>33–35</sup> We download Rrs images at 443, 488, 547, and 667 nm of each month in 2016. These wavelengths are selected because they are the main input data for the OC3 and the CI algorithms that have been used in the OC-CCI processing chain. In total, 48 images, with 4 images for each month, are used with the same image size, 4320 × 8640.

On the other hand, 12 images from the year 2016, with one image for each month, derived from the OC-CCI Chla concentration product, are used as ground truth. All Chla images have the same spatial resolution, the coverage, the map projection, and the size as the Rrs images described above, and as such they can be readily used for establishing the relationship between MODIS images and the OC-CCI Chla concentration estimation. In addition, according to the OC-CCI processing chain, all the possible ocean color satellites should be used. For the Chla concentration product in 2016, MODIS and VIIRS are the only available data to merge due to the service period. Further information about the Chla data can be found in Ref. 36.

### 3.2 Preprocessing

The ground-truth OC-CCI Chla images are logarithmically transformed due to the heavy-tailed distribution of Chla.<sup>37–40</sup> The four input MODIS Rrs images associated with a ground-truth image are scaled to achieve the same range as the log-transformed Chla (Table 3). Comparing with the normalization operation,<sup>41–43</sup> the logarithmic transformation proves in this study to be more effective way to improve the performance of a CNN. Pixels that had been flagged for land, clouds, failure in atmospheric correction, stray light, etc., are assigned the value of NAN to be excluded from your training the algorithm.

An image patch of 15 × 15 (60 km × 60 km) pixels scans the whole Rrs images to generate samples used in this study. Because the Chla image has the same size with the Rrs images, it is convenient to obtain the corresponding Chla concentration located at the patch center. If an image patch contains pixels of NAN value or lies outside the boundary, it will be abandoned. A total of 18,566,240 sample pairs are created, each of which consists of a 3-D matrix of input 4 × 15 × 15 image patch and a float scale value of the output label. Validation is crucial to train a strong CNN model, and nearly 3,000,000 samples are selected for validation and the rest 15,000,000 samples are for training. Moreover, the preprocessing procedure is the same during testing.

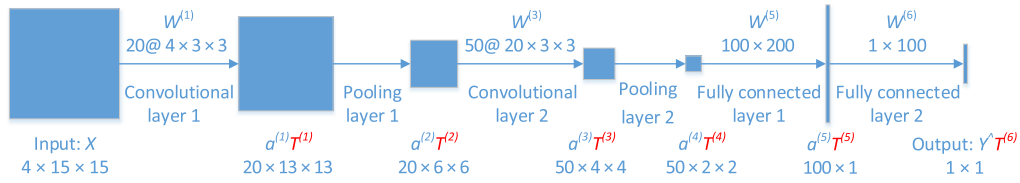
### 3.3 Structure of the Patch-Based Regression CNN

The structure of the patch-based regression CNN is demonstrated in Fig. 3. The main difference from Fig. 1 is the last layer which is specifically designed by outputting only one value to solve

**Table 3** Details of five images before and after preprocessing.

	Before preprocessing				After preprocessing			
	Max	Min	Mean	Mid	Max	Min	Mean	Mid
Rrs 443	0.0658	-0.00096	0.0079	0.0078	6.5752	0	0.78917	0.7762
Rrs 488	0.0738	0.000104	0.006	0.0062	7.384	0.0104	0.59786	0.6154
Rrs 547	0.0635	0.000802	0.0022	0.0019	6.3532	0.0802	0.22099	0.1878
Rrs 667	0.0342	-0.003	0.000225	0.000168	3.4212	0	0.02247	0.0168
Chla	96.3701	0.0012	0.2367	0.1353	1.9839	-2.9345	-0.8374	-0.8687





**Fig. 3** Structure of the patch-based regression CNN. The order of 3-D  $X$  and  $a$  are channel  $\times$  height  $\times$  width. Kernels  $W$  in each convolution layer are four-dimensional and the number before it stands for number of the 3-D kernels, so the order of the four-dimensional kernels  $W$  is number  $\times$  channel  $\times$  height  $\times$  width. The red  $T^n$  is the residual error in the  $n$ 'th layer during backpropagation which has the same size as the corresponding  $a$ .

regression problem. The CNN consists of two convolutional layers, two pooling layers and two fully connected layers. Strategy of how to tune the parameters to get the structure of the CNN will be explained in Sec. 5.

The input  $X$  is image patch with the size of  $4 \times 15 \times 15$ . The output of the convolution layer 1 is 3-D feature map  $a^{(1)}$  with the size of  $20 \times 13 \times 13$ , which is a result of the convolution of input patch with the kernel  $W^{(1)}$  with stride 1. A max-pooling layer is followed to reduce the size to  $20 \times 6 \times 6$ , and the maximum operation is chosen in the pooling layer owing to its simplicity and effectiveness. Another convolution of kernel  $W^{(2)}$  with stride 1, following a max-pooling of process on  $a^{(2)}$ , yielding feature map  $a^{(3)}$  with the size of  $50 \times 4 \times 4$  and  $a^{(4)}$  with the size of  $50 \times 2 \times 2$ . The first fully connected layer takes input of dimension  $(50 \times 2 \times 2) = 200$  and outputs 100 neurons. Unlike employing activation function sigmoids in the classical LeNet, the activation function ReLU,  $\max\{0, z\}$ , is used in each convolutional layer and the first fully connected layer because it helps accelerate convergence in training.<sup>44</sup> Different with the classification problem that the output number of the last fully connected layer is the number of the class, to address the regression here, the last fully connected layer takes as input the 100 neurons and yields one neuron which corresponds to the Chla concentration value.

### 3.4 Training and Testing

After modeling a CNN network, a loss function needs to be defined which will be minimized by training. In place of using softmax loss, which is regularly adopted in classification CNNs, the Euclidean loss is adopted for this regression CNN to minimize the error between the CNN output and the ground-truth Chla concentration provided by the OC-CCI Chla image. The loss function is equal to

$$L(W, B) = \frac{1}{2N} \sum_{n=1}^N (Y_n - \hat{Y}_n)^2, \quad (10)$$

where  $W$  represents the weights in kernels,  $B$  means bias,  $N$  is the number of samples used in each mini-batch which is a small subset of training samples,<sup>45</sup>  $\hat{Y}$  is the output of the CNN with the parameters, and  $Y$  is the ground-truth Chla concentration from OC-CCI Chla image.  $\hat{Y}$  is calculated as

$$\hat{Y} = W^{(6)} \cdot f[W^{(5)} \cdot \text{down}\{f[W^{(3)} * \text{down}\{f[W^{(1)} * X + B^{(1)}]\} + B^{(3)}]\}(\cdot) + B^{(5)}] + B^{(6)}, \quad (11)$$

which is similar with Eq. (9) in Sec. 2 but without the softmax layer. After the loss function defined, the CNN is trained to minimize the cost using stochastic gradient descent (SGD) optimization algorithm.<sup>46,47</sup> The update rule of weights  $W$  in SGD is given as

$$W_{i+1} = W_i + \Delta W_{i+1}, \quad (12)$$

$$\Delta W_{i+1} = \alpha \Delta W_i - \epsilon \left( \frac{\partial L}{\partial W} + \lambda W_i \right), \quad (13)$$

where  $W_{i+1}$  and  $\Delta W_{i+1}$  are weights and weight update at iteration  $i + 1$ , the learning rate  $\epsilon$  decides how much to learn in each step, the momentum  $\alpha$  is motivated by physical perspective for optimization problem to accelerate converge rates, and the weight decay  $\lambda$  helps to avoid overfitting.

In Fig. 3,  $T^n$  is the residual error in the  $n$ 'th layer, which has the same size as the corresponding  $a$ .  $\frac{\partial L}{\partial W}$  will be calculated in each layer from back to head. The  $\frac{\partial L}{\partial W}$  in fully connected layer 2 will be computed as

$$\frac{\partial L}{\partial W^{(6)}} = T^{(6)} \cdot a^{(5)T}, \tag{14}$$

$$T^{(6)} = Y - \hat{Y}. \tag{15}$$

The  $\frac{\partial L}{\partial W}$  in fully connected layer 1 will be computed as

$$\frac{\partial L}{\partial W^{(5)}} = T^{(5)} \cdot a^{(4)T}, \tag{16}$$

$$T^{(5)} = f'[W^{(6)T} \cdot T^{(6)}], \tag{17}$$

where  $f'(\cdot)$  represents the inverse function of ReLU. No weights and bias need to be trained in pooling layers and the  $\frac{\partial L}{\partial W}$  in convolutional layer 2 will be computed as

$$\frac{\partial L}{\partial W^{(3)}} = T^{(3)} \otimes a^{(2)T}, \tag{18}$$

$$T^{(3)} = f'\{\text{up}[T^{(4)}]\}, \tag{19}$$

$$T^{(4)} = W^{(5)T} \cdot T^{(5)}, \tag{20}$$

where  $\otimes$  represents the reverse convolution operation and  $\text{up}(\cdot)$  represents upsampling. The  $\frac{\partial L}{\partial W}$  in convolutional layer 1 will be computed as

$$\frac{\partial L}{\partial W^{(1)}} = T^{(1)} \otimes X^T, \tag{21}$$

$$T^{(1)} = f'\{\text{up}[T^{(2)}]\}. \tag{22}$$

$$T^{(2)} = W^{(3)T} \otimes T^{(3)}. \tag{23}$$

The update rule of bias  $B$  is same as weights  $W$  and the  $\frac{\partial L}{\partial B}$  in each layer will be computed as

$$\frac{\partial L}{\partial B^{(6)}} = T^{(6)}, \tag{24}$$

$$\frac{\partial L}{\partial B^{(5)}} = T^{(5)}, \tag{25}$$

$$\frac{\partial L}{\partial B^{(3)}} = T^{(3)}, \tag{26}$$

$$\frac{\partial L}{\partial B^{(1)}} = T^{(1)}. \tag{27}$$

We train the CNN for 100,000 iterations with mini-batches size of 128. During the training, validating is carried out every 1000 iterations with exploiting 3000 forward passes each iteration on a validating mini-batch size of 1000. All the parameters for this CNN are chosen after comprehensive experiments with different values, and the one with the best validating performance



will be selected. Learning rate, momentum, and weight decay are set to be 0.0001, 0.9, and 0.0005 respectively. Once the patch-based CNN model is trained, Chla concentration will be estimated by applying the trained model on the Rrs images pixel by pixel. Evaluation for the testing is processed by the corresponding Chla images.

### 3.5 Comparison with SVR

For the purpose of comparison, an SVR is adopted to estimate the Chla concentration using the same data source as the CNN method. The SVR solves complex regression problems by calculating the linear regression function in a higher-dimensional feature space where the input data are converted by a nonlinear function. The SVR is implemented using the `fitsvm` in MATLAB. Grid search method is used to tune the soft margin parameter and the Gaussian kernel parameter to maximize performance, and the two tuned parameters are 1 and 2 finally. Unlike the patch-based CNN method, the Chla concentration is performed in a pixel-based manner. Since SVR in MATLAB cannot handle large training data, only 50,000 representative samples are adopted for training.

### 3.6 Numerically Assessment of the Method

To numerically assess the agreement between the Chla concentration values achieved by the CNN model and the ground-truth OC-CCI Chla concentration values, a set of statistical indices are used, which include the coefficient of determination [ $R^2$ ; Eq. (28)], the root-mean-square error [RMSE, Eq. (29)], the mean bias [Eq. (30)], the mean absolute percentage error [MAPE, Eq. (31)], and the mean absolute error [MAE; Eq. (32)].

$$R^2 = \left\{ \frac{\sum_i [(Chla_{CNN_i} - \overline{Chla_{CNN}}) \times (Chla_{OC-CCI_i} - \overline{Chla_{OC-CCI}})]}{N \times \sigma_{Chla_{CNN}} \times \sigma_{Chla_{OC-CCI}}} \right\}^2, \quad (28)$$

$$RMSE = \sqrt{\frac{\sum (Chla_{CNN} - Chla_{OC-CCI})^2}{N}}, \quad (29)$$

$$bias = \frac{\sum (Chla_{CNN} - Chla_{OC-CCI})}{N}, \quad (30)$$

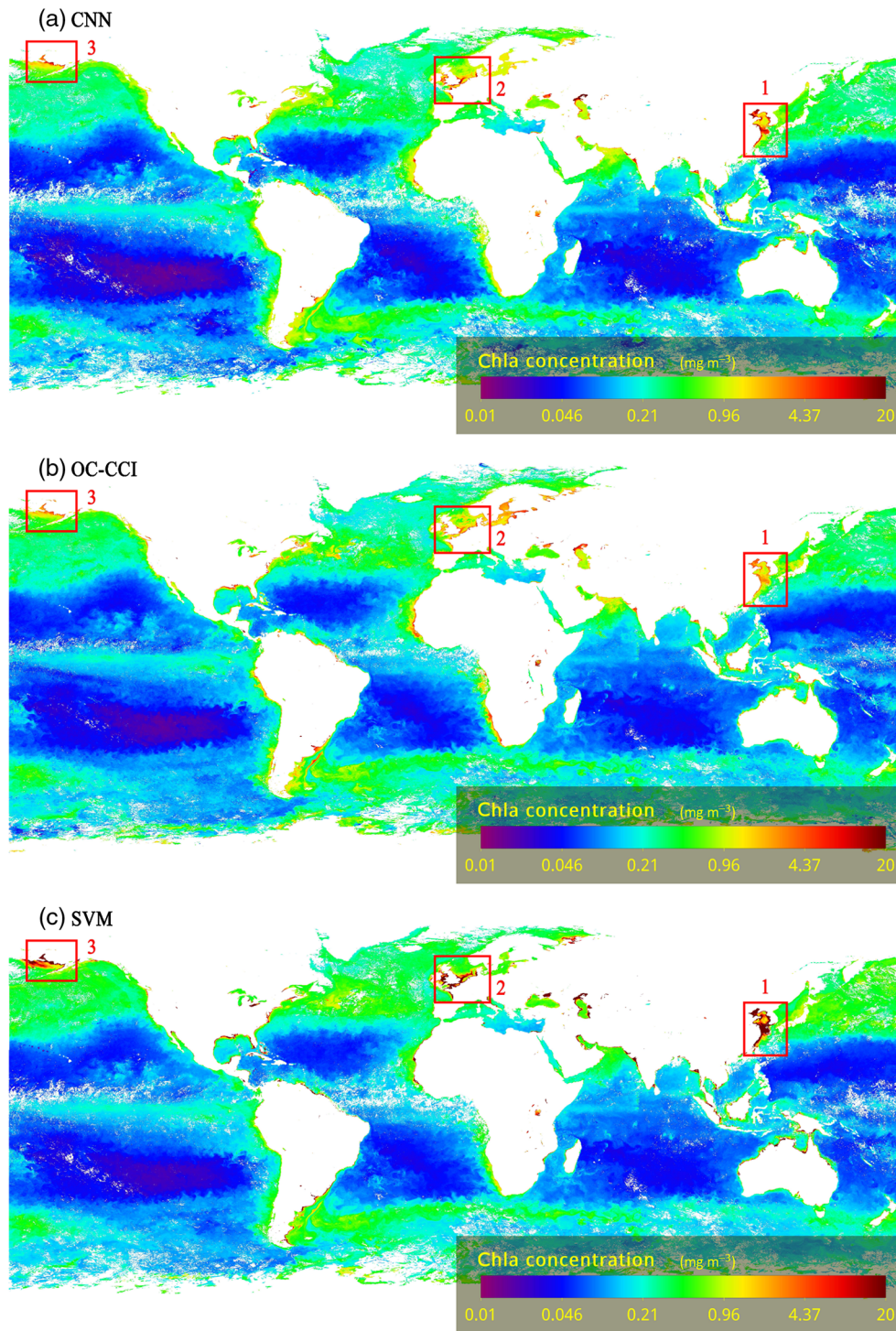
$$MAPE = \sum \left( \frac{Chla_{CNN} - Chla_{OC-CCI}}{Chla_{OC-CCI}} \right) \frac{100}{N}, \quad (31)$$

$$MAE = \frac{\sum |Chla_{CNN} - Chla_{OC-CCI}|}{N}, \quad (32)$$

where  $Chla_{CNN}$  denotes the Chla concentration values estimated by the CNN model and  $Chla_{OC-CCI}$  denotes the ground-truth OC-CCI Chla concentration values. Therefore,  $R^2$  is a measurement of the correlation between the predicted and the ground-truth data sets, and RMSE is used as a measurement of absolute error.<sup>48</sup> The pixels with NAN values in the CNN prediction map or the OC-CCI Chla ground-truth map are not used for calculating the above-described measures. Moreover, because the natural distribution of Chla is lognormal, the evaluation is in log 10 space.<sup>43,49,50</sup>

## 4 Results

Using the patch-based CNN method, 12 monthly global Chla concentration images are predicted, and the prediction of each image takes about 2 h. A one-month Chla concentration map estimated from the Rrs images by the CNN model and the SVR model, as well as the corresponding OC-CCI ground-truth image are shown in Fig. 3. For both SVR and the CNN predictions, values below 0.001 or above 100 mg m<sup>-3</sup> (representing <0.001% of the Chla estimates) are removed to limit the product as same as the OC-CCI ground-truth image. From Figs. 4(a) and



**Fig. 4** Spatial distribution of the monthly global Chla concentration in March 2016 (a) obtained using the CNN method, (b) the OC-CCI data, and (c) obtained using the SVR method. In general, Figs. 4(a) and 4(c) have strong spatial consistency with Fig. 4(b). But in the three red rectangle areas, Fig. 4(c) is much deeper red than Fig. 4(a), which means that heavier overestimation happened in the prediction map achieved by the SVR method than the CNN model.

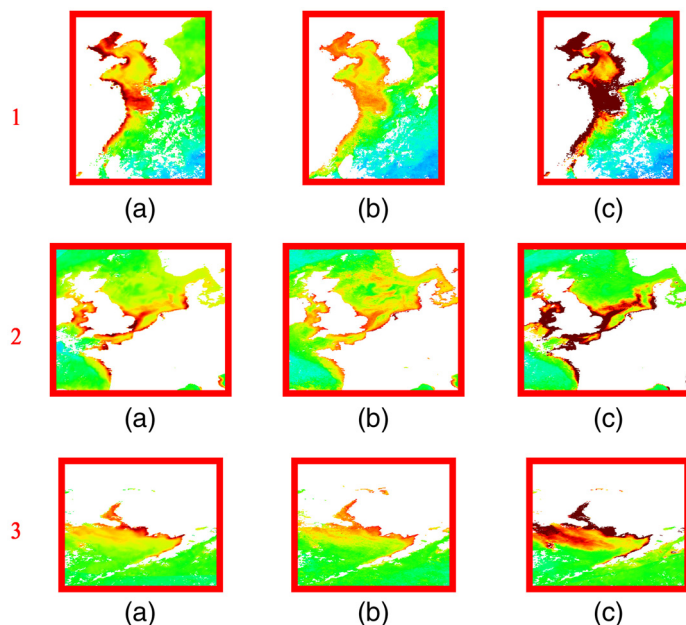
4(b), we can see that the prediction map achieved by the CNN model has a strong spatial consistency with the OC-CCI ground-truth Chla concentration map, which is achieved by assimilating various sources of information, such as the proximity to land, the depth of the ocean, and ocean currents.<sup>51</sup> Figure 6(a) shows the scatter plots of the predicted values achieved by the CNN

**Table 4** Statistical results for 12-month Chla concentration estimated by the CNN model.

Month	$R^2$	RMSE	Bias	MAPE	MAE
1	0.915	0.123	0.008	23.519	0.117
2	0.884	0.148	-0.018	24.997	0.124
3	0.893	0.121	-0.004	21.454	0.153
4	0.914	0.116	0.007	20.020	0.102
5	0.918	0.126	-0.005	18.946	0.151
6	0.912	0.128	0.013	20.669	0.118
7	0.902	0.128	0.038	26.480	0.117
8	0.899	0.134	0.050	26.362	0.123
9	0.883	0.146	-0.032	23.757	0.127
10	0.918	0.114	0.049	25.243	0.112
11	0.895	0.124	0.035	25.966	0.141
12	0.879	0.145	0.054	26.687	0.119
Mean	0.901	0.129	0.016	23.675	0.125

model and ground-truth OC-CCI Chla concentration values. In a scatter plot, good performance means that points are close to the equal-value line. Figure 6(a) shows that the majority of data points distribute around the equal-value line, indicating that the CNN method owns the capacity to predict very accurately the Chla concentration information in OC-CCI data, by learning the complex relationship between the Rrs spectra data and the OC-CCI data.

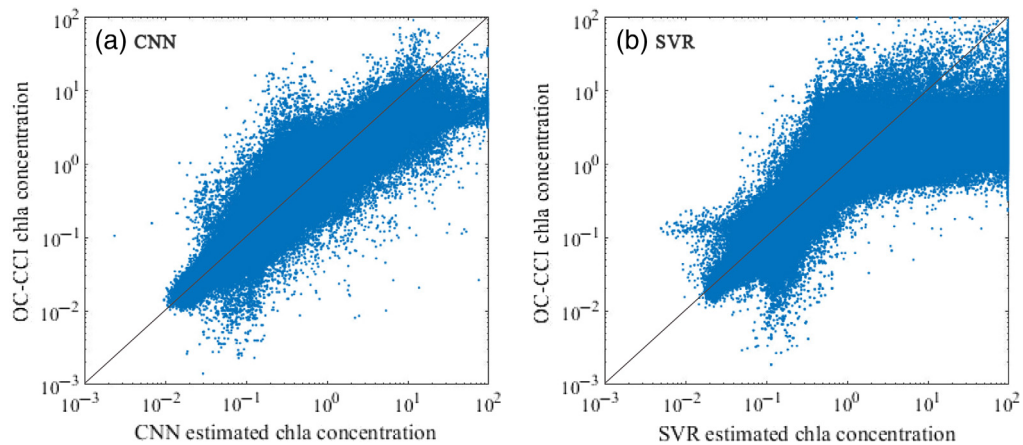
In Table 4, we can see that the  $R^2$  achieved by the CNN model is between 0.879 and 0.918 with the mean of 0.901, indicating a high consistency between the Chla concentration values predicted by the CNN model and the ground-truth OC-CCI values. The other numerical



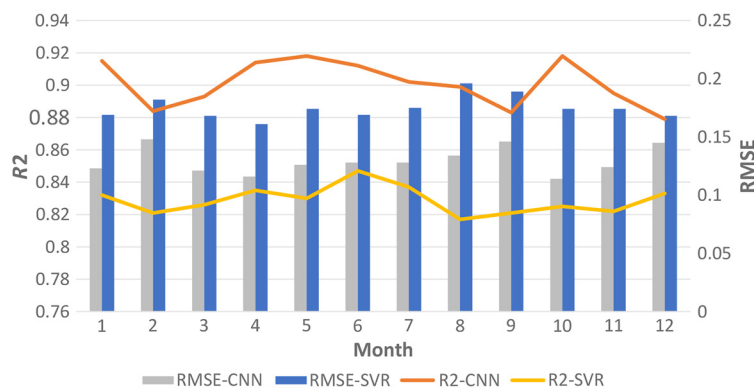
**Fig. 5** Zoomed in regions of the corresponding red rectangle areas in Fig. 4. (a) CNN, (b), OC-CCI, and (c) SVR.

measures also indicate small error and high accuracy of the proposed CNN model, i.e., the overall RMSE of the estimated Chla concentration is 0.129, and the max bias and MAPE are 0.054 and 26.687, respectively. The max and min of MAE are 0.153 and 0.102, respectively, and the mean value is 0.125. We want to highlight the fact that although the proposed CNN model is only trained on the month 1 data, it achieves stably good performance on the other 11 months' data, indicating a strong generalization capability of the CNN model.

Figure 7 shows the RMSE and  $R^2$  performance comparison between the CNN model and the SVR method on 12 months Chla concentration estimation. We can see that the  $R^2$  curve achieved by the CNN model is above the curve of the SVR method, indicating that CNN can outperform SVR over all the 12-month predictions. Moreover, the RMSE achieved by CNN are also consistently lower than SVR. The mean RMSE and  $R^2$  of the 12 months are (0.129, 0.901) and (0.175, 0.828) for CNN and SVR, respectively. Figures 4(c) and 5 shows that the color of some case 2 water areas are much deeper red than Figs. 4(a) and 5, i.e., Yellow Sea and East China Sea (red rectangle 1), English Channel and the south of North Sea (red rectangle 2), and the east of Bering Sea (red rectangle 3), which means that heavier overestimation happened in the prediction map achieved by the SVR method than the CNN model. Figure 6 supports the visual analysis above, i.e., although both the CNN model and the SVR method overestimate in the high Chla concentration areas, the SVR method performs much worse. From Fig. 6, we can also see that the CNN model outperforms the SVR method in the low Chla concentration areas.



**Fig. 6** The scatter plot between (a) the Chla concentration values estimated by the CNN model and the ground-truth OC-CCI Chla concentration values. (b) The values achieved by the SVR method and the OC-CCI values. The SVR approach tends to overestimate when the Chla concentration is high, which is consistent with Figs. 4 and 5 that the SVR results have higher values than the ground truth in high Chla concentration areas.



**Fig. 7** The performance measures comparison of CNN and SVR for 12 months Chla concentration estimation based on RMSE and  $R^2$ .



## 5 Discussion

Based on the above experimental results, we can summarize that the CNN approach can better learn the complex relationship between the Rrs and the OC-CCI Chla values than SVR, demonstrating the possibility that the CNN may be used for building an end-to-end approach for efficient Chla concentration estimation. Song et al.<sup>52</sup> proposed a novel inversion method for rough surface parameters using CNN, which microwave images of rough surfaces are used for training. In Ref. 53, a CNN-based method is employed to establish the solar flare forecasting model, in which forecasting patterns can be learned from line-of-sight magnetograms of solar active regions. Line-of-sight magnetograms of active regions observed by SOHO/MDI and SDO/HMI, and the corresponding soft x-ray solar flares observed by GOES generate the dataset for training. Previously CNN has been used for rough surface parameters and solar flare inversion. Here we demonstrate that it can also be used for Chla concentration inversion by learning the complex nonlinear relationship between the Rrs and the OC-CCI Chla values.

CNNs have a large number of hyperparameters and in this paper the patch size, kernel size, number of kernel, neuron number in the fully connected layer, number of layers, and parameters for training need to tune to obtain good performance for estimating global Chla concentration. Adopting grid search method to obtain the best combination of hyperparameters is very time-consuming. In this experiment, the hyperparameters are tuned in the flowing strategy. Structure of a CNN mentioned in Fig. 1 is the classical LeNet, and the default hyperparameters in this experiment are based on the parameters in Fig. 1. Number of layers will first to be changed manually from Fig. 1 to generate a few experiment groups with different depths. Then other hyperparameters in each group will be tuned to get the best performance CNN model which owns the structure described in Fig. 2.

In this paper, only January 2016 data are used for training and the whole-year 12-month data for testing. Still, CNN works very well, therefore CNN has strong robustness and generalization capability, indicating that it may be used for predicting long-term Chla concentration without the need to fine tune using the new observations. The mean  $R^2$  of CNN is 0.901, which is very close to OC-CCI Chla concentration, suggesting that using MODIS data only, the CNN approach can approximate the OC-CCI products, which is probably due to the high learning capability of the CNN that can fully take advantage of the information in MODIS by exploiting the complex nonlinear relationship between the Rrs and the OC-CCI Chla values. In addition, the OC-CCI Chla concentration dataset is adopted as the ground truth because it is well-recognized Chla concentration product that is produced by assimilating different satellite data through a complex data processing steps. The GPU Nvidia Quadro M2000 is used in this study, with the GPU memory of 4 GB, NVIDIA CUDA cores of 768, and boost clock of 1180 MHz. A more powerful GPU like Nvidia GeForce GTX 1080 Ti with the GPU memory of 11 GB, NVIDIA CUDA cores of 3584, and boost clock of 1582 MHz will help deal with big data. With better GPU, more GPUs, and batch processing during the prediction stage, we can further shorten the processing time.

Although the mean RMSE and  $R^2$  of CNN are acceptable, overestimation of CNN can be found from Figs. 4(a), 5, and 6(a) in the high Chla concentration areas, which may be caused by the imbalance quantities of the case 1 and the case 2 waters training samples. During sample selection, we scan the whole global Chla image to train the CNN model, resulting in the case 1 samples are much more than case 2 samples. More comprehensive studies, like combining sea surface temperature and digital bathymetry model data as input, are required to in the next stage to solve the overestimation problem. Although OC-CCI is the most accessible and reliable Chla concentration product, uncertainties may exist that will affect the performance of model. In the future, if more reliable data appear, further studies will be conducted to train and evaluate the CNN model.

## 6 Conclusion

In this study, a CNN method was applied to MODIS Rrs images to estimate global Chla concentration. The CNN took the patches of four MODIS Rrs images as input and generated the Chla concentration directly. The OC-CCI Chla concentration image was used as ground-truth for

training. A total of 12 monthly global Chla concentration images were produced and the generation of each image takes about 2 h. Qualitative and quantitative analyses were used for evaluation and comparison analysis, and the results implied that the CNN constitutes an accurate, fast, and robust method for the estimation of the global Chla concentration. Considering the big data characteristic of remotely sensed global Chla data, these characteristics of the CNN model is of great importance.

## 7 Appendix

All experiments are conducted on a 64 bits Intel Xeon E5-2640 v3 workstation with 2.6 GHz of clock and 32 GB of RAM memory. A GPU Nvidia Quadro M2000 with 4 GB memory and version 8.0 CUDA. Caffe,<sup>54</sup> a popular deep learning framework especially for CNN which is short for convolutional architecture for fast feature embedding, under Ubuntu 14.04 LTS is used in this paper. Caffe provides a convenient way to implement the CNN models by defining the architecture of CNN such as the number of layers, the type of layer and the strategy for optimization. The preprocessing of the data and the numerical assessment are performed under MATLAB R2014a. The data type of CNN in Caffe for reading image patches and their corresponding Chla values is of the HDF5 format instead of the default Lightning Memory-Mapped Database format to achieve easily storage of multichannel images and float labels.

## Acknowledgments

We would like to thank the anonymous reviewers for their valuable comments, which helped to improve the quality of this paper. The authors declare no conflicts of interest.

## References

1. M. J. Behrenfeld and P. G. Falkowski, "Photosynthetic rates derived from satellite-based chlorophyll concentration," *Limnol. Oceanogr.* **42**(1), 1–20 (1997).
2. M.-E. Carr et al., "A comparison of global estimates of marine primary production from ocean color," *Deep Sea Res. Part II* **53**(5–7), 741–770 (2006).
3. R. P. Stumpf, "Applications of Satellite Ocean color sensors for monitoring and predicting harmful algal blooms," *Hum. Ecol. Risk Assess.* **7**(5), 1363–1368 (2001).
4. Y.-J. Park, K. Ruddick, and G. Lacroix, "Detection of algal blooms in European waters based on satellite chlorophyll data from MERIS and MODIS," *Int. J. Remote Sens.* **31**(24), 6567–6583 (2010).
5. P. Shanmugam, "A new bio-optical algorithm for the remote sensing of algal blooms in complex ocean waters," *J. Geophys. Res. Oceans* **116**(C4) (2011).
6. D. Blondeau-Patissier et al., "A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans," *Prog. Oceanogr.* **123**, 123–144 (2014).
7. J. I. Goes et al., "Warming of the Eurasian landmass is making the Arabian sea more productive," *Science* **308**(5721), 545–547 (2005).
8. C. R. McClain, "A decade of satellite ocean color observations," *Annu. Rev. Mar. Sci.* **1**, 19–42 (2009).
9. W. W. Gregg and N. W. Casey, "Modeling coccolithophores in the global oceans," *Deep Sea Res. Part II* **54**(5–7), 447–477 (2007).
10. G. Lacroix et al., "Validation of the 3D biogeochemical model MIRO&Co with field nutrient and phytoplankton data and MERIS-derived surface chlorophyll a images," *J. Mar. Syst.* **64**(1–4), 66–88 (2007).
11. M. T. McCann, K. H. Jin, and M. Unser, "Convolutional neural networks for inverse problems in imaging: a review," *IEEE Signal Process. Mag.* **34**(6), 85–95 (2017).
12. Y. LeCun et al., "Gradient-based learning applied to document recognition," *Proc. IEEE* **86**(11), 2278–2324 (1998).

13. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Adv. Neural Inf. Process. Syst.*, pp. 1097–1105 (2012).
14. C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 1–9 (2015).
15. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**(7553), 436 (2015).
16. K. He et al., "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 770–778 (2016).
17. L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: a technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.* **4**(2), 22–40 (2016).
18. X. X. Zhu et al., "Deep learning in remote sensing: a comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.* **5**(4), 8–36 (2017).
19. J. E. Ball, D. T. Anderson, and C. S. Chan, "Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community," *J. Appl. Remote Sens.* **11**(4), 042609 (2017).
20. S. Lavender, T. Jackson, and S. Sathyendranath, "The ocean colour climate change initiative: Merging ocean colour observations seamlessly," *Ocean Challenge* **21**, 29–31 (2015).
21. H. R. Gordon and A. Y. Morel, "Remote assessment of ocean color for interpretation of satellite visible imagery: a review," *Lect. Notes Coast. Estuarine Stud.* **4** (1983).
22. J. E. O'Reilly et al., "Ocean color chlorophyll algorithms for SeaWiFS," *J. Geophys. Res. Oceans* **103**(C11), 24937–24953 (1998).
23. C. Hu, Z. Lee, and B. Franz, "Chlorophyll a algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference," *J. Geophys. Res. Oceans* **117**(C1), C01011 (2012).
24. Plymouth Marine Laboratory, "Ocean colour climate change initiative product user guide," 2016, [http://www.esa-oceancolour-cci.org/?q=webfm\\_send/317](http://www.esa-oceancolour-cci.org/?q=webfm_send/317).
25. M. W. Matthews, "A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters," *Int. J. Remote Sens.* **32**(21), 6855–6899 (2011).
26. D. Odermatt et al., "Review of constituent retrieval in optically deep and complex waters from satellite imagery," *Remote Sens. Environ.* **118**, 116–126 (2012).
27. L. Wang, K. A. Scott, and D. A. Clausi, "Sea ice concentration estimation during freeze-up from sar imagery using a convolutional neural network," *Remote Sens.* **9**(5), 408 (2017).
28. L. Wang et al., "Sea ice concentration estimation during melt from dual-pol SAR scenes using deep convolutional neural networks: a case study," *IEEE Trans. Geosci. Remote Sens.* **54**(8), 4524–4533 (2016).
29. F. Hu et al., "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sens.* **7**(11), 14680–14707 (2015).
30. K. Nogueira, O. A. Penatti, and J. A. dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognit.* **61**, 539–556 (2017).
31. E. Maggiori et al., "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.* **55**(2), 645–657 (2017).
32. <http://oceancolor.gsfc.nasa.gov/>.
33. H. R. Gordon and M. Wang, "Influence of oceanic whitecaps on atmospheric correction of ocean-color sensors," *Appl. Opt.* **33**(33), 7754–7763 (1994).
34. Z. Ahmad et al., "New aerosol models for the retrieval of aerosol optical thickness and normalized water-leaving radiances from the SeaWiFS and MODIS sensors over coastal regions and open oceans," *Appl. Opt.* **49**(29), 5545–5560 (2010).
35. S. W. Bailey, B. A. Franz, and P. J. Werdell, "Estimation of near-infrared water-leaving reflectance for satellite ocean color data processing," *Opt. Express* **18**(7), 7521–7527 (2010).
36. <http://www.esa-oceancolour-cci.org/>.
37. J. W. Campbell, "The lognormal distribution as a model for bio-optical variability in the sea," *J. Geophys. Res. Oceans* **100**(C7), 13237–13254 (1995).
38. P. Cipollini et al., "Retrieval of sea water optically active parameters from hyperspectral data by means of generalized radial basis function neural networks," *IEEE Trans. Geosci. Remote Sens.* **39**(7), 1508–1524 (2001).



39. L. Xu, J. Li, and A. Brenning, "A comparative study of different classification techniques for marine oil spill identification using radarsat-1 imagery," *Remote Sens. Environ.* **141**, 14–23 (2014).
40. N. Al-Naimi et al., "Evaluation of satellite retrievals of chlorophyll-a in the Arabian Gulf," *Remote Sens.* **9**(3), 301 (2017).
41. V. Mnih, "Machine learning for aerial image labeling," PhD Thesis, University of Toronto, Canada (2013).
42. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 3431–3440 (2015).
43. R. Alshelhi et al., "Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks," *ISPRS J. Photogramm. Remote Sens.* **130**, 139–149 (2017).
44. V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, pp. 807–814 (2010).
45. N. Mboga et al., "Detection of informal settlements from VHR images using convolutional neural networks," *Remote Sens.* **9**(11), 1106 (2017).
46. L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proc. COMPSTAT*, Springer, pp. 177–186 (2010).
47. Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," *Lect. Notes Comput. Sci.* **7700**, 437–478 (2012).
48. L. G. Vilas, E. Spyrakos, and J. M. T. Palenzuela, "Neural network estimation of chlorophyll a from MERIS full resolution data for the coastal waters of Galician Rias (NW Spain)," *Remote Sens. Environ.* **115**(2), 524–535 (2011).
49. D. A. Siegel et al., "Regional to global assessments of phytoplankton dynamics from the SeaWiFS mission," *Remote Sens. Environ.* **135**, 77–91 (2013).
50. M. Woźniak et al., "Empirical model for phycocyanin concentration estimation as an indicator of cyanobacterial bloom in the optically complex coastal waters of the Baltic Sea," *Remote Sens.* **8**(3), 212 (2016).
51. S. Dasgupta, R. P. Singh, and M. Kafatos, "Comparison of global chlorophyll concentrations using MODIS data," *Adv. Space Res.* **43**(7), 1090–1100 (2009).
52. T. Song et al., "Inversion of rough surface parameters from SAR images using simulation-trained convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.* **15**, 1130–1134 (2018).
53. X. Huang et al., "Deep learning based solar flare forecasting model. I. Results for line-of-sight magnetograms," *Astrophys. J.* **856**(1), 7 (2018).
54. Y. Jia et al., "Caffe: convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, ACM, pp. 675–678 (2014).

**Bowen Yu** received his BSc, MEng, and PhD degrees from China University of Geosciences, Beijing, China, in 2012, 2015, and 2019, respectively, all in remote sensing. He is currently an engineer at China Academy of Electronics and Information Technology. His research interests are in the field of deep learning for remote sensing application.

**Linlin Xu** received his BEng and MSc degrees in geomatics engineering from the China University of Geosciences, Beijing, China, in 2007 and 2010, respectively, and his PhD in geography from the University of Waterloo, Waterloo, Ontario, Canada, in 2014. His research interests include machine learning, deep learning, and remote sensing image processing, especially hyperspectral, and synthetic aperture radar image processing.

**Junhuan Peng** received his PhD in geodesy from the Wuhan University, Wuhan, China, in 2003. Currently, he is a professor with the School of Land Science and Techniques, China University of Geosciences, Beijing, China. His research interests are in the areas of spatial statistics, robust estimation and their associated application in surveying engineering, image geodesy, remote sensing, and satellite geodesy.

**Zhongzheng Hu** received his BEng and MEng from China University of Geosciences, Beijing, China, in 2015 and 2019, both in surveying and mapping. Currently, he is an engineer at China

Center for Resources Satellite Data and Application. His research interests are in the field of deep learning for remote sensing application.

**Alexander Wong** received his BSc degree in computer engineering, his MSc degree in electrical and computer engineering, and his PhD in systems design engineering from the University of Waterloo, Waterloo, Ontario, Canada, in 2005, 2007, and 2010, respectively. Currently, he is an assistant professor with the Department of Systems Design Engineering, University of Waterloo. He has authored refereed journal and conference papers, as well as patents, in various fields, such as computer vision, graphics, image processing, multimedia systems, and wireless communications. His research interests include image processing, computer vision, pattern recognition, and cognitive radio networks, with a focus on biomedical and remote sensing image processing and analysis such as image registration, image denoising and reconstruction, image super-resolution, image segmentation, tracking, and image and video coding and transmissions. He was a recipient of the Outstanding Performance Award, the Engineering Research Excellence Award, the Early Researcher Award from the Ministry of Economic Development and Innovation, the Best Paper Award by the Canadian Image Processing and Pattern Recognition Society, and the Alumni Gold Medal.