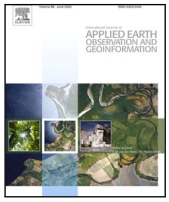
International Journal of Applied Earth Observation and Geoinformation 122 (2023) 103421

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



A three-step machine learning approach for algal bloom detection using stationary RGB camera images

Zhenyu Tan [a,b], Chen Yang [a], Yinguo Qiu [c], Wei Jia [a], Chenxi Gao [a], Hongtao Duan [a,b,c,*]

[a] College of Urban and Environmental Sciences, Northwest University, Xi'an 710127, China
[b] Shaanxi Key Laboratory of Earth Surface System and Environmental Carrying Capacity, Northwest University, Xi'an 710127, China
[c] Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China

ARTICLE INFO

Keywords:
Algal bloom detection
RGB images
Surveillance cameras
Machine learning
Lake Chaohu

ABSTRACT

Stationary surveillance cameras deployed around lakes can provide continuous real-time observations of key water areas for harmful algal bloom (HAB). They can be used to supplement remote sensing-based monitoring in situations that satellites cannot handle. While some cameras were initially installed for other purposes, and the poses are not fixed during operation, hence, detecting HABs remains a challenging task due to the diverse surface features present in image frames. A novel three-step machine learning approach was proposed in this paper to address this problem. The acquired images are initially classified using the first model, and images with certain HABs undergo further examination. A second model is employed to generate a water mask, thereby eliminating interferences from non-water features. Finally, the third model is applied to detect and identify HABs specifically within water areas. The experiments showed that the three steps implemented in sequence can effectively extract distinct HABs from RGB images captured under various shooting poses. The overall pixel-level accuracy, intersection over union, and F1 score reached 0.83, 0.76, and 0.76, respectively, on 1969 images from August to September 2020. The novelty of our approach is attributed to that the combination of the three steps can significantly abate the adverse influence of an external environment; thus, the final detection can be performed with satisfactory accuracy. In practice, the approach was applied in Lake Chaohu and consistently reports the real-time status of HABs along the bank. It exhibits substantial potential for the application in eutrophic lakes to avoid HAB-induced secondary disasters.

1. Introduction

Recent years witnessed the process of eutrophication in lake waters, caused by natural environmental changes and anthropogenic activities (Hou et al., 2022). Algal blooms are a direct consequence of water eutrophication. Excessive algae growth is often harmful; they are thus referred to as harmful algal blooms (HABs). HABs not only release toxins and harmful substances, jeopardizing water safety, but also cause the death of biomass in water, resulting in massive economic losses (Janssen et al., 2019). The dangers of HABs are worse and even fatal for urban lakes. One of the immediate impacts is from the constant malodor emitted by the decaying algae. If a lake is the main source of drinking water, it could even cause illnesses. Hence, HABs should be detected and disposed of timely in important areas, such as water sources and scenic locations. However, detecting HABs in a timely and accurate manner still remains a challenge.

Multispectral remote sensing has been recognized as a cost-effective and efficient monitoring tool for HAB detection (Oyama et al., 2015; Gholizadeh et al., 2016). HABs can be distinguished from normal waters by their distinctive spectral characteristics that normal water shows strong absorption at the red and near-infrared wavelengths, while phytoplankton present in HABs exhibit a distinct reflection (Duan et al., 2017). Various models have been developed based on this fact to detect algal distributions in lakes (Page et al., 2018; Khan et al., 2021), such as the maximum chlorophyll index (MCI) and floating algae index (FAI) (Hu, 2009). However, because of the ever-changing wind, water temperature, and radiation conditions, the algal distribution may vary significantly even on a daily basis (Huang et al., 2013). Commonly used satellite data for large or medium lakes, such as the Moderate Resolution Imaging Spectroradiometer (MODIS), can enable semidiurnal observations but with a low spatial resolution (Duan et al., 2017). However, due to the tradeoff between the spatial and temporal resolutions of remote sensing observations (Tan et al., 2022), satellite-based monitoring cannot meet the demand for small lakes or key regions of large lakes with fast-growing HABs. In this case, unmanned

* Corresponding author at: College of Urban and Environmental Sciences, Northwest University, Xi'an 710127, China.
E-mail addresses: tanzhenyu@nwu.edu.cn (Z. Tan), htduan@niglas.ac.cn (H. Duan).

https://doi.org/10.1016/j.jag.2023.103421
Received 26 December 2022; Received in revised form 3 July 2023; Accepted 8 July 2023
Available online 20 July 2023

aerial vehicles (UAVs) are leveraged to capture images of lakes during particular seasons or in specific regions when and where HABs frequently occur (Kislik et al., 2018). Nevertheless, weather factors, such as clouds and rain, can easily affect optical remote sensing and UVAs. Moreover, UAV-based monitoring necessitates manual control, resulting in substantial labor expenses, and the restricted operational duration further restricts its widespread application.

Meanwhile, stationary cameras can be employed as an auxiliary instrument for real-time monitoring and alarming. First, RGB cameras are comparatively inexpensive; they have already been deployed around important lakes for various purposes, such as detecting illicit fishing activities, enabling their utilization for continuous real-time monitoring throughout the day. In addition, cameras are typically mounted at a low height, rendering them less sensitive to weather conditions. Ongoing efforts have been directed towards harnessing surveillance cameras for environmental monitoring and ecological protection (Wang et al., 2022; Pedrayes et al., 2022). However, unlike ordinary object detection, HABs do not have certain sizes, shapes, and textures; recognition of HABs necessitates contextual understanding within the intricate surroundings of a water body, thus HAB detection through RGB cameras poses a promising yet challenging endeavor.

Recent studies on HAB detection with RGB images mainly employed UAVs because the camera pose can be set appropriately in advance, and the acquired image can be further corrected and calibrated to mitigate the impacts from the drone platform and external environmental factors (Wu et al., 2019). Among the diverse techniques explored for HAB detection, spectral indices have emerged as the prevailing methodology (Wu et al., 2019; Kislik et al., 2018). Moreover, computer vision and deep learning methods have also been explored and applied (Samantaray et al., 2018). These methods function quite well in the case where the image is corrected and calibrated without much environmental interferences. Of course, the methods proposed for UAVs can also be borrowed and applied to stationary cameras. One study established a specialized camera sensor network and integrated the spectral index and machine learning for HAB detection (Wang et al., 2015). However, the cameras in this study were deliberately deployed for HAB detection, capturing imagery solely focused on the water body. In the case of multipurpose monitoring scenarios where camera poses undergo frequent alterations, the complexity of the task significantly amplifies. Consequently, little research has been conducted on this issue.

In this paper, we dedicated to develop a HAB detection method that can function in a situation where cameras work for multiple purposes with unfixed shooting poses in a complex environment. As far as we are concerned, two main challenges need to be addressed: the absence of key spectra regarding the RGB imagery and the interference caused by the complex environment depicted in the image content. Specifically, previous studies demonstrated that the wavelengths in the red and near-infrared are much superior to the RGB-based method (Binding et al., 2013). However, RGB cameras are widely deployed and lacks HAB-sensitive bands. The second challenge stems from the complexity of image content. Many cameras are not installed for HAB monitoring, and the camera poses, controlled by the pan–tilt–zoom (PTZ) parameters, are adjusted dynamically. The pan and tilt parameters indicate the horizontal and vertical angles of the imaging center, and the zoom factor affect the field of view. Unfixed PTZ parameters introduce significant variations in the image content, encompassing diverse surface features at different scales. Moreover, environmental factors also exert certain influences. For example, sunlight, water turbidity, and ripples can make algal blooms vague and indistinct; green plants and object shadows appear to share a similar greenish hue to cyanobacteria. All these factors combined render it difficult to perform the detection.

By introducing advanced machine learning and deep learning technologies into HAB detection, we proposed a three-step approach, including scene classification, water identification, and algal bloom detection, progressively stripping interferences from complex environment, to obtain accurate HAB recognition. With this approach, first, HABs can be located automatically in time without laborious artificial patrol. Second, early warning of HABs can be alarmed through the establishment of threshold criteria, and pollutants can be disposed of as soon as possible. Third, analysis of variations among different time periods offers insights into HAB dynamics, enabling the generation of statistical models for enhanced understanding and prediction of HAB occurrences.

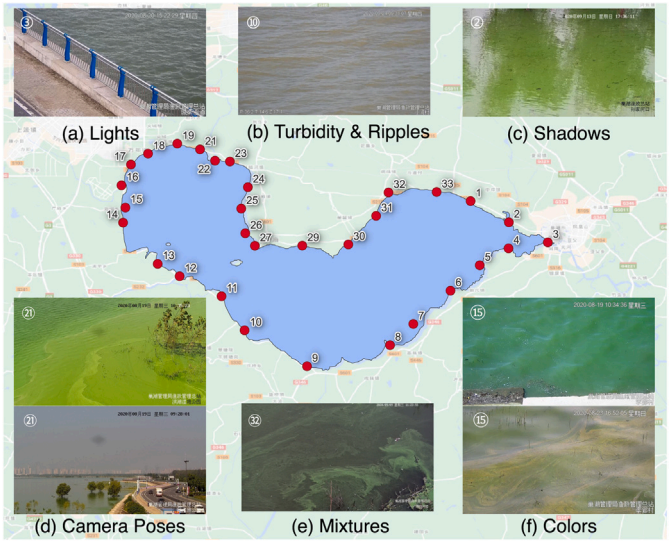


Fig. 1. Examples of RGB images captured by the surveillance cameras in Lake Chaohu (the red dots indicate the camera locations; the associated numbers are the station IDs. Subfigures exhibit some typical examples shot by the cameras; the circled numbers in subfigures correspond to the station IDs). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2. Materials and methods

2.1. Study area and datasets

Lake Chaohu (31°25′~31°43′N, 117°17′~117°52′E) is the fifth-largest freshwater lake in China with an area of 760 km^2 (Zhang et al., 2015). Over the past decades, the lake has been subject to recurrent and significant levels of eutrophication. In response to this issue, the government has undertaken various strategies aimed at mitigating cyanobacteria blooms, such as artificial patrol, remote sensing monitoring, and model prediction. The cameras installed around the lake provide an excellent opportunity to perform real-time inspection of key regions and facilitate routine satellite monitoring, so Lake Chaohu was chosen as our study area.

Fig. 1 illustrates the locations of stationary cameras and some examples of the collected images. The annotated numbers identify the in-situ camera stations. Subfigures presented in Fig. 1 demonstrate the substantial influence of external environmental factors on the appearance of algal blooms in an image. Specifically, first, with the influence of sunlight and water turbidity, it is even difficult for humans to identify whether there are HABs, as shown in Fig. 1(a) and (b). Second, there may be different surface features present in images even acquired from the same spot due to the unfixed shooting angles and focal lengths, as shown in Fig. 1(d). HABs in an image with a wide angle of view are difficult to discern because only a rough overview of the lake is acquired. Consequently, many images with large zoom factors are not qualified for HAB detection. Third, aquatic plants, shadows of trees and buildings, and HABs exhibit similar hue in insufficient sunlight, as shown in Fig. 1(c), which is also a substantial problem without key feature bands. In addition, HABs with different colors and forms may coexist due to the variation in algal species and density, as well as light

intensity, as shown in Fig. 1(f). Some of the HABs may be clustered together covering a large area, and some scatter over a particular region mixing with normal waters, as shown in Fig. 1(e) and (f). Overall, the integration and interaction of all these factors pose significant challenges to identify HABs using plain RGB images.

2.2. Details of the three-step approach

To address these issues, we proposed a novel three-step machine learning approach to decompose and simplify this complicated task, as shown in Fig. 2. First, the RGB images are classified into three categories based on the presence and status of HABs, namely, nonalgal, indeterminate, and algal blooms, to reduce interferences from external environmental factors as much as possible. The nonalgal group does not need additional attention. The indeterminate status means there is a possibility that algal blooms have occurred, but it cannot be verified due to the aforementioned reasons. The retention of this category allows for manual inspection in special cases. Then, attention is only paid to the third category, namely, images exhibiting evident HABs. Second, the water areas are extracted from the images falling into the third category, and the corresponding water masks are generated using the second model. The second step can eliminate interferences caused by objects in the lake and on the shore, be they natural or man-made. Overall, by applying the two initial steps, the environmental surroundings are eliminated, leaving only normal waters and HABs for further identification. This simplification reduces the difficulty of distinguishing them with limited spectral bands. Third, HABs can be identified with the third model in the water areas. In Fig. 2, the three steps are highlighted in bold in the model selection panel. Users must instantiate a concrete model for each step before putting them into effect; hence, appropriate models should be selected and trained in advance to ensure practical efficacy. Scene classification is technically an image classification problem, while water identification and algal bloom detection are semantic segmentation problems. In light of the current advancements in machine learning and deep learning technologies nowadays, some of the state-of-the-art deep convolutional neural networks (CNNs) in the literature and some classic machine learning (CML) models were investigated and tested in the following experiments to instantiate the processing flow.

Images are usually regarded as discrete pixels and rearranged to a two-dimensional matrix mathematically (number of pixels × number of bands) when using CML models. The internal processing of a CML model is also performed on the pixel level according to a series of elaborated rules. The positional relation of pixels is implicit in the matrix, while the CNN model directly operates on the three-dimensional tensor (height × width × bands), taking full advantage of neighborhood information. A CNN generally refers to a class of deep networks that utilize many convolution kernels to extract feature maps at different levels. As the layer becomes deeper, the receptive field of a CNN becomes larger, and the model can perceive wider regions of the input. In other words, the modeling unit of a CNN is image patch, whereas it is pixel for a CML model. The model for scene classification should estimate HAB status based on the whole image content, rather than focusing on individual pixels. Considering that the water body typically occupies a substantial area, it is more natural to identify the water areas in an object-oriented manner, as opposed to a pixel level. In addition, deep learning models have transcended CML models in many applications and are favored in both academia and industry (Voulodimos et al., 2018). Based on the above, the CNN models were chosen and explored to realize scene classification and water identification. As for the third step, since HABs may spread out in a continuous large area or scatter over water surface forming distinct small groups interleaving with normal waters, it is worth investigating both the pixel-based and object-oriented models. Accordingly, the CML and CNN models were explored for the last step. The candidate models are shown in the model selection panel of Fig. 2. Among them, Inception V3 (Szegedy et al., 2016) and MobileNet V3 (Howard et al., 2019) are the commonly-used CNN models for image classification. UNet (Ronneberger et al., 2015), DeepLab V3 (Chen et al., 2018), and MobileNet V3 with an additional segmentation head proposed by Howard et al. (2019) named LR-ASPP were selected for semantic segmentation. Meanwhile, the commonly-used CML models, including artificial neural network (ANN), random forest (RF), and extreme gradient boosting (XGB), were also explored for HAB detection in the third step.

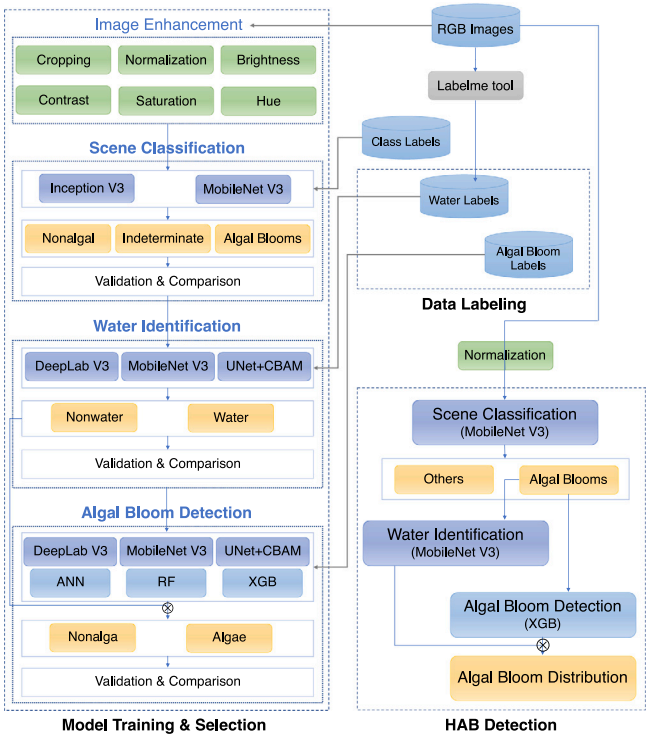


Fig. 2. A three-step machine-learning approach for HAB detection using RGB images (all listed models in each step were tested and the optimal models were employed in practice. The multiplication symbol inside a circle denotes elementwise multiplication).

2.3. Design of experiments

Inception, MobileNet, DeepLab, and UNet with an additional convolutional block attention module (CBAM) from Woo et al. (2018) were all implemented using the *PyTorch* library. The maximum features was empirically set to 960 for all the deep models. The CML models were implemented using the *scikit-learn* library. The ANN was instantiated with five hidden layers and the maximum neurons was set to 256. The key hyperparameter of RF and XGB is the number of trees, which was set to 200 in the experiment. A total of 534 images were collected from August to September 2020. Images were manually classified, resulting in 98 samples labeled as nonalgal, 93 samples as indeterminate, and 343 samples as algal blooms. Images with significant algal blooms were then manually labeled with water and nonwater using the Labelme annotation tool, as shown in Fig. 2. Next, the water areas were further labeled with HAB and nonalgal. Not all pixels were labeled because some areas were too small for labeling or too difficult to determine which label should be assigned to, as illustrated in Fig. 1(e). Image enhancement techniques were performed to increase the sample variability considering the number of labeled images was limited. In the training process of CNNs, images were dynamically cropped to the size of 512 × 512, normalized to the range between 0 and 1, and the brightness, contrast, saturation, and hue were altered randomly. The initial learning rate was set to 0.001 and the batch size was 48. Because the water and nonalgal areas were much larger than the

Table 1
Data sample count and acquisition time for model training, validation, and practical HAB detection.

Experimental phase	Steps	Sample counts	Time
Training & Validation	Scene classification	Nonalgal (98) Indeterminate (93) Algal blooms (343)	From August to September 2020
	Water identification	343	
	Algal bloom detection	343 Additional 1969 for validation	
Practical HAB detection	Including all the steps	69 180	From August to September 2021

nonwater and HAB areas regarding the samples, focal loss proposed by Lin et al. (2017) was employed for deep model training to address the imbalanced sample problem. The focal loss can be formulated as Eq. (1):

$$p_t = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{otherwise} \end{cases}$$
$$\mathcal{L}(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \tag{1}$$

where p denotes the predicted possibility of negative samples and α_t and γ are the hyperparameters to adjust the weights of imbalanced samples. They were empirically set to 0.25 and 2. The labeled image pixels were extracted as inputs to train the CML models and the undersampling technique was used to reduce the amount of samples and balance the sample classes. Overall, there were 16 030 training sample points for CML models. The machine used in the experiments was equipped with 125 G RAM, two Intel Xeon Gold CPUs, and two NVIDIA RTX GPUs. The deep models were trained with GPU acceleration while the CML models were trained on CPUs. Once the models were trained, an additional set of 1969 images were generated with all the labeled samples using the image enhancement technique for model validation. Table 1 provides an overview of the sample count and acquisition time utilized for model training, validation, and practical application of HAB detection.

The model performance was compared via visual observations and quantitative metrics. The evaluation metrics used in this paper include accuracy, precision, recall, and F1 score for scene classification (Grandini et al., 2020) and pixel accuracy, intersection over union (IoU), and F1 score for semantic segmentation (Garcia-Garcia et al., 2018). The calculation of the metrics is listed in the following equations:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{2}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

$$\text{F1} = 2 * \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Presision}} \tag{5}$$

$$\text{IoU} = \frac{|\text{A} \cap \text{B}|}{|\text{A} \cup \text{B}|} \tag{6}$$

where TP and TN denote the number of true positives and negatives, and they are all predicted correctly by a model; FP and FN denote the number of false positives and negatives, and they are all predicted incorrectly. In our study, the positive and negative classes were HABs and nonalgal. The symbols A and B in Eq. (6) represent the positive areas in the prediction and ground truth, respectively. The IoU quantifies the amount of overlap between a predicted object extent and the ground truth.

Three goals were expected to achieve in the following experiments: (1) to select the optimal model for each step; (2) to verify the contribution of each step; and (3) to demonstrate the applicability of the three-step approach. Accordingly, first, the selected CNNs were tested for the first two steps, and both the CML and CNN models were evaluated for the third step. Seventy percent of the labeled samples were used for training, and the rest were used for validation in each step. After the candidate models were determined, the necessity and effectiveness of the three steps were verified by an ablation study. Specifically, three test cases were designed, including the one-step approach with the last HAB detection being performed directly without preceding steps, the two-step approach consisting of water identification and HAB detection, and the complete three-step approach. Finally, the 69 180 images acquired from May to November 2021 were used to demonstrate the applicability of the approach. The images were randomly captured every hour from 5:00 AM to 7:00 PM in Lake Chaohu. Finally, some simple analyses were conducted based on the detection results.

Fig. 3. Typical scene classification results using MobileNet V3.

3. Results and discussion

3.1. Results

After 500 epochs of training, the CNNs were considered to have converged. The quantitative evaluation metrics for scene classification on the validation dataset are listed in Table 2. Statistically, MobileNet and Inception were quite comparable. All the scores were all greater than 0.98, but the learnable parameters of Inception were nearly six times greater than those of MobileNet, and the inference time of MobileNet was two times faster than that of Inception, so MobileNet is recommended for scene classification. Fig. 3 exhibits some examples of the three categories. The first category usually includes two scenarios where there is no HAB in water or the imaging extent is too broad to capture useful information. There might be a possibility that HABs exists in the second category, but it is unlikely to identify them correctly because of water turbidity, lights, shadows, etc. After filtering the two categories, HAB alarming can be triggers only from the third category.

The quantitative evaluation results for water identification on the validation dataset are listed in Table 3. The performance of MobileNet with LR-ASPP and DeepLab were also comparable, with all scores surpassing 0.89, but MobileNet was slightly faster than DeepLab. UNet performed the worst, albeit with all scores greater than 0.82. Moreover,

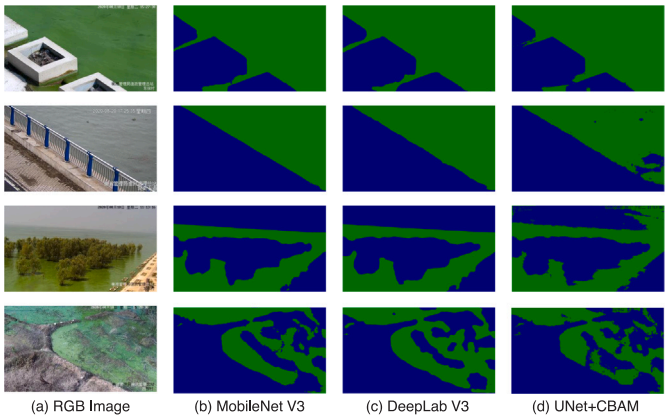


Fig. 4. Comparison of water identification results with the three deep models (the green and blue colors indicate water and nonwater). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Quantitative evaluation for scene classification on the validation dataset.

Model	Parameters	Time (s)	Accuracy	Precision	Recall	F1
MobileNet V3	4 205 875	0.13	0.98	0.99	0.98	0.99
Inception V3	24 351 718	0.27	0.98	0.98	0.99	0.98

Table 3
Quantitative evaluation for water identification on the validation dataset.

Model	Parameters	Time (s)	Accuracy	IoU	F1
LR-ASPP+MobileNet	3 218 308	0.17	0.99	0.97	0.90
DeepLab V3	11 020 594	0.21	0.99	0.96	0.89
UNet+CBAM	19 191 374	0.51	0.95	0.83	0.82

UNet possessed the highest parameter count, resulting in an inference time three times longer than that of MobileNet. Fig. 4 exhibits four cases with the RGB images and the corresponding water segmentation results. The green and blue colors indicate the water and nonwater areas, respectively. The first two rows show that all the models identified the water areas correctly for a regular shoreline. There were distinct water plants in the third row. The three models were all capable of distinguishing water plants. The images in the fourth row exhibit a complex scenario with mixed water, solid, deadwood, and algae. The models also differentiated them approximately. Generally, the results of MobileNet and Inception tended to be more natural with smooth boundaries, while the result of UNet seemed more fragmented with speckles which is often referred to as the "salt-and-pepper" effect. Considering the inference time and detection accuracy, MobileNet with LR-ASPP is recommended for water identification.

The algal bloom detection in the last step was performed only on the water bodies by masking all nonwater areas. Three CMLs and three CNNs were tested on both 30% of the original samples and the enhanced validation dataset with more much images. Table 4 presents the quantitative evaluation results. The two rows for each model correspond to the statistical results on the original small and enhanced large validation datasets. First, MobileNet with LR-ASPP showed slight superiority over other deep models regarding evaluation scores and inference time. Second, the three CML models were comparable regarding evaluation scores, whereas the XGB consumed the least time. Third, MobileNet with LR-ASPP was comparable with the CML models, with all scores greater than 0.62 on the enhanced large dataset. Fourth, the CML models produced slightly higher scores than the CNNs on the large dataset overall. Considering training a CML model was much faster than a CNN, even without hardware acceleration, the CML model is recommended for the third step.

Table 4
Quantitative evaluation for algal bloom detection on the validation datasets.

Model	Time (s)	Accuracy	IoU	F1
LR-ASPP+MobileNet	0.17	0.99 0.77	0.97 0.62	0.90 0.64
DeepLab V3	0.21	0.99 0.71	0.96 0.52	0.89 0.59
UNet+CBAM	0.51	0.95 0.71	0.83 0.51	0.82 0.58
ANN	1.43	0.97 0.77	0.62 0.64	0.61 0.72
RF	2.66	0.97 0.76	0.61 0.62	0.61 0.70
XGB	0.32	0.97 0.76	0.62 0.63	0.61 0.71

Fig. 5 demonstrates some typical HAB detection results using the six models. The first and second rows show the results in a simple environment where there was a clear-cut distinction between HABs and normal waters. All the models can identify HABs and the outcome was highly consistent with the ground truth. The third row shows a situation where water plants were mixed with algae, and all the models also produced satisfactory results. The CNNs tended to treat the water plant as a single object, resulting in its complete masking, while the results of CML models manifested a more fine-grained representation, capturing the precise shape of the plant with greater delicacy. The fourth row exhibits a situation where there are some inconspicuous algal blooms at a distance marked with a red circle. All the models detected distinct HABs nearby, but the small area with HABs in the red circle was not identified using the CNNs. In contrast, the three CML models extracted the inconspicuous HABs. The fifth row exemplifies a complex situation where small areas of normal waters were interspersed among extensive areas affected by of HABs. MobileNet and DeepLab failed to identify the elongated normal water area marked with a yellow circle, while UNet and the other three CML models detected it accurately. The last row shows a situation where the algae accumulated both on the surface and under the water, with all models performing well in this context. Generally, all six models detected large-scale HABs, but the results of CML models showed more fineness in some complex environments.

Moreover, an ablation study was conducted to verify the necessity and effectiveness of the proposed three steps. Taking MobileNet with RL-ASPP and XGB as examples of CNN and CML models, Table 5 showcases the progressive impact on evaluation scores as individual steps are systematically removed. The first row presents the outcomes obtained from each model using the comprehensive three-step approach. The second row illustrates the results after eliminating the first step, which entails the absence of scene classification. The third row depicts the outcomes achieved solely by employing the third step, specifically the direct identification of HABs. As Table 5 shows, the scores decreased significantly for both models by unloading one step each time, thus confirming that each step considerably contributed to the final accuracy. Some typical examples are illustrated in Fig. 6. The column name with a numbered suffix indicates the total steps employed in HAB detection. The first row displays an image with a wide view angle, in which misidentification can occur in the absence of the preceding two steps. A scenario featuring solely turbid water is presented in the second row. The CNN model correctly distinguished turbid waters and algal blooms, but the CML model erroneously classified some turbid waters as HABs. The third row illustrates a scenario characterized by turbid waters and tree shadows. A similar conclusion can be drawn from this row. The last three rows show a situation where green plants were both in water or on land. The CNN model can differentiate between green plants and HABs without the first two steps, while this cannot be realized with the CML model unless water identification is performed in advance.

Based on the results of the above experiments, MobileNet was chosen for scene classification and water identification, and XGB was

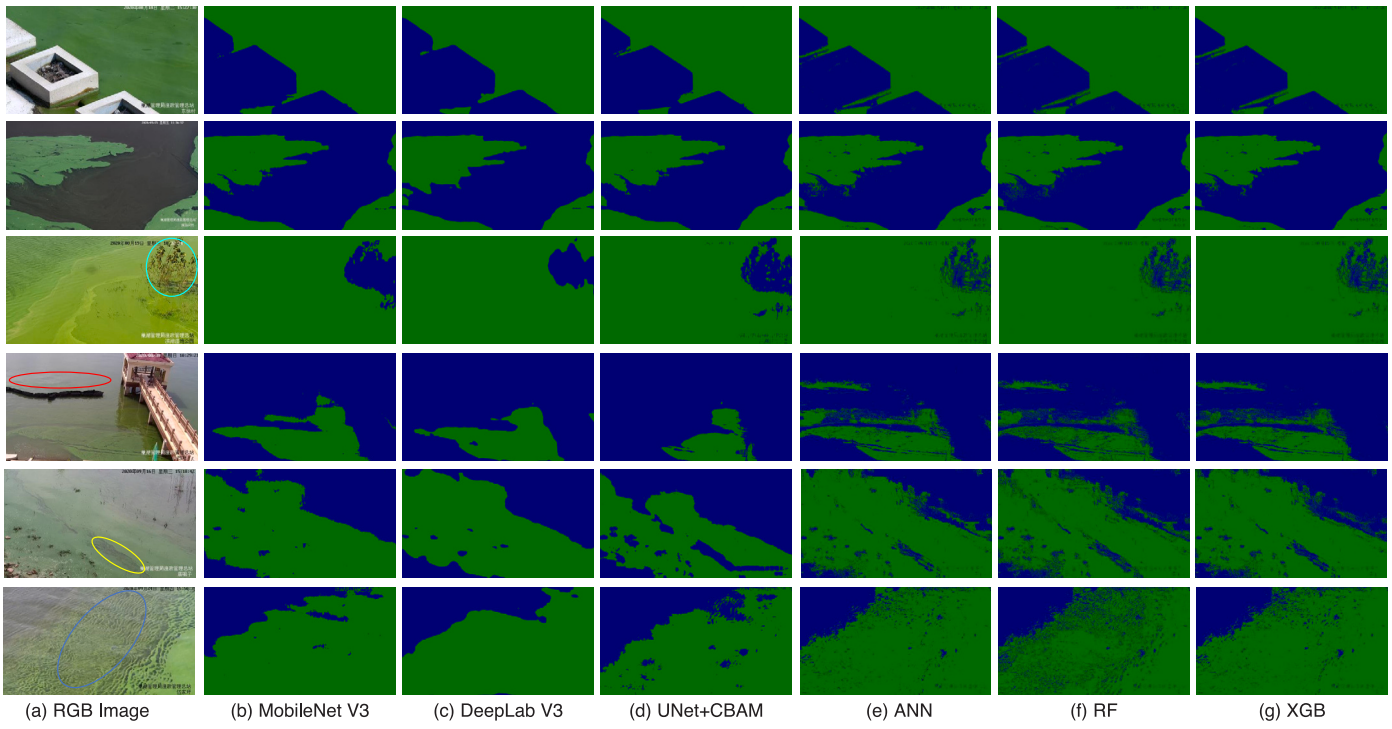


Fig. 5. Comparison of algal bloom detection results with the three CNNs and three CMLs (the green and blue colors indicate algal blooms and nonalgal). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

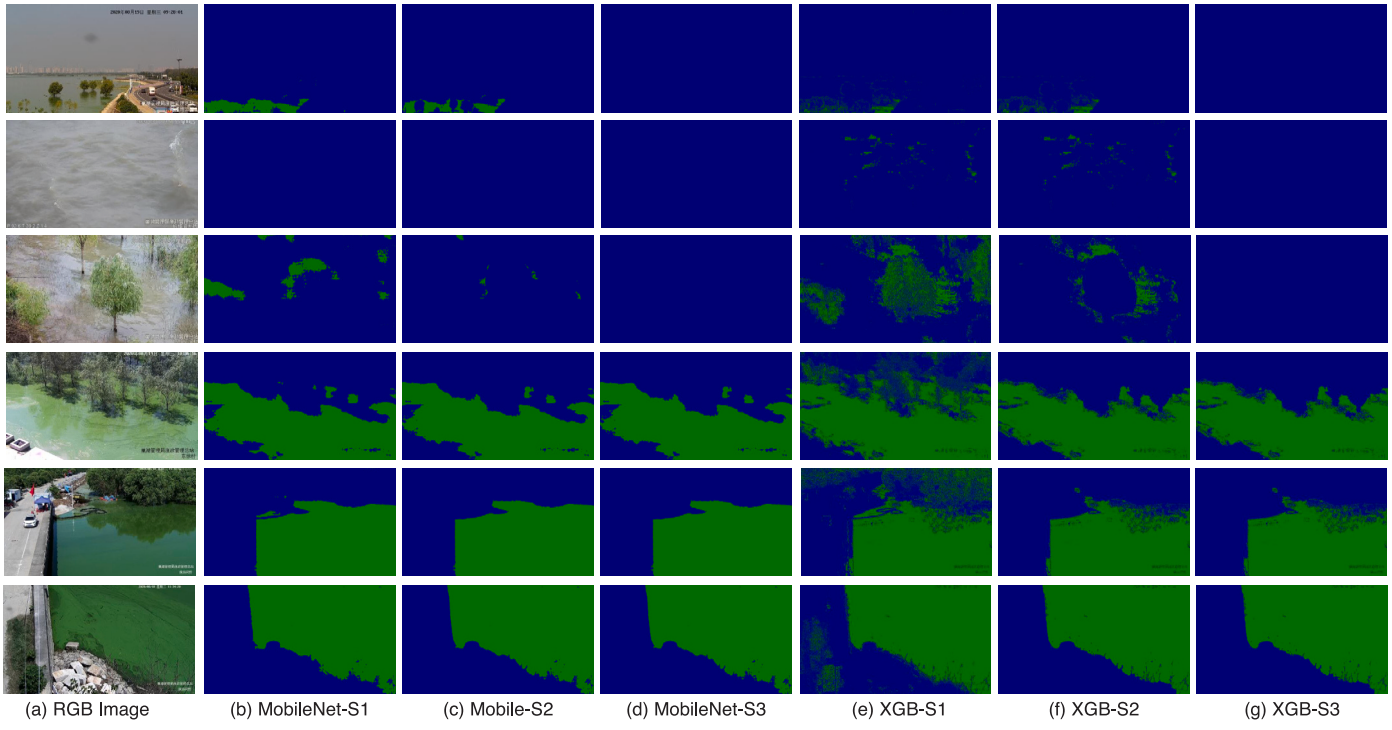


Fig. 6. Results of the ablation study by progressively removing steps from the three-step approach (the column name with a numbered suffix indicates the cumulative number of employed steps). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

chosen for the last HAB detection. The concrete models were then applied in practical HAB detection in Lake Chaohu. Taking 2021 as an example, Fig. 7 shows the statistics of HAB cover ratio along Lake Chaohu. Fig. 7(a) illustrates the mean cover ratio of all stations from April to November. Because camera poses varied constantly, a great many images with a wide field of view could not be used for detailed HAB detection. Consequently, the statistical cover ratio was relatively low, as shown in 7(a). The most serious outbreak of HABs occurred in September, and the corresponding daily mean cover ratio is presented in Fig. 7(b). For the same reason, many values were quite low in

Table 5
Evaluation for the ablation study by progressively removing steps from the whole process.

Model	Time (s)	Accuracy	IoU	F1
LR-ASPP+MobileNet	1.21	0.81	0.74	0.74
	1.19	0.79	0.65	0.68
	0.64	0.77	0.62	0.64
XGB	1.24	0.83	0.76	0.76
	1.17	0.81	0.70	0.72
	0.32	0.76	0.63	0.71

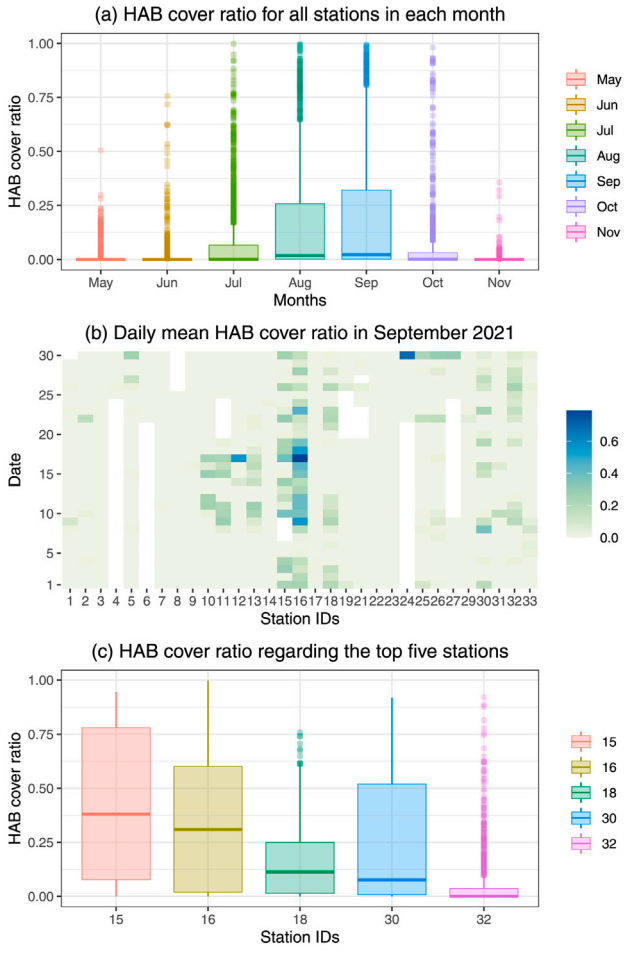


Fig. 7. Statistics of HAB cover ratio in 2021 along Lake Chaohu (a shows the cover ratio for all stations from April to November 2021; b shows the daily mean cover ratio in September; c shows the cover ratio regarding the top five stations).

Fig. 7(b). The hourly cover ratio of the top five stations is illustrated in Fig. 7(c). The top three are all located in northwestern Lake Chaohu, and the mean cover ratio of the top two is far beyond that of the other.

3.2. Discussion

In this paper, we proposed a novel three-step approach for HAB detection using plain RGB images. Different CNN and CML models were exploratively tested for each step. Generally, the CNN can make decisions based on global information with deep layers (Luo et al., 2016), while the CML model made decisions pixel-by-pixel, and the results, of course, were completely pixel-based, revealing many more details of ground features. So the CNN is recommended for the first two steps and the CML model is recommended for the last step. An ablation study was conducted and the results from Table 5 and Fig. 6 revealed that each step served different purposes and truly contributed to the

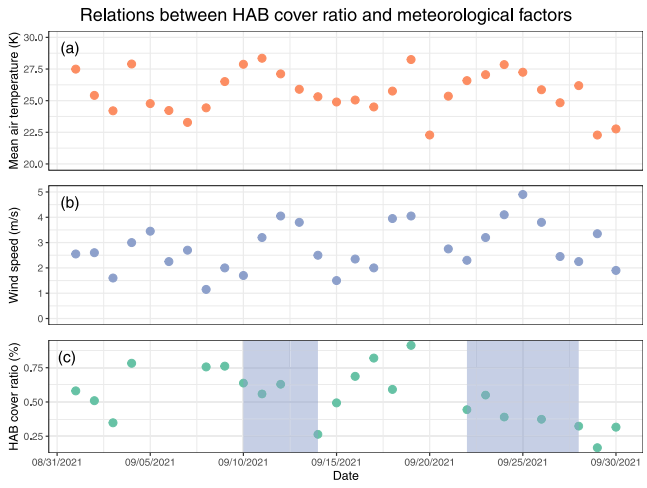


Fig. 8. Relations between HAB cover ratio and meteorological factors at Station 23 in September 2021 (the highlighted regions in purple-blue indicate that the HAB cover ratio is more affected by the wind). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

final result. Scene classification eliminated interferences from various camera poses and external environments, such as sunlight and turbid water, to minimize false detection. Most images without significant HABs were excluded in this step. Water identification was mainly used to remove potential disruptions posed by surface objects both on land and in water, especially green plants. It could ease the difficulties in later HAB detection by preemptively masking water plants, as shown in Figs. 4 and 6. This is rather important for CML models because the CML model makes predictions completely on the RGB values of pixels. The CNNs, in contrast, can make use of the high-level shape, texture, and structure features to conduct inference, so it is capable of distinguishing the water plants and HABs, as shown in Fig. 6. Hence, in the third step, the CNN exhibited lower dependency on the second step than the CML model did, but the fineness of the results was not as good as that of the CML model.

The application in Lake Chaohu demonstrated the feasibility of the proposed approach. However, because cameras were mainly installed along the bank, camera-based monitoring cannot present the whole picture of a lake. Moreover, the cameras initially served for illegal fishing monitoring, and the HAB monitoring center could only receive one image per hour from each station with randomly changed PTZ parameters. Many acquired images were not available for use due to the wide angle of views. If a more accurate HAB distribution and cover rate are required, the camera can be set to capture images in a specific pose at a specific time and cover the whole edge of the lake during the daytime. The statistics showed that the most frequent outbreaks of HABs occurred in August and September, and the northwest section of the lake was the most severe. This conclusion is consistent with previous studies (Zhang et al., 2015; Ma et al., 2021) using remote sensing. In addition, the correlation between HABs and meteorological factors was also explored. Fig. 8 depicts the variations in the mean air temperature, wind speed, and HAB cover ratio at Station 23 in September 2021. Generally, the temporal fluctuation of HAB cover ratio exhibited a concordant pattern with variations in air temperature, as evidenced by a Pearson correlation coefficient (r) of 0.46 and a corresponding p-value of 0.03. The impact of wind on the HAB cover ratio was also observed, although the overall correlation demonstrated minimal strength ($r = -0.01$, p-value $= 0.94$). This is because the HAB cover ratio displayed a heightened sensitivity only to wind in regions delineated by shades of purple-blue where the wind speed was relatively high, but the influence was notably diminished under conditions of weak wind. This phenomenon also supports the conclusion derived

from remote sensing-based monitoring in Lake Chaohu (Zhang et al., 2015, 2021).

Currently, satellite-based monitoring has become a prevailing practice for obtaining comprehensive information regarding HAB conditions in lakes. Advanced inverse models can be even developed to estimate chlorophyll-a or microcystin, enabling quantitative detection of HABs. In this context, the proposed camera-based monitoring approach holds the potential to function as an effective early warning tool, facilitating prompt intervention and control measures against HABs. However, it is crucial to acknowledge that the camera-based approach is limited to capturing visually conspicuous aggregations of HABs. Despite concerted efforts to mitigate the influence of external environmental factors and camera poses through the implementation of the three-step approach, it remains important to recognize that these factors continue to impact the detection process. Furthermore, UVA-based monitoring can be employed in emergency scenarios to map the distribution of HABs in specific areas, albeit necessitating manual control and being constrained by limited cruising time. Each of these approaches possesses its own inherent advantages and disadvantages, and their integration can culminate in a complementary system capable of achieving comprehensive coverage and automated monitoring of HABs in lakes.

4. Conclusion

The application of surveillance cameras provides an economical way to perform real-time monitoring of HABs in lakes. Nevertheless, the limited spectral bands and complex environment pose two challenges when using plain RGB images. In this paper, we proposed a three-step machine-learning-based approach to decompose this complicated task and offered a preliminary solution. First, the collected images were classified into three categories, namely, nonalgal, indeterminate, and algal blooms. Only images with certain algal blooms were considered for HAB detection. Next, the second model for the water identification model was built to distinguish water and nonwater. The second step removed interferences from the objects in nonwater areas. Third, the last model was employed for HAB detection in solely water areas.

The experiments demonstrated that the three-step approach can derive satisfactory HAB detection results. MobileNet and XGB were chosen for the preceding two steps and the last step. The overall pixel accuracy, IoU, and F1 score reached 0.83, 0.76, and 0.76, respectively, on the enhanced dataset. Compared with the one-step model, the three-step approach considerably improved the detection accuracy and reduce the misdetection. The proposed approach functioned quite well in a simple environment. There may be some commissions or omissions in the complex environment occasionally, but most HABs can still be correctly detected even they are affected by lights, shadows, camera poses, etc. Hence, we believe that the approach presented in this paper can supplement the remote sensing approach and it can satisfy the demand for automatic early warning of HABs. In the future, multispectral or multiview cameras could be introduced to further improve the detection accuracy.

CRediT authorship contribution statement

Zhenyu Tan: Conceptualization, Methodology, Implementation, Formal analysis, Writing – original draft. **Chen Yang:** Data curation, Experiment, Validation. **Yinguo Qiu:** Data curation, Investigation. **Wei Jia:** Data curation, Experiment. **Chenxi Gao:** Data curation. **Hongtao Duan:** Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

This research was partly supported by the National Natural Science Foundation of China under Grant 42101342, Grant 41971309, and Grant U2243205; and partly by the Education Department of Shaanxi Province, China under Grant 21JK0928.

Appendix A. Details on model training

The following provides a comprehensive elucidation on the methodologies employed for the preparation of training samples and the training of models to instantiate the procedures for the HAB detection:

(1) To prepare the training samples:
At first, the acquired RGB images need to be classified manually into three categories: nonalgal, indeterminate, and algal blooms. Subsequently, the images categorized as algal blooms should be labeled using the Labelme tool. This process involves manually marking the pixel areas in the image as either water or nonwater. Further, the pixels within the water areas should be labeled as either algal blooms or normal waters. It is important to note that not all pixels are labeled in this step, as some may be difficult to determine. Attention should be given to pixels exhibiting distinctive features. Typically, a split of 70% of the data is allocated for model training, while the remaining is reserved for model validation.

(2) To train the model for scene classification:
The MobileNet is recommended for scene classification. The image samples consisting of the three classes should be fed into the model, and undergo multiple iterations of training. To enhance the diversity of samples, the brightness, contrast, saturation, and hue of the images are randomly adjusted in the training. The images should be dynamically cropped to a uniform size and normalized within the range of 0 to 1 following the image enhancement. Focal loss is recommended for this and following steps due to the imbalanced classes. Stochastic Gradient Descent (SGD) or Adam optimizer can be used to optimize the model parameters. The initial learning rate is set to 0.001. The batch size can set according to the hardware capacity. Once the model's loss and accuracy have reached convergence, the model weights can be frozen and utilized for actual prediction.

(3) To train the model for water identification:
Similarly, it is advisable to employ the MobileNet with LR-ASPP segmentation head for water identification. The input samples for training are the labeled images with water and nonwater. Image enhancement and normalization is also required. The training process for this task bears resemblance to the initial step, as the hyperparameters for training can be set in a similar manner as those used in the preceding stage. However, it is important to note that this training process is entirely independent and distinct from the aforementioned stage.

(4) To train the model for algal bloom detection:
The training process for algal bloom detection is distinctly separate from the preceding two training processes, while sharing similar training procedures. The samples comprised of algal blooms or normal waters are entered into the third model with the dynamic image enhancement. The hyperparameters governing this training process can be established in a manner akin to the previous steps. Once the models are trained completely, the three models can be utilized jointly to perform the practical HAB detection.

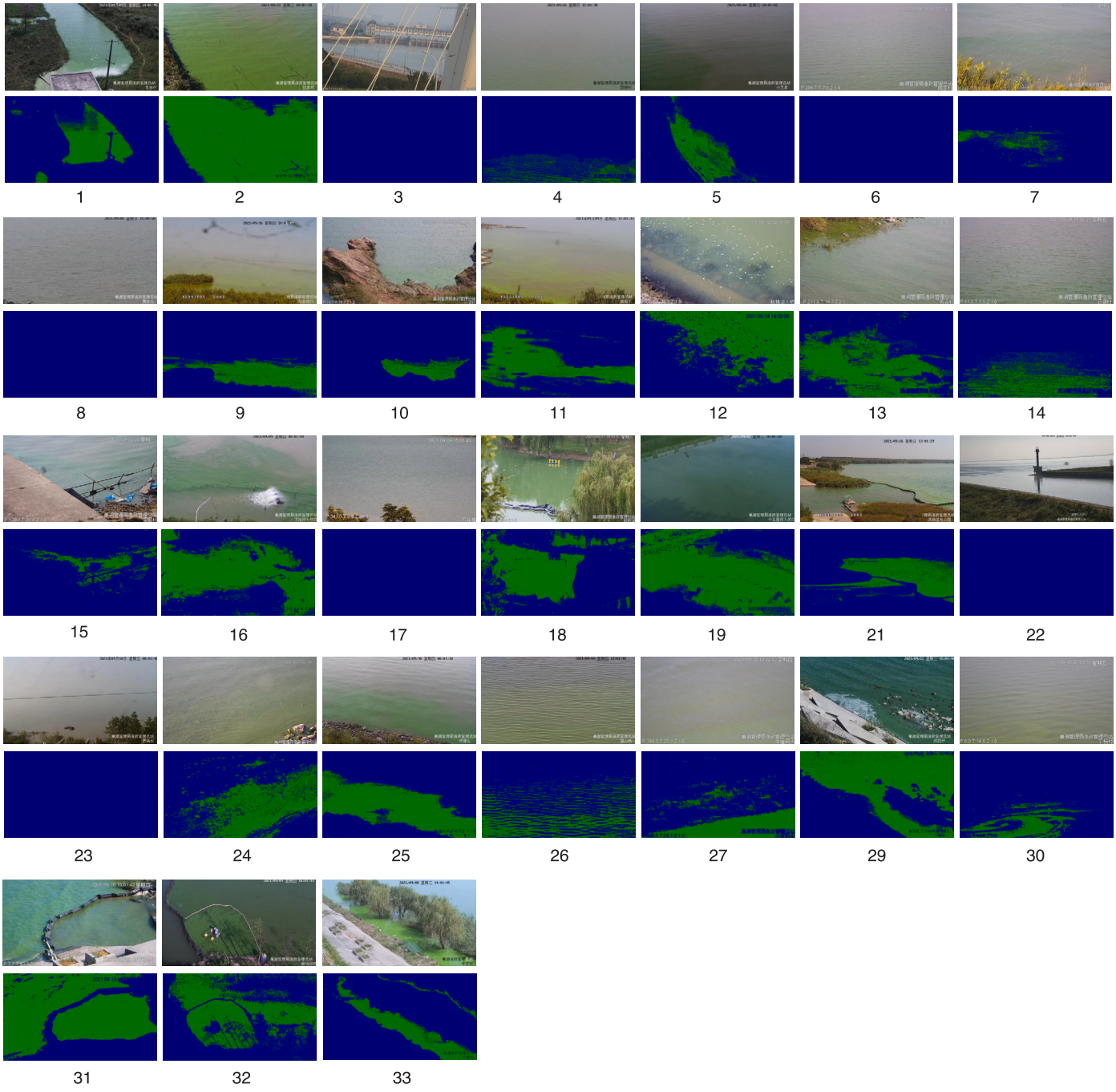


Fig. B.1. Typical examples of HAB detection for each camera station along Lake Chaohu (the number denotes the station ID). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Appendix B. Examples of the application

Typical HAB detection results for each camera station are shown in Fig. B.1. The number below each image denotes the station ID. A rich variety of scenarios are included here. For example, HABs cannot be observed in the images of stations 3, 22, and 23 with a large angle of view. Images of stations 4 and 6 are all filled with wide open water, but one is detected with HABs, while the other is not because of the differences in color shades. Ground objects with similar colors to HABs can be correctly identified, such as in stations 1, 9, and 11. HABs with different color tones can also be identified correctly, such as in stations 11, 19, and 27.

Fig. B.2 presents some typical HAB detection results for station 23 at different times. The annotation below each image denotes the shooting time in the format of month/date: hour. Overall, serious HABs can be observed at station 23, and Fig. B.2 exhibits satisfactory results. However, HAB detection can be influenced by many factors, and there are also limitations. For example, the detected areas of HABs for the same spot are slightly different at 11:00 and 15:00 on August 4th due to the variation in sunlight. The area where the fishing boats just pass by is detected as an algal bloom at 16:00 on September 10th, while it can be correctly identified as normal water at 16:00 on September 14th due to the differences in image contrast. Generally, the three-step approach functions quite well in most cases, but there is inevitable misdetection in some situations caused by image tuning and contrast.

Fig. B.2. Typical examples of HAB detection for station 23 along Lake Chaohu with the maximum HAB cover ratio at different times (the annotation below each image denotes the shooting time).

References

Binding, C., Greenberg, T., Bukata, R., 2013. The MERIS maximum chlorophyll index; its merits and limitations for inland water algal bloom monitoring. J. Great Lakes Res. 39, 100–107. http://dx.doi.org/10.1016/j.jglr.2013.04.005.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2018. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. 40 (4), 834–848. http://dx.doi.org/10.1109/tpami.2017.2699184.

Duan, H., Tao, M., Loiselle, S.A., Zhao, W., Cao, Z., Ma, R., Tang, X., 2017. MODIS observations of cyanobacterial risks in a eutrophic lake: Implications for long-term safety evaluation in drinking-water source. Water Res. 122, 455–470. http://dx.doi.org/10.1016/j.watres.2017.06.022.

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Martinez-Gonzalez, P., Garcia-Rodriguez, J., 2018. A survey on deep learning techniques for image and video semantic segmentation. Appl. Soft Comput. 70, 41–65. http://dx.doi.org/10.1016/j.asoc.2018.05.018.

Gholizadeh, M., Melesse, A., Reddi, L., 2016. A comprehensive review on water quality parameters estimation using remote sensing techniques. Sensors-Basel 16 (8), 1298. http://dx.doi.org/10.3390/s16081298.

Grandini, M., Bagli, E., Visani, G., 2020. Metrics for multi-class classification: An overview. http://dx.doi.org/10.48550/arXiv.2008.05756, arXiv eprints.

Hou, X., Feng, L., Dai, Y., Hu, C., Gibson, L., Tang, J., Lee, Z., Wang, Y., Cai, X., Liu, J., Zheng, Y., Zheng, C., 2022. Global mapping reveals increase in lacustrine algal blooms over the past decade. Nat. Geosci. 15 (2), 130–134. http://dx.doi.org/10.1038/s41561-021-00887-x.

Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L.-C., Tan, M., Chu, G., Vasudevan, V., Zhu, Y., Pang, R., Adam, H., Le, Q., 2019. Searching for MobileNetV3.

In: 2019 IEEE/CVF International Conference on Computer Vision. ICCV, IEEE, pp. 1314–1324. http://dx.doi.org/10.1109/iccv.2019.00140.

Hu, C., 2009. A novel ocean color index to detect floating algae in the global oceans. Remote Sens. Environ. 113 (10), 2118–2129. http://dx.doi.org/10.1016/j.rse.2009.05.012.

Huang, C., Li, Y., Yang, H., Sun, D., Yu, Z., Zhang, Z., Chen, X., Xu, L., 2013. Detection of algal bloom and factors influencing its formation in taihu lake from 2000 to 2011 by MODIS. Environ. Earth Sci. 71 (8), 3705–3714. http://dx.doi.org/10.1007/s12665-013-2764-6.

Janssen, A.B., Janse, J.H., Beusen, A.H., Chang, M., Harrison, J.A., Huttunen, I., Kong, X., Rost, J., Teurlincx, S., Troost, T.A., van Wijk, D., Mooij, W.M., 2019. How to model algal blooms in any lake on earth. Curr. Opin. Green Sustain. Chem. 36, 1–10. http://dx.doi.org/10.1016/j.cosust.2018.09.001.

Khan, R.M., Salehi, B., Mahdianpari, M., Mohammadimanesh, F., Mountrakis, G., Quackenbush, L.J., 2021. A meta-analysis on harmful algal bloom (HAB) detection and monitoring: A remote sensing perspective. Remote Sens. 13 (21), 4347. http://dx.doi.org/10.3390/rs13214347.

Kislik, C., Dronova, I., Kelly, M., 2018. UAVs in support of algal bloom research: A review of current applications and future opportunities. Drones 2 (4), 35. http://dx.doi.org/10.3390/drones2040035.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollar, P., 2017. Focal loss for dense object detection. In: 2017 IEEE International Conference on Computer Vision. ICCV, IEEE, pp. 2980–2988. http://dx.doi.org/10.1109/iccv.2017.324.

Luo, W., Li, Y., Urtasun, R., Zemel, R., 2016. Understanding the effective receptive field in deep convolutional neural networks. In: Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Vol. 29. Curran Associates, Inc., http://dx.doi.org/10.5555/3157382.3157645.

Ma, J., Jin, S., Li, J., He, Y., Shang, W., 2021. Spatio-temporal variations and driving forces of harmful algal blooms in chaohu lake: A multi-source remote sensing approach. Remote Sens. 13 (3), 427. http://dx.doi.org/10.3390/rs13030427.

Oyama, Y., Fukushima, T., Matsushita, B., Matsuzaki, H., Kamiya, K., Kobinata, H., 2015. Monitoring levels of cyanobacterial blooms using the visual cyanobacteria index (VCI) and floating algae index (FAI). Int. J. Appl. Earth Obs. Geoinf. 38, 335–348. http://dx.doi.org/10.1016/j.jag.2015.02.002.

Page, B.P., Kumar, A., Mishra, D.R., 2018. A novel cross-satellite based assessment of the spatio-temporal development of a cyanobacterial harmful algal bloom. Int. J. Appl. Earth Obs. Geoinf. 66, 69–81. http://dx.doi.org/10.1016/j.jag.2017.11.003.

Pedrayes, O.D., Lema, D.G., Usamentiaga, R., García, D.F., 2022. Detection and localization of fugitive emissions in industrial plants using surveillance cameras. Comput. Ind. 142, 103731. http://dx.doi.org/10.1016/j.compind.2022.103731.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing, Cham, pp. 234–241. http://dx.doi.org/10.1007/978-3-319-24574-4_28.

Samantaray, A., Yang, B., Dietz, J.E., Min, B.-C., 2018. Algae detection using computer vision and deep learning. http://dx.doi.org/10.48550/arXiv.1811.10847, arXiv eprints.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, IEEE, pp. 2818–2826. http://dx.doi.org/10.1109/cvpr.2016.308.

Tan, Z., Gao, M., Yuan, J., Jiang, L., Duan, H., 2022. A robust model for MODIS and landsat image fusion considering input noise. IEEE Trans. Geosci. Remote Sens. 60, 1–17. http://dx.doi.org/10.1109/tgrs.2022.3145086.

Voulodimos, A., Doulamis, N., Doulamis, A., Protopapadakis, E., 2018. Deep learning for computer vision: A brief review. Comput. Intell. Neurosci. 2018, 1–13. http://dx.doi.org/10.1155/2018/7068349.

Wang, X., Wang, M., Liu, X., Zhu, L., Glade, T., Chen, M., Zhao, W., Xie, Y., 2022. A novel quality control model of rainfall estimation with videos – a survey based on multi-surveillance cameras. J. Hydrol. 605, 127312. http://dx.doi.org/10.1016/j.jhydrol.2021.127312.

Wang, Z., Zhao, Z., Li, D., Cui, L., Liu, X., 2015. Monitoring algal blooms using active learning camera sensor networks. Int. J. Sens. Netw. 19 (2), 91. http://dx.doi.org/10.1504/ijsnet.2015.071633.

Woo, S., Park, J., Lee, J.-Y., Kweon, I.S., 2018. CBAM: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 3–19. http://dx.doi.org/10.48550/arXiv.1807.06521.

Wu, D., Li, R., Zhang, F., Liu, J., 2019. A review on drone-based harmful algae blooms monitoring. Environ. Monit. Assess. 191 (4), 211. http://dx.doi.org/10.1007/s10661-019-7365-8.

Zhang, Y., Hu, M., Shi, K., Zhang, M., Han, T., Lai, L., Zhan, P., 2021. Sensitivity of phytoplankton to climatic factors in a large shallow lake revealed by column-integrated algal biomass from long-term satellite observations. Water Res. 207, 117786. http://dx.doi.org/10.1016/j.watres.2021.117786.

Zhang, Y., Ma, R., Zhang, M., Duan, H., Loiselle, S., Xu, J., 2015. Fourteen-year record (2000–2013) of the spatial and temporal dynamics of floating algae blooms in lake chaohu, observed from time series of MODIS images. Remote Sens. 7 (8), 10523–10542. http://dx.doi.org/10.3390/rs70810523.