

基于视频监控图像的环巢湖蓝藻实时动态监测*

马腾耀¹, 肖鹏峰^{1**}, 张学良¹, 段洪涛², 邱银国²

(1: 南京大学地理与海洋科学学院, 自然资源部国土卫星遥感应用重点实验室, 江苏省地理信息技术重点实验室, 南京 210023)

(2: 中国科学院南京地理与湖泊研究所, 中国科学院流域地理学重点实验室, 南京 210008)

摘要: 蓝藻的防控与治理是湖泊水环境、水生态管理的重要内容, 实时获取蓝藻的空间分布信息对于降低蓝藻灾害风险具有重要意义。针对地面调查费时费力、卫星遥感监测粒度较粗且时效性不强等问题, 本文提出了一种基于视频监控网络的湖泊蓝藻实时监测技术。基于环巢湖视频监控网络的 33 个功能摄像机, 研究如何从视频图像中实时、准确提取蓝藻的分布信息。为克服不同摄像头的观测角度不一致、光照强度和背景条件不一致等诸多挑战, 在视频图像蓝藻表征分析的基础上, 通过多尺度深度网络进行图像粗粒度分类, 区分蓝藻与浑浊、阴影水体; 基于随机森林进行蓝藻精细化识别, 克服蓝藻的强异质性。最后以渔政站沿岸水体的日均蓝藻覆盖率和月均蓝藻覆盖率为统计单位, 开展了巢湖沿岸蓝藻的动态监测。研究成果可为科学制定蓝藻治理方案提供技术支撑。

关键词: 巢湖; 视频图像蓝藻识别; 多尺度卷积神经网络; 随机森林; 蓝藻动态监测

Real-time monitoring of cyanobacterial blooms dynamics around Lake Chaohu based on video surveillance images*

Ma Tengyao¹, Xiao Pengfeng^{1**}, Zhang Xueliang¹, Duan Hongtao² & Qiu Yinguo²

(1: *Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Key Laboratory for Land Satellite Remote Sensing Applications of Ministry of Natural Resources, School of Geography and Ocean Science, Nanjing University, Nanjing 210023, P.R.China*)

(2: *Key Laboratory of Watershed Geographic Sciences, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, P.R.China*)

Abstract: The prevention and management of cyanobacterial blooms is an important part of lake environment and water ecological management. Real-time acquisition of spatial distribution of cyanobacterial blooms is of great significance for timely salvage and disaster reduction. Aiming at the problems of time-consuming and laborious ground surveys, satellite remote sensing monitoring with a low spatial and temporal resolution, a new method for real-time monitoring of cyanobacterial blooms in lakes using video surveillance network (VSN) was proposed. Based on the 33 cameras of VSN around Lake Chaohu, the study focuses on the real-time and accurate extraction of cyanobacteria distribution information from video images. First, in order to overcome the challenges of different observation angles from different cameras, light intensities and background conditions, the representation of cyanobacterial blooms in video images was analyzed. Then, a multi-scale depth network was used for coarse-grained image classification to distinguish cyanobacteria from turbid and shadowed water; Random forest method was used to finely recognize cyanobacterial blooms to overcome the strong heterogeneity of cyanobacteria. Finally, the distribution information of cyanobacterial blooms was acquired. Based on the average daily and monthly cyanobacteria coverage of the coastal waters of the fishery administration station, monitoring of cyanobacterial blooms dynamics along the Lake Chaohu coast was completed, which can provide technical support for the management of cyanobacterial blooms.

Keywords: Lake Chaohu; cyanobacteria recognition based on video images; multi-scale convolutional neural network; random for-

* 2021-09-16 收稿; 2022-02-19 收修改稿。

安徽省巢湖管理局项目“巢湖蓝藻水华监测预警与模拟分析平台”(2019AMCZ4622)资助。

** 通信作者; E-mail: xiaopf@nju.edu.cn.

est; monitoring of cyanobacterial blooms dynamics

湖泊和水库作为中国大部分地区的主要饮用水源,许多处于富营养化状态,存在蓝藻水华暴发风险^[1-2]. 据《2020 年中国生态环境状况公报》显示,110 个重点监测湖泊中,6 个呈重度或中度富营养化、26 个呈轻度富营养化^[3]. 一方面,蓝藻暴发会导致水体含氧量减少,不仅使水生植物窒息,还使鱼类等其他生物的多样性大大降低^[4]. 另一方面,当水体中包含藻类毒素时,将威胁当地居民的饮水安全;此外在蓝藻消亡时会散发恶臭,对生态环境造成了恶劣影响^[5-6]. 鉴于蓝藻对人类的生产生活影响巨大,如果能对蓝藻的时空变化进行动态监测,实现在暴发前有所预防、暴发时及时打捞、暴发后精准评估,可有效减轻蓝藻的危害.

目前常用的监测方法有基于站点的蓝藻监测^[7-8]、基于遥感的蓝藻识别^[9-10]和反演^[9,11]. 基于站点的蓝藻监测,虽能准确获取诸多水质参数,但耗时耗力,难以满足大范围蓝藻变化监测的需求^[12]. 基于遥感的蓝藻识别是利用数字图像处理方法,从遥感图像中提取蓝藻范围的过程. 蓝藻识别方法有:近红外波段阈值法^[13];近红外波段与红波段的差值法或比值法^[13-14];归一化植被指数(*NDVI*)、增强型植被指数(*EVI*)及浮游藻类指数(*FAI*)等指数类方法^[15-16];区域生长、大津算法等图像分割方法^[17];支持向量机、随机森林等图像分类方法^[18]. 基于遥感的蓝藻反演是利用遥感和实测数据,通过构建蓝藻参数与遥感反射率的经验或半经验模型^[19-20]或生物光学模型^[21],从遥感图像中获取蓝藻参数. 其中,叶绿素 *a* 和藻蓝素是最常用的蓝藻反演参数^[21-22].

蓝藻的变化极为迅速,这对于高时间和高空间分辨率难以兼顾的卫星传感器来说,不能满足蓝藻的变化监测需求^[23]. 目前可通过水下探针传感器或视频监控实现湖泊水环境的实时监测^[24]. 尤其是湖泊的视频监控网络能够提供实时的视频图像,可以较高频率监测蓝藻的时空动态变化^[25]. 研发的基于摄像机的水环境高光谱实时监测与反演系统,推动了蓝藻变化监测技术的不断进步^[26-27]. 本文以环湖视频监控网络为数据源,发展一种基于视频图像的自动化蓝藻识别和动态监测方法. 研究成果将为湖泊蓝藻监测提供一种实时、连续、动态的技术方案,作为“天-空-地”协同观测体系的地面监测环节,能够及时掌握湖泊沿岸蓝藻水华的空间分布和动态变化,实现对重点区域的蓝藻监测^[28].

1 研究区与数据

1.1 研究区

巢湖(31°25′~31°43′N, 117°16′~117°50′E)位于安徽省中部,北接肥东县,西邻合肥市和肥西县,东至巢湖市,南靠庐江县(图 1). 巢湖平均水深 2.89 m,水面面积约 785.4 km²^[29]. 巢湖水系发源于大别山区东麓和浮槎山区东南麓,共有大小河流 35 条从四周呈放射状注入巢湖,其中来自西、北、南三面的河流量占比在 90% 以上,最后经东部唯一出口河流——裕溪河并入长江干流^[30].

巢湖流域属于北亚热带湿润季风气候. 冬季偏北风来自高纬大陆,空气干燥寒冷;夏季东南风来自东南沿海,温度高且湿度大. 多年平均气温为 16℃,平均相对湿度为 76%,平均年降水量为 1178.3 mm. 流域降水主要集中在夏、秋季节,其中 6、7 月的降水量占年降水量的 30%~40%^[31].

随着工业化、城镇化的发展,氮磷污染物排放量不断增加,巢湖水体的富营养化严重. 每年由南淝河、十五里河和派河输入巢湖的污染物占全湖 60% 以上,导致湖体营养盐超标,存在蓝藻暴发风险^[32]. 此外,夏季温度高适宜浮游藻类生长,在东南沿海暖湿气流的支配下,巢湖蓝藻在空间分布上呈现由东南向西北逐渐增加的趋势^[30].

1.2 数据资料

本文使用的视频图像源于环巢湖视频监控网络的 33 个功能摄像机. 摄像机均匀分布于环巢湖沿岸,而且大部分位于河流入湖口等重要水域(图 1),具有较强的空间代表性. 摄像机可变焦距,通过调整 P(Pan, 水平)、T(Tilt, 俯仰)、Z(Zoom, 缩放)3 个姿态参数,能实现摄像机对周边水域的巡视. 为便于蓝藻监测,设置每台摄像机在白天每隔 1 小时自动调整至指定姿态参数,使视角内仅含蓝藻和水体,避免无关要素的干扰. 单台摄像机覆盖面积约为 800~1000 m². 视频图像采集过程中充分考虑光照强度、天气状态、蓝藻形态等影响因素,在各摄像机中收集蓝藻暴发时的图像,尽量使获取的蓝藻图像更全面. 分别在上午(8:00—

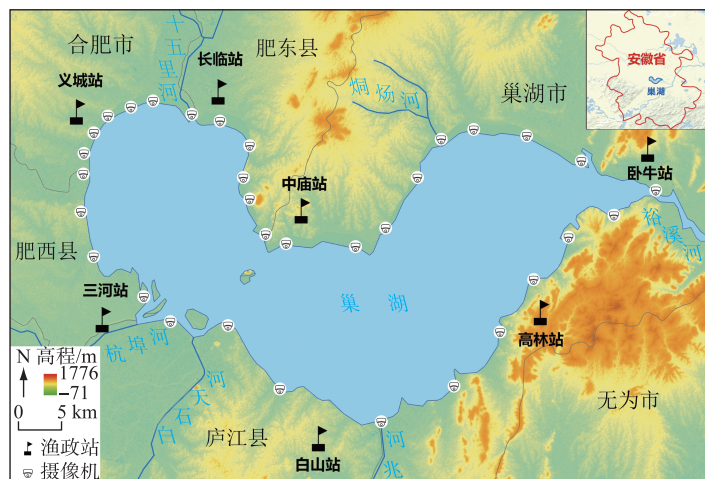


图1 研究区位置图

Fig.1 Location of the study area

11:00)、中午(11:00—14:00)、下午(14:00—17:00)3个时段内进行图像采集,共获得2020年8—10月图像679张。其中上午图像260张,中午图像113张,下午图像306张。蓝藻暴发较频繁的摄像点位获取图像较多,每处摄像机获取图像在10~50张;剩余摄像点位蓝藻水华发生频率低,每处摄像机获取的图像在10张以下。由于采集图像的软件不能并行操作,同一摄像机每天获取图像的时间并不一致。单帧图像的大小为1920列×1080行,图像由红、绿、蓝3波段组成。

巢湖管理局下设7个渔政站,自巢湖市城南的卧牛站开始,沿顺时针方向分布有高林站、白山站、三河站、义城站、长临站、中庙站,33个摄像机由各渔政站管理,每一渔政站覆盖4~7处摄像机(图2)。以渔政站作为统计单元,对其辖区水域的蓝藻识别结果进行动态分析,可实现巢湖沿岸蓝藻水华的动态监测。



图2 巢湖沿岸的部分监控摄像机

Fig.2 Some cameras in the shoreside of Lake Chaohu

2 视频图像的蓝藻识别和监测方法

2.1 视频图像表征分析

不同摄像机的观测角度不同、所处的光照强度和成像条件也不同,在复杂外部环境的影响下,视频图像表现出蓝藻与阴影水体、浑浊水体难区分和蓝藻的异质性高等诸多问题。由于这些复杂成像条件的存在,给视频图像的蓝藻识别和监测带来了极大挑战。为克服不同摄像头的观测角度不一致、光照强度和背景条件不一致等诸多挑战,本文将从蓝藻与水体的混淆程度和蓝藻异质性两方面进行表征分析,研究其对蓝藻识别的影响。蓝藻与水体的混淆程度分析是在蓝藻图像、阴影水体图像和浑浊水体图像中挑选样本(像元),观测各类型样本的直方图分布特征。蓝藻异质性分析是比较直接在图像中挑选蓝藻、水体样本(像元)和将这些样本及所在图像按颜色分成不同类型后样本的直方图分布差异。根据对这两部分内容的分析,确定了本文蓝藻识别与监测的基本思路。

首先,当风力较强时,湖水会产生波浪,与波浪的垂直方向上亮度值呈现高、低依次排列,阴影水体(亮度值低的区域)表现出与蓝藻相似的特征.而且风力对湖水的搅动使水体变浑浊,导致阴影水体、浑浊水体对蓝藻识别造成干扰(图3).



图3 风力作用对图像光谱特征的影响

Fig.3 Effect of wind force on images spectral characteristics

为分析蓝藻与阴影水体、浑浊水体的混淆程度,本文选取了阴影水体 461 个样本(像元)、浑浊水体 505 个样本(像元)、蓝藻 497 个样本(像元)进行对比.分别比较了阴影水体样本、浑浊水体样本与蓝藻样本的分布特征(图4),发现它们在分布上有较多的重叠,说明样本的可分离性差,难以准确识别蓝藻.

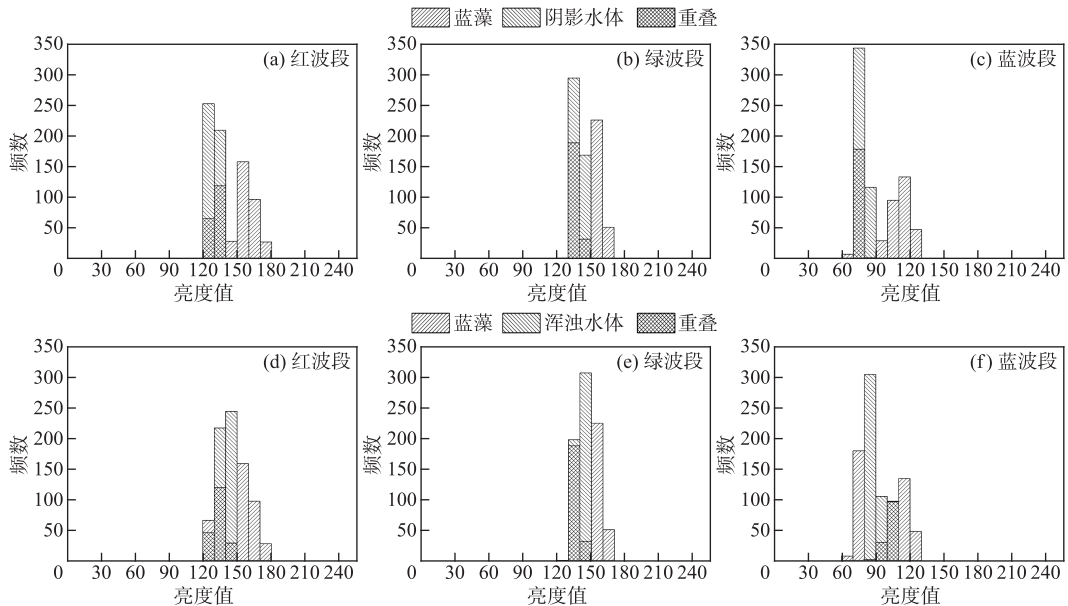


图4 蓝藻与阴影水体、浑浊水体样本的亮度值分布(上:蓝藻与阴影水体,下:蓝藻与浑浊水体)

Fig.4 Lightness distribution of cyanobacteria, shadow water, and turbid water samples

其次,视频图像中的蓝藻具有强异质性.受蓝藻自身特性和环境条件的综合影响,蓝藻在视频图像中呈现不同的颜色,主要包括墨绿色、青色、绿色、黄色和灰色 5 种颜色,分别记为 BG 型、C 型、G 型、Y 型、GR 型图像(图5).

为分析强异质性条件下蓝藻和水体的亮度值差异,以人机交互方式从 100 张典型图像中选取 10463 个样本(像元)进行分析,其中蓝藻样本 4509 个,水体样本 5954 个.样本的亮度值分布特征图显示:红、蓝波段中蓝藻的亮度值稍低于水体的亮度值,绿波段蓝藻和水体的亮度值较为接近(图6).

由图6可以看出,如果直接选取样本进行蓝藻识别很难将图像中的蓝藻与水体进行区分.为解决蓝藻的异质性问题,本文通过人工判别的方式对这 100 张图像进行颜色划分,并将选好的样本归入 BG 型、C 型、G 型、Y 型和 GR 型 5 类,分析各类型图像的亮度值分布特征(图7).

图7显示不同类型图像中蓝藻和水体样本的亮度值分布差异较大.红、蓝波段的蓝藻和水体样本亮度

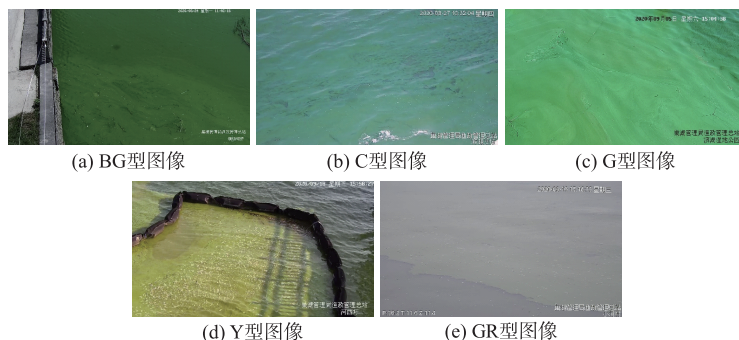


图5 相近时段内5种不同类型的蓝藻图像

Fig.5 Five types of cyanobacteria images on similar time

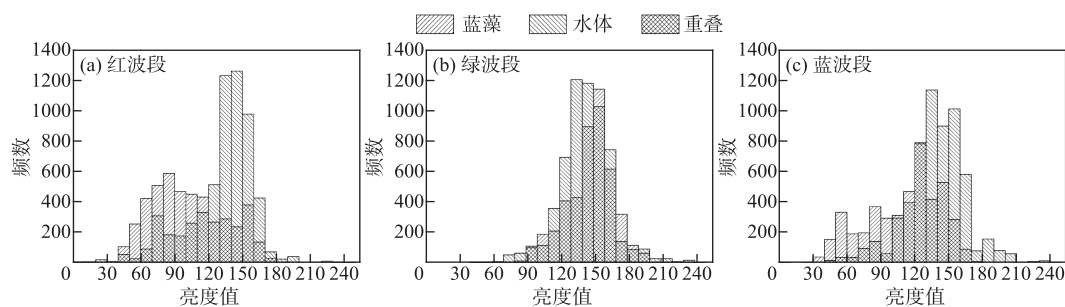


图6 蓝藻和水体样本的亮度值分布

Fig.6 Lightness distribution of cyanobacteria and water samples

值分布在不同类型图像中均出现了明显的“阶梯式”分布. 同一类型图像中蓝藻样本在不同波段的分布特征差异较大. 除 GR 型图像外, 其余各类型图像均有 2 个波段的蓝藻亮度值分布接近, 且与另一波段有明显区别. GR 型图像 3 个波段的蓝藻亮度分布区间则比较接近. 总体上, 利用颜色对图像类型进行划分后, 可将蓝藻、水体样本的亮度分布限制在较小区间内, 增大了蓝藻与水体的区分度. 说明在设计蓝藻识别算法时可先划分图像类型, 然后分不同的类型进行识别, 以此解决蓝藻的异质性问题.

本文使用的视频图像仅有红、绿、蓝 3 波段. 使用绿波段进行表征分析和后续蓝藻识别, 是在缺少近红外波段时, 可作为重要的替代波段, 同时在输入卷积神经网络时, 绿波段具有较丰富的空间信息, 有助于区分蓝藻与阴影水体、浑浊水体.

2.2 视频图像的蓝藻识别方法

视频图像的蓝藻识别方法包括两部分. 首先, 基于多尺度深度网络进行图像粗粒度分类, 估算图像的蓝藻覆盖率以区分蓝藻与阴影水体、浑浊水体. 其次, 利用颜色对图像类型进行划分, 再基于随机森林分别对不同图像类型进行蓝藻精细化识别.

2.2.1 基于多尺度深度网络的图像粗粒度分类 本文借鉴 VGG 分类网络的思路, 形成多尺度深度网络框架^[33]. 首先, 以分块后的样本图像为基准, 通过对该样本的空间范围进行规则化“生长”或“压缩”, 获得不同视域范围的多张图像. 然后将这些图像重采样至与基准图像相同的尺度, 合并成为多尺度特征. 最后将多尺度样本输入卷积神经网络进行训练. 多尺度体现在输入网络的训练样本是多个尺度特征的综合叠加(图 8).

以人工筛选的方式从 679 张图像中选取 67 张各类典型蓝藻图像制作深度神经网络的训练样本. 其中, 上午(8:00—11:00)图像 31 张、中午(11:00—14:00)图像 7 张、下午(14:00—17:00)图像 29 张. 以 40 行×40 列作为分块基准, 单张图像可分成 1296 个样本(图像块, 下同). 然后对样本进行标记. 因分块较大会遗

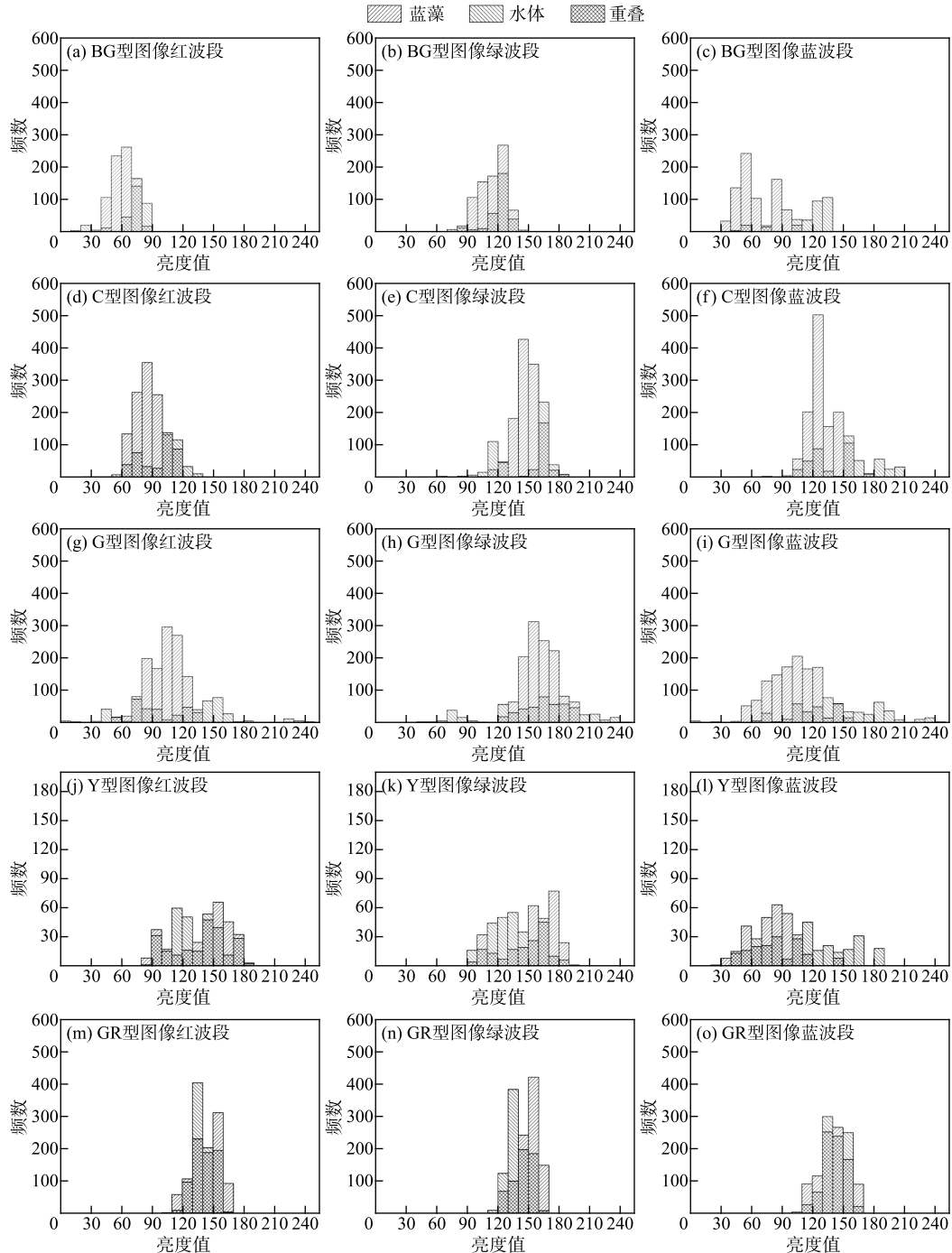


图 7 5 种类型图像的蓝藻和水体样本的亮度值分布

Fig.7 Lightness distribution of cyanobacteria and water samples from five kinds of images

漏小区域的蓝藻,图像蓝藻覆盖率被低估,从而直接被判定为水体图像,本文在样本标记时刻意偏向于蓝藻.若蓝藻所占面积大于 1/4,该样本为蓝藻,标记为 1;否则为水体,标记为 0. 剔除掉不适合作为训练数据的图

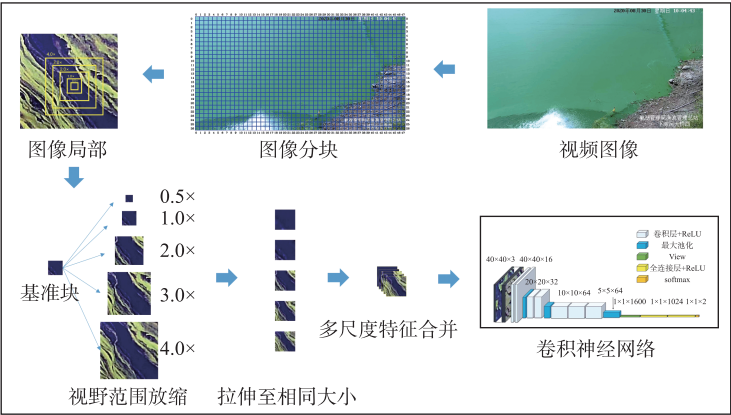


图 8 基于多尺度卷积神经网络的图像分类示意

Fig.8 Schematic diagram of image classification based on multiscale convolutional neural network classification

像块后,最终形成 65762 个样本,并按 6: 4 的比例随机分成训练集 (39363 个样本) 和测试集 (26399 个样本).

网络训练在 PyTorch 环境中进行. 将训练集样本分批次投入网络,每次投入 500 个样本,79 次可完成一个循环. 学习率参数设置为 0.001,并随迭代次数的增加而减小;冲量参数采用经验值 0.9. 共设置了 3000 次循环过程,根据交叉熵损失函数计算损失值.

基于训练好的模型,对所有图像进行推测. 根据推测结果统计每张图像的蓝藻图像块占比,估算每张图像的蓝藻覆盖率. 再基于蓝藻覆盖率的数值,将图像划分成 3 种类型. 第一类是蓝藻覆盖率小于 5% 的图像,图像内以水体为主,包含浑浊水体、阴影水体等情形;第二类是蓝藻覆盖率大于等于 5% 且小于 90% 的图像,图像中蓝藻和水体共存;第三类是蓝藻覆盖率大于等于 90% 的图像,图像中以蓝藻为主.

2.2.2 基于随机森林的蓝藻精细化识别 本文将低蓝藻覆盖率图像作为水体图像,针对高、中蓝藻覆盖率图像分别划分图像类型,将图像分为 BG、C、G、Y 和 GR 5 种类型. 首先,计算待分类图像 3 个波段的亮度均值,并进行两两差分. 然后,求取 3 个差分值的最小值和最大值,据此划分图像类型. GR 型图像的 3 个差分值均比较小,可设定最大阈值 threshold1 首先进行区分. 其余类型图像由 3 波段中差分最小值波段的剩余波段决定. BG 型图像和 G 型图像均由绿波段决定,通过设定绿波段阈值 threshold2 区分两者. 完成图像类型划分后,因此将中、高蓝藻覆盖率图像按类型归并,最后共形成 5 类图像.

随机森林是一种集成学习方法,其作用机制是从所有特征中随机抽取适量特征构建指定数量的决策树,多棵决策树投票决定最终的分类结果^[34]. 由于其较高的分类精度和较快的处理速度已得到广泛应用^[35]. 本文使用机器学习包 Scikit-learn 中的 Random Forest Classifier 函数构建随机森林^[36],需要设置决策树个数和训练

表 1 各类型图像的训练样本

Tab.1 Details of training samples collected for various types of images

图像类型	图像数量	蓝藻样本 (像元)	水体样本 (像元)	总样本 (像元)
BG 型	17	2925	1322	4247
C 型	17	2135	1627	3762
G 型	34	3943	3211	7154
Y 型	24	3152	2766	5918
GR 型	24	2207	2298	4505
合计	116	14362	11224	25586

每棵决策树的特征个数 2 个重要参数. 将每棵决策树的特征个数设定为 2 和 3,将决策树个数的上限设定为 150 棵.

本文使用 116 张图像来选取随机森林的训练样本,其中 BG 型和 C 型图像各 17 张、G 型图像 34 张、Y 型和 GR 型图像各 24 张. 以人机交互方式选取各类型的训练样本 (表 1). 针对上述 5 种类型图像,分别使用各自样本训练随机森林模型,统计随决策树个数增加时平均袋外数据的变化.

2.3 视频图像的蓝藻动态监测

由于单个摄像头的观测范围有限,本文以渔政站为单位对巢湖水域视频图像进行统计分析,实现巢湖沿岸水域的蓝藻动态监测. 首先,获取各摄像头视角内的日

均蓝藻覆盖率. 对各渔政站辖区所有摄像头的日均蓝藻覆盖率进行统计,获取渔政站辖区水域日均蓝藻覆盖率. 由此判定当日渔政站沿岸水域是否暴发蓝藻水华以及暴发的程度. 其次,将日均蓝藻覆盖率作为衡量指标,设定阈值判定当日是否发生蓝藻水华. 考虑到蓝藻覆盖率估算时的误差,本文将发生蓝藻水华的阈值设定为 5%. 若日均蓝藻覆盖率大于 5% 时,则认为发生蓝藻水华. 统计每月蓝藻覆盖率大于 5% 的天数,再除以当月有效观测天数,即可获得月均蓝藻暴发频率,以此分析蓝藻暴发比较严重的月份.

2.4 精度评价指标

本文采用 $F1$ 测度进行精度评价,包括准确率(Precision)、召回率(Recall)、 $F1$ 分数($F1$ -score)和总体精度(OA)4 个指标^[37]. 准确率指识别结果中蓝藻被正确识别的比例,记为 P . 召回率指所有蓝藻样本中蓝藻被正确识别的比例,记为 R . $F1$ 分数是准确率和召回率的协调平均值,记为 $F1$. 总体精度指所有验证样本(蓝藻、水体)被正确识别的比例,记为 OA ,各指标的计算公式如下:

$$P = TP / (TP + FP) \quad (1)$$

$$R = TP / (TP + FN) \quad (2)$$

$$F1 = 2P \times R / (P + R) \quad (3)$$

$$OA = (TP + TN) / (TP + TN + FP + FN) \quad (4)$$

式中, TP 指蓝藻被正确识别的样本数, TN 指水体被正确识别的样本数, FP 指水体被识别为蓝藻的样本数, FN 指蓝藻被识别为水体的样本数.

3 蓝藻识别与动态监测结果

3.1 蓝藻识别参数设置

本文采用图像粗粒度分类和精细化识别相结合的方法进行蓝藻识别. 首先,基于多尺度深度神经网络进行蓝藻粗粒度分类. 神经网络开始训练时损失值为 0.673,迭代 1300 次后接近于 0. 将训练好的神经网络对所有图像进行蓝藻覆盖率估算,并将图像分成高、中、低蓝藻覆盖率 3 种类型. 然后,利用颜色对高、中蓝藻覆盖率图像进行类型划分,再基于随机森林分别对不同图像进行蓝藻精细化识别. 经多次测试,高、中蓝藻覆盖率图像的 $threshold1$ 设定为 10 最优,高蓝藻覆盖率图像的 $threshold2$ 设定为 135,中等蓝藻覆盖率图像的 $threshold2$ 设置为 133.5 时最优. 随机森林使用 50 棵决策树、每棵决策树中随机抽选特征个数为 2 时,分类效果最优. 基于本文提出的蓝藻识别模型,对 2020 年 8—10 月环巢湖视频监控网络的 679 张图像进行蓝藻识别.

3.2 蓝藻识别结果分析

为验证本文模型的优势,利用未参与训练的 295 张图像对蓝藻识别结果进行精度评价. 同时,还加入了对照组实验. 将所有训练样本合并后训练单个随机森林分类器,记为模型 A;对训练样本所在图像仅进行图像粗粒度分类,分为高、中、低蓝藻覆盖率 3 种图像,分别训练各自的随机森林分类器,记为模型 B. 基于相同的验证样本和基分类器,比较模型 A、模型 B 和本文模型的蓝藻识别结果的精度.

以规则抽样的方式选取验证样本. 在每张图像中选取 9 个像元及其 8 邻域的像元作为验证样本. 每张图像共选取 81 个样本(像元),所有图像共计 23895 个样本(像元). 验证样本的标注以人机交互方式进行. 验证样本共包括蓝藻样本 10814 个(像元)、水体样本 13081 个(像元).

图 9 为 5 张不同类型图像的蓝藻识别结果. 在 BG 型图像中,对于左半部分亮度值较暗区域,模型 A、B 的结果中有许多蓝藻未被检测出来,本文模型识别的蓝藻范围更加完整. 在 C 型图像中,水体呈青灰色,与青色的蓝藻容易混淆,模型 A、B 的结果出现大量的误判现象,而本文模型受影响较小. 在 G 型图像中,本文模型识别的蓝藻范围与实际的蓝藻水华更接近. 在 Y 型图像中,由于光照强度不均,在图像右上区域模型 A、B 的误判区域较多,本文模型可减轻光照不均的影响. 在 GR 型图像中,蓝藻与水体亮度值分布接近,本文模型可获得更完整的蓝藻范围.

在不同覆盖率图像的蓝藻识别中,本文模型均取得较好效果(图 10). 低蓝藻覆盖率图像包含了大量的阴影区域,模型 A 的识别结果中该区域被误判为蓝藻,而模型 B 和本文模型则没有. 中蓝藻覆盖率图像中,3 种模型均可将蓝藻大致范围识别出来,但本文模型的识别结果中蓝藻范围更加完整,蓝藻内部的“空洞”较

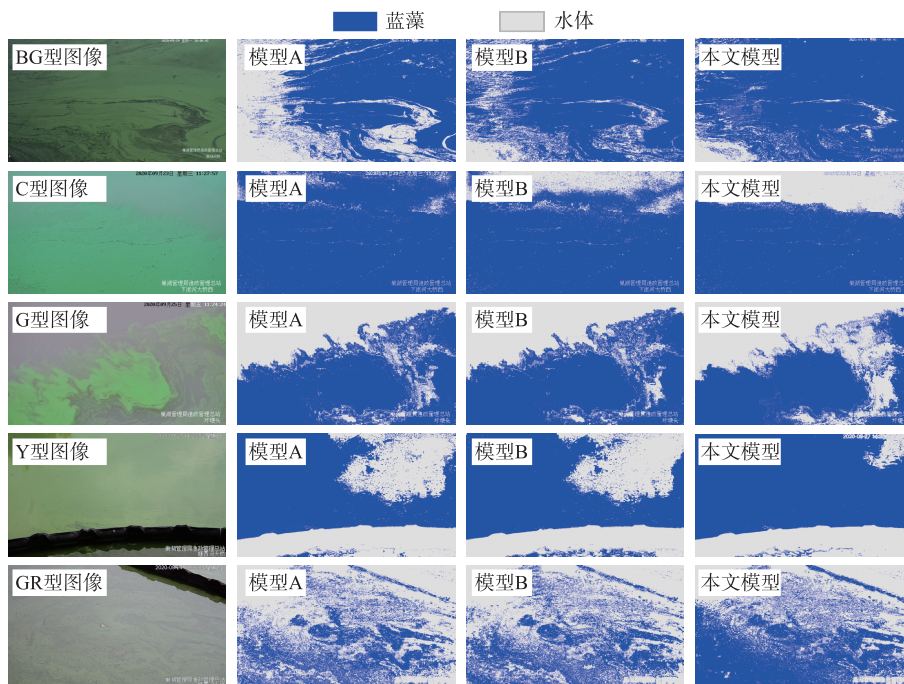


图 9 3 种模型在不同颜色图像上的蓝藻识别结果

Fig.9 Cyanobacteria recognition results of three models on different types of images

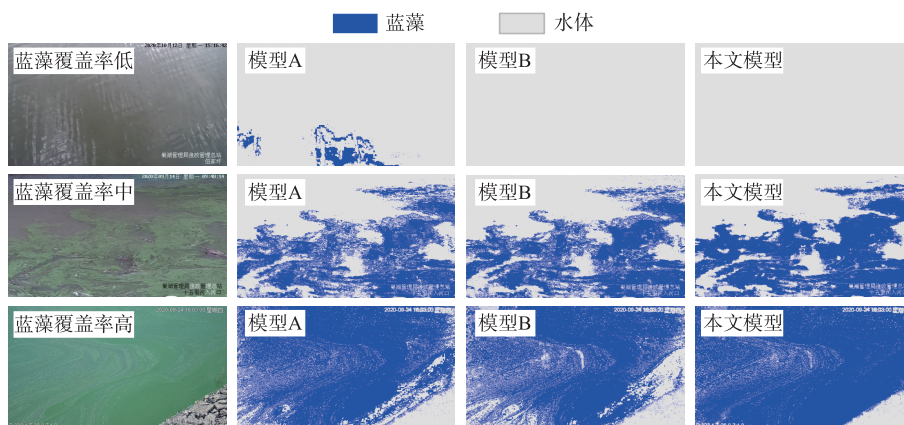


图 10 3 种模型在不同蓝藻覆盖率图像上的蓝藻识别结果

Fig.10 Cyanobacteria recognition results of three models on different cyanobacteria coverage images

少。相比模型 A、B, 在高蓝藻覆盖率图像中本文模型识别的蓝藻范围与真实情况更接近, 尤其是在该图像的左下角和右部, 误判现象更少。

表 2 为精度评价结果。本文模型中蓝藻样本被正确识别的个数 (TP) 为 7653 个, 在 3 种模型中精度最高。模型 B 和本文模型中水体样本被识别为蓝藻 (FP) 的个数分别比模型 A 少 121 和 65 个, 这说明基于蓝藻覆盖率估算的图像粗粒度分类方法发挥了作用, 使阴影水体、浑浊水体与蓝藻的混淆减少。此外本文模型中蓝藻样本被识别为水体 (FN) 的个数是三者中最少的, 说明在蓝藻识别中利用颜色区分图像的类型, 降低蓝藻异质性的策略是有效的。

表 2 3 种模型的蓝藻识别结果精度比较

Tab.2 Accuracy comparison of three models in cyanobacteria recognition results

蓝藻识别模型	TP	FP	FN	TN	<i>P</i>	<i>R</i>	<i>F1</i>	<i>OA</i>
模型 A	7522	904	3292	12177	0.893	0.696	0.782	0.824
模型 B	7483	783	3331	12298	0.905	0.692	0.784	0.828
本文模型	7653	839	3161	12242	0.901	0.708	0.793	0.833

在 4 个评价指标中,本文模型有 3 个指标(召回率、F1 分数、总体精度)是 3 种模型中最高的. 从模型 A 到模型 B,再到本文模型,F1 分数和总体精度的值呈逐渐增加的趋势,说明通过划分图像类型,能够减少蓝藻与阴影水体、浑浊水体的混淆,降低图像中蓝藻异质性,提高蓝藻识别的精度;也说明先图像粗粒度分类、后蓝藻精细化识别的方法优势显著.

3.3 蓝藻动态监测分析

各渔政站沿岸水域的日均蓝藻覆盖率变化趋势显示,7 个渔政站沿岸水域均发生不同程度的蓝藻水华(图 11). 在 8 月 17 日和 9 月 20 日前后,卧牛站有两次较大的蓝藻水华事件,日均蓝藻覆盖率峰值超过 70%. 高林站沿岸水域的蓝藻覆盖率始终处于较低水平,仅在 9 月 5 日左右发生一次小规模蓝藻水华,当日蓝藻覆盖率为 40%. 白山站在 9 月 7 日至 10 月 13 日期间有持续性的蓝藻水华事件,但日均蓝藻覆盖率总体偏低,未引发大面积的蓝藻水华. 三河站沿岸水域 9 月份内蓝藻水华比较严重,包括 3 次持续时间长、覆盖范围广的蓝藻事件;此外在 8 月和 10 月也有少量的蓝藻出现. 义城站的蓝藻暴发程度也较为严重,在 8—10 月暴发了 4 次大规模的蓝藻水华. 在 8 月 30 日和 9 月 26 日开始的 2 次蓝藻水华持续时间较长,日均蓝藻覆盖率达 80% 以上. 长临站沿岸水域每次蓝藻暴发的持续时间短,但暴发强度高. 中庙站沿岸水域的蓝藻水华处于中等水平,8—10 月的大部分时间暴发了蓝藻水华,但日均蓝藻覆盖率较低. 从月份的角度来看,蓝藻水华主要发生在 8—9 月份,日均蓝藻覆盖率大多处于较高水平. 10 月份日均蓝藻覆盖率持续走低,直至蓝藻消退.

据此可知,蓝藻暴发比较严重的水域在西巢湖沿岸的三河站、义城站和长临站. 位于东巢湖西岸的中庙站沿岸水域,虽然蓝藻暴发程度不高,但持续时间比较长,也造成了较大影响. 位于东巢湖的卧牛站、高林站和白山站的蓝藻暴发程度较弱,偶有蓝藻发生. 研究表明,巢湖蓝藻暴发区域主要集中在西巢湖沿岸和东巢湖的西岸,这与遥感图像蓝藻识别获得的巢湖沿岸蓝藻空间分布范围基本一致^[1,30].

在蓝藻暴发频率方面,三河站、义城站、长临站和中庙站沿岸水域 8—10 月均处于较高水平;高林站和白山站沿岸水域在各月份均比较低(表 3). 西巢湖沿岸水域的蓝藻暴发频率比东巢湖整体偏高. 总体来看,各渔政站沿岸水域的蓝藻暴发频率最高的月份是 9 月,其中三河站、义城站和中庙站沿岸水域的暴发频率达 80% 以上. 8—10 月各渔政站沿岸水域的蓝藻暴发频率呈现先增加后降低的趋势. 8 月份中庙站沿岸水域的蓝藻暴发频率为 100%,比真实蓝藻暴发频率高. 这是因为该渔政站 8 月份的统计天数不完整导致的. 在将来的蓝藻动态监测中,将进行更全面地统计,获得更准确的蓝藻暴发频率.

表 3 各渔政站观测的月均蓝藻暴发频率

Tab.3 Monthly average frequency of cyanobacterial blooms of each station observation

渔政站	8 月	9 月	10 月
卧牛站	50.0%	61.5%	10.0%
高林站	0	16.7%	0
白山站	0	54.6%	25.0%
三河站	62.5%	91.7%	70.0%
义城站	75.0%	83.3%	71.4%
长临站	63.6%	46.7%	18.2%
中庙站	100.0%	93.3%	45.5%

4 讨论

4.1 不同环境条件对蓝藻识别精度的影响

在多种外部环境因素的作用下,蓝藻表征具有多样性,这使得蓝藻识别的精度具有不确定性. 因此有必要分析在不同外部环境条件下蓝藻识别的精度. 本文基于 295 张验证图像(23895 个样本像元),分别从不同观测时间、不同观测角度和不同天气状况对蓝藻识别的精度进行分析(图 12). 其中,按观测时间分为

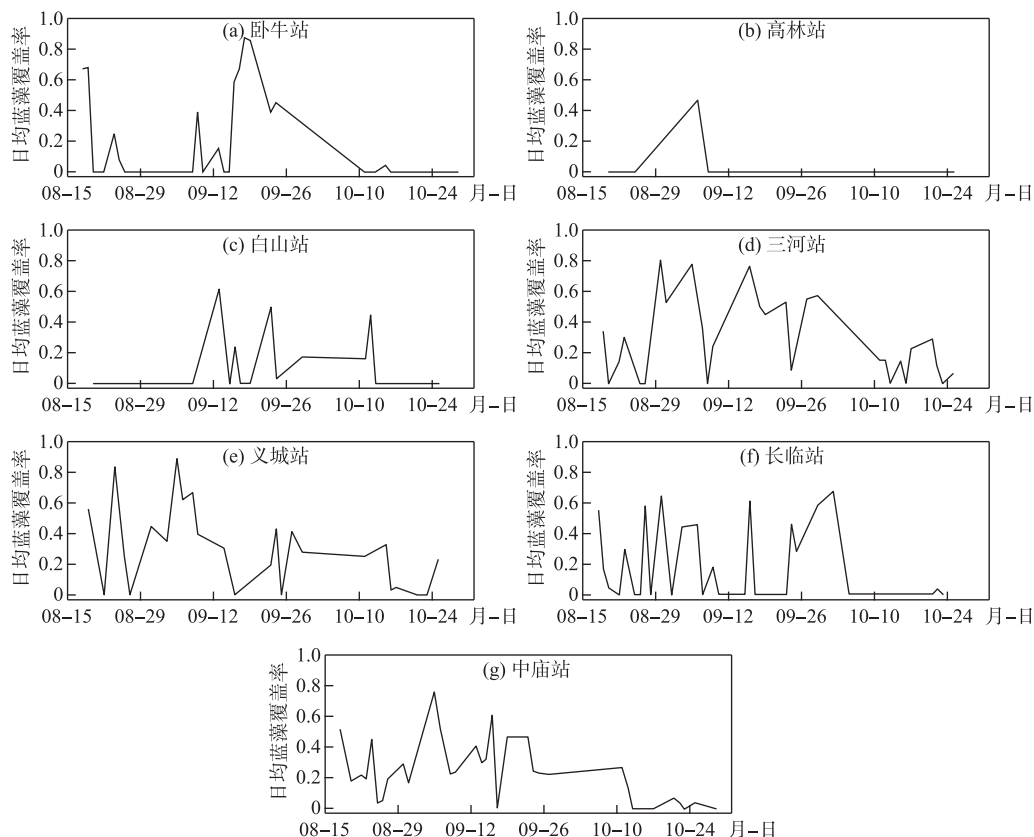


图 11 各渔政站观测的日均蓝藻覆盖率变化趋势

Fig.11 Change trends of the daily average cyanobacteria coverage of each station observation

8:00—11:00、11:00—14:00、14:00—17:00 3 类图像;按观测角度分为朝东、朝南、朝西、朝北 4 类图像;按天气状况分为晴、大风和薄雾 3 类图像。

图 12a 显示各时段图像的蓝藻识别精度比较接近。准确率 P 在 0.87~0.91 之间;召回率 R 在 0.69~0.71 之间; $F1$ 分数和总体精度 OA 均在 0.8 以上,8:00—11:00 的总体精度为 0.9。以上数据说明,本文蓝藻识别模型在各时段内的蓝藻识别结果均比较稳定,而 8:00—11:00 总体精度达到最高水平,识别效果最佳,模型更为适用。总体上,本文蓝藻识别模型不受观测时间变化的影响。

图 12b 显示 4 种观测角度下蓝藻的识别精度表现出明显的差异。其中,观测角度朝西和朝南图像的蓝藻识别精度最高, $F1$ 约为 0.84, OA 约为 0.9;观测角度朝东的图像的精度次之, $F1$ 和 OA 均约为 0.8;观测角度朝北的图像准确率 P 最高,为 0.92,但召回率 R 仅有 0.6,低于平均水平。观察图像发现,朝北观测的摄像机位于稀疏林地居多,导致拍摄图像中包含树木阴影,部分蓝藻未被有效识别;同时,朝北观测的摄像机数量有限(5 处),整体精度较低,具体原因有待进一步研究。在模型的后续改进中,可综合考虑观测角度的差异,提高蓝藻识别的精度。

图 12c 显示不同天气状况下的蓝藻识别结果存在较大差异。晴空条件下 $F1$ 和 OA 均在 0.87 以上,蓝藻识别结果比较理想;但在恶劣天气下,仍有较大局限。大风情况下 $F1$ 仅为 0.6, OA 为 0.8。 OA 高是因为大风情况下,蓝藻发生频率低,而水体识别准确高,导致总体精度较高。薄雾情况下 P 值较高,为 0.91,其他精度指标均比较低,这是因为薄雾条件下蓝藻特征不明显的区域未能有效识别。总体上,对蓝藻识别模型影响最大的外部环境条件是天气因素,处于大风、薄雾天气下蓝藻识别精度比晴空条件下约低 20%。如何在恶劣天气中有效提高蓝藻识别精度将是后续研究的重点。

4.2 对蓝藻监测实践和湖泊环境管理的启示

视频监控网络作为巢湖“天-空-地”协同观测平台的地面监测环节,能够及时掌握湖泊沿岸蓝藻水华的空间分布和动态变化,实现对重点区域的蓝藻监测,在巢湖蓝藻治理实践中可发挥重要作用^[28]. 本文以环巢湖视频监控网络为支撑,发展了一种基于视频图像的自动化蓝藻识别和动态监测方法. 研究成果已在巢湖开展应用,研发的视频图像蓝藻监测系统部署于巢湖蓝藻水华监测预警与模拟分析平台,在巢湖管理局进行业务化运行,并具有一定的推广潜力.

多点位视频监控的有机组合能够有效描述巢湖蓝藻水华特征. 因为单个摄像机拍摄的视频图像范围较小,不能代表巢湖沿岸蓝藻水华的整体水平,所以本文以渔政站为单位开展蓝藻水华的实时监测. 每一渔政站内包含 4~6 个摄像机,统计各摄像机的平均蓝藻覆盖水平,可使监测结果代表巢湖沿岸水域的蓝藻水华程度,在摄像机数量有限时,是一种好的解决途径. 在成本允许的情况下,可提高视频监控网络的分布密度,使监测结果更能代表巢湖的蓝藻水平;同时摄像机中可配置近红外波段,以提高蓝藻识别的能力,更好地为蓝藻识别与动态监测服务.

总体而言,本文提出模型取得了较好的蓝藻识别效果. 在此基础上,可以进一步确定巢湖蓝藻暴发相对频繁的区域和时间,掌握蓝藻暴发的规律. 巢湖“天-空-地”协同观测平台为湖泊治理提供了一种有效解决方案,可为管理层决策提供依据,使蓝藻治理更具有针对性. 因此,应重视视频监控在蓝藻治理中的作用,并基于“天-空-地”协同观测平台建立长效的蓝藻监测机制,为湖泊环境管理服务.

5 结论

为克服不同摄像头的观测角度不一致、光照强度和背景条件不一致等诸多挑战,针对复杂成像环境引起的视频图像蓝藻与浑浊水体和阴影水体难区分、蓝藻异质性强等诸多难题,本文提出了一种视频图像蓝藻识别方法,通过模型集成有效提高了识别的精度,具有一定的方法创新,可为湖泊蓝藻的动态监测提供助力,为制定蓝藻打捞与治理方案提供科学支撑. 本文提出的蓝藻识别模型中,多尺度深度神经网络可以准确估算图像的蓝藻覆盖率,实现蓝藻与浑浊水体、阴影水体的区分,解决不同背景条件对蓝藻识别的影响;基于随机森林的蓝藻精细化识别可以在一定程度上克服蓝藻的异质性;采用先图像粗粒度分类、后蓝藻精细化识别的方法能有效提高蓝藻识别的精度.

对 2020 年 8—10 月巢湖沿岸水域蓝藻动态监测的结果表明,西巢湖沿岸的蓝藻暴发程度普遍比东巢湖严重,在 8—10 月西巢湖沿岸水域蓝藻常处于较高水平. 在蓝藻暴发频率方面,西巢湖沿岸的月均蓝藻暴发频率比东巢湖高,9 月份蓝藻暴发最为频繁. 整体来看,8—10 月巢湖沿岸水域的日均蓝藻覆盖率和月均蓝藻暴发频率均呈现先升高后降低的趋势.

6 参考文献

- [1] Duan HT, Wan NS, Qiu YG *et al.* Discussions and practices on the framework of monitoring system in eutrophic lakes and reservoirs. *J Lake Sci*, 2020, **32**(5): 1396-1405. DOI: 10.18307/2020.0513. [段洪涛, 万能胜, 邱银国等. 富营养化湖库天空地一体化监控平台系统设计与实践. 湖泊科学, 2020, **32**(5): 1396-1405.]

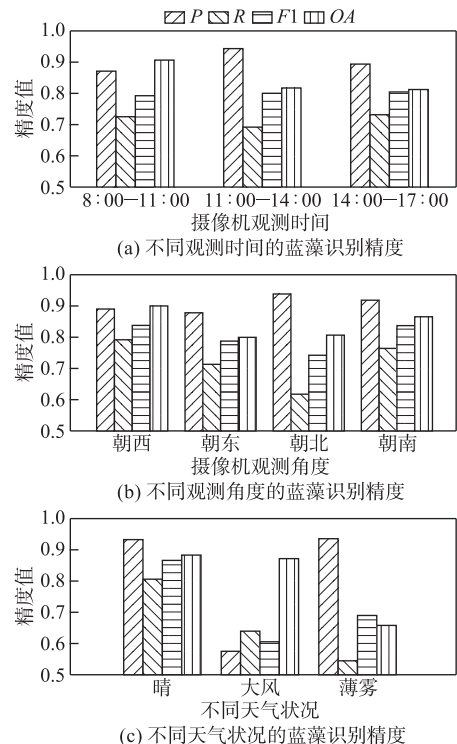


图 12 不同环境条件下的蓝藻识别精度

Fig.12 Accuracy of cyanobacterial blooms recognition under different environmental conditions

- [2] Wu FC, Meng W, Song YH *et al.* Research progress in lake water quality criteria in China. *Acta Scientiae Circumstantiae*, 2008, **28**(12): 2385-2393. DOI: 10.13671/j.hjkxb.2008.12.016. [吴丰昌, 孟伟, 宋永会等. 中国湖泊水环境基准的研究进展. 环境科学学报, 2008, **28**(12): 2385-2393.]
- [3] Ministry of Ecology and Environment of the People's Republic of China. 2020 Bulletin of China's Ecological Environment, 2021. [中华人民共和国生态环境部. 2020 年中国生态环境状况公报, 2021.]
- [4] Paerl HW, Huisman J. Blooms like it hot. *Science*, 2008, **320**(5872): 57-58. DOI: 10.1126/science.1155398.
- [5] Kong FX. The management of lake eutrophication and the control of cyanobacterial bloom. *Jiangsu Science and Technology Information*, 2007, (9): 7-11. [孔繁翔. 湖泊富营养化治理与蓝藻水华控制. 江苏科技信息, 2007, (9): 7-11.]
- [6] Ho JC, Michalak AM, Pahlevan N. Widespread global increase in intense lake phytoplankton blooms since the 1980s. *Nature*, 2019, **574**(7780): 667-670. DOI: 10.1038/s41586-019-1648-7.
- [7] Domingues RB, Sobrino C, Galvão H. Impact of reservoir filling on phytoplankton succession and cyanobacteria blooms in a temperate estuary. *Estuarine, Coastal and Shelf Science*, 2007, **74**(1/2): 31-43. DOI: 10.1016/j.ecss.2007.03.021.
- [8] Ji XY, Zhang D, Li WP *et al.* Operational platform construction of heaven and earth integration algae monitoring in Lake Dianchi. *Environmental Protection*, 2016, **44**(18): 24-27. [嵇晓燕, 张迪, 李文攀等. 湖泊蓝藻天地一体化监测业务化运行平台的构建——以滇池流域为例. 环境保护, 2016, **44**(18): 24-27.]
- [9] Lai L, Zhang YC, Jing YY *et al.* Research progress on remote sensing monitoring of phytoplankton in eutrophic water. *J Lake Sci*, 2021, **33**(5): 1299-1314. DOI: 10.18307/2021.0502. [来莱, 张玉超, 景园媛等. 富营养化水体浮游植物遥感监测研究进展. 湖泊科学, 2021, **33**(5): 1299-1314.]
- [10] Yin ML, Zhang JQ, Lu CH *et al.* A review of real-time monitoring techniques for cyanobacteria. *Value Engineering*, 2019, **38**(22): 208-209. [尹美琳, 张嘉琪, 路春晖等. 蓝藻实时监测技术综述. 价值工程, 2019, **38**(22): 208-209.]
- [11] Wang XY, Yang W. Water quality monitoring and evaluation using remote sensing techniques in China: A systematic review. *Ecosystem Health and Sustainability*, 2019, **5**(1): 47-56. DOI: 10.1080/20964129.2019.1571443.
- [12] Song CQ, Zhan PF, Ma RH. Progress in remote sensing study on lake hydrologic regime. *J Lake Sci*, 2020, **32**(5): 1406-1420. DOI: 10.18307/2020.0514. [宋春桥, 詹鹏飞, 马荣华. 湖泊水情遥感研究进展. 湖泊科学, 2020, **32**(5): 1406-1420.]
- [13] Duan HT, Zhang SX, Zhang YZ. Cyanobacteria bloom monitoring with remote sensing in Lake Taihu. *J Lake Sci*, 2008, **20**(2): 145-152. DOI: 10.18307/2008.0202. [段洪涛, 张寿选, 张渊智. 太湖蓝藻水华遥感监测方法. 湖泊科学, 2008, **20**(2): 145-152.]
- [14] Gower JFR. Red tide monitoring using AVHRR HRPT imagery from a local receiver. *Remote Sensing of Environment*, 1994, **48**(3): 309-318. DOI: 10.1016/0034-4257(94)90005-1.
- [15] Hu CM. A novel ocean color index to detect floating algae in the global oceans. *Remote Sensing of Environment*, 2009, **113**(10): 2118-2129. DOI: 10.1016/j.rse.2009.05.012.
- [16] Zhu L, Wang Q, Wu CQ *et al.* Monitoring and annual statistical analysis of algal blooms in Chaohu based on remote sensing. *Environmental Monitoring of China*, 2013, **29**(2): 162-166. [朱利, 王桥, 吴传庆等. 巢湖水华遥感监测与年度统计分析研究. 中国环境监测, 2013, **29**(2): 162-166.]
- [17] Zhang YC, Ma RH, Duan HT *et al.* A novel algorithm to estimate algal bloom coverage to subpixel resolution in Lake Taihu. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, **7**(7): 3060-3068. DOI: 10.1109/JSTARS.2014.2327076.
- [18] Ananias PHM, Negri RG. Anomalous behaviour detection using one-class support vector machine and remote sensing images: A case study of algal bloom occurrence in inland waters. *International Journal of Digital Earth*, 2021, **14**(7): 921-942. DOI: 10.1080/17538947.2021.1907462.
- [19] El-Alem A, Chokmani K, Laurion I *et al.* Comparative analysis of four models to estimate chlorophyll-a concentration in case-2 waters using MODerate resolution imaging spectroradiometer (MODIS) imagery. *Remote Sensing*, 2012, **4**(8): 2373-2400. DOI: 10.3390/rs4082373.
- [20] Blix K, Eltoft T. Machine learning automatic model selection algorithm for oceanic chlorophyll-a content retrieval. *Remote Sensing*, 2018, **10**(5): 775. DOI: 10.3390/rs10050775.
- [21] Le CF, Li YM, Zha Y *et al.* Remote estimation of chlorophyll a in optically complex waters based on optical classification. *Remote Sensing of Environment*, 2011, **115**(2): 725-737. DOI: 10.1016/j.rse.2010.10.014.

- [22] Duan HT, Ma RH, Zhang YC *et al.* Are algal blooms occurring later in Lake Taihu? Climate local effects outcompete mitigation prevention. *Journal of Plankton Research*, 2014, **36**(3): 866-871. DOI: 10.1093/plankt/ftt132.
- [23] Wang WC, Zou WB. Methods of extraction in high resolution remote sensing image information. *Beijing Surveying and Mapping*, 2013, **111**(4): 1-5. [王伟超, 邹维宝. 高分辨率遥感影像信息提取方法综述. 北京测绘, 2013, **111**(4): 1-5.]
- [24] Zhan XF. Construction and implementation of lake ecological environment monitoring system. *Wireless Internet Technology*, 2019, **16**(3): 66-67. [詹晓芳. 湖泊生态环境监测系统的建设和实现. 无线互联科技, 2019, **16**(3): 66-67.]
- [25] Yang X, Chen XR, Du CX *et al.* Cyanobacteria bloom monitoring system. China, CN103198325A. [杨旭, 陈晓荣, 杜翠霞等. 蓝藻水华监测系统. 中国, CN103198325A.]
- [26] Zhang YL, Shi K, Zhang YB *et al.* The proposal, practice and preliminary application of land-based (ground-based, shore-based) remote sensing of water environment. *National Remote Sensing Bulletin*, 2021, **25**(11): 2163-2172. DOI: 10.11834/jrs.20210473. [张运林, 施坤, 张毅博等. 陆基(地基、岸基)水环境遥感的提出、实践和初步应用. 遥感学报, 2021, **25**(11): 2163-2172.]
- [27] Zhang YL, Zhang YB, Li N *et al.* Capturing the rapid intra-day change of cyanobacteria bloom by land-based hyperspectral remote sensing in Lake Taihu. *J Lake Sci*, 2021, **33**(6): 1951-1960. DOI: 10.18307/2021.0628. [张运林, 张毅博, 李娜等. 利用陆基高光谱遥感捕捉太湖蓝藻水华日内快速变化过程. 湖泊科学, 2021, **33**(6): 1951-1960.]
- [28] Qiu YG, Duan HT, Wan NS *et al.* Design and practice of a platform for monitoring, early-warning and simulation of algal blooms in Lake Chaohu. *J Lake Sci*, 2022, **34**(1): 38-48. DOI: 10.18307/2022.0102. [邱银国, 段洪涛, 万能胜等. 巢湖蓝藻水华监测预警与模拟分析平台设计与实践. 湖泊科学, 2022, **34**(1): 38-48.]
- [29] Anhui Chaohu Administration Bureau. Natural landscape of Lake Chaohu. <http://chglj.hefei.gov.cn/chwh/18009121.html>. [安徽省巢湖管理局. 巢湖自然景观.]
- [30] Yang LB, Lei K, Meng W *et al.* Temporal and spatial changes in nutrients and chlorophyll-*a* in a shallow lake, Lake Chaohu, China: An 11-year investigation. *Journal of Environmental Sciences*, 2013, **25**(6): 1117-1123. DOI: 10.1016/S1001-0742(12)60171-5.
- [31] Han ZC. Comparative analysis of flood characteristics and related factors of Lake Chaohu in 2016 and 2020. *Harnessing the Huaihe River*, 2021, (1): 7-9. [韩志诚. 巢湖2016年、2020年洪水特性及相关因素对比分析. 治淮, 2021, (1): 7-9.]
- [32] Zhang M, Kong FX. The process, spatial and temporal distributions and mitigation strategies of the eutrophication of Lake Chaohu (1984–2013). *J Lake Sci*, 2015, **27**(5): 791-798. DOI: 10.18307/2015.0505. [张民, 孔繁翔. 巢湖富营养化的历程、空间分布与治理策略(1984—2013年). 湖泊科学, 2015, **27**(5): 791-798.]
- [33] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *Computer Vision and Pattern Recognition*, 2014.
- [34] Breiman L. Random forests. *Machine Learning*, 2001.
- [35] Belgiu M, Drăguț L. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, **114**: 24-31. DOI: 10.1016/j.isprsjprs.2016.01.011.
- [36] Pedregosa F, Varoquaux G, Gramfort A *et al.* Scikit-learn; Machine Learning in Python. *Journal of Machine Learning Research*, 2011, **12**(85): 2825-2830.
- [37] Xiao PF, Li CX, Zhu L *et al.* Multitemporal ensemble learning for snow cover extraction from high-spatial-resolution images in mountain areas. *International Journal of Remote Sensing*, 2020, **41**(5): 1668-1691. DOI: 10.1080/01431161.2019.1674458.