引文格式：田晨, 张金龙, 董世元, 等. 基于遥感图像的蓝藻提取算法探析[J]. 地理空间信息, 2023, 21(10): 80-84. doi:10. 3969/j.issn. 1672-4623.2023.10.020

基于遥感图像的蓝藻提取算法探析

田　晨[1]，张金龙[1]，董世元[2*]，张乃祥[2]
（1. 苏州市水利水务信息调度指挥中心，江苏 苏州 215011；2. 苏州中科蓝迪软件技术有限公司，江苏 苏州 215163）

摘　要：基于卫星遥感的蓝藻精细化分类及信息提取在内陆湖泊水环境治理、监测等方面具有重要的研究意义和应用价值。对Sentinel-2卫星图像数据，依托支持向量机分类模型，探析基于像素和面向对象两种分类思想在蓝藻遥感监测领域的适用情况。研究结果表明：①从不同评价指标上看，基于像素和面向对象的分类方法，分类精度相差不大，均高于90%；②从蓝藻空间分布结果上看，基于像素的分类方法在蓝藻零星分布区域有较好的结果，面向对象的分类方法则适用于蓝藻大范围爆发情况。应用面向对象分类方法对2020—2021年阳澄湖蓝藻进行自动提取，蓝藻提取结果能够表征蓝藻实际分布情况，阳澄湖蓝藻爆发区域主要集中于阳澄西湖北部，阳澄西湖南部、阳澄中湖北部区域存在零星蓝藻爆发情况。

关键词：支持向量机；面向对象；蓝藻；哨兵2号
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2023）10-0080-05

Analysis of Cyanobacteria Extraction Algorithm Based on Remote Sensing Images

TIAN Chen[1], ZHANG Jinlong[1], DONG Shiyuan[2], ZHANG Naixiang[2]
（1. Suzhou Water Conservancy and Water Information Dispatching Command Center, Suzhou 215011, China;
2. Suzhou Land Think Software Technology Company Limited of Chinese Academy of Science, Suzhou 215163, China）

Abstract: Fine classification and information extraction of cyanobacteria based on satellite remote sensing have important research significance and application value in inland lake water environment treatment and monitoring. Based on Sentinel-2 satellite image data and support vector machine classification model, we analyzed the application of pixel-based and object-oriented classification ideas in the field of cyanobacteria remote sensing monitoring. The results show that ①from the perspective of different evaluation indicators, these two methods have little difference in classification accuracy, both higher than 90%. ②From the spatial distribution results of cyanobacteria, the pixel-based classification method has better results in the sporadic distribution area of cyanobacteria, while the object-oriented classification method is suitable for the large-scale outbreak of cyanobacteria. The object-oriented classification method was applied to the cyanobacteria extraction in Yangcheng Lake from 2020 to 2021, and the automatic extraction results could represent the actual distribution of cyanobacteria. The areas where sporadic cyanobacterial outbreaks occur in Yangcheng Lake are mainly concentrated in the northern part of Yangcheng West Lake, while sporadic cyanobacterial outbreaks occur in the southern part and the northern part of Yangcheng Central Lake.

Key words: support vector machine, object-oriented, cyanobacteria, Sentinel-2

面向对象的分类方法越来越多的应用于陆域地物分类中，但在水环境地物识别领域面向对象是否适用等问题现有研究资料较少[1-5]。因此，本文利用Sentinel-2遥感图像数据，以江苏省苏州市阳澄湖为研究区域，设计了基于像素和面向对象的两种蓝藻提取方法，其中基于像素的蓝藻提取方法，直接利用支持向量机进行蓝藻提取；面向对象的分类方法，则是先采用简单线性迭代聚类（SLIC）算法对遥感数据进行分割形成超像素对象，然后以对象为最小研究单元利用支持向量机进行蓝藻目标提取。利用多精度评价指标和实际应用情况，对比基于像素方法和面向对象的分类方法在水环境蓝藻提取领域的优缺点，为水环境监测与治理提供技术参考。最后利用面向对象的分类方法，对2020—2021年的阳澄湖蓝藻进行自动提取，并分析阳澄湖蓝藻聚集特点。

1　数据与方法

1.1　研究区概况

阳澄湖位于江苏省苏州市北部，119°51′32"~120°36′10"E，30°55′40"~30°32′58"N之间，阳澄湖东西最

收稿日期：2022-05-05。
项目来源：苏州市水利水务科技项目（2021009）。
第一作者简介：田晨（1982—），硕士，高级工程师，主要从事环境水利学方向工作，E-mail：2634062009@qq.com。
通信作者：董世元（1994—），硕士，主要从事遥感信息提取、机器学习、深度学习等工作，E-mail：dongsyrs@163.com。

大宽度约8 km，南北长17 km，占地约120 km^2。阳澄湖由于富营养化引起的蓝藻爆发是影响水生态安全、生物多样化的重要因素，严重威胁了苏州市、昆山市饮用水的安全[11]。

1.2　数据获取及处理

本文利用Sentinel-2数据进行阳澄湖的水生植被提取，考虑到水环境治理需求，选用Sentinel-2空间分辨率为10 m的波段数据，分别为可见光波段的Band2（496.6 nm）、Band3（560.0 nm）、Band4（664.5 nm），近红外波段的Band8（835.1 nm）。Sentinel-2可通过网站（https：//scihub.copernicus.eu/dhus/#/home）进行下载，将下载的L1C数据，利用官方推荐的Sen2cor软件对其进行大气校正，可得到10 m分辨率的反射率数据。经阳澄湖实地踏勘以及文献调研[6]，蓝藻爆发主要集中在7—10月份之间，因此本文主要针对2020年、2021年的7—10月份遥感数据进行蓝藻监测。获取的无云或少云的覆盖阳澄湖Sentinel-2数据，2020年共计8景，2021年共计5景。

1.3　特征集构建及水体类型划分

参考前人针对蓝藻信息提取方面的研究，发现应用光谱特征可以直接将水体和水生植被进行区分，但蓝藻和其他挺水植物光谱曲线接近，部分还存在同谱异物的情况，另外在蓝藻聚集程度较低的区域，难以将水体和蓝藻进行有效区分。本文选取在植被、蓝藻有较好表征能力的NDVI，以及能够更好区分水体和植被的NDWI作为分类参考依据。特征变量为：Band2、Band3、Band4、Band8、NDVI、NDWI。其中NDVI、NDWI具体计算方法详见表1。

表1　特征变量概述

指标	特征变量	计算公式
植被指数	NDVI	（Band8-Band4）/（Band8+Band4）
水体指数	NDWI	（Band8-Band3）/（Band8+Band3）

研究区域内蓝藻分布较为复杂，本研究结合实地踏勘结果对图像数据进行目视解译。根据研究需要，将阳澄湖湖区范围内地物类型分为水体、蓝藻、其他植物、云4种类型。

1.4　支持向量机

选用对于样本数量依赖性小且应用广泛的支持向量机（SVM）算法[1-2]，作为蓝藻识别模型基础算法。SVM属于监督学习的一种，算法核心思想是：在特征空间上建立一个可以划分不同类别并且几何间隔最大的最优超平面。其优点是：可以进行小样本学习，不涉及概率测度及大数定律，简化分类和回归的运算过程；在训练过程中利用少数支持向量，可以捕捉关键样本，避开冗余信息和少量的错识别信息，模型具有较好的鲁棒性[8]。

1.5　图像分割方法

本文选用在常用超像元分割算法中精度较高的简单线性迭代聚类（SLIC）算法，该算法由Radhakrishna Achanta等在2011年提出，在医学图像上能够清晰分割超像元，算法核心是检索邻域范围内的相似像元作为一个分割单元[7]。

遥感图像与传统数字图像不同，传统图像分辨率高且边界明显，遥感图像分辨率低，但光谱信息丰富。为整合遥感数据的光谱信息，参考遥感时空融合算法中的光谱相似性度量以及距离度量计算方法，以此来评价像元之间的相似性[10-11]。

1.6　精度评价

在模型精度评估过程中，利用定量化指标评估识别模型精度。利用精确率Precision评价各类别识别精度，召回率Recall评价各类别正确识别召回情况，*F*1-Score和Kappa系数评价每期总体分类精度。

精确率Precision计算方法：

$$\text{Precision}=\frac{TP}{FP+TP} \tag{2}$$

式中，*TP*为正确分类的数量，*FP*为将其他地物误识别为某一类别的数量。

召回率Recall计算方法：

$$\text{Recall}=\frac{TP}{TP+FN} \tag{3}$$

式中，*FN*为某一类地物误识别为其他地物的数量。

*F*1－Score(*F*1)，能够综合精确率和召回率2个指标，计算方法：

$$F1=2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{4}$$

Kappa系数计算方法：

$$\text{Kappa}=\frac{\text{Acc}-Pe}{1-Pe} \tag{5}$$

式中，Acc为准确率，即为正确分类的总数量与样本总数的比值。*Pe*具体计算方式为：

$$Pe=\frac{a_1\times b_1+a_2\times b_2+\ldots+a_c\times b_c}{n\times n} \tag{5}$$

式中，*a*为真实样本数量；*b*为预测得到的样本数量；*n*为总数量。

2　技术路线

本文研究路线如图1所示。

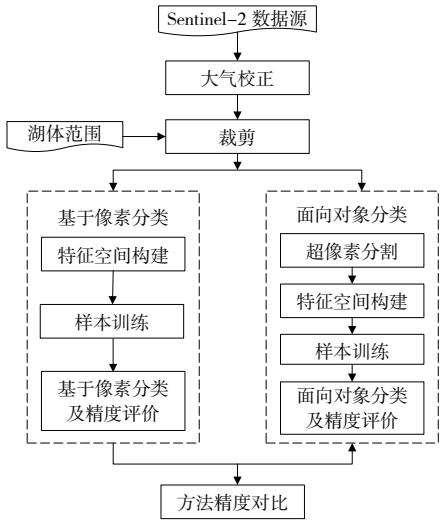
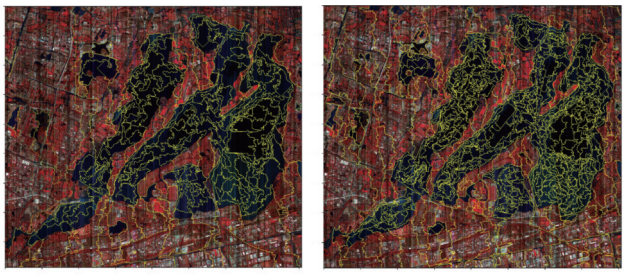


图1 技术路线



a SLIC1 000个超像元分割结果 b SLIC3 000个超像元分割结果

图2 SLIC超像元分割

3 实验结果与分析

3.1 图像分割结果

图2展示了SLIC算法设定超像元个数 S 为1 000、3 000时得到的超像元划分结果，从图中可直观地看出超像元聚类同一类地物对阳澄湖图像进行分割。经测试，超像元数量 S 设置为 2×10^5 个，即每个超像元包含原始数据约18个像元，每个分割对象单元内地物类别较为单一。

3.2 分类结果对比分析

SVM和SLIC-SVM各类别的精确率、召回率、$F1-Score$ 如表2所示，整体来看：水体、云识别精度最高，准确率、召回率、$F1-Score$ 都在96%以上；其他植被和蓝藻的识别精度稍低，但也基本保持在90%以上。对比不同地物类别可以看出：水体SVM方法的精确率高于SLIC-SVM，后者相较于前者约低于5%，两种分类方案召回率相差不大，均接近100%；厚云的SVM和SLIC-SVM的精确率较为一致，接近于100%，但后者相较于前者召回率低4%；蓝藻的识别精确率和召回率两个实验方案相差不大，分别在96%和98%左右；其他植被识别中SVM和SLIC-SVM两者的精确率相差不大，均在99%左右，但方案二SLIC-SVM低于前者约8%。对比不同识别方案，SVM的识别结果优于SLIC-SVM，水体和云有明显的识别特征可实现高精度分类，蓝藻、其他植被分布情况较为复杂分类结果相对较差，但识别精确率和召回率也能保持在90%以上，其中SVM和SLIC-SVM其他植被识别精度差异最大。

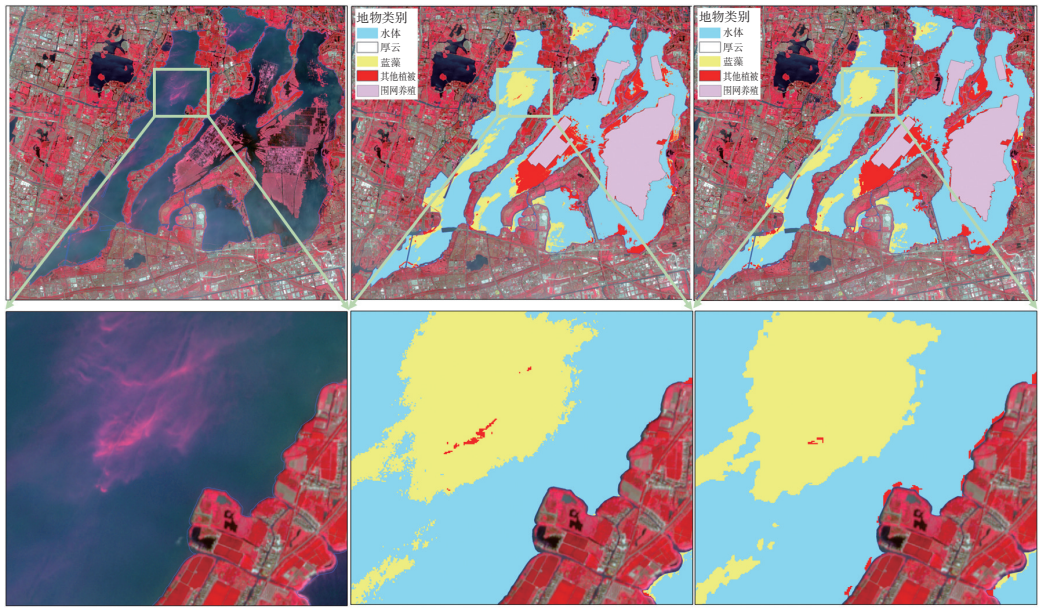
表2 SVM、SLIC-SVM在不同类别分类精确率和召回率的对比/%

变量	精确率		召回率		$F1$	
	SVM	SLIC-SVM	SVM	SLIC-SVM	SVM/%	SLIC-SVM
水体	99.96	94.63	100	99.93	99.97	97.21
云	100	99.32	100	96.05	100	97.66
蓝藻	96.99	96.25	98.14	97.25	97.56	96.75
其他植被	98.53	99.39	97.53	89.45	98.02	94.16

2021-08-26遥感图像数据蓝藻爆发范围较大，利用基于像素方法和面向对象方法对蓝藻进行提取，两种方法Kappa系数分别为：80.44%、82.84%，SLIC-SVM算法精度高于SVM算法。精确率蓝藻最高在93%以上，水体次之，其他植被最低在79%以上，召回率水体最高在97%以上、其他植被次之、蓝藻最低在72%以上。

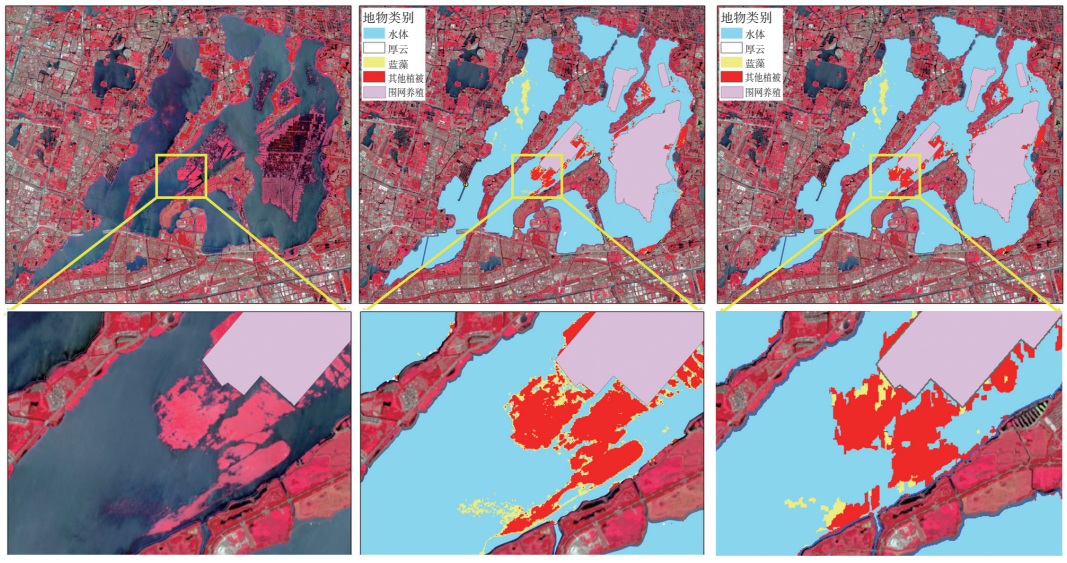
图3b为利用利用指数-阈值方法，并结合人工标注的蓝藻提取成果，图3c、d分别展示了SVM和SLIC-SVM的分类结果，可以从第一行看出3种方法提取的蓝藻信息整体分布情况一致。在第二行细节展示中，人工标注成果存在一定的主观性，如左下角蓝藻浓度聚集程度较低，但标注有蓝藻存在如图3b左下角蓝藻标识区域，而其他浓度聚集程度低区域未标注如图3b左侧区域；SVM算法在蓝藻聚集程度较高的区域将蓝藻误识别为其他植被。利用SLIC-SVM方法分类时，由于前期进行了超像元分割，抵消了部分由于蓝藻聚集程度过高而导致反射率高，造成将蓝藻、水体及其他植被误识别的问题，从而提高了蓝藻识别精度。

图4展示了2020年8月1日的SVM和SLIC-SVM的分类成果，第一行展示了阳澄湖整体分类结果，可以发现图4bSVM和图4cSLIC-SVM的识别蓝藻区域分布基本一致，第二行对比了差异明显的阳澄湖中湖区域，观察到蓝藻零星出现的区域，由于SILV-SVM进行了超像素分割，将零星分布的蓝藻合并为超像元，蓝藻和其他植被片状呈现，有较好的目视效果，但精度稍差，尤其对于零星分布较多的蓝藻，被超像元分割为蓝藻或水体，抹除细小地物的纹理



a 标准假彩色　　b SVM分类成果　　c SILC- SVM分类成果

图3　2021年8月26日识别结果对比



a 标准假彩色　　b SVM分类成果　　c SILC-SVM分类成果

图4　2020年8月1日识别成果对比

信息；SVM是基于逐像元进行分类识别，因此最大程度上保留了原始数据包含的光谱分类信息，能够保持较高的识别精度，但是识别成果易受原始数据噪点影响出现碎斑。

4　2020—2021年蓝藻提取成果

通过1.2节分析可知，基于支持向量机构建面向对象的蓝藻提取方法在蓝藻大范围聚集爆发情况下，具有较好的应用效果。选取2020—2021年有蓝藻爆发且无云的Sentinel-2图像数据，进行蓝藻信息提取，共计提取8期蓝藻空间分布成果图如图5所示。在阳澄湖湖体范围内，蓝藻爆发区域主要集中于阳澄西湖北部，阳澄西湖南部、阳澄中湖北部有蓝藻零星分布，阳澄东湖基本无蓝藻。因此在阳澄湖水环境蓝藻治理过程中，应当以阳澄西湖蓝藻治理打捞为主。

5　结　语

本文以江苏省苏州市阳澄湖为研究区域，利用Sentinel-2遥感图像数据对蓝藻信息进行提取。依托支持向量机分类算法，结合SLIC超像元分割方法对遥感数据进行分割实现面向对象的分类方法。对比传统

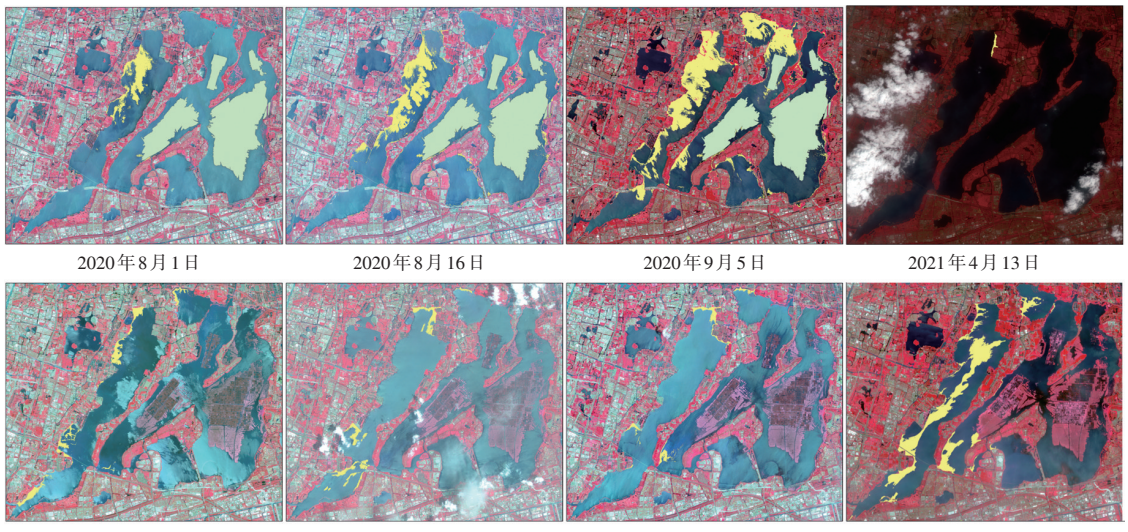


图5 2020–2021年蓝藻空间分布情况

像素和面向对象的2种分类思想在水环境监测中的优缺点，得到以下结论：

1）传统像素的分类思想，其优点是：在地物破碎程度高，纹理边界不清晰的情况下精度较高，如2020年8月1日的分类结果，蓝藻零星分散于水体、其他植被中；其缺点表现，虽然像素级的分类方法能够反映各个像元的特征，以2021年数据为例，得到的结果细碎化程度高，椒盐程度明显，且易受原始数据噪声的影响。

2）面向对象的分类思想，其优点是，能够消除类别之间特征接近和原始数据噪声的问题，解决像素级分类结果的碎斑问题，且蓝藻识别结果边界清晰；其缺点也较为明显，在水环境中，植被空间纹理不明显，例如2020年的分类结果，利用分割算法难以真正的划分有效边界。

3）相较于传统的指数–阈值分类方法，使用机器学习方法进行蓝藻信息提取，避免了人为调节阈值的主观判断，利用机器学习方法提取蓝藻信息，其结果具备较强的客观性。

4）阳澄湖蓝藻空间分布情况，本文获取了2020—2021年的Sentinel–2数据，并进行蓝藻信息提取，发现阳澄湖蓝藻爆发区域主要集中于阳澄西湖北部，且蓝藻爆发较为频繁。

综上所述，近年来较为流行的面向对象分类方法，识别结果边界清晰，但用于地物离散度较高的区域，精细地物易被邻近大块地物概化，但同时能抵消部分“椒盐噪声”的影响。在蓝藻信息提取过程中，面向对象的分类方法容易忽略小部分聚集的蓝藻，但对于零星的图像噪声有较好的抑制性。

参考文献

[1] 李旭青,刘世盟,李龙,等. 基于RF算法优选多时相特征的冬小麦空间分布自动解译[J]. 农业机械学报,2019,50(6):218-225

[2] 李敏,刘国栋,谭凌. 基于随机森林的土地利用分类与景观格局分析[J]. 地理空间信息,2022,20(2):51-56

[3] 胡潭高,潘耀忠,张锦水,等. 基于线性光谱模型和支撑向量机的软硬分类方法[J]. 光谱学与光谱分析, 2011, 31(2):508-511

[4] 赵莲,张锦水,胡潭高,等. 变端元混合像元分解冬小麦种植面积测量方法[J]. 国土资源遥感, 2011,23(1):66-72

[5] 黄邵东,徐伟恒,熊源,等. 结合纹理和空间特征的多光谱图像面向对象茶园提取[J]. 光谱学与光谱分析,2021,41(8):2 565-2 571

[6] 王振华,张鑫月,刘智翔,等. 遥感地物分割的改进格子玻尔兹曼并行模型[J]. 遥感信息,2021,36(4):1-6

[7] 徐兆丰,袁宇欣,彭浩,等. 一种面向对象的城市湖泊遥感提取指数[J]. 地理空间信息,2022,20(1):85-88

[8] 孙淑文,夏霆,杨金艳,等. 苏州湖泊水源地浮游植物群落特征及营养状况分析[J]. 中国农村水利水电,2020(8):51-55

[9] 刘瑞清,李加林,孙超,等. 基于Sentinel-2遥感时间序列植被物候特征的盐城滨海湿地植被分类[J]. 地理学报,2021,76(7):1 680-1 692

[10] 丁世飞,齐丙娟,谭红艳. 支持向量机理论与算法研究综述[J]. 电子科技大学学报,2011,40(1):2-10

[11] 董世元,张文娟,许君一,等. ESTARFM相似像元选取方法的改进研究[J]. 遥感技术与应用,2020,35(1):185-193