

顾天红,杜小玲,李力,等. 基于纹理特征和机器学习的卫星云图分类实验[J]. 中低纬山地气象, 2023, 47(5): 50 - 55.

基于纹理特征和机器学习的卫星云图分类实验

顾天红,杜小玲,李力,朱育雷,张艳梅,吴昌航

(贵州省气象台,贵州 贵阳 550002)

摘 要:准确识别云对提升天气预报和气候预测准确性有着重要意义。传统的阈值法和聚类法很难找出统一通用的阈值标准和方法,随着机器学习在云分类领域的应用和发展,分类速度和分类精度都有了明显提升。该实验对风云二号 G 星的红外云图进行预处理并构建卫星云图样本库,通过提取云图纹理特征再结合支持向量机(SVM)、随机森林(RF)和 XGBoost 分类器实现对“晴空”、“层积云或高积云”、“积雨云”、“密层云”和“卷层云”的分类。实验结果表明:(1)3 种分类器对该实验云分类的平均准确率分别为 RF(62.5%) > XGBoost(61.7%) > SVM(60.0%);(2)3 种分类器对“层积云或高积云”的分类都最好且稳定,平均分类精度均达到了 90% 以上,最高为 91.5%;(3)SVM 对密层云(67.9%)、RF 对卷层云(68.9%)、XGboost 对晴空(68.3%)的分类效果次之,平均分类精度均达 67% 以上。

关键词:灰度共生矩阵;纹理特征;卫星云图;支持向量机;随机森林

中图分类号:P426.5⁺6 **文献标识码:**A

Classificaiton of Satellite Cloud Image Based on Texture Features and Machine Learning

GU Tianhong, DU Xiaoling, LI Li, ZHU Yulei, ZHANG Yanmei, WU Changhang

(Guizhou Meteorological Observatory, Guiyang 550002, China)

Abstract: Accurate identification of clouds is of great significance to improve the accuracy of weather forecast and climate prediction. It is difficult for traditional threshold method and clustering method to find a unified and universal threshold standard and method. With the application and development of machine learning in the field of cloud classification, the classification speed and classification accuracy have been significantly improved. In this experiment, the infrared cloud image of FY - 2G was preprocessed and the satellite cloud image sample library was constructed. The texture features of the cloud image were extracted and combined with the SVM, RF and XGBoost classifiers to realize the classification of "clear sky", "stratocumulus or altocumulus", "cumulonimbus", "dense stratus" and "cirrostratus", and the experimental results show that: (1) the average accuracy of the three classifiers for the cloud classification is RF (62.5%) > XGBoost (61.7%) > SVM (60.0%). (2) The classification of "stratocumulus or altocumulus" is the best and the most stable, and the average classification accuracy reaches more than 90%, the highest is 91.5%. (3) SVM classification of dense stratus cloud(67.9%), RF classification of cirrostratus cloud(68.9%) and XGboost classification of clear sky(68.3%) are the next best, with an average classification accuracy of more than 67%.

Key words: GLCM; texture feature; satellite cloud image; SVM; RF

收稿日期:2022 - 10 - 29

第一作者简介:顾天红(1994—),女,工程师,主要从事短时临近天气预报和决策气象服务工作,E-mail:carolgh@163.com。

通讯作者简介:杜小玲(1970—),女,正研,主要从事中短期天气预报及暴雨、冻雨研究,E-mail:13985141480@163.com。

资助项目:贵州省气象局研究型业务关键技术公关团队(GGTD - 202212);贵州省山地气候与资源重点实验室基金项目(QHLSSLJ[2022] - 12)。

0 引言

卫星云图以其观测范围广、测量物不受测量系统干扰等特性,越来越广泛地应用于常规天气预报和气象科学研究中,在森林防火、资源调查和环境监测等领域也占据着十分重要的地位^[1-2]。云是气象预报中至关重要的观测要素之一,“看云识天气”,不同类型的云对天气的指示作用不同,所以准确识别云对提升天气预报和气候预测准确性有着重要意义^[3-4]。国际云气候计划 (ISCCP) 根据云顶气压将云主要分为高云、中云和低云^[5]; 基于结构形式、范围大小、边界形状、色调、暗影和纹理 6 种判据,在卫星气象学中又分为较为常见的 4 种:层云、卷云、积云和积雨云^[6]。云的自动识别最早由 Shenk 等^[7] 提出,随着图像处理和模式识别技术的飞速发展,基于云图的分类和识别成为新的研究热点。传统的云检测和云分类分为阈值法和聚类法,由于卫星图像随时间、地点、地表而变化,很难找出通用的阈值标准和聚类方法。近年来,机器学习在卫星云图分类识别领域受到关注并迅速发展^[8],经过大量学者的对比研究,发现基于机器学习的分类无论是分类速度还是分类精度都有明显提升。刘扬^[9] 基于阈值法和支持向量机对 7 类云进行了分类;陈晨^[10] 研究对比了极限学习机和支持向量机在卫星云图中的云分类,发现极限学习机在分类速度上有明显优势,但分类精度低于 SVM;邵丽群^[11] 将堆叠极限学习机应用于卫星云图自动分类,发现极

限学习机与支持向量机相比,在泛化能力和学习速度上有明显优势,但分类精度仍低于支持向量机;田文哲等^[12] 采用自适应模糊支持向量机来实现对晴空区、低云、中云、高云及直展云的分类,识别准确率均优于标准支持向量机和传统模糊支持向量机;郑宗生等^[13] 提出基于参数迁移的卫星云图下台风等级的分类方法,发现分类精度明显高于浅层卷积神经网络和源模型;张敏靖等^[14] 结合对抗学习和迁移学习的卷积神经网络进行灾害天气云图分类研究,显著提升了各类灾害天气的分类精度。而卫星云图中的云的分类和识别,实质上是对图像进行切割和特征提取等处理^[15]。纹理是物体结构的反映,分析纹理特征能获取图像综合信息,故可以通过提取云图纹理特征来实现对云的分类识别^[16-17]。

本文采用提取云图的纹理特征参数作为云图样本集的特征向量,再分别用支持向量机、随机森林和 XGboost 分类器进行云的分类识别实验,旨在为云团的追踪识别奠定基础。

1 资料与方法

1.1 资料来源

因目前尚无云图数据库公开可用,本文采用自建云图库。考虑到尽可能降低人为判断识别的主观性,本文参考风云卫星遥感数据服务网站 (<http://satellite.nsmc.org.cn/PortalSite/Default.aspx>) 上 FY-2G 云分类产品 (图 1), 将云分类成以下 5 种:晴空、层积云或高积云、积雨云、密层云和卷层云。

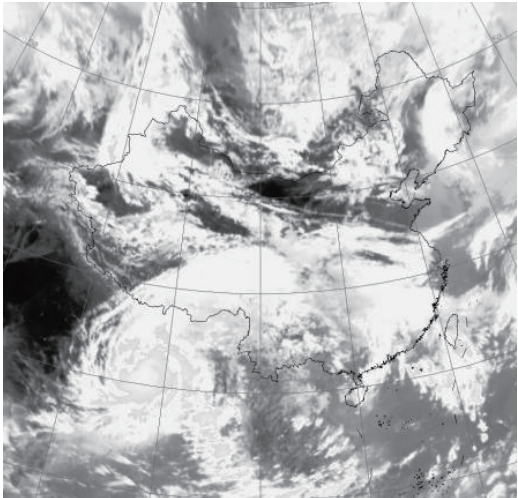
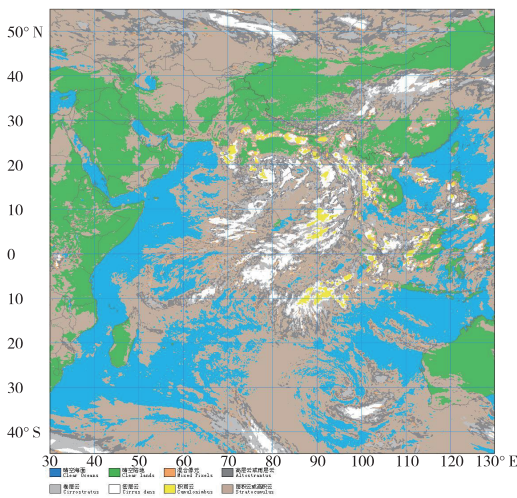


图 1 同一时次 FY-2H 云分类产品与 FY-2G 红外云图对照

Fig. 1 Comparison of FY-2H cloud classification products and FY-2G infrared cloud image at the same time

为使研究对象与观测能力相适应,考虑到红外通道不受日照因素影响,观测连续性佳,本文所用实验数据均来自 2020 年 4—6 月风云二号 G 星的红

外云图 (IR1)。对照同一时次风云卫星遥感数据服务网站 (<http://satellite.nsmc.org.cn/PortalSite/Default.aspx>) 上 FY-2G 云分类产品,对红外云图进行

人为分割分类,统一像素比为 50×50 ,共制作完成了 500 张云型图像,其中晴空、层积云或高积云、积

雨云、密层云和卷层云各 100 张,部分原始数据集如图 2 所示。

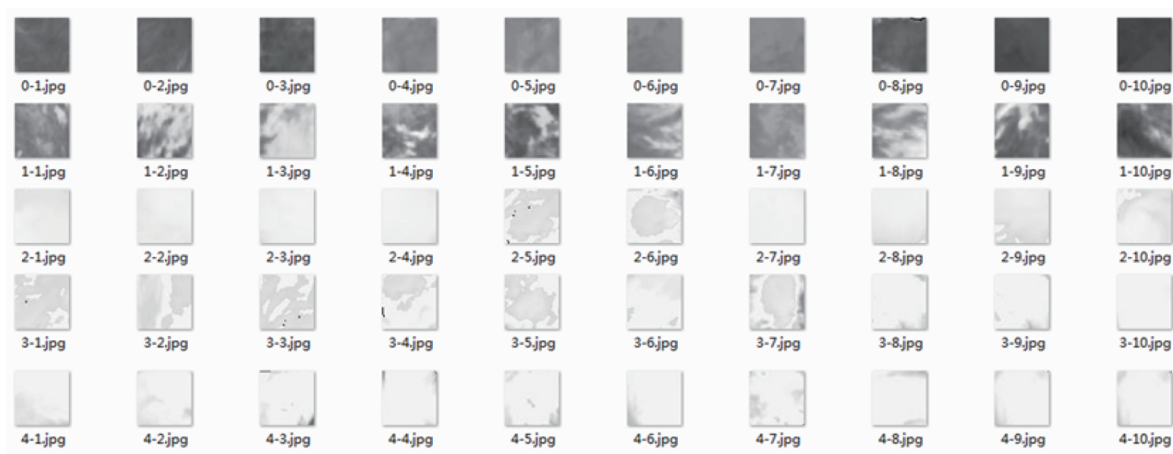


图 2 卫星云图部分原始数据集示意图

Fig. 2 The partial diagram of original satellite cloud image data set

1.2 云图纹理特征及提取方法

目前,提取图像纹理特征的方法主要有 4 种,即基于分布模型参数的模型法、基于纹理基元的结构分析法、基于线性变换和滤波处理的信号处理法和基于图像灰度值的统计分析法^[18]。其中,信号处理法和统计分析法是目前最常用且提取效果较好的 2 种方法^[19]。实践表明,对于类似地形纹理分析^[20]、气象云图分析^[21]等复杂无规律的纹理分析,很多文献中一般使用基于统计分析的典型方法即灰度共

$$P(i, j, d, \theta) = \frac{\#\{[(x_1, y_1), (x_2, y_2)] \in S \mid f(x_1, y_1) = i \& f(x_2, y_2) = j\}}{\#S}$$

式中, d 和 θ 分别为 2 像素点之间的距离和方向(通常取 $0^\circ, 45^\circ, 90^\circ, 135^\circ$),分子代表像素对 (i, j) 的个数,分母代表像素对的总数($\#$ 表示数量),上式得到的 P 是归一化的。将联合概率密度 $P(i, j, d, \theta)$ 写成矩阵形式,即得到 GLCM,其像素对分布示意图如图 3。

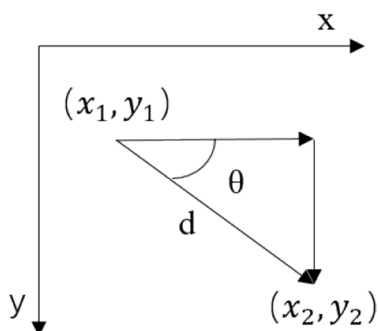


图 3 GLCM 像素对分布示意图

Fig. 3 The distribution diagram of GLCM pixel pair

生矩阵法 (Grey - level Co - occurrence Matrix, GLCM),故本文也采用这一方法进行卫星云图纹理特征分析。

1.2.1 灰度共生矩阵 灰度共生矩阵 (GLCM) 是由 Haralick 提出的一种能够统计图像中每种灰度信息出现频次的纹理特征提取方法^[22],能较为精确地反映图像灰度的方向、相邻距离和变化幅度^[23]。GLCM 定义为图像中 2 个不同像素点灰度分布的联合概率密度 $P(i, j, d, \theta)$ 构成,其计算原理如下^[24]:

1.2.2 基于灰度共生矩阵的纹理属性 尽管 GLCM 能描述图像纹理信息,但为了进行更深入的纹理分析,实际应用中通常不直接使用 GLCM,而是基于 GLCM 再作统计处理,得到的二次统计量再作为图像的纹理特征^[25]。Haralick 定义了 14 种纹理特征参量,但经过 Baraldi 验证,发现对遥感云图特征提取效果最好的 4 种参量为能量、相关性、对比度和熵^[23]。

(1) 能量 (Energy) 是 GLCM 元素值的平方和,反映图像灰度分布均匀性和纹理粗细度。图像纹理越粗,其能量值越大。

$$AMS = \sum_i \sum_j P(i, j)^2$$

(2) 相关性 (Correlation) 是度量 GLCM 元素在行或列上的相似度,反映图像中局部灰度相关性。GLCM 元素越均等,相关性越大。

$$COR = \sum_i \sum_j (i, j)^2 P(i, j)$$

(3) 对比度 (Contrast) 反映图像的清晰度和纹

理沟纹深浅程度。纹理沟纹越深,对比度越大。

$$CON = \frac{\sum_i \sum_j (i - \bar{x})(j - \bar{y})P(i,j)}{\sigma_x \sigma_y}$$

(4)熵(Entropy)是图像随机性的度量,反映图像信息的复杂程度。纹理越复杂,熵越大。

$$ENT = - \sum_i \sum_j P(i,j) \lg P(i,j)$$

2 算法实现

基于 Matlab 编写 GLCM 纹理特征提取代码,基于 Python 3.8 编写云图分类识别代码,其算法步骤如下:

开始→批量读入图像→转化成灰度图像→计算灰度共生矩阵→计算纹理特征统计值(4个方向4个特征值)→计算均值作为最终特征值→基于所提取特征值分别建立 SVM、RF 和 XGBoost 分类模型

→分类效果评估→结束。

3 纹理特征统计分析

利用 GLCM 对晴空、层积云或高积云、积雨云、密层云和卷层云各 100 张云图分别进行纹理特征参数(均为无量纲数)计算,基于箱线图得出对比度、相关性、熵和能量分别在 4 个 θ 方向上的均值的主体范围(图 4)。箱线图的最高点为最大值,最低点为最小值,箱体上框线为 75% 上四分位值,下部框线为 25% 下四分位值,箱中黑横线为中位数,黑点为平均值。综合来看,利用对比度、相关性、熵和能量 4 种纹理特征参数,较容易将层积云或高积云(1)区分出来,而剩下 4 类依靠统计特征难以以人为地区分,故很有必要引入机器学习来进行分类识别。

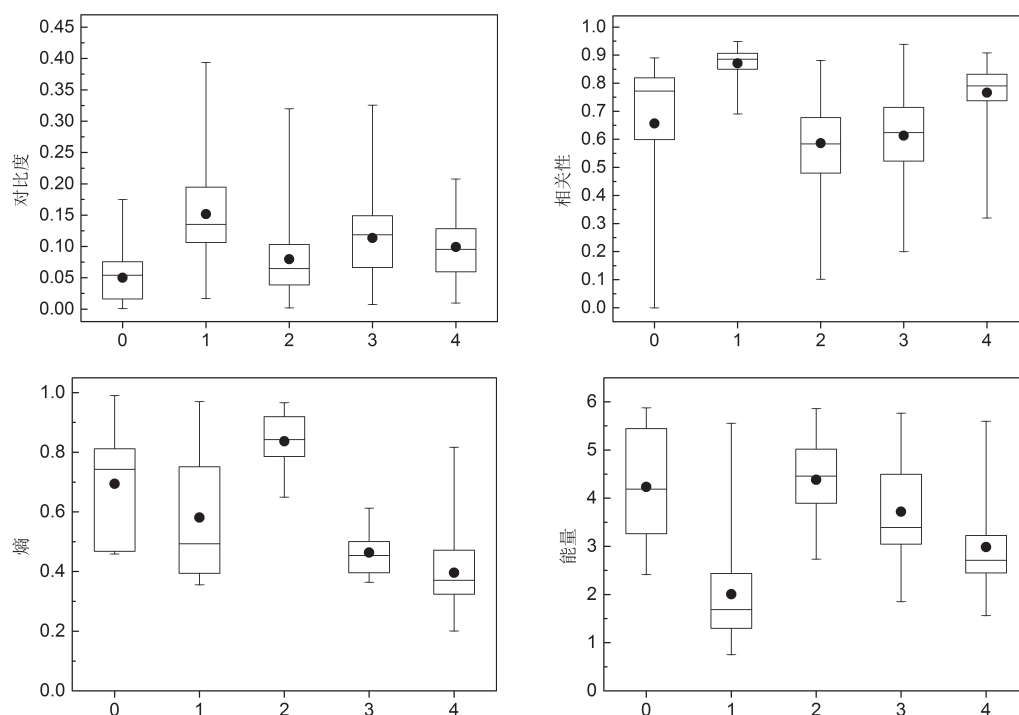


图 4 5 种云分类的对比度(a)、相关性(b)、熵(c)、能量(d)参数箱线图
(图中 0~4 分别为晴空、层积云或高积云、积雨云、密层云、卷层云)

Fig. 4 The box plots of contrast (a), correlation (b), entropy (c) and energy (d) parameters of five cloud classification products (0~4 in the figures are clear sky, stratocumulus or altocumulus, cumulonimbus, dense stratus and cirrostratus)

4 基于机器学习的云分类识别实验

4.1 支持向量机简介

支持向量机(SVM)是 Cortes 和 Vapnik 于 1964 年首次提出的一种基于统计理论用来解决分类问题的分类器^[25],在 20 世纪 90 年代快速发展并不断优化改进,在模式识别中展现出良好的泛化性,在

小样本、高维度和非线性问题上有明显优势^[26]。

4.2 随机森林简介

随机森林是以很多互不关联的决策树为基础构建的一种集成学习(Ensemble Learning)方法,采用 bootstrap 有放回地抽样,让森林里的每棵决策树都对一次抽样进行投票,遵从少数服从多数的原则,综合考虑所有树的决策结果最终得出决策模

型^[27]。随机森林算法因参数少、不易过拟合等优点,被广泛应用于分类和回归问题^[28]。

4.3 XGBoost 简介

XGBoost 以分类回归树 (Classification and Regression Tree, CART) 为基分类器,相关性通过梯度提升来集成,是梯度提升算法的进化版^[29]。XGBoost 对每个特征进行分块 (block) 排序,使其在寻找最佳分裂点的时候能够并行化计算,以提高运行速度^[30]。

4.4 分类结果及分析

构建的云图样本库共 500 个,每一类分别为

表 1 3 种分类器结果对比 (单位:%)

Tab. 1 Comparison of the results of three classifiers (unit:%)

分类器	晴空	层积云或高积云	积雨云	密层云	卷层云	平均准确率
SVM	45.1	90.3	56.1	67.9	48.9	60.0
RF	52.3	91.5	60.0	54.6	68.9	62.5
XGBoost	68.3	90.0	55.6	54.0	51.6	61.7

test Accuracy of SVM:					
		precision	recall	f1-score	support
晴空	0.44	0.47	0.46	17	
层积云或高积云	0.93	0.72	0.81	18	
积雨云	0.56	0.56	0.56	18	
密层云	0.75	0.50	0.60	18	
卷层云	0.48	0.72	0.58	18	
accuracy			0.60	89	
macro avg	0.63	0.59	0.60	89	
weighted avg	0.63	0.60	0.60	89	
test Accuracy of RFC:					
		precision	recall	f1-score	support
晴空	0.57	0.71	0.63	17	
层积云或高积云	0.93	0.72	0.81	18	
积雨云	0.60	0.67	0.63	18	
密层云	0.56	0.50	0.53	18	
卷层云	0.72	0.72	0.72	18	
accuracy			0.66	89	
macro avg	0.68	0.66	0.67	89	
weighted avg	0.68	0.66	0.67	89	
test Accuracy of XGB:					
		precision	recall	f1-score	support
晴空	0.71	0.65	0.68	23	
层积云或高积云	0.90	0.75	0.82	24	
积雨云	0.56	0.41	0.47	22	
密层云	0.56	0.65	0.60	23	
卷层云	0.53	0.73	0.62	22	
accuracy			0.64	114	
macro avg	0.65	0.64	0.64	114	
weighted avg	0.66	0.64	0.64	114	

图 5 Python 3.8 下某次分类结果运行截图

Fig.5 The screenshot of a classification result in Python 3.8

分析可知,RF 对该实验云分类的平均准确率最高,约为 62.5%;XGBoost 次之,约为 61.7%;SVM 分类效果最差,约为 60.0%。其中,SVM 对这 5 种云分类效果依次为:层积云或高积云 (90.3%) > 密层云 (67.9%) > 积雨云 (56.1%) > 卷层云 (48.9%) > 晴空 (45.1%);RF 为:层积云或高积云 (91.5%) > 卷层云 (68.9%) > 积雨云 (60.0%) > 密层云 (54.6%) > 晴空 (52.3%);XGBoost 则为:层积云或高积云 (90.0%) > 晴空 (68.3%) > 积雨云 (60.0%) > 密层云 (54.0%) > 卷层云 (51.6%)。

100 个,分类模型按样本容量的 70% 随机分层抽取作为训练集与测试集,比例为 7:3。样本特征为:“CON”一对比度,“COR”一相关性,“ENT”一熵,“AMS”一能量;样本标签为:“0”一晴空,“1”一层积云或高积云,“2”一积雨云,“3”一密层云,“4”一卷层云。实验构造的 SVM 使用径向基函数 (RBF),C 取 1,g 取“auto”;RF 和 XGBoost 各参数均为默认。3 种分类器对该样本分别进行 10 次训练,最后取其平均准确率作为评估。分类结果见表 1,部分运行结果截图见图 5。

综上,SVM、RF 和 XGBoost 对“层积云或高积云”的分类识别效果都最稳定且最好,均达到了 90% 以上。

5 结论与讨论

本文基于云图纹理特征提取,得出的统计特征参数难以人为区分“晴空”、“层积云或高积云”、“积雨云”、“密层云”和“卷层云”,故结合支持向量机 (SVM)、随机森林 (RF) 和 XGboost 分类器来实现对这 5 类云的分类,得出实验结论如下:

(1) 利用对比度、相关性、熵和能量 4 种纹理特征参数,能较容易地将“层积云或高积云”区分出来,而剩下 4 类依靠统计特征难以人为地区分。

(2) 3 种分类器对该实验云分类的平均准确率分别为 RF (62.5%) > XGBoost (61.7%) > SVM (60.0%)。而在刘扬^[9]的研究中使用的 SVM 云分类器在参数组合为 ($C = 0.5625, g = 0.0$) 时平均准确率最高,为 49.925%。

(3) SVM、RF 和 XGBoost 3 种分类器对本文构建的卫星云图样本库的分类训练结果不一,但对“层积云或高积云”的分类效果都最好且稳定,平均分类精度均达到了 90% 以上,最高为 91.5%;除此之外,SVM 对密层云 (67.9%)、RF 对卷层云 (68.9%)、XGboost 对晴空 (68.3%) 的分类效果次之,平均分类精度均达 67% 以上。而在邵丽群^[11]的研究中,基于光谱特征和纹理特征的堆叠极限学习机 (S-ELMs) 对“层积云或高积云”的分类精度为 96.67%,对“积雨云”为 82.64%,对“晴空”则

为 41.22%。

本实验存在以下问题有待改进:(1)样本容量小;(2)样本特征单一且少;(3)样本库的构建主观性仍难以消除;(4)所选用机器学习分类器未进行调参优化。

参考文献

- [1] 陈健. 卫星云图特征提取与图像显示技术研究及其实现[D]. 哈尔滨:哈尔滨工程大学,2008.
- [2] 姚熠,段莹,李光一,等. 卫星遥感资料在山区洪涝灾害监测及决策服务中的应用[J]. 中低纬山地气象,2022,46(2):76-81.
- [3] 黄琰,赵建,张辉. 望谟县3次特大致洪暴雨卫星云图分析[J]. 贵州气象,2015,39(4):32-36.
- [4] 贾克斌,张亮,刘鹏宇,等. 基于轻量级神经网络的地基云图识别[J]. 北京工业大学学报,2021,47(5):489-499.
- [5] 胡凯,严昊,夏旻,等. 基于迁移学习的卫星云图云分类[J]. 大气科学学报,2017,40(6):856-863.
- [6] 王猛,何丽莉,白洪涛,等. 基于卫星云图的DBSCAN聚类云团分类方法[J]. 吉林大学学报(理学版),2016,54(1):91-99.
- [7] 赵晓利. 气象卫星云图云检测及分类的研究[D]. 广州:华南理工大学,2014.
- [8] 丁叶文. 基于深度学习的卫星云图分割和识别研究[D]. 南京:南京信息工程大学,2020.
- [9] 刘扬. 基于静止气象卫星云图的分类研究[D]. 青岛:中国海洋大学,2011.
- [10] 陈晨. 基于ELM和SVM的卫星云图分类研究[D]. 南昌:南昌航空大学,2014.
- [11] 邵丽群. 基于堆叠极限学习机的卫星云图分类研究[D]. 南昌:南昌航空大学,2016.
- [12] 田文哲,符冉迪,金炜,等. 面向卫星云图云分类的自适应模糊支持向量机[J]. 武汉大学学报(信息科学版),2017,42(4):488-495.
- [13] 郑宗生,胡晨雨,黄冬梅,等. 基于迁移学习及气象卫星云图的台风等级分类研究[J]. 遥感技术与应用,2020,35(1):202-210.
- [14] 张敏靖,白琮,张敬林,等. 基于对抗和迁移学习的灾害天气卫星云图分类[J]. 北京航空航天大学学报,2021,47(3):585-595.
- [15] 刘玉婷,李子阳,李恒凯,等. 融合物候和纹理特征的南方柑橘林地提取方法[J]. 测绘科学,2021,46(9):83-93.
- [16] 李冰洁,李国东. 基于灰度共生矩阵与AdaBoost判别法的冰雹预测[J]. 贵州气象,2017,41(1):35-39.
- [17] 寿亦萱,张颖超,赵忠明,等. 暴雨过程的卫星云图纹理特征研究[J]. 南京气象学院学报,2005(3):337-343.
- [18] 李冰洁. 气象卫星系统的云图自动分类识别研究[D]. 西安:西安科技大学,2019.
- [19] 杨帅. 基于灰度共生矩阵的地震纹理属性分析及其应用[D]. 北京:中国石油大学(北京),2018.
- [20] 张昊,石晓荣,倪亮. 基于分形和纹理特征的红外图像云检测[J]. 计算机仿真,2020,37(8):452-456.
- [21] 刘光宇,黄懿,曹禹,等. 基于灰度共生矩阵的图像纹理特征提取研究[J]. 科技风,2021(12):61-64.
- [22] 胡启明. 基于灰度共生矩阵的地形纹理特征量化研究[D]. 福州:福州大学,2017.
- [23] 王梦薇. 基于MATLAB的医学超声图像处理仿真平台设计与研究[D]. 包头:内蒙古科技大学,2020.
- [24] 窦锦响. 基于改进灰度共生矩阵的图像块复制粘贴篡改检测研究[D]. 兰州:西北师范大学,2020.
- [25] CORTES C, VAPNIK V. Support - Vector networks[J]. Machine Learning,1995(20):273-297.
- [26] 祁亨年. 支持向量机及其应用研究综述[J]. 计算机工程,2004(10):6-9.
- [27] 张玉霞. 基于随机结构森林的桥梁裂缝检测算法研究[D]. 西安:陕西师范大学,2018.
- [28] 王祥,陈发达,刘凯,等. 基于随机森林-支持向量机隧道盾构引起建筑物沉降研究[J]. 土木工程与管理学报,2021,38(1):86-92.
- [29] 苏智杰. 基于XGBoost优化算法的G型恒星分类研究[D]. 南昌:江西财经大学,2021.
- [30] 张泉,罗阿理. 基于XGBOOST的恒星光谱分类特征数值化[J]. 光谱学与光谱分析,2019,39(10):3292-3296.