

基于深度学习的降水现象自动识别方法

李 鹏, 谢婷婷

中国电子科技集团公司第二十七研究所, 河南 郑州

收稿日期: 2021年9月13日; 录用日期: 2021年10月8日; 发布日期: 2021年10月15日

摘 要

基于全国开展的降水天气现象平行观测工作期间积累的数据基础, 收集并整理降水现象仪自动观测数据和人工观测数据, 解析降水现象仪产生的雨滴谱数据文件, 依据降水粒子大小和下落速度分级以及累计时段内粒子数量, 生成与之对应的二维图像, 同时结合降水现象仪平行观测期间的人工观测数据, 利用深度学习图像分类技术进行训练, 建立降水现象自动识别模型, 完成不同降水现象的自动识别。模型训练集和验证集平均准确率均达到86%, 雨的测试准确率达到62.7%, 雨和毛毛雨的总识别率达到89.1%, 雪的测试准确率达到93%, 说明利用深度卷积神经网络对雨滴谱数据生成的雨滴图进行自动降水现象识别方案可行。

关键词

雨滴谱, 降水现象自动识别, 深度学习, 图像分类

Automatic Recognition Method of Precipitation Phenomena Based on Deep Learning

Peng Li, Tingting Xie

The 27th Research Institute of China Electronics Technology Group Corporation, Zhengzhou Henan

Received: Sep. 13th, 2021; accepted: Oct. 8th, 2021; published: Oct. 15th, 2021

Abstract

Based on the data accumulated during the parallel observation of precipitation weather phenomena carried out throughout the country, collect and sort out the automatic observation data and manual observation data of the precipitation phenomometer, analyze the raindrop spectrum data file generated by the precipitation phenomometer to generate the corresponding two-dimensional image. At the same time, combined with the artificial observation data during the parallel observation period of the precipitation phenomenon instrument, the deep learning image

classification technology is used for training, and the automatic recognition model of precipitation phenomenon is established to complete the automatic recognition of different precipitation phenomena. The average accuracy of the model training set and validation set is 86%, the test accuracy of rain is 62.7%, the total recognition rate of rain and drizzle is 89.1%, and the test accuracy of snow is 93%, indicating that the scheme of automatic precipitation recognition based on the raindrop pattern generated by the deep convolutional neural network is feasible.

Keywords

Raindrop Spectrum, Automatic Recognition of Precipitation Phenomena, Deep Learning, Image Classification

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 背景

目前, 全国大部分省(区、直辖市)正开展降水天气现象平行观测工作, 但是从目前全国统计的数据分析发现, 目前安装的降水现象仪对于常见的雨、雪降水现象, 可以捕获到降水过程, 平均捕获率为 70%, 但是错报、漏报率较高; 对于雨夹雪、冰雹这些不常见的天气现象, 捕获率较低, 漏报率较高。现需要统计全国 2242 个台站近几年的真实观测数据, 得到符合各台观测特性的降水类型识别方法。

目前降水现象仪不能满足用户需要, 根本原因在于目前各厂家降水类型识别算法均是依靠国外试验所得经验公式, 按照 32 等级粒径和 32 等级粒速粒子分布情况, 绘制降水类型识别图[1] [2]。在此图中虽然雨区、雪区及冰雹区在尾端有明显的分界线, 而在前端则重叠区域较多; 毛毛雨包含在雨区前端, 雨夹雪区则为雨区和雪区内的中间部分组成[3]。这种方法, 只利用了区域位置信息和简单的粒子数量分级, 并且我国疆域辽阔, 地域性强, 该方法通用性不强, 不能准确分辨天气现象。

本文采用基于深度学习的图像分类方法[4], 首先解析雨滴谱文件, 生成包含粒径、粒速和粒子数量信息的二维图像; 再利用人工观测天气现象对生成的二维图像进行人工标注, 形成数据集; 最后利用深度学习图像分类方法搭建降水现象自动识别模型。

2. 降水现象识别方法

雨滴谱传感器由激光发射装置、激光接收装置、温控装置、处理单元等部分组成, 通过检测通过采样区的降水粒子的尺寸和速度, 获取雨滴谱信息, 区分降水类型[5] [6] [7]。二维图像具有颜色特征、纹理特征、形状特征和空间关系特征。本文把雨滴谱传感器生成的报文按照 32 等级粒径和 32 等级粒速粒子分布的数量, 生成二维图像, 一方面获取了区域位置信息, 另一方面把粒子数量信息转换成像素值, 获取了颜色特征、纹理特征和形状特征。

考虑到目前基于卷积神经网络的深度学习在图像处理领域表现出来的优势及其对特征刻画的优秀表现, 我们的目光就聚焦到了卷积神经网络上, 卷积网络通过自动学习训练数据中的特征, 对特征自动进行抽取和筛选, 从而得到理想的特征提取模型, 相比于人工提取特征, 卷积神经网络排除了人的主观因素[8], 使特征提取更加准确和合理。

本文基于深度学习的降水现象自动识别方法, 首先收集雨滴谱传感器生成的报文, 生成二维图像并

制作成图像分类模型所需的数据集；接着根据降水现象分类需求，搭建神经网络，进行模型训练，生成自动识别模型；最后对模型进行测试，检验模型的可用性。

3. 数据集制作

雨滴谱文件中存储了每日每分钟累计粒子数量。解析雨滴谱文件，每日每分钟生成一幅雨滴图，每日1440张图片，其中当前分钟无天气现象的图片删除，每日最多产生1440张，最少0张。A文件中存储了每日天气现象人工观测结果，精确到分。依据A文件中人工观察结果把雨滴图分类到对应的天气现象类别中。

对50772、50892等23个站提供的雨滴谱文件和A文件进行解析，生成数据集，数据集收集到包含五种天气现象图片435,698张，其中毛毛雨9867张、雨337,595张、雨夹雪9134张、雪79,100张、冰雹2张，详情如表1台站数据汇总表所示。

Table 1. Summary table of station data

表1. 台站数据汇总表

编号	台站号	台站地区	毛毛雨	雨	雨夹雪	雪	冰雹
1	50772			2138	879	9896	
2	50892			6430	554	12,549	2
3	58633			40,815	1241	612	
4	54328	喀左		8229	122	905	
5	50655	五大连池		9330	230	9855	
6	50964	方正	9	5505	742	10,400	
7	50247	黑龙江大兴安岭呼中		8969	580	6884	
8	58550	诸暨		52,613	1271	1317	
9	58457	杭州	3681	17,057	487	1500	
10	58549	金华		24,979	260	985	
11	58642	武义		29,718	412	1451	
12	50659	克东		7139	299	5686	
13	50962	木兰		4070	458	8620	
14	54098	黑龙江省牡丹江市宁安		6307	149	3451	
15	58745			29,832	366		
16	58448	杭州临安	3007	12,444	388	1284	
17	58424	安徽安庆	3170	29,375	233	2479	
18	54597	北京霞云岭		2673		985	
19	54596	北京房山		2471	463	241	
20	56664	云南华坪		10,123			
21	58929	福建安溪		6341			
22	58848	福建连江		10,918			
23	58845	福建罗源		10,119			
		总计:	9867	337,595	9134	79,100	2

虽然样本种类涵盖了全部五种天气现象,但是每一种类样本数量差异较大,雨、雪图片数量多,冰雹图片只有2张,毛毛雨和雨夹雪有1万张左右,且毛毛雨现象只在3个台站出现,样本分布呈现明显的不均衡。挑选了部分具有代表性的二维图像进行分析,详情如下图1五种天气现象二维图所示。从二维图中可以看出,不同的天气现象呈现的二维图像有明显的差异,这为本文采用图像分类方法识别天气现象奠定了基础。



Figure 1. Two-dimensional map of five weather phenomena

图 1. 五种天气现象二维图

基于样本均衡性的考虑,把23个站数据合并在一起组成一个包含4类天气现象的数据集,其中毛毛雨9867张、雨10,000张、雨夹雪9134张、雪11,300张,详情如表2训练数据集所示。

Table 2. Training data set

表 2. 训练数据集

识别类型	数据集大小(张)
毛毛雨	9867
雨	10,100
雨夹雪	9134
雪	10,300

4. 模型搭建与训练

通过分析数据集规模,按照5:1的比例划分训练集和验证集,把数据集中每个样本当作分辨率为(32, 32)

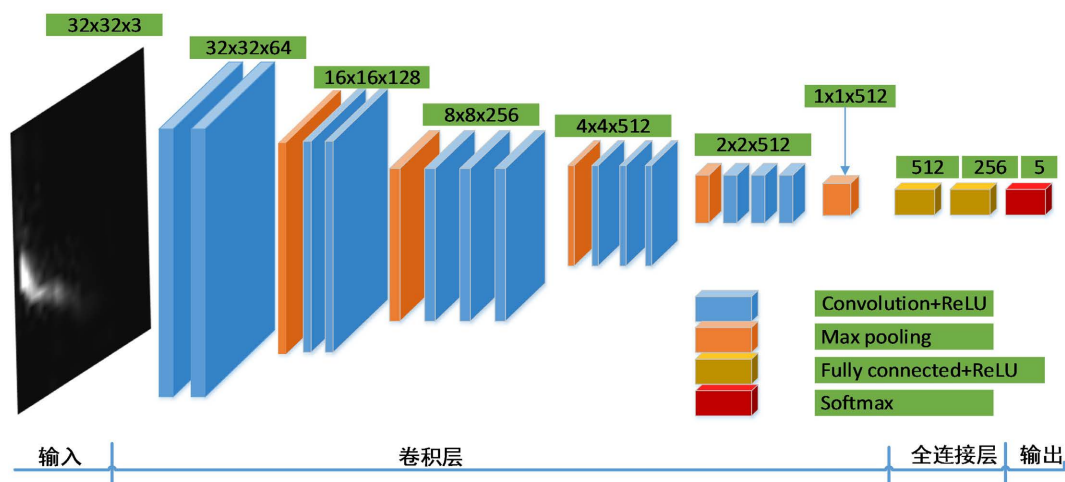


Figure 2. The structure diagram of the five types of precipitation recognition depth model

图 2. 五种降水类型识别深度模型结构图

的图像, 利用图像分类的方法, 针对雨滴谱样本搭建了一个包含 20 层, 15,896,132 个参数的深度网络[9], 上图 2 五种降水类型识别深度模型结构图展示了该网络的结构。该网络输入尺寸(32, 32, 3), 经过多次卷积、池化, 最后通过全连接层变形、压缩, 输出一个(5, 1)的数组。

利用搭建的五种降水类型识别深度模型, 在四万张样本的数据集上展开训练。本实验采用的仿真平台为 64 位 window7 操作系统, 硬件环境为内存 16 G, Intel (R) Xeon (R) E5-1607, GPU NVIDIA Quadro K600。选用的深度学习开源框架为 Tensorflow, 调用 GPU 实现卷积神经网络的并行运算。模型设置训练 epoch = 100, batch_size = 32, 初始学习率采用默认值, 经过 52 小时的迭代学习, 模型训练准确率和验证准确率不再变化, 趋于稳定。下图 3 展示了训练集数据识别的准确率和验证集数据识别的准确率, 随着训练次数增多, 训练平均准确率和验证平均准确率不断上升, 最后稳定在 86%附近。



Figure 3. (a) Recognition accuracy rate of training set; (b) Verification set recognition accuracy

图 3. (a) 训练集识别准确率; (b) 验证集识别准确率

5. 结果及分析

基于样本均衡的考虑, 训练时只用了雨、雪的部分样本, 我们把剩余样本作为测试集, 其中雨的测试样本 320,446 张, 雪的测试样本 67,800 张, 详细情况如表 3 测试数据集所示。利用训练好的最优模型, 对两类测试数据进行测试, 验证下模型的泛化能力。

Table 3. Test data set

表 3. 测试数据集

识别类型	总样本	训练数据集	测试数据集
毛毛雨	9867	9867	0
雨	337,595	10,100	320,446
雨夹雪	9134	9134	0
雪	79,100	10,300	67,800

5.1. 雨样本测试

雨的测试样本 320,446 张, 利用训练的最优模型对雨样测试集进行自动识别, 识别结果毛毛雨 86,682 张, 占总数 26.4%; 雨 198,350 张, 占总数 62.7%; 雨夹雪 30,226 张, 占总数 9.2%; 雪 5188 张, 占总数 1.5%, 详情情况如表 4 雨测试准确率统计表所示。

雨的识别准确率是 62.7%，其中错报率较高的是毛毛雨，达到 26.4%。雨和毛毛雨的总识别率达到 89.1%，说明雨和毛毛雨的样本相似度高，与雨夹雪和雪区分度高。经过分析发现，同一时段人工观测有时候会多个现象同时出现，这个问题直接导致雨的错报率提高。下一步针对这个问题，一方面对数据集做相似度分析，重新筛选样本，提高数据集的可靠性；另一方面做一个二次辨识雨、毛毛雨的模型，提高识别准确率。

Table 4. Accuracy statistics of rain test

表 4. 雨测试准确率统计表

识别类型	样本数	百分率
毛毛雨	86,682	26.4
雨	198,350	62.7
雨夹雪	30,226	9.2
雪	5188	1.5

5.2. 雪样本测试

雪的测试样本 67,800 张，利用训练的最优模型对雪样测试集进行自动识别，识别结果毛毛雨 601 张，占总数 0.8%；雨 331 张，占总数 0.4%；雨夹雪 3785 张，占总数 5.5%；雪 63,083 张，占总数 93%，详细情况如表 5 雪测试准确率统计表所示。

雪的识别准确率是 93%，其中错报率较高的是雨夹雪，达到 5.5%。雪和雨夹雪的总识别率达到 98.5%，说明雪和雨夹雪的样本相似度高，与毛毛雨、雨区分明显。下一步针对这个问题，做一个二次辨识雪、雨夹雪的模型来解决这个问题，识别准确率可以进一步提高。

Table 5. Snow test accuracy rate statistics table

表 5. 雪测试准确率统计表

识别类型	样本数	百分率
毛毛雨	601	0.8
雨	331	0.4
雨夹雪	3785	5.5
雪	63,083	93

5.3. 综合分析

对雨、雪测试结果进行混淆分析结果如表 6 测试结果混淆矩阵统计所示，雨样本被准确识别的概率是 62.7%，其中误判率较高的是毛毛雨，误判雪、雨夹雪的概率低；雪样本被准确识别的概率是 93%，误判率较高的是雨夹雪，误判雨、毛毛雨的概率极小。结果说明该识别模型能够有效区分雨、雪。下一步，对雨、雪再进行二次辨识，即可区分毛毛雨、雨夹雪。即模型第一步先识别雨雪，如果识别为雨，再进一步具体分析是雨还是毛毛雨；如果是雪，再进一步具体分析是雪还是雨夹雪。

Table 6. Confusion matrix statistics of test results

表 6. 测试结果混淆矩阵统计

		识别结果			
		毛毛雨	雨	雨夹雪	雪
真实类别	雨	86,682 (26.4%)	198,350 (62.7%)	30,226 (9.2%)	5188 (1.5%)
	雪	601 (0.8%)	331 (0.4%)	3785 (5.5%)	63,083 (93%)

6. 结论

模型训练集和验证集平均准确率均达到 86%，雨和毛毛雨的总识别率达到 89.1%，雪的识别准确率是 93%，说明利用深度卷积神经网络对降水现象仪数据生成的雨滴图进行自动降水现象识别方案可行，这一方法具有独创性，领先性。随着训练样本数据增多，模型进一步优化，平均准确率会进一步提高。利用深度学习图像分类技术来提高降水现象判定的准确性，从而提高整体自动化观测水平的目标可以实现。

参考文献

- [1] 杜波, 马舒庆, 梁明珠, 刘达新, 张晓宇, 王柏林. 雨滴谱降水现象仪对比观测试验技术应用分析[J]. 气象科技, 2017, 45(6): 995-1002.
- [2] 中国气象局. 地面气象观测规范[M]. 北京: 气象出版社, 2003.
- [3] 文斌青, 陈国强. 浅析 DSG1 型降水现象仪[J]. 农业与技术, 2019, 39(10): 141-142.
- [4] Uchida, K., Tanaka, M. and Okutomi, M. (2018) Coupled Convolution Layer for Convolutional Neural Network. *Neural Networks*, **105**, 197-205. <https://doi.org/10.1016/j.neunet.2018.05.002>
- [5] 文斌青, 陈国强. 浅析 DSG1 型降水现象仪[J]. 农业与技术, 2019(10): 135-136.
- [6] 杨宁, 张晋, 刘钧. 雨滴谱式降水现象仪降水类型判定算法优化探究[J]. 气象科技进展, 2018, 8(6): 89-94.
- [7] 张静, 宋中玲. 降水现象仪与人工观测资料对比分析[J]. 青海气象, 2019(3): 107-112.
- [8] 陆明, 申双和, 王春艳, 等. 基于图像识别技术的夏玉米生育期识别方法初探[J]. 中国农业气象, 2011, 32(3): 423-429.
- [9] Akilan, T., Wu, Q.M.J., Yang, Y., *et al.* (2017) Fusion of Transfer Learning Features and Its Application in Image Classification. 2017 *IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, Windsor, 30 April-3 May 2017, 1-5. <https://doi.org/10.1109/CCECE.2017.7946733>