

基于 ComplexCNN 的龙卷识别算法

董珈畅¹, 王鱼¹, 曾强宇^{1,2,3*}, 黄尚丹¹

1. 成都信息工程大学电子工程学院, 四川 成都 610225

2. 中国气象局龙卷风重点开放实验室, 北京 100000

3. 大气探测重点开放实验室, 四川 成都 610225

DOI: 10.61369/TACS.2026070048

摘 要 : 龙卷风作为一种突发性的强对流天气现象具有极强的破坏力, 对社会造成严重的人员伤亡和经济损失。随着全球极端天气事件的频发, 龙卷风的预警预报成为气象研究的热点问题。目前龙卷风的监测多依赖于天气雷达, 传统非监督学习算法在识别精度和预警时效性方面存在局限性。人工智能技术的快速发展为龙卷风识别提供了新的技术手段。利用 TorNet 数据集和我国龙卷雷达数据, 构造以反射率, 径向速度和速度谱宽的龙卷风数据集, 提出 ComplexCNN 龙卷识别算法, 直接从雷达基数据样本中识别龙卷特征。该模型直接从雷达数据中学习特征, 相比设计特征的龙卷样本简化了数据处理流程, 显著提升了检测概率和临界成功指数, 在江苏省历史龙卷风案例中的平均预警时间达到 20 分钟。

关 键 词 : 龙卷风暴; 天气雷达; 龙卷识别; 深度学习

Tornado Detection Algorithm Based on ComplexCNN

Dong Jiachang¹, Wang Yu¹, Zeng Qiangyu^{1,2,3*}, Huang Shangdan¹

1.College of Electronic Engineering, Chengdu University of Information Technology, Chengdu, Sichuan 610225

2.China Meteorological Administration Tornado Key Open Laboratory, Beijing 100000

3.Key Open Laboratory for Atmospheric Detection, Chengdu, Sichuan 610225

Abstract : Tornadoes are sudden and severe convective weather phenomenon, possessing immense destructive power and potentially causing severe casualties and economic losses. The warning and forecasting of tornadoes has become a hot topic in meteorological research, with the increasing frequency of extreme weather events worldwide. At present, tornado forecasting relies heavily on weather radar, and traditional unsupervised learning algorithms have limitations in tornado detection accuracy and leading time. The rapid development of artificial intelligence technology has provided new technological tools for tornado detection. Using the TorNet dataset and Chinese tornado radar data, a tornado dataset with reflectivity, radial velocity, and velocity spectral width was constructed. The ComplexCNN tornado recognition algorithm was proposed, which directly identifies tornado features from radar base data. This model directly learns features from radar data, simplifying the data processing flow and significantly improving the detection probability and critical success index compared to tornado detection algorithm based on features. The average leading time in historical tornado cases in Jiangsu Province reached 20 minutes.

Keywords : tornado; weather radar; tornado recognition; deep learning

引言

龙卷是具有局地性、突发性的灾害性强对流天气, 由风暴单体云底伸展至地面的漏斗状云产生的强烈的旋风, 风速常在达 40m/s 以上, 极端情况下可超过 100m/s, 影响范围通常在几米到数百米之间, 具有生消迅速、空间尺度小、破坏力极强的特点^[1]。根据龙卷的破坏强度和风速可使用增强藤田级数, 将龙卷分为 EF0~EF5 级^[2]。每年全球发生约 2000 多个龙卷, 其中美国是发生频率最高的国家, 每年龙卷发生超过 1200 次^[3], 中国龙卷发生频率显著低于美国, 约为美国的十分之一, 但中国龙卷多发于经济发达人员稠密的沿海地区, 造成的灾害影响很大^[4]。美国的龙卷通常尺度和强度较大, 超级单体龙卷比较常见, 而中国的龙卷通常尺度较小, 强度相对较弱, 多是单体或多体雷暴下产生的龙卷; 美国龙卷通常在超级单体雷暴或者风暴对流统中形成, 而中国龙卷通常在冷锋、暖锋、飑线和切变线等天气过程中形成^[5]。

项目信息: 国家自然科学基金 (U2342216; 42575154); 四川省自然科学基金 (2026NSFSC0208); 海南省南海气象防灾减灾重点实验室 (SCSF202503)

通信作者: 曾强宇, Email: zqy@cuit.edu.cn

近年来基于人工智能算法龙卷识别开始在我国业务化应用, 有效提高龙卷识别准确率。Wang[6] 提出了基于神经元自适应系统的龙卷自动探测算法 (NFTDA, Neuro-Fuzzy Tornado Detection Algorithm), 利用龙卷的涡旋和谱宽特征结合模糊逻辑算法对美国龙卷进行识别, 识别率达到 86%。XU[7] 利用大涡模拟仿真龙卷碎片特征, 发现了低层龙卷碎片扩散同时降低碎片的速度。QingZ[8] 提出了 ADASYN-LOF algorithm (ALA), 能有效处理龙卷数据集样本不平衡, 提升机器学习龙卷识别算法性能。我国龙卷研究起步较晚, 但近年来在观测技术和识别算法方面取得显著进展。人工智能算法应用于龙卷, 进一步提高了预警预报能力。

一、数据来源与处理

(一) 龙卷数据

我国 2018 年开始对天气雷达进行双偏振升级, 且中国龙卷发生频率较低, 记录的江苏省和广东省 2019-2022 年双偏振雷达的历史数据共计 15 次龙卷风过程。由于多数龙卷风发生在垂直距离 3km 以下^[9], 本文主要使用雷达低层仰角 0.5°、1.5° 和 2.4° 的数据。

TorNet 数据集^[10]用于龙卷风检测和预测, 由林肯实验室联合英伟达构建。TorNet 数据集包含了 10 年的全分辨率、双极化天气雷达数据, 源自美国国家环境信息中心的风暴事件数据库, 保障了龙卷数据的质量和准确性。数据集分为三类样本: 已确认的龙卷风、非龙卷风的龙卷风警报和随机非龙卷风, 这样的分类便于算法学习中提高识别的准确率, 降低误报率。根据官方介绍该数据集包含超过 200,000 组雷达数据, 其中 13,587 个描绘了龙卷风的情况。

(二) 龙卷数据集构建流程

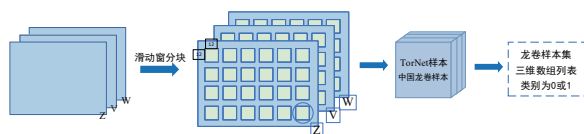


图1 TorNet 龙卷样本制作流程

如图1显示的是龙卷样本集制作流程, 首先按块划分数据, 为取龙卷数据时完整包含整个龙卷路径, 将 TorNet 样本的距离库大小调整成 12×12。随后将针对龙卷位置的三种数据构建维度为 3 (数据类型) × 12 (径向) × 12 (距离单元数) 的数据矩阵。将含有龙卷风涡旋的样本标记为“正样本” (标记为 1), 而不含龙卷风特征的样本则标记为“负样本” (标记为 0)。最终获得了 16,556 个正样本和 82,542 个负样本。为优化模型训练效果, 在训练集中会根据调整正负样本的比例以实现数据平衡。将数据集按照 8:1:1 的比例划分为训练集、验证集和测试集。

二、模型构建与评估

(一) ComplexCNN 龙卷风识别算法

龙卷风识别是中心天气预警的重要方向, 通过识别天气雷达数据, 可及时识别龙卷风并采取应对措施。ComplexCNN 龙卷风识别算法流程包括数据集制作、网络构建、网络训练、参数优化和模型部署等步骤。首先, 利用龙卷数据集, 将数据集划分为训练集、验证集和测试集。其次, 构建网络模型和模型训练, 通

过优化损失函数来调整网络参数。同时为提高模型泛化能力, 可以在训练过程中调整模型的网络参数, 使用学习率调度器在验证集损失不再下降时降低学习率, 调整超参数 (如学习率、批量大小、Dropout 比例等) 和网络结构 (如增加卷积层数量或调整卷积核大小) 搜索最优的模型配置。最后, 要在测试集上评估模型的最终性能, 确保算法具有良好的泛化能力。

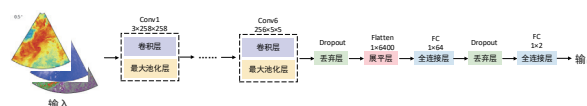


图2 ComplexCNN 模型结构图

如图2所示, ComplexCNN 是一个 21 层的深度卷积神经网络, 其结构包括输入层、多个卷积层、池化层、全连接层和输出层。输入层接收预处理后的龙卷数据集数据, 卷积层通过卷积核提取特征, 每层后接 ReLU 激活函数以引入非线性。池化层用于降低特征图的尺寸, 减少计算量并防止过拟合。全连接层将卷积层提取的特征展开并输入到输出层, 输出层使用 Softmax 激活函数生成最终预测结果。为了提高模型的性能, 在卷积层后加入批归一化层 (Batch Normalization) 以加速训练并提高稳定性, 同时使用 Dropout 层防止过拟合, 有效减少模型对训练数据的过度依赖, 还可以通过随机丢弃神经元增强泛化能力。

模型的输入大小为 (3, 256, 256), 表示 3 个通道, 分别输入反射率, 径向速度, 速度谱宽。输出是一个向量, 表示二分类任务的输出。由于龙卷样本尺寸较小, 使用 3×3 卷积核用于提取龙卷特征。模型包含 6 个卷积层, 每个卷积层使用 3×3 卷积核并带有 padding=2 以保持特征尺寸, 接 ReLU 激活函数和 2×2 的最大池化层逐步下采样。卷积层的通道数从 3 逐渐增加到 256, 通过多层卷积和池化操作提取龙卷数据的多层次特征。在卷积层之后, 模型引入了一个 Dropout 层 (丢弃率 0.5) 减少过拟合风险。提取特征被展为一维向量, 并通过两个全连接层进行分类: 第一个全连接层将 6400 维特征映射到 64 维, 第二个全连接层输出 2 维结果, 对应龙卷识别结果的二分类任务。模型大部分参数集中在全连接层, 尤其是第一个全连接层。通过卷积层提取局部特征, 池化层降低计算复杂度, Dropout 层增强泛化能力, 最终通过全连接层实现分类。

(二) 评估指标

本文使用机器学习中的二分类问题评估龙卷风识别模型的性能, 采用混淆矩阵作为主要评估工具。模型输出预测类别后根据二分类混淆矩阵可得到 ACC、Pre、F1-score、POD、FAR、CSI 和 AUC^[10]。

表1 ComplexCNN 在测试集上的评估指标得分

	ACC ↑	Pre ↑	AUC ↑	F1-score ↑	POD ↑	FAR ↓	CSI ↑
ComplexCNN	0.9778	0.9726	0.9983	0.9773	0.9828	0.0274	0.9219
TDA-RF ^[22]	0.9504	0.9267	0.9476	0.9486	0.9552	0.0733	0.9022

表1展示了ComplexCNN在测试集上的评估指标得分,其中ACC为0.9778,Pre为0.9726,AUC为0.9983,F1-score为0.9773,POD为0.9828,FAR为0.0274,CSI为0.9219。结果表明,模型在整体分类的准确率、正类检测能力及分类性能方面表现较好,由于现有TDA-RF算法。

三、实例测试结果

2020年6月12日,江苏省迎来了入梅后的首场强降雨[23],当日13时49分,高邮湖东岸约2.5至3公里处的高邮镇出现了8级大风,并产生了龙卷风,最大风速达17.5m/s。图3展示了ComplexCNN龙卷风识别算法的识别结果,提前时间为29分钟。随着龙母体风暴的强烈发展与中心聚合,在1.5°仰角的基本反射率图上可清晰看到入流缺口特征。高反射率区主要分布在入流缺口的西侧,在2.4°仰角的回波图中,龙卷区域回波强度超过了50dBZ,表明风暴内部存在剧烈的强对流运动。径向速度图上出现正负速度对(-20m/s和27m/s),这也是龙卷发生的显著特征。

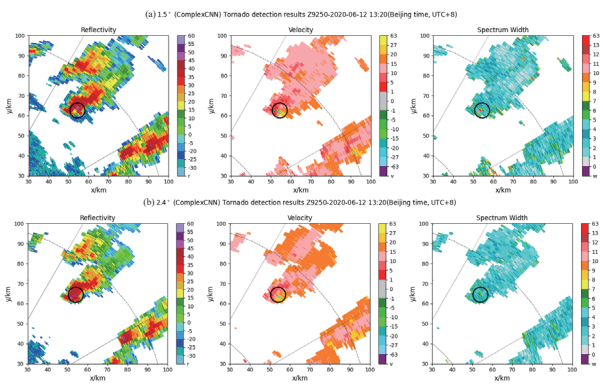


图3 ComplexCNN在Z9250雷达13:20(UTC+8)的识别结果

(a)为1.5° PPI识别结果,(b)为2.4° PPI识别结果

(以模型输出概率最大值的样本为中心,半径1.5km画圈作为龙卷识别的结果)

四、总结

本文介绍了构建龙卷风数据集的方法和ComplexCNN龙卷识别算法的结构设计思路、各部分搭建方法。并利用实际龙卷风案例数据进行了应用测试,得出以下结论:

(1)在江苏省两次龙卷风案例中,ComplexCNN进行了实例测试,平均预警时间为20分钟,最大预警时间达到29分钟,预警时间与龙卷风强度存在关联。

(2)实际测试结果显示,模型评估指标为POD:0.738、CSI:0.646、FAR:0.262,体现了优秀的识别性能。同时使用TorNet数据进行案例测试也能成功识别龙卷风,表明本文算法具备一定的泛用性。

参考文献

[1] 范雯杰, 俞小鼎. 中国龙卷的时空分布特征 [J]. 气象, 2015, 41(07): 793-805.
[2] 闻新宇, 朱清照, 贾喆. 为了龙卷风预警: 从约翰芬利到藤田哲也 [J]. 气象科技进展, 2013, 3(06): 72-6.
[3] BROOKS H E, LEE J W, CRAVEN J P. The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data [J]. Atmospheric Research, 2003, 67: 73-94.
[4] 郑永光, 陶祖钰, 俞小鼎. 强对流天气预报的一些基本问题 [J]. 气象, 2017, 43(06): 641-52.
[5] 王秀明, 俞小鼎. 热带一次致灾龙卷形成物理过程研究 [J]. 气象学报, 2019, 77(03): 387-404.
[6] WANG Y, YU T-Y. Novel tornado detection using an adaptive neuro-fuzzy system with S-band polarimetric weather radar [J]. Journal of Atmospheric and Oceanic Technology, 2015, 32(2): 195-208.
[7] XU Q, NAI K, WEI L, et al. A Vortex Relocation Method for Improving Initial Conditions and Subsequent Predictions of Tornadoes [J]. Weather and Forecasting, 2023, 38(3): 391-9.
[8] QING Z, ZENG Q, WANG H, et al. ADASYN-LOF algorithm for imbalanced tornado samples [J]. Atmosphere, 2022, 13(4): 544.
[9] 黄荣, 黄晴, 屈梅芳, et al. 1713号台风“天鸽”外围龙卷特征及成因分析 [J]. 气象研究与应用, 2018, 39(01): 28-32+90+154.
[10] ZENG Q, QING Z, ZHU M, et al. Application of random forest algorithm on tornado detection [J]. Remote Sensing, 2022, 14(19): 4909.
[11] 唐佳佳, 唐晓文, 徐芬, et al. “0612”高邮龙卷母体风暴演变特征分析及龙卷形成机理初探 [J]. 热带气象学报, 2021, 37(Z1): 824-35.