

文章编号: 2095-2163(2022)12-0187-05

中图分类号: TP391.4

文献标志码: A

# 百度 EasyDL 开发平台在气象装备保障中的应用

涂爱琴<sup>1,2,3,4</sup>, 陈庆亮<sup>3,4</sup>, 于帅<sup>3,4</sup>, 张玉洁<sup>3,4</sup>

(1 中国科学院合肥物质科学研究院 安徽光学精密机械研究所, 中国科学院大气光学重点实验室, 合肥 230031;

2 中国科学技术大学研究生院科学岛分院, 合肥 230026; 3 山东省气象防灾减灾重点实验室, 济南 250031;

4 山东省气象局大气探测技术保障中心, 济南 250031)

**摘要:** 随着深度学习等算法技术的突破,人工智能在气象领域中的应用也越来越广泛。文章分析了人工智能技术在观测环境监控、数据质量核查和设备故障监控等业务中的应用场景,并基于百度 EasyDL 平台创建了雨量筒是否加盖、雨量筒是否有异物及气溶胶观测设备故障监控等模型。结果表明人工智能技术有可能改变现有的装备保障模式,有效提升保障效率和效果。

**关键词:** EasyDL; 深度学习; 气象装备保障; 故障监控; 图像识别; 时序预测

## Application of Baidu EasyDL development platform in meteorological equipment support

TU Aiqin<sup>1,2,3,4</sup>, CHEN Qingliang<sup>3,4</sup>, YU Shuai<sup>3,4</sup>, ZHANG Yujie<sup>3,4</sup>

(1 Key Laboratory of Atmospheric Optics, Anhui Institute of Optics and Fine Mechanics, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China; 2 Science Island Branch of Graduate School, University of Science and Technology of China, Hefei 230026, China; 3 Key Laboratory for Meteorological Disaster Prevention and Mitigation of Shandong, Jinan 250031, China; 4 Ensuring Center of Atmospheric Sounding Technology, Shandong Meteorological Bureau, Jinan 250031, China)

**[Abstract]** With the breakthrough of algorithm technology such as deep learning, the application of artificial intelligence in the field of meteorology is also increasingly widespread. The application scenarios of artificial intelligence technology, such as observation environment monitoring, data quality verification and equipment fault monitoring, are analyzed in this paper. Also, based on Baidu EasyDL platform, models such as whether the rain gauge is covered, whether there are foreign matters in the rain gauge are constructed and the models of fault monitoring of aerosol observation equipment are created. The results show that AI technology could change the existing equipment support mode and effectively improve the ensuring efficiency and effect.

**[Key words]** EasyDL; deep learning; meteorological equipment support; fault monitoring; image recognition; time series prediction

## 0 引言

随着以深度学习为代表的机器学习算法取得突破,人工智能(AI)呈现了加速发展的态势<sup>[1]</sup>。由于机器学习在计算效率、准确性、可移植性、协同性、灵活性和易用性等方面具有较大的优势,近年来已在气象领域获得广泛应用。在地面观测上,朱磊等人<sup>[2]</sup>提出了一种基于小样本学习的地面结露结霜现象检测方法,田治仁等人<sup>[3]</sup>发明了基于图像的雾浓度等级判别方法,黄小玉等人<sup>[4]</sup>建立了冰雪天气现象的识别模型。在天气雷达观测上,主要开展了异常回波识别、雷达回波外推、水凝物相态识别等应用研究<sup>[5-14]</sup>。在数值天气预报上,从数据预处理、数据同化到预报和预测后的处理和校正,都已经进行了机器学习算法的应用研究<sup>[1,15-18]</sup>。此外,人工智

能技术在强对流天气的监测和预警<sup>[19-21]</sup>、卫星资料的应用<sup>[22-24]</sup>、气候预测<sup>[25]</sup>和农业气象<sup>[26]</sup>等方向也都取得了系列研究成果。

综上所述,目前人工智能技术在气象领域的应用主要集中在气象观测、数值天气预报、强对流天气识别预警及卫星资料应用四个方面,但是在气象装备保障方向的应用还尚未开展。论文举例分析了人工智能技术在气象装备保障中的应用场景,并依托百度 EasyDL 开发平台建立实验模型,对应用效果进行验证。

## 1 EasyDL 介绍

EasyDL 是百度推出的人工智能开发平台,内置百度超大规模预训练模型和自研 AutoDL 技术,只需少量数据就能训练出高精度模型,适用于 AI 零算法

基金项目: 山东省气象局面上科研项目(2019sdqxm06,2020sdqxm05,2020sdqxm06)。

作者简介: 涂爱琴(1983-),女,硕士,高级工程师,主要研究方向:气象观测、数据质控和装备保障等。

通讯作者: 涂爱琴 Email: betrue\_tu@sohu.com

收稿日期: 2022-05-18

基础或追求高效率开发的用户。根据应用场景及深度学习的技术方向, EasyDL 产品可分为 6 大类, 分别是: EasyDL 图像、EasyDL 文本、EasyDL 语音、EasyDL OCR、EasyDL 视频和 EasyDL 结构化数据。其中, EasyDL 图像定制基于图像进行多样化分析的 AI 模型, 实现图像内容理解分类、图中物体检测定位等, 适用于图片内容检索、安防监控、工业质检等场景; EasyDL 文本基于百度大脑文心领先的语义理解技术, 提供一整套 NLP 定制与应用能力, 广泛应用于文本分类、文本实体抽取、情感倾向分析及短文本相似度分析等场景; EasyDL 语音定制语音识别模型, 适用于数据采集录入、声音分类等场景; EasyDL OCR 定制化训练文字识别模型, 结构化输出关键字段内容, 适用于证件照电子化审批、财税报销电子化等场景; EasyDL 视频定制化分析视频片段内容、跟踪视频中特定的目标对象, 适用于视频内容审核、人流/车流统计、养殖场牲畜移动轨迹分析等场景; EasyDL 结构化数据旨在帮助用户通过机器学习技术从数据中发现潜在规律, 从而对未来的变化趋势进行预测, 包含表格数据预测和时序预测两个细分产品, 适用于交通流量预测、价格预测等场景。

## 2 应用场景分析

### 2.1 雨量筒加/揭盖检查

每年汛期结束后停止非称重降水观测, 维护人员要给雨量筒加盖保护设备; 汛期开始前则要揭开雨量筒盖, 利于后续的降水观测。目前区域站维护业务大多采取了外包服务, 因此对外包公司是否及时进行了加/揭盖维护需要开展业务检查。由于区域站场地分散, 一般采取抽查的方式, 不但不能实现普查, 且耗费的人力财力巨大。在业务场景安装监控摄像头, 采用定时抓拍或手动抓拍方式采集雨量筒图片, 建立雨量筒加/揭盖检查模型, 就可以实现区域站雨量筒是否加盖和揭盖的智能检查, 提高业务质量。雨量筒有无加盖检测图像如图 1 所示。

### 2.2 雨量筒内是否有异物检查

当雨量筒中出现树叶等异物时, 常常会影响观测结果的准确性, 如带来延时降水、降水偏小和无降水等情况, 给气象服务造成困扰。因此每当预报有天气变化过程时, 保障人员需要在过程前巡查和维护雨量筒; 气象服务人员则需要在过程结束后制作雨情报告等服务材料时, 甄别数据的真实性, 并将存疑的站点数据剔除掉。前者成本高、效率低, 后者难度大、易出错。在业务场景安装监控摄像头, 建立雨量筒有无异

物检查模型, 当预报有天气变化过程时, 保障人员抓拍现场图片并自动识别雨量筒有无异物, 随后仅对识别结果为有异物的站点进行维护, 能大大降低维护成本。当雨情出现后, 气象服务人员在制作服务材料前抓拍现场图片, 并自动识别雨量筒有无异物, 将模型识别为有异物的站点数据剔除掉, 能确保材料的准确性, 有效提高服务质量。雨量筒有无异物检测图像如图 2 所示。



图 1 雨量筒有无加盖图像

Fig. 1 Pictures of rain gauge with or without the cover



图 2 雨量筒有无异物图像

Fig. 2 Pictures of rain gauge with or without foreign matters

### 2.3 设备故障监控

当观测设备出现故障的时候, 常常会在观测数据上有所体现, 如缺测、跳变、数值偏低或偏高等。因此对观测数据进行监控能够实现对气象观测设备故障的监控。通过数据监控设备运行状态的常规方法是设置阈值来检验观测数据是否正常, 但是阈值设置不当会影响监控方法的准确性, 且阈值检测是单点检测, 因观测环境等因素引起的单个数值跳变容易引起设备故障的误判。

另一方面, 当观测数据出现缺测和跳变等异常

时,其观测产品、如时间变化曲线等也往往出现异常,如图 3、图 4 所示。图 3 是正常和典型故障下气溶胶观测数据时间变化曲线图。图 3(a)中,PM<sub>2.5</sub> 观测值长时间大于 PM<sub>10</sub> 观测值,是典型的倒挂故障;图 3(b)中,PM<sub>2.5</sub> 值一直没有变化,是恒值故障;图 3(c)是正常的曲线图。图 4 是正常和典型故障下土壤水分观测数据时间变化曲线图。图 4(a)

是正常曲线;图 4(b)表示 20 公分传感器附近有个菜根引起数据跳变;图 4(c)表示 10 公分传感器故障导致数据偏小且跳变。鉴于正常曲线和典型故障曲线的明显区别,将观测数据按模板制作成数据产品,随后应用基于深度学习的图像识别技术,建立设备故障智能监控模型,可对设备的运行状态进行实时监控和预警。



图 3 正常和典型故障下气溶胶观测数据曲线图

Fig. 3 Curve of aerosol observation data under normal situation and typical faults

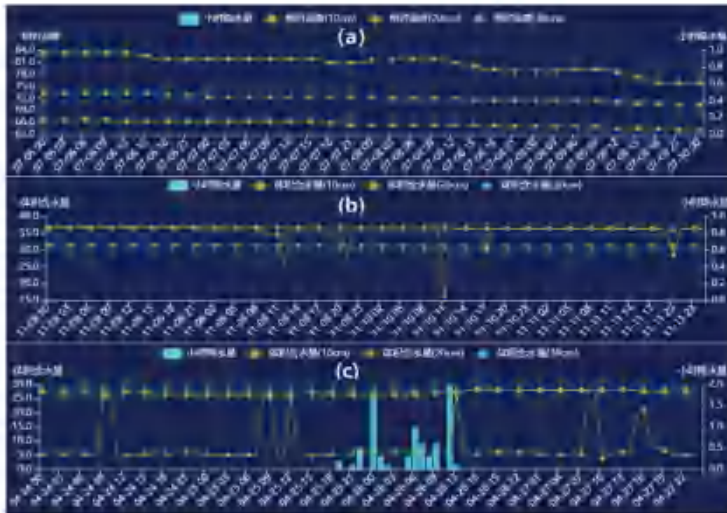


图 4 正常和典型故障下自动土壤水分观测数据曲线图

Fig. 4 Curve of automatic soil moisture observation data under normal situation and typical faults

### 3 建模实验

#### 3.1 建模流程

使用 EasyDL 平台建立 AI 模型的步骤如图 5 所示。在建模之前,需要有效分析和拆解业务需求,然后选择合适的产品类型。如,针对雨量筒有无异物检查的业务需求,可以通过 EasyDL 产品中的图像模

型进行判断;针对预测备件消耗数量的业务需求,则可以通过 EasyDL 产品中的结构化数据来进行建模预测。有的业务需求含有多个业务功能,且每个功能适用的产品不同,这时就需要对业务需求进行拆解,并针对每个业务功能分别建模,综合应用多个模型来满足业务需要。



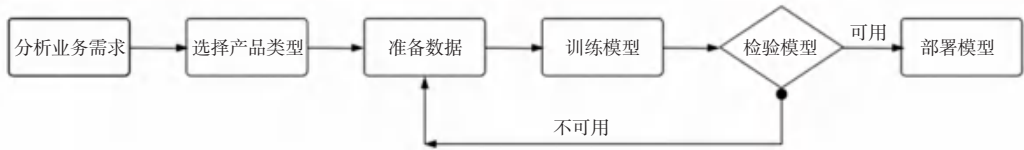


图5 AI模型建立流程

Fig. 5 Flow chart of AI model establishment

研究可知,基于 EasyDL 平台开发,确定了产品类型就相当于确定了算法,因此接下来需要进行的就数据收集工作。采集数据后,可以通过 EasyDL 在线标注工具或线下利用其他标注工具对已有的数据进行标注。如上述雨量筒有无异物的模型,需要将采集图片按照有异物和无异物两类进行分类标注。

数据准备好后,在 EasyDL 平台根据向导填写模型名称、应用场景和功能描述等信息创建模型。模型建好后加入数据集进行训练,根据数据集大小及训练环境等因素,模型训练时间从几分钟到几小时不等。模型训练完成后,百度平台自动生成了评估报告,查看评估报告初步了解模型的效果。为了进一步验证模型效果,可以通过平台提供的校验功能,对模型效果进行测试。测试结果可用的模型就可以申请发布,并在业务系统的运行中调入使用;测试结果不理想的模型则需要增加数据集继续训练,直到测试结果可用后发布。

### 3.2 数据来源

论文基于百度 EasyDL 平台的图像分类产品,通过在区域站安装摄像头,采集了区域站雨量筒图片 500 张,训练了雨量筒是否加盖和雨量筒是否有异物两个模型。通过中国气象局气象探测中心开发的数据质量控制系统-天衡,采集了气溶胶质量浓度时间变化曲线图片 150 张,训练了气溶胶观测设备故障识别模型。

### 3.3 结果分析

3 个模型的检测效果见表 1。表 1 中,准确率是指正确分类的样本数与总样本数之比,越接近 1、模型效果越好。就某类别而言,精确率是正确预测为该类别的样本数与预测为该类别的总样本数之比,指标越高、误识别率越低;召回率是正确预测为该类别的样本数与该类别的样本数之比,指标越高、漏识别率越低; $F_1$ -score 是精确率和召回率的调和平均数,其值越高、模型效果越好。

由表 1 可以看出,雨量筒是否加盖和是否有异物的识别模型效果较好,准确率达到或接近 100%,而气溶胶观测设备故障识别模型由于设置的故障分

类较多,训练的图片样本数量又偏少等原因,准确率偏低,模型需要进一步优化。

表 1 模型效果评估表

模型名称	准确率 $F_1$ -score	精确率	召回率
雨量筒是否加盖识别模型	96.7	97.7	98.0
雨量筒是否有异物识别模型	100.0	100.0	100.0
气溶胶观测设备故障识别模型	73.3	48.4	53.8

## 4 结束语

(1) 百度 EasyDL 开发平台提供了包括数据清洗、数据标注、训练任务配置和调参、模型效果评估和模型部署等一站式服务,具有广适配和高精度等优势,平台设计简约,容易理解,适用于零基础气象业务人员开展人工智能技术的应用研究。

(2) 随着深度学习算法的突破,人工智能技术在气象领域的应用也在飞快发展。在观测保障上,除了智能观测方面的应用,人工智能技术未来在观测业务质量检查和智能保障方面也将会得到大规模应用。如应用图像分类和物体检测技术开展观测环境的监控和预警,应用异常观测产品识别方法开展数据质量控制和设备故障监控工作,应用时序预测模型开展备件消耗数量预测用于备件储备指导等。

(3) 基于百度 EasyDL 平台训练的雨量筒是否加盖、雨量筒是否有异物和气溶胶观测设备故障识别模型验证了深度学习技术应用于气象装备保障领域的可行性。此外,前 2 个模型的精度已经达到业务使用的要求,第 3 个模型仍需要积累数据做进一步的优化。

## 参考文献

- [1] 杜智涛,姜明波,杜晓勇,等. 机器学习在气象领域的应用现状与展望[J]. 气象科技, 2021, 49(06): 930-941.
- [2] 朱磊,张小虎,吴谨. 基于小样本学习的地面结露结霜现象检测方法[J]. 现代电子技术, 2019, 42(18): 130-135.
- [3] 田治仁,张贵峰,李锐海,等. 雾浓度等级判别方法及装置: 中国, CN110705619A[P]. 2020-01-17.
- [4] 黄小玉,张垚,叶成志,等. 地面积雪天气现象人工智能观测识别研究[J]. 气象, 2019, 45(09): 1189-1198.
- [5] 刘昉,李奇临,蒋涌,等. 基于 CNN 的天气雷达异常回波图像识

- 别算法研究[J]. 电子设计工程, 2021, 29(06):74-78,87.
- [6] CHEN Xiaowu, ZHOU Bin, GUO Yu, et al. Structure guided texture inpainting through multi-scale patches and global optimization for image completion[J]. Science China: Information Sciences, 2014, 57:012102:1-012102:16.
- [7] YANG Ling, WANG Yun, WANG Zhongke, et al. A new method based on stacked auto-encoders to identify abnormal weather radar echo images [J]. Eurasip Journal on Wireless Communications and Networking, 2020, 2020(1):177.
- [8] HAN Lei, SUN Juazhen, ZHANG Wei, et al. A machine learning nowcasting method based on real-time reanalysis data[J]. Journal of Geophysical Research: Atmospheres, 2017, 122(7):4038-4051.
- [9] 甘建红,尹麒名,任宇,等.一种基于改进的 TrajGRU 网络的气象雷达回波外推方法;中国, CN112180375A[P]. 2021-01-05.
- [10] 李英睿,李琼,魏加华,等.基于 ConvLSTM 的气象雷达回波外推[J]. 青海大学学报,2021, 39(01):93-100.
- [11] 方巍,庞林,张飞鸿,等.对抗型长短期记忆网络的雷达回波外推算法[J]. 中国图象图形学报, 2021, 26(05):1067-1080.
- [12] 凌青,钟淑鑫,蒙伟光,等.基于注意力机制的时空神经网络雷达回波外推预报方法;中国, CN112446419A[P]. 2021-03-05.
- [13] 冉元波,孙敏,高梦清,等.双偏振天气雷达水凝物识别研究[J]. 成都信息工程学院学报, 2017,32(06):590-596.
- [14] 徐舒扬,吴翀,刘黎平.双偏振雷达水凝物相态识别算法的参数改进[J]. 应用气象学报, 2020, 31(03):350-360.
- [15] CHEVALLIER F, MAHFOUF J F. Evaluation of the jacobians of infrared radiation models for variational data assimilation [J]. Journal of Applied Meteorology, 2010, 40(8):1445-1461.
- [16] 肖林鸿,陈焕盛,陈婷婷,等.基于机器学习的短临预报方法及其在空气质量保障中的应用[J]. 中国环境监测,2021,37(03):66-74.
- [17] 疏杏胜,王子茹,李福威,等.基于机器学习模型的短期降雨多模式集成预报[J]. 南水北调与水利科技(中英文), 2020, 18(01):42-50.
- [18] 孙全德,焦瑞莉,夏江江,等.基于机器学习的数值天气预报风速订正研究[J]. 气象, 2019, 45(03):426-436.
- [19] 周康辉,郑永光,韩雷,等.机器学习在强对流监测预报中的应用进展[J]. 气象, 2021,47(03):274-289.
- [20] 修媛媛,韩雷,冯海磊.基于机器学习方法的强对流天气识别研究[J]. 电子设计工程, 2016, 24(09):4-7.
- [21] 刘新伟,黄武斌,蒋盈沙,等.基于 LightGBM 算法的强对流天气分类识别研究[J]. 高原气象, 2021, 40(04):909-918.
- [22] 邹国良,侯倩,郑宗生,等.面向卫星云图及深度学习的台风等级分类[J]. 遥感信息, 2019, 34(03):1-6.
- [23] 崔林丽,陈昭,于兴兴,等.风云四号卫星东南沿海热带气旋强度深度学习估算[J]. 遥感学报, 2020, 24(07):842-851.
- [24] 阙希,张永宏,曹庭,等.利用多光谱卫星遥感和深度学习方法进行青藏高原积雪辨识[J]. 测绘学报, 2016, 45(10):1210-1221.
- [25] 贺圣平,王会军,李华,等.机器学习的原理及其在气候预测中的潜在应用[J]. 大气科学学报, 2021, 44(01):26-38.
- [26] 郝玲,张佩,史逸民,等.基于机器学习的江苏省冬小麦气象产量客观区划及歉年预测[J]. 江苏农业科学, 2021, 49(12):162-168.

## (上接第 186 页)

- [10] LI Boyi, PENG Xiulian, WANG Zhangyang, et al. An all-in-one network for dehazing and beyond[J]. arXiv preprint arXiv:1707.06543, 2017.
- [11] REN Wenqi, MA Lin, ZHANG Jiawei, et al. Gated fusion network for single image dehazing[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 3253-3261.
- [12] 睢青青,李朝锋,桑庆兵.改进多尺度卷积神经网络的单幅图像去雾方法[J]. 计算机工程与应用, 2019, 55(10):179-185.
- [13] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[J]. Advances in Neural Information Processing Systems, 2014, 27: 2672-2680.
- [14] ZHU Junyan, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 2223-2232.
- [15] ENGIN D, GENC A, KEMAL EKENEL H. Cycle-dehaze: Enhanced cycleGAN for single image dehazing[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City, ; IEEE, 2018: 825-833.
- [16] 肖进胜,申梦瑶,雷俊峰,等.基于生成对抗网络的雾霾场景图像转换算法[J]. 计算机学报, 2020,43(01):165-176.
- [17] POLYKOVSKIY D, NOVIKOV A. Bayesian methods for machine learning[J]. Coursera and National Research University Higher Economics, 2017.
- [18] LEE H Y, TSENG H Y, HUANG J B, et al. Diverse image-to-image translation via disentangled representations [C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich Germany: Springer, 2018: 35-51.
- [19] ODENA A, DUMOULIN V, OLAH C. Deconvolution and checkerboard artifacts[J]. Distill, 2016, 1(10):e3.
- [20] JOHNSON J, ALAHI A, LI Feifei. Perceptual losses for real-time style transfer and super-resolution [C]//European Conference on Computer Vision. Glasgow, UK: Springer, 2016: 694-711.
- [21] LI Boyi, REN Wenqi, FU Dengpan, et al. Benchmarking single-image dehazing and beyond[J]. IEEE Transactions on Image Processing, 2018, 28(1):492-505.
- [22] SILBERMAN N, HOIEM D, KOHLI P, et al. Indoor segmentation and support inference from rgb-d images [C]//European Conference on Computer Vision. Florence, Italy: Springer, 2012: 746-760.
- [23] SCHARSTEIN D, HIRSCHMULLER H, KITAJIMA Y, et al. High-resolution stereo datasets with subpixel-accurate ground truth [M]//JIANG X, HORNEGGER J, KOCH R. Pattern Recognition. GCPR 2014. Lecture Notes in Computer Science(). Cham: Springer, 2014,8573: 31-42.
- [24] QU Yanyun, CHEN Yizi, HUANG Jingying, et al. Enhanced pix2pix dehazing network [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 8152-8160.