工程与管理科学
Engineering and Management Science
第 8 卷◆第 10 期◆版本 1.0◆2026 年
文章类型：论文|刊号（ISSN）：2705-0637(P) / 2705-0645(O)

基于 ConvNeXt-Tiny 与置信度筛选的台站霜图像识别

贺晓冬 [1,2]
1. 易天气（北京）科技有限公司 北京 100054；2. 国家超级计算天津中心 天津 300457

DOI：10.32629/ems.v8i10.22607

［摘　要］固定台站相机可连续记录地表霜，但霜纹理细薄，易与干草、湿土和积雪混淆。本文以 ImageNet 预训练 ConvNeXt-Tiny 输出霜概率，结合类别加权、翻转增强和置信度筛选支持人工复核。数据含霜 529 张、非霜 2341 张、其他场景 130 张，二分类覆盖 119 个站点。站点隔离五折评价中，AP、AUROC 和 F1 分别为 0.683、0.869 和 0.587，召回率和精确率为 0.456 和 0.825；AP 较 ResNet18 提高 0.038。阈值降至 0.1 时，F1 和召回率分别为 0.633 和 0.586，精确率降至 0.687。其他场景中 20.8% 的图像概率不低于 0.5。该方法适合候选筛查，尚不支持无人值守判定。

［关键词］霜识别；气象台站；图像分类；ConvNeXt；深度学习；不确定样本

1 引言

霜由地表温度、湿度和辐射共同作用形成，影响农业、交通和自动气象站判读。固定相机可连续记录地表霜，但人工逐帧检查成本高；积雪图像监测已有研究基础[1-3]，天气现象记录也有规范要求[4]。

霜与干草反光、湿土高亮和弱降雪相似，卷积网络虽能提取复杂背景特征[5-7]，固定台站模型仍可能记忆相机背景和站点差异。

本文比较 ConvNeXt-Tiny 与 ResNet18，评估增强、阈值、置信度和时间支持，并用其他场景检验误判。

2 数据与方法

2.1 数据来源与标签

图像含霜、非霜和其他场景三类，二分类为 529 张霜和 2341 张非霜，覆盖 119 个站点；其他场景 130 张，仅作压力测试。组成与示例见图 1、图 2。

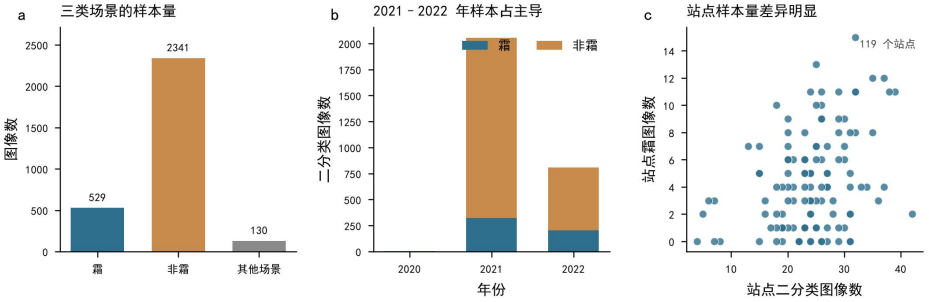

图 1 实验数据组成。a 为霜、非霜和其他场景的样本量；b 为二分类图像的年份分布；c 为各站点总样本量与霜样本量的关系。二分类数据包含 2870 张图像、119 个站点。

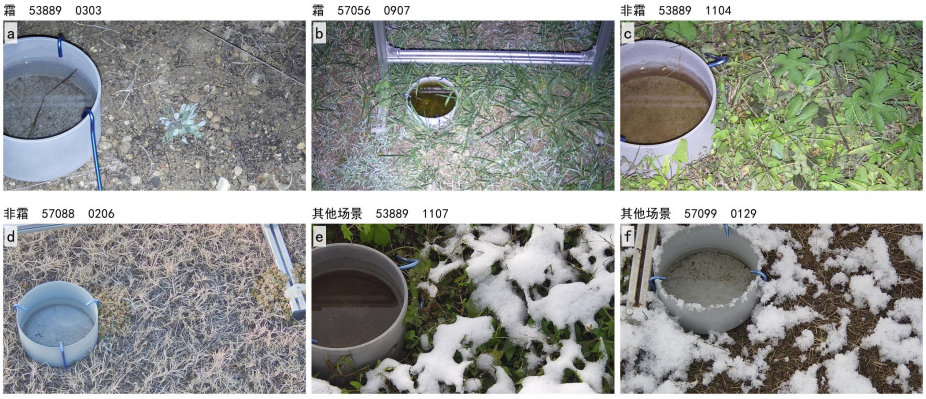

图 2 典型台站场景。a、b 为霜图像；c、d 为非霜图像；e、f 为其他场景图像。其他场景中包含积雪，未并入非霜类。

Copyright © This work is licensed under a Commons Attibution-Non Commercial 4.0 International License.

“霜”未由温度或天气现象记录独立确认；其他场景不参与准确率和 F1 计算。

2.2 霜分类模型

主模型为 ImageNet 预训练 ConvNeXt-Tiny[6,7]，ResNet18[5]为对照。采用水平翻转、旋转、亮度和颜色扰动、模糊及随机擦除[8]，并以频率加权采样[9]和二元交叉熵损失缓解类别不平衡。

训练使用 AdamW，主干学习率 2×10^{-4}，分类头为其 8 倍，批大小 32，余弦退火 12 轮；随后在 2870 张二分类图像上训练 14 轮。

2.3 识别算法与不确定样本处理

采用水平翻转测试增强，默认阈值 0.5 并报告 0.1～0.9 的敏感性。同站点 3 小时支持按高斯权重融合相邻预测，不使用测试标签。以最大类后验置信度 C 描述不确定样本，统计自动判定覆盖率和准确率。

2.4 评价设计

评价采用站点隔离五折交叉验证，各测试折含 105～107 张霜图像和 560～585 张总图像。报告 AP、AUROC、F1、召回率、精确率、特异度和准确率，并以站点 bootstrap 估计 95%置信区间。

3 结果与分析

3.1 数据分布

霜和非霜分别占二分类数据的 18.4%和 81.6%；102 个站点同时含两类，部分站点样本不足 10 张（表 1）。2021 年和 2022 年样本分别占 71.6%和 28.2%。

表 1 台站霜图像数据组成

类别	图像数	站点数	占二分类数据比例
霜	529	102	18.4%
非霜	2341	119	81.6%
其他场景	130	69	不参与
合计有效图像	3000	119	不适用

3.2 模型分类性能

ConvNeXt-Tiny 增强后的 AP、AUROC、F1、召回率和精确率分别为 0.683、0.869、0.587、0.456 和 0.825；ResNet18 分别为 0.645、0.857、0.509、0.386 和 0.747（表 2）。前者 AP 和 F1 分别提高 0.038 和 0.078，AP 的 95%置信区间为 0.006～0.073，但 AUROC 区间包含 0，优势有限。

表 2 两种模型在站点隔离五折评价中的折外结果

模型	参数量	AP	AUROC	F1	召回率	精确率	特异度
ResNet18＋翻转增强	11.2M	0.645	0.857	0.509	0.386	0.747	0.971
ConvNeXt-Tiny 原始模型	27.8M	0.683	0.867	0.592	0.457	0.837	0.980
ConvNeXt-Tiny＋翻转增强	27.8M	0.683	0.869	0.587	0.456	0.825	0.978

0.5 阈值下召回率仅 0.456，超过一半霜图像未检出；整体准确率 0.882 主要受非霜样本数量优势影响。AP、AUROC 和 F1 的 95%置信区间分别为 0.624～0.737、0.845～0.892 和 0.537～0.642。结果见图 3a。

Copyright © This work is licensed under a Commons Attibution-Non Commercial 4.0 International License.

文章类型：论文|刊号（ISSN）：2705-0637(P) / 2705-0645(O)

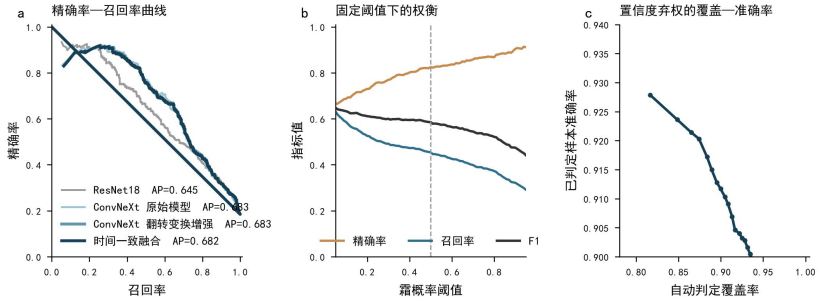

图 3 分类性能、阈值权衡与置信度弃权。a 为各方案的精确率-召回率曲线；b 为阈值对精确率、召回率和 F1 的影响；c 为不同置信度下的自动判定覆盖率与准确率。所有结果来自折外预测。

3.3 阈值与人工复核

阈值降至 0.1 时,F1 和召回率分别升至 0.633 和 0.586，精确率降至 0.687;阈值 0.2 时分别为 0.611、0.526 和 0.730（表 3）。降低阈值可增加候选，但会增加复核负担。

表 3 时间融合概率阈值敏感性

阈值	精确率	召回率	F1
0.1	0.687	0.586	0.633
0.2	0.730	0.526	0.611
0.3	0.769	0.491	0.600
0.5	0.825	0.454	0.585
0.6	0.836	0.425	0.564

接受 C≥0.8 时，自动判定覆盖率和准确率分别为 94.6%和 89.7%；C≥0.9 时为 90.8%和 90.9%。置信度筛选不能替代概率校准。

3.4 时间支持与其他场景压力测试

时间支持的 AP、AUROC 和 F1 分别为 0.682、0.869 和 0.585，未改善主要指标，故不纳入流程。

130 张其他场景中 20.8%的图像概率不低于 0.5，积雪可进入高置信度霜区（图 4）；业务含积雪时需增加积雪类或门控。

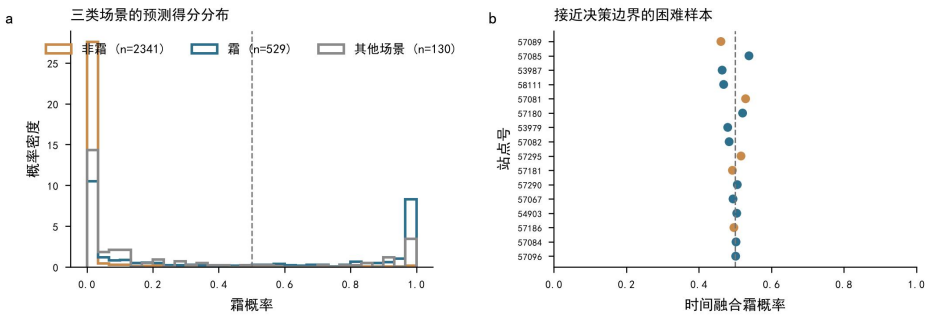

图 4 其他场景压力测试。a 为三类场景的预测概率分布；b 为接近决策边界的困难样本。其他场景未参与训练，也未计算分类准确率。

4 讨论

4.1 ConvNeXt-Tiny 提供了有限但可测量的改进

ConvNeXt-Tiny 的 AP 和 F1 较高，但 AUROC 区间包含 0，不能称为全面优于对照；参数量增至 27.8M，部署需权衡算力。

薄霜纹理、昼夜照明和复杂地表仍使识别困难；后续可

Copyright © This work is licensed under a Commons Attibution-Non Commercial 4.0 International License.

结合地表区域分割或局部块分类。

4.2 阈值、弃权与业务使用边界

0.5 阈值漏检过多，概率更适合候选排序；业务可用双阈值并人工复核中间区间。

按站点分组避免了直接背景泄漏，但同类设备上的结果不能外推至新设备、新视角和新地表类型。

4.3 其他场景与多相态识别的需求

其他场景中约五分之一图像被判为霜，说明二分类不足以覆盖冬季地表相态；应构建霜、积雪、湿土和干地表多类数据，或增加积雪门控。

4.4 局限

数据未提供标注规范和双盲复核结果，且不是连续时间序列；五折评价仍属同一相机体系，概率校准尚未开展，不能证明新站点泛化能力。

5 结论

本文建立固定台站霜图像识别流程。ConvNeXt-Tiny 取得 0.683 的 AP、0.869 的 AUROC 和 0.587 的 F1，AP 和 F1 优于 ResNet18；该方法适合候选筛查，不适合无人值守判定。后续需补充双人复核、概率校准、连续时序和新场景验证。

参考文献

[1] LIM S H, KIM G H, LEE S G, et al. Internet-based visual snow-cover monitoring and measuring system[C]//2006 SICE-ICASE International Joint Conference. IEEE, 2006: 3887-3890.

[2] SHIN I, RHIM S, LEE S G. The efficiency improvement of image-based snow-cover measurement system using histogram projection and thinning method[C]//2007 International Conference on Control, Automation and Systems. IEEE, 2007: 379-383.

[3] RYU H R, YOO K S, LIM Z S. Snow depth measurement using the structured pattern generator[C]//35th Annual Conference of IEEE Industrial Electronics. IEEE, 2009: 2715-2720.

[4] World Meteorological Organization. Guide to instruments and methods of observation: Volume I, measurement of meteorological variables, WMO-No. 8[R]. Geneva: WMO, 2021.

[5] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 770-778.

[6] LIU Z, MAO H, WU C Y, et al. A ConvNet for the 2020s[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2022: 11976-11986.

[7] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009: 248-255.

[8] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data, 2019, 6: 60.

[9] BUDA M, MAKI A, MAZUROWSKI M A. A systematic study of the class imbalance problem in convolutional neural networks[J]. Neural Networks, 2018, 106: 249-259.

基金项目：西城区科技专项项目（XCKJZX-2025-07）；北京市高层次创新创业人才支持计划科技新星计划（202604841147）。

通信作者：贺晓冬（E-mail：hexiaodong@cloudlizard.cn）

Copyright © This work is licensed under a Commons Attibution-Non Commercial 4.0 International License.