

基于深度学习与统计模型的垃圾分类任务研究

张静颜, 袁文涛, 邱敏凤, 芮荣祥*

中央民族大学理学院, 北京

收稿日期: 2026年7月11日; 录用日期: 2026年8月3日; 发布日期: 2026年8月12日

摘要

随着城市生活垃圾产量持续增长, 垃圾分类已成为推进资源循环利用与城市精细化治理的重要环节。针对传统人工分拣效率低、成本高以及智能分类模型在复杂场景下泛化能力不足的问题, 本文基于华为云生活垃圾分类数据集, 构建了一个融合统计模型与深度学习的多模型对比实验框架, 对生活垃圾多分类任务展开研究。本文首先采用HOG-LBP-HSV多手工特征融合提取轮廓、纹理、色彩三个维度的图像特征, 构建支持向量机(SVM)、随机森林(RF)和XGBoost三种传统统计基线模型。随后基于迁移学习策略, 选取ResNet18、MobileNetV3-Large和Swin-Tiny三种深度学习模型进行微调训练。实验中对原始训练集按9:1划分为训练集与验证集, 并将原始测试集作为独立测试集, 统一采用准确率、宏精确率、宏召回率、宏F1值、ROC-AUC以及推理FPS、参数量、FLOPs算力指标等指标对各模型进行综合评估, 并引入ANOVA方差分析验证性能差异显著性, 结合归一化混淆矩阵与高频错分样本完成可解释分析。实验结果表明, 深度学习模型在垃圾分类任务上显著优于传统统计模型。其中, Swin-Tiny在测试集上取得了最优的分类性能, ResNet18和MobileNetV3-Large也表现出较强的特征学习能力与泛化性能。相比之下, SVM、随机森林与XGBoost三个传统模型受限于手工特征表达能力, 在复杂类别识别上性能相对较弱。进一步的混淆矩阵和错误案例分析表明, 模型的主要误差集中于外观相似、材质接近的垃圾类别之间。综上, 本文验证了迁移学习深度学习模型在生活垃圾多分类任务中的有效性, 并为智能垃圾分类系统的模型选型与落地部署提供了实验依据。

关键词

HOG-LBP-HSV, 支持向量机, 随机森林, XGBoost, ResNet18, MobileNetV3-Large, Swin-Tiny

A Study of Deep Learning and Statistical Models for Multiclass Classification of Household Waste

Jingyan Zhang, Wentao Yuan, Minfeng Qiu, Rongxiang Rui*

School of Science, Minzu University of China, Beijing

*通讯作者。

文章引用: 张静颜, 袁文涛, 邱敏凤, 芮荣祥. 基于深度学习与统计模型的垃圾分类任务研究[J]. 统计学与应用, 2026, 15(8): 168-180. DOI: 10.12677/sa.2026.158188

Received: July 11, 2026; accepted: August 3, 2026; published: August 12, 2026

Abstract

With the continuous growth of municipal solid waste, waste classification has become an important part of resource recycling and fine-grained urban governance. To address the low efficiency and high cost of manual sorting, as well as the limited generalization ability of intelligent classification models in complex scenarios, this study constructs a multi-model comparative framework based on a public waste image dataset and investigates the waste multi-classification problem using both statistical machine learning and deep learning methods. This paper first extracts multi-dimensional visual features—including contour, texture, and color information—by fusing HOG-LBP-HSV descriptors, and subsequently trains three classical models: Support Vector Machine (SVM), Random Forest (RF), and XGBoost. Then, three pretrained deep learning models, ResNet18, MobileNetV3-Large, and Swin-Tiny, are fine-tuned under a transfer learning strategy. The original training set is split into training and validation subsets with a ratio of 9:1, while the original test set is used as an independent test set. Comprehensive performance assessment is conducted using multiple metrics, including Accuracy, Macro-Precision, Macro-Recall, Macro-F1, ROC-AUC, as well as computational efficiency indicators such as inference FPS, parameter count, and FLOPs. ANOVA was introduced to verify the significance of performance differences, and an interpretable analysis was completed by combining the normalized confusion matrix and high-frequency misclassified samples. Experimental results show that deep learning models significantly outperform traditional statistical baselines in waste classification. Among all models, Swin-Tiny achieves the best overall performance on the test set, while ResNet18 and MobileNetV3-Large also demonstrate strong feature learning and generalization capabilities. In contrast, SVM, RF, and XGBoost are constrained by handcrafted feature representations and show relatively weaker performance in recognizing visually similar categories. Further confusion matrix and misclassification case analyses indicate that most errors occur between categories with similar appearance and material characteristics. Overall, this study validates the effectiveness of transfer learning-based deep models in municipal waste classification and provides useful evidence for model selection and practical deployment in intelligent waste sorting systems.

Keywords

HOG-LBP-HSV, SVM, Random Forest, XGBoost, ResNet18, MobileNetV3-Large, Swin-Tiny

Copyright © 2026 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. 引言

近年来,我国城市生活垃圾年产量持续攀升,垃圾围城、环境污染、资源浪费等问题日益突出。垃圾分类作为垃圾减量化、无害化、资源化处理的前置关键环节,是推进绿色低碳发展、完善城市治理体系的重要举措。传统垃圾分类高度依赖人工分拣,存在分拣效率低、人工成本高、易错率高、工作环境恶劣等诸多弊端,无法满足现代化垃圾处理的规模化、自动化需求。

随着计算机视觉与人工智能技术的快速迭代,基于图像识别的智能垃圾分类技术逐步落地应用,凭借高效、精准、全天候工作的优势,成为替代人工分拣的核心方案。垃圾分类图像识别研究经历了从传

统机器学习到深度学习的逐步演进。早期研究多基于传统机器学习方法,通过提取图像颜色、纹理、形状等手工特征,结合 SVM、决策树等模型完成垃圾分类,实现了基础的自动分类功能,但在复杂场景下泛化能力较差。Yang 和 Thung 通过构建经典的 TrashNet 垃圾数据集,采用 SIFT 特征结合 SVM 模型开展垃圾分类实验,验证了手工特征加统计分类器在垃圾识别中的可行性[1]。Gaurav 等在垃圾图像分类任务中对比了 SVM、随机森林、梯度提升树等多类统计分类器,证实集成树模型的鲁棒性显著优于单一 SVM,同时该研究指出,仅依靠手工特征存在性能天花板,基于深度预训练特征可大幅提升识别精度[2]。另有 Fotovvatikhah 等撰写的综述系统归纳了 SURF、HOG、LBP 搭配 KNN、SVM 等传统图像分类实验方案,指出高维手工特征存在推理耗时高、不均衡数据集适配差等落地短板[3]。针对传统方法的局限性,Rabano 等指出,传统的浅层手工特征仅能描述图像轮廓与边缘信息,无法捕捉垃圾材质、形态的高层语义特征,多分类精度存在明显天花板,而卷积神经网络(CNN)的引入有效突破了这一瓶颈[4]。Mittal 等率先将浅层 CNN 应用于移动端垃圾分类系统,证实了端到端特征提取的优越性,但浅层网络特征提取能力有限[5]。Aral 等基于迁移学习对比了多款经典预训练 CNN 模型,证实 ResNet 等残差网络可有效缓解梯度消失问题,在垃圾图像分类任务中稳定性更强[6]。为适配嵌入式智能设备落地需求,轻量化 CNN 模型成为研究热点。国内学者肖克江针对国产智能垃圾桶场景验证,MobileNetV3 在保证较高分类精度的同时,计算开销极低,适合算力有限的边缘端垃圾分类设备,兼顾了实用性与经济性[7]。近年来,视觉 Transformer 模型突破了 CNN 局部特征提取的局限,凭借窗口自注意力机制捕捉图像全局特征,在细粒度图像分类任务中性能大幅提升。Huang 等的研究证实,Vision Transformer 架构凭借自注意力机制能够捕捉全局特征,突破了传统 CNN 感受野受限的瓶颈,在 TrashNet 数据集上的垃圾分类准确率显著优于已有 CNN 方法,验证了该架构在自动垃圾分类任务中的优越性[8]。Deng 等人的研究证实,Vision Transformer 依靠全局空间建模能力,对材质相近、形态复杂的垃圾样本识别效果优于 ResNet 等传统 CNN 模型,可有效降低塑料、玻璃、织物等易混淆垃圾的错分率[9]。Swin Transformer 作为轻量化的层级 Transformer 模型,兼顾了全局特征建模能力与计算效率,在生活垃圾多分类任务中具备巨大的应用潜力[10]。

综合现有研究,多数工作仅单一采用 HOG 搭配 SVM 和 RF,缺少 LBP 纹理、HSV 色彩多维度特征融合,且未引入 XGBoost 高性能集成树,传统模型性能上限被严重低估,无法形成有说服力的对照基准。虽然现有深度学习研究已充分证实不同网络架构的性能差异,但多数研究仅单独验证单一模型性能,缺乏轻量化 CNN、经典残差 CNN、Transformer 模型的横向对比,同时缺少与传统统计基线模型的量化对照,无法清晰体现深度学习算法的性能优势与迭代升级规律。本文通过搭建统计基线模型与深度学习模型的完整对比实验体系,首先融合 HOG 轮廓、LBP 纹理、HSV 颜色矩多维手工特征,结合 SVM、随机森林、XGBoost 构建传统统计机器学习基线模型;其次选取 ResNet18、MobileNetV3、Swin-Tiny 三个深度学习模型,基于迁移学习完成模型微调;通过统一分层数据集划分、统一超参数、统一评价指标,在独立测试集上系统对比所有模型的分类精度、均衡性与泛化能力。通过训练曲线、指标对比、ROC-AUC、训练时间、性能-复杂度平衡图、混淆矩阵和错误案例图等多种可视化方式,对模型的性能、判别能力、收敛过程和误分类模式进行综合分析,总结不同模型在生活垃圾多分类任务中的适用场景和性能特点,为生活垃圾智能分类技术的选取与落地提供完整的实验支撑与理论依据。

2. 实验方案设计

2.1. 数据集的选取与划分

本文实验采用华为云垃圾分类数据集,该数据集符合国内生活垃圾分类国家标准,完整覆盖可回收物、厨余垃圾、有害垃圾、其他垃圾四大类别,包含一次性快餐盒、剩饭剩菜、金属易拉罐、废旧电池等日常生活常见废弃物样本。数据集整体分为两大原始文件夹,原始训练文件夹为竞赛官方基础图像样本

库, 包含各类垃圾实景拍摄图片, 样本覆盖不同光照、拍摄角度、背景环境, 原始测试文件夹为后期扩充新增实景图像, 图片拍摄场景、垃圾形态与原始训练集存在差异化分布, 可用于客观检验模型真实泛化能力, 完全满足生活垃圾多分类对比实验需求。

原始数据集分为训练集文件夹与测试集文件夹, 为保证实验严谨性与可复现性, 同时保障全部对比模型实验条件统一、结果可复现, 本文采用分层随机抽样策略对原始训练集进行二次拆分, 并固定随机种子保证划分结果唯一。将原始训练集按照 9:1 分层随机抽样划分为新的训练集与验证集。分层抽样保证各类别样本比例不变, 避免类别分布偏移, 原始测试集数据完全不参与训练与调参, 作为独立测试集, 用于最终泛化性能评估。这最大程度还原模型落地真实识别场景, 防止数据泄露。划分完成后的最终数据集结构为训练集 13,214 张图片, 用于输入模型完成迭代训练, 更新网络和统计分类器的参数, 验证集 1469 张图片, 用于训练过程中实时评估模型性能, 进行超参数调优、最优权重筛选, 测试集 4165 张图片, 用于最终指标评估。本文中所有模型共用同一套分层划分结果和同一套训练、验证和测试样本, 消除数据集样本数量、类别分布差异带来的实验干扰, 确保不同算法之间的性能对比具备公平性和可信度。此外, 为了后续能够进行统计学显著性检验, 本文在对比深度学习模型时, 额外实施了 10 折交叉验证策略获取多组评估指标。

2.2. 模型实验参数设置

针对传统统计基线模型, 本文对 HOG 手工特征提取流程设置统一标准化参数, 将输入图像统一缩放至 64×64 分辨率, 梯度方向划分为 9 个区间, 单个像素单元格尺寸设定为 8×8 , 以 2×2 单元格组合为基础分块单元, 计算过程不输出梯度可视化热力图以降低存储开销。LBP 采用均匀模式算子, 半径设为 1、采样点数为 8, 用以刻画局部纹理分布。HSV 颜色空间特征则提取三通道的一阶均值、二阶方差及三阶颜色矩, 用以描述全局色彩分布。在分类器配置层面, 选取 RBF 核函数实现非线性分类, 开启概率输出功能并固定随机种子为 42 以保证实验可复现。随机森林集成模型内置 100 棵独立决策树, 启用全线程并行加速训练模式, 同步设置相同随机种子消除采样随机性带来的实验偏差。XGBoost 则基于多分类 softmax 损失函数进行优化, 迭代轮次设为 100。数据集拆分环节采用分层抽样策略对原始训练集执行 9:1 划分, 划分随机种子固定取值 42, 分层机制能够严格保留样本的原始分布占比, 避免子集类别失衡对分类结果造成干扰。

所有深度学习模型共用同一套基础实验超参数, 统一设置批次尺寸为 32, 完整训练迭代轮次固定为 10 轮。图像预处理流程区分训练子集与验证、测试子集执行差异化变换, 训练样本引入随机缩放裁剪、随机水平翻转的数据增强策略扩充样本多样性, 验证与测试样本仅执行尺寸缩放与中心裁剪操作, 全部输入图像统一归一化至 224×224 标准尺寸。训练损失函数统一采用多分类交叉熵损失, 参数优化器选用 AdamW 算法, 权重衰减系数设置为 $1e-2$ 实现正则化约束。学习率调度机制采用余弦退火策略, 调度周期取值为 10, 与整体训练轮数保持匹配。实验全程固定数据集划分随机种子为 42, 每轮迭代完整记录训练损失、验证损失、训练集准确率、验证集准确率与实时学习率数值, 全部迭代结束后加载验证集准确率最优的模型权重, 在独立测试集上完成泛化性能评估。

三个深度学习网络依据自身结构特性配置差异化初始学习率, 其中 ResNet18 与 MobileNetV3 的初始学习率为 $1e-4$, Swin-Tiny 架构对学习率变化更为敏感, 为规避训练震荡问题将初始学习率下调至 $5e-5$ 。三个网络训练前均加载 ImageNet 大型通用图像数据集的官方预训练权重完成参数初始化, 仅对网络末端分类输出层进行重构, 调整输出维度以适配本实验数据集包含的垃圾细分类别, 实现通用视觉特征向细分垃圾分类任务的迁移。

2.3. 模型评估指标

为全面客观衡量 SVM、随机森林、XGBoost、ResNet18、MobileNetV3、Swin-Tiny 六类模型在垃圾分类任务中的综合识别性能,避免单一指标造成评价偏差,本文构建多维度分层评价指标体系,分别从全局分类精度、类别均衡识别能力、模型判别区分度、训练收敛稳定性、类别错分分布规律五大维度开展量化分析,全部指标均基于独立测试集样本计算,排除训练、验证集数据干扰,保证评估结果真实反映模型泛化能力。具体而言,采用多维度综合评价指标,包括准确率(Accuracy)、宏精确率(Macro-Precision)、宏召回率(Macro-Recall)、宏 F1 值(Macro-F1)、ROC 曲线与 AUC 值,同时结合训练损失、训练准确率、验证损失、验证准确率分析模型收敛性能,通过混淆矩阵分析类别错分规律。由于本实验包含可回收物、厨余垃圾、有害垃圾、其他垃圾四大不平衡大类,为公平衡量模型对四类样本的均衡识别能力,本文采用宏平均(Macro-average)方式计算四类样本各自的精确率、召回率、F1 值,不受各类样本数量差异影响,客观体现模型对小众小样本垃圾大类的识别水平。

为量化模型对四个垃圾大类分类边界的区分能力,本文基于模型输出的类别预测概率绘制多分类 Macro-ROC 曲线并计算 AUC 值。对于 SVM、随机森林与 XGBoost,通过开启概率输出接口得到样本归属四大类别的后验概率;对于 ResNet18、MobileNetV3、Swin-Tiny 深度学习模型,借助网络末端 Softmax 层输出归一化类别概率分布。AUC 取值区间为[0, 1],数值越趋近于 1,代表模型对不同垃圾大类的判别边界区分效果越好,类别误判概率越低。实验完整记录每一轮迭代过程中的训练损失、验证损失、训练集准确率、验证集准确率,通过损失与精度变化曲线直观分析模型收敛状态,判断是否存在欠拟合、过拟合问题。若训练损失持续下降,同时验证损失逐步趋于平稳无明显回升,说明模型收敛稳定,若验证损失持续升高则判定出现过拟合。本文仅保留验证集准确率峰值对应的最优模型权重开展测试集评估,规避过拟合权重带来的泛化性能偏差。

3. 实验结果与分析

基于华为云四大类生活垃圾数据集,对 SVM、随机森林、XGBoost 三个传统统计基线模型与 ResNet18、MobileNetV3、Swin-Tiny 三个深度学习模型的训练过程、量化指标、判别能力与类别错分规律展开分析,全部实验评估均在独立测试集完成,结合模型训练收敛曲线、多模型的指标对比图、Macro-ROC 曲线、归一化混淆矩阵等可视化图表,从训练稳定性、整体分类精度、类别判别能力、错误案例分析等方面展开对比论证。

3.1. 模型训练收敛过程分析

从图 1 所示的深度学习训练曲线可以看出,三个深度模型均呈现稳定收敛趋势,训练损失持续下降、训练准确率持续上升,验证损失逐步趋于平稳,无明显过拟合与欠拟合现象。余弦退火学习率策略有效降低了训练后期的参数震荡,保证模型能够收敛至最优局部解。从训练损失与验证损失曲线可以看出,三种深度学习模型在训练过程中均表现出较为稳定的下降趋势,表明模型参数能够逐步收敛。其中, Swin-Tiny 的训练损失下降较快,且验证损失同步降低,说明该模型具有更强的特征拟合能力和泛化能力。ResNet18 与 MobileNetV3 的损失曲线也整体平稳,没有出现明显震荡或发散,表明当前学习率设置和优化策略较为合理。所有模型均基于验证集最优权重完成测试,保证泛化性能最优。

在准确率曲线方面,三种模型的训练准确率和验证准确率均随 epoch 增加而逐步提升,且两者之间的差距较小,说明模型未出现严重过拟合。Swin-Tiny 最终在训练和验证阶段都达到较高准确率,反映出其结构在该任务上的适应性更强。MobileNetV3 虽然结构轻量,但准确率提升同样明显,说明轻量化网络在垃圾分类任务中依然具有较好的应用价值。训练结束时 Swin-Tiny 验证集准确率达到全场最高, MobileNetV3 次之, ResNet18 略低,与损失曲线规律完全匹配。

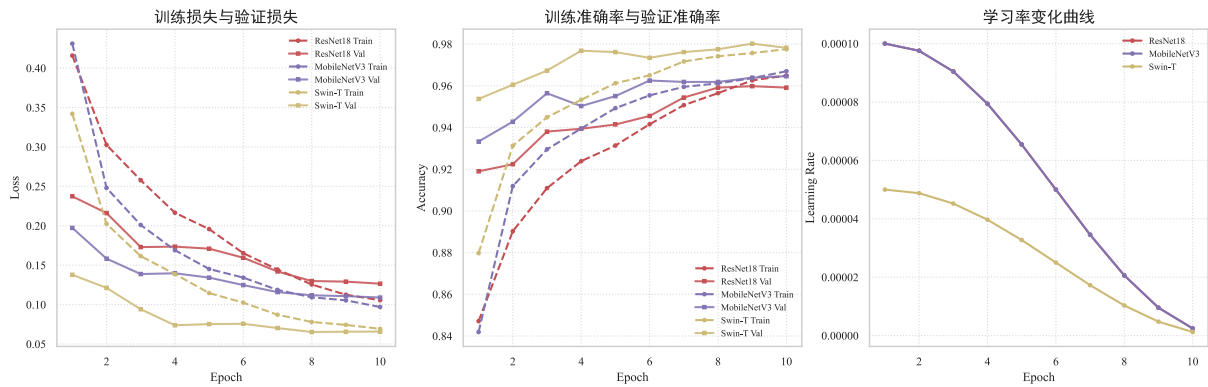


Figure 1. Training process of deep learning models

图 1. 深度学习模型训练过程曲线

右侧学习率衰减曲线为余弦退火调度结果，从图中可以看出，ResNet18、MobileNetV3 和 Swin-Tiny 的学习率均随着训练轮数增加而逐渐下降，符合余弦退火策略的设计目标。学习率的逐步衰减有助于模型在训练前期快速搜索参数空间，在后期进行细致收敛，从而提高最终性能。MobileNetV3、ResNet18 初始学习率设置为 $1e-4$ ，Swin-Tiny 因 Transformer 结构训练敏感性，初始学习率调低至 $5e-5$ ，二者均按余弦曲线平滑衰减至 0。前期较大学习率快速完成基础特征拟合，后期极小学习率精细微调分类层权重，避免训练后期参数震荡，保障收敛稳定。需要说明的是，图中未绘制 SVM、RF、XGBoost 的学习率变化曲线，这是因为传统机器学习模型无迭代梯度下降意义上的动态学习率调度过程，XGBoost 虽在训练时存在固定学习率超参数用于约束单棵决策树更新幅度，但该参数为全局固定超参，不会随训练轮次动态衰减调度，因此无连续变化的学习率曲线可供可视化，而 SVM、随机森林本身不存在学习率相关参数。

在三种深度学习模型中，ResNet18、MobileNetV3 和 Swin-Tiny 均取得了较好的测试表现，说明迁移学习能够有效提升垃圾分类任务的收敛速度与最终性能。ResNet18 通过残差连接实现了较稳定的梯度传播，在中等规模数据集上具有较强的适应性。MobileNetV3 由于参数量较小，在保持较高精度的同时具有较好的计算效率。Swin-Tiny 通过局部窗口注意力机制建模图像上下文关系，通常能够捕捉更丰富的特征表示，因此在复杂类别判别上具备潜在优势。

3.2. 多模型量化指标对比分析

对六个模型分别进行指标计算如下表 1 所示。

Table 1. Performance comparison of different models

表 1. 不同模型的性能对比

模型	准确率	宏精确率	宏召回率	宏 F1	训练耗时 (s)	参数量 (M)	计算量 (GMac)	单图推理 (ms)	推理 (FPS)
SVM	0.6487	0.4861	0.4109	0.4010	910.3	-	-	-	-
RF	0.6130	0.3432	0.3358	0.3296	9.21	-	-	-	-
XGBoost	0.6134	0.4271	0.4206	0.4070	57.06	-	-	-	-
ResNet18	0.9155	0.8598	0.8853	0.8685	1142.71	11.18	1.82	5.35	186.9
MobileNetV3	0.9419	0.8891	0.9218	0.9042	1112.17	4.21	0.23	11.59	86.3
Swin-Tiny	0.9587	0.9221	0.9429	0.9319	1968.55	27.52	3.13	26.29	38.0

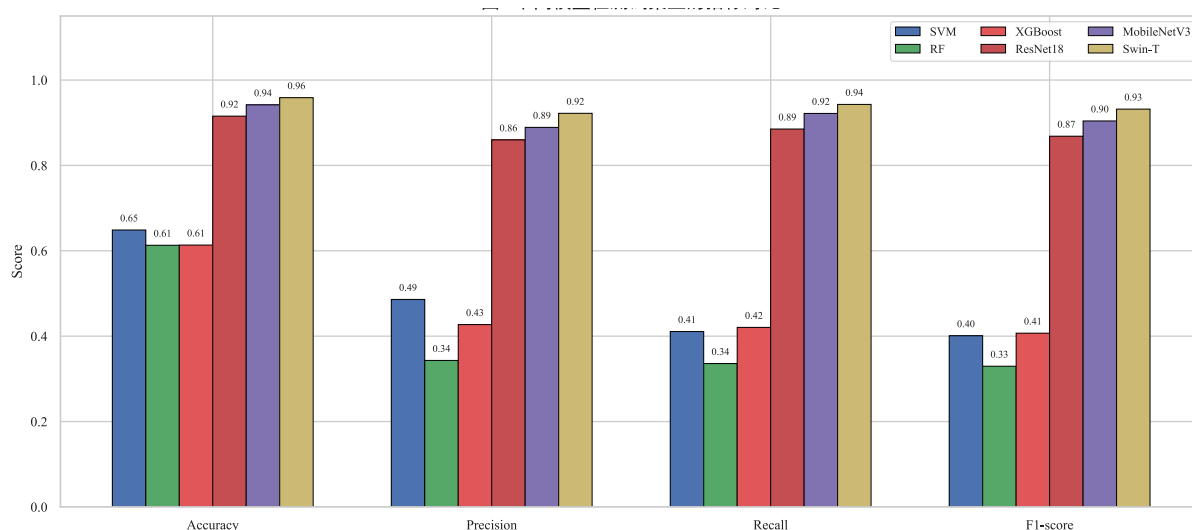


Figure 2. Performance of different models on testing data

图 2. 不同模型在测试集上的性能对比

从 Accuracy、Precision、Recall 和 F1-score 四项指标对比结果图 2 可以看出,深度学习模型整体显著优于三种传统机器学习基线模型,这表明对于垃圾分类图像任务,仅依赖 HOG、LBP、HSV 等手工特征难以充分刻画图像的复杂视觉语义信息,端到端深度特征学习在垃圾分类任务中具有显著优势。其中, Swin-Tiny 在四项指标上均取得了最优结果,表明其在全局上下文建模和局部窗口注意力机制方面具有更强的表示能力,能够更充分地提取图像中的语义特征。MobileNetV3 和 ResNet18 的整体性能接近,但前者在分类精度与召回率等各项指标上均略优于后者,说明 MobileNetV3 作为轻量化网络在分类准确性和泛化能力上取得了更好的平衡。此外, MobileNetV3 作为轻量化模型,在保持较高分类性能的同时,计算成本相对可控,具有较好的工程应用潜力。

传统机器学习模型中, SVM、随机森林和 XGBoost 均能完成基本分类任务,但整体性能受限于特征表达能力。SVM 的整体性能优于随机森林,这说明在当前数据集上,基于 HOG 等特征的 SVM 更适合处理该分类任务。随机森林的四项指标均偏低,反映出基于树模型的集成分类器在面对高维、稀疏且类别边界复杂的图像特征时,难以形成足够稳健的判别边界。两者均无法像深度网络那样自动学习高层语义特征,因此在复杂类别识别中表现相对有限。

在实际部署场景中,垃圾分类模型的推理速度与分类精度之间存在固有的权衡关系,不同应用场景对二者的侧重各不相同。为系统评估各模型在多样化部署条件下的适用场合,本文引入推理 FPS 与测试集准确率两个维度进行联合分析,结果如图 3 所示。

图中依据推理速度将应用场景划分为三个区域,分别为云端高精度场景、本地均衡场景及边缘嵌入式场景。在云端高精度场景中,模型可部署于高性能 GPU 服务器,对推理延迟容忍度较低,但要求尽可能高的分类准确率。Swin-Tiny 以 38.0 FPS 的推理速度与 0.9587 的准确率落于该区域,凭借 Transformer 架构的全局建模能力在精度上显著领先,但受限于较大的参数量与计算量,其推理吞吐量相对有限,适用于对分类精度要求严苛、推理资源充足的云端或中心化服务场景。

在边缘嵌入式场景中,模型需在有限的计算资源下实现实时推理,对推理速度要求极高。ResNet18 以 186.9 FPS 的推理速度位居首位,准确率仍保持在较高水平,是实时性要求极高的边缘场景,如智能垃圾分类分拣流水线等的优选方案。MobileNetV3-Large 在测试集取得 86.3 FPS、0.9419 准确率,相较 ResNet18,

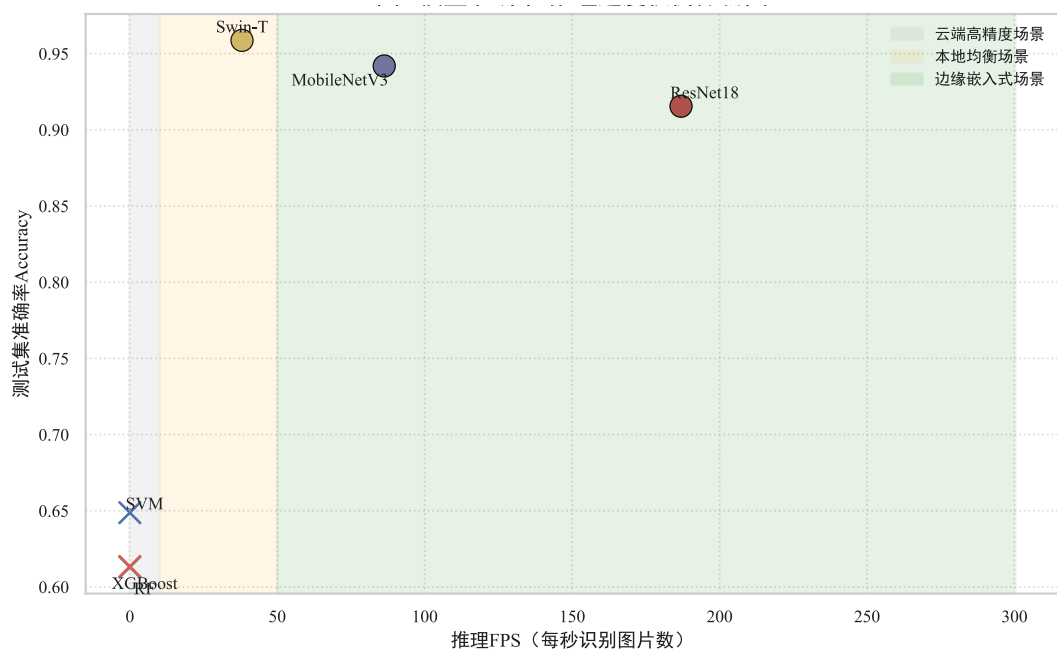


Figure 3. Trade-Off between inference speed and accuracy of different models on testing data
图 3. 不同模型在测试集上的推理速度与准确率权衡图

其推理吞吐量存在明显下降,主要源于轻量化网络内置深度可分离卷积、SE 通道注意力模块与 Hard-Swish 非线性激活函数,该类逐通道运算会增加串行计算耗时,在本实验硬件环境下降低单图推理速度。但该模型参数量仅 4.21 M、理论浮点计算量远低于 ResNet18,模型存储开销更小,可在内存、闪存资源极度受限的低阶嵌入式终端部署,以小幅推理延迟为代价换取更高分类精度,在精度与存储占用之间实现更优平衡,适用于算力有限的移动端智能垃圾桶设备。而 ResNet18 无额外注意力分支,计算流程简洁,推理吞吐量可达 186.9 FPS,更适配需要连续高速分拣的流水线设备。

综合来看, Swin-Tiny 适用于云端高精度推理场景, MobileNetV3-Large 在精度与效率之间实现了最佳平衡,在边缘端具有广泛应用前景, ResNet18 以最高的推理吞吐量满足极限实时性需求。上述分析为垃圾分类系统在实际部署中的模型选型提供了量化依据。

3.3. ROC-AUC 性能分析

图 4 为六类模型基于测试集预测概率绘制的 Macro-ROC 曲线与对应 AUC 数值,曲线包裹面积代表模型对四个垃圾大类的判别边界区分能力。ROC 曲线越贴近左上角、AUC 数值越接近 1,说明模型在较低假阳性率下能够获得更高的真阳性率,代表模型判别区分性能越强。多分类 Macro-ROC 曲线表明,深度学习模型的整体判别能力显著优于传统机器学习模型。三个深度学习模型 ROC 曲线几乎紧贴左上角上边界,AUC 数值均超过 0.98。其中 Swin-Tiny 的 ROC 曲线最接近左上角,AUC 值最高,达到 0.9941,说明其能够在较低误报率下获得更高的真阳性率,模型区分能力最强。MobileNetV3 和 ResNet18 的 ROC 曲线同样表现良好,AUC 维持在较高水平,说明其对各类垃圾样本均具有较强的可分性。该结果证明端到端自动特征学习能够挖掘区分四大垃圾大类的核心判别特征,模型输出的类别置信概率可信度高,类别边界划分清晰,大幅降低大类之间混淆误判概率。相比之下,SVM、随机森林和 XGBoost 的 ROC 曲线明显远离左上角,AUC 值相对较低,随机森林 Macro-AUC 仅 0.6554,SVM 为 0.7345,说明其在多分类判别任务中排序能力不足。这与前述混淆矩阵和总体指标结果保持一致,再次说明手工特征配合传统

分类器在复杂图像任务中的表达能力有限。Macro-ROC 曲线与 AUC 值反映模型整体判别能力，实验结果表明，深度学习模型的曲线覆盖面积显著大于传统统计模型，说明端到端特征学习能够有效提升类别判别边界的区分度。Swin-Tiny 的 AUC 值最高，判别误差最小，对各类别样本的区分能力最为均衡，RF 模型 ROC 曲线贴近对角线，判别不确定性高，易出现错分、误判现象。

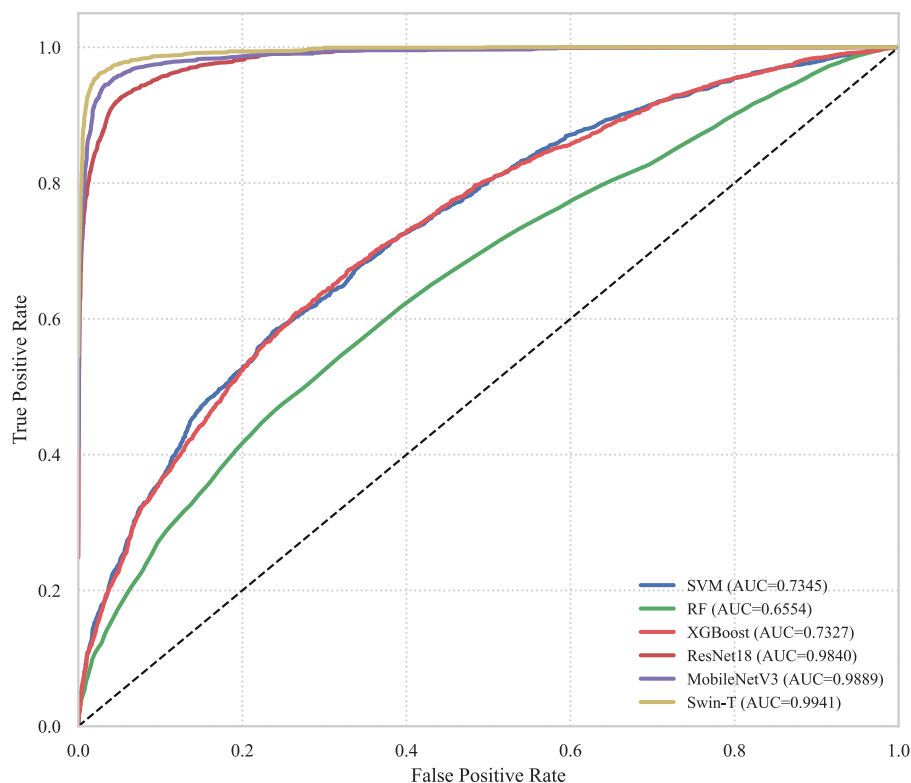


Figure 4. Macro-ROC comparison about different models

图 4. 多模型 Macro-ROC 曲线对比

为验证深度学习模型之间分类性能差异的统计显著性，避免因随机性导致的结论偏差，本文引入单因素方差分析(ANOVA)基于 10 折交叉验证对三种深度模型在测试集上的准确率进行假设检验，并在方差分析通过后进一步采用 Tukey HSD (Honestly Significant Difference)事后多重比较方法进行两两对比。方差分析结果显示，F 统计量为 61.5154，p 值小于 0.001，表明三种深度学习模型在分类准确率上存在极显著的统计学差异，拒绝三组模型性能无差异的原假设。Tukey HSD 事后检验的两两对比结果进一步揭示了差异的具体来源。MobileNetV3 与 ResNet18 之间的均值差为 0.0264，拒绝两者性能相同的原假设，表明 MobileNetV3 在测试集上的准确率显著优于 ResNet18。MobileNetV3 与 Swin-Tiny 之间的均值差为 0.0168，差异同样显著，Swin-Tiny 的准确率显著高于 MobileNetV3。ResNet18 与 Swin-Tiny 之间的均值差为 0.0432，是三者中差异最大的一对，进一步确认了 Swin-Tiny 在垃圾分类任务中的领先地位。该结论排除了实验随机波动对模型排序的干扰，增强了本文对比分析结论的可靠性与说服力。

3.4. 混淆矩阵分析

为进一步评估各模型在不同垃圾类别上的分类细粒度表现，本文在测试集上绘制了归一化混淆矩阵热力图，如图 5 所示。横坐标为模型预测类别，纵坐标为样本真实类别，0、1、2、3 依次对应四大垃圾

基础类别，0 代表其他垃圾、1 代表厨余垃圾、2 代表可回收物、3 代表有害垃圾，色块颜色越深代表该类别正确识别占比越高。从各模型的归一化混淆矩阵可以看出，深度学习模型的对角线区域显著更深，说明其在各类别上的分类准确率较高，误分类较少。其中，ResNet18、MobileNetV3 和 Swin-Tiny 的混淆矩阵几乎呈现明显的对角占优结构，表明这三个模型能够较好地地区分垃圾图像的不同类别，具备较强的特征学习和判别能力。相比之下，SVM、随机森林和 XGBoost 的对角线虽然仍然可见，但非对角线区域存在更明显的颜色分布，说明其对类别边界的区分能力有限，特别是在部分视觉相似类别上更容易发生混淆。

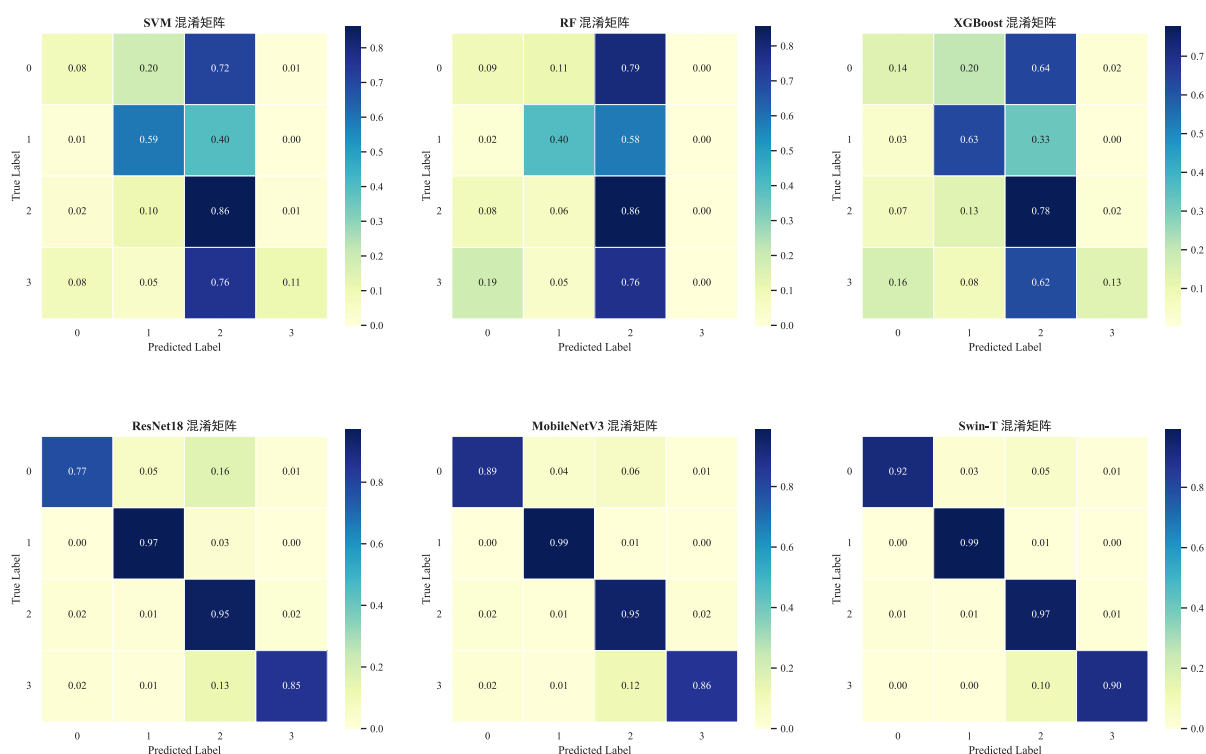


Figure 5. Confusion matrices of different models on the testing dataset

图 5. 各模型在测试集上的归一化混淆矩阵

具体而言，SVM 的误分类主要集中在若干类别之间的交叉区域，这表明 HOG 等边缘与轮廓描述子在捕捉垃圾图像的复杂纹理、材质变化及背景干扰等方面存在明显不足。随机森林的混淆矩阵则进一步显示，其在多个类别上的正确识别概率低于 SVM，尤其在训练样本相对较少或类别特征不显著的类别上，预测结果更容易偏向于样本量占优的主导类别，反映出基于 Bagging 集成的决策树分类器在高维图像特征空间中难以建立稳定且精确的判别边界。XGBoost 虽通过梯度提升在一定程度上缓解了偏差，但整体混淆程度仍较高，其核心错分集中在其他垃圾、有害垃圾被大量误判为可回收物，这与两者外观形态多变、类间边界模糊的特性密切相关。

在深度学习模型中，表现最优的 Swin-Tiny 的归一化混淆矩阵显示，其主要误差集中在视觉特征高度相似的类别之间，是生活垃圾分类任务的共性难点。相比统计模型，深度学习模型能够大幅降低相似品类的错分概率，其中 Swin-Tiny 对细粒度特征的区分能力最优，错分样本数量最少。统计模型由于特征表达单一，各类别均存在大量误判，均衡性较差。归一化混淆矩阵显示，最优模型在多数类别上能够形

成较明显的对角线分布,说明分类结果较为集中,误分类比例较低。但在某些外观相似或材质相近的垃圾类别之间,仍可能存在混淆现象。这说明垃圾分类任务不仅依赖图像表面纹理,还与物体的形状、材质和局部结构有关。MobileNetV3-Large 与 ResNet18 次之,两者均存在少量误判,但整体可控。

总体而言,混淆矩阵结果表明,迁移学习深度模型能够有效提升垃圾分类任务的类别可分性,其中 Swin-Tiny 与 MobileNetV3 的分类结果最为稳定,ResNet18 次之,而传统机器学习模型在复杂视觉场景下表现相对较弱。

3.5. 错误案例分析

结合归一化混淆矩阵统计结果,选取两类全模型普遍存在的高频混淆组合,分别为其他垃圾误分为可回收物、有害垃圾误分为可回收物,传统模型在两类样本上错分占比极高,深度学习模型虽可以大幅缓解,但仍存在少量误判,提取各模型代表性错分样本可视化如图 6 所示。

第一类高频错分为其他垃圾被模型误判为可回收物。一次性塑料餐盒等废弃日用塑料,以及铝制分格饭盒等金属餐具虽归属其他垃圾,但其光滑反光表面、规整几何轮廓等浅层视觉特征,与可回收塑料、金属制品高度趋同。依靠手工浅层特征建模的 SVM、XGBoost 对该类样本识别缺陷突出,超六成其他垃圾样本会被错误划分至可回收物类别。三个深度学习模型大幅降低了错分比例,但对于表面布满污渍、形态破损的轻薄塑料废弃物,仍会产生少量判别偏差。从图 6 可视化样本可见, SVM、RF、XGBoost、MobileNetV3、Swin-Tiny 均存在将塑料餐盒类其他垃圾误判为可回收物的情况,其中 ResNet18 易将泡沫分格餐盒误判为有害垃圾。



Figure 6. Common high-frequency error cases of different models on the testing dataset
图 6. 各模型在测试集上的高频典型错误案例

第二类高频错分为有害垃圾被模型误判为可回收物。废旧金属外壳电池、金属药膏管拥有和金属可回收物一致的金属反光浅层特征，模型仅依靠局部纹理极易混淆。Swin-Tiny 依靠全局注意力捕捉药剂、警示标识等细微区分特征，该类错分样本数量显著少于 ResNet18 与 MobileNetV3。ResNet18 仅依靠局部卷积感受野提取色块纹理，易被局部污渍、纯色塑料表面干扰，判别稳定性较弱。MobileNetV3 轻量化卷积特征提取能力有限，全局上下文建模不足，同样存在明显混淆。Swin-Tiny 依靠分层全局注意力整合全图纹理、色彩、轮廓综合信息，能够兼顾材质属性与外观形态，该类样本错分占比显著低于另外两种 CNN 模型。

综合两类典型错分样本可得出，现有模型判别偏差主要来源于类别间视觉特征重叠、国标分类语义与图像浅层特征不匹配、实景污渍或干燥形态带来的特征干扰。轻量化 CNN 受限于感受野与特征提取容量，更容易被局部浅层纹理误导产生误判。Swin-Tiny 全局自注意力结构可融合长距离上下文信息，弱化干扰特征的负面影响，细粒度区分能力更优，混淆样本数量显著减少。

4. 总结

本文针对国家标准四类生活垃圾图像自动识别任务，构建手工特征融合统计模型与深度学习网络对照实验体系，选取 SVM、随机森林、XGBoost 三个传统机器学习模型，以及 ResNet18、MobileNetV3、Swin-Tiny 三个深度视觉模型开展控制变量下的对比实验，结合模型算力测算、统计显著性检验、错分样本特征可解释分析，形成完整量化论证。结合文献综述梳理的现有研究成果，本文所得实验结论可划分为对已有成果的验证与补充拓展，能够针对性弥补当前垃圾分类视觉识别领域存在的研究空白，明晰本研究的学术价值。

在性能规律层面，本文实验结果验证了现有迁移学习垃圾分类研究形成的统一共识，依托 ImageNet 预训练权重，结合标准化数据增强策略微调的端到端深度学习模型，可自主挖掘图像高层语义判别特征，综合分类性能全面优于依赖人工设计特征的传统机器学习算法。实验结果表明，多数传统统计模型(如随机森林、XGBoost)具备训练耗时短的优势，但 HOG-LBP-HSV 融合手工特征的表达能力存在固有上限，面对实景遮挡、材质外观高度近似等复杂样本时泛化性能显著衰减，细粒度垃圾区分精度存在难以突破的瓶颈，仅可作为严苛的对照基线。与之相对，深度网络能够稳定提取垃圾全局轮廓与局部细粒度判别信息，是提升复杂实景生活垃圾识别精度的有效技术路径，该结论与已有相关文献的研究结论保持一致。

在创新层面，本文从三个方面补充拓展了现有研究缺失的实验内容。第一，针对过往多数研究仅采用单一 HOG 特征与 SVM、随机森林等构建薄弱基线的局限，本文提出 HOG-LBP-HSV 轮廓、纹理、色彩多维度特征融合方案，并引入 XGBoost 梯度提升树搭建性能更强的传统对照基准，定量证实多类手工特征融合可显著抬升传统机器学习模型的识别上限，完善了手工特征体系下传统算法性能边界的量化实证数据，弥补了现有研究传统基线构建不完善、对比说服力不足的空白。第二，现有同类研究多仅汇报分类准确率单一指标，缺乏面向工程落地的算力开销量化分析，本文新增模型参数量、FLOPs 计算量、单图推理耗时、推理 FPS 全套工程指标，量化三个深度网络的精度与算力权衡关系，ResNet18 残差卷积网络训练过程稳定、实景干扰鲁棒性均衡，是兼顾识别精度与硬件开销的通用型基础模型，虽然在本实验的 GPU 环境下 MobileNetV3 的推理帧率不及 ResNet18，但其极小的模型体积和极低的理论计算量，使其在内存和闪存极度受限的 ARM/移动端架构边缘设备上具备更大的落地潜力，Swin-Tiny 分层窗口自注意力机制具备全局上下文建模能力，对易混淆细粒度样本识别效果最优，适用于追求极致精度的工业自动化分拣流水线，但存在计算量大、推理延迟更高的短板，填补了垃圾分类任务分场景模型选型量化依据缺失的研究空白。第三，现有对比类实验普遍缺乏统计学论证与模型误差可解释分析，本文引入单因素 ANOVA 方差分析与 Tukey HSD 多重事后检验，从统计学层面证实三个深度网络的性能差距并非随

机实验波动, 研究结论具备统计可信度, 同时结合归一化混淆矩阵与典型高频错分样本可视化, 从图像纹理色彩、国标分类语义边界、实景噪声干扰三个维度剖析模型判别偏差的内在成因, 揭示轻量化卷积网络易受局部浅层特征误导、窗口注意力 Transformer 可弱化局部干扰特征影响的机理, 完善了垃圾分类模型误差形成机制的可解释性。

综上, 本文在统一数据集和实验配置的条件下, 既充分验证了深度学习识别性能优于手工特征传统模型的结论, 又从传统基线优化、工程算力定量评估、统计严谨性与误差可解释分析三个维度补充了现有文献未覆盖的实验内容, 搭建起兼顾学术论证严谨性与工程落地指导价值的生活垃圾多分类对照实验框架, 可为不同硬件部署场景下智能垃圾分拣系统的模型选型提供完整、可信的量化实证支撑。

参考文献

- [1] Yang, G. and Thung, G. (2016) Classification of Trash for Recyclability Status. CS229 Project Report, 3. <https://cs229.stanford.edu/proj2016/report/>
- [2] Gaurav, A., Gupta, B.B., Arya, V., Attar, R.W., Bansal, S., Alhomoud, A., *et al.* (2025) Smart Waste Classification in IoT-Enabled Smart Cities Using VGG16 and Cat Swarm Optimized Random Forest. *PLOS ONE*, **20**, e0316930. <https://doi.org/10.1371/journal.pone.0316930>
- [3] Fotovvatikhah, F., Ahmady, I., Noor, R.M. and Munir, M.U. (2025) A Systematic Review of AI-Based Techniques for Automated Waste Classification. *Sensors*, **25**, Article No. 3181. <https://doi.org/10.3390/s25103181>
- [4] Rabano, S.L., Cabatuan, M.K., Sybingco, E., Dadios, E.P. and Calilung, E.J. (2018) Common Garbage Classification Using MobileNet. 2018 *IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, Baguio City, 29 November-2 December 2018, 1-4. <https://doi.org/10.1109/hnicem.2018.8666300>
- [5] Mittal, G., Yagnik, K.B., Garg, M. and Krishnan, N.C. (2016) SpotGarbage: Smartphone App to Detect Garbage Using Deep Learning. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Heidelberg, 12-16 September 2016, 940-945. <https://doi.org/10.1145/2971648.2971731>
- [6] Aral, R.A., Keskin, S.R., Kaya, M. and Haciomeroglu, M. (2018) Classification of TrashNet Dataset Based on Deep Learning Models. 2018 *IEEE International Conference on Big Data (Big Data)*, Seattle, 10-13 December 2018, 5058-5062. <https://doi.org/10.1109/bigdata.2018.8622212>
- [7] 肖克江, 陈亮, 高阔, 等. 基于多模态数据融合的边缘设备轻量级垃圾分类方法和系统研究[J]. *工程科学学报*, 2025, 47(9): 1905-1916.
- [8] Huang, K., Lei, H., Jiao, Z. and Zhong, Z. (2021) Recycling Waste Classification Using Vision Transformer on Portable Device. *Sustainability*, **13**, Article No. 11572. <https://doi.org/10.3390/su132111572>
- [9] Deng, S., Fan, A. and Sun, J. (2022) Waste Classification and Management Using Computer Vision. Stanford University.
- [10] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., *et al.* (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 10012-10022. <https://doi.org/10.1109/iccv48922.2021.00986>