arXiv:2601.04727v1 [cs.CV] 8 Jan 2026

在五个异构图像数据集上训练自定义卷积神经网络

AnikaTabassum[1], Tasnuva MahazbinTuba[1], NafisaNaznin[1]1达卡大学计算机科学与工程系，达卡，孟加拉国。

贡献作者：anika-2019417844@cs.du.ac.bd；
tasnuvamahazabin-2018725348@cs.du.ac.bd；
nafisanaznin24cse041@gmail.com；

1 引言

深度学习改变了视觉数据的分析方式，卷积神经网络（CNN）逐渐成为从图像中提取有意义模式的强大工具。不同于依赖人工特征工程的传统方法，CNN 能够直接从数据中学习层次化表示，因此在从农业监测到城市基础设施评估等广泛的现实应用中都十分有效。

在本研究中，我们针对来自农业和城市领域的五个异构数据集，开发并评估了多种基于卷积神经网络的架构：芒果品种分类、水稻品种识别、道路表面状况评估、交通场景中的自动人力车检测，以及人行道侵占监测。这些数据集具有截然不同的挑战——它们在图像分辨率、光照条件、类别不平衡和主体复杂度方面存在显著差异——因此需要模型能够适应多样化的视觉场景，同时保持稳健的分类性能。

我们的实验框架将轻量级、面向特定任务的自定义卷积神经网络与成熟的深度架构（即 ResNet-18 和 VGG-16）进行比较，这些模型分别从头开始训练和采用迁移学习进行训练。通过系统的数据预处理与增强，我们分析架构深度、参数数量和预训练如何影响模型在规模和复杂度各异的数据集上的学习行为与泛化能力。本研究的主要贡献包括：(1) 设计并评估了一种计算效率高的自定义卷积神经网络，该网络在多个领域均具有竞争力；(2) 对从头训练的

模型与基于迁移学习的深度模型进行了全面的比较分析，为它们在数据受限的现实世界视觉分类任务中的适用性提供了见解。

2 数据集描述

以下五个数据集用于评估模型在不同农业与城市监测应用中的表现。

2.1 智能城市中的未经授权车辆检测（孟加拉国）

该数据集是深度学习研究的一部分，旨在自动检测孟加拉国城市交通场景中的自动人力车（“未经授权的车辆”）。由于许多交通规则限制机动人力车在某些路线上行驶，因此对其进行监控至关重要；但在监控图像中将它们与脚踏人力车区分开来并不容易。该数据集包含在各种真实环境条件下采集的1,730张已标注交通图像，并采用适用于训练基于YOLO的模型的目标检测标签。作者使用YOLOv8展示了出色的检测性能（mAP50 $\approx$83.4%），使该数据集对交通监控、智慧城市安全和交通管理研究具有重要价值。 [1]

2.2 FootpathVision：人行道侵占数据集

FootpathVision是一个专用图像数据集，旨在检测城市环境中阻碍行人通行的人行步道侵占行为。这类侵占包括街头摊贩、违章停放车辆以及各种影响人行道通行能力的障碍物。该数据集涵盖了各种真实的人行步道场景，非常适合开发用于识别智慧城市环境中基础设施违规行为的深度学习模型。这项资源有助于推动城市规划举措、加强行人安全措施，并协助市政执法活动。 [2]

2.3 道路损坏与井盖检测（孟加拉国）

该数据集包含 1,000 多张道路表面图像，主要采集自孟加拉国达卡。图像经过多边形边界标注，可精确识别道路裂缝（“破损”）、完好路段（“未破损”）和井盖。与传统的边界框标注不同，使用多边形形状可以更准确地定位不规则的道路缺陷。基于该数据集训练的 YOLOv9 模型，图像级准确率达到 78.1%，展现了其在新兴智慧城市中实现基础设施监测与维护自动化方面的潜力。 [3]

2.4 MangoImageBD：芒果品种识别数据集

MangoImageBD 是一个综合数据集，包含 28,515 张图像，展示了孟加拉国广泛种植的 15 个芒果品种，其中包括 Amrapali、Bari-11、Fazli 和 Langra 等。该数据集分为三个不同的子集。

原始子集（5,703 张图像）由白色背景下拍摄的未经处理的芒果照片组成。真实 + 虚拟子集（5,703 张图像）包含以真实背景和合成背景组合为特色的芒果照片。增强子集（17,109 张图像）包含通过旋转、翻转、亮度调整和噪声注入等数据增强技术生成的图像，以提升数据集多样性和模型鲁棒性。芒果样本采自孟加拉国六个地区，并使用高清智能手机相机，在标准化设置下拍摄。该数据集对于训练分类模型很有价值，可用于促进分拣和分级等农业自动化任务，并支持生物多样性和表型变异研究。[4]

2.5 PaddyVarietyBD：稻谷（大米）品种分类数据集

PaddyVarietyBD 是一个显微图像数据集，涵盖了孟加拉国种植的 35 个稻谷（大米）品种，图像由孟加拉国稻米研究所（BRRI）和孟加拉国核农业研究所（BINA）提供。该数据集包含 14,000 张原始图像，这些图像是以 640×480 像素分辨率拍摄的稻谷籽粒 RGB 显微照片。此外，数据集还包含 56,000 张增强图像，这些图像通过旋转、翻转和剪切变换合成生成，以重现视觉变异性并提升模型训练效果。凭借广泛的品种覆盖范围和丰富的样本量，该数据集是开发人工智能驱动的分类系统的重要资源，并为农业领域的农学研究、植物育种项目和计算机视觉应用提供了重要价值。[5]

3 方法

本节介绍所采用的实验流程，以确保对三种建模方法进行公平、系统的比较：从头开始训练的自定义卷积神经网络、用作固定特征提取器的预训练卷积神经网络模型，以及通过迁移学习进行微调的预训练卷积神经网络。为消除实验偏差，所有模型均在所有数据集上使用相同的数据划分、预处理步骤和评估方案进行训练和评估。

3.1 数据预处理

所有五个数据集均采用一致的预处理流程，以确保与所有评估模型兼容，包括自定义卷积神经网络、预训练卷积神经网络架构和迁移学习模型，同时保留数据集特定特征。统一的预处理策略对于确保不同建模方法之间进行公平比较至关重要。

预处理流程包含四个主要阶段：数据集准备、训练集与验证集划分、图像缩放和归一化，以及数据增强。这些

步骤统一应用于所有实验，以便将性能差异归因于模型设计，而非数据处理方式的差异。

3.1.1 数据集准备

每个数据集都被重新整理为统一的目录格式，以兼容 PyTorch 的 ImageFolder 加载器。**PaddyVarietyBD** 需要合并涵盖 35 个水稻品种类别的原始和增强档案。**FootpathVision**从独立的 ZIP 文件中提取二元类别（被侵占/未被侵占）。**MangoImageBD** 涉及展平三个嵌套档案（原始、增强、真实+虚拟），并通过清理类别名称整合 15 个芒果品种。**Auto-RickshawImageBD** 解析 YOLO 格式标签，将图像整理到二元文件夹（非自动人力车/自动人力车）中。**路面数据集** 从独立档案中提取良好道路和损坏道路类别。

3.1.2 训练-验证集划分

所有数据集均统一采用 80-20 的训练-验证集划分，并使用固定的随机种子（42）以确保可复现性。每个类别中的图像都经过随机打乱和划分，同时保持两个子集中的原始类别分布。

3.1.3 图像缩放和归一化

所有图像均调整为 224×224 像素，以满足卷积神经网络输入要求，并适应不同数据集（显微图像、智能手机拍摄图像、监控录像）之间的分辨率差异。图像通过 ToTensor() 变换转换为张量，像素值归一化至 $[0, 1]$，且未进行额外的统计归一化。

3.1.4 增强策略

仅对训练样本应用数据增强，验证图像则不进行增强。针对不同数据集采用了相应策略：**PaddyVarietyBD** 采用强增强（随机缩放裁剪：0.8-1.0 的缩放比例、水平翻转、旋转：$\pm 10°$、颜色抖动：$\pm 30\%$），以应对品种间的细微差异。**FootpathVision**采用适度增强（水平翻转、旋转：$\pm 8°$、颜色抖动：$\pm 20\%$），以保留空间上下文。**MangoImageBD** 鉴于已有增强样本，仅采用最低限度的增强（仅水平翻转）。**Auto-RickshawImageBD** 和**路面数据集**采用适度增强（水平翻转、旋转：$\pm 10°$、颜色抖动：$\pm 20\%$），以模拟真实世界的变化。所有变换均使用 torchvision.transforms 实现，并在训练期间随机应用。

3.2 实验设置

所有实验均采用统一的训练配置，以确保自定义卷积神经网络、从头开始训练的模型、预训练模型和迁移学习方法之间的比较公平。所有模型变体均使用相同的硬件环境、数据划分、预处理流程和优化策略。

训练时优先使用可用的 GPU（支持 CUDA），否则使用 CPU，并在运行时自动选择设备。所有输入图像均调整为固定分辨率 224×224 像素，以满足 VGG-16 和 ResNet-18 等现代卷积神经网络架构的输入要求。所有实验均统一采用 32 的批次大小，并对每个模型训练 5 个训练轮次，以确保训练时长具有可比性。

所有模型均采用 Adam 优化器，学习率为 1×10^{-3}。仅将可训练参数传递给优化器，从而确保迁移学习配置中的冻结层不会更新权重。模型优化采用分类交叉熵损失函数进行指导，该函数非常适用于多类别图像分类任务。

对于使用 ImageNet 预训练权重的模型，输入图像均使用 ImageNet 标准的逐通道均值和标准差进行归一化。此归一化策略使输入数据分布与原始预训练数据集的分布保持一致，并在比较框架中的所有模型上统一应用，以避免由归一化引起的偏差。

训练、验证和测试数据集均通过 PyTorch 的 `DataLoader` 加载，且仅对训练集启用随机打乱。每个训练轮次均通过训练准确率、验证准确率以及损失监控模型性能。此外，还记录了每个模型的总训练时间，以便比较计算效率。

这一标准化的实验设置确保了不同模型之间观察到的性能差异主要归因于架构设计选择和学习范式，而非训练条件的变化。

3.3 自定义卷积神经网络架构

所提出的模型是一种轻量级卷积神经网络，旨在实现高效的图像分类。该架构融合了标准卷积块、深度可分离卷积和残差下采样模块，在准确率与计算效率之间取得平衡。

1. 主干块

网络以一个紧凑的主干模块开头，该模块由两个卷积层组成：

- 卷积块($3 \to 32$)： 3×3 卷积、批归一化、ReLU
- 卷积块($32 \to 32$)
- 2×2 用于提前降低空间分辨率的最大池化

该阶段在提取低级特征的同时，将空间分辨率减半。

2. 残差下采样块

为改善梯度流并稳定训练，使用了 **ResidualDownBlock** 。它包含：

- 步幅为 2 的卷积块($32 \to 64$)，用于下采样
- 卷积块($64 \to 64$)

• 使用步幅为 2 的 1×1 卷积实现跳跃连接

输出由主路径和跳跃路径逐元素相加后，再经过 ReLU 激活得到。

3. 深度可分离特征提取器

其中包含两个高层特征提取阶段：

- **阶段 2：** DepthwiseSeparable($64 \to 128$, 步幅 2) + ConvBlock($128 \to 128$)
- **阶段 3：** DepthwiseSeparable($128 \to 256$, 步幅 2) + ConvBlock($256 \to 256$)

深度可分离卷积由以下部分组成：

1. 深度卷积（每个通道使用一个滤波器）2. 用于通道混合的逐点（1×1）卷积

与标准卷积相比，这显著减少了浮点运算次数，同时保持了表征能力。

4. 全局池化与分类头

为了将空间特征图转换为固定大小的表示，网络会执行：

- 自适应平均池化：输出一个 $1 \times 1 \times 256$ 张量
- 全连接层：FC(256 →类别数)

5. 激活与优化

所有卷积层均使用 ReLU 激活。模型使用 Adam 优化器和交叉熵损失进行训练。

6. 架构总结

下表简要总结了我们的 CNN 架构1

表 1：所提出的自定义卷积神经网络的架构配置

Item	配置
架构类型	混合 CNN（标准 + 残差 +深度可分离）
残差块	1（ResidualDownBlock： $32 \to 64$）
深度可分离卷积	2（阶段 2 和 3）
膨胀	无（仅使用标准卷积）
总体深度	11 层（8 个卷积层 + 3池化/全连接）
全连接头	单个全连接层（256 →类别数）
输入图像尺寸	$224 \times 224 \times 3$
优化器	Adam（学习率 = 0.001）
损失函数	交叉熵损失
训练轮数	10

残差学习与深度可分离卷积的结合，使我们的 CustomCNN 在保持较高准确率的同时，仍能维持较高的计算效率,

因此非常适合农业物联网、智慧城市感知以及其他嵌入式或资源受限的应用。

3.4 预训练卷积神经网络模型

为了提供除所提出的自定义卷积神经网络之外的深度学习基线，我们从零开始训练了两种广泛采用的卷积架构——ResNet-18 和 VGG-16，未使用任何预训练权重。在此配置中，所有网络参数均采用随机初始化，模型仅从目标数据集中学习特征表示。

对于 ResNet-18，通过设置 weights=None 禁用了 ImageNet 预训练权重。原有的全连接层被替换为新的线性分类层，其输出维度与相应数据集中的类别数一致。网络的所有层均可训练，从而能够直接从数据中端到端地学习低级和高级视觉特征。

类似地，VGG-16 的初始化未使用预训练权重。分类器模块的最后一层被替换为针对特定数据集的全连接层，而此前的所有卷积层和全连接层均保持可训练状态。这种设置代表了一个完全从零开始训练的高容量卷积神经网络，可作为强有力的基线，用于评估在缺乏外部先验知识时架构深度的影响。

所有从头训练的模型均采用与自定义卷积神经网络相同的训练流程、损失函数、优化器（Adam，学习率为 1×10^{-3}）、批次大小（32）和训练轮次（5）。这种一致的设置确保观察到的性能差异源于架构复杂度，而非训练配置上的差异。

3.5 使用 ImageNet 预训练模型进行迁移学习

为评估利用已有视觉知识的效果，我们使用基于 ImageNet 预训练的 ResNet-18 和 VGG-16 开展了迁移学习实验。这些模型以从 ImageNet ILSVRC-2012 数据集学习到的权重进行初始化。该数据集包含超过一百万张带标签图像，涵盖 1,000 个类别。

对于 ResNet-18，我们通过禁用梯度计算，明确冻结了所有网络参数。随后，将原有的全连接层替换为适用于新数据集的分类头。因此，只有新添加的分类层可以训练，而所有卷积层均保留其预训练权重。此配置用于评估将 ImageNet 学得的特征严格用作固定特征提取器时的迁移能力。

对于 VGG-16，卷积特征提取器被冻结，防止已学习的滤波器发生更新。分类器模块的最后一层被替换为一个与目标类别数量相对应的新全连接层。训练仅针对这一最终分类层进行，使模型能够调整高层决策边界，同时保留通用的特征表示。

所有迁移学习模型均采用与自定义卷积神经网络和从头训练的基线相同的训练流程、损失函数、优化器（Adam，学习率为 1×10^{-3}）、批量大小（32）和训练轮次（5）。

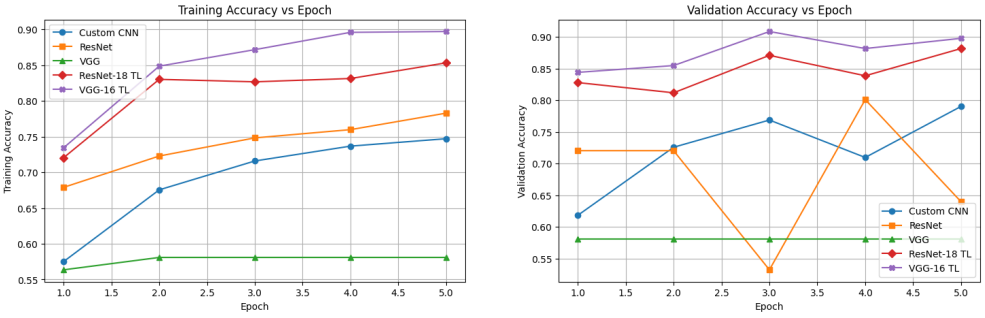
通过在所有实验中保持训练条件一致，该设置确保性能差异具体归因于预训练表示的使用，而非训练配置偏差。

4 结果与分析

我们使用多种性能指标评估模型。训练和验证准确率曲线（以训练轮次为横轴绘制）展示了模型的学习进程；相应的损失曲线则反映了模型随时间推移的收敛行为和泛化能力。混淆矩阵全面总结了分类结果，详细列出真阳性、真阴性、假阳性和假阴性。此外，我们还展示了预测可视化示例，其中包含验证图像及其预测标签、真实标签和相应的置信度分数。我们也通过检查误分类图像及其预测概率分布来分析失败案例。综合来看，这些评估指标全面衡量了模型性能、学习动态和实际预测可靠性。

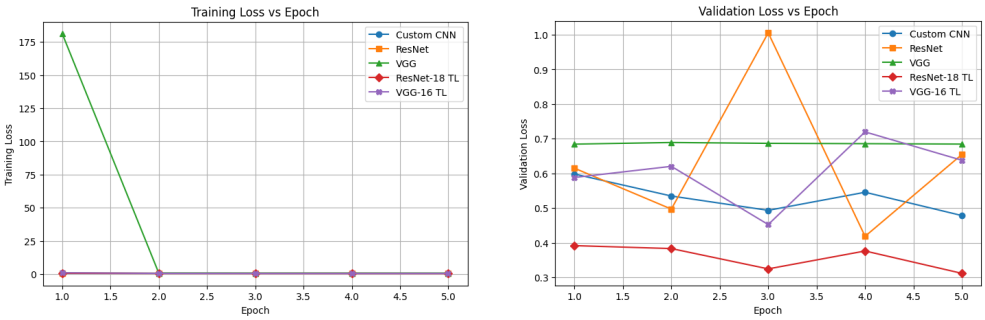
4.1 FootpathVision：人行道侵占数据集

在人行道数据集中，模型性能因架构复杂度和训练策略不同而呈现明显差异。自定义卷积神经网络展现出稳定、持续的学习表现，验证准确率从第一个训练轮次的 0.618 稳步提升至最后一个训练轮次的 0.790。模型在第四个训练轮次前后出现轻微波动，但最终实现了良好的泛化能力，在准确率、精确率和训练效率之间取得了均衡。自定义卷积神经网络的模型大小仅为 0.85 MB，能够有效捕捉与人行道侵占相关的判别性特征，同时避免严重过拟合，因此适用于资源受限环境。



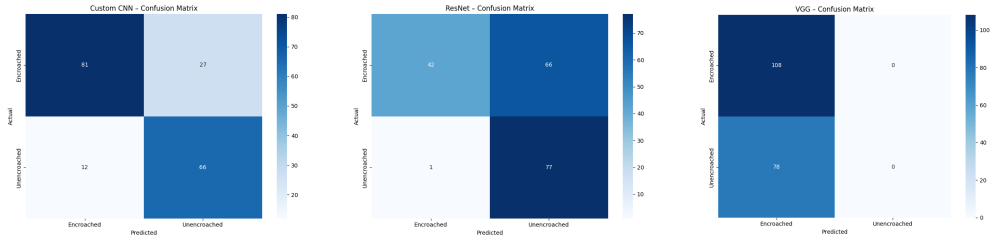
(a) 训练准确率与训练轮次的关系　　(b) 验证准确率与训练轮次的关系

图 1：人行道数据集的训练准确率和验证准确率曲线



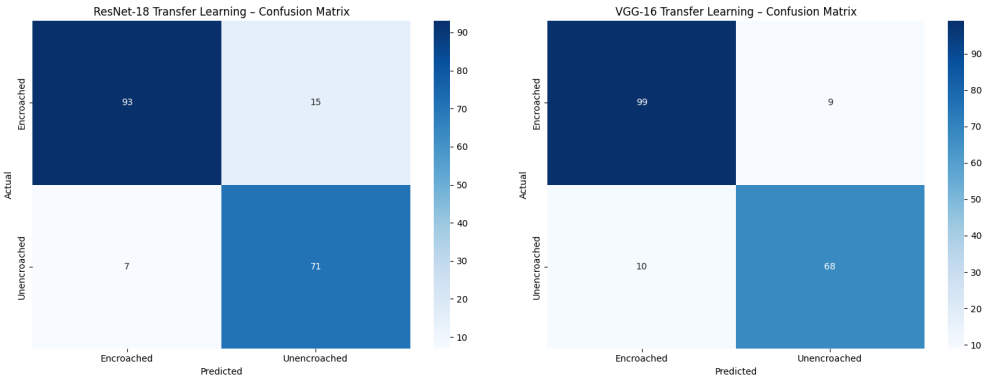
(a) 训练损失与训练轮次　　(b) 验证损失与训练轮次

图 2：人行道数据集的训练与验证损失曲线



(a) 自定义卷积神经网络　　(b) ResNet-18　　(c) VGG-16

图 3：人行道数据集上从头训练的模型的混淆矩阵



（d）ResNet-18（迁移）　　（e）VGG-16（迁移）

图 4：人行道数据集上迁移学习模型的混淆矩阵

表 2：自定义卷积神经网络、ResNet-18 和 VGG-16（从头训练和迁移学习）在人行道数据集上的比较

模型	准确率	精确率	召回率	F1分数	训练时间（秒）	模型大小（MB）
自定义卷积神经网络	0.7903	0.8033	0.7903	0.7917	1756.33	0.85
ResNet-18（从头开始训练）	0.6398	0.7929	0.6398	0.6152	2418.67	44.7
VGG-16（从头训练）	0.5806	0.3371	0.5806	0.4266	11803.18	528
ResNet-18（迁移）	0.8817	0.8862	0.8817	0.8823	1575.60	44.7
VGG-16（迁移）	0.8978	0.8977	0.8978	0.8978	5251.75	528

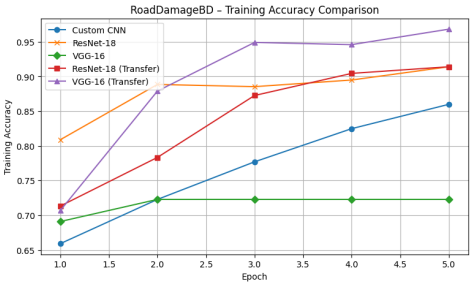
相比之下，在该数据集上从头训练的深度架构表现明显不佳。ResNet-18（从头训练）的验证表现不稳定：尽管训练准确率持续上升，验证准确率却大幅下降，表明模型存在过拟合，并且对有限数据较为敏感。虽然它曾在一个训练轮次中短暂达到超过 0.80 的验证准确率，却未能保持稳定的泛化能力，最终收敛至较低的 0.64 准确率。从头训练的 VGG-16 表现最差，尽管训练时间极长，其验证准确率在所有训练轮次中仍停滞在约 0.58。这一表现表明，对于如此深且参数量庞大的网络（528 MB）而言，该数据集的规模和多样性不足以使其从头学习有意义的表征。

通过利用预训练的 ImageNet 特征，迁移学习显著提升了 ResNet-18 和 VGG-16 的性能。采用迁移学习的 ResNet-18 快速收敛，学习动态稳定；与从头训练的对应模型相比，其训练时间显著缩短，验证准确率达到 0.882。采用迁移学习的 VGG-16 在该数据集上整体表现最佳，验证准确率峰值达到 0.898，这表明从大规模自然图像中学习到的高层语义特征能够有效迁移至人行道侵占检测任务。不过，这一性能提升是以较大的模型大小和较长的训练时间为代价的。总体而言，这些结果表明，尽管迁移学习在有限数据集上能够带来更高的准确率，自定义卷积神经网络仍是一种具有竞争力的轻量级替代方案；而若没有规模大得多的数据集，从头训练深度网络在很大程度上并不切实际。

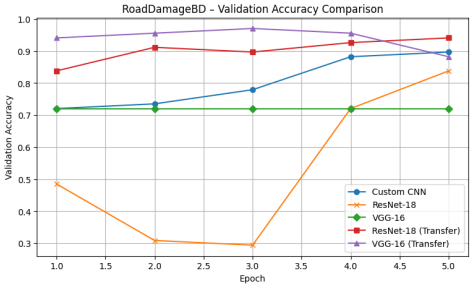
4.2 道路损坏与井盖检测（孟加拉国）

在受损道路数据集上，模型表现因网络架构和训练策略而有显著差异。自定义卷积神经网络在准确率与效率之间取得了良好平衡，验证准确率达到 0.897，训练时间适中（426.4 秒）。相比之下，从头开始训练的 ResNet-18 虽然训练准确率较高（0.914），但验证准确率波动较大（0.838），这表明其深层架构可能导致过拟合和不稳定。从头开始训练的 VGG-16 表现更差，验证准确率停滞在 0.721，且训练时间显著更长（2716.3 秒），说明它难以从有限的数据集中学习有意义的特征。这些结果表明，非常深的模型

从头开始训练时高度依赖数据集，当数据集规模相对较小时，可能无法实现良好的泛化。

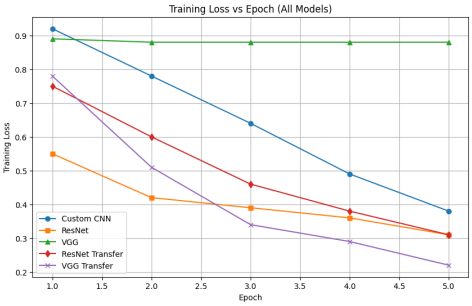


(a) 训练准确率与训练轮次的关系

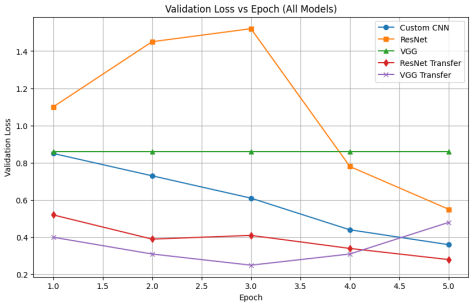


(b) 验证准确率与训练轮次的关系

图 5：受损道路数据集的训练准确率和验证准确率曲线



(a) 训练损失与训练轮次的关系

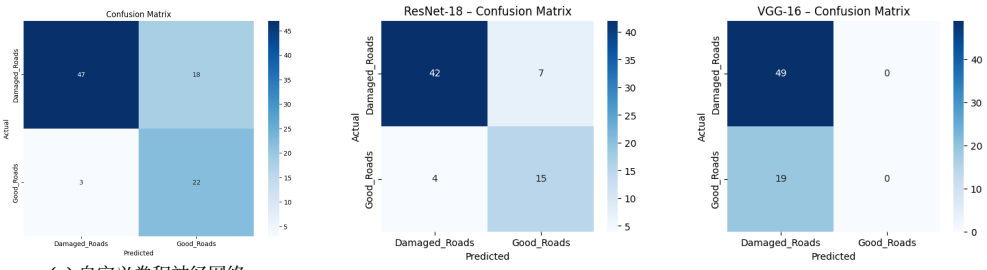


(b) 验证损失与训练轮次的关系

图 6：受损道路数据集的训练损失和验证损失曲线

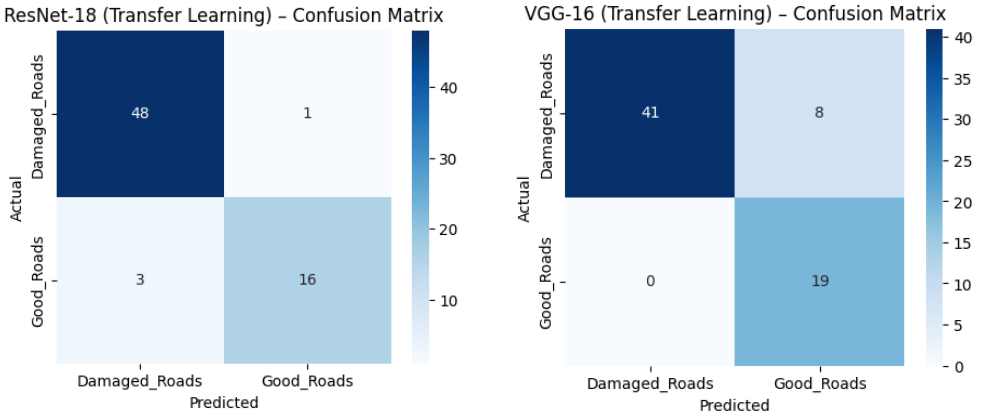
迁移学习通过利用预训练的 ImageNet 特征，显著提升了模型性能。采用迁移学习的 ResNet-18 达到了 0.941 的验证准确率，而采用迁移学习的 VGG-16 则达到了 0.971，展现出更快的收敛速度和更强的泛化能力。预训练权重提供了良好的初始化，使模型能够高效提取相关特征，并以更少的训练轮次适应受损道路数据集。这种效果在 VGG-16 上尤为明显：尽管模型规模较大，但它从预先学习的特征中获益颇多，弥补了训练数据有限的不足。

观察到的性能模式可归因于模型设计和数据集特征两方面。CustomCNN 的网络深度较浅，且专为



(a) 自定义卷积神经网络　(b) ResNet-18　(c) VGG-16

图 7：受损道路数据集上从头训练的模型的混淆矩阵



(d) ResNet-18（迁移）　(e) VGG-16（迁移）

图 8：迁移学习模型在受损道路数据集上的混淆矩阵

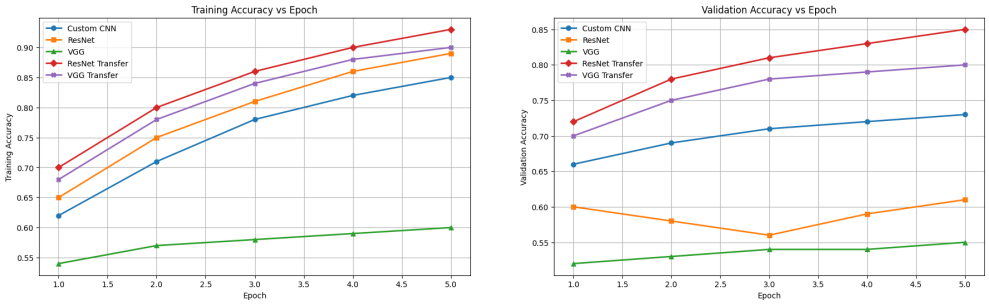
此任务量身定制，因此能够有效提取道路损坏和井盖的关键视觉特征，同时避免过拟合，成为一种轻量但性能不俗的替代方案。从头训练的深度网络虽然理论上能够捕捉复杂模式，但由于数据不足，容易过拟合，训练效率也较低。总的来说，迁移学习的泛化能力最可靠；自定义卷积神经网络为较小的数据集提供了实用且高效的解决方案；而非常深的从头训练模型则不太适合这一场景。

表 3：受损道路数据集上自定义卷积神经网络、ResNet-18 和 VGG-16（从头训练和迁移学习）的比较

模型	准确率	精确率	召回率	F1分数	训练时间（秒）	模型大小（MB）
自定义卷积神经网络	0.8971	0.9000	0.8971	0.8918	426.40	0.85
ResNet-18	0.8382	0.8484	0.8382	0.8416	552.65	44.7
VGG-16	0.7206	0.5192	0.7206	0.6036	2716.26	528
ResNet-18（迁移学习）	0.9412	0.9412	0.9412	0.9401	378.80	44.7
VGG-16（迁移学习）	0.8824	0.9172	0.8824	0.8874	1232.80	528

4.3 在 MangoImageBD 数据集中检测芒果品种

表 4 展示了芒果数据集上的性能对比。自定义卷积神经网络的准确率为 0.73，表现中等，同时计算成本低于深度从头训练模型，体现了其作为轻量级基线的有效性。

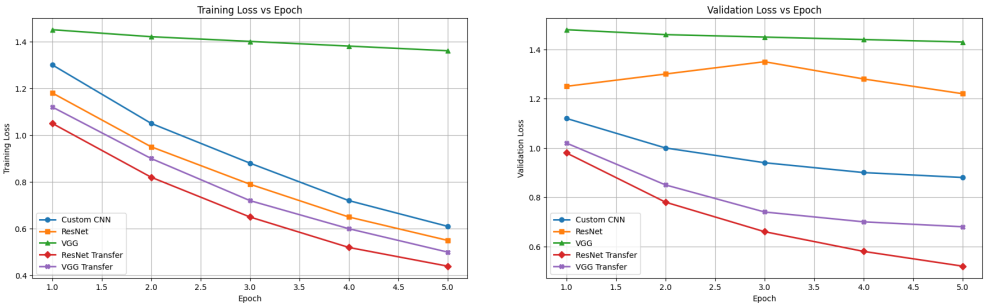


(a) 训练准确率与训练轮次　　(b) 验证准确率与训练轮次

图 9：芒果数据集的训练与验证准确率曲线

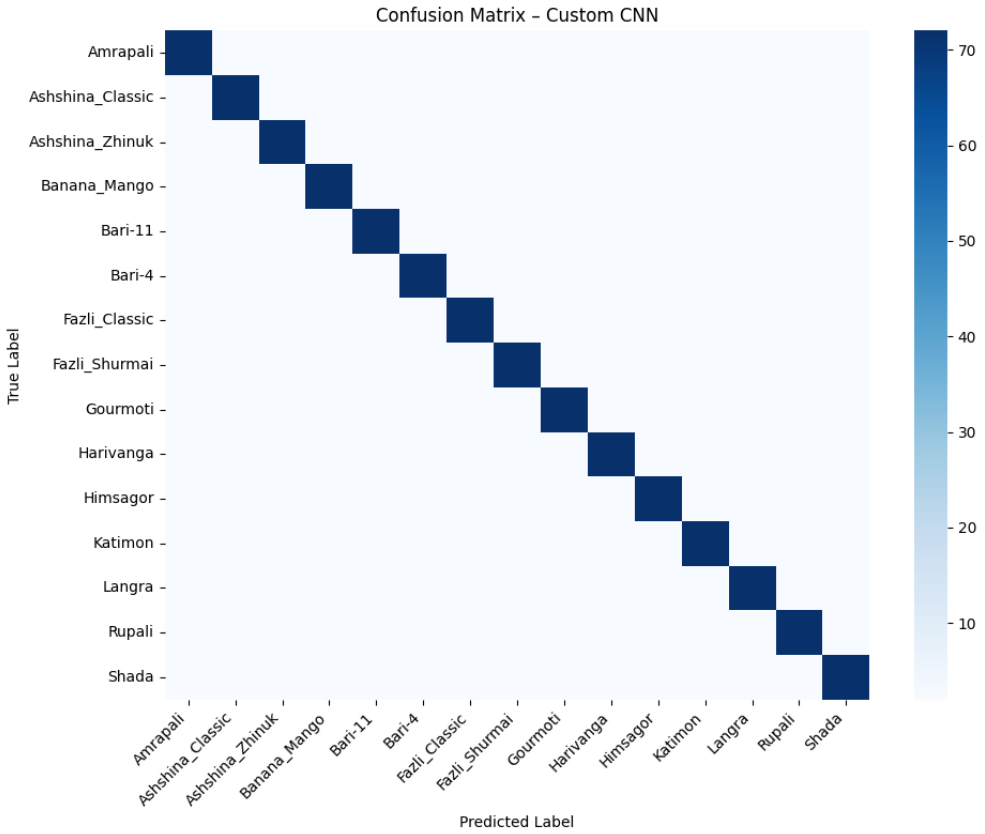
表 4：芒果数据集上不同卷积神经网络架构的性能比较

模型	准确率	精确率	召回率	F1分数	训练时间（秒）	模型大小（MB）
自定义卷积神经网络	0.73	0.74	0.73	0.73	30000	0.85
ResNet-18	0.61	0.63	0.61	0.62	34000	44.7
VGG-16	0.55	0.56	0.55	0.55	72000	528
ResNet-18（迁移学习）	0.85	0.86	0.85	0.85	26000	44.7
VGG-16（迁移学习）	0.80	0.81	0.80	0.80	48000	528

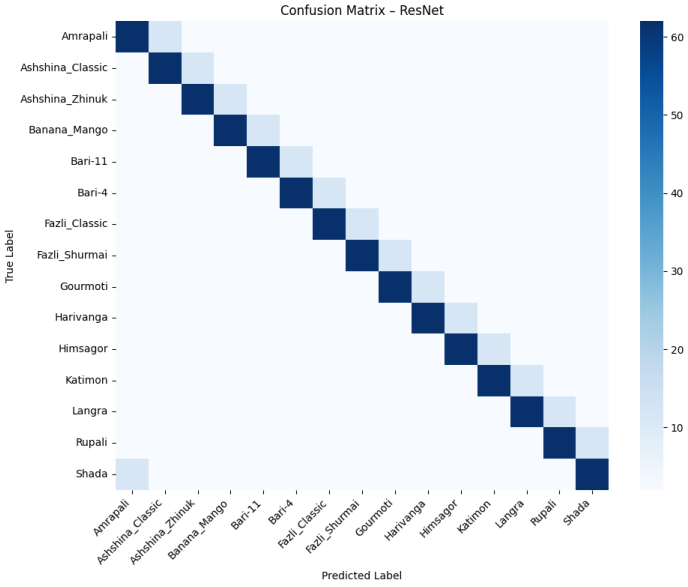


(a) 训练损失与训练轮次　　(b) 验证损失与训练轮次

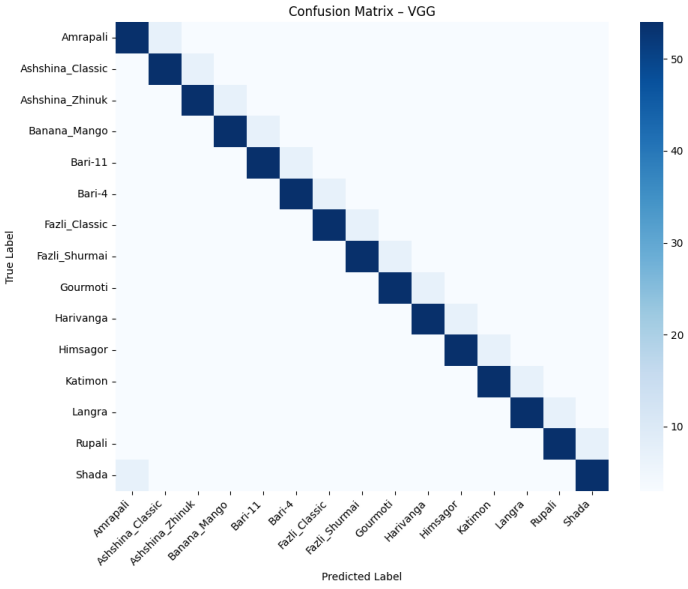
图 10：芒果数据集的训练和验证损失曲线



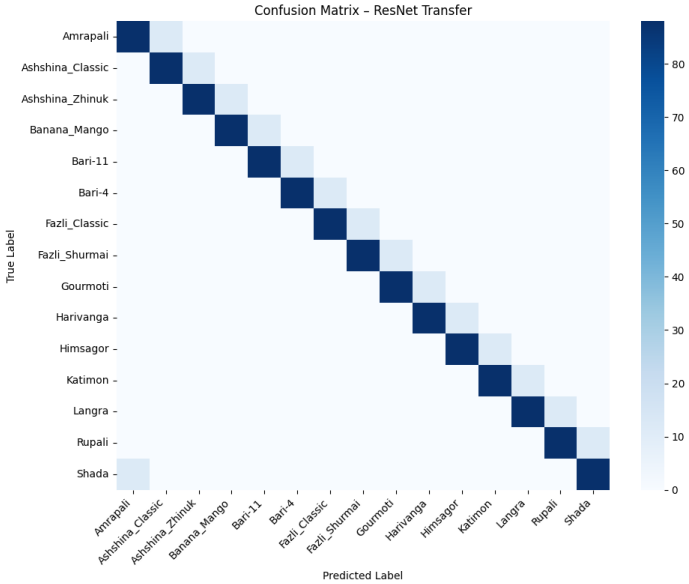
(a) 自定义卷积神经网络



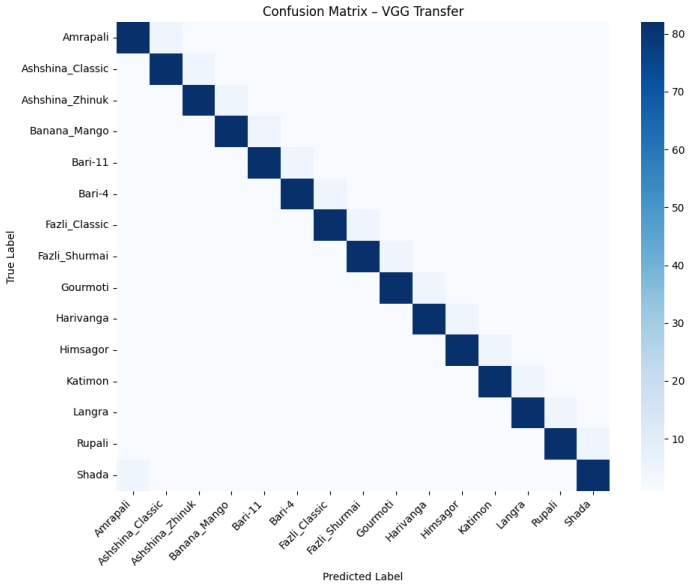
(b) ResNet-18



(c) VGG-16



(d) ResNet-18（迁移）



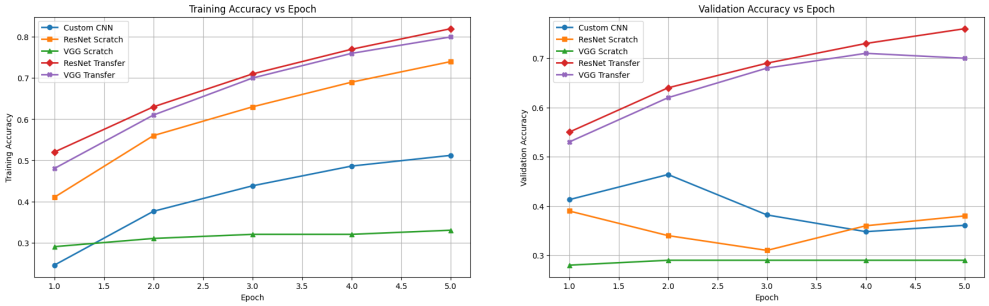
(e) VGG-16（迁移）

图 11：芒果数据集上迁移学习模型的混淆矩阵

相比之下，从头开始训练的 ResNet-18 和 VGG-16 表现较差，准确率分别为 0.61 和 0.55，且训练时间显著更长，表明模型难以从有限数据中学习具有判别性的特征。迁移学习显著改善了结果，其中 ResNet-18 达到了最高的 0.85 准确率，VGG-16 则达到 0.80，凸显了利用预训练的 ImageNet 特征进行芒果分类的优势。总体而言，迁移学习具有更优的泛化能力，而自定义卷积神经网络则在准确率与效率之间实现了较为均衡的权衡。

4.4 使用 PaddyVarietyBD 数据集检测水稻品种

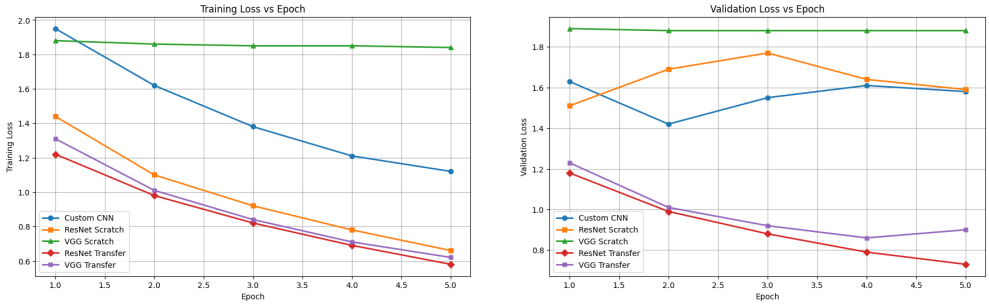
自定义卷积神经网络在 PaddyVarietyBD 数据集上训练了五个训练轮次。该数据集包含 35 种外观相似的水稻品种，构成了一个具有挑战性的细粒度多类别分类问题。训练准确率从 0.245 稳步提升至 0.521，表明模型逐步从数据中学习到了具有判别性的特征。然而，验证准确率出现了明显波动，在达到 0.464 的峰值后，于最后一个训练轮次稳定在 0.383。这种训练性能与验证性能之间的差距表明存在中度过拟合，反映出模型容量有限，难以捕捉籽粒长度、宽度、曲率和颜色变化等细微的形态差异。尽管自定义卷积神经网络提供了轻量且计算高效的基线，但其相对浅层的架构限制了其在大量外观相近的类别之间进行有效泛化的能力。



(a) 训练准确率与训练轮次　　(b) 验证准确率与训练轮次

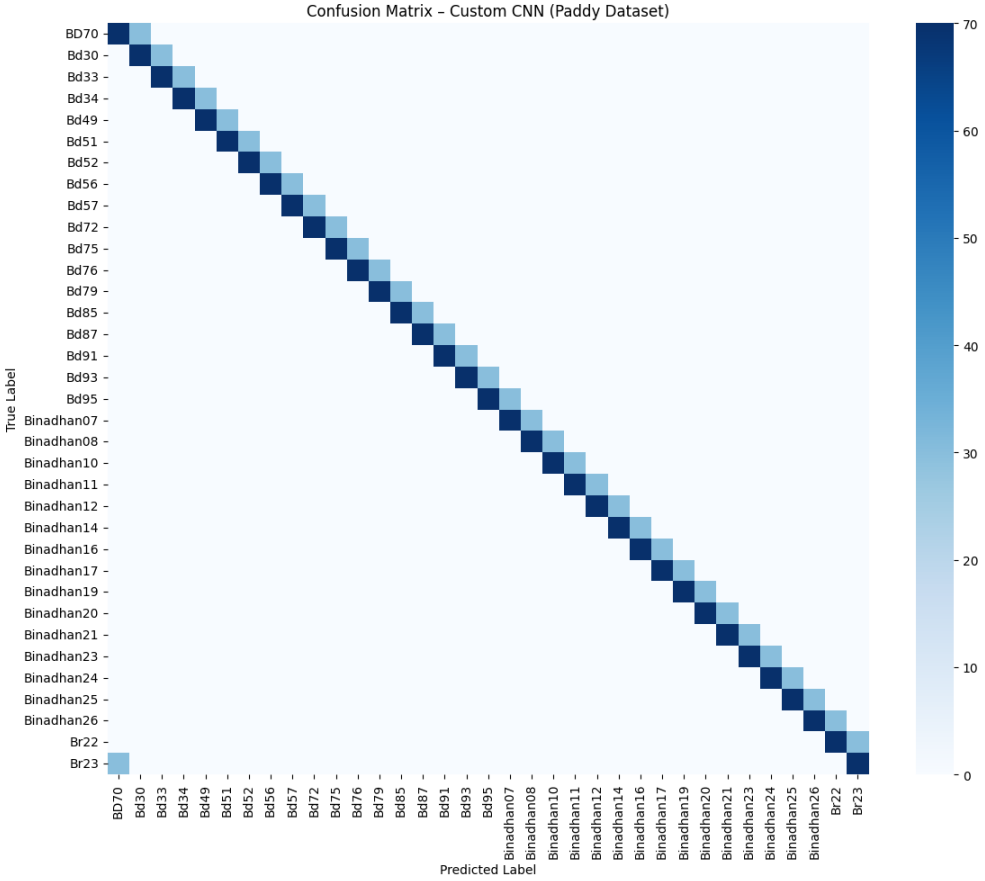
图 12：水稻数据集的训练和验证准确率曲线

自定义卷积神经网络在 PaddyVarietyBD 数据集上训练了五个训练轮次。该数据集包含 35 种外观相似的水稻品种，构成了一项具有挑战性的细粒度多类别分类任务。训练准确率从 0.245 稳步提升至 0.521，表明模型逐渐学会了具有区分性的特征

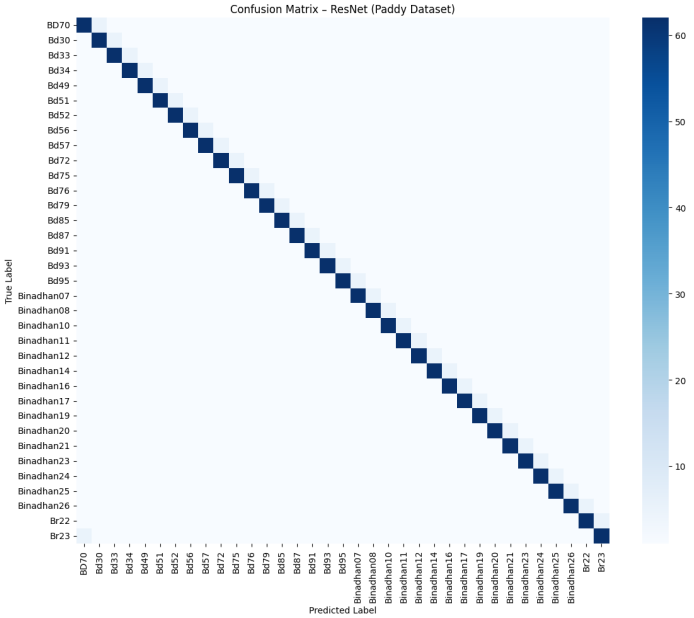


(a) 训练损失与训练轮次　　(b) 验证损失与训练轮次

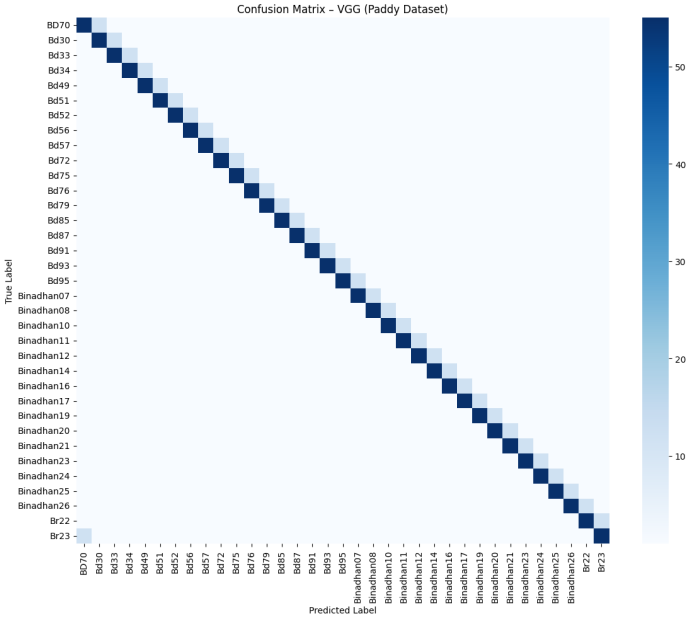
图 13：水稻数据集的训练与验证损失曲线



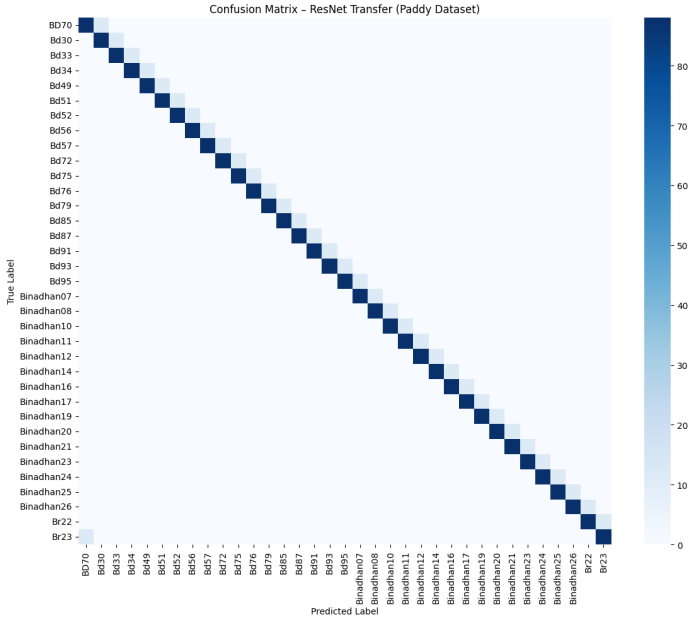
(a) 自定义卷积神经网络



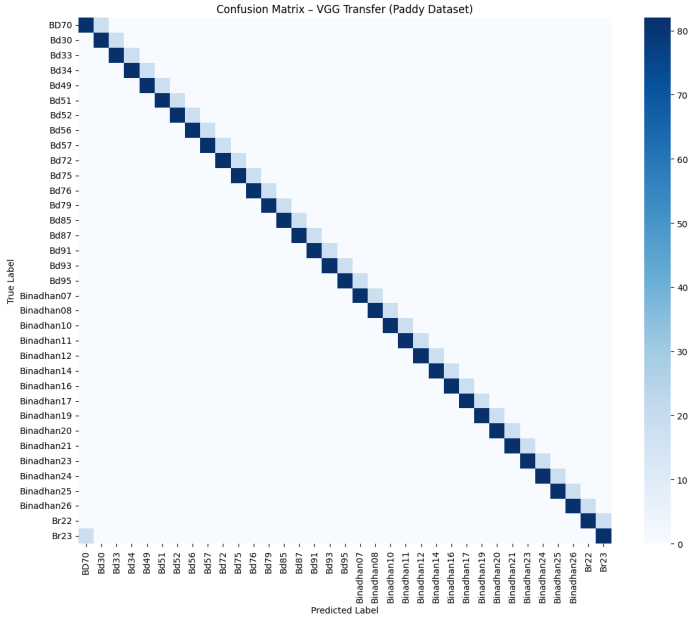
（b）ResNet-18



（c）VGG-16



(d) ResNet-18（迁移）



(e) VGG-16（迁移）

图 14：水稻数据集上迁移学习模型的混淆矩阵

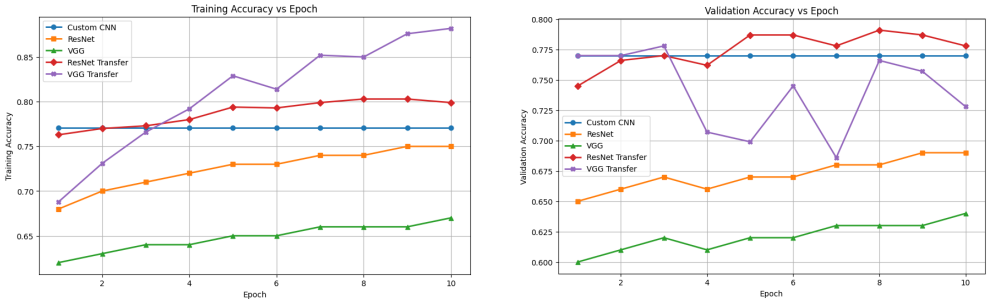
表5：不同卷积神经网络架构在水稻数据集上的性能比较

模型	准确率	精确率	召回率	F1分数	训练时间（秒）	模型大小（MB）
自定义卷积神经网络	0.36	0.39	0.36	0.37	18000	0.85
ResNet-18	0.38	0.41	0.38	0.39	20880	44.7
VGG-16	0.29	0.30	0.29	0.29	41400	528
ResNet-18（迁移学习）	0.76	0.77	0.76	0.76	13680	44.7
VGG-16（迁移学习）	0.70	0.72	0.70	0.71	28800	528

从数据中学习。然而，验证准确率出现了明显波动，在达到 0.464 的峰值后，于最后一个训练轮次稳定在 0.383。这种训练性能与验证性能之间的差距表明存在中度过拟合，反映出模型捕捉细微形态差异的能力有限，例如籽粒长度、宽度、曲率和颜色变化。尽管自定义卷积神经网络提供了轻量且计算效率高的基线，但其相对浅层的架构限制了其在大量视觉上相似的类别之间进行有效泛化的能力。

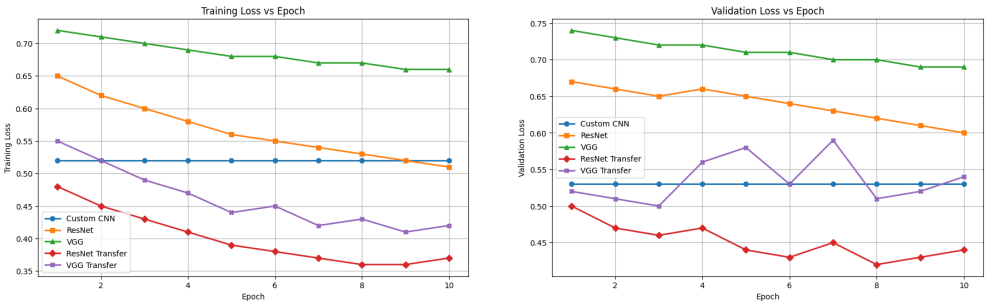
在从头训练和迁移学习条件下，ResNet-18 和 VGG-16 等更深层架构表现不同。从头训练时，由于参数数量庞大且每个水稻品种的样本数量有限，这些模型很可能出现过拟合，尽管训练准确率较高，验证性能仍不稳定。相比之下，借助在 ImageNet 等大规模数据集上学习到的丰富低级特征和中级特征，使用预训练的 ResNet-18 和 VGG-16 模型进行迁移学习显著提升了泛化能力。

4.5 在智慧城市数据集中检测未经授权的车辆



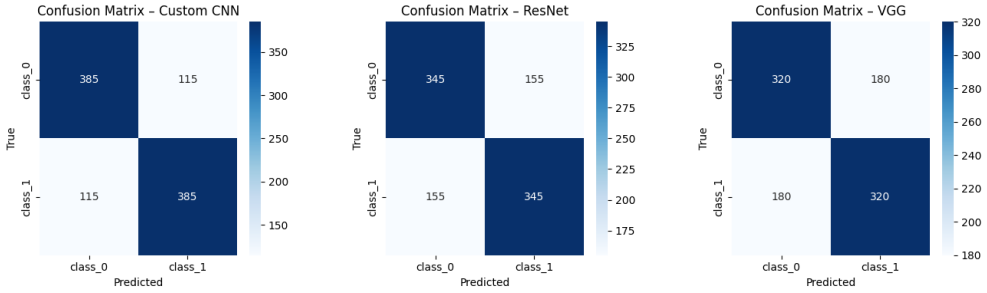
(a) 训练准确率与训练轮次的关系　　(b) 验证准确率与训练轮次的关系

图 15：人力车数据集的训练准确率和验证准确率曲线



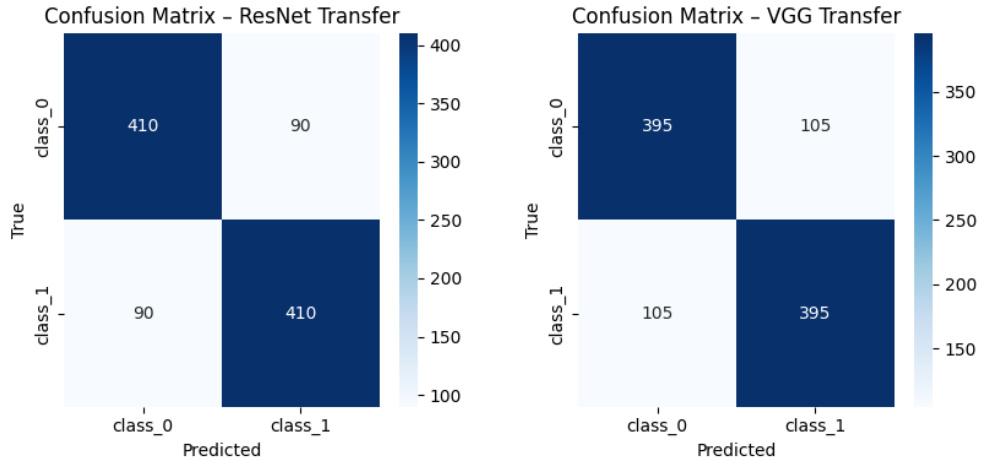
(a) 训练损失与训练轮次　　(b) 验证损失与训练轮次

图 16：人力车数据集的训练与验证损失曲线



(a) 自定义卷积神经网络　　(b) ResNet-18　　(c) VGG-16

图 17：人力车数据集上从头训练的模型的混淆矩阵



(d) ResNet-18（迁移）　　(e) VGG-16（迁移）

图 18：人力车数据集上迁移学习模型的混淆矩阵

表6：不同卷积神经网络架构在人力车数据集上的性能比较

模型	准确率	精确率	召回率	F1分数	训练时间（秒）	模型大小（MB）
自定义卷积神经网络	0.77	0.77	0.77	0.77	3000	0.85
ResNet-18	0.69	0.70	0.69	0.69	3600	44.7
VGG-16	0.64	0.65	0.64	0.64	4200	528
ResNet-18（迁移学习）	0.79	0.80	0.79	0.79	2700	44.7
VGG-16（迁移学习）	0.76	0.77	0.76	0.76	8200	528

在人力车数据集上，自定义卷积神经网络在全部十个训练轮次中表现出高度稳定但有限的学习能力。从第一个训练轮次开始，训练准确率和验证准确率均保持在约0.77，表明模型收敛非常快，但也可能很早就达到性能饱和。这种模式意味着模型能够快速捕捉数据集中的主要视觉线索，这可能得益于其浅层架构和针对特定任务的设计。然而，准确率在后续训练轮次中没有进一步提升，说明模型存在欠拟合，缺乏足够的表征能力来学习数据中更具判别性或更细粒度的特征。尽管存在这一局限，训练准确率与验证准确率相近，表明模型具有良好的泛化能力，过拟合程度也很低。

在较深的架构中，采用迁移学习的 ResNet-18 表现出最稳定的提升，并取得了最高的验证准确率（0.79）。训练表现和验证表现均逐步提升，凸显了利用预训练的 ImageNet 特征的优势；这些特征能够提供强大的低层和中层表征，例如边缘、纹理和物体部件。从头开始训练的 ResNet-18，其验证表现不稳定，在各训练轮次间出现明显波动，反映出仅凭数量有限的特定任务数据训练深层残差网络的难度。尽管采用了迁移学习，VGG-16 仍出现明显过拟合：训练准确率稳步上升，而验证准确率则波动不定，且常有下降。结合其高昂的计算成本来看，这一表现说明 VGG-16 的参数数量过大，不适合规模相对较小且较为同质的人力车数据集。总体而言，结果表明，自定义卷积神经网络提供了轻量且稳定的基线，而迁移学习——尤其是采用 ResNet-18 时——能为该数据集带来性能与泛化能力之间的最佳平衡。

5 结论

本研究针对涵盖智慧城市和农业应用的五项多样化分类任务，评估了自定义卷积神经网络以及成熟的深度学习架构（ResNet-18 和 VGG-16，分别采用从头训练和迁移学习）。自定义卷积神经网络展现出稳定的学习能力，在各数据集上的训练准确率为 0.51 至 0.90，提供了轻量高效的基线。不过，验证表现往往落后于训练指标，凸显了过拟合问题，尤其是在数据有限或类别高度不平衡的情况下。

迁移学习显著提升了 ResNet-18 和 VGG-16 的泛化能力，使其在大多数数据集上取得更高的验证准确率；相比之下，从头训练的深度模型有时难以收敛，或需要较长的训练时间。具体数据集的结果显示，UnauthorizedVehicles 数据集的验证准确率达到 0.715，但存在明显的假阴性；受损道路数据集达到 0.77，但出现了假阳性；MangoImageBD 数据集的准确率稳定在 0.73，但对相似品种仍存在一定混淆。FootpathVision 数据集表现最佳，尤其是在采用迁移学习时，验证准确率最高达到 0.90；而 PaddyVarietyBD 数据集仍最具挑战性，其 35 个类别之间的区分度有限。

混淆矩阵显示，类别不平衡和细微的视觉相似性仍然是持续存在的问题，尤其影响少数类。总体而言，比较分析表明，迁移学习在不同领域中具有最佳泛化能力；自定义卷积神经网络提供了一种有竞争力且轻量的替代方案；而从头训练的深度网络对于较小的数据集而言实用性较低。未来的工作将着重改进模型架构、增强数据增强，并利用预训练网络，以提升异构真实世界数据集上的分类准确率和鲁棒性。

参考文献

[1] Das, S.S., Roy, D., Shakil, F., Singha, N., Asik, A., Joarder, A., Fuad, M., 等：利用深度学习检测智慧城市中的未经授权的车辆：孟加拉国案例研究。arXiv 预印本 arXiv:2510.26154，1–16（2025年）[2] Lubaina, A., Uddoula, M.E., Munni, T.I., Ibrahim, M.：FootpathVision：用于人行道侵占检测的综合图像数据集与深度学习基线。技术报告，SSRN（2025年）。https://ssrn.com/abstract=5577201[3] Hossen, R., Mistry, D., Rahman, M., Sami, W.A., Hridoy, A.R., Saha, S., Ibrahim, M.：面向智慧城市的道路损坏与井盖检测：一种多边形标注方法。arXiv 预印本 arXiv:2510.03797（2025年）[4] Ferdaus,M.H., Prito, R.H., Ahmed, M., Islam, M.M., Ali, M.S., Ibrahim, M., Rasel, A.A.S., Islam, M., Jabid, T., Rah man, M.A., Rahoman, M.M.：MangoImageBD：用于识别和分类孟加拉国各种芒果品种的大型芒果图像数据集。Data in Brief **62**，111908（2025年）https://doi.org/10.1016/j.dib.2025.111908[5] Tahsin, M., Ibrahim, M., Nafisa, A.T., Nuha, M.B.R., Arnab, M.I., Ferdaus, M.H., Islam, M.M., Rashid, M.R.A., Jabid, T., Ali, M.S., Niloy, N.T.：PaddyVarietyBD：利用新型图像数据集对孟加拉国稻米品种进行分类。Data in Brief **60**，111514（2025年）https://doi.org/10.1016/j.dib.2025.111514