doi:10.6041/j.issn.1000-1298.2023.S1.018

基于 MobileViT 模型的小麦收获机喂入密度分类方法

杨志凯[1]　扶兰兰[1]　唐　灿[1]　王发明[1]　倪昕东[1]　陈　度[1,2]
(1.中国农业大学工学院, 北京 100083; 2.智能农业动力装备全国重点实验室, 北京 100083)

摘要: 基于喂入量的作业速度智能化控制技术是优化联合收获机作业效率和质量的重要手段。本文针对传统喂入量自动控制技术时滞明显,在喂入量调整时无法及时适应实际情况的问题,采用基于图像的深度学习方法开展了成熟期小麦植株密度等级分类识别方法研究,通过预先感知作物密度,实现联合收获机作业参数的自动调整。首先基于车载相机和无人机图像构建了小麦植株图像数据集,并细分为低密度、中密度、高密度和特高密度 4 类;其次构建了基于 MobileViT-XS 轻量化网络的密度等级识别模型,利用建立的数据集进行模型的训练和测试;最后将其与 VGG16、GoogLeNet 和 ResNet 进行了比较。结果表明,MobileViT-XS 模型的总体识别准确率达到 91.03%,且单幅图像推理时间仅为 29.5 ms。与 VGG16、ResNet 网络相比,总体识别准确率分别高出 3.51、2.34 个百分点,MobileViT-XS 模型可以较好的完成小麦不同密度等级的分类识别任务,为实时预测小麦喂入密度提供了技术支持。

关键词: 联合收获机; 喂入量; MobileViT; 图像分类; 密度识别
中图分类号: TP391　　**文献标识码**: A　　**文章编号**: 1000-1298(2023)S1-0172-09

Classification Method for Wheat Harvester Feeding Density Based on MobileViT Model

YANG Zhikai[1]　FU Lanlan[1]　TANG Can[1]　WANG Faming[1]　NI Xindong[1]　CHEN Du[1,2]
(1. *College of Engineering*, *China Agricultural University*, *Beijing* 100083, *China*
2. *State Key Laboratory of Intelligent Agricultural Power Equipment*, *Beijing* 100083, *China*)

Abstract: Intelligent control technology of operating speed based on feeding rate is an important means to optimize the efficiency and quality of combine harvester operations. Aiming at the obvious time delay of the traditional feeding rate automatic control technology and the inability to adapt to the actual situation in time when the feeding rate is adjusted. By analyzing the influencing factors of the feeding rate, an image-based deep learning method was used to carry out a research on the classification and recognition method of wheat plant density in the mature stage. By sensing crop density in advance, the operating parameters of the combine harvester can be automatically adjusted. Firstly, a multi-variety and multi-region mature stage wheat plant image dataset was constructed based on vehicle-mounted cameras and UAV images, and classified it into four categories: low density, medium density, high density, and very high density. Next, a density classification recognition model was built based on the lightweight MobileViT - XS network, and trained and tested the model by using the established dataset. Finally, it was compared with VGG16, GoogLeNet, and ResNet. The results showed that the overall recognition accuracy of the MobileViT - XS model reached 91.03%, and the inference time for a single image was 29.5 ms. Compared with VGG16 and ResNet networks, the overall recognition accuracy was 3.51 percentage points and 2.34 percentage points higher respectively. The MobileViT - XS model can effectively accomplish the classification recognition of wheat at different density levels, providing technical support for real-time prediction of wheat feeding rate.

Key words: combine harvester; feeding rate; MobileViT; image classification; density recognition

收稿日期: 2023-05-31　修回日期: 2023-09-01
基金项目: 国家自然科学基金项目(32201687)和国家重点研发计划项目(2022YFD2001203)
作者简介: 杨志凯(1997—),男,硕士生,主要从事农机装备智能测控技术研究,E-mail: 1776456982@qq.com
通信作者: 陈度(1981—),男,教授级高级工程师,主要从事农业自动化及测控技术研究,E-mail: tchendu@cau.edu.cn

0　引言

喂入量是指单位时间内联合收获机加工的作物量，不稳定的喂入量会给联合收获机的工作部件带来变化的载荷和冲击力，导致整机在工作时产生随机振动，进而影响整机的工作稳定性和可靠性[1]。另外，联合收获机的喂入量也会对收获损失率产生影响[2]，稳定的喂入量对联合收获机的性能和效率至关重要，将联合收获机喂入量控制在适宜范围内，可以确保高效率高质量的收获作业。

联合收获机的喂入量与作业割幅、作业速度、作物密度等因素密切相关，由于田间环境的复杂性和不可控性，通过调节联合收获机作业速度[3]及割台高度[4]，以达到合适的喂入量。喂入量自动控制旨在通过合理的控制算法和系统，实现联合收获机喂入量的自动调控[5-6]。传统的喂入量控制方法，多采用接触式传感器实时检测喂入量，根据先前的行为特征来进行控制，因此存在一定的时间延迟，难以实现真正的实时控制[7]。

随着图像技术的不断发展和进步，基于特征工程的机器视觉方法以及基于神经网络的深度学习方法广泛应用于农作物分类识别以及农作物密度检测等任务。在作物分类识别方面，孟志超等[8]提出了 Enhanced VGG16 网络模型，成功解决了以往训练模型权重过大和准确率不高的问题，在油茶幼苗品系上进行测试，该模型达到 98.44% 的准确率。LUO 等[9]首次建立了一套完整的杂草种子图像采集系统，并在此基础上对单株杂草种子图像进行了快速分割，然后对比了 6 种深度卷积神经网络模型。结果显示，GoogLeNet 在分类准确性方面表现最佳。MODI 等[10]在甘蔗田间采集了倒伏的甘蔗图像，并通过数据增强技术将数据集扩大 5 倍。结果表明，ResNet50 模型在该任务上取得了最优效果，其最高精度达到98.5%。在作物密度检测方面，潘静等[11]通过光谱分析水稻冠层密度图像，提取出水稻 2R + G 颜色特征参数值作为喂入密度特征，并在 6 个水稻品种上建立相关模型，探索了图像检测在喂入量预测方面的应用；刘海娟等[12]采用图像分割技术将水稻图像分成各个区域，并提取出与喂入量关系最为密切的颜色特征量，建立函数关系以实现水稻密度的预测；刘仲鹏等[13]从机器视觉角度，利用改进的阈值分割方法获取成熟期小麦的冠层图像，分析实际密度与图像区域像素数之间的函数关系，并建立喂入量随图像像素变化的模型以进行实时调整；SÉBASTIEN 等[14]利用 YOLO v5 模型上的无监督学习方法以及 DeepMAC 分割方法来开发小麦穗计数和麦穗分割系统，并使用第 2 个 RGB 相机来将图像中的麦穗数量转换为麦穗密度。

以上文献在作物密度检测方面取得了显著的成果，但是这些研究针对特定拍摄角度下的作物图像进行检测，难以在实际的收获过程中得到有效应用。本文围绕联合收获机实时在线检测作物喂入密度的需求，利用车载相机实时获取联合收获机前方的图像信息，并通过人工取样的方式获取小麦作物属性数据，进而对图像进行标定。提出基于 MobileViT 轻量化网络模型的小麦密度等级分类方法，并与 VGG16、GoogLeNet、ResNet 等模型进行对比，以期实现小麦喂入密度等级的预测，为联合收获机喂入量自动控制提供参考。

1　材料与方法

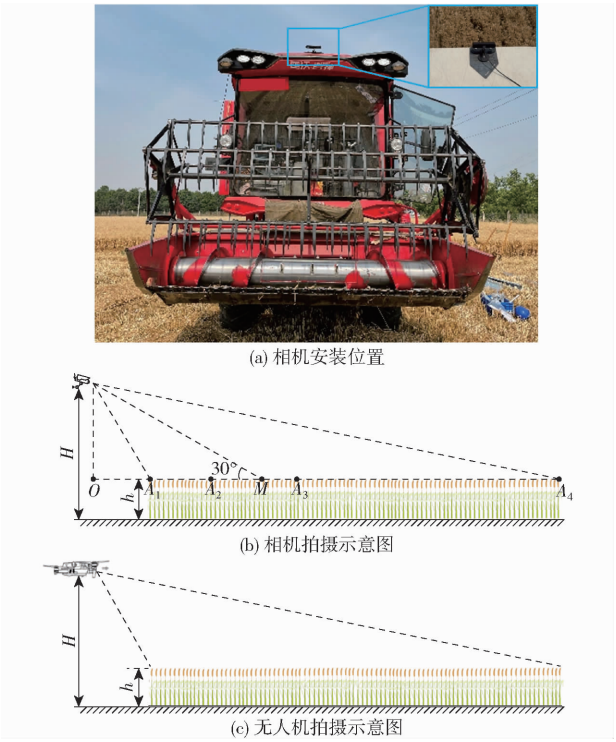
1.1　图像数据来源

小麦收获具有较强的时效性，为了获取收获期不同小麦密度的图像数据，于 2022 年 5—6 月，在湖北省襄阳市、河南省洛阳市、河北省保定市以及北京市国家精准农业示范基地进行田间小麦图像采集。为了收集联合收获机在作业过程中的图像，采用相机随车拍摄和无人机悬停拍摄两种不同的图像采集方式。

相机的安装方式如图 1a 所示，将 ZED2i 型相机安装在联合收获机车顶正上方。为了获得较宽阔的视野，提前预测未收割区域小麦作物密度，将相机与水平地面夹角设置为 30°。ZED2i 型相机是一款深度相机，该相机使用双目视觉技术，可以获取图像中作物的深度信息[15]，相机深度范围可达 0.2 ~ 20 m。图 1b 为相机的拍摄范围，点 A_1 对应图像底部端点，点 A_2 对应图像下三等分点，点 M 对应图像中心点，点 A_3 对应图像上三等分点，点 A_4 对应图像顶部端点，通过深度数据可以得到各点到相机的水平距离。尽管随车采集的图像是较好的数据来源，但其图像采集效率受限于收获机的作业效率。无人机航拍可以快速收集图像数据，其采集效率更高、布署更加灵活。使用大疆御 MAVIC AIR2 无人机来采集车载相机视角下的图像，图 1c 为无人机拍摄示意图，无人机飞行高度（3 550 mm）和拍摄角度与车载相机保持一致，无人机与车载相机先后进行图像数据采集。

1.2　作物样本采集

为了实现小麦图像密度等级分类，进行了小麦参数测定。采用五点法取样的方式对收割区的小麦作物密度进行测定。选取无底无盖的纸板箱放在作物中，取样箱尺寸为 20 cm × 30 cm，取样面积



(a) 相机安装位置

(b) 相机拍摄示意图

(c) 无人机拍摄示意图

图 1 图像采集方式

Fig. 1 Image acquisition method

为 600 cm^2,在测量区域内采集样本。取样时控制留茬高度在 10 cm 以内,收获后将籽粒与秸秆分离并在田间进行称量。将每平方米的籽粒和秸秆总质量记作喂入密度 ρ_s,其计算公式为

$$\rho_s = \frac{M_s}{S_s} \tag{1}$$

式中 ρ_s——收获机喂入密度,kg/m^2

M_s——取样质量,kg

S_s——采样面积,m^2

部分小麦样本数据如表 1 所示。根据喂入密度 ρ_s 来划分小麦密度等级,将小麦密度分为 4 个等级,分类标准如表 2 所示。

表 1 部分小麦作物样本数据

Tab. 1 Sample data of some wheat crops

组号	籽粒质量/kg	秸秆质量/kg	取样质量/kg	采样面积/m^2	喂入密度/(kg·m^{-2})	草谷比
1	0.056 2	0.038 6	0.094 8	0.06	1.58	0.69
2	0.062 1	0.033 9	0.096 0	0.06	1.60	0.55
3	0.077 3	0.048 1	0.125 4	0.06	2.09	0.62
4	0.084 3	0.050 7	0.135 0	0.06	2.25	0.60
5	0.066 8	0.041 2	0.108 0	0.06	1.80	0.62
6	0.075 4	0.044 0	0.119 4	0.06	1.99	0.58
7	0.097 3	0.055 7	0.153 0	0.06	2.55	0.57
8	0.077 1	0.039 9	0.117 0	0.06	1.95	0.52
9	0.069 1	0.050 3	0.119 4	0.06	1.99	0.73
10	0.079 8	0.051 6	0.131 4	0.06	2.19	0.65

表 2 小麦密度等级划分

Tab. 2 Classification of wheat density

密度等级	密度估计	密度 ρ_s/(kg·m^{-2})	特征
1	低	0~1.3	麦穗籽粒干瘪,植株分布稀疏
2	中	1.3~1.9	麦穗籽粒充实,植株分布均匀
3	高	1.9~2.2	麦穗籽粒充盈,植株分布密集
4	特高	>2.2	麦穗籽粒饱满,植株分布茂密

1.3 数据集构建

为了训练小麦作物密度分类识别模型,根据前文的小麦密度划分标准对采集的麦田图像数据进行了分类。图 2 为不同密度等级的小麦图像样本,数据集包含 248 幅低密度图像、424 幅中密度图像、689 幅高密度图像和 176 幅特高密度图像。然而,由于数据集较小且采集的图像尺寸不一致,同时存在样本不均衡问题,高密度图像数量约为特高密度图像数量的 3.9 倍,为了解决数据集较少导致的模型过拟合问题以及样本不均衡导致的模型过度关注多样本类别而缺乏对小样本类别识别能力的问题[16],对数据集进行了多种图像预处理操作,以平衡和扩充数据,从而有效地提高模型的准确率和泛化能力。无人机采集到的图像分辨率为 4 000 像素×2 250 像素,车载相机采集到的图像分辨率为 2 208 像素×1 242 像素,为了方便后续任务,对图像进行按比例缩放,将原始图像均调整为 4 000 像素×2 250 像素。为了减少小麦图像中不必要的信息,增加图像的可用性和多样性,采用九角裁剪[17]方法将原始图像等分为 9 部分,保留中间部分作为感兴趣的区域(图 3)。这样可以有效地去除其他作物行的影响,同时通过图像裁剪将数据集扩充 3 倍,此时数据集中包含 2 586 幅图像,图像尺寸为 1 334 像素×750 像素,随机选取 20% 图像数据作为测试集(513 幅图像),将剩下的数据按 9∶1 划分为训练集和验证集。为了增强深度学习模型的泛化能力和鲁棒性,使得模型可以更好地适应不同的数据分布和应用场景,获得更好的分类结果,对训练集和验证集进行了 1 倍数据增强,包括亮度调整、对比度调整、锐度调整、左右翻转和高斯噪声等方式。通过上述处理,本文构建一个包含 4 659 幅图像的数据集,并将图像数据集划分为训练集、验证集和测试集,用于训练和评估模型。其中,训练集包含 3 732 幅图像,验证集包含 414 幅图像,测试集包含 513 幅图像。

1.4 MobileViT 网络模型

MobileViT[18] 是一个轻量级的视觉 Transformer[19]模型,由于其高效的推理速度和较少的参数量,在移动设备和嵌入式设备上具有广泛的

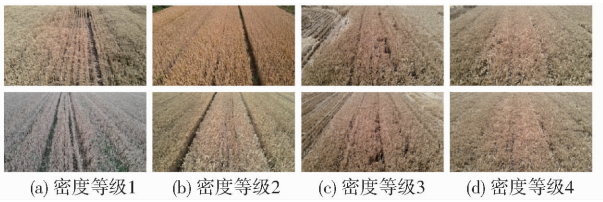


图 2　不同密度等级的小麦图像样本

Fig. 2　Wheat image samples of different density levels

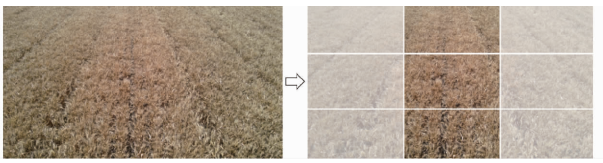


图 3　图像九角裁剪

Fig. 3　Image nine-corner cropping

应用前景。MobileViT 模型结合了 CNN 的高效和 Transformer 的自注意力机制和全局视野，该模型依次由普通卷积、MobileNetV2（MV2）结构[20]、MobileViT block（MViT）结构、全局池化以及全连接层组成，主体结构如图 4 所示。MobileViT 模型的卷积层用于提取输入图像的低级和中级特征，而 Transformer 模块用于提取输入图像的高级语义特征。在 MobileViT 模型中，Transformer 模块的输入是经过卷积层提取的特征图，每个位置的特征向量表示该位置的像素在特征空间中的语义信息，Transformer 模块使用自注意力机制对这些特征向量进行编码，在不丢失全局信息的同时保留局部特征，然后通过前馈网络对特征进行非线性变换，自注意力机制和前馈网络模块协同作用以提高模型的预测精度，最后通过一个全局池化层对 Transformer 模块的输出进行池化操作，将得到的特征向量输入到全连接层进行类别预测。在 MobileViT 中有 3 种不同的配置，分别是 MobileViT－S、MobileViT－XS 和 MobileViT－XXS，这些配置的主要区别在于模型的深度和通道数不同，对应影响模型的精度和计算量。

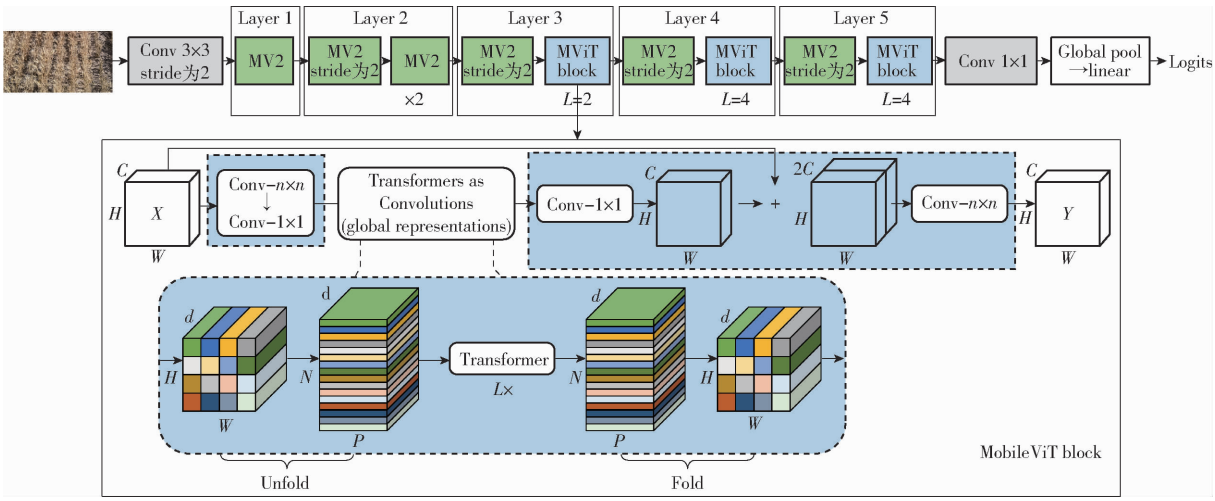


图 4　MobileViT 模型结构图

Fig. 4　MobileViT model structure diagram

MobileViT 模型的核心分别为 MV2 结构和 MViT 结构。在 MV2 结构中，如图 5 所示，当 stride 为 1 且输入特征矩阵与输出特征矩阵形状相同时，有一个捷径分支连接输入和输出，使得输入特征直接与输出特征相加，从而保持特征图大小不变，捷径分支可以避免信息的丢失，并且在训练过程中能够帮助模型更快地收敛。当 stride 为 2 时，输入特征矩阵的通道与输出特征矩阵不同，需要进行下采样。这时 MV2 结构使用步长为 2 的卷积层进行下采样，同时使用一个 1×1 的卷积层来调整通道数，使得下采样后的特征图大小和通道数与输出特征矩阵相同。

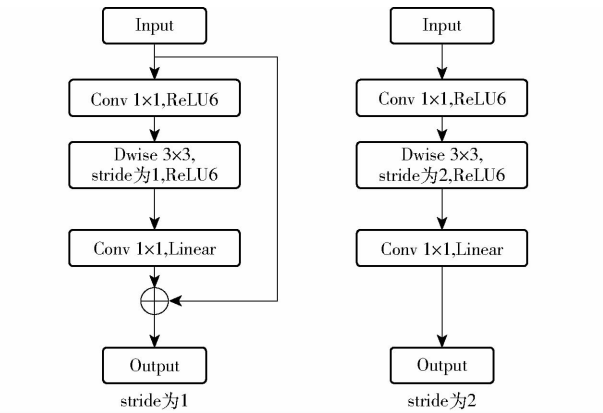


图 5　MV2 结构图

Fig. 5　MV2 structure diagram

MViT 结构首先对经过前期处理的特征图进行局部信息编码，使用卷积核大小为 $n \times n$ 和 1×1 的卷积层进行局部信息提取，特征图通过 Transformers as Convolutions（TaC）进行全局信息编码，Unfold 将 $H \times W \times d$ 的输入特征图展开为 $P \times N \times d$，然后应用 Transformer 进行编码，使用 Fold 将编码后 $P \times N \times d$ 特征图拼接回原始形状 $H \times W \times d$，通过一个 1×1

的卷积核进行通道数的调整,再与原始输入特征图进行拼接完成特征信息的融合,最后使用一个大小为 $n \times n$ 的卷积核进行通道融合,得到最终的输出特征图。

2 模型训练

2.1 试验环境

采用 Windows 操作系统,基于 Pytorch 1.10 深度学习框架和 Python 3.7 编译环境进行网络模型训练与测试。将图像测试样本尺寸设置为 224 × 224,并使用评估指标对小麦密度等级分类模型的检测结果进行评价,从而调整模型参数以达到最优的训练效果。

2.2 超参数设置

学习率是一个非常关键的参数,其取值直接影响模型的收敛效果。如果学习率过大,会导致损失函数直接跳过全局最优点,进而导致损失值过大;而学习率过小,则会大大增加网络的收敛复杂度,并且容易被困在局部最小值中,从而无法找到最优参数。因此,设置合适的学习率对于深度学习模型的训练至关重要。本文采用指数标尺选择了 4 个不同的学习率(0.005、0.000 5、0.001、0.000 1)进行试验[21]。在使用 MobileViT 模型时,其他参数保持一致,将迭代训练次数设置为 30 个周期。以 MobileViT - S 为例,在训练集上完成训练后,其在验证集上准确率如图 6 所示。结果表明,学习率为 0.005 时,模型在验证集上准确率最低;学习率为 0.000 1 时,模型在验证集上准确率整体呈上升趋势,效果最好。

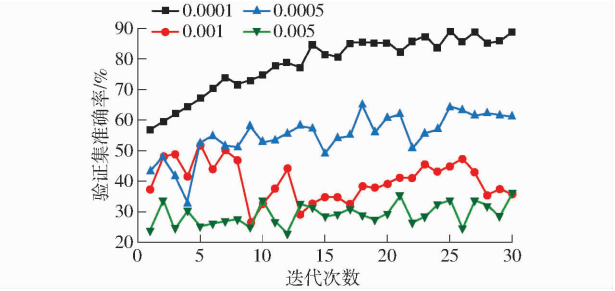


图 6 不同学习率时 MobileViT - S 在验证集上准确率曲线

Fig. 6 Accuracy curves of MobileViT - S on validation set with different learning rates

在深度学习中,存在多种不同的优化器可供选择,如 SGD、Momentum、Adam 等[22]。这些优化器可以分为两大类:第 1 类是学习率不受梯度影响,可以保持不变或按照特定的时间表变化,其中最常见的是随机梯度下降法(SGD);第 2 类是学习率根据梯度变化自适应调整,以尽可能减小全局学习率的影响,其中最常用的是自适应矩估计算法(Adam)。在训练集上训练后,使用不同的优化器在验证集上进行测试,如图 7 所示。结果显示,当优化器为 SGD 时,模型在验证集上准确率上升缓慢;而当优化器为 Adam 时,模型在验证集上准确率迅速提升,并且整体准确率都高于使用 SGD 优化器。因此,在后续的模型训练中,优化器设置为 Adam。

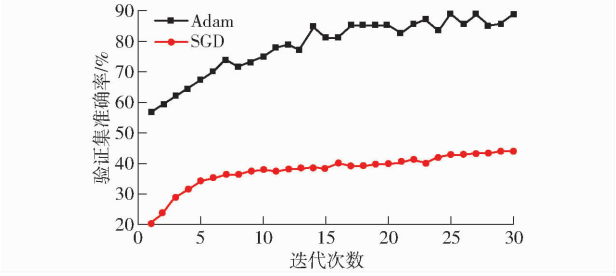


图 7 不同优化器时 MobileViT - S 在验证集上准确率曲线

Fig. 7 Accuracy curves of MobileViT - S on validation set with different optimizers

2.3 评价指标

在本试验中,采用混淆矩阵作为主要的评价标准来衡量模型的性能。通过混淆矩阵,可以计算出准确率、精确率、召回率和特异度等相关指标,以评估模型的训练效果。准确率是预测正确的密度等级数量占所有密度等级数量的百分比。精确率是正确预测小麦密度等级的数量占所有预测为正例的小麦密度实例个数的百分比。召回率是对某小麦密度等级预测为正确数量的百分比。特异度指真正不属于某一小麦密度等级的图像中,被判断为非该小麦密度等级的百分比。本研究中的小麦密度等级数据集由 4 个类别组成,在分类过程中使用 4 类混淆矩阵,如表 3 所示,其中 M1、M2、M3、M4 分别为低密度等级、中密度等级、高密度等级和特高密度等级。T1 ~ T4 表示真实类别为 M1 ~ M4 的图像被正确预测为类别 M1 ~ M4 的数量;F21、F31、F41 表示真实类别为 M2、M3、M4 的图像被误识为类别 M1 的数量;F12、F32、F42 表示真实类别为 M1、M3、M4 的图像被误识为类别 M2 的数量;F13、F23、F43 表示真实类别为 M1、M2、M4 的图像被误识为类别 M3 的数量;F14、F24、F34 表示真实类别为 M1、M2、M3 的图像被误识为类别 M4 的数量。

表 3 小麦密度等级 4 类混淆矩阵

Tab. 3 Confusion matrix of four types of wheat density grades

真实类别	预测类别			
	M1	M2	M3	M4
M1	T1	F12	F13	F14
M2	F21	T2	F23	F24
M3	F31	F32	T3	F34
M4	F41	F42	F43	T4

3　结果分析

为了比较传统卷积模型和轻量化网络模型在小麦图像密度预测中的性能，对所有模型采用相同的输入大小，并在相同的计算环境下运行，使用传统的性能指标来评估模型的有效性。对于构建的分类训练集和验证集，图 8 绘制了各模型在训练过程中的损失值曲线以及各模型在验证过程中的准确率曲线。

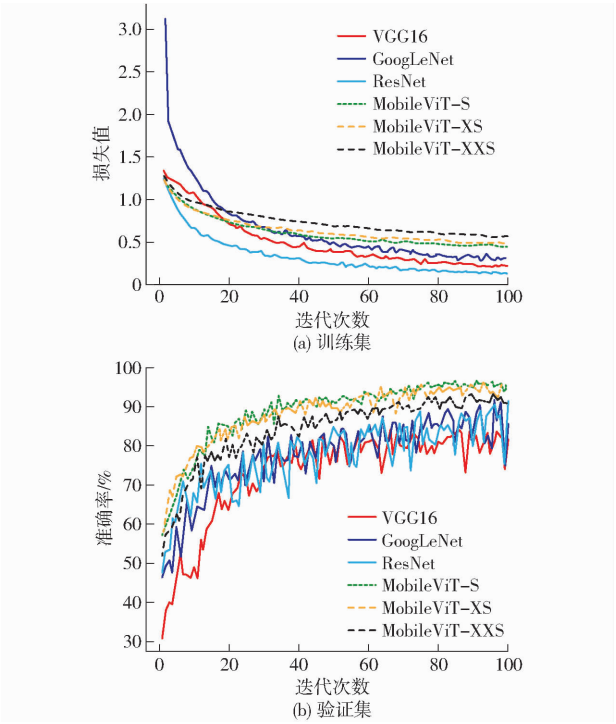


图 8　各模型的损失值和准确率变化曲线

Fig. 8　Loss value and accuracy curves of each model

根据图 8a 所示，模型的训练损失值随着迭代次数的增加逐渐收敛。其中，GoogLeNet 模型的初始损失值最高，高于其他模型，而 ResNet 模型的损失值始终低于其他模型。所有模型在前 20 次迭代中的损失值急剧下降，表明模型的权重更新跨度较大，进入了快速优化阶段；20 ~ 80 次迭代期间，损失值略有下降，说明模型进入了微调阶段并接近最优值；最后 20 次迭代期间，损失曲线趋于稳定，表明模型训练已达到饱和状态，基本上达到最佳性能。MobileViT－S、MobileViT－XS、MobileViT－XXS 模型的损失值在达到稳定状态后一直保持在 0.6 以下。根据图 8b 的结果，VGG16、GoogLeNet 和 ResNet 模型的验证准确率整体呈现较大波动。VGG16 模型在第 66 次迭代时达到最佳准确率 84.50%，但随着迭代次数的增加，验证集的准确性开始降低，此时模型开始过拟合。GoogLeNet 和 ResNet 模型的最佳精度分别在迭代第 96 次和 100 次时出现，分别为 91.47% 和 91.28%。相比之下，MobileViT 模型的验证准确率整体上高于传统卷积模型，并且模型的验证准确率随着迭代次数的增加逐步趋于稳定。MobileViT－S、MobileViT－XS 和 MobileViT－XXS 模型在验证集上的最大准确率分别出现在第 85 次、第 93 次和第 96 次迭代时，分别为 96.71%、96.32% 和 93.22%。

在测试集上，使用 VGG16、GoogLeNet、ResNet、MobileViT－S、MobileViT－XS、MobileViT－XXS 模型对 513 幅不同密度等级的小麦图像进行分类，各模型的整体准确率分别为 87.52%、91.03%、88.69%、88.89%、91.03%、86.74%，其中 GoogLeNet 和 MobileViT－XS 模型的分类效果最好，相比 VGG16、ResNet、MobileViT－S、MobileViT－XXS 分别高 3.51、2.34、2.14、4.29 个百分点。通过混淆矩阵（图 9）可以观察到各模型的误判情况，具体来说，GoogLeNet 模型在第 4 类小麦密度图像上的误判最多，达到 19 幅；VGG16 模型误判类型主要集中在第 3 类小麦密度图像上，共计 25 幅；ResNet 模型同样在第 3 类小麦密度图像上误判数量最高，共计 28 幅。从混淆矩阵图 10 中可以看出，MobileViT－XS 模型的分类错误最少，仅有 46 幅图像被错误分类，而 MobileViT－XXS 模型的分类错误最多，达到了 68 幅图像。MobileViT 模型对第 4 密度等级的小麦图像的分类错误最多，其中 MobileViT－S、MobileViT－XS、MobileViT－XXS 模型分别出现了 14、22、28 幅误分类图像。

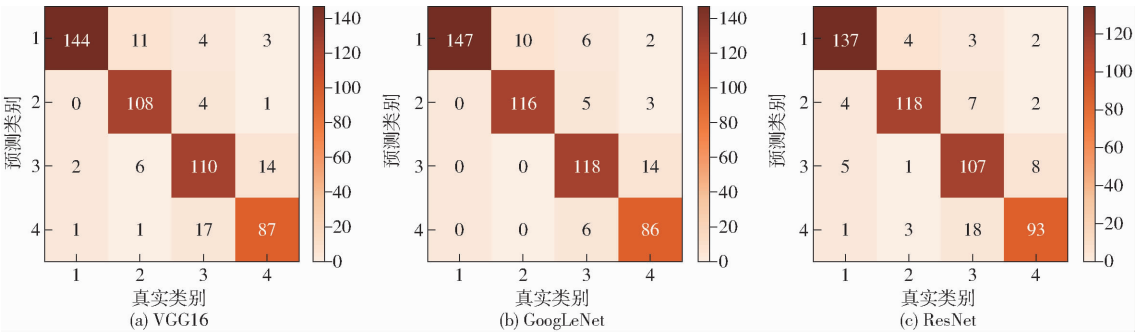


图 9　传统神经网络模型混淆矩阵图

Fig. 9　Confusion matrix diagrams of traditional neural network model

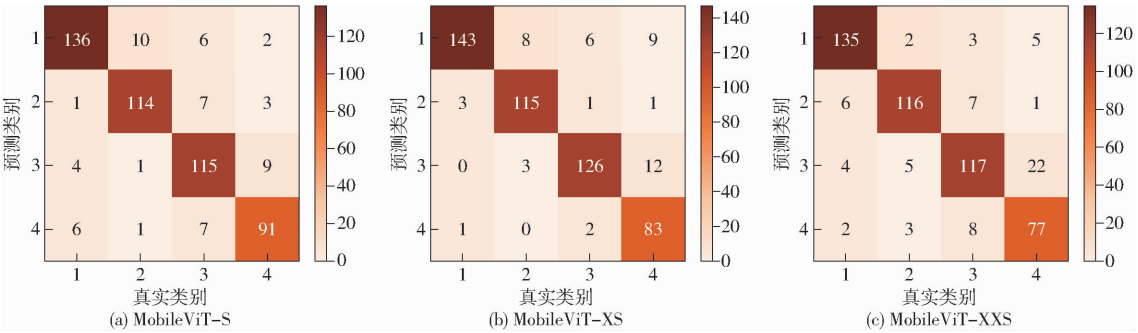


图 10 轻量化模型混淆矩阵图

Fig. 10 Confusion matrix diagrams of lightweight model

为了测试不同模型在处理小麦图像时的效率，选择整个测试集的图像数据进行了 5 次时间测试，并计算了平均测试时间，测试结果如表 4 所示。MobileViT 系列模型的推理时间相对于传统卷积神经网络模型更长。然而，随着模型内存占用量的减小，MobileViT 的计算时间也相应减小。MobileViT－S、MobileViT－XS、MobileViT－XXS 平均识别速度分别为 33.1、33.9、34.7 幅/s。因此，MobileViT 系列模型可以满足实际生产中对预测分类实效性的要求。

表 4 不同模型测试时间

Tab. 4 Test time of different models s

模型	测试 1	测试 2	测试 3	测试 4	测试 5	平均值
VGG16	12.81	12.69	12.64	12.55	12.58	12.65
GoogLeNet	13.91	13.80	13.71	13.89	14.03	13.87
ResNet	13.33	13.24	13.87	13.66	13.51	13.52
MobileViT－S	15.77	15.37	15.27	15.70	15.30	15.48
MobileViT－XS	15.06	15.07	15.06	15.18	15.16	15.11
MobileViT－XXS	14.76	14.87	14.80	14.67	14.81	14.78

针对分类任务，模型的预测精度是其中一个重要的性能指标，综合考虑模型参数量、内存占用量和浮点运算次数（FLOPs）等方面的指标，不同模型的性能测试结果如表 5 所示。MobileViT－XXS 模型具有最小的模型参数量、内存占用量和浮点运算次数，但其识别准确率较低。VGG16 模型具有最快的检测速度，但其参数量和内存占用较大，不利于部署在移动端。MobileViT－XS 模型在这些方面都表现出色，它在测试集上的整体识别准确率达到了 91.03%，该模型所需的内存为 7.5 MB，更适合部署在嵌入式设备上。此外，MobileViT－XS 的识别速度也很快，单幅图像的推理时间只需 29.5 ms，每秒推理图像约 34 幅。

为了评估 MobileViT 模型对不同深度范围的小麦图像的识别效果，按照小麦植株到相机的水平距离分为 3 个区间，区间 1 范围为 1.7～3.5 m、区间 2 范围为 3.5～6.1 m、区间 3 范围为 6.1～15.5 m，将测试集图像分成 3 类并对模型进行进一步评估。MobileViT－S、MobileViT－XS、MobileViT－XXS 在不同深度范围的测试集图像上平均识别准确率分别为 91.23%、90.06%、85.38%，87.72%、92.98%、92.40%，88.30%、85.38%、86.55%。为了进一步了解模型对输入图像中不同区域的关注程度以及模型的决策过程，并进行模型的可解释性分析，使用 Grad－CAM 来可视化模型对图像中不同区域的关注程度[23]。通过图像可视化，可以得到 MobileViT 模型在不同距离的图像中对于高亮区域的分布范围和集中度表现不同。从图 11 可以看出，MobileViT－S、MobileViT－XS 和 MobileViT－XXS 3 种模型在不同距离的图像中，对于高亮区域的分布范围和集中度表现不同。MobileViT－XS 在中距离的图像中高亮区域分布更广且更集中，因此表现出更高的识别准确率。

表 5 不同模型性能测试结果

Tab. 5 Performance test results of different models

模型	模型参数量	内存占用量/MB	浮点运算次数	推理时间/ms	准确率/%
VGG16	$1.342\,8\times10^8$	512.0	1.044×10^6	24.7	87.52
GoogLeNet	1.032×10^7	39.3	1.588×10^6	27.0	91.03
ResNet	2.129×10^7	81.3	1.470×10^6	26.4	88.69
MobileViT－S	4.94×10^6	19.0	1.576×10^6	30.2	88.89
MobileViT－XS	1.93×10^6	7.5	8.18×10^5	29.5	91.03
MobileViT－XXS	9.5×10^5	3.8	3.19×10^5	28.8	86.74

4 结论

（1）针对目前联合收获机喂入量控制存在滞后性的问题，提出了一种作物密度预测方法，通过实际采集不同生长状况和不同密度等级的小麦图像，采用深度学习模型对小麦图像密度等级进行分类，以提前预测联合收获机的喂入密度。

（2）采用车载相机和无人机两种不同的图像采集方式，利用车载相机实时采集小麦图像数据，使用

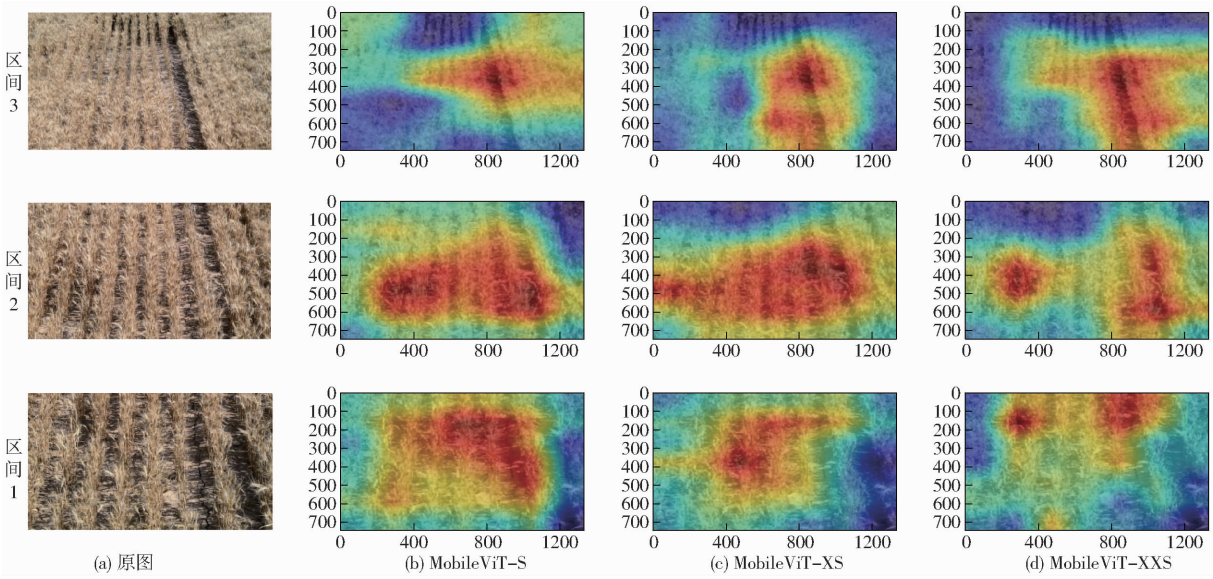


图 11　MobileViT 模型在不同深度图像上的热力对比

Fig. 11　Thermal comparison of MobileViT model on different depth images

无人机在相同高度和角度下拍摄小麦图像以获取更多的图像数据，最终将车载相机图像与无人机图像相结合，构建小麦图像数据集。

(3)提出了一种基于 MobileViT 的轻量化网络模型，并将其与传统卷积模型进行对比，通过检测准确率、检测速度、模型参数量和内存占用量等评价指标对不同模型进行评估。试验结果表明，MobileViT-XS 模型在小麦图像密度等级识别方面表现优异，达到 91.03% 的识别准确率，每秒推理图像约 34 幅，内存占用量为 7.5 MB。该模型更适合部署在移动式设备上，为联合收获机喂入量控制和智能化收割提供了一种有效的解决方案。

参　考　文　献

[1] 丁翰韬，陈树人，周巍伟，等. 喂入量扰动下联合收获机振动特性机理研究[J]. 农业机械学报，2022，53(增刊 2)：20－27，51.
DING Hantao，CHEN Shuren，ZHOU Weiwei，et al. Mechanism analysis of combine harvester's vibration characteristics under feeding interference[J]. Transactions of the Chinese Society for Agricultural Machinery，2022，53(Supp. 2)：20－27，51. (in Chinese)

[2] 陈度，王书茂，康峰，等. 联合收割机喂入量与收获过程损失模型[J]. 农业工程学报，2011，27(9)：18－21.
CHEN Du，WANG Shumao，KANG Feng，et al. Mathematical model of feeding rate and processing loss for combine harvester[J]. Transactions of the CSAE，2011，27(9)：18－21. (in Chinese)

[3] 宁小波，陈进，李耀明，等. 联合收获机前进速度模糊控制系统多目标遗传优化[J]. 农业机械学报，2015，46(5)：68－74.
NING Xiaobo，CHEN Jin，LI Yaoming，et al. Multi-objective genetic algorithm optimization of forward speed of fuzzy control system for combine harvester[J]. Transactions of the Chinese Society for Agricultural Machinery，2015，46(5)：68－74. (in Chinese)

[4] 庄肖波，李耀明. 基于鲁棒反馈线性化的联合收获机割台高度控制策略[J]. 农业机械学报，2020，51(11)：123－130.
ZHUANG Xiaobo，LI Yaoming. Header height control strategy of harvester based on robust feedback linearization[J]. Transactions of the Chinese Society for Agricultural Machinery，2020，51(11)：123－130. (in Chinese)

[5] 尤惠媛，卢文涛. 联合收获机喂入量模糊控制系统研究[J]. 西北农林科技大学学报(自然科学版)，2015，43(5)：229－234.
YOU Huiyuan，LU Wentao. Fuzzy control system for feed quantity of combine harvester[J]. Journal of Northwest A&F University(Natural Science Edition)，2015，43(5)：229－234. (in Chinese)

[6] 李平，冯伟，崔晋波，等. 喂入量自适应技术在小型联合收割机中的应用[J]. 农机化研究，2022，44(5)：157－162.
LI Ping，FENG Wei，CUI Jinbo，et al. Application of feed rate adaptive technology in small combine[J]. Journal of Agricultural Mechanization Research，2022，44(5)：157－162. (in Chinese)

[7] 姬江涛，王荣先，符丽君. 联合收获机喂入量灰色预测模糊 PID 控制[J]. 农业机械学报，2008，39(3)：63－66.
JI Jiangtao，WANG Rongxian，FU Lijun. Grey prediction fuzzy PID control of the feeding quantity in combine[J]. Transactions of the Chinese Society for Agricultural Machinery，2008，39(3)：63－66. (in Chinese)

[8] 孟志超，贺磊盈，杜小强，等. 基于 Enhanced VGG16 的油茶品种分类[J]. 农业工程学报，2022，38(10)：176－181.
MENG Zhichao，HE Leiying，DU Xiaoqiang，et al. Classification of *Camellia oleifera* based on Enhanced VGG16 network[J].

Transactions of the CSAE,2022,38(10):176 - 181. (in Chinese)
[9] LUO T, ZHAO J, GU Y, et al. Classification of weed seeds based on visual images and deep learning[J]. Information Processing in Agriculture,2023,10(1):40 - 51.
[10] MODI R U, CHANDEL A K, CHANDEL N S, et al. State-of-the-art computer vision techniques for automated sugarcane lodging classification[J]. Field Crops Research,2023,291:108797.
[11] 潘静,邵陆寿,王轲. 水稻联合收割机喂入密度检测方法[J]. 农业工程学报,2010,26(8):113 - 116.
PAN Jing, SHAO Lushou, WANG Ke. Detection method on feed density for rice combine[J]. Transactions of the CSAE,2010,26(8):113 - 116. (in Chinese)
[12] 刘海娟,邵陆寿,朱建辉,等. 基于图像处理的水稻成熟期密度检测[J]. 昆明理工大学学报(理工版),2009,34(2):84 - 88.
LIU Haijuan, SHAO Lushou, ZHU Jianhui, et al. Detection of rice density during maturity based on image processing[J]. Journal of Kunming University of Science and Technology(Natural Science Edition),2009,34(2):84 - 88. (in Chinese)
[13] 刘仲鹏,赵建军,刘丽娟. 基于改进算法的小麦收获机作物密度检测研究[J]. 农机化研究,2015,37(8):49 - 52.
LIU Zhongpeng, ZHAO Jianjun, LIU Lijuan. Study on wheat harvester crop density detection during maturity based on the improved algorithms[J]. Journal of Agricultural Mechanization Research,2015,37(8):49 - 52. (in Chinese)
[14] SÉBASTIEN D, ELIAS E, ALEXIS C, et al. Deep learning for wheat ear segmentation and ear density measurement: from heading to maturity[J]. Computers and Electronics in Agriculture,2022,199:107161.
[15] 魏新华,张敏,刘青山,等. 基于双目视觉的田间作物高度和收割边界信息提取[J]. 农业机械学报,2022,53(3):225 - 233.
WEI Xinhua, ZHANG Min, LIU Qingshan, et al. Extraction of crop height and cut-edge information based on binocular vision [J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(3):225 - 233. (in Chinese)
[16] 李艳霞,柴毅,胡友强,等. 不平衡数据分类方法综述[J]. 控制与决策,2019,34(4):673 - 688.
LI Yanxia, CHAI Yi, HU Youqiang, et al. Review of imbalanced data classification methods[J]. Control and Decision,2019,34(4):673 - 688. (in Chinese)
[17] SUHARJIT O, ELWIREHARDJA G N, PRAYOGA J S. Oil palm fresh fruit bunch ripeness classification on mobile devices using deep learning approaches[J]. Computers and Electronics in Agriculture,2021,188:106359.
[18] MEHTA S, RASTEGARI M. MobileViT: light-weight, general-purpose, and mobile-friendly vision transformer[J]. arXiv Preprint, arXiv:2110.02178,2021.
[19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems,2017,30:5998 - 6008.
[20] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018:4510 - 4520.
[21] SMITH L N. Cyclical learning rates for training neural networks[C]// 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE,2017: 464 - 472.
[22] RUDER S. An overview of gradient descent optimization algorithms[J]. arXiv Preprint, arXiv:1609.04747,2016.
[23] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C]// Proceedings of the IEEE International Conference on Computer Vision,2017:618 - 626.