doi:10.6041/j.issn.1000-1298.2026.23.030

基于 FPGA 加速 LLN 的农作物病害识别方法

段瑞枫[1,2]　张添茗[1]　陈羽茜[1]　冯宇航[1]　杨　锋[1,2]　齐建东[1,2]
(1.北京林业大学信息学院,北京 100083;2.河北省智慧国家公园重点实验室,北京 100083)

摘要:农作物病害会导致农作物减产,而高精度病害识别是实施精准病害防控的前提,卷积神经网络(Convolutional neural network,CNN)病害识别是近年来卓有成效的方法。但传统 CNN 对计算资源要求高,难以部署在资源受限的边缘设备。针对这一难题,本文基于现场可编程门阵列(Field programmable gate array,FPGA)平台提出一种高效实时的田间农作物病害识别方案。首先,提出了一种带有跳跃连接的轻量级神经网络(Light link net,LLN),该模型通过采用深度可分离卷积及批归一化层融合等技术将网络参数量压缩到 1.4×10^3,浮点运算次数(Floating point operations,FLOPs)降低至 2.2×10^6,在不同深度的卷积层之间通过跳跃连接引入初始病害特征以保证关键特征提取,并结合知识蒸馏方法训练模型,使得网络识别精度达到了 97.6%。其次,基于 FPGA 平台设计 LLN 加速电路,先运用高层次综合工具(High-level-synthesis,HLS)开发硬件电路,然后综合采用模型量化及特征图分块技术进一步减少模型存储与计算开销。同时,采用双缓冲机制与多种并行度优化策略,显著提高电路运算效率,有效减少推理时延。最后,将优化后的系统部署于 FPGA 开发板 Zynq MZ7030FA,采用 PlantVillage 数据集进行测试,实验室环境下单帧图像推理时间 55.7 ms,运行功耗 3.78 W,为农田病害实时精准检测提供了高效且低成本的解决方案。

关键词:农作物病害识别;轻量化神经网络;FPGA 加速;实时;低成本
中图分类号:S24　　**文献标识码**:A　　**文章编号**:1000-1298(2026)23-0316-13　　**OSID**:

Method for Crop Disease Recognition Based on FPGA-accelerated LLN

Duan Ruifeng[1,2]　Zhang Tianming[1]　Chen Yuxi[1]　Feng Yuhang[1]　Yang Feng[1,2]　Qi Jiandong[1,2]
(1. *School of Information Science and Technology*, *Beijing Forestry University*, *Beijing* 100083, *China*
2. *Hebei Key Laboratory of Smart National Park*, *Beijing* 100083, *China*)

Abstract: Crop diseases can lead to reduced crop yields, and high-precision disease identification is a prerequisite for implementing precise disease control. Convolutional neural network (CNN)-based disease identification has been a highly effective method in recent years. However, traditional CNNs have high requirements for computing resources and are difficult to deploy on resource-constrained edge devices. To address this challenge, it proposed an efficient and real-time field crop disease identification solution based on the field programmable gate array (FPGA) platform. Firstly, a lightweight neural network with skip connections, light link net (LLN), was proposed. This model compressed the network parameters to 1.4×10^3 and reduced the number of floating point operations (FLOPs) to 2.2×10^6 by adopting depthwise separable convolution (DSC) and batch normalization (BN) layer fusion techniques. Skip connections were used between convolution layers of different depths to introduce initial disease features and ensure the extraction of key features. The model was trained by using the knowledge distillation method, achieving a recognition accuracy of 97.6%. Secondly, an LLN acceleration circuit was designed based on the FPGA platform. High-level synthesis (HLS) tools were used to develop the hardware circuit, and model quantization and feature map block techniques were comprehensively adopted to further reduce model storage and computational costs. Meanwhile, a dual-buffer mechanism and multiple parallel optimization strategies were employed to significantly improve the circuit operation efficiency and

收稿日期:2025-08-14　修回日期:2025-10-14
基金项目:中央高校基本科研业务费项目(CGZH202509)和国家重点研发计划项目(2022YFF1302700)
作者简介:段瑞枫(1985—),女,副教授,博士,主要从事林草业智能监测与高效数据传输等研究,E-mail:drffighting2019@bjfu.edu.cn
通信作者:杨锋(1988—),男,副教授,博士,主要从事人工智能与图像识别、林草智能监测研究,E-mail:fengyang@bifu.edu.cn

reduce inference latency. Finally, the optimized system was deployed on the FPGA development board Zynq MZ7030FA and tested by using the PlantVillage dataset. In the laboratory environment, a detection speed of 55.7 ms per frame and a low power consumption of 3.78 W were achieved, providing an efficient and low-cost solution for real-time and accurate detection of crop diseases in the field.

Key words: crop disease recognition; lightweight neural network; FPGA acceleration; real-time; low cost

0　引言

植物病害一直是制约农作物优质高产的重要因素之一，而高精度识别对于病害的高效防治至关重要。由于植物病害种类繁多、症状相似性高，传统人工检测方式存在效率低、主观性强、误诊率高等问题，不仅消耗大量人力资源，且错误的诊断极易导致防治措施失效，进一步加剧经济损失[1]。在此背景下，基于计算机视觉的智能检测技术展现出显著优势，通过自动化、高精度的病害识别，有效降低人工成本，提升检测效率与准确性，为农业病虫害防控提供了智能化解决方案[2]。

随着人工智能技术的蓬勃发展，机器学习方法被广泛应用于农业病害检测。传统的机器学习方法包括：决策树、贝叶斯算法和支持向量机等[3]，这些方法虽理论成熟且经过验证，但普遍存在准确率有限、泛化能力较弱的问题。近年来，深度学习凭借强大的特征提取与模式识别能力，在识别精度和泛化性能上展现显著优势[4-8]。这些研究成果充分验证了深度学习在农业病害识别领域的技术可行性与应用潜力。

尽管深度学习方法已经在实验室环境下的农业病害识别任务中表现优异，但卷积神经网络蕴含大量的参数以及浮点运算次数（Floating point operations，FLOPs）[9]，导致其通常依赖高性能计算服务器实现部署。GPU 是目前主流的深度学习加速器，虽然其拥有很强的计算能力，但高功耗与庞大体积严重限制了该类器件在农田等野外场景的部署可行性[10]，而且 GPU 通常价格高昂，使普通农户难以承受[11]。同时，不同于实验室的农田环境对设备的低功耗和便携性也有更为严格的要求[12]。与 GPU 相比，现场可编程门阵列（Field programmable gate array，FPGA）器件有着低成本、低功耗、高并行度及便于在嵌入式设备中集成等优势[13]。

近年来，基于 FPGA 加速卷积神经网络成为了研究热点，并且已在病害识别领域的网络部署方面取得一些进展。Luo 等[14]设计了轻量级网络 LiteCNN 识别植物病害，精度达到了 95.2%，并通过 FPGA 加速使其可以部署在田间环境中，平均每帧 71 ms。郭冬冬等[15]进一步优化网络，在 MobileNetV2 的基础上改进网络，并部署在 FPGA 上，以每帧 72 ms 的检测速度完成了对芒果外观品级的实时检测。尽管这些研究都成功将神经网络部署在 FPGA 上完成了室外环境检测，但目前这 2 种 FPGA 加速神经网络的方法仍存在识别准确度不高、检测时间较长等问题。这归因于，基于 FPGA 部署神经网络需将串行算法映射为并行电路，性能优越、复杂度高的神经网络将映射成大规模的电路，在低成本、运算资源有限的 FPGA 平台部署时会面临严重挑战[16]。为了得到更准确更迅速的检测效果，可以从以下两个方面进行改进[17]：①压缩或设计性能更好的网络模型，在保证高准确度情况下有更低的复杂度。②创新 FPGA 硬件架构设计，通过优化并行计算逻辑与资源调度策略，进一步缩短计算周期，实现更高效的 FPGA 加速方案。

本文将采用 FPGA 作为硬件平台来部署神经网络为植物病害实时识别提供可行方案。综合考虑模型识别准确性、FPGA 部署可行性与识别速度，首先设计带有跳跃连接的轻量级神经网络 LLN（Light link net），该网络运用深度可分离卷积（Depthwise separable convolution，DSC）、模型量化与批归一化（Batch normalization，BN）层融合技术，有效压缩模型体量。同时，该网络引入跳跃连接机制，确保浅层特征提取的完整性，并结合知识蒸馏方法训练模型，提高识别准确度。在此基础上，采用高层次综合工具（High-level-synthesis，HLS）实现 LLN 网络的 FPGA 电路设计，通过模型量化、特征图分块策略降低计算负荷，使模型易于部署在资源有限的 FPGA 上；再采取双缓冲机制优化访存效率，配合循环展开、循环流水线与数组分割方法设计进一步提高硬件并行处理能力，加快网络在 FPGA 上的推理速度。最后，将优化后的电路部署于 Zynq MZ7030FA 开发板，以期实现农业病害的高精度实时检测，有效平衡农田环境下病害识别的速度与准确性需求。

1　实验平台与检测流程

1.1　Zynq MZ7030FA 开发板

MZ7030FA 是一款高性能、低成本、资源丰富的集成开发板，提供 ARM 和 FPGA 资源环境。图 1 为 MZ7030FA 开发板资源结构，采用 Xilinx Kintex－7

系列的 Zynq XC7Z030 作为主控芯片,该芯片由主频 800 MHz 的双核 ARM Cortex - A9 处理系统(Processing system,PS 端)和可编程逻辑模块(Programmable logic,PL 端)两部分构成,二者通过 AXI 总线进行数据交互[18]。PS 端的 ARM 用于控制图像向 PL 端的输入输出以及 LLN 电路的执行调度,PL 端执行神经网络的推理运算,可通过 PS 端与 PL 端协同工作完成任务。

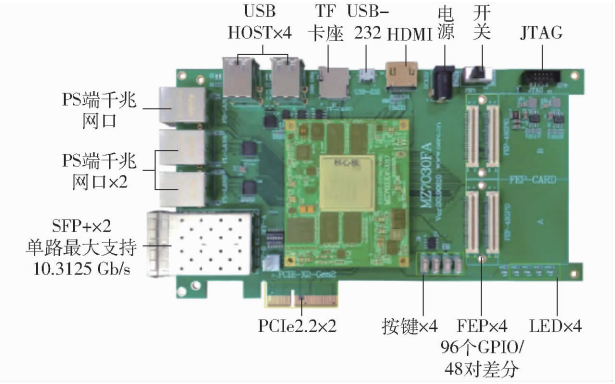

图 1 Zynq MZ7030FA 结构图

Fig. 1 Zynq MZ7030FA structure diagram

该 FPGA 的 PL 端硬件资源参数如下:片上存储器(Block RAM,BRAM)265 块,存储容量共 9.3 Mb,查找表(Look-up table,LUT)78 600 个,触发器(Flip-flop,FF)157 200 个,数字信号处理单元(Digital signal processor,DSP)400 个。

1.2 FPGA 检测流程

基于 FPGA 的农作物病害检测流程采用软硬件协同的方式,检测流程如图 2 所示。软件层面,基于 PyTorch 框架搭建轻量级网络 LLN,通过知识蒸馏训练压缩模型并与其他模型比较性能,验证 LLN 模型符合轻量级且识别精度高的要求后将模型参数保存在 SD 卡中,以供后续 FPGA 在检测时调用。硬件板卡的 PS 端读取保存的网络参数以及病害图像后,通过 AXI 总线将数据传输至 PL 端并调度网络执行;PL 端完成网络的卷积、激活等计算密集型操作,实现前向传播并生成预测结果;最终由 PS 端将预测结果输出至外部接口。该流程通过软硬件协同设计,实现了从模型训练到硬件加速推理的整个过程。

2 网络设计与优化

2.1 数据集及预处理

在实验室环境下训练 LLN 模型,本研究采用的数据来自 Kaggle 网站的 PlantVillage 数据集(植物病害数据集网址:https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset?select=test),

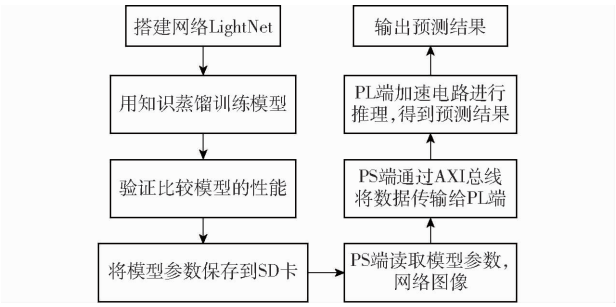

图 2 检测流程图

Fig. 2 Flowchart of system design

并从中挑选了 19 类常见的农作物病害样本。如图 3 所示,包含苹果、葡萄、玉米、马铃薯、番茄等作物的典型病害特征。该数据集共计包含 44 262 幅图像,各病害类别样本量充足且分布均衡,每种病害包含约 2 300 幅图像。为了减小过拟合的风险,并提高模型的准确性及泛化性,本文采用以下方法进行数据预处理:①将图像分辨率统一缩放到 96 像素×96 像素。②随机对部分图像进行翻转,模拟田间图像采集时的拍摄角度差异。③随机对其余图像进行水平或垂直方向的偏移,模拟田间图像采集时的位置偏差。预处理后数据集以比例 8∶2 划分为训练集与测试集,确保模型训练的有效性和评估的客观性。

2.2 LLN 模型设计

考虑到 FPGA 的硬件资源有限,适用于 FPGA 部署的病害识别模型应满足:①具有较高准确度以保证农作物病害的精准识别。②具有较低的参数量与 FLOPs,以减少模型对存储与计算资源的消耗,在进行 FPGA 部署时能有更充足的硬件资源来支持电路优化设计,加快 FPGA 推理速度。

然而神经网络的低复杂度往往会导致网络的准确度下降,针对这一问题,设计了轻量级高精度病害识别网络 LLN。LLN 模型结构及模型中主要的 DSC Block 结构如图 4 所示,首先采用 2 个并行的 DSC Block 完成对病害图像的浅层特征提取,特征拼接后再通过 2 个串行的 DSC Block 进一步完成不同深度的深层特征提取,并且依次插入 3 个最大池化层 Maxpool 来抓取关键特征的同时显著降低了模型的参数。其次,为缓解轻量化网络中特征丢失问题,通过引入多路径跳跃连接结构,实现了浅层特征和深层特征的深度融合,从而提升了模型识别准确度。最后经过全连接层维度映射后输出预测结果。

表 1 给出了本文提出的 LLN 详细结构参数,其中 k、s、p 分别表示卷积核尺寸、步长及填充值。DSC Block 中使用 DSC 代替了普通卷积。DSC 最早由谷歌的 MobileNetV1 提出[19],分为深度卷积和点卷积

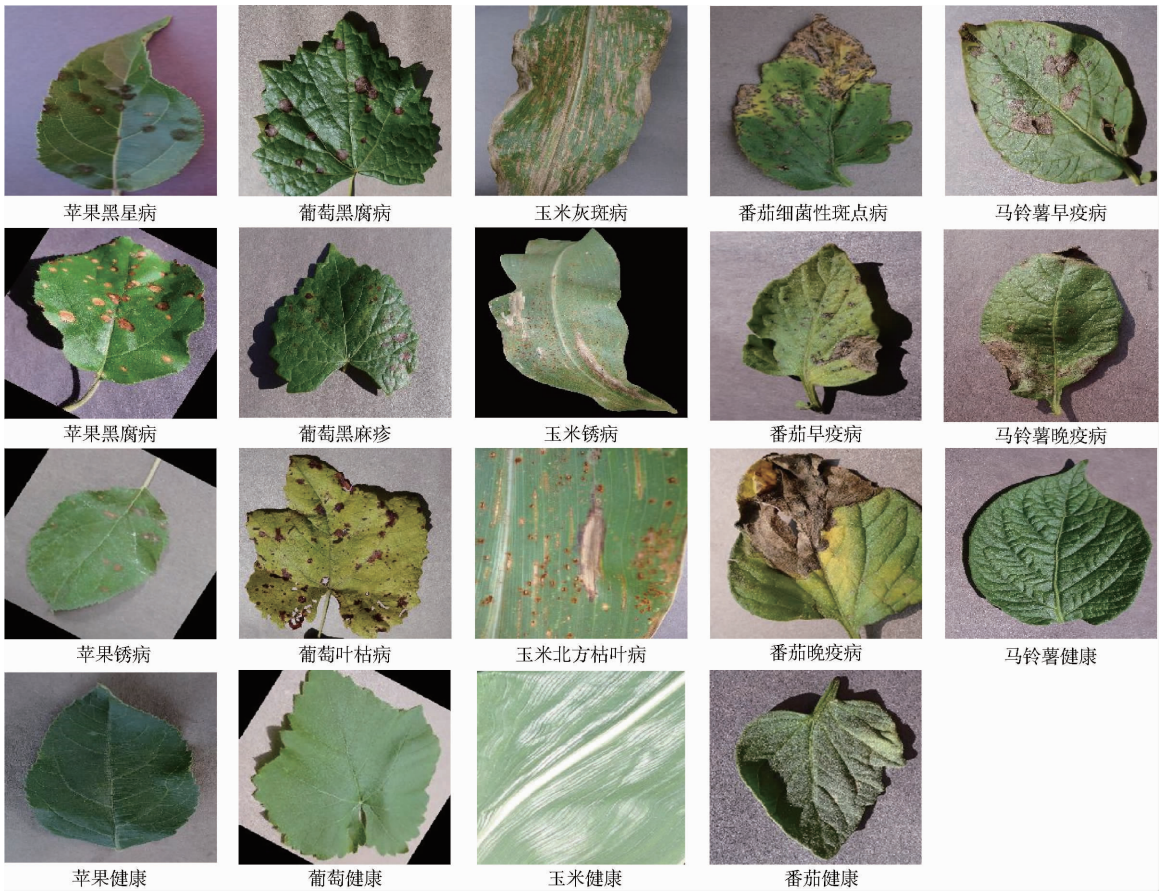

图 3　PlantVillage 数据集植物病害图像示例

Fig. 3　Plant-disease images in PlantVillage dataset

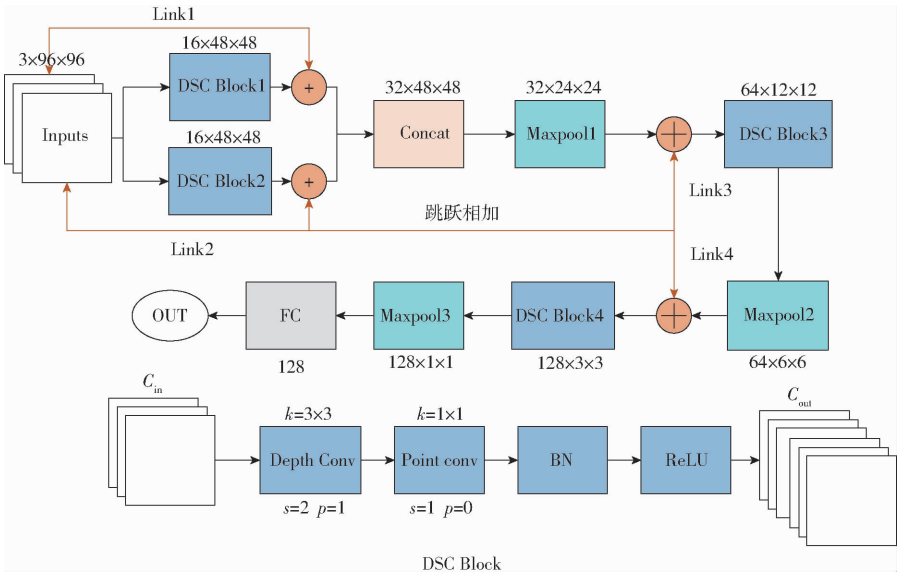

图 4　LLN 模型结构图

Fig. 4　Network structure of LLN

两步，先让每个卷积核独立处理输入特征图的单个通道，再用点卷积组合通道信息，从而大幅减少参数量与 FLOPs。输入数据经过 DSC 后再经过批归一化层与 ReLU 激活层后得到 DSC Block 模块的输出结果。同时，为了实现跳跃连接，LLN 在每个目标卷积块前，先将初始输入特征图通过 1 × 1 的点卷积（对应于表 1 中的 Pwconv $k = 1$）进行维度变换，再通过池化缩放到目标尺寸，最后与目标特征图相加来增强输入，完成跳跃连接，确保浅层特征提取完整，有效提升特征表达能力。综上可知，DSC 模块和跳跃连接的联合应用，使得 LLN 在保证低复杂度的同时，达到了较高识别准确度。此外，本文在训练和

表 1 LLN 模型详细结构参数

Tab. 1 Detailed structural parameters of LLN model

层	操作	输出尺寸
DSC Block1	DSC $k=3,s=2,p=1$	$48^2\times16$
DSC Block2	DSC $k=3,s=2,p=1$	$48^2\times16$
Link1	Pwconv $k=1,s=1,p=0$	$96^2\times16$
	Avgpool $k=2\times2$	$48^2\times16$
	Plus	$48^2\times16$
Link2	Pwconv $k=1,s=1,p=0$	$96^2\times16$
	Avgpool $k=2\times2$	$48^2\times16$
	Plus	$48^2\times16$
Concat	Concat	$48^2\times32$
Maxpool1	Maxpool $k=2\times2$	$24^2\times32$
Link3	Pwconv $k=1,s=1,p=0$	$96^2\times32$
	Avgpool $k=4\times4$	$24^2\times32$
	Plus	$24^2\times32$
DSC Block3	DSC $k=3,s=2,p=1$	$12^2\times64$
Maxpool2	Maxpool $k=2\times2$	$6^2\times64$
Link4	Maxpool $k=4\times4$	$24^2\times3$
	Pwconv $k=1,s=1,p=0$	$24^2\times64$
	Avgpool $k=4\times4$	$6^2\times64$
	Plus	$6^2\times64$
DSC Block4	DSC $k=3,s=2,p=1$	$3^2\times128$
Maxpool3	Maxpool $k=2\times2$	$1^2\times128$
Linear	Dense(128,19)	19

保存网络模型时,采用了知识蒸馏与 BN 层融合的 2 个策略来进一步提升网络识别精度并缩减模型参数量。

2.3 知识蒸馏

为了尽可能地提升识别准确度,本文采用知识蒸馏方法训练模型。知识蒸馏是由 Hinton 等[20]提出的一种压缩模型方法,其思想是将教师网络中的知识通过“蒸馏”的方法传输到学生网络中。本文的教师网络选取 ResNet101[21],而学生网络是提出的 LLN 模型。如图 5 所示,训练过程中教师网络输出软标签,即病害类别的预测概率为

$$q_i=\frac{\exp(v_i)}{\sum_{j=1}^{N}\exp\left(\frac{v_j}{T}\right)}\quad(i=1,2,\cdots,N)\tag{1}$$

式中 N——农作物病害类别数

T——蒸馏温度

v_i、v_j——损失值

q_i——预测概率

学生网络分别输出软、硬标签(病害类别预测值)。学生网络的软标签与教师网络的软标签被用于计算蒸馏损失值 L_{distill}

$$L_{\text{distill}}=\mathrm{KL}(q_{\text{teacher}}\mid q_{\text{student}})\tag{2}$$

$$\mathrm{KL}(P\mid Q)=\sum_x P(x)\ln\frac{P(x)}{Q(x)}\tag{3}$$

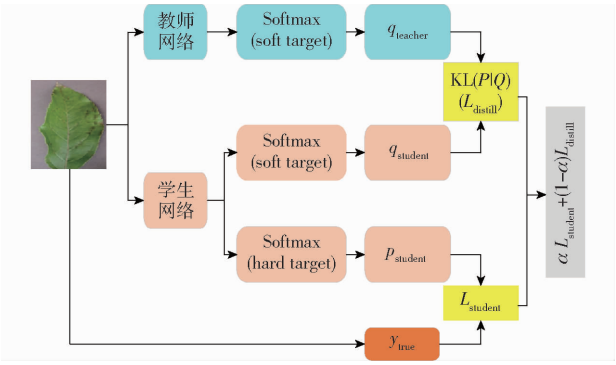

图 5 知识蒸馏示意图

Fig. 5 Knowledge distillation schematic

式中 L_{distill}——蒸馏损失值

q_{student}——学生网络概率分布

q_{teacher}——教师网络概率分布

KL——KL 散度

P、Q——教师、学生模型的软标签

同时采用学生网络的硬标签与真实值计算出学生网络的损失值 L_{student}

$$L_{\text{student}}=H(y_{\text{true}},p_{\text{student}})\tag{4}$$

$$H(P,Q)=-\sum_x P(x)\ln Q(x)\tag{5}$$

式中 L_{student}——学生网络损失值

H——交叉熵损失函数

y_{true}——硬标签

p_{student}——学生模型预测概率

最后将 2 个损失值加权求和获得最终的损失值 L

$$L=\alpha L_{\text{student}}+(1-\alpha)L_{\text{distill}}\tag{6}$$

式中 α——调整两种损失贡献的超参数

综上可知,在知识蒸馏的训练过程中,通过联合应用学生网络本身的损失值和教师与学生网络之间的蒸馏损失值,教师网络 ResNet101 中的知识“蒸馏”到学生网络 LLN 中,优化了 LLN 模型中损失函数的迭代,减少了模型损失,从而提升了模型预测准确度。

2.4 BN 层融合设计

为了提升模型的准确度与泛化能力,通常会在卷积层后添加 BN 层,但 BN 层的加入会带来额外的参数量与 FLOPs。本文采用卷积- BN 层融合技术,融合示意图如图 6 所示。

BN 层融合是将卷积层与 BN 层的计算合并,保留了原 BN 层提升模型性能的优势但不需要花费额外的计算与存储资源。BN 层主要进行归一化运算和线性变化运算,通过逻辑运算将 BN 层与卷积层融合,融合前卷积层和 BN 层的公式分别为

$$\boldsymbol{y}_{\text{conv}}=\boldsymbol{wx}+\boldsymbol{b}\tag{7}$$

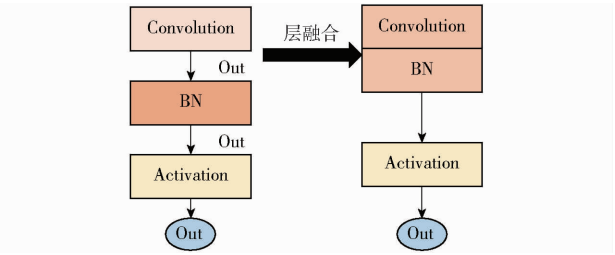

图 6　层融合示意图

Fig. 6　Schematic of layer fusion

$$y_{BN}=\gamma\frac{y_{conv}-\hat{u}}{\sqrt{\sigma^2+\varepsilon}}+\beta \tag{8}$$

式中　y_{conv}——输出向量

x——输入向量

w——卷积层权重参数

b——卷积层偏置参数

y_{BN}——BN 层输出向量

γ——缩放因子

$\hat{u}$——输入均值

σ——输入方差

β——偏移

ε——防止分母为 0 的较小数

融合后总输出为

$$Y=\frac{w\gamma}{\sqrt{\sigma^2+\varepsilon}}x+\frac{\gamma(b-\hat{u})}{\sqrt{\sigma^2+\varepsilon}}+\beta=w'x+b' \tag{9}$$

其中

$$\begin{cases}w'=\dfrac{\gamma}{\sqrt{\sigma^2+\varepsilon}}w\\ b'=\dfrac{\gamma(b-\hat{u})}{\sqrt{\sigma^2+\varepsilon}}+\beta\end{cases} \tag{10}$$

式中　Y——融合后输出向量

w'——融合后的权重参数

b'——融合后的偏置参数

使用融合后的 w' 和 b' 进行卷积运算即可得到原本通过卷积层和 BN 层运算后的输出结果。即实现了融合前后的等价计算。层融合带来的好处有：①融合后的卷积完成融合前两层的效果，减少前向推理的内存访问与浮点运算，降低了推理延迟。②不需要存储 BN 层的参数，减少了存储开销。③设计 FPGA 加速电路时不需要实现 BN 层模块，减轻了硬件资源的负担和电路设计难度。

3　FPGA 部署与优化

3.1　FPGA 架构设计

FPGA 平台上神经网络执行的是网络前向传播部分的计算。本文采用高层次综合工具 Vivado HLS 设计 LLN 的电路，相较于硬件描述语言，该方法允许开发者使用 C/C + + 语言实现网络计算逻辑，并控制数据的输入输出。HLS 工具的综合功能可将这些高级语言代码转换成对应的硬件描述语言（Hardware description language，HDL），极大地提高了开发效率[22]。

本文硬件平台使用的 FPGA 型号为 XC7Z030FFG－2，它的整体架构由 PS 端和 PL 端组成，图 7 给出了 FPGA 加速器架构设计。PS 端接收 SD 卡中存储的网络参数并保存在外部存储 DDR 中；PL 端实现神经网络中各层的计算，二者通过 AXI 总线进行交互。当 FPGA 执行前向传播时，首先由 PS 端 ARM CPU 调用 PL 端神经网络对应层的算子，同时将 DDR 中该层的参数传输到 PL 端，PL 端完成计算后将计算结果回传至 DDR，然后 ARM CPU 再调用下一层的算子，继续执行上述操作直至完成网络的前向传播全过程。由于神经网络的计算几乎都是在 PL 端完成，所以 PL 端网络算子的设计对 FPGA 执行前向推理的速度有着重大影响。

考虑到 FPGA 硬件资源与成本的约束，如果对 LLN 中所有层进行单独设计将造成巨大的资源浪费。为此，本文引入时分复用技术，针对每种类型的计算操作仅设计单个可复用的算子，多次调用预设的算子来实现网络前向传播，从而有效降低硬件资

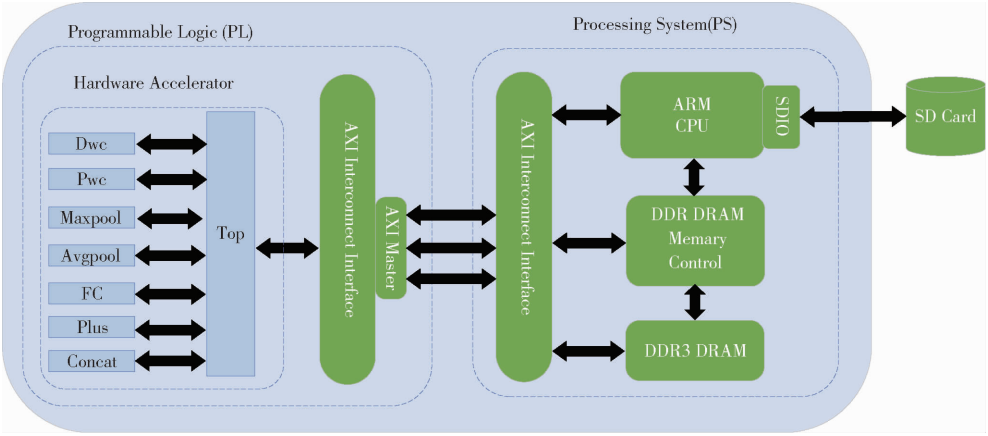

图 7　XC7Z030FFG－2 FPGA 加速器架构

Fig. 7　XC7Z030FFG－2 FPGA accelerator architecture

源消耗。LLN 网络中包括的算子有:深度卷积(Dwc)、点卷积(Pwc)、最大池化(Maxpool)、平均池化(Avgpool)、全连接(FC)、矩阵相加(Plus)及特征图拼接(Concat)。这些算子在 PL 端由顶层模块统一调用,实现了神经网络所需的所有计算功能,PL 端与 PS 端协同操作实现前向传播的完整推理流程。

3.2 模型量化

神经网络通常使用高精度浮点数表示其特征值与权重参数。然而,考虑到 FPGA 片上存储与计算资源有限,直接部署高精度模型会导致硬件资源紧张与运行效率低下。模型量化方法通过使用低位宽的整型数据代替高位宽的浮点型参数,能有效压缩模型的参数规模,缓解 FPGA 部署时的资源压力[23]。

浮点数到定点数的转换公式为

$$v_{\text{fixed16}} = v_{\text{float32}} \times 2^{10} \tag{11}$$

式中 v_{fixed16}——定点数数值

v_{float32}——浮点数数值

为了兼顾模型性能与 FPGA 资源消耗,本文采用 16 位定点数对模型进行量化处理,定点数整数位与小数位安排时应保证其取值范围覆盖模型中所有参数的情况下尽可能地提高精度。提出的 LLN 模型参数最大值不超过 2^5,因此选用高 6 位做整数位,低 10 位做小数位,其表示的整数范围从 -2^5 到 2^5,能表示的小数精度达到了 2^{-10}。图 8 表示 32 位浮点数和 16 位定点数的存储结构。其中 32 位 Float 型浮点数遵循 IEEE754 原则。得益于定点数较高的精度与深度学习良好的鲁棒性,模型量化后的识别准确度不会有明显下降,却能有效地减少硬件资源的开销和访存次数。与 32 位浮点数相比,本文选择的定点数可减少大量的存储与计算资源消耗。

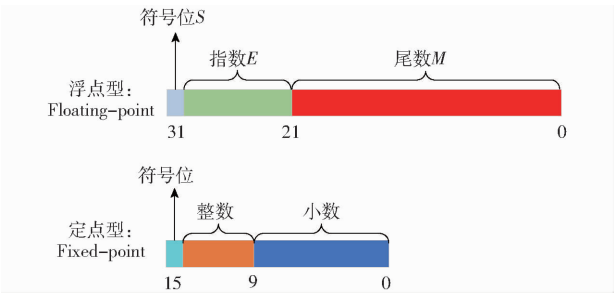

图 8 浮点型和定点型结构示意图

Fig. 8 Schematic of floating-point and fixed-point structures

3.3 特征图分块设计

考虑到 FPGA 片上存储空间有限,难以直接缓存完整的大尺寸输入特征图,因此本文设计 PL 端算子时采用特征图分块策略。如图 9 所示,IFM 是输入为 96 像素 ×96 像素 ×3 通道的特征图,将其分

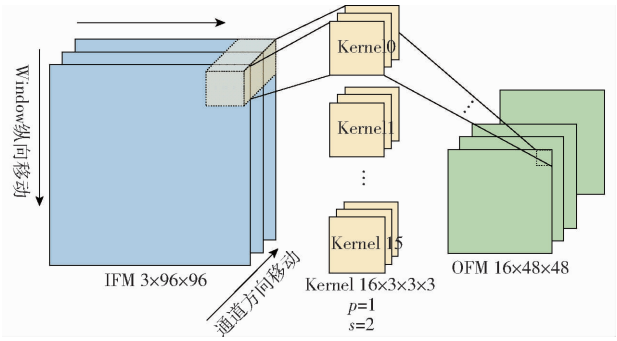

图 9 卷积分块运算示意图

Fig. 9 Schematic of block operation for convolution

割为多个子块分阶段处理,每个子块完成卷积计算后其输出结果将精准映射回原始特征图对应位置,通过多次迭代运算得到完整的输出特征图 OFM,即采用分块计算的策略实现了卷积操作。

该分块策略减少了 FPGA 单次处理的数据规模,有效降低了存储单元与计算资源的瞬时占用,显著缓解硬件资源压力。同时,通过灵活调整分块大小,可动态控制电路的资源消耗与计算周期。较小的分块尺寸能够进一步削减资源开销,但会增加运算迭代次数从而延长推理时间;而较大的分块则可提升计算并行度、缩短推理时长,但对硬件资源要求更高。本文基于硬件平台,针对不同的算子,在硬件资源消耗与推理效率之间进行了有效平衡,设计了高效的 FPGA 电路实现方案。

3.4 FPGA 时延优化策略

在农业智能化应用中,高准确度识别病害仅满足基础需求,识别速度、系统功耗及设备价格均是需要考虑的要素,其中识别速度直接决定了单台设备所能支持的病害图像输入频率和农田面积,进而决定了单位面积农田的病害识别成本,因此是系统评估的重要指标。FPGA 执行前向推理时绝大部分计算都由 PL 端的算子完成,所以算子的设计决定了 FPGA 的推理速度。本文分别从数据高速访存及并行计算两个方面来优化 PL 端算子,有效加快了 FPGA 执行前向传播的速度,减小了识别时延。

3.4.1 双缓冲机制

对于 PL 端算子而言,如果能将所有数据都在 FPGA 片上存储,便可最大限度降低数据传输延迟。但由于 FPGA 板卡存储资源和带宽有限,无法将所有输入数据及中间计算结果都存储在片上缓冲区。所以,数据传输效率成为制约系统性能的关键因素。本文采用了双缓冲机制(乒乓策略),处理流程如图 10 所示。该机制划分了 2 个交替工作的缓冲区,当其中一个缓冲区正在执行数据加载时,另一个缓冲区同步执行算子计算任务。与单缓冲区相比,双缓冲区将数据传输时间与计算时间重叠,有效消除

Single Buffer A

Buffer\Time	Time0	Time1	Time2	Time3	Time4
A	Load ifm0	Compute ifm0	Load ifm1	Compute ifm1	Load ifm2

Double Buffer A&B

Buffer\Time	Time0	Time1	Time2	Time3	Time4
A	Load ifm0	Compute ifm0	Load ifm2	Compute ifm2	Load ifm4
B		Load ifm1	Compute ifm1	Load ifm3	Compute ifm3

图 10　单缓冲与双缓冲机制处理流程图对比

Fig. 10　Comparison of processing flowcharts between single buffer and double buffer mechanisms

加载过程造成的计算等待周期。这一策略充分挖掘了 FPGA 的并行处理能力，显著提升了系统吞吐量。

3.4.2　FPGA 并行度优化

FPGA 芯片拥有丰富的可编程单元，支持并发操作，通过算法优化，实现数据访存、卷积运算中多个算子的并行处理，可以有效加快 FPGA 的前向推理速度。本文主要设计了 3 种优化方法：循环展开优化（UNROLL）、循环流水线优化（PIPELINE）和数组分割优化（ARRAY_PARTITION）。

循环展开优化，通过将循环体复制为多份并行执行单元，显著缩短计算周期。图 11 给出了循环展开示意图。以 LLN 模型中 3×3 卷积运算为例将 3 次迭代的循环展开后，指令被复制为 3 组并行执行，将 3 次循环卷积运算的耗时压缩至单次循环用时，计算效率提升了 2/3。该优化通过资源复用策略，在消耗 3 倍寄存器资源的代价下，有效提升了卷积层的计算效率。

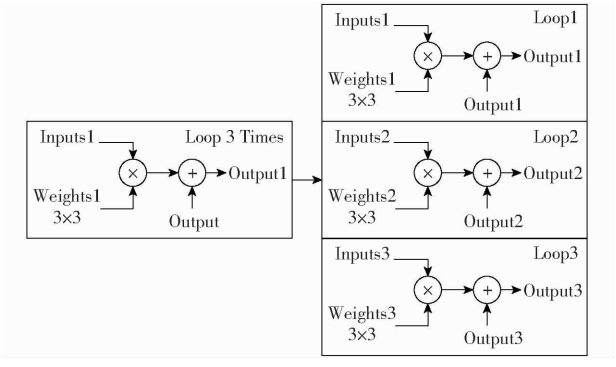

图 11　循环展开示意图

Fig. 11　Unroll schematic

循环流水线优化是指将串行执行的指令重组为并行流水处理，使每个时钟周期均可启动新的计算任务，从而缩短指令周期。图 12 给出了循环流水线优化示意图，针对卷积计算中的多级指令序列，本文采用 4 级流水线架构。与串行执行指令相比，流水线优化后理论吞吐量可提升至优化前的 4 倍。此策略特别适用于 LLN 中密集的 3×3 卷积与逐点卷积操作，有效提升了数据处理的连续性。

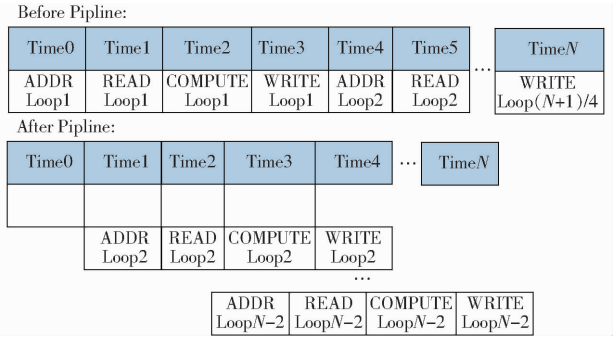

图 12　循环流水线优化前后流程图对比

Fig. 12　Comparison of flowcharts before and after loop pipeline optimization

数组分割优化是对卷积核参数与特征图数据的存储采用块分割策略，将连续存储的多维数组按需求划分为多个小段进行存储。图 13 给出了数据分割示意图，以 LLN 模型中 3×3 的卷积核的存储为例，由于只有 9 个数据，采用完全分割的方式将每个数据独立存储，每个分割后的存储单元可以被不同的处理单元同时访问，从而减少数据访问冲突的概率。该机制与双缓冲机制协同处理极大地降低了片上存储的访问延迟，缩短访存时间，提升系统的吞吐量。

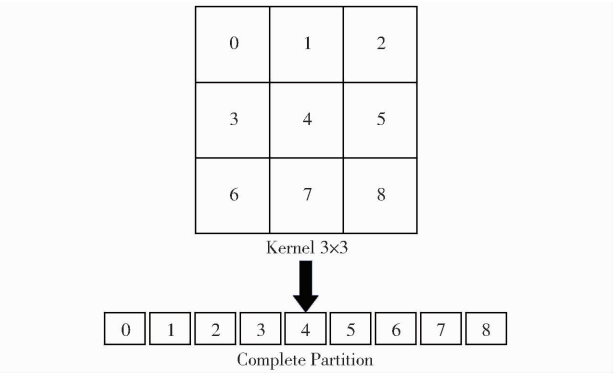

图 13　数组分割示意图

Fig. 13　Schematic of array partition

4　实验

本文的仿真实验软件平台配置为：PyCharm 2023 专业版、深度学习开发框架 PyTorch，使用 Python 3.10 作为开发语言。Vivado HLS 2019.2、Vivado 2019.2 以及 Vitis IDE 2019.2，使用 C++作为开发语言。在超参数中，PyTorch 框架中采用默认参数的 SGD 作为优化器。训练轮次设置为 100 epochs，每个批次大小为 64。初始学习率为 0.01，知识蒸馏的蒸馏温度为 3。硬件平台配置为：AMD Ryzen7 5800H 处理器（主频 3.20 GHz）、单块 NVIDIA GeForce RTX 3070 显卡、32 GB DDR4 内存、1 TB 固态硬盘以及型号为 ZynqMZ7030FA 的 FPGA 板卡。

4.1 病害图像分辨率选择

较高分辨率的输入图像可以提供更多的特征信息,获得更好的识别效果,但是需要更多的 FLOPs,且增加识别时延,而过小分辨率图像会导致特征信息严重丢失从而影响模型判别能力。为探究不同分辨率的病害图像对网络性能及复杂度的影响,本文采用了 4 种不同空间分辨率的输入图像进行对比实验。表 2 给出了不同分辨率的输入图像测试结果。从表 2 可知,随着输入图像分辨率降低,模型 LLN 的 FLOPs 和准确率均会下降。当输入图像分辨率为 224 像素 ×224 像素时,FLOPs 最高达到 1.2×10^7,准确率最高达 98.2%。输入分辨率降至 128 像素 ×128 像素和 96 像素 ×96 像素时,FLOPs 分别降至 3.9×10^6 和 2.2×10^6,准确率仅小幅降低了 0.5 个百分点和 0.6 个百分点。当输入分辨率进一步降至 48 像素 ×48 像素时,FLOPs 虽低至 5×10^5,但准确率大幅下降了 3 个百分点。综上知,96 像素 ×96 像素的图像分辨率在计算效率与精度间实现较好平衡,因此本文选择了 96 像素 ×96 像素分辨率的图像作为模型输入。

表 2 不同分辨率图像输入下知识蒸馏后的测试结果

Tab.2 Test results after knowledge distillation under different resolution image inputs

输入分辨率/(像素 × 像素)	参数量	FLOPs	准确率/%
224 ×224	1.4×10^3	1.2×10^7	98.2
128 ×128	1.4×10^3	3.9×10^6	97.7
96 ×96	1.4×10^3	2.2×10^6	97.6
48 ×48	1.4×10^3	5×10^5	95.2

4.2 LLN 模型评估

4.2.1 PlantVillage 数据集上的模型评估

本研究在 PlantVillage 数据集上对 LLN 网络与现有方法进行对比,所有模型均完成 100 轮训练,表 3 给出了不同模型在参数量、FLOPs 和分类准确率 3 个维度的对比结果。由表 3 可知,以 DenseNet 和 ResNet 为代表的深层网络同时具有高精度与高复杂度的特点,分别凭借密集连接和残差连接的网络架构准确率达到了 99.4% 和 99.5%,但参数量分别高达 1.81×10^7、4.25×10^7,FLOPs 超过 4.3×10^9。尽管这 2 种网络病害识别性能极佳,但是庞大的参数量与 FLOPs 使其难以部署在 FPGA 平台上。与深层网络相比,轻量化网络 MobileNetV3 与 ShuffleNet 虽然准确率稍低一些,但是参数量与 FLOPs 缩减到了不足深层网络的 1/10。然而 MobileNetV3 和 ShuffleNet 在进行 FPGA 部署时,仍面临电路较为庞大的问题,低配置的 FPGA 会导致识别时延过长,无法满足农作物病害识别高帧率的需求,而高配置的 FPGA 成本和功耗都会相应增加。因此,实用农作物病害识别需要设计一种兼具低复杂度和高准确度的神经网络模型,以支持低成本 FPGA 部署和快速精准病害识别。

表 3 PlantVillage 数据集上不同网络模型效果对比

Tab.3 Comparison of performance of different network models on PlantVillage dataset

模型	参数量	FLOPs	准确率/%
DenseNet201[24]	1.81×10^7	4.3×10^9	99.4
ResNet101[25]	4.25×10^7	7.86×10^9	99.5
MobileNetV3[26]	2.68×10^6	2.71×10^8	98.6
ShuffleNet[27]	1.8×10^6	3.28×10^8	99.2
LiteCNN[14]	2.4×10^3	3.2×10^6	92.8
LiteCNN_kd[14]	2.4×10^3	3.2×10^6	94.8
LLN (Proposed)	1.4×10^3	2.2×10^6	95.4
LLN_kd (Proposed)	1.4×10^3	2.2×10^6	97.6

注:_kd 表示采用了知识蒸馏方法。

本文设计的 LLN 模型参数量低至 1.4×10^3,FLOPs 为 2.2×10^6,十分有利于低成本模型部署,当采用知识蒸馏方法训练得到 LLN_kd 后,参数量与 FLOPs 保持不变,但准确率从 95.4% 提升至 97.6%。这是因为蒸馏损失函数促进了有效参数信息的传递,因此提升了模型的识别准确率。类似的性能提升也存在于 LiteCNN 模型采用知识蒸馏前后。尽管 LiteCNN 也兼顾了模型复杂度与识别性能,但本文提出的 LLN 模型,识别性能比前者提升了 2.8 个百分点,且参数量与 FLOPs 分别降低了 42% 和 31%。这主要归因于 LLN 模型中跳跃连接结构多次将浅层特征向量融合入深层网络,既加速了收敛过程,又减少了关键特征在传播中的损耗,有效解决了轻量化模型中常见的特征表达能力不足问题。后文提到的 LLN 模型均指蒸馏后的模型。

即使与识别性能优越的轻量化模型 ShuffleNet 相比,本文提出模型的参数量与 FLOPs 大幅缩减至其的 0.08% 左右,但分类准确率仅降低了 1.6 百分点。综合考虑轻量化部署与识别准确率,LLN 模型在保持极低复杂度的同时达到了 97.6% 的高准确率,在农田环境下的农作物病害实时高精度检测方面具有显著优势。

4.2.2 模型在不同数据集上的效果分析

本文所采用的 PlantVillage 数据集,其图像样本多采集于实验室环境,单一化的背景会降低模型的识别任务难度。为进一步验证本文所提模型在真实田间复杂场景下的识别性能与泛化能力,本研究选取田间实地拍摄的 Paddydoctor 数据集开展补充对比实验。通过旋转、平移等策略对 Paddydoctor 数据

集进行数据增强处理，将 10 000 幅原始图像扩展至 50 000 幅，然后按照比例 8∶2将增强后的数据集划分为训练集与验证集，并基于该划分方案完成模型训练及测试。

实验结果如表 4 所示，在背景更真实、环境更复杂的 Paddydoctor 数据集上，各对比模型的识别准确率均出现不同程度的下降。其中，本文提出的 LLN 模型仍保持 94% 的较高准确率，与轻量级网络 MobileNetV3 持平，略低于 ShuffleNet，但 LLN 模型的复杂度仅为后两者的 1% 左右。与 LLN 复杂度较为接近的 LiteCNN 模型在新数据集上识别准确率下降幅度略小，但其 92.7% 的准确率在所有对比模型中仍处于最低水平。由此可见，在复杂场景的新数据集上，本文提出的 LLN 模型仍保持了较高的识别准确率，且在轻量级网络中表现突出，具备良好的泛化能力。

表 4　2 个数据集上不同网络的识别准确率对比

Tab. 4　Comparison of recognition accuracy of different networks across two datasets　%

网络	PlantVillage	Paddydoctor
LLN_kd	97.6	94.0
LiteCNN_kd	94.8	92.7
ResNet101	99.5	97.0
DenseNet201	99.4	96.7
MobileNetV3	98.6	94.0
ShuffleNet	99.2	95.4

4.3　模型量化对 LLN 电路的影响

为了研究模型量化对 LLN 模型加速性能的影响，本文选取了单精度浮点数 float(32 位)和定点型 ap_fixed(16 位)2 种数据类型进行对比，实验结果如表 5 所示。由表 5 可看出，与 float 类型相比，采用 ap_fixed 类型消耗的各项硬件资源显著减少，其中查找表 LUT 占用从 81% 下降到 59%，有效缓解了硬件部署的资源压力，而模型的识别准确率从 97.6% 略微下降至 97.2%，精度损失基本可忽略。实验结果表明，通过采用更低位宽的定点数代替高位宽的浮点数，模型量化有效减少了网络的参数量与 FLOPs，进而减少了 FPGA 的片上存储与计算资源开销。定点数虽然位数少，但在精准分配整数位与小数位时仍能保持较高的精度，配合神经网络良好的鲁棒性，模型识别准确率几乎不损失。因此，模型量化是提升 FPGA 部署可行性与检测效率的有效方法。

4.4　电路优化前后性能对比

为了提升病害检测速度需要最大程度地利用 FPGA 的硬件资源。本文采用了第 3 节提出的双缓冲机制和多种并行度优化策略，并行度优化具体包括了循环展开、循环流水线与数组分割优化。优化前后的资源用量及单幅图像的平均识别速度如表 6 所示。其中基础电路是仅采用特征图分块和模型量化后的电路，这是因为本研究使用的 FPGA 平台属于低成本芯片，未采用特征图分块时，模型所消耗的硬件资源超过了芯片所能提供的最大资源，无法部署。

表 5　不同数据类型下 LLN 电路性能

Tab. 5　Performance of LLN circuits under different data types

类型	BRAM		LUT		FF		DSP		准确率/%
	数量/块	占比/%	数量/块	占比/%	数量/块	占比/%	数量/块	占比/%	
float	96	36.2	64 328	81	72 461	46	249	62	97.6
ap_fixed	69.5	26	46 729	59	46 979	30	163	41	97.2

表 6　优化前后电路性能对比

Tab. 6　Comparison of circuit performance before and after optimization

优化方式	BRAM		LUT		FF		DSP		单帧图像平均
	数量/块	占比/%	数量/块	占比/%	数量/块	占比/%	数量/块	占比/%	推理时间/ms
基础电路	53.5	20	23 867	30.3	24 819	15.8	93	23.3	118.1
加入双缓冲	55	20.7	24 243	30.8	24 947	15.8	93	23.3	95.9
双缓冲合并并行度优化	69.5	26.2	46 729	59.5	46 979	30	163	40.8	55.7

由表 6 可知，与基础电路相比，加入双缓冲机制后，BRAM 和 LUT 这 2 种硬件资源使用量有小幅上升，这是由于双缓冲机制开辟的新缓冲区需要额外 BRAM 与逻辑控制。因为双缓冲机制有效提升了数据加载效率，电路的推理时间从 118.1 ms 降低到了 95.9 ms。

进一步应用多种并行度优化方案后，各项硬件资源消耗均有显著提高，其中循环展开与循环流水线优化均提高了 FPGA 运行的并行度与瞬时计算量，从而需要更多的逻辑控制与乘法器，因此主要导致 LUT 和 DSP 使用量上升，其中 LUT 占比由双缓冲时的 30.8% 上升至 59.5%，DSP 占比则由 23.3%

上升至 40.8%,二者消耗量几乎翻倍。数组分割优化将连续的数组分开存储到独立的区域中,因此主要带来存储资源消耗的增长。BRAM 的使用量由双缓冲时的 55 块(20.7%)上升到 69.5 块(26.2%)。3 种并行度优化方案在实现过程中均需要触发器 FF 来存储中间结果,共使得 FF 用量从双缓冲时的 15.8% 上升至 30%。电路并行度的提高虽然加剧了各项硬件资源的使用量,但是大幅缩减了电路的推理时间,从加入双缓冲机制时的 95.9 ms 减少到全面优化后的 55.7 ms。以上推理时间均为 FPGA 加速器检测 8 852 幅测试集图像的平均时间。

图 14 展示了采用双缓冲机制与多种并行度优化后 FPGA 加速器的识别速度测试结果,平均推理时间为 55.7 ms,与基础电路的 118.1 ms 相比,全面优化后单帧推理时间减少了 52.8%。实验表明,通过提升内存访问效率与电路并行度,优化后的电路更加充分地利用了 FPGA 的并行操作优势,在合理增加资源占用的同时有效提升了识别速度。值得注意的是,本文设计的电路仍有进一步优化空间,当识别任务升级需要部署更复杂的神经网络时,可以延用或升级本文设计的算子,且提出的访存和并行度优化方案也可以被推广应用,通过在存储、计算资源和推理时间进行折中优化,支撑更复杂网络的部署。

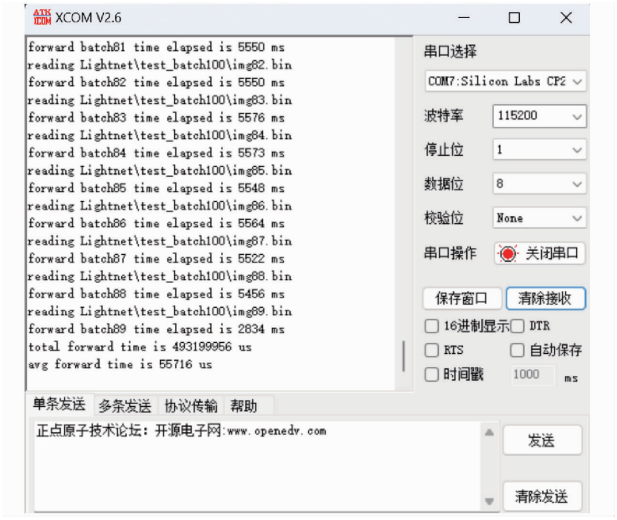

图 14 优化后电路识别速度测试截图

Fig. 14 Screenshot of optimized circuit recognition speed test

4.5 不同硬件平台的实测性能对比

为了综合评估基于 FPGA 加速的 LLN 模型的检测性能,本研究在不同的硬件平台上,分别从准确率、速度、功耗以及成本 4 个维度对 LLN 模型的表现进行对比分析,表 7 给出了测试结果。由表 7 可知,与 CPU 平台相比,FPGA 部署 LLN 的准确率降低了 0.4 个百分点,这是模型量化导致的精度损失,但是 FPGA 的单帧图像推理时间达到 55.7 ms,少于 CPU 的 84.8 ms,同时功耗降低至 3.78 W,仅仅为 CPU 功耗的 31.2%,且 FPGA 成本也远低于 CPU。CPU + GPU 的配置准确率与 CPU 相同并且识别时间最短,仅为 5 ms,虽然其具有优秀的性能,但 61.81 W 的功耗(达到 FPGA 功耗的 16 倍)与高达 9 000 元的价格使其难以实现农田环境下的大面积部署。

得益于跳跃连接和并行分支的使用,LLN 模型获得了比 LiteCNN 更高的识别准确率,并且 LLN 具有更少的参数量和 FLOPs。在模型部署时,因为本文提出的硬件加速方法采用了双缓冲及多种并行优化策略,从而大幅提高了电路的并行度,这使得单位时间内电路的存储和计算增加,因此与 LiteCNN 电路相比,LLN 电路功耗更高,但单帧图像推理时间大大减少。推理速度的加快意味着单张板卡可支持的病害监测面积更大,进而在相同农田面积下减少了所需板卡的数量,达到节约硬件成本的效果。实际应用中可兼顾病害识别速度需求和硬件平台资源,灵活调整电路并行度、实现推理速度和功耗的折中。综上所述,本文设计的基于 FPGA 加速的 LLN 模型在准确率、识别速度、功耗与成本 4 个维度的综合性能优秀,是部署在田间环境下用于农作物病害识别的理想方案。

表 7 不同硬件平台上实测性能对比

Tab. 7 Performance comparison of field tests on different hardware platforms

设备	准确率/%	单帧图像推理时间/ms	功耗/W	成本/元
FPGA(LLN)	97.2	55.7	3.78	1 045
CPU	97.6	84.8	12.10	2 484
CPU + GPU	97.6	5.0	61.81	9 000
FPGA(LiteCNN)	94.8	63.8	2.55	1 045

注:FPGA 为 Zynq MZ7030FA 的 FPGA 开发板;CPU 为 AMD Ryzen7 5800H 处理器(3.2 GHz),系统内存 32 GB,固态硬盘 1 TB;GPU 为 NVIDIA GeForce RTX 3070 显卡(8 GB 显存);CPU + GPU 指包含 CPU、GPU、内存和硬盘等硬件的完整设备。

5 结束语

针对农田环境下植物病害难以快速准确识别的问题,本文采用 FPGA 加速神经网络解决了病害识别网络在边缘设备部署的难题。首先设计了具有较高识别准确度的轻量级网络 LLN,其在 PlantVillage 数据集上达到 97.6%,在 Paddydoctor 数据集上的准确率达到了 94%,具有良好的泛化能力,并极大程度压缩了网络的参数量和 FLOPs,确保了模型可部

署在低成本 FPGA 上。其次，在型号为 MZ7030FA 的 FPGA 开发板上设计并优化 LLN 加速电路，通过采用模型量化、特征图分块及时延优化多个策略，FPGA 平台的单帧图像推理时间达到 55.7 ms，功耗低至 3.78 W。与 CPU + GPU 平台识别相比，本文开发的 FPGA 识别系统虽然检测速度略慢，但是在田间供电受限、设备体积及经济成本约束下，具有巨大的低功耗、低成本及便携性优势。对比同类型 FPGA 加速方案，本文提出的电路优化加速方案，达到了更高的准确率与更快的识别速度。此外，针对不同的网络模型，可以参考本文设计的算子及电路优化方案，完成快速部署。综上所述，本研究工作能为农业病害人工智能识别方法提供参考方案，对农业智能化、自动化具有重要意义。

参 考 文 献

[1] 康振生. 我国植物真菌病害的研究现状及发展策略[J]. 植物保护，2010，36(3)：9 – 12.
Kang Zhensheng. Current research status and development strategies of plant fungal diseases in China[J]. Plant Protection, 2010, 36(3): 9 – 12. (in Chinese)

[2] 杨涛，李晓晓. 机器视觉技术在现代农业生产中的研究进展[J]. 中国农机化学报，2021，42(3)：171 – 181.
Yang Tao, Li Xiaoxiao. Research progress of machine vision technology in modern agricultural production[J]. Journal of Chinese Agricultural Mechanization, 2021, 42(3): 171 – 181. (in Chinese)

[3] 王翔宇，温皓杰，李鑫星，等. 农业主要病害检测与预警技术研究进展分析[J]. 农业机械学报，2016，47(9)：266 – 277.
Wang Xiangyu, Wen Haojie, Li Xinxing, et al. Research progress analysis of detection and early warning technologies for major agricultural diseases[J]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(9): 266 – 277. (in Chinese)

[4] 李振波，刘欣，竺明昊，等. 面向种植农业复杂场景的 YOLO 算法应用综述[J]. 农业机械学报，2026，57(13)：45 – 67.
Li Zhenbo, Liu Xin, Zhu Minghao, et al. Survey of YOLO algorithm applications for complex scenarios in crop agriculture[J]. Transactions of the Chinese Society for Agricultural Machinery, 2026, 57(13): 45 – 67. (in Chinese)

[5] 秦丰，刘东霞，孙炳达，等. 基于深度学习和支持向量机的 4 种苜蓿叶部病害图像识别[J]. 中国农业大学学报，2017，22(7)：123 – 133.
Qin Feng, Liu Dongxia, Sun Bingda, et al. Image recognition of four alfalfa leaf diseases based on deep learning and support vector machine[J]. Journal of China Agricultural University, 2017, 22(7): 123 – 133. (in Chinese)

[6] 姜敏，沈一鸣，张敬尧，等. 基于深度学习的水稻病虫害诊断方法研究[J]. 洛阳理工学院学报(自然科学版)，2019，29(4)：78 – 83.
Jiang Min, Shen Yiming, Zhang Jingyao, et al. Research on rice pest and disease diagnosis method based on deep learning[J]. Journal of Luoyang Institute of Science and Technology (Natural Science Edition), 2019, 29(4): 78 – 83. (in Chinese)

[7] 邵峻青，魏霖静. 基于深度学习的农业病害图像识别的研究[J]. 计算机与数字工程，2024，52(10)：3131 – 3135，3178.
Shao Junqing, Wei Linjing. Research on agricultural disease image recognition based on deep learning[J]. Computer and Digital Engineering, 2024, 52(10): 3131 – 3135, 3178. (in Chinese)

[8] 高发瑞，古华宁，张巧玲，等. 基于农业大数据和深度学习的稻田杂草识别[J]. 江苏农业科学，2024，52(18)：215 – 221.
Gao Farui, Gu Huaning, Zhang Qiaoling, et al. Weed recognition in rice fields based on agricultural big data and deep learning[J]. Jiangsu Agricultural Sciences, 2024, 52(18): 215 – 221. (in Chinese)

[9] Cheng J, Wang P S, Li G, et al. Recent advances in efficient computation of deep convolutional neural networks[J]. Frontiers of Information Technology & Electronic Engineering, 2018, 19: 64 – 77.

[10] 周莹. FPGA 在人工智能深度学习中的应用[J]. 数字技术与应用，2022，40(10)：143 – 145，224.
Zhou Ying. Applications of FPGA in artificial intelligence and deep learning[J]. Digital Technology and Application, 2022, 40(10): 143 – 145, 224. (in Chinese)

[11] Rupanagudi S R, Ranjani B S, Nagaraj P, et al. A novel cloud computing based smart farming system for early detection of borer insects in tomatoes[C] // 2015 International Conference on Communication, Information & Computing Technology (ICCICT). IEEE, 2015: 1 – 6.

[12] Koutayni A, Rybalkin V, Malik J, et al. Real-time energy efficient hand pose estimation: a case study[J]. Sensors, 2020, 20(10): 2828.

[13] Ma Y, Cao Y, Vrudhula S, et al. Automatic compilation of diverse CNNs onto high- performance FPGA accelerators[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2018, 39(2): 424 – 437.

[14] Luo Y, Cai X, Qi J, et al. FPGA-accelerated CNN for real-time plant disease identification[J]. Computers and Electronics in Agriculture, 2023, 207: 107715.

[15] 郭冬冬，蔡祥，齐建东，等. 基于 FPGA 加速 CNN 的低成本芒果外观品质检测装置[J]. 农业工程学报，2024，40(21)：137 – 145.
Guo Dongdong, Cai Xiang, Qi Jiandong, et al. A low-cost mango appearance quality detection device based on FPGA-accelerated CNN[J]. Transactions of the CSAE, 2024, 40(21): 137 – 145. (in Chinese)

[16] 张军阳，王慧丽，郭阳，等. 深度学习相关研究综述[J]. 计算机应用研究，2018，35(7)：1921 – 1928，1936.
Zhang Junyang, Wang Huili, Guo Yang, et al. A review of research related to deep learning[J]. Application Research of

Computers, 2018, 35(7): 1921 - 1928,1936. (in Chinese)
[17] 吴艳霞,梁楷,刘颖,等.深度学习 FPGA 加速器的进展与趋势[J].计算机学报,2019,42(11):2461 - 2480.
Wu Yanxia, Liang Kai, Liu Ying, et al. Progress and trends of deep learning FPGA accelerators[J]. Chinese Journal of Computers, 2019, 42(11): 2461 - 2480. (in Chinese)
[18] 段瑞枫,陈艳,洪凯,等.基于 PCIe 级联网口的农业监测视频高速传输系统研究[J].农业机械学报,2024,55(3):203 - 212.
Duan Ruifeng, Chen Yan, Hong Kai, et al. Research on high-speed transmission system for agricultural monitoring video based on PCIe-level network interface[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(3): 203 - 212. (in Chinese)
[19] Howard A G, Zhu M, Chen B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications[PP/OL]. V1. arXiv(2017 - 04 - 17)[2026 - 09 - 20]. https://doi.org/10.48550/arXiv.1704.04861.
[20] Hinton G E, Vinyals O, Dean J. Distilling the knowledge in a neural network[PP/OL]. V1. arXiv (2015 - 03 - 09)[2026 - 09 - 20]. https://doi.org/10.48550/arXiv.1503.02531.
[21] 卢冶,陈瑶,李涛,等.面向边缘计算的嵌入式 FPGA 卷积神经网络构建方法[J].计算机研究与发展,2018,55(3):551 - 562.
Lu Ye, Chen Yao, Li Tao, et al. A method for constructing embedded fpga convolutional neural networks for edge computing [J]. Journal of Computer Research and Development, 2018, 55(3): 551 - 562. (in Chinese)
[22] 高杨,曹仰杰,段鹏松.神经网络模型轻量化方法综述[J].计算机科学,2024,51(增刊 1):23 - 33.
Gao Yang, Cao Yangjie, Duan Pengsong. A review of neural network model lightweight methods[J]. Computer Science, 2024, 51(S1): 23 - 33. (in Chinese)
[23] 余子健,马德,严晓浪,等.基于 FPGA 的卷积神经网络加速器[J].计算机工程,2017,43(1):109 - 114,119.
Yu Zijian, Ma De, Yan Xiaolang, et al. FPGA-based convolutional neural network accelerator[J]. Computer Engineering, 2017, 43(1): 109 - 114,119. (in Chinese)
[24] Huang G, Liu Z, Van DerMaaten L, et al. Densely connected convolutional networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700 - 4708.
[25] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770 - 778.
[26] 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.
[27] Zhang X, Zhou X, Lin M, et al. An extremely efficient convolutional neural network for mobile devices[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018:6848 - 6856.

(上接第 150 页)

[14] 赵小霞, 李志强. 基于 PLC 和机器视觉的水果自动分级系统研究[J]. 农机化研究, 2021, 43(8): 75 - 79.
Zhao Xiaoxia, Li Zhiqiang. Research on fruit automatic grading system based on PLC and machine vision[J]. Journal of Agricultural Mechanization Research, 2021, 43(8): 75 - 79. (in Chinese)
[15] Sofu M M, Er O, Kayacan M C, et al. Design of an automatic apple sorting system using machine vision[J]. Computers and Electronics in Agriculture, 2016, 127: 395 - 405.
[16] 熊玮, 周礼辰, 余豪, 等. 薄皮果蔬分选技术与设备研究进展与展望[J]. 农业机械学报, 2025, 56(8): 665 - 683.
Xiong Wei, Zhou Lichen, Yu Hao, et al. Review of sorting technologies and equipment for thin-skinned fruits and vegetables [J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(8): 665 - 683. (in Chinese)
[17] Lu Y Z, Lu R F. Using composite sinusoidal patterns in structured-illumination reflectance imaging (SIRI) for enhanced detection of apple bruise[J]. Journal of Food Engineering, 2017, 199: 54 - 64.
[18] Liu X H, Zhang Z, Yu J F, et al. Design, integration, and evaluation of a low-cost system for automatic apple picking and infield sorting[J]. Computers and Electronics in Agriculture, 2025, 239: 110933.
[19] Kumar C S, Subeesh A, Dubey K, et al. Development of an optimally designed real-time automatic citrus fruit grading-sorting machine leveraging computer vision-based adaptive deep learning model [J]. Engineering Applications of Artificial Intelligence, 2023, 120: 105826.
[20] Pothula A K, Zhang Z, Lu R F. Evaluation of a new apple in-field sorting system for fruit singulation, rotation and imaging [J]. Computers and Electronics in Agriculture, 2023, 208: 107789.
[21] Mizushima A, Lu R F. Cost benefits analysis of in-field presorting for the apple industry[J]. Applied Engineering in Agriculture, 2011, 27(1): 33 - 40.
[22] Zhang Z, Pothula A K, Lu R F. Development and preliminary evaluation of a new bin filler for apple harvesting and in-field sorting machine[J]. Transactions of the ASABE, 2017, 60(6): 1839 - 1849.
[23] Flemmer C, Bakker H, Flemmer R. Analysis of the stochastic excursions of tumbling apples[J]. Computers and Electronics in Agriculture, 2021, 188: 106362.