

基于 FPGA 加速 CNN 的低成本芒果外观品质检测装置

郭冬冬¹, 蔡祥^{1,2*}, 齐建东^{1,2}, 罗钊轩¹, 高培文¹, 朱山³

(1. 北京林业大学信息学院, 北京 100083; 2. 国家林业和草原局林业智能信息处理工程技术研究中心, 北京 100083;
3. 湖北华夏水利水电股份有限公司, 荆州 434000)

摘要: 针对卷积神经网络 (CNN) 算法对硬件性能要求高, 难以在成本低、性能受限的边缘设备上部署实现的问题, 该研究综合考虑芒果外观品质检测的准确性, 速度, 功耗, 成本等因素, 设计开发了基于现场可编程门阵列 (FPGA) 加速 CNN 的品质分级检测装置。首先, 设计了一种基于 MobileNetV2 的结构简单且高效的轻量级网络 (Compact MobileNet, CMNet), 通过压缩网络结构降低模型参数数量和计算量, 保证其在准确率可接受前提下, 适合在边缘设备部署。其次, 为了加快 CMNet 在成本和性能受限边缘设备上的执行速度, 通过 BN 层融合和模型量化方法进一步减少模型的存储需求和计算量, 同时设计实现了基于 FPGA 加速 CMNet 网络的硬件电路, 并使用高层次综合 (high-level synthesis, HLS) 优化方法对加速电路进行并行优化。最后, 基于 FPGA 开发板 Zynq Z7-Lite 7020, 设计开发了芒果外观品质分级检测装置, 装置集成 OV5640 摄像头, CMNet 网络加速电路和 HDMI 显示器。在实验室环境下, 将芒果外观品质依据中国芒果农业行业标准 (NY/T 492-2002) 分为 3 个等级, 利用 Dosehri 芒果数据集对检测装置进行效果验证。结果显示本研究的芒果外观品质检测装置准确率达到 96%, 检测速度为每帧 0.072 s, 功耗为 2.6 W。表明该研究实现了一种低成本、低功耗、准确率高和速度快的芒果外观品质分级检测装置, 能够进行芒果品质的实时动态分级检测。

关键词: 卷积神经网络; 芒果; 外观; 品质检测; FPGA 加速; 实时

doi: 10.11975/j.issn.1002-6819.202404087

中图分类号: S24

文献标志码: A

文章编号: 1002-6819(2024)-21-0137-09

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. doi: 10.11975/j.issn.1002-6819.202404087 <http://www.tcsae.org>
GUO Dongdong, CAI Xiang, QI Jiandong, et al. Low-cost grading device for mango appearance quality using accelerated CNN with FPGA[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(21): 137-145. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202404087 <http://www.tcsae.org>

0 引言

芒果品质分级对于实现芒果最大商业价值至关重要, 外观品质是其品质分级的重要指标, 消费者大多通过芒果的外观来判断其品质。当前中国芒果产业的智能化水平较低, 品质分级主要依靠人力。而人工分级存在成本高、效率低且精度受限等问题^[1]。利用计算机视觉技术辅助实现芒果外观品质分级具有重要意义, 可以减少人工成本, 提高效率和准确性^[2]。

用于水果品质检测的传统机器学习方法主要包括 K 最近邻^[3]、支持向量机^[4]和随机森林^[5]等。这些算法理论成熟, 且经过验证, 但存在准确率较低和对环境变化的鲁棒性较差等问题^[6]。随着计算机和人工智能技术的不断进步, 卷积神经网络 (convolutional neural network, CNN) 在果蔬外观品质分级检测应用中也备受关注^[6-7]。COSTA 等^[8]基于 ResNet 网络模型进行番茄外观品质检测, 在测试集上的最优平均精度为 94.6%。陈锋军等^[9]

在 YOLOv7 模型中引入交叉注意力机制以及基于距离和交并比的非极大值抑制算法, 用来检测油茶果成熟度, 达到了 94.85% 的检测准确率。倪建功等^[10]采用知识蒸馏方法训练 CNN 模型进行胡萝卜外观品质等级检测, 使用复杂教师网络指导精简学生网络 Resnet18 训练, 将学生网络的检测准确率提高至 95.8%。张思雨等^[11]提出一种自适应卷积神经网络检测花生质量, 将二次函数与正态分布模型引入网络损失中, 同时从网络的损失与权重方面提出自适应学习率, 对花生常见缺陷实现了 99.7% 的平均识别率。由此可见, 基于 CNN 方法的果蔬外观品质检测具有很高的准确性。但基于 CNN 算法对水果外观品质检测的研究大多集中在改进 CNN 算法以提高检测的准确性, 而对基于 CNN 算法进行水果品质检测的实时性研究以及算法在终端的部署研究则少有报道。

尽管 CNN 算法能够提高水果外观品质检测的准确性, 但该方法具有计算密集和存储密集特性, 通常要求在高性能计算服务器上执行^[12]。而高性能计算服务器价格昂贵, 只有规模较大的企业能够承受其价格成本, 对于普通一线果农, 在生产资料匮乏的现场部署深度学习方法非常困难。文献 [13] 总结了水果外观品质检测系统被广泛应用需要满足的要求: 1) 高精度, 准确性是保证水果品质检测和等级划分的基本要求; 2) 低成本, 较低成本

收稿日期: 2024-04-13 修订日期: 2024-09-27

基金项目: 国家自然科学基金资助项目 (31400621)

作者简介: 郭冬冬, 硕士, 研究方向为智能农业装备研究、FPGA 硬件加速人工智能。Email: 1335617043@qq.com

*通信作者: 蔡祥, 博士, 副教授, 研究方向为智能农业装备研究、FPGA 硬件加速人工智能。Email: caixiang617@163.com

更容易推广；3) 低功耗，低功耗不仅降低系统在使用过程中的成本，而且适合边缘部署；4) 高效性，在保证准确率的前提下保证系统的识别速度。CNN 解决了分级检测的准确性问题，其他问题仍亟待解决，如果能在低成本的嵌入式边缘设备上部署 CNN，将有利于基于 CNN 的品质检测系统推广应用。然而在低成本、资源受限的边缘设备上部署 CNN 网络，即便是轻量级网络，也难以达到系统运行的实时性要求。因此对 CNN 网络进行加速很有必要。GPU 是目前主流的深度学习加速器，具有高并行、高带宽和高速浮点计算能力。但由于 GPU 体积较大、功耗高且价格昂贵，难以边缘部署^[14]。近年来，FPGA 成为加速 CNN 的研究热点，其功耗低，体积小，适用于嵌入式装置，且其硬件功能可定制或修改以适应不同需求^[15]。

综合以上分析，针对在低成本、性能受限的边缘设备上，基于 CNN 的芒果外观品质分级检测难以实现的问题，综合考虑检测速度，准确性，功耗，成本等因素，本文首先基于 MobileNetV2 网络模型，设计 8 层轻量级网络 CMNet (Compact MobileNet)，在确保较高准确率前提下，最大程度降低模型计算量和参数量。其次，使用 BN 层融合和模型量化方法进一步精简网络。同时设计加速 CMNet 网络的 FPGA 硬件电路，使用 HLS 优化方法对硬件电路进行并行优化，加快 CMNet 网络的执行速度。最后，基于 FPGA 开发板 Zynq Z7-Lite 7020，设计开发了芒果外观品质分级检测装置，装置融合 OV5640 模块摄像头驱动电路，CMNet 网络加速电路，和 HDMI 显示驱动电路，实现了芒果图像采集，实时品质检测和结果输出显示功能。

1 材料和方法

1.1 检测装置系统架构设计

芒果外观品质检测装置的整体架构如图 1 所示，主要由 SoC 开发板、摄像头以及 HDMI 显示器组成。开发板采用 Zynq Z7-Lite 7020 作为控制主板，板上使用 ZYNQ7020 作为核心 FPGA 控制芯片，该芯片由主频 1 GHz 的双 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。PL 端主要功能包括芒果图像采集和显示的驱动，以及 CMNet 网络的硬件加速。图像采集采用 OV5640 CMOS 摄像头，通过 SCCB (Serial Camera Control Bus) 协议设置寄存器以控制摄像头。采集图像通过 AXI VDMA 接口电路传输至 PS 端，缓存在 DDR 内存中，后续作为 CMNet 网络处理计算的输入数据。CMNet 的网络参数保存在 SD 卡中，PS 端的 ARM 系统启动后从 SD 卡中读出参数到 DDR 内存，供 CMNet 后续使用。PS 端的 ARM 用于初始化配置摄像头，控制图像数据和网络参数向 PL 端的 CMNet 加速电路加载，控制 CMNet 的执行过程，以及控制将芒果品质检测结果显示在 HDMI 显示器中。

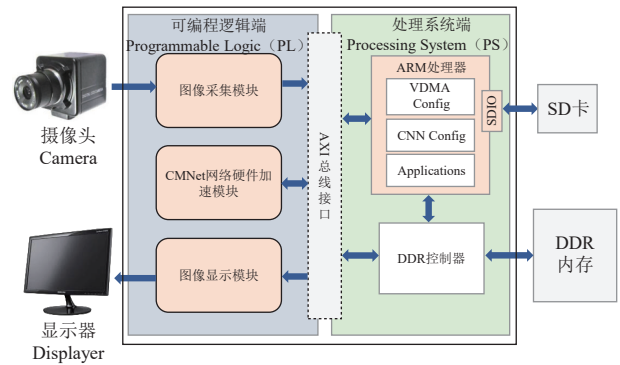


图 1 检测装置系统架构图

Fig.1 System architecture diagram

装置的工作流程如图 2 所示，上电初始化后，ARM 从 SD 卡中将 CMNet 模型参数读至 DRAM 中。启动 PL 端硬件电路，初始化并配置摄像头和 VDMA 模块，开始采集 RGB 图像，并以 DMA 方式将图像数据发送至 DRAM。PS 端 ARM 处理器对图像进行缩放和归一化预处理，再由 ARM 控制将 CMNet 模型参数及图像数据发送至 PL 端 CMNet 网络的加速电路进行计算，计算结果由 AXI 总线传回 PS 端，由 ARM 控制将检测结果叠加采集图像并显示在 HDMI 显示屏。

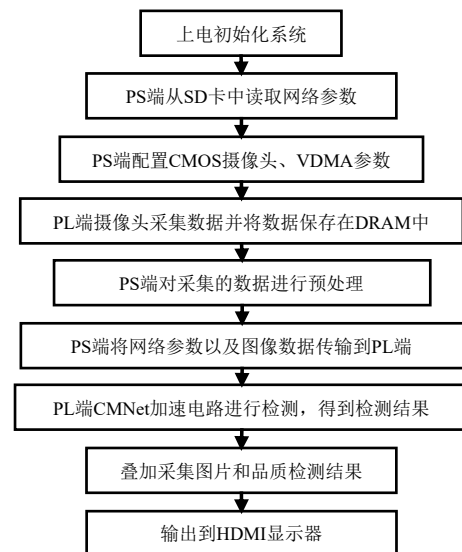


图 2 装置工作流程

Fig.2 Workflow of the device

1.2 网络的设计与优化

1.2.1 CMNet 模型设计

MobileNetV1^[16] 是谷歌提出的轻量级 CNN 网络，它结构简单，运行速度快，使用深度可分离卷积替代普通卷积，降低了网络的参数量和计算量。普通卷积和深度可分离卷积区别如图 3 所示，深度可分离卷积包含逐通道卷积和逐点卷积，其中逐通道卷积将输入通道拆分，使用 1 通道卷积核对不同通道特征进行卷积操作，卷积核数量与输入特征通道数、输出特征通道数相同，此操作不改变特征通道数。逐点卷积将输入特征进行 1×1 卷积，输出特征通道数与 1×1 卷积核个数相等，通过调整卷积核个数调整输出特征的通道数。MobileNetV2^[17] 在

MobileNetV1 基础上, 引入倒残差网络 Inverted Residual 和线性瓶颈结构 Linear Bottlenecks, 缓解 MobilenetV1 网络存在特征/神经元退化的潜在问题, 在精度、模型参数量和计算时间之间更好地取得平衡。

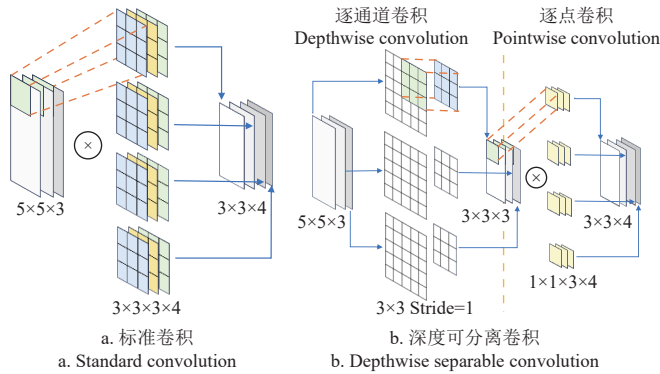


图 3 标准卷积与深度可分离卷积

Fig.3 Standard convolution and depthwise separable convolution

在保证模型具有较高芒果品质检测准确率的前提下, 模型复杂度要设计得尽可能低, 以确保模型的计算复杂度和存储复杂度在常规 FPGA 上能够实现加速。基于此, 本文在 MobileNetV2 基础上, 进一步压缩网络, 设计用于检测芒果品质的 CMNet 模型, 主要改动包括: 1) 降低网络深度, 仅保留 3 个 bottleneck 层; 2) 精简 bottleneck 结构, 统一 bottleneck 中的卷积步长, 全部修改为 2; 3) 修改 MobileNetV2 网络最后的全连接层, 由原来的一层改为两层, 防止特征维度在降维过程中跳跃过大, 造成信息丢失。CMNet 网络的具体结构如表 1 所示。

表 1 网络结构

Table 1 Network structure

输入尺寸 Input size	操作 Operator	输出通道数 Numbers of output channels	步长 Stride
$96^2 \times 3$	conv2 d,3×3	16	2
$48^2 \times 16$	bottleneck,3×3	16	2
$24^2 \times 16$	bottleneck,3×3	24	2
$12^2 \times 24$	bottleneck,3×3	24	2
$6^2 \times 24$	conv2 d,1×1	64	1
$6^2 \times 64$	pool	—	1
$1^2 \times 64$	conv2 d,1×1	32	1
$1^2 \times 32$	conv2 d,1×1	3	1

1.2.2 数据集及预处理

在实验室环境下训练 CMNet 模型, 所用数据为从网络下载的 Dosehri 芒果外观图像数据。依据中国芒果农业行业标准 (NY/T 492-2002) [18], 将芒果外观品质分为特等品、一等品和二等品 3 个等级。图 4 给出了每种品质等级芒果的示例图像。



图 4 不同外观品质等级的芒果图像样本示例

Fig.4 Mango examples with various appearance quality

为了减小模型过拟合风险, 提高模型准确性和泛化能力, 通过以下 4 种方式进行数据集增强: 1) 随机旋转图像角度以模拟拍摄的角度变化; 2) 横向或纵向随机幅度偏移以模拟图像的位置偏移; 3) 随机增强或减弱图像亮度以模拟不同的光照强度; 4) 添加随机噪声以模拟恶劣环境下的拍摄噪声。数据增强后将数据集以 5: 2: 3 比例进行随机划分, 得到训练集 827 张, 验证集 331 张和测试集 497 张。表 2 给出了芒果外观品质分级等级特征以及训练用数据集的详细信息。

表 2 芒果外观品质数据集信息

Table 2 Dataset information for mango appearance quality detection

等级 Quality grade	特征 Feature	图像数量 / 张 Number of images	
		增强前 Before argumentation	增强后 After argumentation
特等品 Premium-grade	外观无畸形, 大小均匀, 具有该品种的典型色泽, 完全均匀着色。	159	557
一等品 First-grade	外观无明显变形, 具有该品种典型色泽, 75% 以上果面均匀着色。	198	572
二等品 Second-grade	允许有不影响产品品质的果形变化, 具有该品种典型色泽, 35% 以上果面均匀着色。	143	526

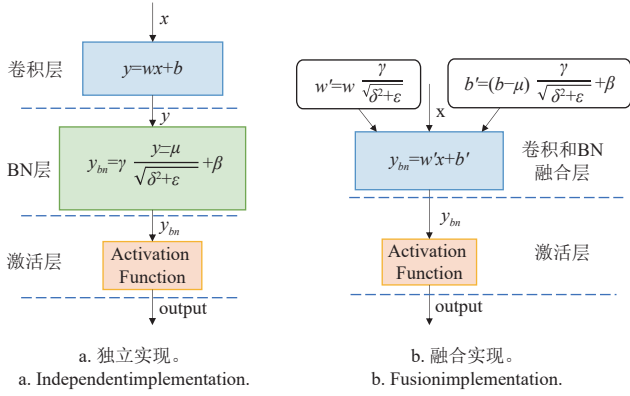
考虑到模型输入图像大小和模型计算量和计算时间正相关, 数据预处理操作还包括将图像剪裁至 96×96 像素大小, 96×96 像素的尺寸大小是根据后续的试验结果确定。

1.2.3 BN (batch normalization) 层融合

批量归一化有助于加速模型收敛和提升模型泛化能力。BN 层的运算如图 5a 所示, 上层模块表示 Conv 层, 中间模块表示 BN 层, 通过 Conv 层和 BN 层的计算方式可以看出, Conv 层与 BN 层均属于线性变换。在设计 BN 层功能电路时考虑到: 1) 单独为 BN 层设计电路会增加读取和存储特征数据的过程, 以及增加 BN 层计算的时间; 2) 单独 BN 层计算电路要消耗额外的 FPGA 硬件资源。BN 层融合是指将 BN 层的线性变换融合进 Conv 层的线性变换中, 仅用 Conv 层电路同时实现卷积和归一化功能, 图 5b 展示了 BN 层融合的原理。在设计加速电路时, BN 层融合的好处显而易见, 减少了计算时间并节约了片上资源。

1.3 芒果品质分级检测装置电路设计

基于 ZYNQ7020 核心控制芯片的 “FPGA+ARM” 架构, 针对系统功能进行软硬件协同设计, 合理规划软硬件功能, 可有效提高芒果品质分级检测的实时性。将计算量大和图像处理相关的操作均部署在 PL 端 FPGA 中, 以硬件电路形式实现, 主要包括芒果图像采集与显示电路, 以及 CMNet 网络加速电路。电路的设计与实现采用 FPGA 高抽象级别工具 HLS, 与 FPGA 硬件描述语言不同, HLS 使用 C/C++语法规则编写代码, 开发过程只关心算法本身, 屏蔽了电路实现细节, 可有效提高开发效率 [19]。



注: x , 输入特征向量; y , 输出特征向量; w , 卷积层权重参数; b , 卷积层偏置; y_{bn} , BN 层输出向量; μ , 输入均值; δ^2 , 输入方差; γ , 缩放因子; β , 偏移; ϵ , 防止分母为 0 的较小数; w' , 融合层权重参数; b' , 融合层偏置参数

Note: x , input feature; y , output feature; w , weight; b , bias; y_{bn} , output feature of BN layer; μ , mean; δ^2 , variance; γ , scale factor; β , shift factor; ϵ , a small constant added for numerical stability; y_{bn} , output feature of fused layer; w' , weight of fused layer; b' , bias of fused layer.

图 5 BN 层不同实现方式

Fig.5 Different implementation methods of BN layer

1.3.1 图像采集与显示电路

图 6 展示的是图像采集及显示电路结构, 主要包括: 1) OV5640 摄像头接口电路, 封装成 OV5640_capture IP 核, 用于获取原始的图像像素数据, 并将数据转为通用并行视频信号; 2) 输入视频信号格式转换电路, 使用 Video In to AXI4-Stream IP 核将通用并行视频信号转换为 AXI4-Stream 接口信号, 以便使用 AXI 总线传输视频图像; 3) 视频数据传输电路, 使用 VDMA IP 核将视频数据基于 AXI 总线以 DMA 方式传输至 PS 端的 DDR 内存。同时设置两个帧缓存空间, 交替存储帧数据, 防止处理图像数据时, 当前图像被覆盖; 4) 图像叠加电路, 使用 Video Mixer IP 核将采集的芒果图像和芒果品质等级检测结果进行叠加, 输出带有检测结果的芒果图像; 5) 输出视频信号格式转换电路, 使用 AXI4-Stream to Video Out IP 核将 AXI4-Stream 格式的视频数据流转换为

视频输出信号; 6) HDMI 接口电路, 封装成 HDMI_trans IP 核, 用于生成 HDMI 视频信号所需的时序信号、控制信号和数据信号, 驱动 HDMI 显示器。其中电路模块 1~3 用于实时采集芒果图像, 模块 4~6 用于结果显示。

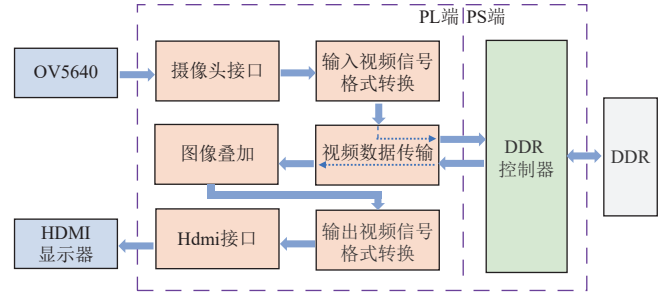


图 6 图像采集与显示电路结构

Fig.6 Image acquisition and result display circuit structure

1.3.2 CMNet 网络加速电路设计

CMNet 网络主要由卷积层、Bottleneck 结构和全连接层构成, 根据层次结构特性, 基于以下两个原因将 CMNet 的加速电路分模块设计: 1) 网络结构每层均会产生中间特征数据, 模块化电路设计可以不在 FPGA 上存储中间特征数据, 将其传输至 DDR 上暂存, 减小对 FPGA 存储资源的要求; 2) 模块化设计电路更有利于调试开发以及神经网络规模的扩展。最终, 将 CMNet 加速电路划分为 4 个电路子模块, 分别为 1) 普通卷积 Conv 电路; 2) 逐通道卷积 Depthwise Conv 电路 DW; 3) 逐点卷积 Pointwise Conv 电路 PW; 4) 池化层电路 Pool。

1.3.3 芒果品质分级检测装置电路结构

将各功能子电路封装成 IP 核, 在 Vivado 工具中将各功能电路的 IP 核进行互联, 得到如图 7 所示的分级检测装置电路结构图。其中 ZYNQ7 Processing System (ARM 处理器) IP 核通过 4 条 AXI HP 总线与各电路 IP 核相连, 用于 PS 端处理器、存储器和 PL 端电路的高吞吐量数据交互。

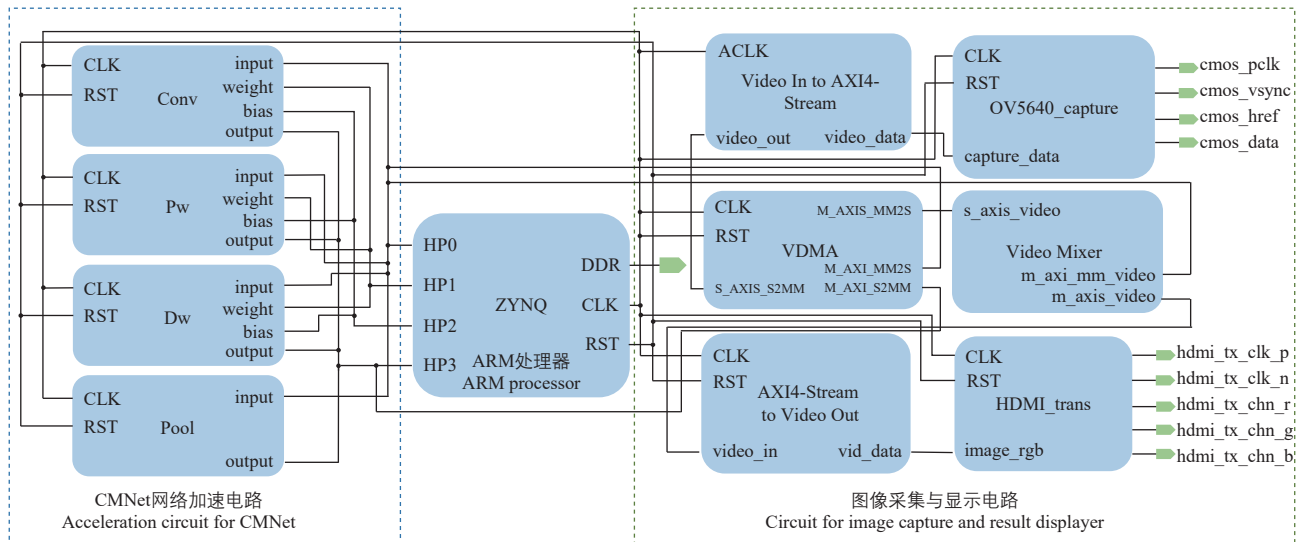


图 7 芒果品质分级检测装置电路结构图

Fig.7 Circuit structure diagram of mango quality detection and grading system

1.4 模型量化

深度学习模型的特征向量和网络参数通常使用高精度浮点型数据表示，数据存储、传输和计算消耗大量硬件资源。模型量化是指以损失小部分模型准确性为代价，将模型输入特征向量和参数使用低精度数据表示，从而降低对存储带宽和计算资源的需求，减小存储访问延迟。尽管模型量化会降低计算精度，但神经网络模型在实施精度上有很强鲁棒性，即使数据表示精度有所降低，模型准确率仍保持在可接受范围^[20]。图 8 展示了浮点型和 16 位定点型数据的存储结构。基于模型的参数取值范围，本研究使用的定点数的整数部分占 6 位，小数部分占 10 位，式 (1) 展示了 32 位浮点数转 16 位定点数（小数位占 10 位）的计算方法，其中， v_{fixed16} 表示定点数大小， v_{float32} 表示浮点数数值。显而易见，采用 16 位定点数表示特征向量和网络参数，存储效率提高了 1 倍，能同时减少访存次数，降低系统能耗和资源开销。

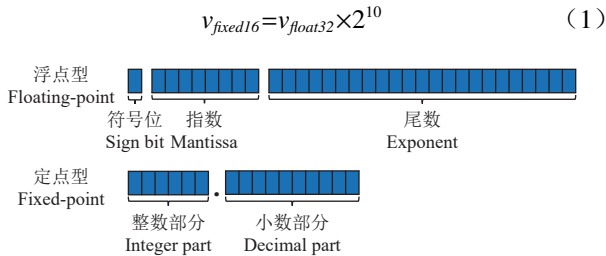


图 8 浮点型和定点型数据存储结构

Fig.8 Storage structure of floating-point and fixed-point data type

1.5 CMNet 网络加速电路并行优化

HLS 显著提高了 FPGA 电路设计效率，但生成的电路在资源占用和执行时间方面存在较大优化空间。CMNet 网络加速电路的各电路模块主要是由嵌套的 for 循环实现，为了进一步提升电路性能，对电路进行并行优化，主要优化方法包括循环展开、循环流水以及数组分割。循环展开和循环流水方法是针对 for 循环进行的优化，其中循环展开将循环体展开多份执行，循环体内指令对应的电路被复制相应的份数，多份电路并行执行，和单一 for 循环相比，减少执行时间。循环流水将循环内顺序执行的指令转换成流水线方式处理，无数据冲突的数据读取，计算和存储操作可同时执行，从而提高数据的吞吐量，加快执行速度。数组分割将数据拆分存储至多个 RAM 中，以多个小空间代替一个大存储空间，减小数据访问冲突概率，提升读写效率及系统吞吐率。这些 HLS 优化方法使用 FPGA 资源面积换取计算时间，能大幅提升卷积运算的计算效率，减少品质检查的计算时延。

2 结果与分析

2.1 检测装置

实验室环境下，基于 FPGA 加速 CNN 的低成本芒果外观品质分级检测装置如图 9 所示，摄像头采集左侧屏幕中的图像，开发板中的 CMNet 网络加速电路检测芒

果品质，右侧屏幕通过 HDMI 线与开发板相连接输出检测结果，整个系统的 FPGA 硬件资源使用情况如表 3 所示。

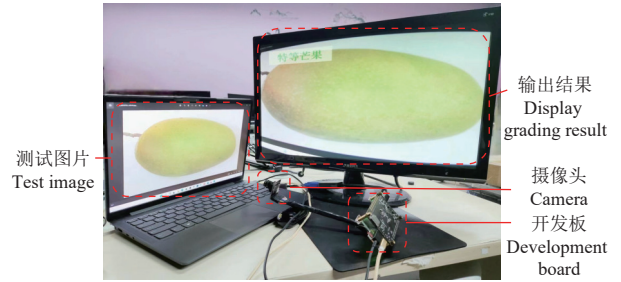


图 9 芒果外观品质分级检测装置

Fig.9 Device for mango appearance quality grading detection

表 3 FPGA 硬件资源使用情况统计

Table 3 Statistics on FPGA hardware resource usage

硬件资源 Resource	使用量 Utilization	可用量 Available	使用率 Utilization rate/%
查找表 Look up table, LUT	30 584	53 200	57.49
寄存器 Flip flop, FF	39 183	106 400	36.83
块 RAM Block random access memory, BRAM	20	140	14.29
数字信号处理器 Digital signal processor, DSP	107	220	48.64

2.2 CMNet 模型评估

将 CMNet 和常见经典网络模型进行对比，从模型大小、参数量、浮点运算次数（floating point operations, FLOPs）及准确率这几方面评估 CMNet 的整体性能，对比结果如表 4 所示。可以看出所有网络均取得了很好的芒果品质检测准确率，其中 AlexNet 网络的准确率最高，达到了 99.0%，CMNet 网络准确率略低于经典网络模型，为 96.4%，仅比 AlexNet 低 2.6 个百分点。尽管如此，CMNet 的网络规模却明显小于经典网络，其参数量（0.015M）以及计算量（0.007 2 G）仅为 AlexNet 的 0.06% 和 1%。由此可见，在保证较高芒果品质检测准确率前提下，规模小的 CMNet 可以更好的在资源受限的 FPGA 上实现。

表 4 网络模型效果对比

Table 4 Comparison of network model effects

模型 Model	模型大小 Model size/kB	参数量 Total parameters/M	计算量 Calculation/G	精度 Accuracy/%
GoogleNet ^[21]	24 058	6.1	1.58	98.8
ResNet-50 ^[22]	9 430	23.6	3.64	98.6
AlexNet ^[23]	97 032	24.9	0.72	99.0
MobileNetV2	9 500	2.3	0.33	98.4
CMNet	65	0.015	0.007 2	96.4

2.3 CMNet 识别效果

CMNet 在测试数据集的分级结果如图 10 所示，混淆矩阵展示了 CMNet 网络对于不同品质等级芒果图片的判断情况。数值表示预测为某种等级的数量，大数值主要集中在混淆矩阵的主对角线，意味着 CMNet 对不同外观品质等级的芒果均可做出正确判断。

图 11 展示了 CMNet 对芒果品质等级分类错误的图片实例。具体来说：1) 图 11a 中上面是标注为特等品质的输入芒果图片，被错误预测为一等品质芒果，下面为

一等品质芒果的典型图片。不难看出,输入图片中芒果存在果柄疤痕特征,与一等品质芒果示例图的局部少量疤痕特征类似,模型难以辨识,导致出现误判。这与训练数据中携带果柄疤痕的特等品质芒果照片数据不充分有关,模型未学习到芒果的果柄特征;2)图 11b 中上面是标注为一等品质芒果,被错误分级为二等品质,下面为二等品质芒果的典型图片。推测模型检测错误的主要原因是输入图片芒果表面存在反光部分,反光部分像素被模型判断为不属于该品质芒果的典型色泽,被误判为瑕疵。这种类型的错误分级主要是芒果反光的训练样本图片较少,模型没有学习到反光特征所导致;3)图 11c 中上面是标注为一等品质芒果,被错误分级为特等品质,下面为特等品质芒果的典型图片。输入图片因存在细小瑕疵被标注为一等品质芒果,但由于 CMNet 网络在检测之前将图片尺寸缩放至 96×96 像素,导致图像中瑕疵特征信息丢失,将一等品质芒果错误分级为特等品质。

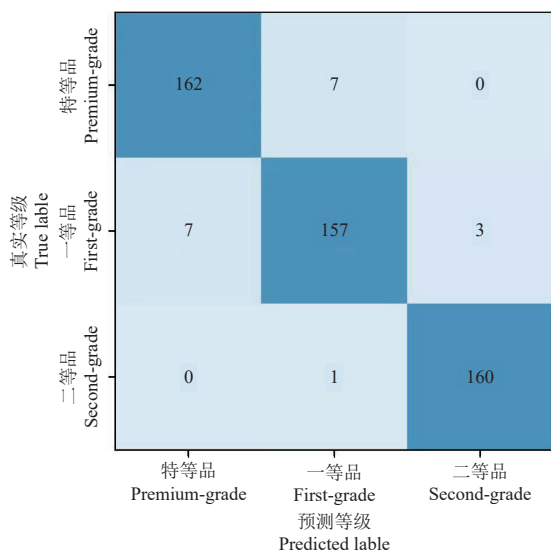


图 10 分级结果混淆矩阵

Fig.10 Classification result confusion matrix

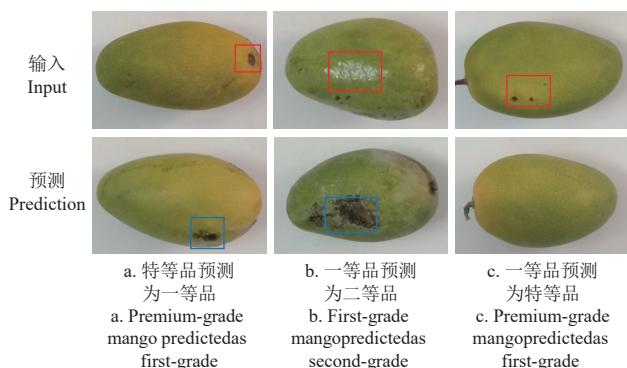


图 11 错误分级示例

Fig.11 Mis-classification Examples

结合 CMNet 网络 96.4% 的分级准确率以及对分级错误芒果实例的观察与分析,不难看出 CMNet 的芒果品质分级检测效果较好,CMNet 检测错误受多种因素影响,主要原因是训练模型的样本数据不够充分,模型对影响芒果品质判断的特殊特征学习不充分。

2.4 不同尺寸图像输入对检测模型的影响

为了研究网络模型的最佳图像输入尺寸,比较了 4 种不同尺寸图像对 CMNet 的影响。结果如表 5 所示,可以看出,保持模型规模大小不变情况下,随着芒果图像尺寸减小,模型的计算量和准确率都随之下降。当输入图像尺寸为 224×224 像素时,模型的计算量和准确率均达到最大,分别为 0.039 2 G 和 98.4%。当图像尺寸减小到 96×96 像素时,网络模型的计算量大幅降低,与尺寸为 224×224 像素时相比,减少了 0.032 G。而芒果品质检测的准确率依然保持在 96% 以上,只下降了 2 个百分点。继续减小图像尺寸至 48×48 像素,模型的准确率明显下降至 90.5%。由此可见,较大尺寸的输入可以提供更多的特征信息,有助于提高检测的准确性,但需要更多的执行时间,较小的输入准确性有所降低,但检测的实时性更好。因此,在确保准确率的情况下,本文将芒果图片尺寸调整到 96×96 像素,确保芒果品质检测的高准确性和较好的实时性。

表 5 不同输入尺寸对网络影响

Table 5 The impact of various input sizes on the network

输入尺寸 Input size	参数量 Parameters/M	计算量 Computation/G	精度 Accuracy/%	速度 Speed/(s·帧 ⁻¹)
$224 \times 224 \times 3$	0.0158	0.0392	98.4	0.349
$128 \times 128 \times 3$	0.0158	0.0128	96.8	0.115
$96 \times 96 \times 3$	0.0158	0.0072	96.4	0.072
$48 \times 48 \times 3$	0.0158	0.0018	90.5	0.029

2.5 模型量化对 CMNet 网络加速电路的影响

为研究模型量化对 CMNet 网络硬件加速电路的影响,对单精度浮点数 float (长 32 bits), 半精度浮点数 half (长 16 bits) 和 16 位 ap_fixed 定点型 3 种数据类型进行实验对比,表 6 展示了 3 种数据类型的资源消耗情况以及模型准确率,不难看出, half 和 ap_fixed 类型的准确率相同,达到了 96%, 仅比 float 类型的准确率低 0.4 个百分点。在资源消耗上, ap_fixed 型占用的资源最少, half 型次之, float 型占用资源最多,且远多于前两种。由此可见,使用 ap_fixed 定点型数据设计加速电路可以保证模型较高的准确率,同时节省大量 FPGA 硬件资源,节省的资源可以用来探索更多电路并行优化措施。

表 6 不同数据类型对电路的影响

Table 6 The impact of various data types on circuits

数据类型 Data type	BRAM/块	LUT/个	FF/个	DSP/个	精度 Accuracy /%
ap_fixed	20.0	30 584	39 183	107	96.0
half	21.3	31 708	45 275	178	96.0
float	28.5	39 416	56 367	213	96.4

注: 1 块 BRAM 的大小为 18K。

Note: 1 BRAM is with 18 K memory.

2.6 电路优化前后对比

为最大程度利用 Zynq Z7-Lite 7 020 开发板 FPGA 片上资源,加快网络的检测速度,使用循环展开、循环流水以及数组分割方法对硬件电路进行了优化。表 7 为优化前后使用的资源情况以及检测电路的时延 Latency。优化后,电路的资源消耗有所增加,但时延明显下降,优化后时延仅为优化前的 28.6%。对比优化前后的资源消

耗和时延，不难看出，并行优化消耗了更多电路资源，换取电路时延的大幅降低，并行优化可有效提高资源的利用率，加快 CMNet 网络检测芒果品质的速度。

表 7 电路优化前后时延和资源消耗对比

优化 Optimization	Latency/s	LUT/个	BRAM/块	FF/个	DSP/个
优化前 Before optimization	0.252	22 216	18	32 747	93
优化后 After optimization	0.072	30 584	20	39 183	107

2.7 与其他计算设备的综合对比

为了综合评价本文设计的基于 FPGA 加速 CMNet 的芒果外观品质分级检测装置的性能，从芒果分级准确性、速度、功耗和成本方面，将其与计算设备 CPU，GPU，树莓派，android 智能手机进行对比，具体结果见表 8。其中 GPU 搭载在 CPU2 上，构成“CPU+GPU”图形工作站，各设备具体型号详见表注。

表 8 不同设备实现分级网络效果对比

Table 8 Comparison of graded network effects achieved by various devices				
设备 Equipment	精度 Accuracy/%	速度 Speed/(s·帧 ⁻¹)	功耗 Power/W	价格 Price/元
FPGA	96.0	0.072	2.6	790
CPU1	96.4	0.12	15.0	2 198
CPU2+GPU	96.4	0.011	95.0	7 000
Raspberry Pi	96.4	0.74	4.0	300
Android	96.4	0.02	8.0	3 999

注：FPGA 为 Zynq Z7-Lite 7 020 开发板；CPU1 为 Intel(R) Core(TM) i5-6200U 2.40 GHz；；CPU2 为 Intel Xeon Bronze 3 106 1.7 GHz×16 系统内存为 64GB；GPU 为 NVIDIA GeForce GTX 1080Ti；树莓派为 3 代 B 型，soc 为 Broadcom BCM2837；Android 设备为三星 S22。
Note: FPGA is ZYNQ Z7-Lite 7 020 developing board, CPU1 is Intel Core i5-6200U 2.40 GHz; CPU2 is Intel Xeon Bronze 3 106 1.7 GHz × 16; GPU is NVIDIA GeForce GTX 1080Ti; Raspberry Pi is 3B (Broadcom BCM2837 SoC); Android is Galaxy S22.

在准确率方面，FPGA 识别的准确率为 96.0%，其余设备均为 96.4%，这是由于所有设备实现的网络均为 CMNet，FPGA 采用 16 位定点型数据，精度略有损失，其余实现均采用 float 型数据。在速度方面，“CPU2+GPU”的推理时间最短，为 0.011 s，往后依次为 Android（0.02 s），FPGA（0.072 s），CPU（0.12 s）和树莓派（0.74 s）。需要说明的是，“CPU2+GPU”的 0.011 s 为批量检测 500 张芒果图片的平均速度。输入单张图片，“CPU2+GPU”的检测时间和 CPU1 的时间大致相当。在功耗方面，“CPU2+ GPU”功耗最大，为 95.0 W，往后依次为 CPU1 15.0 W，安卓设备 8.0 W，树莓派 4.0 W，FPGA 功耗最小，为 2.6 W。功耗数据为品质检测时的瞬时功率，其中 CPU 和树莓派的功率采用 HWiNFO 软件测得，GPU 功率使用 nvidia-smi 程序测得，Android 功率使用 PerfDog 软件测得，FPGA 开发板功耗由功率计直接测得。在价格方面，“CPU2+ GPU”最高，往后依次为安卓设备，CPU1，FPGA 和树莓派，要说明的是 CPU、GPU 仅为芯片价格，其他为设备价格。不难看出，FPGA 与 CPU、GPU 相比，品质检测准确率相当，功耗最低，且体积更小，便携性和移动性更强。FPGA 和树莓派相

比，便携性相当，价格略高，但 FPGA 在功耗和识别速度上优于树莓派。FPGA 与 android 设备相比，速度较慢，但价格和功耗低于 android。综合比较可以看出，FPGA 在速度、精度、功耗、便携性、价格方面的综合性能优于其他设备，是芒果品质分级检测的最优方案。

3 结 论

针对在低成本、性能受限的边缘设备上，芒果外观品质分级检测难以实现的问题，本文综合考虑检测速度，准确性，功耗，成本等因素，设计并实现了基于 FPGA 加速 CNN 的检测装置，实现了芒果外观品质的实时检测。得出主要结论如下：

1) 设计的轻量级网络 CMNet 具有较高的芒果品质检测精度，其检测准确率达到了 96.4%。同时，CMNet 模型具有较低的计算量和参数量，使得 CMNet 网络能够在资源受限的 FPGA 上部署实现。

2) 基于 FPGA 设计的 CMNet 网络硬件加速电路，进行芒果品质检测的速度为每帧 0.072 s，可实现芒果外观品质的实时检测。

3) 基于 FPGA 加速 CMNet 的芒果检测装置，总功耗为 2.6W。与 CPU、GPU 相比，FPGA 检测装置具有巨大的功耗和便携性优势。与便携式树莓派和 android 智能手机相比，FPGA 具有单位成本性能最佳的优势。

本文基于 FPGA 加速 CNN 方法的芒果外观品质分级检测装置，满足边缘环境对设备的成本、性能、供电方式和便携性的各种要求，适用于芒果外观品质的实时检测。另外，本文方法为农业上采用人工智能方法对视频数据进行快速信息提取的类似应用场景，提供了参考解决方案，对农业智能化、自动化具有重要意义。

[参 考 文 献]

[1] 邓秀新. 关于我国水果产业发展若干问题的思考[J]. 果树学报, 2021, 38(1): 121-127.
DENG Xiuxin. Thoughts on the development of China's fruit industry[J]. Journal of Fruit Science, 2021, 38(1): 121-127. (in Chinese with English abstract)

[2] 何文斌, 魏爱云, 明五一, 等. 基于机器视觉的水果品质检测综述[J]. 计算机工程与应用, 2020, 56 (11): 10-16.
HE Wenbin, WEI Aiyun, MING Wuyi, et al. Survey of fruit quality detection based on machine vision. Computer Engineering and Applications, 2020, 56(11): 10-16. (in Chinese with English abstract)

[3] GILL G. Computer vision based model for fruit sorting using K-nearest neighbour classifier[C]// International Conference on Electronic Design Innovations and Technologies (EDIT 2015). Chandigarh: International Journal of Electrical and Electronics Engineering, 2015.

[4] VARJÃO J, BOTELHO G, ALMEIDA T, et al. Citrus fruit quality classification using support vector machines[J]. International Journal of Advanced Engineering Research and

- Science, 2019, 6: 59-65.
- [5] ZAWBAA H M, ABBASS M, HAZMAN M, et al. Automatic fruit classification using random forest algorithm[C]// International Conference on Hybrid Intelligent Systems. New York: IEEE, 2015.
- [6] 田有文, 吴伟, 卢时铅, 等. 深度学习在水果品质检测与分级分类中的应用[J]. 食品科学, 2021, 42(19): 260-270. TIAN Youwen, WU Wei, LU Shiqian, et al. Application of deep learning in fruit quality detection and grading[J]. Food Science, 2021, 42(19): 260-270. (in Chinese with English abstract)
- [7] 蒋雪松, 计恺豪, 姜洪喆, 等. 深度学习在林果品质无损检测中的研究进展[J]. 农业工程学报, 2024, 40(17): 1-16. JIANG Xuesong, JI Kaihao, JIANG Hongzhe, et al. Research progress of non-destructive detection of forest fruit quality using deep learning[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(17): 1-16. (in Chinese with English abstract)
- [8] COSTA A Z D, FIGUEROA H E H, FRACAROLLI J A. Computer vision based detection of external defects on tomatoes using deep learning[J]. Biosystems Engineering, 2020, 190: 131-144.
- [9] 陈锋军, 陈闯, 朱学岩, 等. 基于改进 YOLOv7 的油茶果实成熟度检测[J]. 农业工程学报, 2024, 40(5): 177-186. CHEN Fengjun, CHEN Chuang, ZHU Xueyan, et al. Detection of Camellia oleifera fruit maturity based on improved YOLOv7[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(5): 177-186. (in Chinese with English abstract)
- [10] 倪建功, 李娟, 邓立苗, 等. 基于知识蒸馏的胡萝卜外观品质等级智能检测[J]. 农业工程学报, 2020, 36(18): 181-187. NI Jiangong, LI Juan, DENG Limiao, et al. Intelligent detection of appearance quality of carrot grade using knowledge distillation[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(18): 181-187. (in Chinese with English abstract)
- [11] 张思雨, 张秋菊, 李可. 采用机器视觉与自适应卷积神经网络检测花生仁品质[J]. 农业工程学报, 2020, 36(4): 269-277. ZHANG Siyu, ZHANG Qiuju, LI Ke. Detection of peanut kernel quality based on machine vision and adaptive convolution neural network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(4): 269-277. (in Chinese with English abstract)
- [12] HAN Song, LIU Xingyu, MAO Huizi, et al. EIE: efficient inference engine on compressed deep neural network[C]// In Proceedings of the 43rd International Symposium on Computer Architecture. New York, USA: IEEE Press, 2016: 243-254.
- [13] WANG Y, JIN X, ZHENG J, et al. An energy-efficient classification system for peach ripeness using YOLOv4 and flexible piezoelectric sensor[J]. Computers and Electronics in Agriculture, 2023, 210: 107909.
- [14] KOUTAYNI M R A, RYBALKIN V, MALIK J, et al. Real-time energy efficient hand pose estimation: A case study[J]. Sensors (Basel, Switzerland), 2020, 20(10).
- [15] YMA Yufei, CAO Yu, VRUDHULA Sarma, et al. Automatic Compilation of Diverse CNNs Onto High-Performance FPGA Accelerators[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2020, 39(2): 424-437.
- [16] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. 2017[2024-04-13], <https://doi.org/10.48550/arXiv.1704.04861>.
- [17] 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. Salt Lake City, 2018: 4510-4520.
- [18] 农业部热带农产品质量监督检验测试中心, 广西壮族自治区热带作物研究所. 芒果: NY/T 492-2002[S]. 北京: 中国标准出版社, 2002: 4-5
- [19] MILLON R, FRATI E, RUCCI E. A comparative study between HLS and HDL on SoC for image processing applications[EB/OL]. 2020[2024-04-13]. <https://doi.org/10.48550/arXiv.2012.08320>.
- [20] 杨春, 张睿尧, 黄珑, 等. 深度神经网络模型量化方法综述[J]. 工程科学学报, 2023, 45(10): 1613-1629. YANG Qing, ZHANG Ruiyao, HUANG Long, et al. A survey of quantization methods for deep neural networks[J]. Chinese Journal of Engineering, 2023, 45(10): 1613-1629. (in Chinese with English abstract)
- [21] SZEGEDY C, WEI L, JIA Y, et al. Going deeper with convolutions[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA, 2015: 1-9
- [22] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. 2016: 770-778.
- [23] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 60(6): 84-90.

Low-cost grading device for mango appearance quality using accelerated CNN with FPGA

GUO Dongdong¹, CAI Xiang^{1,2✉}, QI Jiandong^{1,2}, LUO Yuexuan¹, GAO Peiwen¹, ZHU Shan³

(1. School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China; 2. Engineering Research Center for Forestry-oriented Intelligent Information Processing of National Forestry and Grassland Administration, Beijing 100083, China;

3. Hubei Huaxia Water Resources and Hydropower Co., Ltd, Jingzhou 434000, China)

Abstract: A convolutional neural network (CNN) can be expected to grade the mango appearance quality for high accuracy. However, it is still challenging to implement CNN on low-cost and performance-constrained edge devices, due to the computation- and storage-intensive CNN algorithm. This has greatly limited the application and promotion of CNN for mango appearance quality grading. In this study, a low-cost grading device was proposed to balance the grading accuracy, speed, power consumption, and detection cost of mango appearance quality using FPGA-accelerated CNN. Firstly, a lightweight network called “Compact MobileNet” (CMNet) was designed with a simple and efficient structure. The parameter and computational complexity of CMNet were significantly reduced to compress the network structure suitable for the deployment of edge devices with acceptable accuracy. Secondly, Batch normalization (BN) layer fusion and model quantization were used to further reduce the storage requirements and computational complexity of the model, in order to accelerate the execution speed of CMNet on cost- and performance-constrained edge devices. Meanwhile, an FPGA-based hardware accelerator was designed for CMNet. Since the basic function of the accelerator was achieved, the high-level synthesis (HLS) optimization, including “unrolling the for-loop”, “pipelining the for-loop”, and “array partitioning”, were used to optimize the parallelism of the hardware accelerator. Finally, a grading device was developed for the appearance quality of mangoes using an FPGA development board, specifically the “ZYNQ Z7 Lite 7020” model. An OV5640 camera module, the CMNet network acceleration circuit, and an HDMI interface were integrated to enable the mango image collection, real-time detection, and display of appearance quality. A series of experiments were conducted in a laboratory environment, in order to verify the performance of CMNet and the grading device. The Dosehri mango dataset downloaded from the internet was used to train and validate CMNet. The mango appearance quality was classified into three grades, according to China's agricultural industry standard for mango (NY/T 492-2002). The model comparison demonstrated that the CMNet shared a very lightweight model scale, only with 0.015M parameters and 7.2MFlops computations, with a high grading accuracy of 96.4%. Therefore, the CMNet was suitable for deployment on a resource-limited FPGA terminal. The assessment experiments on the FPGA-based accelerator of CMNet demonstrated that the accelerator significantly expedited the quality grading, with a speed of 0.11 s per frame. The accelerator was capable of detecting the mango appearance quality in real time. The performance comparison among various computing devices revealed that the power consumption of the FPGA-based grading device was 2.6 W, which was the lowest among the tested devices. The FPGA-based detection device offered substantial advantages, in terms of power consumption and portability, compared with the CPUs and GPUs. The FPGA excelled in the best unit cost performance, compared with the portable Raspberry Pi devices and Android smartphones. In summary, all experimental results indicate that a low-cost, low-power, high-accuracy, and high-speed grading device was suitable for assessing mango appearance quality. This device can also be effectively utilized for the real-time grading of mango quality in the field. Furthermore, this finding can provide a strong reference for similar agricultural applications, where artificial intelligence can be employed to rapidly extract information from video data.

Keywords: convolutional neural network; mango; appearance; quality grading; FPGA-accelerated; real-time