

文章编号:1007-2780(2026)09-1315-10

基于 ZYNQ 与 YOLO26n 的 PCB 缺陷检测

严 飞^{1,2}, 邵俊楠¹, 王翔宇¹, 彭 澎¹, 刘银萍^{1*}

(1. 南京信息工程大学 自动化学院, 江苏 南京 210044;

2. 江苏省大气环境与装备技术协同创新中心, 江苏 南京 210044)

摘要:印制电路板(Printed Circuit Board, PCB)作为电子设备中的基础部件,其表面缺陷会影响电子产品的可靠性与使用寿命。针对 PCB 缺陷检测模型端侧部署中计算量大、存储开销高及功耗受限等问题,本文设计了一种基于 ZYNQ SoC 平台与 YOLO26n 目标检测模型的 PCB 缺陷检测系统。针对卷积计算瓶颈,本文采用 Vitis HLS 设计卷积加速 IP 核,并结合空间分块、通道并行、循环展开及双缓冲乒乓调度等策略提高计算并行度与数据传输效率;同时采用 16 位 Q7.9 定点量化及扩展位宽累加降低数据存储与传输开销。实验结果表明,全网络 170 个算子均完成浮点与定点路径对齐,卷积拓扑和 Q7.9 定点路径的余弦相似度最小值分别为 0.989 4 和 0.999 8。CPU 浮点与 ZYNQ 定点推理在各类缺陷检测指标上整体保持接近,精度(Precision)、召回率(Recall)和 mAP@0.5 最大差值分别为 0.023、0.024 和 0.012。硬件资源中查找表(Look-Up-Table, LUT)、触发器(Flip-Flop, FF)及数字信号处理器(DSP)使用率分别为 15.42%、7.30% 和 20.64%,整体功耗为 2.615 W,表明系统能够在保持较高检测一致性的同时实现低功耗端侧推理。

关键词:PCB 缺陷检测;ZYNQ;YOLO26n;硬件加速

中图分类号:TP391.41;TP274 **文献标识码:**A **doi:**10.3724/CJLCD.2026-0138 **CSTR:**32172.14.CJLCD.2026-0138

PCB defect detection based on ZYNQ and YOLO26n

YAN Fei^{1,2}, SHAO Junnan¹, WANG Xiangyu¹, PENG Peng¹, LIU Yinping^{1*}

(1. School of Automation, Nanjing University of Information Science and Technology, Nanjing 210044, China;

2. Jiangsu Collaborative Innovation Center for Atmospheric Environment and Equipment Technology,
Nanjing 210044, China)

Abstract: Printed Circuit Boards(PCBs) are fundamental components in electronic devices, and surface defects can affect the reliability and lifespan of electronic products. To address the issues of high computational cost, high storage overhead, and limited power consumption in edge deployment of PCB defect detection models, this paper designs a PCB defect detection system based on the ZYNQ SoC platform and the YOLO26n target detection model. To overcome the bottleneck of convolutional computation, this paper employs a Vitis HLS convolution acceleration IP core and combines spatial partitioning, channel parallelism, loop unrolling, and double-buffered ping-pong scheduling strategies to improve computational parallelism and data transmission efficiency. Simultaneously, 16-bit Q7.9 fixed-point quantization and extended bit-

收稿日期:2026-07-23;**修订日期:**2026-08-20.

基金项目:江苏省产业前瞻与关键核心技术重点项目(No. BE2020006-2);国家自然科学基金(No. 61605083)

Supported by Jiangsu Provincial Key Program for Industry Foresight and Core Technology(No. BE2020006-2);

National Natural Science Foundation of China(No. 61605083)

*通信联系人, E-mail: yinpingliu@nuist.edu.cn

width accumulation are used to reduce data storage and transmission overhead. Experimental results show that all 170 operators in the network achieve floating-point and fixed-point path alignment, and the minimum cosine similarity between the convolutional topology and the Q7.9 fixed-point path are 0.989 4 and 0.999 8, respectively. CPU floating-point and ZYNQ fixed-point inference maintained similar performance across various defect detection metrics, with the largest differences in Precision, Recall, and mAP@0.5 being 0.023, 0.024, and 0.012, respectively. Hardware resource utilization rates of Look-Up-Table (LUT), Flip-Flop (FF), and Digital signal Processor (DSP) are 15.42%, 7.30%, and 20.64%, respectively, with an overall power consumption of 2.615 W. This indicates that the system can achieve low-power edge inference while maintaining high detection consistency.

Key words: PCB defect detection; ZYNQ; YOLO26n; hardware acceleration

1 引言

随着电子产品向小型化和高集成度方向发展,印制线路板(PCB)线路间距不断缩小,制造过程中易出现开路、短路、毛刺等表面缺陷^[1-2]。早期PCB缺陷检测主要采用人工目检、数字图像处理等方法;人工目检受人员经验及主观判断影响难以满足高速生产线需求;图像处理往往依赖稳定的成像条件和较明确的缺陷边界,鲁棒性和泛化能力有限。近年来,高分辨率成像、光学检测及智能视觉技术不断进步^[3-5],推动了PCB缺陷检测向自动化、智能化方向发展。

随着人工智能技术日渐成熟,以R-CNN和YOLO为代表的目标检测模型在PCB缺陷检测中得到了广泛应用^[6-9]。R-CNN采用候选区域生成与目标分类相结合的两阶段检测方式,具有较强的特征提取与定位能力,但计算和存储开销较大;YOLO则采用单阶段检测方式,通过单次前向传播完成类别预测和边界框回归,具有检测速度快、计算效率高优势。近年来,国内外研究提出PCB-YOLO、CDI-YOLO和YOLO-HMC等模型,分别从特征融合、检测头设计和模型结构优化等方面提升检测性能^[10-12]。朱黎颖等^[13]针对复杂背景下的电路板表面焊接缺陷检测问题,提出相应的视觉检测方法以提升复杂场景下缺陷识别能力;杨彦萍等^[14]针对PCB表面缺陷检测任务对YOLO11进行改进,通过优化网络结构提升缺陷检测精度并降低模型复杂度;高雪等^[15]针对小目标、复杂背景及缺陷密集等问题改进DEIM检测算法,进一步增强了模型对PCB缺陷的检测能力。

目前深度学习主流平台包括CPU、GPU、

ASIC和FPGA^[16]。CPU并行计算能力有限,GPU具有较高吞吐率但功耗较高,ASIC虽然具有较高计算能效,但开发周期长,灵活性差。相比之下,ZYNQ平台结合FPGA的硬件可重构性与ARM处理器的通用计算能力,兼具并行性、低功耗和灵活性^[17]。

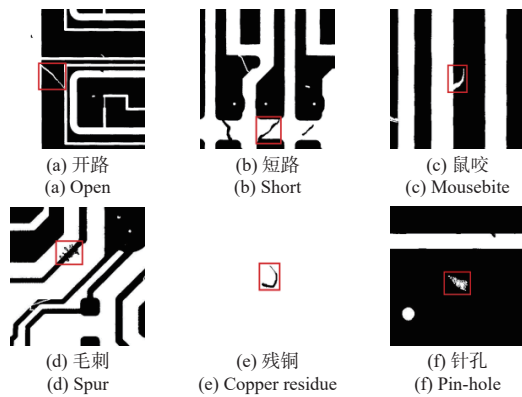
考虑PCB缺陷尺度差异大、细小缺陷特征不明显及端侧部署资源受限等问题,本文选用YOLO26n作为检测模型。其轻量化检测结构能够降低模型计算与后处理开销,在兼顾检测精度的同时提升端侧部署效率^[18]。在此基础上,本文设计了一种基于ZYNQ与YOLO26n的PCB缺陷检测系统:在PC端完成模型训练,并基于处理系统/可编程逻辑(PS/PL)异构架构实现推理加速。针对计算量较大的卷积层,采用HLS设计参数化卷积加速IP,并结合空间分块、通道并行、循环展开和双缓冲调度等策略提升片上计算效率;同时引入16位定点量化降低模型参数和中间特征图的存储开销。

2 检测算法与硬件平台

2.1 DeepPCB数据集

本文采用公开DeepPCB数据集作为PCB缺陷检测实验数据^[19]。数据集包含1 500对PCB图像,将其划分为训练集、验证集和测试集,分别包含1 168、160和172张图像。训练集用于网络参数优化,验证集用于训练过程中的性能评估,测试集用于训练完成后的泛化性能评价。

完整数据集包含6类常见PCB缺陷,如图1所示。

图1 PCB缺陷类型^[19]Fig. 1 PCB defect types^[19]

2.2 YOLO26n 目标检测模型

YOLO26n 整体由主干网络(Backbone)、颈部网络(Neck)和检测头(Head)三部分构成,网络结构如图2所示:输入图像尺寸为 640×640 ,首先由主干网络逐层提取特征,获得包含浅层纹理信息和深层语义信息的多尺度特征;随后颈部网络对不同层级特征进行融合与重组,增强多尺度特征间的信息交互;最终检测头基于融合后的多尺度特征完成边界框回归、目标置信度估计和类别预测,实现对不同尺寸目标的检测。

主干网络主要由卷积层、C3k2模块、SPPF模块和C2PSA模块组成:卷积层负责基础特征提取、通道变换和空间下采样功能;C3k2模块基于

CSP思想构建,通过特征分支、通道融合以及内部残差结构,在降低参数量和计算量的同时增强特征提取能力;SPPF模块通过快速空间金字塔池化扩大网络感受野,增强深层特征的上下文表达;C2PSA模块通过注意力机制强化空间依赖和语义关联,提升高层特征对目标区域的响应能力。

颈部网络采用FPN与PAN相结合的双向特征融合方式:FPN通过自顶向下的特征传递,将高层语义信息融入浅层高分辨率特征;PAN则通过自底向上的信息传递进一步补充定位信息,增强不同尺度特征之间的交互。网络在P3、P4和P5三个尺度上输出检测结果:P3保留较高空间分辨率,主要用于细小开路、短路、针孔等缺陷的定位;P4和P5具有更大的感受野,补充中等及较大缺陷的语义信息。

2.3 ZYNQ平台与HLS

本文选用Xilinx ZYNQ-7000系列XC7Z100-2FFG900-2芯片。该芯片内集成双核ARM Cortex-A9处理器构成的处理系统(Processing System, PS)和可编程逻辑(Programmable Logic, PL)资源,能够实现控制任务与并行计算任务的协同处理。其中PS端负责系统初始化、加速器调度、DDR数据搬运以及检测结果后处理,PL端主要完成卷积、池化、上采样、张量变换及注意力相关运算。

硬件加速模块采用Vitis HLS进行设计。HLS可将C/C++等高级语言描述的算法流程转换为寄存器传输级(Register Transfer Level, RTL)硬件电路。相比传统Verilog/VHDL开发方式,HLS能够从算法层进行硬件描述,降低开发难度并缩短设计周期。其开发流程如图3所

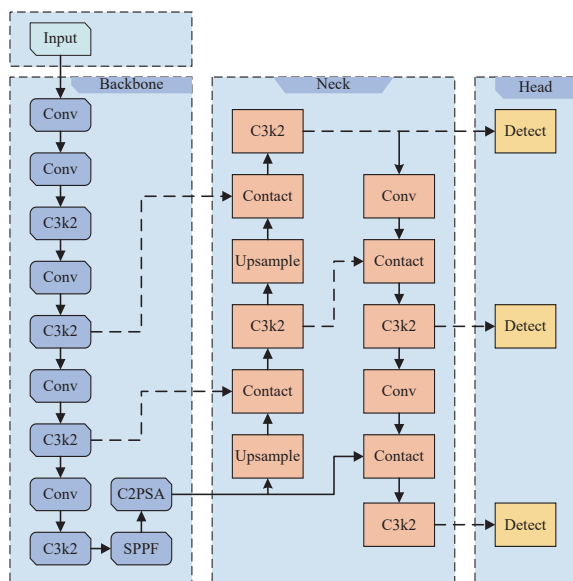


图2 YOLO26n网络结构

Fig. 2 YOLO26n network structure

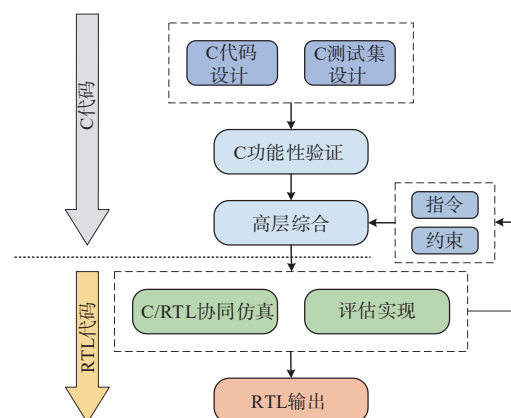


图3 HLS开发设计流程

Fig. 3 HLS development and design process

示:首先通过C代码与测试数据完成功能验证,随后在指令与约束控制下执行高层综合,实现C代码到RTL电路的转换,并通过C/RTL协同仿真进行验证与性能评估;若结果未达到要求,则调整优化指令与约束进行迭代设计。

3 硬件加速器设计

针对YOLO26n网络包含卷积、特征融合及注意力等多类型算子的特点,本文基于Vitis HLS设计了统一的参数化硬件计算单元。顶层模块通过AXI4-Lite接口接收算子类型、输入输出通道数、特征图尺寸、卷积核尺寸等配置参数,并依据操作码动态选择相应的计算路径。该模块覆盖普通卷积、 1×1 逐点卷积、深度卷积、最大池化、2倍上采样、特征拼接、逐元素加法与乘法、注意力计算及特征提取等操作。完整网络由PS端按照预先生成的计算图顺序配置并调用各类算子,PL端每次完成一个算子任务。

3.1 并行计算与循环优化

为满足不同特征提取需求,网络采用多种卷积运算形式。标准卷积通过对输入通道维度以及卷积核空间维度上的数据进行乘积累加,每个输出通道均融合全部输入通道的特征信息,同时完成空间特征提取与通道信息交互;逐点卷积是标准卷积在卷积核尺寸为 1×1 时的特殊形式,仅对同一空间位置的各输入通道进行加权累加,主要用于通道维度映射与特征融合;深度卷积则按照输入通道进行分组,各通道分别采用独立卷积核完成空间卷积,通道之间不进行信息交互。

针对上述不同卷积形式,硬件计算单元采用参数化设计,如图4所示。将输出特征图按照 13×13 的空间区域划分为计算块,输入通道并行度设置为4,输出通道并行度设置为16,形成 4×16 路并行乘加结构。标准卷积和逐点卷积复用同一参数化计算路径,通过配置卷积核尺寸和步长实现不同形式的卷积运算;当采用深度卷积模式时,输入通道与卷积分组一一对应,各通道分别使用独立卷积核进行空间卷积,通道间不进行信息交互。

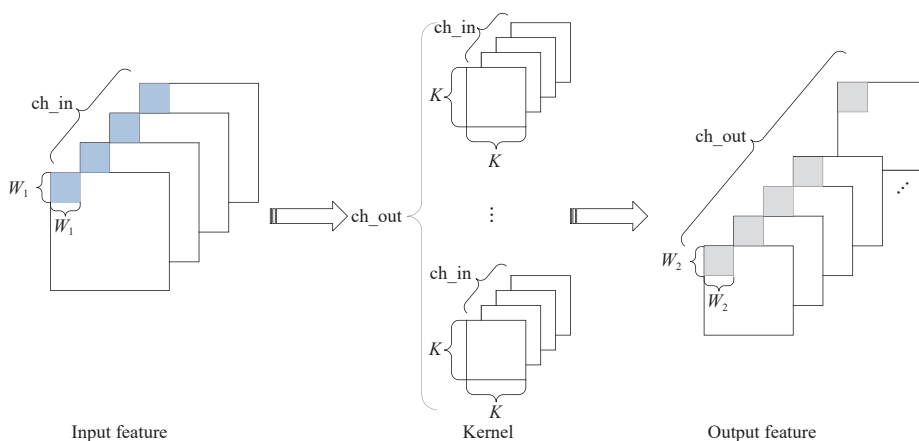


图4 卷积计算的空间分块与通道并行

Fig. 4 Spatial partitioning and channel parallelism of convolution computation

为保证并行计算单元能够连续访问数据,输入特征缓存按输入通道维数进行完全数组分割、权重缓存按输出通道和输入通道维数进行数组分割、输出累加缓存按输出通道维数进行分割。在此基础上,对输出通道和输入通道循环展开,使多个通道乘加操作能够并行执行;对输入数据装载、输出结果写回以及池化、上采样和逐元素运算等循环施加流水线约束,以降低循环控制开

销并提高数据处理吞吐率。

对于空间尺寸或通道数不能被分块参数整除的情况,计算单元通过边界判断和零填充处理尾部空间块及尾部通道,从而使不同网络层能够复用统一的硬件计算结构。

3.2 双缓冲乒乓调度

网络推理过程中会产生大量中间特征图及跨层连接数据,为降低外部存储访问开销,本文

针对不同数据流配置独立的AXI主接口,并在卷积计算单元内部设置片上缓存。PS端按照网络拓扑管理中间特征图及权重数据,PL端根据算子参数读取当前计算所需数据并完成相应计算。

对于卷积运算,输入特征和权重分别采用两组片上缓存构成乒乓缓存结构。对于一个输出空间块,系统首先将初始输入通道块及对应权重

写入第一组缓存,并将偏置写入输出累加缓存。随后将下一输入通道块及其权重写入另一组缓存,并调用卷积计算单元处理上一组缓存中的数据;当前计算完成后切换缓存状态,交替处理后续输入通道块,直至完成全部通道的累加。输入特征和权重缓存的交替调度过程如图5所示。该调度结构能够避免相邻通道块之间的数据覆盖,并提高片上存储空间的复用效率。

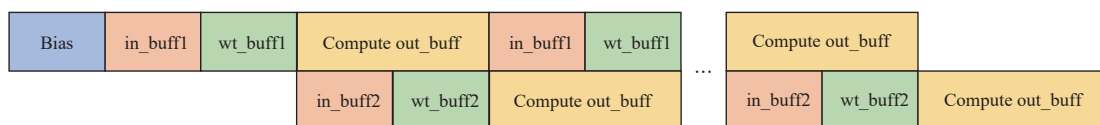


图5 输入特征与权重缓存交替调度

Fig. 5 Interleaved scheduling of input features and weight cache

卷积计算过程中,各输入通道块的部分和累加结果暂存在片上输出缓存中。完成一个输出空间块的全部通道计算后,累加结果经过激活处理并统一写回DDR。为降低输出数据存储带

来的等待开销,采用计算与存储流水机制,在写回当前输出块的同时计算下一输出块,实现计算过程与DDR访问的重叠执行,具体过程如图6所示。

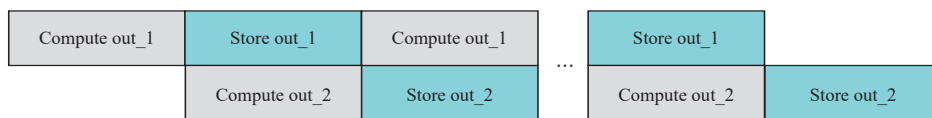


图6 输出累加缓存及写回流程

Fig. 6 Output accumulation buffer and write-back process

整个卷积计算过程按照数据加载、并行计算、结果累加和输出写回的流程协同执行:首先根据卷积参数确定输入特征的读取范围,并将初始输入通道块及对应权重写入片上缓存;随后利用数组分割为并行乘加单元提供独立的数据访问端口,并通过输入通道和输出通道循环展开,同时完成4路输入通道与16路输出通道的乘积累加。当前通道块计算完成后,输入特征和权重缓存交替切换,用于装载和处理后续通道块,部分和则持续保存在输出累加缓存中,直至完成该空间块的全部输入通道计算。最后累加结果经过激活处理后写回外部存储。

上述执行流程中,空间分块确定计算工作集,通道并行和循环展开提高计算并行度,数组分割保障并行访问,双缓存调度实现输入特征和权重分别在两组片上缓存之间交替切换,实现数据装载与计算过程的重叠。

3.3 参数定点化

为降低模型参数及中间特征的存储和计算开销,本文对网络输入、卷积权重、偏置及主要中间特征采用16位Q7.9定点格式表示。定点化过程中,浮点数据经舍入和饱和处理后转换为定点数据,以减小量化偏差并避免数值溢出^[20],具体如图7所示。

针对卷积运算中多通道乘积累加可能导致动态范围不足的问题,卷积运算采用扩展位宽进行累加,使用Q16.16格式保存累加结果,以降低

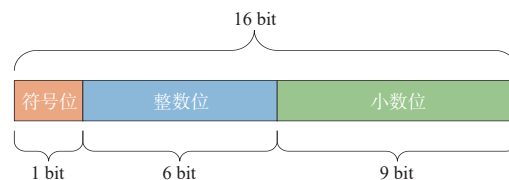


图7 参数定点化

Fig. 7 Parameter quantization

累加过程中的数值溢出风险。卷积计算完成后,再将扩展位宽的累加结果通过舍入和饱和处理量化为 Q7.9 格式,并传递至后续网络层。该处理方式在保证中间计算具有足够动态范围的同时,减少了数据缩窄过程中的精度损失。

SiLU 激活函数及注意力模块中的非线性运算通过 HLS 数学函数实现,并将计算结果量化后输入后续网络层。本文采用以定点计算为主的混合数值实现方式,在降低数据存储与传输开销的同时,控制定点化引入的数值误差。

4 系统设计与实现

本文构建的 PCB 缺陷检测系统由软件端模型训练和硬件端加速部署两部分组成,整体流程如图 8 所示。首先在 PC 端基于 PCB 缺陷数据集完成 YOLO26n 网络训练并获得权重和偏置参数;随后在 ZYNQ 平台上搭建硬件加速系统,通过 Vivado 工具链集成由 HLS 生成的加速器 IP 核;最后由 PS 端和 PL 端协同完成图像输入、卷积计算、结果回传和缺陷检测输出。

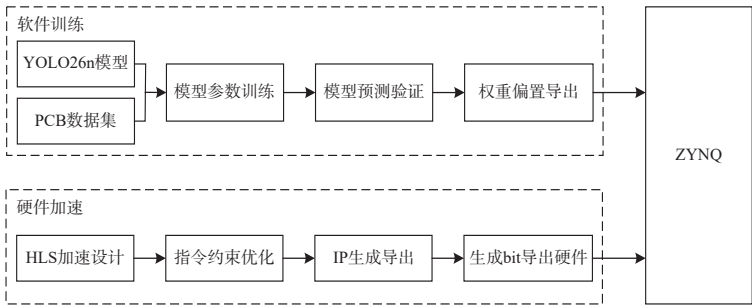


图 8 软硬件设计流程图

Fig. 8 Software and hardware design flowchart

4.1 软件模型训练

为确保 YOLO26n 模型平稳收敛并获得良好性能,本文在训练阶段对迭代轮次、优化器、学习率及正则化等参数进行统一设置:训练轮次为 100 轮、优化器采用 SGD、动量项为 0.937、权重衰减设置为 0.000 5 以抑制过拟合。学习率由 0.001 按线性策略逐步衰减至 0.000 01,使模型在训练初期保持较快收敛速度,并在后期获得更稳定的精度表现。训练过程中采用验证集对模型性能进行评估,并根据验证集表现确定最终模型权重。模型训练完成后在 172 张测试集图像上对模型进行最终性能评估,得到的各类缺陷的检测性能指标如表 1 所示。

4.2 硬件架构设计

图 9 给出了 ZYNQ SoC 系统的整体工作流程。系统运行前将 PC 端训练得到的权重、偏置参数以及待检测 PCB 图像存放至 SD 卡根目录。系统复位并完成初始化后,PS 端把原始图像和网络参数加载到 DDR3 存储器,再经 AXI 总线将待处理数据传输至 PL 端加速核;PL 端利用片上

表 1 缺陷检测性能指标

Tab. 1 Defect detection performance indicators

缺陷类别	Precision	Recall	F1	mAP@0.5
Open	0.574	0.838	0.681	0.759
Short	0.808	0.594	0.685	0.707
Mousebite	0.634	0.761	0.692	0.777
Spur	0.760	0.760	0.760	0.829
Copper	0.727	0.860	0.788	0.876
Pin-hole	0.772	0.674	0.720	0.825

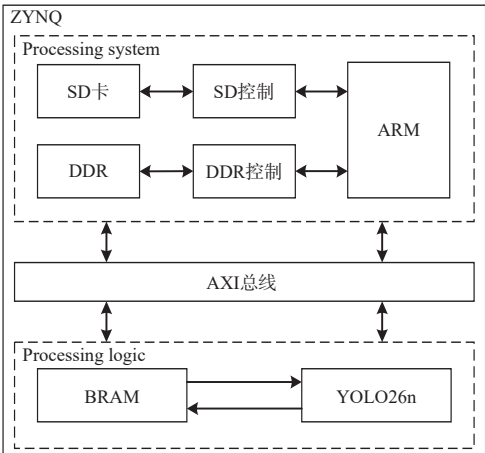


图 9 ZYNQ 框架图

Fig. 9 ZYNQ framework diagram

缓存和并行卷积阵列完成主要计算,计算结果随后通过 AXI 总线写回 DDR3,最后由 PS 端读取并保存检测结果。由于权重参数只需在启动阶段加载一次,连续检测时仅需更新待测图像,可减少重复的数据搬运开销。

PL 端加速模块是系统推理计算的核心。本文首先在 Vitis HLS 环境下使用 C 语言实现顶层统一的参数化硬件计算单元,并在完成功能验证和性能评估后添加接口约束,导出 RTL 级加速

器 IP 核。

随后在 Vivado 中进行块设计(Block Design),如图 10 所示:先添加并配置 ZYNQ IP 核,使能 PS 端所需的 SD 卡接口和 UART 串口,同时配置 PL 端时钟、DDR 存储器接口以及 PS 与 PL 之间的 AXI 互联;再将 HLS 生成的加速器 IP 核接入系统。各模块之间通过 Xilinx AXI SmartConnect IP 完成互联,从而实现 PS 端控制逻辑与 PL 端计算逻辑之间的数据通信。

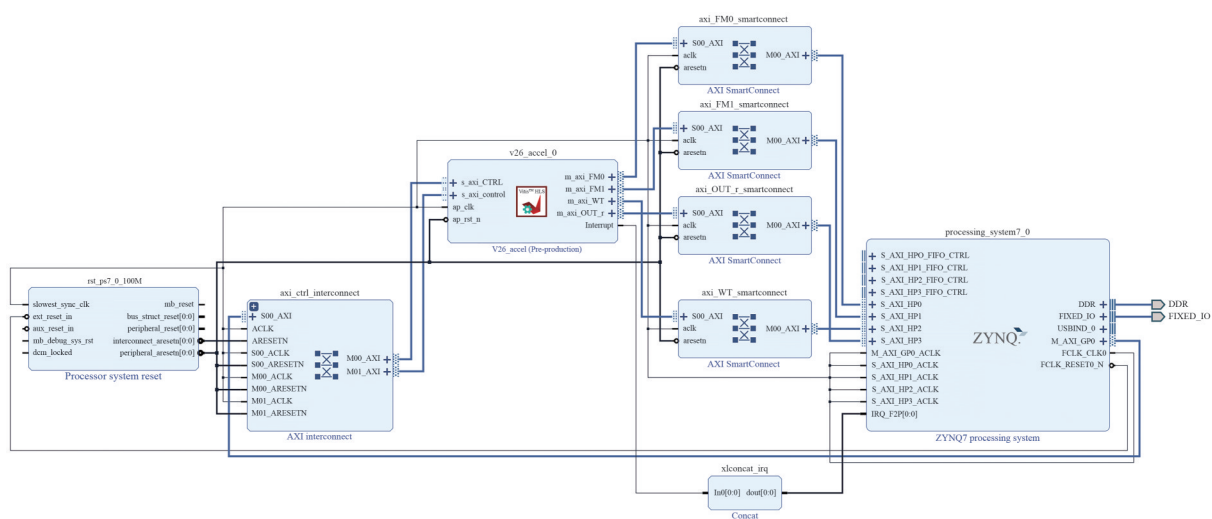


图 10 块设计(Block Design)图

Fig. 10 Block Design diagram

5 实验结果与分析

5.1 实验环境

整体实验分为软件模型训练与硬件系统实现。硬件设计采用 Vivado 2021.1、Vitis 2021.1 和 Vitis HLS 2021.1 工具链,器件为 Xilinx ZYNQ-7000 系列的 XC7Z100-2FFG900-2,PL 端工作时钟约束设置为 100 MHz。模型输入尺寸为 640×640 ,输入数据、卷积权重及主要中间特征采用 Q7.9 格式表示,卷积累加过程采用扩展位宽保存。PS 端负责系统控制、数据传输和检测结果后处理,PL 端负责网络主要计算算子的硬件加速。搭建完成的系统平台如图 11 所示。

5.2 定点化与硬件调度验证

为系统评估定点化及硬件调度对网络计算精度的影响,本文从3个层次开展一致性验证:首

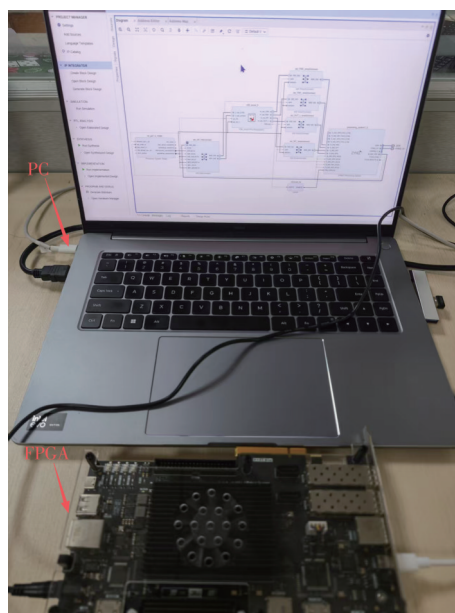


图 11 实验环境

Fig. 11 Experimental environment

先通过对比 PyTorch 模型与调度浮点路径验证模型结构映射的一致性;其次通过对比调度浮点与 Q7.9 路径评估定点化带来的数值误差;最后通过对比 CPU 与 ZYNQ 端到端检测结果,验证完整硬件推理链路的一致性。

如表 2 所示,全网络调度级审计共覆盖 170 个算子,其中卷积运算 102 个,非卷积运算 68 个,涵盖加法、特征拆分、特征拼接、池化、上采样、特征提取及注意力等操作,并完成各算子浮点参考路径与定点计算路径的输出核对。在此基础上,对其中能够与浮点模型层级输出建立对应关系的 96 个卷积算子进一步开展拓扑级核对。其余 6 个卷积算子属于检测头末端输出层,仅纳入调度级

误差统计。

表 3 给出了拓扑级对齐结果,其中,余弦相似度越接近 1,误差越接近 0,表明一致性越高。PyTorch 模型与调度浮点路径的卷积拓扑输出的余弦相似度最小值为 0.989 4,均方根误差为 0.078 7,表明网络结构转换和算子调度具有较高一致性;调度级 Q7.9 对齐结果显示,定点算子输出的余弦相似度最小值为 0.999 8,均方根误差为 0.008 9。结果表明,在本文采用的数据格式、扩展位宽累加以及舍入饱和策略下,定点计算路径与浮点参考路径的中间特征具有较高一致性,量化误差总体处于可控范围。

表 3 拓扑及定点化一致性

Tab. 3 Topological and fixed-point consistency

对比路径	余弦相似度	RMSE
PyTorch-调度浮点	0.989 4	0.078 7
调度浮点-Q7.9	0.999 8	0.008 9

表 2 全网络算子调度

Tab. 2 Network-wide operator scheduling

算子类型	数量
卷积运算	102
加法运算	19
特征拆分	18
多输入特征拼接	13
特征拼接	9
最大池化	3
上采样	2
特征提取	2
注意力运算	2
总计	170

5.3 PCB检测精度分析

为进一步验证软件端与硬件端推理路径的一致性,针对 172 张测试集图像,分别采用 CPU 端浮点路径和 ZYNQ 端 Q7.9 路径进行推理,其检测精度指标对比如表 4 所示。结果表明,两种推理路径在各类缺陷类别上的检测性能指标精度(Precision)、召回率(Recall)和 mAP@0.5 整体保持接近,最大差异分别为 0.023、0.024 和 0.012。

表 4 CPU 与 ZYNQ 推理对比

Tab. 4 Comparison of CPU and ZYNQ inference

缺陷类型	CPU Precision	ZYNQ Precision	Δ Precision	CPU Recall	ZYNQ Recall	Δ Recall	CPU mAP@0.5	ZYNQ mAP@0.5	Δ mAP@0.5
Open	0.574	0.589	+0.015	0.838	0.828	-0.010	0.759	0.763	+0.004
Short	0.808	0.822	+0.014	0.594	0.571	-0.024	0.707	0.705	-0.002
Mousebite	0.634	0.614	-0.020	0.761	0.766	+0.005	0.777	0.765	-0.012
Spur	0.760	0.782	+0.023	0.760	0.743	-0.017	0.829	0.821	-0.007
Copper	0.727	0.750	+0.023	0.860	0.841	-0.018	0.876	0.876	0.000
Pin-hole	0.772	0.778	+0.006	0.674	0.680	+0.006	0.825	0.826	+0.001

在测试集中选取一张图像进行可视化对比,检测结果如图 12 所示,左侧为 CPU 推理结果,右侧为 ZYNQ 检测结果。

两端输出的最高置信度目标均为 Short 缺陷,CPU 浮点路径和 ZYNQ 定点路径的置信度分别

为 0.920 和 0.897,二者置信度绝对差为 0.023,边界框坐标最大绝对差为 0.156 像素。该结果进一步说明,在相同输入图像条件下,Q7.9 定点计算、检测框解码及后处理过程未改变主要缺陷目标的类别判定和空间定位结果,与表 4 所示的整

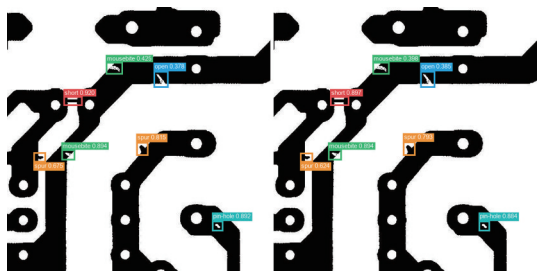


图 12 CPU 端与 ZYNQ 端检测结果对比
Fig. 12 Detection results on the CPU and ZYNQ sides

体检测指标变化相一致。

5.4 硬件资源使用

Vitis HLS 综合结果显示,在 10.00 ns 目标周期约束下,加速模块的估计周期为 7.300 ns,综合使用 BRAM_18K 65 个、数字信号处理器(DSP) 333 个、触发器(FF) 60 616 个和查找表(LUT) 75 305 个。

Vivado 系统级布局布线结果显示,在 100 MHz 时钟约束下,完整设计使用 LUT 42 784 个、FF 40 492 个、BRAM 36 个和 DSP 417 个,对应资源利用率分别为 15.42%、7.30%、4.83% 和 20.64%。

资源使用情况如表 5 所示,从表中资源使用来看,系统的 DSP 利用率相对较高,主要来源于

表 5 资源使用
Tab. 5 Resource utilization

资源类型	已使用	总量	使用率/%
LUT	42 784	277 400	15.42
FF	40 492	554 800	7.30
BRAM	36	755	4.83
DSP	417	2 020	20.64

卷积计算中的并行乘累加操作;LUT 和 FF 利用率分别为 15.42% 和 7.30%,表明计算控制逻辑及流水线寄存器带来的资源开销处于可控范围;BRAM 利用率为 4.83%,表明当前片上存储资源能够满足本文特征缓存和权重分块的存储需求。

ZYNQ 平台的功耗如表 6 所示,片上总功耗为 2.615 W,其中动态功耗为 2.367 W,约占整个系统功耗的 90%,是系统能耗的主要组成部分;静态功耗为 0.248 W,占比约 10%。

表 6 系统功耗
Tab. 6 System power consumption

功耗类型	功耗/W
动态功耗	2.367
静态功耗	0.248
总功耗	2.615

6 结 论

本文基于 ZYNQ 平台实现了 YOLO26n PCB 缺陷检测模型的端侧部署,构建了 PS 与 PL 协同的硬件架构,并采用 Q7.9 定点表示和扩展位宽累加实现主要算子的硬件映射。全网络 170 个算子完成逐算子对齐验证,卷积拓扑和 Q7.9 定点路径的余弦相似度最小值分别为 0.989 4 和 0.999 8。CPU 浮点推理与 ZYNQ 定点推理的检测性能指标 Precision、Recall 和 mAP@0.5 整体保持接近,最大差值分别为 0.023、0.024 和 0.012,验证了定点化推理路径具有较好的一致性,为资源受限嵌入式平台上的 PCB 缺陷检测提供了可行的实现方案。

参 考 文 献

[1] 吴一全,赵朗月,苑玉彬,等. 基于机器视觉的 PCB 缺陷检测算法研究现状及展望[J]. 仪器仪表学报,2022, 43(8):1-17.
WU Y Q, ZHAO L Y, YUAN Y B, et al. Research status and the prospect of PCB defect detection algorithm based on machine vision [J]. *Chinese Journal of Scientific Instrument*, 2022, 43(8): 1-17. (in Chinese)

[2] 徐一奇,肖金球,汪俞成,等. 基于机器视觉的 PCB 表面缺陷检测研究综述[J]. 微电子学与计算机,2025,42(4):1-15.
XU Y Q, XIAO J Q, WANG Y C, et al. A review of research on PCB surface defect detection based on machine vision [J]. *Microelectronics & Computer*, 2025, 42(4): 1-15. (in Chinese)

[3] PARK J, CHOI Y, KWON S, et al. Microsphere-assisted hyperspectral imaging: super-resolution, non-destructive metrology for semiconductor devices [J]. *Light: Science & Applications*, 2024, 13(1): 122.

[4] 刘智,赖庆荣,张天禹,等. 激光多普勒测振技术无损检测果冻橙粒化病[J]. 中国光学(中英文),2025,18(1):160-169.
LIU Z, LAI Q R, ZHANG T Y, et al. Non-destruction detection of jelly orange granulation disease using laser

- Doppler vibrometry [J]. *Chinese Optics*, 2025, 18(1): 160-169. (in Chinese)
- [5] ZHANG Z L, KONG L Y, ZHANG L H Y, *et al.* Structured light meets machine intelligence [J]. *eLight*, 2025, 5(1): 26.
- [6] GIRSHICK R, DONAHUE J, DARRELL T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation [C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 580-587.
- [7] REN S Q, HE K M, GIRSHICK R, *et al.* Faster R-CNN: towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [8] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You only look once: unified, real-time object detection [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779-788.
- [9] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [J/OL]. *arXiv*, 2020: 2004.10934.
- [10] TANG J L, LIU S B, ZHAO D X, *et al.* PCB-YOLO: an improved detection algorithm of PCB surface defects based on YOLOv5 [J]. *Sustainability*, 2023, 15(7): 5963.
- [11] XIAO G S, HOU S L, ZHOU H Y. PCB defect detection algorithm based on CDI-YOLO [J]. *Scientific Reports*, 2024, 14(1): 7351.
- [12] YUAN M H, ZHOU Y B, REN X Y, *et al.* YOLO-HMC: an improved method for PCB surface defect detection [J]. *IEEE Transactions on Instrumentation and Measurement*, 2024, 73: 2001611.
- [13] 朱黎颖,王森,沈爱萍,等. 复杂背景下的电路板表面焊接缺陷视觉检测[J]. 光学精密工程, 2024, 32(14): 2256-2271. ZHU L Y, WANG S, SHEN A P, *et al.* Visual inspection of soldering defects on board surfaces against complex backgrounds [J]. *Optics and Precision Engineering*, 2024, 32(14): 2256-2271. (in Chinese)
- [14] 杨彦萍,高军伟,刘兆龙,等. 基于改进 YOLO11 的印刷电路板表面缺陷检测算法[J]. 液晶与显示, 2025, 40(10): 1545-1556. YANG Y P, GAO J W, LIU Z L, *et al.* Surface defect detection on printed circuit boards based on improved YOLO11 [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(10): 1545-1556. (in Chinese)
- [15] 高雪,袁飞,霍艳飞. 基于改进 DEIM 的印刷电路板缺陷检测算法[J]. 液晶与显示, 2026, 41(5): 700-716. GAO X, YUAN F, HUO Y F. PCB defect detection algorithm based on improved DEIM [J]. *Chinese Journal of Liquid Crystals and Displays*, 2026, 41(5): 700-716. (in Chinese)
- [16] DONG Y B, BAI Y C, ZHANG Q M, *et al.* High-throughput optical neuromorphic graphic processing at millions of images per second [J]. *eLight*, 2025, 5(1): 29.
- [17] MELONI P, CAPOTONDI A, DERIU G, *et al.* NEURAghe: exploiting CPU-FPGA synergies for efficient and flexible CNN inference acceleration on Zynq SoCs [J]. *ACM Transactions on Reconfigurable Technology and Systems*, 2018, 11(3): 18.
- [18] JOCHER G, QIU J, LIU M Y, *et al.* Ultralytics YOLO26: unified real-time end-to-end vision models [J/OL]. *arXiv*, 2026: 2606.03748.
- [19] TANG S L, HE F, HUANG X L, *et al.* Online PCB defect detector on a new PCB defect dataset [J/OL]. *arXiv*, 2019: 1902.06197.
- [20] DAI D Y, ZHANG Y C, ZHANG J H, *et al.* Trainable fixed-point quantization for deep learning acceleration on FPGAs [J/OL]. *arXiv*, 2024: 2401.17544.

作者简介:



严 飞,男,博士,副教授,2013年于中国科学院长春光学精密机械与物理研究所获得博士学位,主要从事三维测量与显示、嵌入式图像处理技术等方面的研究。E-mail: fyan@nuist.edu.cn



刘银萍,女,博士,副教授,2013年于吉林大学获得博士学位,主要从事电磁场计算、安全技术等方面的研究。E-mail: yinpingliu@nuist.edu.cn