2019 年 SEU-Xilinx 国际暑期学校团队项目设计文档

(Project Paper Submission Template)

设计作品名称	基于 DPU 加速的深度学习人脸识别与颜值评估解决方案
参赛队员姓名、学校及所在院系	杜江溯/中山大学 高性能计算协同创新中心 陈志炜/中山大学 电子信息工程学院 古奕康/东莞理工学院 粤台产业科技学院 刘璐/北京大学 软件与微电子学院
参赛队员房间号及桌号	717 室/17 组

第一部分

小组成员分工

（各成员任务分配）

	姓名	任务分配
组长	杜江溯	深度学习人脸学习算法选择和训练，DNNC 映射算法，调试人脸打分系统
组员 1	刘璐	DPU IP 的配置、优化和测试，Petalinux 的配置、SDK 的配置，文档编写，视频制作
组员 2	陈志炜	Ultra96 板卡和镜像调试，调试人脸打分系统，文档编写
组员 3	古奕康	Petalinux、DNNC 的配置，NMS 算法改进成 Soft-NMS 算法

第二部分

设计概述 /Design Introduction

（请概括地描述一下你的设计，包括设计目的、应用领域及适用范围等。撰写过程中应注重突出设计实现的主要/特色功能）

利用深度学习技术实现人脸识别、追踪和颜值打分功能，并利用 DPU+DNNDK 对程序的 CNN 网络部分进行加速，以实现实时性要求。

设计目的：

作为实际应用场景，发挥 FPGA 在计算方面的优势，又实现 AI 领域中人脸识别的具体应用，以良好的硬件资源为载体，在其上实现优化的神经网络模型，以实现现实场景中的实际应用，是对软硬件协同设计的良好落地。

应用领域及适用范围：

基于 DPU 的深度学习人脸识别能够广泛应用于监控所在的各种场景，通过摄像头对于图像的摄取，以及深度学习技术的处理，在基于 DPU 的 FPGA 上进行计算，能够快速的识别并跟踪人脸，进而实现多场景应用，例如对目标人物的快速追踪，对场景内人员的年龄评估颜值打分等更细化的需求。本项目在此次紧张的开发周期中实现了边缘深度学习加速器的设计与应用，实现了基础功能与具体功能的开发，日后可将本项目作为基础项目，实现更细化需求的开发与应用。

第三部分

系统组成及功能说明 /System Construction & Function Description

（请详细说明你作品要实现的所有功能以及如何组建系统以实现该功能，还包括为实现该功能需要用到的所有参数和所有操作的详细说明，必要的地方多用图表形式表述）

1. 本项目用到的硬件设备如下：

序号	设备
1	Ultra96 开发板
2	USB 摄像头
3	USB 键盘
4	MiniDP--HDMI 转换线
5	USB_Hub
6	Micro SD 卡
7	串口调试板（选用）

2. 项目功能模块划分：

序号	项目模块	步骤细分	功能描述
1	人脸检测	使用深度学习框架实现图片的人脸识别	略
		人脸检测模块DPU部署	对模型进行优化，并利用 DNNDK 进行转换。
2	颜值打分	使用深度学习框架实现人脸颜值打分，并训练	略
		人脸检测模块DPU部署	对模型进行优化，并利用 DNNDK 进行转换。
3	项目集成		功能模块串联

3. 具体实现：

如图 1，是项目的总体架构，ARM 处理器对视频流进行处理，生成 CNN 网络可以处理的图片，后将任务加载到 FPGA 中进行计算，返回人脸在图像中的位置，在 ARM 中对人脸进行截取，并 resize 到下一个 CNN 网络的输入大小，输入到下一个 CNN 网络，获取颜值评分。再在 ARM 中利用 opencv 库对图片进行处理并生成视频流输出。

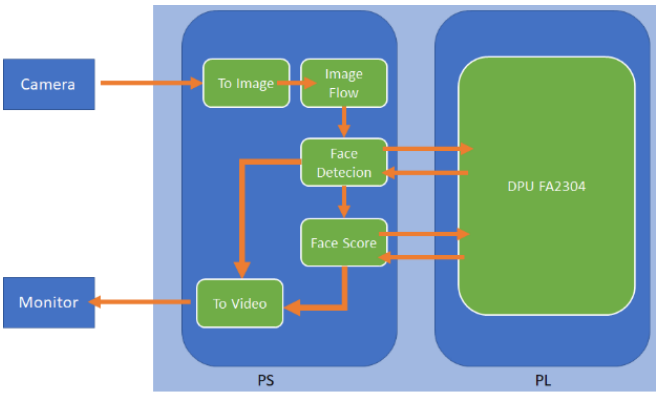

Figure 1 项目架构

基于 DPU 的深度学习快速人脸识别解决方案，是 Xilinx 边缘深度学习加速器的设计与应用的实现，所使用的硬件平台为 Ultra96 开发板，内置卷积结构为 B2304 的 DPU。在本项目中，分为相对独立又相辅相成的两条软硬路来推进整体流程。首先，对于硬件道路，通过使用 vivado 软件布置 DPU 核生成.hdf 文件，在 petalinux 环境中生成适合于 ultra96 的 img 与 sysroot，将镜像 img 存入 SD 卡作为 Ultra96 的系统。同时，对于软件道路，使用 DNNDK 软件，《此处需要 2 句话描述 DNNDK 软件的运行流程》，生成的.elf 文件与编写的.cc 运行文件一同在 XilinxSDK 软件中进行编译生成 img 可执行的.elf 文件，放入 SD 卡中。由此，软件和硬件两条路已经走通，将 ultra96 开机，运行.elf 文件，则可看到运行的效果，实现人脸识别和打分功能。

硬件设计方面，使用 vivado 软件进行 Block Design，针对不同的硬件资源，可以选择不同的 DPU 型号，针对 Ultra96 选择 B2304 核，对新 Block Design 进行综合和实现，并生成比特流，将生成的.hdf 文件放入 Petalinux 环境中，生成适配的镜像。

在硬件设备方面，本项目的主要特点是应用了 Ultra96 的 DPU 核， DPU 是一个专用于卷积神经网络的可硬件设计，可快速高效计算卷积核。Ultra96 开发板是基于 arm 规范的 SoC 开发版，在可编程逻辑中拥有广泛的潜在外围设备和加速引擎。其硬件参数信息大致如下：包含 2GB 的 LPDDR4 内存，16G Micro SD 卡适配，WiFi/蓝牙，MiniDP 口，USB 口等。其中在本项目中，我们使用 16G Micro SD 卡将适配的镜像写入，放入 Ultra96 中，摁下开机键；将 MiniDP 口与 HDMI 相连，则在显示器上可以看到启动画面，随后进入操作系统的终端界面，同时，使用 USB 口插入鼠标/键盘，方便进行操作。在 Ultra96 上也可以接入串口进行调试，亦可以通过 wifi 连接 IP 地址进行调试。

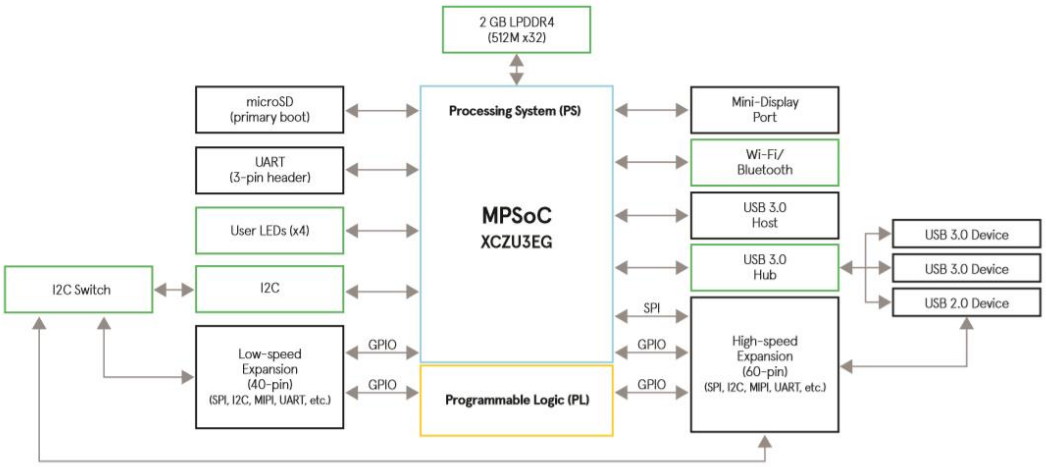

Figure 2 Ultra96 Architecture

软件方面，包括对两个 CNN 网络的实现，优化，训练与后续的 DPU 部署。基于 Github 开源项目，我们分别使用 DenseBox 与 VGG 实现人脸检测与人脸打分。

网络优化与 DPU 部署，以人脸打分为例：

- 人脸打分功能的本质是图片分类，如图 Figure2 是人脸数据集中的人脸数据，图片名标识了图片的标签，也就是颜值，通过深度学习网络学习人脸与颜值的对应关系。
- VGG16 网络的参数量为百 M 级别，其前馈推断计算量对于 Ultra96 来说过于庞大，无法满足实时性的要求，因此我们需要对网络进行优化，首先我们选择 miniVGG 从网络结构的角度来降低参数量，其次我们对输入图片进行了 Resize，如图我们将图片缩小到 32*32*3 的大小，这极大的降低了对资源的需求。之后使用 DNNDK 提供的量化工具，最终的 elf 文件从开始的 400MB 优化到最终的 2MB。

Figure 3

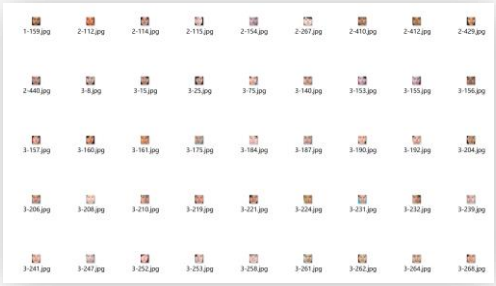

Figure 4

- DNNC 根据网络结构与参数转化 DPU 执行文件（ELF）。
- 难点：由于 DNNDK 的文档不够详细，因此在 decent 与 dnnc 的过程中，对于数据对应关系，网络的对应关系等问题有较多需要尝试。

将 NMS 算法改进成 Soft-NMS 算法：

```
            iarea = (box[i][2] - box[i][0] + 1) * (box[i][3] - box[i][1] + 1);
            jarea = (box[j][2] - box[j][0] + 1) * (box[j][3] - box[j][1] + 1);
            inter = w * h;

            ovr = inter / (iarea + jarea - inter);
            if (method == 0)
                exist_box[j] = ovr >= thres ? false : true;
            else
            {
                score = box[j][4] * exp(-(ovr * ovr) / sigma);
                exist_box[j] = score >= thres ? true : false;
            }
        }
    }

    vector<vector<float>> result;
    result.reserve(keep.size());
```

NMS

Soft-NMS实现

Figure 5 Soft-NMS 改进

- NMS 算法全名为非最大抑制算法，在人脸检测后在多张重复的检测框中选择最优的检测框，是物体检测的后处理步骤。
- 由于 NMS 算法最终选择的依据是通过硬阀值选择，太过粗暴，有学者于 ICCV2017 提出了 Soft-NMS 的算法，指在通过将硬阀值改为软阀值，达到 Soft 的效果。
  论文链接：https://arxiv.org/abs/1704.04503

第三部分

完成情况及性能参数 /Final Design & Performance Parameters

完成情况：分模块完成作品（已实现的功能）：

序号	项目模块	步骤细分	功能描述	完成
1	人脸检测	使用深度学习框架实现图片的人脸识别	略	√
		人脸检测模块 DPU 部署	对模型进行优化，并利用 DNNDK 进行转换。	√
2	颜值打分	使用深度学习框架实现人脸颜值打分，并训练	略	√
		人脸检测模块 DPU 部署	对模型进行优化，并利用 DNNDK 进行转换。	√
3	项目集成		功能模块串联	×

本设计实现基本实现了 DPU 加速的深度学习人脸评分，完成了神经网络算法的训练，由 DNNC 成功编译出.elf。完成了 petalinux 的一些配置。完成了人脸评分应用程序的修改。完成了在 Ultra96 上进行编译和调试。

从结果上，我们看到了我们完成的系统能成功调用 DPU， 即 dpuOpen();成功加载 dpu 内核，创建了 DPUTask ， dpuCreateTask(kernelconv, 0);进入人脸评分网络，face_core(taskconv, image, imageName)。

从图看到 dpuGetOutputTensorChannel()，dpuGetOutputTensorScale()均成功运行，给 DPU 输入了图片 dpuSetInputImage2()，运行 DPU 任务 dpuRunTask(taskConv));但是由于 DNNC 编译深度学习网络模型文件后，还缺了全连接层，由于没有全连接层的权值参数，但在最后的项目集成过程中，由于对 ELF 文件中接口的不熟悉，主要是全连接层不在 DPU 中进行训练的问题，导致最终未能按时完成各个模块的串联工作。

在我们的测试中，人脸检测算法可以实现实时的视频流处理，而人脸打分功能可以实现每秒 500 张 32*32*3 图片的处理，也就是说可以实时的对 20 张人脸进行打分，而对于实际应用，无论是人脸检测还是人脸打分都可以通过处理部分数据的方法来降低对系统的要求，因此我们认为我们的算法能够满足实时性的要求。

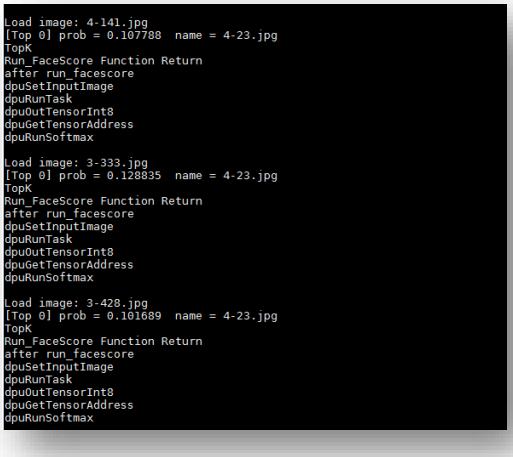

Figure 6 FaceScore 运行过程

第五部分

项目总结 /Conclusions

(项目中所用到的知识点，项目收获及心得)

知识点：

1	Vivado 设计套件中构建硬件平台 将 DPU IP 添加到 IP 目录。最主要的是关于 DPU 之间的 IP 和互联的问题，DPU 使用 3 个时钟，还有 AXI 的一些知识储备。生成比特流，生成.hdf 文件，导出硬件
2	PetaLinux 中生成 Linux 平台，创建 PetaLinux 项目，配置 PetaLinux 以安装 dnndk 文件 构建系统指向.hdf 从 Vivado Design Suite 导出的文件，配置 rootfs，将 DPU 添加到设备树，构建内核和根文件系统，创建启动镜像，创建 sysroot
3	深度神经网络的设计和优化，模型的训练，DNNK 和 DNNC 环境。
4	编写人脸检测应用程序，编写人脸评分应用程序，在板上进行调试。

项目收获：

在这个项目中，根据主办方提供的文档结合组内不同背景的同学发挥各自的专业优势，在不到 7 天的时间内将整个实验流程跑通，并作出神经网络优化，添加具体功能应用等对项目的实现。通过使用 vivado 配置硬件资源，使用 petalinux 生成镜像，使用 DNNDK 生成 decent 和 dnndc，使用 SDK 生成可执行文件，最后在 Ultra96 开发板上运行。

第六部分

源代码/Source code

（作品源码 Github 链接，github 操作步骤见 *Github 项目上传简单教程*）

https://github.com/dujiangsu/FaceScore-Accelerated-by-Xilinx-DPU.git

第七部分（技术总结/心得/笔记等分享）

（可分享的技术文档附件或已发表的文章链接）

陈志炜：

本次暑假学校，过得可是真是充实又忙碌，和小伙伴们天天白天一起调试，天天晚上加班有时还熬夜。见识前几天的技术大牛授课中学习了 HLS，认识到 PYNQ 的生态，还有最吸引的 DPU 加速。本次 DPU 项目过程，可是身体力气的考验和心态的考验，考验身体是否能受得住来连续一周的项目开发所需的体力，有时做项目遇到 bug，长时间调试未果，心态会容易奔溃，心会容易累。但是这也是做技术岗位所必经的路，做技术所必须的具备的。本次学习到了 HLS 在 FPGA 开发的优势，学习发 HLS 的优化以 for 循环，func 函数，array 为核心。尤其是 for， 可以 uroll，pipeline。本次项目学习到许多 xilinx 关于在 FPGA 上部署 DPU 的操作，收获很多。

刘璐：

抱着学习交流的心态从北京到南京，在这最后一天，心里满满的开心与收获却不知从哪里说起。从干货满满的四天讲座，到每日充实且实用的实验，从项目小组成立，到今天一起总结走过的路与风景，我想我收获的不仅是技术上的多方位学习，更是认识志同道合且可爱的人的欢喜。技术上的壁垒有时难以快速突破，但是同伴间简单的快乐却能让人认识到生活的美好，非常荣幸的和这么多有趣又可爱的同学们一起。在短暂的十几天中学习，再次熟悉一系列软硬结合的技术操作流程，也对这个行业有了更多层面的接触与了解，非常感谢老师在培训中的知识的传授，感谢同学们一路一起走过的精彩，感谢南京。