

一种 SDI 视频目标检测及压缩全链路实现方法

王生霄¹ 沈梦雨¹ 高小雨¹
WANG Shengxiao SHEN Mengyu GAO Xiaoyu

摘要

针对现今嵌入式端视频目标检测分离级联式方案处理复杂度高、成本高,难以规模化适用低成本需求的场景,文章设计了一种资源约束型全链路视频采集、视频预处理、目标检测及视频压缩解决方案。该方案基于高集成多功能域态芯片,以 SDI 视频为输入,通过可编程逻辑实现视频解码、视频预处理及后处理;依托片内神经网络处理器(neural processing unit, NPU)资源实现神经网络推理,并利用通用处理器资源实现后处理、视频压缩,结合流水线式高效任务调度方案,实现视频全链路处理。经嵌入式工程化实践表明,该链路解决方案能够以低功耗、低延时,高效地实现视频处理需求,极大节约设计成本。

关键词

SDI; FPGA; 目标检测; 全链路解决方案

doi: 10.3969/j.issn.1672-9528.2026.02.025

0 引言

目前随着战场态势的复杂性变化,对无人装备的智能化提出较高要求,要求在具备图传能力的同时,还应具备自主态势分析、决策能力,能够自主检测地方基地、地方武器装备等信息^[1]。因此无人侦察装备自身需要具备边缘端侧智能处理能力,使无人装备在无线链路中断的情景下也能够对战场态势做出回应。无人装备显著的特点是低功耗、低成本及小型化,而目前在目标检测端到端的实现时,广泛采用多芯片分离式级联处理方案,即视频采集模块、视频预处理模块、神经网络推理模块、视频压缩模块等多级串联分离式处理,由此带来的高成本、处理的高延时、高功耗,这在低成本及小型化的产品设计需求中是不可接受的。

基于此,提出一种资源约束性的全链路目标检测及压缩解决方案,该方案以高集成的集成多功能固态芯片为基础,实现视频采集及目标检测全链路处理,实现无人装备的低功耗、视频处理低延时、低成本设计。

1 架构设计

全链路视频处理方案主要实现视频采集、视频预处理、视频目标检测、视频压缩四个主要功能。

视频采集主要实现外部视频的采集,因此全链路视频处理方案也主要针对 SDI 视频进行预处理,传感器采集 SDI 视频后经过 75 Ω 同轴电缆,在视频采集域使用均衡器对外部视

频进行补偿,然后使用可编程逻辑进行视频解码,完成视频采集。

视频预处理主要通过可编程逻辑实现对单帧视频的裁剪、数据归一化处理等操作,以此达到与模型数据分布的一致性。

视频目标检测主要通过 FPAI 芯片内部集成的神经网络处理芯片实现神经网络模型推理,通过可编程逻辑实现阈值筛选、非极大值抑制等后处理,得到目标在图像中的坐标信息。

视频压缩使用通用处理器内置视频控制单元(video control unit, VCU)硬件资源实现视频 H.264/HEVC 压缩。

该方案通过逻辑资源实现视频的编解码、压缩以及目标检测预处理,极大减少了端至端的处理延时,同时使用单芯片节约了设计成本,降低了产品功耗。如图 1 所示。

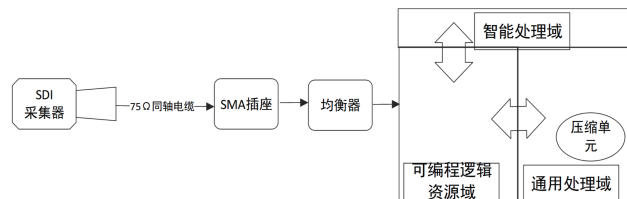


图 1 全链路实现方案

2 视频采集

目前 SDI 视频以传输稳定性高、传输距离远、低成本等优点,常用于低成本无人机吊舱图传解决方案,SDI 视频传输常用的标准^[2]如表 1 所示。小型化吊舱常用的视频传输标准为 3G-SDI,用于传输可见光及红外图像,传输线缆使用 1 根 75 Ω 同轴电缆。

1. 中国航空工业集团公司西安航空计算技术研究所
陕西西安 710068

表 1 SDI 视频标准

序号	标准来源	名称	视频格式	传输速率	时间
1	SMPTE 259M	SD-SDI	480i&576i	270 Mbit/s	1989
2	SMPTE 292M	HD-SDI	720P&1080i	1.485 Gbit/s	1998
3	SMPTE 424M	3G-SDI	1920×1080@60Hz	2.97 Gbit/s	2006
4	SMPTEST-2081	6G-SDI	3840×2160@30Hz	6 Gbit/s	2015
5	SMPTEST-2082	12G-SDI	3840×2160@60Hz	12 Gbit/s	2015

SDI 视频编解码主要有两种实现方案：

方案 1：使用专用的编解码芯片实现，如图 2 所示，即外部串行 SDI 视频经过 75 Ω 同轴电缆传输至 SMA 基座，使用解码芯片实现模拟信号向并行 YCbCr422 或 BT1120 数字信号的转换，对外传输时使用编码芯片将并行的数字视频信号编码至串行的 SDI 视频信号对外输出。典型的解码器芯片有 SEMTECH 公司的 GS2971 芯片，编码芯片 GS2972。该策略的优点是软件及硬件实现方便简单，缺点在于编解码芯片均无国产化芯片可替代，因此对于无国产化要求的项目，可使用该策略实现 SDI 视频的编解码。

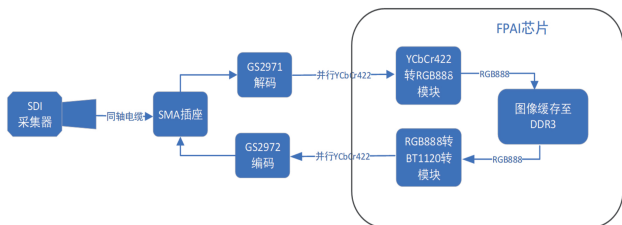


图 2 专用芯片 SDI 编解码方案

方案 2：使用可编程逻辑资源实现 SDI 视频的编解码，如图 3 所示，即外部 SDI 视频经过同轴电缆传输至开发板的 SMA 基座，使用均衡器芯片实现单端信号转为差分信号，均衡器为长距离传输的 SDI 信号提供信号强度均衡补偿，补偿后的差分信号接入 FPGA，在 FPGA 中使用 GTX 收发器实现串行数据的并行化，选用 GTX 收发器主要原因为其采用高性能的调制解调算法，具备较强的错误检测和纠正能力，能够在高噪声环境下稳定传输。GTX 收发器采用成熟的 IP 核实现。至此仅实现数据的接收，接收到数据发送至 FPGA 的下一个硬件解码模块，解码模块采用 SMPTE 的 SDI IP 核进行 SDI 视频的解码操作，解码后的数据为 RGB888 格式，至此完成 SDI 视频的编解码操作。

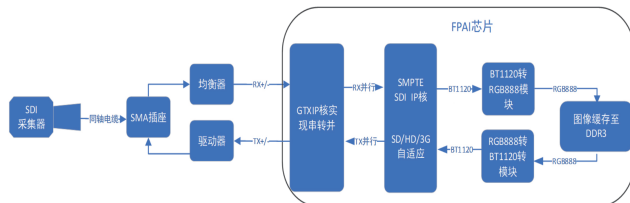


图 3 可编程逻辑实现 SDI 视频编解码

本方案基于一体化、低成本、全链路、低功耗考虑，因此选用集成可编程逻辑资源的 FPAI 芯片实现 SDI 视频的编解码，即采用方案 2 实现 SDI 视频的采集。

3 视频预处理

视频数据采集完成后得到 RGB888 的视频数据格式，由于输入数据格式为 1 080P@60 Hz 视频数据，数据处理速度小于视频输入速度，因此在逻辑端 DDR 内存中缓存一帧视频，当启动预处理模块后，从逻辑端 DDR 中读取一帧图像，单帧图像复制为两份，1 份视频流用于神经网络前向推理，1 份视频流用于后标定检测框及视频压缩。

其中，神经网络前向推理的图像在进行缩放的处理方式应该和模型训练的处理方式保持一致，即采用直接拉伸图像比例而不考虑形变的策略，这样避免为网络带入额外噪声，影响模型的准确性。这一步骤采用 Resize 模块实现。在缩放完成后，需要将图像进行归一化处理，此步骤仍然需要与训练阶段保持一致，这样能够保证视频数据样本分布拟合的一致性，能够提高检测精度，图像归一化操作采用 SCALE 模块实现，预处理后的图像存放在逻辑段的 DDR 中作为 NPU 的前向输入。

用于压缩的视频分支，在逻辑端需要图像大小裁剪为压缩模块可识别的尺寸，裁剪后的图像格式为 RGB565，即将各 RGB 三通道的低位数据截取而后获得。同时在保证图像质量的前提下，为了减少数据传输量，可将 RGB565 格式的图像转化为 YUV422（16 bit），此时数据量可减少 50%，而图像的色度分量能够保证完整，预处理后的视频数据通过片内总线存放在通用处理器端的 DDR 中。

视频预处理均采用可编程逻辑实现，实现过程中涉及多个模块间数据通信，为了保证数据传输的低延时，在逻辑设计时多个模块采用同一时钟域实现。

4 目标检测

4.1 模型准备

针对无人机航拍领域应用，选取 DOTAV2.0 数据集^[3]作为训练样本，该数据集包含来自不同平台以及不同传感器的航拍图像，如 Google Earth、JL-1 卫星拍摄，以及中国资源卫星数据和应用中心的 GF-2 卫星拍摄。数据集包含固定翼飞机、旋翼飞机、港口、桥梁、车辆、跑道等 18 个类别，11 268 个图像和 1 793 658 个实例，每张图像的像素尺寸在 800 px× 800 px 到 20 000 px× 20 000 px 之间。

4.2 算法选用

针对机载嵌入式目标检测处理要求低延时、高效率、高

精度的特点,因此选用单阶段的YOLO系列神经网络模型,而YOLOv5^[4]作为YOLO系列中较为先进的一种目标检测模型,且针对不同应用场景,衍生S、M、L等多种网络模型版本。S网络模型仅有 7.2×10^6 的参数量,常用于资源较低的设备,适合部署于嵌入式端,可以做到实时推理,但较低的网络参数降低了模型推理的精度;M网络模型有 21.2×10^6 的参数量,在计算资源和推理之间取得了平衡;L网络模型有 46.5×10^6 的参数量,具备更深的网络结构,因此对计算资源要求较高,常用于云端服务器的高精度部署。

选用的多功能域态芯片具备27.5TOPs@INT8算力,为了实现端到端检测的实时性,因此选用YOLOv5s网络模型,同时针对航拍数据集中以及无人机吊舱航拍图像中小目标(像素分布在 $32 \text{ px} \times 32 \text{ px}$ 以下像素)分布较多的特点,提高识别精度,对YOLOv5s网络模型从以下几个方面进行改进:

(1) 检测头扩展

YOLOv5s原型网络模型在预测大、中、小目标时,使用 20×20 、 40×40 、 80×80 三种尺度的特征图进行预测,每种尺度特征图分别分配3种锚框,其中 20×20 的感受野较大,可以预测大尺寸目标; 40×40 用于预测中型尺度目标; 80×80 尺寸的特征图感受野相对较小可以预测小目标。

由于航拍数据集中小目标分布较多,大尺度目标分布相对较小,因此取消 20×20 尺度特征图的预测,增加一层 160×160 尺度的检测头,即对 80×80 尺度的特征图进行上采样然后与主干网络中的第二层卷积后得到 160×160 尺度的特征图同时增加预测的锚框尺寸[4,5,8,10,22,18],用于预测远距离广角视角下的航拍 4×4 像素大小的小目标。

(2) 增加注意力机制

由于传统卷积网络仅关注局部信息,而忽略了全局信息,因此为了提高目标检测的准确性,在主干网络中增加通道注意力机制与空间注意力机制,这样使得网络在训练时能够更加关注于目标本身的特征,提高检测的准确性,同时兼顾推理的实时性,因此在网络中不宜增加过多的注意力模块,因此仅在主干网络的最后一个卷积层增加注意力机制,用于聚焦小目标特征。

4.3 模型转换

训练生成的网络模型需要使用FPAI提供的网络模型编译工具将网络模型的数据结构进行重译,操作包括优化算子间的连接关系以及算子融合,数据量化等操作,编译后得到模型Detection.bin。

4.4 神经网络推理

NPU前向推理时,首先需要初始化NPU计算单元,即

调用API接口,打开设备,从存储介质中读取编译量化后的网络模型以及网络结构,并实例化该网络结构,将网络运行中的算子绑定至物理计算设备上,然后从逻辑端的DDR中读取已经预处理过的图像数据进行前向计算,完成前向推理后,输出包含检测信息的最后一层特征图,存入逻辑端的DDR内存中,阈值筛选模块读取逻辑端DDR中的特征图,将低于阈值的检测加过删除,将高于阈值的检测结果保留,筛选后的结果通过内部AXI总线传输至通用处理器端的DDR中保存,用于目标检测后处理。

4.5 检测后处理

通用处理器端应用程序通过读取DDR(通用处理器端)中的预测数据,使用非极大值抑制(non-maximum suppression, NMS)^[5]方法,将冗余检测框剔除,得出最终检测结果,然后使用坐标转换程序将缩减图像的坐标转换为原始图像中目标的坐标,应用程序读取DDR中原始视频数据,将检测坐标标记在原始视频流上,存入通用处理器端DDR中,待VCU调取实现压缩。

5 压缩实现

压缩采用通用处理器端内置硬件VCU编解码引擎实现,该引擎主要实现编解码功能,编码与解码共享硬件资源,为半双工实现,当需求为同时实现多个码流的编码和解码时,需要采用分时复用的策略实现编解码,因此需要着重评估分时复用带来的延时是否能够满足要求,本项目中测试多功能域态芯片的编解码时延。

VCU软件主要为两个层级实现,第一层为驱动层,该层与VCU硬件直接交互,以V4L2视频驱动框架为基础实现;第二层为用户程序层,通过调用驱动API实现对VCU的访问,实现时首先通过调用API接口设置输入与输出的视频格式及压缩标准,然后以内存映射的方式申请缓冲区Buffer,完成输入输出Buffer申请后,即可从内存中将输入数据搬运至VCU进行压缩,然后将压缩后的输出数据循环读取。

6 全链路流水线优化

为了提升全链路检测的性能,即提高图像处理速度,因此将整个全链路按照时间先后的处理速度,分为4部分,即视频读取(含预处理)、神经网络计算、后处理、视频压缩,4个部分采用流水线多线程处理,各处理部分分别运行在自己的线程中,线程之间通过采用任务队列的方式进行通信,如图4所示。

针对不同的网络模型,设计不同的优化策略,对于小网络模型而言,前向计算的时间不再是影响整个处理流程的最大因素,此时耗时集中在图像预处理模块,因此可以通过设

置预处理访问内存的优先级大于前向处理的优先级，将整个计算时间进行平均；对于大网络模型而言，则更注重缓存的申请，因为在整个数据处理的过程中，需要在通用处理器端的 DDR 和逻辑端的 DDR 中申请多个缓存来暂存图像数据，缓存的申请即从初始化阶段开始，直至整个程序运行结束。同时为了保证处理过程中数据不丢帧，需要着重分析整个 Buffer 的申请数量，一个 Buffer 必须等到后处理完成后，才能空闲出来用于存放下一帧图像，在 Buffer 空闲之前视频是不断输入的，此时系统会进入 2 帧图像，因此为了提高处理并行程度，缓存的数量应申请 3 个，此时能够最大化地对处理速度进行优化。

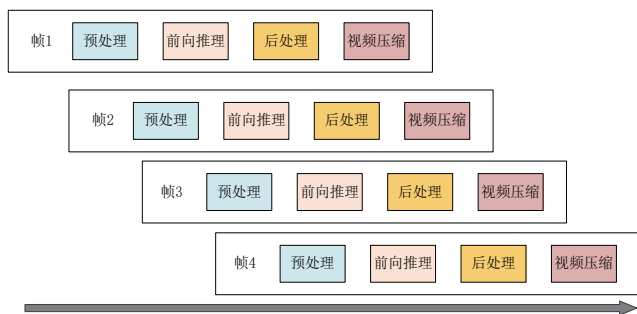


图4 全链路流水线优化处理

7 检测结果

测试使航拍图像,在该图像中共有7种目标类型,分别为 Person、Truck、Bus、Car、Plane、Bike、Tricycle,表2为推理检测准确率,图5为可视化检测结果,表3为各阶段检测耗时。

表 2 航拍检测结果

序号	目标种类	检测准确度 /%
1	Person	90
2	Truck	96
3	Bus	92
4	Car	95
5	Plane	94
6	Bike	95
7	Tricycle	93



图 5 航拍图像可视化检测结果

表 3 视频处理耗时

序号	处理阶段	处理时间 /ms
1	预处理	1
2	NPU 检测	8
3	逻辑后处理	0.2
4	ARM 后处理	15
5	压缩	30

如表 2 及图 5 所示, 优化后的网络模型以较高的检测准确度实现了对目标的检测。基于实验结果表明, 全链路一体化解决方案能够在嵌入式端满足工程化实践应用。

8 结语

本文基于无人机低空航拍目标检测应用场景，针对视频采集，视频解码，视频预处理，目标检测，视频压缩全流程处理，提出一种全链路一体化解决方案，以低功耗，低成本，低延时为目标，基于目前国内高集成度的多功能域态芯片，实现了视频流端到端的处理，极大地提高了边缘侧视频的处理时间，且以软件约束思路充分发挥硬件资源使用率，单芯片片内处理，片内总线互传，多任务统一调度，充分发挥单芯片硬件资源能力，工程化实践表明，该全链路解决方案能够以低功耗、低延时高效地实现视频处理需求，极大节约设计成本，值得多项目推广使用。

参考文献:

- [1]尹欣繁,章贡川,彭先敏,等.军用无人机技术智能化发展及应用[J].国防科技,2018,39(5): 30-34.
- [2]SMPTE ST 259:2008 电视-10比特4:2:2分量和4fsc复合数字信号的串行数字接口[S]. Society of Motion Picture and Television Engineers (SMPTE), 2008.
- [3]XIA G S, BAI X, DING J, et al. DOTA: a large-scale dataset for object detection in aerial images[C/OL]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.Piscataway:IEEE,2018[2025-12-30].<https://ieeexplore.ieee.org/document/8578516>.DOI:10.1109/CVPR.2018.00418.
- [4]YOLOv5:Real-time object detection[M]. Ultralytics LLC, 2020.
- [5]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.

【作者简介】

王生霄(1994—),男,宁夏吴忠人,硕士研究生,工程师,
研究方向:嵌入式计算机、智能计算。

(收稿日期: 2025-08-12 修回日期: 2026-02-03)