

基于AI智算WLAN的 智能视频监控系统设计与实现

Design and Implementation of Intelligent Video Monitoring System Based on AI Computing WLAN

余涛, 曾炜 (中国联通江西分公司, 江西 南昌 330000)
Yu Tao, Zeng Wei (China Unicom Jiangxi Branch, Nanchang 330000, China)

摘要:

设计并实现了一种集成边缘计算能力的AI智算WLAN架构。该架构在无线接入控制器中深度融合了算力为6~30 TOPS的神经网络处理单元加速模块,通过异构计算框架实现视频流处理任务的硬件卸载。系统采用改进的三帧差分法进行运动目标检测,并集成YOLOv5s-INT8轻量化模型实现多模态分析。通过协议栈优化与智能流量调度,构建了扁平化的二层网络拓扑。实验结果表明,在智慧园区等高密度场景下,本系统可实现平均响应延迟 ≤ 150 ms,带宽占用降低65.8%,存储需求减少90%,并显著降低部署空间与运维成本。

Abstract:

It designs and implements an AI computing WLAN architecture integrated with edge computing capability. A neural processing unit (NPU) acceleration module with 6~30 TOPS computing power is deeply integrated into the wireless access controller, and hardware offloading of video stream processing tasks is realized through a heterogeneous computing framework. The system adopts an improved three-frame difference method for moving target detection, and integrates the lightweight YOLOv5s-INT8 model to complete multimodal analysis. A flat two-layer network topology is constructed via protocol stack optimization and intelligent traffic scheduling. Experimental results show that in high-density scenarios such as smart parks, the proposed system achieves an average response delay of ≤ 150 ms, bandwidth consumption reduced by 65.8%, storage demand reduced by 90%, and greatly lowers deployment space and operation & maintenance costs.

Keywords:

Edge computing; WLAN; Video analysis; NPU acceleration; Resource scheduling

关键词:

边缘计算; WLAN; 视频分析; NPU加速; 资源调度
doi: 10.12045/j.issn.1007-3043.2026.07.010
文章编号: 1007-3043(2026)07-0057-04
中图分类号: TN929.5
文献标识码: A
开放科学(资源服务)标识码(OSID): 

引用格式: 余涛, 曾炜. 基于AI智算WLAN的智能视频监控系统设计与实现[J]. 邮电设计技术, 2026(7): 57-60.

1 概述

随着企业数字化转型的深入,高清视频监控与无线局域网已成为提升运营效率与安全保障的关键基础设施^[1]。然而,传统WLAN架构最初并非为传输密集、持续的视频流而设计,其在支持智能视频分析应

用时存在显著瓶颈。一方面,将视频流上传至云端进行分析的模式导致高昂的带宽消耗与难以接受的延迟^[2];另一方面,采用内置GPU/NPU的AI摄像头方案则面临硬件成本高、散热困难、模型更新维护复杂等困境^[3]。

边缘计算技术为上述问题提供了潜在的解决路径,即通过将计算任务下沉至网络边缘,就近处理视频数据^[4]。近期研究进一步强调了云边端协同分析的

收稿日期: 2026-06-10

重要性^[5],以及工业视频网络在资源受限环境下实现智能感知所面临的挑战与进展^[6]。文献[7]提出了多种基于边缘服务器的计算卸载策略,以平衡分析任务的延迟与能耗。然而,这些方案通常将边缘服务器作为独立实体,与网络设备间的通信开销和资源调度冲突问题尚未得到根本解决。

同时,智能WLAN技术(如基于802.11ax标准的OFDMA和BSS Coloring)可有效提升高密度接入场景下的频谱效率和抗干扰能力^[8]。更有前沿研究开始探索生成式AI与无线传感的融合,为构建无线基础模型提供了新的思路^[9]。但现有研究多集中于通信性能优化,未能将AI计算能力以硬件形式深度嵌入网络控制层,难以实现计算与通信资源的统一调度。

本文提出一种深度集成于WLAN接入控制器(AC)的AI智算架构,将边缘计算能力从“外挂”式服务器嵌入到网络核心设备AC中,实现了网络与算力的原生融合,为上述挑战提供了新的解决方案。

2 系统架构设计

2.1 整体架构

如图1所示,本文设计的AI智算WLAN系统架构由数据平面、计算平面和控制平面构成。数据平面负责基础的数据转发与接入管理,全面支持IEEE 802.11ax标准及ONVIF/RTSP等视频传输协议。计算平面作为系统的核心,集成了NPU阵列、视频预处理单元和智能分析引擎。NPU阵列基于多核异构架构,专用于视频流的结构化处理与特征提取。该设计符合边缘智能网关的集成化发展趋势^[10]。控制平面通过智能调度引擎,协同管理网络资源与计算资源,实现动态QoS策略和算力分配(见表1)。

2.2 业务流程

如图2所示,系统的视频智能分析业务流程包含如下步骤。

a) 视频采集与预处理。普通摄像头捕获的原始视频流首先由视频预处理单元进行解码、降噪和关键区域(ROI)提取。此步骤可减少约70%的后续计算量。

b) AI推理与分析。预处理后的数据送入NPU阵列,并行执行目标检测(YOLOv5s-INT8)、行为识别(SlowFast精简版)和异常检测(LSTM+Attention)等任务。该分析引擎具备处理复杂场景的能力,其理念与先进的行业级视觉模型相符^[11]。

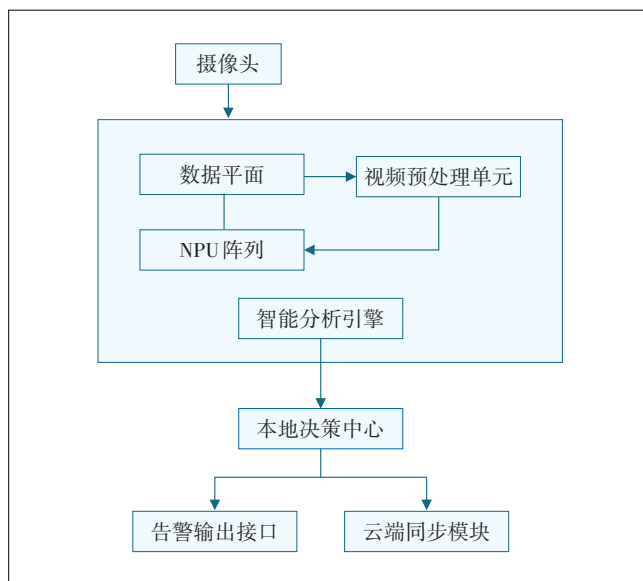


图1 智算WLAN系统架构

表1 系统架构组件功能描述

平面	核心组件	主要功能
数据平面	网络接口模块	数据转发、协议支持
计算平面	NPU阵列	视频流结构化处理
	预处理单元	视频解码、降噪、ROI提取
控制平面	调度引擎	资源协同管理

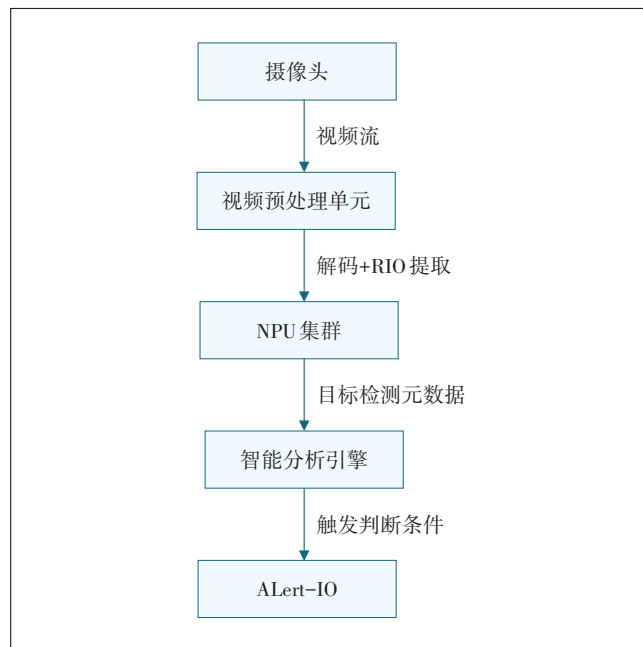


图2 AI智算WLAN的视频业务流程处理

c) 决策与输出。智能分析引擎对结构化数据进行融合分析,实时生成告警事件或控制指令。

3 关键算法与机制

3.1 计算核心动态调度机制

为高效利用NPU的4个计算核心(见表2),本文设计了动态调度策略。其中,核心0与核心1采用流水线并行模式,专用于高优先级的目标检测任务。核心0负责图像预处理,核心1执行模型推理,并通过共享内存实现状态同步,确保任务延迟 $\leq 20\text{ ms}$ 。核心2固定用于行为识别任务,采用FP16精度平衡速度与准确率。核心3作为智能备用核心,支持热备份与突发事件接管,实现资源弹性扩展。

全局调度器基于强化学习算法,以10 ms为周期监控各核心的利用率、任务队列长度和QoS等级,动态分配算力资源。

3.2 网络资源协同调度机制

为实现网络与计算的协同优化,设计了如下调度机制。

表2 NPU核心动态调度策略

核心编号	主要任务	调度策略	性能保障
0,1	目标检测	流水线并行,负载均衡	延迟 $\leq 20\text{ ms}$,故障切换 $< 3\text{ ms}$
2	行为识别	固定分配,低负载时辅助	支持32路并发
3	备用/突发事件	热备份,动态频率调节	响应延迟 $< 5\text{ ms}$

a) 智能QoS分级。根据AI分析结果(如异常事件)动态调整视频流优先级,为关键流分配80%的可用带宽。

b) 虚拟切片管理。将无线频谱划分为8个虚拟切片,关键视频流独占2个高优先级切片,确保传输可靠性。

c) 传输协议增强。采用DMA零拷贝技术,将数据直接从摄像头传感器传输至NPU内存,结合802.11ax的OFDMA机制,将单帧传输延迟优化至8 ms。

4 实验与结果分析

4.1 实验环境

为验证系统性能,在某企业库房部署了原型系统(见图3)。实验环境包括1台AI智算AC控制器、5个802.11ax AP、20台1080P@30fps普通摄像头。该部署场景符合智能工厂管控一体化集成参考模型^[12]及重点行业数字化转型的参考指引^[13]中强调的场景化落地原则。系统加载的算法模型如表3所示。

4.2 性能评估

系统连续运行30天,各项算法的性能指标如表4所示。所有算法均表现出较高的精度和较低的推理延迟,满足实时性要求。其性能优于已有的基于ARM平台的嵌入式监控系统方案^[6]。

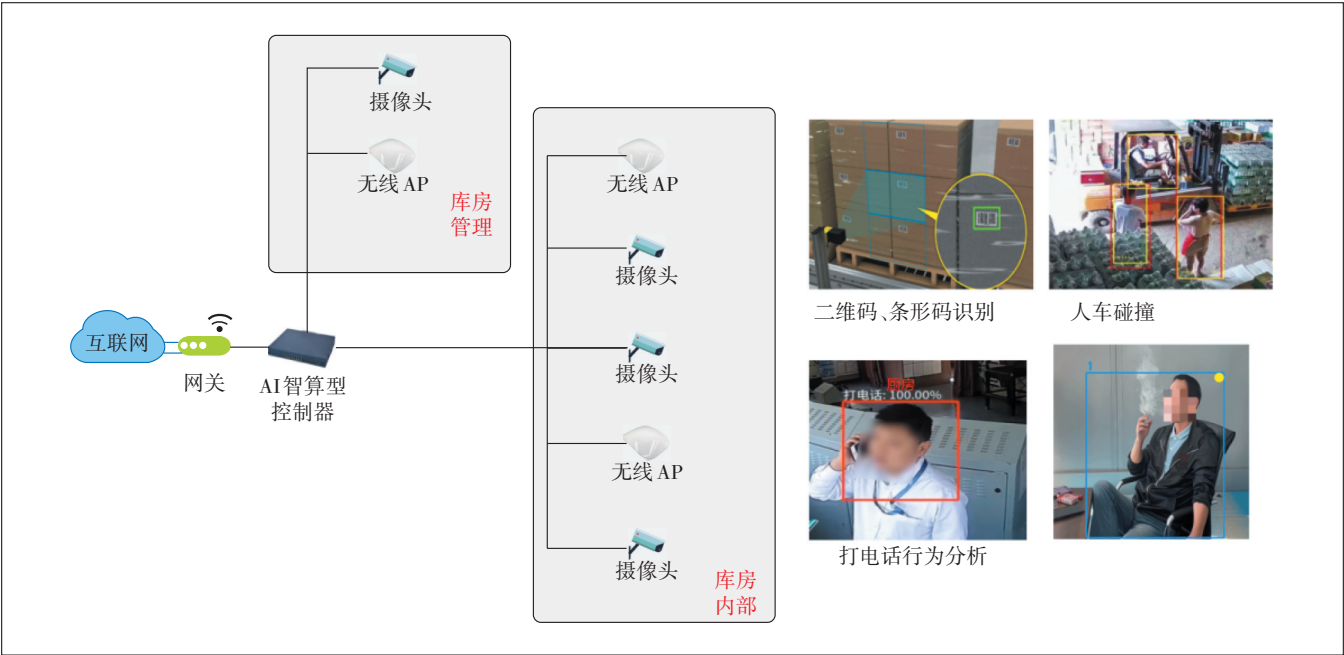


图3 AI智算WLAN在某企业库房数字化监控中的应用拓扑

表3 系统加载的算法模型

分析类型	算法模型	应用目标
目标检测	YOLOv5s-INT8	二维码、车辆识别
行为识别	SlowFast-精简版	抽烟、打电话行为识别
异常检测	LSTM+Attention	人车碰撞检测

表4 算法性能实测指标

检测类型	算法模型	精度/%	推理耗时/ms
二维码识别	DeepQR	99.8(Accuracy)	10
人脸识别	MobileFaceNet	98.8(Accuracy)	10
行为分析	SlowFast	96.3(Accuracy)	32
车辆检测	YOLOv5s-INT8	91.0 (mAP)	8

本方案与传统WLAN方案的对比结果如表5所示。从表5可以看出,本方案在带宽占用、存储需求和响应延迟方面均有显著提升。

表5 与传统WLAN方案的性能对比

性能指标	传统WLAN方案	AI智算WLAN方案	提升幅度
带宽占用率	82%	28%	降低65.8%
存储空间需求	1.2 TB/天	100 GB/天	降低90%
异常事件响应延迟	10 s~20 min	150 ms	延迟从10 s~20 min降至150 ms,提升2个数量级

5 结论与展望

本文设计并实现了一种基于AI智算WLAN的智能视频监控系统。通过将NPU加速模块深度集成到网络控制器中,实现了计算与通信资源的原生融合与协同调度。实验证明,该系统能有效解决传统方案面临的带宽、延迟和成本问题,为高密度视频监控场景提供了高效、可靠的解决方案。此类技术对于提升公共安全应急效率^[14]及推动可持续发展具有积极意义。

未来的研究将集中于以下几个方面。

a) 探索5G融合。研究5G网络与AI智算WLAN的深度融合组网技术,支持更广域的无缝覆盖与移动性管理。

b) 隐私保护增强。研究基于联邦学习的隐私保护型协同分析框架,在发挥数据价值的同时严格保护个人隐私。

c) 多租户资源保障。进一步优化多租户环境下的资源隔离与服务质量保障机制,提升平台的商用化能力。

参考文献:

- [1] 刘灵科,尹岗. 基于DSP运动目标检测系统的设计与实现[J]. 电子测量技术,2008,31(10):159-161.
- [2] 郭峰. 摄像头云监控系统的设计与实现[D]. 上海:东华大学,2015.
- [3] 张秀娟,何秀文,张军元,等. 一种基于移动边缘计算的计算卸载和资源分配方法:CN202310967149.6[P]. 2023-08-03.
- [4] Mao Y Y, You C S, Zhang J, et al. A survey on mobile edge computing: the communication perspective[J]. IEEE Communications Surveys & Tutorials, 2017, 19(4): 2322-2358.
- [5] 童佳慧,李越,李燕君,等. 边云协同的视频分析任务卸载优化策略[J]. 传感技术学报,2025,38(1):128-134.
- [6] 涂静正,温晓婧,陈彩莲,等. 基于边缘计算的工业视频网络智能感知:挑战与进展[J]. 自动化学报,2025,51(8):1715-1738.
- [7] 杨佳,李畅,曾莉. 基于嵌入式技术的智能远程视频监控系统设计研究[J]. 中国新技术新产品,2025(5):7-10.
- [8] IEEE Std 802.11-2020; IEEE standard for information technology--telecommunications and information exchange between systems--local and metropolitan area networks--specific requirements--part 11: wireless lan medium access control (MAC) and physical layer (PHY) specifications[S/OL]. <https://ieeexplore.ieee.org/document/9363693>.
- [9] Yang Z, Chi G X, Wu C S, et al. Generative AI meets wireless sensing: towards wireless foundation model[EB/OL]. [2026-04-09]. <https://arxiv.org/abs/2509.15258>.
- [10] 冯波,汪建业. 可控边缘智能网关的研制及工程应用[J]. 计算机与自主智能研究进展,2025,3(1):31-33.
- [11] Dahua Technology. Xinghan large-scale AI models: V-series (vision models) and M-series (multimodal models)[EB/OL]. [2026-04-09]. <https://www.dahuasecurity.com/Products/keyTechnologies/modalXH>.
- [12] GB/T 45387-2025 离散型智能工厂管控一体化集成参考模型[S].
- [13] 工业和信息化部办公厅. 场景化、图谱化推进重点行业数字化转型的参考指引(2025版)[EB/OL]. [2026-04-09]. <https://gxt.hebei.gov.cn/sme/zc/zxzc/2025111315140949565/>.
- [14] 刘霞. AI系统借普通摄像头实时监测火情[N]. 科技日报,2025-09-19(4).

作者简介:

余涛,工程师,学士,主要从事数据通信工程规划、设计工作;曾炜,助理工程师,学士,主要从事科技创新、移动通信工程规划、设计工作。

