

视频AI分析平台大规模视频分析技术研究

张传刚, 梁秉豪, 袁明明, 王凯, 周正

浪潮通信信息系统有限公司, 山东 济南

收稿日期: 2023年11月26日; 录用日期: 2023年12月21日; 发布日期: 2023年12月29日

摘要

在智慧城市建设持续纵深推进的过程中, 智能视频分析技术已经广泛应用在城市公共安全管理、市容市貌治理和非现场式监督等场景。随着“雪亮工程”视频监控项目在全国范围内大面积推进, 一、二、三类监控点位已经能够覆盖大部分城市公共区域。大量视频监控摄像头的视频数据汇聚接入和智能分析, 对原有的视频监控系统提出了更高的要求。随着人工智能算法架构的不断升级以及预训练大模型等技术逐渐走向成熟, 对于平台资源调度能力, 峰值并发能力都带来了更大的挑战。本文重点介绍了浪潮视频AI分析平台, 平台利用关键帧解码技术, 节省90%的算力资源消耗, 提升了通过云边协同的分布式架构, 将任务流和数据流进行解耦, 同时引入了大模型技术进行视频结构化分析和存储, 大幅提升视频关键内容检索效率。平台实现了对百万级视频数据的实时分析, 对智能分析任务的实时调度, 可以在保证智能视频分析服务质量的同时降低资源消耗, 满足实际应用需求。

关键词

视频分析, 云边协同, 任务调度, 视频监控

Research on Large-Scale Video Analysis Technology of Video AI Analysis Platform

Chuangang Zhang, Binghao Liang, Mingming Yuan, Kai Wang, Zheng Zhou

Inspur Communication Information System Co., Ltd., Jinan Shandong

Received: Nov. 26th, 2023; accepted: Dec. 21st, 2023; published: Dec. 29th, 2023

Abstract

In the process of continuous in-depth promotion of smart city construction, intelligent video anal-

ysis technology has been widely used in urban public safety management, city appearance management and off-site supervision and other scenarios. With the “Xueliang project” video surveillance project in a large area of the country, one, two, three types of monitoring points have been able to cover most of the urban public areas. The convergence of video data access and intelligent analysis of a large number of video surveillance cameras put forward higher requirements for the original video surveillance system. With the continuous upgrading of artificial intelligence algorithm architecture and the gradual maturation of technologies such as pre-training large models, greater challenges have been brought to platform resource scheduling capabilities and peak concurrency capabilities. This paper focuses on the introduction of Inspur video AI analysis platform, which realizes real-time analysis of millions of video data, saves 90% of computing resources and real-time scheduling of intelligent analysis tasks through the distributed architecture of cloud edge collaboration, which can ensure the quality of intelligent video analysis service while reducing resource consumption and meeting the needs of practical applications.

Keywords

Video Analysis, Cloud Edge Collaboration, Task Scheduling, Video Surveillance

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在《中国安防行业“十四五”规划》和《超高清视频产业发展行动计划(2019-2022年)》等国家和行业相关政策驱动下,视频监控、智能安防和计算机视觉等相关行业发展迅猛[1]。在后疫情时代背景下,我国经济持续恢复,智能视频监控领域需求旺盛,政府企业投资额较大。传统人工视频监控方式成本上涨推动平台智能化升级佳速,居民消费升级也进一步促进家用视频监控成为智能视频市场的新增长点。

智能视频分析是视频监控向智能化演进的重要技术,如何从原始的视频数据中通过计算机算法分析并提取出复合人类认知的关键事件和语义信息,是智能视频分析的主要研究内容。郑雪原[2]等人对平台底层架构进行了优化以 Kafka, MPP 数据库, Hbase 和 Spark/Hive 集群为基础,可支持城市级的海量视频监控接入和存储。唐志鸿[3]等人结合了物联网技术采集环境数据,结合视频分析实现了对环保、水利和电力等行业的监管。叶逢春[4]等人通过视频分析技术,实现了对城市管理中的违规经营和违规占道等行为进行分析。目前大部分的智能视频分析应用场景主要通过集中式云化部署和分布式边缘部署两种方式对外提供服务,这两种方式都存在一定的局限性。其中,集中式云化部署适合应用在前端设备分布较为集中而且数量较多的情况,这种方式可以有效降低软硬件系统运维成本,同时提高算力利用效率,但是对于网络资源消耗和隐私数据安全等方面较难满足用户个性化需求。分布式边缘部署提供了更灵活的应用模式,可以保障数据不出局满足对时延要求或数据安全要求较高的场景,在工业互联网和公安执法等领域广泛应用。

针对上述两种方式所存在的局限性,本文提出了一种基于云边协同方式实现的大规模视频分析技术,并将该技术应用在视频 AI 分析平台中。平台以“一云、二库、三中心”的基础架构,建立了云化的算力资源池,以及数据共享仓库和算法仓库快速支撑上层应用的灵活编排。同时将智能视频分析任务解耦为视频解析中心,任务编排中心和智能感知中心,通过“平台与设备”、“应用与算法”双解耦方式实现视频数据和 AI 能力的协同调度,适配多种底层平台和运行环境,支持高并发、大规模的视频分析业务。

2. 视频 AI 分析平台产品架构

2.1. 产品功能架构

视频 AI 分析平台主要包括视频汇聚层、共享存储层、视频分析层和能力应用层(如图 1 所示)。视频汇聚层主要将各类监控摄像头采集到的视频数据和设备信息进行统一的汇聚,完成视频流的实时标准化转发,提供直播和回放等能力。共享存储层主要包含了业务数据和 AI 算法,通过数据共享仓库,对视频、图像和结构化数据进行持久化的存储与共享。通过 AI 算法仓库统一进行算法模型统一存储和版本管理。视频分析层对视频数据进行解析和抽帧处理,通过任务编排模块调用 AI 算法仓库中的预置能力[5],完成对视频内容的理解和告警事件的输出,对 AI 算法与场景进行灵活编排,实现 AI 算法的跨场景高效复用。

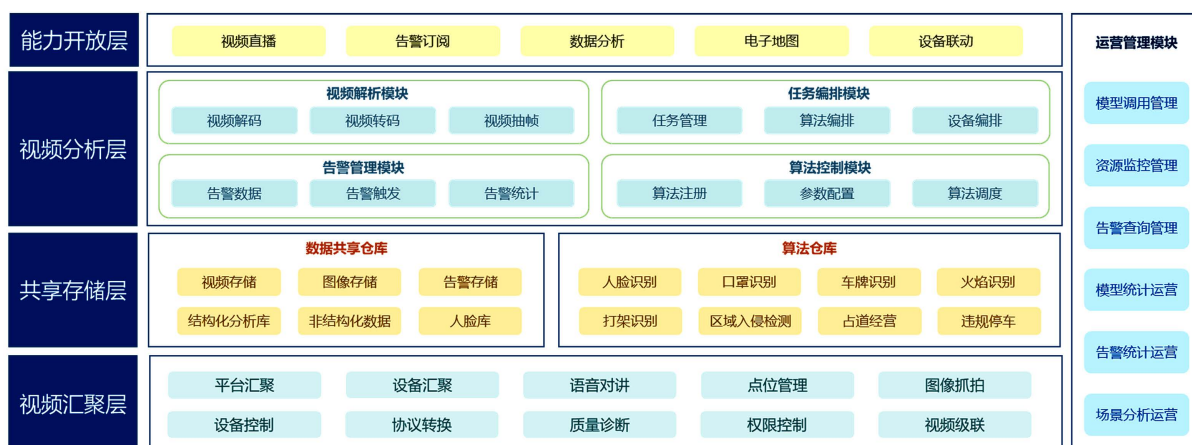


Figure 1. Functional architecture diagram of the video AI analysis platform

图 1. 视频 AI 分析平台功能架构图

视频 AI 分析平台将抽帧、编解码、AI 分析等任务进行容器化封装,实现了算法、算力和调度模块的深度解耦,借助算力网络编排调度技术,完成云边端分布式协同部署,使得平台能够按需混合云和跨域资源池完成部署,提高了业务灵活性。通过运营管理模块实现对智能视频分析全过程管理,将 AI 算法调用量等数据进行全面记录,结合历史调用数据可以及时发现摄像头遮挡、服务中断等异常情况。通过资源监控管理模块实时监控系统资源状态,根据任务量和资源状态完成任务调度和资源弹性伸缩,在夜间或其他非高峰时段增大视频分析事件监控,降低算力资源占用量和系统能耗。

2.2. 产品技术架构

如图 2 所示,视频 AI 分析平台整体技术架构包含应用端、能力调用服务、数据同步服务和视频处理服务四大部分。首先通过在中心和边缘集群分布式部署的视频处理服务,对前端摄像头回传的视频流数据进行解码和抽帧,将采集到的待识别图片存入 redis 集群,同时将分析任务的关键信息存入 rabbitMQ 等待处理。数据同步服务将待执行的任务和图像实时同步至主节点,然后通过能力调用服务完成智能视频分析。

在视频处理服务中,包含了多个工作节点(worker),根据任务信息完成视频拉流、解码和抽帧,然后将抽帧图片存储到 redis 集群,工作节点可通过容器化方式实现弹性伸缩,支持批量部署操作,降低了任务与设备的耦合性,应用端可以根据计算、网络和存储资源的消耗情况实时选择合适的边缘集群进行任务的部署和调度。视频处理服务中的 Server 端负责管理该节点下的硬件资源,从应用端接收到任务后,完成工作节点的任务调度。

数据同步服务主要负责同步各个中心集群和边缘集群中的图像数据和任务信息,使用 rabbitMQ 消息队列进行任务管理,实时将应用端接收到的任务分发到各个独立运行的视频处理服务中进行解析和处理,中心集群及地市集群通过中间件解耦的方式保障系统鲁棒性和健壮性,通过数据同步服务完成集群间的通信和容灾备份。

针对告警信息,平台分别通过数据上报接口和数据查询接口对接第三方系统,对于告警上报接口,通过 MySQL 进行数据存储,保障告警数据及时上报,对于非告警类分析结果和调用记录等数据,平台通过 ES 集群进行存储。ES 集群对全量告警信息和日志数据进行存储,通过任务列表建立索引,实现数据查询和统计分析,将视频处理任务和数据处理任务进行解耦,保证视频处理任务的实时性和数据处理任务的安全性。针对告警图像存储需求,平台结合 MySQL 和 Go-FastDFS 数据库,实现对海量抽帧图像、告警图像和视频数据的分级和高效存储。其中 MySQL 数据库用于存储图像和视频的基本信息和元数据,而 Go-FastDFS 用于存储图像和视频文件。

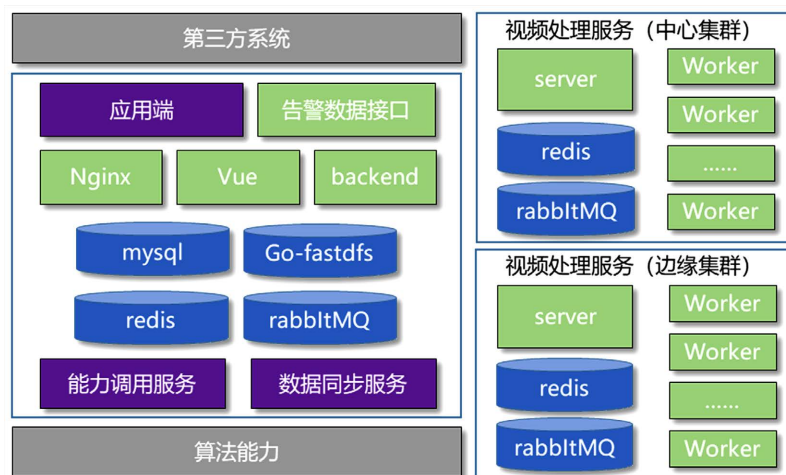


Figure 2. Technical architecture diagram of video AI analysis platform
图 2. 视频 AI 分析平台技术架构图

3. 大规模视频分析技术实现方案

在海量视频接入情况下,视频 AI 分析任务需要消耗大量的算力资源和网络资源。视频 AI 分析平台通过对视频处理、智能分析和告警分发查询等环节进行性能优化,满足用户对百万路视频完成实时分析的业务需求。

3.1. 视频解析中心

视频解析过程需要完成视频流的解码、存储和抽帧等任务,其中抽帧过程需要消耗大量算力资源,导致平台出现性能瓶颈。通过对主流视频监控设备进行分析,发现大部分视频监控设备在视频传输过程中,均采用了 H.264 格式对视频进行编码和压缩,从而节省视频传输过程中的带宽占用。其中主要采用了 GOP 内压缩技术,将视频划分为帧内编码帧(I 帧)、前向预测编码帧(P 帧)和双向预测帧(B 帧)。

利用常规方式对视频进行解码,首先需要搜索 I 帧,然后再对下一个 P 帧进行解码,最后再对中间的 B 帧按顺序依次解码,具体解码顺序如图 3 所示。其中 I 帧包含了图像的完整信息,因此在解码过程中,仅对 I 帧进行解码可以有效节省算力资源,单任务平均 CPU 占用率可以下降到 2%左右,相比于全量解码(CPU 平均占用率 20%)可以节省 90%的算力资源消耗。

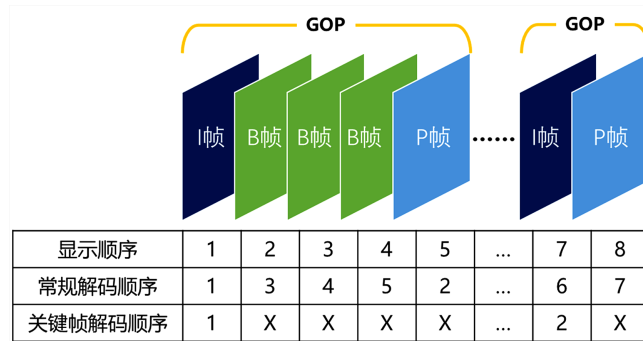


Figure 3. Key frame decoding flow diagram

图 3. 关键帧解码流程示意图

3.2. 任务编排中心

视频 AI 分析平台通过云边协同的分布式架构(如图 4 所示), 实现了 AI 能力快速编排、部署和调度, 支撑业务快速开通和能力复用。通过云侧应用层的控制节点统一管理和调度任务, 通过 Kubernetes 将控制流和数据流进行解耦, 云侧部署的中心节点进行任务的分发和流量控[6] [7], 边缘侧子节点完成抽帧解码处理后将数据实时同步至云侧进行智能分析, 可以降低平台中心节点的负载。通过云边协同处理的方式, 可以将子节点分地域进行部署, 在任务分发过程中可以根据网络情况实时选择视频监控设备所接入的子节点。通过中心节点实现视频解析任务和算力调用任务的灵活编排, 通过将省级云节点和地市边缘子节点实现跨集群协同调度, 实现算力资源弹性伸缩。

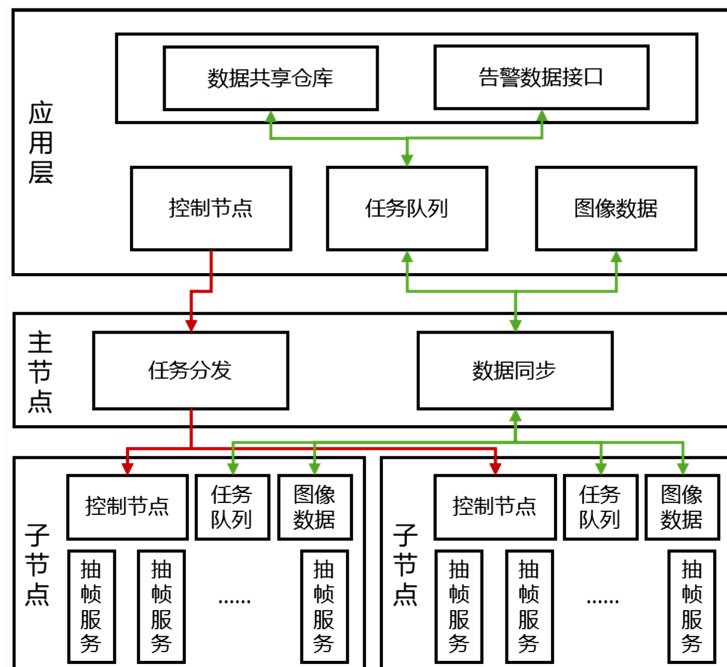


Figure 4. Technology architecture of video analysis center

图 4. 视频解析中心关键技术架构

此外, 不同场景下的视频分析任务主要活跃时间不同, 例如客流统计任务[8]在上下班高峰期需要消耗更多的算力, 违规停车任务主要应用在住宅区附近道路的夜间违停。平台通过潮汐调度根据不同任务

的特点，灵活控制视频分析的周期和时间段，通过合理的调度可以实现算力需求的削峰填谷，提升平台资源利用效率。

3.3. 智能感知中心

智能感知中心支持结构化告警数据查询，告警事件订阅和基于过滤规则的智能推送[9]，告警样本数据自动回流。如图 5 所示，智能查询部分主要基于预训练图生文大模型以及目标检测大模型进行图像理解，将视频数据转化成结构化信息，其中包含图像中的物体类型、数量、位置、颜色、纹理等信息，以及图像中的事件信息。查询过程中可以通过以图搜索的方式，对比输入图像和视频内容的结构化数据之间的相似度，返回相似度较高的图像。此外，也可以通过自然语言交互方式实现告警事件的高效查询，基于大语言模型将查询内容转化为结构化信息进行检索。智能告警部分，可以根据用户设定的告警规则，对事件库中的图像进行二次过滤，实现告警结果的智能化过滤和按需推送。

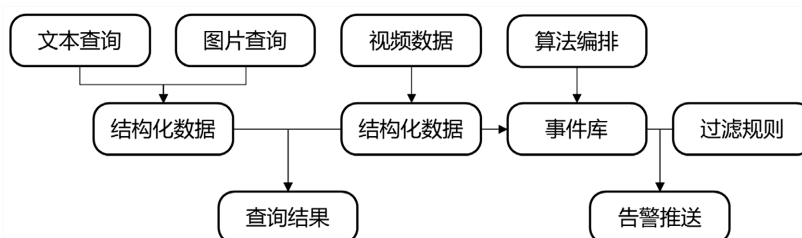


Figure 5. Service processing flow of the intelligent awareness center
图 5. 智能感知中心业务处理流程

3.4. 算法仓库

平台通过算法仓库实现了对视频分析类算法的统一管理[10]，支持云边算法统一管理以及云边算法一键迁移。基于浪潮自研的容器管理平台为云端业务提供了镜像管理和负载均衡等能力，通过 Kubeedge 纳管边缘端业务应用。目前基于本平台的容器化方案，已经兼容了大部分主流边缘算力盒子，实现对边缘设备的统一管理能力，在云端可以对边缘设备上部署的算法进行一键更新[11][12]。平台根据业务需求远程下发算法模型，并对其进行权限和版本等管理，已适配主流人工智能加速芯片，支持部署和调用多种主流深度学习框架开发的计算机视觉类算法模型。目前实现对算法的授权管理以及算法模型的升级和更新等操作，通过本平台提供的通用算法适配工具[13]，将云端训练好的算法模型一键转换为目标边缘设备可运行的版本。基于容器化技术通过“一键自发现”功能实现边缘节点的自动搜索和加载，通过“算法一键部署”功能选择所需的算法插件直接下发至边缘端，满足个性化 AI 算法需求。

4. 总结

本文主要针对海量视频智能分析场景，设计了一套可以支撑百万量级分析任务的视频 AI 分析平台。其中视频解析中心通过关键帧解码技术，大幅提升了图像抽帧效率，任务编排过程中，通过云边协同的方式，将视频抽帧解码和智能分析节点下沉到边缘侧，降低中心节点负载。最后在智能感知中心实现了视频数据的结构化分析，通过以图搜图或者自然语言查询的方式，完成数据检索和查询，通过配置告警规则可以有效过滤多余告警。与现有视频智能分析系统相比，本文所设计和实现的视频 AI 分析平台对整体架构进行了分层解耦，通过中心和边缘节点协同处理的方式，整体接入量和资源消耗相比文献中的方法有较大的性能提升。本文研发的大规模视频分析技术可以有效提升视频监控系统的智能化水平，为智慧城市、城市治理和社会监管等领域提供了一种更有效的技术手段，赋能非现场式监管和无接触式执法，

建立城市管理新模式。

参考文献

- [1] 黄凯奇, 陈晓棠, 康运锋, 等. 智能视频监控技术综述[J]. 计算机学报, 2015, 38(6): 1093-1118.
- [2] 郑雪原. 城市级公共安全视频图像智能分析平台系统架构设计实践[J]. 电子技术与软件工程, 2023(7): 208-211.
- [3] 唐志鸿, 王宏图, 徐若龙. 视频与物联网大数据融合分析应用平台[J]. 江苏通信, 2019, 35(3): 61-63, 73.
- [4] 叶逢春, 杨兴浩. 基于视频智能化分析的智慧城管平台[J]. 电子技术与软件工程, 2021(12): 154-155.
- [5] 聂志勇, 孙占冬. 基于深度学习的智能视频监控系统应用研究[J]. 能源科技, 2023, 21(1): 9-13.
- [6] 王佳, 吴任博, 周小光, 等. 基于云边结合技术的电网智能视频云架构设计与测试[J]. 粘接, 2023, 50(10): 193-196.
- [7] 李成华, 石胜涛, 李孝天, 等. 基于边缘算力协同系统的视频智能分析任务动态调度方法[J/OL]. 电子与信息学报, 2023: 1-11. <http://kns.cnki.net/kcms/detail/11.4494.tn.20230807.1638.003.html>
- [8] 李宇杰, 裴中阳, 李强, 等. 智能视频分析技术在地铁站客流监测中的应用研究[J]. 现代城市轨道交通, 2022(3): 86-92.
- [9] 任鹏飞. 城市级视频监控系统智能分析平台的接口设计[D]: [硕士学位论文]. 南京: 南京邮电大学, 2023.
- [10] Paszke, A., Gross, S., Massa, F., Lerer, A. and Chintala, S. (2019) PyTorch: An Imperative Style, High-Performance Deep Learning Library. <https://arxiv.org/pdf/1912.01703v1.pdf>
- [11] 马艳军, 于佃海, 吴甜, 王海峰. 飞桨: 源于产业实践的开源深度学习平台[J]. 数据与计算发展前沿, 2019, 1(1): 105-115..
- [12] 束柬, 陈剑波. 深度学习平台体系架构及其关键技术[J]. 计算机应用研究, 2023, 40(11): 3353-3357.
- [13] 黄巨涛, 郑杰生, 高尚, 等. 基于云平台的人工智能开源开发平台框架研究[J]. 自动化与仪器仪表, 2020(7): 159-162, 166.