面向大规模智能监控系统的算力共享网络解决方案

作者 姜云杰 吉莉 熊重驰



# 摘要

面向大规模智能监控系统的算力共享网络是面向大规模智能监控系统发展的主旋律，基于边缘计算的算力共享网络在未来大规模视频监控预警市场占绝对优势。随着“智能管道 智慧管网”的迅速推进，将安防监控与运营商的网络、边缘计算、云技术相结合，为视屏监控云网融合、云边协同带来可能。

智能监控系统作为实现“智能管道 智慧管网”系统的重要组成部分受到了社会各界的重视。为提高监控系统效率，降低风险漏报率，面向大规模智能监控系统的算力共享网络已经成为迫切需求。本文聚焦于大规模智能监控系统所面临的设备低电量导致计算能力不足、计算核心故障导致的风险漏报等问题。

针对设备低电量导致的计算能力不足问题，提出一种算力共享网络的架构，围绕着外部供电计算核心部署和分配计算资源，克服传统部署方式在设备低电量时计算核心无法获得足够的工作电压而导致计算能力不足的问题。通过部署区域外部供电计算节点处理区域内因低电量导致无法正常工作的节点的视频数据，部署和新阶段实现区域计算节点的统一调度和管理，实现基于区域和视频源的视频计算资源部署和分配模式。受影响的计算节点只需将摄像机记录的信息片段发送到核心节点，有效解决计算能力不足的问题。

针对计算核心故障导致的风险漏报问题，提出一种算力替代策略。针对监控视频价值密度低的特点，提出一种自适应的视频丢帧策略和基于视频重要程度的替代计算方法。当设备计算核心故障无法响应时，根据丢帧策略选择性丢弃部分摄像头的视频帧，减小其对计算资源的消耗，并使用通讯模块将视频传输至外部供电计算核心代替计算，在保证整个监控系统稳定的同时减少风险漏报的问题。

本文基于以上提出的计算架构与方法，探讨面向大规模智能监控系统的算力共享网络的可行性、有效性和实用性。

# 课题研究背景及意义

长输油气管道安防主要依靠人力巡检巡查，西气东输每年投入的人力巡护成本较高，为了响应国家油气体制改革要求，西气东输公司率先开展了“智能管道 智慧管网”的探索与建设，并在管道智能监控感知威胁态势方面取得了实质性的突破。视频监控技术作为实现“智能管道 智慧管网”的重要技术手段也越来越受到社会各界的重视，同时在各个行业得到了广泛的应用。2010年，为了顺利、安全地举办世博会，上海市大规模增加监控网络的范围和密度，仅在浦东新区就建设了包含一万多个监控点位的高清监控系统；截至2011年，黑龙江投入8亿多建设资金用于监控系统建设，共建设近千个监控中心，二十多万个监控点；武汉市为推进“平安城市”建设，建设了覆盖跨江大桥、主干道和主城区城道口等重点部位的监控系统。

长输油气管道安防主要依靠人力巡检巡查，西气东输每年投入的人力巡护成本较高，为了响应国家油气体制改革要求，西气东输公司率先开展了“智能管道 智慧管网”的探索与建设，并在管道智能监控感知威胁态势方面取得了实质性的突破。视频监控技术作为实现“智能管道 智慧管网”的重要技术手段也越来越受到社会各界的重视，同时在各个行业得到了广泛的应用。2010年，为了顺利、安全地举办世博会，上海市大规模增加监控网络的范围和密度，仅在浦东新区就建设了包含一万多个监控点位的高清监控系统；截至2011年，黑龙江投入8亿多建设资金用于监控系统建设，共建设近千个监控中心，二十多万个监控点；武汉市为推进“平安城市”建设，建设了覆盖跨江大桥、主干道和主城区城道口等重点部位的监控系统。

视频监控系统的发展主要分为以下三个阶段：

第一阶段：以模拟摄像机及模拟录像机为主的模拟视频监控系统，这种监控系统多是以摄像机、录像机、分割器为核心，通过模拟信号输出视频，传输距离短，图像质量差，目前已逐步退出市场。

第二阶段：基于模拟摄像机和数字录像机的视频监控系统，采用模拟信号传输视频，数字方式处理与贮存，属于过渡阶段。

第三阶段：基于网络摄像机的全数字视频监控系统，以数字方式传输视频，传输距离长，图像质量好，目前正得到广泛应用。

以上三代视频监控系统，主要解决视频捕捉、传输、存储等问题，并不具备视频职能分析的功能，真正起到安全防范的作用则需要人为观察视频从而做出安全评估。一种方法是实时观看，建设一个监控中心，将监控点所有的视频流导入监控中心，同时配备多块显示屏来播放视频画面，由工作人员通过观看屏幕上的监控点的视频，事后进行调阅，本质上还是以人观看的方式来查找感兴趣的信息。

然而人工监控具有本身固有的缺陷，例如：消耗大量人力、易使人产生疲劳、可扩展差等等。尤其随着“智能管道 智慧管网”监控系统规模不断扩大，面对数以万计的监控点和海量视频数据，以人工为核心的监控系统越来越无法满足需求。如何解决人工监测带来的种种问题，加强监控系统的事前预警和事后的检索，成为当前摆在“智能管道 智慧管网”发展前的重要问题。

监控界提出了智能视频监控(Intelligent Video Surveillance, IVS)的概念。智能监控即使用计算机强大的计算能力以及图像分析技术对传统的监控视频进行处理、分析和理解，过滤无用信息。目前智能视频监控在计算机视觉领域中备受关注。

相比于传统监控系统，智能监控系统具有以下优势：

1. 全天候的监控能力：使用计算机代替人来监控视频画面，实现对监控画面的不间断分析，彻底改变以往由监控人员对监控画面进行监视和分析的方式。
2. 快速的响应能力：通过设置某些可疑行为的识别规则，在其发生危害之前通知相关人员采取相应措施，为潜在威胁做好准备工作。
3. 扩展视频资源的应用领域：传统视频监控系统只能获取单一的视频数据，而智能监控系统可以提取视频中的各种信息，为其他应用提供支持。

随着“智能管道 智慧管网”监控系统规模的不断扩大，面向大规模监控系统的视频计算对智能监控系统提出了新的挑战：

1. 监控设备电量低导致的算力不足问题。视频监控节点通常部署在野外，由太阳能供电，而监控视频流数据量大，计算核心需要在稳定的工作电压下才能高效运算，如果出现电量不足的情况，计算核心得不到稳定的工作电压，将无法负荷繁重的计算任务，这将导致设备反应缓慢，遇到风险时无法及时汇报。如何解决低电量算力不足问题，是智能监控系统需要解决的首要问题；
2. 计算核心故障导致的风险漏报问题。视频监控摄像头每分每秒都在向计算核心输出大量的视频流数据，计算核心任务十分繁重。当计算核心出现故障（如死锁、过热和短路）时，该区域的风险监控将瘫痪，此时若存在风险，势必会出现风险漏报问题，后果不堪设想。如何保证智能监控系统的稳定性，杜绝因计算核心故障而导致的风险漏报也是亟待解决的问题；

本文针对以上两个问题进行研究，有利于实现智能监控，推进“智能管道 智慧管网”建设， 对公共安全、社会稳定等具有重要意义。

# 国内外发展研究现状

基于边缘计算的智能城市实时监控视频分析目前引起了从业者和研究人员的广泛关注，大量的视频数据被提交到互联网。根据原始视频内容，如何删除冗余视频帧、正确分割视频序列并减少不必要的计算资源消耗是一项具有挑战性的工作。X. Zhang等人提出了一种检测和定位具有运动场形状的视频异常的方法，S. Ding等人利用改进的 Harris-Laplace 时空兴趣点从大视频中识别有趣的片段，然后从中选择关键帧。时空兴趣点可以准确地附着在检测到的目标周围，兴趣区域构建算法可以快速准确地定位包含目标的候选区域。由于图像分类和目标识别的巨大成功，特别是视频监控、目标计数和目标检测，深度学习受到了广泛关注。相比之下，Y. Fang等人提出了空间和时间特征图的计算，然后检测视频显著性。由于资源限制（能源消耗、计算和内存），在智能终端设备上部署这些具有低延迟要求的大型、强大的视频任务仍然是一个挑战。因此，考虑将这些计算任务转移到更强大的边缘服务器或云。然而，云计算模型不适用于需要短期执行的边缘服务，因为将任务卸载到云中心增加了网络往返传输延迟，应用服务请求不会很快得到响应，但是充分利用云中强大的计算和内存资源将减少总响应时间。由于边缘节点和边缘服务器离用户近，可以快速响应用户请求，成为首选的帮手。在边缘服务器上运行计算密集型任务时，需要对多个终端设备资源进行有效管理。目标是平衡精度、能耗、延迟和负载平衡的性能参数。 VideoStorm引入了这些权衡，以在满足准确性和延迟目标的前提下为每个请求选择正确的配置。例如，在 Chameleon中，配置在流视频输入期间在线更新。然而，基于任务划分的不同模型分割点会导致不同的计算延迟。因此，有必要选择合理的切割策略，以最大限度地发挥端边甚至云协同的优势。在将任务卸载到边缘服务器时，我们可以在边缘进行数据预处理，减少冗余、带宽、延迟和对云中心的依赖，同时提高视频分析的效率。为了减少带宽消耗，有学者提出了端边云协同架构和模型压缩，以消除不同环境下的数据传输。例如M. Song等人提出只有从边缘设备推断出的数据传输到云端进行再训练，以减少数据传输，建议在不影响精度的情况下去除冗余数据以减少数据传输。Glimpse 将所有 DL 计算任务迁移到最近的边缘服务器，同时使用更新检测来删除应该卸载的相机帧。如果检测没有变化，将在本地进行帧跟踪。这种过滤增强了系统的处理能力并使移动设备上的实时目标检测成为可能。Vigil提出了一种分布式架构，它巧妙地利用边缘和云之间的处理任务来减少视频监控中的带宽消耗。类似地，VideoEdge提出了边缘和云的分层架构来处理摄像机流，以便在多个目标和约束之间实现更好的权衡，公平地分配资源。我们在智能视频监测系统中提出了一种支持边缘计算的算力共享网络，该网络能够帮助边缘节点之间进行相互协调，实现计算任务卸载、计算任务调度以及能效优化等方面，保证边缘节点正常工作。

# 课题需求分析

在大规模智能视频监控系统的背景下，虽然现存网络利用边缘计算技术，缓解了云中心计算延迟高、带宽不足、安全性低的问题，实现了“就地、就近”提供服务的功能，但是对于智能边缘计算设备所存在的不稳定情况（如：产生死机或掉线等），依然会对整体网络造成巨大影响，一旦出现以上情况，可能会造成该终端节点所连接摄像机中视频数据的丢失，那么对于该监控区域来说，必定存在一定风险，因为智能边缘计算设备无法进行实时的检测和报警工作。

除此以外，考虑到一些智能边缘计算设备因为条件不足，采用太阳能板进行供电。但是如果出现连续阴雨天气导致无法为太阳能板持续提供能量，那么边缘计算节点面临着计算能力下降的问题。在供电不足的情况下，无法对所收集到的视频数据提供正常的处理，往往会出现对异常情况的漏报。针对以上智能边缘计算设备的不稳定性和计算能力不足问题，我们提出以下需求分析：

1. 规避计算能力不足及漏报风险：当边缘计算节点的能量产生大量消耗时，容易导致计算能力下降，可以借助算法共享方法，使得当前节点只需执行少量本地任务，利用丢帧策略将剩余视频数据任务传输给距离当前节点最近的边缘设备帮助计算，要求这些边缘设备是连接电源的稳定节点。除此之外，当边缘计算节点的计算核心出现故障，无法支持自我计算，可能会产生漏报风险，那么此时，边缘计算节点只需要维持传输功能，将所有的视频数据信息传输给最近的稳定节点进行计算，以避免漏报情况。
2. 实现模块化的功能处理：在当前的视频监控系统中，智能边缘计算的“黑盒子”通过一块芯片实现存储、计算和通信等功能，这对于设备的耗能来说，是非常巨大的。因此，为帮助实现后期持续发展，可以采用基于功能的芯片模块化，对于存储、计算和通信三个方面，分别采用不同芯片提供相关功能操作，这样可以避免不同方面所造成的干扰，让每一模块的芯片性能（尤其是计算性能）达到最优。
3. 增强智能边缘计算节点的可靠性：从物理安全方面来说，因为边缘计算节点往往部署在室外环境，缺少一定的保护措施，无法防止一些非法人员在物理设备上进行一些恶意破坏，因此需要对边缘计算节点提供物理上的保护，保证其可靠性。从数据安全方面来说，在将弱性能智能边缘设备的视频数据发送给最近的高性能智能边缘设备时，应当考虑由边缘计算的开放特性所带来的安全与保护问题，避免数据暴露等隐私风险。

# 解决方案

为了实现泛在计算和服务的感知、互联和协同调度，CAN架构体系从逻辑功能上可分为算力服务层、算网管理层、算力资源层、算力路由层和网络资源层。其中，算力路由层包含控制面和转发面，如图3所示。

·算力应用层：承载泛在计算的各类服务及应用，并将用户对业务SLA的请求（包括算力请求等）参数传递给算力路由层。

·算力管理层：完成算力运营、算力服务编排，以及对算力资源和网络资源的管理。该层的具体工作包括对算力资源的感知、度量，以及OAM管理等，实现对终端用户的算网运营以及对算力路由层和网络资源层的管理。

·算力路由层：是CAN的核心。基于抽象后的算网资源，并综合考虑网络状况和计算资源状况。该层可以将业务灵活按需调度到不同的计算资源节点中。

·算力资源层：利用现有计算基础设施提供算力资源。计算基础设施包括单核中央处理器（CPU）、多核CPU，以及CPU+图形处理器（GPU）+现场可编程门阵列（FPGA）等多种计算能力的组合。为满足边缘计算领域多样性计算需求，该层能够提供算力模型、算力应用程序编程接口（API）、算网资源标识等功能。

·网络资源层：利用现有的网络基础设施为网络中的各个角落提供无处不在的网络连接，网络基础设施包括接入网、城域网和骨干网。

其中，算力资源层和网络资源层是CAN的基础设施层，算网管理层和算力路由层是实现算力感知功能体系的两大核心功能模块。基于所定义的五大功能模块，CAN实现了对算网资源的感知、控制和调度。

总之，作为计算网络深度融合的新型网络，CAN以无所不在的网络连接为基础，基于高度分布式的计算节点，通过服务的自动化部署、最优路由和负载均衡，构建算力感知的全新的网络基础设施，真正实现网络的无所不达、算力无处不在、智能无所不及。海量应用、海量功能函数、海量计算资源则构成一个开放的生态。其中，海量的应用能够按需、实时调用不同地方的计算资源，提高计算资源利用效率，最终实现用户体验最优化、计算资源利用率最优化等。

3.1 算力度量和建模

如何构建统一的算力模型是CAN的研究基础。基于算力统一的度量体系，通过对不同计算类型的异构算力资源进行统一抽象描述，形成算力能力模板，可以为算力路由、算力设备管理、算力计费等提供标准的算力度量规则。首先，异构硬件设备通过统一的算力度量和建模，实现对现场可编程门阵列（FPGA）、GPU、CPU等异构物理资源的统一资源描述，从而可以有效地提供计算服务。其次，考虑到计算过程受不同算法的影响，需要对不同的算法如人工智能（AI）、机器学习、神经网络等算法所需的算力进行度量，更有效地了解应用调用算法所需的算力，从而服务于应用。最后，由于用户的不同服务会产生不同的算力需求，需要把用户需求映射为实际所需的算力资源，从而可以使网络更充分有效地感知用户的需求，提高和用户交互的效率。

3.2 算力路由关键技术

算力路由层是CAN的核心功能层，支持对网络、计算、存储等多维资源、服务的感知与通告，从而实现“网络+计算”的联合调度。算力路由层包括算力路由控制技术和算力路由转发技术，这两种技术可以实现业务请求在路由层的按需调度。

算力路由控制面可以通告算力节点的信息并生成算力拓扑，进而生成算力感知的新型路由表。算力路由控制面基于业务需求生成动态、按需的算力调度策略，实现算力感知的算网协同调度。算力路由转发需要通过IP协议/SRv6扩展增强实现网络感知应用、算力需求以及随路OAM管理等功能。算力路由支持网络编程、灵活可扩展的新型数据面，能够实现算力服务的最优体验。

3.3 算力管理关键技术

算力管理包含算力设备的注册、OAM、运营等。统一的管理面可以对网络和算力进行管理和监测，并可生成算力服务合约以及计费策略，以实现对算力的统一运营。基于统一的算力度量体系，通过对不同计算类型进行统一抽象描述，算网管理层能够形成算力能力模板，从而为算力设备的管理、合约和计费以及OAM提供标准的算力度量规则。算力注册需要实现对算力节点的注册、更新和注销，并对相应的路由通告策略进行管理。算力OAM需要实现对算力资源层的算力性能监测控制、算力计费管理等。