

浙江省通信技术前沿论坛

边缘智能高通量视觉处理算法与系统

殷海兵

杭州电子科技大学

2019年11月16日



汇报内容

一

研究背景

二

研究内容

三

实验结果

四

总结与展望

研究背景

监控视频应用广，数据体量大，对智能分析提出巨大挑战



公安：
搜人、布控



综治：
群体感知

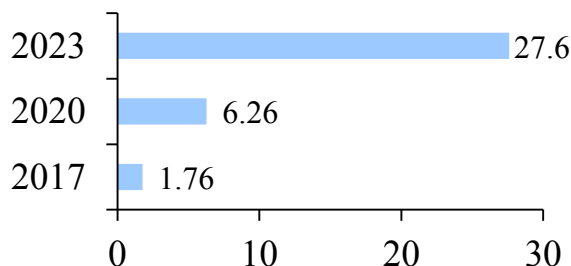


交通：
治堵、搜车

应用领域广泛

数据来源：中投顾问咨询公司

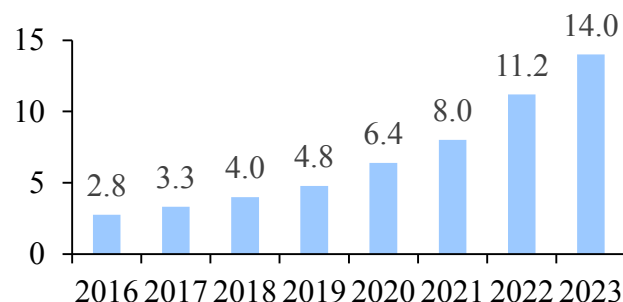
2017-2023年监控摄像头数量预测 (单位：亿台)



摄像头数量增长迅猛

数据来源：IDC咨询公司

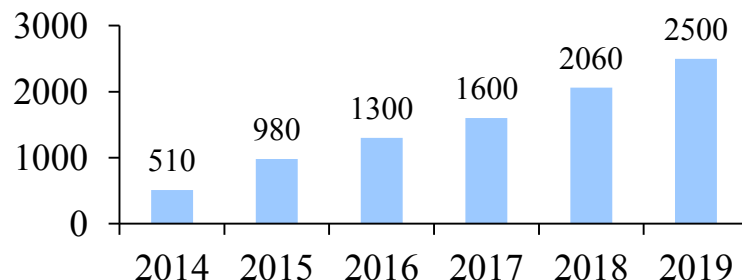
2018-2023年监控视频市场规模预测 (单位：百亿美元)



市场规模巨大

数据来源：中商产业研究院

2014-2019年新增监控摄像头每天产生的视频数据量预测 (单位：PB)



视频数据量激增

数据来源：IHS咨询公司

研究背景

国家战略需求：视频大数据的高效利用（公安、综治、交通）

2016：白银案，横跨28年，11人遇害。



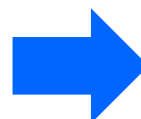
2014：上海外滩踩踏事件造成36人死亡，49人受伤。



2018：北京后厂村路500m堵1小时。



公安：搜人、布控
综治：群体异常感知
交通：搜车、治堵



视频分析理解判断

研究背景

为实现海量摄像头数据智能协同感知
须实现高效视频压缩、智能分析和高效计算



周克华

重庆两次作案：
2004年04月22日
2005年05月16日



2012年08月10日

南京一次作案：
2012年01月06日



昆明一次作案：
2005年10月16日



长沙四次作案：
2009年10月14日
2009年12月04日
2010年10月25日
2011年06月28日

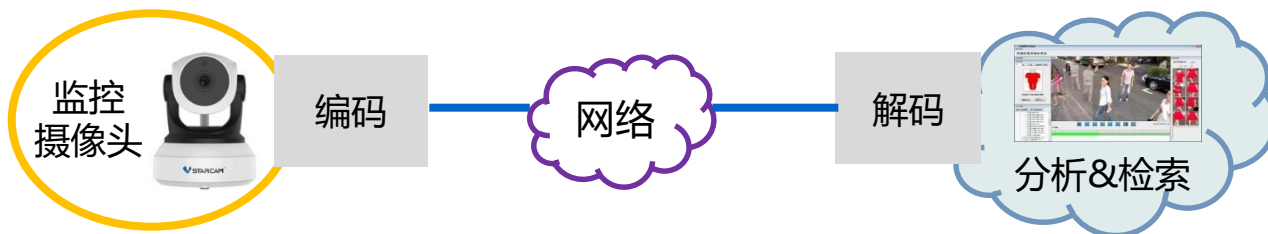
周克华案破获

仅南京就抽调数千名民警进行海量(3500T)监控视频数据人工甄别和查找,相当于看约80万部电影。

海量摄像头视频数据的实时汇聚、分析、感知国内外还处于探索阶段

研究背景

传统终端编码/云端分析视频计算体系存在内在缺陷



1. 数据汇聚带宽巨大
200Gbps
 2. 视频数据存储
17PB/天
- (以中等城市10万高清摄像头为例)

瓶颈

难存储

1. 终端编码失真导致云端分析准确率下降
 2. 算法面临挑战：成像/光照条件、拍摄角度
- 小目标 模糊尺度变化 目标遮挡 场景迁移
(以人脸识别为例下降**60%***)

难检索/难识别

图像特征提取和搜索
计算开销：**450万核/s**
(以中等城市10万高清摄像头为例)

计算速度慢
(成本高)

缺乏统一智能视觉行业标准
(不同平台互联)

平台互联互通难

*W. Gao, et.al, the IEEE 1857 standard empowering smart surveillance systems. *IEEE Intelligent systems*. 2014.



研究背景

□ 问题和需求

➤ 视频大数据整体架构

问题：单路视频智能化成本高，网络传输带宽大，存储成本高

急需：（1）新的视频大数据技术体系：高效端云协同计算架构
（2）软件定义摄像机技术体系

➤ 边缘计算前端设备

问题：前端设备算力有限，智能视频算法性能待提高

急需：（1）高性价比嵌入式计算平台（芯片）
（2）高效智能视频分析算法(检测、跟踪、ReID)

➤ 存储和传输系统

问题：海量摄像头网络带宽大，存储成本太高，有效数据比例低

急需：（1）高效视频编码
（2）智能特征编码
（3）多层次多粒度视频流（视频、图片、半结构化、结构化特征）



研究背景

□ 问题和需求

➤ 大数据分析

问题：百亿级数据响应、系统轻量化、数据挖掘可视化

急需：(1) 高效特征编码标准 (2) 高效视觉特征检索技术

➤ 行业标准

问题：不同厂商平台数据难融合，不同行业不同安全等级数据难融合

急需：视觉特征编码和交换、神经网络表示与模型压缩、高效智能安防场景视频压缩、以及不同厂家设备互联数据交换等标准

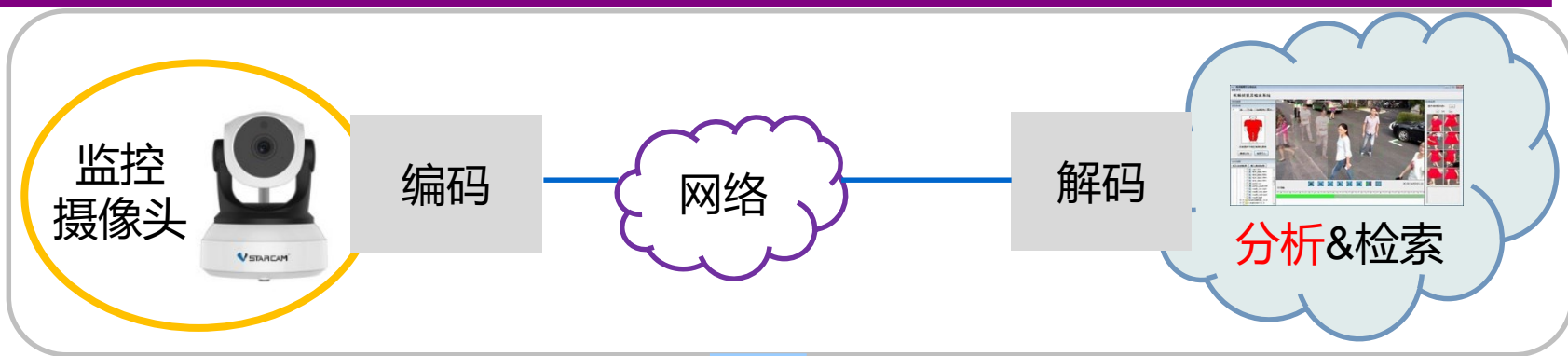
➤ 数据安全

问题：视频数据不够安全

急需：非像素级视觉特征提取/表示/检索技术

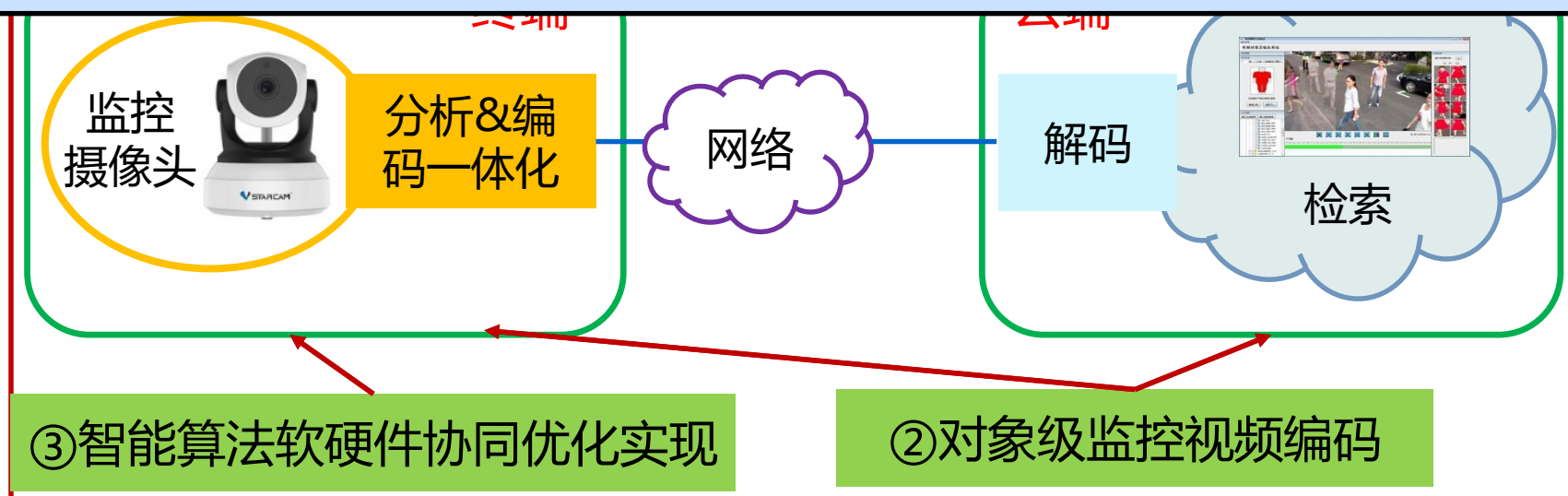
研究背景

终端编码
云端分析



目标：实现分析与编码为一体的端云协同视频智能计算，构建多目标协同优化的算法验证平台

协同视频智能计算





汇报内容

一

研究背景

二

研究内容

三

实验结果

四

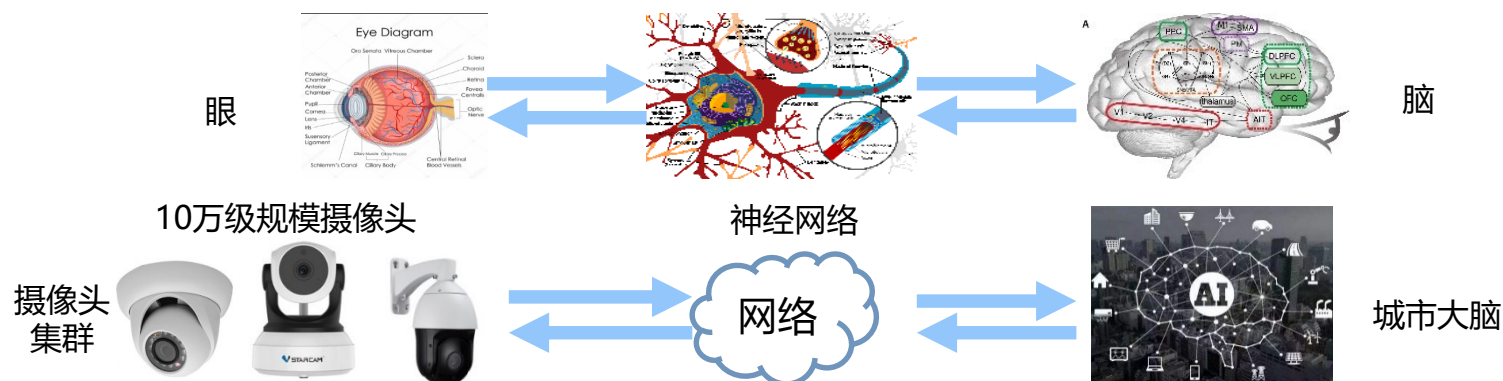
总结与展望

1.1 数字视网膜系统



Martha Farah教授
University of Pennsylvania

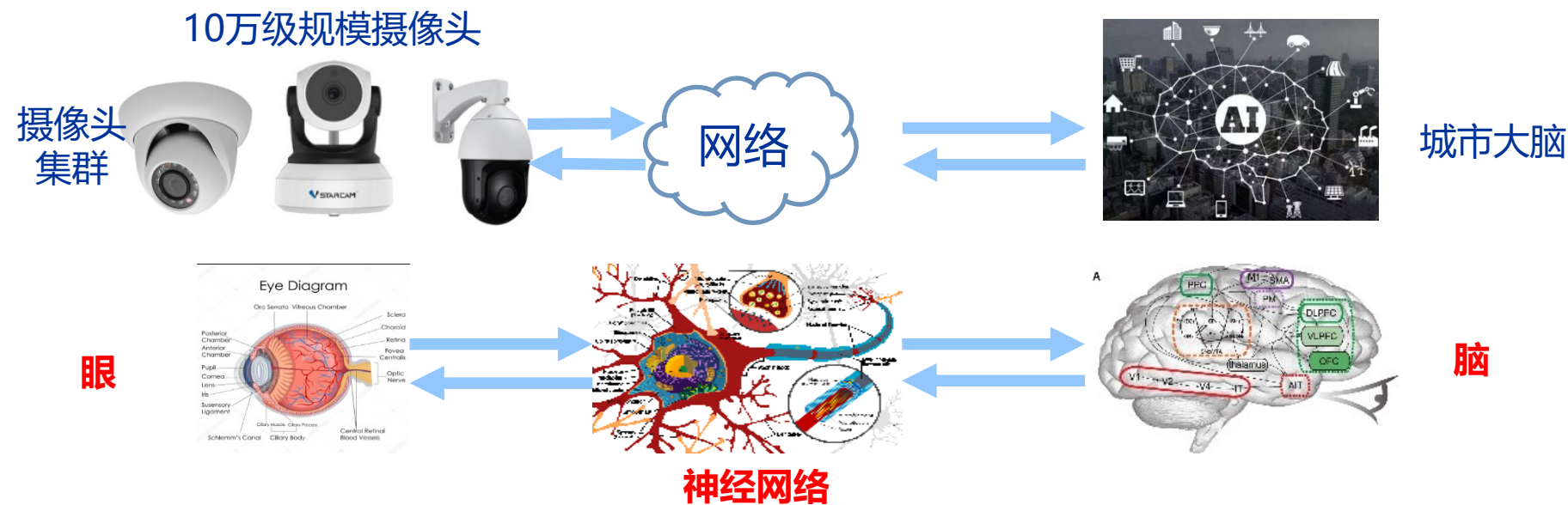
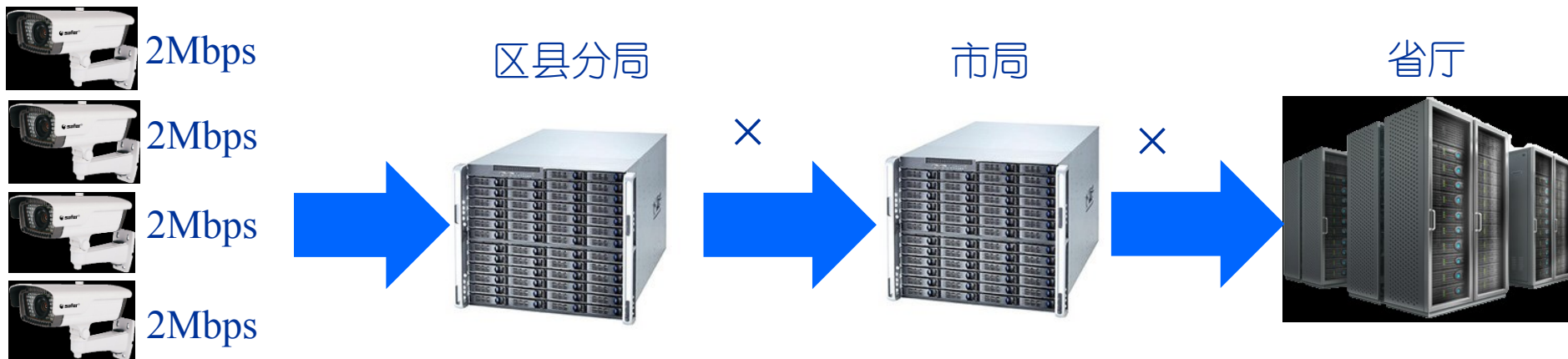
人类视网膜大约拥有十亿视觉细胞，使得人眼成为十亿像素的高清相机。但是人眼和大脑皮层的连接神经元细胞只有一万个，**人眼怎么通过窄带带宽传输高达十亿的视觉信号？** 类比，**城市级海量监控视频信号怎么通过网络汇聚？**





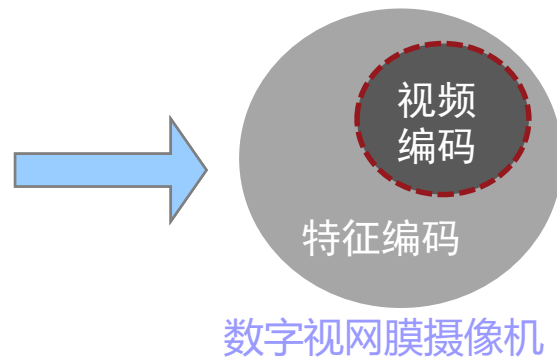
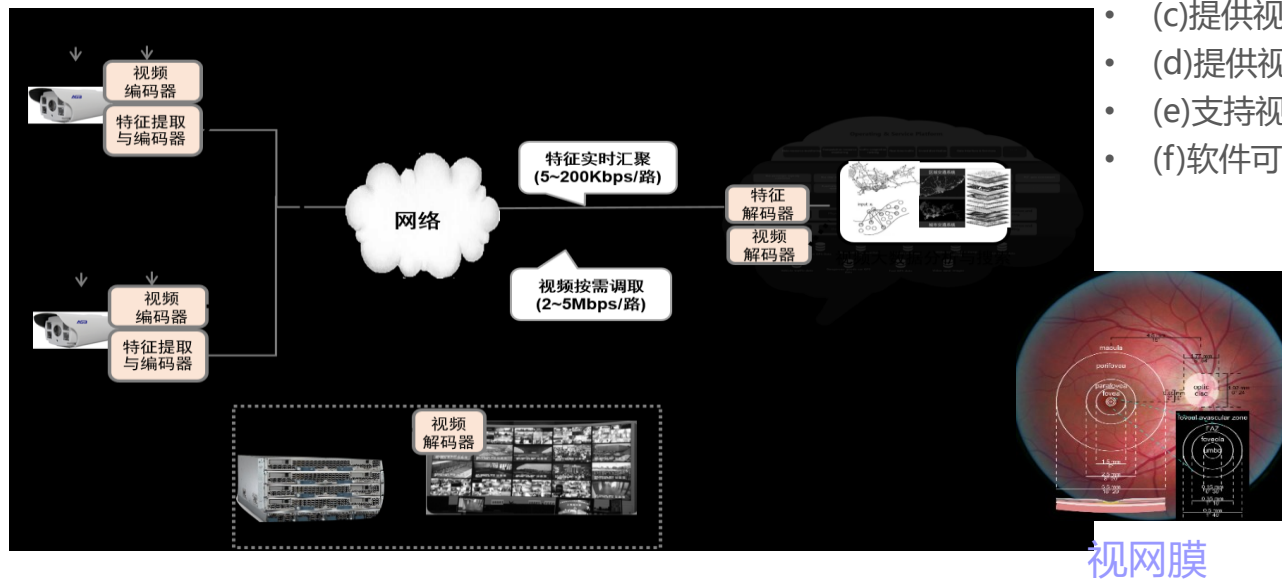
研究内容1：端云协同计算架构

-12-



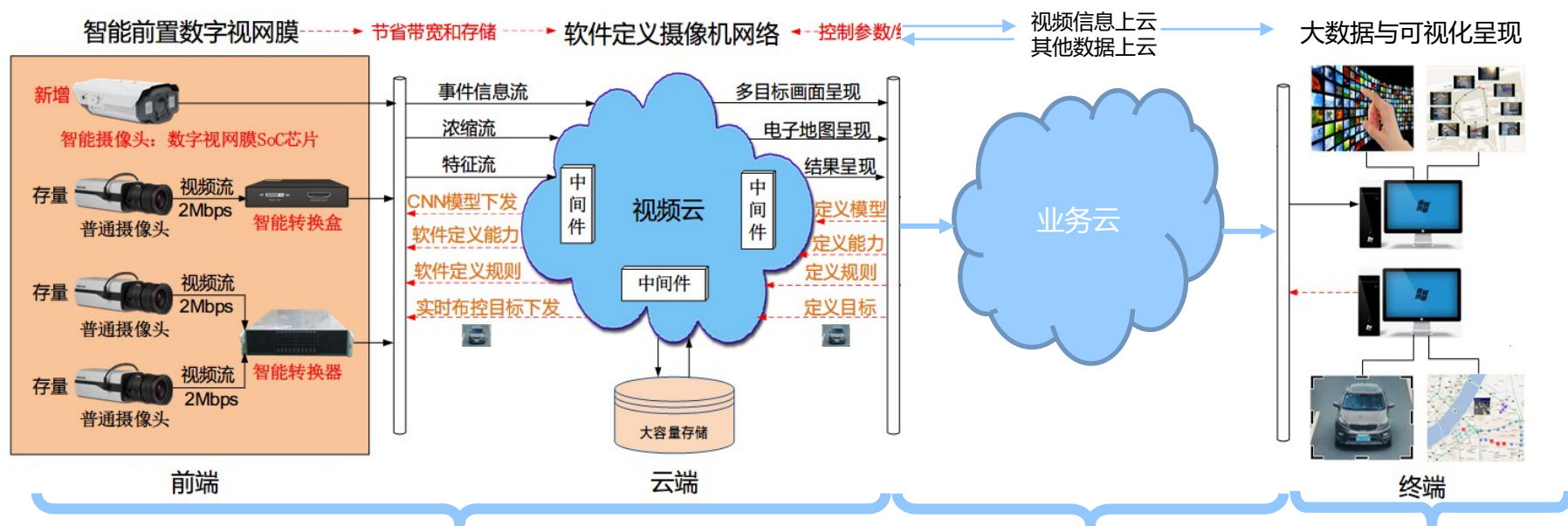
➤ 数字视网膜，必须满足如下条件:

- (a)使用全网统一的时间,
- (b)提供精确地理位置
- (c)提供视频数据的高效编码功能
- (d)提供视频数据的紧凑特征表达
- (e)支持视频编码与特征表达的联合优化
- (f)软件可定义功能



研究内容1：端云协同计算架构

为什么叫数字视网膜，是仿生概念，人眼采集信息到进脑决策判断，信息是逐级递减的，因此在超大规模应用场景中，视频数据也是逐级递减，逐级提取有效信息进云进行业务决策



数字视网膜（城市复眼），提供视觉智能分析，帮助大脑降低计算压力

城市大脑，业务处理

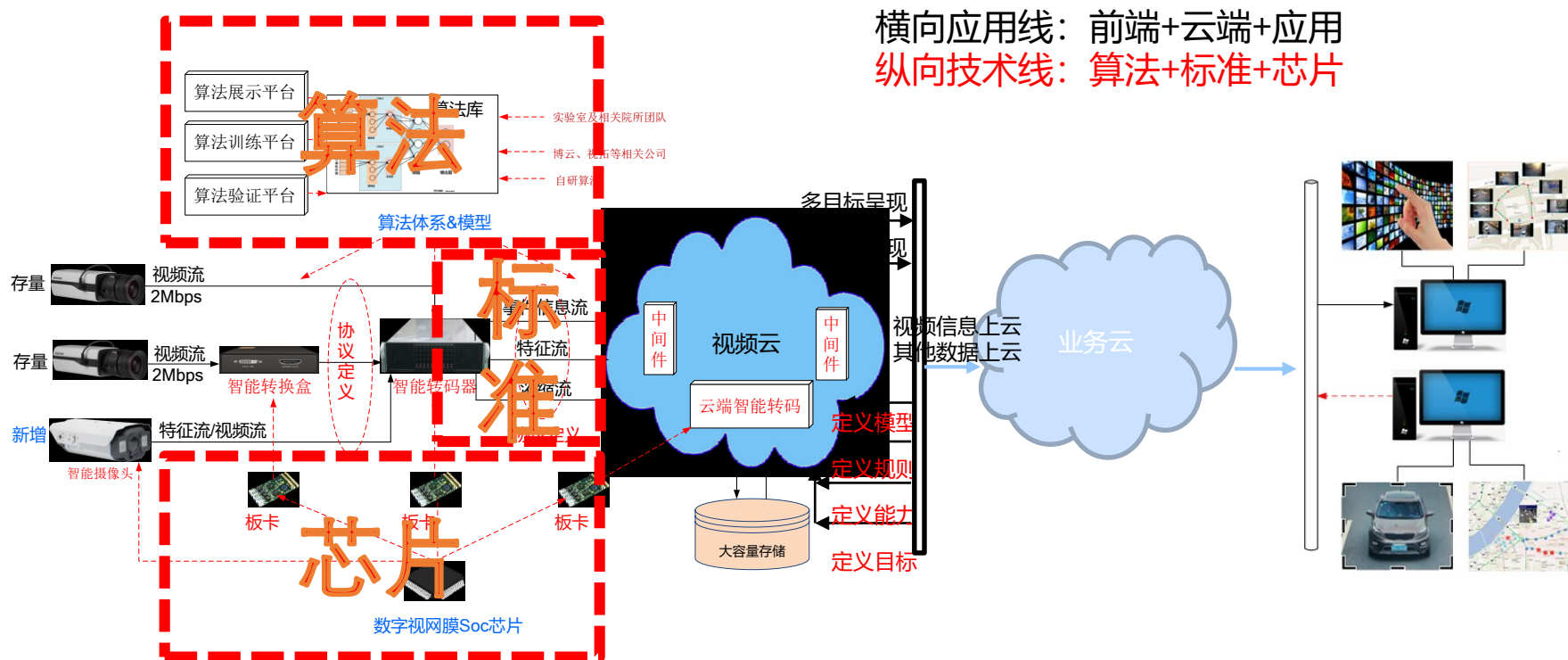
智能提升，节能/增效

仿人类视网膜机理，通过算法和计算对视频数据逐级紧凑表达，为云计算提供高质量的视频信息数据



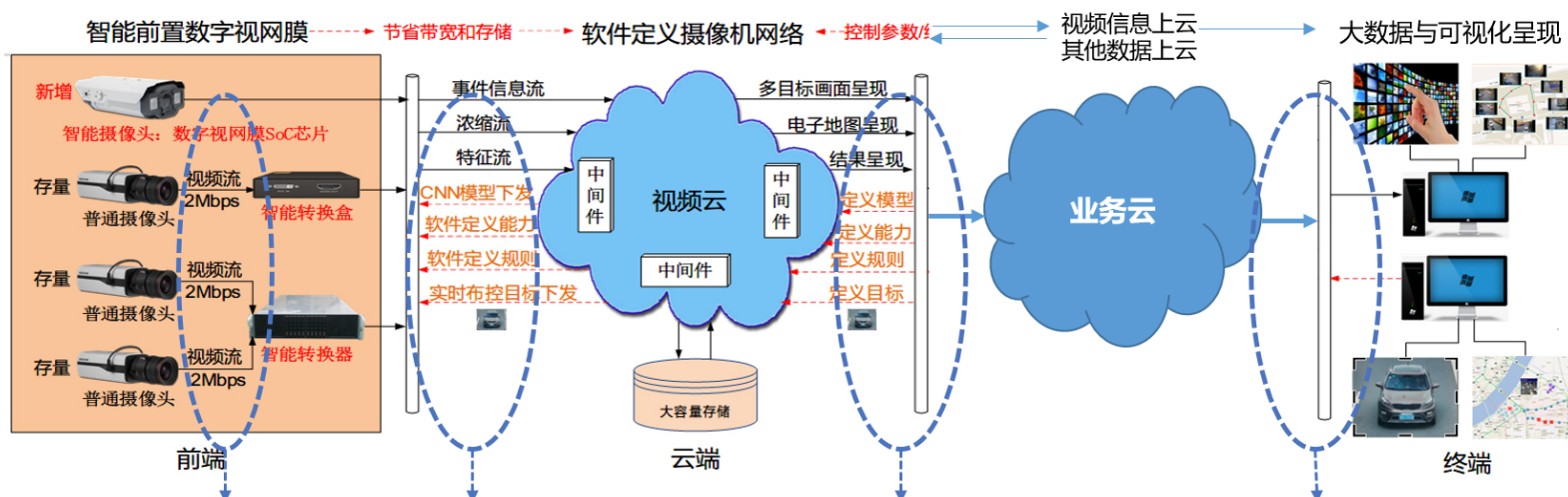
研究内容1：端云协同计算架构

1.2 数字视网膜技术体系



研究内容1：端云协同计算架构

1.3 数字视网膜系统标准体系



新一代人工智能产业技术创新战略联盟



北京大学
PEKING UNIVERSITY



数字音视频编解码技术标准工作组

Audio Video coding Standard Workgroup of China

视频编码标准

AVS2/IEEE 1857.4

特征编码标准

MPEG CDVS

MPEG CDVA

人工智能标准（拟）

神经网络表示与模型压缩标准

深度特征压缩标准

其它系统相关标准（待定）

描述性标准

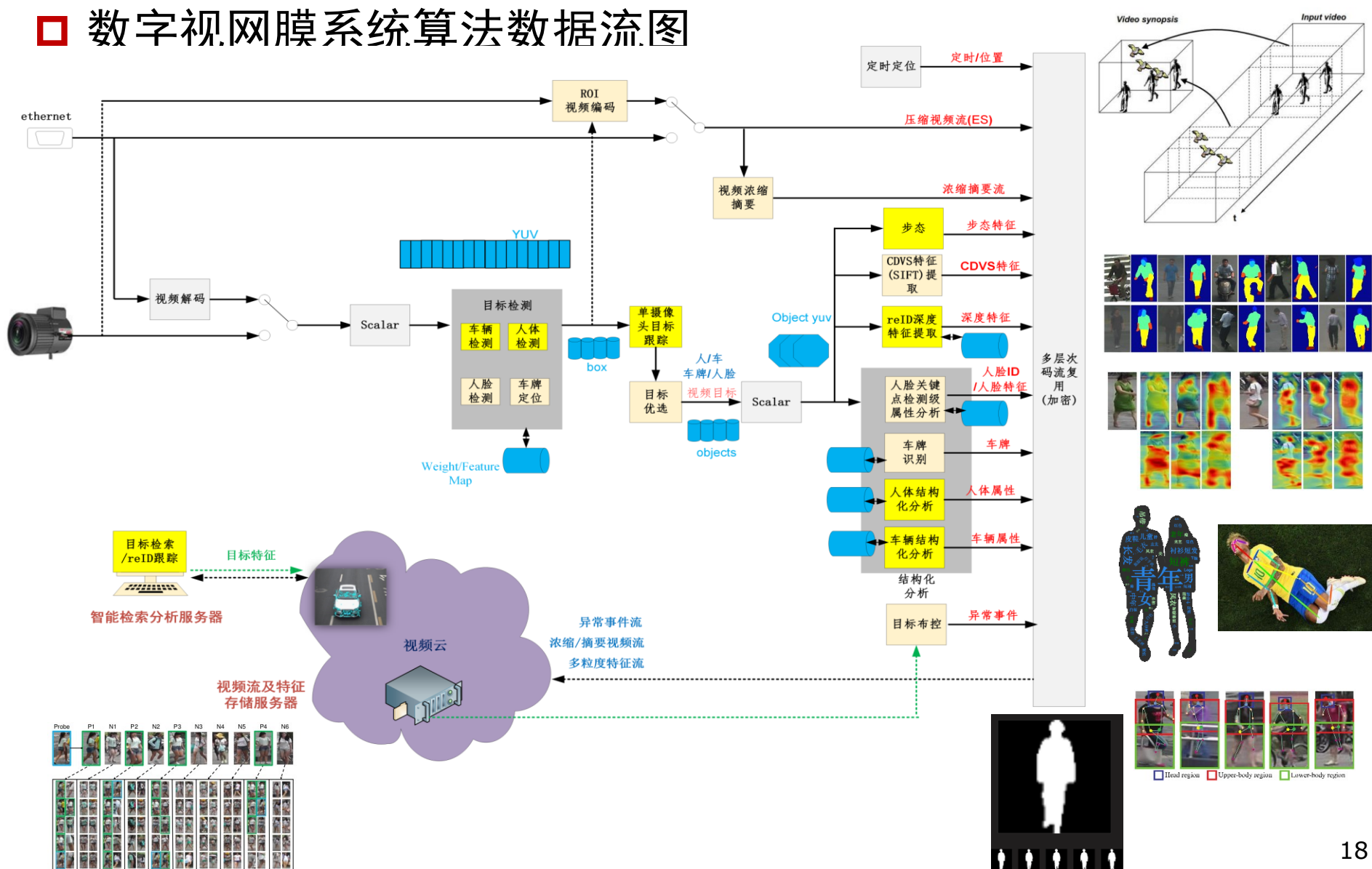


数字视网膜系统算法



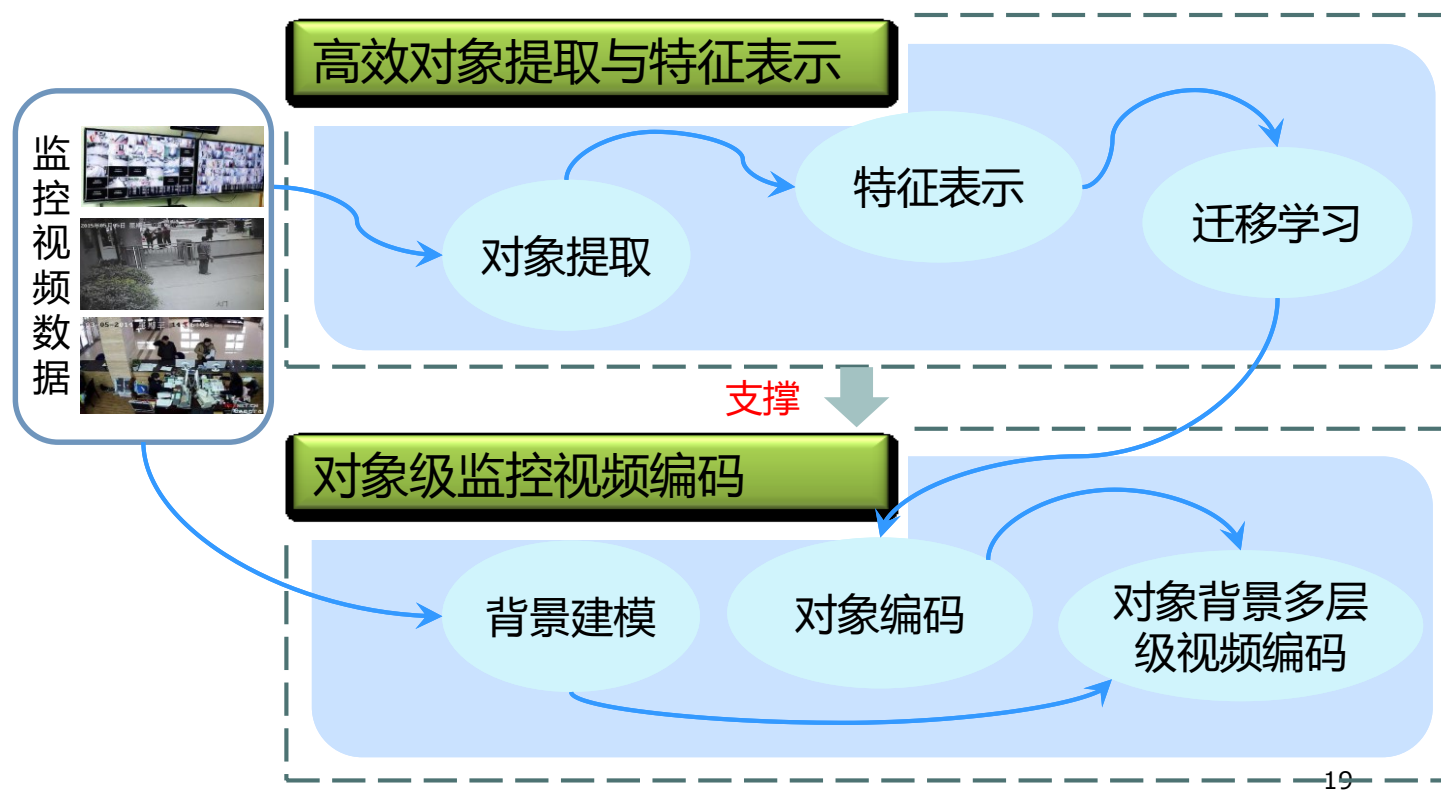
研究内容2:视频编码和分析算法

数字视网膜系统算法数据流图



研究内容2:视频编码和分析算法

□ 主要问题及算法研究点



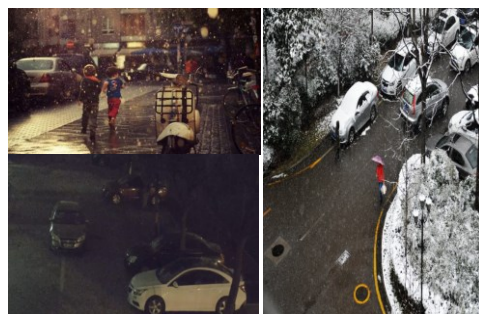


研究内容2:视频编码和分析算法

❑ 主要问题1：高效对象提取与特征表示

主要难点

对象所处
场景复杂

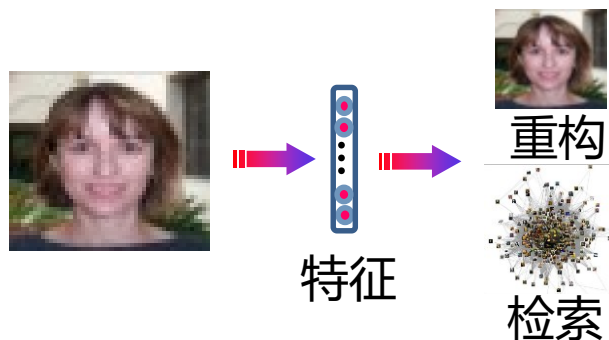


遮挡
光照
距离
...

研究思路

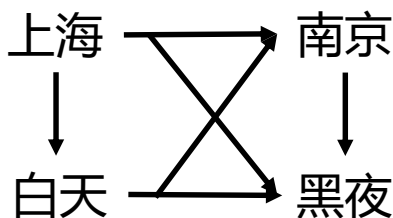
端云协同多视角
关联的目标检测

特征难以同时
满足重构和检索



面向重构和检索
联合优化的特征表示

网络迁移
能力差



源域到目标域
数据的迁移学习



研究内容2:视频编码和分析算法

□ 主要问题2: 对象级监控视频编码

主要难点

背景光照复杂



研究思路

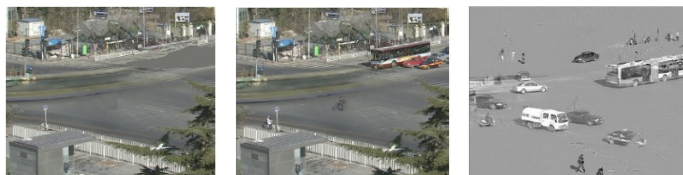
面向监控视频
自适应背景建模

特定对象语义级
压缩表示



基于时空结构化的
特定对象编码

监控视频信息
冗余压缩



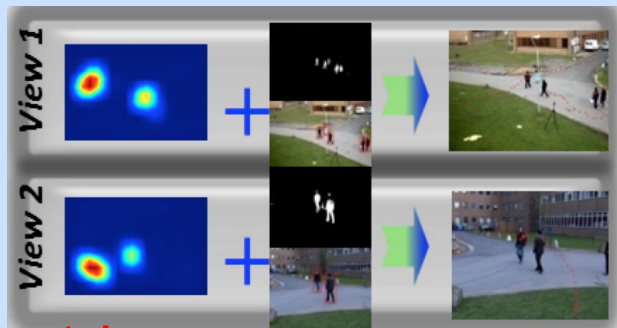
对象背景多层次
监控视频编码

研究内容2:视频编码和分析算法

2.1 端云协同多视角关联的目标检测

技术方案

单视对象检测与跟踪



端 (采集 + 智能分析)

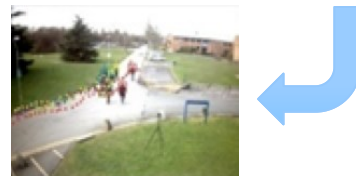


多视运动对象关联与分析



空间关联
轨迹关联
特征关联

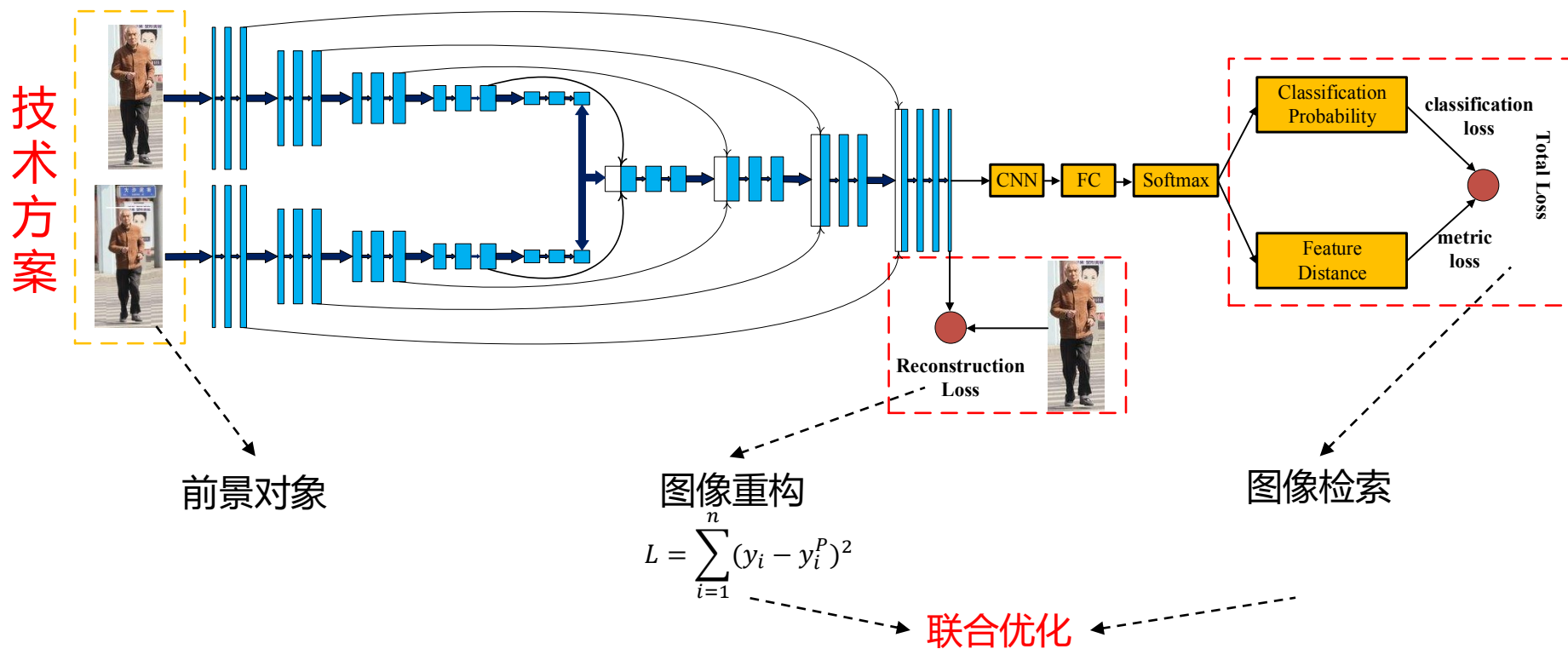
云 (汇聚 + 智能关联)



- 通过云端汇聚多终端多视角信息，利用多终端的互补位置和角度信息对运动对象智能关联与分析，以提高光照、遮挡、背景等复杂场景下目标检测的精度和效率

研究内容2:视频编码和分析算法

2.2 面向重构和检索联合优化的特征表示

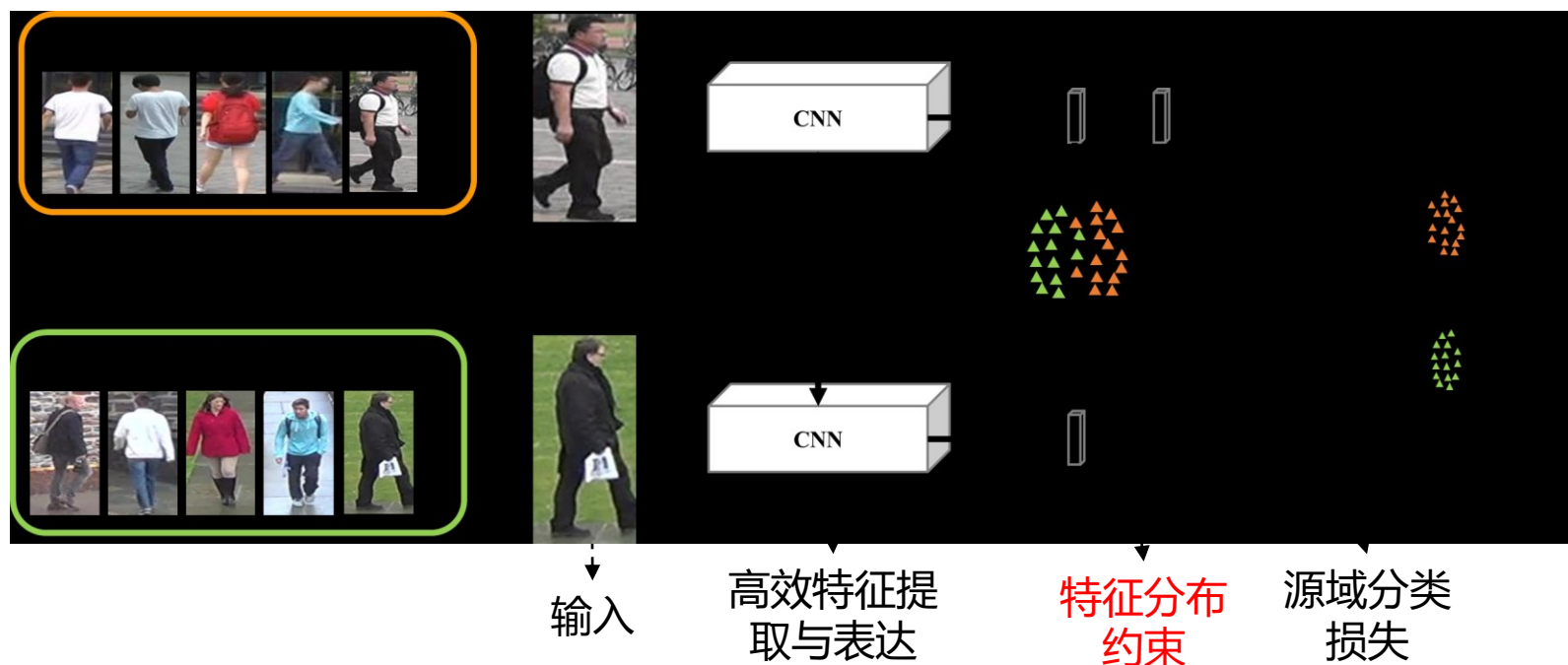


- 提出多任务深度网络模型，建立**双任务驱动的联合优化函数**，提取的特征满足重构和检索需求。

研究内容2:视频编码和分析算法

2.3 源域到目标域数据的迁移学习

技术方案



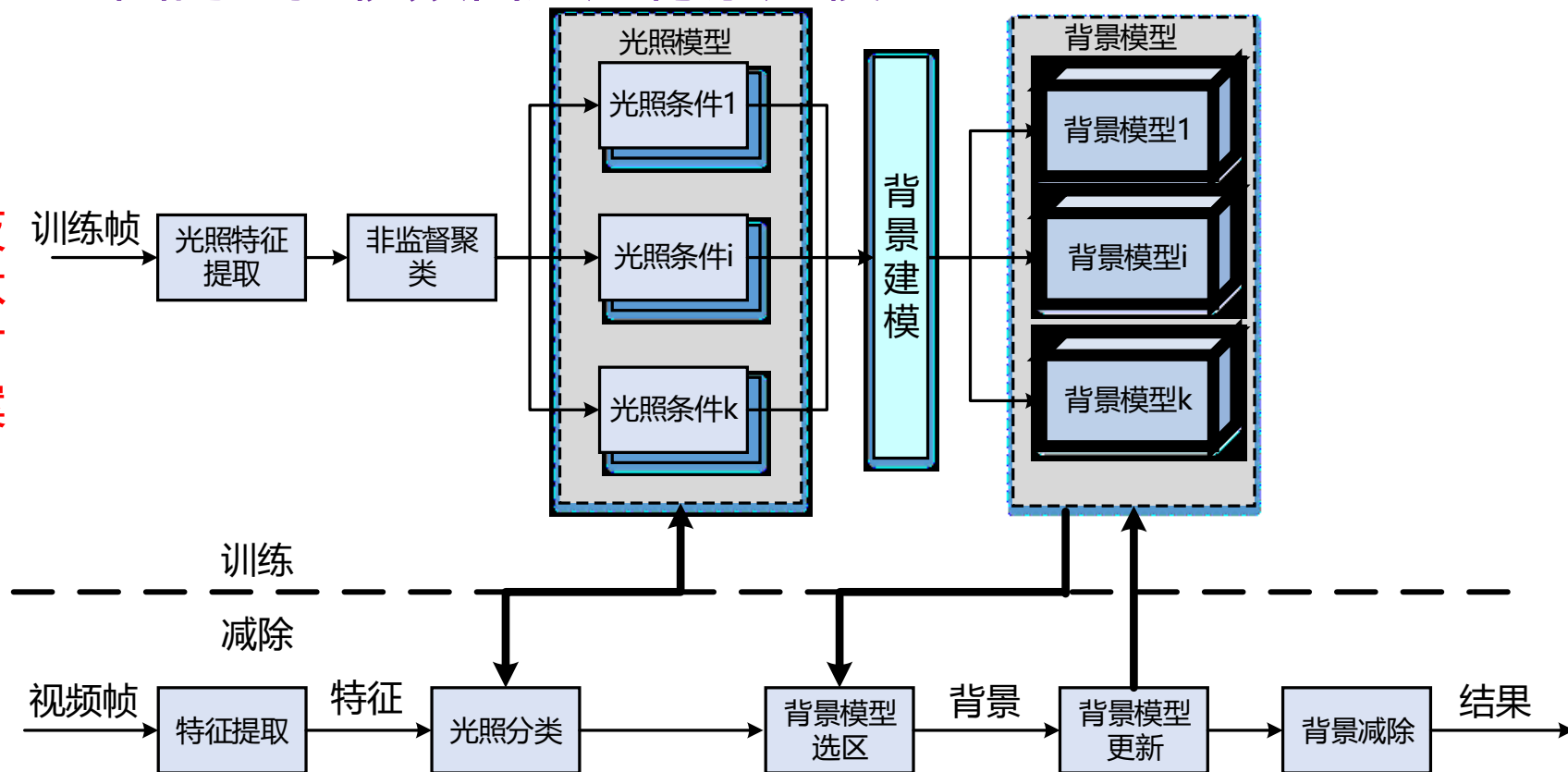
- 通过计算源域的分类损失，学习具有区分性的特征，通过源域与目标域特征分布距离约束，最终将具有区分性的特征表达迁移到目标域。



研究内容2:视频编码和分析算法

2.4 面向监控视频自适应背景建模

技术方案



- 周期提取视频中的光照特征，与训练好的光照聚类模型比对，更新最匹配的背景模型，解决不同光照条件背景建模问题

研究内容2:视频编码和分析算法

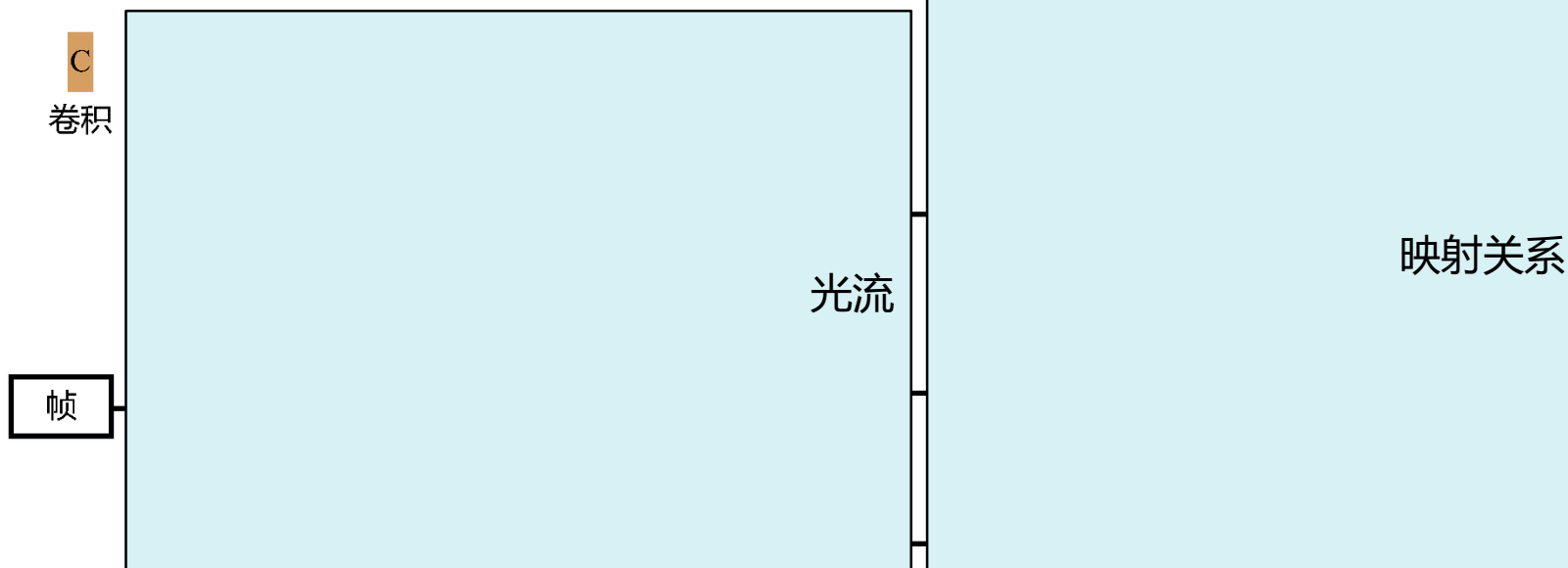
2.5 基于时空结构化的特定对象编码

基于特定对象
姿态的结构化
空域编码



$$P_i = (label_i, x_i, y_i, w_i, h_i, k_{i,x}^1, k_{i,y}^1, k_{i,x}^2, k_{i,y}^2, \dots, k_{i,x}^N, k_{i,y}^N)$$

技术方案



- 构建特定对象姿态的结构化空域编码模型，建立基于光流的帧间映射关系，提出深度特征流映射学习的时域编码方法。

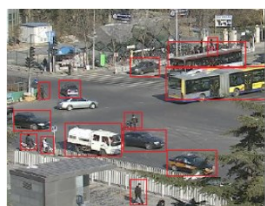
研究内容2:视频编码和分析算法

2.6 对象背景多层次监控视频编码

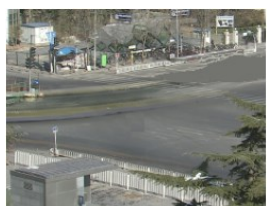
技术方案



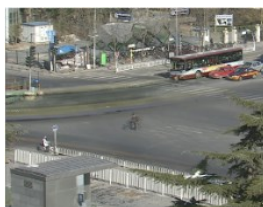
原图



特定语义对象提取



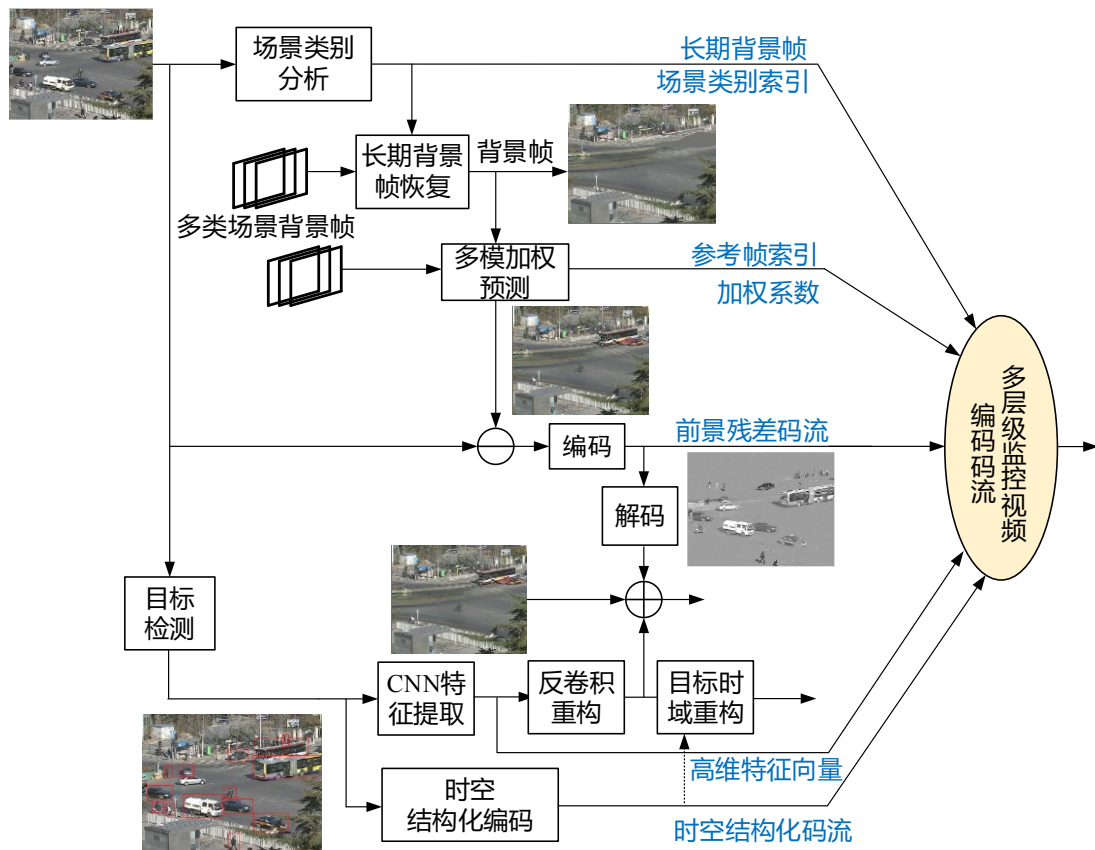
长期背景帧建模



短期背景帧建模



前景预测残差

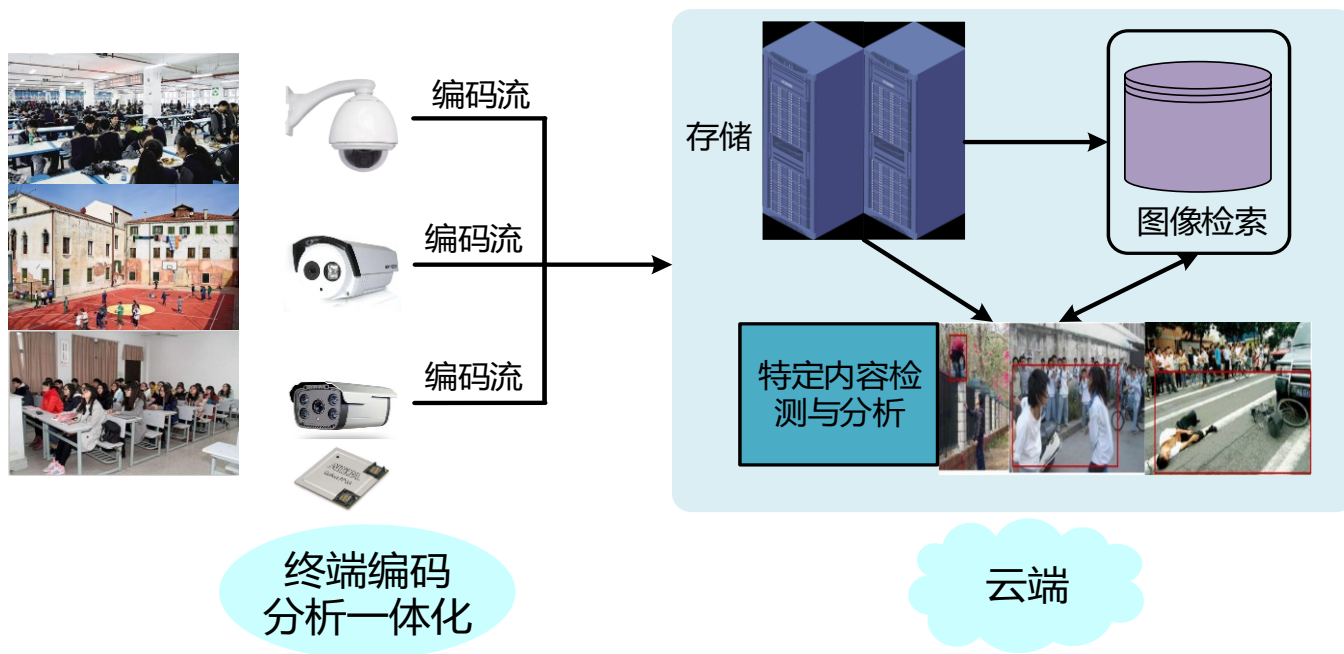


- 提出特定语义对象编码、长期背景帧编码、短期背景帧编码、前景预测残差编码的多层级编码方法，解决监控视频空-时冗余压缩和高效编码问题

研究内容3:应用示范系统

□ 万路视频协同感知应用示范系统

技术方案



- 面向公共安全管理应用需求，构建新一代的终端编码分析一体化，云端视频存储、图像检索、特定内容检测与分析的端云协同视频智能监控原型系统。



汇报内容

一

研究背景

二

研究内容

三

实验结果

四

总结与展望

研究成果



研究结果



研究结果





研究结果

序	任务	子任务	上周指标	本周指标	指标	交付
1	检测	行人、车辆		如上周	/	√
2		车牌、挂件等	mAP:94.5%(IOU:0.5 南京) 挂件: 98.6%(AP@0.5) 遮阳板:100%(AP@0.5)	如上周	挂件:>95% 遮阳板:>93% ^[1]	✗
3		人脸	mAP@0.5: 94.3%(>40*40), 89.8%(<40*40)	如上周	/	√
4	跟踪	SORT 算法	IDF1:/, ID Sw:1017, Hz:5000 fps	如上周		√
5		MHT 算法	IDF1:38%, ID Sw:148, Hz:10-30 fps	如上周		
6		南京算法	IDF1:/, ID Sw:20325, Hz:100 fps	如上周		
7	结构化	行人结构化-性别	性别: 90%(10K),86%(51 lmg),83.94%(现场) 年龄: 66%(Acc) 鞋子: 63%(类型), 68%(颜色) 其他: 85%-99%	性别: 93.65%(570lmg) , 92.16(51lmg) 年龄: 74.91% 鞋子:65.01%(类型),73.3%(颜色) 其他:68.07-100%	检出率:>90% 识别率:>90%	√
8	ReID	行人	监督: 80.4%(Rank-1, Duke) 半监督: 70.8% (Rank-1 Duke) 无监督: 60.5%(Rank-1 Duke) Market1501 82.6(mAP) 93.3(Rank1) 97.8(Rank5) 98.6(Rank10)	如上周	>80%	√
9		车辆	79.8%(mAP) 91%(Rank1) 95.5%(Rank5) 96.6%(Rank10)	如上周	>80%	√
10		非机动车	/	/	>80%	✗
11	语义分割	行人轮廓分割	97.42%(PA, 自有融合), 94.0%(Mean_IOU)	如上周	/	/
12	步态	行人步态识别	62%(Rank-1 海创园), 83.6%(Rank1 Acc, 现场数据训练后,39IDs)	如上周	命中率:>80%	✗
13	分类	跟踪结果筛选	91.27%	如上周	Acc>95%	✗



汇报内容

一

研究背景

二

研究内容

三

实验结果

四

总结与展望



总结与展望

-35-

- **学术研究：系统级端到端算法多目标优化**
- **标准化：数字视网膜系统标准化**
- **应用示范：2000/10000路视频分析系统**



请批评指正！

谢谢

yhbm@hdu.edu.cn