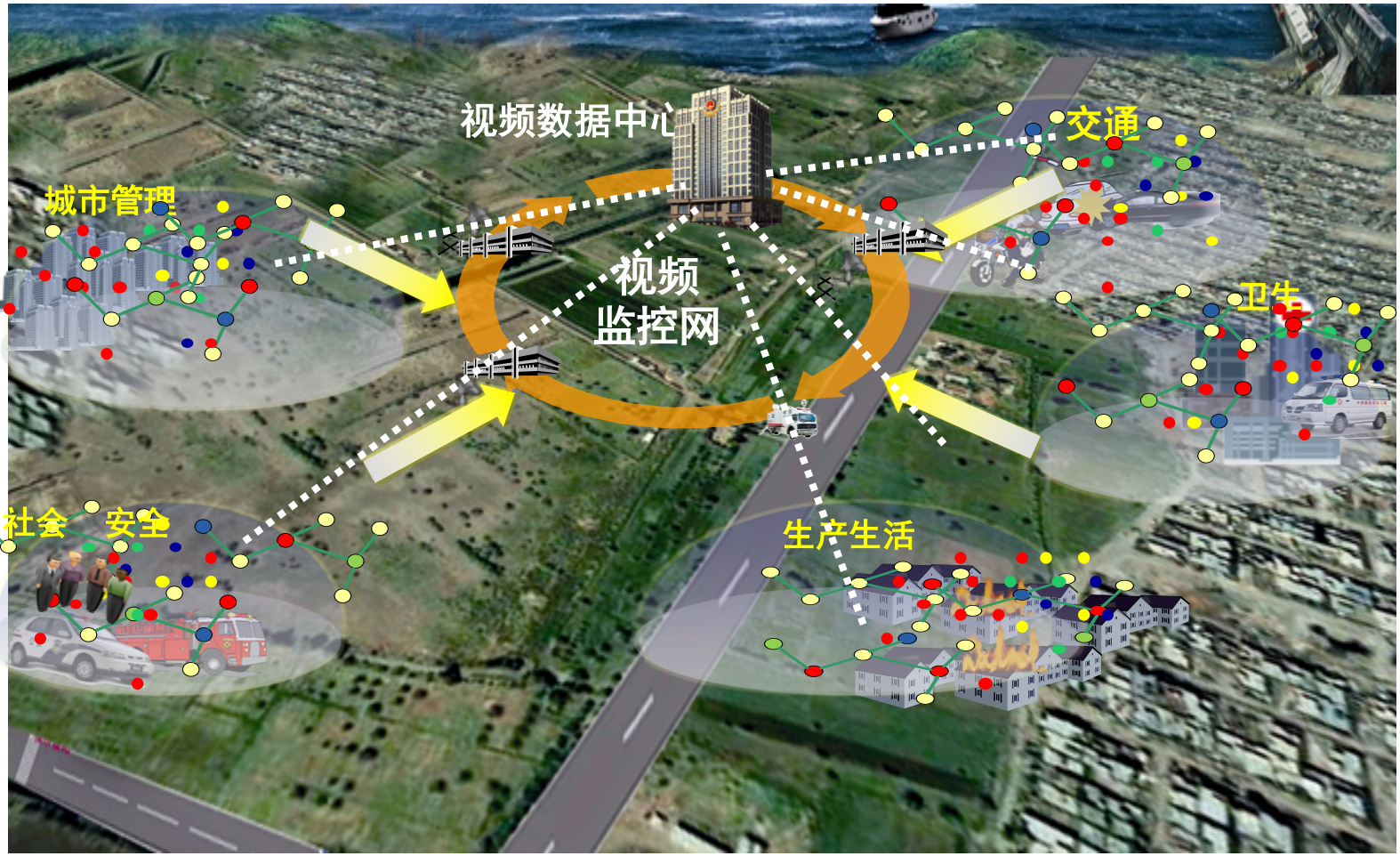
大规模视觉对象的精准搜索与识别

田永鸿
yhtian@pku.edu.cn

北京大学
数字视频编解码技术国家工程实验室

视频监控网络：城市之眼



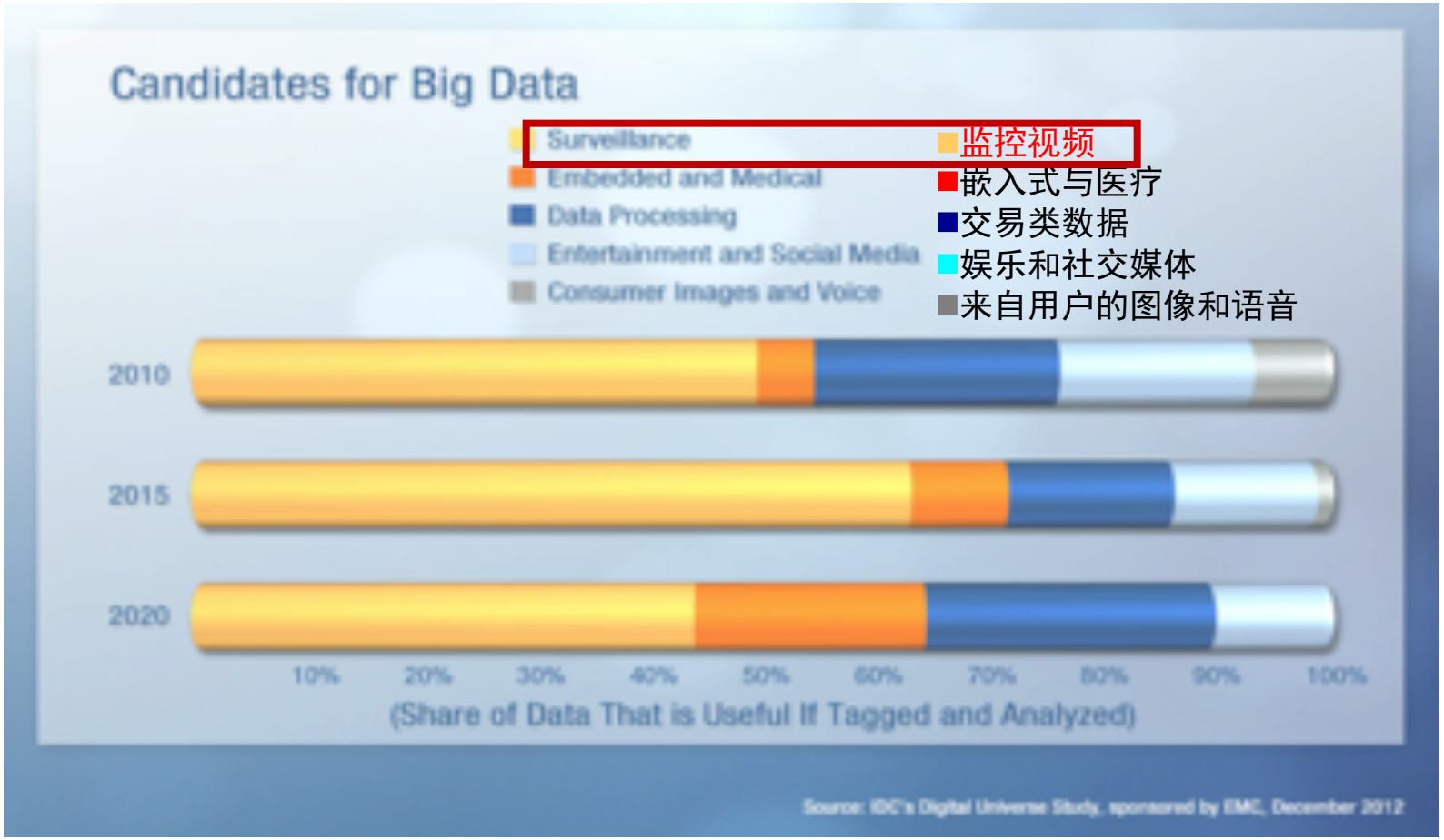
全国已部署超3000万监控摄像头

（7000万？——阿里王坚）

北京大学

监控视频：体量最大的大数据

大数据的一半是监控视频



数据来源：IDC's Universe Study, 2012年12月

关键技术挑战：智能分析识别

视频监控系统现状：智能程度低，综合效能差

有眼（即高清头）←→无珠（即智能分析）

华西都市报 2014年12月12日

车上乱扔竹签 “全球眼”10分钟送来罚单

成都高清球型摄像机将严查车窗垃圾，昨日开出首张取证罚单

华西都市报见习记者 张乾 摄影 郝飞

随手向车外撒洒纸巾、烟头等垃圾，是一种典型的不文明行为。而这种交通违法在瞬时间完成，隐秘性高，即使路面有交警也很难固定证据。但12月11日午后，市民周先生乘坐在同事的车内吃完一顿“关东煮”午饭，随后一扔小竹签，却很快领到了罚单。

这是因为已在成都全面启用的无死角高清球型摄像机，即俗称的“全球眼”，在二环路东四段龙舟路口，清楚地记录下他向车外扔竹签的全过程，最终他被处以50元罚款。这是全球眼启用后，成都交警部门针对向车窗外抛撒物品的违法行为抓拍、开出的首例罚单。昨日，在全市范围内有45名乱扔杂物的乘客或司机受到了处罚。

12月11日，二环路龙舟路口，全球眼抓拍乘客抛物全过

竹签车外一扔 全球眼紧盯

600多套全球眼在成都全面启动，已经有一段时间了。全球眼的终端“成都交警交通违法视频监控中心”，每天都有值守人员监控街头上的隐秘交通违法。

12月11日下午2点10分，值守人员监控二环路东四段龙舟路口的全球眼时，就发现了一辆白色福特汽车行至路口停车等候红绿灯时，有一根小竹签从车内被扔了出来。“我们马上将拍摄下的违法数据固定并上传，同时通知了属地民警。”工作人员介绍。

收到指挥中心指令，成都市交警三分局民警立即在下一个路口将该车拦下。车内坐着的3个人还一头雾水。“我们没有违法司机说。

抛物被拍 乘客挨罚50元

“你们车里的人刚才有没有向车外扔什么东西？”民警询问。车上的3人都摇摇头，否认了。随后，一行人被带往交警三分局。

在看到10多分钟前小竹签划出的清晰弧线，当时坐在副驾的乘客周先生终于开口：“唉，确实是我扔的。”周先生说，他和同事上午去三色路跑业务，由于太忙中午没吃饭，就在便利店买了一些“关东煮”当午饭。“我坐在前排，没留意就顺

立即

别让车门

时间、车辆等特点，采取流动巡逻与固定岗点相结合、区域联勤和错时勤务等方式，形成了严密管控的高压态势，有效挤压车辆乱停放、向车外抛撒物品、遗撒飘散载运物等交通陋习的生存空间。”民警说。

华西都市报记者 李鑫 摄影报道

华西微博:t.huaxi100.com | 华西都市网:www.huaxi100.com | 报料QQ:130069110 | 短信报料:13880035050 | 编辑 向雷 版式 易灵 校对 黄淑琼

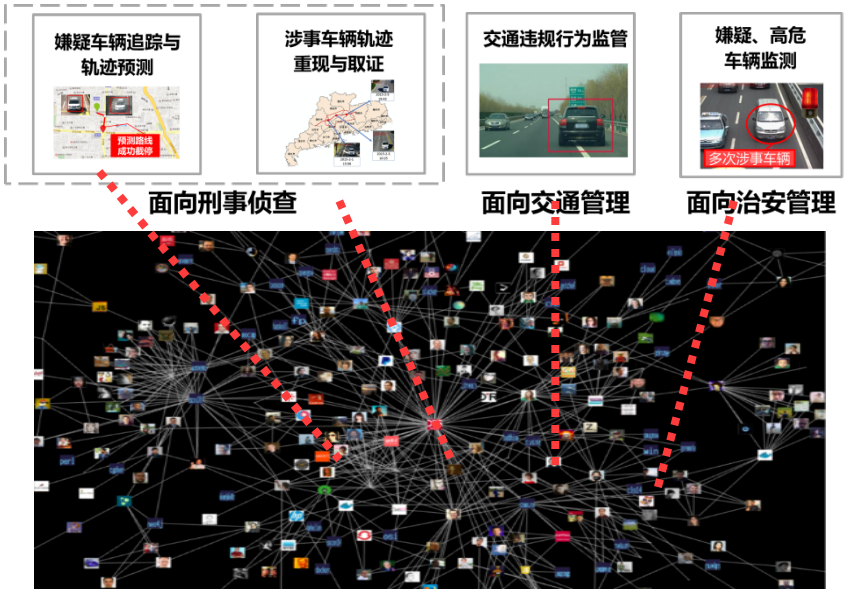
600多套全球眼在成都全面启动，已经有一段时间了。全球眼的终端“成都交警交通违法视频监控中心”，每天都有值守人员监控街头上的隐秘交通违法。

600个摄像头，600个人在后面看？

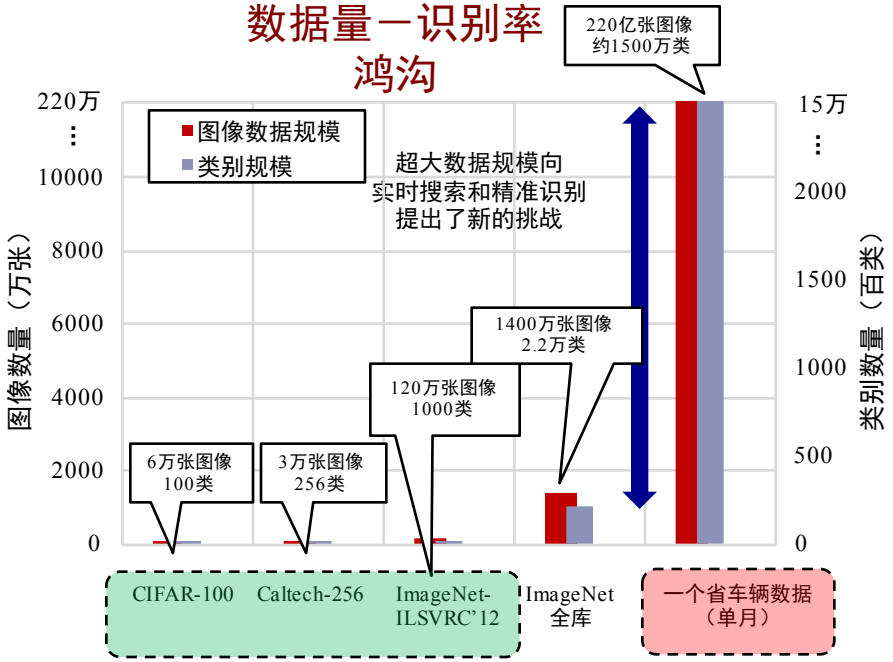
中国一个典型的省...

数据类型	规模	数量质量
全部摄像头	100万	2013年开始高清化，如果像电视质量那样存储，需要2000PB，成本200亿元
治安摄像头	10万	
卡口图像	3千电子卡口 3万电子警察 （预计两年翻番）	卡口图像 1亿幅/天 电子警察 10亿幅/天
人像	1亿人	证件照
车辆	2千万	行驶证/[illegible]险
民警	**10万**	[illegible]？ [illegible]路？）

核心科学问题：大规模视觉对象搜索与识别



图像视频大数据：视觉对象跨域关联与识别



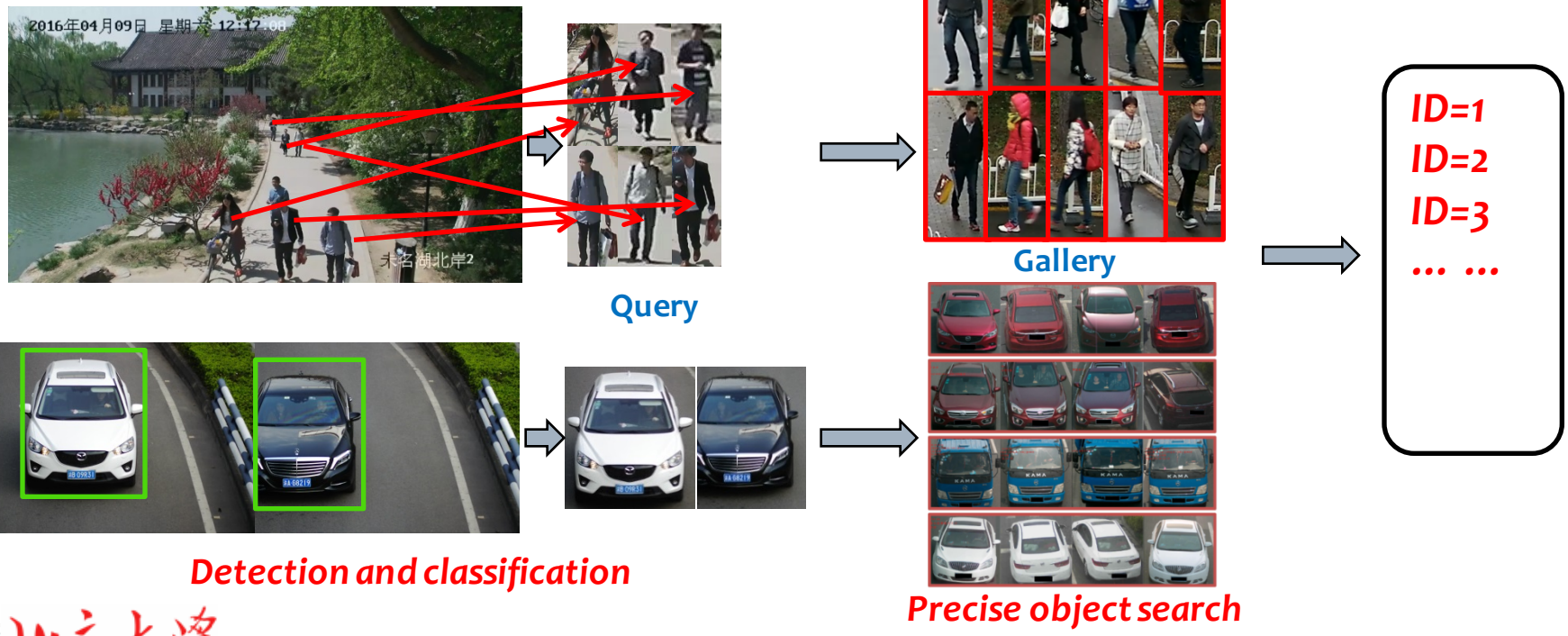
超大数据规模向实时搜索和精准识别提出了新的挑战

100万
监控摄像头

百亿车辆图像 ＋ 千万规模车辆 ＋ 十万行人对象 ⇒ 识不准 搜不快

精准对象搜索 (Precise Object Search)

- 任务：从采集自分布式摄像头网络中的大规模数据中搜索特定的对象
  - Search as **Similarity Ranking (SaS):** NOT for visually similar object
  - Search as **Recognition (SaR):** BUT for exactly the same object

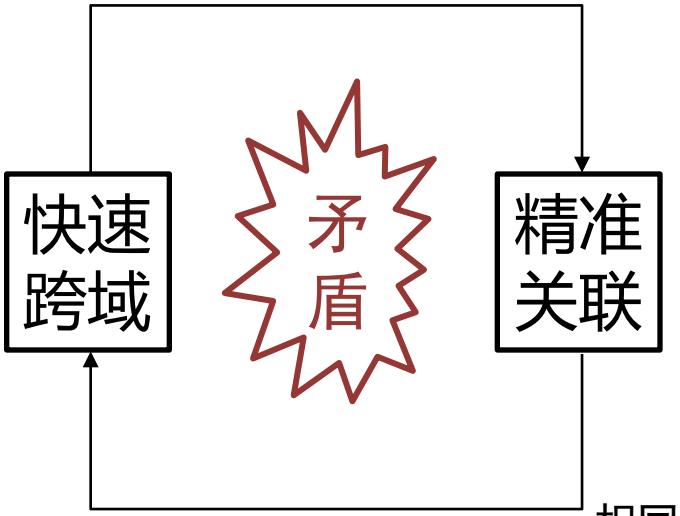


为什么难?

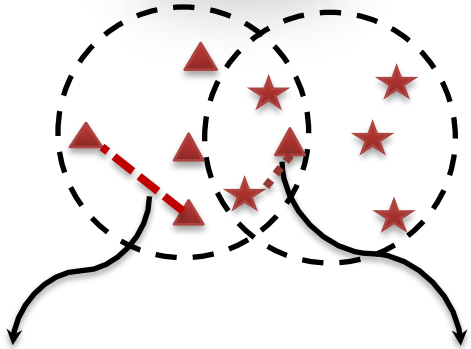
- 需要在统一框架下分析、识别与搜索来自无限制环境下采集的图像视频数据
  - 数据量大
  - 不同的成像视角、光照、环境条件、遮挡、图像质量
  - 嫌疑人/车的视觉表观变化
  - 其他原因（如缺乏足够训练数据）

场景跨度大

时间动态变化



数据规模大



相同视觉对象
呈现多样性

不同视觉对象
呈现相似性

为什么难?

- Twin难题：
  - 孪生子识别，模式识别中经典难题
  - 难以有效区分视觉表观相似对象（如同颜色、同型号的车辆）

Different

Same

为什么难?

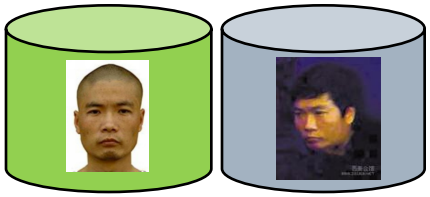
- 不能依赖于强标识信息，如人脸、车牌等
  - 人脸在大多数现实监控摄像头下不可得
  - 嫌疑车辆可能套牌、换牌、遮牌或无牌

Face Image Retrieval Scenario [Li, ICCV2015]

→Pseudo-proposition?

How to search given these pictures?

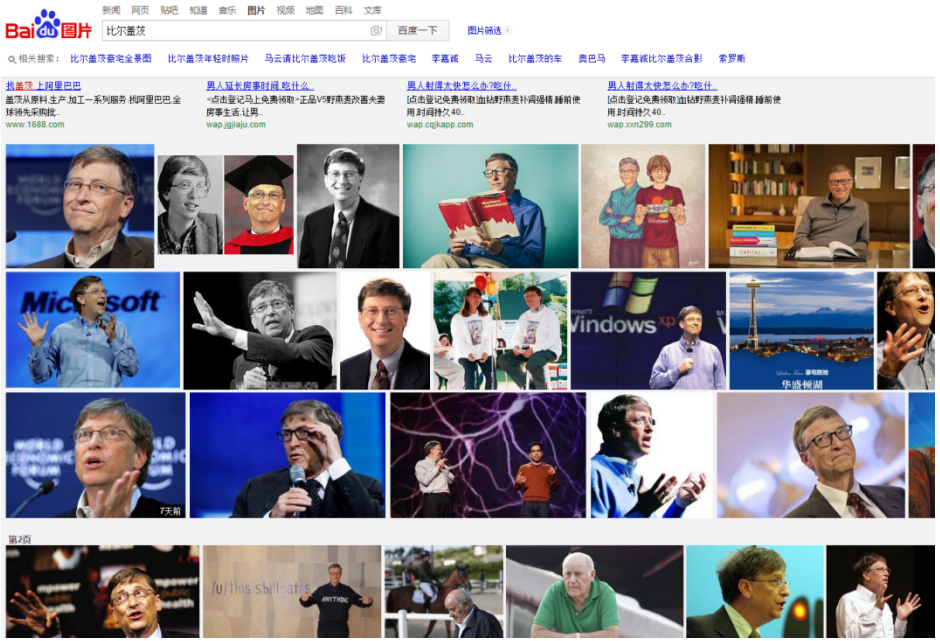
- ✓ *No front face image is available*
- ✓ *With some facial makeups*
- ✓ *Don't know he is who*



ID Face Database

Surveillance Face Database

从搜索到识别



- 精确识别是监控智能的终极目标
- 没有一种识别技术能够在任意环境下达到足够的识别精度：车牌识别、人脸识别?
- 互联网的成功模式告诉我们，搜索可以代替识别

北京大学

案例：搜索可以解决识别问题

Tollgate

Car Monitoring 2

Car Registry Database

Search Engine

Car Monitoring 3

Car Monitoring N

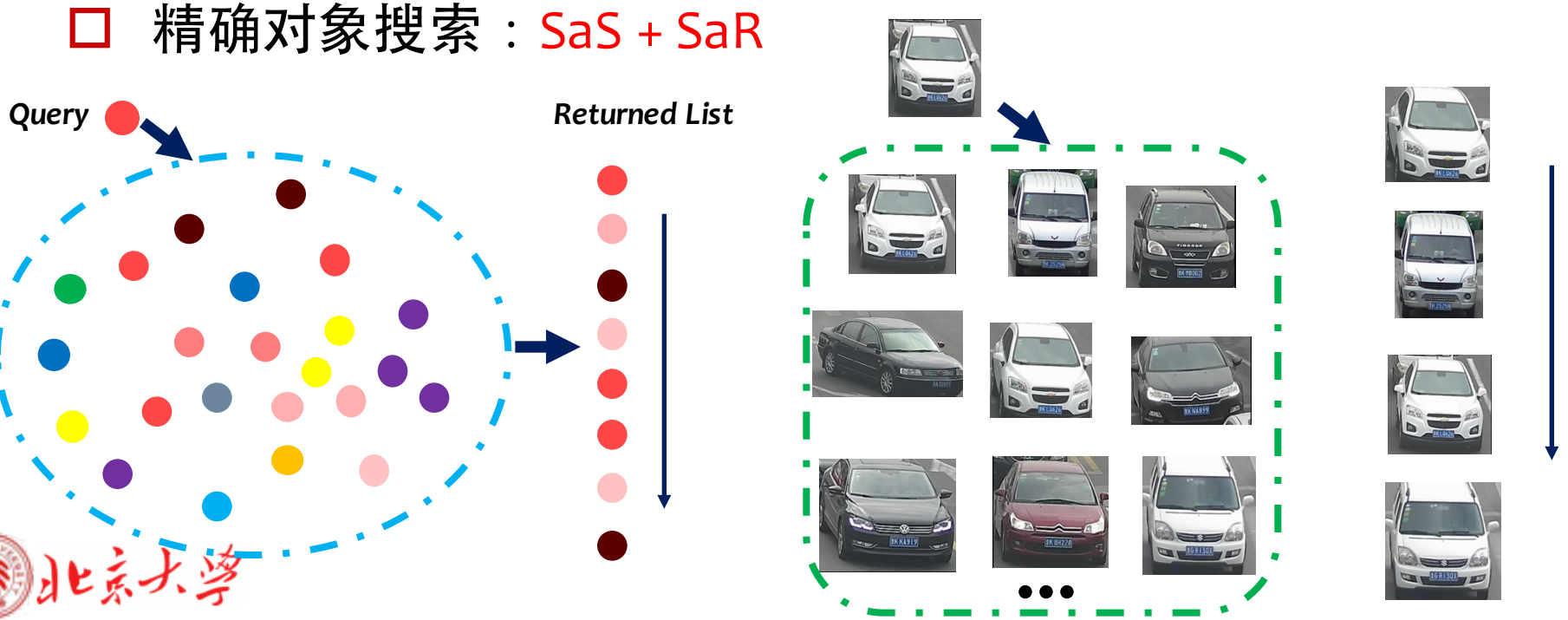
Honda accord

套牌车

Peugeot 206

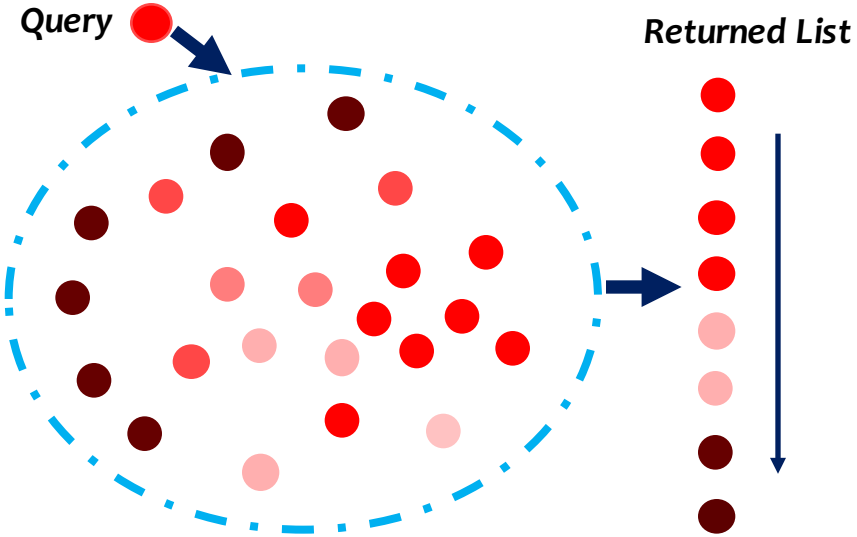
与传统视觉搜索的区别

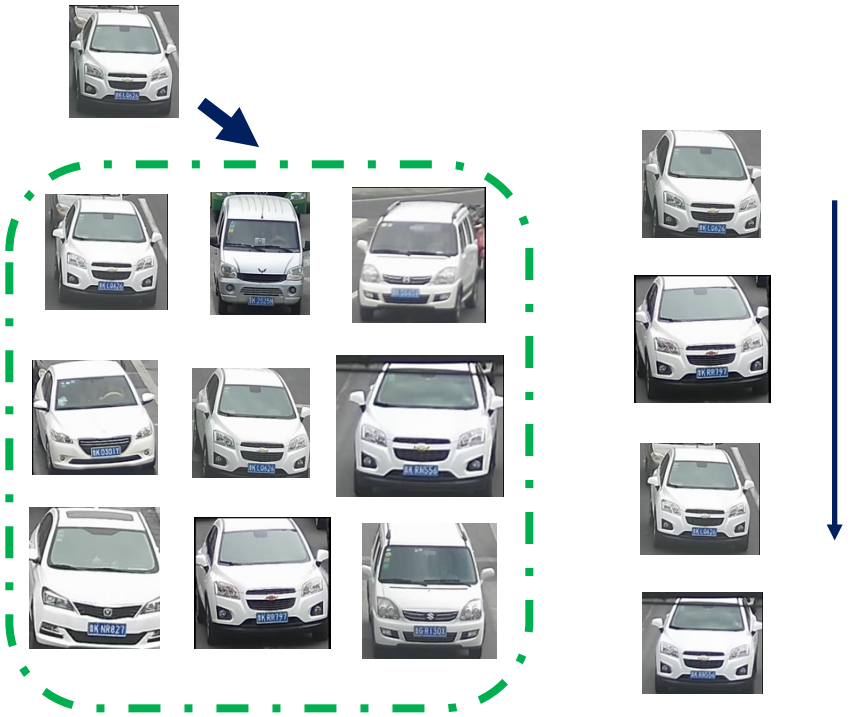
- 传统视觉搜索：目的是找到视觉表观相似的对象（visually similar objects）
  - In most cases, the returned objects that are visually similar (e.g., within the same (sub-)category, having the same attributes such as color) are treated as correct
  - Recent development: Fine-grain search
- 精确对象搜索：SaS + SaR



与对象再标识(Re-ID)的区别

- Re-ID: The task of assigning the same identifier to all the instances of the same object or, more specifically, of the same person, by means of the visual aspects that have been captured and extracted from an image or a video [Vezzani et al. 2013]
  - 主要问题：行人再标识（Person Re-ID）
  - 主要模型：re-identification as matching (RaM)
  - Generalization is the key challenge





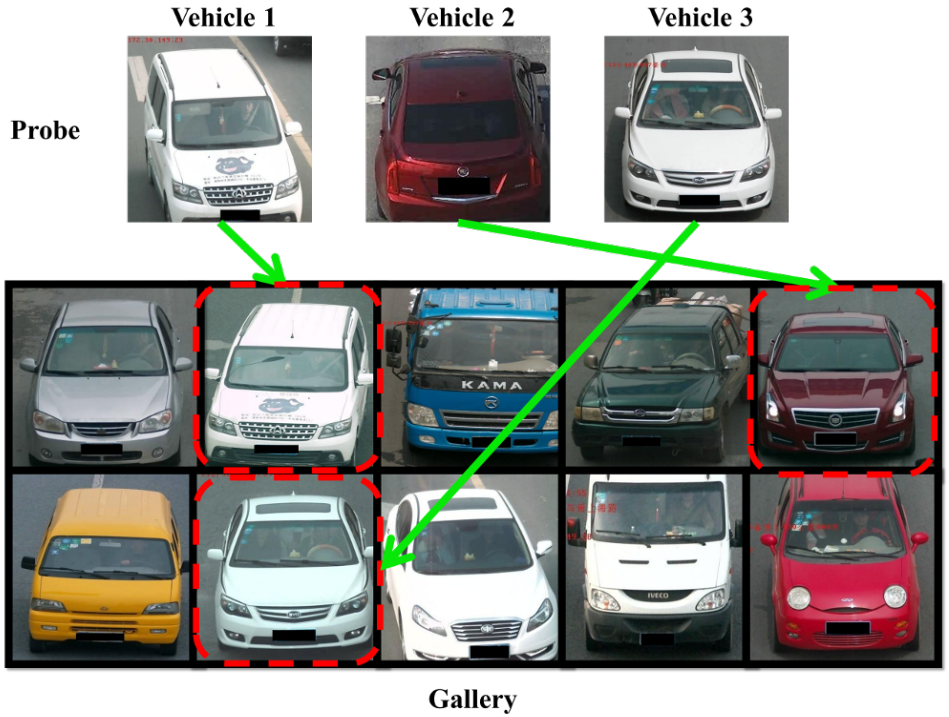
小结：三个紧密相关问题的区别

	Ranking problem	Large database	Identify object	Index structure	Evaluation Metrics
Visual search	✓	✓	✗	✓	mAP
Object Re-id	✓	✗	✓	✗	top-k matching rate
Precise object search	✓	✓	✓	✓	Both

两类典型的精准对象搜索问题

Precise person search

Precise vehicle search

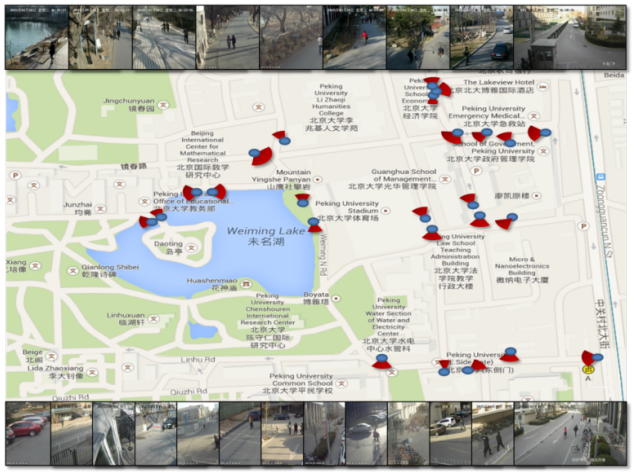


北京大学

PART I：Precise person search

行人Re-ID方法的固有缺陷

- 有监督方法：研究主流→数据依赖性高，缺乏可扩展性
  - 需要为每个摄像头采集的数据标注训练数据
  - 标注时需要肉眼分辨来自不同摄像头的人是否为同一人，非常难
  - 现有行人Re-ID数据集通常较小，最大行人ID数不超过1000（对比：人脸LFW有5749个人的脸，最新结果是在8M个人的人脸数据库预训练）



在北大校园监控网络中，25个摄像头需要标注 300 组摄像头对之间的行人Re-ID 数据。

- 无监督方法：性能普遍较差
  - 没有不同摄像头间标注的匹配行人图像对，现有模型很难学得如何使得在对象表观有显著变化的情况下可健壮标识的特征

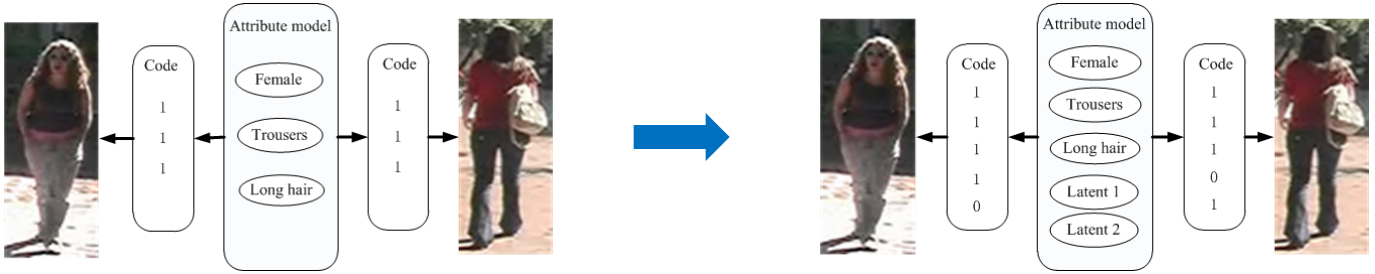
利用属性来解决大规模跨场景行人表征问题

- 跨场景行人表征问题的难点
  - 由于光照、角度等变化，同一行人在不同视角下外观差异大
  - 由于衣服等因素，不同人很可能会有相似的外观
  - 行人表征要求：对视角变化鲁棒，对不同行人高区分性，推广性好

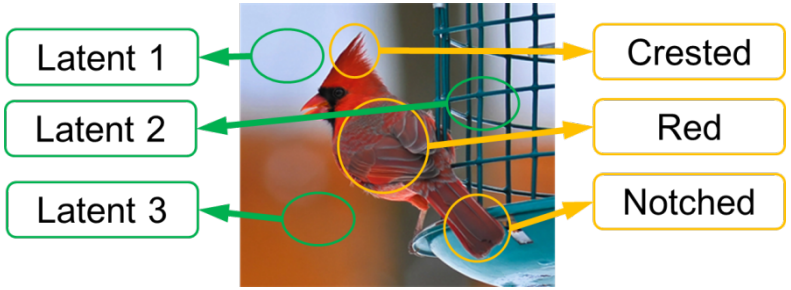
- 属性（Visual Attributes）
  - 一种中层表达，通常具有语义，并且由人工自己定义
  - 优点：对视角鲁棒；
    空间越大，判别越强
    属性模型可共享
  - 缺点：需要人工标注

属性空间扩展：从语义标注到隐性属性

- 隐性属性（Latent Attributes）
  - 隐性属性，同样是一种目标的中层表达方式，未被定义或者无法被显式定义
  - 隐性属性可以帮助预定义属性使得属性具有更好的区分性

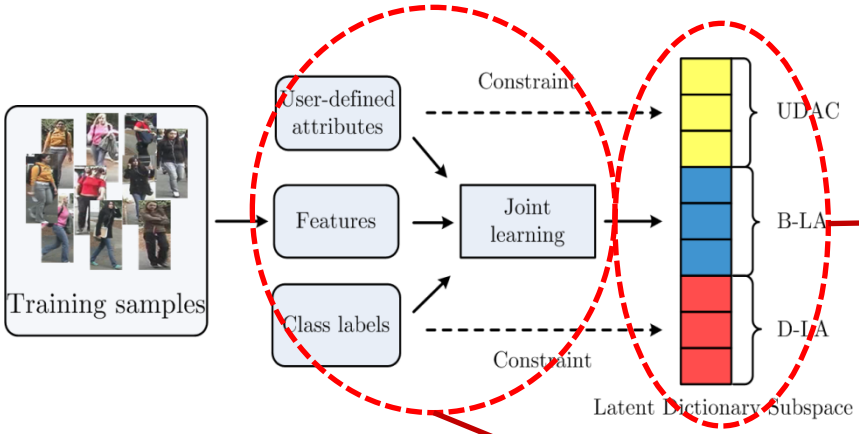


  - 隐性属性可以更好地帮助标注属性的学习



语义标注和隐性属性的联合学习

预定义和隐性属性的联合学习框架



学什么？

1. 预定义属性和隐性属性
2. 隐性属性进一步划分成具有区分性的隐性属性 (D-LA) 和背景隐性属性 (B-LA)。
2. D-LA 用来建模图像中具有区分性的图像信息，对行人Re-ID 有帮助。
3. B-LA 用来建模背景等噪音信息，防止这些信息对有用属性的学习产生干扰。

怎么学？

预定义属性和两种隐性属性建模在同一个子空间里，所以这三种属性可以信息互补，对应地，

UDAC: 代表与预定于属性相关的子空间，它的建模受到预定义属性标注的约束。

D-LA：代表具有区分性的隐性属性的子空间，它在学习中受到训练样本类别信息的约束。

B-LA：代表噪音，在学习过程中没有约束。

语义标注和隐性属性的联合学习

预定义和隐性属性的联合学习模型

字典子空间分解:
- UDAC (D^u): 和预定义属性相关。
- D-LA (D^d):代表具有区分性的隐性属性的子空间。
- B-LA (D^b):代表背景等噪音信息。

阶梯化的重构误差项:
- 第一项的目的是让UDAC和D-LA尽可能的重构训练集的原始特征Y。
- 第二项的目的是为了让 B-LA 去重构UDAC和 D-LA 无法解释的部分，即残差部分。
- 两项联合相当于增加了 UDAC和 D-LA 的优化权重，即保证大多数的图像信息可以被UDAC和D-LA建模。

$$\left[D^u, D^d, D^b, W\right] = \arg\min \left\| Y - D^u X^u - D^d X^d \right\|_F^2 + \left\| Y - D^u X^u - D^d X^d - D^b X^b \right\|_F^2$$
$$+ \beta^2 \left\| X^u - WA \right\|_F^2 + \alpha \sum_{i,j=1}^{n} m_{i,j} \left\| x_i^d - x_j^d \right\|_2^2, \; s.t. \left\| d_i^u \right\|_2^2 \le 1, \left\| d_i^d \right\|_2^2 \le 1, \left\| d_i^b \right\|_2^2 \le 1, \left\| w_i \right\|_2^2 \le 1$$

额外说明
- 在没有预定义属性标注A时，本方法依然有效。
- 在这种情况下，该方法只学D-LA 和 B-LA。

UDAC的约束方式:
- 建立 UDAC 的系数 X^u 与预定义属性的标注矩阵 A 之间的线性约束关系 W。
- 预定义属性本身也可以看做是一个子空间，但预定义属性来源于人为定义，因此不适合直接作为字典子空间的基。
- 因此采用线性关系建模预定义属性空间和字典子空间之间的关系。

D-LA的约束方式
- 利用 Graph Laplacian 项约束 D-LA 的系数。$m_{i,j} = 1$代表样本 i 和样本 j 为同一个人。反之$m_{i,j} = 0$
- 目的在于同一个人的不同图像上拥有相似的D-LA。

语义标注和隐性属性的联合学习

- 模型求解思路

$$\left[D^u, D^d, D^b, W\right] = \arg\min \left\| Y - D^u X^u - D^d X^d \right\|_F^2 + \left\| Y - D^u X^u - D^d X^d - D^b X^b \right\|_F^2$$

$$+\beta^2 \left\| X^u - WA \right\|_F^2 + \alpha \sum_{i,j=1}^{n} m_{i,j} \left\| x_i^d - x_j^d \right\|^2, \; s.t. \left\| d_i^u \right\|_2^2 \le 1, \left\| d_i^d \right\|_2^2 \le 1, \left\| d_i^b \right\|_2^2 \le 1, \left\| w_i \right\|_2^2 \le 1.$$

  - 模型中有大量的未知量, 包括$D^u, D^d, D^b, W, X^u, X^d$和$X^b$ ， 因此无法直接求解。
  - 值得注意的是，原模型的目标函数非负，根据“有下界的单调递减数列必收敛”这一定理，可以通过每次求局部最优迭代求解。
  - 参考字典学习的标准算法，每次固定其他未知项，只更新一个未知量。

利用语义标注和隐性属性表征行人

测试样本的特征向量 y

字典编码

$$\left[x^u, x^d, x^b\right] = \arg\min \left\| y - D^u x^u - D^d x^d - D^b x^b \right\|_2^2 + \gamma\left(\left\|x^u\right\|_2^2 + \left\|x^d\right\|_2^2 + \left\|x^b\right\|_2^2\right)$$

D-LA 向量 x^u
B-LA 向量 x^b

$$\tilde{x} = \left(\tilde{D}'\tilde{D} + 2\gamma I\right)^{-1}\tilde{D}'\tilde{y}.$$

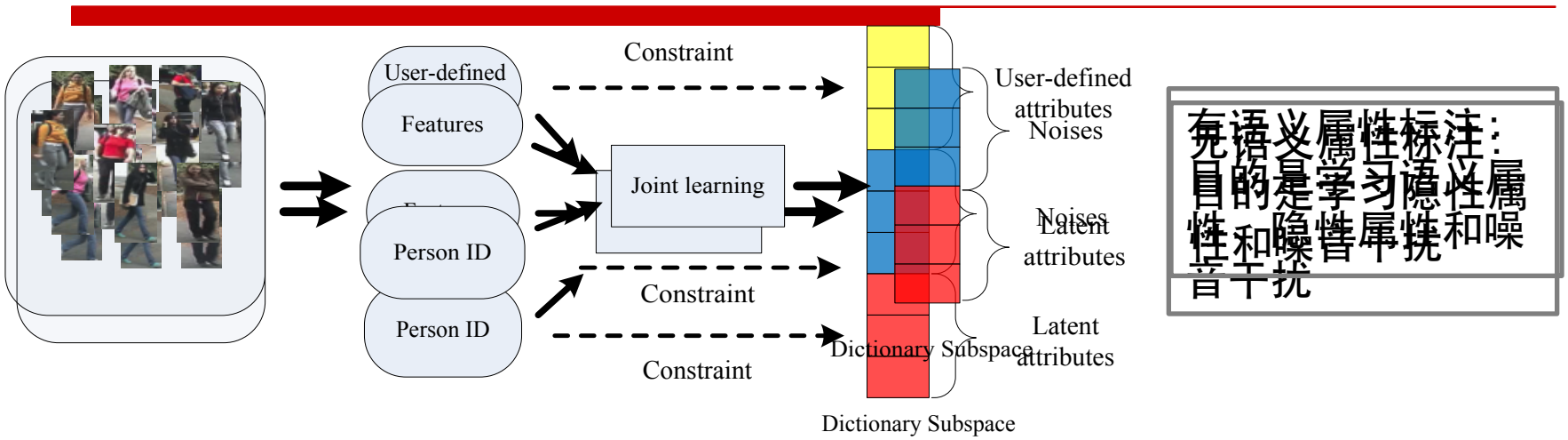
根据 x^u 进行
语义属性转化

$$a = \arg\min \left\| x^u - Wa \right\|_2^2 + \gamma \left\| a \right\|_2^2$$

语义属性向量 a

每一个测试样本y 都可以通过它基于预定义属性的编码向量a和基于D-LA 的编码向量x^d 来表示

没有语义属性标注时...



有语义属性标注：
目的是学习隐性属性和噪音干扰

字典子空间分解:
- D-LA (D^d):代表具有区分性的隐性属性的子空间
- B-LA (D^b):代表背景等噪音信息。

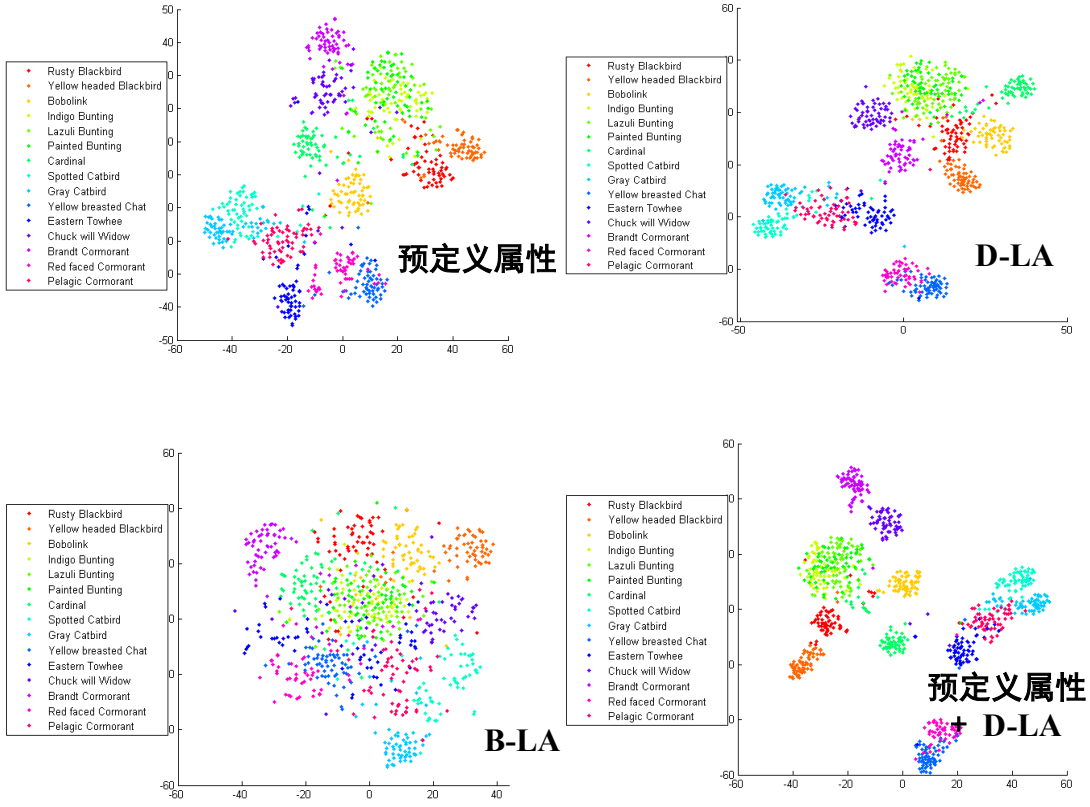
阶梯化的重构误差项:
- 第一项的目的是让D-LA尽可能的重构训练集的原始特征Y。
- 第二项的目的是为了让 B-LA 去重构D-LA 无法解释的部分，即残差部分。
- 两项联合保证大多数的图像信息可以被D-LA 建模。

$$\left[D^d, D^b\right] = \arg\min \left\|Y - D^d X^d\right\|_F^2 + \left\|Y - D^d X^d - D^b X^b\right\|_F^2 + \alpha \sum_{i,j=1}^{n} m_{i,j} \left\|x_i^d - x_j^d\right\|^2,$$

$$s.t. \left\|d_i^d\right\|_2^2 \le 1, \left\|d_i^b\right\|_2^2 \le 1.$$

D-LA的约束方式
- 利用 Graph Laplacian 项约束 D-LA 的系数。$m_{i,j} = 1$ 代表样本 i 和样本 j 为同一个人。反之$m_{i,j} = 0$
- 目的在于同一个人的不同图像上拥有相似的D-LA。

隐性属性的可视化分析



CUB数据集上，测试集里10个种类的D-LA分布图。 从图上表明， 所学到的D-LA 有很明显的区分性。

一些 D-LA 的例子，当没有语义属性标注不存在时，D-LA往往可以发现语义属性。

一些 D-LA 的例子，当语义属性标注存在时，D-LA往往可以发现一些其他很难被定义或者无法被定义的视觉属性。

北京大学

行人Re-ID

即使用于行人 Re-ID，也能达到现有方法的前沿性能

		VIPeR	PRID	Market	CUHK03 (M)	CUKH03(D)
传统方法	MFA (L)	39.6	20.9	45.7	-	
	kLFDA (L)	39.9	21.6	51.4	-	
	XQDA (L)	40.0	-	43.8	52.2	46.3
	NFST (L)	42.5	**29.8**	55.4	58.9	53.7
	Ours_L (L)	**42.3**	25.1	**61.1**	**60.8**	**58.4**
属性方法	aMTL (L)	42.3	18.0	-	-	-
	Ours_U (L)	28.4	16.3	-	-	-
	Ours_All (L)	**45.4**	**26.8**	**-**	**-**	**-**
深度方法	DeepReID				20.6	19.9
	Improved Deep	34.8			54.7	45.0
	FT-JSTL+DGD	38.6	-	-	75.3	
	Zhang et al	43.0	-	-	57.0	51.2
	Wang et al	35.8	-	-	52.2	-
	Ours_L (C)	**51.8**	**37.5**	**65.7**	**77.5**	**66.5**

Ours_L 代表没有语义属性标注，只利用学到的隐性属性去做行人Re-ID。
Ours_All 代表存在语义属性标注，利用语义属性和隐性属性一起做行人Re-ID
（L）代表传统特征，（C）代表深度特征。

实验分析：

- ***所学到的D-LA 对行人Re-ID有帮助。***
- ***在有预定义属性存在时，Ours_U 的性能很差，这说明只利用预定义属性通常区分性不够。***
- ***学到的预定义属性信息对行人Re-ID有帮助，并且预定义属性和隐性属性之间信息互补。***
- ***对不同特征都有效，在深度特征上依然适用。***

属性预测与ZSL分类

- ATT：对未知的图像进行属性预测
- Zero-shot 图像分类(ZSL)：根据属性进行分类，每个类别只有一个属性描述，没有训练样本。

Method	ZSL		ATT	
	AwA	CUB	AwA	CUB
DAP	57.5	-	72.8	61.8
ALE	43.5	18.0	65.7	60.3
UMF	48.6	18.2	-	-
CSHAP	45.6	17.5	74.3	68.7
SSE	76.3	30.4	-	-
SJE	61.9	40.3	-	-
JLSE	80.5	42.1	-	-
SC	72.9	54.5		
MSS	-	56.5		
LatEm	76.1	51.7	-	-
Ours	**82.9**	**56.9**	**79.6**	**78.3**

实验分析：

- ***所提方法不仅可以学习到 D-LA 用于行人 Re-ID任务，也能获得更好地帮助学习预定义属性模型。***

P. Peng, Y. Tian, T. Xiang, Y. Wang, and T. Huang, "Joint learning of semantic and latent attributes," in ECCV, 2016.

回到核心问题：如何减少对训练样本依赖？

- 属性迁移学习
  - **目标数据：**当前需要完成Re-ID 任务的监控场景中收集到的训练数据，无标注信息。
  - **源域：**从其他监控场景中收集到的有标注训练数据。
  - **迁移目的：**从源数据中学到对行人 Re-ID 有帮助的信息，并且跨数据迁移到目标数据上。

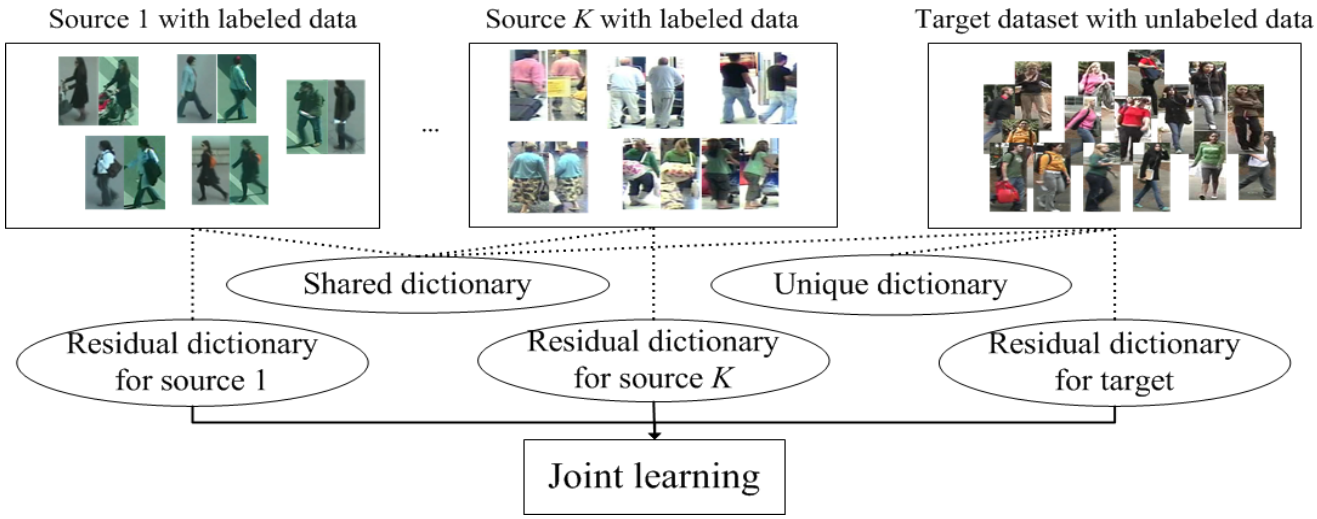
- 优势
  - 减少标注数据的代价
  - 增加已有标注数据的可利用率
  - 增加 Re-ID 方法的适用性

属性迁移学习

- 迁移什么？
  - 属性模型与场景无关，可以做到跨数据集共享。
  - 因此将语义属性和所学的 D-LA 作为迁移信息。
- 怎么迁移？
  - UDAC 和 D-LA 子空间跨数据集共享。
- 如何避免迁移无效信息？
  - 每个数据集上单独构造 B-LA。
  - B-LA建模每个数据集上无法共享的信息。
- 如何保证迁移信息的完备性？
  - 非对称学习模式，模型偏向于目标数据。

属性迁移学习

跨数据集属性迁移学习框架



框架分析：

- 将跨数据集迁移问题建模成非对称的多任务字典学习模型，每一个数据对应一个任务。
- 共享字典用于迁移数据集共享的语义属性和 D-LA
- 残差字典为每个数据集特有，用于建模每个数据集中无法共享的部分。
- 为目标数据集额外增加一个独有字典，使得学到的模型偏向于目标数据集。
- 不同字典通过统一学习得到，使得它们之间有信息互补。

非对称多任务字典学习模型

$$\left[D^u, D^{ds}, D^{du}, D_1^b, \cdots, D_T^b, W\right] = \arg\min \sum_{t=1}^{T-1} \left\|Y_t - D^u X_t^u - D^{ds} X_t^{ds}\right\|_F^2$$

$$+ \sum_{t=1}^{T-1} \left\|Y_t - D^u X_t^u - D^{ds} X_t^{ds} - D_t^b X_t^b\right\|_F^2 + \left\|Y_T - D^u X_T^u - D^{ds} X_T^{ds} - D^{du} X_T^{du}\right\|_F^2$$

$$+ \left\|Y_T - D^u X_T^u - D^{ds} X_T^{ds} - D^{du} X_T^{du} - D^b X_T^b\right\|_F^2 + \alpha \sum_{t=1}^{T} \sum_{i,j=1}^{N_t} m_{t,i,j} \left\|x_{t,i}^{ds} - x_{t,j}^{ds}\right\|^2$$

$$+ \alpha \sum_{i,j=1}^{N_T} m_{T,i,j} \left\|x_{T,i}^{du} - x_{T,j}^{du}\right\|^2 + \alpha \sum_{i,j=1}^{N_T} m_{T,i,j} \left\|x_{T,i}^{u} - x_{T,j}^{u}\right\|^2 + \beta^2 \sum_{t=1}^{T-1} \left\|X_t^u - W A_t\right\|_F^2,$$

$$s.t.\ \left\|d_i^u\right\|_2^2 \le 1,\ \left\|d_i^{ds}\right\|_2^2 \le 1,\ \left\|d_i^{du}\right\|_2^2 \le 1,\ \left\|d_{t,i}^b\right\|_2^2 \le 1, \|w_i\|_2^2 \le 1,\ \forall i, t.$$

一些说明:

- **1,…,T-1 为源任务，T为目标任务。**
- **D^u: 代表与语义属性相关的子空间。**
- **D^{ds}: 代表数据集共享的 D-LA 子空间**
- **D^{du}: 代表目标数据集上独有的 D-LA 子空间。**
- **D_t^b 用于建模背景等每个数据集上特有的无法共享的信息。**

同样当源数据集上不存在预定义属性时，该模型同样适用

算法 5: 基于跨数据集迁移学习的行人Re-ID算法

Input: Y_t; 随机初始化 D^u, D^{ds} D^{du} 和 D_t^b; $X_t^b \to 0$;
Output: D^u, D^{ds},D^{du} ,D_t^b, D^u 和 W $(t = 1, ..., T)$.
while *Non-convergence* **do**
  for $t = 1 \to T$ **do**
    if 源任务 **then**
      固定其他项，求解系数X_t^{ds};
      固定其他项，求解系数X_t^u;
    if 目标任务 **then**
      固定其他项，求解系数X_t^u, X_t^{ds}和 X_t^{du};
    固定其他项，求解系数X_t^b;
  固定其他项，更新字典矩阵D^u;
  固定其他项，更新字典矩阵D^{ds};
  for $t = 1 \to T$ **do**
    if 源任务 **then**
      固定其他项，更新字典矩阵D_t^b;
    if 目标任务 **then**
      固定其他项，更新字典矩阵D^{du};
      固定其他项，更新字典矩阵D_t^b;
    固定其他项，更新字典矩阵W.

北京大学

实验结果

无监督设置下实验

数据集	VIPeR	PRID	CAVIAR	iLID
SDALF	19.9	16.3	-	29.0
eSDC	26.7	-	-	36.8
GTS	25.1	-	-	42.4
ISR	27.0	17.0	29.0	39.5
Our_S	26.9	14.1	34.8	45.7
kLFDA_N	15.9	9.1	32.8	36.9
SA_DA +kLFDA	15.2	8.7	32.1	35.8
AdaRSVM	10.9	4.9	-	
Ours_L	**31.5**	**23.4**	**41.6**	**49.3**
Ours_ALL	**34.6**	**25.6**	**43.4**	**51.1**

Ours_S：本方法的单任务版本

Ours_L代表本方法在学习过程中没有使用预定义属性的标注信息

Ours_ALL表示使用了源数据集上预定义属性

实验分析：

- kLFDA_N 性能很差，证明现有的有监督行人 Re-ID 模型只适用于当前数据集。
- 一般的域迁移算法 SA_DA 不适用于跨数据集迁移学习问题，因为跨数据集迁移比域迁移问题更难。
- 同类型方法AdaRSVM性能较差，证明我们方法的优越性。

实验结果

半监督设置下实验

数据集	VIPeR	PRID	CAVIAR	iLID
SSCDL	25.6	-	49.1	-
kLFDA	27.5	14.1	35.7	41.6
kCCA	24.6	5.3	-	-
kLFDA_N	18.4	12.4	34.8	38.4
cAMT	16.2	13.5	29.1	33.6
Ours_L	**34.1**	**24.9**	**47.3**	**50.2**

各模块贡献分析

数据集	VIPeR	PRID	CAVIAR	iLID
无非对称性	27.2	22.3	38.1	46.5
无阶梯式优化	23.8	18.9	35.7	44.2
完整模型	**31.5**	**24.2**	**41.6**	**49.3**

无监督设置下，各个模块的贡献分析，结果表明每一个设计的模块都发挥了正面作用。

实验分析：

- kLFDA 和 kCCA 性能很差，证明现有的有监督行人 Re-ID 方法需要较多的标注数据。
- 半监督设置下的实验性能都好于无监督设置下的实验性能，说明我们的方法不仅能存源数据获益，还能利用当前数据集的标注信息。
- 与同类型方法cAMT的比较结果，证明我们方法的优越性。

P. Peng, T. Xiang, Y. Wang, M. Pontil, S. Gong, T. Huang, Y. Tian, "Unsupervised Cross-Dataset Transfer Learning for Person Re-identification," in CVPR, 2016.

深度学习应用于行人搜索

- 面临的挑战：缺乏足够规模的数据来训练深度网络
  - 深度网络模型参数较多，需要较多的训练数据
  - 有监督学习的行人再识别算法难以推广
- 解决方案
  - 从大规模有标注数据到小规模无标注数据的无监督迁移学习
  - 深度网络与字典学习方法相结合
  - 在深度网络训练上，利用特征点的KNN Graph建立伪标注（Soft labels），配合有监督分步学习策略来实现无监督迁移学习

Algorithm 1 Soft Label Generation for Unsupervised Deep Re-ID

Input: Training image set $X = \{x_1^a, ..., x_{m_1}^a, x_1^b, ..., x_{m_2}^b\}$, base network ϕ, parameter k
Output: Soft label set $Y = \{y_1^a, ..., y_{m_1}^a, y_1^b, ..., y_{m_2}^b\}$
1: **For each** $y_i \in Y$ **do**:
2: $y_i = \varnothing$;
3: Compute feature set $P = \{\phi(x_1^b), ...\phi(x_{m_2}^b)\}$;
4: **For each** $x_i^a \in X$ **do**:
5: Compute $\phi(x_i^a)$;
6: $y_i^a = \{i\}$;
7: Calculate the k-nearest neighbor set N_i of $\phi(x_i^a)$ in P;
8: **For each** $\phi(x_i^b) \in N_i$ **do**:
9: $y_i^b = y_i^b \cup \{i\}$;

无监督行人Re-ID

- 算法性能在无监督情况下超过大部分有监督模型

	VIPeR	PRID	CUHK01 (N_t=871/485)
SCSP [3]	53.5	-	-
LSSCDL [62]	42.6	-	-
TMA [34]	43.8	-	-
ℓ1 GL [19]	41.5	30.1	-/50.1
Siamese LSTM [50]	42.4	-	-
Metric Ensemble [37]	45.9	-	-
DNS [60]	51.1	40.9	-/69.0
IDLA [1]	34.8	-	65.0/47.5
DGD [54]	38.6	64.0*	-/66.6
MCP-CNN [4]	47.8	22.0	-/53.7
Gated S-CNN [49]	37.8	-	-
EDM [41]	40.9	-	86.6/-
Joint Learning [51]	35.8	-	72.5/-
CAN [28]	-	-	81.0/-

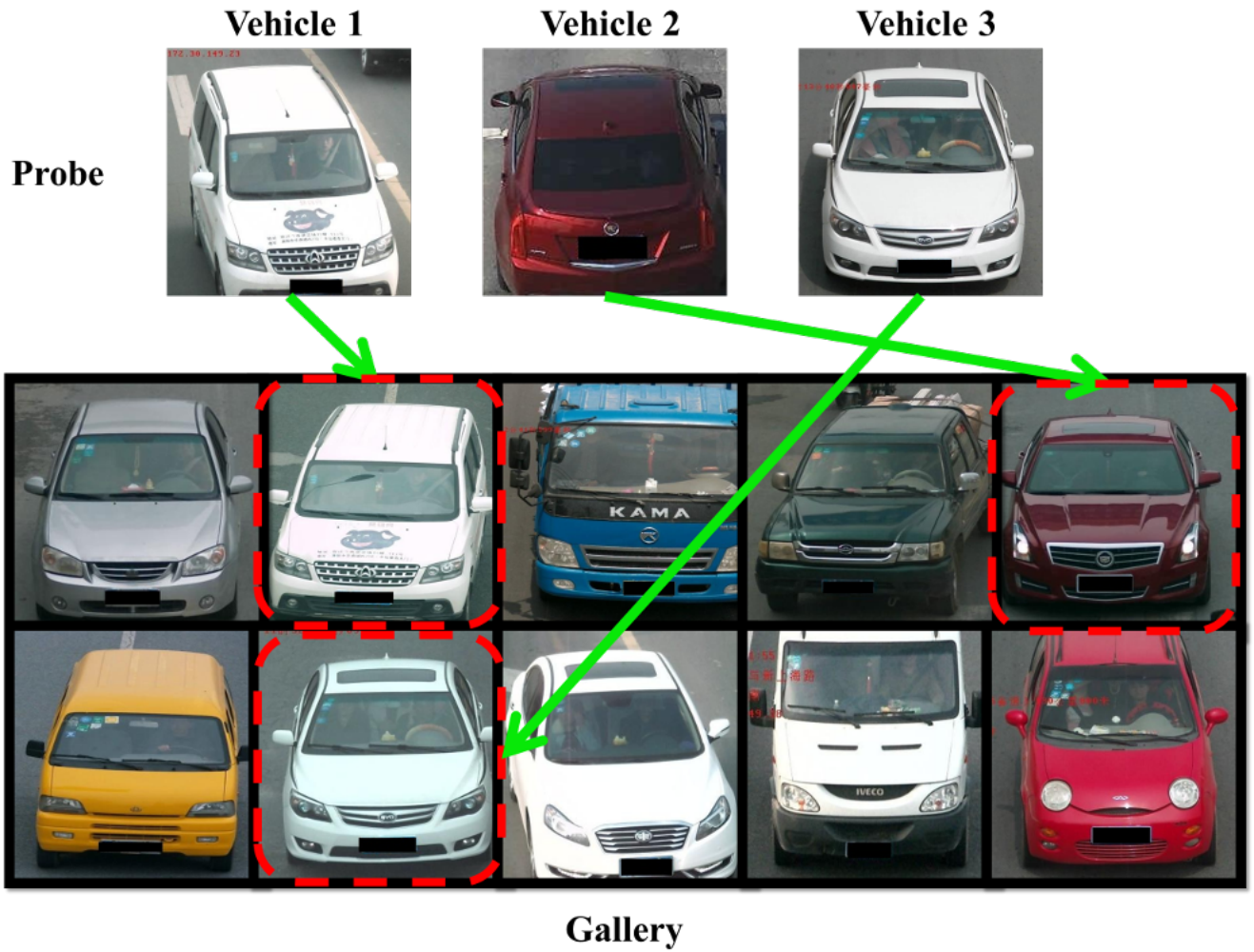
Supervised Result of State-of-the-art Methods

	VIPeR	PRID	CUHK01
DLLR [20]	29.6	21.1	-
CDTL [39]	31.5	24.2	27.1
ℓ1 GL [19]	33.5	25.0	41.0
Ours	**45.1**	**36.2**	**68.8**

Table 7. Unsupervised transfer learning results

Mengyue Geng, Yonghong Tian, et a., Deep Transfer Learning for Person Re-Identification, ICCV 2017, submitted.

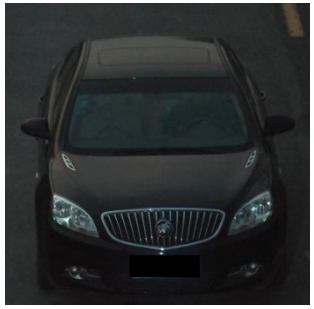
PART II: Precise vehicle search



车辆精准搜索

- NOT an easy task
  - **The Twin Problem**: It is very difficult to distinguish two cars from the same model and with the same color








正常　　视角不同　　光线不同　　检测结果差

车辆精准搜索

- Is it really possible to distinguish two vehicles of the same model and color?
  - Yes, if we can find some discriminative features
  - Attributes help precise vehicle search

大部分行人可以通过属性特征区分（如衣服颜色、性别、头发长度等）

寻找目标的独有特征

车辆精准搜索

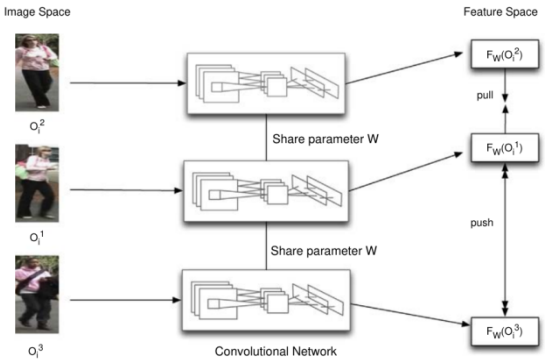
- Appearance-based coarse filtering: low-level hand-crafted features and high-level semantic attributes
- Plate-based accurate search : a Siamese neural network is trained for license plate verification instead of recognizing the characters
- Spatiotemporal relation model : utilized to re-rank vehicles

Liu, Xinchen, et al. "A Deep Learning-Based Approach to Progressive Vehicle Re-identification for Urban Surveillance." ECCV, 2016.

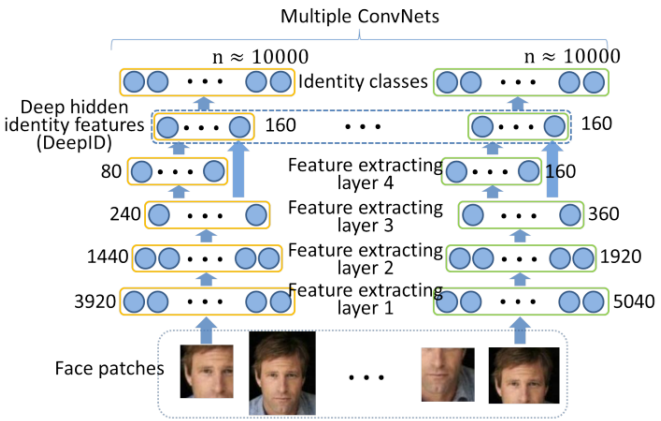
如何在深度网络框架下学习高区分特征?

- 结合深度神经网络与度量学习(Triplet Loss)
  - Deep feature learning with relative distance comparison for person re-identification(Pattern Recognition 2015)
- 结合识别与校验的深度神经网络 (DeepID)
  - Deep Learning Face Representation from Predicting 10,000 Classes (CVPR 2014)

Triplet Loss

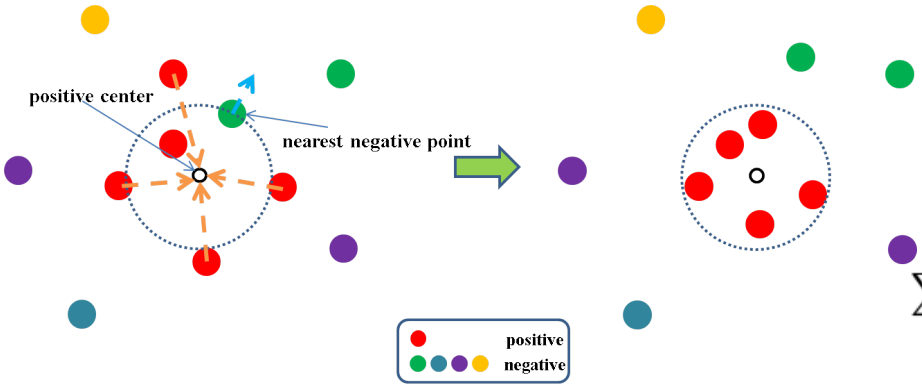


DeepID



如何在深度网络框架下学习高区分特征?

- 深度相对距离学习
  - 把原始图片映射到一个特殊的高维欧式空间，然后利用欧式距离进行相似度度量
  - 双分支网络，同时提取车型类别信息和个体特有特征(混合差分网络)
  - 样本类内和类间的相对距离约束(簇聚类损失函数)



样本间距离约束

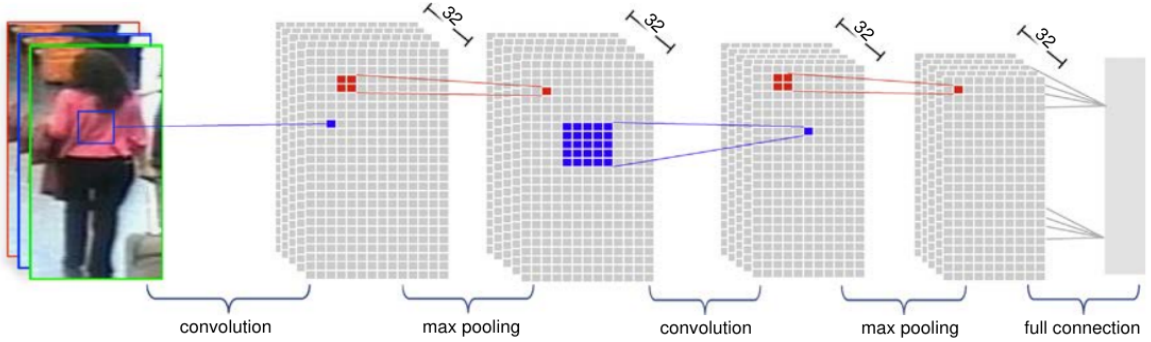
$$c^p = \frac{1}{N^p}\sum_i^{N^p} f(x_i^p)$$

$$L(W, X^p, X^n) = \sum_i^{N^p} \frac{1}{2} max\{0, \|f(x_i^p - c^p)\|_2^2 + \alpha - \|f(x_*^n) - c^p\|_2^2\}$$

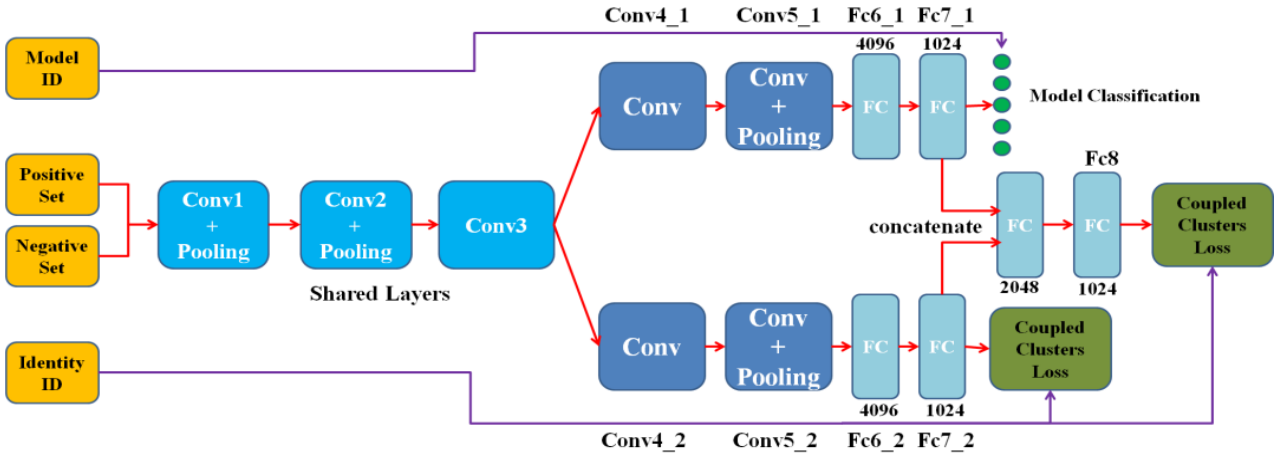
目标函数：类内距离小于类间距离

深度相对距离学习

- 多任务混合差分网络



(a) 传统网络模型要么仅能学习属性特征进行属性分类，要么仅能学习判别性特征进行不同样本间的距离度量



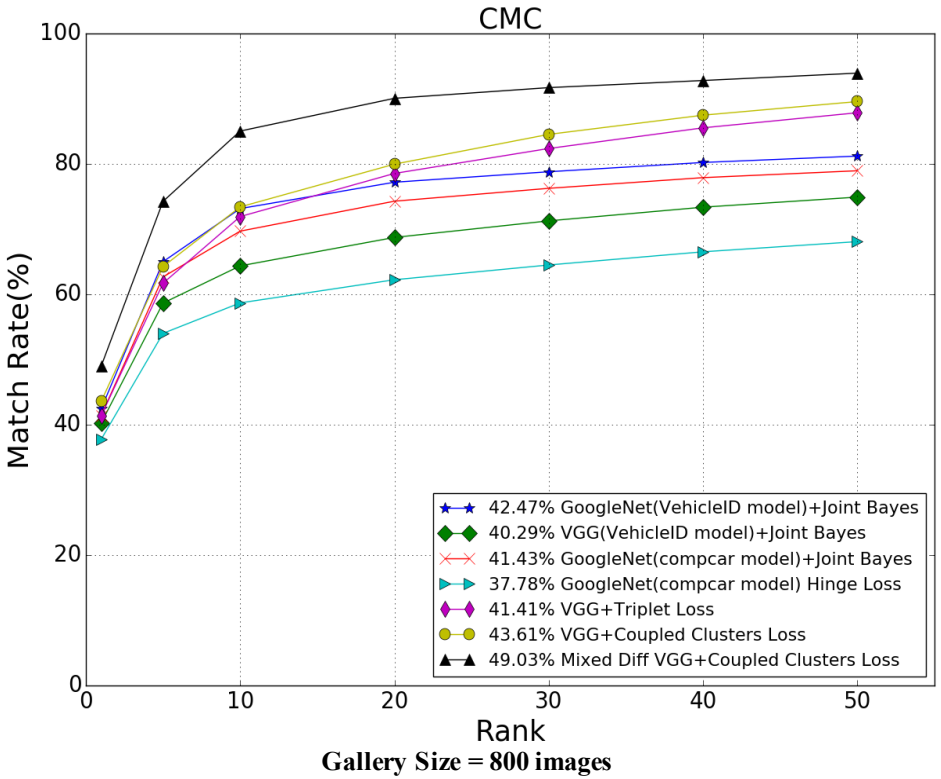
(b)通过一个多任务混合差分网络同时学习属性特征与判别性特征

实验

测试数据集

22万实拍车辆图片（包含2万辆车）

车辆再标识实验



Gallery Size = 800 images

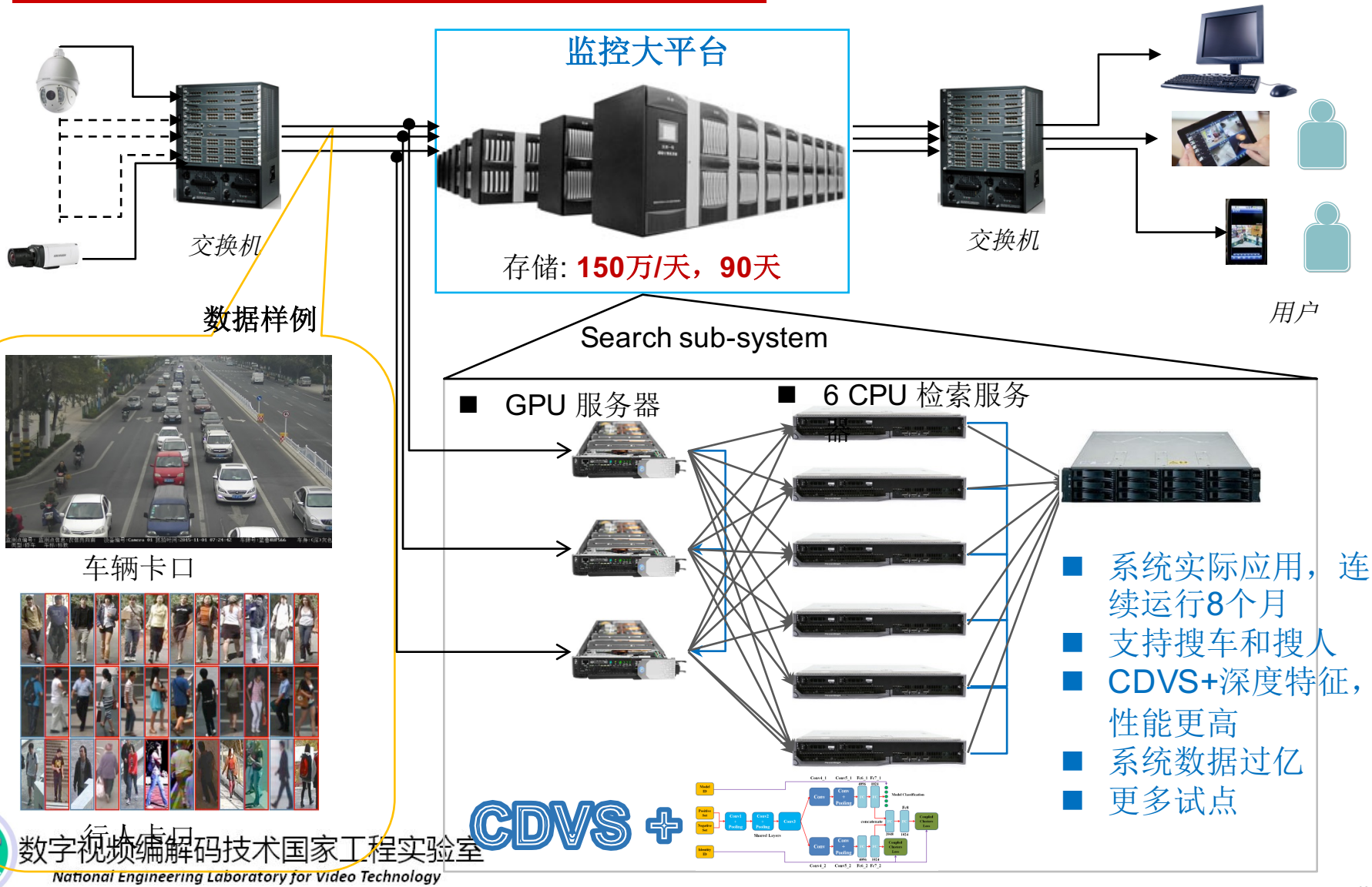
车辆检索实验

Table 4. MAP of Vehicle Retrieval Task

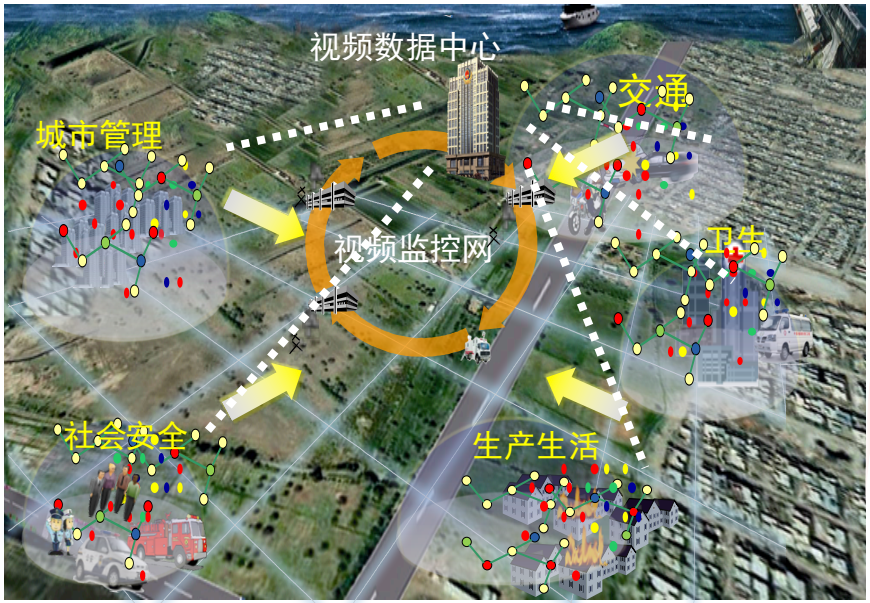
MAP	Small	Medium	Large
VGG+Triplet Loss[5]	0.444	0.391	0.373
VGG+CCL(Ours)	0.492	0.448	0.386
Mixed Diff+CCL(Ours)	0.546	0.481	0.455

Hongye Liu, Yonghong Tian, Yaowei Wang, Lu Pang, Tiejun Huang, Deep Relative Distance Learning: Tell the Difference Between Similar Vehicles, CVPR, 2016, 2167-2175

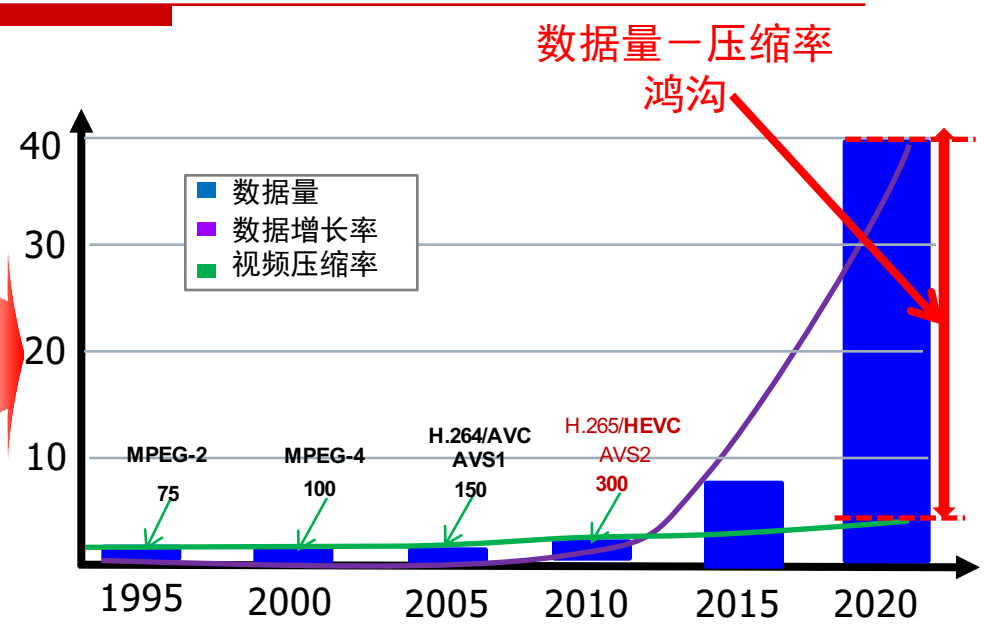
监控目标精准搜索系统


监控大平台
存储: 150万/天，90天
交换机
交换机
用户
数据样例
车辆卡口
行人卡口
Search sub-system
GPU 服务器
6 CPU 检索服务器
CDVS +
系统实际应用，连续运行8个月
支持搜车和搜人
CDVS+深度特征，性能更高
系统数据过亿
更多试点

一直被忽略的首要问题：无法汇聚！



视频大数据：数据量巨大、存储分散



数据量两年增长一倍←→压缩率十年增长一倍

10万监控摄像头
(一个省>100万)

…×10Mbps×1月
= 10 EB
×200万/PB
=7亿元

H.264/AVC

…×5Mbps×1月
=5EB
…
=3.5亿元

H.265/AVS2

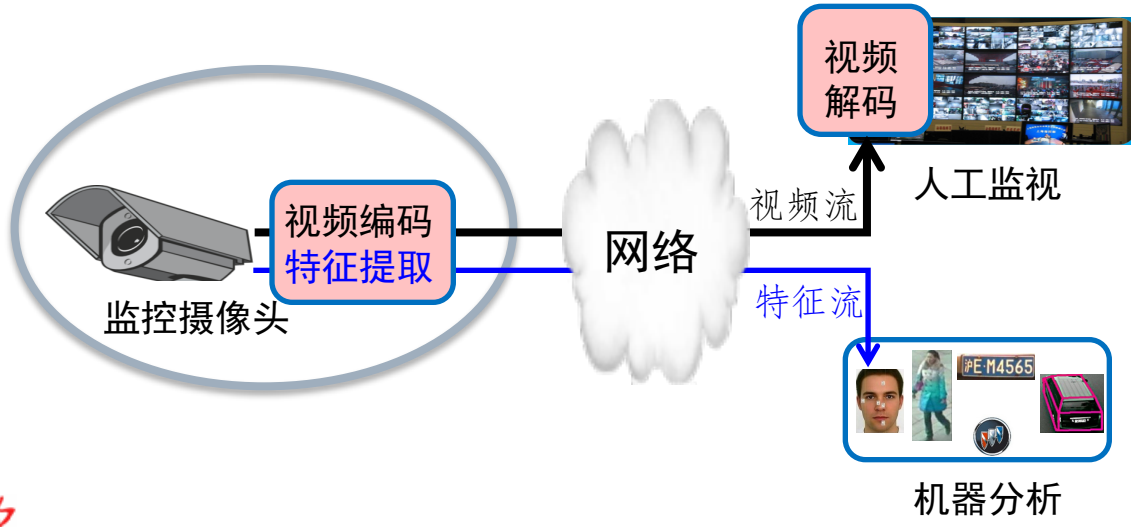
百万路监控视频
存储和汇聚?

“数据大”≠“大数据”

如何支持大规模汇聚分析？

- 视频编码不能完全解决大规模视频汇聚分析问题
  - 即使压缩比达1000:1：每路高清约1.5Mbps
  - 传输10万路高清视频：需要150Gbps

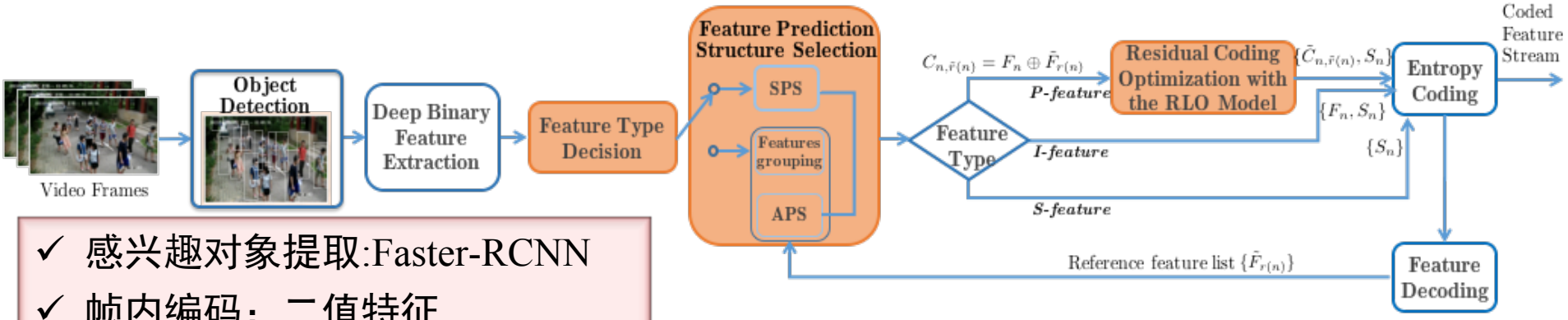
  →五年内不现实！（如某省公安专网带宽2.5Gbps）
- 视频监控的核心手段：机器分析与识别
  - 只需传输分析识别所需特征：数据量小(可压缩至<150K)
  - 处理架构：**先压后解再分析→源端压缩与特征提取同步**



视觉特征的紧凑表示

- 从紧凑特征表示角度出发，解决监控视频大数据分析、识别规模化应用所面临的视觉特征汇聚以及搜索分类任务提出的大规模、高性能、实时性等挑战。

深度特征的紧凑描述学习框架

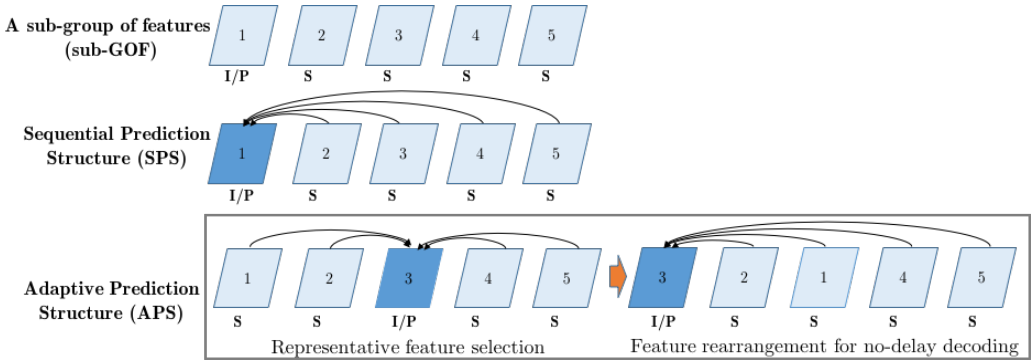


- ✓ 感兴趣对象提取:Faster-RCNN
- ✓ 帧内编码：二值特征

  CNN hash: GoogleNet+Sigmoid

不同特征编码类型

- ✓ I-Feature代表较大的场景变化，需要独立编码(动态GOP长度)
- ✓ P-Feature代表一定程度的场景变化，需要编码残差
- ✓ S-Feature代表较小的场景变化，可以用前一帧的特征表示



北京大学

特征紧凑表示的优化模型

输入：图像视频数据集合 V　　输出：图像视频数据的高效表达 S^*

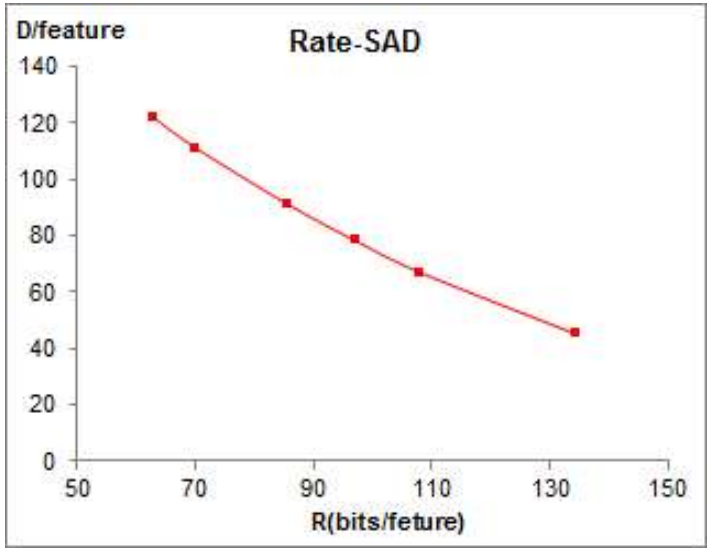
$$S^* = \arg\min_S \mathcal{L}(S|V,T;\Theta_L) + \lambda\mathcal{R}(S;\Theta_R)$$

特征率失准

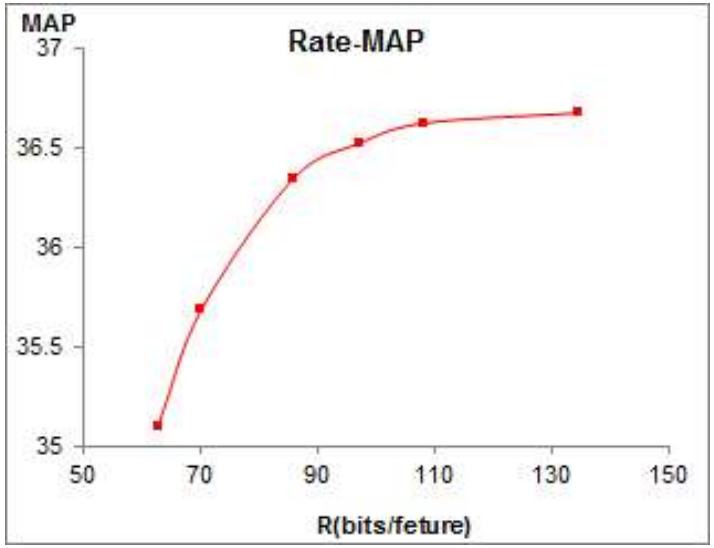
模型：损失函数 $\mathcal{L}$ 表示为压缩导致的检索识别性能损失 $\mathcal{L}(D)$

目标：建立特征精简表达，实现既“准”又“小”

难点：如何建立特征压缩率–检索识别性能损失间的相关模型？



感知性失真度量(SAD)与码率关系

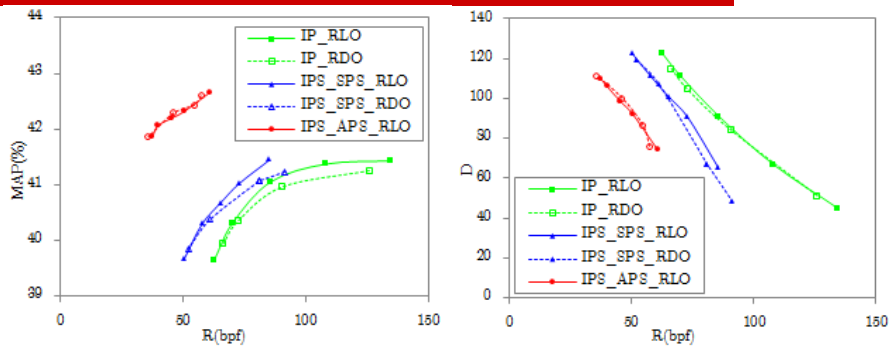


视觉检索性能(MAP)与码率关系

实验结果

CDVA数据集：

MPEG的测试数据集（1200个参考视频，约2M帧）



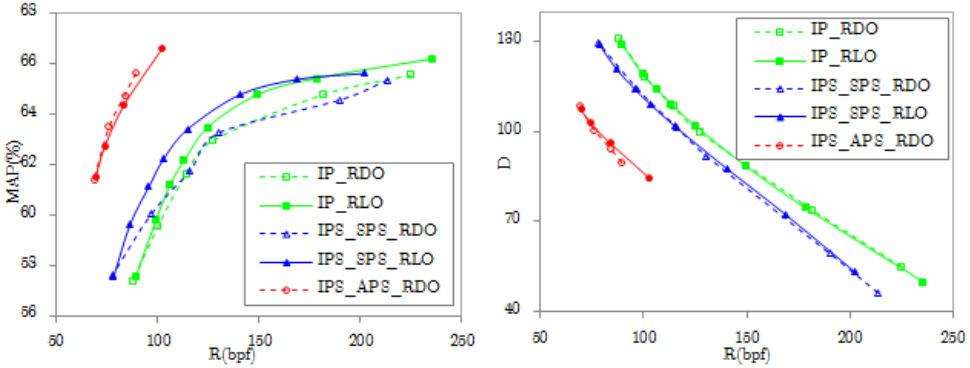
(a) (b)

Fig. 19: Comparisons between RDO and RLO for the CDVA dataset: a) Rate-MAP curves, b) Rate-Distortion curves.

平均每帧50bit，即可达到与无压缩特征（41.43%）更高的检索性能（压缩过程中引入了特征选择）

VFC-1M数据集：

监控视频数据集（60个无压缩视频，约1M对象）



(a) (b)

Fig. 20: Comparisons between RDO and RLO for the VFC-1M dataset: a) Rate-MAP curves, b) Rate-Distortion curves.

平均每帧100bit，即可达到与无压缩特征（54.2%）更高的检索性能（压缩过程中引入了特征选择）

Lin Ding, Yonghong Tian, Yaowei Wang, Hongfei Fan, Tiejun Huang, Rate-Performance-Loss Optimization for Inter-frame Deep Feature Coding from Videos, IEEE TIP, 2017, submitted.

总结

- 精准对象搜索
  - Beyond visual search, and object re-identification
  - Search as recognition (SaR)
  - (Traditional) image search → Fine-grained image search→ Precise object search
  - Compact descriptors for multi-tasks
- 未来研究方向
  - Effectiveness & Efficiency
  - Large-Scale Benchmarking: Billions-scale benchmark dataset
  - Multi-task Feature: More discriminative global and local deep features, for both fine-grained categorization and search
  - Unified Framework: One framework for detection, recognition and search

教育部学位与研究生教育发展中心
China Academic Degrees & Graduate Education Development Center
全国工程专业学位研究生教育指导委员会
数字音视频编解码（AVS）产业技术创新战略联盟

中国科协青少年科技中心
中国智慧城市产业技术创新战略联盟
北京航空航天大学 BEIHANG UNIVERSITY

“全国研究生创新实践系列活动”之：

研究生智慧城市技术与创意设计大赛
智能技术挑战赛

（西南交大@2017）

http://www.smartcity-competition.com.cn
http://www.bigmmchallenge.org

- 比赛方式：在线擂台赛和现场答辩相结合
- 四大任务：行人精准搜索、车辆精准搜索、异常行为检测、无人机飞行场景重建
- 时间节点：5月1日擂台赛启动；8月下旬决赛
- 多点共赢：成果直接可投稿顶级会议与期刊
  企业界对人才的重要衡量标准，华为、海康…
  教育部/科协认可的比赛证书、丰厚奖金…

Thanks!
Q&A?

北京大学