

中国科学院深圳先进技术研究院
深圳市商汤科技有限公司
蒋小可博士后出站答辩

行人重识别的优化研究

导师: 乔宇 闫俊杰
答辩人: 蒋小可

2021.6.8

目录

- 典型场景和挑战
- 两方面优化
 - attention视频ReID：基于现场视频信息的ReID优化
 - 行为模式：基于长期行为模式的ReID优化
- 研究总结
- 学术成果

典型场景和挑战

目标：充分利用能获得的数据，提升ReID准确性

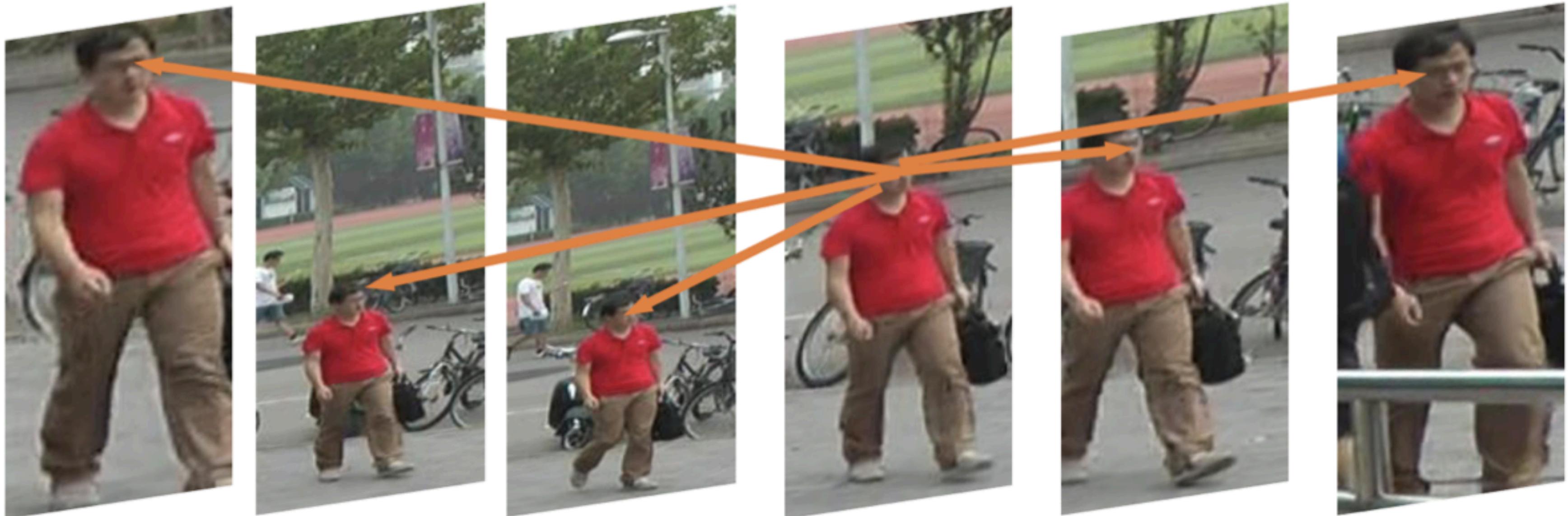
- ReID数据更多，更容易获得，但特征区分力比较弱
 - 行人走动的时候拍摄，没有刻意配合
 - 人体非刚体，不同姿态，光线，摄像机角度
- 原始数据
 - 大部分场景有视频数据
- 系统积累
 - 长时间运行积累行为数据

解决方案：基于长期和短期时空信息的优化

- 现场视频数据：
 - 现场时空信息
 - 视频ReID，多帧的信息，互为补充
 - 可变信息互补：不同角度的人体不同的视觉信息互补，拼凑一个更完整的人体信息，从而获得更好的人体特征
 - 不变信息去噪：帧间信息把人体与背景部分区分开来，从而减少噪声干扰
- 长期行为数据：基于行为模式
 - 长期行为数据
 - 贝叶斯概率约束
 - 某人出现在此时此地的概率
 - 从A到B事件顺序的约束

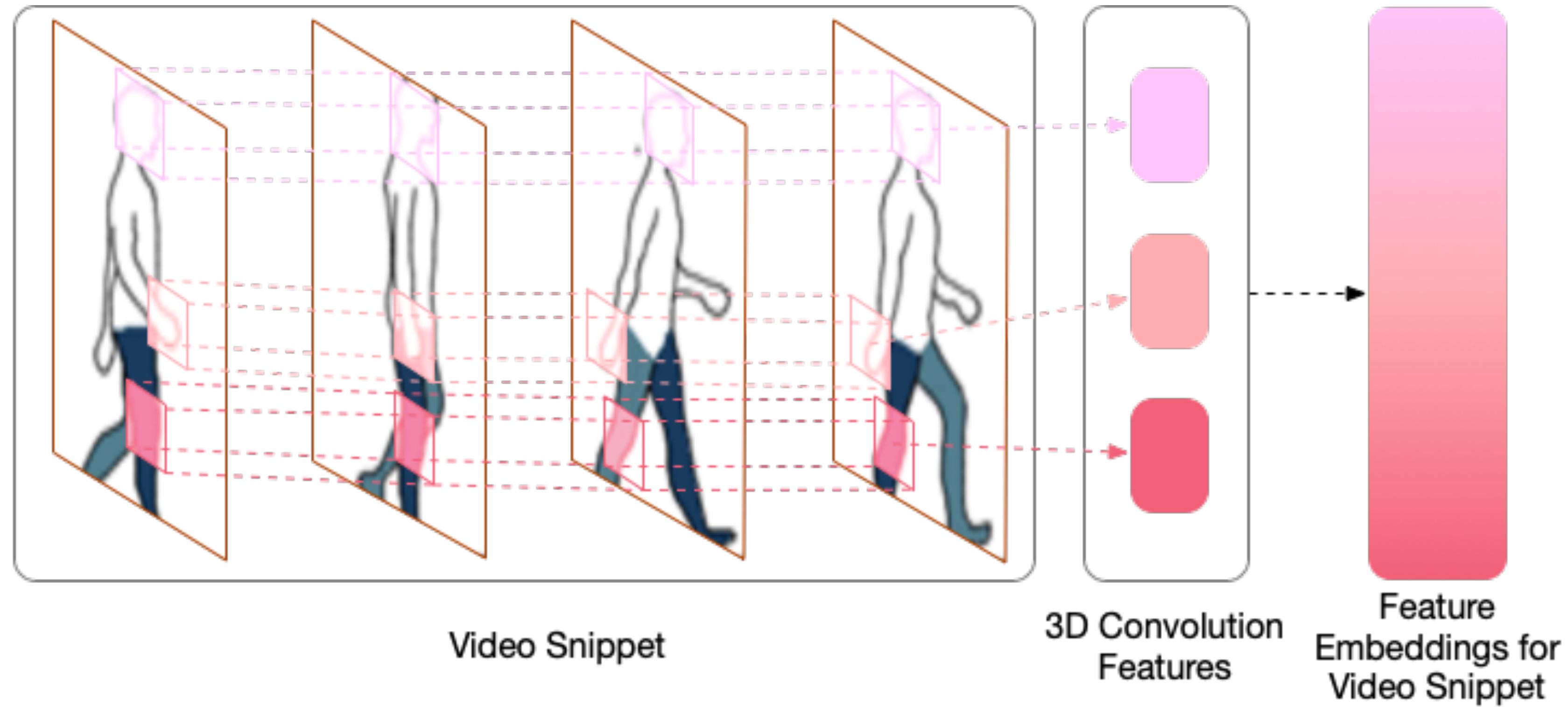
视频ReID

人体关键部位



- 人体部位信息齐全：坐标不同，分布范围广
- 检测人体框不完全准，遮挡，姿态估计也不完全准

解决思路：对齐不同的人体部位



- SSN(self-separated network): 对齐在不同帧（时间）和位置（空间）的人体部位
- 3D卷积：对不同部位分别卷积，提取特征

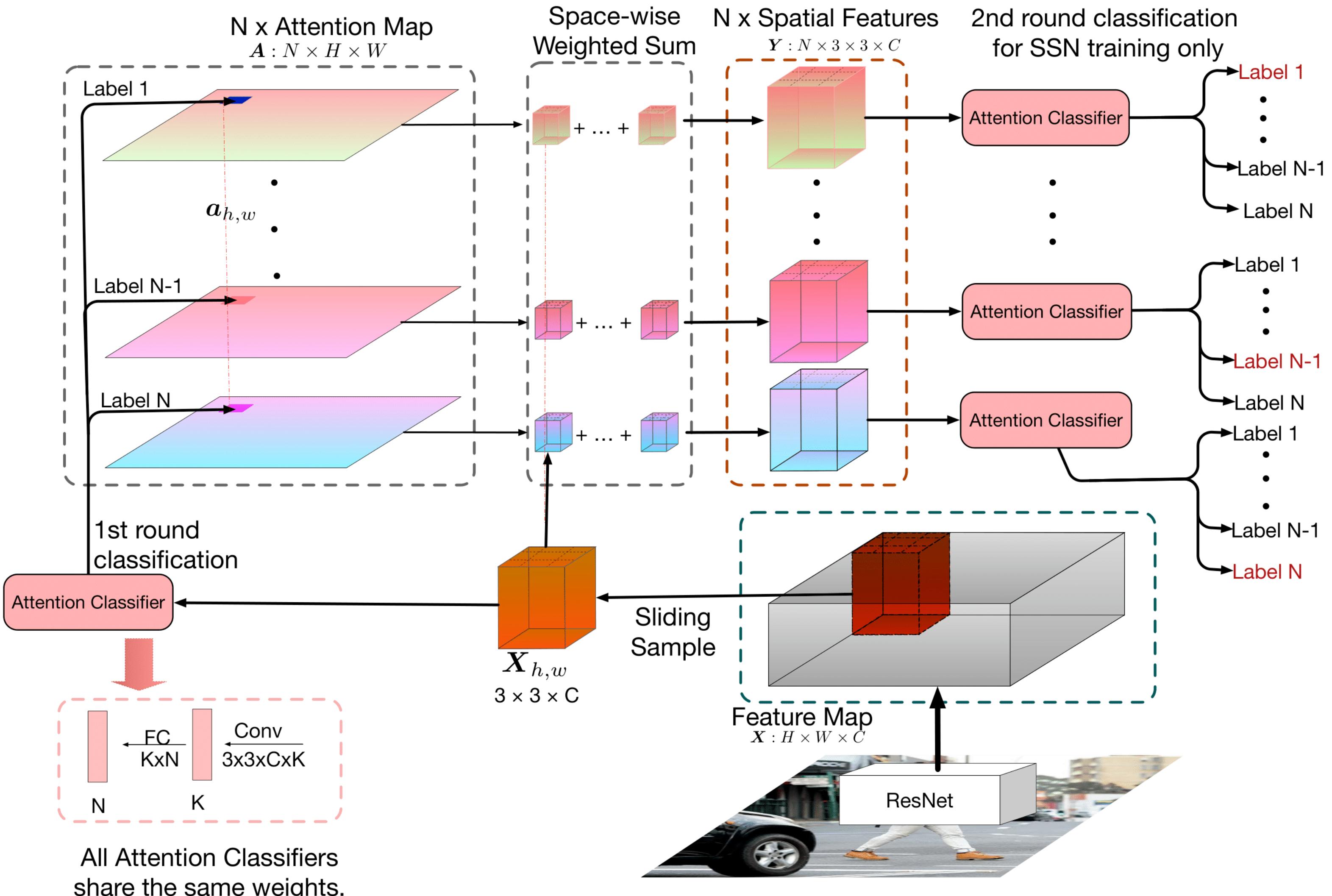
解决思路：对齐不同的人体部位

Attention Classifier

- 关注局部特征，不知道局部在哪里：不需要全局attention，不需要全局的特征
- 若干个关键部位，每个部位应该有有一个attention
- Classifier: 对每一个pixel进行分类
 - 每个分类N个概率输出
 - 对全图分类之后形成attention mask

SSN设计

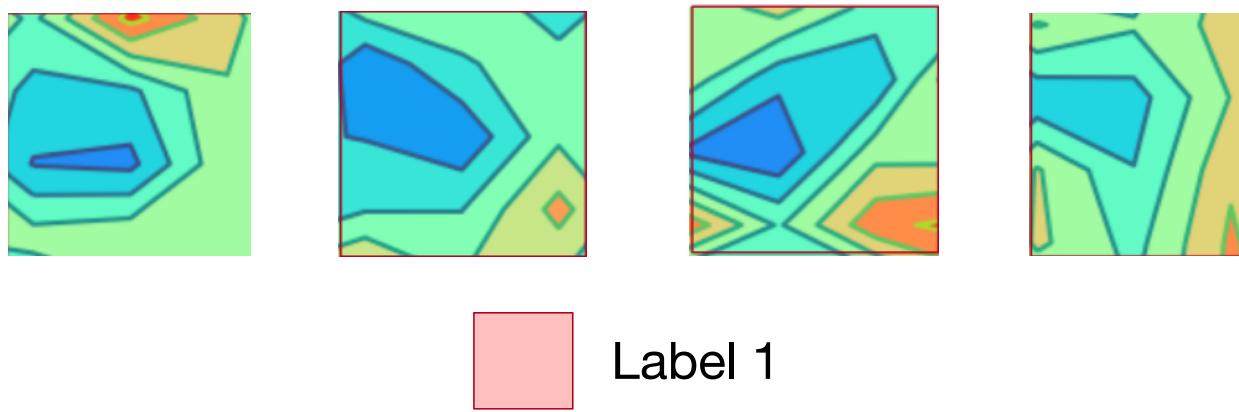
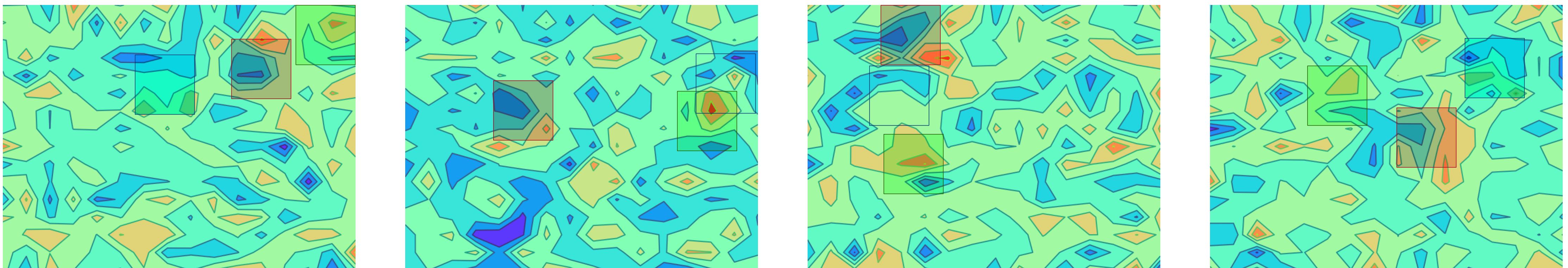
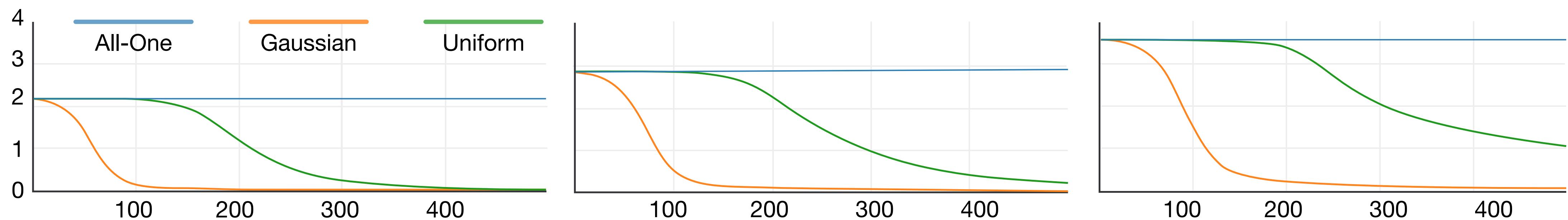
AttentionClassifier 机制 + 两轮分类



- **AttentionClassifier**
 - 全局搜索
 - 局部特征
- **两轮分类：**
 - 两轮分类结果必须是一致的，提供了一致性信号，这样可以采用无监督的方式来训练网络

SSN设计

无监督训练



Label 1 Label 2 Label 3

SSN设计

无监督训练

- 不稳定
- 不准确
- 缺乏目标性



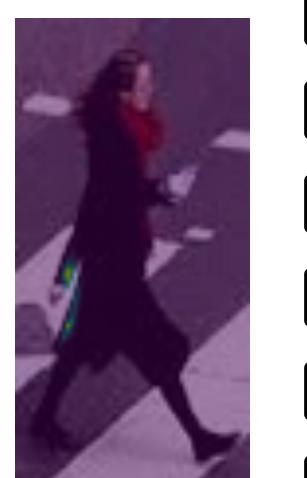
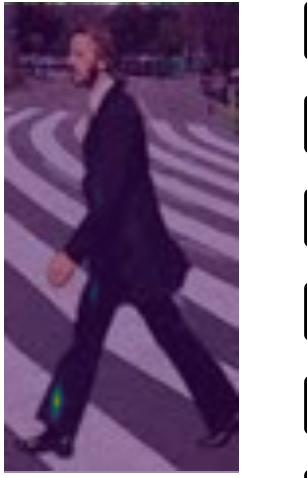
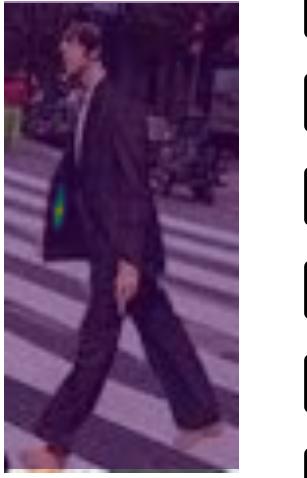
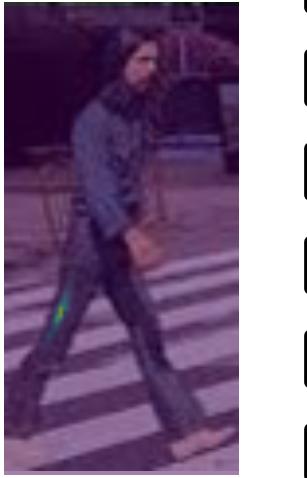
Point 0

Point 1

Point 2

Point 3

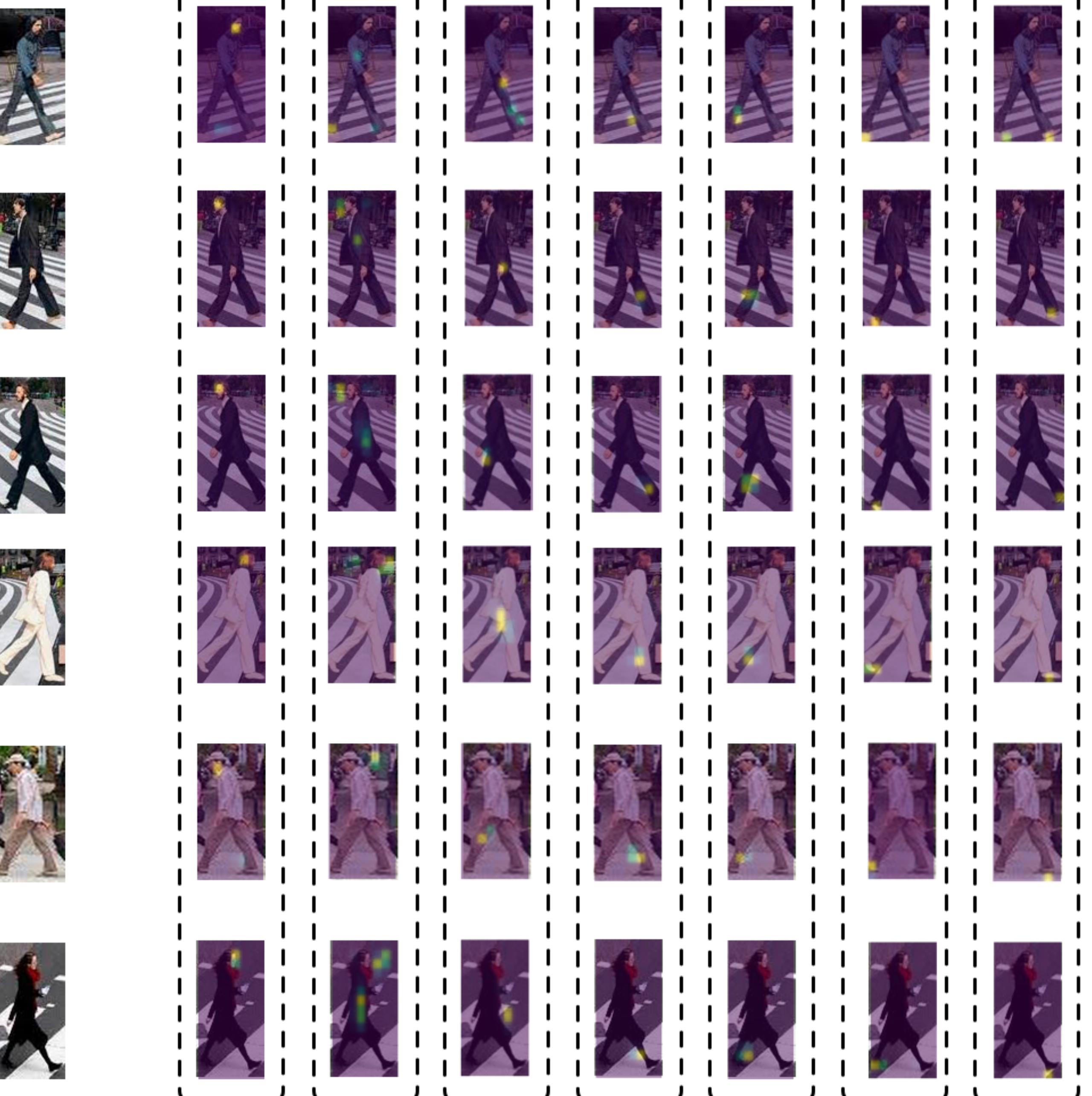
Point 4



SSN设计

有监督训练 & 半监督训练

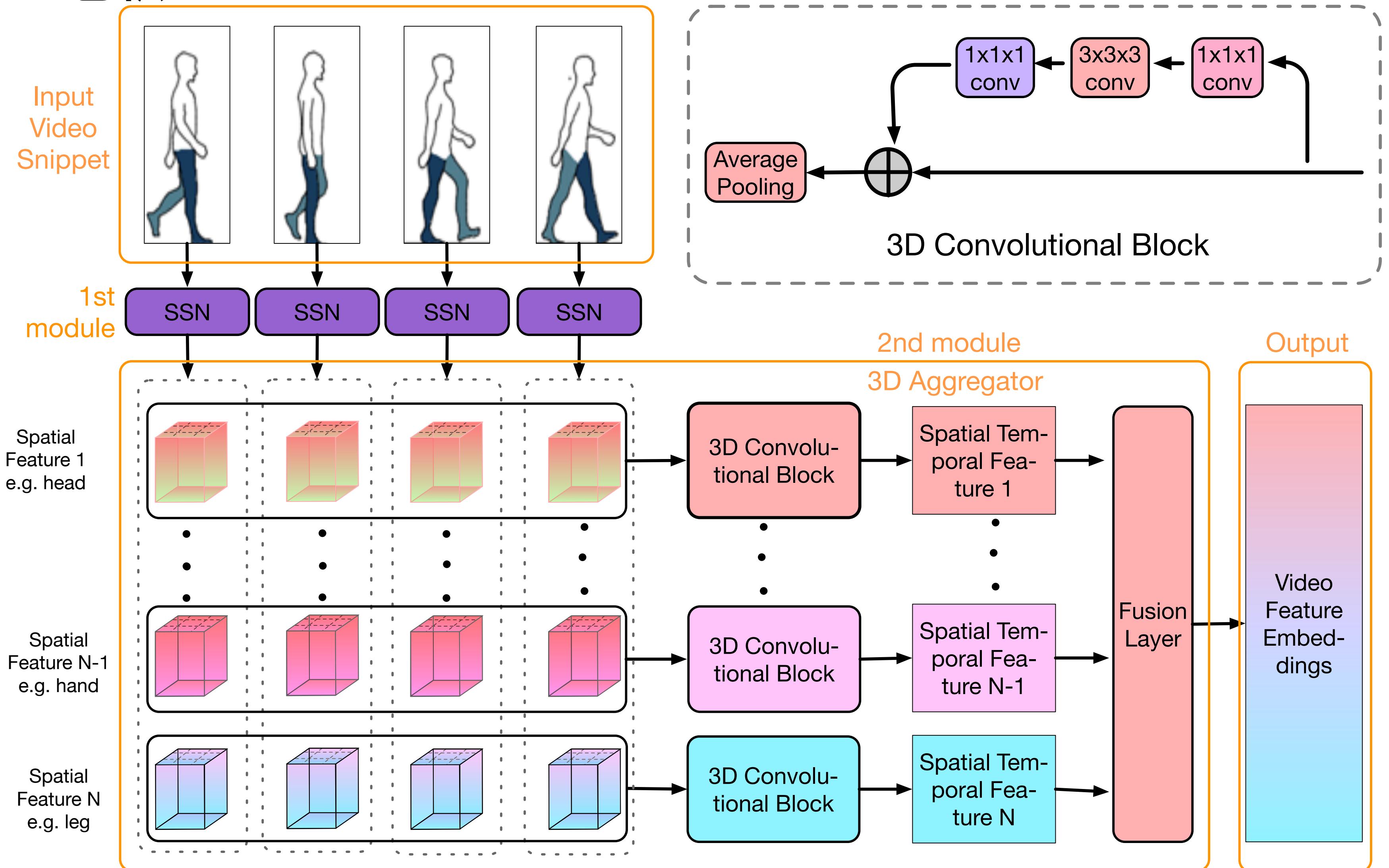
- 稳定
- 准确



3D 卷积的设计

每一个部位做一个3D卷积

- 3D卷积权重
不共享
- FC做融合



训练

难样本在线学习

$$\mathcal{L}_{tri} = \sum_{i=1}^B [m + \max_{f_p \in S_i^+} \frac{\|f_i - f_p\|_2}{\sqrt{d}} - \min_{f_n \in S_i^-} \frac{\|f_i - f_n\|_2}{\sqrt{d}}]_+$$

$$\mathcal{L}_{SSN} = - \sum_{i=1}^N \log \left(\frac{\exp(p_{i,i})}{\sum_{j=1}^N \exp(p_{i,j})} \right)$$

$$\mathcal{L} = \mathcal{L}_{tri} + \lambda \cdot \mathcal{L}_{SSN}$$

3D 卷积的设计

每一个部位做一个3D卷积

- 无监督好于有监督，半监督好于无监督

Learning Strategy	iLIDS-VID		MARS		DukeMTMC-Video	
	top-1	mAP	top-1	mAP	top-1	mAP
Supervised Learning	73.4	75.8	69.8	61.1	86.3	79.6
Unsupervised Learning	83.1	84.2	82.4	67.5	89.9	86.2
Semi-Supervised Learning	88.9	89.2	90.1	86.2	96.8	96.3

- SSN分类器不共享权重高于共享权重20个点

Attention Classifiers	iLIDS-VID		MARS		DukeMTMC-Video	
	top-1	mAP	top-1	mAP	top-1	mAP
NonSharing Weights	69.4	73.2	72.3	70.9	84.9	71.2
Sharing Weights	88.9	89.2	90.1	86.2	96.8	96.3

与现有方法的比较

SOTA

Methods	top-1	top-5	top-10
LFDA ^[49]	32.9	68.5	82.2
KISSME ^[50]	36.5	67.8	78.8
LADF ^[51]	39.0	76.8	89.0
STF3D ^[52]	44.3	71.7	83.7
TDL ^[53]	56.3	87.6	95.6
MARS ^[27]	53.0	81.4	-
SeeForest ^[54]	55.2	86.5	91.0
CNN+RNN ^[55]	58.0	84.0	91.0
Seq-Decision ^[56]	60.2	84.7	91.7
ASTPN ^[57]	62.0	86.0	94.0
QAN ^[58]	68.0	86.8	95.4
RQEN ^[59]	77.1	93.2	97.7
STAN ^[24]	80.2	-	-
Snippet ^[60]	79.8	91.8	-
Snippet+OF ^[60]	85.4	96.7	98.8
VRSTC ^[28]	83.4	95.5	97.7
AP3D ^[23]	86.7	-	-
SSN3D	88.9	97.3	98.8

表 3.3 iLIDS-VID 数据集比较

Methods	top-1	top-5	top-10	mAP
Mars ^[27]	68.3	82.6	89.4	49.3
SeeForest ^[54]	70.6	90.0	97.6	50.7
Seq-Decision ^[56]	71.2	85.7	91.8	-
Latent Parts ^[61]	71.8	86.6	93.0	56.1
QAN ^[58]	73.7	84.9	91.6	51.7
K-reciprocal ^[62]	73.9	-	-	68.5
RQEN ^[59]	77.8	88.8	94.3	71.7
TriNet ^[14]	79.8	91.3	-	67.7
EUG ^[63]	80.8	92.1	96.1	67.4
STAN ^[24]	82.3	-	-	65.8
Snippet ^[60]	81.2	92.1	-	69.4
Snippet+OF ^[60]	86.3	94.7	98.2	76.1
VRSTC ^[28]	88.5	96.5	97.4	82.3
AP3D ^[23]	90.1	-	-	85.1
SSN3D	90.1	96.6	98.0	86.2

表 3.4 MARS 数据集比较

Methods	top-1	top-5	top-10	mAP
EUG ^[63]	83.6	94.6	97.6	78.3
VRSTC ^[28]	95.0	99.1	99.4	93.5
AP3D ^[23]	96.3	-	-	95.6
SSN3D	96.8	98.6	99.4	96.3

表 3.5 DukeMTMC-Video 行人重识别数据集比较

- 达到SOTA水平

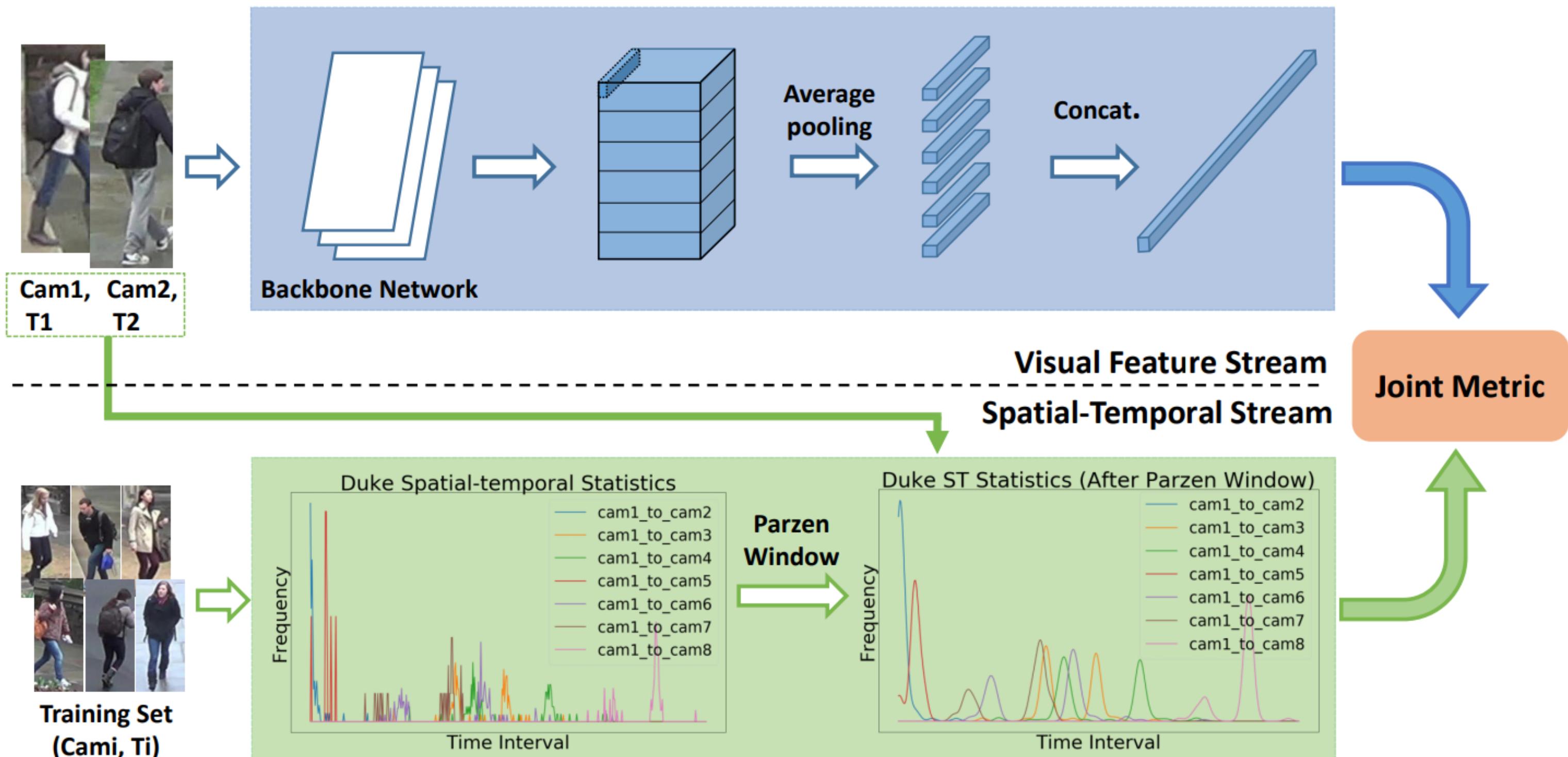
视频ReID小结

- AttentionClassifier机制：全局搜索，局部特征
- 两轮分类的方法：一致性信号可供无监督学习
- SSN+3D的pipeline结合：分别利用了attention机制选择关注点，3D卷积处理含时间维度信息的优势
- 不足：
 - 计算开销大，每一个部位就需要增加一个3D卷积模块
 - 帧间信息用来区分背景&前景，没有用上
 - 除了人类主动选择之外的人体关键点：部分引导+部分自动推导

基于时空行为模式的优化

running system角度：系统运行期间，积累大量历史数据

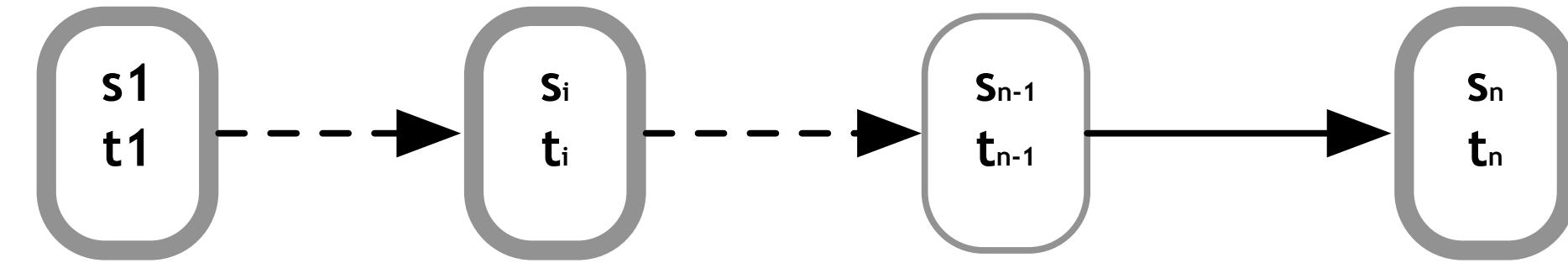
$$h' = \arg \max_h = J(V(f, G_h), P(h, s, t))$$



现有工作: Spatial-Temporal Person Re-identification, AAAI2019

方法改进

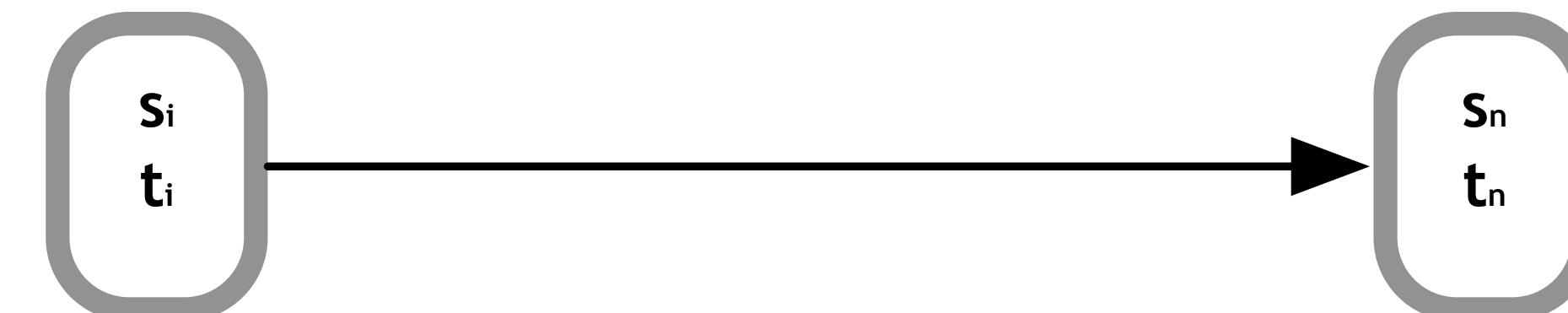
- 真实场景：
 - 历史记录+依靠人脸聚类，人脸人体关联，很多人体获得了特征
 - 增量聚类、实时增量聚类
- 数学模型基础：
 - 有序状态，假设具备马尔科夫性质，当前状态只与上一个状态有关
 - 多个中间状态，但不给定给定中间状态，当前状态只跟上一个状态有关。
 - 方向性：给定每个站点初始概率，方向可互换



{ s_i }为空间位置序列

{ t_i }为时间序列

$$p(n|i) = \sum_{j \in J} P(j|i) \cdot p(n|j)$$



同样适用于人脸识别的场景

简化成车站里的进站、出站两个状态

基于时空行为模式的优化

同样适用于人脸识别系统

- 地铁车站乘车来描述
描述问题
- 某人出现的概率
- 从A到B转移的限制

$$P(h, i, t) = p_t(h|i) = \frac{p_t(h, i)}{p_t(i)}$$

进站

$$P(h, o, t) = p_t(h|o) \cdot p_t(\Delta t^h | i^h, o, h)$$

出站

- 忽略人个体区别，忽略时刻区别

$$p_t(\Delta t^h | i^h, o, h) \approx p_t(\Delta t^h | i^h, o) \approx p(\Delta t^h | i^h, o)$$

近似

以车站乘车作为例子

$$p_t(h|i) = \frac{1}{Z} \sum_r p_t(h|i) \cdot K(r - t), r \in [1, R],$$

这里的 $K(\cdot)$ 是高斯核，即

$$K(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-x^2}{\sigma^2}},$$

Z 是：

$$Z = \sum_t p_t(h|i), t \in [1, R].$$

平滑

$$f(x; \lambda, \gamma) = \frac{1}{1 + \lambda e^{-\gamma x}}.$$

其中 λ 和 γ 是常数。 λ 是平滑因子，而 γ 是收缩因子。

对于进站情况， J 为

$$J(V, P) = f(v, \lambda_1, \gamma_1) \cdot f(p, \lambda_2, \gamma_2).$$

对于出站模型， J 为

$$J(V, P) = f(v, \lambda_1, \gamma_1) \cdot f(p, \lambda_2, \gamma_2) \cdot f(p, \lambda_3, \gamma_3).$$

上面的方法也用于^[5]中，其中定义了联合度量函数。

联合

以车站乘车作为例子

变量名	含义
N	被选择并应用概率模型的候选数
R_1 for $p_t(h s)$	时间段数。定时点也要分割箱
L_1 for $p_t(h s)$	parzen window 长度，它表示可以计算的邻域时间段总数，以进行平滑处理
σ_1 for $p_t(h s)$	Parzen window 平滑系数
R_2 for $P(\Delta t i, o, t)$	时间间隔的数量
R_3 for $P(\Delta t i, o, t)$	成本时间间隔区间的数量
ΔT_{max} for $P(\Delta t i, o, t)$	的最长间隔，如果时间成本长于此，则忽略它
L_2 for $p_t(h s)$	Parzen 窗口的时间长度
L_3 for $p_t(h s)$	成本时间维度的 parzen 窗口长度
σ_2 for $P(\Delta t i, o, t)$	Parzen 窗口平滑系数
$\lambda_1, \lambda_2, \lambda_3, \gamma_1, \gamma_2, \gamma_3$	联合概率的系数

表 4.1 概率优化方法参数表

$$L_{joint} = J(V(q, G_h), P(h, s, t)) - J(V(q, G_n), P(n, s, t))$$

为了摆脱过度拟合，我们添加了一个 L_2 参数归一化损失，最终损失为

$$L_{total} = L_{model} + L_2(\sigma_1, \sigma_2, \lambda_1, \lambda_2, \lambda_3, \gamma_1, \gamma_2, \gamma_3)。$$

我们的目标是找到一组使损失函数最小的参数，即

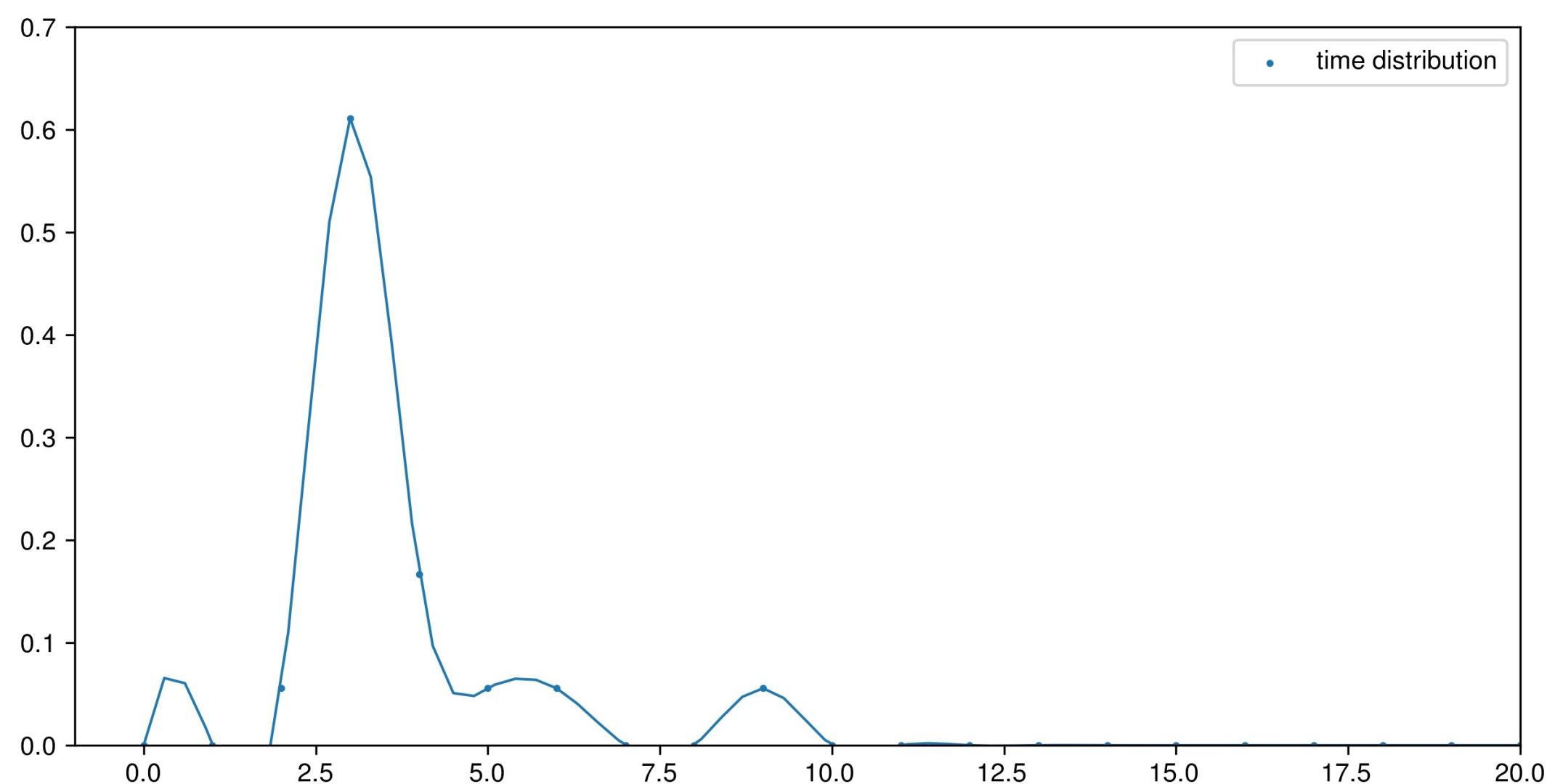
$$\arg \max_{\sigma_1, \sigma_2, \lambda_1, \lambda_2, \lambda_3, \gamma_1, \gamma_2, \gamma_3} (L_{total})。$$

参数表

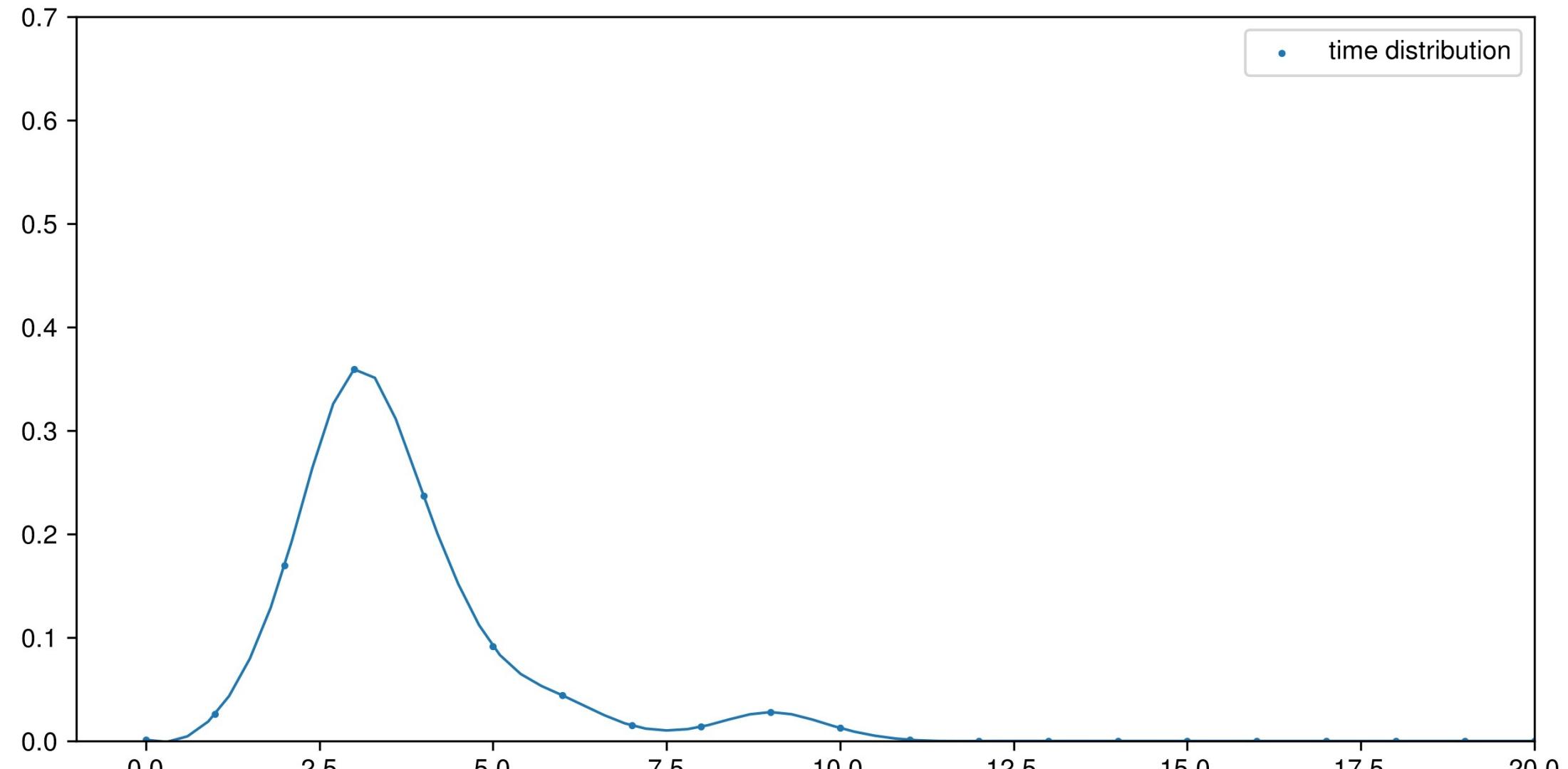
损失和求解

实验结果：平滑效果

对某一位乘客在不同时间段出现在某个车站概率的做平滑



平滑前

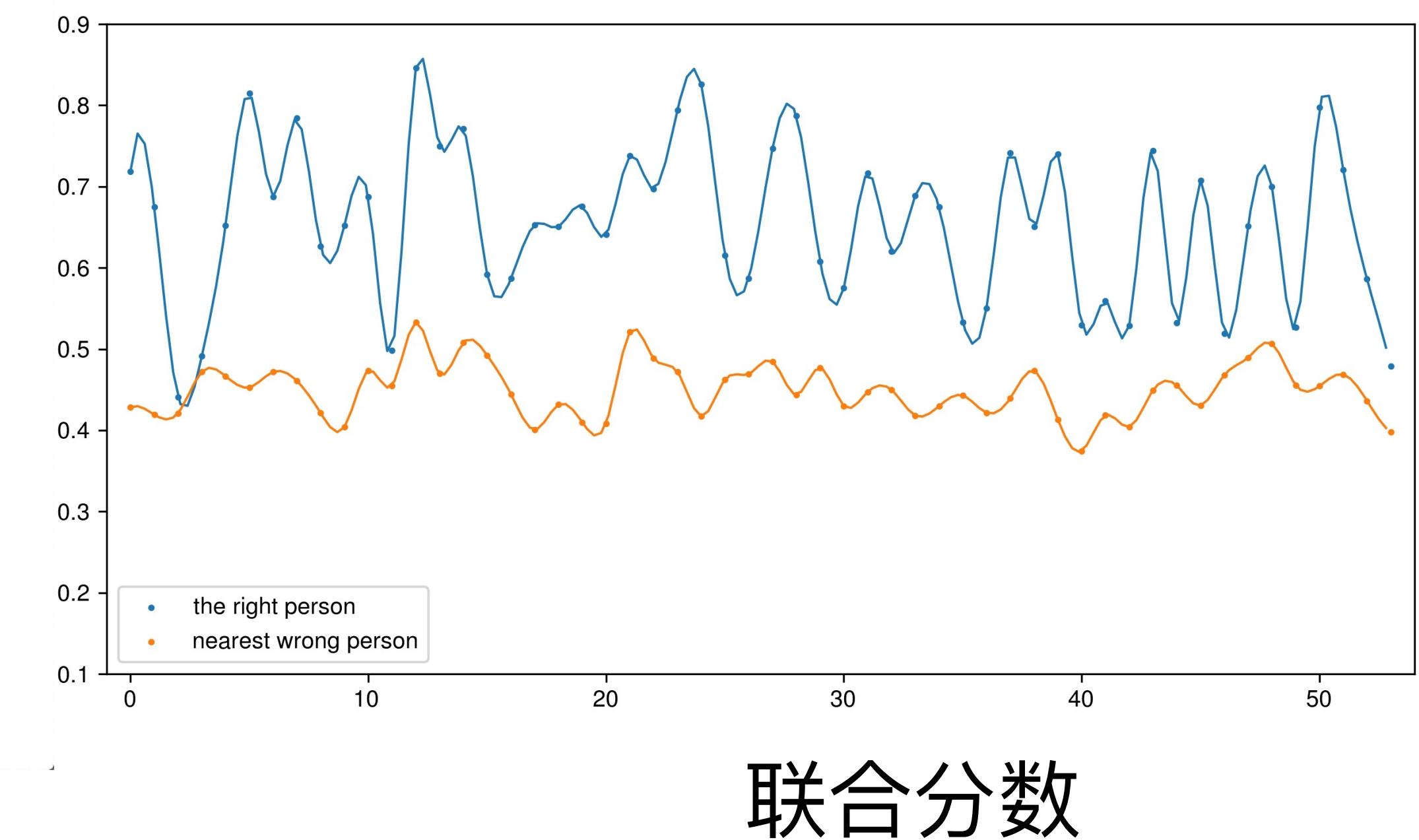
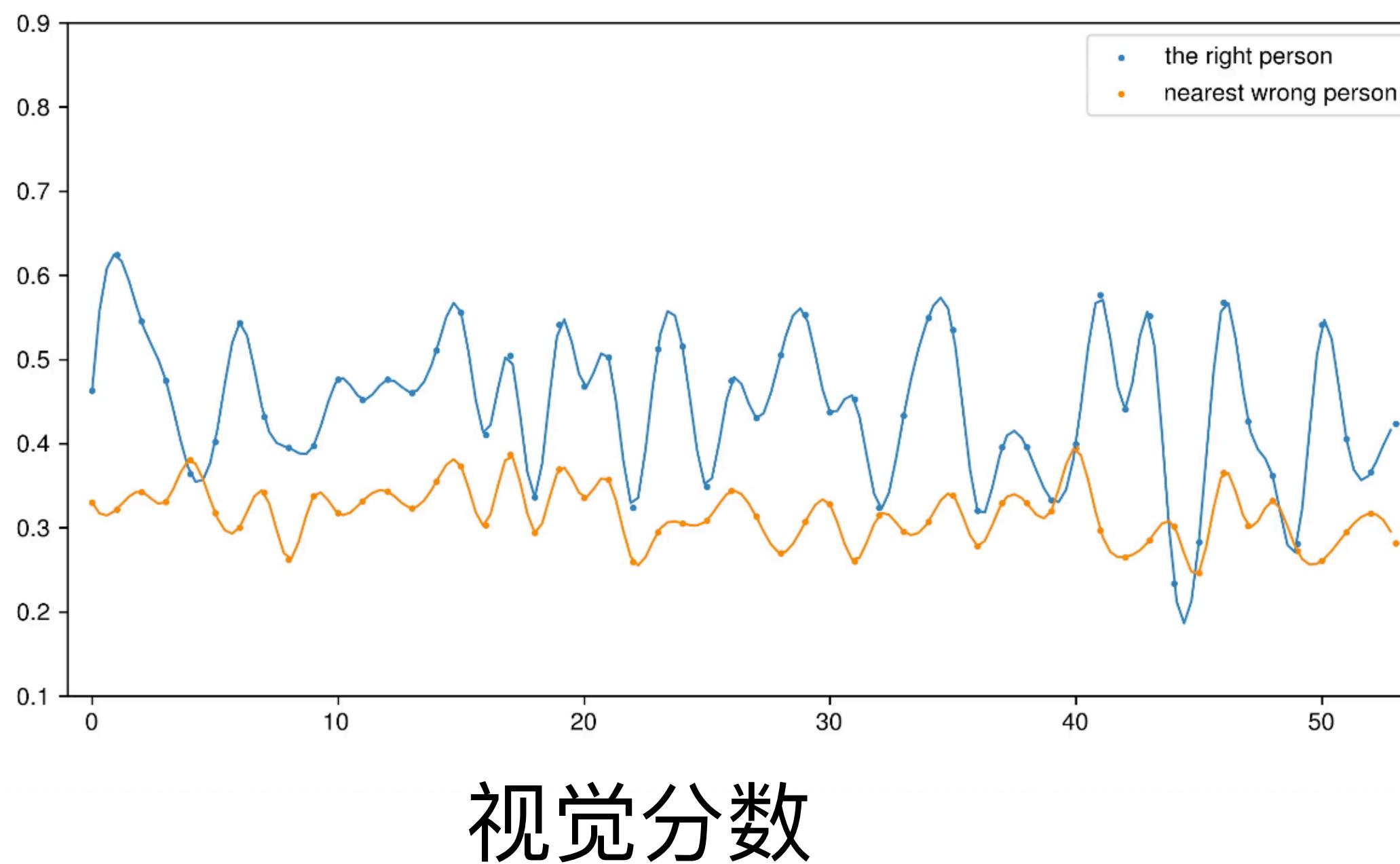


平滑后

实验结果：时空信息辅助

减少33%错误, 平均差从0.13提升到0.21 (+62%)

x轴:人ID, 蓝色线: probe与gt相似度, 黄色线: probe与视觉最近邻相似度



行为模式优化小结

- 基于马尔科夫性质建模、近似、求解、应用
- 与现有工作相比，更加准确和更一般性的模型；参数通过最优化求解
- 不足：产品应用
 - 研究不完整，由于缺乏行人ReID相关的数据，我们把它应用在地铁刷脸的场景，以此来做验证
 - 更一般情况下，关于上一个位置也是根据算法推算出来的，其正确性也是概率的，如何处理？

研究总结

- 基于现场视频时空信息的ReID优化
 - AttentionClassifier & 两轮分类：全局搜索，局部特征，半监督
- 基于长期行为时空信息的ReID优化
 - 基于连续事件，更一般的问题建模，近似，求解和应用
- 两个方向分别反映了博后期间的学习两个不同阶段：
 - DL&数学领域，设计&模型都可以推导
 - 工作计划，投稿计划（中期答辩）顺利执行

学术成果

- 论文
 - Xiaoke Jiang, Yu Qiao, Junjie Yan, Qichen Li, Wanrong Zheng, Dapeng Chen, SSN3D: Self-Separated Network to Align Parts for 3D Convolution in Video Person Re-Identification, *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(2), 1691-1699. Retrieved from <https://ojs.aaai.org/index.php/AAAI/article/view/16262>
- 专利
 - 蒋小可、丁思杰、鲍纪奎、季聪, 一种基于人脸识别与人体关联的地铁逃票监测系统, 申请号: 202011529962.8, 申请日: 2020.12.22
 - 郑莞蓉、蒋小可、鲍纪奎、李启琛、季聪, 基于历史客流大数据的轨道交通人脸识别乘车解决方案, 申请号: 202011132611.3, 申请日: 2020.10.21
 - 孙哲、郑莞蓉、蒋小可、姚兴华、季聪, 一种隔栏递包行为检测解决方案, 申请号: 202110129505.8, 申请日: 2021.1.29
 - 蒋小可、邱达明、马睿、马志友, 一种全景视频动态切流后的画面质量评估方法及装置, CN108833976B, 申请日: 2018.6.27, 授权日: 2020.1.24
 - 蒋小可、马睿、马志友, 全景视频的画面质量显示方法及装置, 公开号: CN108810513B, 申请日: 2018.6.27, 授权公告日: 2020.3.13
 - 蒋小可、王伦、蒋捷, 全景画面生成方法及装置, 公开号: CN10727474B, 申请日: 2017.6.30, 授权日: 2019.6.25
 - 交底书: 郑莞蓉、蒋小可, 基于乘客历史乘车习惯的轨道交通人识别乘车解决方案

**谢谢！
请各位评委老师批评指正**