



# 面向非对称数据场景的 行人重识别关键技术研究

北京大学 贾梦溪

# 提纲

---

## □ 绪论

- 研究背景与意义
- 研究问题与难点
- 本文研究内容

## □ 相关研究工作

## □ 研究内容

- 语义模式集合隐式匹配的局部-完整行人重识别方法
- 半注意力部件关联构建的遮挡-完全行人重识别方法
- 基于跨模态相似性推理的红外-可见光行人重识别方法
- 基于多层次模态对齐的文本-图像行人重识别方法

# 提纲

---

## □ 绪论

- 研究背景与意义
- 研究问题与难点
- 本文研究内容

## □ 相关研究工作

## □ 研究内容

- 语义模式集合隐式匹配的局部-完整行人重识别方法
- 半注意力部件关联构建的遮挡-完全行人重识别方法
- 基于跨模态相似性推理的红外-可见光行人重识别方法
- 基于多层次模态对齐的文本-图像行人重识别方法

# 1.1 研究背景与意义

- 智慧安防产业迅猛发展，视频监控智能化是当前发展趋势

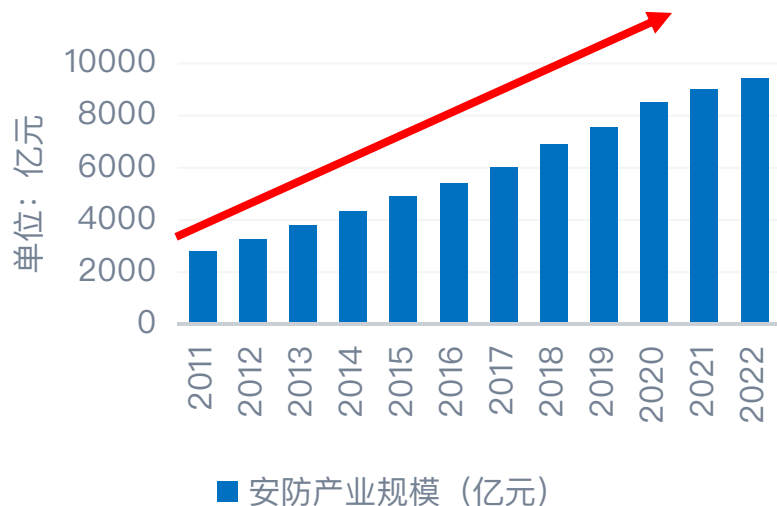


图. 2011-2022年中国安防产业规模（数据来自中国安全防范产品行业协会）

- 2011年到2022年，中国**安防产业规模从2800亿元上升到9421亿元**
- 2023年上半年全国千万级“雪亮工程”及社会治安视频监控项目共46个，**总金额17亿元**

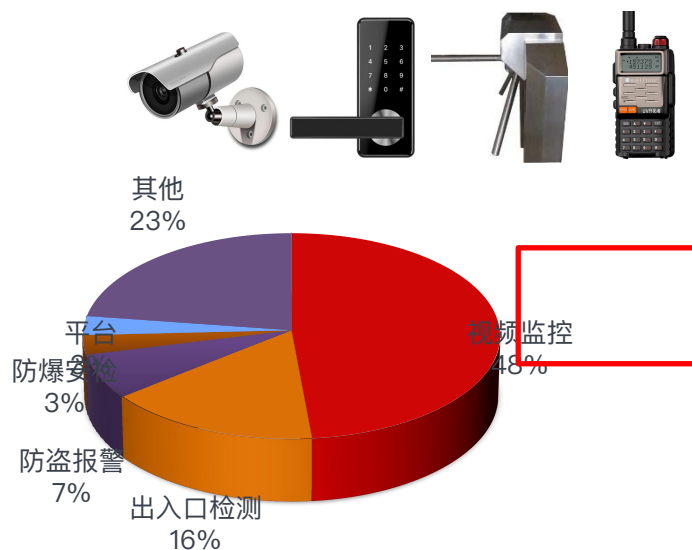


图. 2021年中国智能安防产业结构（数据来自华经产业研究院）

- 智慧安防是AI+安防结合的产物。
- 视频监控产品占整个智能安防产品的市场比重接近50%**

# 1.1 研究背景与意义

- 行人重识别技术
  - 针对视频监控的主要目标：**行人**
  - 自动、快速、精准返回目标行人跨摄像头运动轨迹
- 应用场景：智能安防、智慧社区、公共卫生等



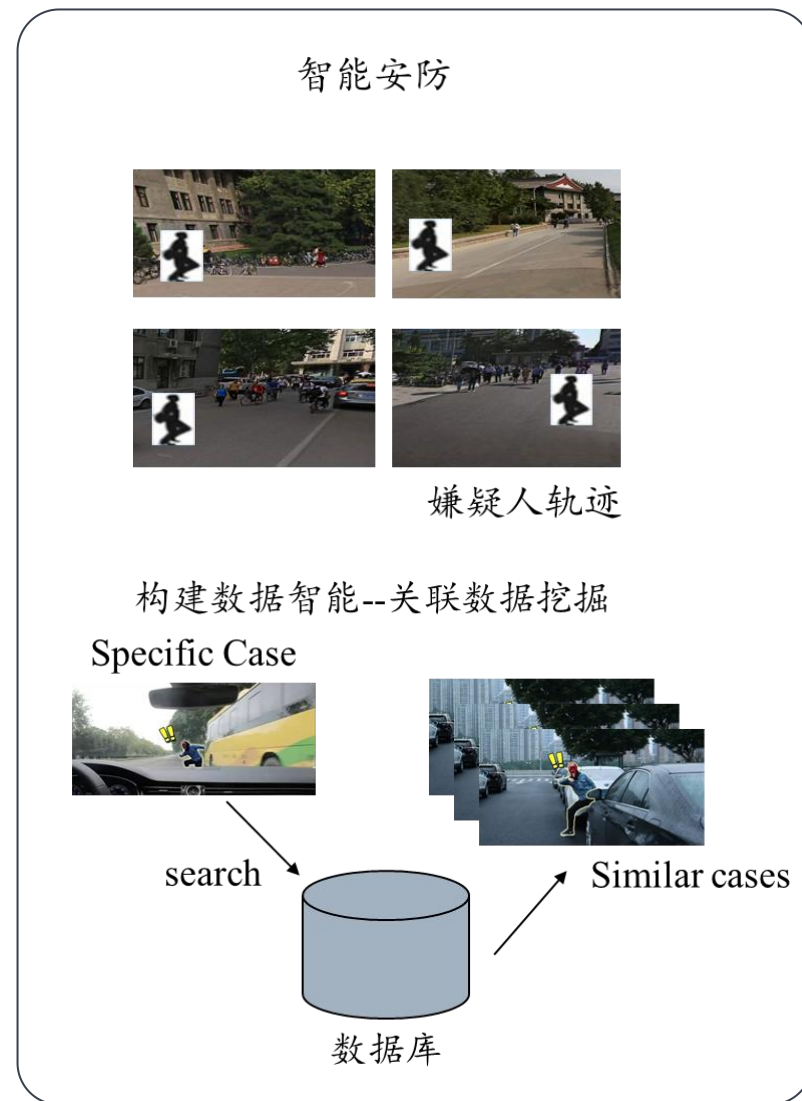
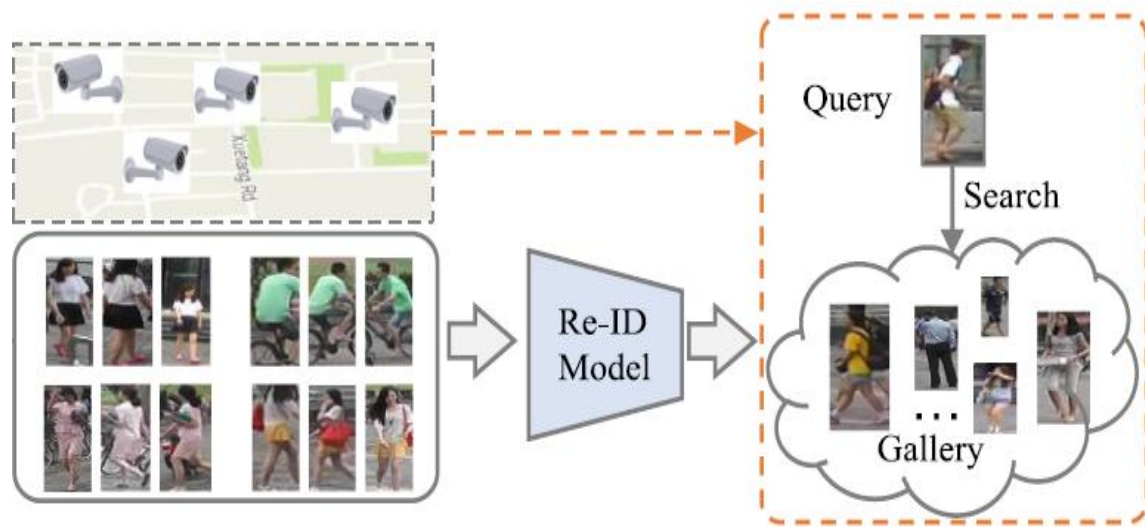
图. 海量数据实时产生



图. 行人重识别典型应用

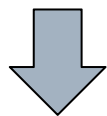
## 1.2 研究问题与难点

- 行人重识别 (Re-ID)
  - 图像检索的子问题
  - 目标表征+度量匹配
- 应用场景：智能安防、数据挖掘



## 1.2 研究问题与难点

- 当前行人重识别高度依赖**完整可见**的查询图像
  - 实际应用中，由于现实开放场景的复杂性以及监控覆盖的局限性，行人信息**来源多样化**，且**高质量完整可见**行人图像难以获取。



行人重识别面临更为复杂的  
**多源非对称数据**的挑战

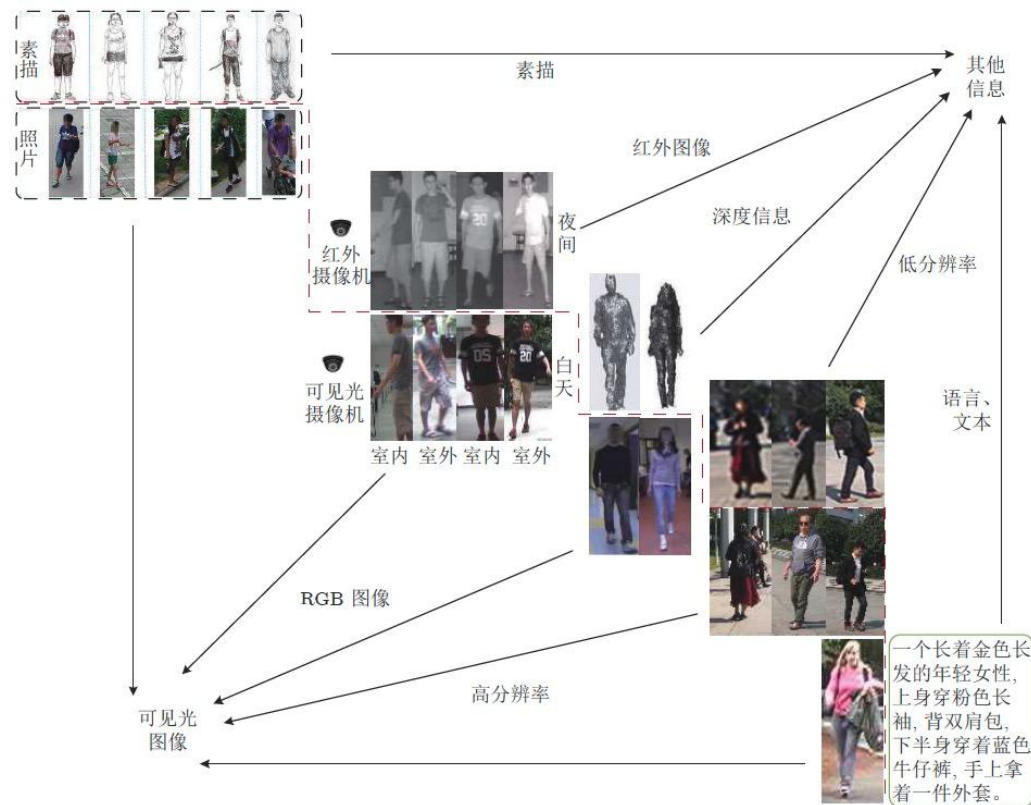


图. 海量多源非对称数据用于行人重识别



# 1.2 研究问题与难点

□ 根据数据源的差异程度，非对称数据场景可以划分成：



图. 四种非对称数据行人重识别典型场景



## 1.2 研究问题与难点

- 行人重识别基础框架通常包含两个子模型
  - **表示学习子模型**：用于表征提取
  - **度量学习子模型**：用于样本相似度量

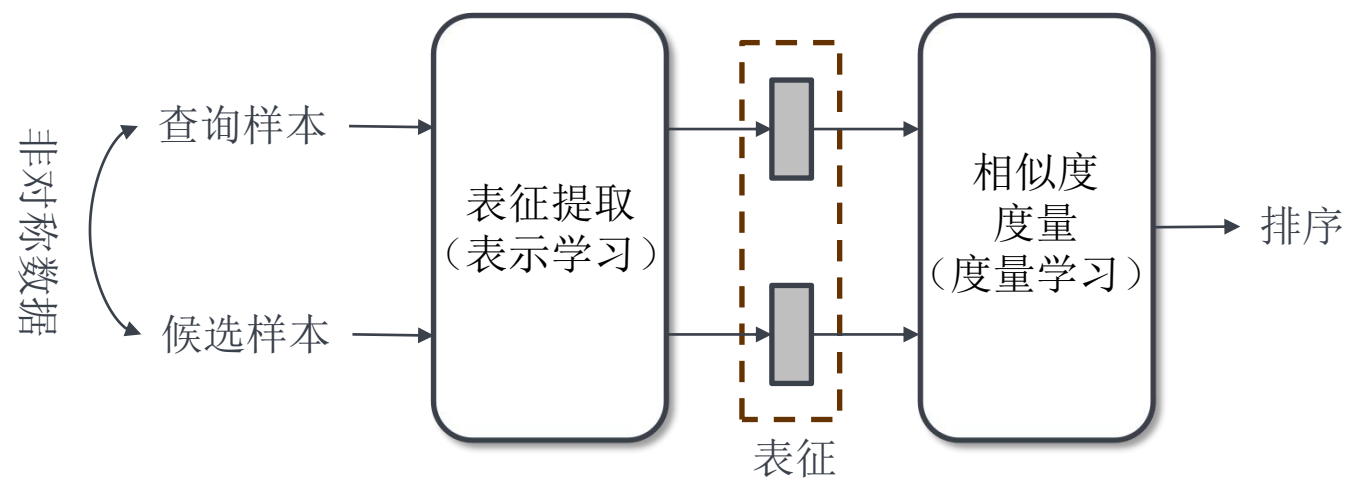


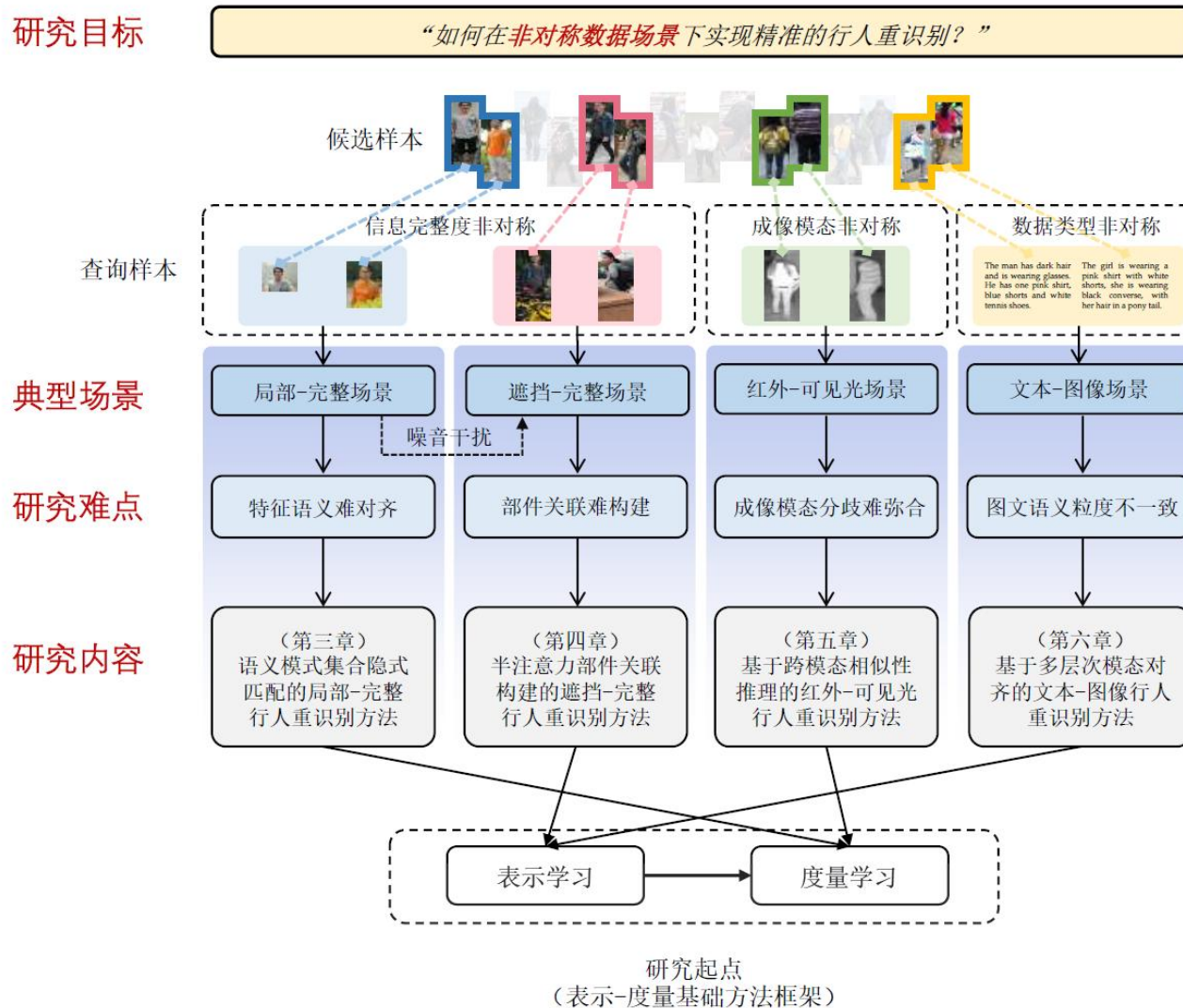
图. 行人重识别基础框架

# 1.3 本文研究内容

- 一个研究目标
  - 如何在非对称数据下实现精准的行人重识别

- 四个研究难点
  - 特征语义难对齐
  - 部件关联难构建
  - 模态分歧难弥合
  - 语义粒度不一致

- 四个研究内容



# 提纲

---

## □ 绪论

- 研究背景与意义
- 研究问题与难点
- 本文研究内容

## □ 相关研究工作

## □ 研究内容

- 语义模式集合隐式匹配的局部-完整行人重识别方法
- 半注意力部件关联构建的遮挡-完全行人重识别方法
- 基于跨模态相似性推理的红外-可见光行人重识别方法
- 基于多层次模态对齐的文本-图像行人重识别方法

## 2. 相关研究工作

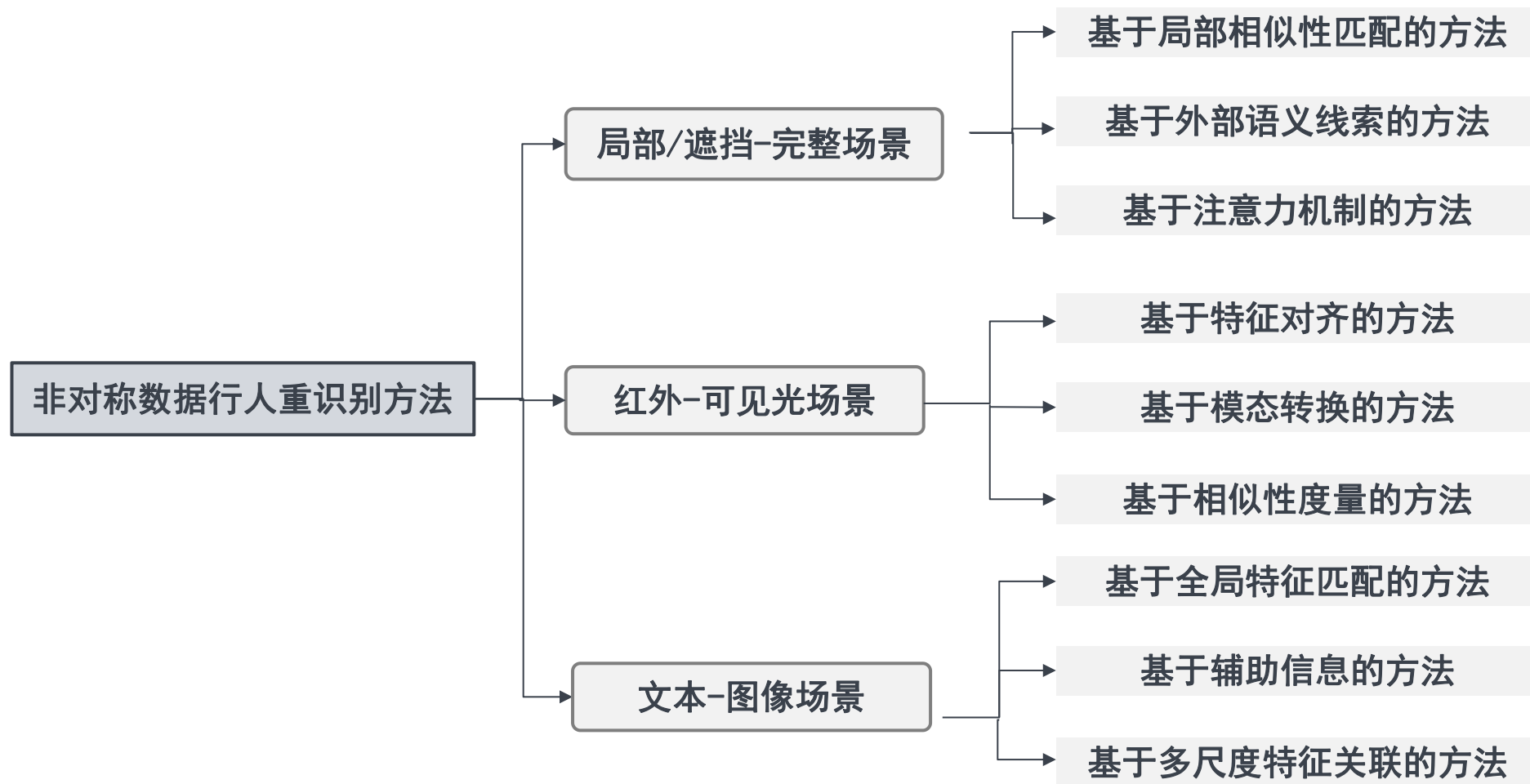
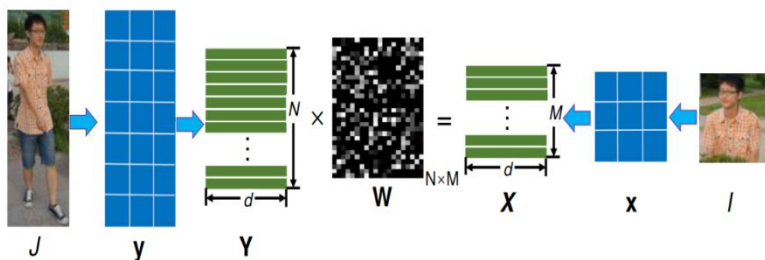


图. 非对称数据行人重识别国内外研究进展

## 2. 相关研究工作

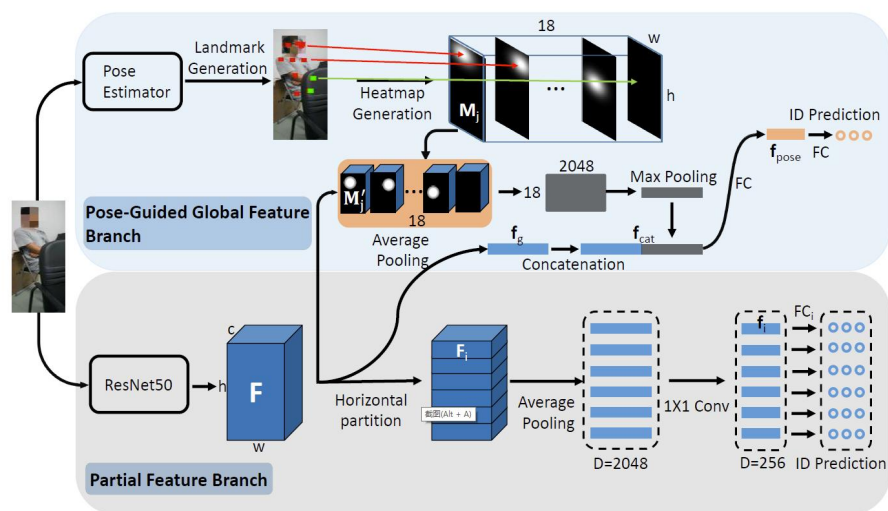
### □ 局部/遮挡-完整场景

基于局部相似性匹配的方法



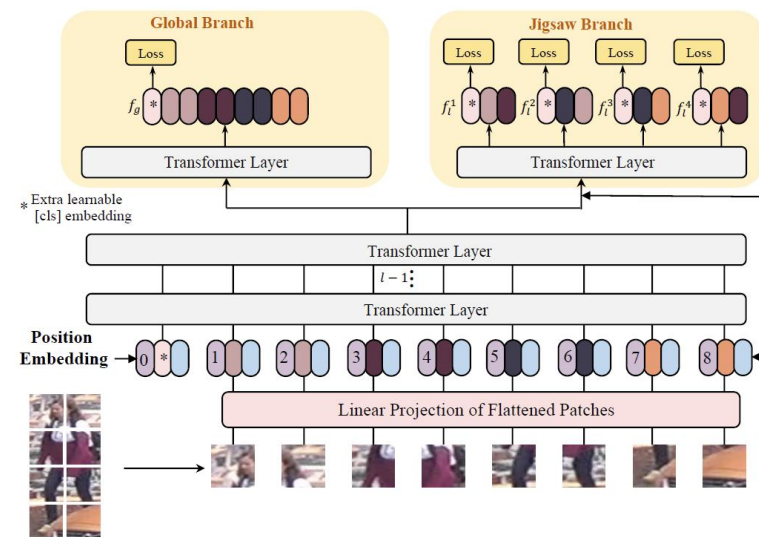
无需监督信息  
大量额外计算开销

基于外部语义线索的方法



先验信息具有可解释性  
领域差异导致推理误差

基于注意力机制的方法

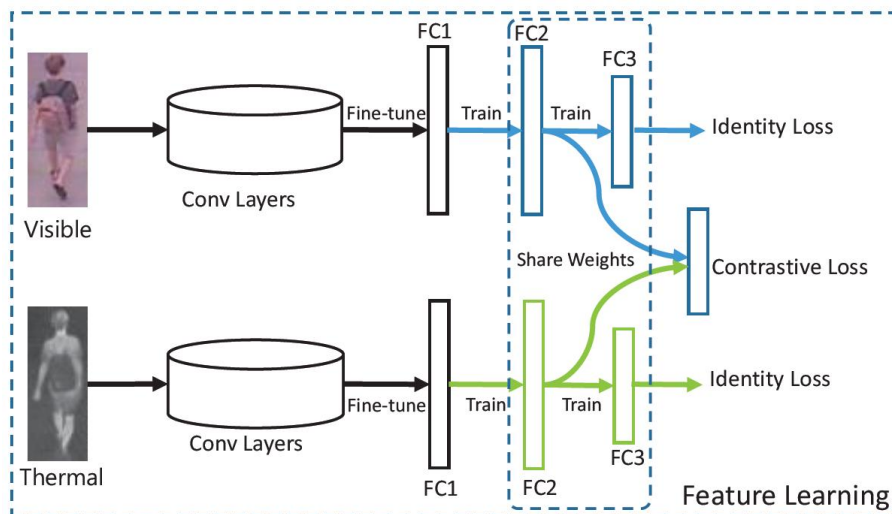


自适应于特定数据  
部件级特征多样性不足

## 2. 相关研究工作

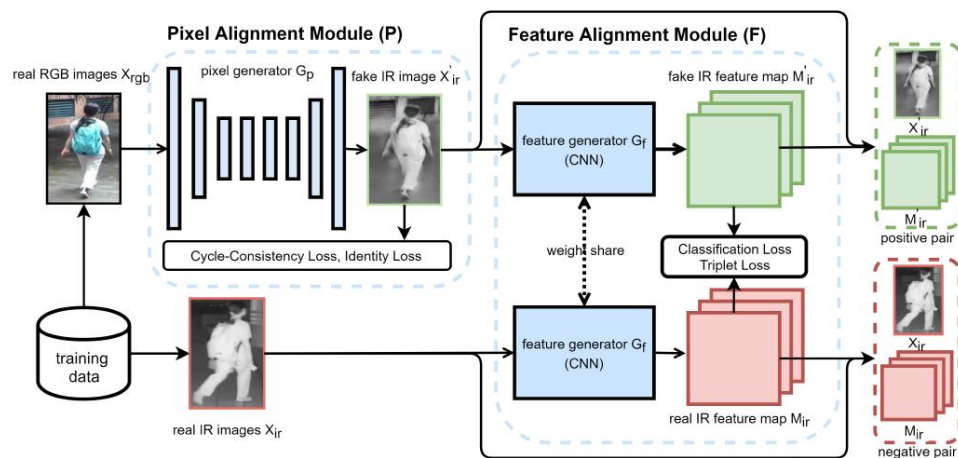
### □ 红外-可见光场景

基于特征对齐的方法



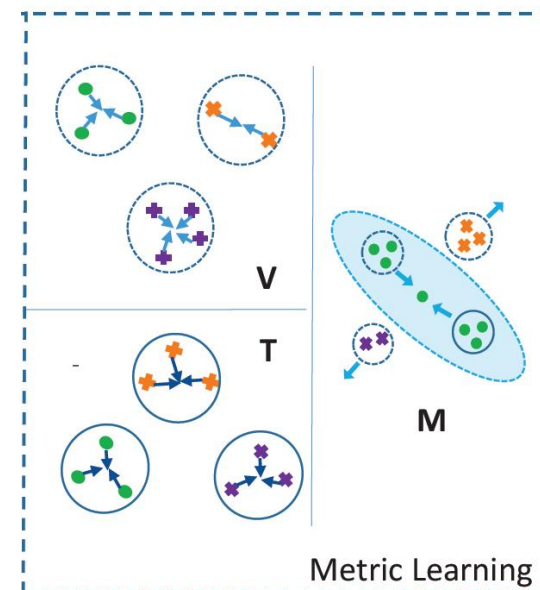
减小模态分布偏差  
未克服固有分歧的影响

基于模态转换的方法



减小像素级模态偏差  
丢失部分判别信息

基于相似性度量的方法



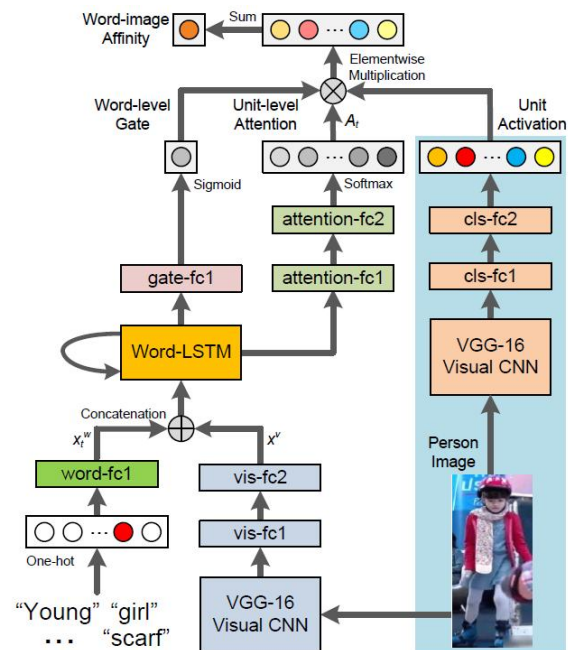
挖掘利用样本分布特性  
计算复杂度较高



# 2. 相关研究工作

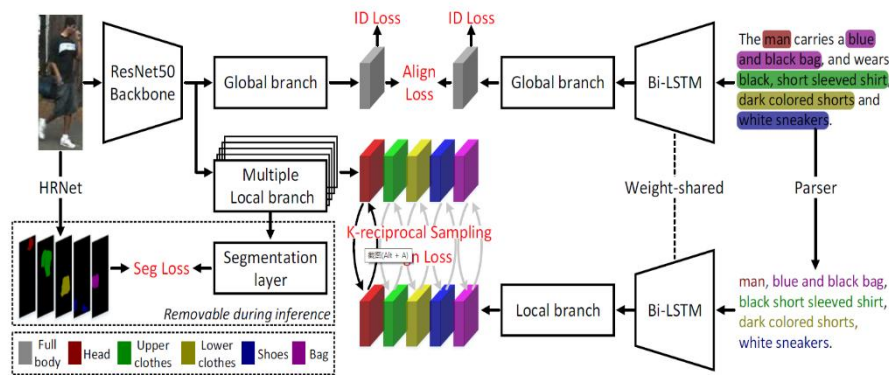
## □ 文本-图像场景

### 基于全局特征匹配的方法



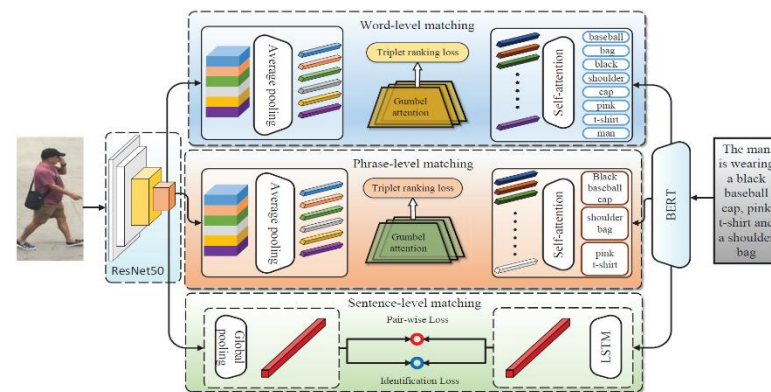
有效利用宏观信息  
忽略细粒度特点

### 基于辅助信息的方法



利用和参考外部信息  
引入误差、效率低

### 基于多尺度特征关联的方法



挖掘利用局部信息  
预处理复杂、受粒度不对齐影响

# 提纲

---

## □ 绪论

- 研究背景与意义
- 研究问题与难点
- 本文研究内容

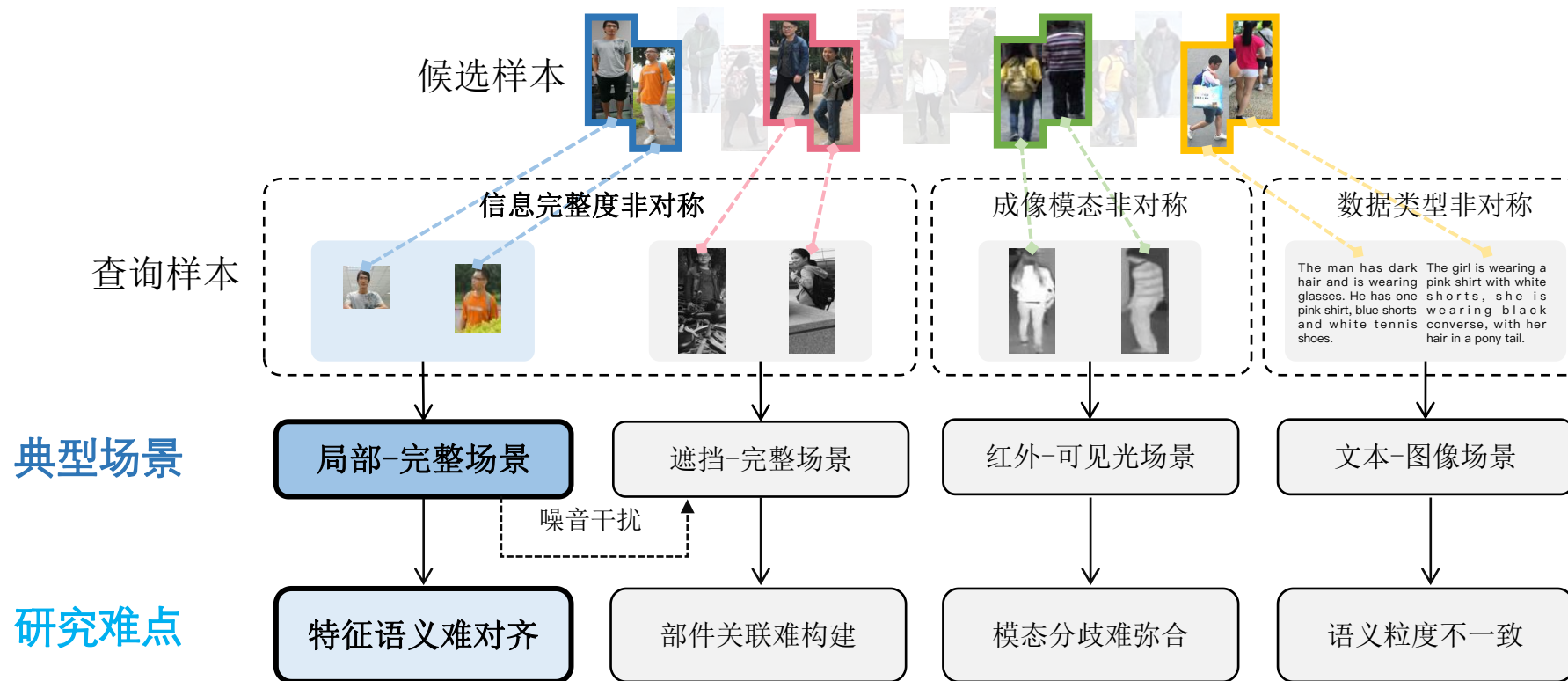
## □ 相关研究工作

## □ 研究内容

- 语义模式集合隐式匹配的局部-完整行人重识别方法
- 半注意力部件关联构建的遮挡-完全行人重识别方法
- 基于跨模态相似性推理的红外-可见光行人重识别方法
- 基于多层次模态对齐的文本-图像行人重识别方法

# 3. 语义模式集合隐式匹配的局部-完整行人重识别方法

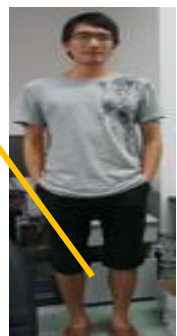
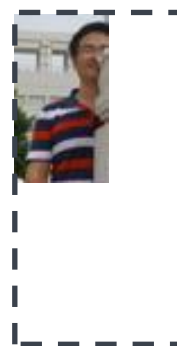
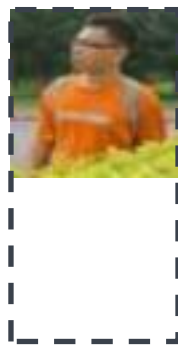
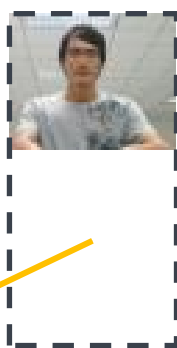
□ 研究内容：探究信息完整度非对称下的局部-完整图像场景



### 3. 语义模式集合隐式匹配的局部-完整行人重识别方法

□ 针对行人外观信息缺失导致的**特征语义难对齐**问题

1. 局部图像缺乏关键的局部外观特征



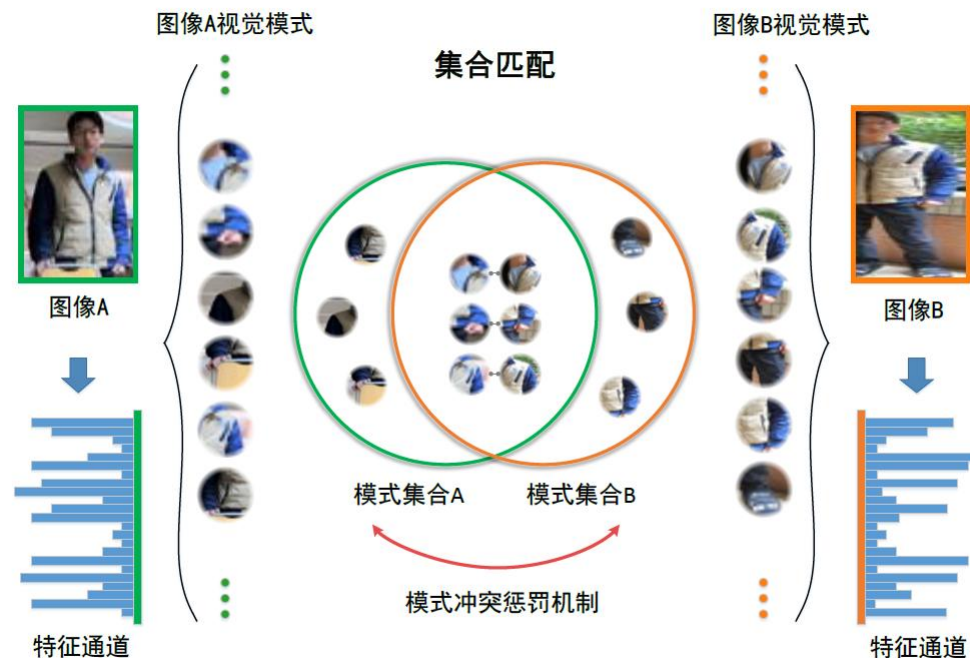
2. 局部图像缺乏轮廓、结构等宏观信息

针对以上难点，引入**集合隐式匹配**思想解决局部-完整场景下的行人度量问题。

### 3. 语义模式集合隐式匹配的局部-完整行人重识别方法

- MoS (Matching on Set) 方法
  - 基于模式集合匹配的行人度量 (创新点1)
    - 特征匹配问题转换为可导化的模式集合匹配
  - 模式冲突惩罚 (创新点2)
    - 挖掘冲突模式对, 进一步修正集合相似性

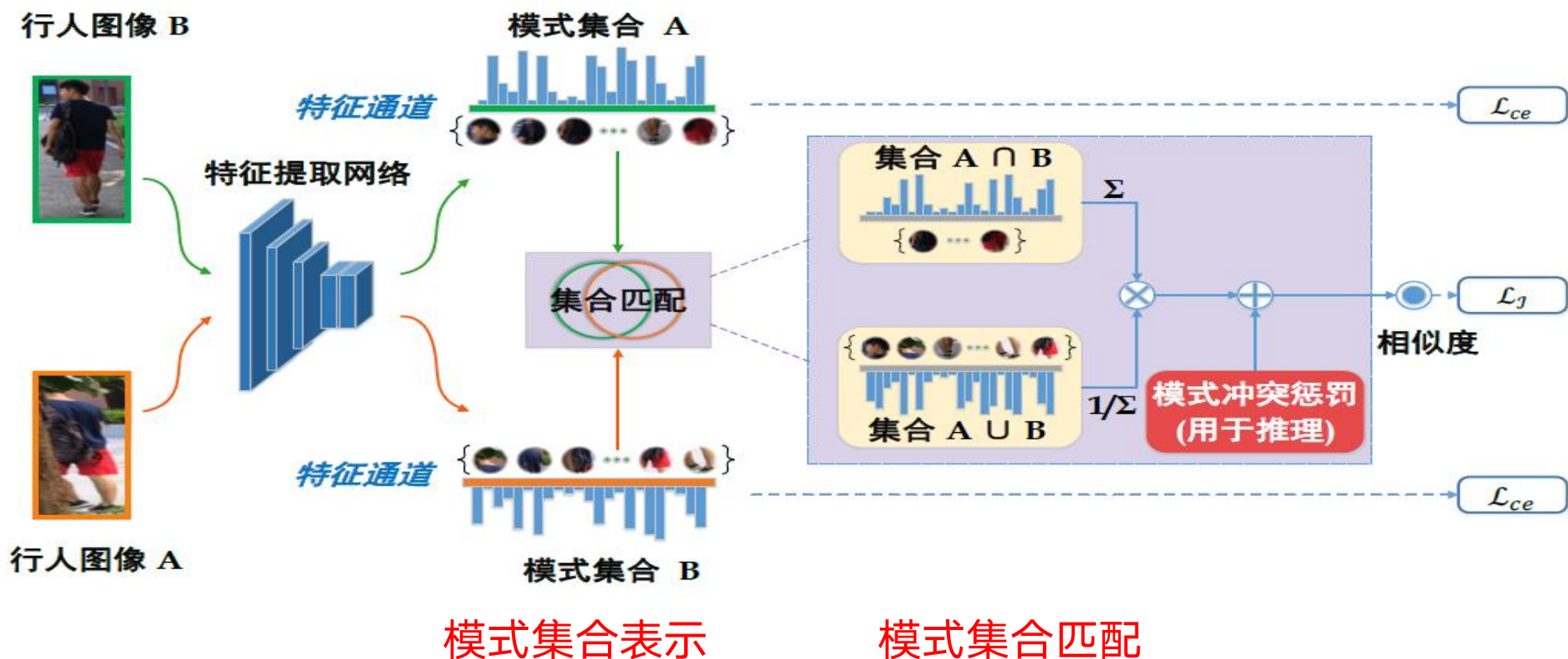
特征通道看作是一个模式,  
样本表示为模式的集合



降低部件匹配复杂度  
 $O(M \times N) \rightarrow O(M + N)$

### 3. 语义模式集合隐式匹配的局部-完整行人重识别方法

- 基于模式集合的行人度量框架（创新点1）
  - 行人表示为模式集合，并利用杰卡德距离实现行人集合匹配
  - 杰卡德距离：交集与并集个数之比。集合交集越大、并集越小，集合的相似性越高





### 3. 语义模式集合隐式匹配的局部-完整行人重识别方法

□ 基于模式集合的行人度量框架（创新点1）

■ 杰卡德距离的可导化学习——实现端到端训练

采用Min/Max操作符表示交并集

$$\mathcal{J}(A, B) = \frac{\sum_{c=1}^C A[c] \wedge B[c]}{\sum_{c=1}^C A[c] \vee B[c]}, \quad \mathcal{J}(g_1, g_2) = \frac{\sum_{c=1}^C \min(g_1[c], g_2[c])}{\sum_{c=1}^C \max(g_1[c], g_2[c])}$$

↓ 可导化处理

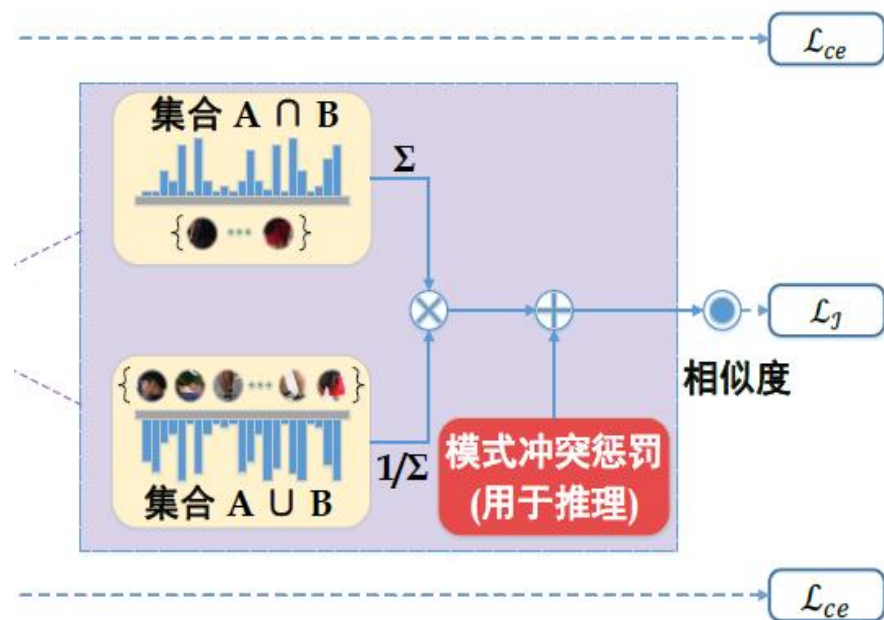
采用softmin/softmax操作符表示交并集

$$\mathcal{J}_s(g_1, g_2) = \frac{\sum_{c=1}^C (w_1^{min}[c] \cdot g_1[c] + w_2^{min}[c] \cdot g_2[c])}{\sum_{c=1}^C (w_1^{max}[c] \cdot g_1[c] + w_2^{max}[c] \cdot g_2[c])}, \quad (5)$$

$$w_k^{min}[c] = \frac{e^{-\tau \cdot g_k[c]}}{\sum_n e^{-\tau \cdot g_n[c]}}, \quad w_k^{max}[c] = \frac{e^{\tau \cdot g_k[c]}}{\sum_n e^{\tau \cdot g_n[c]}}$$

构造杰卡德距离三元组损失

$$\mathcal{L}_J = \sum_{n=1}^N [\delta + \mathcal{D}_J(g_n, g_{n+}) - \mathcal{D}_J(g_n, g_{n-})]_+, \quad \mathcal{D}_J(g_1, g_2) = 1 - \mathcal{J}_s(g_1, g_2),$$



### 3. 语义模式集合隐式匹配的局部-完整行人重识别方法

- 模式冲突惩罚（创新点2）
  - 并集中潜在冲突模式, *e.g.*, 蓝鞋-黑鞋
  - 自动检测冲突模式对, 并减小目标相似性

模式共现矩阵

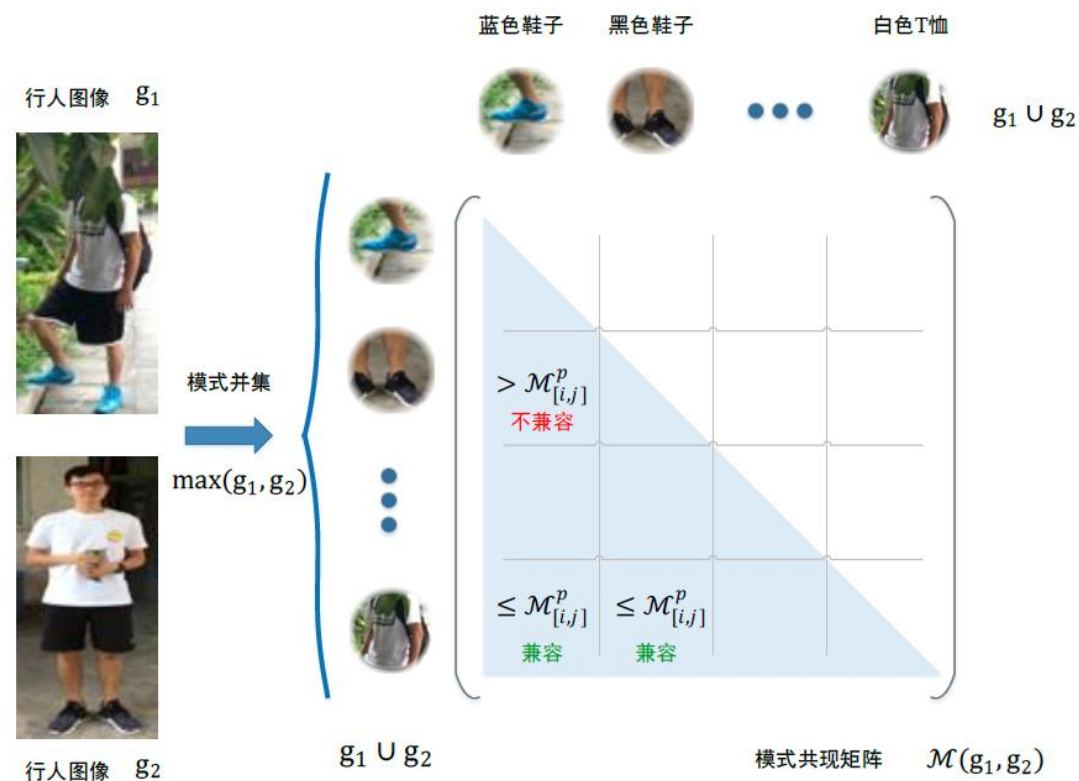
$$\mathcal{M}(\mathbf{g}_1, \mathbf{g}_2) = (\mathbf{g}_1 \cup \mathbf{g}_2)(\mathbf{g}_1 \cup \mathbf{g}_2)^T,$$

模式共现矩阵  $\mathcal{M}_{[i,j]}^p = \max_{t,n} \mathcal{M}_{[i,j]}(\mathbf{g}_t, \mathbf{g}_n),$   
 s. t.  $y_t = y_n,$   
 $t, n = 1, \dots, N.$

修正后相似性

$$\mathcal{S}(\mathbf{x}_1, \mathbf{x}_2) = \mathcal{J}(\mathbf{g}_1, \mathbf{g}_2) - \lambda \mathcal{CP}(\mathbf{g}_1, \mathbf{g}_2),$$

$$\mathcal{CP}(\mathbf{g}_1, \mathbf{g}_2) = \sum_{i,j} \max(0, e^{(\mathcal{M}_{[i,j]}(\mathbf{g}_1, \mathbf{g}_2) - \mathcal{M}_{[i,j]}^p)^{-\varepsilon}} - 1),$$



# 3. 语义模式集合隐式匹配的局部-完整行人重识别方法

## □ 实验证明本方法优于现有先进方法

\*数据集-P-ETHZ

- 85 个行人
- 3,897张局部/完整图像
- 训练测试 1:1
- 指标: Rank-1/mAP

\*数据集Market-1501

\*数据集DukeMTMC-reID

局部匹配方法总体优于  
全局匹配方法

本章方法显著优于现有  
先进方法

方法	Market-1501		DukeMTMC	
	Rank-1	mAP	Rank-1	mAP
IANet <sup>[80]</sup>	94.4	83.1	87.1	73.4
MVPM <sup>[81]</sup>	91.4	80.5	83.4	70.0
DMML <sup>[82]</sup>	93.5	81.6	85.9	73.7
SFT <sup>[83]</sup>	93.4	82.7	86.9	73.2
VCFL <sup>[84]</sup>	89.3	74.5	-	-
Circle <sup>[85]</sup>	94.2	84.9	-	-
PCB <sup>[76]</sup>	92.3	77.4	81.8	66.1
PCB+RPP <sup>[76]</sup>	93.8	81.6	83.3	69.2
AlignedReID <sup>[86]</sup>	91.8	79.3	-	-
DSR <sup>[61]</sup>	83.6	64.3	-	-
VPM <sup>[13]</sup>	93.0	80.8	83.6	72.6
SPReID <sup>[8]</sup>	92.5	81.3	-	-
MGCAM <sup>[59]</sup>	83.8	74.3	46.7	46.0
Pose-transfer <sup>[87]</sup>	87.7	68.9	30.1	28.2
PSE <sup>[88]</sup>	87.7	69.0	27.3	30.2
PGFA <sup>[7]</sup>	91.2	76.8	82.6	65.5
AAANet <sup>[89]</sup>	93.9	82.5	86.4	72.6
HONet <sup>[10]</sup>	94.2	84.9	86.9	75.6
<b>MoS (本章方法)</b>	<b>94.7</b>	<b>86.8</b>	<b>88.7</b>	<b>77.0</b>
DCDS* <sup>[90]</sup>	94.8	85.8	87.5	75.5
ISP* <sup>[5]</sup>	95.3	88.6	89.6	80.0
<b>MoS<sub>wlibn</sub> (本章方法)</b>	<b>95.4</b>	<b>89.0</b>	<b>90.6</b>	<b>80.2</b>

方法	Rank-1	mAP
DIM <sup>[72]</sup>	21.5	14.4
Part Aligned <sup>[73]</sup>	28.8	20.2
HACNN <sup>[74]</sup>	34.4	26.0
Adver Occluded <sup>[75]</sup>	44.5	32.2
PCB <sup>[76]</sup>	42.6	33.7
Part Bilinear <sup>[77]</sup>	36.9	-
FD-GAN <sup>[78]</sup>	40.8	-
PGFA <sup>[7]</sup>	51.4	37.3
HONet <sup>[10]</sup>	55.1	43.8
DSR <sup>[61]</sup>	40.8	30.4
SFR <sup>[79]</sup>	42.3	32.0
<b>MoS (本章方法)</b>	<b>61.0</b>	<b>49.2</b>
ISP* <sup>[5]</sup>	62.8	52.3
<b>MoS<sub>wlibn</sub> (本章方法)</b>	<b>66.6</b>	<b>55.1</b>

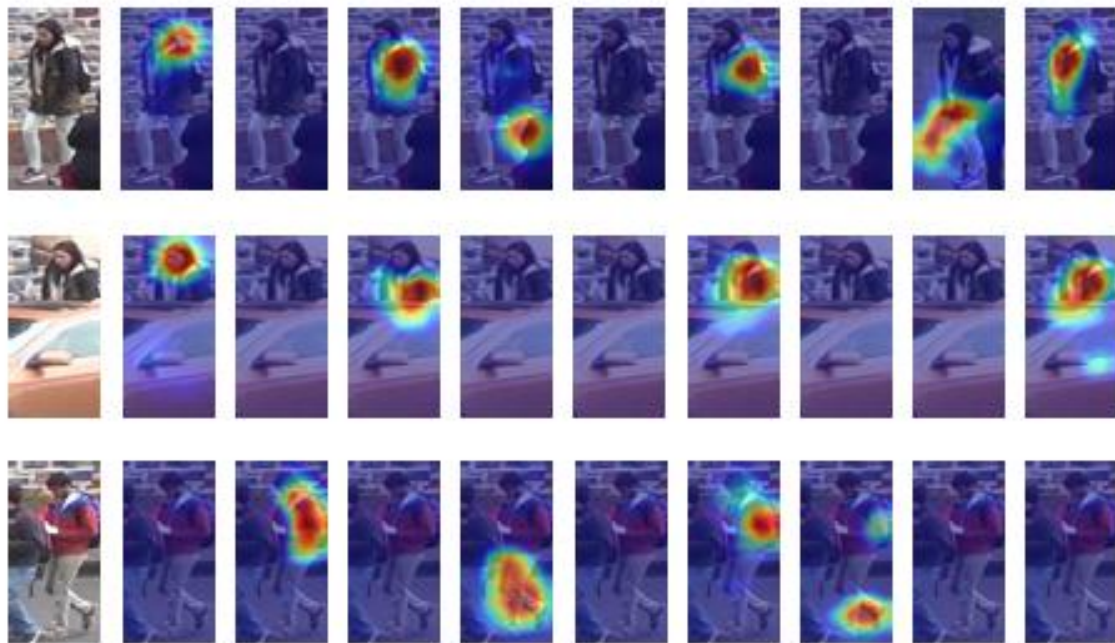
方法	Rank-1	Rank-5	mAP
DGD <sup>[70]</sup>	51.2	81.0	-
SVDNet <sup>[71]</sup>	52.2	79.0	-
REDA <sup>[69]</sup>	54.4	79.1	-
AFPB <sup>[63]</sup>	58.2	84.6	-
<b>MoS (本章方法)</b>	<b>76.2</b>	<b>84.8</b>	<b>64.0</b>
<b>MoS<sub>wlibn</sub> (本章方法)</b>	<b>79.5</b>	<b>85.7</b>	<b>66.8</b>



### 3. 语义模式集合隐式匹配的局部-完整行人重识别方法

#### 可视化分析

- 每行展示一个样本
- 每列表示同一个特征通道的激活位置



本方法各通道可有效表  
征特定的视觉模式

- 左：原图
- 中：传统三元组损失 特征激活图
- 右：杰卡德三元组损失 特征激活图



本章所提出的杰卡德三元组显著  
优于传统三元组损失

# 提纲

---

## □ 绪论

- 研究背景与意义
- 研究问题与难点
- 本文研究内容

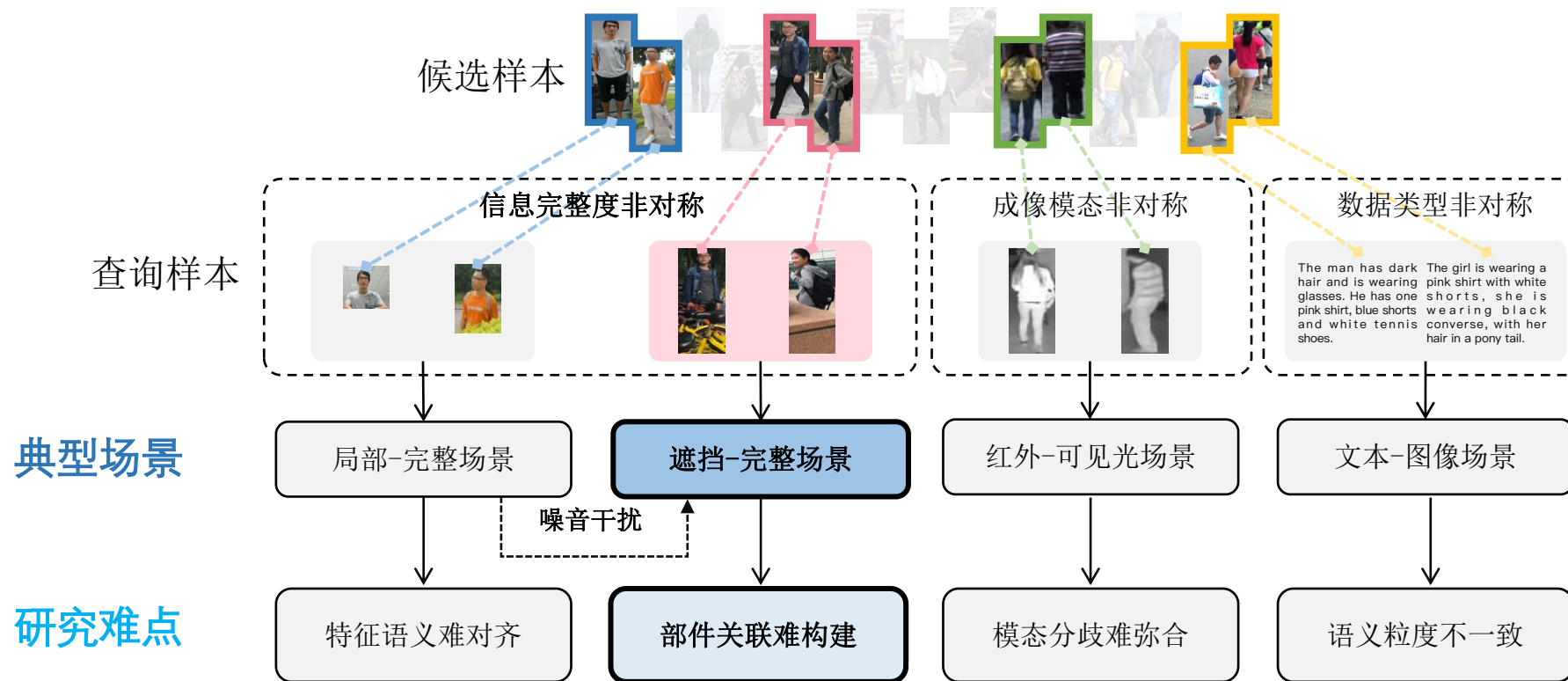
## □ 相关研究工作

## □ 研究内容

- 语义模式集合隐式匹配的局部-完整行人重识别方法
- 半注意力部件关联构建的遮挡-完全行人重识别方法
- 基于跨模态相似性推理的红外-可见光行人重识别方法
- 基于多层次模态对齐的文本-图像行人重识别方法

# 4. 半注意力部件关联构建的遮挡-完整行人重识别方法

□ 研究内容：探究信息完整度非对称下的**遮挡-完整**图像场景





# 4. 半注意力部件关联构建的遮挡-完整行人重识别方法

□ 研究内容：探究信息完整度非对称下的**遮挡-完整**图像场景

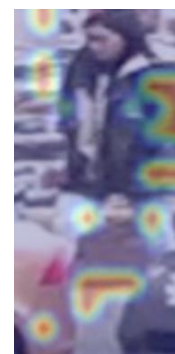


**信息局限：**部分外观不可见  
辨别信息减少

**信息干扰：**遮挡物破坏行人  
各部件拓扑



主流现有方法采用  
**部件关联**  
实现遮挡行人匹配



**注意力模型**  
难以自动学习  
有效部件信息

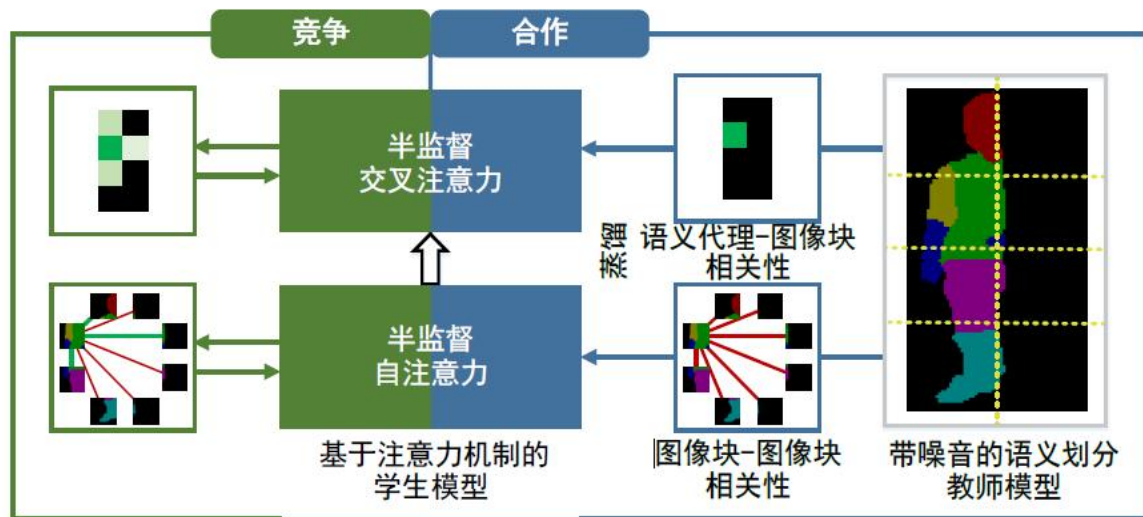


**外部语义模型**  
辅助特征提取  
存在泛化局限

如何有效构建部件特征关联？

# 4. 半注意力部件关联构建的遮挡-完整行人重识别方法

- 动机：针对遮挡条件下**部件特征关联难构建**问题
  - 本文提出SAP (Semi-Attention Partition) 方法
  - 首次引入**竞争和合作师生学习策略**实现有效表征学习——注意力学生模型从带噪音的语义专家模型中学习有效的先验知识



# 4. 半注意力部件关联构建的遮挡-完整行人重识别方法

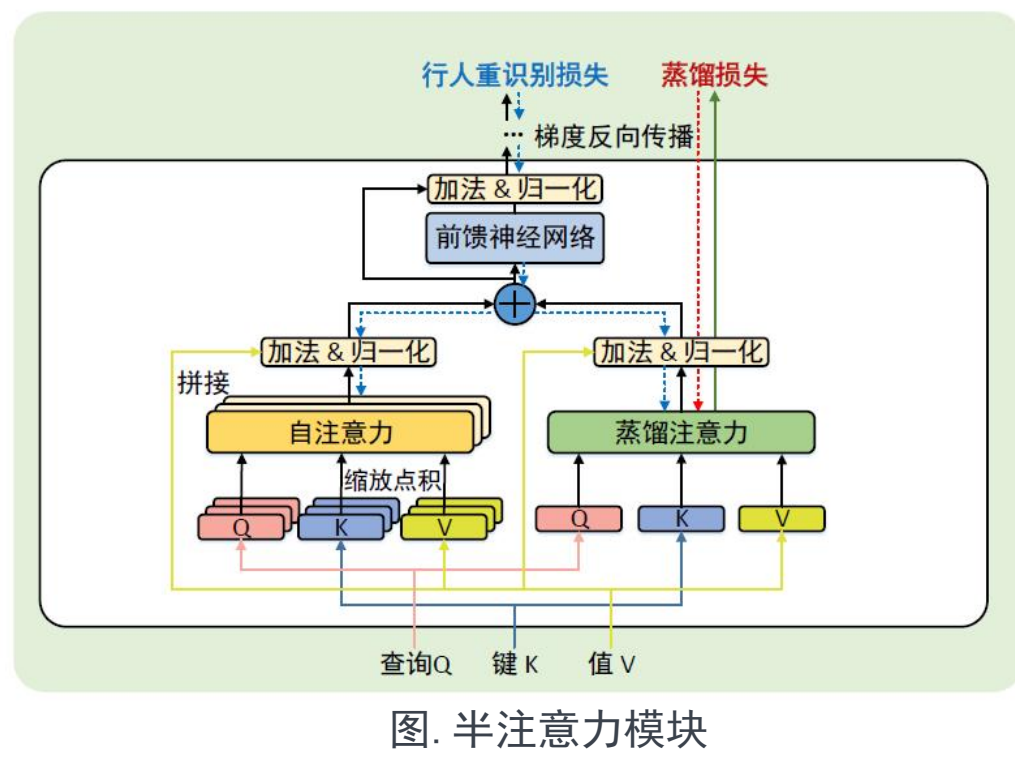
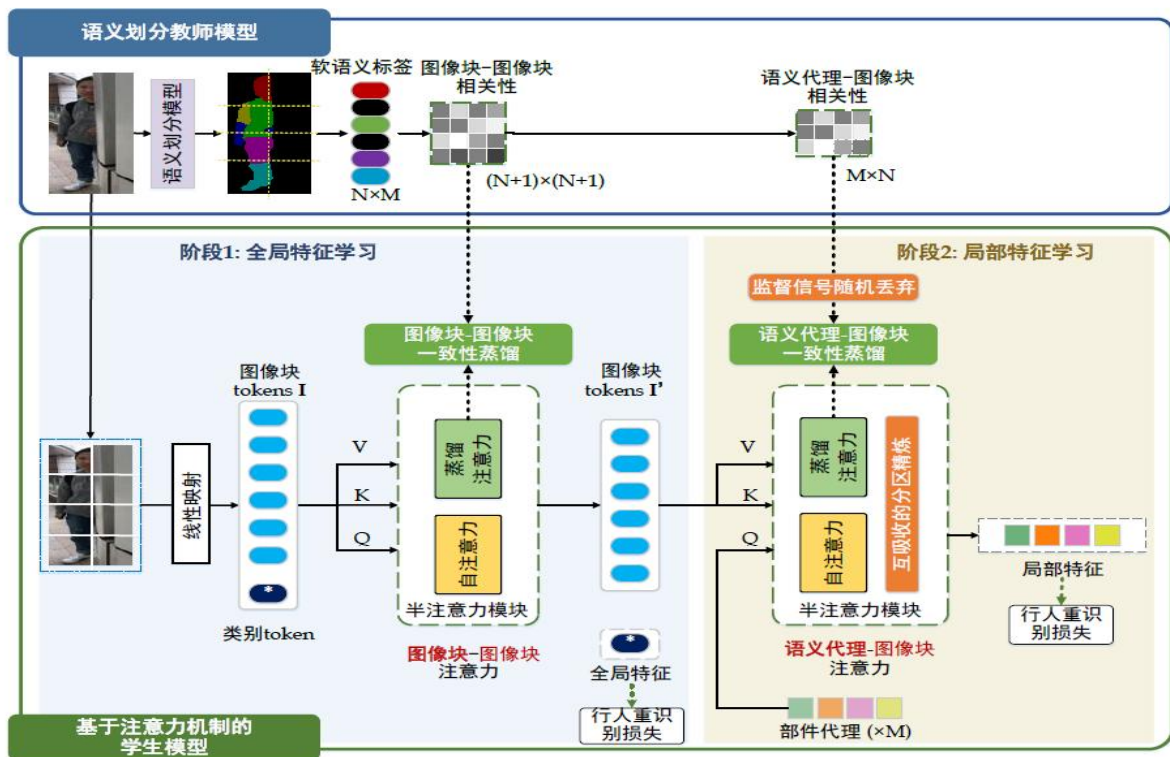
## □ SAP (Semi-Attention Partition) 方法

■ **教师模型**: 语义划分模型 / **学生模型**: Transformer模型

■ **学生-教师学习 (半注意力方式)**

□ **创新点1**: 合作学习 (图像块-图像块一致性 / 代理-图像块一致性)

□ **创新点2**: 竞争学习 (互吸收的分区精炼 / 监督信号随机丢弃)



# 4. 半注意力部件关联构建的遮挡-完整行人重识别方法

## □ 师生合作学习：图像块-图像块一致性 (1/2)

### ■ 图像块-图像块注意力

$$DA_{patch-patch}(\mathbf{I}) = softmax\left(\frac{\mathbf{I}W_q(\mathbf{I}W_k)^T}{\sqrt{D}}\right)\mathbf{I}W_v,$$

### ■ 图像块-图像块级蒸馏

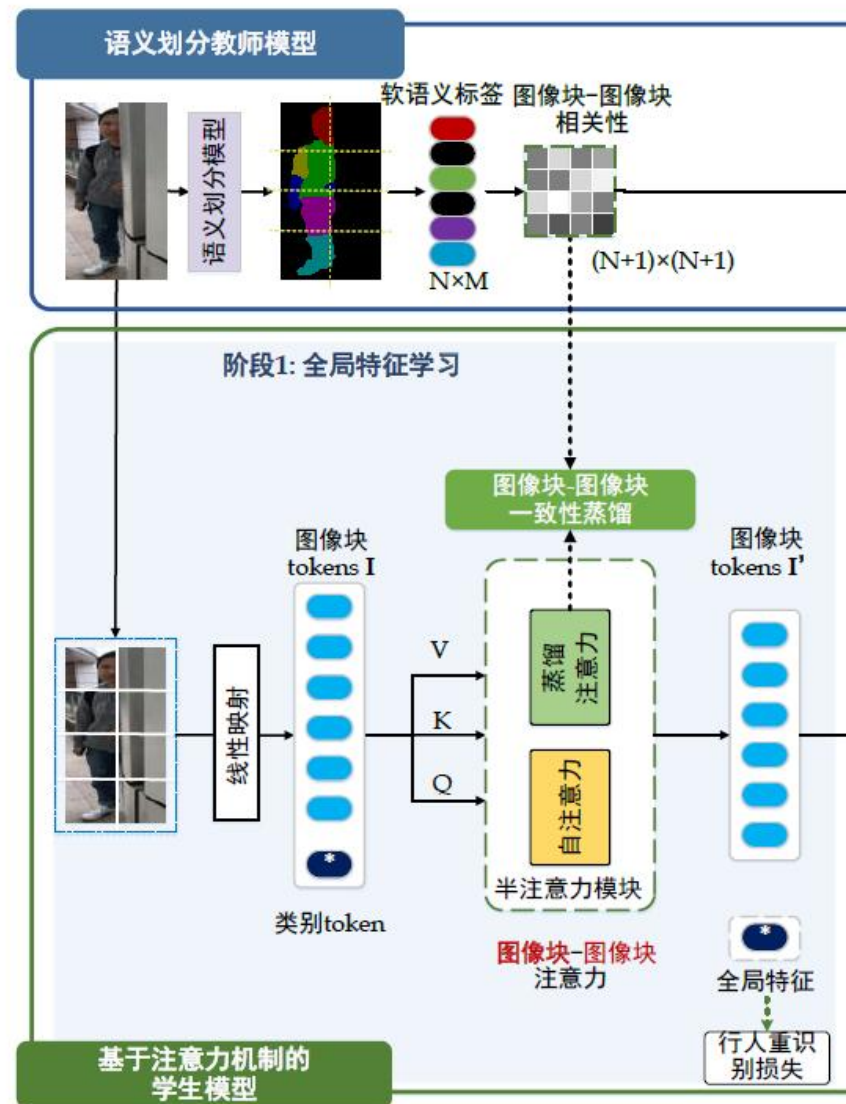
$$\mathcal{L}_{patch} = \alpha \sum softmax(A_{patch}) \log(softmax(\frac{\mathbf{I}W_q(\mathbf{I}W_k)^T}{\sqrt{D}})),$$

教师 (目标): 部件一致性矩阵      学生: 注意力矩阵

## □ $A_{patch}$ 为语义相似度 (蒸馏目标)

$$A_{patch} = \begin{bmatrix} 1 & \mathbf{m}_f^T \\ \mathbf{m}_f & \hat{\mathbf{m}}\hat{\mathbf{m}}^T \end{bmatrix},$$

$$\mathbf{m}_f = \mathbb{I}_{\hat{\mathbf{m}}[B] < \gamma} \in \{0, 1\}^{N \times 1} \quad \hat{\mathbf{m}} = softmax\left(\lambda \sum_{i=1}^P \sum_{j=1}^P \mathcal{G}_{oh}(\mathbf{m})_{[i,j]}\right)$$





# 4. 半注意力部件关联构建的遮挡-完整行人重识别方法

## □ 师生合作学习：代理-图像块一致性 (2/2)

### ■ 代理-图像块注意力

$$DA_{proxy-patch}(Q_{proxy}, I') = \text{softmax}\left(\frac{Q_{proxy}(I'W'_k)^T}{\sqrt{d}}\right)I'W'_v,$$

□  $Q_{proxy}$  为可学习的部件代理（一组向量）

### ■ 代理-图像块级蒸馏

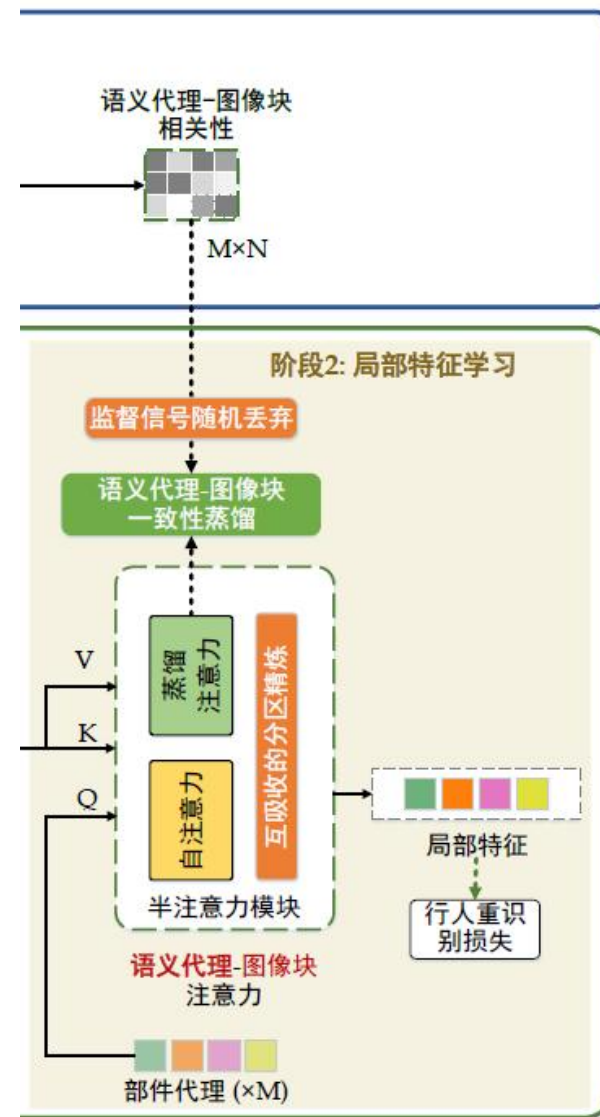
$$\mathcal{L}_{proxy} = \alpha \sum \text{softmax}(\hat{\mathbf{m}}^T) \log(\text{softmax}\left(\frac{Q_{proxy}(I'W'_k)^T}{\sqrt{d}}\right)).$$

教师 (目标)

学生

□  $\hat{\mathbf{m}}^T$  为部件标签矩阵的转置

$$\hat{\mathbf{m}} = \text{softmax}\left(\lambda \sum_{i=1}^P \sum_{j=1}^P \mathcal{G}_{oh}(\mathbf{m})_{[i,j]}\right)$$



## 4. 半注意力部件关联构建的遮挡-完整行人重识别方法

□ 师生竞争学习：两种策略来鼓励学生与有噪声的教师发生合理的竞争

■ 互吸收的分区精炼-适应图像块间的固有相似性

□ 代理→图像块：交叉注意力中代理作为 Q, 图像块作为 K, V

□ 图像块→代理：交叉注意力中图像块作为 Q, 代理作为 K, V

■ 监督信号随机丢弃-稀疏监督 减轻噪声

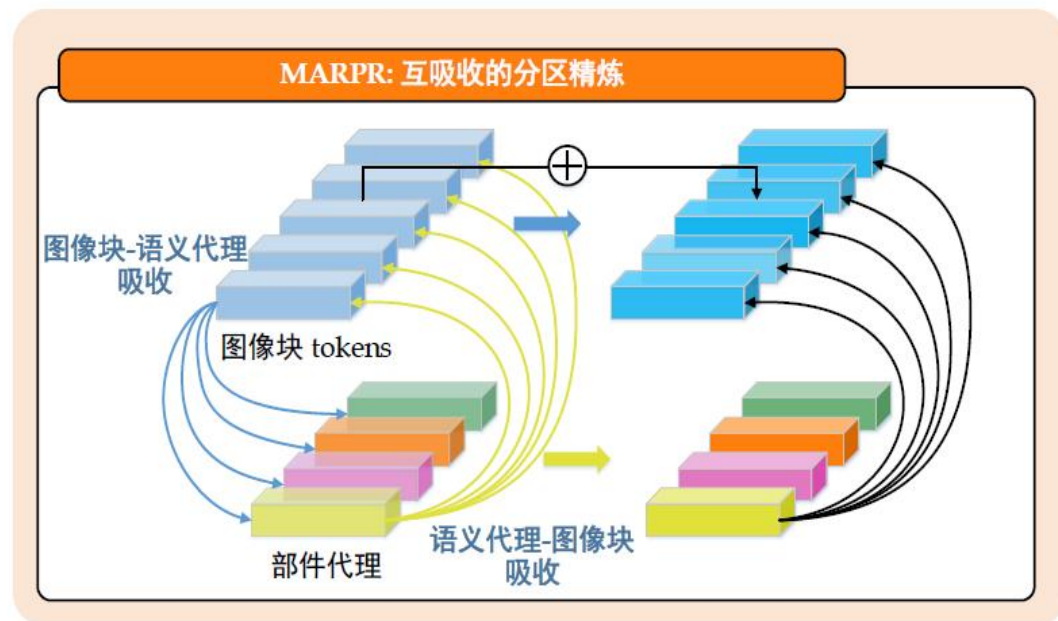
□ 丢弃重复出现的图像块的监督

□ 每个图像块的丢弃概率与其与所有其他图像块的相似度的平均值正相关

$$p(i) = \text{softmax}(\sum \mathbf{T}_{srd}[i])$$

$$\mathbf{T}_{srd} = \mathbf{I}'\mathbf{I}'^T \in \mathbb{R}^{N \times N}$$

[i]表示patch相似性矩阵的第i行。





# 4. 半注意力部件关联构建的遮挡-完整行人重识别方法

□ 实验证明本方法优于现有先进方法

\*数据集

Occluded-Duke

- 15,618张训练图像
- 2,210遮挡查询图像
- 17,661张候选图像

Occluded-REID

- 200个行人
- 2,000张图像
- 每个行人5张遮挡5张完整

在公开基准数据集上显著  
优于现有遮挡-完整行人重  
识别方法

方法	Occluded-Duke		Occluded-REID	
	mAP	Rank-1	mAP	Rank-1
PCB <sup>[76]</sup>	33.7	42.6	38.9	41.3
DSR <sup>[8]</sup>	30.4	40.8	62.8	72.8
SFR <sup>[79]</sup>	32.0	42.3	-	-
FPR <sup>[8]</sup>	-	-	68.0	78.3
ISP <sup>[8]</sup>	52.3	62.8	-	-
Part Bilinear <sup>[77]</sup>	-	36.9	-	-
FD-GAN <sup>[78]</sup>	-	40.8	-	-
PGFA <sup>[7]</sup>	37.3	51.4	-	-
PVPM <sup>[9]</sup>	37.7	47.0	59.5	66.8
HONet <sup>[10]</sup>	43.8	55.1	70.2	80.3
GASM <sup>[92]</sup>	-	-	65.6	74.5
RFCnet <sup>[93]</sup>	54.5	63.9	-	-
LKWS <sup>[6]</sup>	46.3	62.2	71.0	81.0
PAT <sup>[14]</sup>	53.6	64.5	72.1	81.6
Pirt <sup>[15]</sup>	50.9	60.0	-	-
DRL-Net <sup>[16]</sup>	53.9	65.8	-	-
AAformer <sup>[17]</sup>	58.2	67.0	-	-
TransReID <sup>[12]</sup>	59.2	66.4	-	-
PFD <sup>[94]</sup>	61.8	69.5	<b>83.0</b>	81.5
FED <sup>[95]</sup>	56.4	68.1	<b>79.3</b>	<b>86.3</b>
本章方法	<b>62.2</b>	<b>70.0</b>	76.8	83.0

方法	Market-1501		MSMT17	
	mAP	Rank-1	mAP	Rank-1
PCB <sup>[76]</sup>	81.6	93.8	40.4	68.2
MGN <sup>[96]</sup>	86.9	95.7	52.1	76.9
VPM <sup>[13]</sup>	80.8	93.0	-	-
ISP <sup>[8]</sup>	88.6	95.3	-	-
SPReID <sup>[8]</sup>	81.3	92.5	-	-
AANet <sup>[89]</sup>	82.5	93.9	-	-
P <sup>2</sup> -Net <sup>[97]</sup>	85.6	95.2	-	-
PGFA <sup>[7]</sup>	76.8	91.2	-	-
HONet <sup>[10]</sup>	84.9	94.2	-	-
GASM <sup>[92]</sup>	84.7	95.3	52.5	79.5
RFCnet <sup>[93]</sup>	89.2	95.2	60.2	82.0
ABDNet <sup>[98]</sup>	88.3	95.6	60.8	82.3
PAT <sup>[14]</sup>	88.0	95.4	-	-
Pirt <sup>[15]</sup>	86.3	94.1	-	-
DRL-Net <sup>[16]</sup>	86.9	94.7	55.3	78.4
AAformer <sup>[17]</sup>	87.7	95.4	62.6	83.1
TransReID <sup>[12]</sup>	88.9	95.2	67.4	85.3
PFD <sup>[94]</sup>	89.7	95.5	64.4	83.8
FED <sup>[95]</sup>	86.3	95.0	-	-
本章方法	<b>90.5</b>	<b>96.0</b>	<b>67.8</b>	<b>85.7</b>

# 4. 半注意力部件关联构建的遮挡-完整行人重识别方法

## □ 消融实验

方法	mAP	Rank-1	Rank-5
基线	56.0	64.8	79.8
基线 + 图像块半监督	59.6	66.6	81.0
基线 + 代理半监督	60.1	66.4	81.2
基线 + 图像块半监督 + 代理半监督	<b>62.2</b>	<b>70.0</b>	<b>83.0</b>
基线 + 图像块半监督 + 代理半监督 <i>w/o MAPR</i>	60.8	68.0	81.2
基线 + 图像块半监督 + 代理半监督 <i>w/o SSD</i>	61.8	68.9	82.4

详细的消融实验证明各个模块均有性能提升

教师模型	Occluded-Duke		Market-1501		MSMT17	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
基于手工规则	60.6	67.4	90.1	95.7	66.3	83.9
基于姿态估计	61.1	68.2	90.2	95.8	67.1	84.5
基于人体解析	62.2	70.0	90.5	96.0	67.8	85.7

本章方法可与不同的教师模型/预训练数据集兼容，具有较强的鲁棒性

预训练数据集	Occluded-Duke		Market-1501	
	mAP	Rank-1	mAP	Rank-1
ATR <sup>[99]</sup>	61.4	68.8	89.8	95.5
lip <sup>[100]</sup>	61.6	69.1	90.5	95.7
Pascal <sup>[101]</sup>	62.2	70.0	90.5	96.0

# 提纲

---

## □ 绪论

- 研究背景与意义
- 研究问题与难点
- 本文研究内容

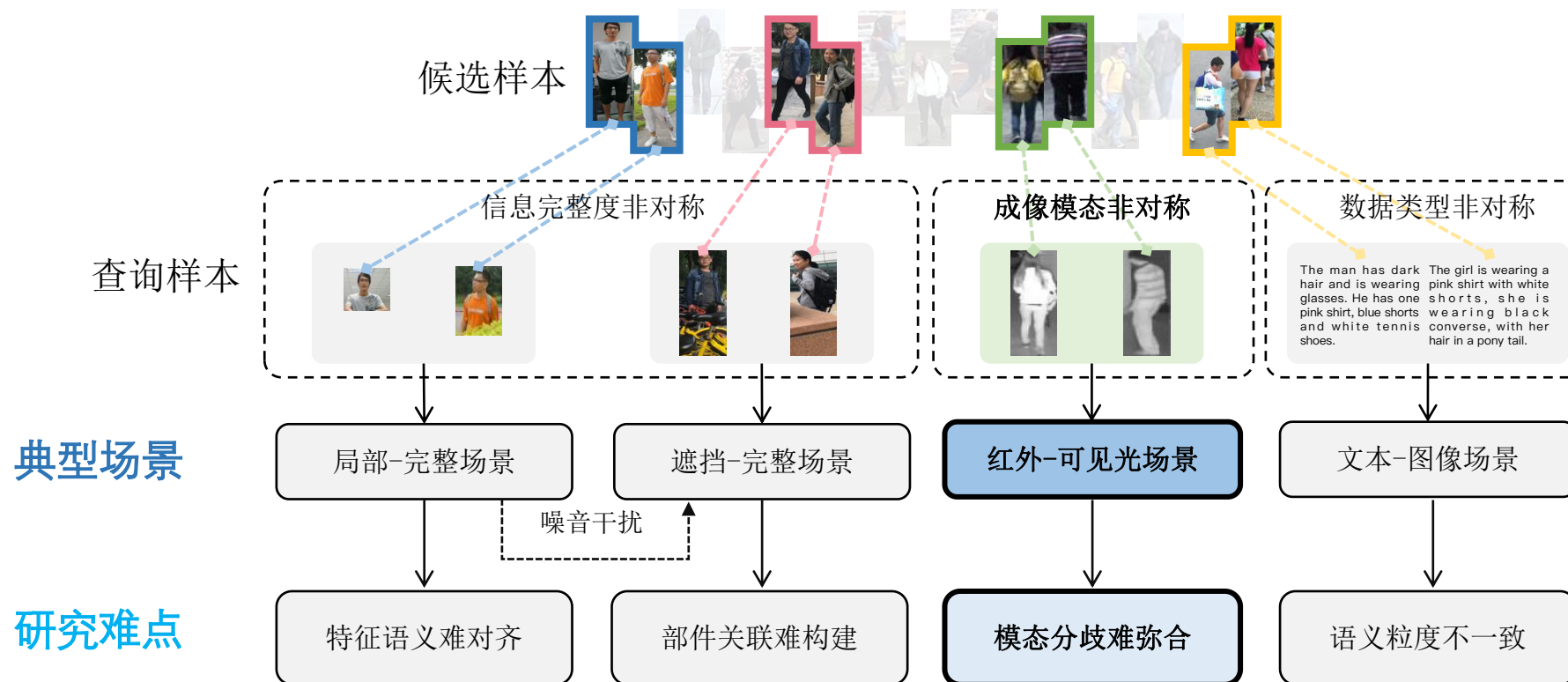
## □ 相关研究工作

## □ 研究内容

- 语义模式集合隐式匹配的局部-完整行人重识别方法
- 半注意力部件关联构建的遮挡-完全行人重识别方法
- 基于跨模态相似性推理的红外-可见光行人重识别方法
- 基于多层次模态对齐的文本-图像行人重识别方法

# 5. 基于跨模态相似性推理的红外-可见光行人重识别方法

□ 研究内容：探究成像模态非对称下的**红外-可见光**图像场景



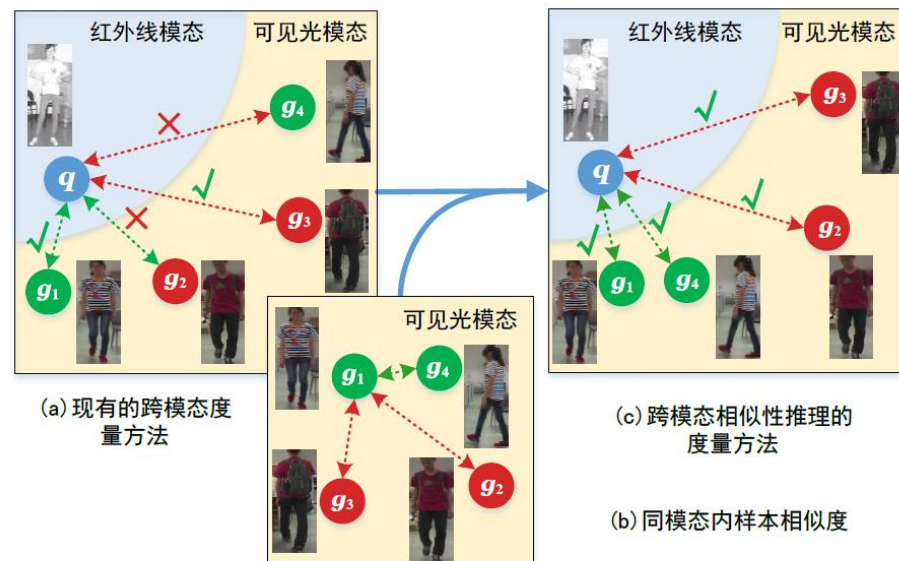
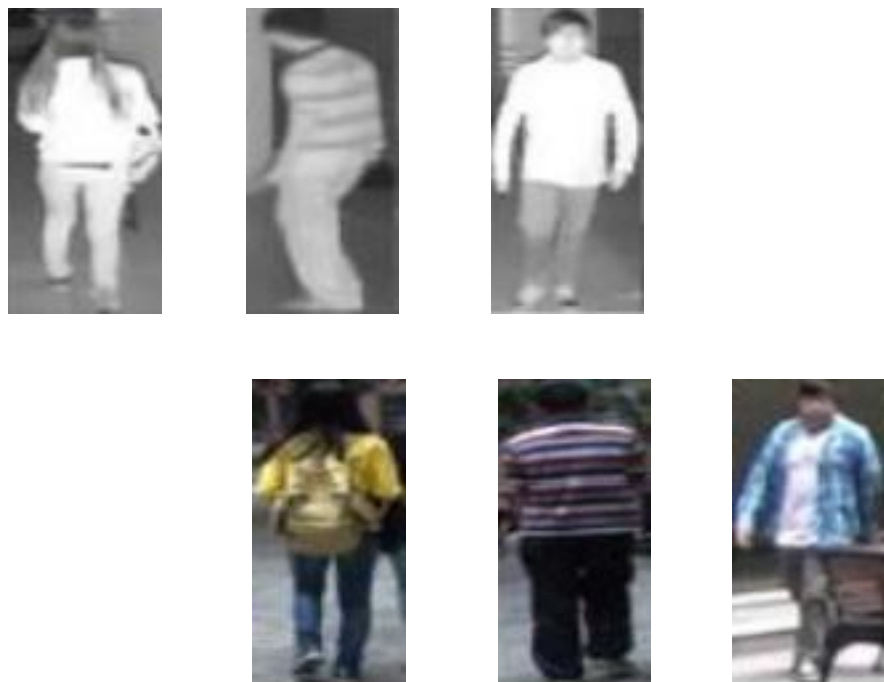
# 5. 基于跨模态相似性推理的红外-可见光行人重识别方法

□ 针对成像模态差异导致的特征分歧难弥合问题

动机：直接度量受限  
红外-可见光图像间图像通道数不同、  
成像原理不同，分歧较大



方法：采用间接度量  
提出将相似性度量问题转化为相似图上的  
最短路径问题



$q$  查询图像 (红外线)       $\longleftrightarrow$  模型预测为正样本  
 $g_1, g_4$  正样本候选图像 (可见光)       $\dashrightarrow$  模型预测为负样本  
 $g_2, g_3$  负样本候选图像 (可见光)       $\checkmark$  正确预测       $\times$  错误预测



# 5. 基于跨模态相似性推理的红外-可见光行人重识别方法

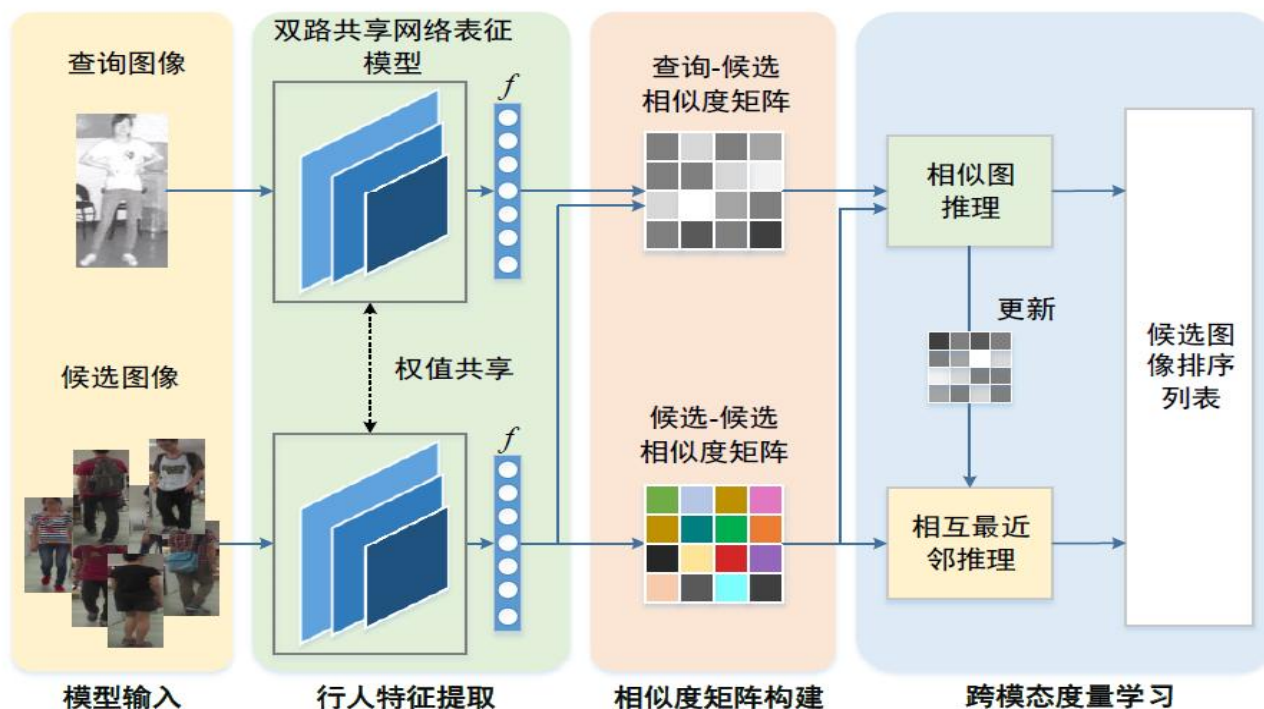
## □ SIM (Similarity Inference Metric) 方法

### ■ 相似图推理 (创新点1)

□ 在测试集上构建相似图，并通过求解样本间的最短路径来度量二者相似性

### ■ 相互最近邻推理 (创新点2)

□ 通过对比同模态下邻近样本集合的相似性，以克服模态间的特征分歧





# 5. 基于跨模态相似性推理的红外-可见光行人重识别方法

## □ 相似图推理（创新点1）

- 构建样本相似图，求解该图上两个样本的最短路径作为二者的相似度量

步骤1：根据初始距离 $D_{q,g}$ ,  $D_{g,g}$ 构建图

$$\begin{cases} \mathcal{E}(q_i, g_j) = D_{q,g}(q_i, g_j) \\ \mathcal{E}(g_j, g_k) = \lambda D_{g,g}(g_j, g_k), \end{cases} \quad (1)$$

步骤2：计算最短路径

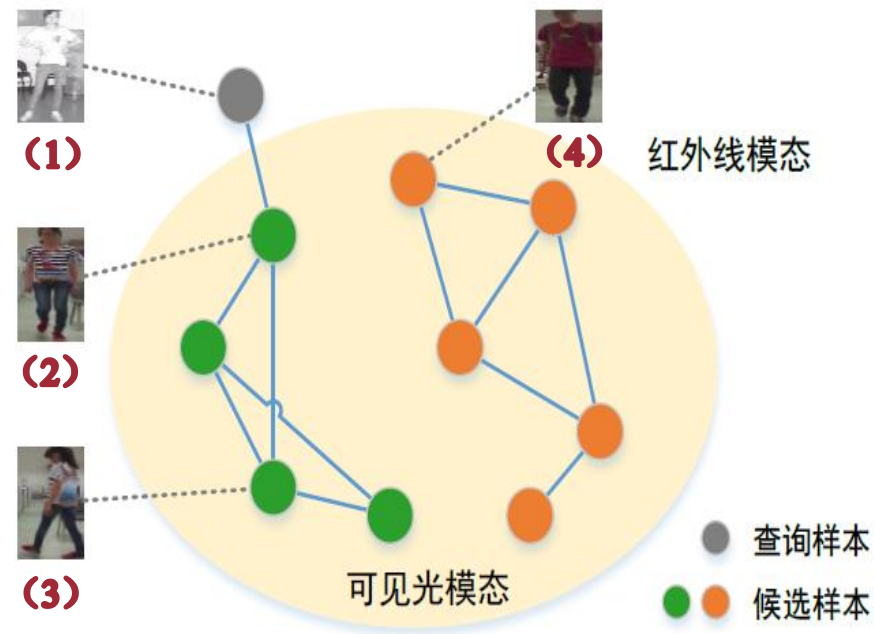
$$d(q_i, g_j) = \min_{\mathcal{P} \in \Omega_{q_i, g_j}} \sum_{t=1}^{n-1} \mathcal{E}(p_t, p_{t+1}). \quad (2)$$

$$\mathcal{E}(g_i, g_j) + \mathcal{E}(g_j, g_t) \geq \mathcal{E}(g_i, g_t), \quad \forall 1 \leq i, j, t \leq N_g. \quad (3)$$

$$d(q_i, g_j) = \min_{1 \leq t \leq N_g} \{\mathcal{E}(q_i, g_t) + \mathcal{E}(g_t, g_j)\}. \quad (4)$$

步骤3：计算Top-K最短路径的平均

$$d_S(q_i, g_j) = \frac{1}{K} \sum_{k=1}^K d^{(k)}(q_i, g_j). \quad (5)$$



# 5. 基于跨模态相似性推理的红外-可见光行人重识别方法

## □ 相互最近邻推理（创新点2）

- 对比最近邻样本集合的集合相似性，作为辅助相似度量
- 通过集合相似性，克服显著的模态间特征分歧

步骤1：构建最近邻集合

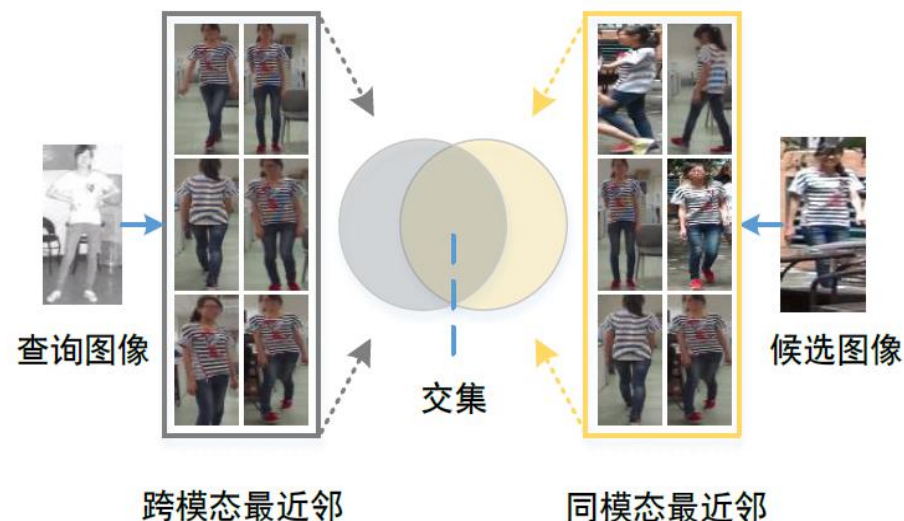
$$\mathcal{N}_c(\mathbf{q}_i, k_q, d_S) = \{g^{(1)}, g^{(2)}, \dots, g^{(k_q)}\}. \quad (6)$$

步骤2：计算集合间的杰卡德距离

$$d_M(\mathbf{q}_i, \mathbf{g}_j) = 1 - \frac{|\mathcal{N}_c(\mathbf{q}_i, k_q, d_S) \cap \mathcal{R}_i^*(\mathbf{g}_j, k_g, D_{g,g})|}{|\mathcal{N}_c(\mathbf{q}_i, k_q, d_S) \cup \mathcal{R}_i^*(\mathbf{g}_j, k_g, D_{g,g})|}. \quad (7)$$

步骤3：最终相似度为相似图与相互最近邻的加权和

$$d_{SIM} = \alpha d_S + (1 - \alpha) d_M. \quad (8)$$



# 5. 基于跨模态相似性推理的红外-可见光行人重识别方法

## □ 实验分析

\*数据集 SYSU-MM01

- 419个行人
- 训练集可见光/红外图像
- 19659/12792张

RegDB

- 412个行人
- 每人10张可见光、10张红外图像

方法	可见光检索红外线		红外线检索可见光	
	mAP	Rank-1	mAP	Rank-1
LOMO <sup>[111]</sup>	2.28	0.85	-	-
HOG <sup>[112]</sup>	10.31	13.49	-	-
GSM <sup>[113]</sup>	15.06	17.28	-	-
One-stream <sup>[22]</sup>	14.02	13.11	-	-
Two-stream <sup>[22]</sup>	12.43	30.36	-	-
Zero-Padding <sup>[22]</sup>	18.90	17.75	17.82	16.63
TONE <sup>[23]</sup>	14.92	16.87	16.98	13.86
HCML	20.08	24.44	22.24	21.70
BDTR <sup>[39]</sup>	31.83	33.47	31.10	32.72
HSME <sup>[40]</sup>	38.80	41.30	37.50	40.70
D-HSME <sup>[40]</sup>	47.00	50.90	46.20	50.20
D <sup>2</sup> RL <sup>[32]</sup>	44.10	43.40	44.10	43.40
PIG <sup>[36]</sup>	49.30	48.50	48.90	48.10
AlignGAN <sup>[33]</sup>	53.60	57.90	53.40	56.30
<b>本章方法</b>	<b>75.29</b>	<b>74.47</b>	<b>78.30</b>	<b>75.24</b>

方法	mAP	Rank-1
LOMO <sup>[111]</sup>	2.28	4.70
HOG <sup>[112]</sup>	2.16	3.82
GSM <sup>[113]</sup>	4.38	6.19
One-stream <sup>[22]</sup>	8.59	16.3
Two-stream <sup>[22]</sup>	8.03	16.4
Zero-Padding <sup>[22]</sup>	10.9	19.2
cmGAN <sup>[24]</sup>	22.27	31.49
PIG <sup>[36]</sup>	29.5	45.1
AlignGAN <sup>[33]</sup>	33.90	51.50
<b>本章方法</b>	<b>60.88</b>	<b>56.93</b>

方法	mAP	Rank-1
AlignGAN* <sup>[33]</sup>	34.56	47.92
AlignGAN + 相似图推理	51.30	52.56
AlignGAN + 相互最近邻推理	50.26	52.33
AlignGAN + 相似度推理 (本章方法)	<b>54.45</b>	<b>52.70</b>
AGW <sup>[2]</sup>	40.03	50.87
AGW + 相似图推理	55.89	52.70
AGW + 相互最近邻推理	51.40	52.93
AGW + 相似度推理 (本章方法)	<b>57.47</b>	<b>53.75</b>

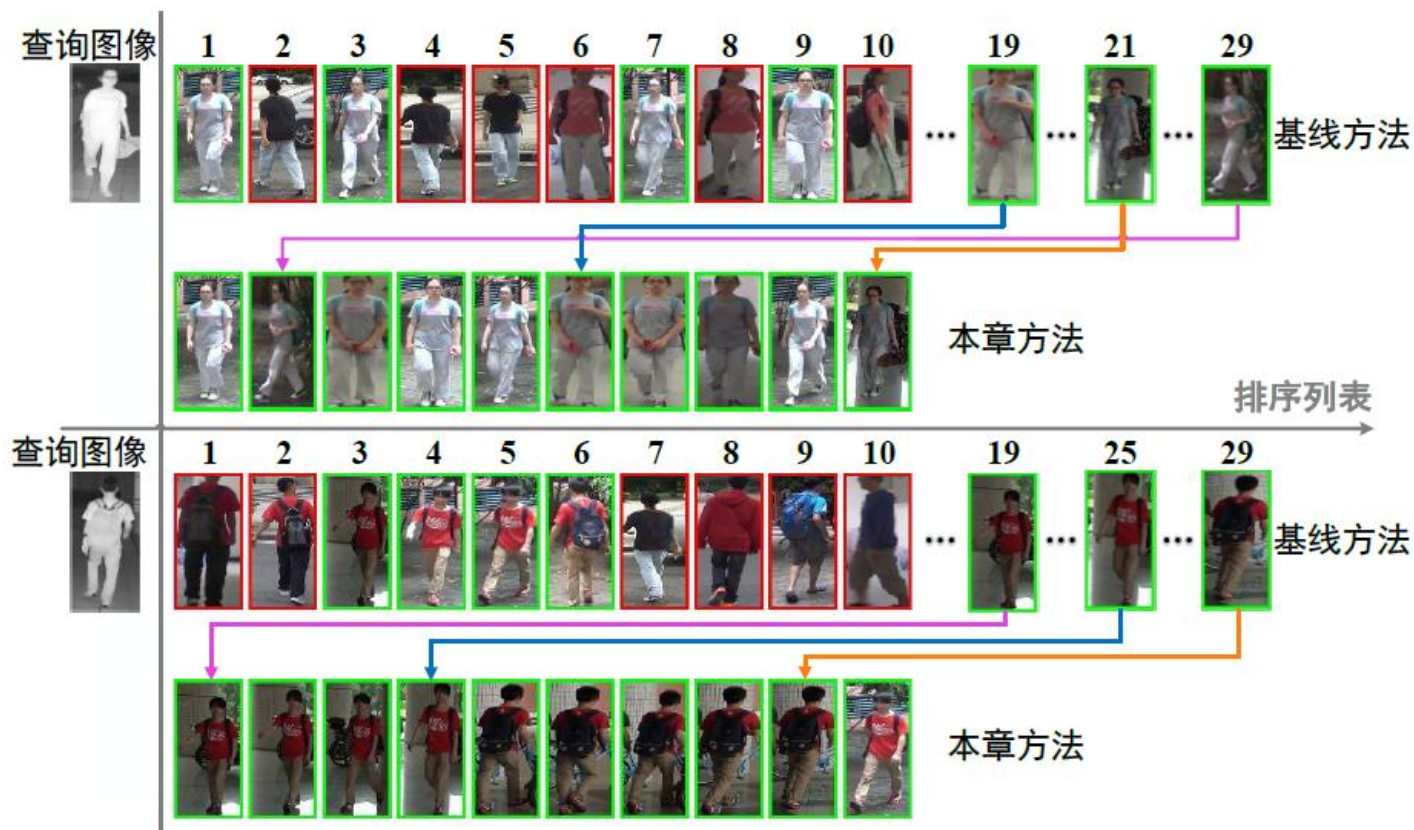
在公开基准数据集上显著优于现有红外-可见光重识别方法

可与多种基线方法兼容，即插入即用显著提升现有性能

# 5. 基于跨模态相似性推理的红外-可见光行人重识别方法

## □ 可视化分析

- 红色框：负样本
- 绿色框：正样本



相似度排序结果表明本方法显著优于基础方法



# 提纲

---

## □ 绪论

- 研究背景与意义
- 研究问题与难点
- 本文研究内容

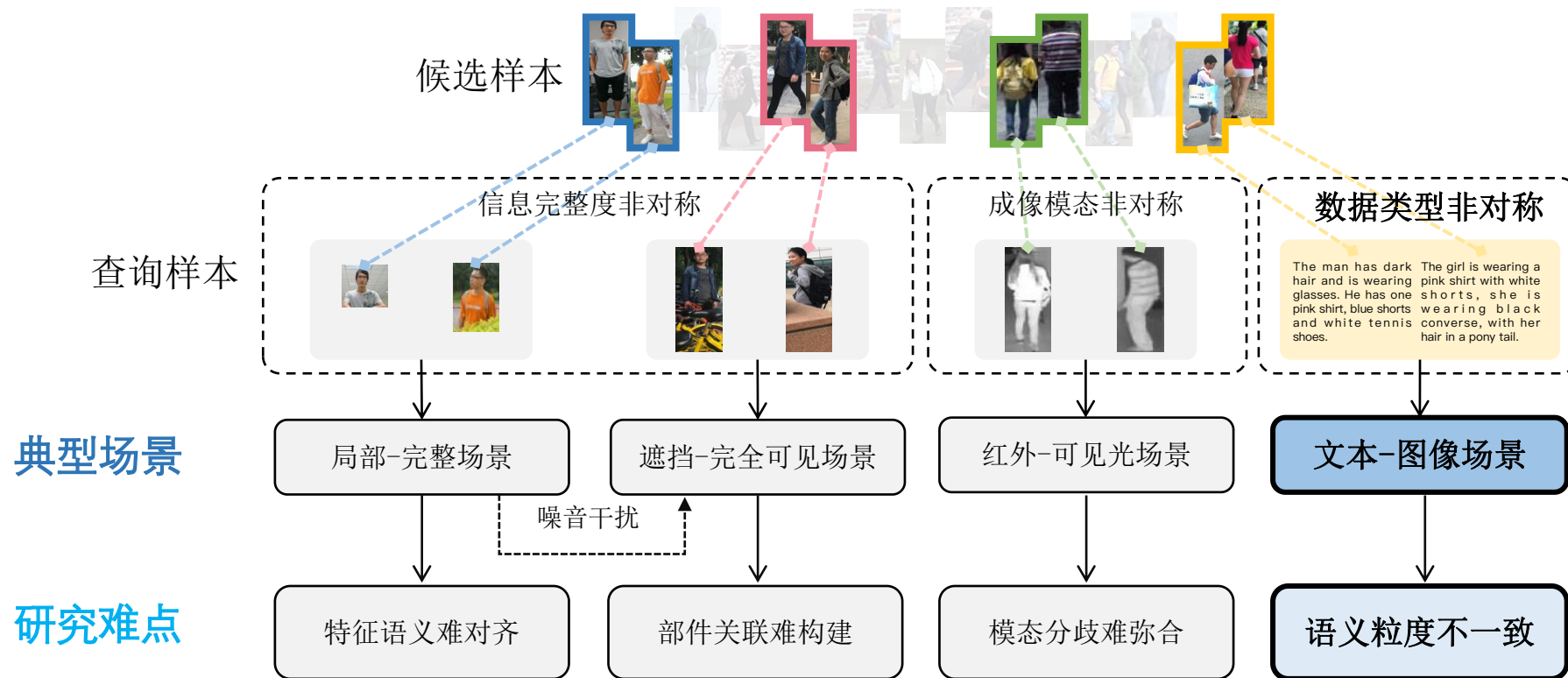
## □ 相关研究工作

## □ 研究内容

- 语义模式集合隐式匹配的局部-完整行人重识别方法
- 半注意力部件关联构建的遮挡-完全行人重识别方法
- 基于跨模态相似性推理的红外-可见光行人重识别方法
- 基于多层次模态对齐的文本-图像行人重识别方法

# 6. 基于多层次模态对齐的文本-图像行人重识别方法

□ 研究内容：探究数据类型非对称下的**文本-图像**场景





# 6. 基于多层次模态对齐的文本-图像行人重识别方法

□ 研究内容：探究数据类型非对称下的**文本-图像**场景

行人图像缺失



文字描述搜索视觉图像

固有差异性



抽象信息  
(粗粒度)

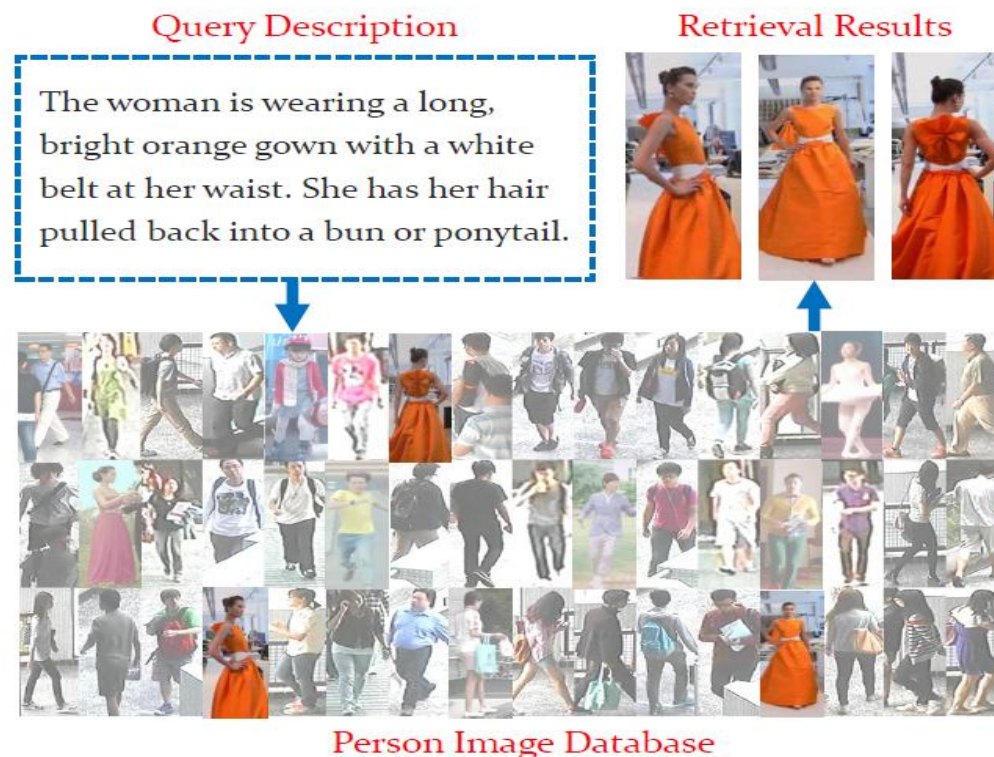


具象信息  
(细粒度)

主要挑战



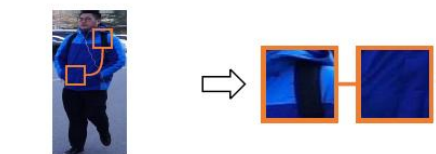
如何解决文本-图像数据类型差异导致的**特征粒度不一致**问题？



## 6. 基于多层次模态对齐的文本-图像行人重识别方法

- 针对文本-图像数据类型差异导致的**特征粒度不一致**问题
  - MCMA (Multi-level Cross-Modality Alignment) 方法
  - 建模多层次的多模态对齐关系（图像块级、实例级、身份级）

“The man is wearing a *blue coat*”



(a) 图像块级不对齐



(b) 实例级不对齐



(c) 身份级不对齐

- ◆ **图像块级不对齐**：即同一语义区域中的**不同图像块**具有不同外观表征。
- ◆ **实例级不对齐**：即同一语义区域在**同一行人的不同实例间**（角度、光照不同）具有不同外观表征。
- ◆ **身份级分不对齐**：即同一语义区域在**不同行人身份**下的具体对象具有不同外观表征。

# 6. 基于多层次模态对齐的文本-图像行人重识别方法

## 创新点1：跨模态掩码重构学习

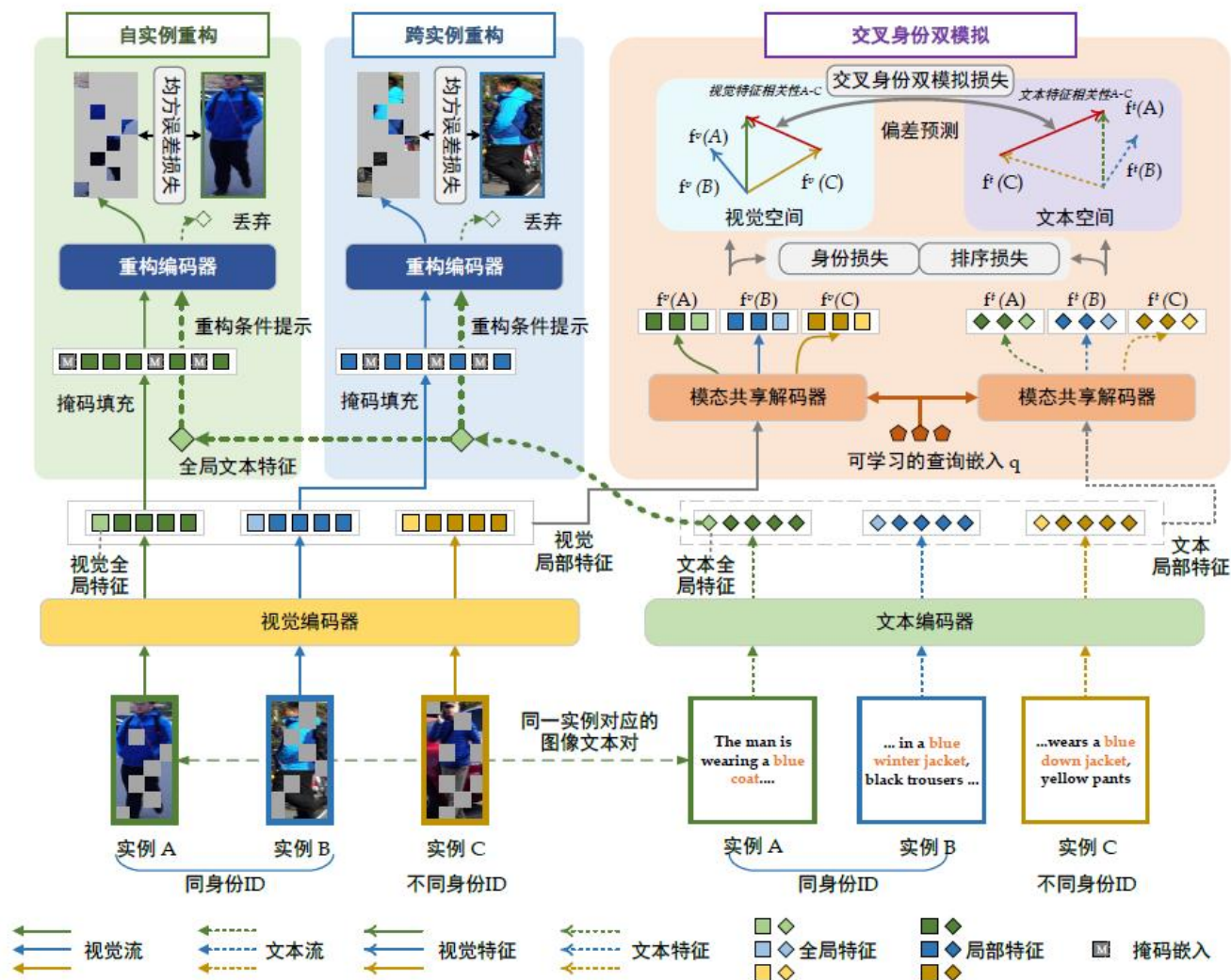
- 引入多层次的**基于文本提示的图像块掩码重构任务**，根据可见的图像块和描述性的文本提示中来学习重建被掩码的图像区域。
- 通过同时在实例内以及实例间进行多层次跨模态重建，来实现图像块与实例级的跨模态一致性。

## 自实例重构

$$\hat{\mathbf{T}}_{A,m}^v = E_{rec}(\{\mathbf{Z}_{A,um}^v, \mathbf{Z}^m, \mathbf{z}_{A,glo}^t\})$$

## 跨实例重构

$$\hat{\mathbf{T}}_{B,m}^v = E_{rec}(\{\mathbf{Z}_{B,um}^v, \mathbf{Z}^m, \mathbf{z}_{A,glo}^t\})$$





# 6. 基于多层次模态对齐的文本-图像行人重识别方法

## □ 创新点2: 交叉身份双模拟学习

- 通过学习一个辅助偏差预测器, 使得不同身份行人的特征间相关性可以在不同模态间无损转换

最大化互信息

$$\arg \max_{\theta^f} \mathbb{I}(\mathbf{r}^V; \mathbf{r}^T),$$

优化InfoNCE目标

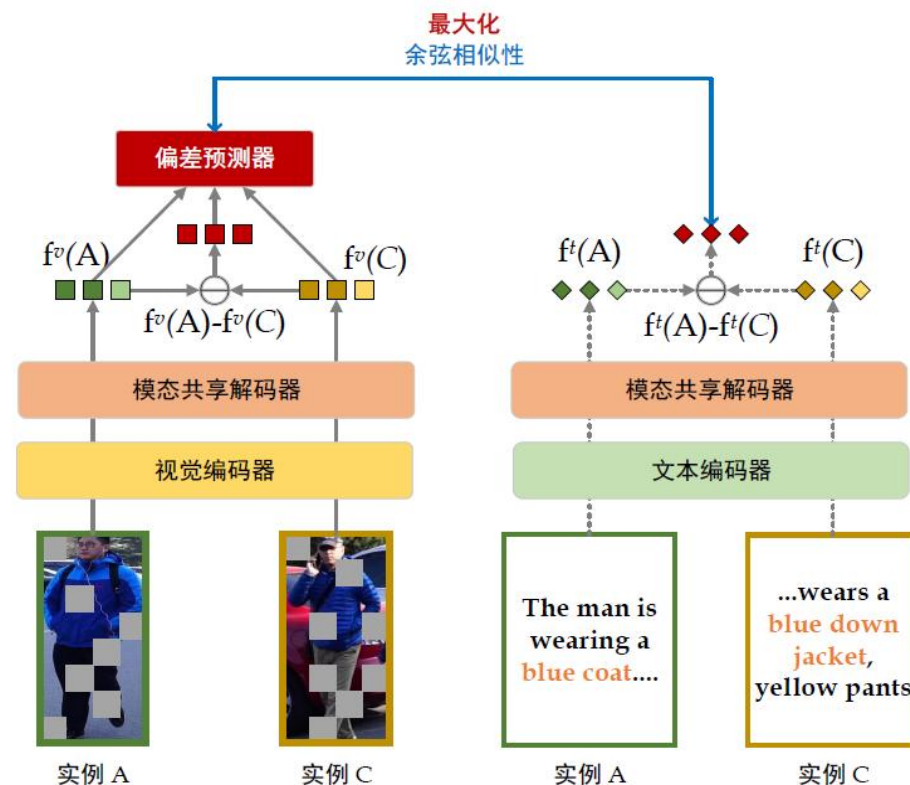
$$\arg \min_{\theta^f, \theta^p} -\mathbb{E}_{a,c \in X} \log \frac{\exp(s(r_{a,c}^V, r_{a,c}^T | \theta^p))}{\sum_{a',c' \in X} \exp(s(r_{a,c}^V, r_{a',c'}^T | \theta^p))},$$

模态间预测

$$s(r_{a,c}^V, r_{a,c}^T | \theta^p) = \frac{\langle r_{a,c}^V, \hat{r}_{a,c}^V \rangle}{\|r_{a,c}^V\| \cdot \|\hat{r}_{a,c}^V\|} + \frac{\langle r_{a,c}^T, \hat{r}_{a,c}^T \rangle}{\|r_{a,c}^T\| \cdot \|\hat{r}_{a,c}^T\|},$$

交叉身份双模拟学习的损失

$$\mathcal{L}_{bisim} = -\log \frac{\exp(s(r_{a,c}^V, r_{a,c}^T))}{\sum_{a',c' \in B} \exp(s(r_{a,c}^V, r_{a',c'}^T))},$$



# 6. 基于多层次模态对齐的文本-图像行人重识别方法

## □ 实验证明本方法优于现有先进方法

### \*数据集

#### CUHK-PEDES

- 训练集 / 测试集
- 图像:34,054 / 3,074
- 文本:68,120 / 6,156

#### ICFG-PEDES

- 训练集 / 测试集
- 图像:34,674 / 19,848
- 文本图像 1:1

#### RSTPReid

- 20,505张图像
- 41,010个文本

主干网络	方法	CUHK-PEDES 数据集		
		Rank-1	Rank-5	Rank-10
ResNet	Dual Path <sup>[50]</sup>	44.40	66.26	75.07
	CMPM/C <sup>[52]</sup>	49.37	-	79.27
	MIA <sup>[119]</sup>	53.10	75.00	82.90
	A-GANet <sup>[134]</sup>	53.14	75.45	82.95
	GALM <sup>[55]</sup>	54.51	77.56	82.97
	TIMAM <sup>[121]</sup>	54.51	77.56	84.78
	TDE <sup>[135]</sup>	55.25	77.46	84.56
	VTA <sup>[136]</sup>	55.32	77.00	84.26
	SCAN <sup>[137]</sup>	55.86	75.97	83.69
	ViTAA <sup>[136]</sup>	55.97	75.84	83.52
	CMAAM <sup>[53]</sup>	56.68	77.18	84.86
	HGAN <sup>[57]</sup>	59.00	79.49	86.62
	NAFS <sup>[138]</sup>	59.94	79.86	86.70
	DSSL <sup>[133]</sup>	59.98	80.41	87.56
	MGEL <sup>[118]</sup>	60.27	80.01	86.74
	SSAN <sup>[139]</sup>	61.37	80.15	86.73
	Han <i>et al</i> <sup>[116]</sup>	61.65	80.98	86.78
	LapsCore <sup>[140]</sup>	63.40	-	87.80
	ISANet <sup>[141]</sup>	63.92	82.15	87.69
	AXM-Net <sup>[142]</sup>	64.44	80.52	86.77
CAIBC <sup>[143]</sup>	64.43	82.87	88.37	
Transformer	SAF <sup>[144]</sup>	64.13	82.62	88.40
	CFine-IMG <sup>[145]</sup>	65.07	83.01	89.00
	LGUR <sup>[146]</sup>	65.25	83.12	89.00
	IVT <sup>[147]</sup>	65.59	83.11	89.21
	MCMA	<b>67.33</b>	<b>84.58</b>	<b>90.00</b>

主干网络	方法	ICFG-PEDES 数据集		
		Rank-1	Rank-5	Rank-10
ResNet	Dual Path <sup>[47]</sup>	38.99	59.44	68.41
	CMPM/C <sup>[49]</sup>	43.51	65.44	74.26
	MIA <sup>[118]</sup>	46.49	67.14	75.18
	SCAN <sup>[145]</sup>	50.05	69.65	77.21
	ViTAA <sup>[135]</sup>	50.98	68.79	75.78
	SSAN <sup>[138]</sup>	54.23	72.63	79.53
	ISANet <sup>[140]</sup>	57.73	75.42	81.72
	CFine-IMG <sup>[144]</sup>	55.69	72.72	79.46
Transformer	IVT <sup>[146]</sup>	56.04	73.60	80.22
	LGUR <sup>[145]</sup>	59.02	75.32	81.56
	本章方法	<b>61.21</b>	<b>77.08</b>	<b>82.98</b>

主干网络	方法	RSTPReid 数据集		
		Rank-1	Rank-5	Rank-10
ResNet	IMG-Net <sup>[147]</sup>	37.60	61.15	73.55
	AMEN <sup>[148]</sup>	38.45	62.40	73.80
	DSSL <sup>[132]</sup>	39.05	62.60	73.95
	SUM <sup>[149]</sup>	41.38	67.48	76.48
	SSAN <sup>[138]</sup>	43.50	67.80	77.15
	CAIBC <sup>[142]</sup>	47.35	69.55	79.00
	LGUR <sup>[145]</sup>	45.55	68.20	77.85
Transformer	CFine-IMG <sup>[144]</sup>	45.85	70.30	78.40
	IVT <sup>[146]</sup>	46.70	70.00	78.80
	本章方法	<b>53.80</b>	<b>74.75</b>	<b>83.00</b>

## 6. 基于多层次模态对齐的文本-图像行人重识别方法

### □ 泛化性分析和消融实验

方法	CUHK-PEDES → ICFG-PEDES		
	Rank-1	Rank-5	Rank-10
Dual Path <sup>[47]</sup>	15.41	29.80	38.19
MIA <sup>[118]</sup>	19.35	36.78	46.42
SCAN <sup>[136]</sup>	21.27	39.26	48.83
SSAN <sup>[138]</sup>	24.72	43.43	53.01
SSAN(w/BERT) <sup>[138]</sup>	29.24	49.00	58.83
LGUR <sup>[145]</sup>	34.25	52.58	60.85
<b>本章方法</b>	<b>37.37</b>	<b>55.30</b>	<b>63.64</b>

方法	ICFG-PEDES → CUHK-PEDES		
	Rank-1	Rank-5	Rank-10
Dual Path <sup>[47]</sup>	7.63	17.14	23.52
MIA <sup>[118]</sup>	10.93	23.77	32.39
SCAN <sup>[136]</sup>	13.63	28.61	37.05
SSAN <sup>[138]</sup>	16.68	33.84	43.00
SSAN(w/BERT) <sup>[138]</sup>	21.07	38.94	48.54
LGUR <sup>[145]</sup>	25.44	44.48	54.39
<b>本章方法</b>	<b>29.55</b>	<b>49.40</b>	<b>58.45</b>

方法	模块			CUHK-PEDES		
	自实例重构	跨实例重构	交叉身份双模拟	Rank-1	Rank-5	Rank-10
基线				59.32	79.61	86.73
基线 + 模态共享解码器				63.69	81.77	87.64
方法 1	✓			65.14	82.10	88.24
方法 2	✓	✓		66.34	83.95	89.12
方法 3			✓	66.33	82.60	88.63
方法 4	✓		✓	67.02	83.92	88.97
<b>本章方法</b>	<b>✓</b>	<b>✓</b>	<b>✓</b>	<b>67.28</b>	<b>84.52</b>	<b>90.00</b>

方法	模块			RSTPReid		
	自实例重构	跨实例重构	交叉身份双模拟	Rank-1	Rank-5	Rank-10
基线				47.75	69.40	78.85
基线 + 模态共享解码器				49.65	70.65	79.00
方法 1	✓			50.75	72.15	81.85
方法 2	✓	✓		52.00	73.85	82.15
方法 3			✓	50.80	72.75	81.80
方法 4	✓		✓	51.80	74.65	82.15
<b>本章方法</b>	<b>✓</b>	<b>✓</b>	<b>✓</b>	<b>53.80</b>	<b>74.75</b>	<b>83.00</b>

在多个层次上学习适应性的跨模态对齐关系，有助于提取更具有外观鲁棒性的多模态特征，提高模型的泛化能力。消融实验验证各个模块的显著有效性。



# 6. 基于多层次模态对齐的文本-图像行人重识别方法

## □ 可视化分析

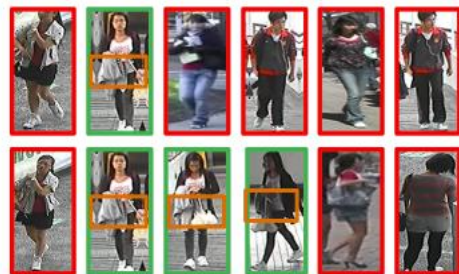
A man wears a **floor length camel colored robe**. He is standing with his back turned.



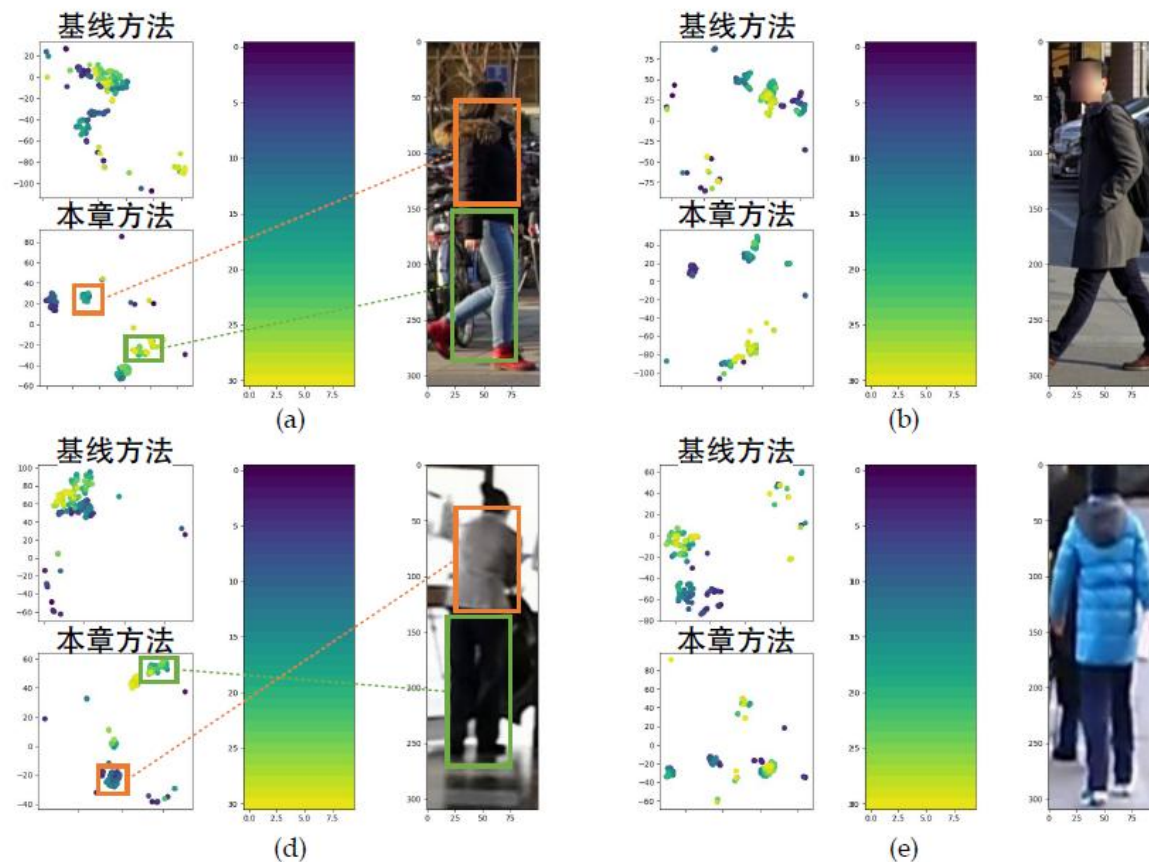
The man wears a blue T-shirt blue jeans and **black and white sneakers**. He walks near a bicycle carrying a black backpack.



The woman was on white sneakers and a red and white T-shirt, she is also **carrying a grey jacket**.



有效改善检索结果排序列表



有效增强局部特征的语义判别性

# 提纲

---

## □ 绪论

- 研究背景与意义
- 研究问题与难点
- 本文研究内容

## □ 相关研究工作

## □ 研究内容

- 语义模式集合隐式匹配的局部-完整行人重识别方法
- 半注意力部件关联构建的遮挡-完全行人重识别方法
- 基于跨模态相似性推理的红外-可见光行人重识别方法
- 基于多层次模态对齐的文本-图像行人重识别方法



**Thanks**