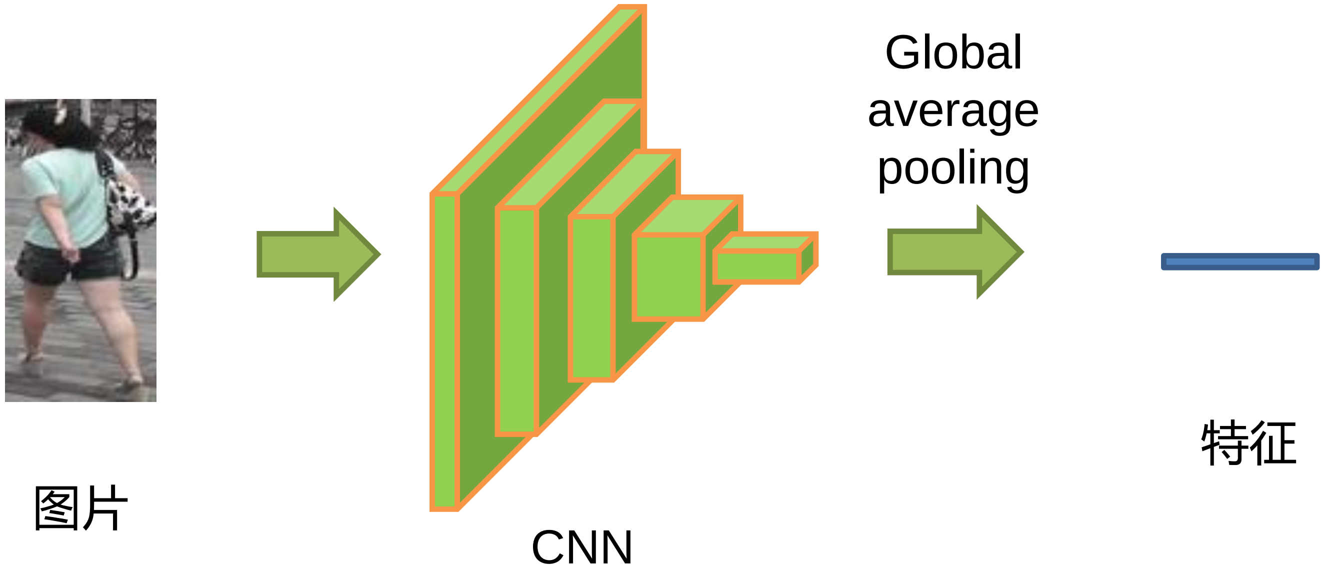
行人重识别——（全局特征和）局部特征

罗浩
浙江大学

基于深度学习的行人重识别

全局特征（Global features）



全局特征是指每一张行人图片的全局信息进行一个特征抽取，这个全局特征没有任何的空间信息

基于深度学习的行人重识别

全局特征（Global features）

噪声区域会对全局特征造成极大的干扰

姿态的不对齐也会是的全局特征无法匹配

全局特征是指每一张行人图片的全局信息进行一个特征抽取，这个全局特征没有任何的空间信息

基于深度学习的行人重识别

局部特征（Global features）

- ★★★ 切片（Stride）
- ★★★ 姿态（Pose）
- ★★ 分割（Segmentation）
- ★ 网格（Grid）

切片

姿态

分割

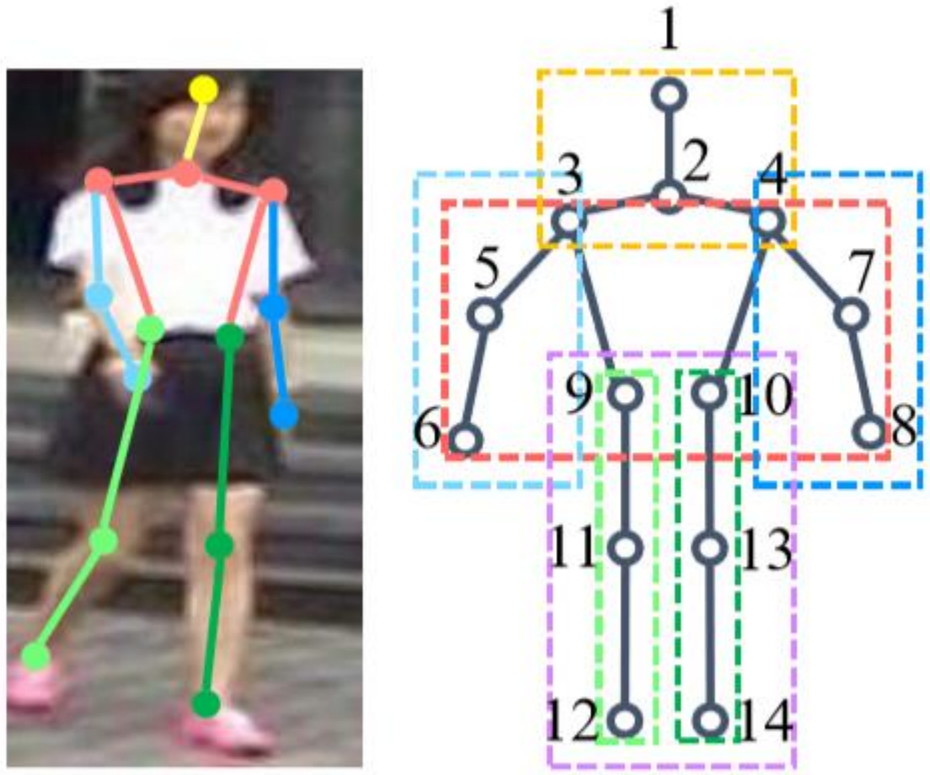
网格

局部特征是指对图像中的某一个区域进行特征提取，最后将多个局部特征融合起来作为最终特征

AI300学院

局部特征

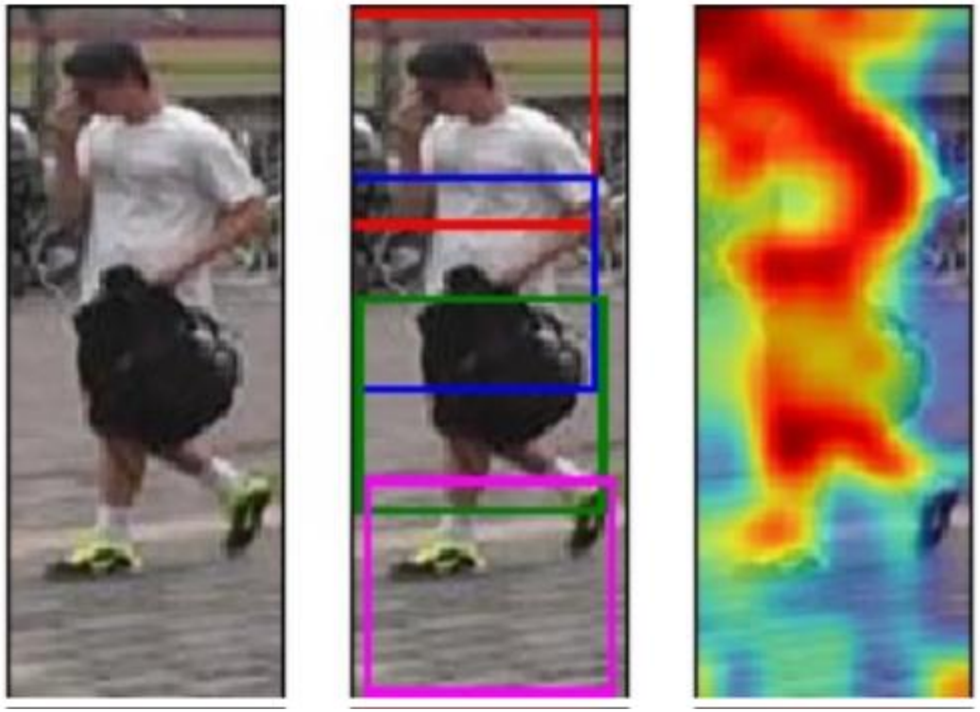
一些概念——姿态



- 通常一个行人会定义14个姿态点（pose/keypoint）
- 两个相邻的姿态点相连形成骨架（skeleton）
- 常用的姿态点估计模型包括：Hourglass、OpenPose、CPM、AlphaPose

局部特征

一些概念——Part&Attention

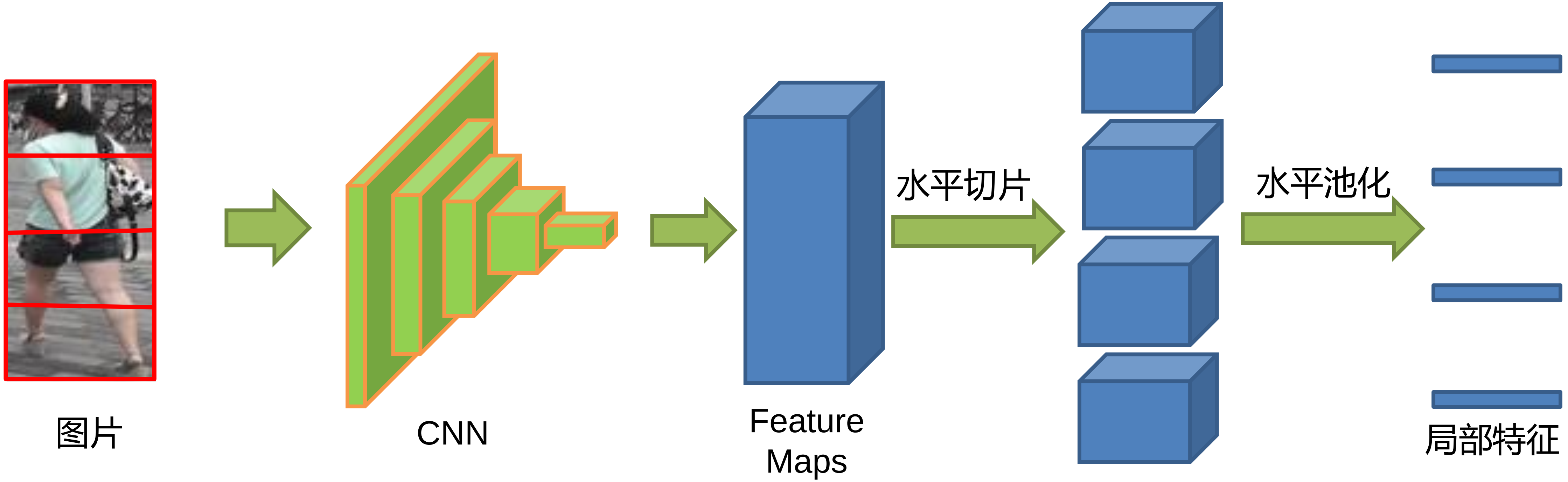


原图　　Part　　Attention

- Part：是指通过一定规则（例如姿态点信息）手工设置的一些矩形框区域
- Attention：是指（在一定的约束条件下）网络自动学习出的比较重要的任意形状区域

局部特征

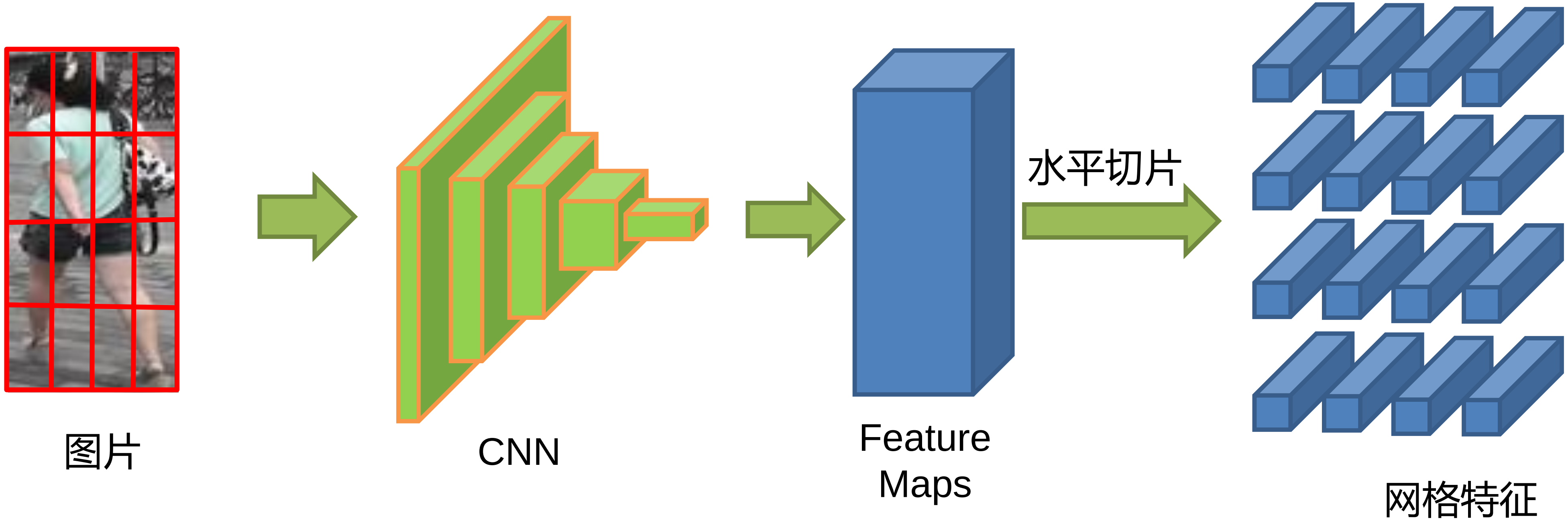
一些概念——水平池化



水平池化是指对于将feature maps进行水平等分，然后再池化得到分块的局部特征

局部特征

一些概念——网格特征



网格特征是指，将H×W×C尺寸的feature map中每个像素的C维特征作为一个网格特征，最终共有H×W个网格特征向量，每个向量的维度为通道数C

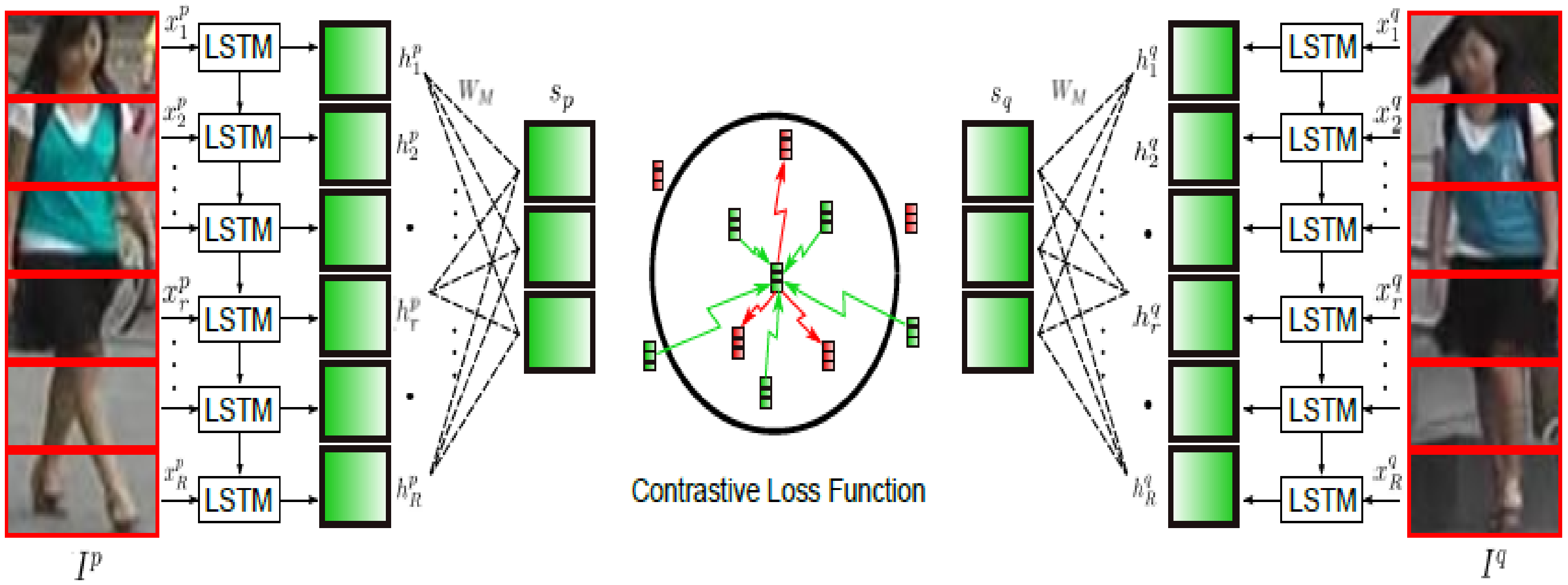
水平切块

代表算法

- Gate Siamese
- AlignedReID
- PCB
- ICNN
- SCPNet

水平切块

Gate Siamese



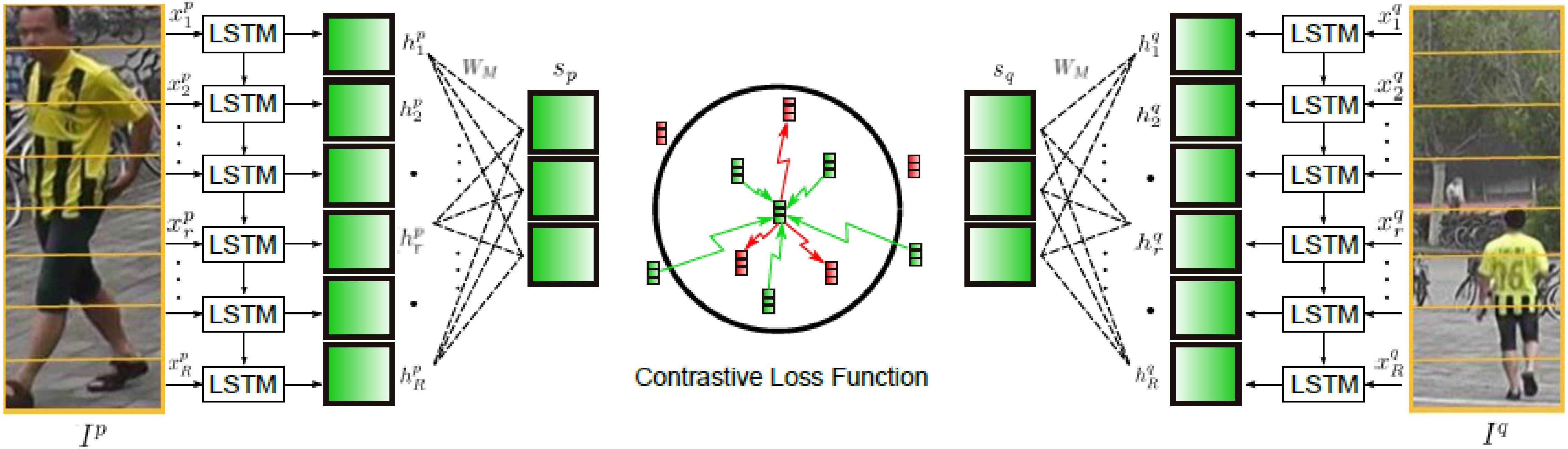
- 每一块图像经过CNN网络得到特征，局部特征按顺序输入到LSTM网络，自动表达为图像最终的特征
- 利用对比损失训练网络
- 水平切块比较早期的工作，目前已经很少使用

[1] Qiqi Xiao, Kelei Cao, Haonan Chen, Fangyue Peng, Chi Zhang. Cross domain knowledge transfer for person re-identification[J]. arXiv preprint arXiv:1611.06026, 2016.
[2] Rahul Rama Varior, Bing Shuai, Jiwen Lu, Dong Xu, Gang Wang. A siamese long short-term memory architecture for human re-identification[C]//European Conference on Computer Vision. Springer, 2016:135–153.

AI300学院

水平切块

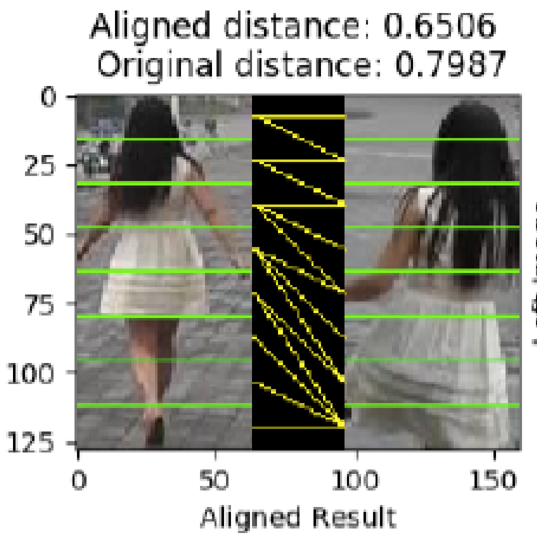
水平切块的不对齐

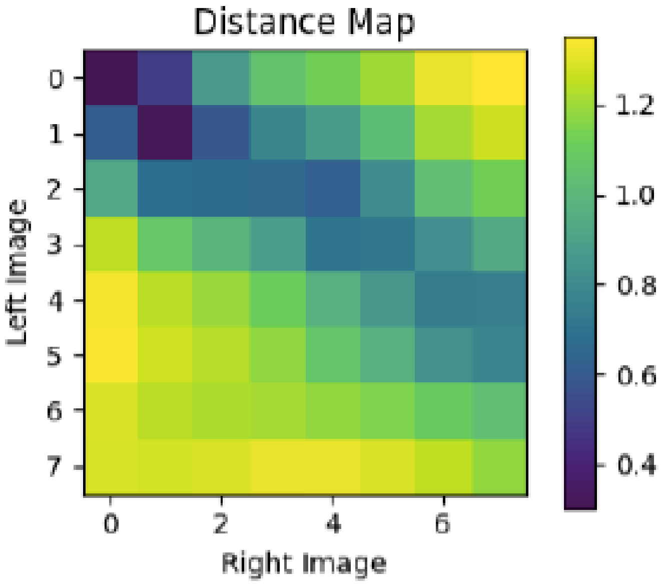


怎么解决姿态不对齐的问题?

水平切块

AlignedReID



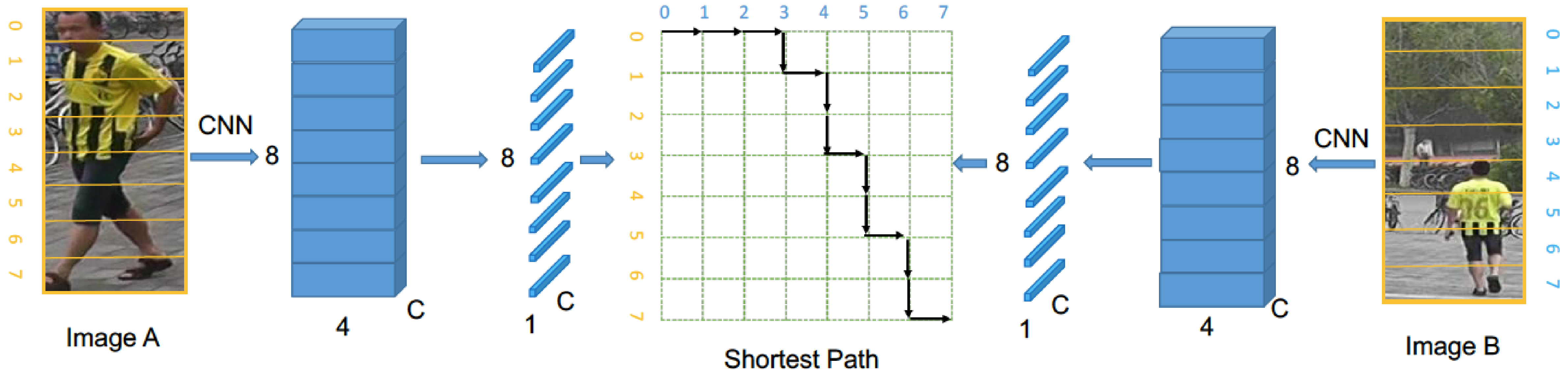


- AlignedReID是旷视科技提出的ReID算法
- 首次在ReID任务上“超越”了人类准确度
- 主要解决姿态不对齐的问题
- 论文中骨架网络为ResNet50
- 在实战教学环节将会实现AlignedReID算法

[1] Zhang X*, Luo H*, Fan X, et al. Alignedreid: Surpassing human-level performance in person re-identification[J]. arXiv preprint arXiv:1711.08184, 2017. (* stands for equal contribution)
[2] Luo H, Jiang W, Zhang, et al. Alignedreid++: Dynamically Matching Local Information for Person Re-Identification. (Under review on Pattern Recognition)

水平切块——AlignedReID

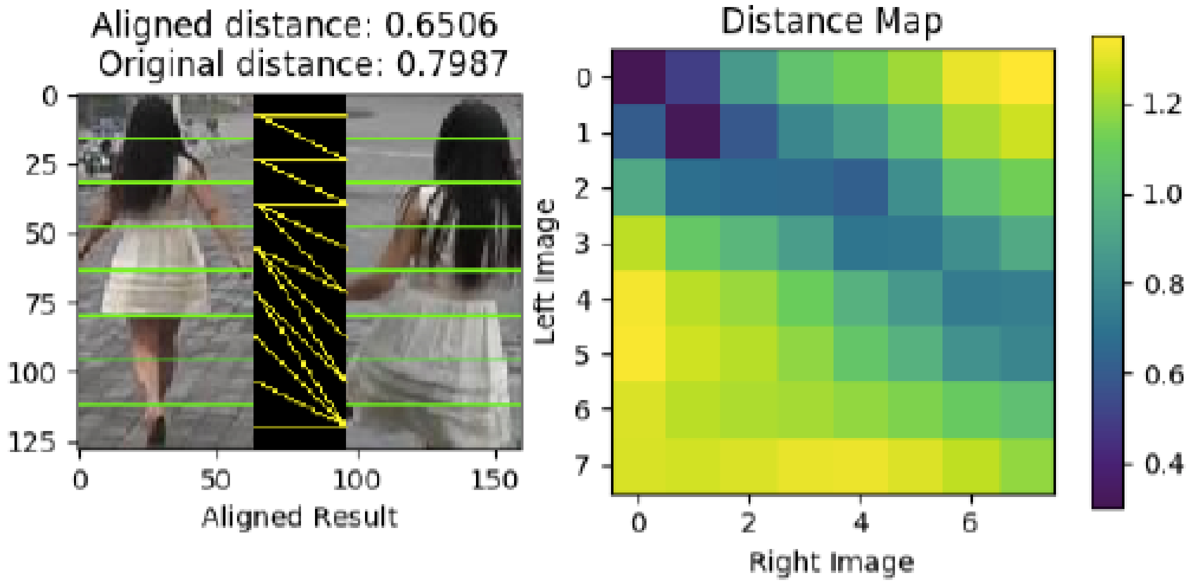
Dynamically Matching Local Information (DMLI)



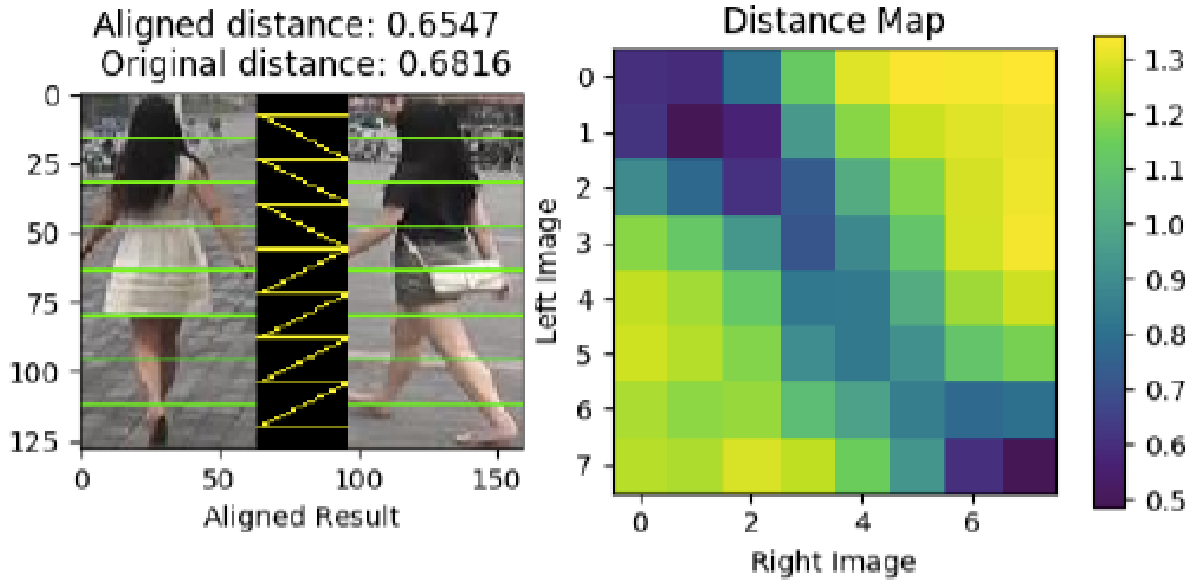
- 加入输入图像为256×128，输出的特征图尺寸为8×4×2048
- 利用水平池化得到8个局部特征，并计算得到一个8×8的距离方阵
- 对齐局部信息不能有跳连（从上到下）
- 利用shortest path 来找到最有的动态连接 (Dynamic Time Warping)

水平切块——AlignedReID

结果示例



(a) Positive pair

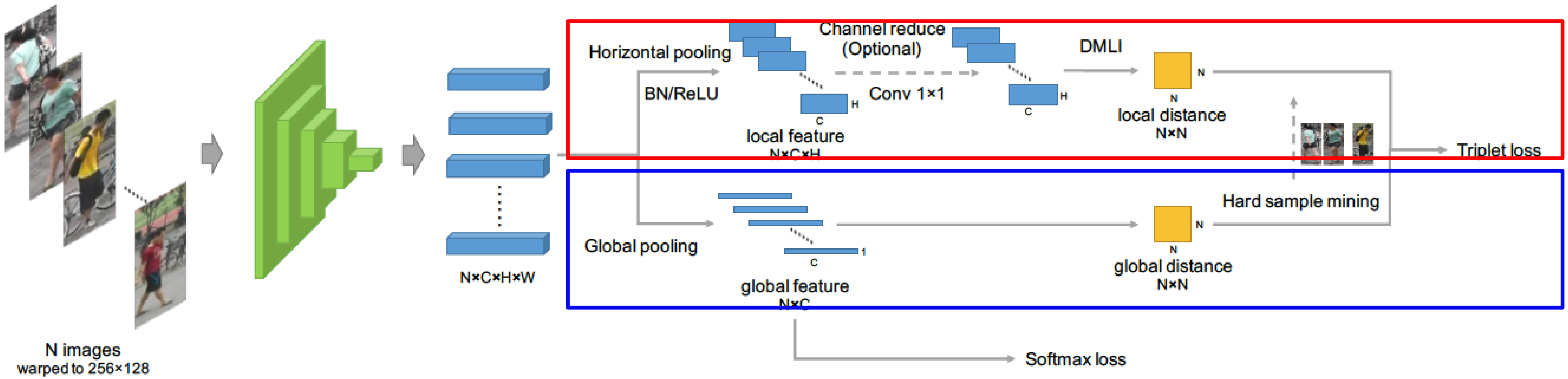


(b) Negative pair

DMLI 可以修正一些检索结果

水平切块——AlignedReID

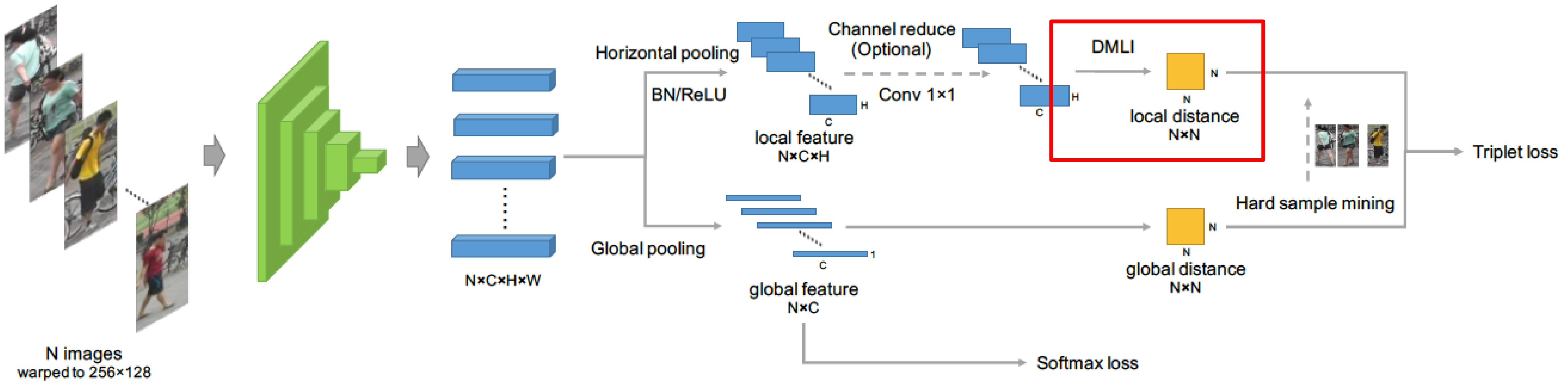
Framework



- 联合全局分支和局部分支一起训练
- 全局分支是一个正常的ReID网络
- 局部分支引入DMLI对齐思想

水平切块——AlignedReID

归一化



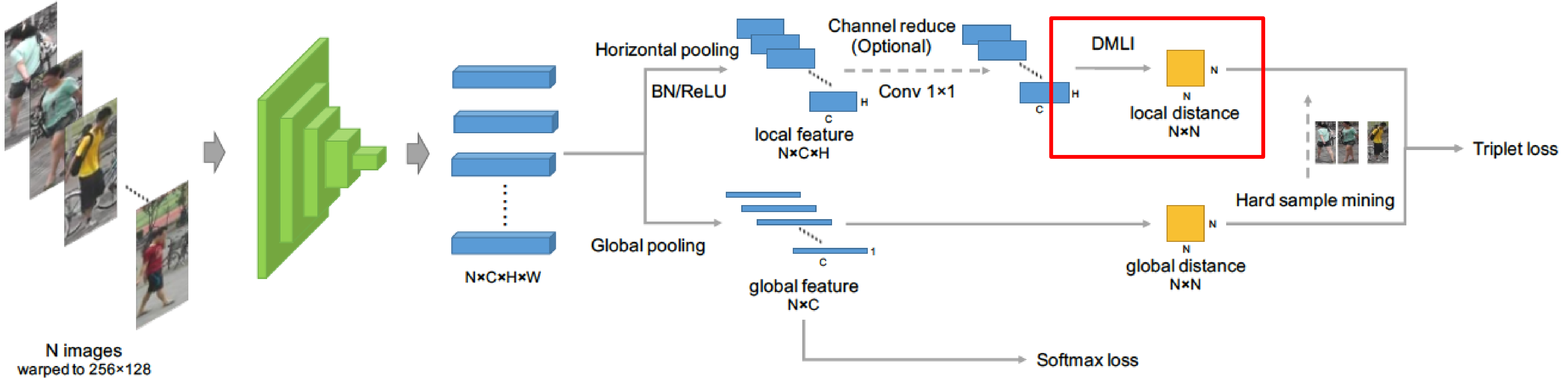
$$d_{i,j} = \frac{e^{||l_A^i - l_B^j||_2} - 1}{e^{||l_A^i - l_B^j||_2} + 1} \quad i,j \in 1,2,3...,H \Rightarrow d_{ij} \propto x$$

$$\frac{\partial d_{ij}}{\partial x} = \frac{2}{e^x + \frac{1}{e^x} + 2} \quad e^x \in [1, e^{\max(x)}] \Rightarrow \frac{\partial d_{ij}}{\partial x} \propto \frac{1}{x}$$

- 计算局部特征两两之间的距离，并归一化
- 使得网络更加关注相似的区域

水平切块——AlignedReID

最短路径距离

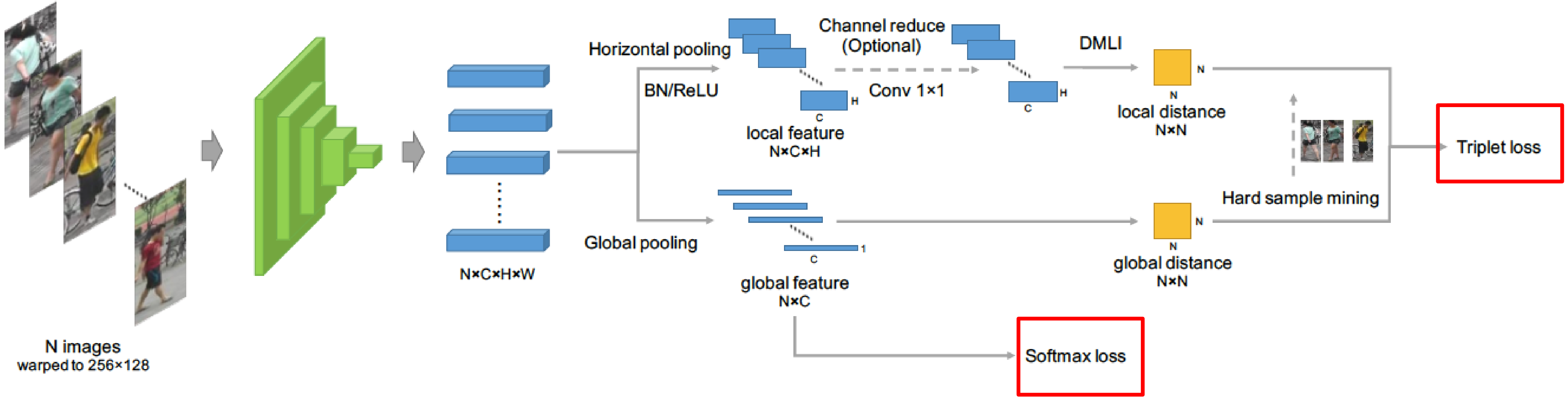


$$S_{i,j} = \begin{cases} d_{i,j} & i=1, j=1 \\ S_{i-1,j} + d_{i,j} & i \neq 1, j=1 \\ S_{i,j-1} + d_{i,j} & i=1, j \neq 1 \\ min(S_{i-1,j}, S_{i,j-1}) + d_{i,j} & i \neq 1, j \neq 1, \end{cases}$$

```
for i in range(m):
  for j in range(n):
    if (i == 0) and (j == 0):
      dist[i][j] = dist_mat[i, j]
    elif (i == 0) and (j > 0):
      dist[i][j] = dist[i][j - 1] + dist_mat[i, j]
    elif (i > 0) and (j == 0):
      dist[i][j] = dist[i - 1][j] + dist_mat[i, j]
    else:
      dist[i][j] = torch.min(dist[i - 1][j], dist[i][j - 1]) + dist_mat[i, j]
dist = dist[-1][-1]
```

水平切块——AlignedReID

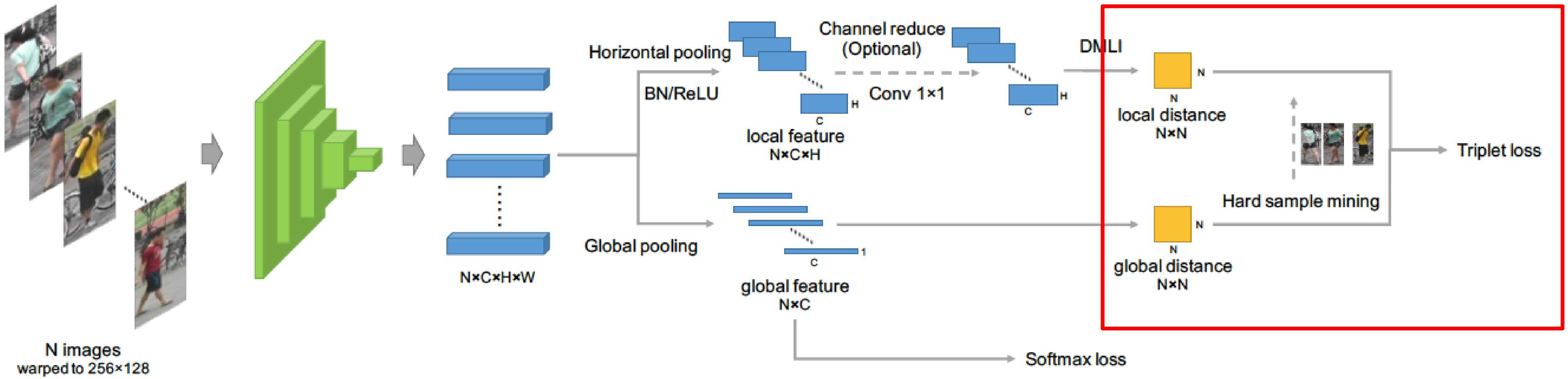
难样本挖掘



- 全局分支计算一个表征学习损失（ID损失）和一个度量学习损失（TriHard损失）
- 局部分支计算一个度量学习损失（TriHard损失）

水平切块——AlignedReID

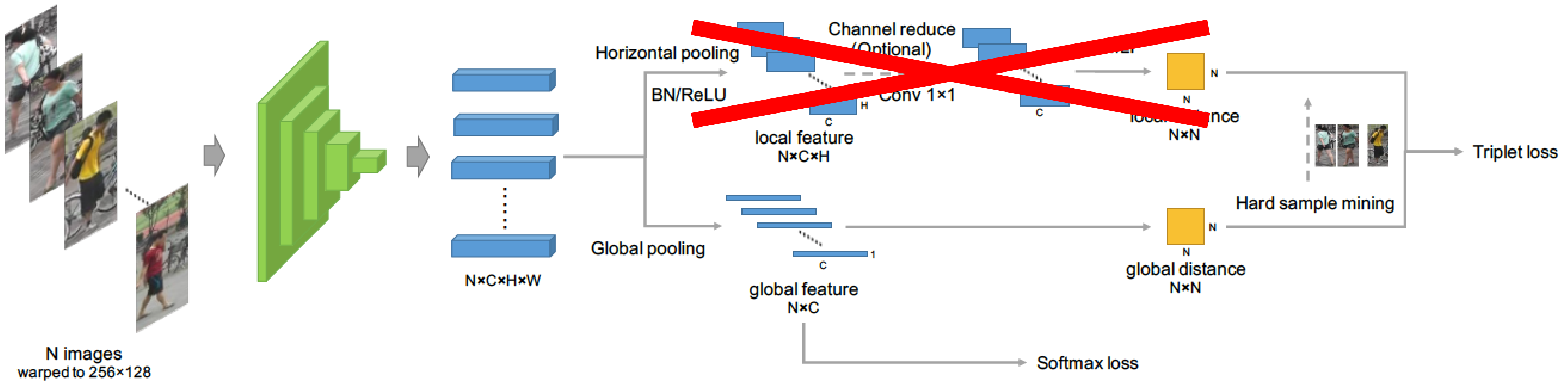
难样本挖掘



- 局部分支和全局分支各自得到一个N×N的距离方阵
- 利用全局分支的距离方阵进行难样本挖掘：速度快、梯度方向一致

水平切块——AlignedReID

测试阶段



- 联合训练使得局部分支会将一部分知识传递给全局分支
- 测试阶段丢弃局部分支，只使用全局特征，不增加额外的计算量
- 测试阶段联合局部分支还能取得进一步的性能提升

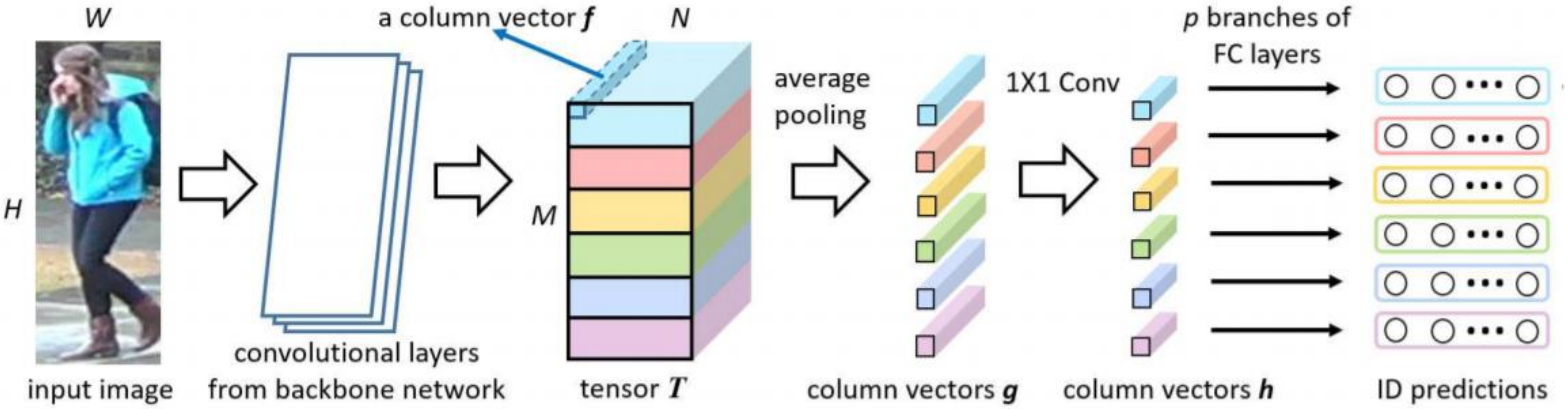
水平切块——AlignedReID

结果

Method		Market1501 mAP	Market1501 r = 1	DukeMTMC mAP	DukeMTMC r = 1	CUHK03 mAP	CUHK03 r=1	MSMT17 mAP	MSMT17 r=1
SVD[38]	ICCV17	62.1	82.3	56.8	76.7	37.3	41.5	-	-
PDC[6]	ICCV17	63.4	84.1	-	-	-	-	29.7	58.0
PCE&ECN[4]	CVPR18	69.0	87.0	62.0	79.8	27.3	30.2	-	-
GLAD[5]	ACMMM17	73.9	89.9	-	-	-	-	34.0	61.4
MGCAM[39]	CVPR18	74.3	83.8	-	-	46.9	46.7	-	-
AWTL[12]	CVPR18	75.7	89.5	63.4	79.8	-	-	-	-
MLFN[40]	CVPR18	74.3	90.0	62.8	81.0	47.8	52.8	-	-
HA-CNN[41]	CVPR18	75.7	91.2	63.8	80.5	38.3	41.7	-	-
Baseline		70.9	86.4	61.0	76.9	56.2	57.3	37.3	61.5
AlignedReID++		77.6	91.0	68.0	80.7	**59.7**	60.9	40.6	66.3
AlignedReID++(LS)		**79.1**	**91.8**	**69.7**	**82.1**	59.6	**61.5**	**43.7**	**69.8**
CamStyle(RK)[42]	CVPR18	71.5	89.5	57.6	78.3	-	-	-	-
DaRe(RK)[43]	CVPR18	82.0	88.3	74.5	80.4	63.6	62.8	-	-
AlignedReID++(RK)		88.5	92.0	81.2	85.2	70.7	67.6	-	-
AlignedReID++(LS,RK)		**89.4**	**92.8**	**82.8**	**86.2**	**70.7**	**67.9**	-	-

水平切块——PCB

Framework



- 输入图像384×192，分成6块；
- 利用ResNet50提取特征，最后24×8的feature map
- 每一行提取一个局部特征，连接一个ReID loss
- 使用的时候把6个局部特征concatenate起来

[1] Sun Y, Zheng L, Yang Y, et al. Beyond Part Models: Person Retrieval with Refined Part Pooling (and A Strong Convolutional Baseline). ECCV, 2018

水平切块——PCB

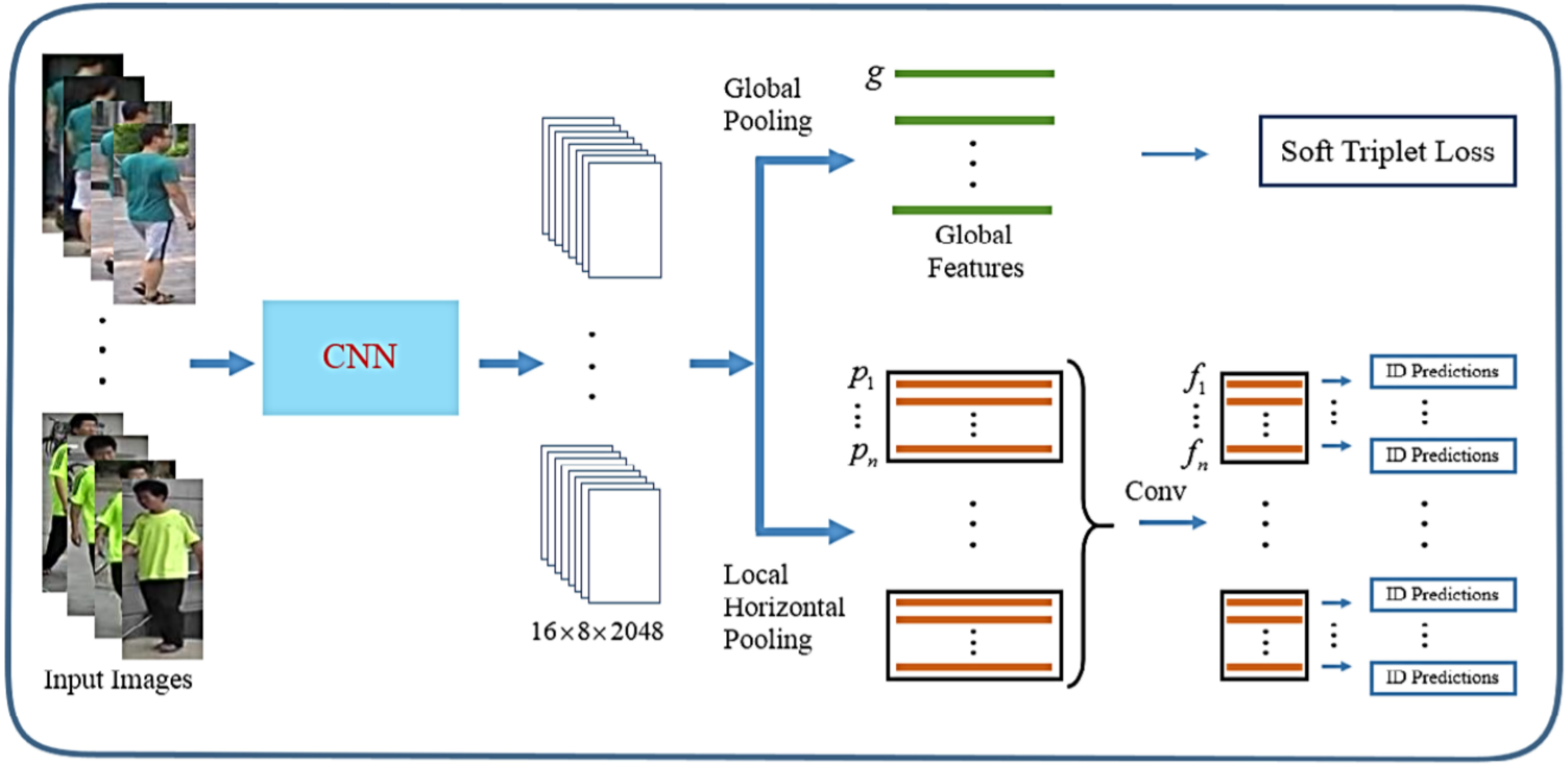
结果

Models	Feature	dim	Market-1501		DukeMTMC-reID		CUHK03	
			R-1	mAP	R-1	mAP	R-1	mAP
IDE	pool5	2048	85.3	68.5	73.2	52.8	43.8	38.9
IDE	FC	256	83.8	67.7	72.4	51.6	43.3	38.3
Variant 1	$\mathcal{G}$	12288	86.7	69.4	73.9	53.2	43.6	38.8
Variant 1	$\mathcal{H}$	1536	85.6	68.3	72.8	52.5	44.1	39.1
Variant 2	$\mathcal{G}$	12288	91.2	75.0	80.2	62.8	52.6	45.8
Variant 2	$\mathcal{H}$	1536	91.0	75.3	80.0	62.6	54.0	47.2
PCB	$\mathcal{G}$	12288	92.3	77.4	81.7	66.1	59.7	53.2
PCB	$\mathcal{H}$	1536	92.4	77.3	81.9	65.3	61.3	54.2
PCB+RPP	$\mathcal{G}$	12288	**93.8**	**81.6**	**83.3**	**69.2**	62.8	56.7
PCB+RPP	$\mathcal{H}$	1536	93.1	81.0	82.9	68.5	**63.7**	**57.5**

- Variant 1 平均所有局部特征，算一个ID loss
- Variant 2 共享FC层参数，算6个ID loss

水平切块——ICNN

Framework

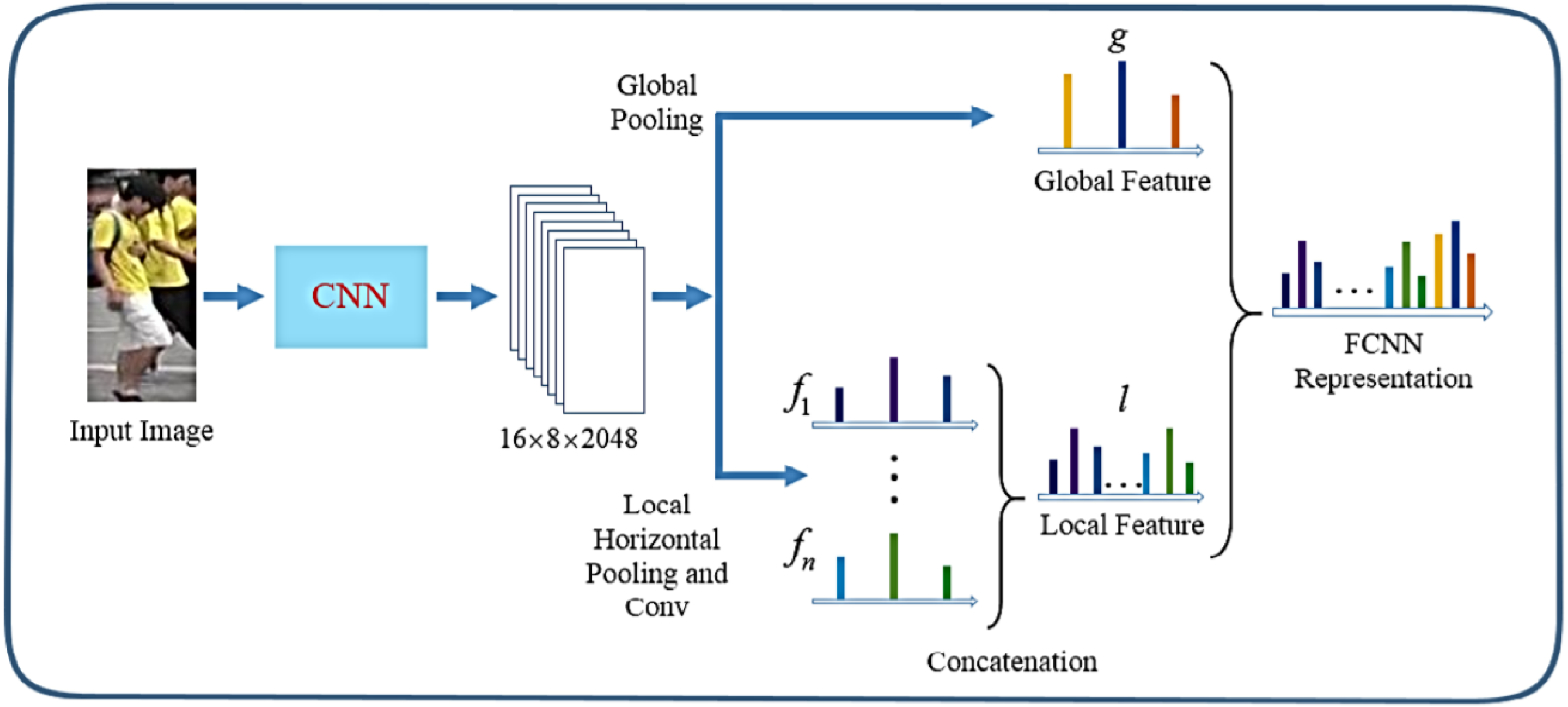


- ICNN ≈ PCB + global branch with triplet loss

[1] Zhang Z, Si T, Liu S. Integration Convolutional Neural Network for Person Re-Identification in Camera Networks[J]. IEEE Access, 2018, 6: 36887-36896.

水平切块——ICNN

Framework



- 最后融合global和local特征，最后一个stride=1

[1] Zhang Z, Si T, Liu S. Integration Convolutional Neural Network for Person Re-Identification in Camera Networks[J]. IEEE Access, 2018, 6: 36887-36896.

水平切块——ICNN

结果

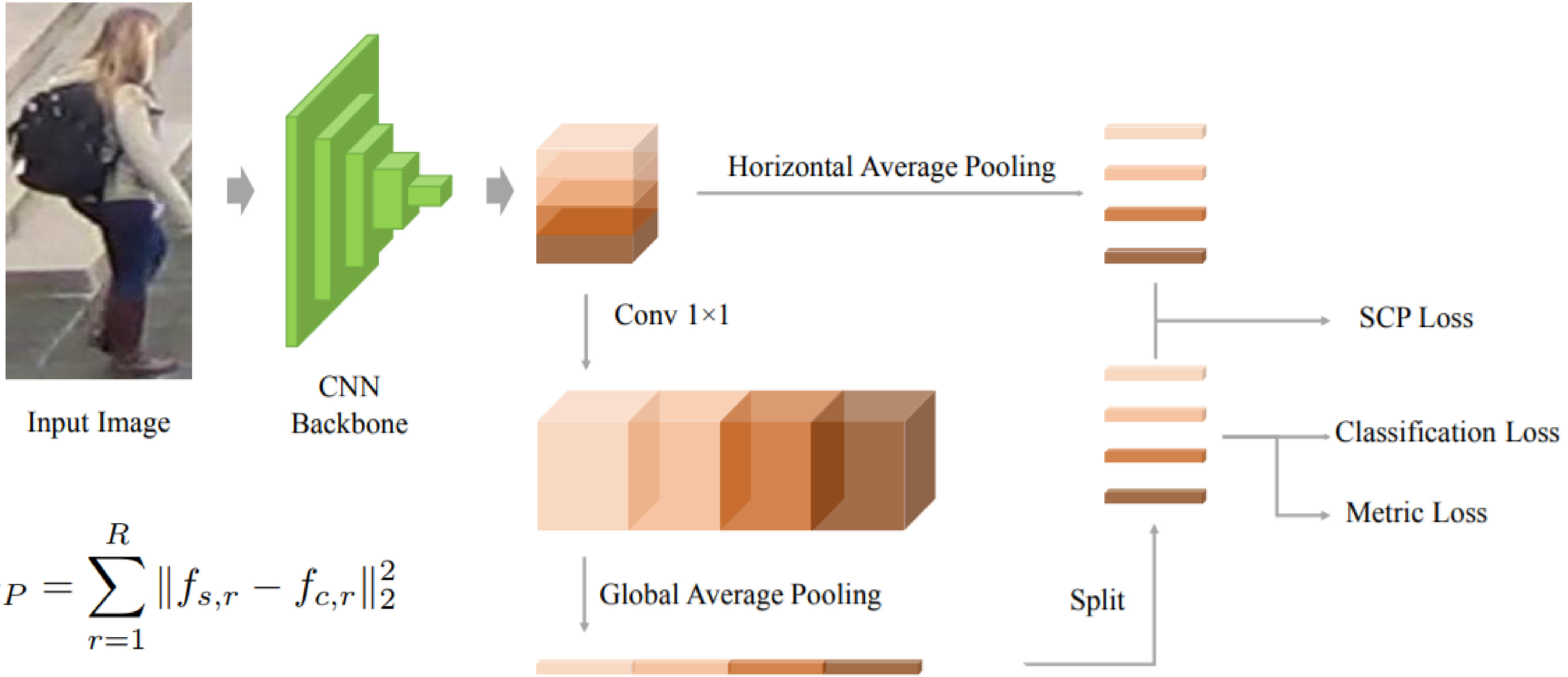
TABLE 2. Comparison of the proposed ICNN with global features, local features and triplet batch hard loss (TBH). Rank-1 (%) and mAP (%) are listed.

Methods	Market1501		CUHK03		DukeMTMC-reID	
	rank-1	mAP	rank-1	mAP	rank-1	mAP
Global features	80.73	63.63	52.36	47.85	75.32	60.78
Local features	89.40	72.06	49.57	45.36	81.51	65.99
TBH	90.78	76.53	60.49	54.79	83.45	70.12
ICNN	**92.13**	**79.01**	**61.36**	**55.78**	**85.32**	**71.20**

[1] Zhang Z, Si T, Liu S. Integration Convolutional Neural Network for Person Re-Identification in Camera Networks[J]. IEEE Access, 2018, 6: 36887-36896.

水平切块——SCPNet

SCPNet



$$L_{SCP} = \sum_{r=1}^{R} \|f_{s,r} - f_{c,r}\|_2^2$$

- 利用spatial part特征连监督channel group特征，将local feature传给global feature

[1] Fan X, Luo H, Jiang W, et al. SCPNet: Spatial-Channel Parallelism Network for Joint Holistic and Partial Person Re-Identification. ACCV, 2018

水平切块——SCPNet

SCPNet

Settings		Results (r=1), $R=4, \lambda=10$					
Model	Branch	Market-1501	DukeMTMC-reID	CUHK03	CUHK-SYSU	Partial ReID	Partial-iLIDs
Baseline	Local	90.7	77.5	90.3	94.5	61.0	83.2
	Global	92.0	80.2	89.6	93.1	60.0	78.2
	Concate	92.8	81.7	92.7	93.6	61.0	84.9
SCPNet		**94.1**	**84.8**	**93.3**	**94.6**	**68.3**	84.9

- 在全身和半身的ReID数据集上均取得了提高

[1] Fan X, Luo H, Jiang W, et al. SCPNet: Spatial-Channel Parallelism Network for Joint Holistic and Partial Person Re-Identification. ACCV, 2018

水平切块

总结

- 将图像进行水平方向的等分，每一个水平切块通过水平池化提取一个特征
- Gate Siamese和AlignedReID通过设计规则融合所有的局部特征计算距离
- PCB、ICNN、SCPNet对每一个局部特征计算一个ReID损失，直接将局部特征拼接起来
- 联合局部特征和全局特征往往能够得到更好的结果

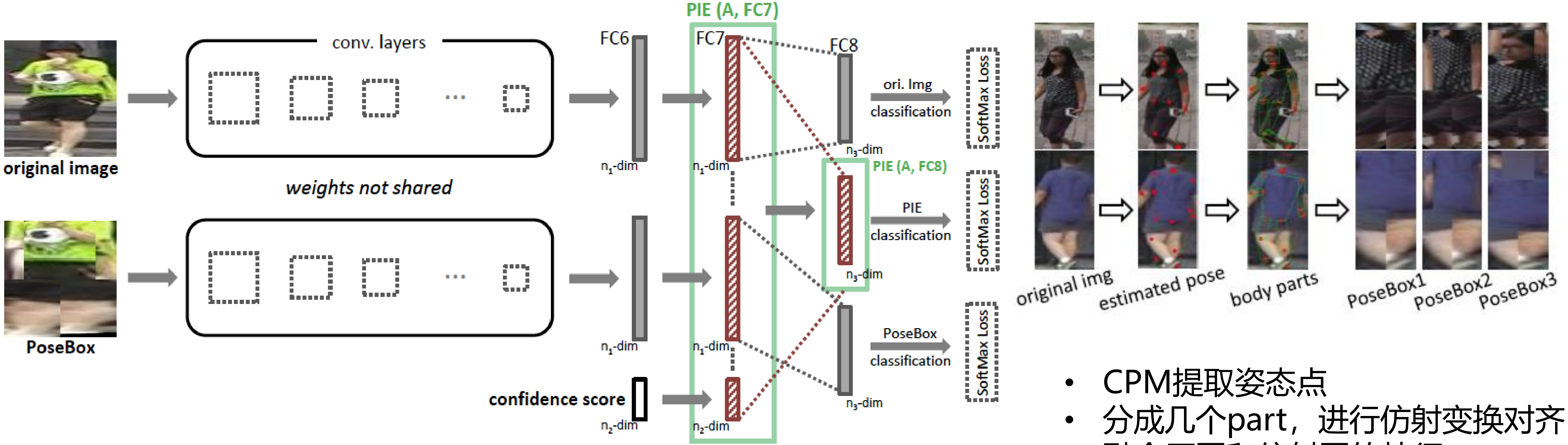
姿态信息

代表算法

- PIE
- Spindle Net
- PDC
- GLAD
- PABP

姿态信息

PIE

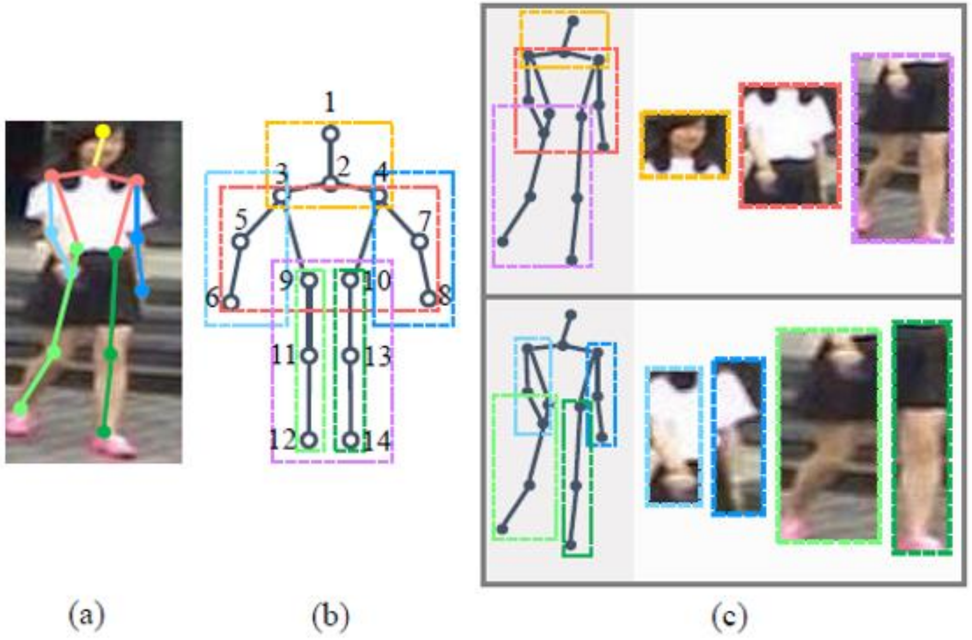


- CPM提取姿态点
- 分成几个part，进行仿射变换对齐
- 融合原图和仿射图的特征
- 采用ID损失训练网络

Liang Zheng, Yujia Huang, Huchuan Lu, Yi Yang. Pose invariant embedding for deep person reidentification[J]. arXiv preprint arXiv:1701.07732, 2017.

姿态信息

Spindle Net

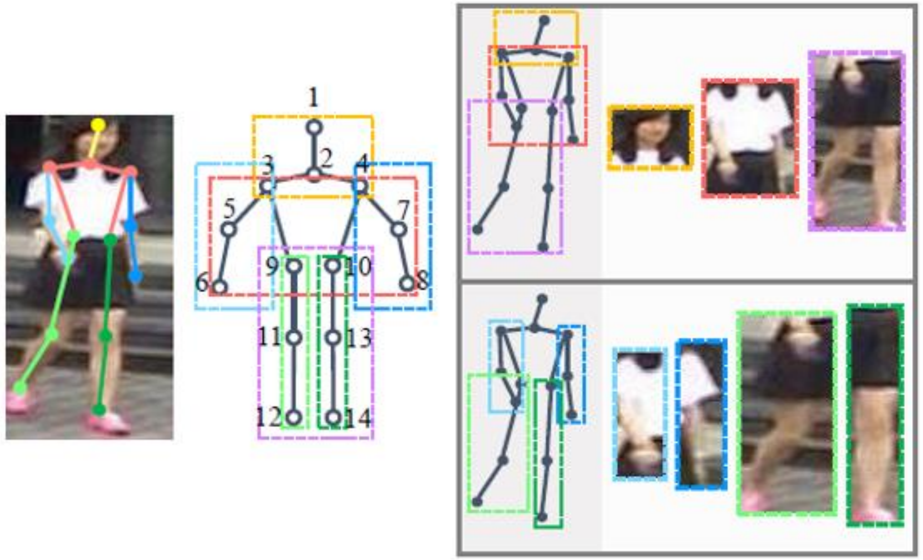


- 利用姿态点估计模型得到14个姿态点
- 分成七个part：
  - 头，上身，下身；
  - 左手，右手，左腿，右腿

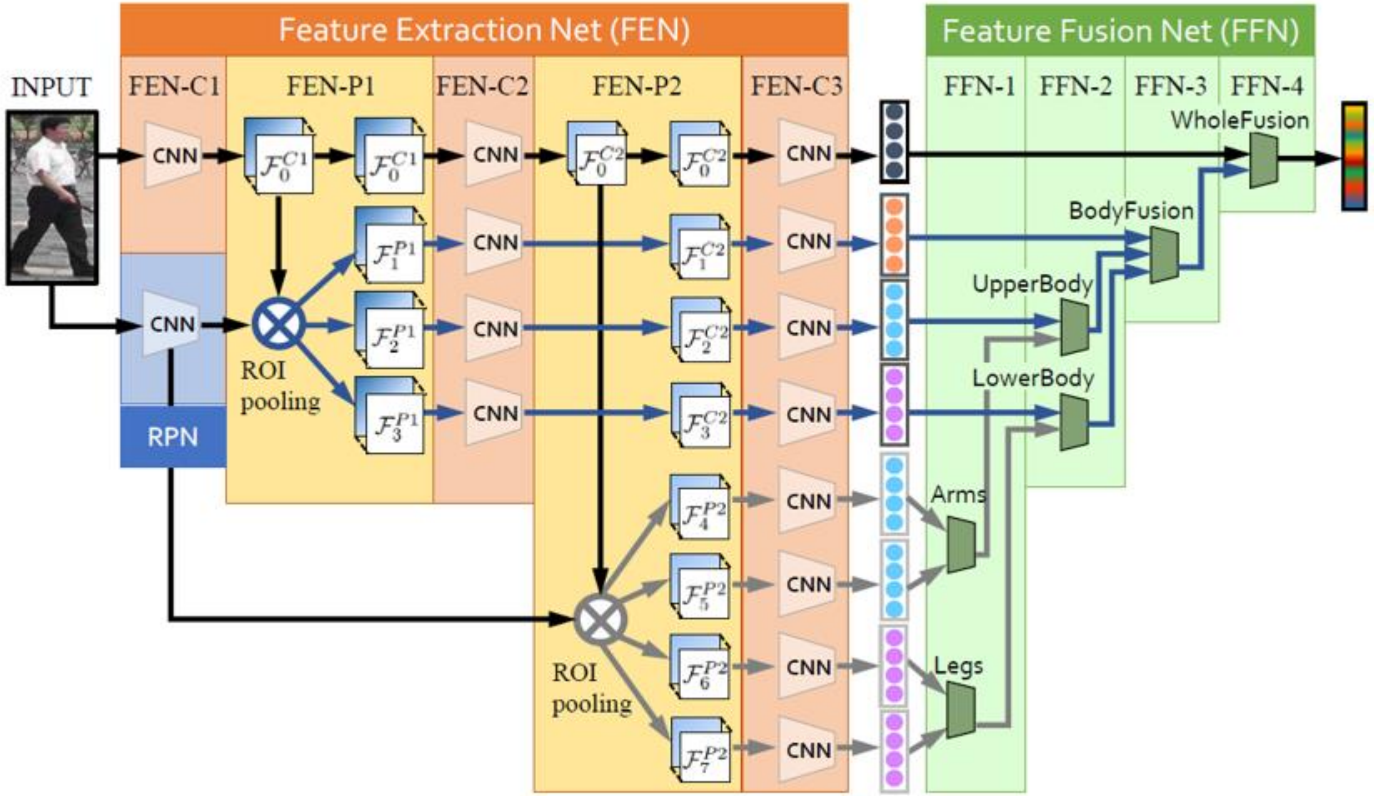
Haiyu Zhao, Maoqing Tian, Shuyang Sun, Jing Shao, Junjie Yan, Shuai Yi, Xiaogang Wang, Xiaoou Tang. Spindle net: Person re-identification with human body region guided feature decomposition and fusion[C]. CVPR, 2017.

姿态信息

Spindle Net



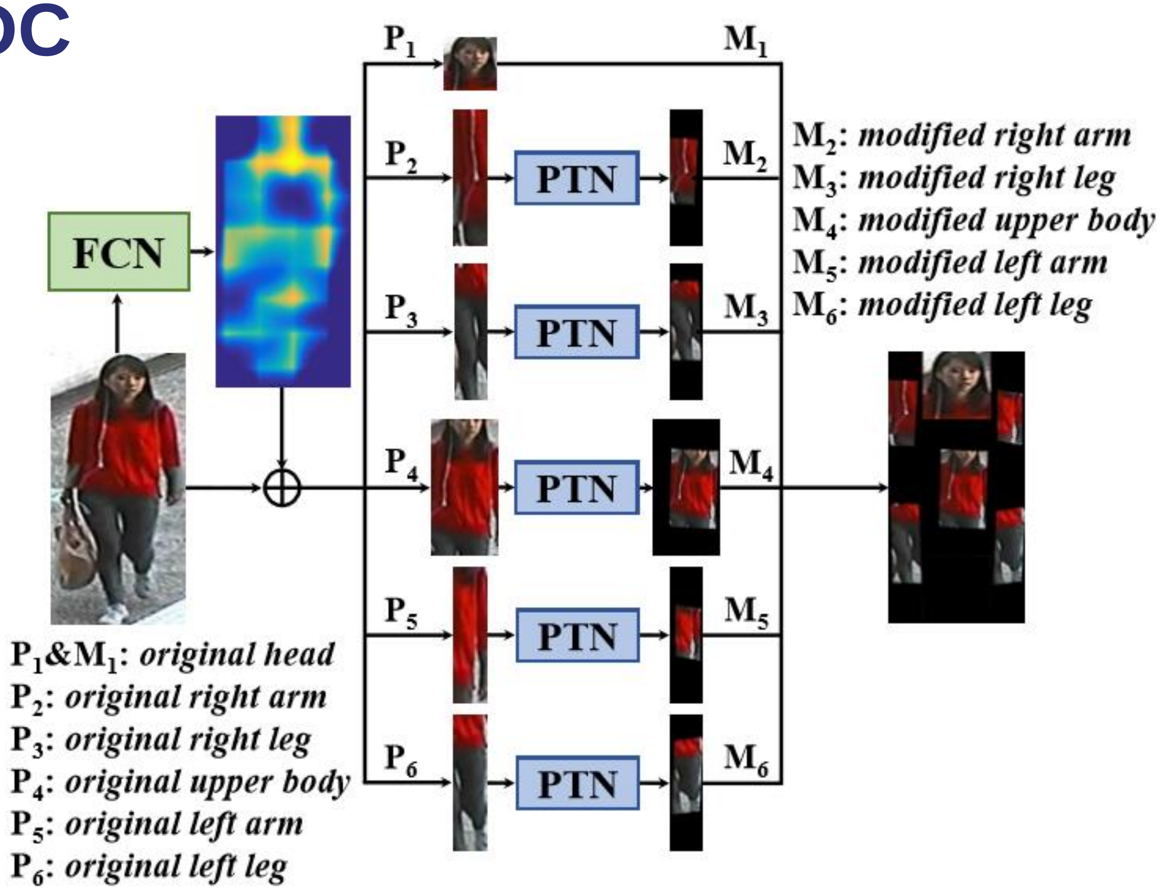
(a) (b) (c)



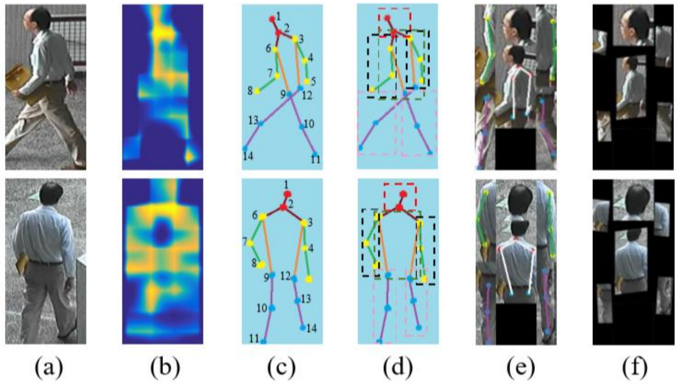
- FEN网络提取特征，FFN网络层次性地融合特征

姿态信息

PDC



- 利用姿态点信息分割为六个part



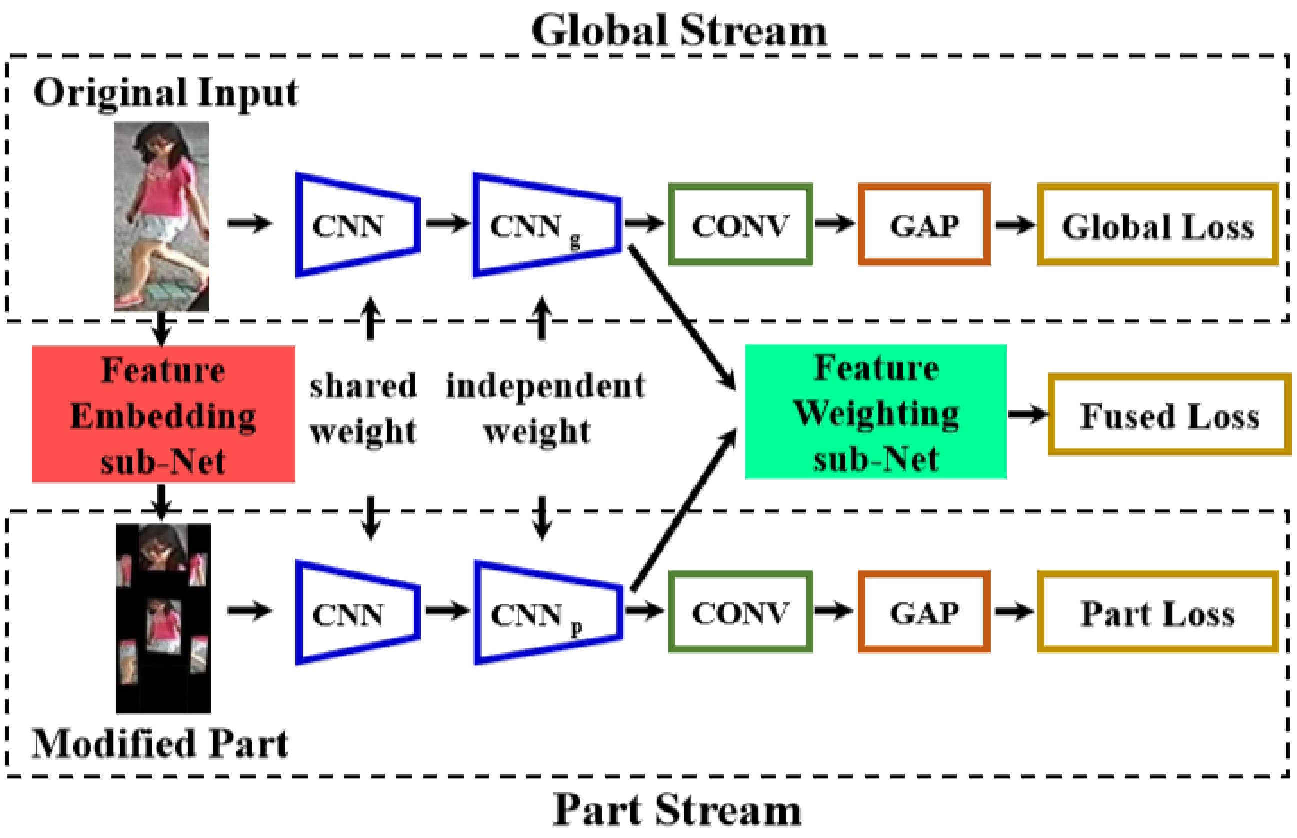
- 改进STN网络为PTN网络，学习仿射变换参数得到modified part image

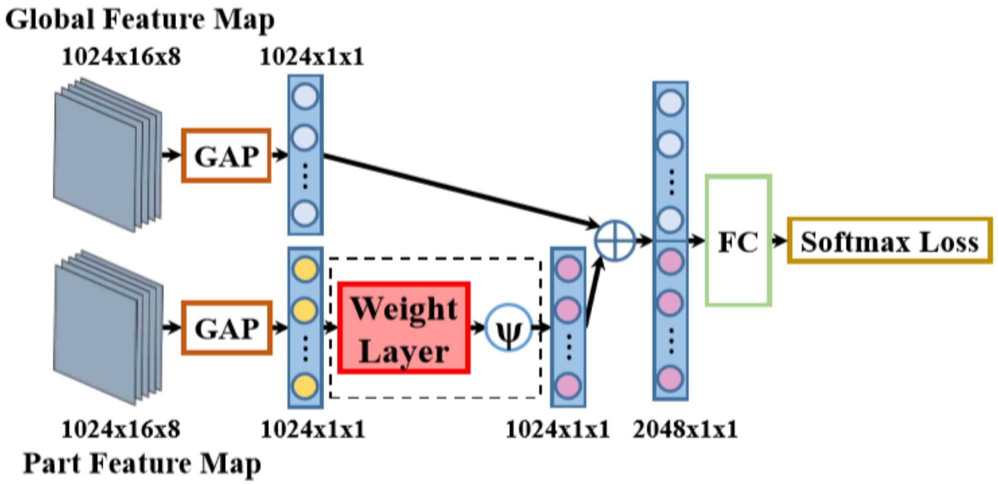
$$\begin{pmatrix} x^s \\ y^s \end{pmatrix} = A_\theta \begin{pmatrix} x^t \\ y^t \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \\ \theta_4 & \theta_5 & \theta_6 \end{bmatrix} \begin{pmatrix} x^t \\ y^t \\ 1 \end{pmatrix}$$

Su C, Li J, Zhang S, et al. Pose-driven deep convolutional model for person re-identification[C]//Computer Vision (ICCV), 2017 IEEE International Conference on. IEEE, 2017: 3980-3989.

姿态信息

PDC





- 融合全局特征和局部特征

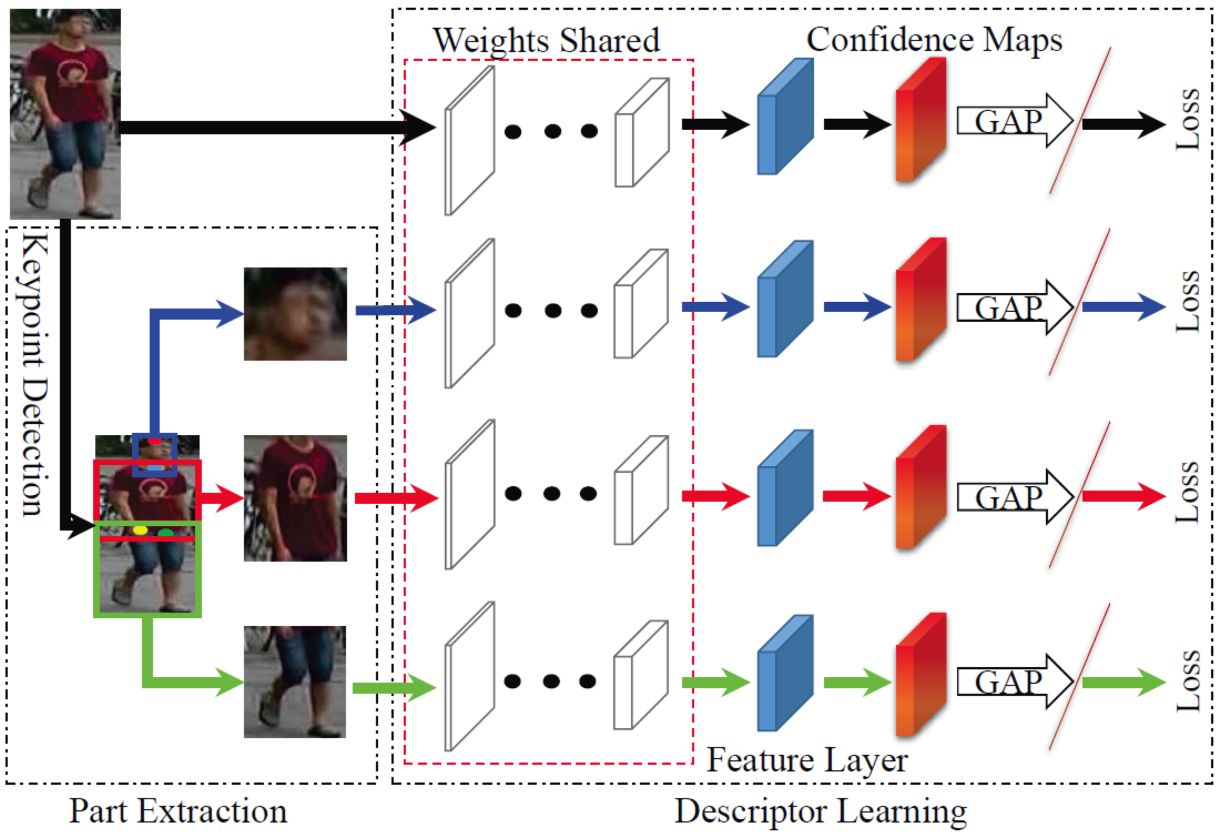
$$F_{fusion} = [F_{global}, tanh(F_{part} \odot W + B)]$$

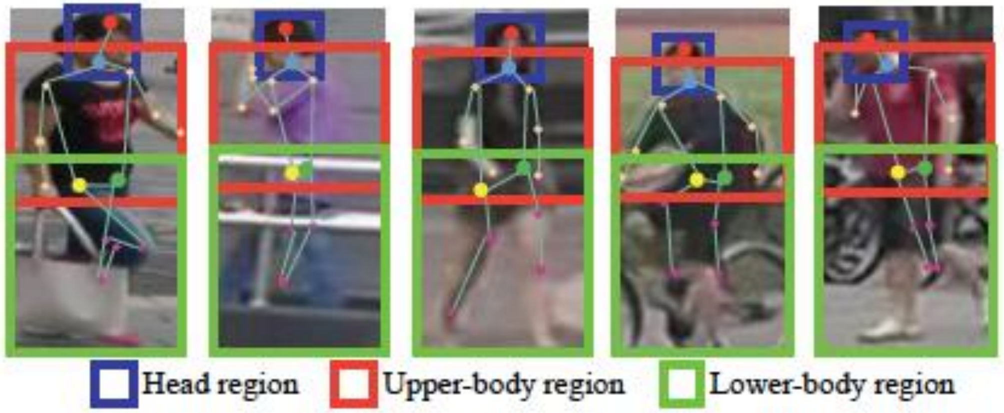
- 计算三个ReID损失
- 浅层网络共享，高层网络独立

Su C, Li J, Zhang S, et al. Pose-driven deep convolutional model for person re-identification[C]//Computer Vision (ICCV), 2017 IEEE International Conference on. IEEE, 2017: 3980-3989.

姿态信息

GLAD





- 分为头，上身，下身三个part
- 融合全局特征和三个part的特征

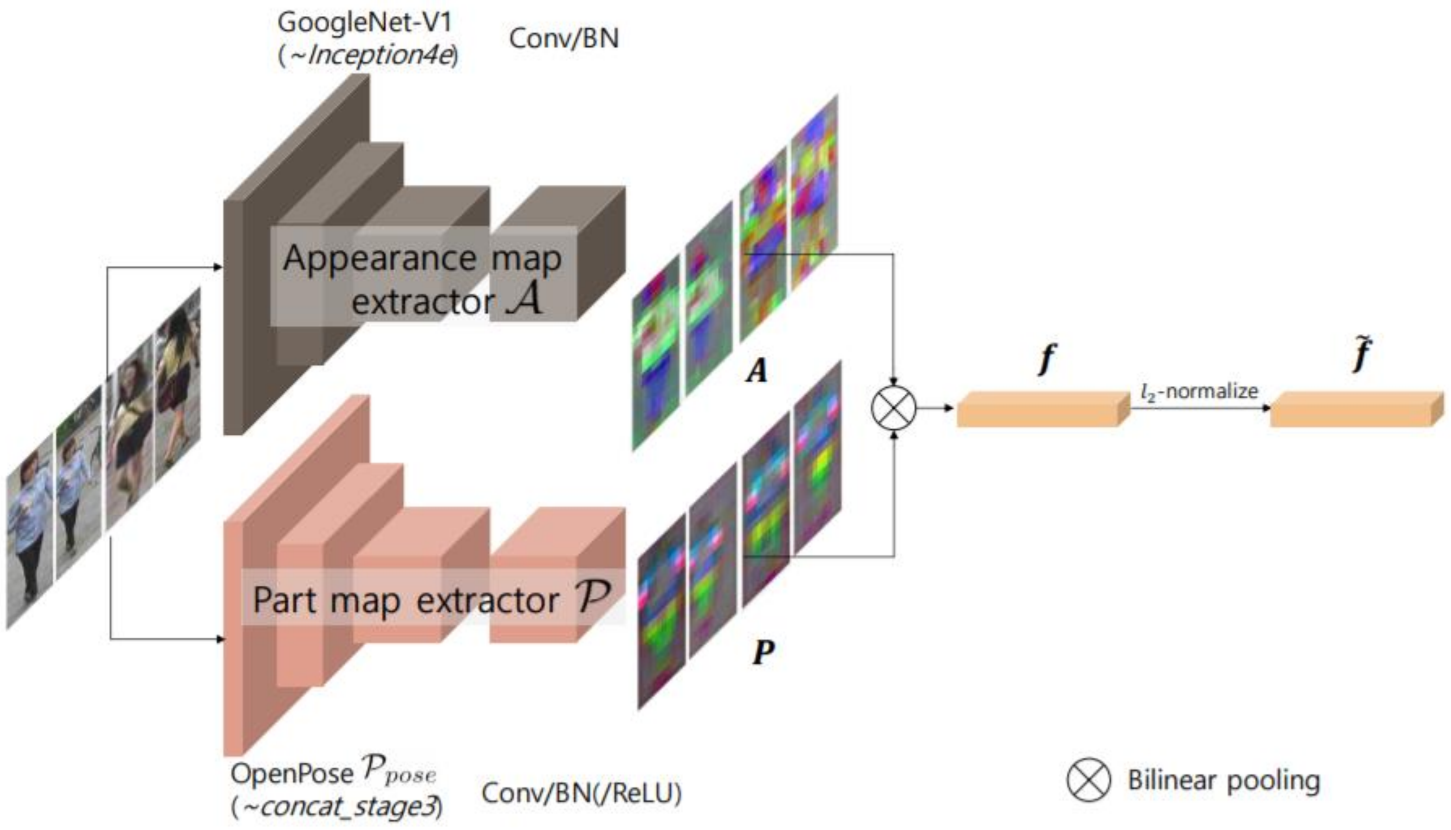
$$\mathbf{f}^{GLAD} = [\mathbf{f}^{G}; \mathbf{f}^{h}; \mathbf{f}^{ub}; \mathbf{f}^{lb}],$$

Longhui Wei, Shiliang Zhang, Hantao Yao, Wen Gao, Qi Tian. Glad: Global-local-alignment descriptor for pedestrian retrieval[J]. arXiv preprint arXiv:1709.04329, 2017.

AI300学院

姿态信息

PABP



- 利用ReID网络提取feature map A
- 利用openpose提取feature map P
- A和P每个对应像素位置的向量进行外积，并向量化

$$\mathbf{f} = \text{pooling}_{xy}\{\mathbf{f}_{xy}\} = \frac{1}{S}\sum_{xy}\mathbf{f}_{xy},$$

$$\mathbf{f}_{xy} = \text{vec}(\mathbf{a}_{xy} \otimes \mathbf{p}_{xy}),$$

$$\text{vec}(\mathbf{a} \otimes \mathbf{p}) = [(p_1\mathbf{a})^\top \ (p_2\mathbf{a})^\top \ \ldots (p_{c_P}\mathbf{a})^\top]^\top$$

- 会激活对应位置的外观特征

Suh Y, Wang J, Tang S, et al. Part-Aligned Bilinear Representations for Person Re-identification. ECCV, 2018.

姿态信息

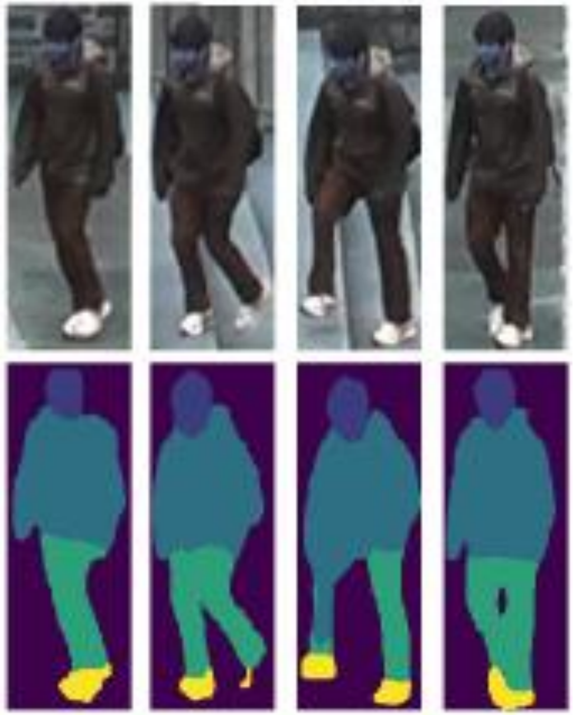
总结

- 利用一个姿态估计模型得到行人的（14个）关键姿态点
- 根据姿态点得到具有语义信息的part区域
- 对于每个part区域提取局部特征
- 联合局部特征和全局特征往往能够得到更好的结果

分割信息

行人分割

粗粒度的分割

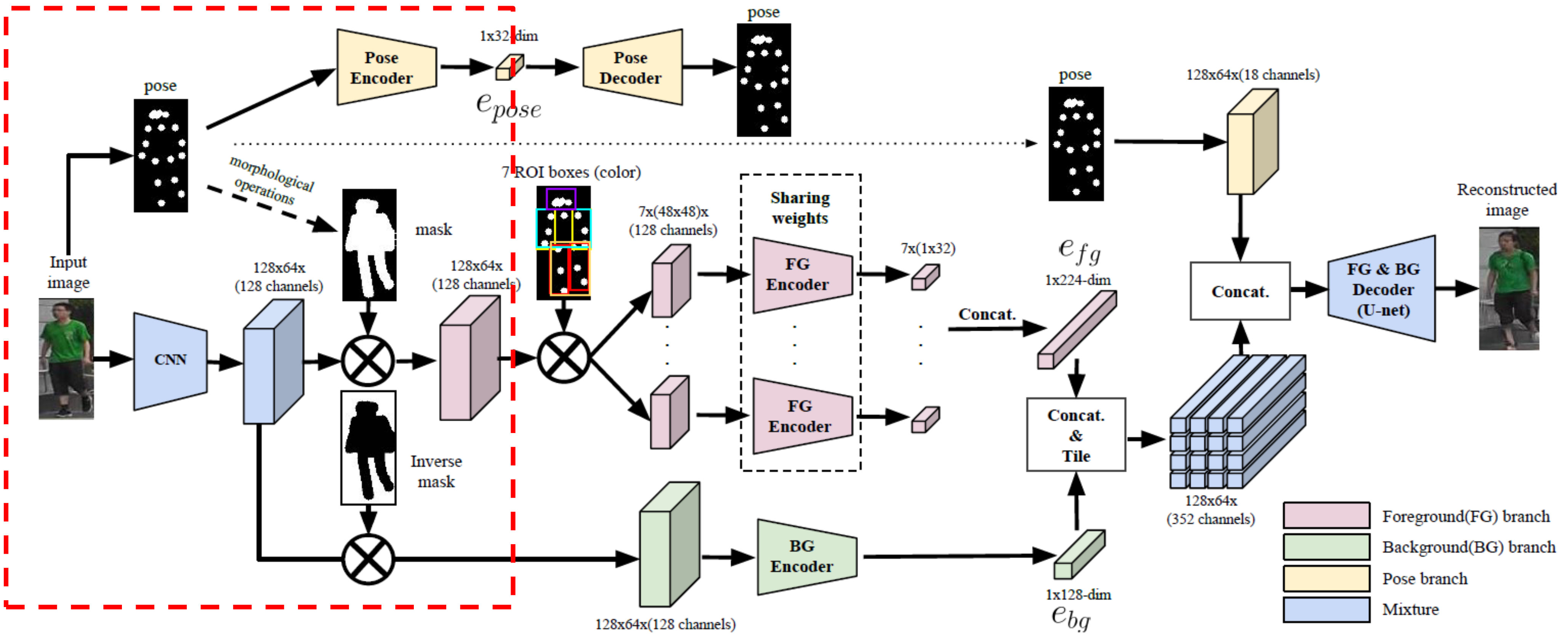


细粒度的分割

- 数据集：
  - COCO
  - LIP
- 分割算法：
  - FCN
  - PSPNet
  - DeepLabV2
  - Mask RCNN

分割信息

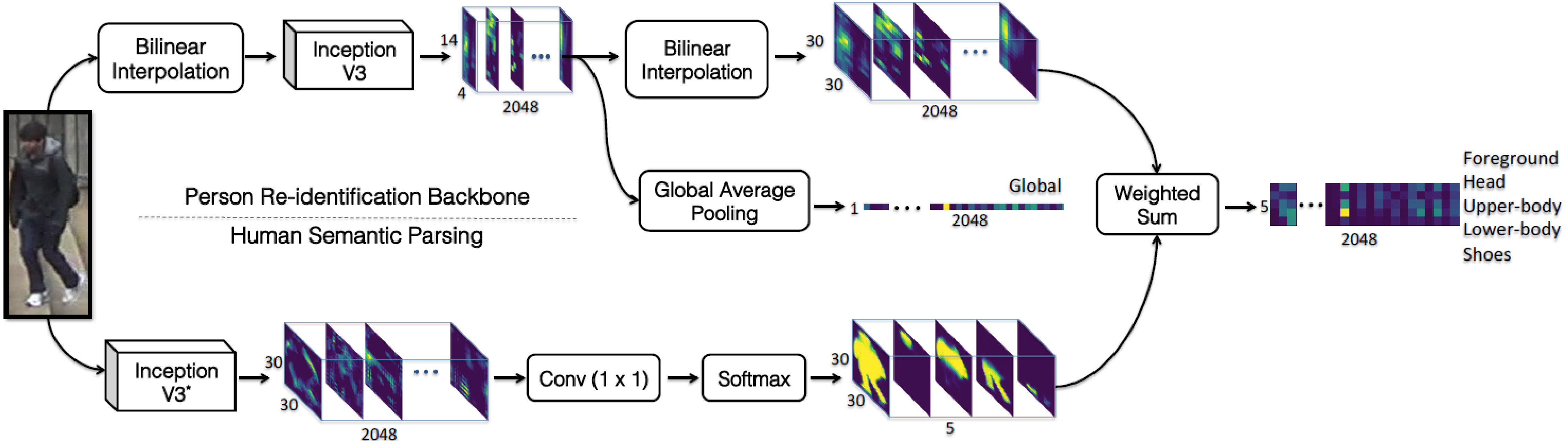
前背景提取



Ma L, Sun Q, Georgoulis S, et al. Disentangled person image generation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 99-108.

分割信息

SPReID



- 一个ReID分支和一个语义分割分支
- 语义分割分支：在LIP数据集上训练
- 一个全局region和四个局部part
- 将语义分割结果作为attention相乘

Kalayeh M M, Basaran E, Gökmen M, et al. Human Semantic Parsing for Person Re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 1062-1071.

分割信息

总结

- 图像语义分割是一种极精细的像素级别part信息
- 图像分割分为粗粒度的行人前景分割和细粒度的肢体语义分割
- 分割结果通常作为图像预处理的Mask或者feature map中的attention相乘
- 目前基于分割的方法没有取得特别广泛的应用

网格特征

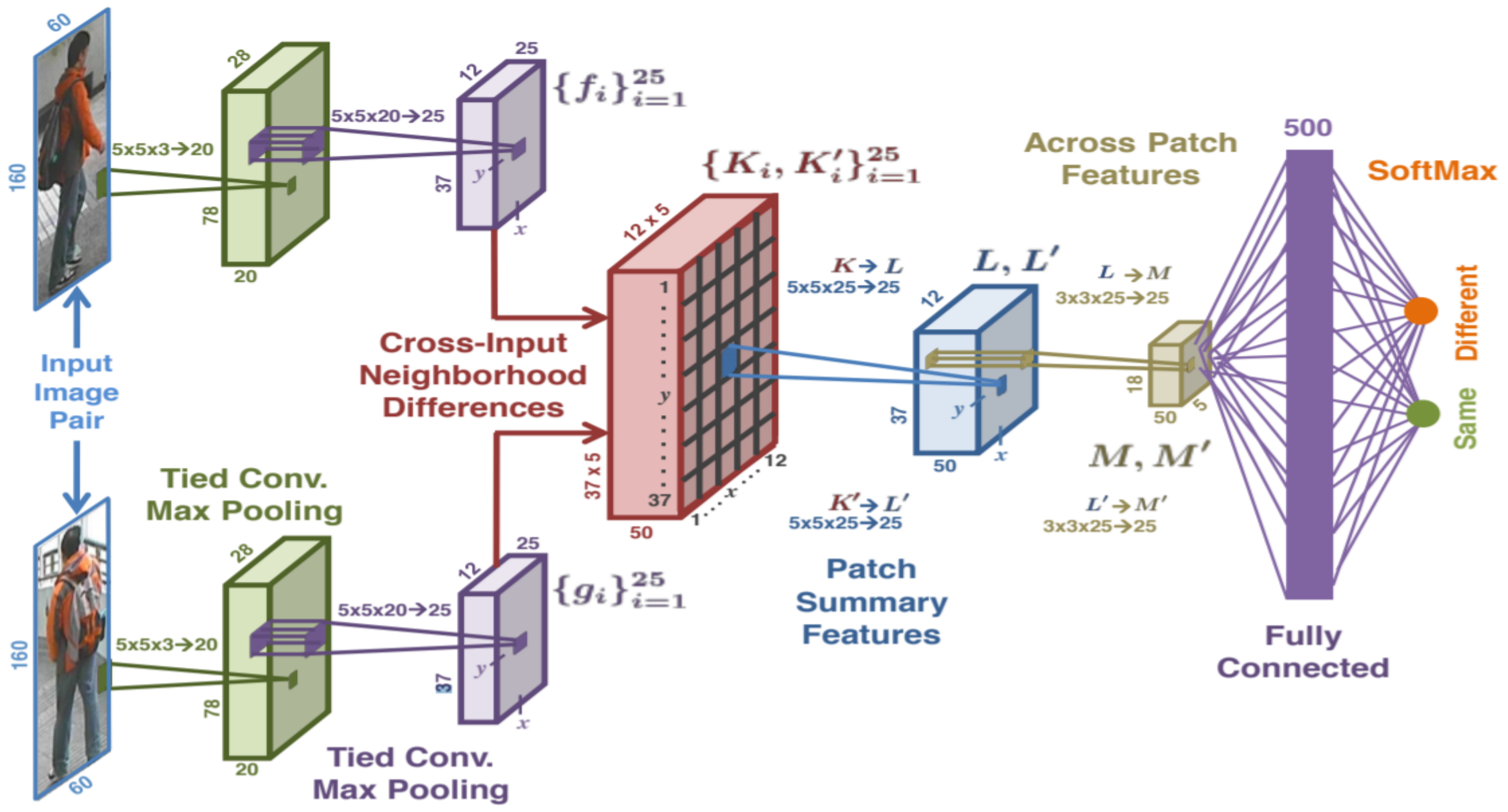
代表算法

- IDLA
- PersonNet
- DSR

AI300学院

网格特征

IDLA



- 骨干网络为Siamese网络
- 计算两幅图5×5网格特征差值

$$K_i(x,y) = f_i(x,y)\mathbb{1}(5,5) - \mathcal{N}[g_i(x,y)]$$

where

$\mathbb{1}(5,5) \in \mathbb{R}^{5\times5}$ is a 5×5 matrix of 1s,

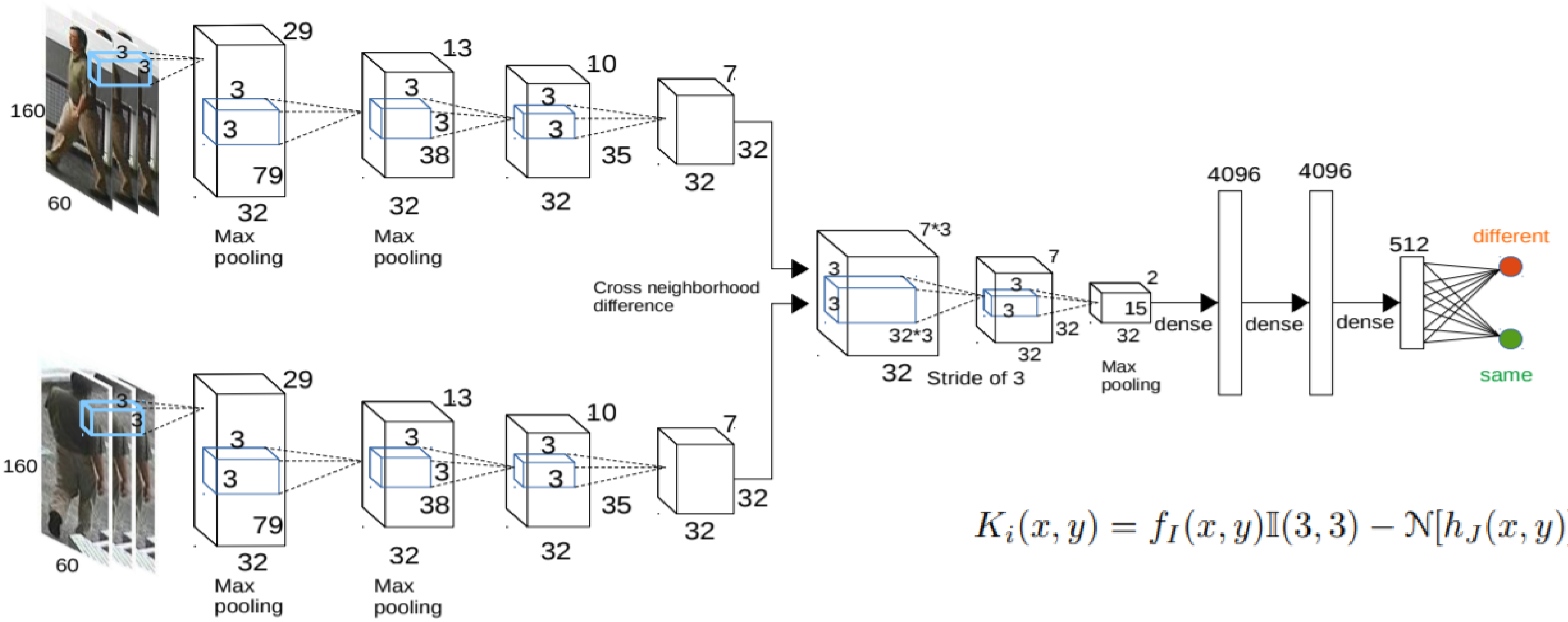
$\mathcal{N}[g_i(x,y)] \in \mathbb{R}^{5\times5}$ is the 5×5 neighborhood of g_i centered at (x,y).

- 交换“主客”分别计算K和K’
- 计算二分类验证损失

Ahmed E, Jones M, Marks T K. An improved deep learning architecture for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3908-3916.

网格特征

PersonNet



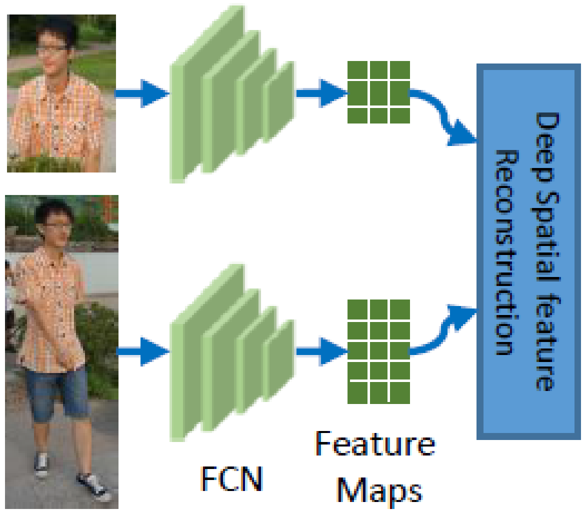
$$K_i(x,y) = f_I(x,y)\mathbb{I}(3,3) - \mathbb{N}[h_J(x,y)]$$

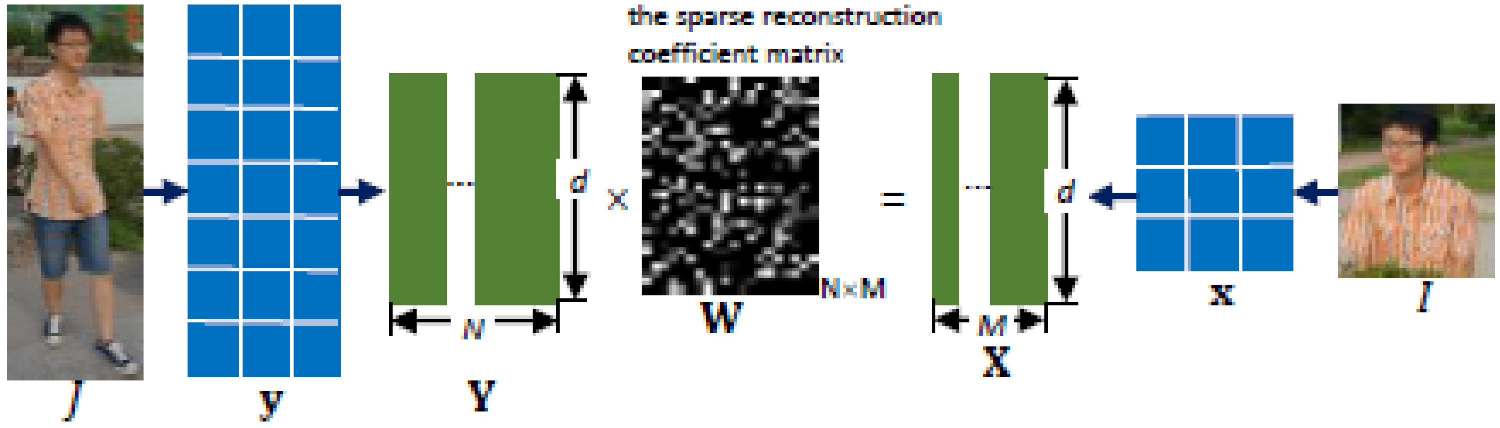
Wu L, Shen C, Hengel A. Personnet: Person re-identification with deep convolutional neural networks[J]. arXiv preprint arXiv:1601.07255, 2016.

AI300学院

网格特征

DSR





$$\mathrm{X} = \{\mathrm{x}_1, \cdots, \mathrm{x}_N\} \in \mathbb{R}^{d \times N}$$

$$\mathrm{Y} = \{\mathrm{y}_1, \cdots, \mathrm{y}_M\} \in \mathbb{R}^{d \times M},$$

$$\min_{\mathrm{w}_n} ||\mathrm{x}_n - \mathrm{Y}\mathrm{w}_n||_2^2 + \beta||\mathrm{w}_n||_1,$$

- 将一副图像的所有网格特征作为一个特征集合
- 对两个特征集合进行稀疏重建得到集合距离

He L, Liang J, Li H, et al. Deep Spatial Feature Reconstruction for Partial Person Re-identification: Alignment-Free Approach[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7073-7082.

网格特征

总结

- 网格特征是比较细粒度的物理区域特征
- 早期工作将网格特征扩展为part特征计算两幅图像的特征图差
- 近期利用网格特征解决partial ReID工作
- 总体而言网格特征并不常用

课后思考

1. 切块类和姿态类ReID方法各自有什么优缺点?

2. 以上大量方法中，哪些思路有利于训练一个高准确高帧率的ReID产品模型?

欢迎关注AI300学院