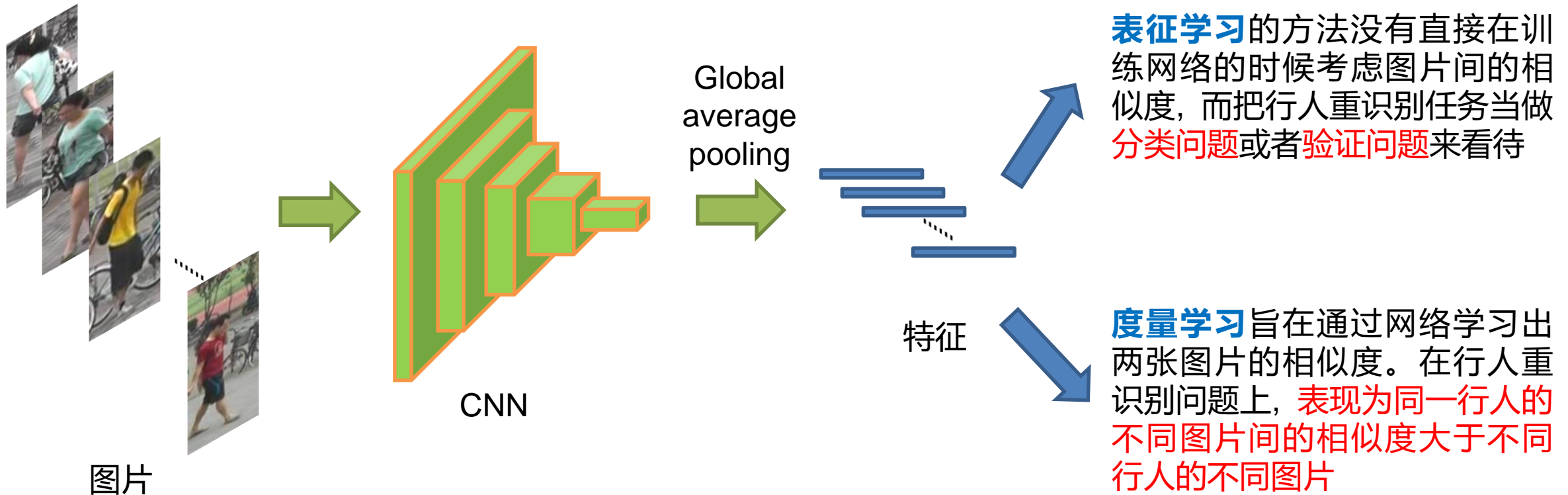


# 行人重识别——表征学习与度量学习

罗浩  
浙江大学

# 基于深度学习的行人重识别

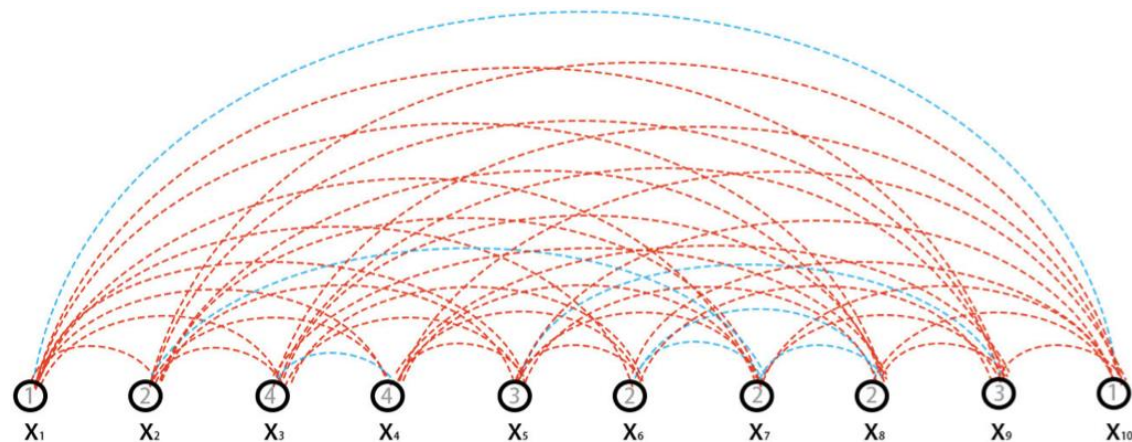
## 根据损失分类



# 表征学习

## 分类损失&验证损失

- 分类损失又称ID损失：利用行人的ID作为训练标签来训练模型，每次只需要输入一张图片

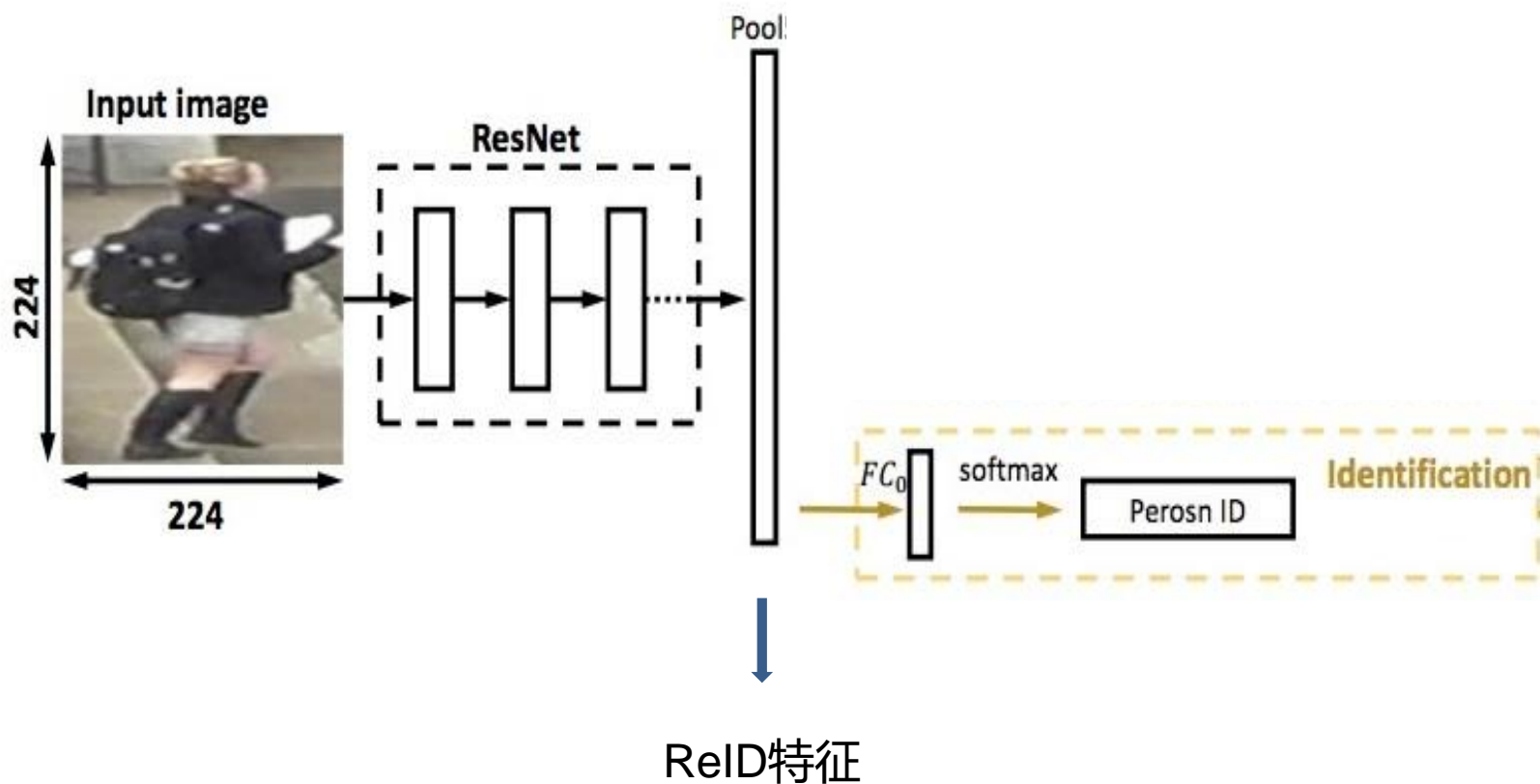


- 验证损失：输入一对（两张）行人图片，让网络来学习这两张图片是否属于同一个行人，等效于二分类问题



# 表征学习

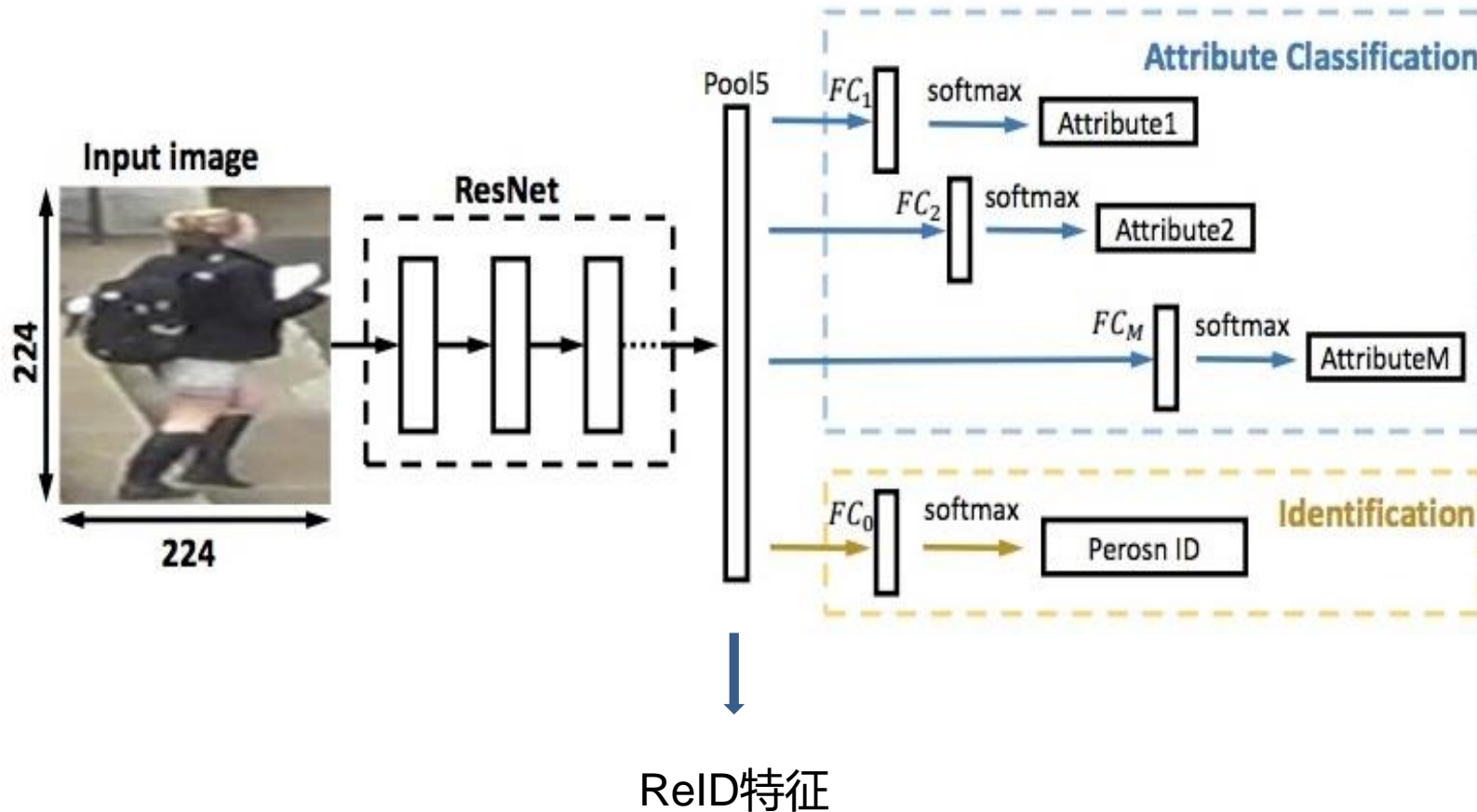
## 分类损失（Classification/Identification Loss）



- 分类损失也叫ID损失，只有ID损失的网络称为ID Embedding网络（IDE网络）
- 训练集中行人的ID数为网络的类别数
- 特征层后接一个分类FC，经过Softmax激活函数计算交叉熵损失
- 测试阶段使用倒数第二层的特征向量进行检索，分类FC层丢弃

# 表征学习

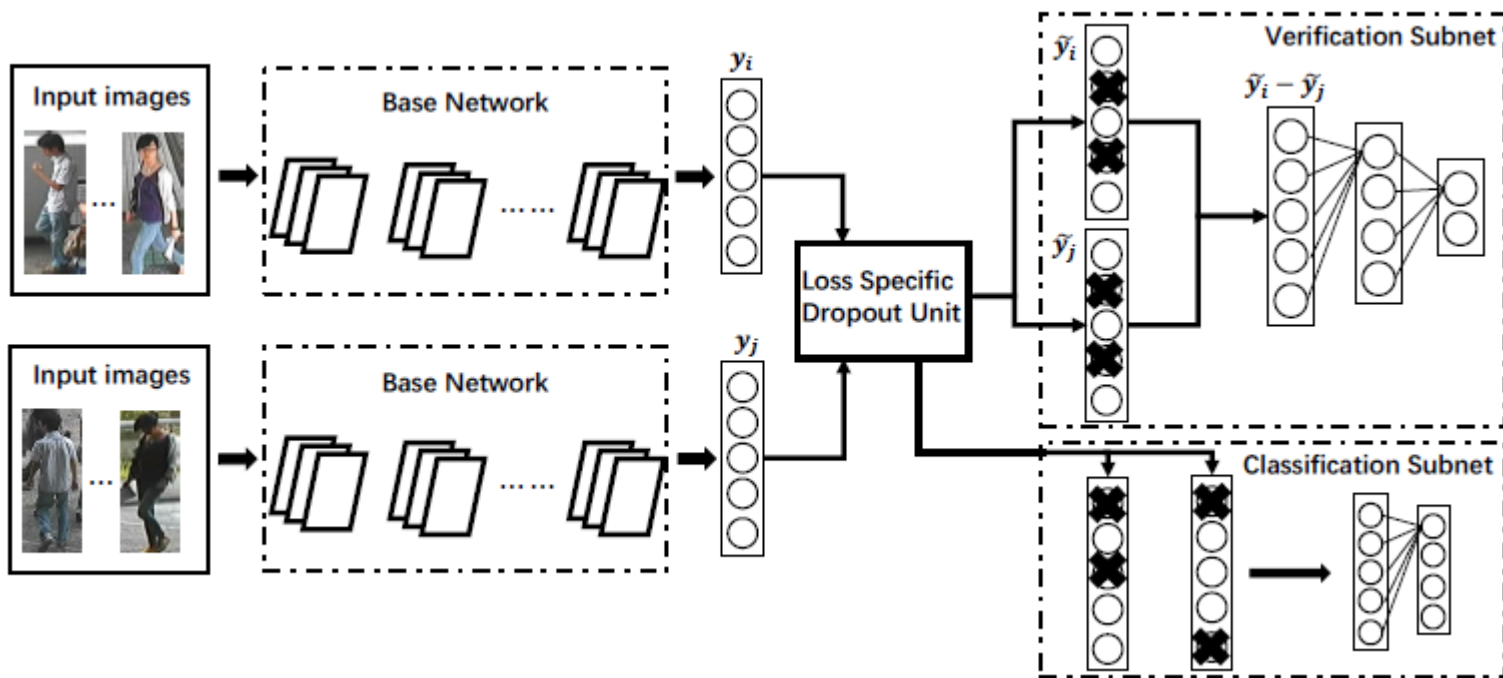
## 属性损失 (Attribute Loss)



- 可以同时连接几个属性分类损失增强ReID特征的性能
- 每一个属性损失都是一个分类的交叉熵
- 可以等效于一个multi-task的网络
- 测试阶段将所有分类FC丢弃, 只使用ReID特征

# 表征学习

## 验证损失 (Verification Loss)



- 每次输入一对（两张）图片，进入同一个Siamese网络提取特征
- 融合两个特征信息计算一个二分类损失（验证损失）
- 训练阶段可以和ID损失一起使用
- 测试阶段输入两张图片，直接判断该两张图片是否属于一个行人

## 总结

- 通过构造网络来直接得到鲁棒的ReID特征，不直接学习图片之间的相似性
- 通常需要额外的FC层来辅导特征学习，测试阶段FC层会被丢弃
- ID损失的FC层维度与ID数量一致，当训练集太大时网络巨大，训练收敛很难
- 验证损失测试的时候需要输入一对图片，识别效率很低
- 表征学习通常而言训练比较稳定，结果易复现
- 表征学习的分布式训练通常比较成熟



## 定义

- 度量学习旨在学习出两张图片的相似性

定义一个映射：

$$f(x) : \mathbb{R}^F \rightarrow \mathbb{R}^D$$

将图片从原始域映射到特征域, 之后再定义一个距离度量函数

$$D(x, y) : \mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}$$

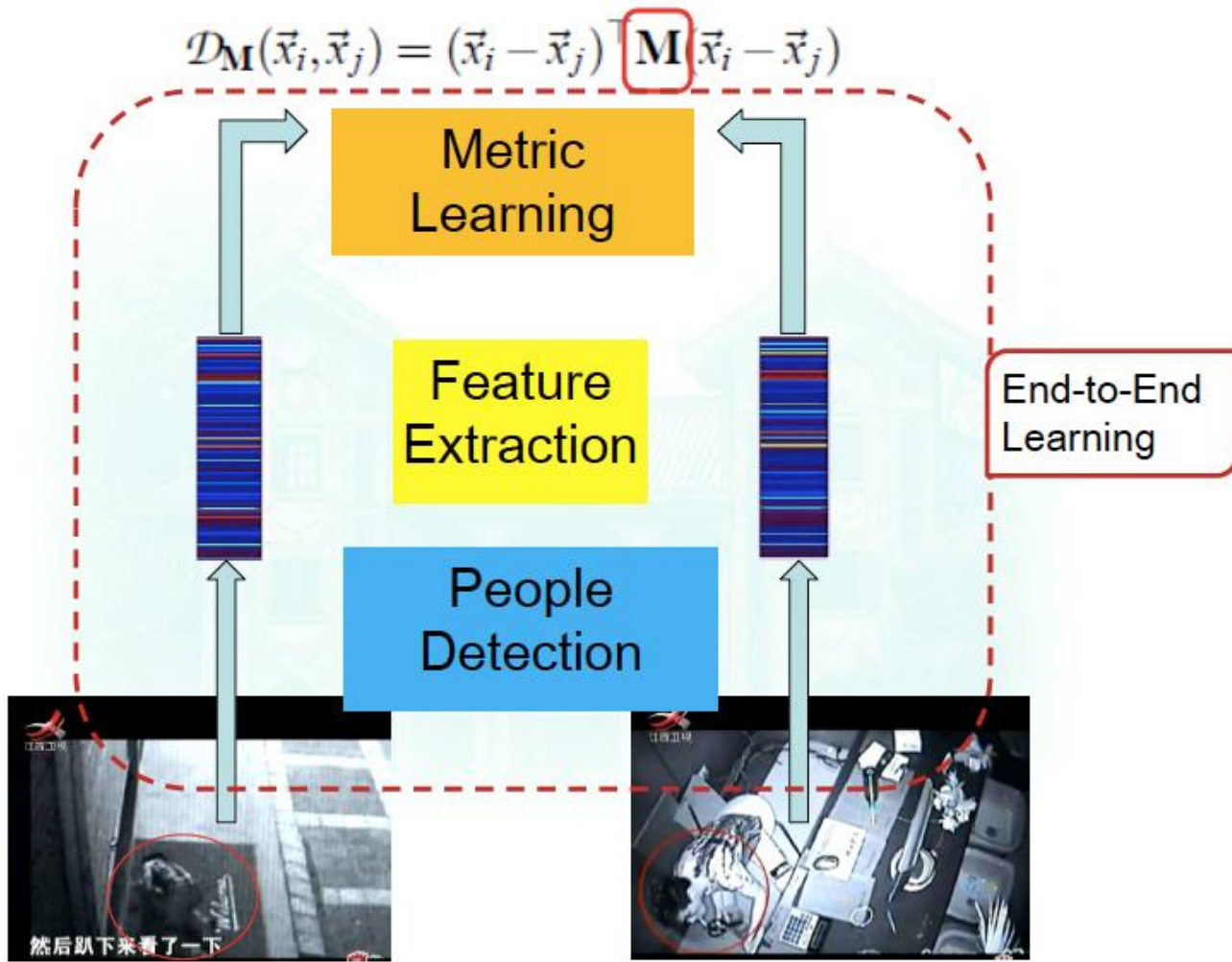
来计算两个特征向量之间的距离。最后通过最小化网络的度量损失, 来寻找一个最优的映射 $f(x)$ , 使得相同行人两张图片（正样本对）的距离尽可能小, 不同行人两张图片（负样本对）的距离尽可能大。而这个映射 $f(x)$ , 就是我们训练得到的深度卷积网络。

- 欧式距离  $d_{I_1, I_2} = \|f_{I_1} - f_{I_2}\|_2$
- 余弦距离  $d_{I_1, I_2} = 1 - \frac{f_{I_1} \cdot f_{I_2}}{\|f_{I_1}\|_2 \|f_{I_2}\|_2}$





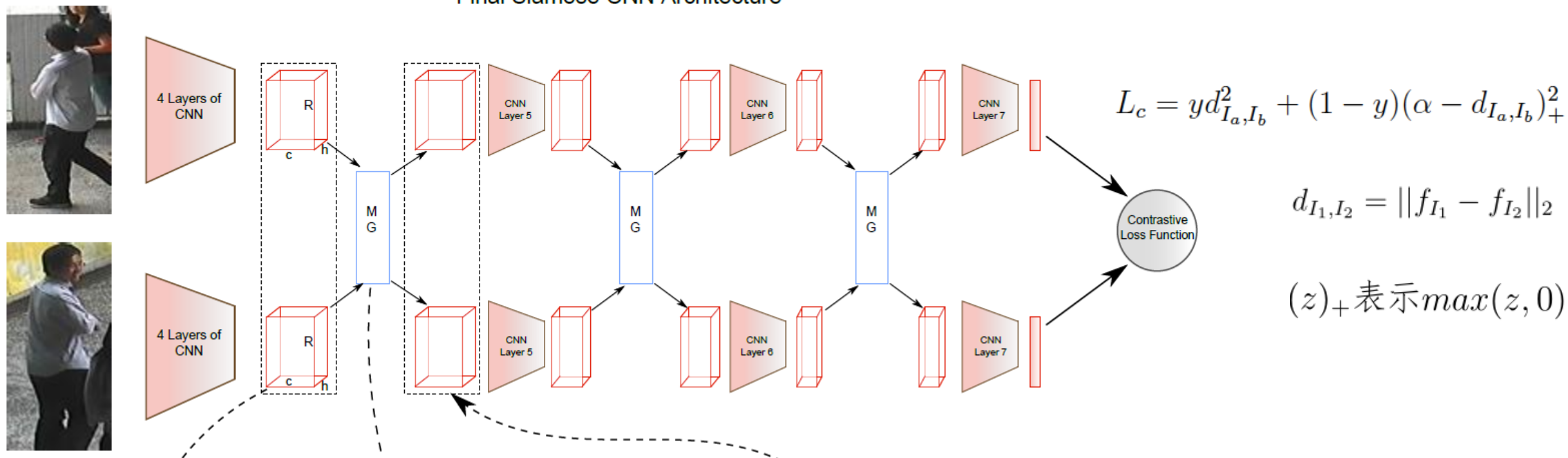
## 概述



- 对比损失 (Contrastive loss)
- 三元组损失 (Triplet loss)
- 改进三元组损失 (Improved triplet loss)
- 四元组损失 (Quadruplet loss)
- TriHard loss

## 对比损失 (Contrastive loss)

Final Siamese CNN Architecture



每次输入一对（两张）图片，拉近正样本对（ $y=1$ ）间距离趋于0，推开负样本对（ $y=0$ ）间距离大于 $\alpha$

## 三元组损失 (Triplet loss)

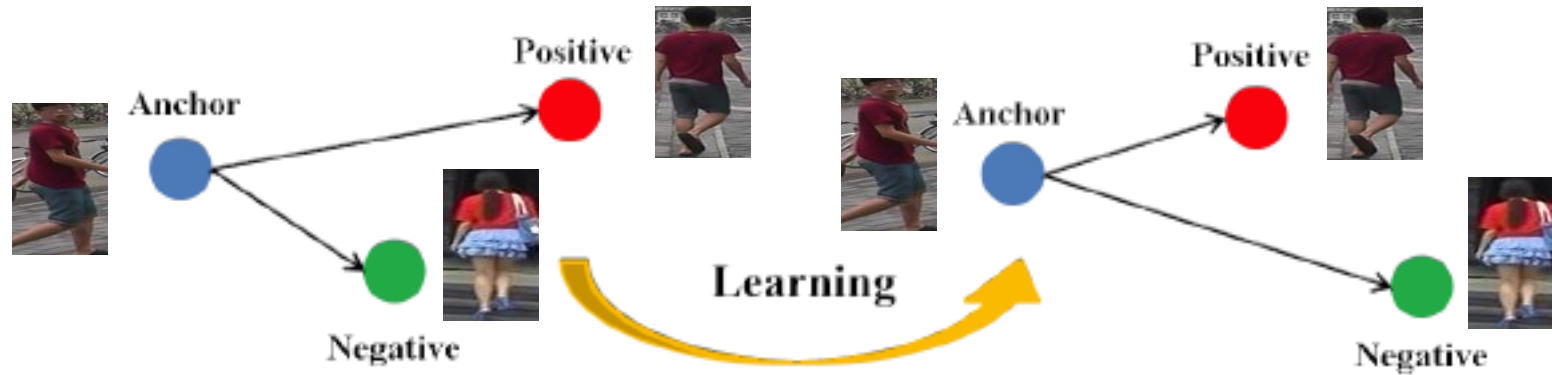


Figure 4. Triplet loss.



$$d_{a,p} + \alpha < d_{a,n}$$



$$L_t = (d_{a,p} - d_{a,n} + \alpha)_+$$

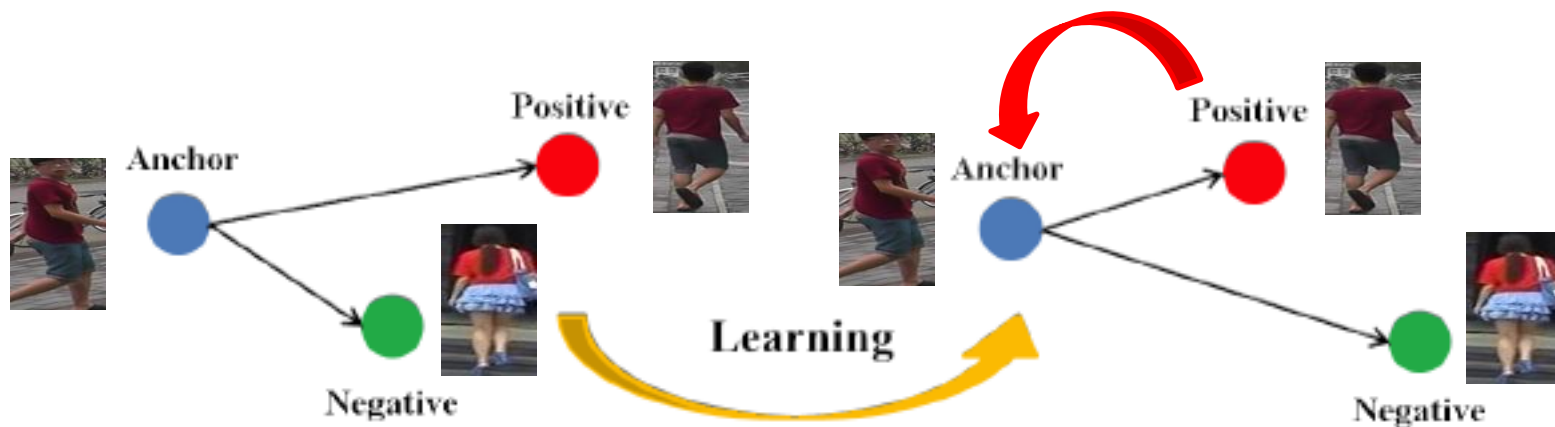
$$\alpha = 0.3$$

$$d_{a,p} = 0.5, d_{a,n} = 0.7, L_t = 0.1$$

$$d_{a,p} = 1.5, d_{a,n} = 1.7, L_t = 0.1$$

# 度量学习

## 改进三元组损失 (Improved triplet loss)



a



p



n

$$L_t = (d_{a,p} - d_{a,n} + \alpha)_+$$

Improved

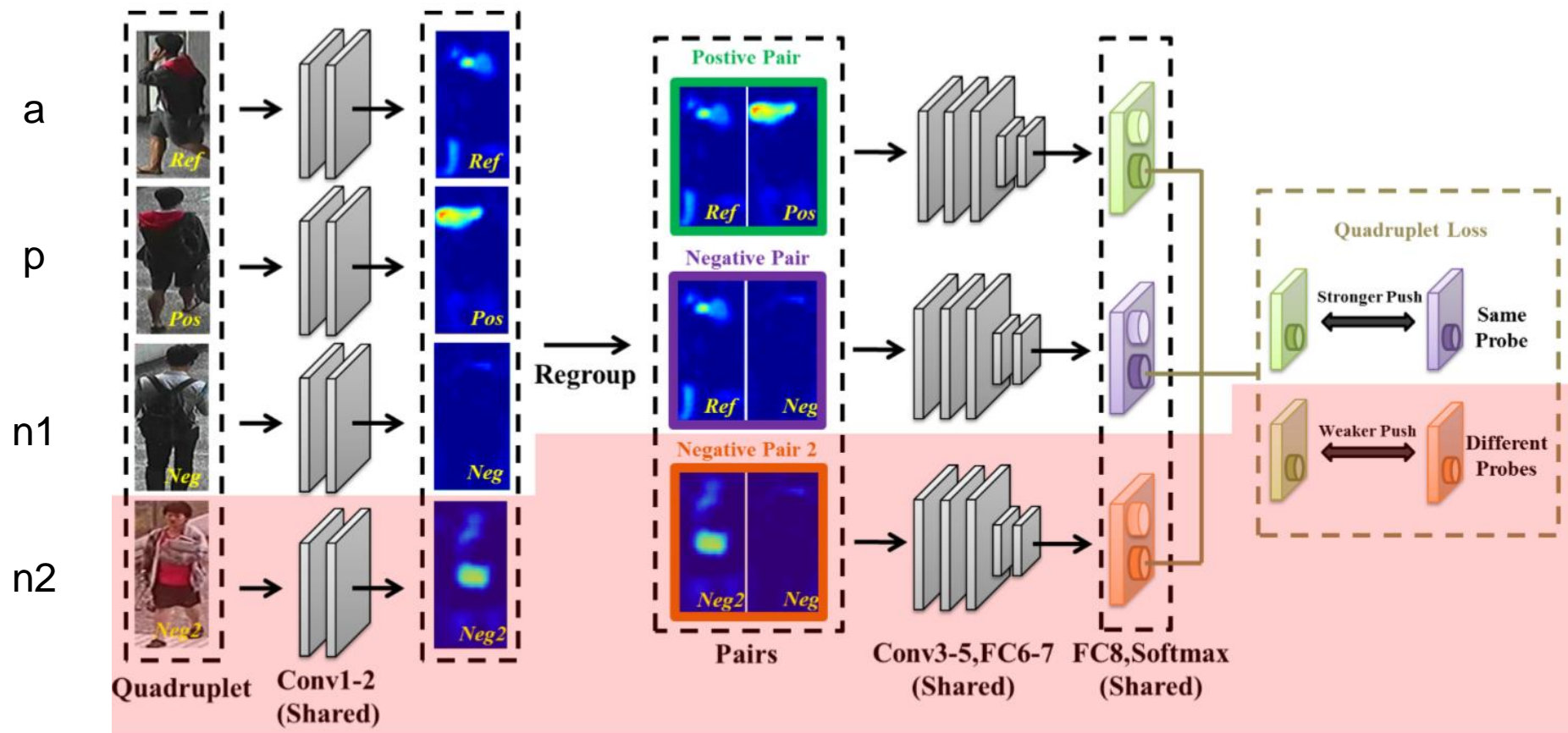
$$L_{it} = d_{a,p} + (d_{a,p} - d_{a,n} + \alpha)_+$$

$$\alpha = 0.3$$

$$d_{a,p} = 0.5, d_{a,n} = 0.7, L_{it} = 0.6$$

$$d_{a,p} = 1.5, d_{a,n} = 1.7, L_{it} = 1.6$$

## 四元组损失 (Quadruplet loss)



- 输入四张图片，一个对正样本，两张负样本，共3个ID
- 第一项为正常的三元组损失
- 第二项为弱推动的三元组损失

$$L_q = (d_{a,p} - d_{a,n1} + \alpha)_+ + (d_{a,p} - d_{n1,n2} + \beta)_+$$



# 度量学习

## 对比

- Contrastive loss

$$L_c = yd_{I_a, I_b}^2 + (1 - y)(\alpha - d_{I_a, I_b})_+^2$$

- Triplet loss

$$L_t = (d_{a,p} - d_{a,n} + \alpha)_+$$

- Improve triplet loss

$$L_{it} = d_{a,p} + (d_{a,p} - d_{a,n} + \alpha)_+$$

- Quadruplet loss

$$L_q = (d_{a,p} - d_{a,n1} + \alpha)_+ + (d_{a,p} - d_{n1,n2} + \beta)_+$$



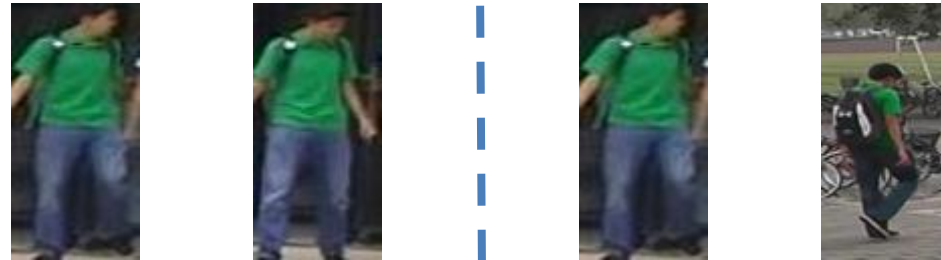
$a$        $p$        $n/n_1$        $n_2$

# 度量学习

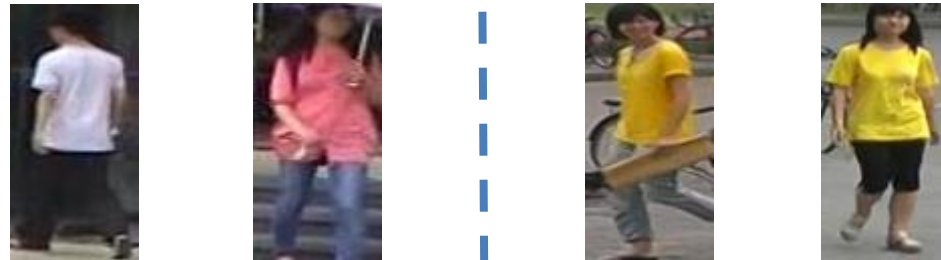
## 难样本挖掘

- 传统的度量学习方法是随机采样组成元组样本，但是这样采集的样本通常是一些**非常容易识别**的样本，不利于训练表达能力强的ReID网络

正样本对



负样本对



简单的样本

困难的样本

# 度量学习

## Triplet loss with hard example mining (TriHard loss)

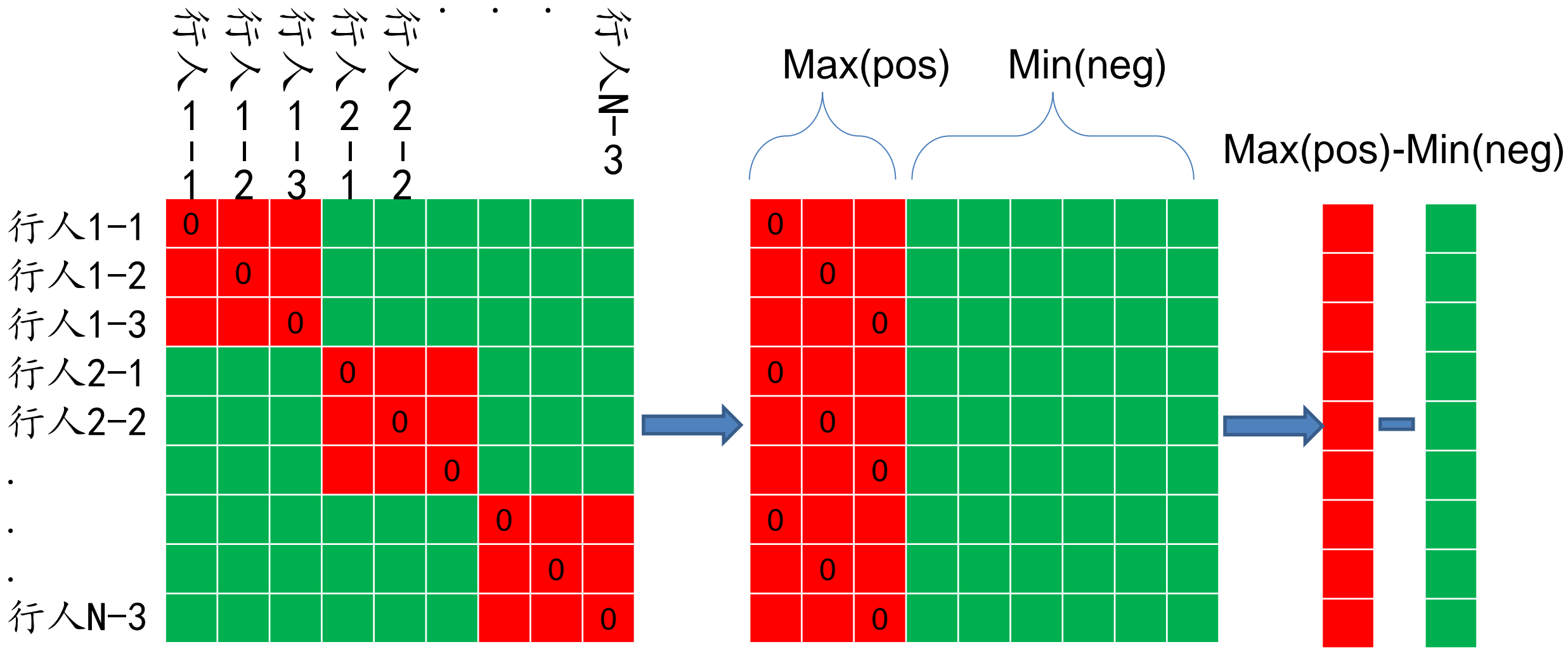
- TriHard loss的核心思想是：对于每一个训练batch挑选P个ID的行人，每个行人，随机挑选K张不同的图片，即一个batch含有P×K张图片。之后对于batch中的每一张图片a，我们可以挑选一个最难的正样本和一个最难的负样本和a组成一个三元组。首先我们定义和a为相同ID的图片集为A，剩下不同ID的图片集为B，则TriHard损失表示为：

$$L_{th} = \frac{1}{P \times K} \sum_{a \in batch} (\max_{p \in A} d_{a,p} - \min_{n \in B} d_{a,n} + \alpha)_+$$

↓  
最难的正  
样本对

↓  
最难的负  
样本对

## TriHard loss实现



## Triplet loss with adaptive weights

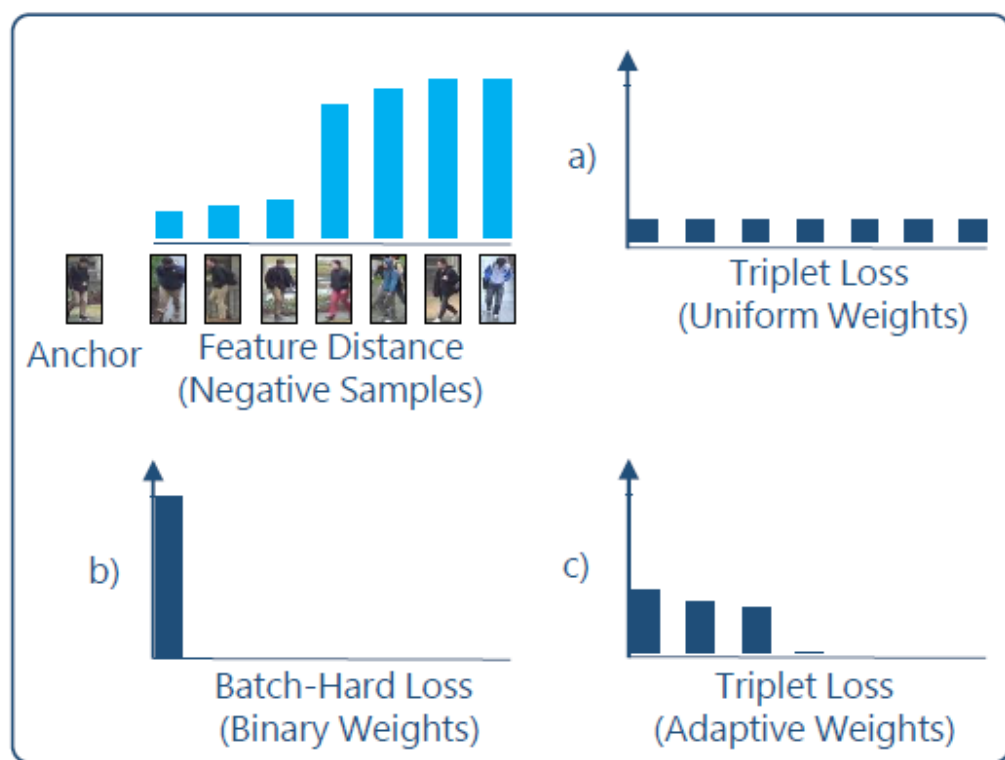


Figure 3. Triplet loss weighing schemes.

$$L_3 = \left[ m + \sum_{x_p \in P(a)} w_p d(x_a, x_p) - \sum_{x_n \in N(a)} w_n d(x_a, x_n) \right]_+$$

$$w_p = \frac{e^{d(x_a, x_p)}}{\sum_{x \in P(a)} e^{d(x_a, x)}} , \quad w_n = \frac{e^{-d(x_a, x_n)}}{\sum_{x \in N(a)} e^{-d(x_a, x)}} .$$

- TriHard loss只考虑了极端样本的信息
- 自适应权重损失根据样本对之间的距离计算自适应权重
- 对于正样本对，距离越大，权重越大；对于负样本对，距离越大，权重越小



## 总结

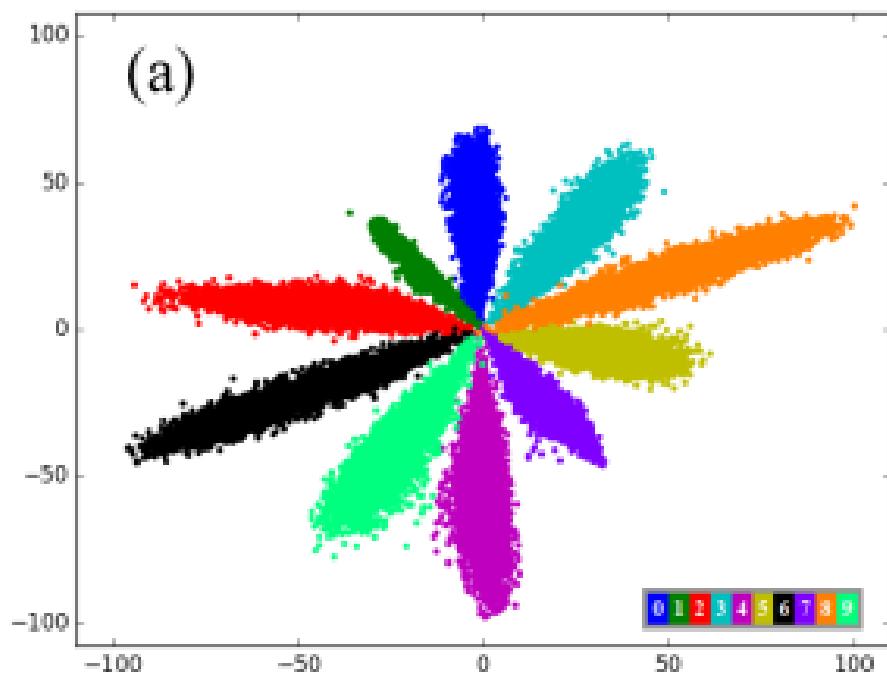
- 通过构造网络来直接学习图片之间的相似性
- 不需要额外的FC层来辅导特征学习
- 网络大小与训练集规模无关，但是数据采样器时间消耗会增加
- TriHard Loss为目前业界度量学习的标杆
- 度量学习通常而言训练比较随机，需要一定的训练经验
- 度量学习的分布式训练不太成熟，通常需要自己实现部分代码

## 一些最新论文

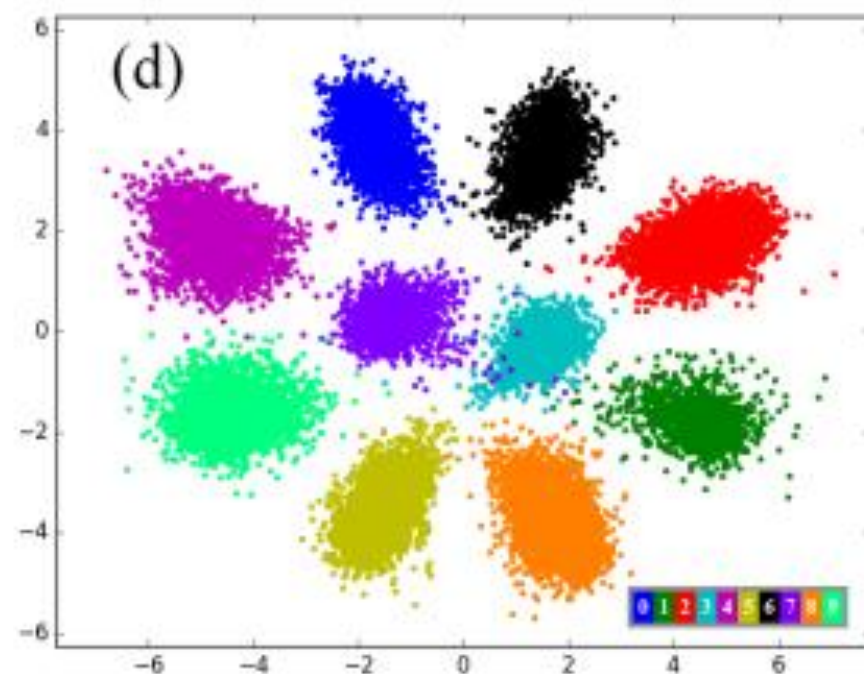
- Zhai Y, Guo X, Lu Y, et al. In Defense of the Classification Loss for Person Re-Identification[J]. arXiv preprint arXiv:1809.05864, 2018.
- Chen K, Chen Y, Han C, et al. Improving Person Re-Identification by Adaptive Hard Sample Mining[C]//2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018: 1638-1642.
- Oh Song H, Xiang Y, Jegelka S, et al. Deep metric learning via lifted structured feature embedding[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 4004-4012.
- Yu R, Dou Z, Bai S, et al. Hard-Aware Point-to-Set Deep Metric for Person Re-identification[J]. arXiv preprint arXiv:1807.11206, 2018.
- Wang C, Zhang Q, Huang C, et al. Manacs: A Multi-task Attentional Network with Curriculum Sampling for Person Re-identification[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 365-381.

# 表征学习&度量学习

## 特征空间对比



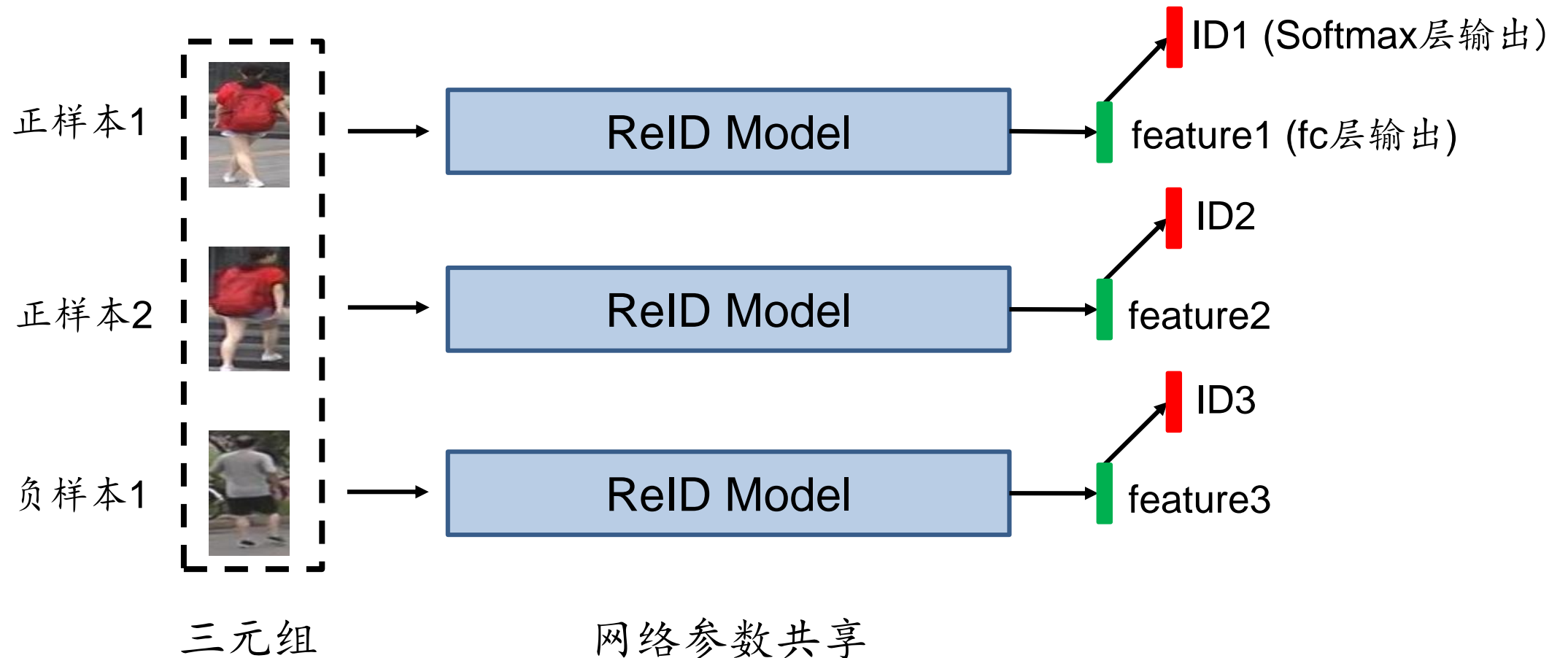
表征学习



度量学习

# 表征学习&度量学习

## 联合训练



- 表征学习损失: ID1, ID2, ID3
- 度量学习损失: f1, f2, f3

# 课后思考

---

1. 是否可以改进ID损失，使其能直接学习相似性度量？
2. 度量学习的分布式训练应该如何实现？



欢迎关注AI300学院

