

# Bag of Tricks and A Strong Baseline for Deep Person Re-identification

Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, Wei Jiang

Zhejiang University, Chinese Academy of Sciences, Xi'an Jiaotong University

## 摘要

本文提出一个ReID中简单且有效的baseline。本文使用一些trick，在只使用全局特征的情况下达到rank1 94.5%和mAP 85.9%的水平。

## 简介

当前大部分的工作都在性能不太好的baseline上进行，本文下网提供一个强力的baseline给研究者和社区来推动ReID的发展。

本文集合了一些高效的训练方法并设计了一个BNNeck。

## 标准baseline

骨干网络是ResNet50，训练中有以下五个步骤：

- 1、使用在ImageNet上预训练好的ResNet50参数，并将最后的全连接层的输出改为 $N$ （训练集中id数）
- 2、每个批次随机采样 $P$ 个人的 $K$ 张图片进行训练，batch size 为 $B = P * K$ ，本文设计 $P = 16, K = 4$
- 3、将每张图片resize到 $256 * 128$ 并向外用0补充10个像素，在随机裁剪到 $256 * 128$
- 4、每张图片以50%的概率水平翻转
- 5、每张图片编码为介于[0,1]的32位浮点数的像素值，然后归一化RGB通道
- 6、模型输出ReID特征 $f$ 和ID预测标签 $p$
- 7、ReID特征 $f$ 用来计算triplet loss，ID预测 $p$ 用来计算cross entropy loss。triplet loss的margin设为0.3

8、使用Adam方法来对模型进行最优化，初始学习率设为0.00035，在40epoch和70epoch时衰减0.1，一共训练120轮

## 训练trick

Fig2(b)显示了这一部分所描述的训练策略

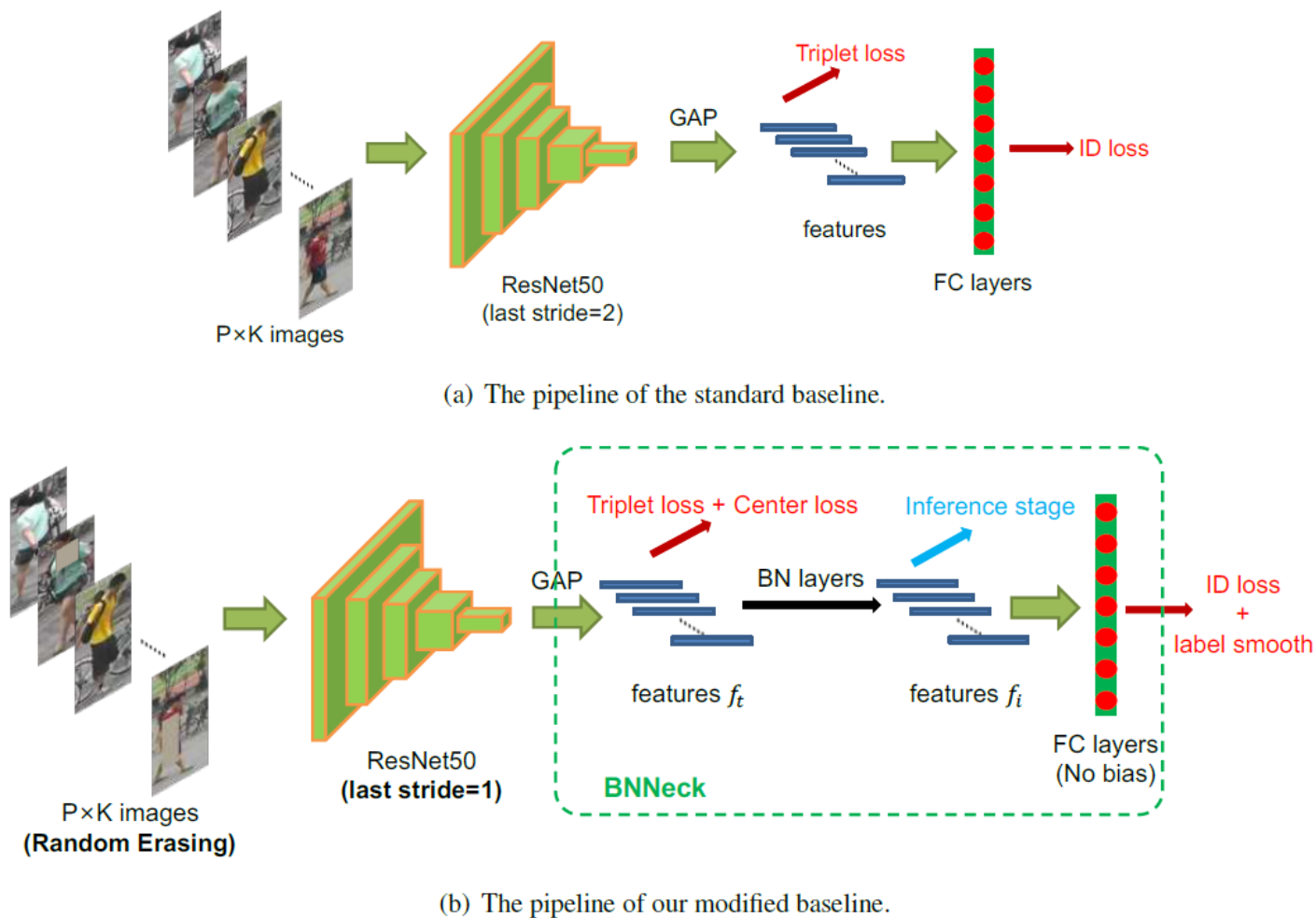


Figure 2. The pipelines of the standard baseline and our modified baseline.

## 预热学习率（Warmup）

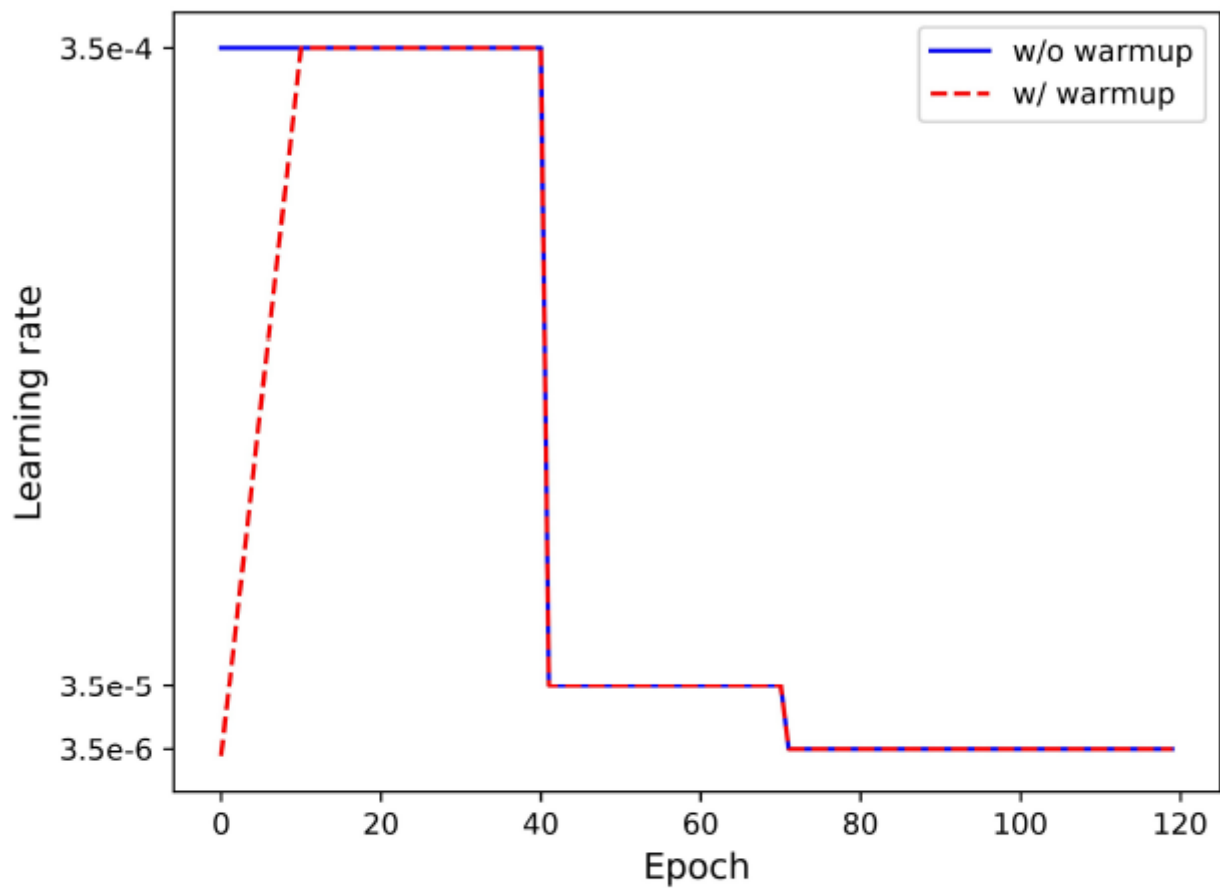


Figure 3. Comparison of learning rate schedules. With warmup strategy, the learning rate is linearly increased in the first 10 epochs.

随机擦除增强(RandomErasing, REA)

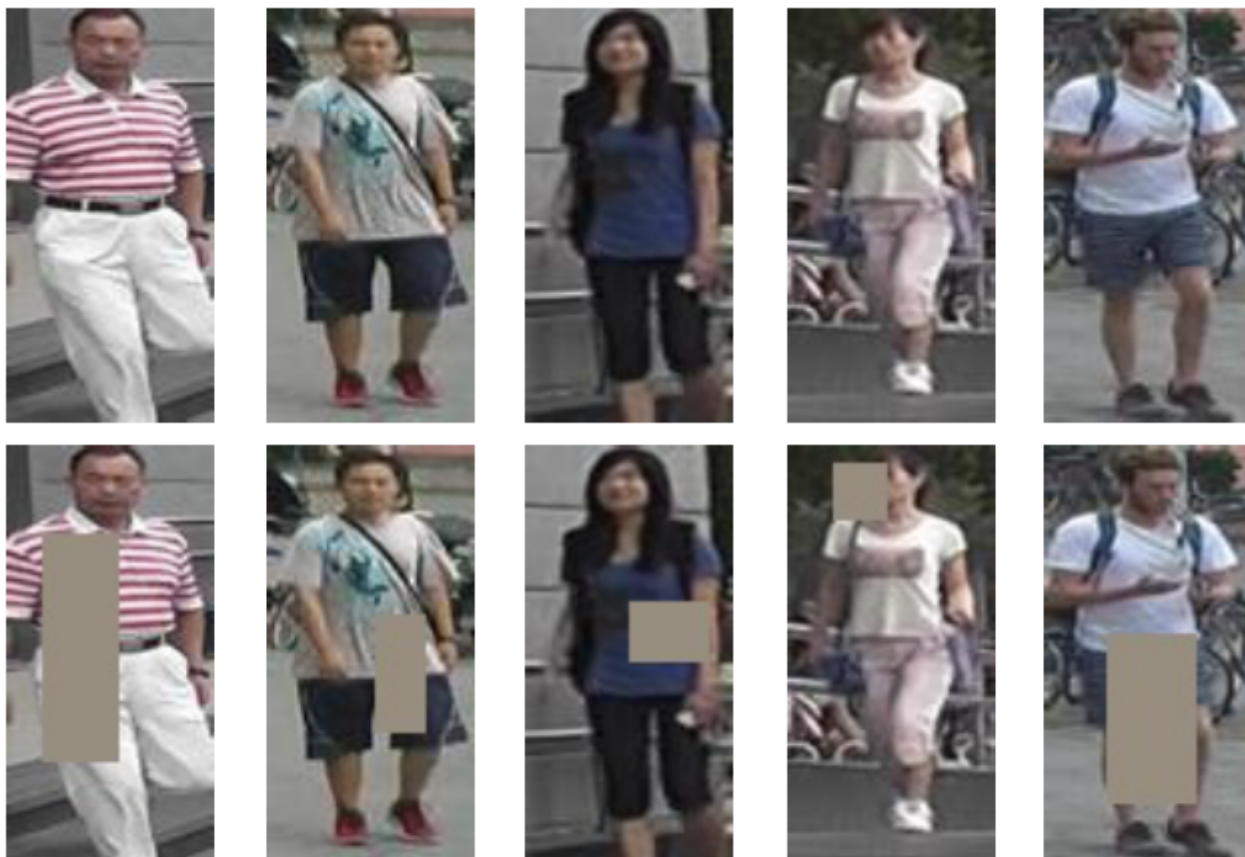


Figure 4. Sampled examples of random erasing augmentation. The first row shows five original training images. The processed images are presented in the second row.

## 标签平滑(LabelSmooth, LS)

ID嵌入（IDE）[25]网络是ReID的基本基线。IDE的最后一层，输出图像>ID预测对数，是一个全连接层，其隐藏大小等于人数N。给定图像，我们将 $y$ 表示为真ID标签，将 $p_i$ 表示为ID预测日志。交叉熵损失的计算公式为：

$$L(ID) = \sum_{i=1}^N -q_i \log(p_i) \begin{cases} q_i = 0, & y \neq i \\ q_i = 1, & y = i \end{cases}$$

由于分类的类别是由人员ID决定的，因此本文将这种损失函数称为ID损失。

文献[17]中提出的标签平滑（LS）是一种广泛使用的方法，用于防止过度拟合分类任务。它将 $q_i$ 的构造更改为：

$$q_i = \begin{cases} 1 - \frac{N-1}{N}\varepsilon, i = y \\ \frac{\varepsilon}{N}, otherwise \end{cases}$$

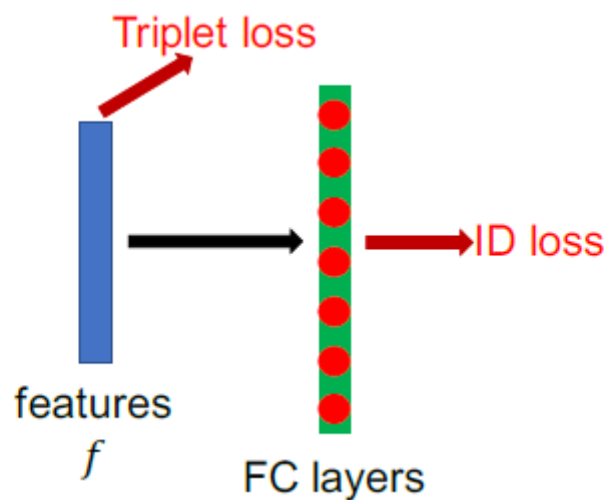
$\varepsilon$ 是一个小常数，可以鼓励模型对训练集失去信心。本文中设置 $\varepsilon = 0.1$

## 最后一层的步长(**Last Stride**)

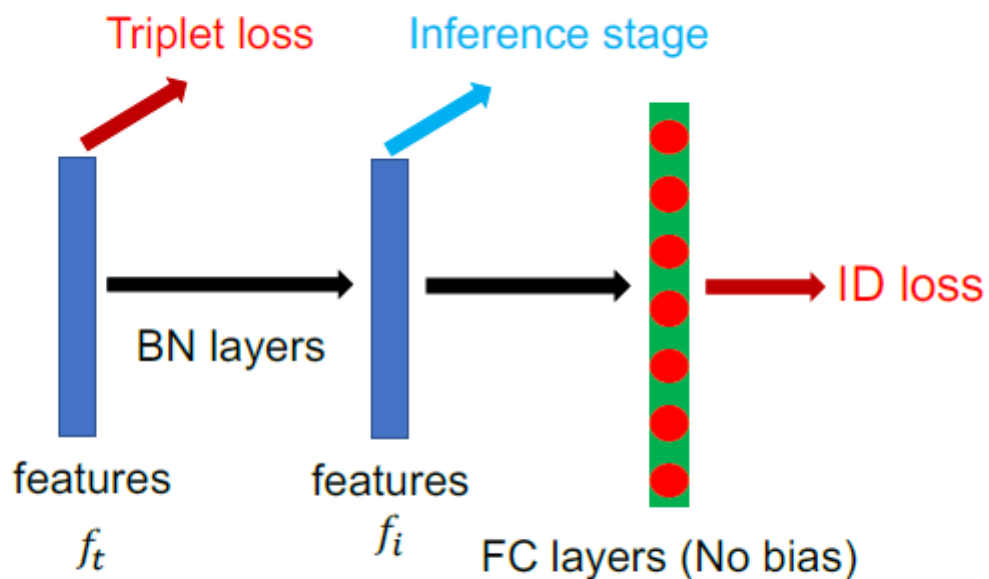
在[16]中，Sun等人。删除了骨干网中最后的空间下采样操作，以增加特征图的大小。

## **BNNeck**

大多数的工作结合了ID loss和triplet loss来训练ReID模型。标准baseline中，ID loss和triplet loss使用同样的特征 $f$ 。但是，这两个损失的目标在嵌入空间中不一致。



(a) The neck of the standard baseline.



(b) Our designed BNNeck. In the inference stage, we choose  $f_i$  following the BN layer to do the retrieval.

Figure 5. Comparison between standard neck and our designed BNNeck.

如图6（a）所示，ID loss构造了几个超平面，以将嵌入空间分成不同的子空间。

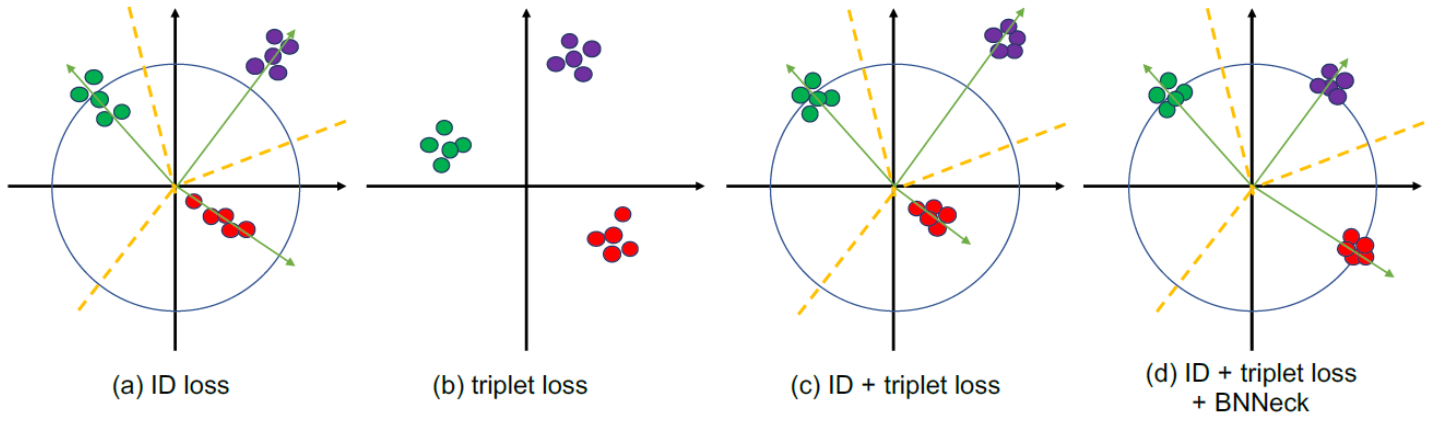


Figure 6. Two-dimensional visualization of sample distribution in the embedding space supervised by (a) ID Loss, (b) Triplet Loss, (c) ID + triplet loss and (d) ID + triplet loss + BNNeck. Points of different colors represent embedding features from different classes. The yellow dotted lines stand for the supposed classification hyperplanes.

每个类的特征分布在不同的子空间。这种情况下cos距离比欧式距离更加啊事故和ID loss的优化。

另一方面，**triplet loss**增强了欧几里得空间内的类内紧密度和类间可分离性。由于三重态损失不能提供全局最优约束，因此类间距离有时小于类内距离。更广泛的做法是把**ID loss**和**triplet loss**结合到一起来进行训练。然而，对于嵌入空间中的图像对，**ID**损失主要是优化余弦距离，而三重态损失则集中在欧几里得距离上。如果我们使用这两个损失来同时优化特征向量，则它们的目标可能不一致。在训练过程中，一种可能的现象是一种损失减少了，而另一种损失正在振荡甚至增加。

为了解决这个问题，文章提出**BNNeck**(Fig5(b))，**BNNeck**只是在特征后加一个BN层，BN层之前的特征定义为 $f_t$ ，BN后的特征 $f_i$ ，训练过程中， $f_t$ 和 $f_i$ 用于计算**triplet loss**和**ID loss**。归一化平衡了 $f_i$ 的每个维度。特征在高球表面附近呈高斯分布。这种分布使**ID**损失更易于收敛。另外，**BNNeck**减少了**ID**损失对 $f_t$ 的约束。**ID**损失的较少约束导致三重态损失更易于同时收敛。归一化保持了属于同一个人的特征的紧凑分布。

由于超球面几乎是围绕坐标轴的原点对称的，因此**BNNeck**的另一招是消除分类器FC层的偏置。

## Center Loss

**Triplet loss**计算方式为：

$$L_{Tri} = [d_p - d_n + \alpha]_+$$

其中 $d_p$ 和 $d_n$ 是正样本对和负样本对的距离， $\alpha$ 是**triplet loss**的边界， $[z]_+$ 等于 $\max(z, 0)$ 。本文中， $\alpha$ 设为0.3。

**Center loss**同时学习了每个类别的深层特征的中心并惩罚了深层特征与其对应的类中心之间的距离，弥补了三重态损失的缺点。其定义为： $L_C = \frac{1}{2} \sum_{j=1}^B \|f_{t_j} - c_{y_j}\|_2^2$ ，其中 $y_j$ 是mini-batch中第j个图片的label， $c_{y_j}$ 是第 $y_j$ 个类的深度特征中心。使中心损失最小化可提高类内紧凑性。本文中模型包括三部分：

$$L = L_{ID} + L_{Triplet} + \beta L_C$$

其中 $\beta$ 是评委center loss的权重，实验中设为0.0005

## 实验结果

### 各个trick的影响（单域）

Model	Market1501		DukeMTMC	
	r = 1	mAP	r = 1	mAP
Baseline-S	87.7	74.0	79.7	63.7
+warmup	88.7	75.2	80.6	65.1
+REA	91.3	79.3	81.5	68.3
+LS	91.4	80.3	82.4	69.3
+stride=1	92.0	81.7	82.6	70.6
+BNNeck	94.1	85.7	86.2	75.9
+center loss	94.5	85.9	86.4	76.4

Table 1. The performance of different models is evaluated on Market1501 and DukeMTMC-reID datasets. Baseline-S stands for the standard baseline introduced in section 2.

### 对BNNeck的分析



Feature	Metric	Market1501		DukeMTMC	
		r = 1	mAP	r = 1	mAP
$f$ (w/o BNNeck)	Euclidean	92.0	81.7	82.6	70.6
$f_t$	Euclidean	94.2	85.5	85.7	74.4
$f_t$	Cosine	94.2	85.7	85.5	74.6
$f_i$	Euclidean	93.8	83.7	86.6	73.0
$f_i$	Cosine	<b>94.1</b>	<b>85.7</b>	<b>86.2</b>	<b>75.9</b>

Table 2. The ablation study of BNNeck.  $f$  (w/o BNNeck) is baseline without BNNeck. BNNeck includes two features  $f_t$  and  $f_i$ . We evaluate the performance of them with Euclidean distance and cosine distance, respectively.

余弦距离度量的 $f_t$ 优于欧几里德距离度量的 $f_t$ 。由于ID损失直接约束了BN层之后的特征，因此 $f_i$ 可以被几个超平面清楚地分开。余弦距离可以测量两个特征向量之间的角度，因此对于 $f_i$ ，余弦距离度量比欧几里德距离度量更合适。但是， $f_t$ 接近triplet loss，并且同时受到ID损失的限制。两种度量标准可以实现类似 $f_t$ 的性能。总的来说，BNNeck可以极大提高ReID模型的性能。

## 每个Trick的影响（跨域）

Table3中列出了跨域的实验，包括warmup，label smooth和BNNeck。

改变步长和center loss看起来没有太大的影响，REA甚至对跨域RedID的性能是负优化。推测REA使得模型更加拟合训练集的数据分布。

Model	M→D		D→M	
	r = 1	mAP	r = 1	mAP
Baseline	24.4	12.9	34.2	14.5
+warmup	26.3	14.1	39.7	17.4
+REA	21.5	10.2	32.5	13.5
+LS	23.2	11.3	36.5	14.9
+stride=1	23.1	11.8	37.1	15.4
+BNNeck	26.7	15.2	47.7	21.6
+center loss	27.5	15.0	47.4	21.4
-REA	41.4	25.7	54.3	25.5

Table 3. The performance of different models is evaluated on cross-domain datasets. M→D means that we train the model on Market1501 and evaluate it on DukeMTMC-reID.

对**SOTA**模型的对比

Type	Method	$N_f$	Market1501		DukeMTMC	
			r = 1	mAP	r = 1	mAP
Pose-guided	GLAD[19]	4	89.9	73.9	-	-
	PIE [23]	3	87.7	69.0	79.8	62.0
	PSE [13]	3	78.7	56.0	-	-
Mask-guided	SPReID [7]	5	92.5	81.3	84.4	71.0
	MaskReID [9]	3	90.0	75.3	78.8	61.9
Stripe-based	AlignedReID [21]	1	90.6	77.7	81.2	67.4
	SCPNet [3]	1	91.2	75.2	80.3	62.6
	PCB [16]	6	93.8	81.6	83.3	69.2
	Pyramid[22]	1	92.8	82.1	-	-
	Pyramid[22]	21	95.7	88.2	89.0	79.0
	BFE[1]	2	94.5	85.0	88.7	75.8
Attention-based	Manacs [18]	1	93.1	82.3	84.9	71.8
	DuATM [14]	1	91.4	76.6	81.2	62.3
	HA-CNN [8]	4	91.2	75.7	80.5	63.8
GAN-based	Camstyle [28]	1	88.1	68.7	75.3	53.5
	PN-GAN [10]	9	89.4	72.6	73.6	53.2
Global feature	IDE [25]	1	79.5	59.9	-	-
	SVDNet [15]	1	82.3	62.1	76.7	56.8
	TriNet[6]	1	84.9	69.1	-	-
	AWTL[12]	1	89.5	75.7	79.8	63.4
	<b>Ours</b>	1	<b>94.5</b>	<b>85.9</b>	<b>86.4</b>	<b>76.4</b>
	<b>Ours(RK)</b>	1	<b>95.4</b>	<b>94.2</b>	<b>90.3</b>	<b>89.1</b>

Table 4. Comparison of state-of-the-art methods.  $N_f$  is the number of features used in the inference stage. RK stands for  $k$ -reciprocal re-ranking method [26]

## 补充实验

### Batch Size的影响

Batch Size $P \times K$	Market1501		DukeMTMC	
	r = 1	mAP	r = 1	mAP
$8 \times 3$	92.6	79.2	84.4	68.1
$8 \times 4$	92.9	80.0	84.7	69.4
$8 \times 6$	93.5	81.6	85.1	70.7
$8 \times 8$	93.9	82.0	85.8	71.5
$16 \times 3$	93.8	83.1	86.8	72.1
$16 \times 4$	93.8	83.7	86.6	73.0
$16 \times 6$	94.0	82.8	85.1	69.9
$16 \times 8$	93.1	81.6	86.7	72.1
$32 \times 3$	94.5	84.1	86.0	71.4
$32 \times 4$	93.2	82.8	86.5	73.1

Table 5. Performance of ReID models with different numbers of batch size.

## 图片大小的影响

Image Size	Market1501		DukeMTMC	
	r = 1	mAP	r = 1	mAP
$256 \times 128$	93.8	83.7	86.6	73.0
$224 \times 224$	94.2	83.3	86.1	72.2
$384 \times 128$	94.0	82.7	86.4	73.2
$384 \times 192$	93.8	83.1	87.1	72.9

Table 6. Performance of ReID models with different image sizes.

## 总结和回顾

本文收集了一些有效的训练技巧，并为 Person ReID 设计了牢固的baseline。为了证明每种技巧对ReID模型性能的影响，对同域和跨域ReID任务进行了大量实验。最后，仅使用全局功能，trong baseline就可以在Market1501上达到94.5%的rank1精度和85.9%的mAP。希望这项工作能够促进ReID在学术界和工业界的研究。

将来，将继续设计更多实验来分析这些trick的原理。例如，当用L2规范化替换BNNeck时，该网络的性能如何？此外，是否可以在strong baselie上扩展一些最先进的方法，例如PCB，MGN和AlignedReID等？更多的可视化也有助于其他人理解这项工作。