

In addition, we also found that some works were unfairly compared with other state-of-the-arts methods. Specifically, the improvements were mainly from training tricks rather than methods themselves. But the training tricks were understated in the paper so that readers ignored them.

it would make the effectiveness of the method exaggerated. We suggest that reviewers need to take into account these tricks when commenting academic papers.

此外，我们还发现，一些作品与其他先进的方法相比是不公平的。具体来说，这些改进主要来自训练技巧，而不是方法本身。但训练技巧在paper上被低估了，所以读者忽略了它们

这将使该方法的有效性被夸大。我们建议审稿人在评论学术论文时考虑这些技巧。

Apart from aforementioned reasons, another consideration is that the industry prefers to simple and effective models rather than **concatenating** lots of local features in the inference stage. In pursuit of high accuracy, researchers in the academic always combine several local features or utilize the semantic information from pose estimation or segmentation models. Such methods bring too much extra consumption. Large features also greatly reduce the speed of retrieval process. Thus, we hope to use some tricks to improve the ability of the ReID model and only use global features to achieve high performance. The purposes of this paper are summarized as follow:

除了上述原因，另一个需要考虑的是，该行业更喜欢简单、高效的模型，而不是在推理过程中**串联**大量的本地特征。为了获得较高的精度，学者们通常会结合局部特征或利用姿态估计或分割模型中的语义信息。这些方法带来了太多的额外消耗。大的特征也大大降低了检索的速度。因此，我们希望使用一些技巧来提高ReID模型的能力，并且只使用全局特性来实现高性能。本文的研究目的如下

这块我没翻译

Moreover, we found different works choose different image sizes and numbers of batch size, as a supplement, we also explore their impacts on model performance. In summary, the contributions of this paper are concluded as follow

此外，我们发现不同的作品选择不同的图像尺寸和批量大小，作为补充，我们也探讨了它们对模型性能的影响。综上所述，本文的贡献总结如下

## 创新点

1. We collect some effective training tricks for person ReID. Among them, we design a new neck structure named as BNNeck. In addition, we evaluate the improvements from each trick on two widely used datasets
2. We provide a strong ReID baseline, which achieves 94.5% and 85.9% mAP on Market1501. it is worth mentioned that the results are obtained with global features provided by ResNet50 backbone. To our best knowledge, it is the best performance acquired by global features in person ReID.
3. As a supplement, we evaluate the influences of the image size and the number of batch size on the performance of ReID models

1. 我们为ReID收集了一些有效的训练技巧。其中，我们设计了一种新的颈部结构，命名为BNNeck。另外，我们在两个广泛使用的数据集上评估每个技巧的改进
2. 我们提供了一个强大的ReID基线，在Market1501上实现了94.5%和85.9%的映射。值得一提的是，这些结果是由ResNet50主干网提供的全局特征得到的。据我们所知，它是global features所

获得的最佳性能person ReID。

### 3. 最后，我们评估了图像大小和batch size 的数量对ReID模型性能的影响

1. 我们为学术界和工业界提供强大而简单的ReID基线；
2. 我们收集并评估了Market1501和DukeMTMC-reID数据集上的一些训练技巧；
3. 我们观察到ID丢失和三联体丢失之间的不一致，并提出了一种名为BNNeck的新型结构；
4. 我们观察到ReID任务忽略了类内紧凑性和类间可分性，并且证明中心丢失可以弥补它。

#### Standard Baseline:

We follow a widely used open-source as our standard baseline. The backbone(主干, 骨骼) of the standard baseline is ResNet50. During the training stage, the pipeline includes following steps :

1. We initialize the ResNet50 with pre-trained parameters on ImageNet and change the dimension of the fully connected layer to N. N denotes the number of identities in the training dataset.
2. We randomly sample P identities and K images of per person to constitute a training batch. Finally the batch size equals to  $B = PK$ . In this paper, we set  $P = 16$  and  $K = 4$ .
3. We resize each image into 256 128 pixels and pad the resized image 10 pixels with zero values. Then randomly crop it into a 256 128 rectangular image.
4. Each image is flipped horizontally with 0.5 probability.
5. Each image is decoded into 32-bit floating point raw pixel values in  $[0; 1]$ . Then we normalize RGB channels by subtracting 0.485, 0.456, 0.406 and dividing by 0.229, 0.224, 0.225, respectively.
6. The model outputs ReID features  $f$  and ID prediction logits  $p$ .
7. ReID features  $f$  is used to calculate triplet loss [6]. ID prediction logits  $p$  is used to calculate cross entropy loss. The margin  $m$  of triplet loss is set to be 0.3.
8. Adam method is adopted to optimize the model. The initial learning rate is set to be 0.00035 and is decreased by 0.1 at the 40th epoch and 70th epoch respectively. Totally there are 120 training epochs.

1. 我们在ImageNet上初始化带有预训练参数的ResNet50，并将全连接层的维数更改为N。N表示训练数据集中的身份数。
2. 我们随机抽取每个人的P个身份和K个图像作为训练样本。最后批处理大小为 $B = PK$ ，本文令 $P = 16$ ,  $K = 4$ 。
3. 我们将每个图像的大小调整为256 128像素，并将调整后的图像填充为10像素，并设置零值。然后随机裁剪成一个256 128矩形图像。
4. 每幅图像水平翻转的概率为0.5。
5. 每个图像被解码成32位浮点原始像素值在 $[0;1]$ 。然后分别减去0.485、0.456、0.406，再除以0.229、0.224、0.225，对RGB通道进行归一化处理。
6. 模型输出ReID特征 $f$ 和ID预测 分对数 $p$ 。
7. ReID 特征  $f$  用于计算triplet loss [6]。利用ID预测 logits  $p$ 计算交叉熵损失。triplet loss 的 margin  $m$ 设定为0.3。
8. 采用Adam法对模型进行优化。初始学习率设定为0.00035，第40期和第70期分别降低0.1。总共有120个训练期。

## 3.培训技巧 (training Tricks)

This section will introduce some effective training tricks in person ReID. Most of such tricks can be expanded on the standard baseline without changing the model architecture. The Fig. 2 (b) shows training strategies and the model architecture appeared in this section.

本节将介绍一些有效的训练技巧。大多数这样的技巧可以在标准基线上扩展，而不需要改变模型架构。图2 (b)为本节中出现的训练策略和模型架构。

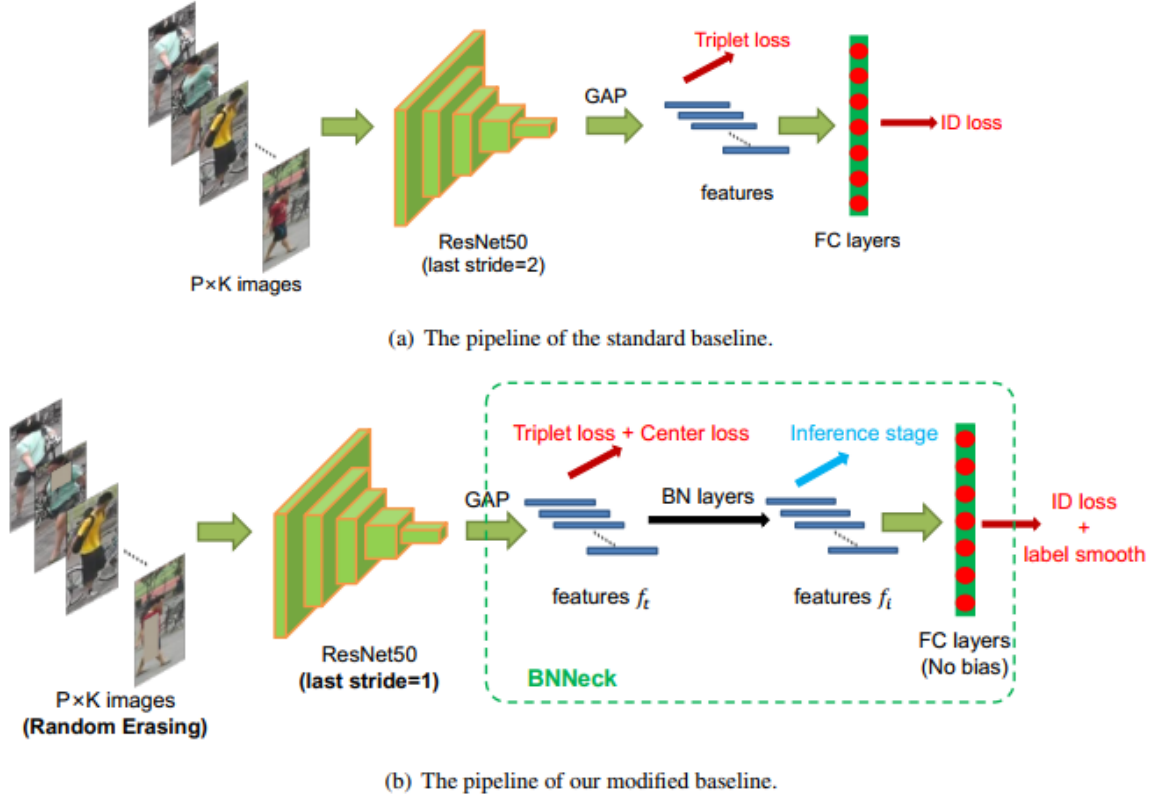


Figure 2. The pipelines of the standard baseline and our modified baseline.

### 3.1. Warmup Learning Rate

Learning rate has a great impact for the performance of a ReID model. Standard baseline is initially trained with a large and constant learning rate. In [2], a warmup strategy is applied to bootstrap the network for better performance. In practice, As shown in Fig. 3, we spent 10 epochs linearly increasing the learning rate from  $3.5 \times 10^{-5}$  to  $3.5 \times 10^{-4}$ . Then, the learning rate is decayed to  $3.5 \times 10^{-5}$  and  $3.5 \times 10^{-6}$  at 40th epoch and 70th epoch respectively. The learning rate  $lr(t)$  at epoch  $t$  is compute as;

$$lr(t) = \begin{cases} 3.5 \times 10^{-5} \times \frac{t}{10} & \text{if } t \leq 10 \\ 3.5 \times 10^{-4} & \text{if } 10 < t \leq 40 \\ 3.5 \times 10^{-5} & \text{if } 40 < t \leq 70 \\ 3.5 \times 10^{-6} & \text{if } 70 < t \leq 120 \end{cases} \quad (1)$$

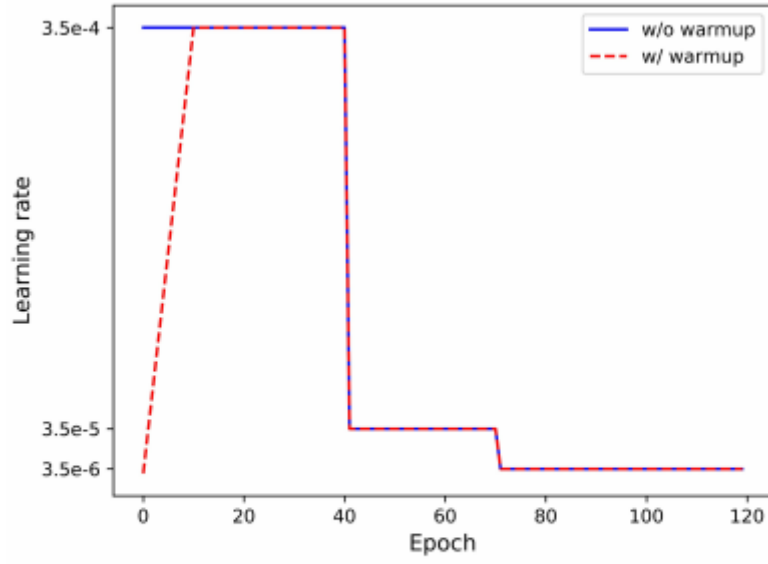


Figure 3. Comparison of learning rate schedules. With warmup strategy, the learning rate is linearly increased in the first 10 epochs.

学习率对ReID模型的性能有很大的影响。标准基线最初以较大且恒定的学习率进行训练。在[2]中，使用预热策略引导网络以获得更好的性能。在实践中，如fig 3所示，我们花费了10个epoch，将学习率从 $3.5 \times 10^{-5}$ 线性增加到 $3.5 \times 10^{-4}$ 。然后分别在第40元和第70元将学习率衰减为 $3.5 \times 10^{-5}$ 和 $3.5 \times 10^{-6}$ 。历元 $t$ 的学习速率 $lr(t)$ 计算为

$$lr(t) = \begin{cases} 3.5 \times 10^{-5} \times \frac{t}{10} & \text{if } t \leq 10 \\ 3.5 \times 10^{-4} & \text{if } 10 < t \leq 40 \\ 3.5 \times 10^{-5} & \text{if } 40 < t \leq 70 \\ 3.5 \times 10^{-6} & \text{if } 70 < t \leq 120 \end{cases} \quad (1)$$

### 3.2. Random Erasing Augmentation (随机擦除增强)

In person ReID, persons in the images are sometimes occluded by other objects. To address the occlusion problem and improve the generalization ability of ReID models, Zhong *et al.* [27] proposed a new data augmentation approach named as Random Erasing Augmentation (REA). In practice, for an image  $I$  in a mini-batch, the probability of it undergoing Random Erasing is  $p_e$ , and the probability of it being kept unchanged is  $1 - p_e$ . Then, REA randomly selects a rectangle region  $I_e$  with size  $(W_e, H_e)$  in image  $I$ , and erases its pixels with random values. Assuming the area of image  $I$  and region  $I_e$  are  $S = W \times H$  and  $S_e = W_e \times H_e$  respectively, we denote  $r_e = \frac{S_e}{S}$  as the area ratio of erasing rectangle region. In addition, the aspect ratio of region  $I_e$  is randomly initialized between  $r_1$  and  $r_2$ . To determine a unique region, REA randomly initializes a point  $\mathcal{P} = (x_e, y_e)$ . If  $x_e + W_e \leq W$  and  $y_e + H_e \leq H$ , we set the region,  $I_e = (x_e, y_e, x_e + W_e, y_e + H_e)$ , as the selected rectangle region. Otherwise we repeat the above process until an appropriate  $I_e$  is selected. With the selected erasing region  $I_e$ , each pixel in  $I_e$  is assigned to the mean value of image  $I$ , respectively.

In this study, we set hyper-parameters to  $p = 0.5, 0.02 < S_e < 0.4, r_1 = 0.3, r_2 = 3.33$ , respectively. Some examples are shown in Fig. 4.



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.

在行人重识别时，有时会被其他物体遮挡。为了解决遮挡问题，提高ReID模型的泛化能力，Zhong等人提出了一种新的数据增强方法——随机消除增强(REA)。在实践中，对于一个小批量的图像 $I$ ，随机擦除的概率为 $p_e$ ，保持不变的概率为 $1 - p_e$ 。然后，REA在图像 $I$ 中随机选择一个大小为 $(W_e, H_e)$ 的矩形区域，并用随机值擦除其像素。假设图像 $I$ 和区域 $I_e$ 的面积分别为 $S = W \times H$ 和 $S_e = W_e \times H_e$ ，我们将 $r_e = S_e/S$ 表示为擦除矩形区域的面积比。另外，区域 $I_e$ 的长宽比在 $r_1$ 和 $r_2$ 之间随机初始化。为了确定一个唯一的区域，REA随机初始化一个点 $\mathcal{P} = (x_e, y_e)$ 。如果 $x_e + W_e \leq W$ 和 $y_e + H_e \leq H$ ，我们将区域 $I_e = (x_e, y_e, x_e + W_e, y_e + H_e)$ 设置为选定的矩形区域。否则，我们重复上述过程，直到一个适当的 $I_e$ 被选择。在选定的消除区域 $I_e$ 中， $I_e$ 中的每个像素分别赋给图像 $I$ 的均值。在本研究中，我们分别将hyper参数设置为 $p = 0.5, 0.02 < S_e < 0.4, r_1 = 0.3, r_2 = 3.33$ 。一些例子如图4所示。



### 3.3 Label Smoothing (标签平滑)

ID Embedding (IDE) [25] network is a basic baseline in person ReID. The last layer of IDE, which outputs the ID prediction logits of images, is a fully-connected layer with a hidden size being equal to numbers of persons  $N$ . Given an image, we denote  $y$  as truth ID label and  $p_i$  as ID prediction logits of class  $i$ . The cross entropy loss is computed as:

$$L(ID) = \sum_{i=1}^N -q_i \log(p_i) \begin{cases} q_i = 0, y \neq i \\ q_i = 1, y = i \end{cases} \quad (2)$$

Because the category of the classification is determined by the person ID, we call such loss function as ID loss in this paper.

Nevertheless, person ReID can be regard as one-shot learning task because person IDs of the testing set have not appeared in the training set. So it is pretty important to prevent the ReID model from overfitting training IDs. Label smoothing (LS) proposed in [17] is a widely used method to prevent overfitting for a classification task. It changes the construction of  $q_i$  to:

$$q_i = \begin{cases} 1 - \frac{N-1}{N}\varepsilon & \text{if } i = y \\ \varepsilon/N & \text{otherwise,} \end{cases} \quad (3)$$

where  $\varepsilon$  is a small constant to encourage the model to be less confident on the training set. In this study,  $\varepsilon$  is set to be 0.1. When the training set is not very large, LS can significantly improve the performance of the model.

ID嵌入(IDE)[25]网络是person ReID的基本基线。IDE的最后一层,输出的ID预测分对数图像,是一个全层隐藏大小等于数字的人 $N$ 。给定一个图像,我们表示 $y$ 作为真实ID标签,  $p_i$  为第  $i$  类的 ID 预测的 对数。交叉熵损失计算 如下

$$L(ID) = \sum_{i=1}^N -q_i \log(p_i) \begin{cases} q_i = 0, y \neq i \\ q_i = 1, y = i \end{cases} \quad (2)$$

由于分类的类别是由人的ID来决定的, 所以我们将这种损失函数称为ID损失。然而, 由于测试集的人id没有出现在训练集中, 所以可以将person ReID视为一次性学习任务, 因此防止ReID模型对训练id进行过拟合非常重要。在[17]中提出的标签平滑(LS)是一种广泛用于防止分类任务过拟合的方法。它改变了  $q_i$  的结构

$$q_i = \begin{cases} 1 - \frac{N-1}{N}\varepsilon & \text{if } i = y \\ \varepsilon/N & \text{otherwise,} \end{cases} \quad (3)$$

$\varepsilon$ 的地方是一个小的常数鼓励模型训练集上的缺乏置信度。在这项研究中, $\varepsilon$ 将是0.1。当训练集不是很大时, LS可以显著提高模型的性能。

### 3.4. Last Stride (最后一步)

Higher spatial resolution always enriches the granularity of feature. In [16], Sun *et al.* removed the last spatial down-sampling operation in the backbone network to increase the size of the feature map. For convenience, we denote the last spatial down-sampling operation in the backbone network as last stride. The last stride of ResNet50 is set to be 2. When fed into a image of  $256 \times 128$  size, the backbone of ResNet50 outputs a feature map with the spatial size of  $8 \times 4$ . If change last stride from 2 to 1, we can get a feature map with higher spatial size ( $16 \times 8$ ). This manipulation only increases very light computation cost and does not involve extra training parameters. However, higher spatial resolution brings significant improvement.

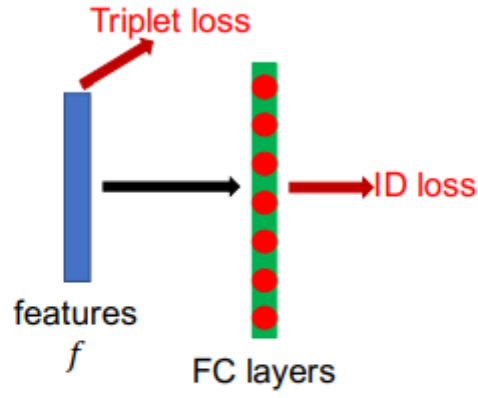
更高的空间分辨率通常会丰富特征的粒度。在[16]中, Sun等人删除了主干网络中的最后一次空间下采样操作, 以增加feature map的大小。为了方便起见, 我们将骨干网络中的最后一次空间下采样操作称为最后一步。ResNet50的最后一步设置为2。当输入一个大小为 $256 \times 128$ 的图像时, ResNet50的主干输出一个空间大小为 $8 \times 4$ 的feature map。如果将最后一步由2步变为1步, 我们可以得到一个空间尺寸更大的feature map( $16 \times 8$ ), 这种操作只会增加非常小的计算量, 并且不需要额外的训练参数。然而, 更高的空间分辨率带来了显著的改善。

### 3.5. BNNeck

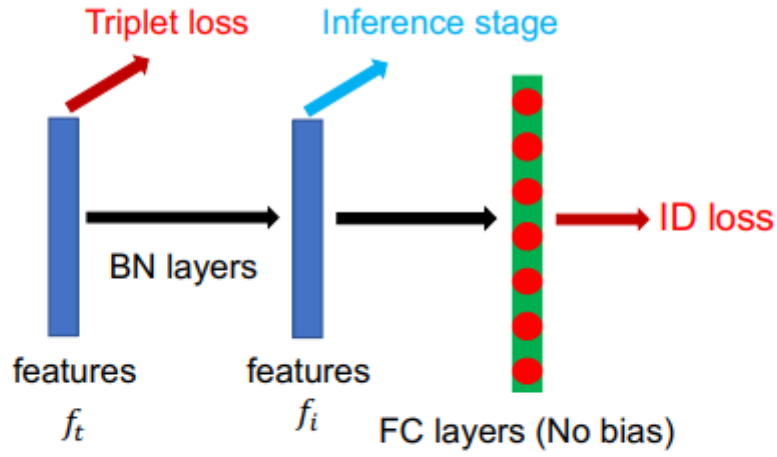
Most of works combined ID loss and triplet loss together to train ReID models. As shown in Fig. 5(a), in the standard baseline, ID loss and triplet loss constrain the same feature  $f$ . However, the targets of these two losses are inconsistent in the embedding space.

As shown in Fig. 6(a), ID loss constructs several hyperplanes to separate the embedding space into different subspaces. The features of each class are distributed in different subspaces. In this case, cosine distance is more suitable than Euclidean distance for the model optimized by ID loss in the inference stage. On the other hand, as shown in 6(b),

大部分工作都是将ID丢失和三重丢失结合起来训练ReID模型。如图5(a)所示, 在标准基线中, ID丢失和三重态丢失约束了相同的特征 $f$ 。但是, 这两个丢失的目标在嵌入空间上是不一致的。如图6(a)所示, ID丢失构造多个超平面, 将嵌入空间划分为不同的子空间。每个类的特性分布在不同的子空间中。在这种情况下, 对于推理阶段的ID丢失优化模型, 余弦距离比欧几里德距离更合适。另一方面, 如6(b)所示



(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.

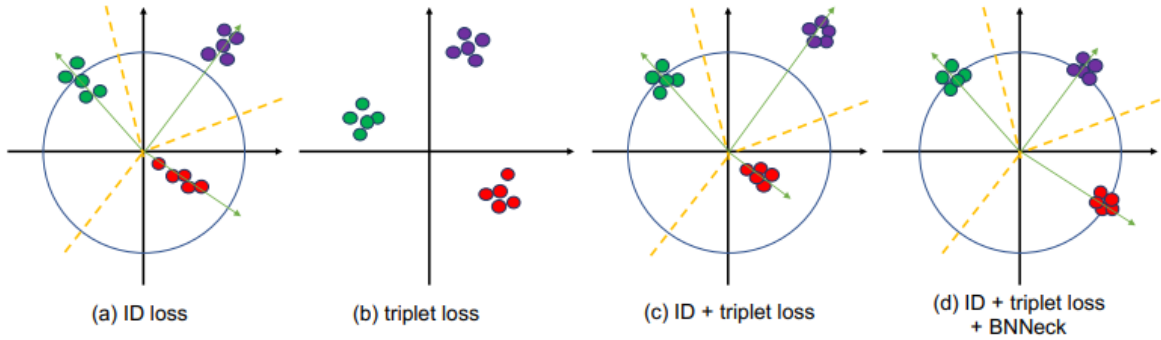


Figure 6. Two-dimensional visualization of sample distribution in the embedding space supervised by (a) ID Loss, (b) Triplet Loss, (c) ID + triplet loss and (d) ID + triplet loss + BNNeck. Points of different colors represent embedding features from different classes. The yellow dotted lines stand for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



triplet loss enhances the intra-class compactness and inter-class separability in the Euclidean space. Because triplet loss can not provide globally optimal constraint, inter-class distance sometimes is smaller than intra-class distance. A widely used method is to combine ID loss and triplet loss to train the model together. This approach let the model learn more discriminative features. Nevertheless, for image pairs in the embedding space, ID loss mainly optimizes the cosine distances while triplet loss focuses on the Euclidean distances. If we use these two losses to simultaneously optimize a feature vector, their goals may be inconsistent. In the training process, a possible phenomenon is that one loss is reduced, while the other loss is oscillating or even increased.

To overcome the aforementioned problem, we design a structure named as BNNeck shown in Fig. 5(b). BNNeck only adds a batch normalization (BN) layer after features (and before classifier FC layers). The feature before the BN layer is denoted as  $f_t$ . We let  $f_t$  pass through a BN layer to acquire the normalized feature  $f_i$ . In the training stage,  $f_t$  and  $f_i$  are used to compute triplet loss and ID loss, respectively. Normalization balances each dimension of  $f_i$ . The features are gaussianly distributed near the surface of the hypersphere. This distribution makes the ID loss easier to converge. In addition, BNNeck reduces the constraint of the ID loss on  $f_t$ . Less constraint from ID loss leads to triplet loss easier to converge at the same time. Thirdly, normalization keeps the compact distribution of features that belong to one same person.

Because the hypersphere is almost symmetric about the origin of the coordinate axis, another trick of BNNeck is removing the bias of classifier FC layer. It constrains the classification hyperplanes to pass through the origin of the coordinate axis. We initialize the FC layer with Kaiming initialization proposed in [4].

In the inference stage, we choose  $f_i$  to do the person ReID task. Cosine distance metric can achieve better performance than Euclidean distance metric. Experimental results in Table. 1 show that BNNeck can improve performance of the ReID model by a large margin.

triplet-loss三重态损失增强了欧氏空间的类内紧性和类间可分性。由于三重态损失不能提供全局最优约束，类间距离有时小于类内距离。一种广泛使用的方法是将身份损失和三重损失结合起来训练模型。这种方法使模型学习到更多的区别性特征。而对于嵌入空间中的图像对，ID损失主要优化余弦距离，triplet-loss主要优化欧氏距离。如果我们使用这两个损失同时优化一个特征向量，它们的目标可能是不一致的。在训练过程中，一个可能的现象是一个损失减少了，而另一个损失却在振荡甚至增加。为了解决上述问题，我们设计了如图5(b)所示的BNNeck结构。BNNeck只在特征之后(在分类器FC层之前)添加了一个batch normalization (BN)层。BN层之前的特征记作  $f_t$ ，我们让  $f_t$  通过 BN 层获得归一化特征  $f_i$ 。在训练阶段， $f_t$  和  $f_i$  分别用于计算三重态损失和ID损失。归一化平衡了  $f_i$  的各个维度。超球面的特征是高斯分布的。这种分布使得ID丢失更容易收敛。此外，BNNeck还减少了ID损失对  $f_t$  的约束，较少的ID损失约束使三重损失同时更容易收敛。第三，归一化保持了属于同一人的特征的紧凑分布。

由于超球面与坐标轴的原点几乎是对称的，所以BNNeck的另一个技巧是消除分级器FC层的偏差。它约束分类超平面通过坐标轴的原点。我们用[4]中提出的 kaiming 初始化来初始化FC层。

在推理阶段，我们选择fi来完成person ReID任务。余弦距离度量比欧几里德距离度量具有更好的性能。实验结果如表1所示，BNNeck可以大幅度提高ReID模型的性能。

### 3.6. Center Loss (中心损失)

Triplet loss is computed as:

$$L_{Tri} = [d_p - d_n + \alpha]_+, \quad (4)$$

where  $d_p$  and  $d_n$  are feature distances of positive pair and negative pair.  $\alpha$  is the margin of triplet loss, and  $[z]_+$  equals to  $\max(z, 0)$ . In this paper,  $\alpha$  is set to 0.3. However, triplet loss only considers the difference between  $d_p$  and  $d_n$  and ignores the absolute values of them. For instance, when  $d_p = 0.3, d_n = 0.5$ , the triplet loss is 0.1. For another case, when  $d_p = 1.3, d_n = 1.5$ , the triplet loss also is 0.1. Triplet loss is determined by two person IDs sampled randomly. It is difficult to ensure that  $d_p < d_n$  in the whole training dataset.

Center loss [20], which simultaneously learns a center for deep features of each class and penalizes the distances between the deep features and their corresponding class centers, makes up for the drawbacks of the triplet loss. The center loss function is formulated as:

$$\mathcal{L}_C = \frac{1}{2} \sum_{j=1}^B \|f_{t_j} - c_{y_j}\|_2^2, \quad (5)$$

where  $y_j$  is the label of the  $j$ th image in a mini-batch.  $c_{y_j}$  denotes the  $y_j$ th class center of deep features.  $B$  is the number of batch size. The formulation effectively characterizes the intra-class variations. Minimizing center loss increases intra-class compactness. Our model totally includes three losses as follow:

$$L = L_{ID} + L_{Triplet} + \beta L_C \quad (6)$$

$\beta$  is the balanced weight of center loss. In our experiments,  $\beta$  is set to be 0.0005.

Triplet loss is computed as:

$$L_{Tri} = [d_p - d_n + \alpha]_+, \quad (4)$$

其中 $d_p$ 和 $d_n$ 是正对和负对的特征距离。 $\alpha$ 是三个一组的保证金损失,和 $[z]_+$ 等价于  $\max(z, 0)$ ,  $\alpha$ 设置为0.3。然而，三重损耗只考虑 $d_p$ 和 $d_n$ 之间的差异，而忽略了它们的绝对值。例如，当 $d_p = 0.3, d_n = 0.5$ 时，三重损耗为0.1。另一种情况下，当 $d_p = 1.3, d_n = 1.5$ 时，三重损耗也是0.1。三重态损耗是由两个随机采样的人的id决定的。很难保证在整个训练数据集中 $d_p < d_n$ 。

中心丢失[20]同时学习每个类的深度特征中心，并惩罚深度特征与其对应的类中心之间的距离，弥补了三重丢失的缺点。中心损失函数表示为

$$\mathcal{L}_C = \frac{1}{2} \sum_{j=1}^B \|f_{t_j} - c_{y_j}\|_2^2, \quad (5)$$

其中  $y_j$  是小批量处理中的第  $j$  个图像的标签。  $c_{yj}$  表示深部特征的第一类中心。  $B$  是批量大小的数量。这个公式有效地描述了阶级内部的变化。最小化中心损失可以增加类内的紧凑性。我们的模型总共包括以下三个损失：

$$L = L_{ID} + L_{Triplet} + \beta L_C \quad (6)$$

$\beta$  是中心的平衡重量损失。在我们的实验中,  $\beta$  将 0.0005。

#### 4. Experimental Results

In this section, we will evaluate our models on Market1501 and DukeMTMC-reID [11] datasets. The Rank-1 accuracy and mean Average Precision (mAP) are reported as evaluation metrics. We add tricks on the standard baseline successively and do not change any training settings. The results of ablation studies present the performance boost from each trick. In order to prevent being misled by overfitting, we also show the results of cross-domain experiments.

在本节中，我们将在Market1501和DukeMTMC-reID[11]数据集上评估我们的模型。以Rank1精度和平均精度(mAP)作为评价指标。我们依次在标准基线上添加技巧，并且不改变任何训练设置。消融研究的结果显示了每个技巧的性能提升。为了避免过拟合的误导，我们还展示了跨域实验的结果。

##### 4.1. Influences of Each Trick (Same domain)

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.

The standard baseline introduced in section 2 achieves 87.7% and 79.7% rank-1 accuracies on Market1501 and DukeMTMC-reID, respectively. The performance of standard baseline is similar with most of baselines reported in other papers. Then, we add warmup strategy, random erasing augmentation, label smoothing, stride change, BNNeck and center loss to the model training process, one by one. Our designed BNNeck boosts more performance than other tricks, especially on DukeMTMC-reID. Finally, these tricks make baseline acquire 94.5% rank-1 accuracy and 85.9% mAP on Market1501. On DukeMTMC-reID, it reaches 86.4% rank-1 accuracy and 76.4% mAP. In other works, these training tricks boost the performance of the standard baseline by more than 10% mAP. In addition, to get such improvement, we only involve an extra BN layer and do not increase training time.

第2节中引入的标准基线对Market1501和DukeMTMC-reID的秩-1精度分别达到87.7%和79.7%。标准基线的性能与其他文献报道的大多数基线相似。然后，我们将热身策略、随机消除增强、标签平滑、步幅变化、BNNeck和中心损失一个接一个地加入到模型训练过程中。我们设计的BNNeck比其他技巧提高更多的性能，特别是在DukeMTMC-reID上。最后，这些技巧使基线获得94.5%的排名-1精度和85.9%的市场地图1501。在DukeMTMC-reID上，其秩1正确率为86.4%，地图正确率为76.4%。在其他工作

中，这些训练技巧将标准基线的性能提高了10%以上。另外，为了得到这样的改善，我们只涉及额外的BN层，不增加训练时间。

#### 4.2. Analysis of BNNeck

In this section, we evaluate the performance of two different features ( $f_t$  and  $f_i$ ) with Euclidean distance metric and cosine distance metric. All models are trained without center loss in Table. 2. We observe that cosine distance metric performs better than Euclidean distance metric for  $f_i$ . Because ID loss directly constrains the features followed

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.

the BN layer,  $f_i$  can be separated by several hyperplanes clearly. The cosine distance can measure the angle between two feature vectors, so cosine distance metric is more suitable than Euclidean distance metric for  $f_i$ . However,  $f_t$  is close to triplet loss and is constrained by ID loss at the same time. Two kinds of metrics achieve similar performance for  $f_t$ .

In overall, BNNeck significantly improve the performance of ReID models. We choose  $f_i$  with cosine distance metric to do the retrieval in the inference stage.

在本节中，我们使用欧氏距离度量和余弦距离度量来评估两个不同特征( $f_t$ 和 $f_i$ )的性能。表2所有模型均经过训练，无中心损失。我们观察到，在 $f_t$ 中，余弦距离度量比欧几里德距离度量表现得更好。由于ID损失直接限制了BN层之后的特征， $f_i$ 可以被多个超平面清晰地分割。余弦距离可以度量两个特征向量之间的夹角，因此余弦距离度量比欧几里德距离度量更适合用于 $f_i$ ，但是 $f_t$ 更接近于三重损失，同时也受到ID损失的限制。两种指标对 $f_t$ 的性能基本相同。总体而言，BNNeck显著提高了ReID模型的性能。在推理阶段，我们选择带余弦距离度量的 $f_i$ 进行检索。

#### 4.3. Influences of Each Trick (Cross domain) 每一trick的影响(跨域)



To further explore effectiveness, we also present the results of cross-domain experiments in Table. 3. In overview, three tricks including warmup strategy, label smoothing and BNNeck significantly boost the cross-domain performance of ReID models. Stride change and center loss seem to have no big impact on the performance. However, REA does harm to models in cross-domain ReID task. In particular, when our modified baseline is trained without REA, it achieves 41.4% and 54.3% rank-1 accuracies on Market1501 and DukeMTMC-reID datasets, respectively. Its performance surpass the ones of the standard baseline by a large margin. We infer that REA masking the regions of training images lets the model learn more knowledge in the training domain. It causes the model to perform worse in

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.

为了进一步探讨其有效性，我们还将跨域实验结果如表3所示。综上所述，热身策略、标签平滑和BNNeck三种技巧显著提高了ReID模型的跨域性能。步幅的变化和中锋的损失似乎对表现没有太大的影响。然而，在跨域ReID任务中，REA对模型有一定的危害。特别是，当我们的修改基线在没有REA的情况下进行训练时，它在Market1501和DukeMTMC-reID数据集上的排名-1正确率分别达到41.4%和54.3%。它的性能大大超过了标准基线的性能。我们推断，REA掩盖训练图像的区域可以让模型在训练领域学到更多的知识。它会导致模型在测试领域的表现变差。



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	Manes [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-arts methods.  $N_f$  is the number of features used in the inference stage. RK stands for  $k$ -reciprocal re-ranking method [26]

#### 4.4. Comparison of State-of-the-Arts

我们比较了强基线与最先进的方法在表4。所有的方法都被分成不同的类型。pyramid[22]实现了两个数据集的惊人性能。但是，它连接了21个不同尺度的局部特征。若仅利用全球特征，则可获得92.8%的rank-1精度和82.1%的market - 1501 mAp。我们的强大基线可以达到94.5%的rank1精度和85.9%的市场地图1501。BFE[1]的性能与我们的强基线类似。但是它结合了两个分支的特性。在所有只使用全局特性的方法中，我们的强基线在 Market1501 和DukeMTMC-reID 上都比 AWTL[12] 强10%以上。通过k-reciprocal（倒数k个）重链方法来提高性能，我们的方法在Market1501和DukeMTMC-reID上分别达到了94.1%和89.1%的mAP。据我们所知，在仅使用全局特性的情况下，我们的基线实现了最佳性能。

#### 5. Supplementary Experiments 补充实验

我们观察到之前的一些工作是用不同数量的批大小或图像大小完成的。在本节中，作为补充，我们将探讨它们对模型性能的影响。

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.

### 5.1. Influences of the Number of Batch Size

The mini-batch of triplet loss includes  $B = P \times K$  images.  $P$  and  $K$  denote the number of different persons and the number of different images per person, respectively. A mini-batch can only contain up to 128 images in one GPU, so that we can not do the experiments with  $P = 32, K = 6$  or  $P = 32, K = 8$ . We removed center loss to clearly find the relation between triplet loss and batch size. The results are present in Table. 5. However, there are not specific conclusions to show the effect of  $B$  on performance. A slight trend we observed is that larger batch size is beneficial for the model performance. We infer that large  $K$  helps to mine hard positive pairs while large  $P$  helps to mining hard negative pairs.

### 5.2. Influences of Image Size

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.

We trained models without center loss and set  $P = 16, K = 4$ . As shown in Table. 6, four models achieve similar performances on both datasets. In our opinion, the image size is not a pretty importance factor for the performance of ReID models.

#### 5.1. Influences of the Number of Batch Size（影响批量大小的因素）

小批量的三重态损失 包括  $B = P * K$  个图像。P和K分别表示每个人的不同数量和不同图像的数量。一个小批处理在一个GPU中最多只能包含128张图像，因此我们不能在  $P = 32, K = 6$  或  $P = 32, K = 8$  的情况下进行实验。我们移除了中心损失，以清楚地找出三重损失与批次大小之间的关系。结果见表5。但是，目前还没有具体的结论来说明B对性能的影响。我们观察到的一个轻微的趋势是，更大的批大小有利于模型的性能。我们推断，大K有助于挖掘硬阳性对，而大P有助于挖掘难的负类对。

#### 5.2. Influences of Image Size

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.

我们训练了无中心损失的模型，令  $P = 16, K = 4$ 。如表6所示，四个模型在两个数据集上实现了类似的性能。在我们看来，图像大小并不是ReID模型性能的一个非常重要的因素。

### 6. Conclusions and Outlooks（结论和展望）

在本文中，我们收集了一些有效的训练技巧，并为ReID设计了一个强基线。为了证明每种技巧对ReID模型性能的影响，我们在相同的域和跨域ReID任务上做了大量的实验。最后，仅使用全局功能，我们的强大基线实现了94.5%的排名-1的准确性和85.9%的市场地图1501。我们希望这项工作能够促进ReID在学术界和工业界的研究。然而，我们工作的目的不是粗略地提高性能。与人脸识别相比，人里德还有很长的路要走。我们认为一些训练技巧可以加速探索，还有很多有效的技巧没有被发现。我们欢迎研究人员与我们分享一些其他有效的技巧。今后我们将根据这项工作对它们进行评估。

In this paper, we collect some effective training tricks and design a strong baseline for person ReID. To demonstrate the influences of each trick on the performance of ReID models, we do a lot of experiments on both same-domain and cross-domain ReID tasks. Finally, only using global features, our strong baseline achieve 94.5% rank-1 accuracy and 85.9% mAP on Market1501. We hope that this work can promote the ReID research in academia and industry.

However, the purpose of our work is not to improve performance roughly. Compared with face recognition, person ReID still has a long way to explore. We think some training tricks can speed up the exploration and there are many effective tricks not discovered. We welcome researchers to share some other effective tricks with us. We will evaluate them based on this work in the future.

## References



global features, our strong baseline achieve 94.5% rank-1 accuracy and 85.9% mAP on Market1501. We hope that this work can promote the ReID research in academia and industry.

However, the purpose of our work is not to improve performance roughly. Compared with face recognition, person ReID still has a long way to explore. We think some training tricks can speed up the exploration and there are many effective tricks not discovered. We welcome researchers to share some other effective tricks with us. We will evaluate them based on this work in the future.

## References

- [1] Z. Dai, M. Chen, S. Zhu, and P. Tan. Batch feature erasing for person re-identification and beyond. *arXiv preprint arXiv:1811.07130*, 2018. 7
- [2] X. Fan, W. Jiang, H. Luo, and M. Fei. Spherereid: Deep hypersphere manifold embedding for person re-identification. *Journal of Visual Communication and Image Representation*, 2019. 3
- [3] X. Fan, H. Luo, X. Zhang, L. He, C. Zhang, and W. Jiang. Scpnnet: Spatial-channel parallelism network for joint holistic and partial person re-identification. *arXiv preprint arXiv:1810.06996*, 2018. 7
- [4] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015. 5
- [5] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2
- [6] A. Hermans, L. Beyer, and B. Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017. 2, 7
- [7] M. M. Kalayeh, E. Basaran, M. Gökmen, M. E. Kamasak, and M. Shah. Human semantic parsing for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1062–1071, 2018. 7
- [8] W. Li, X. Zhu, and S. Gong. Harmonious attention network for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2285–2294, 2018. 7
- [9] L. Qi, J. Huo, L. Wang, Y. Shi, and Y. Gao. Maskreid: A mask based deep ranking neural network for person re-identification. *arXiv preprint arXiv:1804.03864*, 2018. 7
- [10] X. Qian, Y. Fu, T. Xiang, W. Wang, J. Qiu, Y. Wu, Y.-G. Jiang, and X. Xue. Pose-normalized image generation for person re-identification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 650–667, 2018. 7
- [11] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *European Conference on Computer Vision workshop on Benchmarking Multi-Target Tracking*, 2016. 6
- [12] E. Ristani and C. Tomasi. Features for multi-target multi-camera tracking and re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6036–6046, 2018. 7
- [13] M. Saquib Sarfraz, A. Schumann, A. Eberle, and R. Stiefelhagen. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 7
- [14] J. Si, H. Zhang, C.-G. Li, J. Kuen, X. Kong, A. C. Kot, and G. Wang. Dual attention matching network for context-aware feature sequence based person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5363–5372, 2018. 7
- [15] Y. Sun, L. Zheng, W. Deng, and S. Wang. Svdnet for pedestrian retrieval. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3800–3808, 2017. 7
- [16] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 480–496, 2018. 4, 7
- [17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016. 4
- [18] C. Wang, Q. Zhang, C. Huang, W. Liu, and X. Wang. Mancs: A multi-task attentional network with curriculum sampling for person re-identification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 365–381, 2018. 7
- [19] L. Wei, S. Zhang, H. Yao, W. Gao, and Q. Tian. Glad: Global-local-alignment descriptor for pedestrian retrieval. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 420–428. ACM, 2017. 7
- [20] Y. Wen, K. Zhang, Z. Li, and Y. Qiao. A discriminative feature learning approach for deep face recognition. In *European conference on computer vision*, pages 499–515. Springer, 2016. 5
- [21] X. Zhang, H. Luo, X. Fan, W. Xiang, Y. Sun, Q. Xiao, W. Jiang, C. Zhang, and J. Sun. Alignedreid: Surpassing human-level performance in person re-identification. *arXiv preprint arXiv:1711.08184*, 2017. 7
- [22] F. Zheng, X. Sun, X. Jiang, X. Guo, Z. Yu, and F. Huang. A coarse-to-fine pyramidal model for person re-identification via multi-loss dynamic training. *arXiv preprint arXiv:1810.12193*, 2018. 7
- [23] L. Zheng, Y. Huang, H. Lu, and Y. Yang. Pose invariant embedding for deep person re-identification. *arXiv preprint arXiv:1701.07732*, 2017. 7
- [24] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian. Scalable person re-identification: A benchmark. In *Computer Vision, IEEE International Conference*, 2015. 1
- [25] Z. Zheng, L. Zheng, and Y. Yang. A discriminatively learned cnn embedding for person reidentification. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(1):13, 2018. 4, 7
- [26] Z. Zhong, L. Zheng, D. Cao, and S. Li. Re-ranking person re-identification with k-reciprocal encoding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1318–1327, 2017. 7
- [27] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang. Random erasing data augmentation. *arXiv preprint arXiv:1708.04896*, 2017. 3
- [28] Z. Zhong, L. Zheng, Z. Zheng, S. Li, and Y. Yang. Cam-style: A novel data augmentation method for person re-identification. *IEEE Transactions on Image Processing*, 28(3):1176–1190, 2019. 7