# 人流检测的关键问题

###### 人流检测遮挡问题

在目标检测中，遮挡问题是比较常见的，主要分为两种，一种是其他物体对行人的遮挡，这往往会带来目标信息的缺失，进而导致漏检；另一种是行人个体之间的相互遮挡，这往往会引入大量的干扰信息，进而导致更多的虚检。

第一种遮挡，由于目标干扰物体遮挡，而算法只能学习待检测物体的特征，因此第一种遮挡只能通过增加样本来优化检测效果。第二种遮挡又分为类间遮挡和类内遮挡，理解为小孩与大人重叠时，小孩的头部与大人的衣物间遮挡为类间遮挡，类内遮挡理解为是头部重叠，身体部位重叠，衣物重叠等。类内遮挡产生于同类物体，也被称为密集遮挡，由于密集遮挡的两个目标的类别是相同的，所以两个目标之间的特征是相似的，检测器很可能无法定位。重点解决的是后一种情况导致的遮挡问题。

1.1方法一：

从loss的层面提出了一种新颖的解决方案，提出了一个全新的定位技术­——Repulsion Loss（RepLoss）。



图一：RepLoss图示：预测框靠近B会受到惩罚，所以RepLoss可以有效防止预测边界框移向相邻的重叠物体，提升检测器在人群场景中的鲁棒性。

* 1. 方法二：

提出聚集损失（AggLoss）来改进密集遮挡问题







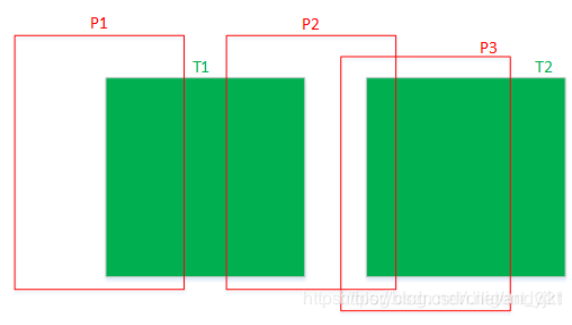
基于Faster-RCNN框架从两个方面解决遮挡问题，针对第二种遮挡情况，为了消除链接处的误检测，提出AggLoss使的指定的相同目标的候选框尽可能的靠近，最小化这些候选框的间隔。



图二：为遮挡情况连接处的误检测

1.3 总结：

问题描述：



图三：两种误检测情况：1、p1,p2,p3均存在，不精准；2、仅p2,p3存在，由于对应的max score的候选框为p3，当p2,p3的IOU均大于阈值时，后处理NMS阶段就会把p2过滤掉，则T1被漏召回了。

结论：解决遮挡问题可以从模型和损失上进行创新，从损失上改进目标遮挡的误检情况，从模型结构上解决非目标的漏检情况。

###### RE-ID去冗余问题

人员重识别去冗余问题是在双目摄像机检测到的视图有重叠部分，要将重叠部分的人员冗余去掉。在本项目中存在的去冗余问题有很多外界因素影响，例如：光线变化，由于相机拍摄时差引起的人员姿势微变化，拍摄视点不同，效果模糊，遮挡和背景杂波在相机视图中的复杂变化等。

2.1 人员重识别系统需要五个主要步骤：

1. 特定场景下原始数据集采集；
2. 候选框生成，边界框通常由行人检测或跟踪算法获得；
3. 训练数据标注：标注跨摄像机标签，由于摄像机之间的差异较大，训练数据标注对于有区别的重识别模型学习必不可少；
4. 训练模型：用标注的图像训练重识别模型；
5. 行人检索：测试阶段进行行人检索。给定一个感兴趣的人(查询)和一个图库集，使用前一阶段学习的重标识模型提取特征表示。通过对计算出的查询到图库的相似性进行排序来获得检索到的排名列表。

2.2 一个封闭空间再识别系统包含三个主要组成部分：

（1）、特征表示学习，侧重于开发特征构建策略；

（2）、深度度量学习，用于设计不同的损失函数或抽样策略的训练目标；

（3）、排名优化，其集中于优化检索到的排名列表。

2.2.1特征表示学习：

a)全局特征，它为每个人图像提取一个全局特征表示向量，而不需要额外的注释提示；

b)局部特征，它聚集部分级别的局部特征，以形成每个人物图像的组合表示；

c)辅助特征，其使用辅助信息，例如属性、GAN生成的图像等，来改进特征表示学习。

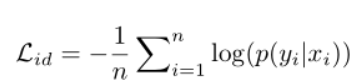
d)视频特征，它使用多个图像帧和时间信息学习基于视频的重标识的视频表示。

2.2.2深度度量学习:

度量学习的作用已经被损失函数设计所取代，以指导特征表示学习。以下面这篇论文为例说明损失函数的设计：A Discriminatively Learned CNN Embedding for Person Re-identification

文献中有三种广泛研究的损失函数及其变体，包括身份损失、验证损失和三重损失。

身份损失：它将人重新身份的训练过程视为一个图像分类问题，即每个身份是一个不同的类。在测试阶段，采用汇集层或嵌入层的输出作为特征提取器。给定一个带有标签yi的输入图像xi，被识别为类别yi的预测概率用softmax函数编码，通过交叉熵计算身份损失：



验证损失：它通过对比损失或二元验证损失来优化成对关系。对比损失改善了相对成对距离比较。验证损失通常与身份丢失相结合，以提高性能：



三重损失。它将重标识模型训练过程视为一个检索排序问题。基本思想是正对之间的距离应该比负对小一个预定的余量。通常，一个三元组包含一个anchor样本xi，一个相同身份的正样品xj，和一个不同身份的负样本xk。带有余量参数的三重态损失由下式表示：



扩展：一些方法还研究了信息三元组挖掘的点对集相似策略。提出了一种用于深度度量学习的点对集合相似度，用点对集合度量代替了点到点的距离。

2.3排名优化

在测试阶段，排序优化对提高检索性能起着至关重要的作用。给定初始排名列表，它通过自动图库间相似性挖掘或人工交互来优化排名顺序。等级/度量融合是另一种利用多个等级列表输入来提高等级性能的流行方法。

2.3.1 方法一：2014 DeepReID: Deep Filter Pairing Neural Network for Person Re-Identification

首先利用卷积核进行特征提取，然后将行人图像分为很多个块输出位移矩阵，这些唯一矩阵编码了不同特征下的块匹配模式。为了提高块匹配的鲁棒性，增加了maxout分组层，最后每个卷积核对应一个特征。其中加入了部件位移和姿势视点的建模，将行人各种姿势和视点的变换看作是零件位移的不同组合，其分布是多模态的。最后使用一个softmax层确定两个输入是否属于同一个人。

2.3.2 方法二：2020 STNReID: Deep Convolutional Networks With Pairwise Spatial Transformer Networks for Partial Person Re-Identification STNReID

提出一种新型的REID框架（STNReID），STNReID包括一个空间转换器网络（STN）模块和一个ReID模块。STN模块利用局部和整体图像的高级ReID特征，并预测二维仿射变换器的参数，如两幅图像的大小调整（缩放）、旋转和反射。然后，STN模块从整体图像中采样仿射图像以匹配部分图像。ReID模块提取仿射图像和部分图像的全局特征，以检索目标人物图像。经过训练的局部图像是从整体的人物数据集生成的。因此，STNReID可以通过标准的整体ReID数据集进行训练，而不需要额外的标记数据。然而，STN模块的性能随着ReID模块功率的增加而降低。这样的两阶段培训可以获得强大的STNReID模型。在推理阶段，STNReID可以将某个局部图像与多个整体图像进行批量匹配。这种端到端的一对多匹配比一对一匹配更有效。

2.3.3 方法三：人群图像通常从不同的视角拍摄，导致了各种各样的视角和尺度的变化。靠近相机的人通常会被捕捉到大量的细节，比如他们的脸，有时他们的整个身体都会被捕捉到。然而，当人们远离相机或从空中拍摄图像时，每个人只表现为一个头部斑点。在这两种情况下，对人的有效检测要求该模型同时在高度语义级别(面孔/身体探测器)上操作，同时还能识别低级的头部斑点模式，使用深度和浅层卷积神经网络的组合实现了这一点。

在模型中，目标是识别低级的头部斑点模式，从远离相机的人产生，使用浅卷积网络。由于Blob检测不需要捕获高级语义，将该网络设计为浅层，深度仅为3个卷积层。每一层都有24个过滤器，内核大小为5×5。为了使该网络预测的空间分辨率与深度网络预测的空间分辨率相等，在每个卷积层之后都使用池化层。浅层网络主要用于检测小的头部斑点。为了确保不会因为最大池化而丢失计数，在浅层网络中使用平均池化层。

2.4 结论：重识别主要进行人员特征提取，并进行匹配检测，难点在于在不同摄像头视角下，人员的视图大小不一致，角度不同，姿态不同，光线不同，在做特征匹配时，要克服这些问题。

# 研究现状和最新成果

###### 1. 研究现状

1.1 目标检测算法

自2006年起，包括目标检测算法在内的计算机视觉领域的各个研究方向都在得到了快速的发展，达到了新的高度，这得益于深度学习的提出。R-CNN概念，即区域范围内的卷积神经网络，其是第一次有研究者成功将深度学习应用在目标检测任务中。R-CNN首先使用卷积神经网络提取预选框内所有图像像素的特征，替代了Haar特征等传统方法，之后使用分类器对预选框的特征进行分类，得到最终的检测结果。R-CNN的思路是先得到预选框，然后对预选框的范围内图像的特征进行分类，这种将整个目标检测算法流程拆分为前后两步的方法被称为Two-stages的目标检测算法。2015年，R-CNN的设计者，Ross Girshick等人，在R-CNN的基础上，提出了Fast R-CNN。Fast R-CNN将深度神经网络的backbone替换为性能更优的VGG-16，检测算法的精度和速度相比前作都有提升。同样是在2015年，Ross Girshick等人又提出了Faster R-CNN。Faster R-CNN是在Fast R-CNN的基础上引入了一个可以同时预测每个待检测目标的边界框在像素坐标系下的坐标和对应的边界框内是否存在目标的置信度得分的全卷积网络，该网络在文章中被称为Region Proposal Network。这一改进使得算法能够更准确的定位待检测的目标，其优越的性能也使得R-CNN系列算法成为了目标检测领域中Two-stages算法的代表。

YOLO系列的第一个版本的算法和Faster R-CNN 在同年被提出，其开辟了不同于R-CNN系列算法将定位检测框的位置和分类检测框内目标类别分为两个步骤的思路。YOLO系列算法是一阶段的目标检测算法，其全称完美地概述了YOLO系列算法的特点，You Only Look Once意为只经过一次操作便能“看到”最终的待检测结果，即预测检测框的位置和目标类别的分类在一步中同时完成，这种思路被称为One-Stage的目标检测算法。除了YOLO系列算法，2017年被Kaiming He 等人提出的Retina Net 网络、2016年被 W Liu等人提出的SSD(Single Shot MultiBox Detector）算法等，都是One-Stages 算法的代表。

上述的目标检测算法，不论是Two-Stages的策略还是 One-Stages的策略，都是基于anchor的，即需要提前设置先验框的位置、大小、和数量。该类目标检测算法通过预测待检测目标的边界框相对于先验框的偏移量来得到待检测目标在图像中的位置。由于一般先验框的数量需要设置的足够多，这样才能保证算法对图像各个位置存在的目标都能得到较精准的检测结果，因此基于anchor的目标检测算法会得到大量的预选框，并对每个预选框中的物体进行分类，另外还需要附加非极大值抑制等后处理操作来过滤掉冗余的预选框。这使得该类算法推理过程效率较低。近年来，有研究者针对上述中基于anchor的目标检测算法存在的问题提出了anchor-freel的概念。Anchor-free 的目标检测算法不再通过预测待检测目标的边界框相对于先验框的偏移量来实现对图像中待检测目标的定位，而是通过直接预测图像中待检测目标的中心点位置以及待检测目标的尺寸信息。这使得anchor-free 的目标检测算法不再需要提前手动设置先验框信息，去除了先验框的约束。2019年由XingyiZhou等人提出的名为CenterNet的anchor-free的目标检测算法，在常见的各类目标检测数据集上的性能都和基于anchor的目标检测算法的最优性能相当，正式开启了目标检测算法的Anchor-free时代。

1.2 行人重识别算法

行人重识别算法，即 Reid算法(Person re-identification)，其算法的目的是将一段视频中所有的行人目标按身份归类，也就是将一段视频中所有身份相同的行人匹配起来。行人重识别算法在当前的计算机视觉算法中落地应用最为广泛。不同于图像风格迁移、表情迁移、动作迁移等，仅仅为了娱乐大众而落地，最终昙花一现的计算机视觉算法，行人重识别算法已经在智能监控、智能安防等领域找到了最适合自己的落地方法，是真正能够通过落地带来价值和收益的计算机视觉算法。2016年 Zhedong Zheng 等人验证了将基于深度神经网络的验证模型和分类模型分别应用在 Reid任务中的可行性，并结合验证模型和分类模型提出了一种能够更好地区分不同行人目标之间的描述符的模型结构。同年，Mengyue Geng 等人将迁移学习(Transfer Learning)的方法引入 Reid 任务，进一步将深度学习和Reid任务结合起来，并提高了Reid算法的性能。

###### 最新成果

张等人提出了一种深度卷积神经网络、多任务、基于块的人群计数方法。该方法甚至适用于将不存在于数据集中的新场景作为输入图像引入的情况。网络是交替训练与人群计数和密度估计图。当一个新的场景被引入输入时，经过训练的网络对图像进行微调，以预测人群数量。从而使网络适应更多的场景

Zhang等人提出了一种多列卷积神经网络，该网络适用于由于透视失真导致的头部大小变化。他们收集了1198张图片和33万张人头。多列体系结构有3列，接收字段的过滤器有大、中、小三种尺寸。因此，由于透视失真的误差可以减少。

Sindagi等[10]提出了一种端到端多任务级联CNN结构，用于人群计数和密度估计。获取密度图是高度精细化的，只受到较低的计数误差。该方法具有两个并行网络、高水平先验阶段和一个密度估计阶段。这些层共享一组初始卷积层。它由两层组成，每层之后是PReLU激活层。两个并行网络由4个卷积层组成，每个卷积层后面都有PReLU激活层。

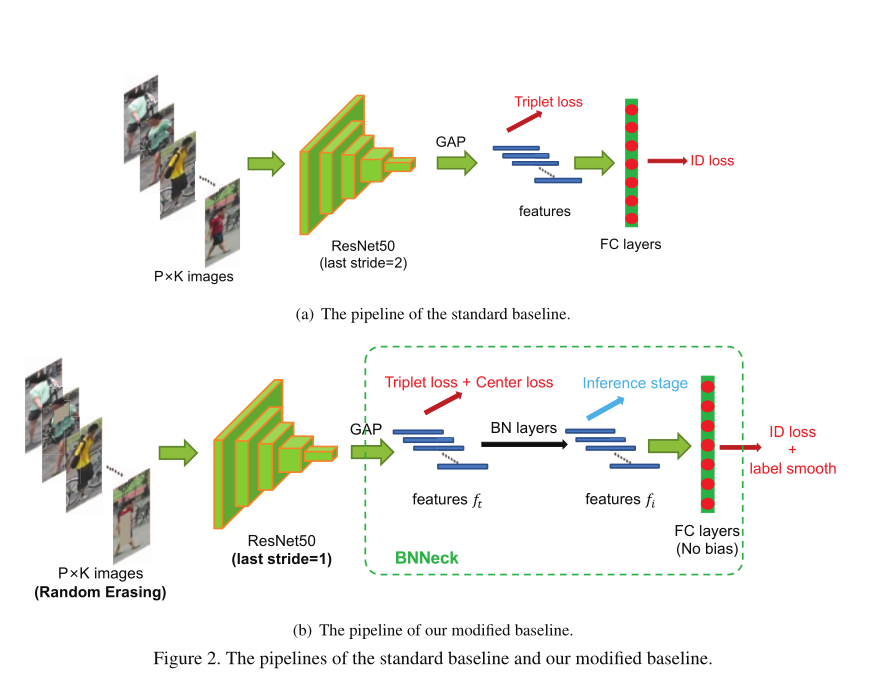
DecideNet是一个端到端的人群统计框架，它利用了检测和回归。在拥挤区域多采用基于回归的方法，而在不太密集的人群场景中采用基于检测的方法更准确。DecideNet分别生成基于检测和基于回归的密度图该体系结构由三个模块组成:RegNet、DetNet、QualityNet。RegNet和DetNet分别提供了两种类型的密度图，密度图将样本放大到输入图像Ii的大小。然后，密度图和Ii被堆叠起来，并作为输入提供给QualityNet。最终得到的密度图Di作为QualityNet的输出。

ADCrowdNet由两个连接的网络组成。第一个是注意力感知网络，叫做注意力地图生成器(AMG)。它检测人群区域并计算交通拥堵的程度。第二种是多尺度可变形网络密度图估计器(DME)。DME生成高质量密度图。ADCrowdNet对各种噪音有很强的抵抗力，能更有效地捕捉人群特征。AMG将图像分为背景图像和人群图像。得到拥挤区域较高值的注意图，它表示拥挤程度。将输入图像与注意图的像素级乘积作为DME网络的输入。提取底层特征前端使用10层经过训练的VGG-16模型。后端使用与初始模块相关的结构，其中多尺度的可变形卷积层有助于克服遮挡，由于透视视图的失真，人群分布的变化。

CP-CNN方法明确地结合了人群图像的全局和局部上下文信息进行人群估计，生成高质量的人群密度图。CP-CNN由四个模块组成。它们是全局上下文估计器(GCE)，局部上下文估计器(LCE)，密度图估计器(DME)和一个融合CNN(FCNN)。该网络使用GCE合并全局上下文，使用LCE合并本地上下文。他们是基于CNN的网络。GCE和LCE将输入图像分为五类。它们是，极高密度(ex-hi)，高密度(hi)，中等密度(med)，低密度(ex-lo)，极低密度(ex-lo)。DME它将输入图像转换为高维特征图。最后，通过GCE和LCE获得上下文信息，并结合DME获得的高维特征图。这样可以生成高质量和高分辨率的密度图。

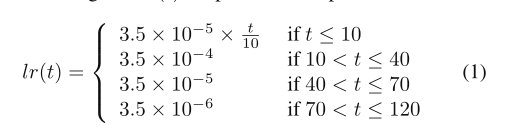
# 训练技巧

本节将介绍一些有效的训练技巧ReID。大多数此类技巧可以在标准基线上扩展，而无需更改模型体系结构。图2（b）显示了本节中出现的训练策略和模型架构。



###### 1. 预热学习率

学习率对ReID模型的性能有很大影响。标准基线最初以较大且恒定的学习率进行训练。在[2]中，采用预热策略引导网络以获得更好的性能。在实践中，如图3所示，我们花了10个epoch，从3.5×10−5线性增加了学习率至3.5×10−4。然后，第40个epoch和第70个epoch的学习速率分别衰减到3.5×10−5和3.5×10−6 。第t个 epoch的学习率lr（t）计算为:



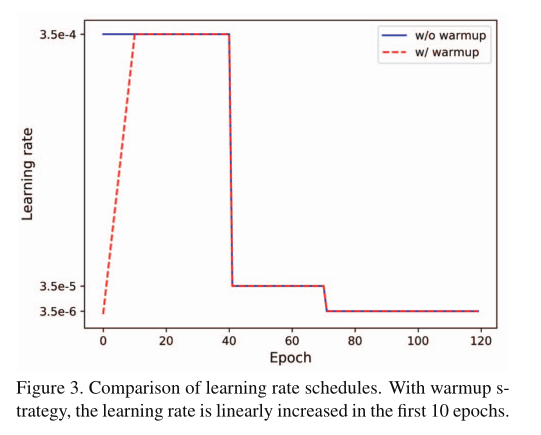


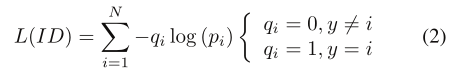
图3：学习率的比较。使用热身策略，前10个阶段的学习率呈线性增加。

###### 2. 随机擦除增强

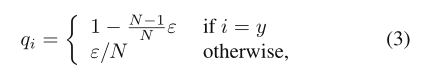
在人物ReID中，图像中的人物有时会被其他物体遮挡。为了解决遮挡问题并提高ReID模型的泛化能力，Zhong等人[1]提出了一种新的数据增强方法，称为随机擦除增强（REA）。实际上，对于小批量中的图像I，其经历随机擦除的概率为pe，保持不变的概率为1−pe。然后，REA随机选择图像I中大小为（We，He）的矩形区域Ie，并用随机值擦除其像素。假设图像I和区域Ie的面积分别为S=W×H和Se=We×He，我们将re=Se/S表示为擦除矩形区域的面积比。此外，区域Ie的纵横比在r1和r2之间随机初始化。为了确定唯一区域，REA随机初始化点P=（xe，ye）。如果xe+We≤ W和ye+He≤ H、 我们将区域Ie=（xe，ye，xe+we，ye+He）设置为选定的矩形区域。否则，我们将重复上述过程，直到选择合适的Ie。对于选定的擦除区域Ie，Ie中的每个像素都被分配给图像I的值的平均值。

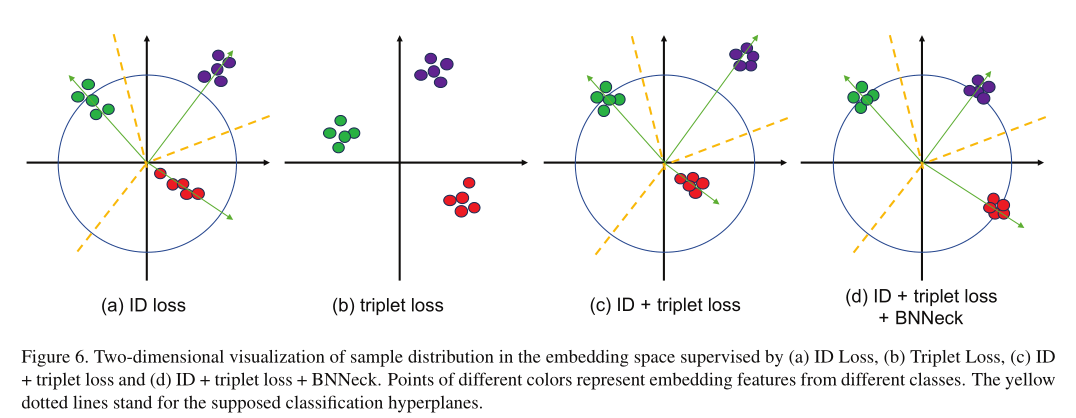
图4：随机擦除增强的示例。第一行显示五个原始训练图像。处理后的图像显示在第二低层。

###### 3. 标签平滑

ID嵌入（IDE）[2]网络是人员ReID的基本基线。IDE的最后一层输出图像的ID预测逻辑，它是一个完全连接的层，隐藏大小等于人数N。给定一幅图像，我们将y表示为真ID标签，将pi表示为i类ID预测逻辑。交叉熵损失计算如下：

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。然而，由于测试集的人员ID没有出现在训练集中，因此人员ReID可以被视为一次性学习任务。因此，防止ReID模型过度拟合训练ID非常重要。文献[3]中提出的标签平滑（LS）是一种广泛使用的方法，用于防止分类任务的过度拟合。它将qi的结构更改为：



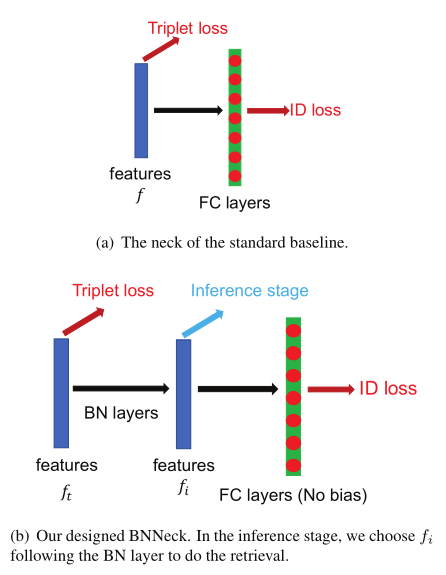
式中，ε是一个小常数，以鼓励模型在训练集上置信不足。当训练集不是很大时，LS可以显著提高模型的性能。

由（a）ID损失、（b）三重态损失、（c）ID+三重态损失和（d）ID+三重态损失+BNNeck监督的嵌入空间中样本分布的二维可视化。不同颜色的点表示不同类的嵌入特征。黄色虚线代表假定的分类超平面

###### 4. 最后一步

较高的空间分辨率通常会丰富特征的粒度。在[4]中，Sun等人删除了骨干网络中的最后一次空间降采样操作，以增加特征图的大小。为方便起见，我们将骨干网络中的最后一次空间下采样操作表示为最后一步。ResNet50的最后一步设置为2。当输入256×128大小的图像时，ResNet50的主干输出空间大小为8×4的特征地图。如果将最后一步从2改为1，可以得到空间尺寸更大（16×8）的特征图。这种操作只会增加很小的计算成本，并且不涉及额外的训练参数。然而，更高的空间分辨率带来了显著的改善。

###### 5. BNNeck

大多数工作将ID损失和三重态损失结合起来训练ReID模型。如图5（a）所示，在标准基线中，ID损失和三重态损失约束相同的特征f。然而，这两种损失的目标在嵌入空间中不一致。如图6（a）所示，ID损失构造了多个超平面，将嵌入空间分隔为不同的子空间。每个类的特征分布在不同的子空间中。在这种情况下，余弦距离比欧几里德距离更适合于推理阶段由ID损失优化的模型。另一方面，如6（b）所示，三重态损失增强了欧氏空间中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ID损失和三重态损失相结合来训练模型。这种方法可以让模型学习更多的鉴别特征。然而，对于嵌入空间中的图像对，ID损失主要优化余弦距离，而三重损失主要优化欧氏距离。如果我们使用这两个损失同时优化特征向量，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个损失减少，而另一个损失振荡甚至增加。

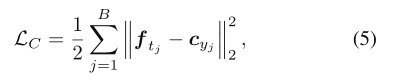
为了克服上述问题，我们设计了一个名为BNNeck的结构，如图5（b）所示。BNNeck仅在特征之后（分类器FC层之前）添加一个批标准化（BN）层。BN层之前的特征表示为ft。我们让ft通过BN层来获取归一化特征fi。在训练阶段，ft和fi分别用于计算三重态损失和ID损失。标准化平衡了fi的各个维度。特征在超球面附近呈高斯分布。这种分布使得ID损失更容易收敛。此外，BNNeck减少了ID损失对ft的约束。ID损失的约束越小，则三重态损失越容易同时收敛。第三，归一化保持属于同一个人的特征的紧凑分布。

由于超球体几乎与坐标轴原点对称，BNNeck的另一个技巧是消除分类器FC层的偏差。它约束分类超平面通过坐标轴的原点。我们使用[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。余弦距离度量可以获得比欧几里德距离度量更好的性能。实验结果见表。1表明BNNeck可以大幅度提高ReID模型的性能。

**6. 中心损失**  
三重态损失计算如下：



其中dp和dn是正对和负对的特征距离。α是三重态损失的裕度，[z] +等于最大值（z，0）。然而，三重态损失只考虑了dp和dn之间的差异，而忽略了它们的绝对值。例如，当dp=0.3，dn=0.5时，三重态损失为0.1。对于另一种情况，当dp=1.3，dn=1.5时，三重态损失也为0.1。三重态损失由随机抽取的两个人ID确定。很难确保整个训练数据集中的dp< dn。中心损失（Center loss）[5]可以同时为每个类的深层特征学习一个中心，并惩罚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。中心损失函数公式如下：

其中，yj是小批量中第j个图像的标签。Cyj代表第yj级深部特征中心。B是批量大小的数量。该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紧凑性。我们的模型共包括以下三种损失：

β是中心损失的平衡权重。

# 数据集

###### Market-1501

是一个公开的专门用于 ReID 的数据集。它的采集通过 6 个摄像头，其中5 个是高清摄像头，拍摄地点在清华大学校园内。里面共包括行人 1501 个，其中检测到的行人框有 32668 个，每个行人至少由 2 个摄像头捕获到，并且在一个摄像头中可能具有多张图像。其中训练集包含图像 12936 张，涵盖人数为 751 人，也就是平均每人有 17.2张图像；测试集包含 19732 张图像，涵盖人数为 750 人，也就是说平均每人对应 26.3 张。在多行人跟踪问题上，跟踪算法与已有的旧轨迹匹配的方式都需要依赖在各帧图片中对行人的定位结果。其中，对于第一次出现的行人，需要生成一个新的跟踪序列号；而对于那些即将离开场景覆盖范围的行人，要对跟踪的轨迹做终止和删除操作。此时，行人对象与检测结果进行匹配的过程可以看成是行人重识别（Person Re-identification，ReID）的过程。在多目标跟踪中，将那些已经产生轨迹了的行人图像集合作为图像库（gallery），而查询图像（query）作为新一轮的行人检测结果，此时，查询图像到检索图像库的过程等价于行人检测与行人轨迹之间关联匹配的过程。

###### WILDTRACK。

在这个数据集中，12米乘36米区域的行人被7台摄像机捕捉到。图像大小为1080×1920，每幅图像标注2帧第二该场景有400个图像，由于有7个摄像头，因此图像总数为2800。平均而言，每帧捕获23.8人，每个人在30.41帧中被看到。WILDTRACK是一个监控录像数据集，记录了苏黎世联邦理工大学主楼外的学生。这些视频是在“无剧本”、“非演员但真实的环境中”拍摄的。总共有7个包含数千名学生的35分钟视频被秘密录制下来，并公开供任何类型的研究使用。在这些视频中，超过1000名学生、教师和旁观者通过土耳其机器人标注他们在每一帧中的位置。由于视频中有大量的学生，“一个受过训练的人平均花10分钟来注释一帧。”WILDTRACK:一个多摄像机高清数据集，用于密集的无脚本行人检测，用于多摄像机行人检测，应用于安全、监视、远程人员识别、机器人、自动驾驶和众包。WILDTRACK是一个监控录像数据集，记录了苏黎世联邦理工大学主楼外的学生。这些视频是在“无剧本”、“非演员但真实的环境中”拍摄的。总共有7个包含数千名学生的35分钟视频被秘密录制下来，并公开供任何类型的研究使用。在这些视频中，超过1000名学生、教师和旁观者通过土耳其机器人标注他们在每一帧中的位置。由于视频中有大量的学生，“一个受过训练的人平均花10分钟来注释一帧。”这些注释随后被用于研究论文WILDTRACK Multi-Camera Person Dataset和WILDTRACK:一个多摄像机高清数据集，用于密集的无脚本行人检测，用于多摄像机行人检测，应用于安全、监视、远程人员识别、机器人、自动驾驶和众包。

###### MultiviewX

该数据集是由Unity engine和PersonX的人体模型生成的合成数据集[22]。与WILDTRACK相同，400帧的大小为1080×1920，注释为每秒2帧。地平面尺寸为16×25m2，略小于WILDTRACK。与WILDTRACK不同，该数据集中使用了6台摄像机，每帧有40人。

###### Duke MTMC (Multi-Target, Multi-Camera)

是杜克大学2014年校园监控录像的数据集，用于视频跟踪系统的研发、人员再识别和低分辨率面部识别。数据集包含超过14小时的同步监控视频，来自8个摄像头，1080p和60fps，超过200万帧，2000名学生步行上下学。校园里安装了8个监控摄像头，专门用来捕捉学生“在课间，行人拥堵的时候”的行为。

PETS数据集包含S0，S1，S2，S3四个子集，S0为训练数据，S1为行人计数和密度估计，S2为行人跟踪，S3为流分析和事件识别。PETS-2009包含7个摄像头采集的有重叠区域的视频，采集对象为英国雷丁大学主校区的学生，摄像头视角较高且均为静态。

###### INRIA Person Dataset

         Inria数据集是最常使用的行人检测数据集。其中正样本（行人）为png格式，负样本为jpg格式。里面的图片分为只有车，只有人，有车有人，无车无人四个类别。图片像素为70\*134，96\*160，64\*128等。具体图像如下图所示，下载链接为<http://pascal.inrialpes.fr/data/human/>

###### CaltechPedestrian Detection Benchmark

加州理工学院的步行数据集包含大约包含10个小时640x480 30Hz的视频。其主要是在一个在行驶在乡村街道的小车上拍摄。视频大约250000帧（在137个约分钟的长段），共有350000个边界框和2300个独特的行人进行了注释。注释包括包围盒和详细的闭塞标签之间的时间对应关系。更多信息可在其PAMI 2012 CVPR 2009标杆的论文获得。具体图像如下图所示，下载链接为<http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/>

###### MIT cbcl (center for biological and computational learning)Pedestrian Data

该数据集主要包含2个部分，一部分为128\*64的包含924个图片的ppm格式的图片，另一部分为从打图中分别切割而出的小图，主要包含胳膊，脑袋，脚，腿，头肩，身体等。具体图像如下图所示，下载链接为<http://cbcl.mit.edu/software-datasets/PedestrianData.html>，需要翻墙才可以。

其他数据集：

