遮挡行人重识别

摘要

当应用于拥挤的公共场所时，人的重新识别 (re-id)会遇到严重的遮挡问题。在本文中， 我们提议利用具有遮挡的人的图像来恢复人全身的图像。这与传统的人重识别问题有很大的不同，在传统问题上，假设人的图像是在没有任何遮挡的情况下被检测到的。因此，我们称这个新问题为遮挡行人重识别。 为了解决这一新问题，我们提出了一种基于深度学习的人身体注意力框架 (AFPB)， 包括 1)遮挡模拟器 (OS)，它自动为全身人图像生成人工遮挡，2)多任务损失，迫使神经网络不仅区分一个人的身份， 而且还 确定一个样本是来自遮挡数据分布还是全身数据分布。 对一个新的遮挡行人重识别数据集和三个现有的基准修改为包括全身人图像和遮挡人图像进行了实验， 表明了该方法的优越性。

关键字：遮挡行人重识别，人体注意力框架，遮挡模拟器，多任务损失

1. 引言

人的重新识别(re-id)旨在通过多个非重叠相机重新识别目标人，这已被应用于加强许多重要公共空间的安全，特别是拥挤的公共空间，例如机场、火车站、购物中心和医院 。 然而，当引导人在这些拥挤的地方重新识别时，遮挡是一个不可避免的问题。 例如，一个人/罪犯可能被现场的其他人遮挡，或诸如汽车、柱子、墙壁等静态障碍物。 考虑到遮挡问题的重要性，寻找一种有效的方法去搜索人全身图像， 通过给出一个人图像与遮挡作为探针，是至关重要的。我们称之为被遮挡的人重新识别问题。

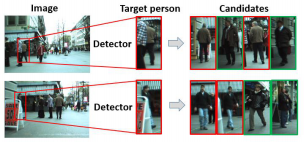


图 1： 被遮挡者重新标识的插图。 左： 视频监控中的原始 图像； 中： 被现有行人探测器检测到的遮挡目标人； 右： 我们的目标是给一个有遮挡的人找回一个全身的人。 红色 包围框表示同一个人， 而绿色表示不同的人。

有三个现实的挑战成为解决遮挡行人重识别的瓶颈。 首先 ， 遮挡不仅会导致目标信息的丢失，还会导致遮挡信息的干扰。具有不同特征的遮挡，如颜色、大小、形状和位置，会使人的重新标识的全局表示恶化。因此，对于被遮挡的人来说， 很难学习到一种鲁棒的特征表示。 其次，人们可以对被遮挡的人的图像使用基于局部/部分的表示。 一种直观的方法是使用身体部位检测器检测非遮挡的身体部位， 然后在画廊中匹配相应的身体部位。 然而，身体检测 器学习需要额外的注释。更糟糕的是， 有时遮挡的身体部位是关键的鉴别部位， 而非遮挡的身体部位则有相似的外观。第三， 由于大多数现有的方法隐含地假设一个人的全身外观是现成的， 而一个有遮挡的人的图像是一个无效的 样本， 很少有公共数据集供遮挡行人重识别去学习一个合适的模型， 特别是用于深度学习。

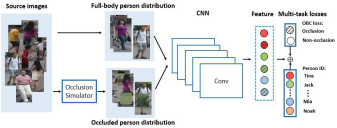


图 2： 概述我们的做法。 给定源图像， 我们首先使用遮挡模拟器生成人工遮挡数据， 然后将源数据和遮挡数据联合编码到具有多任务损失的CNN网络。

为了解决这一具有挑战性的问题， 我们提出了一种新的深度学习框架， 称为人体注意力框架 (AFPB)(fig 2)。 特别是， AFPB由两个组成部分组成。首先，使用遮挡模拟器(OS)通过随机添加背景补丁到全身人图像中自动生成大量人工遮挡的人图像。因此，人工遮挡的人的图像是作为一个人工遮挡集形成的。然后，人工遮挡集和源（全身）集被联合使用， 以学习被遮挡者重新识别的鲁棒特征表示。 第二 ， 多任务损失， 即识别损失和遮挡 /非遮挡二进制分类(OBC)损失，被整合到 AFPB框架中。 令人惊讶的是，简单的OBC损失带来了令人印象深刻的改进， 通过确定一个样本是来自人工遮挡集还是来自源图像。

通过使用遮挡模拟器模块，AFPB方法可以被描述为一个“ 注意”模型。 与传统的注意方案不同，AFPB框架通过观察OS模块产生的不同类型的遮挡人，逐渐关注人的身体。 还可以隐式地将AFPB框架解释为先前知识的编码器， 允许将额外的专家知识集成到深度框架中。 特别是，AFPB的OS模块通过生成大量人工知识相关样本来模拟现实世界来集成遮挡信息， 而OBC损失的目的是确定一个样本是来自被遮挡的人集还是全身人集， 从而将先前的信息编码到框架中。此外，识别损失可以区分给定的人的身份，无论该人是否被遮挡。 这种限制还迫使框架关注人的身体部分。

该方法可以通过具有反向传播的SGD进行端到端的训练， 并且可以很容易地使用公共库(例如 Caffe)来实现。综上，本文主要做出三方面的贡献。

1. 这是第一次尝试定义遮挡行人重识别问题， 这是常 见的现实场景和应用。
2. 对于被遮挡的人重新识别任务，我们提出了一个人体注意框架 (AFPB)，它通过观察不同类型的被遮挡者来逐渐关注人的身体。 遮挡模拟器(OS)用于生成人工样本来编码先验知识，而多任务损失用于学习针对遮挡问题的鲁棒特征表示。
3. 我们创建了一个名Occluded-REID的新数据集和两个被修改的数据集 P-DukeMTMC-reID and P-ETHZ， 它们分别从 DukeMTMC-ReID[1]和 ETHZ[2][中为遮挡行人重识别。](#_bookmark1)
4. 相关工作

典型人物再识别工作主要包括两个步骤：特征提取和度量学习。 第一步旨在提取一种鲁棒和独特的特征表示，它不受光照、视点和遮挡等挑战的影响。 第二步学习度量或子 空间，以便更好地匹配，从而使同一类的距离比不同类的距离更近。

最近，随着深度学习的发展 [3]，有三种网络框架应用于人物重识别，即分类网络、孪生网络和三重态网络。分类网络将人重新 识别问题视为分类问题，由于卷积神经网络 (CNN)在大规模数据集上的优越性能，直接提取判别特征。例如，肖等人。[4]采用域导退 (DGD)方法，共同训练多个领域的分类模型，以提高性能。 孪生网络以图像对作为输入，并使用对比损失计算相似性。 例如，Ahmed等人。 [5]计算图像对的邻域差，以学习指示两个输入图像是否描绘同一个人的度量。 三重网络，由于三个输入图像之间的相似性 ，是一个扩展版本的孪生网络。 基于[6] ， Wang等 .. [7]开发了一个点对点三重态，用于图像对视频人物的重新识别。

虽然人物重识别方法已经得到了很好的研究 [4-11]，但很少有工作初步尝试解决遮挡行人重识别问题。 在与我们有相似想法的作品中，部分人物重识别[12]的目的是将探针部分图像与画廊全身图像相匹配，这是为了给遮挡问题提供一张图片。 然而，他们只关注非封闭的身体部分和全身部分之间的匹配。一些关键问题仍然需要解决：1)现有的行人探测器对被遮挡的人的图像的输出往往包括人的身体的一部分和其他遮挡，而不仅仅是部分身体。也就是说，在 [12]中，必须通过手工裁剪操作来消除遮挡，这是不现实的在实践中。

：2)[12]中提出的基于补丁的匹配方法需要大量的计算，而不考虑注意方案。 我们基于CNN网络的工作不同于 [12]，因为我们直接计算被遮挡的人图像和全身人图像之间的匹配，并提出了一个AFPB框架，通过观看由遮挡模拟器生成的各种被遮挡的人数据来自动关注人的身体。

1. 方法

我们提出了一个深度学习框架，人的身体注意力框架 (AFPB)，如图2所示。AFPB包括两个主要组成部分：遮挡模拟器(OS)和多任务损失。 该遮挡模拟器旨在生成人工遮挡人数据，用于使用源（全身人）数据模拟各种遮挡情况。 接下来，在CNN网络上对全身人数据和遮挡人数据进行多任务损失联合训练，即识别损失和遮挡/非遮挡二值分类(OBC) 损失。 通常，AFPB通过将遮挡的先验信息编码到框架中，迫使特征表示更多地关注人的身体部分。在本节中，我们详细介绍了这些阶段，并可视化了注意结果。

3.1遮挡模拟器 (OS)

正如在第 1节中提到的，随着被遮挡者重识别数据不足的限制，很难为被遮挡者重识别训练一个合适的深度模型。人们会想，是否可以根据现有数据创建大量被遮挡的人重识别数据。 这个问题的一个好主意是从全身人数据中自动生成人工遮挡的行人数据。 这样，不仅可以模拟多种遮挡情况，而且可以给整个系统带来多样化的信息。 基于上述假设，我们设计了一个遮挡模拟器(OS)来生成人工遮挡的人图像。 接下来，我们联合训练一个 CNN网络基于源/全身人数据和被遮挡的人数据。 该网络将更多地关注人的身体部位，通过观看各种遮挡的人的图像在人物重识别中。 具体实 施如下.

假设我们有一个原始的全身数据集X，它由 M个人的 N个图像组成。 设 表示X中的所有样本，其中是第i个人的第j图像，{1，2，···，M}是人。 遮挡模拟器可以表示为图像映射函数 F ： X→Z，其中Z是从真实图像集X生成的人工遮挡人数据集。 映射F是通过一种简单但有效的方法来实现的， 其中来自源图像背景的随机补丁被用作遮挡来覆盖人身体的一部分。 设表示Z中的所有样本，其 中 是第i个人的第j图像。我们最终将X和Z合并成一个组合集。 该过程如算法1所示。

**算法1**：遮挡模拟器

**输入：**全身集合 X(有 N个真实图像)

背景贴片的尺寸s

遮挡面积比范围 [r1 ，r2]

**输出：**人工遮挡集Z，合并集O，

**while N > 0 do**

从X中随机选择一个带有标签y的图像I

选择一个遮挡区域*So ∈* [*SI × r*1*, SI × r*2]

*Sb = s × s ←*从I中选择一个背景贴片

*So ←* resize *Sb*

I\**←* I的一个随机位置被So覆盖

把I\*贴上标签放进Z集

*N* *← N -* 1

**End**

合并集O *←* 合并 X 和 Z

我们的目标是学习一个通用的特征提取器h(·)，它使同一个人的描述符更接近，而不同的描述符更不同。 在我们的框架中，我们训练一个具有识别损失的CNN来识别每个人的身份。当只对全身集合X进行训练时，目标函数是

 (1)

其中f(·) 是识别分类器，(·)是识别损失函数。在从X中生成人工遮挡集Z后，我们将两个集合组合在一起，因此目标函数是



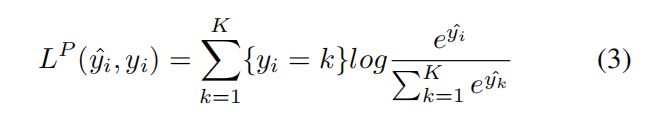
(2)

可以看出，目标函数使 h(xi)和 h(zi)都更接近 yi，因此它会迫使 h(zi)与 h(xi)更相似。直观地解释说，网络已经学会了如何通过观看大量被遮挡的人的数据和源数据来更多地关注关键的人的身体部位，而不是遮挡物或背景。

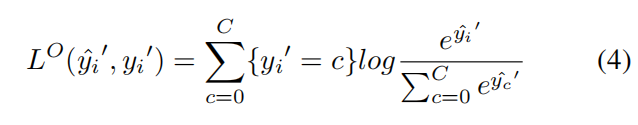
3.2. 多任务损失

随着识别损失，使用遮挡/非遮挡二值分类(OBC)来确定样本是来自遮挡的人分布还是全身人分布，这样我们的框架就可以在区分一个人的身体是否被遮挡的基础上识别出 这个人。 当将两个损失集成到一个统一的框架中时，AFPB方法进一步学习去提取一个遮挡人物重识别的鲁棒的和有识别力的特征表示。这样，OBC损失将遮挡的先验信息编码到框架中。

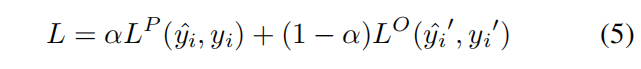
我们将人物重识别视为分类问题，并使用**Softmax**损失作为识别损失。假设原全体集具有K个人，则一个人物重识别分类器的K类Softmax损失如下



其中*y*ˆ*i* 是第k类的第i个训练样本的人物重识别分类器的预测分数。因此，OBC损失如下



其中yˆi’是第i个训练样本的遮挡分类器中的预测分数，c{0，1}其中c=0表示被遮挡者，c=1表示其他的。将识别损失和OBC损失相结合，将多任务损失表述为



其中，是一个超参数，它平衡了两个各自的损失。一般来说，设置α>=0.5是合理的，因为是主角,对的辅助。有了多任务损失，CNN网络具有识别人的判别性，无论一个人是否被遮挡。也就是说，如果一个全身的人图像是可利用的，网络可以利用整个人的结构信息。如果一个人被遮挡，网络可以集中在关键的身体部位。

通过这个过程，我们的框架可以集中在人的身体部分，以学习一个鲁棒的特征表示，在现实的世界里对抗遮挡。Fig.3显示了显着性映射，通过平均池化最后一个卷积层的所有特征映射。这证明了我们的框架可以更多地关注人的身体部位，而不是遮挡物或背景。

1. 实验
   1. 准备

**数据集.**我们在四个数据集上对所提出的方法进行了评估**:Occluded-REID, Partial-REID, P-DukeMTMC-reID and P-ETHZ**每个数据集被组织成两个部分：被遮挡的人图像和全身人图像（见图4）。

**Occluded-REID**是一种由移动摄像机设备捕获的新数据集，由200个被遮挡人的

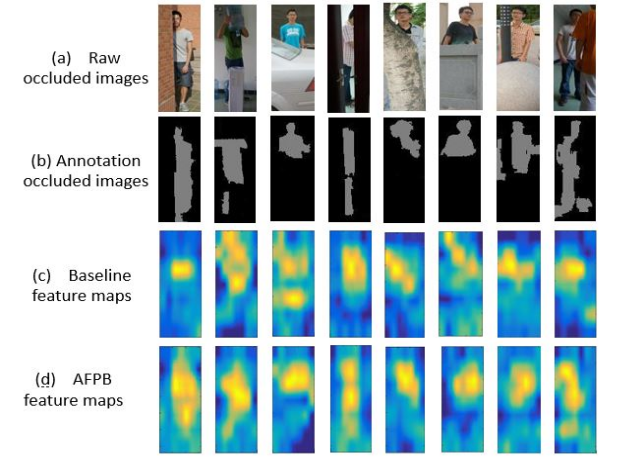


图 3：例子：(a)原始遮挡图像，(b)人工标注人体部位的遮挡图像，(c)基线网络的显着性地映射，(d)AFPB的显着性地映射。突出区域揭示了网络表示关注的部分.

2000张图像组成。每个人有5个全身人像和5个不同类型重度遮挡人像。所有具有不同视点和背景的图像都被调整为128×64。此数据集将在稍后发布。

**P-DukeMTMC-reID** 和 **P-ETHZ** 从

**DukeMTMC-reID** [1]和**ETHZ**[2]中修改而来。

它们[包](#_bookmark1)含在公共场合被不[同](#_bookmark2)类型的遮挡物遮挡的目标人的图像，例如人、行李 、汽车和导板。 我们选择具有全身人图像和遮挡人图像的人 。整理后，在**P-DukeMTMC-reID**中有 24143张1299个人的图像，在**P-ETHZ**中有3897张85个人的图像。这两个数据集也将在稍后发布。

**Partial-REID** 部分人物重识别的第一个数据集[12]，其中包括60个人的900幅 图像，其中5幅全身人图像、5幅部分人图像和5幅被遮挡的人图像。这些图像是在一个大学校园收集的，有不同的视点和遮挡。

**实验环境。**我们以被遮挡的人的图像为探针，以全身人的图像为画廊 ，随机选择一半的身份进行训练，其余的进行测试。 我们报告了在基线网络**ResNet-50**[13]上训练的结果。单次(N=1)和多次(N=2，3，4，5)实验的初始学习率为1e-3，批大小=20，α=0.8，迭代50K。

**评价指标。**在匹配生成中，我们通过L2距离计算每个探针与所有画廊图像之间的相似性。广泛应用的累积匹配特征(CMC)曲线[7]和秩-1率被用来定量评价人的重识别任务。 实验重复10次，以获得平均结果

**数据增强。** 在我们的实验中，我们将所有的图像调整为240×240，并用一个小的随机扰动裁剪224×224的中心区域来增加训练数据。

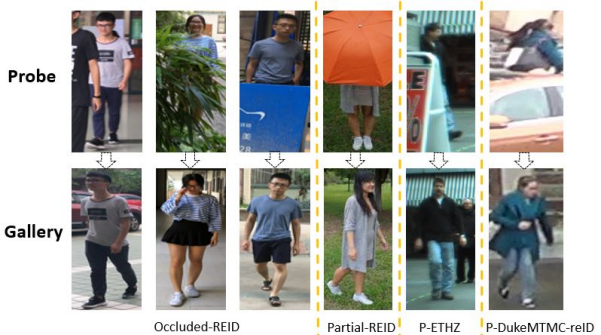


图 4：数据集实例。 上：遮挡人像；下：全身人影像.

4.2. AFPB框架的执行情况

为了评估我们的人身注意力框架 (AFPB)的性能，我们将其与三个网络进行了比较：1)基线网络，在大规模的人重识别数据**MARS**[14]上获得的**ResNet-50**， 2)具有AFPB第一个组成部分的基线网络，遮挡模拟器(OS)，3)具有AFPB第二个组成部分的基线网络，多任务损失，在四个数据集上，**Occluded-REID**，**Partial-REID**，**P-DukeMTMC-reID**和**P-ETHZ**

。如表1和图5所示。我们的AFPB框架显著大幅度优于基线（分别提高了10.9%、8.7%、 7.0%和7.1%，这表明了我们框架的有效性。 此外，具有遮挡模拟器和多任务损失的**ResNet-50**的性能也优于基线，但比我们的 AFPB框架更差。它表明，框架的两个组成部分对提出的的框架作出贡献，它可以通过将它们结合在一起来获得更好的性能，因为 AFPB的两个组成部分形成一种互补关系，以处理各种遮挡的情况。

此外，我们还在两个实验中评估了 AFPB的遮挡意识和注意性能。首先，我们测试了 OBC在训练模型上的准确性。分类准确率分别为88.50%、85.33%、1.75%和73.88%，分别为**Occluded-REID, Partial-REID, P-DukeMTMC-reID and P-ETHZ**，表明OBC损失对AFPB有遮挡意识。 然后，我们计算检测精度[15]，这是显着性映射中的显着区域与我们在**Occluded-REID**和**Partial-REID**上手动注释身体部分的比率（图3）。表2显[示了我们的方法在检测精度上分别超过基线网络](#_bookmark10)6.21%和7.13%，这证实了AFPB对人的身体有优势。

4.3. 与最先进的相比

我们在四个数据集上将我们的方法与最先进的方法进行了比较。表3和图6中。我们收集了七种方法包括 (A)四种手工制作特征

表 1 ： AFPB在单次 /多次射击中的表现

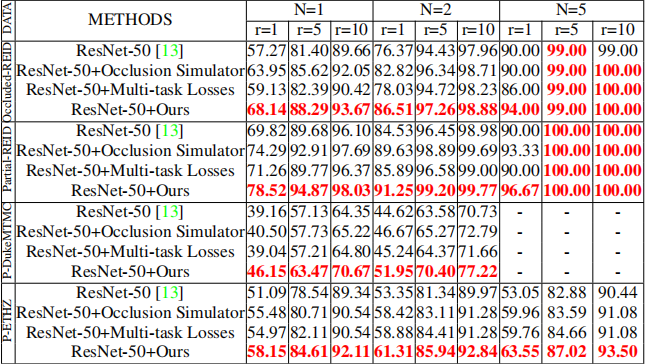
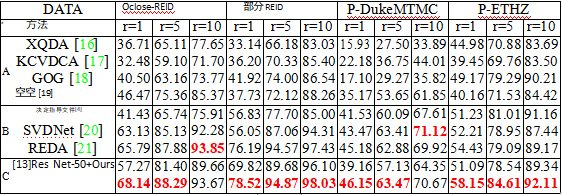


表2 ： **Occluded-REID**，**Partial-REID中**显著性图中显著区域的检测精度



表3 ： 1/5/10等级与最先进水平的比较。

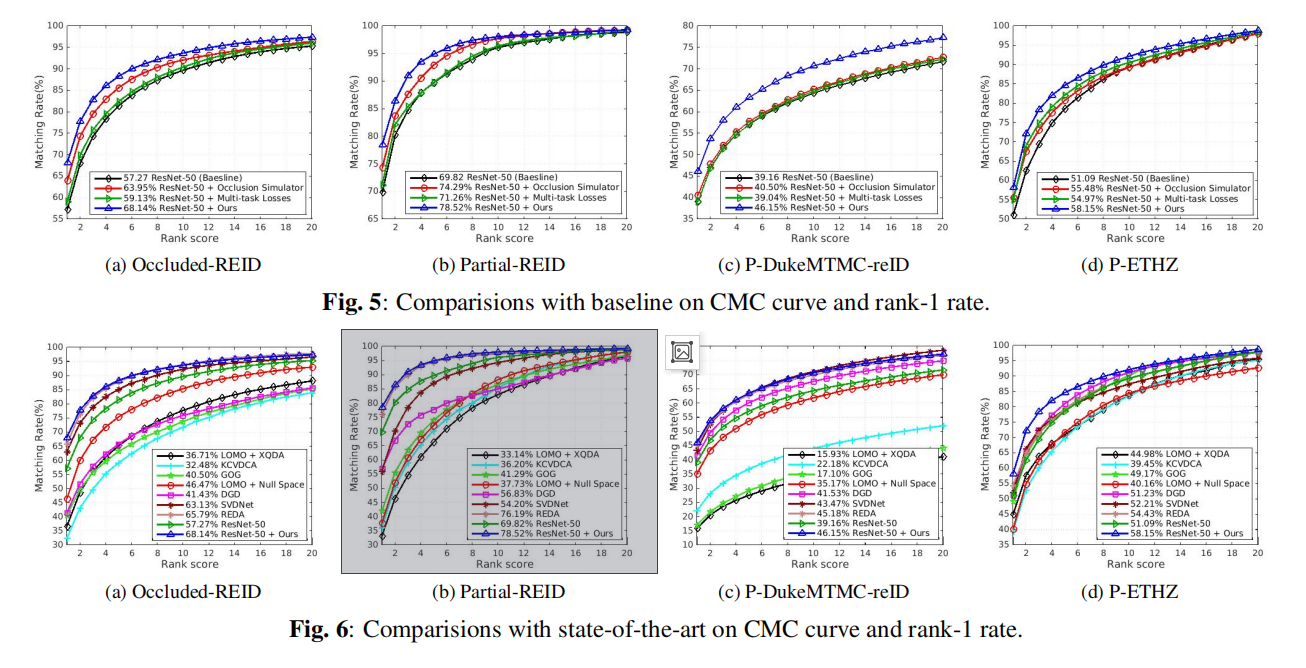


和距离度量的方法和 (B)三种基于深度学习的方法。显然，我们的方法在所有类别中表现最好，并分别在四个数据集上超过排名第二的-1 2.35%, 2.33%, 0.97% and 3.72%。

一般来说，(B)中方法的性能优于(A)中的方 法，因为深度神经网络具有强大的学习能力来自动学习和更新模型。在(B)[4]中的一些方法没有表现出良好的性能， 因为它们的模型关注的是具有全身图像的数据，而不考虑 被遮挡的数据域。不同的是，我们的方法可以学习确定一个人是否被遮挡，并基于遮挡的先验知识提取健壮的特征表示。 通过这种方式，我们的方法显示了在被遮挡的人重识别中的优势。

4.4. 参数分析

我们模型中唯一的一个参数α控制了识别损失和OBC损失之间的权衡。为了探讨两种损失的不同比例的影响，我们进 一步测试了不同α在**Occluded-REID**和**Partial-REID**上的表现。 如图7所示。随着α的增加，表示的性能提高。



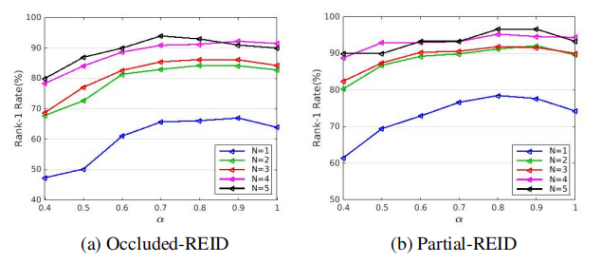


图7：在不同α**Occluded-REID,Partial-REID,** 上的单射和多射实验的秩-1率。

当α在 0.7和 0.9以内时，我们的模型达到了更好的性能，这证实了OBC损失的辅助效果。

1. 结论

在本文中，我们第一次尝试解决被遮挡的人重识别问题。为解决这一问题，**AFPB** 被提出来通过观看各种生成的被遮挡的人图像来学习健壮的特征表示。此外，多任务损失被纳入个人主体注意力框架。 实验结果表明了该方法的有效性和优越性。

1. 参考资料

[1] Z.D. Zheng, L. Zheng, and Y. Yang, “Unlabeled samples generated by gan improve the person re-identifification baseline in vitro,” in *ICCV*, 2017.

[2] A. Ess, B. Leibe, K. Schindler, and L. Van Gool,“A mobile vision system for robust multi-person tracking,”in *CVPR*, 2008.

[3] G.C. Wang, X.H. Xie, J.H. Lai, and J.X. Zhuo, “Deep growing learning,” in ICCV, 2017.

[4] T. Xiao, H.S. Li, W.L. Ouyang, and X.G. Wang, “Learning deep

feature representations with domain guided dropout for person reidentification,” in CVPR, 2016.

[5] E. Ahmed, M. Jones, and T. Marks, “An improved deep learning architecture for person re-identification,” in CVPR, 2015.

[6] S.Y. Ding, L. Lin, G.R. Wang, and H.Y. Chao, “Deep feature learning

with relative distance comparison for person re-identification,” PR,

2015.

[7] G.C. Wang, J.H. Lai, and X.H. Xie, “P2snet: Can an image match

a video for person re-identification in an end-to-end way?,” TCSVT,

2017.

[8] Y.C. Chen, X.T. Zhu, W.S. Zheng, and J.H. Lai, “Person reidentification by camera correlation aware feature augmentation,”

TPAMI, 2018.

[9] S.Z. Chen, C.C. Guo, and J.H. Lai, “Deep ranking for person reidentification via joint representation learning,” TIP, 2016.

[10] S.C. Shi, C.C. Guo, J.H. Lai, S.Z. Chen, and X.J. Hu, “Person reidentification with multi-level adaptive correspondence models,” Neurocomputing, 2015.

[11] C.C. Guo, S.Z. Chen, J.H. Lai, X.J. Hu, and S.C. Shi, “Multi-shot

person re-identification with automatic ambiguity inference and removal,” in ICPR, 2014.

[12] W.S. Zheng, X. Li, T. Xiang, S.C. Liao, J.H. Lai, and S.G. Gong,

“Partial person re-identification,” in ICCV, 2015.

[13] K.M. He, X.Y. Zhang, S.Q. Ren, and J. Sun, “Deep residual learning

for image recognition,” in CVPR, 2016.

[14] L. Zheng, Z. Bie, Y.F. Sun, J.D. Wang, C. Su, S.J. Wang, and Q. Tian,

“Mars: A video benchmark for large-scale person re-identification,”

in ECCV, 2016.

[15] A. Borji, M.M. Cheng, H. Jiang, and J. Li, “Salient object detection:

A benchmark,” TIP, 2015.

[16] S.C. Liao, Y. Hu, X.Y. Zhu, and S.Z. Li, “Person re-identification

by local maximal occurrence representation and metric learning,” in

CVPR, 2015.

[17] Y.C. Chen, W. S. Zheng, and J.H. Lai, “Mirror representation for modeling view-specific transform in person re-identification.,” in IJCAI,

2015.

[18] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato, “Hierarchical gaussian descriptor for person re-identification,” in CVPR, 2016.

[19] L. Zhang, T. Xiang, and S.G. Gong, “Learning a discriminative null

space for person re-identification,” in CVPR, 2016.

[20] Y.F. Sun, L. Zheng, W.J. Deng, and S.J. Wang, “Svdnet for pedestrian

retrieval,” arXiv preprint arXiv:1703.05693, 2017.

[21] Z. Zhong, L. Zheng, G. L. Kang, S.Z. Li, and Y. Yang, “Random

erasing data augmentation,” arXiv preprint arXiv:1708.04896, 2017.