**高阶信息问题：遮挡行人重识别的学习关系和拓扑**

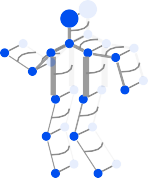
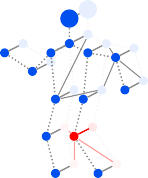
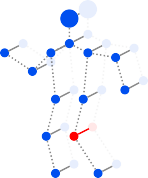
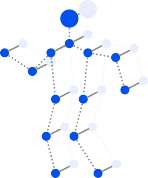
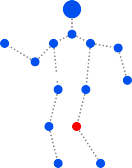
# 摘要

遮挡行人重识别(ReID)旨在将遮挡人图像与不相交摄像机上的整体图像相匹配。在本文中，我们提出了一个新的框架，通过学习高阶关系和拓扑信息的判别特征和鲁棒对齐。首先，我们使用CNN骨干和关键点估计模型来提取语义局部特征。即使如此，被遮挡的图像仍然存在遮挡和离群点。然后，我们将图像的局部特征看作是图的节点，并提出了一种自适应方向图卷积层来传递节点之间的关系信息。所提出的ADGC层可以通过动态学习方向和联动程度自动抑制无意义特征的消息传递。当从两幅图像中对齐两组局部特征时，我们将其视为一个图形匹配问题，并提出了一个交叉图嵌入对齐(CGEA)层，以联合学习和嵌入拓扑信息到局部特征，并直接预测相似性分数。所提出的CGEA层不仅充分利用了图匹配所学的对齐，而且用鲁棒的软匹配代替了敏感的一对一匹配。最后，在遮挡、部分和整体ReID任务上的广泛实验表明了我们提出的方法的有效性。 具体来说，我们的框架在6之前明显优于最先进的框架。在闭塞的杜克数据集上有5%的mAP分数。代码可在https://github.com/wangguanan/HOReID.

# 导言

人的重新识别(ReID)[6,43]旨在匹配一个人的图像跨越不相交的相机，这被广泛用于视频监控、安全和智慧城市。最近，人们提出了[25,39,18,44,16,19,43,11,35]人身份识别的各种方法。然而，它们大多集中在整体图像上，而忽略了被遮挡的图像，这可能更实用和更具挑战性。 如图1(a)所示，人很容易被一些障碍(例如。行李、柜台、拥挤的公共场所、汽车、树木)或走出相机场，导致遮挡图像。因此，有必要将人与遮挡的观察相匹配，这就是所谓的遮挡行人重识别问题[48,26]。

与整体图像匹配相比，由于以下原因，遮挡重识别更具挑战性[45,48]：（1）对于遮挡区域，图像包含的鉴别信息较少，更有可能匹配错误的人。（2）通过部分对部分的匹配，证明了基于部分的特征是有效的[35]。 但它们需要事先严格的人对齐，因此在严重遮挡的情况下不能很好地工作。 最近，提出了许多遮挡/部分人重识别方法[48,49,26,10,8,34,23]，其中大多数只考虑一阶信息进行特征学习和对齐。 例如，预定义的区域[35]、姿态[26]或人类解析[10]用于特征学习和对齐。我们认为，除了一阶信息外，还应该导入高阶信息，并且可能对遮挡行人重识别更有效。



特征学习 一致学习

匹配

软匹配通过合并相关关键点的特征

通用汽车通过边缘到边缘学习拓扑信息

使用图形从边缘学习关系信息

**(c)我们的方法**

来自相同关键点的特征之间的硬匹配

假设关键点准确，特征对齐

提取关键点区域的局部特征

**香草法**

2

2

3

3

**(a)隐藏图像**

2

2

1

1

图 1：说明高阶关系和拓扑信息。（a）在遮挡行人重识别中，关键点受到阻塞(1 2 )和异常值(3 )。(b)Vanilla方法在所有三个阶段都依赖一阶关键点信息，这一点不可靠。(c)我们的方法通过图学习特征以建模关系信息，并通过学习节点到节点和边缘到边缘的对应关系，将对齐视为模型拓扑信息的图形匹配问题。

在图1(a)中，我们可以看到关键点信息受到遮挡(1，2)和异常值(3)。 例如，关键点1 还有2 被遮挡，导致毫无意义的特征。关键点3 是异常值，导致失调。常见的解决方案如图1(b)所示)。它提取关键点区域的局部特征，假设所有关键点都是准确的，并且局部特征很好地对齐。 在这个解决方案中，所有三个阶段都依赖于一阶关键点信息，这并不是很健壮。 在本文中，如图1(c)所示，我们提出了一种新的判别特征和鲁棒对齐框架。 在特征学习阶段，我们将图像的局部特征视作图的节点来学习关系信息。 通过在图中传递消息，遮挡的关键点引起的无意义特征可以通过其邻居有意义的特征来改善。在对齐阶段，我们使用图匹配算法[40]学习鲁棒对齐。 除了与节点对节点的对应关系保持一致外，它还模拟了额外的边对边的对应关系。然后，我们通过构造交叉图像图将对齐信息嵌入到特征中，其中图像的节点消息可以传递给其他图像的节点。 因此，异常点的特征可以通过它在另一幅图像上的相应特征来修复。 最后，我们使用网络来学习由验证损失监督的相似性，而不是用预定义的距离来计算相似性。

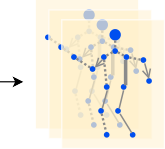
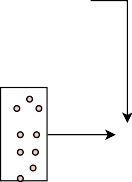
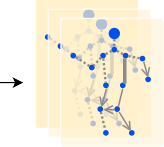
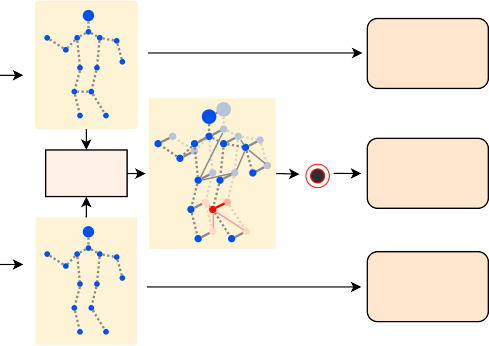
*o*

具体而言，我们提出了一种新的框架，联合建模高阶关系和人-拓扑信息的遮挡行人重新识别。如图2所示，我们的框架包括三个模块，即： 一阶语义模块(S)、高阶关系模块(R)和高阶人体拓扑模块(T)。**(1)**在(S)中，我们利用CNN主干网学习特征映射和人类关键点估计模型学习关键点。然后，我们可以提取相应关键点的语义特征。**(2)**在(R)中，我们将图像的学习语义特征视为图的节点，并提出了一种自适应方向图卷积层(ADGC)来学习和传递边缘特征信息。

*r*

ADGC层可以自动决定每个边缘的方向和程度。 因此，它可以促进语义特征的信息传递，抑制无意义和噪声的信息传递。 最后，学习的节点包含语义和相关信息。**(2)** 在(T)中，我们提出了一种交叉图嵌入对齐(CGEA)层。它以两个图为输入，利用图匹配策略学习两个图之间节点的对应关系，并将学习到的对应关系作为邻接矩阵传递消息。 因此，可以增强相关特征，并将对齐信息嵌入到特征中。 最后，为了避免一对一的硬对齐，我们通过将两个图映射到多元逻辑回归中来预测它们的相似性，并以验证损失进行监督。

本文的主要贡献总结如下：**(1)**提出了一种新的联合建模高阶关系和人体拓扑信息的框架，用于学习遮挡行人重识别的良好和鲁棒对齐特征。据我们所知，这是第一个将这种高阶信息引入到遮挡行人重识别的工作。



局部

局部

局部

局部

**分类损失三重损失**

局部

局部

全局

全局

特征映射

自适应定向图卷积层

**核查损失**

局部

局部

局部

局部

局部

**分类损失三重损失**

局部

全局

**输入**

全局

自适应定向图卷积层

姿势估算器

姿势估算器

CNN

主干

图匹配

相似性分

**S：一阶语义模块 R：高阶关系模块 T：高阶人体拓扑模块**

图2。说明我们提出的框架。它由一阶语义模块S、高阶关系模块R和高阶拓扑模块组成T。模块学习关键点区域的语义局部特征。在此基础上，我们将图像的局部特征视为图的节点，并提出了一种自适应方向图卷积层来传递节点之间的关系信息。在此基础上，我们将对齐问题看作是一个图形匹配问题，并提出了一个跨图嵌入对齐(CGEA)层，将拓扑信息联合学习并嵌入到局部特征中，并直接预测相似性分数。

**(2)**提出了一种自适应有向图卷积层来动态学习图的有向链接，它可以促进语义区域的消息传递，抑制无意义区域的消息传递，如遮挡或异常点。利用该模型，可以更好地模拟遮挡行人重识别的关系信息。 **(3)** 提出了一种具有验证损失的交叉图嵌入对齐(CGEA)层，用于学习特征对齐和预测相似性分数。它们可以避免敏感的硬一对一的人物匹配，并执行一个健壮的软匹配。**(4)**对遮挡、部分和整体重识别数据集的广泛实验结果表明，所提出的模型对最先进的方法具有良好的效果。特别是在occluded-Duke数据集，我们的方法在Rank-1和mAP评分方面明显优于最先进的至少3.7%和6.5。

# 相关工程

**人重新识别。** 人重新识别解决了在不相交的摄像机[6]上匹配行人图像的问题。关键的挑战在于，由于不同的观点、姿势、光照和遮挡而导致的大的类内和小的类间变化。 现有的方法可以分为手工制作的描述符[25,39,18]、度量学习方法[44,16,19]和深度学习算法[43,11,35,36,37,22]。 所有这些ReID方法都集中在匹配整体人图像上，但对遮挡图像的性能不佳，这限制了在实际监控场景中的适用性。

**遮挡行人重识别。** 给定遮挡的探针图像，遮挡行人重识别[48]旨在发现同一人的全身外观在不相交的相机里。

由于信息不完整和空间失调，这项任务更具挑战性。卓等人在[48]使用遮挡/非遮挡二值分类(OBC)损失来区分遮挡图像和整体图像。在他们接下来的工作中，预测了一个显着性地图来突出区分部分，师生学习方案进一步改进了学习特征。 Miao等人在[26]提出了一种基于人的语义关键点的姿态引导特征对齐方法来匹配探针和图库图像的局部补丁。它们使用预先定义的关键点置信度阈值来确定一个身体部分是否被遮挡。Fan等人在[3]使用空间信道并行网络(SCPNet)将部分特征编码到特定的信道，并将整体特征和部分特征融合起来，得到判别特征。罗等人在[23]使用空间变换模块对整体图像进行变换，使其与部分图像对齐，然后计算对齐对的距离。此外，还对部分Re-ID任务的空间对齐进行了一些努力。

**部分人重新识别。**伴随着遮挡的图像，局部图像往往发生由于不完善的检测和离群点的相机视图。与被遮挡的人ReID一样，部分人ReID[45]的目的是将部分探针图像与画廊整体图像相匹配。郑等人在[45]提出了一种全局到局部匹配模型来捕获空间布局信息。何等人在[7]从整体行人中重建部分查询的特征映射，并通过前景-背景掩码进一步改进，以避免背景杂波对[10]的影响。孙等人在[34]中提出了一种可见性感知部分模型(VPM)，它通过自我监督学习感知区域的可见性。

与现有的只使用一阶信息进行特征学习和对齐的遮挡和部分ReID方法不同，我们使用高阶关系和人体拓扑信息用于特征学习和对齐，从而达到更好的性能。

# 提出的方法

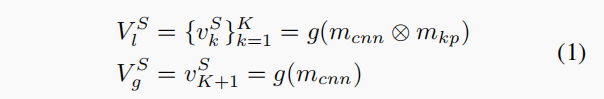
本节介绍了我们提出的框架，包括一阶语义模块(S)提取人类关键点区域的语义特征，高阶关系模块(R)对不同语义局部特征之间的关系信息进行建模，以及高阶人体拓扑模块(T)学习鲁棒对齐和预测两幅图像之间的相似性。这三个模块是

以端到端的方式联合训练的。所提出的方法的概述如图2示。

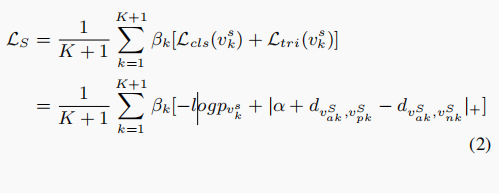
**语义特征提取。** 该模块的目标是提取关键点区域的一阶语义特征，这是由两个线索启发的。首先，基于部分的特性被证明是有效的人ReID[35]。其次，在遮挡/部分ReID[8,34,

10]中，局部特征的精确对齐是必要的。根据上述想法，并在人ReID[43,35,24,4]和人类关键点预测[2,33]的最新发展的启发下，我们利用CNN主干提取不同关键点的局部特征。请注意，虽然人类关键点预测已经达到了很高的精度，但在遮挡/部分图像[17]下，它们仍然具有不满意的性能。这些因素导致关键点位置和置信度不准确。因此，需要以下关系和人体拓扑信息，并将在下一节中讨论。

具体来说，给定一个行人图像x，我们可以得到它的特征映射*mCNN* 和关键点热图*mKP* 通过CNN模型和关键点模型。通过一个叉积()和全局平均池操作()，我们可以得到一组关键点区域的语义局部特征以及一个全局特征。程序即可以用等式(1)来表示，其中K是关键点数，∈，c是通道数。注意*mKP*是通过用Softmax函数对原始关键点热图进行归一化得到的，以防止噪声和异常值。这种简单的操作在实验部分是有效的。



**训练损失。**接下来的[43,11]，我们使用分类和三重损失作为我们的目标在等式(2)中。这里，*βk* =*max*(*mkp*[*k*]) ∈[0*,* 1] 是第k个关键点的置信度。*βK*+1 = 1是全局特征，是特征属于分类器预测的背景真实身份的概率。α 是裕度，，是来自相同身份的正对（，）之间的距离。（，）来自不同的身份。不同局部特征的分类器是不共享的。



**3.1. 高阶关系学习**

虽然我们有不同关键点区域的一阶语义信息，但由于行

人图像不完整，遮挡ReID更具挑战性。因此，有必要利用更多的鉴别特征。我们转向图卷积网络(GCN)方法[1]并尝试对高阶关系信息进行建模。在GCN中，不同关键点区域的语义特征被看作是节点。通过在节点之间传递消息，不仅可以联合考虑一阶语义信息(节点特征)，还可以联合考虑高阶关系信息(边缘特征)。

然而，对于遮挡ReID仍然存在挑战。闭塞区域的特征往往是毫无意义的，甚至是嘈杂的。当在图中传递这些特征时，它会带来更多的噪声，并对遮挡的ReID产生副作用。

因此，我们提出了一种新的自适应方向图卷积层(ADGC)来

动态学习消息传递的方向和程度。利用它，我们可以自动

抑制无意义特征的消息传递，促进语义特征的传递。

自适应定向图卷积层。 一个简单的图卷积层[15]有两

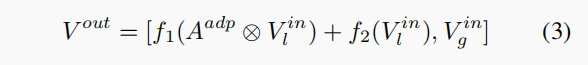
个输入，一个图的邻接矩阵A和所有节点的特征X，输出可以通过以下来计算：

*O* = *XW，*其中是A的归一化版本，W是参数。

我们基于输入特征自适应地学习邻接矩阵（节点的链

接）来改进简单的图卷积层。我们假设给定两个局部特征，

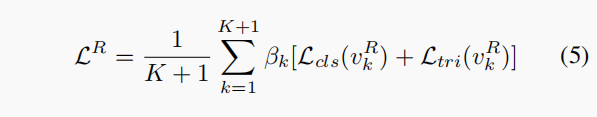
有意义的特征比无意义的特征更类似于全局特征。因此，我们提出了一种自适应有向图卷积(ADGC)层，其输入是全局特征和 K 个局部特征，以及预定义的图(邻接矩阵为A)。我们使用局部特征和全局特征之间的差异动态更新图中所有节点的边权重，得到。然后，通过和之间的乘法可以得到一个简单的图卷积。为了稳定训练，我们将输入的本地特性融合到我们的 ADGC 层的输出中，就像在 ResNet[7]中一样。详情如图 3 所示。我们的自适应有向图卷积(ADGC)层可以用方程（3）表示。其中和是两个未共享的全连接层。



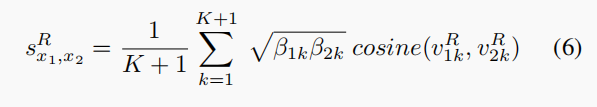
最后，我们实现了我们的高阶关系模块作为 ADGC 层的级联。因此，给定一个图像x，我们可以通过等式(1)得到它的语义特征。则其关系特征可表述如下：



**损失和相似性**。 我们使用分类和三重态损失作为我们的目标，就像在等式(5)中一样。和的定义可在等式(2)中找到。注意是第k个关键点的置信度。



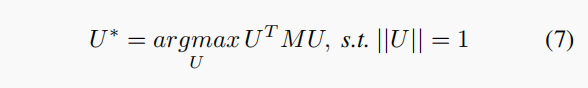
给定两幅图像x1和x2，我们可以通过等式（4）得到它们的关系特征和，然后根据余弦距离计算他们的相似度，如等式（6）。



**3.2. 高阶人体拓扑学习**

基于部分的特性已被证明是非常有效的人ReID[35,34]。 一个简单的对齐策略是直接匹配相同关键点的特征。然而，这种一阶对齐策略不能处理一些坏的情况，如异常值，特别是在严重遮挡的情况下[17]。图形匹配[40,38]自然可以考虑高阶人类拓扑信息。但只能学会一一对应。 这种硬对齐仍然对异常值敏感，并对性能产生不好的影响。在该模块中，我们提出了一种新的跨图嵌入对齐层，它不仅可以充分利用图匹配算法所学的人类拓扑信息，而且可以避免敏感的一对一对齐。

**图形匹配的修订。**从图像x1和x2得到两个图 = (,)和 = (,)。图匹配的目标是学习和之间的匹配矩阵*U ∈ 。*让U ∈[0,1]作一个指示向量然后是和之间的匹配程度。建立一个平方对称正矩阵，使得测量每一对(i，j)∈与(a，b)∈的匹配程度。对于不形成边的对，它们在矩阵中的对应项被设置为 0。对角线条目包含节点-节点分数，而非对角线条目包含边到边的分数。因此，最优匹配可以表述如下：



在[40]之后，我们用一元和成对点特征来参数矩阵 M。 优化过程是通过功率迭代和双随机操作来制定的。 因此，我

们可以在具有随机梯度下降的深度学习框架中优化 U。受页面限制，我们没有显示更多的图形匹配细节，请参考文献[38,

40]。

**具有相似性预测的跨图嵌入对齐层。**我们提出了一种新的交叉图嵌入对齐层(CGEA)，它既考虑了 GM 学习的高阶人体拓扑信息，又避免了敏感的一对一对齐。所提出的CGEA层以两幅图像中的两个子图作为输入，输出嵌入的特征，包括语义特征和人类拓扑引导的对齐特征。

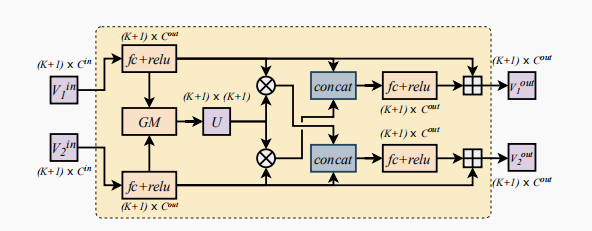
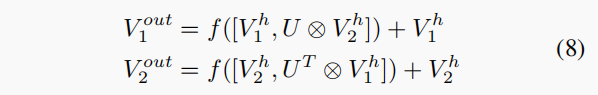


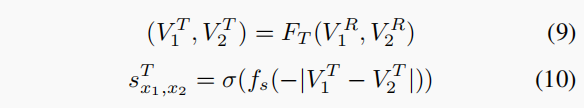
图 4：跨图嵌入对齐层说明。这里，是矩阵乘法， fc+relu 表示完全连接层和恢复线性单元，GM 表示图形匹配操作，U 是学习到的亲和矩阵。详情请参阅正文。

我们提出的 CGEA 层的结构如图 4 所示。它需要两组特征，并输出两组特征。首先，有两组节点和，我们将它们嵌入到具有全连接层和 ReLU 层的隐藏空间中，得到两组隐藏特征和。其次，我们通过公式（7）执行和之间的图匹配，得到和之间的亲和矩阵。在这里，U(i，j)表示和之间的对应关系。最后，输出可以用Eq（8）表示。其中[·，·]表示沿信道维数的级联操作，

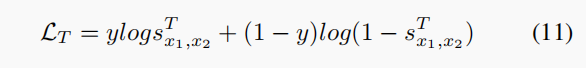
f是一个全连接层。



我们实现了我们具有级联的 CGEA 层和相似性预测层 的高阶拓扑模块(T)。给定一对图像(x1，x2)，我们可以通过Eq（4)得到它们的关系特征，然后通过Eq（9)得到拓扑特征。 在得到拓扑特征对后，我们可以使用 Eq(10)计算它们的相似性，其中|.|是按 元素进行的绝对化操作，是从到1的全连接层，是激活函数。



**验证损失。**我们的高阶人体拓扑模块的损失可以用方程 （11）来表示。，其中 y 是他们的地面真值，y=1，如 果(x1，x2)来自同一个人，否则 y=0。



**4. 训练和推理**

在训练阶段，我们的框架的总体目标函数如 Eq.(12)，其中为相应项的权重。我们通过最小化来端到端训练我们的框架。



对于相似性，给定一对图像(x1，x2)，我们可以从Eq(6)中得到基于相似性的关系信息，从Eq(10)中得到基于相似性的拓扑信息。最后的相似性可以通过联合这两种相似性来计算。

推断时，给定查询图像 xq我们首先计算它与所有画廊图像的相似性 ，并得到它最近的n个近邻。然后用Eq(13)计算最终相似度去提炼top n。

**5. 实验**

**5.1. 实施细节**

**模型架构**。 对于 **CNN** 骨干，与[43]一样，我们通过删

除其全局平均池(GAP)层和全连接层，将**ResNet50**[7]作为 CNN骨干。对于分类器，在[24]之后，我们使用批处理归一化层[13]和全连接层，然后是 **Softmax** 函数。对于人类关键点模型，我们使用预先在COCO数据集[20]训练的 **HR-Net**[33]，一个最先进的关键点模型。该模型预测了 17 个关键点，我们融合了头部区域的所有关键点，得到了最终的 K=14 个关键点，包括头部、肩膀、肘部、手腕、臀部、膝盖和脚踝。

**训练细节。**我们用 Pytorch 实现了我们的框架。图像

被调整到 256×128，并通过随机水平翻转、填充 10 像素，

随机裁剪和随机擦除[47]进行增强。当对遮挡/部分数据集进

行测试时，我们使用额外的颜色抖动增强来避免域方差。批

量大小设置为 64，每人4张图片。在训练阶段，所有三个模块都以端到端的方式联合训练120次，初始化学习率为 3.5e-4，衰减到它的0.1 在30和70个时代。有关实现细节请参考我们的代码。

**评价指标。**我们使用标准指标，如大多数ReID文献，即累积匹配特性(CMC)曲线和平**均精**度(mAP)，以评估不同的人物重新识别模型的质量。所有实验都是在单个查询设置中执行的。

**5.2. 实验结果**

**遮挡数据集的结果。** 我们在两个被遮挡的数据集上评估我们提出的框架，即Occluded-Duke [26] 和 Occluded-ReID [48]。Occluded-Duke是通过从DukeMTMC-reID中留下一些遮挡的图像并过滤掉一些重叠的图像得到的。它包含 15,618 个训练图像、17,661 个图库图像和 2,210 个遮挡查询图像。 Occluded-ReID由移动摄像机捕获，由 200 名被遮挡者的 2000 幅图像组成。 每个身份有五个全身人图像和五个具有不同类型的遮挡的图像。

比较了四种方法，分别是vanilla整体 ReID 方[41,

1. 具有关键点信息[32,5]的整体 ReID 方法、部分 ReID 方法 [45,8,9] 和遮挡 ReID 方法[12,49,10,26]。实验结果见表 2。正如我们所看到的，vanilla整体 ReID 方法和具有关键点 信息的整体方法之间没有显著的差距。 例如，PCB[34]和 FD-GAN[5]都在Occluded-Duke 数据集上获得了大约 40%的 Rank-1 分数，这表明简单地使用关键点信息可能不会显著地有利于 遮挡重识别任务。 对于部分 ReID 和遮挡 ReID 方法，它们都在遮挡数据集上取得了明显的改进。例如，DSR[8] 和 FPR[10]在Occluded-Duke 数据集上分别获得 78.3% 和72.8%的 Rank-1 分数。这表明了遮挡 ReID 和部分 ReID 任务也有类似的困难，即学习鉴别特征和特征对齐。 最后，我们提出的框架在Occluded-Duke 和 Occluded-ReID 数据集上达到了最好的性能，在 Rank-1 评分方面分别为 55.1%和 80.4%，显示了它的有效性。

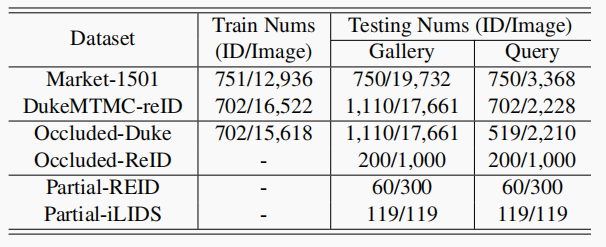


表 1。数据集的详细信息。我们在 6 个公共数据集上对我们提出的方法进行了广泛的评价，包括 2 个整体数据集、2 个遮挡数据集和 2 个部分数据集。

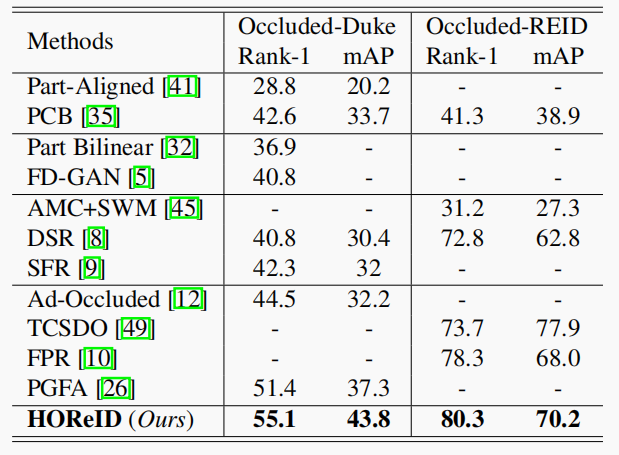


表 2。在两个遮挡数据集上与最先进技术的比较。

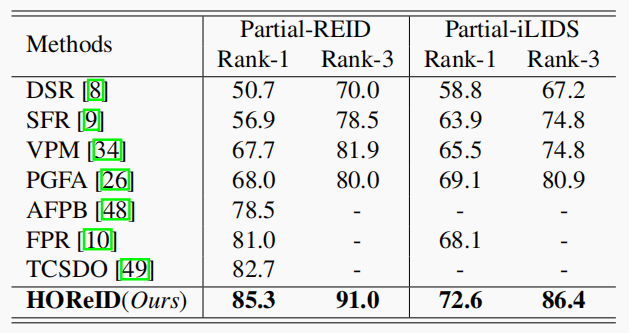


表 3. 在两个部分数据集上与最先进技术的比较。

**部分数据集的结果。**伴随着被遮挡的图像，由于检测 不完善、摄像机视图异常等原因，往往会出现部分图像。为 了进一步评估我们提出的框架，在表 3 中，我**Partial-REID**[45]和**Partial-iLIDS**[8]的结果。**Partial-REID**包括 60 个人的 600 张图像，每人 5 张全身图像和 5 张部分图像，仅用于测试。**Partial-iLIDS**基于**iLIDS[8]**数据集，共包含机场多个不重叠摄像机捕获的 119 人的 238 幅图像，并手动裁剪其遮挡区域。 接下来的[34,10,49]， 由于两个部分数据集太小，我们使用 Market-1501 作为训练 集，两个部分数据集作为测试集。 正如我们所看到的，我 们提出的框架在两个数据集上的Rank-1 评分方面明显优于其他方法至少 2.6%和 4.4。

**整体数据集的结果。** 虽然最近的遮挡/部分 ReID 方法在遮挡/部分数据集上得到了改进，但它们往往无法在整体数据集上获得令人满意的性能。这是由特征学习和对齐过程中的噪声引起的。在这一部分中，我们证明了我们提出的框架也可以在包括 Market-1501 和 DuekMTMTC-reID 在内的整体 ReID 数据集上获得满意的性能。Market-1501[42]包含从 6 个摄像机视点观察到的 1501个身份，19732 个 画廊图像和12936 个训练图像，所有数据集都包含很少被遮挡或部分人图像。 DukeMTMC-reID [28,46]包含 1404 个身份、16522 个训练图像、2228 个查询和 17661 个画廊图像。

具体地 ，我们对两个常见的整体ReID 数据集 Market-1501[42]和 DukeMTMC-reID[28,46]进行了实验，并与3种vanilla ReID方法[35,34,24]、3种人工解析信息的ReID 方法和 4 种关键点信息[31,21,29,26]的整体 ReID 方法进行了比较。实验结果如表 4 所示。正如我们所看到的，3种vanilla 整体 ReID 方法获得了非常有竞争力的性能。 例如， BOT[24]在两个数据集上获得 94.1%和 86.4%的 Rank-1 评分。然而，对于使用外部线索的整体 ReID 方法，例如人为解析和关键点信息的性能更差。 例如，SPREID[14]使用 人工解析信息，在 Market-1501 数据集上只获得 92.5%的 Rannk-1评分。 PFGA[26]使用关键点信息，在 DukeMTMC-re ID 数据集上只获得 82.6%的 Rank-1 分数。这表明，简单地使用外部线索，如人为解析和关键点，可能不会带来在整体 ReID 数据集的改进。 这是由于大多数图像整体 ReID 数据集被很好地检测到，vanilla 整体 ReID 方法足够强大，足以学习鉴别特征。 最后，我们提出了一种自适应方向图卷积 层(ADGC)，它可以抑制噪声特征，并提出了一种避免硬一 对一对齐的跨图嵌入对齐(CGEA)层。 在所提出的 ADGC 和 CGEA 层中，我们的框架在两个整体 ReID 数据集上也取得了类似的性能。 具体来说，我们在 Market-1501 和 DukeMTMC-reID 数据集上获得了大约 94%和 87%的 Rank-1 分数。

**5.3. 模型分析**

**提出的模型的分析。** 在这一部分中，我们分析我们提出的一阶语义模块(S)，高阶关系模块(R)和高阶人体拓扑模块(T)。 实验结果见表 5。首先，在索引1 中，我们删除了所有三个模块，将我们的框架降级为 IDE模型[43]，其中只有一个全局特性Vg可用。 其性能不满足，仅达到 49.9%的Rank-1 分数。 其次，在索引 2 中，当使用一阶语义信息时，性能提高了 2.5%，达到 52.4%的 Rank-1 评分。这表明，关键点的语义信息对 于学习和对齐特征是有用的。 第三，在索引3 中，添加了额外的高阶关系信息，性能进一步提高了1.5%，达到53.9%。 这证明了我们模块 R 的有效性。最后，在索引4 中，我们的完整框架达到了最佳的精度，在 55.1%的 Rank-1 评分下，显示了我们模块 T 的有效性。

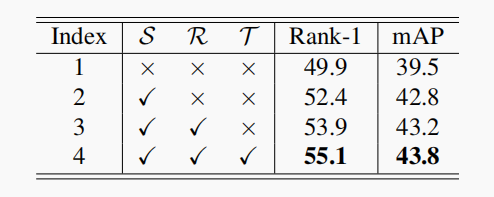


表 5. 分析一阶语义模块(S)，高阶关系模块(R)和高阶人体拓扑模块(T)。实验结果表明了我们提出的三个模块的有效性。

**提出的层的分析。**在这一部分中，我们进一步分析了 语义模块(S)、关系模块(R)和拓扑模块(T)的关键组成部分——关键点置信度(NORM)、自适应方向图卷积(ADGC)层和跨图嵌入对齐(CGEA)层的归一化。具体来说，在删除 NORM 时，直接使用原始的置信度分数。当移除 ADGC 时，在 Eq(3) 中。我们用一个连接起来的固定邻接矩阵替换 ， 就像一个人的拓扑结构一样。 因此，关系模块(S)退化为vanilla

GCN，不能抑制噪声信息。当在 Eq (8)中移除 CGEA 时，我们用一 个全连接的矩阵来代替和 。 即图 1 的每个节点都连接到图 2 的所有节点。 然后，拓扑模块(T)不包含用于特征对齐的高阶人体拓扑信息，并退化为vanilla验证模块。 实验结果见表 6。正如我们所看到的，当删除 NORM、ADGC 或 CGEA 时，性能显著下降 2.6%、1.4%和 0.7%在Rank-1分数上。实验 结果表明，我们提出的 NORM、ADGC 和 CGEA 组分的有效性。

**参数分析。** 我们评估了Eq（13）中参数的影响，即和n。结果如图 5 所示，最优设置为 =0.5，n=8。在分析一个参数时，另一个参数固定在最优值。很明显，当使用不同的  和 n 时，我们的模型稳定地优于基线模型。 实验结果表明，我们提出的框架对不同的权重是鲁棒的。 请注意，这里的表现与表 2 不同，前者达到 57%，后者达到 55%。 这是因为后者是使用平均 10 次计算的，以便进行公平比较。

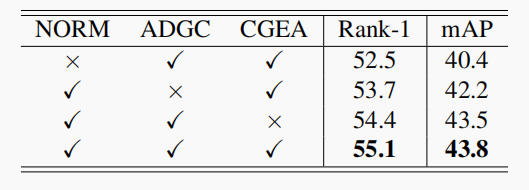


表 6. 关键点置信度(NORM)、自适应方向图卷积层(ADGC)和跨图嵌入对齐(CGEA)层归一化分析。实验结果表明了我们提出的层的有效性。

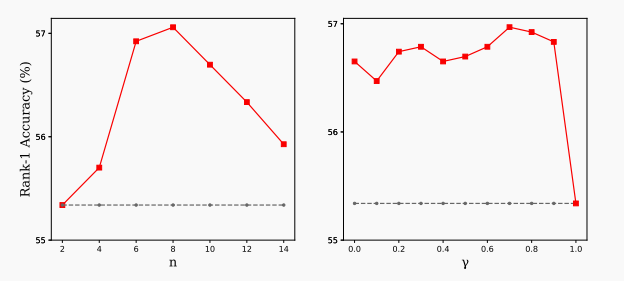


图 5. Eq(13)中参数和 n 的分析。最优值为 =0.5 和 n=8。当分析其中一个时，另一个固定为其最优值。实验结果表明，该模型对不同参数具有较强的鲁棒性。

**6. 结论**

在本文中，我们提出了一个新的框架来学习高阶关系信息的判别特征和拓扑信息的鲁棒对齐。对于学习关系信息，我们将图像的局部特征表示为图的节点，并提出了一种自适应方向图卷积层(ADGC)，以促进语义特征的消息传递，抑制无意义和噪声特征的消息传递。对于学习拓扑信息，我们提出了一种具有验证损耗的跨图嵌入对齐(CGEA)层，它可以避免敏感的硬一对一对齐，并执行健壮的软对齐。最后，在遮挡、部分和整体数据集上的广泛实验证明了我们提出的框架的有效性。

**参考资料**

[1] Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vincius Flores Zambaldi, and Mateusz Malinowski. Relational inductive biases, deep learning, and

graph networks. *arXiv preprint arXiv:1806.01261*, 2018.

[2] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. OpenPose: realtime multi-person 2D pose estimation using Part Affifinity Fields. In *arXiv preprint arXiv:1812.08008*, 2018.

[3] Xing Fan, Hao Luo, Xuan Zhang, Lingxiao He, Chi Zhang, and Wei Jiang. Scpnet: Spatial-channel parallelism network for joint holistic and partial person re-identifification. In *Asian Conference on Computer Vision*, pages 19–34. Springer, 2018.

[4] Yang Fu, Yunchao Wei, Yuqian Zhou, Honghui Shi, Gao Huang, Xinchao Wang, Zhiqiang Yao, and Thomas Huang. Horizontal pyramid matching for person re-identifification. In *Proceedings of the AAAI Conference on Artifificial Intelli*

*gence*, volume 33, pages 8295–8302, 2019.

[5] Yixiao Ge, Zhuowan Li, Haiyu Zhao, Guojun Yin, Shuai Yi, Xiaogang Wang, et al. Fd-gan: Pose-guided feature distilling gan for robust person re-identifification. In *Advances in Neural Information Processing Systems*, pages 1222–1233, 2018.

[6] Shaogang Gong, Marco Cristani, Shuicheng Yan, and

Chen Change Loy. *Person Re-Identifification*. 2014.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[8] Lingxiao He, Jian Liang, Haiqing Li, and Zhenan Sun. Deep spatial feature reconstruction for partial person reidentifification: Alignment-free approach. pages 7073–7082, 2018.

[9] Lingxiao He, Zhenan Sun, Yuhao Zhu, and Yunbo Wang.Recognizing partial biometric patterns. *arXiv preprint arXiv:1810.07399*, 2018.

[10] Lingxiao He, Yinggang Wang, Wu Liu, Xingyu Liao, He Zhao, Zhenan Sun, and Jiashi Feng. Foreground-aware pyramid reconstruction for alignment-free occluded person reidentifification. *arXiv: Computer Vision and Pattern Recognition*, 2019.

[11] Alexander Hermans, Lucas Beyer, and Bastian Leibe. Indefense of the triplet loss for person re-identifification. *arXivpreprint arXiv:1703.07737*, 2017.

[12] Houjing Huang, Dangwei Li, Zhang Zhang, Xiaotang Chen, and Kaiqi Huang. Adversarially occluded samples for person re-identifification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5098–5107, 2018.

[13] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *international conference on machine learning*,pages 448–456, 2015.

[14] Mahdi M Kalayeh, Emrah Basaran, Muhittin Gokmen,

Mustafa E Kamasak, and Mubarak Shah. Human semantic parsing for person re-identifification. In *Proceedings of theIEEE Conference on Computer Vision and Pattern Recognition*, pages 1062–1071, 2018.

[15] Thomas N Kipf and Max Welling. Semi-supervised classi-fification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[16] Martin Koestinger, Martin Hirzer, Paul Wohlhart, Peter M Roth, and Horst Bischof. Large scale metric learning from equivalence constraints. In *2012 IEEE conference on computer vision and pattern recognition*, pages 2288–2295.IEEE, 2012.

[17] Jiefeng Li, Can Wang, Hao Zhu, Yihuan Mao, Hao-ShuFang, and Cewu Lu. Crowdpose: Effificient crowded scenes pose estimation and a new benchmark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10863–10872, 2018.

[18] Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z Li. Person re-identifification by local maximal occurrence representation and metric learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2197–2206, 2015.

[19] Shengcai Liao and Stan Z Li. Effificient psd constrained asymmetric metric learning for person re-identifification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3685–3693, 2015.

[20] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollar, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[21] Jinxian Liu, Bingbing Ni, Yichao Yan, Peng Zhou, Shuo Cheng, and Jianguo Hu. Pose transferrable person reidentifification. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4099–4108, 2018.

[22] Yan Lu, Yue Wu, Bin Liu, Tianzhu Zhang, Baopu Li, Qi Chu, and Nenghai Yu. Cross-modality person re-identifification with shared-specifific feature transfer, 2020.

[23] Hao Luo, Xing Fan, Chi Zhang, and Wei Jiang. Stnreid : Deep convolutional networks with pairwise spatial transformer networks for partial person re-identifification. *arXiv*

*preprint arXiv:1903.07072*, 2019.

[24] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identifification. In *Proceedings of the IEEE Conference on*

*Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[25] Bingpeng Ma, Yu Su, and Frederic Jurie. Covariance descriptor based on bio-inspired features for person reidentifification and face verifification. *Image and Vision Computing*, 32(6-7):379–390, 2014.

[26] Jiaxu Miao, Yu Wu, Ping Liu, Yuhang Ding, and Yi

Yang. Pose-guided feature alignment for occluded person re-identifification. In *ICCV*, 2019.

[27] Lei Qi, Jing Huo, Lei Wang, Yinghuan Shi, and Yang Gao. Maskreid: A mask based deep ranking neural network for person re-identifification. *arXiv preprint arXiv:1804.03864*,

2018.

[28] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set formulti-target, multi-camera tracking. In *European Conference on Computer Vision*, pages 17–35. Springer, 2016.

[29] M Saquib Sarfraz, Arne Schumann, Andreas Eberle, and Rainer Stiefelhagen. A pose-sensitive embedding for person re-identifification with expanded cross neighborhood reranking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 420–429, 2018.

[30] Chunfeng Song, Yan Huang, Wanli Ouyang, and Liang Wang. Mask-guided contrastive attention model for person re-identifification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1179–1188, 2018.

[31] Chi Su, Jianing Li, Shiliang Zhang, Junliang Xing, Wen Gao, and Qi Tian. Pose-driven deep convolutional model for person re-identifification. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 3980–3989, 2017.

[32] Yumin Suh, Jingdong Wang, Siyu Tang, Tao Mei, and Kyoung Mu Lee. Part-aligned bilinear representations for person re-identifification. *european conference on computer vision*, pages 418–437, 2018.

[33] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *CVPR*, 2019.

[34] Yifan Sun, Qin Xu, Yali Li, Chi Zhang, Yikang Li, Shengjin Wang, and Jian Sun. Perceive where to focus: Learning visibility-aware part-level features for partial person reidentifification. pages 393–402, 2019.

[35] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refifined part pooling (and a strong convolutional baseline). In *Proceedings of the European Conference on Computer Vision(ECCV)*, pages 480–496, 2018.

[36] Guan’an Wang, Tianzhu Zhang, Jian Cheng, Si Liu, Yang Yang, and Zengguang Hou. Rgb-infrared cross-modality person re-identifification via joint pixel and feature alignment.

In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.

[37] Guan-An Wang, Tianzhu Zhang, Yang Yang, Jian Cheng, Jianlong Chang, Xu Liang, and Zengguang Hou. Crossmodality paired-images generation for rgb-infrared person re-identifification. In *AAAI-20 AAAI Conference on Artifificial*

*Intelligence*, 2020.

[38] Runzhong Wang, Junchi Yan, and Xiaokang Yang. Learning combinatorial embedding networks for deep graph matching. *arXiv preprint arXiv:1904.00597*, 2019.

[39] Yang Yang, Jimei Yang, Junjie Yan, Shengcai Liao, Dong Yi, and Stan Z Li. Salient color names for person reidentifification. In *European conference on computer vision*, pages 536–551. Springer, 2014.

[40] Andrei Zanfifir and Cristian Sminchisescu. Deep learning of graph matching. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2684–2693,

2018.

[41] Liming Zhao, Xi Li, Yueting Zhuang, and Jingdong Wang.Deeply-learned part-aligned representations for person reidentifification. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 3239–3248, 2017.

[42] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identifification:

A benchmark. In *Proceedings of the IEEE international conference on computer vision*, pages 1116–1124, 2015.

[43] Liang Zheng, Yi Yang, and Alexander G Hauptmann. Person re-identifification: Past, present and future. *arXiv preprint arXiv:1610.02984*, 2016.

[44] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Reidentififi-

cation by relative distance comparison. *IEEE transactions on pattern analysis and machine intelligence*, 35(3):653–668, 2013.

[45] Wei-Shi Zheng, Xiang Li, Tao Xiang, Shengcai Liao, Jianhuang Lai, and Shaogang Gong. Partial person reidentifification. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 4678–4686, 2015.

[46] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identifification baseline in vitro. *arXiv preprint arXiv:1701.07717*, 2017.

[47] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. *arXiv preprint arXiv:1708.04896*, 2017.

[48] Jiaxuan Zhuo, Zeyu Chen, Jianhuang Lai, and Guangcong Wang. Occluded person re-identifification. In *2018 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2018.

[49] Jiaxuan Zhuo, Jianhuang Lai, and Peijia Chen. A novel teacher-student learning framework for occluded person reidentifification. *arXiv preprint arXiv:1907.03253*, 2019.