文章编号：2095-2163(2022)03-0028-06 中图分类号：TP391.4 文献标志码：A

# 基于跨纬度交互注意力机制的行人重识别方法

杨世欣[1]，胡晓光[2]，杜卓群[1]，周峻林[1]，谢佳彧[1]
（1 中国人民公安大学 信息与网络安全学院，北京 100038；2 中国人民公安大学 侦查学院，北京 100038）

**摘 要**：为解决由行人姿态、环境等复杂因素导致的行人特征表达能力弱、识别率低等问题，本文通过对 AlignedReID + +模型进行改进，提出了基于跨纬度交互注意力机制的行人重识别方法。首先，在特征提取部分，将跨纬度交互注意力 Triplet Attention 模块嵌入到 ResNet50 网络中，捕获空间维度和通道维度之间跨纬度的交互信息；其次，引入基于空间特性的视觉激活函数 *Funnel ReLU*，解决激活函数的空间不敏感问题，增强网络模型的非线性表达能力；最后，在 3 个主流的行人重识别数据集 Market1501、DukeMTMC-ReID、CUHK03 上对改进模型进行效能评估，首位命中率 *Rank* - 1 分别提高了 1.6%、1.4%、2.8%，平均精度均值 *mAP* 分别提高了 2.1%、2.3%、3.1%。结果表明所提算法具有良好的性能。

**关键词**：行人重识别；注意力机制；特征提取；视觉激活函数

## Research on person re-identification based on triplet attention mechanism

YANG Shixin[1], HU Xiaoguang[2], DU Zhuoqun[1], ZHOU Junlin[1], XIE Jiayu[1]
(1 School of Information Technology and Cyber security, People's Public Security University of China, Beijing 100038, China; 2 School of Investigation, People's Public Security University of China, Beijing 100038, China)

**【Abstract】** In order to solve the problem of weak person feature expression ability and low recognition rate caused by complex factors such as person attitude and environment, this paper proposes a person re-identificaton method based on cross-latitude interactive attention mechanism by improving AlignedReID + + model. Firstly, in feature extraction, the Triplet Attention module is embedded in ResNet50 network to capture the cross-latitude interaction information between spatial dimensions and channel dimensions; Secondly, Funnel ReLU, a visual activation function based on spatial characteristics, was introduced to alleviate the spatial insensitivity of the activation function and enhance the nonlinear expression ability of the network model; Finally, the effectiveness of the improved model was evaluated on the three mainstream person re-identification datasets Market1501, DukeMTMC-ReID and CUHK03. The first shot Rank-1 increased by 1.6%, 1.4% and 2.8%, the mean average precision mAP increased by 2.1%, 2.3% and 3.1%, respectively. The results show that the proposed algorithm has good performance and can achieve higher recognition accuracy.

**【Key words】** person re-identification; attention mechanism; feature extraction; visual activation function

## 0 引 言

行人重识别(Person Re-identification, ReID)是用计算机视觉技术对多个非重叠摄像机下的不同行人进行检索判断，从而对固定行人进行跟踪的一种有效方法，是计算机视觉领域近年来的研究热点之一，在“平安城市”、“智慧城市”等重大项目建设中扮演着十分重要的角色，具有广泛的应用前景。然而摄像机捕获的行人信息受到视角、光照、分辨率、环境等各种复杂因素的影响，使得大量研究工作都是在寻找鲁棒性更强的行人特征。

随着深度学习在计算机视觉领域的兴起，基于深度学习的方法将特征提取和距离度量紧密结合在一起进行行人重识别，极大推动了行人重识别的发展。依据特征表示方式分为了全局特征和局部特征。全局特征表示方式是对行人图像的整体信息进行特征提取。Wu 等人[1]采用小尺寸卷积滤波器来捕捉行人图像全局特征中的细粒度信息，提出了“PersonNet”的网络结构；Zheng 等人[2]提出一种结合分类损失(identification loss)和验证损失(verification loss)的融合模型，来增强行人图像特征的表达。局部特征表示学习是手动或自动地让网络去提取图像的局部特征，最终的特征由多个局部特征融合而成。常用实现方式有图像水平切片、姿态

基金项目：中国人民公安大学 2021 年度拔尖创新人才培养项目(2021yjsky013)；中国人民公安大学新型犯罪研究专项(2021XXFZ010)；中国人民公安大学公共安全行为科学实验室开放课题(2021SYS03)；上海市现场物证重点实验室开放课题基金(2020XCWZK05)。

作者简介：杨世欣(1996-)，男，硕士研究生，主要研究方向：计算机视觉、行人再识别；胡晓光(1980-)，男，博士，讲师，硕士生导师，主要研究方向：人工智能、计算机视觉。

通讯作者：胡晓光 Email：michael.hu.07@ foxmail.com

收稿日期：2022-01-24

点估计、骨架关键点定位和人体图像分割等。但是通常不会单独使用局部特征，将互补的全局特征与局部特征融合是目前提高网络性能的一个重要分支；Su 等人[3]提出一种结合全局和局部特征的解决姿态变化问题的 PDC（Pose - Driven - Deep Convolutional）模型，利用身体区域线索来学习高效的特征表示以及自适应相似度量；Zhao 等人[4]提出的新型卷积神经网络 SpindleNet，未进行行人对齐，但利用 14 个人体关键姿态点得到具有语义信息的区域，最终将不同尺度的局部特征与全局特征相融合，该模型是基于人体区域引导多阶段特征分解和树结构竞争特征融合的新构想；Zhang 等人[5]提出的另一种融合方法 AlignedReID，先分别计算两幅行人图像的全局特征距离和局部特征距离，再加权求和作为最终结果，亮点在于提出基于局部区域之间联系的动态匹配最小路径算法，用最短路径距离来进行低成本的对齐；在此基础上，Luo 等人[6]提出 AlignedReID + +，采用动态匹配局部信息（DMLI）的方法，不引入额外监督即可自动对齐切片，解决行人不对齐问题。

本文在 AlignedReID + +基础上，对特征提取模块进行改进。以 Resnet50 为基础，通过引入跨维交互注意力 Triplet Attention 来捕捉空间维度和通道维度之间的交互作用；引入一个基于空间特性的视觉激活函数 Funnel ReLU，解决激活函数的空间不敏感问题。

## 1　AlignedReID + +算法

AlignedReID + +算法主要由特征提取和相似度度量两部分组成。在提取特征阶段，把原始大小为 256×128 的行人图像通过 ResNet50 网络进行特征提取，将提取到的特征分别输送给全局分支和局部分支；在相似度度量阶段，分别计算提取的全局特征和局部特征之间的距离。全局距离即全局分支提取到的全局特征的 $L2$ 距离，式（1）：

$$d_g(A,B) = \| \boldsymbol{f}_A - \boldsymbol{f}_B \|_2 \tag{1}$$

其中：$\boldsymbol{f}_A$ 和 $\boldsymbol{f}_B$ 分别表示图像 $A$ 与图像 $B$ 的特征向量。通过计算 $A$、$B$ 两幅图像特征向量之间的 $L2$ 距离来得到两幅图像的相似度。

给定两幅图像的局部特征 $I_A = \{I_{A1}, \cdots, I_{AH}\}$ 和 $I_B = \{I_{B1}, \cdots, I_{BH}\}$，首先通过张量操作 element-wise 转化将距离归一化到 (0,1] 之间，式（2）：

$$d_{i,j} = \frac{e^{\| I_A^i - I_B^j \|_2} - 1}{e^{\| I_A^i - I_B^j \|_2} + 1} \quad i,j \in 1,2,3,\cdots,H \tag{2}$$

其中：$d_{i,j}$ 为 $A$ 图像中第 $i$ 个垂直部分和 $B$ 图像中第 $j$ 个垂直部分之间的距离，$\boldsymbol{D}$ 为距离矩阵。

两幅图像间的局部距离则定义为矩阵 $\boldsymbol{D}$ 中最短路径从（1,1）到（$H$,$H$）的总距离。可以通过动态规划计算，式（3）：

$$S_{i,j} = \begin{cases} d_{i,j} & i = 1, j = 1 \\ S_{i-1,j} + d_{i,j} & i \neq 1, j = 1 \\ S_{i,j-1} + d_{i,j} & i = 1, j \neq 1 \\ \min(S_{i-1,j}, S_{i,j-1}) + d_{i,j} & i \neq 1, j \neq 1 \end{cases} \tag{3}$$

其中：$S_{i,j}$ 是距离矩阵 $\boldsymbol{D}$ 从（1,1）到（$i$,$j$）的最短路径的总距离，$S_{H,H}$ 代表两幅图像之间最终最短路径的总距离，即局部距离，式（4）：

$$d_1(A,B) = S_{H,H} \tag{4}$$

可将两幅图像间总距离表示为局部距离与全局距离之和，式（5）：

$$d(A,B) = d_g(A,B) + \lambda d_1(A,B) \tag{5}$$

其中：$d_1(A,B)$ 为局部距离，$\lambda$ 为平衡全局距离与局部距离的权重系数，此处取值为 1。

训练过程中选用 TriHard 损失作为度量学习的损失，同时全局分支中使用 Softmax 损失来进行多分类，则 AlignedReID + +的总体损失函数，式（6）：

$$L = L_{ID} + L_T^g + L_T^1 \tag{6}$$

其中：$L_{ID}$ 和 $L_T^g$ 分别是全局分支的 Softmax 损失和 Triplet 损失，$L_T^1$ 为局部分支的 Triplet 损失，整体损失为 3 个损失之和。

## 2　改进的方法

本文对 AlignedReID + + 模型框架进行改进，如图 1 所示。

（1）将跨维交互注意力（Triplet Attention，TA）模块[7]引入到特征提取网络 ResNet50 中，使模型更加关注行人图像中的关键区域，抑制无关特征。

（2）引入基于空间特性的视觉激活函数 Funnel ReLU[8]，通过增加一个空间条件，缓解激活函数的空间不敏感问题。

### 2.1　Triplet Attention

注意力机制（Attention Mechanism）的目标是从大量的信息中筛选出对当前任务更有效的细节信息。本文通过引入 Triplet Attention 模块对 AlignedReID + +中的特征提取网络 ResNet50 进行改进，使模型更加关注行人图像中的关键区域。Triplet Attention 是一个几乎无参数、且不涉及任何降维的廉价且有效的注意力机制。其原理是一种基

于三分支结构的跨维度交互(cross dimension interaction)计算注意力权重的新方法,即通过 3 个分支分别捕获输入张量的 $(C,H)$、$(C,W)$ 和 $(H,W)$ 之间的依赖关系。网络结构如图 2 所示。

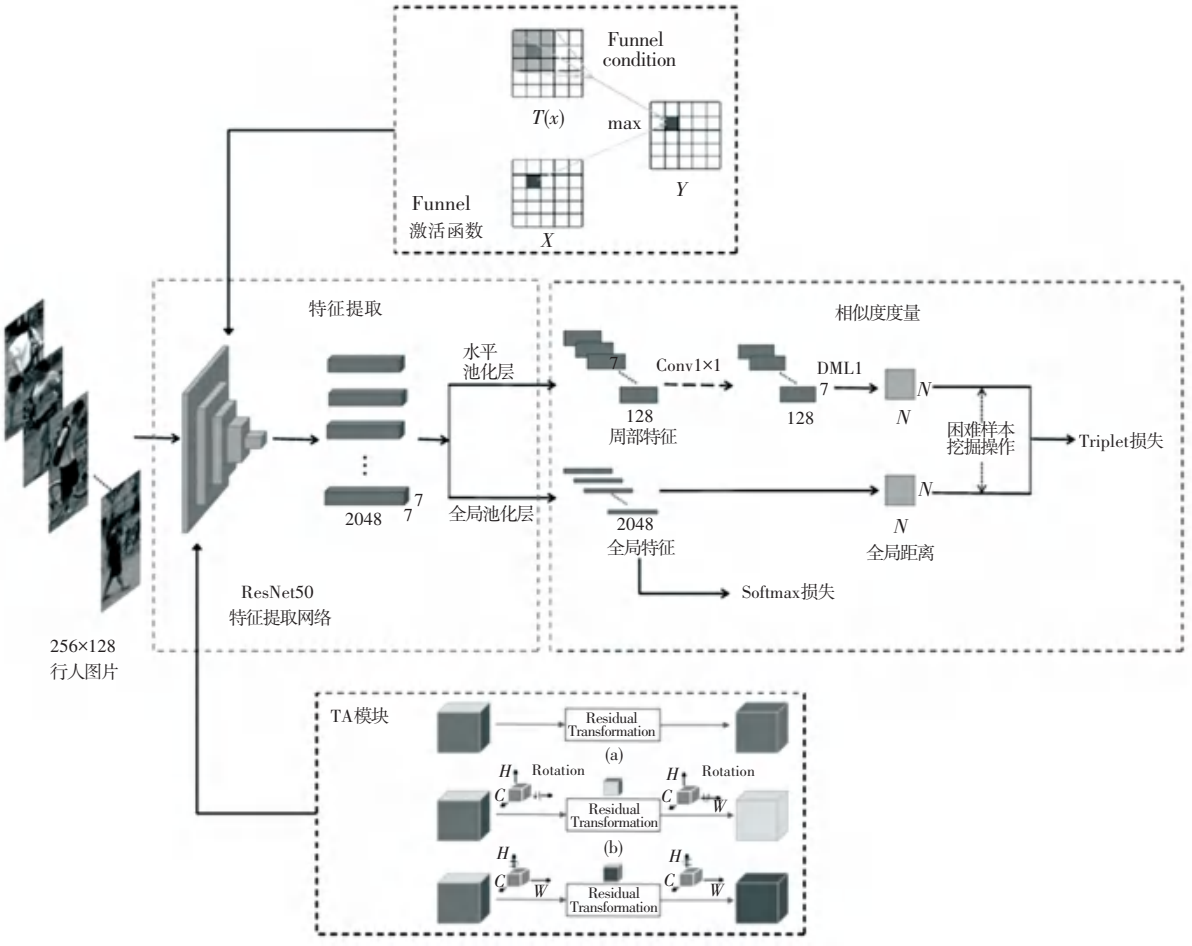



**图 1 基于 AlignedReID + +改进的行人重识别框架图**

**Fig. 1 Improved person re-identification framework based on AlignedReID + +**

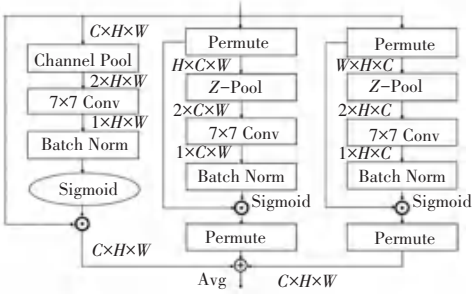



**图 2 Triplet Attention 网络结构图**

**Fig. 2 Network structure diagram of Triplet Attention**

给定一个输入张量 $\boldsymbol{x} \in \mathbb{R}^{C\times H\times W}$,首先,把输入 $\boldsymbol{x}$ 传递给 3 个分支:

(1)第一个分支对通道维度 $C$ 和高度维度 $H$ 之间进行交互,将输入的特征 $\boldsymbol{x}$ 沿 $H$ 轴逆时针旋转 90°得到大小为 $(W\times H\times C)$ 的 $\hat{\boldsymbol{x}}_1$,接着在 $W$ 维度上通过Z-Pool缩减到大小为 $(2\times H\times C)$ 的 $\hat{\boldsymbol{x}}_1^*$,随后通过内核大小为 7×7 的标准卷积层和批量归一化层,从而得到一个中间输出 $(1\times H\times C)$,再通过 Sigmoid 激活层 $(\sigma)$ 生成注意力权重,随后将其应用于 $\hat{\boldsymbol{x}}_1$,最终沿着 $H$ 轴顺时针旋转 90°以保持 $\boldsymbol{x}$ 的原状。

(2)第二个分支对通道维度 $C$ 和宽度维度 $W$ 之间进行交互,将输入特征 $\boldsymbol{x}$ 沿 $W$ 轴逆时针旋转 90°,随后进行与第一分支相同的变换,相应的得到 $\hat{\boldsymbol{x}}_2$、$\hat{\boldsymbol{x}}_2^*$。

(3)第三个分支为空间注意力计算分支,输入张量 $\boldsymbol{x}$ 经过 Channel-Pool 得到大小为(2×H×W)的 $\hat{\boldsymbol{x}}_3$,再接着经过 7×7 的标准卷积和批量归一化层得到 $(1\times H\times W)$,最终通过 Sigmoid 激活层$(\sigma)$ 生成空间注意力权重并应用于输入 $\boldsymbol{x}$。

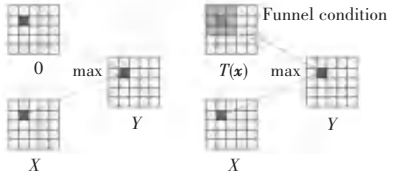
最终对 3 个分支的所有输出特征进行汇总求平均值。

将跨维度交互的 Triplet Attention 模块引入到特

征提取网络 ResNet50 中，使其提取到的行人特征更具有代表性和泛化性。

### 2.2 Funnel 激活函数

激活函数可通过加入非线性因素来解决线性模型表达能力不足的问题。广泛使用的 ReLU、PReLU、Leaky ReLU 等激活函数在语义分割中表现出了对空间信息的不敏感，不能很好的捕捉图片中的空间信息。针对这个问题，本文引入了一种新的基于空间特性的视觉激活函数 Funnel ReLU（FReLU），通过简单的增加一个空间条件，将 ReLU 函数扩展为 2D 激活函数，解决了激活函数的空间不敏感问题，且增加的计算开销不大，如图 3 所示。




（a）ReLU 函数：$\max(\boldsymbol{x},0)$　（b）FReLU 函数：$\max(\boldsymbol{x},T(\boldsymbol{x}))$

**图 3　激活函数示意图**

**Fig. 3　Schematic diagrams of activation function**

FReLU 采用与 ReLU 函数相同的 max(·)，即使用 max(·) 来获得 $\boldsymbol{x}$ 和条件之间的最大值，并通过添加一个视觉漏斗条 $T(\boldsymbol{x})$ 将其扩展到 2D。

FReLU 的表达式(7)~(8)：

$$f(\boldsymbol{x}_c,i,j)=\max(\boldsymbol{x}_c,i,j,T(\boldsymbol{x}_c,i,j)) \tag{7}$$

$$T(\boldsymbol{x}_c,i,j)=\boldsymbol{x}_c^{w},i,j\cdot p_c^{w} \tag{8}$$

其中，以 2D 位置 $(i,j)$ 为像素中心，$f(\cdot)$ 为参数化池化窗口，在第 $c$ 个通道上；$\boldsymbol{x}_c,i,j$ 作参数化池化窗口，在第 $c$ 个通道上；$\boldsymbol{x}_c,i,j$ 作为非线性激活函数 $f(\cdot)$ 的输入像素；$p_c^{w}$ 表示在同一通道中共享的这个参数窗口上的系数；$(\cdot)$ 表示点乘。

FReLU 函数拥有像素级的空间布局能力，通过在激活函数中使用空间条件，将原始 ReLU 更新为一个具有了自适应获取图像局部上下文能力且形式又简单的激活函数，可以轻易的提取图像的空间结构，更加提升了激活函数在行人重识别任务中的精度和鲁棒性。

## 3　实验结果与分析

### 3.1　数据集

为了评估本文所提出的方法，选取行人重识别研究中 3 个主流数据集：Market1501、DukeMTMC-reID、CUHK03。CUHK03 数据集使用 5 对摄像头进行采集，包括 1 467 个不同的行人和 13 164 张图片；Market1501 数据集包括由 6 个摄像头拍摄到的 1 501个行人、32 668 个检测到的行人矩形框。其中训练集有 751 个行人、12 936 张图像，测试集有 750 个行人、19 732 张图像；DukeMTMC-reID 数据集包括 1 404 个行人、36 411 张图像，其中训练集有 702 个行人、16 522 张图像，测试集有 702 个行人、17 661张图像。

### 3.2　实验设置

本实验在 GeForce RTX2080Ti GPU 服务器上搭建了基于 PyTorch 的深度学习框架，选择 ResNet50 作为 Backbone。先将图片分辨率统一为 256×128，然后进行随机擦除等方法及归一化处理，最后将处理过的特征输入到网络中；训练共进行 300 轮，batchsize 设置为 32，初始学习率设置为0.000 2，并且学习率在第 150 个 epoch 时进行衰减，衰减系数为 0.1。Triplet hard loss 中 margin 设置为 0.3。

### 3.3　仿真实验与结果分析

为了验证引入的 TA 模块和 FReLU 激活函数的有效性，将改进后的模型在 CUHK03、Market1501 和 DukeMTMC-reID 3 数据集上进行训练和测试，并遵循通用的评价标准，利用累计匹配特性（Cumulative Match Characteristic Curve，CMC）曲线中的首位命中率 *Rank* - 1 和平均精度均值 *mAP*（mean Average Precision）两个最常用的性能评价指标对网络性能进行评测。全部实验均采用单帧查询模式，采用全局距离加局部距离的结果（Global + DMLI），以及再排序（Re-ranking，RK）后的结果。

将 TA 注意力模块加入到 Baseline 网络中，实验结果见表 1。由表 1 可以看出，模型在 3 个数据集上性能均有所提升。在 Market1501 数据集上 *Rank* - 1 达到了 91.9%，*mAP* 达到了 79.8%，分别提升了 0.9%和 2.2%。在 DukeMTMC-ReID 数据集上性能相差不多，但 *Rank* - 1 也是达到了 81.2%，提升了0.5%。在 CUHK03 数据集上 *Rank* - 1 达到了 62.9%，*mAP* 达到了 60.1%，分别提升了 2.0% 和 0.4%。经过 RK 后，效果尤其明显。在 Market1501 数据集上 *Rank* - 1 和 *mAP* 分别提升了 1.3%和1.9%。在 DukeMTMC-ReID 数据集上 *Rank* - 1 和 *mAP* 分别提升了 0.9%和 1.6%。在 CUHK03 数据集上 *Rank* - 1 和 *mAP* 分别提升了 2.7%和 2.9%。实验证明嵌入 TA 注意力模块可以显著提升模型的效能。

将 FReLU 模块加入到 Baseline 网络中，实验结果见表 2，可以看出对激活函数进行改进之后，模型在 3 个数据集上性能同样得到了显著的提升。在

Market1501 数据集上 *Rank* - 1 达到了 91.5%, *mAP* 达到了 79.6%, 分别提升了 0.5% 和 2.0%。在 DukeMTMC-ReID 数据集上 *Rank* - 1 达到了82.0%, *mAP* 达到了 69.1%,分别提升了 1.3%和 1.1%。在 CUHK03 数据集上性能相差不大,但 *Rank* - 1 也达到了61.1%,提升了0.2%。经过 RK 后,提升效果更为明显。在 Market1501 数据集上 *Rank* - 1 和 *mAP* 分别提升了 1.0%和 1.4%。在 DukeMTMC-ReID 数据集上 *Rank* - 1 和 *mAP* 分别提升了 1.4%和 2.1%。在 CUHK03 数据集上 *Rank* - 1 和 *mAP* 分别提升了 2.2%和 1.9%。实验证明采用视觉激活函数可以显著提升模型的效能。

**表 1 基于 TA 模块改进的实验结果**

**Tab. 1 Improved experimental results based on TA module** %

| Method | Market501 | | DukeMTMC-ReID | | CUHK03 | |
|---|---|---|---|---|---|---|
| | *Rank*1/*Rank*1(RK) | *mAP*/*mAP*(RK) | *Rank*1/*Rank*1(RK) | *mAP*/*mAP*(RK) | *Rank*1/*Rank*1(RK) | *mAP*/*mAP*(RK) |
| Baseline | 91.0/92.0 | 77.6/88.5 | 80.7/85.2 | 68.0/81.2 | 60.9/67.6 | 59.7/70.7 |
| Baseline+TA | 91.9/93.3 | 79.8/90.4 | 81.2/86.1 | 67.3/82.8 | 62.9/70.3 | 60.1/73.6 |

**表 2 基于 FReLU 模块改进的实验结果**

**Tab. 2 Improved experimental results based on FReLU module** %

| Method | Market501 | | DukeMTMC-ReID | | CUHK03 | |
|---|---|---|---|---|---|---|
| | *Rank*1/*Rank*1(*RK*) | *mAP*/*mAP*(*RK*) | *Rank*1/*Rank*1(*RK*) | *mAP*/*mAP*(*RK*) | *Rank*1/*Rank*1(*RK*) | *mAP*/*mAP*(*RK*) |
| Baseline | 91.0/92.0 | 77.6/88.5 | 80.7/85.2 | 68.0/81.2 | 60.9/67.6 | 59.7/70.7 |
| Baseline+FReLU | 91.5/93.0 | 79.6/89.9 | 82.0/86.6 | 69.1/83.3 | 61.1/69.8 | 59.3/72.6 |

将改进后的模型与现有模型进行比较,见表 3。改进后的模型在 Market1501、DukeMTMC - ReID、CUHK03 数据集上的性能均有显著的提升。综上,本文提出的改进方法在行人重识别问题中效果显著。

**表 3 实验结果对比**

**Tab. 3 Comparison of experimental results** %

| Method | Market501 | | DukeMTMC-ReID | | CUHK03 | |
|---|---|---|---|---|---|---|
| | *Rank*1 | *mAP* | *Rank*1 | *mAP* | *Rank*1 | *mAP* |
| SVD[9] | 82.3 | 62.1 | 76.7 | 56.8 | 41.5 | 37.3 |
| PCE&ECN[10] | 87.0 | 69.0 | 79.8 | 62.0 | 30.2 | 27.3 |
| MLFN[11] | 90.0 | 74.3 | 81.0 | 62.8 | 52.8 | 47.8 |
| HA-CNN[12] | 91.2 | 75.7 | 80.5 | 63.8 | 41.7 | 38.3 |
| AlignedReID + + | 91.0 | 77.6 | 80.7 | 68.0 | 60.9 | 59.7 |
| AlignedReID + +(RK) | 92.0 | 88.5 | 85.2 | 81.2 | 67.6 | 70.7 |
| Ours(RK) | 93.6 | 90.6 | 86.6 | 83.5 | 70.4 | 73.8 |

## 4 结束语

本文通过改进 AlignedReID + +网络模型,提出了一种基于跨纬度交互注意力机制的行人重识别方法。在 AlignedReID + +基础上,向特征提取部分嵌入跨纬度交互注意力机制 TA 模块,使网络模型更关注于图像关键特征信息,得到更具鲁棒性的行人特征;同时采用基于空间特性的视觉激活函数 FReLU,通过增添一个空间条件,解决激活函数空间的不敏感问题;最后,与行人重识别最新方法对比,通过在 Market1501、DukeMTMC - ReID、CUHK03 数据集上进行效能评估实验,可以看到改进后的模型鲁棒性更强、精确性更高。

## 参考文献

[1] WU L, SHEN C, HENGEL A V. Person Net: Person re-identification with deep convolutional neural networks[J]. arXiv preprint arXiv: 1601.07255, 2016.

[2] ZHENG Z, ZHENG L, YANG Y. A discriminatively learned cnn embedding for person re-identification[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2017, 14(1): 1-20.

[3] SU C, LI J, ZHANG S, et al. Pose-Driven deep convolutional model for person re-identification[C] //2017 IEEE International Conference on Computer Vision. IEEE, 2017.

[4] ZHAO H, TIAN M, SUN S, et al. Spindle net: Person re-identification with human body region guided feature decomposition and fusion[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1077-1085.

[5] ZHANG X, LUO H, FAN X, et al. AlignedReID: Surpassing human-level performance in person re-identification [EB/OL]. (2018-01-31) [2020-03-02]https://arxiv.org/abs/1711.08184.

[6] LUO H, JIANG W, ZHANG X, et al. AlignedReID + +: Dynamically matching local information for person re-identification [J]. Pattern Recognition, 2019, 94: 53-61.

(下转第 39 页)