第 46 卷 第 5 期 计 算 机 工 程 2020 年 5 月
Vol. 46 No. 5 Computer Engineering May 2020

·图形图像处理· 文章编号：1000-3428(2020)05-0247-07 文献标志码：A 中图分类号：TP311

# 基于姿态引导对齐网络的局部行人再识别

郑 烨，赵杰煜，王 翀，张 毅

（宁波大学 信息科学与工程学院，浙江 宁波 315000）

**摘 要：** 将局部行人再识别中的局部图像与整体图像直接进行比较会产生严重的空间错位，从而导致无法检测到正确目标。针对相同尺寸的行人局部图像与全局图像不匹配问题，提出姿态引导对齐网络（PGAN）模型，将姿态作为辅助信息引入到姿态引导的空间变换模块中，从局部图像与整体图像中提取仿射变换后的行人图像并将其与标准姿态进行对齐，再利用卷积神经网络学习相关特征实现局部行人再识别。实验结果表明，在 Partial-REID 数据集上 PGAN 模型取得 65% 的 Rank-1 准确率，相比直接使用深度卷积神经网络提取全局特征进行匹配的基准模型提高了 3.7%，从而证明其具有良好的局部图像对齐能力及行人再识别效果。

**关键词：** 局部行人再识别；对齐网络；空间变换；姿态；深度卷积神经网络

**开放科学（资源服务）标志码（OSID）：**

**中文引用格式：** 郑烨，赵杰煜，王翀，等. 基于姿态引导对齐网络的局部行人再识别[J]. 计算机工程，2020，46(5)：247-253.

**英文引用格式：** ZHENG Ye，ZHAO Jieyu，WANG Chong，et al. Partial pedestrain re-identification based on pose-guided alignment network[J]. Computer Engineering，2020，46(5)：247-253.

# Partial Pedestrain Re-Identification Based on Pose-Guided Alignment Network

ZHENG Ye，ZHAO Jieyu，WANG Chong，ZHANG Yi

（College of Information Science and Engineering，Ningbo University，Ningbo，Zhejiang 315000，China）

**【Abstract】** In partial pedestrian re-identification，serious spatial misalignment will be caused when the partial image of a pedestrian is directly compared with the holistic image，leading to a failure in target detection. To solve the mismatch of the partial pedestrian image and the holistic image of the same size，this paper proposes a Pose-Guided Alignment Network（PGAN） model. The PGAN firstly introduces the pose into Pose-Guided Spatial Transformation（PST） module as auxiliary information，extracts the pedestrian image after affine transformation from the partial image and holistic image，and compares the pedestrian image with the standard pose. Then the Convolutional Neural Network（CNN） is used to learn the features for partial pedestrian re-identification. Experimental results on the Partial-REID dataset show that the rank-1 accuracy of the PGAN model reaches 65% ，which is 3. 7% higher than that of the baseline model that directly extracts the global features with Deep Convolutional Neural Network（DCNN）. The results demonstrate the proposed model has excellent performance in partial image alignment and pedestrian re-identification.

**【Key words】** partial pedestrian re-identification；alignment network；spatial transformation；pose；Deep Convolutional Neural Network（DCNN）

**DOI：** 10.19678/j.issn.1000-3428.0056642

## 0 概述

目标行人从一个相机视域离开，然后在另一个不重叠的相机视域中再次被识别，这一过程在计算机视觉领域称为行人再识别（Re-ID），其是实现多摄像头跟踪的前提条件，在现实生活中得到广泛应用。目前，对于行人再识别的研究主要集中于整体图像的研究，但在现实场景中，由于遮挡等原因拍摄到的图像不都是完整的图像，也可能存在只有部分身体的局部图像，因此需对局部行人再识别作进一步研究。

深度卷积神经网络（Deep Convolutional Neural Network，DCNN）模型通常会将图像缩放到固定大小作为输入，而尺寸相同的局部图像与整体图像会

**基金项目：** 国家自然科学基金（61603202，61571247）；浙江省自然科学基金重点项目（LZ16F03001，LY17F030002）。

**作者简介：** 郑 烨（1994—），女，硕士研究生，主研方向为计算机视觉；赵杰煜，教授；王 翀，副教授；张 毅，硕士研究生。

**收稿日期：** 2019-11-19 **修回日期：** 2020-01-13 **E-mail：** ye11za@163.com

存在严重的不匹配问题并对特征匹配产生影响。相比而言,预对齐的局部行人图像更适合与整体图像进行匹配。本文提出姿态引导对齐网络(Pose-Guided Alignment Network,PGAN)模型,将人体先验知识引入到对齐网络中,使用空间变换生成与标准姿势对齐的行人图像,并在训练阶段利用姿势信息学习空间变换器的对齐参数。

## 1 相关工作

### 1.1 行人再识别

表征学习被应用于行人识别中以学习人的外貌特征。文献[1-2]使用卷积神经网络(Convolutional Neural Network,CNN)学习全局特征。文献[3-4]将图像分为多个部分提取可区分的局部特征,通过图像水平分割可有效提取变化较少的局部特征。图片切块是一种常见的局部特征提取方式,但其缺点在于对图像对齐的要求较高,如果两幅图像没有上下对齐,那么很可能出现头和上身不对齐的现象,反而使得模型判断错误。行人身体的不匹配问题将严重影响不同图像之间的特征匹配。为应对图像不对齐问题,研究人员将空间变换网络(Spatial Transformation Network,STN)引入到再识别模型中对行人图片进行空间变换对齐行人,还有研究人员将人体解析[5]、姿态估计方法等作为先验知识引入到 Re-ID 模型中对行人图像进行对齐。

#### 1.1.1 基于空间变换的行人再识别

STN[6]是一个空间变换模块,可以引入神经网络以提供空间变换功能,包括平移、缩放、旋转等。STN 是一个小型网络,可以进行标准的反向传播和端到端训练,而不会显着增加训练过程的复杂性。STN 由定位网、网格生成器和采样器组成,定位网获取输入的特征图并输出变换参数,网格生成器计算每个输出像素的原始图像中的位置坐标,采样器生成采样的输出图像。

文献[7]提出一种多尺度上下文感知网络(MSCAN),通过将 STN 与定位损失相结合来提取可变的身体部位,从而减少背景影响并在一定程度上将行人图像对齐,但是定位损失的中心先验约束是基于图像主体完整且图像对齐的前提而提出。文献[8]提出行人对齐网络(PAN),使用 STN 在 Re-ID 深度卷积网络前对齐行人图像,但是 PAN 仅使用 Re-ID 损失对其进行训练,图像对齐效果较差。

#### 1.1.2 基于姿态估计的行人再识别

Spindle Net[9]和 GLAD[10]使用姿势估计算法预测人体关键点,然后学习每个部件的特征并组合部件级的特征以形成最终描述符,以解决姿势变化问题。姿态驱动深度卷积模型(PDC)[11]通过姿态信息裁剪身体区域,然后获得经过旋转和调整大小的身体部位用于姿势变换网络对身体部位进行归一化。文献[12]利用姿态不变特征(PIE)作为行人描述符,利用姿势估计定位关键点,将身体各部件通过仿射变化映射生成 Pose Boex 结构。文献[13]提出一个姿势敏感的行人 Re-ID 模型,将关节信息和粗略方位信息引入到卷积神经网络中学习判别特征,实验结果表明,检测到的关节位置和拍摄视角有助于学习特征。然而这些方法都将姿势估计直接嵌入到模型中,增加了计算成本和模型复杂度。

### 1.2 局部行人再识别

在局部行人再识别中,由于存在只有局部身体可被观测到的局部图像,局部图像与整体图像的匹配是局部行人再识别的一大难题。滑动窗口匹配(Sliding Window Matching,SWM)[14]利用与局部图像大小相同的滑动窗口来搜索每个整体图像上最相似的区域,然而局部匹配的计算代价太大。文献[15]提出一种深度空间特征重构(Deep Spatial Feature Reconstruction,DSR)方案,使用全卷积网络(Full Convolutional Network,FCN)生成具有一定大小的空间特征图,以匹配不同大小的行人图像。与 SWM 方案相比,DSR 方案大幅减少了计算量。文献[16]提出可视性局部模型(Visibility Partial Model,VPM),通过监督学习感知区域的可见性,提取区域级特征并比较两个图像的共享区域。

## 2 基于 PGAN 的局部行人再识别

为解决遮挡和尺度变化问题,本文设计一个姿态引导对齐网络来对齐局部行人,然后学习有效的特征进行行人再识别,整体框架(如图 1 所示)包括以下模块:

1)姿态引导的空间变换(Pose-Guided Spatial Transformation,PST)模块,其进行训练并将整体/部分行人图像转换为对齐的行人图像。该对齐方法是基于完整或部分人体姿态,将人体骨骼作为先验知识,利用姿态估计方法提取每个行人图像的姿态信息。需要注意的是:该过程无需对每个图像进行姿态估计,而是使用一个空间变换器来学习标准姿态和给定标准姿态之间的转换参数。在无需显示姿态信息的情况下,该方法是一种非常有效的推理方法。

2)特征提取模块(ResNet)[17],其作为特征提取器的主干模块,提取行人图像的全局特征。

基于 PGAN 模型,局部图像可以实现与整体图像的匹配。

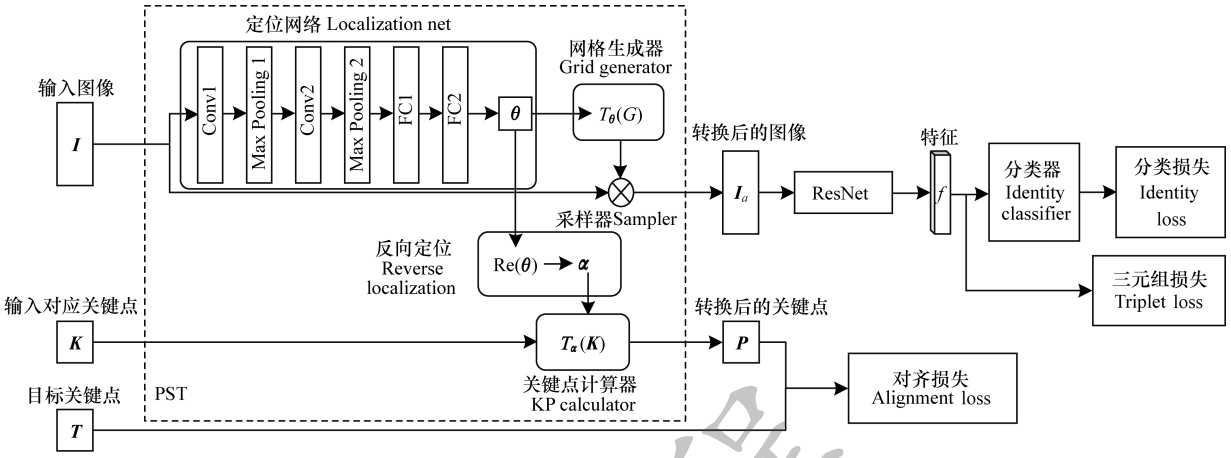



**图 1　姿态引导对齐网络框架**

**Fig. 1　Framework of pose-guided alignment network**

### 2.1　姿态引导的空间变换

PST 模块是 PGAN 中的关键部分,利用姿态信息引导局部行人图像进行空间变换。具体为训练一个空间变换生成一个与目标姿态接近的对齐行人图像,根据人体骨骼关键点确定损失函数。

行人再识别通常使用二维图像作为输入,本文采用仿射变换来变换整体/部分行人图像进行对齐。在 PST 模块中,所有图像被转换成更接近标准姿态的图像。当原始行人图像与对应的转换图像分别表示为 $\boldsymbol{I}$、$\boldsymbol{I}_a$,$\boldsymbol{I}$、$\boldsymbol{I}_a$中的像素可以进一步表示为 $p$、$p_a$,那么 $p$、$p_a$之间的仿射变换为:

$$p_a = \boldsymbol{R}p + \boldsymbol{b} \tag{1}$$

其中,$\boldsymbol{R}$ 是一个与缩放、旋转相关的 2 × 2 参数矩阵,$\boldsymbol{b}$ 是一个与平移相关的 1 × 2 参数向量。

由定位网络、网格生成器和采样器组成的空间转换网络[6]在 PST 模块中被用来进行仿射变换得到仿射变换后的图像$\boldsymbol{I}_a$。输入为原始行人图像 $\boldsymbol{I}$,输出为 $\boldsymbol{\theta}$,包含用于对齐的仿射变换参数。

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} = f_{\text{loc}}(\boldsymbol{I}) \tag{2}$$

其中,$f_{\text{loc}}(\boldsymbol{I})$表示定位网络。定位网络结构如表 1 所示,包括两个卷积层、两个池化层和两个全连接(FC)层,最后一个 FC 层生成仿射转换参数 $\boldsymbol{\theta}$,用于创建网格生成器中的采样网格。

**表 1　定位网络结构**

**Table 1　Structure of localization network**

| 层名称 | 模板数 | 卷积核大小 | 步长 | 填充像素值 | 输出 |
|---|---|---|---|---|---|
| Input | — | — | — | — | 3 × 256 × 128 |
| Conv1 | 32 | 7 × 7 | 2 | 1 | 32 × 128 × 64 |
| Max Pooling 1 | — | 2 × 2 | 2 | — | 32 × 64 × 32 |
| Conv2 | 32 | 3 × 3 | 2 | 1 | 32 × 32 × 16 |
| Max Pooling 2 | — | 2 × 2 | 2 | — | 32 × 16 × 8 |
| Flatten | — | — | — | — | 4 096 |
| FC1 | — | — | — | — | 512 |
| FC2 | — | — | — | — | 6 |

然后采样器从原始行人图像 $\boldsymbol{I}$ 中提取一组采样点,并产生采样输出$\boldsymbol{I}_a$。从仿射图像到原始图像的逐点变换过程如下:

$$\begin{pmatrix} x_i^o \\ y_i^o \end{pmatrix} = \begin{bmatrix} s_x & 0 & t_x \\ 0 & s_y & t_y \end{bmatrix} \begin{pmatrix} x_i^a \\ y_i^a \\ 1 \end{pmatrix} \tag{3}$$

其中,$(x_i^o, y_i^o)^{\mathrm{T}}$ 是原始图像中像素 $p$ 的原坐标,定义为采样点;$(x_i^a, y_i^a)^{\mathrm{T}}$ 是仿射图像上像素$p_a$对应的网格点所在位置坐标。

最终得到仿射变换后的图像$\boldsymbol{I}_a$:

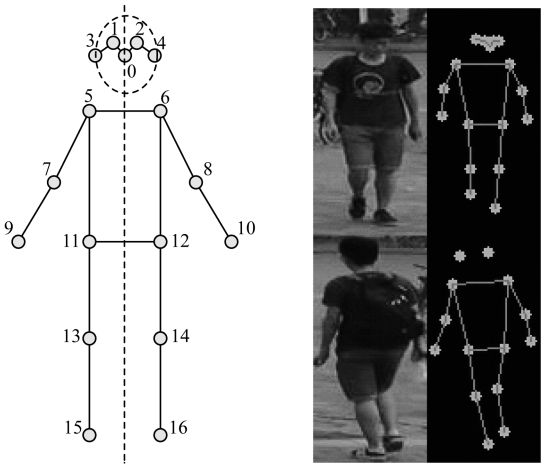
$$\boldsymbol{I}_a = f_{\text{STN}}(\boldsymbol{I}) \tag{4}$$

其中,$f_{\text{STN}}$为空间转换网络。

如果所有行人都有相同的姿势和尺度,则再识别难度会大幅下降。为获得更好的再识别性能,不同 $\boldsymbol{I}$ 的变换图像 $\boldsymbol{I}_a$ 中的行人应具有相似的姿态和尺度,但在原始的 STN 中,变换参数是由网络在没有任何指导的情况下进行学习得到,换言之,其不能确保转换后的图像$I_a$具有所需属性,例如姿态和比例。本文将关键点形式表示的标准姿态作为对齐目标,引导网络学习本文所需的变换,关键点是人体骨骼关节,如图 2 所示,其中包括成对对称的 16 个关节和单个关节,共 17 个关节。关键点的定义如下:

$$\begin{aligned} \boldsymbol{K}_{\text{p}} &= \{\boldsymbol{k}_0, \boldsymbol{k}_1, \cdots, \boldsymbol{k}_i, \cdots, \boldsymbol{k}_N\} = \\ &\quad \{(x_0, y_0), (x_1, y_1), \cdots, (x_i, y_i), \cdots, (x_N, y_N)\} \\ \boldsymbol{K}_{\text{vis}} &= \{v_0, v_1, \cdots, v_i, \cdots, v_N\} \end{aligned} \tag{5}$$

其中,$\boldsymbol{K}_{\text{p}}$、$\boldsymbol{K}_{\text{vis}}$ 分别表示关键点和关键点可见性,$(x_i, y_i)$表示第 $i$ 个关键点的坐标,$v_i$表示其对应的可视分数,$N$ 表示总的关键点个数。




(a)人类骨骼关节 (b)行人图像关键点

**图 2 行人图像骨骼关节关键点**

**Fig. 2 Key points of skeletal joints for pedestrian image**

为评价仿射变换后的图像$\boldsymbol{I}_a$与标准图像的姿态相似性,首先通过 RMPE 姿态估计算法[18]提取$\boldsymbol{I}_a$中人物的关键点 $\boldsymbol{P}$,将这些关键点 $\boldsymbol{P}$ 与标准姿态的目标关键点 $\boldsymbol{T}$ 进行匹配,得到两个姿态之间的相似性。然后将该相似性作为损失项来指导 PGAN 中转换参数的学习过程。需要注意的是:仿射变换在学习阶段可能会极大改变图像中人物的形状。因此,RMPE 姿态估计算法可能无法检测到$\boldsymbol{I}_a$中的关键点。此外,在模型中嵌入姿态估计会大幅增加计算复杂度,因为每个训练元中的每幅图像都需要一个姿态估计,并且该过程不能在 GPU 上并行执行。

由此可知,$\boldsymbol{I}_a$的姿态信息不是直接通过姿态估计得到。实际上,变换后的图像$\boldsymbol{I}_a$无需估计姿态关键点,可以利用输入图像 $\boldsymbol{I}$ 中原始关键点的坐标 $\boldsymbol{K}$ 和 STN 得到的变换参数 $\boldsymbol{\theta}$ 计算得到,并且原始关键点只需在数据准备阶段使用一次 RMPE 姿态估计算法。将原始图像 $\boldsymbol{I}$ 及其姿态信息 $\boldsymbol{K}$ 作为 PST 模块的输入,通过方向定位和关键点计算获得$\boldsymbol{I}_a$中关键点的转换位置 $\boldsymbol{P}$。

$$\boldsymbol{\alpha}=\mathrm{Re}(\boldsymbol{\theta})=\begin{bmatrix}\alpha_{11}&\alpha_{12}&\alpha_{13}\\ \alpha_{21}&\alpha_{22}&\alpha_{23}\end{bmatrix}$$

$$\begin{bmatrix}\alpha_{11}&\alpha_{12}\\ \alpha_{21}&\alpha_{22}\end{bmatrix}=\begin{bmatrix}\theta_{11}&\theta_{12}\\ \theta_{21}&\theta_{22}\end{bmatrix}^{-1},\begin{bmatrix}\alpha_{13}\\ \alpha_{23}\end{bmatrix}=-\begin{bmatrix}\theta_{13}\\ \theta_{23}\end{bmatrix} \tag{6}$$

$$\boldsymbol{P}=\boldsymbol{T}_{\alpha}(\boldsymbol{K}):\begin{pmatrix}x_i^p\\ y_i^p\end{pmatrix}=\begin{bmatrix}\alpha_{11}&\alpha_{12}\\ \alpha_{21}&\alpha_{22}\end{bmatrix}\begin{pmatrix}x_i^k\\ y_i^k\end{pmatrix}+\begin{bmatrix}\alpha_{13}\\ \alpha_{23}\end{bmatrix} \tag{7}$$

其中,$\boldsymbol{\alpha}$是 $\boldsymbol{\theta}$ 对应从 $\boldsymbol{I}$ 到$\boldsymbol{I}_a$反向定位关键点位置的参数,$\boldsymbol{P}$、$\boldsymbol{K}$ 分别是$\boldsymbol{I}_a$和 $\boldsymbol{I}$ 上的关键点,$(x_i^p,y_i^p)^{\mathrm{T}}$、$(x_i^k,y_i^k)^{\mathrm{T}}$ 分别表示 $\boldsymbol{P}$、$\boldsymbol{K}$ 的第 $i$ 个关键点的位置坐标。

为使输入图像与标准姿态对齐,关键点匹配损失定义为两组关键点之间的$L_2$损失之和,使$\boldsymbol{I}_a$ 中每个转换后的关键点与标准姿态的关键点之间尽可能接近。对称的对齐损失函数$L_{\mathrm{Ali}}$为:

$$L_{\mathrm{Ali}}=\mathrm{MSE}(\boldsymbol{P},\boldsymbol{T})=\frac{1}{2N}\sum_{n=0}^{N-1}v_n\times\|\boldsymbol{k}_n-\boldsymbol{t}_n\|_2 \tag{8}$$

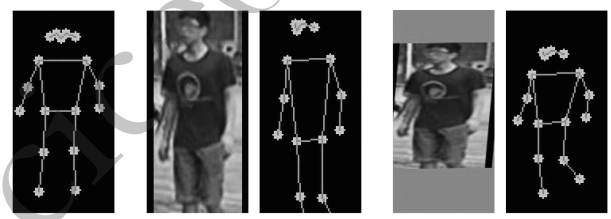
其中,$\boldsymbol{P}$ 和 $\boldsymbol{T}$ 分别表示$\boldsymbol{I}_a$上关键点和目标关键点,$\boldsymbol{k}_n$和$\boldsymbol{t}_n$分别表示 $\boldsymbol{P}$ 和 $\boldsymbol{T}$ 的第 $n$ 个关键点,$v_n$ 表示第$n$ 个目标关键点的可见性分数,$N$ 表示关键点的数量。

如果输入图像是低分辨率或从侧面、背面拍摄的图像,多数姿态估计算法很难区分人体的左右部分。在这种情况下,将所有仿射变换后的关键点与标准姿态匹配可能会造成巨大损失,再加上对 STN 的不当指引,会进一步导致意外的空间变换。为解决上述问题,本文根据人体对称性放宽约束,只选择对称关键点的中心和距离来计算损失,而忽略其他属性。改进的对称对齐损失函数$L_{\mathrm{Ali}}$计算如下:

$$\begin{aligned}&\boldsymbol{P}_m=(\boldsymbol{k}_{p_i}+\boldsymbol{k}_{p_j})/2,P_d=\|\boldsymbol{k}_{p_i}-\boldsymbol{k}_{p_j}\|\\ &\boldsymbol{T}_m=(\boldsymbol{k}_{T_i}+\boldsymbol{k}_{T_j})/2,T_d=\|\boldsymbol{k}_{T_i}-\boldsymbol{k}_{T_j}\|\\ &L_{\mathrm{Ali}}=\mathrm{MSE}(\boldsymbol{P}_m,\boldsymbol{T}_m)+\mathrm{MSE}(P_d,T_d)\end{aligned} \tag{9}$$

其中,$\boldsymbol{P}_m$、$\boldsymbol{T}_m$分别表示对应仿射图像的关键点 $\boldsymbol{P}$ 和标准姿态目标关键点 $\boldsymbol{T}$ 中对称关键点的中心坐标$(x,y)$,$P_d$、$T_d$分别表示 $\boldsymbol{P}$ 和 $\boldsymbol{T}$ 中对称关键点的距离,$\boldsymbol{k}_{p_i}$、$\boldsymbol{k}_{p_j}$和$\boldsymbol{k}_{T_i}$、$\boldsymbol{k}_{T_j}$分别表示 $\boldsymbol{P}$ 和 $\boldsymbol{T}$ 中对称的关键点对。

PST 模块不仅可以单独用于对齐(如图 3、图 4 所示),还可以嵌入到 CNN 模型中进行端到端训练。



(a)标准姿态骨架 (b)训练集上裁剪后的输入图像及其对应关键点 (c)训练过程中变换后的图像及其对应关键点

**图 3 Market-1501 训练集对齐示例**

**Fig. 3 Example of alignment on Market-1501 training set**

(a)输入的测试图像 (b)经过PST对齐后的图像

**图 4 Market-1501 测试集上的对齐结果可视化**

**Fig. 4 Visualization of alignment results on the Market-1501 testing set**

## 2.2 特征提取

ResNet 是目前使用较广泛的 CNN 特征提取网络,本文采用 ResNet-50 作为主干网络,在 PST 模块中提取仿射变换后图像$\boldsymbol{I}_a$的全局特征。

$$\boldsymbol{F}=f_{\mathrm{FE}}(\boldsymbol{I}_a) \tag{10}$$

其中,$f_{FE}(\boldsymbol{I})$是一个特征提取器。

通常用于行人再识别的 softmax 损失$L_{ID}$和三元组损失$L_{Tri}$都被用来训练本文模型,同时利用由全连接层和 softmax 函数组成的分类器来预测输入行人的身份。

$$\boldsymbol{P}_{ID} = \text{softmax}(\boldsymbol{W}^{T}\boldsymbol{F} + \boldsymbol{b})$$

$$L_{ID} = \text{cross-entropy}(\boldsymbol{P}_{ID}, y)$$

$$L_{Tri} = [\| \boldsymbol{F}_i^a - \boldsymbol{F}_i^p \|_2^2 - \| \boldsymbol{F}_i^a - \boldsymbol{F}_i^n \|_2^2 + \beta]_+ \quad (11)$$

其中,$\boldsymbol{P}_{ID}$是 $M$ 个类的预测值分布,$M$ 是身份个数,$y$ 是每个样本的身份信息,$\boldsymbol{F}_i^a$、$\boldsymbol{F}_i^p$、$\boldsymbol{F}_i^n$分别是 anchor 图像、positive 图像和 negative 图像特征,$\beta$ 是三元组损失的边缘,在实验中取值为 0.3。

对于整个网络的训练,本文将结合$L_{Ali}$、$L_{ID}$和$L_{Tri}$作为最终的损失函数,如式(12)所示。通过嵌入 PST 模块使得 PGAN 可以学习对齐特征进行匹配。

$$L = \lambda L_{Ali} + L_{ID} + L_{Tri} \quad (12)$$

其中,超参数 $\lambda$ 在实验中取值为 0.1。

## 3　实验结果与分析

### 3.1　数据集与测试协议

本文模型首先在 Market-1501 数据集上进行训练,然后在 Partial-REID 和 Partial-iLIDS 数据集上进行测试。在数据增强阶段,随机裁剪生成局部图像用于训练。

1)Market-1501[19]包含 6 台摄像机从不同视角拍摄的 1 501 个身份的行人图像 32 368 张。在训练集中,包含 12 936 张 751 个身份的图像。

2)Partial-REID[14]是一个局部行人图像数据集,包含 60 个身份的 600 张行人图像,每个身份有 5 张全身图像和 5 张局部图像。这些图像是在某大学校园从不同视角、背景进行拍摄,并存在不同类型的遮挡情况,每个人的所有局部图像组成 Query 集,而整体行人图像用作 Gallery 集。

3)Partial-iLIDS[20]是一个基于 iLIDS 的局部图像数据集。Partial-iLIDS 共包含 238 张由多个非重叠摄像机捕获的 119 个身份的图像。对于被遮挡的行人,通过剪切每个身份图像的非遮挡区域生成局部图像,构建 Query 集,每个身份的非遮挡图像被选择用来构成 Gallery 集。

本文使用累积匹配曲线(Cumulative Match Characteristic,CMC)的 Rank-1、Rank-3 准确率作为评估指标来衡量模型性能。

### 3.2　PGAN 实现

PGAN 实现过程具体如下:

1)局部图像产生:由于局部行人再识别数据集只提供测试集,因此需要对某些整体数据集进行训练。为学习部分图像的对齐,根据给定范围随机裁剪图像生成整体图像的局部图像,同时对输入关键点做同样处理,使其与输入图像一致。为平衡训练集中整体图像和局部图像的数量,设置整体图像的裁剪概率为 0.5。

2)数据增强:图像大小调整为 256 像素 ×128 像素,并将原始像素值归一化至[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。在训练阶段,在水平方向随机翻转每个图像,填充 10 个零值像素,再将其随机裁剪成一个 256 像素 ×128 像素的图像进行数据增强。

3)网络设定:选择 Re-ID 中常用的 ResNet-50 作为骨干网络。参考文献[21]设置,使用 ImageNet 上预训练的参数初始化 ResNet-50,并将最后一个卷积层的 stride 修改为 1,将全连接层的连接数修改为 $M$,$M$ 表示训练数据集中的身份数。在全连接层前使用 BN bottleneck,PST 中 STN 的参数 $\boldsymbol{\theta}$ 初始化为[1,0,0,0,1,0]。采用 Adam 法对模型进行优化,共有 120 个训练 epoch。初始学习率设定为 $3.5 \times 10^{-4}$,使用 Warmup 方法改变学习速率,在前 10 个 epoch 时将学习速率从 $3.5 \times 10^{-5}$线性增加至 $3.5 \times 10^{-4}$,然后分别在第 40 个 epoch 和第 70 个 epoch 时将学习率除以 10。

4)训练:本文模型分为两个训练阶段。在第一个阶段使用式(9)中的对齐损失$L_{Ali}$在 Market-1501 数据集上预训练 PST 模型;在第二阶段利用预训练的 PST 权值和 ResNet-50 的 ImageNet 上的预训练参数初始化整个 PGAN 模型。

### 3.3　PST 结果可视化

为验证 PST 对局部图像的对齐性,在 Partial-REID 和 Partial-iLIDS 数据集上进行实验。PST 可以学习对齐的空间变换,其在 Partial-REID 数据集上的结果如图 5 所示,结果表明 PST 不仅能对整体图像进行对齐,而且能对局部图像进行准确对齐,验证了 PST 模块的有效性。

(a)整体图像　　(b)局部图像

**图 5　Partial-REID 数据集上的 PST 结果可视化**

**Fig. 5　Visualization of PST results on Partial-REID dataset**

### 3.4 在 Partial-REID 和 Partial-iLIDS 上的测试结果

本文在 Partial-REID 和 Partial-iLIDS 数据集上进行 Rank-1、Rank-3 准确率实验,结果如表 2 所示。

**表 2 Partial-REID、Partial-iLIDS 数据集上的 Rank-1 和 Rank-3 准确率**

**Table 2 Accuracy of Rank-1 and Rank-3 on Partial-REID and Partial-iLIDS datasets** %

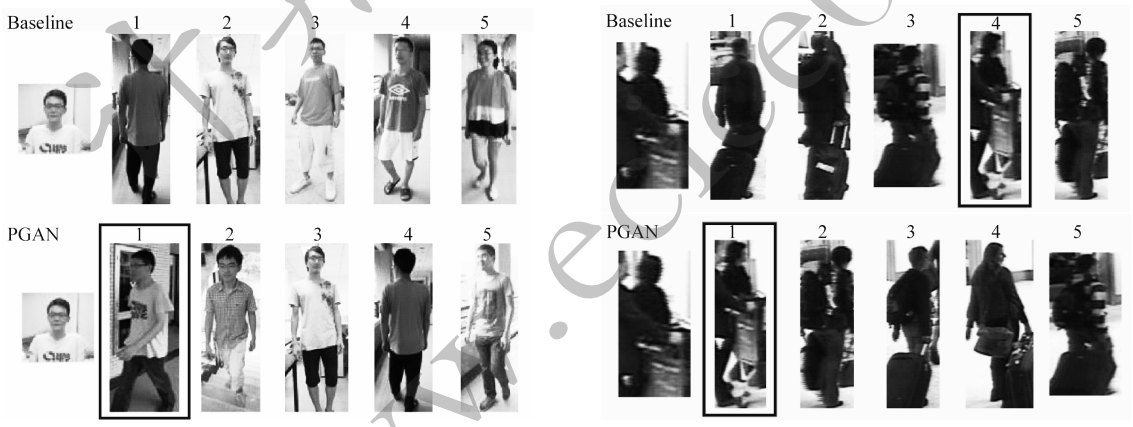
| 模型 | Partial-REID 数据集 | | Partial-iLIDS 数据集 | |
|---|---|---|---|---|
| | Rank-1 | Rank-3 | Rank-1 | Rank-3 |
| AMC + SWM[14] | 37.3 | 46.0 | 21.0 | 32.8 |
| DSR[15] | 50.7 | 70.0 | 58.8 | 67.2 |
| VPM(Bottom)[16] | 53.2 | 73.2 | 53.6 | 62.3 |
| VPM(Top)[16] | 64.3 | 83.6 | 67.2 | 76.5 |
| VPM(Bilateral)[16] | 67.7 | 81.9 | 65.5 | 74.8 |
| Baseline | 61.3 | 76.3 | 56.3 | 74.8 |
| PGAN | 65.0 | 78.0 | 56.3 | 74.1 |

为验证 PST 对局部 Re-ID 的识别效果,本文将 PGAN 与未使用 PST 的 Baseline 模型进行比较。对于局部 Re-ID 数据集,Baseline 在 Partial-REID、Partial-iLIDS 数据集上分别得到 61.3%、56.3% 的 Rank-1 准确率,PGAN 在 Partial-REID 数据集上的 Rank-1 准确率相比 Baseline 提高了 3.7 个百分点。PGAN 相对于 Baseline 的优势在于:PST 进行身体水平的对齐,以处理显著的不对齐问题,但其在 Partial-iLIDS 数据集上没有明显的性能提升,可能的原因为 Partial-iLIDS 数据集中的局部图像保留了大部分身体,其不对齐程度在 CNN 的处理范围内。

PGAN 在两个局部数据集 Partial-REID 和 Partial-iLIDS 上与其他模型进行比较。在 Partial-REID 数据集上,PGAN 的性能相比 AMC + SWM 和 DSR 有较大的优势,但与 VPM 相当。在 Partial-iLIDS 数据集上,PGAN 的性能超越了 AMC + SWM。由于 PGAN 模型嵌入了一个简单的 PST 模块进行局部图像与整体图像的对齐,从而提高局部图像的识别性能。

### 3.5 检索结果可视化

对于卷积神经网络,即使图像中只有行人的局部身体部位但包含一些明显特征的图像,如黄色衣服或者偏移程度较小的图像,其也可以学习高层特征实现图像区分,然而对于没有明显特征、显著不对齐的局部图像很难与整体图像进行匹配。图像行人局部图像的检索结果如图 6 所示。在图 6(a)中,Baseline 无法找到匹配图像,即匹配图像排在第 5 位后,而 PGAN 匹配图像排在第 1 位。在图 6(b)中,对于同一输入图像,Baseline 匹配图像排在第 4 位,而 PGAN 的匹配图像排在第 1 位。实验结果表明,局部图像对齐对局部行人再识别具有较大作用。




(a)Partial-REID数据集 (b)Partial-iLIDS数据集

**图 6 检索结果可视化**

**Fig. 6 Visualization of retrieved results**

## 4 结束语

本文提出一种用于局部行人再识别的姿态引导对齐网络(PGAN)。在 PGAN 中,PST 模块通过姿态信息引导,可对部分行人图像进行有效对齐,只需在数据准备阶段通过姿态估计获取训练数据姿态信息,然后基于 PST 计算得到模型所需姿态信息,使得训练过程更加高效。实验结果表明,PGAN 在局部行人再识别上取得了较好的识别效果,且在训练阶段和推论阶段均未产生额外的计算成本与姿态信息。后续将对结合注意力机制的局部行人再识别模型进行研究,通过抑制背景等干扰信息提高局部行人再识别的准确率。

### 参考文献

[1] LI Wei,ZHAO Rui,XIAO Tong,et al. DeepReID:deep filter pairing neural network for person re-identification[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C.,USA:IEEE Press,2014:152-159.

[2] AHMED E,JONES M,MARKS T K. An improved deep learning architecture for person re-identification[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ,USA:IEEE Press,2015:3908-3916.

[3] CHENG De,GONG Yihong,ZHOU Sanping,et al. Person re-identification by multi-channel parts-based CNN with improved triplet loss function[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ,USA:IEEE Press,2016:1335-1344.

[4] SUN Yifan, ZHENG Liang, YANG Yi, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline)[C]//Proceedings of European Conference on Computer Vision. Washington D. C. ,USA:IEEE Press,2018:480-496.

[5] KALAYEH M M, BASARAN E, GÖKMEN M, et al. Human semantic parsing for person re-identification[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ,USA:IEEE Press, 2018:1062-1071.

[6] JADERBERG M, SIMONYAN K, ZISSERMAN A. Spatial transformer networks[EB/OL]. [2019-10-12]. https://arxiv.org/abs/1506.02025.

[7] LI Dangwei, CHEN Xiaotang, ZHANG Zhang, et al. Learning deep context-aware features over body and latent parts for person re-identification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017:384-393.

[8] ZHENG Zhedong, ZHENG Liang, YANG Yi. Pedestrian alignment network for large-scale person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology,2018,29(10):3037-3045.

[9] ZHAO Haiyu, TIAN Maoqing, SUN Shuyang, et al. Spindle Net: person re-identification with human body region guided feature decomposition and fusion[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ,USA:IEEE Press,2017:1077-1085.

[10] WEI Longhui, ZHANG Shiliang, YAO Hantao, et al. Glad: global-local-alignment descriptor for pedestrian retrieval[C]// Proceedings of the 25th ACM International Conference on Multimedia. New York,USA:ACM Press,2017:420-428.

[11] SU Chi,LI Jianing,ZHANG Shiliang,et al. Pose-driven deep convolutional model for person re-identification[C]// Proceedings of 2017 IEEE International Conference on Computer Vision. Washington D. C. ,USA:IEEE Press,2017: 3980-3989.

[12] ZHENG Liang,HUANG Yujia,LU Huchuan,et al. Pose invariant embedding for deep person re-identification [EB/OL]. [2019-10-12]. https://arxiv.org/abs/1701.07732.

[13] SAQUIB M, SCHUMANN A, EBERLE A, et al. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ,USA:IEEE Press,2018:420-429.

[14] ZHENG Weishi,LI Xiang,XIANG Tao,et al. Partial person re-identification[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. ,USA: IEEE Press,2015:4678-4686.

[15] HE Lingxiao,LIANG Jian,LI Haiqing,et al. Deep spatial feature reconstruction for partial person re-identification: alignment-free approach[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ,USA:IEEE Press,2018:7073-7082.

[16] SUN Yifan,XU Qin,LI Yali,et al. Perceive where to focus: learning visibility-aware part-level features for partial person re-identification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2019:393-402.

[17] HE Kaiming,ZHANG Xiangyu,REN Shaoqing,et al. Deep residual learning for image recognition[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ,USA:IEEE Press,2016:770-778.

[18] FANG Shaoshu,XIE Shuqin,TAI Yuying,et al. RMPE: regional multi-person pose estimation[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. ,USA:IEEE Press,2017:2334-2343.

[19] ZHENG Liang,SHEN Liyue,TIAN Lu,et al. Scalable person re-identification: a benchmark[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. ,USA:IEEE Press,2015:1116-1124.

[20] ZHENG W S,GONG S,XIANG T. Person re-identification by probabilistic relative distance comparison[C]//Proceedings of CVPR'11. Washington D. C., USA: IEEE Press, 2011: 649-656.

[21] LUO Hao, GU Youzhi, LIAO Xingyu, et al. Bag of tricks and a strong baseline for deep person re-identification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops. Washington D. C. ,USA:IEEE Press,2019:1-7.

编辑　陆燕菲