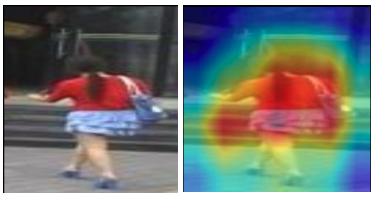
Opto-Electronic Engineering
光 电 工 程

Article

2021 年，第 48 卷，第 5 期

DOI: 10.12086/oee.2021.200388

深度双重注意力的生成与判别联合学习的行人重识别



张晓艳[1]，张宝华[1,3*]，吕晓琪[2,3]，谷　宇[1,3]，
王月明[1,3]，刘　新[1,3]，任　彦[1]，李建军[1,3]

[1]内蒙古科技大学信息工程学院，内蒙古自治区 包头 014010；
[2]内蒙古工业大学信息工程学院，内蒙古自治区 呼和浩特 010051；
[3]内蒙古自治区模式识别与智能图像处理重点实验室，内蒙古自治区 包头 014010

摘要：在行人重识别任务中存在数据集标注难度大，样本量少，特征提取后细节特征缺失等问题。针对以上问题提出深度双重注意力的生成与判别联合学习的行人重识别。首先，构建联合学习框架，将判别模块嵌入生成模块，实现图像生成和判别端到端的训练，及时将生成图像反馈给判别模块，同时优化生成模块与判别模块。其次，通过相邻的通道注意力模块间连接和相邻空间注意力模块间连接，融合所有通道特征和空间特征，构建深度双重注意力模块，将其嵌入教师模型，使模型能更好地提取行人细节身份特征，提高模型识别能力。实验结果表明，该算法在 Market-1501 和 DukeMTMC-ReID 数据集上具有较好的鲁棒性、判别性。

关键词：行人重识别；图像生成；联合学习；注意力机制；深度学习

中图分类号：TP391　　**文献标志码：**A

张晓艳，张宝华，吕晓琪，等. 深度双重注意力的生成与判别联合学习的行人重识别[J]. 光电工程，2021，**48**(5): 200388
Zhang X Y, Zhang B H, Lv X Q, *et al*. The joint discriminative and generative learning for person re-identification of deep dual attention[J]. *Opto-Electron Eng*, 2021, **48**(5): 200388

The joint discriminative and generative learning for person re-identification of deep dual attention

Zhang Xiaoyan[1], Zhang Baohua[1,3*], Lv Xiaoqi[2,3], Gu Yu[1,3], Wang Yueming[1,3],
Liu Xin[1,3], Ren Yan[1], Li Jianjun[1,3]

[1]School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou, Inner Mongolia 014010, China;
[2]School of Information Engineering, Mongolia Industrial University, Huhehaote, Inner Mongolia 010051, China;
[3]Inner Mongolia Key Laboratory of Pattern Recognition and Intelligent Image Processing, Baotou, Inner Mongolia 014010, China

收稿日期：2020-10-20； 收到修改稿日期：2021-03-31
基金项目：国家自然科学基金资助项目(61962046，62001255，61841204)；内蒙古杰青培育项目(2018JQ02)；内蒙古科技计划项目“面向交通大数据智能分析平台的关键技术研究与实现”(202001)；内蒙古草原英才，内蒙古自治区自然科学基金(2019MS06003，2018MS06018)；教育部“春晖计划”合作科研项目(教外司留 1383 号)；内蒙古自治区高等学校科学技术研究项目(NJZY145)资助
作者简介：张晓艳(1995-)，女，硕士研究生，主要从事行人重识别的研究。E-mail：1297118658@qq.com
通信作者：张宝华(1981-)，男，博士，教授，硕士生导师，主要从事数字图像处理及应用、目标识别与跟踪的研究。E-mail：zbh_wj2004@imust.cn

Abstract: In the task of person re-identification, there are problems such as difficulty in labeling datasets, small sample size, and detail feature missing after feature extraction. The joint discriminative and generative learning for person re-identification of the deep dual attention is proposed against the above issues. Firstly, the author constructs a joint learning framework and embeds the discriminative module into the generative module to realize the end-to-end training of image generative and discriminative. Then, the generated pictures are sent to the discriminative module to optimize the generative module and the discriminative module simultaneously. Secondly, according to the connection between the channels of the attention modules and the connection between the attention modules in spaces, it merges all the channel features and spatial features and constructs a deep dual attention module. By embedding the models in the teacher model, the model can better extract the fine-grained features of the objects and improve the recognition ability. The experimental results show that the algorithm has better robustness and discriminative capability on the Market-1501 and the DukeMTMC-ReID datasets.

Keywords: person re-identification; image generative; joint learning; attention; deep learning

1 引 言

行人重识别(Person re-identification, Person ReID)也称行人再识别，在多视角摄像头拍摄的情况下，利用计算机视觉技术判断特定摄像头拍摄的行人图像是否能在大规模行人图像库中检索到相同身份的行人，是图像检索的一类子问题[1]。由于行人重识别应用场景的复杂性，存在视角、遮挡、姿态、尺度和光照变化以及低分辨率等[2]因素的影响，给重识别任务带来极大的挑战。

在传统的行人重识别研究中包括特征提取[3]和距离度量[4]，是基于人工设计的特征，一般应用于小数据集。2014 年以来，随着深度学习的兴起，深度神经网络广泛应用在重识别领域，而小规模数据集无法满足神经网络的需求，且易造成过拟合等问题。Zheng[5]等将生成对抗网络(GAN)应用在重识别领域，提出将无条件 GAN 生成数据融合到训练数据中的半监督模型，解决了训练数据不足的问题。由于数据集之间存在域差异性，使得不同数据集之间训练与测试性能降低。因此，Wei[6]等提出不同数据集之间行人图像的迁移，即保证行人本身前景不变的情况下，将背景风格转换为其他数据集的风格。在行人重识别领域中，姿势的变化也会影响识别的精度，因此，Ge[7]等提出姿态引导的生成对抗网络(pose-guide feature distilling GAN, FD-GAN)，在改变姿态的情况下保持身份特征一致性，通过姿态引导去除冗余特征。Deng[8]等人提出了一种风格迁移学习的框架以及一种生成对抗网络，用无监督学习的方法将有标记图像从源域迁移到目标域，然后通过有监督学习训练迁移图像。然而，上述方法均为数据生成和重识别阶段，是相对独立的，使生成数据利用不充分。

近年来，视觉注意力广泛应用于行人重识别方向。Song[9]等提出一种对比注意模型(mask-guided contrastive attention model, MGCAM)从身体和背景区域对比学习特征。Xu[10]等提出注意力感知组成网络(attention-aware compositional network, AACN)，利用注意力模块获取精确的目标部位以及对全局特征对齐，排除背景干扰。Li[11]等提出协调注意力模型(harmonious attention network for person re-identification, HA-CNN)，共同学习基于像素的软注意力特征和硬注意力特征，将其应用于错位图像。上述注意力的方法均为排除背景噪声干扰，且只考虑单独注意力模块提取的特征。

针对上述方法存在的问题，本文提出基于深度双重注意力的生成与判别联合学习的行人重识别模型。将生成模块与判别模块联合统一[12]，使生成数据在线反馈给判别模块，同时优化生成模块和判别模块，实现模块间端到端的训练。受文献[13-14]启发，提出深度双重注意力模块(DDA)，通过连接相邻注意力模块，促使注意力模块之间信息交流，增强注意力模块提取特征的能力。

2 基本原理

2.1 师生联合网络框架

本文网络框架主要由学生模型和基于深度双重注意力机制的教师模型组成，如图 1 所示。学生模型包括外观编码器(appearance encoder, A_e)，结构编码器(structure encoder, S_e)，解码器(decoder, D_e)，鉴别器(discriminator, D)等。其中外观编码器也是判别模块，即判别模块通过共享外观编码器嵌入生成模块。图像生成方式包括：身份一致的图像重构，交叉身份交叉

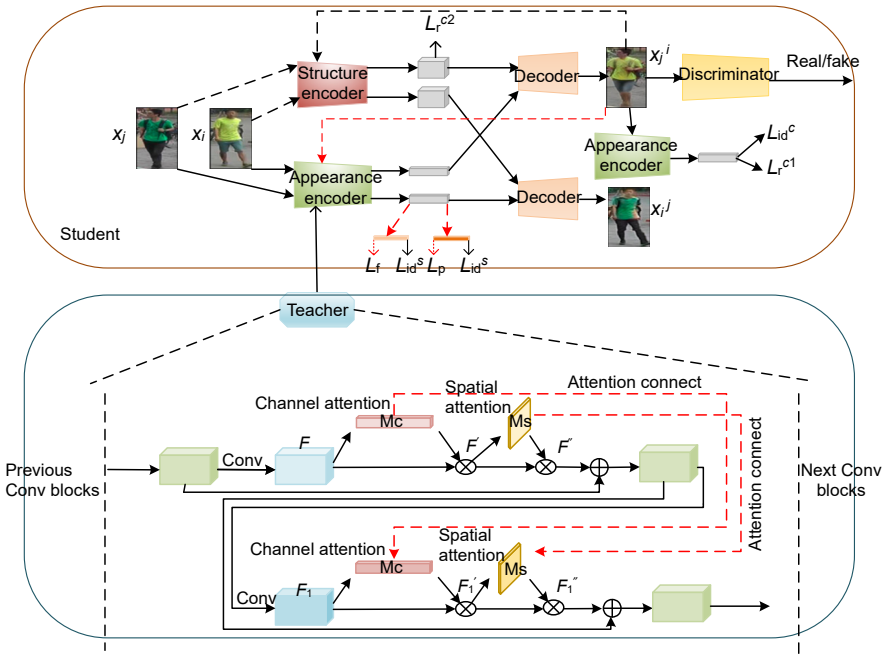


图 1 师生联合网络框架

Fig. 1 Framework for teacher- student network

图像的合成。以上方法均为将图像分别输入外观编码器和结构编码器,输出外观特征向量和结构特征向量,通过解码器交换外观和结构特征向量生成图像[12]。由于学生模型中图像生成和判别是联合统一训练，使得生成图像实时反馈给外观编码器，优化判别模块的同时也改善外观编码器生成的外观特征向量。通过教师模型[15]辅助学生模型学习主要身份特征，将生成的图像作为训练样本。但由于生成的图像相似度高，增加教师模型的识别难度，进而会影响学生模型识别的准确率。为了解决该问题，提出基于深度双重注意力机制的教师模型，该模型由 ResNet50[15]网络和深度双重注意力机制组成。将卷积得到的特征图输入到通道注意力模块，得到具有通道注意力的特征图，作为空间注意力模块的输入，再通过注意力连接网络将同类的注意力模块连接,使各模块间提取的注意力特征融合，提高注意力模块的学习能力，避免信息在传递过程中频繁变化[14]。

2.2 学生模型

用 $X=\{x_i\}_{i=1}^{N}$ 表示真实图像， $Y=\{y_i\}_{i=1}^{N}$ 表示身份标签，N 为图像数目， $y_i\in[1,K]$ ，K 为图像编码在数据集中身份编号。

2.2.1 身份一致的图像重构

身份一致的图像重构即相同身份的一张或两张图像重构。给定一张图像，分别输入到外观编码器和结构编码器，得到外观特征向量和结构特征向量，再通过解码器得到合成图像。相同身份重构的图像使生成器起到正则化的作用。

$$L_{\mathrm{r}}^{\mathrm{imag1}}=\xi[\|x_i-D_{\mathrm{e}}(a_i,\mathrm{s}_i)\|_1] \quad 。 \tag{1}$$

如式(1)所示，该图像的重构采用像素级的损失函数，即若生成的图像与目标图像相同，则像素差为 0。

由于同一个人的不同图像其外观特征相近，且具有相同身份标签。因此，采用式(2)所示的损失函数，缩短相同身份外观特征向量的距离，增大不同身份的外观特征向量。

$$L_{\mathrm{r}}^{\mathrm{imag2}}=\xi[\|x_i-D_{\mathrm{e}}(a_t,\mathrm{s}_i)\|_1] \quad , \tag{2}$$

$$L_{\mathrm{id}}^{s}=\xi[-\log(p(y_i\,|\,x_i))] \quad 。 \tag{3}$$

由于外观特征携带身份信息，因此采用式(3)所示的损失函数， $p(y_i\,|\,x_i)$ 是基于外观特征向量去预测 x_i 属于真实类别 y_i 的概率。

2.2.2 交叉身份交叉图像的合成

交叉身份交叉图像的合成即任意两张不同身份和不同图像进行的重构。合成图像无身份标签，无法采用像素级别的监督。将合成图像重新编码为新的外观特征向量和结构特征向量，利用式(4)、式(5)所示的损失函数计算合成图像和真实图像之间的损失。

$$L_{\mathrm{r}}^{\mathrm{c1}}=\xi[\|a_i-A_{\mathrm{e}}(D_{\mathrm{e}}(a_i,\mathrm{s}_j))\|_1] \quad , \tag{4}$$

$$L_{\mathrm{r}}^{c2}=\xi[\|s_j-S_{\mathrm{e}}(D_{\mathrm{e}}(a_i,s_j))\|_1]\ 。 \tag{5}$$

利用式(6)提供身份监督，让其与提供外观特征向量的真实图像保持身份一致性。

$$L_{\mathrm{id}}^{c}=\xi[-\log(p(y_i\,|\,x_i))]\ 。 \tag{6}$$

利用式(7)使生成数据的分布接近真实数据的分布。

$$L_a=\xi[\log D(x_i)+\log(1-D(D_{\mathrm{e}}(a_i,s_j)))]\ 。 \tag{7}$$

2.2.3 图像判别

判别模块通过共享外观编码器嵌入到图像生成模块中，本文通过融合主要身份特征和细粒度特征对行人图像进行判别。由基于注意力机制的教师模型辅助学生模型学习主要身份特征，学生模型单独学习细粒度特征。

$$L_p=\xi[-\sum_{i=1}^{K}q(k\,|\,x_j^i)\log(\frac{p(x_j^i)}{q(x_j^i)})]\ , \tag{8}$$

$$L_f=\xi[-\log(p(y_i\,|\,x_i))]\ 。 \tag{9}$$

如式(8)所示，学生模型预测 $p(x_j^i)$ 概率分布和教师模型预测的 $q(x_j^i)$ 概率分布之间采用最小化 KL 距离。将合成的图像作为训练样本，迫使学生模型学习与衣附无关的一些细粒度特征，由于外观特征包括身份信息，因此采用式(9)所示的身份损失函数，保证在学习细粒度特征的同时保持身份一致性。

2.3 基于深度双重注意力机制的教师模型

教师模型采用 ResNet50[15]作为基础网络。残差网络加速深度神经网络的训练，提升深度网络的准确率。此外，残差网络在很大程度上避免网络层数的增加而产生的梯度消失或梯度爆炸的问题[16]。将生成图像作为训练样本，无需手动标记行人属性，可自动从合成的图像中采集细节属性。采用师生监督模型，教师模型动态地分配一个软标签给合成图像 x_j^i，外观来自 x_i，结构来自 x_j。由于行人图像相似度高且图像质量差，增加教师模型的识别难度，降低教师模型的辅助学生模型学习主要身份特征的能力，因此引入深度双重注意力机制，帮助教师模型挖掘更深层的身份特征，提高学生模型判别性。

2.3.1 深度双重注意力机制

自我注意力机制在许多视觉任务中表现出优越的效果，但仅考虑了单独注意力模块提取的特征，无法充分融合注意力块之间的特征。受文献[13-14]启发，本文提出了深度双重注意力机制，将相邻的通道注意块与通道注意块、空间注意块与空间注意块之间连接起来，使得注意力模块之间可以互相进行信息交流，联合所有注意力模块进行训练，增强注意力模块学习的能力，挖掘更深的注意力特征。

通道注意块为给定一个特征图 $F\in R^{C\times H\times W}$ 作为输入，首先经过平均池化和最大池化聚合特征映射的空间信息，生成两个不同的空间上下文描述符：F_{avg}^{c} 和 F_{max}^{c}，分别表示平均池化和最大池化。两个描述符送到一个共享网络，以产生通道注意力图 $M_C\in R^{C\times1\times1}$。将共享网络应用于每个描述符之后，使用逐元素求和合并输出特征向量[13]。

通道注意模块的数学式：

$$\begin{aligned}M_C(F)&=\sigma(MLP(\mathrm{AvgPool}(F))+MLP(\mathrm{MaxPool}(F)))\\&=\sigma(W_1(W_0(F_{\mathrm{avg}}^{c}))+W_1(W_0(F_{\mathrm{max}}^{c})))\ 。\end{aligned} \tag{10}$$

将通道注意力输出的特征图作为空间注意力块的输入，使用最大池化和平均池化操作聚合特征映射的通道信息。然后经过卷积层降维，再经过 Sigmoid 函数产生二维空间注意图。空间注意块的计算式：

$$\begin{aligned}M_S(F)&=\sigma(f^{7\times7}([\mathrm{AvgPool}(F);\mathrm{MaxPool}(F)]))\\&=\sigma(f^{7\times7}([F_{\mathrm{avg}}^{s};F_{\mathrm{max}}^{s}]))\ ,\end{aligned} \tag{11}$$

其中：σ 表示 Sigmoid 函数，$W_0\in R^{C/r\times C}$，$W_1=R^{C\times C/r}$，7×7 表示卷积核大小。

总体过程可以概括为

$$F'=M_C(F)\otimes F,\ F''=M_S(F')\otimes F'\ , \tag{12}$$

其中：$\otimes$ 表示逐元素相乘，F'' 是最终的优化输出。

注意力连接网络[14]通过参数化的加法操作将当前注意力特征与之前的注意力特征结合，确保信息在注意力块间以前馈的方式传递，避免信息在传递过程中频繁变动的问题，在不改变模型内部结构的同时，提高注意力模块的学习能力。

通道与通道、空间与空间注意力模块之间的连接函数：

$$f(\alpha T_n,\beta\tilde{F}_n)=\alpha T_n+\beta\tilde{F}_n\ , \tag{13}$$

其中：n 是特征的索引，$\tilde{F}_n$ 可以看作是 T 的增强。α 和 β 均为可学习参数，T 为特征提取器提取的特征，$\tilde{F}_n$ 为前一个注意力模块生成的注意力映射特征图。如果将 α 和 β 分别设置为 1 和 0，此时为普通的注意力模块。

3 实验结果与分析

3.1 数据集评估与评价标准

为了验证提出模型的有效性，本文分别在 Market1501，DukeMTMC-ReID 两个主流公开数据集上进行有效性的验证。Market1501 数据集包含 6 个摄

像头(其中 5 个高清摄像头和 1 个低清摄像头)，共有 1501 个行人的 32668 张图像，其中训练集 751 人，包含 12936 张图像；另外测试集 750 人，包含 19732 张图像。DukeMTMC-ReID 数据集是 DukeMTMC 数据集的一个子集，用于研究行人重识别，该数据集包含 8 个摄像头，共 1404 个行人的 36411 张图像，随机选择 702 个行人的 16522 张图像作为训练集，另外的 702 个行人的 19889 张图像作为测试集。

本次实验使用首位命中率 Rank-1 和平均精度均值 mAP 作为评价指标。

3.2 实验配置

实验基于 PyTorch 1.1 框架，硬件配置采用处理器为 Intel(R) Xeon(R) CPU E5-1650 V4 3.60 GHz，两块 NVIDIA GeForce RTX 2080 Ti 的 GPU，软件环境为 Ubuntu-16.04。本实验中联合网络训练数据的最大迭代次数为 100000 次，每批次的样本数为 8，训练共耗时 22 h。

3.3 实验细节

使用 $c\times h\times w$ 表示特征映射的大小。外观编码器是基于 ResNet50 预训练的 ImageNet 模型，移除全局平均池化层和全连接层，然后添加一个最大池化层输出外观特征向量，采用 SGD 优化器，其学习率设置为 0.002，动能设置为 0.9。编码器和解码器均由 4 个卷积层和 4 个跳跃连接块组成。鉴别器采用多尺度图像输入。结构编码器、解码器、鉴别器使用 Adam 优化器，其学习率设置为 0.0001。

3.4 实验结果分析

教师模型的参数设置对学生模型学习主要特征的能力影响较大，在 ResNet50 基础网络上优化教师模型，在 Market1501 数据集和 DukeMTMC-ReID 数据集上 Rank-1 精度和 mAP 分别为 86.66%、65.14%、81.32%、64.08%。

本实验中当教师模型的参数设置训练次数 epoch 为 60，每批次的样本数为 8、学习率为 0.02 时，结果达到最优。如图 2 所示，当学习率为 0.02 时，Rank-1 精度和 mAP 分别为 90.74%和 75.05%。由图 3 可知，加入双重注意力模块后，会比基准网络多耗时近半小时，是因为双重注意力模块促进基准网络提取通道和空间位置的信息，然后进行特征融合。而在此基础上加入深度注意力连接网络，耗时增加近 1 h，是因为深度注意力连接网络增强了双重注意力模块提取特征的能力，将前一个通道注意力模块的提取特征以前馈方式传递给相邻的通道注意力模块，空间注意力模块同理，最后融合通道特征和空间特征，降低训练速度，提高了提取特征的性能。

当学习率为 0.02 时教师模型最优，为验证所提算法的有效性，分别将双重注意力机制(DA)和深度双重注意力机制(DDA)引入最优的教师模型，进行消融实验。

在 Market1501 数据集和 DukeMTMC-ReID 数据集上，加入双重注意力机制之后，如表 1 所示，相对基准网络识别精度稍有提升。由此可以看出，双重注意力模块能有效地捕捉通道和空间位置特征，对于教师模型的识别效果有相应的提升，使得该模型能更好地关注主要特征。将深度双重注意力机制引入教师模型之后，相对基准网络，在 Market1501 数据集和 DukeMTMC-ReID 数据集上 Rank-1 精度和 mAP 分别提高了 4.04%、9.91%、2.07%和 1.47%。这说明深度连接注意力网络增强了双重注意力模块获取通道和空间位置信息的能力，充分融合了通道特征和空间特征，以挖掘更深层次的特征。将引入深度双重注意力机制

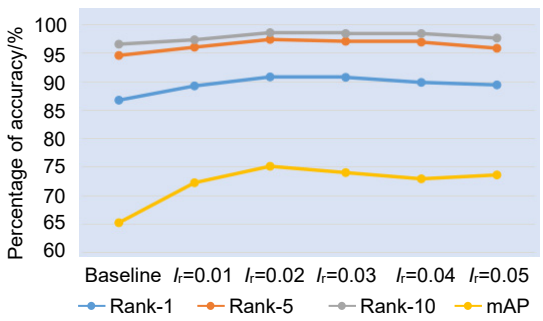


图 2 学习率

Fig. 2 Learning rate

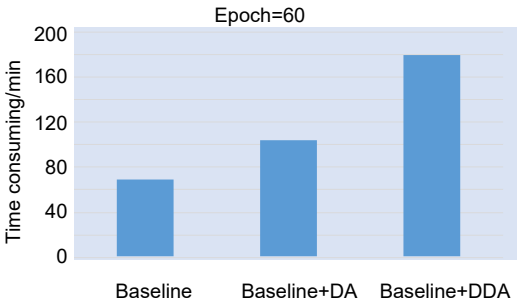


图 3 不同方法的耗时对比

Fig. 3 Time-consuming comparison of different methods

表 1 消融实验

Table 1 Ablation study

Methods	Market1501		DukeMTMC-ReID	
	Rank-1	mAP	Rank-1	mAP
Baseline	86.66	65.14	81.32	64.08
Baseline+DA	87.73	69.57	81.23	64.45
Baseline+DDA	90.74	75.05	83.39	65.55
Prime+DDA	94.15	85.44	85.91	74.52
Ours	94.74	86.39	86.49	75.01

的最优教师模型用于辅助学生模型学习主要特征，如表 1 所示，在 Market1501 数据集和 DukeMTMC-ReID 数据集上 Rank-1 精度和 mAP 分别提升至 94.15%、85.44%、85.91%和 74.52%。由于判别模型是由主要特征的学习和细粒度特征的学习联合作用进行判别，故最终识别结果为在 Market1501 数据集上 Rank-1 精度和 mAP 分别提升至 94.74% 和 86.39%，在 DukeMTMC-ReID 数据集上 Rank-1 精度和 mAP 分别提升至 86.49%和 75.01%。

为验证深度注意力模块的有效性，对加入注意力机制的不同阶段进行可视化对比，如图 4 所示。

图 4(a)为原始输入图像；图 4(b)为基准网络可视化结果，此时该网络所关注的重心仅在其右侧，关注重点较少；图 4(c)在基准网络的基础上加入双重注意力机制，网络关注的重心有所扩大，可以看出注意力模块增加网络所关注的重点；图 4(d)为基准网络结合深度双重注意力机制，此时网络关注的重心聚焦在具有明显区分行人信息的上半身，证明注意力连接网络将各模块间的注意力特征融合，避免了信息传递过程中频繁变动的问题，确保关注重点不变的情况下增加关注范围；图 4(e)为深度双重注意力机制结合教师模型辅助学生模型所学的主要特征信息，此时网络关注的重点范围有所延伸。由此可知，深度双重注意力模块可以使教师模型准确且全面地学习主要身份特征，提高模型的识别精度。

为验证本文算法的优越性，将本文算法与近年来相关算法在两个数据集 Market1501 和 DukeMTMC-ReID 上进行对比，如表 2 所示。相关算法如下文。

1) 注意力相关算法：注意力感知组成网络(attention-aware compositional network, AACN)、协调注意力网络(harmonious attention network, HA-CNN)、局部注意力网络(a part-based attention network, PBAN)。

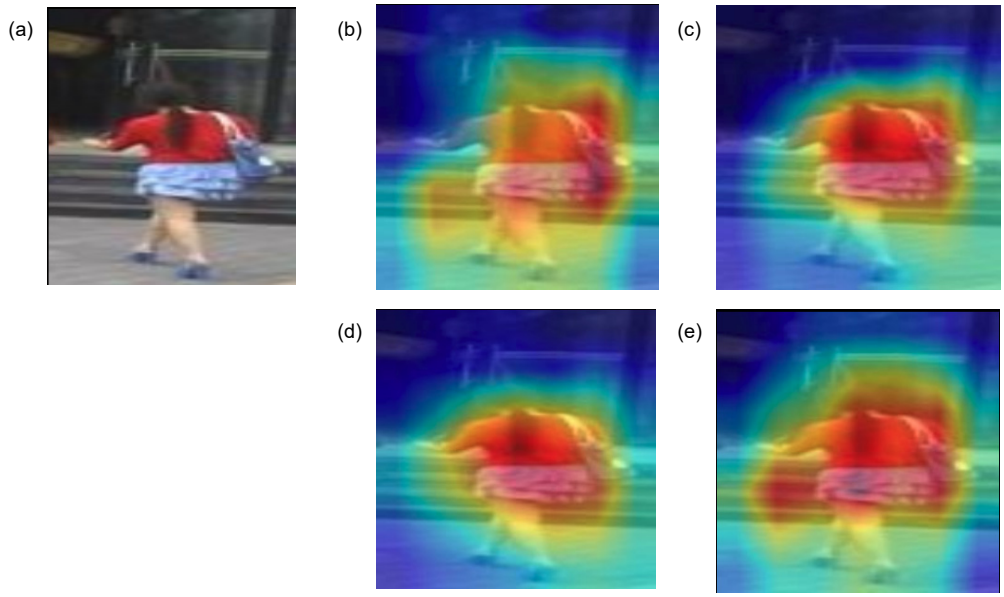


图 4 注意力机制不同阶段可视化对比结果

(a) 输入图像；(b) 基准网络；(c) 加入双重注意力机制；(d) 加入深度双重注意力机制；(e) 教师模型辅助学生模型

Fig. 4 Visual contrast results of different stages of attention mechanism.
(a) Input image; (b) Baseline; (c) Add dual attention mechanism;
(d) Add the deep dual attention mechanism; (e) Teacher model aided student model

表 2 与主流行人重识别方法精度对比

Table 2 Accuracy comparison with the main popular re-identification method

Methods	Market1501		DukeMTMC-ReID	
	Rank-1	mAP	Rank-1	mAP
AACN[10]	85.90	66.87	76.84	58.25
HA-CNN[11]	91.2	75.7	80.5	63.8
PBAN[17]	93.6	81.7	84.7	69.4
GLAD[18]	89.9	73.9	-	-
PGFA[19]	91.2	76.8	82.6	65.5
CBN[20]	91.3	77.3	82.5	67.3
Deep-Person[21]	92.31	79.58	80.90	64.80
VPM[22]	93.0	80.8	83.6	72.6
PN-GAN[23]	89.4	72.6	73.6	53.2
FD-GAN[7]	90.5	77.7	80.0	64.5
Ours	**94.74**	**86.39**	**86.49**	**75.01**

2) 未采用生成数据进行训练的方法：用于行人检索的全局局部对齐描述符(global-local-alignment descriptor for pedestrian retrieval，GLAD)、基于遮挡行人的姿势引导的特征对齐(pose-guided feature alignment for occluded person re-identification，PGFA)。感知重点：学习残缺行人的可视化局部特征(perceive where to focus: learning visibility-aware part-level features for partial person re-identification, VPM)、学习判别性的深度特征(learning discriminative deep features for person re-identification, Deep-Person)、基于相机批量归一化的行人分布差距的再思考(rethinking the distribution gap of person re-identification with camera-based batch normalization, CBN)等。

3) 数据生成和判别相对独立的方法：姿态归一化的图像生成(pose-normalized image generative for person re-identification，PN-GAN)、基于鲁棒行人的姿态引导的特征提取的生成对抗网络(pose-guided feature distilling gan for robust person re-identification, FD-GAN)等。由表中数据可知，本文提出的方法相较于其他主流方法性能明显提高。

相较于关注部分注意力的 AACN 和关注像素的软注意力特征和硬注意力特征的 HA-CNN，PBAN 利用注意机制来缓解错位问题，并利用全局-局部特征的互补效应，稳定地描述行人特征，在两个数据集上精度有效地提高，但 PBAN 无法充分地将注意力模块间信息相互传递。在本方法中，通过注意力连接网络分别将通道注意力模块相互连接和空间注意力模块相互连接，使模型中所有的注意力模块联合训练，提高注意力模块的学习能力。

相较于经典的 GLAD，PGFA 使用关键点信息解决行人遮挡的问题，CBN 解决了相机之间差异问题造成识别精度低的问题，Deep-person 考虑不同部件之间的上下文信息和空间信息，VPM 解决了行人局部识别所造成的空间不对齐的现象，但以上方法无充足的样本量。在本方法中，采用生成的数据进行训练模型，扩充数据样本，提高模型性能。

相较于针对重识别中的姿态归一化而设计的 PN-GAN，FD-GAN 解决了姿态变化的问题，但此方法采用的生成数据和判别是相对独立的两个阶段，无法将生成的图像及时用做训练样本。在本方法中，采用生成数据和判别联合学习的网络，使生成模块和判别模块采用对抗原理相互优化，提高模型的识别能力。

为进一步验证算法的实时性，将该算法与相关算法的在数据集 Market1501 中进行测试对比，如表 3 所示。

由表 3 可知，所提算法识别速度优于 GLAD 和 CBN，但略差于 PGFA，以运行速度换取精度。由于在实时监控系统中，图像检索库也在实时增加，在匹配时考虑新增行人即可，本文匹配单张图像所耗费时间为 0.0162 s，足以满足实时监控的条件。

表 3 算法测试时间对比结果

Table 3 Comparative results of test time of different methods

Methods	Time/s	
	Test time	Per match(19732)
GLAD[18]	368	0.0186
PGFA[19]	315	0.0159
CBN[20]	347	0.0175
Ours	321	0.0162

4 结 论

本文提出的深度双重注意力的生成与判别联合学习的行人重识别，通过联合框架将生成模块与判别模块联合统一，将生成数据在线反馈给判别模块，同时优化生成模块与判别模块，充分利用生成数据。通过引入深度双重注意力模块，使得注意力块之间的信息相互流动，强化注意力块获取通道和空间位置信息的能力，提高教师模型的教学能力，帮助学生模型学习

较深层次的特征，结合细粒度特征之后达到最优性能。通过在 Market1501 和 DukeMTMC-ReID 两个数据集上的实验验证本文提出的方法有效性，相较于其他主流算法有较大地精度提升。

参考文献

[1] Song W R, Zhao Q Q, Chen C H, *et al*. Survey on pedestrian re-identification research[J]. *CAAI Trans Intell Syst*, 2017, **12**(6): 770–780.
宋婉茹, 赵晴晴, 陈昌红, 等. 行人重识别研究综述[J]. 智能系统学报, 2017, **12**(6): 770–780.

[2] Feng X, Du J H, Duan Y N, *et al*. Research on person re-identification based on deep learning[J]. *Appl Res Comput*, 2020, **37**(11): 3220–3226, 3240.
冯霞, 杜佳浩, 段仪浓, 等. 基于深度学习的行人重识别研究综述[J]. 计算机应用研究, 2020, **37**(11): 3220–3226, 3240.

[3] Matsukawa T, Okabe T, Suzuki E, *et al*. Hierarchical Gaussian descriptor for person re-identification[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016: 1363–1372.

[4] Lisanti G, Masi I, Bagdanov A D, *et al*. Person re-identification by iterative re-weighted sparse ranking[J]. *IEEE Trans Pattern Anal Mach Intell*, 2015, **37**(8): 1629–1642.

[5] Zheng Z D, Zheng L, Yang Y. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro[C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*, Venice, Italy, 2017: 3754–3762.

[6] Wei L H, Zhang S L, Gao W, *et al*. Person transfer GAN to bridge domain gap for person re-identification[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018: 79–88.

[7] Ge Y X, Li Z W, Zhao H Y, *et al*. FD-GAN: pose-guided feature distilling GAN for robust person re-identification[C]//*Proceedings of the 32nd International Conference on Neural Information Processing Systems*, Montréal, Canada, 2018: 1222–1233.

[8] Deng W J, Zheng L, Ye Q X, *et al*. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018: 994–1003.

[9] Song C F, Huang Y, Ouyang W L, *et al*. Mask-guided contrastive attention model for person re-identification[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018: 1179–1188.

[10] Xu J, Zhao R, Zhu F, *et al*. Attention-aware compositional network for person re-identification[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018: 2119–2128.

[11] Li W, Zhu X T, Gong S G. Harmonious attention network for person re-identification[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018: 2285–2294.

[12] Zheng Z D, Yang X D, Yu Z D, *et al*. Joint discriminative and generative learning for person re-identification[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 2019: 2138–2147.

[13] Woo S, Park J, Lee J Y, *et al*. CBAM: convolutional block attention module[C]//*Proceedings of the 15th European Conference on Computer Vision*, Munich, Germany, 2018: 3–19.

[14] Ma X, Guo J D, Tang S H, *et al*. DCANet: learning connected attentions for convolutional neural networks[Z]. arXiv: 2007.05099, 2020.

[15] He K M, Zhang X Y, Ren S Q, *et al*. Deep residual learning for image recognition[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016: 770–778.

[16] Guo Y X, Yang W, Liu Q, *et al*. Survey of residual network[J]. *Appl Res Comput*, 2020, **37**(5): 1292–1297.
郭玥秀, 杨伟, 刘琦, 等. 残差网络研究综述[J]. 计算机应用研究, 2020, **37**(5): 1292–1297.

[17] Zhong W L, Jiang L F, Zhang T, *et al*. A part-based attention network for person re-identification[J]. *Multimed Tools Appl*, 2020, **79**(31): 22525–22549.

[18] Wei L H, Zhang S L, Yao H T, *et al*. GLAD: global-local-alignment descriptor for scalable person re-identification[J]. *IEEE Trans Multimed*, 2019, **21**(4): 986–999.

[19] Miao J X, Wu Y, Liu P, *et al*. Pose-guided feature alignment for occluded person re-identification[C]//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*, Seoul, Korea, 2019: 542–551.

[20] Zhuang Z J, Wei L H, Xie L X, *et al*. Rethinking the distribution gap of person re-identification with camera-based batch normalization[C]//*Proceedings of the 16th European Conference on Computer Vision*, Glasgow, UK, 2020: 140–157.

[21] Bai X, Yang M K, Huang T T, *et al*. Deep-person: learning discriminative deep features for person Re-identification[J]. *Pattern Recognit*, 2020, **98**: 107036.

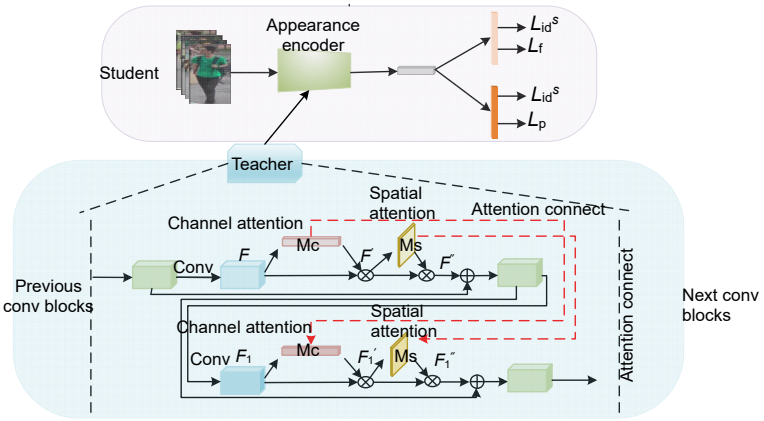
[22] Sun Y F, Xu Q, Li Y L, *et al*. Perceive where to focus: learning visibility-aware part-level features for partial person re-identification[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 2019: 393–402.

[23] Qian X L, Fu Y W, Xiang T, *et al*. Pose-normalized image generation for person re-identification[C]//*Proceedings of the 15th European Conference on Computer Vision*, Munich, Germany, 2018: 650–667.

The joint discriminative and generative learning for person re-identification of deep dual attention

Zhang Xiaoyan[1], Zhang Baohua[1,3*], Lv Xiaoqi[2,3], Gu Yu[1,3],
Wang Yueming[1,3], Liu Xin[1,3], Ren Yan[1], Li Jianjun[1,3]

[1]School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou, Inner Mongolia 014010, China;
[2]School of Information Engineering, Mongolia Industrial University, Huhehaote, Inner Mongolia 010051, China;
[3]Inner Mongolia Key Laboratory of Pattern Recognition and Intelligent Image Processing, Baotou, Inner Mongolia 014010, China



Framework for teacher-student network

Overview: Person re-identification is a technology that uses computer technology to determine whether there is a specific object in an image or video sequence. In the task of person re-identification, there are problems such as difficulty in labeling datasets, small sample size, and detail feature missing after feature extraction.The joint discriminative and generative learning for person re-identification of deep dual attention is proposed against the above issues. Firstly, the author constructs a joint learning framework and embeds the discriminative module into the generative module to realize the end-to-end training of images generative and discriminative. Then the generated pictures are sent to the discriminative module to optimize the generative module and discriminative module simultaneously. Secondly, we construct a deep dual attention module. Through the connection between the channels of the attention modules and the connection between attention modules in spaces, the information collected by the previous attention module is passed to the next attention module. The attention feature is fused with the currently extracted attention feature to ensure that the information between attention modules of the same dimension can flow in a feed-forward manner, effectively avoiding the frequent changes of the information between attention modules. At last, the author merges all channel features and spatial features. Due to the high similarity of the appearance of the images in the dataset, the teacher model recognition becomes more difficult. Therefore, the deep dual attention module is embedded in the teacher model to improve the recognition ability of the teacher model. The teacher model is used to assist the student model to learn the main features. The generated images are used as training samples to force students to learn fine-grained features independent of appearance features. The author merges the main features and fine-grained features to re-identification person. The experimental results show that Rank-1 and mAP used in this paper are superior to the advanced correlation algorithms.

Zhang X Y, Zhang B H, Lv X Q, *et al.* The joint discriminative and generative learning for person re-identification of deep dual attention[J]. *Opto-Electron Eng*, 2021, **48**(5): 200388; DOI: 10.12086/oee.2021.200388

Foundation item: National Natural Science Foundation (61962046, 61663036, 61841204), Inner Mongolia Outstanding Youth Cultivation Project (2018JQ02), Inner Mongolia Science and Technology Planning Project "Research and Realization of Key Technologies for the Intelligent Analysis Platform of Traffic Big Data"(202001), Inner Mongolia Prairie Talent, Inner Mongolia Youth Science and Technology Innovation Talent Project, Inner Mongolia Autonomous Region Natural Science Fund (2015MS0604, 2018MS06018), and Inner Mongolia Autonomous Region Higher Education Science Funded by the Technical Research Project (NJZY145)

* E-mail: zbh_wj2004@imust.cn