E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(2023)05-1346-14

论文引用格式:Yang L L, Lan L, Sun D T, Teng X, Ben X Y and Shen X B. 2023. Newly low-resolution pedestrian re-identification-relevant dataset and its benched method. Journal of Image and Graphics, 28(05):1346-1359(杨露露, 蓝龙, 孙冬婷, 滕霄, 贲晛烨, 沈肖波. 2023. 低分辨率行人重识别数据集及其基准方法. 中国图象图形学报, 28(05):1346-1359)[DOI:10.11834/jig.221082]

低分辨率行人重识别数据集及其基准方法

杨露露[1,2],蓝龙[1,2*],孙冬婷[1,2],滕霄[1,2],贲晛烨[3],沈肖波[4]

1. 国防科技大学计算机学院,长沙　410073; 2. 国防科技大学量子信息研究所兼高性能计算国家重点实验室,长沙　410073;
3. 山东大学信息科学与工程学院,青岛　266237; 4. 南京理工大学计算机科学与工程学院,南京　210094

摘　要:**目的**　行人重识别旨在解决多个非重叠摄像头下行人的查询和识别问题。在很多实际的应用场景中,监控摄像头获取的是低分辨率行人图像,而现有的许多行人重识别方法很少关注真实场景中低分辨率行人相互匹配的问题。为研究该问题,本文收集并标注了一个新的基于枪球摄像头的行人重识别数据集,并基于此设计了一种低分辨率行人重识别模型来提升低分辨率行人匹配性能。**方法**　该数据集由部署在3个不同位置的枪机摄像头和球机摄像头收集裁剪得到,最终形成包含200个有身份标签的行人和320个无身份标签的行人重识别数据集。与同类其他数据集不同,该数据集为每个行人同时提供高分辨率和低分辨率图像。针对低分辨率下的行人匹配难题,本文提出的基准模型考虑了图像超分、行人特征学习以及判别3个方面因素,并设计了相应的超分模块、特征学习模块和特征判别器模块,分别完成低分辨率图像超分、行人特征学习以及行人特征判断。**结果**　提出的基准模型在枪球行人重识别数据集上的实验表明,对比于经典的行人重识别模型,新基准模型在平均精度均值(mean average precision,mAP)和Rank-1指标上分别提高了3.1%和6.1%。**结论**　本文构建了典型的低分辨率行人重识别数据集,为研究低分辨率行人重识别问题提供了重要的数据来源,并基于该数据集研究了低分辨率下行人重识别基础方法。研究表明,提出的基准方法能够有效地解决低分辨行人匹配问题。

关键词:行人重识别;基准数据集;低分辨率(LR);超分辨率(SR);判别器

Newly low-resolution pedestrian re-identification-relevant dataset and its benched method

Yang Lulu[1,2], Lan Long[1,2*], Sun Dongting[1,2], Teng Xiao[1,2], Ben Xianye[3], Shen Xiaobo[4]

1. College of Computer Science, National University of Defense Technology, Changsha　410073, China;
2. Institute for Quantum Information & State Key Laboratory of High Performance Computing, National University of Defense Technology, Changsha　410073, China; 3. College of Information Science and Engineering, Shandong University, Qingdao　266237, China; 4. College of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing　210094, China

Abstract: **Objective** Pedestrian re-identification is focused on multiview non-overlapping-derived problem of querying and identifying the same identity pedestrian. However, such real-world application scenarios are challenged to some camera-

收稿日期:2022-11-04;**修回日期**:2023-02-08;**预印本日期**:2023-02-15

*通信作者:蓝龙　long.lan@nudt.edu.cn

基金项目:国家重点研发计划资助(2020AAA0103501);国家自然科学基金项目(62176126,62271289);广东省基础与应用基础研究基金项目(2022A1515010186)

Supported by: National Key R&D Program of China (2020AAA0103501); National Natural Science Foundation of China (62176126, 62271289); Guangdong Basic and Applied Basic Research Foundation (2022A1515010186)

relevant factors like 1) hardware, 2) shooting distance, 3) angle of view, 4) background clutters, and 5) occlusions. Current surveillance camera-based images can be captured and it is still challenged for its low resolution (LR) as well. In real scenes, pedestrian re-identification (re-id) methods are required to resolve multiple pedestrians-oriented heterogeneous problem for low resolution image further. To deal with the mismatch problem between high resolution (HR) images and LR images, conventional re-id methods are mainly concerned of the cross-resolution matching problem. It is essential to richer mutual-benefited ability between low-resolution gallery images and query images. To improve the low-resolution pedestrian matching performance, we develop a novel of gun-ball camera-based pedestrian re-identification benchmark dataset and a low-resolution pedestrian re-identification benchmark model is designed as well. **Method** This data collection is acquired by the gun and ball system, which is deployed at three intersections. To capture LR images, two of three cameras are placed at each intersection, and the gun camera has a fixed direction and focal length. To obtain high-resolution images more, the other ball camera can be used to tune the focal length and the direction of view according to the target pedestrian position. And, a pedestrian re-identification dataset can be built and sampled by 200 pedestrians-identified categories (the same pedestrian is captured and identified at different locations), and a sample of 320 pedestrian-unidentified categories (pedestrians can be captured under a certain camera only). Each of these pedestrians-related images are all in related to the two aspects of high resolution and low resolution. A pedestrian-identified image can be captured by at least 2 different gun-ball cameras from different places, and a pedestrian-unidentified image can be captured by one gun-ball camera only and it is required to be searched and matched across cameras further. Pedestrian-unidentified images are in relevance to both of LR & HR as well. Some optimal factors are illustrated as mentioned below: 1) a richer and more diverse pedestrian dataset: the gun-ball camera-based pedestrian re-identification dataset can be used to acquire various pedestrian images from intersections. 2) The pedestrian dynamics: each pedestrian image has its temporal information because the gun-ball camera-based pedestrian dataset is captured and cropped from the video stream. Such temporal-based pedestrian images can be used for video-related pedestrian re-identification as well. 3) Other potentials: some pedestrians-unidentified images can be focused on, which can be used to study pedestrian re-identification algorithms in semi-supervised or unsupervised domains, as well as the ground truth of identification systems. That is, given an unknown identity of a pedestrian, its identification system can automatically detect the similar one in the surveillance screen or database. To strengthen the matching problem of LR images of pedestrians, we consider 1) image super-resolution, 2) pedestrian-related feature learning, and 3) discrimination as three key factors in our baseline. Specifically, to optimize each aspect of resolution, pedestrian feature-related learning and discrimination, the baseline is involved of a super-resolution module, a pedestrian re-identification module, and a pedestrian feature discriminator. Therefore, we design a baseline pedestrian re-identification model and it is benched on the 5 following aspects: generator $\boldsymbol{G}$, image discriminator $\boldsymbol{D}_s$, gradient discriminator $\boldsymbol{D}_g$, pedestrian feature extractor $\boldsymbol{F}$, and pedestrian feature discriminator $\boldsymbol{D}_f$. To resolve the problem of low-resolution pedestrian re-identification in real scenes, our proposed model can be used to optimize both of the resolution of pedestrian image and pedestrian discrimination features. **Result** The low-resolution pedestrian re-identification baseline model is demonstrated and experimentally validated on the gun-ball pedestrian re-identification dataset. The mean average precision (mAP) and Rank-1 metrics are improved by 3.1% and 6.1%. **Conclusion** LR-related pedestrian recognition in natural scenes can be facilitated, and its pixel misalignment-derived problem of low quality of generated super-resolved images can be resolved to a certain extent. The dataset and benchmark model are proposed and its potential is in related to pedestrian re-identification and image super-resolution. It provides a data source for the field of low-resolution pedestrian re-identification. Also, the proposed baseline model can be predicted to tackle the low-resolution pedestrian matching problem further.

Key words: pedestrian re-identification; benchmark dataset; low-resolution(LR); super-resolution(SR); discriminator

0 引　言

行人重识别(Lan 等，2020; Liang 等，2021a; Zhang等,2021a)旨在从多个非重叠监控摄像头中搜索或匹配同一个行人,已经广泛应用于安防和视频监控等场景。行人重识别将在一台摄像机视角中观察到的特定行人与另外一台摄像机拍摄到的诸多候

选行人进行比较，自动发现特定行人，从而完成行人在不同摄像头之间的再次识别。但是，在真实的复杂场景中，同一个行人在不同摄像头中的成像存在外貌、尺寸等差异问题，并且由于视角、拍摄距离、身体姿态和遮挡条件的变化，导致拍摄到的行人图像可能会存在低分辨率的情况。相比于高分辨率图像，低分辨率的行人图像包含了更少的身份与细节信息，如果直接对低分辨率行人图像进行相互匹配会造成显著的性能损失（贲晛烨 等，2012；史维东 等，2020；沈庆 等，2020；郑鑫 等，2020）。

现有的许多行人重识别方法通常侧重于解决跨分辨率行人匹配问题，即同一个行人不同分辨率图像之间的相互匹配。近年来涌现了许多跨分辨率行人重识别方法（Adil 等，2020；Jing 等，2017），大致可以分为3类：1）利用超分辨率技术（Wang 等，2018）；2）采用对抗学习方法（Li 等，2019）；3）学习分辨率不变特征表示（Chen 等，2019b）。第1类方法是联合训练超分模型和行人重识别模型，然而这种训练方式会导致梯度不能有效地传播，模型难以收敛。为简便起见，许多现有的基于超分辨率的方法在训练过程中直接对高分辨率图像下采样得到对应的低分辨率图像。这种数据采样的方式并不能使超分模型有效地恢复低分辨率图像的细节特征。在真实场景中，低分辨率图像的产生受光照、噪音背景环境等复杂因素影响。因此，通过下采样方式得到的低分辨率图像无法准确反映真实场景中获取的低分辨率图像情况。第2类方法通常采用对抗学习的思想实现分辨率自适应表示，然而通过这种方法并不能有效地解决不同分辨率行人的相互匹配。第3类方法通常学习低分辨率和高分辨图像共有的特征表示，但是由于低分辨率图像缺失细节信息，从而无法获取细粒度判别特征。

目前，一些行人重识别方法只关注高低分辨率行人图像不匹配的问题。这些方法只考虑了 Probe 集合里的图像是低分辨率的，往往忽略了训练集合和 Gallery 集合里也存在低分辨率图像。低分辨率图像所包含的行人细节信息较少，不利于同一身份的行人相互匹配。许多行人重识别算法尝试采用超分模型恢复图像细节，但是需要足够的高低分辨率图像对训练超分模型。目前，最常见的方式是直接使用原始的行人数据集作为高分辨率图像集，然后下采样原始行人图像得到低分辨率图像集。虽然通过这种采样方式可以用来训练超分模型，但是并不能确保超分模型能有效地学习真实场景下高分辨图像和低分辨率图像之间的映射关系。

以上的研究工作大都采用模拟的低分辨率数据集解决不同分辨率行人之间的相互匹配，本文聚焦于一个更富有挑战性的行人重识别问题，即实际场景中的低分辨行人之间的相互匹配。为研究该问题，本文首先构建了一个基于枪球摄像机的行人数据集。该数据集由部署在3个交叉路口的枪球系统收集得到，如图1所示。每个交叉路口都放置了两台摄像机，其中的枪机摄像头具有固定方向和焦距，拍摄获取低分辨率图像。另一个球机摄像头可以根据目标行人位置，调整焦距和视线方向，从而获得高分辨率图像。枪机摄像头获得的低分辨率图像和球机摄像头拍摄的高分辨率图像如图2所示。

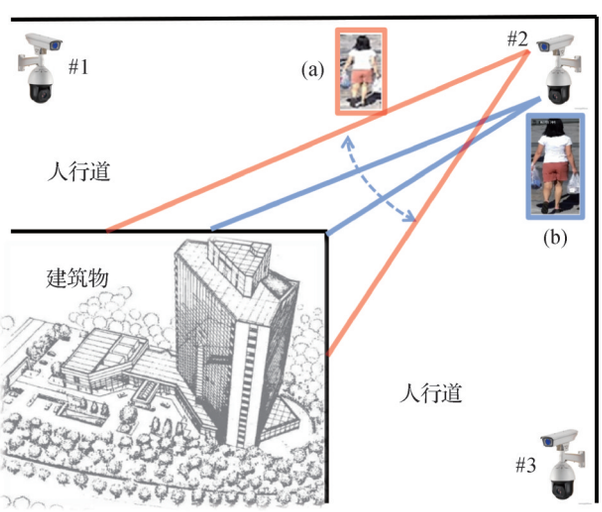

图1　枪球摄像机数据采集系统

Fig. 1　The gun-ball camera system

基于枪球摄像机的行人数据集共包含200个有身份标签行人（同一行人在不同位置被拍摄和识别）和320个无身份标签行人（只在某个摄像头下拍摄的行人），其中每个行人都包含高分辨率和低分辨率图像。有身份标签的行人指的是被至少2台不同的枪球摄像机从不同地方捕获到，无身份标签行人指的是只被1台枪球摄像机拍摄到，无法进行跨相机搜索与匹配，但是每个无身份类别的行人也包含低分辨率和高分辨率图像，从而可以有效地训练超分模型。为了研究真实场景下的低分辨率行人匹配问题，本文提出了一个通用的低分辨率行人重识别基准模型。

(a) 低分辨率图像

(b) 高分辨率图像

图2　基于枪球摄像机的行人数据集示例图像

Fig. 2　Sample images of the gun-ball camera-based person dataset((a)low resolution images; (b)high resolution images)

本文的主要工作包括两个方面:1)从真实场景中构建了一个小型的基于枪球摄像机的行人重识别数据集,其中每个行人具有成对的高分辨率和低分辨率图像,同时被每台摄像机捕获多幅图像。该数据集包含从6台摄像机收集的大约200个有身份标签行人和320个无身份标签行人。这个基于枪球摄像机的行人数据集为未来的低分辨率行人重识别的研究提供了更接近于实际情况的基准。2)基于构建的数据集,设计了一个低分辨率行人重识别基准模型,该基准模型包括超分模块、特征学习模块和特征判别器模块。其中,超分模块由基于Transformer的生成器网络、梯度判别器和图像判别器组成,实现低分辨率图像超分。特征学习模块采用预训练的残差网络,完成行人特征学习。特征判别器模块用于鉴别超分图像和高分辨率图像的行人特征。这个模型可以同时优化行人图像的分辨率和行人判别特征,从而解决实际场景中的低分辨行人识别问题。对比经典的行人重识别模型(Ye等,2022),这个基准模型在基于枪球摄像机的数据上分别将平均精度均值(mean average precision,mAP)和Rank-1指标提高了3.1%和6.1%。

本文是对前期工作(Sun 等,2022)的扩展和创新,系统性地介绍了低分辨率下行人重识别研究,相比于前期工作,主要有两个新贡献:1)更加全面深入地介绍了行人重识别数据集收集方法和使用方式。2)从3个方面对基准方法进行了大幅优化和创新。(1)设计了一个基于Transformer的生成器网络,并基于此提出了一种新的行人重识别模型用于低分辨率行人匹配,新方法在低分辨率数据集上取得了更高的识别精度;(2)扩充了消融实验,更加全面地验证所提模型的有效性;(3)设计了更优的网络训练方法,有效提高了行人图像的分辨率和特征判别效果。

1　相关工作

1.1　行人重识别

行人重识别是计算机视觉领域的重要任务,受到广泛关注并取得了迅猛发展。现有的许多行人重识别方法通过提取更鲁棒的判别性特征,解决行人匹配中存在的视角姿态变化、背景干扰和部分遮挡等各种挑战。这些外在环境的变化和影响,使行人重识别成为一项具有挑战性的任务,越来越多的行人重识别算法聚焦这些困难并提高行人匹配的精度。Liu 等人(2018)设计了一种姿态可转移的行人重识别框架,通过利用姿态转移的样本增强生成特定姿态的图像,从而解决行人匹配中姿态变化的问题。另外一些方法通过引入注意力机制解决行人匹配中的背景干扰问题。Kalaye等人(2018)采用语义解析分割出前景和背景信息,从而减少背景的干扰。针对部分遮挡的问题,Li等人(2021a)提出了一种基于Transformer的编码器和解码器架构,依赖一个完整目标的标签识别遮挡的行人。然而,上述方法大都忽略了真实场景中拍摄的行人会存在低分辨率的问题,不能有效地解决低分辨率行人匹配问题。因此,一些跨分辨率行人重识别方法应运而生。Li等人(2019)提出了一个跨分辨率对抗双网络(cross-resolution adversarial dual network,CAD-Net),利用对抗网络获得分辨率自适应表示并学习恢复低分辨率行人图像的细节。Cheng等人(2020)通过引入一种模型训练正则化方法(inter-task association critic,INTACT),实现超分和行人重识别模型的有效联合训练。Zhang等人(2021b)设计了一个伪孪生网络框架,以减少低分辨和高分辨率图像之间特征分布差异。Munir等人(2021)为了解决跨分辨率图像匹配问题,引入了基于分辨率的特征提取方法学习分辨率不变特征。Wu等人(2022)提出了一个由超分模块和双流特征融合模块构成的超分辨率双流特征融合子网络,其中超分模块恢复图像分辨率,双流特征融合模块减少图像细节的丢失,从而联合优化行人图像的特征细节和提取。Zheng等人(2022)设计

了一种新的联合双边分辨率身份建模的方法，同时进行特定高分辨率身份特征学习、低分辨率身份特征学习和行人重识别优化。然而这些方法只关注跨分辨率行人匹配问题，即低分辨率Query图像和高分辨率Gallery图像之间的相互匹配，而对于真实场景中低分辨率下的Query和Gallery图像相互匹配的研究甚少。

1.2 超分辨率

基于深度学习的方法在图像超分（super-resolution，SR）领域取得了极大成功。图像SR旨在从低分辨率图像中重建高分辨率图像，并学习低分辨率图像和高分辨率图像之间的映射关系。Dong等人（2014）首次使用卷积神经网络解决单幅图像的超分辨率问题。随着深度卷积神经网络（convolutional neural network，CNN）的发展，提出了越来越多的基于CNN的方法。Zhang等人（2018）利用残差和残差（residual in residual，RIR）结构建立了一个非常深的可训练网络。此外，考虑到通道之间的相互依赖关系，该工作还设计了通道注意力机制。Liu等人（2020）提出了一种渐进式多尺度残差网络（progressive multi-scale residual network，PMRN），通过对参数受限的特征进行连续挖掘，解决了单幅图像的超分辨率问题。鉴于SR模型有利于提升低分辨率图像质量，本文在提出的基准模型中采用了改进的SR模型，并在输入图像中融入梯度信息。

1.3 生成对抗网络

生成对抗网络（generative adversarial network，GAN）在许多无监督学习任务中取得了显著成功。随着不断发展，生成对抗网络已广泛应用于语义分割、目标检测和行人重识别领域。对抗网络由生成网络和判别网络组成，生成网络用于生成新样本，判别网络区分真假样本。Makhzani等人（2015）提出了一种通过聚集后验数据进行正则化的对抗式自编码器。Kim等人（2017）设计了一种自动学习并发现跨域关系的生成对抗网络，用于图像风格迁移。借鉴以上基于生成对抗网络算法的成功应用，本文提出的低分辨行人重识别基准模型采用对抗思想，以减少超分行人和高分辨率行人特征分布之间的差异。

2 本文工作

目前，行人重识别领域的许多研究都倾向于关注分辨率不匹配问题，忽略了实际场景下低分辨率行人匹配的问题。此外，现有的许多算法都是直接从公开的行人重识别数据集中通过下采样的方式构建低分辨率行人数据集，模拟真实场景中出现的低分辨率行人。与高分辨率图像相比，低分辨率图像不仅在尺寸上发生了变化，即图像宽高变小，同时在像素上发生了变化，即像素值变低。放大低分辨率图像，图像会变得模糊。而通过下采样的方式获取低分辨率图像虽然能保证尺寸变小，但不能确保像素值是否变低。在现实场景中，低分辨率图像受许多复杂因素影响，如失真、噪音和相机等因素。简单的下采样过程很难模拟出现实世界中的非线性变换。此外，真实场景中一般是通过低清摄像头拍摄获取低分辨率图像。本文为了验证通过下采样方式获取的低分辨率图像和真实场景中所获取的低分辨率图像不同，在消融实验中，设计了5组实验对比其差异。实验结果表明，下采样方式获取的低分辨率数据集训练的模型不能很好地处理真实场景中的低分辨率行人匹配问题。因此，通过这种方式构建的低分辨率行人数据集在效果上并不完全等同于真实场景中出现的低分辨率行人。为此，本文从真实场景中收集了一个低分辨率行人数据集，用以解决低分辨率行人匹配的问题。本文构建的基于枪球摄像机的行人重识别数据集包含了一组具有身份标签的高分辨率和低分辨率图像对。其中的高分辨率和低分辨率图像对用于训练超分模型，身份标签信息为行人重识别模型提供了可监督训练。本文探究了图像超分的潜力，一个有效的超分模型能够从降级的低分辨率图像中生成细节丰富的高分辨图像，缓解Probe图像和Gallery图像之间的匹配问题。为了使超分模型生成的高分辨率图像有益于行人识别，本文通过级联超分和行人重识别模型进行多任务联合学习。

2.1 基于枪球摄像机的行人重识别数据集

基于枪球摄像机的行人重识别数据集由部署在3个交叉路口的枪球摄像机收集，每个交叉路口有1台高清摄像机（球机摄像机）和1台低清摄像机（枪机摄像机）。该数据集包括6台摄像机拍摄的520个不同身份类别的行人。其中200个行人有身份标签，320个行人没有身份标签。每个行人至少被2台摄像机捕获到。同时每个行人不仅具有高分辨率的图像，还具有低分辨率的图像。这个数据集共包括10 424幅图像，每个行人平均有17幅训练图像。数

据集中的每幅行人图像均由真实场景中的摄像机自动拍摄获取。枪球摄像机拍摄的是多帧图像，图像里面不仅包含了目标行人还有其他建筑物、道路和车辆等非目标对象。因此，本文利用ImageMagick图像标注工具，将目标行人从整幅图像中裁剪出来。由于每幅图像中目标行人的大小不一，裁剪出来的图像尺寸大小也不一样。为了训练方便，在训练过程中将所有高分辨率图像的尺寸调整为192×96像素，低分辨率图像尺寸调整为64×32像素。

基于枪球摄像机的行人重识别数据集中每个身份类别的行人在每个摄像头下都具有多幅图像，这将有利于跨摄像头搜索并匹配同身份类别的行人。本文构建的数据集与现有主流数据集存在以下不同。1）现有的一些数据集（如Market501、CUHK03（Chinese University of Hong Kong）和CAVIAR）主要通过捕获大学校园或者购物商场行人图像，而基于枪球摄像机的行人重识别数据集从交叉路口获取各种路人图像，形成了更丰富、更多样化的行人数据集。2）因为基于枪球摄像机的行人数据集是从视频流中捕获并裁剪得到，所以每幅行人图像具有时序信息，可以捕捉到随时间变化的行人动态。这种具有时序特征的行人图像还适用于研究视频行人重识别。3）本文构建的数据集还包括一些身份未标明的行人，可以用于研究半监督或者无监督领域的行人重识别算法，同时也可以模拟现实世界中身份识别系统的工作模式。即给定一幅未知身份的人员图像，身份识别系统将会在监控画面或者数据库中自动检测到该同类人员。本文构建的数据集与现有主流行人数据集Market1501、CUHK03、CAVIAR和VIPeR的对比结果如表1所示。

表1　基于枪球摄像机的行人重识别数据集与其他数据集对比

Table 1　Comparison between the gun-ball camera-based person re-identification and other datasets

数据集	摄像头/个	身份类别/个	图像/幅	高低分辨率图像
Market501	6	1 501	32 217	不包含
CUHK03	10	1 306	13 164	不包含
CAVIAR	2	72	1 220	包含
VIPeR	2	632	1 264	不包含
本文	6	200	10 424	包含

从表1可以看出，本文构建的数据集具有以下优点：1）这是第1个为每个行人同时提供高分辨率图像和低分辨图像的行人重识别数据集。CAVIAR数据集虽然也包含低分辨和高分辨率图像，但是这两个图像是独立拍摄获取的，彼此之间没有对应关系，所以无法直接用于训练超分模型，而本文构建的数据集中每个行人具有对应的高分辨率和低分辨率图像，因此可以通过超分模型学习低分辨率图像和高分辨率图像之间的映射关系。2）整个数据集中的每个行人是由两台不同的摄像机同时拍摄的。因此，每个行人的高分辨率和低分辨率图像之间可能存在像素误对齐的问题。在某种程度上，该数据集也可以应用于研究图像超分领域的像素误对齐问题。因此，本文构建的数据集对其他领域的研究具有重要的参考价值。

2.2　低分辨率行人重识别模型总体架构

本文提出的低分辨率行人重识别基准模型联合学习超分任务和行人重识别任务，整体网络框架如图3所示，包括生成器网络$\boldsymbol{G}$、梯度判别器$\boldsymbol{D}_g$、图像判别器$\boldsymbol{D}_s$、行人特征判别器$\boldsymbol{D}_f$和行人特征提取器$\boldsymbol{F}$。对于输入的低分辨率行人图像，本文基准模型训练目标有：1）将低分辨率图像恢复为高分辨率图像；2）识别并匹配不同摄像机下的同身份行人。

输入的低分辨图像$\boldsymbol{x}_{\mathrm{LR}}$首先经过生成器$\boldsymbol{G}$得到超分图像$\boldsymbol{x}_{\mathrm{SR}}$，然后图像判别器$\boldsymbol{D}_s$区分高分辨率图像$\boldsymbol{x}_{\mathrm{HR}}$和超分图像$\boldsymbol{x}_{\mathrm{SR}}$，同时梯度判别器$\boldsymbol{D}_g$负责鉴别超分图像和高分辨图像梯度图的真假，最后利用特征提取器$\boldsymbol{F}$提取超分图像$\boldsymbol{x}_{\mathrm{SR}}$和高分辨图像$\boldsymbol{x}_{\mathrm{HR}}$的判别特征，并将提取的判别特征输入到特征判别器$\boldsymbol{D}_f$辨别是否来自同一特征分布。

2.3　图像超分辨率模型

本文采用的超分模型由生成器网络$\boldsymbol{G}$、梯度判别器$\boldsymbol{D}_g$和图像判别器$\boldsymbol{D}_s$组成。

2.3.1　生成器网络

为了从低分辨率行人图像中获取高质量的行人图像，本文采用基于SwinIR（swin image restoration）（Liang等，2021b）的生成器网络架构。但是SwinIR只能有效地解决超分领域中像素对齐图像的复原问题，而在像素误对齐图像上使用失效，因此本文对SwinIR模型中的网络结构进行了两方面的改进：1）网络输入同时包含了低分辨率图像的梯度信息。输入的梯度信息可以使网络学习到图像的结构特征和

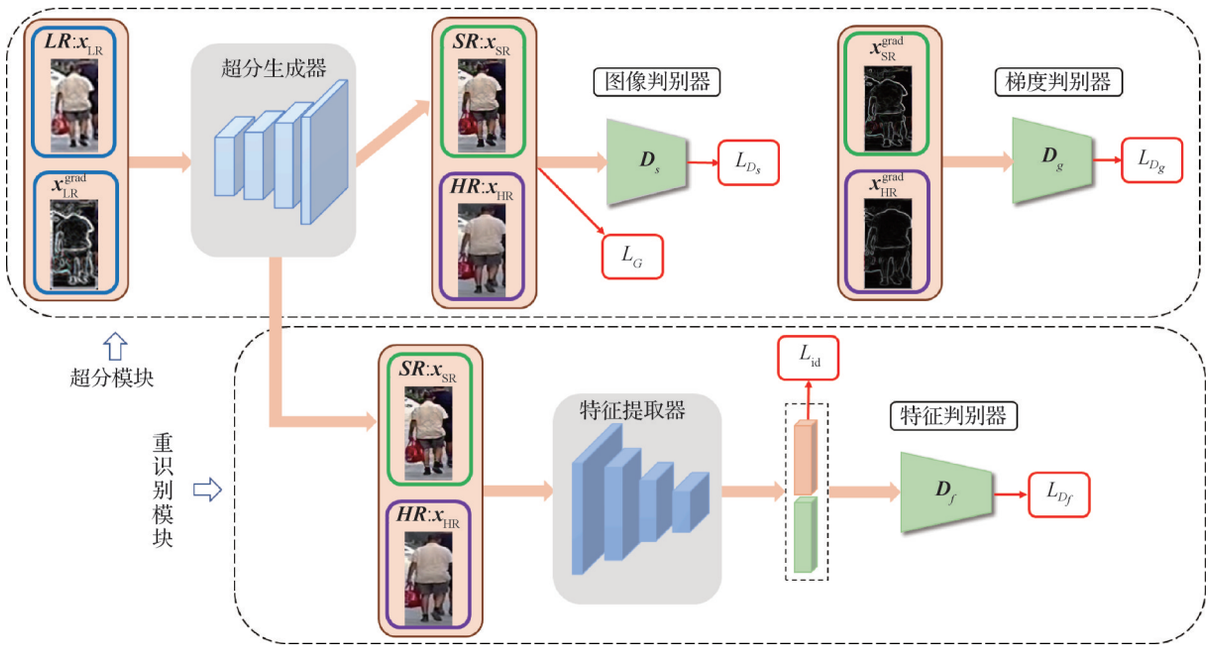

图 3　网络总体结构图

Fig. 3　Overall network structure of the baseline

高频信息，同时结合梯度判别器的使用，能有效地解决像素误对齐的问题。首先低分辨率图像 $\boldsymbol{x}_{\mathrm{LR}}$ 通过梯度函数 $M(\cdot)$（$M(\cdot)$采用固定的3×3大小卷积核对图像的3个颜色通道分别做水平和垂直方向上的卷积操作，再将得到的卷积结果在颜色通道维度上进行拼接得到最终梯度）得到梯度图 $\boldsymbol{x}_{\mathrm{LR}}^{\mathrm{grad}}$；然后 $\boldsymbol{x}_{\mathrm{LR}}$ 和 $\boldsymbol{x}_{\mathrm{LR}}^{\mathrm{grad}}$ 分别输入到一个卷积核大小为3 × 3的卷积层提取浅层特征；最后将两者的浅层特征在通道维度上进行连接操作，并将连接后的特征图作为后续模块的输入。2）为了减少上采样操作所带来的计算量，网络结构的上采样层采用最近邻插值算法增大图像分辨率。生成器网络结构如图4所示，输入图像 $\boldsymbol{x}_{\mathrm{LR}}$ 和 $\boldsymbol{x}_{\mathrm{LR}}^{\mathrm{grad}}$ 首先经过一个3 × 3卷积层获取到浅层特征，并将浅层特征进行通道维度拼接。然后，将拼接特征输入到6个RSTB（residual swin transformer block）（Liang等，2021b）模块和1个3 × 3卷积层提取深层特征。最后，将拼接特征和深层特征相加得到融合特征，并将融合特征输入到上采样层得到最终的高质量图像。

本文采用像素级损失和感知损失。像素级损失最小化超分图像和高分辨率图像之间的像素级误差，同时最小化超分图像梯度图和高分辨率图像梯度图之间的像素级误差。感知损失最小化超分图像和高分辨率图像之间的特征损失。感知特征由预先训练的视觉几何群网络（Visual Geometry Group network-16，VGG-16）提取。目标函数为

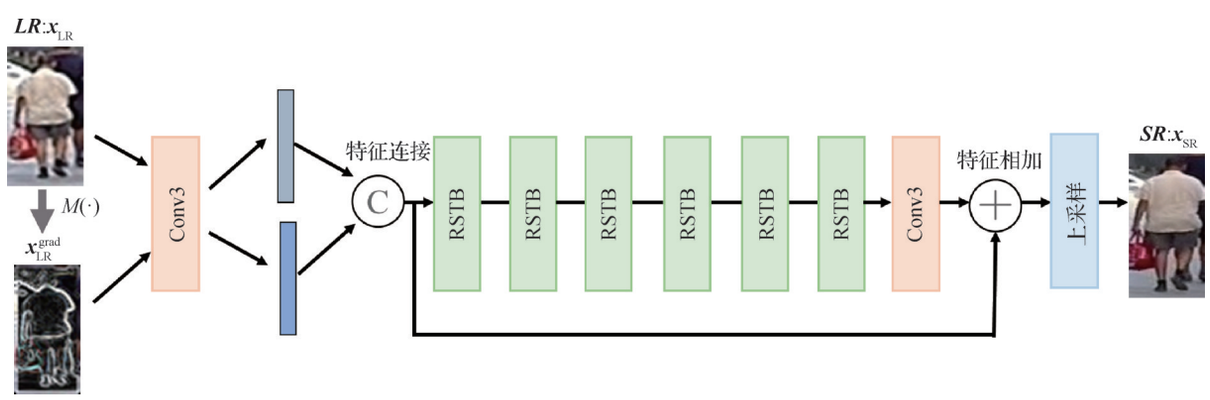

图4　生成器网络

Fig. 4　The generator network

$$L_G = \lambda_{\text{pix}}^{\text{SR}} \left\| G(\boldsymbol{x}_{\text{LR}}) - \boldsymbol{x}_{\text{HR}} \right\|_{L1} + \lambda_{\text{pix}}^{\text{grad}} \left\| \boldsymbol{x}_{\text{SR}}^{\text{grad}} - \boldsymbol{x}_{\text{HR}}^{\text{grad}} \right\|_{L1} + \lambda_{\text{per}}^{\text{SR}} \left\| \phi(G(\boldsymbol{x}_{\text{LR}})) - \phi(\boldsymbol{x}_{\text{HR}}) \right\|_{L1} \tag{1}$$

式中，$\lambda_{\text{pix}}^{\text{SR}}$，$\lambda_{\text{pix}}^{\text{grad}}$，$\lambda_{\text{per}}^{\text{SR}}$是权重参数，$\phi(\cdot)$是感知特征提取函数。在图像、梯度和感知特征的监督下，生成器不仅可以学习到细节信息，还可以避免结构失真。

2.3.2　图像判别器

许多基于生成对抗网络的方法都成功地解决了图像超分领域的问题。本文利用$\boldsymbol{D}_s$区分生成的图像和高分辨率图像，使得两图像之间更相似。优化目标函数为

$$L_{D_s} = -E_{x_{\text{SR}}}[\log(1 - \boldsymbol{D}_s(\boldsymbol{x}_{\text{SR}}))] - E_{x_{\text{HR}}}[\log(\boldsymbol{D}_s(\boldsymbol{x}_{\text{HR}}))] \tag{2}$$

式中，$E_{x_{\text{SR}}}$表示$\log(1 - \boldsymbol{D}_s(\boldsymbol{x}_{\text{SR}}))$的数学期望，$E_{x_{\text{HR}}}$表示$\log(\boldsymbol{D}_s(\boldsymbol{x}_{\text{HR}}))$的数学期望。

2.3.3　梯度判别器

本文构建的数据集是通过人工裁剪视频帧中的行人得到的，因此高分辨率行人和低分辨率行人之间可能存在像素误对齐的问题。为此，本文利用梯度判别器$\boldsymbol{D}_g$来解决像素误对齐问题。梯度判别器$\boldsymbol{D}_g$鉴别高分辨率行人梯度图和超分行人梯度图的真假，可以通过对抗学习监督超分图像的生成，保留完整的细节和结构信息。为了优化$\boldsymbol{D}_g$，最小化目标函数L_{D_g}，具体为

$$L_{D_g} = -E_{x_{\text{SR}}}[\log(1 - \boldsymbol{D}_g(M(\boldsymbol{x}_{\text{SR}})))] - E_{x_{\text{HR}}}[\log(\boldsymbol{D}_g(M(\boldsymbol{x}_{\text{HR}})))] \tag{3}$$

2.4　行人特征提取器

行人重识别根据图像特征判断是否来自于同一个行人。其中Probe集合是待检索的图像集，而Gallery集合是用于匹配的图像集。当要对Probe集合的某个行人图像进行检索，首先需要特征提取器提取待检索行人图像特征，然后再提取Gallery集合中的所有行人图像特征并计算与待检索行人图像特征的距离，并将特征距离按照升序排序，最后特征距离最小的图像即为匹配成功的图像。本文采用在ImageNet数据集上预训练的残差网络ResNet50（residual network 50）（He等，2016）作为行人特征提取器。

许多行人重识别模型利用交叉熵损失函数训练行人特征提取器，目标函数为

$$L_{\text{cross}} = -\log(p_y) \tag{4}$$

式中，y是输入图像的身份标签，p_y是在y类上的预测概率。

为了更好地匹配同身份行人，本文引入了三元组损失函数最小化类间距离，目标函数为

$$L_{\text{triplet}} = \max(d_p - d_n + \alpha, 0) \tag{5}$$

式中，d_p和d_n分别表示正样本和负样本的特征距离，α是一个大于0的常数。本文根据经验知识将α的值设置为0.3。该行人特征提取器的优化目标为

$$L_{\text{id}} = L_{\text{cross}} + L_{\text{triplet}} \tag{6}$$

2.5　行人特征判别器

在图像空间上，利用梯度和图像判别器改善生成的超分图像质量并没有显著地提升行人识别的性能。为此，本文引入了行人特征判别器$\boldsymbol{D}_f$在特征空间上区分超分图像特征和高分辨率图像特征，使这两个特征的分布相似。相对而言，在特征空间上超分相似约束能极大改善行人匹配的性能。研究表明，如果使用二分类损失函数优化行人特征判别器$\boldsymbol{D}_f$和特征生成器（由生成器$\boldsymbol{G}$和特征提取器$\boldsymbol{F}$组成），那么由于特征生成器网络过深可能会造成训练不稳定，因此本文行人特征判别器的最后一层移除了sigmoid，采用基于Wasserstein GAN（Arjovsky等，2017）的判别器损失函数，具体为

$$L_{D_f} = E_{f_{\text{HR}}}[\boldsymbol{D}_f(\boldsymbol{f}_{\text{HR}})] - E_{f_{\text{SR}}}[\boldsymbol{D}_f(\boldsymbol{f}_{\text{SR}})] \tag{7}$$

式中，$\boldsymbol{f}_{\text{HR}}$和$\boldsymbol{f}_{\text{SR}}$分别表示高分辨率和超分行人图像的特征。

2.6　网络训练

在模型训练过程中，L_G损失项和L_{id}损失项可以直接嵌入到GAN的优化中，并且整个模型都保持端到端的可训练。具体步骤如下：

输入：具有身份标签的训练集$\boldsymbol{D} = \{\boldsymbol{x}_{\text{LR}}, \boldsymbol{x}_{\text{HR}}\}$。

输出：网络$\boldsymbol{G}$，$\boldsymbol{F}$，$\boldsymbol{D}_s$，$\boldsymbol{D}_g$，$\boldsymbol{D}_f$。

1）从训练集$\boldsymbol{D}$中随机选取一批数据输入到生成器$\boldsymbol{G}$和行人特征提取器$\boldsymbol{F}$，最后得到输出$\boldsymbol{x}_{\text{SR}}$、$\boldsymbol{f}_{\text{SR}}$和$\boldsymbol{f}_{\text{HR}}$，并利用$L_G + L_{\text{id}}$更新网络$\boldsymbol{G}$和$\boldsymbol{F}$的参数；

2）行人特征判别器$\boldsymbol{D}_f$鉴别$\boldsymbol{f}_{\text{SR}}$和$\boldsymbol{f}_{\text{HR}}$特征的真假，利用$L_{D_f}$更新网络$\boldsymbol{D}_f$的参数；

3）图像判别器鉴别$\boldsymbol{x}_{\text{SR}}$和$\boldsymbol{x}_{\text{HR}}$的真假，利用$L_{D_s}$更新网络$\boldsymbol{D}_s$的参数；

4）梯度判别器鉴别$\boldsymbol{x}_{\text{SR}}$和$\boldsymbol{x}_{\text{HR}}$的梯度图的真假，利用$L_{D_g}$更新网络$\boldsymbol{D}_g$的参数；

5)重复步骤1)—4),直至网络收敛。

3 实 验

3.1 实验设置及评价指标

3.1.1 实验设置

本文的所有实验都基于深度学习框架Pytorch,在显卡为GeForce RTX 2080的Linux操作系统的单机电脑上进行训练。本文在构建的行人重识别数据集上的实验设置是将数据集按照7:3的比例划分为训练集和测试集。图像判别器和梯度判别器采用VGG-16网络结构,特征判别器由4层线性变换层和3层非线性激活层组成,除了最后一层线性变换层,其他线性变换层后面都有非线性激活层。特征判别器采用基于Wasserstein GAN(Arjovsky 等,2017)的损失函数更新网络参数,因此参照Arjovsky等人(2017)的设计,在每次更新特征判别器的参数之前,将参数绝对值限定到[-0.01, 0.01]范围。网络训练的总轮次达到200时,网络基本达到收敛且性能不再上升。采用RMSProp优化器更新生成器和特征提取器网络参数,根据行人重识别训练的经验,本文设置初始化学习率为0.000 35,权重衰减参数为0.000 5。考虑到显存原因,batch size设置为8。所有判别器网络采用初始化学习率为0.000 1的Adam优化器更新网络参数,$\beta_1 = 0.9$,$\beta_2 = 0.999$。根据超分辨率图像训练的经验,在本文实验中设置式(1)中的权重系数,$\lambda_{\text{pix}}^{\text{SR}} = \lambda_{\text{pix}}^{\text{grad}} = 0.01$,$\lambda_{\text{per}}^{\text{SR}} = 1$。

3.1.2 评价指标

实验使用累积匹配特征(cumulative matching characteristic,CMC)和平均精度均值(mAP)作为行人重识别的性能评估指标。CMC曲线表示被查询的行人出现在不同尺寸行人的候选名单中的概率,用来量化性能,CMC@K表示排名在前K位正确匹配的百分比,即本文实验中的Rank-K。

3.2 实验结果与分析

3.2.1 对比实验

为了验证本文提出的基准模型的性能,与采用不同训练的3种方法进行对比,并分别进行定性和定量实验分析,结果如图5和表2所示。图5展示了在基于枪球摄像机的低分辨率行人数据集上进行图像超分的不同结果。由图5可知,采用了超分模块的方法能够有效地提高图像的分辨率。同时也容易观察到,结合了行人重识别模块的联合模型在超分的同时保留了更多的细节信息,因而更加适合行人重识别任务。

对比实验中所采用的超分模型为SwinIR(Liang等,2021b),行人重识别模型为AGW(Ye 等,2022)。第1种方法是行人重识别模型直接训练基于枪球摄像机的低分辨率行人数据集,然后在低分辨率行人数据集上进行测试。从实验结果可以看出,低分辨率的行人识别性能差。第2种方法采用联合训练的方式同时优化级联的超分模型和行人重识别模型。第3种方法是单独训练超分模型和行人重识别模型。首先使用数据集中的高分辨率图像和低分辨率图像训练超分模型,然后低分辨率图像输入到训练好的超分模型中得到超分图像,最后行人重识别模型对超分图像进行训练和测试。本文提出的基准模型是在级联的超分模型和行人重识别模型中又嵌入了特征判别器模块,使得高分辨率行人特征和超分辨率行人特征空间分布更相似。为了验证所提出的基准模型包含的3个网络模块的有效性,以SwinIR作为超分模型改善低分辨率图像的分辨率,同时以AGW为行人重识别模型识别行人特征,设计了3种模型Sole ID、SR+ID和Sole SR与基准模型进行对比。其中,Sole ID直接采用低分辨率图像训练AGW模型;SR+ID使用高低分辨图像对联合训练SwinIR和AGW模型;Sole SR使用高低分辨图像对单独训练SwinIR模型,然后低分辨率图像输入到训练好的SwinIR模型中得到超分图像,最后AGW模型对超分图像进行训练和测试。消融实验结果如表2所示。可以看出,本文模型显著优于Sole ID、SR + ID和Sole SR这3种模型的识别精度。表2的实验结果表明,当同时优化超分模型和行人重识别模型,性能有所提升。但将超分模型和行人重识别模型分开训练,识别精度反而下降。这是因为超分模型的生成器并没有学习到有益于行人识别的细节特征。联合优化的训练方式虽然提高了识别性能,但是仍然显著低于本文方法的识别精度。

实验将所提模型与4种代表性行人重识别方法进行比较,包括BagTricks(Luo 等,2019)、CDNet(combined depth network)(Li 等,2021b)、ABD-Net(attentive but diverse network)(Chen 等,2019a)和NFormer(neighbor transformer network)(Wang 等,2022)。公平起见,所有方法都在本文构建数据集中

图 5　枪球行人数据集的超分图像实例

Fig 5　Examples of super-resolved person images from the gun-ball person dataset

表 2　在枪球行人数据集的低分辨率行人上的性能对比

Table 2　Performance comparison on low-resolution pedestrians on the gun-ball person dataset

/%

模型	mAP	Rank-1	Rank-5	Rank-10
Sole ID	74.8	77.0	84.5	89.9
SR+ID	75.8	80.4	**88.5**	**91.2**
Sole SR	73.4	77.7	86.5	89.2
本文	**77.9**	**83.1**	85.8	90.5

注：加粗字体表示各列最优结果。

表 3　不同方法在枪球行人数据集上的对比结果

Table 3　Comparison results of different methods on the gun-ball pedestrian dataset

/%

模型	mAP	Rank-1	Rank-5	Rank-10
BagTricks	74.8	77.0	84.5	89.9
CDNet	70.6	72.3	79.1	86.5
NFormer	73.5	76.4	**87.2**	**91.2**
ABD-Net	77.2	79.1	86.5	89.2
本文	**77.9**	**83.1**	85.8	90.5

注：加粗字体表示各列最优结果。

的低分辨率图像上进行训练，所有图像输入尺寸为 64×32 像素。表 3 给出了对比结果。与主流的行人重识别方法相比，在 mAP 和 Rank-1 评价指标上，本文模型超过现有主流方法的性能，证明了所提模型能有效解决真实场景中低分辨率下的行人匹配。与 BagTricks、CDNet、ABD-Net 和 NFormer 等 4 种方法相比，本文所提出的基准模型包含了图像超分模块，因此能够将低分辨率行人图像恢复为高分辨率图像，并通过级联超分和行人重识别模型进行多任务联合学习，从而有效提升了低分辨率行人匹配的性能。

3.2.2　消融实验

为了验证真实场景的低分辨率图像无法通过下采样模拟获取，设计了5组实验来对比真实场景中低分辨图像与下采样获得的低分辨率图像之间的差异。5组实验采用相同的训练模型和方法，但是低分辨率训练数据集的获取来自不同方式。第1组实验通过双线性插值的方式下采样高分辨率图像获取低分辨率图像；第2组实验采用双三次插值算法下采样高分辨率图像获得低分辨率图像；第3组实验采用最近邻插值算法下采样高分辨率图像获得低分辨率图像；第4组实验采用区域插值算法下采样高分辨率图像获得低分辨率图像；第5组实验从真实场景中收集低分辨率图像。实验结果都是在真实场景上的低分辨率行人图像上进行测试得到，如图6所示。

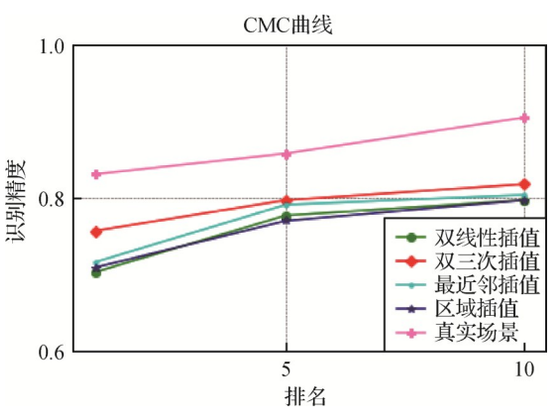

图6　低分辨率图像获取方式不同的消融研究

Fig. 6　Ablation of low-resolution images in different ways

从实验结果可以看出，通过下采样方式获取的低分辨率数据集训练的模型不能很好地处理真实场景中的低分辨率行人匹配。所以，现有的许多解决低分辨率行人匹配的算法可能无法有效地解决真实场景中的低分辨率行人识别问题。这同时也说明采用简单的下采样方法很难模拟真实场景中的非线性变换。

此外，通过选定超分模型各模块和行人特征判别器，探究本文提出的基准模型的训练方法在其他行人特征提取器上的有效性。包括：1）在ImageNet数据集上预训练的残差网络结构ResNet50；2）Szegedy等人（2016）通过修改Inception模块得到的InceptionV4网络结构；3）Zhou等人（2019）设计的一种实现全尺度特征学习的深度行人重识别的全尺度网络OSNet（omni-scale network）；4）在InceptionV3网络结构基础上采用深度可分离卷积替换Inception模块的标准卷积并引入残差结构的Xception（Chollet，2017）。实验结果如表4所示。公平起见，表4中所有采用Joint方式的网络保持相同的训练数据、学习率和优化器。同理，所有采用Raw方式（Raw方式直接在本文提出的低分辨率行人数据集上进行训练和测试）的网络也一样。从实验结果可以看出，本文设计的基准模型网络对于不同的行人特征提取器都是适用的，进一步验证了这个基准模型不仅在残差网络上有效，在其他网络上识别性能也同样得到了极大提升。

表4　不同行人特征提取器的实验结果对比

Table 4　Comparison of different pedestrian feature extractors

/%

特征提取网络	训练方法	mAP	Rank-1	Rank-5	Rank-10
ResNet50	Joint	**77.9**	**83.1**	**85.8**	90.5
	Raw	74.8	77.0	84.5	89.9
InceptionV4	Joint	65.6	73.0	81.8	86.5
	Raw	58.9	64.9	80.4	87.8
OSNet	Joint	57.6	58.8	70.3	77.0
	Raw	45.2	48.0	60.8	65.5
Xception	Joint	69.4	75.7	**85.8**	**92.6**
	Raw	59.5	66.9	77.7	83.8

注：加粗字体表示各列最优结果。

目前，超分模型的生成器都是基于CNN网络或者Transformer网络。因此本文探究了这两种网络类型的生成器模块对识别性能的影响，并在基于枪球摄像机的行人重识别数据集上进行了实验对比，结果如表5所示。可以看出，不论生成器基于何种类型的网络，识别性能都有所提升。但是基于Transformer的生成器模型，性能提升的幅度更大。

4　结　论

针对实际场景中的低分辨率行人重识别问题，本文构建了一个基于枪球摄像机的行人重识别数据

表 5　不同生成器类型的实验结果对比

Table 5　Comparison of experimental results of different generator types

/%

生成器类型	mAP	Rank-1	Rank-5	Rank-10
CNN	77.6	80.4	**87.8**	**92.6**
Transformer	**77.9**	**83.1**	85.8	90.5

注：加粗字体表示各列最优结果。

集，共包含 200 个有身份标签的行人（同一行人在不同位置被拍摄和识别）和 320 个无身份标签的行人（只在某个摄像头下拍摄的行人），其中每个行人都包含高分辨率和低分辨率图像。同时，为低分辨率下的行人匹配设计了一个基准行人重识别模型，由生成器、图像判别器、梯度判别器、行人特征提取器和行人特征判别器构成。该基准模型可以同时优化行人图像的分辨率和行人判别特征，从而解决实际场景中的低分辨行人识别问题。实验结果表明，本文提出的基准模型对比于经典的行人重识别模型，在 mAP 和 Rank-1 指标上分别提高了 3.1% 和 6.1%。因此，相对其他方法，本文方法能更好地解决实际场景中的低分辨率行人识别问题，并在一定程度上解决了由于像素误对齐导致生成的超分图像质量不高的问题。本文所提出的数据集和基准模型不仅可以应用于行人重识别领域，还可以应用于图像超分领域。

目前，本文实验的行人重识别训练部分都是针对所构建数据集中的有身份标签的行人。后续会考虑在识别部分加入无身份标签的行人，从而利用半监督算法解决低分辨率行人匹配的问题。因此，未来拟研究弱监督场景下的行人重识别算法。

参考文献（References）

Adil M, Mamoon S, Zakir A, Manzoor M A and Lian Z C. 2020. Multi scale-adaptive super-resolution person re-identification using GAN. IEEE Access, 8: 177351-177362 [DOI: 10.1109/ACCESS.2020.3023594]

Arjovsky M, Chintala S and Bottou L. 2017. Wasserstein GAN [EB/OL]. [2022-11-04]. https: https://arxiv.org/pdf/1701.07875.pdf

Ben X Y, Xu S and Wang K J. 2012. Review on pedestrian gait feature expression and recognition. Pattern Recognition and Artificial Intelligence, 25(1): 71-81 (贲晛烨, 徐森, 王科俊. 2012. 行人步态的特征表达及识别综述. 模式识别与人工智能, 25(1): 71-81) [DOI: 10.3969/j.issn.1003-6059.2012.01.010]

Chen T L, Ding S J, Xie J Y, Yuan Y, Chen W Y, Yang Y, Ren Z and Wang Z Y. 2019a. ABD-Net: attentive but diverse person re-identification//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 8350-8360 [DOI: 10.1109/ICCV.2019.00844]

Chen Y C, Li Y J, Du X F and Wang Y C F. 2019b. Learning resolution-invariant deep representations for person re-identification. Proceedings of 2019 AAAI Conference on Artificial Intelligence, 33(1): 8215-8222 [DOI: 10.1609/AAAI.v33i01.33018215]

Cheng Z Y, Dong Q, Gong S H and Zhu X T. 2020. Inter-task association critic for cross-resolution person re-identification//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 2602-2612 [DOI: 10.1109/CVPR42600.2020.00268]

Chollet F. 2017. Xception: deep learning with depthwise separable convolutions//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 1800-1807 [DOI: 10.1109/CVPR.2017.195]

Dong C, Loy C C, He K M and Tang X O. 2014. Learning a deep convolutional network for image super-resolution//Proceedings of the 13th European Conference on Computer Vision (ECCV). Zurich, Switzerland: Springer: 184-199 [DOI: 10.1007/978-3-319-10593-2_13]

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Jing X Y, Zhu X K, Wu F, Hu R M, You X G, Wang Y H, Feng H and Yang J Y. 2017. Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning. IEEE Transactions on Image Processing, 26(3): 1363-1378 [DOI: 10.1109/TIP.2017.2651364]

Kalayeh M M, Basaran E, Gökmen M, Kamasak M E and Shah M. 2018. Human semantic parsing for person re-identification//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE: 1062-1071 [DOI: 10.1109/CVPR.2018.00117]

Kim T, Cha M, Kim H, Lee J K and Kim J. 2017. Learning to discover cross-domain relations with generative adversarial networks//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: IMLS: 1857-1865 [DOI: 10.48550/arXiv.1703.05192]

Lan L, Wang X C, Hua G, Huang T S and Tao D C. 2020. Semi-online multi-people tracking by re-identification. International Journal of Computer Vision, 128(7): 1937-1955 [DOI: 10.1007/s11263-020-01314-1]

Li H J, Wu G J and Zheng W S. 2021b. Combined depth space-based architecture search for person re-identification//Proceedings of

2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 6725-6734 [DOI: 10.1109/CVPR46437.2021.00666]

Li Y J, Chen Y C, Lin Y Y, Du X F and Wang Y C F. 2019. Recover and identify: a generative dual model for cross-resolution person re-identification//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 8089-8098 [DOI: 10.1109/ICCV.2019.00818]

Li Y L, He J F, Zhang T Z, Liu X, Zhang Y D and Wu F. 2021a. Diverse part discovery: occluded person re-identification with part-aware transformer//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 2897-2906 [DOI: 10.1109/CVPR46437.2021.00292]

Liang J Y, J. Cao J Z, Sun G L, K. Zhang K, Van Gool L and Timofte R. 2021b. SwinIR: image restoration using swin transformer//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, Canada: IEEE: 1833-1844 [DOI: 10.1109/ICCVW54120.2021.00210]

Liang T Y, Lan L, Zhang X and Luo Z G. 2021a. A generic MOT boosting framework by combining cues from SOT, tracklet and re-identification. Knowledge and Information Systems, 63(8): 2109-2127 [DOI: 10.1007/s10115-021-01576-2]

Liu J X, Ni B B, Yan Y C, Zhou P, Cheng S and Hu J G. 2018. Pose transferrable person re-identification//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 4099-4108 [DOI: 10.1109/CVPR.2018.00431]

Liu Y Q, Zhang X F, Wang S S, Ma S W and Gao W. 2020. Progressive multi-scale residual network for single image super-resolution [EB/OL]. [2022-11-04]. https://arxiv.org/pdf/2007.09552.pdf

Luo H, Gu Y Z, Liao X Y, Lai S Q and Jiang W. 2019. Bag of tricks and a strong baseline for deep person re-identification//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, USA: IEEE: 1487-1495 [DOI: 10.1109/CVPRW.2019.00190]

Makhzani A, Shlens J, Jaitly N, Goodfellow I and Frey B. 2015. Adversarial autoencoders [EB/OL]. [2022-11-04]. https://arxiv.org/pdf/1511.05644.pdf

Munir A, Lyu C, Goossens B, Philips W and Micheloni C. 2021. Resolution based feature distillation for cross resolution person re-identification//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal, Canada: IEEE: 281-289 [DOI: 10.1109/ICCVW54120.2021.00036]

Shen Q, Tian C, Wang J B, Jiao S S and Du L. 2020. Multi-resolution feature attention fusion method for person re-identification. Journal of Image and Graphics, 25(5): 946-955 (沈庆, 田畅, 王家宝, 焦珊珊, 杜麟. 2020. 多分辨率特征注意力融合行人再识别. 中国图象图形学报, 25(5): 946-955) [DOI: 10.11834/jig.190237]

Shi W D, Zhang Y Z, Liu S W, Zhu S D and Bao J N. 2020. Person re-identification based on deformation and occlusion mechanisms. Journal of Image and Graphics, 25(12): 2530-2540 (史维东, 张云洲, 刘双伟, 朱尚栋, 暴吉宁. 2020. 针对形变与遮挡问题的行人再识别. 中国图象图形学报, 25(12): 2530-2540) [DOI: 10.11834/jig.200016]

Sun D T, Yang L L, Lan L and Luo Z G. 2022. Toward to real low-resolution person re-identification: a new dataset and baseline//Proceedings of 2022 IEEE International Conference on Multimedia and Expo (ICME). Taipei, China: IEEE: #9860022 [DOI: 10.1109/ICME52920.2022.9860022]

Szegedy C, Ioffe S, Vanhoucke V and Alemi A. 2016. Inception-v4, Inception-ResNet and the impact of residual connections on learning//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, USA: AAAI: 4278-4284 [DOI: 10.1609/aaai.v31i1.11231]

Wang H C, Shen J Y, Liu Y T, Gao Y and Gavves E. 2022. NFormer: robust person re-identification with neighbor transformer//Proceedings of 2022 IEEE Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 7287-7297 [DOI: 10.48550/arXiv.2204.09331]

Wang Z, Ye M, Yang F, Bai X and Satoh S. 2018. Cascaded SR-GAN for scale-adaptive low resolution person re-identification//Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI). Stockholm, Sweden: AAAI Press: 3891-3897

Wu Z Z, Yu X C, Zhu D L, Pang Q W, Shen S T, Ma T and Zheng J. 2022. SR-DSFF and FENet-ReID: a two-stage approach for cross resolution person re-identification. Computational Intelligence and Neuroscience, 2022: #4398727 [DOI: 10.1155/2022/4398727]

Ye M, Shen J B, Lin G J, Xiang T and Hoi S C H. 2022. Deep learning for person re-identification: a survey and outlook. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(6): 2827-2893 [DOI: 10.1109/TPAMI.2021.3054775]

Zhang G Q, Ge Y, Dong Z C, Wang H, Zheng Y H and Chen S Y. 2021b. Deep high-resolution representation learning for cross-resolution person re-identification. IEEE Transactions on Image Processing, 30: 8913-8925 [DOI: 10.1109/TIP.2021.3120054]

Zhang P, Xu J S, Wu Q, Huang Q and Ben X Y. 2021a. Learning spatial-temporal representations over walking tracklet for long-term person re-identification in the wild. IEEE Transactions on Multimedia, 23: 3562-3576 [DOI: 10.1109/TMM.2020.3028461]

Zhang Y L, Li K P, Li K, Wang L C, Zhong B N and Fu Y. 2018.Image super-resolution using very deep residual channel attention networks//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 294-310 [DOI: 10.1007/978-3-030-01234-2_18]

Zheng W S, Hong J C, Jiao J N, Wu A C, Zhu X T, Gong S G, Qin J Y

and Lai J H. 2022. Joint bilateral-resolution identity modeling for cross-resolution person re-identification. International Journal of Computer Vision, 130(1): 136-156 [DOI: 10.1007/s11263-021-01518-z]

Zheng X, Lin L, Ye M, Wang L and He C L. 2020. Improving person re-identification by attention and multi-attributes. Journal of Image and Graphics, 25(5): 936-945 (郑鑫, 林兰, 叶茂, 王丽, 贺春林. 2020. 结合注意力机制和多属性分类的行人再识别. 中国图象图形学报, 25(5): 936-945) [DOI: 10.11834/jig.190185]

Zhou K Y, Yang Y X, Cavallaro A and Xiang T. 2019. Omni-scale feature learning for person re-identification//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 3701-3711 [DOI: 10.1109/ICCV.2019.00380]

作者简介

杨露露，女，硕士研究生，主要研究方向为计算机视觉。
E-mail: windychanchan@163.com
蓝龙，通信作者，男，副研究员，主要研究方向为计算机视觉。
E-mail: long.lan@nudt.edu.cn
孙冬婷，女，博士研究生，主要研究方向为计算机视觉。
E-mail: dongting_sun@sina.com
滕霄，男，博士研究生，主要研究方向为计算机视觉。
E-mail: tengxiao14@nudt.edu.cn
贲晛烨，女，教授，主要研究方向为模式识别和图像处理。
E-mail: benxianyeye@163.com
沈肖波，男，教授，主要研究方向为机器学习、计算机视觉和数据挖掘。E-mail: xbshen@njust.edu.cn