中图法分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(2023)05-1225-17

论文引用格式:Zhang Q, Lai J H, Xie X H and Chen H X. 2023. A summary on group re-identification. Journal of Image and Graphics, 28(05): 1225-1241(张权, 赖剑煌, 谢晓华, 陈泓栩. 2023. 小股人群重识别研究进展. 中国图象图形学报, 28(05):1225-1241)[DOI:10.11834/jig.220697]

小股人群重识别研究进展

张权[1],赖剑煌[1,2,3,4*],谢晓华[1,3],陈泓栩[1]

1. 中山大学计算机学院,广州　510006; 2. 广州新华学院,广州　510520; 3. 广东省信息安全技术重点实验室,广州　510006; 4. 视频图像智能分析与应用技术公安部重点实验室,广州　510006

摘　要:小股人群重识别旨在将非重叠视域的摄像头网络下具有相同成员的群组图像进行正确的关联。小股人群重识别是传统行人重识别任务的一个重要拓展,在安防监控场景下有着重要的研究意义和应用前景。小股人群重识别所面临的独特挑战在于如何针对群内成员的数量变化和布局变化进行建模,并提取稳定、鲁棒的特征表达。近年来,小股人群重识别引发了研究人员的广泛关注,并获得了快速的发展。本文对小股人群重识别技术的研究进展进行了全面的梳理回顾。首先简要介绍本领域的研究背景,对基本概念、数据集和相关技术进行了简要总结。在此基础上,对多种小股人群重识别算法进行了详细的介绍,并在多个数据集上对前沿算法进行性能对比。最后,对该任务进行展望。整体而言,与行人重识别相比,小股人群重识别的现有方法在具体场景下的特定挑战性能表现欠佳,还需要从数据收集和方法设计两方面进一步探讨。此外,现有的小股人群重识别研究与其他视觉任务的关联性不够紧密,如何协同多任务作业以解决更多业界需求、加速产业落地,需要学术界和工业界共同思考和推动。

关键词:小股人群重识别(GReID);行人重识别;虚拟数据;深度学习;特征学习;度量学习;Transformer

A summary on group re-identification

Zhang Quan[1], Lai Jianhuang[1,2,3,4*], Xie Xiaohua[1,3], Chen Hongxu[1]

1. School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou　510006, China; 2. Guangzhou Xinhua University, Guangzhou　510520, China; 3. Guangdong Province Key Laboratory of Information Security Technology, Guangzhou　510006, China; 4. Key Laboratory of Video and Image Intelligent Analysis and Application Technology, Ministry of Public Security, Guangzhou　510006, China

Abstract: Pedestrians-oriented group re-identification (GReID) analysis is focused on non-overlapped and multi-viewed small groups. To extract stable and robust feature representations, the challenge issue of GReID is to model the temporal changes and intra-group pedestrians. Our summary is reviewed on the growth of GReID critically. First, we review its research domain in related to its basic concepts, technologies and datasets in relevance. To optimize the surveillance in public security, the GReID can monitor and prevent group-based crimes accurately like women and children-oriented kidnapping and trafficking. Due to pedestrians-targeted are severely occluded or even disappeared, it can leverage the appearance features of pedestrians' partners as additional prior information for recognition. Specifically, GReID-based groups are composed of 2 to 8 members. First, the same group can be identified when the identified intersection-over-union(IoU) ratio of member is greater than 60% in the two group images. Then, a variety of GReID algorithms are introduced and

收稿日期:2022-07-07;**修回日期**:2022-09-28;**预印本日期**:2022-10-05

*通信作者:赖剑煌　stsljh@mail.sysu.edu.cn

基金项目:国家自然科学基金项目(62076258);中国公安部科技项目(2019GABJC39)

Supported by: National Natural Science Foundation of China (62076258); Project of Ministry of Public Security of China (2019GABJC39)

tested in detail. The existing works can be categorized from three perspectives: 1) data, 2) method, and 3) label. For data types, the existing methods can be segmented into: real image-based, synthetic images-based, and real video-based methods. The real images-based method is basically focused on the datasets collected from real surveillance scenarios, such as CUHK-SYSU Group (CSG), RoadGroup, iLIDS-MCTS, and etc. These datasets can be used to collect several group images from different camera views of different groups and provide the elements of location information and identification information of member. This supervision information can be used to design discriminative group feature representations. However, it is still more challenging to collect and label the real group datasets than the traditional pedestrian re-identification datasets because the consistent group identity is required to be judged between group images, including member variations and layout variations. The following datasets are proposed based on 3D synthetic images. This type of datasets can generate mass group images with high-quality labels efficiency and effectively. These methods can be used to improve the performances of the model in real datasets through massive synthetic data. The video-based datasets can provide several consecutive frames for each group from the surveillance videos. Researchers can extract the group features according to the potential patio-temporal or intra-group relationships. They can be mainly divided into: traditional methods and deep learning methods. The former one is to design group descriptors and extract group features derived of human experience. However, due to the high dependence on the prior knowledge of expertise, it is unable to describe and generalize all possible situations for group images. The model can construct the representations of group images automatically because the emerging deep learning based methods is beneficial for a large number of data samples, and the discrimination and robustness of the deep models have been significantly improved. Deep learning-based methods can be divided into 1) the feature learning-based, 2) metric learning based, and 3) generative adversarial network (GAN) based. The deep feature learning based methods aim to design a discriminative network structure or a discriminative feature learning strategy. The features-extracted can reflect the group identification of the input images accurately, and it can be robust enough to suppress occlusion, illumination, number and layout variations of intra-group members. Metric learning based methods can be focused on a similarity criterion evaluation between two groups of images. To get high similarity under the designed measurement criteria, even two group images from the same group class have great differences. To optimize small size of the dataset, GAN-based method attempts to expand the dataset scale of the GReID task by style transfer of samples from other related pedestrian re-identification datasets. For its label, the existing methods can be categorized into: supervised and unsupervised. Supervised learning based methods tend to be more competitive because the group labels or the member labels are participated in the entire training process. It often can learn the similarity only for the local area of the group images because labels are not be provided in the unsupervised learning, and cluster methods can be designed to extract the feature representations of the same group class. To sum up, 1) the specific scenarios based GReID is required to be developed from the aspects of data collection and method design further; 2) GReID is still not interrelated to other related visual tasks mutually. Therefore, multiple tasks-collaborated are called to resolve more industry needs, and the implementation of the industry is required to be accelerated for the domain of academia and industry. Furthermore, the data privacy policy-relevant ethic issue needs to be utilized for virtual data and real data in the future.

Key words: group re-identification (GReID); pedestrian re-identification; synthesis data; deep learning; feature learning; metric learning; Transformer

0 引 言

行人重识别(person re-identification)是计算机视觉在安防领域的一个重要任务。该任务旨在一个视域互不重叠的摄像机网络中,将不同摄像机下属于同一身份类别的行人图像进行正确的关联(赵才荣 等,2021)。该任务自从2006年首次在国际计算机视觉与模式识别会议提出以来(Gheissari 等,2006),得到了学术界和工业界的广泛关注。在经过了十几年的快速发展之后,行人重识别的模型框架日趋成熟,相关算法落地初展头角,在我国有着广泛的应用价值,给城市安全和公众财产给予了强有力的保障和支撑。然而,现有的行人重识别技术仅关注对于单个行人研究,对于多个行人同时出现的情况没有过多讨论。由于人类社会存在群居属性以及

社会角色的多样性，人们更倾向于以小群体形式结伴出行，而非独立出现（Mei 等，2020）。

将小规模人群作为重识别目标的任务，简称小股人群重识别（group re-identification，GReID）（Zheng 等，2009）。该任务对于行人重识别中被研究对象的数目进行了拓展，大幅拓展了行人重识别的应用场景，对某些团伙作案的犯罪行为（如绑架）起到了良好的检测、防范和打击作用，对社会安全和人民财产起到了保护作用。

除此之外，小股人群重识别也可以促进行人重识别的性能进一步提高。具体地，由于监控摄像头的环境复杂，单个行人很容易受到遮挡的影响。在严重遮挡的条件下，仅依靠单人的表观特征并不可靠。此时小股人群重识别可以考虑将与之同行的其他群体成员的表观特征作为检索线索，提高行人重识别在复杂环境下的鲁棒性和判别性。

相比于行人重识别，小股人群重识别的研究相对较少。因为小股人群不仅需要面对传统行人重识别中的挑战，还需要解决由于图像中群组结构变化带来的挑战，通常包括群组人数的变化和群组成员位置的变化。近年来，提出了不少围绕小股人群重识别的工作，使该任务在深度学习时代得到了快速发展。Lin 等人（2021）提出了多粒度表示（multi-grained representations，MGR），引入深度网络结构作为成员表观特征的提取器。Huang 等人（2019a）提出了基于域迁移的图神经网络方法（domain transfer graph neural network，DoTGNN），引入对抗生成网络增加群组的多样性。Yan 等人（2020）提出了多注意上下文图匹配（multi-attention context graph，MACG），将图结构与注意力机制相结合提取群组特征，Zhang 等人（2022b）提出了二阶自注意力模型（second-order Transformer，SOT），首次将自注意力模型（Transformer）应用在这项任务中。然而，上述工作都是针对小股人群重识别任务的某一具体问题或者挑战展开技术性研究，对该任务的整体回顾和未来发展缺乏详细的讨论。

本文是对小股人群重识别任务发展至今的综述性研究。具体地，本文对已有工作在数据集、算法和性能等方面进行系统论述，对已有算法和模型进行归纳总结，对未来的研究趋势进行讨论和展望。

1　小股人群重识别问题及技术总览

在一个视域互不重叠的摄像机网络中，小股人群重识别任务旨在将包含多个相同行人的群组样本（图像或者视频序列）进行正确的关联。考虑到任务的复杂性以及实际情况下的群组成员数量，该任务通常考虑人数较少的目标群组，当前的研究通常关注分布在 2～8 人之间的目标人群。如图 1 所示，给定的待查询群组样本是由两个行人组成的群组图像，该任务的目标是在底库图像集中找到所有与之群组身份相同、摄像头不同的群组图像，并按照相似度大小返回检索结果。

图 1　小股人群重识别任务描述

Fig. 1　Illustration of group re-identification

小股人群重识别任务与密集人群视觉任务相比，其不同点在于由于研究对象的数目不同导致所关注的问题不同。在密集人群的视觉任务中，由于个体的表观特征往往难以获得，这些任务更关注整个高密度人群的属性分析，例如群体人数、群体运动趋势、群体行为等。而小股人群重识别由于较少的人群数量，可以得到较为丰富的个体信息，并在此基础上研究如何进行多目标的关系建模，从而提取到关于群体的特征表达。

小股人群重识别任务与行人重识别相比，其独特性在于同一个群组身份呈现群组结构的多样性，即由于群组成员数量的变动和群组成员位置的变动所引发的群组结构变化。其中，关于群组的成员人数，现有研究（Yan 等，2020）一般认为两幅群组图像的共同人数超过 60% 时，即认为这两幅群组图像属于同一个群组类别。

小股人群重识别按照数据类型、标签设置和模型策略的不同可以划分为不同子任务，如图 2 所示，但是本质都是在围绕如何针对上述两种变动对群组关系进行建模和描述。

根据群组数据类型的不同，可分为基于静态图

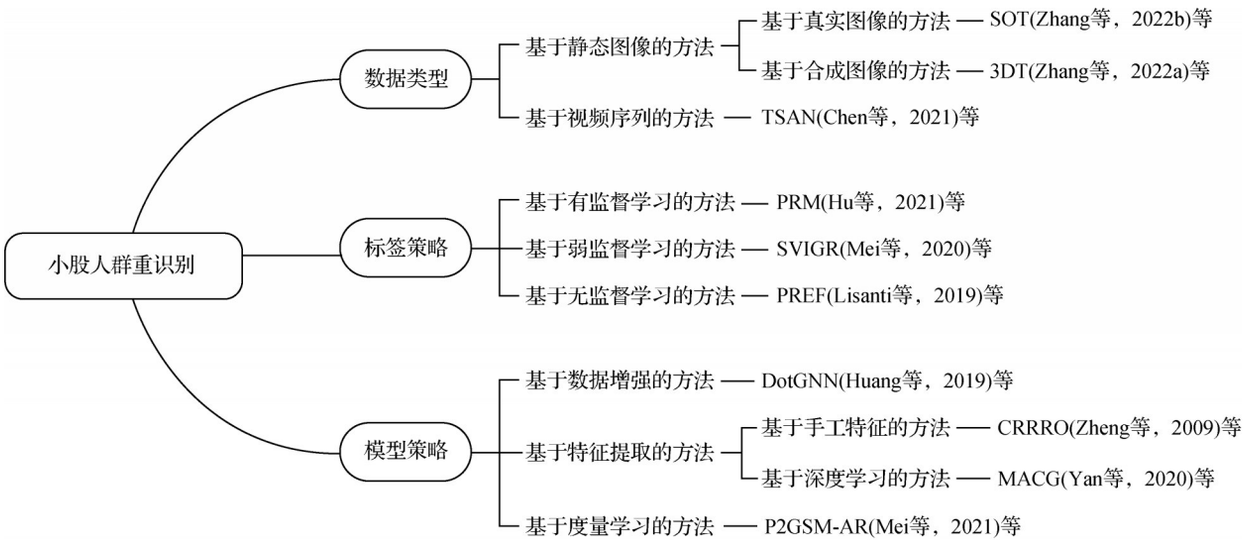

图2　小股人群重识别技术总结

Fig. 2　Technology summary of group re-identification

像的方法和基于视频序列的方法。基于静态图像的方法包括Zheng等人(2009)提出的中心矩形环比占用描述符(center rectangular ring ratio-occurrence descriptor, CRRRO)、Cai等人(2010)提出的协方差描述符(covariance)、MACG(Yan等，2020)和SOT(second-order Transformer)(Zhang等，2022b)等。基于视频序列的方法包括Chen等人(2021)提出的双流注意力网络(two-stream attentive network, TSAN)等。

根据群组标签设置的不同，可分为基于有监督的方法、基于弱监督的方法和基于无监督的方法。基于有监督的方法包括Xu等人(2019b)提出的混合注意力模型(hybrid attention model, HAM)、MACG(Yan等，2020)、SOT(Zhang等，2022b)、Zhu等人(2021)提出的群组上下文图神经网络(group context graph neural networks, GCGNN)和Zhang等人(2022a)提出的3维自注意力模型(3D Transformer, 3DT)等。基于弱监督的方法包括DotGNN(domain-transferred graph neural network)(Huang等，2019a)和Mei等人(2020)提出的孪生群组检索方法(siamese verification-identification-based group retrieval, SVIGR)等。基于无监督的方法包括CRRRO(Zheng等，2009)、Covariance(Cai等，2010)和Lisanti等人(2017，2019)提出的基于编码特征的残差表示方法(pooling residuals of encoded features，PREF)等。

基于模型策略的不同，可分为基于数据增强的方法、基于特征提取的方法和基于度量学习的方法。基于数据增强的方法包括Huang等人(2019b)构造的成对表征的迁移学习网络(transferred single and couple representation learning network，TSCN)、DotGNN(Huang等，2019a)、DotSCN(Huang等，2021)和Liu等人(2021)提出的单特征注意力网络(single feature attention learning network, SFALN)。基于特征提取的方法包括CRRRO和MACG。基于度量学习的方法包括Xiao等人(2018)提出的利用和集成多粒度信息(leveraging and integrating multi-grain information，LIMI)的方法、Xu等人(2019b)提出的最小二乘残差距离(least squares residual distance, LSRD)，SVIGR(siamese verification-identification-based group retrieval)(Mei等，2020)和Mei等人(2021)提出的基于人与群组相似性匹配的歧义消除方法(person-to-group similarity matching based ambiguity removal，P2GSM-AR)。其中，特征提取又可以进一步划分为基于人工设计的手工特征方法和基于数据驱动的深度学习方法。前者包括CRRRO-BRO(Zheng等，2009)、Covariance描述符(Cai等，2010)、Salamon等人(2015)提出的基于软生物特征的方法(soft-biometric characteristics，SBC)和PREF(Lisanti等，2017)。后者包括HAM模型(Xu等，2019b)、MACG(Yan等，2020)、GCGNN(Zhu等，2021)、Hu等人(2021)提出的部分相关均值模型(part relational mean model，PRM)、SOT(Zhang等，2022b)和3DT(3D transformer)模型(Zhang等，2022a)。

尽管小股人群重识别的细分子任务众多，但是

现阶段仍处于起步阶段，大量的任务仍集中在基于图像和有监督学习的小股人群重识别方法，这也是本综述重点阐述的内容。

2　小股人群重识别相关数据集

为了支持小股人群重识别的研究，研究人员已经构建了许多群组数据集。本节主要按照数据集的数据类型分别进行介绍。

2.1　静态图像数据集

2.1.1　真实图像数据集

iLIDS-MCTS（the imagery library for intelligent detection systems，multiple camera tracking scenario）数据集（Zheng 等，2009）是最早公开的小股人群重识别数据集之一，该数据集是在机场大厅中的一个多摄像机监控网络在机场繁忙时期拍摄的。该数据集提取了两个互不重叠的摄像头下的群组图像，共包含 64 个群组身份和 274 幅经过裁剪的群组图像。大多数组有 4 幅图像，来自不同的摄像头或来自同一摄像头但在不同时间不同位置。这些群组图像的大小不同。如果需要得知群组成员的相关信息，还需要研究者使用额外的检测算法。

CASIA（Institute of Automation，Chinese Academy of Sciences）群组数据集（Cai 等，2010）和 OULU 群组数据集（Cai 等，2010）同时提出。其中，CASIA 群组数据集由两个室外的摄像头采集得到，共收集了 44 对（即 88 幅）群组图像。OULU 群组数据集由 5 个室内的摄像头采集得到，共收集了 20 对（即 40 幅）群组图像。这两个数据集提供了更加完整的相机成像画面而非经过裁剪的。因此，两个数据集的实验结果指出，如果利用头肩检测算法得到多于两个行人的检测结果，则将行人区域定义为包含群组图像的前景区域，其余为背景区域。

ETHZGroup（ETH-zurich Group）数据集（Salamon 等，2015）是在 ETHZ 数据集（Ess 等，2007）上进一步发展得到的。该群组数据集共标注了 141 个群组身份，213 个群组成员身份以及 72 幅群组图像。该数据集着重强调个人和群体在不同场景下的反复出现。统计表明，ETHZGroup 数据集中有 29 个群组成员和 11 个群体反复出现，具体地，约有 13.61% 的个人和 7.8% 的群体至少出现两次。

ZeCSS（zebra crossing summer scenario）数据集（Zhu 等，2016）是由公路斑马线两端的两个非重叠视域摄像头拍摄得到。共包括 189 个群组类别以及 514 幅经过裁剪的群组图像。该数据集保证每个群组在每个摄像头下至少出现一次，除此之外，该数据集仅考虑群组成员数目在 2～4 人之间的群组。

QUEUE02 数据集（Koperski 等，2017）用于模拟两个地点之间的群体排队场景，由两个不同场景下的摄像头采集，这两个场景的光照条件差异非常显著。该数据集共标注了 23 个行人身份和 15 个群组身份。与之前所有的群组数据集不同，该数据集提供了群组图像的拍摄时间（即时间戳），使得更多基于时空关联的方法可以被考虑。

MUSEUM 数据集（Lisanti 等，2017）是在博物馆大厅通过 4 个摄像头采集的，其中摄像头视域的重叠很小或没有重叠。该数据集包含 524 幅群组图像，由不同人数的群组成的 18 个群组类别。每个群组大约有 30 幅图像。

OGRE（outdoor groups re-identification）数据集（Lisanti 等，2017），由停车场场景下的 3 个不相交的摄像头采集。该数据集包含 39 个群组身份，并且在不同时刻和不同天气条件下采集了大约 2 500 幅图像。OGRE 数据集是通过一种弱监督方法标注的，即给定一个手动选择的群组区域，通过运行 KCF（kernelized correlation filter）跟踪器（Henriques 等，2015）获得后续检测结果。此外，该数据集还涵盖了严重的视角变化和大量的群组成员自遮挡。

RoadGroup 数据集（Xiao 等，2018）是在一个人流密集的马路上由两个视域互不重叠的摄像头拍摄。摄像机所拍摄的群组包含了 162 个群组身份，每个群组包含两幅经过裁剪的图像。与之前的工作不同，RoadGroup 和 DukeGroup 数据集首次定义了在群组成员数量变动与群组类别的关系。即只要两幅群组图像的共同成员超过 60%，即认为这两幅图像属于同一个群组身份类别。

SYSUGroup（Sun Yat-sen University Group）数据集（Mei 等，2020）是在大学校园场景下完成采集的。采集设备由 8 个高清摄像头组成（摄像头 1～3 为 1 920 × 1 080 像素，摄像头 4～8 为 1 280 × 1 080 像素），场景包括室内及室外。该数据集邀请了 524 位志愿者完成数据集构建，组成了 208 个不同类别的群组，共采集了 7 071 幅群组图像。其中每个群组内的成员数均在 2～6 人之间，一旦群组人数确定，其

在所有的采集场景下的人数均固定不变。在采集完成之后，所有图像均调整为256×256像素的分辨率。

CSG（CUHK-SYSU Group）数据集（Yan等，2020）是在行人搜索数据集CUHK-SYSU（Xiao等，2017）的基础上发展而来的。CSG共包含3 839幅群组图像、1 558个群组身份以及大约3 500个成员身份标签。这些图像一部分来自监控数据，另一部分来自影视剧等非监控场景。每个群组包含的人数在2～8人之间。其中有28%的群组图像出现了遮挡。为了进一步增加难度，CSG数据集在测试阶段的底库集合中又额外添加了5 000幅图像作为干扰图像。

BRT（bus rapid transit）数据集（Hu等，2021）是基于某城市公交线路图进行采集的，数据集所涵盖的时空范围相较于之前的数据集都有了显著拓展。该数据集在53个视像头下采集完成，共标注了400个群组，3 210幅群组图像。

OWGA-Campus（open world and group ambiguity campus）数据集（Mei等，2021）是对SYSUGroup数据集的拓展。上文指出，SYSUGroup数据集并不包含群组成员的变动，是相对干净和理想的状态。OWGA-Campus数据集在此基础上，考虑了群组成员数量的变动以及路人的随机加入与离开。该数据集由9个高清摄像头采集，共包含210个群组类别和10 808幅群组图像。每个群组有1幅标准的群组图像作为参考，并在测试阶段当作查询图像，其余在底库中的图像是群组人数变动的图像。

2.1.2　合成图像数据集

2.1.1节展示的数据均是来自真实场景下的群组图像，其好处是能够准确地反映某一地域下小股人群的运动规律，然而其缺点也十分明显。首先，群组重识别的数据集采集难度更大，因为需要标注人员手工挖掘存在组关系的多个行人；其次，群组重识别的标注难度更困难，因为不仅需要对每个成员给出单独的身份标签，还需要对整个群组进行标注。在进行跨摄像头的群组关联时，所需要付出的标注成本更是成倍增长。最后是由于数据隐私政策，对于行人数据的采集、标注、公开都提出了更高更严的标准，使得大规模公开数据集的制作难以进行。

得益于虚拟动画引擎技术的发展，基于合成数据的研究成为了一个新兴的任务，并在传统的行人重识别领域被证明了是一种低成本且有效的方式，例如PersonX（Sun和Zheng，2019）、RandPerson（Wang等，2020）和UnRealPerson（Zhang等，2021）等。因此，City1M数据集（Zhang等，2022a）首次利用虚拟动画技术将合成数据引用于小股人群重识别任务中。City1M数据集构建了一个基于城市中心街区的场景，选择了路边8个视域互不重叠的摄像头作为采集设备。同时，City1M构建了45 000个不同的行人模型，基于此构建了11 500个不同的群组类别，共采集了1 840 000幅全画幅的群组图像，每幅图像的分辨率均为1 920×1 080像素。City1M是首个小股人群重识别领域百万级规模的数据集。为了增加数据集的多样性，City1M包括了尺度变化、光照变化、昼夜变化、遮挡与自遮挡以及群组内的人数变化和位置变化。

2.2　视频序列数据集

DukeGroupVid数据集（Xu等，2019a）是深度学习时代首个基于视频序列的小股人群重识别数据集。该数据集由8个摄像头组成，共收集了371个群组类别和890段轨迹序列，每段序列包含数量不固定的视频帧，范围从12帧～6 444帧不等。每个群组类别出现在2～4个摄像头下。由于小股人群重识别的研究在初期阶段，基于图像的任务尚未良好解决，因此基于视频的小股人群数据集目前还没有得到过多的关注。

2.3　小结

小股人群重识别的数据集主要可分为静态图像数据集和视频序列数据集，目前已有的静态图像数据集较多且具有不同的风格，而视频序列数据集较少且风格单一。静态图像数据集又可分为真实图像数据集和合成图像数据集，其中已有的小股人群重识别数据集以真实图像数据集为主。表1给出了已知群组数据集的统计结果。图3给出了部分数据集的实例，其中人脸数据已经被保护。

3　小股人群重识别算法

本节从数据、标签和模型3个维度对现有的小股人群重识别方法进行介绍。需要指出的是，这3个维度并不是彼此孤立的，其中基于图像的有监督深度学习是目前发展相对快速的方向。

表 1　小股人群重识别数据集内容摘录

Table 1　Abstract of group re-identification datasets

数据集分类	数据集名称	年份	群组身份/个	行人身份/个	群组图像/幅	摄像头/个	备注
静态图像数据集	iLIDS-MCTS	2008	64	–	274	2	
	CASIAGroup	2010	–	–	88	2	
	OULUGroup	2010	–	–	40	5	
	ETHZGroup	2015	141	213	72	–	
	ZeCSS	2016	189	–	514	2	
	QUEUE02	2017	15	23	–	2	带有时间戳标签
	MUSEUM	2017	18	–	524	4	
	OGRE	2017	39	–	~ 2 500	–	
	RoadGroup	2018	162	–	324	2	
	SYSUGroup	2020	208	524	7 071	8	
	CSG	2020	1 558	3 500	3 839	Vary	包含影视剧等场景
	BRT	2021	400	–	3 210	53	
	OWGA-Campus	2021	210	–	10 808	9	
视频数据集	DukeGroupVid	2019	371	–	–	8	首个基于视频的数据集
合成数据集	City1M	2022	11 500	45 000	1 840 000	8	首个百万级虚拟数据集

注：“–”表示暂无相关数据。

(a) OGRE　(b) MUSEUM　(c) CSG

(d) City1M　(e) RoadGroup　(f) SYSUGroup

图 3　小股人群重识别数据集展示

Fig. 3　Visualization of some group re-identification datasets

((a) OGRE; (b) MUSEUM; (c) CSG; (d) City1M; (e) RoadGroup; (f) SYSUGroup)

3.1　数据类型

3.1.1　基于静态图像的方法

基于静态图像的方法是现阶段最常用的小股人群重识别任务下的研究对象。具体地，给定一幅孤立的包含群组成员的图像(无论是全局的还是经过裁剪的，后文统称为群组图像)，需要根据其建模群

组的表观特征和群组的布局特征。早期的方法是通过设计特征描述子在全图范围内提取群组特征，例如CRRRO（Zheng等，2009）和Covariance（Cai等，2010）等。在深度学习时代，主流的研究流程是对图像中出现的行人进行检测并裁剪，以提取准确的行人特征，并在此基础上，对群组的结构和特征进行建模，例如基于卷积神经网络的方法MACG（Yan等，2020）和基于无卷积自注意力模型的方法SOT（Zhang等，2022b）等。

除了上述的研究任务（判别性群组特征建模），基于合成图像的方法的额外研究任务是如何利用虚拟图像辅助模型在真实图像上的性能表现。Zhang等人（2022a）通过实验发现，由于现有的小股人群重识别数据集规模普遍不大，利用百万级合成数据集City1M进行预训练，然后直接迁移到现有的真实图像数据集中进行测试，已经实现了具有竞争力的性能表现。

3.1.2　基于视频序列的方法

基于视频序列的方法旨在将某个群组一系列连续帧作为输入，并对其进行群组特征建模。相比于单帧图像，基于视频序列的输入提供了更多的群组知识，对于某些单帧出现严重遮挡的不理想情况，该任务可以很好地通过前后帧之间的关系进行补充和重建。目前有DukeGroupVid数据集支持相关的研究工作。其中，Xu等人（2019a）设计了一个结合卷积神经网络和循环神经网络的统一框架。具体地，该方法先利用卷积神经网络对输入的每一帧进行特征提取，然后利用循环神经网络对每帧中具有判别性的特征进行挖掘和聚合，最终得到完整的关于输入序列的群组特征表达。在此基础上，Chen等人（2021）对上述方法进一步完善，分别从个人群体和个人处学习信息，设计得到了双流注意网络TSAN，该网络主要考虑了视频序列中空间特征与时间特征的融合。

3.2　标签策略

3.2.1　基于有监督学习的方法

基于有监督的学习方法是该任务目前主流的研究策略。在训练阶段，数据集会提供所有的标签信息，包括但不限于成员身份标签、成员位置、群组身份标签和群组区域等。一般而言，将使用了“群组标签+成员标签”的组合称为有监督学习。例如，HAM（Xu等，2019b）、MACG（Yan等，2020）和SOT（Zhang等，2022b）等方法在模型的训练阶段对这两方面的标签信息都进行了充分利用。模型可以在此基础上进一步选择使用其他的标签信息，例如使用群组身份标签进行监督学习。GCGNN（Zhu等，2021）使用群组成员在图像中的2维位置坐标进行群组关系建模；3DT（Zhang等，2022a）在合成数据集City1M上使用数据集提供的成员3维坐标来建模群组布局关系，对于仅提供2维坐标的真实数据集，3DT先利用已有的单目深度估计算法重构成员的3维位置坐标，然后再进行群组布局关系建模。除此之外，还有使用额外信息的方法。例如Koperski等人（2017）在QUEUE02数据集上考虑了群组图像之间的时空关联性，利用群组在摄像头之间的运动时间来修正某些困难样本仅基于表观特征的错误关联。

3.2.2　基于弱监督学习的方法

基于全监督的方法需要大量人力标注成本，因此部分方法通过仅使用一部分身份标签的策略来构建模型。这些仅使用了群组身份标签或者成员身份标签的方法称为弱监督学习方法。例如，DotGNN（Huang等，2019a）仅使用群组标签来设计和训练模型，SVIGR（Mei等，2020）分别评测了仅使用群组身份或者成员身份两种情况下的性能。SVIGR的实验结果表明，两种标签都可以使模型关注到判别性特征，使用个人身份标签的效果会略优于使用群组身份标签的效果。

3.2.3　基于无监督学习的方法

基于无监督学习的方法仅使用数据集中提供的群组图像，然后自发地根据其内在相似度对相同群组身份的图像进行关联。基于传统手工设计的特征描述符不依赖数据集中的标签，因此可以归属为无监督学习方法，例如CRRRO（Zheng等，2009）、Covariance（Cai等，2010）和PREF（Lisanti等，2017，2019）等。

3.3　模型策略

与已有识别任务类似，小股人群重识别的基本流程主要分为3个阶段，如图4所示。第1阶段是数据增强，即对数据集提供的有限样本进行数量扩充，在实例级层面增加群组的多样化表达；第2阶段是特征提取，即设计相应的特征提取网络或者特征描述符对输入群组进行建模和特征抽取；第3阶段是特征度量，即设计特定的度量空间或者特定的特征相似性准则对上一阶段抽取得到的特征进行同类聚

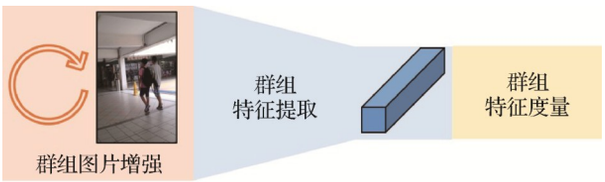

图 4　小股人群重识别的任务流程

Fig. 4　Pipeline of group re-identification

合和异类分离，这样使得方法在测试阶段能够对未知类别的群组依然保持良好的内聚性。现有的主流方法本质上从这 3 个层面出发来提升小股人群重识别的性能表现。

3.3.1　基于数据增强的方法

这类方法主要是从生成式对抗网络的角度出发，通过已有的行人数据集拓展小股人群重识别数据集的样本规模。

TSCN（Huang 等，2019b）首先利用风格迁移的对抗生成网络将传统行人重识别数据集中的行人图像迁移成为小股人群重识别中的行人风格，然后手工对迁移后的行人图像进行群组构造（均仅包含两个成员，并将两幅行人图像进行拼接即完成群组构造）；之后对于每个群组，先单独提取单人表观特征，然后将单人表观特征的差作为群组特征表达。在测试阶段，如果出现多于两个人的群组，则将行人的表观特征两两相减作为群组的关系特征。

DotGNN（Huang 等，2019a）同样是利用基于对抗生成网络的域迁移算法实现行人风格从已有行人重识别数据集到小股人群重识别数据集的转换。在进行手动群组的构造时，DotGNN 考虑了群组内成员数量更多元化的情况，分别生成了 2～6 人不等的群组。除此之外，在构造群组时，DotGNN 还设计了群组变动的情况，即在一个成员稳定不变的基础群组中，随机引入额外的干扰成员并且不引起群组类别的改变。在建模群组特征时，也采用了图神经网络群组进行建模和特征提取。

DotSCN（Huang 等，2021）是对 TSCN 的改进和拓展。该方法在数据增强方面与 TSCN 是类似的，在对群组特征进行建模和提取时，提出了新的特征融合方式。通过对排序—距离曲线的观察，提出了基于 L 型曲线的单人和成对特征的融合策略。

SFALN 模型（Liu 等，2021）利用循环生成网络模型实现传统行人数据集与小股人群数据集之间的图像风格迁移。然后利用将深度残差网络与非局部注意力模块相结合的网络模型提取群组图像特征。

3.3.2　基于手工特征的方法

基于手工特征的方法是早期小股人群重识别研究中常见的方式之一，这类方法不依赖海量数据，与早期的数据集规模是契合的，也给后来的研究提供了良好的思路和方向。

CRRRO-BRO（Zheng 等，2009）描述符是早期的研究工作之一。其中，CRRRO 是一种近似旋转不变的描述符，从图像的中心开始，以矩形环带的方式沿径向向外展开。接下来，在每个矩形环区域内构造基于直方图特征表达，最终得到整幅图像的特征。但是，CRRRO 仍然不能很好地处理成员在群体中大幅度的非中心旋转变化。同时 CRRRO 也没有利用任何局部结构信息，因此又进一步提出了 BRO 描述符。BRO 基于全图的均匀切块，并对每一个子块提取相应的局部特征。

Covariance 描述符（Cai 等，2010）也是基于手工特征的方法，早期应用在目标检测等多个视觉任务中。该描述子的优良性质在于协方差矩阵中编码了足够多的判别信息。此外，它还以一种较低的维数将多种异构特征自然地组合在一起。在任务中，Covariance 描述符对于群组图像中每一个像素的空间位置和 RGB 值进行编码，得到一个低维度的协方差矩阵作为群组图像的特征表达，并用基于协方差矩阵的特征值来衡量群组图像之间的相似度。

SBC 方法（Salamon 等，2015）是一种用户交互式的特征描述子，是一个基于手工定义的软生物特征。用户能够选择描述属于小股人群的每个成员的软生物统计学属性的颜色。该方法使用颜色距离度量和新颖的自适应阈值进行检索匹配。实验结果表明，该方法能够帮助重识别过程在没有训练数据的情况下对最可能的结果进行排序，并且可以在没有先前图像的情况下进行扩展。

BSC + CM（boosted salience channels + consistent matching）方法（Zhu 等，2016）是一个两阶段的基于手工特征的群组描述方法。在第 1 阶段使用增强式显著性通道 BSC 的方法，在第 2 阶段进行一致性匹配 CM。其中，BSC 首先将群组图像利用稠密网格均匀分为若干个子块，并对每一个子块用无监督显著性算法得到每个子块的显著性分数（得分区间在 0～1 之间），然后利用阈值过滤保留存在显著行人的子块，简称为显著性块集合，为后一步处理做准

备。在一致性匹配CM阶段，对于一对群组图像中的两个显著性块集合进行相似度匹配，并根据基于对应块的相似度构造出两个显著性块集合的相似度，作为两个群组图像之间的距离。

PREF方法（Lisanti等，2017）是基于稀疏字典编码的手工特征方法。该方法首先在单个行人的数据集上进行稀疏字典的学习，然后在群组数据集上利用学习好的字段构造群组特征。在字典的学习过程中，首先将单人行人的图像均分切分为若干个正方形子块。对每一个子块提取在RGB，LAB，HS等3种颜色空间下的直方图特征。然后对每一个子块学习稀疏的字典表达，即希望用少数的字典中的原子表达，可以最大程度地重构出输入块的颜色特征。在求解字典的原子表达时，PREF使用了最小角回归（least angle regression，LARS）的迭代优化方法（Efron等，2004）。在测试阶段，对于群组图像的每一个子块，计算与对应字典原子的系数作为特征，最后将所有块的系数特征执行平均值或最大值，并取其主成分分析下的最相关分量作为群组特征表达。PREF方法后续又得到进一步完善（Lisanti等，2019），但核心思想仍然是一致的。

3.3.3　基于深度特征的方法

深度学习在越来越多的视觉任务中发挥着不可忽视的作用，因此研究人员也考虑将其应用在小股人群重识别任务中。其中主要包括卷积神经网络、图神经网络以及自注意力模型Transformer的单独使用及组合。

HAM模型（Xu等，2019b）旨在解决基于静态图像的小股人群重识别任务中的空间结构变化问题，HAM是基于卷积神经网络设计的，并在此基础上同时加入了空间注意力和通道注意力机制，使得模型更加关注能够提取判别性特征的区域。

MACG（Yan等，2020）模型是一个基于多级注意的上下文图结构学习框架，它是卷积神经网络与图神经网络的组合模型。具体而言，MACG将每个小股人群建模为上下文图，其中节点指的是组内成员。上下文图表示有两个优点。首先，可以通过节点的特征聚合获得图层级的群组表示，这从本质上解决了群组布局和群组成员变化的挑战。其次，通过使用图神经网络，组中的节点特征可以受益于上下文信息，这大幅促进了信息通过图边的传播。基于此，小股人群重识别任务可以建模为图层级特征学习任务。此外，MACG还提出了一种多层次的注意机制来学习小股人群的上下文信息图。对于节点级表示，提出了图内和图间注意模块，分别对同一幅图内和不同图间的上下文信息进行编码。MACG进一步构建了更高层次的注意机制，用于聚合节点级特征，以获得最终的图级表示。在测试阶段，一幅群组图像的图级特征表达直接用于特征检索。

GCGNN（Zhu等，2021）对每一个群组建模为空间K-最近邻图（spatial K-nearest neighbor graph，SKNNG），并在此基础上提出了一个组上下文图神经网络用于群组特征表示学习。GCGNN基于一个先验观察，即在不同的摄像机下，相距较近的群内成员比相距较远的组内成员更可能保持在同一组中。具体而言，GCGNN根据群组图像中成员的位置信息坐标构建SKNNG，其中每个节点仅与最相邻的K个成员存在邻接关系。对于所构建图中的每个节点（每个成员），GCGNN学习相应的一个特征嵌入，该嵌入从相邻节点聚合上下文信息。GCGNN根据图的属性，包括节点的度和空间关系属性，设计了多个邻域聚合加权核。最后，GCGNN计算组成员关联的两个图的节点嵌入之间的相似度得分，并将所有链接节点对的相似度得分相加，得到两个图之间的匹配得分。

PRM方法（Hu等，2021）是一个基于深度特征的负平均关系和算术平均描述符，用于缓解小股人群重识别中的群组关系变化。其中，负平均关系描述符可以用来描述组成员外观上的差异。该方法将由负平均关系描述符获得的特征和由算术平均描述符获得的特征输入到交叉熵损失函数中进行监督学习，然后应用梯度下降算法对目标函数进行优化，得到12个描述群体特征的分类器。

SOT模型（Zhang等，2022b）从单幅图像的群组关系出发，指出已有的方法针对每幅群组图像仅考虑了图像中所提供的群组关系，忽略了对于未知群组关系的建模。基于此，SOT设计了一种基于不确定性建模方法，将每个群组图像看做一个分布而不是固定值，然后通过随机采样挖掘出潜在组。基于潜在的和原始的群组关系，不确定性建模可以更好地学习类别边界。此外，SOT还提出了一种二阶自注意力模型，一种新颖的无卷积深度网络模型Transformer。其动机来自于自注意力模型的位置编码结构与小股人群重识别任务是相契合的。SOT由

成员内模块和成员间模块组成。具体而言，成员内模块提取每个成员的一阶Token特征作为成员表观特征，然后成员间模块通过上述一阶Token提取二阶Token特征，作为群组特征表达。

3DT模型（Zhang等，2022a）考虑了现有的基于成像平面的位置信息是具有歧义的，并不能准确反映成员在真实3维场景下的位置关系。基于此，3DT模型通过3维布局建模克服了上述不足。具体而言，3DT模型是一种基于自注意力模型Transformer的网络结构。对于没有3维位置标注的数据集，3DT首先利用单目深度估计算法成员之间的相对3D布局关系，得到归一化的布局空间，然后应用采样和量化沿3个维度预设一系列布局标记，并根据成员在其中的位置坐标选择相应的标记作为每个成员的布局特征。

3.3.4　基于度量学习的方法

小股行人重识别本质上是一个开集任务，在测试阶段并不看重模型的分类能力，而是特征是否能够表现出显著的类内聚集与类间分散。因此通过度量学习调整特征之间的距离，或者设计更加契合度量的特征空间也是解决小股人群重识别的有效方案。

LIMI（Xiao等，2018）试图通过特征度量的方式来解决群组关系的变动，即处于变动状态下的群组特征或许会不同，但是其相似度应是稳定不变的。基于此，LIMI考虑了多粒度的群组信息，即个人以及群体图像中由2个成员和3个成员组成的子群体。LIMI通过计算多粒度特征来描述多粒度对象的表观和空间特征，并评估每个成员在小股人群识别中的重要性权重，从而缓解群组关系动态变化的干扰。LIMI使用基于多粒度表示和重要性权重的多阶匹配过程来计算最优分组匹配。此外，LIMI根据当前匹配结果动态更新重要性权重，然后计算新的最优分组匹配。这两个步骤是迭代执行的，产生最终的匹配结果。之后，原作者团队又将LIMI进一步拓展成为MGR（Lin等，2021），其中主要是对特征提取的方式进行了完善，但对于特征度量的思想是一致的。

LSRD（Xu等，2019b）提出了一种基于最小二乘算法的最小二乘剩余距离。LSRD可以利用最小二乘法得到的拟合函数的残差，实现学习群组图像对之间的度量。

SVIGR模型（Mei等，2020）在使用孪生网络模型提取成员特征之后，设计基于最小距离的匹配原则，构造群组距离向量并实现了“成员—群组”匹配问题。即衡量两个群组的特征距离时，用两个群组中特征距离最近的两个成员之间的距离作为整体的群组特征距离。

P2GSM-AR（Mei等，2021）考虑在群组成员存在干扰的条件下进行相似度衡量。与SVIGR类似，P2GSM-AR也从单人特征距离的角度出发去衡量群组之间的距离，并在此基础上考虑了更复杂的成员与成员之间的关系，其核心思想是寻找两个群组图像中最相似成员的子集合的距离作为群组距离的计算依据。

3.4　小结

小股行人重识别的算法在数据类型上可分为基于静态图像的方法和基于视频序列的方法；在标签策略上可分为有监督、弱监督和无监督的方法；在模型策略上可分为基于数据增强、基于特征提取和基于度量学习的方法。基于特征提取的方法又可细分为基于手工特征的方法和基于深度学习的方法。现有小股行人重识别算法的技术分类概览如表2所示。

4　代表算法在现有数据集的性能

4.1　性能评估指标

与行人重识别相同，小股人群重识别在模型测试阶段主要通过平均准确率均值（mean average precision，mAP）和前k位命中率（Rank-k）进行衡量。其中mAP和Rank-1是衡量模型性能最关键，也是学术界和工业界最关心的两个性能指标。mAP衡量底库中与待查询图像身份相同的群组图像是否都能尽可能地出现在排序结果的靠前位置的水平；Rank-1（也称首位命中率）衡量待查询图像的首位检索结果是否与其身份相同。

4.2　性能评测结果

现有的前沿代表算法在现有数据集上的模型性能结果统计如表3所示。值得注意的是，并非所有的数据集在提出之后都得到广泛使用。例如，早期的小股人群数据集CASIAGroup和ONLUGroup，后来都鲜有方法用其进行性能评测和消融实验分析。以静态图像为例，公认广泛使用的数据集包括iLIDS-MCTS、RoadGroup、DukeGroup和CSG。

表2　现有方法技术分类概览

Table 2　Abstract of the types of existing methods

方法	文献	数据类型		标签策略			模型策略		
		图像	视频	全监督学习	弱监督学习	无监督学习	数据增强	特征提取	特征度量
CRRRO-BRO	Zheng等人(2009)	√	×	×	×	√	×	√	×
Covariance	Cai等人(2010)	√	×	×	×	√	×	√	×
SBC	Salamon等人(2015)	√	×	×	×	√	×	√	×
BSC+CM	Zhu等人(2016)	√	×	×	×	√	×	√	√
PREF	Lisanti等人(2017)	√	×	×	×	√	×	√	×
LIMI	Xiao等人(2018)	√	×	×	√	×	×	×	√
MGR	Lin等人(2021)	√	×	√	×	×	×	×	√
TSCN	Huang等人(2019b)	√	×	√	×	×	√	×	×
DoTGNN	Huang等人(2019a)	√	×	×	√	×	√	×	×
LSRD	Xu等人(2019b)	√	×	√	×	×	×	×	√
HAM	Xu等人(2019b)	√	×	√	×	×	×	√	×
STFNet	Xu等人(2019a)	×	√	×	√	×	×	√	×
DoTSCN	Huang等人(2021)	√	×	√	×	×	√	×	×
MACG	Yan等人(2020)	√	×	√	×	×	×	√	×
GCGNN	Zhu等人(2021)	√	×	×	√	×	×	√	×
SVIGR	Mei等人(2020)	√	×	×	√	×	×	×	√
TSAN	Chen等人(2021)	×	√	×	√	×	×	√	×
PGSM-AR	Mei等人(2021)	√	×	×	√	×	×	×	√
PRM	Hu等人(2021)	√	×	×	√	×	×	√	×
SFALN	Liu等人(2021)	√	×	×	√	×	√	×	×
SOT	Zhang等人(2022b)	√	×	√	×	×	×	√	×
3DT	Zhang等人(2022a)	√	×	√	×	×	×	√	×

注："√"表示该方法属于对应的类型和策略，"×"表示该方法不属于对应的类型和策略。

表3　不同数据集的前沿代表算法摘录

Table 3　Abstract of the state-of-the-art methods on different datasets

/%

数据集	前沿方法	Rank-1	Rank-5	Rank-10	mAP
iLIDS-MCTS	SVIGR(Mei等,2020)	46.2	71.8	–	42.1
CASIAGroup	Covariance(Cai等,2010)	80.0	–	100.0	–
ONLUGroup	Covariance(Cai等,2010)	60.0	–	100.0	–
ETHZGroup	SBC(Salamon 等,2015)	82.3	–	–	–
ZeCSS	BSC+CM(Zhu等,2016)	36.4	–	–	–
QUEUE02	App-SIM(Koperski等,2017)	55.5	100.0	–	–
MUSEUM	PREF(Lisanti等,2017)	25.8	50.2	–	–
OGRE	PREF(Lisanti等,2017)	15.1	–	41.6	–
DukeGroup	3DT+ (Zhang等,2022a)	90.9	99.9	99.9	94.1
RoadGroup	3DT+ (Zhang等,2022a)	93.8	97.5	98.8	94.8
DukeGroup383	LSRD(Xu 等,2019b)	84.8	91.4	–	76.1
DukeGroupVid	TSAN(Chen等,2021)	82.6	97.9	–	85.6
SYSUGroup	SVIGR(Mei等,2020)	94.8	98.6	99.9	80.2
CSG	3DT+(Zhang等,2022a)	95.1	97.7	98.6	94.4

注："–"表示该方法暂无该指标数据。

一些现有的前沿算法在部分主流数据集上的结果比较如表4所示。可以看到，基于深度学习的方法已经成为解决小股人群重识别的主流趋势，基于手工特征的方法已经日渐式微。在4个主流数据集上，均是基于深度学习的模型实现了最佳性能。基于手工特征的方法在早期的小规模群组数据集中还有所尝试，但在最新的数据集上已经无法与深度学习抗衡。基于手工特征的方法在小数据集上表现尚可，其中一部分原因是因为小数据集中所包含的群组关系、背景干扰都是相对有限的。手工特征设计者可以在设计之初针对这些特点给出相应的解决方案，然而面对大规模数据集时，数据集中的群组变化、背景变化等挑战已经远远超过手工特征算子的表达能力。此时，基于数据驱动的深度学习方法成为了主流研究趋势。这不仅是本任务中才出现的现象，而是在很多视觉任务中都表现出的通用结论。然而，深度学习并非完美，在小数据集上深度学习往往容易过拟合。因此，如何将两者结合才是更重要的研究任务。更进一步地，在基于深度学习的方法中，目前还没有完全统一的框架，基于不同思路的深度学习模型在小股人群重识别任务中都发挥了不错的性能。在iLIDS-MCTS数据集上，SVIGR方法通过卷积神经网络和群组相似度度量的方法达到了当前最好性能；在RoadGroup、DukeGroup和CSG这3个数据集上，基于自注意力模型Transformer和大规模预训练的方法3DT+取得了一致的最好性能。

表4　主流数据集的算法性能评测

Table 4　Performance evaluation of methods for mainstream datasets

/%

方法	文献	CSG数据集				RoadGroup数据集				iLIDS-MCTS数据集			
		Rank1	Rank5	Rank10	mAP	Rank1	Rank5	Rank10	mAP	Rank1	Rank5	Rank10	mAP
CRRRO-BRO	Zheng等人(2009)	10.4	25.8	37.5	–	17.8	34.6	48.1	–	23.3	54	69.8	–
Covariance	Cai等人(2010)	16.5	34.1	47.9	–	38	61	73.1	–	26.5	52.5	66	–
PREF	Lisanti等人(2017)	19.2	36.4	51.8	–	43	68.7	77.9	–	30.6	55.3	67	–
BSC+CM	Zhu等人(2016)	24.6	38.5	55.1	–	58.6	80.6	87.4	–	32	59.1	72.3	–
LIMI	Xiao等人(2018)	–	–	–	–	72.3	90.6	94.1	–	37.9	64.5	79.4	–
DotGNN	Huang等人(2019a)	–	–	–	–	74.1	90.1	92.6	–	–	–	–	–
GCGNN	Zhu等人(2021)	–	–	–	–	81.7	94.3	96.5	–	41.9	68.1	**86.9**	–
SVIGR	Mei等人(2020)	–	–	–	–	87.8	92.7	–	89.2	**46.2**	**71.8**	–	**42.1**
MACG	Yan等人(2020)	63.2	75.4	79.7	–	84.5	95	96.9	–	45.1	70.4	84.9	–
MGR	Lin等人(2021)	57.8	71.6	76.5	–	80.2	93.8	96.3	–	38.8	65.7	82.5	–
SOT	Zhang等人(2022b)	91.7	96.5	97.6	90.7	86.4	96.3	98.8	91.3	–	–	–	–
3DT	Zhang等人(2022a)	92.9	97.3	98.1	92.1	91.4	97.5	98.8	94.3	–	–	–	–
3DT+	Zhang等人(2022a)	**95.1**	**97.7**	**98.6**	**94.4**	**93.8**	**97.5**	**98.8**	**94.8**	–	–	–	–

注：加粗字体表示各列最优结果，“–”表示该指标暂无相关数据。

4.3 小结

本节主要论述了小股人群重识别的模型在数据集上的性能评估指标以及近年来主流方法在主流数据集上的性能评测结果。

5 讨论与展望

5.1 讨论

小股人群重识别是智能监控领域中非常重要的任务之一，是对传统行人重识别任务的新兴拓展方向。随着社会安防的需求升级，这一任务得到了越来越多的关注，产生了相应的研究成果，对于推动该任务的发展起到了积极作用。本文对小股人群重识别的相关研究进行了全面的回顾与介绍，对前沿算法进行讨论，形成观点如下：

1）小股人群重识别的独有研究核心是对于群组结构的建模与描述。这是该任务区别于传统单目标重识别任务的独特之处。通过回顾不难发现，主流的研究都是围绕这一问题展开，早期的手工特征描述子从全图提取群组特征；基于卷积神经网络和图神经网络的方法从特征层面先构造单人特征，并进一步在基础上构造群组特征；后来基于自注意力模型的方法也是沿袭这个研究思路并加以改进。因此，小股人群重识别的方法对于多目标的关系建模是关键的研究内容之一。

2）基于深度学习的方法成为当前研究的主流方向。当前，小股人群重识别技术所面临的挑战更多，其群组结构的变化更加复杂，其场景更加多样性。因此，基于手工特征的方法表达能力有限，非常依赖设计者对于具体场景的特点分析。在面对海量的数据集时，手工描述符往往很难给出精确的描述。然而，基于数据驱动的深度学习则完全发挥了模型优势与大规模数据优势，自发地去挖掘数据内部的关联。然而，这并非完全否定手工特征的作用。在数据集规模较小时，基于深度学习的模型容易出现过拟合的现象，因此手工特征的设计思路可以被深度学习所借鉴，通过正则化的方式减少模型参数的搜索量；在深度学习的模型设计时，研究人员并非全部依赖深度学习自身的拟合能力，而是利用手工特征的思想，在关键环节指导深度模型的学习，降低模型的拟合难度。例如在PRM模型（Hu等，2021）中，利用差分特征这一传统特征提取的思想，深度学习模型获得了更具有判别性的群组特征。在数据集场景较为复杂时，可以通过时序先验信息先对全部数据进行粗筛，然后再用深度学习的方法获得最终结果。综上所述，深度学习已经成为该任务的主流研究方式，并且在此基础上与传统方法相结合将会为这一任务带来更多的促进和启发。

3）基于深度学习的方法没有形成完全统一的处理框架。目前小股人群重识别技术处于快速发展的阶段，从各种不同角度出发的深度学习模型层出不穷，例如基于卷积神经网络的、基于特征度量学习的、基于生成对抗网络的、基于自注意力模型的，等等。然而，在面对一个具体场景时，目前没有形成完全统一且十分完善的解决框架，也没有形成定论哪种方法会更好。需要指出的是，此处所提出的“统一框架”是指小股人群重识别模型在实际应用下还没有形成统一的选择标准，目前仍处在一个百花齐放的快速发展阶段。一般而言，基于生成对抗网络的网络通常用于模型离线训练的数据增强阶段；基于度量学习的方法倾向于仅在训练时使用。在实际场景的使用中，本文倾向于选择仅使用特征提取的网络模型，在此基础上会选择符合使用大规模数据集训练，提取特征维度短，依赖额外先验少，检索速度实时的深度模型，这也是以上各类深度学习的方法可以考虑改进和优化的方向。当然这一倾向不是绝对的，随着技术的发展也会及时更迭。

5.2 展望

随着小股人群重识别的发展，也有一些潜在的问题和发展方向值得注意。

1）小股人群重识别的现有研究对于现实场景下的种种挑战关注过少。例如，行人重识别技术会不可避免地受到昼夜更替的光照变化，因此针对跨模态下的行人重识别受到广泛关注；行人重识别会受到成像分辨率的影响，因此提出了跨分辨率的行人重识别；行人重识别还会受到成员表观遮挡的严重影响，因此面向遮挡场景的行人重识别成为研究子任务之一。与行人重识别的广泛研究相比，小股行人重识别中在面向无监督的研究、面向视频的研究都相对过少；面向跨模态、不同分辨率以及遮挡的研究等几乎依然处于空白的状态。因此，该任务一方面可以采集相应场景下的数据集，给相关研究创造研究基础和研究环境；另一方面，该任务可以更多地从单人行人重识别中借鉴相关的前沿算法，并针对

小股人群重识别任务在不同场景下的独特挑战展开相关的研究。

2)小股人群重识别的现有研究与已有视觉任务的协同和关联不够紧密。小股人群重识别作为一个新兴子方向,应积极探讨与其他视觉任务相结合的可能性,以期通过本任务推动其他任务的性能发展,亦或借鉴其他任务的解决框架来改进本任务的方法流程。例如,是否可以将小股人群重识别与其他群体视觉任务相结合以起到相互协同促进的结果;是否可以将小股人群重识别与其他3维场景理解的方法相结合,以便更好地分析群组的关系等,这些都是值得关注和思考的问题。

3)随着数据隐私政策的增强,小股人群重识别的研究难度加剧。随着全球对于人类隐私数据保护政策的加强,小股人群重识别的研究也受到了一定的限制。DukeMTMC(duke multi-target multi-camera)数据集(Gou等,2017)已经由于隐私政策被禁止使用。在小股人群重识别的发展过程中,有相当一部分的数据集是基于DukeMTMC数据集进一步改造而提出的,使得小股人群重识别任务在部分数据集上存在被禁止研究和使用的风险。针对这一现状,部分研究者已经给出了初步尝试,即在City1M数据集(Zhang等,2022a)中通过图形学生成虚拟数据,以免去对于现场场景隐私的获取。这一思路也启发该任务领域研究如何通过图形学和3D技术产生脱敏数据。同时这也启发本文思考如何建立虚拟数据与真实数据的联系,如何综合两种数据的优势从而进一步推送该任务的发展。首先,真实数据可以指导虚拟数据的质量和语义更加符合现实场景,使得虚拟数据更加具有研究价值;其次,低成本的虚拟数据集可以迁移到真实数据集以丰富现有数据集的表达模式;最后,在当下的元宇宙热潮中,小股人群重识别方法也有着非常广阔的应用前景。因为在元宇宙中也存在着相应的小群体活动,所以小股人群重识别方法对元宇宙中的公众财产和安全保护也会发挥积极正面的作用。此外,一些基于其他模态传感器的方法,可以在规避敏感行人数据的同时,完成对于小股群体的身份识别,在未来也有非常重要的研究价值。

4)小股人群重识别的学术研究和业界需求还未形成良好互动。小股人群重识别目前处于学术界广泛关注、工业界关注不足的状态。这对于小股人群重识别这一面向应用的任务是不友好的。因此,该领域的研究人员应积极推动学术界与工业界的协同合作。学术界通过业界的实际情况来调整该任务所要面对的场景与实际问题,同时业界通过学术界前沿算法找到该任务可以落地的业务场景。通过产学结合,共同推动小股人群重识别任务的发展,加速该任务的落地。更进一步地,工业界可根据最新的前沿算法,个性化设计相应的硬件设备和软件系统模型,实现统一的部署与设备分发,为该任务快速、高效和可信分布式部署提供可能性与保障。

6 结　语

本文对小股人群重识别技术的研究进展进行了全面的梳理回顾,对于该任务的基本概念、数据集、相关技术及未来发展趋势进行了简要总结。在此基础上,对多种小股人群重识别算法进行了详细介绍,并在多个数据集上对前沿算法进行了性能对比。小股人群重识别任务具有重要的研究意义和应用价值,值得学术界和工业界更多的关注和推动。

参考文献(References)

Cai Y H, Takala V and Pietikainen M. 2010. Matching groups of people by covariance descriptor//Proceedings of the 20th International Conference on Pattern Recognition. Istanbul, Turkey: IEEE: 2744-2747 [DOI: 10.1109/ICPR.2010.672]

Chen L, Yang H, Xu Q L and Gao Z Y. 2021. Harmonious attention network for person re-identification via complementarity between groups and individuals. Neurocomputing, 453: 766-776 [DOI: 10.1016/j.neucom.2020.07.118]

Efron B, Hastie T, Johnstone I and Tibshirani R. 2004. Least angle regression. The Annals of Statistics, 32(2): 407-499 [DOI: 10.1214/009053604000000067]

Ess A, Leibe B and Van Gool L. 2007. Depth and appearance for mobile scene analysis//Proceedings of the 11th IEEE International Conference on Computer Vision. Rio de Janeiro, Brazil: IEEE: 1-8 [DOI: 10.1109/ICCV.2007.4409092]

Gheissari N, Sebastian T B and Hartley R. 2006. Person reidentification using spatiotemporal appearance//Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE: 1528-1535 [DOI: 10.1109/CVPR.2006.223]

Gou M R, Karanam S, Liu W Q, Camps O and Radke R J. 2017. DukeMTMC4ReID: a large-scale multi-camera person re-identification

dataset//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, USA: IEEE: 1425-1434 [DOI: 10.1109/CVPRW.2017.185]

Henriques J F, Caseiro R, Martins P and Batista J. 2015. High-speed tracking with kernelized correlation filter. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37 (3): 583-596 [DOI: 10.1109/TPAMI.2014.2345390]

Hu P, Zheng H W and Zheng W S. 2021. Part relational mean model for group re-identification. IEEE Access, 9: 46265-46279 [DOI: 10.1109/ACCESS.2021.3065984]

Huang Z L, Wang Z, Hu W, Lin C W and Satoh S. 2019a. DoT-GNN: domain-transferred graph neural network for group re-identification//Proceedings of the 27th ACM International Conference on Multimedia. Nice, France: ACM: 1888-1896 [DOI: 10.1145/3343031.3351027]

Huang Z L, Wang Z, Hung T Y, Satoh S and Lin C W. 2019b. Group re-identification via transferred representation and adaptive fusion//Proceedings of 2019 IEEE International Conference on Multimedia Big Data. Singapore, Singapore: IEEE: 128-132 [DOI: 10.1109/BigMM.2019.00-34]

Huang Z L, Wang Z, Tsai C C, Satoh S and Lin C W. 2021. DotSCN: group re-identification via domain-transferred single and couple representation learning. IEEE Transactions on Circuits and Systems for Video Technology, 31(7): 2739-2750 [DOI: 10.1109/TCSVT.2020.3031303]

Koperski M, Bak S and Carr P. 2017. Groups re-identification with temporal context//Proceedings of 2017 ACM on International Conference on Multimedia Retrieval. Bucharest, Romania: ACM: 209-217 [DOI: 10.1145/3078971.3078978]

Lin W Y, Li Y X, Xiao H, See J, Zou J N, Xiong H K, Wang J D and Mei T. 2021. Group reidentification with multigrained matching and integration. IEEE Transactions on Cybernetics, 51(3): 1478-1492 [DOI: 10.1109/TCYB.2019.2917713]

Lisanti G, Martinel N, Del Bimbo A and Foresti G. 2017. Group re-identification via unsupervised transfer of sparse features encoding//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 2468-2477 [DOI: 10.1109/ICCV.2017.268]

Lisanti G, Martinel N, Micheloni C, Del Bimbo A and Foresti G L. 2019. From person to group re-identification via unsupervised transfer of sparse features. Image and Vision Computing, 83-84: 29-38 [DOI: 10.1016/j.imavis.2019.02.009]

Liu X H, Yu L S and Lai J H. 2021. Group re-identification based on single feature attention learning network (SFALN)//Proceedings of the 4th Chinese Conference on Pattern Recognition and Computer Vision. Beijing, China: Springer: 554-563 [DOI: 10.1007/978-3-030-88004-0_45]

Mei L, Lai J H, Feng Z X and Xie X H. 2020. From pedestrian to group retrieval via siamese network and correlation. Neurocomputing, 412: 447-460 [DOI: 10.1016/j.neucom.2020.06.055]

Mei L, Lai J H, Feng Z X and Xie X H. 2021. Open-world group retrieval with ambiguity removal: a benchmark//Proceedings of the 25th International Conference on Pattern Recognition. Milan, Italy: IEEE: 584-591 [DOI: 10.1109/ICPR48806.2021.9412734]

Salamon N Z, Junior J C J and Musse S R. 2015. A user-based framework for group re-identification in still images//2015 IEEE International Symposium on Multimedia. Miami, USA: IEEE: 315-318 [DOI: 10.1109/ISM.2015.41]

Sun X X and Zheng L. 2019. Dissecting person re-identification from the viewpoint of viewpoint//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 608-617 [DOI: 10.1109/CVPR.2019.00070]

Wang Y N, Liao S C and Shao L. 2020. Surpassing real-world source training data: random 3D characters for generalizable person re-identification//Proceedings of the 28th ACM International Conference on Multimedia. Seattle, USA: ACM: 3422-3430 [DOI: 10.1145/3394171.3413815]

Xiao H, Lin W Y, Sheng B, Lu K, Yan J C, Wang J D, Ding E R, Zhang Y H and Xiong H K. 2018. Group re-identification: leveraging and integrating multi-grain information//Proceedings of the 26th ACM International Conference on Multimedia. Seoul, Korea (South): ACM: 192-200 [DOI: 10.1145/3240508.3240539]

Xiao T, Li S, Wang B C, Lin L and Wang X G. 2017. Joint detection and identification feature learning for person search//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 3376-3385 [DOI: 10.1109/CVPR.2017.360]

Xu Q L, Yang H and Chen L. 2019a. Spatial-temporal fusion network with residual learning and attention mechanism: a benchmark for video-based group Re-ID//Proceedings of the 2nd Chinese Conference on Pattern Recognition and Computer Vision. Xi'an, China: Springer: 492-504 [DOI: 10.1007/978-3-030-31654-9_42]

Xu Q L, Yang H, Chen L and Zhai G T. 2019b. Group re-identification with hybrid attention model and residual distance//Proceedings of 2019 IEEE International Conference on Image Processing. Taipei, China: IEEE: 1217-1221 [DOI: 10.1109/ICIP.2019.8803758]

Yan Y C, Qin J, Ni B B, Chen J X, Liu L, Zhu F, Zheng W S, Yang X K and Shao L. 2020. Learning multi-attention context graph for group-based re-identification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(6): 7001-7018 [DOI: 10.1109/TPAMI.2020.3032542]

Zhang Q, Dang K H, Lai J H, Feng Z X and Xie X H. 2022a. Modeling 3D layout for group re-identification//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 7502-7510 [DOI: 10.1109/CVPR52688.2022.00736]

Zhang Q, Lai J H, Feng Z X and Xie X H. 2022b. Uncertainty modeling with second-order transformer for group re-identification. Proceed-

ings of the AAAI Conference on Artificial Intelligence, 36(3): 3318-3325[DOI: 10.1609/aaai.v36i3.20241]

Zhang T Y, Xie L X, Wei L H, Zhuang Z J, Zhang Y F, Li B and Tian Q. 2021. UnrealPerson: an adaptive pipeline towards costless person re-identification//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 11501-11510[DOI: 10.1109/CVPR46437.2021.01134]

Zhao C R, Qi D, Dou S G, Tu Y P, Sun T L, Bai S, Jiang X Y, Bai X and Miao D Q. 2021. Key technology for intelligent video surveillance: a review of person re-identification. Scientia Sinica Informationis, 51(12): 1979-2015 (赵才荣, 齐鼎, 窦曙光, 涂远鹏, 孙添力, 柏松, 蒋忻洋, 白翔, 苗夺谦. 2021. 智能视频监控关键技术: 行人再识别研究综述. 中国科学: 信息科学, 51(12): 1979-2015)[DOI: 10.1360/SSI-2021-0211]

Zheng W S, Gong S G and Xiang T. 2009. Associating groups of people//Proceedings of 2019 British Machine Vision Conference. London, UK: BMVA Press: #23[DOI: 10.5244/c.23.23]

Zhu F, Chu Q and Yu N H. 2016. Consistent matching based on boosted salience channels for group re-identification//Proceedings of 2016 IEEE International Conference on Image Processing. Phoenix, USA: IEEE: 4279-4283[DOI: 10.1109/ICIP.2016.7533167]

Zhu J, Yang H, Lin W Y, Liu N, Wang J and Zhang W J. 2021. Group re-identification with group context graph neural networks. IEEE Transactions on Multimedia, 23: 2614-2626[DOI: 10.1109/TMM.2020.3013531]

作者简介

张权，男，博士研究生，主要研究方向为计算机视觉与模式识别。E-mail: zhangq48@mail2.sysu.edu.cn

赖剑煌，通信作者，男，教授，博士生导师，主要研究方向为计算机视觉与模式识别。E-mail: stsljh@mail.sysu.edu.cn

谢晓华，男，副教授，博士生导师，主要研究方向为计算机视觉与模式识别。E-mail: xiexiaoh6@mail.sysu.edu.cn

陈泓栩，男，硕士研究生，主要研究方向为计算机视觉。E-mail: chenhx87@mail2.sysu.edu.cn