

基于特征分层级联网络的目标快速搜索方法

唐浩漾¹, 吴聪¹, 刘洋^{1*}, 翟玉翔¹, 杨东方²

¹西安邮电大学自动化学院, 陕西 西安 710121;

²火箭军工程大学导弹工程学院, 陕西 西安 710025

摘要 图像目标搜索是智能安防、区域监控等领域的一项关键技术。近年来,随着视频数据的爆炸式增长,如何在海量视频数据中进行感兴趣目标的快速搜索成为智能视频处理领域亟待解决的关键问题。本文提出了一种基于底层特征和深度学习特征分层级联的目标快速智能搜索方法。首先,采用目标智能检测技术对视频数据中的感兴趣类型目标进行智能检测;其次,针对数量众多的目标检测结果,设计基于底层特征相关性度量的门控机制,实现对目标框的筛选;最后,利用高维深度学习特征完成对目标的准确识别。实验结果表明,所提方法在 Market1501 数据集上的重识别精度(mAP)达到 88.5%,同时网络推理过程的计算量与 YOLO 和重识别网络直接级联的目标搜索方法相比降低了 81.75%,显著提升了目标搜索速度。

关键词 机器视觉; 目标搜索; 底层特征; 行人重识别; 特征分层级联

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP202259.0215005

Fast Target Search Method Based on Feature Hierarchical Cascade Network

Tang Haoyang¹, Wu Cong¹, Liu Yang^{1*}, Zhai Yuxiang¹, Yang Dongfang²

¹Automation College, Xi'an University of Posts & Telecommunications, Xi'an, Shaanxi 710121, China;

²Missile Engineering College, Rocket Force University of Engineering, Xi'an, Shaanxi 710025, China

Abstract Image target search is a key technology in the fields of intelligent security and regional surveillance. In recent years, with the explosive growth of video data, how to quickly search for objects of interest in massive amounts of video data has become a key issue to be solved in the current intelligent video processing field. A fast and intelligent search method based on the hierarchical cascade of bottom-level features and deep learning features is proposed. First, the intelligent target detection technology is used to intelligently detect the types of targets of interest in the video data. Second, for a large number of target detection results, a gating mechanism based on the bottom-level feature correlation measurement is designed to achieve the screening of target frames. Finally, high-dimensional deep learning features are used to complete the accurate recognition of the target. The experimental results show that the proposed method has a mean average precision (mAP) of 88.5% on the Market1501 dataset. At the same time, the calculation amount of the network reasoning process is reduced by 81.75% compared with YOLO and re-identification network direct cascade target search method, which significantly improves the target search speed.

Key words machine vision; target search; bottom-level feature; person re-identification; feature hierarchical cascade

收稿日期: 2021-02-25; 修回日期: 2021-03-29; 录用日期: 2021-04-14

基金项目: 国家自然科学基金(61673017, 61403398)、国家博士后基金面上项目(2019M3643)、陕西省自然科学基金(2017JM6077)、陕西省重点研发计划(2018ZDXM-GY-039)、西安市科技计划(21RGZN0020)

通信作者: *liuyang@xupt.edu.cn

1 引言

目标搜索旨在利用摄像头实现公共区域目标的准确定位和识别,在视频监控、智能安防、大范围目标搜索、公共安全、边防警戒等军事领域和民用领域得到了广泛关注^[1]。但实际应用过程中,由于待识别目标数量较多,将所有目标送入重识别网络进行匹配关联计算会占用较多的时间,而影响目标搜索的效率。因此,面向海量监控数据源的目标快速搜索技术成为当前国内外研究的热点^[2-4]。

目标搜索是将目标检测^[5-7]和目标重识别^[8-13](Re-ID)相结合的一项关键技术。目标检测的任务是找出图像中感兴趣目标,确定其大小和位置,而目标重识别则是判断给定图像(gallery)中是否存在特定目标(query),归属于图像检索问题。目前典型的目标检测算法如 R-CNN 系列^[14-15]、YOLO^[16]、SSD^[17]等都有广泛的应用^[18-19],而基于深度学习的目标重识别研究主要集中在如何提取有效的鲁棒特征。当前目标重识别特征提取的主要方式可分为全局特征提取和局部特征提取两类。全局特征提取重点在于网络模型的设计和训练技巧的使用,以此提高网络的特征表达能力,如 Ahmed 等^[20]提出了一个新的网络模型来学习图像特征并比较特征图相近区域的特征相似度;Xiao 等^[21]利用卷积神经网络对混合数据集进行训练,提出多域引导退出策略增强特征的鲁棒性,提高了模型的泛化能力;Luo 等^[22]利用随机擦除、标签平滑等有效的训练技巧构造一个只使用全局特征的基准模型,在公共数据集上获得了较好性能。而局部特征提取则关注目标图像中更为重要的局部区域信息,包括对目标图像的分割^[23]以及语义信息^[24]的提取,局部特征提取取得了较高的识别精度,但复杂的网络极大降低了网络推理速度。

相较于直接利用裁剪好的图像作为输入的目

标重识别任务而言,目标搜索更加贴近于实际任务。当前,基于深度学习的目标搜索方法主要可以分为两类:1)对检测网络与重识别网络共享部分特征进行联合训练的单步法(one-stage),如文献^[25]对目标检测与重识别网络共享部分特征进行联合训练,利用卷积神经网络预测候选目标位置后进一步提取重识别特征,实现端到端的目标搜索一体框架。但这种直接共享特征的方法忽略了目标检测关注的是目标之间的共性特征,而目标重识别关注的目标之间的个性特征,导致无法平衡检测与重识别之间的矛盾,因此该方法的检测与重识别的准确率较低。2)单独训练目标检测与目标重识别网络的两步法(two-stage),该方法可以避免两任务之间的竞争而导致的精度问题,但将两个网络直接级联的方式存在计算量大、实时性差等问题,很难应用于人群密集的实际场景中。

为解决当前目标搜索任务存在精度低、计算量大、实时性差等问题,本文提出了一种基于特征分层级联的目标快速搜索方法。与现有两步法搜索方法相比,本文在其重识别网络头部添加底层特征筛选模块,利用底层特征剔除与特定目标差异较大的目标候选框,减少了重识别网络提取深度学习特征计算量,实现了深度学习特征与底层特征的分层互补。实验结果表明,利用底层特征和深度学习特征分层级联的方式在提升目标搜索精度的同时降低了搜索时间,提高了搜索效率。

2 基于特征分层级联网络的目标识别方法

所提方法使用当前性能较好、应用较为广泛的 YOLOv3 目标检测算法和文献^[22]中提出的目标重识别方法架构作为目标搜索框架,在此基础上,构造一个目标底层特征提取与匹配模块,所提方法具体流程如图 1 所示。此小节将从底层特征提取与

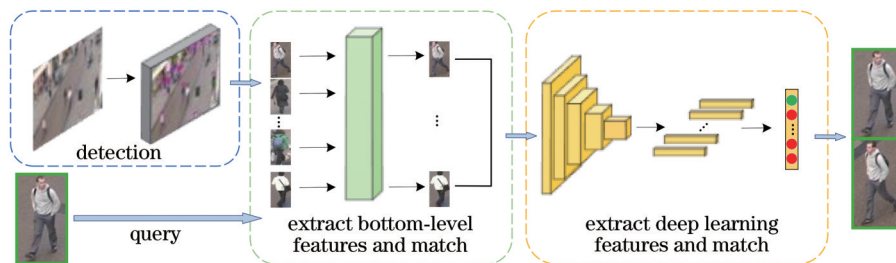


图 1 基于特征分层级联网络的目标快速搜索方法流程图

Fig. 1 Flow chart of fast target search method based on feature hierarchical cascade network

匹配、特征分层级联的目标重识别网络、网络损失函数 3 个方面进行详细介绍。

2.1 底层特征提取与匹配

在广场、车站等人群较为密集的公共区域,将检测网络与重识别模型进行简单级联,无差别地提取所有检测目标的重识别特征会产生巨大计算量,

极大地影响搜索的速度和精度。当待搜索目标(query)与部分待匹配目标(gallery)存在较大差异时,提取这类检测框的重识别特征是不必要的,如图 2 所示。为此,所提方法利用有效的底层特征对待匹配目标进行初步筛选,减少无意义检测框带来的网络计算量和推理时间。



图 2 待搜索目标与待匹配目标图例。(a)待搜索目标;(b)~(e)待匹配目标
Fig. 2 Legends of query and gallery. (a) Query; (b)-(e) galleries

通过目标检测算法对原图中行人目标进行定位,并将检测框送入特征分层级联网络进行筛选识别。由于图像的底层特征主要分布在低维,所提方法利用图像的低维信息,包括颜色均值、方差、斜度来表征目标的底层外观特征。图像 HSV 各通道的均值 E_i 、方差 σ_i 和斜度 s_i 信息作为底层特征时,可以描述为

$$E_i = \frac{1}{N} \sum_{j=1}^N p_{ij}, \quad (1)$$

$$\sigma_i = \left[\frac{1}{N} \sum_{j=1}^N (p_{ij} - E_i)^2 \right]^{\frac{1}{2}}, \quad (2)$$

$$s_i = \left[\frac{1}{N} \sum_{j=1}^N (p_{ij} - E_i)^3 \right]^{\frac{1}{3}}, \quad (3)$$

式中: N 表示图像中像素数量之和; p_{ij} 表示第 i 个颜

色通道上第 j 个像素的像素值; E_i 表示第 i 个颜色通道上所有像素的均值; σ_i 表示第 i 个颜色通道上所有像素的标准差; s_i 表示第 i 个颜色通道上所有像素斜度的立方根。

受目标运动引起的姿态、背景变化等因素干扰,提取目标的全局底层特征进行计算匹配时,存在较大误差,从而导致搜索精度和效率降低。因此,所提方法将大小为 (w, h) 的检测框水平切片为 4 部分,并从中截取宽度为 a 像素的局部区域,即以 4 块大小为 $\left(a, \frac{h}{4}\right)$ 的局部图像块替代全局图像,如图 3 所示。每一部分的底层特征 F_k 可以表示为

$$F_k = [E_H, \sigma_H, s_H, E_S, \sigma_S, s_S, E_V, \sigma_V, s_V], k = 1, 2, 3, 4. \quad (4)$$

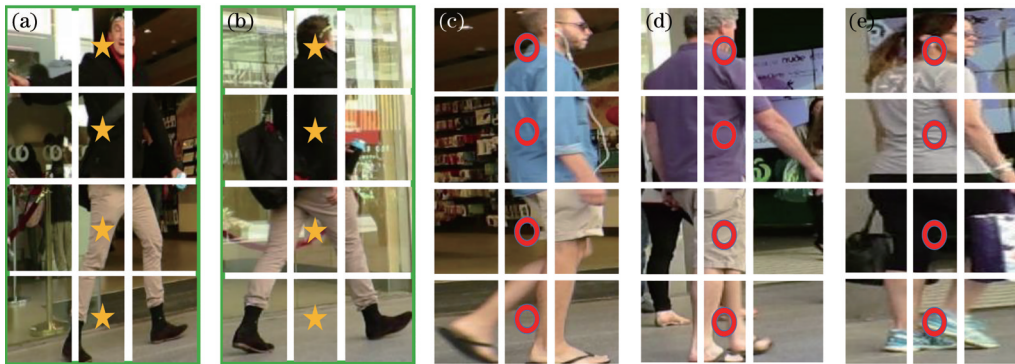


图 3 局部位置底层特征提取图例。(a)待搜索目标;(b)~(e)待匹配目标
Fig. 3 Legends of bottom-level features extraction for local location. (a) Query; (b)-(e) galleries

在度量待匹配目标与待搜索目标之间相似度时,给予目标图像块各部分不同权重使模型关注更具可区分性底层特征的位置。相似度匹配度量表达式为

$$D_k = \left\| \mathbf{F}_k^{(q)}(E_i, \sigma_i, s_i) - \mathbf{F}_k^{(g)}(E_i, \sigma_i, s_i) \right\|_2, \quad (5)$$

$$D = \alpha D_1 + \beta D_2 + \gamma D_3 + \eta D_4, \quad (6)$$

式中: D_k 表示第 k 部分之间的相似度; $\mathbf{F}_k^{(q)}, \mathbf{F}_k^{(g)}$ 分别表

示待搜索目标和待匹配目标第 k 部分的底层特征; $\alpha, \beta, \gamma, \eta$ 表示各部分权重; D 为最终度量匹配结果。在实验中, $\alpha, \beta, \gamma, \eta$ 分别取 0.15, 0.4, 0.4, 0.05。

2.2 特征分层级联的目标重识别网络

所提特征分层级联网络由底层特征提取与匹配模块、深度学习特征提取与匹配模块两部分组成,并设计了网络门控控制网络推理的进程与检测框特征的选择,如图 4 所示。

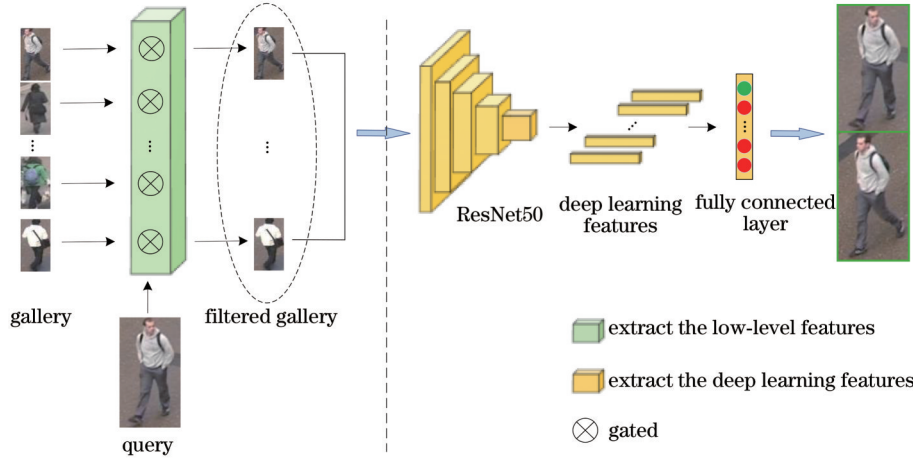


图 4 特征分层级联的目标重识别网络结构图

Fig. 4 Structure diagram of target re-identification network with hierarchical cascade features

在提取目标的底层特征后,对图 3 中的重点局部区域进行加权,加权匹配相似度作为门控筛选的依据。当相似度小于阈值 θ 时,判定为相似目标,门控关闭并进一步提取目标的深度学习特征;否则门控打开,忽略该目标。在深度学习特征提取模块中,以文献[22]的工作为基准,将输入图像大小调整为 256×128 ,采用 ResNet50 作为特征提取的骨干网络,并将最后一个下采样的步长由 2 变为 1,以丰富特征图的细粒度。另外,在特征层与全连接层之间添加批量归一化层(BN)以平衡特征层的各个维度,使得网络在训练阶段更容易收敛的同时保证属于同一个人的特征分布更加紧凑,提高网络预测的准确性。

2.3 特征分层级联的目标重识别网络损失函数

所提分层特征包括底层特征和深度学习特征,但底层特征提取与网络的训练阶段无关,因而网络的损失函数仅用来约束深度学习特征的提取和分类。在训练阶段,使用三元组损失函数^[26]约束高维重识别特征,增强目标的类内紧凑性和类间可分性,可描述为

$$L_{Tri} = \sum_j \left[\left\| \mathbf{f}(x_j^a) - \mathbf{f}(x_j^b) \right\|_2^2 - \left\| \mathbf{f}(x_j^a) - \mathbf{f}(x_j^n) \right\|_2^2 + \varphi \right]_+, \quad (7)$$

式中: x_j^a 表示第 j 个的样本 a ; x_j^b 表示与 a 同一类别的正样本; x_j^n 表示与 a 不同类别的负样本; φ 是一个常量;左边二范数表示与样本 a 同类的类内距离;右边二范数表示与样本 a 不同类别的类间距离。在训练阶段,使用梯度下降法使得损失 L_{Tri} 不断下降,即类内距离不断下降,类间距离不断提升,使得类间距离大于类内距离。

另外,为了优化三元组损失函数不能提供全局最优约束,存在类间距离小于类内距离的问题,将结合身份预测的交叉熵损失函数一同训练模型,使得网络学习到更有区别性的特征。网络总的损失函数为

$$L_{tot} = L_{Tri} + L_{ID}, \quad (8)$$

式中: L_{Tri} 为三元组损失函数损失; L_{ID} 为交叉熵损失。

3 实 验

为了验证所提算法的性能,在公开数据集上对推理速度、计算量及重识别精度(mAP)进行了测试,并与 YOLO 与重识别直接级联的目标搜索方法进行对比分析。

3.1 数据集介绍

目前,主流的目标检测数据集有 PASCAL VOC、ILSVRC、COCO 等,而目标重识别常用的数据集包括 Market1501、DukeMTMC-reid、MSMT17 等。目标重识别数据集中包含由多个摄像机从不同角度拍摄到的行人目标,经手工裁剪并标定出检测框,如图 5 所示。

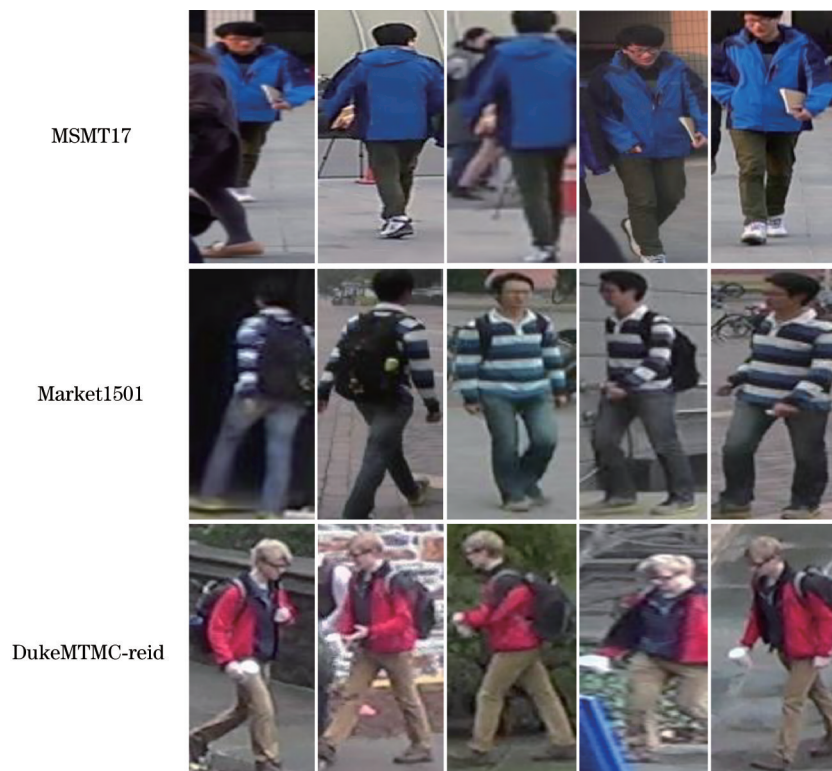


图 5 目标重识别数据集图像实例

Fig. 5 Examples of target re-identification datasets

3.2 主要实验参数说明

对检测模型和目标重识别模型分别进行单独训练,重点考虑目标重识别模型和特征分层级联的有效性。联合训练数据集中目标框大小不一,在训练时将输入图像尺寸调整为 256 pixel \times 128 pixel。模型的初始学习率为 0.0003, batch size 设置为 64, 训练次数为 120。为了保证重识别的精度,所提方法允许目标检测阶段存在部分误检,为此将目标检测的置信度阈值设置为 0.1,非极大值抑制的阈值为 0.4。

3.3 指标说明和实验结果

为了验证所提方法的有效性,从浮点运算次数 (FLOPs) 和 mAP 两个指标上评价模型性能。

FLOPs 实际上是模型中乘法和加法的运算次数,卷积层的浮点运算次数不仅取决于卷积核大小

为了使得网络具有较好的泛化能力,将 3 个目标重识别数据集统一为 Market1501 格式进行联合训练,其中训练集中共有 2493 个不同的行人目标,共计 61479 张图像。训练阶段,所提算法利用随机擦除技术在训练集图片上随机选择一块矩形区域,将该区域的像素替换为随机值,以解决遮挡问题,提高模型鲁棒性。

和输入输出通道数,还取决于特征图的大小。对于一个卷积核大小为 $K \times K$ 的普通卷积层,它的 FLOPs 可以表示为

$$F_{\text{conv}} = 2HW(C_{\text{in}}K^2 + 1)C_{\text{out}}, \quad (9)$$

式中: C_{in} 和 C_{out} 分别为特征图输入通道数和输出通道数; H 和 W 为输出特征图的高度和宽度。

网络模型中,全连接层不存在权值共享,它的 FLOPs 可以表示为

$$F_{\text{FC}} = (2I - 1)O, \quad (10)$$

式中: I 和 O 分别表示为输入神经元数量和输出神经元数量。

在 40 人以上的同一测试目标场景下,利用同一训练条件对比未加入底层特征算法与所提算法的推理速度和 FLOPs,实验结果如表 1 所示。

从表 1 可以看出:在实验数据不变的条件下,

表 1 网络推理速度与 FLOPs 对比

Table 1 Comparison of network speed and FLOPs

Algorithm	Speed / (frame·s ⁻¹)	FLOPs
Baseline	7	183.6
Proposed algorithm	13	33.5

Baseline 对 1500 frame 的测试数据进行推理的速度约为 7 frame·s⁻¹, FLOPs 为 183.6; 所提算法的推理速度为 13 frame·s⁻¹, FLOPs 为 33.5。实验结果表明, 对比未加入底层特征算法, 所提算法的推理速度提高了约 85.7%, 同时网络的计算量减少了 81.75%。

为了验证所设计的底层特征模块对目标搜索算法有效性的影响, 对比未加入底层特征算法与所



图 6 所提算法结果可视化

Fig. 6 Visualization results of proposed algorithm

实验结果表明, 所提算法能够利用加权底层特征有效去除背景冗余信息的干扰, 筛选出与待搜索目标相似检测框, 再利用深度学习特征实现快速、准确的识别, 提高了目标搜索的效率。

4 结 论

提出了一种基于特征分层级联的目标快速搜索方法, 该方法利用底层特征对目标检测框进行预选匹配, 然后使用深度学习模型对目标框进行再识别, 在保证搜索准确率的情况下, 有效地解决了目标搜索在密集人群中实时性差、计算量大的问题。另外, 针对检测框中存在大量的冗余干扰信息导致匹配精度低的问题, 对检测框进行局部加权匹配, 使得模型重点关注目标局部区域特征, 降低背景冗余信息的干扰。实验结果表明, 所提算法能够有效地对目标进行筛选, 提高了目标搜索精度和效率, 具有较强的实际应用价值。

提算法的 mAP, 实验结果如表 2 所示。

从表 2 可以看出, 在实验数据不变的条件下, 未加入底层特征筛选模块的 Baseline 在 Market1501 数据集和 DukeMTMC-reid 数据集的 mAP 分别为 86.1% 和 76.4%, 而所提算法的 mAP 分别为 88.5% 和 78.2%, 与 Baseline 相比分别提高了 2.4 个百分点和 1.8 个百分点。图 6 为所提算法可视化结果, 从图 6 中可以看出, 所提算法能够在复杂场景下实现对目标的准确识别搜索。

表 2 不同算法的 mAP 对比

Table 2 Comparison of mAP of different algorithms unit: %

Algorithm	Market1501	DukeMTMC-reid
Baseline	86.1	76.4
Proposed algorithm	88.5	78.2

参 考 文 献

- [1] Xu Y L, Ma B P, Huang R, et al. Person search in a scene by jointly modeling people commonness and person uniqueness[C]//Proceedings of the 22nd ACM international conference on Multimedia, November 3-7, Orlando, FL, USA. New York: ACM, 2014: 937-940.
- [2] Li S, Xiao T, Li H S, et al. Person search with natural language description[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5187-5196.
- [3] Chen D, Zhang S S, Ouyang W L, et al. Person search via a mask-guided two-stream CNN model [M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 764-781.
- [4] Xiao J M, Xie Y C, Tillo T, et al. IAN: the

- individual aggregation network for person search[J]. *Pattern Recognition*, 2019, 87: 332-340.
- [5] Sharma K U, Thakur N V. A review and an approach for object detection in images[J]. *International Journal of Computational Vision and Robotics*, 2017, 7(1/2): 196.
- [6] Zhao Z Q, Zheng P, Xu S T, et al. Object detection with deep learning: a review[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, 30(11): 3212-3232.
- [7] Luo H L, Chen H K. Survey of object detection based on deep learning[J]. *Acta Electronica Sinica*, 2020, 48(6): 1230-1239.
罗会兰, 陈鸿坤. 基于深度学习的目标检测研究综述[J]. *电子学报*, 2020, 48(6): 1230-1239.
- [8] Saghafi M A, Hussain A, Zaman H B, et al. Review of person re-identification techniques[J]. *IET Computer Vision*, 2014, 8(6): 455-474.
- [9] Chen S Z, Guo C C, Lai J H. Deep ranking for person re-identification via joint representation learning[J]. *IEEE Transactions on Image Processing*, 2016, 25(5): 2353-2367.
- [10] Sun Y F, Zheng L, Yang Y, et al. Beyond part models: person retrieval with refined part pooling (and A strong convolutional baseline)[M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018, 11208: 501-518.
- [11] Luo H, Jiang W, Fan X, et al. A survey on deep learning based person Re-identification[J]. *Acta Automatica Sinica*, 2019, 45(11): 2032-2049.
罗浩, 姜伟, 范星, 等. 基于深度学习的行人重识别研究进展[J]. *自动化学报*, 2019, 45(11): 2032-2049.
- [12] Bi X J, Wang H. Person Re-identification based on view information embedding[J]. *Acta Optica Sinica*, 2019, 39(6): 0615007.
毕晓君, 汪灏. 基于视角信息嵌入的行人重识别[J]. *光学学报*, 2019, 39(6): 0615007.
- [13] Zhang T, Yi Z M, Li X, et al. Improved algorithm for person re-identification based on global features[J]. *Laser & Optoelectronics Progress*, 2020, 57(24): 241503.
张涛, 易争明, 李璇, 等. 一种基于全局特征的行人重识别改进算法[J]. *激光与光电子学进展*, 2020, 57(24): 241503.
- [14] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1440-1448.
- [15] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [16] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [17] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9905: 21-37.
- [18] Liu X, Chen J, Yang D F, et al. Scene-coupled intelligent multi-task detection algorithm for air-to-ground remote sensing image[J]. *Acta Optica Sinica*, 2018, 38(12): 1215008.
刘星, 陈坚, 杨东方, 等. 场景耦合的空对地多任务遥感影像智能检测算法[J]. *光学学报*, 2018, 38(12): 1215008.
- [19] Li Z Q, Zhu R F, Ma J Y, et al. Airport detection method combined with continuous learning of residual-based network on remote sensing image[J]. *Acta Optica Sinica*, 2020, 40(16): 1628005.
李竺强, 朱瑞飞, 马经宇, 等. 联合连续学习的残差网络遥感影像机场目标检测方法[J]. *光学学报*, 2020, 40(16): 1628005.
- [20] Ahmed E, Jones M, Marks T K. An improved deep learning architecture for person re-identification[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 3908-3916.
- [21] Xiao T, Li H S, Ouyang W L, et al. Learning deep feature representations with domain guided dropout for person re-identification[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1249-1258.
- [22] Luo H, Jiang W, Gu Y Z, et al. A strong baseline and batch normalization neck for deep person Re-identification[J]. *IEEE Transactions on Multimedia*, 2020, 22(10): 2597-2609.
- [23] Zhang X, Luo H, Fan X, et al. AlignedReID: surpassing human-level performance in person re-identification[EB/OL]. (2017-11-22)[2021-01-15].

<https://arxiv.org/abs/1711.08184>.

- [24] Zhao H Y, Tian M Q, Sun S Y, et al. Spindle net: person re-identification with human body region guided feature decomposition and fusion[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 907-915.
- [25] Xiao T, Li S, Wang B C, et al. Joint detection and identification feature learning for person search[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 3376-3385.
- [26] Schroff F, Kalenichenko D, Philbin J. FaceNet: a unified embedding for face recognition and clustering[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 815-823.