

DOI:10.19651/j.cnki.emt.2005576

基于分组特征赋权的动态视角图像特征融合*

张佳琦¹ 张金芝¹ 楼亮亮²

(1. 上海大学 特种光纤与光接入网重点实验室特种光纤与先进通信国际合作联合实验室 上海 200444;
2. 中国科学院 上海微系统与信息技术研究所 无线传感网与通信重点实验室 上海 200050)

摘要: 为了解决移动机器人在目标识别过程中捕获的图像存在多目标干扰和单一视角下目标特征有限,进而导致识别准确率低的问题,提出一种基于分组特征赋权的动态视角图像特征融合方法,该方法通过递进式 K 均值聚类,对多目标特征进行赋权分组,并且利用 LSTM 网络实现动态视角下连续图像特征的融合,从而达到提高目标识别的准确率的目的。验证结果表明,在 Market-1501 数据集上的首位识别率达到了 93.80%,平均准确率达到 89.13%,具有较好的实验效果。

关键词: 分组特征赋权;特征融合;特征提取;目标识别;长短期神经网络

中图分类号: TP391 文献标识码: A 国家标准学科分类代码: 510.1050

Feature fusion of dynamic visual angle images based on grouping feature weight

Zhang Jiaqi¹ Zhang Jinyi¹ Lou Liangliang²

(1. Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, Shanghai University, Shanghai 200444, China;
2. Key Laboratory of Wireless Sensor Network & Communication, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China)

Abstract: In order to solve the problem of low recognition accuracy, the image captured by mobile robot in the process of target recognition has multi-target interference and the target feature is limited in a single perspective. In this paper, a method of feature fusion of dynamic visual angle images based on grouping feature weight weights is proposed. In this method, multiple target features are weighted and grouped by progressive K-means clustering, and continuous image features are fused from dynamic perspective by LSTM network, so as to improve the accuracy of target recognition. The verification results show that the first recognition rate on Market-1501 data set reaches 93.80%, and the average accuracy reaches 89.13%, with good experimental results.

Keywords: grouping feature weight; feature fusion; feature extraction; target recognition; long short term memory network

0 引言

近年来,目标识别成为机器人学领域一个重要的研究课题^[1-2],对于工作在人居环境的移动机器人来说,单视角下的特征对目标描述不准确、不充分,并且在识别过程中,同一图像中往往存在多目标相互干扰,影响了目标识别的准确率^[3]。针对这些问题,现有的解决方案有:陈鸿昶等^[4]提出基于注意力模型的目标识别神经网络模型,通过在特征提取阶段增添注意力机制,提取显著性目标特征,进而在

特征识别阶段,依据目标特征显著性进行分级识别;Luo等^[5]提出动态匹配目标图像方法,通过全局特征和局部特征的共同的学习,提高特征匹配的准确率;张继等^[6]将特征表达、相似性度量和目标匹配重排序相结合,通过奇异值分解降低相似目标特征之间的相关性,优化深度学习过程,并且利用改进的 Softmax 损失函数使相似目标之间拥有更大的类间距离和更小的类内距离。

上述的这些方案从增强目标特征的代表能力,优化识别效果的角度出发,力图在多目标中识别出正确的目标。

收稿日期:2020-12-21

* 基金项目:高等学校学科创新引智计划(111)资助(D20031)、上海市教委重点学科资助项目(J50104)资助

• 144 •

但是,存在这样两个问题:1)只对单个目标的特征进行基于注意力机制提取,赋予单个目标不同特征相应的权值信息,缺乏对相似目标特征的处理,导致相似目标发生干扰,不利于提高首位准确率和平均准确率;2)考虑到相似目标特征类间距离和类内距离,但是对单个目标特征未进行权值分配,导致一些不重要特征影响整体特征的效果,进而也影响首位准确率和平均准确率。基于以上分析,本文提出基于分组特征赋权的动态视角图像特征融合研究,相关研究从多目标分组特征赋权和动态视角下连续图像特征融合两个角度展开。前者通过递进式K均值聚类,对多目标特征进行分组,得到多组质心,并用质心特征替代全组多目标特征进行相似度判断,从而完成分组特征赋权,降低了目标识别的复杂度,提高了首位识别率;后者利用长短期神经网络(long short term memory networks, LSTM)网络实现连续图像特征拼接,通过记忆门和遗忘门完成目标特征选择与融合,解决目标平均识别准确率低的问题。验证结果表明,本文方法在 Market-1501 数据集上的首位识别率达到了 93.80%,平均准确率达到 89.13%,具有较好的实验效果。

1 多目标分组特征赋权

移动机器人在目标识别的过程中,捕获的图像可能存在多目标相互干扰^[7-9],为了减少干扰,识别出正确的目标,本文采用多目标特征提取与递进式K均值聚类分组与特征赋权相结合的方法。

1.1 目标特征提取

卷积神经网络(convolutional neural networks, CNN)

是深度学习中提取图像特征的主要方法,其能直接对图像进行处理,相较于传统的模式特征提取方法,提取到的特征具有更好的鲁棒性^[10-12]。本文借鉴 CNN 网络结构,同时考虑到池化层为了从卷积激活层输出中提取出更加有效的特征,会对卷积激活层输出特征进行缩放,去除冗余特征和不相关的特征,但是也会导致有效特征的丢失,为了能够获得目标更加全面的特征信息,图1中采用3个级联的卷积激活层,尽可能充分地提取出目标的特征信息,池化层部分采用平均池化,相较于其他池化方法能够更好地筛选较优特征^[13-14]。 t 时刻,当输入的图像中存在 n 个目标 $\{T_i \mid i=0,1,2,\dots,n\}$ 卷积后得到的特征分别为 f_i^1, f_i^2 和 f_i^3 ,如式(1)~(3)所示,经池化后得到的特征如式(4)所示。

$$f_i^1 = A(C_1 \cdot T_i + b_1) \tag{1}$$

$$f_i^2 = A(C_2 \cdot f_i^1 + b_2) \tag{2}$$

$$f_i^3 = A(C_3 \cdot f_i^2 + b_3) \tag{3}$$

$$V_i = Average(f_i^3) \tag{4}$$

式中: C_1, C_2 和 C_3 表示卷积核的大小; b_1, b_2 和 b_3 为偏置项; A 为激励函数。为了提高梯度下降的速度,防止梯度消失问题,本文采用线性整流函数(rectified linear unit, ReLU)作为激励函数。经池化后的特征 $V_i = [x_1, x_2, \dots, x_m]^T$ 经过全连接层后得到的最终输出为 $L_i = CV_i + d = [x_1, x_2, \dots, x_k]^T$,是 $m \times k$ 的矩阵。若输入存在 n 个相似目标,则输出 $L = [L_1, L_2, \dots, L_n]^T$ ($n \in N, N$ 表示自然数),即得到的多目标特征向量矩阵。

1.2 递进式K均值聚类分组与特征赋权

图像中 n 个相似目标平行经过特征提取网络,得到 $L = [L_1, L_2, \dots, L_n]^T$,但是经 CNN 网络提取的特征维数

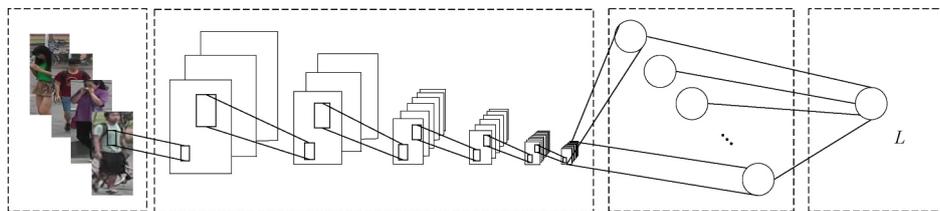


图1 CNN 特征提取图

较高^[15-16], n 个相似目标依次与正确目标进行相似性比较会导致许多冗余的计算,造成计算资源的浪费。如果特征 L 中包含 n 个目标特征,则第 c 个目标特征可以表示为 \vec{l}_c , 每一个 \vec{l}_c 中包含 m 个特征,则 L 为 $n \times m$ 的矩阵,记为 $L_{n \times m}$,第 i 行第 j 列可以表示为 l_{ij} 。本文提出的递进式K均值聚类算法,采用二分组的思路,分为质心收敛的两组 β_1 和 β_2 ,并以每组的均值作为该组的质心 \vec{l}_{mod} ,质心的计算公式如式(5)所示,之后递进式进行上述过程,直至找到与正确目标相似度最高的目标 l_{max} 。

$$\vec{l}_{mod}(\beta_i) = \frac{1}{len(\beta_i)} \sum_{l_i \in \beta_i} l_i \tag{5}$$

递进式K均值聚类每次分为质心收敛的两组 β_1 和 β_2 , β_1 和 β_2 组内应该满足两两元素之间尽可能的相似,这里使用欧几里得距离进行相似性衡量^[17],如式(6)所示。

$$d(\vec{l}_x, \vec{l}_y) = \sqrt{\sum_{z=0}^m (l_{x,z} - l_{y,z})^2} \tag{6}$$

如果 $d(\vec{l}_{mod1}, \vec{l}_{real}) > d(\vec{l}_{mod2}, \vec{l}_{real})$,则 β_1 组的权重 w_1 ,大于 β_2 组的权重 w_2 ,反之则 β_1 组的权重 w_1 大于 β_2 组的权重 w_2 ,完成特征赋权的过程,之后对权重较大的组进行递进式K均值聚类,重复上面的过程,每次递进式找出下一次分组结果,直到组内只剩唯一一个目标 l_{max} ,即与正确目标相似度最高的目标,若其与真实值的相似度大于

阈值 ϵ , 则结束上述过程, 否则依次回转上述流程, 直至找到满足阈值条件的目标 $l_{\max o}$ 。

2 动态视角下连续图像特征融合

在得到与正确目标相似度最高的目标 l_{\max} 后, 考虑到 l_{\max} 是与其他相似目标相对比较得到的, 并不能证明 l_{\max} 与正确目标是完全相同的, 可能因为目标的某些属性如性别、头发、年龄、穿着等因素导致的误匹配, 因此, 为了进一步提高目标的判决正确率, 需要对目标动态连续捕获多幅图像, 并利用同一图像不同视角下特征的互补性进行自适应融合, 本文中融合网络架构采用 LSTM 网络, 网络模型如图 2 所示。该网络是递归神经网络 (recurrent neural network, RNN) 的衍生, 它能对同一目标的连续多幅图像

进行空间相关性建模, 即依据前 $n-1$ 幅图像学习第 n 幅图像特征表示, 并通过记忆门和遗忘门完成目标特征选择与融合, 获取更多的鉴别性特征, 如图 3 所示。

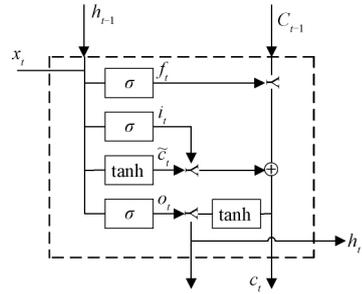


图 2 LSTM 网络模型

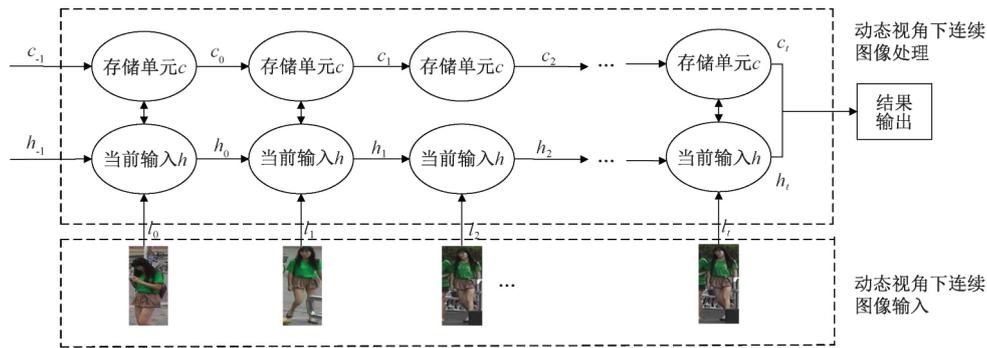


图 3 连续图像特征融合

LSTM 是一种可以自适应可变长度输入的神经网络^[18-19], 它主要由存储单元 c_t 、遗忘门 f_t 、输入门 i_t 和输出门 o_t 组成。存储单元用于存储第 n 帧的特征信息; 输入门用于输入新的特征信息, 并为融合前后时间帧中的特征信息注入新的特征; 遗忘门用于前一时间帧作用下当前时间帧需要遗忘的特征信息; 输出门保存并输出经过 LSTM 网络特征自适应融合后更新的特征信息。

当第 n 帧中输入的特征为 l^* 时, 存储单元 c_t 、遗忘门 f_t 、输入门 i_t 和输出门 o_t 依次将进行更新, 更新的公式为:

$$\begin{cases} f_t = \sigma(W_f[h_{t-1}, l^*] + b_f) \\ i_t = \sigma(W_i[h_{t-1}, l^*] + b_i) \\ \tilde{c}_t = \tanh(W_c[h_{t-1}, l^*] + b_c) \\ o_t = \sigma(W_o[h_{t-1}, l^*] + b_o) \\ h_t = o_t * \tanh(c_t) \end{cases} \quad (7)$$

式中: h_t 为 LSTM 网络模型的最终输出; $\sigma(\cdot)$ 是 sigmoid 激活函数, 它是一个值在 $[0, 1]$ 之间的一个连续单调的阈值函数; $\tanh(\cdot)$ 是双曲正切激活函数, 具有迭代速度快, 收敛时间短的优点; W_f, W_i, W_c 和 W_o 依次为遗忘门、输入门、存储单元和输出门的权重矩阵; b_f, b_i, b_c 和 b_o 依次为遗忘门、输入门、存储单元和输出门的偏置参数。

经过式(7)对动态视角下连续图像特征进行更新融合, 去除冗余和不相关特征, 有利于提高目标识别的平均准确率。

3 验证与分析

3.1 数据集与验证环境配置

本文采用清华大学收集的 Market-1501 数据集, 该数据集已经划分训练集和测试集, 共 1 501 个目标, 32 668 张图像, 其中训练集中包含 751 个目标, 每个目标平均有 17 张不同视角的图像, 共 12 936 个图像, 测试集中包含 750 个目标, 每个目标平均有 26 张不同视角的图像, 共 19 732 个图像。

本文硬件平台是基于 Windows10 操作系统, 显卡为 NVIDIA 1050Ti, 软件平台基于 Pycharm 实现, 采用 pytorch1.4.0 神经网络框架构建, python 版本为 3.7。

在训练过程中, 将目标图像调整为 256×128 输入网络, 训练批尺寸为 16, 每次训练采用随机采样方式, 学习率为 0.01, 并采用随机梯度下降法优化网络。

3.2 验证结果分析

为了验证本文提出的基于分组特征赋权的动态视角图像特征融合的方法的有效性, 本文将从两个方面展开: 1) 验证多目标分组特征赋权的效果, 当输入图像中包含多个目标时, 观察首位识别率的指标; 2) 验证动态视角连续图像融合对目标识别的影响, 将单视角图像输入和多视角图像输入对结果的影响进行比较, 验证平均准确率。首

先,为验证多目标分组特征赋权的效果,将多个相似目标的图像作为输入,经过特征提取网络,进行特征提取,所有图像特征组成集合 L ,之后将 L 进行递进式K均值聚类分组与特征赋权算法进行运算,找到与正确目标相似度最高的目标 l_{max} ,这可用首位识别率(Rank-1)指标进行衡量,Rank- n 表示特征相似度最靠前的 n 张图有正确结果的概率。同时为了与文献[4]、文献[5]和文献[6]进行更好的比较,这里增加指标平均精确均值(mean average precision, mAP),它代表每一个目标类别的平均精度的均值,它的表达式如式(8)所示。

$$mAP = \frac{\sum_{k=0}^{k=C} \sum P_k}{C} \quad (8)$$

式中: $\sum P_k$ 表示第 k 个目标类别精确度的值; I_C 表示第 c 个类别的目标数量。具体比较结构可以如表1所示,从表1中可以看出,Rank-1和mAP两个指标综合来看本文方法具有明显的优势,Rank-1虽然与文献[4]不相上下,但是mAP高于文献[5]和文献[6]至少4%,mAP虽然比文献[5]略低,但是通过下一步动态视角连续图像融合,能够进一步提高mAP指标。

表1 比较结果

	Rank-1/%	mAP/%
目标属性分级识别 ^[4]	93.10	76.20
动态特征匹配 ^[5]	89.40	92.80
奇异值分解与目标对齐 ^[6]	83.41	65.09
本文方法	93.80	89.13

其次,为验证动态视角连续图像融合对目标识别的影响,将同一目标的多视角图像按照时间序列依次输入,并依据式(7)对输入特征进行动态视角下图像特征融合。验证结果如图4所示,其中左边是动态视角下捕获的连续图像,右边是在测试集中找到的TOP-6高匹配图像用黑色实



(a) 动态视角下捕获的图像

(b) 识别出的TOP-6相似图像

图4 验证结果

线框包裹的图像表示与输入是相同的目标,黑色实线框之外的则是不同的目标。从图4中可以看出,随着获取同一目标多视角图像越多,目标mAP越高,可以达到89.13%。

4 结 论

本文提出的基于分组特征赋权的动态视角图像特征融合方法,分为多目标分组特征赋权和动态视角下连续图像特征融合两个部分,前者通过递进式K均值聚类,对多目标特征进行分组,得到多组质心,并用质心特征替代全组多目标特征进行相似度判断,从而完成分组特征赋权,降低了目标识别的复杂度;后者利用LSTM网络实现连续图像特征拼接,通过记忆门和遗忘门完成目标特征选择与融合。

需要注意的是,本文处理的图像为清晰的目标多视角图像,并未考虑遮挡、目标快速运动、小目标图像等极端条件,为了增强目标识别的适应性和鲁棒性,未来可以从极端条件建模和对目标更加精细化的处理,进一步将目标识别技术落地,适应各种复杂的情况。

参考文献

- [1] HUANG T, CHEN Y F, YAO B J, et al. Adversarial attacks on deep-learning-based radar range profile target recognition[J]. Information Sciences, 2020, 531(8): 159-176.
- [2] 储泽楠, 王伟, 蔡胜宇. 基于OpenMV的视觉识别系统在六足机器人中的应用[J]. 河南科技, 2019(34): 26-29.
- [3] 韩玉洁, 曹杰, 刘琨, 等. 基于改进YOLO的无人机对地多目标检测[J]. 电子测量技术, 2020, 43(21): 19-24.
- [4] 陈鸿昶, 吴彦丞, 李邵梅, 等. 基于行人属性分级识别的行人再识别[J]. 电子与信息学报, 2019, 41(9): 2239-2246.
- [5] LUO H, JIANG W, ZHANG X, et al. Alignedreid++: Dynamically matching local information for person reidentification[J]. Pattern Recognition, 2019, 94(5): 53-61.
- [6] 张继, 金翠, 王洪元, 等. 基于奇异值分解行人对齐网络的行人重识别[J]. 山东大学学报(工学版), 2019, 49(5): 91-97.
- [7] KOBER V I, RUCHAY A N, KARNAUKHOV V N. Tracking multiple objects with locally adaptive generalized optimum correlation filters[J]. Journal of Communications Technology and Electronics, 2020, 65(6): 16-724.
- [8] KE B, ZHENG H C, CHEN L R, et al. Multi-object tracking by joint detection and identification learning[J]. Neural Processing Letters, 2019, 50(1): 283-296.
- [9] 孟庆鑫, 孟庆磊, 闫帅. 基于时空信息的运动目标识别算

- 法[J]. 电子测量与仪器学报, 2019, 33(2): 154-160.
- [10] WONG F K, HU H F. Adaptive learning feature pyramid for object detection [J]. IET Computer Vision, 2019, 13(8): 742-748.
- [11] ZHI S F, LIU Y X, LI X, et al. Toward real-time 3D object recognition: A lightweight volumetric CNN framework using multitask learning[J]. Computers & Graphics, 2018, 71(10): 199-207.
- [12] WANG J C, LI G Y. Accelerate proposal generation in R-CNN methods for fast pedestrian extraction[J]. The Electronic Library, 2019, 37(3): 435-453.
- [13] ABOOZAR T, GEORGINA C, MCGINNITY T M. AdaBoost-CNN: An adaptive boosting algorithm for convolutional neural networks to classify multi-class imbalanced datasets using transfer learning [J]. Neurocomputing, 2020, 404(3): 351-366.
- [14] 张婷, 钱丽萍, 汪立东, 等. 基于多层卷积模型的恶意 URL 特征自动提取[J]. 计算机工程与设计, 2020, 41(7): 1821-1828.
- [15] MCLAUGHLIN N, RINCON J M D, MILLER P. Recurrent convolutional network for video-based person re-identification [C]. Computer Vision and Pattern Recognition (CVPR) Las Vegas, NV: Institute of Electrical and Electronics Engineers(IEEE), 2016.
- [16] SARANYA R, POONGODI C, DEVARAJ S, et al. Novel deep learning model for facial expression recognition based on maximum boosted CNN and LSTM[J]. IET Image Processing, 2020, 14(7): 1373-1381.
- [17] 臧艳辉, 席运江, 赵雪章. 基于 MapReduce 的分治 k 均值聚类方法[J]. 计算机工程与设计, 2020, 41(5): 1345-1351.
- [18] 张金艺, 秦政, 林羽晨, 等. 面向伴随型机器人同步运动的递进式步态时相检测算法[J]. 仪器仪表学报, 2020, 41(1): 113-120.
- [19] HOU Q Q, WANG J J, BAI R B, et al. Face alignment recurrent network[J]. Pattern Recognition, 2018(74): 448-458.

作者简介

张佳琦, 硕士研究生, 主要研究方向为图像处理、特征融合。

E-mail: zjq15366322672@163.com

张金艺, 硕士生导师, 主要研究方向为深度学习、图像处理、伴随型机器人研究。

E-mail: zhangjinyi@shu.edu.cn

楼亮亮, 博士, 主要研究方向为物联网感知技术、无线传感器网络。

E-mail: lou0908@126.com