

文章编号: 2095-2163(2020)07-0203-03

中图分类号: TP18; TP391.41

文献标志码: A

# 基于 SDAE 深度学习的多目标检测与跟踪研究

左国才<sup>1</sup>, 苏秀芝<sup>1</sup>, 陈明丽<sup>2</sup>, 张 珏<sup>3</sup>, 吴小平<sup>1</sup>

(1 湖南软件职业学院 软件与信息工程学院, 湖南 湘潭 411100 ;2 新化县楚怡工业学校, 湖南 新化 417600;

3 湖南铁道职业技术学院 轨道交通电务技术学院, 湖南 株洲 412001)

**摘要:** 运动目标的检测与跟踪是智能交通、智能监控等领域的重要组成部分, 尤其是行人、车辆的检测与跟踪, 对于行人行车安全、流量监控等都有着重要的意义。由于复杂交通场景中多目标检测与跟踪的难度较高, 人工设计的目标特征无法满足复杂的多目标检测与跟踪。因此, 本文提出一种基于堆栈式去噪自编码器深度学习框架的多目标检测与跟踪算法, 利用海量的图片数据集训练深度学习模型, 提取目标深度特征, 用于多目标检测与跟踪。实验结果表明, 基于堆栈式去噪自编码器深度学习框架的多目标检测与跟踪算法, 提高了多目标检测的准确性, 实现了更加鲁棒的多目标跟踪效果。

**关键词:** 深度学习; 堆栈式去噪自编码器; 多目标检测与跟踪

## Research on multi-target detection and tracking based on SDAE deep learning

ZUO Guocai<sup>1</sup>, SU Xiuzhi<sup>1</sup>, CHEN Mingli<sup>2</sup>, ZHANG Jue<sup>3</sup>, WU Xiaoping<sup>1</sup>

(1 School of Software and Information Engineering, Hunan Software Vocational Institute, Xiangtan Hunan 411100, China;

2 Xinhua Chuyi Industrial School, Xinhua Hunan 417600, China; 3 Rail Transit and Communication Technology College,

Hunan Railway Vocational and Technical College, Zhuzhou Hunan 412001, China)

**[Abstract]** Moving target detection and tracking is an important part of intelligent transportation, intelligent monitoring and other fields. Especially the detection and tracking of pedestrians and vehicles are of great significance for pedestrian traffic safety and traffic monitoring. Due to the difficulty of multi-target detection and tracking in complex traffic scenes, the target features designed manually cannot meet the requirements of complex multi-target detection and tracking. Therefore, this paper proposes a multi-target detection and tracking algorithm based on stack de-noising self-encoder depth learning framework. Using the massive picture data set to train the depth learning model to extract the depth feature of the target, which is used in the multi-target detection and tracking. The experimental results show that the multi-target detection and tracking algorithm based on the stack de-noising self-encoder depth learning framework improves the accuracy of the multi-target detection and achieves a more robust multi-target tracking effect.

**[Key words]** Deep Learning, Stacked Denoising Autoencoders; multi-target detection and tracking

## 0 引言

多目标检测与跟踪是目前计算机视觉领域中最火热、最具挑战性的方向之一<sup>[1]</sup>。智能交通研究在国内外起步较早, 很多专家进行了诸多有价值的研究。例如: 机动车辆车牌检测识别、机动车辆流量统计、机动车辆车速测算等。由于当时技术的局限性, 智能交通技术并没有真正应用于实现生活中, 实用价值较低。但是智能交通的重要性得到了大家一致认可, 投入大量的资金和人力开发智能交通系统<sup>[2]</sup>。

Wax 最早提出了多目标检测跟踪。在多目标检测跟踪模型的设计中, 为了解决多目标检测跟踪过程中存在的难题, 提出了很多相关的算法。如: 基于多假设跟踪算法(MHT)和基于粒子滤波的跟踪框架。由于深度学习在图像分类方面取得的显著效

果, 提出了基于深度学习的单目标跟踪算法。其中, 自编码器在图像识别、语音识别、自然语言处理等领域已经广泛应用。基于深度学习框架的深度特征在多目标检测与跟踪上的应用研究具有理论和实践意义。由于传统人工设计目标特征无法满足复杂交通场景中的行人多目标检测与跟踪, 因此本文采用堆栈式去噪自编码器(SDAE)深度学习框架提取目标深度特征, 设计基于深度学习的多目标检测与跟踪模型。

## 1 SDAE 相关理论

Vincent 在 2010 年提出堆栈式去噪自编码器(Stacked Denoising Autoencoders, SDAE)算法。堆栈式去噪自编码器算法在相关改进后, 算法处理的精度更高, 具有更加强大的抗击噪声的能力。SDAE 深度

**基金项目:** 湖南省自然科学基金(2020JJ7007); 湘潭市科技局指导性计划项目(ZDX-CG2019013)。

**作者简介:** 左国才(1978-), 女, 硕士, 教授, 高级工程师, 主要研究方向: 机器视觉、深度学习。

**收稿日期:** 2020-06-01

学习网络模型通过对运动目标的初始特征进行抽象,可以获得更加深层次的目标特征,即:深度特征。

使用堆栈式去噪自编码器构建基于 SDAE 深度学习框架的多目标检测与跟踪模型,可以采取以下方式:将堆栈式去噪自编码器最后的输出层连接到回归层,将堆栈式去噪自编码器逐层进行无监督的训练,上一层网络提取到的深度高层特征的输出作为下一层网络的输入,完成网络参数的初始化操作,之后使用带标签数据进行有监督训练网络模型,参数微调,将堆栈式去噪自编码器中最后层网络模型的输出作为网络模型回归层的输入<sup>[3]</sup>。本文设计的基于 SDAE 深度学习网络模型采用这种方式。

SDAE 的思想就是将多个 DAE 堆叠在一起形成一个深度的架构。只有在训练的时候才会对输入加噪,一旦训练完成,就不需要再加噪,SDAE 网络模型结构<sup>[3]</sup>如图 1 所示。

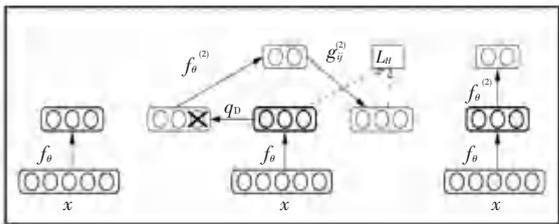


图 1 SDAE 网络模型结构

Fig. 1 SDAE network model structure

## 2 基于 SDAE 深度学习框架的多目标检测与跟踪算法

### 2.1 基于 SDAE 模型提取深度特征

得到预训练模型后,经过参数微调可以得到目标高层次深度特征。在完成无监督的网络模型预训练后,在堆栈式去噪自编码器最后一层加上回归层,将带有标签数据训练网络模型对 SDAE 深度学习网络模型微调参数,得到 SDAE 深度学习网络模型。

本文采用堆栈式去噪自编码器网络模型来提取图像高层抽象特征。离线进行无监督训练 SDAE 深度学习网络模型,使得 SDAE 模型能够有效地提取到图像高层抽象特征。

通过加入噪声,经过编码,如式(1)所示:

$$c_i = f(Wx_i' + b). \quad (1)$$

再经过解码,重构原始样本,如式(2)所示:

$$x_i = f(W'c_i + b'). \quad (2)$$

其损失函数如式(3)所示:

$$\min_{W, W', b, b'} \sum_{i=1}^k \|x_i - x_i'\|_2^2 + \lambda (\|W\|_F^2 + \|W'\|_F^2). \quad (3)$$

### 2.2 多目标检测

在基于深度学习框架 SDAE 网络模型的顶层增

加分类器 softmax,将带有标签的数据有监督训练基于 SDAE 深度学习网络模型,对网络参数微调,使得 SDAE 深度学习模型可以完成运动目标分类。

多目标检测算法流程:

基于深度学习框架 SDAE 网络模型对多运动目标检测的步骤如下:

(1)原始的训练数据集数据完成相关预处理。

(2)初始化 SDAE 深度学习网络参数,从训练数据中选取已完成修正的参数作为 SDAE 网络模型的初始输入。

(3)将数据集中的数据分为训练集和测试集,随机选取数据集中 80%的数据作为 SDAE 网络模型训练数据,20%作为 SDAE 网络模型的测试数据,测试模型的鲁棒性。

(4)将在数据集中选取的 80%的训练集输入 SDAE 深度学习模型,并且微调 SDAE 深度学习网络模型参数,利用样本数据和标签对网络进行有监督的微调。

(5)将在数据集中选取的 20%的测试集输入到已经训练完成的 SDAE 深度学习模型,测试训练好的 SDAE 深度学习网络模型的拟合度及鲁棒性。

(6)数据集中的训练数据和测试数据加上噪声,重复训练与测试 SDAE 深度学习模型,完成 SDAE 深度学习模型抗噪能力的检测操作。

多目标检测流程如图 2 所示。

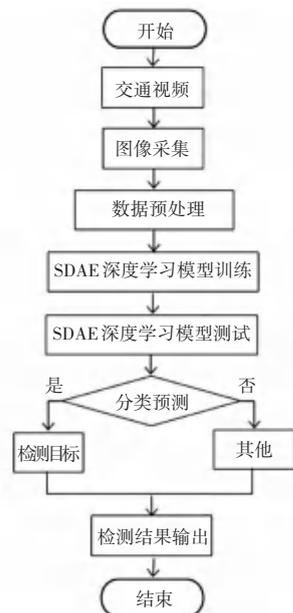


图 2 多目标检测流程图

Fig. 2 Flow chart of multi-target detection

### 2.3 多目标跟踪

将多个运动目标的跟踪转化成估计最相似的粒

子图像块。训练一个基于堆栈式去噪自编码器的多类运动目标的深度学习网络模型,实现多个运动目标的跟踪,利用粒子滤波技术和 Softmax 分类器完成多个运动目标的跟踪。

本文提出一种基于堆栈式去噪自编码器深度框架的多目标跟踪网络模型。算法流程:

(1) 利用拥有海量图像的数据集进行离线无监督训练基于 SDAE 深度学习框架的网络模型,以此获得大量图像中的一般特征。

(2) 在多运动目标跟踪时,在线有监督微调参数,有效提取运动目标的深度特征,可以实现更加鲁棒的多目标跟踪。

(3) 增加状态记忆模块,利用粒子滤波技术,实现多运动目标深度特征的自动在线学习和更新 SDAE 深度学习模型,实现稳定的多运动目标跟踪。多目标跟踪流程如图 3 所示。

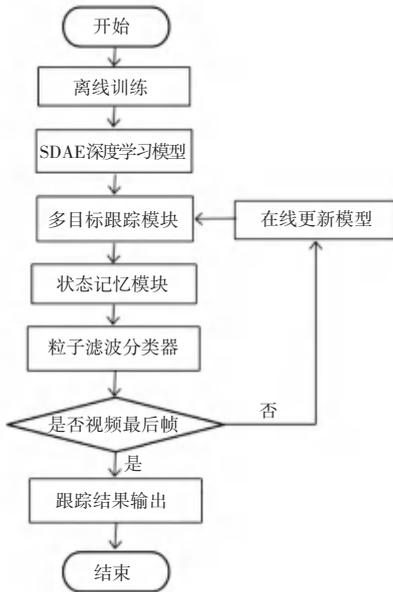


图 3 多目标跟踪流程图

Fig. 3 Flow chart of multi-target tracking

### 3 实验结果分析

实验环境: Windows7, 128G 内存, GPU 为 GTX1080, TensorFlow1.4, Python3.6。

基于 SDAE 深度学习的检测与跟踪模型在复杂交通场景下的视频序列上取得了较好的检测与跟踪效果,并且能够适应各种环境的变化,在大多数的视频序列中能够表现出较好的性能。本文选取了背景杂乱的视频序列进行目标检测与跟踪实验,基于 SDAE 深度学习的检测与跟踪模型能够准确地检测出复杂场景中的多个行人,实现鲁棒的多目标跟踪,检测与跟踪结果如图 4~图 6 所示。

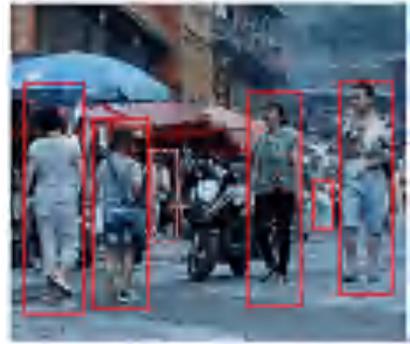


图 4 多目标检测与跟踪结果 1

Fig. 4 Multi target detection and tracking result 1



图 5 多目标检测与跟踪结果 2

Fig. 5 Multi target detection and tracking result 2



图 6 多目标检测与跟踪结果 3

Fig. 6 Multi target detection and tracking result 3

### 4 结束语

本文提出了一种基于 SDAE 深度学习框架的多目标检测与跟踪模型,该模型通过了复杂交通场景视频序列中多目标的检测与跟踪验证,实验结果表明,本文提出的基于 SDAE 深度学习框架的多目标检测与跟踪模型能够提取目标高层深度特征,能够在复杂的背景下准确地检测出多个目标,适应多目标检测与跟踪所需的深度特征,并且取得较好的检测效果与稳定的跟踪效果。

### 参考文献

[1] 尹宏鹏,陈波,柴毅,等. 基于视觉的目标检测与跟踪综述[J]. Acta Automatica Sinica, 2016, 42(10): 1466-1489.

[2] 武迹淳. 智能交通系统发展与展望[J]. 工业 b, 2015(21): 33-35.

[3] VINCENTP, LAROCHELLE H, LAJOIE I, et al. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion[J]. Journal of Machine Learning Research, 2010, 11(12): 3371-3408