

文章编号: 2095-2163(2020)02-0232-05

中图分类号: TP391.4

文献标志码: A

基于 CNN 和 LSTM 混合模型的车辆行为检测方法

王 硕, 王岩松, 王孝兰

(上海工程技术大学 机械与汽车工程学院, 上海 201620)

摘要: 为了解决复杂道路环境下前方车辆行为识别率低、实时性差的问题, 提出一种基于 CNN 和 LSTM 混合模型的车辆行为检测方法。首先, 利用 YOLOv3 目标检测算法与 Kalman 滤波对前方车辆视频中的车辆目标进行实时检测与跟踪, 获取目标大小、位置、轨迹等信息, 其次将检测跟踪得到的特征信息输入 LSTM 网络建立时间序列模型, 最终识别前方车辆行为。通过实验验证了本文方法的有效性。

关键词: 行为识别; YOLOv3; 智能车辆; LSTM; 卷积神经网络

Vehicle behavior detection method based on CNN and LSTM hybrid model

WANG Shuo, WANG Yansong, WANG Xiaolan

(School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

[Abstract] In order to solve the problem of low vehicle behavior recognition rate and poor real-time performance in complex road environment, a vehicle behavior detection method based on CNN and LSTM hybrid model is proposed. Firstly, the YOLOv3 target detection algorithm and Kalman filter are used to detect and track the vehicle target in the forward vehicle video in real time, and obtain the target size, position, trajectory and other information. Secondly, the feature information detected by the tracking is input into the LSTM network to establish a time series model. Finally, the paper identifies the behavior of the vehicle ahead. The effectiveness of the proposed method is verified by experiments.

[Key words] behavior recognition; YOLOv3; intelligent vehicle; LSTM; CNN

0 引言

随着汽车行业科学技术的发展, 智能车辆成为未来汽车行业发展的必然趋势^[1]。自动驾驶作为智能车辆的关键技术, 引起了研究者的广泛关注。其中, 环境感知技术通过传感器探测车辆周围环境, 为智能车决策部分提供信息, 是自动驾驶技术的重要前提。车辆作为道路环境的主体, 其行为姿态的识别对自身车辆周围环境理解具有重要影响^[2]。

车辆行为的分类主要采用隐马尔科夫模型 (Hidden Markov Model, HMM)、贝叶斯网络、BP 神经网络等方法。文献[3]采用贝叶斯网络对每个典型的轨迹建模进行分类。Regine 等人^[4]将当前场景与数据库中场景进行对比, 选出最相似场景的基于案件推理的在线学习行为预测方法。黄慧玲等人^[5]提出一种基于前方车辆行为识别的碰撞预警系统, 通过梯度方向直方图和支持向量机来训练前方车辆检测模型, 并结合 Kalman 滤波对车辆进行跟踪, 最后通过 HMM 算法建立车辆行为识别模型, 完成前方车辆行为的识别。传统方法多存在检测和识

别精度低的问题, 并且传统方法的使用场景多是环境简单的道路路口或高速车道等固定场景, 无法满足实际需求。

随着深度学习技术的发展, 循环神经网络 (RNN)^[6]被用于处理行为识别问题。但 RNN 存在梯度消失和梯度爆炸等问题, 标准的 RNN 网络对长序列的学习效果不佳。Hochreiter 等人^[7]对 RNN 进行改进, 在 1997 年提出长短期记忆网络 (Long Short-Term Memory, LSTM), 在长序列的处理上表现更好。Donahue 等人^[8]提出一种长效递归卷积神经网络 (LRCN), 初次将 CNN 与 LSTM 结合, 用于视频识别、分类等领域。随后越来越多的研究者采用这种方式对人体行为进行识别, 通过 CNN 模型提取特征, 并采用 LSTM 进行训练分类^[9-10]。文献[11]通过双卷积网络并行对关键帧的特征信息提取, 利用 LSTM 网络建立序列模型, 完成车辆行为动态识别, 但存在实时性差与不能识别多台车辆行为的问题。车辆在实际行驶中会遇到阴影、光斑、畸变等因素影响, 导致传统方法的车辆检测误检率高、目标偏差大

基金项目: 国家自然科学基金 (51675324)。

作者简介: 王 硕 (1995-), 男, 硕士研究生, 主要研究方向: 智能车辆环境感知; 王岩松 (1971-), 男, 博士, 教授, 主要研究方向: 车辆系统动力学; 王孝兰 (1985-), 女, 博士, 讲师, 主要研究方向: 智能车辆导航控制。

收稿日期: 2019-11-24

与定位误差大的问题, 基于深度学习的方法对硬件要求较高, 实时性低。同时车辆之间的遮挡也会导致跟踪丢失目标, 使得车辆轨迹难以跟踪, 参数获取缺乏完整性和准确性, 影响最后的行为识别精度。

本文针对上述情况提出一种基于 CNN 和 LSTM 混合模型的车辆行为检测方法, 分别选择基于 YOLOv3^[12] 的车辆识别方法, 采用 Kalman^[13] 滤波器跟踪车辆融合 LSTM 网络构建车辆行为识别模型。通过试验, 验证本文提出方法的有效性和准确性。

1 车辆检测与跟踪

1.1 YOLOv3 原理

YOLO 算法将目标检测问题作为回归问题处理, 一步完成分类与定位任务, 直接预测目标的类别与位置, 减少检测时间的同时提升了目标检测精度。YOLOv3 算法是在 YOLO 的基础上进行改进, 增加到 3 种检测尺度, 有效地增加了算法对不同尺寸物体的检测精度, 同时采用多尺度训练的方式增强了鲁棒性, 并在 DarkNet19 基础上引入残差网络提出新的特征提取网络 DarkNet53。其网络结构如图 1 所示, DarkNet53 主要由一系列 1×1 和 3×3 的卷积层组成, 每个卷积层后跟一个 Batch Normalization 层和一个 LeakyReLU 层, 进行 5 次下采样提取特征, 采用 FPN 的思想, 利用多尺度的特征增加特征丰富度, 从而达到对不同尺寸目标的精确检测。

	类型	滤波器	尺寸	输出	
1×	卷积层	32	3×3	416×416	
	卷积层	64	3×3/2	208×208	
	卷积层	32	1×1		
	卷积层	64	3×3		
	残差层			208×208	
	卷积层	128	3×3/2	104×104	
	2×	卷积层	64	1×1	
		卷积层	128	3×3	
		残差层			104×104
		卷积层	256	3×3/2	52×52
8×	卷积层	128	1×1		
	卷积层	256	3×3		
	残差层			52×52	
	卷积层	512	3×3/2	26×26	
8×	卷积层	256	1×1		
	卷积层	512	3×3		
	残差层			26×26	
	卷积层	1 024	3×3/2	13×13	
4×	卷积层	512	1×1		
	卷积层	1 024	3×3		
	残差层			13×13	
	平均池化		全局		
	连接数	1 000			
	分类				

图 1 YOLOv3 网络结构

Fig. 1 YOLOv3 network structure

在目标检测阶段, YOLOv3 首先将图片缩放到 416×416 的大小, 并划分为 $S \times S$ 的等大单元格。当待检测目标中心落入单元格内, 该单元格预测目标 B 个候选框的位置信息以及目标类别的条件概率,

即中心坐标 (x, y) 和目标的宽 w 和高 h , 记为 t_x, t_y, t_w, t_h 。目标中心所在单元格相对图像左上角有偏移 (c_x, c_y) , p_w 和 p_h 代表该单元格对应候选框的宽高, 计算出候选框的坐标如下:

$$b_x = \sigma(t_x) + c_x, \quad (1)$$

$$b_y = \sigma(t_y) + c_y, \quad (2)$$

$$b_w = p_w e^{t_w}, \quad (3)$$

$$b_h = p_h e^{t_h}. \quad (4)$$

当多个候选框检测到相同目标时, 通过对置信度设定阈值筛选出高分的候选框, 并利用非极大值抑制处理得到最终的预测结果。

1.2 车辆检测与跟踪模型

复杂环境下的车辆检测会受到遮挡影响, 导致跟踪丢失目标, 使得车辆轨迹难以跟踪, 参数获取缺乏完整性和准确性, 影响最后的行为识别精度。针对上述情况, 利用 Kalman 滤波器对检测到的车辆目标大小 (w, h) 、位置 (x, y) 等确切信息进行跟踪预测。

Kalman 滤波具体的原理和过程分析如下:

$$\hat{X}_{k|k-1} = A\hat{X}_{k-1|k-1} + BU_k, \quad (5)$$

$$P_{k|k-1} = AP_{k-1|k-1}A^T + Q, \quad (6)$$

$$K_k = P_{k|k-1}H^T(R + HP_{k|k-1}H^T)^{-1}, \quad (7)$$

$$X_{k|k} = X_{k|k-1} + K_k(Z_k - H\hat{X}_{k|k-1}), \quad (8)$$

$$P_{k|k} = P_{k|k-1} - K_kHP_{k|k-1}. \quad (9)$$

其中, A, B 表示系统参数; H 表示观测系统参数; Q, R 分别为系统噪声和观测系统噪声的方差。

在跟踪阶段, 运动车辆单位时间在图像平面上近似为线性的匀速运动模型, 以初始帧中车辆目标检测的信息初始化 Kalman 滤波器的状态向量 $\hat{X} = [x_0, y_0, xv_0, yv_0]^T$, (x_0, y_0) 表示目标区域的质心位置信息, (xv_0, yv_0) 表示质心的速度信息, 完成遮挡情况下的车辆目标跟踪与预测, 保证轨迹信息的完整性与准确性。结果如图 2 所示。

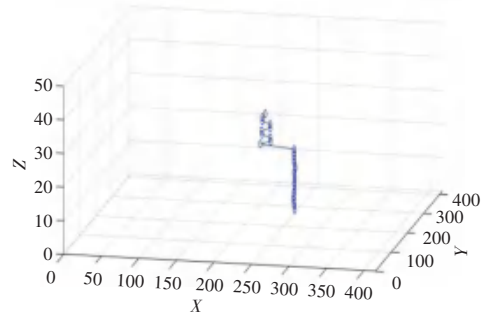


图 2 前方车辆运动轨迹

Fig. 2 Forward vehicle trajectory

2 前方车辆行为识别

2.1 LSTM 网络

循环神经网络具有强烈的时间序列性,其当前时刻的输出由之前时刻的状态和当前时刻的输入决定,保留了时间维度上的信息。LSTM 是循环神经网络的一种,为了解决一般循环神经网络梯度消失只能对短时间序列建模的问题,LSTM 引入输入门、遗忘门、输出门三个部分更新系统状态,可以有效处理时间序列中间隔和延迟较长的信息。其网络结构如图 3 所示。

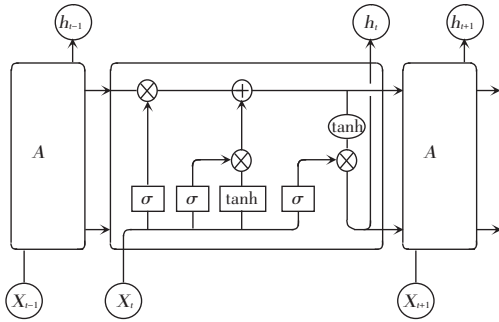


图 3 LSTM 原理图

Fig. 3 LSTM schematic

LSTM 通过输入门 i_t 、遗忘门 f_t 、输出门 o_t 丢弃时间序列中的无用的信息并传递后续时刻所需信息。其中, i_t 和 o_t 控制信息的流入和流出, f_t 代表遗忘门的输出,具体公式为:

$$f_t = \text{sigmoid}(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (10)$$

$$i_t = \text{sigmoid}(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (11)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \quad (12)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \quad (13)$$

$$o_t = \text{sigmoid}(W_o \cdot [h_{t-1}, x_t] + b_o), \quad (14)$$

$$h_t = o_t * \tanh(C_t). \quad (15)$$

其中, x_t, h_t 表示当前时刻单元的输入和输出; C_t 表示更新后的单元; W_i, W_f, W_c, W_o 与 b_i, b_f, b_c, b_o 分别表示加权项与偏置项。

2.2 车辆行为识别模型

利用 YOLOv3 目标检测算法与 Kalman 滤波对前方车辆视频中的车辆目标进行实时检测与跟踪,结合 LSTM 网络构建前方车辆行为识别模型。

本文网络结构模型如图 4 所示。首先,对输入的视频利用训练的 YOLOv3 车辆检测模型逐帧检测,提取前方车辆目标,得到目标类别、检测框尺寸和目标中心点坐标;其次,结合 kalman 滤波对前方车辆视频中的车辆目标进行实时检测与跟踪,获取目标大小、位置、轨迹等信息,获取准确的车辆运动

轨迹特征;再次,将车辆运动轨迹特征输入到 LSTM 网络框架中,建立时间序列模型,分析序列间特征,最终识别前方车辆的行为。

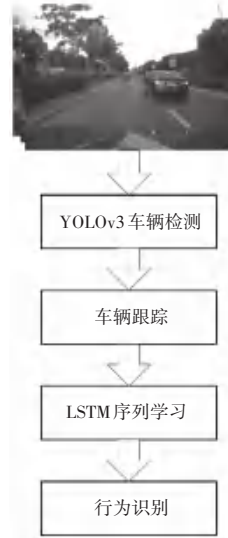


图 4 前方车辆行为识别模型

Fig. 4 Forward vehicle behavior recognition model

3 实验结果

3.1 实验平台与数据集

实验平台配置:Windows 10 操作系统,CPU 为 Inter Core i5-8400,内存为 16 GB,GPU 为 NVIDIA GeForce GTX1070Ti,深度学习框架为 Tensorflow。

本文采用自行搭建的车载试验平台所采集的视频数据集,数据采集设备固定于车辆前挡风玻璃正前方,在自身车辆不变道的情况下,采集前方车辆的行驶行为。数据集包含城区、高速、乡村和学校等场景采集的真实视频数据,每段视频中包含车辆、行人和电动车等各种目标,还包括光照变化、树木阴影和目标被各种程度的遮挡与截断等情况。根据实际应用场景,车辆视频数据分为 7 类别,分别为直行、左转、右转、左变道、右变道、左切入和右切入。车辆行为数据集如图 5 所示,本文选取 300 个视频作为实验数据,其中 80% 作为训练集,20% 作为验证集。

3.2 实验结果

经过训练得到前车行为识别网络模型,其实验结果见如下。

图 6 为车辆行为识别结果。图 6(a)中靠近本车的车辆在接下来的 5 s 向右侧变道,远方车辆保持直行,采用本文方法识别出 2 台目标车辆,并进行轨迹跟踪。识别结果显示,正确地识别出了不同车辆的右变道与直行行为。图 6(b)识别出 3 台目标车辆并进行跟踪,采用本文方法正确识别出 3 辆车的直行行为。

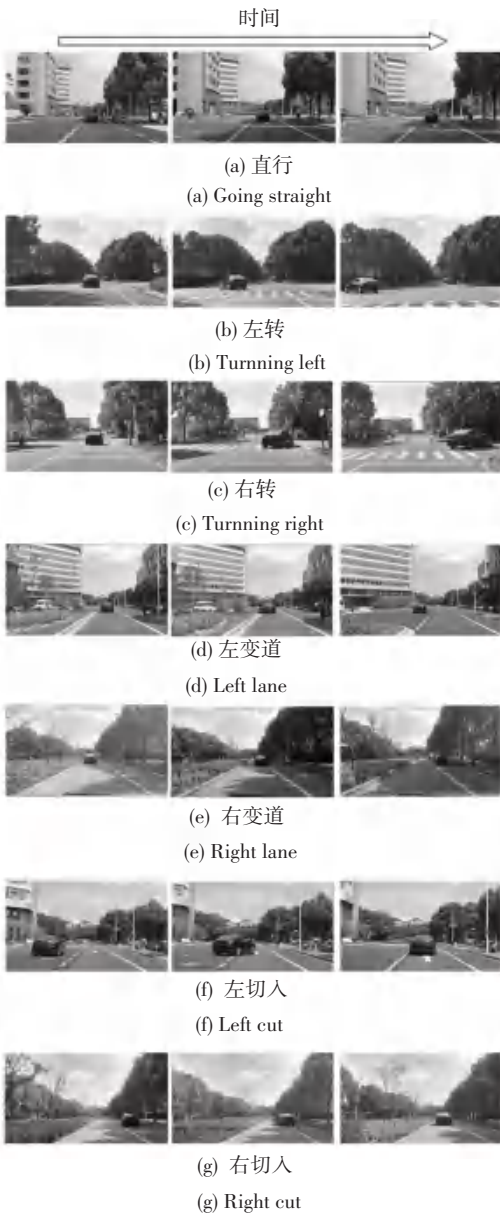


图 5 车辆行为数据集

Fig. 5 Vehicle behavior data set



图 6 前方车辆行为识别结果

Fig. 6 Forward vehicle behavior recognition results

表 1 为前车直行、左转、右转、左变道、右变道、左切入和右切入行为识别结果统计表。由表 1 分析可得, 直行行为的识别正确率较高, 而右转和右变道行为识别的正确率略低, 主要原因是车辆变道与转向操作前期轨迹特征相似, 易误检为前车左转和右转行为, 对于左切入和右切入两种行为, 本文方法可以做到有效识别。

表 1 车辆行为识别结果

Tab. 1 Vehicle behavior recognition results

前车行为	测试样本数量	正确率/%
直行	58	94.8
左转	37	91.9
右转	36	88.9
左变道	45	93.3
右变道	38	89.5
左切入	44	90.9
右切入	42	92.8

4 结束语

本文提出一种基于 CNN 和 LSTM 混合模型的车辆行为检测方法, 首先训练 YOLOv3 网络模型进行前方车辆识别, 然后使用 Kalman 滤波跟踪检测车辆; 最后训练 LSTM 网络模型对车辆行为进行建模, 完成前方车辆行为识别。实验结果表明, 根据前方车辆运动轨迹可在复杂环境下识别多车行为。但是, 本文的研究存在自身车辆保持直行的局限性, 下一步的重点应放在自身车辆行为变化的情况下识别前方车辆行为上。

参考文献

[1] 《中国公路学报》编辑部. 中国汽车工程学术研究成果综述·2017 [J]. 中国公路学报, 2017, 30(6): 1.

[2] 张艳辉, 徐坤, 郑春花, 等. 智能电动汽车信息感知技术研究进展[J]. 仪器仪表学报, 2017, 38(4): 794.

[3] KASPER D, WEIDL G, DANG T, et al. Object-oriented Bayesian networks for detection of lane change maneuvers [J]. IEEE Intelligent Transportation Systems Magazine, 2012, 4(3): 19.

[4] REGINE G, DEUSCH H, SEELIGER F, et al. A learning concept for behavior prediction at intersections[C]//2014 IEEE Intelligent Vehicles Symposium Proceedings. MI, USA: IEEE, 2014: 939.

[5] 黄慧玲, 杨明, 王春香, 等. 基于前方车辆行为识别的碰撞预警系统[J]. 华中科技大学学报(自然科学版), 2015, 43(S1): 117.

[6] RUMELHART D E, HINTON G E, WILLIAMS R J, et al. Learning representations by back-propagating errors[J]. Nature, 1986, 323: 533.

[7] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural computation, 1997, 9(8): 1735.

(下转第 239 页)