



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2020

09

VOL.25

ISSN1006-8961
CN11-3758/TB



遥感边缘智能技术 P1719



第25卷第9期 (总第293期)

2020年9月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿, 均视为同意在本刊网站及CNKI等全文数据库出版, 所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用, 所有内容, 未经许可, 不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院

主办单位 中国科学院空天信息创新研究院

中国图象图形学学会

北京应用物理与计算数学研究所

主 编 顾行发

编辑出版 《中国图象图形学报》编辑出版委员会

通信地址 北京市海淀区北四环西路19号

邮 编 100190

电子信箱 jig@radi.ac.cn

电 话 010-58887035

网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号

总 发 行 北京报刊发行局

订 购 全国各地邮局

海外发行 中国国际图书贸易集团有限公司

(邮政信箱: 北京399信箱 邮编: 100048)

印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian

Monthly, Started in 1996

Superintended by Chinese Academy of Sciences

Sponsored by Aerospace Information Research Institute, CAS

China Society of Image and Graphics

Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Gu Xingfa

Editor, Publisher Editorial and Publishing Board of Journal of
Image and Graphics

Address No. 19, North 4th Ring Road West, Haidian District,
Beijing, P. R. China

Zip code 100190

E-mail jig@radi.ac.cn

Telephone 010-58887035

Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals

Domestic All Local Post Offices in China

Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China))

Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB

ISSN 1006-8961

CODEN ZTTXFZ

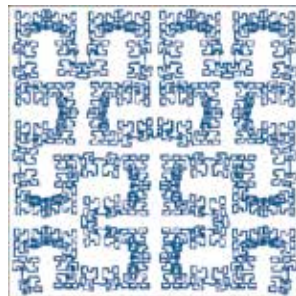
国外发行代号 M1406

国内邮发代号 82-831

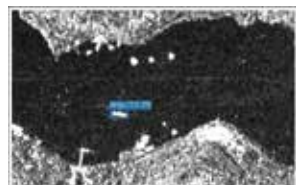
国内定价 60.00元



第一人称视角下的社会力优化多行人跟踪(第1869页)



分形曲线生成的频域方法(第1904页)



双向特征融合的数据自适应SAR图像舰船目标检测模型(第1943页)

学者观点

遥感边缘智能技术研究进展及挑战

孙显, 梁伟, 刁文辉, 曹志颖, 冯瑛超, 王冰, 付琨 1719

综述

红外弱小目标检测算法综述

李俊宏, 张萍, 王晓玮, 黄世泽 1739

目标检测中的尺度变换应用综述

申奉臻, 张萍, 罗金, 刘松阳, 冯世杰 1754

图像处理和编码

分层特征融合注意力网络图像超分辨率重建

雷鹏程, 刘丛, 唐坚刚, 彭敦陆 1773

适用小样本的无参考水下视频质量评价方法

宋巍, 刘诗梦, 黄冬梅, 王文娟, 王建 1787

图像分析和识别

结合面料属性和触觉感测的织物识别

邢寅初, 刘骊, 付晓东, 刘利军, 黄青松 1800

结合GAN的轻量级模糊车牌识别算法

段宾, 符祥, 江毅, 曾接贤 1813

非贪婪的鲁棒性度量学习算法

曾凡霞, 张文生 1825

不规则像素簇显著性检测算法

李明旭, 翟东海 1837

全局区域相异度阈值驱动构建的稀疏尺度集模型

王浩宇, 沈占锋, 柯映明, 许泽宇, 李硕, 焦淑慧 1848

图像理解和计算机视觉

分裂合并运动分割的多运动视觉里程计方法

王晨捷, 张云, 赵青, 王伟, 尹露, 罗斌, 张良培 1859

第一人称视角下的社会力优化多行人跟踪

杨廷召, 刘骊, 付晓东, 刘利军, 黄青松 1869

引入辅助损失的多场景车道线检测

陈立潮, 徐秀芝, 曹建芳, 潘理虎 1882

计算机图形学

旋转骨架法在二值图像分形维数计算中的应用

纪佑军, 何杰, 韩海水, 程忠钊, 曾保全 1894

分形曲线生成的频域方法

陈伟, 乔洁雯, 周晨 1904

医学图像处理

融合残差注意力机制的UNet视盘分割

侯向丹, 赵一浩, 刘洪普, 郭鸿湧, 于习欣, 丁梦园 1915

深层聚合残差密集网络的超声图像左心室分割

吴宣言, 缙新科, 朱子重, 魏域林, 王凯 1930

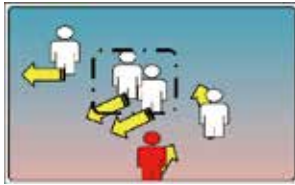
遥感图像处理

双向特征融合的数据自适应SAR图像舰船目标检测模型

张筱晗, 姚力波, 吕亚飞, 简涛, 赵志伟, 藏洁 1943

CONTENTS

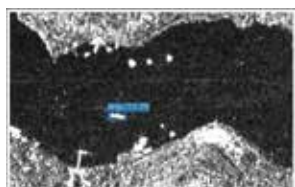
JOURNAL OF IMAGE AND GRAPHICS



Multi-pedestrian tracking optimized by social force model under first-person perspective(P1869)



Fractal curve generation method based on the frequency domain (P1904)



Data-adaptive single-shot ship detector with a bidirectional feature fusion module for SAR images (P1943)

Scholar View

Progress and challenges of remote sensing edge intelligence technology

Sun Xian, Liang Wei, Diao Wenhui, Cao Zhiying, Feng Yingchao, Wang Bing, Fu Kun 1719

Review

Infrared small-target detection algorithms: a survey

Li Junhong, Zhang Ping, Wang Xiaowei, Huang Shize 1739

Scale changing in general object detection: a survey

Shen Fengcan, Zhang Ping, Luo Jin, Liu Songyang, Feng Shijie 1754

Image Processing and Coding

Hierarchical feature fusion attention network for image super-resolution reconstruction

Lei Pengcheng, Liu Cong, Tang Jiangang, Peng Dunlu 1773

Non-reference underwater video quality assessment method for small size samples

Song Wei, Liu Shimeng, Huang Dongmei, Wang Wenjuan, Wang Jian 1787

Image Analysis and Recognition

Cloth recognition based on fabric properties and tactile sensing

Xing Yinchu, Liu Li, Fu Xiaodong, Liu Lijun, Huang Qingsong 1800

Lightweight blurred car plate recognition method combined with generated images

Duan Bin, Fu Xiang, Jiang Yi, Zeng Jiexian 1813

Robust distance metric learning algorithm based on nongreedy solution

Zeng Fanxia, Zhang Wensheng 1825

Significance detection method with irregular pixel clusters

Li Mingxu, Zhai Donghai 1837

Sparse scale set model based on global regional dissimilarity threshold

Wang Haoyu, Shen Zhanfeng, Ke Yingming, Xu Zeyu, Li Shuo, Jiao Shuhui 1848

Image Understanding and Computer Vision

Multi-motion visual odometry based on split-merged motion segmentation

Wang Chenjie, Zhang Yun, Zhao Qing, Wang Wei, Yin Lu, Luo Bin, Zhang Liangpei 1859

Multi-pedestrian tracking optimized by social force model under first-person perspective

Yang Tingzhao, Liu Li, Fu Xiaodong, Liu Lijun, Huang Qingsong 1869

Multi-scenario lane line detection with auxiliary loss

Chen Lichao, Xu Xiuzhi, Cao Jianfang, Pan Lihu 1882

Computer Graphics

Application of skeleton rotating method in fractal dimension calculate of binary image

Ji Youjun, He Jie, Han Haishui, Cheng Zhongzhao, Zeng Baoquan 1894

Fractal curve generation method based on the frequency domain

Chen Wei, Qiao Jiewen, Zhou Chen 1904

Medical Image Processing

Optic disk segmentation by combining UNet and residual attention mechanism

Hou Xiangdan, Zhao Yihao, Liu Hongpu, Guo Hongyong, Yu Xixin, Ding Mengyuan 1915

Left ventricular segmentation on ultrasound images using deep layer aggregation for residual dense networks

Wu Xuanyan, Gou Xinke, Zhu Zizhong, Wei Yulin, Wang Kai 1930

Remote Sensing Image Processing

Data-adaptive single-shot ship detector with a bidirectional feature fusion module for SAR images

Zhang Xiaohan, Yao Libo, Lyu Yafei, Jian Tao, Zhao Zhiwei, Zang Jie 1943

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2020)09-1882-12

论文引用格式: Chen L C, Xu X Z, Cao J F and Pan L H. 2020. Multi-scenario lane line detection with auxiliary loss. Journal of Image and Graphics, 25(09):1882-1893 [陈立潮, 徐秀芝, 曹建芳, 潘理虎. 2020. 引入辅助损失的多场景车道线检测. 中国图象图形学报, 25(09):1882-1893] [DOI: 10.11834/jig.190646]

引入辅助损失的多场景车道线检测

陈立潮¹, 徐秀芝¹, 曹建芳^{1,2}, 潘理虎¹

1. 太原科技大学计算机科学与技术学院, 太原 030024; 2. 忻州师范学院计算机系, 忻州 034099

摘要: **目的** 为解决实时车辆驾驶中因物体遮挡、光照变化和阴影干扰等多场景环境影响造成的车道线检测实时性和准确性不佳的问题,提出一种引入辅助损失的车道线检测模型。**方法** 该模型改进了有效的残差分解网络(efficient residual factorized network, ERFNet),在 ERFNet 的编码器之后加入车道预测分支和辅助训练分支,使得解码阶段与车道预测分支、辅助训练分支并列,并且在辅助训练分支的卷积层之后,利用双线性插值来匹配输入图像的分辨率,从而对 4 条车道线和图像背景进行分类。通过计算辅助损失,将辅助损失以一定的权重协同语义分割损失、车道预测损失进行反向传播,较好地解决了梯度消失问题。语义分割得到每条车道线的概率分布图,分别在每条车道线的概率分布图上按行找出概率大于特定阈值的最大点的坐标,并按一定规则选取相应的坐标点,形成拟合的车道线。**结果** 经过在 CULane 公共数据集上实验测试,模型在正常场景的 F1 指标为 91.85%,与空间卷积神经网络(spatial convolutional neural network, SCNN)模型相比,提高了 1.25%,比其他场景分别提高了 1%~7%;9 种场景的 F1 平均值为 73.76%,比目前最好的残差网络——101-自注意力蒸馏(ResNet-101-self attention distillation, R-101-SAD)模型(71.80%)高出 1.96%。在单个 GPU 上测试,每幅图像的平均运行时间缩短至原来的 1/13,模型的参数量减少至原来的 1/10。与平均运行时间最短的车道线检测模型 ENet——自注意力蒸馏(ENet-self attention distillation, ENet-SAD)相比,单幅图像的平均运行时间减短了 2.3 ms。**结论** 在物体遮挡、光照变化、阴影干扰等多种复杂场景下,对于实时驾驶车辆而言,本文模型具有准确性高和实时性好等特点。

关键词: 多场景车道线检测;语义分割网络;辅助损失;梯度消失;CULane 数据集

Multi-scenario lane line detection with auxiliary loss

Chen Lichao¹, Xu Xiuzhi¹, Cao Jianfang^{1,2}, Pan Lihu¹

1. School of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China;

2. Department of Computer, Xinzhou Teachers University, Xinzhou 034099, China

Abstract: Objective In a real-time driving process, the vehicle must be positioned to complete the basic tasks of horizontal and vertical control. The premise of vehicle positioning problems is to understand road information. Road information includes all kinds of traffic signs, among which lane line is an important pavement information in the road scene. Such information is crucial for lane maintaining, departure warning, and path planning; it is also important in research on advanced driving assistance systems. Therefore, lane line detection has become an important topic in real-time vehicle driving. Road scene images can be obtained using a vehicle camera, lidar, and other equipments, thus making it easy to obtain

收稿日期: 2020-01-03; 修回日期: 2020-03-09; 预印本日期: 2020-03-16

基金项目: 山西省自然科学基金项目(201901D111258); 山西省应用基础研究项目(201801D221179)

Supported by: Natural Science Foundation of Shanxi Province, China (201901D111258); Applied Basic Research Project of Shanxi Province, China (201801D221179)

lane line images. However, lane line detection suffers from some difficulties. Traditional lane line detection methods usually design features manually. Starting from the bottom features, such as color, brightness, shape, and gray level, the method involves image processing via denoising, binarization, and graying. Then, the lane line features are extracted by combining edge detection, Hough transform, color threshold setting, perspective transform, and other methods. Afterward, the lane lines are fitted by straight or curve line models. These methods are simple and easy to implement, but the accuracy of lane line detection is poor under the influence of multiscene environment conditions, such as object occlusion, light change, and shadow interference; moreover, the manual design of features is time consuming and thus cannot meet the real-time requirements of vehicle driving. To solve these problems, this study proposes a lane detection model named efficient residual factorized network-auxiliary loss (ERFNet-AL), which embeds an auxiliary loss. **Method** The model improves the semantic segmentation network of ERFNet. After the encoder of ERFNet, a lane prediction branch and an auxiliary training branch are added to make the decoding phase parallel with the lane prediction and auxiliary training branches. After the convolution layer of the auxiliary training branch, bilinear interpolation is used to match the resolution of input images to classify four lane lines and the background of images. The training set images in the dataset are sent to the lane line detection model after preprocessing, such as clipping, rotating, scaling, and normalization. The features are extracted through the semantic segmentation network ERFNet, thereby obtaining the probability distribution of each lane line. The auxiliary training branch uses a convolution operation to extract features, and bilinear interpolation is used to replace the deconvolution layer after the convolution layer to match the resolution of the input images and classify the four lane lines and background. After using convolution, batch normalization, dropout layers, and other operations, the lane line prediction branch predicts the existence of lane lines or virtual lane lines and outputs the probability value of lane line classification. An output probability value greater than 0.5 indicates the existence of lane lines. If at least one lane line exists, then a probability distribution map of the corresponding lane lines must be determined. On the probability distribution map of each lane line, the coordinates of the largest point with a probability greater than a specific threshold is identified by row, and the corresponding coordinate points is selected in accordance with the rules of selecting points in the SCNN (spatial convolutional neural network) model. If the number of points found are greater than 2, then these points are connected to form a fitted lane line. Then, the cross-entropy loss between the predicted value of the auxiliary training branch output and the real label is calculated and used as the auxiliary loss. The weight of all four lane lines is 1, and the weight of the background is 0.4. The auxiliary, semantic segmentation, and lane prediction losses are weighted and summed in accordance with a certain weight, and the network parameters are adjusted via backpropagation. Among them, the total loss includes the main, auxiliary, and lane prediction losses. During training, the weights of ERFNet on the Cityscapes dataset are used as pretraining weights. During training, the model with the largest mean intersection over union is taken as the best model. **Result** After testing in nine scenarios of the CULane public dataset, the F1 index of the model in the normal scenario is found to be 91.85%, which is a 1.25% increase compared with that of the SCNN model (90.6%). Moreover, the F1 index in seven scenes, including crowded, night, no line, shadow, arrow, dazzle light, and curve scenarios, is increased by 1%~7%; the total average F1 value in nine scenarios is 73.76%, which is 1.96% higher than the best ResNet-101-self-attention distillation (SAD) model; the average run time of each image is 11.1 ms, which is 11 times shorter than the average running time of the spatial CNN model when tested on a single GPU of GeForce GTX 1080; the parameter quantity of the model is only 2.49 MB, which is 7.3 times less than that of the SCNN model. On the CULane dataset, ENet with SAD is the lane line detection model with the shortest average run time of a single-image test. The average run time of this model is 13.4 ms, whereas that of our model is 11.1 ms. Compared with ENet with SAD, the average running time is reduced by 2.3 ms. When detecting lane lines at a crossroad scenario, the number of false positive is large, which may be due to the large number of lane lines at crossroads, whereas only four lane lines are detected in our experiment. **Conclusion** In various complex scenarios, such as object occlusion, lighting changes, and shadow interference, the model is minimally affected by the environment for real-time driving vehicles, and its accuracy and real-time performance are improved. The next work will aim to increase the number of lane lines, optimize the model, and improve the model's detection performance at crossroads.

Key words: multi-scenario lane line detection; semantic segmentation network; auxiliary loss; gradient disappear; CULane dataset

0 引言

随着人工智能的发展及 5G 网络的出现,智能驾驶汽车应运而生,但由于部分车联网和智能驾驶技术尚不成熟,根基尚未稳固,使得智能驾驶的安全问题日渐突出。为此,研究者提出了许多相应的方法和技术,在这些技术中,道路感知和车道标记检测在帮助驾驶员避免错误方面起着至关重要的作用(Xing 等, 2018),而高级驾驶辅助系统(advanced driving assistant system, ADAS)在获取车辆内部及外部的信息方面具有显著优势,可以辅助驾驶员分析道路场景,检测路面信息。

车道线作为路面主要交通标志,它的检测技术对于车道偏离预警系统、车道保持系统或轨迹规划决策都至关重要,能显著提高汽车行驶的安全性(赵文明和张海文, 2017)。基于视觉的车道线检测主要有基于传统方法的检测方法和基于深度学习的检测方法。

传统的车道线检测算法一般包括车道图像获取、图像预处理、车道线特征提取、车道线拟合和车道线输出等方面。车道图像主要通过车载摄像机的拍摄等方式获取,经过图像灰度化、感兴趣区域划分和图像逆透视变换等预处理后,利用滑动窗口搜索和 Hough 变换等方法提取车道线的边缘和形状等特征,再利用直线或曲线模型拟合车道线,最终输出车道线。其中,基于多帧叠加和窗口搜索的快速车道检测方法(陈涵深 等, 2018),在图像预处理时,通过逆透视变换将图像感兴趣区域转换成鸟瞰图,结合多帧叠加将 RGB 图像转化为二值图。在图像识别时,采用滑动窗口搜索提取车道线,最后选用不同的车道线模型,结合最小二乘法拟合车道线。城市轿车路口背景提取与车道标定算法(高飞 等, 2016),融合了均值与帧间差分方法,进行背景稳定与更新,利用 Canny 算子及 Hough 直线检测得到各类车道线,采用透视变换和聚类分析等建立数学模型,达到车道线标定的目的。这些传统的车道线检测方法在光线变化、路面积水和阴影干扰等场景下实时性不好,准确率不佳。传统方法通过单目视觉实时获取前方道路信息并对其进行处理和检测,但由于道路环境复杂、光照不均及障碍物引起的边缘断裂和亮度不连续导致图像难以获得连续的边缘特

性(凌诗韵 等, 2017)。

为解决多场景下传统车道线检测算法存在的问题,国内外学者开始研究基于深度学习的车道线检测模型, Pan 等人(2018)提出跨行和列的像素间空间消息传递的网络(spatial convolutional neural network, SCNN),将传统的层接层(layer-by-layer)的卷积连接形式转换为特征图中片连片(slice-by-slice)的卷积形式,与空间卷积神经网络不同,它通过专门设计的 CNN 结构传播空间信息,将特征图的行或者列视为一层,并应用卷积、非线性激活以及求和运算从上到下、从下到上、从左到右和从右到左依次计算,使得信息可以在同一层结构中传递,有效地提取空间信息。该方法适用于检测像车道线这种具有长距离连续形状,并且有极强的空间关系但外观线索较差的目标; Hou 等人(2019)提出一种采用自我关注蒸馏的轻量级神经网络检测车道线,在不需要任何额外的监督和标记的情况下,允许模型从自身中学习并获得实质性的改进。这些算法一定程度上克服了传统算法的一些缺点,但检测的实时性和准确性仍差强人意。Chen 和 Chen(2017), Ghafoorian 等人(2018)将语义分割网络广泛用于车道线检测,给图像中的每个像素都分配了相应的标签以指示其是否属于车道。

鉴于此,本文提出一种引入辅助损失的语义分割网络检测车道线,既减少了参数和运行时间,又提高了检测指标,对城市复杂场景下的车道线检测具有很好的实时性和准确性,对 ADAS 系统的研发具有一定的指导意义。

1 改进的车道线检测

1.1 ERFNet 语义分割网络

有效的残差分解网络(efficient residual factorized network, ERFNet)(Romera 等, 2017, 2018)是针对智能车辆自动驾驶提出的一种结合残差连接和分解卷积的语义分割网络。

ERFNet 与语义分割网络(semantic segmentation network, SegNet)(Badrinarayanan 等, 2017)类似,采用基于产生下采样特征映射的编码器段和对特征映射进行上采样,以匹配输入分辨率的后续解码器段进行语义分割(陈昌宇, 2018)。模型共 23 层,其中,1—16 层为编码(encoder)阶段,17—23 层为解码(decoder)阶段。

编码阶段通过两个模块:下采样模块和非瓶颈—1维(non-bottleneck-1 dimension, Non-bt-1D)模块提取特征。

下采样模块在通道维度上拼接最大池化和卷积层(convolution, Conv),采用“加宽”网络策略融合更多的特征图特征(石永涛,2018),下采样模块如图1所示,图中 s 表示步长, w_i 和 w_o 分别表示本模块的输入和输出特征图的通道数。

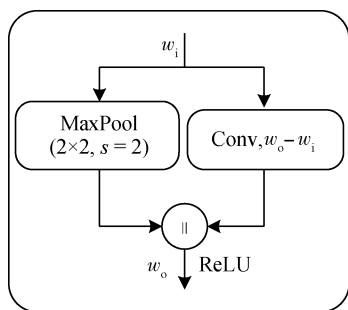


图1 下采样模块

Fig.1 Downsampler module

Non-bt-1D 模块是利用非瓶颈设计构建的带1维分解内核的卷积,它将常规残差模块中两个 3×3 的卷积核分解为两组 3×1 和 1×3 的1维卷积,并利用残差连接将输入特征和卷积后的特征图进行融合,强化了网络的表达能力。常规残差和 Non-bt-1D 模块对比如图2所示,图中 w 表示输出特征图的数量。

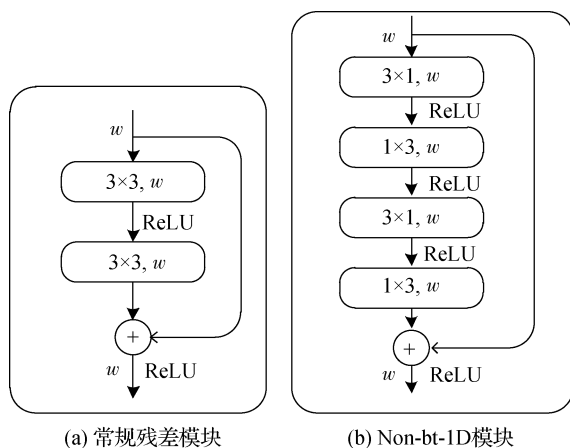


图2 常规残差模块与 Non-bt-1D 模块对比

Fig.2 Comparison between conventional residual module and Non-bt-1D module((a) conventional residual module; (b) Non-bt-1D module)

Non-bt-1D 模块中的部分层运用空洞卷积,扩张率分别为2、4、8、16,以便收集更多的背景,从而提

高检测效果。

解码阶段利用反卷积进行上采样以匹配输入的分辨率,此阶段反卷积(deconvolution, Deconv)和 Non-bt-1D 模块交替进行,利用深度神经网络从大量标注的图像数据中提取出图像特征和语义信息,再根据这些信息进行学习、推理原始图像中像素的类别,通过端到端训练的方式对每个像素进行分类,以像素分类的方式达到语义分割的目标(田萱等,2019)。

用 ERFNet 语义分割的前后对比图如图3所示。ERFNet 模型最初用于对 Cityscapes 数据集(Cordts 等,2016)的城市场景图像进行语义分割,在此数据集上,模型的计算速度要比 SegNet 等其他模型快几个数量级。基于该模型在计算速度上的优越性,将该网络模型用于车道线检测公共数据集上,借助语义分割网络,给图像中每一个像素分配一个预定义的表示其语义类别的标签(Csurka 和 Perronnin,2011)。

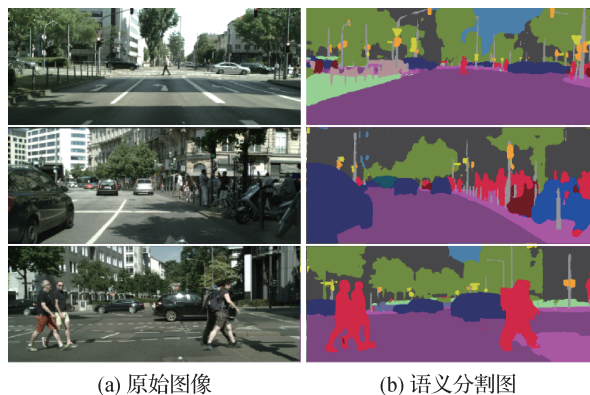


图3 语义分割效果图

Fig.3 Effect images of semantic segmentation

((a) original images; (b) images of semantic segmentation)

1.2 改进之处

1.2.1 引入辅助损失

深度学习的核心是根据训练数据得到学习器的权重。损失函数用来估量模型的预测值和真实值之间的不一致程度。训练过程需要根据损失函数反向传播来调整学习器的权重,损失函数越小,模型的性能就越好。然而,利用反向传播算法训练深度卷积神经网络时,存在梯度消失问题,越前面的隐藏层,参数的损失梯度值越小,使得后面的隐藏层训练得比较好,而前面的隐藏层几乎不能训练,导致神经网络的收敛速度不理想(朱威等,2017),影响参数的

优化。因此,解决梯度消失问题成为深度神经网络训练的重中之重。

Trinh 等人(2018)提出通过辅助损失提升循环神经网络(recurrent neural network, RNN)学习长期依赖关系的能力。该方法介绍了一种无监督辅助损失,可以重建预测锚点前后的一部分随机序列来解决梯度消失问题,优化模型参数。鉴于此,提出在ERFNet模型的编码器之后额外添加辅助训练分支来解决梯度消失问题,优化学习过程。该分支的输入通道为128,输出通道为5(对应背景和4条车道线5个类别),卷积核为 1×1 ,输出层利用双线性插值(bilinear),将图像放大8倍,以匹配输入的分辨率。

在道路场景语义分割时,由于图像的背景和车道线像素分布不均匀,采用交叉熵损失函数对模型参数求梯度可以较快地更新网络参数值(姜立标和台啟龙,2019)。交叉熵损失(cross entropy loss)等于负对数似然损失(negative log-likelihood loss, NLLLoss)与softmax取对数后的和。NLLLoss计算为

$$L(\mathbf{x}, l) = -x_l \quad (1)$$

式中, $l \in (0, C-1)$,是对应的标签, $\mathbf{x}$ 是输入向量, C 为类别数, $\mathbf{x} \in \mathbf{R}^N$, N 是 $\mathbf{x}$ 的维度。

交叉熵损失需要对输出层输出的softmax的值取对数,再与NLLLoss的计算结果相加,计算为

$$L(\mathbf{x}, l) = -w_l \ln \frac{e^{x_l}}{\sum_{j=1}^N e^{x_j}} = w_l [-x_l + \ln \sum_{j=1}^N e^{x_j}] \quad (2)$$

式中, $w \in \mathbf{R}^C$,表示各类别的权重;下标 $l \in (0, C-1)$,是对应的标签, N 是输入向量 $\mathbf{x}$ 的维度, j 为类别, $j \in (1, N)$ 。

计算该分支的交叉熵损失,将此损失作为辅助损失,辅助语义分割分支,更新参数,解决梯度消失问题。在训练过程中,语义分割分支的损失为主要损失,辅助训练的损失以一定的权重存在。

1.2.2 增加车道预测分支

车道线检测与仅需要边界框的常见物体检测任务不同,它需要精确地预测车道线。通过在ERFNet的编码器之后添加车道线预测分支小网络,该小网络与ERFNet编码器后添加的辅助训练分支、解码

器并行,用于预测车道线是否存在,包括空洞卷积层(dilated convolution, DtConv)、批标准化(batch normalization, BN)、失活层(dropout)、卷积层、最大池化层(maxpooling)和两个全连接层(fully connected layers, FC)。卷积层采用空洞卷积层,空洞卷积的扩张率为4,输入通道是128,输出通道是32,用 3×3 的卷积核,步长为1,边距为4。为了计算的稳定性,在BN层的分母上添加一个较小值 ε ,失活层以0.1的概率随机失活,卷积层用 1×1 的卷积核,输出通道为5,最大池化采用 2×2 大小的窗口,第一个全连接层的输入通道为3965,输出通道为128,最后的全连接层输出通道为4,代表4条车道线。若车道线存在的概率大于0.5,则表示存在车道线或虚拟车道线,标记为1,否则标记为0。实验中采用的数据集是在SCNN模型下提出的,且在SCNN模型采用的取点规则上对车道线的拟合效果较好,为体现改进后模型的有效性,对存在的每条车道线,按照与SCNN模型相同的取点规则进行取点,在相应的车道线概率图上每隔20行寻找对应的点,得到阈值大于0.5中概率值最大点的坐标,然后通过3次样条函数连接这些点,得到最终拟合的车道线。

2 车道线检测模型与流程

2.1 车道线检测模型

改进的车道线模型包括语义分割、辅助训练和车道预测3个分支。语义分割分支以ERFNet模型为原型,在ERFNet模型的编码器阶段后,分别添加辅助训练和车道预测分支。ERFNet模型的解码器分支与辅助训练分支、车道预测分支并列进行,辅助训练分支用于对4条车道线和背景进行分类,车道线预测分支用来判断车道线是否存在。该模型称为有效的残差分解网络—辅助损失(efficient residual factorized network-auxiliary loss, ERFNet-AL)。模型结构如图4所示。

2.2 车道线检测流程

车道线检测流程主要包括图像预处理、车道线检测和车道线拟合3个环节。图像预处理后,用ERFNet的编码器—解码器进行语义分割得到车道线概率分布图。用车道线预测分支判断是否存在车道线或虚拟车道线,若存在,则按所选的规则取点连

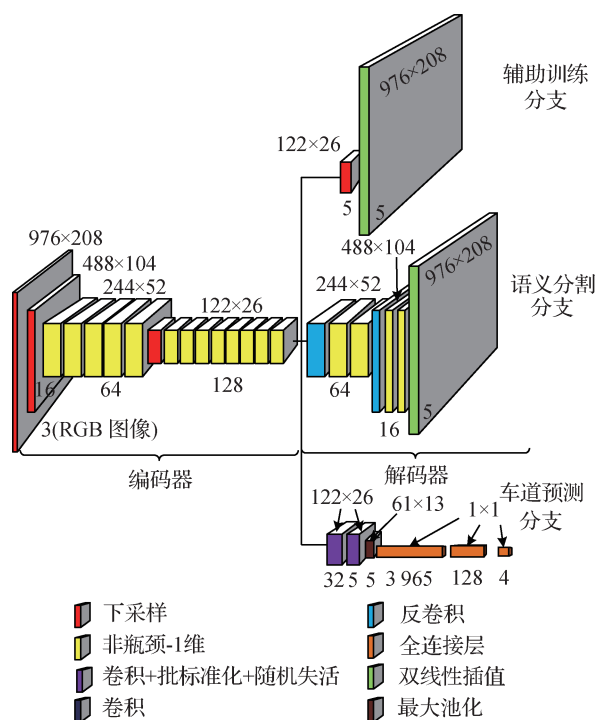


图4 网络模型图

Fig. 4 Network model diagram

线,拟合车道线。辅助训练分支只用在训练阶段,用来调整网络参数,优化网络模型。车道线检测流程如图5所示。

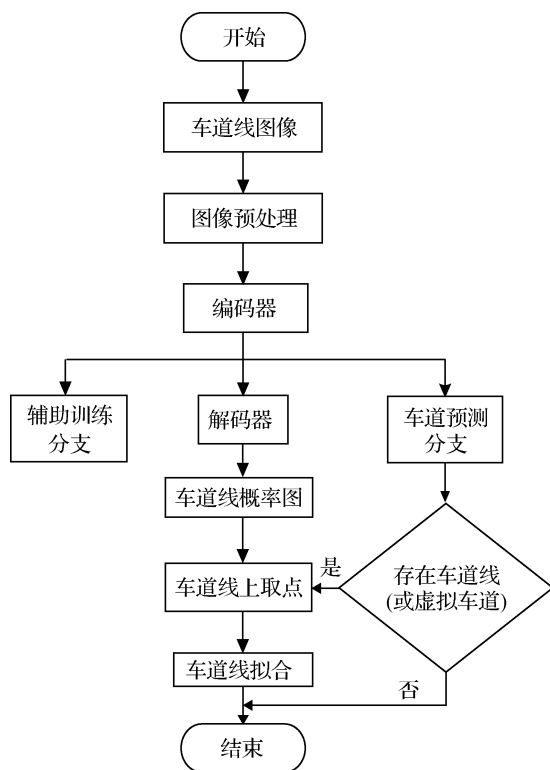


图5 车道线检测流程图

Fig. 5 Flow chart of lane line detection

3 实验结果及分析

3.1 实验环境及数据集处理

实验通过 Python 结合 MATLAB 实现。采用 pytorch 框架,在 ubuntu 16.04 LTS 64 位系统下实现,处理器为 Intel® Core™ i7-8 7500CPU @ 2.20 GHz × 12,训练阶段的图形处理器为两个 GeForce GTX 1080 GPU,测试阶段用 1 个 GPU,每个内存大小为 12 GB。

实验采用公开数据集 CULane (Pan 等,2018),收集了超过 55 h 的视频,提取了 133 235 帧。将 88 880 帧用于训练集,9 675 帧用于验证集,34 680 帧用于测试集。图像分辨率为 1 640 × 590 像素。对于每一帧,用 3 次样条手动标注车道。每幅图像都有一个 .txt 注释文件,其中每一行给出车道标志上关键点的 x, y 坐标。而训练的真实标签,每行的格式为“输入图像,每个像素的标签,4 个由 0 或 1 组合成的数字(表示从左到右存在的 4 个车道标记,若为 1110,则表示存在左一、左二、右一 3 条车道线,右二车道线不存在)”。复杂场景图像包括正常(normal)、拥挤(crowded)、夜晚(night)、不存在车道线(no line)、阴影(shadow)、箭头(arrow)、炫目的灯光(dazzle light)、曲线(curve)和交叉路口(cross-road)等 9 种情况,9 种场景及其所占比例如图 6 所示。

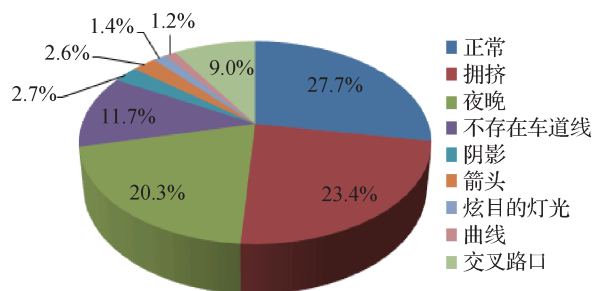


图6 9种场景及其所占比例

Fig. 6 Nine scenes and their proportion

车道线场景图像对杂波和噪声等环境较为敏感,因此车道线检测中图像预处理环节对检测效果的影响显著。实验前,先将图像由 RGB 通道图像转换为 BGR 通道的图像,以使用开源的计算机视觉库(open source computer vision library, OpenCV)读取图像。将所有图像裁剪为 976 × 208 像素,对训练

图像通过随机变换缩放、随机裁剪、随机旋转进行数据增强,并进行归一化处理,对验证图像进行缩放、裁剪和归一化处理,对测试图像进行缩放和归一化处理。

3.2 评价指标

该实验将车道线标记为宽度为 30 像素的线,并计算真实车道线和预测车道线之间的交并比(intersection over union, IoU)。若 IoU 大于 0.5,则该车道线为真正例的数目,最终采用调和平均值 F1 作为评价指标,在交叉路口场景下的真实标签不含有车道线的标注,采用假正例的数目作为指标,假正例的数目越小,效果越好。查准率、查全率、查准率和查全率的调和平均值 F1 以及准确率分别计算为

$$P = \frac{TP}{TP + FP} \quad (3)$$

$$R = \frac{TP}{TP + FN} \quad (4)$$

$$F_1 = 2 \frac{P \times R}{P + R} \quad (5)$$

$$A = \frac{TP + TN}{TP + FN + FP + TN} \quad (6)$$

式中, FN 、 TN 分别为假反例、真反例的数目, TP 、 FP 分别为真正例、假正例的数目, P 为查准率, R 为查全率, F_1 为 F1 指标, A 为准确率。

3.3 模型训练

实验利用迁移学习,将 Cityscapes 数据集上的

权重作为初始化权重。采用 batch_size 为 6 进行批训练,初始学习率为 0.01,衰减系数及学习率为

$$d_y = l_r \times (1 - p/p_s)^{0.9} \quad (7)$$

$$l'_r = l_r \times d_y \quad (8)$$

式中, d_y 表示学习率的衰减系数, l_r 为初始学习率, p 为当前训练次数, p_s 为总次数,这里设置为 15 次, l'_r 为更新后的学习率。

学习率变化曲线如图 7 所示,可以看出,最终学习率为 0.001。

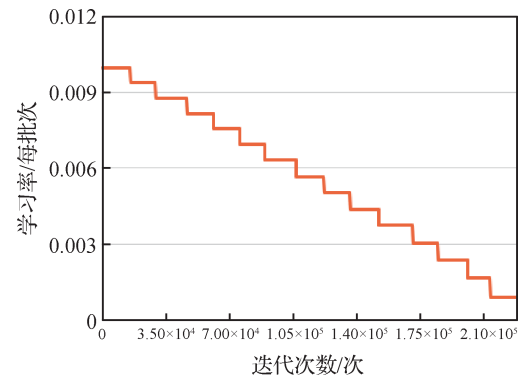


图 7 学习率曲线

Fig. 7 Curve of learning rate

语义分割、辅助训练和车道预测分支每批次的平均损失如图 8 所示。可以看出,语义分割损失较稳定,辅助训练和车道预测的损失收敛较快,当迭代到 207.4×10^3 次时,损失值达到最低。

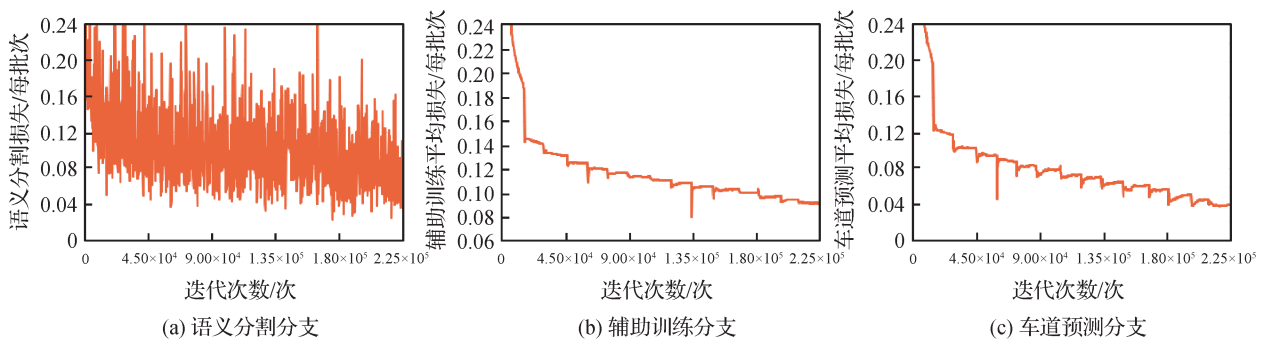


图 8 模型分支平均损失曲线

Fig. 8 Curves of average loss of model branches

((a) semantic segmentation branch; (b) auxiliary training branch; (c) lane forecast branch)

语义分割用于对车道线场景图像的背景和车道线准确分类,为了更准确地检测小物体,通常为较少的数据设置更大的损失权重,方便神经网络准确检测小物体,所以,在计算总损失时,语义分割损失的权重设置为 1,车道预测分支、辅助

训练分支主要是在语义分割编码器的基础上对车道线进行分类,一定程度上避免了因语义分割分类不准确而造成车道线分类不正确的问题,通过多次实验发现,将车道预测损失和辅助训练损失的权重都设置为 0.1 时,语义分割损失、车道

预测损失和辅助训练损失三者加权求和后的总损失达到最低, 收敛较快。总损失曲线如图9所示。

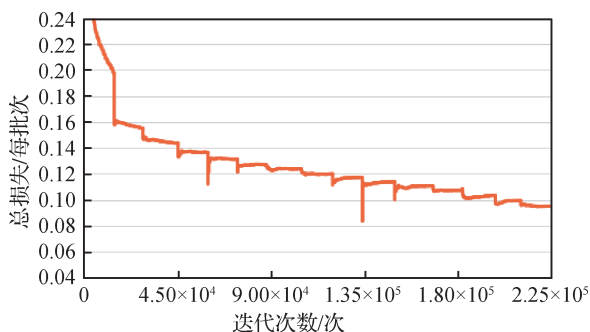
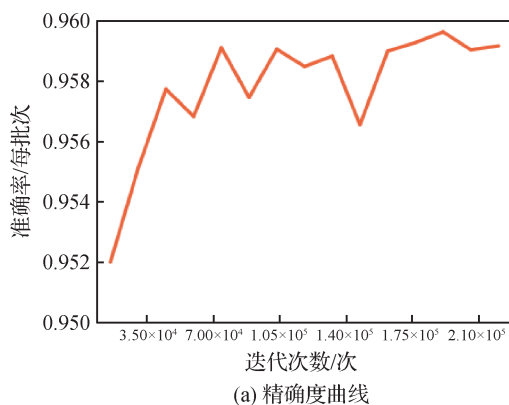


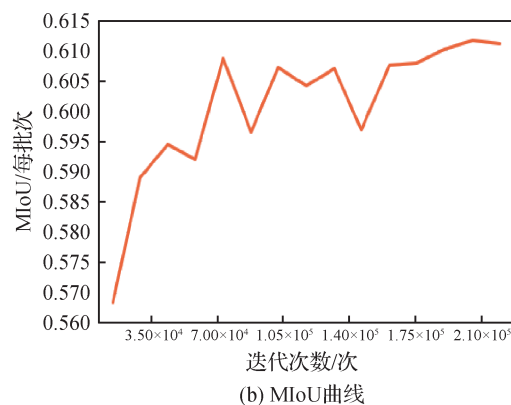
图9 总损失曲线

Fig. 9 Curve of total losses

通过计算每个类别上预测值与真实标签的IoU, 求得每个类别的平均交并比 (mean intersection over union, MIoU) (Garcia-Garcia 等, 2017) 为



(a) 精确度曲线



(b) MIoU曲线

图10 精确度和 MIoU 曲线

Fig. 10 Curves of precision and MIoU ((a) curves of precision; (b) curves of MIoU)

3.4 性能测试及对比

3.4.1 模型结构的有效性

提出的 ERFNet-AL 模型在测试集上的查准率 P 、查全率 R 及调和平均值 $F1$ 如表1所示, 可以看出, 正常场景的 $F1$ 值达到了 91.85%, 9 种场景 (包括交叉路口) 的平均值为 73.76%。

将 ERFNet-AL 模型的测试结果与 SCNN 模型、残差网络-34-自注意力蒸馏 (ResNet-34-self attention distillation, R-34-SAD)、R-101-SAD、ENet-SAD 模型 (Hou 等, 2019) 在 $F1$ 指标、 FP (仅限交叉路口)、运行时间 (处理 100 幅图像的平均时间) 和参数量等方面进行对比。

$F1$ 指标对比结果如图 11 所示, 图例中的 AL 表

$$M = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{i=0}^k p_{ji} - p_{ii}} = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{FN + FP + TP} \quad (9)$$

式中, M 为平均交并比, i 表示真实值, j 表示预测值, p_{ij} 表示将 i 预测为 j 的概率, $k+1$ 是类别个数 (包含空类)。

在验证集上进行验证, 精确度和 MIoU 变化曲线如图 10 所示。可以看出, 随着训练次数的增加, 验证集上的精确度先增后减再增, 最后趋于稳定状态, 当迭代次数达到 207.4×10^3 次时, 验证集上的精确度为 95.9%; 而 MIoU 值随着训练次数的增加, 逐渐稳定在 0.610 附近, 当迭代到 207.4×10^3 次时, 验证集上的 MIoU 达到了 0.6118, MIoU 值达到最大, 此次训练的模型为最优模型。

示加入了辅助损失, 横坐标表示场景, 对应数据集中除交叉路口以外的 8 种场景。可以看出, ERFNet-AL 模型在 8 种场景下的 $F1$ 指标都高于其他模型, $F1$ 平均值也明显高于其他模型, 可见该模型在车道线检测上具有一定优势。

ERFNet-AL 在交叉路口和非交叉路口场景的 FP 指标, 以及 9 种场景的平均 FP 如图 12 所示, 可以看出, 在占比例较大的拥挤、不存在车道线等场景, FP 较大。

各模型在交叉路口场景的 FP 指标如图 13 所示。可以看出, ERFNet-AL 模型在交叉路口的 FP 与其他模型的 FP 指标存在一定差距, 表明该模型在交叉路口的检测具有一定局限性, 需要进一步改进。

表 1 模型的指标对比
Table 1 Comparison of indicators of the model

场景	查准率	查全率	调和平均值 /%
正常	91.93	91.77	91.85
拥挤	72.84	71.12	71.97
夜晚	70.06	67.26	68.63
不存在车道线	48.98	44.33	46.54
阴影	74.62	71.77	73.17
箭头	88.45	85.45	86.92
炫目的灯光	67.41	64.33	65.84
曲线	75.63	62.20	68.26
平均	74.07	73.45	73.76

注:加粗字体为每列最优结果。

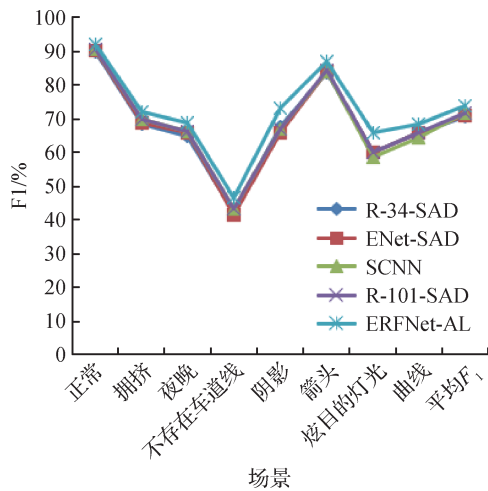


图 11 不同场景的 F1 指标对比

Fig. 11 Comparison of F1 indexes in different scenes

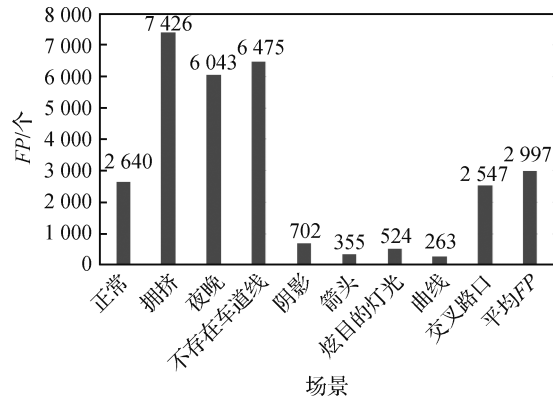


图 12 不同场景的 FP 指标对比

Fig. 12 Comparison of FP indexes in different scenes

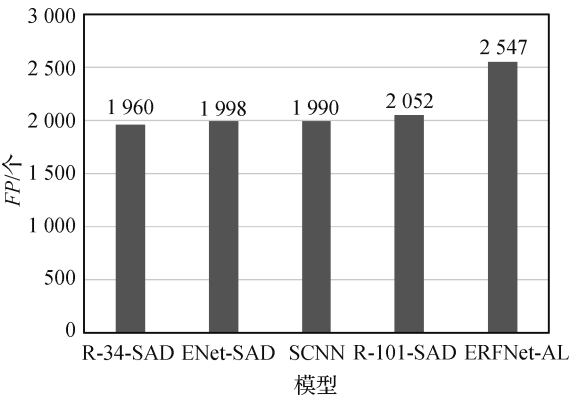


图 13 不同模型在交叉路口的 FP 指标对比

Fig. 13 Comparison of FP indexes of different models at the crossroads

各模型的运行时间和参数量的对比如表 2 所示,可以看出,ERFNet-AL 模型的参数量较少,运行时间最短。

表 2 不同模型的性能对比

Table 2 Comparison of performances of different models

模型	运行时间/ms	参数量/MB
R-34-SAD	50.50	22.70
ENet-SAD	13.40	0.98
SCNN	133.50	20.72
R-101-SAD	171.20	52.53
ERFNet-AL	11.10	2.49

注:加粗字体表示每列最佳结果。

从 F1 指标、FP、运行时间和参数量等指标对比可知,ERFNet-AL 模型在车道线检测上具有一定的优越性,可以满足车辆驾驶中对车道线检测的实时性和准确性要求。

3.4.2 辅助损失的有效性

采用控制变量法对加入和去掉辅助损失的模型的测试结果进行对比。交叉路口场景记录 FP 的数目,其余场景记录 F1 值,并计算 9 种场景的 F1 平均值,结果如表 3 所示。可以看出,加入辅助损失后,ERFNet-AL 模型的参数量为 2.49 MB;运行时间为 11.10 ms。与不加辅助损失相比,模型的参数量和运行时间几乎不变。在正常、拥挤、不存在车道线、阴影、箭头等场景的 F1 指标明显提高,交叉路口的 FP 指标明显下降,9 种场景 F1 的平均值也有所提高,表明加入辅助损失有助于车道线的检测。

表3 有无辅助损失性能对比
Table 3 Comparison of performances
with or without auxiliary loss

场景	ERFNet	ERFNet-AL
交叉路口/个	2 705	2 547
正常/%	91.63	91.85
拥挤/%	71.56	71.97
夜晚/%	68.77	68.63
不存在车道线/%	46.31	46.54
阴影/%	71.53	73.17
箭头/%	86.65	86.92
炫目的灯光/%	66.32	65.84
曲线/%	68.27	68.26
平均 F1/%	73.44	73.76
运行时间/ms	11.10	11.10
参数量/mB	2.49	2.49

注:加粗字体为每行最优结果。

3.4.3 检测结果的可视化

在测试集的每种场景中随机选取2幅图像运用提出的 ERFNet-AL 模型检测车道线拟合效果,4条车道线按从左到右依次用蓝、绿、红、墨绿色的线拟合,9种场景的原图与车道线拟合效果图对比如图14所示。可以看出,ERFNet-AL模型在正常、拥挤、箭头、炫目的灯光和曲线场景的拟合效果较好,在夜晚、不存在车道线、阴影和交叉路口场景的拟合效果需要进一步改善。

4 结 论

传统的车道线检测算法在物体遮挡、光线变化和阴影干扰等场景,检测的实时性和准确性不佳,针对这些问题,提出了嵌入辅助损失的多场景车道线检测模型——ERFNet-AL。在原有的 ERFNet 模型上增加辅助训练分支和车道预测分支,对9种实时场景的车道线图像进行语义分割,同时拟合出车道

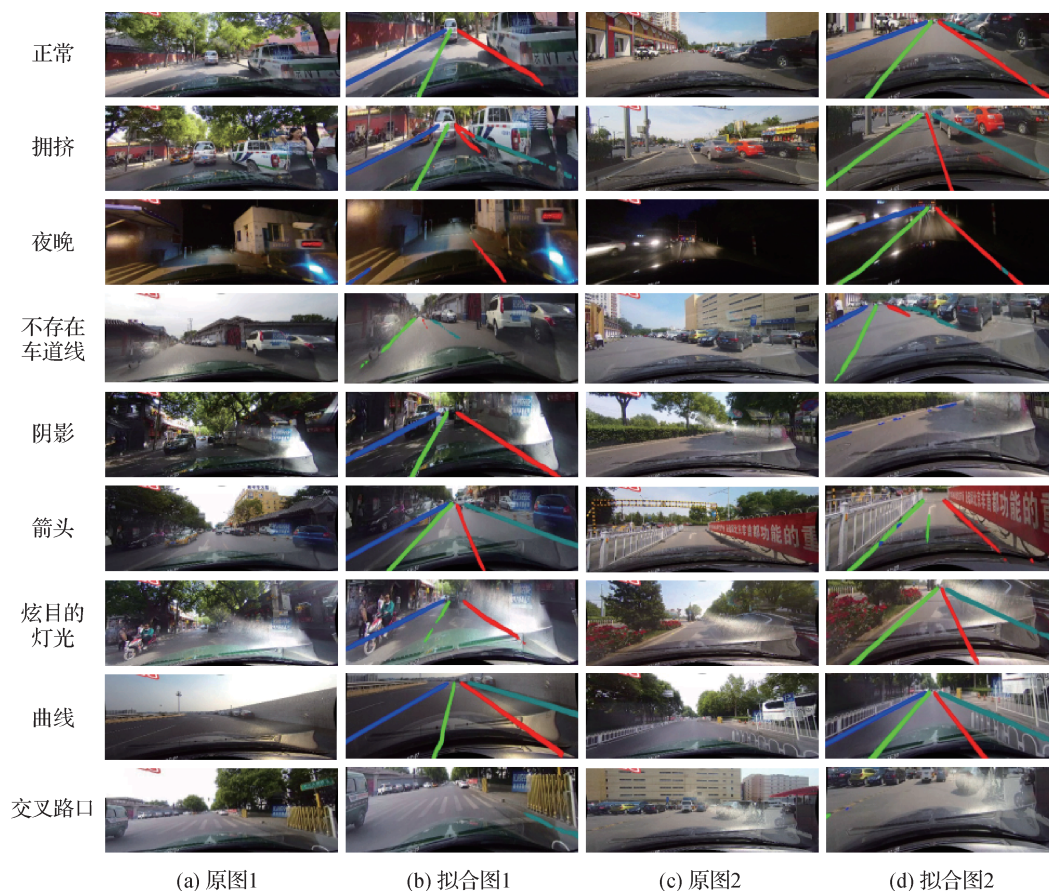


图14 车道线原图和拟合图

Fig. 14 Original images and fitted images of lane line

((a) the 1st original image; (b) the 1st fitting image; (c) the 2nd original image; (d) the 2nd fitting image)

线。主要贡献是以 ERFNet 网络为基础,在编码器之后分别添加辅助训练分支和车道预测分支,利用辅助损失,解决了梯度消失问题,优化了学习过程。在 CULane 公共数据集上,与 SCNN、R-34-SAD、R-101-SAD 和 ENet-SAD 模型的测试结果相比,本文模型在各场景(除交叉路口)的 F1 指标均有所提升;模型的参数量较少,仅有 2.49 MB;测试单幅图像的平均运行时间最短。表明在物体遮挡、光照变化和阴影干扰等场景,本文模型具有良好的抗干扰能力。

实验的不足之处是:在交叉路口场景,模型的检测效果不太理想,原因是交叉路口的车道线多于 4 条,而实验中的模型只关注 4 条车道线。

下一步的工作是增加车道线的检测数目以及寻找标注交叉路口的车道线的方法,使之更好地适用于交叉路口的车道线检测。

参考文献 (References)

- Badrinarayanan V, Kendall A and Cipolla R. 2017. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12): 2481-2495 [DOI: 10.1109/tpami.2016.2644615]
- Chen C Y. 2018. Research of Highway Rainy Weather Detection Based on Deep Semantic Segmentation. Chengdu: Southwest Jiaotong University (陈昌宇. 2018. 基于深度分割网络的高速公路监控视频雨天检测算法研究. 成都: 西南交通大学)
- Chen H S, Yao M H, Chen Z H and Yang Z. 2018. Efficient method of lane detection based on multi-frame blending and windows searching. *Computer Science*, 45(10): 255-260 (陈涵深, 姚明海, 陈志浩, 杨圳. 2018. 基于多帧叠加和窗口搜索的快速车道检测. 计算机科学, 45(10): 255-260)
- Chen Z and Chen Z J. 2017. RBNET: a deep neural network for unified road and road boundary detection//Proceedings of the 24th International Conference on Neural Information Processing. Guangzhou: Springer: 677-687 [DOI: 10.1007/978-3-319-70087-8_70]
- Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, Franke U, Roth S and Schiele B. 2016. The cityscapes dataset for semantic urban scene understanding//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE: 3213-3223 [DOI: 10.1109/cvpr.2016.350]
- Csurka G and Perronnin F. 2011. An efficient approach to semantic segmentation. *International Journal of Computer Vision*, 95(2): 198-212 [DOI: 10.1007/s11263-010-0344-8]
- Gao F, Mei K C, Gao Y, Lu S F and Xiao G. 2016. Algorithm of intersection background extraction and driveway calibration. *Journal of Image and Graphics*, 21(6): 734-744 (高飞, 梅凯城, 高炎, 卢书芳, 肖刚. 2016. 城市交叉路口背景提取与车道标定算法. 中国图象图形学报, 21(6): 734-744) [DOI: 10.11834/jig.20160606]
- Garcia-Garcia A, Orts-Escolano S, Oprea S, Villena-Martinez V and Garcia-Rodriguez J. 2017. A review on deep learning techniques applied to semantic segmentation [EB/OL]. [2020-01-03]. <https://arxiv.org/pdf/1704.06857.pdf>
- Ghafoorian M, Nugteren C, Baka N Booi O and Hofmann M. 2018. EL-GAN: embedding loss driven generative adversarial networks for lane detection//Proceedings of 2018 European Conference on Computer Vision. Munich: Springer: 256-272 [DOI: 10.1007/978-3-030-11009-3_15]
- Hou Y N, Ma Z, Liu C X and Loy C C. 2019. Learning lightweight lane detection CNNs by self attention distillation//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE: 1013-1021 [DOI: 10.1109/ICCV.2019.00110]
- Jiang L B and Tai Q L. 2019. The lane line detection in complex scene based on instance segmentation. *Machine Design and Manufacturing Engineering*, 48(5): 113-118 (姜立标, 台敬龙. 2019. 基于实例分割方法的复杂场景下车道线检测. 机械设计与制造工程, 48(5): 113-118) [DOI: 10.3969/j.issn.2095-509X.2019.05.027]
- Ling S Y, Ma Y, Huang C R and Zhai W L. 2017. Research on improved urban environment road detection algorithm based on Hough transform. *Machine Design and Manufacturing Engineering*, 46(12): 71-75 (凌诗韵, 马乐, 黄楚然, 翟伟良. 2017. 基于 Hough 变换的城市环境道路识别优化算法研究. 机械设计与制造工程, 46(12): 71-75) [DOI: 10.3969/j.issn.2095-509X.2017.12.017]
- Pan X G, Shi J P, Luo P, Wang X G and Tang X O. 2018. Spatial as deep: spatial CNN for traffic scene understanding [EB/OL]. [2020-01-03]. <https://arxiv.org/pdf/1712.06080.pdf>
- Romera E, Álvarez J M, Bergasa L M and Roberto A. 2017. Efficient ConvNet for real-time semantic segmentation [EB/OL]. [2020-01-03]. <https://arxiv.org/pdf/1712.06080.pdf>
- Romera E, Álvarez J M, Bergasa L M and Roberto A. 2018. ERFNet: efficient residual factorized ConvNet for real-time semantic segmentation. *IEEE Transactions on Intelligent Transportation Systems*, 19(1): 263-272 [DOI: 10.1109/TITS.2017.2750080]
- Shi Y T. 2018. Research on semantic segmentation of road scene based on deep neural networks. Chengdu: Southwest Jiaotong University (石永涛. 2018. 基于深度神经网络的道路场景语义分割研究. 成都: 西南交通大学)
- Tian X, Wang L and Ding Q. 2019. Review of image semantic segmentation based on deep learning. *Journal of Software*, 30(2): 440-468 (田萱, 王亮, 丁琪. 2019. 基于深度学习的图像语义分割方法

综述. 软件学报, 30(2): 440-468 [DOI: 10.13328/j.cnki.jos.005659]

Trinh T H, Dai A M, Luong M T and Le Q V. 2018. Learning longer-term dependencies in RNNs with auxiliary losses[EB/OL]. [2020-01-03]. <https://arxiv.org/pdf/1803.00144.pdf>

Xing Y, Lv C, Chen L, Wang H J, Wang H, Cao D P, Velenis E and Wang F Y. 2018. Advances in vision-based lane detection: algorithms, integration, assessment, and perspectives on ACP-based parallel vision. IEEE/CAA Journal of Automatica Sinica, 5(3): 645-661 [DOI: 10.1109/JAS.2018.7511063]

Zhao W M and Zhang H W. 2017. A research on lane maintenance assist system based on machine vision. Digital Technology and Application, (11): 63-64 (赵文明, 张海文. 2017. 基于机器视觉的车道保持辅助系统研究. 数字技术与应用, (11): 63-64) [DOI: 10.19695/j.cnki.cn12-1369.2017.11.035]

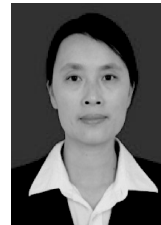
Zhu W, Qu J Y and Wu R B. 2017. Straight convolutional neural networks algorithm based on batch normalization for image classification. Journal of Computer-Aided Design and Computer Graphics, 29(9): 1650-1657 (朱威, 屈景怡, 吴仁彪. 2017. 结合批归一化的直通卷积神经网络图像分类算法. 计算机辅助设计与图形学学报, 29(9): 1650-1657) [DOI: 10.3969/j.issn.1003-9775.2017.09.008]

作者简介



陈立潮, 1961年生, 男, 教授, 主要研究方向为大数据软件工程、智能图像信息处理。

E-mail: chen_lc@263.net



曹建芳, 通信作者, 女, 教授, 主要研究方向为数字图像理解、大数据技术。

E-mail: kcxjdj122@126.com

徐秀芝, 女, 硕士研究生, 主要研究方向为智能图像信息处理。E-mail: 2627084603@qq.com

潘理虎, 男, 副教授, 主要研究方向为智能信息处理。

E-mail: panlh@tyust.edu.cn