

光电工程

Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

轴注意力引导的锚点分类车道线检测

罗鑫, 黄影平, 梁振明

引用本文:

罗鑫, 黄影平, 梁振明. 轴注意力引导的锚点分类车道线检测[J]. 光电工程, 2023, 50(7): 230079.

Luo X, Huang Y P, Liang Z M. Axial attention-guided anchor classification lane detection[J]. *Opto-Electron Eng*, 2023, 50(7): 230079.

<https://doi.org/10.12086/oe.2023.230079>

收稿日期: 2023-04-07; 修改日期: 2023-06-27; 录用日期: 2023-07-11

相关论文

基于语义分割的实时车道线检测方法

张冲, 黄影平, 郭志阳, 杨静怡

光电工程 2022, 49(5): 210378 doi: 10.12086/oe.2022.210378

基于点云与图像交叉融合的道路分割方法

张莹, 黄影平, 郭志阳, 张冲

光电工程 2021, 48(12): 210340 doi: 10.12086/oe.2021.210340

雾对基于可见光的车辆定位性能的研究

张颖, 杨晶, 杨玉峰

光电工程 2020, 47(4): 190661 doi: 10.12086/oe.2020.190661

更多相关论文见光电期刊集群网站 



<http://cn.ojournal.org/oe>



 OE_Journal



Website

DOI: 10.12086/oe.2023.230079

轴注意力引导的锚点分类 车道线检测

罗鑫, 黄影平*, 梁振明

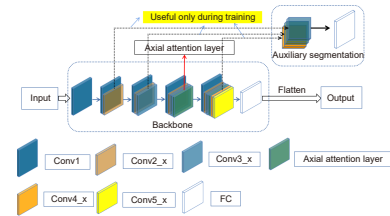
上海理工大学光电信息与计算机工程学院, 上海 200093

摘要: 由于车道线的多样性以及交通场景的复杂性等问题, 车道线检测是一项具有挑战性的任务。其主要表现在当车辆行驶在拥堵、夜晚、弯道等车道线不清晰或被遮挡的道路上时, 现有检测方法的检测结果并不理想。本文基于检测方法的框架提出了一种轴注意力引导的锚点分类车道线检测方法来解决两个问题。首先是车道线不清晰或缺失时存在的视觉线索缺失问题。其次是锚点分类时用混合锚点上的稀疏坐标表示车道线带来的特征信息缺失问题, 从而导致检测精度下降, 所以通过在骨干网络中添加轴注意力层来聚焦行向和列向的显著特征来提高精度。在 TuSimple 和 CULane 两个数据集上进行了大量实验。实验结果表明, 本文方法在各种条件下都具有鲁棒性, 同时与现有的先进方法相比, 在检测精度和速度方面都表现出综合优势。

关键词: 车道线检测; 锚点分类; 轴注意力; 自动驾驶

中图分类号: TP391.4

文献标志码: A



罗鑫, 黄影平, 梁振明. 轴注意力引导的锚点分类车道线检测 [J]. 光电工程, 2023, 50(7): 230079

Luo X, Huang Y P, Liang Z M. Axial attention-guided anchor classification lane detection[J]. *Opto-Electron Eng*, 2023, 50(7): 230079

Axial attention-guided anchor classification lane detection

Luo Xin, Huang Yingping*, Liang Zhenming

School of Opto-Electronic and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

Abstract: Lane detection is a challenging task due to the diversity of lane lines and the complexity of traffic scenes. The detection results of the existing detection methods are not ideal when the vehicle is driving in congestion, at night, and the lane lines are not clear or blocked on the road such as curves. Based on the framework of detection methods, a method that axial attention-guided anchor classification lane detection is proposed to solve two problems. The first is the problem of missing visual cues when lane lines are unclear or missing. The second problem is the lack of feature information caused by using sparse coordinates on mixed anchors, which leads to a decline of detection accuracy. Therefore, an axial attention layer is added to the backbone network to focus on prominent features of the row and column directions to improve the accuracy. Extensive experiments are conducted on the TuSimple and CULane datasets. Experimental results show that the proposed method is robust under various conditions while showing comprehensive advantages in terms of detection accuracy and speed compared with existing advanced methods.

Keywords: lane detection; anchor classification; axial attention; autonomous driving

收稿日期: 2023-04-07; 修回日期: 2023-06-27; 录用日期: 2023-07-11

基金项目: 上海市自然科学基金项目 (20ZR1437900); 国家自然科学基金面上项目 (61374197)

*通信作者: 黄影平, huangyingping@usst.edu.cn。

版权所有©2023 中国科学院光电技术研究所

1 引言

车道线检测是自动驾驶汽车在环境感知方面的重要功能。虽然关于车道线检测算法的研究由来已久,但是现有算法在实际应用中仍然面临诸多挑战。例如在拥堵、夜晚以及弯道等交通环境复杂的场景下,车道线会出现缺失、模糊或者被遮挡的情况,现有车道线检测模型对这些场景中的车道线检测效果不佳。

现有的车道线检测方法可以分为三种:经典方法,基于深度学习的方法以及基于激光雷达的方法。近年来,由于卷积神经网络强大的表征能力,基于深度学习的方法成为车道线检测的主流方法。基于深度学习的方法主要分为基于分割、基于检测和基于参数曲线三类。基于分割方法将车道线检测考虑成图像分割任务,通过对场景图像中的每个像素分配语义标签来区分车道线和背景。基于检测的方法通常是车道线转化为块或点等可学习的结构化表示,再使用检测模型对这些结构特征进行检测,以此作为车道线的表示。基于参数曲线的方法将车道曲线表达为一些抽象的多项式曲线,再使用曲线拟合方法来匹配车道线。基于分割的方法精度相对较高但计算成本太大,基于检测的方法虽然速度较快却在精度上有所欠缺,

基于曲线拟合的方法由于曲线模型的抽象性质,其性能无论在精度还是效率上都要落后于基于检测和分割的方法。为解决上述问题,本文遵循基于检测的方法框架,提出了一种将轴注意力与锚点分类结合的方法来检测车道线。基本思想是把车道线分段为断续的点块,将车道线检测问题转化为对车道线锚点的检测问题。在实施过程中,采用行锚点和列锚点取代逐像素点的分割不仅可以提高车道线检测速度,还可以改善车道线视觉线索缺失的问题。在网络结构上,将轴注意力机制添加到特征提取网络中,可以更有效地提取特征并过滤掉多余特征,以改善基于检测方法精度较低的问题。这项工作的主要贡献可以概括如下:

1) 利用行锚点和列锚点的方法取代逐像素点分割,有效提高了车道线检测的速度。另外,研究表明,该方法可以改善当车辆行驶在车道标志不明显或缺失的道路上时(如拥堵、夜晚和弯道)存在的无视觉线索问题。

2) 通过在骨干网络 ResNet 中添加轴注意力层可以更好地提取车道线特征,从而改善基于检测的方法

检测车道线时存在的精度较低问题。

3) 我们在 TuSimple 数据集^[1]和 CULane 数据集^[2]上进行了大量的实验来验证提出的方法能够提高车道线检测性能。实验结果表明,该方法在精度和效率上都取得了较好的效果。

2 相关工作

2.1 车道线检测方法

2.1.1 经典方法

经典方法主要是从图像中获取具有特定特征的识别线段并使用预定义的曲线模型来对这些特征线段进行拟合,其中主流方法有基于霍夫变换的车道线检测,基于俯视图变换的车道线检测等。例如 Huang 等人^[3]对经过透视反变换得到的俯视图图像进行处理,然后利用卡尔曼滤波器实现对车道线的跟踪。经典方法的思想一般都是通过图像处理得到视觉线索来检测车道线,当视觉线索丢失时,则需要尝试一些后处理解决方案,如马尔可夫模型和条件随机场模型^[4],在机器学习领域也可以采用模板匹配,支持向量机等算法^[5]。尽管在这方面已有大量的研究,但由于驾驶环境的复杂性和多样性,经典车道线检测方法具有较大的局限性。例如,在交通场景比较混乱的情况下,不仅有较大的图像噪声而且车道线视觉线索容易丢失,而上述后处理方法不能良好解决车道线视觉线索丢失问题,这就最终导致经典算法难以实现精确且快速的车道线检测任务。

2.1.2 基于深度学习的方法

基于深度学习的方法主要是利用网络模型自动学习车道线特征并对其拟合。该类方法因对图像噪声具有良好的鲁棒性,近年来受到越来越多的关注。其大致可以分为三类:基于分割、基于检测和基于参数曲线。

基于分割的方法利用深度神经网络强大的特征提取能力对车道线特征进行提取,将车道线检测考虑成图像分割任务,即为场景图像中的每个像素分配语义标签,并通过像素级别的分类对车道线和背景做区分。该类方法凭借较高的精度一直活跃在车道线检测领域。其中比较有代表性的工作是 Pan 等人^[2]提出的 SCNN,其将传统的逐层连接转换为在特征图中逐层连接的形式。它能够在图像像素中实现行和列之间的信息传递,适用于车道线等远距离物体的检测。为了提高检测速度, Hou^[6]等人提出了一个自注意力蒸馏(SAD)模块

来聚合上下文信息, 该模块虽然允许使用更轻量级的主干来检测车道线, 但是计算成本仍然较高。Neven 等人^[7]通过将分割方法与聚类算法相结合, 将车道线检测任务视为实例分割问题。然而, 该方法在车道线标识缺失或被遮挡的情况下效果不佳。为解决这一问题, Guo 等人^[8]提出一种实例分割方法并开发了一种双注意力网络 DALaneNet 用于车道线检测, 其中空间注意力学习像素之间的位置关系, 通道注意力学习重要通道的特征。尽管取得了良好的效果, 但是计算资源进一步加大。张冲等人^[9]设计了一个基于编码解码结构的轻量级车道线分割网络, 针对卷积神经网络计算量大的问题, 引入深度可分离卷积来替代普通卷积以减少卷积运算量。然而, 计算量仍然较大。总而言之, 基于分割的方法通过逐像素点检测车道线特征能够实现高精度检测, 但是由于计算成本和时间成本比较高导致检测效率较低。

基于检测的方法通常是车道线段转化为块或点等可学习的结构化表示, 并使用检测模型对这些块或点进行检测, 以此作为车道线表示。在基于检测的方法中, 随着 Yoo 等人^[10]首次提出对车道线做基于行的车道标记并检测之后, Qin 等人^[11]将车道线检测过程看作是在全局图像上使用大区域的基于行的选择问题。它很好地解决了速度慢的问题, 但对弯道处车道线检测效果不佳。为了解决弯道处车道线检测问题, Qin 等人^[12]提出基于列的选择方法并取得了一定的效果, 但是在 CULane 数据集上的精度与基于分割的方法仍然存在差距。Zheng 等人^[13]指出车道线具有特定的局部信息, 需要详细的低层特征才能精确定位, 所以文章提出跨层细化网络 (CLRNet) 希望在车道线检测中充分利用高层和低层特征, 首先利用高层语义特征检测车道线, 然后基于低层特征进行细化。LaneATT^[14]网络指出当车道线不清晰或者缺失时, 全局信息对于车道线检测是很重要的, 为此文章提出在结构化表示的块或点之间应用注意力机制来获取全局信息, 在各种数据集上都取得了良好的效果。

另外还有基于参数曲线的方法, 其基本思想是将车道线赋予一种合适的数学模型, 并基于该模型对车道线进行拟合, 基本原理就是根据车道线的几何特征为车道线匹配合适的曲线模型。与最先进的基于分割和基于检测的方法不同, 基于参数曲线的方法可以自然地学习整体车道线表示^[15]。直接对车道线整体进行学习显然要比将整体车道线视为像素点的集合或者将

车道线转化为可学习的结构化表示更加方便, 但是实际场景中的车道线错综复杂, 对参数曲线的多项式系数要求极高, 而抽象的多项式系数很难通过计算机去学习。之前的研究表明, 该类方法的性能落后于设计良好的基于分割和基于检测的方法。

2.1.3 基于激光雷达的方法

基于激光雷达的车道线检测方法主要利用激光雷达对物体的反射强度信息不同来得到车道线的特征信息, 然后对车道线进行拟合。主要方法有四种, 一是根据激光雷达回波宽度, 一般路面和车道线的回波宽度会有明显差异; 二是根据激光雷达反射强度信息形成的灰度图, 或者根据强度信息与高程信息配合, 过滤出无效信息; 三是激光雷达与高精度地图配合, 在检测车道线的同时还可以进行自行车定位; 四是激光雷达能够利用物体反射信息不同的特性来获取路沿高度信息, 假设道路宽度是已知的, 那么可以先检测出路沿再根据距离推算出车道线位置^[16]。该类方法可以很好地解决车道线不完整或者被遮挡的问题以及由于外部环境条件引起的图像质量较差的问题, 检测速度快且检测精度高。但是激光雷达的费用昂贵, 从工程造价方面考虑, 不宜使用在车道线检测上。

2.2 注意力机制

注意力机制类似于人类的感知系统, 能够充分利用局部信息关注场景中的显著特征。人类在看东西的时候一般不会既看整体又看细节, 往往只会根据需求注意并观察特定的一部分, 将其应用在神经网络中就是一种权重参数的分配机制, 将计算资源分配给更重要的任务。通过引入注意力机制, 在众多的输入信息中聚焦于对当前任务更为关键的信息, 降低对其他信息的关注度, 甚至过滤掉无关信息, 就可以解决信息过载问题, 并提高任务处理的效率和准确性。

从原理上讲, 注意力机制主要分为空间注意力模型、通道注意力模型和空间与通道混合注意力模型。空间注意力就是寻找网络中最重要的部分进行处理, 也就是说不是图像中所有的区域对任务的贡献都是同样重要的, 只有与任务相关的区域才是需要关心的; 通道注意力通过赋予每个通道不同的权重, 以增强判断重要特征通道的能力; 在混合注意力机制中, 通道注意力和空间注意力可以通过串联或者并联的方式进行组合。Wang 等人^[17]提出了一种带有空间注意力机制的剩余注意力网络, 使用下采样和上采样进行图像分类。SeNet^[18]通过使用全局平均池化对通道依赖进行

建模, 以强调重要的特征映射。在这些工作的影响下, CBAM^[19]、DANet^[20] 等系列工作将空间注意力和通道注意力融合为一个模块, 在很多领域都证明了他们工作的有效性。受这些成功的注意力机制应用的启发, 我们提出将轴注意力机制应用于深度卷积神经网络来提取车道线的显著特征, 从而改善基于检测的方法存在的精度低问题。轴注意力的具体原理详见 3.3 节。

3 本文方法

本文遵循基于检测方法的框架将车道线段转化为可学习的结构化表示, 再对这些可学习的结构化表示进行检测并分类, 以此作为车道线表示。首先对图像进行预处理, 预处理主要是针对数据集中的图像和标签做了一些处理。对于图像主要是平移操作, 小幅度旋转等; 车道线标签存在长短不一的情况, 需要对标签进行延伸。具体来说, 就是针对标注数据中部分车道线标签中断的情况, 通过一个线性拟合来得到延伸的标签。预处理完成后将图像中包含车道线的部分从行或者列进行划分并将其网格化, 得到可学习的结构化表示, 我们将其称为锚点, 以行划分称为行锚, 其对应的网格称为行锚点, 以列划分称为列锚, 其对应的网格称为列锚点。然后在骨干网络 ResNet 中添加轴注意力机制对车道线行锚点和列锚点进行检测, 将检测结果输入到 MLP 分类器对车道线锚点进行分类, 其中左前车道和右前车道在行维度上划分并分类, 左侧车道和右侧车道在列维度上划分并分类。MLP 分类器有两个输出分支, 存在分支用于判断是否有车道线锚点存在, 定位分支用于确定车道线锚点的位置, 最终通过两个分支的合并得到车道线的检测结果。网络架构详见 3.2 节。

3.1 锚点分类

3.1.1 基本思想

首先需要将车道线段转化为可学习的结构化表示, 即锚点。我们将图像中包含车道线的部分 (即感兴趣区域) 划分为特定数量的行或者列, 并将每一行或每一列分成许多单元格。假设图像大小为 $H \times W$, 通常预定义划分的行或者列的数量远远小于图像的 H 和 W 。网格化之后, 整体的车道线被划分为许多包含车道线标识的断续单元格分布在每一行或每一列, 再从所有单元格中预测出最有可能包含车道线标识的单元格。这样车道线的检测就可以描述为在预定义的行向或列向单元格中选择包含车道线标识的单元格, 然后将属于同一条车道线标识的特定单元格进行合并归类。以行锚点为例, 首先将图像中感兴趣区域划分为特定数量的行, 即行锚。为了表示车道线的具体位置点, 我们将每个行划分为许多单元格, 每一个单元格即一个行锚点。如图 1 所示。其基本原理是对于每一行, 模型预测出最有可能包含车道标记的单元格, 每一行中如果存在多条车道线的标记的话就需要使用不同单元格来区分。行锚点分类就是把所有预定义行中包含车道线标识的单元格检测出来并分类, 属于同一车道线的单元格为一类。

与行锚点类似, 列锚点的处理方法是图像中的感兴趣区域划分为一系列位于预定义列上的垂直位置点, 同样将图像在每一列上划分出许多单元格, 每一个单元格就是一个列锚点。再在预定义好的列锚上选择特定单元格 (即包含车道线标识的列锚点) 进行分类, 即列锚点分类。将图像分别从行和列两个维度来划分锚点主要是因为行锚适合处理垂直化车道, 而列锚适合处理水平化车道。接下来我们将进一步分析和讨论

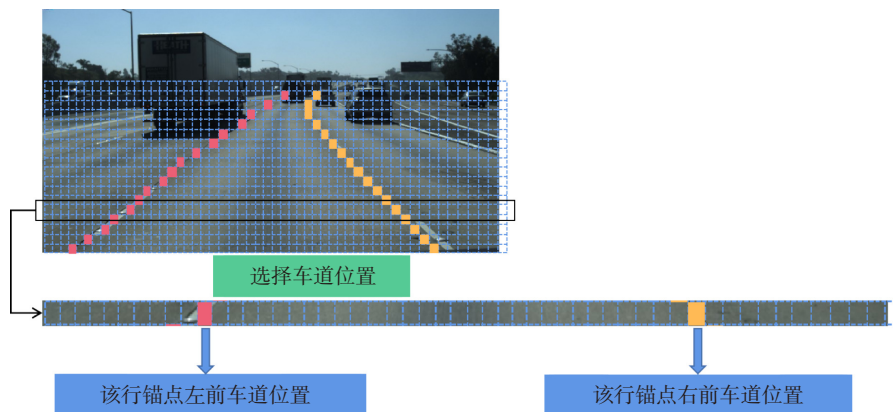


图 1 行锚点划分示意图

Fig. 1 Schematic diagram of the anchor division in a row

如何根据需求选择适当的行锚和列锚, 并解释为什么需要混合使用它们。

3.1.2 行锚和列锚的选择

经过预处理后, 一条完整的车道线会被划分为许多具有车道线标识的单元格分布在所有单元格中。如果一条完整车道线的所有标识单元格在每一行或每一列的分布越好的话, 那么后续对这条车道线的检测效果越好, 也就是说一条车道线在预定义的行锚或列锚产生的交点越多, 后续检测效果越好。所以我们需要对不同车道线判断它适合使用行锚点分类还是列锚点分类。以四条车道线为例, 如图 2(a) 所示, 将其定义为左侧车道、左前车道、右前车道和右侧车道。

当车辆行驶在笔直的道路时, 视角中的车道线近似于直线且趋于垂直化, 尤其是左前车道和右前车道, 如图 2(b) 所示。为了确定这两条车道线适合行锚还是列锚, 我们将这两条车道线分别与预定义的行锚和列锚相交, 从图 2(b) 中可以看出这两条车道线与行锚有 4 个交点, 与列锚有 3 个交点, 行锚效果要略优于列锚效果。在这种情况下, 使用行锚点效果将会更

好。然而, 当车辆行驶在弯道时, 视线里的车道线会趋于水平化, 尤其是左侧车道和右侧车道, 如图 2(c) 所示。同样将这两条车道线分别与预定义的行锚和列锚相交, 可以发现这两条车道线与行锚只有 2 个交点, 而与列锚有 3 或 4 个交点, 这时列锚与完整车道线产生的交点多于行锚与完整车道线产生的交点, 所以我们考虑对左侧车道和右侧车道采用列锚点分类。

行锚不适合弯道处车道主要是因为弯道处的车道线在视角中会趋于水平化, 如果对趋于水平化的车道线继续使用行锚分类会在车道线定位时产生较大的误差, 如图 3 所示。角度 θ 越小则表示越趋于水平化, 此时行锚上的放大误差带就会越大, 车道线的预测值与真值的偏差也就越严重。

如果只使用行锚, 虽然能良好检测直线行驶时的垂直化车道, 但是对水平化车道 (左侧车道和右侧车道) 的检测效果不佳, 只使用列锚的话尽管可以较好检测水平化车道, 但是常见的垂直化车道 (左前车道和右前车道) 反而难以有效检测。总之, 不同的锚点系统对车道线有不同的效果。具体来说, 由于垂直化车道在水平方向上的差异较大, 行锚与车道线相交更

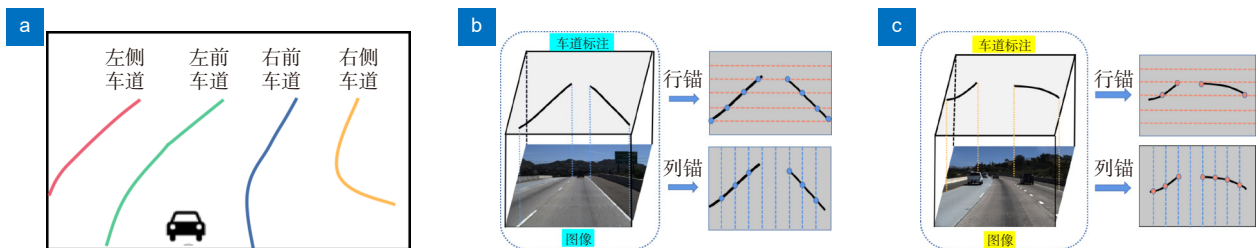


图 2 车道线定义及行锚与列锚选择说明。(a) CULane 数据集^[2]对车道线的定义; (b) 左前车道和右前车道; (c) 左侧车道和右侧车道

Fig. 2 Description of lane line definition and selection of row anchor and column anchor. (a) The definition of lane line in CULane dataset^[2]; (b) Left ego lane and right ego lane; (c) Left side lane and right side lane

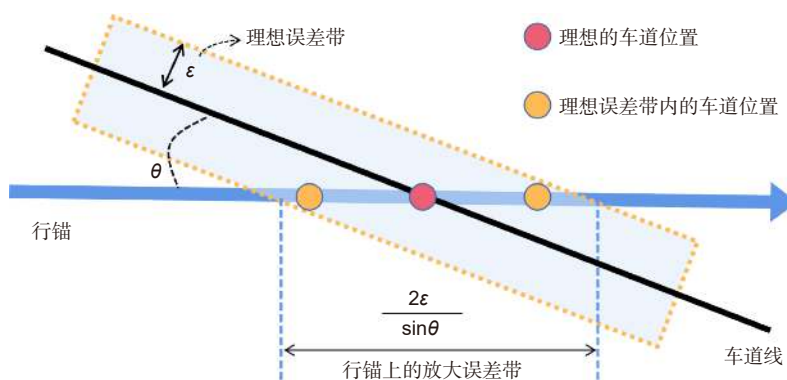


图 3 定位误差产生示意图

Fig. 3 Schematic diagram of positioning error generation

容易形成有车道线标识的锚点, 因此行锚更适用于趋于垂直化的车道。对于趋于水平化的车道线也是如此, 利用列锚能实现更好的检测效果。所以, 本文将行锚和列锚混合使用, 对左前车道和右前车道使用行锚分类, 对左侧车道和右侧车道使用列锚分类。后续实验也表明, 对不同类型的车道线采用不同的锚比单独使用行锚更有效。

该方法将车道线检测问题转化为对锚点的检测问题, 比基于分割方法速度更快。另外在全局图像上划分锚点具有更大的感受野, 更利于获取全局信息, 从而可以改善车辆行驶在拥堵、夜晚和弯道等道路上时存在的无视觉线索问题。然而, 这种行锚点和列锚点的稀疏结构化表示会丢失部分特征信息, 为此我们提出在骨干网络中融合轴注意力机制来聚焦显著特征, 以此来改善特征信息丢失的问题, 提高检测精度。

3.2 网络架构

本文采用的网络框架如图 4 所示。首先利用骨干网络对图像中的车道线锚点特征进行检测。然后将骨干网络提取出来的特征图做平铺操作后输入到 MLP 分类器中进行分类。该分类器有两个输出分支, 即存在分支和定位分支。存在分支用来判断输入特征图中四条车道线是否都存在, 定位分支则对输入特征图中的每条车道线的锚点进行定位, 将同属于一条车道线的锚点归为一类。最终将存在分支和定位分支进行汇总, 排除掉不存在的车道线就得到了车道线的输出。

骨干网络由 ResNet-34 和轴注意力模块组成, 如图 5 所示。ResNet 由 He 等人^[21] 于 2015 年提出, 其基本思想是利用 bottleneck block 来实现更深的网络。其内部的 bottleneck block 使用了跳跃连接, 缓解了深度神经网络中增加深度带来的梯度消失问题。输入图像先经过一个初始块(由卷积核 7×7、步长为 2 的卷积层, 批量规范化层 (BN) 和 ReLU 激活层组成), 然后依次使用一个最大池化层和四个中间卷积层来逐渐降低分辨率。四个中间卷积层分别由 n 个 bottleneck block 构成, bottleneck block 由卷积核分别为 1×1、3×3、1×1 的三个卷积层组成。ResNet 有多种体系结构, 本文采用 ResNet-34, 四个中间卷积层的 bottleneck block 数目 n 分别为 3、3、5、2。我们在中间卷积层 2 和中间卷积层 3 之间添加了轴注意力层, 另外有三个中间卷积层的特征图会被应用到辅助分割模块中, 辅助分割模块只用于训练过程, 目的是提高卷积层特征提取能力。最后经过所有的中间卷积层后再通过全连接层就得到了特征图输出。

3.3 轴注意力层

如图 5 所示, 轴注意力层被添加到中间卷积层 2 和中间卷积层 3 之间, 目的是通过在特征提取时聚焦显著特征来改善基于检测的方法存在的特征信息丢失问题。轴注意力机制由 Wang 等人^[22] 首次提出, 其核心思想是将二维注意力依次沿着竖直方向和水平方向分解为两个一维注意力来执行更大区域甚至全局区域的注意力。然后 Ramachandran 等人^[23] 将轴注意力应

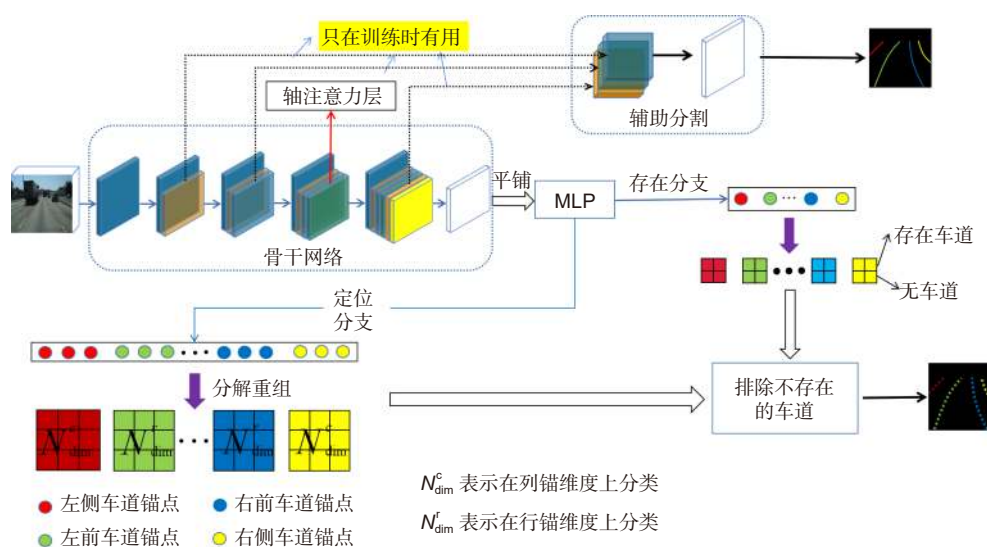


图 4 网络架构的说明

Fig. 4 Description of the network architecture

用到了 Transform 架构中, 并实现了良好的效果。考虑到本文车道线检测任务从行锚和列锚两个角度出发, 所以在特征提取骨干网络中将轴注意力与 ResNet 结合以此来聚焦行向和列向的显著特征。轴注意力结构示意图如图 6 所示。

具体来说, 首先在图像的 W 轴上定义一个轴向注意力层, 简单地将其定义为一维的自我注意力层, 并对 H 轴使用类似的定义。一个自我注意力层沿一个特定的轴传播信息, 并建立起该方向的特征信息联系。

为了获取更多特征信息, 分别对 W 轴和 H 轴连续使用了两个采用多头注意力机制的轴向注意力层。轴注意力原理如图 7 所示。假设现在需要得到图中黑色点与其他位置点的特征关系, 先计算沿 H 轴方向的关系特征点, 再计算沿 W 轴方法的关系特征点, 即图中红色点, 都计算完之后则认为黑色点的特征重构完成, 再计算下一位置点。后续实验表明, 在特征提取网络中添加轴注意力机制可以有效提高车道线检测模型的性能。

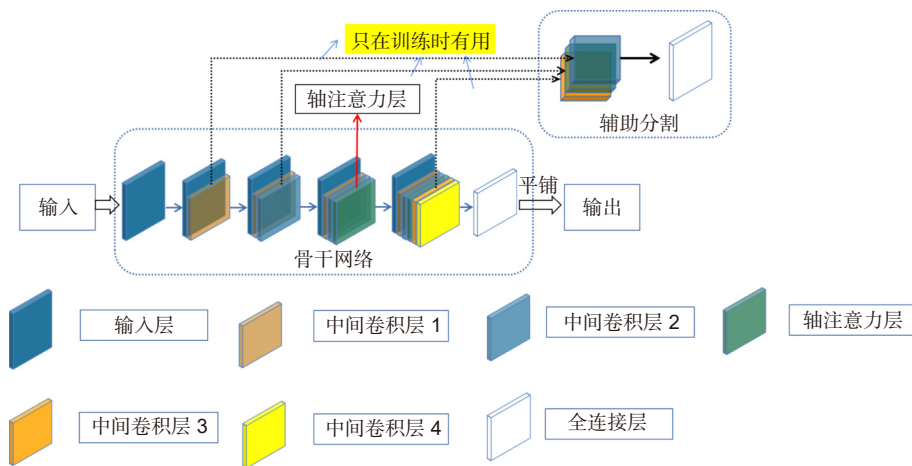


图 5 骨干网络细节图

Fig. 5 Details of the backbone network

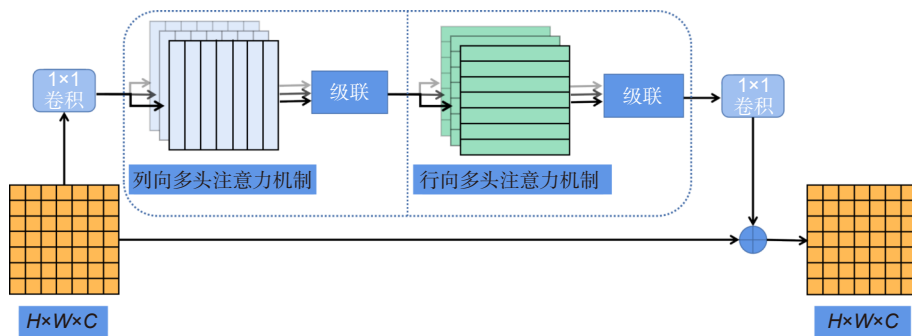


图 6 轴注意力结构示意图。其中包括了行向和列向的两个多头注意力机制

Fig. 6 Schematic diagram of the attention structure of the axial. It includes two multi-head attention mechanisms

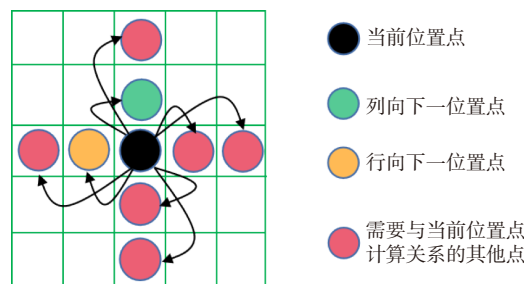


图 7 轴注意力原理解释图

Fig. 7 Schematic diagram of axial attention

4 实验和结果分析

4.1 实验设置

4.1.1 数据集

为了评估本文方法, 我们在两个广泛使用的数据集上进行了实验: TuSimple 数据集和 CULane 数据集。关于两个数据集的详细信息见表 1。

TuSimple 数据集是在高速公路光照条件稳定的情况下采集的数据集, 常用于测试自动驾驶算法。它由 6408 张图像组成, 其中 3626 张标注训练图像和 2782 张测试图像, 分辨率大小为 1280 × 720。CULane 数据集由 133235 张图像组成, 其中包括九种不同的场景, 图像分辨率为 1640 × 590, 场景更加复杂, 检测难度更大。

4.1.2 参数设置

针对不同的数据集, 由于其图像分辨率不同, 所以在对图像进行预处理时参数设置是不同的, 不同数据集的参数设置如表 2 所示。

对于 TuSimple 数据集, 对左前车道和右前车道做行锚分类时将图像中感兴趣区域划分为 56 行, 每一行包含 100 个单元格, 对左侧车道和右侧车道做列锚分类时将图像中感兴趣区域划分为 40 行, 每一行包含 100 个单元格。对于 CULane 数据集, 对左前车道和右前车道做行锚分类时将图像划分为 18 行, 每一行包含 200 个单元格, 对左侧车道和右侧车道做列锚分类时将图像划分为 40 行, 每一行包含 100 个单元格。每一行和每一列的单元格数量即行锚点和列锚点数量。

4.1.3 评估指标

两个数据集的官方评估指标是不同的, 对于 TuSimple 数据集, 评估指标是 Acc、FP 和 FN。Acc 的计算公式为:

$$Acc = \frac{\sum_{clip} C_{clip}}{\sum_{clip} S_{clip}}, \quad (1)$$

其中, C_{clip} 是一帧中正确预测的车道点的数量, 而 S_{clip} 是一帧中真值点的总数。FP 和 FN 分别为假阳性和假阴性, 其计算公式如下:

$$FP = \frac{F_{pred}}{N_{pred}}, \quad (2)$$

$$FN = \frac{M_{pred}}{N_{gt}}, \quad (3)$$

其中, F_{pred} 为错误预测的车道数, N_{pred} 为预测的车道数, M_{pred} 为漏检的真实车道数, N_{gt} 为所有真实车道数。将车道线预测正确率小于 0.8 的分段车道线定义为预测错误车道线。

对于 CULane 数据集, 评估指标是 F1, 我们对 Tusimple 数据集也使用了 F1 评估指标。计算公式如式 (4):

$$F1 = \frac{2 \times P_{precision} \times R_{recall}}{P_{precision} + R_{recall}}, \quad (4)$$

其中, $P_{precision} = \frac{TP}{TP + FP}$, $R_{recall} = \frac{TP}{TP + FN}$ 。

4.1.4 环境与训练参数

算法在 24 GB 显存的 NVIDIA 3090 GPU 上实现。操作系统为 Ubuntu 18.04, 开发平台为 PyCharm2021,

表 1 数据集描述

Table 1 Datasets description

数据集	总数据	训练集	验证集	测试集	分辨率	车道数	环境	场景
TuSimple	6408	3268	358	2782	1280×720	≤5	1	高速公路
CULane	133235	88880	9675	34680	1640×590	≤4	9	城区和高速公路

表 2 不同数据集上的参数设置

Table 2 Hyperparameter settings on different datasets.

数据集	TuSimple	CULane
行数量	56	18
列数量	40	40
每一行单元格数量(行锚点数量)	100	200
每一列单元格数量(列锚点数量)	100	100
使用行锚分类车道线数量	2	2
使用列锚分类车道线数量	2	2

使用 PyTorch 深度学习框架。对于 TuSimple 数据集我们使用基础学习率为 0.05 的 SGD 优化器, 权重衰减率为 0.0001, 迭代次数为 100, 每 20 次迭代后使用验证集测试模型训练的效果并返回指标参数。CULane 数据集使用基础学习率为 0.05 的 SGD 优化器, 权重衰减率为 0.0001, 迭代次数为 50, 每次迭代后使用验证集测试模型训练的效果并返回指标参数。要注意的是 TuSimple 是单 GPU, 而 CULane 是 8GPU, 如果对 CULane 数据集进行单 GPU 训练, 学习率需要调整为原学习率的八分之一。

4.2 消融实验

为了明确列锚和轴注意力机制的作用, 我们做了四组消融实验, 分别是仅使用行锚, 行锚与列锚相结合, 在行锚的基础上添加轴注意力机制, 以及将行锚、列锚与轴注意力机制相结合。使用各自训练好的模型在两个数据集下测试, 得到表 3 中的结果参数。

表 3 消融实验结果
Table 3 Ablation results

行锚	列锚	轴注意力	精度/TuSimple	精度/CULane
√			95.55	64.72
√	√		95.89	71.34
√		√	95.91	65.61
√	√	√	95.92	73.05

从表 3 可以看出, 单一的行锚测试结果在两个数据集上的精度都是最低的, 在 TuSimple 上精度只有 95.55, CULane 数据集上只有 64.72。单独添加列锚后, TuSimple 数据集上精度提高了 0.34, CULane 数

据集上精度提高了 6.62。单独添加轴注意力之后, TuSimple 数据集上精度提高了 0.36, CULane 数据集上精度提高了 0.89。而既添加列锚又添加轴注意力之后, 两个数据集的精度达到了最优, TuSimple 数据集上精度提高了 0.37, CULane 数据集上精度提高了 8.33。这些结果表明对车道线使用混合锚点进行检测是有效的, 相较于单一的行锚检测, 混合锚点检测可以大幅提高检测性能, 同时添加轴注意力机制后也可以提高模型的检测精度。

4.3 性能评估及与其他方法的比较

在两个数据集上的模型训练好后, 我们进行了测试评估并与其他方法作比较。对于 TuSimple 数据集, 使用了 SCNN^[2]、SAD^[6]、LaneNet^[7]、DALaneNet^[8]、BezierLaneNet^[15]、E2E^[10]、LaneATT^[14]、UFLD^[11]、UFLDv2^[12] 九种方法进行对比。实验比较了 F1、Acc、FP、FN 四项指标, 如表 4 所示。

由于 TuSimple 数据集在光照条件良好条件下采集且交通场景并不复杂, 目前的车道线检测模型在该数据集上都能实现较高精度的检测, 所以各个方法的评估结果相差不大。对比 F1 指标, 整体来说基于检测的方法要优于基于分割的方法, 本文方法达到了 96.64%, 比分割方法中的 SCNN^[2] 和 SAD^[6] 分别提升 0.67% 和 0.72%。对比 Acc 指标, 本文方法达到了 95.92%, 在基于分割的方法中只比 DALaneNet^[8] 高 0.06%, 在检测方法中分别比 LaneATT^[14]、UFLD^[11] 方法要提升 0.35%, 比基于参数曲线的方法 BezierLaneNet^[15] 提升 0.27%。FP 和 FN 是假阳性和假阴性, 指标越低表示效果越好。对比 FP 指标, 本文方法在基于检测方法中最低, 比分割方法中最低

表 4 在 TuSimple 数据集上与其他方法的比较
Table 4 Comparison with other methods on the TuSimple dataset

方法	F1/%	Acc/%	FP/%	FN/%	
基于分割的方法	SCNN ^[2]	95.97	96.53	6.17	1.80
	SAD ^[6]	95.92	96.64	6.02	2.05
	LaneNet ^[7]	N/A	96.40	7.80	2.44
	DALaneNet ^[8]	N/A	95.86	8.20	3.16
基于参数曲线的方法	BezierLaneNet (ResNet34) ^[15]	N/A	95.65	5.10	3.90
	E2E (ResNet34) ^[10]	N/A	96.22	3.21	4.28
	LaneATT (ResNet18) ^[14]	96.71	95.57	3.56	3.01
	UFLD (ResNet34) ^[11]	N/A	95.55	19.35	4.30
	UFLDv2 (ResNet18) ^[12]	96.11	95.92	3.16	4.59
Ours (ResNet34)	96.64	95.92	2.41	4.29	

注: N/A表示相关论文没有提及该内容。

的 SAD^[6] 还要低 3.61%，比基于参数曲线的方法要低 2.69%，在所有参与比较的方法中，本文效果最好。对比 FN 指标，本文方法虽然没有基于分割的方法和基于参数曲线的方法效果好，但是仍然与基于检测的方法相差不大。总体来说，本文方法在该数据集上的各项指标都获得了与最先进的方法相当的性能。

对于 CULane 数据集，使用 SCNN^[2]、SAD^[6]、CLRNet^[13]、BezierLaneNet^[15]、E2E^[10]、LaneATT^[14]、UFLD^[11] 和 UFLDv2^[12] 八种方法进行比较。主要对数据集自定义的 F1 评价指标和 FPS 进行比较，结果如表 5 所示。前九列为各方法在 CULane 数据集的九种不同场景的检测结果，倒数第二列为九种场景的平均值，即 F1 指标，最后一列为 FPS 指标。本文方法在 CULane 数据集上的评估指标 F1 达到了 76.0%，FPS 达到了 171 帧。相比基于分割的方法，本文方法 F1 指标比 SCNN^[2] 方法高出 4.4%，比 SAD^[6] 方法高出 5.2%，同时 FPS 要远高于以上两个分割方法。本文方法与基于参数曲线方法相比，F1 指标高出 2.3%。在基于检测的方法中 F1 指标位居前列，仅比 CLRNet^[13] 方法低 0.9%，比 LaneATT (ResNet34)^[14] 低 0.7%，同时 FPS 指标也可以实现对车道线的实时检测。具体到九种场景，本文方法在九种场景中的检测性能达到了较高的水准，基本上只与 CLRNet (ResNet34)^[13] 和 LaneATT (ResNet34)^[14] 有细微差距。

以上结果证明，本文方法不仅具备基于检测的方法的快速性，还达到了与基于分割的方法相当的精度，在各种条件下都具有一定的鲁棒性。

4.4 典型场景下可视化结果分析

图 8 为本文方法在两个数据集中典型场景下的检测结果。每个数据集都展示为两行，第一行和第二行是 CULane 数据集中的图像，第三行和第四行是 TuSimple 数据集中的图像。

在图 8 中，第一行图像主要展示了在拥堵情况下的车道线检测效果。可以看到拥堵时原图像中车道线存在被车辆遮挡的情况，如图中白色椭圆所示；第二行主要展示了在夜晚情况下的车道线检测效果。图中也存在车道线被车辆遮挡的情况，同时由于夜晚光照条件不好还存在车道线模糊的情况，如图中红色椭圆所示；第三行和第四行主要展示了对弯道附近车道线的检测效果。如图中黄色椭圆框所示，车道线在弯道附近同样存在不清晰和被遮挡的情况。对于这些典型场景下存在的车道线视觉线索不好的情况，通过对其可视化结果分析，可以发现本文方法最终对上述拥堵、夜晚和弯道情况下的车道线都实现了良好的检测，尤其是原图像中椭圆框标记出的车道线部分，证明了本文方法可以良好解决当车辆行驶在车道线不清晰或缺失的道路上时存在的视觉线索缺失问题。

表 5 CULane 测试集 F1 的比较
Table 5 Comparison of F1 on CULane dataset

方法	Normal	Crowd	Dazzle	Shadow	No-line	Arrow	Curve	Cross	Night	Average	FPS
基于分割的方法											
SCNN ^[2]	90.6	69.7	58.5	66.9	43.4	84.1	64.4	1990	66.1	71.6	8
SAD ^[6]	90.1	68.8	60.2	65.9	41.6	84.0	65.7	1998	66.0	70.8	75
基于参数曲线的方法											
BezierLaneNet (ResNet18) ^[15]	90.2	71.6	62.5	70.9	45.3	84.1	59.0	996	68.7	73.7	N/A
基于检测的方法											
E2E (ResNet34) ^[10]	90.4	69.9	61.5	68.1	45.0	83.7	69.8	2077	63.2	71.5	N/A
CLRNet (ResNet34) ^[13]	93.3	78.3	73.7	79.7	53.1	90.3	71.6	1321	75.1	76.9	103
LaneATT (ResNet18) ^[14]	91.1	73.0	65.7	70.9	48.4	85.5	68.4	1170	69.0	75.1	250
LaneATT (ResNet34) ^[14]	92.1	75.0	66.5	78.2	49.4	88.4	67.7	1330	70.7	76.7	171
UFLD (ResNet18) ^[11]	89.3	68.0	62.2	63.0	40.7	83.5	58.2	1743	62.9	69.7	323
UFLD (ResNet34) ^[11]	89.5	68.7	57.2	69.2	41.7	84.7	59.3	2037	65.4	70.9	175
UFLDv2 (ResNet18) ^[12]	92.0	74.0	63.2	72.4	45.0	87.7	69.0	1998	69.8	75.0	330
Ours (ResNet34)	92.6	74.9	65.6	75.5	49.0	88.2	69.8	1864	70.9	76.0	171

注：N/A表示相关论文没有提及该内容。

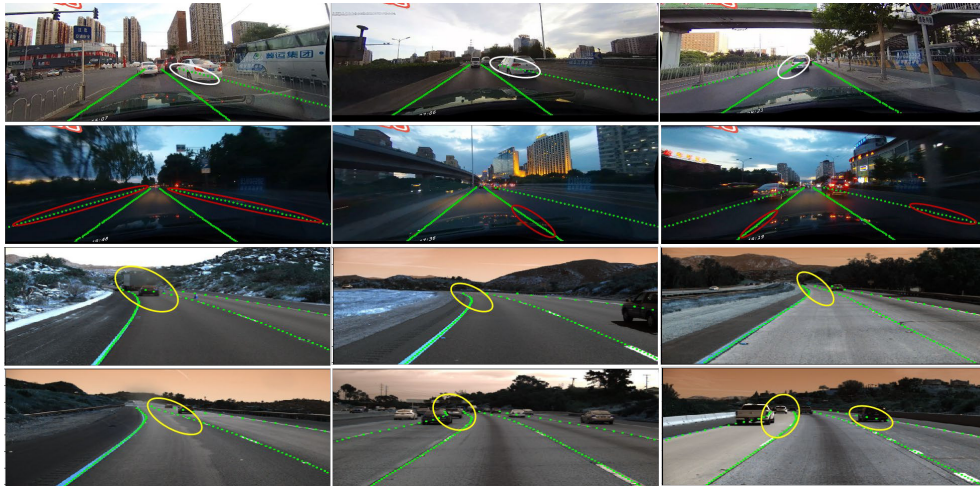


图 8 CULane 和 TuSimple 数据集的可视化
Fig. 8 Visualization of the CULane and TuSimple dataset

5 总结

本文提出了一种轴注意力引导的锚点分类车道线检测方法。首先,在图像上利用行锚和列锚来取代逐像素点的分割方法可以通过更大的感受野获取到全局信息,从而解决视觉线索缺失的问题,同时在模型性能上减少了计算量,提升了检测速度。与基于分割的方法相比,锚点分类通过用一系列(行和列)混合锚点上的稀疏坐标表示车道线会丢失部分信息,而利用轴注意力可以聚焦行向和列向的显著特征,从而改善基于检测方法存在的检测精度较低的问题。本文方法具有较高的精度和较快的速度,在 TuSimple 和 CULane 数据集上实现了良好的效果,同时与其他方法相比,我们的方法具有综合优势。然而,作为一种基于单传感器的检测方法难以在更加复杂的现实场景中(如雨天路面、污染路面)实现高精度检测,后续研究可以通过多传感器融合(如激光雷达融合视觉)、引入车辆运动信息的先验约束等方法来完成更为复杂场景下的车道线检测任务。

参考文献

- [1] The tusimple lane challenge[EB/OL]. (2018-10-20). <https://github.com/TuSimple/tusimple-benchmark/issues/3>.
- [2] Pan X G, Shi J P, Luo P, et al. Spatial as deep: spatial CNN for traffic scene understanding[C]//*Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018: 891.
- [3] Huang Y P, Li Y W, Hu X, et al. Lane detection based on inverse perspective transformation and Kalman filter[J]. *KSII Trans Internet Inf Syst*, 2018, 12(2): 643-661.
- [4] Laskar Z, Kannala J. Context aware query image representation for particular object retrieval[C]//*Proceedings of the 20th Scandinavian Conference on Image Analysis*, 2017: 88-99. https://doi.org/10.1007/978-3-319-59129-2_8.
- [5] Zhao H S, Zhang Y, Liu S, et al. PSANet: point-wise spatial attention network for scene parsing[C]//*Proceedings of the 15th European Conference on Computer Vision*, 2018: 270-286. https://doi.org/10.1007/978-3-030-01240-3_17.
- [6] Hou Y N, Ma Z, Liu C X, et al. Learning lightweight lane detection CNNs by self attention distillation[C]//*2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019: 1013-1021. <https://doi.org/10.1109/ICCV.2019.00110>.
- [7] Neven D, De Brabandere B, Georgoulis S, et al. Towards end-to-end lane detection: an instance segmentation approach[C]//*2018 IEEE Intelligent Vehicles Symposium*, 2018: 286-291. <https://doi.org/10.1109/IVS.2018.8500547>.
- [8] Guo Z Y, Huang Y P, Wei H J, et al. DALaneNet: a dual attention instance segmentation network for real-time lane detection[J]. *IEEE Sensors J*, 2021, 21(19): 21730-21739.
- [9] Zhang C, Huang Y P, Guo Z Y, et al. Real-time lane detection method based on semantic segmentation[J]. *Opto-Electron Eng*, 2022, 49(5): 210378. 张冲, 黄影平, 郭志阳, 等. 基于语义分割的实时车道线检测方法[J]. *光电工程*, 2022, 49(5): 210378.
- [10] Yoo S, Lee H S, Myeong H, et al. End-to-end lane marker detection via row-wise classification[C]//*2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020: 4335-4343.
- [11] Qin Z Q, Wang H Y, Li X. Ultra fast structure-aware deep lane detection[C]//*Proceedings of the 16th European Conference on Computer Vision*, 2020: 276-291. https://doi.org/10.1007/978-3-030-58586-0_17.
- [12] Qin Z Q, Zhang P Y, Li X. Ultra fast deep lane detection with hybrid anchor driven ordinal classification[J]. *IEEE Trans Pattern Anal Mach Intell*, 2022. <https://doi.org/10.1109/TPAMI.2022.3182097>.
- [13] Zheng T, Huang Y F, Liu Y, et al. CLRNet: cross layer refinement network for lane detection[C]//*2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022: 888-897. <https://doi.org/10.1109/CVPR52688.2022.00097>.

- [14] Tabelini L, Berriel R, Paixão T M, et al. Keep your eyes on the lane: real-time attention-guided lane detection[C]. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021: 294–302. <https://doi.org/10.1109/CVPR46437.2021.00036>.
- [15] Feng Z Y, Guo S H, Tan X, et al. Rethinking efficient lane detection via curve modeling[C]//*2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022: 17041–17049. <https://doi.org/10.1109/CVPR52688.2022.01655>.
- [16] Felguera-Martin D, Gonzalez-Partida J T, Almorox-Gonzalez P, et al. Vehicular traffic surveillance and road lane detection using radar interferometry[J]. *IEEE Trans Veh Technol*, 2012, 61(3): 959–970.
- [17] Wang F, Jiang M Q, Qian C, et al. Residual attention network for image classification[C]//*2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 6450–6458. <https://doi.org/10.1109/CVPR.2017.683>.
- [18] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//*2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 7132–7141. <https://doi.org/10.1109/CVPR.2018.00745>.
- [19] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//*Proceedings of the 15th European Conference on Computer Vision*, 2018: 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.
- [20] Fu J, Liu J, Tian H J, et al. Dual attention network for scene segmentation[C]//*2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 3141–3149. <https://doi.org/10.1109/CVPR.2019.00326>.
- [21] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//*2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016: 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
- [22] Wang H Y, Zhu Y K, Green B, et al. Axial-deeplab: stand-alone axial-attention for panoptic segmentation[C]//*Proceedings of the 16th European Conference on Computer Vision*, 2020: 108–126. https://doi.org/10.1007/978-3-030-58548-8_7.
- [23] Ramachandran P, Parmar N, Vaswani A, et al. Stand-alone self-attention in vision models[C]//*Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 2019: 7.

作者简介



罗鑫 (1998-), 男, 硕士研究生, 主要从事计算机视觉的研究。

E-mail: lx15195975156@163.com



【通信作者】黄影平 (1966-), 男, 教授, 主要从事汽车电子、计算机视觉的研究。

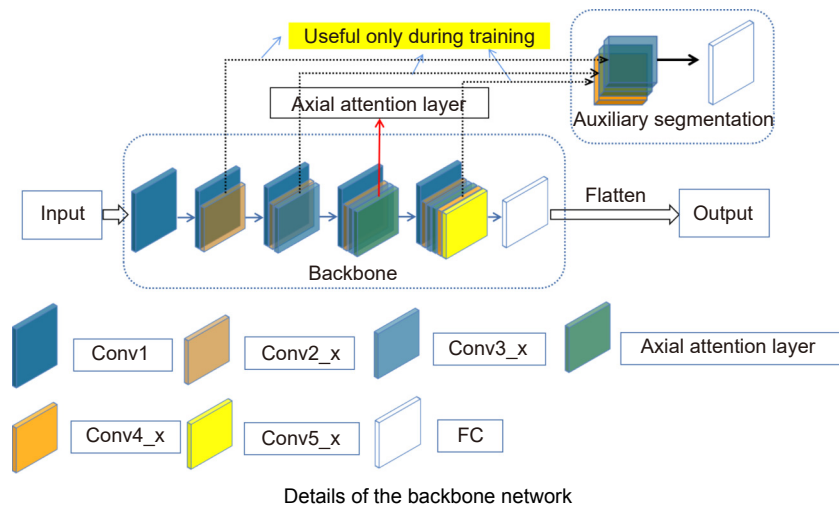
E-mail: huangyingping@usst.edu.cn



扫描二维码, 获取PDF全文

Axial attention-guided anchor classification lane detection

Luo Xin, Huang Yingping*, Liang Zhenming



Overview: Lane detection is an important function of environment perception for autonomous vehicles. Although lane detection algorithms have been studied for a long time, existing algorithms still face many challenges in practical applications, mainly reflected in their unsatisfactory detection results when vehicles travel on roads with unclear or occluded lane lines such as in congestion, at night, or on curves. In recent years, deep learning-based methods have attracted more and more attention in lane detection because of their excellent robustness to image noise. These methods can be roughly divided into three categories: segment-based, detection-based, and parametric curve-based. Segmentation-based methods can achieve high-precision detection by detecting lane features pixel by pixel but have low detection efficiency due to high computational cost and time consumption. Detection-based methods usually convert the lane segments into learnable structural representations such as blocks or points, and then detect these structural features as lane lines. This method has the advantages of high speed and a strong ability to handle straight lanes, but their detection accuracy is obviously inferior to the segmentation-based methods. The performance of parametric curve-based methods lags behind well-designed segmentation-based and detection-based methods because the abstract polynomial coefficients are difficult for computers to learn. Following the framework of detection-based methods, a method that axial attention-guided anchor classification lane detection is proposed. The basic idea is to segment the lane into intermittent point blocks and transform the lane detection problem into the detection of lane anchor points. In the implementation process, replacing the pixel-by-pixel segmentation with a row anchor and column anchor can not only improve the lane detection speed but also improve the problem of missing visual cues of lane lines. In terms of network structure, adding the axial attention mechanism to the feature extraction network can more effectively extract anchor features and filter out redundant features, thereby improving the accuracy problem of detection-based methods. We conducted extensive experiments on two datasets, TuSimple and CULane, and the experimental results show that the proposed method has good robustness under various road conditions, especially in the case of occlusion. Compared with the existing models, it has comprehensive advantages in detection accuracy and speed. However as a detection method reliant on a single sensor, it remains challenging to achieve high-accuracy detection in highly complex real-world scenes, like rainy and polluted roads. Subsequent studies might achieve lane detection in more demanding environments by fusing multiple sensors together, such as laser radar and vision, and by incorporating prior constraints on vehicle motion.

Luo X, Huang Y P, Liang Z M. Axial attention-guided anchor classification lane detection[J]. *Opto-Electron Eng*, 2023, 50(7): 230079; DOI: 10.12086/oe.2023.230079

Foundation item: Project supported by the Natural Science Foundation of Shanghai (20ZR1437900), and National Natural Science Foundation of China (61374197)

School of Opto-Electronic and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

* E-mail: huangyingping@usst.edu.cn