西安交通大学

毕业设计（论文）

题　目　基于快速特征移动和极化注意力机制的车道线检测方法

软件　学院　软件　专业　82　班

学生姓名　张允执

学　　号　2184214444

指导教师　白秀秀

设计所在单位　软件学院

2022　年　6　月

系（ 专业 ） _____
系 (专业) 主任 _____
批 准 日 期 _____

西 安 交 通 大 学

毕业设计(论文)任务书

软件 学院 软件 专业 82 班 学生 张允执

毕业设计(论文)课题 基于神经网络的车道线检测方法研究

毕业设计(论文)工作自 2022 年 2 月 20 日起至 2022 年 6 月 16 日止

毕业设计(论文)进行地点：西安交通大学软件学院西小楼

课题的背景、意义及培养目标

车道线检测是自动驾驶中的一个基本问题，在车辆实时定位、行车路线规划、车道保持辅助以及自适应巡航控制等应用中发挥着至关重要的作用。传统的车道线检测方法通常依靠手工特征提取和需要手动调整滤波算子的算法来拟合线型，工作量大且鲁棒性差。当前基于深度神经网络的车道线检测方法大致可以分为基于分割的方法、基于锚框的方法、行分类检测方法和参数预测方法。基于分割的方法最为常见也取得了令人印象深刻的性能，但检测效率较低。基于锚框的方法从目标检测领域延伸而来，通过改进锚框的形式取得不错的性能，但固定锚框形状导致线型自由度较低。行分类检测方法充分利用车道线形状先验信息，预测车道线在每行的位置，在准确率和效率方面都取得最先进的性能。与上面预测点的方法不同，参数预测方法直接输出由曲线方程表示的参数线，但在精度方面有些欠缺。

设计(论文)的原始数据与资料

1）有关深度神经网络的最新研究参考资料

2）有关道路线检测的研究积累

3）有关道路线的开源代码

4）用于验证算法的道路线检测的数据集

课题的主要任务

1）学习和理解深度卷积神经网络模型的工作原理；

2）学习基于深度神经网络的道路线检测模型；

3）对现有模型的基础上对模型加以改进；

4）采用公开数据集对设计的方法进行验证和测试。

课题的基本要求(工程设计类题应有技术经济分析要求)

通过对道路线检测的研究和学习，了解该问题的研究热点和难点，以及这项任务的发展趋势。为了实现本毕业设计的目标，要求学生充分调查相关资料和文献，掌握该任务的方法和基本思想，并具备实现核心算法的能力。在现有模型的基础上，分析模型特性，实现对模型的改进。

完成任务后提交的书面材料要求(图纸规格、数量，论文字数，外文翻译字数等)

1）图纸规格：参考 2022 届本科毕业设计（论文）工作手册

2）论文字数：论文正文字数不少于 15000 字

3）外文翻译字数：不少于 3000 字

4）其他材料：基于深度神经网络的道路线检测代码

主要参考文献

[1] Qin, Zequn, Huanyu Wang and Xi Li. “Ultra Fast Structure-aware Deep Lane Detection.” ECCV (2020).

[2] Liu, Lizhe, Xiaohao Chen, Siyu Zhu and Ping Tan. “CondLaneNet: a Top-to-down Lane Detection Framework Based on Conditional Convolution.” ICCV (2021).

[3] Tabelini, Lucas, Rodrigo Berriel, Thiago Meireles Paixão, Claudine Santos Badue, Alberto Ferreira de Souza and Thiago Oliveira-Santos. “Keep your Eyes on the Lane: Real-time Attention-guided Lane Detection.” CVPR (2021).

[4] Su, Jinming, Chao Chen, Ke Zhang, Jun Luo, Xiaoming Wei and Xiaolin Wei. “Structure Guided Lane Detection.” IJCAI (2021).

指导教师

接受设计(论文)任务日期

（注：由指导教师填写）　　学生签名：

摘 要

车道线检测是无人驾驶的一个子任务。车道线检测算法采用计算机视觉中的卷积神经网络模型。车道线检测的输入是汽车前置摄像头送入的交通情况的照片，在复杂的路况下，输出照片上对应的车道线掩码。激光雷达和其他高精度传感器成本高昂是自动驾驶难以商业化的一大原因，而使用视觉模型辅助自动驾驶系统的决策则廉价得多。不仅如此，负责车道线检测的视觉模型可以和传感器相辅相成，互相校正车辆行驶过程中的道路信息，保证了自动驾驶系统的安全性和可靠性。

本文所用的 baseline 是 RESA，对车道线的掩码进行端到端的像素级预测。针对原 baseline 像素级预测计算开销大实时性很差的问题，本文基于它提出的新的模型能在降低一半计算量开销的情况下保持原有的精度。 此外，即使是三四线城市，交通状况会非常复杂，其中，行人和车辆遮挡车道线的问题首当其冲，为神经网络推理车道线的位置带来了困难。针对遮挡的问题，原 baseline 提出了新颖的信息传递的模块 RESA。此模块对遮挡场景切实有效，但有堆叠参数之嫌。本文提出的模型在其他模块的加持下，使用 baseline 一半的参数量就达到了原模型相同的精度。最后，车道线的结构是细长的，空间跨度大。以 ResNet 为代表的卷积网络是小卷积核堆叠的模型，感受野十分有限，难以建模空间上的长距离依赖关系。而进行全局信息建模的 Transformer 又无法实时处理汽车前置摄像头送入的高分辨率图片。针对卷积难以建模长距离依赖的问题，本文将注意力机制嵌入 ResNet 中，让 ResNet 拥有全局感受野，建模车道线中空间信息的长距离依赖关系。

本文提出了快速特征移动聚集模块 Fast-FSA(Fast Feature Shift Aggregator)在计算量和参数量降低一半的情况下保持了 Baseline 的精度，甚至是它对抗遮挡的效果。本文对 ResNet 进行的改进，赋予 ResNet 全局感受野能再让模型在精度上提高 2%。

关 键 词：车道线检测；注意力机制；卷积神经网络；

ABSTRACT

Lane detection is a sub-task of auto-driving. Lane detection adopts convolutional neural network in computer vision. Lane detection, which accepts the input of images sent by the camera in front of the car, should output the corresponding lane masks of the input even in complex traffic condition. However, the exorbitant price of the laser radar prevents the auto-driving from commercialization. With the helper of the lane detection, which alleviates the exorbitant price, can also complement the sensor to correct the road information in the process of vehicle driving, ensuring the safety and reliability of the auto-driving.

The baseline used in this thesis is RESA, which conducts end-to-end pixel-wise prediction for the lane masks. To reduce the high computational overhead and speed up the poor real-time performance of the baseline, a new model proposed in this paper can reduce the computational overhead by half with no accuracy falling. In addition, even in the small cities, the traffic is still very complicated so that the problem of pedestrians and vehicles blocking the lane line makes a big trouble for the network doing inference. To alleviate the severe occlusion, the baseline proposes a module called RESA , which has a novel way of the aggregation of spatial information in forward pass. RESA is effective in occlusion scenes, but it has the problem of stacking parameters with huge redundancy. The model proposed in this thesis reduces unnecessary redundant parameters and increases the diversity of modules to achieve the same F1 score as the original model. Finally, the structure of the lane is slender and has a large space span. The convolutional network, such as ResNet, is a model of small kernels stacked, whose receptive field is very limited, has difficulty in modeling long-range dependencies of the spatial information. Transformer, which conducts global information modeling, cannot process high-resolution images sent by the front-facing camera in real time. To solve the problem of modeling long-range dependencies in CNN, my thesis embedded the attention mechanism into ResNet, enables a global receptive field for ResNet and models the long-range dependencies of spatial information of the lanes.

Fast-FSA(Fast Feature Shift Aggregator), a new module proposed in my thesis, compare to the baseline, reduces the FOLPs by half with no F1 scores falling. Then we give ResNet the global receptive field to improve the F1 score by another 2%.

KEY WORDS: Lane detection; CNN; attention mechanism;

目　录

1 绪论 1
1.1 研究背景与意义 1
1.2 国内外研究现状 2
1.3 本文主要研究内容 2
1.4 本文组织结构 3
2 相关工作 4
2.1 车道线检测的常用方法 4
2.2 多尺度特征融合 4
2.2.1 多尺度特征融合的相关研究 4
2.2.1 多尺度特征融合在车道线任务中的应用 6
2.3 注意力机制 6
2.3.1 注意力机制的相关研究 6
2.3.2 注意力机制在车道线检测中的应用 7
2.4 本章小结 7
3 基于特征循环移动的抗遮挡车道线检测方法 8
3.1 问题分析 8
3.1.1 RESA 的不足 8
3.1.2 YOLOF 的不足 8
3.2 RESA 9
3.2.1 Baseline 网络介绍 9
3.2.2 RESA 的前向传播 10
3.3 基于空洞编码的车道线检测方法 YOLOF 11
3.4 基于快速特征移动的车道线检测方法：Fast-FSA 11
3.4.1 Fast-FSA 的前向传播 12
3.5 损失函数 13
3.6 实验与结果分析 14
3.7 本章小结 14
4 基于极化注意力机制的车道线检测方法 15
4.1 问题分析 15
4.2 极化注意力机制 PSA 的前向传播 15
4.3 坐标注意力机制 CA 的前向传播 17
4.4 实验与结果分析 18
4.4.1 CA 和 Fast-FSA 的对比实验 18

4.4.2 实验数据汇总 19
4.4.3 实验细节 21
4.4.4 定性实验 22
4.5 本章小节 22
5 结论与展望 23
5.1 结论 23
5.2 展望 23
参考文献 24

1 绪论

1.1 研究背景与意义

汽车产业是我国的支柱产业之一，但在发动机和相关配件方面缺乏国际竞争力。中国要想在汽车制造方面赶超欧美，就要在新能源汽车领域弯道超车。新能源汽车是国家重点扶持的科技项目。在我国政府多年的补贴下，相关车企的发展蒸蒸日上。但我国的汽车制造，不能停留在硬件层面，软件方面，也要有所创新。如果能普及自动驾驶，大量的司机会成为工作条件更加舒适的安全员，相关的算法工程师、硬件工程师、车辆工程师等高端就业的供给能显著提高。自动驾驶的目的之一就是避免司机的疲劳或一时的疏忽酿成的悲剧，其二是运用在复杂的物流场景中，节约人力成本。本文研究的车道线检测，让自动驾驶的车辆保持在车道线以内，辅助自动驾驶系统做出更加安全智能的决策。

车道线检测系自动驾驶的一个子任务，输入是汽车摄像头输入的高分辨率图片，然后经过计算机视觉的相关算法，将道路线进行标注，输出每一条车道线在输入图像上对应的掩码。最终目的是辅助自动驾驶系统的其他传感器做出决策。此外，为了保证乘客的生命安全，车道线检测必须能应对各种复杂的路况，比如雨雪，强光，黑夜，遮挡等情况，并兼顾实时性。

经济上促进了软硬件和汽车相关产业的发展，归属于国家智能制造的一部分，对带动国家上下游产业，提供新的就业，供给更多的高端岗位有重要意义。

本文的目的是基于视觉算法，改进 baseline[1]模型，提出新的基于分割的车道线检测模型。

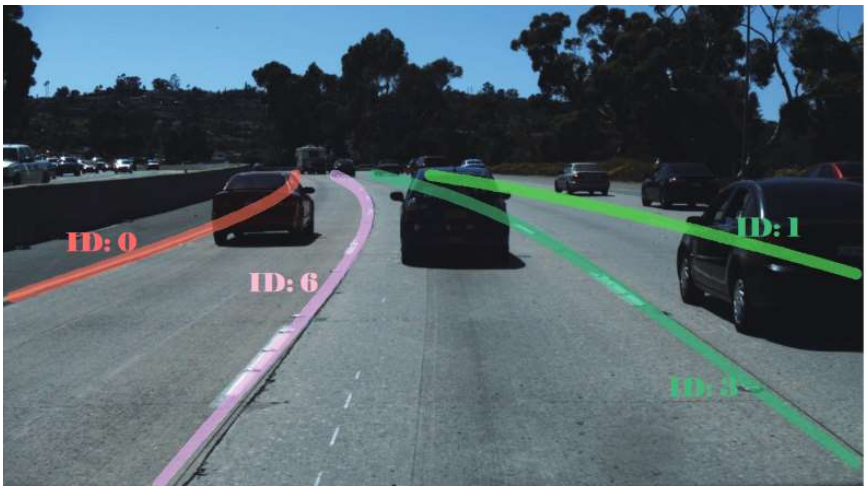

图 1-1 车道线检测效果示意

1.2 国内外研究现状

车道线检测方法分几类：基于图像分割和基于 GAN 的检测方法，对车道线掩码进行像素级预测，精度上有很大的提升潜力，但计算量偏高。基于锚的检测方法，进行车道线实例级别的检测，能兼顾实时性和检测精度，但自由度不高。基于关键点的检测是近两年兴起的，泛化性、实时性都很好，但预处理的过程非常复杂。另外还有基于行分类的检测方法，实时性可以很高，通常用于检测直线。

2020 年左右，以 ResNet18[2]或 ResNet34 为 backbone 的轻量模型，F1 值能达到 75，如果增大模型的参数，采用 ResNet50 乃至 ResNet122 系列，可以将 F1 值扩展到 77，其中 LaneATT[4]采用线锚的方式，兼顾了精度和实时性。到了 2021 年，基于条件卷积的 CondLane[5]把 F1 值提升到了 79，基于关键点检测的 FOLOLane[6]则首次进行了跨数据集的研究，展示了惊人的泛化性。2022，在 FOLOLane 的基础上，更快的 GANet[7]的性能超过 CondLane，其轻量模型的 fps 也能达到 100 以上，与此同时，CLRNet[8]则成为新的 SOTA，首次将 F1 值扩展到了 80。三年以来，精度的增长率大约为 1%，仅净增加 3%。不过，F1 值首次突破了 80 的心理大关，大大鼓舞了相关从业人员。

提升模型在所有场景下的泛化性终究是一条内卷的道路，所以有人另辟蹊径，专注于提升模型的实时性，其中 SwiftLane[9]达到了 400+fps，而很多要兼顾精度的模型却只有可怜的 30fps，SwiftLane 在检测直线的方面，性能优于此前实时性 300fps 的 UFast[10]模型，这样的模型计算资源消耗小，速度快，适合部署在笔直的高速路上。

也有人做车道线的场景特化，通过 GAN 或其他图像处理算法生成特定场景的训练集，让模型在黑夜和恶劣自然条件下保持高精度，这都是算法落地的硬性需求。CondLane 则是针对车道线交织在一起难以分离的复杂的拓扑结构提出了特化算法。FOLOLane 则首次提出跨数据集的研究，以拓展模型的泛化性。

1.3 本文主要研究内容

本文对基于分割的车道线检测模型的改进含 3 个方面。其一，基线模型 RESA[1]对抗遮挡的模块存在参数暴力堆叠的现象，本文提出的模型旨在保留原模型对抗遮挡的信息传递过程的同时避免冗余低效的计算，以达到减少计算资源的消耗并提高实时性的目的。其二，ResNet 卷积核过小，有效感受野非常小，难以建模空间信息的长距离依赖关系，本文改进了 ResNet 模型，给予其全局感受野以建模长距离依赖。其三，FPN[11]网络复杂的分支结构难以训练，影响实时性，本文提出，只用 backbone 产出的最后一级输出，加上 FPN 的分治策略就能在保

持实时性的同时达到不错的效果。实验结果表明，本文做的工作是切实有效的。

1.4 本文组织结构

第一章 绪论部分简要地补充了摘要提及的研究内容和经济影响。描述了相关研究成果的发展历史，指出了基线模型的一些痛点和改进方向。

第二章 相关工作指出车道线检测方法的多样性后着重介绍了 SOTA 模型和高精度模型都在用的多尺度特征融合，提到了一些注意力机制对本文的启发。

第三章 基于特征循环移动的抗遮挡车道线检测方法详细地介绍了 baseline 模型 RESA，和模块 YOLOF[12]，基于这两个模块才能提出本文的新模块 Fast-FSA, 此模块能快速检测遮挡场景。

第四章 基于极化注意力机制的车道线检测提出要对 ResNet 难有全局信息交互的短板做出补偿，进而嵌入注意力机制 CA[32]和 PSA[31]来改进 ResNet。此外，所有的实验数据会在这一章进行一个汇总，方便读者观察实验结果。

第五章 结论与展望重申了本文模型的优势，指出了一些比较宽泛的模型优化方向。

2 相关工作

2.1 车道线检测的常用方法

车道线具有鲜明的几何结构，这使得研究车道线的方法变得十分灵活。主要有以下几种研究方法：基于图像分割、基于关键点聚类、基于目标检测、基于曲线参数方程、基于 GAN 等。

基于分割的网络如 SCNN[22]在当年也堪称高精度模型，使用了特征图的移动来对抗遮挡，缺点是像素级的密集预测计算量偏高。RESA 则吸收了 SCNN 对抗遮挡问题的信息传递方式，减少了 SCNN 的计算量，在 2020 年也达到了不错的效果。LaneAF[14]把亲和力场的概念迁移了过来，提出了鲁棒的分割模型。

基于关键点聚类的 FOLOLane 提出了自底向上的架构恢复车道线，但车道线恢复是迭代式的，效率较低。GANet 在 FOLOLane 的基础上，提出预测关键点离起始点的偏移量能通过索引的方式高效聚类车道线，达到了更强的效果。

基于目标检测的 Line-CNN[23]首先提出了不同于 bounding box 的线锚概念，让研究者能把成熟的目标相关技术迁移到车道线检测中。LaneATT 再此基础上通过车道线实例级别的注意力机制，达到了不错的精度和实时性。

基于曲线参数方程的模型并不多，不过预测的参数量小，实时性高。基于 Transformer[3]的 LSTR[24]也能在 TuSimple[25]这样的简单数据集上达到不错的效果。

基于 GAN 的 EL-GAN[27]认为，基于 GAN 生成的车道线更加接近真实的形状，更能保留车道线的整体结构。CycleGAN[26]则增强了模型夜间的表现。

2.2 多尺度特征融合

2.2.1 多尺度特征融合的相关研究

FPN：识别尺度差异大的物体是视觉任务的挑战之一。车道线有长有短，具有一定的尺度差异。FPN 就是为了解决这个问题而存在。SOTA 模型 CLRNet 和先进的模型 GANet 证明，FPN 对目标检测算法的提升具有显著性。FPN 这些年不断迭代进化具有了如图所示的非常多的复杂结构。FPN 的缺点也很明显，它空间开销大，复杂的分支结构更是影响推理速度，而且难以训练。在追求实时性的车道线检测任务中应该根据实际情况对其结构有所取舍。

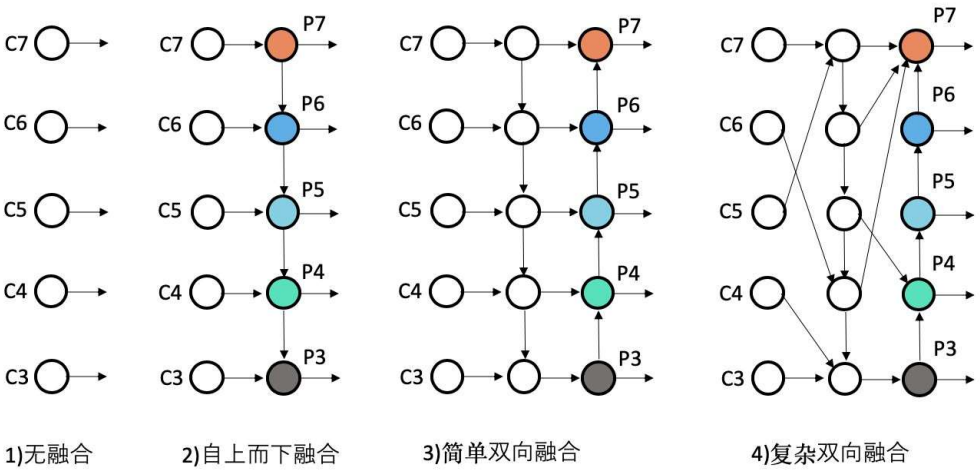

图 2-1 FPN 的常见结构

YOLOF：2021 年的这篇 CVPR 做了大量实验表明，FPN 网络中，分治策略对网络优化问题的影响显著大于特征融合，而特征融合给网络带来的复杂分支结构正好会降低网络的实时性。基于以上思想，作者试图保留分治策略，去除 FPN 的网络复杂的分支结构，在优先保证实时性的前提下保持检测的精度。本文会基于 YOLOF 对 FPN 的改进思想改善 RESA 的性能。

DLA-34：DLA-34[13]是 SOTA 模型 CLRNet 和模型 LaneAF 使用的 backbone，该模型在一定程度上可以被定义为树状的 FPN 网络，在同等参数量的前提下，DLA 系列比 ResNet 系列具有更强的拟合能力。传统的卷积网络只是往更深更广的方向生长，这样生长的网络本质是堆叠参数，没有改变网络层与层之间，模块与模块之间的信息传递方式。U-Net[15]中经典的 skip connection 虽然缓解了上采样过程中信息丢失的问题，但这样的连接方式让解码器使用的信息还是太浅层了，也就是说，一些信息的编码级别不够高，信息难以跨层利用。DLA 系列针对以上问题提出更有利于信息跨层融合的树状网络。本文的研究也会对网络的 backbone 进行改进。

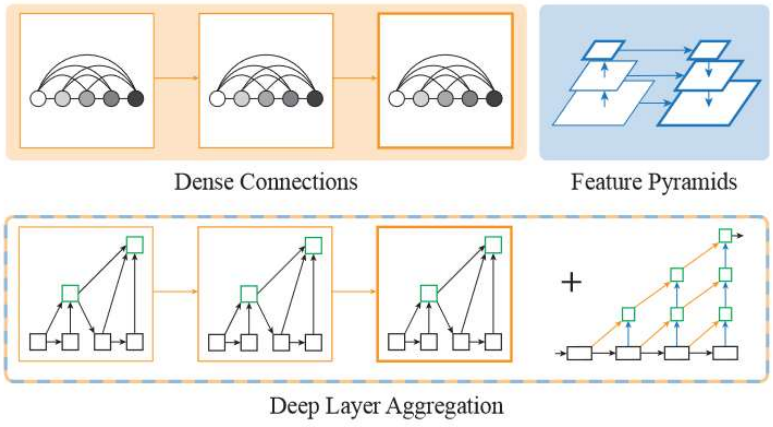

图 2-2 DLA 介绍

2.2.1 多尺度特征融合在车道线任务中的应用

CLRNet（SOTA F1:80.47）：车道线具有的简单的结构，但原始数据中属于低层次特征，车辆行人天气都会对车道线特征提取造成困难。CNN 的特点是将低层次的空间信息逐步转化为高层次的语义信息并储存在通道维度中。无遮挡的车道线的几何结构相对固定，用传统算法求图像梯度或多或少都能捕捉到车道线的特征，而遮挡位置却非常随机，且常见，这就需要 CNN 从遮挡中推理出空间信息，这是一个相对困难的问题。当然，CNN 网络越是深层，在没有模型衰退的情况下，推理出的未遮挡车道线形状信息就越丰富，故可以从深层向浅层进行一个特征融合。这样自顶向下的结构，就能让 CNN 修复出被遮挡的车道线。该方法达到了 SOTA，是正确应用 FPN 的典范。然而 FPN 网络难以收敛，因此本文在应对遮挡问题时，会采用其他方式。

GANet（F1:79.6）：GANet 仅次于 SOTA，直接实现了 FPN 结构，保留了 8 倍、16 倍、32 倍的下采样结果。下采样倍率大了，就和有误差，GANet 在 L1 损失函数中，补偿了这个误差。也就是说，在多尺度融合时，下采样倍率高的张量应该加上某种补偿机制以应对空间信息的丢失。这种思想是值得借鉴的。

CurveLane[16]（F1:74.8）：CurveLane 是早年专门针对弯曲车道线提出的模型。CurveLane 认为，只要关键点选取的足够多，车道线就能像素描画圆一样用关键点切出一条曲线。该模型预测三个尺寸的关键点图，并把这些点进行融合，该过程使用 NMS 降噪，当不同尺寸的关键点图在预测的过程中产生冲突时，优先考虑大尺寸的预测图。这是因为作者相信浅层信息保留了更多空间信息，越高的下采样倍率信息丢失越严重。

UFAST（300fps）：UFAST 是专门检测笔直车道线的模型，速度显著高于其他模型，此模型的优势不在于精度。FPN 如果能训练出来，精度当然会提高，但它的分支结构会摧毁网络的实时性。UFAST 就只在训练时用 FPN，做测试时，变成单分支结构。也就是说，FPN 网络会变成训练 UFAST 模型的监督者去做图像分割，辅助行分类。这也是对 FPN 实时性的改善。

2.3 注意力机制

2.3.1 注意力机制的相关研究

SENet[17]：上文提到过，卷积神经网络缺少全局信息的提取能力，要想获得更大的感受野，必须让网络加深。加深网络的过程中，空间信息会逐步转化为语义信息并储存在通道维度中。然而，网络的通道维度之间缺乏信息的交互，SENet 使用了空间尺度上的全局平均池化代表对应的通道维度信息，再让这些信息经过

全连接层，一种叫 bottleneck 的结构，在通道间建立自注意力关系。

ECANet[18] :SENet 的 bottleneck 结构会将通道信息投影到比通道维度更低的维度。原本只是为了降低计算量，增加网络非线性，但这种结构与相关神经元之间并不能形成一一对应的关系。文章尝试去掉了投影到低维的操作，并将全连接矩阵换成了对角矩阵，发现性能优于 SENet。这篇文章让人思考，提出新的注意力机制模块时，哪些信息应该被保留，哪些信息应该被池化。本文选用的注意力机制也会注意这些问题。

SKNet[19]:SENet 是通道注意力机制，SKNet 是基于卷积核的注意力机制。涉及多尺度问题时，应用于多分支卷积网络中，且和 Inception[28]网络的多尺度选择不同，SKNet 会自己选择用什么尺寸的卷积。SKNet 的实验结果非常有趣，当检测目标的尺寸变大时，大尺寸的卷积核注意力权重会上升。但这一现象只存在于网络的浅层中，网络层数越深，随着空间信息的丢失和感受野的变大，大核卷积和小核卷积的注意力权重将会持平。这个工作为本文在哪里采用大核卷积提供理论依据。

CBAM[20] :光有通道维度的注意力机制是不够的，CBAM 简洁而高效地对空间信息进行了编码，对池化的选取也不限于均值池化。CBAM 使用 7x7 的卷积同时编码均值池化和最大值池化的空间信息。通道维度中也做了类似地处理。本文选用的注意力机制会用不同的方式对空间信息进行编码。

2.3.2 注意力机制在车道线检测中的应用

VIL-100[21]：本文最大的贡献是公开了用于视频检测车道线的数据集。然而 VIL-100 对注意力机制的应用到达了令人发指的程度，不过也少有的将视频分割的方法迁移到了车道线检测任务中。将视频中相邻的帧之间进行了 non-local 运算，当前的帧编码成查询 Q，相邻的帧编码出 K 和 V 进行注意力运算。

LaneATT：LaneATT 使用 ResNet 将每条候选车道线的特征张量池化了出来，然后使用车道线实例级别的注意力机制。

2.4 本章小结

首先介绍了车道线检测的常见研究方法是：基于分割、关键点、目标检测、参数方程、GAN 等方法。之后介绍了多尺度特征融合，这是因为先进的工作都使用了多尺度特征。最后介绍了常见的注意力机制，以及在车道线任务中的应用。

3 基于特征循环移动的抗遮挡车道线检测方法

ResNet 囿于感受野的局限性，不加深网络层数，就难以获得更大的感受野，难以建模车道线这样的长距离依赖。此外，车道线的遮挡问题十分严重，CNN 从特征图恢复车道线时，被遮挡的部分需要根据相邻的未遮挡的车道线推理出来。本文的方法 Fast-FSA(Fast Feature Shift Aggregator)，能聚集特征图相邻切片的信息，这样，被遮挡部分影响的切片，就能根据相邻切片的信息修补出来。相比于 Baseline 所用到的 RESA 方法，本文提出的模型更快速。

3.1 问题分析

3.1.1 RESA 的不足

RESA 有暴力堆叠参数之嫌。当 RESA 每一个方向的迭代次数是 4 的话，那么就需要 16 个卷积核，这些卷积核的大小在实现中被设置成为了 1x9 和 9x1。RESA 的基础模块很像 ResNet 中无正则化的残差模块，先将特征图进行水平或竖直方向移动 1/4 尺度后通过卷积层和 ReLU 函数激活后与原输入求和。卷积的 kernel 的大小 1x9 和 9x1，单调的堆叠了 16 次。

RESA 缺少正则化和残差连接。在 CULane 这个只有万张图片的数据集下，非常容易造成过拟合的现象，如果不在图像预处理或是损失函数的地方下下功夫，会造成模型性能的衰退。

RESA 的计算量过大。其一体现在 neck 部分的 RESA 模块，总共堆叠了 16 个大小为 9 的卷积，其二体现在它的上采样模块，加上了 BUSD 上采样后，只要批量大小设置的稍大，训练过程中会直接让 GPU 开销翻倍。因此，用类似 U-Net 或其他做高精度语义分割模型的结构跑实时性任务是不明智的。车道线检测必须有一定的实时性，单调重复的堆叠参数，精度终究会达到饱和，速度则一直下降，造成了非常严重的资源浪费。本文则通过增加模块的多样性解决这个问题。

3.1.2 YOLOF 的不足

YOLOF 的感受野偏小且难以处理遮挡问题。YOLOF 所用到的目标检测数据集中，囿于专业不一样，没有为遮挡问题特别考虑过应对方案，后续实验表明 YOLOF 处理遮挡的时候并不如 RESA。此外，YOLOF 所采用的一些卷积核的感受野还是太小了，毕竟它不是做图像分割的，无需为了恢复空间维度的信息而减

小下采样倍率，因此 YOLOF 的下采样倍率比 RESA 高很多，对感受野大小的需求没分割那么强，原论文的中心思想仅仅是拿它解决多尺度问题，在具体的车道线任务中，将 YOLOF“本地化”才能让这个模型更上一层楼。

这两个模块都有缺点，RESA 堆叠网络层数让计算量陡增，YOLOF 感受野不够大，处理遮挡问题时没有较好的应对方案。因此，本文提出使用 YOLOF 的结构实现 RESA 的思想就能消除原有孤立的两个模块的缺点。

3.2 RESA

3.2.1 Baseline 网络介绍

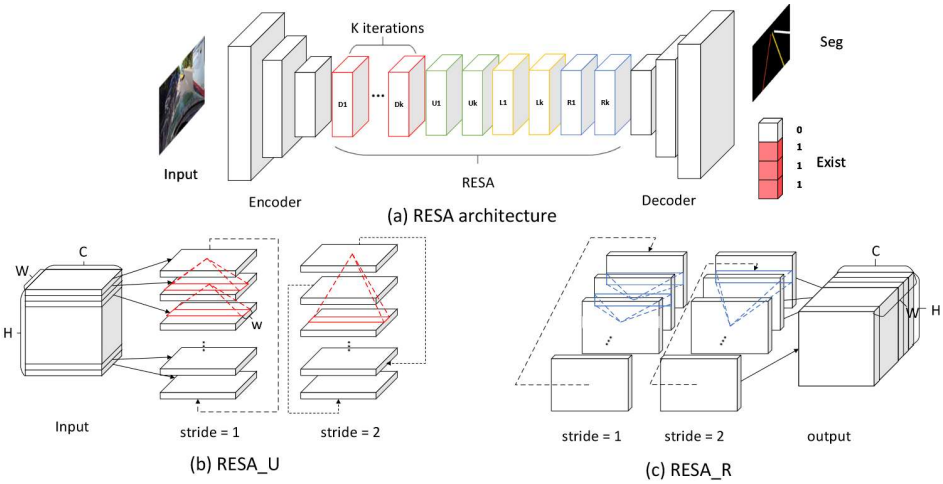

图 3-1 RESA 的网络结构

上图的(a)是 RESA 的整体架构，有三个组成部分，分别是 Encoder，RESA 和 Decoder。(b)则表示在 RESA_U 的模块中，在俯视的 2d 平面上看，特征图信息的传递方向是循环地从下往上传递。(c)则表示 RESA_R 模块中，特征图信息的传递方向是循环地从左向右传递。

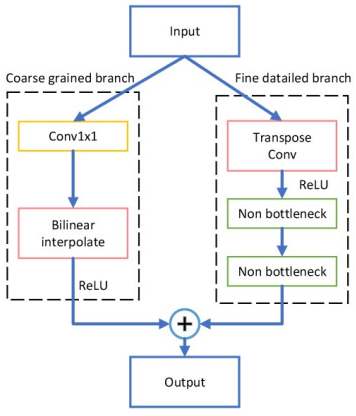

图 3-2 RESA 的网络的 BUSD 上采样模块

Encoer：这是一个视觉模型常用的特征提取器，比如 ResNet。RESA 是基于分割的车道线标注器，在空间尺度上，保有的信息就多一些，下采样倍率仅为 8。图像最初步的特征会被提取送入 RESA。

RESA：全称为循环特征移位的特征聚集器，用来聚集空间尺度的信息来对抗遮挡。在 RESA 的每一次迭代中，都有 4 个可选的信息传递方向，它们分别是从上到下、从下到上、从左往右、从右往左。这四个方向地、或垂直或水平地把特征聚集在一起。每个信息传递的方向会经历 K 次迭代来保证特征图的每一个切片在信息传递的过程中被覆盖，以便网络推理出遮挡的信息。

Decoder：基于图像分割的架构就要有上采样模块输出和特征图尺度相同的掩码。RESA 采用了双侧上采样模块。每一个模块上采样两次，成为两侧。空间尺度相当于恢复了原图的八分之一。每一个上采样模块有两个分支，一个分支大体恢复图像的轮廓，另一个分支恢复它的细节。

Head：在进行上采样之后，RESA 模块输出的特征图会送入全连接层 Head 并以概率分布的形式预测每一条车道线的存在性，对每一条车道线的存在和不存在做二分类预测。模型结合 Head 预测车道线的存在性的信息和 Decoder 输出的掩码信息来预测车道线。这样，车道线的掩码就会逐像素的输出。

3.2.2 RESA 的前向传播

对于给定的张量输入$X \in R^{C\times H\times W}$，其中$C$代表输入的通道数，$H$代表特征图的高度，$W$是特征图的宽度。经历一次 RESA 的迭代后，在空间尺度$R^{H\times W}$上把特征图X的切片顺序进行打乱，指的是在水平和竖直方向进行循环移动。$X^k_{c,i,j}$的上标k代表当前特征图迭代的次数，c,i,j分别表示通道维度的索引，高度维的索引和宽度维的索引。这样，RESA 的前向传播过程可以被描述为：

$$Z^k_{c,i,j} = \sum_{m,n} F_{m,c,n} \cdot X^k_{m,(i+s_k) mod H, j+n-1} \tag{3-2-1}$$

$$Z^k_{c,i,j} = \sum_{m,n} F_{m,c,n} \cdot X^k_{m,i+n-1,(j+s_k) mod W} \tag{3-2-2}$$

$$X^{k'}_{c,i,j} = X^k_{c,i,j} + f\left(Z^k_{c,i,j}\right) \tag{3-2-3}$$

$$s_k = \frac{L}{2^{K-k}}, k = 0,1,\dots,K-1 where K = \lfloor log_2 L \rfloor \tag{3-2-4}$$

上述公式中，f是非线性的激活函数比如 ReLU。公式（3-2-3）中X的单引号上标表示X进行了 inplace 的运算。s_k是偏移量。F是卷积操作，它的输入和输出的通道数都是C，保持不变。

3.3 基于空洞编码的车道线检测方法 YOLOF

YOLOF 是你(You)只(Only)看(Look)一级(One-Level)特征(Feature)的缩写，FPN 网络过于复杂，影响推理速度，模型难以收敛，但又想要多尺度的特征融合，所以就有了空洞编码层。

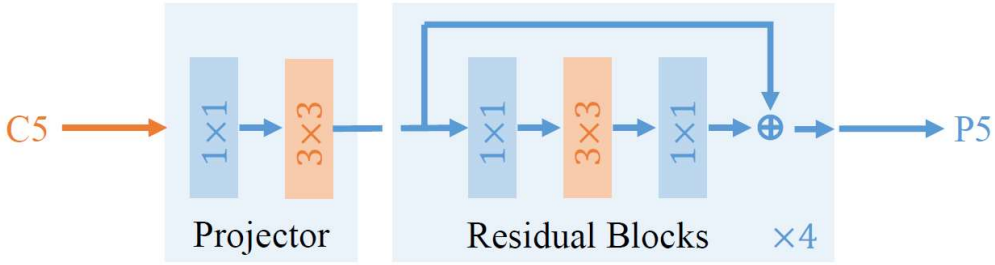

图 3-3 YOLOF 的结构图

上图阐明了 YOLOF 中空洞编码层的架构，图中的 1x1 和 3x3 都是卷积核的大小，x4 则说明有四个后继残差模块。所有的残差模块后面都会跟着一个正则化的层和一个 ReLU 函数激活。在投影层 Projector，则会使用卷积层和批量正则化层。

在原论文实现的过程中，原有的 YOLOF 输入是 backbone 下采样 32 倍的特征图，其通道数为 512。先经过 Projector 进行投影，1x1 的卷积会在通道维度上削减 4 倍以降低计算量，然后再通过 3x3 的卷积。它为了解决目标检测中的多尺度问题，4 个残差模块的 3x3 卷积都采用了扩张率不同的空洞卷积组成不同的感受野。最后，如果有需求，则可以用一个 1x1 的卷积把通道维度恢复成 512。

3.4 基于快速特征移动的车道线检测方法：Fast-FSA

如果把检测网络的结构分为 input、backbone、neck、head 的话，那么本文提出的新模块是在 neck 上进行改进。

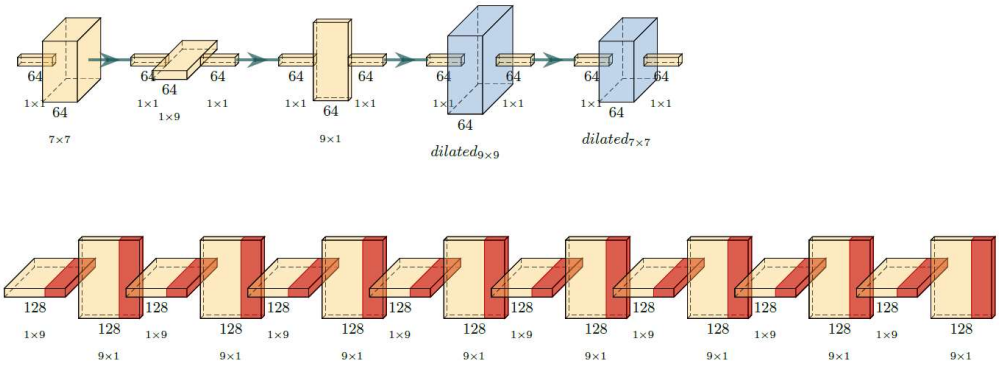

图 3-4 RESA(下)和 Fast-FSA(上)的结构图

如图3-4所示。上半部分是本文提出的Fast-FSA的结构图，下半部分是RESA的结构图。其中的方形的形状能体现卷积核的形状。图中足以见得，RESA的结构是单调的，是模块的重复堆叠。本文采用的结构在卷积核的感受野上具有多样性。蓝色的代表空洞卷积，去除空洞部分后核的有效面积都是9。Fast-FSA模块中，前两个卷积核是投影层Projector，之后，每三个卷积核为一组，隶属于一个残差块，所以有四个残差块。Fast-FSA的残差块与残差块之间，用箭头做了串联。当然，图中只画出了Fast-FSA的卷积层，它的每一个残差块都会紧随一个正则化的层和一个ReLU函数激活。当然，既然Fast-FSA的残差块的名字与ResNet的残差块名字相同，残差的位置也是一样的，会在激活函数之前加上输入的恒等映射。Fast-FSA计算的通道维度是64，RESA计算的通道维度是128。

3.4.1 Fast-FSA的前向传播

对于给定的张量输入$X \in R^{C\times H\times W}$，其中$C$代表输入的通道数，$H$代表特征图的高度，$W$是特征图的宽度：

$$Z = projector(X) \tag{3-4-1}$$

$$identity = Z \tag{3-4-2}$$

$$Z_{c,h,w} = Z_{c,(h+h//4)\ mod\ H,(w+w//4)\ mod\ W} \tag{3-4-3}$$

$$Z = conv_{1\times 1}(Z) \tag{3-4-4}$$

$$Z = conv_{1\times 9}(Z) \tag{3-4-5}$$

$$Z = conv_{1\times 1}(Z) \tag{3-4-6}$$

$$Z = BatchNorm(Z) \tag{3-4-7}$$

$$Z = Z + identity \tag{3-4-8}$$

$$Z = ReLU(Z) \tag{3-4-9}$$

上述公式只展示了从输入到投影层和第一个残差模块的计算过程，第一个残差模块的最终输出在公式(3-4-9)中。“//”代表整数类型的除法。“=”是赋值符号。其中：$h = 0,1,2,\dots,H-1, w = 0,1,2,\dots,W-1, c = 0,1,2,\dots,C-1$，是索引。变量的下表是相关切片的索引运算。Projector模块在图3-4(上)中有显示。

基于YOLOF的架构和RESA的思想，本文提出了新的模块，Fast-FSA。首先在Projector模块，保留在通道维度上做投影的1x1卷积以降低计算量。紧随的3x3的卷积层会被换成感受野更大的7x7的卷积层。理由是基于分割的网络中，其实在空间维度的下采样仅仅为8倍，YOLOF中则能达到16和32倍，所以在几何级数缩小的特征图上3x3的卷积显得没那么小。分割的话为了有一个更好的局部性，让空间维度的信息保留地更加全面，让卷积核更好地聚集信息，为此特意换成了7x7的卷积核。

另外一个改进就是在它的Residual Block上，前向传播过程会和作为baseline

的 RESA 缝合一下，经过卷积核大小为 9 的卷积之前会做循环移动来实现 RESA 的思想。不过，与 RESA 不同的是，RESA 模块堆叠的深度为 16，因此总共会循环移动 16 次，每一次只会往四个方向中的一个方向循环移动，这其实也是没必要的，对机器来说也是单调重复的劳动，有故意堆叠参数之嫌。为了实时性，并对这一套流水线做出简化，我们的模块 Fast-FSA 每次都会同时往两个方向循环移动，每次移动 1/4 个图像尺度。YOLOF 中的每一次迭代会经过一个残差模块，这样循环移动 4 次，图像切片的所有信息都会得到覆盖和传递，从而助力模型推理出遮挡的信息。另外，反向传播的过程中，新的索引和原来的索引对应的梯度也是不变的，能够正常传播。

最后，Fast-FSA 为了更加适应空间尺度上下采样倍率比较小的细长特征图 $X \in R^{128\times36\times100}$，而且注意到 RESA 采用了 1x9 的卷积核恐怕也是考虑到了特征图比较细长的特点，我们采用 1x9 的卷积核则更能在图像循环移动的过程中保证信息的传递，将卷积的局部性发挥得更好。而 YOLOF 中还是存有 3x3 且扩张率为 1 的卷积，它的感受野远远不够大。因此 YOLOF 中不符合特征图尺寸需要的卷积核尺寸会被淘汰。最终，卷积核的尺寸选择为 1x9,9x1,3x3(dilation=4)，3x3(dilation=3)。

3.5 损失函数

Fast-FSA 的损失函数是 Dice loss[33]。

Focal Loss 不稳定，如果不进行平滑处理，在密集像素点的预测计算中容易出现加法溢出和对数溢出，因此，本文的损失函数是 Dice Loss，这个损失函数用于计算两个张量的相似度：

$$s = \frac{2|X \cap Y|}{|X| + |Y|} \tag{3-5-1}$$

其中，$X \cap Y$代表样本X和样本Y的交集，$|X|, |Y|$分别是 X，Y 元素的个数。于是，Dice loss 就可以被描述为：

$$d = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \tag{3-5-2}$$

理解了上述的公式，Dice loss 的实现也要因地制宜。对于网络预测输出$X \in R^{C\times H\times W}$, 和对应的真实掩码$Y \in R^{C\times H\times W}$, 将他们展平后，令：

$$a = X \cdot Y \tag{3-5-3}$$

$$b = X \cdot X + 0.01 \tag{3-5-4}$$

$$c = Y \cdot Y + 0.01 \tag{3-5-5}$$

$$d = 1 - \frac{2 \cdot a}{b + c} \tag{3-5-6}$$

$$loss = \bar{d} \tag{3-5-7}$$

最后的损失函数取了均值，公式(3-5-4)和公式(3-5-5)主要是为了防止分母是0 和一些极端的情况导致数据类型的溢出。毕竟 CULane[22]的有些图片并没有车道线。

3.6 实验与结果分析

表 3-1 本文的模块 Fast-FSA 在 CULane 数据集上的 F1 值。Cross 的指标是 FP。

	RESA-ResNet18[1]	**ResNet18+Fast-FSA**
Normal	0.875	**0.909**
Crowd	0.658	**0.701**
Night	0.637	**0.672**
Noline	0.388	**0.446**
Shadow	0.619	**0.689**
Arrow	0.816	**0.863**
Hlight	0.546	**0.615**
Curve	0.631	**0.667**
Cross(FP)	2141	**1822**
F1	0.683	**0.726**

在模型什么都不加的情况下，F1 值只有 0.683，因此，本文提出的模块，Fast-FSA 是有效的。不仅如此，除了普通场景和曲线场景，本文的模块对其他困难场景的提升效果显著。如果你看不懂什么是 F1 精度，或者不明白该数据集，请参考随后的章节。

3.7 本章小结

如果把目标检测的网络结构分为 input、backbone、neck、head，那么，本文提出的新模块 Fast-FSA 用在网络的 neck 部分，功能是收集整理 backbone 的信息。

Fast-FSA 有 FPN 多尺度分而治之的思想，但为了实时性舍弃了多分支结构。Fast-FSA 用 YOLOF 的结构实现了 RESA 前向传播的思想，以应对遮挡问题，改进了 RESA 计算的流水线，计算量比 RESA 小得多。受第二章注意力机制相关工作的影响，选用了大核卷积代替小核卷积。最后又介绍了损失函数，然后用实验证明了本文提出的模块是有效的。

4 基于极化注意力机制的车道线检测方法

本文将极化(Polarized)注意力(Self-Attention)机制嵌入 ResNet 网络中，为 ResNet 赋予全局感受野，增加 backbone 的信息交互能力，使其建模车道线的长距离依赖。

4.1 问题分析

ResNet 由小核卷积堆叠而成。小核卷积的感受野很小。车道线的结构是细长的，分割网络的下采样倍率较低，保留了不少空间信息。在大量空间信息的保留之下，小核卷积难以像 Transformer 一样对全局信息进行建模。如果不进行全局信息建模，编码的空间信息之间会缺乏信息交互。反映在实际问题中的表现如下：如果原图片存在遮挡信息，经过 ResNet 的编码后，对应的特征图的区域需要从另外的空间区域得到相邻车道线的信息以推理出被遮挡的信息。还有，如果天气状况有变，黑夜和强光，图像的亮度会整体下降或上升，解码出的特征图对应像素值也会随这个趋势变化。照片因为雨雪的原因变得模糊，车道线对应的梯度会被平滑掉，也需要全局信息恢复梯度。ResNet 由于感受野不足，难以对全局信息做出判断。因而需要改进 ResNet 的结构。如图 4-1 所示，本文在 ResNet 中嵌入两种注意力机制并赋予 ResNet 全局感受野。

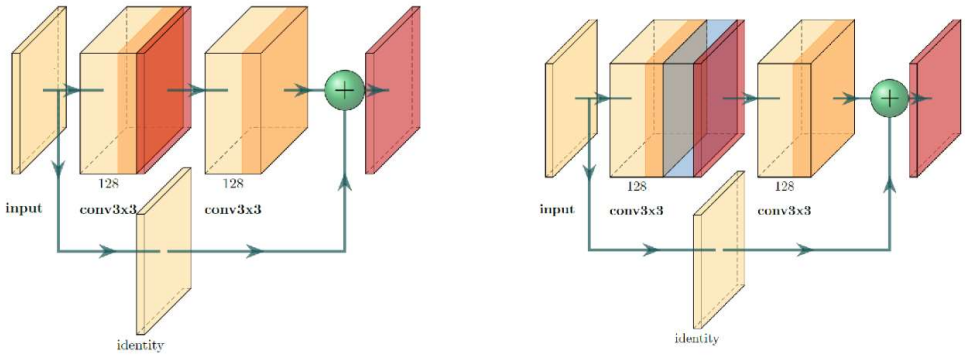

图 4-1 Residual Block(左)和本文的 PSA Block(右)的结构图

4.2 极化注意力机制 PSA 的前向传播

极化注意力机制 PSA 是相关工作中介绍的 CBAM 的一个变体。给定一个输入特征图：$X \in R^{C\times H\times W}$，那么它就有两个尺度信息，第一个尺度信息是$X_{channel} \in R^{C}$，代表通道维度，第二个尺度信息是空间尺度$X_{spatial} \in R^{H\times W}$，这两个尺度是

正交的。那么 PSA 的两个分支做的第一件事就是降维：在输入特征图的一个尺度上保持高分辨率，在与其正交的尺度的所有维度会被展开合并然后用卷积投影到 1，于是，输入特征图就在一个尺度上塌陷了，降低了计算量。在另一个与其正交的尺度上保有了高分辨率。然后过一层 softmax 和正则化并用 sigmoid 函数激活。下列公式展示 PSA 其中一个池化分支的前向传播过程：

$$V = \sigma_v\left(conv_v(X)\right) \tag{4-2-1}$$

$$Q = softmax(\sigma_q\left(conv_q(X)\right)) \tag{4-2-2}$$

$$att = sigmoid(proj(V \times Q^T)) \tag{4-2-3}$$

$$X = att \cdot X \tag{4-2-4}$$

其中以空间尺度的池化为例，$conv_v(\cdot)$代表将输入的张量通道减半，以降低计算量。而σ_v是投影变换，会将空间尺度的维度合并之后展开。而$conv_q$会把通道维度的形状投影到 1，σ_q也是投影变换，和σ_v的作用相同。需要注意的是，×是做矩阵乘法，而·是做内积。$proj(\cdot)$做投影变换并整合信息，里面包含了 LayerNorm 和 ReLU 函数激活，最后会使用卷积将输入投影到与X相同的维度，以便做内积。在通道维度的池化也是相同的前向传播过程，于是输入 PSA 模块的特征图X就会经历两次注意力进行加权。

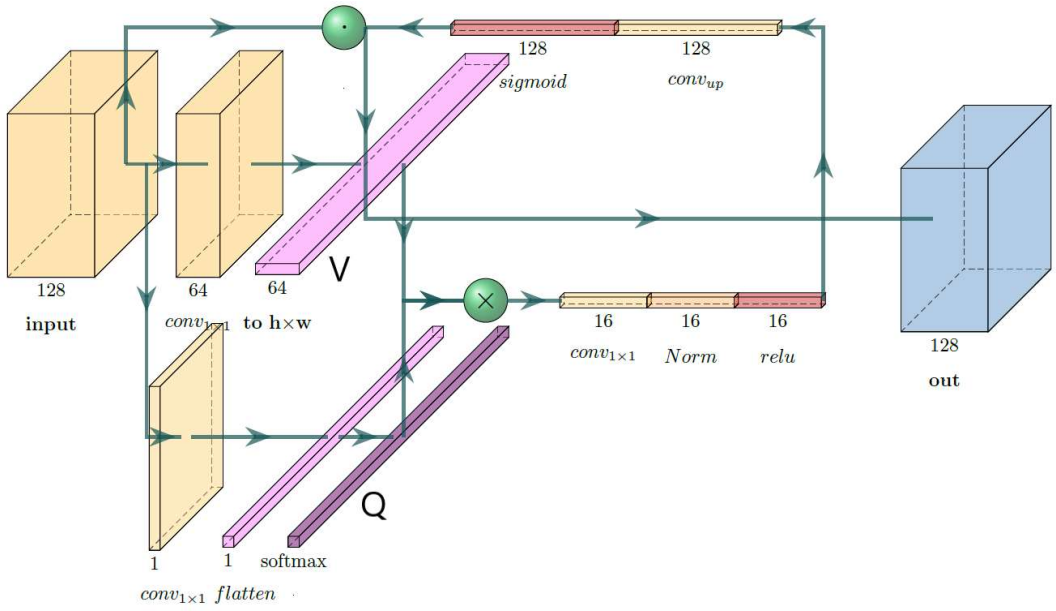

图 4-2 PSA Block 的空间池化过程

如图 4-2，PSA 模块中，这些长方体代表张量，其形状的变化对应模块计算过程中张量形状的变化。空间池化的输入 input 的通道数是 128。到了实现部分，我们把 PSA 模块加到 ResNet 中每一个残差模块里的第一个 3x3 卷积之后。另外，PSA 模块的通道维度池化过程也是类似的，但不同在于，输入张量维度塌陷

过程发生在与其空间池化过程中塌陷的维度正交的维度上，比如Q进行的就不是通道维度的全局平均池化，而是空间维度的全局平均池化。

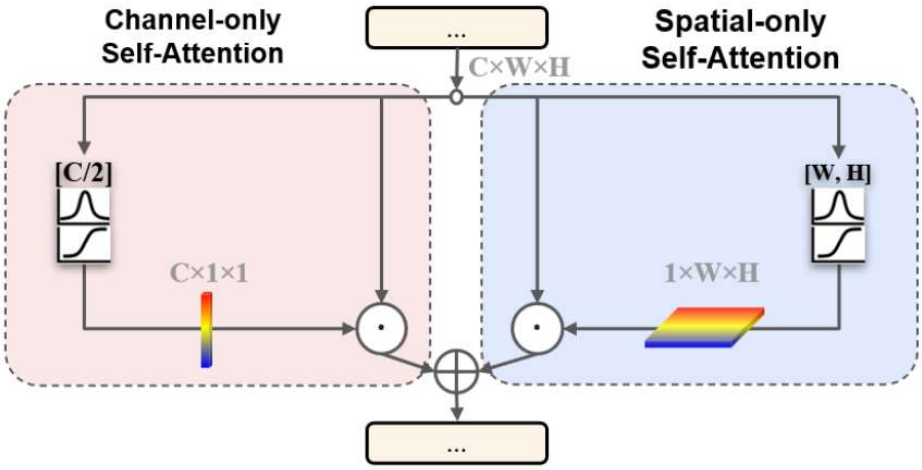

图 4-3 极化注意力机制

4.3 坐标注意力机制 CA 的前向传播

第一步，坐标信息的嵌入：通道注意力使用全局平均池化，缺点是难以保留位置信息，位置信息的保留恰恰对视觉任务非常重要。为了保留精确位置信息，建模空间信息的长距离交互，可以对二维的全局池化在 x 轴和 y 轴上分解成 2 个一维的池化。给定输入$X \in R^{C\times H\times W}$，分别在 H 维度和 W 维度上有两个池化的核$(H,1),(1,W)$，它们分别在水平和垂直的方向上进行编码。所以，在高度 h 上第 c 个通道对应的输出是：

$$z_c^h(h)=\frac{1}{W}\sum_{0\le i<W}x_c(h,i) \tag{4-4-1}$$

类似的，第 c 个通道在宽度 w 上对应的输出是：

$$z_c^w(w)=\frac{1}{H}\sum_{0\le j<H}x_c(j,w) \tag{4-4-2}$$

输入的特征图经过上述的两个变换会在空间尺度的两个不同的方向聚集信息。就能得到含有方向信息的注意力特征图，这个操作保留了更多的空间信息，和 SENet 大不一样。这两个变换也能帮助模型去捕捉空间上的长距离依赖。

第二步，坐标注意力的生成：本小节的公式(4-4-1)和公式(4-4-2)能让编码的时候开启全局感受野，得到精准的位置信息。第二步就是解决如何高效地利用这些信息编码。设计第二步的时候，遵循 3 个准则，第一个准则是变换应该是简单廉价的，没有过度的开销。第二个准则是让模块高效利用位置信息，模型感兴趣

的信号，都会准确地放大。最后，不能仅仅捕捉空间信息而放弃通道信息，通道间也要有自注意力机制进行信息的交互。于是，在得到了公式(4-4-1)和公式(4-4-2)的输出后，会把z^h, z^w拼接在一起，让它们通过一个共享的 1x1 卷积F_1来聚合信息：

$$f = \delta\Big(F_1([z^h, z^w])\Big) \tag{4-4-3}$$

符号[·,·]代表特征图在空间尺度上的拼接，δ是非线性的激活函数，$f \in R^{C/r\times(H+W)}$是中间产出，编码了水平和竖直方向上的空间信，将通道降低 r 倍用于减少计算量。随后，又会沿两个方向分解f，得到$f^h \in R^{C/r\times H}$，$f^w \in R^{C/r\times W}$，另外 2 个 1x1 的卷积操作F_h, F_w会充分利用f^h，f^w的信息去得到与输入X通道数相同的 2 个输出：

$$g^h = \sigma\Big(F_h(f^h)\Big) \tag{4-4-4}$$

$$g^w = \sigma\big(F_w(f^w)\big) \tag{4-4-5}$$

σ是 sigmoid 函数，本小节的公式(4-4-4)和公式(4-4-5)都会被用作注意力的权重，所以，坐标注意力机制模块的输出Y每一个通道 c 和位置 i，j 对应的$y_c(i,j)$是：

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{4-4-6}$$

4.4 实验与结果分析

4.4.1 CA 和 Fast-FSA 的对比实验

表 4−1 Fast-FSA 在 CULane 上和 CA 的对比实验。Cross 采用的指标是 FP。UFAST 和 SCNN 是其他模型。

	RESA-ResNet18[1]	ResNet34-UFAST[10]	SCNN[22]	**ResNet18+CA (ours)**	**ResNet18+Fast-FSA (ours)**
Normal	0.875	0.899	0.906	0.907	**0.909**
Crowd	0.658	0.685	0.697	0.687	**0.701**
Night	0.637	0.646	0.661	0.665	**0.672**
Noline	0.388	0.422	0.434	0.415	**0.446**
Shadow	0.619	**0.677**	**0.669**	**0.712**	0.689
Arrow	0.816	0.838	0.841	0.853	**0.863**
Hlight	0.546	0.595	0.585	0.596	**0.615**
Curve	0.631	**0.695**	0.644	0.670	0.667

Cross(FP)	2141	2037	1990	2070	**1822**
F1	0.683	0.723	0.716	0.715	**0.726**

为了对比本文的模块 Fast-FSA 效果究竟有多强，特别选取了去年微软亚洲研究院出品的注意力机制做对比实验。这两个模块加的位置是相同的，足以见得，本文提出的模块在基于分割的场景下，效果显著超越注意力机制。

4.4.2 实验数据汇总

表 4-2 本文的模型和 RESA 在 CULane 上的 F1 值的相关消融实验。Cross 采用的指标是 FP。

	ResNet34+RESA[1]	+BUSD[1]	RESA-ResNet18[1]	**+CA(ours)**	**+Fast-FSA (ours)**	**+ Fast-FSA +PSA (ours)**
Normal	0.908	0.919	0.875	0.907	0.909	**0.923**
Crowd	0.703	**0.724**	0.658	0.687	0.701	0.719
Night	0.678	**0.698**	0.637	0.665	0.672	0.688
Noline	0.437	**0.463**	0.388	0.415	0.446	0.453
Shadow	0.620	**0.720**	0.619	0.712	0.689	0.671
Arrow	0.862	0.881	0.816	0.853	0.863	**0.883**
Hlight	0.579	**0.665**	0.546	0.596	0.615	0.623
Curve	0.641	0.686	0.631	0.670	0.667	**0.705**
Cross(FP)	**884**	1896	2141	2070	1822	1495
F1	0.733	**0.745**	0.683	0.715	0.726	0.742

为了验证 Fast-FSA 的有效性，与在模型中位置相同的注意力机制 CA 做了对比实验。为了验证 PSA 的有效性，在 Fast-FSA 模块基础上，使用 PSA 改进了 backbone。实验结果表明，backbone 感受野的增加，能全面提升模型的表现。

+BUSD 全称是 ResNet34+BUSD，同理，+CA，+Fast-FSA，+Fast-FSA+PSA 是 ResNet18+CA，ResNet18+Fast-FSA，ResNet18+Fast-FSA+PSA 的缩写。本文的模型会进行字体加粗。Backbone 的参数之所以选的不一样，是为了证明基线 RESA 在计算上存在冗余、暴力堆叠参数的现象。

表 4-3 模型的实时性

模型	FLOPs/G	参数量/M	F1
ResNet34+RESA	88.271	24.287	0.733
ResNet34+RESA+BUSD（baseline）	95.934	24.526	**0.745**
ResNet18+Fast-FSA	**43.303**	**12.018**	0.726

ResNet18+Fast-FSA +PSA**(ours)**	47.221	13.198	0.742

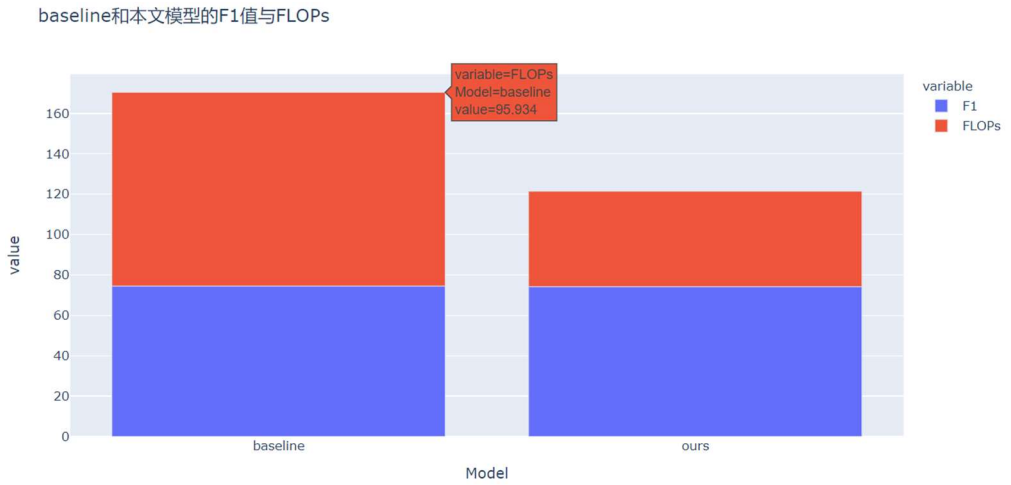

图 4-4baseline 和本文模型的 FLOPs

如图 4-1 所示，红色代表 FLOPs，蓝色是 F1，baseline 的红色部分比本文的模型高很多，而精度几乎相同。

实时性的表格进一步验证了基线 RESA 存在冗余计算的问题。本文的模型相对于 baseline 参数量和计算量是腰斩的。但 F1 能达到相同的水平。在让参数量减半的情况下，保持模型的精度，并不是神乎其技，仅仅是注意到了基线模型存在参数堆叠的问题，然后通过减少参数堆叠，增加模型的多样性以达到相同的性能。

表 4-4 LaneATT 使用本文的模块在 CULane 的 F1 值与其他模型的对比。Cross 采用的指标是 FP。

	LaneATT[4]	SpinNet[30]	R-18-E2E[29]	CycleGAN[26]	**+CA(ours)**	**+Fast-FSA (ours)**
Normal	0.905	0.905	0.900	**0.918**	0.912	0.912
Crowd	0.721	0.717	0.697	0.718	**0.736**	0.726
Night	0.682	0.680	0.633	0.694	**0.704**	0.690
Noline	0.474	0.432	0.432	0.461	0.481	**0.490**
Shadow	0.680	0.729	0.625	**0.762**	0.731	0.748
Arrow	0.854	0.850	0.832	**0.878**	0.869	0.868
Hlight	0.641	0.620	0.602	0.664	**0.689**	0.667
Curve	0.632	0.507	**0.703**	0.671	0.637	0.636
Cross(FP)	1170	NULL	1822	2346	1080	**1070**
F1	0.745	0.742	0.708	0.739	**0.757**	0.754

为了验证本文模块的适用性，进行了跨 baseline 的验证。新的 baseline 选用

了基于锚的模型 LaneATT。+CA，+Fast-FSA 的全称是：LaneATT+CA 和 LaneATT+Fast-FSA。SpinNet、R-18-E2E、CycleGAN 是 LaneATT 发布当年用于对比的其他车道线检测模型。可以观察到，无论是注意力机制 CA 还是本文提出的 Fast-FSA 模块，加在新的 baseline 上效果都是立竿见影的，体现了本文提出模块即插即用的特性。

所有的实验都是在数据集 CULane 上进行的。在 CULane 上，判断模型好坏的指标是 F1 值和FP：

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4-5-1}$$

$$Precision = \frac{TP}{TP + FP} \tag{4-5-2}$$

$$Recall = \frac{TP}{TP + FN} \tag{4-5-3}$$

公式(4-5-1)到(4-5-3)中，真的正样本TP，假的负样本就是FN，假的正样本是FP。值得注意的是，场景 Cross 没有车道线，因此没有用 F1 值，该场景下的FP越小越好。SpinNet 没有给出FP，记为 NULL。

4.4.3 实验细节

本文的代码环境是 linux 系统，显卡是 RTX2060ti，深度学习框架选用 pytorch。RESA 模型和本文的模型选取训练 18 个 epoch 的结果。LaneATT 上进行的实验会选取第 15 个 epoch 训练出来的结果。批量大小控制成 16。本文的模型和 RESA 模型的优化器是随机梯度下降，起始的学习率为 0.025，权重衰减系数为10^{-4}, 动量大小为 0.9。LaneATT 模型则选用 Adam 优化器，初始学习率为 0.0003。本文的模型和 RESA 在图像预处理部分相同，由于数据集的图片的照片拍到了地平线，地平线上方没有车道线，因此 240 个像素高度的图片和标签会被截断。图片也会被缩小为[288,800]的大小。我们的模型用于训练用的图片会经历一次正则化，三个通道上送入的均值为：103.939,116.779,123.68，另外，还会对训练用的图片做随机旋转。其他的部分请参考相关开源代码。

Fast-FSA 会将 backbone 输入的通道维度削减到 64。PSA 模块计算的中间过程通道也会降低一半。CA 模块会将通道数削减 32 倍。

4.4.4 定性实验

图 4-5 ResNet18(左)和 ResNet18+Fast-FSA+PSA(右)的可视化

加上本文提出的模块后，能在微弱的视野和远光灯干扰下检测出难以检测道路线（第一行）。相比于左图，本文的模型检测时，偏好两旁的车道线（第二行和第三行）。还能对车辆的遮挡有巨幅的提升（第四行），并非常漂亮地推理出了中线。总体上，本文的模型还未进行曲线回归，就对遮挡有巨幅提升。

4.5 本章小节

如果把目标检测的网络结构分为 input、backbone、neck、head，那么，本章的改进作用在网络的 backbone 部分，本文的改进作用在了 backbone 和 neck 部分。backbone 加载了预训练的参数。

针对 ResNet 没有全局视野的问题，本文的改进赋予了 ResNet 全局感受野。本文在 ResNet 中嵌入了两个注意力机制。其中 PSA 模块加到 ResNet 中每一个残差模块里的第一个 3x3 卷积之后。CA 模块会贴在 ResNet 模块的后面。最后进行了实验数据的汇总。

5 结论与展望

5.1 结论

RESA 的信息传递方式拥有良好的抗遮挡能力，缺点是计算量太大，所以本文用 YOLOF 的结构实现了 RESA 信息传递的方式，提出了新的模块 Fast-FSA，用于模型的 neck 部分整合 backbone 的信息对遮挡也起到了良好的效果在分割任务上的效果显著超过注意力机制。

ResNet 不加深网络就没有全局感受野，就难以建模长距离依赖，本文用两种注意力机制赋予 ResNet 全局感受野。其中一个注意力机制 PSA 嵌入到了 ResNet 的残差块中，另一个 CA 贴在了 ResNet 后面，达到了不错的效果。

本文的模型整合了上述改进措施，在精度不下降的情况下相比 RESA 减少了一半的计算量。而且本文提出的模块 Fast-FSA 不仅适用于基于分割的 RESA，也适用于基于目标检测的 LaneATT，即插即用，尤其在分割任务上的效果显著超越注意力机制。

5.2 展望

本文贡献了新的模块 Fast-FSA 用于对抗遮挡，改进了 ResNet。Fast-FSA 模块的多尺度特征融合思想出于对实时性的追求，仅仅使用了特征图的一层特征。高精度方法基本上都用到了 backbone 的多级特征，甚至是图神经网络。此外，高精度模型中，几乎没有将车道线检测当成视频检测任务，可能是这样做会极大降低实时性，所以没有多少研究者跟这个方向。本文提出的模型仅仅跑了一个数据集，如果有可能的话，还是希望在不同的数据集上观察结果，甚至跨数据集观察模型的泛化性。此外，基于分割的车道线检测模型可能自由度太高，预测的车道线形状观感上不如使用曲线回归出来的车道线，原因可能是取点的时候，没有考虑到车道线关键点之间的几何信息有何联系，建立这类联系的方法，在基于关键点检测的车道线标注模型中经常用到。

参考文献

[1] Zheng, Tu, Hao Fang, Yi Zhang, Wenjian Tang, Zheng Yang, Haifeng Liu, and Deng Cai. “RESA: Recurrent Feature-Shift Aggregator for Lane Detection.” *ArXiv:2008.13719 [Cs]*, March 25, 2021.

[2] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Deep Residual Learning for Image Recognition.” *ArXiv:1512.03385 [Cs]*, December 10, 2015.

[3] Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, et al. “An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale.” *ArXiv:2010.11929 [Cs]*, June 3, 2021.

[4] Lucas Tabelini, Rodrigo Berriel, Thiago M. Paixao, Claudine Badue, Alberto F. De Souza, and Thiago Oliveira-Santos. Keep your eyes on the lane: Real-time attention-guided lane detection. In CVPR, 2021.

[5] Lizhe Liu, Xiaohao Chen, Siyu Zhu, and Ping Tan. Condlanenet: A top-to-down lane detection framework based on conditional convolution. In ICCV, 2021.

[6] “Focus on Local: Detecting Lane Marker from Bottom Up via Key Point.” In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 14117–25. Nashville, TN, USA: IEEE, 2021.

[7] Wang, Jinsheng, Yinchao Ma, Shaofei Huang, Tianrui Hui, Fei Wang, Chen Qian, and Tianzhu Zhang. “A Keypoint-Based Global Association Network for Lane Detection,” n.d., 9. In CVPR, 2022.

[8] Zheng, Tu, Yifei Huang, Yang Liu, Wenjian Tang, Zheng Yang, Deng Cai, and Xiaofei He. “CLRNet: Cross Layer Refinement Network for Lane Detection,” n.d., 10. In CVPR, 2022.

[9] Jayasinghe, Oshada, Damith Anhettigama, Sahan Hemachandra, Shenali Kariyawasam, Ranga Rodrigo, and Peshala Jayasekara. “SwiftLane: Towards Fast and Efficient Lane Detection.” *ArXiv:2110.11779 [Cs]*, October 22, 2021.

[10] Qin, Zequn and Wang, Huanyu and Li, Xi. (2020). Ultra Fast Structure-aware Deep Lane Detection. The European Conference on Computer Vision (ECCV).

[11] Lin, T. Y. , et al. "Feature Pyramid Networks for Object Detection." 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) IEEE Computer Society, 2017.

[12] Chen, Qiang, Yingming Wang, Tong Yang, Xiangyu Zhang, Jian Cheng, and Jian Sun. (2021) “You Only Look One-Level Feature.” IEEE Conference on Computer Vision and Pattern Recognition

[13] Yu, Fisher, Dequan Wang, Evan Shelhamer, and Trevor Darrell. (2018) “Deep Layer Aggregation.” IEEE Conference on Computer Vision and Pattern Recognition

[14] Abualsaud, Hala, Sean Liu, David Lu, Kenny Situ, Akshay Rangesh, and Mohan M. Trivedi. “LaneAF: Robust Multi-Lane Detection with Affinity Fields.” *ArXiv:2103.12040 [Cs]*, August 19, 2021.

[15] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015.

[16] Xu, Hang, Shaoju Wang, Xinyue Cai, Wei Zhang, Xiaodan Liang, and Zhenguo Li. “CurveLane-NAS: Unifying Lane-Sensitive Architecture Search and Adaptive Point Blending.” *ArXiv:2007.12147 [Cs]*, July 23, 2020.

[17] Jie, H. , et al. "Squeeze-and-Excitation Networks." IEEE Transactions on Pattern Analysis and Machine Intelligence PP.99(2017).
[18] Wang, Q. , Wu, B. , Zhu, P. , Li, P. , & Hu, Q. . (2020). ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE.
[19] Li, Xiang, Wenhai Wang, Xiaolin Hu, and Jian Yang. “Selective Kernel Networks.” 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)
[20] Woo, Sanghyun, Jongchan Park, Joon-Young Lee, and In So Kweon. “CBAM: Convolutional Block Attention Module.” arXiv, July 18, 2018.
[21] Zhang, Yujun, Lei Zhu, Wei Feng, Huazhu Fu, Mingqian Wang, Qingxia Li, Cheng Li, and Song Wang. “VIL-100: A New Dataset and A Baseline Model for Video Instance Lane Detection.” *ArXiv:2108.08482 [Cs]*, August 18, 2021.
[22] Pan, Xingang, Jianping Shi, Ping Luo, Xiaogang Wang, and Xiaoou Tang. “Spatial As Deep: Spatial CNN for Traffic Scene Understanding.” arXiv, December 17, 2017.
[23] X. Li, J. Li, X. Hu and J. Yang, "Line-CNN: End-to-End Traffic Line Detection With Line Proposal Unit," in IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 1, pp. 248-258, Jan. 2020, doi: 10.1109/TITS.2019.2890870.
[24] Liu, Ruijin, Zejian Yuan, Tie Liu, and Zhiliang Xiong. “End-to-End Lane Shape Prediction with Transformers.” arXiv, November 28, 2020.
[25] Tusimple benchmark. https://github.com/TuSimple/tusimple-benchmark
[26] T. Liu, Z. Chen, Y. Yang, Z. Wu and H. Li, "Lane Detection in Low-light Conditions Using an Efficient Data Enhancement: Light Conditions Style Transfer," 2020 IEEE Intelligent Vehicles Symposium (IV), 2020, pp. 1394-1399, doi: 10.1109/IV47402.2020.9304613.
[27] Ghafoorian, Mohsen, Cedric Nugteren, Nóra Baka, Olaf Booij, and Michael Hofmann. “EL-GAN: Embedding Loss Driven Generative Adversarial Networks for Lane Detection.” In *Computer Vision – ECCV 2018 Workshops*, edited by Laura Leal-Taixé and Stefan Roth, 11129:256–72. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019.
[28] Szegedy, Christian, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi. “Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning.” arXiv, August 23, 2016.
[29] Seungwoo Yoo, Hee Seok Lee, Heesoo Myeong, Sungrack Yun, Hyoungwoo Park, Janghoon Cho, and Duck Hoon Kim. End-to-End Lane Marker Detection via Row-wise Classification. In IEEE CVPR Workshop, 2020.
[30] Ruochen Fan, Xuanrun Wang, Qibin Hou, Hanchao Liu, and Tai-Jiang Mu. SpinNet: Spinning Convolutional Network for Lane Boundary Detection. Computational Visual Media, 5(4):417–428, 2019.
[31] Liu, Huajun, Fuqiang Liu, Xinyi Fan, and Dong Huang. “Polarized Self-Attention: Towards High-Quality Pixel-Wise Regression.” arXiv, July 8, 2021.
[32] Hou, Qibin, Daquan Zhou, and Jiashi Feng. “Coordinate Attention for Efficient Mobile Network Design.” In 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 13708–17. Nashville, TN, USA: IEEE, 2021.
[33] Milletari, Fausto, Nassir Navab, and Seyed-Ahmad Ahmadi. (2016). V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016 Fourth International Conference on 3D Vision (3DV). IEEE.