

doi: 10.19562/j.chinasae.qcgc.2024.11.008

地下矿道无人车可行驶区域检测算法*

陈志军^{1,2}, 王朝伟^{1,2}, 吴超仲^{1,3,4}, 钱 闯¹, 吴怀主⁵, 申广俊⁵

(1. 武汉理工大学智能交通系统研究中心, 武汉 430063; 2. 武汉理工大学计算机与人工智能学院, 武汉 430070;

3. 交通信息与安全教育部工程研究中心, 武汉 430063; 4. 湖北文理学院, 襄阳 441053;

5. 东风汽车有限公司东风商用车技术中心, 武汉 430056)

[摘要] 地下矿道可行驶区域检测是地下矿山自动驾驶系统的关键感知技术,然而地下矿道光照强度低、工况复杂的特点给该任务带来极大挑战。鉴于此,本文提出一种地下矿道可行驶区域检测算法。首先,为解决地下矿道细节退化导致图像特征难以提取的问题,提出一种双分支特征提取骨干网络;然后,针对地下矿道可行驶区域检测不完整问题,提出一种自适应多尺度空间空洞池化金字塔特征增强模块;最后,为解决地下矿道边界提取不准确的问题,设计一种双分支通道注意力机制融合模块。在自制矿道可行驶区域数据集上进行实验,相较于 Deeplabv3+、UNet、DDRNet-23、PIDNet, 本文算法取得最佳效果,在 MIoU 分数上分别提升 2.07、2.39、1.87、1.92 个百分点,在 mAcc 分数上分别提升 1.78、2.45、1.84、1.86。本文算法已成功应用于地下无人驾驶矿车,验证了其在真实矿道场景下的有效性。

关键词: 自动驾驶; 井工无人矿车; 可行驶区域检测; 语义分割

Algorithm for Detecting Free Space in Underground Mine Tunnels for Autonomous Vehicles

Chen Zhijun^{1,2}, Wang Chaowei^{1,2}, Wu Chaozhong^{1,3,4}, Qian Chuang¹,
Wu Huaizhu⁵ & Shen Guangjun⁵

1. Intelligent Transportation Systems Research Center, Wuhan University of Technology, Wuhan 430063;

2. School of Computer Science and Artificial Intelligence, Wuhan University of Technology, Wuhan 430070;

3. Engineering Research Center of Transportation Information and Safety, Ministry of Education, Wuhan 430063;

4. Hubei University of Arts and Science, Xiangyang 441053;

5. Dongfeng Commercial Vehicle Technical Center, Dongfeng Commercial Vehicle Co., Ltd., Wuhan 430056

[Abstract] The detection of free space in underground mine tunnels is the key sensing technology for underground mining autonomous driving systems. However, the characteristics of low illumination and complex working environment inside the tunnels bring great challenges to this task. In view of this, in this paper an algorithm for detecting free space in underground mine tunnels is proposed. Firstly, a dual-branch feature extraction backbone network is proposed to solve the problem of difficulty in extracting image features caused by the degradation of tunnel details. Secondly, for the problem of incomplete detection of drivable areas in underground mining tunnels, an adaptive multi-scale atrous spatial pyramid pooling feature enhancement module is proposed. Finally, a dual-branch channel attention mechanism fusion module is developed to solve the problem of inaccurate boundary extraction in the underground mine tunnels. The experiments are conducted on a self-made dataset specifically designed for underground mine tunnels. The results show that the proposed algorithm surpasses other existing methods such as Dee-

* 国家自然科学基金(52332010, 52072288)、湖北省重点研发计划项目(2022BAA078)和武汉市科技计划项目(2023010402040022)资助。

原稿收到日期为 2024 年 06 月 07 日,修改稿收到日期为 2024 年 07 月 09 日。

通信作者: 吴超仲, 教授, 博士, E-mail: wucz@whut.edu.cn。

plabv3+, UNet, DDRNet-23, and PIDNet, with an increase of 2.07, 2.39, 1.87, and 1.92 percentage points in terms of MIoU scores, and 1.78, 2.45, 1.84, and 1.86 in terms of mAcc scores, respectively. The effectiveness of the proposed algorithm has been validated through its successful application in real mine tunnel scenarios, particularly for underground mining autonomous driving vehicles.

Keywords: autonomous driving; unmanned mine truck; free space detection; semantic segmentation

前言

随着我国国民经济由高速增长阶段转向高质量发展阶段,煤炭需求增速放缓、产能过剩以及煤炭工业结构性调整等问题凸显,智慧矿山的智能生产模式和轻资产运营方式将成为传统矿山行业的突破口。

在智慧矿山构建的过程中,地下无人矿车是重要的一环。无人矿车的应用能够减少人员进入危险区域,降低事故风险,且无人矿车能够全天候连续作业,有效提高生产效率。

可行驶区域检测在地下无人驾驶中扮演关键角色,通过精准感知车辆可安全行驶的区域,可以为路径生成提供参考,保障车辆运行安全。

在可行驶区域检测领域,相机是最常用的传感器,较于激光雷达,价格便宜,且可以提供高分辨率的影像信息。因此本文提出基于相机的地下矿道可行驶区域检测方法,用语义分割的方法给图像中的每个像素分配可行驶区域和背景类标签,实现对地下矿道图像的分析解释。

随着深度学习的发展,从全卷积网络(fully convolutional networks, FCN)^[1]开始,深度学习方法在语义分割领域逐渐占据主导地位。深度学习方法用于语义分割主要可以分成两个阶段。

第1阶段主要关注模型准确性,通过堆叠大量深层的卷积神经网络或Transformer结构,取得很高精度。Unet^[2-4]使用编码器和解码器结构,并引入跳跃连接,在图像分割等领域取得广泛应用。Poudel等^[5]调整网络结构,所提出模型能够在嵌入式设备上高效计算。Chen等^[6-9]系列通过在网络中集成具有不同膨胀率的扩张卷积,扩大感受野,实现对先前工作的巨大改进。金字塔场景解析网络(pyramid scene parsing network, PSPNet)^[10]引入金字塔模块来解析多尺度上下文信息。Strudel等^[11]提出基于Transformer模型的语义分割方法,该方法在第一层和整个网络中建模全局上下文,以帮助解决图

像分割中个别图像块的歧义。Cheng等^[12]将语义分割和全景分割任务统一为掩码分类问题,简化语义和全景分割任务。Zhao等^[13]利用预训练的文本到图像扩散模型提供视觉感知任务中的语义信息,表现出显著的效果。

在第2个阶段,为兼顾实时性,研究人员付出了大量努力来重新设计网络架构,这一阶段工作可以概括为轻量级编码解码器和多分支网络。Paszke等^[14]通过采用轻量级解码器并在早期阶段对特征图进行下采样,大大提高推理速度。Howard等^[15-17]系列用深度可分离卷积代替传统卷积,降低整体模型复杂度。Gamal等^[18]采用移动端主干网络^[19-20],结合信道混洗和组卷积,以降低计算成本。Yu等^[21]提出一种双分支网络架构,该架构包含两个具有不同深度的分支,用于上下文嵌入和细节解析。然后,基于该架构的一些方法被提出,以提高其表示能力和降低模型复杂性。Pan等^[22]引入双边连接,以增强上下文和细节分支之间的信息交换,达到实时语义分割的优秀结果。Xu等^[23]借鉴自动化工程领域的PID(proportional integral derivative)概念,并提出一种三分支网络架构,并利用边界来监督上下文和详细特征的融合。

尽管现有算法在许多任务已取得很好效果,但矿道可行驶区域检测依然面临以下问题:

(1)地下矿道工况复杂,且光照强度低会导致图像特征退化,而这些语义分割模型大多针对城市道路设计,无法有效提取矿道特征。

(2)地下矿道可行驶区域和背景相似度高,且空间信息难以获取,上述模型容易将可行驶区域内部的像素点识别为背景,导致可行驶区域内部检测不完整。

(3)地下矿道中墙壁和道路之间并无明显的分界线,在道路和墙壁的边界处,这些方法会出现严重失真问题。

针对上述问题,本文提出以下解决方案:

(1)提出一种轻量化双分支特征提取骨干网络,

提高网络特征提取能力;

(2)提出一种自适应多尺度空间空洞池化金字塔特征增强网络,对两个分支的特征进行增强,使得网络能够充分理解地下矿道可行驶区域与背景的区别;

(3)设计一种双分支通道注意力机制融合模块,实现高低分辨率两条通道边界信息的特征融合。

为验证本文方法,在矿道可行驶区域数据集进行实验。结果显示,在所有模型中,本文提出的模型取得最高精度。

1 矿道可行驶区域检测方法设计

1.1 框架概览

地下矿道光照条件差、工况复杂,可行驶区域和背景相似度高,墙壁与道路分界不明显,现有分割方法无法准确区分。鉴于此,提出一种基于双分支特征融合的地下矿道可行驶区域实时检测网络(real-time dual-branch fusion network for detection of free space in mining tunnels, DDMNet)。整体框架如图1所示,由特征提取模块、特征融合增强模块以及掩码

构建模块组成。

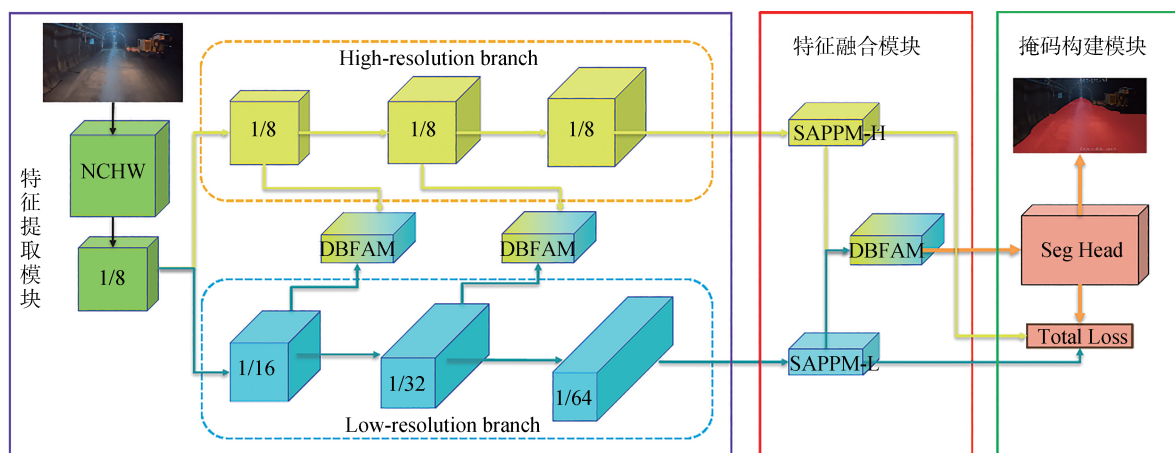
特征提取模块采用轻量化双分支融合骨干网络(mobile dual-branch fusion network, MDFNet)进行特征提取。特征融合增强模块采用自适应多尺度空间空洞池化金字塔(self-adaption atrous spatial pyramid pooling module, SAPPM)和双分支通道注意力融合模块(dual branch spatial feature attention fusion module, DBFAM)进行特征增强和融合。掩码构建模块使用一个简单分割头和损失函数实现可行驶区域检测下游任务。

1.2 轻量化双分支融合骨干特征提取网络

1.2.1 轻量化双分支融合网络架构

地下矿道光照条件差、工况复杂,使得特征细节退化,难以提取。为此,本文提出一种轻量化双分支融合骨干网络MDFNet,以增强特征提取能力。

MDFNet架构如图1左侧紫色方框内所示,由高分辨率特征分支和低分辨率特征分支组成。高分辨率分支的输入特征图大小为图像的1/8,能够提供丰富的矿道场景语义信息。低分辨率分支的输入特征图大小为图像的1/16,能够提供地下矿道边界空间引导信息。



注:图中的DBFAM均属于特征融合模块,为使流程图看起来清晰,故未将所有DBFAM放入特征融合模块所在红色方框。

图1 DDMNet框架图

1.2.2 网络模块细节

MDFNet算法的细节如表1所示。首先在主干分支采用多个顺序倒残差模块对原始特征图进行8倍下采样;然后使用普通bottleneck block对低分辨率分支进行下采样,使用大卷积核bottleneck block对高分辨率分支进行特征提取;分支间使用DBFAM模块进行双向特征融合;最后在每个分支的末尾添加一个bottleneck block,同时使用SAPPM进行多尺

度特征聚合。

MDFNet中使用的Inverted Residual Block是一种常见模块结构,具体如图2(a)所示。使用 1×1 卷积增加通道数,通过分组卷积减少参数量,可以将计算量降低到普通卷积的 $1/8 \sim 1/9$ 。

MDFNet还采用大量Bottleneck Block^[24]结构,具体如图2(c)所示。该结构通过 1×1 的卷积减少通道数,然后使用 3×3 的卷积核进行特征提取,最后再使

表1 MDFNet结构参数

层名	输出特征图尺寸($H \times W \times C$)		网络层主要结构	
L1	128×128×128		Inverted Residual Block×1, stride=2	
L2	64×64×256		Inverted Residual Block×1, stride=2	
L3	32×32×64		Inverted Residual Block×4	
L4	16×16×128	32×32×64	bottleneck block×2	大核 bottleneck block×2
	DBFAM		DBFAM 模块×1	
L5	8×8×256	32×32×96	bottleneck block×2	大核 bottleneck block×2
	DBFAM		DBFAM 模块×1	
L6	4×4×384	32×32×128	bottleneck block×2	bottleneck block×2
	SAPPM-L	SAPPM-H	SAPPM-L×1	SAPPM-H×1
	DBFAM		DBFAM	

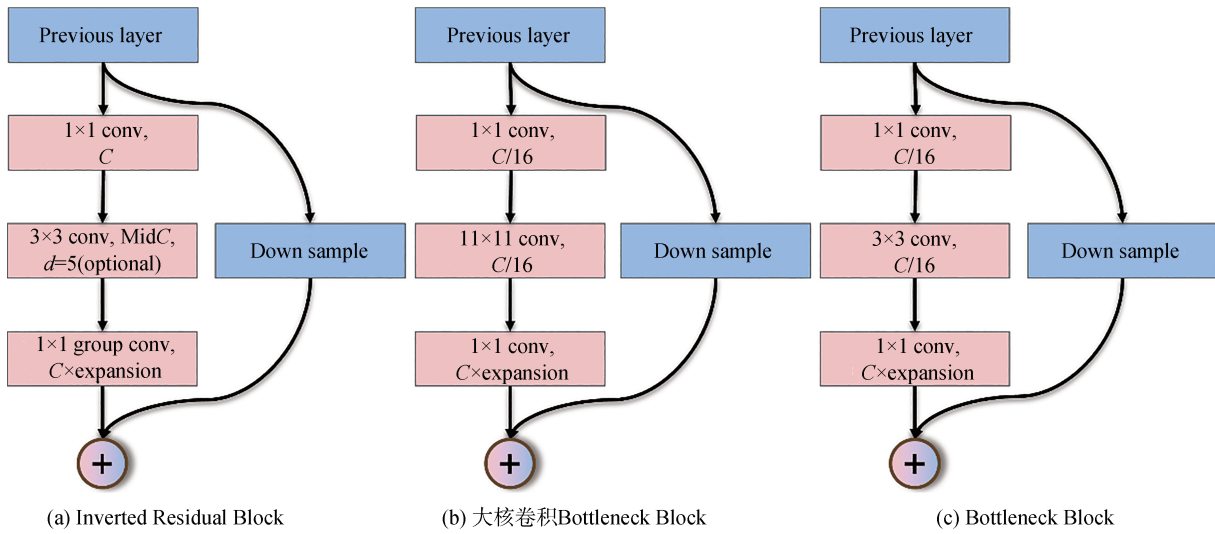


图2 MDFNet基本模块

用 1×1 的卷积核来调整通道数。同时,该结构还引入残差连接,以缓解梯度消失问题。残差学习的计算公式为

$$y = F(x, \mathbf{W}_i) + x \quad (1)$$

式中: $F(x, \mathbf{W}_i)$ 是权重为 \mathbf{W}_i 的残差映射输出; x 是残差模块的输入; y 是残差模块的输出。

较大卷积核可以增加感受野范围,基于此,本文对Bottleneck Block进行改进,如图2(b)所示。使用 11×11 卷积核替代 3×3 卷积核,将通道数量缩减为原来的 $1/16$,以平衡参数量。将该模块应用在高分辨率分支,以提高模型对全局信息的感知能力。

1.3 自适应多尺度空间空洞池化金字塔结构

如图3(a)所示,在地下矿道可行驶区域检测任务中,经常出现可行驶区域内部检测不完整的现象,造成该现象的原因有:

(1)低照度场景下,地下矿道背景与可行驶区域特征相似,不易区分;

(2)网络没有充分学习到可行驶区域的空间位置信息。

多尺度特征信息的融合可以增强模型提取到的特征,并使模型学习到空间位置信息,解决内部检测不完整问题。鉴于此,基于特征金字塔^[25],针对两个分支设计了自适应多尺度空间空洞池化金字塔(self-adaption atrous spatial pyramid pooling module, SAPPM),如图4所示,其计算公式为

$$y_i = \begin{cases} C_{1 \times 1}(x), & i = 1 \\ C_{3 \times 3}(U(C_{1 \times 1}(P_{i,k}, p)) + y_{i-1}), & 1 < i \leq n \end{cases} \quad (2)$$

式中: $C_{1 \times 1}$ 表示 1×1 卷积; $C_{3 \times 3}$ 表示 3×3 卷积操作; U 代表上采样; $P_{i,k}$ 代表第 i 个分支池化核为 k 的最大池化; p 表示池化的填充大小,默认 $p = (k/2) + 1$; y_i 表示第 i 个分支输出。

低分辨率分支通过一个 1×1 卷积映射作为残差连接通道,采用步长分别为1、2、4、0,池化核大小分别

为1、3、5和全局的最大池化,生成不同大小的特征图,最后将所有特征图拼接并压缩回原始通道数量,称为SAPPM-L。由于高分辨率分支的特征图尺度更大,因此自适应地调整卷积核大小,计算过程与SAPPM-L相同,称为SPAPPM-H。具体参数详见表2。

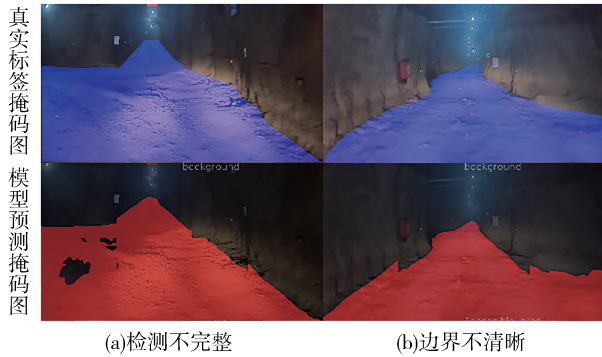


图3 可行驶区域检测常见问题

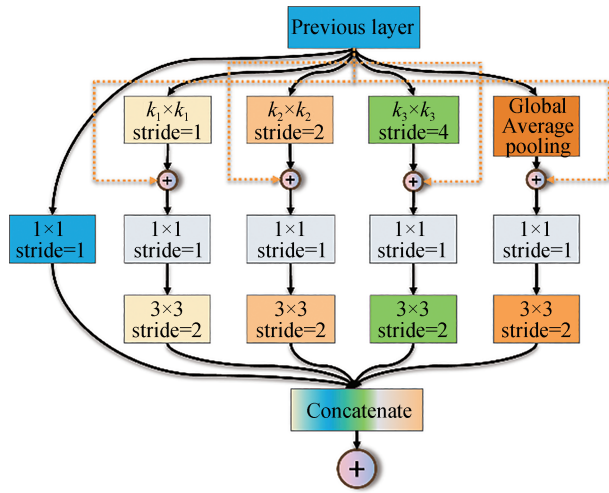


图4 SAPPM结构图

表2 SAPPM详细结构

方法	池化核大小	填充	步长
SAPPM-L	[1, 1, 3, 5, global]	[0, 1, 2, 4, 0]	[1, 1, 2, 4, global]
SAPPM-H	[1, 5, 9, 13, global]	[0, 2, 4, 6, 0]	[1, 1, 2, 4, global]

1.4 双分支通道注意力融合模块

如图3(b)所示,在地下矿道可行驶区域检测任务中,经常出现边界不清晰现象,其原因是可行驶区域边界与背景分界不明显。

双分支网络的低分辨率分支可以捕获边界的高层次语义特征以提供空间引导,高分辨率分支则保留局部细节特征以防止边界细节丢失,融合两个分支后的特征同时具备细节特征和高层语义特征,有助于解决边界不清晰问题。

为此,基于通道注意力机制^[26],设计一种双分支通道注意力融合模块(dual branch spatial feature attention fusion module, DBFAM),实现两个分支间的特征融合,如图5所示,包括分支内特征融合、高到低融合和低到高融合3个部分。

分支内特征融合指对每个分支内的特征图进行融合。使用1×1卷积核降低通道数,然后使用3×3的卷积核进行特征提取,然后使用1×1卷积核将通道数还原,同时伴随一个通道注意力残差连接。

如图5所示,高到低融合将高分辨率分支特征融合进低分辨率特征分支。通过步幅为2的3×3卷积核进行下采样,然后再逐点相加。而低到高融合则是将低分辨率分支特征融合进高分辨率特征分支。先通过1×1卷积压缩通道,然后通过双线性插值进行上采样,最后再逐点相加,其计算公式为

$$\begin{cases} y_{low} = R(y'_{high} + M_{c1}(C(x_1))) \\ y_{high} = R(y'_{low} + M_{c2}(C(x_2))) \end{cases} \quad (3)$$

式中: y_{low} 表示低分辨率分支融合高分辨率特征后的输出; R 表示Relu激活函数; y'_{high} 表示高分辨率融合之前的中间值; y_{high} 表示高分辨率分支融合低分辨率特征后的输出; x_1 和 x_2 分别表示两个分支的输入; C 表示卷积操作; M_{c1} 和 M_{c2} 分别代表两个分支的通道注意力; y'_{low} 表示低分辨率通道融合之前的中间值。

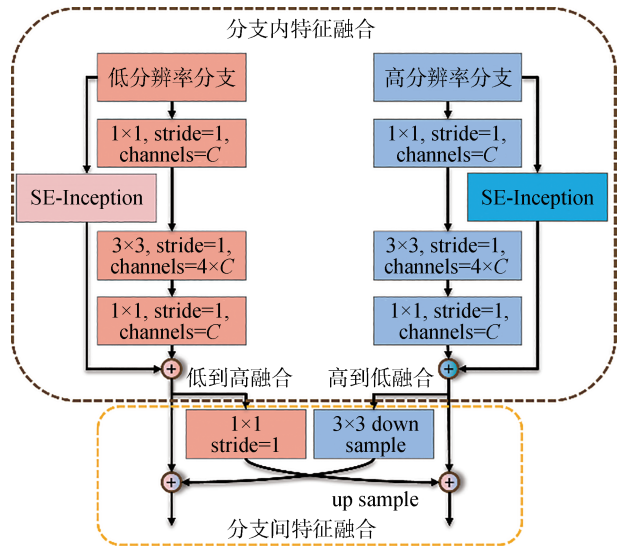


图5 DBFAM结构图

通道注意力结构如图6所示。将卷积后的特征图与其自身再经空间卷积后获得的1×1×C特征图相乘,该1×1×C特征图包含全局信息,它便是通道注意力,其计算公式为

$$M_c(F) = \sigma(MLP(AugPool(F) + MaxPool(F))) \quad (4)$$

式中: σ 表示sigmoid激活函数; $M_c(F)$ 表示最后一维通道注意力权重矩阵。

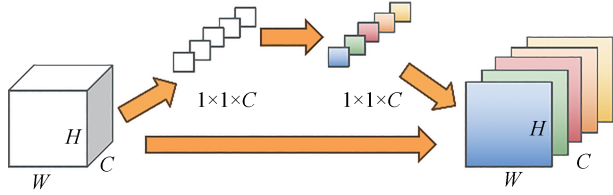


图6 通道注意力结构图

1.5 分割掩码头部设计

1.5.1 分割头设计

分割头结构如图7所示。首先使用 3×3 卷积对上一层特征图进行特征提取,然后再使用 1×1 大小的卷积核将特征图的通道数调整为2(可行驶区域与背景),通过双线性插值对特征图进行上采样,得到最终预测图。

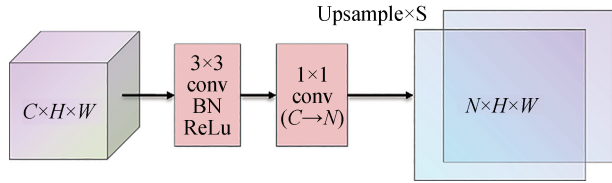


图7 分割头设计

1.5.2 损失函数设计

如图1所示,在高分辨率分支和低分辨率分支末尾引入两个辅助损失函数,主损失函数位于DBFAM模块之后。训练时,启用两个辅助分支,测试时只使用主分支预测。总训练损失计算如下:

$$L_f = L_M + \alpha L_{low} + \beta L_{high} \quad (5)$$

式中: L 代表OhemCrossEntropy^[27]损失函数; L_f 代表总损失; L_M 代表主分支损失; α 代表低分辨率分支损失权重; L_{low} 代表低分辨率分支损失; β 代表高分辨率分支损失权重; L_{high} 代表高分辨率分支损失。

在本文中,由于背景区域远大于可行驶区域,负类样本远多于正类样本,数据分布不均衡。为此,采用OhemCrossEntropyLoss作为损失函数,以选择具有较高损失值的困难样本进行梯度更新,定义如下:

$$OCELoss = -\frac{1}{N} \sum_{i=1}^N \begin{cases} \log(p_i), & y_i = 1 \\ \log(1 - p_i), & y_i = 0 \\ 0, & \text{其他} \end{cases} \quad (6)$$

式中: $OCELoss$ 表示OhemCrossEntropyLoss损失函

数; N 表示批次; p_i 为模型预测目标类的概率; y_i 为真实标签。

2 实验验证

2.1 矿道可行驶区域数据集

利用如图8所示的无人矿货车前视相机在湖北省某地下矿区长约10 km的地下矿道采集图像98 030帧。经筛选,保留3 000帧质量较高的图片,使用labelme工具进行像素级标注,分为可行驶区域和背景区域两个类别。按照7:1:2的比例随机将数据集划分为训练集、验证集和测试集。



图8 数据采集所采用的无人实验车

如图9和图10所示,将测试集分为困难版本(test-hard)和常规版本(test-normal)。将困难版本(test-hard)分为雾气、低照明强度、噪声光、渗水及

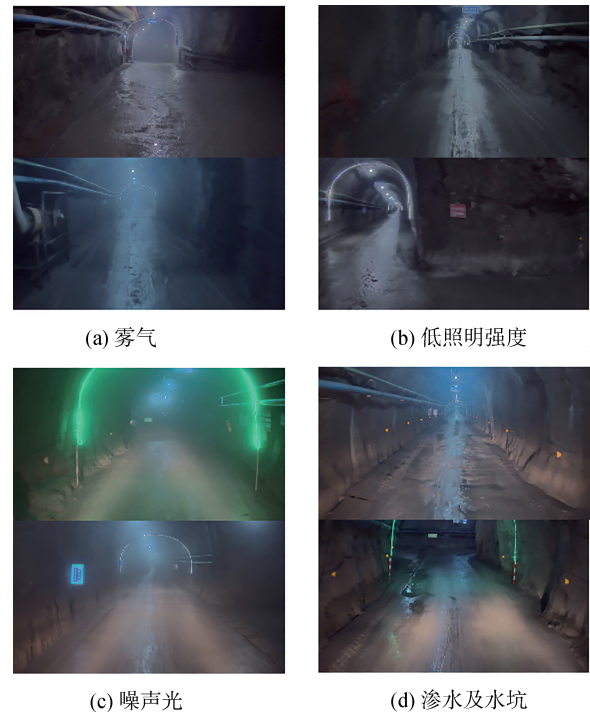


图9 矿道可行驶区域数据集困难版本

水坑4类状况。将常规版本分为长直道、会车区域、弯道、斜坡、上料卸料区5类工况。

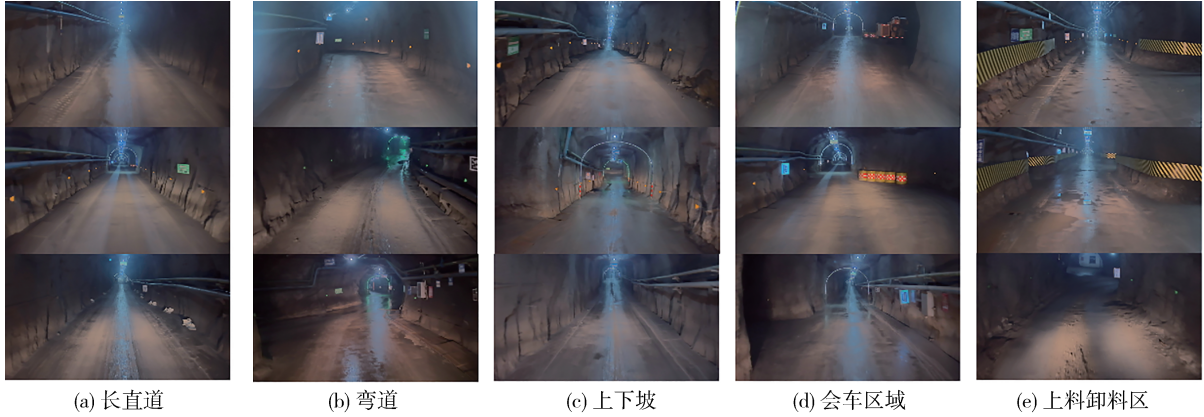


图10 矿道可行驶区域数据集常规版本

2.2 实验配置及性能评估指标

2.2.1 实验环境配置及超参数设置

本文实验运行在 Windows 11 操作系统的计算机上,使用一块 Nvidia RTX Laptop 4060 8GB 显卡对提出的语义分割模型以及主流语义分割模型进行训练。采用 PyTorch 深度学习框架, CUDA 版本为 12.0, PyTorch 版本为 1.11.0。使用 SGD 优化器,学习率采用多项式衰减策略,学习率计算公式为

$$l_r = \max\left(l_{r_initial} \times \left(1 - \frac{it}{it_{max}}\right)^{power}, l_{r_min}\right) \quad (7)$$

式中: l_r 表示学习率; $l_{r_initial}$ 代表初始学习率,设为0.01; l_{r_min} 表示最小学习率,设为0.0001; $power$ 表示衰减幂,设为0.9; it 表示迭代次数,设为60000。批次大小设为8,采用混合精度训练器,模型输入图像原始尺寸为256×256,采用随机缩放、随机裁剪等数据增强策略。

2.2.2 性能评估指标

本实验采用平均交并比(mean intersection over union, MIoU)、类均准确率(mean accuracy, mAcc)、每秒帧数(frame per second, FPS)、参数量(Params)和每秒浮点操作数(floating-point operations per second, FLOPs)对模型进行量化评价。MIoU和mAcc综合体现检测精度;FPS体现检测速度;Params与FLOPs体现模型规模,对训练耗时与处理速度产生影响。MIoU、mAcc和FPS计算如下:

$$\begin{cases} MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{TP+FN+FP} \\ mAcc = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{TP+FP} \\ FPS = \frac{Frames}{Time} \end{cases} \quad (8)$$

式中: k 表示数据集类别数目; TP 表示预测正确的像素正样本数; FN 为预测错误的像素负样本数; FP 为预测错误的像素正样本数; $Frames$ 为总检测帧数; $Time$ 为总检测时间。

2.3 对比实验设计

2.3.1 常规环境下模型精度对比

将本文所提算法与Fast-SCNN、Deeplabv3plus-R18、Deeplabv3plus-R50、UNet3+ FCN、MobileNetv3+ASPP、DDNet-23、PIDNet-S等先进方法在矿道可行驶区域数据集常规版本上进行比较,结果如表3所示。

表3 常规环境下不同模型精度对比 %

方法	MIoU(normal)	mAcc(normal)
SegNet	87.31	89.65
Fast-scnn	93.21	95.65
Deeplabv3+-r18+ FCN	93.97	96.26
Deeplabv3+-50+ FCN	94.03	96.31
MobileNetv3+ASPP	93.57	95.89
UNet + FCN	93.71	95.66
DDNet-23	94.23	96.25
PIDNet-S	94.18	96.23
DDMNet(ours)	96.10	98.09

本文算法在MIoU和mAcc两个指标上分别取得96.10%和98.09%的分数,表现最佳,相较目前最先进的方法分别提高1.87和1.78个百分点。

本节对所有对比方法输出结果进行可视化,结果如图11所示。可以看到,在相同的场景下,Deeplabv3plus-R18和UNet + FCN存在右侧边界划分不精确的问题,而DDNet-23和PIDNet-S误将墙壁划分为可行驶区域,本文提出的DDMNet避免了

这些问题,展现更加准确和稳定的结果。

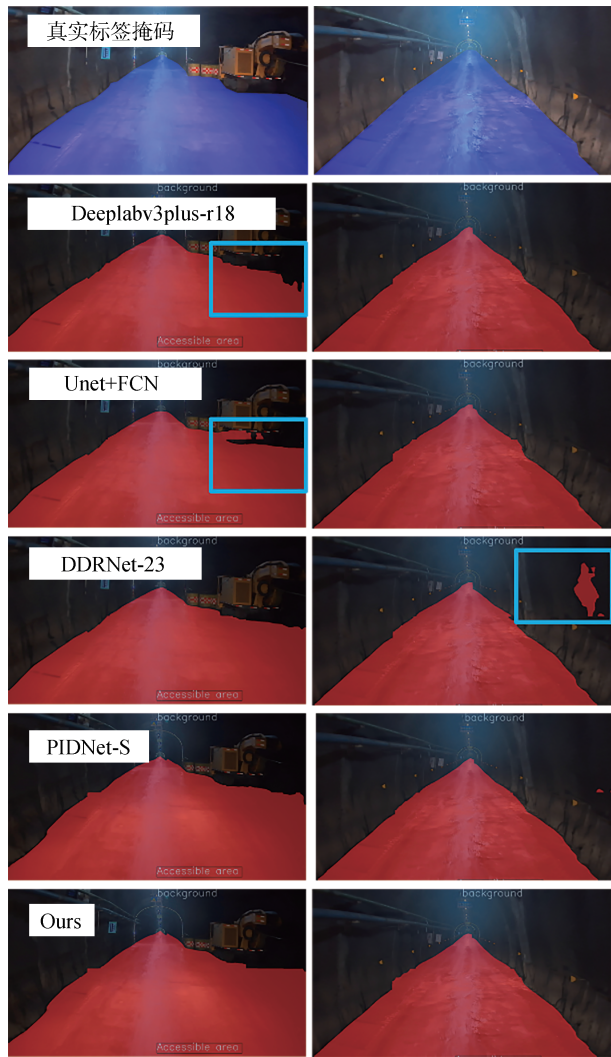


图11 常规环境下不同模型输出可视化结果

2.3.2 极端环境下模型精度对比

地下矿区存在环境光照条件差、噪声光以及水雾、水坑等特点,为验证算法对这些极端环境的鲁棒性,本节在测试集困难版本上与其他方法进行比较,结果如表4所示。

本文提出的算法在MIoU和mAcc两个指标上分别取得93.30%和95.22%的分数,相较于目前最先进的方法,分别提高1.98和2.08个百分点。

同时,本节在矿道可行驶区域数据集的困难版本测试集进行输出可视化,结果如图12所示。在噪声光、雾气、水坑影响下,模型给出和真实标签几乎一致的预测。在低照度下,可以观察到模型较为准确地给出了可行驶区域的预测,但在近处区域与真实值有微小偏差。

表4 极端环境下不同模型分割精度 %

方法	MIoU(hard)	mAcc(hard)
SegNet	82.11	84.81
Fast-senn	86.41	89.23
Deeplabv3+-r18+ FCN	89.88	91.86
Deeplabv3+-50+ FCN	90.83	92.15
MobileNetv3+ASPP	89.14	91.29
UNet + FCN	89.67	91.85
DDRNet-23	91.32	93.10
PIDNet-S	91.04	93.14
DDMNet(ours)	93.30	95.22

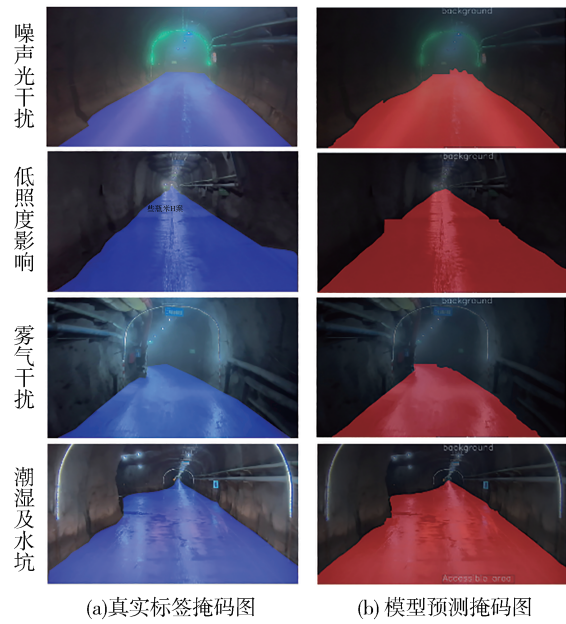


图12 常规版本测试集和困难版本分割可视化结果

综上所述,矿道中的一些极端复杂环境,如雾气会对可行驶区域的检测造成轻微影响,但模型总体上表现出很强的鲁棒性。

2.3.3 不同工况下模型精度对比

将测试集划分为长直道、会车区域、弯道、斜坡、上料卸料区5类工况,对模型效果进行测试。测试结果如表5所示,在各工况下,模型均能取得93.78%以上MIoU分数以及94.83%以上的mAcc分数。

表5 不同工况下的分割精度 %

工况	MIoU(hard)	mAcc(hard)
长直道	96.94	98.64
弯道	95.68	97.74
斜坡	96.17	96.21
会车区域	95.30	95.42
上料卸料区	93.78	94.83

不同工况的模型输出如图 13 所示,可以看到,在不同工况下,模型均表现良好,并无明显差异。综

上,本文提出的模型在各类工况下均具有很好的稳定性。

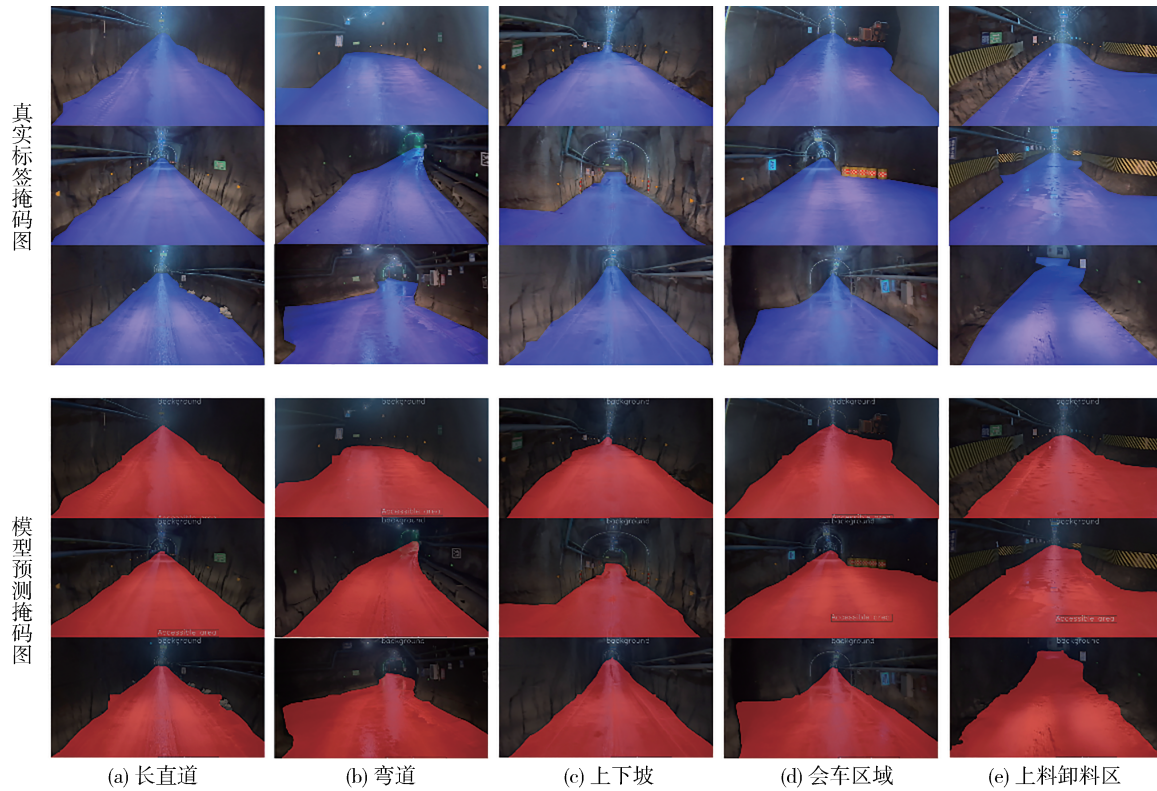


图 13 不同工况下模型输出可视化结果

2.3.4 模型推理性能对比

本节对模型的 Float、Param 和 FPS 与主流方法进行对比。实验结果如表 6 所示。模型针对井下检测任务的实时性和轻量化设计,使得模型的数量控制在 15M 之内,在所有方法中,模型的数量并非最低,但合理的分支结构设计使得模型运算并行性更好,在未使用任何 GPU 加速设备的情况下,模型在检测速度指标 FPS 上取得 12.24 的分数,仅次于轻量化网络结构 MobileNetv3+ASPP。综上,模型在速度和精度上取得最佳平衡效果。

2.4 消融实验设计

为验证所提改进方法的有效性,以 DDRNet 为基准方法,设计以下 4 组消融实验。

(1) MDFNet: 使用 MDFNet 骨干网络代替 DDRNet 骨干网络。

(2) MDFNet+SAPPM: 在实验 1 基础上使用本文提出的 SAPPM 代替 DDRNet 原有特征增强模块。

(3) MDFNet+DBFAM: 在实验 1 基础上引入本文提出的 DBFAM 模块。

(4) MDFNet+SAPPM+DBFAM: 在实验 1 基础上

表 6 模型推理性能对比结果

方法	Float	Param	FPS
SegNet	2.860G	29.5M	4.22
Fast-scnn	0.232G	1.4M	11.84
Deeplabv3+-r18+ FCN	13.565G	12.3M	10.71
Deeplabv3+-50+ FCN	44.121G	41.2M	5.39
MobileNetv3+ASPP	3.487G	4.4M	14.67
UNet + FCN	50.738G	30.0M	9.33
DDRNet-23	1.14G	20.1M	11.72
PIDNet-S	1.821G	7.6M	8.08
DDMNet(ours)	1.731G	14.6M	12.24

使用本文提出的 SAPPM 代替 DDRNet 原有特征增强模块,且引入本文提出的 DBFAM 模块。

消融实验结果如表 7 所示,消融实验 1 仅将 DDRNet 骨干网络更换为 MDFNet, MIoU 提升 1.02 个百分点, FPS 提升 0.68。这表明,相较于 DDRNet, MDFNet 骨干网络设计能够有效提取地下矿道特征,且计算速度略有提升。

根据消融实验 2 和消融实验 3, MDFNet+SAPPM 与 MDFNet+DBFAM 的两两组合分别使 MIoU 提升

表7 消融实验结果

实验方法	MDFNet	SAPPM	DBFAM	MIoU/%	FPS
DDRNet				94.23	11.72
MDFNet	√			95.35	12.40
MDFNet+SAPPM	√	√		95.70	12.32
MDFNet+DBFAM	√		√	95.59	12.20
MDFNet+SAPPM+DBFAM	√	√	√	96.10	12.24

0.35和0.24个百分点,FPS下降0.08和0.20。可见特征增强网络SAPPM的设计使得模型能够从多尺度理解深层特征图的空间语义信息,从而减少网络检测不完整问题。双分支通道注意力融合模块的设计使模型实现高分辨率和低分辨率两个通道特征的自适应融合,从而提高边界处像素识别的准确率。且SAPPM和DBFAM的设计未对模型推理速度造成太大影响。

根据消融实验4,在MDFNet基础之上引入SAPPM与DBFAM的组合之后,模型MIoU提升0.75个百分点,这个增幅大于单独引入SAPPM与DBFAM的MIoU增幅之和,可见SAPPM与DBFAM的组合会有相互促进作用。

2.5 可行驶区域检测实车验证

2.5.1 不同工况下的实车验证

为测试模型在实际矿道的通用性,利用如图8所示的新能源无人矿货车在地下矿区未曾采集过数据的路径段使用车身上前置固定定位相机进行可行驶区域实车检测。

不同工况下的实际检测结果如图14所示。上坡和弯道路段车速为8 km/h,直道和会车路段车速为15 km/h。可以看出,在实际地下矿道,即使训练数据集不包含该路径段,本文的算法仍能够准确地检测出可行驶区域。

总的来说,地下矿山的地下矿道环境和工况类似,而本研究的数据集涵盖了长直道、弯道、上下坡、会车区域、上料区、卸料区各类工况和极端环境,因此基于本文数据集训练出来的模型,在矿山场景具有良好通用性。

2.5.2 不同车速下的实车验证

地下矿区自动驾驶车速通常保持在8~20 km/h之间。为验证本文算法是否满足矿下自动驾驶车速要求,在3个不同的路径段,分别以10、15、20 km/h的速度展开实车验证,实际检测结果如图15所示。在不同车速下,模型均能稳定检测出可行驶区域,满足目前矿下的自动驾驶需求。

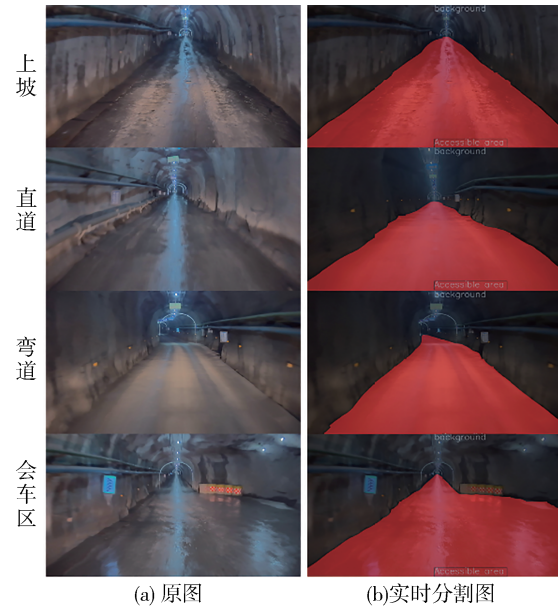


图14 实车验证效果图

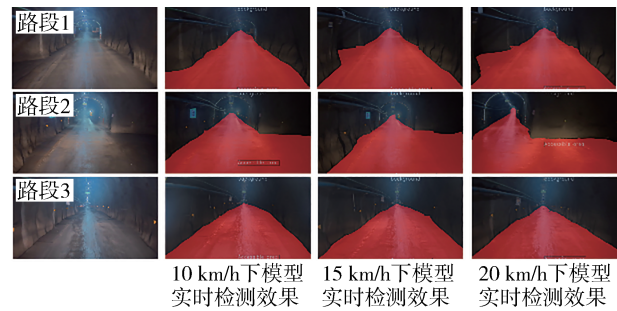


图15 实车验证效果图

3 结论

针对地下矿道可行驶区域检测中由于光照强度低、工况复杂而造成的不准确问题,提出一种基于语义分割的矿道可行驶区域实时检测网络。在真实地下矿道场景下,本文制作地下矿道可行驶区域数据集,并进行实车验证。结果表明,本文提出的网络模型达到最高的精度,具备良好的鲁棒性和泛化能力。然而,在光照强度低、大雾等极端复杂场景下,观察到模型的检测精度会有一定下降,带来未知风险。因此,未来的工作将着重于从极端环境的角度优化模型算法,以进一步提高模型的鲁棒性和可靠性。

参考文献

- [1] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015:

- 3431–3440.
- [2] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation [C]. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18*. Springer International Publishing, 2015: 234–241.
- [3] ZHOU Z, RAHMAN SIDDIQUEE M M, TAJBAKHS N, et al. Unet++: a nested u-net architecture for medical image segmentation [C]. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*. Springer International Publishing, 2018: 3–11.
- [4] HUANG H, LIN L, TONG R, et al. UNet 3+: a full-scale connected uNet for medical image segmentation [C]. *ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020: 1055–1059.
- [5] POUDEL R P K, LIWICKI S, CIPOLLA R. Fast-SCNN: fast semantic segmentation network [J]. *arXiv preprint arXiv: 1902.04502*, 2019.
- [6] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs [J]. *arXiv preprint arXiv: 1412.7062*, 2014.
- [7] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 40 (4) : 834–848.
- [8] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation [J]. *arXiv preprint arXiv: 1706.05587*, 2017.
- [9] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 801–818.
- [10] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 2881–2890.
- [11] STRUDEL R, GARCIA R, LAPTEV I, et al. Segformer: transformer for semantic segmentation [C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 7262–7272.
- [12] CHENG B, MISRA I, SCHWING A G, et al. Masked-attention mask transformer for universal image segmentation [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 1290–1299.
- [13] ZHAO W, RAO Y, LIU Z, et al. Unleashing text-to-image diffusion models for visual perception [C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023: 5729–5739.
- [14] PASZKE A, CHAURASIA A, KIM S, et al. ENet: a deep neural network architecture for real-time semantic segmentation [J]. *arXiv preprint arXiv: 1606.02147*, 2016.
- [15] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [J]. *arXiv preprint arXiv: 1704.04861*, 2017.
- [16] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: inverted residuals and linear bottlenecks [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 4510–4520.
- [17] HOWARD A, SANDLER M, CHU G, et al. Searching for mobileNetV3 [C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 1314–1324.
- [18] GAMAL M, SIAM M, ABDEL-RAZEK M. ShuffleSeg: real-time semantic segmentation network [J]. *arXiv preprint arXiv: 1803.03816*, 2018.
- [19] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 6848–6856.
- [20] MA N, ZHANG X, ZHENG H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design [C]. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 116–131.
- [21] YU C, WANG J, PENG C, et al. BiSeNet: bilateral segmentation network for real-time semantic segmentation [C]. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 325–341.
- [22] PAN H, HONG Y, SUN W, et al. Deep dual-resolution networks for real-time and accurate semantic segmentation of traffic scenes [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 24(3): 3448–3460.
- [23] XU J, XIONG Z, BHATTACHARYYA S P. PIDNet: a real-time semantic segmentation network inspired by PID controllers [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 19529–19539.
- [24] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770–778.
- [25] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37 (9): 1904–1916.
- [26] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 7132–7141.
- [27] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training region-based object detectors with online hard example mining [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 761–769.