

DOI:10.19651/j.cnki.emt.2212451

基于双模态融合的线缆图像语义分割方法研究^{*}

曹国群 刘桂雄

(华南理工大学机械与汽车工程学院 广州 510640)

摘要: 线缆敷设时需要严格控制最小弯曲半径,线缆敷设图像准确分割是控制弯曲半径的基础,传统视觉方法、经典语义分割方法对复杂环境下线缆细长特征目标分割效果不佳。本文提出一种基于改进双模态融合语义分割网络ESANet的线缆语义分割方法,使用高效的SAGate代替ESANet中RGB-D Fusion模块完成双模态特征校正与融合任务,融合特征分别同时参与后续两种模态的特征提取,实现细长特征线缆掩膜的准确分割。通过采集不同姿态的线缆RGB及对应深度图像进行实验,结果表明本文改进的ESANet网络对线缆等细长特征目标有较好分割效果,较ESANet模型分割精度(mIoU)提升了3.99%,较RGB单模态语义分割网络SwiftNet精度提升7.68%,该方法可以推广到其它具有细长特征的目标分割任务中。

关键词: 语义分割;双模态融合;深度学习;线缆

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 460

Research on semantic segmentation of cable image based on bimodal fusion

Cao Guoqun Liu Guixiong

(School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510640, China)

Abstract: Minimum bending radius needs to be strictly controlled when laying cables. Accurate segmentation of cable laying image is the basis of controlling bending radius. Traditional visual and classical semantics segmentation methods do not work well for target segmentation of long and thin cables in complex environment. This paper presents a new cable semantics segmentation method based on improved dual-mode fusion semantics for ESANet network. Instead of the RGB-D Fusion module in ESANet, an efficient SAGate is used to complete the dual-mode feature correction and fusion tasks. The fused features participate in the feature extraction of the subsequent two modes at the same time to achieve accurate segmentation of the thin feature cable mask. By collecting RGB and corresponding depth images of cables with different postures, the results show that the improved ESANet network has a good segmentation effect on slender feature targets such as cables, which is 3.99% higher than Net model segmentation accuracy (mIoU), and 7.68% higher than SwiftNet single-mode semantics segmentation network of RGB. This method can be extended to other target segmentation tasks with slender feature.

Keywords: semantic segmentation; dual-modal fusion; deep learning; cable

0 引言

线缆作为电力输送、信号传输等重要载体,其敷设弯曲半径有严格限制^[1-2],轨道车辆^[3]、飞机^[4]等装备中的线缆过弯会造成绝缘层破损、线芯破股,从而增加安全隐患^[5]。线缆弯曲半径测量主要分为人工与机器视觉分割测量两种,其中人工测量方法主要存在操作人员要求高、劳动强度大问题;机器视觉分割测量方法是从线缆图像中准确分割出线缆轮廓,再对其进行进一步处理得出线缆弯曲半径,开展线

缆敷设图像准确分割方法研究是线缆弯曲半径计算的前提、基础。

机器视觉分割包括传统视觉分割、深度学习语义分割两类,其中传统视觉分割方法主要通过控制颜色通道阈值、图像区域操作算法完成线缆轮廓区域分割,但在复杂背景下分割效果不佳^[6-7];经典机器视觉语义分割仅依靠RGB单模态特征,对线缆这类细长特征物体分割效果较差,还难以满足实际应用要求^[8-9]。目前,RGB+深度双模态特征机器视觉语义分割成为学者研究关注热点^[10],通常是使用多

收稿日期:2022-12-21

^{*} 基金项目:广东省重点领域研发计划项目(2019B010154003)资助

个并行特征提取网络分别提取 RGB 模态特征、深度特征,再进行轮廓分割。浙江大学研究注意力互补网络 (attention complementary network, ACNet),通过多个注意力模块选择性地从 RGB、深度分支收集特征并进行融合,使用 3 个并行特征提取网络分别对 RGB 信息、深度信息及融合信息提取特征,对特征图融合后统一解码生成分割掩膜,对 NYUv2 数据集 mIoU 达到 48.3%^[11];北京大学针对 RGB 与深度模态之间特征的差异性,提出一种用于双模态特征校正、融合的分隔聚合门结构 (separation-and-aggregation gate, SAGate),利用两种模态互补信息完成特征校正与融合,在 NYUv2 数据集上 mIoU 达到 52.4%^[12];Seichter 等^[13]研究一种搭载在移动终端的 RGB-D 语义分割网络 ESANet,多级融合深度特征指导 RGB 特征提取,在侧重实时推理基础上 NYUv2 数据集上 mIoU 达到 50.3%;文献[14]基于 ESANet 框架设计双模态特征融合模块 (attention mechanism bimodal fusion, AMBF),融合不同阶段特征跳跃连接参与解码过程,在同等实验条件下 NYUv2 数据集上 mIoU 较 ESANet 提高 1.9%;文献[15]在 ESANet 网络基础上提出基于邻域度量关系的 RGB-D 语义分割模型 NMRNet,在原有挤压激励模块后添加残差精炼模块,引导网络更加关注与任务目标相关的特征信息,在 NYUv2 数据集上较不使用残差精炼模块 mIoU 仅提高

0.9%。ESANet 及现有的改进网络仅对 RGB 和深度两种模态单独处理再简单融合,未充分利用模态间互补信息提高特征的可信度。此外,有学者将融合模态扩展到深度以外的热红外、偏振等不同模态,辅助 RGB 模态提高语义分割精度^[16]。

本文针对细长特征的线缆传统 RGB 单模态语义分割效果差等问题,采用 RGB-D 双模态语义分割以互补特征加强特征表示,选用高效的 ESANet 作为基础网络,并对网络部分模块调整,线缆细长特征提高线缆细长特征下分割精度。

1 基于双模态融合的线缆图像语义分割方法

方案考虑线缆一般敷设在狭小昏暗环境中,颜色多样、形态各异,且线缆细长特征仅依靠 RGB 单模态信息难以实现准确语义分割,采用 RGB+深度双模态融合的语义分割方法,利用两种模态的互补信息进行特征校正与融合并参与后续特征提取准确的线缆轮廓掩膜,再提取掩膜的中心线作为线缆弯曲半径的测量特征曲线。图 1 为线缆轮廓掩膜分割方案,由深度相机采集线缆图像的 RGB 图像、深度图像,将所得图像输入改进 ESANet 网络中生成线缆目标掩膜,再对掩膜提取特征曲线,改进 ESANet 网络是整个算法核心,要实现线缆特征提取、模态特征融合与掩膜分割任务。

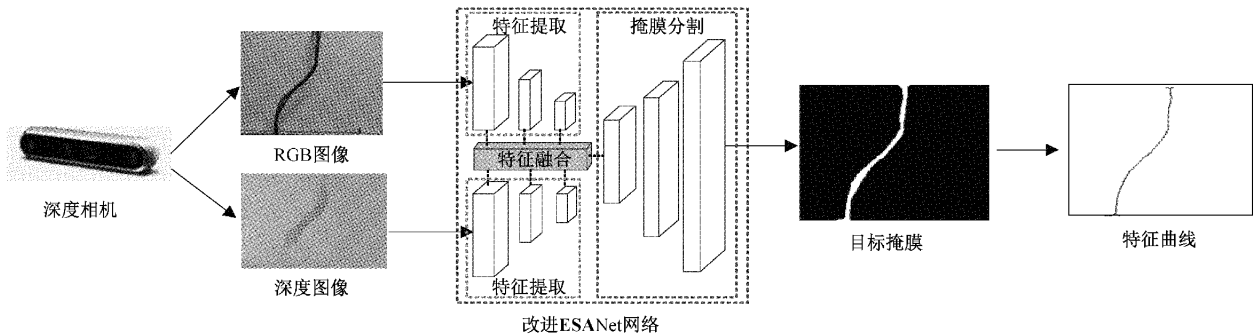


图1 线缆轮廓掩膜分割方案

1.1 改进 ESANet 网络模型

语义分割模型采用 Encoder-Decoder 模型框架,采用编码器 Encoder 提取特征,解码器 Decoder 再将提取到具有高级语义信息特征上采样到原图像尺寸得到掩膜,完成目标分割功能。一般 ESANet 网络模型整体思路是在 RGB 单模态语义分割网络 SwiftNet^[17]基础上添加深度模态,两种模态采用对称并行特征提取网络,在编码器提取特征过程中多阶段融合 RGB 模态、深度模态,再由解码器统一解码至原图像尺寸掩膜图像,重点在于 RGB 和深度两种模态之间的融合方法。ESANet 及当前几个改进网络大多只关注模态内的突出特征,在特征融合模块中为 RGB 模态、深度模态分别生成对应权重,两种模态特征图直接加权相加得到融合特征,未能利用两种模态间互补信息,且模态融合时受输入特征噪声影响大,若利用模态间特征互补信息对两种模态

的特征进行校正再融合将更有利于提升融合质量。

为此,本文提出图 2 所示的改进 ESANet 模型结构。RGB 模态、深度模态特征提取网络采用并行两个 Resnet50,用分隔聚合门 SAGate 模块代替 RGB-D Fusion 模块完成跨模态特征校正与融合功能,SAGate 输出同时作为 RGB、深度模态下一阶段特征提取输入(不同于一般仅输入 RGB 下一阶段特征提取中)。其中,RGB 模态、深度模态的 Resnet50 分别输入 RGB 三通道彩色图像、float16 编码的单通道深度图像;Resnet50 含 5 个特征提取层^[18],特征图尺寸逐步缩小,提取出高级语义特征;每个特征提取层后有 SAGate 模块对 RGB 模态、深度模态特征进行校正与融合,前 4 阶段融合特征与融合前 RGB 输入特征、深度特征分别对应下一阶段特征提取网络的输入特征,第 5 阶段融合特征经过一个上下文模块后作为解码器

Decoder 的输入; Decoder 由 3 个解码单元、2 个学习上采样单元 Learned up. $\times 2$ 组成, 每个解码单元将特征图的分辨率上采样 2 倍, 整个解码单元后接一个 3×3 卷积组合相邻特征, 最后两个 Learned up. $\times 2$ 单元将特征图语义信息

映射到输出掩膜图中。Decoder 与 Encoder 之间采用跳跃连接, 通过每个解码单元接收特征提取过程中对应维度融合特征, 使解码过程包含更多输入图像原始特征, 兼顾全局语义信息, 有助于提高分割准确性。

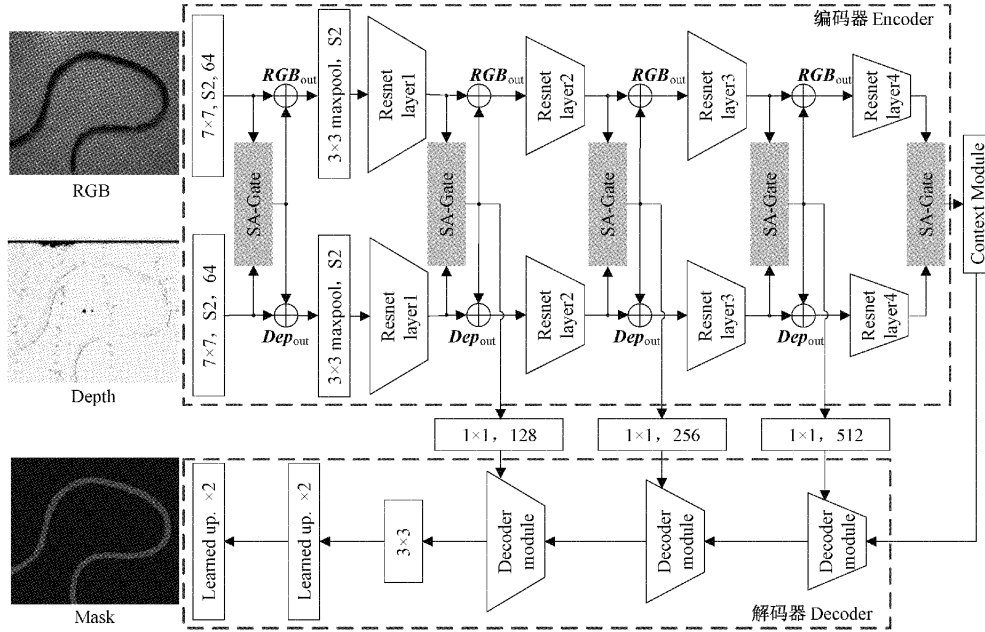


图 2 改进 ESANet 模型结构

1.2 应用 SAGate 跨模态特征校正与融合实现

同一对象不同模态关注特征不同, RGB 模态关注对象的颜色、纹理等特征, 容易受光照强度、色彩对比度影响; 深度模态关注对象位置关系、物理距离等特征, 容易受到深度相机成像原理限制而导致深度图中的对象边缘出现信息缺失情况。利用 RGB、深度两种模态的互补信息进行模态间相互校正与融合, 有利于提高特征提取准确度, 从

而提高光照复杂条件下目标分割精度。SAGate 模块完成特征校正、特征融合两方面任务, 图 3 为应用 SAGate 完成跨模态特征校正与融合原理框图, 首先利用两种模态特征生成跨模态特征校正注意力向量提取输入特征图的高置信度特征, 利用两种模态高置信度特征相互完成特征校正功能, 再利用融合校正后特征生成融合权重矩阵完成特征跨模态融合功能, 实现 RGB、深度双模态特征校正与融合。

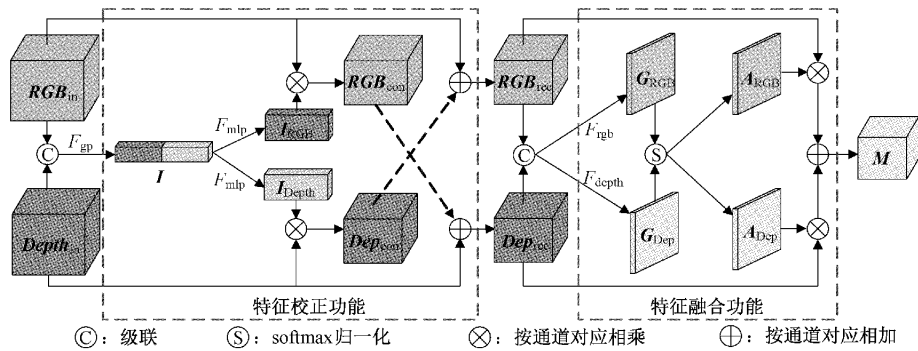


图 3 SAGate 特征校正与融合原理框图

下面对 SAGate 完成校准与融合功能的关键公式进行推导。

1) 特征校正功能实现

对输入的 RGB 特征 RGB_{in} 、深度特征 $Depth_{in}$ 级联 \parallel , 再全局平均池化 F_{gp} 得到跨模态注意力向量 I , 跨模态注意力向量 I 经过多层感知机网络 F_{mlp} , 归一化得到 RGB 模

态、深度模态特征注意力向量 I_{RGB} 、 I_{Depth} , 即:

$$I = F_{gp}(RGB_{in} \parallel Depth_{in}) \quad (1)$$

$$[I_{RGB}, I_{Depth}] = \text{Sigmoid}[F_{mlp}(I)] \quad (2)$$

然后, I_{RGB} 与 RGB_{in} 、 I_{Depth} 与 $Depth_{in}$ 分别按通道对应相乘 \otimes , 求得 RGB_{in} 、 $Depth_{in}$ 全局高置信度特征 RGB_{con} 、 $Depth_{con}$; RGB_{in} 与 $Depth_{con}$ 、 $Depth_{in}$ 与 RGB_{con} 逐通道按像

素相加 \oplus ,得到 RGB、深度模态校正后全局特征 RGB_{rec} 、 $Depth_{rec}$:

$$\begin{cases} RGB_{con} = I_{RGB} \otimes RGB_{in} \\ Depth_{con} = I_{Depth} \otimes Depth_{in} \end{cases} \quad (3)$$

$$\begin{cases} RGB_{rec} = RGB_{in} \oplus Depth_{con} \\ Depth_{rec} = Depth_{in} \oplus RGB_{con} \end{cases} \quad (4)$$

2) 特征融合功能实现

将校正后特征 RGB_{rec} 、 $Depth_{rec}$ 级联 \parallel , 通过卷积将级联特征展平成两个二维特征图 G_{RGB} 、 G_{Depth} , 即:

$$[G_{RGB}, G_{Depth}] = \text{Conv}(RGB_{rec} \parallel Depth_{rec}) \quad (5)$$

将 G_{RGB} 、 G_{Depth} 通过 softmax 函数逐像素归一化, 得二维权重矩阵 A_{RGB} 、 A_{Depth} , 它们在 (i, j) 处满足:

$$A_{RGB}(i, j) + A_{Depth}(i, j) = 1 \quad (6)$$

将 RGB_{rec} 、 $Depth_{rec}$ 逐通道加权相加, 得到跨模态融合特征 M , 并跳跃连接至对应解码单元参与解码过程, M 分别与 RGB_{in} 、 $Depth_{in}$ 逐通道逐像素平均, 得 RGB、深度特征提取网络阶段输出特征 RGB_{out} 、 $Depth_{out}$:

$$M = (A_{RGB} \otimes RGB_{rec}) \oplus (A_{Depth} \otimes Depth_{rec}) \quad (7)$$

$$RGB_{out} = \frac{1}{2}(RGB_{in} \oplus M); Depth_{out} = \frac{1}{2}(Depth_{in} \oplus M) \quad (8)$$

这完成 SAGate 校准与融合流程。

2 实验与分析

本文采用 realsense D455 相机采集 100 种不同颜色、规格、形态的线缆图像共 614 份, 其中训练集 492 份, 验证集 122 份, 每份图像包含 RGB 图像、相对应的深度图像, RGB 图像以三通道彩色图像形式保存, 深度图像以 float16 数据格式的单通道图像形式保存。

训练过程中, 令输入图像的总像素数量为 N , 真实标签值为 y_i , 预测结果概率分布 \hat{y}_i , 则交叉熵损失函数为:

$$Loss = - \sum_{i=0}^N y_i \log \hat{y}_i \quad (9)$$

在线缆分割任务中, 可采用交并比 (IoU) 评价每类标签的分割精度以及平均交并比 (mIoU) 评价整个模型分割性能, IoU 计算式为:

$$IoU = \frac{TP}{TP + FN + FP}; mIoU = \frac{1}{k} \sum_{i=1}^k IoU_i \quad (10)$$

式中: TP 为预测结果与标签均是目标的像素数量, FN 为标签是目标但预测结果不是目标的像素数量, FP 为标签不是目标但预测结果是目标的像素数量; k 为模型分割类别数, IoU_i 为第 i 个模型交并比。

使用 GTX 3090 配置迭代 500 次数, $Loss$ 已经到较小值, 求得较佳的改进 ESANet 网络模型权重值; 用训练得到权重得到网络模型, 对待测线缆做预测分割。图 4 为模型训练指标收敛情况, 曲线①、②分别为 $mIoU$ 、 $Loss$ 随迭代次数变化曲线。

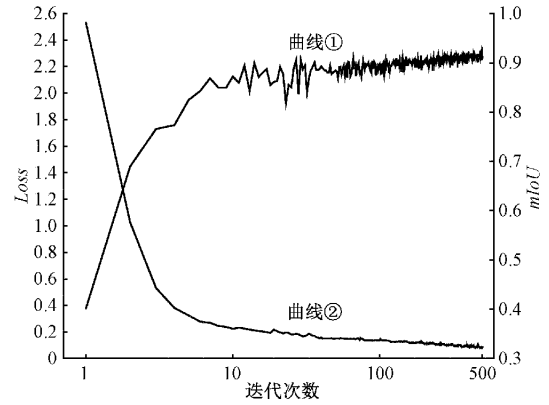


图 4 模型训练指标收敛情况

为了验证本文方法实验效果, 采用相同数据集、相同超参数设置, 用 RGB 单模态语义分割网络 SwiftNet、ESANet 网络与本文方法做对比实验, 表 1 为实验结果对比, 图 5 为待测线缆分割结果图。

表 1 3 种分割模型实验结果表

模型	FPS	mIoU/%
SwiftNet ^[17]	76.7	85.32
ESANet ^[13]	47.1	89.01
本文方法	45.9	93.00

可以看出, 本文改进 ESANet 模型线缆分割精度 $mIoU=93.00\%$ 、图像处理帧率 $FPS=45.9$, 较 ESANet 模型 FPS 仅下降 2.5%, $mIoU$ 提升 3.99%; 较 SwiftNet 网络 $mIoU$ 大幅度提升, 达 7.68%。本文改进 ESANet 模型分割效果比较明显。

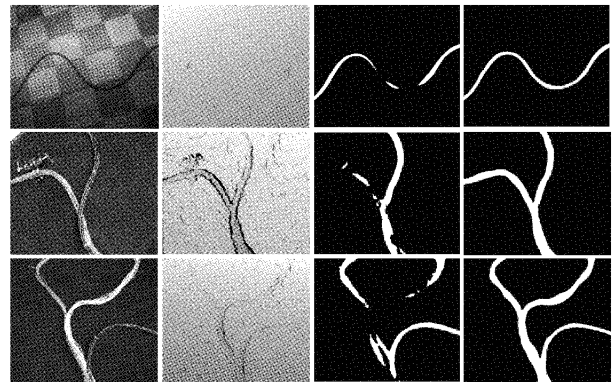


图 5 待测线缆分割结果图

3 结论

针对线缆形态多变、颜色各异等特点, 本文提出使用改进 ESANet 进行线缆掩膜分割方法, 采用更精细的跨模态融合模块 SAGate 代替原有的 RGB-D Fusion 模块, 利用融合特征同时参与 RGB 和深度两种模态后续的特征提取

过程;SAGate 提取 RGB、深度两种模态中高置信度特征,利用两种模态互补的高置信度特征相互校正与融合,使得提取到特征更具有语义信息;通过融合 RGB 和深度双模态特征及融合特征与解码器模块跳跃连接方式,进一步提高语义分割边界精度。实验表明,本文方法对线缆分割精度(mIoU)达到 93.00%,较 ESANet 方法 FPS 仅下降 2.5%的情况下 mIoU 提升了 3.99%,本文改进 ESANet 模型具有很好的分割效果。

参考文献

- [1] 国家市场监督管理总局,国家标准化管理委员会. 塑料绝缘控制电缆:GB/T 9330-2020[S]. 北京:中国标准出版社,2020.
- [2] IEC. IEC 62440-2008 Electric cables with a rated voltage not exceeding 450/750 V-Guide to use[S]. Geneva, 2008.
- [3] 罗权,易兵,王杰. 带工艺约束柔性体模拟的轨道车辆布线方法[J]. 铁道科学与工程学报,2022,19(10): 3070-3078.
- [4] 李炜,韩永进. 飞机现役线缆状态及寿命评估研究[J]. 电线电缆,2020(6):5-7.
- [5] 王发麟,廖文和,郭宇,等. 线缆虚拟装配关键技术研究现状及其发展[J]. 中国机械工程,2016,27(6): 839-851.
- [6] BAKER L, MILLS S, LANGLOTZ T, et al. Power line detection using hough transform and line tracing techniques[C]. IEEE, 2016, DOI:10.1109/IVCNZ. 2016. 7804438.
- [7] 汪成龙,陈国壮,马秋立,等. 基于 OpenCV 的电线颜色识别系统软件[J]. 电子测量技术. 2018,41(16): 91-97.
- [8] DE G D, PALLI G, DI S L. Let's take a walk on superpixels graphs: Deformable linear objects segmentation and model estimation[C]. 2018, DOI: 10.1007/978-3-030-20890-5_42.
- [9] BO L A, CHENG C B, SD C, et al. Transmission line detection in aerial images: An instance segmentation approach based on multitask neural networks[J]. Signal Processing: Image Communication, 2021, DOI: 10.1016/j.image.2021.116278.
- [10] XING Y, WANG J, ZENG G. Malleable 2.5 d convolution: Learning receptive fields along the depth-axis for RGB-D scene parsing [C]. European Conference on Computer Vision. Springer, Cham, 2020: 555-571.
- [11] HU X X, YANG K L, FEI L, et al. ACNet: Attention based network to exploit complementary features for RGBD semantic segmentation[C]. 2019 IEEE International Conference on Image Processing (ICIP), 2019.
- [12] CHEN X, LIN K Y, WANG J, et al. Bi-directional cross-modality feature propagation with separation-and-aggregation gate for RGB-D semantic segmentation[C]. European Conference on Computer Vision. Springer, Cham, 2020: 561-577.
- [13] SEICHTER D, KÖHLER M, LEWANDOWSKI B, et al. Efficient RGB-D semantic segmentation for indoor scene analysis [C]. 2021 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2021: 13525-13531.
- [14] 罗益琳,方艳红,李鑫,等. RGB-D 双模态特征融合语义分割[J]. 计算机工程与应用,2023,59(7): 222-231.
- [15] 张健,陈焯恒,朱世强,等. 一种基于邻域度量关系的 RGB-D 融合语义分割算法[J]. 机器人,2023,45(2): 156-165.
- [16] LIU H, ZHANG J, YANG K, et al. CMX: Cross-modal fusion for RGB-X semantic segmentation with transformers[J]. ArXiv Preprint, 2022, ArXiv:2203.04838.
- [17] ORSIC M, KRESO I, BEVANDIC P, et al. In defense of pre-trained imagenet architectures for real-time semantic segmentation of road-driving images [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 12607-12616.
- [18] 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.

作者简介

曹国群,硕士研究生,主要研究方向为现代传感技术与机器视觉。

E-mail: gq_cao@foxmail.com

刘桂雄,教授,主要研究方向为先进传感与仪器研究。

E-mail: megxliu@scut.edu.cn