

引用格式: 马文琪, 石颀, 吴宏杰. 深度卷积神经网络语义分割综述 [J]. 微电子学与计算机, 2023, 40(9): 55-64. [MA W Q, SHI J, WU H J. Survey on semantic segmentation using deep convolutional neural networks[J]. Microelectronics & Computer, 2023, 40(9): 55-64.] DOI: 10.19304/J.ISSN1000-7180.2022.0825

深度卷积神经网络语义分割综述

马文琪¹, 石颀¹, 吴宏杰²

(1 苏州科技大学 电子与信息工程学院, 江苏 苏州 215009;

2 江苏省建筑智慧节能重点实验室, 江苏 苏州 215009)

摘要: 得益于深度卷积神经网络在特征提取和语义理解的强大能力, 基于深度神经网络的语义分割技术逐渐成为计算机视觉研究的热点课题。在无人驾驶、医学图像, 甚至是虚拟交互、增强现实等领域都需要精确高效的语义分割技术。语义分割从图像像素级理解出发, 为每个像素分配单独的类别标签。针对基于深度神经网络的语义分割技术, 根据技术特性的差异, 从编码-解码架构、多尺度目标融合、卷积优化、注意力机制、传统-深度结合、策略融合方面展开, 对现有模型的优缺点进行梳理和分析, 并当前主流语义分割方法在公共数据集实验结果进行对比, 总结了该领域当前面临的挑战以及对未来研究方向的展望。

关键词: 深度学习; 语义分割; 卷积神经网络; 计算机视觉

中图分类号: TP391.41

文献标识码: A

文章编号: 1000-7180(2023)09-0055-10

Survey on semantic segmentation using deep convolutional neural networks

MA Wenqi¹, SHI Jie¹, WU Hongjie²

(1 Suzhou University of Science and Technology, Suzhou 215009, China;

2 Jiangsu Provincial Key Laboratory of Building Intelligent Energy Conservation,
Suzhou 215009, China)

Abstract: Benefiting from the powerful ability of deep convolutional neural network in feature extraction and semantic understanding, semantic segmentation technology based on deep neural network has gradually become a hot topic in computer vision research. Accurate and efficient semantic segmentation techniques are needed in the fields of unmanned driving, medical images, virtual interaction, augmented reality and so on. Semantic segmentation starts from pixel-level understanding of the image and assigns a separate category label to each pixel. Aiming at the semantic segmentation technology based on deep neural network, according to the differences in technical characteristics, the advantages and disadvantages of existing models are sorted out and analyzed from the aspects of encoder-decoder architecture, multi-scale target fusion, convolution optimization, attention mechanism, traditional-deep combination, and strategy fusion. The current mainstream semantic segmentation methods are compared in the experimental results of public datasets. Finally, the current challenges and future research directions in this field were summarized.

Key words: Deep learning; Semantic segmentation; Convolutional neural network; Computer vision

1 引言

图像语义分割技术是指根据图像的灰度、色彩、纹理等特征,为图像中的每个像素分配单独的类别标签,使图像被分割成若干视觉意义上的特定的、具有独特性质的区域^[1].随着智能化生活的推进,语义分割技术在无人驾驶、医学图像处理、视频监控,甚至是虚拟交互、增强现实等领域都发挥着日益重要的作用.

传统语义分割方法包括基于阈值^[1-3]、候选区域^[4-6]、边缘检测^[7-9]、聚类^[10-12]、图论^[13-15]等利用数学理论,以及使用纹理基元森林^[16]或者随机森林^[17]等机器学习方法来构建像素分类器.但由于硬件设备的计算能力有限,传统方法在分割图像的颜色、形状、纹理等低级语义特征时效率较低、分割时间较长,并且难以识别相互遮挡的物体,导致分割精度较低.

随着硬件性能的升级和深度学习的兴起,深度卷积神经网络(Deep Convolutional Neural Networks, DCNN)^[18]强大的分析与处理能力日益突出.2015年Long等人提出全卷积网络(Fully Convolutional Networks, FCN)^[19],将DCNN应用于语义分割,从此语义分割正式进入以DCNN为主的发展时代,能够以端到端的训练方式在图像语义层面提取并学习图像特征,使网络主动推理每个像素的语义信息并分类,获得更高的分割精度和计算效率.

现有的语义分割相关的综述文献^[20-22],对整个语义分割领域进行了全面的概述性介绍,涵盖了早期几乎所有流行的语义分割方法和多维的数据模式;文献^[23-24],针对语义分割模型训练数据集质量参差不齐的问题,从全监督和弱监督的训练方式进行分析.为了更好的分析当前深度神经网络语义分割的发展情况,本文基于技术特性的不同对该领域的发展情况进行补充和丰富,并分析了每类技术的优缺点,并在常用数据集上对比各模型的实验结果.最后讨论了该领域当前及未来发展亟待解决的问题.

2 深度卷积神经网络的语义分割技术

深度卷积神经网络语义分割是指利用DCNN学习预先标注好的图像数据集,并利用学习到的语义特征进行像素级的分类.常用的DCNN包括首次尝试分组卷积的AlexNet^[25]、利用小卷积核替代大卷积核的VGGNet^[26]、使用多个不同尺寸卷积核的

GoogleNet^[27]、利于更深层网络训练的残差网络ResNet^[28]、卷积同时考虑图像特征区域和通道的Xception^[29].

由于DCNN中最大池化和下采样操作使得图像尺寸随网络加深而减小,导致细粒度信息的丢失.为此研究者们针对以上问题提出了不同的解决方法.本文依据技术特性将其分为:编码-解码架构、多尺度目标融合、卷积优化、注意力机制、传统-深度结合、策略融合,分析并总结了每类方法的优缺点,具体如表1所示.

2.1 编码-解码架构

为了解决池化操作造成的特征细节丢失的问题,研究者们引入由编码器和解码器组成编码-解码架构.编码器利用多个卷积和池化操作,提取浅层的位置特征和深层的语义特征;解码器利用上采样操作恢复图像的细节和空间信息.

2015年Noh等人^[30]提出对称的编码-解码架构的语义分割模型DeconvNet学习图像的语义特征,但该模型中参数过多的全连接层会增加模型训练难度.为此,Badinara等人^[31]提出架构相似且去除全连接层的SegNet网络,在获得更高精度的同时节省训练内存.

为了解决深层级像素的位置信息易丢失的问题,提高分类边界置信度,Ronneberger等人^[32]提出架构对称的网络U-Net,利用跳层连接将对应层级的特征相连,充分融合浅层细节与深层语义.但U-Net的卷积过程中没有padding操作,使输入与输出大小不同,且只能处理2D图像.为此Milletari等人^[33]在适当使用padding操作,提出可用于3D图像分割的网络V-Net. U-Net++^[34]和U-Net3+^[35]同样优化U-Net,可以更充分的捕获全尺度细粒度语义和粗粒度语义的特征.

由于对称结构参数较多,实时性较差的问题,Arani等人^[36]设计了非对称架构的实时语义分割网络RGNet,简化解码器规模,并结合适配器保存和细化多级信息,兼顾实时性和精度.Wang等人^[37]提出编码器较大、解码器较小的轻量级模型LEDNet,编码器使用ResNet提取特征,利用新的残差模块,减小网络规模;解码器中引入注意力金字塔提取更加丰富的特征,降低网络的复杂度.Romera等人^[38]提出架构更简化的ERFNet,编码器交错使用空洞卷积获取更多的语义信息;解码器包含上采样操作和大量的深度可分离卷积,极大的减少网络的参数.

表 1 图像语义分割代表方法分析与总结

Tab. 1 Analysis and Summary of Image Semantic Segmentation methods

技术特性	代表算法	时间	主要特征	优点	缺点
编码-解码架构	DeconvNet ^[30]	2015	对称架构; 最大池化激活值; 转置卷积层稠密特征		
	SegNet ^[31]	2016	对称架构; 去除全连接层		
	U-Net ^[32]	2015	对称架构; 跳跃连接	避免深度神经网络因池化造成的特征细节丢失的情况;	目标边界信息提取不清晰; 计算参数量较多;
	LEDNet ^[37]	2019	非对称架构; 通道分离和混洗操作	解码网络能有效恢复目标的特征维度和空间信息	分割精度欠佳, 不满足实际工程
	RGPNet ^[36]	2022	轻量级解码器; 适配器融合编码器和解码器信息		
	ERFNet ^[38]	2018	非对称架构; 残差链接		
	PSPNet ^[40]	2017	特征金字塔结构, 多尺度特征融合		
多尺度目标融合	ForkNet ^[41]	2022	孪生特征金字塔网络	充分利用图像的多尺度特征, 改善特征提取的粗糙结果;	部分边界信息丢失; 需要辅助模块做处理
	文献[42]	2021	纹理增强模块; 金字塔纹理特征提取模块	融合全局上下文语义特征	
	RefineNet ^[43]	2016	多路径细化不同尺度特征		
	BiSeNet ^[44]	2018	细节分支和语义分支 两路分支结构		
卷积优化	DeepLabv1 ^[46]	2016	空洞卷积; 完全连接的条件随机场作为后处理		
	DeepLabv2 ^[47]	2017	不同速率的空洞卷积; 空洞空间金字塔池化		
	DeepLabv3 ^[48]	2017	可级联和并行使用的空洞卷积模块	扩大特征提取感受野; 保护特征位置语义信息	性能较普通卷积欠佳; 局部信息连续性略差
	DeepLabv3+ ^[49]	2018	深度可分离卷积模块		
	文献[50]	2018	分解卷积; 平滑空洞卷积		
	文献[51]	2018	密集上采样卷积; 混合扩展卷积		
注意力机制	PSANet ^[53]	2018	自适应注意力掩码机制		
	DANet ^[54]	2019	双向注意力机制	突出主体目标特征, 使模型有效提取上下文重要特征	不易捕获特征位置信息; 计算量较大
	文献[55]	2020	交叉路径注意力模块		
	文献[56]	2022	注意力聚合模块		
	OCNet ^[57]	2021	自注意力机制		
传统-深度结合	文献[58]	2015	结合马尔科夫随机场		
	CRFasRNN ^[60]	2016	结合条件随机场		
	Graph-FCN ^[61]	2020	结合图网络模型	充分利用全局上下文特征语义信息和位置信息	计算量大, 内存消耗大, 实时性差
	文献[62]	2021	结合区域感知学习		
	CEU-Net ^[63]	2022	结合聚类算法		
	文献[64]	2022	结合边缘检测		
多策略融合	文献[65]	2020	编码-解码架构; 可变注意力金字塔		
	文献[66]	2022	编码-解码架构; 空间金字塔; 通道注意力; 边缘检测	捕获上下文有利特征, 减少计算量和内存消耗	主要目标部分边界信息丢失
	ReSTR ^[68]	2022	自注意力机制; 多模态融合编码器		
	PCANet ^[69]	2020	金字塔卷积注意力网络		

2.2 多尺度目标融合

深层语义特征预测效果好,但分辨率低、细节丢失严重;浅层语义特征分辨率高、细节丰富,但语义类别预测效果差.因此许多研究者利用不同层的信息差异性,融合多尺度信息补充分割细节.

最常见的多尺度目标融合方式是特征金字塔结构(Feature Pyramid Network,FPN)^[39],通过多次采样使不同大小的目标在不同尺度有相应的特征表示,再融合不同尺度信息预测不同大小的目标.Zhao 等人^[40]提出金字塔场景解析网络 PSPNet,能够融合四个不同尺度的特征信息,在捕获上下文语义信息的同时保持全局特征的权重,并能够在不增加计算量的情况下与局部 FCN 特征同时训练.He 等人^[41]提出孪生特征金字塔网络 ForkNet,将 FPN 应用于特征提取阶段,在每个层次生成具有强大语义的特征表示.Zhu 等人^[42]提出特征金字塔纹理特征提取模块,丰富特征的纹理细节,并且利用量化和技术算子,捕捉和描述低层级纹理信息,将连续纹理量化为多级强度.

另一种多尺度目标融合是多路分支结构,即在不同的支路中处理感受野各异的特征图,再互相融合补充出空间轮廓信息.2017 年 Lin 等人^[43]等人提出两通路实时语义网络 RenfineNet,利用上通路和下通路聚合多级别特征,形成利于梯度传播和模型训练的长、短射程,解决了浅层特征利用率较低的问题,强化了高级语义特征.Yu 等人^[44]为了平衡模型的分割精度和推理速度,提出两分支架构网络双边分割网络 BiSeNet,同时解决特征提取的空间信息缺失和感受野缩小的问题.

2.3 优化卷积操作

卷积操作利用像素值的加权平均代替,提取出抽象化、符号化的语义特征信息.但由于卷积步长的存在,图像的特征信息会随着网络加深而丢失.因此许多研究者通过优化卷积操作,提高语义分割的性能.

空洞卷积^[45](Atrous Convolution)在标准卷积中注入空洞,以此扩大卷积核的感受野,使得每个卷积核的输出都包含较大范围的特征信息,并且保持输出时特征映射的尺寸.Chen 等人提出的 Deeplab 系列的 v1^[46]和 v2^[47],利用空洞卷积增加特征提取的感受野,缓解因重复池化和卷积造成图像的位置信息无法修复的情况.随后在 v3^[48]中优化空洞卷积,使之够级联或并行使用,利用不同膨胀率空洞卷积在串行模块中获取更大的感受野^[49].

由于堆叠的空洞卷积会受网格会产生伪影效应,

即卷积后的特征图像出现不存在于输入图像的伪特征,文献^[50]提出分解卷积,使用交互层平滑空洞卷积,或者在空洞卷积之前插入可分离和共享的卷积平滑空洞卷积,解决了网格伪影效应;Wang 等人^[51]则设计混合卷积,利用密集上采样卷积取代解码器的双线性上采样,捕获更多的细节;利用混合扩展卷积,减轻编码器的空洞卷积的伪影效应,扩大感受野聚合全局信息.

受 Inception 模块^[27]的启发,Chollet 等人^[29]利用深度可分离卷积(Depthwise Separable Convolution, DSP)训练 DCNN. DSP 包含逐通道卷积和逐点卷积,同时捕获空间信息和通道信息. Deeplabv3+^[49]中广泛使用 DSP 提高模型的边界特征表达能力,并减少计算量;Cao 等人^[52]则使用 DSP 作为基础组件应用于轻量级的语义分割网络,以实现盲道和人行横道图像的高效语义识别,帮助盲人感知周围环境.

2.4 注意力机制

注意力机制能够增强分割的主体信息,抑制不相关信息,从而提高网络分割精度. DeepLabv1^[46]针对全卷积语义分割网络感知区域小的问题,将多尺度注意力模块与多尺度结构进行联合训练,使得网络能够更好的学习多尺度特征在图像像素位置的软权重.

针对卷积核的物理结构容易导致神经网络信息流被约束在一定的范围内,从而造成复杂场景难以理解的问题,Zhao 等人^[53]提出 PSANet 语义分割网络,使用点状空间注意力(Point-wise Spatial Attention)来解决局部区域限制问题,使得每个像素都可以与其他位置的像素建立联系,从而丰富上下文信息.为了能够充分捕捉图像的上下文语义特征,增加特征空间语义信息和特征维度语义信息的相关性,Fu 等人^[54]提出双注意力网络 DANet,结合位置注意力模块和通道注意力模块在局部特征上建模丰富的上下文关系,进一步改进特征表示,从而获取更精确的分割效果.

此外还有 Huang 等人^[55]利用交叉注意力模块,探索交叉路经周围像素的上下文信息;Li 等人^[56]提出了注意力聚合模块,通过注意力引到的特征聚合来增强网络对语义特征的学习;Yuan 等人^[57]受自注意力机制影响设计了 OCNet,通过提高特征图的分辨率来提高特征空间准确性,并且聚焦于物体上下文估计和物体上下文聚合,来优化最终的预测结果.

2.5 传统-深度结合

传统语义分割方法,计算简单、对均匀的联通目标有较好的分割结果,但对于多个物体需要多次运行. DCNN 能够端到端训练模型主动提取图像各层

级的特征,却缺乏上下文建模的长范围依赖关系.将传统方法与 DCNN 网络结合也是语义分割的热门研究方向.

文献 [58] 将概率图模型马尔可夫随机场(Markov Random Field,MRF)^[59]与 DCNN 结合设计了深度解析网络,利用 MRF 获取图像中的上下文语义信息并进行关系建模,以表示更多的成对项,因此该模型能在减少网络参数的同时,提高分割效果.条件随机场(Conditional Random Field,CRF)是一种特殊的 MRF,DeepLabv1^[46]和 DeepLabv2^[47]中利用 CRF 作为后处理操作,促进模型捕获低层的细粒度信息,避免分割网络由于图像的空间不变性造成细节的丢失,Zheng 等人^[60]在模型 CRFasRNN 中将 CRF 的推理过程设计为循环神经网络的结构,并将该结构融入到 DCNN 中,使得网络能够同时优化各个阶段的参数,并且降低预处理和后处理的时间. Lu 等人^[61]设计模型 Graph-FCN,同样利用图网络模型与 DCNN 结合,在网络中间的特征层构造一个图网络模型,可以更好的保存细节信息.

为了考虑整个数据集的语义关系,Hu 等人^[62]提出区域感知对比学习,通过构建区域中心来表示每幅图像的不同类别的特征,使得模型在区域级别处理所有的输入数据,节省了更多的计算内存. Soucy 等人^[63]提出结合聚类方法的 CEU-Net 模型,检测图像中存在的重叠和分散的类别组成; Ghandorh 等人^[64]结合边缘检测技术,使用分割后的掩膜生成的图像边缘,获得更清晰的像素分割图,提高整体语义分割分割精度.

2.6 多策略融合

多策略融合则是将编码-解码架构、多层特征融合、优化卷积操作、注意力机制、传统-深度结合技术综合使用,能极大程度提高语义分割的性能.

Wu 等人^[65]针对编码-解码架构在固定集合过渡效果较差的问题,提出动态注意力网络,将可变形注意力模块和融合注意力模块应用到编码-解码架构当中,使得长、短期信息更有效地反向传播,提升网络捕获上下文信息的能力.

Zhang 等人^[66]遵循编码-解码架构设计了混合深度卷积神经网络,利用 Swin Transformer^[67]编码长程空间依赖关系,利用空洞空间金字塔池化捕获多尺度信息,利用通道注意力模块增强特征,利用边缘检测分支约束边缘. Kim 等人^[68]同样利用 Transformer 设计模型 ReSTR,结合注意力机制和编码器融合各种模

态的特征,获取模态中实体之间的长程依赖关系,能够灵活且自适应的交互语义信息.

为了有效地捕获长程依赖关系,融合不同层次的特征,Sang 等人^[69]提出金字塔卷积注意力网络 PCANet,利用金字塔空洞注意力模块捕获长程依赖关系,利用卷积注意力细化模块,为低层级的特征提供全局上下文,为高层级特征提供局部细节.

3 算法实验结果对比

编码-解码架构虽然模型能减少特征细节丢失,能有效恢复图像的特征维度和空间信息,但仍然存在边界特征的提取效果欠佳,以及计算量巨大、训练周期长的问题;多尺度目标融合技术可以融合全局上下文信息,但仍需辅助模块;卷积优化技术能扩大特征感受野并保护位置信息,但对局部信息不敏感,缺乏连续性;注意力机制技术能突出分割主体特征,但易忽略特征的位置信息;传统-深度结合技术能充分考虑全局特征的语义和位置信息,但巨大的计算量和内存消耗仍影响算法的实时性.

为了衡量语义分割算法的性能,可以从硬件方面的计算效率和内存占用比例进行分析,也可以通过数学理论分析的方法给出不同模型的计算量和参数量,再进行比较.虽然深度卷积神经网络极大的推进着语义分割技术的发展,但由于不同的语义分割网络在优化方向、应用场景和技术重心的不同,因此无法单一从硬件方面的使用效率来衡量模型的整体性能.但可以根据定量精度指标平均交并比(Mean Intersection over Union,MIOU)横向对比不同技术特性的语义分割模型的性能^[20],如公式(1),其中 $k+1$ 为类别数, N_{ii} 表示真阳性(True Positives), N_{ij} 表示假阳性(False Positives), N_{ji} 表示假阴性(False Negatives), N_{jj} 表示真阴性(True Negatives).

$$MIOU = \frac{1}{k+1} \sum_{i=0}^k \frac{N_{ii}}{\sum_{j=0}^k N_{ij} + \sum_{j=0}^k N_{ji} - N_{ii}} \quad (1)$$

为此,综合调研不同论文报告的实验结果,根据定量精度指标 MIOU 横向对比不同技术的语义分割模型的性能.将前文所述不同技术特性的语义分割方法在常用的一般场景数据集 PASCAL VOC^[70]、PASCAL Context^[71]和 ADE20K^[72],以及城市场景数据集 Cityspaces^[73]和 CamVid^[74]上的实验结果作比较,如表 2 所示,其中“—”则表示未有相关论文报告以该算法在对应数据集上进行实验.

表 2 语义分割算法性能比较
Tab. 2 Performance Comparison of Semantic segmentation algorithms

技术特性	代表算法	时间	基础网络	分割精度MIOU%				
				PASCAL VOC	PASCAL Context	Cityspaces	CamVid	ADE20K
编码-解码架构	DeconvNet ^[30]	2015	VGG16	69.6 ^[30]	—	—	59.77 ^[31]	—
	SegNet ^[31]	2016	VGG16	—	—	57.0 ^[31]	60.10 ^[31]	21.6 ^[43]
	LEDNet ^[37]	2019	ResNet	—	—	69.2 ^[37]	—	—
	RGPNet ^[36]	2022	ResNet101	—	—	81.9 ^[36]	69.2 ^[36]	—
	ERFNet ^[38]	2018	—	—	—	68.0 ^[38]	—	—
多尺度目标融合	PSPNet ^[40]	2017	ResNet101	82.6 ^[40]	—	78.4 ^[40]	69.1 ^[55]	—
	文献[42]	2021	ResNet101	—	55.8 ^[42]	82.3 ^[42]	—	46.48 ^[42]
	RefineNet ^[43]	2016	ResNet101	82.4 ^[43]	47.1 ^[43]	73.6 ^[38]	—	40.20 ^[43]
	BiSeNet ^[44]	2018	ResNet18	—	—	77.7 ^[44]	68.7 ^[44]	—
	DeepLabv1 ^[46]	2016	VGG16	71.6 ^[46]	39.6 ^[47]	63.1 ^[47]	54.7 ^[31]	—
优化卷积操作	DeepLabv2 ^[47]	2017	ResNet101	79.7 ^[47]	45.7 ^[47]	70.4 ^[47]	—	—
	DeepLabv3 ^[48]	2017	ResNet101	85.7 ^[48]	—	81.3 ^[48]	—	—
	DeepLabv3+ ^[49]	2018	Xception	87.8 ^[49]	—	82.1 ^[49]	—	45.5 ^[62]
	文献[50]	2018	—	71.6 ^[50]	—	69.2 ^[50]	—	—
	文献[51]	2018	ResNet101	83.1 ^[51]	—	80.1 ^[51]	—	—
注意力机制	PSANet ^[53]	2018	ResNet101	85.7 ^[53]	—	80.1 ^[53]	—	43.77 ^[53]
	DANet ^[54]	2019	ResNet101	82.6 ^[54]	52.6 ^[54]	81.5 ^[54]	—	45.22 ^[57]
	文献[55]	2020	ResNet101	—	—	80.5 ^[55]	79.1 ^[55]	45.76 ^[55]
	文献[56]	2022	ResNet34	—	—	—	—	—
	OCNet ^[57]	2021	ResNet101	—	54.0 ^[57]	81.9 ^[57]	—	45.40 ^[57]
传统-深度结合	文献[58]	2015	VGG16	74.1 ^[58]	—	66.8 ^[40]	—	—
	CRFasRNN ^[60]	2016	—	72.0 ^[60]	39.28 ^[60]	62.5 ^[40]	—	—
	Graph-FCN ^[61]	2020	—	65.91 ^[61]	—	—	—	—
	文献[62]	2021	ResNet101	—	—	82.3 ^[62]	—	—
策略融合	ReSTR ^[68]	2022	—	—	—	—	—	—
	PCANet ^[69]	2020	ResNet101	84.2 ^[69]	—	80.1 ^[69]	—	—

4 结束语

本文主要针对当前深度卷积神经网络语义分割技术的发展情况,从技术差异的角度归纳分类并进行了较为详细的分析和总结.随着深度卷积神经网络的不断推进,在模型分割精确度、运行速度和复杂性方面仍需大量的研究工作.结合当前语义分割研究现状,提出下述值得挑战和深入研究的方向:

(1)如何更好地理解语义分割模型?

尽管深度语义分割模型在特征提取方面体现了较好的性能,但深度模型到底在学习什么?研究者如何解释学习到的特征?虽然存在相关技术能够可视化模型提取的特征,但尚未有相关研究针对语义分割模型底层特征学习方式行表述.对模型理论方面的理解,可以更好地开发和改进泛化性强的语义分割模型.

(2)如何获得高质量的语义分割数据集?

语义分割模型性能与数据集质量成正相关.现有算法大都依赖人工标记的数据集,费时费力,且数据集规模较小.虽然当前很多研究者提出基于半监督和弱监督的分割方法,以弱注释方式进行深度学习,但准确性较全监督学习方法相差甚远.因此获得高质量的语义分割数据集仍是值得深入研究的方向.

(3)如何平衡语义分割的精确度和速度?

实时语义分割在机器人交互、自动驾驶等领域发挥着重要作用.而基于深度学习的语义分割更注重于分割精确度而不是速度.当前主流的实时语义分割方法仍然缺乏较高的精度,需要更好的方法和途径平衡精确度和速度.

(4)如何充分利用上下文特征信息提高模型准确性?

有研究者利用 CRF 作为后处理,提出端到端的方法提高语义分割的准确性.也有研究者通过多尺度特征融合在准确性方面获得显著性的进展.图像包含丰富的上下文特征信息,如何更好的利用这些信息,仍需要更多的研究.

(5)如何更好的轻量化模型?

当前主流分割模型在推理阶段仍需大量内存,许多研究者为了在提高精度的同时应用于特定设备上,利用模型压缩技术,或者训练一个复杂模型再用知识整理技术压缩,以减小体积更适用于移动设备.因此在保持精度的同时能轻量化模型体积,将会是未来基于深度卷积语义分割模型的发展方向.

参考文献:

- [1] OTSU N. A threshold selection method from gray-level histograms[J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1979, 9(1): 62-66. DOI: 10.1109/TS-MC.1979.4310076.
- [2] WESZKA J S. A survey of threshold selection techniques[J]. *Computer Graphics and Image Processing*, 1978, 7(2): 259-265. DOI: 10.1016/0146-664X(78)90116-8.
- [3] BATENBURG K J, SIJBERS J. Optimal threshold selection for tomogram segmentation by projection distance minimization[J]. *IEEE Transactions on Medical Imaging*, 2008, 28(5): 676-686. DOI: 10.1109/TMI.2008.2010437.
- [4] ADAMS R, BISCHOF L. Seeded region growing [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1994, 16(6): 641-647. DOI: 10.1109/34.295913.
- [5] ANGULO J, JEULIN D. Stochastic watershed segmentation[C]//8th ISMM 2007. Rio de Janeiro: ISMM, 2007: 265-276.
- [6] HARIHARAN B, ARBELÁEZ P, GIRSHICK R, et al. Simultaneous detection and segmentation[C]//Proceedings of the 13th European Conference on Computer Vision. Zurich: Springer, 2014: 297-312. DOI: 10.1007/978-3-319-10584-0_20.
- [7] SHRIVAKSHAN G T, CHANDRASEKAR C. A comparison of various edge detection techniques used in image processing[J]. *International Journal of Computer Science Issues*, 2012, 9(5): 269-276.
- [8] IRVIN S. An isotropic 3×3 image gradient operator[J]. *Machine Vision for Three-dimensional Sciences*, 1990: 1-5.
- [9] PREWITT J M S, MENDELSON M L. The analysis of cell images[J]. *Annals of the New York Academy of Sciences*, 1966, 128(3): 1035-1053. DOI: 10.1111/j.1749-6632.1965.tb11715.x.
- [10] FRIEDMAN N, GEIGER D, GOLDSZMIDT M. Bayesian network classifiers[J]. *Machine Learning*, 1997, 29(2): 131-163. DOI: 10.1023/A:1007465528199.
- [11] KOLLER D, FRIEDMAN N. Probabilistic graphical models: principles and techniques[M]. Cambridge: The MIT Press, 2009: 1208.
- [12] FUKUNAGA K, HOSTETLER L. The estimation of the gradient of a density function with applications in pattern recognition[J]. *IEEE Transactions on Information Theory*, 1975, 21(1): 32-40. DOI: 10.1109/TIT.1975.1055330.
- [13] BOYKOV Y, JOLLY M P. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images[C]//Proceedings Eighth IEEE International Conference on Computer Vision. Vancouver: IEEE, 2001: 105-112. DOI: 10.1109/ICCV.2001.937505.
- [14] ROTHERC, KOLMOGOROVV, BLAKEA. "GrabCut": interactive foreground extraction using iterated graph cuts[J]. *ACM Transactions on Graphics*, 2004, 23(3): 309-314. DOI: 10.1145/1015706.1015720.
- [15] GREIG D M, PORTEOUS B T, SEHEULT A H. Exact maximum *a posteriori* estimation for binary images[J]. *Journal of the Royal Statistical Society: Series B*, 1989, 51(2): 271-279. DOI: 10.1111/j.2517-6161.1989.tb01764.x.
- [16] SHOTTON J, JOHNSON M, CIPOLLAR R. Semantic texton forests for image categorization and segmentation[C]//Proceedings of 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA: IEEE, 2008: 1-8. DOI: 10.1109/CVPR.2008.4587503.
- [17] MAHAPATRA D. Analyzing training information from

- random forests for improved image segmentation[J]. *IEEE Transactions on Image Processing*, 2014, 23(4): 1504-1512. DOI: [10.1109/TIP.2014.2305073](https://doi.org/10.1109/TIP.2014.2305073).
- [18] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. *计算机学报*, 2017, 40(6): 1229-1251. DOI: [10.11897/SP.J.1016.2017.01229](https://doi.org/10.11897/SP.J.1016.2017.01229).
ZHOU F Y, JIN L P, DONG J. Review of convolutional neural network[J]. *Chinese Journal of Computers*, 2017, 40(6): 1229-1251. DOI: [10.11897/SP.J.1016.2017.01229](https://doi.org/10.11897/SP.J.1016.2017.01229).
- [19] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 3431-3440. DOI: [10.1109/CVPR.2015.7298965](https://doi.org/10.1109/CVPR.2015.7298965).
- [20] GARCIA-GARCIA A, ORTS-ESCOLANO S, OPREA S, et al. A review on deep learning techniques applied to semantic segmentation[J]. arXiv: 1704.06857, 2017. DOI: [10.48550/arXiv.1704.06857](https://doi.org/10.48550/arXiv.1704.06857).
- [21] YU H S, YANG Z G, TAN L, et al. Methods and datasets on semantic segmentation: a review[J]. *Neurocomputing*, 2018, 304: 82-103. DOI: [10.1016/j.neucom.2018.03.037](https://doi.org/10.1016/j.neucom.2018.03.037).
- [22] LATEEF F, RUICHEK Y. Survey on semantic segmentation using deep learning techniques[J]. *Neurocomputing*, 2019, 338: 321-348. DOI: [10.1016/j.neucom.2019.02.003](https://doi.org/10.1016/j.neucom.2019.02.003).
- [23] 田莹, 王亮, 丁琪. 基于深度学习的图像语义分割方法综述[J]. *软件学报*, 2019, 30(2): 440-468. DOI: [10.13328/j.cnki.jos.005659](https://doi.org/10.13328/j.cnki.jos.005659).
TIAN X, WANG L, DING Q. Review of image semantic segmentation based on deep learning[J]. *Journal of Software*, 2019, 30(2): 440-468. DOI: [10.13328/j.cnki.jos.005659](https://doi.org/10.13328/j.cnki.jos.005659).
- [24] 青晨, 禹晶, 肖创柏, 等. 深度卷积神经网络图像语义分割研究进展[J]. *中国图象图形学报*, 2020, 25(6): 1069-1090. DOI: [10.11834/jig.190355](https://doi.org/10.11834/jig.190355).
QING C, YU J, XIAO C B, et al. Deep convolutional neural network for semantic image segmentation[J]. *Journal of Image and Graphics*, 2020, 25(6): 1069-1090. DOI: [10.11834/jig.190355](https://doi.org/10.11834/jig.190355).
- [25] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe: ACM, 2012: 1097-1105. DOI: [10.5555/2999134.2999257](https://doi.org/10.5555/2999134.2999257).
- [26] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//3rd International Conference on Learning Representations. San Diego: ICLR, 2015.
- [27] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 1-9. DOI: [10.1109/CVPR.2015.7298594](https://doi.org/10.1109/CVPR.2015.7298594).
- [28] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770-778. DOI: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [29] CHOLLET F. Xception: deep learning with Depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 1800-1807. DOI: [10.1109/CVPR.2017.195](https://doi.org/10.1109/CVPR.2017.195).
- [30] NOH H, HONG S, HAN B. Learning deconvolution network for semantic segmentation[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1520-1528. DOI: [10.1109/ICCV.2015.178](https://doi.org/10.1109/ICCV.2015.178).
- [31] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481-2495. DOI: [10.1109/TPAMI.2016.2644615](https://doi.org/10.1109/TPAMI.2016.2644615).
- [32] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich: Springer, 2015: 234-241. DOI: [10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28).
- [33] MILLETARI F, NAVAB N, AHMADI S A. V-Net: fully convolutional neural networks for volumetric medical image segmentation[C]//2016 Fourth International Conference on 3D Vision. Stanford: IEEE, 2016: 565-571. DOI: [10.1109/3DV.2016.79](https://doi.org/10.1109/3DV.2016.79).
- [34] ZHOU Z W, RAHMAN SIDDIQUEE M M, TAJBAKHS H N, et al. UNet++: A nested u-net architecture for medical image segmentation[C]//4th International Workshop on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Granada: Springer, 2018: 3-11. DOI: [10.1007/978-3-030-00889-5_1](https://doi.org/10.1007/978-3-030-00889-5_1).
- [35] HUANG H M, LIN L F, TONG R F, et al. UNet 3+: A full-scale connected UNet for medical image segmentation[C]//2020 IEEE International Conference on Acoustics, Speech and Signal Processing. Barcelona: IEEE, 2020: 1055-1059. DOI: [10.1109/ICASSP40776.2020.9053405](https://doi.org/10.1109/ICASSP40776.2020.9053405).
- [36] ARANI E, MARZBAN S, PATA A, et al. RGPNet: a real-time general purpose semantic segmentation[C]//2021 IEEE Winter Conference on Applications of Com-

- puter Vision. Waikoloa: IEEE, 2021: 3008-3017. DOI: [10.1109/WACV48630.2021.00305](https://doi.org/10.1109/WACV48630.2021.00305).
- [37] WANG Y, ZHOU Q, LIU J, et al. Lednet: A lightweight encoder-decoder network for real-time semantic segmentation[C]//2019 IEEE International Conference on Image Processing (ICIP). Taipei, China: IEEE, 2019: 1860-1864. DOI: [10.1109/ICIP.2019.8803154](https://doi.org/10.1109/ICIP.2019.8803154).
- [38] ROMERA E, ALVAREZ J M, BERGASA L M, et al. ERFNet: efficient residual factorized ConvNet for real-time semantic segmentation[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2018, 19(1): 263-272. DOI: [10.1109/TITS.2017.2750080](https://doi.org/10.1109/TITS.2017.2750080).
- [39] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 936-944. DOI: [10.1109/CVPR.2017.106](https://doi.org/10.1109/CVPR.2017.106).
- [40] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 6230-6239. DOI: [10.1109/CVPR.2017.660](https://doi.org/10.1109/CVPR.2017.660).
- [41] HE H M, CHEN Y R, LI M C, et al. ForkNet: strong semantic feature representation and subregion supervision for accurate remote sensing change detection[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022, 15: 2142-2153. DOI: [10.1109/JSTARS.2022.3152775](https://doi.org/10.1109/JSTARS.2022.3152775).
- [42] ZHU L Y, JI D Y, ZHU S P. Learning statistical texture for semantic segmentation[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 12532-12541. DOI: [10.1109/CVPR46437.2021.01235](https://doi.org/10.1109/CVPR46437.2021.01235).
- [43] LIN G, MILAN A, SHEN C H, et al. RefineNet: multi-path refinement networks for high-resolution semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 5168-5177. DOI: [10.1109/CVPR.2017.549](https://doi.org/10.1109/CVPR.2017.549).
- [44] YU C Q, WANG J B, PENG C, et al. BiSeNet: bilateral segmentation network for real-time semantic segmentation[C]//Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018: 334-349. DOI: [10.1007/978-3-030-01261-8_20](https://doi.org/10.1007/978-3-030-01261-8_20).
- [45] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[C]//4th International Conference on Learning Representations. San Juan: ICLR, 2016.
- [46] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[C]//3rd International Conference on Learning Representations. San Diego: ICLR, 2015.
- [47] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected CRFs[C]. arXiv: 1606.00915, 2016. DOI: [10.48550/arXiv.1606.00915](https://doi.org/10.48550/arXiv.1606.00915).
- [48] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Re-thinking atrous convolution for semantic image segmentation[C]. arXiv: 1706.05587, 2017. DOI: [10.48550/arXiv.1706.05587](https://doi.org/10.48550/arXiv.1706.05587).
- [49] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018: 833-851. DOI: [10.1007/978-3-030-01234-2](https://doi.org/10.1007/978-3-030-01234-2).
- [50] WANG Z Y, JI S W. Smoothed dilated convolutions for improved dense prediction[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London: ACM, 2018: 2486-2495. DOI: [10.1145/3219819.3219944](https://doi.org/10.1145/3219819.3219944).
- [51] WANG P Q, CHEN P F, YUAN Y, et al. Understanding convolution for semantic segmentation[C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Lake Tahoe: IEEE, 2018: 1451-1460. DOI: [10.1109/WACV.2018.00163](https://doi.org/10.1109/WACV.2018.00163).
- [52] CAO Z C, XU X W, HU B, et al. Rapid detection of blind roads and crosswalks by using a lightweight semantic segmentation network[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 22(10): 6188-6197. DOI: [10.1109/TITS.2020.2989129](https://doi.org/10.1109/TITS.2020.2989129).
- [53] ZHAO H S, ZHANG Y, LIU S, et al. PSANet: point-wise spatial attention network for scene parsing[C]//15th European Conference on Computer Vision. Munich: Springer, 2018: 270-286. DOI: [10.1007/978-3-030-01240-3_17](https://doi.org/10.1007/978-3-030-01240-3_17).
- [54] 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. Long Beach: IEEE, 2019: 3141-3149. DOI: [10.1109/CVPR.2019.00326](https://doi.org/10.1109/CVPR.2019.00326).
- [55] HUANG Z L, WANG X G, WEI Y C, et al. CCNet: criss-cross attention for semantic segmentation[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 603-612. DOI: [10.1109/ICCV.2019.00069](https://doi.org/10.1109/ICCV.2019.00069).
- [56] LI R, WANG L B, ZHANG C, et al. A²-FPN for semantic segmentation of fine-resolution remotely sensed images[J]. *International Journal of Remote Sensing*, 2022, 43(3): 1131-1155. DOI: [10.1080/01431161.2022.2030071](https://doi.org/10.1080/01431161.2022.2030071).
- [57] YUAN Y H, HUANG L, GUO J Y, et al. OCNet: object context network for scene parsing[J]. arXiv: 1809.00916, 2018. DOI: [10.48550/arXiv.1809.00916](https://doi.org/10.48550/arXiv.1809.00916).

- [58] LIU Z W, LI X X, LUO P, et al. Semantic image segmentation via deep parsing network[C]//2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1377-1385. DOI: [10.1109/ICCV.2015.162](https://doi.org/10.1109/ICCV.2015.162).
- [59] 徐胜军, 韩九强, 刘光辉. 基于马尔可夫随机场的图像分割方法综述[J]. *计算机应用研究*, 2013, 30(9): 2576-2582. DOI: [10.3969/j.issn.1001-3695.2013.09.004](https://doi.org/10.3969/j.issn.1001-3695.2013.09.004).
XU SHENGJUN, HAN JIUQIANG, LIU GUANGHUI. Survey of image segmentation methods based on Markov random fields[J]. *Application Research of Computers*, 2013, 30(9): 2576-2582. DOI: [10.3969/j.issn.1001-3695.2013.09.004](https://doi.org/10.3969/j.issn.1001-3695.2013.09.004).
- [60] ZHENG S, JAYASUMANA S, ROMERA-PAREDES B, et al. Conditional random fields as recurrent neural networks[C]//2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1529-1537. DOI: [10.1109/ICCV.2015.179](https://doi.org/10.1109/ICCV.2015.179).
- [61] LU Y, CHEN Y R, ZHAO D B, et al. Graph-FCN for image semantic segmentation[C]//16th International Symposium on Advances in Neural Networks. Moscow: Springer, 2019: 97-105. DOI: [10.1007/978-3-030-22796-8_11](https://doi.org/10.1007/978-3-030-22796-8_11).
- [62] HU H Z, CUI J S, WANG L W. Region-aware contrastive learning for semantic segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 16271-16281. DOI: [10.1109/ICCV48922.2021.01598](https://doi.org/10.1109/ICCV48922.2021.01598).
- [63] SOUCY N, SEKEH S Y. CEU-Net: ensemble semantic segmentation of hyperspectral images using clustering[J]. *Journal of Big Data*, 2023, 10: 43. DOI: [10.1186/S40537-023-00718-3](https://doi.org/10.1186/S40537-023-00718-3).
- [64] GHANDORH H, BOULILA W, MASOOD S, et al. Semantic segmentation and edge detection—approach to road detection in very high resolution satellite images[J]. *Remote Sensing*, 2022, 14(3): 613. DOI: [10.3390/rs14030613](https://doi.org/10.3390/rs14030613).
- [65] WU F, CHEN F, JING X Y, et al. Dynamic attention network for semantic segmentation[J]. *Neurocomputing*, 2020, 384: 182-191. DOI: [10.1016/j.neucom.2019.12.042](https://doi.org/10.1016/j.neucom.2019.12.042).
- [66] ZHANG C, JIANG W S, ZHANG Y, et al. Transformer and CNN hybrid deep neural network for semantic segmentation of very-high-resolution remote sensing imagery[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 4408820. DOI: [10.1109/TGRS.2022.3144894](https://doi.org/10.1109/TGRS.2022.3144894).
- [67] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 9992-10002. DOI: [10.1109/ICCV48922.2021.00986](https://doi.org/10.1109/ICCV48922.2021.00986).
- [68] KIM N, KIM D, KWAK S, et al. ReSTR: convolution-free referring image segmentation using transformers[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 18124-18133. DOI: [10.1109/CVPR52688.2022.01761](https://doi.org/10.1109/CVPR52688.2022.01761).
- [69] SANG H W, ZHOU Q H, ZHAO Y. PCANet: pyramid convolutional attention network for semantic segmentation[J]. *Image and Vision Computing*, 2020, 103: 103997. DOI: [10.1016/j.imavis.2020.103997](https://doi.org/10.1016/j.imavis.2020.103997).
- [70] EVERINGHAM M, ESLAMI S M A, VAN GOOL L, et al. The PASCAL visual object classes challenge: a retrospective[J]. *International Journal of Computer Vision*, 2015, 111(1): 98-136. DOI: [10.1007/s11263-014-0733-5](https://doi.org/10.1007/s11263-014-0733-5).
- [71] MOTTAGHI R, CHEN X J, LIU X B, et al. The role of context for object detection and semantic segmentation in the wild[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 891-898. DOI: [10.1109/CVPR.2014.119](https://doi.org/10.1109/CVPR.2014.119).
- [72] CHEN X J, MOTTAGHI R, LIU X B, et al. Detect what you can: detecting and representing objects using holistic models and body parts[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 1979-1986. DOI: [10.1109/CVPR.2014.254](https://doi.org/10.1109/CVPR.2014.254).
- [73] ZHOU B L, ZHAO H, PUIG X, et al. Semantic understanding of scenes through the ADE20K dataset[J]. *arXiv*: 1608.05442, 2016. DOI: [10.48550/arXiv.1608.05442](https://doi.org/10.48550/arXiv.1608.05442).
- [74] BROSTOW G J, FAUQUEUR J, CIPOLLA R. Semantic object classes in video: A high-definition ground truth database[J]. *Pattern Recognition Letters*, 2009, 30(2): 88-97. DOI: [10.1016/j.patrec.2008.04.005](https://doi.org/10.1016/j.patrec.2008.04.005).

作者简介:

马文琪 女,(1999-),硕士研究生.研究方向为视觉检测及应用.

石 颀(通讯作者) 男,(1978-),博士,研究员.研究方向为深度学习与模式识别. E-mail: 2546@usts.edu.cn.

吴宏杰 男,(1977-),博士,教授.研究方向为人工智能及应用.