

## 深度神经网络图像语义分割方法综述

徐 辉<sup>1,2</sup>, 祝玉华<sup>1,2,3+</sup>, 甄 彤<sup>1,2</sup>, 李智慧<sup>1,2</sup>

1. 粮食信息处理与控制教育部重点实验室(河南工业大学), 郑州 450001

2. 河南工业大学 信息科学与工程学院, 郑州 450001

3. 黄河水利职业技术学院, 河南 开封 475000

+ 通信作者 E-mail: 13014528786@163.com

**摘 要:** 图像语义分割是计算机视觉领域近年来的热点研究课题, 随着深度学习技术的兴起, 图像语义分割与深度学习技术进行融合发展, 取得了显著的进步, 在无人驾驶、智能安防、智能机器人、人机交互等真实场景中应用广泛。首先对应用于图像语义分割的几种深度神经网络模型进行简单介绍, 接着详细阐述了现有主流的基于深度神经网络的图像语义分割方法, 依据实现技术的区别对图像语义分割方法进行分类, 并对每类方法中代表性算法的技术特点、优势和不足进行分析与总结。之后归纳了图像语义分割常用的大规模公共数据集和性能评价指标, 并在此基础上对经典的语义分割方法的实验结果进行了对比, 最后对语义分割领域未来可行的研究方向进行展望。

**关键词:** 计算机视觉; 图像语义分割; 深度神经网络

**文献标志码:** A **中图分类号:** TP391

## Survey of Image Semantic Segmentation Methods Based on Deep Neural Network

XU Hui<sup>1,2</sup>, ZHU Yuhua<sup>1,2,3+</sup>, ZHEN Tong<sup>1,2</sup>, LI Zhihui<sup>1,2</sup>

1. Key Laboratory of Grain Information Processing and Control (Henan University of Technology), Ministry of Education, Zhengzhou 450001, China

2. College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China

3. Yellow River Conservancy Technical Institute, Kaifeng, Henan 475000, China

**Abstract:** Image semantic segmentation is a hot research topic in the field of computer vision in recent years. With the rise of deep learning technology, image semantic segmentation and deep learning technology are integrated and developed, which has made significant progress. It is widely used in practical scenarios such as unmanned driving, intelligent security, intelligent robot, human-computer interaction. Firstly, several deep neural network models for image semantic segmentation are introduced, and then the existing mainstream deep neural network-based image semantic segmentation methods are introduced. According to the differences of implementation technologies, image semantic segmentation methods are classified, and the technical characteristics, advantages and disadvantages of representative algorithms are analyzed and summarized. After that, the common datasets and performance evaluation indexes of image semantic segmentation are summarized, and the experimental results of classic semantic segmentation

**基金项目:** 国家重点研发计划(2018YFD0401404)。

This work was supported by the National Key Research and Development Program of China (2018YFD0401404).

**收稿日期:** 2020-04-16 **修回日期:** 2020-08-21

methods are compared on this basis. Finally, the future feasible research directions in the field of semantic segmentation are prospected.

**Key words:** computer vision; image semantic segmentation; deep neural network

图像分割是计算机视觉研究的关键,亦是理解图像内容的基石技术,其发展历程从基于边缘检测、基于阈值、基于区域等方法的传统图像分割,逐渐发展到当今主流的基于深度神经网络的图像语义分割,在技术水平上取得了巨大的进步。传统的图像分割受计算机算力的限制,只能提取图像的纹理信息、颜色、形状等低层特征对图像进行分割,且需要人工设计特征,因而分割准度不高。而随着计算机软硬件的更新换代、深度学习技术的日益强大,语义分割也进入新的发展时期。许多研究学者将深度学习引入图像语义分割领域,利用深度神经网络从输入数据中自动学习特征,能够提取图像的低层、中层和高层特征,实现对图像目标端到端的像素级分类,极大地提高了语义分割的精度和效率。

图像语义分割是像素级别的图像识别和理解,即依据图像所含语义信息对每个像素进行分类标注,将图像分割为语义含义相同的若干区域,并使用不同的颜色标记不同区域,以此将每个区域的类别标注结果可视化。目前语义分割技术已经发展得较为成熟,在自动驾驶、医学图像分析、增强现实等实际场景中应用广泛。

现有的一些相关综述文章<sup>[1-3]</sup>已经对图像语义分割的发展历史、技术革新进行了较为完善的总结,但如文献[2-3]对所列举的语义分割方法的设计思路、优缺点等论述不足,文献[1]则缺乏对语义分割最新研究成果和算法的解读。针对这些问题,本文踵事增华,梳理和总结了现有基于深度神经网络的代表性语义分割方法的基本思路与关键技术,按照实现技术的不同将所述方法进行分类,并分析每类方法的优缺点。此外,还归纳了近两年计算机视觉领域顶级会

议/期刊上部分优秀语义分割算法。最后基于权威的数据集和评价指标对这些算法的性能进行对比,并对语义分割技术未来的热门发展方向进行了展望。

## 1 语义分割常用神经网络

### 1.1 卷积神经网络

卷积神经网络(convolutional neural network, CNN)<sup>[4]</sup>是基于传统神经网络改进的层级网络,通常由卷积层、池化层、全连接层构成,结构如图1所示。卷积层:通过卷积运算进行图像降维以及提取图像的局部特征。池化(也称下采样)层:将卷积操作提取的特征划分为大小相同的 $n$ 个区域,然后取其平均值或最大值特征(通过平均池化或最大值池化操作)表示区域的特征,以降低特征维度,防止过拟合。全连接层:将所有局部特征整合得到完整的特征图,最后利用softmax函数进行图像分类。经典网络模型有AlexNet<sup>[5]</sup>、VGGNet<sup>[6]</sup>、ResNet<sup>[7]</sup>和GoogLeNet<sup>[8]</sup>等。

### 1.2 循环神经网络

循环神经网络(recurrent neural networks, RNN)<sup>[9]</sup>适用于处理序列数据,按时间展开的标准结构如图2所示。其中 $x_t$ 是 $t$ 时刻输入层的值, $h_t$ 是 $t$ 时刻隐藏层的值,从结构图中可以看出 $h_t$ 的值不仅取决于当前时刻的输入 $x_t$ ,还依赖先前时刻隐藏层的值 $h_{t-1}$ 。同样的, $t$ 时刻隐藏层的值 $h_t$ 除了传输到当前时刻的输出层 $o_t$ ,也用于下一时刻隐藏层的学习。因而,RNN网络具有历史记忆能力,可以递归处理历史信息 and 建模序列数据。

### 1.3 生成对抗网络

生成对抗网络(generative adversarial networks,

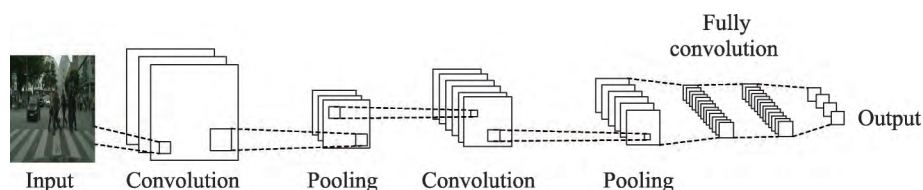


Fig.1 Network architecture of CNN

图1 CNN网络架构

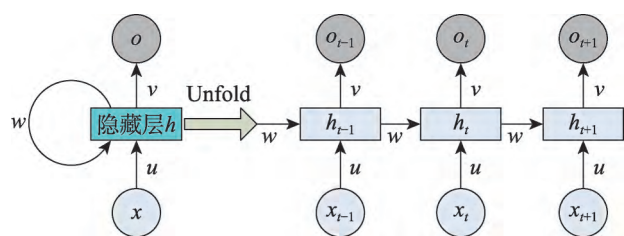


Fig.2 Network expansion diagram of RNN

图2 RNN时间线展开图

GAN)<sup>[10]</sup>是近两年最火爆的深度网络架构,采用无监督学习进行训练,由生成器和判别器两部分组成,如图3所示。随机噪声经生成器处理生成预测图,然后将其作为伪样本和真值标注图像组成的真实样本一起送入到判别器中。判别器的功能是对输入的真实样本、伪样本进行学习和真假判断,并把判断结果反馈给生成器,指导生成器进行参数的优化、更新。生成器和判别器之间进行迭代对抗训练,逐渐提升网络的分割精度。

## 2 基于深度神经网络的图像语义分割

2015年Long等人<sup>[11]</sup>提出了全卷积网络(fully convolutional networks, FCN),架构如图4所示。FCN将图像级的分类网络拓展为像素级分类网络,并且实

现了端到端的网络训练,是将深度学习应用于图像语义分割的开山之作。

FCN网络极大地推动语义分割的发展,但仍然存在许多不足,后续许多文献提出了改进的方法。本文综述了其中的代表性方法,依据技术理念的区别将其分为六类:基于空洞卷积的方法、基于编解码的方法、基于特征融合的方法、基于RNN的方法、基于注意力机制的方法、基于GAN的方法。然后分析和总结了每类方法的优缺点以及其中典型算法的技术特点,如表1所示。

### 2.1 基于空洞卷积的方法

FCN网络在下采样时造成特征图感受野变小,图像的部分空间信息丢失,此外还缺乏对图像上下文信息的利用。针对这些问题,Chen等人<sup>[12]</sup>提出DeepLab V1网络,结构如图5所示。其创新性地将深度卷积神经网络(deep convolutional neural network, DCNN)的部分卷积层替换为空洞卷积(atrous/dilated convolution,也称扩张卷积)<sup>[13]</sup>,在不增加参数的同时增大了感受野,从而获得更多的特征信息。此外,在DCNN的最后一层添加全连接条件随机场(fully connected conditional random field, FCCRF)来增强捕获图像细节信息的能力,实现目标的精确定位。后续,

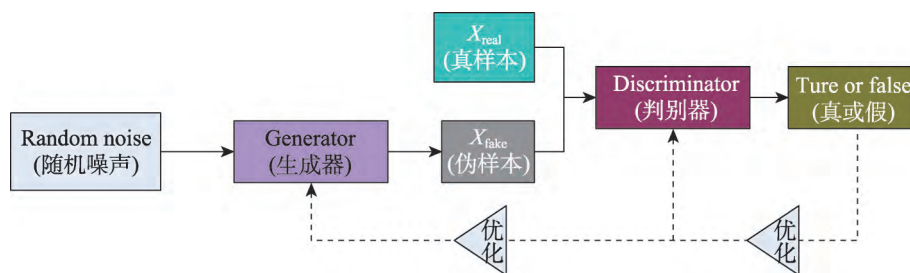


Fig.3 Network architecture of GAN

图3 GAN网络架构

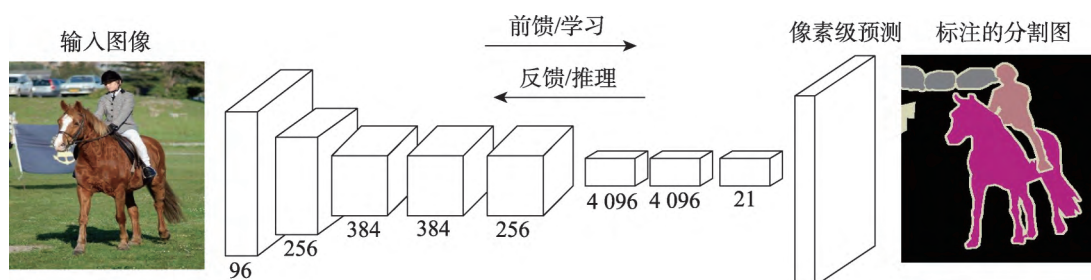


Fig.4 Network architecture of FCN

图4 FCN网络架构



Table 1 Analysis and summary of image semantic segmentation methods

表1 图像语义分割方法的分析与总结

分类	主流算法	算法特点	方法的优势	方法的不足
基于空洞卷积的方法	DeepLab V1	使用空洞卷积提取特征,利用全连接CRF捕获图像细节信息	扩大了感受野,可以提取更多的特征信息	计算成本较高,像素部分位置信息丢失,影响特征图的局部一致性
	DeepLab V2	引入ASPP模块捕获多尺度的上下文信息		
	DeepLab V3	在ASPP模块中添加批次正则化层,以改进ASPP模块,此外去除了FCCRF		
	DeepLab V3+	在DeepLab V3基础上引入解码模块,提高物体边界分割的准确度		
基于编解码的方法	U-Net	使用跳跃连接将编码网络的特征图拼接 to 相对应的解码网络的特征图	在FCN基础上,加入解码网络,有效融合低层和高层特征,恢复图像的空间维度和边界信息	网络结构复杂,参数数量多,物体边界分割效果不佳
	SegNet	解码器层使用对应编码器层存储的最大池化索引对特征图进行上采样,增强了图像边界定位的准确度		
	ENet	采用较大编码器和较小解码器结构,减少了模型的参数数量的同时保持良好的分割准确度		
	LEDNet	非对称编码器-解码器结构结合注意力机制,降低了网络的复杂度		
基于特征融合的方法	ParseNet	使用全局平均池化提取全局特征,并将局部特征与全局特征融合	有效获取图像的上下文信息,并提高了对图像局部、全局特征的利用	会丢失目标部分边界细节信息
	RefineNet	利用RefineNet模块有效融合低层特征和高层特征		
	PSPNet	引入空间金字塔模块来聚合不同尺度的特征,以有效捕获上下文信息		
基于RNN的方法	ReSeg	级联多个RNN模型,获取图像的上下文信息	可以记忆历史信息,易于提取像素的序列特征和获取全局上下文信息	处理远距离序列信息时存在梯度消失问题,可能造成部分像素信息的丢失,导致像素的错误分类
	2D LSTM	使用4个独立且不同方向的LSTM获取全局上下文信息		
	Graph-LSTM	基于超像素构造图拓扑结构,高效地将图像上下文信息传输到各个图节点		
	DA-RNN	利用数据间的相关性来融合不同时刻视频帧的特征,提升标注的准确度		
基于注意力机制的方法	PAN	注意力机制与空间金字塔结合,来提取准确且密集的特征进行像素标注	能够有效捕获重要的上下文信息,模型结构简单,参数量少	需要考虑每个像素之间的关系,因此计算量较大。此外,获取图像位置信息的能力不足
	DANet	利用自注意力机制整合图像的局部特征,捕获上下文依赖关系		
	CCNet	引入纵横交叉注意模块,利用像素间长距离依赖关系捕获全局上下文信息		
基于GAN的方法	文献[32]	生成器与判别器之间进行对抗训练,提高分割图与真值标注图的一致性	增强了远程空间标签的连续性,提高语义标注准确度的同时没有增加模型的复杂度	网络训练困难且不稳定,处理大规模数据时可解释性差
	文献[33]	利用GAN网络进行半监督分割,减少语义标注工作量的同时良好保持了分割的准确度		
	文献[54]			

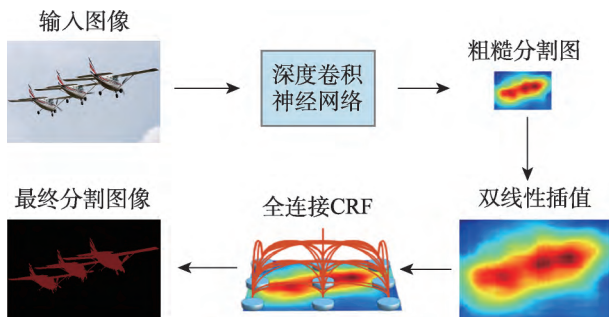


Fig.5 Workflow of DeepLab V1

图5 DeepLab V1网络处理流程

Chen等人对DeepLab V1进行扩展提出了DeepLab V2<sup>[14]</sup>,其将空洞卷积和空间金字塔池化模型结合,提出了带孔空间金字塔池化(atrous spatial pyramid pooling,ASPP)模块。ASPP模块使用多个不同采样率的空洞卷积来获取不同尺度的特征,并将特征进行融合以获取上下文信息,实现多尺度目标的处理。最后使用FCCRF优化边界分割效果。

2017年,Chen等人在DeepLab V1、V2的基础上提出了DeepLab V3<sup>[15]</sup>,其在ASPP模块中增加了批正则化(batch normalization)层,改进了ASPP模块。同

时将串行/并行连接的空洞卷积与改进的ASPP模块以串行连接的方式结合,以更有效地提取特征和捕获全局上下文信息,增强处理多尺度物体的能力。DeepLab V3相比DeepLabV1、V2去除了FCCRF,但性能进一步提高。针对DeepLab V3生成的预测图稀疏、空洞卷积造成的边界信息丢失等问题,Chen等人提出了DeepLab V3+<sup>[16]</sup>。其基于DeepLab V3设计编码网络,来编码全局上下文信息,引入解码网络来恢复目标的边界细节信息。此外,在ASPP模块和解码网络中添加可分离深度卷积层,提高了网络的运行速率和鲁棒性,并大幅提升了分割准度。

## 2.2 基于编解码的方法

解决池化操作导致的图像分辨率降低、像素空间信息丢失问题的另一个方法是引入编码器-解码器结构。编码器通常由多个卷积层和池化层组成,作用是从原图中获取含有位置信息和语义信息的特征图。而解码器通常由反卷积层和反池化层构成,作用是恢复特征图中丢失的空间维度和位置信息,生成稠密的预测图。

FCN在上采样时使用反卷积和跳跃连接解决位置信息丢失问题,但是最终分割结果依然较粗糙。文献[17]基于FCN提出了SegNet,采用对称的编码器-解码器结构,如图6所示。编码器基于VGG-16网络但移除了全连接层,降低了模型复杂度。其中池化层还存储特征图最大池化的位置信息。解码器由上采样层和反卷积层组成,依据相应编码器层存储的最大池化索引对输入特征进行上采样,有效减少图像位置信息的丢失,从而生成密集的特征图。最后将特征图送入Softmax分类器中进行像素分类。SegNet减少了内存占用和参数量,提高了语义分割

的准度,但对小尺寸目标进行上采样时,仍会丢失部分边界信息。

U-Net<sup>[18]</sup>网络结构与SegNet相似,采用的是“U型”的编码器-解码器结构,主要应用于医学图像分析领域。独特之处在于将编码器中低分辨率特征图通过跳跃连接直接拼接到对应解码器上采样生成的特征图,从而有效融合了低层的细节信息和高层的像素分类信息,实现更精确的分割。

传统的编解码器结构复杂、参数数量多,因此后续许多研究学者对编解码器的结构进行优化。如ENet<sup>[19]</sup>网络采用大的编码器和较小的解码器,简化了解码器结构。同时,通过低阶近似运算分解卷积操作,以减少参数量,提高语义分割的实时性。此外,Wang等人<sup>[20]</sup>提出轻量化网络LEDNet,也采用非对称的编解码器结构。该网络中编码器基于ReNet,但改进了残差模块,从而增强了特征表达能力,减小了网络规模。解码器中引入注意力机制预测每个像素的语义标签,进一步减少了网络的计算量。实验结果表明,该网络实现了分割速度与准度的良好平衡。

## 2.3 基于特征融合的方法

FCN网络通常提取图像的局部特征进行像素分类,缺乏对图像全局特征、上下文信息的利用,因而导致分割结果较为粗糙。基于特征融合的方法通过融合不同尺度、不同层次的特征,提高获取图像上下文信息的能力。此外,能够将图像的局部特征和全局特征连接来优化分割结果。

ParseNet<sup>[21]</sup>网络首先通过全局平均池化来获取图像的全局特征,然后将全局特征与局部特征进行融合,并用于分类器的学习,从而有效利用了全局上下文信息,提高了语义分割的准度。

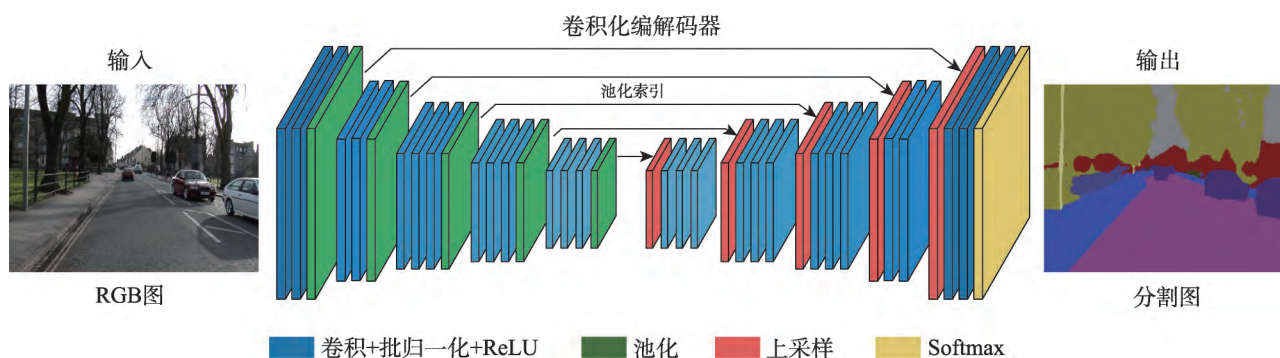


Fig.6 Network architecture of SegNet

图6 SegNet网络架构

为了解决空间信息丢失问题, Lin 等人<sup>[22]</sup>提出了多路径细化网络(简称 RefineNet), 能够将多尺度的特征融合得到精细的分割结果。该网络首先把原始图像输入 CNN 网络中生成 4 个不同尺度、不同分辨率的特征图, 再把 4 个特征图分别送入对应的 4 个精细化模块(RefineNet block)中。RefineNet block 由基于残差连接设计的多个组件构成, 可以将粗糙的高层特征和细粒度的低层特征进行连接, 并有效融合了下采样过程丢失的特征信息, 从而良好保留像素的空间信息, 生成高分辨的分割图。

针对 FCN 在场景分割时存在的上下文关系推断能力不足问题, Zhao 等人<sup>[23]</sup>提出了金字塔场景解析网络(简称 PSPNet), 结构如图 7 所示。该网络首先使用添加了扩张卷积的 ResNet 网络提取特征, 之后将特征输入到金字塔池化模块(pyramid pooling module)中。金字塔池化模块通过对输入特征进行 4 种不同尺度的池化, 得到 4 个不同层级的特征图; 然后对不同层级的特征图进行上采样恢复到池化前的大小, 并与池化前的特征进行拼接; 最后通过卷积操作生成最终的预测图。PSPNet 网络通过融合不同尺度的特征, 有效利用了局部和全局上下文信息。此外, 在

基础网络训练过程中添加辅助损失函数, 降低优化的难度, 从而实现了高质量的像素级场景解析。

## 2.4 基于 RNN 的方法

解决 FCN 网络忽略像素间的关联性, 对全局上下文信息利用不足问题的另一个方法是利用 RNN 网络进行语义分割。RNN 网络具有记忆历史信息特性, 可以利用先前时刻的信息指导下一时刻的输出, 充分考虑了图像中像素之间的相关性, 有利于提取像素的序列信息和建模语义依赖关系, 促进对全局上下文信息的利用。

受 RNN 网络启发, 文献[24]基于由 RNN 改进的图像分类模型 ReNet<sup>[25]</sup>提出了 ReSeg 网络, 结构如图 8 所示。该网络先将原始图像输入预训练的 VGG-16 网络进行特征提取; 然后把特征送入到多个 ReNet 层中。ReNet 层由 4 个 RNN 组成, 通过在水平和垂直两个方向扫描图像, 有效获取图像的全局特征和上下文信息; 最后使用若干由反卷积网络构成的上采样层恢复特征图的分辨率。ReSeg 网络处理分割任务时十分高效、灵活, 但计算量较大, 对部分类别的像素的分割效果不佳。

传统 RNN 网络无法学习远距离节点的信息, 存在

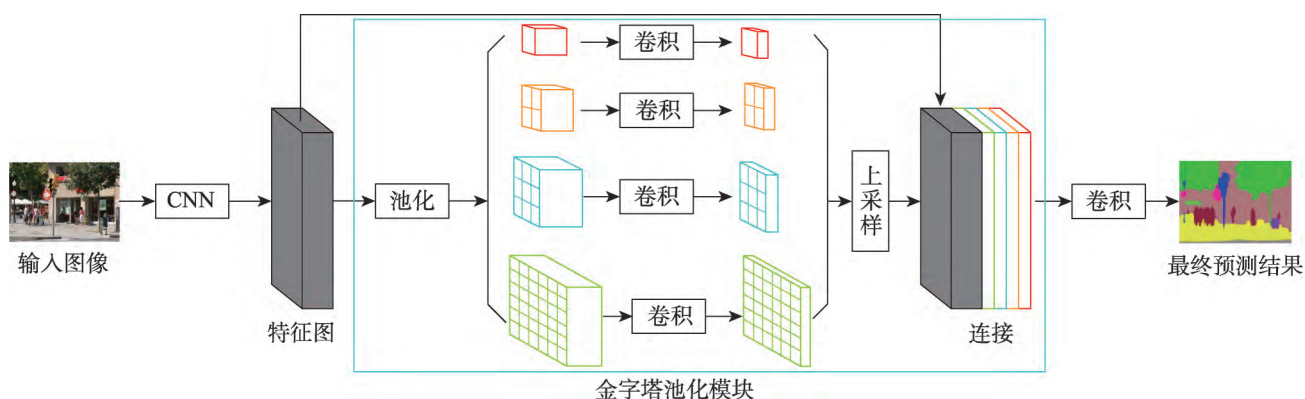


Fig.7 Workflow of PSPNet

图 7 PSPNet 网络处理流程

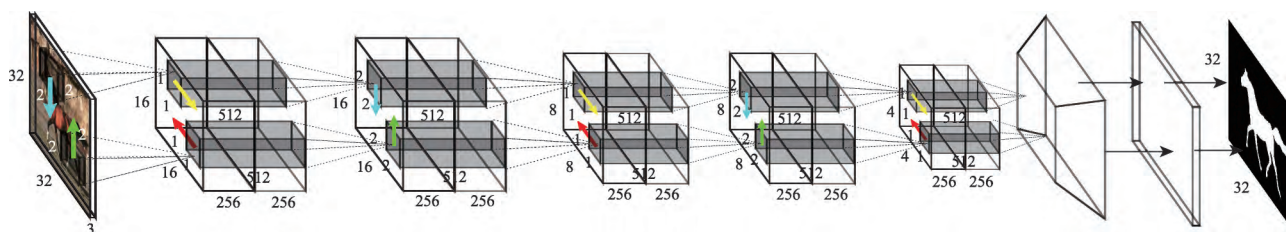


Fig.8 Network architecture of ReSeg

图 8 ReSeg 网络架构



长期依赖问题。针对这一情况,后续在RNN的基础上衍生出长短期记忆(long short-term memory, LSTM)<sup>[26]</sup>网络和门控循环单元(gated recurrent unit, GRU)<sup>[27]</sup>。Byeon 等人<sup>[28]</sup>提出了二维长短期记忆循环神经网络(2D LSTM)。该网络将输入图像分成若干个非折叠窗口,然后送入4个独立的LSTM记忆单元,利用LSTM的远程记忆特性来捕获图像的局部、全局上下文信息,以较低的计算复杂度实现了先进的性能。普通的LSTM网络处理时需要将图像固定且均匀地划分为块,会造成图像属性利用不充分。因此,Liang 等人<sup>[29]</sup>提出 Graph-LSTM,将传统LSTM从序列数据拓展到图结构数据,将超像素作为图的节点,基于超像素及其空间连接自适应地构造一个无向图拓扑结构。Graph-LSTM利用图拓扑结构可以高效地将图像的上下文信息传输到图中所有超像素节点,降低冗余计算量的同时良好保留了目标的边界信息,从而增强了特征表示。

RNN网络在三维场景解析中亦有应用。Xiang 等人<sup>[30]</sup>提出了数据关联循环神经网络(data associated recurrent neural networks, DA-RNN),在RNN网络中引入一个新的数据关联递归单元(data associated recurrent unit, DA-RU),用于融合先前视频帧和当前视频帧的特征,以对RGB-D视频的每一帧进行语义标注,之后把网络的输出与KinectFusion技术结合,将语义标签融入到三维空间中,实现三维场景重建。

## 2.5 基于注意力机制的方法

注意力机制<sup>[31]</sup>的基本原理是为特征图中不同的区域分配不同的权重,即提取特征图中关注对象的有用信息,同时抑制无用信息,从而实现高效的特征提取,并降低了网络训练难度。其次,注意力机制有助于获取全局上下文信息,促进语义分割准度的提升。

金字塔注意力网络(pyramid attention network, PAN)<sup>[32]</sup>将注意力机制和空间金字塔结合起来获取密集的特征和语义信息,其创新工作在于:(1)提出特征金字塔注意模块(feature pyramid attention, FPA),利用金字塔模型融合不同尺度的上下文信息,同时将像素级注意力应用于高层特征,以增强像素分类的一致性。(2)提出全局注意上采样模块(global attention upsampling, GAU),利用全局平均池化来提取高层特征的上下文信息,用于指导低层特征恢复像素的细节信息,从而促进分割精度的提升。

针对传统的FCN网络提取的特征导致的目标错误分类问题,Fu 等人<sup>[33]</sup>提出双注意网络(dual attention network, DANet)。该网络先基于ResNet模型提取特征图,然后将特征图输入到并行连接的位置注意力模块和通道注意力模块。其中位置注意力模块利用自注意力机制捕获特征图中任意两个位置间的空间依赖关系,通道注意力模块利用自注意力机制捕获通道特征图之间的依赖关系。最后使用卷积操作将两个模块的输出进行聚合,使网络可以在局部特征上建模全局上下文依赖关系,并进一步增强了特征表示。

DANet网络虽然提升了分割的准度,但模型计算量较大。文献[34]提出纵横交叉网络(criss-cross network, CCNet),设计了纵横交叉注意力(criss-cross attention, CCA)模块,通过特征加权求和来建模像素间的远距离依赖关系,以在水平和垂直方向上捕获图像的全局上下文信息。该网络使用较少的GPU内存和计算成本,达到了先进的分割效果。

## 2.6 基于GAN的方法

传统的基于FCN的语义分割方法常使用CRF(conditional random field)、金字塔模型来提升算法的分割准度,但是存在模型结构复杂、计算成本高等问题。而基于GAN网络的语义分割方法通过对抗训练优化模型性能,在不增加网络复杂度的情况下,增强了像素预测结果的一致性,进而促进语义分割准度的提高。

经生成对抗网络的启发,Luc 等人<sup>[35]</sup>首次将生成对抗网络应用于图像语义分割。首先利用分割网络生成预测图,然后将预测图和真实标签图输入到判别网络中判断真假。两个网络之间进行迭代训练,提高了预测图与真实标签图之间的高阶一致性。

全监督语义分割是对每个像素进行语义标注,时间成本高且难度大。为了减少标注时间和代价,文献[36]将GAN应用于半监督学习,利用少量的标注数据和大量的未标注数据进行语义分割,并利用条件生成对抗网络<sup>[37]</sup>生成高质量的样本,以提升像素分类的一致性。无独有偶,文献[38]也利用GAN进行半监督语义分割,以有效减少人工标注的工作量。不同点在于利用未标注的数据训练分割网络时,将判别网络的输出作为自学习信号,用于补偿图像语义标注的缺失和优化分割网络。

### 3 语义分割算法的性能分析与对比

随着语义分割技术的发展,越来越多的分割算法被提出,继而需要对这些算法的性能进行标准化、系统化的评估。而评估不同算法的性能需要建立在大型的数据集和统一的评价指标上。因此,本章对图像语义分割实验中常用的公共数据集和性能评估指标进行了归纳,并以此为标准对本文所述算法的性能进行对比。

#### 3.1 图像语义分割常用数据集

本节整理了常用的大型公共数据集,并对各个数据集的基本信息、主要应用场景进行了汇总,如表2所示。

Stanford background<sup>[39]</sup>: 斯坦福大学发布的一个室外场景数据集,数据主要来自 LabelMe、MSRC、PASCAL VOC 等公共数据集,包含 715 张图像,像素约为 320×240。数据集有 8 个类别,包括道路、树木、草、水、建筑物、山脉、天空和前景物体。

SiftFlow<sup>[40]</sup>: 该数据集是 LabelMe 数据集的子集,包含 33 个语义类别,共有 2 688 张标注的图像,大部分图像取自 8 种不同的户外场景(山脉、海滩、街道、

城市等),每张图像的像素值为 256×256。

KITTI<sup>[41]</sup>: 自动驾驶场景最常用数据集,包含从德国乡村、城市和高速公路采集的图像,适用于道路检测、视觉测距、三维物体检测和跟踪等挑战。原始数据集没有提供真实的语义标注,后来 Alvarez 等人<sup>[42]</sup>、Zhang 等人<sup>[43]</sup>和 Ros 等人<sup>[44]</sup>为其中部分类的图像添加了语义标注。

SBD(semantic boundaries dataset)<sup>[45]</sup>: SBD 数据集对 PASCAL VOC 数据集进行了拓展,并继承 PASCAL VOC 中的 11 355 张语义标注图像。数据集分为训练集(8 498 张图像)和验证集(2 857 张图像),实际应用中已逐渐替代 PASCAL VOC 数据集。

PASCAL VOC(PASCAL visual object classes)<sup>[46]</sup>: 最初是为目标检测任务创立的国际竞赛,从 2005 年一直发展到 2012 年,产生了一系列高质量的数据,目前 PASCAL VOC 2012 是最常用的。数据集中共有 21 个类别(含背景),包括人、动物、交通工具、室内物品等。

PASCAL Context<sup>[47]</sup>: 在 PASCAL VOC 数据集的基础上进行了扩展,总共有 540 个类,包含 10 103 张

Table 2 Common datasets for semantic segmentation

表 2 语义分割常用公共数据集

应用场景	数据集	时间	简要介绍	类别数	训练集	验证集	测试集
室外场景	Stanford background	2009	大型户外场景解析数据集之一,包含 715 张图像,像素值为 320×240	8	725	—	—
	SiftFlow	2011	数据集包含 2 688 张完全标注的图像,其中大部分图像来自 8 种不同的户外场景	33	2 688	—	—
驾驶场景	KITTI-Ros	2015	智能机器人和自动驾驶领域常用数据集之一,所有图片都是真彩色 png 格式,每张图像有不同程度的遮挡和截断	11	170	—	46
	KITTI-Zhang	2015		10	140	—	112
多种场景	SBD	2011	PASCAL VOC 数据集的拓展,对其中未标注的图像提供含有像素级和实例级信息的语义标注	21	8 498	2 857	—
	PASCAL VOC 2012	2012	数据集由 1 个真实标注的图像数据集和 5 种不同挑战赛的数据组成	21	1 464	1 449	1 452
	PASCAL-Context	2014	PASCAL VOC 2010 数据集的拓展,其中 10 103 张图像进行了像素级标注	540	10 103	10 103	9 637
	MS COCO	2014	数据集主要由室内和室外场景的图像组成,其中每张图片都包含多个类别和实例目标	81	82 783	40 504	81 434
	ADE20K	2017	数据集包含室内、室外场景中的 150 种物体类别和超过 25 000 张图像	151	20 210	2 000	3 352
城市街道场景	Cityscapes	2015	一个大规模城市道路和交通数据集,包含了 8 大类别 30 个种类的像素级、实例级标注图像	30	22 973	500	—
室内场景	SUN RGB-D	2015	应用于场景解析任务,主要包含由 4 款 RGB-D 传感器捕获的 10 000 张 RGB-D 图像	37	2 666	2 619	5 050



语义标注的图像。该数据集类别繁多,但其中许多类比较稀疏,因此在评估语义分割算法性能时,通常使用出现频率最高的59个类作为语义标签。

MS COCO(Microsoft common objects in context)<sup>[48]</sup>:由微软公司发布,数据集中图像种类丰富,其包含的图像大多选取自复杂的室内和室外场景,常用于图像识别、语义分割任务。数据集含有80个类别,328 000幅图像、2 500 000个对象实例。

ADE20K<sup>[49]</sup>:MIT推出的用于场景解析、分割、目标识别、语义理解的数据集,该数据集中共拥有超过25 000张图像、150个语义类别,并对图像中的目标进行了密集的像素标注。其中训练集有20 210张图像,验证集有2 000张图像。

Cityscapes<sup>[50]</sup>:主要用于城市街道场景解析的大型数据集,包含了从50个城市获取的不同时间段、不同背景和场景布局下的高分辨率图像,数据集为部分图像提供了语义、实例标注,其中精细标注的图像约5 000张,粗略标注的图像20 000张。

SUNRGB-D<sup>[51]</sup>:主要应用于场景解析任务,数据集由4款RGB-D传感器采集的图像和NYU depthv2、SUN3D等数据集组成。共包括10 335张室内场景图像、146 617个二维多边形标注、58 657个三维边界框标注以及不同场景的布局信息、类别信息。

### 3.2 图像语义分割常用评价指标

目前学术界通常从运行时间、内存占用、准确度三方面衡量语义分割算法的性能,因准确度最具客观性和公正性,所以本节着重介绍图像语义分割准确度的评价指标。主要包括像素准确率(pixel accuracy, PA)<sup>[52]</sup>、交并比(intersection over union, IoU)<sup>[52]</sup>、平均交并比(mean intersection over union, mIoU)<sup>[52]</sup>等。其中mIoU简洁且代表性强,是图像语义分割实验评测中最常用的指标。PA、IoU、mIoU三个评价指标的定义、计算公式的详细阐述如下。

(1)PA:表示图像中正确分类的像素数目与像素总数的比率。

$$PA = \frac{\sum_{i=1}^k p_{ii}}{\sum_{i=1}^k t_i}$$

(2)IoU:预测图集合和真实标注图集合的交集与这两个集合的并集的比率,图像语义分割中常表示分割图与原始图像真值之间的重合程度。

$$IoU = \frac{\sum_{i=1}^k p_{ii}}{t_i + \sum_{j=1}^k (p_{ji} - p_{ii})}$$

(3)mIoU:图像像素每个类的IoU值累加后的平均值。

$$mIoU = \frac{1}{k} \sum_{i=1}^k \frac{p_{ii}}{t_i + \sum_{j=1}^k (p_{ji} - p_{ii})}$$

式中, $k$ 表示像素的类别数; $p_{ii}$ 表示实际类别为*i*、预测的类别也为*i*的像素的数目; $t_i$ 表示类别为*i*的像素的总数; $p_{ji}$ 表示实际类别为*i*、预测的类别为*j*的像素的数目。

### 3.3 语义分割算法性能对比

本文中所述语义分割算法的关键技术以及在PASCAL VOC 2012、CityScapes等常用数据集上的mIoU实验测试结果对比,如表3所示。

## 4 总结与展望

本文简述了语义分割中常用的几种深度神经网络,将基于深度神经网络的语义分割方法依据所用技术的区别划分为基于空洞卷积的方法、基于编解码的方法等六类方法,并对这六类方法中的主流算法进行细致的研究和总结。在对语义分割技术深入研究后,发现语义分割技术依然存在许多挑战和可行的研究方向。

#### (1)实时语义分割

目前许多语义分割算法已经实现了良好的分割精度,但运行速度却无法达到实时性的要求,实现实时语义分割的难点在于要在保证分割准度的前提下提升分割速度。目前的主流解决方法:一是采用高效且轻量化的主干网络;二是采用多分支网络结构;三是裁剪输入图像的尺寸。如RGPNet<sup>[54]</sup>网络采用非对称编解码结构,使网络模型轻量化,提高了分割精度并减少了计算成本。旷视公司提出的DFANet<sup>[53]</sup>网络通过充分利用深度聚合的高层特征以及轻量化的编解码器结构,实现了实时分割。目前越来越多的实际应用场景要求在更短的响应时间达到精确的分割结果(如无人驾驶、安防监控),因此未来的语义分割将在保持高精度的同时,探索如何进一步提高分割速度。

#### (2)三维点云数据的语义分割

基于深度学习的语义分割已经在一维、二维数

Table 3 Comparison of experimental results of image semantic segmentation methods

表3 图像语义分割方法的实验结果对比

分类	主流算法	时间	关键技术	数据集	mIoU/%
基于空洞卷积的方法	DeepLab V1	2014	空洞卷积、FCCRF	PASCAL VOC 2012	71.6
				CityScapes	63.1
	DeepLab V2	2016	ASPP模块	PASCAL VOC 2012	79.7
				CityScapes	70.4
	DeepLab V3	2017	改进的ASPP模块	PASCAL VOC 2012	86.9
				CityScapes	81.3
基于编解码的方法	DeepLab V3+	2018	编码器解码器结构 Xception	PASCAL VOC 2012	89.0
				CityScapes	82.1
	U-Net	2015	特征拼接、转置卷积	—	—
	SegNet	2015	上采样、池化层索引	CamVid	60.1
	ENet	2016	分解卷积层	CityScapes	58.3
				CamVid	51.3
基于特征融合的方法	LEDNet	2020	ReNet、Attention mechanism	CityScapes	70.6
	ParseNet	2016	全局平均池化	PASCAL VOC 2012	69.8
	RefineNet	2016	RefineNet block	PASCAL VOC 2012	83.4
				CityScapes	73.6
	PSPNet	2016	Pyramid pooling module	PASCAL VOC 2012	85.4
				CityScapes	81.2
基于RNN的方法	ReSeg	2016	ReNet、RNN	MS COCO	82.6
				CamVid	—
	2D LSTM	2015	LSTM	SIFT Flow	—
	Graph-LSTM	2016	图拓扑结构、LSTM	PASCALPersonPart	60.16
	DA-RNN	2017	DA-RU、RNN	SUN RGB-D	90.10
				—	—
基于注意力机制的方法	PAN	2018	Attention mechanism	PASCAL VOC 2012	84.0
			Spatial pyramid structure	CityScapes	78.6
	DANet	2019	Self-attention mechanism	PASCAL VOC 2012	82.6
				CityScapes	81.5
	CCNet	2019	Criss-cross attention module	CityScapes	81.4
				—	—
基于GAN的方法	文献[32]	2016	GAN	PASCAL VOC 2012	54.3
	文献[33]	2016	半监督学习、GAN	CityScapes	67.8
	文献[53]	2018	半监督学习、GAN	PASCAL VOC 2012	—

据的处理上取得了很大的成功,而针对点云等三维数据的语义分割近几年却发展缓慢。2017年文献[55]提出的 PointNet 网络首次实现在非规则点云数据上进行图像分割,2020年文献[56]提出 RandLA-Net 网络,通过逐渐增加每个点的感受野来更好地学习点云的几何结构,实现了大规模三维点云的处理。虽说近两年三维点云数据的语义分割研究火热,但依然存在许多挑战。如现有的三维数据集规模都较小,而点云语义分割是建立在大规模三维数据的基础上,因此如何建立完善的大型三维数据集是一个不

小的难点。此外,如何将稀疏的、无序的、非结构化的点云数据结构化、离散化,以促进点云数据分割准度的提升也是未来研究的热点方向。

### (3) 基于图卷积网络的语义分割

卷积神经网络在规则的欧式数据(如图像、文本等)处理中取得了卓越的成就,但无法处理非欧式数据(如图数据)。图卷积网络将卷积神经网络从欧式数据拓展到图结构数据,实现端到端地学习图节点的特征信息和结构信息,并有效增强了特征表示能力,是图数据处理的最佳手段。文献[57]将图像网格数

据拓展为图结构数据,利用图卷积网络解决图节点的分类问题,减少了局部位置信息的丢失,取得了良好的分割结果。目前,许多真实场景都是以图数据的形式存在的,且图数据建模在视觉任务中的重要性日益凸显(如三维点云数据分割)。因此,基于图卷积网络进行语义分割是一个极具潜力的研究方向。

## 参考文献:

- [1] TIAN X, WANG L, DING Q. Review of image semantic segmentation based on deep learning[J]. Journal of Software, 2019, 30(2): 440-468.  
田萱, 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.
- [2] LIANG X Y, LUO C, QUAN J C, et al. Research on progress of image semantic segmentation based on deep learning[J]. Computer Engineering and Applications, 2020, 56(2): 18-28.  
梁新宇, 罗晨, 权冀川, 等. 基于深度学习的图像语义分割技术研究进展[J]. 计算机工程与应用, 2020, 56(2): 18-28.
- [3] KUANG H Y, WU J J. Survey of image semantic segmentation based on deep learning[J]. Computer Engineering and Applications, 2019, 55(19): 12-21.  
邝辉宇, 吴俊君. 基于深度学习的图像语义分割技术研究综述[J]. 计算机工程与应用, 2019, 55(19): 12-21.
- [4] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [5] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [6] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv:1409.1556, 2014.
- [7] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, Jun 7-12, 2015. Washington: IEEE Computer Society, 2015: 1-9.
- [8] HE K M, ZHANG X Y, REN S Q, et al. Deep residual-learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Jun 27-30, 2016. Washington: IEEE Computer Society, 2016: 770-778.
- [9] LIPTON Z C, BERKOWITZ J, ELKAN C. A critical review of recurrent neural networks for sequence learning[J]. arXiv:1506.00019, 2015.
- [10] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. arXiv:1406.2661, 2014.
- [11] LONG J, SHEHMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, Jun 7-12, 2015. Washington: IEEE Computer Society, 2015: 3431-3440.
- [12] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[J]. arXiv:1412.7062, 2014.
- [13] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[J]. arXiv:1511.07122, 2015.
- [14] 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.
- [15] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv:1706.05587, 2017.
- [16] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//LNCS 11211: Proceedings of the 15th European Conference on Computer Vision, Munich, Sep 8-14, 2018. Berlin, Heidelberg: Springer, 2018: 833-851.
- [17] 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.
- [18] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//LNCS 9351: Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Oct 5-9, 2015. Berlin, Heidelberg: Springer, 2015: 234-241.
- [19] PASZKE A, CHAURASIA A, KIM S, et al. Enet: a deep neural network architecture for real-time semantic segmentation[J]. arXiv:1606.02147, 2016.
- [20] WANG Y, ZHOU Q, LIU J, et al. Lednet: a lightweight encoder-decoder network for real-time semantic segmentation[C]//Proceedings of the 2019 IEEE International Conference on Image Processing, Taipei, China, Sep 22-25, 2019: 1860-1864.
- [21] LIU W, RABINOVICH A, BERGA C. ParseNet: looking wider to see better[J]. arXiv:1506.04579, 2015.
- [22] LIN G S, MILAN A, SHENC H, et al. Refinenet: multi-path refinement networks for high-resolution semantic seg-



- mentation[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Jul 21-26, 2017. Washington: IEEE Computer Society, 2017: 1925-1934.
- [23] ZHAO H S, SHI J P, QI X J, et al. Pyramids cene parsing network[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Jul 21-26, 2017. Washington: IEEE Computer Society, 2017: 6230-6239.
- [24] VISIN F, REMORO A, CHO K, et al. ReSeg: a recurrent neural network-based model for semantic segmentation[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, Jun 26-Jul 1, 2016. Washington: IEEE Computer Society, 2016: 426-433.
- [25] VISIN F, KASTNER K, CHO K, et al. ReNet: a recurrent neural network based alternative to convolutional networks [J]. arXiv:1505.00393, 2015.
- [26] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [27] CHO K, MERRIENBOER B V, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv:1406.1078, 2014.
- [28] BYEON W, BREUEL T M, RAUE F, et al. Scene labeling with LSTM recurrent neural networks[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, Jun 7-12, 2015. Washington: IEEE Computer Society, 2015: 3547-3555.
- [29] LIANG X D, SHEN X H, FENG J S, et al. Semantic object parsing with graph LSTM[C]//LNCS 9905: Proceedings of the 2016 European Conference on Computer Vision, Amsterdam, Oct 11-14, 2016. Berlin, Heidelberg: Springer, 2016: 125-143.
- [30] XIANG Y, FOX D. DA-RNN: semantic mapping with data associated recurrent neural networks[J]. arXiv:1703.03098, 2017.
- [31] MNIH V, HEES N, GRAVES A, et al. Recurrent models of visual attention[C]//Proceedings of the Annual Conference on Neural Information Processing Systems, Montreal, Dec 8-13, 2014. Red Hook: Curran Associates, 2014: 2204-2212.
- [32] LI H C, XIONG P F, AN J, et al. Pyramid attention network for semantic segmentation[J]. arXiv:1805.10180, 2018.
- [33] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation[J]. arXiv:1809.02983, 2018.
- [34] HUANG Z L, WANG X G, HUANG L C, et al. CCNET: criss-cross attention for semantic segmentation[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Oct 27-Nov 2, 2019. Piscataway: IEEE, 2019: 603-612.
- [35] LUC P, COUPRIE C, CHINTALA S, et al. Semantic segmentation using adversarial networks[J]. arXiv:1611.08408, 2016.
- [36] SOULY N, SPAMPINATO C, SHAH M. Semi and weakly supervised semantic segmentation using generative adversarial network[J]. arXiv:1703.02382, 2017.
- [37] MIRZA M, OSINDERO S. Conditional generative adversarial nets[J]. Computer Science, 2014, 27(8): 2672-2680.
- [38] HUNG W C, TSAI H Y, LIOU Y T, et al. Adversarial learning for semi-supervised semantic segmentation[J]. arXiv:1802.07934, 2018.
- [39] GOULD S, FULTON R, KOLLER D. Decomposing a scene into geometric and semantically consistent regions[C]//Proceedings of the IEEE 12th International Conference on Computer Vision, Kyoto, Sep 27-Oct 4, 2009. Washington: IEEE Computer Society, 2009: 1-8.
- [40] LIU C, YUEN J, TORRALBA A. Nonparametric scene parsing: label transfer via dense scene alignment[C]//Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, Jun 20-25, 2009. Washington: IEEE Computer Society, 2009: 1972-1979.
- [41] GEIGER A, LENZ P, STILLER C, et al. Vision meets robotics: the KITTI dataset[J]. The International Journal of Robotics Research, 2013, 32(11): 1231-1237.
- [42] ALVAREZ J M, GEVERS T, LECUN Y, et al. Road scene segmentation from a single image[C]//LNCS 7578: Proceedings of the 2012 European Conference on Computer Vision, Florence, Oct 7-13, 2012. Berlin, Heidelberg: Springer, 2012: 376-389.
- [43] ZHANG R, CANDRA S A, VETTER K, et al. Sensor fusion for semantic segmentation of urban scenes[C]//Proceedings of the 2015 IEEE International Conference on Robotics and Automation, Seattle, May 26-30, 2015. Piscataway: IEEE, 2015: 1850-1857.
- [44] ROS G, RAMOS S, GRANADOS M, et al. Vision-based offline-online perception paradigm for autonomous driving [C]//Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision, Waikoloa, Jan 5-9, 2015. Washington: IEEE Computer Society, 2015: 231-238.
- [45] HARIHARAN B, ARBELAEZ P, BOURDEV L D, et al. Semantic contours from inverse detectors[C]//Proceedings of the 2011 IEEE International Conference on Computer Vision, Barcelona, Nov 6-13, 2011. Washington: IEEE Computer Society, 2011: 991-998.

- [46] EVERINGHAM M, ESLAMI S A, VAN G L, et al. The pascal visual object classes challenge: a retrospective[J]. International Journal on Computer Vision, 2014, 11(1): 98-136.
- [47] MOTTAGHI R, CHEN X, LIU X, et al. The role of context for object detection and semantic segmentation in the wild [C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, Jun 23-28, 2014. Washington: IEEE Computer Society, 2014: 891-898.
- [48] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: common objects in context[C]//LNCS 8693: Proceedings of the 13th European Conference on Computer Vision, Zurich, Sep 6-12, 2014. Berlin, Heidelberg: Springer, 2014: 740-755.
- [49] ZHOU B L, ZHAO H, PUIG X, et al. Scene parsing through ADE20K dataset[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Jul 21-26, 2017. Washington: IEEE Computer Society, 2017: 5122-5130.
- [50] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Jun 27-30, 2016. Washington: IEEE Computer Society, 2016: 3213-3223.
- [51] SONG S, LICHTENBERG S P, XIAO J. SUN RGB-D: a RGB-D scene understanding benchmark suite[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, Jun 7-12, 2015. Washington: IEEE Computer Society, 2015: 567-576.
- [52] GARCIA-GARCIA A, ORTOS-ESCOLANO S, OPREA S O, et al. A review on deep learning techniques applied to semantic segmentation[J]. arXiv:1704.06857, 2017.
- [53] LI H C, XIONG P F, FAN H Q, et al. DFANet: deep feature aggregation for real-time semantic segmentation[J]. arXiv: 1904.02216, 2019.
- [54] ARANI E, MARZBAN S, PATA A, et al. RGPNet: a real-time general purpose semantic segmentation[J]. arXiv:1912.01394, 2019.
- [55] CHARLES R Q, SU H, KAICHUN M, et al. PointNet:

deep learning on point sets for 3D classification and segmentation[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Jul 21-26, 2017. Washington: IEEE Computer Society, 2017: 77-85.

- [56] HU Q Y, YANG B, ROSA S, et al. RandLA-Net: efficient semantic segmentation of large-scale pointclouds[J]. arXiv: 1911.11236, 2019.

- [57] LU Y, CHEN Y Y, ZHAO D B, et al. Graph-FCN for image semantic segmentation[J]. arXiv:2001.00335, 2020.



**徐辉**(1997—),男,河南信阳人,硕士研究生, CCF 学生会员,主要研究方向为计算机视觉、图像处理。

**XU Hui**, born in 1997, M.S. candidate, student member of CCF. His research interests include computer vision and image processing.



**祝玉华**(1965—),女,博士,教授,主要研究方向为地图制图学与地理信息工程、科学计算可视化。

**ZHU Yuhua**, born in 1965, Ph.D., professor. Her research interests include cartography, geographic information engineering and scientific computing visualization.



**甄彤**(1964—),男,教授,主要研究方向为人工智能、智能控制。

**ZHEN Tong**, born in 1964, professor. His research interests include artificial intelligence and intelligent control.



**李智慧**(1975—),女,讲师,主要研究方向为智能检测、智能控制。

**LI Zhihui**, born in 1975, lecturer. Her research interests include intelligent detection and intelligent control.