

基于特征融合的道路场景语义分割综述

曾志远 高鹏举

华北水利水电大学信息工程学院

DOI:10.12238/bd.v6i2.3888

[摘要] 基于深度学习的图像语义分割是近些年来计算机领域的重要研究内容,是实现场景感知和场景理解的重要方法之一。在对近年来的道路场景语义分割方法归纳分析后,首先对道路场景语义分割的发展进行介绍;其次根据一些道路场景语义分割中,特征融合形式的不同,将基于特征融合的道路场景语义分割方法划分为基于编码器-解码器结构、基于空间金字塔结构、基于多路径形式和基于注意力机制来分别进行阐述;然后对一些常用的道路场景数据集进行介绍;最后对道路场景语义分割以后的研究方向进行了展望。

[关键词] 计算机视觉;深度学习;语义分割;场景理解;特征融合

中图分类号: TP391 **文献标识码:** A

Overview of semantic segmentation of road scene based on Feature Fusion

Zhiyuan Zeng Pengju Gao

School of information engineering, North China University of water resources and hydropower

[Abstract] Image semantic segmentation based on deep learning is an important research content in computer field in recent years. It is one of the important methods to realize scene perception and scene understanding. After summarizing and analyzing the semantic segmentation methods of road scenes in recent years, the development of semantic segmentation of road scenes is introduced firstly. Secondly, according to the different feature fusion forms in semantic segmentation of some road scenes, the semantic segmentation methods based on feature fusion path scene are divided into encoder-decoder structure, spatial pyramid structure, multiple path form and attention mechanism. Then some common road scene data sets are introduced. Finally, the future research direction of road scene semantic segmentation is prospected.

[Key words] computer vision; deep learning; semantic segmentation; scenario understanding; feature fusion

引言

随着深度学习的不断发展,图像语义分割在实际生活中的一些领域大放异彩,例如自动驾驶汽车、智慧安防、室内导航等领域。而针对道路场景的语义分割技术^[1,2]正是实现自动驾驶汽车对场景正确理解不可或缺的一环。道路场景语义分割是在像素级别上进行图像分类,是为获取到的车辆行驶道路场景图像中的每个像素分配一个相对应的类别,同一类别的对象不会进行区分^[3]。

Long等人于2015年提出全卷积神经网络(Fully Convolutional Networks for semantic segmentation, FCN),自此道路场景语义分割正式进入快速发展阶段。但是由于FCN在对道路场景进行语义分割过程中,对每个像素分配相应的类别时没有能够充分认识到像素和像素之间可能存在的关系;而且对道路场景图像中的一些细节特征敏感度不足,无法有效提取和利用更多的细节特征,从而导致分割的结果不够精细。针对这些问题,研

究者们提出了各种解决方法,如使用全连接条件随机场(Fully connected conditional random field,简称FCCRF),空洞卷积(Atrous convolution)等方法。而本文则主要对使用特征融合方式来解决此问题的道路场景语义分割方法进行分析和总结。

1 基于特征融合的道路场景语义分割方法

1.1 基于编码器-解码器结构

在通过使用编码器-解码器结构来进行特征融合的道路语义分割方法中,要数U-Net网络结构最为经典。其网络结构主要由编码器部分来提取图像的上下文信息以及解码器部分来恢复图像的空间细节特征。针对编码器和解码器中提取到的深层次特征和浅层次特征无法充分利用的问题,U-Net在编码器和解码器部分使用了跳跃结构,通过对编码器和解码器中的特征进行融合,实现了深层次特征和浅层次特征的有效融合和利用,进一步提高了模型的分割精细度。

虽然U-Net网络结构通过使用跳跃结构使得深层次的语义

特征能够和浅层次的空间细节特征进行特征融合,有效提升了分割精度,但是由于在编码器部分提取到的特征图在与解码器部分提取到的特征图进行特征融合时,尽管其特征图的尺寸和维度相同,但特征图中蕴含的语义特征却并不相同,无法真正保证特征图在进行特征融合时是真正的最佳的特征融合。为了有效缓解这种限制,Zongwei Zhou等人于2018年提出了U-Net++网络模型。与U-Net不同的是,U-Net++中使用的跳跃结构在解码器中需要与来自编码器的特征图进行特征融合的部分设置了不同尺寸包含不同语义信息的特征图,从而帮助在进行特征融合时可以选择以何种特征图进行有效的特征融合。

1.2 基于空间金字塔结构

不同于使用编码器-解码器结构来实现特征融合,空间金字塔池化结构则通过以不同的感受野来提取多个尺寸内的上下文信息,之后再对提取到的不同尺度的特征信息进行有效特征融合。PSPNet通过在其网络结构中使用金字塔池化模块(Pyramid pooling module),实现了对处在不同区域上的上下文特征信息的聚合,对模型能够有效获取足够的全局特征信息提供了有利帮助,并且通过合理有效的利用全局特征信息,进一步提高了模型对道路场景的分割精度。

为了在网络模型提取特征的过程中获得更加丰富的语义特征信息,Chen等人于2016年在VGG16的基础上提出和构建了DeepLab v1网络结构,DeepLab v1通过将基础卷积神经网络VGG6中的部分标准卷积替换为空洞卷积,以此来达到扩大感受野,获取更多丰富的语义信息。DeepLab v1中使用的多尺度预测方法虽然有助于提取和利用多尺度信息,但是其起到的具体效果并不明显,而且使用多尺度预测的方法速度也比较慢。针对这一现象,Chen等在DeepLab v1的基础上进行改进并提出了新的卷积神经网络—DeepLab V2。

DeepLabV2提出和设计了空洞空间金字塔池化模块(Atrous Spatial Pyramid Pooling,简称ASPP)。ASPP模块通过使用采样率分别为6,12,18,24的空洞卷积对输入的特征图进行并行卷积操作,相当于使用多个不同比例来提取图像的上下文信息,最后将使用不同采样率的空洞卷积提取到的特征图进行特征融合。DeepLabV2通过使用ASPP模块来代替多尺度预测方法,不仅加快了网络的计算速度,同时更进一步的提高了模型对图像上下文信息的有效提取。

DeepLabV2在利用ASPP模块实现多尺度信息提取的同时也存在着一些问题,ASPP模块每次进行卷积操作时只计算了部分像素点,采样不够密集,这样做的结果会导致丢失大量的特征细节信息。考虑到这个问题的出现,Maoke Yang等人于2018年将DenseNet中的密集连接的思想与ASPP模块相结合,提出了新的空间金字塔模块—密集空洞空间金字塔池化模块(Dense Atrous Spatial Pyramid Pooling,简称DenseASPP)。DenseASPP通过密集连接的方式使得每一层空洞卷积的输入特征图都是经过此卷积层前面所有卷积层的输出和输入特征图进行特征融合后的特征图。通过这样的方式,能够在扩大感受野的同时,增加

所计算像素点的数量,从而能够更好的提取更多的图像上下文特征信息。

1.3 基于多路径形式

在RefineNet网络结构中,最初输入的图片在经过一系列的卷积层处理后,输出分别是原图尺寸大小1/4、1/8、1/16和1/32的特征图,此后特征图进入到Residual Conv Unit模块中经过两个残差结构的处理提取到新的特征图,之后在Multi-resolution Fusion模块中对不同尺寸的特征图进行特征融合,最后进入到Chained Residual Pooling模块中,通过使用侧边分支上的一系列池化操作来提取背景信息,最后再通过一个残差模块来输出新的特征图。RefineNet结构中通过使用一系列的残差结构对深层次特征进行上采样处理,之后再与浅层次特征进行特征融合,以此来实现对图像上下文信息的有效利用。

Hengshuang Zhao等人于2017年提出并构建了一种多路径级联式的网络结构—ICNet。在不同分辨率的输入图像中,低分辨率的图像通过一系列卷积操作之后获得相对较为粗糙的语义特征图,之后通过级联特征融合模块(Cascade Feature Fusion Unit,简称CFF)和级联标签指导模块(Cascade Label Guidance)来帮助深层次特征选取有价值的中层次特征。通过将提取低分辨率图像特征信息的快速效率和提取高分辨率图像特征信息的高质量相结合,实现了ICNet对道路场景的实时语义分割。

1.4 基于注意力机制

应用在卷积神经网络中的注意力机制则往往是以额外的模块或者是神经网络的形式出现的,其能够有目的的选择输入信息的某些有价值的部分,或者是通过对输入的信息进行不同部分分配不同权重的方式来筛选出有价值的信息。

SENet针对特征图通道之间可能存在着的关系,提出和设计了SE模块(Squeeze and Excitation)。SE模块能够帮助神经网络模型有目的的去学习各个特征通道的重要程度并为之分配权重,并根据各个通道特征的重要程度来对有价值的特征信息增加其权重,对价值不大或者是存在的噪声信息降低其权重。

Sanghyun Woo等人于2018年提出了一种轻量型的注意力模块(Convolutional Block Attention Module,简称CBAM)。相比于SENet网络只使用通道注意力机制来筛选有价值的通道特征信息,CBAM模块则在关注特征图各通道之间的相互依赖关系的同时,又关注了空间各个位置之间可能存在着的关系。通过这种方法能够很好的筛选出对分割任务有价值的特征信息,从而提高模型的分割精度。

Jun Fu等人于2018年在CBAM的基础上提出了双重注意网络—DANet。DANet结构中的位置注意力模块(Position Attention Module)通过对特征图中处于不同空间位置的每个像素点的特征进行加权求和来有目的对特征进行特征融合,这样做后所有处在不同空间位置上的特征都会彼此之间存在关联,通过这种方法能够有效学习到各个特征之间的空间相关性。而DANet结构中的通道注意力模块通过对特征图的所有通道之间相应地特征进行聚合,达到有目的有选择地加强通道之间存在关联关系特征的重要程度,通过

这种通道注意力来学习和利用特征在通道之间地相关性。

2 道路场景数据集

常用的道路场景数据集如表1所示:分别为2009年的CamVid数据集,2013年的KITTI数据集,2016年的Cityscapes数据集,2018年的BDD100K数据集。

表1 常用道路场景数据集

数据集	年份	类别数目	数据总量	包含区域	环境情况
CamVid	2009	32	700	Europe	Day
KITTI	2013	10		Germany and America	Day
Cityscapes	2016	34	20000	Germany, Switzerland and France	Spring, summer, and autumn
BDD100K	2018	10	10000	Multiple cities around the world	Various scenes

3 展望

尽管近些年来各种道路场景语义分割方法不断被提出并取得了瞩目的成绩,但其在未来的很长时间内仍有着很大的发展空间。道路场景语义分割在未来的发展方向可能主要有以下几个方面:

(1)实时道路场景语义分割。目前大多数道路场景语义分割方法主要以提升语义分割精度为主要标准,在分割效率上关注度略显不足,而对于真正能用于无人驾驶技术的道路场景语义

分割方法,必须要求高质量的分割精度和高效的分割效率。所以针对实时性的道路场景语义分割方法^[4]必然会是未来的一个重要研究方向。

(2)三维道路场景语义分割。自动驾驶技术必须能够做到对道路场景上的真实物体做到快速的判别和划分,现实中真实的物体往往是三维的,而目前大多数道路场景语义分割都是对二维的道路场景图片进行语义分割。所以三维道路场景语义分割会是未来的一个重要研究方向。

[参考文献]

[1]商建东,刘艳青,高需.多尺度特征提取的道路场景语义分割[J].计算机应用与软件,2021,38(11):174-178.

[2]岳师怡.面向雾天道路场景的高精度语义分割算法[D].天津大学,2019.

[3]王龙飞,严春满.道路场景语义分割综述[J].激光与光电子学进展,2021,58(12):120.

[4]黄福蓉.用于实时道路场景的语义分割算法CBR-ENet[J].中国电子科学研究院学报,2021,16(03):277-284.

作者简介:

曾志远(1997--),男,汉族,河南省驻马店市泌阳县人,硕士研究生,学生,研究方向:深度学习、图像语义分割。

高鹏举(1999--),男,汉族,河南省周口市西华县人,硕士研究生,学生,研究方向:机器学习、大数据预测分析、图像语义分割。

中国知网数据库简介:

CNKI介绍

国家知识基础设施(National Knowledge Infrastructure, NKI)的概念由世界银行《1998年度世界发展报告》提出。1999年3月,以全面打通知识生产、传播、扩散与利用各环节信息通道,打造支持全国各行业知识创新、学习和应用的交流合作平台为总目标,王明亮提出建设中国知识基础设施工程(China National Knowledge Infrastructure, CNKI),并被列为清华大学重点项目。

CNKI 1.0

CNKI 1.0是在建成《中国知识资源总库》基础工程后,从文献信息服务转向知识服务的一个重要转型。CNKI 1.0目标是面向特定行业领域知识需求进行系统化和定制化知识组织,构建基于内容内在关联的“知网节”,并进行基于知识发现的知识元及其关联关系挖掘,代表了中国知网服务知识创新与知识学习、支持科学决策的产业战略发展方向。

CNKI 2.0

在CNKI 1.0基本建成以后,中国知网充分总结近五年行业知识服务的经验教训,以全面应用大数据与人工智能技术打造知识创新服务业为新起点,CNKI工程跨入了2.0时代。CNKI 2.0目标是将CNKI 1.0基于公共知识整合提供的知识服务,深化到与各行业机构知识创新的过程与结果相结合,通过更为精准、系统、完备的显性管理,以及嵌入工作与学习具体过程的隐性知识管理,提供面向问题的知识服务和激发群体智慧的协同研究平台。其重要标志是建成“世界知识大数据(WKBD)”、建成各单位充分利用“世界知识大数据”进行内外脑协同创新、协同学习的知识基础设施(NKI)、启动“百行知识创新服务工程”、全方位服务中国世界一流科技期刊建设及共建“双一流数字图书馆”。