



结合边缘检测的语义分割算法

王 因^{1,2}, 侯志强^{1,2}, 赵梦琦^{1,2}, 余旺盛³, 马素刚^{1,2}

(1.西安邮电大学 计算机学院,西安 710121; 2.西安邮电大学 陕西省网络数据分析与智能处理重点实验室,西安 710121;
3.空军工程大学 信息与导航学院,西安 710077)

摘 要: 针对语义分割中目标边缘模糊与分割不准确的问题,提出一种结合边缘检测的语义分割网络。整个网络由边缘检测网络和语义分割网络并行组成。利用边缘检测网络与语义分割网络分别提取图像的边缘特征和初步的语义分割特征,通过特征融合模块将边缘特征和语义分割特征进行融合,得到最终的语义分割结果。在 CamVid 数据集和 Cityscapes 数据集上的实验结果表明,与 SegNet 算法相比,该算法平均交并比分别提升了 1.5 和 1.8 个百分点,验证了所提算法的有效性。

关键词: 卷积神经网络;全卷积网络;语义分割;边缘检测;特征融合

开放科学(资源服务)标志码(OSID):



中文引用格式:王因,侯志强,赵梦琦,等.结合边缘检测的语义分割算法[J].计算机工程,2021,47(7):257-265.

英文引用格式:WANG N, HOU Z Q, ZHAO M Q, et al. Semantic segmentation algorithm combined with edge detection[J]. Computer Engineering, 2021, 47(7): 257-265.

Semantic Segmentation Algorithm Combined with Edge Detection

WANG Nan^{1,2}, HOU Zhiqiang^{1,2}, ZHAO Mengqi^{1,2}, YU Wangsheng³, MA Sugang^{1,2}

(1.College of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121, China;
2.Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an University of Posts and Telecommunications, Xi'an 710121, China; 3.Information and Navigation Institute, Air Force Engineering University, Xi'an 710077, China)

[Abstract] To address the fuzzy and inaccurate target edges in semantic segmentation, a semantic segmentation network combined with edge detection is proposed. The network consists of an edge detection network and a semantic segmentation network. The edge detection network is used to extract the edge features of the image, while the semantic segmentation network is used to extract the preliminary semantic segmentation features. The edge features and the semantic segmentation features are fused through the feature fusion module to obtain the final semantic segmentation result. The experimental results show that compared with the SegNet algorithm, the proposed algorithm improves the mIoU by 1.5 percentage points on the CamVid dataset and 1.8 percentage points on the Cityscapes dataset. The effectiveness of the proposed algorithm is verified.

[Key words] convolutional neural network; Fully Convolutional Network (FCN); semantic segmentation; edge detection; feature fusion

DOI: 10.19678/j.issn.1000-3428.0058167

0 概述

图像分割是将图像中有意义或感兴趣的目标提取出来,用于图像后续处理。传统的图像分割方法根据图像的低层特征信息,例如颜色、纹理、亮度、形状等提取出感兴趣的区域。但在某些复杂情况下,传统方法提取的特征识别度低,分割性能也较差。

因此,许多研究者尝试将图像的高级语义信息引入到传统的分割过程中,从而得到图像的语义分割^[1]。图像语义分割^[2]是计算机视觉中一个重要的研究内容,并且广泛应用于自动驾驶、场景识别、无人机应用等领域^[3-5]。

传统的图像分割方法常见的有 N-cut 算法^[6]、分水岭算法^[7]、SLIC 算法^[8]等。N-cut 算法^[6]通过计算

基金项目:国家自然科学基金(61473309,61703423)。

作者简介:王 因(1996—),女,硕士研究生,主研方向为计算机视觉、图像分割;侯志强,教授;赵梦琦,硕士研究生;余旺盛,讲师;马素刚,博士研究生。

收稿日期:2020-04-24 修回日期:2020-06-03 E-mail:wnan1020@163.com

像素与像素之间的关系得到权重参数对图像进行划分,但这种图划分方法需要对一张图像进行多次划分,并且无法对图像中颜色纹理等较为相似的部分进行分割,导致分割结果不佳;分水岭算法^[7]是一种典型的区域分割方法,虽然实现简单,并且能够很好地提取图像轮廓信息,但容易产生过分割问题;SLIC算法^[8]不仅可以分割彩色图像,而且能够分割灰度图像,它使用像素之间特征的相似性对像素进行分组,并用少量的超像素代替大量的像素来表达图像特征,从而会降低后续图像处理的复杂性,但通常作为其他分割算法的预处理步骤,并且不容易预先确定超像素的个数。

近年来,基于深度学习的图像语义分割方法^[9]在语义分割领域中表现出色,分割效果与传统的方法相比有了明显提高。尤其是全卷积网络(Fully Convolution Network, FCN)^[10]有效地提升了语义分割的精度,因此出现了很多基于FCN的改进算法。BADRINARAYANAN等^[11]提出了SegNet网络用于语义分割,SegNet网络是一种编解码结构的分割网络,具有更高的分割精度及更快的分割速度,但这种高效的处理会使图像丢失许多高频细节信息,导致目标边缘模糊以及分割结果中的细节丢失;PASZKE等^[12]提出的ENet算法是基于改进SegNet而来的,虽然提供了较高的分割速度,但同时也影响了分割精度,导致分割精度低;YU等^[13]提出了空洞卷积的思想,在不使用池化操作的情况下会使感受野指数变大,获得全局特征信息,但随之运算量也增大;ZHAO等^[14]提出的金字塔池化模块对不同区域的上下文进行聚合,提升了网络利用全局上下文信息的能力;CHEN等^[15]提出了空洞金字塔池化(Atrous Spatial Pyramid Pooling, ASPP)模块,可以增大空间分辨率,同时不改变感受野效果;CHEN等^[16]提出了DeepLab-v3结构,在ASPP模块中加入了批量归一化(Batch Normalization, BN)层,设计了串行和并行的空洞卷积模块,采用多种不同的空洞率来获取多尺度的特征信息,并将所有特征融合一起;之后,CHEN等^[17]提出DeepLab-v3+结构,在DeepLab-v3基础上增加一个解码器模块,构成一个编解码结构的语义分割模型。综上所述,现有方法大多使用空洞卷积、特征融合等方法来提升语义分割效果,但这类方法需要高分辨率的特征图,计算量较大,而且容易丢失边缘细节信息。

针对边缘细节信息丢失的问题,CHEN等^[18]提出通过条件随机场(Conditional Random Field, CRF)对FCN得到的分割结果进行细节上的优化,但这种处理方法只是利用颜色信息和纹理信息等低层特征来修正分割结果,无法学习目标边缘与输入图像之间的对应关系。HUANG等^[19]在原有的语义分割FCN结构上构建了边界检测分支,输入图像先进入语义分割模块,从分割模块中构建检测分支来获取

目标形状细节信息,但会导致学习的边缘信息不完整,影响语义分割模块的分割效果。

为进一步解决分割中边缘模糊与分割不准确的问题,可考虑结合边缘检测的方法进行语义分割。边缘检测可以得到局部像素灰度的突变,获得封闭或者开放的边缘,提供丰富的低层边缘信息,从而弥补语义分割中的边缘信息丢失。早期边缘检测方法使用Sobel边缘检测算子^[20]、Canny边缘检测算子^[21]等方法检测边缘。随着深度学习的发展,很多方法采用卷积神经网络进行边缘检测,如N⁴-Fields^[22]、Deep Edge^[23]、Deep Contour^[24]等方法。XIE等^[25]提出了整体嵌套边缘检测(Holistically-Nested Edge Detection, HED)算法进行边缘检测,该网络实现了第1个端到端的边缘检测模型;LIU等^[26]提出了基于更丰富特征的边缘检测(Richer Convolutional Features for Edge Detection, RCF),RCF首先将网络中每个阶段的特征进行了元素相加操作,然后进行特征融合,RCF利用了网络中所有卷积层的特征,是目前较好的边缘检测算法。近年来,也出现了一些结合边缘检测的语义分割方法^[27-28],但该方法目前尚不多见。

综上所述,本文提出了一种结合边缘检测网络的语义分割算法。在语义分割网络基础上,并联了边缘检测网络,构建一种结合边缘检测的语义分割模型,利用边缘特征改善语义分割中的目标边缘模糊问题,为语义分割提供丰富准确的边缘信息。设计一个特征融合模块,将边缘特征和语义分割特征进行concat融合,并对融合后的特征进行多层卷积操作,最终在CamVid数据集和Cityscapes数据集上进行实验。

1 本文算法

本文提出一种结合边缘信息的语义分割算法,以此来改善语义分割中边缘模糊、分割不准确等问题,提升语义分割精度。

1.1 网络整体结构

本文所提出的结合边缘检测的语义分割模型采用一种并行结构,如图1所示,主要由边缘检测子网络、语义分割子网络和特征融合模块3个部分组成(彩图效果见《计算机工程》官网HTML版)。图1中的虚线框为边缘检测子网络,该网络能捕获和学习图像的边缘信息,有助于获取目标更多的细节信息,从而可以为语义分割提供更精细的边缘信息。图1中的点线框为语义分割子网络,该网络对输入图像进行下采样,再进行上采样来提取图像的区域特征,得到初步的语义分割特征,但是大量的细节信息在下采样时被丢失。模型的后端是特征融合模块,即图1中点划线框所包含的部分,将边缘特征和语义分割特征进行卷积操作融合,利用边缘信息更强的边缘特征来改善语义分割中边缘模糊、分割不准确等问题。

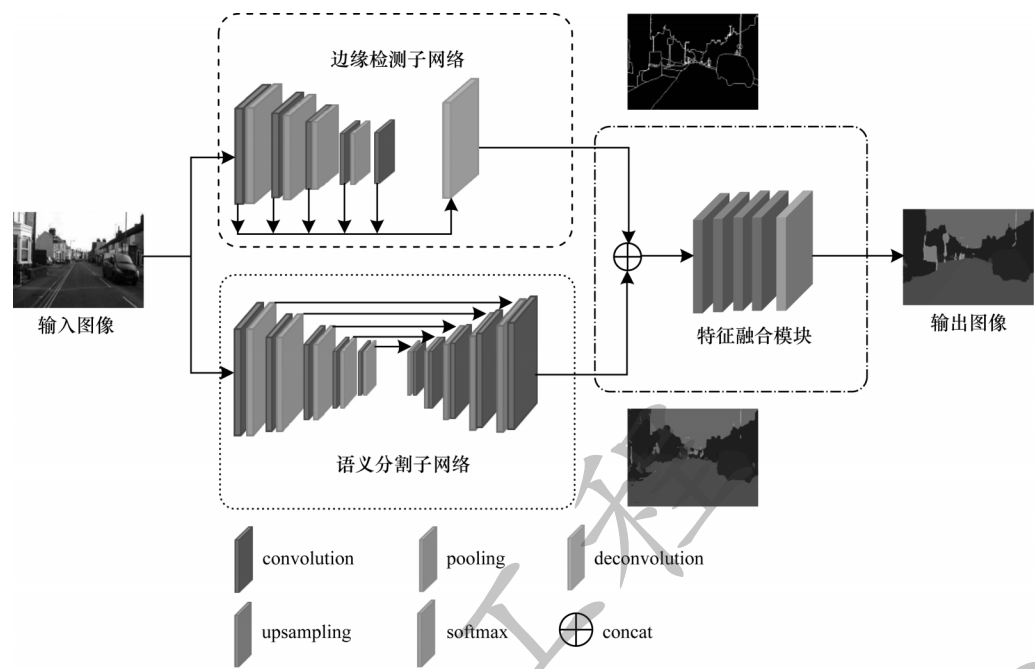


图 1 语义分割算法模型结构

Fig.1 Model structure of semantic segmentation algorithm

1.2 边缘检测子网络

边缘检测子网络的详细网络结构如图 2 所示。在边缘检测子网络^[26]中,整个网络的卷积层分为 5 个阶段,每个阶段由卷积层和池化层组成,然后去掉第 5 阶

段的池化层,最后将第 5 阶段的卷积层使用空洞率为 2 的空洞卷积^[12],空洞卷积的使用有效地避免了池化操作后对高层特征信息的丢失问题,更好地捕获了目标的轮廓特征,使得高层特征中的边缘更加清晰和完整。

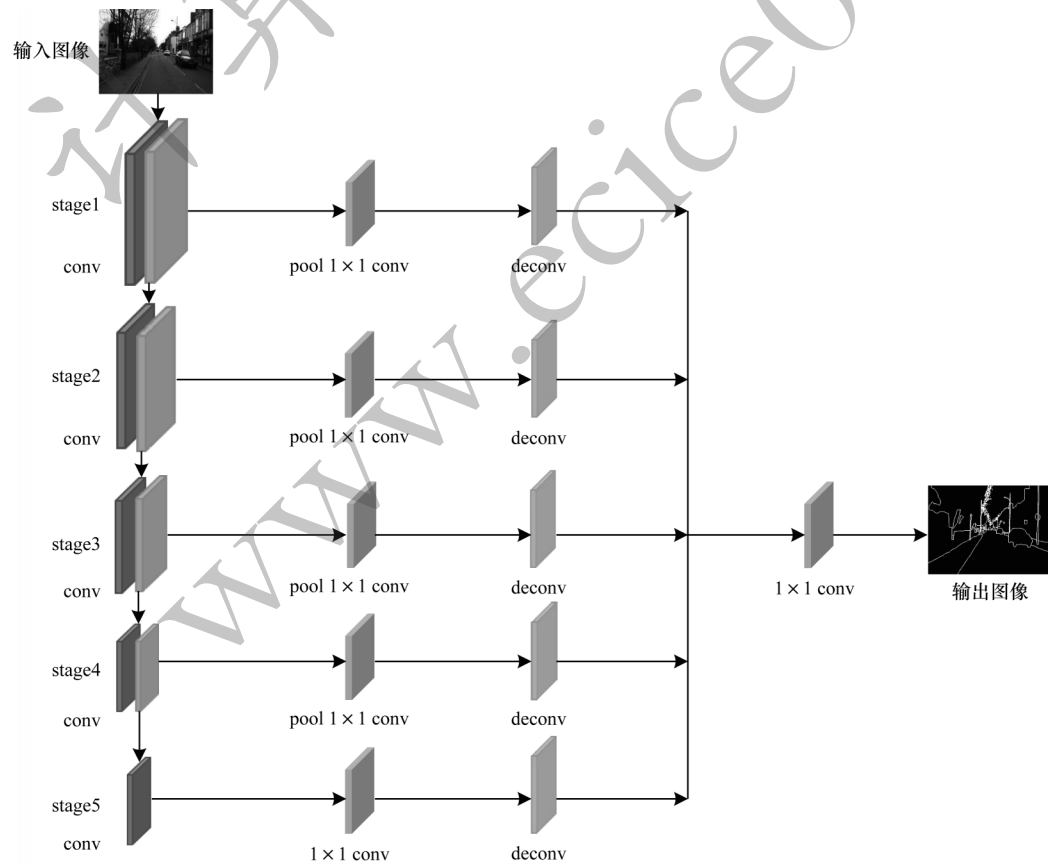


图 2 边缘检测子网络的结构

Fig.2 Structure of edge detection sub-network

5 个阶段的结构如表 1 所示。

表 1 边缘检测子网络 5 个阶段的结构
Table 1 Structure of edge detection sub-network
five-stage

层名	核大小及相应操作
stage1(conv1+relu1)	[3,3,64]
stage1(pool1)	[2,2,max]
stage2(conv2+relu2)	[3,3,128]
stage2(pool2)	[2,2,max]
stage3(conv3+relu3)	[3,3,256]
stage3(pool3)	[2,2,max]
stage4(conv4+relu4)	[3,3,512]
stage4(pool4)	[2,2,max]
stage5(dila+conv5+relu5)	[2,3,3,512]

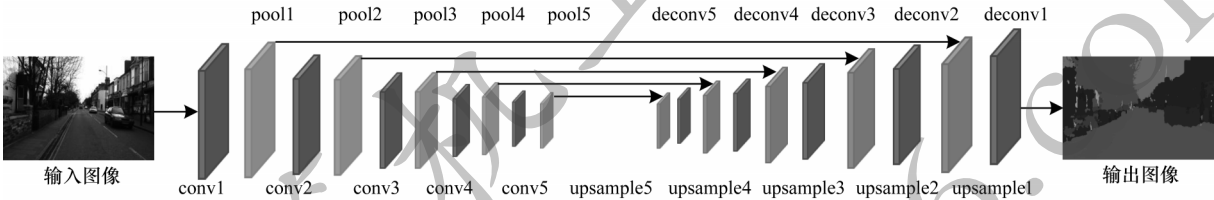


图 3 语义分割子网络的结构

Fig.3 Structure of semantic segmentation sub-network

表 2 编码阶段的网络结构
Table 2 Network structure of encoding phase

层名	核大小及相应操作
conv1(conv+BN+relu)	[3,3,64]
pool1	保留池化索引 pool1_indices
conv2(conv+BN+relu)	[3,3,128]
pool2	保留池化索引 pool2_indices
conv3(conv+BN+relu)	[3,3,256]
pool3	保留池化索引 pool3_indices
conv4(conv+BN+relu)	[3,3,512]
pool4	保留池化索引 pool4_indices
conv5(conv+BN+relu)	[3,3,512]
pool5	保留池化索引 pool5_indices

解码器阶段是由对应于每个编码器的解码器组成的,解码器首先使用从相应的编码器存储的最大池化索引来获取目标信息与大致的位置信息,并对缩小后的特征图进行上采样,这种上采样方式可以减少训练的参数数量以及减小池化对信息的损失。然后对上采样后的图像进行卷积处理,弥补编码器阶段池化层对目标造成的细节损失。解码器各层的网络结构如表 3 所示。

在每个阶段的卷积层之后连接一个核大小为 1×1 卷积层,然后通过反卷积将特征图进行上采样到原图像大小分辨率,最后利用一个 1×1 的卷积层将 5 个特征图融合。该网络能捕获和学习图像的边缘信息,可以为语义分割提供更精细的边缘信息。

1.3 语义分割子网络

在语义分割子网络中,其结构采用的是经典的编码器和解码器结构^[11]。语义分割子网络的详细网络结构如图 3 所示。编码阶段沿用 VGG16 网络模型,由一连串的卷积层、池化层和 BN 层组成。卷积层负责获取图像特征,池化层对图像进行下采样将特征传送到下一层,因为在最大池化过程中会损失一些信息,这里会存储最大池化索引,保存最大池化过程中的位置信息,用于在上采样阶段进行恢复,BN 层主要对图像的分布进行归一化,加速学习。编码器阶段主要对图像进行特征提取。编码器各层的网络结构如表 2 所示。

表 3 解码阶段的网络结构

Table 3 Network structure of decoding phase

层名	核大小及相应操作
upsample5	使用池化索引 pool5_indices
deconv5(conv+BN+relu)	[3,3,512]
upsample4	使用池化索引 pool4_indices
deconv4(conv+BN+relu)	[3,3,512]
upsample3	使用池化索引 pool3_indices
deconv3(conv+BN+relu)	[3,3,256]
upsample2	使用池化索引 pool2_indices
deconv2(conv+BN+relu)	[3,3,128]
upsample1	使用池化索引 pool1_indices
deconv1(conv+BN+relu)	[3,3,64]

1.4 特征融合模块

在本文模型中,当从边缘检测子网络获得边缘信息和语义分割子网络获得语义信息时,需要对 2 路特征进行融合。因为边缘检测子网络和语义分割子网络得到的特征表示不同,边缘检测子网络得到的结果更多地表示图像边缘和细节特征,语义分割子网络得到的结果更多的地示图像的区域特征,所以在进行 2 路特征融合时,简单地将 2 路特征进行加权融合,并不能充分地利用 2 路特征,达到预期效果。因此,本文提出了一个特征融合模块来融合边缘检测子网络特征和语义分割子网络特征。

在选择特征融合的方式上,主要有 add 融合方式和 concat 融合方式 2 种^[29]。add 融合方式对上下 2 路特征图对应位置元素的值进行相加,但这种方式在融合过程中会损失原始特征的信息,不能体现特征之间的互补性,所得到的结果也并不理想;concat 融合方式对上下 2 路特征直接级联融合,将不同通道进行合并,这种方式不仅避免了 add 融合方式对特征信息造成的损失,而且对上下 2 路的特征进行通道

数的合并,并结合前面所得到的特征得到后续的特征。

本文设计的特征融合模块包括使用 concat 特征融合方式对边缘特征和语义分割特征进行初步融合 2 个部分;然后对初步融合后的特征进行多层卷积操作,多层卷积由 4 层卷积层组成,其特征图的个数分别为 64、128、256 和 512,卷积核的大小为 3×3 。特征融合模块的详细结构如图 4 所示。

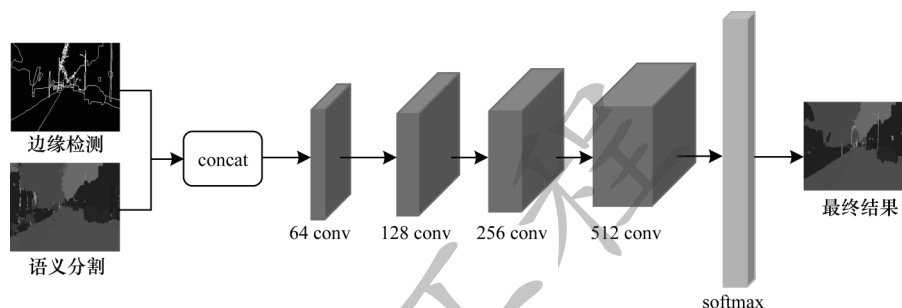


图4 特征融合模块

Fig.4 Feature fusion module

具体的融合过程为:首先将边缘检测子网络和语义分割子网络输出的特征进行 concat 方式融合;然后进行多层卷积操作,利用卷积去学习融合 2 路特征;最后再进行分类,得到最终的分割图。

在对特征融合方式选择上,本文进行实验比较确定选择 concat 特征融合方式。首先将得到的边缘特征和语义分割特征分别使用 add 方式和 concat 方式进行初步融合;然后对融合后的特征进行多层卷积操作;最后进行分类得到最终分割结果。实验结果表明,采用 add 融合方式和 concat 融合方式对融合结果有不同影响,如图 5 所示。

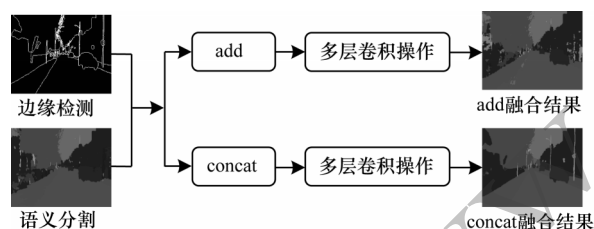


图5 不同融合方式对分割结果的影响

Fig.5 Impact of different fusion methods on segmentation results

不融合、add 融合和 concat 融合方式的平均交并比(mIoU)分别为 55.6%、55.7% 和 57.1%。

1.5 算法步骤

本文针对分割任务中目标边缘模糊、分割不准确等问题,提出一种结合边缘检测的语义分割算法。考虑边缘检测与语义分割的结合,在语义分割网络基础上并联一个边缘检测网络来学习目标的边缘信息,再将边缘特征和语义分割特征进行融合,综合利用 2 路数据的特征信息,得到最终的语义分割结果。

本文算法的主要步骤如下:

步骤 1 输入图像。

步骤 2 将图像分别送入边缘检测子网络和语义分割子网络,得到图像边缘的特征和图像初步的语义分割特征。

步骤 3 把得到的 2 类特征按 concat 方式融合。

步骤 4 将初步融合后的特征进行多层卷积操作,利用卷积融合 2 个子网络的特征。

步骤 5 将步骤 4 得到的特征融合结果进行 softmax 分类,得到最终的语义分割图像。

本文算法流程如图 6 所示。从图 6 可以看出,本文算法改善了语义分割中边缘不清晰、分割任务中目标边缘模糊等问题,提高了分割精度。

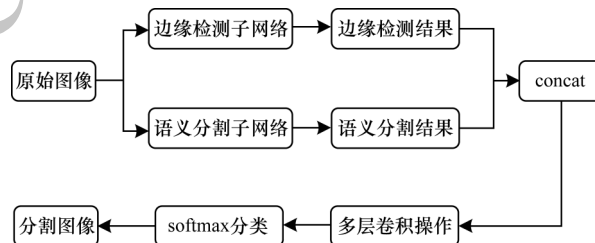


图6 本文算法流程

Fig.6 Algorithm procedure of this paper

2 实验

本文算法实验基于 Caffe 框架,系统为 Linux16.04,使用的 GPU 是 NVIDIA GTX 1080Ti,并在 CamVid 数据集和 Cityscapes 数据集上进行了实验。

2.1 数据集及评价指标

CamVid 数据集是从驾驶汽车的角度来看的一个街景数据集。它总共包含 701 张图像,其中 367 张用于训练,101 张用于验证,233 张用于测试。图像

分辨率为960像素×720像素,数据集包含11个语义类别。在SegNet^[10]中将其进行处理,将图像分辨率改为480像素×360像素,所以使用修改后的数据集验证模型的有效性。Cityscapes是另一个大型城市街道场景数据集,在语义分割领域中广泛使用。它有5 000张具有高质量精细标注的图像和20 000张具有粗略标注的图像。在本文的实验中,只使用具有精细标注的图像,它包含2 975张用于训练的精细注释图像、1 525张用于测试的图像和500张用于验证的图像。为了实验方便,将图像分辨率由2 018像素×1 024像素改为480像素×360像素,其中每个像素都被注释为预定义的19个类。

本文采用的评价指标为平均交并比(mean Intersection over Union, mIoU)和每秒处理帧数(Frames Per Second, FPS),这2种评价指标为当前语义分割中使用较多的标准度量。

mIoU计算2个集合为真实值和预测值的交集和并集之比,用于评价算法精度。IoU是每一个类别的交集与并集之比,而mIoU则是所有类别的平均IoU。其计算公式如下:

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

其中: k 是前景对象的个数; p_{ij} 是指原本属于第 i 类却被分类到第 j 类的像素的数量。

FPS用于评价算法速度,其计算公式如下:

$$F_{FPS} = \frac{N}{\sum_j T_j} \quad (2)$$

其中: N 为图像数量; T_j 为算法处理第 j 张图像的时间。

2.2 实验方法

本文提出的结合边缘信息的语义分割模型使用

端到端的联合训练方式。在模型训练过程中,使用SGD优化器,初始学习率设置为0.001,将batchsize设置为2, momentum设置为0.9, weight_decay设置为0.000 5,最大迭代次数为100 000。在训练过程中,本模型涉及边缘检测和语义分割2种分类网络,因此采用2种损失函数。

边缘检测网络是一个对像素点进行是否为边界的二分类问题。边缘检测网络的损失函数采用常用的交叉熵损失函数,其可以定义为:

$$l_{edge}^{loss} = -\frac{1}{n} \sum y_{edgelab} \lg y_{edgepre} \quad (3)$$

其中: $y_{edgelab}$ 代表真实值; $y_{edgepre}$ 代表预测值; n 代表图像的总像素点数量。

语义分割网络通常将图像的每个像素点的分类看作是一个多分类问题,因此同样采用交叉熵损失函数,可以定义为:

$$l_{seg}^{loss} = -\frac{1}{n} \sum y_{seglab} \lg y_{segpre} \quad (4)$$

其中: y_{seglab} 代表真实值; y_{segpre} 代表预测值; n 代表图像的总像素点数量。

在实验中,因为涉及语义分割和边缘检测2种分类网络,除了需要已有的语义分割标注外,还需要数据集的边缘标注图像,本文并不需要额外地进行边缘标注,而是利用已有的语义分割标注生成边缘标注图像。

2.3 实验结果与分析

2.3.1 CamVid数据集

基于CamVid数据集,采用未结合边缘检测的语义分割算法SegNet^[11]、ENet^[12]与本文算法进行对比,分析算法的性能表现。图7所示为本文算法与其他对比算法的实验结果。

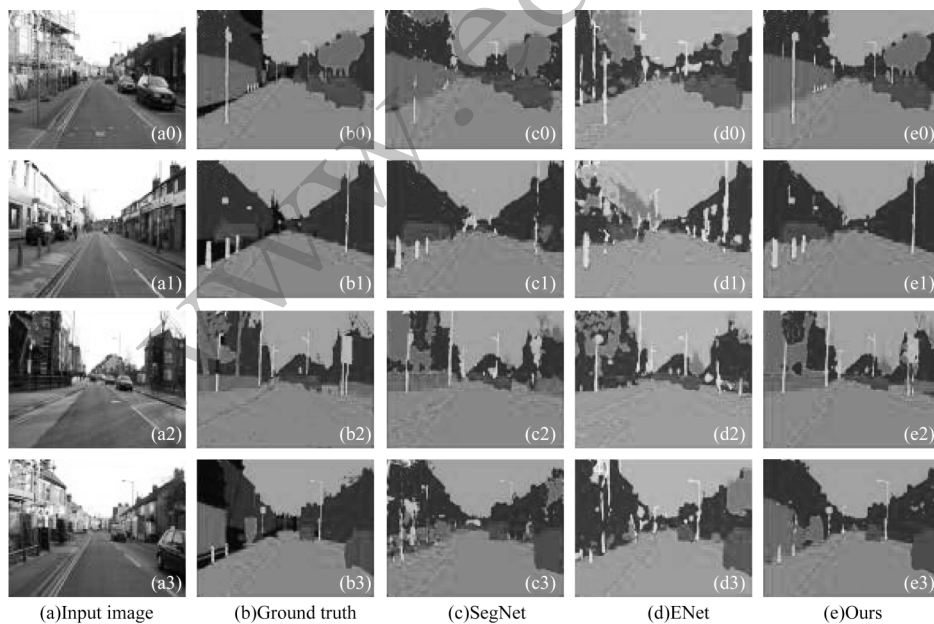


图7 不同算法分割结果对比1

Fig.7 Comparison of different algorithms segmentation results1

从图7可以看出,本文算法在道路、车辆、路灯杆、指示牌等的边界分割更为精准。在图7(a0)中,本文算法对路灯杆的分割更加连续和清晰;在图7(a1)中,本文算法可以清晰地分割出车辆旁边的行人以及远处车辆之间的粘连;在图7(a2)、图7(a3)中,不仅在道路、路灯杆处边缘更连续,更清晰,而且也较完整地分割出路边指示牌的边缘。

表4是本文方法与其他语义分割方法在相同的实验环境下的分割效果对比。从表4可以看出,本文算法在分割精度上高于其他2种算法,在分割速度上,本文算法相较于SegNet^[11]算法没有明显下降,虽然ENet^[12]算法的分割速度要优于本文算法,但分割精度远低于本文算法。

表4 不同算法在CamVid数据集上的对比

Table 4 Comparison of different algorithms on CamVid datasets

算法	基础模型	mIoU/%	算法速度/(frames·s ⁻¹)
SegNet ^[11] 算法	VGG16	55.6	5.8
ENet ^[12] 算法	From Scratch	51.3	14.3
本文算法	VGG16	57.1	4.8

2.3.2 Cityscapes数据集

基于Cityscapes数据集,同样采用未结合边缘检测的语义分割算法SegNet^[11]、ENet^[12]与本文的算法进行对比,分析算法的性能表现。图8所示为本文算法与其他对比算法的实验结果。

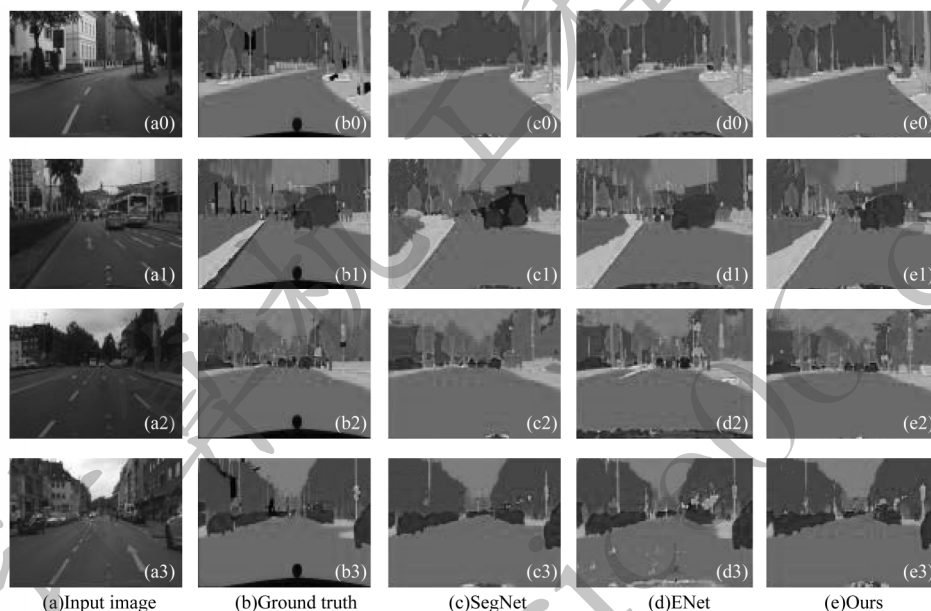


图8 不同算法分割结果对比2

Fig.8 Comparison of different algorithms segmentation results 2

从图8可以看出,在图8(a0)中,本文算法结果可以更好地将路灯杆等细小目标分割出来;在图8(a1)中,SegNet^[11]的分割结果图8(c1)将大巴车分割错误,而本文算法分割正确,虽然ENet^[12]的结果图8(c2)也分割正确,但本文分割结果相比于ENet^[12],道路分割更加平滑清晰;在图8(a2)中,本文算法结果相比于SegNet^[11]、ENet^[12],可以明显地分割出来交通灯,改善了路边行人粘连的问题;在图8(a3)中,SegNet^[11]、ENet^[12]不能清晰地分割出路中间的行人,而本文算法可以很好地分割出来。

表5是本文算法与其他语义分割算法在相同的实验环境下的分割效果对比。从表5可以看出,本文算法在分割精度上高于其他算法,在分割速度上,本文算法相较于SegNet^[11]算法没有明显下降,虽然ENet^[12]算法的分割速度要优于本文算法,但本文算法分割精度要高于ENet^[12]算法。

表5 不同算法在Cityscapes数据集上的对比

Table 5 Comparison of different algorithms on Cityscapes datasets

算法	基础模型	mIoU/%	算法速度/(frames·s ⁻¹)
SegNet ^[11] 算法	VGG16	56.1	5.8
ENet ^[12] 算法	From Scratch	57.6	14.3
本文算法	VGG16	57.9	4.8

通过对以上实验结果的分析,证明了该方法在分割精度方面的有效性,并且与未考虑边缘检测的SegNet^[11]算法相比,本文算法在CamVid数据集和Cityscapes数据集上的性能分别提升了1.5和1.8个百分点。

2.3.3 与现有算法的比较与分析

本文算法采用结合边缘检测的方法改进语义分割中边缘模糊问题,与已有的边缘检测与语义分割相结合的ESNet^[27]算法比较结果如表6所示。

表6 与ESNet算法的比较结果

Table 6 Comparison result with ESNet algorithm

算法	输入尺寸/像素	mIoU/%	算法速度/(frame·s ⁻¹)
ESNet ^[27] 算法	1 024×512	60.1	53.0
本文算法	480×360	57.9	4.8

本文算法与已有的边缘检测与语义分割相结合的方法主要有以下不同:

1)网络结构不同。本文采用的语义分割网络为SegNet^[11]网络,ESNet^[27]采用的语义分割网络为ESPNetV2^[30]网络,由于选择的语义分割baseline的差异,ESPNetV2的分割效果比SegNet的分割效果要好;本文采用的边缘检测网络为已有的边缘检测算法RCF^[26],而ESNet采用MobileNetV2^[31]作为边缘骨架,虽然MobileNetV2作为轻量级网络,在速度上占有优势,但本文选用的边缘检测网络结构简单,使用了结构中所有卷积层的特征,利用了更多的有用信息,可以提供更好的边缘效果。

2)输入图像分辨率不同。在ESNet方法中,输入图像分辨率为1 024像素×512像素;而本文算法由于baseline的限制,输入图像分辨率为480像素×360像素,不同的图像分辨率会带来不同的分割结果,分辨率越高,分割结果越好。

3)实验环境上的不同。本文实验使用的显卡是1台NVIDIA GTX 1080Ti,而ESNet使用了4台NVIDIA Tesla P40显卡,可以得到更大更稳定的性能。由此,本文算法精度与ESNet方法相比略低。

通过以上的比较与分析,验证了本文算法的有效性。

3 结束语

本文提出一种结合边缘检测的语义分割算法。整个网络由边缘检测网络和语义分割网络并行组成,边缘检测网络用来提取图像的边缘特征,语义分割网络用来提取初步的语义分割特征,后端将2路特征进行融合,以获得最终的语义分割结果。实验结果表明,结合边缘检测的语义分割算法可以有效改善边缘模糊、分割不准确等问题,能够得到较理想的语义分割结果。下一步将研究结合边缘检测的语义分割方法来提高分割精度并加快分割速度。

参考文献

- [1] 魏云超,赵耀. 基于DCNN的图像语义分割综述[J]. 北京交通大学学报,2016,40(4):82-91.
WEI Y C, ZHAO Y. A review on image semantic segmentation based on DCNN[J]. Journal of Beijing Jiaotong University, 2016, 40(4):82-91. (in Chinese)
- [2] 易盟,隋立春. 基于改进全卷积神经网络的航拍图像语义分类方法[J]. 计算机工程, 2017, 43(10): 216-221.
- [3] YI M, SUI L C. Aerial image semantic classification method based on improved full convolution neural network[J]. Computer Engineering, 2017, 43(10): 216-221. (in Chinese)
- [4] SIAM M, ELKERDAWY S, JAGERSAND M, et al. Deep semantic segmentation for automated driving: taxonomy, roadmap and challenges[C]//Proceedings of the 20th IEEE International Conference on Intelligent Transportation Systems. Washington D. C., USA: IEEE Press, 2017: 1-8.
- [5] 曾碧,黄文. 一种融合多特征聚类集成的室内点云分割方法[J]. 计算机工程, 2018, 44(3): 281-286.
ZENG B, HUANG W. An indoor point cloud segmentation method combining multi-feature clustering integration[J]. Computer Engineering, 2018, 44(3): 281-286. (in Chinese)
- [6] 于旭宅,王瑞瑞,陈伟杰. 改进分水岭算法在无人机遥感影像树冠分割中的应用[J]. 福建农林大学学报(自然科学版), 2018, 47(4): 428-434.
YU X Z, WANG R R, CHEN W J. Forest canopy segmentation of UAV remote sensing images using improved watershed algorithm[J]. Journal of Fujian Agriculture and Forestry University (Natural Science Edition), 2018, 47(4): 428-434. (in Chinese)
- [7] SHI J, MALIK J. Normalized cuts and image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 888-905.
- [8] VINCENT L, SOILLE P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1991, 13(6): 583-598.
- [9] ACHANTA R, SHAJI A, SMITH K, et al. SLIC superpixels compared to state-of-the-art superpixel methods[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2274-2282.
- [10] 罗嗣卿,张志超,岳琪. 基于改进SEGNET模型的图像语义分割[J]. 计算机工程, 2021, 47(4): 256-261.
LUO Siqing, ZHANG Zhichao, YUE Qi. Semantic image segmentation based on improved SEGNET model[J]. Computer Engineering, 2021, 47(4): 256-261. (in Chinese)
- [11] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015: 3431-3440.
- [12] 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.
- [13] PASZKE A, CHAURASIA A, KIM S, et al. ENet: a deep neural network architecture for real-time semantic segmentation[EB/OL]. [2020-03-20]. <https://arxiv.org/abs/1606.02147>.
- [14] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions [EB/OL]. [2020-03-20]. <https://arxiv.org/abs/1511.07122>.

- [14] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2017: 2881-2890.
- [15] 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.
- [16] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. [2020-03-20]. <https://arxiv.org/abs/1706.05587>.
- [17] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]// Proceedings of IEEE ECCV'18. Washington D. C. , USA: IEEE Press, 2018: 801-808.
- [18] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[EB/OL]. [2020-03-20]. <https://arxiv.org/abs/1412.7062>.
- [19] HUANG Q, XIA C, ZHENG W, et al. Object boundary guided semantic segmentation [C]//Proceedings of IEEE ACCV'16. Washington D. C. , USA: IEEE Press, 2016: 197-212.
- [20] KITTLER J. On the accuracy of the Sobel edge detector [J]. Image and Vision Computing, 1983, 1(1): 37-42.
- [21] CANNY J. A computational approach to edge detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, 8(6): 679-698.
- [22] GANIN Y, LEMPITSKY V. N²-Fields: neural network nearest neighbor fields for image transforms [EB/OL]. [2020-03-20]. <https://arxiv.org/abs/1406.6558>.
- [23] BERTASIUS G, SHI J, TORRESANI L. DeepEdge: a multi-scale bifurcated deep network for top-down contour detection [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2015: 4380-4389.
- [24] SHEN W, WANG X, WANG Y, et al. DeepContour: a deep convolutional feature learned by positive-sharing loss for contour detection [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2015: 3982-3991.
- [25] XIE S, TU Z. Holistically-nested edge detection [C]// Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2015: 1395-1403.
- [26] LIU Y, CHENG M M, HU X, et al. Richer convolutional features for edge detection [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2017: 3000-3009.
- [27] LYU H, FU H, HU X, et al. ESNet: edge-based segmentation network for real-time semantic segmentation in traffic scenes [C]//Proceedings of IEEE International Conference on Image Processing. Washington D. C. , USA: IEEE Press, 2019: 1855-1859.
- [28] 杨威, 张建林, 徐智勇, 等. 一种改进的 FocalLoss 在语义分割上的应用 [J]. 半导体光电, 2019, 40(4): 555-559.
- [29] YANG E, ZHANG J L, XU Z Y, et al. An improved Focal Loss function for semantic segmentation [J]. Semiconductor Optoelectronics, 2019, 40(4): 555-559. (in Chinese)
- [30] CAO G, XIE X, YANG W, et al. Feature-fused SSD: fast detection for small objects [C]//Proceedings of the 9th International Conference on Graphic and Image Processing. Washington D. C. , USA: IEEE Press, 2018: 106-151.
- [31] SACHIN M. ESPNetv2: Light-weight, efficient power, and general purpose convolutional neural network [EB/OL]. [2020-03-20]. <https://arxiv.org/abs/1811.11431>.
- [32] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks [EB/OL]. [2020-03-20]. <https://arxiv.org/abs/1801.04381>.