基于全卷积神经网络的非对称并行语义分割模型

李宝奇,贺昱曜,何灵蛟,强 伟
(西北工业大学航海学院,陕西西安 710072)

摘 要: 针对RGB图像具有丰富的色彩细节特征,红外图像对目标轮廓、尺寸、边界等外形特征有较高敏感度的特点,提出了一种非对称并行语义分割模型APFCN(Asymmetric Parallelism Fully Convolutional Networks). APFCN上路设计了一个卷积核尺寸非统一的五层空洞卷积网络来提取红外图像目标高层轮廓特征;下路沿用卷积加池化网络提取RGB图像三个尺度上的细节特征;后端将红外图像高层特征与RGB图像三个尺度的细节特征进行融合,并将4倍上采样后的融合特征作为语义分割输出. 结果表明,APFCN在像素精度和交并比等方面均优于FCN(输入为RGB图像或红外图像),适用于背景一致下地面目标的语义分割任务.

关键词: 语义分割; 全卷积神经网络; 非对称并行全卷积神经网络; 空洞卷积; 空洞率
中图分类号: TP183 **文献标识码:** A **文章编号:** 0372-2112 (2019)05-1058-07
电子学报URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2019.05.012

Asymmetric Parallel Semantic Segmentation Model Based on Full Convolutional Neural Network

LI Bao-qi, HE Yu-yao, HE Ling-jiao, QIANG Wei
(*School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, Shaanxi* 710072, *China*)

Abstract: Aiming at that RGB image is rich in color details of scene and infrared image is sensitive to outline、size and boundary of target, a novel semantic segmentation model APFCN (Asymmetric Parallelism Fully Convolutional Networks) is proposed. In the upper part of APFCN, a five layer dilation convolution network, where the five kernel sizes are not uniform, is designed used to extract the high-level targets contour features of infrared image. In the lower part of APFCN, a classical CNN network is used to extract three scale features of RGB images. At the back of APFCN, the high level features of the infrared image are fused with the three scale features of the RGB image, and the fused features after 4 times upper sampling is used as the semantic segmentation output of APFCN. The results show that APFCN is better than FCN (input RGB image or infrared image) in PA (Pixel Accuracy) and MIoU (Mean Intersection over Union). APFCN is suitable for the semantic segmentation task of ground targets with consistent background colors.

Key words: semantic segmentation; fully convolution neural network; asymmetric parallelism fully convolutional networks; dilation convolution; dilation rate

1 引言

图像语义分割(场景解析)是在像素层面上识别和分割图像内各个物体的技术,广泛应用于图像理解、内容检索和自动驾驶等领域,是计算机视觉的研究热点问题之一. 语义分割方法通常在标准的数据集上进行研究,例如PASCAL VOC2012、MS COCO、Cityscapes和ADE20K等,这样做的好处是方便实验结果的比较. 虽然上述语义分割数据集中目标种类众多,但目标与背景颜色一般存在明显区分,这有助于目标与背景的分离. 当目标与背景颜色接近(目标隐匿在背景中)或者不同目标之间颜色接近,由于目标与背景之间不存在清晰的轮廓界限,同时目标与背景的纹理特征也相近无几,在像素层次上对图像内目标进行语义分割是困难的.

经典语义分割方法是建立在条件随机场下以保证邻居结点的标识一致性. He等人[1]提出了条件随机场下基于像素点的语义分割方法,该方法利用神经网络训练像素点的颜色特征获取先验模型,然后求解一个条件随机场下的全局能量函数完成语义分割. 由于像素点的局部特征不包含物体的全局统计信息,该方法准确性较差. 同时,以像素点为计算结点的计算量大,效率也偏低. 为此,Yang等人[2]提出了基于超像素的条件

收稿日期:2018-06-04;修回日期:2019-01-17;责任编辑:马兰英
基金项目:国家自然科学基金(No.61271143)

随机场下的语义分割方法. 相比于基于像素点的方法,基于超像素的方法计算结点少,因此效率较高. 然而超像素算法存在过度分割问题,过度分割提取不到有价值的全局特征,该方法分割出的场景比较粗糙. 为了改善语义分割的精度,多层超像素下的语义分割方法[3~6]以及结合像素点和超像素的高阶条件随机场下的语义分割方法[7~9]被提出. 不过,这类改进方法只是将像素点或像素块与简单的特征表达建立联系,同时特征表达和分割过程又是分开的,因此在语义分割的精度和速度上依然难以满足实际需求.

卷积神经网络 CNN[10,11]是一种端到端的深度学习模型[12,13],它从多层的网络结构中自动提取不同层次的特征,这些特征来自于海量的数据,本身具有高度的概括性,能更好的反映物体的特点. Farabet 等人[14]首次提出了基于 CNN 的语义分割模型. 该模型上路首先使用拉普拉斯金字塔对输入图像进行多尺度分解,然后将分解后的图像送入对应的 CNN 网络中进行训练;下路对输入图像进行超像素分割,然后利用条件随机场将像素块与上路的输出结果进行整合. 虽然利用 CNN 提升了图像语义分割的精度,但模型计算复杂度较高. 为此,Long 等人[15,16]提出了一种真正端对端的图像语义分割模型 FCN(Fully Convolutional Networks). FCN 将 CNN 中的全连接层转换成卷积层,通过上采样操作将 CNN 特征恢复成输入图像大小,以 Ground Truth 作为监督信息,让网络直接做像素级别的识别和分割. 虽然融合了三个尺度特征的 FCN-8s 比只用一个尺度的 FCN-32s 的语义分割结果好了很多,但 FCN 的分割边缘还是不够精细. 其中最重要的原因是由于池化层的参数是非可训练的(在整个训练过程中池化层的参数固定不变),输入图像在经过多次池化操作后存在信息大量丢失的问题.

空洞卷积[17]是在卷积层引入了一个称为空洞率(Dilation Rate)的新参数,通过改变该参数来调节卷积核内单元的间距从而同步实现特征提取和特征降维. Chen 等人[18]提出了一种基于空洞卷积网络的语义分割模型. 该模型利用空洞卷积替代卷积加池化操纵来提取输入图像的多尺度特征,然后通过特征融合提高语义分割的精度. 但空洞卷积网络的设计应避免网格问题(Gridding Problem). 网格问题就是在空洞卷积的设计过程中,由于空洞率选取不当致使空洞滤波器叠加操作后存在覆盖不完整,进而导致小目标信息丢失. 为此,Wang 等人[19]提出了一种固定卷积核尺寸下的空洞卷积网络空洞率选取策略 HDC(Hybrid Dilated Convolution)来设计空洞卷积网络以实现对输入图像的覆盖.

RGB 图像和红外图像均是对真实场景的再现. RGB 图像利用可见光传感器依据物体反射率的不同进行成像,具有较好的颜色信息,反应场景的真实情况. 但由于目标与背景颜色接近,图像目标没有明显的轮廓特征. 红外图像依据物体的温度或辐射率不同进行成像,能更好的体现目标的边缘特性,但由于自身的成像原理,红外图像具有噪声大、对比度低、视觉效果差等问题[20]. 对于背景一致的语义分割任务,RGB 图像虽能提供丰富的目标颜色细节信息,但因目标与背景颜色接近,其语义分割结果会出现目标边缘不完整的问题;红外图像虽能提供清晰的目标轮廓特征,但因红外图像的空间分辨率低,其语义分割会出现精度低的问题. 由于红外图像和可见光图像分别反映了场景的温度(或辐射率)和照度分布信息,二者有机结合可以增强场景信息间的互补性,减少场景的不确定性[21],提取红外图像的轮廓特征作为 RGB 图像的补充是改善背景一致下地面目标语义分割的精度的有效方法.

在上述分析的基础上,本文提出了一个非对称并行语义分割模型 APFCN(Asymmetric parallelism Fully Convolutional Networks). 根据 RGB 图像和红外图像的特点,卷积加池化网络适合于提取 RGB 图像目标细节信息,而空洞卷积网络适合于提取红外图像目标轮廓信息. APFCN 上路用空洞卷积网络提取红外图像目标轮廓信息,下路用卷积和池化网络提取 RGB 图像的多尺度细节信息,最后通过融合红外图像目标轮廓信息与 RGB 图像的多尺度细节信息来改善背景一致下地面目标语义分割精度. 其中,为了提取高质量的红外图像目标轮廓信息,本文提出了一种卷积核尺寸非固定的空洞率选取策略 NHDC(Ununified HDC). NHDC 在 HDC 基础上将空洞卷积网络卷积核尺寸按从小到大重新设计,并根据每一层 RDF(Resulting Dilated Flter)来重新确定该层空洞率的大小. 由 NHDC 策略设计的空洞卷积网络能在保证覆盖图像的条件下尽可能多的从前端网络获取更多有用的原始信息,从而生成区分度更高的红外图像高层特征,进一步提高 APFCN 语义分割精度.

2　基于 FCN 的非对称并行语义分割模型

鉴于 FCN 对于背景一致下地面目标 RGB 图像语义分割精度较低,本文提出了一种非对称并行语义分割模型 APFCN. APFCN 上路利用五层空洞卷积网络提取红外图像轮廓特征,五层空洞卷积网络的卷积核与空洞率根据 NHDC 策略设计;下路沿用 FCN 中的卷积加池化网络提取 RGB 图像多尺度细节信息;最后通过融合两类特征来改善地面目标语义分割精度,其中红外高层特征与 RGB 高层特征采用加权融合的方式.

2.1　APFCN 结构设计

APFCN 采用上下非对称结构,如图 1 所示. APFCN 上路的输入为红外图像,使用空洞卷积网络提取红外图像高层特征. 空洞卷积通过设置空洞率能自动的调节卷

积核的尺寸,因此空洞卷积一次操作等同于卷积加池化两次操作,重要的是空洞卷积内所有参数均是可训练的.这就可以有效避免多次池化操作后高层特征信息丢失和无法恢复的问题,同时由于空洞卷积核尺寸大,因此能更好的捕获红外图像目标的轮廓特征,使得高层特征中的目标轮廓更加清晰和完整. APFCN 红外图像经五次空洞卷积后,输出的特征图(dilated conv5)变为原始图像的1/32,空洞网络的设计原理见 2.2 节.

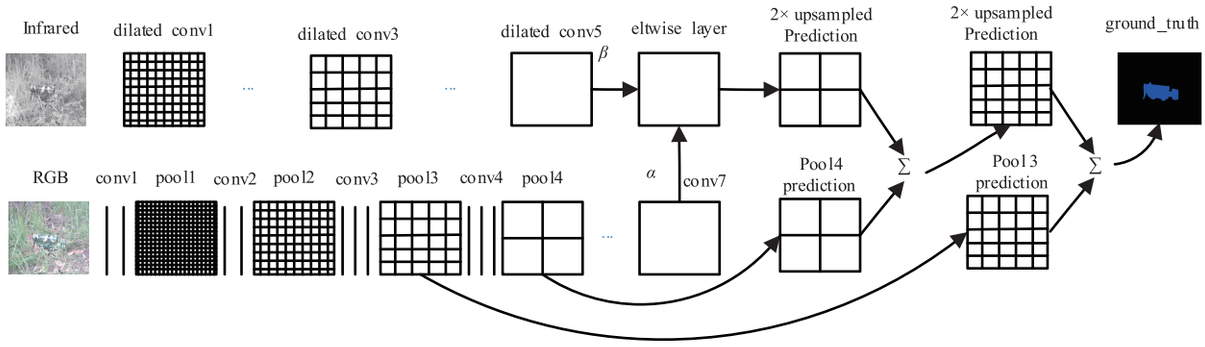

图1 APFCN网络结构

APFCN 下路的输入为 RGB 图像,依然沿用卷积加池化的方式进行提取. 卷积加池化的组合方式能确保卷积神经网络中卷积核的尺寸相对较小,较小的卷积核有助于捕获 RGB 图像的细节信息. 尤其是在池化次数少的条件下,特征图包含丰富的细节信息,而且丢失的信息是可以恢复的. APFCN 依然利用 pool3、pool4 和 conv7 三个尺度的特征[16].

为了保证 dilated conv5 与 conv7 的特征图输出一致,对 dilated conv5 的数量和尺寸进行修正;接着将融合后的特征 eltwise 上采样放大 4 倍,与 pool4 后的特征图进行对应相加融合,之后再对累加结果执行上采样操作放大 2 倍;然后再与 pool3 的特征图累加并上采样放大 4 倍,获得与原图大小一致的输出;最后在 Ground Truth 的约束下,利用 BP(Back Propagation)算法对 APFCN 网络两路参数进行调节直至得到满意的语义分割精度[16].

2.2 空洞卷积网络设计-红外图像高层特征提取

空洞卷积网络的设计应避免网格问题[19]. 网格问题就是空洞卷积网络每一层 RDF 叠加操作后无法实现对输入图像的完全覆盖,因此会造成信息的丢失. 为了解决这个问题,文献[19]提出了一种统一卷积核尺寸下的空洞卷积网络空洞率选取策略 HDC. HDC 卷积核的尺寸是统一的,它通过为每一层网络设置不同的空洞率来改变每层 RDF 的大小和覆盖程度,进而保证叠加后的 RDF 对输入特征的无空洞覆盖. 空洞卷积网络高层特征由前端网络逐层决定,在网络前端使用较小的卷积核能捕获特征图更丰富的信息,在网络端使用较大的卷积核能更好的整合前端信息,有利于生成区分度更高的高层特征. HDC 策略符合上述原则,只不过在统一卷积核尺寸的条件下,卷积核尺寸会相对较大,也就是说第一个空洞卷积层的卷积核尺寸不利于提取更多输入图像的特征. 同时,后端网络 RDF 虽然覆盖范围广,但该层网络的空洞率非常高,过高的空洞率不利于高层信息的整合输出.

本文的思想是在 HDC 空洞率选取策略的基础上(确定每层 RDF 的覆盖范围),通过减小前端网络卷积核尺寸和放大后端卷积核尺寸来改善空洞卷积网络的性能. 基于上述思想,提出了一个可变卷积核尺寸的空洞卷积网络空洞率选取策略 NHDC,并以此来指导空洞卷积网络的设计. NHDC 首先在已知空洞卷积网络层数和卷积核大小的前提下利用 HDC 策略确定空洞卷积网络每一层空洞率的大小;接着在保证空洞卷积网络每层 RDF 固定不变的条件下,给空洞卷积网络每层按从小到大重新设计卷积核尺寸;然后根据 RDF 反推每一层空洞率的大小,给出非固定尺寸卷积核空洞卷积网络的空洞率. NHDC 空洞率选取策略的数学描述如下.

首先确定空洞卷积网络每一层的空洞率. 对于一个层数为 n,卷积核尺寸为$[k_1,k_2,\cdots,k_i,\cdots,k_n]$,空洞率为$[r_1,r_2,\cdots,r_i,\cdots,r_n]$的空洞卷积网络,空洞卷积网络第 i 层最大空洞率 M_i 定义为:

$$M_i=\max[M_{i+1}+1-2r_i,M_{i+1}+1-2(M_{i+1}+1-r_i),r_i] \tag{1}$$

其中 M_i 为第 i 个空洞卷积层的最大空洞率,r_i 为第 i 个空洞卷积层的空洞率. 式(1)同时需要保证 $M_2\leq k_2$,即第 2 个空洞卷积层的最大空洞率不大于该层卷积核的尺寸. 对于卷积核尺寸均为 3 * 3 的三层空洞卷积网络,当 $M_2=2$ 时,$\boldsymbol{r}=[1,2,5]$满足条件;当而 $M_2=5$ 时,$\boldsymbol{r}=[1,2,9]$不满足条件. 除此以外,空洞卷积网络的空洞率不能有大于 1 的公约数. 比如 [2,4,6] 不是一个好的三层空洞卷积,依然会出现网格问题.

其次,将原空洞卷积网络卷积核$[k_1,k_2,\cdots,k_i,\cdots,k_n]$修订为$[\overline{k_1},\overline{k_2},\cdots,\overline{k_i},\cdots,\overline{k_n}]$,其中 $\overline{k_1}\leq\overline{k_2}\leq\cdots\leq\overline{k_n}$. 对

于一个卷积核为 $k*k$ 的空洞卷积层,该层实际空洞滤波器(RDF)尺寸与空洞率之间的关系如下:

$$k_{rdf}=k+(k-1)(r-1) \tag{2}$$

其中 k_{rdf} 为该层 RDF 尺寸,k 为该层卷积核尺寸,r 为该层空洞率大小. 例如,一个卷积核尺寸为 $3*3$,空洞率 $r=2$ 的空洞卷积层, RDF 的实际覆盖范围为 $5*5$,即 $k_{rdf}=5$.

最后,依据空洞卷积网络每层 RDF 尺寸不变的原则,结合式(2),对 $[\overline{k_1},\overline{k_2},\cdots,\overline{k_i},\cdots,\overline{k_n}]$ 重新计算每一层的空洞率,新的空洞率定义为:

$$\overline{r_i}=\frac{(k_{rdf}^i-\overline{k_i})}{(\overline{k_i}-1)}+1 \tag{3}$$

其中 $\overline{r_i}$ 为优化后的第 i 个空洞卷积层的空洞率,k_{rdf}^i 为第 i 个空洞卷积层的 RDF 尺寸,$\overline{k_i}$ 为修订后的第 i 个空洞卷积层的卷积核尺寸. 对于空洞卷积网络第一个卷积核 $\overline{k_1}$,当 $\overline{r_i}=1$ 时,该层的空洞率保持不变,由该层引入的特征图尺寸的增量由后续空洞网络抵消.

表 1　五层空洞卷积网络层参数设置

Name	dilated conv1	dilated conv2	dilated conv3	dilated conv4	dilated conv5
Kernel Size	3*3	7*7	9*9	11*11	15*15
Output	32	64	128	256	512
Dilation	1	7	9	11	7

为了提取尺寸为输入图像 1/32 的高层特征,为红外图像设计了一个五层的空洞卷积网络. 依据 NHDC 空洞率选取策略,网络初始卷积核尺寸为 $9*9$,初始空洞率为[1,3,5,9,13],经过 NHDC 策略得到的五个空洞卷积层的参数如表 1 所示. 新的空洞卷积网络前端卷积核的尺寸变小,网络后端的卷积核变大,有利于提取高质量的红外图像高层轮廓特征.

2.3　红外图像特征与 RGB 图像特征的融合方式

图像融合根据融合处理所处的阶段不同可以分为三个等级,即像素级融合、特征级融合和决策级融合[22]. FCN 网络中 pool3、pool4 和 conv7 三个尺度的特征采用的是像素级的相加融合(对应特征值直接相加),为了获取相同的特征图尺寸需要对上层特征进行上采样和 Crop 修正[15]. 例如,对 pool4 和 conv7 进行融合,需要对 conv7 进行上采样和适当剪裁. 在 APFCN 网络,是对 dilated conv5、pool3、pool4 和 conv7 四个特征进行融合,其中 dilated conv5 是新增特征,与 conv7 属于同一尺度(dilated conv5 本质是与 conv7 数量和尺寸相同的特征图). 因此,同样是在像素级层次上对 dilated conv5 与 conv7 的特征值进行融合处理. 对高层特征而言,重要的是提取目标的全局信息(轮廓特征). 由于红外图像本身的目标轮廓比较清晰,同时空洞卷积网络(APFCN 上路)中的参数完全可训练,因此由空洞卷积网络得到的红外高层特征(dilated conv5)比由卷积加池化网络(APFCN 下路)得到的 RGB 图像高层特征(conv7)更能反映目标的全局信息. 对 APFCN 而言,用 dilated conv5 替换 conv7 同样能获得性能上的提升.

由于红外图像和 RGB 图像是目标两个视角的表达,代表了目标两个不同的属性-红外图像侧重于目标轮廓,RGB 图像侧重于目标细节. 同样的,dilated conv5 和 conv7 也是对两种表达的进一步生成和体现. 换句话说,dilated conv5 和 conv7 同样不是替代关系而是互补关系[21]. 为此,APFCN 通过融合 dilated conv5 和 conv7 来获得更高的性能. 考虑 dilated conv5 的 conv7 的重要程度不同,本文对其采用加权融合方式,并选取 $\alpha=0.60$,$\beta=0.40$ 的融合系数,其中 α 是 conv7 特征的融合系数,β 是 dilated conv5 的融合系数,$\alpha+\beta=1$. 获取 dilated conv5 和 conv7 的融合特征后,按照 FCN 网络的融合方式[15]将其与 pool3 和 pool4 进行融合.

3　仿真试验

为了客观评价本文语义分割方法 APFCN,首先建立背景一致条件下地面目标的 RGB 图像语义分割数据集和红外图像语义分割数据集,并以 PA(Pixel Aaccuracy)、MPA(Mean Pixel Accuracy)和 MIoU(Mean Intersection over Union)作为语义分割模型性能的定量评价指标[16],其中 MIoU 为主要评价指标,PA 和 MPA 为辅助评价指标. 为了验证 APFCN 对背景一致地面目标语义分割的有效性,以及空洞率选取策略(APFCN 上路网络)和高层特征(dilated conv5 和 conv7)融合方式对 APFCN 语义分割精度的影响. 设计实验 1,以背景一致条件下地面目标 RGB 图像在 FCN 模型下的输出结果为参考,比较分析 APFCN 和 FCN 的语义分割精度. 设计实验 2,以融合系数为 $\alpha=0.00$,$\beta=1.00$ 的 APFCN 为研究对象,比较 NHDC 和 HDC 两种空洞率选取策略对其语义分割精度的影响. 设计实验 3,以空洞率策略为 NHDC 的 APFCN 为研究对象,比较 conv7 和 dilated conv5 不同融合权重对其语义分割精度的影响. 实验采用的 FCN 和 APFCN 由 caffe 工具箱设计,采用 GPU(Titan X)计算方式,并利用 cuDNN 进行加速处理.

3.1　实验数据集

为了真实反映背景一致条件下的地面目标场景,实验选用五种与地面背景颜色接近的目标为语义分割数据集的样本源,其中包括坦克-a、坦克-b、吉普车、导弹、直升机. 为了建立背景一致条件下地面目标 RGB 图像语义分割数据集和红外图像语义分割数据集,利用彩色/红外摄像机对五种地面目标分别进行图像采集. 背景一致条件下地面目标 RGB 图像和红外图像语义分

割数据集如表 2 所示,其中 data-RGB 表示 RGB 图像数据集,data-Infrared 表示红外图像数据集. 按相同的顺序从两个数据集中选取 900 幅图像作为语义分割训练数据集,100 幅图像作为语义分割测试数据集. 同一场景同一目标获取的红外和彩色图像共享 Ground Truth, Ground Truth 由 label me 标注.

表 2 地面目标语义分割数据集组成

	data-RGB(幅)	data-Infrared(幅)
tank-a	200	200
tank-b	200	200
jeep	200	200
missile	200	200
helicopter	200	100
总计	1000	1000

3.2 实验 1:APFCN 和 FCN 语义分割精度的比较

本实验比较 APFCN 和 FCN 对背景一致条件下地面目标的语义分割精度. 实验以输入 RGB 图像(data-RGB)的 FCN 为参考,记作:FCN-RGB. APFCN 的上路和下路的输入分别为红外图像(data-Infrared) 和 RGB 图像(data-RGB),空洞率选取策略为 NHDC,RGB 图像高层特征和红外图像高层特征的融合系数为 $\alpha=0.60$, $\beta=0.40$. 同时为了比较 RGB 图像和红外图像在地面目标语义分割中的不同特点,设计 FCN-INF,即输入红外图像(data-Infrared)的 FCN. 记录语义分割模型迭代 20000 次时的 PA、MPA 和 MIoU 的数值,并将其作为语义分割模型的定量评价指标.(见表 3)

表 3 APFCN 与 FCN 语义分割精度比较

	PA	MPA	MIoU
FCN-RGB	0.939	0.745	0.708
FCN-INF	0.905	0.522	0.477
APFCN	0.958	0.909	0.799

从表 3 可以发现,对于背景一致条件下的地面目标语义分割任务,APFCN 在三种定量评价指标中明显优于 FCN-RGB,其中 PA 提升 0.019、MPA 提升 0.164、MIoU 提升 0.091. APFCN 更是大幅度优于 FCN-INF,其中 PA 提升 0.053、MPA 提升 0.387、MIoU 提升 0.322. 从实验数据来看,APFACN 比 FCN 更适合背景一致条件下的地面目标语义分割任务.

为了更直观的说明 APFCN 对背景一致条件下地面目标语义分割的有效性,用训练 20000 次的 APFCN、FCN-RGB 和 RGB-INF 模型分别对 3 种地面目标图像进行语义分割,效果如图 2 所示.

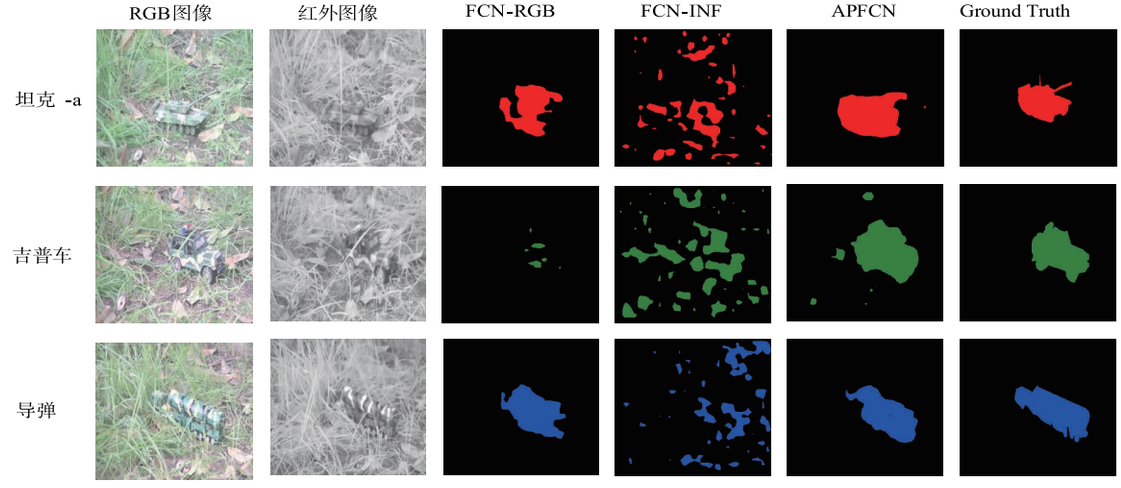

图2 FCN和APFCN语义分割效果图

从图 2 可以发现,在以 RGB 图像为输入的 FCN-RGB 语义分割效果图中,图像目标轮廓不完整,尤其是对吉普车,FCN 并没有将其从 RGB 图像中分割和识别出来;在以红外图像为输入的 FCN-INF 语义分割效果图中,图像目标出现了碎片化的现象,同时背景的部分内容也被分割识别成图像目标. 在以红外图像和 RGB 图像为输入的 APFCN 语义分割效果图中,图像目标具有较清晰和完整的轮廓,同时目标与背景实现了较好的区分. APFCN 通过融合红外图像和 RGB 图像高层特征增强场景信息间的互补性来改善背景一致条件下的地面目标语义分割效果.

3.3 实验 2:空洞率选取对 APFCN 性能的影响

本实验比较 NHDC 和 HDC 两种空洞率选取策略对 APFCN 语义分割精度的影响. 为了更好的比较 NHDC 和 HDC 两种空洞率选取策略,实验以 $\alpha=0.00$, $\beta=1.00$ 下(用红外高层特征 dilated conv5 替代 RGB 高层特征 conv7)的 APFCN 为研究对象,上路和下路的输入分别为 data-Infrared 和 data-RGB 两个数据集. 空洞卷积网络的层数为 5 层,HDC 策略使用的卷积核尺寸为 $9*9$,空洞率为[1,3,5,9,11];NHDC 策略使用的卷

积核尺寸为[3,7,9,11,13],空洞率为[1,7,9,11,7].记录语义分割模型迭代20000次时的PA、MPA和MIoU的数值,并将其作为模型的定量分析指标.

表4　HDC和NHDC对APFCN语义分割精度的影响

	k	r	PA	MPA	MIoU
HDC	[9,9,9,9,9]	[1,3,5,9,11]	0.942	0.822	0.726
NHDC	[3,7,9,11,15]	[1,7,9,11,7]	0.946	0.888	0.758

从表4可以看出,在PA、MPA和MIoU三个指标上NHDC比HDC分别提升0.004、0.066和0.032.由NHDC设计的空洞卷积网络提取的红外图像高层特征质量更好,能进一步优化APFCN的语义分割效果.

3.4　实验3:不同融合方式对APFCN性能的影响

本实验比较conv7和dilated conv5采用不同系数进行融合对APFCN语义分割精度的影响.实验以空洞率选取策略为NHDC的APFCN模型为对象,融合系数在$\alpha+\beta=1$的约束下,按0.1步长依次取值,记录APFCN模型迭代20000次输出的PA、MPA和MIoU的数值,并将其作为融合方式的定量评价指标.结果如表5所示.

表5　不同融合系数对APFCN语义分割精度的影响

α	β	PA	MPA	MIoU
0.00	1.00	0.946	0.888	0.758
0.10	0.90	0.885	0.847	0.651
0.20	0.80	0.919	0.698	0.736
0.30	0.70	0.908	0.892	0.710
0.40	0.60	0.925	0.886	0.726
0.50	0.50	0.936	0.924	0.748
0.60	0.40	0.958	0.909	0.799
0.70	0.30	0.953	0.874	0.777
0.80	0.20	0.959	0.854	0.791
0.90	0.10	0.956	0.846	0.771
1.00	0.00	0.939	0.745	0.708

从表5可以看出,APFCN在$\alpha=0.60,\beta=0.40$条件下的语义分割的精度是最高的.比较$\alpha=1.00,\beta=0.00$和$\alpha=0.00,\beta=1.00$两种条件下的APFCN输出结果,可以发现$\alpha=0.00,\beta=1.00$时的三个评价指标高于$\alpha=1.00,\beta=0.00$,结果表明经空洞卷积网络得到的红外图像高层特征dilated conv5的质量高于经卷积加池化操作得到的RGB图像高层特征conv7.虽然dilated conv5质量高于conv7,但用dilated conv5替代conv7,APFCN并没有获取最优的语义分割精度.主要是由于dilated conv5与conv7不是替代的关系,而是互补的关系.从表5的其他组数据可以发现,并不是任意的融合比例都会带来APFCN性能的提升.当RGB图像高层特征conv7的融合系数$\alpha\in[0.10,0.50]$,APFCN的语义分割精度低于$\alpha=0.00,\beta=1.00$时的APFCN语义分割精度.当conv7的融合系数$\alpha\in[0.60,0.90]$,APFCN的语义分割精度高于$\alpha=0.00,\beta=1.00$时APFCN语义分割精度.对APFCN而言,红外图像高层特征和RGB图像高层特征合理的融合关系是:RGB图像的比重大于红外图像.高层特征中的目标轮廓信息和细节颜色信息对生成高质量的语义分割结果都是有益的,合理的融合比例会改善语义分割的效果.

4　结论

背景一致下地面目标语义分割具有重要的理论研究和实际应用价值.结合RGB图像和红外图像,本文提出了一种新的语义分割模型APFCN.APFCN有效解决了背景一致的地面目标语义分割边缘模糊的问题,并经理论分析和仿真实验证明了APFCN的有效性.同时提出了一种新的空洞率选取策略NHDC,NHDC用于设计全覆盖空洞卷积网络来提取高质量的红外图像高层特征,并实验验证了该策略的有效性.

虽然APFCN比FCN对背景一致地面目标图像语义分割精度高,但对主体存在外部的构件的目标(如坦克的炮管),APFCN语义分割的效果还需要提升.下一步的研究重点包括:(1)研究更合理的空洞卷积网络设计方法;(2)进一步优化APFCN的模型结构,提高模型的语义分割精度.

参考文献

[1] He X,Zemel R S,Carreiraperpiñán M Á. Multiscale conditional random fields for image labeling[A]. Proceedings of Computer Vision and Pattern Recognition[C]. Washington:IEEE,2004. 695 – 702.

[2] Yang L,Meer P,Foran D J. Multiple class segmentation using A unified framework over mean-shift patches[A]. Proceedings of Computer Vision and Pattern Recognition[C]. Minneapolis:IEEE,2007. 1 – 8.

[3] Pantofaru C,Schmid C,Hebert M. Object recognition by integrating multiple image segmentations[A]. Proceedings of European Conference on Computer Vision[C]. Berlin:Springer,2008. 481 – 494.

[4] Gould S,Fulton R,Koller D. Decomposing a scene into geometric and semantically consistent regions[A]. Proceedings of International Conference on Computer Vision[C]. Kyoto:IEEE,2009. 1 – 8.

[5] Kumar M P,Koller D. Efficiently selecting regions for scene understanding[A]. Proceedings of Computer Vision and Pattern Recognition[C]. San Francisco:IEEE,2010. 3217 – 3224.

[6] Jain A,Gupta A,Davis L S. Piecing together the segmentation jigsaw using context[A]. Proceedings of Computer Vision and Pattern Recognition[C]. Colorado:IEEE,2011. 2001 – 2008.

[7] Kohli P,Ladický L,Torr P H S. Robust higher order poten-

tials for enforcing label consistency[A]. Proceedings of Computer Vision and Pattern Recognition[C]. Anchorage: IEEE,2008. 1 – 8.

[8] Russell C. Associative hierarchical CRFs for object class image segmentation[A]. Proceedings of International Conference on Computer Vision[C]. Kyoto: IEEE,2009. 739 – 746.

[9] Ladicky L,Russell C,Kohli P,et al. Graph cut based inference with co-occurrence statistics[A]. Proceedings of European Conference on Computer Vision[C]. Berlin: Springer,2010. 239 – 253.

[10] Hinton G. Where do features come from ? [J]. Cognitive Science,2014,38(6):1078 – 101.

[11] Lecun Y,Bengio Y,Hinton G. Deep learning[J]. Nature,2015,521(7553):436 – 444.

[12] Krizhevsky A,Sutskever I,Hinton G E. ImageNet classification with deep convolutional neural networks[A]. Proceedings of Neural Information Processing Systems[C]. Doho: ACM,2012. 1097 – 1105.

[13] Szegedy C,Liu W,Jia Y,et al. Going deeper with convolutions[A]. Proceedings of Computer Vision and Pattern Recognition[C]. Boston: IEEE,2015. 1 – 9.

[14] Farabet C,Couprie C,Najman L,et al. Learning hierarchical features for scene labeling[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2013,35(8): 1915 – 1929.

[15] Long J,Shelhamer E,Darrell T. Fully convolutional networks for semantic segmentation[A]. Proceedings of Computer Vision and Pattern Recognition[C]. Boston: IEEE,2015. 3431 – 3440.

[16] Long J,Shelhamer E,Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2014,39(4):640 – 651.

[17] Yu F,Koltun V. Multi-scale context aggregation by dilated convolutions[A]. Proceedings of International Conference on Learning Representations[C]. San Juan: ICLR,2016. arXiv:1511. 07122.

[18] 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,2018,40(4):834 – 848.

[19] Wang P,Chen P,Yuan Y,et al. Understanding convolution for semantic segmentation[A]. Proceedings of Computer Vision and Pattern Recognition[C]. Lake Tahoe: IEEE,2018. 1451 – 1460.

[20] 周则明,胡彪,孟勇,陈超迁,罗其祥. 基于流形特征与形状先验的红外直升机图像分割[J]. 电子学报,2018,46(4):834 – 839.
ZHOU Ze-ming,HU Biao,MENG Yong,CHEN Chao-qian,LUO Qi-xiang. Infrared helicopter image segmentation based on manifold feature and shape priori[J]. Acta Electronica Sinica,2018,46(4):834 – 839. (in Chinese)

[21] 廉蔺,李国辉,张军,涂丹. 基于参数分步估计的红外与可见光图像自动配准算法[J]. 电子学报,2012,40(9): 1829 – 1838.
LIAN Lin,LI Guo-hui,ZHANG Jun,TU Dan. An automatic algorithm for infrared and visible image registration based on pParameter Step Estimation[J]. Acta Electronica Sinica,2012,40(9):1829 – 1838. (in Chinese)

[22] 韩崇昭,朱洪艳,段战胜. 多源信息融合[M]. 北京:清华大学出版社,2010. 394 – 449.
Han Chong-zhao, Zhu Hong-yan, Duan Zhan-sheng. Multi-Source Informatin Fusion[M]. Beijing: Tsinghua University Press,2010. 394 – 449. (in Chinese)

作者简介

李宝奇(通信作者) 男,1985 年 12 月生,天津宝坻人. 现于西北工业大学航海学院攻读博士学位,研究方向为目标检测、识别和跟踪,深度学习理论.
E-mail:bqli@ mail. nwpu. edu. cn

贺昱曜 男,1956 年生,陕西富平人. 教授,西北工业大学博士生导师,主要研究方向:精确制导预仿真,智能控制与智能优化理论,图像处理理论与算法.
E-mail:heyyao@ nwpu. edu. cn

何灵蛟 男,1994 年 12 月生,甘肃会宁人. 现于西北工业大学航海学院攻读硕士学位,研究方向为图像增强、图像语义分割及目标检测与识别.
E-mail:helingjiao@ mail. nwpu. edu. cn

强 伟 男,1986 年 12 月生,陕西延安人. 现于西北工业大学航海学院攻读硕士学位,研究方向为图像分类、图像语义分割及目标检测与识别等深度学习理论.
E-mail:xgd2017qw@ mail. nwpu. edu. cn