

基于双分支融合注意力机制的图像分割算法^①



李绍华¹, 于俊洋¹, 郑珂², 翟锐¹

¹(河南大学 软件学院, 开封 475004)

²(河南大学 经济学院, 开封 475004)

通信作者: 郑珂, E-mail: 15134166@qq.com

摘要: 针对 DeepLabV3+ 在特征提取阶段忽略了不同尺度特征重要程度出现的部分细节信息损失导致图像分割不细致, 提出一种融合双分支特征提取和注意力机制的改进算法. ResNet101 骨干网络初步提取出的特征图作为注意力机制的输入特征, 解决了网络退化及梯度消失的问题, 也能够捕获到被 DeepLabV3+ 忽略的图像细节信息; 设计双分支特征提取机制扩大特征提取能力, 细化图像边缘信息以优化网络对不同尺度特征关注不均的问题; 同时, 联合采用交叉熵损失和类别不平衡函数两种损失函数作为损失函数, 通过聚焦于前景样本降低背景的影响, 提高算法分割精度. 实验结果表明, 改进算法在 PASCAL VOC 2012 和 CityScapes 数据集上的平均交并比 (*MIoU*) 值分别达到了 79.92% 和 68.59%, 与经典算法和基于 DeepLabV3+ 改进的算法相比, 特征提取的准确性有所提高, 分割效果更优.

关键词: 深度学习; 语义分割; 注意力机制; 双分支特征提取

引用格式: 李绍华, 于俊洋, 郑珂, 翟锐. 基于双分支融合注意力机制的图像分割算法. 计算机系统应用, 2023, 32(5): 212-219. <http://www.c-s-a.org.cn/1003-3254/9107.html>

Image Segmentation Algorithm Based on Dual-branch Fusion Attention Mechanism

LI Shao-Hua¹, YU Jun-Yang¹, ZHENG Ke², ZHAI Rui¹

¹(School of Software, Henan University, Kaifeng 475004, China)

²(School of Economics, Henan University, Kaifeng 475004, China)

Abstract: DeepLabV3+ ignores the loss of part of detail information due to the importance of features at different scales in the feature extraction stage, which results in imprecise image segmentation. In response, this study proposes an improved algorithm integrating dual-branch feature extraction and attention mechanism. The feature map extracted by the ResNet101 backbone network is used as the input feature of the attention mechanism, which solves the problems of network degradation and gradient disappearance and also captures the image details ignored by DeepLabV3+. The dual-branch feature extraction mechanism expands the feature extraction capability and refines the image edge information to optimize the uneven attention of the network to features at different scales. At the same time, the CE loss function and the Dice loss function are jointly used to reduce the influence of background by focusing on foreground samples and improve segmentation accuracy. The experimental results show that the mean intersection over union (*MIoU*) of the improved algorithm on the PASCAL VOC 2012 and CityScapes datasets reaches 79.92% and 68.59%, respectively. Compared with the classical algorithm and other improved algorithms based on DeepLabV3+, the proposed algorithm obtains a better segmentation effect.

Key words: deep learning; semantic segmentation; attention mechanism; dual-branch feature extraction

① 基金项目: 河南省科技研发项目 (212102210078); 河南省科技攻关项目 (222102210229)

收稿时间: 2022-10-27; 修改时间: 2022-11-29, 2022-12-23; 采用时间: 2023-01-06; csa 在线出版时间: 2023-03-30

CNKI 网络首发时间: 2023-03-30

图像分割是将一幅图像按照一定的标准分割成不同区域的过程,作为计算机视觉领域的一个重要分支,已经取得了一定的进展,并在医疗、自动驾驶等领域得到了广泛应用.语义分割算法主要分为传统的分割算法和基于深度学习的分割算法^[1].

传统分割算法实现简单计算量小,基于阈值的分割方法利用图像的不同灰度阈值将前景物体和背景分离出来,但对于复杂图像,阈值分割效果不佳.基于区域的分割方法将具有相似特征的像素组合起来形成区域,从而将图像分割成若干不同区域.基于边缘检测的分割方法将边缘像素点连接起来形成边缘轮廓,从而达到分割的目的.基于聚类的分割方法是基于事物之间的相似性作为划分标准,即将具有相似特征的像素点聚集在同一区域,完成分割.随着深度学习的不断发展,传统的语义分割效果明显不如基于深度学习的语义分割效果.

深度学习的出现为语义分割带来了新的思路,FCN^[2]将全连接层改为全卷积层,全卷积层可以取任意大小的图像作为输入,并使用上采样来解决卷积核池对图像大小的影响.U-Net^[3]在FCN的基础上增加了跳跃链接,在医学图像分割方面具有优势.SegNet^[4]是一种基于编码和解码的图像分割网络,解码部分对低分辨率特征图像进行上采样操作,避免了FCN中上采样的消耗.DeepLab^[5-8]模型提出了空洞卷积,使得多尺度分割对象得到更好的分割效果,其中DeepLabV3+^[8]是在对之前的模型进行优化的基础上,使用空洞卷积来平衡精度和运行时间,但同时膨胀率较大的空洞卷积会忽略小尺度目标,造成信息丢失.基于上述分析,本文

对DeepLabV3+模型做出以下改进.

1) 设计多维注意力机制捕捉骨干网络输出的深层特征从而避免信息丢失,保留更多有效的细节信息.

2) 经过DCNN结构得到的拼接特征图输入到双分支特征提取机制中,进一步地提取到被忽略的小尺度目标,使深层特征图更好地与浅层特征图融合.

3) 分析交叉熵损失和类别不平衡函数的优缺点,设计一种联合损失函数,解决对不同尺度目标关注度不均的问题,提升分割效果.

1 相关理论研究

1.1 DeepLabV3+

DeepLab模型提出了空洞卷积的思想,可以在不损失分辨率的情况下扩大感受野,同时获得多尺度的上下文信息,并在此基础上不断更新优化,其中DeepLabV3+分割效果最好,其网络结构如图1所示.编码器包含带有空洞卷积(atrous spatial pyramid pooling, ASPP)的深度卷积网络(deep convolutional network, DCNN),骨干网络对输入的原始图像进行初步处理,提取图像特征,图像依次变为原图的1/4、1/8、1/16大小,之后将1/16大小的特征图先输入到DCNN结构中,分别进行卷积、膨胀率为{6, 12, 18}的空洞卷积,以及全局平均池化,再拼接卷积压缩特征输入到解码器部分,4倍上采样后得到1/4大小的深层特征图(deep feature map).同时,1/4大小的浅层特征图(low-level feature map)直接输入到解码器部分,通过卷积调整通道数后与深层特征图叠加融合,进行深度可分离卷积得到最终的预测图.

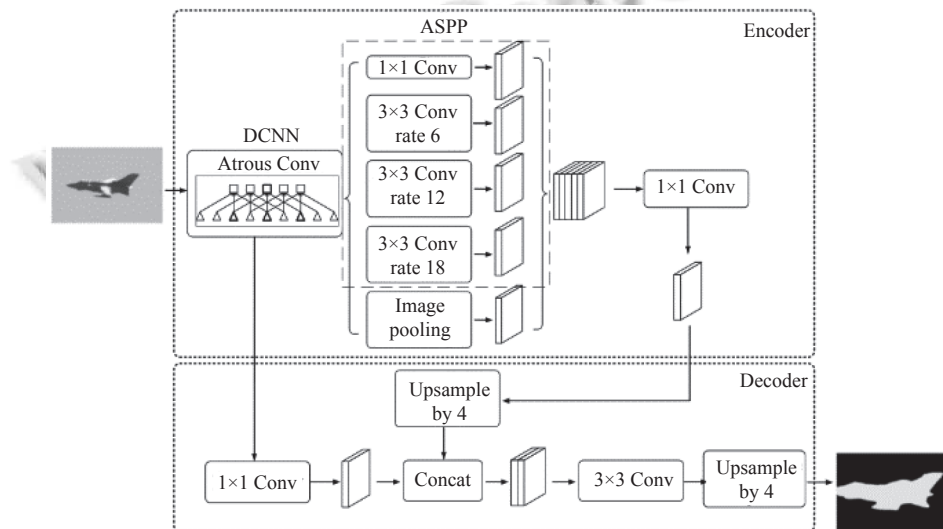


图1 DeepLabV3+网络结构图

1.2 注意力机制

注意力机制用于对前期输入的图像信息进行加权, 聚焦到某些局部信息, 以此提升训练的准确性, 主要分为通道注意力、空间注意力以及混合注意力。

通道注意力结构如图2所示, 输入的特征图 $F (H \times W \times C)$ 经过基于宽和高的最大池化 ($MaxPool$) 和平均池化 ($AvgPool$), 得到两个 $1 \times 1 \times C$ 的特征图去关注到有意义的特征, 分别送入一个共享连接层 ($shared MLP$) 获得加权特征, 并进行加和操作以及 Sigmoid 激活操作得到通道注意力特征 M_c , 数学表达式如式 (1) 所示, 其中 σ 代表 Sigmoid 激活函数。

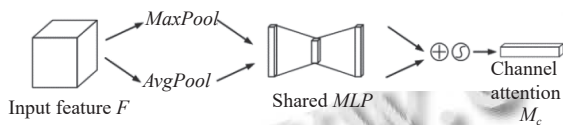


图2 通道注意力机制结构图

空间注意力结构如图3所示, 输入特征 $F (H \times W \times C)$ 做一个基于通道的最大池化和平均池化操作得到两个 $H \times W \times 1$ 的特征图, 通过 7×7 卷积操作降维以及 Sigmoid 激活后得到空间注意力特征 M_s , 数学表达式如式 (2) 所示, 其中 σ 代表 Sigmoid 激活函数, $f^{7 \times 7}$ 为 7×7 卷积操作。

$$M_c(F) = \sigma(MLP(MaxPool(F)) + MLP(AvgPool(F))) \quad (1)$$

$$M_s(F) = \sigma(f^{7 \times 7}([MaxPool(M_c(F)); AvgPool(F)])) \quad (2)$$

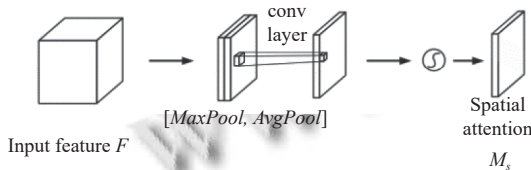


图3 空间注意力机制结构图

2 基于 DeepLabV3+ 的改进算法

为了解决图像信息丢失和细节分割不准确的问题, 本文提出一种基于 DeepLabV3+ 的改进算法, 网络结构如图4所示。采用 ResNet101 网络加载预训练模型提取浅层和深层特征; 对 DCNN 模块进行改进, 增加注意力机制减少空洞卷积带来的小尺度目标丢失, 同时

聚焦更丰富的细节信息; 设计双分支特征提取机制扩大特征提取能力, 对空洞卷积和注意力机制的输出拼接特征图再次细化; 最后联合两种损失函数的优缺点互补同时作为改进算法的损失函数, 从而提升分割效果。

2.1 多维注意力机制

为了关注到小尺度目标, 提高对图像中存在的高频有用信息的获得能力, 本文算法设计一种多维注意力机制, 结合通道和空间两个维度聚焦空洞卷积丢失的小尺度目标。对于需要接收和处理大量数据的图像分割算法而言, 通常只需要关注少部分信息。经过骨干网络初步提取出的特征先输入到通道注意力机制处理其通道分配关系, 按照重要程度分配权重; 再传入到空间注意力机制中进行学习, 使模型更加关注起决定作用的区域。通道和空间注意力同时作用, 可以获取到更多需要关注的细节信息, 抑制其他无用信息, 降低模型对背景的关注, 突出图像中的前景物体, 提升图像分割效果。

另外, 原 DCNN 结构中输出的特征通道数都由 2048 调整为 256, 因此在新增注意力机制之后也加入卷积调整通道数, 以便和其他 5 层特征图融合, 结构如图5所示。InputFeature 为输入到注意力机制的 $1/16$ 大小的特征图, RefineFeature 为经过通道和空间注意力模块处理的特征图, 再进行 1×1 卷积调整通道数, 得到 1 层输出特征图。

改进后的 DCNN 结构分别进行标准卷积, 膨胀率为 {6, 12, 18} 的空洞卷积, 全局平均池化层和新增注意力机制层, 得到输出特征图。假设改进后的 DCNN 结构输入图像为 F_{in} , 混合注意力机制输出如式 (3) 所示, 其中 $f^{1 \times 1}$ 为 1×1 卷积, $M_c(F_{in})$ 是以 F_{in} 作为输入代入式 (1) 中得到的通道注意力权重, $M_s(F_{in} \cdot M_c(F_{in}))$ 是以 $F_{in} \cdot M_c(F_{in})$ 作为输入代入式 (2) 得到的空间注意力机制权重。

$$O = f^{1 \times 1}(F_{in} \cdot M_c(F_{in}) \cdot M_s(F_{in} \cdot M_c(F_{in}))) \quad (3)$$

2.2 双分支特征提取机制

为了提取到更加丰富的特征信息, 提高目标区域分割精度, 本文算法设计了一种双分支特征提取方式, 依次提高模型对图像信息的提取能力, 使训练效果更佳, 双分支结构如图6所示, C_{out} 代表各部分特征图的输出通道数。

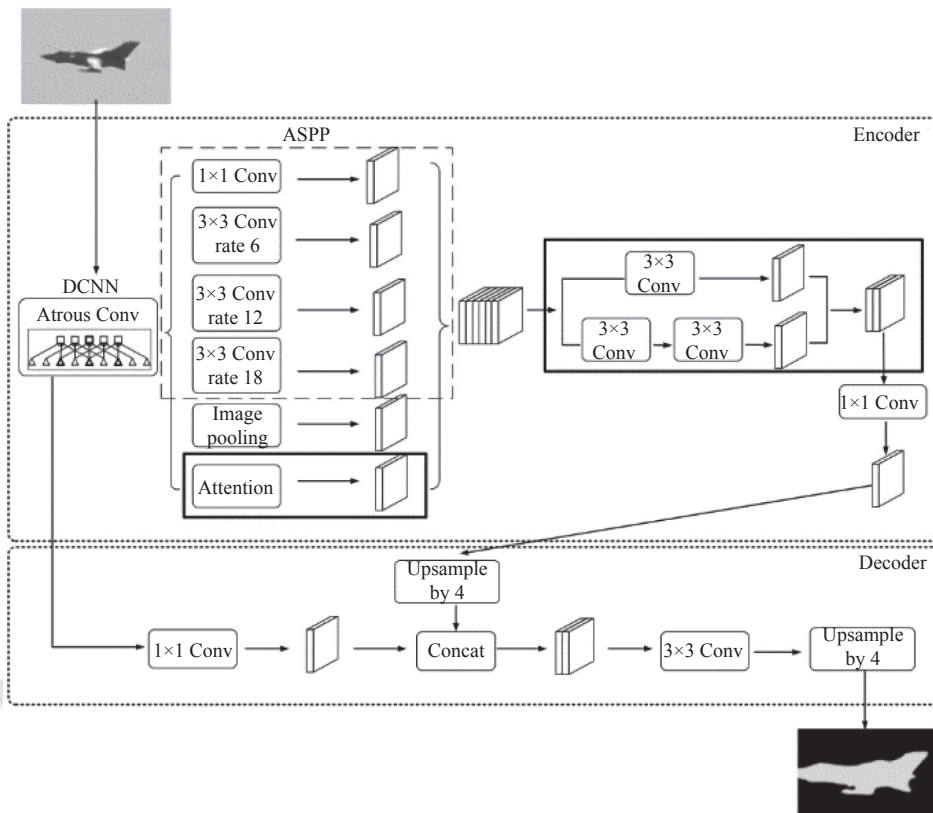


图4 改进网络结构图

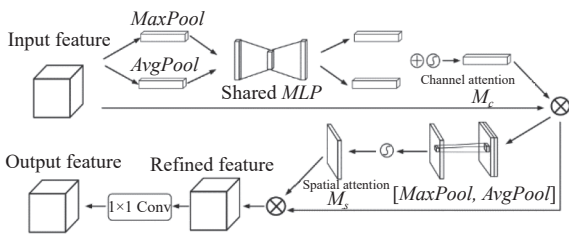


图5 新增注意力机制结构图

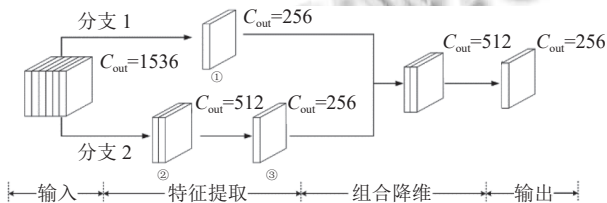


图6 双分支特征提取结构图

相较于原算法的单一卷积, 双分支卷积扩大感受野, 增加了局部上下文信息, 进一步获取到前景物体的细节信息. 同时考虑到网络模型大小, 训练速度及时间等原因, 双分支与更多分支特征提取相比效果无较大差异, 但参数量和计算量更少. 因此本文采用双分支对

输入特征进行同步特征提取, 扩大特征提取通道, 提取更加充分的信息, 从而达到更精准的分割效果.

第1层为输入层. 将DCNN模块中得到6层拼接特征图作为输入, 分别送到分支1和分支2进行特征提取, 此时通道数为 $256 \times 6 = 1536$.

第2层为特征提取层. 对第1层输出的特征图进行卷积操作以提取特征, 分支1使用 3×3 卷积将通道数直接从1536调整为256, 该过程如式(4)所示:

$$F_1(Y_1) = \rho(W_{3 \times 3}^1 \cdot Y_1 + b_1) \quad (4)$$

其中, $F_1(Y_1)$ 为输出特征图, $W_{3 \times 3}^1$ 为通道1使用 3×3 卷积的权重, Y_1 为输入特征图, b_1 为①的偏置值, ρ 为激活函数.

分支2的输入同样是第1层的输出, 其中②的 3×3 卷积将通道数1536调整为512, 再送入③中, 继续卷积调整为256, 分段卷积进行特征提取, 能够关注到被忽略的细节信息, 该过程如式(5)和式(6)所示.

$$F_{21}(Y_2) = \rho(W_{3 \times 3}^{21} \cdot Y_2 + b_{21}) \quad (5)$$

$$F_{22}(F_{21}) = \rho(W_{3 \times 3}^{22} \cdot F_{21} + b_{22}) \quad (6)$$

其中, $F_{21}(Y_2)$ 和 $F_{22}(F_{21})$ 分别为分支2中②和③输出特征图, $W_{3 \times 3}^{21}$ 和 $W_{3 \times 3}^{22}$ 分别为②和③的权重, Y_2 和 F_{21} 分

别为②和③的输入特征图, b_{21} 和 b_{22} 分别为②和③的偏置值, ρ 为激活函数.

第3层为组合降维层, 第2层两个分支得到的特征图作为输入, 使用 Concat 函数进行组合拼接, 此时通道数为 $256 \times 2 = 512$.

第4层为输出层. 第3层的输出作为输入, 经过 1×1 卷积降维得到通道数为 256 的输出特征图.

2.3 联合损失函数

为解决 DeepLabV3+使用 CE-Loss (交叉熵损失) 出现部分细节信息丢失的问题, 本文算法联合使用 CE-Loss 和 Dice-Loss (类别不平衡函数) 作为损失函数. Dice-Loss 能减少过拟合情况的出现, 但会带来训练误差曲线混乱等问题, CE-Loss 梯度形式更优. 但当分割前景数量远远小于背景像素数量时, 会导致模型偏向于背景. 而本文设计的联合损失函数, 保留 CE-Loss 梯度形式的同时引入 Dice-Loss 优化前景目标和背景不平衡的问题.

CE-Loss 表示为真实标签和预测概率之间的差异, 数学表达式如式 (7) 所示, 其中 M 表示类别数, y_c 是一个向量, 元素只有 0 和 1 两种取值, p_c 表示预测样本属于类别 c 的概率. Dice-Loss 通常用于计算真实值和预测值的相似度, 数学表达式如式 (8) 所示, 其中 TP 表示被预测为正样本的正样本, FP 表示被预测为正样本的负样本, FN 表示被预测为负样本的正样本.

本文所提损失函数数学表达式如式 (9) 所示, 其中 μ 表示 CE-Loss 中的梯度系数, 即对不同类别的梯度分配, 取值为 $[0, 1]$, 当梯度系数达到最大值 1 时, 进行梯度回转, 将各类别的梯度系数都规范在 $[0, 1]$ 内, 以此平衡各类别的关注度. λ 表示 Dice-Loss 聚焦前景物体的 Dice 系数, 取值范围为 $[0, 1]$, 提升前景物体的 Dice 系数, 降低背景的 Dice 系数, 优化模型对前景物体的偏向. \circ 表示带有梯度系数的 CE-Loss 函数以及带有 Dice 系数的 Dice-Loss 函数的组合.

$$CELoss = - \sum_{c=1}^M y_c \log(p_c) \quad (7)$$

$$DiceLoss = 1 - \frac{2TP}{2TP + FN + FP} \quad (8)$$

$$Loss = \mu CELoss \circ \lambda DiceLoss \quad (9)$$

在 VOC 数据集上的训练变化如图 7 所示, 带圆点实线代表本文算法训练分割精度的变化, 带圆点虚线

代表本文算法损失函数变化, 实线代表 DeepLabV3+ 训练分割精度的变化, 虚线代表 DeepLabV3+ 损失函数变化. 由图 7 可知, DeepLabV3+ 和改进算法的损失函数值都随迭代次数的增加而逐渐稳定, 且算法精度也在逐步上升, 最终均收敛于 50 代左右.

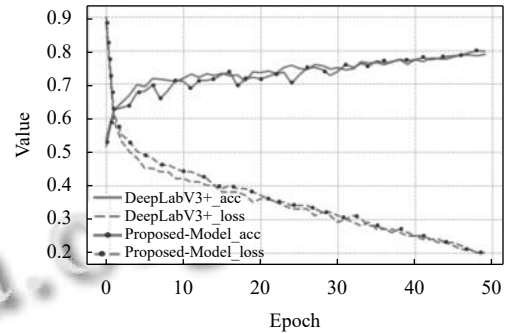


图 7 损失函数及分割精度在 VOC 数据集上的变化图

2.4 算法描述

本文基于双分支融合注意力机制的改进 DeepLabV3+ 图像分割算法的处理过程如算法 1.

算法 1. 双分支融合注意力机制的图像分割算法

- 1) 将固定大小的图片输入到改进算法中;
- 2) 特征提取器 ResNet101 网络对该图片进行初步提取, 得到图像有用信息生成 feature map, 分别传输给 DCNN 结构和 Decoder 结构;
- 3) 进入 DCNN 的 1/16 特征图分别进行标准卷积、空洞卷积、多注意力机制和全局平均池化提取特征. 得到的 6 层特征图拼接后输入双分支进一步提取特征, 再经过 1×1 卷积进行多尺度融合形成 deep feature map;
- 4) 进入 Decoder 中的 1/4 特征图通过 1×1 卷积形成 low-level feature map;
- 5) 采用双线性插值的方法对 deep feature map 进行上采样, 并与 low-level feature map 进行拼接. 然后通过深度可分离卷积得到信息更多的 feature map;
- 6) 对融合后的特征图再上采样, 输出 predicted, 完成图像分割.

3 实验分析

为了验证本文算法在图像分割方向取得了更好的性能, 本文在两个不同的数据集上进行对比实验, 包括 PASCAL VOC 2012 数据集^[9] 和 CityScapes 数据集^[10].

3.1 实验配置及评价指标

实验使用 Ubuntu 20.04 操作系统, NVIDIA Tesla V100 GPU, 32 GB 内存. 基于 PyTorch 框架分别在 PASCAL VOC 2012 和 CityScapes 数据集上训练、验证和测试. PASCAL VOC 2012 数据集和 CityScapes 数据集作为图像处理的公开数据集, 主要用于图像分类、

检测和分割。

本文使用图像分割中常见的3个评价指标,平均交并比($MIoU$)、频权交并比($FWIoU$)和类均准确率(MPA)来对模型的分割效果进行量化评价, $MIoU$ 、 $FWIoU$ 以及 MPA 值越高则代表模型分割效果越好。假设数据集包含背景在内共有 $k+1$ 个类别, p_{ij} 表示属于 i 类但被预测为 j 类的像素数量, p_{ji} 表示属于 j 类但被预测为 i 类的像素数量, p_{ii} 表示被正确预测的像素数量。 $MIoU$ 是模型对某一类别预测结果和真实值的交集与并集的比值,求和再平均的结果,数学表达式如式(10)所示:

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (10)$$

频权交并比 $FWIoU$ 是在 $MIoU$ 基础上,根据类别出现的频率分配不同的权重,数学表达式如式(11)所示:

$$FWIoU = \frac{1}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}} \sum_{i=0}^k \frac{\sum_{j=0}^k p_{ij} p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (11)$$

像素准确率 PA 是预测类别正确的像素数占总像素数的比例,累加求平均后得到类均准确率 MPA ,数学表达式如式(12)所示:

$$MPA = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}} \quad (12)$$

3.2 PASCAL VOC 2012 数据集

PASCAL VOC 2012 增强版数据集,训练图像有10582张,验证图像有1449张,测试图像有1456张,包括背景在内共有21个类别,数据量满足训练要求,如图8所示为VOC数据集典型图像。使用VOC数据集训练过程设置批处理大小为8,输入图片大小为 512×512 ,初始学习率设为0.0035,并采用多项式衰减,最终学习率降为0.0001,动量设为0.9,总共迭代50次。

为了更准确地衡量本文算法性能,并验证其在分割算法中的有效性和优越性,将经典算法、现有基于DeepLabV3+的改进算法与本文算法在PASCAL VOC 2012数据集上的表现进行量化对比,其中Backbone表示选用哪种骨干网络进行预训练,结果如表1所示。

相比其他分割算法,本文算法具有更高的 $MIoU$ 值和 $FWIoU$ 值。



图8 VOC数据集展示图

表1 不同算法在PASCAL VOC 2012数据集上的性能对比(%)

算法	Backbone	$MIoU$	$FWIoU$	MPA
FCN-8s	VGG16	62.20	79.70	71.04
DeepLabV2	ResNet101	77.69	87.44	85.50
DeepLabV3	ResNet101	78.05	88.56	85.95
DeepLabV3+	ResNet101	78.78	90.13	87.85
文献[11]	Xception	75.90	84.89	83.33
文献[12]	ResNet101	79.48	—	95.36
文献[13]	Xception	77.34	87.02	85.10
本文算法	ResNet101	79.92	90.73	88.58

表2为FCN-8s、DeepLabV2、DeepLabV3+和本文算法在PASCAL VOC 2012数据集上测试各类别的IoU值的对比。融入双分支特征提取和注意力机制的改进算法关注于前景细节信息,提升了小尺度目标的分割效果。由表中数据可以看出,本文算法在多数类别上的分割表现最优。

同时,本文也将DeepLabV3+与本文算法在PASCAL VOC 2012数据集的对比效果以图片形式展现,更直观地表达了本文算法的优越性,如图9所示。Original表示原图,即输入的图片,Labeled为标签图,评价指标是根据模型预测图与标签图之间的相似程度来计算。第1幅图片中,对牛进行分割,DeepLabV3+模型也将大致轮廓分割出来,但本文算法对趴在地上的牛头轮廓分割得更细致。第2幅图中本文算法与DeepLabV3+模型相比,对于飞机侧翼轮廓的预测更加精细。第3幅图中,DeepLabV3+模型对人的分割不够全面,

轮廓边缘分割得较粗糙, 本文算法预测图明显边缘更加平滑. 第4幅图中, DeepLabV3+模型将人腿预测为了马匹, 本文算法将人腿分割出来, 效果更加明显. 综上所述可得, 改进后的算法具有更好的分割效果, 能更精细地分割物体轮廓, 对于多物体图像, 效果也更优.

表2 PASCAL VOC 2012 数据集
各类别的IoU值对比(%)

种类	FCN-8s	DeepLabV2	DeepLabV3+	本文算法
人	84.4	85.9	82.2	83.9
猫	79.6	79.7	76.3	78.9
狗	82.9	83.7	85.9	86.1
羊	85.0	85.0	90.1	89.8
牛	81.7	81.9	89.0	91.0
马	79.8	80.1	76.3	78.7
鸟	80.2	79.6	82.1	83.2
自行车	61.9	63.6	63.9	66.6
摩托车	82.8	84.9	88.5	88.2
汽车	85.6	84.2	80.4	79.9
公交车	78.3	77.6	70.6	69.8
火车	73.2	72.8	79.4	80.5
飞机	89.1	90.0	85.9	86.3
船	51.1	57.8	59.2	59.6
餐桌	48.6	54.6	59.2	65.1
椅子	33.5	32.6	65.4	70.4
沙发	49.8	50.4	54.0	54.6
盆栽	47.5	51.8	66.2	70.9
瓶子	73.9	77.6	75.0	80.1
电视机	78.7	75.3	77.6	80.7
<i>MIoU</i>	62.2	77.69	78.78	79.92

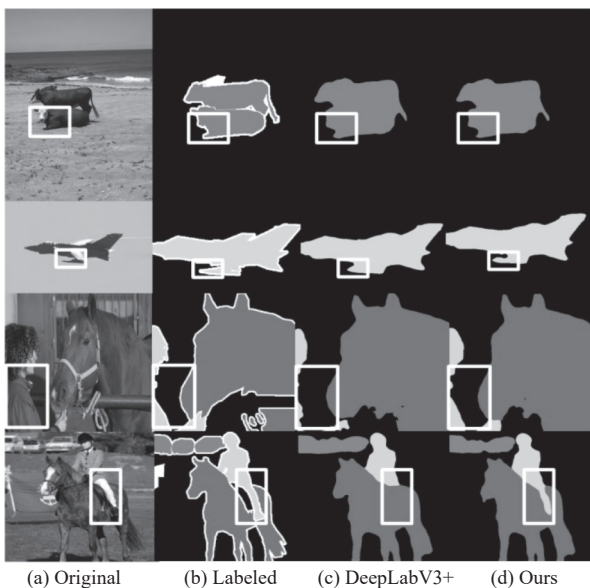


图9 与 DeepLabV3+在 VOC 验证集上的对比结果图

3.3 CityScapes 数据集

CityScapes 数据集为城市景观数据集, 训练图像有

2975 张, 验证图像 500 张, 测试图像 1525 张, 包括背景在内共有 19 个类别. 使用 CityScapes 数据集训练过程, 初始学习率设为 0.001, 总共迭代 200 次, 其他设置与 VOC 数据集一致.

为充分表明本文算法的普遍适用性, 进行泛化实验, 将经典算法、现有基于 DeepLabV3+改进算法与本文算法在 CityScapes 数据集上的表现进行量化对比, 结果如表3所示, 其中 Backbone 表示算法选用哪种骨干网络进行预训练.

表3 不同算法在 CityScapes 数据集上的性能对比(%)

算法	Backbone	<i>MIoU</i>	<i>FWIoU</i>	<i>MPA</i>
FCN-8s	VGG16	62.10	85.78	74.76
DeepLabV2	ResNet101	66.60	89.80	75.12
DeepLabV3	ResNet101	67.12	91.87	75.70
DeepLabV3+	ResNet101	67.94	92.11	75.93
文献[11]	Xception	67.05	91.54	75.88
文献[12]	ResNet101	68.14	92.47	76.80
文献[13]	Xception	66.83	90.25	75.52
本文算法	ResNet101	68.58	92.65	77.44

本文也将 DeepLabV3+与本文算法在 CityScapes 数据集的对比效果以图片形式展现, 如图10所示. 其中 Original 表示原图, Labeled 为标签图. 根据图中标示可以看出, 本文算法比 DeepLabV3+分割得更为细致, 再次证明了其可行性与有效性.

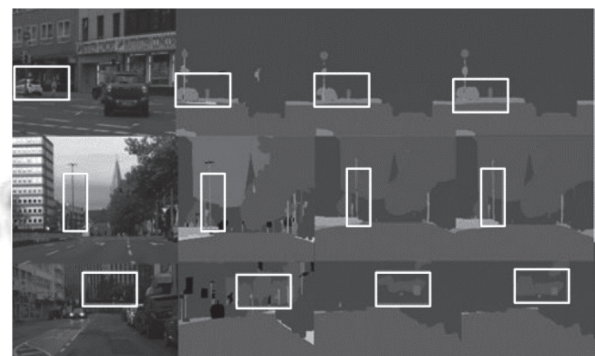


图10 与 DeepLabV3+在 CityScapes 验证集上的对比结果图

3.4 消融实验

为了验证是新增注意力机制和双分支特征提取同时在发挥作用, 本文也在 VOC 数据集和 CityScapes 数据集上做了消融实验, 如表4和表5所示, 其中 Params 为评估模型计算成本的量化度量指标, 即参数量, 模型中的参数量仅针对模型参数而言, 与使用数据集无关, 因此该指标在 VOC 数据集和 CityScapes 数据集上表现一致.

表4 VOC数据集的消融实验结果

注意力机制	双分支	Params (G)	MIoU (%)	FWIoU (%)	MPA (%)
—	—	41.28	78.78	90.13	87.85
√	—	42.40	78.89	90.22	87.17
—	√	41.92	79.10	90.31	88.41
√	√	41.60	79.92	90.73	88.58

表5 CityScapes数据集的消融实验结果

注意力机制	双分支	Params (G)	MIoU (%)	FWIoU (%)	MPA (%)
—	—	41.28	67.94	92.11	75.99
√	—	42.40	68.35	92.24	76.00
—	√	41.92	68.37	92.46	76.26
√	√	41.60	68.59	92.65	77.44

如表4所示,在VOC数据集上只添加注意力机制, *MIoU*、*FWIoU*和*MPA*分别提升0.11%、0.09%和0.32%,属于正常误差范围,可以说注意力机制并未发挥作用。只加入双分支特征提取, *MIoU*、*FWIoU*和*MPA*分别提升0.32%、0.18%和0.56%,有小幅度提升,由于双分支扩大了感受野,能获取到更丰富的特征。而同时加入注意力机制和双分支, *MIoU*、*FWIoU*和*MPA*分别提升1.14%、0.6%和0.73%。这是因为加入注意力机制后模型的参数量增加,网络结构也变得复杂,而加入双分支后,特征提取分支数增加,卷积核带有注意力权重的特征做池化操作以减少参数量,使得注意力机制发挥最佳效果;同时注意力机制中的1×1卷积能够对通道数降维,减少双分支产生的参数量。

如表5所示,在CityScapes数据集上只添加注意力机制, *MIoU*、*FWIoU*和*MPA*分别提高0.41%、0.13%和0.01%,只加入双分支特征提取, *MIoU*、*FWIoU*和*MPA*分别提高0.43%、0.35%和0.27%。而同时加入注意力机制和双分支特征提取, *MIoU*、*FWIoU*和*MPA*分别提高0.65%、0.54%和1.45%,两个消融实验表明,只有注意力机制和双分支同时起作用才能提升分割效果。

4 结论与展望

DeepLabV3+中膨胀率较大的空洞卷积会引起图像信息丢失和细节分割不准确的问题,本文引入注意力机制,结合通道和空间注意力权重捕捉丢失的细节信息,再输入到双分支特征提取机制利用分段卷积以扩大感受野获取更丰富的特征并降低注意力机制的参数量,最后融合两种损失函数减少前景与背景关注不平衡的可能。另外在实验过程中发现,ResNet101作为骨干网络的分割效果还是不够精准,在未来工作中,我们将进一步去优化或者替换骨干网络,尽可能保证内

存不上升的情况下,达到更好的分割效果。

参考文献

- 黄鹏,郑淇,梁超. 图像分割方法综述. 武汉大学学报(理学版), 2020, 66(6): 519–531. [doi: 10.14188/j.1671-8836.2019.0002]
- Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640–650. [doi: 10.1109/TPAMI.2016.2572683]
- Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention*. Munich: Springer, 2015. 234–241.
- Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481–2495. [doi: 10.1109/TPAMI.2016.2644615]
- Chen LC, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs. *Computer Science*, 2014, (4): 357–361.
- Chen LC, Papandreou G, Kokkinos I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834–848. [doi: 10.1109/TPAMI.2017.2699184]
- Chen LC, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587, 2017.
- Chen LC, Zhu YK, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the 15th European Conference on Computer Vision*. Munich: Springer, 2018. 833–851.
- Lee J, Yi JH, Shin C, et al. BBAM: Bounding box attribution map for weakly supervised semantic and instance segmentation. *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville: IEEE, 2021. 2643–2651.
- Fan MY, Lai SQ, Huang JS, et al. Rethinking BiSeNet for real-time semantic segmentation. *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville: IEEE, 2021. 9711–9720.
- 杨贞,彭小宝,朱强强,等. 基于DeepLabV3 Plus的自适应注意力机制图像分割算法. *计算机应用*, 2022, 42(1): 230–238.
- 席一帆,孙乐乐,何立明,等. 基于改进DeepLabV3+网络的语义分割. *计算机系统应用*, 2020, 29(9): 178–183. [doi: 10.15888/j.cnki.csa.007541]
- Bai YQ, Zheng YF, Tian H. Semantic segmentation method of road scene based on DeepLabV3+ and attention mechanism. *Journal of Measurement Science and Instrumentation*, 2021, 12(4): 412–422.

(校对责编:牛欣悦)