文章编号:2095-7386(2023)06-0087-06
DOI:10.3969/j.issn.2095-7386.2023.06.012

一种卷积注意力和混合池化结合的DeepLabV3+语义分割算法

孔祥明,付暄然,李 丹,文国知
(武汉轻工大学 电气与电子工程学院,武汉 430023)

摘 要:为了提升图像分割的精确度,解决特征细节易丢失和目标边界模糊等问题,提出了一种结合卷积注意力和混合池化的DeepLabV3+语义分割算法。用ResNet101网络代替原来的骨干网络Xception,采用卷积注意力机制模块来增强卷积神经网络对图像的关注能力,使用混合池化对卷积层中提取的特征进行挑选,加入深度可分离卷积以减少卷积的参数量,同时利用交叉熵作为损失函数。实验表明,所提算法使图像分割的精确度得到了提高,特征细节信息的损失明显减少,图像分割的平均交并比提升了3.6%。该算法有效解决了边界模糊的问题,为提高目标识别的准确度提供了新的思路。

关键词:DeepLabV3+;图像分割;卷积注意力;混合池化
中图分类号:TP 391 **文献标识码**:A

Research on DeepLabV3+ image semantic segmentation algorithm combining convolutional attention and mix pooling

KONG Xiangming, FU Xuanran, LI Dan, WEN Guozhi
(School of Electrical and Electronic Engineering, Wuhan Polytechnic University, Wuhan 430023, China)

Abstract: In order to improve the accuracy of image segmentation and solve the problems of easy loss of feature details and blurred target boundary, this paper proposes a DeepLabV3+ semantic segmentation algorithm combining convolutional attention and mix pooling. The ResNet101 network is used to replace the original backbone network Xception, the convolutional attention mechanism module is used to enhance the ability of convolutional neural network to focus on images, the features extracted from the convolutional layer are selected by mixing pooling, the depth separable convolution is added to reduce the number of convolution parameters, and the cross entropy is used as a loss function. Experiments show that the proposed algorithm improves the precision of image segmentation, significantly reduces the loss of feature detail information, and increases the average intersection union ratio of image segmentation by 3.6%. The algorithm effectively solves the problem of boundary ambiguity and provides a new idea for improving the accuracy of target recognition.

Key words: DeepLabV3+; image segmentation; convolutional attention; mix pooling

---

收稿日期:2023-10-09.
作者简介:孔祥明(1998—),男,硕士研究生,E-mail:1016197756@qq.com.
通信作者:文国知(1973—),男,博士,副教授,E-mail:wwenguozhi@whpu.edu.cn.
基金项目:湖北省教育厅科学技术研究计划青年项目(D20201603,Q20210608);武汉轻工大学校级科研项目(2021Y37).

1 引言

图像分割是计算机视觉技术中的关键步骤，在计算机视觉领域有着重要的地位，图像识别、医学影像、人工智能等领域都离不开图像分割技术。图像分割是指将图像分成若干互不重叠的子区域，使得同一个子区域内的特征具有一定相似性而不同子区域间的特征呈现较为明显的差异。图像分割是图像识别、场景解析、对象检测等任务的预处理，是计算机视觉中的一项基础任务[1]。传统的图像分割任务大多是基于机器学习方法，且是基于阈值化的图像分割算法[2-4]，通常将图像灰度直方图按不同的灰度阈值进行分类，以达到图像分割的目的，但该方法易受图像噪声的影响，从而导致分割精度和鲁棒性较低[5]。

基于深度学习的图像分割利用卷积神经网络(CNN)，通过端到端的方式推理每个像素的语义信息并实现有意义图形区域的分类，在特征学习和表达能力方面优势明显[6]。由于 CNN 技术的飞速发展，基于深度学习的图像分割成为近年来的研究热点，各种基于 CNN 的语义分割网络相继被提出，比如 FCN[7]、PSPNet[8]、U-Net[9]、DeepLabv3＋[10]等，目前较为热门的就是 DeepLabv3＋。DeepLabv3＋带有空洞卷积的空间金字塔池化(ASPP)，能利用不同空洞率的空洞卷积扩大感受野来获取多尺度信息，但是其图像分割的精确度不高，易丢失许多特征细节信息，导致目标边界模糊，分割效果不好。

针对上述问题，马冬梅等[11]提出一种改进 DeepLabV3＋的高效语义分割算法，通过 MSP 拓展 ASPP 模块，捕获了丰富的多尺度信息；引入 ECA 模块，使网络的解码器能够有选择性地提取更有用的特征，更好地恢复目标边界信息，相比原始模型 MIoU 提升了 1.94%。张小国等[12]提出了一种融合通道注意力机制和空间注意力机制的 FDA-DeepLab 图像语义分割网络，相比原始模型的 MIoU 值提高了 1.2%，多尺度输入时 MIoU 值提高了 1.9%。上述研究表明，引入注意力机制对 DeepLabV3＋的性能确实有提升，但图像分割的精确度依然不高。基于此，笔者在 DeepLabV3＋的基础上，提出了一种结合卷积注意力[13]和混合池化的 DeepLabV3＋语义分割算法，利用卷积注意力机制来提高卷积神经网络对图像信息的关注能力，而混合池化结合了最大池化和平均池化的方法，可以很好地解决过拟合的问题。同时采用交叉熵(Cross Entropy)损失函数，通过相似性度量，使模型预测结果更接近真实标记，从而提高图像分割的精确度。交叉熵损失函数可以避免学习率的下降，更好地优化模型参数。

2 改进的语义分割算法

2.1 卷积注意力机制

卷积注意力机制模块(CBAM)由 Woo S 等[13]提出，融合了通道注意力和空间注意力的模式(见图 1)，用以提高卷积神经网络对图形信息的注意水平。

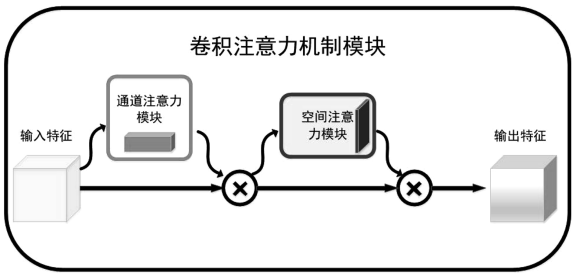

图 1 卷积注意力机制模块

Fig. 1 CBAM

通道注意力模块如图 2 所示。使用全局平均池化(GAP)和全局最大池化(GMP)分别来获得所有通道的全局统计数据，并利用两层全连接层来学习通道的权重。接着，系统会把经过处理后生成的两个结果进行相加，再使用 Sigmoid 函数将权重归一化到 0 和 1 之间，对每个通道进行缩放。最后，将缩放后的通道特征与原始特征相乘，以产生具有增强通道重要性的特征。

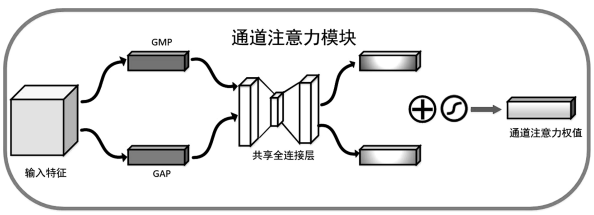

图 2 通道注意力模块

Fig. 2 Channel attention module

空间注意力模块如图 3 所示。使用最大池化(Max Pooling)和平均池化(Avg Pooling)来获取每个空间位置的最大值和平均值。通过最大池化和平均池化后，可获得两张特征图，再将两张特征图加以

拼接，并通过卷积层和 Sigmoid 函数来学习每个空间位置的权重。最后，将权重应用于特征图上的每个空间位置，以产生具有增强空间重要性的特征。

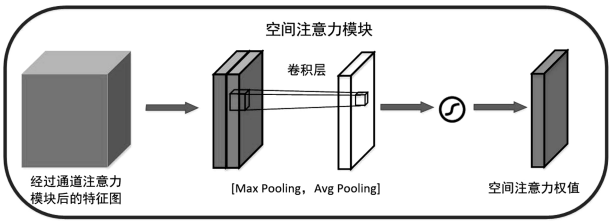

图 3　空间注意力模块

Fig. 3　Spatial attention module

笔者采用的混合池化(Mix Pooling)结合了最大池化与平均池化两种池化方法，在训练过程中随机选择其中一种池化方法来预防过拟合。计算公式如式(1)：

$$y_{kij}=\lambda\cdot\max_{\langle p,q\rangle\in R_{ij}}x_{kpq}+(1-\lambda)\cdot\frac{1}{|R_{ij}|}\sum_{\langle p,q\rangle\in R_{ij}}x_{kpq} \tag{1}$$

式中，λ 是 0 或 1 的随机值，加号左边为最大池化，右边为平均池化。

2.2　网络结构

文中所提算法主要是改进 DeepLabV3＋的网络结构，采用 ResNet101[14] 模型作为骨干网络，其结构较为简单、易于训练，拥有更好的性能，能够解决梯度消失的问题，同时添加了注意力机制模块和混合池化模块。该算法的网络结构如图 4 所示。

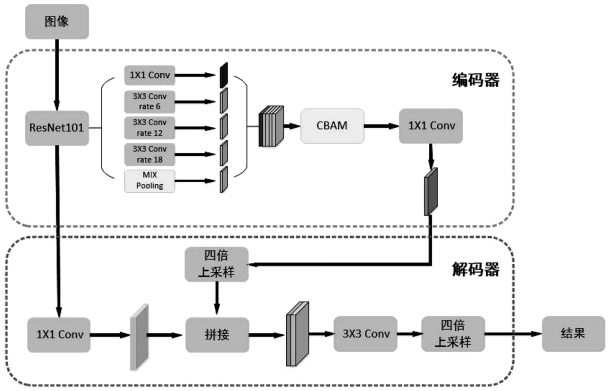

图 4　改进的 DeepLabV3＋网络结构图

Fig. 4　Improved DeepLabV3＋ network diagram

在编码器部分，输入图像经过主干网络 ResNet101，通过主干网络中的深度可分离卷积层完成特征提取。采集后的特征图像进入改进的 ASPP 模块，通过 ASPP 模块内 1×1 卷积、空洞卷积层(空洞率大小依次为 6、12、18)和 Mix Pooling 处理后拼接融合；然后经过 CBAM 模块，增强卷积神经网络对目标的关注和细节的捕捉能力；最后经过 1×1 卷积进行降维。在解码器部分，经过骨干网络 ResNet101 提取出的低级特征图作为解码器中的输入特征，将提取出来的特征图通过 1×1 卷积后，与编码器中通过四倍上采样获取的高级特征拼接。最后利用 3×3 卷积和四倍上采样操作，使分割结果恢复到原来的图像尺寸，得到分割效果图。

3　实验方法和评价指标

3.1　实验数据及参数配置

本次实验的数据集为 PASCAL VOC2012，来自计算机视觉挑战赛 PASCAL VOC，数据集有1 464 张训练集图片、1 449 张验证集图片，共 21 个类别。实验运行在 PyTorch 深度学习框架上，操作系统是 Windows10，显卡是 NVIDIA RTX3090(24 GB)，CPU 是 Intel Platinum8350C。采用 SGD 动量优化器，将学习率设置为 0.01，权重衰减为 $1e^{-4}$，动量为 0.9，学习策略为 poly，迭代训练 400 次。

3.2　评价指标

实验主要通过平均交并比(MIoU)、平均像素精确度(MPA)以及损失值(Loss)来考量算法性能。MIoU 是将实际值与预测值的交集和并集进行计算得到的比例，用以展示预测成果和原本图像标签的匹配程度。累加每一个种类的 IoU 并求平均，最后得到的数值就是最终的 MIoU，计算公式如式(2)：

$$MIoU=\frac{1}{n+1}\sum_{i=0}^{n}\left(\frac{p_{ii}}{\sum_{j=0}^{n}p_{ij}+\sum_{j=0}^{n}p_{ji}-p_{ii}}\right) \tag{2}$$

式中，n 指的是标签识别的类型，$n+1$ 则表述的是包括背景在内的所有类型。p_{ji} 指像素真实值为 j 但预测成 i 的数量；p_{ij} 指像素的真实值为 i 但是预测为 j 的数量；p_{ii} 指像素真实值和预测值都为 i 的数量。关于 MIoU 的数值，它在 0 至 1 之间变动，数值大小可以直接反映图像分割的精确度，数值越接近 1，其精确度越高。

计算每一类别中被准确分类的像素比例，然后求取所有类别的平均数，这就是 MPA。MPA 的值在 0 至 1 之间变动，其值越接近 1，图像分类准确性越好，计算公式如式(3)：

$$MPA=\frac{1}{n+1}\sum_{i=0}^{n}\frac{p_{ii}}{\sum_{j=0}^{n}p_{ij}} \tag{3}$$

4 结果与分析

各种组合方式的分割效果如表 1 所示。对比表中①和③两组数据后发现，把 DeepLabV3＋主干网络 Xception 换成 ResNet101，MIoU 值上升了 2.6%；对比③和④的结果发现，加入CBAM后，图像分割的精确度得到了提升，MIoU 值增长了 0.5%；对比④和⑥的结果发现，CBAM 和 Mix Pooling 的叠加使用让 MIoU 再上升 0.3%；从⑥和⑦的结果对照来看，采用 Cross Entropy 方式使得 MIoU 的数值上升了 0.2%；通过②和⑥的实验结果可以看出，主干网络选择 ResNet101 时，组合的效果更好。

表 1 不同组合方法的分割效果

Table 1 Segmentation effects of different combinations methods

组数	DeepLabV3＋	ResNet101	卷积注意力机制	混合池化	交叉熵损失	平均交并比/%
①	√					72.0
②	√		√	√		72.7
③		√				74.6
④		√	√			75.1
⑤		√		√		74.7
⑥		√	√	√		75.4
⑦		√	√	√	√	75.6

比较改进前和改进后 DeepLabV3＋算法的 MPA 来验证文中的算法性能，具体的结果见表 2。可以看出，本文中的算法不仅使 MIoU 的值提升了 3.6个百分点，MPA 也增加了 1.4 个百分点。

表 2 对比 DeepLabV3＋网络改进前后的各项指标变化情况

Table 2 Compare the changes of various indicators before and after the improvement of the DeepLabV3＋ network

算法	平均像素精确度/%	平均交并比/%
DeepLabV3＋	82.5	72.0
文中算法	83.9	75.6

为了展示改进算法的优越性，绘制了改进前后算法的平均交并比曲线图和损失曲线图。平均交并比曲线图如图 5 所示，虚线为 DeepLabV3＋算法的 MIoU 值变化，实线为本文所提算法的 MIoU 变化。迭代 50 次后，两种算法的 MIoU 均趋于平稳，略有上升，DeepLabV3＋的最终效果达到 72%。文中模型迭代 10 次后，MIoU 明显高于 DeepLabV3＋，最终达到 75.6%。损失曲线图如图 6 所示，同样迭代 10 次后，文中算法的损失值明显低于 DeepLabV3＋，DeepLabV3＋的最终损失值在 0.06 左右，而文中算法的最终损失值在 0.02 左右，改进的算法损失值有明显下降。

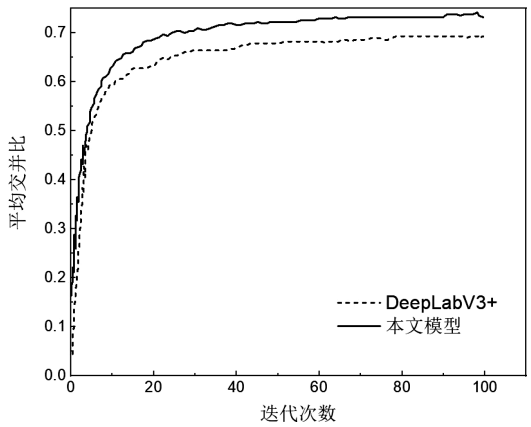

图 5 改进前后的平均交并比曲线图

Fig. 5 Improved the before and after average intersection and union ratio curves

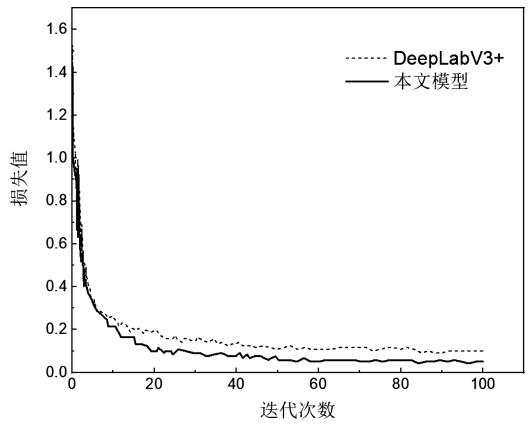

图 6 改进前后的损失曲线图

Fig. 6 Improved before and after loss graphs

为了更直观地对比 DeepLabv3＋改进前后的分割性能，对下面四组分割图进行可视化对比分析，如图 7 所示。从第一行中可发现，原模型对摩托车的后车轮分割不清晰，而文中模型对分割效果有明显的改进。第二行受杂草影响，原模型对绵羊手臂和尾巴部分的分割效果粗糙，而文中模型不受杂草影响，分割效果较好。第三行远处的飞机由于画质模糊且目标较小，原模型存在漏分割的现象，而文中模型的分割效果虽然不如标签一样精准，但还是能够正确分割出来。第四行的背景环境较为复杂，导致原本模型分割效果极差，鸟腿丢失且目标四周也有部分丢失，本文模型虽然鸟腿部分稍有丢失，但整体分割效果比原模型有明显提升。实验结果表明，文中的算法明显改善了目标边界模糊的问题。

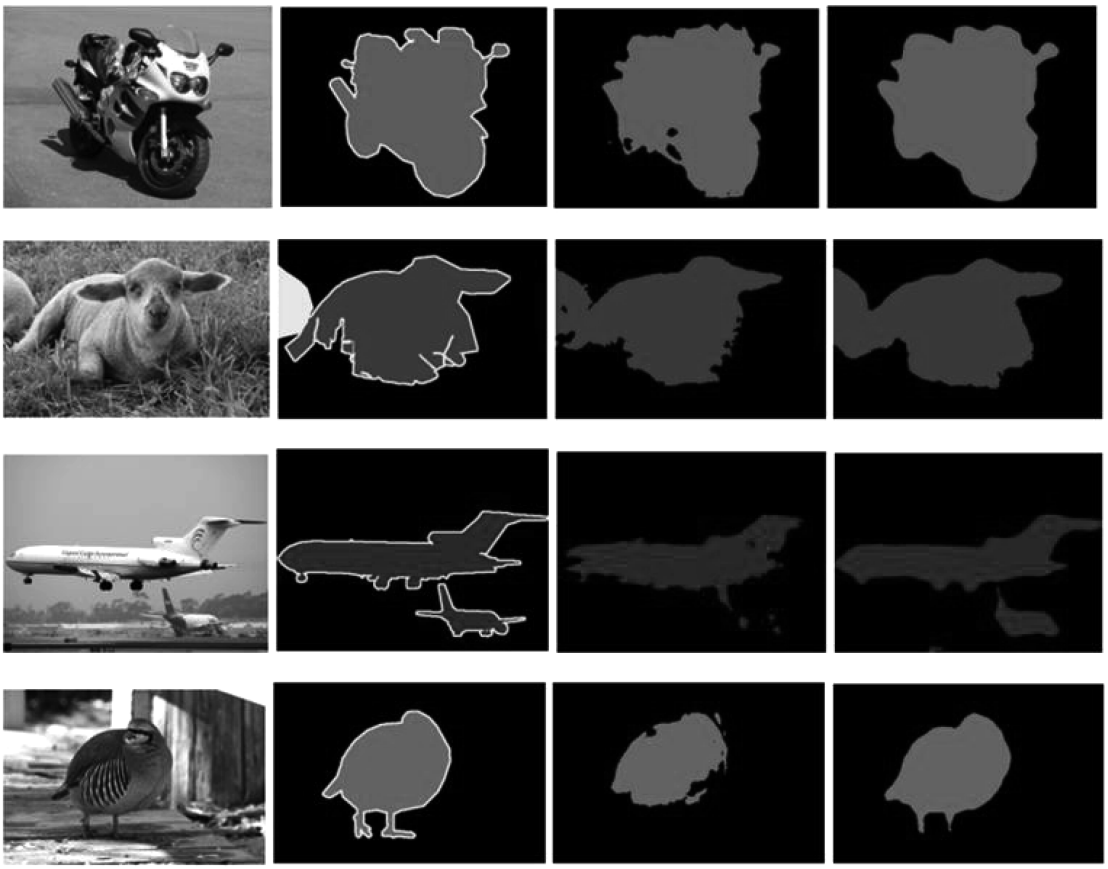

图 7　PASCAL VOC2012 分割结果图

Fig. 7　PASCAL VOC2012 segmentation result graph

5　结论

笔者介绍了一种基于 DeepLabV3＋的图像语义分割算法，结合了卷积注意力机制与混合池化方法，通过引入 CBAM 注意力机制，融合了通道注意力模块和空间注意力模块，推断出注意力图，然后与输入的特征图相乘以完成自动调节的特征优化，获取了更高质量的特征图。混合池化结合了最大池化和平均池化方法，可以更有效地从卷积层筛选特征。该算法还使用了深度可分离卷积代替空洞卷积，大幅度减少了参数量、提高了计算速度，此外，还运用了交叉熵损失函数来防止学习率下降。实验结果表明，与 DeepLabV3＋相比，文中的算法明显提高了图像分割的精确度，图像分割的平均交并比提升了 3.6%。虽然算法的性能有所提高，但在复杂背景下的分割效果不是很理想，后续将进行进一步的研究。

参考文献：

[1]　周莉莉，姜枫. 图像分割方法综述研究[J/OL]. [2016-11-24]. http://www.arocmag.com/article/02-2017-07-064.html.

[2]　Ivanovs M，Ozols K，Dobrajs A，et al. Improving Semantic Segmentation of Urban Scenes for Self-driving Cars with Synthetic Images [J]. Sensors (S1424-8220)，2022，22 (6)：

2252.

[3] Kontschieder P, Bulo S R, Bischof H, et al. Structured Class-labels in Random Forests for Semantic Image Labelling[C]//2011 International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011: 2190-2197.

[4] van den Heuvel M, Mandl R, Hulshoff Pol H. Normalized cut group clustering of resting-state FMRI data[J]. PLoS One, 2008 Apr 23; 3(4): e2001.

[5] 周华平，邓彬. 融合多层次特征的 deeplabV3＋轻量级图像分割算法[J/OL]. 计算机工程与应用，2023-09-18: 1-9. wafang data. com. cn.

[6] 张鑫，姚庆安，赵健，等. 全卷积神经网络图像语义分割方法综述[J]. 计算机工程与应用，2022，58(8): 45-57.

[7] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [J] IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.

[8] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 6230-6239.

[9] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C]// Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. Cham: Springer, 2015: 234-241.

[10] Chen L C, Zhu Y, Papandreou G, et al. Encoder-Decoder with Atrous Separable Convolution for Se-mantic Image Segmentation [C]// Computer Vision - ECCV 2018. Cham: Springer, 2018: 833-851.

[11] 马冬梅，李鹏辉，黄欣悦，等. 改进 DeepLabV3＋的高效语义分割[J]. 计算机工程与科学，2022，44(04): 737-745.

[12] 张小国，丁立早，刘亚飞，等. 基于双注意力模块的 FDA-DeepLab 语义分割网络[J]. 东南大学学报(自然科学版)，2022，52(06): 1145-1151.

[13] Woo S, Park J. CBAM: Convolutional block attention module[C]// Computer Vision - ECCV 2018. Berlin: Springer, 2018: 3-19.

[14] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE, 2016: 770-778.