

柯寅, 陈丹. 基于改进 DeepLabv3+网络的图像语义分割方法[J]. 智能计算机与应用, 2025, 15(4): 17-24. DOI: 10.20169/j.issn.2095-2163.250403

基于改进 DeepLabv3+网络的图像语义分割方法

柯寅, 陈丹

(福州大学 电气工程与自动化学院, 福州 350108)

摘要: 针对 DeepLabv3+网络对多个目标图像的分割性能不够优越、容易丢失图像的细节信息、产生分割断裂等问题, 提出一种基于改进 DeepLabv3+网络的图像语义分割方法。首先, 在空洞空间金字塔池化 ASPP 模块中添加空洞卷积分支, 并将不同空洞卷积分支与输入特征拼接, 实现不同感受野下的多通道特征信息融合; 其次, 在 ASPP 模块后引入 PSA 注意力机制, 减少特征提取时的信息损失。实验结果表明, 与原 DeepLabv3+网络相比, 改进 DeepLabv3+网络模型在 PASCAL VOC 2012 数据集上 *MIoU* 总体提高了 1.62%, 在自制数据集上 *MIoU* 总体提高了 0.63%, 验证了改进 DeepLabv3+网络模型的良好分割性能以及在现实场景下的可行性。

关键词: 语义分割; DeepLabv3+; 特征信息融合; PSA 注意力机制

中图分类号: TP389.1 **文献标志码:** A **文章编号:** 2095-2163(2025)04-0017-08

Image semantic segmentation method based on improved DeepLabv3+ network

KE Yin, CHEN Dan

(College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China)

Abstract: Aiming at the problem that the image segmentation performance of DeepLabv3+ network is poor for multiple targets, image details is easy to be lost and segmentation breaks could be caused, an image semantic segmentation method based on improved DeepLabv3+ network is proposed. Firstly, the void convolution branches are added to the atrous space pyramid pooling ASPP module, and different atrous convolution branches are spliced with input features to achieve multi-channel feature information fusion under different receptive fields. Secondly, PSA attention mechanism is introduced after ASPP module to reduce the information loss during feature extraction. The experimental results show that compared with the original DeepLabv3+ network, the improved DeepLabv3+ network model has an overall increase of 1.62% in *MIoU* on PASCAL VOC 2012 dataset and 0.63% in *MIoU* on self-made dataset. This simulation verifies the good segmentation performance of DeepLabv3+ network model and the feasibility in real scenarios.

Key words: semantic segmentation; DeepLabv3+; feature information fusion; PSA attention mechanism

0 引言

图像语义分割技术作为计算机视觉技术的重要研究内容之一, 在自动驾驶、人脸分割、地质检测、精准农业等领域有着广泛的应用。现如今, 随着深度学习技术的迅速发展, 国内外研究学者将深度学习技术与语义分割技术结合, 提出了一系列基于深度学习的语义分割方法, 并取得了显著的成功。经典的语义分割网络主要有 FCN^[1]、DeepLabv1^[2]、U-net^[3]、SegNet^[4]、DeepLabv2^[5]、RefineNet^[6]、DeepLabv3^[7]、DeepLabv3+^[8]等。其中, FCN 网络首

次将端到端的卷积神经网络应用到语义分割中, 实现了像素级的分割任务。该方法接受来自网络更深处的深层特征, 因而具有更高级别的语义信息, 且可以接受任意大小的输入图像^[9], 比传统方法更加高效。但由于 FCN 中池化层的存在, 降低了特征图的分辨率, 丢失了空间上的详细信息^[10]。DeepLabv1 网络通过引入空洞卷积 (atrous/dilated convolution) 来增加感受野的大小, 并使用全连接条件随机场优化细节丢失问题。DeepLabv2 网络首次提出了空洞空间金字塔池化 (Atrous Spatial Pyramid Pooling, ASPP) 模块, 利用多个不同采样率空洞卷积分支

基金项目: 国家自然科学基金面上项目 (61973085); 福建省自然科学基金面上项目 (2022J01114)。

作者简介: 陈丹 (1977—), 女, 博士, 副教授, 主要研究方向: 基于网络的机器人遥操作系统, 机器人伺服控制, 预测控制, 机器视觉。

通信作者: 柯寅 (1998—), 男, 硕士研究生, 主要研究方向: 计算机视觉。Email: 210127006@fzu.edu.cn。

收稿日期: 2023-10-08

获得更丰富的多尺度上下文信息。DeepLabv3 网络通过调整空洞卷积率、添加批量归一化 (Batch Normalization, BN) 层和全局平均池化等方式改进 ASPP 模块。DeepLabv3+网络使用改进的 Xception 模块作为网络的骨干,减少参数量,同时结合“编码器-解码器”结构,进一步提高了网络的分割性能。

DeepLab 系列网络能够较好地克服下采样导致的细节丢失、卷积导致的分辨率降低等问题,具备较好的泛化能力和特征提取能力。但 DeepLabv3+网络的参数较多、计算量大,容易丢失一部分位置处的像素信息,降低特征图的局部一致性^[11]。同时,DeepLabv3+网络对多个目标,尤其是对相似目标容易出现误判,而且在输出图像中还存在分割断裂以及分割空洞等问题^[12]。

针对 DeepLabv3+网络存在的问题,大量学者对其进行了改进与研究。文献[13]使用 Group Normalization (GN) 替换原 DeepLabv3+网络中的 BN 层,使得网络能在显存受限的情况下以小 batch size 训练模型,同时保持较好的分割精度。文献[14]提出了 MS-DeepLabv3+网络,通过在编码区增加全特征提取通道、在解码区增加多尺度特征提取通道来聚合多尺度特征和捕捉低层次特征,并通过引入卷积注意力机制模块 (Convolutional Block Attention Module, CBAM) 和 ECA (Efficient Channel Attention) 注意力模块提高算法的准确性。文献[15]在原 DeepLabv3+网络的主干网络 ResNet101 中引入坐标注意力机制 (Coordinate Attention), 并设计了加权 Dice 损失和交叉熵损失 (Binary Cross Entropy Loss, BCE Loss) 结合的优化损失函数来优化样本不均衡的问题,提高了分割精度。文献[16]将 Transformer 引入 DeepLabv3+模型,提出了一种用于无人机影像农作物分类的并行分支结构,提高了作物全局信息的提取能力。文献[17]通过在 ASPP 模块中添加空洞卷积分支、调整空洞卷积率来提高网络对目标边缘的分割能力,并在 DeepLabv3+网络两次上采样前添加卷积注意力机制模块 CBAM, 以此提高图像像素信息的利用率。文献[18]将 ECA 模块插入原 DeepLabv3+骨干网络的残差模块中,增加了基于特征金字塔网络的特征融合分支,融合了不同级别的特征图,取得了更好的分割性能。文献[19]提出一种多尺度输入的 DeepLabv3+网络,使模型的参数规模大大减小,在提高分割精度和节省计算资源方面取得了较好的效果。文献[20]在 DeepLabv3+网络中引入了轻量级的注意力机制,减少网络深度带来的参数

过多和特征冗余问题,提高复杂图像的分割精度,并采用递归级联方法优化高频和低频特征的融合,挖掘潜在的语义信息,减少时间和空间的消耗。

针对以上研究内容,本文提出一种基于改进 DeepLabv3+网络的图像语义分割方法。首先,对原 DeepLabv3+网络模型中的 ASPP 模块进行改进,实现不同感受野下的多通道信息融合,提高网络的特征提取能力;其次,在 ASPP 模块后添加 PSA^[21] (Polarized Self-Attention) 注意力机制,构成 ASPP-PSA 模块,用于减少特征提取时的信息损失;最后,本文在 PASCAL VOC 2012^[22]数据集上设计消融实验与对比实验,并在自制的数据集上设计对比实验。实验结果表明,本文改进 DeepLabv3+网络模型对多个目标图像具有更好的分割效果,能够增强对图像细节信息的提取能力,不仅较好地改善了分割断裂等问题,在实际场景中也依然具有良好的分割性能。

1 DeepLabv3+网络结构原理

DeepLabv3+网络结构分为 Encoder 和 Decoder 两个部分。其中,Encoder 部分主要用于特征提取,采用 Xception 作为主干网络,然后是带有空洞卷积的 ASPP 模块,用于引入多尺度信息。输入图像先经过主干网络提取特征信息,其中的浅层特征传入 Decoder 模块,深层特征传入 ASPP 模块。深层特征经过 ASPP 模块处理后得到 5 个大小相同且通道数均为 256 的特征层,然后将其进行拼接并经过 1×1 卷积模块处理,得到高层特征层。Decoder 部分主要用于特征融合和分类,是将浅层特征与高层特征进一步融合,提升边界信息的分割精度。对 Encoder 中获得的高层特征层进行 4 倍上采样处理调整成大小为 128×128 的高层特征层,再与经过 1×1 卷积模块处理得到的浅层特征层拼接,得到同时包含深层特征与浅层特征信息的融合特征层,然后经过 3×3 卷积模块进行特征提取,最后进行 4 倍上采样处理,得到与原图相同大小的输出图像。

2 DeepLabv3+网络结构改进

为提高网络的特征提取能力,减少特征提取时的信息损失,本文对 DeepLabv3+网络做以下 2 点改进:

(1) 对原 DeepLabv3+网络模型中的 ASPP 模块进行改进。在 ASPP 模块中增加一个空洞卷积率为 3 的空洞卷积分支,并依次将空洞卷积率为 3、6、12 分支的输出与传入 ASPP 的深层特征融合,分别作为空洞卷积率为 6、12、18 分支的输入,实现不同感受野下

的多通道信息融合,提高网络的特征提取能力。

(2) 在改进 ASPP 模块后引入 PSA 注意力机制,构成 ASPP-PSA 模块,将通道和空间分支上的

注意力应用于拼接后的特征层上,保持特征层中较高维度的信息,减少信息损失。

改进后的 DeepLabv3+网络结构如图 1 所示。

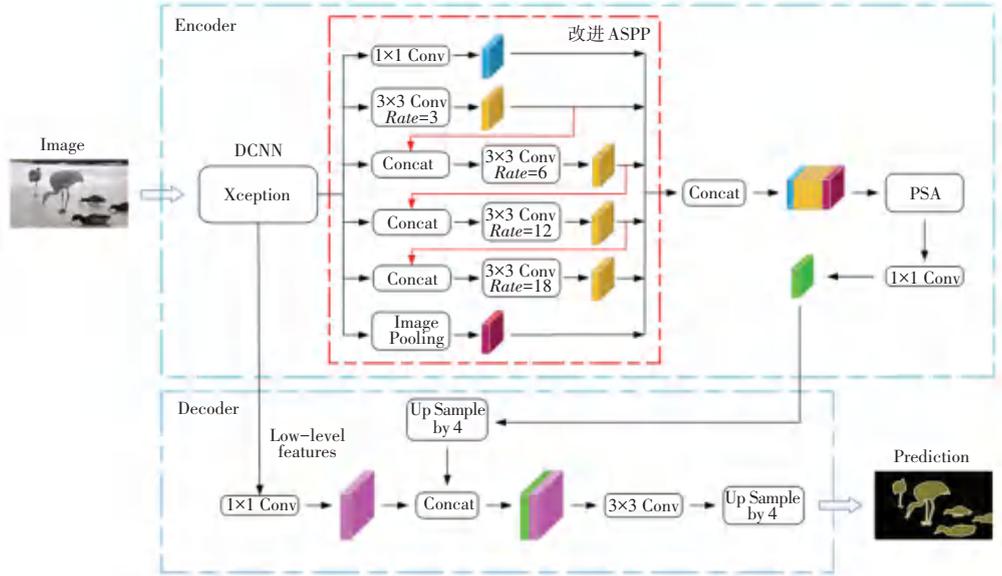


图 1 改进 DeepLabv3+网络结构图

Fig. 1 Improved DeepLabv3+ network structure diagram

2.1 改进 ASPP 模块

空洞空间金字塔池化 ASPP 首次于 DeepLabv2 网络中提出,其主要用于对所给定的输入以不同采样率的空洞卷积进行并行采样,提取不同尺度的上下文信息。相比较而言,DeepLabv3+网络中的 ASPP 模块已经做了将部分空洞卷积模块替成池化模块、调整空洞卷积率等方面改进,并在此基础上构建“编码器-解码器”结构,使网络的性能有了很大的提高。DeepLabv3+网络中的 ASPP 主要包括以下 5 个结构模块:1x1 卷积模块、膨胀率大小分别为 6、

12、18 的空洞卷积模块以及全局平均池化模块。由于过大的空洞卷积率在进行卷积运算时会增加信息丢失的概率,降低边界特征的提取效果,因此本文首先在 ASPP 模块中增加一个空洞卷积率为 3 的空洞卷积分支,用于弥补信息的丢失。然后将膨胀率为 3、6、12 分支的输出与传入 ASPP 模块的深层特征融合,分别作为膨胀率为 6、12、18 分支的输入,实现不同感受野下的多通道信息融合,以此提高 ASPP 模块的特征提取能力。改进后的 ASPP 模块结构如图 2 所示。

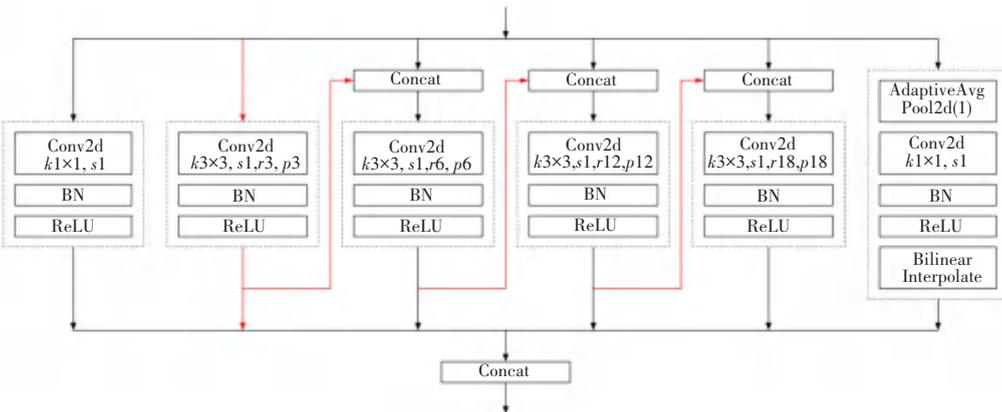


图 2 改进 ASPP 模块结构图

Fig. 2 Improved ASPP module structure diagram

通道特征信息融合实现过程如下:

大小可由如下公式得出:

(1) 输入特征 X 经过空洞卷积的等效卷积核 K'

$$K' = k + (k - 1)(r - 1) \quad (1)$$

其中, k 表示原始卷积核大小, r 表示空洞卷积率参数。

经过空洞卷积的输出特征大小 W_{out} 的公式为:

$$W_{out} = (W_{in} - K' + 2P) / S + 1 \quad (2)$$

其中, W_{in} 表示输入特征 X 的大小; K' 表示等效卷积核大小, 可由式(1)计算得出; S 表示步长; P 表示 padding 的像素数。

(2) 对于改进 ASPP 模块中空洞卷积率为 3 的空洞卷积分支而言, 其原始卷积核大小 $k = 3$, 空洞卷积率参数 $r = 3$, 步长 $S = 1$, padding 的像素数 $P = 3$, 经过式(1)和式(2)计算可得:

$$K' = 7, W_{3out} = W_{3in}$$

因而输入特征与输出特征的大小相等, 该分支的输出特征能与传入 ASPP 模块的输入特征融合, 作为空洞卷积率为 6 的空洞卷积分支的输入, 实现了不同感受野下的通道信息融合。以此类推, 空洞卷积率为 6 和 12 分支的输出能与传入 ASPP 模块的输入特征融合, 分别作为空洞卷积率为 12 和 18 分支的输入。

2.2 PSA 注意力机制基本原理

注意力机制能够使卷积神经网络在训练时注意更为重要的信息, 是实现网络自适应注意的方式之一。在神经网络中添加注意力机制能够提高模型的性能和泛化能力, 因而注意力机制已经被引入到许多计算机视觉任务中。

本文通过引入 PSA 注意力机制来提高网络模型的分割性能。相比于其他注意力机制, PSA 注意力机制在通道和空间上并未进行大幅压缩, 保持着较高的维度, 因而能够使降维过程所带来的信息损失比较小。同时, PSA 注意力机制采用细粒度回归输出分布的非线性函数, 因而能够在细粒度的像素级任务上保持比较好的性能。结合 PSA 注意力机制的上述优点, 本文在 ASPP 模块后引入 PSA 注意力机制, 用于弥补信息的潜在损失, 提高网络模型的性能。

PSA 的输出方式有并行输出和顺序输出两种, 结构原理分别如图 3(a) 和图 3(b) 所示。PSA 算法主要分为 2 条分支, 其中一条分支做通道维度的自注意力机制 (Channel-only Self-Attention), 另一条分支做空间维度的自注意力机制 (Spatial-only Self-Attention)。输入特征图 $X = [x_1, x_2, \dots, x_c]$, $X \in R^{C \times H \times W}$ 。经过 2 条分支得到结果后进行融合, 得到 PSA 的输出, 具体实现过程如下。

(1) 在通道分支上。整个通道模块定义为 $A^{ch}(X)$, $A^{ch}(X) \in R^{C \times 1 \times 1}$, 则:

$$A^{ch}(X) = F_{SG} [W_{z1} \theta_1 ((\sigma_1(W_v(X)) \times F_{SM}(\sigma_2(W_q(X)))))] \quad (3)$$

其中, W_q, W_v, W_z 分别表示 1×1 卷积层; σ_1 和 σ_2 表示 2 个张量整形算子; “ \times ” 表示矩阵点积运算; F_{SM} 表示 Softmax 算子, 对应的公式为:

$$F_{SM}(X) = \sum_{j=1}^{N_p} \frac{e^{x_j}}{\sum_{m=1}^{N_p} e^{x_m}} x^j \quad (4)$$

输入特征图 X 经过通道模块 $A^{ch}(X)$ 后得到的通道注意力输出 Z^{ch} 表示为:

$$A^{ch}(X) \Theta^{ch} X, X \in R^{C \times H \times W} \quad (5)$$

其中, Θ^{ch} 表示一个通道乘法运算符。

(2) 在空间分支上。整个空间模块定义为 $A^{sp}(X)$, $A^{sp}(X) \in R^{1 \times H \times W}$, 则:

$$A^{sp}(X) = F_{SG} [\sigma_3(F_{SM}(\sigma_1(F_{CP}(W_q(X)))) \times \sigma_2(W_v(X)))] \quad (6)$$

其中, W_q, W_v 分别表示 1×1 卷积层; $\sigma_1, \sigma_2, \sigma_3$ 表示 3 个张量整形算子; F_{SM} 表示 Softmax 算子; “ \times ” 表示矩阵点积运算; F_{CP} 表示全局池化算子, 具体公式为:

$$F_{CP}(X) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X(:, i, j) \quad (7)$$

输入特征图 X 经过通道模块 $A^{sp}(X)$ 后得到的空间注意力输出 Z^{sp} 表示为:

$$Z^{sp} = A^{sp}(X) \Theta^{sp} X, X \in R^{C \times H \times W} \quad (8)$$

其中, Θ^{sp} 表示一个空间乘法运算符。

(3) 通道分支与空间分支融合得到 PSA 算法的输出, 融合的方式有 2 种。

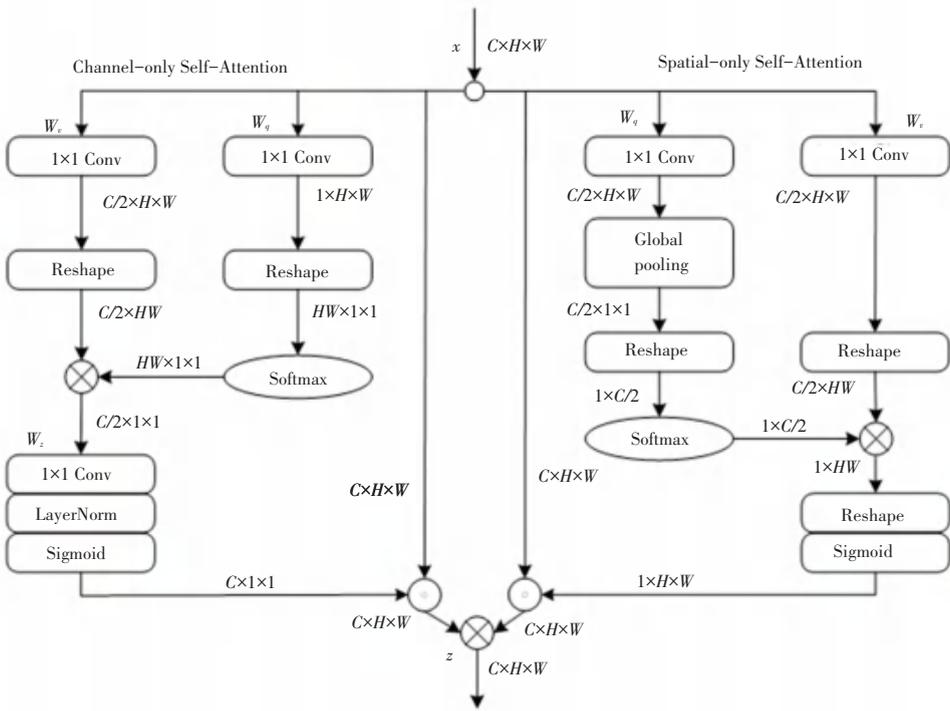
① 以并行布局的方式输出, 将通道分支上的输出与空间分支上的输出相加求得, 计算公式如下:

$$PSA_p(X) = Z^{ch} + Z^{sp} = A^{ch}(X) \Theta^{ch} X + A^{sp}(X) \Theta^{sp} X \quad (9)$$

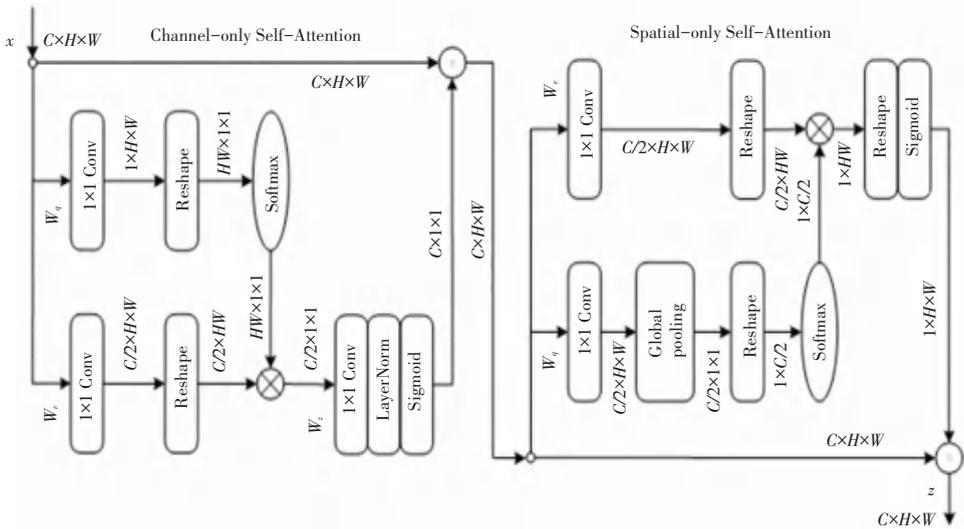
② 以顺序布局的方式输出, 将通道分支上的输出作为空间分支上的输入, 经过计算后求得, 计算公式如下:

$$PSA_s(X) = Z^{sp}(Z^{ch}) = A^{sp}(A^{ch}(X) \Theta^{ch} X) \Theta^{sp} A^{ch}(X) \Theta^{ch} X \quad (10)$$

经过主干网络提取的深层特征经过改进 ASPP 模块后得到 6 个大小、通道数都相同的输出特征层, 此后嵌入 PSA 注意力机制, 将通道和空间分支上的注意力应用于拼接后的特征层上, 保持特征层中较高维度的信息, 保留了特征层中感兴趣区域的位置信息, 减少信息损失, 最后再通过 1×1 卷积模块得到感受野更大、全局信息和语义信息更加丰富的高层特征层。



(a) 并行布局输出的 PSA 算法原理图



(b) 顺序布局输出的 PSA 算法原理图

图 3 2 种布局输出的 PSA 算法原理图

Fig. 3 Schematic diagram of PSA algorithm for two layout outputs

3 实验

3.1 实验介绍

本文的实验在 Windows10 系统上完成, 系统 CPU 型号为 Intel(R) Core(TM) i5-10400F CPU @ 2.90 GHz, 系统 GPU 型号为 NVIDIA GeForce RTX 3080, 编程采用 Python 语言, 模型基于 PyTorch 深度学习框架搭建。

本文首先在 PASCAL VOC 2012 数据集上设计

消融实验和对比实验, 验证改进网络模型的性能, 然后在自制的数据集上设计对比实验, 评估本文改进 DeepLabv3+ 网络模型在实际场景中的分割性能。PASCAL VOC 2012 数据集包含背景、人物、飞机等共 21 个类别, 其中训练集有 1464 张图片, 验证集有 1449 张图片, 测试集有 1456 张图片。本文的自制数据集包含背景、老虎钳、剪刀、螺丝刀、固体胶共 5 个类别, 以 2 个、3 个、4 个物体相互组合、随机遮挡的方式构建共 600 张图像, 其中训练集有 500 张图

片,验证集有50张图片,测试集有50张图片。部分自制数据集图像如图4所示。



图4 自制数据集图像

Fig. 4 Images of self-made dataset

3.2 评估指标

本文实验的评估指标采用语义分割标准度量平均交并比(Mean Intersection over Union, $MIoU$),通过对比网络改进前后 $MIoU$ 值的大小衡量网络模型的分割性能。

$MIoU$ 用于计算所有类别的预测值与真实值集合的交集与并集之比的平均值,计算公式如下:

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

其中, k 表示所有前景的类别数; $k+1$ 表示所有前景加上背景的类别数; P_{ii} 表示将第 i 类像素预测为第 i 类像素的数目; P_{ij} 表示将第 i 类像素预测为第 j 类像素的数目; P_{ji} 表示将第 j 类像素预测为第 i 类像素的数目。

3.3 PASCAL VOC 2012 数据集消融实验验证

为了验证本文改进方法对 DeepLabv3+网络模型性能的影响,首先在 PASCAL VOC 2012 数据集上设计消融实验验证,通过控制变量的方法分别验证改进 ASPP 模块、添加 PSA 注意力机制对网络模型性能的影响。

3.3.1 改进 ASPP 模块

本实验验证改进 ASPP 模块对网络模型性能的影响,实验结果见表1。

表1 改进 ASPP 模块对网络模型性能的影响

Table 1 Impact of improved ASPP modules on network model performance

实验	训练方式	$MIoU$ / %
1	Xception + ASPP	77.59
2	Xception +改进 ASPP	79.03

从表1可以看出,实验1采用原 DeepLabv3+网络进行训练,得到的 $MIoU$ 值为 77.59%;实验2采用 Xception 作为主干网络与改进 ASPP 模块进行训练,得到的 $MIoU$ 值为 79.03%。对比实验1和实验

2可知,改进 ASPP 模块使 $MIoU$ 值提高了 1.44%,因而改进 ASPP 模块能够提高网络模型的性能。

3.3.2 添加 PSA 注意力机制

本实验验证添加 PSA 注意力机制对网络模型性能的影响,实验结果见表2。

表2 添加 PSA 注意力机制对网络模型性能的影响

Table 2 Effect of adding PSA attention mechanism on network model performance

实验	训练方式	$MIoU$ / %
3	Xception + ASPP + PSA	78.65
4	Xception +改进 ASPP + PSA	79.21

从表2可以看出,实验3采用 Xception 作为主干网络、ASPP 模块与添加 PSA 注意力机制进行训练,得到的 $MIoU$ 值为 78.56%;实验4采用 Xception 作为主干网络、改进 ASPP 模块与添加 PSA 注意力机制进行训练,得到的 $MIoU$ 值为 79.21%。首先,对比实验1和实验3可知,添加 PSA 注意力机制使 $MIoU$ 值提高了 1.06%,因而添加 PSA 注意力机制能够提高网络模型的性能。然后,对比实验2和实验4可知,在同样都采用改进 ASPP 模块进行训练的情况下,添加 PSA 注意力机制使 $MIoU$ 值提高了 0.18%,进一步验证了添加 PSA 注意力机制的可行性。此外,对比实验2和实验3可知,采用改进 ASPP 模块训练得到的 $MIoU$ 值比添加 PSA 注意力机制训练得到的 $MIoU$ 值高 0.38%,进一步验证了改进 ASPP 模块的优越性。最后,对比实验1和实验4可知,本文的改进 DeepLabv3+网络模型相比较于原网络模型, $MIoU$ 总值提高了 1.62%,验证了本文的改进方法能够整体提高 DeepLabv3+网络模型的性能。

3.4 PASCAL VOC 2012 数据集对比实验验证

为进一步验证改进 DeepLabv3+网络模型性能的优越性,本文在 PASCAL VOC 2012 数据集上与其他语义分割算法设计对比实验验证,实验结果见表3。

表3 不同语义分割算法在 PASCAL VOC 2012 数据集上的性能对比

Table 3 Performance comparison of different semantic segmentation algorithms on PASCAL VOC 2012 dataset

实验	网络模型	$MIoU$ / %
5	FCN	68.95
6	LR-ASPP	71.23
7	RefineNet	73.32
8	PSPNet	74.78
9	DeepLabv3	76.12
10	DeepLabv3+	77.59
11	本文改进网络	79.21

对比表 3 内所有实验的 $MIoU$ 值可知,改进网络模型的性能要明显优于其他语义分割算法的性能,进一步验证了本文的改进 DeepLabv3+网络模型性能的优越性。

结合以上实验内容,分别采用原 DeepLabv3+网络和改进 DeepLabv3+网络对 PASCAL VOC 2012 测

试集进行预测,部分预测图像如图 5 所示。

由图 5 分析可知,改进 DeepLabv3+网络模型的分割效果要优于原网络模型,并且对于多个目标的图像具有更好的分割性能,同时也能更高效地提取出图像中的小尺寸细节信息,从而能较好地改善分割断裂问题。

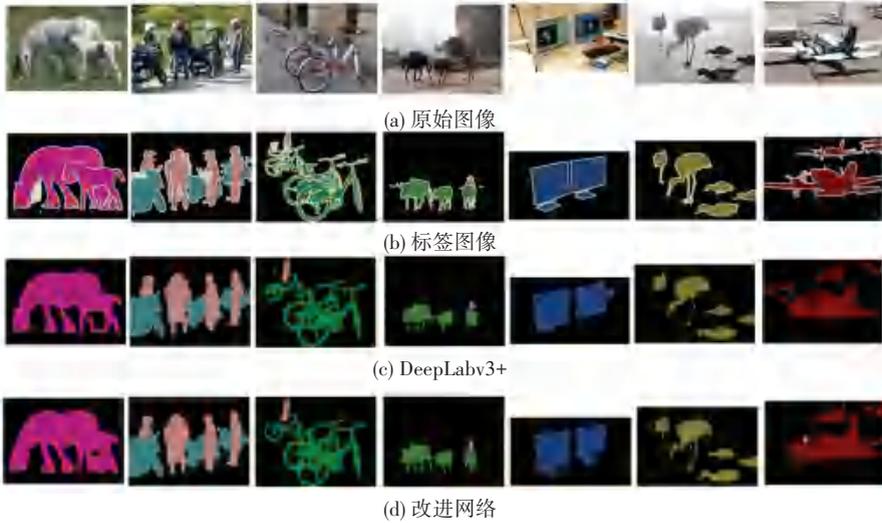


图 5 PASCAL VOC 2012 测试集的预测图像

Fig. 5 Prediction images of the PASCAL VOC 2012 test dataset

3.5 自制数据集对比实验验证

为评估改进 DeepLabv3+网络模型在实际场景中的分割性能,本文在自制数据集上设计对比实验,实验结果见表 4。

表 4 网络模型改进前后在自制数据集上的性能对比

Table 4 Performance comparison of network model on self-made dataset before and after improvement

实验	训练方式	$MIoU$ / %
12	DeepLabv3+	88.56
13	改进 DeepLabv3+	89.19

从表 4 可以看出,实验 12 采用原 DeepLabv3+网络模型进行训练,得到的 $MIoU$ 值为 88.56%;实验 13 采用改进 DeepLabv3+网络模型进行训练,得到的 $MIoU$ 值为 89.19%。对比实验 12 和实验 13 可知,本文的改进网络相比较于原网络, $MIoU$ 总值提高了 0.63%,验证了本文的改进 DeepLabv3+网络模型在实际场景中也依然具有较好的分割性能。

结合以上对比实验,分别采用原 DeepLabv3+网络和改进 DeepLabv3+网络对自制的测试集进行预测。为避免实验的偶然性,分别选用测试集中 2 个、3 个、4 个物体相互组合、随机遮挡的图片进行实验,部分预测图像如图 6 所示。

将图 6(c)和图 6(d)中对 4 个目标的预测图像

放大,得到图 7。从图 7 可以看出,当螺丝刀尖部与老虎钳重叠时,原网络模型对螺丝刀尖部分割失败,而改进网络模型可以提取到螺丝刀尖部的特征,并且能够得到一部分正确的分割结果。由此能够进一步验证本文的改进 DeepLabv3+网络模型对图像中的小尺寸细节信息具有更好的提取能力,对多个目标图像的分割性能也更为优异,这就证实了本文的改进 DeepLabv3+网络模型在实际场景中的可行性。

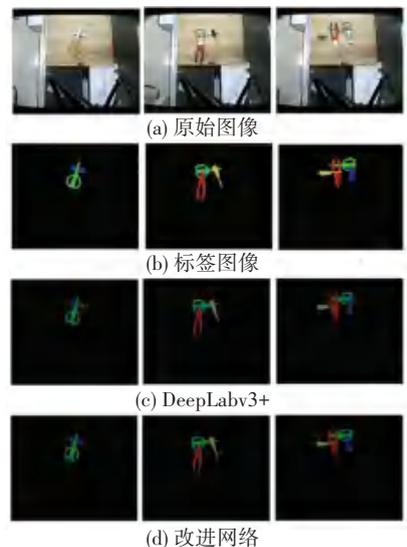


图 6 自制测试集的预测图像

Fig. 6 Prediction images of self-made test dataset

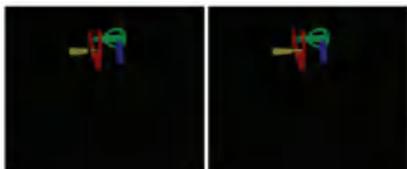


图7 4个目标的预测图像

Fig. 7 Prediction images of the four targets

4 结束语

针对 DeepLabv3+网络对多个目标图像的分割性能欠佳、容易丢失图像的细节信息,产生分割断裂等问题,本文提出一种基于改进 DeepLabv3+网络的语义分割方法。首先,对原 DeepLabv3+网络模型中的 ASPP 模块进行改进,在 ASPP 模块中增加一个空洞卷积率为 3 的空洞卷积分支,并依次将空洞卷积率为 3、6、12 分支的输出与传入 ASPP 的深层特征融合,分别作为空洞卷积率为 6、12、18 分支的输入,实现不同感受野下的多通道信息融合,提高网络的特征提取能力。其次,在 ASPP 模块后引入 PSA 注意力机制,将通道和空间分支上的注意力应用于拼接后的特征层上,保留了特征层中感兴趣区域的位置信息,减少信息损失。最后,在 PASCAL VOC 2012 数据集上设计消融实验,对比实验与测试实验。实验结果表明,改进后的 DeepLabv3+网络模型 *MIoU* 值总体提高了 1.62%,并且优于其他语义分割方法。同时,改进网络模型对于多个目标的图像具有更好的分割性能,对于图像中的小尺寸细节信息也有更好的提取能力,能够较好地改善分割断裂问题。接下来,本文在自制的数据集上设计对比实验与测试实验,实验结果证实了本文的改进 DeepLabv3+网络模型在实际场景中也依然具有良好的分割性能。在未来的研究中,将结合实际工程项目,针对更为复杂的实际工业环境做出进一步的研究与改进。

参考文献

- [1] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [J]. arXiv preprint arXiv, 1411.4038, 2015.
- [2] CHEN L C, PAPANDEOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs [J]. arXiv preprint arXiv, 1412.7062, 2014.
- [3] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C]// 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015: 234-241.
- [4] 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.
- [5] CHEN L C, PAPANDEOU 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.
- [6] LIN Guosheng, MILAN A, SHEN Chunhua, et al. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 1925-1934.
- [7] CHEN L C, PAPANDEOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation [J]. arXiv preprint arXiv, 1706.05587, 2017.
- [8] CHEN L C, ZHU Y, PAPANDEOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2018: 801-818.
- [9] YU Y, WANG C, FU Q, et al. Techniques and challenges of image segmentation: A review [J]. Electronics, 2023, 12(5): 1199.
- [10] EMEK S B, GUZEL M S, BOSTANCI G E, et al. Deep-learning-based approaches for semantic segmentation of natural scene images: A review [J]. Electronics, 2023, 12(12): 2730.
- [11] 徐辉, 祝玉华, 甄彤, 等. 深度神经网络图像语义分割方法综述 [J]. 计算机科学与探索, 2021, 15(1): 47-59.
- [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(6): 1145-1151.
- [13] 严广宇, 刘正熙, 熊运余, 等. 基于目标检测和语义分割共享单车类别与违停检测 [J]. 计算机应用研究, 2020, 37(10): 3175-3179.
- [14] 柳林, 马泽鹏, 孙毅, 等. 基于 MS-DeepLabV3+ 的街景语义分割及城市多维特征识别 [J]. 武汉大学学报(信息科学版), 2024, 49(3): 343-354.
- [15] 范瑶瑶, 王兴芬, 刘亚辉. 改进 Deeplabv3+网络的钢板表面缺陷检测研究 [J]. 计算机工程与应用, 2023, 59(16): 150-158.
- [16] 郭金, 宋廷强, 孙媛媛, 等. 改进 Deeplabv3+ 的双注意力融合作物分类方法 [J]. 计算机工程与应用, 2024, 60(8): 110-120.
- [17] WANG Ping, LI Chuanxue, YANG Qiang, et al. Environment understanding algorithm for substation inspection robot based on improved DeepLab V3+ [J]. Journal of Imaging, 2022, 8(10): 257.
- [18] YUAN Hongbo, ZHU Jiajun, WANG Qifan, et al. An improved DeepLab v3+ deep learning network applied to the segmentation of grape leaf black rot spots [J]. Frontiers in Plant Science, 2022, 13: 795410.
- [19] WANG Jing, LIU Xiuping. Medical image recognition and segmentation of pathological slices of gastric cancer based on Deeplab v3 + neural network [J]. Computer Methods and Programs in Biomedicine, 2021, 207: 106210.
- [20] LIU Chengqi, SU Jie, WANG Longhe, et al. LA-DeepLab V3+: A novel counting network for pigs [J]. Agriculture, 2022, 12(2): 284.
- [21] LIU Huajun, LIU Fuqiang, FAN Xinyi, et al. Polarized self-attention: Towards high-quality pixel-wise regression [J]. arXiv preprint arXiv, 2107.00782, 2021.
- [22] EVERINGHAM M, ESLAMI S M A, GOOL V L, et al. The pascal visual object classes challenge: A retrospective [J]. International Journal of Computer Vision, 2015, 111: 98-136.