

基于 DeepLabV3+ 与注意力机制相结合的图像语义分割

邱云飞, 温金燕*

辽宁工程技术大学软件学院, 辽宁 葫芦岛 125105

摘要 基于 DeepLabV3+ 进行图像分割时, 在特征提取阶段忽略了不同级别的特征图中存在的特征重要程度不同, 丢失了大量的细节信息, 致使分割效果不佳。针对该问题, 提出了一种基于 DeepLabV3+ 与注意力机制相结合的图像语义分割算法。在骨干网络 Xception 模型中提取两条低级特征作为解码器的输入特征, 提高特征提取的准确性; 采用通道注意力模块有效融合高级特征, 获取丰富的上下文信息; 采用空间注意力模块提取低级特征, 过滤背景信息, 减少细节信息的丢失; 采用深度可分离卷积代替空洞卷积有效降低参数量, 提高计算速度; 同时, 采用焦点损失作为损失函数通过降低内部加权, 提高最终的分割效果。实验结果表明, 所提算法在 PASCAL VOC 2012 数据集上的平均交并比 (mIoU) 值达到了 84.44%, 与传统算法和基于 DeepLabV3+ 改进的算法相比, 有效提高了特征提取的准确性, 减少了特征细节信息的损失, 对最终的分割效果有了较好的提升。

关键词 图像处理; 图像分割; DeepLabV3+; Xception 模型; 注意力机制; 空间注意力; 通道注意力

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP202259.0410008

Image Semantic Segmentation Based on Combination of DeepLabV3+ and Attention Mechanism

Qiu Yunfei, Wen Jinyan*

School of Software, Liaoning Technical University, Huludao, Liaoning 125105, China

Abstract In image segmentation based on DeepLabV3+, the different importance of features in different levels of feature images are ignored in the feature extraction stage, and a large amount of details are lost, resulting in poor segmentation effect. To solve this problem, an image semantic segmentation algorithm based on the combination of DeepLabV3+ and attention mechanism is proposed. Two low-level features are extracted in the backbone network Xception model as input features of the decoder to improve the accuracy of feature extraction. The channel attention module is used to effectively integrate high-level features and obtain rich context information. The spatial attention module is used to extract low-level features and filter background information to reduce the loss of details. The depthwise separable convolution is substituted for void convolution to effectively reduce the amount of parameters and improve the calculation speed. At the same time, the focus loss is used as the loss function to improve the final segmentation effect by reducing the internal weighting. Experimental results show that the mean intersection over union (mIoU) value of the proposed algorithm on PASCAL VOC 2012 dataset reaches 84.44%. Compared with the traditional algorithm and the improved algorithm based on DeepLabV3+, the proposed algorithm effectively improves the accuracy of feature extraction, reduces the loss of feature details, and improves the final segmentation effect.

Key words image processing; image segmentation; DeepLabV3+; Xception model; attention mechanism; spatial attention; channel attention

收稿日期: 2021-03-01; 修回日期: 2021-03-18; 录用日期: 2021-03-23

通信作者: *1004985072@qq.com

1 引言

图像中的不同元素是由各种像素组合在一起形成的,因此将这些像素按元素分类的方法称为图像语义分割。基于深度学习的图像语义分割是近些年的热点研究,随着大量原本在图像分类、目标检测、自然语言处理等领域取得成功的深度学习方法被改进、迁移到语义分割领域,图像的语义分割技术取得了较大的突破,逐渐改变了各个行业的发展趋势^[1-3]。

卷积神经网络(CNN)的应用使图像语义分割得到了快速的发展,各种基于卷积神经网络的语义分割网络结构相继被提出。目前,关于图像分割领域的研究大致有3种。1)通过改进全卷积网络的结构^[4-7],结合更深层次的神经网络来提高分割性能。谭光鸿等^[8]提出了一种基于卷积神经网络的低参数量图像分割算法,该算法通过改进深层次的神经网络,运用多尺度空洞卷积,增加规模标准化层,优化网络,在降低参数量的同时提高分割效果。2)基于编解码架构^[9-10]采用多种方法提取特征信息,提高特征图的分辨率从而提升分割效果。Tian等^[11]提出了一种基于DeepLabV3+架构改进的DUpsampling算法,该算法在解码模块采用了一种数据相关型的上采样方法DUpsampling来代替双线性插值,在恢复特征图大小的同时,提升了分割精度,降低了计算复杂度。3)采用注意力机制^[12-15]对目标特征信息进行建模,以突出特征图细节信息从而提升分割效果。Li等^[16]提出了一种用于图像分割的金字塔注意力网络(PAN),该网络通过结合特征金字塔注意力网络(FPA)和全局注意力上采样网络(GAU)替代空洞空间卷积池化金字塔(ASPP)结构,进行特征提取。

目前,DeepLab算法^[17-20]是较为主流的分割算法,该算法引入多尺度特征,通过设置不同参数的卷积层和池化层,提取不同尺度的特征图进行融合,提升了整个网络的性能。同时特征通道中存在着重要的空间信息和通道信息,采用注意力机制对目标特征信息进行建模,通过卷积网络池化得到特征通道,以突出特征图细节信息,提升分割效果,也是目前针对图像分割算法研究的热点之一。因此,本文提出了一种基于DeepLabV3+与注意力机制相结合的图像语义分割算法。所提算法在DeepLabV3+的基础上引入注意力机制,根据每个通道中存在特征的重要程度,针对不同的特征引入

不同的注意力机制。对从骨干网络中提取出的高级特征引入通道注意力,对提取出的低级特征引入空间注意力。在经过不同的注意力机制处理后,高级特征和低级特征具有互补意识,可以更好地融合在一起,保留更多有效的特征信息,提高特征提取准确度。同时采用焦点损失(focal loss)作为损失函数,解决不同类别的样本数量不均衡和不同对象的区分难度不同的问题,提高目标识别的准确度,提升分割效率。

2 相关理论基础

2.1 DeepLabV3+原理

DeepLabV3+是典型的语义分割网络架构之一,在DeepLabV3的基础上增加了一个简单而有效的解码器模块来细化分割结果。DeepLabV3+的编码器部分通过骨干网络Xception模型中不同通道的深度可分离卷积层提取图像特征信息,利用空间金字塔池化模块中不同速率的并行空洞卷积获取高层语义信息,并通过 1×1 卷积进行通道压缩;解码器部分将骨干网络中提取出的低级特征与经过4倍双线性插值上采样的高级特征进行融合,再利用 3×3 卷积恢复空间信息和4倍双线性插值上采样精细目标边界,得到分割结果。

在DeepLabV3+算法实现过程中,所提算法采用深度可分离卷积替换原有的最大池化层,将深度可分离卷积应用于实际的空间金字塔池和解码器模块中,将depthwise Conv和pointwise Conv相结合代替空洞卷积,有效降低了计算复杂度,从而实现了更快、更强的编解码器网络。

2.2 Xception模型

Xception模型是Chollet^[21]提出的一种基于深度可分离卷积的网络,是在Inception v3的基础上,把Inception模块替换为深度可分离卷积,然后结合ResNet的跳跃连接实现的。Xception模型在Russakovsky等^[22]提出的ImageNet上表现出了很好的图像分类结果,而且计算速度很快。Chen等^[20]提出了一种具有可分离卷积的编解码器用于语义图像分割网络,该编解码器引入Xception模型来完成语义图像分割的任务,并通过结合Abadi等^[23]提出的TensorFlow深度可分离卷积,对Xception模型进行了改进,利用深度可分离卷积替代空洞卷积提取更加密集的特征图,Xception模型结构如图1所示。

在原模型的基础上增加更多的层,不修改入口

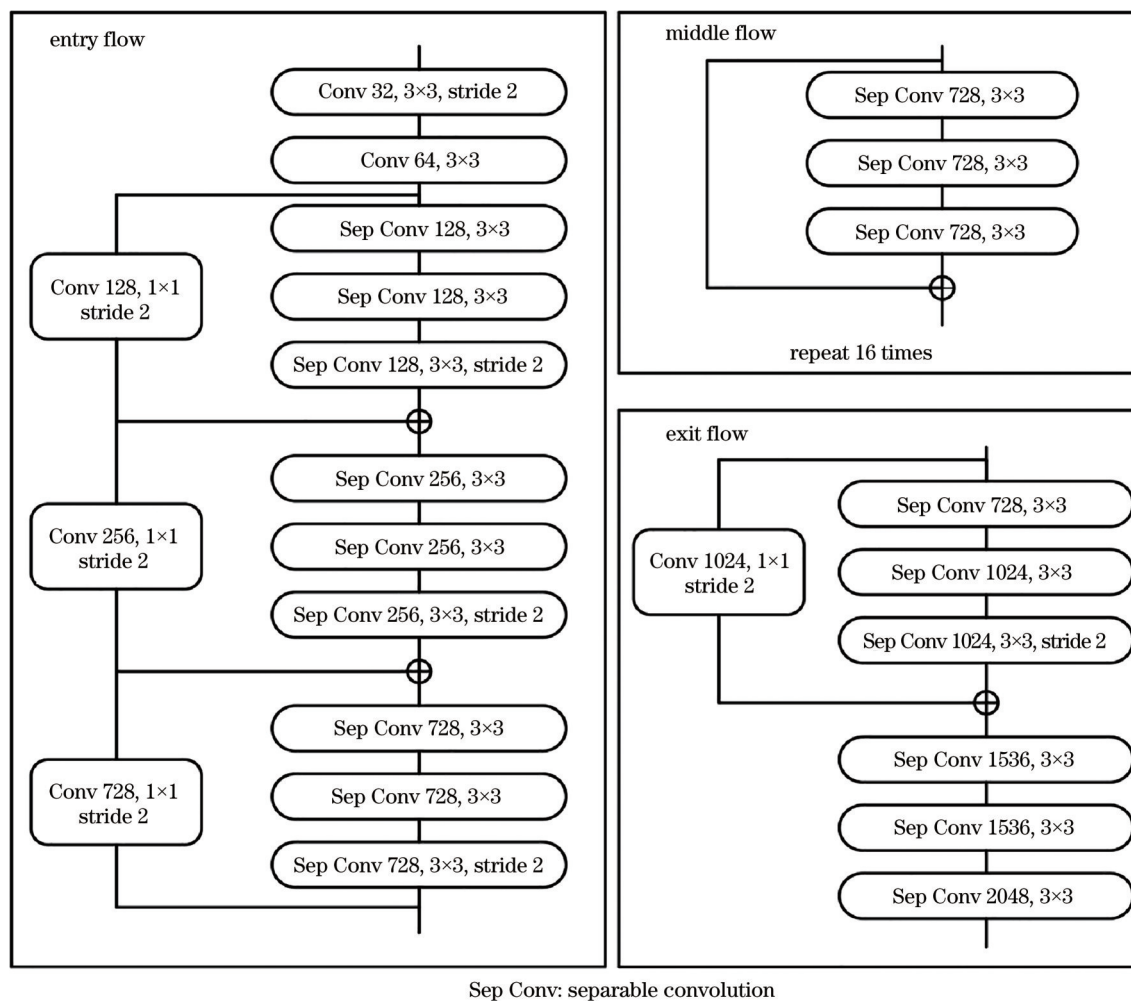


图1 Xception模块结构图

Fig. 1 Structure diagram of Xception module

流网络结构实现了快速计算且存储效率较高;所有最大池化操作都被跨步的深度可分离卷积代替,目的是应用空洞可分离卷积通过任意的分辨率来提取特征图;在每次进行 3×3 深度卷积之后,添加额外的批处理归一化和ReLU激活。

2.3 交叉熵损失函数

利用深度学习方法进行图像分割时,交叉熵可以在神经网络中作为损失函数,用来衡量预测值与真实值之间的相似度。损失函数的选择对于任何架构学习正确的目标都是至关重要的。交叉熵主要用于度量两个概率分布间的差异性信息,可用在语言模型中表示文本识别的难度,也可用在神经网络中作为损失函数。以交叉熵作为损失函数可以衡量真实标记与训练后模型的预测标记之间的相似性,还有一个优势是Sigmoid函数在梯度下降时能避免均方误差损失函数学习速率降低的问题。DeepLabV3+网络原始模型中使用的交叉熵(CE)

损失函数的表达式为

$$L_{ce}(p_i) = -\log p_i, \quad (1)$$

式中: p_i 为预测结果对应标签的概率。

3 基于改进DeepLabV3+的图像语义分割算法

3.1 总体框架

基于改进DeepLabV3+的图像语义分割算法,主要是通过两个部分相结合的方式实现的:改进DeepLabV3+网络模型和注意力机制模块,该算法的网络结构如图2所示。

所提算法基于编码器-解码器结构,采用Xception模型作为骨干网络,具体实现过程如下。在编码器部分:输入图像经过Xception模型中不同通道的深度可分离卷积层进行特征提取,提取后的特征图分别经过ASPP模块和通道注意力(CA)模块进行处理,利用ASPP模块中的 1×1 卷积、空洞

Sigmoid 操作; fc 表示 FC 层; δ 表示 ReLU 函数。该块的最终输出 \tilde{f}^h 通过使用 CA 加权上下文感知金字塔特征获得。

$$\tilde{f}^h = F_{CA} \cdot f^h. \quad (3)$$

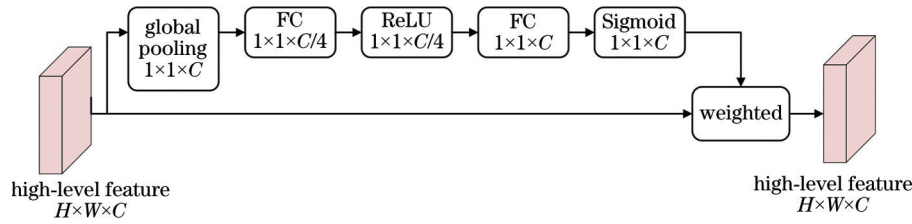


图 3 通道注意力结构图

Fig. 3 Structure diagram of channel attention

3.2.2 空间注意力

在卷积神经网络中,图像特征在网络的不同阶段有着不同的表现能力。虽然从浅层网络可以得到目标位置、边缘等细节特征信息,但是所得信息中存在大量的背景特征信息,很大程度上影响图像分割效果。将网络中获取到的低级特征,直接作为解码阶段的输入信息,引入大量背景特征的同时影响分割结果。为了解决这一问题,采用空间注意力自适应地将低级特征与高级特征相结合,利用高级特征过滤背景信息,空间注意力更多地关注前景区域,聚焦于目标区域特征,有助于生成更加有效的特征图,提高分割效果。

将低级特征表示为 $f^l \in \mathbf{R}^{W \times H \times C}$, 空间位置集合 $R = \{(x, y) | x = 1, \dots, W; y = 1, \dots, H\}$, 其中 (x, y) 为低级特征的空间坐标。为了增加感受野,获得全

局信息的同时不增加参数,类似于 Peng 等^[25]提出的方法,通过全局卷积网络改进语义分割,应用两层卷积,一层核为 $1 \times k$, 另一层核为 $k \times 1$, 用于高级特征获取空间关注点,具体结构如图 4 所示;接着,采用 Sigmoid 运算,对映射到 $[0, 1]$ 的已编码的空间特征图进行归一化处理。

$$C_1 = \text{Conv}_2 \left[\text{Conv}_1(\tilde{f}^h, W_{s1}^1), W_{s1}^2 \right], \quad (4)$$

$$C_2 = \text{Conv}_1 \left[\text{Conv}_2(\tilde{f}^h, W_{s1}^1), W_{s2}^2 \right], \quad (5)$$

$$F_{SA}(\tilde{f}^h, W_s) = \sigma_2(C_1 + C_2), \quad (6)$$

式中: W_s 为空间注意力模块中的参数; σ_2 为 Sigmoid 运算; Conv_1 和 Conv_2 分别为 $1 \times k \times C$ 和 $k \times 1 \times C$ 的对流层。实验中,设置 $k=7$, 块的最终输出 \tilde{f}^l 是通过 SA 加权 f^l 得到的。

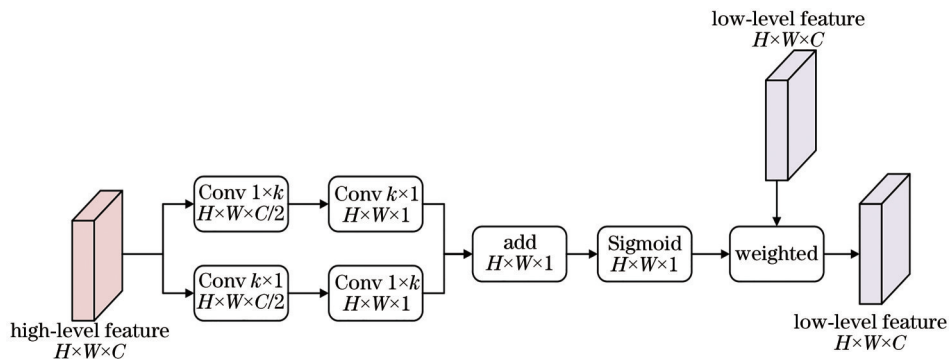


图 4 空间注意力结构图

Fig. 4 Structure diagram of spatial attention

3.3 损失函数

若某类目标的样本相比其他类在数量上占据极大优势,则将该数据集视为不平衡的数据集。这种不平衡将导致两个问题:训练效率低,由于大多数样本都是简单的目标,这些样本在训练中提供给模型不太有用的信息;简单的样本数量上的优势会

影响模型的训练,使模型性能退化。Lin 等^[26]提出的焦点损失(focal loss)函数,通过降低内部加权来解决类别不平衡的问题,该函数侧重于使用困难样本稀疏的数据集来训练,即使简单样本数量很大,但它们对总损失的贡献却很小。

目标物体的种类多,且同一类型的物体大小和

形状也各不相同,其中更包含极少个别突出的物体,CE损失函数不能很好地平衡较少样本的学习,因此引入 focal loss 作为损失函数来解决分割任务中存在的样本不平衡问题。focal loss 是在交叉熵函数的基础上进行改进的,通过修改交叉熵函数和添加样本难度权重调节因子 $(1-p_i)^\gamma$,来减缓样本类别不平衡和样本分类难度不平衡的问题,提升模型精确度,数学表达式为

$$L_{FL}(p_i) = -(1-p_i)^\gamma \log p_i. \quad (7)$$

实际上,也可以增加一个类别权重 α , (7)式改写为

$$L_{FL}(p_i) = -\alpha_i(1-p_i)^\gamma \log p_i, \quad (8)$$

式中: α 为类别间(0-1二分类)的权重参数; $(1-p_i)^\gamma$ 为简单/困难样本调节因子, γ 为聚焦参数。当对某类别预测准确时,即 p_i 接近于1时, $(1-p_i)^\gamma$ 的值接近于0;当对某类别预测不准确时,即 p_i 接近0时, $(1-p_i)^\gamma$ 的值接近于1。设置 $\gamma=2, \alpha=0.25$ 。

4 实验结果与分析

4.1 实验数据及实验参数配置

在 PASCAL VOC 2012 基准数据集上对所提算法进行了相关实验,该数据集是由国际计算机视觉挑战赛发布的用于图像分类、检测或语义分割的数据集,包含 20 个前景对象类和一个背景类,包括人、动物、交通车辆和室内家居用品。数据集中共有 1464 张训练集图像、1449 张验证集图像和 1456 张测试集图像。

实验是在深度学习框架 TensorFlow 上实现的。实验所用操作系统是 Windows 10,显卡是 NVIDIA Quadro P2000。使用 ImageNet-1k 预先训练的 Xception 通过空洞卷积提取密集的特征图,采用 Nesterov 动量优化器, poly 学习策略。实验中将图像裁剪为 513×513 大小进行训练,在刚开始训练的过程中,先采用较小的学习率,以实现平滑启动。设置初始学习率为 1×10^{-4} ,动量为 0.9,迭代训练 50000 次。

4.2 实验评价指标

使用平均交并比(mIoU)、平均像素精度(MPA)和损失值作为图像语义分割算法性能的评价指标。mIoU 通过计算真实值和预测值两个集合的交集来计算预测结果与原始图像中真值的重合程度。

$$R_{mIoU} = \frac{1}{n+1} \sum_{i=0}^n \frac{p_{ii}}{\sum_{j=0}^n p_{ij} + \sum_{j=0}^n p_{ji} - p_{ii}}, \quad (9)$$

式中: n 表示标签标记的类别; $n+1$ 表示包含空类或背景的总类别; p_{ii} 表示实际为 i 类预测为 i 类的像素数量; p_{ij} 表示实际为 i 类但预测为 j 类的像素数量; p_{ji} 表示实际为 j 类但预测为 i 类的像素数量。mIoU的取值范围为 $[0, 1]$,mIoU的值越大,说明预测的分割图越准确。

平均像素精度(MPA)是正确分类的像素数量与所有像素数量的比值后的平均值,表达式为

$$R_{MPA} = \frac{1}{n+1} \sum_{i=0}^n \frac{p_{ii}}{\sum_{j=0}^n p_{ij}}. \quad (10)$$

4.3 实验结果与分析

实验过程中,引入注意力机制提高特征图的准确度,采用通道注意力模块和空间金字塔池化模块提取高级特征,采用空间注意力模块将低级特征和高级特征有效融合,过滤背景信息,减少细节信息的丢失。在空间注意力模块,对高级特征应用两层卷积, $1 \times k$ 和 $k \times 1$,用于获取空间关注点。不同 k 值得到的分割结果如表1所示。从表1中可以看出,对特征图进行不同的卷积操作,分割结果大不相同,在 $k=7$ 时,mIoU最高。因此,最终选取 $k=7$ 的训练模型进行验证。

表 1 不同 k 值的 mIoU 值
Table 1 mIoU values of different k values

k	mIoU / %
1	82.26
3	83.52
5	83.92
7	84.44
9	83.63
11	83.85

在 PASCAL VOC 2012 数据集上进行的实验主要分为两个方面。1) 与不同的图像语义分割算法(PSPNet^[6]、DANet^[27]、DeepLabV3+^[20]和 SANet 算法^[28])的对比;2) 与基于 DeepLabV3+改进的图像分割算法(DUpsampling 算法^[11]和文献[29-31]中的算法)进行对比。实验结果如表2所示。从表2中可以看出,所提算法 mIoU 值相对于 PSPNet 算法提高了 5.88 个百分点,比 DANet 算法提高了 3.48 个百分点,比 DeepLabV3+ 算法提高了 3.30 个百分点,比 SANet 算法提高了 1.24 个百分点;与基于 DeepLabV3+改进的 DUpsampling 算法相比提高了

表 2 不同图像语义分割算法的对比结果

Table 2 Comparison results of different image semantic segmentation algorithms

Algorithm	mIoU / %
PSPNet	78.56
DANet	80.96
DeepLabV3+	81.14
DUpsampling	82.03
Algorithm of [29]	82.26
Algorithm of [30]	82.94
SANet	83.20
Algorithm of [31]	83.40
Proposed algorithm	84.44

2.41个百分点,与其他基于DeepLabV3+改进的图像语义分割算法相比,比文献[29]中的算法提高了2.18个百分点,比文献[30]中的算法提高了1.50个百分点,比文献[31]中的算法提高了1.04个百分点。

为了更好地验证所提算法性能,还比较了改进前后DeepLabV3+算法的MPA,结果如表3所示。从表3中可以看出,所提算法不仅在mIoU值上有所提升,MPA相比于DeepLabV3+也提升了3.38个百分点。

在DeepLabV3+算法的基础上,进行了消融实验,验证改进结果。不同组合方法的图像语义分割结果如表4所示。从表4中①和②的结果对比可以看出,对Xception模型进行微调,加入一条低级特征的提取路径的操作使mIoU值提高了0.42个百分

表 3 改进前后DeepLabV3+网络的不同指标对比

Table 3 Comparison of different indexes of DeepLabV3+ network before and after improvement

Algorithm	MPA / %	mIoU / %
DeepLabV3+	89.12	81.14
Proposed algorithm	92.50	84.44

点;①和④的结果对比可以看出,在原网络的基础上引入注意力机制,mIoU值提高了2.16个百分点,表明注意力机制有效提高了特征提取的准确性;④和⑤的结果对比可以看出,多路径提取低级特征可以使mIoU值再增加0.62个百分点;⑤和⑥的结果对比可以看出,focal loss的使用让mIoU值有了0.52个百分点的改善。实验过程中,也对通道注意力和空间注意力做了消融实验,从①和③的实验结果可以看出,当只使用通道注意力对高级特征进行处理而不使用空间注意力时,实验结果并没有明显的提高,说明引入注意力机制必须同时对高级特征和低级特征进行处理,才能有效提高特征提取的准确度。

为了验证所提算法的抗噪能力,对PASCAL VOC 2012数据集在训练模型之前增加了椒盐噪声,将DeepLabV3+算法和所提算法进行了对比实验,输入图像加噪后的分割结果如图5所示。实验结果显示,加噪前后采用DeepLabV3+算法进行图像分割得到的精确度差值为1.96个百分点,加噪前后采用所提算法得到的精确度差值为0.64个百分点,所提算法差值明显小于DeepLabV3+算法,表明

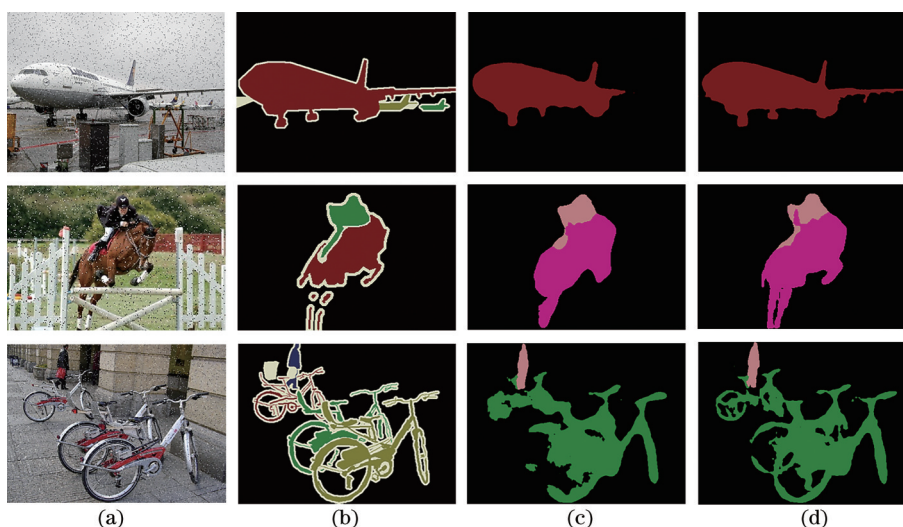


图 5 加噪后分割结果对比图。(a)加噪后图像;(b)图像标签;(c)加噪后DeepLabV3+分割结果;(d)加噪后所提算法分割结果

Fig. 5 Comparison of segmentation results after adding noise. (a) Images after adding noise; (b) image labels; (c) DeepLabV3+ segmentation results after adding noise; (d) segmentation results of proposed algorithm after adding noise

表 4 不同组合方法的图像语义分割结果评估

Table 4 Pedestrian detection accuracy of different combination methods

No.	methods				Focal loss	mIoU /%	Time /s
	DeepLabV3+	Xception	CA	SA			
①	✓				81.14	0.074	
②	✓	✓			81.56	0.075	
③			✓		81.18	0.075	
④	✓		✓	✓	83.30	0.076	
⑤	✓	✓	✓	✓	83.92	0.076	
⑥	✓	✓	✓	✓	84.44	0.077	

所提算法鲁棒性较好,模型更加稳定。对比实验中,针对 Xception 模型提取的不同特征图,也进行了一组对比实验,对图像增加噪声后,分别验证了从 Xception 模型中提取一条特征图和提取两条特征图的分割效果。实验结果显示,后者比前者得到的精确度提高了 0.51 个百分点,表明所提算法可以更好地提取特征信息,减少细节信息的损失,提高分割精度。

DeepLabV3+ 算法和所提算法的训练损失值曲线图如图 6 所示,该图是通过 Tensorboard 可视化展示出来的,横坐标为训练次数,纵坐标为损失值,上边曲线为 DeepLabV3+ 算法的训练损失值变化,下边曲线为所提算法的训练损失值变化。从图 6 及损失日志文件可知,模型训练开始时,损失

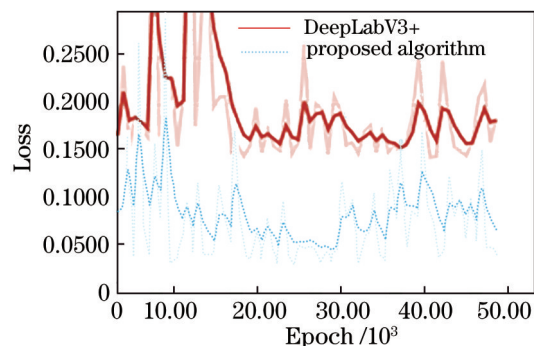


图 6 不同方法的损失值

值忽高忽低,随着训练次数的增多,损失值曲线幅度逐渐平稳。所提算法与原始算法相比,损失值整体降低幅度较大,说明所提算法可有效降低特征信息的损失,提高最终分割的效果。

对 PASCAL VOC 2012 数据集测试集中的测试集采用了不同算法进行分割,实验对比结果如图 7 所示。从图中可以看出,所提算法与 DeepLabV3+ 算法和 DUpsamling 算法相比,分割目标的边界更加清晰; DeepLabV3+ 算法在分割自行车时明显存在分割不均衡问题;相比于 DUpsamling 算法,所提算法的分割结果明显更均衡。实验结果表明,所提算法对于目标背景的边界分割效果,有了显著的提升,细化了目标边界,提高了目标物体的分割效果,具有更好的物体分辨能力。

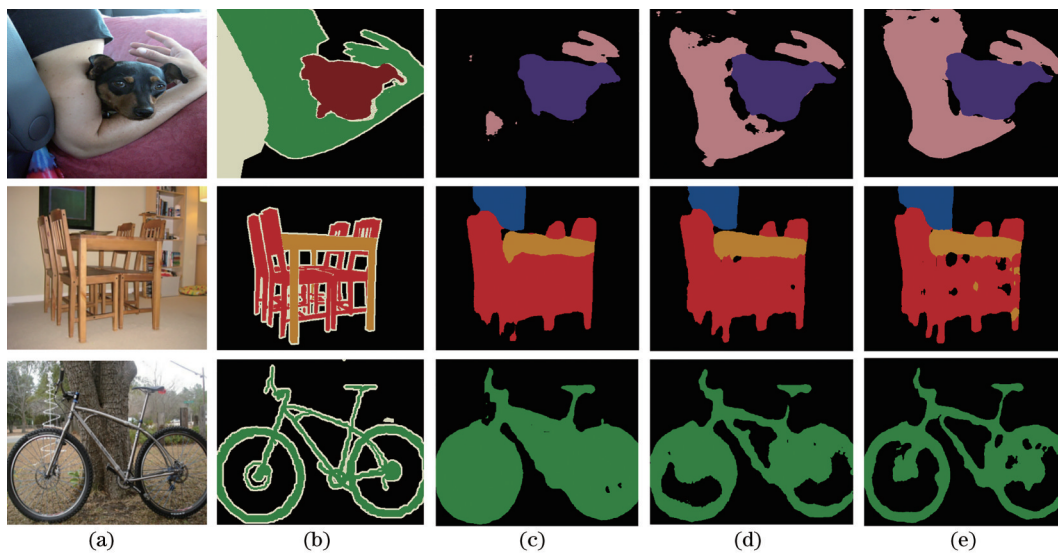


图 7 分割结果对比图。(a)原始图像;(b)图像标签;(c) DeepLabV3+ 分割结果;(d) DUpsamling 分割结果;(e)所提算法分割结果

Fig. 7 Comparison of segmentation results . (a) Original images; (b) image labels; (c) segmentation results of DeepLabV3 + ; (d) segmentation results of DUpsamling; (e) segmentation results of proposed algorithm

5 结 论

提出了一种基于 DeepLabV3+ 与注意力机制相结合的图像语义分割算法。通过在骨干网络 Xception 中新增一条低级特征的提取路径,增加特征信息,扩大感受野,在不影响网络速度的同时,提高算法的准确度;引入注意力机制,通过将通道注意力和空间注意力分别应用于提取高级特征和低级特征,将二者有效地融合,得到丰富的上下文信息,过滤背景信息,获取更加精细的特征图,同时运用深度可分离卷积代替空洞卷积;为了解决不同类别的样本数量不均衡问题和不同对象的区分难度不同的问题,采用焦点损失函数代替交叉熵损失函数,以减少特征细节信息的损失。在 PASCAL VOC 2012 上数据集进行了相关实验,实验结果表明,所提算法相比 DeepLabV3+ 算法对图像分割准确度有了很大的提升;与其他图像分割算法及基于 DeepLabV3+ 改进的图像语义分割算法相比,所提算法对特征提取的精确度有所上升,分割精度不劣于其他最新图像分割算法。所提算法虽然有了较好的提升,但是对于更加复杂场景下的图像分割效果略差,因此下一步的研究重点会放在提高复杂场景下图像分割的精确度上。

参 考 文 献

- [1] Wang E D, Qi K, Li X P, et al. Semantic segmentation of remote sensing image based on neural network[J]. Acta Optica Sinica, 2019, 39(12): 1210001. 王恩德, 齐凯, 李学鹏, 等. 基于神经网络的遥感图像语义分割方法[J]. 光学学报, 2019, 39(12): 1210001.
- [2] Hu X X, Yang K L, Fei L, et al. ACNET: attention based network to exploit complementary features for RGBD semantic segmentation[C]//2019 IEEE International Conference on Image Processing (ICIP), September 22-25, 2019, Taipei, Taiwan, China. New York: IEEE Press, 2019: 1440-1444.
- [3] Zhang S, Li Y P. Retinal vascular image segmentation based on improved HED network[J]. Acta Optica Sinica, 2020, 40(6): 0610002. 张赛, 李艳萍. 基于改进 HED 网络的视网膜血管图像分割[J]. 光学学报, 2020, 40(6): 0610002.
- [4] Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1520-1528.
- [5] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
- [6] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6230-6239.
- [7] Wang P Q, Chen P F, Yuan Y, et al. Understanding convolution for semantic segmentation [C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV), March 12-15, 2018, Lake Tahoe, NV, USA. New York: IEEE Press, 2018: 1451-1460.
- [8] Tan G H, Hou J, Han Y P, et al. Low-parameter real-time image segmentation algorithm based on convolutional neural network[J]. Laser & Optoelectronics Progress, 2019, 56(9): 091003. 谭光鸿, 侯进, 韩雁鹏, 等. 基于卷积神经网络的低参数量实时图像分割算法[J]. 激光与光电子学进展, 2019, 56(9): 091003.
- [9] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.
- [10] Chaurasia A, Culurciello E. LinkNet: exploiting encoder representations for efficient semantic segmentation[C]//2017 IEEE Visual Communications and Image Processing (VCIP), December 10-13, 2017, St. Petersburg, FL, USA. New York: IEEE Press, 2017: 17614349.
- [11] Tian Z, He T, Shen C H, et al. Decoders matter for semantic segmentation: data-dependent decoding enables flexible feature aggregation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3121-3130.
- [12] Wang F, Jiang M Q, Qian C, et al. Residual attention network for image classification[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6450-6458.
- [13] Fu J L, Zheng H L, Mei T. Look closer to see better: Recurrent attention convolutional neural

- network for fine-grained image recognition[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 4476-4484.
- [14] Fu J, Liu J, Tian H J, et al. Dual attention network for scene segmentation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3141-3149.
- [15] Cao J X, Chen Q, Guo J, et al. Attention-guided context feature pyramid network for object detection [EB/OL]. (2020-05-23) [2021-02-25]. <https://arxiv.org/abs/2005.11475v1>.
- [16] Li H C, Xiong P F, An J, et al. Pyramid attention network for semantic segmentation[EB/OL]. (2018-05-25)[2021-02-25]. <https://arxiv.org/abs/1805.10180>.
- [17] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[EB/OL]. (2014-12-22) [2021-02-20]. <https://arxiv.org/abs/1412.7062v2>.
- [18] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. (2017-06-17) [2021-02-20]. <https://arxiv.org/abs/1706.05587>.
- [19] 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.
- [20] Chen L C, Zhu Y K, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 833-851.
- [21] Chollet F. Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1800-1807.
- [22] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [23] Abadi M, Agarwal A, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems[EB/OL]. (2016-03-14) [2021-02-20]. <https://arxiv.org/abs/1603.04467>.
- [24] Zhao T, Wu X Q. Pyramid feature attention network for saliency detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3080-3089.
- [25] Peng C, Zhang X Y, Yu G, et al. Large kernel matters: improve semantic segmentation by global convolutional network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1743-1751.
- [26] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.
- [27] Fu J, Liu J, Tian H J, et al. Dual attention network for scene segmentation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3141-3149.
- [28] Zhong Z L, Lin Z Q, Bidart R, et al. Squeeze-and-attention networks for semantic segmentation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 13062-13071.
- [29] Xu C, Wang L. Image semantic segmentation method based on improved DeepLabv3+ network [J]. Laser & Optoelectronics Progress, 2021, 58(16): 1610008.
徐聪, 王丽. 基于改进DeepLabv3+网络的图像语义分割方法[J]. 激光与光电子学进展, 2021, 58(16): 1610008.
- [30] Liu W X, Shu Y Z, Tang X M, et al. Remote sensing image segmentation using dual attention mechanism Deeplabv3+ algorithm[J]. Tropical Geography, 2020, 40(2): 303-313.
刘文祥, 舒远仲, 唐小敏, 等. 采用双注意力机制 Deeplabv3+算法的遥感影像语义分割[J]. 热带地理, 2020, 40(2): 303-313.
- [31] Ren F L, He X, Wei Z H, et al. Semantic segmentation based on DeepLabV3+ and superpixel optimization[J]. Optics and Precision Engineering, 2019, 27(12): 2722-2729.
任风雷, 何昕, 魏仲慧, 等. 基于DeepLabV3+与超像素优化的语义分割[J]. 光学精密工程, 2019, 27(12): 2722-2729.