激光与光电子学进展 先进成像

DECANet:基于改进DeepLabv3+的图像语义分割方法

唐璐[1,2],万良[1,2*],王婷婷[1,2],李树胜[1,2]

[1]贵州大学计算机科学与技术学院,贵州 贵阳 550025;

[2]贵州大学计算机软件与理论研究所,贵州 贵阳 550025

摘要 在图像的语义分割任务中,不同对象之间像素值存在差异,导致现有的网络模型在图像语义分割过程中丢失图像局部细节信息。针对上述问题,提出一种图像语义分割方法(DECANet)。首先,引入通道注意力网络模块,通过对所有通道的依赖关系进行建模提高网络的表达能力,选择性地学习并强化通道特征,提取有用信息,抑制无用信息。其次,利用改进的空洞空间金字塔池化(ASPP)结构,对提取到的图像卷积特征进行多尺度融合,减少图像细节信息丢失,且在权重参数不改变的情况下提取语义像素位置信息,加快模型的收敛速度。最后,DECANet在PASCAL VOC2012和Cityscapes数据集上的平均交并比分别达81.08%和76%,与现有的先进网络模型相比,检测性能更优,可以有效地捕获局部细节信息,减少图像语义像素分类错误。

关键词 图像语义分割;注意力机制;空洞空间金字塔池化;多尺度融合

中图分类号 TP391 **文献标志码** A **DOI**:10.3788/LOP212704

DECANet: Image Semantic Segmentation Method Based on Improved DeepLabv3+

Tang Lu[1,2], Wan Liang[1,2*], Wang Tingting[1,2], Li Shusheng[1,2]

[1]*College of Computer Science and Technology, Guizhou University, Guiyang* 550025, *Guizhou, China*;

[2]*Institute of Computer Software and Theory, Guizhou University, Guiyang* 550025, *Guizhou, China*

Abstract The variation in pixel values between different objects during semantic segmentation of images leads to the loss of local image details in existing network models. An image semantic segmentation method (DECANet) is proposed to solve this problem. First, a channel attention network module is introduced to improve network clarity by modeling the dependencies of all channels, selectively learning and reinforcing channel features, and extracting useful information to suppress useless data. Second, using an improved atrous space pyramidal pooling (ASPP) structure, the extracted image convolutional features are multiscale fused to reduce the loss of image detail information, and the semantic pixel location information is extracted without increasing the weight parameters to speed up the model's convergence. Finally, the mean intersection over union of the proposed method reaches 81.08% and 76% on PASCAL VOC2012 and Cityscapes datasets, respectively. The detection performance of the DECANet is superior to the existing state-of-the-art network models, which can effectively capture local detail information and reduce image semantic pixel classification errors.

Key words image semantic segmentation; attention mechanism; atrous space pyramidal pooling (ASPP); multi-scale fusion

1 引 言

图像语义分割[1-3]是计算机视觉领域的一个重要分支,被广泛应用于自动驾驶和医学影像等场景。它是对图像像素逐一进行分类,从而解析图像深层语义信息的过程。深度卷积神经网络(DCNN)[4]在图像语义分割中应用广泛,Long等[5]提出了全卷积神经网络(FCN),但经过一系列卷积操作后,网络处理的图像分辨率不断降低,容易造成图像局部细节中像素丢失。Yu等[6]提出的空洞卷积(atrous convolution)聚合多尺度上下文信息,不会丢失图像分辨率,增大了感受野的范围,获得了密集预测结果。DeconvNet[7]使用堆叠的

收稿日期:2021-10-11;**修回日期**:2021-11-29;**录用日期**:2021-12-21;**网络首发日期**:2022-01-03

基金项目:国家自然科学基金(62062020)

通信作者:*wanliangtr@163.com

反卷积层逐步恢复对全分辨率的预测。Wang等[8]通过密集的上采样卷积(DUC)生成像素级预测结果。DeepLabv1[9]引入空洞卷积,在扩大感受野的同时没有增加参数量,采用全连接条件随机场(CRF)优化边界,能够更好地预测图像中的目标边界信息。DeepLabv2[10]对空洞卷积与空间金字塔池化(SPP)算法[11]进行融合,构建了空洞空间金字塔池化(ASPP)算法,该算法采用不同空洞率的空洞卷积对输入的特征信息图进行并行操作。DeepLabv3[12]舍弃了CRF,在ASPP中加入全局平均池化,分割准确率优于前两种算法,进一步提升了性能。DeepLabv3+[13]对空洞卷积和深度可分离卷积进行融合后替换标准卷积,使得模型的参数量得到了一定的减少,同时也增大了感受野,在DeepLabv3基础上提升了分割效果。BiSeNet[14]通过双边结构提取空间信息和上下文信息。APCNet[15]利用不同尺度下的特征构建多尺度特征,获得融合多尺度上下文信息的语义特征表示。DFANet[16]通过将提取到的特征信息重复使用,进行语义层和空间层的特征信息融合[17]。Wavesnet[18]通过在特征图下采样时应用离散小波变换(DWT)来提取数据细节,在上采样时采用逆向离散小波变换(IDWT)来恢复细节。N-DeepLabv3+[19]能够有效提高小尺度目标关注度,缓解目标误分割问题。王鑫等[20]提出了基于八度卷积的实时语义分割网络,该网络解决了在处理颜色变化小的像素区域时存在的空间冗余问题,在保证精度的同时减少了模型计算量。这些方法同时捕获低级空间信息和高级上下文信息,但准确性仍有进一步提高的空间。

注意力机制[21-22]能够忽略不相关信息而获取重要的特征信息。压缩激励网络(SENet)[23]中的SE模块可以学习不同通道上的特征映射权重,将重要的特征映射凸显出来,忽略掉不重要的特征映射。ECANet[24]为了降低模型参数量舍弃了SE模块中的全连接(FC)层,用一维卷积进行代替,同时舍弃了通道维数减少后再恢复的步骤。

随着网络层数的加深,DeepLabv3+网络模型在提取图像特征信息时往往会丢失特征图中的细节信息,从而出现部分微小的物体被误分的情况。本文的贡献如下。1)提出一种新的编码器,该编码器结构主要由DCNN、ECANet模块和并联的DS_ASPP模块组成。首先,利用DCNN提取图像的特征信息。其次,运用ECANet模块加强模型提取各个特征通道信息的能力,将高低层次的语义特征信息有效融合在一起,从而使得细微物体中各个对象之间的几何纹理特征信息得到完好保存,降低像素级分类错误率,优化网络性能,实现更加细致的图像分割效果。最后,利用并联的DS_ASPP模块对特征图进行多尺度处理,然后对得到的特征进行融合,从而高效地捕捉物体的局部细节信息,解决图像像素之间信息利用率不高的问题。2)提出DECANet模型,其提取到的图像特征图中的像素级信息更加精准。同时在PASCAL VOC2012和Cityscapes两个大规模数据集上对所提方法与其他最先进方法进行对比,所提方法表现优异。

2　DeepLabv3+模型

DeepLabv3+网络模型如图1所示,主要由编码器和解码器组成[25-26],该模型的编码器主要有骨干网络ResNet101[27]和ASPP模块。

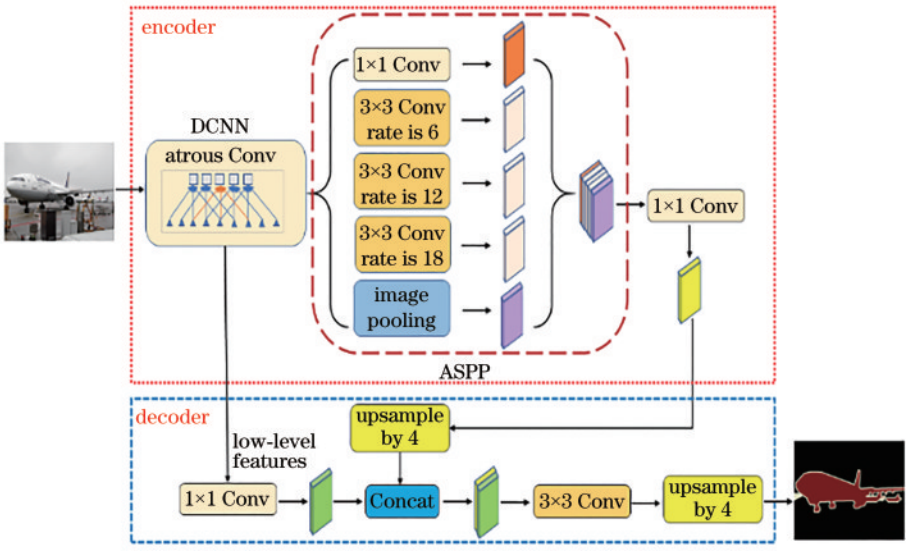

图1　DeepLabv3+模型的结构

Fig. 1　Structure of the DeepLabv3+ model

在编码器中,采用骨干网络ResNet101提取图像特征,获取低层次细节信息,然后将低层次细节信息输入到ASPP模块中。ASPP模块主要对不同扩张率的空洞卷积层通过并联的方式进行连接,其中全局平

均池化的作用主要是提取全局信息。ASPP模块利用骨干网络得到的信息进行多尺度采样，生成多尺度的特征图，在编码器尾部对多尺度的高级语义特征图在特征通道维度进行拼接，最后通过1×1的卷积压缩通道数。在解码器中，采用双线性插值4倍上采样后，对结果与骨干网络ResNet101得到的低级语义信息的特征图进行跨层融合，捕获浅层特征包含的细节信息，进一步丰富图像的语义信息和细节信息；用两个3×3卷积提取特征，特征图尺寸通过双线性插值4倍上采样逐步恢复到原始图像大小，产生最终的语义分割预测结果图。

3　DeepLabv3+模型改进

3.1　DECANet模型

在DeepLabv3+模型基础上提出一种改进的图像语义分割模型(DECANet)，旨在解决在进行图像语义分割过程中图像局部细节信息丢失的问题，提升语义分割精度。DECANet整体模型架构如图2所示。

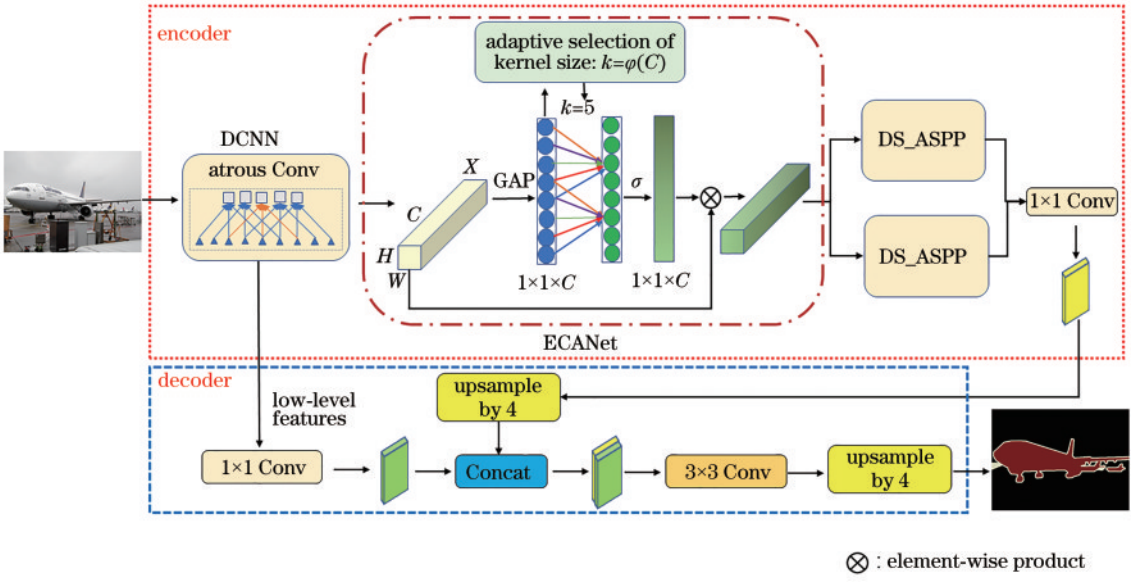

图2　DECANet模型的结构

Fig. 2　Structure of the DECANet model

在编码器引入有效的通道注意力网络(ECANet)模块，图像通过DCNN处理后得到特征图，将特征图输入到ECANet可以加强各个通道中像素间的联系。ECANet模块利用通道之间的关联性加强网络模型选择有效目标特征的能力，使参数的利用率得到大幅度提高，提升了分类精度从而优化图像分割效果。利用深度可分离卷积与空洞卷积构建深度可分离空洞卷积(DSAConv)，将ASPP模块中标准卷积全部更换成深度可分离空洞卷积，在很大程度上可以降低网络模型在训练中产生的参数量，并且可以提升网络模型的分割精度，大大提高训练效率。其次，对ASPP模块进行微调，即将ReLU函数替换为Leaky ReLU函数，加入batchnorm等操作对模型进行优化处理，称改进的ASPP模块为DS_ASPP模块。将经过ECANet模块处理的特征图输入并联的DS_ASPP模块中进行多尺度目标分割，生成多尺度的特征图。最后对多尺度的高级语义特征图在通道维度上进行拼接融合，融合结果通过1×1的卷积降低特征通道数。

3.2　注意力机制模块

图像经过DCNN操作之后，特征图中的各个通道都保留着不同的特征信息，这些特征会影响图像的分割结果。在图像分割过程中，卷积核会采用相同的权重对各个通道进行处理，这种处理方式会导致获取的目标特征信息不准确。目前基于FCN的方法[28-29]通过预训练的DCNN可以获得相对高层次的上下文语义特征信息，从而有效地分割相对较大的目标物体，但对于不明显的小目标则分割得不准确。针对这些问题，本文引入ECANet模块，该模块主要捕捉图像的局部跨通道交互信息，为不同位置的像素建立关联性，增强对特征图的表征能力；并利用具有自适应核大小的一维卷积，使得模型自适应地为同一场景下的不同物体选择合适的分割范围，在较低的模型复杂度下取得了明显的性能提升。

1D卷积特征矩阵为

$$\begin{bmatrix} w^{1,1} & \cdots & w^{1,k} & \cdots & \cdots & 0 \\ 0 & w^{2,2} & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & w^{C,C-k+1} & \cdots & w^{C,C} \end{bmatrix}。\tag{1}$$

ECANet利用矩阵$\boldsymbol{w}_k$学习通道注意力，矩阵$\boldsymbol{w}_k$有$k\times C$个参数。

$$\boldsymbol{y}=\frac{1}{H\times W}\sum_{i=1}^{W}\sum_{j=1}^{H}x_{ij},\tag{2}$$

式中：$y\in\mathbf{R}^{1\times1\times C}$代表输入特征$x\in\mathbf{R}^{H\times W\times C}$的全局平均池化结果。通道交互的尺度由卷积核的大小$k$决定，而$\sigma$和$\omega$分别是Sigmoid函数和每个通道的权重。

$$\omega_i=\sigma\left(\sum_{j=1}^{k} w^j y_i^j\right), y_i^j\in\Omega_i^j, \tag{3}$$

式中：Ω_i^j表示y_i的k个邻域通道。

在1D卷积上采用自适应选择1D卷积核大小的方法，即卷积神经网络共享权重，使每一组的权重大小一样，参数数量从$k\times C$（其中C为通道数）缩减为k。

$$\omega=\sigma\left[\mathrm{C1D}_k(y)\right], \tag{4}$$

式中：C1D代表1D-CNN。C和k存在映射ϕ，卷积核数量为2的k次方，公式为

$$C=\phi(k)=2^{(\gamma\times k-b)}。 \tag{5}$$

因此，给定通道维数C，卷积核大小k为

$$k=\psi(C)=\left|\frac{\log_2 C}{\gamma}+\frac{b}{\gamma}\right|_{\mathrm{odd}}, \tag{6}$$

式中：$|t|_{\mathrm{odd}}$表示距离t最近的奇数；$\gamma=2$，$b=1$。

在DeepLabv3+中引入注意力机制模块，根据各个特征通道承载信息量的多少来对目标分割精度进行评判，同时附加各个特征通道的权重系数，有针对性地加强特征学习。这样做主要是为了凸显出对分割结果有重要作用的特征信息，抑制冗余的通道信息，从而提高模型整体的学习能力和泛化能力。

3.3　深度可分离空洞卷积

深度可分离卷积在降低模型参数量的同时可以保证模型的精度且提升计算速率。它将标准卷积（如图3所示）分解为逐深度卷积（depthwise convolution）（如图4所示）和逐点卷积（pointwise convolution）（如图5所示）。在标准卷积中，各个通道采用相同的卷积核，同时对输入通道中每个卷积核执行卷积运算，不同的卷积核用于提取不同方面的特征。深度可分离卷积中，在深度卷积的各个通道上采用不同的卷积核提取不同的特征，其中每个卷积核只对应一个特征通道，但

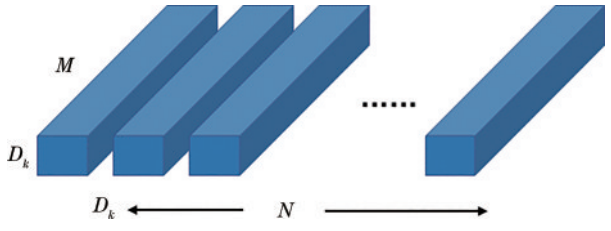

图3　标准卷积

Fig. 3　Standard convolution

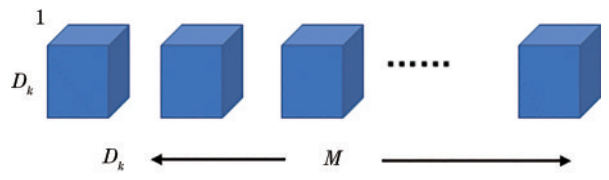

图4　逐深度卷积

Fig. 4　Depthwise convolution

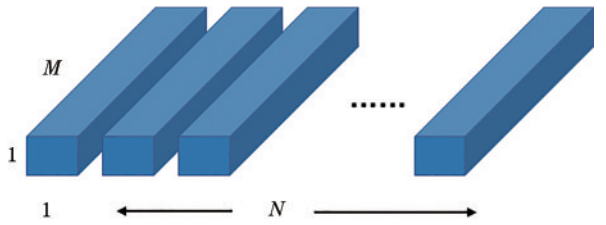

图5　逐点卷积

Fig. 5　Pointwise convolution

是这样对于某个通道来说，就只提取了一部分的特征。因此在此基础上进行逐点卷积，用1×1的卷积核对提取特征后的特征图再次提取不同方面的特征，最终产生和普通卷积相同的输出特征图。深度可分离卷积在提取特征图特征信息的过程中，会导致图像的分辨率降低，可以通过在逐深度卷积过程中引入空洞卷积对其进行处理，将最后得到的卷积称为深度可分离空洞卷积。深度可分离卷积能够大大降低网络模型参数量，保证模型精度且提升计算速率。空洞卷积在卷积核元素之间加入一些空格（零）来增强卷积核操作，在扩大感受野的同时能够获得更加密集的数据，不丢失特征图的分辨率，且保持像素的相对空间位置不变。扩大感受野可以提高对语义分割任务中细小物体的识别。用深度可分离空洞卷积替换ASPP模块中的标准卷积，能够有效地提升对分割物体的预测精度，提高网络的训练速率[30]。

4　实验

4.1　数据集介绍

PASCAL VOC2012是计算机视觉中图像语义分割领域知名的公共标准数据集，总共有21个分割种类，分别为20个物体对象和1个背景类别标签，包含人、动物、生活用品等。实验使用PASCAL VOC2012增强版数据集[31]，其中训练图像有10582张，验证图像有1449张，测试图像有1456张，输入图片大小为513×513。

Cityscapes是城市街道场景的语义分割图片数据集[32]。它主要包含来自50个城市的不同街道场景，拥有5000张在城市环境中具有高质量像素级标注的街景图像，图片分辨率为1024×2048像素，共有19个语义类别。其中训练集2975张图像，验证集500张图像，测试集1525张图像，输入图片大小为769×769。

4.2　实验设计

本实验基于Pytorch网络框架实现，实验配置如表1所示。

学习率使用随迭代频次增长而逐渐减少的“poly”[33]，表达式为

$$r=r_{\mathrm{base}}\times\left(1-\frac{T}{T_{\max}}\right)^{p_{\mathrm{power}}}。 \tag{7}$$

数据集PASCAL VOC2012和Cityscapes的初始

表 1　机器软硬件配置

Table 1　Machine software and hardware configurations

Project	Detail
CPU	AMD EPYC 7742 64-Core Processor
RAM	32G
GPU	NVIDIA Tesla A100 40G
Operating system	Ubuntu 18.04.1
CUDA	Cuda 11.0
Data processing	Python 3.6

学习率(r_{base})分别为 0.007 和 0.01，正在训练时的迭代频次为 T，最大迭代次数 T_{max} 为 30000，r 为当前迭代频次的学习率。优化器使用带动量的随机梯度下降法(SGD)，p_{power} 和动量都为 0.9，权重衰减率为 0.0005。

在图像分割领域，平均交并比(MIoU)是衡量图像分割精度的重要指标，实验使用 MIoU 作为标准评价语义分割算法的准确率。交并比是预测结果与真实语义结果的交集除以它们的并集的结果，平均交并比表示所有物体类别交并比的平均值。IoU 和 MIoU 的定义分别为

$$R_{\mathrm{IoU}}=\frac{f_{mm}}{\sum_{n=0}^{s} f_{mn}+\sum_{n=0}^{s} f_{nm}-f_{mm}}, \tag{8}$$

$$R_{\mathrm{MIoU}}=\frac{1}{s+1}\sum_{m=0}^{s} R_{\mathrm{IoU}\,m}, \tag{9}$$

式中：类别数为 $s+1$；f_{mn} 表示真实值为 m 类像素但预测为 n 类像素点；f_{nm} 表示真实值为 n 类像素但预测为 m 类像素点；f_{mm} 表示预测正确的 m 类像素点。

4.3　实验结果与分析

4.3.1　不同算法下损失函数对比图

图 6 是 DeepLabv3+和 DECANet 网络训练 30000 次后，两者的损失函数值。由此可知：随着训练次数的不断增加，DECANet 模型损失值一直减小且逐渐趋于平缓；DeepLabv3+模型的损失值在减小的过程中波动性比较大，模型的性能没有 DECANet 模型稳定。损失值波动越平缓，表明 DECANet 模型的拟合效果越好，鲁棒性也越好。

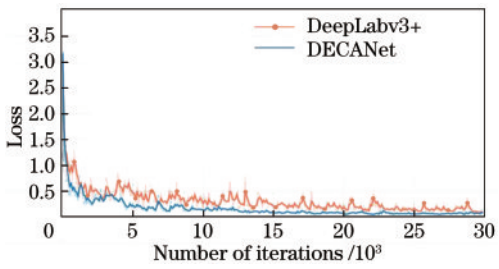

图 6　不同算法下实现语义分割的损失函数图

Fig. 6　Loss function graph for implementing semantic segmentation under different algorithms

4.3.2　在 PASCAL VOC2012 上的实验分析结果

研究通道注意力机制加入前后对网络性能的影响。实验结果如表 2 所示，其中 ECANet 为有效通道注意力网络，PS 为金字塔尺度。首先，用单一尺度来验证通道注意力机制的有效性，实验结果表明，加入通道注意力后，单尺度模型性能提高了 2.42 个百分点。然后，使用不同采样率的空洞卷积来构建多尺度特征，加入 ECANet 模块后，网络模型性能最大提高了 2.23 个百分点。因此可以得知，ECANet 模块有助于捕捉目标特征信息，提高网络的分割性能。

表 2　在 PASCAL VOC2012 验证集上的对比结果

Table 2　Comparison results on PASCAL VOC2012 validation set

Backbone	PS	ECANet	MIoU /%
ResNet-101	1		70.31
		√	72.73
SENet-101	1,6,12,18		76.77
		√	78.96
ResNet-101	1,6,12,18		78.85
		√	81.08

如表 3 所示，对 DS_ASPP 模块进行并联处理时，简称模型为 DECANet_1；当 DS_ASPP 模块与 ECANet 模块进行并联处理时，简称模型为

表 3　不同模型的 IoU

Table 3　IoU of different models　　unit: %

Category	DECANet_1	DECANet_2	DECANet_3	DECANet
background	94.66	94.85	95.13	95.10
bicycle	44.24	44.46	44.15	44.74
boat	77.45	71.12	77.20	73.92
bus	95.51	95.20	95.81	96.03
cat	92.07	94.54	94.13	94.07
cow	90.34	88.66	92.20	92.31
dog	89.50	91.45	90.41	90.09
motorbike	86.60	87.65	85.93	87.21
pottedplant	67.09	68.04	68.03	67.05
sofa	50.61	48.65	59.57	55.46
tvmonitor	75.75	79.31	75.54	79.93
aeroplane	92.48	92.92	92.81	93.09
bird	91.65	91.21	89.59	90.97
bottle	80.24	82.26	81.97	81.19
car	88.86	89.84	88.70	89.78
chair	42.67	44.09	47.65	51.09
diningtable	56.79	53.91	55.39	57.08
horse	87.10	87.89	90.19	91.70
person	88.05	88.56	88.24	88.71
sheep	91.37	89.74	91.45	91.91
train	86.07	90.05	92.75	91.34
MIoU	79.48	79.73	80.80	81.08

DECANet_2;将ECANet模块与DS_ASPP模块进行串联处理时,简称模型为DECANet_3;DS_ASPP进行并联处理,将ECANet模块与其进行串联处理,简称模型为DECANet。表3显示了不同位置的ECANet模块和DS_ASPP模块以不同的组合形式相结合后对图像语义分割精度的影响,在PASCAL VOC2012数据集上分别实现了79.48%、79.73%、80.80%、81.08%的MIoU,实验效果均优于DeepLabv3+网络。从提出的4种分割方法中可以得知:DECANet方法是最好的,能够有效地提高语义分割的精度。本文在实验过程中使用取得最优的策略进行训练和验证评估。

为了验证所提算法精度优于其他图像语义分割算法精度,基于数据集PASCAL VOC2012,不同语义分割算法的对比结果如表4所示。

表4 不同算法在PASCAL VOC2012验证集上的MIoU值对比结果

Table 4 Comparison results of MIoU values of different algorithms on PASCAL VOC2012 validation set

Method	Backbone network	MIoU /%
FCN[5]	VGG16	62.20
DeepLabv1[9]	VGG16	68.70
DeepLabv2[10]	ResNet101	71.60
DeconvNet[7]	VGG16	72.50
DeepLabv3[12]	ResNet101	77.21
DeepLabv3+[13]	ResNet101	78.85
APCNet[15]	ResNet101	80.71
WaveSNet[18]	ResNet101	79.90
DECANet_1	ResNet101	79.48
DECANet_2	ResNet101	79.73
DECANet_3	ResNet101	80.80
DECANet	ResNet101	81.08

PASCAL VOC2012数据集的可视化结果如图7所示。从图7的第一行可以看出,DeepLabv3+网络没有将鹤的腿部信息完整地分割出来;第二行中,DeepLabv3+网络没有将左侧的瓶子完整地分割出来;第三行中,马的腿和臀部被误分类为其他类别;第四行中,第二只鹦鹉的尾巴没有被识别出来,图像背景部分误分类为第三只鹦鹉的尾巴,导致鹦鹉尾巴的边缘分割不准确;第五行中,自行车的把手没有被准确地分割出来;第六行中,主要是飞机的尾翼没有被完整地识别出来;第七行中,整体人物模样已经被分割出来,但是对于人物衣服的轮廓分割很模糊;第八行的马,主要和第二行一样,对腿部的分割明显存在丢失。综上可以得知,目前DeepLabv3+网络分割出来的物体不是很完整,对物体某些细节性部位分割得不是很清晰,会存在将某一对象误分类为另一对象的现象。改进后的DECANet模型主要解决模型局部分割不准确的问题,对鸟、瓶子、人、马、自行车、飞机等不同种类物体,都获得了较好的分割效果,可以更完整地勾画出鸟、瓶子、人和飞机等物体的轮廓;且对于比较微小的对象,有着更加细致的分割,整体分割效果相对较优。由此可知,对物体对象中的大部分区域,DECANet方法可以有效恢复高级特征丢失的细节信息,与DeepLabv3+相比,分割效果比较显著,与原图融合的效果整体上较优,最终获得比较好的语义分割结果。

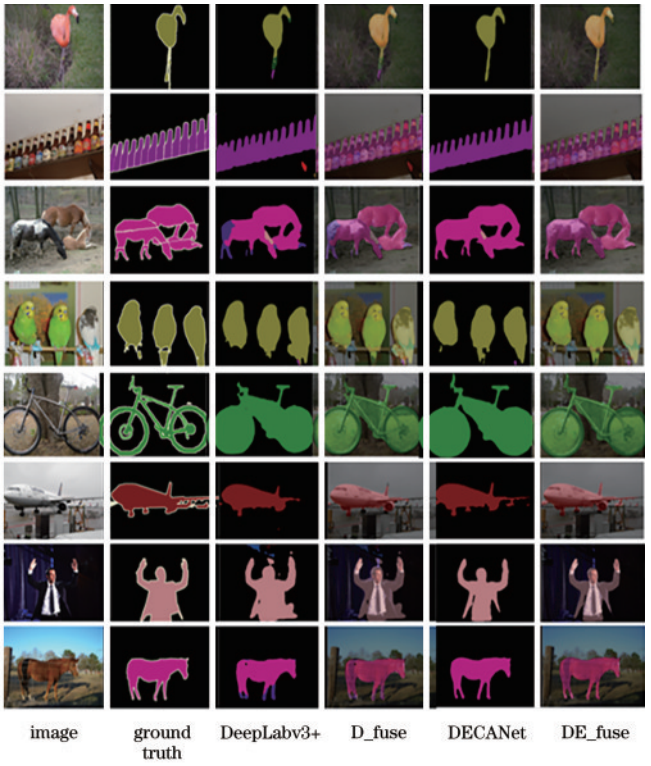

图7 在PASCAL VOC2012验证集上的可视化结果

Fig. 7 Visualization results on PASCAL VOC2012 validation set

4.3.3 在Cityscapes上的实验分析结果

DeepLabv3+网络模型与所提DECANet模型的语义分割实验结果如表5所示。由表5可以看出:与DeepLabv3+语义分割算法相比,所提模型对大部分物体对象都有着更好的分割准确性,语义分割精度得到了进一步的提高。

最后,为了进一步验证DECANet方法的泛化性,将所提算法精度与其他图像语义分割算法精度进行了对比,基于Cityscapes数据集,不同语义分割算法的对比结果如表6所示。

城市景观数据集的可视化结果如图8所示。对比方框标记区域,从第一行的可视化结果可以得知:DeepLabv3+网络对图像中交通标志物公交车站台的预测精度不高,导致公交车站台的图像分割信息丢失严重;DECANet模型通过提取丰富的图像语义特征信息,对图像中的交通标志物公交车站台具有更强的预测能力,并对其边界信息进行了优化,优化后的物体接近于真实标签图。从第二行的图像可视化结果可以得

表 5　两种算法的 IoU

Table 5　IoU of two algorithms　　unit: %

Category	DeepLabv3+	DECANet
road	0.978	0.980
sidewalk	0.826	0.842
building	0.917	0.921
wall	0.488	0.483
fence	0.556	0.568
pole	0.591	0.637
traffic light	0.654	0.694
traffic sign	0.746	0.780
vegetation	0.920	0.921
terrain	0.623	0.618
sky	0.944	0.949
person	0.802	0.817
rider	0.594	0.633
car	0.943	0.946
truck	0.745	0.747
bus	0.789	0.837
train	0.637	0.666
motorcycle	0.602	0.648
bicycle	0.752	0.764
MIoU	0.742	0.760

知：DeepLabv3+网络对右侧路面分割的结果存在一定的信息丢失，没有准确识别出右侧路面地形类别；

表 6　不同算法在 Cityscapes 验证集上的 MIoU 值对比结果

Table 6　Comparison results of MIoU value of different algorithms on Cityscapes validation set

Method	MIoU /%
FCN[5]	63.1
DeepLabv1[9]	63.1
BiSenet[14]	69.0
DFANet[16]	71.3
DeepLabv3+[13]	73.9
N-Deeplabv3+[19]	74.6
DECANet	76.0

DECANet 模型准确地识别了右侧路面地形，整体分割效果优于 DeepLabv3+分割效果。从第三行的可视化结果可以得知：DeepLabv3+网络对图像中的中小尺寸黄色交通标志物和路面信息的分割效果欠佳；但所提模型能够有效地将中小尺寸黄色交通标志物和路面信息完整分割出来。从第四行的图像可视化结果可以得知：DeepLabv3+网络将道路场景的右侧部分错误地分类为其他类别，对整体的预测结果存在较大误差；DECANet 模型则准确地预测出了图像道路场景部分，解决了误分类的问题。由图 8 可知，DECANet 方法分割效果更好，物体局部细节信息突出较为明显，在图像整体语义分割效果上具有一定提升。

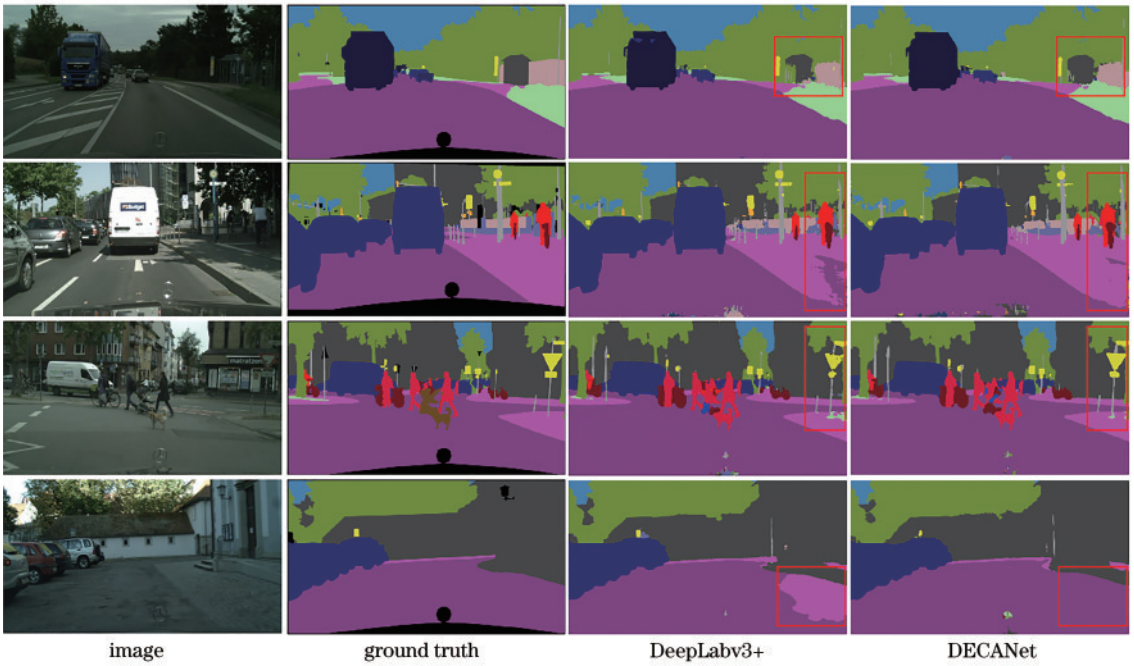

图 8　在 Cityscapes 验证集上的可视化结果

Fig. 8　Visualization results on Cityscapes validation set

5　结　　论

提出的 DECANet 模型方法可以有效地捕获图像特征信息，通过对各个特征通道的相关性进行建模，捕获高层次的上下文信息，通过捕获到的全局信息可以选择性地学习对分类结果有用的特征信息并剔除没有用的特征。所提方法对细节信息的处理更加完善，进一步优化了图像语义分割中的局部分支，出现错误分类的情况比较少，优化了分类的效果，准确地划分出了物体的种类，最终在物体整体语义分割效果上具有一

定提升。大量的实验结果证实所提方法在两个具有挑战性数据集上的有效性。未来进一步的工作是提升物体总体的分割精度，主要聚焦在对物体边缘部分的分割，最后将物体的局部细节分割信息和边缘细节分割信息结合起来，促使网络生成一个语义目标更清晰、纹理细节更丰富的图像分割效果。

参 考 文 献

[1] 王龙飞，严春满. 道路场景语义分割综述[J]. 激光与光电子学进展，2021，58(12)：1200002.
Wang L F，Yan C M. Review on semantic segmentation of road scenes[J]. Laser & Optoelectronics Progress，2021，58(12)：1200002.

[2] 蔡雨，黄学功，张志安，等. 基于特征融合的实时语义分割算法[J]. 激光与光电子学进展，2020，57(2)：021011.
Cai Y，Huang X G，Zhang Z A，et al. Real-time semantic segmentation algorithm based on feature fusion technology[J]. Laser & Optoelectronics Progress，2020，57(2)：021011.

[3] Jiang H J，Wang R P，Shan S G，et al. Adaptive metric learning for zero-shot recognition[J]. IEEE Signal Processing Letters，2019，26(9)：1270-1274.

[4] Lecun Y，Bottou L，Bengio Y，et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE，1998，86(11)：2278-2324.

[5] Long J，Shelhamer E，Darrell T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition，June 7-12，2015，Boston，MA，USA. New York：IEEE Press，2015：3431-3440.

[6] Yu F，Koltun V. Multi-scale context aggregation by dilated convolutions[EB/OL]. (2015-11-23)[2021-04-05]. https://arxiv.org/abs/1511.07122v1.

[7] Noh H，Hong S，Han B. Learning deconvolution network for semantic segmentation[C]//2015 IEEE International Conference on Computer Vision，December 7-13，2015，Santiago，Chile. New York：IEEE Press，2015：1520-1528.

[8] 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，March 12-15，2018，Lake Tahoe，NV，USA. New York：IEEE Press，2018：1451-1460.

[9] 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)[2021-05-04]. https://arxiv.org/abs/1412.7062.

[10] 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.

[11] He K M，Zhang X Y，Ren S Q，et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2015，37(9)：1904-1916.

[12] Chen L C，Papandreou G，Schroff F，et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. (2017-06-17)[2021-04-05]. https://arxiv.org/abs/1706.05587.

[13] 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.

[14] Yu C Q，Wang J B，Peng C，et al. BiSeNet：bilateral segmentation network for real-time semantic segmentation[M]//Ferrari V，Hebert M，Sminchisescu C，et al. Computer Vision-ECCV 2018. Lecture notes in computer science. Cham：Springer，2018，11217：334-349.

[15] He J J，Deng Z Y，Zhou L，et al. Adaptive pyramid context network for semantic 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，2019：7511-7520.

[16] Li H C，Xiong P F，Fan H Q，et al. DFANet：deep feature aggregation for real-time semantic 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，2019：9514-9523.

[17] 郭梦利，阮顺领，卢才武，等. 基于改进DeepLabv3+网络的露天矿路网提取方法[J]. 激光与光电子学进展，2021，58(22)：2228005.
Guo M L，Ruan S L，Lu C W，et al. Road extraction method of open-pit mine based on improved DeepLabv3+ network[J]. Laser & Optoelectronics Progress，2021，58(22)：2228005.

[18] Li Q F，Shen L L. WaveSNet：wavelet integrated deep networks for image segmentation[EB/OL]. (2020-05-29)[2021-05-06]. https://arxiv.org/abs/2005.14461.

[19] 孟俊熙，张莉，曹洋，等. 基于Deeplabv3+的图像语义分割算法优化研究[J]. 激光与光电子学进展，2022，59(16)：1610009.
Meng J X，Zhang L，Cao Y，et al. Research on optimization of image semantic segmentation algorithms based on Deeplabv3+ [J]. Laser & Optoelectronics Progress，2022，59(16)：1610009.

[20] 王鑫，吴开军. 基于八度卷积设计的实时语义分割网络[J]. 激光与光电子学进展，2022，59(8)：0810015.
Wang X，Wu K J. Real-time semantic segmentation network based on octave convolution[J]. Laser & Optoelectronics Progress，2022，59(08)：0810015.

[21] Fu J，Liu J，Jiang J，et al. Scene segmentation with dual relation-aware attention network[J]. IEEE Transactions on Neural Networks and Learning Systems，2021，32(6)：2547-2560.

[22] Yu C Q，Wang J B，Gao C X，et al. Context prior for scene 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: 12413-12422.

[23] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.

[24] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.

[25] 张哲晗, 方薇, 杜丽丽, 等. 基于编码-解码卷积神经网络的遥感图像语义分割[J]. 光学学报, 2020, 40(3): 0310001.
Zhang Z H, Fang W, Du L L, et al. Semantic segmentation of remote sensing image based on encoder-decoder convolutional neural network[J]. Acta Optica Sinica, 2020, 40(3): 0310001.

[26] 郭列, 张团善, 孙威振, 等. 融合空间注意力机制的图像语义描述算法[J]. 激光与光电子学进展, 2021, 58(12): 1210030.
Guo L, Zhang T S, Sun W Z, et al. Image semantic description algorithm with integrated spatial attention mechanism[J]. Laser & Optoelectronics Progress, 2021, 58(12): 1210030.

[27] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[28] 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.

[29] 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.

[30] 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.
Xu C, Wang L. Image semantic segmentation method based on improved DeepLabv3+ network[J]. Laser & Optoelectronics Progress, 2021, 58(16): 1610008.

[31] Everingham M, Gool L, Williams C K I, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

[32] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 3213-3223.

[33] Liu W, Rabinovich A, Berg A C. Parsenet: looking wider to see better[EB/OL]. (2015-06-15)[2021-05-06]. https://arxiv.org/abs/1506.04579.