

文章编号:1007-2780(2022)09-1199-10

A-LinkNet: 注意力与空间信息融合的 语义分割网络

杜敏敏, 司马海峰*

(河南理工大学 计算机学院, 河南 焦作 454000)

摘要: 针对道路图像语义分割中上下文信息不足以及空间细节信息易丢失等问题, 本文提出一种基于 LinkNet 模型的实时分割方法。首先, 在编码区域构建一种新的注意力机制, 捕获道路图像的位置以及通道依赖, 增加目标特征的提取能力。然后, 在中心区域引入空洞空间金字塔池化模型, 在不影响图像分辨率的情况下捕获更加丰富的多尺度特征。在通用数据库上的实验结果表明, 所提方法在 Cityscapes 数据集上 MIOU 达到了 64.78%, 与 LinkNet 相比较, 提高了 5.01%, 同时对于细小目标物体以及边界分割视觉效果有明显的改善, 分割准确率获得了较大提升。

关键词: 语义分割; 注意力机制; 空洞空间金字塔池化; LinkNet

中图分类号: TP391 **文献标识码:** A **doi:** 10.37188/CJLCD.2022-0046

A-LinkNet: semantic segmentation network based on attention and spatial information fusion

DU Min-min, SIMA Hai-feng*

(School of Computer Science and Technology, Henan Polytechnic University,
Jiaozuo 454000, China)

Abstract: To address insufficient contextual information and partial loss of spatial detail information in semantic segmentation of road images, a real-time segmentation method is proposed based on LinkNet. Firstly, a new attention mechanism is constructed in the encoding to capture the location and channel dependence of road images to increase the extraction capability of target features. Then, an atrous spatial pyramid pooling is introduced in the central region to capture richer multi-scale features without affecting image resolution. The experimental results on the general database show that the proposed method achieves 64.78% MIOU on the Cityscapes dataset, which is 5.01% higher in comparison with LinkNet. And it can significantly improve the visual effect of fine target objects and boundary segmentation, and the segmentation accuracy has been greatly improved.

Key words: image semantic segmentation; attention mechanism; atrous spatial pyramid pooling; LinkNet

收稿日期: 2022-02-05; 修订日期: 2022-03-03.

基金项目: 国家自然科学基金 (No. 61602157); 河南省科技攻关项目 (No. 202102210167)

Supported by National Natural Science Foundation of China (No. 61602157); Key Scientific and Technological Project in Henan Province (No. 202102210167)

*通信联系人, E-mail: smhf@hpu.edu.cn

1 引言

作为图像理解的基础,图像语义分割是指根据语义信息将图像划分为不同的区域,而相同的语义区域具有相同的颜色^[1]。其中,道路场景的图像语义分割作为自动驾驶的核心技术,近几年成为研究的热点之一。然而,由于受到光照、天气等各种外界因素的影响,道路图像中物体具有位置关系复杂和种类变化大的特点,使得城市道路图像实时分割变成一项具有挑战性的任务^[2]。一方面,由于道路图像涵盖目标种类多且存在很多细小目标,例如栏杆、交通指示灯等,增加了图像语义分割的难度。另一方面,同一物体在不同图像中呈现出不同的大小,如何保存多尺度信息成为分割的关键。

随着神经网络的发展,图像语义分割进入新的时代,分割速度与精度得到了大幅提高。2014年,FCN^[3]的出现将图像语义分割带入一个新的时代。该模型使用卷积层替换传统网络中的全连接层,然后通过上采样将提取的特征还原成原图大小,极大地促进了语义分割的发展。但是该方法对于复杂场景图像的分割并不理想,因此后续很多模型都是在FCN的基础上进行改进。

2015年,Ronneberger等人^[4]提出的U-Net网络将编码器-解码器结构应用到图像语义分割中,给图像语义分割带来创新。通过跳跃连接将编码器与解码器直接相连,实现了编码层的特征复用,一定程度上弥补了FCN细节特征丢失的问题,但是忽视了不同特征的重要程度。2017年,Badrinarayanan等人^[5]提出了用于自动驾驶的深层语义分割模型,该模型与FCN结构相似,使用VGG-16的卷积层作为编码网络并保留特征图的最大池化索引,在解码器中利用最大池化索引进行上采样,提高了网络性能。虽然上述结构能取得较好的分割精度,但也存在参数多、计算量大的问题,因此研究者在如何提高编码器-解码器结构的分割速度上也进行了大量的探索。ENet^[6]、LEDNet^[7]模型使用非对称的编码器-解码器结构,减少了参数量,有效地提高了语义分割的速度。目前,编码器-解码器模型在图像处理领域得到广泛应用。

此外,在深度神经网络中,普遍使用池化操

作进行下采样,虽然可以扩大感受野,但是在上采样过程中容易导致精度的损失。针对这一问题,Yu等人^[8]提出了空洞卷积(Atrous Convolution)思想,在保证不改变图像分辨率的前提下增大感受野。但是如果在较深层网络使用采样率较大的空洞卷积,会导致“网格效应”,造成局部特征的丢失。为了解决这一问题,Wang等人^[9]提出了HDC结构,与普通空洞卷积不同,该结构采用了具有不同膨胀率的空间卷积来保证感受野的连续性。而文献[10]则提出了空洞空间金字塔池化模型(Atrous Spatial Pyramid Pooling, ASPP),使用一组具有不同膨胀率的空间卷积并行连接来捕获图像的上下文信息,提高模型的性能。2017年,Chen等人^[11]又在文献[10]的基础上,在ASPP模块中加入了图片级特征,对输入特征做全局平均池化,然后与并行的空洞卷积相融合,可以更好地捕获全局上下文信息。

上述网络结构在一定程度上提高了语义分割网络的精度,但是没有考虑到不同特征信息之间的依赖程度,不能很好地区分特征的重要性,注意力机制的出现则很好地解决了这一问题。注意力机制的主要思想是为图片中的关键特征赋予一层新的权重,从而使神经网络能够学习到图片中需要关注的地方,增加上下文的相关性。2018年,Hu等人^[12]提出了一种全新的可以嵌套到任意网络的子结构SENet,该结构通过网络自主学习来获取不同通道的权值,从而根据不同的权值表示不同特征通道的重要性,建模各个通道之间的依赖性。文献[13]在SENet的基础上提出一个简单有效的注意力模块CBAM,与SENet不同的是,该模块结合了空间和通道两个维度上的注意力机制,取得了更好的分割结果。Fu等人^[14]又在CBAM的基础上提出一种适用于自然场景图像的DANet,通过结合自注意力机制,并行连接空间注意力机制以及通道注意力机制,最后将两个模块的输出相加,进一步提高了分割结果的准确度。

Chaurasia等^[15]提出的LinkNet是一种有效的实时图像语义分割网络,结合残差结构^[16]以及Unet模型形成对称式的编码器-解码器结构,将解码器与对应的编码器直接相连,在保证分割速度的同时获得了较高的分割准确率。自此,多个

基于LinkNet的网络被相继提出。Zhou等人^[17]提出了D-LinkNet,它是一种被用于高分辨率卫星道路图像提取的模型,通过结合LinkNet、预训练编码器以及扩张卷积进行道路提取,在一定程度上解决了道路连通性问题。文献[18]又在D-LinkNet的基础上进行改进,使用DenseNet^[19]代替ResNet,减少了网络的参数并且提高了道路提取的准确度。

受以上分割模型的启发,本文结合编码器-解码器结构以及注意力机制的优势,提出了一种基于编码器-解码器结构的道路图像语义分割网络(A-LinkNet)。由于LinkNet编码器与解码器直接相连,导致大量背景特征被引入,对有效特征的提取造成了干扰,因此A-LinkNet在LinkNet的每个编码块后接入注意力模块,增加对有效特征的提取,抑制对无效特征的响应,使得分割过程中更加注意目标特征的提取,避免在解码阶段引入过多的背景特征。同时,由于LinkNet不具有提取多尺度上下文特征的能力,所以对于细小目标以及边界的分割准确度不高,我们引入了空洞空间金字塔池化模块(ASPP),利用多个并联的具有不同膨胀率的卷积操作来

捕获多尺度的上下文信息,融合图像的多尺度特征,进一步提高模型分割的准确率。

2 A-LinkNet 模型

2.1 A-LinkNet 网络结构

本文所提模型主要分为3部分:编码区、中心区以及解码区,其结构图如图1所示。其中,编码区由残差块以及注意力模块串联组合而成。对于输入图像,首先通过一个初始卷积模块对输入图片进行降维,该模块由一个卷积核大小为7、步长为2的卷积层以及一个卷积核大小为3,步长为2的最大池化层组成。然后连接4个残差块,残差块结构如图2所示。与LinkNet编码器不同的是,我们在前3个残差块后分别连接一个注意力模块,使用位置注意力机制对特征图中的每个像素点进行建模,增加目标特征的响应能力,同时使用通道注意力机制对特征图中的不同通道特征进行建模,对不同的通道赋予不同的权重,增加对有利通道的响应,然后将得到的通道特征与位置特征进行特征融合,得到注意力模块的输出。

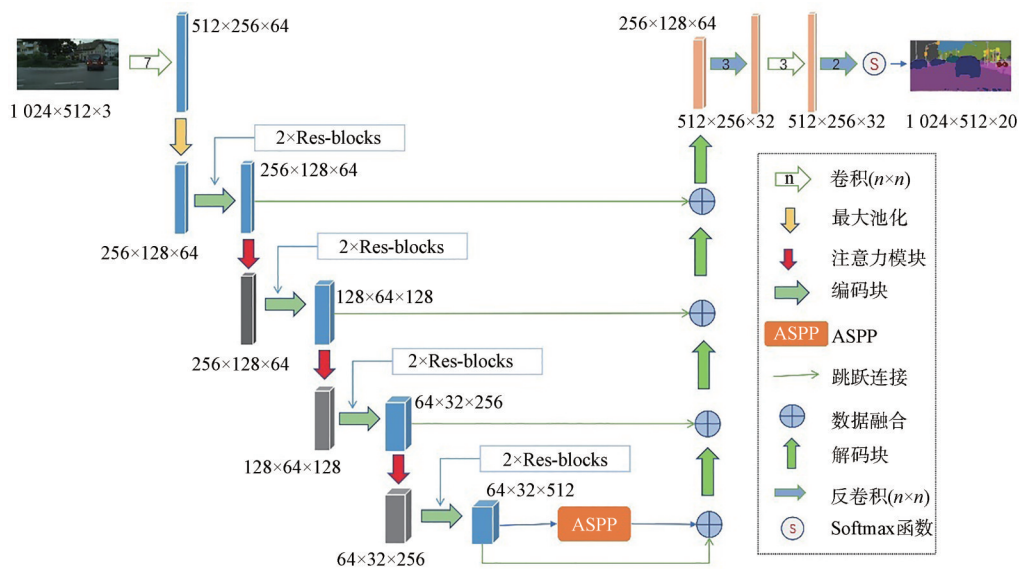


图1 A-LinkNet结构

Fig. 1 Structure of A-LinkNet

中心区在LinkNet的基础上引入了ASPP模块,该模型最早是在DeepLabv2网络^[10]中被提出,因为其在处理多尺度特征提取中表现优异,

而后在图像语义分割中得到广泛应用,其结构如图3所示。ASPP模块由一个卷积核大小为1的卷积层、3个卷积核大小为3、膨胀率为6、12、

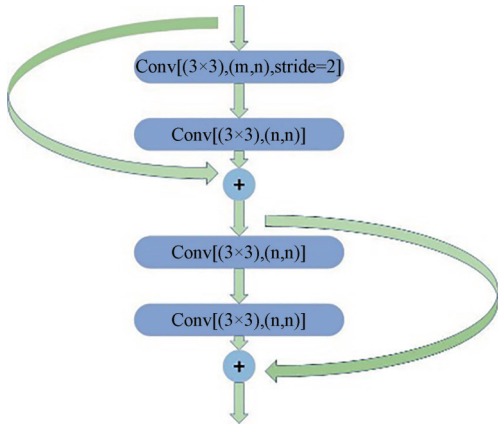


图2 残差块结构

Fig. 2 Structure of residual blocks

18的空洞卷积层以及一个全局平均池化层并联而成,之后将得到的特征图在通道维度上进行特征融合,最后通过一个 1×1 卷积操作来降低特征通道数。通过使用空洞卷积,可以在不影响图像分辨率的前提下增大图像的感受野,更好地捕获上下文信息。设置不同的空洞率可以得到不同尺度的特征图,更好地提取多尺度信息。最后增加全局平均池化,将图像级特征融进ASPP中,提供位置信息。

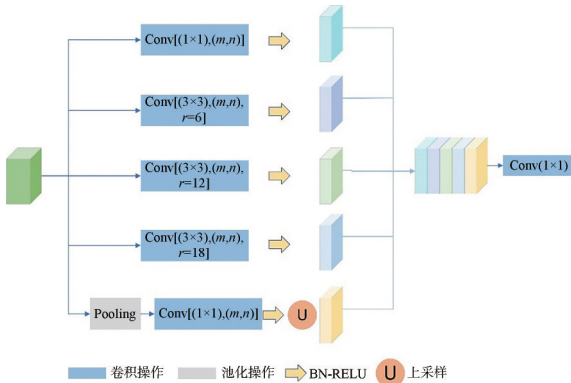


图3 空洞空间金字塔池化模块

Fig. 3 Atrous spatial pyramid pooling module

解码区的结构和LinkNet相同,由4个解码块串联组成,每个解码块的结构如图4所示。该结构使用两个卷积核为1的卷积提高网络的计算效率,使用一个 3×3 的反卷积进行上采样,逐步恢复特征图大小。此外,直接将解码器与编码器相连接,与编码阶段的特征信息进行融合,可以更好地保留空间信息,最终得到与输入图像大小相同的分割图像。

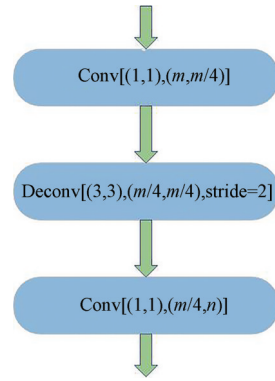


图4 解码块结构

Fig. 4 Structure of decoder blocks

2.2 注意力模块

在道路图像中,由于物体种类比较多,类间差异比较大,图像的背景比较复杂,给分割带来了一定的难度。在深度卷积网络中,较浅层中的特征图分辨率高,可以学到轮廓、边界等低级的空间特征。随着网络的加深,可以得到语义更加丰富的高层特征,但是由于低层特征中含有大量背景特征信息,会给深层特征的学习带来干扰。因此,受到SENet^[12]以及DANet^[14]的启发,我们提出了基于位置和通道的注意力模块,可以从空间相对位置以及不同通道间的依赖中捕获全局上下文信息,对特征进行加权,减少背景特征的权重,增加目标特征的权重,抑制无用信息的干扰。

注意力模块由位置注意力机制和通道注意力机制组成,结构如图5所示。其中,上半部分表示位置注意力模块,下半部分表示通道注意力模块,最后将两个模块的结果进行特征融合。具体来说,位置注意力机制通过对任意两个像素点进行建模来表示两个位置的相关性,这样具有相似特征的两个像素可以相互提升,得到较高的权重。通道注意力机制通过网络学习得到不同的通道权重,进而增加有效通道的响应。

对于输入特征 A ,经过卷积核大小为1的卷积层进行降维操作,得到与输入特征 A 尺寸相同的特征 B 、 C ;将特征 B 进行转置后与特征 C 进行矩阵相乘,进而建模特征图中任意两个像素点之间的相似性,然后通过Softmax函数计算得到相应的位置注意力图 P ,如式(1)所示:

$$P_{ji} = \frac{\exp(B_i \times C_j)}{\sum_{i=1}^N \exp(B_i \times C_j)}, \quad (1)$$

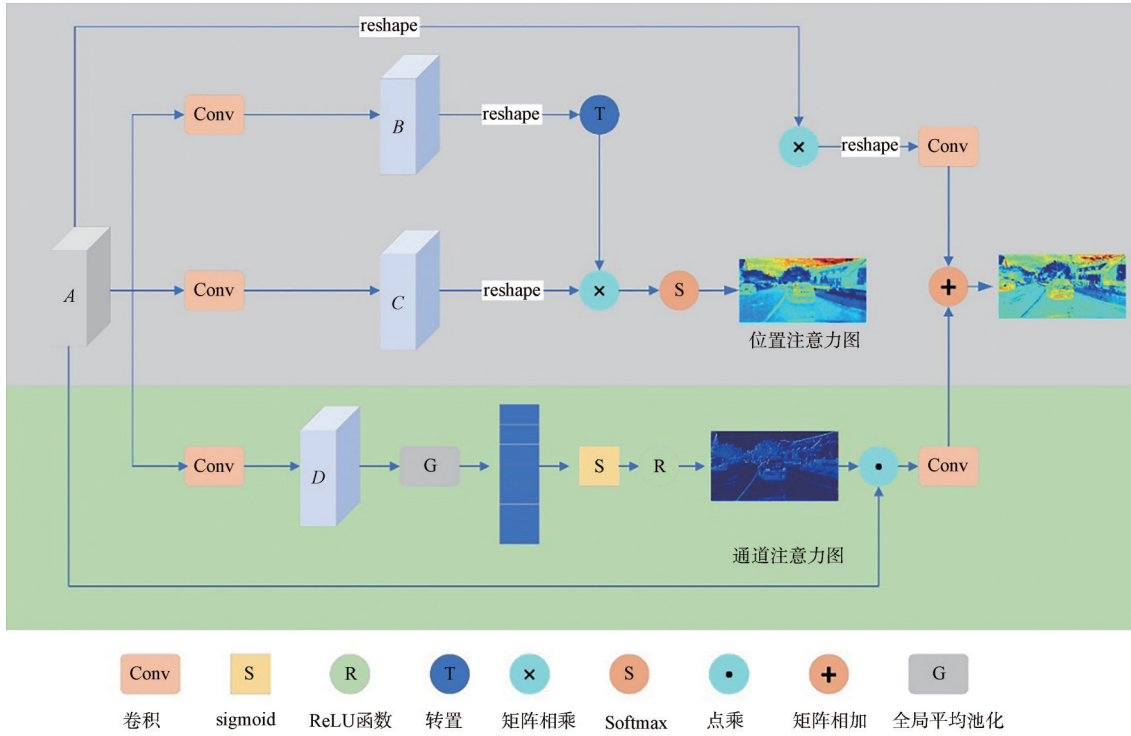


图5 本文提出的注意力模块

Fig. 5 Attention module proposed in this paper

其中, $P_{ji} \in R^{N \times N}$ 表示特征图中第 j 个像素与第 i 个点之间的特征相似性, $B, C \in R^{N \times T}$, $N = H \times W$, H, W 分别表示特征图的高和宽, T 表示通道数。最后将注意力图 P 与输入特征 A 进行相乘得到处理后的位置注意力特征 P_i , 如式(2)所示:

$$P_i = \alpha \sum_{j=1}^N \sum_{i=1}^N P_{ji} \times A_j, \quad (2)$$

其中, α 表示尺度系数, 初始化为 0, 逐渐学习得到更大的权重。

同样, 输入特征 A 经过卷积核大小为 1 的卷积层进行了降维操作, 得到新的特征 D , 然后对特征 D 进行全局平均池化, 得到可以表征各通道的全局信息; 然后连接两个全连接层来建模通道间的非线性相关性, 第一个全连接层后我们使用 ReLU 函数来提取更多的非线性特征, 第二个全连接层后我们使用 Sigmoid 函数获得通道注意力图 H , 具体计算方法如式(3)所示:

$$H_{ij} = \frac{1}{1 + \exp(-\gamma D_{ij}^1)}, \quad (3)$$

其中, D_{ij}^1 表示特征 D 经过全局平均池化后的特征图, γ 表示 ReLU 函数。最后将通道注意力图与输入特征 A 进行相乘得到处理后的位置注意力

特征 H_i , 如式(4)所示:

$$H_i = H_{ij} \cdot A. \quad (4)$$

最终, 使用 1×1 卷积将得到的位置注意力特征与通道注意力特征还原至输入维度, 最后进行特征融合, 得到注意力输出特征。

3 实验结果与分析

3.1 数据集及实验设置

本文实验采用的是 Cityscapes 数据集^[20], 该数据集主要包含来自 50 个不同大型城市的街道场景, 其中包含 5 000 张带有精细标注的图像以及 20 000 张带有粗糙标注的图像。我们使用精细标注的图像进行实验, 其中 2 975 张用作训练集, 500 张用作验证集, 剩余 1 525 张为测试集。数据集共包含 30 个类别, 其中的 19 个类别被选定用于训练和评估。

本文实验均在内存为 48 G 的 NVIDIA Quadro RTX 8 000 GPU 上进行训练, 采用 pytorch 1.2.0 框架。模型优化器选用 Adam 算法^[21], 初始学习率设置为 $5e-4$, 并使用交叉熵损失函数优化所有模型。输入图片大小为 $1\,024 \times 512$, 批处理大

小为 16, epoch 设置为 200。

本文采用平均交并比 (Mean Intersection over Union, MIoU) 作为评估指标, 表示分割结果与真实值的重合度, 是目前语义分割领域最常用的评价指标之一。其计算公式如下:

$$MIoU = \frac{1}{N} \sum_{i=1}^N \frac{TP}{FN + FP + TP}, \quad (5)$$

其中, N 表示类别总数, TP 、 FN 、 FP 、 TN 分别表示真正例、假反例、假正例、真反例。

3.2 对比实验

3.2.1 ASPP 模块实验

为了验证 ASPP 模块不同膨胀率组合的效果, 本文设计了多组 ASPP 结构进行实验, 结果如表 1 所示。

在相同的计算条件下, ASPP 模块可以提供更大的感受野, 采用不同膨胀率的空洞卷积可以捕获不同尺度的信息。随着感受野的增大, 模型的性能逐渐提高。在表 1 中, 可以看到 ASPP 模块的增加对模型的性能有一定的提高, 与 LinkNet 原始模型比较, MIoU 分别提高了 0.86%、1.69%、

表 1 不同膨胀率组合对比实验

Tab. 1 Comparative experiment of different dilation rate combinations

模型	骨干网络	ASPP 结构	感受野	MIoU/%
LinkNet	Resnet18	—	—	59.77
		[1, 3, 6]	13	60.63
		[3, 6, 12]	25	61.46
		[6, 12, 18]	37	61.83
		[12, 18, 24]	49	60.95

注: 黑体数字代表 MIoU 最优值

2.06% 以及 1.18%。但是如果膨胀率过大, 会导致卷积核跑到 padding 区域, 产生无意义的权重, 导致性能下降。本文采用 [12, 18, 24] 组合的 ASPP 模块与 [6, 12, 18] 组合相比, MIoU 并未有效提高。因此, 在本文后续实验中, 选择膨胀率为 [6, 12, 18] 的空洞卷积来提高模型的准确度。

3.2.2 消融实验

为了验证注意力机制以及 ASPP 模块的有效性, 本文在 Cityscapes 数据集上进行了消融对比实验, 具体实验设置以及结果如表 2 所示。

表 2 在 Cityscapes 数据集上的消融实验

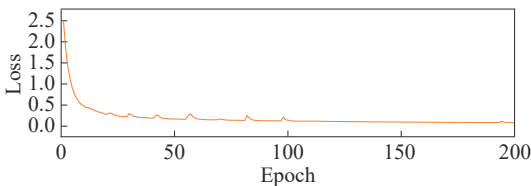
Tab. 2 Ablation experiment on the Cityscapes dataset

模型	骨干网络	注意力机制	ASPP 模块	GFLOPs	参数量/M	MIoU/%
LinkNet	Resnet18	×	×	20.66	11.54	59.77
		✓	×	21.17	11.64	62.04
		×	✓	22.78	15.80	61.83
		✓	✓	23.28	15.91	64.78

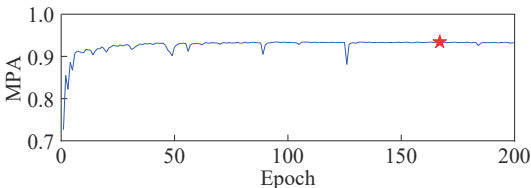
注: 黑体数字代表 MIoU 最优值

从表 2 可以看出, 使用 LinkNet 原始模型得到的 MIoU 为 59.77%, 与原始模型相比较, 仅加入注意力机制的模型分割结果为 62.04%, 增加了 2.27%; 仅加入 ASPP 模块, 虽然参数量和计算量有所增加, 但是 MIoU 提高了 2.06%; 同时加入注意力机制和 ASPP 模块, 最终的分割结果为 64.78%。可以看出, 本文方法有效提高了道路语义分割准确率。

实验过程中的 Loss 曲线如图 6(a) 所示, 可以看出, 在经过约 100 个 epoch 之后, 该模型可以很好地实现收敛。同时, 我们统计了训练过程中的平均像素准确度 (Mean Pixel Accuracy, MPA), 如图 6(b) 所示。随着 epoch 的增加, MPA 逐渐增大, 在 167 个 epoch 时, 模型的平均像素准确度达



(a) Loss 曲线
(a) Loss curve



(b) MPA 曲线
(b) MPA curve

图 6 训练曲线图

Fig. 6 Curves of training

到最大,为93.42%。

3.3 结果与分析

为了验证所提方法的实验效果,与其他方法在相同的软硬件环境下进行实验,采用相同的实验参数在Cityscapes验证集上进行对比。对比方法选择了FCN^[3]、SegNet^[5]与ENet^[6]。FCN是图像语义分割领域的经典模型,SegNet是编码器-解码器结构的代表,ENet则是轻量级模型的代表。实时性对比结果如表3所示,MIoU对比结果如表4所示。

表3 实时性能对比实验

Tab. 3 Real-time performance comparison

模型	GFLOPs	参数量/M	FPS	MIoU/%
FCN ^[3]	136.2	—	1.43	58.61
ENet ^[6]	3.8	0.4	40.88	61.79
SegNet ^[5]	286	29.5	12.67	62.71
Ours	23.28	15.91	26.29	64.78

注:黑体数字代表MIoU最优值

表4 本文方法与其他网络在Cityscapes验证集上MIoU值对比结果

Tab. 4 Comparative of MIoU values between the method in this paper and others on the Cityscapes validation set

类别	FCN ^[3]	ENet ^[6]	SegNet ^[5]	Ours
Road	95.62	95.47	95.61	96.54
Sidewalk	75.77	77.21	76.09	75.23
Building	87.54	88.32	91.49	88.85
Wall	43.31	44.17	43.93	44.69
Fence	39.58	42.01	40.37	47.55
Pole	41.99	41.58	44.41	42.73
Traffic Light	31.05	33.47	35.04	46.57
Traffic Sign	48.14	56.55	52.09	62.67
Vegetation	88.48	84.89	85.88	89.09
Terrain	51.02	52.08	51.69	50.41
Sky	87.57	89.93	91.09	91.97
Person	56.95	64.54	66.80	67.07
Rider	43.39	41.19	40.26	42.53
Car	87.53	91.83	92.91	93.71
Truck	41.84	56.87	66.95	56.69
Bus	59.57	63.36	66.12	75.93
Train	40.61	45.06	50.35	54.02
Motorcycle	32.27	35.21	36.79	39.68
Bicycle	64.47	66.22	63.5	65.84
MIoU	58.61	61.79	62.71	64.78

注:黑体数字代表MIoU最优值

从表3可以看出,FCN网络的分割速度最差;SegNet的分割速度有所提高,但其参数量庞大;ENet作为轻量级网络的代表,具有良好的实时性。本文所提算法虽然实时性不及ENet,但是对比FCN、SegNet,分割速度大幅提高,并且MIoU高于ENet。

从表4可以看出,本文所提方法MIoU结果为64.78%,与其他方法相比,模型性能有了一定提高,比经典的FCN提高了6.17%。另外,我们统计了每一类的IoU,所提方法虽然没有达到每一类都是最优,但是对于多数类别表现最佳。由于注意力机制以及ASPP模块的增加捕获了更加丰富的上下文信息,对于信号灯、信号标志等细小物体的分割,MIoU获得大幅提高。

为了更加直观地对比实验结果,我们在图7中对部分分割结果图像进行了展示。在第1行中,FCN网络对于公共汽车的分割较为粗糙,边缘预测不够突出;SegNet网络对于公共汽车边界的分割有所提高,但是在车顶上方出现了干扰信息,有几处出现识别错误;而本文提出的方法对于公共汽车的轮廓预测更加清晰,对于车身的识别也比较完整。在第2行中,对比FCN网络以及SegNet网络对于图片左侧人行横道的识别,本文所提方法的效果更加完整。在第3行中,对于交通信号灯以及栏杆等细小物体的分割,FCN表现较差,分割边界不清晰,不能完整识别杆状物;SegNet网络分割效果较好,但仍有些细节处理不清晰。在第4行中,对于行人的分割,FCN网络以及SegNet网络分割效果粗糙,边界不明显,识别不完整;本文所提方法对于行人的识别比较完整,边界处较为清晰。在第5行中,对于草等水平植被的识别,FCN网络的分割结果最差,出现多处识别错误,将水平植被误识为树木等垂直植被;SegNet网络将水平植被误识为人行道;而本文所提方法分割相对完整。在第6行中,FCN网络以及SegNet网络对于建筑物的分割效果较差,出现大面积识别错误;而本文所提方法分割结果较为准确。另外,对于交通标志的识别,本文所提方法识别形状更为准确。在第7行中,对于自行车以及骑行者的识别,本文所提方法较为形象,分割结果更为完整。相比较而言,本文所提方法能更加准确地分割细小物体,边界更加清

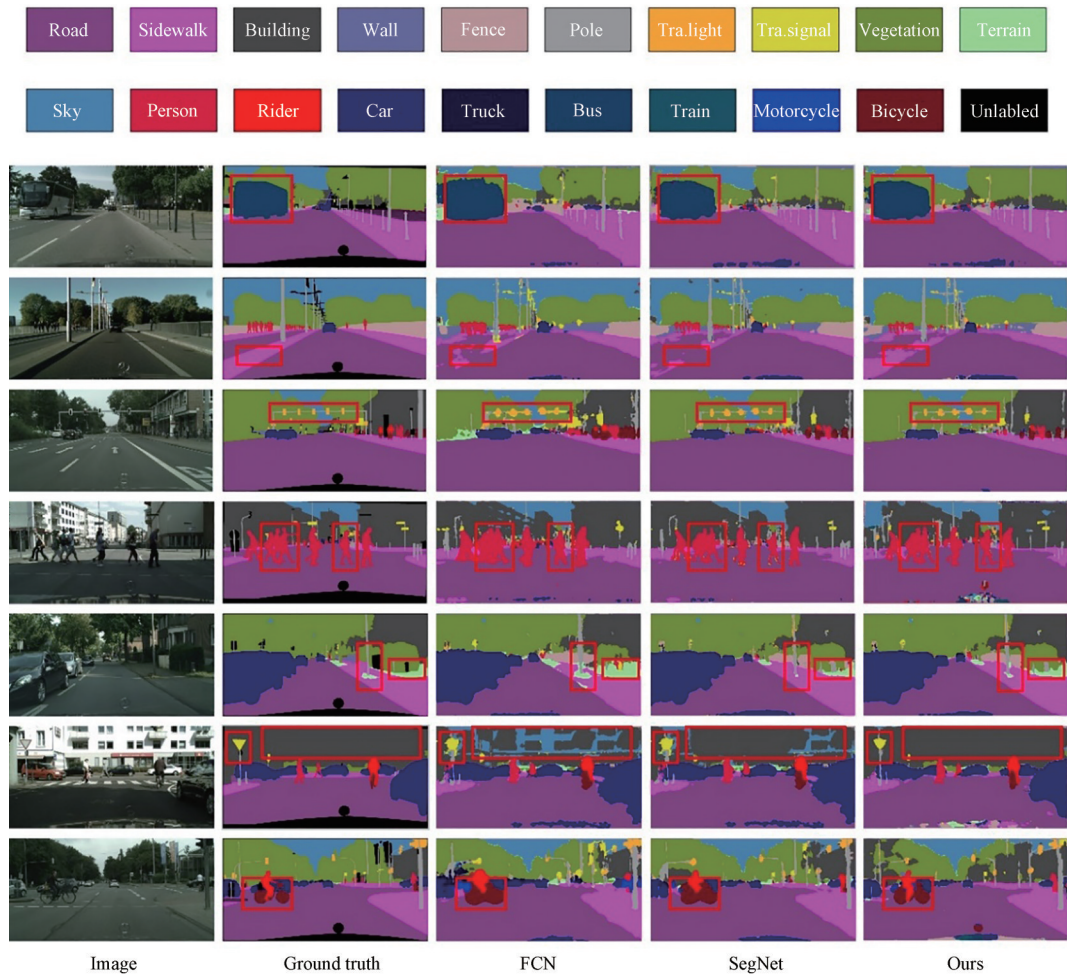


图 7 不同模型在 Cityscapes 验证集上的分割结果
Fig. 7 Segmentation results of different models on the Cityscapes validation set

晰,形状更加完整。总的来说,本文所提方法在 Cityscapes 验证集上可以预测出更优良的分割结果。

4 结 论

本文在 LinkNet 分割网络的基础上进行改进,可以较好地弥补 LinkNet 在道路图像分割中的不足。具体来说,在编码器-解码器的结构上加入注意力机制,分别对位置以及通道维度的特征进行建模,挖掘道路图像在空间维度以及通道

维度的上下文依赖信息,提高模型对有效特征的提取能力,避免无效信息的干扰。另外,在中心区域引入空洞空间金字塔池化模块,通过使用不同膨胀率的空洞卷积融合道路图像的多尺度特征,提取更加丰富的上下文信息,进一步提高模型分割的准确度。本文模型在 Cityscapes 数据集上分割结果表现优异,与其他网络相比较,MIoU 达到了 64.78%,且分割结果中目标特征更加完整,边界更加清晰。在后续的研究中,会考虑优化损失函数,探索更高效的语义分割网络,保证模型在分割速度与分割精度上有更好的平衡。

参 考 文 献 :

[1] 王曦,于鸣,任洪斌. UNET 与 FPN 相结合的遥感图像语义分割[J]. 液晶与显示, 2021, 36(3): 475-483.
WANG X, YU M, REN H E. Remote sensing image semantic segmentation combining UNET and FPN [J]. Chinese Journal of Liquid Crystals and Displays, 2021, 36(3): 475-483. (in Chinese)

- [2] 毛威. 基于深度学习的城市街景图像语义分割研究[D]. 成都:西南交通大学,2019.
MAO W. Semantic segmentation research of urban scenes based on deep learning [D]. Chengdu: Southwest Jiaotong University, 2019. (in Chinese)
- [3] LONG J, SHELHAMER E, DARREL T. Fully convolutional networks for semantic segmentation [C]//*Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA: IEEE, 2015: 3431-3440.
- [4] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]//*Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich: Springer, 2015: 234-241.
- [5] 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.
- [6] PASZKE A, CHAURASIA A, KIM S, *et al.* ENet: a deep neural network architecture for real-time semantic segmentation [J]. *arXiv*, 2016:1606.02147.
- [7] WANG Y, ZHOU Q, LIU J, *et al.* Lednet: a lightweight encoder-decoder network for real-time semantic segmentation [C]//*Proceedings of 2019 IEEE International Conference on Image Processing (ICIP)*. Taipei, China: IEEE, 2019: 1860-1864.
- [8] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions [J]. *arXiv*, 2015:1511.07122.
- [9] WANG P Q, CHEN P F, YUAN Y, *et al.* Understanding convolution for semantic segmentation [C]//*Proceedings of 2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Lake Tahoe: IEEE, 2018: 1451-1460.
- [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] CHEN L C, PAPANDREOU G, SCHROFF F, *et al.* Rethinking atrous convolution for semantic image segmentation [J]. *arXiv preprint arXiv:1706.05587*, 2017.
- [12] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018: 7132-7141.
- [13] WOO S, PARK J, LEE J Y, *et al.* CBAM: convolutional block attention module [C]//*Proceedings of the 15th European Conference on Computer Vision*. Munich: Springer, 2018: 3-19.
- [14] FU J, LIU J, TIAN H J, *et al.* Dual attention network for scene segmentation [C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019: 3141-3149.
- [15] CHAURASIA A, CULURCIELLO E. LinkNet: exploiting encoder representations for efficient semantic segmentation [C]//*Proceedings of 2017 IEEE Visual Communications and Image Processing (VCIP)*. St. Petersburg: IEEE, 2017: 1-4.
- [16] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 770-778.
- [17] ZHOU L C, ZHANG C, WU M. D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction [C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Salt Lake City: IEEE, 2018: 182-186.
- [18] PENG B, LI Y X, FAN K L, *et al.* New network based on D-Linknet and densenet for high resolution satellite imagery road extraction [C]//*Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium*. Yokohama: IEEE, 2019: 3939-3942.
- [19] HUANG G, LIU Z, MAATEN L VAN DER, *et al.* Densely connected convolutional networks [C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017: 2261-2269.
- [20] CORDTS M, OMRAN M, RAMOS S, *et al.* The cityscapes dataset for semantic urban scene understanding [C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE,

2016: 3213-3223.

[21] KINGMA D P, BA J. Adam: a method for stochastic optimization [J]. *arXiv*, 2014:1412. 6980.

作者简介:



杜敏敏(1996—),女,山东泰安人,硕士研究生,2019年于聊城大学获得学士学位,主要从事图像分割、模式识别方面的研究。E-mail: duminmin0131@163.com



司马海峰(1982—),男,河南濮阳人,博士,副教授,2015年于北京理工大学获得博士学位,主要从事图像处理、机器学习与人工智能方面的研究。E-mail: smhf@hpu.edu.cn