DOI:10.19651/j.cnki.emt.2211361

基于改进 U-Net 的街景图像语义分割方法*

徐晓龙 俞晓春 何晓佳 张 卓 万至达
(河海大学物联网工程学院 常州 213022)

摘 要: 为提升多尺度目标的分割效果,增强特征提取能力,提出了一种基于双重注意力机制的改进 U-Net 街景图像语义分割方法。在 U-Net 编码阶段的第 5 个卷积块之后,添加特征金字塔注意力模块,提取多尺度特征,融合上下文信息,增强目标语义特征。在解码阶段不再采用 U-Net 的特征拼接方法,而是设计了一个空间域-通道域联合注意力模块,接收来自跳跃连接的低层特征图和来自前一个注意力模块的高层特征图。在 Cityscapes 数据集上的实验结果表明,引入的注意力模块可有效提升街景图像分割精度,与 PSPNet、FCN 等方法相比,分割性能指标 mIoU 提升了 2.0%~9.6%。

关键词: 语义分割;注意力机制;卷积网络;多尺度特征;上下文信息
中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Semantic segmentation method of street view image based on improved U-Net

Xu Xiaolong Yu Xiaochun He Xiaojia Zhang Zhuo Wan Zhida
(College of Internet of Things Engineering, Hohai University, Changzhou 213022, China)

Abstract: An improved U-Net street image semantic segmentation method based on a dual attention mechanism is proposed to improve the segmentation effect of multi-scale targets and enhance the feature extraction ability. After the fifth convolutional block in the U-Net encoding stage, the feature pyramid attention module is added to extract multi-scale features, fuse contextual information, and enhance the target semantic features. Instead of using the feature stitching method of U-Net in the decoding stage, a joint spatial domain-channel domain attention module is designed to receive the low-level feature maps from the jump connection and the high-level feature maps from the previous attention module. Experimental results on the Cityscapes dataset show that the introduced attention module can effectively improve the street view image segmentation accuracy, and the segmentation performance metric mIoU improves by 2.0~9.6 percentage points compared with methods such as PSPNet and FCN.

Keywords: semantic segmentation;attention mechanism;convolutional network;multi-scale features;contextual information

0 引 言

语义分割是一项高层次、精细化的任务,图像中的每一个像素都需要标记语义信息,来实现像素级别的预测。目前,语义分割已广泛应用于自动驾驶、医疗影像分析、地理信息系统等领域[1]。对街景图像进行语义分割,可以把道路、草地、车辆、行人等完全分类,使车载智能系统自动检测前方路况信息,感知车辆所处环境,判断交通状况,为自动驾驶等人工智能应用场景提供技术基础[2]。传统的分割方法根据实现的原理不同,可以分为基于阈值、边缘检测、区域、图论和聚类法的图像分割算法等,但是由于背景复杂,成像质量较低,目标特征模糊等干扰,始终较难实现目标和背景区域间准确高效的划分。尤其在传统的语义分割方法中,一个分类器只能匹配特定的类别设计,在面临较多类别的情况时,难免会造成增加计算复杂度、提高训练难度及降低分类精度等问题。

自从 AlexNet[3] 在机器视觉领域引入了深度学习并获得 ImageNet 挑战赛冠军之后,国内外专家学者逐渐开始研究深度学习技术,并取得了丰硕的研究成果,由于深度卷积神经网络在语义分割任务上的卓越表现,使其成为语义分割的主流[4-6]。

监督深度学习网络凭借自身错误率极低的优势,已经

收稿日期:2022-09-12
* 基金项目:国家重点研发计划(2018YFC0407101)、国家自然科学基金(61671202)项目资助

成为了深度学习最为常用的一类网络。Long 等[5]提出了全卷积网络(fully convolutional networks, FCN),用卷积层替代全连接层,使得整个网络全部由卷积层组成,避免了全连接层在计算时造成的冗余,也解决了图像尺度的变换问题,但 FCN 只使用单一层次的特征图,图像经过池化操作后,特征图分辨率降低,部分像素空间位置信息丢失,导致最终的语义分割效果表现不佳。

Badrinarayanan 等[7]提出了一种用于图像分割的深度卷积编码器-解码器架构(a deep convolutional encoder-decoder architecture for image segmentation, SegNet),将自然语言处理中的编码器和解码器引入图像分割领域,编码器通过卷积和下采样操作对特征进行提取,同时保存池化过程中提取的每个特征点位置信息;解码器根据高维压缩的图像信息,结合池化过程中特征点的位置进行上采样图像恢复,得到分割预测图像。Ronneberger 等[8]提出一种用于生物医学图像分割的卷积网络(convolutional networks for biomedical image segmentation, U-Net),编码器和解码器为 U 型对称结构,编码网络对图像进行特征提取,得到高层语义特征图;解码网络采用转置卷积逐步恢复图像信息,得到与输入图像分辨率一致的语义分割结果。对应层的编码器和解码器通过跳跃连接,实现特征融合,保证解码网络利用更多的信息恢复图像,提升了语义分割的准确率。Lin 等[9]提出的用于高分辨率语义分割的多路径细化网络(multi-path refinement networks,RefineNet)将残差模块引入 U-Net 网络中,将对应层的解码器和编码器通过残差块连接,进行多尺度融合得到语义分割的结果。吴量等[10]在 U-Net 的基础上引入残差结构,在解码阶段,增加一条并行的膨胀卷积特征提取模块,网络结合改进后的通道和空间注意力机制,使得网络在提取特征时更加专注某些特征层和空间区域,提升了医学图像的分割效果,但分割过程未能有效地考虑图像上下文信息,无法充分利用丰富的空间位置信息,导致局部特征和全局特征的利用率失衡。

Wu 等[11]提出一种轻量级上下文引导网络(light-weight context guided network, CGNet),设计了一个学习局部特征和周围环境联合特征的模块,使得网络可以在所有阶段捕获上下文信息,CGNet 比同时期的双边分割网络(bilateral segmentation network, BiSeNet)更快但精度低。Romera 等[12]提出一种用于实时语义分割的高效残差分解卷积网络(efficient residual factorized ConvNet, ERFNet),使用残差连接和分解卷积来提高语义分割的准确率,减少了参数、计算量,提高了运行速度的同时最终的运行结果没有发生太大的变换。

为提升多尺度目标的分割效果,增强模型的特征提取能力,保留图像细节特征,提出了一种基于双重注意力机制的改进 U-Net 街景图像语义分割方法。在编码阶段,利用特征金字塔注意力模块,提取多尺度特征,融合上下文信息,凸显目标的高层特征,增强目标语义特征;在解码阶段,利用空间域-通道域联合注意力模块改变了 U 型网络跳跃连接的融合机制,在模块内预先进行空间、语义特征的增强及转置卷积上采样的操作,之后融合低层特征图的空间信息和高层特征图的语义信息,互相补足,通过模块级联的方式得到精细的分割结果。消融实验结果表明,双重注意力模块可以有效提升模型的分割效果;对比实验结果表明,该方法有效提升了街景图像的分割精度,与金字塔场景分析网络(pyramid scene parsing network, PSPNet)、FCN 等方法相比,分割性能指标 mIoU 提升了 2.0%～9.6%。

1　算法设计

对于多尺度问题,通常通过增加不同尺度目标的图像,使得不同尺度的图像在数据集中均匀分布的方式,实现网络中的权重对不同尺度目标识别分割性能的提升。为提高现有方法在多尺度特征提取、上下文信息融合、特征增强等方面的性能及语义分割精度,在不额外增加太多计算量的条件下,引入特征金字塔注意力模块(feature pyramid attention module,FPAM)和空间域-通道域联合注意力模块(spatial channel joint attention module,SCAM),提出了基于改进 U-Net 的街景图像语义分割方法,模型结构如图 1 所示。

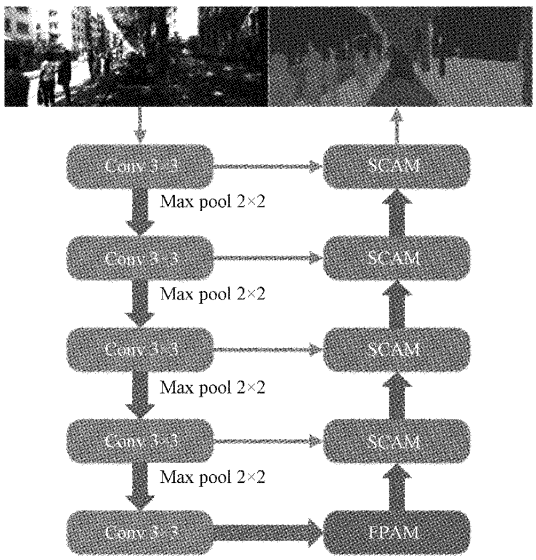


图 1　基于改进 U-Net 的街景图像语义分割模型结构

1.1　网络模型的改进

模型主体为 U-Net 编解码结构,一共 4 次编码、4 次解码。编码层越高,语义信息越丰富,编码层越低,空间信息越丰富。在编码阶段的第 5 个卷积块之后,添加了特征金字塔注意力模块 FPAM。该模块通过不同的卷积核提取目标的多尺度特征;融合了 1×1 卷积层和全局平均池化层提取的特征,以提供全局的上下文信息。在解码阶段,保持 U-Net 跳跃连接的融合机制,将多个空间域-通道域联合注意力模块 SCAM 串联,代替原有解码层,进行空间特征和

语义特征的增强和融合,实现图像的精细化恢复。

在图 1 中,输入图像维度为 1 024×512×3,每一个卷积块表示两次 3×3 的卷积操作,提取特征图,改变通道数,利用 padding 填充使卷积操作不改变图像分辨率。最大池化操作将特征图长宽减半,减小分辨率。利用 FPAM 增强第 5 个卷积块输出的特征图,输入到 SCAM。SCAM 同时处理来自卷积块的低层特征图和前一个模块的高层特征图,利用转置卷积恢复高层特征图的分辨率。最终输出图像分辨率与原图一致,通道数为分割类别数。

1.2 注意力机制的运用

在图像处理领域,注意力机制主要用来增强目标信息和抑制背景信息,通常以注意力模块的形式嵌入在深度网络中,对输入图像进行权重调整,关注目标区域特征,抑制背景区域干扰[13]。

1) 特征金字塔注意力模块

特征金字塔注意力模块可以融合不同分辨率的特征图,提取特征图的多尺度特征;融合高层特征图的上下文信息,提取像素级的高层语义特征,以凸显目标特征、提高分割性能。编码阶段添加的 FPAM 在高层网络叠加多个较小卷积核的卷积层,为平衡小卷积核特殊情况加入图片级特征,融合不同分辨率中的特征图,如此增强了特征图的多尺度特征,扩大了感受野范围。特征金字塔注意力模块的结构如图 2 所示。第 1 条分路的 3 条支路采用卷积核进行了两次卷积操作,降低特征图的分辨率,实现对多尺度特征的提取,通过双线性插值的方式恢复特征图分辨率到输入图像大小,最后逐像素相加融合 3 条支路的特征图,使提取出的高层像素级特征更接近真实特征图像的语义。第 2 条分路在通道方向上对图像特征进行聚合,之后和第 1 条分路输出的特征图相乘,可以加强分割目标位置的权重,得到增强目标特征的图像。第 3 条分路采用了全局平均池化层对图像进行下采样,通过卷积层进行特征融合,最后通过反卷积获得和输入图像大小一致的输出特征图,并和前两条分路融合的结果作逐像素相加得到最终的融合特征图。

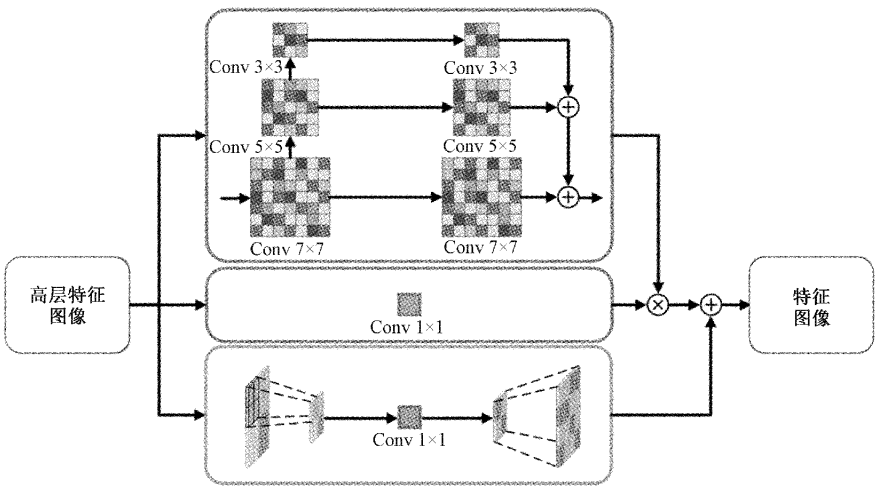


图 2 特征金字塔注意力模块结构

2) 空间域-通道域联合注意力模块

特征金字塔注意力模块通过对高层特征的上下文信息和多尺度特征的提取,提高了网络语义分割的性能,但面临图像对比度低、边缘模糊的情况时,较难实现对高层语义特征的精准提取,因此需要对高层特征图像进行高精度的像素级图像恢复。常用的分割网络如 SegNet、Refinenet、提拉米苏网络结构[14]等网络以 U 型网络为主干网络,叠加多种模块,重用特征信息,但是无法有效利用全局特征下目标像素之间的相互关系。空间域-通道域联合注意力模块保持 U-Net 跳跃连接的融合机制,将低层特征图和高层特征图作为输入,进行空间特征和语义特征的增强和融合,实现对高层特征的精细化图像恢复。空间域-通道域联合注意力模块的结构如图 3 所示。该模块首先使用转置卷积对高层特征图进行上采样操作,将图像恢复到和低层特征图一致的大小;然后将其作为输入,分别送入空间域注意力模块和通道域注意力模块,增强空间特征和语义特征,之后将二者的输出相加融合。低层特征图作为该模块的另一个输入,由于其空间信息更为精确,所以只对其进行空间域注意力模块的空间特征增强。将上述的两个输出相加融合,最后通过一个卷积层得到特征增强的特征图。在图像恢复的过程中,将多个空间域-通道域联合注意力模块串联,就可以得到精确的像素级图像分割结果。

(1) 空间域注意力模块

向空间域注意力模块引入自注意力机制,继而实现空间域中长范围内上下文信息的提取和融合,增强特征图像的空间信息,提高图像分割边缘的准确性。空间域注意力模块的结构如图 4 所示。其中 A 为输入的特征图像,$A\in R^{C\times H\times W}$,$C$ 为通道数,H 为图像的高度,W 为图像的宽度,通过 1×1 的卷积层,得到 3 个图像分别为 B、C、D。将 B、

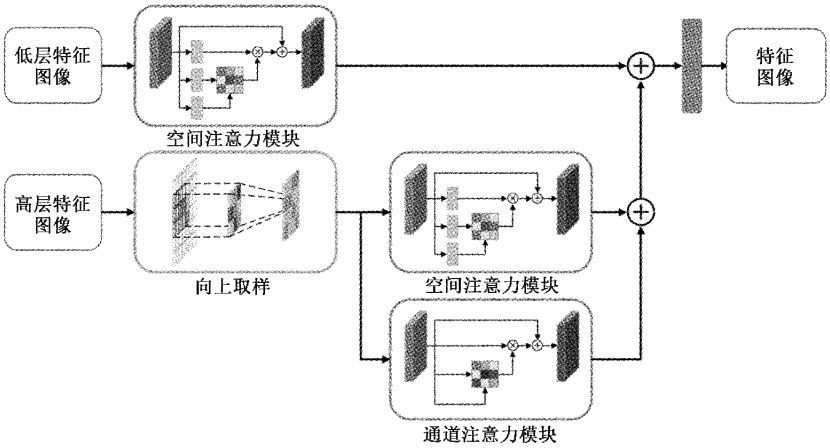


图 3　空间域-通道域联合注意力模块结构

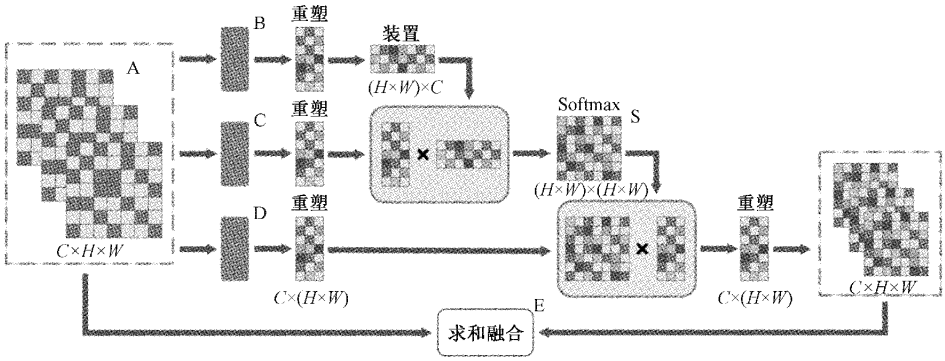


图 4　空间域注意力模块结构

C、D 进行 reshape 操作，使其维度变为$\{B,C,D\}\in R^{C\times N}$，$N=H\times W$。对 B 的转置和 C 做矩阵乘法，然后通过 Softmax 激活函数，得到空间域注意力的权重矩阵 $\boldsymbol{S}$，$\boldsymbol{S}\in R^{N\times N}$，$\boldsymbol{S}$ 的任一元素 S_{ij} 可由式(1)计算。

$$S_{ij}=\frac{e^{(B_i\cdot C_j)}}{\sum_{i=1}^{N}e^{(B_i\cdot C_j)}} \tag{1}$$

其中，S_{ij} 表示 B 图中的 i 位置对于 C 图中的 j 位置的依赖关系，B_i 和 C_i 为对应位置的像素值，两个位置的依赖关系只由像素固有特征决定。对 $\boldsymbol{S}$ 的转置和 D 做矩阵乘法，得到一个 $C\times N$ 的矩阵，然后进行 reshape 操作，得到 $C\times H\times W$ 维度的矩阵，即将权重矩阵作用于特征图得到特征增强后的图像。

最后将特征增强后的特征图和输入特征图 A 相融合，得到空间特征增强的特征图 $E\in \mathrm{R}^{C\times H\times W}$，$E$ 的任一元素 E_j 可由式(2)计算。

$$E_j=\alpha\sum_{i=1}^{N}(S_{ji}D_i)+A_j \tag{2}$$

其中，α 为学习参数，初始化为 0。空间域注意力模块输出特征图 E 的每一个像素点，都是输入特征图和空间位置依赖关系的加权融合，可以实现对特征图中目标空间信息的增强。

(2) 通道域注意力模块

通道域注意力模块先计算不同通道之间的相互依赖关系，然后加权融合，将自注意力机制应用在通道域中，可以增强通道域表征的语义信息。通道域注意力模块的结构如图 5 所示。通道域和空间域注意力模块结构基本相同，A 为输入的特征图像，$A\in R^{C\times H\times W}$。通道域注意力模块不做卷积层处理，因为卷积层会在通道之间进行特征融合，会改变通道之间的相互依赖关系，不利于通道域注意力模块对通道特征的提取。将 A 直接作为 B、C、D 特征图输入到后续的处理中，这样可以保持不同通道之间的相互关系。将 B、C、D 做 reshape 操作，转换为$\{B,C,D\}\in R^{C\times N}$，$N=H\times W$，对 B 的转置和 C 作矩阵乘法，然后通过 Softmax 激活函数，得到通道域注意力的权重矩阵 $\boldsymbol{X}$，$\boldsymbol{X}\in R^{C\times C}$，$\boldsymbol{X}$ 的任一元素 X_{ij} 可由式(3)计算。

$$X_{ij}=\frac{e^{(B_i\cdot C_j)}}{\sum_{i=1}^{C}e^{(B_i\cdot C_j)}} \tag{3}$$

其中，X_{ij} 表示 B 图中的第 i 个通道对于 C 图中的第 j 个通道的依赖关系，B_i 和 C_i 为对应位置的像素值。对 $\boldsymbol{X}$ 的转置和 D 进行矩阵乘法，得到一个 $C\times N$ 的矩阵，然后进行 reshape 操作，得到 $C\times H\times W$ 维度的矩阵，即将权重矩阵作用于特征图得到特征增强后的图像。最后将特征

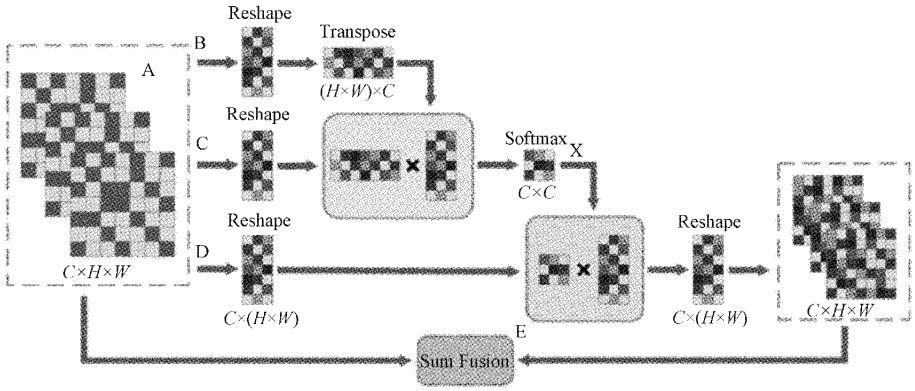


图 5 通道域注意力模块结构

增强后的特征图和输入的原始特征图 A 相融合,得到通道特征增强的特征图 $E \in \mathrm{R}^{C\times H\times W}$,$E$ 的任一元素 E_j 可由式(4)计算。

$$E_j = \beta \sum_{i=1}^{N} (X_{ji} D_i) + A_j \tag{4}$$

其中,β 为学习参数,初始化为 0。通道域注意力模块输出的特征图 E 的每一个通道特征,都是输入特征图和通道依赖关系的加权融合。

2 实验与分析

2.1 实验环境配置、数据集和评价指标

实验以 PyTorch 深度学习框架为基础,搭建基于改进 U-Net 的街景图像语义分割模型,并进行相关实验,实验环境如表 1 所示。

表 1 实验环境

项目	设置
系统	Ubuntu 16.04
CPU	Intel Xeon E5-2680,3.3 GHz
RAM	128 GB
GPU	Nvidia TITAN V
GPU 对应的驱动	Nvidia Driver 435.21,CUDA 10.1
相关软件	深度学习框架 PyTorch 1.3.1
	图像处理工具 OpenCV 4.1
	编程语言 Python 3.6

实验在公开数据集 Cityscapes 上进行,该数据集为多分类标注的全景分割数据集,标注较为准确,降低了标注误差对实验结果的影响。该数据集是来自德国、法国等多个城市的街道场景图片,包含道路、草地、汽车等 30 个语义分类,实验应用其中 19 个类别作为网络的分割任务。该数据集精细化标注的图片共有 5 000 张,其中训练集 2 975 张,验证集 500 张,测试集 1 525 张,图片的分辨率为 2 048×1 024。多分类分割结果用不同的灰度表示不同的种类。

图像分割的评价指标采用交并比 IoU (intersection over union) 和平均交并比 mIoU (mean intersection over union)。IoU 为图像分割预测区域和真实标注区域的交集与并集的比值,代表预测区域和真实区域的重叠率;mIoU 为全部类别预测区域和真实区域的交集与并集的比值的平均值。IoU 和 mIoU 的数值越大,代表目标分割精度越高,分割模型的性能越好,计算方法如式(5)、(6)所示。

$$IoU = \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{5}$$

$$mIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{6}$$

其中,k 表示类别的数目,p_{ii} 表示真实分类为 i 被分类为 i 的像素数目,p_{ij} 表示真实分类为 i 被分类为 j 的像素数目,p_{ji} 表示真实分类为 j 被分类为 i 的像素数目。

2.2 实验过程和结果分析

首先在 Cityscapes 数据集上,对提出的方法进行消融实验,验证特征金字塔注意力模块和空间域-通道域联合注意力模块对分割性能的提升;然后将本方法和主流的分割方法进行对比,验证本方法的分割性能处于较高水平。在进行训练的时候,各个实验均采用控制变量的原则,均使用表 2 所示的训练参数。

表 2 训练参数设置

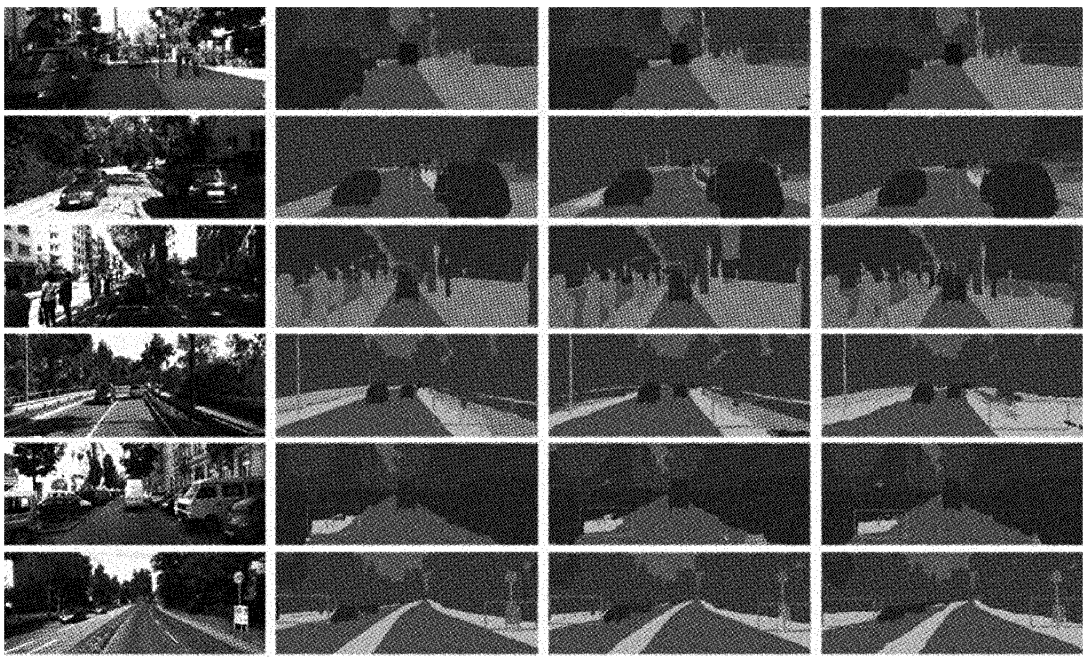
项目	设置
Batch Size	6
Epochs	100
初始学习率	0.05
学习率衰减	0.5/(20epoch)
损失函数	交叉熵损失函数
Optimizer 优化器	Adam 优化器

1) 消融实验

实验采用 Cityscapes 数据集,为了防止显卡内存容量

溢出，将输入图像的分辨率降低到 1 024×512，并且将 Batch size 设置为 6。消融实验是在改进网络上，分别删除特征金字塔注意力模块 FPAM 和空间域-通道域联合注意力模块 SCAM，并对其进行实验和结果分析。

消融实验效果如图 6 所示，从左到右依次为原始输入图像、本方法实验效果图、没有 FPAM 的实验效果图、没有 SCAM 的实验效果图。由图 6(b)和(c)的对比可以看出，FPAM 融合了多尺度特征和上下文信息，增强了高层特征的像素级语义信息，在图中的表现就是提高了对自行车、草地、标志牌的分类性能；由图 6(b)和(d)的对比可以看出，SCAM 增强了低层的空间特征和高层的语义特征，对图像进行了精细化恢复，在图中的表现就是提高了对栏杆、道路的边缘分割效果；由图 6(c)和(d)的对比可以看出，SCAM 和 FPAM 在不同的地方对分割效果都有一定程度的提升，但是总体分割效果的提升上 SCAM 更为明显。



(a) 原始图像　(b) 本方法　(c) 不加FPAM　(d) 不加SCAM

图 6　消融实验效果图

表 3 可以看出，引入注意力模块后，语义分割模型的分割精度评价指标 mIoU 得到了较大的提升，其中引入特征金字塔注意力模块 FPAM 使得分割精度提高了 1.4%，引入空间域-通道域联合注意力模块 SCAM 使得分割精度提高了 2.8%。从客观的分割性能评价指标可以看出，FPAM 和 SCAM 对分割效果都有一定的提升，SCAM 相较于 FPAM 提升效果更为明显，和主观分析一致。

表 3　消融实验评价指标

FPAM	SCAM	训练集	验证集	mIoU/%
		√		66.3
√		√		67.7
	√	√		69.1
√	√	√		70.3
√	√	√	√	73.5

2) 本文方法和主流分割方法的对比实验

本文和主流分割方法在 Cityscapes 数据集上的对比实验结果如表 4 所示。由对比实验可以看出：本文方法在

表 4　本方法与主流分割方法对比实验结果　%

类别	FCN	PSPNet	CGNet	ERFNet	Ours
Road	96.4	97.8	95.9	97.7	**97.8**
Sidewalk	71.7	80.7	73.9	**81.0**	**81.0**
Building	84.6	90.2	89.9	89.8	**91.0**
Wall	27.1	47.9	43.9	42.5	**51.3**
Fence	28.8	48.1	46.0	48.0	**50.6**
Pole	43.2	56.4	52.9	56.2	**58.3**
Traffic light	39.2	61.8	55.9	59.8	**63.0**
Traffic sign	34.4	67.0	63.8	65.3	**68.5**
Vegetation	89.3	92.0	91.7	91.4	**92.3**
Terrain	61.3	69.5	68.3	68.2	**71.3**
Sky	92.7	**94.3**	94.1	94.2	94.2
Person	65.7	**80.3**	76.7	76.8	80.1
Rider	46.4	59.2	54.2	57.1	**59.6**
Car	91.0	93.7	91.3	92.8	**93.8**
Truck	**57.0**	46.0	41.3	50.8	48.4
Bus	**70.3**	57.1	55.9	60.1	68.1
Train	**56.5**	35.0	32.8	51.8	42.1
Motorcycle	40.9	50.4	41.1	47.3	**52.4**
Bicycle	52.6	66.8	60.9	61.7	**67.8**
mIoU	60.5	68.1	64.8	68.0	**70.1**

14 个类别中的分割评价指标(IoU)较好,而且本网络并不依赖网络预训练、图像后处理操作,仅在训练集的训练优化下,mIoU 就可以达到 70.1%,相较于其他方法有 2.0%~9.6%的提升。

FCN 忽略了下采样时的信息损失,仅当目标尺寸偏大时有较好的分割效果,精度比本方法低 9.6%;ERFNet[15] 使用残差连接和分解卷积来提高分割精度,但仍比本方法低 2.1%;PSPNet 引入了金字塔池化来提取多尺度信息,CGNet[16] 融合了上下文信息,一定程度上提高了分割精度,但仍比本文方法低 2.0%、5.3%。本文提出的方法,通过引入双重注意力模块,不仅可以提取多尺度特征,融合上下文信息,而且可以增强并融合低、高层特征,取得了更好的分割效果。

3 结 论

为更好提取多尺度特征、融合上下文信息和增强特征,提高街景图像分割精度,提出了一种基于改进 U-Net 的街景图像语义分割方法。首先设计了特征金字塔注意力模块,提取多尺度特征,融合上下文信息,得到了精准的像素级高层语义特征图,凸显了目标的高层特征;然后设计了空间域-通道域联合注意力模块,增强并融合低层特征图的空间信息和高层特征图的语义信息,得到了精细化的图像恢复结果。实验结果表明,两种注意力模块,均可以不同程度地提高模型分割精度;引入双重注意力模块的改进 U-Net 方法在与其它方法的对比中处于较高水平,提高了街景图像语义分割精度。本文引入注意力模块提升街景图像语义分割效果的思路也可用于其他场景的语义分割中。

参考文献

[1] 胡 涛,李卫华,秦先祥. 图像语义分割方法综述[J]. 测控技术,2019,38(7): 8-12.

[2] 郭旭. 人工智能视角下的无人驾驶技术分析与展望[J]. 电子世界, 2017(20):66-67.

[3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 25: 1097-1105.

[4] 陈小波. 基于深度学习的语义分割算法研究[D]. 成都:电子科技大学,2020.

[5] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4): 640-651.

[6] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE,2017: 2881-2890.

[7] 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.

[8] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]. International Conference on Medical Image Computing and Computer-assisted Intervention. Berlin:Springer, 2015: 234-241.

[9] LIN G, MILAN A, SHEN C, et al. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York: IEEE,2017: 1925-1934.

[10] 吴量,付殿臣,程超. 基于 Unet 的多注意力脑肿瘤图像分割算法[J]. 计算机技术与发展,2021,31(12): 85-91.

[11] WU T, TANG S, ZHANG R, et al. CGNet: A light-weight context guided network for semantic segmentation [J]. IEEE Transactions on Image Processing, 2021, 30: 1169-1179.

[12] ROMERA E, ALVAREZ J M, BERGASA L M, et al. ERFNet: Efficient residual factorized convNet for real-time semantic segmentation [J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 10(1): 1-10.

[13] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York: IEEE,2018: 7132-7141.

[14] JEGOU S, DROZDZAL M, VAZQUEZ D, et al. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017: 11-19.

[15] YE D X, HAN R B. Image semantic segmentation method based on improved ERFNet model[J]. The Journal of Engineering,2022,2022(2):180-190.

[16] YAO Y, ZHANG Z, NI X, et al. CGNet: Detecting computer-generated images based on transfer learning with attention module[J]. Signal Processing: Image Communication,2022,105.

作者简介

徐晓龙,硕士,高级实验师,主要研究方向为机器视觉、嵌入式系统设计、实验室管理等。
E-mail: xuxl@hhuc.edu.cn

俞晓春,硕士研究生,主要研究方向为目标检测。
E-mail:1476498374@qq.com

何晓佳,硕士研究生,主要研究方向为目标检测。
E-mail:1101314519@qq.com

张卓,硕士,高级实验师,主要研究方向为机器视觉。
E-mail: zhangz@hhu.edu.cn

万至达,硕士研究生,主要研究方向为目标检测。
E-mail:2576498444@qq.com