ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 37,No. 2,Mar. 2022,pp. 426－436
DOI:10. 16337/j. 1004-9037. 2022. 02. 015

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于强化语义流场和多级特征融合的道路场景分割方法

项建弘[1,2]，刘　茁[1,2]，王霖郁[1,2]，钟　瑜[3]

(1. 哈尔滨工程大学信息与通信工程学院，哈尔滨 150001；2. 哈尔滨工程大学先进船舶通信与信息技术重点实验室，哈尔滨 150001；3. 中国西南电子技术研究所，成都 610036)

摘　要：自动驾驶是目前计算机视觉任务中难度较大的一类任务，而道路场景下的语义分割是自动驾驶的核心技术之一。本文针对经典分割网络中分辨率恢复方式简单，导致细节信息不完整、目标边缘模糊的问题，提出一种基于强化语义流场的上采样方法。该方法通过学习相邻特征图之间的语义流场，使生成图语义信息更细致，边界处更清晰。同时针对道路场景中目标尺度变化处理困难、小目标难以识别的问题，提出一种新的多级特征融合方法，充分融合深层语义信息与浅层细节信息，以适应不同尺度的目标。本文采用CamVid为数据集进行实验，并进行数据增强。实验表明本文提出的两种方法均显著提升了准确度，整体网络与PSPNet、Deeplabv3＋等多种模型相比，准确率更高，分割效果更接近真实值。

关键词：深度学习；语义分割；道路场景；强化语义流场；多级特征融合

中图分类号：TP391.41;TP18　　**文献标志码**：A

Enhance Semantic Flow Field and Multilevel Feature Fusion Network for Road Scene Segmentation

XIANG Jianhong[1,2], LIU Zhuo[1,2], WANG Linyu[1,2], ZHONG Yu[3]

(1. College of Information & Communication Engineering, Harbin Engineering University, Harbin 150001, China.; 2. Key Laboratory of Advanced Ship Communication and Information Technology, Harbin Engineering University, Harbin 150001, China; 3. Southwest Electronic Technology Research Institute of China, Chengdu 610036, China)

Abstract: Automatic driving is one of the most difficult tasks in computer vision, and semantic segmentation in road scenes is one of the core technologies of automatic driving. This paper proposes an upsampling method based on enhanced semantic flow field, which can make the semantic information of the generated graph more detailed and the boundary clearer by learning the semantic flow field between adjacent feature graphs. At the same time, aiming at the difficulty of processing target scale changes and identifying small targets in road scenes, a new multi-level feature fusion method is proposed, which fully integrates deep semantic information and shallow detail information to adapt to targets of different scales. In this paper, CamVid is taken as the data set and data enhancement is carried out. Experiments show that both methods proposed in this paper bring significant improvement in accuracy. Compared with PSPNet, Deeplabv3＋ and other models, the overall network has higher accuracy and the segmentation effect is closer to the real value.

Key words: deep learning; semantic segmentation; road scene; enhance semantic flow field; multilevel feature fusion

收稿日期：2021-08-26；**修订日期**：2021-11-05

引　言

语义分割是计算机视觉中一项具有挑战性的基本任务,其目的是通过为所有像素分配密集的标签,将场景图像解析并分割成与语义类别相关联的不同图像区域。该技术的研究可广泛应用于自动驾驶、无人机控制与应用、医学图像辅助分析、图像搜索和增强现实等领域。由于道路场景环境复杂多变,目标种类繁多,信息非常复杂,给语义分割任务带来了极大的挑战,分割模型需要准确地划分出行人、车辆、周围建筑物和道路,以便向机动车辆提供准确的信息。因此,性能良好的网络模型是决定当前道路场景能否正确划分和信息能否正确反馈的关键。本文将从设计多级特征融合网络和强化语义流场上采样两方面,提升网络模型的分割性能。

高层语义信息和低层空间信息本身具有差异性,采用直接相融的方式无法弥补低级特征的分辨率和高级特征的语义差距。本文提出多级特征融合的方式,向低级特征中引入更多的语义信息,向高级特征中引入更多的高分辨率信息,增强特征的丰富度,提升网络的识别能力。同时,由于道路场景中,同一类别的目标往往会有不同的尺度,分割任务通常需要获得多尺度的特征信息处理目标的尺度变化问题,因此在多级特征融合方法中引入并行的不同空洞率的空洞卷积,捕获不同尺度感受野下的语义信息,多个不同尺度感受野的叠加,可以编码多尺度的空间上下文信息,加强多级特征融合网络对不同尺度目标信息的处理能力。通过传统的上采样方法不断增大特征图的分辨率,直到恢复到原始输入图像大小的方式通常会使分割结果图的空间细节信息不够完整,目标边缘分割模糊,很难达到精细化分割的效果。本文提出基于强化语义流场的上采样模块代替传统的上采样方法,高效地学习2个相邻特征图之间的流场,有效地将高层特征融合到高分辨率的低层特征图中,生成高度语义化和细节信息丰富的高分辨率特征图。

1　相关工作

1.1　特征融合

所有层次的特征都有助于语义分割。编码部分输出的高级特征描述了输入图像的语义信息,有助于图像区域的类别识别,即解决像素“是什么”的问题。而低级特征描述了输入图像的位置信息,解决的是像素“在哪里”的问题,对于准确预测边界或细节至关重要。如何将二者有效地融合是一个值得探究的问题。经典分割工作全卷积网络(Full convolutional network, FCN)[1]和U-Net[2]都采用直接相加的策略融合高低层特征,物体检测中常用的特征金字塔网络[3]也采用了该策略。然而简单地将低级特征与高级特征进行相加会引入噪声,导致不同层级特征难以进行更好地融合,分割结果差强人意。为了改善特征融合的效果,很多工作网络提出了不同的优化特征融合策略。DFANet[4]编码阶段由3个风格一致的骨干网络构成,在解码阶段对同一骨干网络的高级特征与低级特征进行跨层融合,通过下一级网络来优化上一级网络的输出特征,同时对不同骨干网络获取的高级特征和低级特征分别进行融合;RefineNet[5]提出多路径细化网络,在每个上采样阶段引入了一个复杂的细化模块,利用低级视觉特征来细化高级语义特征,然后采用跨层连接,融合高级语义特征和低级视觉特征以产生高分辨率分割图;BiSeNet[6]双分支结构在融合空间分支和语义分支的特征时,考虑到高层语义信息与低层空间信息的差异,引入特征融合模块(Feature fusion module,FFM)来有效地跨层融合特征,对于通道合并连接后的特征,通过计算权重向量对其进行重新加权,实现特征选择和组合,进一步优化融合的特征;ExFuse[7]网络在特征融合时,将高于当前层级的全部特征进行通道域的融合,然后通过语义嵌入分支(Semantic embedding branch,SEB)与当前层级的特征进行逐像素相乘,弥补低级与高级特征图之间的语义与分辨率的差距,完善特征融合过程。

1.2 多尺度信息

语义分割任务通常需要多尺度特征信息来产生高质量的结果，模型如PSPNet[8]、DeepLabv3[9]和Deeplabv3+[10]，利用多尺度信息在多个场景分割基准上获得优异的结果。为了捕捉在多个尺度上的信息，PSPNet[8]应用了一个金字塔池化模块(Pyramid pooling module，PPM)，该模块包含多个不同比例的平均池化层以收集不同尺度的有效上下文信息。DeepLabv3[9]受图像金字塔启发，使用多个具有不同采样率的并行空洞卷积来捕获不同感受野的上下文信息。Deeplabv3+[10]在其基础上，设计了一个具有全局平均池的空洞空间金字塔池化模块(Atrous spatial Pyramid pooling，ASPP)以捕捉图像的全局背景。

1.3 分辨率增大

原始输入图像通过卷积神经网络进行语义分割时，最终输出的分割结果图像与原始输入图像的尺寸应当保持一致。特征提取过程中，多个阶段的池化和卷积操作通常会将输出的特征图分辨率大小减小为原来的$\frac{1}{32}$，导致损失许多精细的图像结构。如何更好地恢复高分辨率特征图成为提升语义分割精度的关键。线性插值是语义分割中提高特征图的分辨率的常用方法，即在原有图像像素的基础上，在像素点之间采用合适的插值算法插入新的元素，基于双线性插值的上采样方法计算量较小，方法易于实现，不需要对参数值进行训练。然而这种上采样方式很容易导致空间细节特征被忽略，甚至丢失一些目标，不同区域之间边界模糊，难以实现精细化分割。

基于反卷积的上采样操作，是另一种生成高分辨率特征图的策略。反卷积也叫转置卷积，通过卷积运算实现分辨率还原，由于卷积中含有权重参数，所以反卷积是一种可学习的方法，通过参数的学习实现输出图像尽可能与原图像相似。在Y-Net[11]、EANet[12]、LRUNet[13]和Chen[14]等提出的网络中都有采用反卷积来提高小特征图的分辨率。相比于插值的方法，反卷积操作对特征图的还原效果有所改善，能够一定程度恢复部分丢失的特征信息，但仍然无法将丢失的浅层视觉特征恢复到令人满意的效果。

2 方法实现

2.1 多级特征融合模块

高层语义信息和低层空间信息的不相容性是困扰语义分割领域的重要问题之一。一些网络在特征融合过程中会保留低层特征，然后将其与上采样后的高层特征融合，但这种直接融合的方式会损害网络的性能。本文提出了多级特征融合模块(Multilevel feature fusion module，MFFM)，其网络结构如图1所示。

MFFM的融合策略是先逐渐将所有更高级别的特征融合后再与当前特征进行融合，不断地向低层特征中引入语义信息，同时向高层特征中引入细节信息，实现低级特征与高级特征的对齐，使得融合后

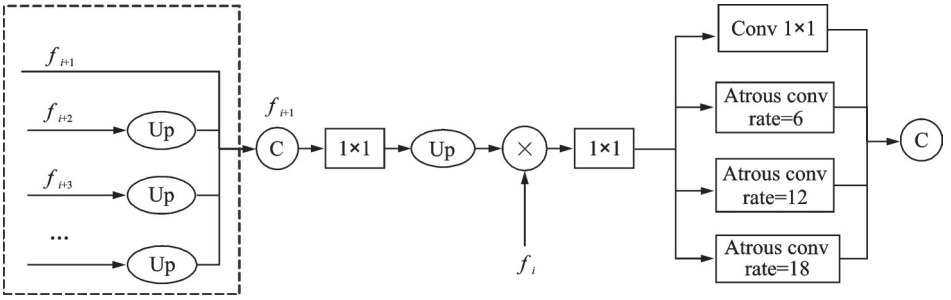

图1 MFFM网络结构图

Fig.1 MFFM module structure

的特征中尽可能多地包含两种信息。这种融合方法可以表述为

$$f_{i+1}=\mathrm{cat}(f_{i+1}+\mathrm{upsample}(f_{i+2})+\cdots+\mathrm{upsample}(f_{i+n})) \tag{1}$$

$$f_i=\mathrm{conv}(\mathrm{upsample}(f_{i+1}))*f_i \tag{2}$$

对于当前特征f_i,将所有高于f_{i+1}的特征上采样到f_{i+1}的分辨率大小,然后与f_{i+1}一起进行通道域的叠加,得到新的高层特征f_{i+1},接下来通过1个stride=1,padding=1的3×3卷积,压缩f_{i+1}的通道数使其与低级特征f_i相同,由于低级特征f_i的尺寸是f_{i+1}的2倍,所以调整通道数目后再将f_{i+1}进行2倍上采样,使这两种不同层级的信息可以进行逐像素相乘操作以达到对齐。

对融合后的特征通过1×1卷积进行通道压缩,来减少通道冗余信息和减轻计算负担,然后采用多路不同采样率大小的空洞卷积进行特征提取,分别是一个卷积核为1×1的标准卷积和3个采样率分别为6、12、18的空洞卷积,最后将得到的4路输出特征进行通道域上的叠加融合。由于目标的尺度变化问题是另一个影响分割结果的重要因素,不同采样率大小的空洞卷积能有效捕获多个尺度的感受野,编码更多维度的语义信息,进而增强处理目标物体尺度变化的能力。MFFM不仅具有使得低级特征与高级特征可以更好融合的功能,而且具有处理多尺度物体分割的作用。

2.2 语义流场上采样模块

与传统的直接对解码器输出的特征图进行双线性插值、反卷积等操作来恢复特征图尺寸大小不同,Li等[15]提出了一种基于语义流场的对齐模块(Flow alignment module,FAM)代替传统的上采样方法,学习相邻层特征映射之间的语义流场,使高级特征被很好地融合到具有高分辨率的低级特征图中。

为了解决在恢复特征图分辨率过程中,由于现有上采样方式的不理想,导致分割目标的边界不够清晰、区域细节信息不完整的问题,本文在Li等[15]的研究基础上提出一种基于强化语义流场的特征对齐模块(Channel-attention flow alignment module,CFAM)代替传统的上采样方法,并将其以更有效的方式嵌入到网络模型内。CFAM将来自相邻级别的特征图作为输入,通过评估它们之间的差异来学习这2个不同分辨率的网络层之间的语义流场,最后利用该语义流场将低分辨率特征图扭曲为高分辨率特征图,解决语义由深层到浅层的无效传播引起的低精度问题。CFAM模块的结构如图2所示。

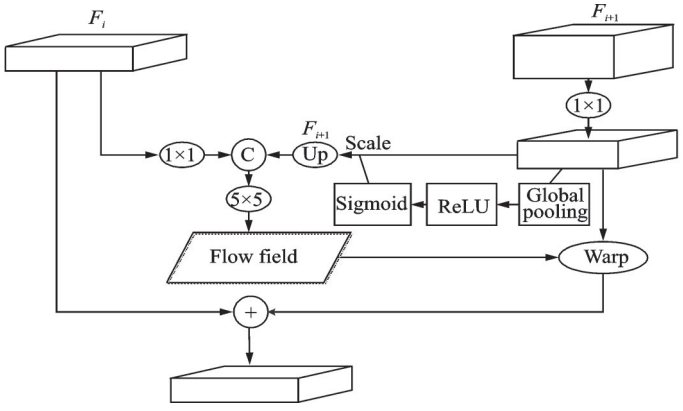

图2 CFAM结构图

Fig.2 CFAM module structure

2.2.1 语义流场生成

语义流场是通过将高分辨率特征和低分辨率特征有效结合而生成的,该流场将给出关于有效对齐这2个特征图的动态指示。如何更有效地预测网络内部的语义流场对两个相邻层的特征对齐至关重要。

在生成强化语义流场时，通过1×1卷积将来自不同层级输入的特征图压缩到相同的通道深度，使相邻的2个特征图F_i和F_{i+1}具有相同的通道数目，对通道调整后的小分辨率特征图F_{i+1}进行特征选择。首先通过对特征图F_{i+1}进行全局池化操作，生成含有全局上下文信息的$1\times1\times C$特征向量对特征的学习进行指导。然后使用ReLU激活函数对特征图的尺度进行平衡，防止梯度弥散，同时令数据范围分布更加合理，将所生成的特征向量通过Sigmoid函数进行范围压缩，得到1组在0～1之间的数值，即数据的概率分布。最后将这组概率分布与F_{i+1}相乘，加强重要特征，抑制无关特征，得到新的F_{i+1}。再通过一个双线性插值层将特征加强后的F_{i+1}上采样到与高分辨率特征F_i相同的大小，连接F_i和F_{i+1}，将连接得到的特征用作语义流场的学习。最后应用5×5的卷积层对合并的特征进行密集提取，得到具有更强表征能力的语义流场$\Delta_i\in\mathbf{R}^{H_i\times W_i\times 2}$。上述步骤可以表示为

$$\Delta_i=\mathrm{conv}(\mathrm{cat}(F_i,F_{i+1})) \tag{3}$$

式中：cat(•)表示通道域的合并操作；conv是5×5卷积层。这里卷积核的大小是经过多次实验测试所得。

2.2.2 特征生成

在得到二维的语义流场Δ_i后，利用该语义流场对低分辨率特征图F_{i+1}进行特征扭曲映射，得到新的高分辨率特征图F_i，具体结构如图3所示。图3中最上方的二维网格即为语义流场，也就是偏移量，该偏移量为高分辨率特征降采样到低分辨率语义特征提供了采样索引。根据这个索引Δ_i，通过简单的加法运算，将高分辨率特征图空间网格上的每个位置p_i对应到相邻的低分辨率特征图空间网格上的位置点p_{i+1}，而由于低分辨率特征图和流场Δ_i之间存在分辨率差距，对应的映射网格及其偏移应该减半，即图中第1幅图到第2幅图的过程，紫色点为位置偏移，红色点即为对应得到的p_{i+1}，对应关系的计算公式为

$$p_{i+1}=\frac{p_i+\Delta_i(p_i)}{2} \tag{4}$$

式中：$\Delta_i(p_i)$为对应的坐标偏移；2代表采样倍率。

然后根据双线性采样机制，对p_{i+1}的4个邻域（左上、右上、左下和右下）进行线性插值，即图中的蓝色点部分，以$\tilde{F}_{i+1}(p_i)$近似CFAM的最终输出，即图3中最下方网格内的红色点部分，公式为

$$\tilde{F}_{i+1}(p_i)=F_i(p_{i+1})=\sum_{p\in N(p_{i+1})}w_pF_{i+1}(p) \tag{5}$$

式中：$N(p_{i+1})$表示F_{i+1}中点p_{i+1}的邻域；w_p表示由扭曲网格的距离估计的双线性核权重。

这种根据学习的语义流场定义领域，将低分辨率特征图中像素线性插值到高分辨率特征图的方法，明确建立特征图之间的对应关系，能够解决高低层特征空间上的误对齐问题，使语义信息更加有效地从深层传递到浅层。

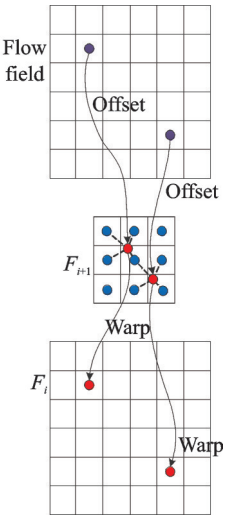

图3 特征生成

Fig.3 Warp procedure

2.3 网络整体架构

图4是基于强化语义流场和多级特征融合网络MCFNet的整体架构，可以看作是一个编码器-解码器结构。通过迁移学习，利用预先训练好的ResNet101[16]模型作为主干网络进行特征提取，提供不同级别的特征表示。将编码由浅层到深层划分为4个阶段，第1阶段输出为f_1，分辨率为输入特征的1/4；第2阶段输出为f_2，分辨率为1/8；第3阶段f_3分辨率为1/16；第4阶段f_4为1/32。其中越深层次的输出特征，感受野就越大，包含的语义信息越丰富，较浅层次的输出特征，分辨率更高，包含更多的空间细节和位置信息。

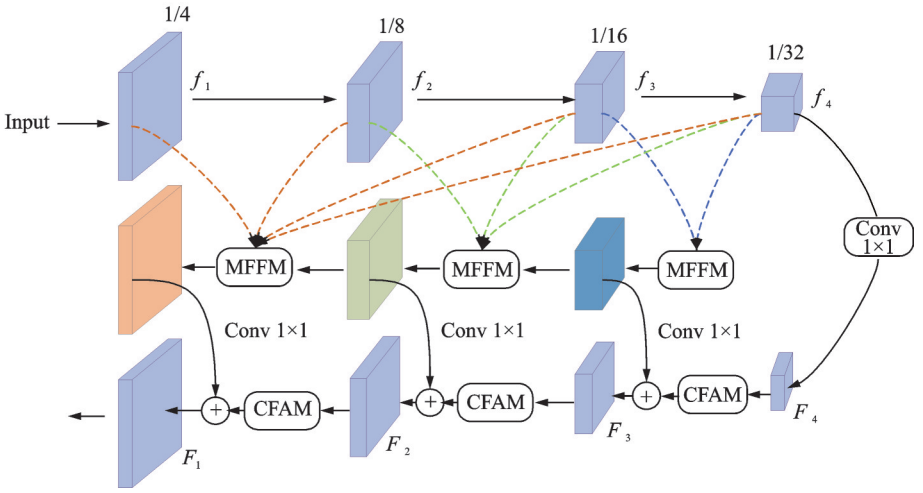

图4 MCFNet整体网络结构

Fig.4 Overall network architecture of MCFNet

2.3.1 编码部分

待分割图片输入主干网络进行特征提取后,通过2.1节提出的MFFM模块将不同层级输出的特征与比该层层级更低的特征不断地在通道域上进行融合,使低级特征与高级特征更高效地对齐,减少噪声的引入,同时增强信息的表征能力和丰富度。为了减少参数量和计算开销,进行多级特征融合后,通过1×1卷积将来自MFFM的特征以及阶段4的特征f_4压缩到相同的通道维度,即像素的最终类别数12,在降低计算负担及通道冗余的同时,为后续解码部分的特征相加融合作准备。

2.3.2 解码部分

在解码部分,采用2.2节提出的基于强化语义流场的上采样模块CFAM来代替传统的上采样方法,将通道数目调整后的高层特征图F_4上采样到原图1/8的分辨率,然后与编码部分对应同等尺寸的多级融合特征进行像素相加融合,得到解码输出特征F_3,重复此过程,直到得到分辨率恢复为原图1/4的F_1,最后对F_1进行4倍上采样操作,使分辨率恢复至原始输入特征图大小,得到最终的预测结果,结束整个分割过程。

3 实验与分析

3.1 实验设置和评估指标

实验是在Windows 10操作系统下进行的,所应用的深度学习框架为Pytorch,版本为1.9.0,硬件配置为英伟达2060显卡,英特尔酷睿i5-9400、显存6 GB。每次实验使用相同的超参数,输入的每个小批次(batch_size)为2,初始的学习率为0.001,迭代次数(Epoch)为200。每当经过50个Epoch时,学习率降低一半,以防止训练时损失函数震荡导致不易于收敛的问题。使用Adam算法作为梯度下降优化算法。

实验在CamVid数据集上进行。CamVid是一个道路场景数据集,由12个类别的701幅具有高质量像素级注释的图像组成,包括367幅训练图像、101幅验证图像和233幅测试图像,分辨率为960像素×720像素。最后使用标准的平均交并比和图像分类正确率来报告分割精度。

3.2 消融实验

3.2.1 MFFM模块消融

以ResNet101网络作为主干网络，不增加任何前文提出的优化模块，采用FCN的解码方式，从f_4开始，逐级通过线性插值法2倍上采样高层特征，进行通道数目调整后与对应的低级特征图在通道域上叠加，直到恢复特征图分辨率至原始输入图像的1/4，最后将通道数目调整至类别数，并4倍上采样得到最终的分割结果图。在该基准网络的基础上，分别测试语义嵌入分支SEB和本文提出的MFFM两种融合方式的效果，对训练后的模型进行测试得到结果如表1所示。表1中mIoU为平均交并比，Acc为图像分类准确度，Δa为变化值。可以看到采用SEB的融合方式后，mIoU和Acc分别比基准网络提高了1.39%和1.31%，而采用本文提出的MFFM融合方式后，mIoU和Acc分别比SEB又提高了0.92%和0.86%，相比于基准网络精确度有了显著的提升，可见本文提出的不同层级特征融合策略的有效性。

表1 不同特征融合方式的对比实验

Table 1 Comparison of different feature fusion methods %

Method	Backbone	mIoU	Δa	Acc	Δa
Baseline	ResNet101	62.64	—	63.17	—
Baseline+SEB	ResNet101	64.03	1.39↑	64.48	1.31↑
Baseline+MFFM	ResNet101	64.95	0.92↑	65.34	0.86↑

分割的可视化效果图如图5所示。可以看到对于输入的原始图片，采用多级融合后生成的可视化分割结果更精细，对图像中的小目标物体，如行人及宠物的识别能力更强，这同样可以证明本文提出的多级特征融合网络通过加强不同层级特征的利用率，提升了网络对像素的分类能力；同时利用并行的多尺度空洞卷积对不同层级进一步进行特征提取，有效地获取多尺度上下文信息，以适应不同尺度的目标；增强特征的丰富度和表达能力，有效地缓解了较深网络层中的空间信息不足和小目标易丢失的问题，对分割效果有所改善。

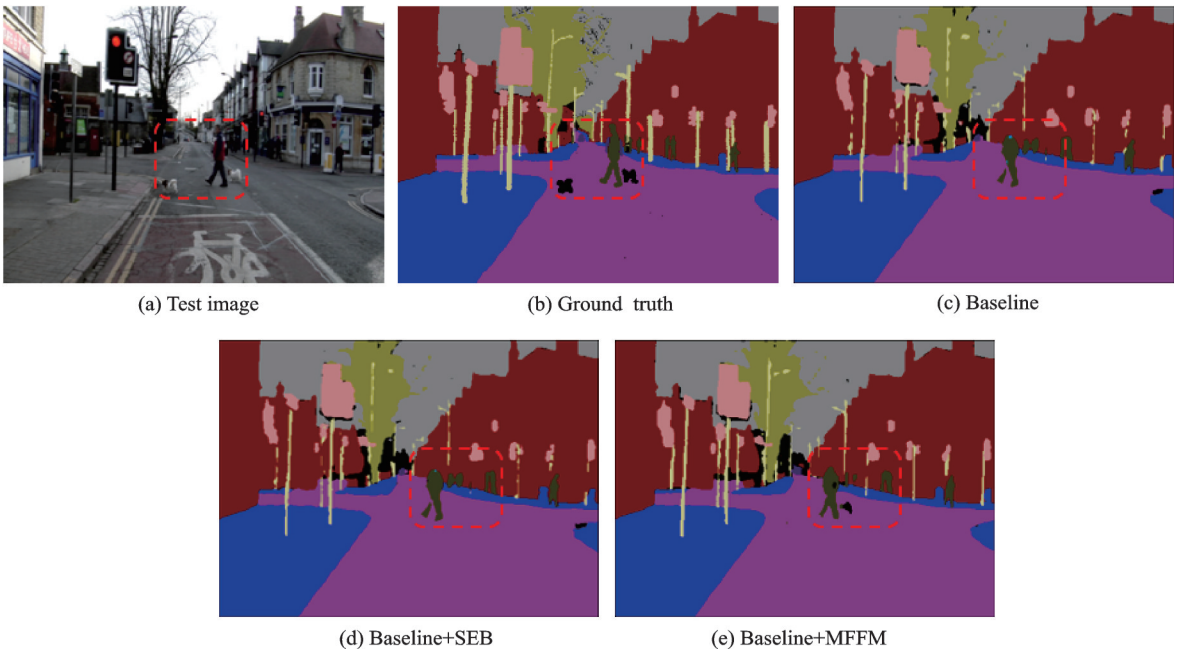

图5 不同特征融合方式的结果对比图

Fig.5 Scene parsing results based on different feature fusion methods

3.2.2 CFAM模块消融实验

本节使用相同的基准网络,MFFM的解码方式,分别针对CFAM的内部结构和CFAM上采样方式的有效性进行多组实验对比。

CFAM模块在生成语义流场时,需要通过卷积对相邻层合并的特征图进行特征提取,并将通道维数压缩至二维。本节对多种卷积核大小进行了尝试,测试结果如表2所示。实验发现使用3×3大小的卷积时,相比于1×1的卷积,mIoU和Acc分别提升了0.54%和0.63%;使用5×5大小的卷积时,相比于3×3的卷积,mIoU和Acc又分别提升了0.42%和0.29%。可见当卷积核较小时,提取到的特征信息感受野不足,增大卷积核的尺寸可以增强语义流场的表征能力,从而提升分割精度。但当使用更大的7×7卷积核时,分割精度不升反降,可见当卷积核过大时,会导致提取特征的空间细节信息损失,生成的语义流场不够密集,从而使后续对相邻层的特征映射效果不佳。5×5大小的卷积核在二者之间取得了最好的均衡,得到的分割精度最高。

表2 语义流场中卷积核大小的性能比较

Table 2 Comparison of different convolution kernel sizes in semantic flow field %

卷积核	mIoU	Δa	Acc	Δa
$K=1$	66.41	—	66.93	—
$K=3$	66.95	0.54↑	67.56	0.63↑
$K=5$	**67.37**	**0.42**↑	**67.85**	**0.29**↑
$K=7$	66.68	0.69↓	67.37	0.43↓

在确定CFAM的内部结构后,为了进一步探讨CFAM的效果,分别使用双线性插值、反卷积、FAM以及2.2节提出的CFAM作为网络解码部分的上采样方法,实验结果如表3所示。可以看到基于语义流场的上采样方法相较于传统的双线性插值和反卷积方法,分割精度有非常显著的提升,而基于强化语义流场的上采样方法通过对生成流场的加强与优化,又进一步提升了分割精度。

表3 采用不同上采样方式的性能比较

Table 3 Comparison of different upsampling methods %

方法	mIoU	Δa	Acc	Δa
双线性插值	64.95	—	65.34	—
反卷积	65.42	0.47↑	65.89	0.55↑
FAM	66.41	1.19↑	67.23	1.34↑
CFAM	**67.37**	**0.76**↑	**67.85**	**0.62**↑

4种上采样方法的实际分割效果如图6所示,可以直观地看到使用CFAM后,分割的道路、汽车与标示牌都与标签值更相近,细节结构更加清晰,边界分割效果更加平滑,误分类减少,可见设计的CFAM相比其他上采样方式获得了更好的分割效果。

(a) Test image (b) Ground truth (c) Bilinear interpolation

(d) Deconvolution (e) FAM (f) CFAM

图6 不同上采样方法的分割结果图

Fig.6 Scene parsing results based on different upsampling methods

3.2.3 整体网络各模块消融实验

整体网络各组件的消融结果如表4所示。可以观察到在基准网络上单独增加多级特征融合模块后，mIoU和Acc分别提升了2.31%和2.17%，这表明MFFM可以更高效地融合不同层次的特征，大幅地提升分割精度。在基准网络上单独增加CFAM模块后，mIoU和Acc分别提升了2.78%和2.71%，表明了基于强化语义流场CFAM上采样方法的显著有效性。在使用MFFM的基础上，采用CFAM的上采样方式，再次将mIoU提高2.42%，Acc提高2.51%。根据以上数据综合分析，每个模块的添加都使得分割精度指标大幅提升，为网络性能带来了显著的增强。

表4 各模块效果的消融

Table 4 Ablation on the effect of each module %

方 法	mIoU	Δa	Acc	Δa
Baseline	62.64	—	63.17	—
Baseline +MFFM	64.95	2.31↑	65.34	2.17↑
Baseline + CFAM	65.42	2.78↑	65.88	2.71↑
Baseline +MFFM +CFAM	67.37	2.42↑	67.85	2.51↑

3.3 与其他语义分割方法对比

将本文所设计的MCFNet网络和与近年来提出的多个道路场景的语义分割模型进行指标对比，选取使用相同骨干网络的GCN[11]、ExFuse[7]、PSPNet[8]和Deeplabv3+[9]进行更客观的实验对比，在相同的实验参数设置下实验数据如表5所示。可以明显看出，本文提出的网络模型MCFNet的mIoU和Acc指标明显高于其他网络模型，MCFNet网络在精确度上的表现最好。分割的实际结果图如图7所示，可以看到，本文提出的网络相比于其他几种模型，对汽车、行人、道路和交通标志等的边缘分割效果有较为明显的提升，边界明显更加完整和连续，同时能够更准确且完整地分割出整个目标物体的形状，如小目标物体宠物以及大型建筑，说明MCFNet显著增强了网络多尺度性能，能更好处理不同尺度的目标，提升不同尺度物体的分割精度。综合以上数据的分析可以证明MCFNet网络在CamVid道路场景数据集上的分割精度更高，效果更优。

表5 在CamVid数据集上的与不同模型的对比

Table 5 Comparison on CamVid set with different models %

方 法	主干网络	mIoU	Acc
GCN[11]	ResNet101	55.63	57.42
ExFuse[7]	ResNet101	60.77	62.86
PSPNet[8]	ResNet101	64.82	65.25
Deeplabv3+[10]	ResNet101	65.29	66.14
MCFNet	**ResNet101**	**67.37**	**67.85**

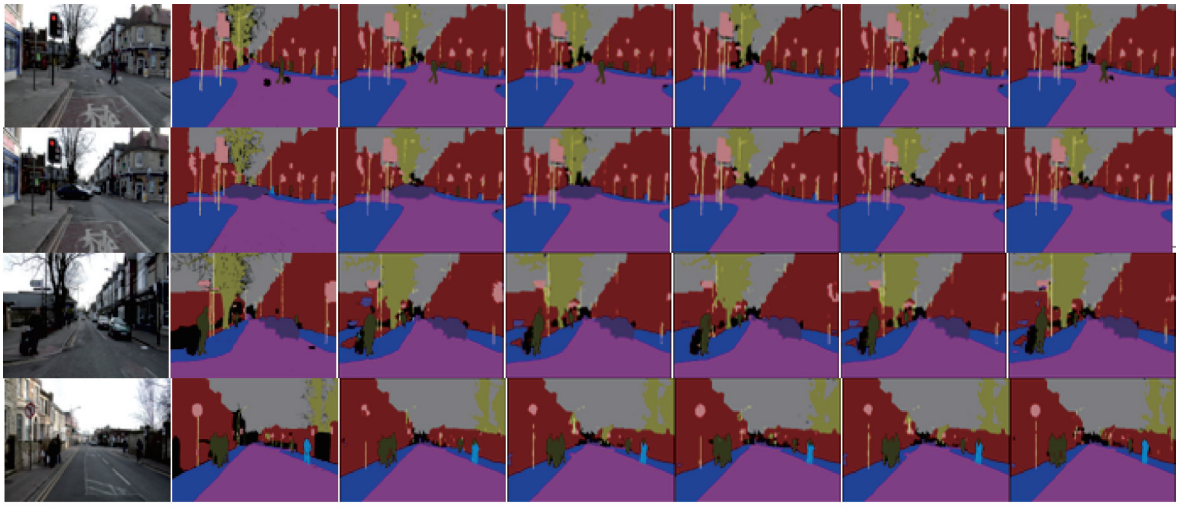

图 7 不同模型的分割结果对比图

Fig.7 Scene parsing results based on different network models

4 结束语

针对道路场景解析,本文提出了多级特征融合模块使高级特征和低级特征更有效地融合,同时获取丰富的空间细节和多尺度的感受野,更有利于解决道路交通场景中的物体尺度变换问题,并且在解码过程中设计了一种基于强化语义流场的特征对齐方式代替传统的上采样方法,生成高质量的高分辨率特征图,使分割结果细节结构更加清晰,边界分割效果更加平滑。通过对实验结果的分析,证明了本文设计的方法有助于提升分割精度的有效性。然而本文模型在轻量化和实时性上有待提高,也是接下来继续研究的方向。

参考文献:

[1] 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.

[2] RONNEBERGER O, FISCHER P, BROX T. U-Net:Convolutional networks for biomedical image segmentation[J]. Medical Image Computingand Computer-Assisted Intercentin,2015, 9351: 241.

[3] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature Pyramid networks for object detection[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu:USA, 2017: 936-944.

[4] LI Hanchao, XIONG Pengfei, FAN Haoqiang, et al. DFANet: Deep feature aggregation for real-time semantic segmentation [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Los Angeles : IEEE, 2019: 9514-9523.

[5] LIN G H, MILAN A, SHEN C H, et al. RefineNet:Multi-path refinement networks for high resolution semantic segmentation [C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017: 5168-5177.

[6] YU C, WANG J, PENG C, et al. BiSeNet: Bilateral segmentation network for real-time semantic segmentation[C]// Proceedings of European Conference on Computer Vision. Munich: Springer ,2018: 325-341.

[7] ZHANG Zhenli, ZHANG Xiangyu, PENG Chao, et al. ExFuse: Enhancing feature fusion for semantic segmentation[C]// Proceedings of European Conference on Computer Vision. Munich: Springer, 2018: 273-288.

[8] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]//Proceedings of the 2017 IEEE Conference on Computer

Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017: 6230-6239.

[9] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. (2017-02-10)[2021-05-26]. https://arxiv.org/abs/1706. 05587.

[10] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of European Conference on Computer Vision. Munich: Springer, 2018: 833-851.

[11] LAN Hengrong, JIANG Daohuai, YANG Changchun, et al. Y-Net: Hybrid deep learning image reconstruction for photoacoustic tomography in VIVO[J]. Photoacoustics, 2020. DOI:10.1016/j.pacs.2020.100197.

[12] 佘玉龙,张晓龙,程若勤,等.基于边缘关注模型的语义分割方法[J].计算机应用,2021,41(2): 343-349.
SHE Yulong, ZHANG Xiaolong, CHENG Ruoqin, et al. Semantic segmentation method based on edge concern model[J]. Computer Application, 2021, 41(2): 343-349.

[13] 何康辉,肖志勇. LRUNet:轻量级脑肿瘤快速语义分割网络[J].中国图像图形学报,2021,26(9): 2233-2242.
HE Kanghui, XIAO Zhiyong. LRUNet: Fast semantic segmentation network for lightweight brain tumors[J]. Journal of Image and Graphics, 2021, 26(9): 2233-2242.

[14] 陈泽斌,罗文婷,李林.基于改进U-Net模型的路面裂缝智能识别[J].数据采集与处理 ,2020,35(2): 260-269.
CHEN Zebin, LUO Wenting, LI Lin. Intelligent pavement crack recognition based on improved U-Net model[J]. Journal of Data Acquisition and Processing, 2020, 35(2): 260-269.

[15] LI Xiangtai, YOU Ansheng, ZHU Zhen, et al. Semantic flow for fast and accurate scene parsing[C]//Proceedings of European Conference on Computer Vision. Glasgow UK: Springer Cham, 2020: 775-793.

[16] HE K M, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770-778.

作者简介:

项建弘(1977-),男,副教授,研究方向:5G无线通信、人工智能与深度学习、卫星通信、智能天线、自适应信号处理,E-mail:xiangjianhong@hrbeu.edu.cn。

钟瑜(1979-),男,高级工程师,研究方向:嵌入式信号处理平台、无线通信算法,E-mail:jade.zhong@hotmail.com。

刘苗(1997-),女,硕士研究生,研究方向:深度学习、语义分割,E-mail:1655636280@qq.com。

王霖郁(1977-),**通信作者**,女,副教授,研究方向:5G移动通信、宽带卫星通信、智能天线、自适应信号处理,E-mail:wanglinyu@hrbeu.edu.cn。

(编辑:刘彦东)