联合多连接特征编解码与小波池化的轻量级语义分割

易清明[1,2]，王 渝[1]，石 敏[1]，骆爱文[1*]

（1. 暨南大学 信息科学技术学院，广州 510632；2. 泰斗微电子科技有限公司，广州 510663）

摘要 语义分割是当前场景理解领域的基础技术之一。现存的语义分割网络通常结构复杂、参数量大、图像特征信息损失过多和计算效率低。针对以上问题，基于编–解码器框架和离散小波变换，设计了一个联合多连接特征编解码与小波池化的轻量级语义分割网络 MLWP-Net（Multi-Link Wavelet-Pooled Network），在编码阶段利用多连接策略并结合深度可分离卷积、空洞卷积和通道压缩设计了轻量级特征提取瓶颈结构，并设计了低频混合小波池化操作替代传统的下采样操作，有效降低编码过程造成的信息丢失；在解码阶段，设计了多分支并行空洞卷积解码器以融合多级特征并行实现图像分辨率的恢复。实验结果表明，MLWP-Net 仅以 0.74 MB 的参数量在数据集 Cityscapes 和 CamVid 上分别达到 74.1% 和 68.2% mIoU 的分割精度，验证了该算法的有效性。

关键词 实时语义分割；轻量级神经网络；多连接特征融合；小波池化；多分支空洞卷积

中图分类号 TP391 **文献标志码** A DOI 10.12178/1001-0548.2023124

Lightweight Semantic Segmentation by Combining Multi-Link Feature Codec with Wavelet Pooling

YI Qingming[1,2], WANG Yu[1], SHI Min[1], and LUO Aiwen[1*]

(1. School of Information Science and Technology, Jinan University, Guangzhou 510632, China;

2. Taidou Microelectronic Science and Technology Co., Ltd., Guangzhou 510663, China)

Abstract Semantic segmentation is currently one of the basic technologies in the field of scene understanding. Existing semantic segmentation networks usually result in complex structures, a large number of parameters, excessive loss of image feature information, and low computational efficiency. To address these problems, this work proposes a lightweight semantic segmentation network named MLWP-Net (Multi-Link Wavelet-Pooled Network) which combines features with multiple connections and wavelet pooling based on the encoder-decoder framework and Discrete Wavelet Transform (DWT). In the encoding phase, a lightweight feature extraction bottleneck is designed by combining with the depthwise separable convolution, dilated convolution, and channel compression, using a multi-link strategy to fuse multi-level features; besides, a low-frequency-mixed wavelet pooling operation is employed to replace the traditional downsampling operation for effectively reducing the information loss during the encoding process. In the decoding stage, a multi-branch parallel dilated convolutional decoder is designed to fuse multiple features linked to the different layers in the encoder to recover the image resolution in parallel. The experimental results show that our MLWP-Net achieves 74.1% and 68.2% mIoU segmentation accuracy on the datasets of Cityscapes and Camvid with only 0.74M parameters, which demonstrates its effectiveness for semantic segmentation.

Key words real-time semantic segmentation; lightweight neural network; multi-link feature fusion; wavelet pooling; multi-branch dilated convolution

收稿日期：2023−04−25；修回日期：2023−11−11

基金项目：国家自然科学基金（62002134）；广东省基础与应用基础研究基金（2020A1515110645, 2023A1515010834）；广东省普通高校新型半导体与器件重点实验室项目（2021KSY001）；羊城创新创业领军人才支持计划（2019019）；广东省科技创新战略专项（大学生科技创新培育）（pdjh2023b0061）

作者简介：易清明，博士，教授，主要从事多媒体信息处理方面的研究。

*通信作者 E-mail: luoaiwen@jnu.edu.cn

语义分割技术作为计算机视觉的一部分，目的是为图像中的每个像素分配类别标签，被广泛应用于工业自动化[1]、医疗图像[2]等场景解析领域。尤其在面向自动驾驶的城市交通场景中，高效的语义分割模型可以对道路做出实时场景解析，为路径规划、避让行人障碍等提供有效的辅助信息。然而真实应用场景中往往要求语义分割网络同时具有较高的分割精度和较快的计算速度，这对语义分割的准确性和实时性均提出较高的要求，因此亟须研究出一种能够在分割精度和计算成本之间实现较好权衡的语义分割算法。

现有的提高语义分割准确度的策略大多是加大网络的深度，以期获得更加丰富的图像特征信息。目前分割效果较好的语义分割网络，如 SegNet[3]、DeepLabV3+[4]、RefineNet[5]等都有较高的准确率。但这些网络算法具有较大的模型参数量和较高的计算复杂度，进而影响分割效率。为了将语义分割技术实现落地应用并获得实时处理图像信息的效果，轻量级神经网络设计成为实时语义分割任务的一个重要研究目标。

现有的轻量级网络如 ENet[6]、ERFNet[7]、LEDNet[8]、CGNet[9]、DABNet[10]、FRNet[11]等，其参数量都已经控制在 1 MB 以下。其中，ENet 和 SegNet 是两大经典的轻量化模型，通过采用非对称的编解码结构和通道裁剪策略，SegNet 的参数量仅为 0.36 MB，而 ERFNet 利用非瓶颈残差结构并将标准卷积替换为非对称卷积，降低参数量的同时获得了很好的分割精度。而 Xception[12]使用深度可分离卷积替代标准卷积，增加网络深度的同时还减少了参数量。在 Xception 的基础上，MobileNet[13]引入深度可分离卷积和残差模块来实现模型的压缩和推理的加速，减少卷积操作带来的参数量和计算量的同时保持了较好的分割性能。相较之下，ShuffleNet[14]运用通道混洗的策略，通过转置、分组卷积、通道乱序的方法来促进信息流动，精简模块的同时提高了计算效率。尽管以上网络在参数量方面较小，并保证了一定的分割精度，但仍然难以满足真实场景中的应用需求[15]。

此外，为了降低特征的维度并保留有效信息，现有的大多数语义分割网络均采用下采样池化操作，如最大池化、平均池化、随机池化等。但池化操作往往会使得图像分辨率下降，导致图像特征信息丢失。尽管已有研究者对池化操作的特征信息丢失问题进行了改进，如采用带步长的卷积替代池化操作或采用低通滤波去除高频特征之后再进行下采样操作[16]，但此类操作或增加计算量，或影响网络的特征表达能力。而离散小波变换（DWT）以其强大的时频分析能力，被广泛应用于信号与图像处理领域[17]。随着深度学习的不断发展，越来越多的研究也将其应用于卷积神经网络（CNN）的优化中。如将 DWT 应用到编−解码器中，降低参数量的同时提高网络的运算速度[18]；或将其结合残差网络，利用小波变换提高图像的恢复能力[19]；或将其与注意力机制结合，加强对不同频率分量的特征注意力[20]。然而，现有的小波变换与 CNN 的组合方法并未充分发挥其多通道分频的优势，仍然具有较大的改进空间。

综合以上分析，本文提出了一个联合多连接特征编解码与小波池化的轻量级语义分割网络（Multi-Link Wavelet-Pooled Network, MLWP-Net），包括：轻量化的逐步特征融合模块（Progressive Feature Fusion, PFF）；基于小波变换理论的低频混合小波池化操作（Low-frequency-mixed Wavelet Pooling, LWP），用于实现高效的下采样操作；以及多分支并行空洞卷积解码器（Multi-Branch Parallel Dilated Convolutional Decoder, MPDCD）。

1　多连接特征编解码与小波池化网络

1.1　整体网络结构介绍

本文提出的基于编码器−解码器框架的 MLWP-Net 整体网络结构如图 1 所示。其中，编码器的主要组成包括初始标准卷积模块（Initial Block）、逐步特征融合（PFF）瓶颈结构以及低频混合小波池化（LWP）模块，解码器主要由 MPDCD 构成。其中，初始标准卷积模块用于对原始图像特征信息的提取；拼接（Concat）操作将不同分辨率的原始输入图像与不同卷积层的特征图进行拼接。本文所提的 MLWP-Net 基于多连接思想，从不同深度的网络中获得不同尺度的特征和上下文信息，结合离散小波变换原理实现小波池化，在降低特征图分辨率的同时尽可能减少信息丢失，扩大网络通道数，并在解码时同样基于多连接思想接入主干网络不同层级的特征信息进行多分支特征融合，实现对图像的准确分割解码。

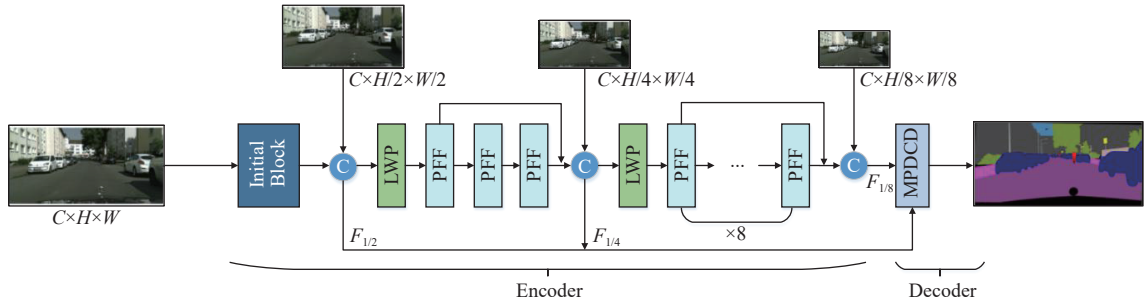
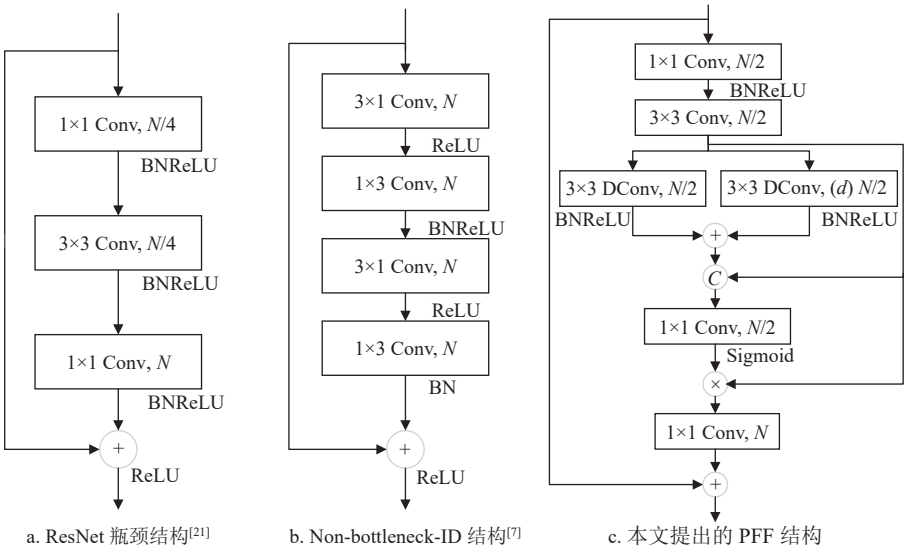


图 1 MLWP-Net 网络的整体结构

1.2 逐步特征融合（PFF）

现存语义分割网络中存在网络整体参数量较大、计算复杂度高、计算速度慢等诸多问题，且随着网络模型深度的加深会带来网络退化问题（如图 2a 中 ResNet[21]），或者，网络模型整体参数量虽小且计算速度快，但其特征提取能力不足（如图 2b 所示的 Non-bottleneck-1D 结构[7]），因此，大多数现有网络模型难以在分割准确性、计算速度、参数量三者之间实现较好的权衡。



a. ResNet 瓶颈结构[21] b. Non-bottleneck-ID 结构[7] c. 本文提出的 PFF 结构

图 2 不同 block 结构的对比

本文利用深度可分离卷积、空洞卷积和通道压缩策略设计了一个轻量级逐步特征融合的特征提取瓶颈结构 PFF，如图 2c 所示，以获得更高的分割精度、更快的推理速度和更低的计算复杂度。不同于图 2b 的 Non-bottleneck-1D 在输入端使用 1×3 和 3×1 卷积模块代替一个标准 3×3 的卷积模块，本文首先在输入端应用 1×1 卷积来进行通道压缩，从而减少计算量，然后运用 3×3 标准卷积来提取局部特征信息，再用两个并行特征提取分支获取多尺度上下文特征信息，并将 3×3 卷积之后获取的局部信息与并行分支获取的多尺度上下文 Concat 起来，从而弥补由于网络层数加深带来的特征信息丢失问题。同时，并行分支中的其中一条支路采用带有空洞率（d=3, 5, 7, 9）的 3×3 深度可分离卷积，增加了网络的深度并扩大网络的感受野。通过 1×1 卷积将通道再次压缩至 1/2 后，利用 Sigmoid 计算通道注意力权重大小，与 3×3 标准卷积后的特征图进行乘积运算，聚合网络学习到的局部特征信息和多尺度特征信息，并筛选出重要的上下文特征信息。最后利用 1×1 卷积恢复通道数，与图 2a 中 ResNet 的瓶颈结构相似，MLWP-Net 也利用残差结构补充原始输入图像的上下文信息，从而解决网络深度加深带来的退化问题，提高特征表达能力。

相较于现存其他特征提取模块而言，本文设计的逐步特征融合 PFF 模块在结构上使用多连接策略，可对局部信息和多尺度上下文信息进行有效聚合，从而提高语义分割网络的特征提取能力；该模

块使用通道压缩策略，实现模型的轻量化。

1.3　低频混合小波池化（LWP）

在特征提取过程中，产生的误差主要来源于两个方面：池化区域大小受限造成的估计值方差增大和卷积层参数误差造成估计均值的偏移。平均池化能够通过计算池化区域内的平均值，更多地保留图像的背景信息，从而减小第一种误差；最大池化能够通过计算池化区域内的最大值并记录该最大值所在输入数据中的位置，从而减小第二种误差，用于提取特征纹理或高频边缘细节。然而上述两种直接下采样的池化操作均忽略了高频特征信息和低频特征信息在空间域与通道域的位置分布差异，容易造成在频域特征间的混叠效应[22]，而且还会导致图像中部分特征信息丢失或弱化，小波变换是在傅里叶的基础上发展而来，可以扩展到时频域进行图像分析，离散小波变换可以捕获特征图的频率和位置信息，有利于保留纹理细节。

为了充分利用特征图的低频图像信息和高频边缘信息。本文采用 Haar 小波基函数进行多分支的低频混合小波池化，实现下采样效果。Haar 小波基函数作为具备紧支性的正交小波函数，其正交性有利于对图像特征的精确重构；其对称性使得小波滤波呈线性相位，有利于提高网络的推理速度；且 Haar 小波计算复杂度低，不会对网络实时性造成太大影响。因此，本文充分结合小波变换的优势，利用不同频率特征信息，从空间域和通道域两个维度减少传统下采样操作导致的信息丢失问题，设计了一个高效的低频混合小波池化（Low-Frequency-Mixed Wavelet Pooling, LWP）模块，如图 3 所示。首先在空间域上对输入特征图进行离散小波变换（DWT）处理，将其分解为低频系数 LL（主要图像信息）、水平方向的高频系数 HL、垂直方向的高频系数 LH 和对角方向的高频系数 HH（细节图像信息）。然后，将低频系数 LL 分别与 3 个高频系数叠加后，再通过逆小波变换 IDWT 实现信息重构，即对相应空间域维度上的特征信息进行更新，分别获得重构系数 LL_1、LH_1、HL_1、HH_1。此外，为了能够结合最大池化与平均池化的优点并减小图像特征尺寸，每类重构特征分别采取 3×3 卷积 $f_{\text{conv3}\times3}$ 进行特征学习后再进行混合池化操作，得到新的重构系数 LL_2、LH_2、HL_2、HH_2。由于分组重构过程每类特征都包含了低频特征，即在空间域上具有一定的空间特征信息，从而减少了后面卷积操作和池化操作造成的信息损失。最后，将 4 类特征拼接起来并通过 1×1 卷积 $f_{\text{conv1}\times1}$ 恢复其通道数，得到最终经过小波分组重构和池化操作之后的特征输出，完成下采样操作，4 个并行分支的计算过程为：

$$\begin{cases} \mathrm{LL}_2 = \mathrm{MixPool}(f_{\text{conv3}\times3}(\mathrm{IDWT}(\mathrm{LL}))) \\ \mathrm{LH}_2 = \mathrm{MixPool}(f_{\text{conv3}\times3}(\mathrm{IDWT}(\mathrm{LL}+\mathrm{LH}))) \\ \mathrm{HL}_2 = \mathrm{MixPool}(f_{\text{conv3}\times3}(\mathrm{IDWT}(\mathrm{LL}+\mathrm{HL}))) \\ \mathrm{HH}_2 = \mathrm{MixPool}(f_{\text{conv3}\times3}(\mathrm{IDWT}(\mathrm{LL}+\mathrm{HH}))) \end{cases} \tag{1}$$

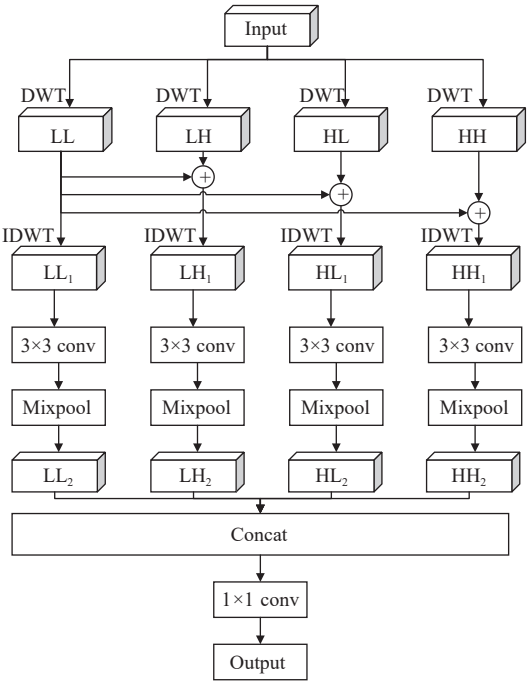


图 3　低频混合小波池化（LWP）

1.4　编码器构建

基于以上两个高效的逐步特征融合模块 PFF 和低频混合小波池化 LWP 模块，本文构建了一个有效融合多层次特征信息的编码器，实现轻量化且高精度特征提取。

如图 1 所示，在编码器初始阶段，MLWP-Net 使用 3 个 3×3 标准卷积对原始图像进行预处理，其中为了改变原始输入图像的尺寸大小。经过初始阶段提取的特征图像与原始输入的 1/2 分辨率图像进行特征融合 Concat，保持初始阶段提取的图像特征并补充细节信息，从而提高分割能力；使用 LWP 对输出特征图进行下采样操作，在降低特征图尺寸的同时尽可能保留全部特征信息。再使用 3 个 PFF 模块进行多尺度上下文特征信息提取，然后对输出特征图与原始输入的 1/4 分辨率图像和首个 PFF 的输出图像进行第二次特征融合，

从而弥补不同网络深度的特征信息。为了进一步压缩模型，再次使用 LWP 对特征图进行下采样，并将经过下采样后的特征图输入 8 个 PFF 模块中，进一步提取深层次的特征信息。最后，将最后一个 PFF 的输出特征图像与原始输入的 1/8 分辨率图像和第一个 PFF 的输出图像进行第三次融合，从而获得编码器的最终输出结果。

1.5 多分支并行空洞卷积解码器（MPDCD）构建

为了更有效地融合不同层级网络的特征信息，利用编码器所提取的特征信息来指导像素级分类，从而完成解码，本文设计了一种多分支并行的空洞卷积解码器（MPDCD），如图 4 所示。

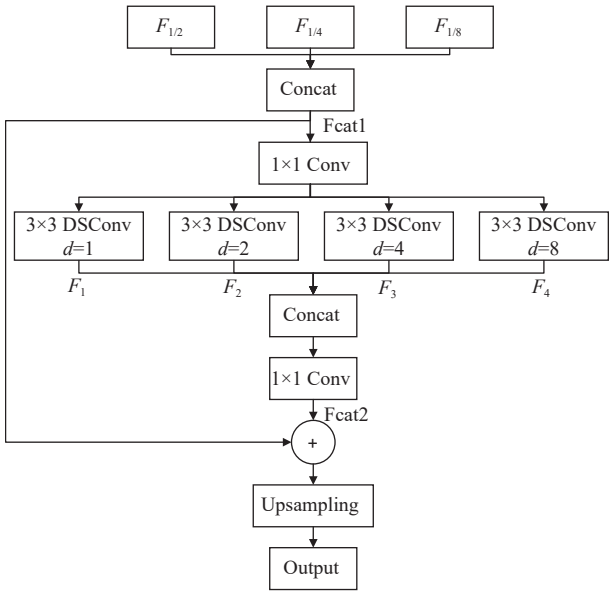


图 4 多分支空洞卷积解码器（MPDCD）

DeepLab 网络中的空洞空间池化金字塔（Atrous Spatial Pyramid Pooling, ASPP）[23] 模块只利用编码器中最后一个输出特征图进行解码，MPDCD 将主干网络中多个层级的特征输出 $F_{1/2}$、$F_{1/4}$、$F_{1/8}$ 融合后作为输入 F_{cat1}，其次运用 1×1 卷积 $\text{f}_{\text{conv1×1}}$ 实现通道混洗，加强不同通道间的信息交流，并采用多分支扩展的方式，利用不同空洞率（d=1, 2, 4, 8）的深度可分离卷积 f_{DSConv} 再次进行特征学习，其中较小空洞率的深度可分离卷积有利于对小目标的识别，较大空洞率的深度可分离卷积能够在保持特征图尺寸不变的情况下增大特征感受野，从而完成多尺度上下文信息结合，得到多分支扩展融合后的输出 F_{cat2}。为了进一步提高特征图恢复精度，再次结合 3 个不同层级的特征融合，共同指导图像特征信息的恢复，最后利用上采样操作将其恢复至原始输入图像的尺寸。经过 MPDCD 解码后，生成最终的预测结果 y，计算过程为：

$$\begin{cases} F_{\text{cat1}} = f_{\text{concat}}(F_{1/2}, F_{1/4}, F_{1/8}) \\ F_{\text{cat2}} = f_{\text{conv1×1}}(f_{\text{concat}}(F_1, F_2, F_3, F_4)) \\ y = f_{\text{upsampling}}(F_{\text{cat1}} + F_{\text{cat2}}) \end{cases} \quad (2)$$

式中，$F_{1/2}$、$F_{1/4}$、$F_{1/8}$ 分别为主干网络中编码器不同层级的输出特征；F_1、F_2、F_3、F_4 分别为经过不同空洞率的深度可分离卷积之后的输出特征结果。

2 实验分析

2.1 数据集与实验参数

2.1.1 实验数据集

本次实验使用无人驾驶环境下的公开数据集 Cityscapes[24] 和 CamVid[25] 进行实验。Cityscapes 数据集是一个大型城市道路场景语义分割数据集，该数据集由 5 000 张像素级标注图像组成，图像分辨率为 1 024×2 048，包含 19 个类别，其中训练集为 2 975 张图像，验证集为 500 张图像，测试集为 1 525 张图像。CamVid 数据集是一个从驾驶汽车角度拍摄的交通场景数据集，共计 701 张图像，图像分辨率为 720×960，包含 11 个类别，其中训练集为 367 张图像，验证集为 101 张图像，测试集为 233 张图像。

2.1.2 参数设置

本文提出的 MLWP-Net 均在 CUDA11.4 的 Pytorch 深度学习框架下使用单个 RTX3090 GPU 进行训练和测试。训练时，采用的优化策略为基于动量的批随机梯度下降，动量 momentum 设置为 0.9，权重衰减为 1.0×10^{-4}，初始学习率为 4.5×10^{-2}，Power 系数为 0.9，使用交叉熵损失函数计算损失。训练过程中不采用其他预训练模型，对输入图像使用随机镜像和随机尺度方式进行预处理，基于 Cityscapes 测试集的实验使用 512×1 024 分辨率的图像；基于 CamVid 测试集的实验使用 360×480 分辨率的图像。

2.2 消融实验

为了验证所提 MLWP-Net 以及各个模块的有效性，本文设计了多个与模块对应的消融实验，并基于 Cityscapes 数据集，将 MLWP-Net 与现有的实时性语义分割网络进行对比。记录由不同模块构成的骨干网络的分割精度（mIoU）、浮点运算量（GFLOPs）以及参数量（Params），实验结果如表 1 所示。

表 1　不同卷积模块在 MLWP-Net 上的实验结果

Bottleneck	mIoU/%	GFLOPs	Params/MB
ResNet[21]	60.4	38.4	**0.57**
Non-bt-1D[7]	71.8	35.6	0.99
本文PFF	**73.5**	**18.1**	0.74

2.2.1　PFF 模块消融实验

为了验证所提多连接逐步特征融合的瓶颈模块 PFF 的有效性，本文将 ResNet 的瓶颈模块、ERFNet 的 Non-bt-1D 模块以及本文 MLWP-Net 的 PFF 模块分别作为编码器的主要特征提取模块，并基于 Cityscapes 验证集中分辨率为 512×1 024 的图像进行实验对比。编码器的输出结果仍然采用 MLWP-Net 的解码器进行解码。

从表 1 中可以看出，ResNet 以最低 0.57 MB 的参数量实现了 60.4% 的分割精度，但其浮点计算量却较高，分割精度难以满足真实道路场景的应用需要。而 Non-bt-1D 受限于其非瓶颈结构，相比 Bottleneck 构成的骨干网而言，其参数量增加了 0.42 MB，但其分割精度却提升到了 71.8%。而由 PFF 模块构成的 MLWP-Net 网络相比 Bottleneck 构成的骨干网仅以 0.2 MB 不到的参数量提升了 13.1% 的分割精度；相比 Non-bt-1D 构成的骨干网而言，PFF 模块不仅拥有更低的参数量和更少的计算复杂度，还提高了近 2% 的分割精度。以上消融实验表明，由 PFF 模块构成的骨干网络具有轻量化、更低计算复杂度和更强的特征提取能力的优点，满足真实道路场景下的语义分割应用需求。

2.2.2　LWP 消融实验

为了验证本文所提的低频混合小波池化模块 LWP 在图像下采样过程的有效性，将 LWP 分别应用到 ERFNet、DABNet、ESNet 中替代传统下采样 DownSample 模块，记录不同网络在使用 LWP 替换原有下采样操作前后网络评价指标的变化情况，实验结果如表 2 所示。

表 2　LWP 在 Cityscapes 验证集的实验结果

Network	LWP	mIoU/%	GFLOPs	Params/MB
ERFNet[7]	—	68.0	35.4	2.06
	√	72.8	35.9	2.04
DABNet[10]	—	70.1	13.7	0.76
	√	70.6	14.4	0.65
ESNet[26]	—	70.7	32.1	1.66
	√	72.0	32.7	1.63
MLWP-Net	—	72.3	17.6	0.84
	√	73.5	18.1	0.74

从表 2 中可以看出，不同网络在使用 LWP 操作后，虽然浮点计算量有轻微的上升，但各个网络的参数量均有所下降，且分割精度均有不同程度的提高。具体地，在使用 LWP 操作后，ERFNet、DABNet、ESNet 在分割精度上分别增加了 4.8%、0.5%、1.3% 的 mIoU。以上消融实验结果表明，低频混合小波池化模块不仅能够抑制传统下采样操作导致的信息丢失问题，并且能够提高分割精度，同时还可以将 LWP 嵌入到多种不同结构的卷积神经网络中，证实了该操作的有效性和通用性。

2.2.3　MPDCD 消融实验

为了验证所提 MPDCD 解码器的图像特征恢复能力，本文将 MPDCD 解码器分别应用到 CGNet、FRNet 中替代网络原有解码器，分别记录不同网络使用 MPDCD 解码器前后在 Cityscapes 验证集上的分割精度、浮点计算量和参数量的变化情况，实验结果如表 3 所示。

表 3　MPDCD 在 Cityscapes 验证集的实验结果

Module	MPDCD	mIoU/%	GFLOPs	Params/MB
CGNet[9]	—	64.8	9.24	0.49
	√	70.1	15.51	0.51
FRNet[11]	—	70.4	16.97	1.01
	√	71.1	23.27	1.03
MLWP-Net	—	72.1	13.60	0.73
	√	73.5	18.10	0.74

从表 3 中可以看出，CGNet 和 FRNet 在使用 MPDCD 解码器替代网络原有解码器后，其网络参数量均基本保持不变，而 mIoU 却有较大提升，分别增加了 5.3% 和 0.7%。相比较而言，虽然多分支空洞卷积结构会增加部分计算复杂度，但网络分割精度却有较大提升，证明了 MPDCD 解码器能够提升不同网络对分割目标的空间信息特征恢复能力，具有轻量化、高准确率的优点。

2.2.4　小波基函数消融实验

小波基函数的选用需要结合信号本身的特点及其对网络分割精度、推理速度的影响。为了选取合适的小波基函数，本文分别针对不同的离散小波基函数进行实验对比。

从表 4 中可以看到，在网络其他条件均相同的情况下，所选用的 Haar 小波基函数能够使网络在 CamVid 数据集上实现 68.2% 的 mIoU，而选用 db1、rbio1.1、bior1.1 小波基函数的网络实现分割精度分别为 67.86%、67.38%、67.24%mIoU。Haar 小波是

具有紧支性和对称性的正交小波函数，其正交性有利于对图像特征信息的重构，且对称性有利于提高网络算法的推理速度。以上消融实验结果证明本文选用 Haar 小波作为基函数能够使网络在分割精度上更具优势。

表 4　不同小波基函数在 CamVid 验证集的实验结果

Method	正交性	对称性	紧支性	Speed/fps	mIoU/%
Haar	有	对称	有	95.0	68.2
db1	有	近似对称	有	94.8	67.8
rbio1.1	无	对称	无	95.3	67.4
bior1.1	无	不对称	有	92.1	67.6

2.3　MLWP-Net 与现有网络的整体性能对比

为了检验 MLWP-Net 的整体性能，本文将其与其他轻量化实时性语义分割网络在 Cityscapes 数据集和 CamVid 数据集上进行性能对比，结果见表 5 和表 6 所示。

表 5　不同模型在 Cityscapes 测试集的实验结果

Method	Pretrain	Input Size	mIoU/%	Params/MB	Speed/fps
SegNet[3]	ImageNet	360×640	56.1	29.50	38.2
RefineNet[5]	ImageNet	512×1 024	73.6	118.1	9.1
SQNet[27]	ImageNet	512×1 024	59.8	16.3	25.7
BiseNetV2[28]	No	512×1 024	73.6	6.20	51.0
ENet[6]	No	512×1 024	58.3	**0.36**	27.4
ERFNet[7]	No	512×1 024	68.0	2.10	41.9
LEDNet[8]	No	512×1 024	69.2	0.95	59.6
CGNet[9]	No	512×1 024	64.8	0.49	65.6
DABNet[10]	No	512×1 024	70.1	0.76	102.0
FRNet[11]	No	512×1 024	70.4	1.01	127.0
ESNet[26]	No	512×1 024	70.7	1.66	63.0
EDANet[29]	No	512×1 024	67.3	0.68	105.5
ESPNet[30]	No	512×1 024	60.3	**0.36**	146.0
ContextNet[31]	No	1 024×2 048	66.1	0.85	57.7
Fast-SCNN[32]	No	1 024×2 048	68.0	1.10	67.1
DFANet[33]	ImageNet	1 024×1 024	71.3	7.80	100.0
LRNNet[34]	No	512×1 024	72.2	0.68	71.0
AGLNet[35]	No	512×1 024	71.3	1.12	52.0
DDPNet[36]	No	768×1 536	74.0	2.52	85.4
CSRNet-light[37]	ResNet18	512×1 024	74.0	—	56.0
LETNet[38]	No	512×1 024	72.8	0.95	**150.0**
MLWP-Net	No	512×1 024	**74.1**	0.74	85.6

可以看出，在没有任何预训练的情况下，当输入图像分辨率为 512×1 024 时，MLWP-Net 仅以 0.74 MB 的参数量在 Cityscapes 数据集上实现了 74.1% 的分割精度；当输入图像分辨率为 360×480 时，MLWP-Net 仅以 0.74 MB 的参数量在 CamVid 数据集上实现了最高 68.2% 的分割精度，其分割效果优于其他网络，MLWP-Net 以更少的参数量实现了更高的分割精度。同时，作为轻量化网络，MLWP-Net 的推理速度也远远超过实时性分割网络的要求。

表 6　不同模型在 CamVid 测试集的实验结果

Method	Pretrain	Input Size	mIoU/%	Params/MB	Speed/fps
SegNet[3]	ImageNet	360×480	55.6	29.50	49.80
ENet[6]	No	360×480	51.3	0.36	105.70
LEDNet[8]	No	360×480	66.6	0.95	109.60
CGNet[9]	No	360×480	65.6	0.50	112.0
DABNet[10]	No	360×480	66.4	0.76	117.0
EDANet[29]	No	360×480	66.4	0.68	232.2
ESPNet[30]	No	360×480	55.6	**0.36**	**297.6**
DFANet[33]	No	720×960	64.7	7.80	120.0
LRNNet[34]	No	360×480	67.6	0.67	83.0
DDPNet[36]	No	360×480	67.3	1.10	—
MLWP-Net	No	360×480	**68.2**	0.74	95.0

此外，本文不仅对多个网络在 Cityscapes 数据集上进行分割精度、参数量和推理速度的对比，还列出了多个网络在 Cityscapes 测试集中每个类别对应的 mIoU，如表 7 所示。从表 7 中可以看出，相比其他网络，MLWP-Net 对于道路（Roa）、人行道（Sid）、墙面（Bui）等 14 种物体类别均达到最高的分割精度，这表明 MLWP-Net 利用其多连接的逐步特征融合模块 PFF 和多分支空洞卷积特征融合解码器 MPDCD 能够进一步加强对图像语义信息的提取能力和对图像中细小目标空间细节的恢复能力，MLWP-Net 对该数据集内的卡车和自行车等类别的分割效果相对现有最佳模型有微小的下降，但仍然具有可比拟的分割精度。

为了更加直观地探究和评价 MLWP-Net 网络对城市交通道路场景的分割效果，本文选择有代表性的轻量化实时性网络 CGNet、EDANet、ESNet、SQNet 与 MLWP-Net 的可视化分割效果进行比较与分析。本文在 Cityscapes 数据集中随机抽取了 4 张样本图像进行了分割效果的细节对比，如图 5 所示。图 5 第一行场景和第四行场景中，MLWP-Net 精确地定位柱子和交通灯的边界并进行分割，相比其他网络更加准确地实现了分割边界的连续性和分割的准确性，这得益于本文提出的多连接逐步特征融合瓶颈模块 PFF，有利于对图像边缘上下文信息进行准确提取；第二行场景中，其余网络模型虽然分割出卡车和人，但其交界处出现了误分类和分割不连续的问题，而 MLWP-Net 不仅能够正确地将不同类别分割，且不会受到不同类别的影响。这依赖于 MPDCD 解码器的多尺度特性，能够充分

利用不同尺度上下文信息进行类别分割，这对复杂场景下的语义分割有较大的提升效果。从图 5 第三行场景中可以看到远处的人物很小且人行道很长，太小或太长的目标都给分割增加了难度，与其他模型分割的细节对比可以看出，MLWP-Net 不仅能够分割出很长的人行道，对细小的人物也能够实现准确分割。随着图像分辨率下降，下采样操作容易造成信息丢失，而本文采用的 LWP 操作能尽可能减少信息丢失的同时完成下采样操作，从而实现对远处细小人物的准确分割。

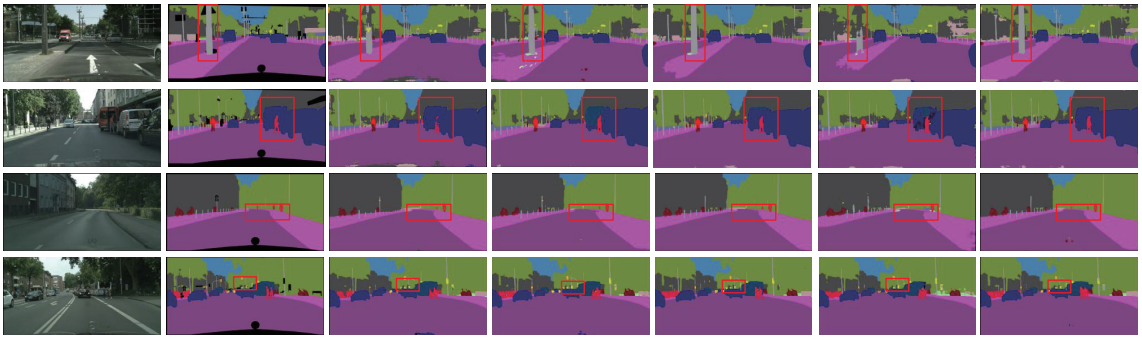


图 5　不同模型在 Cityscapes 数据集上的语义分割结果

表 7　不同模型在 Cityscapes 测试集上的预分类结果　　%

Method	Roa	Sid	Bui	Wal	Fen	Pol	TLi	TSi	Veg	Ter	Sky	Ped	Rid	Car	Tru	Bus	Tra	Mot	Bic	Class	Cat
SegNet[3]	96.4	73.2	84.0	28.4	29.0	35.7	39.8	45.1	87.0	63.8	91.8	62.8	42.8	89.3	38.1	43.1	44.1	35.8	51.9	57.0	79.1
ENet[6]	96.3	74.2	75.0	32.2	33.2	43.4	34.1	44.0	88.6	61.4	90.6	65.5	38.4	90.6	36.9	50.5	48.1	38.8	55.4	58.3	80.4
ERFNet[7]	97.2	80.0	89.5	41.6	45.3	56.4	60.5	64.6	91.4	68.7	94.2	76.1	56.4	92.4	45.7	60.6	27.0	48.7	61.8	66.3	85.2
LEDNet[8]	98.1	79.5	91.6	47.7	49.9	**62.8**	61.3	**72.8**	92.6	61.2	94.9	76.2	53.7	90.9	**64.4**	64.0	52.7	44.4	**71.6**	70.6	87.1
CGNet[9]	95.5	78.7	88.1	40.0	43.0	54.1	59.8	63.9	89.6	67.6	92.9	74.9	54.9	90.2	44.1	59.5	25.2	47.3	60.2	64.8	85.7
DABNet[10]	97.9	82.0	90.6	45.5	50.1	59.3	63.5	67.7	91.8	70.1	92.8	78.1	57.8	93.7	52.8	63.7	56.0	51.3	66.8	70.1	87.0
ESNet[26]	98.1	80.4	**92.4**	48.3	49.2	61.5	62.5	72.3	92.5	61.5	94.4	76.6	53.2	94.4	62.5	74.3	52.4	45.5	71.4	70.7	87.4
SQNet[27]	96.9	75.4	87.9	31.6	35.7	50.9	52.0	61.7	90.9	65.8	93.0	73.8	42.6	91.5	18.8	41.2	33.3	34.0	59.9	59.8	84.3
EDANet[29]	97.8	80.6	89.5	42.0	46.0	52.3	59.8	65.0	91.4	68.7	93.6	75.7	54.3	92.4	40.9	58.7	56.0	50.2	64.0	67.3	85.8
ESPNet[30]	97.0	77.5	76.2	35.0	36.1	45.0	35.6	46.3	90.8	63.2	92.6	67.0	40.9	92.3	38.1	52.5	50.1	41.8	57.2	60.3	82.2
LAANet[39]	97.9	82.9	91.0	47.5	51.5	59.3	66.0	70.3	92.3	69.9	94.7	81.8	61.4	94.2	58.6	74.5	55.1	54.3	69.4	73.6	88.4
MLWP-Net	**98.1**	**83.4**	91.7	**55.4**	**52.5**	62.1	**67.1**	71.8	**92.7**	**70.0**	**95.0**	**83.1**	**63.3**	**94.7**	60.2	**75.7**	**62.5**	**56.9**	71.3	**74.1**	**89.0**

综合以上可视化结果分析，本文所提 MLWP-Net 能够尽可能实现分割边缘的准确性和连续性，进而实现对细小目标和连续大目标的准确分割，取得了较好的分割效果。

3　结束语

本文提出了一种渐进式特征融合与低频混合小波池化结合的轻量化语义分割网络 MLWP-Net，解决了现有语义分割网络中存在的特征信息提取不足和网络参数量较大等问题。一方面，在编码器端主要设计了轻量化的多连接逐步特征融合 PFF 模块和通用型的低频混合小波池化 LWP 操作，应用前者实现了上下文信息的有效聚合，从而高效地提取图像特征；应用后者解决了现有网络中下采样操作导致的特征信息丢失问题，高效地完成下采样操作，并可插入其他分割网络中进行下采样操作。另一方面，提出了多分支空洞卷积特征融合 MPDCD 解码器，有效结合多尺度上下文特征实现图像空间信息的高效恢复。

与现存流行的实时语义分割网络对比，MLWP-Net 在保证高精度的前提下，大幅度减少了模型参数量，对移动终端领域有很好的应用前景，尤其适用于对准确性和时效性要求较高的自动驾驶中的道路场景分割任务中。

参考文献

[1] PENG B. Research on operation stability evaluation of industrial automation system based on improved deep learning[J]. International Journal of Manufacturing

Technology and Management, 2022, 36(2/3/4): 141.
[2] 孔令军, 王茜雯, 包云超, 等. 基于深度学习的医疗图像分割综述[J]. 无线电通信技术, 2021, 47(2): 121-130.
KONG L J, WANG Q W, BAO Y C, et al. A survey on medical image segmentation based on deep learning[J]. Radio Communications Technology, 2021, 47(2): 121-130.
[3] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Trans Pattern Anal Mach Intell, 2017, 39(12): 2481-2495.
[4] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//European Conference on Computer Vision. Cham: Springer, 2018: 833-851.
[5] LIN G S, MILAN A, SHEN C H, 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.
[6] PASZKE A, CHAURASIA A, KIM S, et al. ENet: A deep neural network architecture for real-time semantic segmentation[EB/OL]. [2023-05-21]. https://arxiv.org/abs/1606.02147.
[7] 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, 2018, 19(1): 263-272.
[8] WANG Y, ZHOU Q, LIU J, et al. Lednet: A lightweight encoder-decoder network for real-time semantic segmentation[C]//Proceedings of the IEEE International Conference on Image Processing. New York: IEEE, 2019: 1860-1864.
[9] WU T, TANG S, ZHANG R, et al. CGNet: A light-weight context guided network for semantic segmentation[J]. IEEE Trans Image Process, 2021, 30: 1169-1179.
[10] LI G, YUN I, KIM J, et al. DABNet: Depth-wise asymmetric bottleneck for real-time semantic segmentation[EB/OL]. [2023-05-22]. https://arxiv.org/pdf/1907.11357.pdf.
[11] LU M X, CHEN Z X, JONATHAN WU Q M, et al. FRNet: Factorized and regular blocks network for semantic segmentation in road scene[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(4): 3522-3530.
[12] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 1251-1258.
[13] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. [2023-05-22]. https://arxiv.org/pdf/1704.04861.pdf.
[14] ZHANG X Y, ZHOU X Y, LIN M X, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 6848-6856.
[15] 马宇, 张丽果, 杜慧敏, 等. 卷积神经网络的交通标志语义分割[J]. 计算机科学与探索, 2021, 15(6): 1114-1121.
MA Y, ZHANG L G, DU H M, et al. Traffic sign semantic segmentation based on convolutional neural network[J]. Journal of Frontiers of Computer Science and Technology, 2021, 15(6): 1114-1121.
[16] SPRINGENBERG J T, DOSOVITSKIY A, BROX T, et al. Striving for simplicity: The all convolutional net[EB/OL]. [2023-05-21]. https://arxiv.org/pdf/1412.6806.pdf.
[17] JAMALI A. Comparing the performance and application of wavelet transform in digital image processing segmentation[EB/OL]. [2023-05-22]. http://dx.doi.org/10.2139/ssrn.4554509.
[18] RAMAMONJISOA M, FIRMAN M, WATSON J, et al. Single image depth prediction with wavelet decomposition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2021: 11089-11098.
[19] LIU P J, ZHANG H Z, ZHANG K, et al. Multi-level wavelet-CNN for image restoration[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New York: IEEE, 2018: 773-782.
[20] XUE S K, QIU W Y, LIU F, et al. Wavelet-based residual attention network for image super-resolution[J]. Neurocomputing, 2020, 382: 116-126.
[21] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 770-778.
[22] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 2117-2125.
[23] 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.
[24] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 3213-3223.
[25] BROSTOW G J, FAUQUEUR J, CIPOLLA R. Semantic object classes in video: A high-definition ground truth database[J]. Pattern Recognition Letters, 2009, 30(2): 88-97.
[26] WANG Y, ZHOU Q, XIONG J, et al. ESNet: An efficient symmetric network for real-time semantic segmentation [M]//Pattern Recognition and Computer Vision. Cham: Springer International Publishing, 2019.
[27] TREML M, ARJONA-MEDINA J A, UNTERTHINER T, et al. Speeding up semantic segmentation for autonomous driving[EB/OL] . [2023-05-21]. https://www.researchgate.net/publication/309935608_Speeding_up_Semantic_

Segmentation_for_Autonomous_Driving.

[28] YU C Q, GAO C X, WANG J B, et al. BiSeNet V2: Bilateral network with guided aggregation for real-time semantic segmentation[J]. International Journal of Computer Vision, 2021, 129(11): 3051-3068.

[29] LO S Y, HANG H M, CHAN S W, et al. Efficient dense modules of asymmetric convolution for real-time semantic segmentation[C]//Proceedings of the Proceedings of the 1st ACM International Conference on Multimedia in Asia. New York: ACM, 2019: 1-6.

[30] MEHTA S, RASTEGARI M, CASPI A, et al. ESPNet: Efficient spatial pyramid of dilated convolutions for semantic segmentation[M]//Computer Vision – ECCV 2018. Cham: Springer International Publishing, 2018.

[31] POUDEL R P K, BONDE U, LIWICKI S, et al. Contextnet: Exploring context and detail for semantic segmentation in real-time[EB/OL]. [2023-05-21]. https://arvix.org/abs/1805.04554.

[32] POUDEL R P K, LIWICKI S, CIPOLLA R. Fast-scnn: Fast semantic segmentation network[EB/OL]. [2023-05-21]. https://arvix.org/abs/1902.04502.

[33] LI H C, XIONG P F, FAN H Q, et al. DFANet: Deep feature aggregation for real-time semantic segmentation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 9522-9531.

[34] JIANG W H, XIE Z Z, LI Y Y, et al. LRNNET: A light-weighted network with efficient reduced non-local operation for real-time semantic segmentation[C]// Proceedings of the IEEE International Conference on Multimedia & Expo Workshops. New York: IEEE, 2020: 1-6.

[35] ZHOU Q, WANG Y, FAN Y W, et al. AGLNet: Towards real-time semantic segmentation of self-driving images via attention-guided lightweight network[J]. Applied Soft Computing, 2020, 96: 106682.

[36] GE R J, HE Y T, XIA C, et al. DDPNet: A novel dual-domain parallel network for low-dose CT reconstruction [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2022: 748-757.

[37] XIONG J J, PO L M, YU W Y, et al. CSRNet: Cascaded selective resolution network for real-time semantic segmentation[J]. Expert Systems with Applications, 2023, 211: 118537.

[38] XU G A, LI J C, GAO G W, et al. Lightweight real-time semantic segmentation network with efficient transformer and CNN[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(12): 15897-15906.

[39] ZHANG X L, DU B C, WU Z Y, et al. LAANet: Lightweight attention-guided asymmetric network for real-time semantic segmentation[J]. Neural Computing and Applications, 2022, 34(5): 3573-3587.

编　辑　税　红