文章编号： 1671-7449(2024)06-0575-11

去伪影空洞残差与双向注意力金字塔增强的图像级联语义分割算法

王国刚，罗香莲

（山西大学 物理电子工程学院，山西 太原 030006）

摘　要： 针对ICNet空洞卷积引入网格伪影，金字塔池化模块难以充分融合不同尺度特征层间的上下文信息，同一尺度特征和跨尺度特征均交互不足的问题，提出去伪影空洞残差与双向注意力金字塔增强的图像级联语义分割算法。该算法提出去伪影空洞残差网络，消除空洞卷积引入的网格伪影，加强网络的表征能力；构建双向金字塔池化模块，沿自顶向下的路径将语义信息传递到低层，并沿自底向上的路径将位置信息传递到高层；构造注意力金字塔增强模块，创建自注意力、向上注意力、向下注意力机制，建立跨空间与跨尺度的特征交互。Cityscapes数据集上的实验结果表明，所提算法的总体精确率、平均像素精确度、平均交并比分别比ICNet提高了0.67%、5.58%、4.40%，而且该算法的以上3个评价指标均优于其它8种主流比对算法。

关键词： 深度学习；语义分割；ICNet；双向金字塔；注意力金字塔增强

中图分类号： TP391.4　　**文献标识码：** A　　**doi:** 10.3969/j.issn.1671-7449.2024088

引用格式： 王国刚，罗香莲. 去伪影空洞残差与双向注意力金字塔增强的图像级联语义分割算法[J]. 测试技术学报，2024，38(6)：575-585.

WANG Guogang，LUO Xianglian. Image cascade semantic segmentation algorithm based on deartifacting dilated residual and bidirectional attention pyramid enhancement[J]. Journal of Test and Measurement Technology，2024，38(6)：575-585.

Image Cascade Semantic Segmentation Algorithm Based on Deartifacting Dilated Residual and Bidirectional Attention Pyramid Enhancement

WANG Guogang，LUO Xianglian

（College of Physics and Electronic Engineering，Shanxi University，Taiyuan 030006，China）

Abstract： A novel algorithm named as image cascade semantic segmentation algorithm based on deartifacting dilated residual and bidirectional attention pyramid enhancement is proposed to avoid the disadvantages of an introduction of gridding artifacts for dilated convolutions，the difficulty in fully integrating context information among different scale feature layers for the pyramid pooling module，and lacking sufficient interaction for the same scale features and the cross-scale features in ICNet algorithm. The characterization ability of the network is strengthened since the deartifacting dilated residual network is constructed to remove gridding artifacts

收稿日期： 2024-03-05

基金项目： 国家自然科学基金资助项目(11804209)；山西省自然科学基金资助项目(201901D111031，201901D211173)；山西省高校科技创新计划(2019L0064，2020L0051)

作者简介： 王国刚(1977—)，男，副教授，博士，主要从事图像处理与视频分析研究。E-mail：kingguogang@sxu.edu.cn。

introduced by dilated convolutions. After building the bidirectional pyramid pooling module, the semantic information is transmitted to the lower level along a top-down path while the location information is conveyed to the higher level along a bottom-up path. Moreover, the feature interactions for both cross-space and cross-scale are established by the attention pyramid enhancement module, which introduces self-attention, upward attention, and downward attention mechanism. On the Cityscapes dataset, the experimental results show that the overall accuracy, the mean pixel accuracy and the mean intersection over union of the proposed algorithm have been improved by 0.67%, 5.58% and 4.40% respectively compared to ICNet. Compared with the state-of-the-art methods, the presented method achieves outperformance on the above evaluation indexes.

Key words: deep learning; semantic segmentation; ICNet; bidirectional pyramid; attention pyramid enhancement

0 引 言

作为计算机视觉领域的一项关键技术，语义分割的目的是预测图像中每个像素点对应的语义类别标签。目前，语义分割在遥感测绘[1-3]、无人驾驶[4-7]、医疗影像分析[8-10]等领域得到了广泛应用。

基于机器学习的图像分割方法主要包括基于边缘、阈值、区域、聚类和图论的分割方法。该类方法通过人工经验提取图像特征，复杂场景下分割精度低，难以满足实际场景中多语义类别的需求。近年来，卷积神经网络在计算机视觉(Computer Vision，CV)领域表现出了强大的特征学习能力。2015 年，Long 等[11]提出全卷积神经网络(Fully Convolutional Networks，FCN)，该网络将深度学习引入图像语义分割领域，通过端到端的方式构建语义分割模型。

当前主流语义分割算法大多采用类似FCN的框架，大致可分为两类：一类算法使用编解码结构，通过编码器提取语义信息，再经解码器恢复空间分辨率，如SegNet[12]和GCN(Global Convolutional Network)[13]算法，该类算法一定程度上弥补了空间位置信息的损失，但缺乏足够的上下文信息，导致分割精度较低；另一类算法通过池化或卷积形式的金字塔获取多尺度上下文信息以提高分割准确率，如PSPNet[14]和OCNet[15]算法，该类算法往往模型复杂，推理速度慢。

ICNet[16]通过低分辨率分支获得语义信息，并增加中、高分辨率分支以获取细节信息，这在保证分割性能的同时，提高了推理速度。然而，ICNet空洞卷积引入网格伪影，金字塔池化模块难以充分融合不同尺度特征层间的上下文信息，同一尺度特征和跨尺度特征均交互不足。

针对以上问题，提出去伪影空洞残差与双向注意力金字塔增强的图像级联语义分割算法(Image Cascade Semantic Segmentation Algorithm Based on Deartifacting Dilated Residual and Bidirectional Attention Pyramid Enhancement，IDB)。该算法构建去伪影空洞残差网络，去除空洞卷积引入的网格伪影，提高网络的表征能力；构造双向金字塔池化模块，引入自顶而下和自底向上的路径以加强不同尺度特征层之间上下文信息的融合；提出注意力金字塔增强模块，创建自注意力、向上注意力、向下注意力机制，实现跨空间与跨尺度的特征交互。实验结果表明，IDB 算法在Cityscapes数据集上的总体精确率、平均像素精确度、平均交并比均优于其它8种主流对比算法。

1 相关模型

作为一种典型的实时语义分割模型，ICNet在PSPNet的基础上采用低、中、高三分辨率分支策略来降低时间成本，其结构如图 1 所示。ICNet的低分辨率分支利用PSPNet获得粗略特征，中分辨率分支与高分辨率分支使用少量卷积层获取细节特征。金字塔池化模块(Pyramid Pooling Module，PPM)和级联特征融合模块(Cascade Feature Fusion，CFF)融合各支路特征，再经预测网络得到语义分割图。

2 IDB方法

2.1 去伪影空洞残差网络

ICNet在ResNet50的后两个模块res-4和res-5中分别连续采用扩张率为2和4的空洞卷积。设 F 为特征图，K 为卷积核，d 为扩张率，则空洞卷

积可定义为

$$(F*K)(p)=\sum_{a+db=p}F(a)K(b), \qquad (1)$$

式中：b、a和p分别为卷积层、输入特征图和输出特征图中的点。

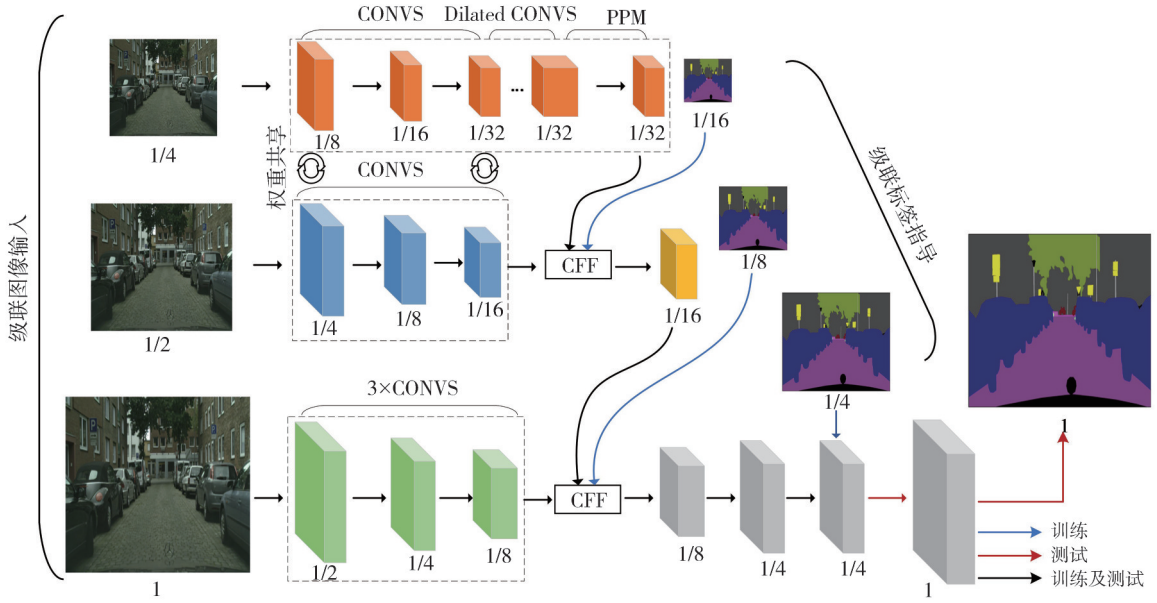

图 1　ICNet网络架构

Fig. 1　ICNet network structure

空洞卷积在卷积核的相邻权值之间填充了零值，这扩充了特征图上卷积核的计算范围，从而提取了更大区域的特征信息。

虽然空洞卷积可以扩大感受野，但有可能会引入网格伪影，如图 2 所示。图 2(a) 绿色部分表示输入的单个像素，图 2(b) 表示扩张率为2的空洞卷积，图 2(c) 展示了网格伪影。ResNet50 在7×7卷积层后采用最大池化层。最大池化层导致的高幅高频效应向后续特征层传播，这使网格伪影问题更严重。

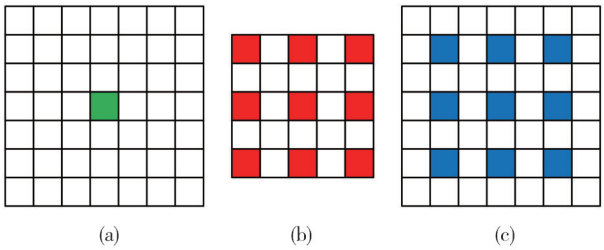

图 2　网格伪影

Fig. 2　A gridding artifact

为消除网格效应，提高分割精度，IDB采用去伪影空洞残差网络(Deartifacting Dilated Residual Network，DDRN)提取特征。若s、d、n、c分别表示步长、扩张率、重复次数和输出通道数，则DDRN的网络结构可由表 1 表示。DDRN将最大池化层替换为两个3×3的卷积层以消除高幅高频效应。此外，DDRN在网络末端添加扩张率为2和1的两个卷积层以滤除混叠伪影。

表 1　去伪影空洞残差网络结构

Tab. 1　Deartifacting dilated residual network structure

层号	操作	c	n	d	s
1	Conv7×7	16	1	1	1
2	Conv3×3	16	1	1	1
3	Conv3×3	32	1	1	2
4	Bottleneck	256	3	1	2
5	Bottleneck	512	4	1	2
6	Bottleneck	1 024	6	2	1
7	Bottleneck	2 048	3	4	1
8	Conv3×3	512	1	2	1
9	Conv3×3	512	1	1	1

通过替换最大池化层和添加不同扩张率的卷积层，DDRN消除了网格伪影，加强了网络的表征能力，提升了IDB模型的分割性能。

2.2　双向金字塔池化

PPM拼接各池化特征图可获取不同区域的特征信息，但缺乏足够的轮廓细节，而且不同尺度特征图之间的上下文信息交互不足。针对此问题，IDB构建双向金字塔池化模块(Bidirectional Pyramid Pooling Module，BPPM)，如图 3 所示。该模块的基本单元BPPM层，沿自顶向下的路径将语义信息传递到低层，并沿自底向上的路径将位置信息传递到高层。这使不同尺度特征间的语

义信息和位置信息得到了充分融合，提高了 IDB 算法的分割准确率。

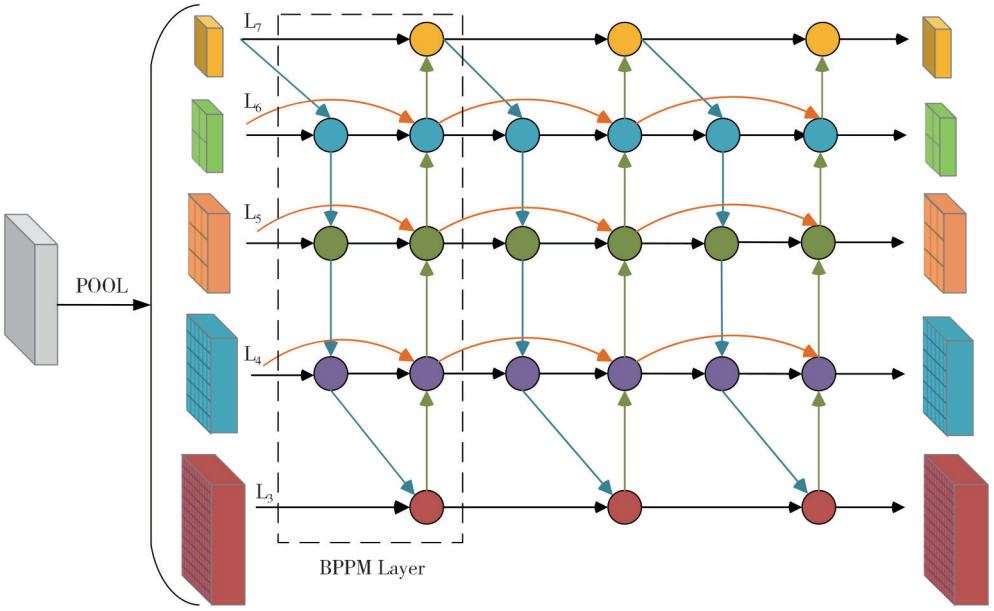

图 3 双向金字塔池化模块

Fig. 3 Bidirectional pyramid pool module

基于不同尺度子区域，BPPM 对输入特征图进行全局平均池化得到 12×12，6×6，3×3，2×2，1×1 的池化特征图 L_3、L_4、L_5、L_6、L_7。将 L_3、L_4、L_5、L_6、L_7 作为输入，第一个 BPPM 层沿自上而下的路径得到中间特征层 L_i^{td}，再自下而上逐层下采样并融合同层特征以得到输出特征层 L_j^{out}，如式(2)～(6)所示。

$$L_6^{\mathrm{td}}=Conv_{3\times3}\left(\frac{\omega_{6,\mathrm{in}}\cdot L_6+\omega_{7,\mathrm{in}}\cdot \mathrm{Re}\,size_{\mathrm{up}}(L_7)}{\omega_{6,\mathrm{in}}+\omega_{7,\mathrm{in}}+\varepsilon}\right),\tag{2}$$

$$L_i^{\mathrm{td}}=Conv_{3\times3}\left(\frac{\omega_{i,\mathrm{in}}\cdot L_i+\omega_{i+1,\mathrm{td}}\cdot \mathrm{Re}\,size_{\mathrm{up}}(L_{i+1}^{\mathrm{td}})}{\omega_{i,\mathrm{in}}+\omega_{i+1,\mathrm{td}}+\varepsilon}\right),\tag{3}$$

式中：$\omega_{i,\mathrm{in}}$，$\omega_{i,\mathrm{td}}$ 分别为第 i 输入层和中间层权重；$\mathrm{Re}\,size_{\mathrm{up}}(\cdot)$ 和 $Conv_{3\times3}(\cdot)$ 分别表示上采样和 3×3 卷积；ε 为 0.000 1；$i\in\{3,4,5\}$。

$$L_3^{\mathrm{out}}=L_3^{\mathrm{td}},\tag{4}$$

$$L_j^{\mathrm{out}}=Conv_{3\times3}\left(\frac{\omega'_{j,\mathrm{in}}\cdot L_j+\omega'_{j,\mathrm{td}}\cdot L_j^{\mathrm{td}}+\omega'_{j-1,\mathrm{out}}\cdot \mathrm{Re}\,size_{\mathrm{down}}(L_{j-1}^{\mathrm{out}})}{\omega'_{j,\mathrm{in}}+\omega'_{j,\mathrm{td}}+\omega'_{j-1,\mathrm{out}}+\varepsilon}\right),\tag{5}$$

$$L_7^{\mathrm{out}}=Conv_{3\times3}\left(\frac{\omega'_{7,\mathrm{in}}\cdot L_7+\omega'_{6,\mathrm{out}}\cdot \mathrm{Re}\,size_{\mathrm{down}}(L_6^{\mathrm{out}})}{\omega'_{7,\mathrm{in}}+\omega'_{6,\mathrm{out}}+\varepsilon}\right),\tag{6}$$

式中：$\omega'_{j,\mathrm{in}}$，$\omega'_{j,\mathrm{td}}$，$\omega'_{j,\mathrm{out}}$ 分别为第 j 输入层、中间层、输出层权重；$\mathrm{Re}\,size_{\mathrm{down}}(\cdot)$ 表示下采样；$j\in\{4,5,6\}$。

BPPM 包含 3 个 BPPM 层。与首个 BPPM 层操作类似，BPPM 依次经第 2、3 层后得到输出特征图。

IDB 构建双向金字塔池化模块，该模块引入池化特征图，并双向加权以融合多尺度特征。这获取了更多轮廓细节，增强了不同尺度特征图之间上下文信息的交互。

2.3 注意力金字塔增强模块

PPM 通过自适应全局平均池化得到不同尺度的特征图，但同一尺度特征和跨尺度特征均交互不足。针对此问题，构建注意力金字塔增强模块（Attention Pyramid Enhancement，APE）。该模块创建自注意力（Self-Attention，SA）、向下注意力（Downward Attention，DA）和向上注意力（Upward Attention，UA）3 种机制。SA 捕捉同一尺度特征图上共现的目标特征；DA 将高层特征图中的语义信息传递到低层；UA 把低层特征图中的位置信息融入到高层。

2.3.1 自注意力机制

$\boldsymbol{X}'$、$\boldsymbol{X}$ 分别为输入特征图和调整尺寸后的输入特征图，设 $\boldsymbol{X}'\in\mathrm{R}^{C\times H\times W}$、$\boldsymbol{X}\in\mathrm{R}^{C\times HW}$。若 $C'\times C$ 的可学习矩阵 $\boldsymbol{W}_q$、$\boldsymbol{W}_k$、$\boldsymbol{W}_v$ 分别表示查询、键、值的权重矩阵，则查询、键、值矩阵可分别表示为 $\boldsymbol{Q}=\boldsymbol{W}_q\boldsymbol{X}$、$\boldsymbol{K}=\boldsymbol{W}_k\boldsymbol{X}$、$\boldsymbol{V}=\boldsymbol{W}_v\boldsymbol{X}$。

令

$$\boldsymbol{Q}=(\boldsymbol{q}_i)=\begin{pmatrix}\boldsymbol{q}_1\\ \vdots\\ \boldsymbol{q}_{\frac{C'}{N}}\\ \vdots\\ \boldsymbol{q}_{\frac{C'}{N}\times(n-1)+1}\\ \vdots\\ \boldsymbol{q}_{\frac{C'}{N}\times n}\\ \vdots\\ \boldsymbol{q}_{\frac{C'}{N}\times(N-1)+1}\\ \vdots\\ \boldsymbol{q}_{C'}\end{pmatrix}=\begin{pmatrix}\boldsymbol{Q}_1\\ \vdots\\ \boldsymbol{Q}_n\\ \vdots\\ \boldsymbol{Q}_N\end{pmatrix},$$

$$\boldsymbol{K}=(\boldsymbol{k}_j)=\begin{pmatrix}\boldsymbol{k}_1\\ \vdots\\ \boldsymbol{k}_{\frac{C'}{N}}\\ \vdots\\ \boldsymbol{k}_{\frac{C'}{N}\times(n-1)+1}\\ \vdots\\ \boldsymbol{k}_{\frac{C'}{N}\times n}\\ \vdots\\ \boldsymbol{k}_{\frac{C'}{N}\times(N-1)+1}\\ \vdots\\ \boldsymbol{k}_{C'}\end{pmatrix}=\begin{pmatrix}\boldsymbol{K}_1\\ \vdots\\ \boldsymbol{K}_n\\ \vdots\\ \boldsymbol{K}_N\end{pmatrix}。\tag{7}$$

则归一化的注意力图可由式(8)给出。从而可得SA的输出特征图，如式(9)～(10)所示。

$$f_{\mathrm{SA}}(\boldsymbol{Q},\boldsymbol{K},N)=(\pi_1,\cdots,\pi_n,\cdots,\pi_N)\cdot\left(softmax\begin{bmatrix}\boldsymbol{Q}_1^{\mathrm{T}}\boldsymbol{K}_1\\ \vdots\\ \boldsymbol{Q}_n^{\mathrm{T}}\boldsymbol{K}_n\\ \vdots\\ \boldsymbol{Q}_N^{\mathrm{T}}\boldsymbol{K}_N\end{bmatrix}\right),\tag{8}$$

$$\boldsymbol{X}_{ss}'=f_{\mathrm{SA}}(\boldsymbol{Q},\boldsymbol{K},N)\cdot\boldsymbol{V}^{\mathrm{T}},\tag{9}$$

$$\boldsymbol{X}_{ss}=reshape(\boldsymbol{W}_s\cdot(\boldsymbol{X}_{ss}')^{\mathrm{T}}),\tag{10}$$

式中：π_n为第n个可学习的聚合权重；$\boldsymbol{W}_s$为可学习的$C\times C'$权重矩阵；$\boldsymbol{X}_{ss}\in\mathrm{R}^{C\times H\times W}$。

2.3.2　向下注意力机制

$\boldsymbol{X}_t''$表示输入的高层特征图，$\boldsymbol{X}_t''$上采样后的特征图为$\boldsymbol{X}_t'$，$\boldsymbol{X}_t'$调整尺寸后的特征图为$\boldsymbol{X}_t$。$\boldsymbol{X}_d'$表示输入的低层特征图，$\boldsymbol{X}_d'$调整尺寸后的特征图为$\boldsymbol{X}_d$。设$\boldsymbol{X}_t''\in\mathrm{R}^{C\times H_t\times W_t}$、$\boldsymbol{X}_t'\in\mathrm{R}^{C\times H_d\times W_d}$、$\boldsymbol{X}_t\in\mathrm{R}^{C\times H_dW_d}$、$\boldsymbol{X}_d'\in\mathrm{R}^{C\times H_d\times W_d}$、$\boldsymbol{X}_d\in\mathrm{R}^{C\times H_dW_d}$。若查询、键、值的权重矩阵分别由可学习的$C'\times C$的矩阵$\boldsymbol{W}_q$、$\boldsymbol{W}_k$、$\boldsymbol{W}_v$表示，则查询、键、值矩阵可分别由$\boldsymbol{Q}=\boldsymbol{W}_q\boldsymbol{X}_d$、$\boldsymbol{K}=\boldsymbol{W}_k\boldsymbol{X}_t$、$\boldsymbol{V}=\boldsymbol{W}_v\boldsymbol{X}_t$表示。

一般而言，不同尺度的特征图提取不同的上下文信息或语义信息。特征图语义信息不同时，负欧式距离表达相似度比点积更有效[17]。对$\boldsymbol{Q}$、$\boldsymbol{K}$采用式(7)所示的分块，使用负欧式距离进行相似度计算，可得归一化的注意力图，如式(11)所示。于是可得DA的输出特征图，如式(12)～式(13)所示。

$$f_{DA}(\boldsymbol{Q},\boldsymbol{K},N)=(\pi_1,\cdots,\pi_n,\cdots,\pi_N)\cdot\left(softmax\begin{bmatrix}-\|\boldsymbol{Q}_1^{\mathrm{T}}-\boldsymbol{K}_1^{\mathrm{T}}\|^2\\ \vdots\\ -\|\boldsymbol{Q}_n^{\mathrm{T}}-\boldsymbol{K}_n^{\mathrm{T}}\|^2\\ \vdots\\ -\|\boldsymbol{Q}_N^{\mathrm{T}}-\boldsymbol{K}_N^{\mathrm{T}}\|^2\end{bmatrix}\right),\tag{11}$$

$$\boldsymbol{X}_{sd}'=f_{\mathrm{DA}}(\boldsymbol{Q},\boldsymbol{K},N)\cdot\boldsymbol{V}^{\mathrm{T}},\tag{12}$$

$$\boldsymbol{X}_{sd}=reshape(\boldsymbol{W}_d\cdot(\boldsymbol{X}_{sd}')^{\mathrm{T}}),\tag{13}$$

式中：π_n为第n个可学习的聚合权重；$\boldsymbol{W}_d$为可学习的$C\times C'$权重矩阵；$\boldsymbol{X}_{sd}\in\mathrm{R}^{C\times H_d\times W_d}$。

2.3.3　向上注意力机制

$\boldsymbol{K}$、$\boldsymbol{V}$均表示输入的低层特征图，$\boldsymbol{Q}$表示输入的高层特征图，$\boldsymbol{Q}_i$为$\boldsymbol{Q}$的第i个通道。设$\boldsymbol{Q}\in\mathrm{R}^{C\times H_t\times W_t}$，$\boldsymbol{K}\in\mathrm{R}^{C\times H_d\times W_d}$，$\boldsymbol{V}\in\mathrm{R}^{C\times H_d\times W_d}$。$\boldsymbol{K}$经全局平均池化得到权重$\boldsymbol{W}$，如式(14)所示。$\boldsymbol{W}$的第$i$个分量$W_i$与$\boldsymbol{Q}_i$数乘、拼接、卷积后得到$\boldsymbol{Q}_w$。$\boldsymbol{V}$下采样后与$\boldsymbol{Q}_w$求和，再经3×3卷积得到UA的输出$\boldsymbol{X}_{st}$，见式(15)～式(16)。

$$\boldsymbol{W}=GAP(\boldsymbol{K}),\tag{14}$$

$$\boldsymbol{Q}_w=Conv_{3\times3}(concat(W_i\cdot\boldsymbol{Q}_i)),\tag{15}$$

$$\boldsymbol{X}_{st}=Conv_{3\times3}(\boldsymbol{Q}_w+SConv_{3\times3}(\boldsymbol{V})),\tag{16}$$

式中：$\boldsymbol{W}\in\mathrm{R}^{C\times1}$，$\boldsymbol{Q}_w$、$\boldsymbol{X}_{st}\in\mathrm{R}^{C\times H_t\times W_t}$；$Conv_{3\times3}$和$SConv_{3\times3}$分别表示3×3卷积和带步长的3×3卷积。

2.3.4　注意力金字塔增强

如图4所示，$\boldsymbol{X}_1$、$\boldsymbol{X}_2$、$\boldsymbol{X}_3$、$\boldsymbol{X}_4$、$\boldsymbol{X}_5$表示APE模块的输入特征图。$\boldsymbol{X}_1$经SA、UA作用后得$\boldsymbol{Y}_1$，$\boldsymbol{X}_5$经SA、DA作用后得$\boldsymbol{Y}_5$。$\boldsymbol{X}_2$、$\boldsymbol{X}_3$、$\boldsymbol{X}_4$分别经SA、DA、UA作用后得$\boldsymbol{Y}_2$、$\boldsymbol{Y}_3$、$\boldsymbol{Y}_4$。从下到上，依次堆叠$\boldsymbol{Y}_1$、$\boldsymbol{Y}_2$、$\boldsymbol{Y}_3$、$\boldsymbol{Y}_4$、$\boldsymbol{Y}_5$中的第i层得到$\boldsymbol{Z}_i$。自上而下依次拼接$\boldsymbol{X}_i$与$\boldsymbol{Z}_i$得到$\boldsymbol{C}_i$，$\boldsymbol{C}_i$经卷积得输出特征图$\boldsymbol{O}_i$。

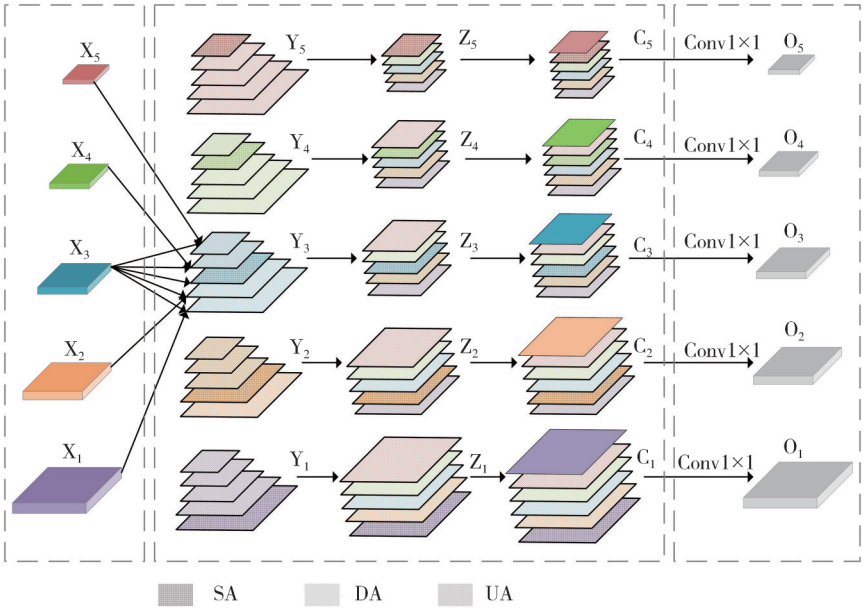

图 4 APE模块

Fig. 4 APE module

APE创建SA、DA、UA机制以建立跨空间与跨尺度的特征交互，这丰富了特征图的上下文信息，提高了模型的分割性能。

2.4 IDB网络模型

IDB网络包含低、中、高3个分辨率分支，如图5所示。低、中分辨率分支构建DDRN网络，低分辨率分支创建BPPM和APE模块。DDRN消除空洞卷积引入的网格伪影，提高网络的表征能力；BPPM引入自顶而下和自底向上的路径，加强多尺度特征的融合；APE构造自注意力、向上注意力、向下注意力机制，建立跨空间与跨尺度的特征交互。

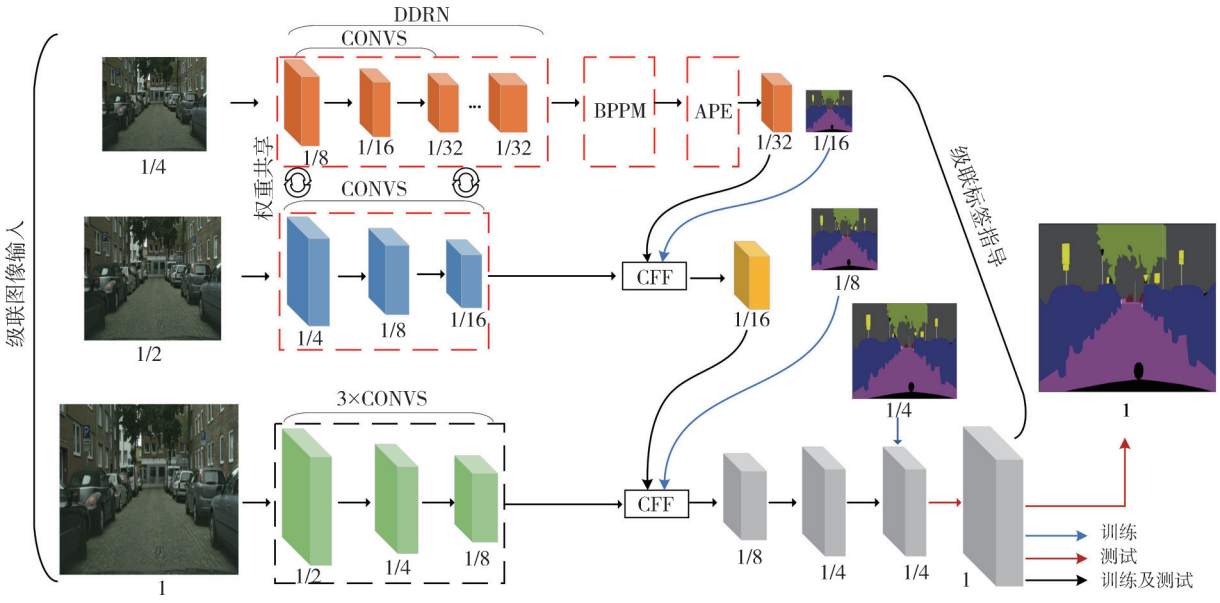

图 5 IDB网络模型

Fig. 5 IDB network model

3 实验结果与分析

3.1 数据集与评价指标

实验采用Cityscapes数据集评估模型性能。该数据集包含2 975幅训练图像、500幅验证图像和1 525幅测试图像，共有19个语义类别。实验采用的测试集为验证集。

评价指标为总体精确率[18]（Overall Accuracy，OA）、平均像素精确度（Mean Pixel Accuracy，

MPA）、平均交并比（Mean Intersection over Union，MIoU）、频率权重交并比（Frequency Weighted Intersection over Union，FWIoU）和FPS。各评价指标的数学表达式如式（17）～式（21）所示。

$$OA=\frac{\sum_{i=0}^{k}p_{ii}}{\sum_{i=0}^{k}\sum_{j=0}^{k}p_{ij}},\tag{17}$$

$$MPA=\frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}},\tag{18}$$

$$MIoU=\frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}},\tag{19}$$

$$FWIoU=\frac{1}{\sum_{i=0}^{k}\sum_{j=0}^{k}p_{ij}}\sum_{i=0}^{k}\frac{\sum_{j=0}^{k}p_{ij}p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}},\tag{20}$$

式中：$k+1$为类别总数；p_{ij}表示实际类别为i的像素被标记为类别j的像素数量。

$$FPS=\frac{N}{\sum_{j=1}^{N}T_j},\tag{21}$$

式中：N为图像数量；T_j为处理第j幅图像所用的时间。

3.2　实验设置

硬件配置：Intel（R）Xeon（R）Platinum CPU，NVIDIA GeForce RTX 3090 GPU。深度学习框架为pytorch1.10，GPU加速库为CUDA11.3和Cudnn8.2.0。

训练阶段，选取SGD（Stochastic Gradient Descent）优化器更新参数。动量和权重衰减分别设置为0.9和0.000 1。初始学习率为0.01，利用poly策略动态调整学习率。批处理大小为8，epoch设置为200。

3.3　定量分析

为评估IDB算法性能，在Cityscapes数据集上与ENet[19]，CGNet[20]，LEDNet[21]，ICNet[16]，DUNet[22]，BiSeNet[23]，EncNet[24]，OCNet[15]8种主流算法进行对比，各类别IoU如表2所示。由表2可知，IDB算法6个目标类别的IoU值为最优，4个目标类别的IoU值为次优。且IDB算法的MIoU值为72.1%，分别比ENet、CGNet、LEDNet、ICNet、DUNet、BiSeNet、EncNet、OCNet算法的MIoU值提高了13.75、8.19、7.06、3.04、2.84、2.39、1.26、1.11百分点。

表2　9种算法的各类别交并比

Tab.2　Each class IoU of nine algorithms

类别	*IoU*/%								
	ENet	CGNet	LEDNet	ICNet	DUNet	BiSeNet	EncNet	OCNet	IDB
road	96.92	97.02	97.13	97.32	96.95	97.52	96.84	96.98	**97.70**
swalk	77.11	77.56	77.80	79.13	80.02	81.03	79.14	79.69	**82.08**
build	88.85	89.39	89.24	89.64	89.96	90.37	90.25	90.23	**91.11**
wall	33.64	40.95	46.60	50.30	42.82	43.09	32.54	34.71	**52.98**
fence	41.30	46.38	47.78	48.45	48.70	**50.92**	49.62	50.33	50.57
pole	52.03	52.12	50.94	48.51	**62.78**	53.28	58.47	57.05	54.94
tlight	42.78	53.55	49.93	53.37	67.89	60.36	**68.66**	67.24	59.49
sign	61.62	63.77	61.77	64.53	**76.01**	70.38	75.01	74.60	72.43
veg	90.46	90.40	90.35	90.89	**91.75**	91.29	91.62	91.48	91.61
terrain	55.64	57.14	55.37	60.53	59.47	60.06	**61.72**	60.09	59.80
sky	93.74	93.23	92.75	93.42	93.20	93.87	93.38	92.48	**94.11**
person	68.58	71.35	70.58	72.79	**81.01**	74.83	79.48	78.77	76.07
rider	36.70	45.94	47.12	47.48	**58.50**	50.69	57.08	56.21	55.25
car	90.57	90.75	90.72	92.46	92.80	92.83	93.25	93.06	**93.70**
tuck	32.73	44.89	53.64	**72.78**	42.41	63.96	57.92	51.11	71.05
bus	45.58	61.59	65.06	76.62	66.08	75.35	75.66	**79.85**	78.03
train	27.03	31.12	41.26	60.84	35.08	56.61	53.18	**68.86**	67.06
mbike	11.12	39.66	42.65	44.24	54.21	47.72	**56.04**	50.54	49.92
bike	62.35	67.52	65.07	68.89	**76.27**	70.23	76.02	75.52	71.94
MIoU/%	58.35	63.91	65.04	69.06	69.26	69.71	70.84	70.99	**72.10**

注：加粗、带下划线字体分别表示各行最优与次优值。

表 3 给出了 9 种算法各类别 PA 值的对比结果。由表 3 可知，IDB 算法 5 个目标类别的 PA 值为最优，3 个目标类别的 PA 值为次优。且 IDB 算法的 MPA 值达到了 80.05%，相较于 ENet、CGNet、LEDNet、ICNet、DUNet、BiSeNet、EncNet、OCNet 算法，分别提高了 13.06、6.18、4.96、4.23、0.94、1.68、1.73、1.4 百分点。

表 3 9种算法的各类别像素准确率

Tab. 3 Each class PA of nine algorithms

类别	*PA*/%								
	ENet	CGNet	LEDNet	ICNet	DUNet	BiSeNet	EncNet	OCNet	IDB
road	98.62	98.59	98.66	98.82	98.35	98.67	98.56	98.42	**99.13**
swalk	86.76	87.89	88.31	88.83	89.03	**90.62**	87.61	89.60	89.04
build	95.94	94.90	94.69	**96.55**	96.14	95.58	96.10	96.04	96.08
wall	36.50	48.14	53.03	57.93	51.43	50.03	34.98	37.22	**63.04**
fence	54.83	58.76	**65.75**	56.10	61.33	63.45	58.69	59.25	58.30
pole	62.08	64.17	63.94	56.84	71.43	63.86	**71.58**	70.10	65.46
tlight	50.05	66.94	59.08	60.74	78.87	72.25	**80.09**	77.20	68.93
sign	67.39	75.50	73.09	70.45	85.15	78.02	83.92	**85.23**	80.88
veg	95.54	95.39	95.38	96.05	95.44	96.36	96.44	96.13	**96.45**
terrain	65.77	69.26	62.75	68.44	**75.45**	68.17	69.62	69.68	69.03
sky	96.98	97.12	97.52	96.03	96.67	**97.77**	96.95	96.99	97.32
person	81.90	85.22	83.56	84.46	89.62	87.29	90.51	**91.14**	86.78
rider	44.16	59.86	61.09	58.40	**71.23**	64.60	68.30	67.45	70.10
car	95.86	96.11	95.61	95.56	97.00	96.66	97.08	96.74	**97.41**
tuck	40.42	55.89	65.99	**78.87**	55.28	73.79	60.60	53.80	76.08
bus	58.09	83.46	76.61	82.45	72.57	**89.54**	85.21	84.09	85.49
train	46.87	36.61	56.02	63.56	64.62	60.62	58.83	71.78	**74.66**
mbike	13.05	48.98	52.62	48.20	**66.87**	57.48	65.98	65.26	60.38
bike	82.06	80.79	83.01	82.36	86.66	84.27	86.96	**88.31**	86.49
MPA/%	66.99	73.87	75.09	75.82	79.11	78.37	78.32	78.65	**80.05**

注：加粗、带下划线字体分别表示各行最优与次优值。

表 4 给出了 9 种算法在 4 项评价指标上的结果。由表 4 可知，与 ICNet 基准算法相比，IDB 算法的 OA、MPA、MIoU 分别提高了 0.63、4.23、3.04 百分点。由表 4 还可看出，IDB 算法的 OA、MPA、MIoU 均优于其它对比算法。

表 4 9种算法客观评价指标对比

Tab4 Comparison of objective evaluation indices of nine algorithms

Models	OA/%	MPA/%	MIoU/%	FPS/(帧·s^{-1})
ENet	93.78	66.99	58.35	119.49
CGNet	94.09	73.87	63.91	87.18
LEDNet	94.10	75.09	65.04	90.44
ICNet	94.62	75.82	69.06	62.76
DUNet	94.80	79.11	69.26	105.11
BiSeNet	94.89	78.37	69.71	**134.53**
EncNet	94.86	78.32	70.84	93.15
OCNet	94.83	78.65	70.99	75.88
IDB	**95.25**	**80.05**	**72.10**	20.48

注：加粗、带下划线字体分别表示各列最优与次优值。

3.4 定性分析

为定性分析 IDB 算法的分割性能，选取 5 幅图像，并与 ENet、CGNet、LEDNet、ICNet、DUNet、BiSeNet、EncNet、OCNet 算法进行对比。实验结果如图 6 所示，黄色框为效果提升明显部分。

由第 1 列对比图可知，ENet、DUNet、EncNet、OCNet 算法对卡车存在错误分割问题；ENet、CGNet、LEDNet、ICNet、BiSeNet 算法将行人的一部分错误分割成卡车。相比之下，IDB 算法对卡车和行人的分割效果更好。

由第 2 列对比图可知，IDB 算法对汽车尾部的分割效果优于 ENet、CGNet、LEDNet、ICNet、DUNet、BiSeNet、EncNet、OCNet 算法，更接近标签图。

由第 3 列对比图可知，IDB 以外的 8 种算法不同程度地将车道错误分割成人行道。ENet、CGNet、LEDNet、DUNet、BiSeNet、EncNet、OCNet 算法将骑行者的一部分错误分割成行人。相比之下，IDB 算法对车道和骑行者的分割结果更准确。

由第 4 列对比图可知，除 IDB 算法外，其余算法均将右边的骑行者错误分割成行人。ENet、CGNet、LEDNet、DUNet、BiSeNet、EncNet、OCNet 算法对

左边的骑行者存在错误分割现象。只有IDB算法没有错误分割问题，分割效果优于其它算法。

由第5列对比图可知，ENet、CGNet、LEDNet、DUNet、BiSeNet、EncNet、OCNet算法对汽车存在错误分割现象。ICNet算法虽然能分割出汽车，但在车轮轮廓处，IDB算法的分割结果与标签图更接近。

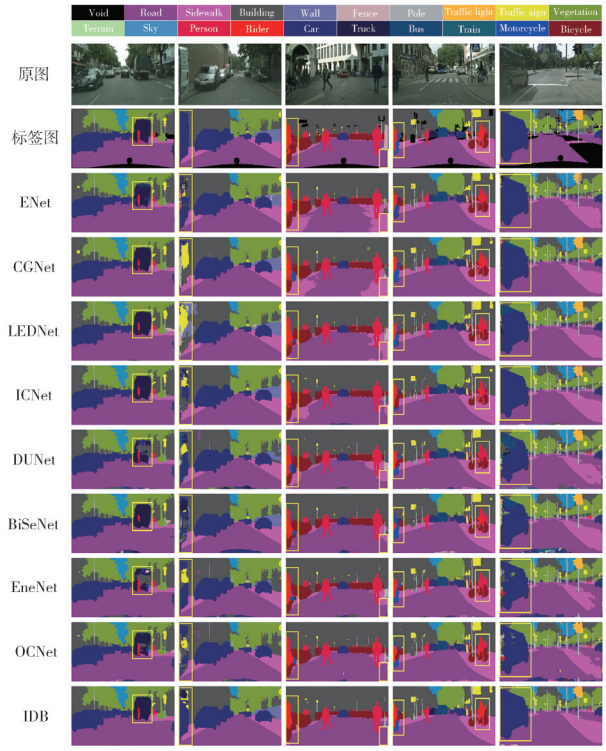

图 6　9种算法在Cityscapes数据集上的分割结果

Fig. 6　Segmentation results of nine algorithms on the Cityscapes dataset

3.5　消融实验

为验证IDB算法中DDRN、BPPM和APE模块的有效性，进行了4组消融实验，结果如表5～表8所示。

表 5　IDB算法的消融实验

Tab. 5　Ablation experiment of the IDB algorithm

	APE	BPPM	DDRN	*OA*/%	*MIoU*/%	*FWIoU*/%
1				94.62	69.06	90.15
2	√			94.74	69.74	90.38
3		√		94.71	67.61	90.36
4			√	95.03	70.64	90.87
5	√	√		94.82	69.08	90.58
6	√		√	95.12	71.72	91.06
7		√	√	95.23	71.62	**91.30**
8	√	√	√	**95.25**	**72.10**	**91.30**

注：加粗字体表示各列最优值。

表 6　DDRN的消融实验

Tab. 6　Ablation experiment of the DDRN

Backbone	*OA*/%	*MIoU*/%	*FWIoU*/%
ResNext-50	93.98	67.93	89.08
ResNest-50	94.27	68.21	89.64
Dilated-ResNet-50	94.62	69.06	90.15
Xception	94.66	70.09	90.29
DDRN	**95.03**	**70.64**	**90.87**

注：加粗字体表示各列最优值。

表 7　BPPM的消融实验

Tab. 7　Ablation experiment of the BPPM

方法	Backbone	*OA*/%	*MIoU*/%	*FWIoU*/%
PPM	DDRN	95.03	70.64	90.87
PPM+PANet	DDRN	94.94	70.96	90.77
PPM+BiFPN	DDRN	95.17	71.58	91.15
BPPM	DDRN	**95.23**	**71.62**	**91.30**

注：加粗字体表示各列最优值。

表5给出了DDRN、BPPM和APE模块的消融实验结果。由表中第1、8行可看出，IDB算法的OA、MIoU、FWIoU分别比ICNet提高了0.63、3.04、1.15百分点。由表中第5、8行可看出，缺少DDRN时，OA、MIoU、FWIoU分别下降了0.43、3.02、0.72百分点，这表明DDRN消除了空洞卷积引入的网格伪影，提高了网络的表征能力。由表中第6、8行可知，缺少BPPM时，OA、MIoU、FWIoU分别下降了0.13、0.38、0.24百分点，这说明BPPM通过自顶向下和自底向上的路径加强了多尺度特征的融合能力。由表中第7、8行可知，缺少APE模块时，OA、MIoU分别下降了0.02、0.48百分点，这说明APE模块通过创建的自注意力、向上注意力、向下注意力机制提高了跨空间与跨尺度的特征交互能力。

表6给出了DDRN的消融实验结果，表中第3行为基准网络。由表6可知，与采用ResNext-50、ResNest-50、Dilated-ResNet-50、Xception的网络相比，采用DDRN网络的OA分别提高了1.05、0.76、0.41、0.37百分点，MIoU分别提升了2.71、2.43、1.58、0.55百分点，FWIoU分别提升了1.79、1.23、0.72、0.58百分点。可以看出，5种骨干网络中，Backbone为DDRN的网络分割性能最好，验证了DDRN模块的有效性。

表7给出了BPPM的消融实验结果。由表7可知，相较于PPM、PPM+PANet、PPM+BiFPN，采用BPPM分割方法的OA分别提升了0.20、0.29、0.06百分点，MIoU分别提高了0.98、0.66、0.04百分点，FWIoU分别提高了0.43、0.53、0.15百分点。可以看出，4种方法中，BPPM的特征融合效果最佳，验证了BPPM的有效性。

表8给出了APE模块的消融实验结果。表中第1～4行、第5～8行的骨干网络分别采用Dilated-ResNet-50和DDRN。由表中第1～4行可知，与ICNet、BPA[25]和FPN相比，采用APE分割方法的MPA分别提高了1.38、0.2、0.14百分点，MIoU分别提高了0.68、0.52、0.31百分点。由表中第5～8行可看出，与第5、6、7行相比，采用APE分割方法的OA分别提高了0.02、0.26、0.17百分点，MIoU分别提升了0.48、0.86、0.54百分点。综上可知，各模块中，APE的特征融合效果最好，证明了APE模块的有效性。

表8 APE模块的消融实验

Tab. 8 Ablation experiment of the APE module

方法	Backbone	*OA*/%	*MPA*/%	*MIoU*/%	*FWIoU*/%
ICNet	Dilated-ResNet-50	94.62	75.82	69.06	90.15
ICNet+BPA	Dilated-ResNet-50	94.72	77.0	69.22	90.36
ICNet+FPN	Dilated-ResNet-50	94.78	77.06	69.43	90.46
ICNet+APE	Dilated-ResNet-50	94.74	77.2	69.74	90.38
BPPM	DDRN	95.23	**80.61**	71.62	**91.30**
BPPM+BPA	DDRN	94.99	79.56	71.24	90.83
BPPM+FPN	DDRN	95.08	79.69	71.56	91.02
BPPM+APE	DDRN	**95.25**	80.05	**72.10**	**91.30**

注：加粗字体表示各列最优值。

4 结 论

本文提出一种去伪影空洞残差与双向注意力金字塔增强的图像级联语义分割算法。该算法构建去伪影空洞残差网络，消除空洞卷积引入的网格伪影，增强网络的特征提取能力；构造双向金字塔池化模块，引入自顶向下和自底向上的路径，充分融合多尺度特征；设计注意力金字塔增强模块，通过自注意力、向上注意力、向下注意力机制建立跨空间与跨尺度的特征交互。实验结果表明，Cityscapes数据集上所提算法的总体精确率、平均像素精确度、平均交并比均高于其它8种主流算法；与基准算法相比，以上3种评价指标分别提高了0.67%、5.58%、4.40%。

参考文献：

[1] JIANG B, AN X, XU S, et al. Intelligent image semantic segmentation: a review through deep learning techniques for remote sensing image analysis[J]. Journal of the Indian Society of Remote Sensing, 2023, 51(9): 1865-1878.

[2] SUN Y, ZHENG W. HRNet-and PSPNet-based multiband semantic segmentation of remote sensing images [J]. Neural Computing and Applications, 2023, 35(12): 8667-8675.

[3] SU Z, LI W, MA Z, et al. An improved U-Net method for the semantic segmentation of remote sensing images[J]. Applied Intelligence, 2022, 52(3): 3276-3288.

[4] HUANG T, SONG S, LIU Q, et al. A novel multi-exposure fusion approach for enhancing visual semantic segmentation of autonomous driving[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2023, 237(7): 1652-1667.

[5] FAN J, GAO B, GE Q, et al. SegTransConv: transformer and CNN hybrid method for real-time semantic segmentation of autonomous vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(2): 1586-1601.

[6] YASMIN S, DURRANI M Y, GILLANI S, et al. Small obstacles detection on roads scenes using semantic segmentation for the safe navigation of autonomous vehicles[J]. Journal of Electronic Imaging, 2022, 31(6): 061806.

[7] JULIUS FUSIC S, HARIHARAN K, SITHARTHAN R, et al. Scene terrain classification for autonomous vehicle navigation based on semantic segmentation method [J]. Transactions of the Institute of Measurement and Control, 2022, 44(13): 2574-2587.

[8] RAHMAN M M, MARCULESCU R. Multi-scale hierarchical vision transformer with cascaded attention decoding for medical image segmentation[C]//Medical Imaging with Deep Learning, PMLR, 2024: 1526-1544.

[9] HATAMIZADEH A, TANG Y, NATH V, et al. UNETR: transformers for 3D medical image segmentation [C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022: 1748-1758.

[10] TYAGI P, MASSON P, JAIN A, et al. Automated knowledge transfer for medical image segmentation using deep learning [C]//2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE), 2024: 954-960.

[11] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation [C]//IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017: 640-651.

[12] 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.

[13] PENG C, ZHANG X, YU G, et al. Large kernel matters--improve semantic segmentation by global convolutional network [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 1743-1751.

[14] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6230-6239.

[15] YUAN Y, HUANG L, GUO J, et al. OCNet: object context for semantic segmentation [J]. International Journal of Computer Vision, 2021, 129 (8): 2375-2398.

[16] ZHAO H, QI X, SHEN X, et al. ICNet for real-time semantic segmentation on high-resolution images [C]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018: 418-434.

[17] ZHANG Y, HARE J, PRÜGEL-BENNETT A. Learning to count objects in natural images for visual question answering [J/OL]. [2024-03-02]. arXiv: 1802.05766v1. https://doi.org/10.48550/arXiv.1802.05766.

[18] 彭秀平, 仝其胜, 林洪彬, 等. 一种面向散乱点云语义分割的深度残差-特征金字塔网络框架[J]. 自动化学报, 2021, 47(12): 2831-2840.
PENG Xiuping, TONG Qisheng, LIN Hongbin, et al. A deep residual-feature pyramid network framework for scattered point cloud semantic segmentation [J]. Acta Automatica Sinica, 2021, 47(12): 2831-2840. (in Chinese)

[19] PASZKE A, CHAURASIA A, KIM S, et al. ENet: a deep neural network architecture for real-time semantic segmentation [J/OL]. [2024-03-02]. arXiv: 1606.02147v1. http://arxiv.org/abs/1606.02147v1.

[20] WU T, TANG S, ZHANG R, et al. Cgnet: a light-weight context guided network for semantic segmentation [J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2021, 30: 1169-1179.

[21] WANG Y, ZHOU Q, LIU J, et al. Lednet: A light-weight encoder-decoder network for real-time semantic segmentation [C]// IEEE International Conference on Image Processing (ICIP), 2019: 1860-1864.

[22] TIAN Z, HE T, SHEN C, et al. Decoders matter for semantic segmentation: data-dependent decoding enables flexible feature aggregation [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 3121-3130.

[23] YU C, WANG J, PENG C, et al. BiSeNet: bilateral segmentation network for real-time semantic segmentation [C]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018: 334-349.

[24] ZHANG H, DANA K, SHI J, et al. Context encoding for semantic segmentation [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 7151-7160.

[25] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 8759-8768.