第 58 卷 第 8 期/2021 年 4 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于区域自我注意力的实时语义分割网络

鲍海龙，万敏*，刘忠祥，秦勉，崔浩宇

西南石油大学机电工程学院，四川 成都 610500

摘要 高精度的语义分割结果往往依赖于丰富的空间语义信息与细节信息，但这两者的计算量均较大。为了解决该问题，通过分析图像局部像素具有的相似性，提出了一种基于区域自我注意力的实时语义分割网络。该网络可分别通过一个区域级的自我注意力模块和一个局部交互通道注意力模块计算出特征信息的区域级关联性和通道注意力信息，然后以较少的计算量获取丰富的注意力信息。在 Cityscapes 数据集上的实验结果表明，相比现有的实时分割网络，本网络的分割精度更高、速度更快。

关键词 图像处理；语义分割；卷积神经网络；注意力机制

中图分类号 TP391.4 **文献标志码** A **doi**：10.3788/LOP202158.0810018

Real-Time Semantic Segmentation Network Based on Regional Self-Attention

Bao Hailong, Wan Min*, Liu Zhongxiang, Qin Mian, Cui Haoyu

School of Mechatronic Engineering, Southwest Petroleum University, Chengdu, Sichuan 610500, *China*

Abstract High accuracy results of semantic segmentation often rely on rich spatial semantic information and detailed information, but both incurring high computational costs. In order to solve this problem, we propose a real-time semantic segmentation network based on regional self-attention by observing the similarity of local pixels in the image. The network can calculate the regional correlation of feature information and channel attention information through a regional self-attention module and a local interactive channel attention module. Then, it obtains rich attention information with less calculation. The experimental results on the Cityscapes dataset show that the segmentation accuracy and speed of the network are higher than the existing real-time segmentation network.

Key words image processing; semantic segmentation; convolutional neural networks; attention mechanism

OCIS codes 100.4996; 100.2960; 200.4260

1 引 言

随着深度学习的快速发展，图像处理任务如光学领域的红外图像、光谱图像处理任务[1-2]、以摄像头为主要应用载体的自动驾驶、人物识别和遥感图像分割等计算机视觉任务[3-6]都得到了快速发展。语义分割是计算机视觉任务的一项核心技术，目的是将图像分割成几组具有特定语义类别的区域，属于像素级的密集分类问题[7]。语义分割可用于红外图像的分割，实现全天候的图像分析与理解，也可应用于现实街道场景的分割，实现自动驾驶的环境感知等任务。对于自动驾驶等快速移动场景，网络的分割速率和精度十分重要，要想获得高精度的分割结果，分割网络必须要获取足够多的语义信息与细节信息[8-9]。但这两者均需要通过加深网络参数或提高输入图像的分辨率实现，导致网络的计算量过大、分割效率过低[10-11]。

自我注意力(SA)机制[12-13]是计算机视觉领域中用来获取长距离语义信息的方法，能很大程度地加深网络对整个特征图的理解。但该方法需要计算

收稿日期：2020-08-05；**修回日期**：2020-09-02；**录用日期**：2020-09-14

* **E-mail**：2264696759@qq.com

出特征图中两两特征点之间的关系，以得到任意特征点对当前特征点的关系权值，该过程的计算量为 $O(N^2C)$（$N=H\times W$，H 和 W 分别为特征图的长和宽，C 为特征的通道数），且计算量随特征图尺寸的增加呈 2 次平方增长关系（如特征图尺寸由 N 增至 aN，计算量则由 N^2C 增至 a^2N^2C），不适用于实时网络的搭建。虽然可通过池化等下采样方式降低图像的分辨率、减少计算量，但会丢失特征图中大量的语义信息，尤其对于低分辨率的高层特征，不利于网络性能的提升[14]。

实际运用中，图像的局部像素分布是具有相似性的，同一区域或同一类别一般拥有相似甚至相同的像素值，传统 SA 机制[12-13]遍历计算所有特征点的两两关联性是冗余且不必要的。因此，本文提出了一个轻量级的区域 SA（RSA）模块，在不损失特征信息的情况下，将特征图通过缩放因子 r 进行区域缩放，将传统 SA 机制的像素级关联性计算转变为区域级关联性计算，从而将计算量减少为 $O(N^2C)/r^2$。随后又提出了一个轻量级的局部通道交互注意力（LCIA）模块，可在不降维、不损失通道信息的情况下，提高网络性能。基于 RSA 和 LCIA 模块，搭建了一个编码器-解码器形式的实时分割网络，利用编码器提取不同阶段的图像特征信息；再利用 RSA 模块对每一阶段的特征进行二次处理，加强网络对每一层信息的全局理解；最后在解码器中结合 LCIA 模块对每一阶段的信息进行有效融合，依次恢复图像的尺寸与细节信息。

2　网络框架的设计

2.1　自我注意力机制

SA 机制可获取所有特征点之间的两两关联性，计算出一个特征点对其他所有特征点的加权影响，从而得到更全面的语义信息，可表示为

$$\boldsymbol{Y}=f(\boldsymbol{Q},\boldsymbol{K}^{\mathrm{T}})\cdot\boldsymbol{V},\tag{1}$$

式中，$\boldsymbol{Y}$ 为 SA 机制的输出，f 为相似度计算函数，T 为矩阵的转置操作，$\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 为原特征图 $\boldsymbol{X}\in\boldsymbol{R}^{C\times H\times W}$ 分别通过 3 个不同的 1×1 卷积得到的相关特征图，3 者的结构与 $\boldsymbol{X}$ 相同。其中，$\boldsymbol{V}$ 包含了原有像素的语义信息，$\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 通过 $f(\boldsymbol{Q},\boldsymbol{K}^{\mathrm{T}})$ 计算出两两特征点之间的关联性，同时结合 Softmax 函数得到注意力图（Attention map），可表示为

$$Y_{j,i}=\frac{\exp(X_i\cdot X_j)}{\sum_{i=1}^{n}\exp(X_i\cdot X_j)},\tag{2}$$

式中，X_i 为特征图 $\boldsymbol{Q}$ 中的第 i 个像素，X_j 为特征图 $\boldsymbol{K}^{\mathrm{T}}$ 中的第 j 个像素，n 为 $\boldsymbol{Q}$ 与 $\boldsymbol{K}^{\mathrm{T}}$ 的像素数量，$Y_{j,i}$ 为像素 i 对像素 j 的影响，两者越相似，则影响值越大[12]。为方便计算所有空间上的像素点，将上述 3 个相关特征图通过矩阵平铺处理得到 $\boldsymbol{X}\in\boldsymbol{R}^{C\times N}$，$f(\boldsymbol{Q},\boldsymbol{K}^{\mathrm{T}})$ 对应的矩阵行列计算式为 $(N,C)\cdot(C,N)$，计算量为 $O(N^2C)$。可以看出，计算量很大，且随特征图尺寸的增加呈 2 次平方增长关系。

2.2　区域级的注意力模块

图 1 为实际处理的图像，可以看出，局部区域中的相邻像素往往是同一类别且有着相似甚至相同的像素值。对于这些相似的像素点，其获得的全局关联性也应是相似的。因此，通过遍历计算所有单个特征点之间的关联性得到注意力信息是冗余且不必要的。

可利用局部区域内相邻像素具有相似性的特点减少 SA 机制的计算量，设计出一个轻量级的 RSA 模块。RSA 模块可在不损失特征信息的情况下减少计算量，并得到相应的注意力信息，结构如图 2(a)所示。RSA 模块包含像素移位（PS）和反向像素移位（R-PS）两个核心操作，如图 2(b)所示。

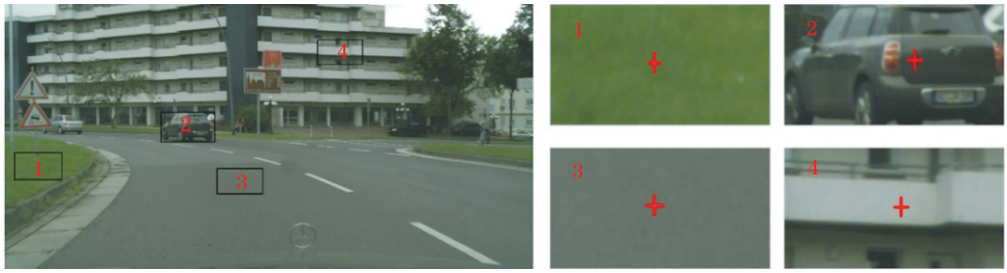

图 1　局部区域的像素分布

Fig. 1　Pixel distribution of the local area

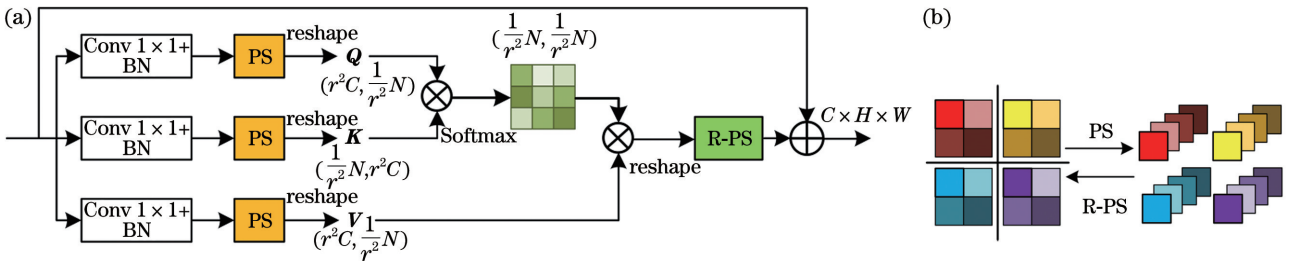

图 2　RSA 模块的结构。(a) RSA 模块;(b) PS 与 R-PS 模块

Fig. 2　Structure of the RSA module. (a) RSA module; (b) PS and R-PS modules

首先,利用 3 个不同的 1×1 卷积(Conv 1×1)与批标准化(BN)处理输入特征图。然后,利用 PS 模块控制缩放因子 r,将一个像素 i 四周 r^2-1 个相似像素点移位至与自身同一通道的相邻位置,即将一片区域的特征点整理在同一通道上,从而在不损失特征信息的同时降低图像的分辨率。R-PS 为 PS 的反向操作,可将通道上移位后的 r^2-1 个像素点还原到像素 i 周围原来的空间位置。基于 PS 和 R-PS 可将(1)式转换为

$$\boldsymbol{Y}=f\left[X_{\mathrm{PS}}(\boldsymbol{Q}),X_{\mathrm{PS}}(\boldsymbol{K}^{\mathrm{T}})\right]\cdot X_{\mathrm{PS}}(\boldsymbol{V}),\tag{3}$$

将(2)式转换为

$$Y_{v,u}=\frac{\exp\left(\sum_{k=1}^{r^2}X_{u,k}\cdot X_{v,k}\right)}{\sum_{u=1}^{n}\exp\left(\sum_{k=1}^{r^2}X_{u,k}\cdot X_{v,k}\right)},\tag{4}$$

式中,X_{PS} 为 PS 操作,$X_{u,k}$ 和 $X_{v,k}$ 为需要计算的两个缩放区域,k 为两个区域对应位置的第 k 个像素点,$Y_{v,u}$ 为区域 $X_{u,k}$ 对 $X_{v,k}$ 的影响力。可以看出,(4)式通过计算对应位置像素点的关联性得到区域 $X_{u,k}$ 和 $X_{v,k}$ 的关联性。在计算量方面,特征图 $\boldsymbol{X}\in\boldsymbol{R}^{C\times H\times W}$ 通过缩放因子为 r 的 PS 模块得到 $\boldsymbol{X}\in\boldsymbol{R}^{r^2C\times\frac{1}{r}H\times\frac{1}{r}W}$,经过矩阵平铺后得到 $\boldsymbol{X}\in\boldsymbol{R}^{r^2C\times\frac{1}{r^2}N}$,相应的矩阵行列计算式为 $\left(\frac{1}{r^2}N,r^2C\right)\cdot\left(r^2C,\frac{1}{r^2}N\right)$,计算量减少为 $\frac{1}{r^2}O(N^2C)$,如 $r=4$ 时,计算量就减少为原来的 1/16。因此,可通过控制缩放因子 r 减少 SA 机制的计算量。

完成区域级的全局特征关系计算后,先采用矩阵反向操作将特征图恢复为空间维度上的二维图像,然后采用 R-PS 模块将通道 C 上移位的像素点恢复到原来所在的空间位置,即将 $\boldsymbol{X}\in\boldsymbol{R}^{r^2C\times\frac{1}{r}H\times\frac{1}{r}W}$ 变回 $\boldsymbol{X}\in\boldsymbol{R}^{C\times H\times W}$,恢复特征图的原始尺寸。为了保证细节信息的完整性,将输入和输出进行相加融合,形成残差连接。

2.3　局部通道交互注意力模块

通道注意力机制可为特征图的每一通道获取到相应的权值信息,提高网络的表达能力。现有通道注意力机制如 SENet(Squeeze-and excitation networks)[15]和卷积块注意模块(CBAM)[16],均利用全连接计算得到权值信息,通常会用通道降维操作(减至原始图像尺寸的 1/16)减少全连接的计算负担。与空间降维类似,通道降维同样会损失大量的语义信息,且捕捉所有通道信息之间的依赖是低效且不必要的。考虑到 CBAM 可通过局部卷积的方式获取空间注意力信息,设计了 LCIA 模块,通过少量的参数计算提升网络的性能,LCIA 模块的结构如图 3 所示。

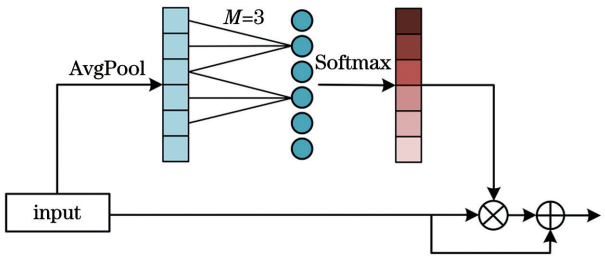

图 3　LCIA 模块的结构

Fig. 3　Structure of the LCIA module

从图 3 可以发现,进行全局平均池化(AvgPool)后,ICLA 模块没有进行通道降维处理,保证了信息的完整性。不同于全连接,ICLA 模块利用长度为 M 的一维卷积模块,仅提取当前通道与其相邻 $M-1$ 个局部通道生成注意力信息,可表示为

$$Y_o=\delta\left(\sum_{m=1}^{M}X_{o,m}\cdot L_m\right),\tag{5}$$

式中,L 为大小为 M 的一维卷积核,可聚合 M 个相邻局部通道值,$X_{o,m}$ 为输入特征第 o 个通道的第 m 个相邻通道,δ 为 Softmax 激活函数,Y_o 为 M 个局部通道对当前通道特征的注意力信息。卷积本身的权值共享特性,使整个 LCIA 模块仅有 M 个参数,计算量为 $O(MC)$,保证了网络的效率,且仅采用部

分相邻通道信息(如 $M=3$)也能带来明显的性能增益。

2.4 网络结构

结合 RSA 与 LCIA 模块并采用编码器-解码器结构搭建了分割网络框架,如图 4(a)所示。解码器部分用结合 18 层的小型残差网络 ResNet-18[17]作为骨架网络获取图像的特征信息,共 5 个阶段,每个阶段均对图像进行 1 次下采样,最后网络输出的特征图尺寸为原始图像尺寸的 1/32,在第 2、3、4、5 阶段对特征进行处理。对于每一阶段的特征信息,首先,利用一个尺寸为 3×3 的卷积模块对特征进行局部处理以融合局部特征信息;然后,结合空洞卷积(DConv 3×3)[18]提高网络感受野,每一个卷积模块后都接一个 BN 与修正线性单元(ReLU)激活函数;其次,利用 RSA 模块获取特征信息的区域级全局关联性。考虑到不同阶段特征有着不同的分辨率和局部相似性,将第 2、3、4、5 阶段 RSA 模块的缩放率分别设置为(4,4,2,1)。在解码器部分结合 LCIA 模块,用图 4(b)所示的特征融合模块(FFM)依次恢复图像的分辨率和细节信息。

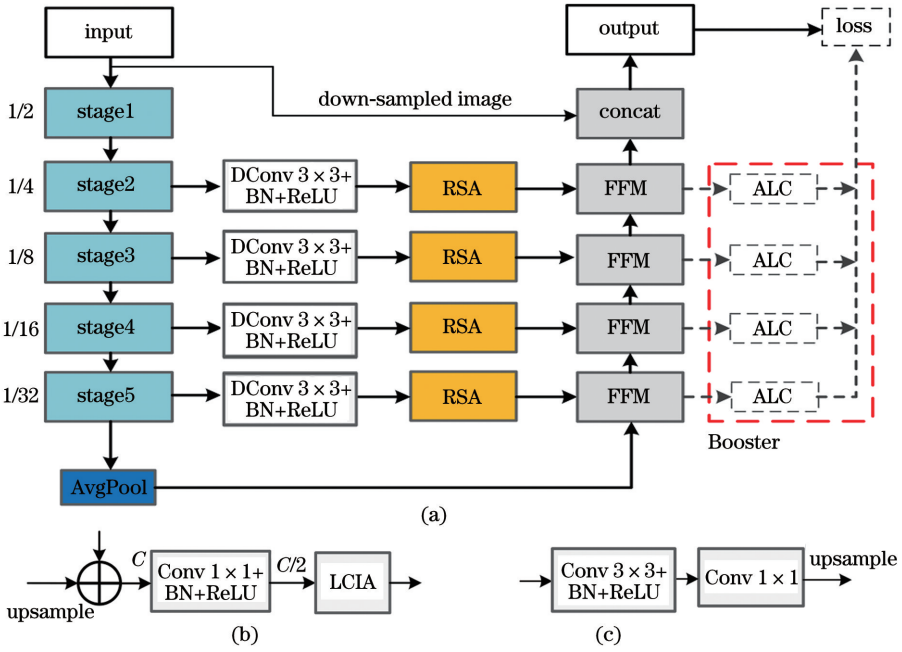

图 4　本网络的结构。(a)网络结构;(b)特征融合模块;(c)辅助损失分类器

Fig. 4　Structure of our network. (a) Network structure; (b) feature fusion module; (c) auxiliary loss classifier

为进一步提升分割效果,设计出一个强化训练 Booster[19]模块,即在解码器每一阶段设置一个辅助损失分类器(ALC)对初始的分割结果进行监督学习,如图 4(c)所示。Booster 模块可在训练阶段增强网络的特征表达能力,且在测试使用时不会参与计算,从而在不影响网络分割效率的情况下提升网络的分割准确度。

3 实验结果

3.1 实验设置

通过 Cityscapes 数据集[20]验证本网络的有效性,Cityscapes 数据集包括 50 个不同城市中的街道场景图像,共 5000 张精标注的图像,其中,2975 张用以训练,500 张用于验证,1525 张用于测试。基于精标注的图像数据进行实验,用包含 19 类物体的图像进行训练和测试。实验环境:软件环境为 Pytorch 深度学习框架,显卡为 1080ti。实验过程中,用随机梯度下降(SGD)算法优化收敛过程;采用 poly 学习率衰减策略,初始学习率为 e^{-2},权值衰减率为 e^{-4},动量为 0.9。损失函数为交叉熵损失函数,批量大小为 10。为增强模型的学习能力,对数据集进行增强处理,包括随机镜像、随机尺寸缩放等,其中缩放范围为{0.75, 1.0, 1.5, 1.75, 2.0}。用平均交互比(MIoU)衡量网络的分割精度,用每秒传输帧数(FPS)衡量网络的分割效率。

3.2 验证实验

3.2.1 缩放率对比实验

RSA 模块可获取有效的区域级特征关联性,但不同阶段的特征图有不同的分辨率,低层特征图的分辨率较大,有较为粗糙的语义信息和更广的相似性,高层信息则相反[21]。因此,对不同阶段分别设置不同的缩放率进行对比实验,结果如表1所示。

表 1 缩放率的对比实验

Table 1 Comparison experiment of the zoom ratio

Serial number	Network	MIoU /%	FPS /frame
1	(1,1,1,1)	71.9	10
2	(4,2,2,1)	71.7	109
3	(4,4,2,1)	71.7	120
4	(8,4,2,1)	71.6	133

其中，第 1 组第 2、3、4、5 阶段 RSA 模块的缩放率为(1,1,1,1)，表示不对特征图进行区域缩放，即原始的 SA 机制[13]，其 MIoU 为 71.9%，FPS 仅为 10 frame，无法满足实时分割的需求。进行缩放处理后，第 2、3、4 组参数的 MIoU 分别为 71.7%、71.7%和 71.6%，FPS 分别为 109，120，133 frame。可以看出，在几乎不影响分割精度的情况下，RSA 模块极大提高了网络的分割速度。

3.2.2 消融实验

为验证 RSA 模块对网络表达能力的提升，进行了消融实验，结果如表 2 所示。可以发现，不采用 RSA 模块时，网络的 MIoU 为 68.4%，FPS 为 158 frame；采用 RSA 模块后，网络的 MIoU 为 71.7%，FPS 为 120 frame。相比直接对特征信息进行融合处理，RSA 模块能帮助网络捕捉到更清晰的特征关联性和长距离信息，提高网络的表达能力。添加 LCIA 模块后，网络的 MIoU 为 72.3%，FPS 为 115 frame。与 CBAM 相比，LCIA 模块以更快的分割速度取得了与其相近的分割精度；与 SENet 相比，LCIA 模块则以更快的分割速度取得了比其更高的分割精度，这表明通道的降维同样会影响网络性能，且仅通过局部的通道交互信息就能获取有效的注意力信息，提高网络性能。结合 Booster 增强训练后，网络的 MIoU 上升到 73.1%，且没有影响分割速度，这表明结合辅助损失训练可有效增强网络的表达能力。

表 2 消融实验的结果

Table 2 Results of the ablation experiments

Network	MIoU /%	FPS /frame
Original network	68.4	158
RSA	71.7	120
RSA+LCIA	72.3	115
RSA+CBAM	72.5	80
RSA+SENet	72.1	102
RSA+LCIA + Booster	73.1	115

为验证 RSA 模块对特征信息的保留能力，分别选择平均池化和最大池化进行对比实验，两者的下采样率与 RSA 模块相同，通过线性插值恢复图像的原始尺寸，结果如表 3 所示。可以发现，相比最大池化和平均池化，采用 RSA 模块后，在相近的速度下网络的分割效果更好。这表明池化过程中的信息丢失对网络的表达能力是有害的，而 RSA 模块可更有效地保留特征信息，进一步说明特征信息的完整性对网络的重要性。

表 3 下采样方式的对比实验

Table 3 Comparison experiment of the down-sampling method

Network	MIoU /%	FPS /frame
AvgPool	70.5	126
MaxPool	70.4	132
RSA	71.7	120

3.3 对比实验

选取几种常见的分割网络[22-28]与本网络的性能进行对比，其中 ENet、ESPNet、ERFNet 与 DABNet 没有采用骨干网络，ICNet 和 DFANet 分别采用预训练网络 PSPNet50 和 XceptionA 作为骨干网络。输入图像的分辨率均为 512 pixel×1024 pixel，结果如表 4 所示。可以发现，引入轻量型网络 ResNet-18 后 FPS 分别为 115 和 126 frame 时，本网络在测试集上的 MIoU 分别为 72.1%和 71.8%，在分割准确度和分割速率上均优于其他实时分割网络。

表 4 不同网络的实验结果

Table 4 Experimental results of different networks

Network	Pretrain	MIoU /%	FPS /frame
ENet	no	58.3	76
ESPNet	no	60.3	112
ERFNet	no	68.0	41.7
ICNet	PSPNet50	69.5	30
DABNet	no	70.1	104
DFANet*	Xception A	70.3	–
DFANet	Xception A	71.3	100
Ours	ResNet-18	72.1/71.8	115/126

为更直观地展现本网络的优越性，选取部分分割结果并对其进行可视化处理，同时与 ERFNet 进行可视化对比，结果如图 5 所示。可以看出，本网络在局部区域上可取得更精细的分割效果，对于细小物体可进行更有效的分割，整体分割结果中的类内不一致和类间不一致情况较少。

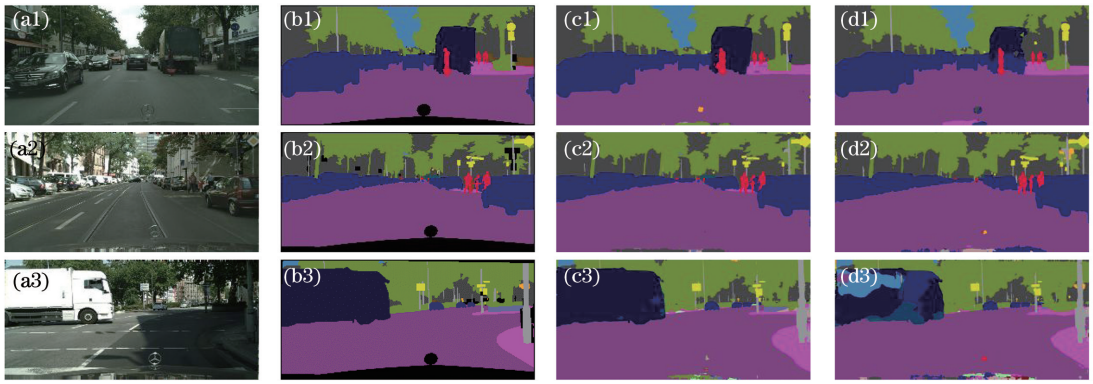

图 5 Cityscapes 数据集的可视化结果。(a)原始图像;(b)真实的分割结果;(c)本网络的分割结果;(d) ERFNet 的分割结果

Fig. 5 Visualization results of the Cityscapes dataset. (a) Original image; (b) real segmentation result; (c) segmentation result of our network; (d) segmentation result of the ERFNet

4 结 论

基于局部像素分布的相似性,设计了一个轻量级 RSA 模块,可在不损失特征信息的情况下,获取全局信息的区域级关联性;且不需要遍历计算所有特征点的两两关联性,极大降低了 SA 机制的计算量,提高网络的分割效率。随后提出了一个 LCIA 模块,仅通过相邻局部通道就能获取有效的通道注意力信息,且避免了通道降维操作,保留了通道信息的完整性。基于 RSA 和 LCIA 模块,搭建了一个编码器-解码器结构的实时语义分割网络,利用 RSA 模块提取每一阶段特征的区域关联性,加强网络的表达能力;在解码器部分结合 LCIA 模块,提升网络性能。实验结果表明,相比其他网络,本网络有更优分割结果和分割效率。

参 考 文 献

[1] Tang C Y, Pu S L, Ye P Z, et al. Fusion of low-illuminance visible and near-infrared images based on convolutional neural networks[J]. Acta Optica Sinica, 2020, 40(16): 1610001.
唐超影, 浦世亮, 叶鹏钊, 等. 基于卷积神经网络的低照度可见光与近红外图像融合[J]. 光学学报, 2020, 40(16): 1610001.

[2] Kong F Q, Zhou Y B, Shen Q, et al. End-to-end multispectral image compression using convolutional neural network[J]. Chinese Journal of Lasers, 2019, 46(10): 1009001.
孔繁锵, 周永波, 沈秋, 等. 基于卷积神经网络的端到端多光谱图像压缩方法[J]. 中国激光, 2019, 46(10): 1009001.

[3] Liu H, Peng L, Wen J W. Multi-scale aware pedestrian detection algorithm based on improved full convolutional network[J]. Laser & Optoelectronics Progress, 2018, 55(9): 091504.
刘辉, 彭力, 闻继伟. 基于改进全卷积网络的多尺度感知行人检测算法[J]. 激光与光电子学进展, 2018, 55(9): 091504.

[4] He Y H, Wang H, Zhang B. Color-based road detection in urban traffic scenes[J]. IEEE Transactions on Intelligent Transportation Systems, 2004, 5(4): 309-318.

[5] Yao L S, Xu G M, Zhao F. Facial expression recognition based on local feature fusion of convolutional neural network[J]. Laser & Optoelectronics Progress, 2020, 57(4): 041513.
姚丽莎, 徐国明, 赵凤. 基于卷积神经网络局部特征融合的人脸表情识别[J]. 激光与光电子学进展, 2020, 57(4): 041513.

[6] Zhang Z H, Fang W, Du L L, et al. Semantic segmentation of remote sensing image based on encoder-decoder convolutional neural network[J]. Acta Optica Sinica, 2020, 40(3): 0310001.
张哲晗, 方薇, 杜丽丽, 等. 基于编码-解码卷积神经网络的遥感图像语义分割[J]. 光学学报, 2020, 40(3): 0310001.

[7] Zhang X F, Liu J, Shi Z S, et al. Review of deep learning-based semantic segmentation[J]. Laser & Optoelectronics Progress, 2019, 56(15): 150003.
张祥甫, 刘健, 石章松, 等. 基于深度学习的语义分割问题研究综述[J]. 激光与光电子学进展, 2019, 56(15): 150003.

[8] Lin G S, Milan A, Shen C H, et al. RefineNet: multi-path refinement networks for high-resolution semantic segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5168-5177.

[9] Peng C, Zhang X Y, Yu G, et al. Large kernel matters: improve semantic segmentation by global convolutional network[C]//2017 IEEE Conference on

Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1743-1751.

[10] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6230-6239.

[11] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation [EB/OL]. [2019-12-09]. https://arxiv.org/abs/1706.05587.

[12] Wang X L, Girshick R, Gupta A, et al. Non-local neural networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7794-7803.

[13] Fu J, Liu J, Tian H J, et al. Dual attention network for scene segmentation [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3141-3149.

[14] Yu C Q, Wang J B, Peng C, et al. Learning a discriminative feature network for semantic segmentation [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 1857-1866.

[15] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks [EB/OL]. [2020-07-20]. https://arxiv.org/abs/1709.01507.

[16] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module [EB/OL]. [2020-07-25]. https://arxiv.org/abs/1807.06521.

[17] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[18] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [EB/OL]. [2020-07-21]. https://arxiv.org/abs/1606.00915.

[19] Mehta S, Rastegari M, Shapiro L, et al. ESPNetv2: a light-weight, power efficient, and general purpose convolutional neural network [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 9182-9192.

[20] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 3213-3223.

[21] Yu C Q, Wang J B, Peng C, et al. BiSeNet: bilateral segmentation network for real-time semantic segmentation[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer Vision-ECCV 2018. Lecture Notes in Computer Science. Cham: Springer, 2018, 11217: 334-349.

[22] Paszke A, Chaurasia A, Kim S, et al. ENet: a deep neural network architecture for real-time semantic segmentation [EB/OL]. [2020-07-23]. https://arxiv.org/abs/1606.02147.

[23] Mehta S, Rastegari M, Caspi A, et al. ESPNet: efficient spatial pyramid of dilated convolutions for semantic segmentation [M] // Ferrari V, Hebert M, Sminchisescu C, et al. Computer Vision-ECCV 2018. Lecture Notes in Computer Science. Cham: Springer, 2018, 11214: 561-580.

[24] Romera E, Álvarez 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.

[25] Zhao H S, Qi X J, Shen X Y, et al. ICNet for real-time semantic segmentation on high-resolution images [M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer Vision-ECCV 2018. Lecture Notes in Computer Science. Cham: Springer, 2018, 11207: 418-434.

[26] Wang Y, Zhou Q, Liu J, et al. Lednet: a lightweight encoder-decoder network for real-time semantic segmentation [C] // 2019 IEEE International Conference on Image Processing (ICIP), September 22-25, 2019, Taipei, Taiwan, China. New York: IEEE Press, 2019: 1860-1864.

[27] Li G, Yun I, Kim J, et al. DABNet: depth-wise asymmetric bottleneck for real-time semantic segmentation [EB/OL]. [2020-07-22]. https://arxiv.org/abs/1907.11357.

[28] Li H C, Xiong P F, Fan H Q, et al. DFANet: deep feature aggregation for real-time semantic segmentation [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 9514-9523.