Artificial Intelligence and Robotics Research 人工智能与机器人研究, 2024, 13(1), 19-29
Published Online February 2024 in Hans. https://www.hanspub.org/journal/airr
https://doi.org/10.12677/airr.2024.131003

Hans汉斯

多尺度信息融合的实时语义分割网络

胡家虎，佘玉梅

云南民族大学数学与计算机科学学院，云南 昆明

收稿日期：2023年11月14日；录用日期：2024年1月29日；发布日期：2024年2月5日

摘　要

在自动驾驶、无人机等处理器资源受限的任务中，需要考虑模型的参数量和运算速度，并确保较好的准确性。一些语义分割模型采用并行式结构提取多尺度信息时，使用深度可分离卷积或分组卷积替换常规卷积来降低计算量。但这些操作存在增加网络延迟，降低推理速度的问题。基于此问题，提出一个基于编码器－解码器的实时语义分割模型。编码器阶段，使用部分卷积结合扩张卷积构建不同的并行式模块，用于提取不同阶段的多尺度信息。解码器阶段，使用融合上采样特征的方式。模型在Cityscapes和CamVid数据集上进行实验，平均交并比分别为71.3%和66.8%，运行速度分别为97帧/s和98帧/s，结果表明该模型在分割精度和运行速度之间达到较好平衡。

关键词

实时语义分割，部分卷积，多尺度特征，编解码器结构

Real-Time Semantic Segmentation Network Based on Multi-Scale Information

Jiahu Hu, Yumei She

School of Mathematics and Computer Science, Yunnan Minzu University, Kunming Yunnan

Received: Nov. 14th, 2023; accepted: Jan. 29th, 2024; published: Feb. 5th, 2024

Abstract

In tasks with limited processor resources such as autonomous driving and UAV, it is necessary to consider the number of parameters and operation speed of the model, and ensure good accuracy. When some semantic segmentation models adopt a parallel structure to extract multi-scale information, they use depth wise separable convolution or grouped convolution to replace conventional convolution to reduce computational complexity. However, these operations have the problem of increasing network delay and reducing inference speed. To solve this problem, a real-time

文章引用: 胡家虎, 佘玉梅. 多尺度信息融合的实时语义分割网络[J]. 人工智能与机器人研究, 2024, 13(1): 19-29.
DOI: 10.12677/airr.2024.131003

semantic segmentation model based on encoder-decoder is proposed. In the encoder stage, partial convolution combined with dilated convolution was used to construct different parallel modules for extracting multi-scale information at different stages. In the decoder stage, the up sampled features are fused. The model is tested on Cityscapes and CamVid datasets, the MIU is 71.3%and 66.8%respectively, and the running speed is 97 frames/s and 98 frames/s respectively. The results show that the model achieves a good balance between segmentation accuracy and running speed.

Keywords

Real-Time Semantic Segmentation, Partial Convolution, Multi-Scale Information, Codec Structure

1. 引言

图像语义分割(Image Semantic Segmentation)作为计算机视觉中一个重要基础性研究方向，对图像进行处理时可以实现计算机自动分割并识别出图像内容。其目的是对图像中的每个像素，按照语义类别分割为不同的区域并标记成不同的颜色，获得具有像素语义标注的图像。近年来，图像语义分割技术在自动驾驶、无人机、医学影像分析等领域都有着重要的应用价值。这些应用要求对图像进行实时语义分割，即网络的运行以每秒至少 30 帧的速度对图像进行处理，从而实现对场景的实时感知与理解。实时语义分割网络要求在处理器资源受限的情况下，构建一种“轻量且高效”的模型对图像进行快速处理并能保证准确性，因此实时语义分割面临更大的计算效率和速度的挑战。

Long [1]等提出了全卷积网络(FCN)，将卷积神经网络[2] [3] [4]的全连接层替换为卷积层，解决了语义级别的分割问题。但因其在下采样中丢失了大量的信息，导致最后的分割结果粗糙。为了防止丢失空间细节信息，Ronneberge [5]等提出了 U-Net 编码器–解码器网络，使用跳跃连接融合编码器中的各层特征信息。为了减少在下采样时空间信息的丢失，Chen 等[6] [7] [8] [9]提出了 DeepLab 系列网络，Depplabv1 利用空洞卷积层取代普通卷积层，无需下采样而增加大网络的感受野；Depplabv2 引入空洞金字塔池化结构(ASPP)以整合多尺度特征信息，提高分割适应性。Deeplabv3、Deeplabv3+，优化了 ASPP 模块和网络结构，在精度和速度方面表现优异。随着深度学习发展，一些模型不断增加网络层数和参数数量，使网络分割精度提高。但这些模型存在计算复杂度高，导致运行速度低的问题，难以应用于动态场景中的实时语义分割任务。因此，研究实时语义分割以实现对动态场景的低延迟感知非常必要，对于低延时较高精度的实时语义分割网络模型成为了研究的重点。

Zhao 提出的 ICNet [10]通过图像金字塔技术实现三种不同尺度的图像输入，有效融合多尺度特征信息提高分割效果。Li 提出的 DFANet [11]模型在主干网络部分使用修改后的 Xception [12]用于特征提取，并以级联方式聚合特征。SFNet [13]模型在编码器中使用 ResNet [14]、ShuffleNetV2 [15]作为主干网络用于特征提取，并提出了 FAM 来学习解码器中的语义信息。Yu 等提出的 BiSeNet [16]在上文下文路径中使用 Xception [12]作为特征提取主干，并设计了一个特征融合模块(FFM)来合并特征信息。这些轻量级语义分割网络模型的主干网络都是采用图像分类模型的轻量级主干网络，但这些专用于分类模型的轻量级主干网络不能完全适用于语义分割模型。一些工作便设计专用语义分割的轻量级主干网络结构，如 Paszke

[17]提出的 ENet 通过引入稀疏性和设计更浅层的结构作为网络的主干，降低模型参数量加速运算。Li 提出的 DABNet [18]利用一种新颖的深度不对称瓶颈(DAB)模块来构建主干用于特征提取，它能够创建足够的感受野并密集地利用上下文信息。

对于语义分割这种精确到像素级别分类的任务，多尺度特征很重要。DABNet，Regseg [19]在模型设计时使用并行式结构结合扩张卷积[6]用于提取特征的多尺度信息，并且 DABNet 使用深度可分离卷积[20]，Regseg 使用分组卷积[21]来降低计算量，但这些操作会存在增加内存访问量，导致网络的延迟增加，推理速度降低的问题。Chen 在 FasterNet [22]中引入了简单快速有效的部分卷积(Partial convolution, Pconv)，通过同时减少冗余计算和内存访问，更有效地提取空间特征。受此启发，为了解决上述问题，本文使用部分卷积结合扩张卷积构建一个并行式模块，用于提取语义分割的多尺度特征信息，基于此模块构建了一个简单高效的实时语义分割模型，并通过对比实验验证了其有效性。

2. 网络结构介绍

2.1. 模型整体结构

本文提出的并行式部分扩张卷积网络(Parallel partially dilated convolutional networks, PPDNet)的整体结构图如图 1 所示，采用编码器—解码器架构的方式，网络结构简单，能够进行端到端的训练。编码器阶段包含 Stemblock、Stage $i(i=1,2,3)$ 四个部分，Stemblock 用于特征提取的初始处理，由三种不同的 PPDblock 以级联的方式构建的 Stage $i(i=1,2,3)$ 用于提取不同阶段的多尺度特征信息，在三个 Stage 阶段后面单独使用 3×3 卷积作为下采样单元，获得更大的感受野，捕获更丰富的上下文信息。解码器阶段采用类似于 FCN 的简单结构，首先上采样较高阶段的特征图，然后与较低阶段的特征图进行连接，最后使用 SegHead 处理连接好的特征图进行最终预测。图 1 中的 3 个维度分别是通道数、高度和宽度，N 表示数据集中标签类型的数量。PPDNet 详细的网络结构参数如表 1 所示。

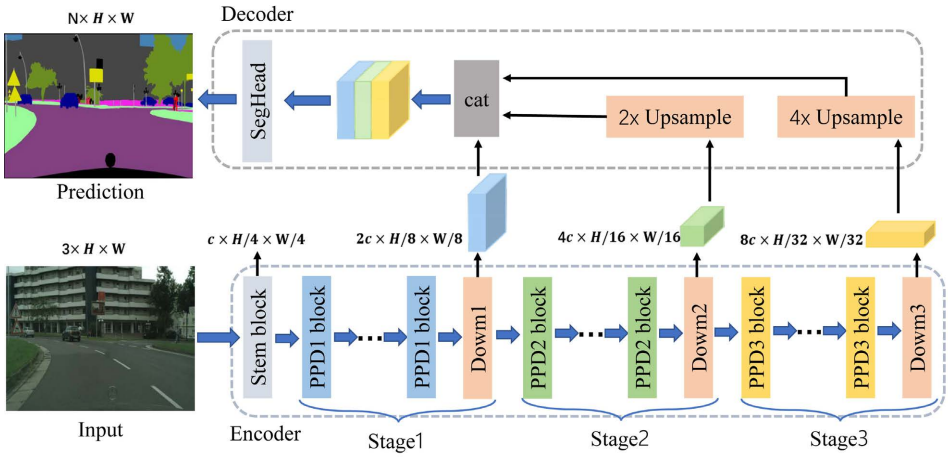

Figure 1. PPDNet structure
图 1. PPDNet 结构

2.2. 编码器阶段

2.2.1. Stem 模块

编码器初步特征提取采用 BiSeNetV2 [23]中 Stem 模块，如图 2 所示。它在两个分支使用不同方式缩小特征表示，然后将两个分支的输出特征连接起来重新加权，学习更具代表性的特征作为输出，该结构

具有高效的计算成本和有效的特征表达能力。由于初始特征提取时，通道维度过低，MobileNetV2 [24]指出使用非线性运算会对低维特征造成信息损失。故本文删除了 Stem 模块中1×1卷积的激活层与左边分支3×3卷积的归一化层和激活层，以避免信息损失。

Table 1. PPDNet structure parameters
表 1. PPDNet 结构参数

阶段	比例	输出通道数	重复数
输入	1	3	
Stem 模块	1/4	32	1
PPD1 模块	1/4	32	9
下采样	1/8	64	1
PPD2 模块	1/8	64	6
下采样	1/16	128	1
PPD3 模块	1/16	128	3
下采样	1/32	256	1
Cat	1/8	448	
解码器 3 x 3Conv	1/8	128	1
解码器 1 x 1Conv	1/8	19	1
参数量		2.55 M	

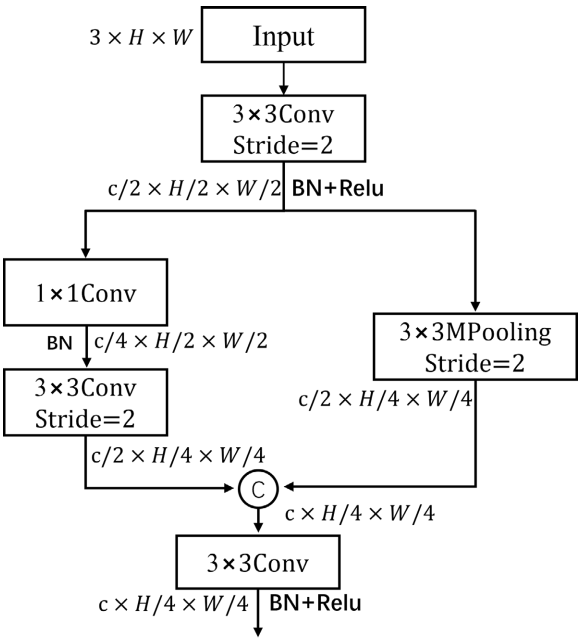

Figure 2. Stem module
图 2. Stem 模块

2.2.2. PPD 模块

FasterNet [22]引入了简单有效的部分卷积(Pconv)，如图 3 所示。对于输入的三维特征图$C\times H\times W$，它仅对部分输入通道上系统地应用常规卷积，而其余通道保持不变，通过同时减少冗余计算和内存访问，更有效地提取空间特征。其中C,H,W分别表示输入特征图的通道数、高度、宽度；c_p表示对输入进行常规卷积的通道数量，本文取$c_p=c/4$；Identity 表示特征保持不变。

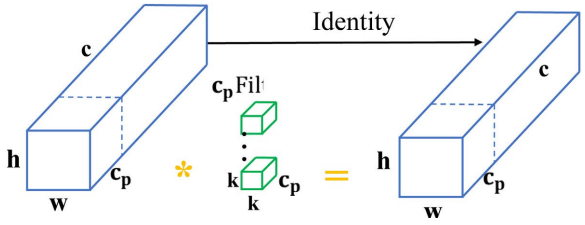

Figure 3. Partial convolution
图 3. 部分卷积

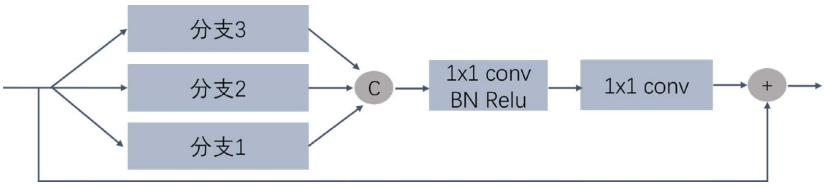

Figure 4. PPD module
图 4. PPD 模块

并行结构可以结合不同扩张率的卷积提取多尺度特征信息，但这种结构会增加计算量和内存访问量，导致推理速度变慢。一些工作[18] [19]使用深度可分离卷积和分组卷积操作来降低计算量，但这些操作又进一步增加了内存访问量，不能有效提高模型的推理速度。

故本文使用部分卷积结合扩张卷积构建一个并行式结构，称为并行式部分扩张卷积模块(PPDblock)，如图 4 所示。该模块主要用于网络模型在各个阶段的多尺度特征提取，不仅可以降低计算量和内存访问量，达到平衡的效果；还可以融合输入特征减少由扩张卷积丢失的信息。

PPDblock 的计算公式如式(1)所示。

$$\begin{aligned} f_1 &= cat\left\{b_1(x), b_2(x), b_3(x)\right\} \\ f_2 &= \mathrm{Re}\,lu\left(BN\left(1\times 1Conv\left(f_1\right)\right)\right) \\ out &= 1\times 1Conv\left(f_2\right) \end{aligned} \tag{1}$$

其中，$b_i\left(i=1,2,3,\right)$ 表示三个分支的卷积操作，x 为输入特征。

考虑到语义分割模型在不同阶段需要用到不同感受野需求的特征信息，如果在 PPDblock 的每个分支都使用部分扩张卷积，则对应感受野需求的特征信息提取过少，导致模型的分割效果不佳。故在由浅到高的三个阶段中，设计了三种的对应 PDDblock，如表 2 所示。通过多阶段获取不同层次特征再融合，提高特征的多样性和表达力。这三个模块在不同阶段使用不同组合的各类卷积，逐步减少计算量的同时保证语义特征的丰富性。第一阶段保留更多细节，第三阶段更加高效，并且后面通过消融实验验证了该模块设计的有效性。

Table 2. Details of each stage branch
表 2. 各阶段模块分支详情

阶段	模块	分支 1	分支 2	分支 3	重复数
Stage1	PPD1	3 * 3conv D = 1	3 * 3Pconv D = 3	3 * 3Pconv D = 5	9
Stage2	PPD2	3 * 3Pconv D=1	3 * 3conv D = 3	3 * 3Pconv D = 5	6
Stage3	PPD3	3 * 3Pconv D = 1	3 * 3Pconv D = 3	3 * 3conv D = 5	3

2.3. 编码器阶段

解码器阶段采用类似于 FCN 的结构，如图 1 所示。首先把 Stage2 和 Stage3 下采样的特征图使用双线性插值上采样，然后与 Stage1 下采样的特征图拼接，最后使用 SegHead 进行预测。SegHead 结构如图 5 所示，3×3 卷积合并特征图，并加入归一化和激活函数提升特征表达能力；1×1 卷积缩减通道，输出预测的类别数，最后对特征图使用双线性插值上采样到输入大小。解码器的公式如式(2)所示：

$$\begin{aligned} f_1 &= cat\left\{Up_4\left(D_3\right), Up_2\left(D_2\right), D_1\right\} \\ f_2 &= \mathrm{Re}\,lu\left(BN\left(3\times 3Conv\left(f_1\right)\right)\right) \\ Predict &= Up_8\left(1\times 1Conv\left(f\right)_2\right) \end{aligned} \tag{2}$$

其中，$D_i\left(i=1,2,3\right)$ 表示由 $Stage_i\left(i=1,2,3\right)$ 下采样后的特征图，$Up_i\left(i=2,4,8\right)$ 分别表示上采样 2 倍、4 倍、8 倍。

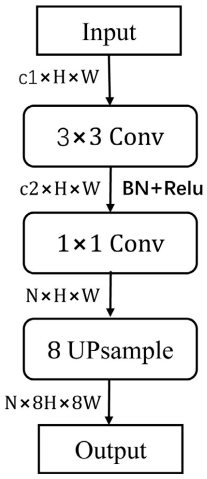

Figure 5. SegHead module
图 5. SegHead 模块

3. 实验与分析

3.1. 实验数据集

本文实验使用 2 个语义分割领域的数据集 Cityscapes [25]和 CamVid [26]。Cityscapes 是一个城市街道语义分割数据集，图像的分辨率为 1024×2048，19 个类别用于语义分割。数据集划分为训练集、验证集、测试集，分别包含 2975、500、1525 张图像。CamVid 是从驾驶骑车的角度拍摄的道路场景数据集，图像的分辨率为 720×960，11 个类别用于语义分割。数据集划分为训练集、验证集、测试集，分别包含 367、101、233 张图片。

3.2. 训练设置

3.2.1. 实验参数

本实验首先进行权重随机初始化，其余参数设置如表 3 所示。

3.2.2. 数据增强

为了扩充数据集、缓解样本不平衡问题，本文对训练集进行数据增强，采用随机水平翻转、均值减法、随机缩放和随机裁剪方式，随机缩放的比例为 $\left\{0.75,1.00,1.25,1.50,1.75,2.00\right\}$。在 Cityscapes 数据集的

训练中，将输入图片的分辨率随机裁剪为 512 × 1024 像素；在 CamVid 数据集的训练中，将输入图片的分辨率随机裁剪为360×480 像素。

Table 3. Parameter Setting
表 3. 参数设置

训练参数	大小或方法
Learning rate (学习率)	0.02
Batch size (批次)	8
Optimizer (优化器)	随机梯度下降(SGD)
Momentum (动量)	0.9
Weight decay (权重衰减)	$1e^{-4}$
Power (幂次)	0.9
Iter (迭代次数)	400

3.3. 评价指标

3.3.1. 平均交并比

平均交并比(Mean Intersection over Union, MIoU)：语义分割任务中最常用的评价指标，表示真实值与预测值的交集和并集的集合，公式如式(3)。

$$MIoU = \frac{1}{k}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij} + \sum_{i=0}^{k}p_{ji} - p_{ii}} \tag{3}$$

其中，p_{ii} 表示属于第 i 类被正确分到第 i 类的像素点，p_{ij} 表示属于第 i 类却被分到第 j 类的像素点，p_{ji} 表示属于第 j 类却被分到第 i 类的像素点，k 表示类别数。

3.3.2. 每秒传输帧数

每秒传输帧数(Framersper Second, fps)：fps 是衡量速度的指标，即图像的刷新频率。目标网络每秒可以处理或检测多少帧，为时间的倒数，公式如式(4)。

$$fps = \frac{1}{T} \tag{4}$$

3.4. 实验结果与分析

3.4.1. 消融实验

为了验证本文所提出的结构有效性，使用与本文相同的编码器－解码器网络框架。以常规扩张卷积搭建三分支模块为基准，通过对深度可分离扩张卷积，分组扩张卷积，以及本文提出的结构进行了消融实验验证。实验在 Cityscapes 训练集上进行训练，在测试集进行测试，结果如表 4 所示。

Table 4. Results of ablation experiment
表 4. 消融实验结果

卷积方式	MIoU%	fps	参数量×10^{-6}
Dconv	71.5	86	3.95
DWconv + Dconv	66.8	144	1.74
Gconv + Dconv	68.3	104	2.27
Pconv + Dconv	67.8	125	1.46
本文	**71.3**	**97**	**2.55**

实验结果表明，在相同编码器–解码器网络框架下，各分支中使用不同卷积方式的效果对比。表 4 可以看出，用于提取多尺度信息的三分支模块，在各分支使用常规扩张卷积可以使模型达到不错的分割效果和运行速度，但模型参数量偏高。虽然使用深度可分离卷积或分组卷积结合扩张卷积使模型参数量减少 57%和 43%，提高了模型的推理速度，但 MIoU 下降了 4.7%和 3.2%，使模型不能达到有效平衡。部分卷积结合扩张卷积的方式使 MIoU 下降 3.7%，但参数量却降低了 63%。考虑到在各阶段的所有分支都使用部分卷积，会导致信息提取太少，使得分割效果不佳。于是本文提出了特定于各阶段的三分支模块，与基准模型相比，MIoU 只降低 0.2%，并且在相似的推理速度下，参数量下降 35%，使模型能都达到有效的平衡。

3.4.2. 对比实验

为进一步验证所提出实时语义分割网络的有效性，在 Cityscapes 数据集上进行分割效果可视化，如图 6 所示，从左到右为数据集的原图，真实标签，预测图。与一些先进实时语义分割网络在 Cityscapes 数据集上和 CamVid 数据集上进行性能对比，本模型没有进行预训练和使用额外的数据集进行预训练，对比结果如表 5 和表 6 所示。

Figure 6. Cityscapes segmentation visualization
图 6. Cityscapes 分割效果可视化

Table 5. Comparison results on Cityscapes
表 5. 在 Cityscapes 上的对比结果

实时语义分割网络	MIOU%	fps	参数量
ENet [17]	58.3	77	0.36
ESPNet [27]	60.3	112	0.36
CGNet [28]	64.8	50	0.50
ERFNet [29]	68.0	42	2.10
BiseNet [16]	68.4	106	5.80
DABNet [18]	70.1	104	0.76
LEDNet [30]	70.6	30	26.50
ICNet [13]	70.6	71	0.92
PPDNet	**71.3**	**97**	**2.55**

在 Cityscapes 训练集和验证集上对 PPDNet 进行训练，并在测试集上进行得到验证结果。实验结果由

表 4 可以看出，本文所提出的 PPDNet 网络模型在 Cityscapes 数据集上的 MIoU 达到了 71.3%，运行速度达到 97 帧/s，在 MIoU 评价指标方面优于所对比的实时语义分割网络模型。在网络效率方面，PPDNet 的运行速度分别比 CGNet 和 LEDNet 快 194%和 323%，虽然 ESPNet 和 BiseNet 比 PPDNet 的运行速度快 115%和 109%，但其 MIoU 下降了 10%和 2.9%。在网络参数量方面，PPDNet 和 ERFNet 在相似参数量的情况下，PPDNet 的 MIoU 提高了 3.3%，且运行速度比 ERFNet 快 217%。

Table 6. Comparison results on CamVid
表 6. 在 CamVid 上的对比结果

实时语义分割网络	MIOU%	fps
ENet [17]	51.3	61
CGNet [28]	65.6	50
BiseNet [16]	65.5	
DABNet [18]	66.4	112
PPDNet	**66.8**	**98**

在 CamVid 训练集和验证集上对 PPDNet 进行训练，并在测试集上进行得到验证结果。实验结果由表 6 可以看出，本文所提出的 PPDNet 网络模型在 CamVid 数据集上的 MIoU 达到了 66.8%，运行速度达到 98 帧/s，在 MIoU 评价指标方面优于所对比的实时语义分割网络模型。

4. 总结

文使用部分卷积结合扩张卷积的操作构建了一个用于多尺度信息提取的不同 PPD 模块，基于此模块采用编码器－解码器结构，提出一种简单高效的 PPDNet 用于实时语义分割。通过在 Cityscapes 和 CamVid 数据集上的实验结果表明，PPDNet 的效果优于大多数实时语义分割网络，充分表明了 PPDNet 在分割精度、网络规模和运行速度方面达到了较好的平衡。在后续工作中将对网络结构进行继续优化，进一步提高网络的精度和运行速度。

参考文献

[1] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. 2015 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Boston, 7-12 June 2015, 3431-3440. https://doi.org/10.1109/CVPR.2015.7298965

[2] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, **25**, 1097-1105.

[3] Simonyan, K. and Zisserman, A. (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. 3*rd International Conference on Learning Representations* (*ICLR* 2015), San Diego, 7-9 May 2015, 1-14.

[4] Szegedy, C., Liu, W., Jia, Y., *et al.* (2014) Going Deeper with Convolutions. 2015 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Boston, 7-12 June 2015, 1-9. https://doi.org/10.1109/CVPR.2015.7298594

[5] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI* 2015, Springer, Cham, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[6] Chen, L.C., Papandreou, G., Kokkinos, I., *et al.* (2014) Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. *Computer Science*, **4**, 357-361.

[7] Chen, L.C., Papandreou, G., Kokkinos, I., *et al.* (2018) DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **40**, 834-848. https://doi.org/10.1109/TPAMI.2017.2699184

[8] Chen, L.C., Papandreou, G., Schroff, F., *et al.* (2023) Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv: 1706.05587.

[9] Chen, L.C., Zhu, Y.K., Papandreou, G., *et al.* (2018) Encoder-Decoder with Atrous Separable Convolution for Seman-

tic Image Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV* 2018, Springer, Cham, 833-851. https://doi.org/10.1007/978-3-030-01234-2_49

[10] Zhao, H.S., Qi, X.J., Shen, X., Shi, J. and Jia, J. (2018) Icnet for Real-Time Semantic Segmentation on High-Resolution Images. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV* 2018, Springer, Cham, 418-434. https://doi.org/10.1007/978-3-030-01219-9_25

[11] Li, H.C., Xiong, P.F., Fan, H.Q. and Sun, J. (2019) Dfanet: Deep Feature Aggregation for Real-Time Semantic Segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 9522-9531. https://doi.org/10.1109/CVPR.2019.00975

[12] Chollet, F. (2017) Xception: Deep Learning with Depthwise Separable Convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 1800-1807. https://doi.org/10.1109/CVPR.2017.195

[13] Li, X.T., You, A.S., Zhu, Z., *et al.* (2002) Semantic Flow for Fast and Accurate Scene Parsing. In: Vedaldi, A., Bischof, H., Brox, T. and Frahm, J.M., Eds., *Computer Vision—ECCV* 2020, Springer, Cham, 775-793.

[14] He, K.M., Zhang, X.Y., Ren, S.Q., and Sun, J. (2016) Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/CVPR.2016.90

[15] Ma, N.N., Zhang, X.Y., Zheng, H.T. and Su, J. (2018) Shufflenetv2: Practical Guidelines for Efficient CNN Architecture Design. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV* 2018, Springer, Cham, 122-138.

[16] Yu, C.Q., Wang, J.B., *et al.* (2018) BiSeNet: Bilateral Segmentation Network for Real-Time Semantic Segmentation. . In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV* 2018, Springer, Cham, 334-349.

[17] Paszke, A., Chaurasia, A., Kim, S. and Culurciello, E. (2016) ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation. arXiv: 1606.02147.

[18] Li, G., Yun, I.Y., Kim, J. and Kim, J. (2019) Dabnet: Depth-Wise Asymmetric Bottleneck for Real-Time Semantic Segmentation. arXiv: 1907.11357.

[19] Gao, R. (2021) Rethinking Dilated Convolution for Real-time Semantic Segmentation. arXiv: 2111.09957.

[20] Howard, A.G., Zhu, M.L., *et al.* (2017) MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv: 1704.04861.

[21] Xie, S.N., Girshick, R., *et al.* (2023) Aggregated Residual Transformations for Deep Neural Networks. arXiv: 1611.05431.

[22] Chen, J.R., Kao, S.H., *et al.* (2023) Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Vancouver, 17-24 June 2023, 12021-12031. https://doi.org/10.1109/CVPR52729.2023.01157

[23] Yu, C.Q., Gao, C.X., *et al.* (2021) Bisenet V2: Bilateral Network with Guided Aggregation for Real-Time Semantic Segmentation. *International Journal of Computer Vision*, **129**, 3051-3068. https://doi.org/10.1007/s11263-021-01515-2

[24] Sandler, M., Howard, A., Zhu, M, L., *et al.* (2018) MobileNetV2: Inverted Residuals and Linear Bottlenecks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 4510-4520. https://doi.org/10.1109/CVPR.2018.00474

[25] Cordts, M., Omran, M., Ramos, S., *et al.* (2016) The Cityscapes Dataset for Semantic Urban Scene Understanding. 2016 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 3213-3223. https://doi.org/10.1109/CVPR.2016.350

[26] Brostow, G.J., Shotton, J., Fauqueur, J., *et al.* (2008) Segmentation and Recognition Using Structure from Motion Point Clouds. In: Forsyth, D., Torr, P. and Zisserman, A., Eds., *Computer Vision—ECCV* 2008, Springer, Berlin, 44-57. https://doi.org/10.1007/978-3-540-88682-2_5

[27] Mehta, S., Rastegari, M., Caspi, A., *et al.* (2018) ESPNet: Efficient Spatial Pyramid of Dilated Convolutions for Semantic Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV* 2018, Springer, Cham, 552-568. https://doi.org/10.1007/978-3-030-01249-6_34

[28] Wu, T.Y., Tang, S., Zhang, R., *et al.* (2021) CGNet: A Light-Weight Context Guided Network for Semantic Segmentation. *IEEE Transactions on Image Processing*, **30**, 1169-1179. https://doi.org/10.1109/TIP.2020.3042065

[29] Romera, E., Alvarez, J.M., Bergasa, L.M., *et al.* (2017) ERFNet: Efficient Residual Factorized Convnet for Real-Time Semantic Segmentation. *IEEE Transactions on Intelligent Transportation Systems*, **19**, 263-272. https://doi.org/10.1109/TITS.2017.2750080

[30] Wang, Y., Zhou, Q., Liu, J., *et al*. (2019) Lednet: A Lightweight Encoder-Decoder Network for Real-Time Semantic Segmentation. *Proceedings of the IEEE International Conference on Image Processing*, Taipei, 22-25 September 2019, 1860-1864. https://doi.org/10.1109/ICIP.2019.8803154