基于卷积神经网络的低参数量实时图像分割算法

谭光鸿，侯进*，韩雁鹏，罗朔
西南交通大学信息科学与技术学院，四川 成都 611756

摘要 提出了一种低参数量实时图像语义分割网络模型 Atrous-squeezeseg。模型在最低参数量为 2.1×10^7 时的运算帧率为 45.3 frame/s，像素点准确度与均交并比分别可达到 59.5%与 62.9%。同时，嵌入式设备 NVIDIA TX2 的运算帧率可达 8.3 frame/s。实验结果表明，相比于其他分割算法，所提模型的速度和参数量均得到了提升。

关键词 图像处理；图像分割；实时图像；低参数量；卷积模块；多尺度特征
中图分类号 TP391 **文献标识码** A **doi**：10.3788/LOP56.091003

Low-Parameter Real-Time Image Segmentation Algorithm Based on Convolutional Neural Network

Tan Guanghong, Hou Jin*, Han Yanpeng, Luo Shuo
School of Information Science and Technology, Southwest Jiaotong University, Chengdu, Sichuan 611756, *China*

Abstract We propose a real-time image semantic segmentation network model, which is named as Atrous-squeezeseg. Under the condition that the minimum parameter of the model is 2.1×10^7, the operation frame rate is 45.3 frame/s, and the pixel point accuracy and mean intersection over union can reach 59.5% and 62.9%, respectively. At the same time, in the embedded device NVIDIA TX2, the operate frame rate is up to 8.3 frame/s. The experimental results show that, compared with other segmentation algorithms, the speed and parameter quantity of the proposed model are increased.

Key words image processing; image segmentation; real-time image; low number of parameters; convolution module; multiscale feature
OCIS codes 100.3020; 100.4996; 170.3010

1 引 言

随着卷积神经网络(CNN)的不断发展，出现了诸多高精度的网络模型。Simonyan 等[1]提出的 VGGnet 是在 Alexnet[2]8 层卷积层的基础上提升到了 19 层；He 等[3]提出的 Resident Network (ResNet)，解决了模型层数加深导致梯度消失的技术难题，其均为 CNN 的阶段性成果。CNN 被广泛地运用在图像处理领域，如图像识别[4]、目标检测[5]和图像分割[6-7]。其中，图像语义分割解决的是语义和位置问题，分别用于预测类别信息和确定位置信息。Shelhamer 等[8]根据全卷积神经网络(FCN)提出了一种端到端的技术从而实现语义图像分割，首次尝试将全卷积神经网络运用到图像语义分割中。FCN 能将生成与输入图像成比例的分数映射，其中每个像素代表一个分类目标。Badrinarayanan 等[9]提出了逐像素语义分割 Segnet 网络，Gamal 等[10]与宋青松等[11]沿用 FCN 的思路对图像进行分割。但上述模型的编译码器过于复杂，参数量多，存在较大冗余，难以被高效地部署。

近年来，对于深度学习的研究呈现良好的发展趋势，随着数据的增加，其能覆盖到的场景也越来越丰富，而学者们也在不断探索使用更精简的网络去消化庞大的数据，同时能够保障网络较高的精准度以便于实际部署[12]。尽管图像分割在诸多应用场景中都具有明显的优势，但目前对实时语义分割网

收稿日期：2018-10-22；**修回日期**：2018-11-29；**录用日期**：2018-11-30
基金项目：浙江大学 CAD&CG 国家重点实验室开放课题(A1923)、成都市科技项目(2015-HM01-00050-SF)
* **E-mail**：jhou@swjtu.edu.cn

络的研究仍较少。实现图像分割网络实际部署的基本要求包括:1) 速度。网络能实现实时或更高的推算速度,以分担硬件控制回路的延时压力。2) 准确度。准确度是判断网络优劣的核心要求,在理想情况下应该实现 100%的预测准确度。3)模型规模。较小的网络模型规模能带来更高效的分布式训练,降低部署时的系统内存占用量。

综上所述,本文提出了一种基于低参数量的实时图像分割网络 Atrous-squeezeseg,该网络不仅能满足分割精度的要求,而且还能降低模型参数量,且能部署在嵌入式设备 NVIDIA TX2 中,其帧率可达 8.3 frame/s。还设计了一种参数量低、融合多尺度特征的网络模块 Atrous-Fire,并在分割网络中使用该模块搭建了编码器,配合跳跃结构向译码器传递高分辨率特征图。本文模型与文献[8]中的分割模型相比,参数量下降了 68.1%,帧率提高了 53.0%,分割图像像素准确率与均交并比(MIU)分别可达 59.5%与 62.9%。

2 模型优化

Shi 等[13]提出的 N-cut 算法是基于图论的分割标准来衡量图像分割质量,将图像分割问题转化为求特征向量的问题。CNN 的运用使得基于深度学习的图像语义分割算法的表现效果得到了大幅度提升。如文献[8]中运用全卷积层对图像中每一个像素点进行分类,最终有效地对图像进行了分割,但模型参数量大、计算成本高、运算速度过慢、实际部署时开销过大等问题一直阻碍着网络模型的实时处理和有效部署。

为将模型轻松部署在嵌入式中,应设计一种稳定性能更强的网络模块来降低模型冗余、提高推算速度。本文模块通过多尺度卷积方式、网络增强优化搭建了一种基于 CNN 的低参数量实时图像分割算法。

2.1 低参数模型

深度学习模型在保证准确度的前提下实现在嵌入式设备中部署,已然成为当前热门的研究方向。而 SqueezeNet[12] 能够在保证精准度达到 Alexnet 的基础上将模型规模压缩为 1/50。该网络采用模块化的设计,既保证了模型的精度,又压缩了模型的参数量,为模型部署在嵌入式设备中提供了可行性的研究思路。模块化同样也是 CNN 宏观架构中新兴的研究领域,有效的模块化设计能提升网络的结构性和稳定性。但 SqueezeNet 中的模块设计未全方位考虑模块的稳定性,导致网络易出现过拟合、网络准确度难以提升等问题。

因此,为了使模型具有低参数量,且能更好地运用在图像分割模型中,本文改进了 SqueezeNet 中的 Fire 模块,为所提图像分割模型设计了 Atrous-Fire 模块。使用 1×1 卷积核对图像进行特征提取,通过添加批量标准化层(BN)[14-15]来抑制网络过拟合,并采用感受野尺度可控的卷积方式,优化图像分割效果。

2.2 多尺度卷积与网络优化

在卷积神经网络中卷积层主要负责提取图像特征,池化层通过下采样简化网络计算的复杂度。随着网络层的加深,单位像素包含原始输入图像的信息比重加大,网络的感受野也将加大,但降低特征图的分辨率将会造成部分原始图像信息流失。故传统的卷积神经网络为了提高卷积的感受视野,采用较大的卷积核,导致模型参数量增加,易造成模型冗余,不利于部署。

本文为了在不增加模型参数量的基础上提高网络的感受视野,在网络模块中使用多尺度的空洞卷积[16]来改善上述问题。空洞卷积能改变卷积核的感受野,且能在不增加卷积核参数量的同时扩大卷积网络的感受野。

图 1(a)为经典 3×3 卷积核的感受野范围;图 1(b)、(c)分别为 $R_{\mathrm{rate}}=2$、3 时与 3×3 卷积核具有相同参数量的空洞卷积核,感受野分别增大到 7×7 和 11×11。感受野大小的计算公式为

$$v=[(K_{\mathrm{size}}+1)\times(R_{\mathrm{rate}}-1)+K_{\mathrm{size}}]^2, \quad (1)$$

式中:K_{size}为卷积核尺寸,通过 R_{rate}控制空洞卷积感受野范围 v。采用不同 R_{rate}结构的多尺度空洞卷积组,增强网络对图像的全局特征和细节特征的敏感度,并控制模型参数量不会增加。

深度学习的主要研究热点是如何优化深度学习模型,提升模型性能,其中网络模型在模型训练中拟合了训练数据中的噪声和没有代表性的特征,使得验证和测试样本精度是否降低成为棘手的问题。为了抑制网络的过拟合,在网络中加入了批量标准化层。批量标准化层在训练过程中可使在同一个数据批次中的所有样本都被关联在一起,网络不会从单一训练样本中生成确定性结果。同时,样本在超平面上被变换重构,每次重构方向的大小均有不同,增强了样本的多样性。

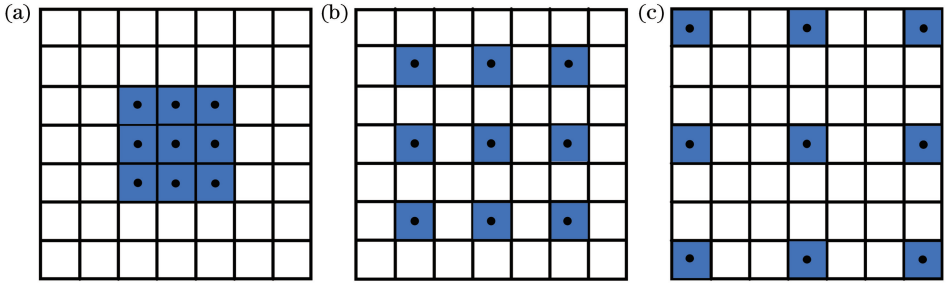

图 1 卷积核。(a)经典卷积核;(b)空洞卷积核 $R_{rate}=2$;(c)空洞卷积核 $R_{rate}=3$

Fig. 1 Convolution kernel. (a) Classical convolution kernel; (b) dilated convolution kernel $R_{rate}=2$; (c) dilated convolution kernel $R_{rate}=3$

3 Atrous-squeezeseg 网络模块设计

3.1 Atrous-Fire 模块结构

文献[12]设计的 SqueezeNet 网络中使用了 Fire 模块结构。该结构主要使用了 3 种策略来保证低参数量并维持模型的准确度:1) 在 Squeeze 层中使用 1×1 卷积核替换 3×3 卷积核,因为 3×3 卷积核的参数是 1×1 卷积核的 9 倍;2) 减少 Expand 层中 3×3 卷积核的输入通道数,模块输出通道由 3×3 卷积核通道与 1×1 卷积核通道叠加融合;3) 网络后期进行下采样,使卷积层有较大的特征图。

由于 SqueezeNet 网络设置的卷积核较小,因此卷积层感受野较小,对图像的全局和细节特征的敏感度不够,无法进行多尺度特征的提取,且模块的稳定性较差。在模型训练实验中发现,即使在该网络训练的中后期加入了 Dropout 机制来抑制网络过拟合,但还是易发生过拟合现象,最终导致网络的准确度无法进一步提高。

针对 Fire 模块存在的缺点,设计了稳定性更佳的 Atrous-Fire 模块,如图 2 所示。Atrous-Fire 模块通过使用多尺度的空洞卷积来扩展卷积核的感受视野且能保证模块的参数量不增加,并在网络模块中添加批量标准化层来抑制网络过拟合,提高了模块的稳定性和训练效率。该模块结构为:1) 对 Squeeze 层输出特征图进行批量标准化处理,优化 Squeeze 层结构的稳定性;2) 输入到包含有 1×1 卷积核和感受视野可变的 3×3 空洞卷积 Expand 层中,该层的主要功能是融合使用不同感受野卷积核提取的多尺度特征。

Atrous-Fire 模块保持了模型的低参数量、扩大了模型卷积核的感受视野,同时批量标准化层能优化网络训练过程。实验证明,相比于 Fire 模块,使用本文模块的网络能加快模型的收敛速度,且能有效地抑制网络过拟合,提高网络精度。

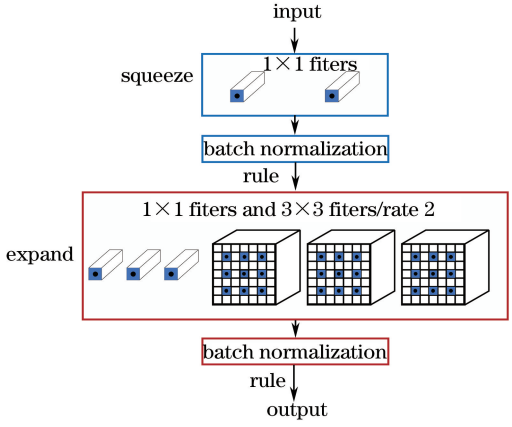

图 2 Atrous-Fire 模块结构

Fig. 2 Atrous-Fire modular structure

3.2 空洞卷积感受野

尽管空洞卷积能扩大卷积感受野、降低网络参数量,但使用不合理的空洞卷积组将会导致模型输出分割图出现栅格现象,影响模型精度。为防止出现栅格现象,本文空洞卷积组采用 R_{rate} 不可约的锯齿结构,如图 3(a)所示。由锯齿结构空洞卷积组可得到无栅格现象的初始特征图[图 3(b)]。而当 R_{rate} 设计为包含公约数的非锯齿结构时,就可能出现栅格现象,如特征图 3(c)所示。栅格现象出现的原因是不连续的空洞卷积核在提取图像特征时出现了遗漏点,从而导致了细节特征流失。针对上述问题,本文在编码器中使用了不可约的锯齿状空洞卷积 R_{rate} 结构,多尺度的空洞卷积核扩展了感受野,增强了卷积核对细节特征的灵敏性,同时连续的卷积核在滑动过程中能够关联特征图像素内联特性。

4 Atrous-squeezeseg 网络架构

用于图像语义分割的模型网络架构 Arous-squeezeseg 主要包括3部分:1)编码器架构(EA)

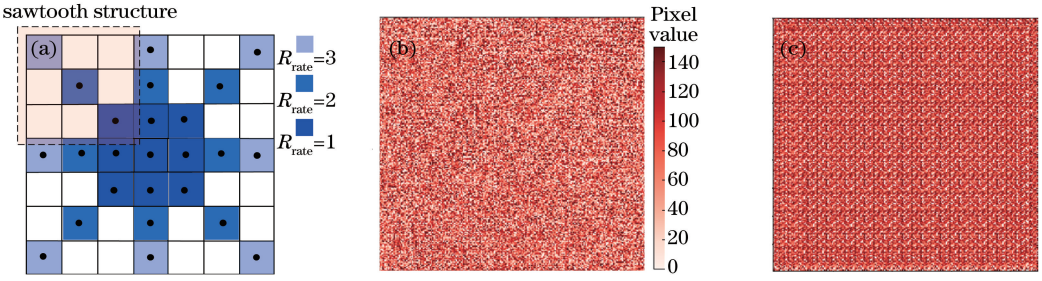

图 3 空洞卷积核与初始特征图。(a)锯齿结构卷积核;(b)无栅格特征图;(c)栅格特征图

Fig. 3 Dilated convolution kernel and initial characteristic graphs. (a) Sawtooth structure convolution kernel; (b) no grid feature graph; (c) grid feature graph

提取图像内部特征;2) 跳跃结构(SS)连接低级特征和高级特征的网络特征图从而提高模型的分割精度;3) 译码器架构(DA)对低分辨率特征图进行反卷积上采样运算并分类图像中每个像素点。

1) EA。本文编码器架构主要由 Atrous-Fire 模块构建,如图 4 中蓝色虚线框所示。首先,将分辨率为 $H\times W$ 的原始 RGB 图像经过卷积核大小为 7×7的卷积层从而获得初级特征,再使用最大值池化降低特征图分辨率。随后,初始特征图进入包含 3 组 Atrous-Fire 模块的 Atrous-Fire1 结构中。在每个 Atrous-Fire 模块中使用较多的 1×1 的卷积核对图像的每个像素点进行特征提取,使用感受野尺度不同的空洞卷积提取特征图中深层多尺度特征,提高模型对全局与细节特征的灵敏度。编码器 Atrous-Fire 模块在输出前融合不同卷积感受野的多尺度特征图,以此增强网络提取特征的能力。编码器网络的各层输出规模、模块中 squeeze(S1)与 expand(E1/E3)层通道数,以及使用锯齿状的空洞卷积尺度 R_{rate} 的具体参数设置如表 1 所示。在编码器与译码器之间添加 score layer,使用 1×1 卷积核提取低分辨率特征图中每个像素点的细节特征,该层的输出通道数设置与分割类别数 C 相同。

表 1 编码器参数

Table 1 Encoder parameters

Layer name	Output size	Squeeze (S1)	Expand (E1/E3)	R_{rate}
Input image	224×224×3			
Conv1	112×112×64			
Maxpool1	56×56×64			
Atrous-Fire1 (3×)	56×56×256	16	32	2/5/7
Maxpool2	28×28×256			
Atrous-Fire2 (3×)	28×28×256	32	64	2/3/5
Maxpool3	14×14×256			
Atrous-Fire3 (3×)	14×14×256	64	128	2/3/5
Atrous-Fire4 (2×)	14×14×512	128	256	1/2
Atrous-Fire4 (2×)	14×14×512	128	256	1/1

2) SS。网络中设计了 Atrous-Fires-S 跳跃结构,通过连接网络中的低级特征和高级特征,提高网络对低级的细节特征的灵敏性,以此提高分割精度。在跳跃结构中从 Atrous-Fire1、Atrous-Fire2 和 Atrous-Fire4 结构中各取出一层作为跳跃结构的输入层。使用 Atrous-Fire 模块作为跳跃连接层,并令 Expand 层输出通道数和模块输入通道数相同,叠加到 deconv3、deconv2 和 deconv3 反卷积特征图中,如图 4 所示。通过跳跃结构融合了网络的初级特征和高级特征,保证了特征信息不被流失。

3) DA。图像分割网络的译码器的主要工作是对低分辨率特征图进行反卷积,通过上采样恢复分辨率获得语义分割图,如图 4 中红色虚线框所示。与 U-net[17]、FCN[8]网络中译码方法不同,为了强化特征图在译码器反卷积过程中的图像特征信息,本文译码器采用跳跃结构叠加融合 Atrous-Fire 与 deconv 的 3 组特征图,通过上采样可得到 $H\times W$ 输出特征图。最后使用 deconv3 层将输出特征图重新塑造为 $H\times W\times C$(C 为分割类别数),并应用像素级 argmax 操作,来获得最终的语义分割图,公式如下:

$$o_p=\arg\max(\mathrm{deconv3}_{H\times W\times C})。\tag{2}$$

为了提高分割类别的准确度,优化网络训练,本文模型对 $\mathrm{deconv3}_{H\times W\times C}$ 使用 softmax 函数计算特征图中每个像素点属于某一类分割类别的概率,再利用交叉熵作为模型训练优化时的损失值函数,公式如下:

$$\mathrm{soft\ max}(x_i)=\exp(x_i)/\Big[\sum_j^C\exp(x_j)\Big],\tag{3}$$

$$H'_y(y)=-\sum_i y'_i\log_2(y_i),\tag{4}$$

式中:x_i 为模型 $\mathrm{deconv3}_{H\times W\times C}$ 的输出值;$\sum_j^C\exp(x_j)$ 为待分割类别中某一类像素点之和;y'_i 为标签中的第 i 类;softmax 归一化输出向量中对应的分量为 y_i。由(4)式可知,当模型分类越准确,y_i

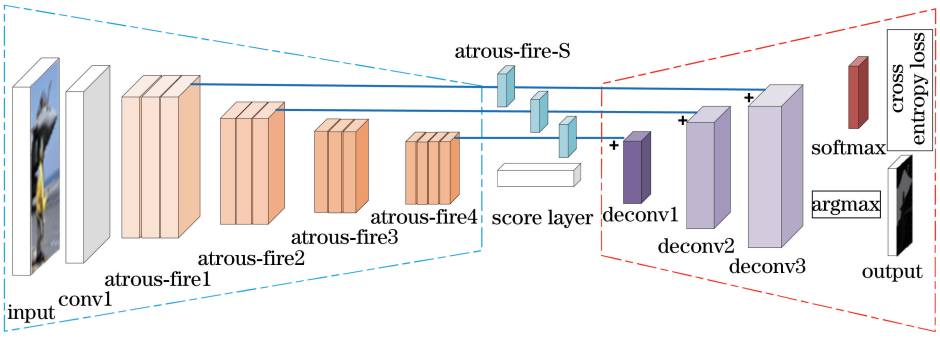

图 4 Atrous-squeezeseg 网络结构

Fig. 4 Network structure of Atrous-squeezeseg

对应的分量就会越接近 1，损失函数 $H'_y(y)$ 就会越小，差异性也就越小。

5 实验结果与分析

本文实验模型的训练、验证和测试均由 TensorFlow[18] 框架搭建，并使用 cuDNN[19] 内核计算。硬件设备主要包括高性能的工作站主机和嵌入式设备 NVIDIA TX2。其中工作站配置为 Inter © CoreTM i7-6800K CPU @ 3.40 GHz，GTX 1080Ti 显卡，负责模型的训练、验证与测试。NVIDIA TX2 为嵌入式级的硬件配置，主要负责部署和测试模型，其中图形处理器(GPU)采用基于 Nvidia Pascal 架构的 GPU 核，CPU 为双 64 位 Nvidia Denver2，四核 ARM A57。

5.1 数据集与预处理

所提模型主要在 ADE20K 数据集[20] 中训练、验证和测试，该数据集中包含多种分割场景，共 40 千张，其中包括室内、道路、建筑、车辆和行人等，共 150 类语义类别。数据集包含两类图像，一类为场景的彩色图像，另一类为对场景进行标注后的标签图。为了使模型能够得到充分学习，提高模型的稳健性，抑制过拟合，本文对训练数据样本进行了数据增强处理，包括翻转图像、改变亮度、添加噪声等。

5.2 评价指标

为了保证实验数据的严谨性，采用文献[21]中严格的评价指标对模型进行评估。本文评价指标主要包括像素准确率(PA)、MIU、模型规模大小，以及计算速度。具体计算公式为

$$T_i = \sum_{j=0}^{c} P_{ij}, \tag{5}$$

$$P_A = \sum_{i=0}^{c} p_{ii} / \sum_{i=0}^{c} T_i, \tag{6}$$

$$M_{IU} = \frac{1}{c} \sum_{i=0}^{c} p_{ii} / \left(T_i + \sum_{j=0}^{c} p_{ji} - p_{ii}\right), \tag{7}$$

式中：P_{ij} 为属于类别 i 的像素点被预测为类别 j 的像素点数量；c 为类别总数；T_i 为类别 i 的像素点的总数量之和。据此可得(6)式预测正确的像素数量占总像素的比例 P_A。在(7)式中通过计算每个类的真实像素值和预测值的交集与并集之比，并求取平均值即为 M_{IU}。

5.3 对比实验结果与分析

为了验证本文模型的优越性，分别与不同的分割网络模型进行对比，实验数据输入大小和维度与表 1 保持一致。对比模型主要包括文献[8]中以 VGG16[1] 网络作为编码器配合 FCN 进行语义分割的 VGG16+FCN 模型，以及使用文献[12]中的网络 SqueezeNet 作为编码器与 FCN 共同完成的图像分割模型 SqueezeNet+FCN。同时，为了验证本文模型的先进性，实验中添加了由 Atrous-Fire 模块搭建的模型结合 FCN 实现的分割模型 Squeezeseg+FCN。

表 2 描述了不同模型完成训练后的模型参数量、M_{IU}，以及主要类别的 Intersection over Union (IU)值。由表 2 可知，所提模型 Atrous-squeezeseg 的参数量仅为 VGG16+FCN 模型的 31.8%。且相比于由 Fire 模块搭建的 Squeezeseg+FCN 网络分割评价指标 M_{IU} 提升了 5%，同时保持了网络的参数量不变。由此可知，本文设计的 Atrous-Fire 模块参数量低，提取特征的能力更强。

为了分析所提 Atrous-Fire 模块的优越性，单独设计了两组对比实验：1) Atrous-squeezeseg (without dilated)模型舍去 Expand 层中空洞卷积，但保留批量标准化层；2) Atrous-squeezeseg (without BN)模型在Expand层中保留空洞卷积，

表 2 不同语义分割模型的参数量与 M_{IU}

Table 2 Number of parameters of different semantic segmentation models and M_{IU}

Method	Number of parameters	M_{IU}	Building	Sky	Car	Tree	Road	Person	Floor	Wall
Atrous-squeezeseg	21.09	62.9	67.5	84.0	61.4	58.1	64.7	49.1	60.4	58.5
Squeezeseg+FCN	54.65	55.9	61.8	85.8	48.4	51.8	61.5	32.3	53.8	52.2
VGG16+FCN	66.21	63.2	68.3	86.8	61.1	58.2	66.0	48.5	58.3	57.4
SqueezeNet+FCN	54.65	50.5	46.7	83.8	44.7	51.8	55.5	28.0	47.3	46.7
Atrous-squeezeseg (without dilated)	21.09	50.6	51.1	83.5	41.2	43.8	53.8	29.7	51.4	50.1
Atrous-squeezeseg (without BN)	21.09	51.6	51.3	83.5	43.6	45.8	58.0	29.8	50.1	51.3

但不再对 Squeeze(S1)层和 Expand(E1/E3)层进行批量标准化处理。由该组对比实验可知，多尺度空洞卷积和批量标准化处理两种优化算法既可提升分割网络性能，又能保持模型的低参数量。因此，使用本文模块搭建的模型不仅能提升网络性能、优化分割效果，还能维持网络的低参数量。

所提模型不仅能在 M_{IU} 指标中保持较好水平，而且 P_A 指标与不同设备中运算帧率（FPS）也表现较好。表 3 描述了模型在 GTX 1080Ti 主机与嵌入式设备 NVIDIA TX2 中 FPS 的不同表现，以及像素准确率 P_A。其中 VGG16[1]+FCN[8]模型的 FPS 值在不同设备中表现不佳，特别是在 NVIDIA TX2 中运行速度仅为 1.9 frame/s，且模型参数过大导致内存占用较大，系统出现了卡顿现象，P_A 值为 59.8%。然而本文设计的 Atrous-squeezeseg 模型在不同设备中表现效果良好，在 GTX 1080Ti 主机中运算速度能超过实时要求，且由于模型参数占用 NVIDIA TX2 内存空间较小、执行效果流畅，P_A 值高达 59.5%。

表 3 P_A 与不同设备中模型的 FPS 值

Table 3 P_A and FPS of model in different devices

Method	FPS /(frame·s^{-1})		P_A/%
	GTX 1080Ti	NVIDIA TX2	
Atrous-squeezeseg	45.3	8.3	59.5
Squeezeseg+FCN	39.5	4.2	59.3
VGG16+FCN	29.6	1.9	59.8
SqueezeNet+FCN	46.6	4.5	55.6
Atrous-squeezeseg (without dilated)	45.6	8.4	56.1
Atrous-squeezeseg (without BN)	56.2	9.2	57.3

图 5 为模型训练过程中的损失值曲线图。由 6 种模型的训练实验结果可知，本文的 Atrous-squeezeseg 模型与 SqueezeNet+FCN 模型经过 100 万次的迭代步数训练，损失值最终收敛到与 VGG16+FCN 模型同样的水平，但 SqueezeNet+FCN 模型在训练初期的曲线振荡幅度较大。由 Atrous-squeezeseg(without BN)模型曲线可知，对 Atrous-Fire 模块各层进行批量标准化可使得网络训练损失值曲线较为平滑，振荡幅度较小。而未添加批量标准化层的模型 Atrous-squeezeseg(without dilated)的曲线振荡较为严重，曲线收敛速度较慢。综上所述，本文在 Atrous-Fire 模块中使用批量标准化能增强训练数据，优化模型训练过程。

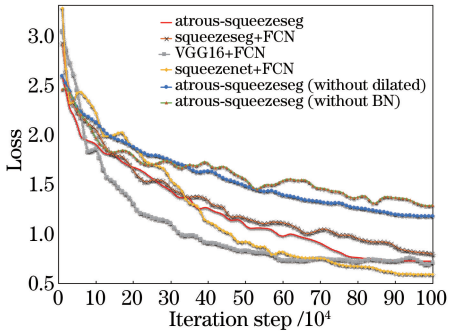

图 5 训练损失值曲线图

Fig. 5 Training loss value curves

图 6 为各种模型在训练过程中对验证数据集进行交叉验证而得到的损失值曲线。由图 6 可知，由 Fire 模块组建的 SqueezeNet+FCN 模型与 Atrous-squeezeseg(without BN)模型损失值曲线均呈先下降再上升的过拟合现象。其他 4 组模型的验证损失值曲线在训练过程中并未发生过拟合现象，且最终都能有效收敛。据此可知，所提 Atrous-Fire 模块中的批量标准化层在优化模型、提高精准度上具有较大贡献。Atrous-squeezeseg(without dilated)未通过多尺度空洞卷积组提取特征，其曲线收敛效果明显较差。

图 7 展示了在 ADE20K 数据集中多种场景下

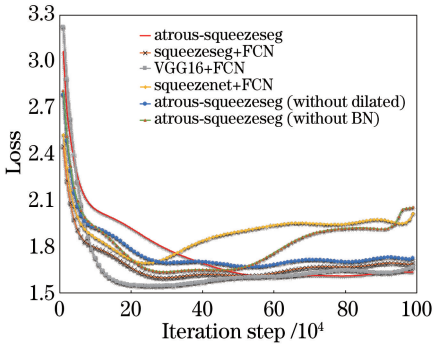

图 6 验证损失值曲线

Fig. 6 Validation loss value curves

所提模型的分割效果，主要为室内、道路的效果对比图。在室内场景中，模型 VGG16＋FCN 将窗户归类为墙，且台灯的分割效果不连续，后 3 组模型的分割效果也都存在着不同程度的缺陷，而所提 Atrous-squeezeseg 模型则表现出了较好的分割效果，能够清晰分割出窗户、台灯和画框，且分割边缘较为平滑，噪声点少。在户外道路场景中，所提模型也能够较为精确地从图像中分割出车辆、路面和行人。而从 Atrous-squeezeseg（without BN）模型的分割效果图来看，其出现了较大的失真，不能准确地对图像像素进行有效归类。

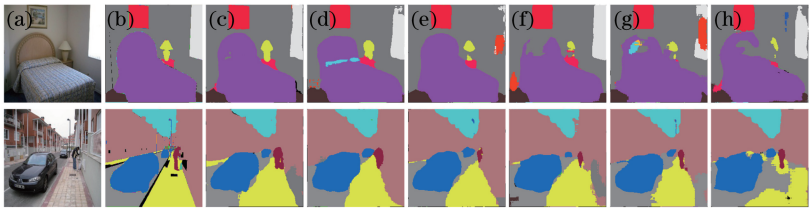

图 7 ADE20K 效果对比图。(a)原图像；(b)分割标注图；(c)所提算法；(d) Squeezeseg＋FCN；(e) VGG16＋FCN；(f) SqueezeNet＋FCN；(g)无空洞；(h)无批量标准化处理

Fig. 7 Effect comparison of ADE20K. (a) Original images; (b) ground truth; (c) proposed algorithm; (d) Squeezeseg＋FCN; (e) VGG16＋FCN; (f) SqueezeNet＋FCN; (g) without dilated; (h) without BN

为了验证本文模型在其他数据集合场景的泛化性，在 PASCAL VOC 数据集中进行了验证，该数据集包含 20 种分类目标和 1 个背景类。在 PASCAL VOC 数据集中，本文模型与文献[8]模型的 P_A 值分别为 60.1％和 62.1％，且分割效果较为平滑，如图 8 所示。其中，模型的参数设置和输入尺度与训练 ADE20K 数据集保持一致。

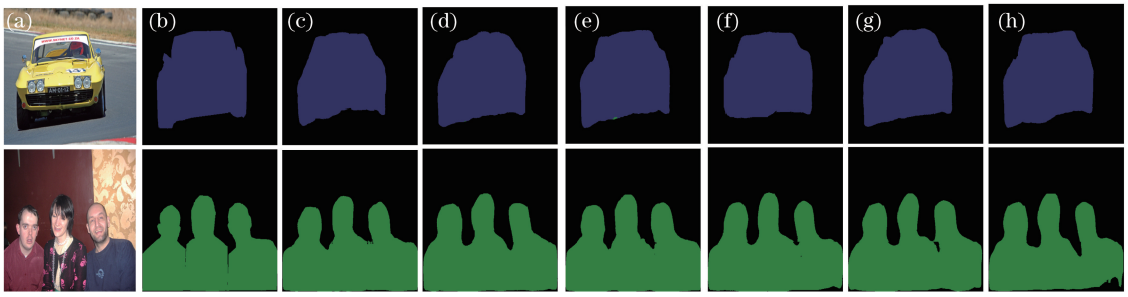

图 8 PASCAL VOC 效果对比图。(a)原图像；(b)分割标注图；(c)所提算法；(d) Squeezeseg＋FCN；(e) VGG16＋FCN；(f) SqueezeNet＋FCN；(g)无空洞；(h)无批量标准化处理

Fig. 8 Effect comparison of PASCAL VOC. (a) Original images; (b) ground truth; (c) proposed algorithm; (d) Squeezeseg＋FCN; (e) VGG16＋FCN; (f) SqueezeNet＋FCN; (g) without dilated; (h) without BN

6 结 论

提出了基于卷积神经网络的低参数实时图像语义分割模型 Atrous-squeezeseg。该模型以 Atrous-Fire 模块为核心，在模块中使用多尺度的空洞卷积方式，并添加了批量标准化层来优化网络。实验结果证明，本文模型不仅能在保障模型精度的前提下，降低模型的参数量，提高模型的输出帧率，而且还能在低参数量时达到较好的分割效果，分割后的图像准确度高，边缘清晰平滑，噪声点少，速度快，易部署在嵌入式设备 NVIDIA TX2 中。但该模型也存在着一些不足，如模型训练时损失曲线收敛速度较慢，在学习率较大时曲线振荡明显等。后续的研究工作将主要针对这些不足，继续优化模型网络，从而进一

步提高模型对语义分割的精准度。

参 考 文 献

[1] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2014.

[2] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]// Proceedings of the 25th International Conference on Neural, 2012, 1: 1097-1105.

[3] He K M, Zhang X Y, Ren S Q, *et al*. Deep residual learning for image recognition[C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[4] Zhang X N, Zhong X, Zhu R F, *et al*. Scene classification of remote sensing images based on integrated convolutional neural networks[J]. Acta Optica Sinica, 2018, 38(11): 1128001.
张晓男, 钟兴, 朱瑞飞, 等. 基于集成卷积神经网络的遥感影像场景分类[J]. 光学学报, 2018, 38(11): 1128001.

[5] Ye G L, Sun S Y, Gao K J, *et al*. Nighttime pedestrian detection based on faster region convolution neural network [J]. Laser & Optoelectronics Progress, 2017, 54(8): 081003.
叶国林, 孙韶媛, 高凯珺, 等. 基于加速区域卷积神经网络的夜间行人检测研究[J]. 激光与光电子学进展, 2017, 54(8): 081003.

[6] Wu C Y, Yi B S, Zhang Y G, *et al*. Retinal vessel image segmentation based on improved convolutional neural network[J]. Acta Optica Sinica, 2018, 38(11): 1111004.
吴晨玥, 易本顺, 章云港, 等. 基于改进卷积神经网络的视网膜血管图像分割[J]. 光学学报, 2018, 38(11): 1111004.

[7] Guo C C, Yu F Q, Chen Y. Image semantic segmentation based on convolutional neural network feature and improved superpixel matching[J]. Laser & Optoelectronics Progress, 2018, 55(8): 081005.
郭呈呈, 于凤芹, 陈莹. 基于卷积神经网络特征和改进超像素匹配的图像语义分割[J]. 激光与光电子学进展, 2018, 55(8): 081005.

[8] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.

[9] 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.

[10] Gamal M, Siam M, Abdel-Razek M. ShuffleSeg: real-time semantic segmentation network[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[11] Song Q S, Zhang C, Chen Y, *et al*. Road segmentation using full convolutional neural networks with conditional random fields [J]. Journal of Tsinghua University (Science and Technology), 2018, 58(8): 725-731.
宋青松, 张超, 陈禹, 等. 组合全卷积神经网络和条件随机场的道路分割[J]. 清华大学学报(自然科学版), 2018, 58(8): 725-731.

[12] Iandola F N, Han S, Moskewicz M W, *et al*. SqueezeNet: Alexnet-level accuracy with 50× fewer parameters and <0.5 MB model size[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[13] Shi J B, Malik J. Normalized cuts and image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 888-905.

[14] Zhu W, Qu J Y, Wu R B. Straight convolutional neural networks algorithm based on batch normalization for image classification[J]. Journal of Computer-Aided Design & Computer Graphics, 2017, 29(9): 1650-1657.
朱威, 屈景怡, 吴仁彪. 结合批归一化的直通卷积神经网络图像分类算法[J]. 计算机辅助设计与图形学学报, 2017, 29(9): 1650-1657.

[15] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]// Proceedings of the 32nd International Conference on Machine Learning, 2015, 37: 448-456.

[16] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2015.

[17] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[C]// IEEE Conference on Medical Image Computing and Computer-Assited Intervention, 2015: 234-241.

[18] Abadi M, Agarwal A, Barham P, *et al*. Tensorflow: large-scale machine learning on

heterogeneous distributed systems[J]. arXiv preprint arXiv:1603.04467,2016.

[19] Chetlur S, Woolley C, Vandermersch P, *et al*. cuDNN: efficient primitives for deep learning[J]. arXiv:1410.0759, 2014.

[20] Zhou B L, Zhao H, Puig X, *et al*. Semantic understanding of scenes through the ADE20K dataset [J]. International Journal of Computer Vision, 2019, 127(3): 302-321.

[21] Garcia-Garcia A, Orts-Escolano S, Oprea S, *et al*. A survey on deep learning techniques for image and video semantic segmentation [J]. Applied Soft Computing, 2018, 70: 41-65.