基于膨胀卷积平滑及轻型上采样的实时语义分割

程晓悦**，赵龙章*，胡穹，史家鹏
南京工业大学电气工程与控制科学学院，江苏 南京 211816

摘要 针对轻量级网络在语义分割速度上较快，但精度较低的问题，在轻量级网络基础上，提出了基于膨胀卷积平滑及轻型上采样的实时语义分割方法。为了提高分割速度，采用结合结构化知识蒸馏的轻量级网络 ResNeXt-18 作为特征提取网络。设计了膨胀卷积平滑模块及轻型上采样模块，用来提高语义分割的精确度。为验证所提方法的有效性，在 Cityscapes 数据集及 CamVid 数据集上进行评估，在 Cityscapes 数据集上得到了速度为 40.2 frame/s，精度为 76.8%，参数量仅为 1.18×10^7 的结果。实验表明，本文提出的实时语义分割方法在保持方法实时性的同时可以得到较好的分割准确度，具有一定的实用价值。

关键词 图像处理；实时语义分割；轻量级网络；知识蒸馏；膨胀卷积；轻型上采样
中图分类号 TP391.4 **文献标志码** A **doi**：10.3788/LOP57.021017

Real-Time Semantic Segmentation Based on Dilated Convolution Smoothing and Lightweight Up-Sampling

Cheng Xiaoyue**, Zhao Longzhang*, Hu Qiong, Shi Jiapeng
College of Electrical Engineering and Control Science, Nanjing Tech University, Nanjing, Jiangsu 211816, *China*

Abstract In lightweight networks, the speed of semantic segmentation is high but the accuracy is low. On the basis of lightweight networks, a real-time semantic segmentation method based on dilated convolution smoothing and lightweight up-sampling is proposed. To improve segmentation speed, a lightweight network, ResNeXt-18, with structured knowledge distillation is used as feature extraction network. To improve the segmentation accuracy, a dilated convolution smoothing module and a lightweight up-sampling module are designed. To verify the effectiveness of the proposed method, the evaluations are carried out using the Cityscapes and CamVid datasets, obtaining the speed of 40.2 frame/s and the segmentation accuracy of 76.8%, with a parameter count of 1.18×10^7. The experimental results demonstrate that the proposed method can obtain high segmentation accuracy while maintaining its high-speed real-time performance; as such, it has certain practical value.

Key words image processing; real-time semantic segmentation; lightweight network; knowledge distillation; dilated convolution; lightweight up-sampling
OCIS codes 100.2960; 150.4065; 150.0155

1 引 言

全卷积神经网络(FCN)[1]的提出解除了网络需要大小固定输出的限制，使得网络可以对图像进行像素级分割，极大地推动了图像语义分割的发展，并在此涌现出一大批思路新颖、性能优异的分割算法。在此基础上，PSPNet[2]利用金字塔场景解析(PSP)模块来捕获图像的全局上下文，增强了网络获取全局信息的能力，提高了网络的分割精度。2016 年，Chen 等[3]的 DeepLab 网络使用空洞空间金字塔池化(ASPP)挖掘多尺度特征，虽然增强了分割精度，但是膨胀卷积[4]粗糙的下采样特征损失了部分细节。利用残差连接将下采样层与网络层相结合构成的 RefineNet[5]，增加了多路径细化结构以改进预测网络，实现了高精度的分割。在 ShuffleNet[6]中，使用一种逐点群卷积有效地降低了计算复杂度，提高

收稿日期：2019-05-31；**修回日期**：2019-07-09；**录用日期**：2019-07-15
基金项目：国家自然科学基金(61403189,61773200)
* **E-mail**：yzzlz22@njtech.edu.cn；** **E-mail**：861210578@qq.com

了分割速度。为了提高网络的计算效率以及减少网络参数，基于 CondenseNet[7] 制定了一项新的训练策略，将重要连接置于分组卷积中，并修剪冗余的连接。深度可分离卷积[8-9]通过分割常规卷积来降低了计算复杂度。基于 ICNet[10] 提出的一种自定义编码器，可处理带有共享参数的图像金字塔，并在图像进入解码器之前对图像进行多尺度融合，使用图像级联加速分割。使用残差连接和分组卷积，ERFNet[11] 将残差块重新定义，使得参数减少了33%。轻量级网络 ENet[12] 提供了极高的分割速度，但是放弃了模型的最后阶段以追求紧凑的框架，由此导致模型的感受野难以覆盖大型物体。加入通道注意块的 DFN[13] 网络，提升了类内一致性，实现了特征选择。文献[14]使用更深的网络结构，结合密集条件随机场的后处理方法，得到了较高的分割精度。利用轻量级网络结构以及跳过连接来实现快速语义分割的方法有 SegNet[15]。为了更加充分获取特征，文献[16]将残差学习与密集网络相结合，增加短连接，取得了较高的分割准确率。由于直接提取图像特征时存有分辨率损失、边缘信息损失的问题，利用基于双通道的卷积神经网络[17]提取特征信息时融合了深、浅两个通道，有效改善了特征损失问题。一种多尺度膨胀卷积结合深层神经网络[18]方法，虽然可以提升影像云的识别精度，但也带来了大的计算量和参数量；文献[19-20]中引入额外的卷积层或堆叠扩张的卷积层，虽然能够实现分割精度的微小改进，但在膨胀卷积中通过增加大量的可训练参数来克服信息损失，这使得模型耗费了大量资源。

为了提高网络分割精度的同时不增加参数量，本文以在 ImageNet 上预训练的 ResNeXt-18 网络为基干网络，结合结构化知识蒸馏[21]方法，引入像素损失函数，进行网络的知识蒸馏。设计了膨胀卷积平滑模块，有效地解决了采样时信息丢失的问题。通过轻型上采样模块恢复预测分辨率，并使用 Cityscapes 数据集对本文提出的实时语义分割方法进行验证。

2 本文工作

2.1 膨胀卷积平滑模块(DCSM)

在语义分割任务中，感受野越大，输出的信息越多[22]。膨胀卷积[23]支持感受器的指数扩展而不损失分辨率和覆盖范围，在分割任务中能有效增大感受野，并且不改变原始内核大小。虽然膨胀卷积可以在不同尺度上捕获特征，但是膨胀卷积采用的是稀疏采样，仅考虑几个点来判断图像的潜在特征，故当出现坏点或者其他干扰时，采样分割效果不佳。为了解决此类问题，对膨胀卷积进行改进：在输入通道上使用插值滤波器捕获更多局部信息，而不是直接进行膨胀卷积；通过组合相邻像素信息对膨胀卷积进行平滑[24]。

2.1.1 插值方法

由于采样时采样点具有不规则、排列偏移的性质，常见的下采样需要计算图像像素的双线性插值[25]，计算方法为

$$x(p)=\sum_{q}G(q,p)\cdot x(q)\ , \tag{1}$$

式中：q 为输入特征图中，在采样点周围相邻像素点的集合；p 为采样点；$G(\cdot)$ 为双线性插值核，$G(q,p)=g(q_x,p_x)\cdot g(q_y,p_y)$，$g(a,b)=\max(0,1-|a-b|)$，$a$、$b$ 为实数。p_x 和 p_y 为采样点 p 的坐标，q_x 和 q_y 为采样点周围相邻像素点坐标。

对于输入滤波器，一般方法是计算周围相邻像素点的平均值，设较远处的像素权重为零，计算公式为

$$v(x,y)=\begin{cases}\dfrac{1}{r^2}, & \text{if } |x|,|y|<r/2\\ 0, & \text{otherwise}\end{cases}, \tag{2}$$

式中：x，y，r 分别表示像素点横坐标、像素点纵坐标、像素点与相邻像素点间的距离。

由于像素的直接相邻点应该比远处的像素点更强烈地影响扩张卷积输出，因此使用高斯滤波器，将更多的权重放在靠近滤波器中心的像素上，较少的权重放在滤波器边缘的像素点上。满足以上要求的函数是方差为 σ^2 的高斯函数，具体表达式为

$$v(x,y)=\begin{cases}\dfrac{1}{2\pi\sigma^2}\exp\leqslant\left[-\dfrac{1}{2\sigma^2}(x^2+y^2)\right], & \\ \qquad \text{if } |x|,|y|<r/2 & \\ 0, \text{ otherwise} & \end{cases}, \tag{3}$$

式中：σ 为一个可以根据经验选择的固定参数。经过多次实验比较，本研究选择 σ 为 4。为了可以使用梯度下降进行参数优化，在网络中训练时，采用的文献[26]中的插值方法，并学习输入滤波器的参数，来实现深度可分离卷积[27]。

2.1.2 平滑方法

为了捕获更多的信息，先在扩张卷积输入中组合了相邻像素信息，再使用大小为 r_f 的插值滤波器 f_v，对每个像素周围的像素信息进行组合，可表示为

$$y(i)=\sum_{k=1}^{K}\left[\sum_{n=-r_f/2}^{r_f/2}x(i-r_f\cdot k-n)v(n)\right]w(k), \tag{4}$$

式中：$\left(\frac{r_f}{2}\right)$表示插值滤波器大小的一半；$K$ 为扩张率；v 为差值滤波器；w 为权重；n 为卷积核大小；i 为像素点。

经过平滑的卷积如图 1 所示。对于一个核为 3×3、扩张率为 5 的卷积，利用大小为 r_f 的恒定滤波器 f_v 对其进行平滑处理，得到了可以捕获更多信息的膨胀卷积。更形象的平滑效果如图 2 所示。

本文插值方法及平滑方法在只增加较少的参数量情况下，有效地克服了信息丢失，扩大了感受野，以较低的计算成本提高了扩张卷积的性能，能够抵抗局部噪声并编码更多的局部空间信息。

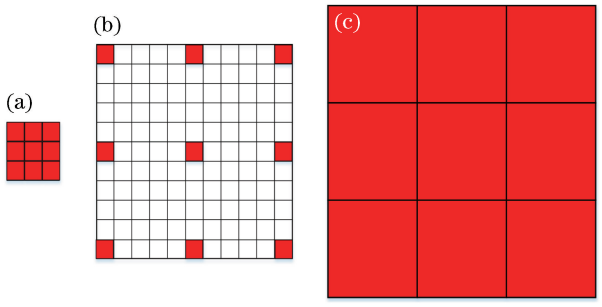

图 1 不同的卷积类型。(a)普通卷积；(b)膨胀卷积；(c)平滑后的膨胀卷积

Fig. 1 Different convolution types. (a) Ordinary convolution; (b) dilated convolution; (c) dilated convolution after smoothing

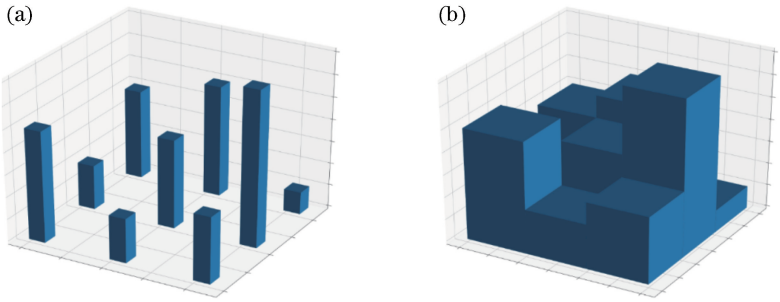

图 2 膨胀卷积与平滑效果[24]。(a)膨胀卷积；(b)平滑后的膨胀卷积

Fig. 2 Dilated convolution and smoothing effect[24]. (a) Dilated convolution; (b) dilated convolution after smoothing

2.2 轻型上采样模块(HC-LUM)

图像可以在具有简单编码器的条件下进行语义分割，但是编码器会逐渐减小分辨率并增加特征映射的数量。一般为了不减少分辨率，通过添加像素方式的损失函数，来获得预测[28]。像素预测的分辨率低易导致不能分割出小物体、模型感知域不足，从而导致难以对大物体进行像素分割等问题。膨胀卷积[4]、学习上采样[1]、横向连接[29]、分辨率金字塔[28]等方法可以改善此类问题，但由于计算量较大难以达到实时性。因此，为了改善分辨率减小的问题，提出了一种具有横向连接的轻型上采样模块，该模块在本文模型下可以进行有效准确的上采样。

编码器将输入图像转换为语义多样的视觉特征，解码器将这些特征上采样到输入分辨率。具有横向连接的上采样模块[29-31]如图 3 所示，有两个特征输入：低分辨率特征及横向特征。低分辨率特征通过双线性插值的方法，使其分辨率与横向特征的分辨率相同，然后在对应子采样的残差块处进行混合[32]。不同于文献[31]，本文上采样模块使用的卷积为经过平滑的膨胀卷积，图 3 中用 DCS(dilated convolution smoothing)表示，图中将第二个 ReLU

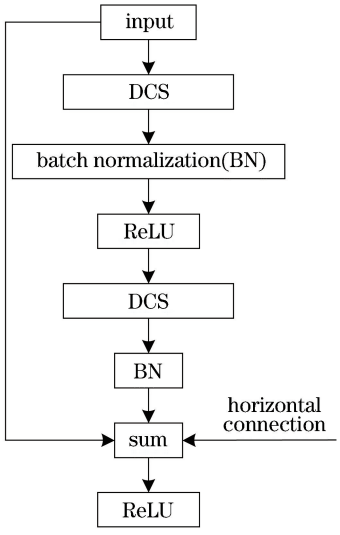

图 3 卷积块内残差单元结构图

Fig. 3 Structure of residual unit in convolution block

输出转到下一个残差块。

图 4 为加入 HC-LUM 的网络结构图，其中 SPP 表示空间金字塔池化层，1×1 瓶颈块表示瓶颈层，在网络中训练卷积组中的参数，并通过 SPP 及若干 1×1 瓶颈块将其传递到 HC-LUM。

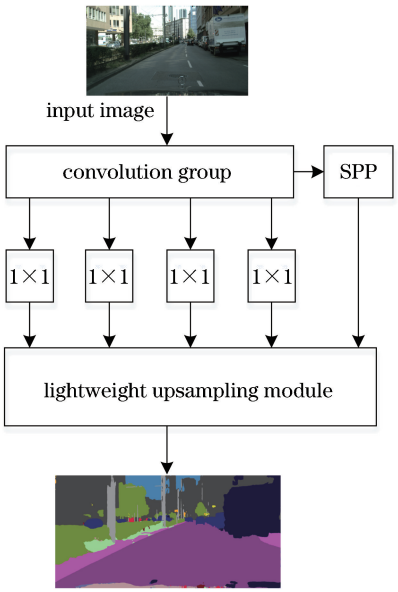

图 4 加入 HC-LUM 的网络结构图

Fig. 4 Structural diagram of network with HC-LUM

2.3 特征提取网络

采用 ImageNet 预训练的 ResNeXt-18 作为特征提取网络，与文献[31]不同的是，本文在 ResNeXt-18 中加入了适用于图像语义分割的结构化知识蒸馏[21,33]方法，在网络计算量不变的前提下可以提高网络的分割精度。知识蒸馏即在高精度、模型复杂度高的网络下训练出子网络，子网络的特点是参数量少、计算量小。保证子网络与原网络输出结果的一致性的关键在于：设计训练中的损失函数。

首先是单个像素损失，一般将图像分割问题看作为单个像素标记问题的集合；其次，为了保证图像中的局部一致性，设计了组合像素损失；最后，为了对图像进行整体结构相似性度量，设计了整体损失函数。

引入的单个像素损失函数为

$$l_{\mathrm{pi\_S}}=\frac{1}{W'\times H'}\sum_{i\in R}\xi_{\mathrm{KL}}(q_i^{\mathrm{S}}\parallel q_i^{\mathrm{T}}),\tag{5}$$

式中：q_i^{S} 表示从子网络 S 产生的第 i 个像素的类概率；q_i^{T} 表示从原网络 T 产生的第 i 个像素的类概率；$\xi_{\mathrm{KL}}(\cdot)$ 为两个概率之间的 Kullback-Leibler 发散度[34]，R 表示所有像素；W' 和 H' 分别表示图像的宽度与高度。

引入的组合像素损失函数为

$$l_{\mathrm{pa\_S}}=\frac{1}{(W'\times H')^2}\sum_{i\in R}\sum_{j\in R}(a_{ij}^{\mathrm{S}}-a_{ij}^{\mathrm{T}})^2,\tag{6}$$

式中：a_{ij}^{T} 表示从原网络 T 中产生的第 i 个像素与第 j 个像素间的相似性；a_{ij}^{S} 表示从子网络 S 中产生的第 i 个像素与第 j 个像素间的相似性。根据特征 $\boldsymbol{f}_i$ 和 $\boldsymbol{f}_j$，组内两像素间的相似性可表示为

$$a_{ij}=\boldsymbol{f}_i^{\mathrm{T}}\boldsymbol{f}_j/(\parallel\boldsymbol{f}_i\parallel_2\parallel\boldsymbol{f}_j\parallel_2)。\tag{7}$$

引入的整体损失函数为

$$l_{\mathrm{ho\_S}}(D)=E_{Q^{\mathrm{S}}\sim p_{\mathrm{S}}(Q^{\mathrm{S}})}[D(Q^{\mathrm{s}}\mid I)]-E_{Q^{\mathrm{T}}\sim p_{\mathrm{T}}(Q^{\mathrm{T}})}[D(Q^{\mathrm{T}}\mid I)],\tag{8}$$

式中：I 为输入的 RGB 图像；Q^{S} 为子网络预测分割图，看作伪样本；Q^{T} 为原网络预测分割图，看作真实样本；$E[\cdot]$为期望算子；$D(\cdot)$为一个嵌入网络，具有 5 个卷积的全卷积神经网络。采用条件对抗性学习，希望 Q^{S} 与 Q^{T} 有相似处，并用 Wasserstein 距离[35] $[Q^{\mathrm{S}}\sim p_{\mathrm{S}}(Q^{\mathrm{S}})$ 和 $Q^{\mathrm{T}}\sim p_{\mathrm{T}}(Q^{\mathrm{T}})]$评估实际分布与假分布之间的差异。在最后三层间插入两个自注意模块以捕获结构信息[36-37]，用于产生表示输入图像和分割图匹配程度的整体嵌入。

3 实验分析

3.1 数据集及评价标准

Cityscapes 数据集[38]是在白天、晴朗天气下从驾驶员的角度拍摄的高分辨率图像的集合。提供了 5000 张高质量像素级精细注释图像和 20000 张粗略注释图像，包括 2975 个训练数据，500 个验证数据和 1525 个带有 19 类用于评估的类别标签的测试数据。Cityscapes 数据集是图像语义分割领域中使用广泛的城市路面场景数据集，实验中使用精细注释的数据，未使用粗略注释图像。CamVid 数据集[39]由英国剑桥大学提供，是第一个具有对象类语义标签的数据集。在评估标准方面，采用平均交并比(MIOU)评估分割精度，使用帧率 f_{ps} 评估分割速度。

3.2 实现过程及结果

本文的实验环境是 Tensorflow1.9，CUDA 9.0，cuDNN7.5，工作站配置 Inter © Core TM i7-6800K CPU@3.4GHz，GTX 1080Ti 显卡，内存 128G，8 张显卡运行。训练时将图像转换为 1024 pixel × 2048 pixel作为输入，并应用常规数据增强方法，包括图像裁剪、随机翻转、高斯噪声添加图像抖动。分割网络采用小批量随机梯度下降(SGD)训练，迭代次数为 4000 次，动量为 0.9，权值衰减为 0.0005。基于本文语义分割网络：在 Cityscapes 数据集上得到了 MIOU 为 76.8%，帧频为 40.2 frame/s，参数量为 1.18×10^7 的结果；在 CamVid 数据集上得到了

MIOU 为 65.3%，帧频为 34.2 frame/s 的结果，满足实时性要求。

在得到最终结果之前，实验还将本文方法与传统 ResNeXt-18 网络＋普通膨胀卷积(D-Cov)的分割方法、传统 ResNeXt-18 网络＋DCSM 的方法、传统 ResNeXt-18 网络＋DCSM＋HC-LUM 方法进行了对比，结果显示，本文方法在精度及速度上均优于以上其他方法，具体结果如表 1、表 2 所示。

表 1　Cityscapes 数据集上 4 种分割方法对应的结果

Table 1　Corresponding results of four segmentation methods on Cityscapes dataset

Method	MIOU /%	Frame rate /(frame·s^{-1})	Parameter /10^7
ResNeXt-18＋D-Cov	72.3	36.3	1.25
ResNeXt-18＋DCSM	73.7	36.1	1.39
ResNeXt-18＋DCSM＋HC-LUM	73.9	35.9	1.18
Proposed	**76.8**	**40.2**	**1.18**

表 2　CamVid 数据集上 4 种分割方法对应的结果

Table 2　Corresponding results of four segmentation methods on CamVid dataset

Method	MIOU / %	Frame rate / (frame·s^{-1})
ResNeXt-18＋D-Cov	64.2	33.9
ResNeXt-18＋DCSM	64.7	33.7
ResNeXt-18＋DCSM＋HC-LUM	65.1	32.5
Proposed	**65.3**	**34.2**

表 3、表 4 是本文方法与其他语义分割方法在相同的硬件配置下的分割效果对比，可以看到，在 Cityscapes 数据集上，本文方法在分割精度上高于其他网络，在速度上 ESPNet 为几个网络中最快，但分割精度远低于本文方法。在 CamVid 数据集上，虽然在分割精度上 PSPNet 较高，但是 PSPNet 没有达到实时性分割的要求。

表 3　本文方法与其他分割网络的对比(Cityscapes 数据集)

Table 3　Comparison of proposed method with other segmentation networks (Cityscapes dataset)

Method	MIOU / %	Frame rate / (frame·s^{-1})
ICNet	69.5	30.3
Two-column Net	72.9	14.7
LadderDenseNet	72.82	31.0
ESPNet	60.3	**112**
ERFNet	68.0	11.2
GUNet	70.4	37.3
Proposed	**76.8**	40.2

除了表 3、表 4 的对比，还给出了在 Cityscapes 数据集上，本文分割网络与其他分割网络在具体 19 个类别中的分割结果对比图，如图 5 所示。从图中可以看到，除了 sidewalk、train 两类物体的分割精度略低于其他网络外，其他各类物体的分割精度都不低于其他网络。

为了表明知识蒸馏方法在保持计算复杂度不变的同时，还可以提高分割精度的效果，以及知识蒸馏方法在其他网络中的普遍适用性，在 Cityscapes 数据集上进行了如图 6 所示的实验对比，可以看出，各网络的参数量保持不变，但是分割精度有了不同程度的提升，证明了知识蒸馏方法的有效性。

表 4　本文方法与其他分割网络的对比(CamVid 数据集)

Table 4　Comparison of proposed method with other segmentation networks (CamVid dataset)

Method	MIOU / %	Frame rate / (frame·s^{-1})
ICNet	67.1	27.8
PSPNet	**69.1**	5.4
Dilation10	65.3	4.4
SegNet	46.4	4.6
ERFNet	59.4	10.1
GUNet	61.8	31.3
Proposed	65.3	**34.2**

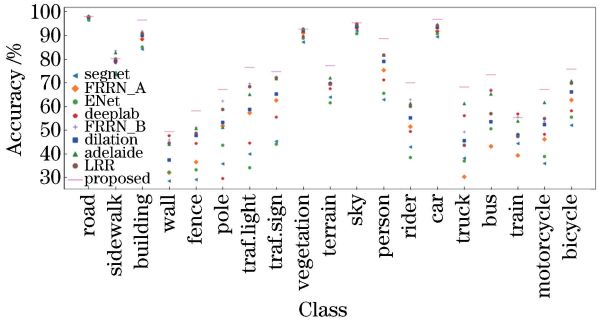

图 5　本文分割方法与其他方法在 19 类物体分割精度的对比

Fig. 5　Comparison of segmentation accuracy between proposed segmentation method and other methods for 19 types of objects

轻量级网络在速度上有很大的优势，层数越少的网络，分割速度越快。但是在分割精度上，往往是深度网络获得的精度较高。为了比较轻量级网络与深度网络的分割精度，做了图 7 所示的对比实验。

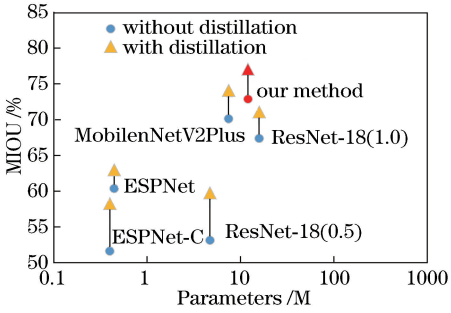

图 6 知识蒸馏方法对各分割网络精度的影响示意图

Fig. 6 Influence of knowledge distillation method on accuracy of each segmentation network

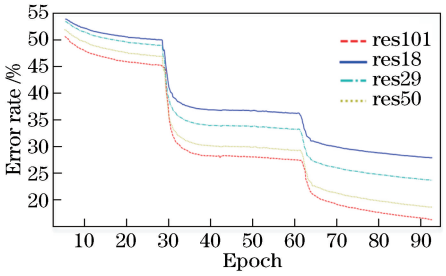

图 7 不同网络层数的分割精度对比

Fig. 7 Comparison of segmentation accuracy for different network layers

在对比实验中，除了网络的层数不同，其他的参数设置、硬件平台保持不变。可以看到，101层的网络错误率最小，层数最少的 18 层网络，错误率最高。

图 8 为使用本文分割方法得出的分割图，图 8(b)中对部分细节进行放大，以方便观察分割结果。

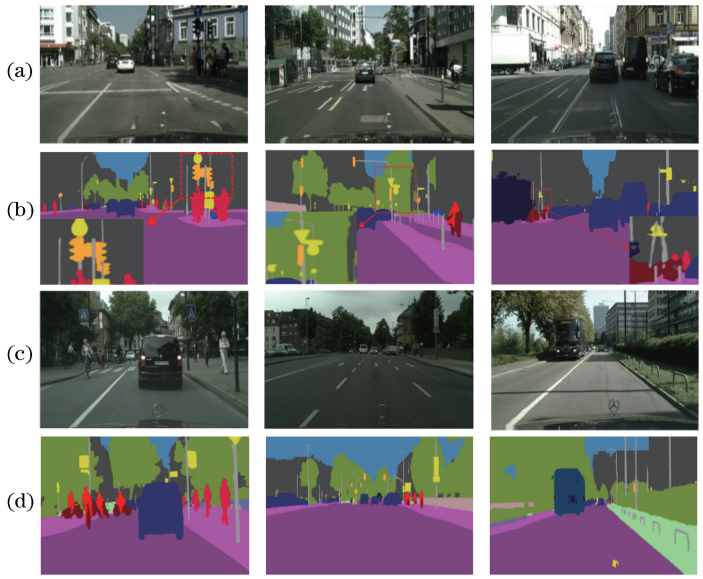

图 8 本文方法分割结果。(a)组图 1；(b)组图 1 的分割结果图(包含部分细节放大图)；(c)组图 2；(d)组图 2 的分割结果图

Fig. 8 Segmentation results of proposed method. (a) Original images 1; (b) segmentation results of original images 1 (including enlarged images of partial detail); (c) original images 2; (d) segmentation results of original images 2

通过上述实验可知，所提网络可以有效地完成实时语义分割的任务。在 ImageNet 上预训练的 ResNeXt 网络的基础上，增加经过平滑的膨胀卷积及轻型上采样模块来进行优化，尽可能地减少网络参数量，通过知识蒸馏方法进一步提高了网络的分割精度，有效地平衡了分割网络的速度和精度。通过与现有分割网络的对比实验，可以看到所提网络在速度和精度上的优越性。

4 结 论

本文分割网络主要有三个创新点，分别是扩张卷积的平滑、轻型上采样模块及特征提取网络的知识蒸馏。在 Cityscapes 数据集上得到了 76.8% 的 MIOU，有效地提高了分割网络的精确度，适用于实时语义分割。膨胀卷积在保持特征图的空间分辨率方面起着重要作用，但是引入的扩张带来了较高的计算复杂度，虽然对膨胀卷积进行了平滑改进，但是其计算复杂度并没有降低。因此，在下一步的研究中，将试图找到可以替代膨胀卷积的方法，在保持分割精确度和速度的情况下，降低计算复杂度。

参 考 文 献

[1] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 3431-3440.

[2] 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, 2017: 6230-6239.

[3] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[J/OL]. (2016-06-07)[2019-05-30]. https://arxiv.org/abs/1412.7062.

[4] 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.

[5] 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, 2017: 5168-5177.

[6] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 6848-6856.

[7] Huang G, Liu S C, van der Maaten L, et al. CondenseNet: an efficient DenseNet using learned group convolutions[J/OL]. (2018-06-07)[2019-05-30]. https://arxiv.org/abs/1711.09224.

[8] Sifre L, Mallat S. Rigid-motion scattering for texture classification[J/OL]. (2014-03-07)[2019-05-30]. https://arxiv.org/abs/1403.1687.

[9] Wang M, Liu B, Foroosh H. Factorized convolutional neural networks[C]// Proceedings of the IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 545-553.

[10] Zhao H S, Qi X J, Shen X Y, et al. ICNet for real-time semantic segmentation on high-resolution images[J/OL]. (2018-08-20)[2019-05-30]. https://arxiv.org/abs/1704.08545.

[11] 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.

[12] Paszke A, Chaurasia A, Kim S, et al. Enet: a deep neural network architecture for real-time semantic segmentation[J/OL]. (2016-06-07)[2019-05-30]. https://arxiv.org/abs/1606.02147.

[13] 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, 2018: 1857-1866.

[14] Hu M Y, Zhang Y Z, Qin C, et al. Semantic map construction based on deep convolutional neural network[J]. ROBOT, 2019, 41(4): 452-463.
胡美玉, 张云洲, 秦操, 等. 基于深度卷积神经网络的语义地图构建[J]. 机器人, 2019, 41(4): 452-463.

[15] 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.

[16] 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.

[17] Li S M, Lei G Q, Fan R. Depth map super-resolution based on two-channel convolutional neural network[J]. Acta Optica Sinica, 2018, 38(10): 1010002.
李素梅, 雷国庆, 范如. 基于双通道卷积神经网络的深度图超分辨研究[J]. 光学学报, 2018, 38(10): 1010002.

[18] Gao L, Song W D, Tan H, et al. Cloud detection based on multi-scale dilation convolutional neural network for ZY-3 satellite remote sensing imagery[J]. Acta Optica Sinica, 2019, 39(1): 0104002.
高琳, 宋伟东, 谭海, 等. 多尺度膨胀卷积神经网络资源三号卫星影像云识别[J]. 光学学报, 2019, 39(1): 0104002.

[19] Wang P Q, Chen P F, Yuan Y, et al. Understanding convolution for semantic segmentation[C]// 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), March 12-15, 2018, Lake Tahoe, NV, USA. New York: IEEE, 2018: 1451-1460.

[20] Wang Z Y, Ji S W. Smoothed dilated convolutions for improved dense prediction[C]// Proceedings of

the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '18, August 19-23, 2018, London, United Kingdom. New York: ACM, 2018: 2486-2495.

[21] Liu Y F, Chen K, Liu C, et al. Structured knowledge distillation for semantic segmentation[J/OL]. (2019-03-12)[2019-05-30]. https://arxiv.org/abs/1903.04197.

[22] Huang T H, Nie Z Y, Wang Q G, Li S, Yan L C, Guo D S. Image real-time semantic segmentation based on block adaptive feature fusion [J/OL]. Acta Automatica Sinica:1-12[2019-04-15].
黄庭鸿, 聂卓赟, 王庆国, 李帅, 晏来成, 郭东生. 基于区块自适应特征融合的图像实时语义分割[J/OL]. 自动化学报: 1-12[2019-04-15].

[23] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[J/OL]. (2016-04-30)[2019-05-30]. https://arxiv.org/abs/1511.07122.

[24] Ziegler T, Fritsche M, Kuhn L, et al. Efficient smoothing of dilated convolutions for image segmentation[J/OL]. (2019-03-19)[2019-05-30]. https://arxiv.org/abs/1903.07992.

[25] Bai J, Hao P H, Chen S H. Traffic scene understanding using image semantic segmentation with an improved lightweight convolutional-neural-network [J]. Journal of Automotive Safety and Energy, 2018, 9(4): 433-440.
白傑, 郝培涵, 陈思汉. 用轻量化卷积神经网络图像语义分割的交通场景理解[J]. 汽车安全与节能学报, 2018, 9(4): 433-440.

[26] Jaderberg M, Simonyan K, Zisserman A, et al. Spatial transformer networks[J/OL]. (2016-02-04)[2019-05-30]. https://arxiv.org/abs/1506.02025.

[27] Chollet F. Xception: deep learning with depthwise separable convolutions[J/OL]. (2017-04-04)[2019-05-30]. https://arxiv.org/abs/1610.02357.

[28] Farabet C, Couprie C, Najman L, et al. Learning hierarchical features for scene labeling [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8): 1915-1929.

[29] Rasmus A, Berglund M, Honkala M, et al. Semi-supervised learning with ladder networks [C] // Advances in Neural Information Processing Systems, December 7-12, 2015, Montreal, Quebec, Canada. Canada: NIPS, 2015: 3546-3554.

[30] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[M] // Navab N, Hornegger J, Wells W, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.

[31] Or ić M, Krešo I, Bevandić P, et al. In defense of pre-trained ImageNet architectures for real-time semantic segmentation of road-driving images [J/OL]. (2019-04-12)[2019-05-30]. https://arxiv.org/abs/1903.08469?context=cs.CV.

[32] Lu W C, Pang Y W, He Y Q, et al. Real-time and accurate semantic segmentation based on separable residual modules [J]. Laser & Optoelectronics Progress, 2019, 56(5): 051005.
路文超, 庞彦伟, 何宇清, 等. 基于可分离残差模块的精确实时语义分割[J]. 激光与光电子学进展, 2019, 56(5): 051005.

[33] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network[J/OL]. (2015-03-09)[2019-05-30]. https://arxiv.org/abs/1503.02531.

[34] Kullback S, Leibler R A. On information and sufficiency [J]. The Annals of Mathematical Statistics, 1951, 22(1): 79-86.

[35] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of Wasserstein GANs [C] // Advances in Neural Information Processing Systems, December 4-9, 2017, Long Beach, CA, USA. Canada: NIPS, 2017: 5767-5777.

[36] Zhang H, Goodfellow I, Metaxas D, et al. Self-attention generative adversarial networks [J/OL]. (2019-06-14)[2019-06-30]. https://arxiv.org/abs/1805.08318.

[37] Zhao H, Zhang Y, Liu S, et al. Psanet: point-wise spatial attention network for scene parsing [C] // Proceedings of the European Conference on Computer Vision (ECCV), September 8-14, 2018, Munich, Germany. New York: IEEE, 2018: 267-283.

[38] 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, June 26-July 1, 2016, Las Vegas, Nevada. New York: IEEE, 2016: 3213-3223.

[39] Brostow G J, Shotton J, Fauqueur J, et al. Segmentation and recognition using structure from motion point clouds [M] // Forsyth D, Torr P, Zisserman A. Computer vision-ECCV 2008. Lecture notes in computer science. Berlin, Heidelberg: Springer, 2008, 5302: 44-57.