计算机科学
COMPUTER SCIENCE

基于多路径特征提取的实时语义分割方法

程成，降爱莲

引用本文

程成，降爱莲. 基于多路径特征提取的实时语义分割方法[J]. 计算机科学, 2022, 49(7): 120-126.

CHENG Cheng, JIANG Ai-lian. Real-time Semantic Segmentation Method Based on Multi-path Feature Extraction [J]. Computer Science, 2022, 49(7): 120-126.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210500157

基于多路径特征提取的实时语义分割方法

程 成 降爱莲
太原理工大学信息与计算机学院 山西 晋中 030600
(2462074653@qq.com)

摘 要 深度学习在图像语义分割领域的应用极大地提升了分割精确度,但由于深度学习网络在速度、内存等方面的限制,其并不能直接应用于嵌入式设备进行实时分割。针对语义分割模型存在的的网络结构复杂和计算开销巨大的问题,提出了结合边缘检测算法的多路径特征提取的实时语义分割算法。模型通过 Sobel 算子、Scharr 算子和 Laplacian 算子对图像的轮廓信息进行提取。算法设计了空间路径提取图像的空间位置信息、语义路径提取图像高级语义信息,以及通过边缘检测路径提取图像中具有代表性的纹理特征,并采用 Ghost 轻量化模块来减少模型参数量,提高算法的分割速度。在 480 像素×360 像素的 CamVid 数据集上的实验结果表明,在 3 种边缘检测算子上,模型的分割准确率均能得到有效提升,尤其是在加入 3×3 大小的 Sobel 算子下算法的性能提升最为明显,在 CamVid 测试集图像处理速度为 349 frames/s 的基础上,分割精度达到了 42.9%。所提算法在分割精度和分割速度上均取得了较好的效果,在实时性和准确性之间达到了很好的平衡。

关键词:深度学习;边缘检测;实时语义分割;特征融合;多特征提取
中图法分类号 TP391

Real-time Semantic Segmentation Method Based on Multi-path Feature Extraction

CHENG Cheng and JIANG Ai-lian
College of Information and Computer, Taiyuan University of Technology, Jinzhong, Shanxi 030600, China

Abstract The application of deep learning in the field of image semantic segmentation has greatly improved the accuracy of segmentation, but due to the limitations of speed and memory, these models can not be directly applied to embedded devices for real-time segmentation. Aiming at the problems of complex network structure and huge computation cost of semantic segmentation model, a real-time semantic segmentation algorithm based on multi-path feature extraction combined with edge detection algorithm is proposed. The model uses Sobel operator, Scharr operator and Laplacian operator to extract the contour information of the image. The algorithm designs the spatial path to extract the spatial position information of the image, designs the semantic path to extract the advanced semantic information of the image, and uses the edge detection path to extract the representative texture features of the image. The ghost lightweight module is used to reduce the amount of model parameters and improve the segmentation speed of the algorithm. Experimental results on 480 pixel and 360 pixel CamVid dataset show that the segmentation accuracy of the model can be improved on the three edge detection operators, especially when the Sobel operator with the size of 3×3 is added, the performance of the algorithm is improved most obviously, and the segmentation accuracy reaches 42.9% on the basis of the image processing speed of 349 frames/s on CamVid test set. Both the segmentation accuracy and segmentation speed achieve good results, and achieve a good balance between real-time and accuracy.

Keywords Deep learning, Edge detection, Semantic segmentation in real-time, Feature fusion, Multi-feature extraction

1 引言

图像语义分割是对图像的像素级分类任务,在无人驾驶、机器人视觉等领域[1-2]有着重要的作用,这些任务对道路界限和障碍物检测有着高精度的需求。在语义分割过程中,由于相邻的像素的图像信息过于相似,造成分割边缘不准确、同一幅图像中不同类别或实例的像素不均衡的问题。现在基于深度学习的分割模型虽然在分割结果上有着高准确度,但是往往需要大量的计算资源,无法达到实时分割的效果。

近年来,深度神经网络[3]逐渐成为图像语义分割任务的首选方法。Long 等[4]于 2015 年提出了全卷积网络(Fully Convolutional Network, FCN),FCN 将分类网络改为全卷积神经网络,采用跳跃连接和反卷积操作进行端到端的像素预测。Ronneberger 等[5]提出了 U-Net 模型结构,U-Net 采用

到稿日期:2021-05-24 返修日期:2021-09-07
基金项目:山西省回国留学人员科研资助项目(2017-051)
This work was suppoted by the Scientific Research Funding Project for Returned Overseas Scholars in Shanxi Province(2017-051).
通信作者:降爱莲(ailianjiang@126.com)

了完全对称的编码器与解码器，通过编码器实现对图像特征的提取，利用解码器对图像像素进行分类。Zhao 等[6]于 2017 年提出的 PSPNet 采用空间金字塔池化模块（Spatial Pyramid Pooling，SPP）聚合不同区域的上下文信息得到全局上下文信息。Chen 等[7]提出了 DeepLab 算法，将空洞卷积和空间金字塔池化方法相结合，提出了空洞空间金字塔池化（Atrous Spatial Pyramid Pooling，ASPP）模块，该模块通过整合多尺度特征，扩大感受野，对空间变换具有较强的不变性，以此提高分割精度。

随着分割任务存在的问题被逐个击破，研究人员开始关注分割图像的速度，力求将其应用于嵌入式设备或实时的分割任务中，这就要求实时分割模型减少大量计算资源的消耗。Paszke 等[8]提出了 ENet 实时分割算法，ENet 由一个较大的编码器和一个较小的解码器组成，在图像进行解码输出的过程中大幅减少参数量，并通过减少特征提取过程中的通道数，提升了算法前向推理的速度。Yu 等[9]提出的 BiseNet 网络采用了双边的实时语义分割模型，通过结合双路径信息，在分割精度和速度上都有着很好的效果。2019 年，Li 等提出了 DFANet[10]网络，该网络采用深度特征聚合的方法，解决了高分辨率图像上的实时语义分割问题。2020 年，Yu 等在 BiseNet 的基础上提出了 BiseNet V2 网络，该网络设计了新的融合层来增强双边路径信息的融合，并通过增强训练策略提升了分割性能。

本文提出了一种既能保证分割实时性又能保证分割准确率的网络结构，采用边缘检测算法对图像信息进行提取，并将所提取的信息和原图下采样所提取的特征图进行融合。该网络通过 3 条特征路径提取了图像的空间位置信息、语义信息和图像边缘信息，并在语义信息的提取过程中采用了 Ghost 轻量化模块[11]，提升了网络分割图像的速度。本文算法有以下 3 个特点：

（1）结合边缘检测算子提取图像中的边缘信息，为算法预测输出提供了边缘特征信息；

（2）采用 3 条特征提取路径，结合图像的空间位置信息、语义信息和边缘信息推动深度学习模型对图像的理解；

（3）引入 Ghost 轻量化模块，在不减少特征数量的情况下极大地降低了计算复杂度。

2　相关工作

2.1　语义分割方法

语义分割模型的高精确度取决于它们的主干网络，这些模型的主干网络有两个主要的结构：1）以空洞卷积为主要特征提取手段的网络主干，其减少了下采样操作次数，在不增加参数的情况下通过不同的空洞率来扩大感受野；2）编码器-解码器结构[12]，编码器逐渐降低特征图的空间维度，解码器逐步修复物体的细节和空间细节。编码器和解码器之间通常存在跳跃连接，因此能帮助解码器更好地修复目标的细节。U-Net 是这种方法中最常用的结构。

图像分割准确度的提高，既需要丰富的空间信息，又需要较大的感受野来感应不同尺寸的目标，目前常用的方法有：1）将普通卷积操作替换为空洞卷积；2）在下采样过程中将特征图密集连接；3）采用编-解码结构结合上下文信息进行推理等。上述方法的基本思想都是尽可能多地融合图像特征信息来对图片目标进行辨别；然而模型为达到高精准度，需要花费高计算成本，且分割图像速度过慢，阻碍了图像语义分割在现实落地。

2.2　模型轻量化方法

为了满足嵌入式设备、自动驾驶等技术对语义分割的实时性需求，研究人员提出了分解卷积操作、模型剪枝和权重量化的模型轻量化技术。1）分解卷积操作：普通的卷积操作同时对特征图的所有通道的信息进行计算。Inception[13]，Xception[14]和 MobileNet[15]等通过深度可分离卷积降低了计算复杂度。深度可分离卷积将标准的卷积操作分解成对特征图的分组卷积和 1×1 的点卷积，极大地减少了参数量并降低了计算复杂度。2）模型剪枝：挑选出模型中不重要的参数并将其剔除，但不对模型的效果造成太大的影响。在剔除不重要的参数之后，通过重新训练恢复模型性能。因此，找到一种有效的评价参数重要性的手段，在这个方法中尤为重要。目前，基于模型裁剪的方法[16]是最为简单有效的模型压缩方法。3）权重量化：通过减少表示各个权重所需的比特数来压缩原始网络。例如，8 bit 近似算法将 32 bit 的梯度和激活值压缩到 8 bit 大小，通过 GPU 集群测试模型和数据的并行化性能，可以在保证模型预测精度的条件下取得两倍的数据传输速度。

3　多路径特征提取方法

为了使分割模型兼顾分割速度和分割精度，本文将从这两个方面来分别阐述。为了提升模型分割的准确度，语义分割要尽可能地提取图像各个层次的信息。为了更好地对图像信息进行补充，本文使用边缘检测算法提取更为细致的边缘信息，我们将不同算子生成的图像分别与原图结合后输入模型。此外，为了更好地弥补在上采样过程中空间信息的缺失，将具有空间位置信息的高分辨率特征图融入网络中进行预测。在模型处理图片速度方面，在一个训练好的神经网络中，一般都是通过一系列的卷积操作来提取图像特征，这会产生大量的特征图，其中通常包含冗余的特征图，本文采用了 Ghost 模块，可以使用更少的参数来生成更多的特征图。基于上述方法，本文构建了一种可以充分提取特征以及加快分割速度的网络，如图 1 所示。图 1 上半部分通过语义路径和空间路径分别提取原图的位置信息和语义信息，下半部分为对边缘检测算法生成的图像提取边缘信息，最后通过特征融合模块（Feature Fusion Module，FFM）将 3 条路径所提取的特征进行融合预测。

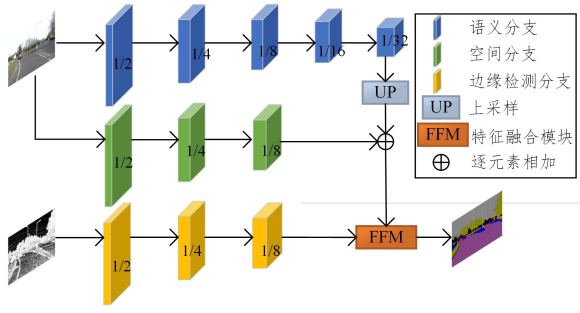


图 1　模型结构图

Fig. 1　Diagram of model structure

3.1 边缘检测算法

边缘提取是通过不同的算子对图像进行不同的提取。边缘检测的主要工具是边缘检测模板，本文采用 3 种边缘检测模板，分别是 Sobel 算子、Scharr 算子和 Laplacian 算子。

3.1.1 Sobel 算子

Sobel 检测算子对灰度渐变和噪声较多的图像处理效果较好，是一种离散性差分算子，用来计算图像亮度的灰度近似值。在图像上使用此算子，将会产生对应的灰度矢量或法矢量。

3.1.2 Scharr 算子

Scharr 算子是对 Sobel 算子差异性的增强，因此两者在检测图像边缘的原理和使用方式上相同。Scharr 算子的边缘检测滤波尺寸为 3×3，可以通过将滤波器中的权重系数放大来增大像素值间的差异。

3.1.3 Laplacian 算子

Laplacian 算子是一种二阶导数算子，将在边缘处产生一个陡峭的零交叉。Laplacian 算子是各向同性的，能对任何走向的界线和线条进行锐化，无方向性，这是该算子区别于其他算法的最大优点所在。

3.2 多路径提取方法

图像分割准确度的提高需要尽可能多地融合多角度的图像信息来对图片目标进行预测。PSPNet 网络采用金字塔池化，其利用空间金字塔池化模块适应多尺度的特征图，融合不同角度提取的特征。ICNet[17] 使用了级联的图像输入，即低、中、高分辨率图像，采用了级联的特征融合单元，训练时使用了级联的标签监督。基于此，本文提出了空间路径、语义路径、边缘检测路径这 3 条特征提取路径来获取多层次的图像特征。

3.2.1 空间路径

在语义分割任务中，可以通过空洞卷积来保持特征图的高分辨率，以此得到足够的空间信息。如图 2 所示，本文提出了一条可以编码空间信息的特征提取路径，其输入为原始图像与边缘检测生成图像在通道方向拼接之后的 4 通道数据。空间特征提取路径只进行了 3 次下采样，即提取出了原始输入数据 1/8 大小的特征图，获得了较高分辨率的特征图，为后续预测图的输出提供了空间位置信息。

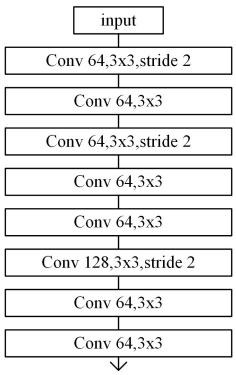


图 2 空间路径

Fig. 2 Space path

3.2.2 语义路径

在语义分割任务中，对图像进行密集预测要求输出特征图的感受野足够大，以确保在预测过程中没有忽略重要的信息。一般地，网络越深，感受野越大，分割预测效果越好。为了获得足够大的感受野，最简单的方法是在卷积操作中设置大卷积核。DeepLab 网络采用空洞空间金字塔池化模块，在不降低特征图分辨率的前提下获得了巨大的感受野。以上方法表明图像分割准确率的提高需要足够大的感受野。本文结合上述经验设计了如图 3 所示的语义路径，其输入数据为原始图像与边缘检测生成图进行通道融合的 4 通道数据，采用轻量化模块 Ghost 对输入数据进行 5 次下采样。为保证网络轻量化，在各个下采样过程中特征图的通道数不能过多。最终，在快速提取出图像中高级语义信息的同时，得到足够大的感受野。

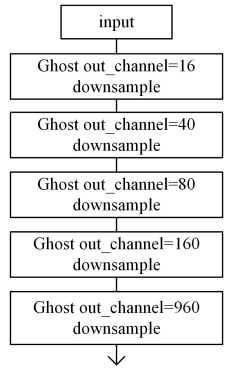


图 3 语义路径

Fig. 3 Semantic path

3.2.3 边缘检测路径

本文设计的边缘检测路径如图 4 所示，为实现对边缘检测生成图像的特征提取，我们使用深度可分离卷积替换标准卷积操作，将标准卷积操作进行了拆分。首先对特征图的每一个通道单独进行计算，然后通过 1×1 的点卷积对上一步输出的特征图在深度方向上进行加权组合，生成新的特征图，最后通过多个深度可分离卷积对已经处理过的输入图像提取出更具代表性的纹理特征。

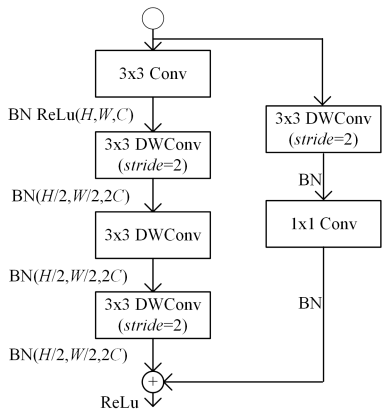


图 4 边缘检测路径

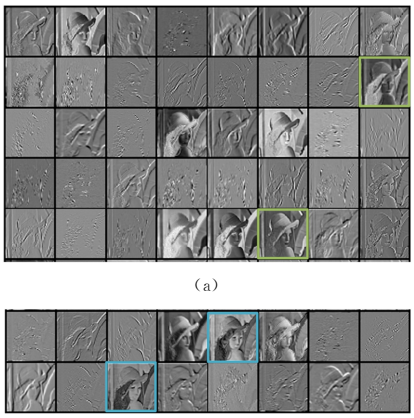
Fig. 4 Edge detection path

3.3 Ghost 模块

为加快深度学习的前向推理速度，一些研究人员提出了模型的压缩方法，如剪枝、量化、知识蒸馏等；还有一些则着重于高效的网络结构设计，如 MobileNet，ShuffleNet[18] 等。本文采用了一种轻量化的神经网络基础单元 Ghost 模块来提高

模型的前向推理速度。

一个训练好的深度神经网络通常会包含丰富的特征图,以保证对输入数据有全面的理解,但也因此产生了大量的冗余特征。图 5 为通过两次卷积产生的特征图,图 5(a)和图 5(b)分别对相似的特征图进行标记。可以看出,即使在网络特征提取的前几层就出现相似的特征图,在后续层次的卷积过程中也不可避免地会产生大量冗余。



(a)

(b)

图 5　特征图可视化

Fig. 5　Feature map visualization

如图 6 所示,在一般卷积操作中,给定输入数据 $X\in\mathbb{R}^{c\times h\times w}$,其中 c 是输入通道数,h 和 w 是高度和宽度,对于生成 n 个特征图的任意卷积的运算可以表示为:

$$Y=X*f+b \tag{1}$$

其中,$*$ 是卷积运算,b 是偏差项,$Y\in\mathbb{R}^{h'\times w'\times n}$ 是具有 n 个通道的输出特征图,h' 和 w' 为输出数据的高度与宽度,卷积核为 $f\in\mathbb{R}^{c\times k\times k\times n}$,$k\times k$ 是卷积核 f 的内核大小。由于 n 和通道数 c 在 CNNs 的后续操作中越来越大,因此所需的 FLOPs 数量也会增加。

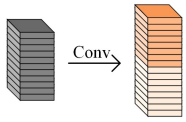


图 6　普通卷积

Fig. 6　Common convolution

针对上述卷积过程中产生的巨大计算量,Ghost 模块可以使用更少的参数来生成更多的特征图。图 7 给出了 Ghost 模块的变换过程,具体操作为将深度神经网络中的普通卷积层分成两个部分。首先通过普通卷积生成固定数量的特征图,然后对每个特征图进行线性运算 Φ,生成多种相似的特征图。与普通卷积神经网络相比,在不更改输出特征图大小的情况下,Ghost 模块减少了大量的参数,降低了计算复杂度。

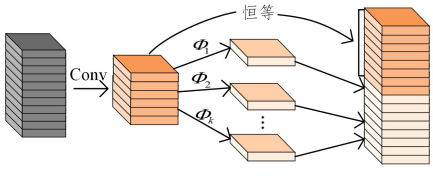


图 7　Ghost 模块

Fig. 7　Ghost module

4　实验及结果分析

4.1　实验数据

本文主要采用 CamVid 数据集作为实验数据。CamVid 数据集的分辨率为 480 像素×360 像素,其中包含 367 张训练图片、101 张验证图片以及 233 张测试数据。共有 11 个不同的类别,如建筑、树、天空、汽车等。该数据集分割样本少且像素分辨率低,分割难度大。此外,本文还采用了 Cityscapes 数据集对模型效果再次进行了论证。

4.2　实验环境

实验显卡型号为 NVIDIA RTX 2080Ti,使用随机梯度下降算法(Stochastic Gradient Descent,SGD),初始学习率设置为 0.0005,在训练过程中采用 Poly 学习策略,power 为 0.9,权重衰减系数为 0.003,训练轮数为 10000。

4.3　评价指标

本文实验中使用平均交并比(mean Intersection-over-Union,mIoU)和每秒处理帧数(Frames Per Second,FPS)作为评估算法分割准确率和处理速度的指标,使用浮点运算数(Floating Point Operations,FLOPs)和参数量来衡量模型计算的复杂度和规模。

mIoU 是真实值和预测值集合的交集与并集之比,用于评价算法的分割能力,其计算式如式(2)所示:

$$mIoU=\frac{1}{k+1}\sum_{i=0}^{k}\frac{TP}{FN+FP-TP} \tag{2}$$

其中,k 表示类别的数量,TP 表示预测为真正的数量,FN 表示预测为假负的数量,FP 表示预测为假正的数量。

FPS 用于评价算法速度,其表达式如式(3)所示:

$$FPS=\frac{N}{\sum_{j}^{N}T_j} \tag{3}$$

其中,N 表示图像数量,T_j 表示算法处理第 j 张图像的时间。

FLOPs 是浮点运算数,参数量是算法中包含的参数数量,用来衡量算法的复杂度。对于卷积操作,FLOPs 和参数量的计算式如式(4)、式(5)所示:

$$FLOPs=2HW(C_{\text{in}}K^2+1)C_{\text{out}} \tag{4}$$

$$Parameters=K^2\times C_{\text{in}}\times C_{\text{out}} \tag{5}$$

其中,C_{in} 是卷积层输入数据的通道数,C_{out} 是卷积层输出数据的通道数,K 是卷积核大小,H 和 W 分别为输入数据的高度和宽度。

4.4　算法性能分析与比较

首先,本文分别对 Sobel 算子、Scharr 算子以及 Laplacian 算子生成的图像输入本文模型进行对比,从分割速度和精度两方面进行有效评估;然后与模型 FCN,UNet,Enet,LedNet[19] 和 BiseNet 进行分割精度和速度上的比较;最后通过可视化结果进一步分析算法的分割性能。

4.4.1　边缘检测算子的有效性评估

为了验证边缘检测算子的有效性,本文设计了不同算子生成的图像在所提模型上的对比实验。首先移除所提模型的边缘检测路径,将此模型记为 Baseline,然后分别对具有边缘

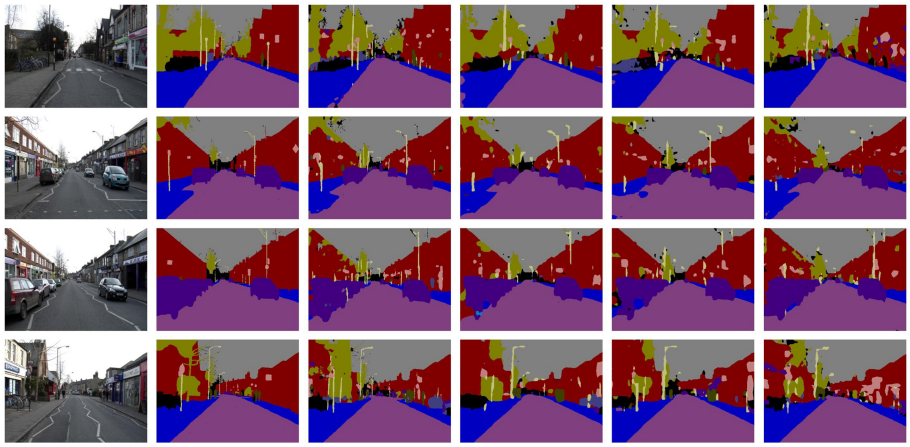
检测路径的网络进行评估，根据边缘检测算子的不同，将它们分别记为 Base+Sobel_3，Base+Sobel_5，Base+Scharr，Base+Laplacian。其中，Sobel_3 和 Sobel_5 分别代表 3×3 与 5×5 大小的 Sobel 算子生成的图像。表 1 列出了在 NVIDIA RTX 2080Ti 的实验环境下的消融实验结果。在 CamVid 数据集上的实验结果表明，3×3 大小的 Sobel 算子对模型分割精度提升最大，最多比 Baseline 模型提高了 3.1%；5×5 的 Sobel 算子和 Scharr 算子都有不错的效果，相比 Baseline 分别提升了 1.5%和 2.1%；而 Laplacian 算子对模型的提升效果较差，仅提升了 0.9%。Baseline 处理每张图像的运算时间比其他模型减少了 0.2 ms，证明边缘检测算子可以在很短的运算时间的前提下提高模型的分割准确率。

表 1 各个边缘检测算子有效性评估的对比

Table 1 Comparison of effectiveness evaluation of each edge detection operator

Method	*mIoU*/%	*t*/ms
Baseline	39.8	2.6
Base+Sobel_3	42.9	2.8
Base+Sobel_5	41.3	2.8
Base+Scharr	41.9	2.8
Base+Laplacian	40.7	2.8

图 8 给出了未进行边缘检测算子处理的 Baseline 和 3 种算子加入网络之后对图像处理后的可视化效果，可以看出，边缘检测算子对原图中各个目标分割界限更为清晰，由此说明边缘检测算子可以有效提高分割准确率。



(a)原图 (b)Groud Truth (c)Baseline (d)Sobel_3 (e)Scharr (f) Laplacian

图 8 本文方法在 CamVid 测试集上的可视化效果对比

Fig. 8 Comparison of visual effects of the proposed method on CamVid test set

本文算法对图像的多个角度进行了特征提取。为了更好地融合多方面提取的特征，本文提出了特征融合模块来结合原始图像提取的特征和边缘检测算子提取的图像特征，如图 9 所示。由表 2 可以看出，加入 FFM 模块之后，算法的精度最多提高了 1.1%，说明 FFM 模块能够很好地结合特征信息。

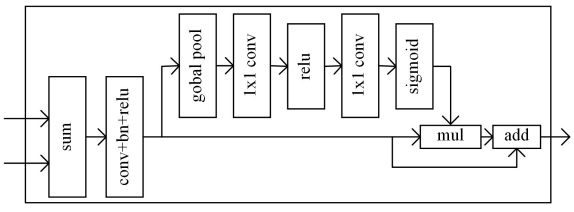


图 9 特征融合模块

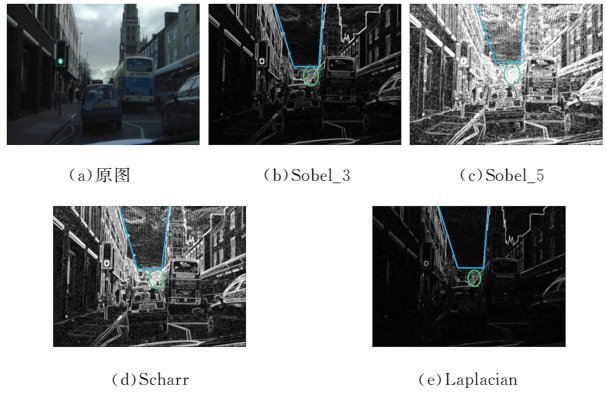
Fig. 9 Feature fusion module

表 2 FFM 模块的有效性评估对比

Table 2 Effectiveness evaluation and comparison of FFM

Method	*mIoU*/%	*t*/ms
Base+Sobel_3	41.8	2.8
Base+Sobel_3+FFM	42.9	2.8
Base+Scharr	41.2	2.8
Base+Scharr+FFM	41.9	2.8

图 10 给出了原图和 Sobel_3 算子、Sobel_5 算子、Scharr 算子以及 Laplacian 算子对图像处理之后的对比情况。在 4 种算子的对比中，分别用圆圈和直线标注了树木和天空区域。其中 Sobel_5 和 Scharr 算子提取的树木和天空轮廓信息更丰富，对物体的边缘信息更为敏感，但是语义分割强调不同物体间的差异性，过多的轮廓信息会造成物体本身划分不准确，进而导致各个目标之间的划分不准确。从图 10(e)中可以看出，Laplacian 算子对图像边缘信息的提取较少，对边缘信息不太敏感，可供分割的特征较少，使得分割效果不佳。Sobel_3 算子相比其他 3 种算子更强调提取各个目标之间的边缘信息，并且剔除了不相关的信息，保留了图像重要的结构属性。



(a)原图 (b)Sobel_3 (c)Sobel_5

(d)Scharr (e)Laplacian

图 10 边缘检测处理图

Fig. 10 Edge detection processing diagram

4.4.2　算法性能分析与比较检测

为验证本文算法的有效性，将其与 FCN，SegNet[20]，Enet，LinkNet 和 DfaNet 在 CamVid 数据集上进行分割精度与速度的对比，统一采用 mIoU 和 FPS 作为衡量指标。为了保证模型效果的有效性，将所有模型的训练参数设置一致，不对包括本文算法在内的任何算法进行针对性调优。

实验使用 360 像素×480 像素大小的 CamVid 数据集进行训练测试。表 3 列出了 NVIDIA RTX 2080Ti 显卡环境下的结果，图 11 为各个网络在 mIoU 指标上的柱形图。其中，本文算法在数据集上的处理速度为 349 frames/s，虽然分割速度慢于 BiseNet 模型，但是 mIoU 比 BiseNet 提高了 2.2%；本文算法在分割精度和速度上全面超越了 LedNet 和 Enet。因此，本文算法在分割准确率和处理图片速度上优于现有的实时分割算法。与 FCN 以及 U-Net 相比，本文算法在速度上有着很大的优势，可以确保在嵌入式设备上达到实时分割的效果。综合分析，本文算法实现了精确、高效的分割。

表 3　不同算法的性能对比

Table 3　Performance comparison of different algorithms

Method	*mIoU*/%	*t*/ms	FPS/(frames/s)
FCN	45.1	8.4	119.4
U-Net	47.8	14.6	68.7
BiseNet	40.7	2.2	457.6
Enet	40.2	3.8	305.8
LedNet	38.3	3.3	301.0
Ours	42.9	2.8	349.0

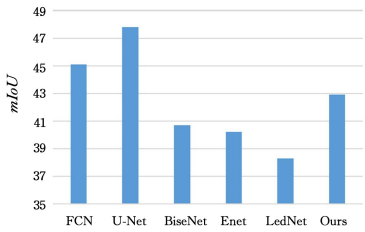


图 11　本文方法和其他方法的 mIoU 指标效果对比

Fig. 11　Comparison of the proposed method and other methods on mIoU indicators

对于实时语义分割模型，本文除了从分割精度和速度两个指标来判断模型的好坏，还考虑了模型的复杂度，即参数的个数和计算量。参数个数表示存储模型所需的存储空间；计算量表示输入数据在模型中需要的计算量，即模型进行分割所需的计算力。表 4 中，FLOPs 是通过输入 352 像素×480 像素的数据计算所得的计算力。可以看出，Enet 和 LedNet 算法参数量和所需计算力都要小于本文模型，但是本文算法兼顾了分割速度和精度，总体性能优于这两种算法，而 BiseNet 模型在 FLOPs 和参数量上的消耗都比本文算法大，无法适用于一些低算力的嵌入式设备。

表 4　不同模型的复杂度对比

Table 4　Complexity comparison of different models

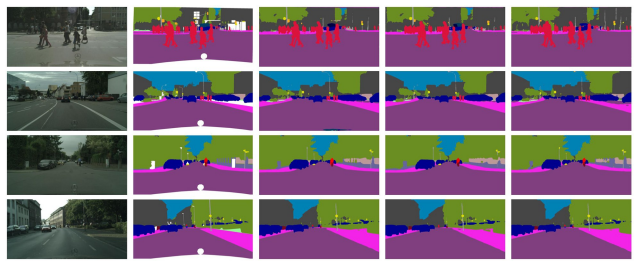
Method	*FLOPs*	*Parameters*
LedNet	1.34	0.92×10^6
ENet	0.79	0.35×10^6
BiseNet	4.90	13.39×10^6
Ours	2.90	3.86×10^6

由于 CamVid 数据集样本较小，为了更有效地说明本文算法的有效性，将本网络和 BiseNet 以及 ICNet 在 Cityscapes 数据集上进行 mIoU 和 FLOPs 两个指标的进行比较。表 5 中，FLOPs 为输入 2048 像素×1024 像素大小的 Cityscapes 数据图像计算所得的结果。可以看出，在输入大分辨率数据时，深度学习网络所消耗的计算力急剧增加，相比 ICNet 和 BiseNet 网络，本文网络的计算消耗较小，分割准确率也更高，在 Cityscapes 数据集上的分割准确率达到了 69.6%。图 12 给出了 3 种算法在 Cityscapes 数据集上的可视化效果。

表 5　不同模型在 Cityscapes 数据集上的对比

Table 5　Comparison of different models on Cityscape dataset

Method	*FLOPs*	*mIoU*/%
ICNet	121.22	67.7
BiseNet	60.86	68.4
Ours	35.90	69.6

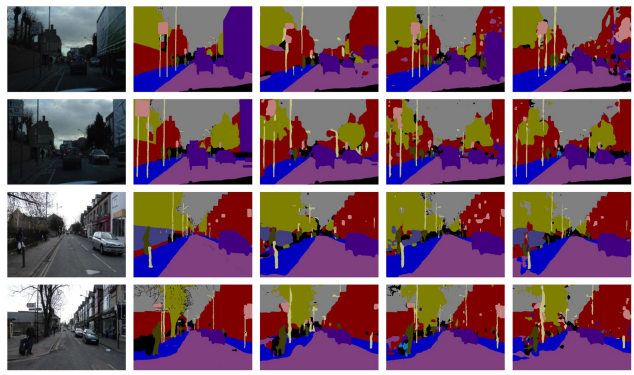


(a)原图　(b)Groud Truth　(c)Ours　(d)ICNet　(e)BiseNet

图 12　本文方法和其他方法在 Cityscapes 数据集上的可视化效果对比

Fig. 12　Comparison of visualization effects of the proposed method and other methods on Cityscapes dataset

在图 13 所示的可视化效果对比中，对于只有 360 像素×480 像素分辨率的图像，Enet 和 LedNet 分割出的目标对象包含着大量的其他目标信息；本文算法由于结合了边缘检测生成图像的特征，可以有效地对各个物体进行划分，分割出更完整的目标，可视化效果更好。



(a)原图　(b)Groud Truth　(c)Ours　(d)ICNet　(e)BiseNet

图 13　本文方法和其他方法在 CamVid 数据集上的可视化效果对比

Fig. 13　Comparison of visualization effects of the proposed method and other methods on CamVid test set

结束语　为满足实时语义分割任务的需求，本文提出了基于多路径特征提取的轻量化语义分割算法。首先利用边缘检测剔除图像中的噪声信息，保留图像重要的边缘和位置特征；其次，通过 3 条特征提取路径对图像进行特征提取，并

结合提取出的原始数据的高级语义信息、低层次的纹理信息以及边缘检测生成图像的特征信息；最后，通过特征融合模块将多方面信息融合进行推理预测，提高了模型的分割精度。本文将 Ghost 模块应用于语义特征路径的特征提取，提高了模型的整体分割速度。最后在 CamVid 和 Cityscapes 数据集上进行了一系列对比实验，实验结果表明本文算法兼顾了分割速度与精度。本文网络在语义特征提取过程中采用 Ghost 模块作为提高模型处理速度的方法。下一步将采取剪枝和知识蒸馏的方法对模型参数进行压缩，进一步提升网络的分割速度。

参 考 文 献

[1] CHEN Q S, TAO Y, SHEN F H, et al. Semantic segmentation of images based on contextual structure [J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2020, 32(2): 287-294.

[2] CHAO Y, XIAO N F. Research on Robotic Grasping Methods Based on Semantic Segmentation and Brain Computer Interface [J]. Journal of Chongqing University of Technology (Natural Science), 2020, 34(3): 128-136, 151.

[3] LECUN Y, BOTTOU L. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[4] LONG J, SHEKHAMER E, DARRELL T. Fully Convolutional Networks for Semantic Segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4): 640-651.

[5] RONNEBERGER O, FISCHER P, BROX T, et al. U-net: Convolutional networks for biomedical image segmentation [J]. Medical Image Computing and Computer Assisted Intervention, 2015, 28(4): 234-241.

[6] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Hawaii: IEEE Press, 2017: 2881-2890.

[7] 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 & Machine Intelligence, 2016, 40(4): 834-848.

[8] PASZKE A, CHAURASIA A, KIM A. ENet: a deep neural network architecture for real-time semantic segmentation [J]. arXiv: 1606.02147, 2016.

[9] YU C, WANG J, PENG C, et al. Bisenet: Bilateral segmentation network for real-time semantic segmentation [C]// Proceedings of the European Conference on Computer Vision. Munich, Germany, 2018: 325-341.

[10] LI H, XIONG P, FAN H, et al. Dfanet: Deep feature aggregation for real-time semantic segmentation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seoul: IEEE Press, 2019: 9522-9531.

[11] HAN K, WANG Y, TIAN Q, et al. GhostNet: More Features From Cheap Operations [C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE Press, 2020: 1577-1586.

[12] LIN G S, MILAN A, SHEN C H, et al. Refinenet: multi-path refinement networks for high-resolution semantic segmentation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 5168-5177.

[13] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, 2016: 2818-2826.

[14] CHOLLET F. Xception: deep learning with depth wise separable convolutions [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 1800-1807.

[15] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: inverted residuals and linear bottlenecks. conference [C]// IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 4510-4520.

[16] ZHUANG L, LI J, SHEN Z, et al. Learning Efficient Convolutional Networks through Network Slimming [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 2755-2763.

[17] ZHAO H, QI X, SHEN X, et al. Icnet for real-time semantic segmentation on high-resolution images [C]// Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany, 2018: 405-420.

[18] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: Practical guidelines for efficient CNN architecture design [C]// Proceedings of the European Conference on Computer Vision. Munich, Germany, 2018: 116-131.

[19] 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). Taipei: IEEE Press, 2019: 126-172.

[20] 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.

CHENG Cheng, born in 1996, postgraduate, is a member of China Computer Federation. His main research interests include deep learning and semantic segmentation.

JIANG Ai-lian, born in 1969, Ph.D, associate professor, is a member of China Computer Federation. Her main research interests include big data analysis and processing, feature selection, artificial intelligence and computer vision.

（责任编辑：柯颖）