

文章编号 1004-924X(2023)08-1217-11

基于改进 BiSeNet 的实时图像语义分割

任凤雷^{1,2}, 杨璐^{1,2*}, 周海波^{1,2}, 张诗雨^{1,2}, 何昕³, 徐文学⁴

- (1. 天津理工大学 天津市先进机电系统设计与智能控制重点实验室, 天津 300384;
2. 天津理工大学 机电工程国家级实验教学示范中心, 天津 300384;
3. 中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130033;
4. 天津卓越信通科技有限公司, 天津 300384)

摘要: 为了提升图像语义分割算法的性能, 使其同时满足准确性和实时性需求, 本文提出了一种基于改进 BiSeNet 的实时图像语义分割算法。首先, 通过使双分支网络头部共享以消除 BiSeNet 网络结构部分通道和参数的冗余, 同时有效提取图像的浅层特征; 然后, 将上述共享网络拆分为由细节分支和语义分支组成的双分支网络, 并分别用于提取空间细节信息和语义上下文信息; 此外, 在语义分支尾部引入通道和空间注意力机制以增强特征表达能力, 通过使用双注意力机制对 BiSeNet 算法进行优化以更有效地提取语义上下文特征; 最后, 对细节分支和语义分支的特征进行融合并通过上采样操作恢复至输入图像分辨率大小以实现图像语义分割。本文算法在 Cityscapes 数据集以 95.3 FPS 的实时性表现达到 77.2% mIoU 的准确性; 在 CamVid 数据集以 179.1 FPS 的实时性表现达到 73.8% mIoU 的准确性。实验结果表明, 本文算法在实时性和准确性方面获得了很好的平衡, 其语义分割性能相较于 BiSeNet 算法及其它现有算法得到了显著的提升。

关键词: 语义分割; 注意力机制; 实时性; 深度学习

中图分类号: TP394.1 **文献标识码:** A **doi:** 10.37188/OPE.20233108.1217

Real-time semantic segmentation based on improved BiSeNet

REN Fenglei^{1,2}, YANG Lu^{1,2*}, ZHOU Haibo^{1,2}, ZHANG Shiyu^{1,2}, HE Xin³, XU Wenxue⁴

- (1. Tianjin Key Laboratory for Advanced Mechatronic System Design and Intelligent Control, School of Mechanical Engineering, Tianjin University of Technology, Tianjin 300384, China;
2. National Demonstration Center for Experimental Mechanical and Electrical Engineering Education, Tianjin University of Technology, Tianjin 300384, China;
3. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China;
4. Transcend Communication Technology Tianjin Co., Ltd, Tianjin 300384, China)

* Corresponding author, E-mail: yanglu8206@163.com

Abstract: To improve the performance of image semantic segmentation on accuracy and efficiency for practical applications, in this study, we propose a real-time semantic segmentation algorithm based on im-

收稿日期: 2022-09-12; **修订日期:** 2022-10-01.

基金项目: 国家自然科学基金资助项目 (No. 51275209); 天津市自然科学基金重点项目资助 (No. 17JCZDJC30400); 广东省重点领域研发计划资助项目 (No. 2019B090922002)

proved BiSeNet. First, the redundancy of certain channels and parameters of BiSeNet is eliminated by sharing the heads of dual branches, and the affluent shallow features are effectively extracted at the same time. Subsequently, the shared layers are divided into dual branches, namely, the detail branch and the semantic branch, which are used to extract detailed spatial information and contextual semantic information, respectively. Furthermore, both the channel attention mechanism and spatial attention mechanism are introduced into the tail of the semantic branch to enhance the feature representation; thus the BiSeNet is optimized by using dual attention mechanisms to extract contextual semantic features more effectively. Finally, the features of the detail branch and semantic branch are fused and up-sampled to the resolution of the input image to obtain semantic segmentation. Our proposed algorithm achieves 77.2% mIoU on accuracy with real-time performance of 95.3 FPS on Cityscapes dataset and 73.8% mIoU on accuracy with real-time performance of 179.1 FPS on CamVid dataset. The experiments demonstrate that our proposed semantic segmentation algorithm achieves a good trade-off between accuracy and efficiency. Furthermore, the performance of semantic segmentation is significantly improved compared with BiSeNet and other existing algorithms.

Key words: semantic segmentation; attention mechanism; real time; deep learning

1 引言

图像语义分割作为计算机视觉领域的一项重要技术,旨在将每一个图像像素分类为相应的语义类别,其在自动驾驶、医学检测、机器人导航、场景解析、人机交互等领域均有着极为广泛的应用^[1-5]。

近年来,随着人工智能的飞速发展,深度学习技术被越来越广泛地应用到图像语义分割领域并取得了相对传统图像分割算法更加出色的成果。如全卷积网络(Fully Convolutional Networks, FCN)^[6]最早尝试将深度学习引入图像语义分割领域,并取得了优异的语义分割性能。DeepLab算法^[7]引入扩张卷积(Dilated Convolution)来保持特征图的分辨率,同时有效增大卷积核的感受野从而提升语义分割准确性。在此基础上不断涌现出的针对DeepLab算法的优化改进版本^[8-10]使语义分割的准确性得到了逐步提升。HRNet^[11]通过保持图像的高分辨率,同时并行的进行下采样以提取不同分辨率的特征,实现了高准确率的图像语义分割。然而,上述算法实时性较差,难以满足如自动驾驶等对实时性要求较高的领域的实际应用需求。

为了满足算法的实际应用部署需求,实时语义分割算法得到了越来越广泛的关注。ENet算

法^[12]通过设计轻量级网络结构在损失部分准确性的前提下增强了语义分割算法的实时性。SwiftNet算法^[13]分别使用低分辨率输入以获取语义特征和高分辨率输入以获取充分的细节信息。DFANet算法^[14]通过引入深度可分离卷积来提升算法的运行效率,并通过聚合多尺度特征来保证算法的准确性,其在Cityscapes数据集获得了71.3%的语义分割准确性和超过100FPS的实时性表现。M Fan等人^[15]提出了一种新的Short-Term Dense Concatenate (STDC)模块,并使用此模块取得了出色的实时性表现。FANet算法^[16]通过设计高效的注意力模块,将注意力机制引入了实时语义分割算法。BiSeNet系列算法^[17-18]提出了一种轻量级的双分支网络结构,分别用来提取图像的空间细节信息和语义上下文信息,此算法获得了出色的实时性和准确性表现,并成为目前最流行的实时语义分割算法之一。然而,BiSeNet系列算法的双分支结构在头部存在部分通道和参数的冗余,这会对算法的实时性产生一定程度的消极影响;此外,BiSeNet系列算法仅借鉴了通道注意力机制的思路,并没有充分发挥注意力机制的优势。总结来说,上述算法的语义分割性能仍然有待提升,如何平衡算法的准确性和实时性成为算法实际应用部署亟需解决的问题。

针对上述问题,本文拟在 BiSeNet 语义分割算法的基础上对其进行优化改进,以提升算法的性能,使实时性和准确性得到更好的平衡。本文的主要贡献总结如下:(1)通过共享双分支网络头部消除部分通道和参数的冗余以进一步提升算法的效率。(2)通过在语义分支中引入通道和空间注意力机制增强特征表达能力以保障算法的准确性。(3)本文算法在 Cityscapes 数据集和 CamVid 数据集的语义分割性能相较于 BiSeNet 算法及其他现有先进算法均得到了显著提升,证明了本文算法的优越性。

2 相关工作

2.1 图像语义分割

图像语义分割算法通过将图像分割成代表不同语义类别的像素区块,实现对场景的语义感知和解析。大量研究工作^[19-22]表明,作为一种密集预测视觉算法,高分辨率特征下的空间细节信息和低分辨率特征下的语义上下文信息对于获得高准确性语义分割结果都是必不可少的。

针对图像语义分割算法的实时性方面,目前主流算法^[17-20]通常使用双分支网络结构,即细节分支和语义分支,以保证算法准确性的同时提升算法的运行效率。此外,众多高效的轻量级骨干网络结构^[23-26]可用于进行网络结构的设计,从而进一步提升算法的推理速度。图像语义分割示意图如图1所示。



图1 图像语义分割示意图

Fig. 1 Diagram of semantic segmentation

2.2 注意力机制

注意力机制(Attention Mechanism)最初被用于自然语言处理领域,其在本质上是一种对资源进行自动重新分配的机制,由于与人类对外界事物的观察机制相似,如今已经在图像识别、目标检测、图像分割等不同视觉任务中得到了有效应用^[27-28]。注意力机制通过对输入特征各个部分

赋予不同的权重因子,从而引导网络模型主动提取更加关键和重要的特征而忽略无效的特征,使网络模型的特征表达能力得到增强^[35-36]。

3 本文算法

BiSeNet 算法(左图)和本文算法(右图)的网络结构简图如图2所示。

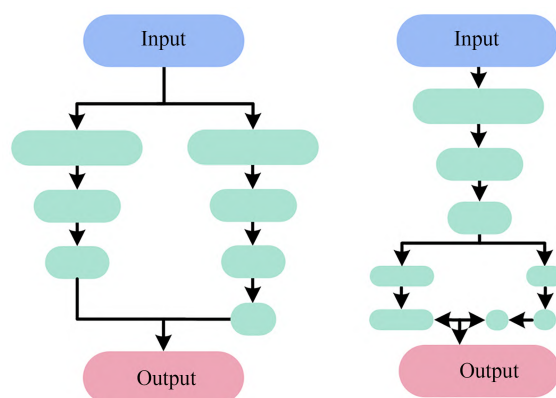


图2 BiSeNet算法及本文算法网络结构简图

Fig. 2 Network comparison of BiSeNet and our algorithm

本文算法拟针对 BiSeNet 语义分割算法进行优化改进,首先对图像的浅层特征进行提取并被后续的双分支网络共享;然后在上述基础上进行网络结构的拆分,得到细节分支和语义分支,并分别提取空间细节特征和语义上下文特征;最后将上述双分支的特征进行融合并上采样得到语义分割结果。本文所提出的基于改进 BiSeNet 的实时图像语义分割算法框图如图3所示(彩图见期刊电子版),其中,紫色虚框表示头部共享网络部分;绿色虚框表示拆分双分支网络部分,由粉色块表示的细节分支和黄色块表示的语义分支组成;PAM 表示 Position Attention Mechanism,即空间注意力机制;CAM 表示 Channel Attention Mechanism,即通道注意力机制;FFM 表示 Feature Fusing Module,即特征融合模块,特征图旁的数字表示其相对于输入图像分辨率大小的比值。

3.1 网络头部共享

BiSeNet 算法使用双分支网络结构分别提取输入图像高分辨率特征下的空间细节信息和低分辨率特征下的语义上下文信息以提升算法的

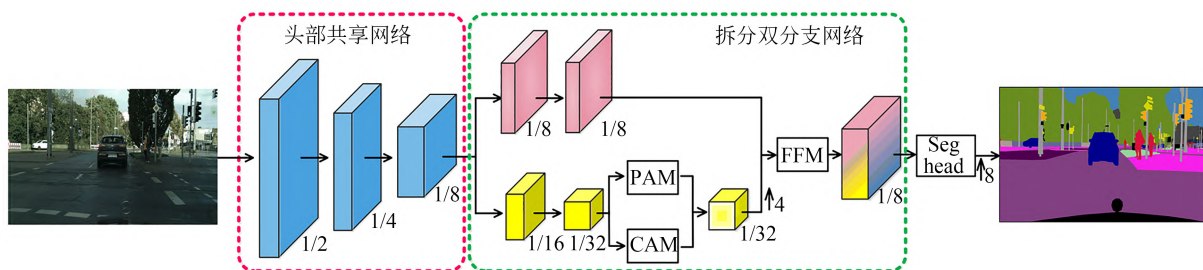


图3 本文图像语义分割算法示意图

Fig. 3 Diagram of our proposed semantic segmentation algorithm

语义分割性能。然而,其双分支结构在头部存在部分通道和参数的冗余,消除上述冗余有利于在几乎不损失分割精度的前提下提升算法的实时性。因此,本文算法对其进行改进,通过对双分支网络结构的头部进行共享减少其通道和参数的冗余。

为了兼顾网络的准确性和实时性,我们使用轻量级的残差网络 ResNet-18^[23]作为骨干网络进行算法结构的设计。通过使用步长为2的卷积操作和池化操作,将特征图分辨率降采样至输入图像的1/8。由此,提取得到丰富的图像特征并被后续的双分支网络共享。

3.2 拆分双分支网络

基于上述头部共享网络,可以初步提取得到图像的浅层特征。继而,本文算法通过将头部共享网络拆分为双分支网络以增强特征的表达能力,从而提升算法的语义分割准确性。其中,双分支网络由细节分支和语义分支组成,分别用于进一步提取高分辨率特征下的空间细节信息和低分辨率特征下的语义上下文信息。

3.2.1 细节分支

如图3所示,为了避免如物体边缘等图像细节信息的丢失导致语义分割准确率下降,本文算法双分支网络细节分支不采用任何下采样操作,而使用连续两次的步长为1的 3×3 卷积操作、批归一化操作^[29]和ReLU激活函数^[30]来提取丰富的空间细节信息,同时保持较高的特征图分辨率,即细节分支的特征图分辨率和网络头部共享部分输出的特征分辨率相同,均为输入图像的1/8。

3.2.2 语义分支

本文算法设计和细节分支平行的语义分支来提取丰富的语义上下文特征。为此,语义分支

通过对网络头部共享部分输出的特征进行下采样操作以获取低分辨率特征图并同时增大网络的感受野。因此,本文算法通过连续使用两次步长为2的卷积操作、批归一化操作和ReLU激活函数,最终将特征图分辨率降采样为输入图像的1/32。

值得一提的是,BiSeNet算法选择在语义分支中借鉴通道注意力机制^[28],通过使用全局平均池化操作作为特征通道赋予不同权重从而增大网络感受野并获取全局上下文信息。然而,上述注意力机制只针对特征的通道维度进行了权重分配,并没有涉及空间维度。

鉴于此,本文算法选择并行引入空间注意力机制和通道注意力机制,从而同时在空间维度和通道维度建模各个特征维度的重要程度,根据学习到的权重因子引导神经网络更有效地提取语义上下文信息。如图4所示,本文借鉴文献[31]的思路,针对空间注意力机制,对于上述语义分支提取到的特征图 $A \in R^{C \times H \times W}$,通过如下运算过程:

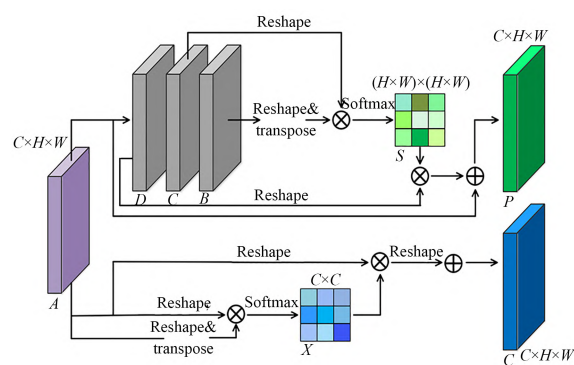


图4 空间和通道注意力机制示意图

Fig. 4 Diagram of position and channel attention mechanisms

$$S_{ji} = \frac{\exp(B_i \cdot C_j)}{\sum_{i=1}^N \exp(B_i \cdot C_j)}, \quad (1)$$

$$P_j = \alpha \sum_{i=1}^N (S_{ji} D_i) + A_j, \quad (2)$$

其中: B, C 和 D 均由 A 经过卷积和reshape操作得到, $S \in R^{(H \times W) \times (H \times W)}$ 表示根据空间相似性得到的权重图,并被用于得到 $P \in R^{C \times H \times W}$,权重因子 α 初始值设置为0,并随着网络的训练不断更新。由此,特征图中任意空间位置的特征值会被特征图所有的特征值及其对应位置权重的加权和所替代,由此,即可得到空间维度的全局上下文信息。相应的,针对空间注意力机制,有:

$$X_{ji} = \frac{\exp(A_i \cdot A_j)}{\sum_{i=1}^C \exp(A_i \cdot A_j)}, \quad (3)$$

$$C_j = \beta \sum_{i=1}^C (X_{ji} A_i) + A_j, \quad (4)$$

其中: $X \in R^{C \times C}$ 表示根据各特征通道的相似性得到的权重图,并被用于得到 $C \in R^{C \times H \times W}$,权重因子 β 初始值设置为0,并随着网络的训练不断更新。由此,即可在通道维度对特征进行增强。

上述通道注意力机制和空间注意力机制可以分别从通道维度和空间维度对特征进行增强,从而使网络更好的对语义上下文特征进行提取。值得注意的是,本文将上述空间和通道注意力机制嵌入到语义分支的尾部,由于其特征图分辨率仅为输入图像的1/32,因此不会增加过多的计算量,从而可以保证算法的实时性。

3.3 双分支特征融合

在上述基础上,本文算法根据文献[18]的思路设计特征融合网络,以对细节分支提取到的空间细节特征和语义分支提取到的语义上下文特征进行融合。鉴于双分支特征的互补性,此特征融合网络采用双向融合的方式,即对语义分支特征进行上采样并与细节分支特征进行交互,同时对细节分支特征进行下采样并与语义分支特征进行交互。特征融合网络示意图如图5所示。

最后,融合后的特征经过一个简单的分割头模块并通过双线性插值上采样操作输出最终的语义分割结果。该分割头模块由卷积核为 3×3 的卷积、批归一化、ReLU非线性激活函数和卷积

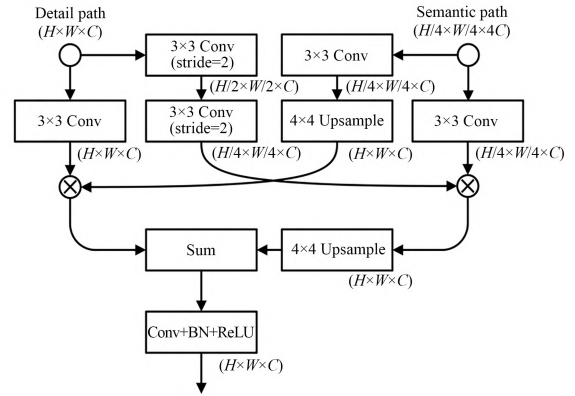


图5 特征融合网络示意图

Fig. 5 Diagram of feature fusing module

核为 1×1 的卷积操作组成,其输出维度设置为待分割的语义类别数。

4 实验与结果分析

针对本文所提出的基于改进 BiSeNet 的实时图像语义分割算法,本节对其进行了大量实验以验证所提出算法的有效性。其中,本文实验采用 Cityscapes 和 CamVid 图像语义分割数据集,实验基于 Ubuntu 18.04 OS, NVIDIA GeForce 2080Ti GPU 硬件平台。算法实现方面,使用 Pytorch 深度学习框架进行算法模型的实现和训练。

4.1 数据集简介

4.1.1 Cityscapes 数据集

Cityscapes 数据集^[32]是图像语义分割领域最流行的公共数据集之一。该数据集主要聚焦城市的街道驾驶场景。Cityscapes 数据集提供了 5 000 张精确标注了像素级语义分类的图像和 20 000 张粗略标注的图像。本文实验只使用精确标注的图像进行算法的验证,其中 2 975 张图像用作训练集,500 张图像用作验证集,1 525 张图像用作测试集。该数据集共包含汽车、道路等 19 个语义类别,所有图像的分辨率均为 $2\,048 \times 1\,024$ 。

4.1.2 CamVid 数据集

CamVid 数据集全称为 The Cambridge-driving Labeled Video Database^[33],该数据集由剑桥大学发布,针对城市交通道路自动驾驶场景,主要包含 701 张截取自视频序列的图像,通常将其

中的 367 张图像用作训练集, 101 张图像用作验证集, 其余的 233 张图像用作测试集。该数据集一般使用汽车、行人、建筑物等 11 种语义类别进行分割精度的评估, 所有的图像分辨率为 960×720 。

4.2 训练策略简介

借鉴 BiSeNet, 本文使用小批量随机梯度下降 (mini-batch SGD) 作为优化器进行网络模型的训练, 设置初始学习率为 $2.5e^{-2}$, momentum 为 0.9。采用“poly”策略, 使学习率按照迭代次数依次衰减 $(1 - \frac{iter}{\max_iter})^{\text{power}}$, 其中 power 设置值为 0.9。采用随机水平翻转、随机裁剪等方法进行数据增强。

4.3 评价指标

算法的准确性方面, 参考数据集要求, 本文使用平均交并比, 即 (Intersection-over-Union, mIoU) 作为评价指标进行衡量, 其定义如下式:

$$IoU = \frac{P_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}}, \quad (5)$$

$$mIoU = \frac{\sum_{i=0}^k IoU_i}{k+1}, \quad (6)$$

其中: $k+1$ 表示数据集定义的语义类别总数, i 表示真实值, j 表示预测值。

算法的实时性方面, 本文使用帧率, 即 Frames Per Second (FPS) 作为评价标准进行衡量, 通过统计网络前向推理 500 次的时间并求其平均值进而计算得到算法的帧率。

4.4 消融实验

如前文所述, 本文使用头部共享网络初步提取得到图像浅层特征, 并对输入图像进行下采样, 继而使用拆分双分支网络, 分别设计细节分支和语义分支并各自提取丰富的高分辨率特征下的空间细节信息和低分辨率特征下的语义上下文信息。为了验证各组成模块的有效性, 本文在 Cityscapes 数据集进行了如下消融实验, 实验结果如表 1 所示, 其中, HSM 表示 Head Shared Module, 即头部共享网络; SDBM 表示 Split Dual Branch Module, 即拆分双分支网络; DB 表示 Detail Branch, 即细节分支; SB 表示 Semantic Branch, 即语义分支。由表可知, 单独使用头部

共享网络进行图像语义分割仅可获得 58.4% 的准确率, 而通过加入细节分支和语义分支, 分别可以带来 7.3% 和 13.4% 的准确率提升。当同时使用头部共享网络和由细节分支及语义分支组成的拆分双分支网络进行语义分割时, 其准确率可以达到 77.5%。上述消融实验充分验证了头部共享网络和拆分双分支网络的有效性, 且细节分支和语义分支有助于使网络结构有效提取足量互补的特征来提升语义分割算法的性能。

表 1 各组成模块消融实验结果

Tab. 1 Ablation study of different modules

HSM	SDBM		mIoU/%
	DB	SB	
✓			58.4
✓	✓		65.7
✓		✓	71.8
✓	✓	✓	77.5

本文提出在语义分支中引入空间注意力机制和通道注意力机制引导神经网络更有效地提取语义上下文信息。为了验证其有效性, 本文在 Cityscapes 数据集进一步设计了如下消融实验, 其实验结果如表 2 所示, 其中, PAM 和 CAM 分别表示空间注意力机制和通道注意力机制; Parallel 表示两种注意力机制并行使用; Series 表示两种注意力机制串行使用。由表可知, 空间注意力机制和通道注意力机制可以分别在空间维度和通道维度对特征进行增强, 从而提升算法的语义分割准确性。其中, 当并行使用空间注意力机制和通道注意力机制时, 可以使算法的语义分割

表 2 空间及通道注意力机制消融实验结果

Tab. 2 Ablation study of position and channel mechanisms

PAM	CAM	Fusion mode	mIoU/%
			72.9
✓			76.4
	✓		75.8
✓	✓	Parallel	77.5
✓	✓	Series	77.3

准确性从 72.9% 提升至 77.5%。

针对双分支网络的特征融合,为了验证本文算法所使用的特征融合方式的有效性,我们将其与双通道特征相加(summation)、双通道特征拼接(concatenation)两种方式进行了对比,对比结果如表 3 所示,其中 FFM 表示 Feature Fusing Module,即本文所使用的特征融合模块。由表可知,本文所使用的特征融合模块相比较简单的特征相加或拼接能够更好地对细节信息和语义上下文信息进行融合。

表 3 特征融合方式消融实验结果
Tab. 3 Ablation study of feature fusion

Sum	Con	FFM	mIoU/%
✓			76.3
	✓		76.9
		✓	77.5

4.5 定性分析

基于所提出的图像语义分割算法,本文分别在 Cityscapes 数据集和 CamVid 数据集上对其进行了定性实验验证。本文算法在上述两个数据集的定性实验结果分别如图 6 和图 7 所示。为了更好地体现本文算法相对于 BiSeNet 算法的提升,本文同时列举了上述两种算法的语义分割定性结果,其中,第 1 列为输入图像;第 2 列为 BiSeNet 算法的语义分割结果图像;第 3 列为本文算法的语义分割结果;第 4 列为数据集提供的 ground-truth 图像,即准确标注有各语义类别的分割结果图像。

由上述实验结果可以看出,本文所提出的基于改进 BiSeNet 的实时图像语义分割算法可以在不同的道路交通场景下实现对输入图像的语义分割,且其相对于 BiSeNet 算法具有更好的分割性能,语义分割准确性得到了提升。

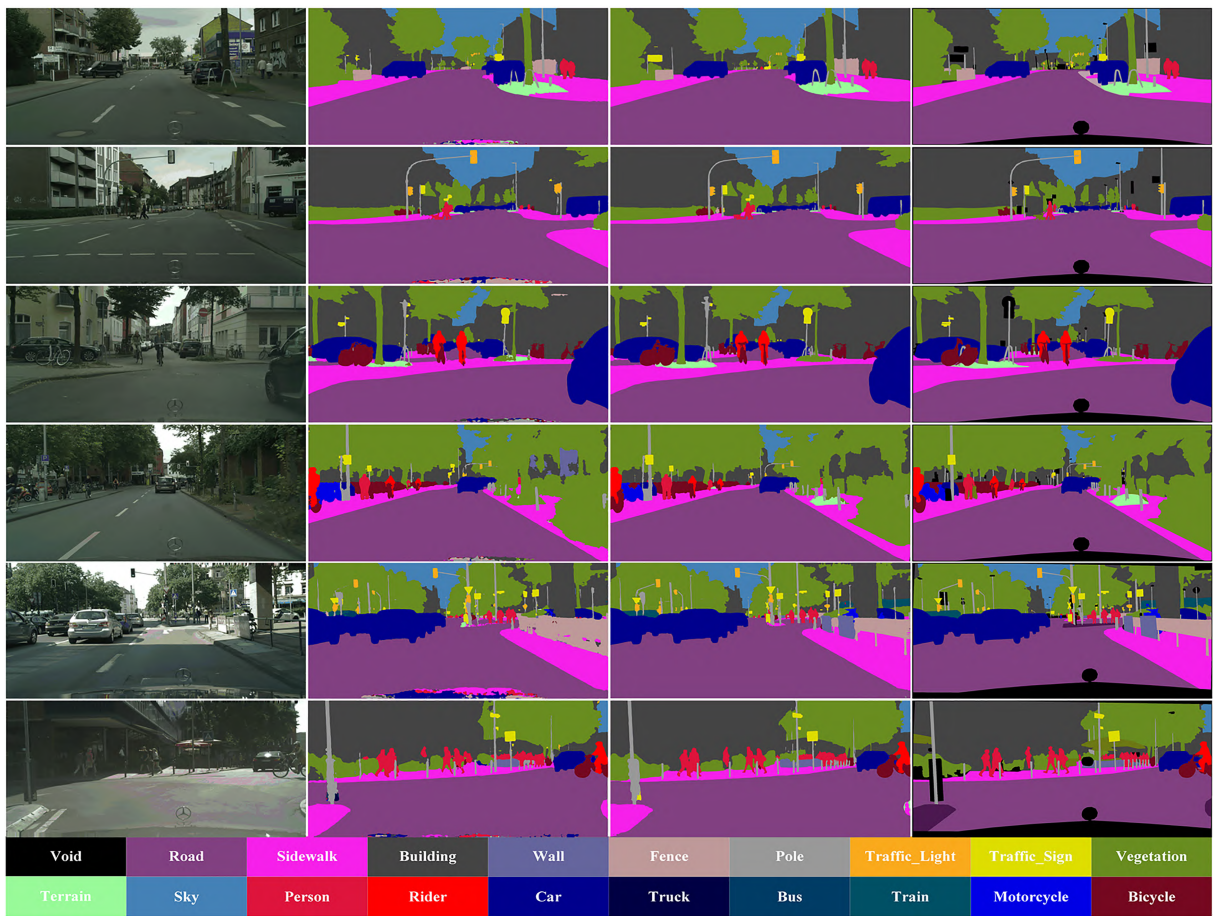


图 6 Cityscapes 数据集定性实验结果
Fig. 6 Qualitative results on Cityscapes

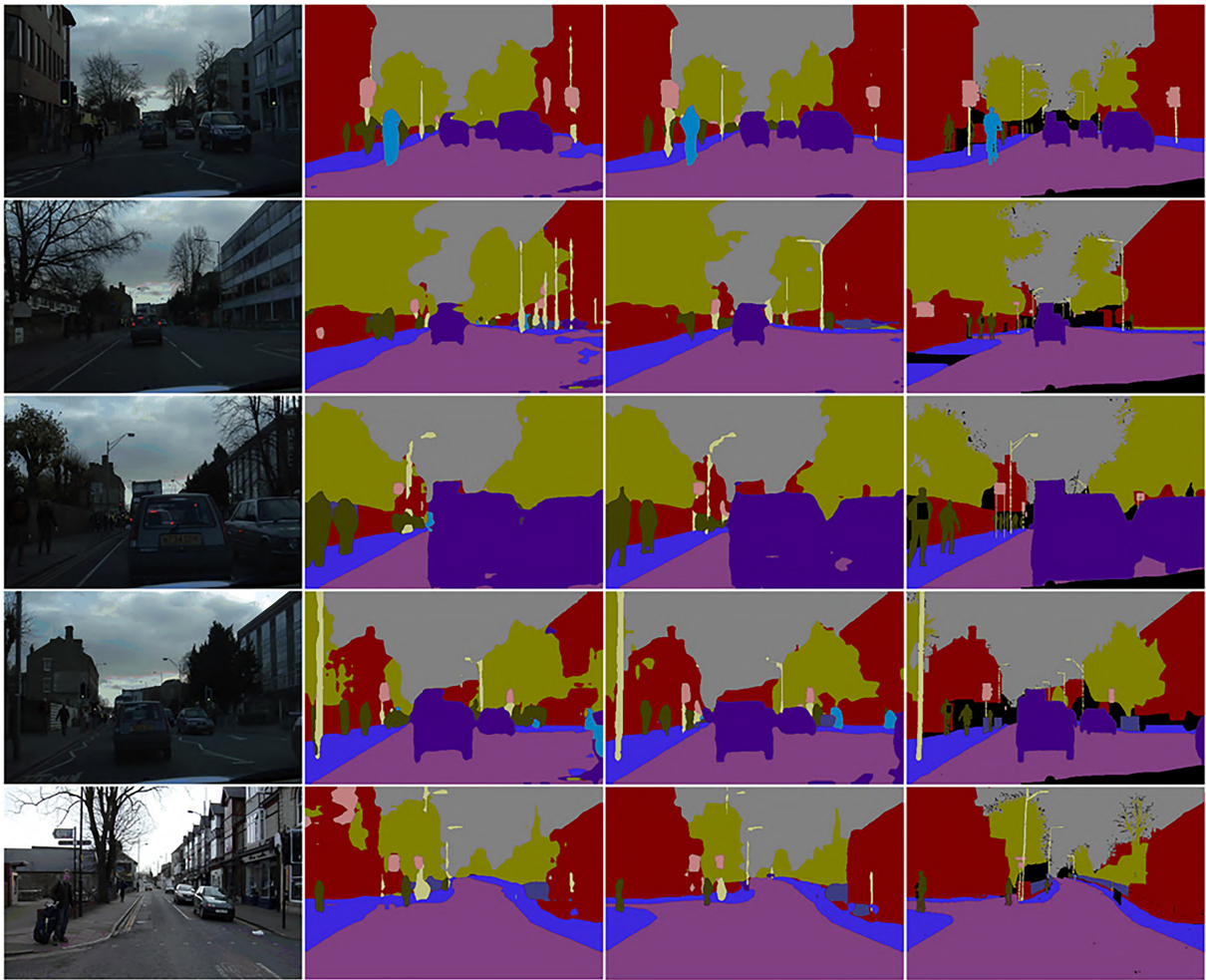


图7 CamVid数据集定性实验结果
Fig. 7 Qualitative results on CamVid

4.6 定量分析

针对 Cityscapes 数据集和 CamVid 数据集，本文算法和目前较为先进的图像语义分割算法的定量实验结果分别如表 4 和表 5 所示。

表 4 Cityscapes 数据集定量实验结果
Tab. 4 Quantitative results on Cityscapes

Method	mIoU/%		FPS
	Val	Test	
SegNet ^[34]	—	56.1	16.7
ENet ^[12]	—	58.3	76.9
SwiftNet ^[13]	75.4	75.5	39.9
BiSeNet ^[17]	74.8	74.7	65.5
BiSeNetV2 ^[18]	73.4	72.6	156
Fast-SCNN ^[20]	68.6	68.0	123.5
CABiNet ^[22]	76.6	75.9	76.5
Our Method	77.5	77.2	95.3

表 5 CamVid 数据集定量实验结果
Tab. 5 Quantitative results on CamVid

Method	mIoU/%	FPS
ENet ^[12]	51.3	61.2
SwiftNet ^[13]	72.6	—
BiSeNet ^[17]	68.7	116.3
BiSeNetV2 ^[18]	72.4	124.5
Our Method	73.8	179.1

由表可知,本文算法的语义分割性能达到了目前领先的水平,相比较 BiSeNet 等其他流行的实时语义分割算法在具有更好的实时性表现的同时,其准确性得到了一定提升,具体地,本文算法在 Cityscapes 数据集以 95.3 FPS 的实时性表现达到 77.2% mIoU 的准确性;在 CamVid 数据集以 179.1 FPS 的实时性表现达到 73.8% mIoU

的准确性。表明本文算法可以在准确性和实时性方面达到更好的平衡。

5 结 论

本文根据语义分割算法的实际应用需求,提出了一种基于改进BiSeNet的实时图像语义分割算法。本文算法在BiSeNet算法的基础上针对其存在的不足进行优化改进,首先,通过对双分支网络结构的头部进行共享减少BiSeNet算法存在

的部分通道和参数的冗余;其次,拆分为由细节分支和语义分支组成的双分支网络,并在语义分支中并行引入空间注意力机制和通道注意力机制以更有效地提取语义上下文信息;最后,融合细节分支和语义分支的特征并通过上采样操作实现对图像的语义分割。实验结果表明,相比较BiSeNet等现有语义分割算法,本文算法在准确性和实时性方面达到了更好的平衡,其语义分割性能得到了显著提升。本文算法可以基本满足图像语义分割的实际应用部署需求。

参考文献:

- [1] 王中宇,倪显扬,尚振东. 利用卷积神经网络的自动驾驶场景语义分割[J]. 光学精密工程, 2019, 27(11): 2429-2438.
WANG ZH Y, NI X Y, SHANG ZH D. Autonomous driving semantic segmentation with convolution neural networks [J]. *Opt. Precision Eng.*, 2019, 27(11): 2429-2438. (in Chinese)
- [2] RONNEBERGER O, FISCHER P, BROX T. *U-Net: Convolutional Networks For Biomedical Image Segmentation* [M]. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015: 234-241.
- [3] CHEN J, LU Y, YU Q, et al. *TransUNet: Transformers Make Strong Encoders For Medical Image Segmentation* [EB/OL]. 2021: arXiv: 2102.04306. <https://arxiv.org/abs/2102.04306>
- [4] 王雅男,王挺峰,田玉珍,等. 基于改进的局部表面凸性算法三维点云分割[J]. 中国光学, 2017, 10(3): 348-354.
WANG Y N, WANG T F, TIAN Y Z, et al. Improved local convexity algorithm of segmentation for 3D point cloud [J]. *Chinese Optics*, 2017, 10(3): 348-354. (in Chinese)
- [5] REN F L, ZHOU H B, YANG L, et al. ADPNet: Attention based dual path network for lane detection [J]. *Journal of Visual Communication and Image Representation*, 2022, 87: 103574.
- [6] 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. IEEE, 2015: 3431-3440.
- [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 and Machine Intelligence*, 2018, 40(4): 834-848.
- [8] CHEN L C, PAPANDREOU G, SCHROFF F, et al. *Rethinking Atrous Convolution for Semantic Image Segmentation* [EB/OL]. 2017: arXiv: 1706.05587. <https://arxiv.org/abs/1706.05587>
- [9] CHEN L C, ZHU Y K, PAPANDREOU G, et al. *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation* [M]. Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 833-851.
- [10] 任风雷,何昕,魏仲慧,等. 基于DeepLabV3+与超像素优化的语义分割[J]. 光学精密工程, 2019, 27(12): 2722-2729.
REN F L, HE X, WEI ZH H, et al. Semantic segmentation based on DeepLabV3+ and superpixel optimization [J]. *Opt. Precision Eng.*, 2019, 27(12): 2722-2729. (in Chinese)
- [11] WANG J D, SUN K, CHENG T H, et al. Deep high-resolution representation learning for visual recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(10): 3349-3364.
- [12] PASZKE A, CHAURASIA A, KIM S, et al. *ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation* [EB/OL]. 2016: arXiv: 1606.02147. <https://arxiv.org/abs/1606.02147>
- [13] WANG H C, JIANG X L, REN H B, et al. SwiftNet: real-time video object segmentation

- [C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 20-25, 2021, Nashville, TN, USA. IEEE, 2021: 1296-1305.
- [14] 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. IEEE, 2020: 9514-9523.
- [15] FAN M Y, LAI S Q, HUANG J S, *et al.* Rethinking BiSeNet for Real-Time Semantic Segmentation [C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 20-25, 2021, Nashville, TN, USA. IEEE, 2021: 9711-9720.
- [16] HU P, PERAZZI F, HEILBRON F C, *et al.* Real-time semantic segmentation with fast attention [J]. *IEEE Robotics and Automation Letters*, 2021, 6(1): 263-270.
- [17] YU C Q, WANG J B, PENG C, *et al.* BiSeNet: Bilateral Segmentation Network for Real-Time Semantic Segmentation [M]. Computer Vision - EC-CV 2018. Cham: Springer International Publishing, 2018: 334-349.
- [18] YU C Q, GAO C X, WANG J B, *et al.* BiSeNet V2: bilateral network with guided aggregation for real-time semantic segmentation [J]. *International Journal of Computer Vision*, 2021, 129 (11) : 3051-3068.
- [19] HONG Y, PAN H, SUN W, *et al.* Deep Dual-Resolution Networks for Real-Time and Accurate Semantic Segmentation of Road Scenes [EB/OL]. 2021: arXiv: 2101.06085. <https://arxiv.org/abs/2101.06085>
- [20] POUDEL R P K, LIWICKI S, CIPOLLA R. Fast-SCNN: Fast Semantic Segmentation Network [EB/OL]. 2019: arXiv: 1902.04502. <https://arxiv.org/abs/1902.04502>
- [21] 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. IEEE, 2017: 6230-6239.
- [22] KUMAAR S, LYU Y, NEX F, *et al.* CABiNet: Efficient Context Aggregation Network for Low-Latency Semantic Segmentation [C]. 2021 *IEEE International Conference on Robotics and Automation (ICRA)*. May 30-June 5, 2021, Xi'an, China. IEEE, 2021: 13517-13524.
- [23] 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. IEEE, 2016: 770-778.
- [24] SANDLER M, HOWARD A, ZHU M L, *et al.* MobileNetV2: Inverted Residuals and Linear Bottlenecks [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018. Salt Lake City, UT. IEEE, 2018: 4510-4520.
- [25] 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. IEEE, 2018: 6848-6856.
- [26] CHOLLET F. Xception: deep learning with depthwise separable convolutions [C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 1800-1807.
- [27] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need [J]. *Advances in neural information processing systems*, 2017, 30.
- [28] HU J, SHEN L, ALBANIE S, *et al.* Squeeze-and-excitation networks [C]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. April 29, 2019, IEEE, 2019: 2011-2023.
- [29] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]. *Proceedings of the 32nd International Conference on International Conference on Machine Learning-Volume 37*. New York: ACM, 2015: 448-456.
- [30] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks [C]. *Proceedings of the fourteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings*, 2011: 315-323.
- [31] 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. IEEE, 2020: 3141-3149.
- [32] 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)*. IEEE, 2016: 3213-3223.
- [33] BROSTOW GJ, FAUQUEUR J, CIPOLLA R. Semantic object classes in video: a high-definition ground truth database [J]. *Pattern Recognition Letters*, 2009, 30(2): 88-97.
- [34] 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.
- [35] 王曦, 于鸣, 任洪斌. UNET与FPN相结合的遥感图像语义分割 [J]. *液晶与显示*, 2021, 36(3): 475-483.
- WANG X, YU M, REN H E. Remote sensing image semantic segmentation combining UNET and FPN [J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(3): 475-483. (in Chinese)
- [36] 沈言善, 王阿川. 基于深度学习的遥感图像地物分割方法 [J]. *液晶与显示*, 2021, 36(5): 733-740.
- SHEN Y SH, WANG A CH. Remote sensing image feature segmentation method based on deep learning [J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(5): 733-740. (in Chinese)

作者简介:



任凤雷(1991—),男,河北沧州人,博士,讲师,2015年于吉林大学获学士学位,2020年于中国科学院长春光学精密机械与物理研究所获博士学位,主要从事数字图像处理,自动驾驶视觉环境感知方面的研究。E-mail:ren-fenglei15@mails.ucas.edu.cn; ren-fl@email.tjut.edu.cn

通讯作者:



杨璐(1982—),女,天津人,博士,副教授,硕士生导师,2005年于哈尔滨工业大学获得学士学位,2007年于吉林大学获得硕士学位,2011年于吉林大学获得博士学位,主要从事计算机视觉、机器学习等方面的研究。E-mail: yanglu8206@163.com