

DOI: 10.13382/j.jemi.B2205751

融合注意力机制与重影特征映射的无人机交通场景目标轻量级语义分割*

樊博 高玮玮 单明陶 方宇

(上海工程技术大学机械与汽车工程学院 上海 201620)

摘要:针对轻量级语义分割算法应用于无人机高分辨率交通场景图像分割时存在边缘信息模糊、小目标特征提取准确性较差的问题,提出一种融合注意力机制与重影特征映射的轻量级语义分割算法。首先在 BiSeNet V2 算法语义分支 8 倍和 16 倍下采样过程嵌入混合注意力模块,重新分配深层特征图权重,增强局部关键特征提取能力;然后采用重影特征映射单元优化传统卷积层,进一步降低运算成本;最后使用动态阈值损失函数监督训练,调节高损失困难样本训练权重。利用 UAVid 数据集对改进后的算法进行训练并测试,发现算法平均交并比(mean intersection over union, mIoU)为 52.7%,较改进前的模型提升 7.8%,且当输入图像尺寸为 1 280×736 时推理速度达到 81.6 FPS,满足实时分割要求。结果表明,该算法能较好地适应复杂交通场景,有效改善边缘信息模糊和小目标分割准确性较差的问题。

关键词: 语义分割;无人机交通场景;轻量化;注意力机制;重影特征映射;损失函数

中图分类号: TP751.1 **文献标识码:** A **国家标准学科分类代码:** 510.40

Lightweight semantic segmentation of UAV traffic scene objects combining attention mechanism and ghost feature mapping

Fan Bo Gao Weiwei Shan Mingtao Fang Yu

(School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

Abstract: To solve the problems of blurred edge information and poor accuracy of small targets feature extraction when the lightweight semantic segmentation algorithm is applied to the UAV high-resolution traffic scenes image segmentation, a lightweight semantic segmentation algorithm combining attention mechanism and ghost feature mapping is proposed. Firstly, the hybrid attention module is embedded in the semantic branch 8-fold and 16-fold down-sampling process of the BiSeNet V2 to redistribute the weights of the deep feature maps and enhance the local key feature extraction ability. Then the ghost feature mapping unit is used to optimize the traditional convolution layers to further reduce the computational cost. Finally, the dynamic threshold loss function is applied to supervise the training, adjusting the training weights of the high-loss difficult samples. Using the UAVid dataset to train and test the improved algorithm, it is found that the mIoU is 52.7%, which is 7.8% higher than the BiSeNet V2. When the input images size is 1 280×736, the inference speed can reach 73.6 FPS, meeting the real-time segmentation requirements. The results show that the algorithm can be well adapted to complex traffic scenes, and can effectively improve the problems of blurred edge information and poor accuracy of small objects.

Keywords: semantic segmentation; UAV traffic scenes; lightweight; attention mechanism; ghost feature mapping; loss function

0 引言

随着交通运输业蓬勃发展,城市交通基础设施日益完善,交通高速发展的同时也给人们出行带来极大的安

全隐患,全国各地交通事故屡见不鲜,严重的交通拥堵和事故给交通监管带来极大压力^[1]。无人机作为一个灵活轻巧的空中搭载平台,在公路巡查、违章抓拍、流量监测、应急指挥以及勘察取证等交通管理方面得到广泛应用^[2]。近年来人工智能技术日新月异,将深度学习图像

识别方法应用于交通管理成为可能。语义分割作为图像识别一项关键技术,也得到迅速发展,FCN^[3]、U-Net^[4]、DeepLab^[5]、PSPNet^[6]等语义分割网络被广泛应用。这些网络模型往往因为参数量大,在实际应用时存在很大的局限性。随着现实生产活动对快速响应要求提高,ICNet^[7]、DFANet^[8]、BiSeNet^[9]等实时语义分割算法得到广泛关注。尽管这些轻量语义分割网络能够提高实时推理速度,但常常为提升效率而牺牲准确性,在复杂场景应用效果不佳。目前关于交通场景语义分割研究大多聚焦于地面车载视角,面向低空无人机飞行视角的研究较少。石敏等^[10]采用基于双注意力机制的轻量语义分割网络 DALNet 应用于城市交通场景图像分割,取得较好的效果;刘步实^[11]设计一种包含多尺度分支和级联特征融合单元的轻量级 SP-ICNet 语义分割模型实现对道路场景环境的快速感知;王晓霏等^[12]使用结合拓扑结构和全局上下文感知的 TGCA-Net 网络完成无人机遥感影像道路提取任务;王之博等^[13]基于 Mask R-CNN 框架对遥感无人机影像中交通标志进行分割识别。

交通场景复杂多变,所含物体类别多,无人机视角下拍摄的图像中各种植被、建筑、汽车以及行人等目标存在遮挡重叠现象,边界信息不明显,且汽车和行人等小目标在图像中占有很少像素比例,特征提取过程中的池化等操作在扩大感受野时会降低图像分辨率,容易丢失小目标特征信息,因此分割效果并不理想。这些都是轻量化语义分割模型应用于无人机交通场景所需解决的关键问题。

目前轻量化语义分割算法多采用多分支结构,不同路径分别负责提取不同层次的特征信息并进行融合,能较好地适用于细节信息丰富、语义类别复杂的场景,且由于两条分支并行计算,分割效率也能得到保证,故选择多分支分割网络中效果较好的 BiSeNet V2^[14] 网络应用于交通场景图像语义分割研究。BiSeNet V2 的双分支结构虽然能够一定程度上提取目标差异化特征,但因忽视特征图不同通道和位置特征的区别导致在面对复杂交通场景时分割结果较为粗糙,存在边缘信息不明显、小目标特征提取准确度较低等问题。为此,在 BiSeNet V2 网络语义分支下采样特征提取过程引入注意力机制提高局部关键特征提取能力,且采用重影特征映射单元降低模型计算量以适应高分辨率航拍图像输入,同时使用动态阈值损失函数调节高损失困难样本训练权重,从而构建更加精准高效的无人机交通场景语义分割模型。

1 改进 BiSeNet V2 算法的无人机交通场景语义分割模型

BiSeNet V2 算法是在 BiSeNet 算法的基础上改进而

来,其不同路径负责获取图像不同层次特征信息,取得分割精度与速度的平衡。BiSeNet V2 算法采用并行的双分支结构:细节分支(detail branch)和语义分支(semantics branch),其分别针对特征提取时空信息丢失和图像上下文信息缺失的问题而设计;其中细节分支使用包含多个 3×3 的树状卷积层以保留丰富空间细节特征;而语义分支采用茎模块(stem block)进行快速下采样,通过多级聚集膨胀层(gather-and-expansion layer, GE)级联获取深层语义特征,同时在末端设计上下文嵌入模块(context embedding block, CE)获取全局上下文信息。为融合不同尺度特征,双边引导聚合层(bilateral guided aggregation layer)通过平均池化(average pooling)和上采样(upsampling)加权融合两条分支信息。由于无人机交通场景下获得的图像具有细节信息丰富、语义类别复杂的特点,想要获得较佳的分割效果,需要根据交通场景图像特征对 BiSeNet V2 进行改进。首先在 BiSeNet V2 特征提取过程嵌入混合注意力模块,增强模型局部特征提取能力;然后引入重影特征映射单元优化卷积运算过程,适应模型高分辨率航拍图像输入要求;最后采用动态阈值损失函数进一步提高模型对小目标样本的识别能力。

1.1 嵌入混合注意力模块优化特征提取

无人机视角下的交通场景图像具有目标类别复杂、边界不清晰且汽车及行人等目标像素占比少的特点,这给图像分割带来极大的困难。由于语义分支通过 Stem 模块中步长为 4 和 GE 模块中步长为 2 的卷积层快速降低特征图分辨率,同时为兼顾运算效率使用少量卷积核,因而深层特征图中仍保留大量场景细节信息,而 CE 模块中的全局平均池化(global average pooling)并不能针对性地关注特征图中局部关键信息,导致算法应用于复杂交通场景图像分割时存在结果边缘模糊,小目标分割效果不理想等问题。

混合注意力模块^[15]包括通道注意力模块(channel attention module, CAM)、空间注意力模块(spatial attention module, SAM)以及残差连接,其整体结构如图 1 所示。混合注意力模块能够获取全局特征中具有关键作用的局部特征信息,其中通道注意力关注特征图通道权重信息,空间注意力关注特征图像素点权重信息,而残差连接在不增加额外参数和计算复杂度的情况下增强特征输出。在 BiSeNet V2 特征提取过程嵌入混合注意力模块能够对特征图权重进行重新分配,更好地提取局部精细化特征,改善物体边缘细节,提升对行人等小目标分割性能。由于在细节分支嵌入混合注意力模块会增加模型计算成本,且细节分支与语义分支融合前其特征图分辨率为原始图像的 $1/8$,其中包含丰富的空间细节信息,故仅在 BiSeNet V2 语义分支 8 倍和 16 倍下采样过程嵌入混合注意力模块。

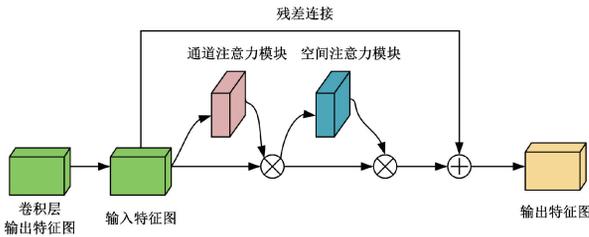


图1 混合注意力模块
Fig.1 Hybrid attention module

1.2 引入重影特征映射单元优化卷积运算

无人机低空航拍图像分辨率较高,分割前为减少计算复杂度往往会缩放图像尺寸,导致图像中关键目标像素数量减少,细节损失严重,影响算法分割精度,故可引入重影特征映射单元优化卷积运算过程,以适应高分辨率图像实时分割要求。重影特征映射单元^[16]通过应用一系列线性变换降低特征提取过程中的卷积运算量,其过程可分为两个部分:第1部分利用卷积核大小为 3×3 传统卷积生成少量原始特征图,第2部分通过 3×3 深度卷积对原始特征图映射处理,最后将两部分特征图拼接输出,结构如图2所示。由于BiSeNet V2算法通过细节分支和语义分支中的Stem模块与GE模块级联提取图像特征信息,故引入重影特征映射单元对其进行优化能够取得显著效果。其中细节分支通过3个卷积块提取图像浅层特征,每个卷积块使用多个 3×3 标准卷积,使用重影特征映射单元对其进行优化;而语义分支的Stem模块下采样过程分两部分进行,其运算量主要来自每部分的卷积计算,对其使用重影特征映射单元进行优化,优化后的Stem模块结构如图3所示;GE模块通过 3×3 卷积层以聚集特征信息并扩展到高维空间,在此基础上使用重影特征映射单元进行优化能够有效降低运算量,优化后的GE模块结构如图4所示。在BiSeNet V2特征提取过程中使用重影特征映射单元,在保证模型特征提取能力的同时进一步使算法轻量化,满足高分辨率航拍图像实时分割要求。

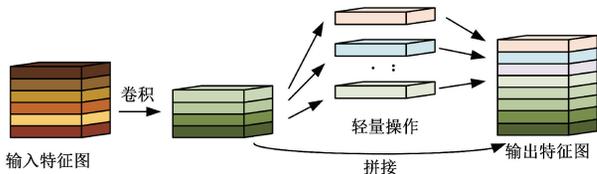


图2 重影特征映射单元
Fig.2 Ghost feature mapping module

1.3 损失函数改进

BiSeNet V2 利用主损失函数加 4 个辅助损失函数来

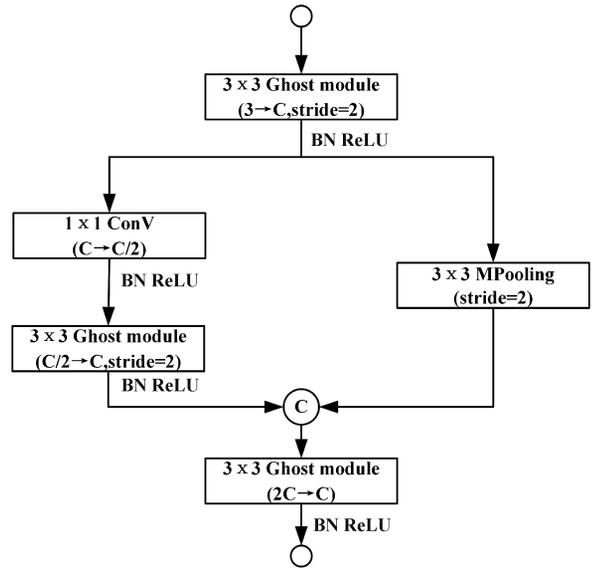


图3 使用重影特征映射单元的 Stem 模块
Fig.3 Stem block with ghost feature mapping module

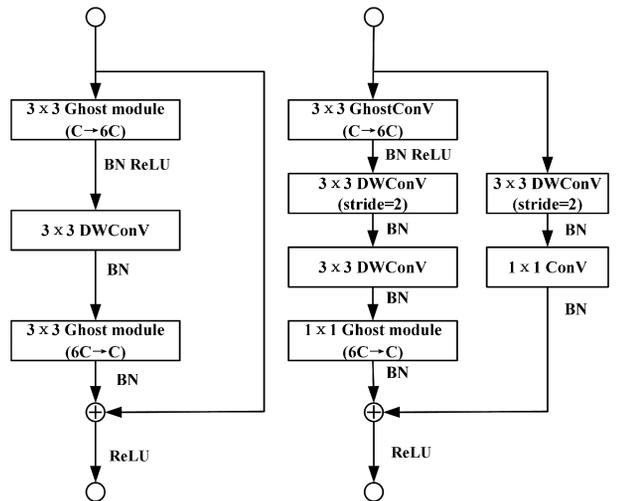


图4 使用重影特征映射单元的 GE 模块
Fig.4 GE block with ghost feature mapping module

监督网络训练,所有损失函数都是交叉熵损失函数。网络训练时根据裁剪图像大小,每个裁剪图片都可能具有数千个标记的像素需要预测,然而许多像素点很容易被区分出来,尤其是位于语义区域的中心部分,继续从这些像素学习很难提高语义分割效果^[17],基于以上考虑采用动态阈值损失函数,迫使网络在训练期间针对性关注像素比例较少、区域独立的小目标,改进后的损失函数为:

$$l = - \frac{1}{\sum_{i=1}^M \sum_{j=1}^C \varepsilon(y_i = j, p_{ij} < \lambda)} \times \left(\sum_{i=1}^M \sum_{j=1}^C \varepsilon(y_i = j, p_{ij} < \lambda) \log p_{ij} \right) \quad (1)$$

其中, M 表示像素点总数量; C 表示总类别数目; y_i 表示第 i 个像素点标签值; $\varepsilon(\cdot)$ 为示性函数, 满足条件时函数值取 1, 否则为 0; p_j 表示第 i 个像素点预测为第 j 类的概率值; λ 表示动态阈值, 取值随模型表现动态变化。

改进后的算法在 BiSeNet V2 语义分支 8 倍和 16 倍下采样过程嵌入混合注意力模块, 且采用包含卷积核大小为 3×3 的标准卷积与深度卷积的重影特征映射单元优化传统卷积过程, 同时使用动态阈值损失函数监督网络训练, 其整体结构如图 5 所示。

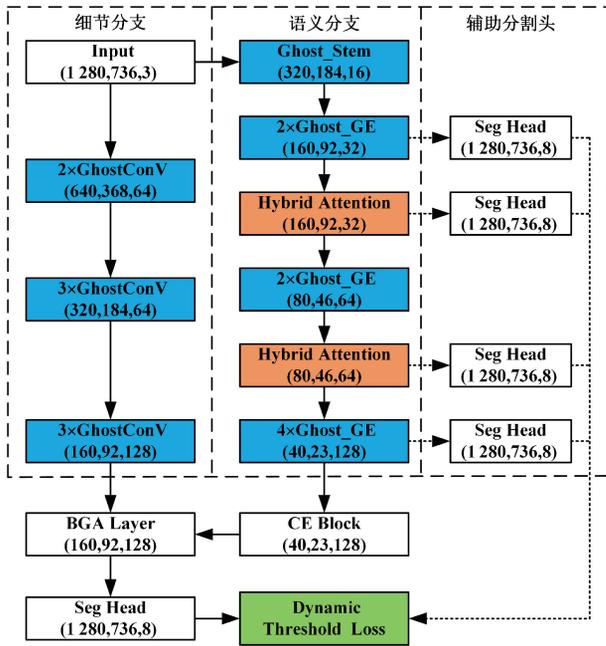


图 5 改进后的 BiSeNet V2 网络框架

Fig. 5 Framework of the improved BiSeNet V2 network

2 实验与结果分析

2.1 实验平台及数据

实验基于深度学习框架 PyTorch 实现, 实验平台配置为: 使用计算机处理器 Intel Core i5-11500, 其运行内存为 16 GB; 使用独立显卡 Nvidia RTX3060, 其显存为 12 GB; 使用 Cuda10.2 和 Cudnn7.51 加速 GPU 运算; 使用 Numpy 等 Python 包进行数据处理; 使用 Matplotlib 库完成分割结果可视化等操作。

实验过程采用公开的无人机语义分割数据集 UAVid^[18] 对算法进行训练和验证。该数据集共有 420 张无人机高分辨率图像, 其中训练集、验证集和测试集的图像数量分别为 200 张、70 张和 150 张, 包含植被、建筑、汽车、行人等 8 个物体类别。训练过程对数据集实施预处理, 对图像采取随机裁剪、随机缩放以及水平翻转等数据

增强方式提高模型泛化能力。训练过程参数设置为: UAVid 2020 训练集图像原始尺寸为 $3\ 840\ \text{pixels} \times 2\ 160\ \text{pixels}$, 经随机裁剪后尺寸为 $1\ 280\ \text{pixels} \times 736\ \text{pixels}$; 随机缩放的尺度列表为 $(1.00, 1.25, 1.50, 1.75, 2.00)$ 。

2.2 性能指标

轻量语义分割算法应用于复杂交通场景图像分割时, 应结合实际情况既关注算法精度又注重算法运行效率。衡量语义分割算法精度的指标有像素精度 (pixel accuracy, PA)、平均像素精度 (mean pixel accuracy, mPA) 和平均交并比 (mean intersection over union, mIoU) 等标准, 其中 mPA 和 mIoU 都是基于图像全局水平的性能指标, 而 mIoU 更能反映交通场景图像中各目标的区域一致性, 真实表示算法实际分割情况, 因此选择 mIoU 作为评价算法精度的性能指标, 其计算公式如式 (2) 所示。另外, 结合使用实时推理速度 (frames per second, FPS) 指标表示算法实际运行效率, 其计算公式如式 (3) 所示。

$$mIoU = \frac{1}{k} \sum_{i=1}^k \frac{p_{ii}}{t_i + \sum_{j=1}^k (p_{ji} - p_{ii})} \quad (2)$$

$$f_{FPS} = \frac{N}{\sum_{j=1}^N T_j} \quad (3)$$

式中: k 代表像素类别数; t_i 代表类别为 i 像素总数量; p_{ii} 代表实际类别为 i , 预测类别为 i 像素数量; p_{ji} 代表实际类别为 j , 预测类别为 i 的像素数目; N 表示图像数量, T_j 表示算法处理第 j 幅图像的时间。

2.3 实验结果与分析

1) 实验参数设置

模型训练过程采用动态阈值损失函数监督网络训练, 训练时自动选取有多多样性和高损失的困难像素样本作为训练样本, 其中困难像素样本选取方式为: 按损失值降序排序后, 若大于阈值的样本数多于初始设定值, 则选择所有大于阈值的样本; 否则, 按设定值依次选择样本。动态阈值为第 $512 \times \text{batchsize}$ 个困难像素样本所属语义类别预测概率值, 其中 batchsize 表示单批次训练图像样本数量。样本数量设定值公式如式 (4) 所示, 其中 size_0 和 size_1 分别表示输入图像的宽和高。

$$n = \frac{\text{batchsize} \times \text{size}_0 \times \text{size}_1}{16} \quad (4)$$

模型训练参数设置为: 批尺寸 (batch size) 设置为 5; 训练次数 (epoch) 设置 200 次; 初始学习率设置为 0.001, 100 次后学习率减小至 0.000 1; 采用 Adam 优化器更新网络训练权重。

2) 实验结果评估

为验证混合注意力模块、重影特征映射单元以及动态阈值损失函数3种组分对模型整体性能提升的有效性,以BiSeNet V2作为基准网络,在UAVid训练集上进行模型训练后,于测试集上进行实验,采用mIoU和FPS指标评估算法性能,各实验参数设置保持一致。具体实验结果数据如表1所示。

表1 不同组分在测试集上的性能评估

Table 1 Performance evaluation of different components on the test set

混合注意力模块	×	√	×	√	√	
重影特征映射单元	×	×	√	√	√	
动态阈值损失函数	×	×	×	×	√	
IoU/%	障碍物	48.3	53.4	46.8	52.9	53.2
	建筑	78.9	79.4	78.8	77.8	77.5
	道路	60.2	69.5	58.7	65.4	68.7
	停放汽车	24.4	33.7	24.2	29.9	34.4
	树木	68.1	73.5	66.8	71.4	72.3
	低矮植被	43.3	64.9	41.5	59.4	62.1
	行人	7.9	15.4	6.4	11.2	13.8
	行驶汽车	27.9	38.8	29.7	32.8	39.4
	mIoU/%	44.9	53.6	44.1	50.1	52.7
FPS	65.2	62.4	83.7	81.3	81.6	

(注:“×”表示未添加,“√”表示添加)

具体说明如下:表1第1列数据是以BiSeNet V2为基准网络,不包含各部分的实验结果;第2、3列表示同样以BiSeNet V2为基准网络,分别对混合注意力模块、重影特征映射单元部分的有效性验证;第4、5列分别为对动态阈值损失函数部分的有效性验证以及包含各部分的完整网络的实验数据。由表1数据发现:引入混合注意力模块后,各个类别交并比(intersection over union, IoU)均有较大幅度提升,其中停放汽车、行驶汽车和行人等小目标交并比指标分别提升9.3%、10.9%和7.5%,说明引入混合注意力模块的模型特征提取能力得到增强,能够在复杂交通场景中更好地保留小目标特征信息;引入重影特征映射单元后,mIoU仅降低约0.8%,而模型实时推理速度提高18.5 FPS,表明包含重影特征映射单元的模型能够在基本保证分割准确率的同时大幅提高分割速度;由表1第4、5列数据对比发现,引入动态阈值损失函数后,各个类别IoU均有不同程度的提升,而停放汽车、行驶汽车和行人等小目标IoU分别提升4.5%、6.6%和2.6%,表示包含动态阈值损失函数的模型能够进一步提高对小目标的分割能力,提升模型训练效果;由表1第1、5列数据对比可知,包含各组分的完整模型mIoU提高约7.8%,推理速度提高约16.4 FPS,表明改进后的模型能够结合各部分的优点,达到分割精

度与速度的平衡。

为验证算法的先进性,以mIoU和前向推理速度FPS为评价指标,在相同的参数设置下,将改进后算法分别与FCN-8s^[3]、PSPNet^[6]、SwiftNet^[19]、Fast-SCNN^[20]、BiSeNet^[9]算法分割结果进行对比,结果如表2所示。由表2数据发现:在分割精度方面,改进后的算法相比FCN-8s、PSPNet、SwiftNet、Fast-SCNN及BiSeNet分别高出9.2%、5.0%、0.6%、9.9%以及1.3%;而在推理速度方面,改进后的算法相比其余算法分别高出78、50.3、31.2、13.0以及29.2 FPS,能够保持一定的速度优势。因此,改进后的轻量级语义分割算法能够充分利用混合注意力模块、重影特征映射单元以及动态阈值损失函数的优点,实现算法精度与速度的平衡,从而满足无人机交通场景实时语义分割要求。

表2 不同算法分割性能对比

Table 2 Comparison of segmentation performance of different algorithms

算法	Backbone	Parameter/M	mIoU/%	FPS
FCN-8s	VGG16	132.5	43.5	3.6
PSPNet	ResNet18	24.1	47.7	31.3
SwiftNet	ResNet18	11.9	52.1	50.4
Fast-SCNN	—	2.6	42.8	68.6
BiSeNet	ResNet18	13.6	51.4	52.4
Ours	—	12.7	52.7	81.6

改进后的算法与BiSeNet V2在无人机交通场景下的图像分割结果对比如图6所示。在无人机交通场景图像中,既有轮廓较大的对象,如大片相连的植物、建筑物等,又有轮廓较小的对象,如道路上行驶的汽车、行人等。由图6可以看出,BiSeNet V2存在植被等大轮廓对象分割边缘不完整,汽车等小轮廓对象丢失的情况。相对而言,改进后的算法分割结果边缘较为清晰,细节特征明显,内部孤立点较少,汽车以及行人等小目标也能较为明显地识别,证明其更加适用于分辨率高且细节特征多的无人机交通场景。

3 结论

无人机视角下交通场景图像中各语义类别存在遮挡重叠现象,边界信息不明显,且汽车和行人等目标像素占比少,直接应用BiSeNet V2算法进行分割时结果存在边缘模糊、内部孤立点多、小目标分割效果不理想的问题。针对以上问题,提出一种融合注意力机制与重影特征映射的轻量化语义分割算法。该算法首先在语义分支8倍

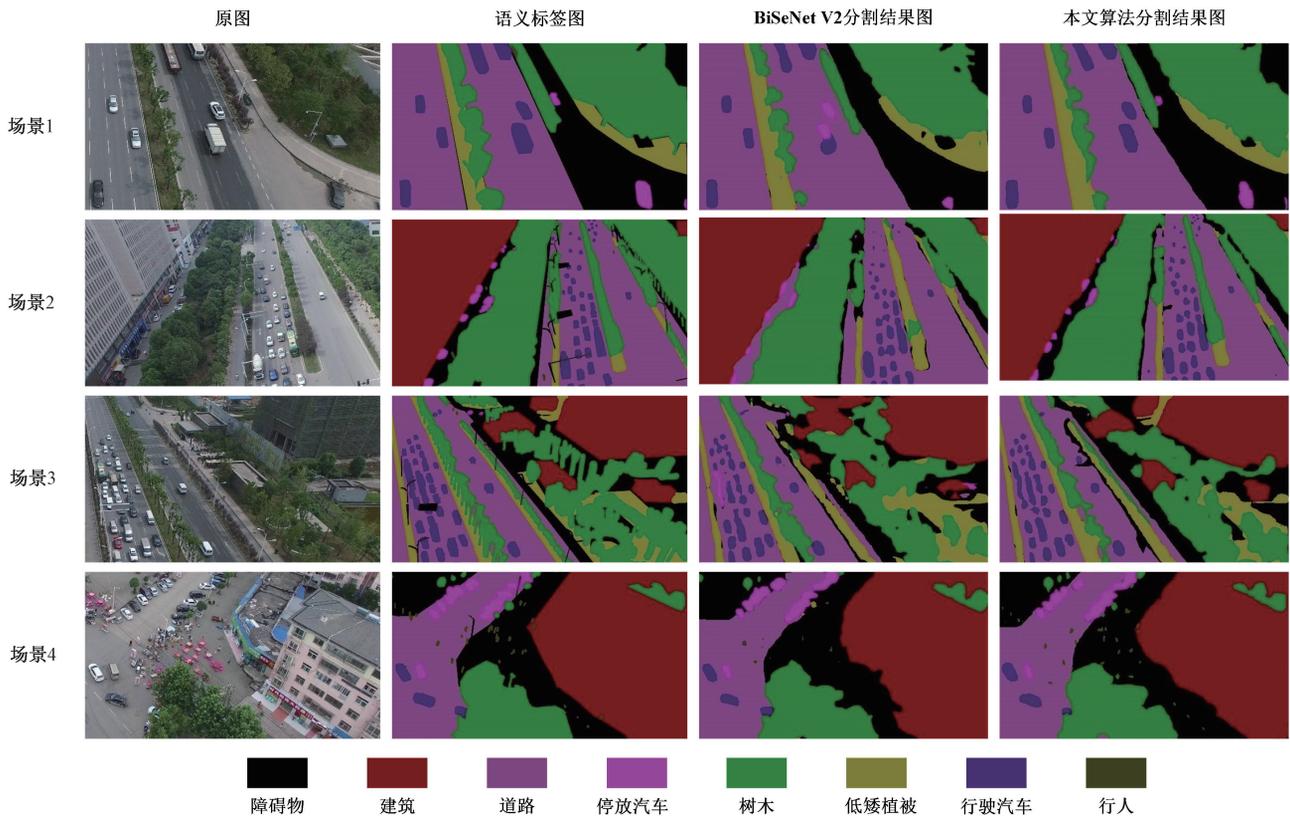


图 6 UAVID 数据可视化结果

Fig. 6 The visualization results on UAVID

和 16 倍下采样过程嵌入混合注意力模块以重新分配特征图权重,增强网络局部关键特征提取能力;且出于网络轻量化目的,采用重影特征映射模块优化卷积层,降低模型计算成本;同时使用动态阈值损失函数调节高损失困难样本训练权重,迫使网络在训练期间更关注像素比例较少、区域独立的小目标特征。利用 UAVID 数据集进行测试,实验结果表明改进后的算法在分割精度和速度上均有较大提升,表明改进后的网络能够适应复杂交通场景,在满足实时分割要求的同时有效改善边缘信息模糊和小目标分割准确性较差的问题。由于无人机航拍图像在低空和高空的图像特征尺度差异较大^[21],而本文算法没有考虑物体多尺度特征差异,下一步将对此展开研究。

参考文献

- [1] 殷礼胜,魏帅康,孙双晨,等. 基于 FEEMD-SAPSO-BiLSTM 组合模型的短时交通流预测[J]. 电子测量与仪器学报, 2021, 35(10): 72-81.
YIN L SH, WEI SH K, SUN SH CH, et al. Short-term traffic flow forecast based on FEEMD-SAPSO-BiLSTM combined model[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(10): 72-81.

- [2] 董明明. 5G 移动通信下无人机高速公路巡检防碰撞技术研究[J]. 电子测量技术, 2021, 44(5): 41-45.
DONG M M. Research on UAV expressway patrol anti-collision technology under 5G mobile communication[J]. Electronic Measurement Technology, 2021, 44(5): 41-45.
- [3] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.
- [4] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234-241.
- [5] 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, 2017, 40(4): 834-848.
- [6] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing

- network [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2881-2890.
- [7] 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, 2018: 405-420.
- [8] LI H, XIONG P, FAN H, et al. DFANet: Deep feature aggregation for real-time semantic segmentation [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 9522-9531.
- [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, 2018: 325-341.
- [10] 石敏, 沈佳林, 易清明, 等. 快速超轻量城市交通场景语义分割 [J]. 计算机科学与探索, 2022, 16(10): 2377-2386.
- SHI M, SHEN J L, YI Q M, et al. Rapid and ultra-lightweight semantic segmentation in urban traffic scene [J]. Journal of Frontiers of Computer Science and Technology, 2022, 16(10): 2377-2386.
- [11] 刘步实. 复杂交通场景下轻量化视觉感知方法研究 [D]. 北京: 北京交通大学, 2021.
- LIU B SH. Research on lightweight visual perception methods in complex traffic scenes [D]. Beijing: Beijing Jiaotong University, 2021.
- [12] 王晓霏, 叶虎平, 廖小罕, 等. 一种深度学习的无人机影像道路自动提取方法 [J]. 测绘科学, 2021, 46(11): 106-113.
- WANG X F, YE H P, LIAO X H, et al. An automatic road extraction method from UAV imagery based on deep learning [J]. Science of Surveying and Mapping, 2021, 46(11): 106-113.
- [13] 王之博, 赵双明. 基于 Mask R-CNN 的道路交通标志识别 [J]. 测绘地理信息, 2022, 47(3): 119-122.
- WANG ZH B, ZHAO SH M. Road traffic sign recognition based on Mask R-CNN [J]. Journal of Geomatics, 2022, 47(3): 119-122.
- [14] YU C, GAO C, WANG J, 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.
- [15] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [C]. Proceedings of the European Conference on Computer Vision, 2018: 3-19.
- [16] HAN K, WANG Y, TIAN Q, et al. GhostNet: More features from cheap operations [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580-1589.
- [17] WU Z, SHEN C, HENGEL A. Bridging category-level and instance-level semantic image segmentation [J]. arXiv preprint arXiv:1605.06885, 2016.
- [18] LYU Y, VOSSelman G, XIA G S, et al. UAVid: A semantic segmentation dataset for UAV imagery [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 165: 108-119.
- [19] WANG H, JIANG X, REN H, et al. SwiftNet: Real-time video object segmentation [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 1296-1305.
- [20] POUDEL R P K, LIWICHI S, CIPOLLA R. Fast-SCNN: Fast semantic segmentation network [J]. arXiv preprint arXiv:1902.04502, 2019.
- [21] 陶孟卫, 姚宇威, 元海文, 等. 无人机自主降落视觉标识设计及位姿测量方法 [J]. 仪器仪表学报, 2022, 43(5): 155-164.
- TAO M W, YAO Y W, YUAN H W, et al. Visual target design and pose measurement method for UAV autonomous landing [J]. Chinese Journal of Scientific Instrument, 2022, 43(5): 155-164.

作者简介



樊博, 2020 年于西华大学获得学士学位, 现为上海工程技术大学硕士研究生, 主要研究方向为图像处理。

E-mail: fanbo0918@foxmail.com

Fan Bo received his B. Sc. degree from Xihua University in 2020. Now he is a M. Sc. candidate in Shanghai University of Engineering Science. His main research interest includes image processing.



高玮玮 (通信作者), 2014 年于南京航空航天大学获得博士学位, 现为上海工程技术大学讲师, 主要研究方向为机器视觉、图像处理、模式识别。

E-mail: gww03020234@sina.com

Gao Weiwei (Corresponding author) received her Ph. D. degree from Nanjing University of Aeronautics and Astronautics in 2014. Now she is a lecturer in Shanghai University of Engineering Science. Her main research interests include machine vision, image processing and pattern recognition.



单明陶, 2018 年于太原工业学院获得学士学位, 2022 年于上海工程技术大学获得硕士学位, 主要研究方向为计算机视觉、数字图像处理、深度学习。

E-mail: fmapleaves@163.com

Shan Mingtao received his B. Sc. degree from Taiyuan Institute of Technology in 2018, M. Sc. degree from Shanghai University of Engineering Science in 2022. His main research interests include computer vision, digital image processing and deep learning.



方宇, 2006 年于哈尔滨工业大学获得博士学位, 现为上海工程技术大学教授, 主要研究方向为机器视觉与图像检测、特种机器人及智能装备。

E-mail: fangyu_hit@126.com

Fang Yu received his Ph. D. degree from Harbin Institute of Technology in 2006. Now he is a professor in Shanghai University of Engineering Science. His main research interests include machine vision, image detection, special robots and intelligent equipment.