

基于红外热成像的夜间农田实时语义分割

易 诗, 李俊杰, 贾 勇

(成都理工大学信息科学与技术学院(网络安全学院、牛津布鲁克斯学院), 成都 610059)

摘 要: 农田环境实时语义分割是构成智能农机的视觉环境感知的重要环节, 夜间农田语义分割可以使智能农机在夜间通过视觉感知农田环境进行全天候作业, 而夜间无光环境下, 可见光摄像头成像效果较差, 将造成语义分割精度的下降。为保证夜间农田环境下红外图像语义分割的精度与实时性, 该研究提出了一种适用于红外图像的红外实时双边语义分割网络(Infrared Real-time Bilateral Semantic Segmentation Network, IR-BiSeNet), 根据红外图像分辨率低, 细节模糊的特点该网络在实时双边语义分割网络(Bilateral Semantic Segmentation Net, BiSeNet)结构基础上进行改进, 在其空间路径上, 进一步融合红外图像低层特征, 在该网络构架中的注意力提升模块、特征融合模块上使用全局最大池化层替换全局平均池化层以保留红外图像纹理细节信息。为验证提出方法的有效性, 通过在夜间使用红外热成像采集的农田数据集上进行试验, 数据集分割目标包括田地、行人、植物、障碍物、背景。经试验验证, 提出方法在夜间农田红外数据集上达到了85.1%的平均交并比(Mean Intersection over Union, MIoU), 同时达到40帧/s的处理速度, 满足对夜间农田的实时语义分割。

关键词: 智能农机; 语义分割; 红外热成像; 红外实时双边语义分割网络; 夜间农田数据集

doi: 10.11975/j.issn.1002-6819.2020.18.021

中图分类号: TN919.5

文献标志码: A

文章编号: 1002-6819(2020)-18-0174-07

易诗, 李俊杰, 贾勇. 基于红外热成像的夜间农田实时语义分割[J]. 农业工程学报, 2020, 36(18): 174-180. doi: 10.11975/j.issn.1002-6819.2020.18.021 <http://www.tcsae.org>
Yi Shi, Li Junjie, Jia Yong. Real-time semantic segmentation of farmland at night using infrared thermal imaging[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(18): 174-180. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2020.18.021 <http://www.tcsae.org>

0 引 言

对于智能农机的视觉导航与环境感知, 语义分割能够起到通过视觉图像理解农机周围环境的作用。然而在夜间或雨雾烟尘等成像条件恶劣的情况下, 单一使用可见光图像进行语义分割将导致智能农机视觉感知能力下降。红外热成像系统成像原理为物体的温差, 其不依赖光源, 受天气影响小, 探测距离远^[1]。对夜间无光, 存在雨雾等环境下成像更为清晰, 稳定。使用红外热成像图像进行农田环境实时语义分割可以协助智能农机在可见光成像条件不理想情况下感知所处的农田环境, 使其在特定季节, 特定环境进行全天候作业, 提高农业生产效率与智能化程度。

目前实时语义分割技术主要用于自动驾驶领域, 可以帮助自动驾驶汽车通过可见光视觉传感器分割车道、车辆、人和其他感兴趣目标或区域, 以理解车辆所处环境。近年来应用于自动驾驶的语义分割方法在城市街道环境的可见光图像上取得了良好的效果。Yang等^[2]提出通过稠密金字塔特征提取结构(Dense Atrous Spatial Pyramid Pooling, DenseASPP)改进语义分割精度的语义

分割网络, 通过金字塔特征提取结构优化了城市街景的语义分割精度, Chen等^[3]提出通过编解码网络结构(Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation)提升语义分割精度网络, 通过编解码结构进一步提升了自动驾驶环境下城市街景的语义分割精度, Wu等^[4]提出的上下文栅格语义分割网络(Context Grided Network, CGNet)使用轻量化构架与上下文联合特征提高了街景语义分割的实时性, Li等^[5]提出的基于轻量级主干网络, 多尺度特征提取的实时性语义分割网络(Deep Feature Aggregation for real-time semantic segmentation Net, DFANet)分割城市道路街景, 在各个公开城市道路街景数据集上取得了语义分割精度与实时性的较好均衡。Yu等^[6]提出的双边语义分割网络(Bilateral Semantic segmentation Net, BiSeNet), 该网络采用空间路径与上下文路径双边结构, 兼顾了语义分割精度与实时性, 同时网络结构简单利于后期优化。

自动驾驶实时语义分割需要考虑夜间无光环境与雨雾等恶劣气候条件下的成像系统, 因此, 运用红外热成像图像进行夜间环境自动驾驶语义分割的研究工作近年来取得一定进展。王晨等^[7]以深度卷积神经网络为基础, 结合条件随机场后处理优化模型, 搭建端到端的红外语义分割算法框架并进行训练, 对夜间街景红外图像进行语义分割能实现图像的像素级分类, 并获得较高的预测精度。从而获得红外图像中景物的形状、种类、位置分布等信息, 实现红外场景的语义理解。吴骏逸等^[8]针对夜

收稿日期: 2020-04-27 修订日期: 2020-07-29

基金项目: 国家自然科学基金项目(61771096); 国家大学生创新创业项目(201910616129)

作者简介: 易诗, 高级实验师, 主要从事人工智能, 红外图像处理研究。

Email: 549745481@qq.com

间道路场景解析困难的问题,提出了一种联合可见光与红外热像图实现夜间场景语义分割的方法。对双谱图像进行自适应直方图均衡及双边滤波,并利用基于双谱图像信息的稠密条件随机场对语义分割结果进行优化,对夜间道路场景进行更准确的解析。

相对城市道路的自动驾驶语义分割技术,智能农机视觉感知中的语义分割技术发展较晚,且农田环境地形更为复杂,目前运用于农田环境中的实时语义分割处于起步阶段,当前研究工作较少。李云伍等^[9]研制了一种在丘陵山区田间道路上自主行驶的转运车及其视觉导航系统。在其视觉导航模块中,针对丘陵山区复杂的田间道路场景,提出了基于改进空洞卷积神经网络的田间道路场景图像语义分割模型,以有效地对丘陵地区田间地形进行分割。尚未出现使用红外热成像图像的夜间农田环境实时语义分割技术方面研究,而使用红外图像对夜间农田复杂环境进行实时语义分割,将提高夜间智能农机的视觉感知能力与环境理解能力,有助于智能农机全天候智能化作业。使用红外热成像进行夜间田地环境实时语义分割主要涉及以下 4 个问题:1) 农田环境的复杂性。2) 红外热成像图像分辨率低,细节特征模糊,缺乏色彩特征。3) 语义分割网络模型计算量大,难以达到实时性要求。4) 缺乏公开夜间农田环境红外数据集。为解决以上问题,本文使用户外热像仪采集制作适用于夜间农田环境视觉导航的红外数据集,结合红外热成像夜间对田地环境的成像与基于深度学习的实时语义分割提出了一种夜间农田实时语义分割方法。针对红外图像分辨率低,细节特征模糊,缺乏色彩特征特点^[10]以及智能农机视觉导航的实时性要求,在双边语义分割网络结构基础上改进提出适用于红外图像的红外实时双边语义分割网络(Infrared Real-time Bilateral Semantic segmentation Net, IR-BiSeNet)。该网络根据红外图像特点进行改进,改进方法包括:1) 根据参考文献[11]中提到的全局最大池化层相对全局平均池化层更能有效保留图像纹理信息,因此在网络的注意力提升模块,特征融合模块上使用全局最大池化层替换全局平均池化层以保留红外图像纹理细节信息。2) 在获取空间信息路径上进一步融合红外图像低层特征。

1 材料与方法

1.1 数据获取

为进行夜间农田实时语义分割,需要大量夜间环境下农田红外热成像图像建立数据集,而目前无公开的夜间农田环境红外热成像数据集,本文使用与智能农机相类似的采集平台,即搭载于车辆前端的云台+红外热成像仪进行夜间环境农田的红外数据采集。通过该采集平台获取夜间农田红外数据集包括各种夜间复杂农田环境的红外图像。为采集建立适用于智能农机夜间视觉导航中语义分割的红外数据集,使用该平台在 2019 年秋季夜间野外进行实地采集,采集地点为中国四川省农业大县梓潼县,盐亭县,剑阁县以上区域的地形主要为丘陵山地农田,具有中国西南地区丘陵地形农田环境地形

起伏,田间障碍物较多,植被种类较繁杂的典型性。

实地采集的农田环境红外数据集包括 60 余个夜间农田典型场景下的 9 000 张实拍红外热成像图像。为降低对计算机显存的需求,将采集到的图像像素全部缩放为 512 像素×512 像素。

1.2 数据处理

为保证网络模型训练的鲁棒性与泛化性,需通过数据集增广方法获取更大量的数据集。红外数据集的增广根据文献[12]方法,将采集数据集进行水平、垂直方向上的反转以及 38°~56°的随机旋转处理,数据集增广方法如图 1 所示,采用这 3 种方式,每种方式可扩充 2 倍的数据量,对每个图像样本随机采用上述一种增广方式,由此将数据扩充为原始数据集的 2 倍。按 7.5:1.5:1 比例划分训练集、验证集与测试集。其中验证集与测试集使用原始图像。

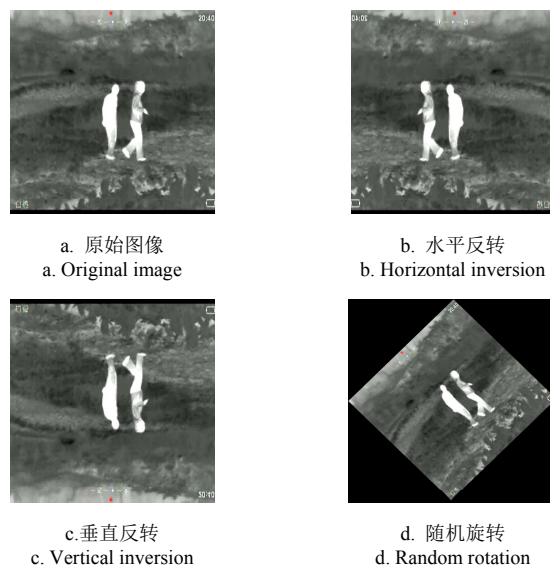


图 1 红外数据集增广方法

Fig.1 Infrared dataset augmentation method

1.3 数据标注

根据智能农机视觉导航中主要需要分割目标,将数据集中夜间田间对象的分割类别分为:行人、植被、障碍物、田地、背景 5 个类别。在图像缩放后使用 Labelme 工具^[13]进行手工标注。语义分割任务在标注分割类别时,要求对需要分割的每个类别赋予不同颜色掩膜标签(Mask),因此每种类别掩膜标签对应不同标注色的 RGB 值对应的标注色如表 1 所示,类别标注图像如图 2 所示。

表 1 各类别对应标注色

Table 1 Label color of each category

编号 No.	类别 Category	红, 绿, 蓝通道值 RGB channel value		
		R	G	B
0	背景	0	0	0
1	行人	0	0	128
2	植被	139	117	0
3	障碍物	160	32	240
4	田地	0	139	0

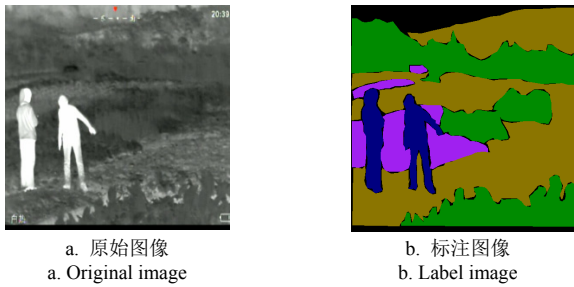


图2 类别标注图像

Fig.2 Category label image

1.4 农田红外图像语义分割方法

本研究提出适用于红外图像的红外实时双边语义分割网络 (IR-BiSeNet)，该网络根据红外图像分辨率低，细节模糊的特点在实时双边语义分割网络 (BiSeNet) 结构基础上进行改进，使其适用于夜间农田环境下红外图像的语义分割。

1.4.1 实时双边语义分割网络结构

实时双边语义分割网络 (BiSeNet) 结构如图 3 所示，为保证不牺牲空间信息的前提下实现快速的实时分割，实时双边语义分割网络 (BiSeNet) 结构划分为 2 个分支：空间路径 (Spatial Path, SP) 与上下文路径 (Context Path, CP)。

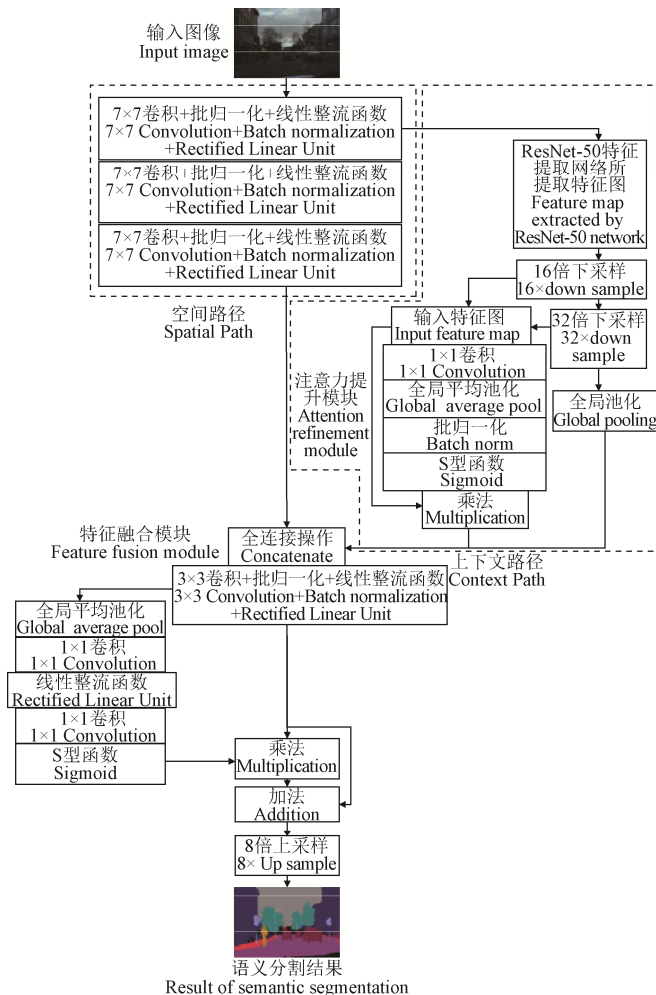


图3 BiSeNet 网络结构

Fig.3 Structure of BiSeNet

SP 结构用于提取高分辨率特征图，获取精确的空间信息。CP 结构用于获得大的感受野，为保证实时性，减少计算量，采用轻量特征提取网络，如浅层残差网络 (Residual Network-50, ResNet-50)，结合全局池化操作，合并 ResNet-50 的中间结果 (16 倍下采样、32 倍下采样) 与全局池化的输出，作为该部分输出。

BiSeNet 在上下文路径中使用注意力提升模块 (Attention refinement model, Arm) 以提高精度，注意力提升模型结构如图 3 所示，其首先使用全局平均池化获取注意力向量，对该向量进行 1×1 卷积操作、归一化和非线性操作，并将该输出与原特征图相乘。对 ResNet-50 不同层特征进行提升，且并未增加过多参数与计算量。

BiSeNet 使用特征融合模块 (Feature fusion module, Ffm) 以结合 SP 与 CP 输出的特征。特征融合模块结构如图 3 所示，该模块首先将 2 个路径输出进行连接，并将连接的特征进行 3×3 卷积、批归一化、线性整流操作，将该输出全局平均池化获取特征向量，经过 1×1 卷积、线性整流函数、S 型函数。计算连接特征，如式 (1) 所示

$$X = Y + S \cdot Y \quad (1)$$

式中 Y 为 3×3 卷积+批归一化+线性整流输出， S 为 S 型函数非线性输出，得到 FFM 的输出特征 X 。

1.4.2 红外实时双边语义分割网络结构

在实时双边语义分割网络结构基础上，本文针对夜间农田红外图像语义分割提出红外实时双边语义分割网络结构 (IR-BiSeNet)。

在实时双边语义分割网络空间路径上，根据参考文献 [14] 处理红外图像的思路，提出了一种通过增加池化层，反卷积层，全连接层的方式融合红外图像低层特征，从而使网络能更好恢复红外图像空间分辨率，提供更大的感受视野。红外实时双边语义分割网络结构如图 4 所示。对实时双边语义分割网络上下文路径中的注意力提升模型以及融合两路特征的特征融合模块，根据红外图像细节模糊、对比度低的特点，在模块网络构架的每个池化层部分使用全局最大池化层替换全局平均池化层以保留红外图像纹理细节信息。

考虑到红外图像特点以及系统实时性，上下文路径采用 ResNet-50 特征提取网络，结合全局池化操作，合并 ResNet-50 中间结果 (16 倍下采样、32 倍下采样) 与全局池化的输出。

红外实时双边语义分割网络在上下文路径中为提高精度所使用的红外注意力提升模块 (IR-Attention refinement model, IR-Arm) 如图 4 所示。

红外实时双边语义分割网络融合空间路径与上下文路径输出的特征所使用的红外特征融合模块 (IR-Feature fusion module, IR-Ffm) 如图 4 所示。

1.4.3 红外实时双边语义分割网络损失函数

本文提出的红外实时双边语义分割网络使用主损失函数监督整个网络的输出，此外采用附加损失函数监督上下文路径输出 [15-17]。

主损失函数 L_s 采用 Softmax 损失函数, 如式 (2) 所示

$$L_s = -\frac{1}{N} \sum_i \log_2 \left(\frac{e^{p_i}}{\sum_j e^{p_j}} \right) \quad (2)$$

式中 N 为需要预测的类别总数, p_i 为样本 i 的网络预测输出, p_j 为样本 j 的网络预测输出。

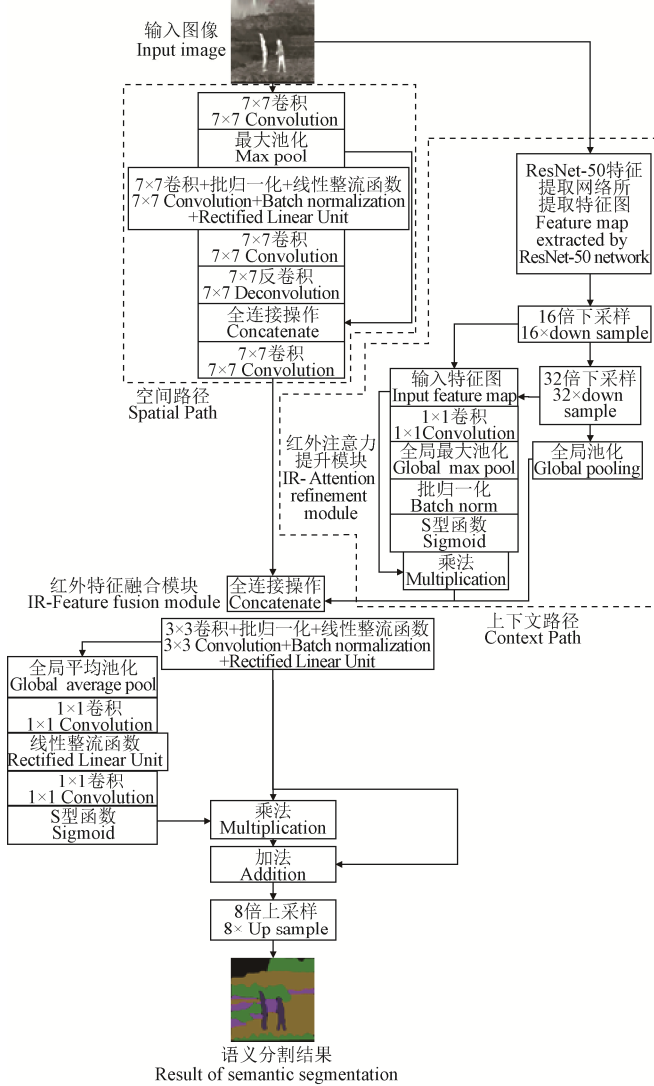


图 4 红外实时双边语义分割网络结构

Fig.4 Structure of infrared real-time bilateral semantic segmentation net

监督上下文路径输出的辅助损失函数, 如式 (3) 所示

$$L = l_p + \sum_{i=2}^3 l_i(X_i) \quad (3)$$

式中 l_p 是全连接层的输出损失函数, l_i 为第 i 阶段的辅助损失函数其输入为 X_i , X_i 为 ResNet-50 网络第 i 阶段特征值。 L 为联合损失函数^[18-21]。

1.5 性能评价指标

为验证本文提出方法对夜间农田红外图像分割的有效性, 设图像中分割类别为 $k+1$, p_{ij} 为本属于 i 类但被预测为 j 类的像素数量, p_{ji} 为本属于 j 类但被预测为 i 类的像素数量, p_{ii} 为真实像素数量, 采用评价实时语义分割的指标如下:

1) 平均交并比 (Mean Intersection over Union, MIoU), 即预测区域和实际区域交集除以预测区域和实际区域的并集^[22-23], 计算如式 (4) 所示

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (4)$$

2) 像素精度 (Pixel Accuracy, PA), 标注正确的像素占总像素的比例^[24], 计算如式 (5)

$$PA = \frac{\sum_{i=0}^k p_{ii}}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}} \quad (5)$$

3) 平均像素精度 (Mean Pixel Accuracy, MPA), 所有类的平均像素精度^[24-26], 计算如式 (6)

$$MPA = \frac{1}{k+1} \frac{\sum_{i=0}^k p_{ii}}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}} \quad (6)$$

4) 帧率 (Frames Per Second, FPS), 衡量语义分割算法实时性。计算如式 (7)

$$FPS = \frac{N_f}{T} \quad (7)$$

式中 N_f 为视频帧数, T 为消耗时间, s。

2 结果与分析

2.1 模型训练

本文提出的红外实时双边语义分割网络 (IR BiSeNet) 在 Win10+tensorflow1.9.0+CUDA9.2+VS2017+opencv4.0/Core i7-8750H 2.2Ghz 处理器+16GB 内存+Geforce GTX 1080 8GB 显卡软/硬件平台上使用相同夜间农田语义分割红外数据集进行训练, 采用 ResNet-50 预训练模型初始化部分参数, RMSprop 优化器优化所有参数, 设置学习率为 0.000 1, 衰减率为 0.995。随机初始化深度网络各层参数。训练批次设置为 200, 一次训练所选取的样本数设置为 5^[27-30], 在 120 批次 (120 epochs) 训练后生成最佳语义分割模型。

2.2 方法对比

为验证本文提出方法对夜间农田红外图像分割的优势, 将本文提出方法与现行 5 种可见光环境下代表性语义分割框架 (BiSeNet、DenseASPP、DeeplabV3+、DFANet、CGNet) 在同一数据集下, 采用相同的训练参数进行训练, 所生成的模型进行对比测试。

对于每一类需要分割的目标交并比 (IoU), IR-BiSeNet 相对其他 5 种方法在测试集上测试结果统计如表 2 所示。

由每一类别需分割目标交并比测试结果可知, IR-BiSeNet 相对其他 5 种方法, 对 5 类夜间农田环境需分割的目标上均取得最高的交并比, 对背景、行人、植被、障碍物、农田的语义分割分别达到 75.3%、88.6%、.3%、86.2%、85.6% 的交并比。因此, 本文提出的红外实时双边语义分割网络对夜间农田红外图像中各类目标的语义分割精度上具有优势。

综合测试环节, 使用测试集中实拍的夜间农田红外图像进行对比测试, 从像素精度 (PA)、平均像素精

度 (MPA)、平均交并比 (MIoU)、帧率 (FPS) 进行对比。本文提出方法相对其他 5 种方法的对比测试结果如表 3 所示。

表 2 不同方法下不同类别的目标交并比 (IoU) 对比

Table 2 Comparison of Intersection over Union (IoU) of different categories under different methods %

语义分割方法 Semantic segmentation method	背景 Background	行人 Pedestrian	植被 Weed	障碍物 Obstacle	田地 Field
CGNet	15.6	55.2	40.6	45.3	54.2
DFANet	63.5	85.5	77.2	74.5	83.2
DeeplabV3+	65.0	86.0	78.8	77.5	84.5
DenseASPP	63.0	85.2	77.9	76.8	82.8
BiSeNet	66.1	80.3	75.1	75.8	83.2
IR-BiSeNet	75.3	88.6	85.3	86.2	85.6

注: CGNet 为上下文栅格语义分割网络; DFANet 为多尺度特征提取的实时性语义分割网络; DeeplabV3+ 为编解码金字塔结构网络语义分割网络; DenseASPP 为稠密金字塔特征语义分割网络; BiSeNet 为双边语义分割网络; IR-BiSeNet 为红外实时双边语义分割网络。

Note: CGNet: Context Grided Network; DFANet: Deep Feature Aggregation for real-time semantic segmentation Net; DeeplabV3+: Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation; DenseASPP: Dense Atrous Spatial Pyramid Pooling; BiSeNet: Bilateral Semantic Segmentation Net; IR-BiSeNet: Infrared Real-time Bilateral Semantic Segmentation Network.

由表 3 的对比测试指标分析可以看出, 在实时性方面, 本文提出的 IR-BiSeNet 网络与 DenseASPP 网络和 DeeplabV3+ 网络相比, 帧率 (FPS) 分别提高 4 和 28 帧/s。由于加深了网络深度, 低于改进前 BiSeNet 网络 2.5 帧/s, 同时低于对实时性特别优化的 DFANet 网络, CGNet 网络

28 和 13 帧/s, 文献[4-6]中提出, 帧率达 25 帧/s 以上即可认定算法具备实时性, 因此本文提出方法可满足语义分割的实时性要求。但在夜间农田语义分割精度方面, 本文提出的 IR-BiSeNet 网络在像素精度 (PA), 平均像素精度 (MPA), 平均交并比 (MIoU) 3 项指标上与其他 5 种语义分割方法对比具有明显优势, 例如与 BiSeNet 网络相比, 本文提出方法在像素精度上提高了 8.4 个百分点, 在平均像素精度上提高了 10.3 个百分点, 在平均交并比上提高了 9.8 个百分点。

表 3 不同方法性能指标对比

Table 3 Comparison of evaluation metrics under different methods

语义分割方法 Semantic segmentation method	像素精度 Pixel accuracy/%	平均像素精度 Mean pixel accuracy/%	平均交并比 Mean intersection over union/%	帧率 Frames per second/(帧·s ⁻¹)
CGNet	60.5	50.2	58.5	53.0
DFANet	77.3	72.5	75.1	68.0
DeeplabV3+	78.5	74.5	76.2	12.0
DenseASPP	76.2	71.5	74.1	36.0
BiSeNet	78.1	73.2	75.3	42.5
IR-BiSeNet	86.5	83.5	85.1	40.0

IR-BiSeNet 保留红外图像纹理特征方面做了优化, 能够保留更多夜间农田红外图像语义分割细节。与其他 5 种方法对测试集中夜间农田环境红外图像分割结果对比如图 5 所示, 由图 5 可以看出对于夜间农田环境的各类分割目标细节分割效果, IR-BiSeNet 分割结果相对其他方法分割效果更好, 更接近于真实标注图像。

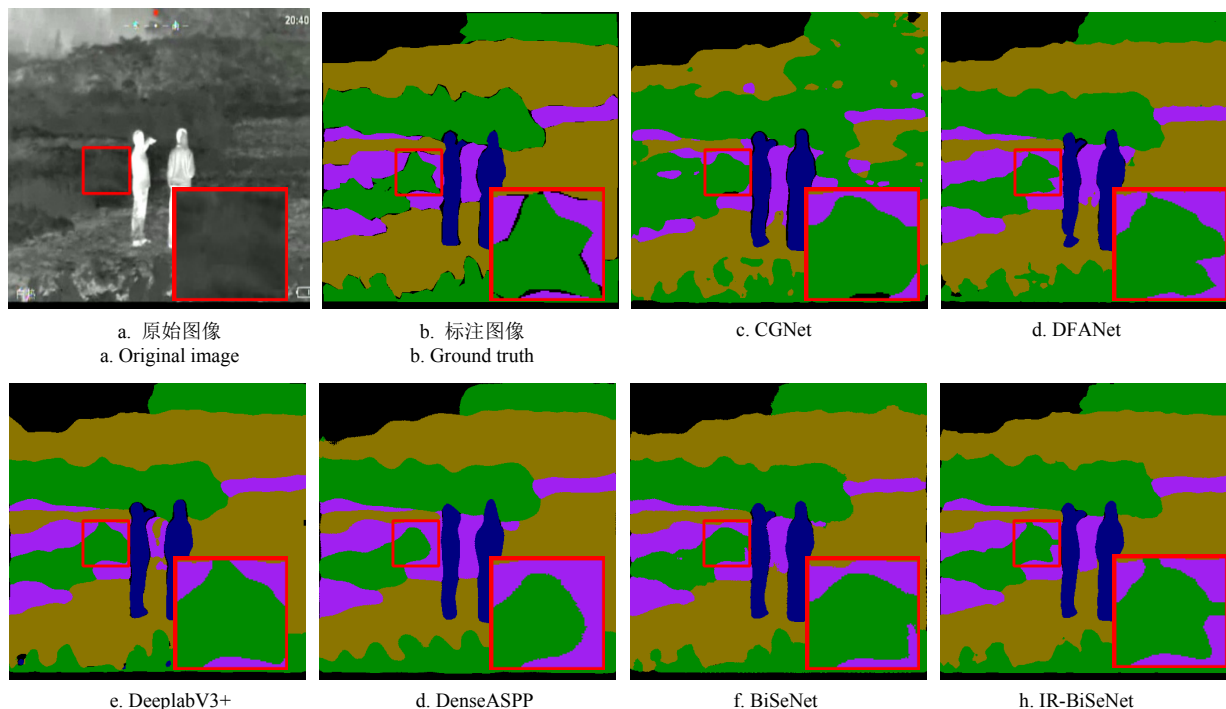


图 5 不同语义分割方法效果对比

Fig.5 Comparison of the effects of different semantic segmentation methods

3 结 论

针对智能农机夜间视觉导航与视觉感知, 本文提出了一种基于红外热成像的夜间农田实时语义分割网络 (IR-BiSeNet), 该网络在实时双边语义分割网络

(BiSeNet) 结构基础上, 在其空间信息路径上进一步融合红外图像低层特征, 在其注意提升模型、特征融合模块上使用全局最大池化层替换全局平均池化层。采集制作了夜间农田红外数据集, 在该数据集上, 本文方法达到了 85.1% 的平均交并比 (MIoU), 同时达到 40 帧/s 的处

理速度, 满足对夜间农田的实时语义分割。在未来工作中, 将采用更高效的语义分割网络针对红外图像特点进行改进, 同时扩充夜间农田红外数据集, 以达到更好的实时语义分割效果。

[参 考 文 献]

- [1] 崔美玉. 论红外热像仪的应用领域及技术特点[J]. 中国安防, 2014(12): 90-93.
Cui Meiyu. On the application field and technical characteristics of infrared thermal imager[J]. China Security, 2014(12): 90-93. (in Chinese with English abstract)
- [2] Yang Maoke, Yu Kun, Chi Zhang, et al. DenseASPP for semantic segmentation in street scenes[C]. IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 3684-3692.
- [3] Chen L C, Papandreou G, Schro F, et al. Rethinking atrous convolution for semantic image segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018.
- [4] Wu Tianyi, Tang Sheng. CGNet: A Light-weight context guided network for semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018.
- [5] Li Hanchao, Xiong Pengfei, Fan Haoqiang, et al. DFANet: Deep feature aggregation for real-time semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019.
- [6] Yu Changqian, Wang Jingbo, Peng Chao, et al. BiSeNet: bilateral segmentation network for real-time semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018.
- [7] 王晨, 汤心溢. 基于深度卷积神经网络的红外场景理解算法[J]. 红外技术, 2019, 39(8): 728-733.
Wang Chen, Tang Xinyi. Infrared scene understanding algorithm based on deep convolutional neural network[J]. Infrared Technology, 2019, 39(8): 728-733. (in Chinese with English abstract)
- [8] 吴骏逸, 谷小婧, 顾幸生. 基于可见光/红外图像的夜间道路场景语义分割[J]. 华东理工大学学报: 自然科学版, 2019, 45(2): 301-309.
Wu Junyi, Gu Xiaojing, Gu Xingsheng. Night road scene semantic segmentation based on visible and infrared thermal images[J]. Journal of East China University of Science and Technology: Natural Science Edition, 2019, 45(2): 301-309. (in Chinese with English abstract)
- [9] 李云伍, 徐俊杰, 刘得雄, 等. 基于改进空洞卷积神经网络的丘陵山区田间道路场景识别[J]. 农业工程学报, 2019, 35(7): 150-159.
Li Yunwu, Xu Junjie, Liu Dexiong, et al. Field road scene recognition in hilly regions based on improved dilated convolutional networks[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(7): 150-159. (in Chinese with English abstract)
- [10] 易诗, 李欣荣, 吴志娟, 等. 基于红外热成像与改进 YOLOV3 的夜间野兔监测方法[J]. 农业工程学报, 2019, 35(19): 223-229.
Yi Shi, Li Xinrong, Wu Zhijuan, et al. Night hare detection method based on infrared thermal imaging and improved YOLOV3[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(19): 223-229. (in Chinese with English abstract)
- [11] 昌杰. 基于深度神经网络的肿瘤图像分析与处理[D]. 合肥: 中国科学技术大学, 2019.
Chang Jie. Tumor Image Analysis and Processing Based on Deep Neural Network[D]. Hefei: University of Science and Technology of China, 2019. (in Chinese with English abstract)
- [12] Li Xiangyuan, Cheng Cai, Zhang Ruifei, et al. Deep cascaded convolutional models for cattle pose estimation[J]. Computers and Electronics in Agriculture. 2019, 164: 104885.
- [13] Russell B C, Torralba A, Murphy K P, et al. LabelMe: A database and web-based tool for image annotation[J]. International Journal of Computer Vision, 2008, 77(1-3): 157-173.
- [14] He Zewei, Cao Yanpeng, Dong Yafei et al. Single-image-based nonuniformity correction of uncooled long-wave infrared detectors: A deep-learning approach[J]. Applied Optics. 2018, 57(18):155-164.
- [15] Amorim W P, Tetila E C, Pistori H, et al. Semi-supervised learning with convolutional neural networks for UAV images automatic recognition[J]. Computers and Electronics in Agriculture. 2019, 164: 104932.
- [16] Barth R, Ijsselmuiden J, Hemming J, et al. Synthetic bootstrapping of convolutional neural networks for semantic plant part segmentation[J]. Computers and Electronics in Agriculture. 2019, 161: 291-304.
- [17] Axel-Christian G, Moulay A. Deep learning enhancement of infrared face images using generative adversarial networks[J]. Applied Optics, 2018, 57(18):98-105.
- [18] Satoru K, Adam M, Abhijit M, et al. Three-dimensional integral imaging and object detection using long-wave infrared imaging[J]. Applied Optics, 2017, 56(9): 120-126.
- [19] Kuang Xiaodong, Sui Xiubao, Liu Yuan, et al. Single infrared image enhancement using a deep convolutional neural network[J]. Neurocomputing, 2019, 332: 119-128.
- [20] 李云伍, 徐俊杰, 王铭枫, 等. 丘陵山区田间道路自动驾驶转运车及其视觉导航系统研制[J]. 农业工程学报, 2019, 35(1): 52-61.
Li Yunwu, Xu Junjie, Wang Mingfeng, et al. Development of autonomous driving transfer trolley on field roads and its visual navigation system for hilly areas[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(1): 52-61. (in Chinese with English abstract)
- [21] Liu Zihao, Jia Xiaojun, Xu Xinsheng. Study of shrimp recognition methods using smart networks[J]. Computers and Electronics in Agriculture, 2019, 165: 104926.
- [22] Tian Mengxiao, Guo Hao, Chen Hong, et al. Automated pig counting using deep learning[J]. Computers and Electronics in Agriculture., 2019, <https://doi.org/10.1016/j.compag.2019.05.049>.
- [23] 孙哲, 张春龙, 葛鲁镇, 等. 基于 Faster R-CNN 的田间西兰花幼苗图像检测方法[J]. 农业机械学报, 2019, 50(7): 216-221.
Sun Zhe, Zhang Chunlong, Ge Luzhen, et al. Image detection method for broccoli seedlings in the field based on Faster R-CNN[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(7): 216-221. (in Chinese with English abstract)
- [24] Zhang Shanwen, Zhang Subing, Zhang Chuanlei, et al. Cucumber leaf disease identification with global pooling dilated convolutional neural network[J]. Computers and

- Electronics in Agriculture, 2019, 162: 422-430.
- [25] Kounalakis T, Triantafyllidis G A, Nalpantidis L. Deep learning-based visual recognition of rumex for robotic precision farming[J]. Computers and Electronics in Agriculture, 2019, 165: 104973.
- [26] Kapoor A J, Fan H, Sardar M S. Intelligent detection using convolutional neural network (ID-CNN)[C]. IOP Conference Series: Earth and Environmental Science, Hubei: IOP Publishing, 2019.
- [27] Li X, Orchard M T. New edge-directed interpolation[J]. IEEE Transactions on Image Processing, 2001, 10(10): 1521-1527.
- [28] Yu Yang, Zhang Kailiang, Yang Li, et al. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN[J]. Computers and Electronics in Agriculture, 2019, 163: 104846.
- [29] Ye Xujun, Sakai K, Manago M, et al. Prediction of citrus yield from airborne hyperspectral imagery[J]. Precision Agriculture, 2007, 8(3): 111-125.
- [30] Lin Tsung Yi, Maire M, Belongie S, et al. Microsoft COCO: Common objects in context[C]. Proceedings of the 13th European Conference on Computer Vision, New York, USA: Springer, 2014: 740-755.

Real-time semantic segmentation of farmland at night using infrared thermal imaging

Yi Shi, Li Junjie, Jia Yong

(College of Information Science and Technology (College of Cyber Security, College of Oxford Brookes), Chengdu University of Technology, Chengdu 610059, China)

Abstract: In intelligent agricultural machinery, automatic navigation, and visual perception technology have been developed rapidly in recent years, and they also play a vital role in intelligent modern agriculture. Therefore, real-time semantic segmentation of farmland environment become an important part of visual environment perception in the intelligent agricultural machinery. The visible light sensing equipment is mainly used for image collection. However, particularly in the dark environment at night, the deficient imaging effect of visible light cameras can result in a decrease in the accuracy of semantic segmentation. Infrared thermal imaging can offer an alternatively way in this case, due to this technology uses the temperature difference of the object for imaging, rather than the light source. Therefore, the infrared thermal imaging can be used to clearly capture the image in the dark night, rain, mist, smoke, and other visible light sensing equipment that is not suitable. In this study, a method for real-time semantic segmentation of infrared images of farmland environment at night was proposed using the infrared thermal imaging system. An infrared real-time bilateral semantic segmentation network (IR-BiSeNet) was also addressed suitable for infrared images, in order to ensure the accuracy and real-time performance of infrared image semantic segmentation in the farmland environment at night. According to the characteristics of low resolution and fuzzy details of infrared images, the network was improved based on the BiSeNet structure, and the low-level features of infrared images were further integrated in its spatial path. In the network, the global maximum pooling layer was used to replace the global average pooling layer in the attention enhancement and the feature fusion module, in order to preserve the texture details of infrared image. The infrared farmland data was collected by the infrared thermal imaging to create a dataset at night, thereby to train a semantic segmentation model suitable for the farmland environment in this case. The segmentation targets of dataset included the fields, pedestrians, plants, obstacles, backgrounds, using the data augmentation to produce the dataset of infrared night farmland. Five representative semantic segmentation methods were selected to verify the proposed method, including BiSeNet, DenseASPP, DeeplabV3+, DFANet, and CGNet. Experimental results showed that the proposed method can achieved the mean intersection over union of 85.1%, and the processing speed of 40 frames/s. The method proposed in this study can be used the infrared thermal imaging to perform real-time farmland environment semantic segmentation at night, which can greatly improve the visual perception of intelligent agricultural machinery at night.

Keywords: intelligent agricultural machinery; semantic segmentation; infrared thermal imaging; infrared real-time bilateral semantic segmentation net; farmland dataset at night