

DOI:10.19651/j.cnki.emt.2208764

一种用于车辆图像分割的 MSSA-UNet 模型*

赵红爱^{1,2} 王旭智^{1,2} 万旺根^{1,2}

(1.上海大学通信与信息工程学院 上海 200072; 2.上海大学智慧城市研究院 上海 200072)

摘要: 针对实际交通场景下的车辆图像分割方法存在模糊、效果差的问题,本文以 UNet 神经网络模型为基础,提出了一种融合多尺度模块和空间注意力机制的 MSSA-UNet 模型。在编解码阶段,采用空洞卷积构建多尺度模块,改善卷积层感受野大小受限的同时输出包含多尺度的特征信息。在上采样前,引入空间注意力机制来弥补采样过程中的局部信息丢失问题,提高特征还原能力。结合交叉熵损失与 Dice 损失,优化网络学习和训练过程,提高模型的分割精度。实验结果表明,本文提出的 MSSA-UNet 模型对于车辆图像分割任务在 IoU 评价指标达到 83.48%,较改进前准确度提升了 2.28%,模型预测值和真实值更接近,分割效果更好,有效提升了模型的分割性能。

关键词: 空洞卷积;多尺度;空间注意力;图像分割;卷积神经网络

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 510.60

A MSSA-UNet model for vehicle image segmentation

Zhao Hong'ai^{1,2} Wang Xuzhi^{1,2} Wan Wanggen^{1,2}

(1. School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China;
2. Institute of Smart City, Shanghai University, Shanghai 200072, China)

Abstract: In view of the problems of ambiguous and poor effect of vehicle image segmentation methods in actual traffic scenarios, this paper proposes a MSSA-UNet model that integrates multi-scale modules and spatial attention mechanism based on the UNet neural network model. In the encoding and decoding stage, dilated convolution is used to build a multi-scale module to improve the limited receptive field size of the convolutional layer while the output contains multi-scale feature information. Before up-sampling, a spatial attention mechanism is introduced to compensate for the problem of local information loss during the sampling process and improve the feature restoration ability. Combined with cross entropy loss and Dice loss, the network learning and training process is optimized, and the segmentation accuracy of the model is improved. The experimental results show that the MSSA-UNet model proposed in this paper achieves 83.48% in the IoU evaluation index for vehicle image segmentation tasks, which is 2.28% higher than the accuracy before improvement, the predicted value of the model is closer to the real value, and the segmentation effect is better, which effectively improves the segmentation performance of the model.

Keywords: dilate convolution; multi-scale; spatial attention; image segmentation; convolutional neural network

0 引言

随着经济社会快速发展,车辆数量迅猛增加使城市交通环境日益复杂,建立智能交通系统^[1]对于实现交通管理智能化和维护公共安全有着十分重要的意义。车辆检测与分割作为车辆识别、跟踪的重要基础,是智能交通系统中非常关键的技术。图像分割在医疗检查、机器人以及自动驾驶等领域有着广阔的应用前景,逐渐成为计算机视觉领域的热点研究方向。相比于目标检测^[2],图像分割实现难度

更大,是一项更为细致的感兴趣区域挖掘任务。

车辆分割任务是指将图像分为背景和车辆子区域的过程。目前,主流的图像分割模型多是基于深度学习的方法创建的^[3]。与传统的像素或超像素分类^[4]、几何活动轮廓^[5]等方法使用手工制作特征不同,深度学习方法能够自动从图像中学习目标复杂的特征,并通过端到端的方式实现分割任务。Long 等^[6]去除了卷积神经网络中的全连接层,提出了全卷积神经网络 (fully convolution network, FCN)。FCN 通过卷积和池化操作对输入图像特征进行提

收稿日期:2022-01-04

* 基金项目:安徽省自然科学基金(1908085MF178)、安徽省重点研究和开发计划项目(202104b11020031)、中国博士后基金(2020M681264)项目资助

取,在特征还原阶段采用反卷积进行上采样,得到了较好的分割结果。FCN模型的输入和输出均为图像,具有输入原始图像并端到端输出分割结果的优点。但是由于其图像还原过程简单,容易导致分割结果不够细致、边缘模糊等问题。Ronneberger等^[7]基于FCN提出了U型神经网络UNet,其左右对称的编解码(encoder-decoder)结构用来获取图像的上下文信息和预测位置信息,通过跳跃连接(skip connection)充分结合了浅层空间特征和深层语义特征,加强了解码器与编码器之间的联系,使网络对细节特征更加敏感,获得了很好的像素级定位性能。此后,众多研究者将UNet网络结构应用于不同领域的分割任务,并提出了各种改进方案。TernausNet^[8]采用VGG11^[9]作为UNet的编码器,通过微调(fine-tuning)技术初始化网络编码器权重,轻松提升了模型性能。同时,预训练网络不仅缩短了训练时间,还可以有效防止过拟合问题。Chen等相继提出了DeepLabv1^[10]、DeepLabv2^[11]以及DeepLabv3^[12]模型,这些方法的主要贡献可总结为:利用空洞卷积^[13]有效扩大卷积的感受野以获得更多的上下文信息;提出空洞空间金字塔池化(atrous spatial pyramid pooling,ASPP)在不同尺度上鲁棒地分割图像;结合深度卷积神经网络和概率图模型改进样本边界的定位问题。这些贡献在大量实验上被证明具有实用价值,进一步推进了图像分割在深度学习领域的发展。在车辆图像分割任务上,高凯珺等^[14]提出了一种无人车夜视图像语义分割方法,采用卷积-反卷积网络完成了像素到像素的学习和训练,得到分割模型。张乐等^[15]提出了一种基于全卷积网络对图像中的车辆进行分割的方法,减少部分卷积层以获得更精细的边缘分类结果,并将车辆目标的分割问题转化为基于像素的二分类问题,削减了其他类别目标的干扰。

现有的车辆图像分割算法大多以目标车辆的全局特征作为分割对象,即从图像中分割出整个车辆模型。然而,由路网监控摄像头采集的车辆图像往往因摄像头角度不同、车辆姿态变化而产生较大的视角差异性,全局特征不足以表达样本图像间的判别信息。局部特征通常包含了区别身份不同但外观相似车辆的显著性信息,合理提取局部特征能够有效提升图像分类、再识别等任务的准确性。本文围绕实际交通场景下的车辆图像局部区域分割任务,针对现有分割方法存在区域边界模糊、分割效果差的问题,从多尺度特征提取和注意力信息增强两个方面对UNet进行改进。与主流的车辆图像分割算法进行对比,证明了所提算法有效提高分割效果,且模型的预测值和真实值更接近。

1 研究方法

本文所提方法的整体架构主要参考了UNet模型,该模型是图像分割领域最常用的分割架构之一。UNet框架主要由编码器、解码器和跳跃连接设计组成。其中,编码器由一系列的卷积层和池化层构成,通过线性加权计算将图像特征逐渐抽象化;解码器将抽取到的高维特征映射到像素空间;跳跃连接的目的是传递对应层次的特征图并在通道维度上进行拼接,使解码器共享了浅层特征。

本文在UNet网络架构的基础上,采用多尺度模块代替标准卷积序列并合理引入空间注意力机制提出了融合多尺度模块和空间注意力机制的UNet(Multi-Scale Spatial Attention UNet,MSSA-UNet)模型,在很大程度上增强了网络的特征表达和细粒度感知能力。通过联合损失函数优化网络的学习和训练过程,提高了模型的分割精度。MSSA-UNet模型的网络架构如图1所示。

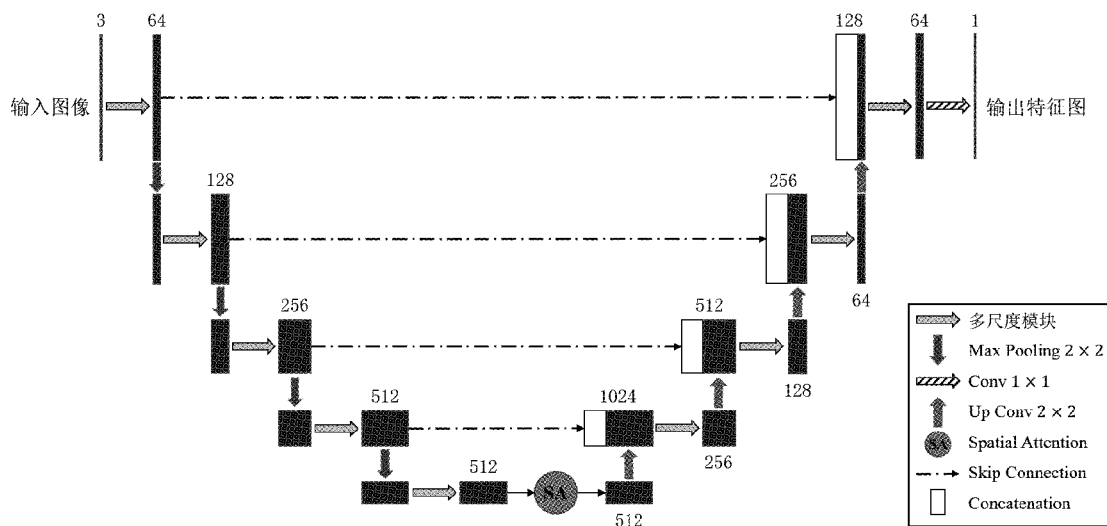


图1 融合多尺度模块和空间注意力机制的UNet模型结构示意图

1.1 MSSA-UNet 框架

MSSA-UNet模型沿用U型对称结构设计,由一个收

缩路径(左边,也称编码器)和一个扩张路径(右边,也称解码器)组成。收缩路径主要采用多尺度模块、下采样模块

和空间注意力机制构建。其中,多尺度模块的卷积序列包含标准卷积核和空洞卷积核,并且每个卷积层后都加入了批归一化层(batch normalization, BN)和修正线性单元(rectified linear unit, ReLU)激活层。输入特征(张量)进入编码阶段,首先通过一个多尺度模块,采用不同感受野的卷积提取特征,该操作不改变特征尺寸大小但将通道数增大。接下来通过第一个 2 倍下采样模块,采用一个 2×2 的最大池化得到尺寸大小为输入特征 $1/2$ 的输出特征。经过上述 4 次操作,得到的特征分辨率降低了 16 倍,且通道数不断增加。随后,将压缩后输出特征经过空间域注意力机制,让网络自动分配权重以获取更加有效的特征信息。扩张路径由多尺度模块和上采样模块组成,并与收缩路径在结构上形成左右对称。当编码器输出特征进入解码器,首先通过上采样模块采用反卷积操作扩大特征尺寸,并通过跳跃连接结构与收缩路径每一层的特征图进行融合,充分结合浅层空间信息和深层语义信息。多尺度模块进一步对融合特征进行卷积计算,尽可能地恢复特征图的细节信息。最后,使用 1 个输出通道数为目标类别数的 1×1 卷积实现像素类别预测,并输出分割结果。

多尺度模块的加入使得网络中嵌入了多个不同感受野的卷积块,可以对提取的特征进行多样化处理,对于需要详细空间信息的密集预测任务很有帮助。因此,多尺度模块不仅可以提高分割性能,而且在训练过程中更有利于网络的学习。空间注意力机制加在了最后一层编码器输出和第一层解码器输入前,加权的特征图能够使网络在上采样过程中更加关注与任务相关的信息,从而提升特征恢复的效果。

1.2 多尺度模块设计

尽管对称结构的设计利用了浅层特征和深层特征的融合信息,但 UNet 模型在特征提取部分并没有考虑到卷积层的感受野大小对分割性能的影响,不同大小的感受野在车辆图像的颜色、纹理上具有不同的敏感度,这使得网络提取到的特征图在语义上存在一定的差异性。本文设计的多尺度模块及其实现细节如图 2 所示。

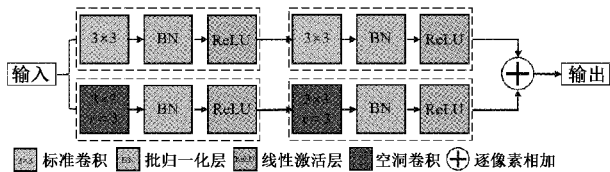


图 2 多尺度模块结构示意图

输入经过感受野不同的卷积核加权运算后融合,该融合特征作为编码器的输出被用于跳跃连接部分的特征拼接。标准卷积核和空洞卷积核拥有不同的感受野,它们得到的特征也是有差异的,这种结构使得编码器提取的特征更丰富,有助于解码器恢复特征,对于小目标区域的预测任务也更有效。

感受野是指神经元对原图像或特征图的感受范围大小,即感受野越大表示该神经元能“看到”的原始图像范围越大。多尺度模块可以轻松嵌入到网络结构中,结合了标准卷积核和感受野更大的空洞卷积核使网络能够获得多尺度的特征信息,在一定程度上增加了特征的多样性。1 个大小为 3×3 ,扩张率 $r = 3$ 的空洞卷积的感受野范围可以达到 7×7 。空洞卷积是通过调整扩张率来控制感受野范围的,相比于相同感受野大小的标准卷积具有参数量少的优点。假设空洞卷积核的大小为 $k \times k$,扩张率为 r ,则该空洞卷积的感受野 k' 的计算方式可被阐述为:

$$k' = k + (k - 1) \times (r - 1) \quad (1)$$

1.3 空间注意力机制

注意力机制是一种常被研究者们嵌入到深度学习模型中的特殊机制,是模仿人类注意力提出的一种有效解决方案,它能够根据贡献大小将原本平均分配的资源进行重新分配^[16]。在神经网络的设计中,注意力机制所要分配的资源就是权重。空间注意力机制是目前主流注意力应用的一种,被广泛应用于图像处理领域中,能够根据任务的需求而更加关注图像中的目标区域。通过对相关性高的区域施加更高的权重使任务需要的目标区域更为突出,即网络能够学习更有价值的信息,而抑制与任务无关的特征信息。空间注意力在空间维度进行变换并保留原始特征的关键信息,其结构如图 3 所示。

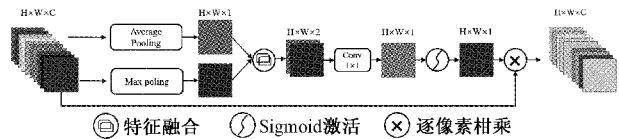


图 3 空间注意力结构示意图

给定输入特征图为 $x_{input} \in R^{H \times W \times C}$,首先通过全局平均池化(global average pooling, GAP)和全局最大池化(global max pooling, GMP)后分别得到 $x_{GAP} \in R^{H \times W \times C}$ 和 $x_{GMP} \in R^{H \times W \times C}$ 。然后以拼接的方式对 x_{GAP} 和 x_{GMP} 进行特征融合,随后需要通过 1×1 卷积 $conv^{1 \times 1}$ 将融合后的二维特征 $x_{concat} \in R^{H \times W \times 2}$ 转换为一维特征。为了得到注意力权重图,使用 Sigmoid 激活函数 σ 进行非线性变换后得到每一个特征值大小范围为 $(0, 1)$ 的注意力权重图 $\alpha \in R^{H \times W \times 1}$ 。最终利用权重图 α 对输入 x_{input} 施加权重后得到空间注意力机制的输出结果 $x_{output} \in R^{H \times W \times C}$ 。其中 C 是特征图的通道数, H 和 W 分别表示特征图的长度和宽度。空间注意力的具体实现方法如下:

$$x_{concat} = Concat(x_{GAP} + x_{GMP}) \quad (2)$$

$$x_{output} = x_{input} \cdot \sigma[conv^{1 \times 1}(x_{concat})] \quad (3)$$

1.4 损失函数设计

在图像分割领域中,常用的损失函数为交叉熵损失(cross entropy loss)函数。研究表明,交叉熵损失采用了类间竞争机制,擅于学习类间信息。但是仅仅使用交叉熵

损失容易使网络的训练被背景区域主导,导致网络难以学习车辆图像的局部区域特征,降低训练效果。因此,本文结合 Dice 损失^[17]设计了联合损失函数,实现细节如下所示:

$$Loss_{Dice} = 1 - \frac{(1 + \beta^2) \sum_i^N p_i g_i + \eta}{\beta^2 \sum_i^N p_i + \sum_i^N g_i + \eta} \quad (4)$$

式中: N 是像素的数量, p_i 代表预测值, g_i 代表真实值, η, β 为参数项,分别设置为 $\eta = 1 \times 10^{-7}, \beta = 1$ 。最终的联合损失函数可表示为:

$$Loss = Loss_{CE} + Loss_{Dice} \quad (5)$$

2 实验与结果分析

2.1 实验环境与实施细节

本文实验所使用的服务器软硬件配置如下:CPU 型号为 Intel® Xeon(R) Gold 5118 CPU@2.3 GHz, Ubuntu 16.04 操作系统, GPU 型号为 NVIDIA GeForce GTX 1080 Ti 12 GB, 深度学习框架为 Pytorch1.6.0, Python3.6.13。

在实验实施细节方面,文中涉及的模型在标注的训练集上进行 60 轮 (Epoch) 训练,训练过程中批处理大小 (batch-size) 设置为 8,验证过程中 batch-size 设置为 1,并采用均方根传递 (RMSProp) 算法优化损失函数,初始学习率设为 0.001,平滑常数 (也称衰减速率) 设为 0.9。

2.2 实验数据集

本文选用公开数据集 VeRi776^[18] 作为实验数据源, VeRi776 数据集是由北京邮电大学构建的用于车辆再识别任务的数据集,使用了 20 个道路监控摄像头采集总共 776 辆车,约 50 000 张图像。数据集中的每张车辆图像通过 2~18 个摄像头在不同视角、光照强度等情况下拍摄。在本文的研究中,车辆可视为一个刚体,由于车辆底部不可见且两侧具有较高的外观相似性,因此将车辆的局部区域定义为前、后、顶部以及侧边四类,且对不同区域进行分别标注。此外,使用 Meng 等^[19] 提供的标注信息对训练集中的 3 165 张图片进行标注,设置目标区域像素点为 [1, 2, 3, 4], 分别对应以上 4 个类别,背景像素点类别设置为 0,生成了用于训练与测试的 mask 图像。其中,对每个车辆样本选取了至少包含 4 个不同视角方向的图像,并且按照 9:1 的比例划分为训练集和测试集。针对训练集中的图像采用随机缩放、亮度增强和图像填充等数据增强方法来提高模型的泛化能力,减少过拟合现象。

2.3 评价指标

为评价所提模型的分割性能,本文从主观评价和客观评价两方面对实验结果进行对比与分析。主观评价主要从人眼观察上对模型的整体分割效果进行感知,客观评价采用图像分割领域中广泛使用的交并比 (intersection over union, IoU) 作为评价指标。本文研究的车辆图像分割任务可被视为像素级别的分类任务,即通过模型预测所有像

素点,并根据预测的判断结果将像素点分类为预定义的目标类和背景类。其中,目标类属于要分割的像素,是正类 (Positive),背景类表示除目标类外的其他像素,是负类 (Negative)。将图像的真实标签与预测标签进行比较,可以得到如图 4 所示的评价关系。

	真实值	Positive	Negative
预测值			
Positive		True Positive (TP) 真阳性	True Positive (FP) 假阳性
Negative		False Negative (FN) 假阴性	False Negative (TN) 假阴性

图 4 真实标签与预测标签比较关系示意图

图中真阳性 (TP) 表示实际所属类别和预测类别均为正类;假阳性 (FP) 表示实际类别为负类,而预测类别为正类;假阴性 (FN) 表示实际类别为正类,而预测类别为负类;真阴性 (TN) 表示实际类别和预测类别均为负类。IoU 也被称为 Jaccard 相似度 (Jaccard similarity),是计算两个集合的交集与并集之比,不仅简洁且极具代表性,根据上述真实标签和预测结果的比较关系,可将 IoU 的表达式量化为如下公式:

$$IoU = \frac{TP}{TP + FN + FP} \quad (6)$$

2.4 实验结果与分析

本文选取了现阶段主流的、分割水平较高的图像分割模型 UNet、DeepLabv3 和所提出的模型进行实验,并对比其分割结果,分析其性能的优劣。同时,为探究设计的多尺度模块具备良好的改进性能,针对仅引入多尺度模块构建的多尺度 UNet (multi-scale UNet, MS-UNet) 模型设置消融实验进行求证。对比实验部分,本文按照上述的实验环境和实施细节,分别在实验数据集上训练了 UNet、DeepLabv3 和本文设计的 MSSA-UNet 模型。3 种不同模型的分割精度如表 1 所示,本文所提出的分割模型在评价指标 IoU 比 UNet、DeepLabv3 分别高出 2.28%、1.73%,分割准确率有着明显提升。

表 1 对比实验结果

分割模型	IoU/%
UNet ^[7]	81.20
DeepLabv3 ^[12]	81.75
MSSA-UNet (本文算法)	83.48

为充分验证多尺度模块对网络的影响力,本文还对 MS-UNet 模型进行了训练,具体的消融实验结果如表 2 所示。从表中可以看出,在标准度量 IoU 上 MS-UNet 模型比原始 UNet 模型高出了 1.74%,证明多尺度模块可以显著提升分割精度。此外,空间注意力机制辅助了网络的特征还原过程,使模型在分割精度上得到进一步提升。

表 2 消融实验结果

分割模型	空间注意力机制	多尺度模块	IoU/%
U-Net ^[7]	×	×	81.20
MS-UNet	×	✓	82.94
MSSA-UNet	✓	✓	83.48

不同模型在车辆图像上的分割结果如图 5 所示,图中第一列表示原始图像,第二列表示真实标注图像,之后的每一列分别对应了不同模型的分割结果。从图中可以看出,U-Net 采用了相同感受野大小的卷积核进行特征提取,模型的分割结果存在明显的过度分割和边缘模糊问题。虽然 DeepLabv3 采用更大的感受野改善了占比较大区域的分割效果,但对于小面积区域依然存在明显的漏分割问题。利用多尺度模块构建的 MS-UNet 模型对比于原始 UNet 模型在车辆图像不同区域间的分割边界上表现更加准确,而相比于 DeepLabv3 也弥补了占比较小区域的漏分割缺陷。由于在特征提取阶段充分利用了多尺度信息,网络的特征表达能力显著提升,分割结果更

好。引入空间注意力机制后,网络在前向传播过程中更加关注重要信息,像素点噪声更少,预测值和真实值也更接近。实验结果证明,利用不同感受野大小的卷积序列构建的多尺度模块对网络结构有着重要影响,可以提取更丰富的语义信息,生成更详细的特征图。通过将多尺度模块嵌入到 UNet 网络结构中,具有不同感受野的卷积层对提取的特征进行多样化处理,网络的特征表达能力明显提高,有效缓解了分割边界模糊和小目标区域容易漏分割的问题,输出了更好的分割结果。由于网络采用包含编解码器的 U 型结构,网络的采样过程在一定程度上造成空间信息丢失,为弥补这一缺陷并尽量轻量化网络结构,合理引入空间注意力机制帮助网络学习更有价值的信息,使空间内不同位置对于目标任务的重要性得以体现,优化了网络的特征还原过程。Dice Loss 有助于学习类别分布,帮助解决像素类别不平衡问题。联合损失函数的设计也让网络在反向传播过程中对难以学习的样本进行稳定优化,使得模型更加关注困难样本,提升了模型的准确性和稳定性。

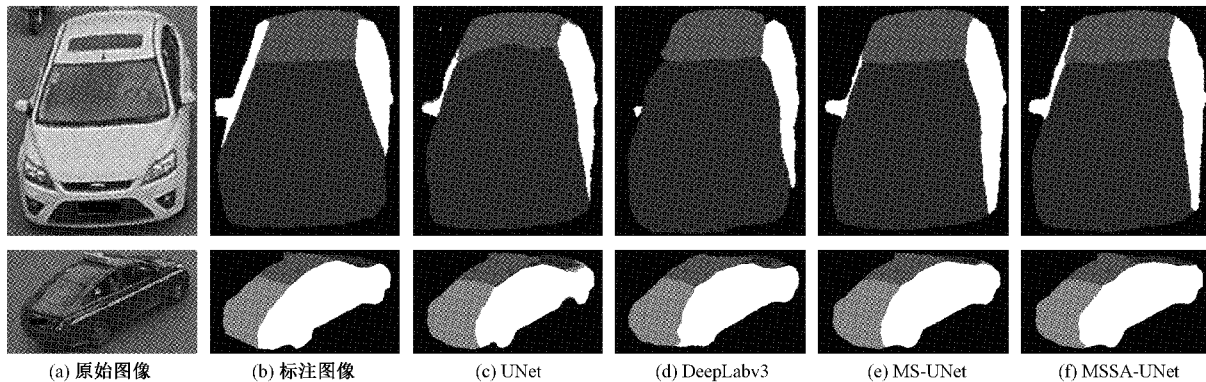


图 5 不同模型在车辆图像上的分割结果

3 结 论

车辆图像分割是车辆识别与跟踪系统中的一个关键组成部分,分割是否准确将直接影响识别与跟踪系统的整体性能。为了提高实际交通场景下车辆图像局部区域的分割精度,本文提出了一种融合多尺度模块和空间注意力机制的 MSSA-UNet 模型。针对原始 UNet 网络进行改进,采用由参数量少且感受野更大的空洞卷积组成卷积序列来构建多尺度模块,对输入图像的特征信息进行多样化处理,使网络具有更强的特征提取能力。另外,在编解码结构上引入空间注意力机制,在轻量化网络的同时让网络在特征权重分配上更加合理,弥补了空间信息丢失问题,有效提高网络的特征恢复能力。实验在主观和客观两方面评价上证明,应用 MSSA-UNet 模型进行车辆图像的局部区域分割性能更好,结果可视化效果和客观评价指标都优于其他分割模型。随着多尺度模块和空间注意力机制的加入,网络的参数数量和训练时间也随之增加,如何有效

轻量化网络结构是未来需要研究的一个方向。

参考文献

- [1] RUDSKOY A, ILIN I, PROKHOROV A. Digital twins in the intelligent transport systems [J]. Transportation Research Procedia, 2021, 54: 927-935.
- [2] 蓝金辉,王迪,申小盼. 卷积神经网络在视觉图像检测的研究进展[J]. 仪器仪表学报, 2020, 41(4): 167-182.
- [3] 谷峥岩,魏利胜. 基于深度学习网络的轴承工件自动检测[J]. 电子测量与仪器学报, 2021, 35(9): 80-88.
- [4] KITRUNGROTSAKUL T, HAN X H, CHEN Y W. Liver segmentation using superpixel-based graph cuts and restricted regions of shape constrains[C]. IEEE International Conference on Image Processing (ICIP), 2015: 3368-3371.
- [5] LIU Y, CAPTUR G, MOON J C, et al. Distance regularized two level sets for segmentation of left and

- right ventricles from cine-MR [J]. *Magnetic Resonance Imaging*, 2016, 34(5): 699-706.
- [6] 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.
- [7] 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*. Springer, Cham, 2015: 234-241.
- [8] IGLOVIKOV V, SHVETS A. Terausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation [J]. *ArXiv PrePrint*, 2018, ArXiv: 1801.05746.
- [9] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. *ArXiv PrePrint*, 2014, ArXiv: 1409.1556.
- [10] CHEN L C, PAPANDEOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs [J]. *Computer Science*, 2014(4): 357-361.
- [11] CHEN L C, PAPANDEOU 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.
- [12] CHEN L C, ZHU Y, PAPANDEOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[J]. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 801-818.
- [13] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[J]. *ArXiv PrePrint*, 2015, ArXiv: 1511.07122.
- [14] 高凯珺,孙韶媛,姚广顺,等.基于深度学习的无人车夜视图像语义分割[J]. *应用光学*, 2017, 38(3): 421-428.
- [15] 张乐,张志梅,刘堃,等.基于全卷积神经网络复杂场景的车辆分割研究[J]. *青岛大学学报(工程技术版)*, 2019, 34(2): 13-20.
- [16] ZHONG Z, LIN Z Q, BIDART R, et al. Squeeze-and-attention networks for semantic segmentation[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 13065-13074.
- [17] MILLETARI F, NAVAB N, AHMADI S A. V-Net: Fully convolutional neural networks for volumetric medical image segmentation [C]. *2016 Fourth International Conference on 3D Vision(3DV)*, 2016: 565-571.
- [18] LIU X, LIU W, MEI T, et al. PROVID: Progressive and multimodal vehicle reidentification for large-scale urban surveillance [J]. *IEEE Transactions on Multimedia*, 2018, 20(3): 645-658.
- [19] MENG D, LI L, LIU X, et al. Parsing-based view-aware embedding network for vehicle re-identification [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 7103-7112.

作者简介

赵红爱, 硕士, 主要研究方向为计算机视觉、车辆检测、分割和再识别。

E-mail: zhaohongai@shu.edu.cn

王旭智, 副教授, 主要研究方向为视频图像后处理、三维图像重建渲染。

E-mail: wangxzw@shu.edu.cn

万旺根, 教授, 博士生导师, 主要研究方向为计算机图形学、信号处理和数据挖掘。

E-mail: wanwg@staff.shu.edu.cn