

田间道路改进 UNet 分割方法

杨丽丽, 陈 炎, 田伟泽, 徐媛媛, 欧非凡, 吴才聪^{*}

(中国农业大学信息与电气工程学院, 北京 100083)

摘 要: 为了保证自动驾驶农机的安全行驶, 需要对农田道路进行高精度识别。该研究以北京市大兴区榆垓镇为研究地点, 构建了农田道路图像数据集, 使用开源标注工具 Labelme 软件进行图像标注, 以 UNet 为基本网络结构, 针对分割过程中存在的道路边缘和远处道路分割效果较差等现象, 提出了 3 个改进方向: 在编码器网络中添加残差连接, 增加网络复杂度; 使用池化卷积融合结构完成下采样, 增加可训练参数以减少信息损失。试验结果表明, 使用 ACBlock (Asymmetric Convolution Block, ACBlock) 和 DACBlock (Dilated Asymmetric Convolution Block, DACBlock) 替换 UNet 中的卷积核, 增加了卷积核“骨架”结构的权重和卷积核的感受野, 提高了远处道路及道路边缘的分割效果, 农田道路分割的交并比值为 85.03%, 相较于原 UNet 提高了 6.52 个百分点, 且高于 ResUNet、UNet3+ 等网络。农机行驶速度在 20 km/h 左右, 该研究网络对于 1 280×720 像素大小的图片平均推理时间为 163 ms, 符合农机自动驾驶时间复杂度要求。该研究提高了自动驾驶农机对农田道路的感知能力, 为安全行驶提供了信息支持。

关键词: 图像分割; 机器视觉; 深度学习; 田间道路; 自动驾驶

doi: 10.11975/j.issn.1002-6819.2021.09.021

中图分类号: TP391.4

文献标志码: A

文章编号: 1002-6819(2021)-09-0185-07

杨丽丽, 陈炎, 田伟泽, 等. 田间道路改进 UNet 分割方法[J]. 农业工程学报, 2021, 37(9): 185-191. doi: 10.11975/j.issn.1002-6819.2021.09.021 <http://www.tcsae.org>

Yang Lili, Chen Yan, Tian Weize, et al. Field road segmentation method based on improved UNet[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(9): 185-191. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2021.09.021 <http://www.tcsae.org>

0 引 言

农业是国民经济发展的基础产业, 目前国内农业生产面临着劳动力不足和生产力低下等问题, 发展农业机械的自动驾驶技术成为了解决这些问题的关键手段^[1]。农机多需要在机库和农田之间往来行驶, 农田道路缺乏导航线, 路边常有杂草和泥土覆盖, 这就需要自动驾驶农机准确识别道路区域, 避免陷入两侧沟渠, 同时躲避道路上的障碍物, 实现安全行驶。机器视觉系统具有价格低廉, 可靠性高的优点, 基于视觉原理进行道路识别, 提取可行驶区域, 对于自动驾驶农机的发展尤为重要。

道路识别通过对采集到的道路图像进行语义分割, 可以看作是像素级的分类任务, 即对每一个像素进行背景和道路的二分类任务, 从而提取可行驶区域。目前对结构化的城市道路识别研究较多, Das 等^[2]使用手工提取特征的方法完成了道路识别, 但这种方法参数调整较为复杂, 鲁棒性较差。Long 等^[3]提出了全卷积网络 (Fully Convolutional Networks, FCN), 第一次将深度学习用于道路识别, 在复杂的城市道路场景中实现了道路区域提取, 但 FCN 仅使用一次上采样将提取的高级语义特征恢

复至原尺寸大小, 对道路边缘分割精度较低。Ronneberger 等^[4]在 UNet 网络中提出了编码解码结构, 解码结构中使用多步上采样结构, 同时融合了下采样过程中的特征, 增强了网络的边缘分割能力。Zhang 等^[5]在 UNet 结构中加入跳跃性的残差连接^[6], 促进了信息的传播, 增加了网络的可训练参数, 在道路分割中减少了噪声点的产生, 提高了多条车道交汇处的道路分割精度。Huang 等^[7]进一步改进了残差连接结构, 将下采样过程中的不同大小的特征图以残差结构连接到了上采样过程中, 将低级语义特征与高级语义特征相融合, 进一步提升了分割效果。为了提高卷积核的感受野, Chen 等^[8-13]将空洞卷积与空间金字塔池化 (Spatial Pyramid Pooling, SPP) 相结合, 对于道路中的障碍物的边缘分割具有良好的效果。不同于结构化的城市道路, 农田道路缺乏明显的车道线等交通标识, 半结构化和非结构化的道路边界区分不明显, 形状复杂多样, 道路两旁的树木多存在遮挡光照的情况, 这些都增大了道路识别的难度。在农田道路特征明显的路段, 使用色彩信息^[14-15]可以实现良好的分割效果, 但这种方法对于光照极其敏感, 缺乏对环境变化的适应性。针对丘陵地区田间道路蜿蜒曲折、坡度大、杂草覆盖多的特点, 李云伍等^[16]使用深度学习的方法, 在 FCN 中使用不同扩张系数的空洞卷积进行级联, 提高了道路边缘形状的分割效果。

本文采集农田道路图像构建数据集, 针对农田道路边界模糊和环境复杂多变等特点, 以 UNet 作为基本结构进行改进, 在农田道路场景下进行道路区域识别,

收稿日期: 2021-01-06 修订日期: 2021-02-07

基金项目: 国家重点研发计划项目(2016YFB0501805)

作者简介: 杨丽丽, 副教授, 博士, 研究方向为计算机网络与智能信息处理的应用。Email: llyang@cau.edu.cn

*通信作者: 吴才聪, 副教授, 博士, 研究方向为无人驾驶与协同作业、农机作业大数据挖掘和农机导航与位置服务的应用。Email: wucc@cau.edu.cn

为后续的自动驾驶农机自主导航奠定基础。

1 农田间道路数据采集

采集设备选取 Stereolabs ZED 双目相机, 水平视场为 90° , 垂直视场为 60° 。该相机可同步采集左、右图像, 图像分辨率为 1280×720 (像素), 帧率为 30 帧/s, 采集软件为相机自带的 ZED Explorer。选取双目相机的左目图像进行数据集的标注和训练。使用 Matlab R2019b 中的 Stereo Camera Calibrator 工具箱对双目相机进行标定, 计算可得左、右相机焦距、内外参数、旋转矩阵和平移矩阵等参数。数据采集于北京市大兴区榆垓镇。北京地区四季分明, 夏秋季节草木茂盛, 春冬季节草木凋零, 为实现数据采集过程中尽可能包含实际生产活动中的不同场景、天气及光照条件, 提高数据集的适用性, 同时更加客观地评估深度学习网络, 选取环境差异较大的春季和夏季进行图像采集, 光照条件包括光照良好、树荫遮挡和背光, 天气情况包括晴天、多云和雨天, 双目相机安装在农机上, 安装位置保证农机行驶过程中无遮挡, 安装高度 1.2 m。数据采集时农机行驶速度约为 5 km/h, 采集路段的农田间道路包括半结构化道路与非结构化道路, 道路宽度为 3~5 m。采集视频时长约 8 h, 从视频中截取图片制作数据集, 筛选路段为机库至农田的田间道路区段, 选择清晰图像, 不考虑农机震动造成的运动模糊, 共筛选出 1 600 张图片, 以 4:1 的比例划分训练集和测试集。使用开源标注软件 Labelme 进行图像标注, 标注得到农田间道路与背景的二值图像, 作为真值图像。

2 改进的 UNet 网络结构

UNet 由编码器网络和解码器网络组成, 编码器用于信息的提取, 解码器用于将特征图尺寸恢复至原图像大小, 网络结构如图 1 所示。在农田间道路分割任务中, UNet 中采用最大池化的下采样方式, 只选取 2×2 (像素) 范围内的最大值代替整个区域, 下采样后特征图宽度和高度为原来的一半, 这对于道路特征信息损失较多。基于此, 本文对 UNet 进行改进, 改进后的网络结构如图 2d 所示。输入图像首先经由 2 个大小为 3 的非对称卷积核 (Asymmetric Convolution Block, ACBlock) [17] 连续卷积提取道路特征, 卷积过程中使用“零值填充” [18]

保证特征图尺寸大小不变, 同时将通道数扩展为 64, 相比于普通卷积核, 具有更好的特征提取能力。在编码器网络中, 每经过 2 次卷积操作, 增加残差结构将输入路径中的农田间道路特征图以像素点相加的方式与输出路径直接相连, 每个卷积操作的后面都会紧跟一个批量标准化处理 [19]。采用池化卷积融合结构 (the Fusion Block of Max Pooling and Convolution, FPCBlock) 代替原 UNet 中最大池化层进行下采样, 将下采样和连续卷积重复 4 次, 最终输出特征图尺寸相对缩小为输入尺寸的 $1/16$, 通道数为 1 024。在最后 2 层中采用空洞非对称卷积 (Dilated Asymmetric Convolution Block, DACBlock) 进行第 1 次卷积。改进后的 UNet 解码器包含 4 次上采样, 上采样的方式为转置卷积, 上采样过程中的特征与对应的下采样特征在通道层进行拼接融合, 最后经由 1×1 卷积将通道数降至 2, 输出农田间道路与背景的二分类结果。

2.1 残差连接

“零值填充”通过对图像四周填充数值为 0 的像素, 可以保持卷积过程中尺寸大小不变, 但这种填充方式不利于图像边缘处特征的提取, 在农田间道路分割中降低了道路边缘形状的分割精度, 残差连接中的增强路径可以有效地改善这个问题。改进的 UNet 在编码器网络中添加了残差连接, 即在输出层加入了一条输入层的恒等映射, 对于神经网络反向传播过程中的梯度消失 [20] 的问题也有所改善。残差连接在不增加额外训练参数的前提下增加了网络复杂度, 训练过程可以产生更多的扰动以模拟复杂的农田间道路环境, 从而提高网络对农田间道路分割的准确率。

2.2 池化卷积融合结构 (FPCBlock)

原始 UNet 采用的下采样结构为最大池化层, 用最大池化的选取范围内的最大值用以代替整个区域的像素值。最大池化层忽略了特征的位置, 具有保持特征平移不变性的优点 [4,11,21-22], 对于处于图像中不同位置的道路都可以保持较好的识别效果, 但由于没有可训练参数, 最大池化层舍弃了 $3/4$ 的像素值, 经过 4 次池化得到的高级语义特征图, 理论上已无法恢复出原图像中小于 16 像素的特征, 损失了较多的有用信息, 从而导致远处小面积的道路区域分割精度较差。

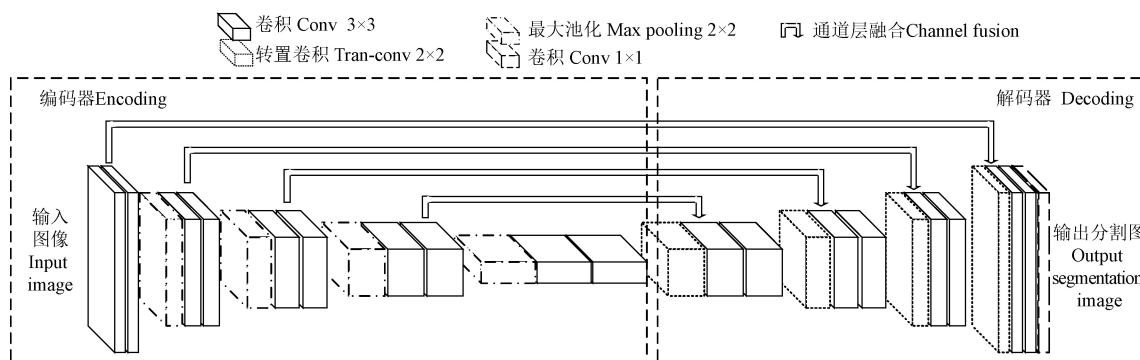


图 1 UNet 网络结构

Fig.1 UNet network structure

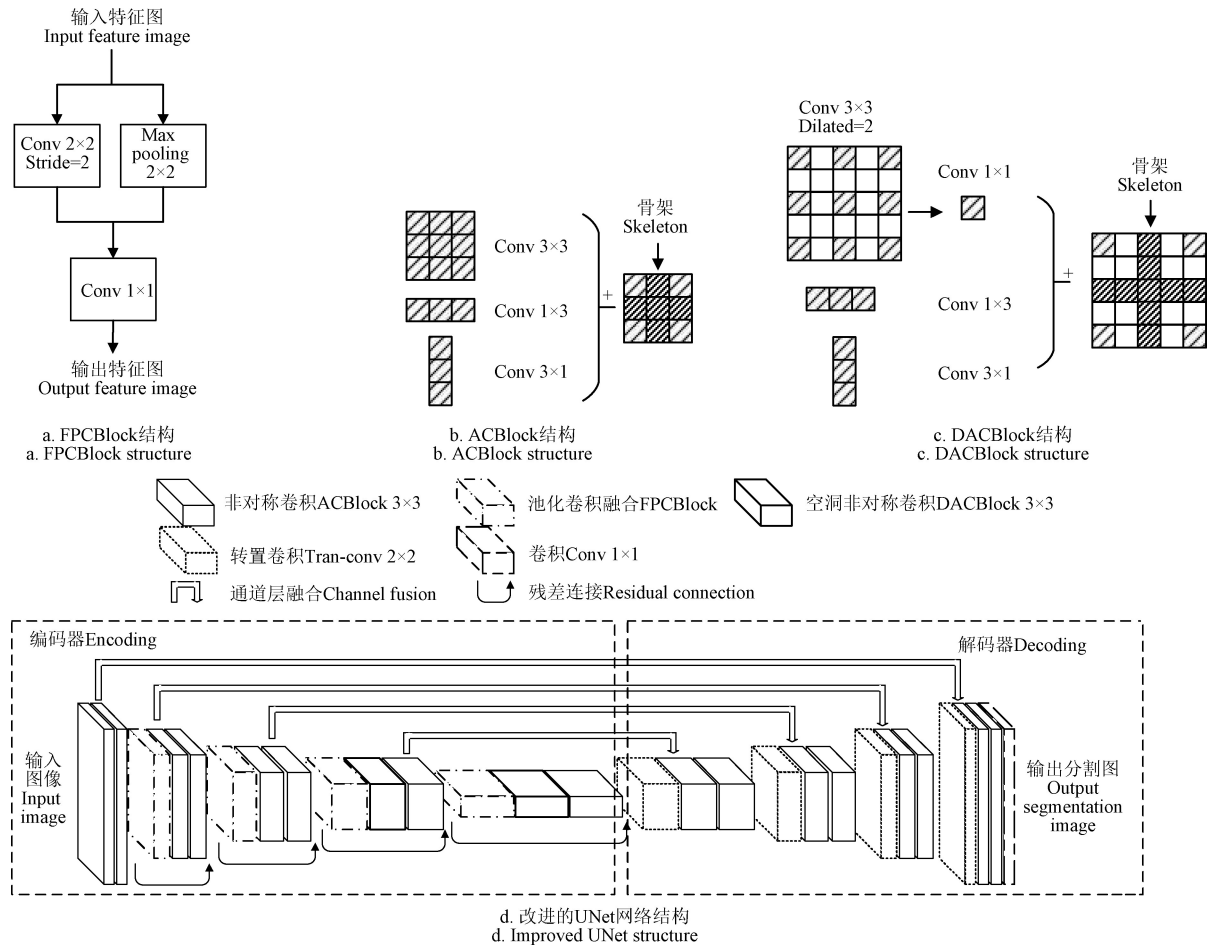


图2 改进的 UNet 网络及各模块结构

Fig.2 Improved UNet structure and each modules' structures

卷积操作的下采样方式具有可训练参数,经过非线性运算从而提取图像中的语义特征^[5,23]。结合最大池化关注于明显特征的提取,而卷积则更有利于整体特征重建^[24]的特性,本文提出了 FPCBlock 结构进行下采样,输入特征图在 FPCBlock 中分别进行大小和步长都为 2 的卷积和池化操作,将得到的结果在特征通道层进行融合,最后采用 1×1 卷积进行降维输出,如图 2a 所示。FPCBlock 的输出层与输入层相比,特征图的尺寸减半,而通道数加倍,保证了卷积过程中道路信息的完整性^[4]。改进的 UNet 网络使用 FPCBlock 替代原始 UNet 下采样过程中的最大池化层,结合了卷积和最大池化的优点,提高了网络对于复杂路况的提取能力,同时保持了在通道层融合结构中编码器和解码器特征图尺寸的一致性。

2.3 ACBlock 与 DACBlock

ACBlock 相比于普通的卷积核,在卷积过程中权重分布不均匀,在中心点附近的“骨架”结构上具有更高的权重,形成非对称的结构,可以提取更多的有效特征,改进的 UNet 使用 ACBlock 替换了原网络中的普通卷积核,通过增加“骨架”结构位置的权重提高了卷积核对于道路特征的提取能力。经过批量标准化融合和分支融合^[17],将 ACBlock 转换为与普通的卷积核相同大小的结构,使改进的 UNet 模型推理时间没有增加。

在 ACBlock 基础上,本文提出了扩大感受野的

DACBlock 结构, DACBlock 中将 3×3 卷积核替换为扩张率为 2 的空洞卷积,另外 2 个分支保持不变,如图 2c 所示。空洞卷积提高了卷积核的感受野,可以获取更多的上下文信息,在神经网络中通过训练得到道路与路边杂草、农田等的某种“联系”,可以更好地提高道路分割精度。但由于空洞卷积中存在空白的间隔,导致分割出的道路图像会出现离散的像素点区域,形成“棋盘效应”^[25]。DACBlock 中先进行空洞卷积,再进行 1×1 卷积的平滑处理,同时另外 2 个分支卷积填补“骨架”结构的空白,提高了获取特征的连续性,有效地改善了“棋盘效应”的问题。改进的 UNet 在下采样的第 4 层和第 5 层中使用 DACBlock 进行第 1 次卷积,增加了卷积核的感受野,从而提高道路识别精度。因为 DACBlock 增加了“骨架”结构处的参数,导致推理时间有所增加。

ACBlock 和 DACBlock 仅对网络中的卷积操作进行修改,未修改网络整体结构,无须修改输入输出和超参数等设置。

3 试验结果与分析

3.1 网络训练

本文提出的改进的 UNet 网络基于 PyTorch 实现,训练数据集共 1 600 张图片,训练集与测试集分别为 1 280 与 320 张图片。选择 Adam 算法进行端到端的训练,初始

学习率设置为 0.000 01, 采用指数衰减的方式动态调整学习率。训练集损失函数使用基于熵值的二分类损失, 计算公式为

$$\text{loss} = -\sum [y \log \hat{y} + (1-y) \log(1-\hat{y})] \quad (1)$$

式中 \hat{y} 为此像素点被预测为道路的概率值, 取值范围为 0~1, y 为此像素点的真值, 若此像素点为道路, 则定义为 1, 否则为 0。验证集损失函数计算方式如下:

$$\text{dice} = \frac{2|A \cap B|}{|A| + |B|} \quad (2)$$

式中 A 、 B 分别为预测道路区域和真值道路区域的面积, 即区域范围内像素点的个数。Batchsize 设置为 8, 默认训练轮数为 500, 使用早停 (Early Stopping Round) 方法, 当连续 10 轮训练中验证集损失函数不下降时, 即停止训练, 早停方法可以避免过拟合的产生, 从而提高模型对于复杂环境的适应性。采用交并比 (Intersection Over Union, IOU) 作为测试集评价标准, 计算公式如下:

$$\text{IOU} = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

3.2 改进网络结构评估

本文使用的模型预测平台 GPU 配置为 GTX 1650, 内存为 4G, 为了评估各个改进结构的效果进行对比试验, 以 UNet 为基础网络, 逐步添加改进结构, 构建 6 组不同的神经网络, 试验如表 1 所示。试验 1 为原始 UNet 网络, 试验 2 在试验 1 的基础上添加了残差连接, 试验 3, 4 在试验 2 的基础上, 分别将下采样方式替换为卷积和 FPCBlock, 试验 5, 6 在试验 4 的基础上, 将卷积核分别替换为 ACBlock 和 DACBlock。通过不同网络的对比, 验证本文不同改进结构的提升效果。训练过程使用农田道路训练集进行训练, 训练完成后在测试集进行测试, 将在测试集上的预测结果与 Labelme 软件标注得到的真值比较, 计算得到 IOU 以评价网络的精度, 网络推理时间为预测过程中平均每张图片的处理时间。

表 1 不同改进网络结构的效果对比

Table 1 Comparison of effects of different improved network structures

试验序号 Test No	残差连接 Residual connection	下采样 Down sampling			卷积核 Convolution kernel			推理时间 Inference time/ms	交并比 Intersection Over Union, IOU/%
		最大池化 Max pooling	卷积 CONV	池化卷积 FPCBlock	卷积 CONV	非对称卷积 ACBlock	空洞非对称卷积 DACBlock		
1	-	√	-	-	√	-	-	147	78.51
2	√	√	-	-	√	-	-	147	79.23
3	√	-	√	-	√	-	-	154	81.45
4	√	-	-	√	√	-	-	158	82.86
5	√	-	-	√	-	√	-	158	83.97
6	√	-	-	√	-	√	√	163	85.03

由表 1 可以看出, 使用 UNet 网络进行农田道路分割, IOU 仅为 78.51%, 本文提出的改进 UNet 网络 IOU 提高了 6.52 个百分点, 达到 85.03%, 由于 FPCBlock 和 DACBlock 中增加了训练参数, 导致模型参数数量和推理时间有所增加。对比试验 2, 3 可以看出, 使用卷积下采样比最大池化下采样 IOU 提高了 2.22 个百分点, 这是因为卷积核在作用区域使用非线性计算提取了特征, 训练过程中通过梯度下降不断提高特征提取能力, 增加了模型训练参数, 减少了下采样过程中的信息损失。对比试验 3, 4 可以看出, FPCBlock 比卷积下采样的 IOU 提高了 1.41 个百分点, 这是因为 FPCBlock 在卷积下采样的基础上, 将卷积与最大池化得到的特征进行融合, 仅增加了 1×1 卷积核中的少量参数, 使模型预测 IOU 进一步得到提升。对比试验 4, 5 可以看出, ACBlock 比普通卷积 IOU 提高了 1.11 个百分点, 这是因为使用 ACBlock 替换普通卷积核后, 增加了“骨架”结构的权重, 这对于图像的旋转变形具有良好的识别效果, ACBlock 卷积核大小与普通卷积核相同, 因此模型参数数量和推理时间与原结构保持一致。对比试验 5, 6 可以看出, DACBlock 比 ACBlock IOU 提高了 1.06 个百分点, 这是因为 DACBlock 在 ACBlock 的基础上使用空洞卷积, 同时在另外 2 个分支中填补了空洞卷积“骨架”结构中的空白部分, 在保持特征图尺寸不变的

前提下, 提高了高级语义特征的感受野, 进一步提高了分割精度。

为了更直观地表示网络改进效果, 选取表 1 中的试验 1、2、4 和 6 共 4 组不同的网络, 分别代表原始 UNet, 和在 UNet 基础上逐步添加残差连接、FPCBlock、ACBlock 和 DACBlock 结构, 选取不同条件的农田道路图像作为输入图像, 以 Labelme 软件标注得到的二值图像作为真值图像, 网络预测结果对比如图 3 所示。

由图 3 可以看出, 春季和夏季农田道路图像中, 背景差异较大, 改进后的 UNet 都能进行较好的道路识别; 多云天气下, 由于杂草遮挡的原因, 在道路交汇处识别精度较差; 雨后存在积水的道路, 由于存在镜面反射, 对积水处存在一定程度的误识别; 在光照良好和傍晚、树荫等光照较弱的情况下, 可以较好地完成道路分割, 保障了农机的安全行驶。

3.3 与其他神经网络对比

为了进一步评价改进 UNet 对于田间道路分割的性能, 使用 UNet^[4]、ResUNet^[5]、DeeplabV3^[10]、DeeplabV3+^[11]、UNet3+^[7], 在农田道路训练集上进行训练后, 然后在测试集上进行测试得到推理时间和 IOU, 并与本文提出的改进 UNet 进行对比, 以对比不同神经网络在农田道路分割时的精度和时间复杂度, 比较结果如表 2 所示。

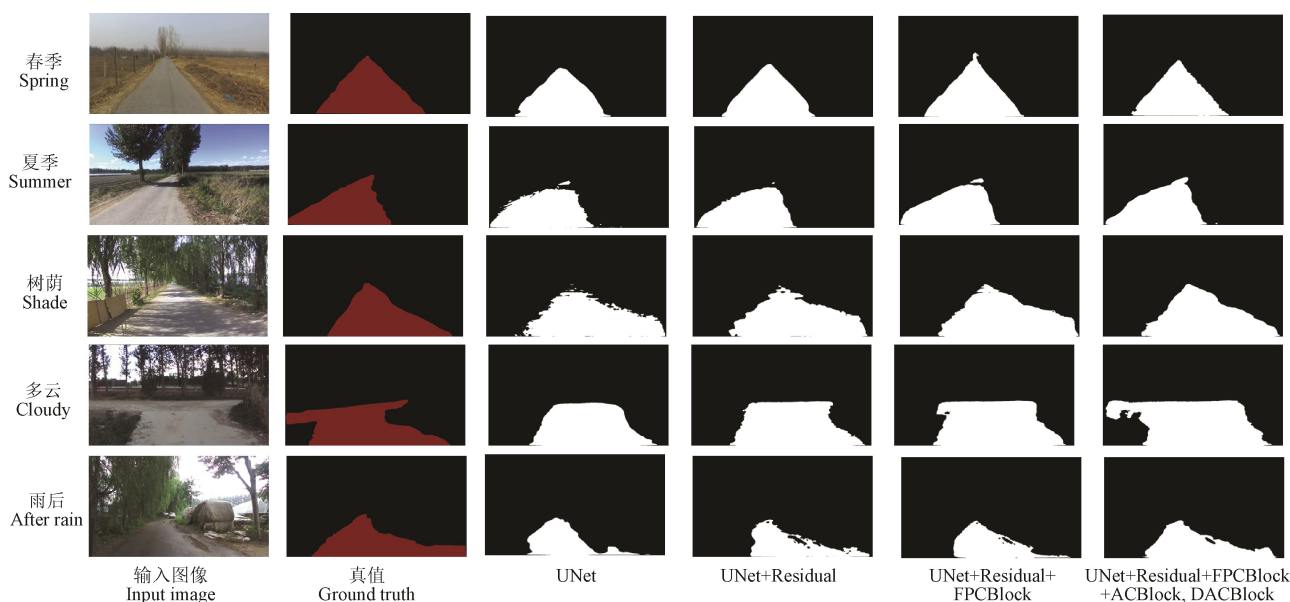


图 3 不同场景下的网络预测效果对比

Fig.3 Comparison of network prediction effects under different scenarios

表 2 不同网络结构的田间道路分割效果对比

Table 2 Comparison of field road segmentation effect with different network structures

网络 Network	推理时间 Inference time/ms	IOU/%
UNet	147	78.51
ResUnet	147	79.28
DeeplabV3	155	82.65
DeeplabV3+	176	83.78
UNet 3+	136	84.87
Improved UNet	163	85.03

从表 2 可以看出，本文提出的改进的 UNet 网络，在分割准确率 IOU 要优于其他网络。对比针对编码器网络和解码器网络的改进效果，表 1 中的试验 2 仅在编码器网络中增加了残差连接，IOU 为 79.23%，而 ResUnet 在编码器和解码器网络都增加了残差连接，IOU 为 79.28%，可以看出，改进解码器网络模型提升效果不明显，与前人结论一致^[24]。

在农田道路分割任务中，道路区域在图像中的分布位置呈现出近处占比高，远处占比小的特点，道路边缘缺乏规则的分界线，这些都增大了道路分割的难度。在农田道路分割任务中，分割精度对于自动驾驶尤为重要。本文提出的改进的 UNet 分割 IOU 为 85.03%，相比其他网络精度较高，网络推理时间平均每张图片为 163 ms，推理速度低于 ResUnet，UNet 3+ 等网络。农机在农田道路上的行驶速度在 20 km/h 左右，表 2 中所有网络推理时间都能满足自动驾驶农机对于道路识别的速度要求。相比于其他网络，本文提出的改进 UNet 在推理时间满足要求的同时识别效果最优，适合自动驾驶农机的田间道路识别。

4 结 论

在农田道路分割任务中，本文以 UNet 为基本网络结构，提出了以下 3 个改进方向：

1) 在解码器网络中每两个卷积操作之间建立一条恒等映射通道，通过像素点相加的方式建立残差连接，改善了训练过程中梯度消失的问题，跳跃连接结构促进了网络中信息传播效率，增加了网络复杂度。

2) 建立 FPCBlock，替代 UNet 网络中最大池化层。该结构结合了卷积和最大池化操作的优点，通过使用可训练参数进行下采样，减少了信息的损失。

3) 使用 ACBlock 和 DACBlock 替换 UNet 中的 3×3 卷积核，ACBlock 提高了卷积核的使用效率，DACBlock 使用空洞卷积提高了卷积核的感受野，同时填补了空洞卷积“骨架”结构中的空白部分，改善了“棋盘效应”的问题。ACBlock 和 DACBlock 使用方便，可以直接替换网络中的普通卷积核，无须修改超参数设置，便于进行网络的训练。

本文提出的改进的 UNet 网络，分割 IOU 为 85.03%，相较于原 UNet 提高了 6.52 个百分点，且高于 ResUnet、UNet 3+ 等网络，平均每张图片推理时间为 163 ms，符合自动驾驶农机的时间要求。

[参 考 文 献]

- [1] Han S F, He Y, Fang H. Recent development in automatic guidance and autonomous vehicle for agriculture: A Review[J]. Journal of Zhejiang University: Agriculture & Life Sciences, 2018, 44(4): 381-391, 515.
韩树丰，何勇，方慧. 农机自动驾驶及无人驾驶车辆的发展综述[J]. 浙江大学学报：农业与生命科学版，2018，44(4): 381-391，515. (in English with Chinese abstract)
- [2] Das S, Mirmaline T T, Varghese K. Use of salient Features for the design of a multistage framework to extract roads from high-resolution multispectral satellite images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2011, 49(10): 3906-3931.

- [3] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4): 640-651.
- [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. Springer, 2015: 234-241.
- [5] Zhang Z, Liu Q, Wang Y. Road extraction by deep residual U-Net[J]. IEEE Geoscience & Remote Sensing Letters, 2017, 32(99): 1-5.
- [6] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// Conference on Computer Vision and Pattern Recognition, IEEE 2016: 5410-5418.
- [7] Huang H, Lin L, Tong R, et al. UNet 3+: A full-scale connected UNet for medical image segmentation[C]// International Conference on Acoustics, Speech and Signal Processing, IEEE 2020: 1717-1722..
- [8] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[J]. Computer Science, 2014, 25(4): 357-361.
- [9] 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.
- [10] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[C]// Computer Vision and Pattern Recognition, IEEE, 2017: 3061-3070.
- [11] Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]// Computer Vision and Pattern Recognition, IEEE, 2018: 4040-4048.
- [12] 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.
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).
- [13] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9): 1904-1916.
- [14] 宋怀波, 何东健, 辛湘俊. 基于机器视觉的非结构化道路检测与障碍物识别方法[J]. 农业工程学报, 2011, 27(6): 225-230.
Song Huaibo, He Dongjian, Xin Xiangjun. Unstructured road detection and obstacle recognition algorithm based on machine vision[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2011, 27(6): 225-230. (in Chinese with English abstract)
- [15] 张雄楚, 陈兵旗, 李景彬, 等. 红枣收获机视觉导航路径检测[J]. 农业工程学报, 2020, 36(13): 133-140.
Zhang Xiongchu, Chen Bingqi, Li Jingbin, et al. Path detection of visual navigation for jujube harvesters[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(13): 133-140. (in Chinese with English abstract)
- [16] 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)
- [17] Ding X, Guo Y, Ding G, et al. ACNet: Strengthening the kernel skeletons for powerful CNN via asymmetric convolution blocks[C] International Conference on Computer Vision, IEEE 2019: 1911-1920.
- [18] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [19] Wu S, Li G Q, Deng L, et al. L1-Norm batch normalization for efficient training of deep neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 30(7): 2043-2051.
- [20] Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25(2): 1097-1105.
- [21] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks[J]. Journal of Machine Learning Research, 2010, 9(2): 249-256.
- [22] 陈进, 韩梦娜, 练毅, 等. 基于 U-Net 模型的含杂水稻籽粒图像分割[J]. 农业工程学报, 2020, 36(10): 174-180.
Chen Jin, Han Mengna, Lian Yi, et al. Segmentation of impurity rice grain images based on U-Net model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(10): 174-180. (in Chinese with English abstract)
- [23] Yang X, Li X, Ye Y, et al. Road detection and centerline extraction via deep recurrent convolutional neural network U-Net[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(9): 7209-7220.
- [24] Minaee S, Boykov Y, Porikli F, et al. Image segmentation using deep learning: A survey[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 29(10): 4550-4568.
- [25] Gao L, Song W, Dai J, et al. Road extraction from high-resolution remote sensing imagery using refined deep residual convolutional neural network[J]. Remote Sensing, 2019, 11(5): 552-557.

Field road segmentation method based on improved UNet

Yang Lili, Chen Yan, Tian Weize, Xu Yuanyuan, Ou Feifan, Wu Caicong^{*}

(College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China)

Abstract: Automatic driving of agricultural machinery has drawn much more attention in recent years, particularly with the development of precision farming and the improvement of sensor technologies. Four parts of autonomous driving are positioning, perception, decision-making, and control system. In perception, the road recognition aims to extract the drivable area for the safe driving of agricultural machinery. However, there are no obvious lane markings or signs for field roads, while the road borders are in irregular shape, often shaded by trees. All of these features make it difficult for field road identification, unlike structured urban road. In road recognition, semantic segmentation on the collected road images is a binary classification task of background and road for each pixel to extract the drivable area. In this study, the data in spring and summer was collected in the Yufa Town, Daxing District, Beijing of China. A stereo camera was fixed on the agricultural machine to collect image data. The fixed position ensured that the camera was firm and reliable without being obscured during driving. The fixed height was set to 1.2 m. The driving speed of agricultural machinery was about 5 km/h during data collection. The field roads included semi-structured and unstructured roads. The sunny day was selected to collect data. The collecting time was about 4 hours, and a total of 1 600 pictures were captured. The training and test set were divided into the ratio of 4:1. The open-source software Labelme was used for image labeling. UNet was selected as the basic network, due to its simplicity and suitability for binary classification. A better performance was achieved when training on a small data set. Three improvements were also proposed for the UNet. 1) An identity mapping channel was established between every two convolutions, and the residual was constructed by adding pixels. The residual connection was used to alleviate the gradient disappearance and explosion during training, while easy the training of deep neural networks. 2) A fusion convolutional structure and the maximum pooling were established to replace the maximum pooling layer in the UNet. The useful information in the original image was maximized when halving feature map, where the segmentation of small area features was improved significantly. The inference time of the model was much longer because much more convolution operation increased the training parameters. 3) An asymmetric convolution structure was used in ACBlock, where the weight of the "skeleton" structure increased to improve the efficiency of feature extraction in the convolution kernel. Inspired by ACBlock, DACBlock was proposed using the dilated convolution, which further expanded the receptive field of the convolution feature map. ACBlock and DACBlock were used to replace the 3×3 convolution kernel in UNet. As such, the segmentation accuracy of road edge shapes was improved significantly. The hierarchical fusion and batch normalization were used in the inference stage to maintain that the number of parameters and inference time were all the same as the original structure. The improved UNet presented an IOU value of 85.03% for the field road segmentation, higher than the original UNet, ResUNet, and UNet3+. The recognition accuracy was relatively lower under cloudy weather in road junctions, due to insufficient light and occlusion. There was always water in the middle of the road after rain, where a certain degree of reflection occurred on the water under the mirror reflection. Therefore, the water increased the error of road segmentation. In the case of good or weak light in the evening and shade, the road segmentation was performed better for the safe driving of agricultural machinery. The segmentation accuracies of remote roads and road edges were also significantly better than those of other networks. Moreover, the average inference time of the model was 163 ms, meeting the time requirements of automatic driving in agricultural machinery.

Keywords: image segmentation; machine vision; deep learning; field roads; automatic driving