

文章编号: 2095-2163(2021)11-0148-06

中图分类号: TP391

文献标志码: A

# 非结构化道路分割算法研究

赵晋燕<sup>1</sup>, 罗素云<sup>1</sup>, 陈杨钟<sup>2</sup>

(1 上海工程技术大学 机械与汽车工程学院, 上海 201620; 2 大工科技(上海)有限公司, 上海 200000)

**摘要:** 现阶段无人驾驶汽车主要依靠视觉技术来完成车辆的环境感知,在道路识别领域,视觉技术能准确识别出道路可行驶区域。为了增强无人驾驶车辆在非结构化道路对场景区域的判别能力,本文基于 SegNet、ENet、UNet 3 种在多分类任务中取得较好成果的分割网络架构,通过对模型网络或参数的调整和修改,提出一种能很好应用到非结构化道路区域分割问题的分割模型。通过拍摄并制作标签数据集,采取不同的评价指标进行分析,得到最佳的道路区域分割模型,用于预测非结构化道路的可行行驶区域。实验证明,相比较于传统的非结构化道路分割的区域生长模型,本文提出的分割模型在分割精度上有明显提升。

**关键词:** 非结构化道路; SegNet; ENet; UNet; 区域分割

## Research on unstructured road segmentation algorithm

ZHAO Jinyan<sup>1</sup>, LUO Suyun<sup>1</sup>, CHEN Yangzhong<sup>2</sup>

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

2 Dagong Technology (Shanghai) Co., Ltd., Shanghai 200000, China)

**[Abstract]** At present, driverless vehicles mainly rely on vision technology to complete the vehicle environment perception. In the field of road recognition, the task of vision technology is to accurately identify the drivable area of the road. In order to enhance the discrimination ability of driverless vehicles in unstructured road for scene area, based on SegNet, ENet and UNet, this paper proposes a segmentation model for unstructured road image segmentation by adjusting and modifying the model network or parameters. By shooting and making label data sets, and taking different evaluation indexes for analysis, the best road region segmentation model is obtained to predict the driving area of unstructured road. Experimental results show that compared with the traditional region growing model of unstructured road segmentation, the segmentation accuracy of the proposed model is significantly improved.

**[Key words]** unstructured road; SegNet; ENet; UNet; image segmentation

## 0 引言

无人驾驶这一新兴产业带动着计算机技术、机器视觉、图像处理等相关算法研究的快速发展。无人驾驶实现无人控制的前提在于对行驶环境准确的分析,对道路的识别,即需得到可行行驶区域。道路区域一般划分为结构化道路和非结构化道路。结构化道路具有确定的道路材料、道路宽度以及明显的车道线,基于车道线的研究已进入成熟阶段;而对于非结构化道路的研究,由于会受到天气、道路材料、阴影、积水等环境因素影响,准确分割出道路区域是至关重要的。

图像分割就是在一幅图像中,把目标从背景中分离出来。分割原理是根据图像的灰度、纹理、形状等特征将图像划分成互不相交的若干区域,同时保证区域内部在分割特征上保持各向同性,区域与区

域之间在分割特征上保持各向异性<sup>[1]</sup>。常见的分割算法主要分为4类:基于阈值的分割方法、基于边缘检测的分割方法、基于区域的分割方法以及基于深度学习的分割方法。

基于阈值的分割算法根据图像的灰度特征以及分割类别,确定一个或多个最佳的阈值,将图像的逐个像素与确定的类别相比较,得到分割结果,应用最广泛的选取阈值的方法是最大类间差法(又称大津法(OTSU))<sup>[2]</sup>;基于边缘检测的分割方法是有局限性的,其原理是基于一幅图像中不同物体间边缘的灰度值梯度较大的特点,在图像上寻找图像灰度变化跳变点的集合,将这些点的集合认定为图像的边缘,将所有的边缘点连接起来形成闭环,图像就被分割成了不同的区域,目前普遍采用的检测边缘的方法有 Sobel 边缘检测算法、Laplace 边缘检测算法、Canny 边缘检测算法等等;区域分割算法又分为两

**作者简介:** 赵晋燕(1997-),男,硕士研究生,主要研究方向:机器视觉、图像处理;罗素云(1975-),女,硕士,副教授,主要研究方向:无人驾驶汽车环境感知及控制;陈杨钟(1983-),男,硕士,工程师,主要研究方向:控制理论与控制工程。

收稿日期: 2021-05-31

哈尔滨工业大学主办 ◆ 专题设计与应用

类:区域分裂合并和区域生长。区域分裂合并算法的基本思想是先确定一个分裂合并的准则,即区域特征一致性的测度,当图像中某个区域的特征不一致时就将该区域分裂成 4 个相等的子区域,当相邻的子区域满足一致性特征时则将其合成一个大区域,直至所有区域不再满足分裂合并的条件为止;区域生长算法是根据事先定义的准则将像素或者子区域聚合成更大区域的过程,其基本思想是从一组生长点开始(生长点可以是单个像素,也可以是某个小区域),将与该生长点性质相似的相邻像素或者区域与生长点合并,形成新的生长点,重复此过程直到不能生长为止。生长点和相似区域的相似性判断依据可以是灰度值、纹理、颜色等图像信息。基于深度学习的分割方法是在全卷积网络(Fully Convolutional Networks(FCN))基础上产生的一系列分割网络模型,如 SegNet、U-Net、Mask-RCNN、Deeplab 等<sup>[3]</sup>。卷积神经网络分割模型根据分割级别分为 patch-wise 和 pixel-wise classification<sup>[4]</sup>。

本文基于在多分类任务中取得较好成果的 3 个分割网络模型,通过对模型参数的调整和修改,应用到非结构化道路区域的分割问题上,采取不同的评价指标进行分析,同时与传统的区域生长模型的分割效果进行比较,得到最佳的道路区域分割模型用于预测非结构化道路的可行驶区域。

## 1 卷积神经网络模型

全卷积神经网络是第一个被用于图像语义分割的深度神经网络框架,后续的卷积神经网络模型都是在此基础上进行改进优化的。深度神经网络结构,一般主要由 4 部分组成,即卷积层、池化层、全连接层和非线性激活单元。卷积层对图片做卷积操作,得到相应的特征图;池化层进行降采样处理,保留主要特征的同时减少参数的计算量;全连接层将卷积层产生的特征图映射成一个固定长度的特征向量,从而解决图像的分类问题。

分割首要考虑的是减少采样过程的分割精度丢失问题,提高分割精度。目前解决下采样中边缘信息丢失问题的主流方法有两种:一是 FCN 中添加编码层的 feature map,如 UNet;二是通过保留编码网络中最大池化过程中最大值的索引,并借此在解码网络中生成稀疏的上采样 feature map,如 SegNet、ENet。本文通过修改网络或调试参数的方式比较了这 3 种网络模型在非结构化道路场景的分割结果。

### 1.1 SegNet 网络模型

SegNet 网络模型是 2016 年 Badrinarayanan 等人提出的一种对称式编码器-解码器(encoder-decoder)框架,其网络结构如图 1 所示<sup>[5]</sup>。SegNet 模型本质基于 FCN 模型,利用 VGG16 网络改进得到的,SegNet 的新颖之处在于解码器对其较低分辨率的输入特征图进行上采样的方式<sup>[6]</sup>。具体地说,解码器网络通过保持对应编码器层的 maxpooling 索引来更新特征映射,解码器使用了相应编码器的最大池化步骤中计算的池化索引来执行非线性上采样,这种方法消除了学习上采样的需要。经上采样后的特征图是稀疏的,因此随后使用可训练的卷积核进行卷积操作,生成密集的特征图。

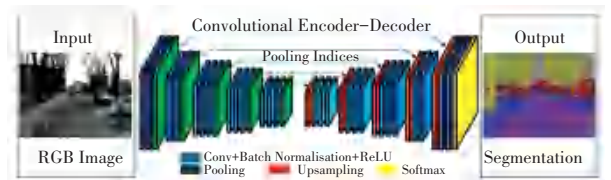


图 1 SegNet 网络结构

Fig. 1 SegNet network architecture

VGG16 网络具有很大的参数量,本文针对非结构化道路分割,选择采用基于 ResNet 模型的 SegNet 网络,也就是说用 ResNet50 网络取代了原来 SegNet 的基于 VGG16 的编码网络,极大的减少了网络参数量,加快了网络的训练速度,网络的配置见表 1。

表 1 SegNet 网络配置

Tab. 1 SegNet network configuration

项	卷积层	卷积核大小/ 特征图数量	图像尺寸 ( $H \times W \times C$ )
输入图像			512×512×3
编码阶段	第一层	3×3/64	256×256×64
	第二层	3×3/128	128×128×128
	第三层	3×3/256	64×64×256
	第四层	3×3/512	32×32×512
	第五层	3×3/512	16×16×512
解码阶段	第六层	3×3/512	32×32×512
	第七层	3×3/256	64×64×256
	第八层	3×3/128	128×128×128
	第九层	3×3/64	256×256×64
	第十层	3×3	256×256×2
Softmax 层			

### 1.2 ENet 网络模型

ENet,不同于对称式编码解码的 SegNet 模型,主要利用残差模块形成的一种非对称式的卷积网络,采用了普通卷积、膨胀卷积、转置卷积等不同卷积形式<sup>[7]</sup>。网络配置见表 2。

表 2 ENet 网络配置  
Tab. 2 ENet network configuration

项	卷积层	类型	图像尺寸 ( $H \times W \times C$ )	
输入图像			512×512×3	
	Initial block		256×256×16	
编码阶段	bottleneck1.0	Downsampler block	128×128×64	
	bottleneck1.x(x:1 2 3 4)		128×128×64	
	bottleneck2.0	Downsampler block	64×64×128	
	bottleneck2.1		64×64×128	
	bottleneck2.2	dilated 2	64×64×128	
	bottleneck2.3	asymmetric 5	64×64×128	
	bottleneck2.4	dilated4	64×64×128	
	bottleneck2.5		64×64×128	
	bottleneck2.6	dilation 8	64×64×128	
	bottleneck2.7	asymmetric 5	64×64×128	
	bottleneck2.8	dilation 16	64×64×128	
	bottleneck3.0-3.7 与 bottleneck2.1-2.8 结构相同			64×64×128
	解码网络	bottleneck4.0	Deconvolution (upsampling)	128×128×64
bottleneck4.1			128×128×64	
bottleneck4.2			128×128×64	
bottleneck5.0		Deconvolution (upsampling)	256×256×16	
bottleneck5.1			256×256×16	
fullconv			512×512×2	

ENet 网络结构包括一个偏大的编码网络和一个较小的解码网络,突出特点在于 bottleneck 模块的改变,如图 2 所示。

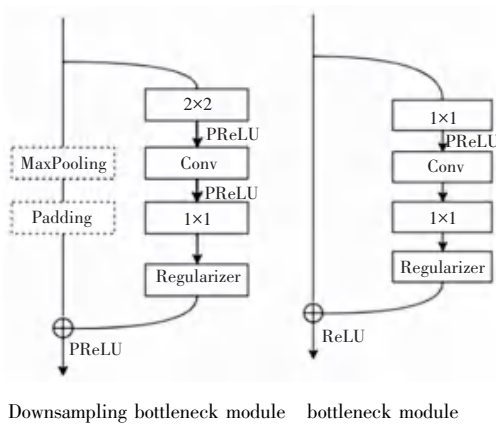


图 2 bottleneck 模块

Fig. 2 Bottleneck module

下采样的 bottleneck 模块主线包括 3 个卷积层,先是 2×2,步幅为 2 的卷积映射做降采样;然后是卷积,这里的卷积层对应不同功能分别采用 3 种不同的

卷积,Conv 普通卷积、Asymmetric 分解卷积、Dilated 空洞卷积,后面再接一个 1×1 做升维。每个卷积层后均接 Batch Norm 和 PReLU。辅线包括最大池化和 Padding 层,最大池化负责提取上下文信息,Padding 负责填充通道,达到后续的残差融合,融合后再接 PReLU。非下采样的 bottleneck 主线包括 3 个卷积层,先是 1×1 映射,用于减少维度;然后是主卷积,有 3 种可能,Conv 普通卷积,Asymmetric 分解卷积, Dilated 空洞卷积,后面再接一个 1×1 的做升维。每个卷积层后均接 Batch Norm 和 PReLU,辅线直接恒等映射,因为只有下采样才会增加通道数,故这里不需要 padding 层,融合后再接 PReLU。

### 1.3 UNet 网络模型

UNet 是基于 FCN 网络改进的一种对称式编码器-解码器 (encoder-decoder) 框架,是 2015 年德国弗莱堡大学计算机科学系和生物信号研究中心 Olaf Ronneberger 等人针对生物医学图像提出的分割网络<sup>[8]</sup>。

UNet 网络模型如图 3 所示,可以分为 3 部分:

Encoder 网络、Decoder 网络和 Predict 网络。Encoder 网络又称主干提取网络,位于结构图左侧,网络包括 5 层,每层包含两个  $3 \times 3$  的卷积层 (ReLU)、 $2 \times 2$  的 max pooling 层 (stride = 2), 每层之间通过一个下采样层连接。从图 3 可以看出,通过主干特征提取得到 5 个有效特征层,在 Decoder 网络利用得到的 5 个有效特征层进行特征融合。Decoder 网络又称加强特征提取网络,位于结构图右侧,每一层由一个  $2 \times 2$  的上采样卷积层 (ReLU)、Concatenation (crop 对应的 Encoder 层的输出 feature map,与 Decoder 层的上采样结果相加)、2 个  $3 \times 3$  的卷积层 (ReLU) 构成。最后一层 Predict 网络,通过一个  $1 \times 1$  卷积将最后得到的有效特征层中的每个特征点进行分类。

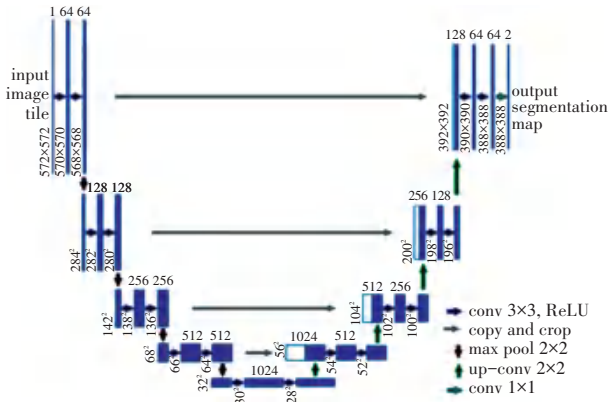


图 3 UNet 网络结构

Fig. 3 UNet network architecture

## 2 实验结果及分析

### 2.1 数据集

不同于常见基准数据集,包括 KITTI、CityScapes、DeepDrive、ApolloScape、Oxford robocar 和 CamVid 等,其代表了自动驾驶汽车可能在其中运行的城市环境,对于需要在非结构化、非公路环境中进行视觉感知的应用程序,现有的数据集存在明显的差距。为此,本文制作了复杂环境的数据集。

非结构化道路包括很多种类,比如沙漠路、泥土路、水泥路等等,为了验证得到的模型具有更加可靠的性能,通过实景拍摄并制作了包括多种情景的数据集。该数据集共有 3 000 张图片,并利用 labelme 对图像进行手动标注。部分标注图如图 4 所示,第一列是采集到图像,包含了常见的泥土路、水泥路、砖石路、雨后道路、岔路口等复杂的道路环境;第二列是标注后得到的二值图像,黑色为背景区域,红色为道路区域;第三列是标注图像。

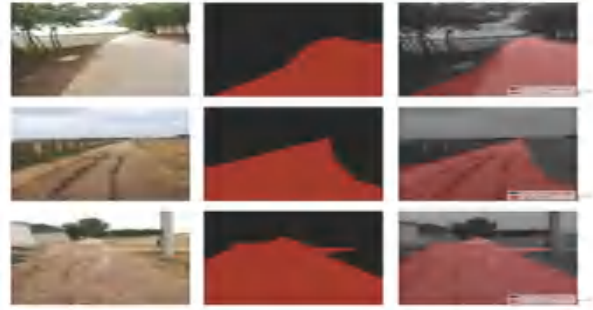


图 4 标注图像

Fig. 4 Annotated images

### 2.2 评价指标

图像分割模型的评价指标需要从多方面性能考虑,如执行时间、内存占用 (内存使用率)、分割精度等。本文的研究重点在于准确分割出非结构化道路图像中的可行使区域,主要对比了不同模型在分割精度上的性能。现阶段语义分割都是基于 pixel-wise 分类,本文采用了两种主要的衡量指标:MPA (Mean Pixel Accuracy 均像素精度) 和 MIoU (Mean Intersection over Union 均交并比)。

假设分割类别为  $k$  类,总计有  $k + 1$  分类 (其中含一个背景类别),  $i$  表示真实值,  $j$  表示预测值。  $P_{ii}$  就表示  $TP$  (true positives),  $P_{ij}$  与  $P_{ji}$  分别表示  $FP$  (false positives) 与  $FN$  (false negatives)。

(1)  $MPA$ : 每个类别正确的像素总数与相对应类别总数的比率相加后求均值,如公式 (1) 所示:

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

(2)  $MIoU$ : 计算两个集合真实值 (ground truth) 和预测值 (predicted segmentation) 的交集和并集之比。这个比例可以变形为正真比上真正、假负、假正 (并集) 之和,在每个类上计算  $MIoU$ , 之后平均,如图 5 所示。

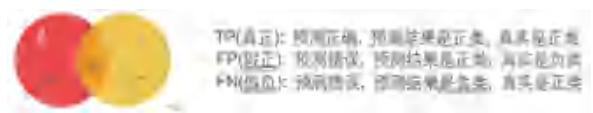


图 5 交并比图示

Fig. 5 Intersection ratio diagram

在图 5 中,  $A$  为真实值, 占据了图像的一定区域;  $B$  为预测值, 也在图像的一定区域。中间的真正部分 ( $TP$ ) 就是真实值和预测值的交, 图像有颜色的部分就是真实值和预测值的并。  $MIoU$  就是计算每一类交并比的平均, 如公式 (2) 所示。

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ij} + \sum_{j=0}^k P_{ji} - P_{ii}} \quad (2)$$

### 2.3 结果分析

本文实验均是在 Ubuntu 系统下进行的,模型的训练、测试和预测基于以下配置环境:Ubuntu16.04.7+CuDa 11.0.207+CuDnn +Pytorch。实验对比了3种分割网络与传统区域生长算法在不同道路环境的分割效果,如图6所示。可以看出,如图6的第一行图片所示,在存在岔路口的区域,只有 UNet 网络可以准确的分割出岔道口的小的道路区域;如图6的第

二行所示,在存在弯道道路情况下,ENet 网络未分割出弯道区域后的道路区域,SegNet 模型、UNet 均可以准确分割;如图6的第三、第四行所示,在泥土材质、水泥材质的直线道路,3种分割网络都可以准确识别出可行使区域,而传统区域生长算法存在漏分割区域、分割区域存在孔洞等问题。

为了进一步说明3种分割网络的优劣,采用 MPA、MIoU 两种评价指标进行定量说明,见表3,可以看出 UNet 网络在不同指标下都取得了优于另外两种网络的结果。

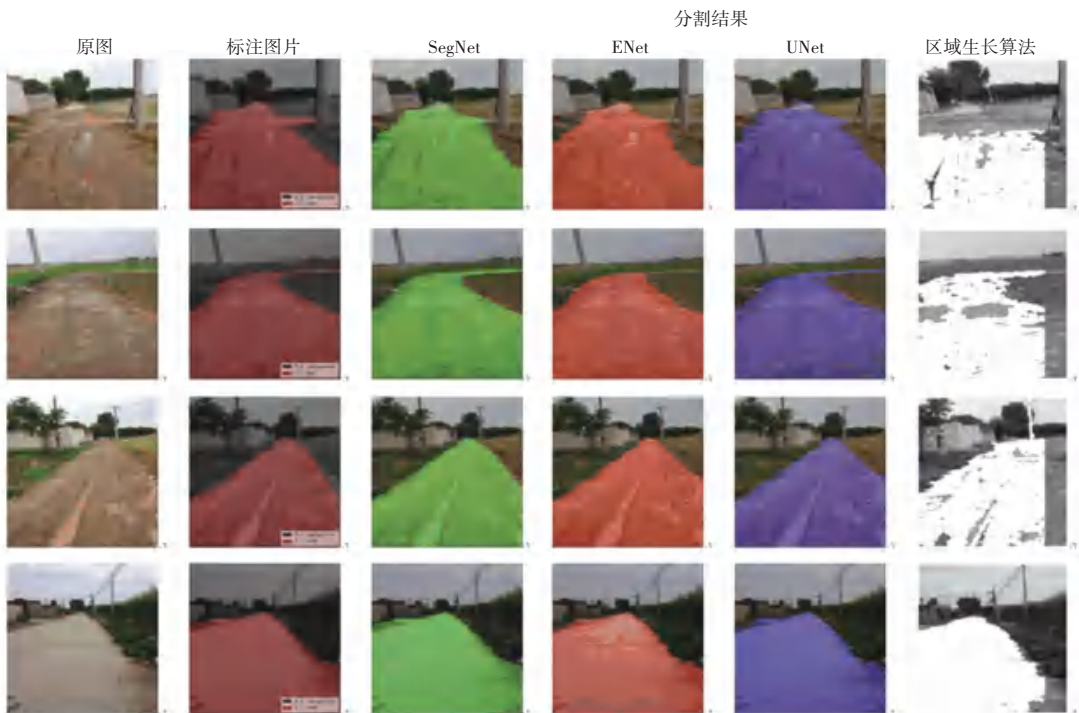


图6 不同模型的分割结果

Fig. 6 Segmentation results of different models

表3 数据集性能结果

Tab. 3 Data set performance results

	Model	Accuracy(MPA)	MIoU
1	SegNet	97.25	94.57
2	ENet	96.99	94.13
3	UNet	97.53	95.11

### 3 结束语

本文通过制作数据集并标注,调整3种分割网络的参数,比较了不同模型在非结构化道路上的分割性能,并通过量化指标加以比较,且与传统应用于非结构化道路的区域生长算法进行了效果上的比较,得到了适用于非结构化道路区域的分割网络,准确的可行使区域分割为下一步可行使区域的边界拟

合研究奠定了基础。

### 参考文献

- [1] 章毓晋. 图像分割[M]. 北京科学出版社.2001.
- [2] OTSU N. A Threshold Selection Method from Gray - Scale Histograms[J]. IEEE Trans.on SMC, 1979, 9.
- [3] 林敏, 陈建新, 陈哲亮. 基于边缘检测与分裂合并的图像分割算法[J]. 计算机技术与应用, 2011, 37(7): 130-133.
- [4] LAW T Y, HENG P A. Automated extraction of bronchus from 3D CT images of lung based on genetic algorithm and 3D region growing[J]. Proceedings of SPIE-The International Society for Optical Engineering, 2000, 3979:906-916.
- [5] RANFT B, STILLER C. The Role of Machine Vision for Intelligent Vehicles[J]. IEEE Transactions on Intelligent Vehicles, 2016:8-19.

(下转第156页)