

doi:10.11835/j.issn.1000-582X.2020.03.009

基于 SegNet 的非结构道路可行驶区域语义分割

张凯航, 冀 杰, 蒋 骆, 周显林

(西南大学 工程技术学院, 重庆 400715)

摘要: 为了增强自动驾驶车辆对非结构化道路中可行驶区域的场景理解能力, 基于 SegNet 深度学习网络结构, 提出了一种针对非结构道路的可行驶区域语义分割方法。在传统的卷积神经网络基础上, 构建编码-解码深度卷积神经网络, 用于自动习得图片中非结构化道路的特征, 通过在数据集上进行训练和学习, 得到图像语义分割模型, 可直接用该模型预测非结构化道路中的可行驶区域, 实现自动驾驶车辆在非结构道路中行驶时的环境感知。实验结果表明, 研究方法分割效果和精确度提升明显, Dice 相似度和 Jaccard 相似系数均可达 80% 以上。

关键词: 自动驾驶; 非结构化道路; 深度学习; 语义分割

中图分类号: U461

文献标志码: A

文章编号: 1000-582X(2020)03-079-09

The semantic segmentation of driving regions on unstructured road based on segnet architecture

ZHANG Kaihang, JI Jie, JIANG Luo, ZHOU Xianlin

(College of Engineering and Technology, Southwest University, Chongqing 400715, P. R. China)

Abstract: To improve the autonomous vehicle's ability to understand the scene of unstructured road driving regions, a semantic segmentation method of unstructured road for autonomous vehicle based on SegNet architecture is proposed. Deep convolutional encoder-decoder architecture is formed by traditional convolutional neural networks, and it can learn the feature map of unstructured roads automatically. By learning and training in the datasets, image semantic segmentation model can be acquired and used to predict the feasible driving area of unstructured roads directly, which is important for autonomous vehicle's scene understanding. The proposed approach outperforms in precision and segmentation consequent, while Dice coefficient reaches more than 80%.

Keywords: autonomous driving; unstructured road; deep learning; semantic segmentation

为实现智能车辆的自动驾驶功能, 需要将道路分割为可行驶区域和非行驶区域, 从而为驾驶行为决策以及路径规划与控制提供基础道路信息。近年来, 很多学者及科研机构已经在结构单一、纹理显著、表面平滑

收稿日期: 2019-07-21

基金项目: 国家自然科学基金青年基金项目(61304189); 中央高校基本业务费专项资金重点项目(XDJK2015B028); 重庆市工业和信息化重点实验室 2019 年度开放课题(19AKC8)。

Supported by National Natural Science Foundation of China (61304189), Fundamental Research Funds for the Central Universities (XDJK2019B053), Open Foundation of Chongqing Key Laboratory (19AKC8)

作者简介: 张凯航(1997—), 男, 西南大学硕士, 主要从事自动驾驶场景识别研究。

通讯作者: 冀杰, 男, 西南大学副教授, 主要从事智能车辆环境感知、驾驶行为决策与运动综合控制方向研究; (E-mail) jijie@swu.edu.cn。

以及边缘线明显的结构化道路中实现了道路区域分割^[1-3]。然而,在乡村或者山区等道路场景差异较大的非结构道路中,可行驶道路区域的检测与分割依然是一项具有挑战性的任务。这是由于非结构化道路具有以下特点:

- 1) 缺乏易于识别的车道线,没有明晰的道路边界,道路形状不规则,难以识别;
- 2) 路面的等级较低,道路周边环境较为原始且差异性显著。

语义分割也叫做图片标注,是指将图片中的每个像素进行准确分类的图像分割技术。传统方法针对非结构道路的语义分割任务,大多集中于通过机器学习的方法实现。文献[4]使用 Gabor 算子提取路面纹理特征,利用得到的方向特征值找出灭点,划分出可行驶区域。文献[5]首先生成道路 RGB 图像的熵图像并计算出熵图像直方图的最小差值,然后以此差值作为阈值分割道路图像,并使用区域生长方法提取出道路可行驶区域。文献[6]采用基于小波变换的方法提取图像边缘,结合高斯函数设定阈值去除非道路边缘点,使用基于斜率和截距的 K-means 聚类算法实现道路拟合。以上方法在特定场景下能够实现道路区域分割,但容易受到外部因素的影响,如颜色特征受光照影响强烈,纹理特征在泥泞路面变得难以提取,这些变化使得上述方法难以满足全场景自动驾驶的鲁棒性要求。

近年来,在场景识别、物体检测、图像深度感知等任务中,基于深度神经网络的方法都有着优异的表现。在 ImageNet^[7] 大规模视觉识别挑战赛中,不断涌现的深度学习网络结构如 AlexNet^[8]、VGGNet^[9]、GoogleNet^[10]和 ResNet^[11]等,使得 1000 类别的图像分类任务 Top-5 错误率降低至 5%。随后这些优秀的深度学习网络结构逐步运用在自动驾驶车辆的环境感知中,采用对预训练好的深度学习模型进行微调(fine-tuning)的方式,实现道路场景的语义分割。

普通的深度神经网络结构(CNN,convolutional neural networks),主要由卷积层、池化层、全连接层和非线性激活单元组成,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。

在此基础上,加州伯克利大学的 Long J 等人提出了全卷积神经网络^[12](FCN,fully convolutional networks),将 CNN 结构末端的全连接层用 1×1 的卷积层替代,得到低分辨率的预测结果,之后采用上采样的方式,将分辨率低的粗略图像恢复到原图的分辨率,首次实现了端到端的语义分割。后来的学者基于 FCN 的思想提出了更加优秀的网络结构,文献[13]提出了在整个网络的后端加入条件随机场(CRF,conditional random fields),以细化粗糙的卷积输出。文献[14]提出了空洞卷积理论,通过扩大卷积核,提取到更多的语义信息。文献[15]中首次加入了基于随机森林算法构成的卷积核,提高了语义分割的准确率。基于不断改进的 FCN 语义分割方法^[16-17]能够在完成城市道路图像多维特征的提取,实现道路的语义分割。然而,FCN 网络结构存在一些不足,采样会造成细节信息的丢失,并且没有充分考虑像素与像素之间的关系,会导致分割结果缺乏空间一致性。

SegNet^[18]与上述网络结构相比,具有网络识别精度高,结构轻巧,模型参数量小,迭代速度快的优点。因此,在该网络基础上,对参数进行修改和调整,得到适用于非结构道路的可行驶区域检测模型。

1 深度神经网络结构

SegNet 是一种以深度卷积为基础,融合编码-解码结构(encoder-decoder)的深度学习网络。编码器网络和解码器网络的对称结构构成了 SegNet 的主要部分,除此之外还有一些输出层。编码器网络由用于图像分类的 VGG16 网络的前 13 层组成,对应图 1 对称结构的左半部分,与完整的 VGG16 网络相比减少了 3 层,这是由于整个网络结构移除了用在特征提取层之间的全连接层,支持在编码器的深层网络输出中保留更高分辨率的特征图,同时大幅度减少训练时参数的数量,参数个数由原来的 1.3×10^9 个降低到 1.400×10^4 个。编码器网络由 5 个层组构成,对应图 1 中对称部分的右半部分,每个层组都包含卷积层,批量归一化^[19](BN, batch normalization)处理层,线性整流函数(ReLu,rectified linear unit)激活层,池化层。图像每经过一个编码器层组,图像的尺寸缩小为原来的一半,依次历遍所有层组以实现有效的数据压缩和特征提取。解码器网络可以看作编码器网络的反向过程,解码器层组构成与编码器层组类似,区别在于进行卷积和池化时,解码

器网络使用保存的最大池化索引进行上采样,得到稀疏的特征图,而后再将特征图与可训练的解码滤波器族卷积得到致密的特征图,最后将高维的特征图输入 Softmax 层,得到每个像素属于道路的概率。这样设计的编码-解码深度神经网络易于学习,可以接受任意大小尺寸的图片输入,在对缩小的图片进行还原的同时能最大限度的保留像素的位置信息,降低错误的分类,适合用于非结构场景中的道路语义分割任务。整个深度神经网络的结构和配置如图 1 和表 1 所示。

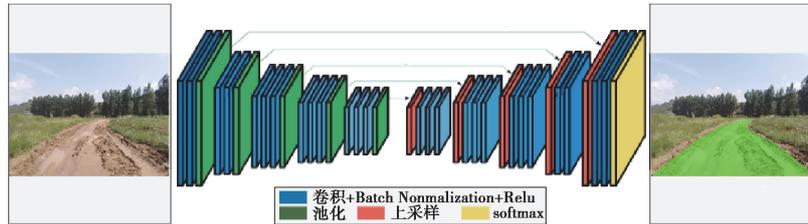


图 1 编码-解码网络结构

Fig. 1 The encoder-decoder network architecture

表 1 网络配置

Table 1 The configuration of network

项	卷积层	卷积核大小/ 特征图数量	输出图像尺寸 ($H \times W \times C$)
输入图像			$512 \times 384 \times 3$
编码器结构	第一层	$3 \times 3 / 32$	$256 \times 192 \times 32$
	第二层	$3 \times 3 / 64$	$128 \times 96 \times 64$
	第三层	$3 \times 3 / 128$	$56 \times 48 \times 128$
	第四层	$3 \times 3 / 256$	$28 \times 24 \times 256$
	第五层	$3 \times 3 / 512$	$14 \times 12 \times 512$
解码器结构	第六层	$3 \times 3 / 256$	$28 \times 24 \times 256$
	第七层	$3 \times 3 / 128$	$56 \times 48 \times 128$
	第八层	$3 \times 3 / 64$	$128 \times 96 \times 64$
	第九层	$3 \times 3 / 32$	$256 \times 192 \times 32$
	第十层	$3 \times 3 / 16$	$512 \times 384 \times 3$
分类	Softmax 层		

1.1 编码器网络算法

以编码器网络层组的第一层组为例,编码器接受大小为 $X1 \times H1 \times M1$ 的输入 $F1$, 经过归一化处理和线性整流函数的作用,进入卷积层执行卷积操作,卷积完成之后紧接着进行最大池化处理,依次循环,直至遍历所有层组。定义编码器层数为 L ,具体算法如图 2:

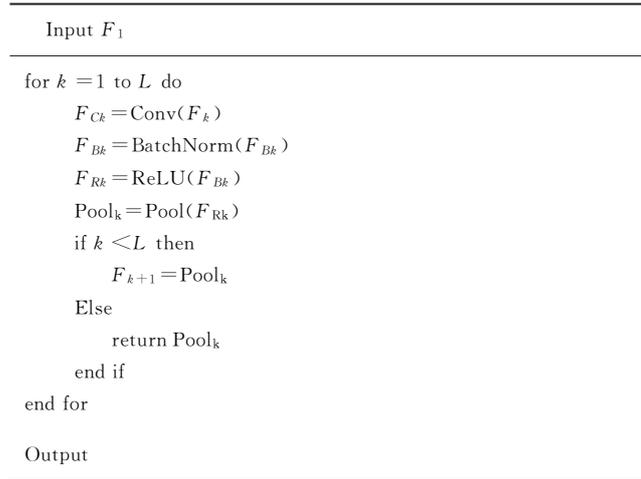


图 2 编码器网络算法

Fig. 2 The encoder network algorithm

该算法完成编码器的相关运算,提取图片中非结构道路的特征,输出大小为 $X_2 \times H_2 \times M_2$ 的特征图,同时该结果作为解码器网络的输入。

1.2 解码器网络算法

同样以解码器网络的第一层组为例,解码器接受编码器网络最后一层的输出 Pool_L 和保存的最大池化索引 $\text{Indice}[p]$ 作为输入,算法和解码器网络算法类似。具体算法如图 3 所示:

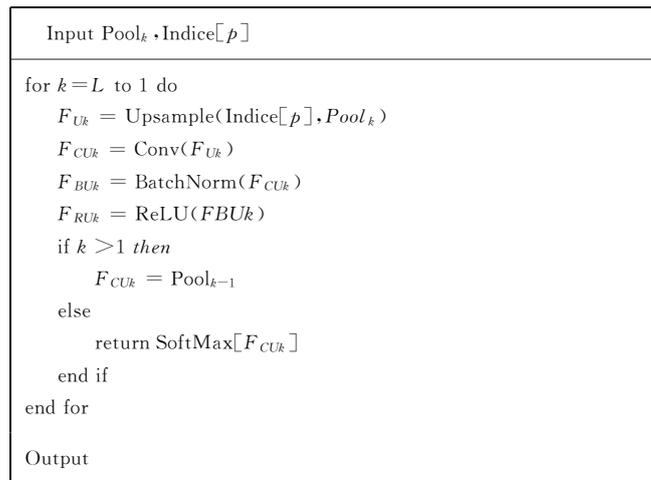


图 3 解码器网络算法

Fig. 3 The decoder network algorithm

输出结果即为语义分割的最终结果。

1.3 批量归一化处理算法

批量归一化处理,使得输入下一层网络的数据具有和原始输入数据相似的分布,加快深度卷积神经网络达到收敛的速度,保证整个神经网络的表征能力。定义批尺寸大小为 m ,批尺寸内包含的数据为 $B = \{x_1, x_2, \dots, x_m\}$,整个处理过程可以看作对 $x_1 \dots m$ 做线性变换得到 $y_1 \dots m$,即

$$BN_{\gamma, \beta}: x_{1 \dots m} \rightarrow y_{1 \dots m}, \quad (1)$$

作为下一层输入的 y 由以下方法得出

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i, \quad (2)$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2, \quad (3)$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}, \quad (4)$$

$$y_i = \gamma \hat{x}_i + \beta, \quad (5)$$

式中: μ_B, σ_B^2 为批尺寸中数据的标准差和方差; \hat{x}_i 为归一化后的值, 其中 ϵ 是一个常数(趋于 0), 用于保证计算方差时的稳定性; γ 和 β 对归一化后的值进行缩放和平移, 是批量归一化处理中需要学习的参数。

1.4 损失函数

研究选取了如下形式的损失函数: 定义 $Y = \{y_1, y_2, \dots, y_n\}$ 是图片的每个像素点 $D = \{x_1, x_2, \dots, x_n\}$ 所对应的标签类别, 其中 $y_i = 1$ 代表像素属于道路, $y_i = 0$ 代表像素属于非道路(其中 $0 < i < N$, N 为图片中像素的数量), 深度神经网络的输出为每个像素点属于道路的概率, 结果可看作二分类的条件分布模型, 训练过程等同于最大化像素点对应的属于正确分类的概率, 可用交叉熵函数表示整个过程的损失函数

$$l(\theta, D) = - \sum_{i=1}^{|D|} \log(P(\gamma = y^{(i)} | x^{(i)}, \theta)), \quad (6)$$

式中: $P(\gamma = y^{(i)} | x^{(i)})$ 代表输入数据 $x^{(i)}$ 的分类属于其正确类别 $y^{(i)}$ 的概率。 θ 代表整个网络所要学习的权重。 $P(\gamma = y^{(i)} | x^{(i)})$ 可用如下方法计算得出

$$P(\gamma = y^{(i)} | x^{(i)}, \theta) = \frac{1}{1 + \exp f(\theta, x_i)}, \quad (7)$$

式中, $f(\theta, x_i)$ 代表在训练中学习的深层特征。选择随机梯度下降算法对网络进行优化, 对于不同卷积层的损失函数的导数可以用链式法则来计算。损失函数关于 W_k 的偏导数可以表示为

$$\frac{\partial l}{\partial W_k} = - \sum_{i=1}^{|D|} f(\theta, x_i) (1\{y^{(i)} = 1\} - P(\gamma = y^{(i)} | x^{(i)}, \theta)), \quad (8)$$

其中, $1\{y^{(i)} = 1\}$ 为示性函数, 当 $y^{(i)} = 1$ 时取 1, 否则取 0。通过进行反向传播, 在训练中不断更新网络的权重。

2 数据集

在深度神经网络的参数训练过程中, 采用具有像素级别标注的语义分割数据集, 能够使道路识别模型取得更好的效果。因此, 选取了专门为非结构化道路制作的开源数据集。该数据集由驾驶员操纵一台顶部安装有单目摄像机的汽车, 在实际的非结构化道路场景中匀速行驶拍摄而来, 摄像机以 20 帧/s 的速率采集了约 6 000 张图片, 图片的像素尺寸为 512×384 。经过筛选, 剔除其中不合格的图片, 对满足数据集要求的合格图片进行标注, 将其划分为路面和非路面两种类别, 路面用绿色的像素表示, 非路面则不做修改。按 6:2:2 比例, 将标注好的数据集分为训练集、验证集、测试集 3 个部分。训练集用来训练模型, 验证集用来确定与网络结构和模型复杂程度有关的参数, 测试集用来对已经训练好的模型进行性能评估。

实际场景中搜集来的图片往往不能满足数据在多样性方面的要求, 单纯依靠未经数据预处理的数据集训练模型, 学习到的特征不够丰富, 无法实现更普遍的路面识别。因此, 需要对有限的图像数据集进行数据预处理, 扩充数据集的深度和广度, 提升深度神经网络的鲁棒性。选取了一些较为常见的预处理措施, 在原有图片的基础上改变图像的几何学特征和纹理结构, 为训练出更好的语义分割模型奠定基础。这些措施包括:

1) 对图片进行水平镜像翻转, 增加整个网络在方向上的不变性;

2) 物体在图像中的清晰程度取决于场景的照明和相机的灵敏度。随机增加或减小图像亮度, 能够提高对环境光照情况变化的适应能力;

3) 在每个 RGB 像素中添加小的随机噪声, 如高斯噪声, 有助于获得对某些相机失真的不变性。

经过预处理的图片如图 4 所示。



(a) 图表示对图片进行水平翻转 (b) 图表示添加高斯噪声 (c) 减小亮度 (d) 增加亮度

图 4 预处理结果

Fig. 4 The results of preprocessed data

3 实验结果

训练过程在 Ubuntu16.04+cuda8.0+cudnn6.0 系统环境下进行,显卡配置 GTX1080ti。使用 Keras 编写神经网络程序。用 Xavier 初始化权重,采用随机梯度下降算法进行优化,初始学习率设定为 10^{-4} ,之后衰减到 10^{-6} ,批尺寸大小为 128,为了防止模型过拟合,在神经网络结构中加入 dropout 层,将 dropout rate 设定为 0.5,并添加 L2 正则化($\lambda=0.0001$)。

在 60 000 次迭代完成后,损失函数的值不再变化,对训练好的模型进行测试。除了在测试集验证模型效果外,还选取了不同于测试集的真实场景做测试,并与基于文献[20]的方法得出的在 KITTI 数据集^[21]中对城市道路图像进行语义分割的结果做比较。具体实验结果如图 3 所示,(a)图的第一二组为在测试集进行测试得到的语义分割结果,第二组图片降低了图片亮度,用于检验模型在光照不足的情况下是否仍然能够进行语义分割,(b)图第三、四组图片是对测试集以外的真实场景中实际采集图片进行语义分割的结果,第四组同样采取了减小亮度的措施,最后一组(c)图作为对比,是对 KITTI 城市道路数据集进行语义分割的结果。图片中使用绿色像素表示语义分割结果中的实际可行驶区域,可以直观的看出,方法在对非结构道路进行可行驶区域检测时,在数据集图片数量较少的情况下,训练出的模型仍然可以很好地识别路面,将非道路识别为道路的错误分类很少,语义分割的均匀性和一致性较好。

除了使用目测法直接观察分割效果外,还选择了语义分割中常用的 2 个评价指标,Dice 相似度和 Jaccard 相似系数(也叫做重叠度),用数值精确的检验神经网络对非结构道路进行语义分割的效果,Dice 相似度也被称为 F1 度量,由精确度和召回率 2 个参数决定。精确度和召回率的定义如下

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (9)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (10)$$

式中, TP , FP 和 FN 分别代表着预测为道路的道路样本,预测为道路的非道路样本,预测为非道路的道路样本。Dice 相似度的定义如下

$$\text{Dice} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (11)$$

第二个指标 Jaccard 相似系数经常使用在计算机视觉领域中,用于描述集合之间的相似度。其定义如下

$$\text{Jaccard} = \frac{TP}{TP + FP + FN}, \quad (12)$$

Dice 相似度和 Jaccard 相似系数能较为一致的反映语义分割效果的好坏,不同之处在于 Jaccard 相似系数对于错误的分割类型更为敏感,惩罚力度更大。2 个指标数值越高,代表语义分割效果越好。在测试集中使用所有图片的分割结果计算这 2 个指标,如图 5 所示并与在城市道路集 KITTI 数据集上用神经网络

实现的道路语义分割结果进行比较。结果如表 2 所示,训练出的模型对于非结构化道路的分割结果与城市道路分割结果相比,二者在 Dice 相似度和 Jaccard 相似系数上差距不明显,这反映出预测结果和真实标定的路面基本接近。高精度度表示分类正确的像素占总像素的比例较大,分割效果好。高召回率意味着模型具有很好的鲁棒性,在环境和光照条件变化时仍然能够精确地完成语义分割任务。

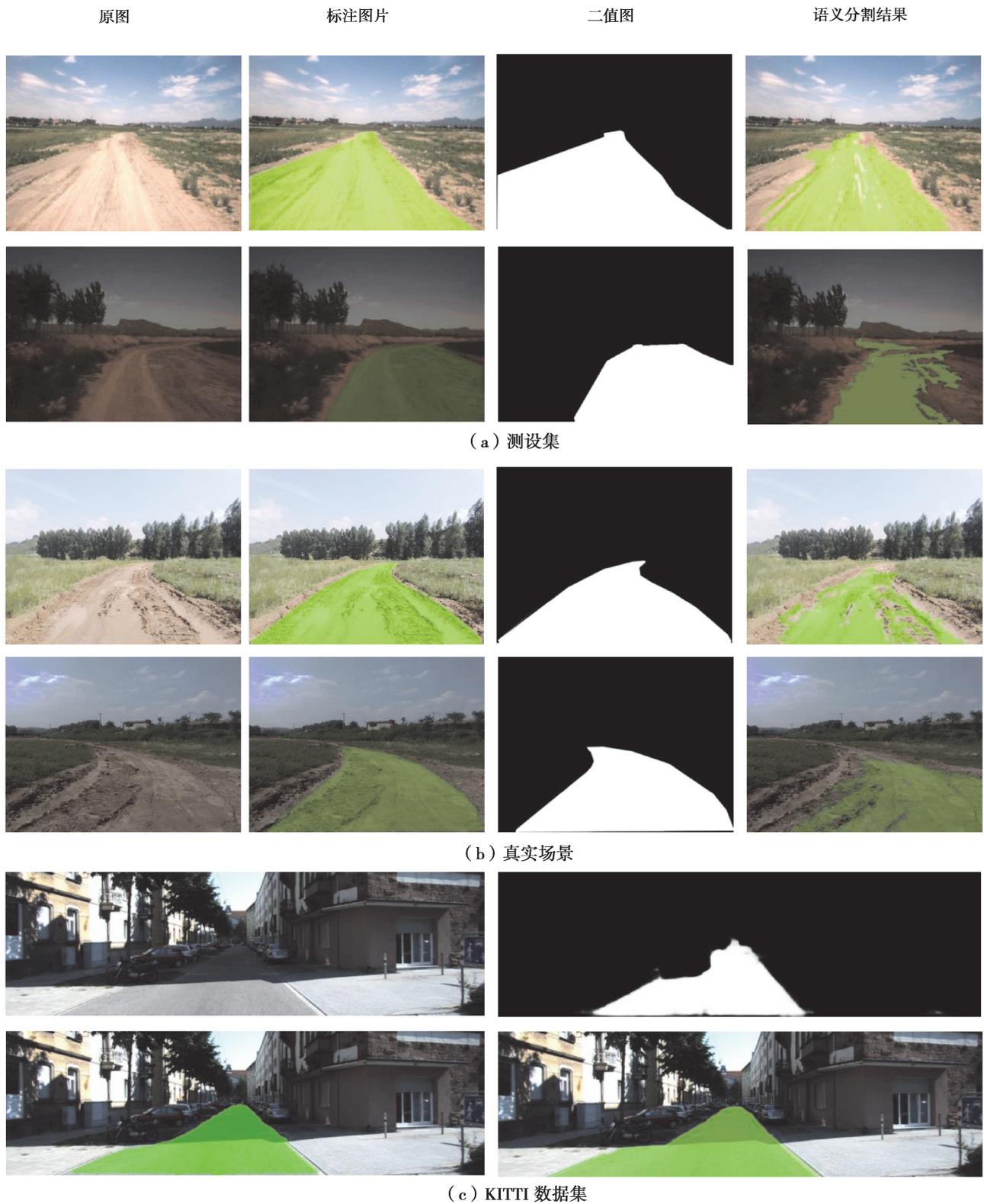


图 5 实验结果

Fig 5 The experimental results

表 2 结果对比

Table 2 Comparison of results

参数	Dice 相似度/%	Jaccard 相似系数/%	精确度/%	召回率/%
非结构化道路	0.865	0.832	0.82	0.75
结构化道路	0.920	0.915	0.91	0.93

图 5 的结果在识别弯道路面时出现了误差,这是因为数据采集过程中,车辆的行驶路面多为平直道路和弯道半径较小的道路,弯道路面的样本数不足,学习的特征不充分,造成少部分像素点的错误匹配。同时,该深度学习语义分割算法网络的结果要与图像中的每一个像素点相匹配,因此会造成分割结果边缘不光滑的问题,影响分割精度。

4 结 论

针对基于视觉的非结构化道路可行驶区域检测,在 SegNet 的基础上提出了一种更加有效的编码—解码结构的深度神经网络。该网络不需要人工设计特征,通过深度神经网络自己学习特征,实现了对道路图像的语义分割,并且当环境发生变化时,仍然能够很好的识别路面。表明该深度学习模型的深层架构具有从已有图片中获得有意义的特征并将其映射成提前设置好的语义标签能力。如果能进一步提高数据集中的图片数量,并且针对性使用不同的环境图像训练模型,如沙石路面和潮湿路面等,将会取得更好的语义分割结果。这将为以后进一步提升自动驾驶车辆的环境感知能力奠定良好的基础。

参考文献:

- [1] 王科, 黄智, 钟志华. 基于不定 Bezier 变形模板的城市道路检测算法[J]. 机械工程学报, 2013, 49(8): 143-150.
WANG Ke, HUANG Zhi, ZHONG Zhihua. Algorithm for urban road detection based on uncertain bezier deformable template[J]. Journal of Mechanical Engineering, 2013, 49(8): 143-150. (in Chinese)
- [2] 曹婷, 王欢. 基于二值空间线特征的道路检测方法[J]. 计算机工程与应用, 2018, 54(6): 161-167.
CAO Ting, WANG Huan. Road detection based on binary spatial ray feature[J]. Computer Engineering and Applications, 2018, 54(6): 161-167. (in Chinese)
- [3] 李晓航, 郭佳, 彭富伦, 等. 一种改进的基于 Hough 变换的道路图像检测方法[J]. 应用光学, 2016, 37(2): 229-234.
LI Xiaohang, GUO Jia, PENG Fulun, et al. Improved algorithm of road detection based on Hough transform[J]. Journal of Applied Optics, 2016, 37(2): 229-234. (in Chinese)
- [4] 刘富, 袁雨桐, 李洋. 基于纹理特征的非结构化道路分割算法[J]. 计算机应用, 2015, 35(S2): 271-273.
LIU Fu, YUAN Yutong, LI Yang. Segmentation algorithm of unstructured road based on texture feature[J]. Journal of Computer Applications, 2015, 35(S2): 271-273. (in Chinese)
- [5] 吴骅跃, 段里仁. 基于 RGB 熵和改进区域生长的非结构化道路识别方法[J]. 吉林大学学报, 2019, 3: 727-735.
WU Huaye, DUAN Liren. Unstructured road detection method based on RGB entropy and improved region growing [J]. Journal of Jilin University, 2019, 3: 727-735.
- [6] 熊思, 李磊民, 黄玉清. 基于小波变换和 K-means 的非结构化道路检测[J]. 计算机工程, 2014, 40(2): 158-161.
XIONG Si, LI Leimin, HUANG Yuqing. Unstructured road detection based on wavelet transform and K-means[J]. Computer Engineering, 2014, 40(2): 158-161. (in Chinese)
- [7] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database[C/OL]. 2009 IEEE Conference on Computer Vision and Pattern Recognition, New York, USA: IEEE, 2009 (2009-08-18) [2019-09-25]. <https://ieeexplore.ieee.org/document/5206848>.
- [8] Alex K, Ilya S, Geoffrey E H. Image net classification with deep convolutional neural networks[J]. Neural Information Processing Systems, 2012, 141(5): 1097-1105.
- [9] Karen S, Andrew Z. Very deep convolutional networks for large-scale image recognition[J/OL]. Computer Science, 2014 [2019-09-25]. https://www.researchgate.net/publication/265385906_Very_Deep_Convolutional_Networks_for_Large

Scale_Image_Recognition.

- [10] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C/OL]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York, USA: IEEE, 2015 (2015-10-15) [2019-09-25]. <https://ieeexplore.ieee.org/document/7298594>.
- [11] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C/OL]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York, USA: IEEE, 2016 (2016-12-12) [2019-09-25]. <https://ieeexplore.ieee.org/document/7780459>.
- [12] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C/OL]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York, USA: IEEE, 2015 (2015-10-15) [2019-09-25]. <https://ieeexplore.ieee.org/document/7298965>.
- [13] 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, 2018, 40(4): 834-848.
- [14] Fisher Yu, Vladlen Koltun. Multi-scale context aggregation by dilated convolutions[J/OL]. arXiv: Computer Vision and Pattern Recognition, 2015 [2019-09-25]. <https://arxiv.org/abs/1511.07122>.
- [15] Ravi D, Bober M, Farinella G M, et al. Semantic segmentation of images exploiting DCT based features and random forest [J]. Pattern Recognition, 2016, 52: 260-273.
- [16] Ranft B, Stiller C. The role of machine vision for intelligent vehicles[J]. IEEE Transactions on Intelligent Vehicles, 2016, 1(1): 8-19.
- [17] Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation[C/OL]. 2015 IEEE International Conference on Computer Vision (ICCV). New York, USA: IEEE, 2015 (2015-02-18) [2019-09-25]. <https://ieeexplore.ieee.org/document/7410535>.
- [18] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
- [19] Sergey Ioffe, Christian Szegedy. Batch normalization: accelerating deep network training by reducing Internal covariate shift[J/OL]. arXiv: Learning, 2015 [2019-09-25]. <https://arxiv.org/abs/1502.03167>.
- [20] Teichmann M, Weber M, Zöllner M, et al. MultiNet: real-time joint semantic reasoning for autonomous driving[C]// 2018 IEEE Intelligent Vehicles Symposium (IV). New York, USA: IEEE, 2018: 1013-1020.
- [21] Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: The KITTI dataset[J]. International Journal of Robotics Research, 2013, 32(11): 1231-1237.

(编辑 侯 湘)