基于卷积神经网络的交通场景语义分割方法研究

李琳辉，钱波，连静，郑伟娜，周雅夫
（大连理工大学工业装备结构分析国家重点实验室运载工程与力学学部汽车工程学院，辽宁 大连 116024）

摘 要：为提高交通场景的语义分割精度，提出一种基于 RGB-D 图像和卷积神经网络的分割方法。首先，基于半全局立体匹配算法获取视差图 D，并将其与 RGB 图像融合成四通道 RGB-D 图像，以建立样本库；其次，对于 2 种不同结构的卷积神经网络，分别采用 2 种不同的学习率调整策略对网络进行训练；最后，对训练得到的网络进行测试及对比分析。实验结果表明，基于 RGB-D 图像的交通场景语义分割算法得到的分割精度高于基于 RGB 图像的分割算法。

关键词：深度学习；卷积神经网络；交通场景；语义分割；视差图
中图分类号：U495，TP391.4
文献标识码：A
doi: 10.11959/j.issn.1000-436x.2018053

Study on traffic scene semantic segmentation method based on convolutional neural network

LI Linhui, QIAN Bo, LIAN Jing, ZHENG Weina, ZHOU Yafu
School of Automotive Engineering, Faculty of Vehicle Engineering and Mechanics,
State Key Laboratory of Structural Analysis for Industrial Equipment, Dalian University of Technology, Dalian 116024, China

Abstract: In order to improve the semantic segmentation accuracy of traffic scene, a segmentation method was proposed based on RGB-D image and convolutional neural network. Firstly, on the basis of semi-global stereo matching algorithm, the disparity map was obtained, and the sample library was established by fusing the disparity map D and RGB image into the four-channel RGB-D image. Then, with two different structures, the networks were trained by using two different learning rate adjustment strategy respectively. Finally, the traffic scene semantic segmentation test was carried out with RGB-D image as the input, and the results were compared with the segmentation method based on RGB image. The experimental results show that the proposed traffic scene segmentation algorithm based on RGB-D image can achieve higher semantic segmentation accuracy than that based on RGB image.

Key words: deep learning, convolutional neural network, traffic scene, semantic segmentation, disparity map

1 引言

无人驾驶是汽车领域的研究热点之一，提高无人驾驶系统智能化程度的关键技术之一是具备对交通场景准确有效的认知。

目前，比较成熟的交通场景分类主要针对环境中的特定目标进行识别，多数属于二分类范畴，如路面识别[1]、车辆识别、行人识别等，用到的方法主要为浅层学习方法，如支持向量机、AdaBoost 等。近几年，深度学习[2]的研究取得了突破性进展，并被广泛应用于图像领域。使用深度学习方法能够较好地解决多分类问题，特别适用于复杂的自然数

收稿日期：2017-07-03；**修回日期**：2018-03-16
通信作者：连静，lianjing@dlut.edu.cn
基金项目：国家自然科学基金资助项目（No.51775082, No.61473057, No.61203171）；中央高校基本科研业务费专项基金资助项目（No.DUT15LK13, No.DUT17LAB11）
Foundation Items: The National Natural Science Foundation of China (No.51775082, No.61473057, No.61203171), The Central University Basic Business Expenses Special Funding for Scientific Research Projects (No.DUT15LK13, No. DUT17LAB11)

据，包括交通场景图像数据。随着 GPU 并行计算的发展，使用深度学习方法造成计算量大的问题得到解决，从而使面向复杂交通环境的像素级别场景分割成为可能。

深度学习已被验证能够提高目标识别[3]和图像语义分割[4,5]的精度，具有代表性的深度网络包括 AlexNet[6]、VGGNet[7]和 GoogLeNet[8]等，它们在图像的单标签分类问题上取得了较好的成绩，对 1 000 类图像分类的 Top-5 错误率均在 8%以内，是近年来 ImageNet[9]图像分类大赛的主要解决方案。在此基础上，针对图像的语义分割问题，Long 等[10]提出了一种基于全卷积网络（FCN）的语义分割方法，对目前的图像分类网络进行了修改，将全连接层改为卷积层，使其学习到的特征适用于图像的语义分割任务；针对更为复杂的室外交通场景，Badrinarayanan 等[11,12]提出了一种卷积神经网络，用来实现交通场景图像的语义分割，通过最大非线性上采样方法得到与输入图像分辨率相同的语义分割结果；Noh 等[13]利用反卷积和上采样方法实现了图像的语义分割任务，在细小物体的语义分割问题上取得了较好的结果。

随着面扫描激光、立体视觉、红外体视等深度传感器的发展，获取图像的深度信息变得越来越容易，基于 RGB-D 图像的语义分割研究成为未来的发展趋势之一。目前，RGB-D 数据集主要用于室内场景的语义分割，例如，Silberman 等[14]制作了 RGB-D 室内场景数据集 NYUv2，考虑到物体之间的支撑关系，提出了基于 RGB-D 图像的室内场景语义分割算法；Gupta 等[15]在室内物体检测算法的基础上提出了基于 RGB-D 的室内场景语义分割算法。相关研究[16,17]表明，基于 RGB-D 图像的室内场景分割相比RGB图像具有更高的分类准确度和环境适应性，可以为基于RGB-D图像的室外场景分割提供借鉴。对室外的交通环境而言，场景复杂多变且需要获取更远距离的深度信息，相应地，也急需更为有效的深度信息获取方法及深度学习方法。

基于以上分析，本文从视差图获取和深度学习 2 个角度入手，提出一种基于 RGB-D 卷积神经网络的交通场景语义分割方法。首先，研究一种基于扫描线最优的半全局立体匹配算法，并通过快速全局图像平滑方法获取连续的视差图；然后，从 KITTI[18]的 Stereo2012 双目视觉数据集中选取具有代表性的交通场景图像，通过上述立体匹配算法获取对应的视差图 D，将左图 RGB 图像和对应的视差图 D 融合成四通道 RGB-D 图像，并将物体分为 7 个类别：天空、建筑、路面、路边界、树木、草坪、车辆，以左图 RGB 图像作为样本对每个像素所属类别进行标注；最后，使用 RGB-D 四通道图像对 2 种不同结构的卷积神经网络进行训练和测试，并与基于 RGB 三通道图像方法的测试结果进行对比分析，结果表明使用 RGB-D 四通道图像训练得到的网络在交通场景的语义分割任务上能够获得更高的分割精度。

2 基于立体视觉的视差图获取

通过立体视觉的立体匹配步骤，可以获得包含所拍摄场景三维信息的视差图，视差图的精度越高，卷积神经网络从视差图中能够提取到的物体特征信息越丰富。因此，立体匹配的精度直接影响着语义分割精度。

立体视觉匹配算法可分为 3 类：局部匹配算法、半全局匹配算法、全局匹配算法，这 3 种匹配算法的匹配精度依次增高，但匹配所消耗的时间也依次增高。考虑到算法的实时性要求，且半全局匹配算法的精度接近于全局匹配算法，本文通过半全局匹配算法[19]来计算获取视差图，并通过一种基于最小二乘法的快速全局图像平滑方法[20]获取视差值更加连续的视差图，算法的基本步骤如下。

1) 采用基于窗口的局部算法计算单个像素点的灰度相似性匹配代价。

2) 通过多个方向扫描线上基于平滑约束的方法对匹配代价进行聚合，建立一个全局的能量函数。

3) 采用胜者为王方法选取使能量函数最小的视差值，并通过二次曲线拟合估计亚像素级别的视差。

4) 分别根据左右视图生成的视差图剔除异常点，使其符合一致性约束，消除遮挡带来的误匹配。

5) 采用基于最小二乘法的快速全局图像平滑方法对视差图进行滤波处理，填充没有匹配到的像素点，获得视差值更为连续、更为平滑的视差图。

步骤 1) 中的基于窗口的局部算法采用 5×5 的窗口，灰度相似性采用灰度差的绝对值叠加方法计算。步骤 2) 中采用了扫描线最优算法的思想，沿 8 路不同方向的扫描线分别计算匹配代价，采用多个方向的一维平滑约束合并的方式来逼近图像平面内二维的平滑约束。步骤 5) 是获取视差图的关键，通过前面的步骤得到的视差图比较粗糙，如图 1(b)所

示，包含一些未匹配的像素点，且物体边界较为粗糙，通过步骤 5)，可以得到更为平滑的视差图，如图 1(c)所示，较好地保留物体的边缘、轮廓信息。

图 1 为 KITTI 数据集中一对立体视觉图像的匹配结果。在视差图中，灰度值越大的点对应的视差值越大，即越亮的点距离相机越近。其中，图 1(b)为未经平滑处理的视差图，可以看出，其中存在一些未匹配的像素点，物体的边缘信息比较模糊，图 1(c)为平滑处理后的视差图，可以看出，经平滑处理后的视差图较好地保留了物体的边缘、轮廓信息，为基于 RGB-D 图像的交通场景语义分割奠定了基础。

(a) 原图

(b) 未平滑处理的视差图

(c) 平滑处理后的视差图

图 1　立体视觉图像的匹配结果

3　交通场景语义分割

3.1　RGB-D 样本库的建立

在具有代表性的交通场景数据集中[18,21,22]，KITTI 是目前最大的道路场景数据集，其中包含了通过立体视觉相机拍摄的交通场景图像，场景中包括天空、路面、树木、车辆等多种类别，便于立体匹配算法验证及网络训练。

首先，从 KITTI 的 Stereo2012 子数据集中选取具有代表性的交通场景立体图像，并将交通场景分为 7 个类别：天空、建筑、路面、路边界、树木、草坪、车辆，类别的标签从 0 到 6，其他类别的标签为 7，不参与反向传播时权值的更新计算。以立体视觉图像中的左侧 RGB 图像为样本，对图像的每个像素所属类别进行标注，将其作为训练的标签。然后，基于前述立体匹配算法，获取左右图像对应的视差图 D。最后，将左图 RGB 图像和视差图 D 融合成四通道 RGB-D 图像。最终建立的样本库包含训练集、验证集和测试集。

3.2　网络训练

本文基于 SegNet[12]和 SegNet-Basic[11]网络来实现交通场景 RGB-D 图像的语义分割。SegNet 和 SegNet-Basic 具有不同的网络结构，其中，SegNet 包含 26 个卷积层、5 个下采样层和 5 个上采样层，SegNet-Basic 包含 8 个卷积层、4 个下采样层和 4 个上采样层。这 2 种网络架构均能够进行端到端的训练，且相对于其他网络架构[10]，在交通场景的语义分割上，SegNet 和 SegNet-Basic 能够获得较高的语义分割精度，且使用训练好的模型进行语义分割测试的实时性较好。

采用小批量训练的方法进行网络的训练，每次选取一定数量的样本图像送入网络进行前向传播，得到每个像素点的输出误差，然后计算该小批量样本图像上所有像素点的输出误差和的平均值，作为网络的输出误差，即训练误差，并根据最小化训练误差的方法来更新网络的权值参数。其中，采用交叉熵损失函数[10]来计算网络的训练误差，计算式为

$$P(x=k)=\frac{\exp(a_k)}{\sum_i \exp(a_i)},\ i=0,1,\cdots,K-1 \qquad (1)$$

$$L=-\frac{1}{N}\sum_i \ln\left[P\left(x=k\right)\right],\ i=0,1,\cdots,N-1 \qquad (2)$$

其中，$P(x=k)$ 为像素点 x 属于其类别 k 的概率，a_i 为第 i 个类别的特征值，由最后一层卷积层得到，K 为分类的类别数量，N 为一个批量上所有像素点的数量，L 为网络最终输出的训练误差值。由于在训练集上各个类别所占的像素数量相差较大，如天空、路面等像素点所占的像素数量较多，因此，采用中值频率平衡[23]方法来计算不同类别的实际误差值，其计算式为

$$\lambda_i=\frac{m}{n_i},\ i=1,2,\cdots,K \qquad (3)$$

其中，λ_i 为第 i 个类别的误差值权重，n_i 为训练集上第 i 个类别所占像素的数量，m 为各个类别所占像素数量的中值。优化后的训练误差计算式为

$$L=-\frac{1}{N}\sum_i \lambda_i \ln\left[P\left(x=k\right)\right],\ i=0,1,\cdots,N-1 \quad (4)$$

在反向传播更新网络权值参数阶段，采用随机梯度下降法[24]来更新网络的权值参数，其通过负梯度 $\nabla L(\boldsymbol{W})$ 和上一次的权值更新值的线性组合来更新权值，计算式为

$$\boldsymbol{V}_{t+1}=\mu\boldsymbol{V}_t-\alpha\nabla L(\boldsymbol{W}_t) \quad (5)$$

$$\boldsymbol{W}_{t+1}=\boldsymbol{W}_t+\boldsymbol{V}_{t+1} \quad (6)$$

其中，$\boldsymbol{W}_t$ 是第 t 次迭代计算时的权值矩阵，$\boldsymbol{V}_t$ 是第 t 次迭代计算时的权值更新值，α 是负梯度的基础学习率，μ 是权值更新值 $\boldsymbol{V}_t$ 的权重，用来加权之前梯度方向对现在梯度下降方向的影响，这 2 个值一般根据经验设定。通常在迭代计算过程中，需要对基础学习率进行调整，常用的调整策略为 fixed 和 step，使用 fixed 方式时，在迭代计算过程中基础学习率保持不变；使用 step 方式时，实际的基础学习率 β 和 α 之间的关系为

$$\beta=\alpha g^{floor\left[\frac{a}{b}\right]} \quad (7)$$

其中，a 是当前迭代次数，b 为基础学习率更新的步长，g 为基础学习率缩放因子，*floor* 为上取整函数。

为了减少网络训练时陷入局部最小值的概率，验证算法的可扩展性和顽健性，本文使用 fixed 和 step 这 2 种方法对网络进行训练，将 α 设为 0.01，μ 设为 0.9，使用 step 学习策略时，将 b 设为 2 000，g 设为 0.1，即每进行 2 000 次迭代，基础学习率更新为上次的 0.1 倍。

4 实验结果与分析

本文算法的具体实现使用的是深度学习框架 Caffe[25]，网络的训练与测试均在 Caffe 环境下完成。实验的硬件环境为 Intel Xeon E5-2620 中央处理器，NVIDIA TITAN X 显卡；软件环境为 Ubuntu 14.04 LTS 操作系统，cuda7.5，cudnn v2。该配置是目前深度学习计算的主流配置。网络训练及权值调整流程如图 2 所示，具体步骤如下。

1) 初始化网络权值参数。

2) 读取训练图片数据，进行网络的前向传播，并输出在训练数据上的误差。

3) 判断是否达到训练次数，如果未达到训练次数，根据得到的误差进行网络权值梯度的计算，并进行反向传播更新网络的权值参数，执行步骤 2)；如果达到训练次数，则停止训练。

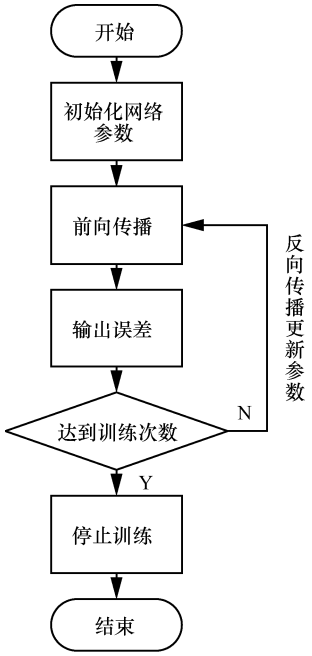


图 2 网络训练及权值调整流程

在语义分割网络的训练与测试中，小批量的大小设置为 4，即每次选取 4 张图片进行迭代计算，每 400 次迭代计算后在验证集上验证一次精确率直至训练误差值开始收敛。

语义分割精度通常有 2 种评判方法，即全局精确率和平均精确率。全局精确率是分类正确的像素点占数据集全部像素点的百分比，全局精确率越高，图像分割结果越平滑；平均精确率是所有类别预测精确率的平均值，与样本中每个类别所占像素点的比例有关，其中，每个类别分割精度为分类正确像素点占该类别所有像素点的比例。语义分割的最终目的是获得平滑的语义预测，所以本文选取在验证集上全局精确率最高的一次迭代计算得到每个类别的分割精度作为此次训练的最终结果。

为了对比不同网络、不同学习率策略及不同数据源输入对交通场景语义分割精度的影响，针对 SegNet 和 SegNet-Basic 网络，分别选择 RGB 和 RGB-D 图像，采用 fixed 和 step 这 2 种学习率调整策略对网络进行训练，得到不同类别的分类精度统计如表 1 和表 2 所示。为了得到网络的训练误差和精确率的收敛情况，以使用 fixed 学习率策略时 SegNet 网络训练误差和验证集分割精确率为例，其迭代过程中的变化趋势如图 3 所示。通过分析，可

表 1　采用 fixed 学习率所得语义分割精度

网络模型	图像格式	分割精度							平均精确率	全局精确率
		天空	建筑	路面	人行道	树木	草坪	车辆		
SegNet	RGB	0.952	0.744	0.964	0.658	0.877	0.562	0.880	0.805	0.858
	RGB-D	0.964	0.778	0.969	0.661	0.869	0.597	0.911	0.821	0.875
SegNet-Basic	RGB	0.941	0.815	0.946	0.677	0.815	0.630	0.856	0.811	0.859
	RGB-D	0.937	0.866	0.956	0.768	0.792	0.709	0.892	0.846	0.870

表 2　采用 step 学习率所得语义分割精度

网络模型	图像格式	分割精度							平均精确率	全局精确率
		天空	建筑	路面	人行道	树木	草坪	车辆		
SegNet	RGB	0.946	0.779	0.941	0.652	0.830	0.533	0.844	0.789	0.842
	RGB-D	0.949	0.809	0.944	0.677	0.812	0.522	0.877	0.799	0.852
SegNet-Basic	RGB	0.943	0.783	0.902	0.700	0.818	0.735	0.865	0.821	0.840
	RGB-D	0.936	0.789	0.915	0.757	0.830	0.718	0.866	0.830	0.855

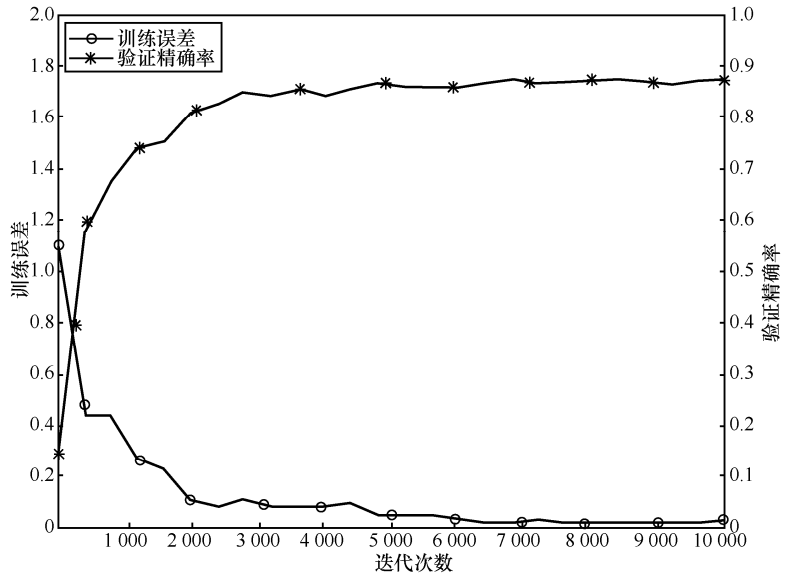


图 3　训练误差和验证精确率曲线

以得到以下结论。

1) 2 种网络均在收敛的基础上取得了良好的分割精度。

2) 对于相同的网络结构，使用 fixed 学习率策略得到的全局精确率高于使用 step 学习率策略时得到的全局精确率。

3) 天空、路面等所占像素点数量较多的类别，分割精度较高；草坪、人行道等所占像素点数量较少的类别，分割精度较低。

以 fixed 学习率策略得到的分割精度为例，将基于 RGB-D 图像和基于 RGB 图像得到的结果进行对比分析，得到以下结论。

1) 针对 RGB 和 RGB-D 图像，SegNet 得到的全局精确率分别为 0.858、0.875，SegNet-Basic 的到的全局精确率分别为 0.859、0.87，因此，使用 RGB-D 图像作为网络输入使 SegNet 和 SegNet-

Basic 网络的全局精确率分别提高了 0.017、0.011，平均精确率分别提高了 0.016、0.035。

2) 在建筑、路面、人行道、草坪、车辆这 5 个类别的语义分割精度上，基于 RGB-D 图像的方法得到的精度均高于基于 RGB 图像的方法，对于 SegNet 网络，以上 5 个类别的分割精度分别提高了 0.034、0.005、0.003、0.035、0.031；对于 SegNet-Basic 网络，以上 5 个类别的分割精度分别提高了 0.051、0.010、0.091、0.079、0.036。

通过以上对比分析可以得到，使用同一种深度网络时，在相同的训练参数下，基于 RGB-D 图像的方法较基于 RGB 图像的方法能够得到较高的全局精确率和平均精确率。

图 4 给出了测试集上部分交通场景图像的视差

(a) 原图

(b) 视差图

(c) 标签

(d) RGB 语义分割结果

(e) RGB-D 语义分割结果

图 4 测试集部分样本的语义分割结果

图和语义分割结果，其中，图 4(c)为图像对应的标签，作为对比基准来评定图像的语义分割效果。图 4(d)为使用 RGB 三通道图像作为网络输入时得到的语义分割结果，与图 4(c)中的标签相比，其语义分割结果存在相对较大的噪声输出。图 4(e)为使用 RGB-D 四通道图像作为网络输入时得到的语义分割结果，通过将图 4(d)与图 4(e)进行对比可以看出，图 4(e)中的语义分割结果噪声较小，更加接近图 4(c)中图像的标签。这说明视差图 D 的引入在一定程度上减少了分类噪声，能够得到更加平滑的语义分割结果。

5　结束语

本文提出一种基于卷积神经网络的交通场景语义分割方法。通过半全局立体匹配和快速全局图像平滑方法获取更加平滑的交通场景视差图 D，将视差图 D 与 RGB 图像融合成 RGB-D 四通道图像，作为网络的输入；将交通场景分为 7 个类别，采用不同结构的卷积神经网络和不同的学习率策略对网络进行训练。在 KITTI 数据集下的实验结果表明，所提方法能够实现像素级别的交通场景语义分割并具有良好的顽健性和可扩展性。通过与以 RGB 图像为输入的交通场景分割方法的对比分析表明，本文提出的基于 RGB-D 图像和卷积神经网络的交通场景分割算法具有更高的语义分割精度，为进一步实现无人驾驶和提高车载环境认知的智能化程度奠定了良好基础。

参考文献：

[1] ANBALAGAN T, GOWRISHANKAR C, SHANMUGAM A. SVM based road surface detection to improve performance of ABS[J]. Journal of Theoretical & Applied Information Technology, 2013, 51(2): 234-239.

[2] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.

[3] 高常鑫, 桑农. 基于深度学习的高分辨率遥感影像目标检测[J]. 测绘通报, 2014(S1):108-111.
GAO C X, SANG N. Deep learning for object detection in remote sensing image[J]. Bulletin of Surveying and Mapping, 2014(S1): 108-111.

[4] 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.
GAO K J, SUN S Y, YAO G S, et al. Semantic segmentation of night vision images for unmanned vehicles based on deep learning[J]. Journal of Applied Optics, 2017, 38(3):421-428.

[5] 刘丹, 刘学军, 王美珍. 一种多尺度 CNN 的图像语义分割算法[J]. 遥感信息, 2017, 32(1):57-64.
LIU D, LIU X J, WANG M Z. Semantic segmentation with multi-scale convolutional neural network[J]. Remote Sensing Information, 2017, 32(1):57-64.

[6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 25(2): 1-9.

[7] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. ArXiv Preprint, ArXiv: 1409. 1556, 2014.

[8] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//IEEE Conference on Computer Vision and Pattern Recognition. 2014: 1-9.

[9] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C]// IEEE Computer Vision and Pattern Recognition.2009:248-255.

[10] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]// IEEE Computer Vision and Pattern Recognition. 2015: 3431-3440.

[11] BADRINARAYANAN V, HANDA A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling[J]. ArXiv Preprint, ArXiv: 1505. 07293, 2015.

[12] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99): 1.

[13] NOH H, HONG S, HAN B. Learning deconvolution network for semantic segmentation[C]//IEEE International Conference on Computer Vision. 2015: 1520-1528.

[14] SILBERMAN N, HOIEM D, KOHLI P, et al. Indoor segmentation and support inference from RGBD images[C]//European Conference on Computer Vision. 2012: 746-760.

[15] GUPTA S, GIRSHICK R, ARBELÁEZ P, et al. Learning rich features from RGB-D images for object detection and segmentation[C]// European Conference on Computer Vision. 2014: 345-360.

[16] SHAO T, XU W, ZHOU K, et al. An interactive approach to semantic modeling of indoor scenes with an RGBD camera[J]. ACM Transactions on Graphics, 2012, 31(6): 439-445.

[17] FILLIAT D, BATTESTI E, BAZEILLE S, et al. RGBD object recognition and visual texture classification for indoor semantic mapping[C]//2012 IEEE International Conference on Technologies for Practical Robot Applications.2012: 127-132.

[18] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]//IEEE Conference on Computer Vision and Pattern Recognition.2012: 3354-3361.

[19] LI L, HUANG H, QIAN B, et al. Vehicle detection method based on mean shift clustering[J]. Journal of Intelligent & Fuzzy Systems, 2016, 31(3):1355-1363.

[20] MIN D, CHOI S, LU J, et al. Fast global image smoothing based on weighted least squares[J]. IEEE Transactions on Image Processing a Publication of the IEEE Signal Processing Society, 2014, 23(12): 5638-5653.

[21] 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.

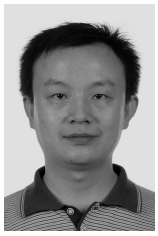
[22] GOULD S, FULTON R, KOLLER D. Decomposing a scene into geometric and semantically consistent regions[C]// IEEE International Conference on Computer Vision.2009:1-8.

[23] EIGEN D, FERGUS R. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture[C]// IEEE International Conference on Computer Vision. 2015: 2650-2658.

[24] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[25] JIA Y, SHELHAMER E, DONAHUE J, et al. Caffe: convolutional architecture for fast feature embedding[C]//The 22nd ACM International Conference on Multimedia. 2014: 675-678.

[作者简介]



李琳辉（1981-），男，河南辉县人，博士，大连理工大学副教授，主要研究方向为汽车安全辅助驾驶、智能车辆及基于视觉传感器的车载环境感知等。

钱波（1991-），男，江苏宿迁人，大连理工大学硕士生，主要研究方向为图像语义分割、立体视觉。

连静（1980-），女，吉林公主岭人，博士，大连理工大学副教授，主要研究方向为智能电动汽车、新能源汽车动力总成及整车控制等。

郑伟娜（1994-），女，山东日照人，大连理工大学硕士生，主要研究方向为交通场景语义分割、图像理解。

周雅夫（1962-），男，辽宁铁岭人，大连理工大学教授，主要研究方向为新能源汽车智能化技术、车载信息采集与远程监控技术、电动汽车整车匹配设计与控制技术、电动汽车驱动电机及其控制技术等。