

基于 RGBD 图像和卷积神经网络的快速道路检测

曲 磊^{1,2}, 王康如^{1,2}, 陈利利¹, 李嘉茂¹, 张晓林¹

¹中国科学院上海微系统与信息技术研究所仿生视觉系统实验室, 上海 200050;

²中国科学院大学, 北京 100049

摘要 针对移动平台有限的计算资源以及基于彩色图像的道路检测方法在极端光照情况下及路面类型变化时存在的不足, 提出了一种融合彩色图像和视差图像的基于 9 层卷积神经网络的快速道路检测算法。提出一种数据输入层预处理方法, 将视差图变换为视差梯度图以强化地面特征, 降低网络深度需求。所提两种网络结构为双通道后融合网络和单通道前融合网络, 分别用于卷积特征分析和快速道路检测。实验使用 KITTI 道路检测数据集并人为划分为普通和困难两组数据集, 对该算法进行实验对比和分析, 结果表明: 与基于彩色图像的卷积神经网络方法相比, 该算法在普通数据集上最大 F_1 指标 ($\text{Max}F_1$) 提升了 1.61%, 在困难数据集上 $\text{Max}F_1$ 提升了 11.58%, 算法检测速度可达 26 frame/s, 可有效克服光照、阴影、路面类型变化等影响。

关键词 图像处理; 道路检测; 卷积神经网络; 视差图像; 立体视觉

中图分类号 TP391.4 **文献标识码** A

doi: 10.3788/AOS201737.1010003

Fast Road Detection Based on RGBD Images and Convolutional Neural Network

Qu Lei^{1,2}, Wang Kangru^{1,2}, Chen Lili¹, Li Jiamao¹, Zhang Xiaolin¹

¹Laboratory of Bionic Vision System, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China;

²University of Chinese Academy of Sciences, Beijing 100049, China

Abstract The road detection method based on color image exists problems under the extreme lighting conditions and changing road surface types, and the computing resource in moving platform is limited. So, based on the 9-layer convolutional neural network, a fast road detection algorithm is proposed to mix the color image and the disparity images. A new preprocessing method is applied in the data input layer, which can transform the disparity images to disparity gradient maps so as to enhance the representation of roads and reduce the demand for network depth. Two proposed networks are proposed including a double-path convolutional neural network which is used to analyze the characteristics of the convolutional neural network, and a single-path convolutional neural network which is applied to detect the road rapidly. The performance of the proposed algorithm is experimentally compared and analyzed on the KITTI road detection dataset which is divided into a common database and a difficult database artificially. The result demonstrates that, compared with the convolutional neural network method based on color images, the $\text{Max}F_1$ measures on the common database and difficult database improve by 1.61% and 11.58%, respectively, and the detection speed can be 26 frame/s. The proposed algorithm can overcome the impact of the lighting, shadow and the changing road surface effectively.

Key words image processing; road detection; convolutional neural network; disparity image; stereo vision

OCIS codes 100.4996; 110.2970; 330.1400

收稿日期: 2017-04-21; 收到修改稿日期: 2017-05-26

基金项目: 国家自然科学基金(61671014, 61601448)、上海市科技人才计划项目(14YF1407300)

作者简介: 曲 磊(1989—), 男, 博士研究生, 主要从事计算机视觉方面的研究。E-mail: qulei@mail.sim.ac.cn

导师简介: 张晓林(1963—), 男, 博士, 研究员, 博士生导师, 主要从事机器视觉方面的研究。

E-mail: xlzhang@mail.sim.ac.cn

1 引 言

道路检测是汽车辅助驾驶、移动机器人、智慧交通等领域最基础、最重要的研究内容之一。传统的道路检测算法主要分为基于道路边缘提取的道路检测和基于区域分割的道路检测。前者主要利用道路边缘信息提取道路边界线或者消失点,如王文锋等^[1]利用局部道路边缘的直线检测和方向一致性判别准则,提出一种基于平行边缘的道路识别算法。后者主要利用道路颜色、纹理等信息对道路图像进行像素级区域分割、聚类等,如段志刚等^[2]对彩色光照不变图像进行分割和聚类,并通过投票函数及道路判别准则提取道路区域。近年来,随着卷积神经网络的发展,基于卷积神经网络的算法已在道路检测算法评测中居于领先地位^[3]。Mendes等^[4]在Krizhevsky等^[5]的基础上提出9层全卷积神经网络实现了道路的快速检测,不仅能保证检测精度,而且检测速度在KITTI道路检测^[3]中领先;Teichmann等^[6]在VGG16的基础上采用端到端和多任务联合训练的方式,在实时情况下,实现了道路的精准检测。

随着网络深度的不断增加,道路检测精度越来越高,但网络计算的复杂度与资源占用率也越来越高,现有的移动平台计算能力无法满足深度网络的要求。另一方面,传统的基于卷积神经网络的道路检测算法大多仅以彩色图片作为网络输入。基于彩色图像的道路检测容易受到地面阴影、路面反光、路面类型变化、动态复杂环境等影响,从而导致路面、建筑、车体等误检或漏检^[4,7]。因此,在使用较浅网络的情况下,同时提升检测速度与检测精度,显得尤为重要。

近年来,Gupta等^[8]提出的HHA方法在基于RGBD图像的物体检测与语义分割领域得以广泛使用。HHA方法将深度图像变换为水平视差、地面高度和重力夹角三通道图表示,并输入卷积网络以提取高层特征,在提升检测精度的同时,降低了网络深度的需求。然而,HHA方法需要转化为三种映射图,不仅计算复杂,而且假设了地面的位置,因此并不适合道路检测。本文利用视差图像的相对梯度信息代替绝对的深度信息,提出一种简单有效的地面梯度表征方法,将视差图表征为单通道的视差梯度图作为网络输入,该方法可用于神经网络的预处理环节,能方便地融入网络的数据输入层。另外,区别于HHA等使用深度图的方法,该方法直接使用视差图像,因此避免了标定相机外部参数和计算三维空间坐标的额外计算量。

为解决传统的基于彩色图像的道路检测方法在极端光照情况下和路面类型变化时存在的问题,本文提出了一种综合运用彩色图像和视差图像的基于卷积神经网络的道路快速检测算法。根据视差图中地面的先验结构信息,提出基于视差梯度图来学习路面的高层特征。采用了两种网络模型,即双通道后融合网络和单通道前融合网络。双通道后融合网络在卷积层部分设计了两个卷积通道,分别输入颜色图和对应的视差梯度图,并采用后融合方法^[9]进行网络融合,这种双通道后融合结构有利于对卷积层输出响应进行可视化,以对比和分析视差梯度图的影响与作用机理;而单通道前融合网络则有利于提升检测速度,比较适合移动平台部署。实验使用流行的KITTI道路检测标准数据集^[3],定性和定量地评估算法性能。

2 算法原理

2.1 视差梯度图

基于Hu等^[10]提出的 uv 视差原理,地面在 v 视差图中被映射为一条倾斜直线,称为地面相关线,形如:

$$d(v) = \alpha \cdot v + \beta, \quad (1)$$

式中 $d(v)$ 表示视差值, v 是图像像素的纵坐标, α 和 β 是与相机外部参数以及地面坡度相关的参数。因此,地面的视差值由近到远呈现近似均匀的递减趋势。基于地面的这种先验特点,提出利用视差梯度信息来表征地面,以弥补基于彩色图像的道路检测方法在光照和道路类型变化时存在的不足。

视差梯度图的每个像素存储该中心像素与周围邻域像素的视差梯度。由 uv 视差原理^[10]可推得:

$$\text{grad } d(u, v) = \gamma \cdot u + \tau \cdot v, \quad (2)$$

式中 $\text{grad } d(u, v)$ 表示点 (u, v) 的梯度, γ 和 τ 是与相机外部参数以及地面坡度或者物体表面朝向相关的参数,即图像的局部梯度信息仅与相机外部参数以及地面坡度或者物体表面朝向相关。为节约内存空间,受局部二值模式算法^[11]启发,本文算法以8 bit表示中心像素与周围8邻域像素的视差梯度,以二进制位0或1表示视差梯度的增量关系,而不保存具体的梯度幅值,梯度为负则为0,反之为1,则每个像素的8位二进制

数值都恰好表示了该点的局部视差梯度信息。与 Sobel 方法相比,本文算法侧重保留视差增量关系,梯度幅值并不重要,且一次计算就可记录各个方向,而后者侧重计算梯度幅值绝对值,需要多个模板才能表征所有方向。

图 1 为典型的道路场景模拟图,包括 3 种不同坡度的路面与 7 种不同朝向的物体表面,自右向左为汽车的前进方向,车载双目相机如图 1 中坐标系所示,与之对应的模拟视差图、视差梯度图以及二值化后的视差梯度图,分别如图 2(a)~(j)中第 1~3 行所示,视差范围为 0~255。图 2(a)为水平道路,图 2(b)为上坡道路,图 2(c)为下坡道路,从图 2(a)~(c)可以看出,道路平面呈现明显且密集的横向条纹,条纹的密集程度与道路的坡度有关,上坡路横向条纹变稀疏,下坡路横向条纹变稠密。图 2(f)为车辆前方前倾平面,图 2(g)为后倾平面,从图 2(f)、(g)可以看出,前后倾平面呈现与道路平面相似的横向条纹,但是条纹的密度以及条纹的亮度都有所差别,前倾平面的视差梯度图中条纹亮度较亮,而后倾平面的视差梯度图中条纹亮度较暗,与地面类似,但是条纹密度比地面稀疏。图 2(d)为车辆左侧平面,图 2(e)为右侧平面,图 2(h)为车辆左侧斜平面,图 2(i)为右侧斜平面,从图 2(d)、(e)和图 2(h)、(i)可以看出,左侧、右侧平面以及左斜、右斜平面的视差纹理图都是竖条纹,和地面有本质区别,同时彼此之间在条纹亮度以及条纹密度方面也有区别,左侧、右侧平面条纹比较稠密,而左斜、右斜平面条纹比较稀疏。另一方面,左侧、左斜平面条纹亮度较亮,而右侧、右斜平面条纹亮度较暗。竖直平面则如图 2(j)所示,呈现为白色无条纹平面。总而言之,经过本文算法预处理的视差梯度图中,路面与物体表面在条纹亮度、密度方面呈现明显不同,这主要与相机外部参数以及地面坡度或者物体表面朝向相关,因此提出将彩色图片与对应的视差梯度图共同作为卷积神经网络的输入。

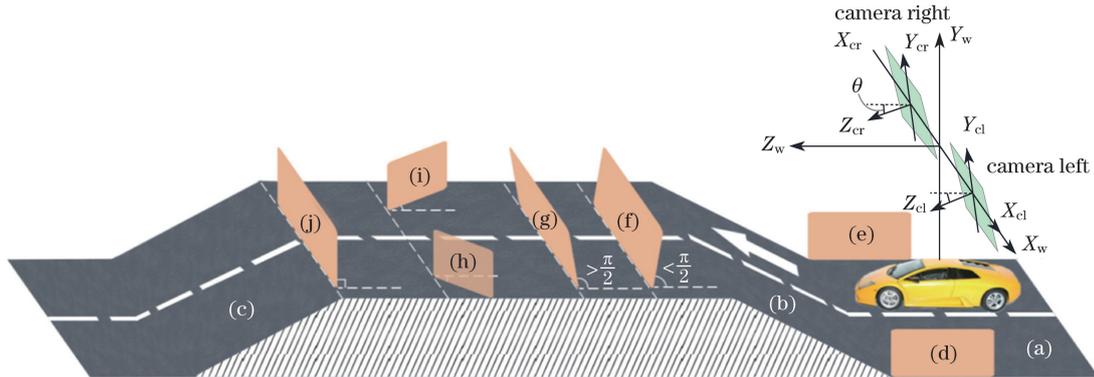


图 1 典型道路场景模拟图。(a)~(c)不同坡度的路面;(d)~(j)不同朝向的物体表面

Fig. 1 Simulation of typical road scenes. (a)-(c) Road surfaces with different longitudinal slopes; (d)-(j) object surfaces with different orientations

2.2 卷积网络结构

卷积神经网络包括两种网络模型,即双通道后融合网络与单通道前融合网络。其中,双通道后融合网络的整个网络流程如图 3 所示,在卷积层部分设计了两个卷积通道,分别输入彩色图和对应的梯度图,并采用后融合方法进行网络融合,这种网络结构不仅有利于拓宽网络结构,提升检测性能,而且有利于分析梯度图对道路检测的影响,避免彩色图像的影响。单通道前融合网络则仅使用一个通道,在数据输入时将彩色图和对应的梯度图合成 4 通道图像,这种方法有利于压缩网络复杂度,提高检测速度,更适合应用于移动平台。

网络输入是一个 $66 \text{ pixel} \times 66 \text{ pixel}$ 的 4 通道图像块,包括 3 通道的彩色图和单通道的视差梯度图。输出是图像块中心 $4 \text{ pixel} \times 4 \text{ pixel}$ 区域的类别。与文献[4]相同,也集中地将所有图像尺寸缩放为 $621 \text{ pixel} \times 187 \text{ pixel}$,每隔 $4 \text{ pixel} \times 4 \text{ pixel}$ 步长提取一个图像块。不同的是,由于双目视差图像无法准确计算边缘处(盲区)的视差值,因此,本文算法只对图像的中央区域进行道路检测。

整个网络共 9 层,包括 4 层卷积层,2 层池化层和 2 层全连接层,并使用 Softmax 函数输出。输入的 4 通道 $66 \text{ pixel} \times 66 \text{ pixel}$ 图像块被分别输入两个通路,即彩色通路和视差通路。两个通路结构相同,每个通路都是从一个 $32 \times 3 \times 3$ 卷积层(32 个 3×3 的卷积滤波器,步长 1)开始,汇聚得到 32 个 $15 \text{ pixel} \times 15 \text{ pixel}$ 的特征映射图,经过两层全连接网络,最后使用 Softmax 函数输出二分类结果。其中,所有卷积层和第一全连接层都配有修正线性单元(ReLU)激活函数^[5,12]。为避免过拟合,全连接层都配有随机丢弃(Dropout^[5,13])层(概率为 0.5)。

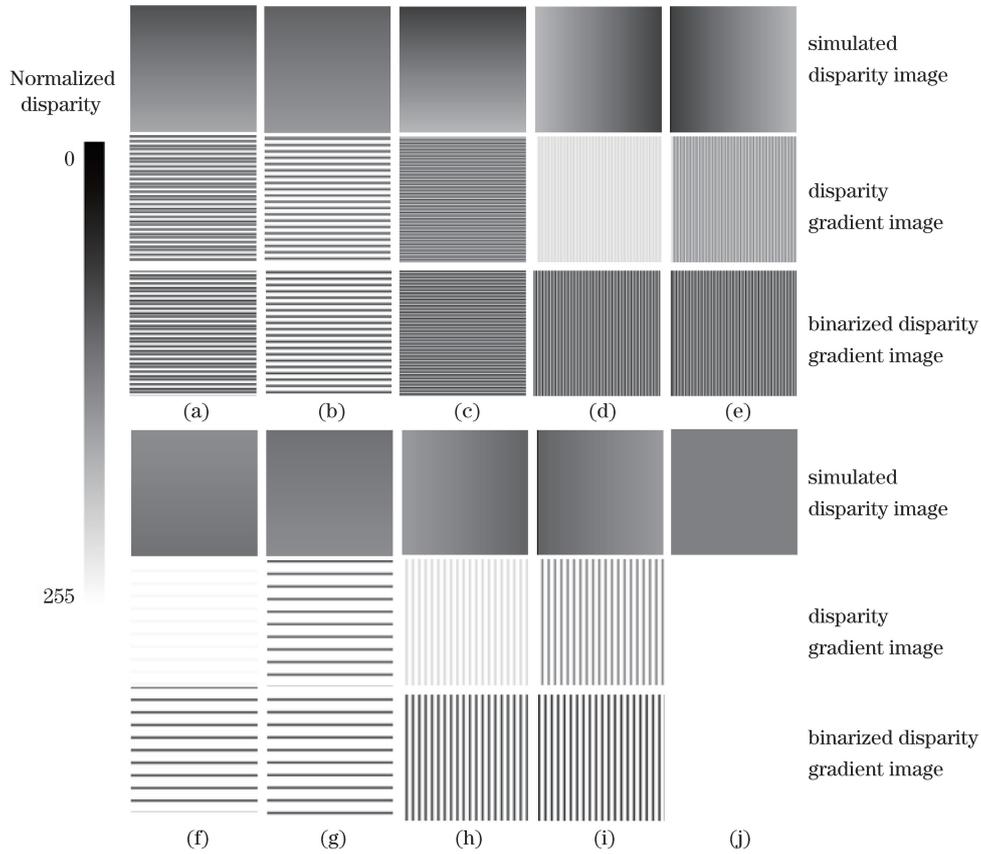


图 2 模拟视差图、视差梯度图以及二值化视差梯度图。(a)水平道路;(b)上坡道路;(c)下坡道路;(d)车辆左侧平面;(e)车辆右侧平面;(f)车辆前倾平面;(g)车辆后倾平面;(h)车辆左侧斜平面;(i)车辆右侧斜平面;(j)竖直平面

Fig. 2 Simulated disparity images, disparity gradient images and binarized disparity gradient images. (a) Level road; (b) uphill road; (c) downhill road; (d) left-side plane of the vehicle; (e) right-side plane of the vehicle; (f) forward plane of the vehicle; (g) backward plane of the vehicle; (h) left oblique plane of the vehicle; (i) right oblique plane of the vehicle; (j) vertical plane

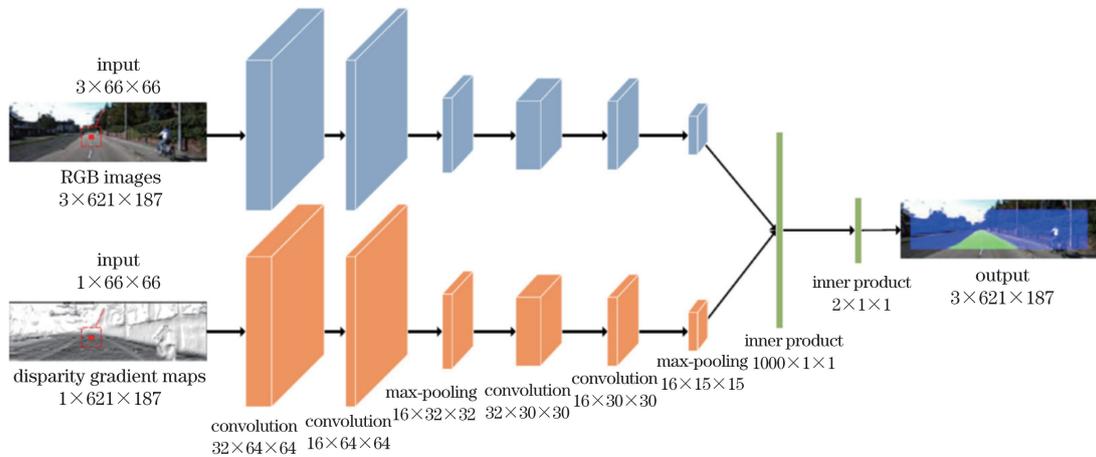


图 3 双通道后融合卷积神经网络的流程图

Fig. 3 Flowchart of double-path convolutional neural network with late fusion

3 实验过程与结果

3.1 数据集

本文实验数据使用 KITTI 道路检测标准数据集^[3],包括 289 张训练图片和 290 张测试图片。KITTI 道路检测数据集包括三类道路:有标记的单行道,无标记的单行道和有标记的多行道,是目前最流行的道路检

测数据集之一。使用 Zbontar 等^[14]提出的立体匹配算法计算视差图。

与文献[4]方法相同,将 289 张训练图片分为含 260 张的训练集和含 29 张的验证集。所有定量的评价结果都是基于验证集获得的。本次实验使用低分辨率图片,所有的图片尺度都被缩放为 621 pixel×187 pixel。如此可以加速训练过程,节约内存。在每张图片中遍历选取 66 pixel×66 pixel 的图像块作为样本。每个样本都由 3 通道的 RGB 图像和单通道的梯度图像以及一个对应的类别标签构成。与文献[4]方法相同,只使用那些图像块中心 4 pixel×4 pixel 区域内像素是同一类的作为样本数据,并对正样本采用垂直翻转的数据扩增方法,从而增加正样本量以平衡数据。

本次实验分为普通性与困难性两组实验。普通性实验采用随机的方法划分训练集和验证集进行训练与评价,而困难性实验则使用文献[4]方法评价结果中较差的 29 张图片以构成验证集,其余构成训练集。因此,本次实验包括普通性和困难性两组训练集和验证集,其中困难性实验更能体现出本文算法对检测性能的巨大提升。

3.2 训练

以文献[4]方法作为对比,进行以下 6 组实验,其中 RGBDT-IN 和 RGB-DT-IN 是本文提出的最终算法:

- 1) RGB:即文献[4]方法,以 RGB 彩色图作为输入,等同于本文算法仅使用彩色通路。
- 2) RGBD:采用单通道前融合网络结构,以 RGB 彩色图和单通道视差图构成的 4 通道图作为输入。
- 3) RGBDT:采用单通道前融合网络结构,以 RGB 彩色图和单通道视差梯度图构成的 4 通道图作为输入。
- 4) RGB-DT:采用双通道后融合网络结构,以 RGB 彩色图和单通道视差梯度图构成的 4 通道图作为输入。
- 5) RGBDT-IN:采用单通道前融合网络结构,以 RGB 彩色图和单通道视差梯度图构成的 4 通道图作为输入,对正样本进行垂直翻转以增加样本量。
- 6) RGB-DT-IN:采用双通道后融合网络结构,以 RGB 彩色图和单通道视差梯度图构成的 4 通道图作为输入,对正样本进行垂直翻转以增加样本量。

所有实验使用批梯度下降法(MBGD)训练网络模型,批大小为 100。初始学习率为 0.01,每轮更新率 0.96,即 $l := l \times 0.96$,其中 l 是学习率,每次训练持续 30 轮。动量为 0.9,权重衰减为 0.0001。基于 Caffe 框架^[15],使用了 cuDNN v4 库。实验平台为 Intel Xeon E5-2630@2.4 GHz,内存 64 GB,图形处理器为 NVIDIA Tesla K40c。

3.3 定性评价

图 4 所示为部分实验结果,分别展示了文献[4]方法和本文算法在 KITTI 道路检测测试集上的对比实验结果,绿色标记为检测到的道路区域。其中图 4(a)、(c)为采用文献[4]方法的结果,图 4(b)、(d)为采用本文 RGB-DT-IN 算法的结果。前 5 行分别列举了在强光照、阴影情况下的对比实验结果,后 5 行分别列举了路面类型变化、动态复杂环境下的实验结果。可以发现,在强光照情况下,图 4(a)、(b)的第 1 行中路面中央区域几乎呈亮白色,文献[4]方法已经无法检测,而本文算法则可以准确检测;阴影情况下,图 4(c)、(d)的第 4 行,浓密树荫导致地面严重阴影,文献[4]方法只能检测到稀疏的部分区域,而本文算法则可准确检测到全部路面;路面类型变化情况下,图 4(a)、(b)的第 9 行,对于呈网格状的砖铺道路,文献[4]方法只能检测远处网格纹理不明显的道路,而对于近处道路则出现大片漏检情况,而本文算法都能准确检测;复杂环境情况下,图 4(c)、(d)的第 7、8 行道路右侧加油站区域以及第 10 行左侧的墙壁区域,本文算法都能准确应对。定性测试结果表明,本文算法能够有效抵抗光照变化、阴影、道路类型变化的影响,减少路面、建筑等被误检或漏检的情况。更重要的是,针对图 4(a)、(b)后 5 行出现的在训练集中不存在的道路类型,本文算法较好地回答了文献[4]的疑问,表明即使在不同的道路类型上训练,借助立体视觉的方法,依然可以较好地解决路面类型变化的问题。

3.4 定量评价

采用多种指标以衡量实验结果,包括最大 F_1 指标($\text{Max}F_1$)、交并比(IoU)、精确率(Pre)、召回率(Rec)、假阳率(FPR)、假阴率(FNR)。所有定量评价结果都是基于验证集获得的。与文献[4]方法类似,每组实验

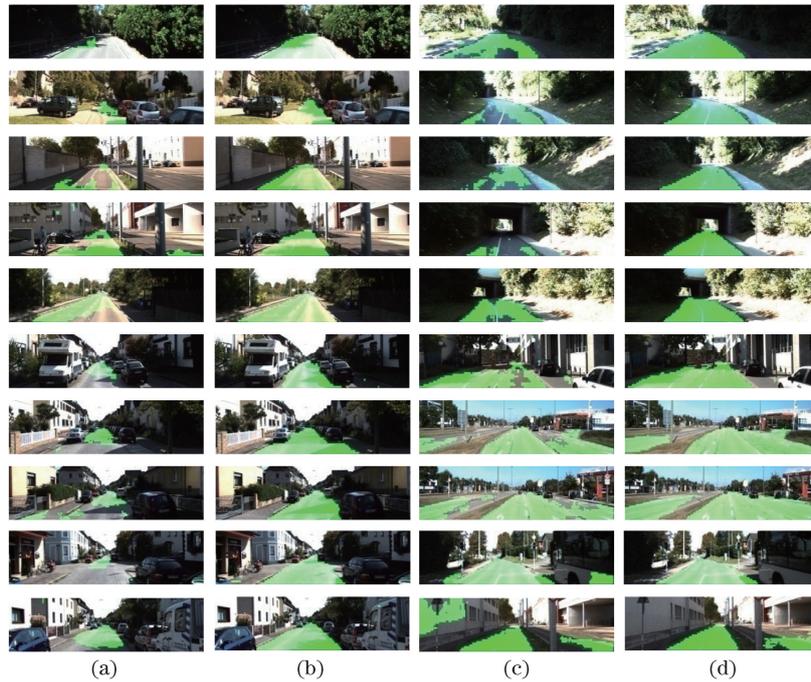


图 4 KITTI 道路检测测试集实验结果图。(a)(c)文献[4]方法;(b)(d)本文的 RGB-DT-IN 算法
Fig. 4 Experimental results of KITTI road detection testing dataset. (a)(c) Method of Ref. [4];
(b)(d) proposed RGB-DT-IN algorithm

重复多次,选择 MaxF_1 值最大的模型作为结果。

实验结果如表 1、2 所示,分别展示了 6 组实验在普通与困难验证集上的评价结果。对比实验 1、2、3 的结果表明,直接使用视差图作为网络输入并不能稳定地提升检测效果,虽然视差图的加入使得网络在困难数据集上 MaxF_1 指标提升了 4.58%,但在普通数据集上反而降低了 1.51%,因此网络并没有很好地学习到地面特征,可能需要更深的网络才能直接从视差图中学习到地面特征。相比而言,RGBDT 算法使得网络性能得到稳步提高,在普通数据集上 MaxF_1 提升 1.20%,在困难数据集上 MaxF_1 提升了 10.12%,说明 RGBDT 算法能够显著地增强地面特征,降低网络深度需求,只需较浅网络就能达到类似深层卷积网络的效果。

表 1 不同实验方法在普通数据集上的评价结果

Table 1 Evaluation results of different methods on the common dataset

No.	Method	MaxF_1	IoU	Pre	Rec	FPR	FNR	%
1	RGB	93.76	88.75	95.75	92.50	0.62	7.50	
2	RGBD	92.25	86.57	88.31	97.92	2.26	2.08	
3	RGBDT	94.96	90.96	94.95	95.68	0.80	4.32	
4	RGB-DT	95.27	91.56	93.96	97.36	1.00	2.64	
5	RGBDT-IN	95.18	91.46	94.21	96.97	0.96	3.03	
6	RGB-DT-IN	95.37	91.73	95.00	96.42	0.78	3.58	

表 2 不同实验方法在困难数据集上的评价结果

Table 2 Evaluation results of different methods on the difficult dataset

No.	Method	MaxF_1	IoU	Pre	Rec	FPR	FNR	%
1	RGB	83.43	75.35	86.03	83.19	1.82	16.81	
2	RGBD	88.01	79.84	84.51	94.25	2.81	5.75	
3	RGBDT	93.55	88.72	93.38	94.81	1.02	5.19	
4	RGB-DT	94.12	89.37	95.83	93.26	0.61	6.74	
5	RGBDT-IN	94.32	89.97	94.39	95.30	0.92	4.70	
6	RGB-DT-IN	95.01	91.11	94.62	96.25	0.86	3.75	

实验 3、4 和实验 5、6 的对比结果表明:双通道后融合网络可以稳定地提升检测效果,这主要得益于网络容量的增大。实验 3、5 和实验 4、6 的对比结果表明,采用垂直翻转的数据扩增也有益于网络性能提升。与文献[4]方法相比,在普通数据集上,RGB-DT-IN 算法在 Pre 降低 0.75%的情况下,Rec 提升了 3.92%;在 FPR 增大 0.16%的情况下,FNR 降低了 3.92%。这表明 RGB-DT-IN 算法明显减少了地面的漏检情况,能够更准确地检测出地面区域。在普通数据集上,利用 RGB-DT-IN 算法 MaxF₁ 提升了 1.61%,IoU 提升了 2.98%。在困难数据集上,利用 RGB-DT-IN 算法所有指标都大幅提升,MaxF₁ 提升了 11.58%,IoU 提升了 15.76%,基本达到普通数据集上的评价结果。这表明 RGB-DT-IN 算法能够较好克服文献[4]方法存在的不足,对光照、阴影、路面类型变化等情况有较强稳健性。

3.5 预测时间

为了进一步提升检测速度,将全连接网络改为全卷积网络,将网络中的连接层使用相同数量参数的卷积层代替,其中,第一全连接层使用 $1000 \times 15 \times 15$ 卷积层代替,第二全连接层使用 $2 \times 1 \times 1$ 卷积层代替。

网络预测时间如表 3 所示。其中,前馈时间仅包含网络前馈的耗时,而预测时间包括读取图片、网络前馈和输出预测结果的整个时长。第 1、2 行分别为文献[4]方法在原论文中的耗时与本文复现该方法的实际耗时。对比第 2、3 行数据表明,与文献[4]方法相比,加入视差梯度图以及单通道前融合网络几乎没有增加时间成本,但检测精度却得到较大提高。RGBDT-IN 和 RGB-DT-IN 的检测速度分别达到 26 frame/s 和 15 frame/s,前者具有较快的检测速度,而后者具有较高精度。

表 3 实验方法耗时统计

Table 3 Processing time of methods

Method	Forward time	Prediction time
Ref. [4]	0.025	0.032
RGB	0.033	0.037
RGBDT-IN	0.034	0.038
RGB-DT-IN	0.063	0.067

4 网络可视化分析

卷积网络的可视化对网络优化具有重要作用。利用双通道后融合卷积网络分别可视化了几种典型的道路与非道路图像的卷积输出响应。如图 5、6 所示,分别对应三组输入和输出,图 5、6 中的(a)为输入的 4 通道彩色图与梯度图,图 5、6 中的(b)为汇聚层输出结果,共包括 32 个卷积输出响应,其中,前 16 个对应彩色通道卷积输出,后 16 个对应视差梯度通道卷积输出。

从图 5、6 中的(b)前 16 个卷积输出可以看出,彩色通道的卷积层对边缘与颜色敏感,而对路面则几乎没有响应。其中,第 11 组卷积核对边缘变化具有较强响应,如图 5 第一组道路中心线、第二组左上角的道路边缘以及第三组右上角的道路边缘,在卷积输出上都有较强响应,另外,在图 6 第一组的轨道上、第二组人体周边等也有较强响应;第 15 组卷积核对亮白色具有较强响应,其图 5 第一组的道路中心线、第三组左侧的道路分界线等都有较强响应,另外,在图 6 第二组的腿部反光处以及第三组的天空处也有较强响应。

从图 5、6 中的(b)后 16 个卷积输出可以看出,视差梯度通道的卷积层对梯度图纹理具有明显的辨识能力。其中,第 1、4、6、9、10、11、12、13 组卷积核对水平纹理具有较强响应,尤其是第 4 组卷积核对水平纹理的响应最强,如图 5 第一组道路、第二组右下方道路、第三组左下方道路以及图 6 第一组下方道路都呈现很强的响应;而第 2、3、5、7、8、14 组卷积核则对竖直纹理具有较强敏感性,如图 6 第二组的腿部以及第三组的墙和树木处,都有较强响应;第 16 组卷积核则对斜纹理具有较强敏感性,如图 5 第二组左上角边缘、第三组右上角边缘以及图 6 第一组左上角边缘等都有较强响应。另外,发现竖直敏感性卷积核在竖直物体上的卷积结果几乎与彩色卷积核的结果相同,如图 6 第二组腿部以及第三组墙和树木处,这似乎意味着本文算法可应用于物体检测领域。

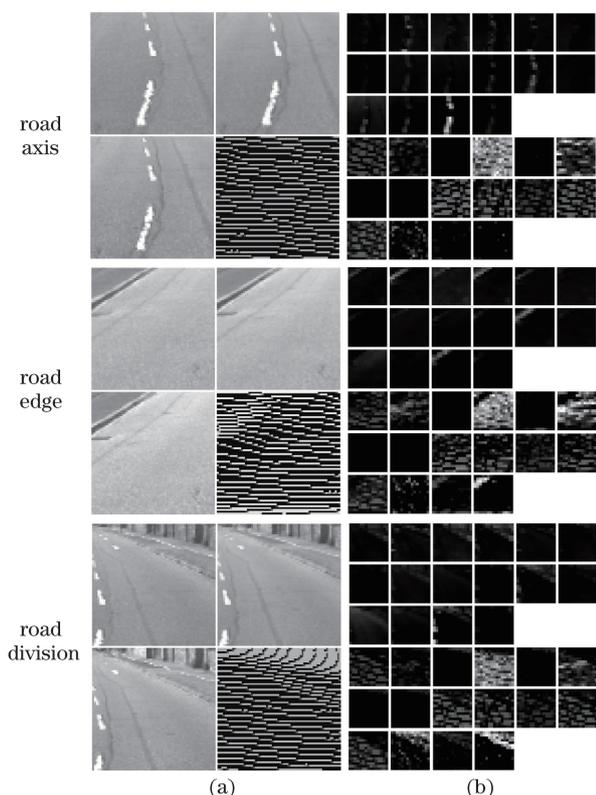


图5 道路图像的卷积输出可视化。

(a)输入的4通道彩色图与梯度图;(b)汇聚层输出结果

Fig. 5 Convolutional output visualization of road images.

(a) Color and gradient maps of input four channel;

(b) output results of convergence layer

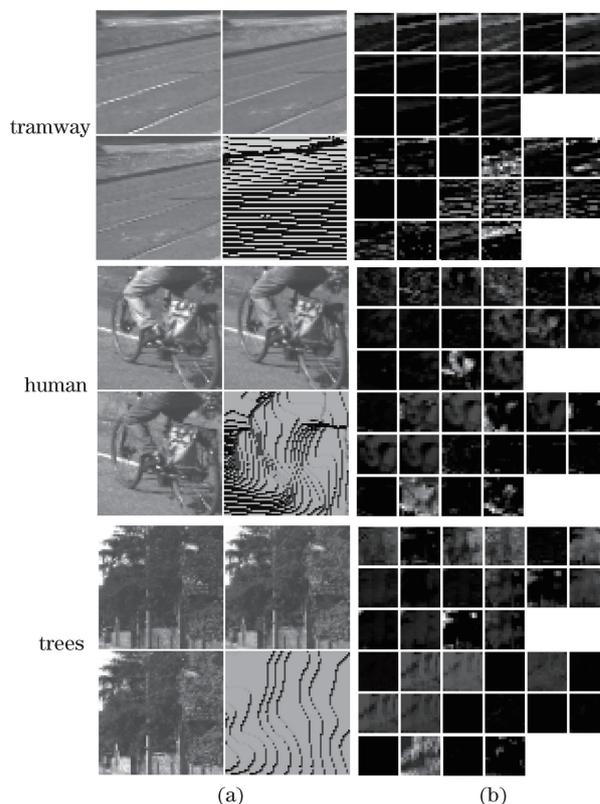


图6 非道路图像的卷积输出可视化。

(a)输入的4通道彩色图与梯度图;(b)汇聚层输出结果

Fig. 6 Convolutional output visualization of non-road images.

(a) Color and gradient maps of input four channel;

(b) output results of convergence layer

5 结 论

提出了一种综合运用彩色图像和视差梯度图的基于卷积融合神经网络的快速道路检测算法。利用视差图中道路视差规律性变化的先验信息,将视差图预处理为视差梯度图作为网络输入。该算法充分利用了视差图像的梯度信息。为进行深入的特征分析,对卷积层输出响应进行了可视化,对比和分析了颜色与视差梯度对道路检测的影响。结果表明,本文算法更有利于增强路面特征,从而降低对网络深度的需求,有利于移动平台部署。

设计了快速的单通道前融合卷积神经网络和双通道后融合卷积神经网络,在卷积层部分设计了两个卷积通道,分别输入彩色图和对应的视差梯度图,并采用后融合方法进行网络融合。不仅提升了网络容量,提高了检测精度,而且具有独立研究和分析彩色图和视差梯度图的作用,避免了相互之间的跨通道影响。

在KITTI道路检测标准数据集上重点对比和分析了视差梯度图与网络结构对检测性能的影响。与文献[4]方法相比,RGB-DT-IN算法在普通数据集上 MaxF_1 提升了1.61%;在困难数据集上 MaxF_1 提升了11.58%。在保证检测精度的同时,RGB-DT-IN算法检测速度可达26 frame/s,网络参数模型也只有14 M,比较合适移动平台使用。实验证明,本文算法可以稳定而有效地克服光照、阴影和道路类型变化等对道路检测的影响,减少路面、建筑等被误检或漏检的情况发生。

当然,本文算法也存在一定的漏分、错分小区域问题,并且由于视差盲区的存在,并不适用于整图中的道路检测,如何避免视差盲区对基于RGBD图像的道路检测的影响,仍需要进一步研究解决。

参 考 文 献

- [1] Wang Wenfeng, Ding Weili, Li Yong, *et al.* An efficient road detection algorithm based on parallel edges[J]. Acta

Optica Sinica, 2015, 35(7): 0715001.

王文锋, 丁伟利, 李 勇, 等. 一种高效的基于平行边缘的道路识别算法[J]. 光学学报, 2015, 35(7): 0715001.

- [2] Duan Zhigang, Li Yong, Wang Ende, *et al.* Road and navigation line detection algorithm from shadow image based on the illumination invariant image[J]. Acta Optica Sinica, 2016, 36(12): 1215004.
段志刚, 李勇, 王恩德, 等. 基于光照不变图像的阴影图像道路及导航线提取算法[J]. 光学学报, 2016, 36(12): 1215004.
- [3] Urtasun R, Lenz P, Geiger A. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012: 3354-3361.
- [4] Mendes C C T, Fr Mont V, Wolf D F. Exploiting fully convolutional neural networks for fast road detection[C]. IEEE International Conference on Robotics and Automation (ICRA), 2016: 3174-3179.
- [5] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]. International Conference on Neural Information Processing Systems (NIPS), 2012, 25(2): 1097-1105.
- [6] Teichmann M, Weber M, Zoellner M, *et al.* MultiNet: real-time joint semantic reasoning for autonomous driving[J]. Computer Vision and Pattern Recognition, 2016.
- [7] Oliveira G L, Burgard W, Brox T. Efficient deep models for monocular road segmentation[C]. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016: 4885-4891.
- [8] Gupta S, Girshick R, Arbel Ez P, *et al.* Learning rich features from RGB-D images for object detection and segmentation[C]. 13th European Conference on Computer Vision (ECCV), 2014: 345-360.
- [9] Chen X Z, Ma H M, Wan J, *et al.* Multi-view 3D object detection network for autonomous driving[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [10] Hu Z, Uchimura K. U-V-disparity: an efficient algorithm for stereovision based scene analysis[C]. IEEE Intelligent Vehicles Symposium (IVS), 2005: 48-54.
- [11] Ojala T, Pietikainen M, Harwood D. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions[C]. 12th International Conference on Pattern Recognition (ICPR), 1994: 582-585.
- [12] Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines[C]. 27th International Conference on Machine Learning (ICML), 2010: 807-814.
- [13] Srivastava N, Hinton G, Krizhevsky A, *et al.* Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [14] Zbontar J, Lecun Y. Computing the stereo matching cost with a convolutional neural network[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 1592-1599.
- [15] Jia Y, Shelhamer E, Donahue J, *et al.* Caffe: convolutional architecture for fast feature embedding[J]. Proceedings of the 22nd ACM International Conference on Multimedia, 2014: 675-678.