文章编号:1007－6735(2020)06－0558－08　DOI: 10.13255/j.cnki.jusst.20191109001

基于深度学习的障碍物检测与深度估计

仇旭阳，　黄影平，　郭志阳，　胡　兴
（上海理工大学 光电信息与计算机工程学院，上海 200093）

摘要：提出了一种针对交通场景的基于深度学习的障碍物检测与深度估计方法。该方法对现有的YOLOv3 模型进行改进，使用 DenseNet 网络代替原网络尺度较小的传输层，得到一种新的障碍物检测模型 Dense-YOLO。然后采用立体匹配模型 PSMNet 得到双目图像的视差图，根据双目测距原理对被测目标深度进行估计。在 KITTI 数据集和实际交通场景中的实验结果表明，与 YOLOv3 模型相比，Dense-YOLO 模型有效地提高了交通场景中障碍物检测的可靠性和正确率，对轿车、行人、骑行者和卡车这 4 类障碍物检测的平均精确率（average precision, AP）提高了 3%～5%，平均精确率均值（mean average precision, mAP）提高了约 4%。障碍物深度估计结果与真实值的平均相对误差约为 3%。

关键词：障碍物检测；深度估计；立体视觉；深度学习；卷积神经网络
中图分类号：TP 391　　**文献标志码：**A

Obstacle detection and depth estimation using deep learning approaches

QIU Xuyang,　HUANG Yingping,　GUO Zhiyang,　HU Xing
(*School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China*)

Abstract: An obstacle detection and depth estimation method based on deep learning for traffic scenarios was proposed. The method modifies the existing YOLOv3 model by replacing its small-scale transmission layer with DenseNet network, and obtains a new obstacle detection network Dense-YOLO. Then, the disparity map of binocular images was obtained by using the stereo matching network model PSMNet, and the depth of detected obstacles was estimated according to the binocular ranging principle. Extensive experiments were conducted on the KITTI dataset and the actual traffic images, and the results show that Dense-YOLO effectively improves the reliability and accuracy of the obstacle detection in traffic scenarios compared to YOLOv3. The average precision (AP) on the four-class of obstacles including car, pedestrian, cyclist and truck is increased by 3% to 5%, and the mean average precision (mAP) is increased by about 4%.The average relative error between the estimated depth of the obstacle and the true value is about 3%.

收稿日期: 2019−11−09
第一作者: 仇旭阳（1995−），男，硕士研究生．研究方向：深度学习、障碍物检测．E-mail：shawn956@163.com
通信作者: 黄影平（1966−），男，教授．研究方向：智能汽车障碍物探测辨识．E-mail：huangyingping@usst.edu.cn

Keywords: *obstacle detection*; *depth estimation*; *stereovision*; *deep-learning*; *convolutional neural network*

从复杂的交通场景图像中准确地检测、辨识障碍物并实现深度估计是智能汽车研究的一个重要课题[1]。随着深度学习的发展，目标检测和识别进入到一个新的阶段。与传统的使用人工设计的特征，如 LBP(local binary patterns)[2], HOG(histogram of oriented gradient)[3] 等对目标进行特征提取，然后将提取到的特征输入到贝叶斯(Bayes)[4]，SVM(support vector machine)[5] 等分类器进行检测和识别的机器学习方法不同，基于深度卷积网络的检测方法具有强大的特征提取能力，能够更好地学习目标的特征，实现更高精度的检测。目前基于深度学习的目标检测方法主要分为基于候选区域的双步(two-stage)检测方法，如 R-CNN[6], Fast R-CNN[7]，Faster R-CNN[8] 等。这类算法将检测阶段分为两步，首先提取图片中目标可能存在的候选区域，然后再将提取的候选区域输入到卷积神经网络中进行特征提取，最后再进行检测和边界框修正[9]。这类算法的特点是检测速度较慢但检测精确度较高。基于边界框回归的单步(one-stage)目标检测算法如 SSD[10]，YOLO(you only look once)[11]，YOLOv2[12]，YOLOv3[13] 等，这类算法不再需要区域建议阶段，直接由网络产生目标的类别概率和位置坐标，经过单次检测即可直接得到最终的检测结果，因此，具有更快的检测速度。其中，YOLOv3 不仅检测精度较高，而且速度较快。但是，对于复杂交通场景中较小的障碍物，检测效果不佳。而且这些基于深度学习的障碍物检测方法，一般采用单目视觉方法，不能检测出目标的深度信息。

对障碍物的深度估计可以使用激光雷达、毫米波雷达、超声波等传感器。与使用传感器测距的方法相比，基于双目视觉的测距方法具有信息丰富、成本低的优点。双目测距方法是利用左右图像进行匹配，获取视差，计算目标的三维信息。在复杂交通场景中，对障碍物进行深度估计，只需要获取场景的图像，然后映射到视差图中，根据双目视觉的原理就可以估计出目标的距离。而目标检测算法可以识别和定位障碍物，从而获取障碍物的坐标信息。为了获取精度高的视差图，本文使用一个端到端的立体匹配网络 PSMNet[14]，其由空间金字塔池化模块和 3D 卷积神经网络模块组成，具有利用全局环境信息寻找不确定区域(遮挡区域、弱纹理区域等)左右图一致性的能力。

针对现有的基于深度学习方法的目标检测存在的问题，本文以 YOLOv3 为基础框架，对其进行改进，使用密集连接卷积网络(DenseNet)[15] 替换 YOLOv3 中尺度较小的传输层，加强特征融合和重用，提出了改进后的 Dense-YOLO 网络，提高了在复杂交通场景中对较小障碍物的检测精确度。在检测的基础上，采用立体匹配网络 PSMNet 获取双目图像的视差图，对 Dense-YOLO 网络检测到的目标区域利用立体视觉原理进行深度估计，实现了对各类障碍物检测的同时提供其距离信息。

1　方法总体框架

本研究的主要任务是检测交通场景下的障碍物并计算出其距离。为了提高障碍物的检测和测距的精度，本文结合深度学习和双目视觉构建一个障碍物检测和测距系统。图 1 是系统的总体框架。

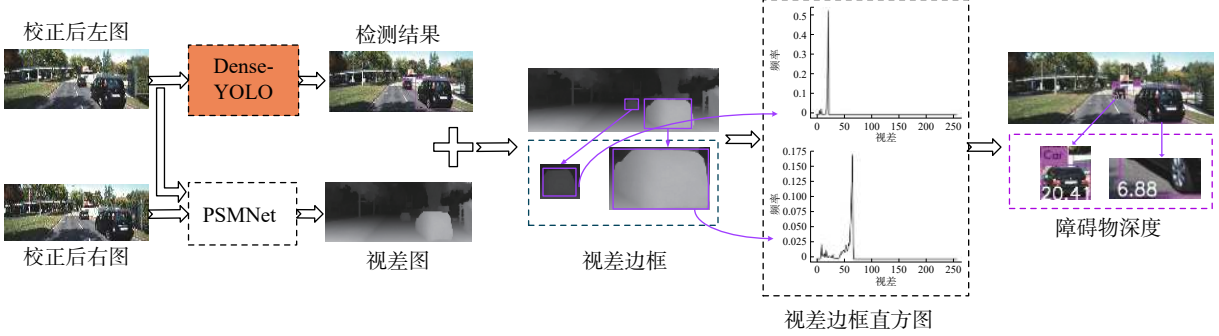


图 1　方法总体框架

Fig.1　Framework of the proposed method

如图1所示，首先以校正后的双目图像作为输入，为了提高目标检测的精度，使用以YOLOv3为基础的改进的Dense-YOLO网络获得高效的检测结果。然后利用PSMNet立体匹配算法得到图像的视差图，并将检测得到的障碍物目标框映射到视差图中。为了排除背景等非障碍物视差对测距精确度的影响，使用视差直方图来统计边界框中每个像素的视差，选取频率最大的视差值作为障碍物的视差。最后利用双目视觉测距原理计算出每个障碍物的距离。

2 障碍物检测

在YOLOv3的基础上使用密集连接卷积网络(DenseNet)进行改进，得到改进后的深度卷积神经网络Dense-YOLO。

2.1 YOLOv3

YOLOv3网络是在YOLO和YOLOv2网络基础上改进而来的。YOLOv3的检测原理如图2所示。t_x，t_y分别为边界框的横坐标和纵坐标，t_w，t_h分别为边界框的宽度和高度。

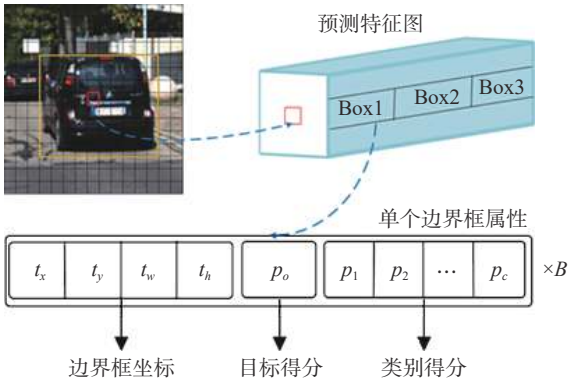


图2 YOLOv3检测原理

Fig. 2 YOLOv3 detection principle diagram

网络将输入图像划分为$S\times S$个网格，如果目标真实边界框的中心落在网格中，那么，该网格负责检测该目标。每个网格预测B个边界框、1个目标置信度得分p_o和C个类别得分$(p_1, p_2, \cdots, p_C)$。其中，目标置信度c定义为

$$c = p_r \times IoU_{\text{pred}}^{\text{truth}}, \quad p_r \in \{0,1\} \tag{1}$$

置信度反映了网格是否包含对象，以及在包含对象时预测的边界框的准确性。当目标在网格中时，$p_r = 1$；否则，$p_r = 0$。$IoU_{\text{pred}}^{\text{truth}}$用于表示真实边界框与预测边界框之间的重合度。当多个包围框检测到同一个目标时，YOLOv3使用非极大值抑制(NMS)方法来选择置信度得分最高的边界框。

2.2 DenseNet

改进目标检测网络的主要方向为增加网络深度或增加网络宽度，密集连接卷积网络(DenseNet)则是从特征入手，通过对特征的极致利用达到更好的效果和更少的参数。

密集连接网络在前馈模式下将当前层与前面所有层连接起来，因此，第l层会接收$x_0, x_1, \cdots, x_{l-1}$层的特征映射作为输入，这样就可以实现层与层之间最大程度的信息传输。

$$x_l = [x_0, x_1, \cdots, x_{l-1}] \tag{2}$$

式中：x_l为前$1-l$层特征图的拼接结果；$[x_0, x_1, \cdots, x_{l-1}]$为$x_0, x_1, \cdots, x_{l-1}$层的特征图拼接，即作通道的合并。

2.3 Dense-YOLO

在神经网络训练过程中，由于卷积和下采样的存在，使得特征向量不断减少，在传输过程中丢失了很多特征信息，这就可能会导致对障碍物检测效果不好，出现漏检或者错检。而DenseNet网络可以有效地利用上一层或者前几层特征信息，同时当前层的特征图也会作为后面每一层的输入，从而促进特征重用。所以，本文在YOLOv3网络的基础上用密集连接网络对网络进行修改，使用DenseNet代替尺度较小的传输层，将当前层之前的多层特征作为输入传递到当前层，重复利用各层特征信息，强化特征传播，并且还可以减少网络计算量。本文所设计的网络结构如图3所示，$H_l(\cdot)$为处理拼接特征图的函数。网络各层参数的设置如图4所示。

图3中，本文使用的非线性变换函数$H_l(\cdot)$是一个组合操作，其包括一系列批量归一化(BN)、整流线性单元(ReLU)和卷积(Conv)操作。在分辨率为30×30的特征层中，x_l由64个子特征层组成。$H_1(\cdot)$先对x_0进行BN-ReLU-Conv(1×1)非线性计算，然后再进行BN-ReLU-Conv(1×1)非线性运算。H_2也对拼接的特征图$[x_0, x_1]$执行上面相同的操作。H_2, H_3以相同的方式完成上述操作。最后，由$[x_0, x_1, x_2, x_3]$拼接的特征图和尺寸为30×30×512的特征层继续向前传播。在分辨率为16×16的特征层中，x_l由128个子特征层组成。按照上述方法进行特征层拼接和特征传播，最后将所有特征层拼接成尺寸为15×15×1 024的特征层前向传播。改进

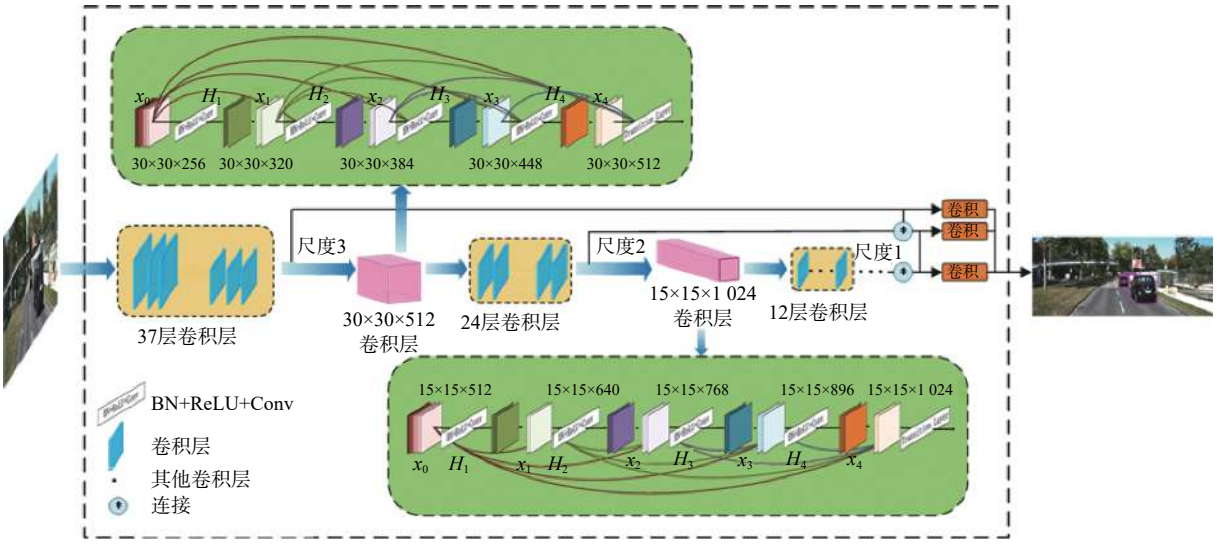


图 3　Dense-YOLO 网络结构图

Fig.3　Dense-YOLO network structure

	Type	Filters	Size	Output
	Convolutional	32	3×3	480×480
	Convolutional	64	3×3/2	240×240
1×	Convolutional	32	1×1	
	Convolutional	64	3×3	
	Residual			240×240
	Convolutional	128	3×3/2	120×120
2×	Convolutional	64	1×1	
	Convolutional	128	3×3	
	Residual			120×120
	Convolutional	256	3×3/2	60×60
8×	Convolutional	128	1×1	
	Convolutional	256	3×3	
	Residual			60×60
	DenseNet	256	3×3/2	30×30
4×	DenseNet	64	1×1	
	DenseNet	64	3×3	30×30
	OutPut	512		30×30
8×	Convolutional	256	1×1	
	Convolutional	512	3×3	
	Residual			30×30
	DenseNet	512	3×3/2	15×15
4×	DenseNet	128	1×1	
	DenseNet	128	3×3	15×15
	OutPut	1 024		15×15
4×	Convolutional	512	1×1	
	Convolutional	1 024	3×3	
	Residual			15×15
	Avgpool		Global	
	Connected		1 000	
	Softmax			

Scale3
Scale2
Scale1
Conv
Conv
Conv
Dense-YOLO detection

图 4　Dense-YOLO 网络参数

Fig. 4　Network parameters of Dense-YOLO

的检测模型预测了 3 种不同尺度的边界框：15×15，30×30，60×60，并能较好地实现车辆前方的轿车、行人、骑行者、卡车等障碍物的检测。

3　障碍物深度估计

3.1　立体视觉原理

双目立体视觉是计算机视觉的一个重要分支。双目视觉遵循人眼观察物体的原理，利用 2 台摄像机从不同角度观察同一目标物体，同时获得两幅图像，经过极线校正、匹配等过程后可以计算图像像素点的视差，进而获得目标的三维坐标信息[16]。双目视觉模型如图 5 所示。

在图 5 中，2 个相机水平放置，焦距均为 f。o_{L}和O_{R}分别是左右相机的光心，2 个光心之间的距离称为基线 B。$O_{\mathrm{L}}O_l$和$O_{\mathrm{R}}O_r$是 2 条平行线，分

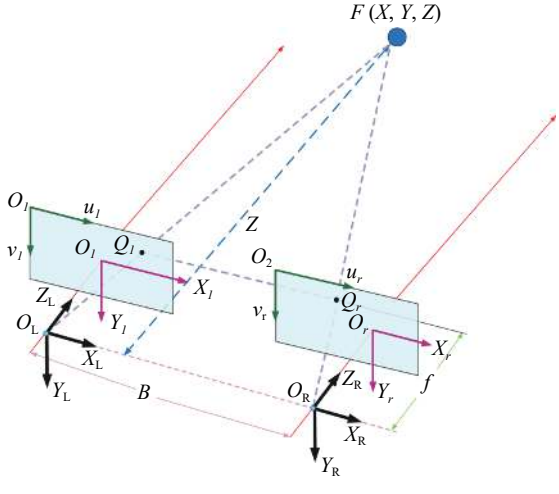


图 5　平行双目视觉模型

Fig. 5　Parallel binocular vision model

别表示左、右相机的光轴。$F(X,Y,Z)$是世界坐标系里的一个目标点，它透过透镜的投影中心投影到像平面上的点$Q_l(x_l,y_l)$和点$Q_r(x_r,y_r)$。将世界坐标原点设为光心O_{L}。然后通过简单的数学模型，可以得到三维世界坐标的表达式：$O_{\mathrm{L}}(0,0,0)$，$O_{\mathrm{R}}(B,0,0)$，$Q_l(x_l,y_l,f)$，$Q_r(x_r+B,y_r,f)$。对于世界坐标中的任意一点F，Q_l和Q_r的y值相等，即$y_l=y_r$。$\triangle FQ_lQ_r$和$\triangle FQ_{\mathrm{L}}Q_{\mathrm{R}}$相似。根据相似三角形原理，可以得到如下关系式：

$$\frac{B-(x_l-x_r)}{B}=\frac{Z-f}{Z} \tag{3}$$

x_l-x_r为视差，令$d=x_l-x_r$，式(5)可改写为

$$\frac{B-d}{B}=\frac{Z-f}{Z} \tag{4}$$

即

$$Z=\frac{fB}{d} \tag{5}$$

如式(5)所示，如果能够获取左、右图像对应像素点的视差，就可以估计出目标的深度。

3.2 立体匹配预测距离

立体匹配是利用校正后的左、右图像获得场景视差图的过程。考虑到需要稠密视差图来提取整个目标的像素点，利用PSMNet算法获取视差图。为了消除背景视差对测量距离的影响，本文利用视差直方图来描述检测边框中的视差的分布。首先将视差值量化到[0，255]，并计算每一个视差的频率f(disparity)。在直方图中，障碍物视差的频率f(disparity)是较大的，并且应该是一个峰值。然后选取每一个直方图的峰值所对应的视差作为障碍物的视差。最后根据式(5)可得到障碍物的深度。

4 实 验

4.1 数据集

采用智能车计算机视觉算法评测数据集KITTI[17]中的图片，并对原有的8类标签信息进行处理，保留实验需要的4个类别标签，即轿车、行人、骑行者和卡车，并将Van归并到Car类。同时选取该数据集中7481张图像作为实验数据。为了使得网络充分学习待检测目标的特征，通过车载Point Grey相机采集了实际场景中的3000张图像，并使用YOLO_mark标注工具进行标注。最后将2个数据集混合作为实验的数据集，并随机选择80%的图片作为训练集，其余20%的图片用于测试。

4.2 训练

网络参数设置如表1所示。实验条件为：Intel Xeon(R)Silver4110CPU，主频2.1GHz，内存32GB；NVIDIA GeForce GTX 1080Ti GPU。

表1 Dense-YOLO的初始化参数

Tab.1 Initialization parameters of Dense-YOLO

图片尺寸	批处理	动量	学习率	权重衰减项	训练次数
480×480	8	0.9	0.001	0.0005	50000

在确定训练参数后对YOLOv3和改进的YOLOv3分别进行训练。训练次数设置为50000次，学习率在35000次后下降到0.0001，在45000次后下降到0.00001，使损失函数进一步收敛。同时利用旋转、调整饱和度和色调等方法对数据集中的图像进行增强和扩充，以增强模型的鲁棒性。

4.3 评价指标

现介绍评价模型有效性的相关指标。

精确率和召回率定义为

$$P=\frac{TP}{TP+FP} \tag{6}$$

$$R=\frac{TP}{TP+FN} \tag{7}$$

式中：TP表示真阳性数；FP表示假阳性数；FN表示假阴性数。

交并比IoU是定义目标检测精度的标准。通过计算预测边界框与真实边界框的重叠率，对模型的性能进行评估。IoU的定义为

$$IoU=\frac{s_{\mathrm{i}}}{s_{\mathrm{u}}} \tag{8}$$

式中：s_{i}为预测边界框与真实边界框的交集面积；s_{u}为2个边界框的并集面积。

IoU重叠阈值采用与KITTI相同的标准，IoU对轿车的检测阈值为0.7，对行人、骑行者和卡车的检测阈值为0.5。

平均精确度AP也用来评估模型的性能。AP即P–R曲线下的面积，其定义为

$$AP=\int_0^1 P(R)\mathrm{d}R \tag{9}$$

式中，$P(R)$为不同召回率所对应的精确率。

AP值越大，性能越好。而对于每一个类别的AP求均值即可得到平均精度均值mAP。

4.4　结果分析

4.4.1　障碍物检测结果分析

Dense-YOLO 与 YOLOv3 检测精度对比数据如表 2 所示。对比表 2 中 2 种检测算法的检测结果可知，本文提出的算法在测试数据集中各类障碍物检测的 *AP* 和 *mAP* 与 YOLOv3 相比分别提高了约 3%～5% 和 4%。其中，对于轿车的检测平均准确度最高，为 91.33%；对卡车的检测平均准确度最低，为 79.83%。

表 2　YOLOv3 和 Dense-YOLO 对 4 类障碍物检测结果对比

Tab.2　Comparison between the detection results of the four-class of obstacles by YOLOv3 and Dense-YOLO　%

算　法	*AP*				*mAP*
	轿车	行人	骑行者	卡车	
YOLOv3	88.48	82.92	77.41	75.37	81.05
Dense-YOLO	91.33	86.64	81.52	79.83	84.83

图 6 分别展示了测试集中对轿车、行人、骑行者、卡车这 4 类障碍物采用 Dense-YOLO 和 YOLOv3 这 2 种模型进行测试的精确率−召回率曲线。通过比较曲线下的面积，可以看出本文方法（Dense-YOLO）在 4 个类别中均取得了更好的检测效果。

图 7 是 YOLOv3 和本文方法障碍物检测结果对比图。图 7（a）选取 KITTI 数据集中的 2 个场景，左边是 YOLOv3 检测效果，右边是本文方法检测结果。从图 7 中可以看出，对于第一个场景，YOLOv3 漏检了一个骑行者；对于第二个场景，YOLOv3 漏检了一辆车，而本文方法没有出现这种情况。图 7（b）是在校园中采集的 2 个交通场景图，对于第一个场景，YOLOv3 漏检了一辆车、一个骑行者和 2 个行人等尺寸较小的障碍物；对于第二个场景，YOLOv3 漏检了 2 个行人。本文方法同样能准确地检测出 YOLOv3 漏检的障碍物。因此，与 YOLOv3 算法相比，本文方法无论是在 KITTI 数据集中还是实际交通场景中，都有良好的检测效果。

4.4.2　障碍物深度估计结果分析

定性分析：使用本研究方法对 KITTI 数据集进行深度估计，并将结果与来自 KITTI 数据集的真实值进行比较。图 8 给出了 KITTI 数据集中 2 个场景障碍物检测的结果和被检测障碍物的估计

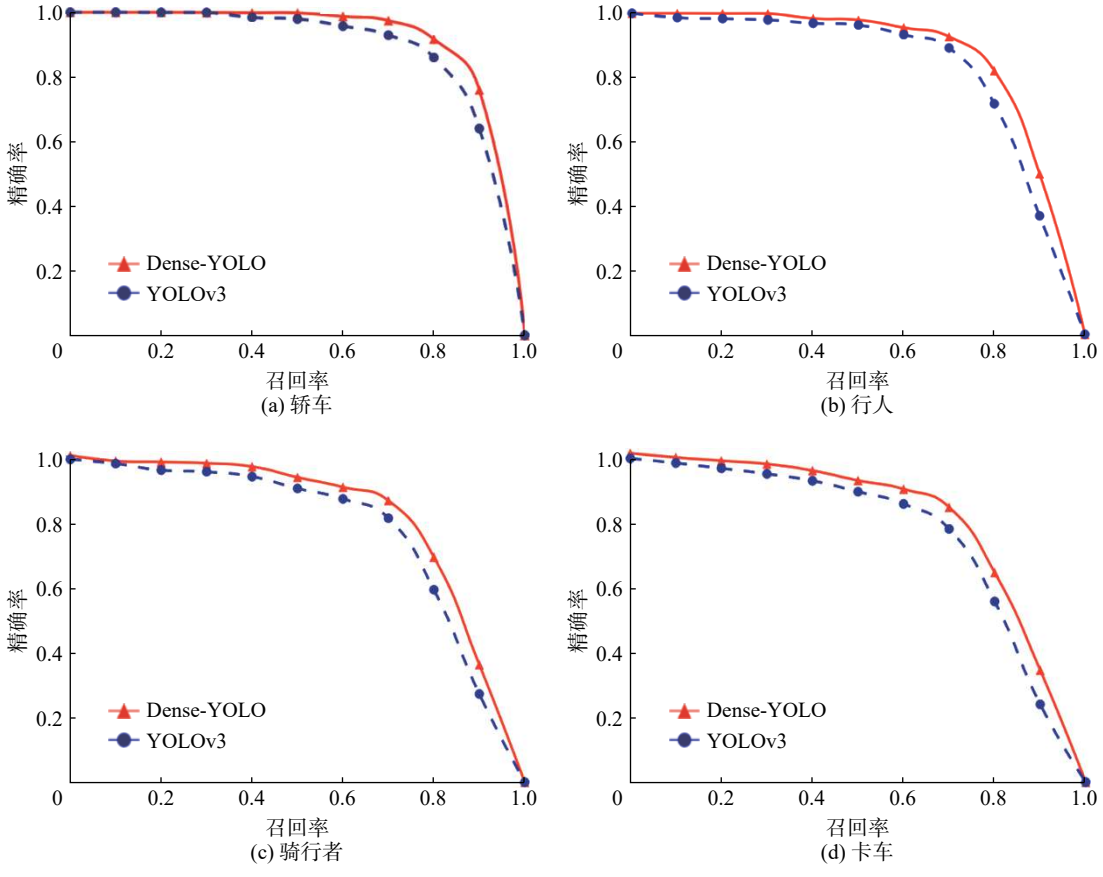


图 6　Dense-YOLO 和 YOLOv3 对 4 类障碍物的 *P-R* 曲线图

Fig.6　*P-R* curves using Dense-YOLO and YOLOv3 to detect the four classes of obstacles

YOLOv3 Dense-YOLO

(a) KITTI 数据集图片中的2个场景

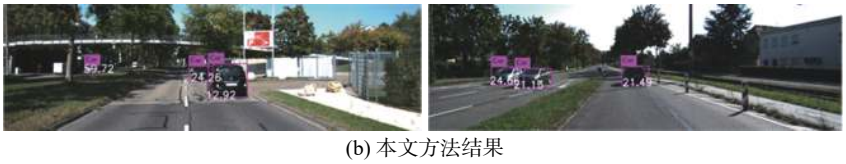
YOLOv3 Dense-YOLO

(b) 校园场景

图 7 障碍物检测结果：轿车、行人、骑行者、卡车

Fig.7 Results of obstacle detection: car, pedestrian, cyclist, truck

(a) 真实值



(b) 本文方法结果

(c) PSMNet 生成的视差图

图 8 方法对 KITTI 数据集的深度估计

Fig.8 Depth prediction of KITTI by the following method

深度和真实值。图 8(a)是障碍物的距离真实值，图 8(b)是 Dense-YOLO 检测结果以及本文方法所估计出的被测障碍物深度。从图中可以看出，Dense-YOLO 能准确地检测和定位到障碍物，并且能识别出障碍物的类别。在障碍物检测结果的基础上，本文的深度估计方法能够较好地估计出障碍物的距离，具体的对比结果如表 3 所示。

定量分析：对本研究方法估计的深度的相对误差进行了性能评价，相对误差定义为

$$e_r = \frac{|d_p - d_t|}{d_t} \times 100\% \qquad (10)$$

表 3 估计深度与真实值比较

Tab.3 Comparison between the estimated and ground-truth values of depth

编号	真实深度/m	估计深度/m	偏差/m	相对误差/%
1	56.65	59.72	3.07	5.42
2	24.91	24.26	0.65	2.61
3	12.92	12.78	0.14	1.10
4	24.02	24.66	0.64	2.66
5	20.62	21.13	0.51	2.47
6	21.01	21.49	0.48	2.28

式中：d_p为本方法估计出的障碍物深度；d_t为真实值。

表 3 给出了上述 2 个场景中被测目标的深度估计结果以及与真实值的比较。由表 3 可知，当障碍物深度真实值在 10 m 左右时，相对误差约为 1%；当障碍物深度真实值在 20 m 左右时，相对误差约为 2.4%；当障碍物深度真实值在 25 m 左右时，相对误差约为 2.7%。随着真实值的增加，相对误差也在增加。

5　结束语

为了满足自动驾驶车辆安全智能的需求，提出了一种交通场景中的障碍物检测和深度估计方法。首先，使用 DenseNet 网络对 YOLOv3 网络进行改进，得到一个新的障碍物检测网络（Dense-YOLO），对输入的图片进行检测和识别，得到障碍物的类别和边界框；然后，结合立体匹配模型 PSMNet 获取视差图，将目标框映射到视差图中，根据双目测距原理估计出障碍物的深度。在 KITTI 数据集和真实交通场景的实验结果表明，本文方法的检测精度优于 YOLOv3 算法，并且对被测障碍物的深度估计结果接近于真实值。本文方法对单个图像测试的时间成本分别约为 0.276 s。综上，本文方法基本满足自动驾驶场景中障碍物的检测与深度估计。

参考文献：

[1] MANCINI M, COSTANTE G, VALIGI P, et al. Fast robust monocular depth estimation for Obstacle Detection with fully convolutional networks[C]//Proceedings of 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems. Daejeon, South Korea: IEEE, 2016: 4296–4303.

[2] CIOCCA G, CUSANO C, SCHETTINI R. Image orientation detection using LBP-based features and logistic regression[J]. Multimedia Tools and Applications, 2015, 74(9): 3013–3034.

[3] LIANG C W, JUANG C F. Moving object classification using local shape and HOG features in wavelet-transformed space with hierarchical SVM classifiers[J]. Applied Soft Computing, 2015, 28: 483–497.

[4] WU Z F, XU Q, LI J A, et al. Passive indoor localization based on CSI and naive Bayes classification[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2018, 48(9): 1566–1577.

[5] ZHANG Y, LI B H, LU H C, et al. Sample-specific SVM learning for person re-identification[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 1278–1287.

[6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014: 580–587.

[7] GIRSHICK R. Fast R-CNN[C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015: 1440–1448.

[8] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149.

[9] 张泽苗, 霍欢, 赵逢禹. 深层卷积神经网络的目标检测算法综述 [J]. 小型微型计算机系统, 2019, 40(9): 1825–1831.

[10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016: 21–37.

[11] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 779–788.

[12] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 6517–6525.

[13] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement[EB/OL]. [2018-04-08]. https://arxiv.org/abs/1804.02767.

[14] CHANG J R, CHEN Y S. Pyramid stereo matching network[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 5410–5418.

[15] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 2261–2269.

[16] YANG L, WANG B Q, ZHANG R H, et al. Analysis on location accuracy for the binocular stereo vision system[J]. IEEE Photonics Journal, 2018, 10(1): 7800316.

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

（编辑：石　瑛）