第 58 卷 第 2 期/2021 年 1 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于双目视觉的三维车辆检测算法

于洁潇，张美琪*，苏育挺

天津大学电气自动化与信息工程学院，天津 300072

摘要 立体区域卷积神经网络(Stereo R-CNN)算法具有准确、高效的特点，在一定场景下的检测性能较好，但对于远景目标的检测仍有一定的提升空间。为了提升双目视觉算法的车辆检测精度，提出一种改进的 Stereo R-CNN 算法。该算法将确定性网络(DetNet)作为骨干网络，以增强网络对远景目标的检测；针对左右目视图的潜在关键点，建立了左右视图关键点一致性损失函数，以提高选取潜在关键点的位置精度，进而提高车辆的检测准确性。在 KITTI 数据集上的实验结果表明，本算法的性能优于 Stereo R-CNN，在二维、三维检测任务上的平均精度提升了 1%～3%。

关键词 机器视觉；三维目标检测；左右关键点一致性；车辆检测

中图分类号 TP39 **文献标志码** A **doi**：10.3788/LOP202158.0215004

Three-Dimensional Vehicle Detection Algorithm Based on Binocular Vision

Yu Jiexiao, Zhang Meiqi*, Su Yuting

School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, *China*

Abstract Stereo region-convolutional neural networks (Stereo R-CNN) algorithm has the characteristics of accuracy and efficiency. It has better detection performance in certain scenes, but there is still room for improvement in the detection of distant targets. In order to improve the vehicle detection accuracy of the binocular vision algorithm, an improved Stereo R-CNN algorithm is proposed in this paper. The algorithm uses deterministic network (DetNet) as the backbone network to enhance the network's detection of long-term targets; for the potential key points of the left and right eye views, the consistency loss function of the key points of the left and right views is established to improve the location accuracy of the potential key points, and then improve the accuracy of vehicle detection. Experimental results on the KITTI data set show that the performance of the algorithm is better than Stereo R-CNN, and the average accuracies of two-dimensional and three-dimensional detection tasks are improved by 1%–3%.

Key words machine vision; three-dimensional object detection; left-right keypoints consistency; vehicle detection

OCIS codes 150.0155; 330.1400; 100.6890

1 引 言

目标检测是计算机视觉领域的重要组成和研究对象，目前二维(2D)目标检测算法的准确率和检测速度均达到了较高的水平[1]，因此，三维(3D)目标检测逐渐成为视觉领域的重点研究对象[2]。对 3D 目标检测的研究在实际应用有重要意义，如无人驾驶技术中 3D 目标检测主要应用在对障碍物的识别阶段(感知阶段)，通过结合感知端采集的信息感知路况，以采取相应的制动措施。

目前 3D 目标检测算法大多是基于激光雷达(LiDAR)、双目摄像头和单目摄像头这三种设备。其中，LiDAR 是最常见的设备，且基于 LiDAR 的算法准确率极高。但 LiDAR 价格高昂、容易受天气变化尤其是雨雪天气的影响、容易伤害人眼[3]，难以在大众市场中进行推广。单目摄像头的价格相对较

收稿日期：2020-05-19；**修回日期**：2020-06-17；**录用日期**：2020-07-07

基金项目：国家自然科学基金(61572356)

* **E-mail**：zhangmeiqi@tju.edu.cn

低、且不受天气影响，但对于 3D 目标检测的误差较大。双目摄像头同时考虑了精度、成本和效率等因素，综合性优于前两种设备[4]，同时可以得到相对精确的深度值[3]。其中，基于快速区域卷积神经网络(Fast R-CNN)的三维目标提议(3DOP)[5]算法通过探究立体图像并将能量函数最小化，以生成 3D 目标候选区域。然后基于这些候选区域，用卷积神经网络(CNN)回归 3D 框的坐标和目标大小。Li 等[6]提出一种基于 2D 检测的 3D 检测算法，可在动态场景下追踪 3D 目标，并将动作轨迹和动态目标光速平差法(BA)相结合，构建了一种语义追踪系统。立体区域卷积神经网络(Stereo R-CNN)[7]是一种基于双目摄像头的 3D 目标检测算法，该算法用 101-残差网络(ResNet-101)和特征金字塔网络(FPN)作为骨干网络提取特征，扩展了 Faster R-CNN[8]，且在 KITTI 数据集上的准确率高于基于双目摄像头的检测算法。

传统基于视觉的目标检测算法包括维奥拉-琼斯检测器(Viola Jones detectors)[9-10]、梯度直方图(HOG)[11]、DPM(Deformable part-based model)[12]。维奥拉-琼斯检测器用滑窗进行检测，并采用积分图像计算 Haar 特征，用 Adaboost 算法进行特征选择，通过检测级联的方式降低计算开销。HOG 在图像的局部细胞单元上采集各像素点的梯度或边缘方向直方图，并对重叠的局部进行对比度归一化。DPM 算法扩展了 HOG，将检测任务分解成检测对象不同部分的集合。相比 CNN，传统检测算法不具有普适性，且对复杂多变的背景鲁棒性较弱[13]。端到端的检测算法如单步多框检测(SSD)[14]、YOLO(You only look once)算法[15]省去生成候选框的过程，提高了检测速度，但检测准确率不高。非端到端的算法如 R-CNN、Fast R-CNN 和 Faster R-CNN 算法通过区域建议生成候选区域，并进行特征提取、分类和回归[16]。其中，R-CNN 算法会发生信息丢失或重复提取特征，导致计算浪费；Fast R-CNN 算法解决了这些问题，但存在候选区域提取时间过长和端到端测试无法实现的问题[17]；而 Faster R-CNN 可解决上述所有问题。因此，基于 Faster R-CNN 的 Stereo R-CNN 算法在目标检测的准确性方面具有更明显的优势。

本文提出一种基于 Stereo R-CNN 的改进网络，为了提高 2D 远景检测的准确率，用具有更大感受野的确定型网络(DetNet)替换原骨干网络 ResNet-101；为了提高 3D 检测的准确率，增加右目图像的关键点，利用左右关键点的一致性，建立左右视图关键点一致性损失函数，修正关键点的位置。最后在 KITTI 数据集[18]上验证了本算法的有效性。

2 相关背景

Stereo R-CNN 是在 Faster R-CNN 基础上进行扩展，将检测设备从单目变为双目，将网络架构从单路变为双路，在 KITTI 数据集上的检测速度和精度较高，可为立体视觉的研究提供新思路。Stereo R-CNN 首先将输入的左目和右目图像传递到主干网络 ResNet-101 和 FPN[19]中，然后提取图像特征，便于后续处理；再将提取的图像特征传递到网络的核心——区域建议网络(RPN)，RPN 运用滑窗选择感兴趣区域(RoI)和非极大值抑制(NMS)选出最佳结果。实验使用双目摄像头，需对左右目视图进行处理，因此用双路 Stereo RPN 替换 RPN，生成左目和右目的 RoI 区域并对这两个区域进行 NMS 处理得到候选区域。其中，训练时选取 Stereo RPN 输入准确率前 2000 名的 RoI 候选区域，测试时选取 Stereo RPN 输入准确率前 300 名的 RoI 候选区域。

RoI Align 将 Stereo RPN 输入的特征图回归成所需的大小，以便进行立体回归，即输入到两个相连且伴有线性修正单元(ReLU)层的全连接层，以提取语义信息。立体回归包括 4 个子分支，分别用于预测物体的类别、立体框、维度以及视角角度。仅使用左目的特征图进行关键点预测，且只预测 4 个关键点用于立体估计。通过基于 R-CNN 的掩模(Mask R-CNN[20])预测的关键点进行 3D 框估计，然后将结果传递给稠密 3D 框对齐(Dense 3D box alignment)进行像素级别的精度匹配，对结果进行进一步修正。网络的具体流程如图 1 所示。

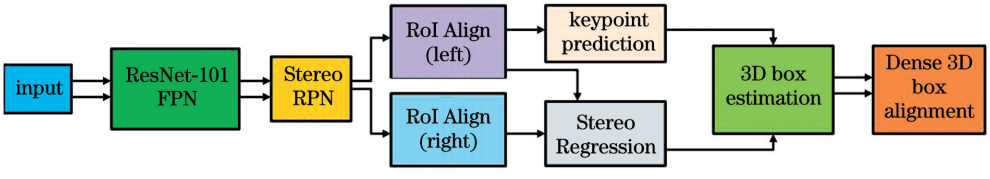

图 1 Stereo R-CNN 的流程图

Fig. 1 Flow chart of Stereo R-CNN

网络训练时的多任务损失函数可表示为

$$L=w_{\mathrm{cls}}^{\mathrm{p}}L_{\mathrm{cls}}^{\mathrm{p}}+w_{\mathrm{reg}}^{\mathrm{p}}L_{\mathrm{reg}}^{\mathrm{p}}+w_{\mathrm{cls}}^{\mathrm{r}}L_{\mathrm{cls}}^{\mathrm{r}}+w_{\mathrm{box}}^{\mathrm{r}}L_{\mathrm{box}}^{\mathrm{r}}+w_{\alpha}^{\mathrm{r}}L_{\alpha}^{\mathrm{r}}+w_{\mathrm{dim}}^{\mathrm{r}}L_{\mathrm{dim}}^{\mathrm{r}}+w_{\mathrm{key\text{-}l}}^{\mathrm{r}}L_{\mathrm{key\text{-}l}}^{\mathrm{r}},\tag{1}$$

式中，$w_{\mathrm{cls}}^{\mathrm{p}}$ 和 $L_{\mathrm{cls}}^{\mathrm{p}}$ 分别为 RPN 中的分类权重和损失，$w_{\mathrm{reg}}^{\mathrm{p}}$ 和 $L_{\mathrm{reg}}^{\mathrm{p}}$ 分别为 RPN 中回归任务的权重和损失，$w_{\mathrm{cls}}^{\mathrm{r}}$ 和 $L_{\mathrm{cls}}^{\mathrm{r}}$ 分别为 R-CNN 中的分类权重和损失，$w_{\mathrm{box}}^{\mathrm{r}}$ 和 $L_{\mathrm{box}}^{\mathrm{r}}$ 分别为 R-CNN 中框定任务的权重和损失，w_{α}^{r} 和 L_{α}^{r} 分别为 R-CNN 中的视角角度权重和损失，$w_{\mathrm{dim}}^{\mathrm{r}}$ 和 $L_{\mathrm{dim}}^{\mathrm{r}}$ 分别为 R-CNN 中的维度权重和损失，$w_{\mathrm{key\text{-}l}}^{\mathrm{r}}$ 和 $L_{\mathrm{key\text{-}l}}^{\mathrm{r}}$ 分别为 R-CNN 中左关键点的权重和损失。

3　网络模型

3.1　DetNet 提取特征

在 KITTI 数据集上进行 2D 检测时，Stereo R-CNN 对离双目摄像头较近车辆的检测精度较高，对远离双目摄像头的车辆容易遗漏。原因是 Stereo R-CNN 使用的骨干网络 ResNet-101 具有简易、高效等特点，但提取小物体（实验中的远景）的信息能力较弱。为了解决该问题，用 DetNet[21] 作为骨干网络。DetNet 保留了 ResNet 的阶段 1～阶段 4，改变阶段 5 并增加阶段 6。其中阶段 1～阶段 4 由残差模块组成，阶段 5 和阶段 6 如图 2 所示，由 A、B 瓶颈模块组成，A、B 的结构如表 1 所示。

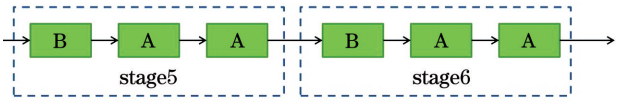

图 2　DetNet 的阶段 5 和 6

Fig. 2　Stages 5, 6 of DetNet

表 1　瓶颈模块 A、B 的具体结构

Table 1　Detail structure of bottleneck blocks A and B

Bottleneck block	Detail structure
A	keneral size 1×1 conv, 256 channels keneral size 3×3 conv, dilate2, 256 channels keneral size 1×1 conv, 256 channels
B	keneral size 1×1 conv, 256 channels keneral size 3×3 conv, dilate2, 256 channels + keneral size 1×1conv keneral size 1×1 conv, 256 channels

DetNet 在检测大物体和小物体时都能得到更高的准确率，虽然 DetNet 在 ResNet-101 的基础上增加了新的阶段，扩大了感受野，但增加的扩张瓶颈结构并不影响运行速度。同时，由于 DetNet 是基于 ResNet 的改进网络，能与原网络很好的融合。因此，实验将 ResNet-101 替换为 DetNet。

3.2　左右关键点一致性

左右视差一致性（Left-right disparity consistency）[22] 作为一个无监督的网络，其中心思想是利用左右目视图和校准信息之间的关系以及左右视差图的一致性，获得更准确的深度信息。通过将左目视图与右目视图进行匹配，从而判断视差图、得到深度图。

Stereo R-CNN 的 3D 检测受制于关键点的位置，而 Stereo R-CNN 中的关键点仅通过左目视图选取的，忽略了右目视图的信息。因此，实验增加了右目视图支路，获取右目视图对应的关键点，并对左目和右目视图的对应关键点进行匹配，以修正左目视图的关键点位置。为了引入左右关键点一致性，在网络中加入了关键点在右目特征图的选取，因此网络的多任务损失函数可表示为

$$L=w_{\mathrm{cls}}^{\mathrm{p}}L_{\mathrm{cls}}^{\mathrm{p}}+w_{\mathrm{reg}}^{\mathrm{p}}L_{\mathrm{reg}}^{\mathrm{p}}+w_{\mathrm{cls}}^{\mathrm{r}}L_{\mathrm{cls}}^{\mathrm{r}}+w_{\mathrm{box}}^{\mathrm{r}}L_{\mathrm{box}}^{\mathrm{r}}+w_{\alpha}^{\mathrm{r}}L_{\alpha}^{\mathrm{r}}+w_{\mathrm{dim}}^{\mathrm{r}}L_{\mathrm{dim}}^{\mathrm{r}}+\beta w_{\mathrm{key\text{-}l}}^{\mathrm{r}}L_{\mathrm{key\text{-}l}}^{\mathrm{r}}+\gamma w_{\mathrm{key\text{-}r}}^{\mathrm{r}}L_{\mathrm{key\text{-}r}}^{\mathrm{r}},\tag{2}$$

式中，$w_{\mathrm{key\text{-}l}}^{\mathrm{r}}$ 和 $w_{\mathrm{key\text{-}r}}^{\mathrm{r}}$ 分别为 R-CNN 中左、右关键点的权重，$L_{\mathrm{key\text{-}l}}^{\mathrm{r}}$ 和 $L_{\mathrm{key\text{-}r}}^{\mathrm{r}}$ 分别为 R-CNN 中左、右关键点的损失，β、γ 分别为左、右关键点的系数，且 $\beta+\gamma=1$。根据文献[23]，可通过不确定性加权得到每项损失的权重。本网络的结构如图 3 所示。

4　分析与讨论

4.1　实验设置

实验使用的数据集为 KITTI，该数据集可模拟驾驶所处的真实环境，主要用于评测立体图像、光流、视觉测距等[18]。数据集中的图像包括市区、乡村、高速公路等场景。实验使用 KITTI 系列的 3D Object Detection Evaluation 2017，共 7481 组样本，每组样本包括左右相机获得的图像、标签及标定信

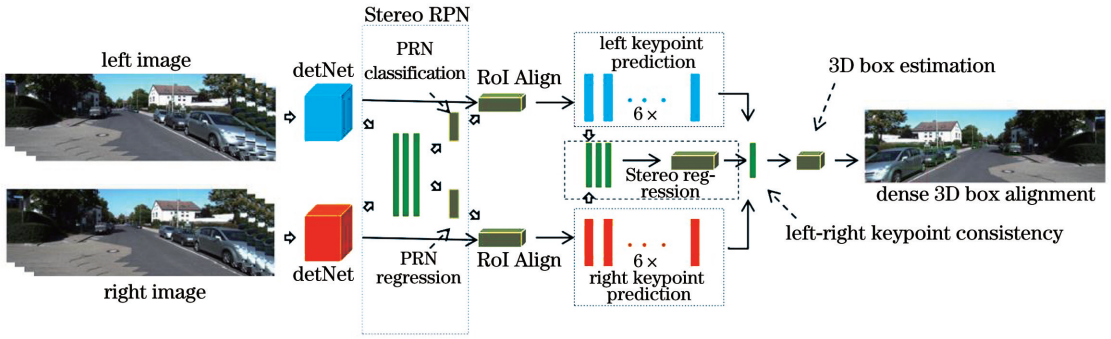

图 3　Stereo R-CNN 左右一致网络的流程

Fig. 3　Flow chart of Stereo R-CNN left and right consistency network

息。每张图像的大小为 1242 pixel×375 pixel，标签中包括目标的类别、遮挡程度、观察角度、2 维边界框、3 维尺寸及位置等信息。目标的类别包括汽车(car)、面包车(van)、卡车(truck)等。根据图像中目标的遮挡程度以及大小，将数据划分为简单(easy)、中等(mode)和困难(hard)三个等级。

实验中，将原始 7481 组数据随机并粗略分成数量相近的两组，其中一组用于训练，另一组用于测试。评估过程中，只考虑汽车这一标签，实验环境为 Ubuntu16.04，软件为 Pytorch 平台，显卡为 NVIDIA Corporation GP102[GeForce GTX 1080 Ti]。训练过程中，将 512 个采样的 RoI 区域作为一批进行处理，利用随机梯度下降(SGD)法进行训练，初始学习率为 0.001，每 5 个 epochs 将学习率变为原来的 0.1 倍。一共训练 20 epochs，大约耗时两天，参数 $\beta=0.8$、$\gamma=0.2$。

4.2　评估方法

实验的评估指标包括 2D 检测平均精度、3D 检测平均精度(包括鸟瞰检测平均精度、3D 边框检测平均精度)。各项指标对真正例(TP)的定义如下。

1) 2D 检测：当左目视图预测值与左目视图真实值之比大于给定阈值时，判定左目视图二维检测为 TP；当右目视图预测值与右目视图真实值之比大于给定阈值时，判定右目视图二维检测为 TP；当左右目图像选定框属于同一目标时，判定双目视觉二维检测为 TP。

2) 鸟瞰检测：俯视图的预测框与真实框的重叠面积与真实面积之比大于给定阈值时为 TP。

3) 3D 边框检测：3D 边框检测的预测框与真实框重叠体积之比大于给定阈值时为 TP，给定阈值一般用交并比(IoU)表示。

4.3　实验结果及分析

1) 2D 目标检测。通过引入 DetNet，扩大检测的感受野，既能检测远景目标，又能检测近景目标。得到交并比为 0.7 的二维检测准确率，如表 2 所示，Stereo R-CNN、DetNet only 和 Ours 分别表示原 Stereo R-CNN 的 2D 检测、将骨干网络替换为 DetNet 的 2D 检测、将骨干网络替换为 DetNet 的同时保持左右关键点一致性的 2D 检测结果。图 4(a1)～图 4(a3)分别为 Stereo R-CNN 的左目、右目、鸟瞰检测结果，图 4(b1)～图 4(b3)分别为本网络的左目、右目、鸟瞰检测结果。可以发现，引入 DetNet 对于适中和困难模式的识别率有轻微提升，对于容易模式没有明显优势，且引入左右关键点一致性对于 2D 检测的结果没有明显影响。整体来说，本算法的检测效果更佳。

表 2　2D 检测的平均准确率

Table 2　Average precision of 2D detection　　unit：%

Algorithm	Left			Right			Stereo		
	Easy	Mode	Hard	Easy	Mode	Hard	Easy	Mode	Hard
Faster R-CNN	98.57	89.01	71.54	–	–	–	–	–	–
Stereo R-CNN	98.73	88.48	71.26	98.71	88.50	71.28	98.53	88.27	71.14
DetNet only	98.01	88.52	73.44	98.80	88.32	72.67	98.63	88.44	73.16
Ours	98.50	88.92	73.63	98.34	88.25	73.10	98.44	88.76	73.18

2) 3D 目标检测。为了便于立体框的回归，且不影响运行速度，只对左目视图进行 3D 检测。实

图 4　不同算法的检测结果。(a)Stereo R-CNN;(b)本算法

Fig. 4　Detection results of different algorithms. (a) Stereo R-CNN; (b) our algorithm

验结果表明,引入右关键点后,利用左右关键点一致性修正关键点的位置精度,能提升 3D 检测的准确性。表 3 为检测交并比阈值为 0.5 时,3 次 3D 检测实验的平均值。表 4 为交并比为 0.7 时,3 次实验的平均值。可以发现,跟 2D 检测相似,引入 DetNet 作为骨干网络对 3D 检测的提升效果不明显,特别是对于简单模式没有明显的提升,对于适中和困难模式有轻微提升。引入左右关键点一致性后,鸟瞰评估以及 3D 边框评估的平均准确率在简单、适中、困难三个模式下,约有 1%～3%的提升。

表 3　3D 检测平均准确率(IoU 为 0.5)

Table 3　Average precision of 3D detection (IoU is 0.5)

unit: %

Algorithm	Bird's eye view			3D box		
	Easy	Mode	Hard	Easy	Mode	Hard
3DOP[5]	55.04	41.25	34.55	46.04	34.63	30.09
Stereo R-CNN	87.13	74.11	58.93	85.84	66.28	57.24
DetNet only	87.02	74.26	58.98	85.12	65.85	57.81
Ours	88.63	75.90	59.46	86.66	66.39	58.18

表 4　3D 检测平均准确率(IoU 为 0.7)

Table 4　Average precision of 3D detection (IoU is 0.7)

unit: %

Algorithm	Bird's eye view			3D box		
	Easy	Mode	Hard	Easy	Mode	Hard
3DOP[5]	12.63	9.49	7.59	6.55	5.07	4.10
Stereo R-CNN	68.50	48.30	41.47	54.11	36.69	31.07
DetNet only	67.42	49.21	42.68	52.12	34.85	31.81
Ours	69.63	49.90	43.46	55.66	36.39	32.14

5　结　　论

针对 Stereo R-CNN 检测远景时容易发生漏检问题,提出了用 DetNet 作为骨干网络,以提升网络对远景的检测效果;为了提高 3D 检测准确率,增加右目图像的关键点,对左右目相应的关键点进行一致性匹配,以修正关键点的位置,从而提高 3D 目标检测的准确性。在 KITTI 数据集上的实验结果表明,相比其他算法,本算法能在一定程度上提高检测的准确性。

参　考　文　献

[1] Wang W F, Jin J, Chen J M. Rapid detection algorithm for small objects based on receptive field block[J]. Laser & Optoelectronics Progress, 2020, 57(2): 021501.
王伟锋, 金杰, 陈景明. 基于感受野的快速小目标检测算法[J]. 激光与光电子学进展, 2020, 57(2): 021501.

[2] Jiang J H, Chang C C, Chen T S. An efficient Huffman decoding method based on pattern partition and look-up table[C]//Fifth Asia-Pacific Conference on Communications, and Fourth Optoelectronics and Communications Conference on Communications, October 18-22, 2001, Beijing, China. New York: IEEE, 1999, 2: 904-907.

[3] Kahn S, Bockholt U, Kuijper A, et al. Towards precise real-time 3D difference detection for industrial applications[J]. Computers in Industry, 2013, 64(9): 1115-1128.

[4] Huang P C, Jiang J Y, Yang B. Research status and progress of binocular stereo vision [J]. Optical Instruments, 2018, 40(4): 81-86.
黄鹏程, 江剑宇, 杨波. 双目立体视觉的研究现状及进展[J]. 光学仪器, 2018, 40(4): 81-86.

[5] Chen X Z, Kundu K, Zhu Y K, et al. 3D object proposals using stereo imagery for accurate object class detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(5):

1259-1272.

[6] Li P L, Qin T, Shen S J. Stereo vision-based semantic 3D object and ego-motion tracking for autonomous driving[M]// Ferrari V, Hebert M, Sminchisescu C, et al. Computer Vision-ECCV 2018. Lecture Notes in Computer Science. Springer, Cham, 2018, 11206: 664-679.

[7] Li P L, Chen X Z, Shen S J. Stereo R-CNN based 3D object detection for autonomous driving[C]// 2019 IEEE/Conference on Computer Vision and Pattern Recognition, June 15-20, 2019, Long Beach, CA, USA. New York: IEEE, 2019: 7636-7644.

[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] Viola P A, Jones M J. Rapid object detection using a boosted cascade of simple features[C]// 2001 IEEE Conference on Computer Vision and Pattern Recognition, December 8-14, 2001, Kauai, HI, USA. New York: IEEE, 2001: I.

[10] Viola P, Jones M J. Robust real-time face detection [J]. International Journal of Computer Vision, 2004, 57(2): 137-154.

[11] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 20-25, 2005, San Diego, CA, USA. New York: IEEE, 2005: 886-893.

[12] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C]// 2008 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2008, Anchorage, AK, USA. New York: IEEE, 2008: 1-8.

[13] Li Q W, Zhou Y Q, Ma Y P, et al. Salient object detection method based on binocular vision[J]. Acta Optica Sinica, 2018, 38(3): 0315002.
李庆武, 周亚琴, 马云鹏, 等. 基于双目视觉的显著性目标检测方法[J]. 光学学报, 2018, 38(3): 0315002.

[14] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer Vision-ECCV 2016. Lecture Notes in Computer Science. Cham: Springer, 2016, 9905: 21-37.

[15] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[EB/OL]. [2020-05-02]. https://arxiv.org/abs/1506.02640.

[16] Qiao T, Su H S, Liu G H, et al. Object detection algorithm based on improved feature extraction network[J]. Laser & Optoelectronics Progress, 2019, 56(23): 231008.
乔婷, 苏寒松, 刘高华, 等. 基于改进的特征提取网络的目标检测算法[J]. 激光与光电子学进展, 2019, 56(23): 231008.

[17] Zhou B, Li R X, Shang Z H, et al. Object detection algorithm based on improved faster R-CNN[J]. Laser & Optoelectronics Progress, 2020, 57(10): 101009.
周兵, 李润鑫, 尚振宏, 等. 基于改进的 Faster R-CNN 目标检测算法[J]. 激光与光电子学进展, 2020, 57(10): 101009.

[18] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI, USA. New York: IEEE, 2012: 3354-3361.

[19] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 936-944.

[20] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN [C]// 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2980-2988.

[21] Li Z M, Peng C, Yu G, et al. DetNet: design backbone for object detection[M]// Ferrari V, Hebert M, Sminchisescu C, et al. Computer Vision-ECCV 2018. Lecture Notes in Computer Science. Springer, Cham, 2018, 11213: 339-354.

[22] Godard C, Aodha O M, Brostow G J, Unsupervised monocular depth estimation with left-right consistency[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, June 21-26, 2017, Honolulu, HI. New York: IEEE, 2017: 6602-6611.

[23] Cipolla R, Gal Y, Kendall A. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics[C]// 2018 IEEE Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 7482-7491.