



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象 图形学报

2021
01
VOL.26

ISSN1006-8961
CN11-3758/TB





第26卷第1期 (总第297期)

2021年1月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿, 均视为同意在本刊网站及CNKI等全文数据库出版, 所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用, 所有内容, 未经许可, 不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@radi.ac.cn
电 话 010-58887035
网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian

Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@radi.ac.cn
Telephone 010-58887035
Website www.cjig.cn

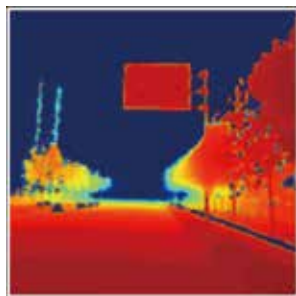
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元



面向智能驾驶的平行视觉感知：基本概念、框架与应用
(第0067页)



Edge-guided GAN: 边界信息引导的深度图像修复
(第0186页)



利用边缘计算的多车协同激光雷达SLAM(第0218页)

序言	王飞跃, 张云泉 I
编者按	II

综述

面向智能驾驶测试的仿真场景构建技术综述	
任秉韬, 邓伟文, 白雪松, 李江坤, 纵瑞雪, 朱冰, 丁娟	0001
自动驾驶软件测试技术研究综述	
冯洋, 夏志龙, 郭安, 陈振宇	0013
强化学习的自动驾驶控制技术研究进展	
潘峰, 鲍泓	0028
自动驾驶地图的数据标准比较研究	
詹骄, 郭迟, 雷婷婷, 屈宜琪, 吴杭彬, 刘经南	0036
视觉感知的端到端自动驾驶运动规划综述	
刘旒菲, 胡学敏, 陈国文, 刘士豪, 陈龙	0049

平行驾驶

面向智能驾驶的平行视觉感知：基本概念、框架与应用	
李轩, 王飞跃	0067
平行视觉的基本框架与关键算法	
张慧, 李轩, 王飞跃	0082

目标检测与跟踪

自动驾驶场景的尺度感知实时行人检测	
徐歆恺, 马岩, 钱旭, 张龔	0093
实时视觉目标跟踪与视频对象分割多任务框架	
李瀚, 刘坤华, 刘嘉杰, 张晓晔	0101
提升预测框定位稳定性的视频目标检测	
郝腾龙, 李熙莹	0113
道路结构特征下的车道线智能检测	
张翔, 唐小林, 黄岩军	0123
适用全速域大曲率路径的自动驾驶跟踪算法	
张龔, 郑颖, 鲍泓	0135
伪3D卷积神经网络与注意力机制结合的疲劳驾驶检测	
庄员, 戚湧	0143
基于眼部自商图—梯度图共生矩阵的疲劳驾驶检测	
潘剑凯, 柳政卿, 王秋成	0154

自动驾驶场景感知与仿真

结合局部平面参数预测的无监督单目图像深度估计	
周大可, 田径, 杨欣	0165
深度纯追随的拟人化无人驾驶转向控制模型	
单云霄, 黄润辉, 何泽, 龚志豪, 景民, 邹雪松	0176
Edge-guided GAN: 边界信息引导的深度图像修复	
刘坤华, 王雪辉, 谢玉婷, 胡坚耀	0186
引入概率分布的深度神经网络贪婪剪枝	
胡骏, 黄启鹏, 刘嘉昕, 刘威, 袁淮, 赵宏	0198

高精地图构建与SLAM

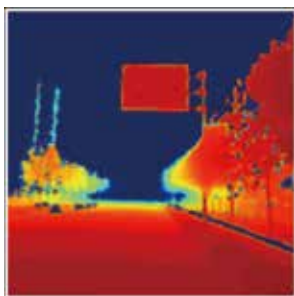
开放道路中匹配高精度地图的在线相机外参标定	
廖文龙, 赵华卿, 严骏驰	0208
利用边缘计算的多车协同激光雷达SLAM	
崔明月, 钟仕鹏, 刘思瑶, 李博洋, 吴成昊, 黄凯	0218

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Parallel visual perception for intelligent driving: basic concept, framework and application (P0067)



Edge-guided GAN: a depth image inpainting approach guided by edge information (P0186)



Cooperative LiDAR SLAM for multi-vehicles based on edge computing (P0218)

Review

- Technologies of virtual scenario construction for intelligent driving testing
Ren Bingtao, Deng Weiwen, Bai Xuesong, Li Jiangkun, Zong Ruixue, Zhu Bing, Ding Juan 0001
- Survey of testing techniques of autonomous driving software
Feng Yang, Xia Zhilong, Guo An, Chen Zhenyu 0013
- Research progress of automatic driving control technology based on reinforcement learning
Pan Feng, Bao Hong 0028
- Comparative study on data standards of autonomous driving map
Zhan Jiao, Guo Chi, Lei Tingting, Qu Yiqi, Wu Hangbin, Liu Jingnan 0036
- Review of end-to-end motion planning for autonomous driving with visual perception
Liu Yifei, Hu Xuemin, Chen Guowen, Liu Shihao, Chen Long 0049

Parallel Driving

- Parallel visual perception for intelligent driving: basic concept, framework and application
Li Xuan, Wang Feiyue 0067
- The basic framework and key algorithms of parallel vision
Zhang Hui, Li Xuan, Wang Feiyue 0082

Object Detection and Tracking

- Scale-aware EfficientDet: real-time pedestrian detection algorithm for automated driving
Xu Xinkai, Ma Yan, Qian Xu, Zhang Yan 0093
- Multitask framework for video object tracking and segmentation combined with multi-scale inter-frame information
Li Han, Liu Kunhua, Liu Jiajie, Zhang Xiaoye 0101
- Video object detection method for improving the stability of bounding box
Hao Tenglong, Li Xiying 0113
- Intelligent detection of lane based on road structure characteristics
Zhang Xiang, Tang Xiaolin, Huang Yanjun 0123
- Autonomous vehicle tracking algorithm for high curvature path in full speed range
Zhang Yan, Zheng Ying, Bao Hong 0135
- Driving fatigue detection based on pseudo 3D convolutional neural network and attention mechanisms
Zhuang Yuan, Qi Yong 0143
- Fatigue driving detection based on ocular self-quotient image and gradient image co-occurrence matrix
Pan Jiankai, Liu Zhengqing, Wang Qiucheng 0154

Scene Perception and Simulation of Autonomous Driving

- Unsupervised monocular image depth estimation based on the prediction of local plane parameters
Zhou Dake, Tian Jing, Yang Xin 0165
- Human-like steering model for autonomous driving based on deep pure pursuit method
Shan Yunxiao, Huang Runhui, He Ze, Gong Zhihao, Jing Min, Zou Xuesong 0176
- Edge-guided GAN: a depth image inpainting approach guided by edge information
Liu Kunhua, Wang Xuehui, Xie Yuting, Hu Jianyao 0186
- Greedy pruning of deep neural networks fused with probability distribution
Hu Jun, Huang Qipeng, Liu Jiaxin, Liu Wei, Yuan Huai, Zhao Hong 0198

High-Definition Map Construction and SLAM

- Online extrinsic camera calibration based on high-definition map matching on public roadway
Liao Wenlong, Zhao Huaqing, Yan Junchi 0208
- Cooperative LiDAR SLAM for multi-vehicles based on edge computing
Cui Mingyue, Zhong Shipeng, Liu Siyao, Li Boyang, Wu Chenghao, Huang Kai 0218

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2021)01-0082-11

论文引用格式: Zhang H, Li X and Wang F Y. 2021. The basic framework and key algorithms of parallel vision. Journal of Image and Graphics, 26(01): 0082-0092 (张慧, 李轩, 王飞跃. 2021. 平行视觉的基本框架与关键算法. 中国图象图形学报, 26(01): 0082-0092) [DOI:10.11834/jig.200400]

平行视觉的基本框架与关键算法

张慧^{1,2}, 李轩^{3,4}, 王飞跃¹

1. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室, 北京 100190;

2. 腾讯科技(北京)有限公司, 北京 100080; 3. 鹏城实验室, 深圳 518055; 4. 北京理工大学, 北京 100081

摘要: 目的 随着计算机与人工智能的快速发展, 视觉感知技术突飞猛进。然而, 以深度学习为主的视觉感知方法依赖于大规模多样性的数据集, 因此, 本文提出了基于平行学习的视觉分析框架——平行视觉, 它通过大量精细标注的人工图像来给视觉算法补充足够的图像数据, 从而将计算机变成计算智能的“实验室”。方法 首先人工图像系统模拟实际图像中可能出现的成像条件, 利用系统内部参数自动得到标注信息, 获取符合要求的人工图像数据; 然后使用预测学习设计视觉感知模型, 利用计算实验方法在人工图像系统生成的大量图像数据上进行各种实验, 方便地研究复杂环境条件等困难场景对视觉感知模型的影响, 使一些实际中的不可控因素转变为可控因素, 增加视觉模型的可解释性; 最后通过指示学习反馈优化模型参数, 利用视觉感知模型在实际场景下存在的困难来指导其在人工场景的训练, 以实际与人工虚实互动的方式, 在线学习和优化视觉感知模型。由于已经有大量研究人员致力于构建人工场景并生成大量虚拟图像, 因此本文采用已构建的这些人工场景图像, 并对实际场景图像进行翻转、裁剪、缩放等数据扩充, 然后以计算实验和预测学习为重点, 开展了相关的应用实例研究。结果 在 SYNTHIA (synthetic collection of imagery and annotations), Virtual KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute) 和 VIPER (visual perception benchmark) 数据集上进行的大量实验表明, 本文方法能够有效地克服数据集分布差异对模型泛化能力的影响, 性能优于同期最好的方法, 比如在 SYNTHIA 数据集上检测和分割性能分别提升了 3.8% 和 2.7%。结论 平行视觉是视觉计算领域的一个重要研究方向, 通过与深度学习的结合, 将推动越来越多的智能视觉系统发展成熟并走向应用。

关键词: 计算机视觉; 平行学习; 平行视觉; 视觉感知模型; 实例分割; 目标检测

The basic framework and key algorithms of parallel vision

Zhang Hui^{1,2}, Li Xuan^{3,4}, Wang Feiyue¹

1. The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese

Academy of Sciences, Beijing 100190, China; 2. Tencent Technology (Beijing) Company, Limited, Beijing 100080, China;

3. Peng Cheng Laboratory, Shenzhen 518055, China; 4. Beijing Institute of Technology, Beijing 100081, China

Abstract: Objective Computer vision makes the camera and computer the “eyes” of the computer, which can have the abilities of segmentation, classification, recognition, tracking and decision-making. In recent years, computer vision technology has been widely used in intelligent transportation, unmanned driving, robot navigation, intelligent video monitoring, and many other fields. At present, the camera has become the most commonly used sensing equipment in automatic driving and smart cities, generating massive image and video data. We can realize real-time analysis and processing of these data only by relying on computer vision technology. We can detect all kinds of objects in real time and obtain their position and

收稿日期: 2020-07-20; 修回日期: 2020-10-22; 预印本日期: 2020-10-29

基金项目: 国家自然科学基金项目 (61876011, U1811463)

Supported by: National Natural Science Foundation of China (61876011, U1811463)

motion states accurately from the image video. However, the actual scene has a very high complexity. Many complex factors interweave together, which poses a great challenge to the visual computing system. At present, computer vision technology is mainly based on the deep learning method through large-scale data-driven mechanisms. Sufficient data are needed due to the heavy dependence of its training algorithm mechanism on datasets. However, collecting and labeling large-scale image data from actual scenes are time-consuming and labor-intensive tasks, and usually, only small-scale and limited diversity of image data can be obtained. For example, Microsoft common objects in context (MS COCO), a popular dataset used for instance segmentation tasks, has a size of about 300 000 and mainly 91 categories. Expressing the complexity of reality and simulate the real situation is difficult. The model trained on the limited dataset will lack practical significance, because the dataset is not large enough to represent the real data distribution and cannot guarantee the effectiveness of practical application. **Method** The theory of social computing and parallel systems is proposed based on artificial systems, computational experiments, and parallel execution (ACP). The ACP methodology plays an essential role in modeling and control of complex systems. A virtual artificial society is constructed to connect the virtual and the real world through parallel management. On the basis of the existing facts, artificial system is used to model the behavior of complex systems by using advanced computing experiments and then analyze its behavior and interact with reality to obtain a better operating system than reality. To address the bottleneck of deep learning in the field of computer vision, this paper proposes parallel vision, a visual analysis framework based on parallel learning. Parallel vision is an intelligent visual perception framework that is an extension of the ACP methodology into the computer vision field. In the framework of parallel vision, large-scale realistic artificial images can be obtained easily to give support to the vision algorithm with enough well-labeled image data. In this way, the computer can be turned into a “laboratory” of computational intelligence. First, the artificial image system simulates the imaging conditions that may appear in the actual image, uses the internal parameters of the system to automatically obtain the annotation information, and obtains the required artificial images. Then, we use the predictive learning method to design the visual perception model, and then we use the computational experiment method to conduct experiments. Various experiments are conducted on a rich supply of image data generated in the artificial image system. Studying the influence of difficult scenes such as complex environmental conditions on the visual perception model is convenient; thus, some uncontrollable factors in practice can be transformed into controllable factors, and the interpretability of the visual model is increased. Finally, we use prescriptive learning method to optimize model parameters. The difficulty of the visual perception model in the actual scene can be used to guide the model training in the artificial scene. We learn and optimize the visual perception model online through virtual-real interaction. This paper also conducted an application case study to preliminarily demonstrate the effectiveness of the proposed framework. This case can work over synthetic images with accurate annotations and real images without any labels. The virtual-real interaction guides the model to learn useful information from synthetic data while keeping consistent with real data. We first analyze the data distribution discrepancy from a probabilistic perspective and divide it into image-level and instance-level discrepancies. Then, we design two components to align these discrepancies, i. e., global-level alignment and local-level alignment. Furthermore, a consistency alignment component is proposed to encourage the consistency between the global-level and the local-level alignment components. **Result** We evaluate the proposed approach on the real Cityscapes dataset by adapting from virtual SYNTHIA (synthetic collection of imagery and annotations), Virtual KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute), and VIPER (visual perception benchmark) datasets. Experimental results demonstrate that it achieves significantly better performance than state-of-the-art methods. **Conclusion** Parallel vision is an important research direction in the field of visual computing. Through combination with deep learning, more and more intelligent vision systems will be developed and applied.

Key words: computer vision; parallel learning; parallel vision; visual perception model; instance segmentation; object detection

0 引言

计算机视觉是让摄像机、电脑这些科技设备成

为计算机的“眼睛”,让其可以具有人类的双眼所具有的分割、分类、识别、跟踪和判别决策等功能。近年来,计算机视觉技术广泛应用于智能交通、无人驾驶、机器人导航和智能视频监控等诸多领域(Wiley

和 Lucas, 2018; 张慧 等, 2017; Luo 等, 2020)。摄像机成为自动驾驶以及智慧城市中最常用的感知设备 (Luo 等, 2018, 2020), 可以产生海量的图像视频数据, 必须依靠计算机视觉技术才能实现对这些数据的实时分析与处理, 从图像视频中实时准确地检测出各类目标, 获取它们的位置和运动状态。但是实际场景具有极高的复杂性, 许多复杂因素交织在一起, 给视觉计算系统带来了严峻考验 (Goyette 等, 2014)。

当前计算机视觉技术主要是以深度学习方法为主, 通过大规模数据驱动的机制, 挖掘数据中蕴含的潜在规律。由于其训练算法机制对数据集的重度依赖, 需要有足够的数据。但是, 从实际场景中采集和标注大规模图像数据非常费时耗力, 通常只能获得小规模且多样性受限的图像数据集。比如针对实例分割任务的通用数据集 MS COCO (Microsoft common objects incontext) (Lin 等, 2014), 样本数在 30 万左右, 主要有 91 个类别, 很难表达出现实的复杂程度。Yuille 和 Liu (2018) 从组合学的角度解释了真实世界图像的“组合爆炸”。

如果没有指数意义上的大数据集, 就很难模拟真实情况。而在有限的数据集上训练/测试出来的模型, 会缺乏现实意义, 因为数据集不够大, 代表不了真实的数据分布, 无法保证实际应用时的有效性。

中国科学院自动化研究所王飞跃研究员于 2004 年提出基于 ACP (A 指人工系统 artificial system; C 指计算实验 computational experiments; P 指虚实系统的平行执行 parallel execution) 方法的社会计算和平行系统理论 (王飞跃, 2004), 构建了一个虚拟的人工社会, 以平行管理的方式连接起虚拟和现实两个世界。在对已有事实认识的基础上, 利用先进计算手段, 借助人工系统对复杂系统的行为进行“实验”, 进而对其行为进行分析, 虚实交互, 得出比“现实”更优的运行系统。

随着平行系统理论不断丰富和完善, 其在智能交通 (吕宜生 等, 2019; Zhang 等, 2018)、智能医疗 (王飞跃 等, 2017)、视觉感知 (Wang 等, 2017; Zhang 等, 2019)、智慧教育 (Wang 等, 2019) 和无人驾驶 (刘腾 等, 2019) 等领域取得了良好效果。针对当前深度学习在计算机视觉领域存在的瓶颈, 本文提出基于平行学习的视觉分析框架, 平行学习由描述学习、预测学习与指示学习构成 (李力 等, 2017), 用于解决在复杂场景下存在的标注数据难以获取、人工

标注不准确、模型选择和优化依赖技巧等问题。采用已构建的人工图像系统来进行实际数据的扩增和人工图像生成, 然后以计算实验与预测学习为重点, 使用预测学习方法得到视觉感知模型, 并通过指示学习反馈优化模型参数。通过上述平行学习方法, 使数据形成闭环, 持续不断迭代优化和提升视觉算法的精准度。

1 平行视觉的理论框架与基本方法

为了解决进一步解决复杂环境中视觉感知模型的科学难题, 本文提出如图 1 所示的平行视觉分析框架, 实现基于实际图像分布的数据扩充和人工场景数据生成, 并开展基于数据驱动的计算实验, 设计和优化视觉算法。基于实际和人工之间的虚实互动, 引入平行学习, 持续反馈优化视觉感知模型。

1.1 人工图像系统

为了解决实际图像多样性不足和标注困难等问题, 本文提出构建基于软件定义的人工图像系统。接下来具体说明获取人工图像的基本方法。正是实际数据的手工注释高昂成本激励了人工图像的研发, 使用已有开源仿真模拟器或商业游戏引擎 (例如: Unity、3DS MAX、OpenGL 和 Google 3D 等) 可以完成人工图像的获取。具体来说, 人工图像由许多要素构成, 通过现场调研、克隆和视频采集等方式收集相关数据, 使用 3 维建模软件以及开源模型资源包, 目的在于提高建模精度和速度, 做到高效和逼真的建模过程。在完成场景建模后, 需要为场景设置虚拟摄像机, 模拟实际摄像机物理参数来生成人工图像序列, 整个阶段将自动生成精确、多标注的视觉数据, 标注图像可以保障视觉算法采用监督的模式进行训练。详细标注包括: 目标检测、目标跟踪、语义分割、实例分割、深度信息和光流等。上述工作为有效获取大规模多样性人工图像奠定了基础。此外, 研究发现人工图像中具有灵活设置成像条件的潜力, 可以改变的成像条件包括 (但不限于): 1) 摄像头位置 (角度和高度); 2) 天气、季节和照明变换等; 3) 目标物体状态 (速度、加速度和运动轨迹等)。

依据所述方法, 研究者获取了 SYNTHIA (synthetic collection of imagery and annotations) (Ros 等, 2016)、Virtual KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute) (Gaidon 等, 2016)

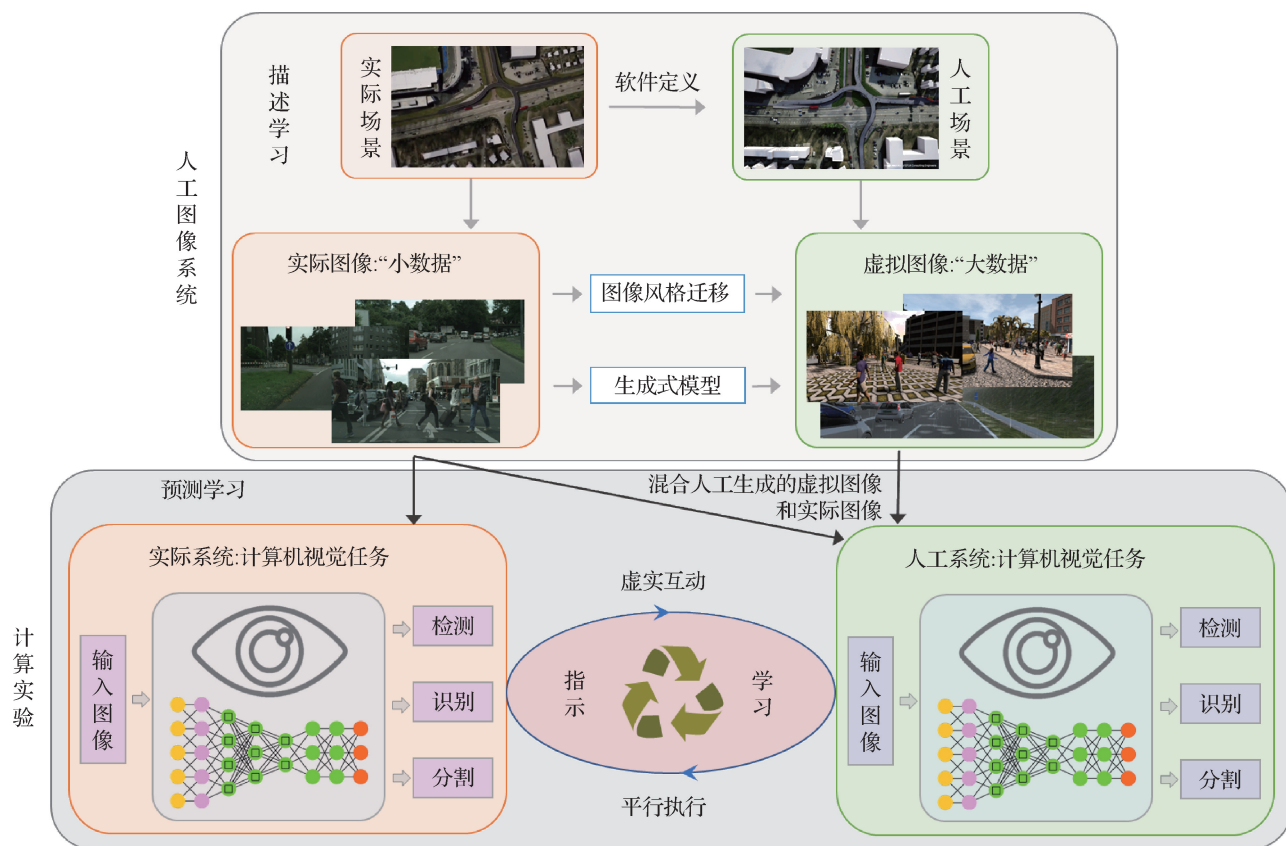


图1 平行视觉分析框架

Fig. 1 The framework of parallel vision

和 ParallelEye (Li 等, 2018) 等人工图像。研究结果表明人工图像在语义分割、目标跟踪和目标检测等视觉任务上能发挥重要作用。其中, SYNTHIA 是在虚拟城市中收集的人工图像集 (SYNTHIA-Rand 和 SYNTHIA-Seqs 系列), 自动生成语义分割标注。SYNTHIA-Rand 是在城市中由虚拟摄像头随机收集的图像, 共 13 400 帧; SYNTHIA-Seqs 是城市中产生的连续视频序列, 分为 4 个季节, 共 200 000 帧。Virtual KITTI 采用克隆的方式完成人工图像获取, 包含多种标注信息和不同成像条件。ParallelEye 采用现场调研和开源地图信息获取了北京市中关村地区的路网信息, 使用快速建模方式模拟构建了北京城市场景, 自动获取 ParallelEye 数据集。该数据集由 40 251 帧图像组成, 共包含 6 类标注信息和 8 种成像条件 (包括: 雨天、雾天、摄像头角度和光照变化)。人工图像的优势如下: 1) 与现有实际图像数据相比, 人工图像收集更加高效, 生成的标注信息也更加准确; 2) 人工图像能够轻易地扩大规模和多样性, 可以更为全面地对视觉算法进

行训练和评估。

1.2 计算智能与预测学习

通过人工图像系统实现数据扩充和数据生成后, 进一步研究基于计算实验的视觉感知模型设计与优化方法, 从而实现智能视觉感知模型。计算实验是在人工场景的基础上进行的一系列模拟实验。由于资源的有限性以及环境的复杂多变, 视觉感知方法很难在各种不同的场景下开展全面有效的实验评估, 因此在实际应用中可能发生意外意想不到的结果。若要增强视觉模型的鲁棒性, 必须在复杂多变的环境下进行全面充分的实验。通过利用人工场景来模拟复杂多变的实际环境, 在各种各样的人工场景中进行计算实验, 来优化和评估视觉算法。该计算实验的核心是预测学习, 即用机器给真实环境建模, 仿真预测感兴趣的未知信息。

首先, 设计更有效的视觉感知模型结构。对于检测、识别和分割, 利用预测学习方法设计新的深度学习模型, 提高模型的特征表达能力和计算效率, 使模型具有更高的精度。当前性能比较好的视觉检测

模型有 Faster R-CNN (region-convolutional neural network)、RetinaNet 和 Mask R-CNN 等 (Zhang 等, 2020)。Ren 等人 (2015) 提出区域建议网络 (region proposal network, RPN), 并且把 RPN 和 Fast R-CNN 融合到一个统一的网络 (称为 Faster R-CNN)。Lin 等人 (2017) 提出了一个新的损失函数来解决类别不平衡的问题, 该函数通过减少容易分类的样本的权重, 使得模型在训练时更专注难分类的样本, 从而改善模型的优化方向, 并提出新的单阶段检测器 RetinaNet 来验证该损失函数的有效性。He 等人 (2017) 基于 Faster R-CNN 架构提出了新的卷积网络 Mask R-CNN。它在 Faster R-CNN 的基础上, 延伸出了一个分割分支, 能够从单幅图像中同时检测目标边框及像素区域。但是这些模型都是在 Pascal VOC (pattern analysis, statistical modelling and computational learning visual object classes) 或 MS COCO 数据集上进行训练的, 其中的图像数据从网络上搜集得到。由于网络空间和实际物理空间并不等价 (Torralba 和 Efros, 2011), 并且网络空间中存在困难场景样本缺失的问题, 因此训练的模型很难直接应用于实际场景。人工图像系统可以仿真模拟各种各样复杂变化的环境条件, 比如交通事故图像数据、恶劣天气和低照度条件的图像数据等, 从而生成大规模多样性的数据集, 作为实际数据集的补充。

然后利用计算实验方法在人工图像系统生成的大量图像数据上进行各种实验, 方便地研究复杂环境条件等困难场景对视觉感知模型的影响, 大大减少实际中的不可控因素, 增加视觉模型的可解释性和鲁棒性。利用人工场景数据集和实际场景数据集联合训练视觉感知模型, 提高它在复杂场景下的性能。在进行模型评估时, 可以先在人工场景数据集上做实验评估, 衡量所提出的视觉感知方法在不同场景下的性能; 然后再把基于人工场景数据集优化的模型应用在实际测试集上, 并根据实际数据集上存在的性能缺陷来反馈优化视觉模型, 最终实现复杂场景下的智能视觉感知与理解。

1.3 基于平行执行的闭环优化

在计算实验的基础上, 将视觉模型在实际场景与人工场景中平行执行, 通过实际与人工之间的虚实互动, 来不断优化提升视觉模型的性能。由于环境的复杂多变性, 在实际应用时视觉模型会存在很多困难和挑战, 需要在应用过程中不断调节和优化,

基于这种虚实互动的平行执行方法来不断优化视觉模型的性能, 使其在复杂的场景中能够进行有效的视觉感知与理解。

平行执行的最大特色把实际系统和人工系统紧密地联合在一起, 形成一个闭环。在丰富的虚拟对象模型数据的基础上, 人工系统可以构建各种有实际意义的场景, 其中不仅包含一般实际场景, 还包含很多困难场景, 困难场景即实际场景中很少发生或很难收集复现的实际场景。通过人工系统和实际系统的交互优化, 提升视觉模型的主动感知能力。在平行执行步骤, 基于指示学习的思想, 通过视觉感知模型在实际场景下存在的困难来指导其在人工场景的训练, 以实际与人工虚实互动的方式, 在线学习和优化视觉模型, 不断提高视觉模型在当前场景中的运行效果, 同时为应对未来场景做好充分的准备。

2 基于平行视觉框架的应用实例

在平行视觉框架下, 本文提出了基于虚实交互的视觉模型, 如图 2 所示, 用于解决图像标签缺失时的实例分割任务的问题。训练实例分割模型需要大规模多样性的数据集, 但是手动标注像素级标签信息非常费时耗力, 并且由于图像细节模糊, 在低照度条件下人类标注逐像素标签是非常困难的。因此使用由计算机自动生成标注信息的虚拟数据集和未标注的真实图像来完成模型的训练。首先利用人工场景来模拟复杂挑战的实际场景, 获取大规模多样性的虚拟数据集, 自动生成详细且精确的标注信息。然后将人工“大数据”和实际“小数据”相结合, 对视觉模型进行虚实互动的学习与评估。基于这种虚实交互的方法, 指导模型更关注实际场景数据和人工场景数据之间的公共特征, 从而减少数据分布差异带来的不利影响, 有效提升视觉模型的检测和分割效果。

2.1 基于人工图像系统的数据生成

平行视觉框架的人工图像系统旨在生成大量具有丰富多样性的人工场景图像并对实际场景图像进行扩增。通过生成虚拟图像来弥补实际场景中不充分的图像, 从而在减少金钱和人力投资的同时提供精确的标注信息。一种构建虚拟数据集的普遍方法是在游戏引擎 (例如 Unity3D) 中模仿实际场景, SYNTHIA (Ros 等, 2016) 和 Virtual KITTI (Gaidon 等, 2016) 都是采用这种方式构建的, 在不同的视

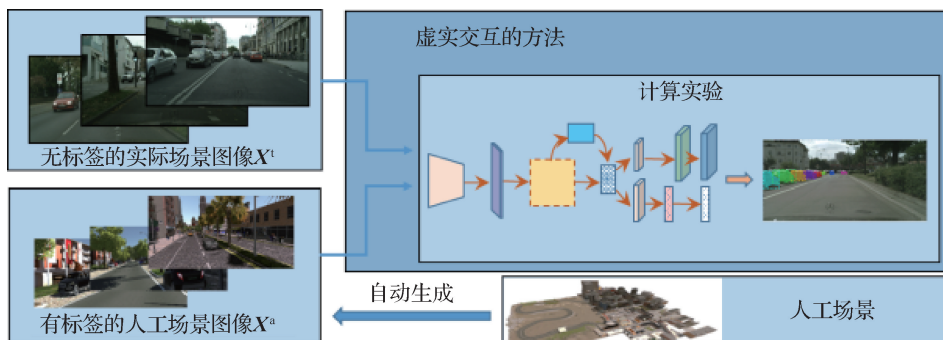


图2 基于平行视觉框架的实例分割模型概述

Fig. 2 Overview of the proposed model based on parallel perception framework

角、季节、天气和照明条件下提供各种各样的图像并自动生成精确的标注信息。Richter 等人(2017)在玩游戏时直接从计算机游戏中收集图像,其中包含从视频游戏中提取的大量帧,以及用于高级和低级视觉任务的准确标签。由于已有大量研究人员致力于构建人工场景并生成大量虚拟图像,因此本文采用已构建的这些人工场景图像,并对实际场景图像进行翻转、裁剪和缩放等数据扩充。

2.2 基于平行视觉的实例分割模型

应用实例使用平行视觉来改进基于区域建议的实例分割模型 Mask R-CNN,使其能解决实际场景图像标签缺失时的实例分割任务问题。计算实验步骤包含两个平行的模型,并且两个模型之间的参数彼此共享,如图3所示。由于实际场景图像没有精确的边界框和像素级标签,因此只能根据输入的人工场景图像计算 Mask R-CNN 的相关检测和分割损失。训练

出的模型很容易在虚拟图像上过拟合,而无法很好地在实际场景中泛化。为了解决这个问题,需要利用实际场景的图像特征来反馈优化视觉模型。来自人工场景和实际场景的图像组成分批的训练数据用于微调所提出的模型。其中人工场景图像用于训练模型以实现准确的检测性能,实际场景图像用于帮助模型专注于人工场景和实际场景之间的公共特征信息。基于指示学习的方法,本文设计了3个新的对齐模块:全局级别对齐模块、局部级别对齐模块和一致性对齐模块。通过这3个对齐模块的共同训练,指导学习到的模型参数更关注于人工场景和实际场景之间的公共特征,从而使得模型获得更好的性能。

全局级别对齐模块和局部级别对齐模块是两个域分类器,用于判断输入数据是来自于哪个数据域,然后采用对抗的思想训练网络,以致于域分类器不能区分特征是来自于人工场景还是实际场景,使得

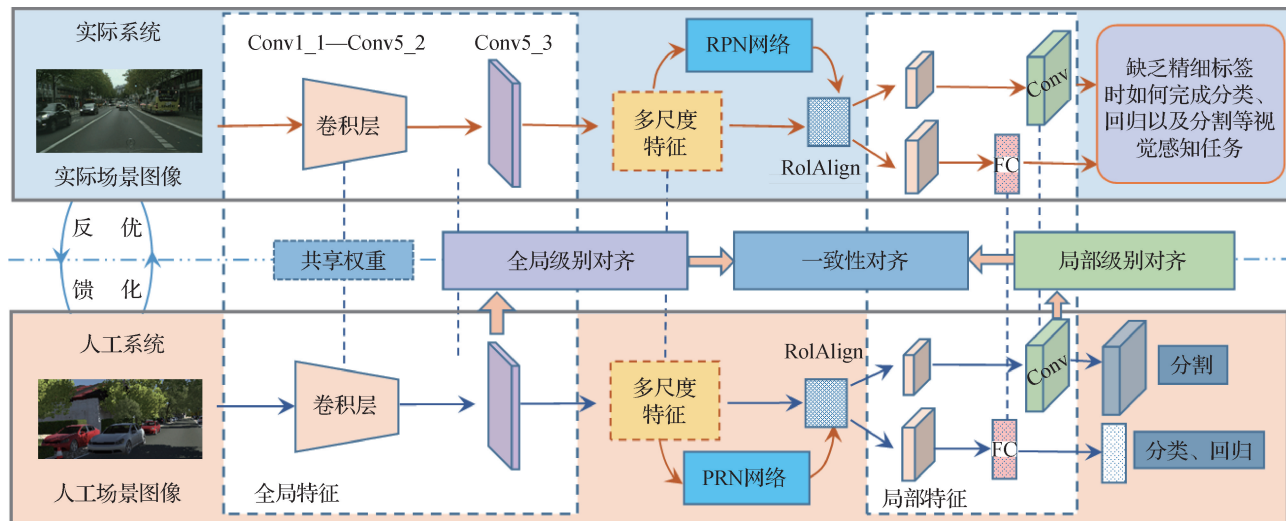


图3 基于平行视觉框架的实例分割模型

Fig. 3 The proposed model based on parallel perception framework

获得的特征空间是两个域的公共特征。一致性对齐模块用于保证全局和局部级别对齐模块预测结果的一致性。在详细介绍对齐模块之前,首先介绍经典的实现对抗思想的方法——梯度反转层 (gradient reverse layer, GRL)。该层在正向过程中保持输入不变,在反向梯度传递过程中将其乘以某个负常数 $-\alpha$ 来反转梯度的符号。对于全局级别对齐模块,本文使用域分类器来减少两个不同数据域之间的分布差异。GRL 层之前的部分类似于对抗网络的生成器,使学习的不变特征对于域分类器而言尽可能地难以区分,保证模型在实际场景中具有良好的泛化能力,GRL 层之后的部分类似于对抗网络的判别器,判别图像特征是来自于人工场景还是实际场景。局部级别对齐模块也是使用类似的方式来学习不变的局部区域特征,以便在给定包含目标区域的情况下,无法区分该区域是来自人工场景还是实际场景。一致性对齐方式用于鼓励不同类型的特征表示之间的一致性。

在训练和测试阶段都使用人工系统中的预测模型来输出检测和像素级分割结果,新增的对齐模块仅在训练阶段用于反馈优化视觉模型。具体来说,在训练阶段,优化预测模型的参数来最小化它们在训练数据集上的检测和分割损失,同时优化 GRL 层之前的特征映射部分的参数来最小化预测模型的损失并最大化 GRL 层之后的域判别器的损失。

2.3 实验结果与分析

本文使用 SYNTHIA (Ros 等,2016),Virtual KITTI (Gaidon 等,2016)和 VIPER (visual perception benchmark) (Richter 等,2017)分别作为人工场景数据集,以及 Cityscapes (Cordts 等,2016)训练集作为训练阶

段无标签的实际场景数据集,并用 Cityscapes 验证集来验证虚实交互的实例分割方法。所有实验使用平均精度 (average precision, AP) 来评估目标检测和分割的性能,训练阶段使用随机梯度下降来更新模型的参数,其动量为 0.9,权重衰减因子为 0.000 5,学习率设置为 0.000 1,梯度反传因子 α 设置为 0.1。

2.3.1 各个模块的有效性分析

在反馈优化的过程中,本文提出了 3 个新的对齐模块:全局级别对齐模块、局部级别对齐模块和一致性对齐模块。通过这 3 个对齐模块的共同训练,指导学习到的模型参数更关注于人工场景和实际场景之间的公共特征,从而使得模型很容易泛化到实际场景中。首先分析了基于平行视觉的实例分割模型中 3 个对齐模块的有效性,如表 1 所示,其中“GLA”表示全局级别对齐方式,“LLA”表示局部级别对齐方式,“CA”表示一致性对齐方式,“/”前后的数值表示检测/分割的性能,mAP (mean average precision) 为 IoU (intersection over union) 设置成 0.50 时计算的所有类别精度的平均值。平行视觉优化前的基准方法检测精度达到 30.3%,分割精度达到 22.5%。从表 1 中可以看出,所有引入不同级别对齐的优化方法都比不使用它的方法表现更好,这证明了所提出的基于平行视觉的实例分割模型的有效性。例如,使用全局级别对齐或局部级别对齐的方法检测精度分别达到 32.3% 和 31.9%。因为虚拟的 SYNTHIA 与实际的 Cityscapes 数据集之间存在分布差异,如果不使用任何对齐方式,模型很容易在虚拟图像上过拟合。因此,不同级别的对齐方式在减少数据分布差异方面起着关键作用。当使用所有的对齐模块

表 1 不同对齐方式的实验结果
Table 1 Experimental results of different-level alignment

方法									
GLA	LLA	CA	汽车	公交车	摩托车	自行车	行人	骑车的人	mAP
			45.5/39.7	43.7/42.1	8.9/6.0	24.3/9.3	37.9/27.9	21.4/10.0	30.3/22.5
✓			47.3/40.5	40.6/39.2	16.4/11.0	24.9/10.1	38.0/27.5	25.8/10.2	32.3/23.1
	✓		46.3/39.8	42.2/41.0	14.7/11.9	24.9/9.7	38.4/28.0	24.8/11.9	31.9/23.7
✓	✓		46.7/40.1	44.2/42.4	17.1/14.5	25.7/12.1	37.8/27.7	29.6/ 14.5	33.5/25.2
	✓	✓	48.0/41.4	42.1/40.7	15.1/11.1	25.1/12.3	37.6/27.2	29.0/10.7	32.8/23.9
✓	✓	✓	48.7/41.9	44.7/43.7	16.7/12.6	26.2/12.6	37.4/27.1	30.3/13.5	34.1/25.2

注:加粗字体表示每列最优结果。

时,基于平行视觉的实例分割模型可以获得最佳性能,检测精度可以达到 34.1%,分割精度达到 25.2%,比基准方法分别提高了 3.8% 和 2.7%。

2.3.2 各个数据集上的实验结果

为了对本文方法进行全面评估,在 SYNTHIA 和 Cityscapes 上进行了更多实验,如表 2 所示。在虚拟的 SYNTHIA 数据集上训练的实例分割模型(表示为“S only”)无法很好地应用在实际场景中,其总体检测性能从 44.0% (如果在真实的 Cityscapes 数据集上进行训练,表示为“C only”)下降到 30.3%。分割性能也存在同样的现象,“C only”模型的 mAP 为 35.0%,而“S only”模型的 mAP 为 22.5%。通过使用本文提出的基于平行视觉的实例分割模型,检测和分割 mAP 分别达到了 34.1% 和 25.2%,这比“S only”模型分别提高了 3.8% 和 2.7%,缩小了“S only”和“C only”模型之间的性能差距。

表 2 SYNTHIA 和 Cityscapes 数据集的实验结果

Table 2 The results on SYNTHIA and Cityscapes

方法	mAP/%	
	检测	分割
S only	30.3	22.5
C only	44.0	35.0
本文	34.1	25.2

注:加粗字体表示最优结果。

在 Virtual KITTI 和 Cityscapes 上的实验结果如表 3 所示。基准模型 Mask R-CNN 对于汽车进行检测和分割的 AP_{50} (IoU 为 0.5 时得到的平均精度)分别为 46.9% 和 42.6%。具有全局级别对齐的方法将 AP_{50} 分别提高了 3.7% 和 2.4%。当进一步增加

表 3 Virtual KITTI 和 Cityscapes 数据集的实验结果

Table 3 The results on Virtual KITTI and Cityscapes

不同方式	基准方法	本文方法				
		/%				
全局级别对齐	✓	✓				✓
局部级别对齐		✓	✓	✓	✓	✓
一致性对齐				✓	✓	
检测: AP_{50}	46.9	50.6	52.2	47.3	48.5	54.3
分割: AP_{50}	42.6	45.0	46.1	42.3	43.5	47.4

注:加粗字体表示最优结果。

局部级别对齐时,本文方法将 AP_{50} 分别提高了 5.3% 和 3.5%。使用所有对齐模块的方式可以获得最佳的性能,在实例分割任务上的 AP_{50} 达到 47.4%,比平行视觉优化前的基准方法高 4.8%。在目标检测任务上,比基准方法提高了 7.4%。

为了进一步验证基于平行视觉的实例分割模型的有效性和鲁棒性,本文使用 VIPER 数据集进行了实验,如表 4 所示。在检测和分割任务中,基准方法的 mAP 分别为 31.3% 和 26.5%。使用全局级别对齐的方法分别达到了 32.0% 和 26.9%。当进一步使用局部级别对齐时,本文方法分别达到了 33.6% 和 27.4%。这表明本文方法可以有效地减少不同层次的数据分布差异。当全局级别对齐、局部级别对齐和一致性对齐一起使用时,mAP 分别提升至 34.1% 和 29.2%,比基准方法分别提高了 2.8% 和 2.7%。这些结果表明,基于平行视觉的实例分割模型可以显著提升模型的检测和分割性能。

表 4 VIPER 和 Cityscapes 数据集的实验结果

Table 4 The results on VIPER and Cityscapes

不同方式	基准方法	本文方法				
		/%				
全局级别对齐	✓	✓				✓
局部级别对齐		✓	✓	✓	✓	✓
一致性对齐				✓	✓	
检测:mAP	31.3	32.0	33.6	31.8	32.2	34.1
分割:mAP	26.5	26.9	27.4	26.6	27.3	29.2

注:加粗字体表示最优结果。

进一步将本文方法与最新的弱监督和领域适应方法进行了比较,如表 5 所示。所有方法均使用虚拟的 SYNTHIA 数据集,除了 Weakly-Sup (Saleh 等, 2017) 不使用虚拟图像, Saleh 等人 (2018) 使用 GTA5 (Grand Theft Auto V) 作为背景类, VEIS (virtual environment for instance segmentation) 作为前景。Weakly-Sup (Saleh 等, 2017) 依赖于使用真实图像的类别标签,相比之下,本文方法可以处理不包含任何真实图像标注信息的场景,借助不同级别的对齐优化方法来获得更好的分割性能。ROAD-Net (Chen 等, 2018) 通过使用目标导向的蒸馏模块和空间感知的适应模块来解决显著的分布差异,本文方法从特征层面应用不同的对齐方法,实现了 8.3% 的提

表 5 与域自适应和弱监督方法的比较

Table 5 Comparison with domain adaptation and weakly-supervised methods

方法	汽车	公交车	摩托车	自行车	行人	骑车的人	mIoU
Weakly-Sup(Saleh 等,2017)	41.2	16.6	15.7	14.9	16.2	6.1	18.5
Dilation Frontend(Yu 和 Koltun,2016)	47.3	3.9	0.1	0.0	51.1	1.5	17.3
FCNs in the wild(Hoffman 等,2016)	54.0	3.2	0.2	0.6	51.2	3.8	18.8
Curriculum(Zhang 等,2017)	43.2	20.7	0.7	13.1	47.1	8.2	22.2
ROAD-Net(Chen 等,2018)	67.8	14.5	7.0	23.8	44.5	16.6	29.0
Saleh(Saleh 等,2018)	76.2	31.6	18.1	9.8	52.2	28.5	36.1
本文	63.9	15.8	24.6	43.8	51.1	24.7	37.3

注:加粗字体表示最优结果。

升。Saleh 等人(2018)将 Mask R-CNN 产生的前景掩膜与 DeepLab 语义分割网络产生的逐像素预测相结合,达到了 36.1% 的 mIoU(mean IoU),比本文方法低 1.2%。

3 思考与展望

3.1 模型可解释性

自 2012 年以来,计算机视觉领域不断涌现出很多激动人心的研究成果,例如,人脸识别、目标识别与分类等方面的性能已接近甚至超过人类视觉系统,这主要得益于深度学习技术的发展。但是,将深度学习模型应用于视觉任务时还缺乏足够的理论支撑,学习到的模型可解释性较弱。本文提出了基于平行学习的视觉分析框架,通过人工图像系统构建的人工场景来模拟外界的动态变化信息,并评估其对视觉模型参数的影响,尽管它可以在一定程度上增加模型的可解释性,但是仍然具有局限性,对于如何选择模型、如何确定模型的深度以及深度学习的本质等基本问题还没有给出很好的解释。该问题在智能交通场景下的自动驾驶尤为突出,基于深度学习的视觉模型难以理解其中依据,存在安全隐患,尤其是发生事故后的责任限定。而且,加上了对抗扰动的输入可以误导自动驾驶汽车的感知系统,使其在分类道路标志时出错,从而可能造成灾难性的后果。例如腾讯科恩安全实验室发现:在道路特定位置贴上几个贴纸,可以让处在自动驾驶模式的汽车并入反向车道。因此,必须进一步完善深度学习理论,为设计模型结构、加速模型训练、提高模型性能、因果关系以及可解释性等提供指导。

3.2 动态平行视觉

当前计算机视觉的很多工作主要关注静态场景,比较缺乏对动态场景的处理,例如人脸识别大部分也是基于静态场景来展开研究的。尽管动态目标的跟踪、检测、分析、行为的识别与理解等相关工作也在研究,但并没有考虑整个场景的动态特性,还没有上升到一个系统化的水平。基于动态场景的视觉任务是基于静态图像的视觉任务的直接扩展,本文提出的平行视觉分析框架可以推广应用于动态视觉。动态视觉存在的很大问题是视频图像的采集困难。视频图像一般是在户外(或室内,但是采集条件比较差)获取的,通常没有用户的配合,所以视频图像经常会有很大的光照、姿态变化,还会有各种各样的遮挡和模糊等,导致视频图像质量不高。平行视觉分析框架的人工图像系统可以利用人工场景模拟复杂挑战的实际环境,提供大量质量较高的视频数据,作为实际视频图像的补充。

3.3 多源视觉任务

智能车辆的车载传感器通常由摄像机、激光雷达和 GPS 等组成。摄像机能够感知来自周围世界的颜色和纹理信息,然而,它们的探测范围有限,在有限的光照或恶劣的天气条件下表现不佳;激光雷达提供精确的距离信息,并且能够探测到小物体,它在夜间也能很好地工作,但不提供颜色信息。因此,基于这些多源数据的平行视觉能够提高视觉模型的鲁棒性与准确性。

4 结 论

本文介绍了平行视觉的基本思想,并提出基于

平行学习的视觉分析框架,首先使用描述学习解决人工图像生成和扩增,然后使用预测学习解决视觉感知模型的选择、设计和优化,并通过指示学习进行虚实交互,反馈优化模型参数。通过上述平行学习方法,使数据形成闭环,持续不断迭代优化和提升视觉算法的准确率。本文概述系统理论框架部分已有相关工作,并通过对已有工作调研分析给出具体思考与该领域未来研究展望。需要指出的是,本文的应用实例研究还处于初步验证阶段,在视觉计算领域的平行视觉理论还需要进一步发展和普及,有待更完备的理论、算法和软件平台的建立和使用,需要更多的研究和工程人员的关注,并实际应用于视觉计算的各个领域。

参考文献 (References)

- Chen Y H, Li W and Van Gool L. 2018. ROAD: reality oriented adaptation for semantic segmentation of urban scenes//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE: 7892-7901 [DOI: 10.1109/CVPR.2018.00823]
- Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, Franke U, Roth S and Schiele B. 2016. The cityscapes dataset for semantic urban scene understanding//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE: 3213-3223 [DOI: 10.1109/CVPR.2016.350]
- Gaidon A, Wang Q, Cabon Y and Vig E. 2016. Virtual worlds as proxy for multi-object tracking analysis//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE: 4340-4349 [DOI: 10.1109/CVPR.2016.470]
- Goyette N, Jodoin P M, Porikli F, Konrad J and Ishwar P. 2014. A novel video dataset for change detection benchmarking. IEEE Transactions on Image Processing, 23 (11): 4663-4679 [DOI: 10.1109/TIP.2014.2346013]
- He K M, Gkioxari G, Dollár P and Girshick R. 2017. Mask R-CNN//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy; IEEE: 2980-2988 [DOI: 10.1109/ICCV.2017.322]
- Hoffman J, Wang D, Yu F and Darrell T. 2016. FCNs in the wild: pixel-level adversarial and constraint-based adaptation [EB/OL]. [2020-06-20]. <https://arxiv.org/pdf/1612.02649.pdf>
- Li L, Lin Y L, Cao D P, Zheng N N and Wang F Y. 2017. Parallel learning—a new framework for machine learning. Acta Automatica Sinica, 43(1): 1-8 (李力, 林懿伦, 曹东璞, 郑南宁, 王飞跃. 2017. 平行学习——机器学习的一个新型理论框架. 自动化学报, 43(1): 1-8) [DOI: 10.16383/j.aas.2017.y000001]
- Li X, Wang K F, Tian Y L, Yan L, Deng F and Wang F Y. 2018. The paralleleye dataset: a large collection of virtual images for traffic vision research. IEEE Transactions on Intelligent Transportation Systems, 20(6): 2072-2084 [DOI: 10.1109/ITITS.2018.2857566]
- Lin T Y, Goyal P, Girshick R, He K M and Dollár P. 2017. Focal loss for dense object detection//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy; IEEE: 2980-2988 [DOI: 10.1109/ICCV.2017.324]
- Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P and Zitnick C L. 2014. Microsoft COCO: common objects in context//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland; Springer: 740-755 [DOI: 10.1007/978-3-319-10602-1_48]
- Liu T, Wang X, Xing Y, Gao Y, Tian B and Chen L. 2019. Research on digital quadruplets in cyber-physical-social space-based parallel driving. Chinese Journal of Intelligent Science and Technology, 1(1): 40-51 (刘腾, 王晓, 邢阳, 高玉, 田滨, 陈龙. 2019. 基于数字四胞胎的平行驾驶系统及应用. 智能科学与技术学报, 1(1): 40-51) [DOI: 10.11959/j.issn.2096-6652.201902]
- Lyu Y S, Chen Y Y, Jin J C, Li Z J, Ye P J and Zhu F H. 2019. Parallel transportation: virtual-real interaction for intelligent traffic management and control. Chinese Journal of Intelligent Science and Technology, 1(1): 21-33 (吕宜生, 陈圆圆, 金峻臣, 李镇江, 叶佩军, 朱凤华. 2019. 平行交通: 虚实互动的智能交通管理与控制. 智能科学与技术学报, 1(1): 21-33) [DOI: 10.11959/j.issn.2096-6652.201908]
- Luo G Y, Yuan Q, Zhou H B, Cheng N, Liu Z H, Yang F C, and Shen X S. 2018. Cooperative vehicular content distribution in edge computing assisted 5G-VANET. China Communications, 15(7): 1-17 [DOI: 10.1109/CC.2018.8424578]
- Luo G Y, Zhang H, He H B, Li J L and Wang F Y. 2020. Multiagent adversarial collaborative learning via mean-field theory [EB/OL]. [2020-07-05]. <https://ieeexplore.ieee.org/document/9238422>
- Luo G Y, Zhou H B, Cheng N, Yuan Q, Li J L, Yang F C and Shen X S. 2019. Software defined cooperative data sharing in edge computing assisted 5G-VANET [EB/OL]. [2020-07-05]. <https://ieeexplore.ieee.org/document/8897045>
- Ren S Q, He K M, Girshick R and Sun J. 2015. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]
- Richter S R, Hayder Z and Koltun V. 2017. Playing for benchmarks//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy; IEEE: 2213-2222 [DOI: 10.1109/ICCV.2017.243]
- Ros G, Sellart L, Materzynska J, Vazquez D and Lopez A M. 2016. The synthia dataset: a large collection of synthetic images for semantic segmentation of urban scenes//Proceedings of 2016 IEEE Confer-

- ence on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE: 3234-3243 [DOI: 10.1109/CVPR.2016.352]
- Saleh F S, Aliakbarian M S, Salzmann M, Petersson L and Alvarez J M. 2017. Bringing background into the foreground: making all classes equal in weakly-supervised video semantic segmentation//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy; IEEE: 2125-2135 [DOI: 10.1109/ICCV.2017.232]
- Saleh F S, Aliakbarian M S, Salzmann M, Petersson L and Alvarez J M. 2018. Effective use of synthetic data for urban scene semantic segmentation//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany; Springer: 86-103 [DOI: 10.1007/978-3-030-01216-8_6]
- Torrallba A and Efros A A. 2011. Unbiased look at dataset bias//Proceedings of 2011 IEEE International Conference on Computer Vision. Providence, USA; IEEE: 1521-1528 [DOI: 10.1109/CVPR.2011.5995347]
- Wang F Y. 2004. Artificial societies, computational experiments, and parallel systems: a discussion on computational theory of complex social-economic systems. *Complex Systems and Complexity Science*, 1(4): 25-35 (王飞跃. 2004. 人工社会、计算实验、平行系统——关于复杂社会经济系统计算研究的讨论. *复杂系统与复杂性科学*, 1(4): 25-35) [DOI: 10.3969/j.issn.1672-3813.2004.04.002]
- Wang F Y, Li C G, Guo Y Y, Wang J, Wang X, Qiu T Y, Meng X B and Shi X B. 2017. Parallel gout: an ACP-based system framework for gout diagnosis and treatment. *Pattern Recognition and Artificial Intelligence*, 30(12): 1057-1068 (王飞跃, 李长贵, 国元元, 王静, 王晓, 邱天雨, 孟祥冰, 施小博. 2017. 平行高特: 基于ACP的平行痛风诊疗系统框架. *模式识别与人工智能*, 30(12): 1057-1068) [DOI: 10.16451/j.cnki.issn1003-6059.201712001]
- Wang F Y, Tang Y, Liu X W and Yuan Y. 2019. Social education: opportunities and challenges in cyber-physical-social space. *IEEE Transactions on Computational Social Systems*, 6(2): 191-196 [DOI: 10.1109/TCSS.2019.2905941]
- Wang K F, Gou C and Wang F Y. 2016. Parallel vision: an ACP-based approach to intelligent vision computing. *Acta Automatica Sinica*, 42(10): 1490-1500 (王坤峰, 苟超, 王飞跃. 2016. 平行视觉: 基于ACP的智能视觉计算方法. *自动化学报*, 42(10): 1490-1500) [DOI: 10.16383/j.aas.2016.c160604]
- Wang K F, Gou C, Zheng N N, Rehg J M and Wang F Y. 2017. Parallel vision for perception and understanding of complex scenes: methods, framework, and perspectives. *Artificial Intelligence Review*, 48(3): 299-329 [DOI: 10.1007/s10462-017-9569-z]
- Wiley V and Lucas T. 2018. Computer vision and image processing: a paper review. *International Journal of Artificial Intelligence Research*, 2(1): 28-36 [DOI: 10.29099/ijair.v2i1.42]
- Yu F and Koltun V. 2016. Multi-scale context aggregation by dilated convolutions [EB/OL]. [2020-06-20]. <https://arxiv.org/pdf/1511.07122.pdf>
- Yuille A L and Liu C X. 2020. Deep nets: what have they ever done for Vision? [EB/OL]. [2020-06-20]. <https://arxiv.org/pdf/1805.04025.pdf>
- Zhang H, Kang D Q, He H B and Wang F Y. 2020. APLNet: attention-enhanced progressive learning network. *Neurocomputing*, 371: 166-176 [DOI: 10.1016/j.neucom.2019.08.086]
- Zhang H, Tian Y L, Wang K F, Zhang W S and Wang F Y. 2019. Mask SSD: an effective single-stage approach to object instance segmentation. *IEEE Transactions on Image Processing*, 29: 2078-2093 [DOI: 10.1109/TIP.2019.2947806]
- Zhang H, Wang K F, Tian Y L, Gou C and Wang F Y. 2018. MFR-CNN: incorporating multi-scale features and global information for traffic object detection. *IEEE Transactions on Vehicular Technology*, 67(9): 8019-8030 [DOI: 10.1109/TVT.2018.2843394]
- Zhang H, Wang K F and Wang F Y. 2017. Advances and perspectives on applications of deep learning in visual object detection. *Acta Automatica Sinica*, 43(8): 1289-1305 (张慧, 王坤峰, 王飞跃. 2017. 深度学习在目标视觉检测中的应用进展与展望. *自动化学报*, 43(8): 1289-1305) [DOI: 10.16383/j.aas.2017.c160822]
- Zhang Y, David P and Gong B. 2017. Curriculum domain adaptation for semantic segmentation of urban scenes//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy; IEEE: 2020-2030 [DOI: 10.1109/ICCV.2017.223]

作者简介



张慧, 1993年生, 女, 博士研究生, 主要研究方向为计算机视觉、模式识别、智能交通系统。

E-mail: zhanghui2015@ia.ac.cn



王飞跃, 通信作者, 男, 研究员, 主要研究方向为智能系统和复杂系统的建模、分析与控制。

E-mail: feiyue.wang@ia.ac.cn

李轩, 男, 博士研究生, 主要研究方向为智能交通、计算机视觉、图像合成、目标视觉检测。E-mail: lixuan0125@126.com