

# 计算机视觉前沿与深度学习

特邀编辑: 王 涛<sup>1</sup> 查红彬<sup>2</sup>

<sup>1</sup>爱奇艺公司

<sup>2</sup>北京大学

关键词: 计算机视觉 深度学习

计算机视觉通常是指用摄像机和计算机代替人眼对目标进行识别、跟踪/测量来实现对客观三维世界的理解。计算机视觉既是科学领域中富有挑战性的理论研究,也是工程领域中的重要应用,在图像检索、安全监控、人机交互、医疗诊断和机器人等领域具有广阔的应用前景。美国和欧洲等先进国家将计算机视觉列为对经济和科学有广泛影响的重大基本问题,计算机视觉也是“谷歌大脑”、“百度大脑”等研究计划中的核心项目。

计算机视觉作为一门学科始于20世纪60年代。随着个人计算机的普及,计算机视觉在80年代取得了重要进展。最近10年,随着计算机性能的大幅提升和互联网的快速发展,新的视觉特征、大数据、稀疏低秩、深度学习等技术的不断涌现,使计算机视觉又迎来了一次突飞猛进的发展,开辟出许多新的研究领域。国内高校与科研单位在计算机

视觉研究中投入巨大,在IEEE模式分析与机器智能汇刊(*IEEE Transactions on Pattern Analysis and Machine Intelligence*, *IEEE TPAMI*)、计算机视觉国际期刊(*International Journal of Computer Vision*, *IJCV*)、IEEE图像处理汇刊(*IEEE Transactions on Image Processing*, *IEEE TIP*)、IEEE国际计算机视觉大会(*IEEE International Conference on Computer Vision*, *IEEE ICCV*)和IEEE国际计算机视觉与模式识别会议(*IEEE Conference on Computer Vision and Pattern Recognition*, *IEEE CVPR*)等顶级国际期刊和会议上发表了许多重要学术论文,产生了许多国际一流的研究成果。其中最受到关注的研究是深度学习,而深度学习领域发表的论文70%以上是关于视觉图像识别方面的。

为了更好地开展学术交流,推动国内计算机视觉学科发展,进一步提升我国计算机视觉研究

在国际领域的影响力,中国计算机学会成立了“计算机视觉专业组”。在本期专题中,计算机视觉专业组特别邀请了多位著名的视觉专家从不同角度撰文,介绍计算机视觉前沿与深度学习研究方面的最新进展。

香港中文大学助理教授王晓刚、博士孙祎、教授汤晓鸥共同撰写的《从统一子空间分析到联合深度学习:人脸识别的十年历程》文章,回顾了人脸识别近十年的发展历程。他们的团队使用深度学习开发了DeepID2+系统,在人脸识别最受关注的LFW(labeled faces in the wild)<sup>1</sup>数据集上取得了人脸确认任务的世界第一,识别率99.47%。深度学习在人脸识别上的巨大成功,并非只是利用复杂模型拟合数据集。DeepID2+系统的神经元响应有很多重要的性质,比如它是中度稀疏的,对人物身份和人脸属性有很强的选择性,对局部遮挡具有良好的鲁棒性。这些性

<sup>1</sup> 标注过的户外脸部测试数据集。

质通过大数据训练自动获得,而非通过对模型显式地加入各种约束。这些研究成果启发我们进一步探索大脑是否也具有类似的性质,不断开拓新的研究方向。

中国科学院计算技术研究所研究员山世光等人撰写的文章《深度学习在人脸分析与识别中的应用》概述了中科院计算所在深度学习、人脸感知方面的工作,内容包括面向人脸识别和表情识别的特征学习,由粗到精的多阶段深度非线性人脸形状提取,姿态鲁棒的人脸特征渐进深度学习等,总结了在没有大规模人脸数据的条件下应用深度学习模型的有关经验。他们的方法在IEEE International Conference on Automatic Face and Gesture Recognition (IEEE 表情和手势自动识别国际会议, IEEE FG 2015)组织的“Point and shoot face recognition challenge”(PaSC)评测中取得了优异的成绩。

北京大学教授林宙辰和上海科技大学教授马毅的《信号与数据处理中的低秩模型》一文,介绍了稀疏低秩模型、优化算法和背景建模、图像对齐、显著区域检测等多个代表性应用。如何鲁棒、高效地处理高维数据是一个重大挑战。幸运的是,高维数据并不是毫无结构的,它们经常分布在低维流形附近。基于低秩的模型是近年涌现出来的鲁棒、高效处理高维数据的新工具,受稀疏表示(sparse representation)和压缩感知(compressed sensing)理

论的推动,系统地发展出了二阶稀疏性的理论与应用。

中国科学院自动化研究所副研究员黄永祯和研究员王亮在《视觉局部特征的表达学习》一文中介绍了局部特征的监督和非监督表达学习,通过卷积神经网络(Convolutional Neural Networks, CNN)监督模型进行局部特征表达学习后,在百度人形分割图像测试集上达到了近87%的分割精度。该文章认为卷积神经网络中的卷积操作是一种特殊的非监督局部特征编码操作,反映了滤波器与图像中每个局部区域之间的相似性。对局部特征编码的深入研究启发我们设计新的操作来代替卷积,开发出性能更好的深度模型。深度学习也启发我们研究深度的特征袋(Bag-of-Features, BoF)模型,使之在局部特征表达学习中获得更大的感知,研究出更好的局部特征编码算法。

百度公司资深研发工程师都大龙、余凯等作者在文章《基于深度学习的图像识别进展:百度的若干实践》中介绍了百度公司在图像分类、物体检测、光学字符识别(Optical Character Recognition, OCR)和自然语言描述图片等领域取得的一流研究成果和多层次并行分布式深度学习平台PADDLE的架构,深度学习技术被成功应用到百度识图、百度翻译和手机百度等众多产品中。人工智能的本质特征是学习的能力。作者通过大量的研发实践,总结了图像扰动、结构化损失函

数、参数稀疏化等颇具价值的知识经验,指出了深度学习在增强学习,大规模弱标注数据训练,低层、高层视觉结合,硬件加速等问题的研究方向。

计算机视觉的终极目标是“建立一个智能的系统,能够让计算机和机器人像人一样看懂世界,也可能超越人类,比人类更能看懂这个世界”。深度学习,作为人工智能的一种形式,通过组合低层特征形成具有抽象表示的深层神经网络,模拟人脑的思维进行感知、识别和记忆,突破了低层特征到高层语义理解的障碍,极大地提升了机器在视觉特征的提取、语义分析和理解方面的智能处理水平。随着计算机技术和人工智能的发展,期待计算机视觉研究取得更大的突破,在社会生活中得到更加广泛的应用! ■



**王 涛**

CCF高级会员、计算机视觉专业组副主任、本刊编委。爱奇艺公司首席科学家。主要研究方向为计算机视觉、多媒体分析和数据挖掘等。  
wtao@qiyi.com



**查红彬**

CCF高级会员、计算机视觉专业组副主任。北京大学教授。主要研究方向为计算机视觉、智能人机交互。  
zha@cis.pku.edu.cn