



计算机视觉中的深度学习专题简介

庞彦伟^{1*}, 白翔², 章国锋³

1. 天津大学, 天津 300072

2. 华中科技大学, 武汉 430074

3. 浙江大学, 杭州 310058

* 通信作者. E-mail: pyw@tju.edu.cn

自从 2012 年深度学习算法在图像分类任务中取得举世瞩目的成绩后, 深度学习逐渐取代了传统的统计学习成为计算机视觉的主流框架和方法, 呈现出巨大的优势, 在世界范围引起学术界和工业界的广泛关注. 基于深度学习的计算机视觉 (特别是视觉智能感知与决策) 已经是我国《新一代人工智能发展规划》的重要研究内容, 在智能驾驶、无人系统、机器人、人机交互、视频监控、生物特征识别等领域发挥着越来越重要的作用. 然而, 为了满足智能视觉感知与决策等计算机视觉系统在处理复杂不可控等场景下的高性能、高效率、高智能等具有现实意义的技术要求, 还需要在深度学习理论和算法上持续进行更多创新性工作. 在此背景下, *SCIENCE CHINA Information Sciences* 计划从 2019 年 62 卷第 12 期起持续组织出版 “Special Focus on Deep Learning for Computer Vision” (计算机视觉中的深度学习专题), 旨在展示和报道有关基于深度学习的计算机视觉的新思想、新框架、新方法, 不断提高计算机视觉系统的精度、效率和智能化.

经过高效率、高质量的评审, 本专题首批录用了 4 篇论文, 涵盖了人体分割解析、行人再识别、文本检测和三维物体检测等具有挑战性的任务. 具体地, “Feature context learning for human parsing” (Huang 等) 提出了通用性很强的特征上下文模块 (FCM), 当嵌入到多种现有算法中时均可显著提升人体解析的性能. “Uncertainty-optimized deep learning model for small-scale person re-identification” (Zhao 等) 通过采用多级 dropout 和改进的蒙特卡洛 (Monte Carlo) 策略明显缓解了过拟合问题, 并因此显著提高了行人再识别的识别率. “Irregular scene text detection via attention guided border labeling” (Chen 等) 提出了一种注意力导引的模型, 在文本区域加权边界处的标签化中发挥了重要作用, 得到了较好的不规则场景文本检测效果. “RPNET: attention region proposal network for 3D object detection” (Ye 等) 提出了一种基于注意力机制的候选区域选择网络 (RPN), 具备利用形状统计信息的作用, 提高了从点云数据中检测三维物体的能力. 需要指出的是, 虽然严格地讲基于点云数据的物体检测并未直接利用视觉数据, 但该任务和计算机视觉具有密切的联系, 具有相互促进的作用, 且在智能驾驶等任务中二者通常会进行协同感知. 因此, 将基于点云数据的三维物体检测列入本专题的范畴.

在端到端深度卷积神经网络框架下, 上述论文分别从注意力机制 (Ye 等, Chen 等)、上下文利用 (Huang 等) 和采样策略 (Zhao 等) 等不同角度, 分别解决或缓解了分割不一致性、小样本重识别、不规则场景文本检测、形状先验信息缺失等问题, 显著提升了不同计算机视觉任务的性能. 本专题今后将持续报道这方面的新进展.

引用格式: 庞彦伟, 白翔, 章国锋. 计算机视觉中的深度学习专题简介. 中国科学: 信息科学, 2019, 49: 1659, doi: 10.1360/SSI-2019-0280