



计算机视觉中的深度学习专题 (2020) 简介

白翔^{1*}, 庞彦伟², 章国锋³

1. 华中科技大学, 武汉 430074

2. 天津大学, 天津 300072

3. 浙江大学, 杭州 310058

* 通信作者. E-mail: xbai@hust.edu.cn

深度学习技术在许多研究领域都取得了巨大的成功, 尤其在计算机视觉领域取得了显著的进展. *SCIENCE CHINA Information Sciences* 从 2019 年起持续组织出版 “Special Focus on Deep Learning for Computer Vision” (计算机视觉中的深度学习专题), 旨在报道利用深度学习解决重要视觉任务的新思路.

经过高效率、高质量的评审, 本专题第二批共录用了 6 篇论文和 4 篇短文, 涵盖了语义分割、目标检测、图像合成、文本识别、年龄估计等有挑战性的重要视觉任务, 发表在 *SCIENCE CHINA Information Sciences* 2020 年 63 卷第 2 期上.

“Progressive rectification network for irregular text recognition” (Gao 等) 提出了一种渐进式矫正网络, 可将不规则形状的文本迭代转换为水平排列的规则视图, 从而显著提升了文本识别的精度.

“Ordinal distribution regression for gait-based age estimation” (Zhu 等) 通过尝试将年龄的顺序关系作为一个重要线索, 设计了一种新的序数分布损失函数, 能有效用于基于步态的年龄估计. 该方法具有一定的通用性, 同样适用于基于人脸的年龄估计.

“FACLSTM: ConvLSTM with focused attention for scene text recognition” (Wang 等) 提出了一种命名为 ConvLSTM 文本识别模型, 能够从二维空间直接识别场景文本. 在此基础上, 该方法引入了注意力机制和字符中心位置来进一步对文本识别精度进行了提升.

“CGNet: cross-guidance network for semantic segmentation” (Zhang 和 Pang) 提出了一个新的网络框架 CGNet, 能够同时进行语义分割、边缘检测和显著性检测三种视觉任务. 在边缘检测与显著性检测子网络的引导下, CGNet 的语义分割性能能获得显著提升.

“SynthText3D: synthesizing scene text images from 3D virtual worlds” (Liao 等) 提出了一种通过三维虚拟世界产生场景文本图像的特别思路. 通过三维虚拟环境产生的文本图像数据具有理想的视觉效

引用格式: 白翔, 庞彦伟, 章国锋. 计算机视觉中的深度学习专题 (2020) 简介. 中国科学: 信息科学, 2020, 50: 303-304, doi: 10.1360/SSI-2020-0021

果, 涵盖了透视变换、光照与遮挡等多种变化, 可用于训练更鲁棒的文本检测模型.

“Preserving details in semantics-aware context for scene parsing” (Ma 等) 通过嵌入丰富低层语义信息来改进语义分割网络的空间解码过程. 该方法能够更精确地获取语义分割所需要的细节特征.

此外, 4 篇短文也带来了在不同视觉任务种的积极进展. Gao 等提出了一种判别式自动编码器, 能够学习得到更加鲁棒的特征表示. Wang 等探讨了跨模态商品图像搜索中的注意力机制. Cui 等研究了基于单步检测网络的小目标定位问题. Wang 等介绍了如何有效利用三维混合形状信息来提升人脸表情识别精度.

我们对所有提交稿件的作者表示衷心的感谢, 同时我们也对所有匿名审稿人致以最深切的谢意.