Special focus on deep learning for computer vision

Articles you may be interested in

Special focus on deep learning for computer vision
SCIENCE CHINA Information Sciences **62**, 220100 (2019);

Special Focus on Deep Learning for Computer Vision
SCIENTIA SINICA Informationis **50**, 303 (2020);

Special focus on deep learning in remote sensing image processing
SCIENCE CHINA Information Sciences **63**, 140300 (2020);

Special focus on learning and real-time optimization of automotive powertrain systems
SCIENCE CHINA Information Sciences **61**, 070200 (2018);

A computer vision system for defect discrimination and grading in tomatoes using machine learning and image processing
Aritificial Intelligence in Agriculture **2**, 28 (2019);

• EDITORIAL •

Special focus on deep learning for computer vision*

Deep learning has achieved great successes in many research areas. In particular, remarkable progresses have been made in the field of computer vision. This special focus, which will also appear in the next few issues, aims at communicating the new ideas on applying deep learning to solve the critical vision tasks.

In this special focus, six research papers and four letters are accepted after high quality review. These papers cover a variety of important vision tasks: semantic segmentation, object detection, image synthesis, image retrieval, OCR, age estimation, etc. More specifically, there are two articles on semantic segmentation (Zhang and Pang, Ma et al.), two articles on scene text recognition (Gao et al., Wang et al.), one article on text image synthesis (Liao et al.), one article on gait-based age estimation (Zhu et al.), one letter on deep feature learning (Gao et al.), one letter on product image retrieval (Wang et al.), one letter on object detection (Cui et al.), and one letter on facial expression recognition (Wang et al.).

All six research papers achieve the significant progresses in their corresponding vision tasks.

(1) In "Progressive rectification network for irregular text recognition", Gao et al. propose a progressive rectification network (PRN) for iteratively transforming irregular scene text into a front-horizontal view, resulting in the significant performance improvement of scene text recognition.

(2) In "Ordinal distribution regression for gait-based age estimation", by considering the ordinal relationship of ages as an important cue, Zhu et al. design a neural network for gait-based age estimation by a new loss function termed as ordinal distribution loss. This general method is not only limited to gait-based age estimation, but also can be used for face-based age estimation.

(3) In "FACLSTM: ConvLSTM with focused attention for scene text recognition", Wang et al. tackle scene text recognition problem from a spatiotemporal prediction perspective. They propose the ConvLSTM model for reading scene text from 2D space, by which attention mechanism and character center masks are further adopted for enhancing the recognition performance.

(4) In "CGNet: cross-guidance network for semantic segmentation", Zhang and Pang introduce a unified framework named cross guidance network (CGNet) for simultaneously extracting segmentation, edge, and salient features. With the guidance of edge and saliency detection network, more discriminative features are learned with CGNet for obviously enhancing the performance of semantic segmentation.

(5) In "SynthText3D: synthesizing scene text images from 3D virtual worlds", Liao et al. propose an unconventional approach for generating scene text images from the 3D virtual worlds. The synthetic images produced from 3D virtual worlds yield realistic visual effects, including complex perspective transforms, various illuminations, and occlusions, which can be used for training a stronger scene text detector.

(6) In "Preserving details in semantics-aware context for scene parsing", Ma et al. attempt to improve the spatial decoding process through embedding possibly lost low level information in a simple yet effective manner. This method well captures the fine image details, which are difficult to be handled by the FCN-based pipelines for semantic segmentation.

Additionally, the four letters show their promising progresses in different vision tasks. Gao et al. present a discriminative stacked autoencoder (DSA) for learning a more robust feature representation.

Wang et al. discuss the attention mechanism for cross-domain product image retrieval. Cui et al. study the problem of localizing small objects with a single shot detector. Wang et al. propose to effectively leverage 3D blendshape information for facial expression recognition.

We would like to express our sincere appreciation to all the authors for submitting their manuscripts. Moreover, we express our deepest gratitude to all the anonymous reviewers for delivering high-quality and timely reviews.

Guest Editors:
Xiang BAI
Huazhong University of Science and Technology, China
Yanwei PANG
Tianjin University, China
Guofeng ZHANG
Zhejiang University, China