

## Special focus on deep learning for computer vision\*

Deep learning-based computer vision plays an increasingly important role in developing the high-performance intelligent perception systems such as self-driving vehicles, unmanned surface vehicles, robots, intelligent human-machine interactions, and visual surveillance. However, more effort has to be done to improve the performance so that deep learning-based computer vision can satisfy the rigorous demands of the intelligent perception systems. This special focus, which will also appear in next few issues, aims at collecting new ideas about deep learning-based computer vision to improve the performance in terms of accuracy and/or efficiency.

After high-quality review, the first group of four articles have been accepted. These articles cover the four domains of human parsing, person re-identification, text detection, and 3D object detection, which are deemed to be the most challenging ones when applying computer vision tasks. These articles attempt to solve the problems of parsing inconsistency (Huang et al.), re-identifying persons with small-sample dataset (Zhao et al.), detecting irregular scene texts (Chen et al.), and missing shape priors (Ye et al.). Within the framework of end-to-end deep learning (specifically, deep convolutional neural networks), the methods in these articles handle the problems from the point of view of attention mechanism (Ye et al., Chen et al.), context utilization (Huang et al.), or sampling strategy (Zhao et al.).

All four articles significantly improve the performance of their respective corresponding computer vision tasks. The Feature Context Module (FCM) proposed by Huang et al. is very general and can be integrated into the existing algorithms with significant performance gains in human parsing. By employing a multi-level dropout method and an improved Monte Carlo strategy, the method proposed by Zhao et al. is able to alleviate the over-fitting problem and hence can significantly improve the recognition rates of person-re-identification. Owing to a positive role of the proposed attention guided model in labelling the weighted borders of the text regions, the method proposed by Chen et al. performs well in detecting irregular scene texts. Ye et al. proposed an attention RPN to utilize the prior of statistic shape for successfully detecting 3D objects from the point cloud.

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 review comments.

Guest Editors:

Yanwei PANG

*Tianjin University, China*

Xiang BAI

*Huazhong University of Science and Technology, China*

Guofeng ZHANG

*Zhejiang University, China*

\***Citation** Pang Y W, Bai X, Zhang G F. Special focus on deep learning for computer vision. *Sci China Inf Sci*, 2019, 62(12): 220100, <https://doi.org/10.1007/s11432-019-2701-8>