International Journal of Computer Vision (2020) 128:1160–1161
https://doi.org/10.1007/s11263-020-01324-z

GUEST EDITORIAL

Special Issue on Deep Learning for Robotic Vision

Anelia Angelova[1] · Gustavo Carneiro[2] · Niko Sünderhauf[3] · Jürgen Leitner[4]

Published online: 28 March 2020

In this special issue we consider the topic of Robotic Vision in the context of deep learning techniques. Vision for robotics, while closely related to classical computer vision, and often using it for inspiration, focuses on specific vision tasks which are required for robot navigation and interaction. It has distinct characteristics specific to the robotics domain, for example, working with multiple or diverse sensors, or incorporating sensing from different viewpoints. Robot vision has the opportunity to use self-supervision, e.g., the robot's own actions or movements can be used as vision-based supervision. It also has its specific challenges related to robotics, for example, being able to perceive visual input in challenging non-structured environments, or in adverse weather conditions, being able transfer methods from simulated to real environments (i.e., sim2real transfer), and others. Robotic Vision is an indispensable part of robotics as many robotics-specific applications, such as learning to navigate, localize, or act in the environment, heavily rely on vision and thus are dependent on advances in automatic visual understanding.

This special issue is based on the Deep Learning for Robotic Vision Workshop we organised at CVPR 2017 but was also open to other submissions.

The special issue received 29 submissions total, of which 9 were accepted. The review process followed the high standards of IJCV.

✉ Anelia Angelova
anelia@google.com

Gustavo Carneiro
gustavo.carneiro@adelaide.edu.au

Niko Sünderhauf
niko.suenderhauf@qut.edu.au

Jürgen Leitner
juxi@lyro.io

[1] Robotics at Google, Mountain View, CA, USA

[2] University of Adelaide, Adelaide, Australia

[3] Queensland University of Technology, Brisbane, Australia

[4] LYRO Robotics Pty Ltd, Brisbane, Australia

Accepted publications

Cognitive Mapping and Planning for Visual Navigation by Saurabh Gupta et al. learns a neural map representation from first-person view, which works jointly with a differentiable planning algorithm for a full end-to-end vision-based navigation.

SeDAR: Reading Floorplans Like a Human. Using Deep Learning to Enable Human-Inspired Localisation by Oscar Alejandro Mendez Maldonado et al. presents an algorithm for Monte Carlo-based localization which uses semantic information by constructing high-level semantic maps, called floor plans. The paper shows successful localization from semantic maps only and from semantics and range information.

Image-Based Geo-Localization Using Satellite Imagery by Sixing Hu and Gim Hee Lee uses ground-to-aerial image matching for geo-referenced satellite map localization, utilizing data observations from widely differing viewpoints. To utilize a video stream of inputs, a particle filter-based localization enforces consistency across consecutive frames.

Self-Supervised Model Adaptation for Multimodal Semantic Segmentation by Abhinav Valada et al. proposes multi-modal fusion for semantic scene segmentation. With a novel module for self-calibration of features, it dynamically adapts the fusion of these multi-modal features, and provides more robust performance.

Deep Multicameral Decoding for Localizing Unoccluded Object Instances from a Single RGB Image, by Matthieu Grard et al. proposes a multicameral method for occlusion-aware instance-sensitive segmentation. To test the proposed approach, the paper introduces Mikado: a new synthetic data set of dense homogeneous object layouts that has more instances and inter-instance occlusions per image than previously proposed public data sets.

Curriculum Model Adaptation with Synthetic and Real Data for Semantic Foggy Scene Understanding by Dengxin Dai et al. learns semantic scene segmentation in images with fog, by a sim2real transfer using simulated light fog images and real unsupervised dense fog images. The algorithm gradually adapts the semantic segmentation model,

initially trained on clear-weather images, to images with dense fog, and shows improved semantic segmentation in images of dense real fog.

Learning 3D Shape Completion under Weak Supervision by David Stutz and Andreas Geiger proposes a learning-based shape completion algorithm from 3D point cloud scans which avoids slow 3D shape optimization. The approach is able to work with weak supervision, and is competitive to state-of-the-art fully supervised baselines.

Semi-Supervised Semantic Mapping through Label Propagation with Semantic Texture Meshes by Radu Alexandru Rosu et al. generates a 3D mesh representation of the scene with learned semantic annotation which enables spatial and temporal consistency of semantic labels from multiple views.

Model-based Robot Imitation with Future Image Similarity by Alan Wu et al. learns robot action policies from future image similarity in videos and without robot trials. The experiments demonstrate a ground mobility robot reaching target objects by using these policies, in both real and simulated environments.

Together these papers offer a glimpse into a number of topics of Robotic Vision and open the door for more development in the field.

We sincerely thank all the reviewers and all authors for their hard and dedicated work.