



## Editorial: Special Issue on Machine Vision with Deep Learning

Ling Shao<sup>1</sup> · Hubert P. H. Shum<sup>2</sup> · Timothy Hospedales<sup>3</sup>

Published online: 9 March 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

This special edition of the International Journal of Computer Vision is devoted to a selection of the best papers submitted to the 2018 British Machine Vision Conference (BMVC) held at Northumbria University, Newcastle upon Tyne, UK. This year, BMVC attracted a total of 862 full paper submissions, which is the highest number in the history of BMVC. Papers were assigned to Area Chairs (AC) and reviewers based on both AC and reviewer biddings on papers, CMT subject areas, and the Toronto Paper Matching System (TPMS) publication history. There were 426 reviewers and 52 area chairs involved in this process, generously donating their time. Every paper was handled by two ACs and each received just over three reviews on average. Of the 862 submissions, just 255 were accepted for presentation in BMVC 2018, which is a 29.5% acceptance rate. Of the accepted papers, 37 were accepted as oral. Based on the reviewers' reports of the papers and feedback obtained at the conference, we invited the authors of 14 papers to submit extended versions to this Special Issue. These were subjected to the rigorous review process of IJCV, and resulted in 8 accepted papers.

The papers presented in this issue offer a snapshot of some of the best work in contemporary deep learning-based machine vision. Overall, they cover topics of: (1) fundamentals of CNN architectures, (2) image processing with deep learning, (3) deep 3D shape image analysis, (4) deep networks for analysing humans. Below is a summary of the papers accepted.

The first two papers address fundamental aspects of convolutional neural network architectures. “*Learning on the Edge: Investigating Boundary Filters in CNNs*” by

Innamorati et al., tackles the issue image boundary handling in convolutional neural networks. While boundaries have conventionally been dealt with by ad-hoc heuristics, they show how to learn the boundary handling strategy itself using a set of explicit boundary filters. This enables the extrapolation strategy at boundaries to be trained in a task-specific way. On the other hand, in “*A Simple and Light-weight Attention Module for Convolutional Neural Networks*”, Park et al. explored the use of attention in convolutional neural networks. They proposed a simple self-contained module called Bottleneck Attention Module (BAM). This module can be trained jointly with the backbone network, producing a channel and position-wise attention map that leads to improved performance across a variety of image analysis tasks and backbone networks.

The next group of papers apply neural networks to image processing tasks. In “*Simultaneous Deep Stereo Matching and Dehazing with Feature Attention*”, Song et al. address producing dense correspondences of stereo images in hazy conditions. They proposed a convolutional neural network that exploits complementary cues from jointly performed stereo matching and dehazing, leading to improved performance in both tasks. In “*Pixelated Semantic Colorization*”, Zhao et al. address inferring colorized versions of grayscale photographs. By exploiting learned semantic segmentation of objects, colorisation is performed in a spatially accurate and semantically plausible manner.

Deep learning is also increasingly been applied to 3D computer vision tasks. In “*Learning Single-image 3D Reconstruction by Generative Modelling of Shape, Pose and Shading*”, Henderson and Ferrari proposed a unified framework for both class-specific 3D reconstruction from a single image and the generation of new 3D shape samples. The system can be trained purely from 2D images, without pose annotations and with only a single view per instance. In “*Modeling Human Motion with Quaternion-based Neural Networks*”, Pavllo et al. addressed the limitations of both joint rotation and joint position representations of 3D human pose sequences. They propose the use of quaternions to represent rotations of the skeleton, together with a loss function

✉ Ling Shao  
ling.shao@ieee.org

Hubert P. H. Shum  
hubert.shum@northumbria.ac.uk

Timothy Hospedales  
t.hospedales@ed.ac.uk

<sup>1</sup> Inception Institute of Artificial Intelligence, Abu Dhabi, UAE

<sup>2</sup> Northumbria University, Newcastle upon Tyne, UK

<sup>3</sup> University of Edinburgh, Edinburgh, UK

that preforms forward kinematics to penalize absolute position errors instead of angle errors.

The final group of papers address human-centric analysis, which is an increasingly topical application area within machine vision. In “*Learning Multi-human Optical Flow*”, Ranjan et al. considered training optical flow models specifically for human motion. The paper introduces a human optical flow dataset, and shows that training optical flow networks specifically for human motion leads to improved accuracy compared to generic flow models. In “*Text2Sign: Towards Sign Language Production using Neural Machine Translation and Generative Adversarial Networks*”, Stoll et al. presented a system to automatically produce sign language videos from spoken language sentences. They adapted Neural Machine Translation with a Motion Graph to translate

spoken language sentences into sign pose sequences. The pose information generated was then used to condition a Generative Adversarial Network that produced photo-realistic sign language video sequence.

We would like to take this opportunity to thank the authors for submitting their high-quality research to BMVC and to this special issue. We also express our gratitude to the reviewers, who have given extensive feedback for improving the quality of the submitted manuscripts, as well as the Springer editorial staff, who have provided significant support on the production of this special issue.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.