Fatih Porikli, Shiguang Shan, Cees Snoek,
Rahul Sukthankar, and Xiaogang Wang

Deep Learning for Visual Understanding

In the past decade, there has been a transformative and permanent revolution in computer vision cultivated by the reinvigorated adoption of deep learning for visual understanding tasks. Driven by the increasing availability of large annotated data sets, efficient training techniques, and faster computational platforms, deep-learning-based solutions have been progressively employed in a broader spectrum of applications from image classification to activity recognition.

Deep learning, in general, refers to a range of artificial neural networks that consist of multiple layers, mimicking the structure and cognitive process of the human brain. Instead of relying on handcrafted features, they allow the acquisition of knowledge directly from data. They regress intricate objective functions in a nested hierarchy, where more sophisticated representations with larger receptive fields computed in terms of less abstract ones with localized supports. Deep learning also makes it possible to incorporate explicit domain knowledge and replace a large variety of conventional algorithmic blocks with trainable differentiable modules. These all give deep learning an exceptional power and flexibility in modeling the relationship between the input data and target output.

Efforts are now shifting toward the remaining challenges. For instance, the majority of current methods have been designed to solve supervised learning problems where data comes with its labeled attributes and how to reliably apply deep learning to unsupervised settings in a similar degree of success is an active area of research. Similarly, recent efforts aim at working with small data, focusing on how to take advantage of large quantities of unlabeled examples as well as with a few labeled samples.

Another area where deep agents may play a significant role is to integrate positive and negative rewards into deep learning to choose the actions that yield the best cumulative reward by interacting with the environment. Also, the fusion of multimodal and structured data into existing deep-learning models would open up more extended application domains.

This special issue of *IEEE Signal Processing Magazine* (*SPM*) is therefore devoted to providing survey articles on the latest advances in deep learning for visual understanding. Its objective is to encourage a diverse audience of researchers and enthusiasts toward an effective participation in the solution of analogous problems in other signal processing fields by inseminating similar ideas.

The range of articles in this two-part special issue indicates the breadth of the computer vision discipline. (Part two will be published in January 2018.) Many fundamental areas are surveyed from the computer vision perspective, including

- reinforcement learning
- learning with limited and no supervision (unsupervised learning)
- weakly supervised learning
- zero- and few-shot learning
- domain adaptation
- multimodal learning
- metric learning
- generative adversarial networks
- recurrent networks
- regression with Bayesian networks
- model compression and robustness.

In addition, in-depth overviews of several deep-learning-based computer vision applications are provided, including

- inverse problems such as superresolution and image enhancement
- picture quality prediction
- saliency detection
- image and video segmentation with conditional random fields
- image-to-text generation
- visual question answering
- face image analytics.

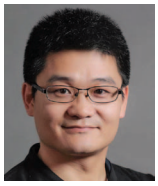
We would like to wholeheartedly thank all of the contributing authors and reviewers of this special issue. We also sincerely appreciate *SPM*'s editor-in-chief, Prof. Min Wu, Managing Editor Jessica Welsh, and the entire magazine's editorial staff for their extremely valuable support.

Meet the guest editors

Fatih Porikli (fatih.porikli@anu.edu.au) received his B.Sc. degree in electrical engineering from Bilkent University, Turkey, in 1992 and his Ph.D. degree in electrical and computer engineering from

Digital Object Identifier 10.1109/MSP.2017.2744538
Date of publication: 13 November 2017

New York University in 2002. He is an IEEE Fellow and a professor at Australian National University. He is also the chief scientist at Huawei, Santa Clara, California. Previously, he served as the Computer Vision Research group leader at National ICT Australia and distinguished scientist at Mitsubishi Electric Research Laboratories. His research interests include computer vision and machine learning with commercial applications in autonomous vehicles, video surveillance, visual inspection, robotics, and medical systems. He received the R&D100 Scientist of the Year Award in 2006, won five Best Paper Awards at IEEE conferences, and invented 71 patents.



Shiguang Shan (sgshan@ict.ac.cn) received his B.S.E. and M.S.E. degrees in computer science from Harbin Institute of Technology, China, in 1997 and 1999, respectively. He received his Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences (CAS), Beijing, in 2004, where he has been a full professor since 2010 and is now the deputy director of the CAS Key Lab of Intelligent Information Processing. His research interests include computer vision, pattern recognition, and machine learning. He has published more than 200 papers in these areas. He served as area chair for many international conferences and is an associate editor of several journals, including *IEEE Transactions on Image Processing*, *Computer Vision and Image Understanding*, *Neurocomputing*, and *Pattern Recognition Letters*.

Cees Snoek (cgmsnoek@uva.nl) received the M.Sc. degree in business information systems in 2000 and the Ph.D. degree in computer science in 2005, both from the University of Amsterdam, The Netherlands. He is currently a director of the QUVA Lab, the joint research lab of Qualcomm and the University of Amsterdam, on deep learning and computer vision. He is also a principal engineer/manager at Qualcomm and an associate professor at the University of Amsterdam. His research interests focus on video and image recognition. He has published more than 200 refereed book chapters, journal, and conference papers. He received a Veni Talent Award, a Fulbright Junior Scholarship, a Vidi Talent Award, and The Netherlands Prize for Computer Science Research, all for research excellence.

Rahul Sukthankar (rahulsukthankar@gmail.com) received his B.S.E. degree in computer science from Princeton University, New Jersey, in 1991 and his Ph.D. degree in robotics from Carnegie Mellon, Pittsburgh, Pennsylvania, in 1997. He leads research efforts in computer vision, machine learning, and robotics at Google. He is also an adjunct research professor with the Robotics Institute at Carnegie Mellon and courtesy faculty at the University of Central Florida. Previously, he was a senior principal researcher at Intel Labs, a senior researcher at HP/Compaq Labs, and a research scientist at Just Research. He has organized several workshops and conferences and currently serves as the editor-in-chief of *Machine Vision and Applications*.

Xiaogang Wang (xgwang@ee.cuhk.edu.hk) received his bachelor's degree in electronic engineering and information science from the Special Class of Gifted Young at the University of Science and Technology of China in 2001, his M.Phil. degree in information engineering from the Chinese University of Hong Kong in 2004, and his Ph.D. degree in computer science from the Massachusetts Institute of Technology in 2009. He has been an associate professor in the Department of Electronic Engineering at the Chinese University of Hong Kong since August 2009. He received the PAMI Young Research Award Honorable Mention in 2016. He is the associate editor of *Image and Visual Computing Journal*, *Computer Vision and Image Understanding*, and *IEEE Transactions on Circuit Systems and Video Technology*.

SP

Community Voices

(continued from page 16)

Massachusetts Institute of Technology in honor of Al Oppenheim's 80th birthday, with the goal of bringing together experts in industry and academia to think progressively and speculate about the future of the field moving forward.

Over a dozen speakers provided a range of thought-provoking insights about the continued impact of the field in the decades ahead, in terms of applications, mathematics for new algorithms, and new implementation technologies. We would love to share this with those in the signal processing community who were unable to attend. A collection of video recordings and thoughts from the symposium will be available at https://futureofsp.eecs.mit.edu/. It was an exciting event, and we hope that the videos continue to stimulate further creative discussion within the community!

Thomas A. Baran *(tom.baran@gmail.com) is a cofounder and chief executive officer of Lumii and a research affiliate at the Massachusetts Institute of Technology.*

Reference

[1] A. Zoubir, "Interdisciplinary research: A catalyst for innovation" [From the Editor], *IEEE Signal Process. Mag.*, vol. 29, no. 3, pp. 2–4, 2012.

SP