Guest Editors' Introduction: Special Section on Learning Deep Architectures

Samy Bengio, Li Deng, Hugo Larochelle, Honglak Lee, and Ruslan Salakhutdinov

THE past seven years have seen a resurgence of research in the design of deep architecture models and learning algorithms, i.e., methods that rely on the extraction of a multilayer representation of the data. Often referred to as deep learning, this topic of research has been building on and contributing to many different research topics, such as neural networks, graphical models, feature learning, unsupervised learning, optimization, pattern recognition, and signal processing. Deep learning is also motivated and inspired by neuroscience and has had a tremendous impact on various applications such as computer vision, speech recognition, and natural language processing. The clearly multidisciplinary nature of deep learning led to a call for papers for a special issue dedicated to learning deep architectures, which would provide a forum for the latest advances on the subject. Associate Editor in Chief (AEIC) Max Welling took the initiative for the special issue in an earlier attempt to collaborate with Editor In Chief of the IEEE Signal Processing Magazine (Li Deng) for a joint special issue. We were five guest editors to oversee the task of selecting among the submissions the eight papers that were included in this special section. We were assisted by a great team of dedicated reviewers. Former Editor in Chief (EIC) Ramin Zabih and current EIC David Forsyth have greatly assisted in the process of realizing this special section. We all thank them for their crucial role in making this special section a success.

The guest editors were not allowed to submit their own papers to this special section. However, it was decided that our paper submissions on deep learning would be handled by AEIC Max Welling (who was not a guest editor) and go through the regular *TPAMI* review process. The last two papers in the list below resulted from this special arrangement.

The first three papers in this special section present insightful overviews of general topics related to deep learning: representation learning, transformation learning, and the primate visual cortex.

The first paper in our special section, "Representation Learning: A Review and New Perspectives," written by Yoshua Bengio, Aaron Courville, and Pascal Vincent, reviews the recent work in deep learning, framing it in terms of representation learning. The paper argues that shallow representations can often entangle and hide the different underlying factors of variability behind the observed surface data, and deep learning has the capacity to disentangle such factors using powerful distributed representations, which are combinatorial in nature. The paper also provides interesting connections among deep learning, manifold learning, and learning based on density estimation.

A short overview of recent work on learning to relate (a pair of) images is also presented in "Learning to Relate Images" by Roland Memisevic. The focus of the work is to learn local relational latent features via multiplicative interactions (using three factors that correspond to latent features and a pair of images). Specifically, the paper proposes a gated 3-way sparse coding model to learn a dictionary of transformations between two or more image patches, and it provides insightful connections to other related methods, such as mapping units, motion energy models, and squared pooling. This work further provides an interesting geometric interpretation of learned transformation dictionary, which represents a set of 2D subspace rotations in eigenspace. The paper demonstrates that interesting filters (related to transformations such as translation and rotation) can be learned from pairs of transformed images. We look forward to exciting real-world applications of relational feature learning in the near future.

The primate visual cortex, which has always been part of the inspiration behind deep learning research, is the subject of the next paper in this special section. The paper "Deep Hierarchies in the Primate Visual Cortex: What Can We Learn For Computer Vision?" by Norbert Kruger, Peter Janssen, Sinan Kalkan, Markus Lappe, Ales Leonardis, Justus Piater, Antonio J. Rodriguez-Sanchez, and Laurenz Wiskott is a great overview of known facts from the primate visual cortex that could help researchers in computer vision interested in deep learning approaches to their field. It shows how the primate visual system is organized in a deep hierarchy, in contrast to the flatter classical computer vision approach, with the hope that more interaction between biological and computer vision researchers may lead to better designs of actual computer vision systems.

- *S. Bengio is with Google Research, 1600 Amphitheatre Parkway, Mountain View, CA 94043. E-mail: bengio@google.com.*
- *L. Deng is with Microsoft Research, One Microsoft Way, Redmond, WA 98052. E-mail: deng@microsoft.com.*
- *H. Larochelle is with the Département d'Informatique, Université de Sherbrooke, 2500 boul. de l'Université, Sherbrooke, QC, J1K 2R1, Canada. E-mail: hugo.larochelle@usherbrooke.ca.*
- *H. Lee is with the Department of Electrical Engineering and Computer Science, University of Michigan, 2260 Hayward St., Ann Arbor, MI 48109. E-mail: honglak@eecs.umich.edu.*
- *R. Salakhutdinov is wth the Departments of Computer Science and Statistics, University of Toronto, 6 King's College Rd., Toronto, ON, M5S 3G4, Canada. E-mail: rsalakhu@cs.toronto.edu.*

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org.

The next three papers in this section then move on to presenting novel deep architectures, designed in the context of image classification and object recognition.

The paper "Invariant Scattering Convolution Networks" by Joan Bruna and Stéphane Mallat describes a new deep convolutional architecture, which can extract a representation that is shown to be locally invariant to translations and stable to deformations. It can also preserve high-frequency information within the input image, which is important in a classification setting. Unlike for deep convolutional models, however, the filters or convolutional kernel are not learned. This paper is a really interesting demonstration of how signal processing ideas can help in designing and theoretically understanding deep architectures.

One of the fundamental challenges of deep learning research is how to learn the structure of the underlying model, including the number of layers and number of hidden variables per layer. The paper "Deep Learning with Hierarchical Convolutional Factor Analysis" by Bo Chen, Gungor Polatkan, Guillermo Sapiro, David Blei, David Dunson, and Lawrence Carin takes a first step toward achieving this goal. It introduces a hierarchical convolutional factor-analysis model, where the factor scores from layer l serve as the inputs to layer l+1 with a max-pooling operation. In particular, a key characteristic of the proposed model is that the number of hidden variables, or dictionary elements, at each layer is inferred from the data using a beta-Bernoulli implementation of the Indian buffet process. The authors further present an empirical evaluation on several image-processing applications.

Next, the paper "Scaling Up Spike-and-Slab Models for Unsupervised Feature Learning," written by Ian Goodfellow, Aaron Courville, and Yoshua Bengio, describes an interesting new deep architecture for object recognition. The (shallow) spike-and-slab sparse coding architecture developed by the authors in previous work has been extended to its deep variant, called the partially directed deep Boltzmann machine. Appropriate learning and inference procedures are described, and a demonstration is given of the ability of the architecture to scale up to very large problem sizes, even when the number of labeled examples is low.

While image classification and object recognition is frequently the focus of deep learning research, this special issue also received papers that present deep architecture systems going beyond the typical object recognition setup.

The paper "Learning Hierarchical Features for Scene Labeling" by Camille Couprie, Clement Farabet, Laurent Najman, and Yann Lecun describes a novel application of convolutional neural networks for scene labeling that jointly segments images and categorizes image regions. The proposed method first introduces a multiscale convolutional neural network (CNN) to automatically learn a feature hierarchy from raw images for pixel-wise label prediction. It then introduces a hierarchical tree-based segmentation algorithm ("optimal label purity cover") to integrate the pixel-level outputs of a CNN from multiple scales. The proposed approach is very fast and achieves state-of-the-art results on several publicly available scene labeling datasets with varying complexity. Overall, this paper shows a compelling demonstration that a deep architecture is potentially capable of solving a very hard computer vision problem that traditionally has been handled only by more hand-engineered methods.

A new application for deep learning to the medical imaging domain is also presented in "Stacked Autoencoders for Unsupervised Feature Learning and Multiple Organ Detection in a Pilot Study Using 4D Patient Data" by Hoo-Chang Shin, Matthew R. Orton, David J. Collins, Simon J. Doran, and Martin O. Leach. They developed an approach to detect organs from magnetic resonance medical images. Medical imaging is a little-explored domain application for deep architecture models. It poses specific challenges, such as the presence of multiple modalities, the lack of labeled data, and the modeling of a variety of abnormalities in imaged tissues. The authors present a thorough empirical analysis of the success of their approach, based on unsupervised learning of sparse, stacked autoencoders.

Finally, we mention two other papers that were submitted by guest editors but were handled through the regular *TPAMI* review process.

The paper "Tensor Deep Stacking Networks" by Brian Hutchinson, Li Deng, and Dong Yu reports in detail the development and evaluation of a new type of deep architecture (T-DSN) and the related learning algorithm. The T-DSN makes effective use of its weight tensors to efficiently represent higher-order statistics in the hidden features, and it shifts the major computation in learning the network weights into a convex optimization problem amenable to a straightforward parallel implementation. An experimental evaluation on three tasks (image classification, phone classification, and phone recognition) demonstrates the importance of sufficient depth in the T-DSN and the symmetry in the two hidden layers' structure within each block of the T-DSN for achieving low error rates in all three tasks.

The paper "Learning with Hierarchical-Deep Models" by Ruslan Salakhutdinov, Joshua Tenenbaum, and Antonio Torralba introduces an interesting compositional learning architecture that integrates deep learning models with structured hierarchical Bayesian models. Their model is capable of learning novel concepts from a few training examples by learning low-level generic features, high-level part-like features that capture correlations among low-level features, and a hierarchy of super-classes for sharing priors over the high-level features that are typical of different kinds of concepts. The authors further evaluate their model on three different perceptual domains, including CIFAR-100 object recognition, handwritten character recognition, and human motion capture datasets.

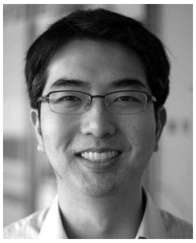
The papers in this section illustrate well the variety of topics and methods behind deep learning research. We hope that by gathering a variety of both insightful review and novel research papers, this special section will help in consolidating this multidisciplinarity and further stimulate innovative research in the field.

Samy Bengio
Li Deng
Hugo Larochelle
Honglak Lee
Ruslan Salakhutdinov
Guest Editors

Samy Bengio received the PhD degree in computer science from the University of Montreal in 1993. He has been a research scientist at Google since 2007. Before that, he was a senior researcher in statistical machine learning at the IDIAP Research Institute since 1999. His most recent research interests are in machine learning, in particular large scale online learning, learning to rank, image ranking and annotation, music information retrieval, and deep learning. He is action editor of the *Journal of Machine Learning Research* and on the editorial board of the *Machine Learning Journal*. He was an associate editor of the *Journal of Computational Statistics*, general chair of the Workshops on Machine Learning for Multimodal Interactions (MLMI '04, '05, and '06), program chair of the IEEE Workshop on Neural Networks for Signal Processing (NNSP '02), and on the program committee of several international conferences such as NIPS, ICML, and ECML. More information can be found on his website: http://bengio.abracadoudou.com.

Li Deng received the PhD degree in electrical and computer engineering from the University of Wisconsin–Madison. He joined the Department of Electrical and Computer Engineering, University of Waterloo, Ontario, Canada, in 1989 as an assistant professor and became a full professor with tenure there in 1996. In 1999, he joined Microsoft Research (MSR), Redmond, Washington, as a Senior Researcher, where he is currently a Principal Researcher. Since 2000, he has also been an affiliate full professor and graduate committee member in the Department of Electrical Engineering at the University of Washington, Seattle. Prior to MSR, he also worked or taught at the Massachusetts Institute of Technology, ATR Interpreting Telecommunications Research Laboratory (Kyoto, Japan), and HKUST. His current research activities include deep learning and machine intelligence, automatic speech and speaker recognition, spoken language understanding, speech-to-speech translation, machine translation, information retrieval, statistical signal processing, and human speech production and perception, and noise robust speech processing. He has been granted more than 60 patents in acoustics/audio, speech/language technology, and machine learning. He is a fellow of the Acoustical Society of America, a fellow of the IEEE, and a fellow of ISCA. He served on the Board of Governors of the IEEE Signal Processing Society (2008-2010). More recently, he served as Editor in Chief for the *IEEE Signal Processing Magazine* (2009-2011), which ranked first in 2010 and 2011 among all IEEE publications in terms of its impact factor and for which he received the 2011 IEEE SPS Meritorious Service Award. He currently serves as editor-in-chief for the *IEEE Transactions on Audio, Speech, and Language Processing*.

Hugo Larochelle received the PhD degree in computer science from the University of Montreal in 2009. He is now an assistant professor at the University of Sherbrooke, Canada. He specializes in the development of deep and probabilistic neural networks for high-dimensional and structured data, with a focus on AI-related problems such as natural language processing and computer vision. He is currently an associate editor for the *IEEE Pattern Analysis and Machine Intelligence* (*TPAMI*) and a member of the program committee for the NIPS 2013 conference. He also served on the senior program committee of UAI 2011. He received a Notable Paper Award at the AISTATS 2011 conference and a Google Faculty Research Award.



Honglak Lee received the PhD degree from the Computer Science Department at Stanford University in 2010, advised by Andrew Ng. He is now an assistant professor of computer science and engineering at the University of Michigan, Ann Arbor. His primary research interests lie in machine learning, which spans deep learning, unsupervised and semi-supervised learning, transfer learning, graphical models, and optimization. He also works on application problems in computer vision, audio recognition, and other perception problems. He has coorganized workshops and tutorials related to deep learning at NIPS, ICML, CVPR, and AAAI, and he has served as an area chair for ICML 2013. His work received best paper awards at ICML and CEAS. He received a Google Faculty Research Award and was selected by *IEEE Intelligent Systems* as one of the AI's Top 10 to Watch.

Ruslan Salakhutdinov received the PhD degree in computer science from the University of Toronto in 2009. After spending two postdoctoral years at the Massachusetts Institute of Technology Artificial Intelligence Lab, he joined the University of Toronto as an assistant professor in the Departments of Statistics and Computer Science. His primary interests lie in artificial intelligence, machine learning, deep learning, and large-scale optimization. He is an action editor of the *Journal of Machine Learning Research* and served on the program committees for several learning conferences, including NIPS, UAI, and ICML. He is an Alfred P. Sloan Research Fellow and Microsoft Research New Faculty Fellow, a recipient of the Early Researcher Award, Connaught New Researcher Award, and is a Scholar of the Canadian Institute for Advanced Research.