Guest Editorial
Introduction to the Special Section on Efficient Network Design for Convergence of Deep Learning and Edge Computing

CONSIDERING the distribution and heterogeneity of edge computing system, it brings great challenges to the design of efficient neural networks for edge computing. The current network design less considers the scenarios and frameworks where the model is to be deployed, and do not regard the design of efficient models for edge computing systems as specific research topics. Therefore, the efficient deep neural network design should be deeply investigated on edge computing scenarios.

The special section of "Efficient Network Design for Convergence of Deep Learning and Edge Computing" focused on the state-of-art neural network design for convergence of deep learning and edge computing. Thanks to the extensive efforts of the reviewers and the great support from the Editor-in-Chief Dr. Jianwei Huang, we were able to accept 5 contributed articles covering several important topics, from the distributed long short-term memory neural networks, to the multiple gradient descent design, to the resource-constrained neural architecture search on edge devices, cloud versus edge deployment strategies, and orthogonal super greedy learning. A brief review follows:

First, we would like to introduce the significance and potential impact of the feature article "DLSTM: Distributed long short-term memory neural networks for the Internet of Things" by Wen *et al.*, in which a new distributed sequential learning framework based on the model structure marginal decomposition technique and the multiscale learning technique was proposed to lay momentous foundation for lightweight distributed collaborative learning as well as pioneers one fresh utility pathway for deep learning among IoT. In this DLSTM, a novel cloud-edge-end collaborative computing architecture was built to decouple the structure of Long Short-Term Memory model layer by layer, breaking through the limited calculation capacity bottleneck of intelligent edge devices where the spatiotemporal large-scale data learning capability is significantly promoted via the distributed local memory transmission couple with the centralized global feature extraction. This paper contains potential practical value in respectable future IoT intelligent scenarios, such as emotion recognition in smart home, multi-area environmental detection in smart environmental protection, automatic driving in smart traffic, video monitoring in smart security, to name just a few.

Multi-task learning technique is widely utilized in machine learning modeling where commonalities and differences across multiple tasks are exploited. Zhou *et al.* in "A multiple gradient descent design for multi-task learning on edge computing: Multi-objective machine learning approach" introduced a multi-gradient descent algorithm for the multi-objective machine learning problem by which an innovative gradient-based optimization is leveraged to converge to an optimal solution of the Pareto set.

Lyu *et al.* in "Resource-constrained neural architecture search on edge devices" employed multi-objective Neural Architecture Search (NAS) on the resource-constrained edge devices. The framework was proposed for multi-objective NAS on edge device, which comprehensively considers the performance and real-world efficiency. The improved MobileNet-V2 search space also strikes the scalability and practicality, thus a series of Pareto-optimal architectures are received. Benefits from the directness and specialization during search procedure, the experiment on JETSON NANO shows the comparable result with the state-of-the-art models on ImageNet.

Ammar *et al.* in "Cloud versus edge deployment strategies of real-time face recognition inference" presented a real-world case study on deploying a face recognition application using MTCNN detector and FaceNet recognizer. Considering challenges faced to decide on the best deployment strategy, three inference architectures were proposed for the deployment, including cloud-based, edge-based, and hybrid. Furthermore, the performance of face recognition inference was evaluated on different cloud-based and edge-based GPU platforms.

Yan *et al.* in "Orthogonal super greedy learning for sparse feedforward neural networks" proposed an Orthogonal Super Greedy learning (OSGL) method for hidden neurons selection. The OSGL selects more than one hidden neurons from a given network structure in a greedy strategy until an adequate sparse network has been constructed. Theoretical analyses show it can reach the optimal learning rate.

In summary, the collected articles not only offer innovative application scenarios but also shed light on the underlying principles of efficient network design for convergence of deep learning and edge computing. We hope that this timely special section will trigger more future work in the emerging area.

Digital Object Identifier 10.1109/TNSE.2021.3133084

2327-4697 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

SHIPING WEN, *Guest Editor*
Australian Artificial Intelligence Institute
University of Technology Sydney
Sydney, NSW 2007, Australia
(e-mail: shiping.wen@uts.edu.au)

TINGWEN HUANG, *Guest Editor*
Texas A&M University-Qatar
Doha, Qatar
(e-mail: tingwen.huang@qatar.tamu.edu)

BJÖRN W. SCHULLER, *Guest Editor*
Imperial College London
London, SW7 2BX, U.K.
(e-mail: schuller@tum.de)

AHMAD TAHER AZAR, *Guest Editor*
Prince Sultan University
Riyadh 12435, Saudi Arabia
(e-mail: ahmad_t_azar@ieee.org)

Shiping Wen (Senior Member, IEEE) received the M.Eng. degree in control science and engineering from the School of Automation, Wuhan University of Technology, Wuhan, China, in 2010, and the Ph.D. degree in control science and engineering from the School of Automation, Huazhong University of Science and Technology, Wuhan, China, in 2013. He is currently a Professor with the Australian Artificial Intelligence Institute, University of Technology Sydney, Sydney, NSW, Australia. His research interests include memristor-based neural networks, deep learning, computer vision, and their applications in medical informatics, *et al.* In 2018 and 2020, he was listed as a Clarivate Analytics Highly Cited Researcher in the Cross-Field, respectively. He was the recipient of the 2017 Young Investigator Award of the Asian Pacific Neural Network Association and the 2015 Chinese Association of Artificial Intelligence Outstanding Ph.D. Dissertation Award. He is currently an Associate Editor for the *Knowledge-Based Systems*, IEEE ACCESS, and *Neural Processing Letters* and was the Leading Guest Editor of special issues IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, *Sustainable Cities and Society, Environmental Research Letters, et al*. He was also the general/publication chair or a member of the Technical Programming Committee for various international conferences.

Tingwen Huang (Fellow, IEEE) received the B.S. degree from Southwest Normal University (now Southwest University), Chongqing, China, 1990, the M.S. degree from Sichuan University, Chengdu, China, 1993, and the Ph.D. degree from Texas A&M University, College Station, TX, USA, 2002. He is currently a Professor with Texas A&M University at Qatar, Doha, Qatar. After graduated from Texas A&M University, he was a Visiting Assistant Professor there. Then, he joined Texas A & M University at Qatar, as an Assistant Professor in August 2003 and then he was promoted to a Professor in 2013. He has authored or coauthored more than 300 peer-review reputable journal papers, including more than 100 papers in IEEE Transactions. His research interests include neural networks based computational intelligence, distributed control and optimization, nonlinear dynamics, and applications in smart grids. He is currently an Associate Editor for five journals, which include the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON CYBERNETICS, *Neural Networks*, *Cognitive Computation*, and *Journal of Control and Decision*, and was the Guest Editor of some special issues in international journals, such as the *Neural Processing Letters*, *Neurocomputing*, and *Cognitive Computation*. He was elevated to an IEEE Fellow in 2018. Since 2018, he has been listed as a Clarivate Analytics Highly Cited Researcher. He is the President of Asia Pacific Neural Networks Society. He was also the general or program chair for various international conferences.

Björn W. Schuller (Fellow, IEEE) received the Diploma, Doctoral, and Habilitation degrees in electrical engineering and information technology from the Technical University of Munich (TUM), Munich, Germany. He is currently an Adjunct Teaching Professor of machine intelligence and signal processing with TUM. He is also a Full Professor of artificial intelligence and the Head of GLAM with Imperial College London, U.K., a Full Professor and the Chair of Embedded Intelligence for Health Care and Wellbeing with the University of Augsburg, Augsburg, Germany, the co-founding CEO and current CSO of audEERING – an Audio Intelligence company based near Munich and in Berlin, Germany, and a permanent Visiting Professor with HIT, China, amongst other Professorships and Affiliations. Previous stays include a Full Professor with the University of Passau, Germany, and a Researcher with Joanneum Research, Graz, Austria, and the CNRS-LIMSI, Orsay, France. He has coauthored more than 900 publications (more than 30k citations, h-index = 81). He is a Golden Core Awardee of the IEEE Computer Society, a Fellow of the ISCA, President-Emeritus of the AAAC, and a Senior Member of the ACM. He is the Field Chief Editor of the *Frontiers in Digital Health* and was the Editor-in-Chief of the IEEE TRANSACTIONS ON AFFECTIVE COMPUTING amongst manifold further commitments and service to the community. His has more than 30 awards, include having been honoured as one of 40 extraordinary scientists under the age of 40 by the WEF in 2015. He was the Coordinator/PI in 15+ European Projects, is an ERC Starting Grantee, and consultant of companies, such as Barclays, GN, Huawei, and Samsung.

Ahmad Taher Azar (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees from the Faculty of Engineering, Cairo University, Egypt, in 2006 and 2009, respectively, and he got his postdoctoral from USA. He is currently a Research Professor with Prince Sultan University, Riyadh, Saudi Arabia. He is also a Professor with the Faculty of Computers and Artificial intelligence, Benha University, Egypt. Prof. Azar is the Editor in Chief of the *International Journal of System Dynamics Applications* and *International Journal of Service Science, Management, Engineering, and Technology* published by IGI Global, USA. Also, he is the Editor in Chief of the *International Journal of Intelligent Engineering Informatics*, Inderscience Publishers, Olney, U.K. He was an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS from 2013 to 2017 and an Associate Editor for the *ISA Transactios, Elsevier* from 2018 to 2020. He is currently an Associate Editor for the IEEE SYSTEMS JOURNAL and *Human-centric Computing and Information Sciences, Springer*. In February 2020, Prof. Azar was awarded the Egyptian Distinguished Order of the first class from Egyptian President. Prof. Ahmad Azar is the Chair of the IEEE Computational Intelligence Society (CIS) Egypt Chapter, the Vice Chair of the IEEE Computational Intelligence Society Interdisciplinary Emergent Technologies Task Force, Vice-Chair Research Activities of IEEE Robotics and Automation Society Egypt Chapter, Committee Member of IEEE CIS Task Force on Fuzzy Logic in Medical Sciences. He is also the Vice-President (North) of System Dynamics Africa Regional Chapter and an Academic Member of the IEEE Systems, Man, and Cybernetics Society Technical Committee on Computational Collective Intelligence.