K. Ota, M.S. Dao, V. Mezaris, F.G.B. De Natale, "Introduction to Special Issue on Deep Learning for Mobile Multimedia" (Editorial), ACM Transactions on Multimedia Computing Communications and Applications (TOMM), vol. 13, no. 3s, June 2017. DOI:10.1145/3088340. The published paper can be found at http://dl.acm.org/citation.cfm?id=3088340

Introduction to Special Issue on Deep Learning for Mobile Multimedia

Deep learning (DL) has become a crucial technology in the field of multimedia computing. It offers a powerful instrument to automatically produce high-level abstractions of complex multimedia data, which can be exploited in a number of applications including object detection and recognition, speech-to-text, media retrieval, multimodal data analysis, and so on. The joint availability of affordable large-scale parallel processing architectures and effective open-source codes implementing learning and inference algorithms, attracted a huge interest on DL within the research community, bringing to the development of a number of well-performing technologies, increasingly transferred to real-world applications. In recent years, the possibility of implementing DL technologies on mobile devices gains significant attention. Any mobile device that holds some sensing and processing capability, has the potential to become a smart object capable of learning and acting, either stand-alone or interconnected with other intelligent objects. In this context, DL not only can boost the performance of mobile multimedia applications availably, but could also pave the way towards more sophisticated uses of mobile devices.

The path towards these exciting future scenarios, however, entangles a number of important research challenges. The fundamental DL technologies, including deep neural network architectures, training and inference methods, and so on, are hardly adapted to the capabilities of mobile and wireless multimedia environments. Therefore, new generations of mobile processors and chipsets are required to support intensive and parallel computation; small footprint learning algorithms have to be developed to fit lower computation and lower power consumption requirements; new models of collaborative and distributed processing are needed to deal with higher-complexity tasks; and a number of other challenges have to be addressed in order to ensure reliable, efficient, and realtime DL technologies for mobile multimedia.

The first paper of this special issue, authored by the Guest Editors, presents a survey on DL over mobile, starting from a general overview of neural models, and then focusing on hardware and software methodologies and tools for efficient learning and inference in the presence of limited resources. The paper also reviews some of the most interesting applications so far proposed in mobile DL and highlights future research directions.

The following papers introduce a variety of application-driven DL solutions that advance the state-of-the-art in DL methodologies for mobile multimedia. Lorenzo Seidenari, Claudio Baecchi, Tiberio Uricchio, Andrea Ferracani, Marco Bertini and Alberto Del Bimbo, in the paper “Deep artwork detection and retrieval for automatic context aware audio guides” address the problem of creating a smart audio guide that adapts to the actions and interests of museum visitors. The proposed smart audio guide is backed by a computer vision system capable of working in real-time on a mobile device, coupled with audio and motion sensors. A compact Convolutional Neural Network (CNN) is adopted to perform object classification and localization. To improve the recognition accuracy, the authors perform additional video processing using shape-based filtering, artwork tracking and temporal filtering. Several experiments have been conducted to prove its effectiveness

The paper “Mobile multi-food recognition using deep learning”, authored by Parisa Pouladzadeh and Shervin Shirmohammadi, proposes a mobile food recognition system that uses the picture of the food, taken by the user's mobile device, to recognize multiple food items in the same meal and, eventually, estimate calories and nutrition facts of the meal. Since the application requires an almost realtime response, more demanding components of the system (i.e., food recognition and calories estimation) are offloaded to the cloud. Simulations have been conducted to support the studies.

The paper "Enhancing transmission collision detection for distributed TDMA in vehicular networks", authored by Sailesh Bharati, Hassan Aboubakr Omar and Weihua Zhuang, proposes a method to improve the D-TDMA performance, by differentiating transmission failures due to poor channel or transmission collisions. The proposed method is based on the application of a Markov chain model to estimate the channel state when a transmission failure occurs. The parameters of the Markov model are dynamically updated by each node (i.e., vehicle or road-side unit) based on the information conveyed by safety messages that are periodically received from neighboring nodes. Additionally, looking at the D-TDMA protocol headers of received messages, a node approximately determines the error in estimating the channel state, and then uses it to further improve subsequent channel state estimations.

The paper "Spott: on-the-spot e-commerce for television using deep learning based video analysis techniques", by Florian Vandecasteele, Karel Vandenbroucke, Dimitri Schuurman and Steven Verstockt, presents an innovative second-screen mobile multimedia application, which provides the consumer with relevant information on objects (e.g., clothing, furniture, food) that they see and like while watching TV programs. The authors show how DL-based video analysis techniques facilitate video summarization, semantic key-frame clustering, and similarity-based object retrieval. They also give an insight on user trials that have been performed to evaluate and optimize the user experience.

"A tucker deep computation model for mobile multimedia feature learning", authored by Qingchen Zhang, Laurence T. Yang, Xingang Liu, Zhikui Chen and Peng Li, proposes a deep computation model by using the Tucker decomposition, to compress the weight tensors in the full-connected layer for multimedia feature learning. Evaluation results show that the proposed deep computation model can achieve a large parameters reduction and speed-up with a limited loss of accuracy in multimedia feature learning.

Overall, the collection of articles in this special issue provides a stimulating overview on some of the more interesting research directions and applications in DL for mobile multimedia. We hope that researchers active in the relevant scientific areas could benefit and get inspiration from these works. We sincerely thank all the authors who submitted their papers to the special issue, as well as the reviewers who contributed with their comments and observations to the high quality of selected articles. Last but not least, we thank editorial members of ACM TOMM, Prof. Alberto Del Bimbo, Prof. Shervin Shirmohammadi, and Prof. Stefano Berretti, for their help and trust to realize this special issue.

Guest Editors
KAORU OTA, Muroran Institute of Technology
MINH SON DAO, Universiti Teknologi Brunei
VASILEIOS MEZARIS, ITI-Centre for Research and Technology Hellas
FRANCESCO G.B. DE NATALE, DISI-University of Trent