EDITORIAL

Editorial for the ICMR 2019 special issue

K. Selçuk Candan[1] · Marco Bertini[2] · Xiao-Yong Wei[3] · Lexing Xie[4]

As the readers of this journal will attest, effectively and efficiently retrieving multimedia information is one of the most critical, yet challenging, tasks in a broad spectrum of application domains and multimedia retrieval continues to be one of the fastest growing research areas in multimedia computing. Along with this journal, the ACM International Conference on Multimedia Retrieval (ICMR) is the premier forum of knowledge exchange for researchers and practitioners of multimedia retrieval algorithms, tools, and systems. The scope of the conference, which emerged through the fusion of the long-running and very successful conference series, ACM International Conference on Image and Video Retrieval (CIVR) and ACM International Conference on Multimedia Information Retrieval (MIR), includes core topics in multimedia retrieval and recommendation (such as large-scale and multimodal indexing), as well as the broader set of topics (such as mixed reality user interaction) that must be addressed to ensure that multimedia retrieval technologies are of practical use in real-world use cases.

In 2019, the 20th edition of this conference series was organized in Ottawa, Canada. The ICMR'19 conference program included a broad range of papers presenting significant and innovative research in multimedia retrieval and related fields. For this special issue, we selected four papers that represent some of the most creative ways researchers in our community are pushing the state of the art by addressing new problems or proposing insightful solutions: "Hierarchical attentive deep neural networks for semantic music annotation through multiple music representations" by Qianqian Wang, Feng Su, and Yuyang Wang; "ContextNet: representation and exploration for painting classification and retrieval in context" by Noa Garcia, Benjamin Renoust,

and Yuta Nakashima; "Characterization and classification of semantic image-text relations" by Christian Otto, Matthias Springstein, Avishek Anand, and Ralph Ewerth; and "The Focus–Aspect–Value model for predicting subjective visual attributes" by Philipp Blandfort, Tushar Karayil, Jörn Hees, and Andreas Dengel. These papers were candidates to the best paper award or received particularly good comments and marks from the reviewers and highlight the depth and richness of the technical and application scopes of the conference.

In "Hierarchical attentive deep neural networks for semantic music annotation through multiple music representations," Wang et al. focus on the problem of automatically associating semantics tags on music pieces. The authors argue, and experimentally show, that this challenging task can be best achieved through leveraging multiple complementary representations supported by a dual-state LSTM network that captures correlations between representation channels and an attention mechanism that helps identify appropriate aggregation weights.

"ContextNet: representation and exploration for painting classification and retrieval in context," by Garcia et al. focuses on the challenge of automatic art analysis, in particular on the difficult task of identifying and representing subtle relationships among artistic attributes. The authors show that painting classification and retrieval tasks can be substantially improved by neural network-based models that not only integrate content and the style, but also encode relationships between different artistic attributes captured through multi-task learning and knowledge graphs.

In "Characterization and classification of semantic image-text relations," Otto et al. argue that visual and textual modalities are both critical in capturing complementary information on various application domains, including entertainment, news, advertisements, science, or education. Building on this argument, authors present a set of metrics that help quantify and categorize various complex image–text relationships and present a deep learning system to predict these three metrics, along with eight image–text classes.

✉ K. Selçuk Candan
candan@asu.edu

1 Arizona State University, Tempe, USA

2 University of Florence, Florence, Italy

3 Sichuan University, Chengdu, China

4 Australian National University, Canberra, Australia

The final article, "The Focus–Aspect–Value model for predicting subjective visual attributes," included in this special issue presents a novel model for capturing subjective tendencies in visual interpretation, with applications in subjective information retrieval, affective image captioning, or answering subjective questions about images. The model breaks down the process of subjective image interpretation by disentangling three components of subjectivity: "focus" (i.e., the center of attention), "aspect" (i.e., the dimension of interest), and "value" (i.e., the result of the evaluation). Authors presented and evaluated several deep learning models and have shown that the proposed approach, especially using TensorFusion, significantly improves zero-shot learning to predict unseen noun-attribute combinations, by effectively incorporating higher-order interactions between image features and noun context.

All papers presented in ICMR conferences are selected through a robust reviewing process by experts in the multimedia retrieval field to ensure relevance, technical quality, and novelty. The above four articles selected for inclusion in this special issue went through an additional round of shepherding process to further ensure that they are substantial extensions of the articles included in the ICMR2019 proceedings. Therefore, this special issue would not have been possible without the contributions of the program committee members, reviewers, and shepherds, in addition to the high-quality technical works of the authors. We also thank and acknowledge the assistance of the IJMIR editorial office during the preparation of the special issue.

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