

## Guest Editorial: Representation Learning for Multimedia Data Understanding

Yan Yan<sup>1</sup> · Zhigang Ma<sup>2</sup> · Bingbing Ni<sup>3</sup>

Published online: 28 July 2016

© Springer Science+Business Media New York 2016

During the past decade, there is an increasing popularity of video on-demand due to the exponential growth of the user generated videos and the prevailing videos sharing communities such as YouTube, Hulu, etc. Semantically, understanding the content of these multimedia data can substantially enhance applications based on the large-scale multimedia data. However, the performance of multimedia understanding system is heavily dependent on the choice of multimedia data representation. Therefore, developing optimal feature representation for multimedia data is the crucial step for the multimedia data understanding.

This special issue serves as a forum for researchers all over the world to discuss their works and recent advances in representation learning methods and its applications in multimedia analysis. This issue consists of 17 papers, which are briefly discussed as follows.

There are two papers addressed multimedia search and retrieval in this issue, i.e., (1) “Multiple Kernel Visual-Auditory Representation Learning for Retrieval” ([10.1007/s11042-016-3294-5](https://doi.org/10.1007/s11042-016-3294-5)) and (2) “Unsupervised Multi-Graph Cross-Modal Hashing for Large-Scale Multimedia Retrieval” ([10.1007/s11042-016-3432-0](https://doi.org/10.1007/s11042-016-3432-0)).

Moreover, novel machine learning approach for representation learning is important for multimedia data understanding. This issue consists of three papers about this topic, i.e., (1) “Graph-based representation learning for automatic human motion segmentation” ([10.1007/s11042-016-3480-5](https://doi.org/10.1007/s11042-016-3480-5)), (2) “A collaborative recommender system enhanced with particle swarm optimization technique” ([10.1007/s11042-016-3481-4](https://doi.org/10.1007/s11042-016-3481-4)), and (3) “Spatial-Dictionary for Collaborative Representation Classification of Hyperspectral Images” ([10.1007/s11042-015-3098-z](https://doi.org/10.1007/s11042-015-3098-z)).

---

✉ Yan Yan  
yan@disi.unitn.it

<sup>1</sup> Department of Information Engineering and Computer Science, University of Trento, Trento, Italy

<sup>2</sup> Department of Computer Science, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA

<sup>3</sup> School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China

Representation learning via deep learning framework is a hot topic these days. This issue consists of two papers about this topic, i.e., (1) “A Deep Semantic Framework for Multimodal Representation Learning” ([10.1007/s11042-016-3380-8](https://doi.org/10.1007/s11042-016-3380-8)) and (2) “Occluded Vehicle Detection with Local Connected Deep Model” ([10.1007/s11042-015-3141-0](https://doi.org/10.1007/s11042-015-3141-0)).

There are many interesting real-world multimedia applications based on representation learning. This issue consists of ten papers about this topic, i.e., (1) “Sub-event Recognition and Summarization for Structured Scenario Photos” ([10.1007/s11042-016-3346-x](https://doi.org/10.1007/s11042-016-3346-x)), (2) “People-flow counting in complex environments by combining depth and color information” ([10.1007/s11042-016-3344-z](https://doi.org/10.1007/s11042-016-3344-z)), (3) “Graph Modeling and Mining Methods for Brain Images” ([10.1007/s11042-016-3482-3](https://doi.org/10.1007/s11042-016-3482-3)), (4) “Lagrangian Twin Support Vector Regression and Genetic Algorithm based Robust Grayscale Image Watermarking” ([10.1007/s11042-016-3381-7](https://doi.org/10.1007/s11042-016-3381-7)), (5) “Transfer Useful Knowledge for Headpose Estimation from Low Resolution Images” ([10.1007/s11042-016-3297-2](https://doi.org/10.1007/s11042-016-3297-2)), (6) “A Rapid Method for Detecting Objects with Rectangular Structures Based on Line Correspondences” ([10.1007/s11042-016-3345-y](https://doi.org/10.1007/s11042-016-3345-y)), (7) “A real-time object tracking via L2-RLS and compressed Haar-Like features matching” ([10.1007/s11042-016-3356-8](https://doi.org/10.1007/s11042-016-3356-8)), (8) “Local Abnormal Behavior Detection Based on Optical Flow and Spatio-temporal Gradient” ([10.1007/s11042-015-3122-3](https://doi.org/10.1007/s11042-015-3122-3)), (9) “Feature Pattern Based Representation of Multimedia Documents for Efficient Knowledge Discovery” ([10.1007/s11042-016-3434-y](https://doi.org/10.1007/s11042-016-3434-y)), and (10) “Tracklet Association Based Multi-target Tracking” ([10.1007/s11042-015-3238-5](https://doi.org/10.1007/s11042-015-3238-5)).

All these 17 papers cover a wide range of methods and applications about representation learning for multimedia data understanding. This special issue serves as a forum for researchers all over the world to discuss their works and recent advances in representation learning for multimedia data understanding. We hope this issue appeal to both the experts in the field as well as to those who wish a snapshot of the current breadth of practical multimedia data understanding.