



Editorial

Advances Techniques in Computer Vision and Multimedia

Yang Wang

School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China; yangwang@hfut.edu.cn

Computer vision has experienced significant advancements and great success in areas closely related to human society, which aims to enable computer systems to automatically see, recognize, and understand the visual world by simulating the mechanism of human vision. Multimedia has also changed our lifestyles and is becoming an indispensable part of our daily lives; it mainly includes emerging computing methods for handling various multi-modal media (picture, text, audio, video, etc.) [1] generated by ubiquitous multimedia sensors and infrastructures, including retrieval of multimedia data, analysis of multimedia contents, methodology based on deep learning, and practical multimedia applications.

This summary includes some of the recent advancements in computer vision and multimedia. The Special Issue presents four articles after a careful peer review process. The addressed topics cover recommendations on deep multi-view semantic similarity learning, cross-modal retrieval to multimedia, price tag data analysis in computer vision, and musical internet.

Mining and analyzing massive amounts of network information to provide users with accurate and fast recommendation information has become a hot and difficult topic. However, common social network-based collaborative filtering algorithms suffer from problems such as low recommendation performance and cold start due to high data sparsity and uneven distribution. In addition, these algorithms do not effectively consider implicit trust relationships between users. To address these problems, Song et al. [2] propose a collaborative filtering recommendation algorithm based on GraphSAGE, namely GraphSAGE-CF. More specifically, the algorithm adopts GraphSAGE to learn low-dimensional feature representations of the global and local structures of user nodes in social networks and then captures the implicit trust relationships between users via the feature representations learned by GraphSAGE. Finally, the comprehensive evaluation shows the scores of users and implicit users on related items and predicts the scores of users on target items.

Cross-modal retrieval aims to search for samples of one modality via the queries of other modalities. However, two main challenges, i.e., heterogeneity gap and semantic interaction across different modalities, have not been solved efficaciously. To overcome these challenges, Cai et al. [3] present a novel end-to-end framework called the Dual Attention Generative Adversarial Network (DA-GAN). This technique is an adversarial semantic representation model with a dual attention mechanism, i.e., intra-modal attention and inter-modal attention. Specifically, intra-modal attention is utilized to focus on the important semantic feature within a modality, whereas inter-modal attention is used to explore the semantic interaction between different modalities, which can effectively learn high-level semantic interaction across different modalities. Moreover, a dual adversarial learning strategy that learns a modality-consistent representation is proposed to reduce the heterogeneity gap.

To find an optimal solution for price tag data analysis, Laptev et al. [4] compare neural networks, including Unet, MobileNetV2, VGG16, and YOLOv4-tiny, for image segmentation as part of this study. The neural networks considered are trained on an individual dataset collected by the authors. The study reveals that the optimal neural network for tag analysis using segmentation is YOLOv4-tiny. In addition, this paper covers an automatic image–text recognition approach using EasyOCR API.



Citation: Wang, Y. Advances Techniques in Computer Vision and Multimedia. *Future Internet* **2023**, *15*, 294. <https://doi.org/10.3390/fi15090294>

Received: 30 August 2023
Accepted: 30 August 2023
Published: 1 September 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keller et al. [5] introduce a new perspective on musical interaction tailored to a specific class of sonic resources: impact sounds. This study is informed by the field of ubiquitous music (ubimus) and engages with the demands of artistic practices. Using a series of deployments of a low-cost and highly flexible network-based prototype, the Dynamic Drum Collective, the authors exemplify the limitations and specific contributions of the banging interaction. Three components of this new design strategy, i.e., adaptive interaction, mid-air techniques, and timbre-led design, target the development of creative action metaphors that use resources available in everyday settings. The techniques involving the use of sonic gridwork yield positive outcomes. The subjects choose sonic materials that approach a full rendition of the proposed soundtrack when combined with their actions on the prototype. The results of the study highlight subjects' reliance on visual feedback as a non-exclusive strategy to handle both temporal organization and collaboration.

Funding: The research is supported by National Natural Science Foundation of China under grant number 62172136 and U21A20470.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Wang, Y. Survey on Deep Multi-modal Data Analytics: Collaboration, Rivalry, and Fusion. *ACM Trans. Multimed. Comput. Commun. Appl.* **2021**, *17*, 10. [[CrossRef](#)]
2. Song, J.; Song, J.; Yuan, X.; He, X.; Zhu, X. Graph representation-based deep multi-view semantic similarity learning model for recommendation. *Future Internet* **2022**, *14*, 32. [[CrossRef](#)]
3. Cai, L.; Zhu, L.; Zhang, H.; Zhu, X. Da-gan: Dual attention generative adversarial network for cross-modal retrieval. *Future Internet* **2022**, *14*, 43. [[CrossRef](#)]
4. Laptev, P.; Litovkin, S.; Davydenko, S.; Konev, A.; Kostyuchenko, E.; Shelupanov, A. Neural Network-Based Price Tag Data Analysis. *Future Internet* **2022**, *14*, 88. [[CrossRef](#)]
5. Keller, D.; Yaseen, A.; Timoney, J.; Chakraborty, S.; Lazzarini, V. Banging Interaction: A Ubimus-Design Strategy for the Musical Internet. *Future Internet* **2023**, *15*, 125. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.