

# Introduction to the Special Issue on Towards Multimedia Theoretical and Algorithmic Solutions

This article serves as the foreword to our special issue, *Towards Multimedia Theoretical and Algorithmic Solutions*, organized in collaboration with **ACM Transactions on Multimedia Computing, Communications, and Applications**. It outlines the aims of the special issue, the focus is on novel theoretical and algorithmic solutions across the domains of multimedia computing, communications, and applications, as well as innovative technical contributions to areas such as social good and healthcare.

The call for papers attracted a number of submissions, all of which were extended versions of papers previously published at **The 6th ACM International Conference on Multimedia in Asia (ACM Multimedia Asia 2024)**. This introduction reveals an overview of the topics covered by the accepted articles.

The mission of this special issue is to highlight the state-of-the-art in multimedia computing by bringing together puissant researchers and practitioners in this heartening field, with extreme emphasis on the selected papers from the ACM MM Asia 2024 conference gleefully organized in Auckland New Zealand. The special issue also encourages visionary articles on new and emerging topics, papers presenting novel ideas with promising preliminary results, and application-oriented studies that make innovative technical and scientific contributions to multimedia research.

In this context, the special issue called for innovative algorithms and groundbreaking approaches addressing topics including, but not limited to: (1) Multimedia and Computer Vision; (2) Multimodal Data Analysis; (3) Deep Learning for Multimedia; (4) Emotional and Social Signals in Multimedia; (5) Multimedia Search and Recommendation; (6) Multimedia Human-Computer Interaction and Quality of Experience; (7) Music and Audio Processing in Multimedia; (8) Multimedia Systems; (9) Multimedia Virtual and Augmented Reality; (10) Multimedia for Healthcare, etc.

Accordingly, these topics are of inspiring interest to multimedia community, and the call attracted significant attention. We received a plethora of inquiries and recommend potential authors submit their manuscript to this special issue. Each submitted article was reviewed by at least three expert reviewers and underwent at least one round of revision. Among the accepted papers, the concentration of 11 papers is on multimodal models and data fusion in multimedia computations, covering the research areas such as image captioning, speech emotion recognition, facial emotion analysis, human action recognition, image-text retrieval, image fusion, action fusion, and acoustic feature fusion.

Convolutional Neural Networks (CNNs) and attention mechanisms remain central and fundamental to deep learning-based multimedia computing, communications, and applications. Four articles in this special issue unfold the recent progress in this area, addressing problems such as image binarization, medical image segmentation, and CT image denoising.

In lieu of CNNs combined with attention models, other neural network architectures are explored in this special issue for multimedia applications, including Spiking Neural Networks (SNN), Graph Convolutional Networks (GCN), and Action Detection Networks (ADN), etc. Three articles reflect the advantages of these prevalent architectures in the state-of-the-art multimedia computations.

Generative AI has emerged as a tremendously influential research direction in multimedia computing. In this special issue, four articles investigate generative AI techniques applied to the research areas such

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM 1551-6865/2026/9-ART

<https://doi.org/10.1145/3842761>

as video streaming using reinforcement learning, the reality–virtuality continuum, and automatic music transition.

Multimedia watermarking and privacy protection remain enduring topics in multimedia research. The security-related studies are playing increasingly important role in multimedia data safeguard, particularly in the era of large language models (LLMs) and AI-powered autonomous systems. This special issue encapsulates two articles dedicated to the prominent challenges.

In conclusion, this special issue attracted intense interests from the multimedia computing community, once again highlighting the significance of multimedia theoretical and algorithmic solutions and the immense impact on shaping the future of this frontier and vibrant field.

Ruili Wang

Massey University, Auckland, New Zealand

Wei Qi Yan

Auckland University of technology, Auckland, New Zealand