

Editorial

Editorial for Special Issue on Invited Papers from APSIPA ASC 2024

Jiantao Zhou¹, Yuanfang Guo² and Jin-Yu Tian³

¹*University of Macau, Macau, China; jtzhou@um.edu.mo*

²*Beihang University, China*

³*Macau University of Science and Technology, Macau, China*

We are delighted to present this special issue of *APSIPA Transactions on Signal and Information Processing*, featuring a selection of outstanding papers from the 2024 APSIPA Annual Summit and Conference (APSIPA ASC 2024). This issue reflects the spirit of innovation and international collaboration that defined the conference, showcasing significant research and emerging trends in signal processing and artificial intelligence.

Held in Macau, China, APSIPA ASC 2024 welcomed 525 attendees from 23 countries and regions. The event fostered high-quality scientific exchange and highlighted the latest advancements in the field. Following a rigorous peer-review process, this issue includes four exceptional articles highly recommended by our editors and reviewers.

The first paper, “Image-Text Retrieval via Green Explainable Multi-modal Alignment (GEMMA),” by Tsung-Shan Yang *et al.*, tackles the computational cost and lack of transparency in image-text retrieval models. GEMMA uses frozen, pre-trained encoders and lightweight linear projections to reduce trainable parameters to just 3% of a full model. Its stage-wise alignment process—global, image cluster, and text cluster—leverages K-means clustering and Discriminant Feature Testing to select informative tokens. GEMMA delivers superior performance on Flickr30k and MS-COCO datasets and generalizes well to unpaired data, offering clear visual insights into its alignment mechanism.

The second paper, “ViP-CBM: A Low-parameter and Interpretable Concept Bottleneck Model Using Visual-projected Embeddings,” by Ji Qi *et al.*, introduces a low-parameter Concept Bottleneck Model for multi-label classification. By projecting visual features into a unified embedding space and applying a simple binary rule for concept classification, ViP-CBM reduces training

parameters by over 50% compared to other CBMs. It achieves competitive performance on CUB and AwA2 benchmarks and enhances interpretability through visualization of its embedding space, enabling efficient human intervention.

The third paper, “Multi-step Prediction and Control of Hierarchical Emotion Distribution in Text-to-speech Synthesis,” by Sho Inoue *et al.*, addresses fine-grained control of emotion rendering in TTS. Their hierarchical emotion distribution module quantifies emotion variance at utterance, word, and phoneme levels. By predicting emotion variance in a multi-step manner, the framework refines local emotional variations using global context. Integrated into a variance adaptor and compatible with various TTS systems, the approach significantly improves emotional expressiveness and control across speech granularities.

The fourth paper, “Music Similarity Representation Learning Focusing on Individual Instruments with Source Separation and Human Preference,” by Takehiro Imamura *et al.*, proposes a method for learning music similarity representations for individual instruments. Using source separation and human preference data, the model improves performance without requiring clean instrument stems. Techniques such as end-to-end fine-tuning, multi-task learning, and perception-aware tuning enhance both objective and perceptual similarity, with the best results achieved through combined source separation and human-guided fine-tuning.

This special issue exemplifies the depth and diversity of research presented at APSIPA ASC 2024. The selected papers demonstrate innovative approaches to pressing challenges in signal processing and AI, offering practical solutions with strong interpretability and performance. We hope this collection inspires further exploration and collaboration across the global research community.

Guest Editors

Jiantao Zhou, Yuanfang Guo and Jin-Yu Tian