Call for Papers

IEEE Transactions on Circuits and Systems for Video Technology

Special Issue on Video Transformers

Summary

With the development of the Internet and technology, millions of videos are uploaded to the social platform every day. Video usage has exploded and is now one of our main communication channels. Effective video analysis and processing approaches impose great opportunities for many practical applications, such as video understanding, recommendation, matching, compression, and generation, which play important roles in public security, social media, entertainment, healthcare, etc. However, due to the specific video structure (such as spatial and temporal coherence) and large dimensionality, it is very challenging to effectively analyze the videos.

Transformer models have achieved great success in the past few years, especially on language and image-related tasks. Benefitting from the self-attention operation, transformers can well model long-range interactions and require minimal inductive biases, which also makes it a promising tool for solving video-related tasks but demands some adaptations and specific network designs. Moreover, video usually consists of multiple modalities, such as audio, text, and image. By using similar processing blocks, the transformer can well process the input of different modalities as well as their cross-modal interactions, showing its good flexibility. Besides, it can be easily extended to large capacity networks and large-scale datasets, demonstrating its great potential in video-related tasks.

This special issue seeks high-quality and original contributions towards advancing the architecture, theory, and algorithmic design of video transformers. We envision original and well-motivated adaptations of transformer models for video tasks and efforts towards improving their accuracy, robustness, and efficiency. The special issue will provide a timely collection of recent advances to benefit the researchers and practitioners working in the broad research field of multimedia analysis, computer vision, and machine intelligence.

Scope

This special issue seeks original contributions from, but not limited to, the following topics:

- Novel transformer-based methods for high-level video understanding such as video-based activity recognition, object detection, segmentation, tracking, summarization, localization, and pose estimation

- Novel transformer-based approaches for low-level video processing tasks such as video deblurring, de-raining, denoising, compression, and so on
- Unsupervised, weakly supervised, and semi-supervised learning for video with transformer models
- Efficient transformer architectures, including novel mechanisms for self-attention
- Transformer-based multi-modal learning that incorporates visual data with text, audio, and knowledge graphs
- Hybrid network designs combining the strengths of transformer models with convolutional and graph-based models
- Novel transformer models for large-scale video pretraining
- Novel transformer models for video generation
- Theoretical insights into transformer-based models

Important Dates:

Open for submissions: 1 January 2023
Submissions due: 1 March 2023
Preliminary notification: 1 May 2023
Revisions due: 1 July 2023
Notification: 1 October 2023
Final manuscripts due: 1 November 2023
Publication (tentative): December 2023

Guest Editors:

Dr. Liqiang Nie, Professor, Harbin Institute of Technology (Shenzhen), China (email: nieliqiang@gmail.com)
Dr. Jianlong Wu, Associate Professor, Harbin Institute of Technology (Shenzhen), China (email: wujianlong@hit.edu.cn)
Dr. Nicu Sebe, Professor, University of Trento, Italy (email: sebe@disi.unitn.it)
Dr. Kiyoharu Aizawa, Professor, The University of Tokyo (email: aizawa@hal.t.u-tokyo.ac.jp)