IEEE Transactions on Circuits and Systems for Video Technology Special Issue on Large Language Models (LLMs) for Video Understanding

Summary:

The rise of Large Language Models (LLMs) marks a major step in language understanding, bringing us closer to Artificial General Intelligence (AGI). Techniques like Chain of Thought, Reinforcement Learning from Human Feedback (RLHF), and Instruction Tuning have boosted LLMs' ability to handle complex tasks. Their versatility in vision-language tasks, without extensive retraining, has gained attention in the computer vision field.

In video understanding, LLMs can process large spatio-temporal data, improving tasks like summarization and sentiment analysis. However, challenges remain, such as enhancing cross-frame context, measuring real-world impact, and improving efficiency for large-scale video data. Tackling these issues could lead to major advancements in AI-driven video comprehension. The objective of this dedicated special issue is to foster advanced research in this domain and present a timely compilation of contributions that can prove valuable to both researchers and practitioners.

We welcome high-quality original submissions addressing important novel theories, methods, applications, and insights centred on LLMs for video understanding. The list of possible topics includes, but is not limited to:

- Implications of using LLMs for video classification, action recognition, object detection and tracking, segmentation, captioning, and other video understanding tasks.
- Zero/few-shot video representation learning through pre-training strategies of LLMs, such as self-supervised learning, unsupervised learning, and multi-task learning.
- Technical advances in Multi-modal foundation models, including vision-language foundation models, video-language foundation models, and vision-language-action foundation models.
- Applications of LLMs with video understanding across various industries and interdisciplinary fields, such as intelligent manufacturing, robotics, smart city, biomedicine, and geography.
- Exploring the capability of combining LLMs with the diffusion model to enhance accessibility and diversity in the generating or editing of video content.
- Overcoming the technical obstacles associated with utilizing LLMs for video comprehension, encompassing concerns regarding explainability and security.

Important Dates:

Submission deadline: December 1, 2024
First review notification: January 20, 2025
Revision submission due: March 1, 2025
Second round review: March 1 to April 20, 2025
Notification of acceptance/rejection: May 1, 2025

Guest Editors:

Jungong Han, University of Sheffield, UK, jungonghan77@gmail.com
Liqi Yan, Hangzhou Dianzi University, China, lqyan18@fudan.edu.cn
Zheng (Thomas) Tang, NVIDIA, USA, tangzhengthomas@gmail.com
Dongfang Liu, Rochester Institute of Technology, USA, dongfang.liu@rit.edu
Zhuang Shao, Newcastle University, UK, zhuang.shao@newcastle.ac.uk
Roland Goecke, University of New South Wales Canberra, Australia, r.goecke@unsw.edu.au

Digital Object Identifier 10.1109/TCSVT.2024.3471329