

Call for Papers

IEEE Transactions on Multimedia Special Issue on

“Pre-trained Models for Multi-modality Understanding”

Pre-trained models (PTMs) have significantly boosted the performance on a broad range of natural language processing (NLP) topics such as neural machine translation, question answering, named entity recognition, text summarization, dialogue, and sentiment analysis. The great success of PTMs in NLP has motivated the researchers in multimedia community to explore pre-trained models for multi-modality understanding. As widely recognized by the community, representation learning is the core problem for multi-modality understanding. Pre-trained models provide better initialization parameters to ameliorate feature representation for downstream tasks. As a result, the paradigm of leveraging pre-trained models has demonstrated exemplary performance to date.

Despite attracting a surge of research interests, some core issues for pre-training technology, such as multi-modality data alignment, pretext tasks designing, and efficient pre-trained models, are still open problems. Besides, the research on how to better exploit pre-trained models for downstream tasks even lags far behind that of pre-training technologies. As the main efforts are devoted to pre-training, the common practice of leveraging pre-trained models is simply fine-tuning. The discussion of whether “pre-training and fine-tuning” is the best paradigm of leveraging pre-trained models for multi-modality understanding is still under-explored.

This special issue is focused on two main aspects: (a) exploration of pre-training algorithms, and (b) investigation on how to make the best of pre-trained models for downstream tasks. We prospect original contribution towards theories, algorithms, model architecture design, as well as novel applications of pre-trained models for multi-modality understanding. The special issue will provide a timely collection of recent advances to benefit the researchers and practitioners working in the broad research field of the multimedia community. Topics of interest include (but are not limited to):

- Novel pre-training framework
- New network architectures for multi-modality understanding
- Prompt tuning for leveraging pre-trained models
- Network adapters for leveraging pre-trained models
- Visual perception based multi-modal pre-trained models
- Image and video synthesis/generation based on multi-modal pre-trained models
- Vision-language understanding
- Multi-modality fusion
- Open-set problems for multi-modality understanding
- New dataset for multi-modality understanding

Prospective authors should submit their manuscripts following the IEEE TMM guidelines at <https://signalprocessingsociety.org/publications-resources/manuscript-submission-guidelines-sp-magazine-feature-special-issue-articles>. Authors should submit a PDF version of their complete manuscript to <https://mc.manuscriptcentral.com/tmm-ieee> according to the following schedule:

- Submission deadline: January 15, 2023
- First Review: April 1, 2023
- Revisions due: June 1, 2023
- Second Review: August 15, 2023
- Final Manuscripts: September 15, 2023
- Publication date: September 30, 2023

Guest Editors:

- Wengang Zhou: University of Science and Technology of China, China (email: zhwg@ustc.edu.cn)
- Jiajun Deng: The University of Sydney, Australia (email: djiajun1206@gmail.com)
- Niculae Sebe: University of Trento, Italy (email: sebe@disi.unitn.it)
- Qi Tian: Huawei, China (email: tian.qi1@huawei.com)
- Alan L. Yuille: Johns Hopkins University, USA (email: ayuille1@jhu.edu)
- Concetto Spampinato: Università degli Studi di Catania (email: concetto.spampinato@unict.it)
- Zakia Hammal: Carnegie Mellon University (email: zhammal@andrew.cmu.edu)