

Video Understanding and Generation with Multimodal Foundation Models

Dr. Ming-Hsuan Yang

Professor, Electrical Engineering and Computer Science, University of California at Merced, USA

Monday, Mar 24, 2025 10:00am
Auditorium 106 at IIS new Building



Abstract

Recent advances in vision and language models have significantly improved visual understanding and generation tasks. In this talk, I will present our latest research on designing effective tokenizers for transformers and our efforts to adapt frozen large language models for diverse vision tasks. These tasks include visual classification, video-text retrieval, visual captioning, visual question answering, visual grounding, video generation, stylization, outpainting, and video-to-audio conversion. If time permits, I will also discuss our recent findings in dynamic 3D vision.

Biography

Ming-Hsuan Yang is a Professor at the University of California, Merced, and a Research Scientist at Google DeepMind. He has received numerous prestigious awards, including the Google Faculty Award (2009), the NSF CAREER Award (2012), and the Nvidia Pioneer Research Award (2017, 2018). He received the Best Paper Honorable Mention at UIST 2017, CVPR 2018, and ACCV 2018, the Longuet-Higgins Prize for Test of Time at CVPR 2023, and Best Paper at ICML 2024. Yang is an Associate Editor-in-Chief of IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) and an Associate Editor for the International Journal of Computer Vision (IJCV). Previously, he was the Editor-in-Chief of Computer Vision and Image Understanding (CVIU) and Program Co-Chair for ICCV 2019. He is a Fellow of IEEE, ACM, and AAAI.

