KEYNOTE TALK
Wednesday, October 23, 2024, at 9:00am

Perception from Generation
Gang Hua
Dolby Laboratories

Abstract: Deep generative models have shown great capability in learning representations for downstream machine perception tasks. To this end, we will present two pieces of our recent research work on 1) how representations learnt from a text-to-video generative model enables spatiotemporal consistent video understanding, and 2) how representations learnt from an implicit auto-encoder can better generalize for 3D perception from point clouds. We will explain in detail why such generative representation could facilitate corresponding perception tasks. We argue that with the capacity of deep generative model increases to fit the distribution of large dataset, it is the right time for us to revisit the analysis by synthesis Bayes regime for computer vision. The analysis by synthesis approach was considered "ahead of its time" mainly due to the lack of generative models that can very well model the prior distribution of "world". To this end, we will present our most recent work on dexterous grasping, where a unified generative diffusion model is proposed along with a physics-based discriminator to achieve new state-of-the-art following the analysis by synthesis approach. We conclude the talk with some of our own reflections on future research directions of Generative AI, and if time allows, discuss the notion of "human-in-control" AI for media generation.

Speaker Bio-Sketch: Gang Hua is the Vice President of Multimodal Experiences Research Lab at Dolby Laboratories. His research focuses on computer vision, pattern recognition, machine learning, robotics, towards Artificial General Intelligence, with primary applications in cloud and edge intelligence, and currently with a focus on media and entertainment. Before that, he was CTO of Convenience Bee, and Chief Scientist of its research branch, Wormpex AI Research. He served in various roles at Microsoft (2015-18) as the Science/Technical Adviser to the CVP of the Computer Vision Group, Director of Computer Vision Science Team in Redmond and Taipei ATL, and Senior Principal Researcher/Research Manager at Microsoft Research. He was an Associate Professor at Stevens Institute of Technology (2011-15). During 2014-15, he took leave and worked on the Amazon-Go project. He was a Visiting Researcher (2011-14) and a Research Staff Member (2010-11) at IBM Research T. J. Watson Center, a Senior Researcher (2009-10) at Nokia Research Center Hollywood, and a Scientist (2006-09) at Microsoft Live labs Research. He received a Ph.D. degree in Electrical and Computer Engineering from Northwestern University in 2006, and a M.S. in pattern recognition and intelligence system from Xi'an Jiaotong University (XJTU) in 2002. He was selected to the Special Class for the Gifted Young of XJTU in 1994 and received a B.S. in Electrical Engineering in 1999. He is the recipient of the 2015 IAPR Young Biometrics Investigator Award. He is an IEEE Fellow, an IAPR Fellow, and an ACM Distinguished Scientist. He has published over 260 peer reviewed papers in top journals and conferences. To date, he holds 19 US patents and has 15 more patents pending. He was a Program Chair for CVPR2019 and CVPR2022.