Intelligent Network Infrastructure for Extended Reality with AI-System Co-design

Dr. Bo Chen, University of Illinois at Urbana-Champaign

Abstract: Network infrastructure for Extended Reality (XR) encompasses an end-to-end design of content management, data transmission, and interactive rendering. However, traditional network infrastructure suffers from sub-optimal designs such as discrete content representation, handcrafted compression, and redundancy-based fault tolerance, which fail to deliver photo-realistic, smooth, and reliable XR experiences. Artificial Intelligence (AI) offers promising solutions to these problems by being generic, jointly optimized, and generative, but AI-based systems often struggle to meet practical requirements for XR such as throughput and latency.

In this talk, I will showcase how to tackle this challenge with the intelligent network infrastructure for XR, which is featured by a co-design of AI and systems: (1) Adapting system optimization based on the systematic abstraction of evolving AI models, and (2) Customizing AI designs according to the analysis of the XR system's context information. I will first present NeRFHub, a framework designed to serve emerging, photo-realistic AI-driven representations (*e.g.,* Neural Radiance Fields) to mobile devices. The key challenge is the prohibitive demand for computation and bandwidth resources of AI-driven representations. This research introduces a novel way for adapting AI-driven representations by formulating a constrained multi-objective optimization (MOO) problem, which is jointly solved with classic design space exploration techniques and fine-tuning of AI models. Furthermore, I will introduce LiFteR, a bandwidth-efficient video streaming system powered by AI-based video compression. The key limitation of existing AI-based video compression is its slow decoding speed, leading to issues such as rebuffering. LiFteR accelerates AI-based compression with a binary-tree-based workflow and a multi-frame codec design, enabling parallel acceleration of frame processing without compromising bandwidth efficiency.

Biography: Bo Chen is a postdoctoral researcher at University of Illinois at Urbana-Champaign (UIUC). He obtained his Ph.D. from UIUC advised by Prof. Klara Nahrstedt in 2022. Previously, he received his BS degree from Shanghai Jiao Tong University in 2016. His research interests include networking, machine learning systems, virtual reality, and mobile computing. His research has resulted in more than 20 publications and has been recognized by top conferences such as USENIX NSDI, ACM MobiCom, ACM MobiSys, and ACM SenSys. He received the best student paper award in ACM MMSys and the best paper award in IEEE ISM. He worked as the publication chair for IEEE SECON 2023 and NSF Workshop on Sustainable Computing for Sustainability 2024.