

Chao Huang

250 Hutchison Road, University of Rochester, Rochester, NY 14627
✉ chuang65@cs.rochester.edu • 📞 +1 (585) 910-9360 • 🌐 <http://wikichao.github.io>

EDUCATION

University of Rochester

Ph.D. in Computer Science

Jan 2021 – Present

Rochester, NY

Nanjing University

B.Eng. in Electronic Science and Engineering

Sep 2015 – Jun 2019

Nanjing, China

RESEARCH INTERESTS

Multimodal Learning (Audio-Visual Learning), 3D Vision, Learning from Egocentric Videos

RESEARCH EXPERIENCE

University of Rochester

Advisor: Chenliang Xu

Jan 2021 – Present

Rochester, NY

Project: **Egocentric Audio-Visual Scene Understanding**

- Motivation: In egocentric recordings, a visual-only system struggles to fully decode the surrounding information due to the limited Field-of-View (FoV). Instead, audio provides stable and persistent signals associated with the depicted events, which makes the integration of auditory and visual signals indispensable for egocentric video understanding.
- We formalize a new task: Egocentric Audio-Visual Object Localization, to explore learning fine-grained audio-visual association in egocentric videos. Concretely, we propose a new framework including mitigating out-of-view sounds with visual guidance and addressing egomotion with a geometry-aware temporal modeling module.
- We annotate and create the *Epic Sounding Object dataset* for evaluation purposes. Our method outperforms other third-person view localization approaches and obtain more precise localization maps.

Project: **Learning to Synthesize Audio-Visual Scene**

- We define novel audio-visual scene synthesis as a task to synthesize a target video, including visual frames and the corresponding spatial audio, along an arbitrary camera trajectory from given source videos and trajectories.
- We introduce a novel acoustic-aware audio generation method to encode prior knowledge of sound propagation and propose a coordinate transformation mechanism for effective direction expression. In addition, we introduce a binaural audio augmentation method to improve the acoustic supervision.
- We focus on scene synthesis in the real-world and quantitatively/qualitatively demonstrate advantages of our method.

The Chinese University of Hong Kong

Advisor: Chi-Wing Fu

Jul 2019 – Dec 2020

Shatin, HK

Project: **Non-local Part-aware Point Cloud Denoising**

- Explored non-local semantically-related features by formulating the non-local learning unit (NLU) to aggregate features on the whole 3D shape, and designed the graph attention module to consider feature similarity and focus on aggregating features with more similar underlying geometries.
- Progressively separated the noise features from the cleaned features by cascading a series of NLU and trained the network to encourage the points to become more evenly located on the underlying surface with shape-wise and part-wise regularization.
- Conducted extensive experiments to quantitatively and qualitatively evaluate our method, and demonstrated its superiority over state-of-the-art method.

Nanjing University

Advisor: Zhan Ma

Sep 2018 – Jun 2019

Nanjing, China

Project: **Extreme Image Compression**

- Adopted generative adversarial optimization for extreme image compression that can be used in situations such as communication in depth and web snapshot
- Proposed a novel Multi-Scale AutoEncoder framework, that constructed a coarse-to-fine image coding pipeline to better preserve the global structure and local details
- Reconstructed the images with an acceptable perceptual quality at a low bitrate (like 0.03bpp), which is a difficult case for other popular image codecs

**OTHER WORK
EXPERIENCE**

Virtual Reality and Imaging Research Centre, the Chinese University of Hong Kong

- Research Assistant, Jul 2019 – Dec 2020
 - Explored high-level 3D vision especially point cloud understanding for better shape generation and editing
 - Developed novel deep neural network based algorithms for low-level 3D point cloud processing tasks like point cloud upsampling and denoising

Aqueti(China) Technology Inc., Co., Suzhou, Jiangsu, China

- Assistant Research and Develop Engineer, Jun 2018 – Aug 2018
 - Established a multi-view camera system and collected image data from different objects and scenes to establish a training dataset
 - Developed novel convolutional neural network based algorithms for the array cameras' data processing pipeline

PUBLICATIONS

- [1] **Chao Huang**, Yapeng Tian, Anurag Kumar, and Chenliang Xu. Egocentric Audio-Visual Object Localization. *In Submission*, 2023.
- [2] Susan Liang, **Chao Huang**, Yapeng Tian, Anurag Kumar, and Chenliang Xu. AV-NeRF: Learning Neural Fields for Real-World Audio-Visual Scene Synthesis. *In Submission*, 2023.
- [3] **Chao Huang**, Yapeng Tian, Anurag Kumar, and Chenliang Xu. Audio-Visual Object Localization in Egocentric Videos. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2022.
- [4] **Chao Huang***, Ruihui Li*, Xianzhi Li, and Chi-wing Fu. Non-local Part-Aware Point Cloud Denoising. *Tech report, arXiv*, 2020. (* joint 1st authors)
- [5] **Chao Huang**, Haojie Liu, Tong Chen, Qiu Shen, and Zhan Ma. Extreme Image Compression via Multiscale Autoencoders with Generative Adversarial Optimization. In *IEEE VCIP oral presentation*, 2019.
- [6] Xuefei Yan, David J Brady, Weiping Zhang, Changzhi Yu, Yulin Jiang, Jianqiang Wang, **Chao Huang**, Zian Li, Zhan Ma. Compressive Sampling for Array Cameras. In *SIAM Journal on Imaging Science*, 2021.

**PROGRAMMING
SKILLS**

- Language: MATLAB, C/C++, Python
- Frameworks: Pytorch, Tensorflow