Xichen Pan

xichenpan.cn

E-mail : xcpan.mail@gmail.com
Mobile : +86 186 535 00448

Education

- **University of Waterloo** — Waterloo, ON
  Ph.D. in Computer Science, advised by Prof. Wenhu Chen — *2022 – 2027 (expected)*
  affiliated with the **Vector Institute**
- **Shanghai Jiao Tong University** — Shanghai, China
  B.Eng. in Computer Science (Outstanding Graduate), advised by Prof. Zhouhan Lin — *Sept. 2018 – June 2022*
  Overall: 88.42/100, Major: 91.29/100

Research Intersects

- **Multimodal Deep Learning**: Interested in building multimodal deep learning systems (including audio-visual and vision-language), especially multimodal representation learning and multimodal pretraining.
- **Speech Recognition**: Interested in speech recognition decoding algorithms, especially building efficient and accurate training and inference methods.
- **Knowledge Representation**: Interested in knowledge representation learning and knowledge base QA.

Publications & Manuscripts

- Xichen Pan, Zekai Li, Yichen Gong, Xinbing Wang, and Zhouhan Lin. **Towards Diverse Lip Reading Representations**, *EMNLP 2022 under review*
- Xichen Pan. **Multimodal Audio-Visual Speech Recognition System Based On Pre-trained Models**, *Bachelor Thesis at Shanghai Jiao Tong University (Best Thesis Award, 1st/150)* [news]
- Xichen Pan, Peiyu Chen, Yichen Gong, Helong Zhou, Xinbing Wang, and Zhouhan Lin. **Leveraging Unimodal Self-Supervised Learning for Multimodal Audio-Visual Speech Recognition**, *ACL 2022 Main Conference* [pdf]

Experience

- **Horizon Robotics** — Beijing, China
  Towards Diverse Lip Reading Representations — *Apr. 2021 – July 2022*
  Research Intern, mentored by Yichen Gong
  - Improved the diversity of lip reading representations by using attention mask to maintain and incorporate contextual information. Solved the over-smoothing problem of Transformer in word-level lip reading.
  - The proposed method achieved new state-of-the-art performances on Lip Reading in the Wild (LRW) in both audio-only, visual-only, and audio-visual settings.
- **John Hopcroft Center for Computer Science, Shanghai Jiao Tong University** — Shanghai, China
  Leveraging Unimodal Self-Supervised Learning for Multimodal AVSR — *Apr. – Sept. 2021*
  Research Intern, advised by Prof. Zhouhan Lin
  - Employed audio and visual self-supervised large-scale pre-training to improve audio-visual speech recognition, achieved a word error rate (WER) of 2.6% on Lip Reading Sentences 2 (LRS2), raising the state-of-the-art performances with a relative improvement of 30%
  - The proposed audio-only and visual-only models gained significant improvement and reached a WER of 2.7% and 43.2%, respectively. Models' noise Robustness also improved greatly due to the extra self-supervised pre-train.
  - Successfully integrate unimodal pre-trained models into a multimodal scenario for the first time, significantly reduced the need of labeled aligned data in the multimodal training process
- **NSF Center for Big Learning, University of Florida** — Gainesville, FL
  Improving Question Answering using EncyclopediaNet — *July – Sept. 2020*
  Research Intern, advised by Prof. Dapeng Oliver Wu
  - Constructed EncyclopediaNet using facts as nodes and multi-hop if-then reasoning as edges
  - Extracted the 5W1H information of simple sentences using a BERT-based semantic role labeling model to structure the nodes, structured information can be utilized to better match the questions and nodes

Service

- **Technical Consultant at Horizon Robotics (End-to-End AVSR Algorithm)** — *July 2022 – Present*