Xichen Pan

xichenpan.cn

E-mail : Flash321@sjtu.edu.cn
Mobile : +86 186 535 00448

Education

- **Shanghai Jiao Tong University** — Shanghai, China
  Bachelor of Engineering in Computer Science and Technology — *Sept. 2018 – June 2022 (expected)*
  Overall: 88.26/100, **Major:** 91.19/100

Research Intersects

Speech (Recognition), Multimodal Deep Learning (Audio-Visual), Self-Supervised Learning, Question Answering

Manuscripts

- **Xichen Pan**, Peiyu Chen, Yichen Gong, Helong Zhou, Xinbing Wang, and Zhouhan Lin. "Leveraging Uni-Modal Self-Supervised Learning for Multimodal Audio-visual Speech Recognition", ***ACL*2022 Main Conference** [pdf]
- **Xichen Pan**, Yichen Gong and Zhouhan Lin. "Multi-task Learning for Word-level Lip Reading", **MM** in preparation

Experience

- **John Hopcroft Center for Computer Science, Shanghai Jiao Tong University** — Shanghai, China
  Audio-Visual Speech Recognition (AVSR) Open-Source Toolkit — *Feb. 2022 – Present*
  Research & Development Intern, advised by Prof. Zhouhan Lin
  - Designed and implemented an AVSR toolkit involving datasets, data pipelines, NN modules and evaluation pipelines
  - Developed a hybrid connectionist temporal classification (CTC)/attention batch decoder
- **Horizon Robotics** — Beijing, China
  Multi-task Learning for Word-level Lip Reading — *Apr. 2021 – Present*
  Machine Learning Engineer Intern, mentored by Yichen Gong
  - Improved visual word classification by training along with CTC loss for visual speech recognition, achieved an accuracy of 89.1% on the Lip Reading in the Wild (LRW) dataset, with a 0.7% improvement over the prior art
  - Improved the short sequence classification performance of Transformer by using word boundaries as attention mask
- **John Hopcroft Center for Computer Science, Shanghai Jiao Tong University** — Shanghai, China
  Leveraging Uni-Modal 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 the Lip Reading Sentences 2 (LRS2) dataset, 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.
  - Accomplished the first work to successfully integrate uni-modal pre-trained models into a multimodal scenario, 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

Selected Projects

- **Curriculum Learning for Sparse Drug-Target Interaction Mining**: Introduced curriculum learning to solve sparse drug-target interaction prediction problem by reducing the amount of positive samples in each iteration; compared to baseline, achieved an absolute F1 score improvement of 8.8% on the highly unbalanced test set
- **MIPS-like Multi-cycle Pipeline Processor**: Designed a MIPS-like multi-cycle pipeline processor using Verilog that can execute 31 instructions and supports stall, forwarding, interrupt, and exception
- **Xconey**: Developed an IIoT software that can extract the main information from G-code and communicate with CNC machine tools, improved productive efficiency in the workshops of several colleges and universities

Selected Awards

- **Academic Excellence Scholarship (*2019, 2020, 2021*)**

Skills

- **Programming Languages**: C/C++, Python
- **Packages**: PyTorch, PyTorch Lightning, fairseq, transformers, OpenCV, h5py, NumPy, pandas, PyQt5