Xichen Pan

xichenpan.com

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

Education

- **Shanghai Jiao Tong University (SJTU)** Shanghai, China
  Bachelor of Engineering in Computer Science (**Outstanding Graduate**) *Sept. 2018 – June 2022*
  Advised by Prof. Zhouhan Lin
  Overall: 88.42/100, Major: 91.29/100

Research Intersects

- **Multimodal Learning**: Multimodal understanding and generation with a focus on vision, language, and audio modalities, especially representation learning and self-supervised pre-training

Publications & Manuscripts

- Xichen Pan, Pengda Qin, Yuhong Li, Hui Xue, and Wenhu Chen. **Synthesizing Coherent Story with Auto-Regressive Latent Diffusion Models**, *CVPR 2023 Under Review* [pdf]
- Xichen Pan, Zekai Li, Yichen Gong, Xinbing Wang, and Zhouhan Lin. **Towards Diverse Lip Reading Representations**, *ICASSP 2023 Under Review* [pdf]
- 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

- **Microsoft Research** Beijing, China
  Multimodal GPT *Nov. 2022 – Present*
  StarBridge Research Assistant, mentored by Li Dong
  - Participated in developing multimodal GPT, a decoding model that can understand vision-language input and generate interleaved natural language responses and illustrations.
- **Alibaba Group** Beijing, China
  Synthesizing Coherent Story with Auto-Regressive Latent Diffusion Models *Sept. – Nov. 2022*
  Research Intern, mentored by Pengda Qin
  - Proposed a history-aware auto-regressive conditioned latent diffusion model named AR-LDM, which first successfully leverages diffusion models for story visualization and continuation with relative FID score improvements of 70% and 20% over previous SoTA, respectively.
  - Introduced the VIST dataset and showed AR-LDM is capable of real-world visual story synthesis.
  - Proposed a simple but efficient adaptation method, allowing AR-LDM to generalize to unseen characters.
- **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 an attention mask to maintain and incorporate contextual information. The proposed method alleviated the over-smoothing problem of Transformer in word-level lip reading, achieving new SoTA audio-visual speech recognition performance on the Lip Reading in the Wild (LRW) dataset.

- **John Hopcroft Center for Computer Science, SJTU** Shanghai, China
  Leveraging Unimodal Self-Supervised Learning for Multimodal AVSR *Apr. – Sept. 2021*

  Research Intern, advised by Prof. Zhouhan Lin

  - Successfully leveraged unimodal self-supervised pre-training for multimodal audio-visual speech recognition for the first time, achieved a word error rate (WER) of 2.6% on the Lip Reading Sentences 2 (LRS2) dataset, raising the SoTA performances with a relative improvement of 30%. The proposed audio-only and visual-only models also reached WERs of 2.7% and 43.2%, respectively.
  - Significantly improved models' noise robustness, as well as reduced the need for labeled aligned data through the use of self-supervised pre-training.

- **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 (who, what, when, where, why, how) information of simple sentences using a BERT-based semantic role labeling model to structure the nodes, which can be utilized to better represent the questions and nodes

- **Data Driven Software Technology Lab, SJTU** Shanghai, China
  An AI-based Approach to Check Coding Style *July – Dec. 2019*

  Research Intern, advised by Prof. Yuting Chen

  - Developed a coding style automatic generator using Pylint and PyQt5 that can generate personalized Python coding style configuration files based on existing code, reducing the need for manually setting up code style checking rules.

Selected Projects

- **Chinese Stable Diffusion**: A Chinese version Stable Diffusion finetuned on the Zero-Corpus dataset, which contains 23 million text-image pairs. A finetuned RoBERTa-base CLIP text model is utilized for Chinese language encoding.
- **Computer Science Masters Application 752 ★**: A webpage built for choosing computer science master programs in north america. This page is powered by Material for MkDocs and supports collaboration through pull requests and GitHub actions.

Media Exposures

- **Synthesizing Coherent Story with Auto-Regressive Latent Diffusion Models**, Synced

Selected Awards

- **Best Thesis Award (*2022*)**
- **Academic Excellence Scholarship (*2019, 2020, 2021*)**

Skills

- **Programming Languages**: C/C++, Python
- **Packages**: PyTorch, Lightning, Transformers, Diffusers, fairseq, WandB, Hydra, OpenCV, h5py, NumPy, PyQt5