I-CHUN CHEN

iotbwu0101@gmail.com
https://github.com/wazenmai
https://blog.wazenmai.com

My research interests broadly lie in the field of image/text/music AI and model compression. Recently I'm working on Large Language Model (LLM) compression, specifically on Mixture-of-Experts (MoE) architecture.

EDUCATION

National Tsing Hua University *September 2022 - July 2024*
Master's degree. Department of Computer Science.
- **Overall GPA**: 4.23/4.3

Technische Universität Dresden, Germany *March 2022 - August 2022*
Exchange Student of Computer Science.

National Tsing Hua University *September 2018 - August 2022*
Bachelor's degree. Department of Computer Science.
- **Overall GPA**: 3.93/4.3

PUBLICATIONS

Retraining-Free Merging of Sparse Mixture-of-Experts via Hierarchical Clustering
https://arxiv.org/abs/2410.08589
https://github.com/wazenmai/HC-SMoE
- Submitted to The Thirteenth International Conference on Learning Representations (ICLR 2025) as first author.
- Reducing parameter of experts in MoE in retraining-free, task-agnostic manner.
- Utilize hierarchical clustering to group similar experts with a merging process within ten minutes.
- Evaluate Mixtral 8x7B and Qwen1.5-MoE on eight language benchmarks.

MidiBERT-Piano: Large-scale Pre-training for Symbolic Music Understanding
https://arxiv.org/abs/2107.05223
https://github.com/wazenmai/MIDI-BERT
- Accepted by Journal of Creative Music Systems (JCMS), 2024.
- Apply BERT into general musical tasks such as melody completion, composer classification and emotion classification using Pytorch.
- Research for two kinds of music representation, REMI and CP and study the performance difference between those results.

WORK EXPERIENCE

Cloud Communication Technology LTD. *February 2022 - August 2023*
iOS and Android App Engineer
- Implement LINE message API and login service in Community Web Flutter app.
- Implement version detection, encrypted qrcode scanner, Google Cloud Storage Image upload and deeplink function in TABF Exam Camera Flutter app.
- Embed Google face detector in camera so we can detect the face angle and position in real-time for ID Photo camera Flutter app.
- Design and Implement backend API and database CRUD operations for ID Photo camera with typescript , Koa and PostgreSQL.

Research Center for Information Technology Innovation, Academia Sinica *July 2020 - August 2020*
Music & AI Lab Summer Intern
- Learn symbolic music format representation and encoding.
- Present paper in NLP transformer family, including Transformer, BERT, RoBERTa, ELECTRA, GPT-2, and T5.
- Modify "Pop Music Transformer" from single-track piano music generation to multi-track pop music generation, the model can either generate songs by itself or continuation after human-given music.

PROJECTS

2022 AI-CUP Explanation Information Tagging of Natural Language Understanding Competition
https://github.com/wazenmai/DataMining-Team5
- Given the talk between two people and one of them's attitude (agree or disagree), we need to find the important statement among the sentences they said to support their main point and attitude.
- View problem as token classification task and fine-tune on BERT, DistillBERT and RoBERTa, also as question answering task and fine-tune on BERT.
- Got score 0.835978 (1 is best and 0 is worst) from token classification RoBERTa with **30th place** in private leader board.

NTHU OAuth Decaptcha
https://chrome.google.com/webstore/detail/nthu-oauth-decaptcha/mflpajkffpiibelpmffonolenndbgogp
https://github.com/wazenmai/NTHU-OAuth-Decaptcha
- Train a CNN model to recognize the number on the captcha image, get 98.9% accuracy on testing set.
- Collect training data by ourselves by posting request with captcha id then we can easily get lots of captcha images.
- There are 356 users using our decaptcha extension.

Parallel Pitch Tracking Algorithm
https://github.com/wazenmai/Parallel-Pitch-Tracking
- Implement ACF (Auto-Correlated Funciton) with MPI and CUDA to recognize pitch in the wav file.
- About 100x speedup with the largest testing data by CUDA optimization.

Massive Data Analysis
https://github.com/wazenmai/Massive-Data-Analysis
- Learn topics in massive data analysis, including MapReduce, similarity search, frequent-itemset mining, managing advertising and recommendation systems, etc.
- Using MapReduce in Python Spark on matrix multiplication, pagerank, kemans, and finding similarity articles.

TECHNICAL SKILLS

Programming:	C, C++, Python, Dart, Javascript, CUDA
Software & Tools:	**ML-Related Framework:** Pytorch, Huggingface-transformers
	Frontend Framework: Flutter, VueJS
	Cloud Service: GCP(Vision API, Cloud Storage), Firebase
	Database: PostgreSQL, MongoDB

EXTRACURRICULAR

Machine Learning Study Group *November 2019 - August 2021*
- Watch and discuss ML-related videos every week, as well as implement algorithms and read papers.
- Including course of Machine Learning on Coursera by Andrew Ng, CS231n: Deep Learning of Computer Vision at Stanford University, Reinforcement Learning course by DeepMind David Silver, and CS294-112: Deep Reinforcement Learning at UC Berkely by Sergey Levine.

LeetCode Study Group - Convener *February 2021 - June 2021*
- Form a 16 students group and invite lecturer to explain 6 problems every 2 weeks.
- Practice Leetcode problems by topics, including dynamic programming, binary search, stack, etc.
- Got **the Excellent Award** of NTHU 2021 winter semester study group.
https://justin0u0.notion.site/LeetCode-57d3fc3c39714bf7b37429a590a85824

2020 Meichu Hackathon *October 2020*
- Design an app that can help people with different perspective to know each other more and reduce cognitive estrangement and hatred between different groups.
- Randomly match people that has different perspective to one thing to chat, design question for them to talk.
- Present various news that people can comment beyond the article and show their attitude, we would separate and compare different attitude comments to users.
- Got **rank 3** in social enterprise group.