Mr. Zihan Zhang

Mobile: +86 18954311726; +1 6672038885| Email: zzhan343@jh.edu

Add: 3501 St. Paul Street, Baltimore, MD, USA

EDUCATION

Master's Program | **ECE Department | Johns Hopkins University,** Baltimore, USA *Aug. 2024 - May. 2026*

- **Major:** Electrical and Computer Engineering
- **GPA:** 4.0/4.0
- **Core Courses:** Machine Learning for Medical Application, Machine Perception, Audio Signal Processing, Machine Learning for Signal Processing, Applied Mathematics, etc.

Bachelor's Degree | **Aulin College | Northeast Forestry University,** Harbin, China *Sep. 2020 - Jun. 2024*

- **Major:** Computer Science and Technology (Sino-foreign cooperative education program)
- **GPA:** 3.45/5.0 (86.15/100)
- **Core Courses:** Algorithms and data structures, Programming techniques (JAVA), Python, Mathematics for Computer Science, Linux System and Shell Programming, Algorithms and Data Structures, Machine Learning, Software Engineering, C++ (Coursera), etc.

PUBLICATION & PATENTS

- **Paper:** ***"Filter, Correlate, Compress: Training-Free Token Reduction for MLLM Acceleration"***, as the third author, has been submitted to ***ICCV25,*** *03/2025*
- **Paper:** ***A Machine Learning Algorithm for Insect Identification****, a*s the first author, was included in the International Conference on Software Engineering and Machine Learning (CONF-SEML 2023), *04/2023*
- **Patent:** ***Small Sample Forestry Pest Identification Based on Deep Learning****, 03/2023*

RESEARCH PROJECTS

FiCoCo – Training-Free Token Reduction for MLLM Acceleration *sep.2024 – Mar. 2025*

Advisor: Dr. Siteng Huang, Alibaba DAMO Academy

- *Developed a Filter-Correlate-Compress framew*ork to remove redundant visual tokens in multimodal large language models. Relative Paper has already been submitted to ***ICCV 25***.
- Designed two variants: *FiCoCo-V* for vision encoders and *FiCoCo-L* for LLM decoders, utilizing dynamic multi-to-multi token correlation and weighted compression.
- Achieved up to 5.7×/14.7× FLOPs reduction with over 92% performance retention on LLaVA-1.5-7B and LLaVA-NeXT-7B, while improving throughput by up to 2.08×.

VAR2Vid – Training-Free Extension of VAR-Based Video Generation Model *sep.2024 – Now*

Advisor: Dr. Siteng Huang, Alibaba DAMO Academy

- Based on VAR image generation models, extend them to video generation models in a training-free manner.
- Inspired by Text2Video-Zero and FateZero, integrated temporal attention modules into the image generation pipeline to ensure fine detail discrimination at low scale and maintain temporal consistency at high scale.

Epilepsy Prediction Algorithms Based on Deep Learning *sep.2024 – Dec. 2024*

Advisor: Associate Prof. Najim Dehak, Johns Hopkins University

- introduces a hybrid model that combines CapsNet for spatial features with a bidirectional LSTM for temporal patterns, improving prediction accuracy and stability for clinical early warnings.
- Based on CHB-MIT EEG data, Using CapsNet extracts spatial features while the bidirectional LSTM captures temporal dynamics. The model achieved an accuracy of 96.03%, an AUC of 0.9922, an F1-score of 0.9605, and a recall of 0.9639 in five-fold cross-validation.

Generating Medical Reports from Medical Imaging Data Using LLMs *Feb.2025 –Now*

Advisor: Assist Research Professor. Laureano Moro Velazquez, Johns Hopkins University

- Integrates CNN/ViT with a large language model to automatically convert medical imaging data into structured reports, significantly reducing report generation time while maintaining high accuracy
- Through domain-specific model fine-tuning and an efficient data preprocessing pipeline, the system lays a solid foundation for future multimodal data integration and model interpretability improvements

Research on Small Sample Forestry Pest Identification Based on Deep Learning

Advisor: Associate Prof. Ji Mingyu *May 2022 - June 2023*

- Aimed to tackle the forestry pest issue, which poses a substantial threat to ecosystems. Designed a machine learning algorithm for efficient pest identification, overcoming the limitations of human visual recognition.
- Leveraged a pre-trained ResNet-18 neural network on the ImageNet dataset as the backbone for the prototypical network, enhancing image recognition capabilities.

INTERNSHIP EXPERIENCES

iFlytek Co. Ltd. **Shenzhen, China**

Position: Intern at AIoT Business Department *Jul. 3, 2023-Oct. 3, 2023*

- Utilized the pytest framework and .sh scripts to implement automated testing across multiple AIoT products;
- developed a posture detection algorithm based on YOLO v8 for Smart desk and chair Product.

AWARDS & HONORS

- ***National Excellent Project Conclusion Award***, National Innovation Program for College Students, June 2023
- ***First-Class Scholarship***, University Level, 2023***(5%)***
- ***Second-Class Scholarship***, University Level, 2022 & 2021***(10%)***