Boyu Guan

email: guanboyu2022@ia.ac.cn
Phone (WeChat) : 18640473542
Personal Website: https://boyuguan.github.io · Location: Beijing

EDUCATION

Institute of Automation, CAS, Beijing, *PhD* Sep. 2022 –Present

- Supervised by Prof. Chengqing Zong; expected graduation: June 2027
- Research: NLP, Multimodal LLMs, Video-Guided Machine Translation
- GPA: 3.6 / 4; completed core CS courses and CSAPP with labs

Northeastern University, B.S. in Mathematics, Shenyang, China Sep. 2018 –Jun. 2022

- GPA: 3.76 / 5.00, top 3/31; admitted to Ph.D. program via recommendation
- Received multiple honors including the Second-Class Scholarship for Outstanding Students and the Second-Class Scholarship from Northwest Institute for Nonferrous Metal Research

Academic Research

Key Technologies in Large-Scale Multilingual Multimodal Neural Machine Translation, NSFC Key Project 2024.01 - Present

- **Main Research Direction**: Focused on video-guided multimodal machine translation, exploring how multimodal large language models can address critical issues in the field, such as data scarcity, high algorithmic cost, and unstable performance.
- **Publications**:
  - *TriFine: A Large-Scale Dataset of Vision-Audio-Subtitle for Tri-Modal Machine Translation and Benchmark with Fine-Grained Annotated Tags*
    - First Author, **COLING 2025**, CCF B, Oral Presentation (9.8%).
    - Proposed the first large-scale tri-modal dataset TriFine, covering vision, audio, and subtitles, with seven categories of fine-grained tags.
    - Introduced a model-agnostic method FIAT based on fine-grained multimodal enhanced inputs, significantly improving translation performance and efficiency.
  - *SHIFT: Selected Helpful Informative Frame for Video-guided Machine Translation*
    - First Author, **EMNLP 2025 Main Conference**, CCF B, Top-tier conference in NLP.
    - Proposed a lightweight and pluggable modality-adaptive framework SHIFT, which adaptively selects key frames or pure text for each video-text sample through clustering and selection modules.
    - Improved translation quality while greatly reducing multimodal computation cost and increasing throughput.
- **Patent Applications**:
  - Patent filed: *A Video Machine Translation Method and Device Integrating Fine-Grained Multimodal Information*, Second Inventor (Advisor First), currently under substantive examination.
  - Patent filed: *An Adaptive Key Frame Selection Method for Video Machine Translation*, Second Inventor (Advisor First), currently under substantive examination.

Internship Experience

Biren Technology, Software Engineering Intern, Beijing Feb. 2023 –Aug. 2023

- Optimized pre-training and inference pipelines for LLaMA, LLaMA2, and ChatGLM; responsible for operator extraction, migration, and adaptation
- Focused on activation checkpointing and 3D parallelism; customized frameworks like DeepSpeed, Megatron-DeepSpeed, and Transformers for performance tuning