# 黄文轩

✉ osilly0616@gmail.com · ☎ +86-153-9960-7206 · https://github.com/Osilly

Google Scholar: `https://scholar.google.com/citations?user=6Ys6HgsAAAAJ`

## 教育背景

**三峡大学** 2019.09-2023.06
计算机与信息学院 本科

**华东师范大学** 2023.09-至今
计算机科学与技术学院 硕士（研二）

## 研究领域

- **Reasoning MLLM**
- **MLLM**
- Model Acceleration
- AI4Geophysic

## 学术论文（（共同）第一作者）

### Current Research Interesting

- **[Reasoning MLLM] Vision-R1: Incentivizing Reasoning Capability in Multimodal Large Language Models.**
  - Submission to **ICCV 2025**, first author.
  - This is the first paper to explore how to effectively use RL for MLLMs and introduce Vision-R1, a reasoning MLLM that leverages cold-start initialization and RL training to incentivize reasoning capability.
  - arXiv: `https://arxiv.org/abs/2503.06749`
  - [Star 500+] Github repo: `https://github.com/Osilly/Vision-R1`

### Accepted Paper

- **[Efficient MLLM] Dynamic-LLaVA: Efficient Multimodal Large Language Models via Dynamic Vision-language Context Sparsification.**
  - Accepted to **ICLR 2025**, first author.
  - Dynamic-LLaVA is the first MLLM acceleration framework that simultaneously sparsifies both vision and language contexts while integrating inference efficiency optimization across different MLLM inference modes into a unified framework.
  - arXiv: `https://arxiv.org/abs/2412.00876`
  - Github repo: `https://github.com/Osilly/dynamic_llava`
- **[Transformer Training Acceleration] A General and Efficient Training for Transformer via Token Expansion.**
  - Accepted to **CVPR 2024**, first author.
  - We proposed one plug-and-play Transformer training acceleration framework, without twisting the original training hyper-parameters, architecture, and introducing additional training strategies.
  - arXiv: `https://arxiv.org/abs/2404.00672`
  - Github repo: `https://github.com/Osilly/TokenExpansion`
- **[AI4Geophysic] An Intelligent First Arrival Picking Method of Microseismic Signals Based on the Small Sample Expansion.**
  - Accepted to **IEEE Transactions on Geoscience and Remote Sensing (TGRS)**, first author.
  - We proposed one GAN to generation the microseismic samples under unsupervised conditions to expand the microseismic data having a limited number of samples. Then we use the enhanced first arrival picking network to improve the accuracy of first arrivals of low SNR microseismic signals.
  - Paper link: `https://ieeexplore.ieee.org/abstract/document/10972295`
  - Github repo: `https://github.com/Osilly/G-LA-MSG-and-AOG-PSPNet`

### Under-review Paper

- **[CNN Inference Acceleration] An Filter Pruning for Efficient CNNs via Knowledge-driven Differential Filter Sampler.**
  - Submission to **IJCV** (major revision), student first author (first author is my advisor).
  - We proposed a unified CNN pruning framework directly optimized in an end-to-end manner in combination with global pruning constraint.
  - arXiv link: `https://arxiv.org/abs/2307.00198`
  - Github repo: `https://github.com/0silly/KDFS`
- **[Transformer Training Acceleration (journal version of ToE)] Feature Sparsification Training Paradigm: Toward Fast and Memory-Efficient General Transformer Training.**
  - Submission to **TPAMI**, first author.
  - arXiv: `https://arxiv.org/abs/2404.00672`
  - Github repo: `https://github.com/0silly/TokenExpansion`
- **[MLLM] LLaVA-RadZ: Can Multimodal Large Language Models Effectively Tackle Zero-shot Radiology Recognition?**
  - Submission to **ICCV 2025**, co-first author (second).
  - *Label in paper: **Wenxuan Huang** proposed the main idea and designed the experiments, contributing to the discussion of this paper. **Bangyan Li** refined and finalized the idea, implemented the code and experiments, and was responsible for writing the manuscript.*
  - Solving the problem of that MLLM cannot effectively tackle zero-shot radiology recognition.
  - arXiv: `https://arxiv.org/abs/2503.07487`
- **[MLLM] TimeSoccer: An End-to-End Multimodal Large Language Model for Soccer Commentary Generation.**
  - Submission to **ACMMM 2025**, co-first author (second).
  - We propose the first end-to-end MLLM for soccer commentary generation, specifically designed for Single-anchor Dense Video Captioning (SDVC) in full-match soccer videos. The model jointly predicts timestamps and generates captions in a single pass, enabling global context modeling over 45-minute matches.
  - arXiv: `https://arxiv.org/abs/2503.07487`
- **[Image Editing Benchmark] Comp-Edit: Benchmarking Complex Instruction-guided Image Editing.**
  - Submission to **NeurIPS 2025**, co-first author (second).
  - We propose one complex image editing benchmark.
- **[CLIP Inference Acceleration] CLIP-Map: Structured Matrix Adaptation for Parameter-Efficient CLIP Compression.**
  - Submission to **NeurIPS 2025**, co-first author (second).
  - We propose a framework that maps the parameters of CLIP to a smaller representation, thereby accelerating inference.

## 成绩与荣誉

- 本科专业排名第一（1/56）
- 2020-2021 学年国家奖学金

## 算法竞赛

- 2020 年蓝桥杯 B 组 C/C++ 程序设计大赛国家二等奖
- 2021 年蓝桥杯 B 组 C/C++ 程序设计大赛国家二等奖
- 2021 年中国高校计算机大赛-团体程序设计天梯赛个人国家二等奖
- 2022 年中国高校计算机大赛-团体程序设计天梯赛团队国家二等奖