Hao Yin

yinhnavi@mail.ustc.edu.cn | ustc-hyin | Google Scholar | Homepage | +86 18051050608

Education

University of Science and Technology of China, M.Sc. in Data Science — Sep 2022 – Jun 2026
School of Artificial Intelligence and Data Science GPA: 3.96/4.30 (2/31)

China University of Mining and Technology, B.Sc. in Applied Mathematics — Sep 2018 – Jun 2022
School of Mathematics GPA: 4.47/5.00 (2/185)

Selected Publications

*Publications are listed in reverse chronological order. * denotes equal contribution.*

Video-OPD: Efficient Post-Training of Multimodal Large Language Models for Temporal Video Grounding via On-Policy Distillation **ICML 2026 (Under Review)**

- **Authors:** Jiaze Li*, **Hao Yin***, Haoran Xu*, Boshen Xu, Jian Luan, *et al.*
- **Paper:** arxiv.org/abs/2602.02994
- **Summery:** We propose Video-OPD, an efficient post-training framework for temporal video grounding that maintains the on-policy property to mitigate distributional shift and converts sparse, episode-level feedback into fine-grained, step-wise learning signals, achieving faster convergence and outperforming GRPO methods.

REVISOR: Beyond Textual Reflection, Towards Multimodal Introspective Reasoning in Long-Form Video Understanding **CVPR 2026 (Poster)**

- **Authors:** Jiaze Li*, **Hao Yin***, Wenhui Tan*, Jingyang Chen*, Jian Luan, *et al.*
- **Paper:** arxiv.org/abs/2511.13026
- **Summery:** We propose REVISOR, a tool-augmented multimodal reflection framework that enables MLLMs to perform introspective reasoning across both visual and textual modalities in long-form videos, aligning reasoning with video evidence via a dual-attribution reward mechanism, achieving state-of-the-art performance across multiple benchmarks.

The Mirage of Performance Gains: Why Contrastive Decoding Fails to Address Multimodal Hallucination **NeurIPS 2025 (Poster)**

- **Authors:** **Hao Yin**, Guangzong Si, Zilei Wang
- **Paper:** arxiv.org/abs/2504.10020
- **Code:** github.com/ustc-hyin/cd_rethink
- **Summery:** We argue that **contrastive decoding does not mitigate the hallucination problem**. The observed performance gains are largely driven by two misleading factors: (1) crude, unidirectional adjustments to the output distribution, and (2) the inherent adaptive plausibility constraint, which collapses sampling strategy into greedy search.

Lifting the Veil on Visual Information Flow in MLLMs: Unlocking Pathways to Faster Inference **CVPR 2025 (Poster)**

- **Authors:** **Hao Yin**, Guangzong Si, Zilei Wang
- **Paper:** arxiv.org/abs/2503.13108
- **Code:** github.com/ustc-hyin/HiMAP
- **Summery:** We uncover **a shift in the dominant flow of visual information within MLLMs**: (1) in the shallow layers, strong interactions occur between image tokens and instruction tokens, with most visual information being infused into the instruction tokens to construct cross-modal semantic representations; (2) in the deeper layers, image tokens primarily interact among themselves, consolidating the remaining visual information to refine semantic representations within the visual modality.

ClearSight: Visual Signal Enhancement for Object Hallucination Mitigation in Multimodal Large Language Models **CVPR 2025 (Poster)**

- **Authors:** **Hao Yin**, Guangzong Si, Zilei Wang
- **Paper:** arxiv.org/abs/2503.13107
- **Code:** github.com/ustc-hyin/ClearSight
- **Summery:** Hallucination mitigation via contrastive decoding faces two main limitations: (1) indiscriminate suppression of language priors can degrade the coherence and accuracy of generated content, and (2) processing contrastive inputs adds computational overhead, slowing inference. We introduce Visual Amplification Fusion (VAF), a plug-and-play technique that enhances attention to visual signals in the model's middle layers, where modality fusion predominantly occurs, effectively addressing both challenges.

Research Experience

Researcher (Master's Program), USTC – Hefei, Anhui, China Jan 2024 – Jun 2025

Research on improving inference efficiency and reducing hallucinations in MLLMs.

- Proposed a hierarchical modality-aware pruning strategy, motivated by the observation that the dominant visual information flow shifts with model depth, leading to substantial improvements in inference speed.
- Addressed object hallucination by strengthening visual attention in critical layers, effectively circumventing the inherent drawbacks of contrastive decoding, such as degradation in output quality and inference latency.
- Identified and experimentally verified that contrastive decoding does not mitigate hallucinations; observed performance gains on benchmarks like POPE are largely driven by misleading factors.
- Core findings published in three first-author papers, with two accepted at CVPR 2025 and one at NeurIPS 2026.

Research Intern, Tencent Technology – Beijing, China Sep 2025 – Dec 2025

Research on enhancing the image captioning capabilities of MLLMs through reinforcement learning strategies.

- Developed a compact MLLM using supervised fine-tuning and reinforcement learning to generate highly precise, context-aware, and structured image captions, advancing real-world visual understanding.
- Proposed a co-evolutionary adversarial framework to enhance MLLM image captioning by jointly optimizing a captioning model and a question-generation model within a self-reinforcing feedback loop, driving progressive improvements in descriptive fidelity and informational completeness.

Research Intern, Xiaomi Technology – Beijing, China Jan 2026 – Apr 2026

Research on enhancing reasoning capabilities of video foundation models through post-training strategies.

- Proposed a tool-augmented MLLM reasoning framework that enables introspective reasoning across both visual and textual modalities, significantly improving long-form video understanding.
- Developed Video-OPD, an efficient post-training framework for temporal video grounding that converts sparse rewards into dense step-wise signals, accelerating convergence while outperforming existing GRPO methods.
- Research outcomes led to two papers, one accepted at CVPR 2026 and the other under review at ICML 2026.

Honors and Awards

National Scholarship, China University of Mining and Technology Oct 2019

- Ranked 1st out of 160 undergraduate students (Top 0.6%)

National Scholarship, China University of Mining and Technology Oct 2020

- Ranked 1st out of 185 undergraduate students (Top 0.5%)

Silver Medalist, Kaggle Competition – Cassava Leaf Disease Classification Feb 2021

- Placed 192nd out of 3,900 teams (Top 4.9%)

Silver Medalist, Kaggle Competition – Jane Street Market Prediction Aug 2021

- Placed 155th out of 4,245 teams (Top 3.6%)