Fangyuan Mao

Computer Science, Institute of Computing Technology, Chinese Academy of Sciences

fangyuanmaocs@gmail.com | Website | Github | Scholar

Research Interests

My research focuses on **Multimodal Generative Models**. I am interested in building intelligent and efficient AI systems that bridge diverse modalities and enable downstream tasks. My prior work spans video[1,6,9,10] and image[2,8] generation, cross-modal perception for autonomous driving[3,5], and efficient model design[4,7].

Education

Institute of Computing Technology, Chinese Academy of Sciences — Beijng, China
M.Phil. in Computer Science, GPA: 3.84/4.0, Advisor: Prof. Yu Hu — *09/2023 - Present*

Zhejiang University — Hangzhou, China
B.Sc. in Geophysics, GPA: 3.93/4.0, Ranking 1st, Advisor: Prof. Yixian Xu — *09/2019 - 06/2023*

Publications & Under Review Manuscripts

[1] Mao, F.†, Hao, A.†, et al. *Omni-Effects*: Unified and Spatially-Controllable Visual Effects Generation. **AAAI 2026**. [*Paper*][*Project*][*Code*, 160+ stars]

[2] Mao, F., Mei, J., et al. PID: Physics-Informed Diffusion Model for Infrared Image Generation. **Pattern Recognition** (JCR Q1, IF=7.9).[*Paper*][*Code*, 140+stars]

[3] Mao, F.†, Wang, S.†, et al. UNIV: Unified Foundation Model for Infrared and Visible Modalities. **CVPR 2026**. Under Review. [*Paper*][*Code*]

[4] Hu, K., Gao, J., **Mao, F.**, et al. Disassembling Convolutional Segmentation Network. **International Jounral of Computer Vision** (JCR Q1, IF=15.5).[*Paper*]

[5] Liu, F., Mei, J., **Mao, F.**, et al. CoreNet: Cross-Modal 4D Radar Denoising Network with LiDAR Supervision for Autonomous Driving. **IROS 2025**. [*Paper*][*Code*]

[6] Wu, M., Zhu, J., Feng, X., Chen, C., Zhu, C., Song, B., **Mao, F.**, et al. ImagerySearch: Adaptive Test-Time Search for Video Generation Beyond Semantic Dependency. **AAAI 2026**. [*Paper*] [*Code*]

[7] Lu, S., **Mao, F.**, et al. SEA: Hierarchically Searching Efficient Adapters for Pre-trained Models. **Neural Networks**. Under Review.

[8] Min, C., Mei, J., Kong, F., **Mao, F.**, et al. GenDet: Colored Bounding Box Generation via Diffusion Models for Object Detection. **IJCV** Under Review.

[9] Feng, X., Zhu, J., Wu, M., Chen, C., **Mao, F.**, et al. MiGA: Make Train-Free Infinite Frame Generation Great Again for Consistent Long Videos. **ICLR 2026**. Under Review.

[10] Chen, C., Zhu, J., Feng, X., Huang, N., Wu, M., **Mao, F.**, et al. S^2-Guidance: Stochastic Self Guidance for Training-Free Enhancement of Diffusion Models. **ICLR 2026**. Under Review.[*Paper*]

Research Experience

Research Intern, Kuaishou Technology, Kling Team — *12/2025 - Present*

Hosts: Shun Lu

1. Video Generation Acceleration: Explore **Distribution-Matching Distillation (DMD)** in video generation acceleration and efficient **Pyramid-based generation** methods.

Research Intern, AMAP Alibaba Group, Machine Learning Team *04/2025 - 10/2025*

Hosts: Xiangxiang Chu, Jiahong Wu

1. *Omni-Effects.* Contributed to the development of a unified framework for **Visual Effects (VFX) video synthesis**, enabling prompt-guided and spatially controllable composite VFX. Designed and implemented **LoRA-based Mixture of Experts (LoRA-MoE)** for optimizing multi-VFX training. Developed **Spatial-Aware Prompting (SAP)** to embed spatial information into text tokens, enhancing VFX localization precision. Introduced **Independent Information Flow (IIF)** to isolate control signals and prevent unintended blending. Curated the most comprehensive VFX dataset **Omni-VFX** and designed an evaluation framework. *Project Page*

Highlights: HuggingFace Daily Paper, 160+ stars on Github, accepted by *AAAI 2026*.

Institute of Computing Technology, Autonomous Navigation Group *09/2023 - Present*

Advisors: Yu Hu, Jilin Mei

Research focuses on **infrared modalities** for autonomous driving, including visible-to-infrared translation, unified pre-training for RGB and IR, and end-to-end perception in off-road environments.

1. PID: Physics-Informed Diffusion Model for Infrared Image Generation. Introduced **physical constraints** for visible-to-infrared translation. The method enforces physical consistency during generation and introduces an efficient infrared decomposition. Achieved state-of-the-art performance. Published in *Pattern Recognition*. *Paper Link*, 140+ stars on GitHub.

2. UNIV: Unified Foundation Model for Infrared and Visible Modalities. UNIV unifies visible and infrared perception by constructing a shared semantic feature space that removes modal bias. It uses a frozen RGB encoder to derive patch-level semantic relations as pseudo labels, which act as anchors for Patch Cross-modal Contrastive Learning. By **pulling together semantically related RGB–IR patches and pushing apart unrelated ones**, the model learns semantics-driven representations with strong cross-modal alignment and inter-class separability.*Paper Link*.

Zhejiang University, Vision Intelligence and Pattern Analysis Lab *03/2022 - 06/2022*

Advisors: Mingli Song, Zunlei Feng

1. Disassembling Convolutional Segmentation Network into category-aware kernels using forward activation and backward gradient attribution, enabling customizable segmentation tasks without retraining while maintaining competitive performance. The project was accepted by *International Journal of Computer Vision*. *Paper Link*

Zhejiang University, Geophysics Lab *01/2023 - 06/2023*

Advisors: Yixian Xu, Bo Yang

1. A Diffusion-Based Method for Two-Stage Recovery of Magnetic Anomaly Data. Pioneered the application of diffusion models to geophysical magnetic field data inpainting. The project was accepted by *Big Data and Earth System*. *Paper Link*

Selected Awards And Fellowships

Yifangda Financial Technology Master's Award(Top 2%)	*2025*
Outstanding graduate of Zhejiang University	*2023*
Zhejiang Provincial Government Scholarship (Top 3%)	*2020, 2021*
Zhejiang University Academic Second Scholarship(Top 8%)	*2020, 2021*
Shufeng Yang Scholarship (Top 3%)	*2020, 2021*

Skills

Language: Mandarin Chinese(native), English(TOEFL: 96)