

RONGYAO FANG

🌐rongyaofang.github.io ✉rongyaofang@gmail.com 📞(+86)158-8850-6776

EDUCATION

The Chinese University of Hong Kong

Sept.2021 - 2025(Exp.)

Ph.D. candidate at MMLab, Department of Electronic Engineering.

Supervisor: Prof. Hongsheng Li and Prof. Xiaogang Wang.

Topics: Multimodal Large Language Model, AIGC, Artificial General Intelligence.

Shanghai Jiao Tong University

Sept.2016 - July 2020

B.Eng., School of Electronic Information and Electrical Engineering.

Major: Information Engineering (Artificial Intelligence track).

Ranking: 1st/157.

Research: Independent researcher under the supervision of Prof. Bingbing Ni.

Massachusetts Institute of Technology

July 2019 - March 2020

Computer Science and Artificial Intelligence Laboratory (CSAIL).

Research: Independent visiting scholar under the supervision of Prof. Dina Katabi .

RESEARCH INTERESTS

My research targets **AGI for visual perception and generation**. I focus on developing integrated systems that **perceive, understand, and generate** visual content through advanced computer vision techniques.

INTERNS

SenseTime

Feb.2024 -

Topics: Advanced multi-modal large language model.

Shanghai AI Laboratory

June 2022 - Apr. 2023

Topics: Representation learning and vision perception.

PROJECTS

Unified MLLM Combining Image Understanding and Generation

Feb.2024

Role: Conceptualization, Experimentation, Implementation, and Writing

• Proposing a unified multimodal large language model framework that integrates multi-granular visual generation and understanding capabilities. It excels at a range of visual tasks such as diverse text-to-image generation, precise image editing, conditional image generation, and multimodal understanding, balancing the trade-off between diversity and controllability in visual generation tasks.

Zero-Shot Scalable Image Synthesis Across Resolutions with FouriScale

Nov.2023

Role: Conceptualization, Experimentation, and Implementation

• FouriScale enabled **zero-shot** scalable high-quality image synthesis **across resolutions and aspect ratios for any stable diffusion model**. It integrates **dilation and low-pass filtering** to maintain structural and scale consistency in these high-resolution images. Pre-trained diffusion models were enhanced with frequency domain analysis for preserving structural integrity across varying resolutions without retraining.

InstructSeq: A Unified Framework for Instruction-Driven Multi-modal Tasks

July 2023

Role: Conceptualization, Experimentation, Implementation, and Writing

• Developed the InstructSeq framework, integrating multimodal transformers for unified vision and language task processing. Engineered key components including visual and text encoders, and an autoregressive transformer for dynamic output generation. Innovated **instruction-based task handling for enhanced flexibility and generalizability** in AI applications without task-specific tuning.

FeatAug-DETR: Enhancing Object Detection with Feature Augmentation

Nov. 2022

Role: Conceptualization, Experimentation, Implementation, and Writing

• Developed FeatAug-DETR, an innovative approach that **augments image feature maps** instead of raw images, **accelerating DETR training and improving detection performance**. Ensured the augmentation methods can be seamlessly integrated with existing DETR models as a **plug-and-play solution**. Provided a versatile technique for enhancing object detection capabilities across various challenging scenarios.

PUBLICATIONS

PUMA: Empowering Unified MLLM with Multi-Granular Visual Generation

Rongyao Fang, Chengqi Duan, Kun Wang, Hao Li, Hao Tian, Xingyu Zeng, Rui Zhao, Jifeng Dai, Hongsheng Li, Xihui Liu. In submission.

FouriScale: A Frequency Perspective on Training-Free High-Resolution Image Synthesis

Linjiang Huang*, **Rongyao Fang***, Aiping Zhang, Guanglu Song, Si Liu, Yu Liu, Hongsheng Li.

European Conference on Computer Vision (**ECCV 2024**) **Accepted** ([Link](#)).

InstructSeq: Unifying Vision Tasks with Instruction-conditioned Multi-modal Sequence Generation

Rongyao Fang, Shilin Yan, Zhaoyang Huang, Jingqiu Zhou, Hao Tian, Jifeng Dai, Hongsheng Li.

In submission to Transactions on Multimedia (**TMM**) ([Link](#)).

FeatAug-DETR: Enriching One-to-Many Matching for DETRs with Feature Augmentation

Rongyao Fang, Peng Gao, Aojun Zhou, Yingjie Cai, Si Liu, Jifeng Dai, Hongsheng Li.

Transactions on Pattern Analysis and Machine Intelligence (**TPAMI**) **Accepted** ([Link](#)).

Tip-Adapter: Training-free CLIP-Adapter for Better Vision-Language Modeling

Renrui Zhang*, **Rongyao Fang***, Peng Gao*, Wei Zhang, Kunchang Li, Jifeng Dai, Yu Qiao, Hongsheng Li.

European Conference on Computer Vision (**ECCV 2022**) ([Link](#)).

Clip-adapter: Better vision-language models with feature adapters

Peng Gao, Shijie Geng, Renrui Zhang, Teli Ma, **Rongyao Fang**, Yongfeng Zhang, Hongsheng Li, Yu Qiao.

International Journal of Computer Vision (**IJCV**) ([Link](#)).

Learning Longterm Representations for Person Re-Identification Using Radio Signals

Lijie Fan*, Tianhong Li*, **Rongyao Fang***, Rumen Hristov, Yuan Yuan, Dina Katabi.

IEEE Conference on Computer Vision and Pattern Recognition (**CVPR 2020**) ([Link](#)).

Probabilistic Radiomics: Ambiguous Diagnosis with Controllable Shape Analysis

Jiancheng Yang*, **Rongyao Fang***, Bingbing Ni, Yamin Li, Yi Xu, Linguo Li.

Medical Image Computing and Computer Assisted Intervention (**MICCAI 2019**) ([Link](#)).

HONORS AND AWARDS

Hong Kong PhD Fellowship

Research Grants Council (RGC) of Hong Kong.

Sept. 2021

Outstanding Graduates of Shanghai

TOP 1%, Shanghai Municipal Education Commission.

July 2020

National Scholarship

TOP 1%, Ministry of Education of P.R.China.

2017 & 2018

Zhiyuan College Honors Scholarship

TOP 5%, Zhiyuan College, Shanghai Jiao Tong University.

2017 & 2018

TECHNICAL SKILLS

Programming Languages: Python, MATLAB, C/C++, Java

Libraries and Tools: PyTorch, PyTorch Lightning, Accelerate, Transformers, DeepSpeed, et al.