

Xinyu Huang (黄新宇)

Updated March 23, 2025

Email: xinyuhuang20@fudan.edu.cn **Phone:** (+86) 18121308479

GitHub: github.com/xinyu1205

HomePage: xinyu1205.github.io

Publications: [Scholar](#)

Brief Bio

From 2020 to 2024, I researched on visual perception and created the [Recognize Anything Model Family](#), a series of open-source and powerful image perception models that exceed CLIP by more than 20 points in fine-grained perception (at OPPO Research Institute & IDEA).

Since 2024, I focus on the full process construction of SOTA large multi-modal models, and lead the pretraining stage of large multimodal models at TikTok Ai Innovation Center, which exceed Qwen2.5-VL.

Education

Nanyang Technological University

Singapore

Visiting Ph.D. in [MMLab](#)

2025, March–2025, September

Mentors: Prof. [Ziwei Liu](#)

Fudan University

Shanghai, China

Ph.D. in Computer Science

2020, September–2025, June

Mentors: Prof. [Rui Feng](#) and Prof. Yuejie Zhang

China University of Geosciences

Wuhan, China

Bachelor of Computer Science

2016, September–2020, June

Mentors: Prof. Maocai Wang. GPA: Rank 2/69, Postgraduate Recommendation

Representative Works

[The Recognize Anything Model \(RAM\) Family \(3.1K+ Stars\)](#)

A series of open-source and powerful image perception models, which **exceed CLIP by more than 20 points at fine-grained perception**.

• [Tag2Text Vision-Language Model](#)

Tag2Text: Guiding Vision-Language Model via Image Tagging

Xinyu Huang, Youcai Zhang, Jinyu Ma, Weiwei Tian, Rui Feng, Yuejie Zhang, Yaqian Li, Yandong Guo, Lei Zhang. *ICLR 2024*.

• [Recognize Anything Model \(RAM\)](#)

Recognize Anything: A Strong Image Tagging Model

Youcai Zhang*, Xinyu Huang*, Jinyu Ma*, Zhaoyang Li*, Zhaochuan Luo, Yanchun Xie, Yuzhuo Qin, Tong Luo, Yaqian Li, Shilong Liu, Yandong Guo, Lei Zhang. *CVPR 2024, Multimodal Foundation Models Workshop*. (*equal contribution)

• [Recognize Anything Plus Model \(RAM++\)](#)

Open-Set Image Tagging with Multi-Grained Text Supervision

Xinyu Huang, Yi-Jie Huang, Youcai Zhang, Weiwei Tian, Rui Feng, Yuejie Zhang, Yanchun Xie, Yaqian Li, Lei Zhang. *Arxiv* 2024.

Other Works

- **IDEA: Increasing Text Diversity via Online Multi-Label Recognition for Vision-Language Pre-training**

Xinyu Huang, Youcai Zhang, Ying Cheng, Weiwei Tian, Ruiwei Zhao, Rui Feng, Yuejie Zhang, Yaqian Li, Yandong Guo, Xiaobo Zhang. *ACM MM* 2022.

- **Simple and Robust Loss Design for Multi-Label Learning with Missing Labels**

Youcai Zhang*, Yuhao Cheng*, Xinyu Huang*, Fei Wen, Rui Feng, Yaqian Li, Yandong Guo. *arXiv* 2022. (*equal contribution)

- **A Medical Multimodal Large Language Model for Pediatric Pneumonia**

Weiwei Tian, Xinyu Huang, Tianhao Cheng, Wen He, Jinwu Fang, Rui Feng, Daoying Geng, Xiaobo Zhang. *IEEE Journal of Biomedical and Health Informatics*, under review.

- **MOSMOS: Multi-Organ Segmentation Facilitated by Medical Report Supervision**

Weiwei Tian, Xinyu Huang, Junlin Hou, Caiyue Ren, Longquan Jiang, Ruiwei Zhao, Gang Jin, Rui Feng, Yuejie Zhang, Daoying Geng. *Biomedical Signal Processing and Control*

- **Principle-Guided Multimodal Reasoning with Minimal Human Demonstrations**

Changkai Ji, Xinyu Huang, Yuejie Zhang, Ying Cheng, Rui Feng, Xiaobo Zhang. *ICASSP* 2024, under review.

- **Predicting Occult Lymph Node Metastasis in Solid-Predominantly Invasive Lung Adenocarcinoma across Multiple Centers Using Radiomics-Deep Learning Fusion Model**

Weiwei Tian, Qinqin Yan, Xinyu Huang, Rui Feng, Fei Shan, Daoying Geng, Zhiyong Zhang. *Cancer Imaging*.

- **Grounded SAM: Assembling Open-World Models for Diverse Visual Tasks**

Tianhe Ren, Shilong Liu, Ailing Zeng, Jing Lin, Kunchang Li, He Cao, Jiayu Chen, Xinyu Huang, Yukang Chen, Feng Yan, Zhaoyang Zeng, Hao Zhang, Feng Li, Jie Yang, Hongyang Li, Qing Jiang, Lei Zhang. *ICCV* 2023 Demo.

Work Experience

Research Intern, at OPPO Research Institute.

Shanghai

Mentors: Youcai Zhang, Yandong Guo

March. 2021 – March. 2024

Led the research on image tagging of foundation vision models (RAM++, RAM).

Led the research on image tagging to guide vision-language models (Tag2Text, IDEA).

Led the research on image tagging learning with missing labels (Hill, SPLC).

Visiting Intern, with International Digital Economy Academy (IDEA).

Mentor: [Prof. Lei Zhang](#)

November. 2022 – November. 2023

Major contributor of foundation visual understanding model ([Grounded-Segment-Anything](#)).

Research Intern, at Tik Tok AI Innovation Center. Shanghai, Singapore

Mentors: [Wei Li](#), [Ziwei Liu](#), [Zejun Ma](#)

March. 2024 – Present

Lead the pretraining stage of large multimodal language model, which exceed Qwen2.5-VL.

Research on multi-modal reasoning.

Projects

[Recognize-Anything](#), Leader & Owner, GitHub Trending with **3.1K+** stars.

[Grounded-Segment-Anything](#), Co-Leader, GitHub Trending with **15.9K+** stars.

[Recognize-Anything Demo](#), Leader & Owner, Hugging Face with **290+** likes.

Grand-Challenge of COVID-19 CT Image Segmentation, rank 54/1016.

Leadership

Leading students in [Cross-Media Intelligent Computing Lab of Fudan University](#), including 2 Ph.D. students and 1 Master student.

Skills & Others

Programming: Python, C/C++, Bash

Misc: PyTorch, Tensorflow, LaTeX, Markdown

Reviewer: IJCV, CVPR, ICLR, WACV, ACM MM

Teaching Assistant: Data Structure

Talk: [AI New Youth](#), [AI Time 2023](#), Invited talk at Huawei Cloud, Invited talk at Fudan Hosted by Prof. Yuejie Zhang