

Xinyu Huang (黄新宇)

Updated July 4, 2024

Email: xinyuhuang20@fudan.edu.cn **Phone:** (+86) 18121308479
GitHub: github.com/xinyu1205 **HomePage:** xinyu1205.github.io **Publications:** [Scholar](#)

Research Interests Computer Vision, Multi-Modality, Visual Recognition

Education

Fudan University	Shanghai, China
Ph.D. in Computer Science	2020 – Present
Mentors: Prof. Rui Feng and Prof. Yuejie Zhang	
China University of Geosciences	Wuhan, China
Bachelor of Computer Science and Technology	2016 – 2020
Mentors: Prof. Maocai Wang. GPA: Rank 2/69, Postgraduate Recommendation	

Representative Works [The Recognize Anything Model \(RAM\) Family \(2.4K+ Stars\)](#)
A series of open-source and powerful image recognition models.

- [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. *ECCV 2024 under review.*

- [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)

- [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.*

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)

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

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

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

Weiwei Tian, **Xinyu Huang**, Junlin Hou, Caiyue Ren, Longquan Jiang, Rui-Wei Zhao, Gang Jin, Rui Feng, Yuejie Zhang, Daoying Geng. *Computers in Biology and Medicine, 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, China

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.

Mentor: [Wei Li](#), [Zejun Ma](#), [Chunyuan Li](#) March. 2024 – Present

Led the research on multimodal alignment large language model.

Major contributor of video multimodal large language models (LLaVA-Video).

Projects

[Recognize-Anything](#), Leader & Owner, GitHub Trending with **2.4K+** stars.

[Grounded-Segment-Anything](#), Co-Leader, GitHub Trending with **13K+** stars.

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: WACV 2023, CVPR 2024, WACV 2024

Talk: [AI New Youth](#), [AI Time 2023](#), Invited talk at Huawei Cloud