Hao Wang

Ph.D. Candidate in Computer Vision, Multimodal Foundation Models, and Agentic AI

Email: wanghao9610@gmail.com | Website: wanghao9610.github.io | WeChat: wangh9610

HCP Lab, Sun Yat-sen University & Pengcheng Laboratory

Expected graduation: December 2026

Open to industry research roles and collaborations

PROFILE

Ph.D. candidate focused on open-ended visual perception, multimodal foundation models, and agentic AI. First-author work includes X2SAM, X-SAM (AAAI 2026), OV-DINO, and TMANet (ICIP 2021), spanning unified image/video segmentation, open-vocabulary detection, and temporal pixel-level understanding. I am interested in connecting foundation models with reliable dense perception for practical vision systems.

RESEARCH CONTRIBUTIONS

- Built a first-author research line on open-ended perception, from video semantic segmentation to open-vocabulary detection and multimodal any-segmentation.
- Developed multimodal segmentation systems that accept both natural-language instructions and visual prompts, bridging high-level reasoning and dense mask prediction.
- Explored unified image/video segmentation with memory mechanisms for temporally consistent object-level and pixel-level understanding.
- Released project pages and code for representative works, emphasizing reproducible and open research practice.

FIRST-AUTHOR PUBLICATIONS

X2SAM: Any Segmentation in Images and Videos **arXiv, 2026**

Hao Wang, Limeng Qiao, Chi Zhang, Guanglu Wan, Lin Ma, Xiangyuan Lan, Xiaodan Liang

Unified segmentation MLLM for images and videos, supporting conversational instructions, visual prompts, and temporally consistent mask memory. Extends any-segmentation from static images to videos while preserving dense pixel-level responses. Paper | Code

X-SAM: From Segment Anything to Any Segmentation **AAAI, 2026**

Hao Wang, Limeng Qiao, Zequn Jie, Zhijian Huang, Chengjian Feng, Qingfang Zheng, Lin Ma, Xiangyuan Lan, Xiaodan Liang

Unified MLLM framework extending segment-anything capabilities to any segmentation and pixel-level perceptual understanding. Supports flexible segmentation through multimodal inputs and instruction-driven interaction. Paper | Code

OV-DINO: Unified Open-Vocabulary Detection with Language-Aware Selective Fusion **arXiv, 2024**

Hao Wang, Pengzhen Ren, Zequn Jie, Xiao Dong, Chengjian Feng, Yinlong Qian, Lin Ma, Dongmei Jiang, Yaowei Wang, Xiangyuan Lan, Xiaodan Liang

Unified open-vocabulary detector pre-trained on diverse large-scale datasets with language-aware selective fusion. Designed to improve category generalization by aligning visual detection with language semantics. Paper | Code

TMANet: Temporal Memory Attention for Video Semantic Segmentation **ICIP, 2021**

Hao Wang, Weining Wang, Jing Liu

Temporal memory attention for long-range video semantic segmentation without optical-flow prediction. Paper | Code

RESEARCH EXPERIENCE

Meituan M17-MM **2025.01 – Present**

Research Intern

- Developed X2SAM, a unified any-segmentation MLLM for images and videos with conversational instructions and visual prompts.
- Investigated mask memory for preserving object-level consistency in video segmentation and interactive perception.

Meituan Vision Intelligence Department **2022.07 – 2025.01**

Research Intern

- Contributed to OV-DINO and X-SAM, advancing open-vocabulary recognition and any segmentation.
- Studied language-aware fusion and multimodal supervision for scalable visual recognition and segmentation.

Selected Applied Research Internships **2019.09 – 2021.08**

Tencent AI Platform Department; Huawei Photo Processing Department

- Conducted applied AI platform and computer vision projects across earlier industry internships.

EDUCATION

Sun Yat-sen University & Pengcheng Laboratory **2022.09 – Present**

Ph.D. student, School of Intelligent Systems Engineering

- Co-supervised by Prof. Xiaodan Liang and Associate Prof. Xiangyuan Lan.
- Research areas: multimodal foundation models, open-ended visual understanding, image/video segmentation, and agentic AI.
- Expected to graduate in December 2026; actively seeking research positions in industry.

University of Chinese Academy of Sciences & Institute of Automation, CAS **2019.09 – 2022.06**

Master student, School of Artificial Intelligence

- Supervised by Prof. Jing Liu.

Beijing Jiaotong University **2015.09 – 2019.06**

Bachelor student, School of Electronic and Information Engineering

ADDITIONAL PUBLICATION

WL-MSR: Watch and Listen for Multimodal Subtitle Recognition **ICASSP, 2023**

Jiawei Liu, **Hao Wang**, Weining Wang, Xingjian He, Jing Liu

Transformer-based multimodal subtitle recognition using OCR and ASR information with mask/crop strategies and multi-level identity embeddings.

PROFESSIONAL SERVICE

Conference Reviewer	AAAI 2026, ICCV 2023, ECCV 2024
Journal Reviewer	Proceedings of the IEEE

AWARDS

2021.09 1st place in the 1st VSPW Challenge Workshop, ICCV 2021.