

# 王豪

计算机视觉、多模态基础模型与智能体方向博士研究生

电子邮箱: wanghao9610@gmail.com | 个人主页: wanghao9610.github.io | 微信: wangh9610

中山大学 HCP 实验室 & 鹏城实验室

预计毕业时间: 2026 年 12 月

寻求工业界研究岗位与科研合作机会

## 个人简介

博士研究生, 研究方向聚焦开放场景视觉感知、多模态基础模型与智能体模型。一作工作包括 X2SAM、X-SAM (AAAI 2026)、OV-DINO 与 TMANet (ICIP 2021), 覆盖图像/视频统一分割、开放词汇检测与时序一致的像素级理解。目前主要关注如何将基础模型与可靠的密集感知能力连接起来, 服务于实际视觉应用。

## 研究亮点

- 建立以第一作者为主的开放场景感知研究线, 覆盖视频语义分割、开放词汇检测与多模态任意分割。
- 构建同时支持自然语言指令和视觉提示的多模态分割框架, 连接高层语义推理与密集掩码预测。
- 探索带记忆机制的统一图像/视频分割方法, 实现跨帧时序一致的对象级和像素级理解。
- 多项代表性工作发布项目主页与代码, 重视可复现和开放共享的研究实践。

## 一作论文

### X2SAM: Any Segmentation in Images and Videos

arXiv, 2026

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

面向图像与视频的统一分割多模态大模型, 支持对话式指令、视觉提示和时序一致的掩码记忆。该工作将任意分割能力从静态图像扩展到视频, 同时保持密集像素级响应。论文 | 代码

### 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

统一的多模态大模型框架, 将 Segment Anything 的能力扩展到任意分割与像素级感知理解。支持通过多模态输入和指令交互实现灵活分割。论文 | 代码

### 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

在多源大规模数据上预训练的统一开放词汇检测器, 提出语言感知的选择性融合机制。该方法通过对齐视觉检测与语言语义提升类别泛化能力。论文 | 代码

### TMANet: Temporal Memory Attention for Video Semantic Segmentation

ICIP, 2021

Hao Wang, Weining Wang, Jing Liu

面向长程视频语义分割的时序记忆注意力方法, 无需光流预测即可建模跨帧上下文。论文 | 代码

## 研究经历

### 美团 M17-MM

2025.01 – 至今

研究实习生

- 研发 X2SAM: 面向图像和视频的统一任意分割多模态大模型, 支持对话式指令和视觉提示。
- 研究掩码记忆机制, 提升视频分割和交互式感知中的对象级一致性。

### 美团视觉智能部

2022.07 – 2025.01

研究实习生

- 参与研发 OV-DINO 与 X-SAM, 推进开放词汇识别和任意分割能力。
- 研究语言感知融合和多模态监督, 用于可扩展视觉识别与分割。

### 早期应用研究实习

2019.09 – 2021.08

腾讯 AI 平台部; 华为照片处理部

- 从事 AI 平台相关场景与计算机视觉应用项目研究。

## 教育经历

### 中山大学 & 鹏城实验室

2022.09 – 至今

博士研究生, 智能工程学院

- 指导老师为梁小丹教授、蓝湘源副教授。
- 研究方向为多模态基础模型、开放场景视觉理解、图像/视频分割与多模态智能体。
- 预计 2026 年 12 月毕业, 正在积极寻求工业界研究岗位。

### 中国科学院大学 & 中国科学院自动化研究所

2019.09 – 2022.06

硕士研究生, 人工智能学院

- 导师: 刘静教授。

### 北京交通大学

2015.09 – 2019.06

本科生, 电子信息工程学院

## 合作论文

### WL-MSR: Watch and Listen for Multimodal Subtitle Recognition

ICASSP, 2023

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

基于 Transformer 的多模态字幕识别方法, 融合 OCR 与 ASR 信息, 并使用掩码/裁剪策略和多层级身份嵌入。

## 学术服务

会议审稿人

AAAI 2026, ICCV 2023, ECCV 2024

期刊审稿人

Proceedings of the IEEE

## 获奖经历

2021.09

ICCV 2021 第 1 届 VSPW Challenge Workshop 第一名。