

Zifu Wan

<https://zifuwan.github.io/>

[✉ zifuw@andrew.cmu.edu](mailto:zifuw@andrew.cmu.edu)

[☎ \(+1\) 412-638-9902](tel:+14126389902)

EDUCATION

-
- Dalian University of Technology, Dalian, China** 09/2019 – 06/2023
• *B.Eng. in Artificial Intelligence, School of Artificial Intelligence* GPA: 89.4/100
- Carnegie Mellon University, Pittsburgh, USA** 08/2023 – Present
• *M.Sc. in Robotics, Robotics Institute, School of Computer Science* GPA: 4.08/4.0

SELECTED PUBLICATIONS (* indicates equal contribution)

-
- **Zifu Wan***, Ce Zhang*, Zhehan Kan, Martin Q. Ma, Simon Stepputtis, Deva Ramanan, Russ Salakhutdinov, Louis-Philippe Morency, Katia Sycara, Yaqi Xie. Incorporating Generative Feedback for Mitigating Hallucinations in Large Vision-Language Models. In submission, 2024.
 - **Zifu Wan**, Pingping Zhang, Yuhao Wang, Silong Yong, Simon Stepputtis, Katia Sycara, Yaqi Xie. Sigma: Siamese Mamba Network for Multi-Modal Semantic Segmentation. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2025. [\[PDF\]](#)[\[Code\]](#)[\[Website\]](#)
 - **Zifu Wan**, Yaqi Xie, Ce Zhang, Zhiqiu Lin, Zihan Wang, Simon Stepputtis, Deva Ramanan, Katia Sycara. InstructPart: Affordance-based Part Segmentation from Language Instruction. In Submission, also at *AAAI Workshop on Public Sector LLMs*, 2024. [\[PDF\]](#)[\[Website\]](#)
 - Tianyu Yan, **Zifu Wan**, Xinhao Deng, Pingping Zhang, Yang Liu, Huchuan Lu. MAS-SAM: Segment Any Marine Animal with Aggregated Features. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2024. [\[PDF\]](#)[\[Code\]](#)
 - **Zifu Wan***, Tianyu Yan*, Pingping Zhang, Gong Cheng, Huchuan L. TransY-Net: Learning Fully Transformer Networks for Change Detection of Remote Sensing Images. In *IEEE Transactions on Geoscience and Remote Sensing (IEEE TGRS)*, 2023. [\[PDF\]](#)
 - Tianyu Yan, **Zifu Wan**, Pingping Zhang. Fully Transformer Network for Change Detection of Remote Sensing Images. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, 2022. (With Student Travel Grant) [\[PDF\]](#)[\[Code\]](#)

RESEARCH PROJECTS

-
- Mitigating Hallucinations in Large Vision Language Models (LVLM) with Generative Feedback** 07/2024 – 10/2024
• Discovered the potential of generative models in mitigating hallucinations in LVLMs at both the response and token levels.
• Proposed a training-free algorithm for LVLMs that enhances the accuracy of responses by integrating generative feedback.
• Outperformed other methods in five benchmarks, covering wide areas from image captioning to visual question answering.
- Sigma: Siamese Mamba Network for Multi-Modal Semantic Segmentation** 02/2023 – 04/2024
• Proposed the first successful application of state space models, specifically Mamba, in multi-modal semantic segmentation.
• Introduced a Mamba-based fusion method and a channel-aware decoder, to extract and integrate information seamlessly.
• Outperformed other methods on four RGB-Thermal and RGB-Depth benchmarks with superior accuracy and efficiency.
- InstructPart: Affordance-based Part Segmentation from Language Instruction** 09/2023 – 01/2024
• Presented the largest real-world task-oriented part segmentation benchmark with hand-labeled instructions and masks.
• Evaluated state-of-the-art vision language models on the benchmark and revealed their limitations.
• Developed a baseline method built upon SOTA foundation models, achieving over a 30% improvement in IoU metrics.
- Adapting Segment Anything Model to Marine Animal Segmentation** 07/2023 – 11/2023
• Developed an adapter-informed SAM Encoder with a hypermap extraction module for marine animal feature extraction.
• Proposed a progressive prediction decoder to capture a wide range of global cues and local details.
• Consistently outperformed other methods on four marine animal segmentation benchmarks.
- Multimodal Renal Tumor CT Image Detection and Segmentation** 02/2023 – 08/2023
• Introduced a large-scale multimodal benchmark (over 20,000 images from four modalities) for renal tumor identification.
• Benchmarked fully-supervised detection methods, achieving an average of 71.3% in AP₅₀₋₉₅ metric.
• Developed a weakly-supervised method for CT image segmentation for computer-aided diagnosis usage.
- Transformer-Based Diffusion Modeling for Change Detection of Remote Sensing Images** 03/2023 – 07/2023
• Developed a Swin-Transformer-based diffusion probabilistic model for remote sensing image feature extraction.

- Designed a spatial-aware self-distillation method for multi-level feature enhancement.
- Surpassed other methods on three change detection benchmarks, achieving superior remote sensing image generation.

Fully Transformer Network for Change Detection of Remote Sensing Images 04/2022 – 04/2023

- Proposed a learning framework for global feature extraction and multi-level feature combination in a pyramid manner.
- Introduced a pyramid structure grafted with a progressive attention module to further improve the feature representation.
- Outperformed most SOTA methods on four public change detection benchmarks with superior efficiency.

Siamese Attentive Convolutional Network for Effective Remote Sensing Image Change Detection 06/2022 – 09/2022

- Proposed to jointly utilize channel-wise and spatial-wise attention mechanisms to aggregate multi-level features.
- Introduced a method to select more discriminating features in a feature difference view during the decoding phase.
- Outperformed SOTA methods on three change detection benchmarks across five metrics.

Semantic Map Construction for Outdoor Mobile Robots 08/2022 – 02/2023

- Implemented Simultaneous Localization and Mapping (SLAM) algorithms on the robotic operating system.
- Designed a derivative algorithm of LeGO-LOAM to enhance the feature-matching phase with extra semantic information.
- Deployed the algorithm to a Robomaster AI robot and reached better performance in the mapping process.

Hand Gesture Controlled Robot Car Based on NVIDIA Jetbot platform 11/2021 – 12/2021

- Embedded a gesture recognition algorithm (based on ResNet) in the NVIDIA Jetson Nano platform.
- Introduced a hand gesture analysis method with real-time image thresholding and key-point extraction.
- Deployed the designed algorithm on an NVIDIA JetBot car, achieving 100% success in gesture control.

WORK EXPERIENCE

Research Assistant in Advanced Agent Robotics Technology Lab, Carnegie Mellon University 08/2023 – Present
Advisor: Prof. [Katia Sycara](#), IEEE/AAAI Fellow; Topic: Robust/Distributed Perception Pittsburgh, USA

- Introduced a comprehensive benchmark for task-oriented part segmentation and affordance learning.
- Developed a Mamba-based fusion method for efficient multi-modal semantic segmentation.
- Proposed a diffusion-based algorithm to mitigate hallucinations in large vision-language models.

Research Intern in AI Innovation Center, Midea Group 02/2023 – 07/2023
Advisor: Dr. [Ning Liu](#); Topic: Model Compression, Efficient Neural Network Design Beijing, China

- Deployed detection/segmentation/3D reconstruction/Automatic Speech Recognition TVM models on internal platform.
- Designed a light-weight semantic segmentation algorithm with magnitude-based pruning and multi-level distillation.
- Optimized ONNX graph before deploying on NVIDIA Jetson Nano platform with TensorRT acceleration.

R&D Intern in Sangfor Technologies Inc. 07/2022 – 09/2022
Topic: Object Detection, Image Classification, Server Communication Design Changsha, China

- Developed a hybrid algorithm combining YOLOv5 and YOLOv7 to identify 16 types of web design elements in UI drafts.
- Utilized PaddleOCR for text identification and designed a lightweight CNN to classify 204 categories of UI icons.
- Assisted in deploying the designed algorithms on the server, which were later in use by UI designers.

Research Assistant in ME Department, Huazhong University of Science and Technology 08/2021 – 09/2021
Advisor: Prof. [Yihua Kang](#); Topic: Signal Processing, Nondestructive Testing (NDT) Wuhan, China

- Studied the relationship between each signal feature and lift-off values extracted from Pulsed Eddy Current (PEC) signals.
- Developed a steel hardness classification method by eliminating the lift-off effect with combined signal features.
- Outperformed traditional methods by over 30% in accuracy for hub bearing groove surface hardness classification.

PERSONAL SKILLS

- **Language Skills:** Chinese: Native; English: Fluent (IELTS 7.5: L 8.0 R 9.0 W 6.5 S 6.5); GRE: 328 (V 160 Q 168 AW 4)
- **Computer Skills:** Python, OpenCV, Pytorch, C++, ROS (Robot Operating System), Linux, LaTeX

ACADEMIC SERVICES

- Journal Reviewer, *International Journal of Computer Vision (IJCV)*
- Journal Reviewer, *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*
- Conference Reviewer, *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2025
- Conference Reviewer, *ACM Multimedia (MM)*, 2024
- Conference Reviewer, *IEEE International Conference on Multimedia and Expo (ICME)*, 2024