華東師範大學
EAST CHINA NORMAL
UNIVERSITY

Junjie Wen

☎ 19946224988 ✉ tsunami1999@163.com

✉ 51255901019@stu.ecnu.edu.cn

Education Experiences

East China Normal University (ECNU) 2022.09 – Now

- M.S. in Computer Science and Technology
- GPA:3.61/4
- Advisor: Chaomin Shen
- Main Courses：Computer Vision、Natural Language Processing、Advanced Engineering Mathematics

South West Jiaotong University (SWJTU) 2018.09 – 2022.06

- B.S. in Software Engineer
- GPA: 3.35/4
- Main Courses：Operating System、Database、Data Mining、Algorithm

Internship experience

Artificial Intellegience Research Center in Midea Group 2023.04 – Now

Embodied Artificial Intelligence Intern Shanghai

2023.04–2023.10: Decoupling multi-modal large models to enhance robot command understanding.

1. **Using multimodal large models to enhance low-level robot motion planning.** Multimodal large models naturally excel in extracting cross-modal features between images and language, aiding robots in better understanding current scenes and instructions. Inspired by the dual-stream hypothesis in cognitive systems, we propose an object-centric enhancement approach. By leveraging open-source multimodal large models, we augment original commands to include richer positional information, facilitating more effective interaction between robotic arms and target objects.
2. **Optimizing the planning efficiency of the high-level planning system.** Robot tasks vary in complexity, with simple tasks typically not requiring enhancement or planning by multimodal large models due to their time-consuming local inference. Therefore, we propose the RFST framework, aimed at automatically classifying tasks. Simple tasks are executed directly through end-to-end low-level models, while complex tasks undergo inference and planning with multimodal large models before being executed step-by-step. This approach aims to improve efficiency and reduce computational costs.

2023.11–2024.06: Exploring end-to-end Visual-Language-Action (VLA) models and scalability.

1. **Building an end-to-end Visual-Language-Action (VLA) model based on VLM and policy models.** In robotics, a common criticism is the generalization problem, where slight changes such as minor shifts in perspective or variations in lighting significantly affect model performance. To enhance the generalization of robotic models, we propose the MuRo-VLA model, which leverages the strong cross-modal semantic understanding capabilities of VLM and policy models. This approach addresses numerous generalization issues in robotics. By utilizing open-source VLM architectures and weights and employing LoRA fine-tuning techniques, only a minimal number of parameters are involved in training.
2. **Exploring the scalability of downstream policy models in the field of robotics.** In the MuRo model mentioned above, we experimented with VLM models ranging from approximately 400 million to approximately 1.4 billion parameters, observing clear scalability where larger models exhibit stronger performance. The size of the downstream policy models used remained unchanged. To further enhance the construction of VLA models, we propose ScaleDP, aimed at exploring the inherent scalability of downstream Policy models

based on the structure of diffusion policy. Model sizes ranged from 10 million to 1 billion parameters, and their performance adhered to the Scaling law.

Publications / Preprints

(∗ *Equal Contribution*)

1. **TinyVLA:Towards Fast, Data-Efficient Vision-Language-Action Models for Robotic Manipulation**
   Junjie Wen*, Yichen Zhu*, Zhiyuan Xu, Jinming Li, Minjie Zhu,Kun Wu, Ning Liu, Chaomin Shen, Yaxin Peng, Feifei Feng, Jian Tang
   submitted to RA-L 2025 [paper] [project page]
2. **Scalable Diffusion Policy: Scale Up DiffusionPolicy via Transformers for Visuomotor Learning**
   Minjie Zhu*, Yichen Zhu*, **Junjie Wen**, Jinming Li, Zhiyuan Xu, Ning Liu, Chaomin Shen, Yaxin Peng, Feifei Feng, Jian Tang
   submitted to ICRA 2025 [paper] [project page]
3. **Discrete Policy: Learning Disentangled Action Space for Multi-Task Robotic Manipulation**
   Kun Wu*, Yichen Zhu*, **Junjie Wen**, jinming Li, Ning Liu, Zhiyuan Xu, Qinru Qiu, Jian Tang
   submitted to ICRA 2025
4. **Object-centric instruction augmentation for robotic manipulation**
   Junjie Wen∗, Yichen Zhu∗, Minjie Zhu, Jinming Li, Zhiyuan Xu, Zhengping Che, Chaomin Shen, Yaxin Peng, Dong Liu, FeifeiFeng, Jian Tang
   ICRA 2024 Oral [paper] [project page]
5. **Language-Conditioned Robotic Manipulation with Fast and Slow Thinking**
   Minjie Zhu∗, Yichen Zhu∗, Jinming Li, **Junjie Wen**, Zhiyuan Xu, Zhengping Che, Chaomin Shen, Yaxin Peng, Dong Liu, FeifeiFeng, Jian Tang
   ICRA 2024 Oral [paper] [project page]
6. **A Survey on Robotics with Foundation Models: toward Embodied AI**
   Zhiyuan Xu, Kun Wu, **Junjie Wen**, Jinming Li, Ning Liu, Zhengping Che, Jian Tang
   Arxiv [paper]

SKILLS

- **Programming language:**Python, C, C++, C#, java
- **English:**CET-4:584, CET-6:508
- **Deeplearning tools:**Pytorch, Tensorflow, jax