CAN CUI

Google Scholar ⋄ Github ⋄ Homepage
Phone: (+86) 153-1611-6391 ⋄ Email: che.cuican@gmail.com

EDUCATION

Shanghai Jiao Tong University, Shanghai, China   Sep. 2020 - Jun. 2023
Master of Engineering in Electronic Information
Thesis: Multi-Label Remote Sensing Image Retrieval and Its Application
GPA: 3.56/4.0

Donghua University, Shanghai, China   Sep. 2016 - Jun. 2020
Bachelor of Engineering in Automation
GPA: 3.56/5.0

EMPLOYMENT

Machine Intelligence Lab (MiLAB), Westlake University   Aug. 2023 - Present
Research Assistant
Focuses on multimodal learning, efficient training/inference, and robot learning.

PUBLICATIONS

[1] **Can Cui**, P. Ding, and et al., "RoboSE: A Simple yet Effective Dual System for Robot Learning," 2025, Under revision (Previously submitted to CVPR 2025), [Link].

[2] B. Jia*, P. Ding*, **Can Cui***, and et al., "Score and Distribution Matching Policy: Advanced accelerated Visuomotor Policies via matched distillation," in *arXiv preprint arXiv:2412.09265*, 2024, [Link].

[3] **Can Cui**, S. Huang, and et al., "ProFD: Prompt-Guided Feature Disentangling for Occluded Person Re-Identification," in *ACM Multimedia (ACM MM)*, 2024, [Link].

[4] **Can Cui**, H. Huo, and T. Fang, "Deep Hashing with Multi-Central Ranking Loss for Multi-Label Image Retrieval," *IEEE Signal Processing Letters (SPL)*, 2023, [Link].

[5] W. Song, H. Zhao, P. Ding, **Can Cui**, and et al., "GeRM: A Generalist Robotic Model with Mixture-of-experts for Quadruped Robot," in *IEEE International Conference on Intelligent Robots and Systems(IROS)*, 2024, [Link].

RESEARCH EXPERIENCE

A Simple yet Effective Dual System for Robot Learning [1]   Sep. 2024 - Present
Supervisors: Prof. Donglin Wang   Westlake University

Dual-system methods integrate Multimodal-Large-Language Models (MLLMs) with small action policy models in Vision-Language-Action (VLA) systems, **balancing generalization and response speed**.

· Analyzed and validated some important design details of previous methods, and proposed a simple yet effective dual-system for robot learning, called **RoboSE**.

· Employed the two-stage training strategy, the **pre-alignment** stage aligns the output embeddings of the MLLM with the feature space of the policy.

· Proposed using **prompt tuning** to approximately fine-tune the high-level MLLM and keeping its parameters frozen, ensuring training efficiency while enhancing generalization and performance.

· Proposed an auxiliary task that projects the output token of MLLM into the action space, compelling the MLLM to engage in **multimodal reasoning**.

· Conducted extensive experiments to demonstrate that RoboSE significantly outperforms existing state-of-the-art methods, and highlighted the critical importance of each component within the model with solid ablation study.

Real-time Diffusion Policy via Distribution Matching Distillation [2]   Jun. 2024 - Present
Supervisors: Prof. Donglin Wang   Westlake University

Diffusion Policy excels in imitation learning, enabling robots to master complex tasks by modeling action distributions. However, its **long inference time** hampers its use in real-world robotic applications.

· Proposed the Distribution Matching Policy (**DM Policy**), an innovative imitation learning algorithm that builds upon pre-trained Diffusion Policy and utilizes two teacher models to guide the optimization of generator.
· Employed **KL divergence** as objective function to align generated distribution of action with the target distribution, overcame stepwise error accumulation of consistency distillation.
· Evaluated on 57 robotic tasks in simulation, achieving a **6× boost in inference speed** and enhanced action quality, balancing speed and precision in robotic action generation.

ProFD: Prompt-Guided Feature Disentangling for Occluded Person ReID[3]   Nov. 2023 - May 2024
Supervisors: Dr. Siteng Huang and Prof. Donglin Wang   Westlake University

ProFD utilizes CLIP to **incorporate textual priors** in Occluded Person ReID, a pure vision task, addressing the two critical problems, missing part information and noisy pseudo-label, in previous methods.

· Designed **part-specific prompts** to introduce the textual prior knowledge of body parts, and utilized spatial-level alignment to align these embeddings with feature map outputted by vision encoder in spatial level.
· Proposed a **hybrid-attention decoder**, combining spatial-aware and semantic-aware attention to generate well-aligned part features, enhancing model performance in occluded scenarios.
· Proposed a **self-distillation strategy** to retain the pre-trained knowledge of CLIP, avoiding catastrophic forgetting and improving generalization.
· Conducted extensive evaluations to validate the effectiveness of ProFD, achieving state-of-the-art results on multiple public Occluded Person ReID datasets.

Deep Hashing with MCR Loss for Multi-Label Image Retrieval [4]   Oct. 2022 - Feb. 2023
Supervisors: Prof. Hong Huo and Prof. Tao Fang   Shanghai Jiao Tong University

To tackle **position conflicts** of proxy-based method and preserve **multi-level similarity** for multi-label image retrieval, we proposed MCR Loss.

· Proposed a **multi-central similarity loss** with learnable hash centers to directly optimize the hashing space, effectively minimizing quantization errors.
· Proposed a **smooth-WAP ranking loss**, an approximation of the weighted Average Precision, to preserve multi-level similarities in the hashing space and address the embedding conflicts caused by proxy-based supervision.
· Conducted extensive experiments to validate the effectiveness of MCR Loss, and achieved state-of-the-art performance on two widely-used benchmark datasets.

ACHIEVEMENTS

The Second Prize Scholarship, awarded by Shanghai Jiao Tong University	Fall 2020 - Summer 2023
Outstanding Graduates , awarded by Donghua University	Summer 2020
Academic Excellence Prize, awarded by Donghua University	Fall 2019
Outstanding Engineer Scholarship, awarded by Donghua University	Fall 2018
National Second Prize, the 13th "NXP Cup" National Smart Car Design Competition	Fall 2018

SKILLSET

Programming Languages: Python, C/C++, JavaScript.
Tools & Platforms: Pytorch, Hugging Face, Docker, Matlab, Git, Vue, Node.js, flask, Linux.
Machine Learning: Vision-Language-Action Model, Generative Model (Diffusion Model, GAN), Large Language Model, CLIP, Parameter Efficient Fine Tuning (LoRA, Prompt Tuning, Adapter), Model Acceleration (Mamba, Mixture-Of-Depth, Consistency Model).
Languages: Chinese(native), English (IELTS: 6.5, with L:7.0, R:6.5, W:6.5, S:6.0).