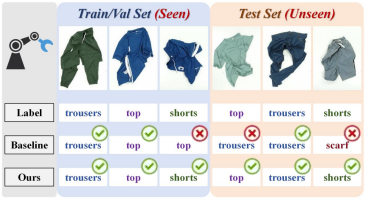
Generalizable Category-Level Topological Structure Learning for Clothing Recognition in Robotic Grasping

Xingyu Zhu[1,2], Yan Wu[3], Zhiwen Tu[1,2], Haifeng Zhong[1,2], Yixing Gao[1,2,*]

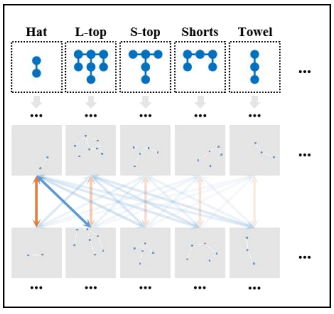
[1] School of Artificial Intelligence, Jilin University, China. [2] Engineering Research Center of Knowledge-Driven Human-Machine Intelligence, Ministry of Education, China.
[3] Robotics and Autonomous Systems Division, Institute for Infocomm Research (I²R), A*STAR, Singapore. *Corresponding author. Email: gaoyixing@jlu.edu.cn

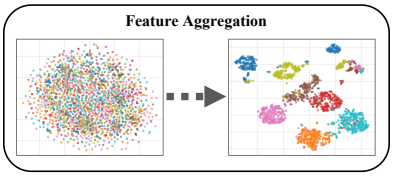
Motivation



- Clothing **deformation features** are highly coupled and rely on **color and texture**.
- Lack of **clothing category-level structural feature learning** process, **low model generalization**.

Topological Structure Representation and Optimization

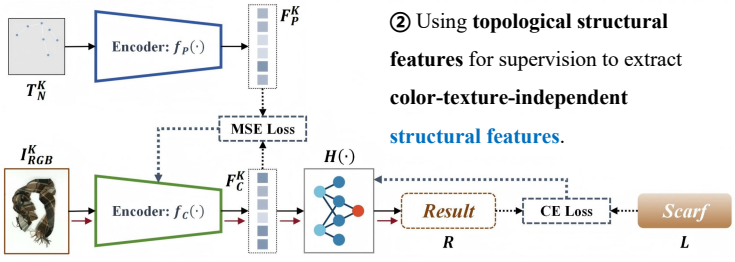




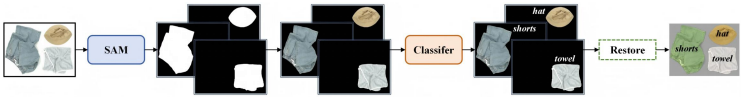
Training Process:

① Extract the **topological structure features** of clothing based on **contrastive learning**;

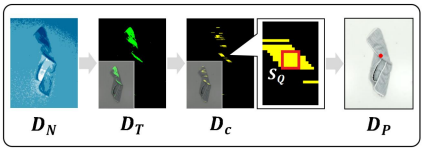
② Using **topological structural features** for supervision to extract **color-texture-independent structural features**.



Multi-clothing Classification Framework



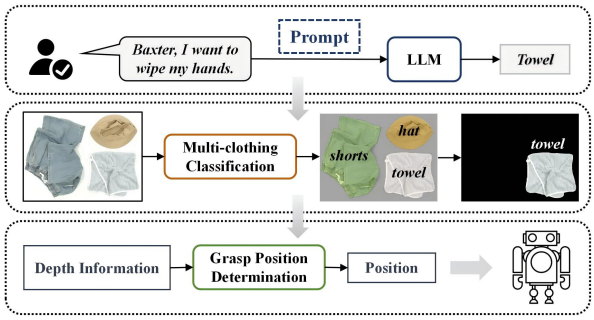
Grasping Position Estimation



- **Depth-Image Protrusion Filtering**
- **Max-Square Grasp Point Lock-on**

Humans are accustomed to looking for **raised wrinkles** on the clothing surface to achieve suitable grasping. The core of our method is to search for **suitable wrinkles** on the clothing surface.

Robotic Clothing Grasping System



- **System Inputs:** **Text data** entered by the user; **RGB images** for multi-clothing classification; **depth information** for grasp position estimation.
- **LLM** is used to analyze user instructions; **regional category information** is output based on a **Multi-Clothing Classification Framework**; and based on the proposed **Grasping Position Estimation Method**, **grasping position information** is output and sent to the robot to complete the user-specified clothing grasping task.

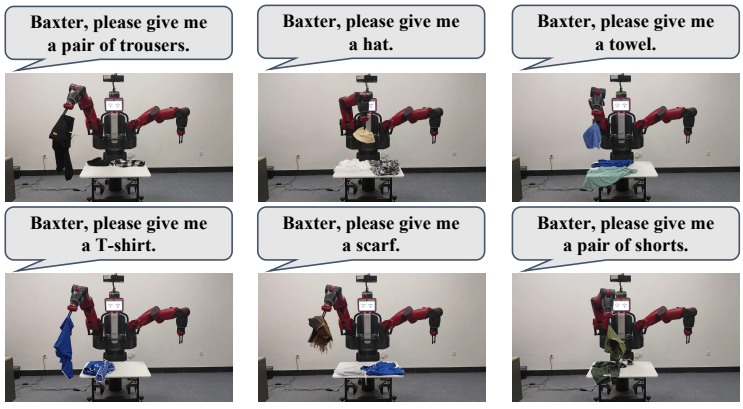
Comparative Analysis

Methods	Seen ↑	Unseen ↑
Shehawy et al.[3]	76.67%	74.52%
Chen et al.[4]	53.81%	52.86%
Ours	**89.29%**	**79.29%**

Methods	Seen ↑	Unseen ↑
GarFusion [2]	76.67%	65.00%
Ours	**89.29%**	**79.29%**

Methods	Classification		Grasping	
	Seen ↑	Unseen ↑	Seen ↑	Unseen ↑
MobileNetV2 [1]	97.11%	53.48%	90.72%	53.57%
Ours	**94.84%**	**84.01%**	**89.29%**	**79.29%**

Real-World Robotic Grasping



- ✓ Our dataset includes **7 categories** and **38 different clothing items**.
- ✓ This **system** enables the **Baxter** to grasp any clothing item based on **user instructions** while demonstrating **strong generalization to unseen targets**.

[1] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4510–4520, 2018.
[2] L. Huang, T. Yang, R. Jiang, X. Tian, F. Zhou, and Y. Chen, "Deforming garment classification with shallow temporal extraction and tree-based fusion," IEEE Robotics and Automation Letters, pp. 1114–1121, 2023.
[3] H. Shehawy, P. Rocco, and A. M. Zanchettin, "Estimating a garment grasping point for robot," in IEEE International Conference on Advanced Robotics (ICAR), pp. 707–714, 2021.
[4] W. Chen, D. Lee, D. Chappell, and N. Rojas, "Learning to grasp clothing structural regions for garment manipulation tasks," in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4889–4895, 2023.

This research was supported by the National Natural Science Foundation of China under Grant Nos. 62203184 and W2421093, and the International Cooperation Project of Jilin Province under Grant No. 20250205079GH.