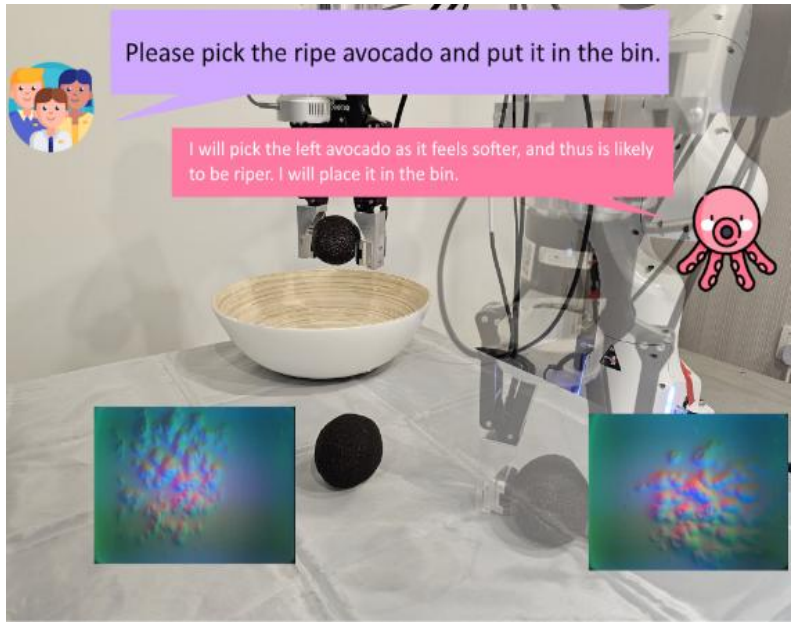
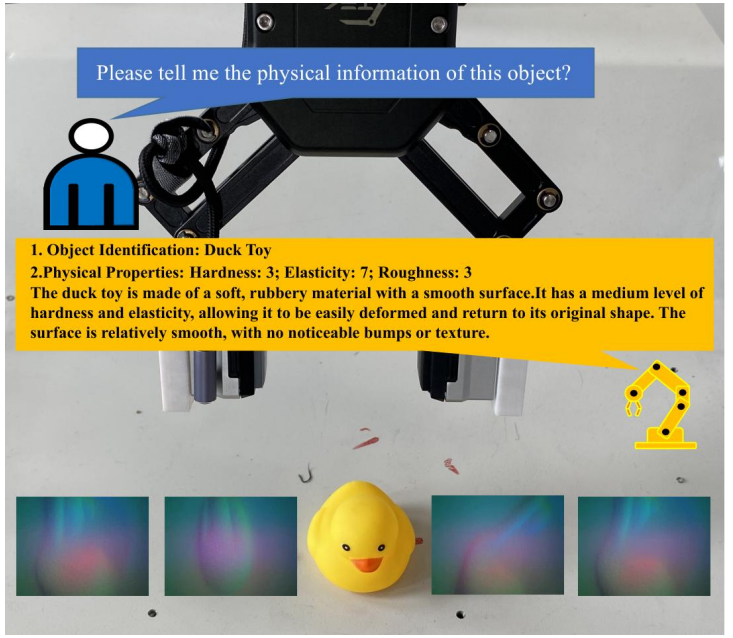
Octopi-X: Robotic Perception with a Large Tactile-Vision-Language Model for Physical Property Inference

Zexiang Guo[1]*, Hengxiang Chen[1]*, Xinheng Mai[1]*, Qiusang Qiu[1], Gan Ma[2], Zhanat Kappassov[3], Qiang Li[1]†, Nutan Chen[4]





Project Homepage:

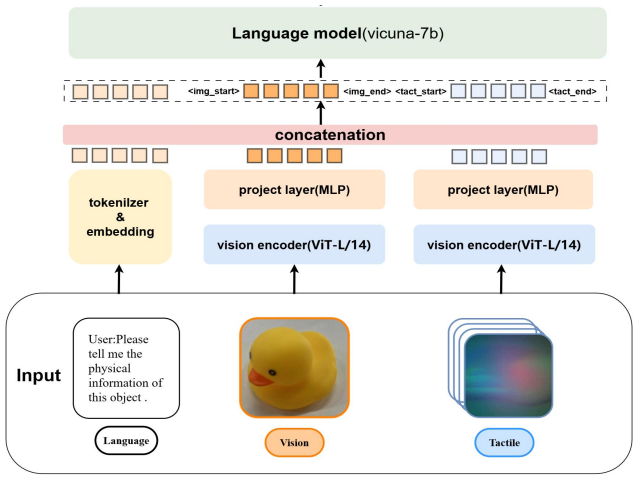
Hardness, elasiticity and bumpiness ✓
Only Tactile-Language reasoning ✗
Pair-level reasoning ✗
Zero-Shot Generalization Ability ✗

Hardness, elasiticity and roughness ✓
Tactile-Vision-Language reasoning ✓
Instance-level reasoning ✓
Zero-Shot Generalization Ability ✓

深圳技术大学 Shenzhen Technology University

Dexterous Robotics Lab

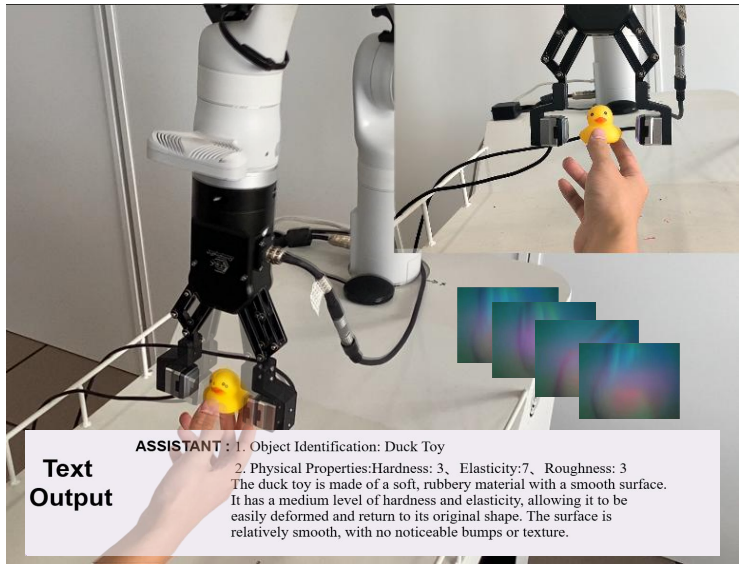
(i) Methodologies



Reference Score Range:

Property	Score Range	Characterization	Example Materials
Hardness	1–2	Extremely soft	Cotton, sponge
	3–4	Soft	Rubber ball, soft plastic toy
	5–6	Medium	Plastic container, shoe sole
	7–8	Hard	Wood, ceramic plate
	9–10	Extremely hard	Metal, diamond
Elasticity	1–2	Minimal elasticity	Clay, dry sponge, wooden ruler
	3–4	Low elasticity	Rubber eraser, hard plastic, book cover
	5–6	Medium elasticity	Foam ball, silicone, thick rubber mat
	7–8	High elasticity	Rubber band, bouncy ball, yoga mat
	9–10	Maximum elasticity	Trampoline surface, latex sheet, inflated balloon
Roughness	1–2	Extremely smooth	Glass, polished marble
	3–4	Smooth	Plastic surface, ceramic mug
	5–6	Medium texture	Paper, leather, cardboard
	7–8	Rough	Sandpaper, concrete, bark of a tree
	9–10	Extremely rough	Gravel, coarse fabric, pumice stone

Reasoning → Action



(ii) Zero-Shot Generalization- Spearman Rank Correlation (ρ) with Ground Truth Measurements

35 unseen objects

Attribute	Method	Correlation Coefficient	P-value
Hardness	Octopi-ViTaL	**0.501**	**0.005**
	Octopi-ViTaL (vision only)	0.307	0.099
	Octopi (fine-grained)	0.307	0.099
	Octopi (original)	0.015	0.935
Elasticity	Octopi-ViTaL	**0.530**	**0.003**
	Octopi-ViTaL (vision only)	0.452	0.012
	Octopi (fine-grained)	0.053	0.781
	Octopi (original)	-0.060	0.753
Roughness	Octopi-ViTaL	**0.643**	**0.0001**
	Octopi-ViTaL (vision only)	0.413	0.023
	Octopi (fine-grained)	-0.010	0.959
	Octopi (original)	0.118	0.534

(iii) Conclusions

i) Integrated vision, tactile, and language inputs → Improved property inference.

ii) Our structured prompting → Achieved strong results on hardness, elasticity, and roughness.

iv) Demonstrated zero-shot generalization across unseen objects.

Project Homepage:

Yu et al., "Octopi: Object Property Reasoning with Large Tactile-Language Models," RSS 2024.