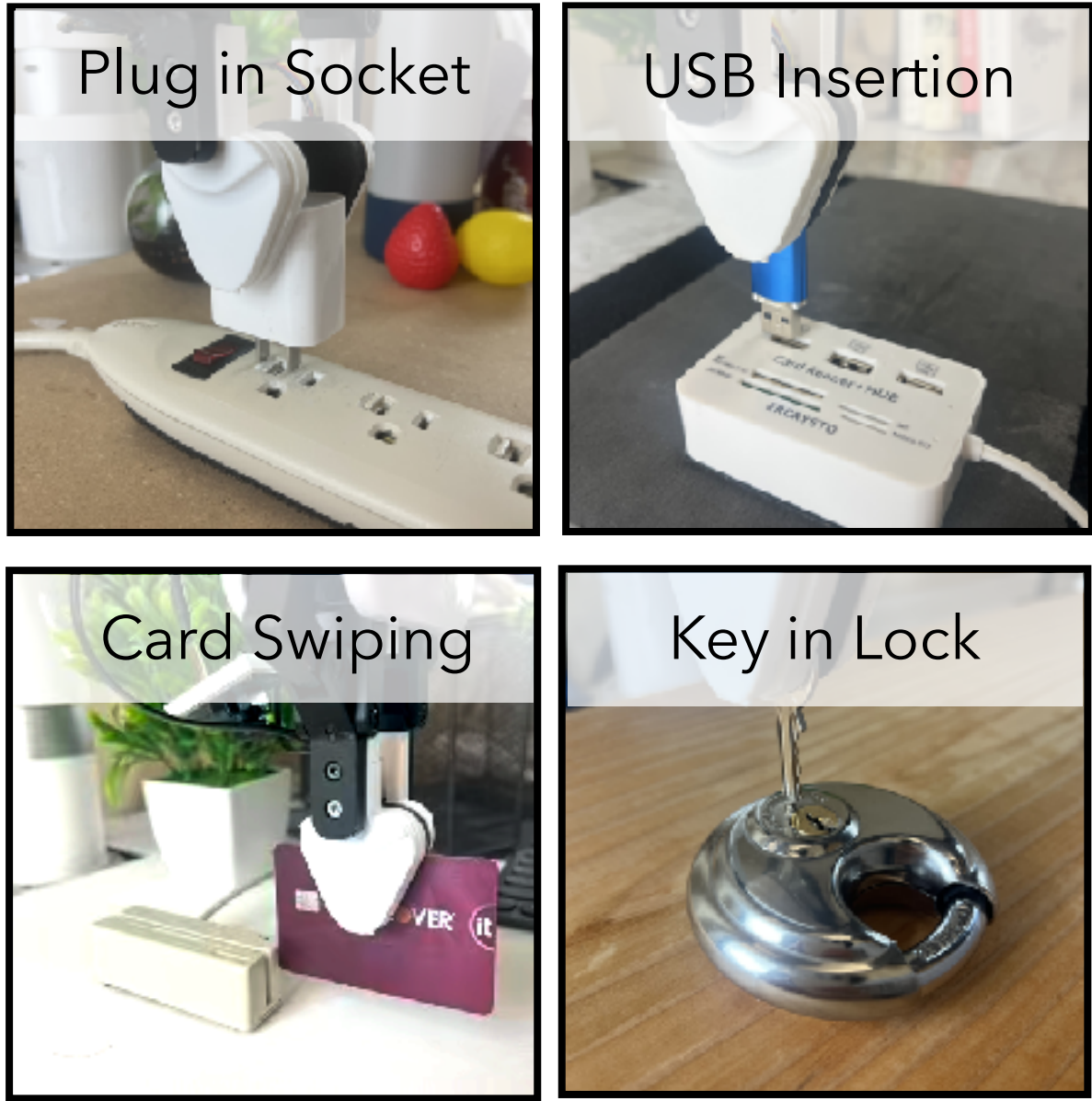
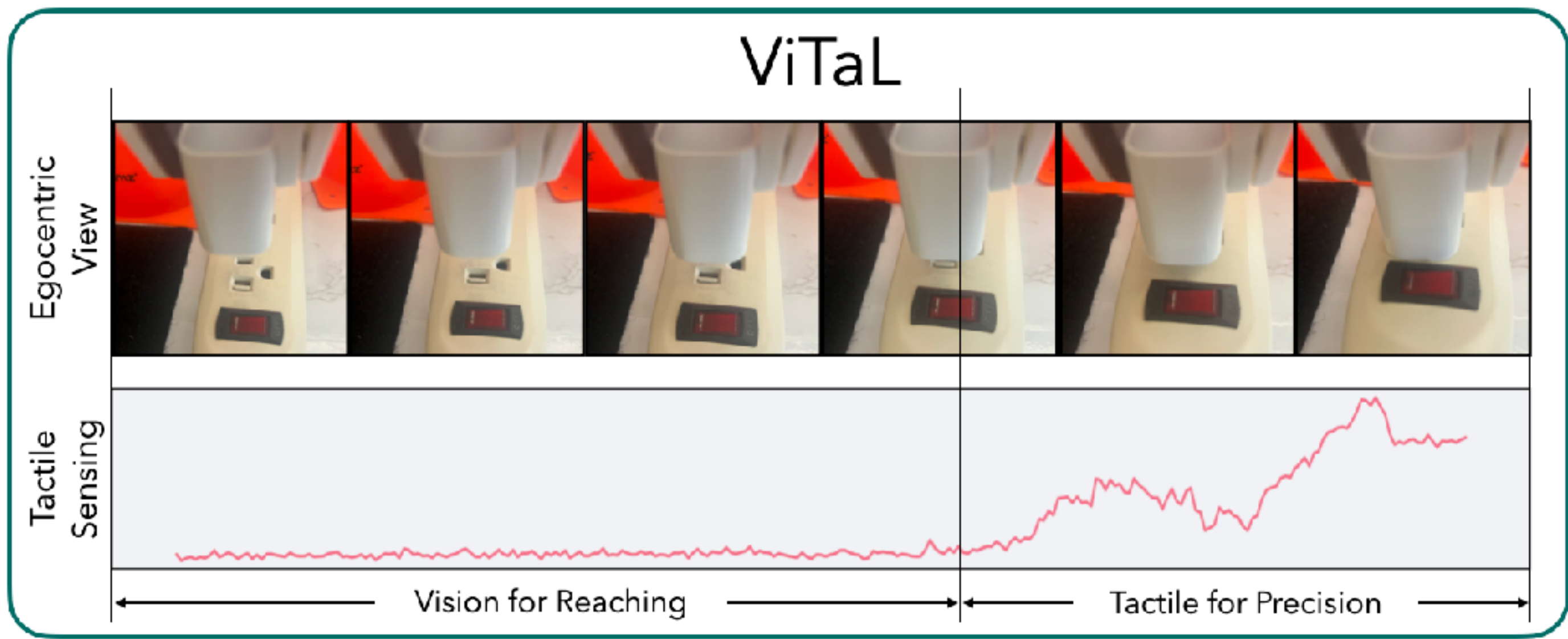


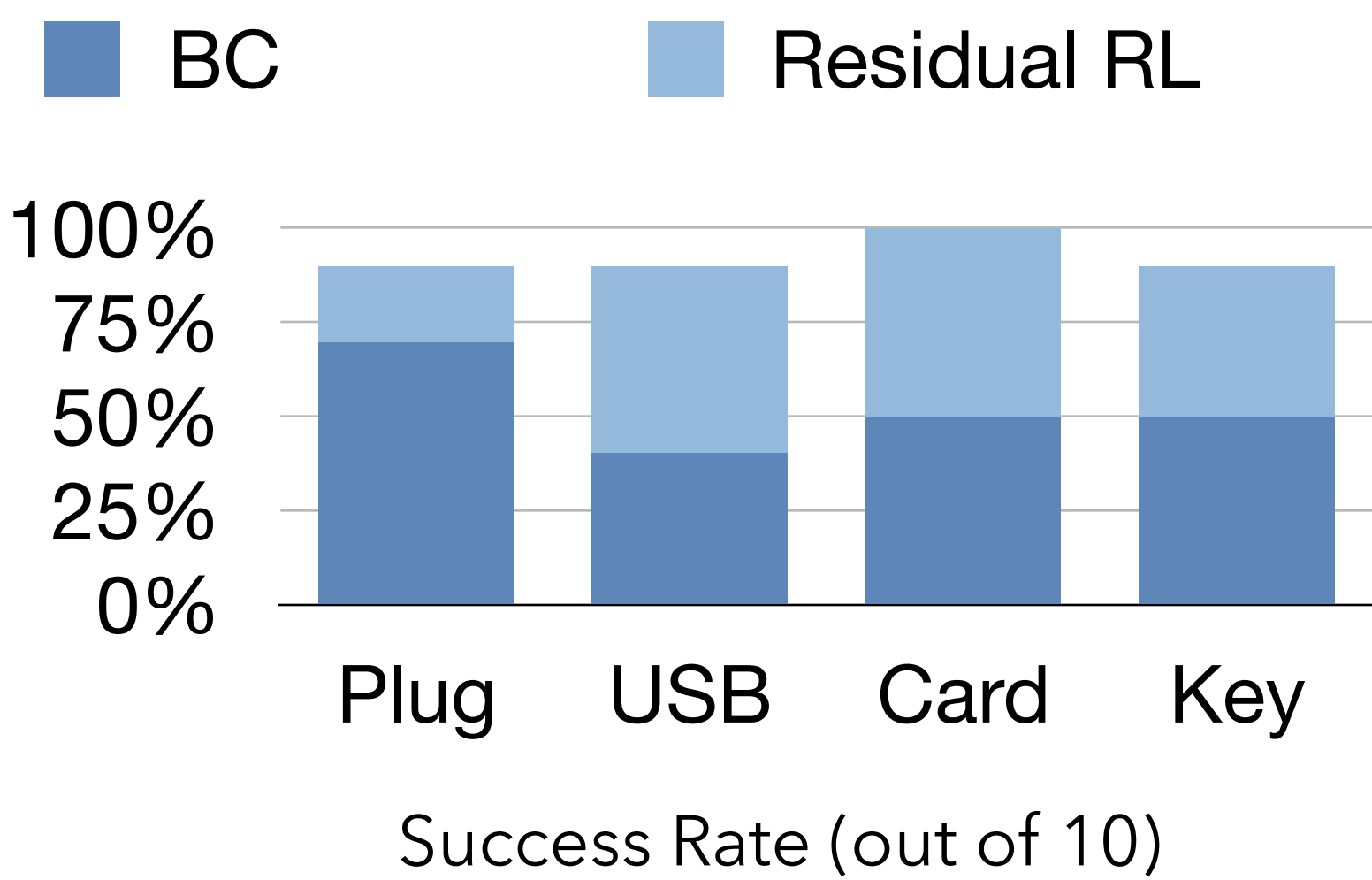
Touch begins where vision ends: Generalizable policies for contact-rich manipulation

Zifan Zhao[‡], Siddhant Haldar[‡], Jinda Cui[‡], Lerrel Pinto[‡], Raunaq Bhirangi[‡]

Vision for Reaching, Tactile for Precision

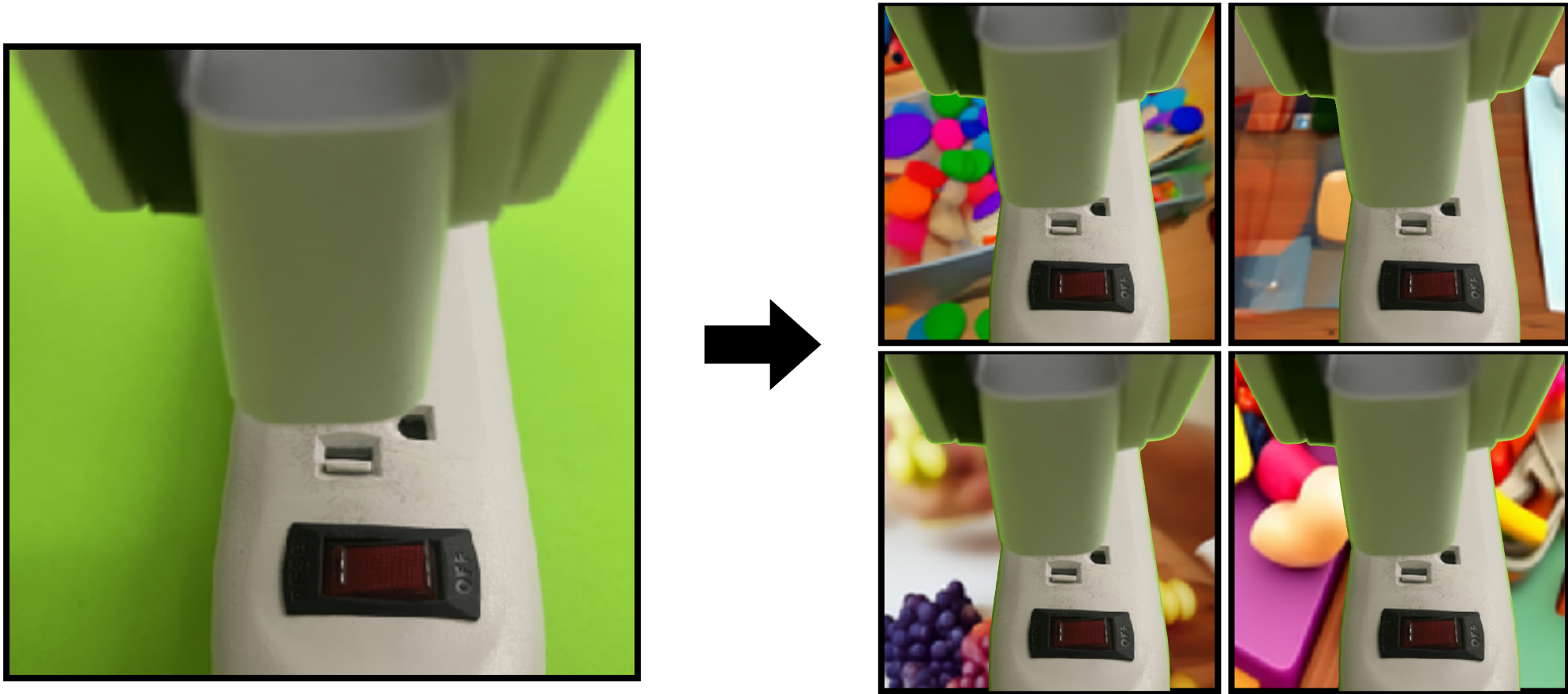


Robust manipulation on millimeter-scale precision



Generalizability to unseen environments

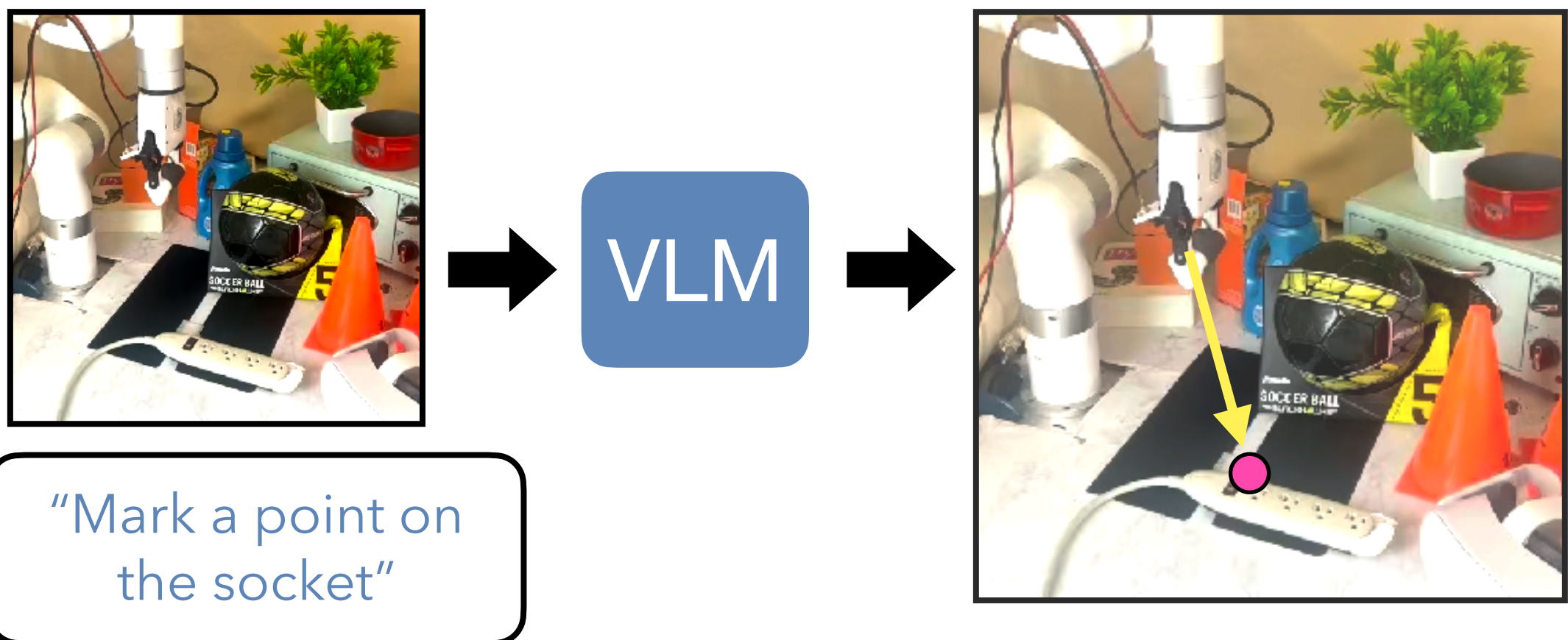
(A) Semantic Augmentation



(B) VisuoTactile BC + Residual RL

Policy	Average SR (%)	Generalizability
BAKU	35	✗
VISK	47.5	✗
RLPD	5	✗
ViTaL	92.5	✓

(C) VLM for Scene-level Reasoning and Localization



Key Takeaways: ViTaL uses VLM for scene-level object localization and VisuoTactile Local policy for fine-grained manipulation, achieving **~90%** success in unseen, distractor-rich environments.



More videos !