ICML 2026 POSTER

Tabero: Learning Gentle Manipulation with Closed-Loop Force Feedback from Vision, Touch, and Language

Qiwei Wu[1,2] · Rui Zhang[1] · Xin Xiang[2] · Tao Li[1] · Weihua Zhang[1] · Junjie Lai[1, ✉] · Renjing Xu[2, ✉]

[1]NVIDIA · [2]Hong Kong University of Science and Technology (Guangzhou)

Paper & Code

TABERO — **A tactile benchmark for gentle manipulation**

Unified evaluation suite for vision-tactile-language data, multi-model force-regime comparison, and gentleness-aware assessment.

5	4	20 Hz
synchronized modalities	gentleness metrics (AG, MG, AA, MA)	tactile-force closed loop

1 Why Gentle Manipulation?

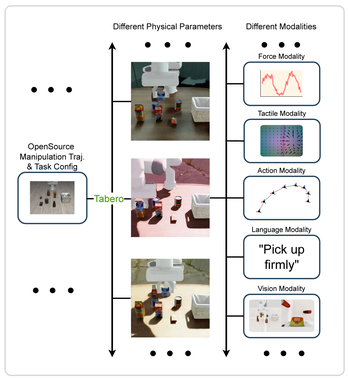
VLA models lack tactile modality — they cannot regulate contact forces. Without force feedback, policies frequently **damage objects**. Existing benchmarks ignore *how* an interaction unfolds, measuring only binary success.

- **Data scarcity:** Real tactile data is costly; simulators lack high-fidelity tactile signals.
- **No gentleness metric:** No benchmark measures whether excessive force was applied.

Can VLAs regulate contact forces for **gentler manipulation** without sacrificing task success?

2 Tabero: Data Pipeline

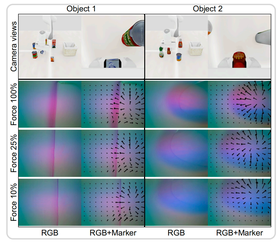
Tabero **repurposes open-source trajectories** (e.g., LIBERO) by replaying them in a high-fidelity tactile simulator, generating aligned multimodal data **without new hardware**.



Cross-platform pipeline: MuJoCo → Isaac Lab replay produces synchronized RGB-D, tactile images, marker displacement fields, and 6D contact forces at 20 Hz.

3 Multimodal Data & Force Diversity

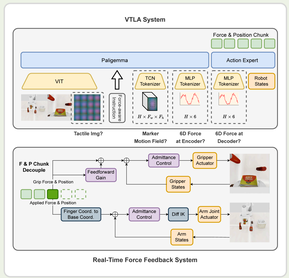
Five synchronized modalities via GPU rendering: tactile images (Taxim), 11 × 9 marker fields, and 6D contact forces. Varying (K_p, K_d) creates diverse force profiles with language augmented accordingly.



Tactile data at 100%, 25%, 10% force: RGB images + marker overlays show distinct contact patterns.

4 Tabero-VTLA ★ KEY

A TCN encodes marker fields into conditional tokens; the policy jointly predicts **gripper poses and force setpoints**, tracked by a decoupled admittance-based low-level controller.



System overview: Tactile inputs fuse with vision + language; decoupled force-position controller tracks predicted commands with real-time feedback.

5 Main Results ★ Headline

9 LIBERO Object tasks across three datasets. Field+FS = marker field + force supervision.

Metric	A (100% / 25%)	B (100% / 10%)	C (100% / 10% bin.)
Success Rate	0.87 / 0.79	0.86 / 0.52	0.86 / 0.87
Avg Grip Force (N) ↓	**31.3 / 8.5**	**32.4 / 3.7**	35.8 / 34.6
Max Grip Force (N) ↓	61.3 / 19.9	62.7 / 11.4	65.7 / 65.0

73% ↓	89% ↓	0.52
AG reduction (25% force)	AG reduction (10% force)	gentle SR at 10% force

Without tactile feedback (C), force modulation fails completely. Field+FS enables precise force tracking under extreme 10% regimes.

6 Ablation: Tactile Modalities

Dataset B (100% / 10% force). F/G = Firm/Gentle.

Model	F SR	G SR	F AG	G AG
None (no tactile)	0.00	0.00	0.0	0.0
Img	0.37	0.01	3.0	1.1
Field	0.40	0.01	2.9	2.0
Force E	0.40	0.01	2.5	1.8
FS (force sup. only)	0.82	0.45	30.4	3.1
Force E+FS / Img+FS	0.84–0.87	0.48–0.49	30.3–30.6	3.4–3.6
Field+FS	**0.86**	**0.52**	32.4	**3.7**

Force supervision is critical: without it, tactile representations alone barely succeed. Field+FS achieves the best gentle success with the lowest grip force.

0.00 → 0.52	32.4 → 3.7	×88%
G SR gain with Field+FS	AG reduction gentle	force cut vs. no tactile

7 Semantic Force & Metrics

Can Tabero-VTLA map **unseen adverbs** to appropriate force?

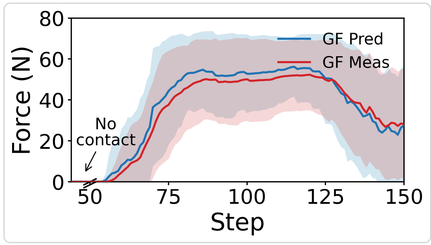
Adverb	SR	AG ↓	MG ↓
firmly, tightly	0.86	32.4	62.7
gently, softly	0.52	3.7	11.4
lightly (OOD)	0.61	14.5	40.8
forcefully (OOD)	0.64	18.1	49.7

OOD adverbs produce **intermediate forces** aligned with semantics — zero-shot compositional understanding.

Beyond success rate: We introduce **MG, AG, MA, AA** — process-aware metrics that quantify *how* an interaction unfolds, not just whether the task completed.

8 Controller Ablation

Isolating the low-level controller on **Dataset A**: we ablate feedforward and admittance terms to measure their contribution to force tracking accuracy.



Grip force tracking (predicted blue, measured red): (a) 100%, (b) 25% hybrid, (c) 25% no feedforward, (d) 25% no admittance. Both feedforward and admittance terms are essential.

Gentle grasping requires **both** sensory awareness (tactile feedback) **and** compliant actuation (hybrid controller). With FF + admittance ✓ | No FF ✗ | No admittance ✗

9 Takeaways

IDEA. Tactile feedback is essential — vision alone cannot regulate contact forces for gentle manipulation.

METHOD. Decoupled force-position control: VLA models reason about *how* to touch, not just *where*.

RESULT. Force-aware policies modulate grip force across regimes while maintaining high task success.

PRACTICAL. Scalable pipeline: repurpose any open-source dataset to add tactile supervision, no new hardware.

HKUST · NVIDIA