FAST

Efficient Action Tokenization for Vision-Language-Action Models

Pertsch et al. · Physical Intelligence · 2025 · arXiv:2501.09747

The problem

- Binning: a token per dimension per step
- Shirt folding: 700 tokens a second
- Next token ≈ copy of the last
- No progress at 20 Hz and 50 Hz

so

The tokenizer

Quantile-normalize to [−1, 1]

↓

DCT on each action dimension

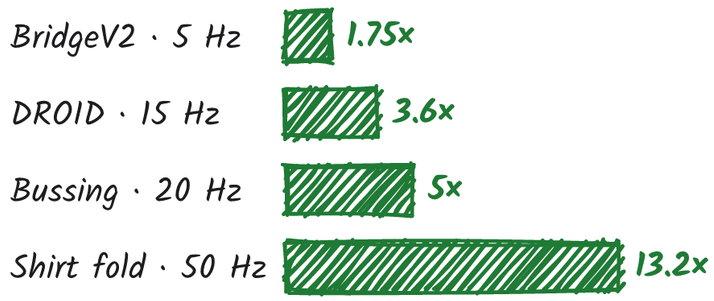
↓

Scale ×10, round: mostly zeros

↓

Low frequencies first → BPE

Fewer tokens



naive ÷ FAST tokens per 1 s chunk · Table I

FAST+ (universal)

- One vocabulary, fit on ~1M chunks
- ≥2× fewer tokens on 14 unseen sets
- Matches per-dataset FAST

Zero-shot DROID

- Trained on 75k successful episodes
- Unseen test scenes, no extra tuning
- 44 scored trials (Table II)

π0-FAST vs diffusion π0

π0-FAST (tokens)

- Matches π0 on generalist tasks
- 3× fewer steps on bussing
- ~750 ms per 1 s chunk

vs

π0 (diffusion)

- Typically ≤100 ms on a 4090
- 10 steps, 300M action expert
- Often ignores DROID prompts

≈20–30

FAST tokens per arm per 1 s chunk (Table I)

5×

fewer GPU hours to train π0-FAST

Takeaway

- Compress actions before tokenizing
- Plain next-token VLAs then learn dexterous 50 Hz tasks
- Limits: slow decoding, and policies tested only on static arms

 · paper CC BY 4.0 · sketchnote by Pavel Bushuyeu

Active listening · FAST

Write while you listen. One line per box is enough.

Pertsch et al. · Physical Intelligence · 2025 · arXiv:2501.09747

Date _____ Presenter _____ Role _____

Their main claim, in my words

The evidence they show

What I would push on

My question for the room

How it connects to my work

Listen for

- ☐ Why binning breaks at high frequency (spline toy task)
- ☐ Which steps lose information: rounding, not DCT or BPE
- ☐ Which results use FAST, which universal FAST+
- ☐ Does it hold on a second backbone (OpenVLA)?
- ☐ What the 5× measures, and what got slower

Ask the room

1. ~750 ms per 1 s chunk: which tasks break first?
2. Why would token decoding follow language better?
3. What policy test would show FAST+ transfers?

Page, video and FAQ:
bushuyeu.com/read-along-fast