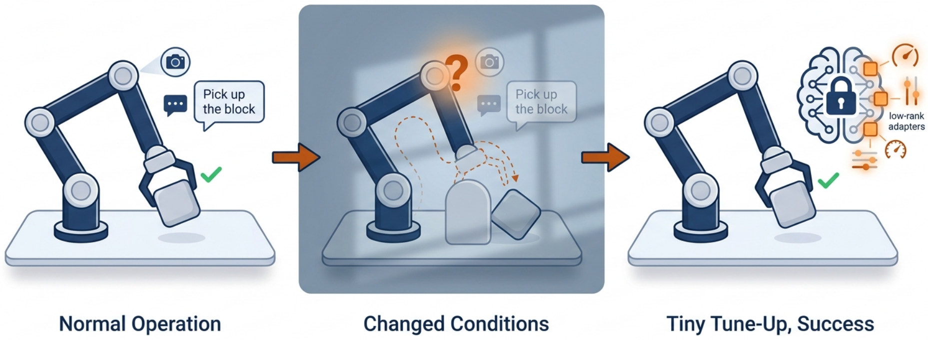
AARHUS UNIVERSITY
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Date: June 7, 2026

A3 Lab

MASTER'S PROJECT AVAILABLE

Parameter-Efficient Test-Time Adaptation for Tiny Vision-Language-Action Models



Motivation

Vision-language-action (VLA) models offer a new route to flexible robotic manipulation: instead of low-level programming, an operator describes a task in natural language and the model maps perception and language directly to robot actions [1]. This suits small-batch and SME manufacturing, where objects, poses, and workcell conditions change frequently. However, VLA models are essentially *static* at deployment — they produce plausible actions, but plausibility does not guarantee reliability when the scene shifts, and the policy can become uncertain or simply wrong. Test-time adaptation (TTA) recovers much of this lost performance by adapting on unlabeled test observations [4, 5], but standard TTA back-propagates through the *entire* encoder, making it memory-intensive and impractical on the consumer/edge GPUs typical of affordable robots. Tiny VLAs such as SmolVLA (450M parameters) [1] run on a single consumer GPU, making them an ideal substrate for a lightweight, deployable adaptation scheme — and the whole problem can be studied reproducibly in simulation, with *no physical robot or digital twin*.

Objective & Methodology

This thesis develops a **Parameter-Efficient, Test-Time Adaptation** framework for a tiny VLA. Using SmolVLA on the public **LIBERO** benchmark [2] — whose suites are explicitly built around distribution shifts in object type, spatial layout, and goals — the aim is to recover robustness under deployment shift on a small update budget.

Key Tasks:

- **Baseline & shift characterisation:** Run SmolVLA on the LIBERO suites and (optionally) inject controlled visual corruptions — lighting, blur, occlusion — into the renderings to create explicit deployment-shift conditions. Establish baselines and quantify the performance drop.
- **Efficient adaptation:** Implement and compare parameter-efficient methods restricted to the test-time phase such as Low-Rank Adaptation (LoRA) [3], or learnable prompts driven by unsupervised signals.
- **Benchmarking:** Quantify the accuracy–efficiency trade-off: task success under shift vs. trainable parameters, peak VRAM during the update, and adaptation/inference latency — against a frozen baseline and full-encoder TTA.

Requirements

- Strong programming skills in Python and PyTorch.
- Solid understanding of Deep Learning, particularly transformers and Vision-Language Models.
- Familiarity with parameter-efficient fine-tuning (LoRA), domain adaptation, or test-time adaptation is a strong plus.
- Interest in robotics / embodied AI (e.g. the Hugging Face LeRobot ecosystem) is welcome.

References

[1] Shukor, M. *et al.*, "SmolVLA: A Vision-Language-Action Model for Affordable and Efficient Robotics," *arXiv:2506.01844*, 2025.
[2] Liu, B. *et al.*, "LIBERO: Benchmarking Knowledge Transfer for Lifelong Robot Learning," *NeurIPS Datasets and Benchmarks Track*, 2023.
[3] Hu, E. J. *et al.*, "LoRA: Low-Rank Adaptation of Large Language Models," *ICLR*, 2022.
[4] Tomar, D., Vray, G., Thiran, J.-P., Bozorgtabar, B., "Un-mixing Test-Time Normalization Statistics: Combatting Label Temporal Correlation," *ICLR*, 2024.
[5] Vray, G. *et al.*, "ReservoirTTA: Prolonged Test-Time Adaptation for Evolving and Recurring Domains," *NeurIPS*, 2025.

Application

If you are interested in this project, please email your CV and a brief transcript to:

Dr. Behzad Bozorgtabar at **behzad@ece.au.dk**.