Yash Thube

Pune, India | thubeyash09@gmail.com | github.io

Research & applied interests

Generalist embodied agents that learn representations of spatial, temporal, physics dynamics, and interact with the physical world.

Broadly - **Model based RL** (representation learning, temporal abstraction, long horizon planning, robot learning), **3D scene understanding** (spatial, causal), **Continual and open ended learning**, Architectures (VLMs, VLAs, World models)

Tools - PyTorch, OpenCV, HuggingFace (Transformers, TRL, Diffusers, PEFT), Git, NumPy, TorchVision, TransformerLens, Gym, PyBullet, Pillow, MatPlotlib.

Experience

- **Machine Learning Engineer**, Hudl India – Pune, MH   03/2024 – 11/2024

  Implemented YOLOv8 based pipeline for real time player detection and tracking in sports videos. Fine-tuned VLMs on custom sports datasets for semantic video understanding. Achieved 92% precision in crowded scenes using DeepSORT multi-object tracking.

- **Deep Learning Collaborator**, Omdena – remote,   10/2023 – 01/2024

  Architected transformer-based pipeline for mental health sentiment analysis on social platforms. Implemented RAG with fine-tuned LLMs and developed audio classification models using CNNs for emotional tone detection in voice content.

- **Computer Vision Collaborator**, AI Accelerator Institute – remote,   02/2023 – 06/2023

  Worked on segmentation and self supervised learning for vision.

- **Technical Writer**, InPlainEnglish | Towards AI – remote   04/2022 – 05/2023

  Distilled complex technologies into clear and accessible content with primary focus on ML and AWS.

Projects & Research

- **MATS (arXiv preprint)** – A behavioral audit toolkit to detect pathological truth bias in VLMs, experiments include activation patching to causally localize failures in cross-attention layers and pooled representations across LLaVA, CLIP, and Qwen-VL architectures.

- **Paper Implementations** – PyTorch implementations of research papers, aimed at deepening my understanding of the underlying concepts.

- **Multimodal/VLMs Research Hub** – A technical resource for researchers exploring Vision-Language Models (VLMs) and multimodal learning, featuring seminal papers/models, datasets, benchmarks, ethical challenges, and research directions.

- **Task-aware SAM LoRA** – PyTorch pipeline that uses a hypernetwork to generate task-specific LoRA adapters for Meta's Segment Anything Model from natural language prompts, targeted segmentation on COCO instances and benchmarked mIoU via pycocotools. Currently

Github

Education

Savitribai Phule Pune University (SPPU)   Expected - 2026
B.E. Computer Engineering | Pune, India