Yash Thube

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

Research interests & Technologies

Generalist embodied machines that intuitively understand space, time & physics, and interact with the physical world.

Broadly - World Models (extending VLMs/VLAMs to learned dynamics), 3D scene understanding (spatial, causal), long-horizon planning, reinforcement & open ended learning

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

Experience

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

  Enhanced sports video performance tracking accuracy by 35% through deep learning model development. Automated video classification using computer vision (SVM, CNN), reducing manual review time by 60% and streamlining workflows.

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

  Analyzed social media's mental health impact and subsequently designed/implemented solutions for healthier online interactions using language models, RAG, prompt engineering, NLP, and audio processing.

- **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 AWS, ML and Serverless technologies.

Projects

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

- **LatentLinter** – A plug-and-play PyTorch library that adds statistical regularization to gradient-based planners in model-based RL. It detects when planned trajectories leave the training manifold and penalizes them, acting as a safety check for optimizer stability.

- **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.

Github

Education

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