Efficient and Robust VLA Models for Real-Time Robotic Manipulation: Week 8 Update

Shriyadita Medavaram

Faculty Mentor: Dr. Yuzhang Shang

Student Mentor: Lam Nguyen

Project Recap

- Goal: Compress Pi0.5 to make smaller faster VLA models
- Two parts
  - Run Pi0.5 model on robot arm
  - Compress Pi0.5 LIBERO checkpoint



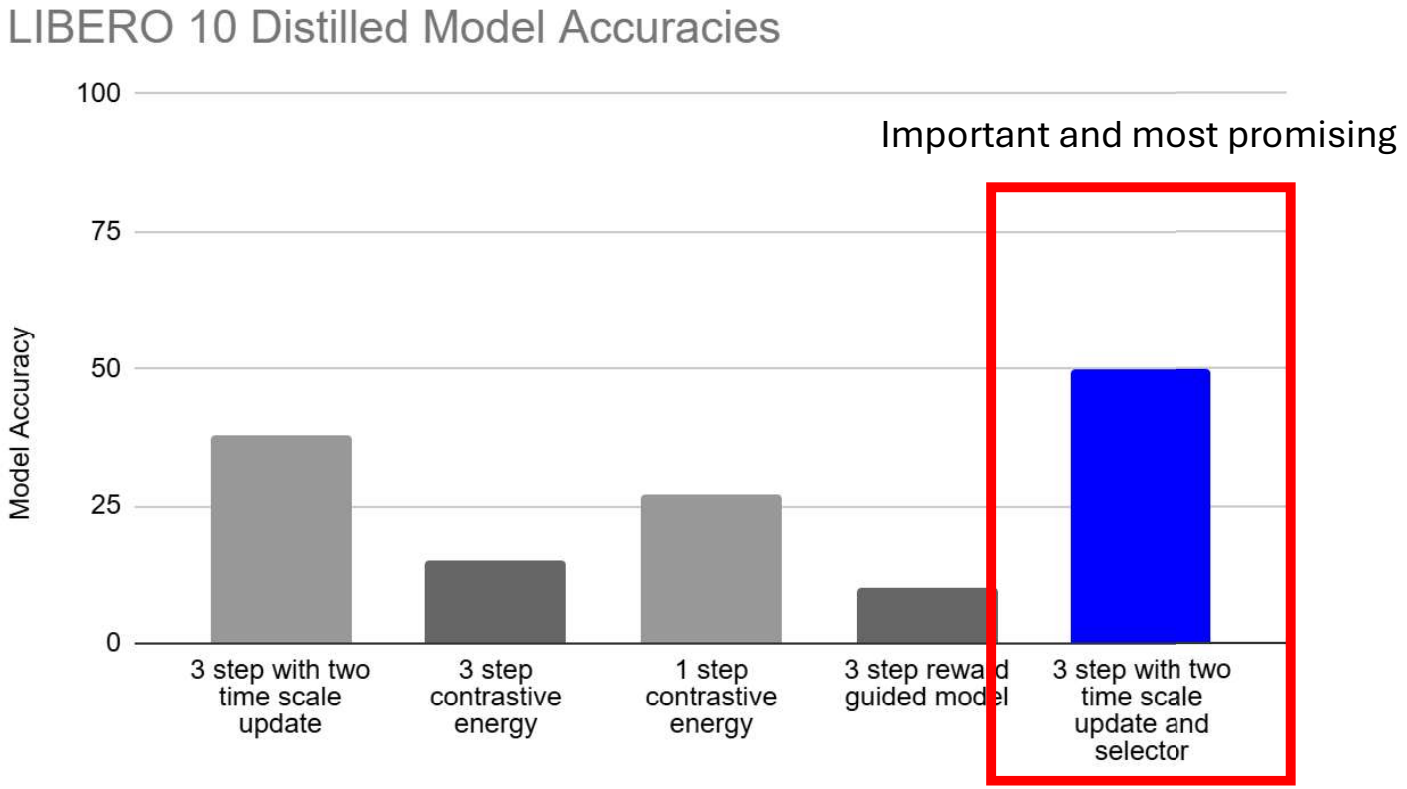
Update: Robot Arm Progress

- Picks up smoother
- Slight runner modifications
- Robot Arm accuracy:
  - Average: 65%
- Future work: further training on more specific tasks

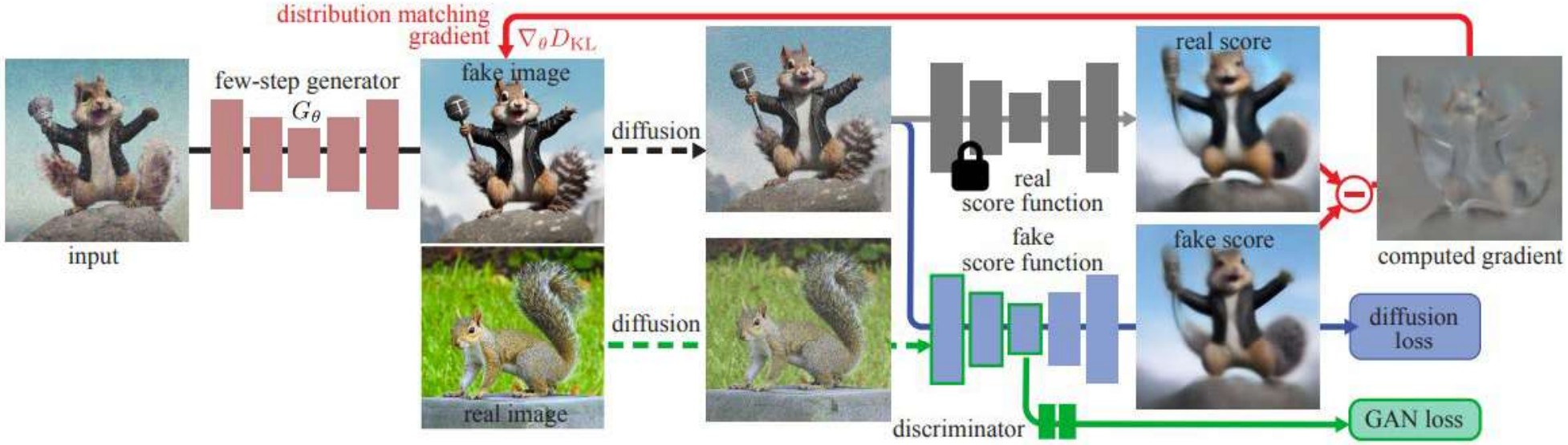


Compression on Pi0.5 with LIBERO

Trained about 15 models this week, of which most promising is listed here:

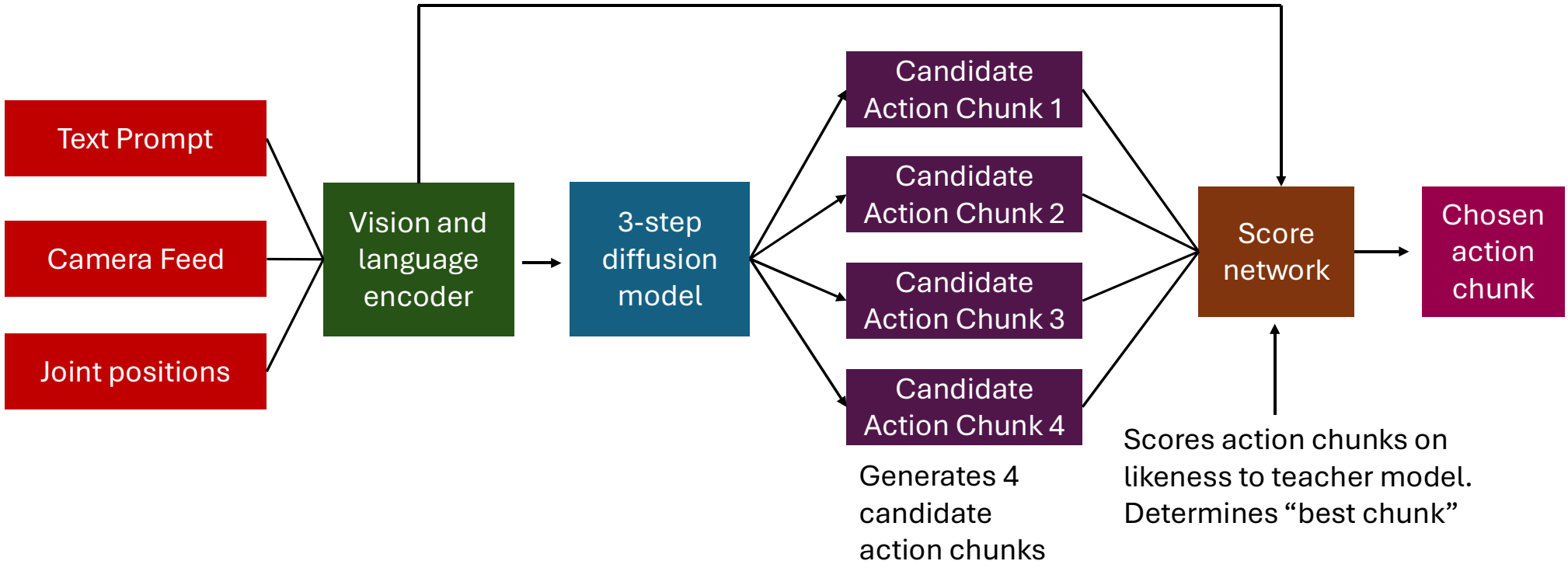


Recap: Distribution Matching Distillation (Training)



From Improved Distribution Matching Distillation for Fast Image Synthesis paper, Tianwei yin et al. MIT, Presented in NeurIPS 2024

Improved Model Inference Architecture



Highlighted Evaluation Results

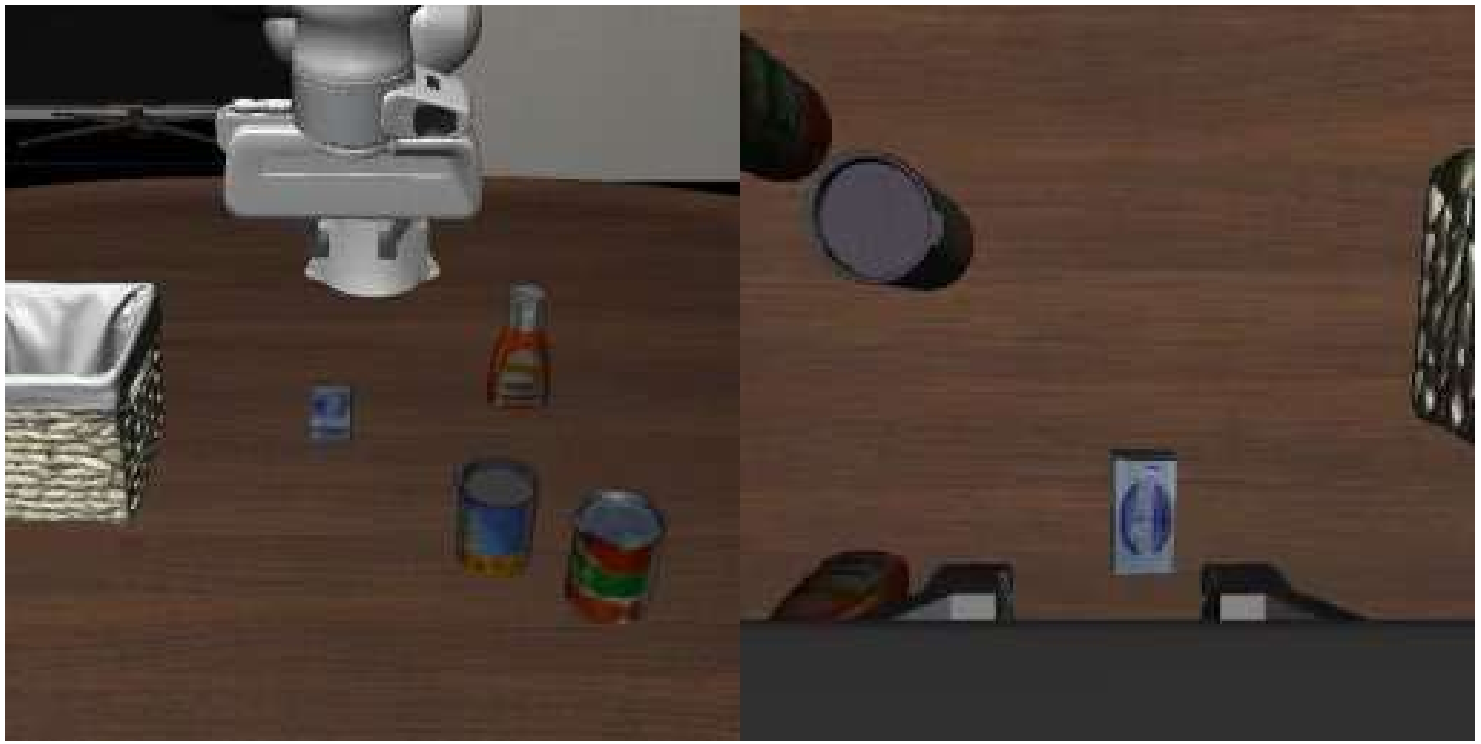
- Does well on specific tasks, fails on others:
  - Commonly failed:
    - “Put the white mug on the plate and put the chocolate pudding to the right of the plate”
  - Commonly passed:
    - “Pick up the book and place it in the back compartment of the caddy”
    - “Put both the cream cheese box and the butter in the basket”

Highlighted Evaluation Results



Prompt: “Put the white mug on the left plate and put the yellow mug on the right plate”

Highlighted Evaluation Results



Prompt: “Put both the alphabet soup and the tomato sauce in the basket”

Model Weaknesses

- Spatial awareness
  - Recognizes the object, but fails to gauge distance and knocks it over
  - Struggles to recognize edges and fails in pick up process
- Language Understanding
  - Gets confused about what to put where
  - Need better understanding of word relationships
- Poor Diffusion process

Next Steps: Improve accuracy