Week 5 Update

Shriyadita Medavaram

Faculty Mentor: Dr. Yuzhang Shang

Project Overview

- Project Title: Efficient and Robust VLA Models for Real-Time Robotic Manipulation
- Issue: Current VLA models are too
  - Slow
  - Computationally expensive
- Goal: Make smaller faster VLA models



Robot Arm

- Issue with dataset
  - Data from wrong arm
  - Joint control
- Resolved dataset issue
  - 240 episodes
  - 4 cans
  - Will retrain model



Video from dataset (not autonomous!)

One-step Diffusion with Distribution Matching Distillation

Tianwei Yin[1] Michaël Gharbi[2] Richard Zhang[2] Eli Shechtman[2]
Frédo Durand[1] William T. Freeman[1] Taesung Park[2]

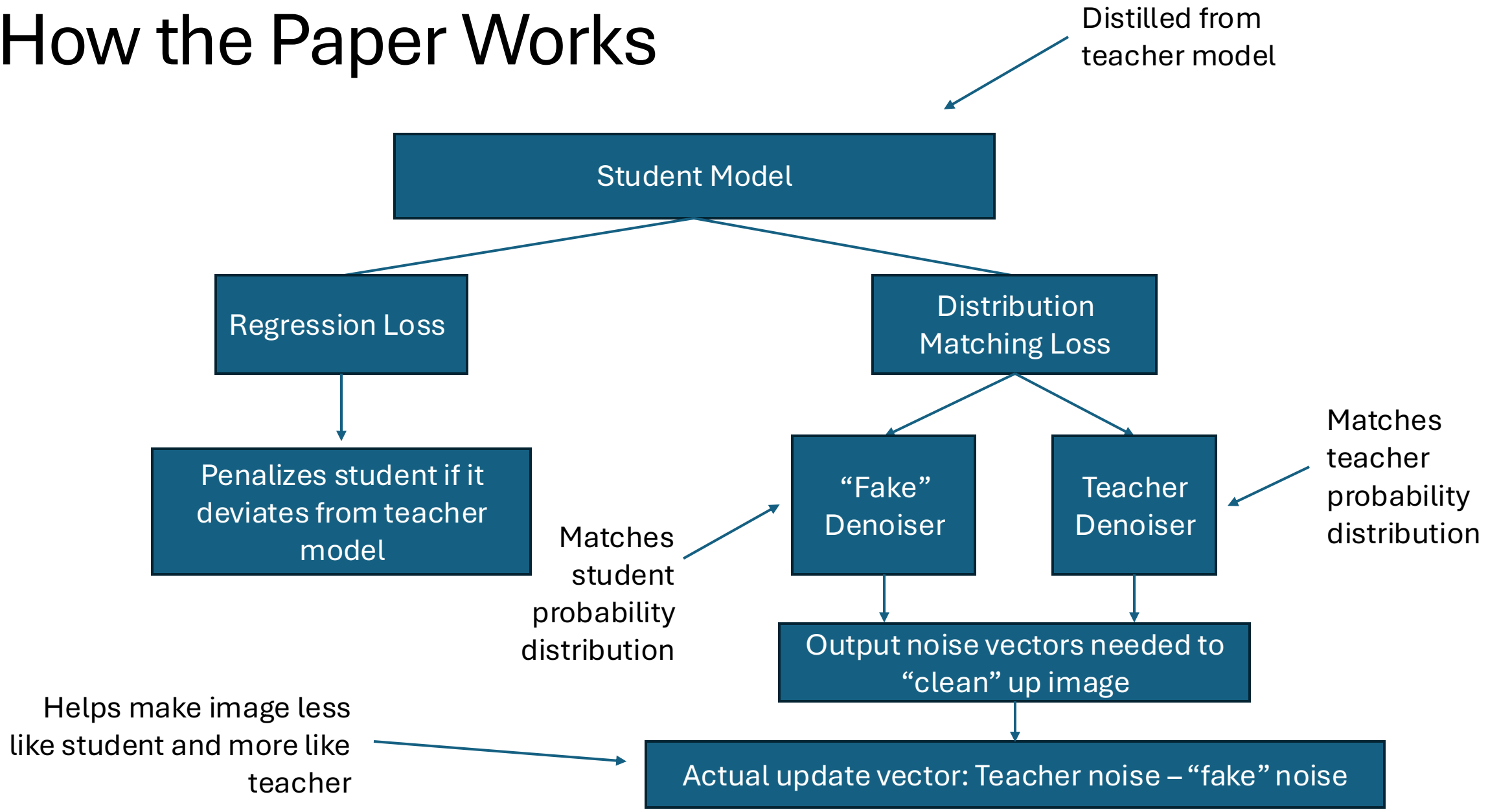
[1]Massachusetts Institute of Technology [2]Adobe Research

https://tianweiy.github.io/dmd/

One Step Diffusion w Distribution Matching Distillation (CVPR-2024)

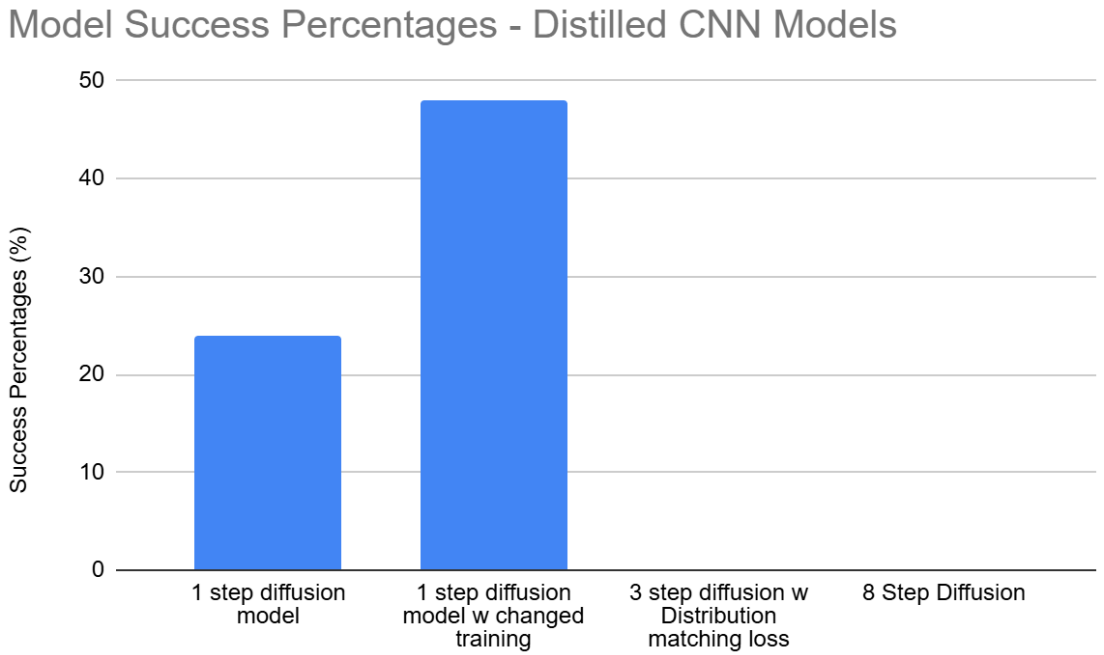
- Compresses image diffusion models
- Single step diffusion (up to 4 steps)
- Applicable to our VLA model as diffusion is major bottleneck

How the Paper Works



Application to VLA (Compression Experiments)

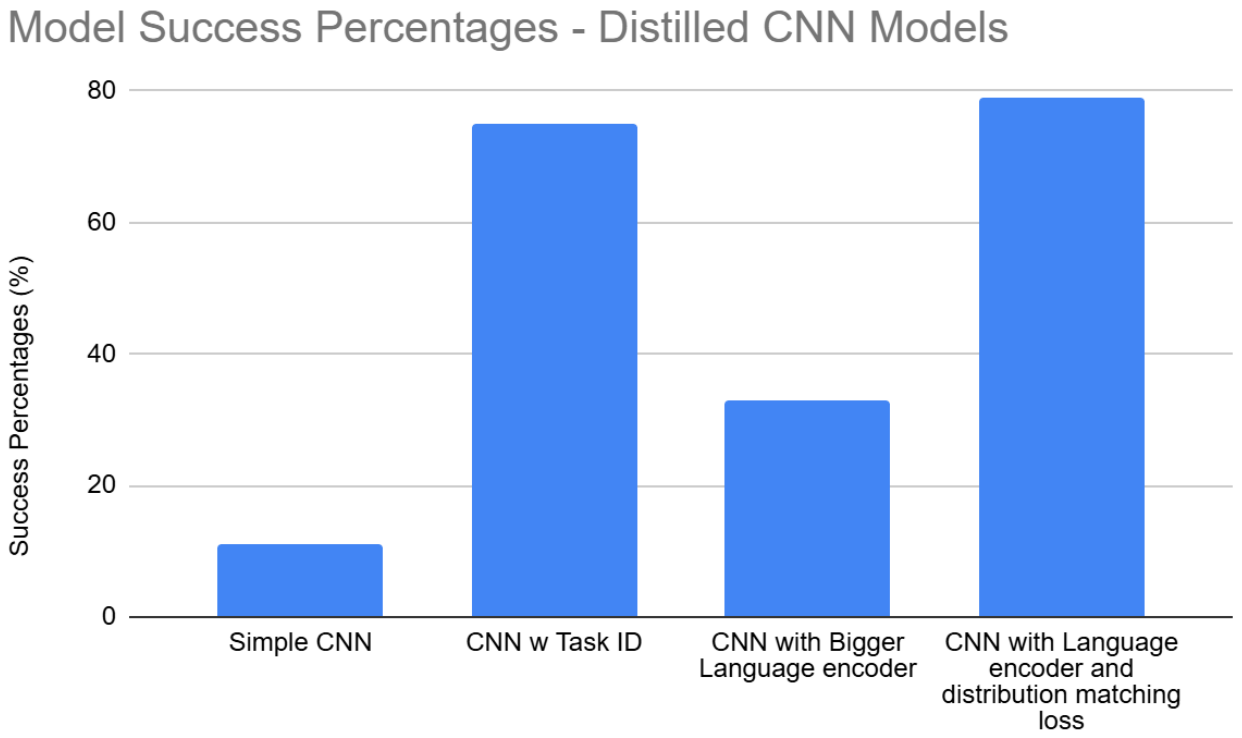
- Trained numerous models and evaluated against LIBERO (benchmark)
- Diffusion Models
  - 1 step diffusion model – 24%
  - 1 step diffusion model w changed training – 48%
  - 3 step diffusion w Distribution matching loss - ~0%
  - 8 Step Diffusion -~0%



* About 10M parameters, Pi0.5 is 3.3 billion (300 times smaller)

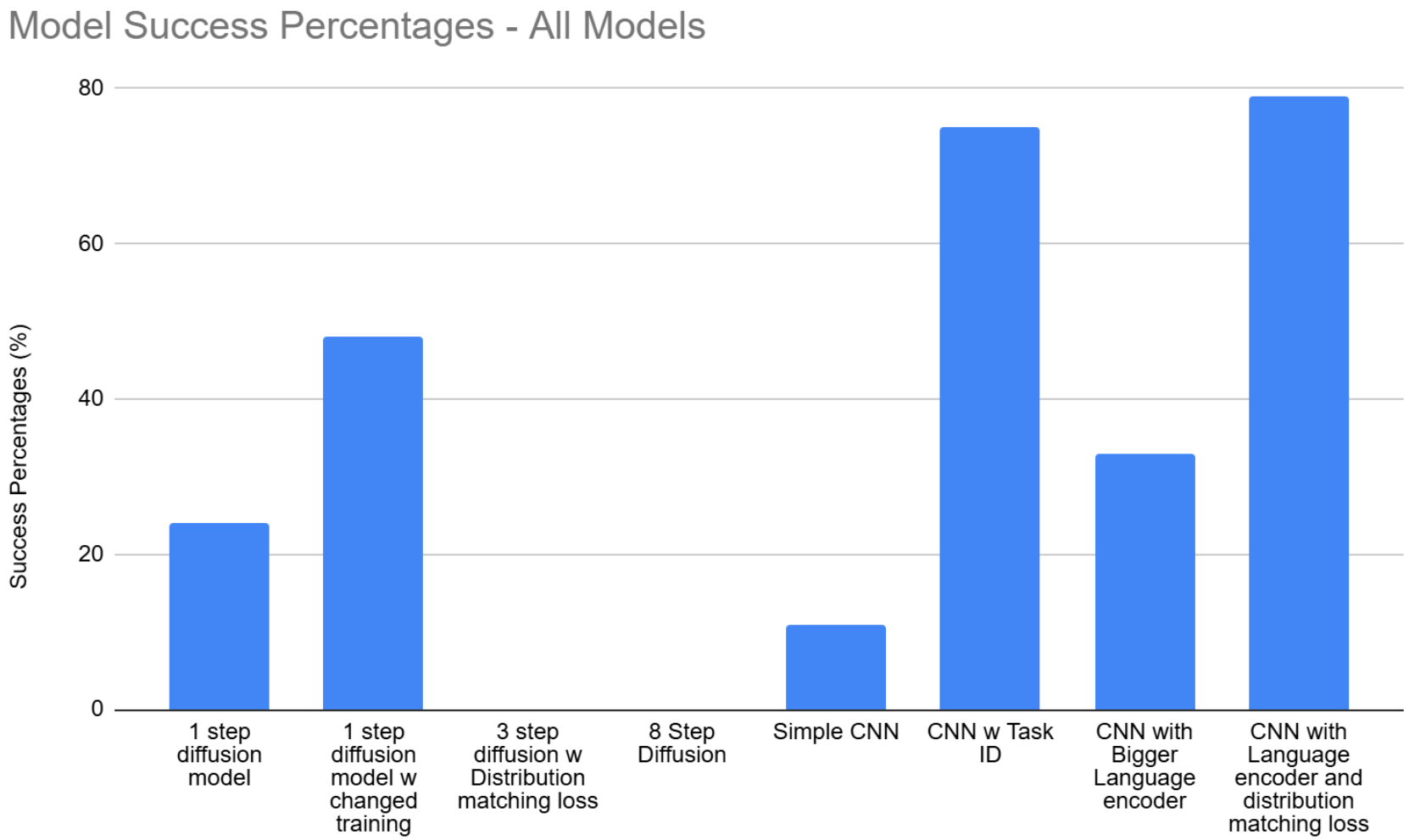
Application to VLA (Compression Experiments)

- CNN Models (Same inputs and encoders)
  - Simple CNN – 11%
  - CNN w Task ID – 75%
  - CNN with larger language encoder – 33%
  - CNN with Language encoder and distribution matching loss – 79%



* About 10M parameters, Pi0.5 is 3.3 billion (300 times smaller)

Overall Performances



Next Steps

- Train Pi0.5 with dataset and evaluate
- Improve compression models