Week 5 Presentation

Shriyadita Medavaram

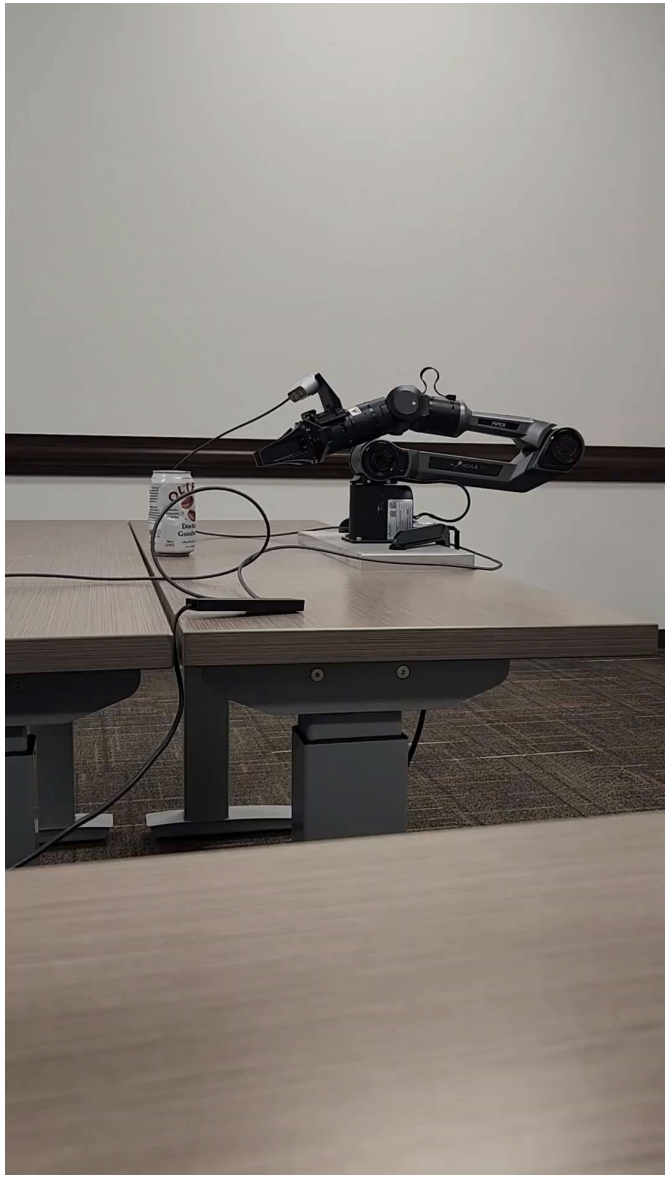
Project Overview

- Efficient and Robust VLA Models for Real Time Robotic Manipulation
  - Problem: Current open source VLA models are
    - Slow
    - Computationally Expensive
  - Goal:
    - Smaller faster VLA model for robots
    - Can run easily on smaller computers
    - Can deal with interference



Schedule

- Week 3: Load Pi-0.5 and Libero on clusters and evaluate
- Week 2-3: Look into Pi-0.5 model and find bottleneck (Run profiles)
- Week 4-6: Run compression experiments
- Week 7: Action-Head Distillation, Full Libero evaluation
- Week 8: Make poster and release code



Real Time Manipulation

- Approaches but does not reach
- Currently on trained checkpoint with 5k steps

Issues With Real Time Manipulation

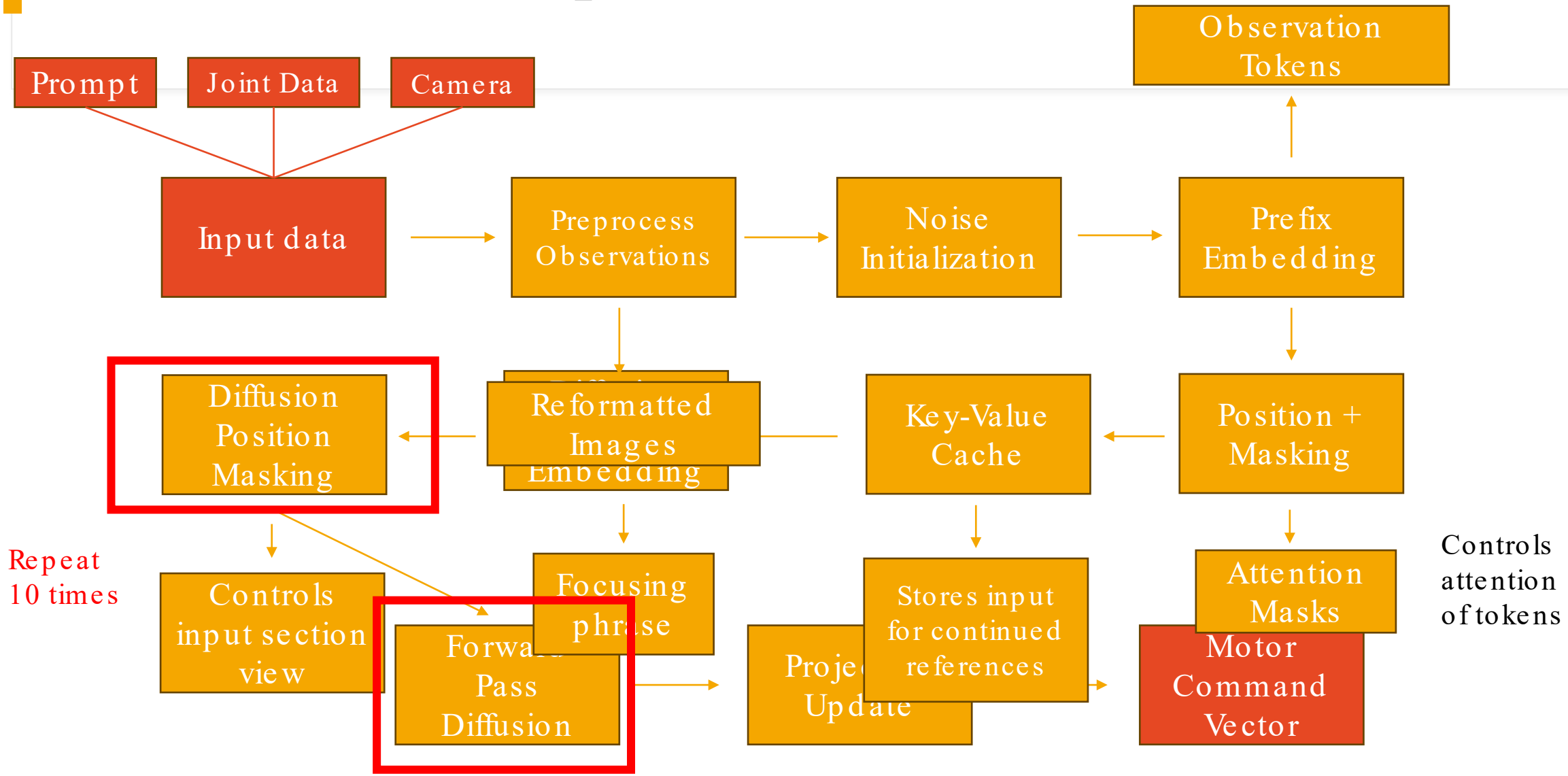
Implemented Fixes

- Units
- Starting position
- Several checkpoints

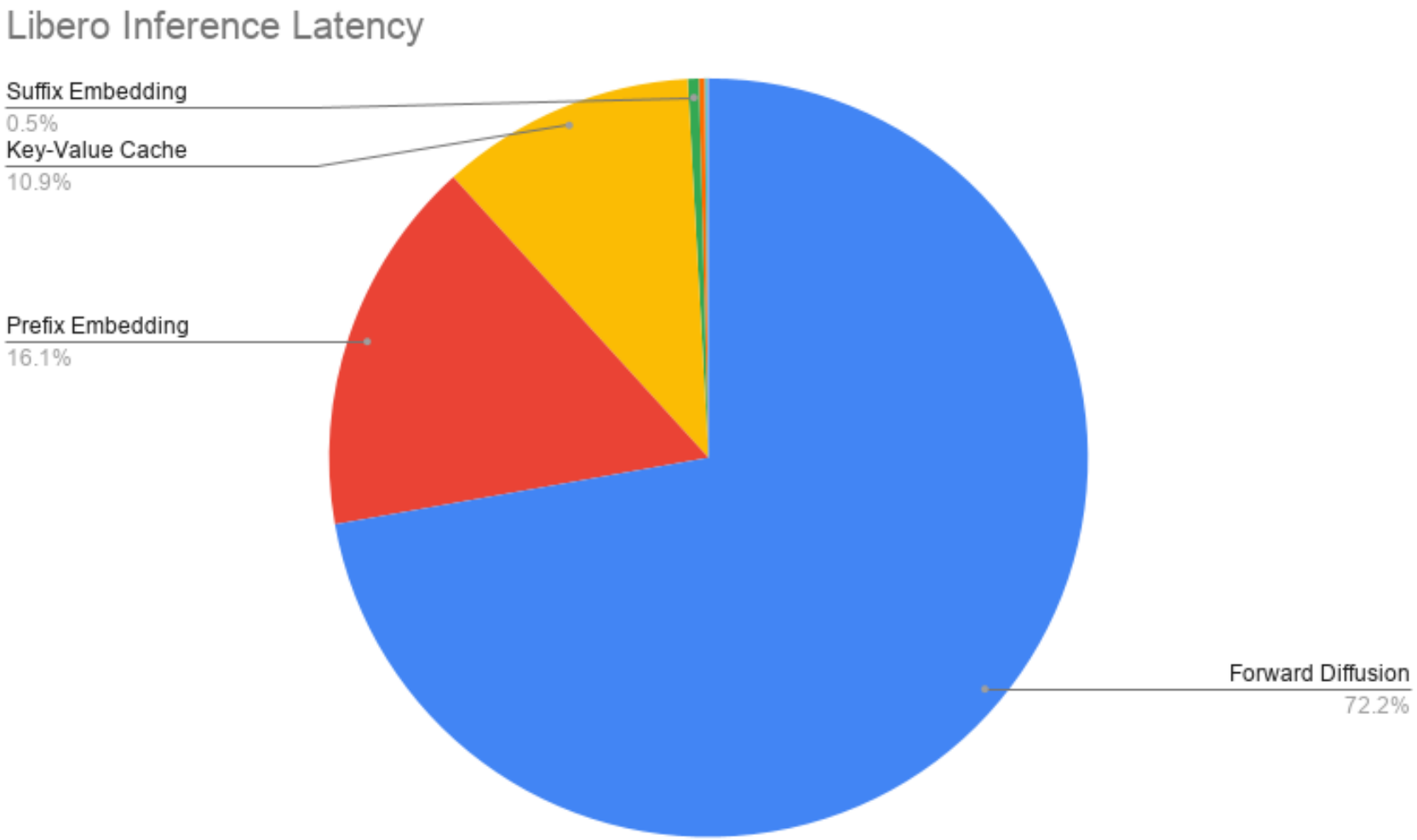
Future Fixes

- Labeling dataset (Success or Failure)
- More data
- Retrain Model

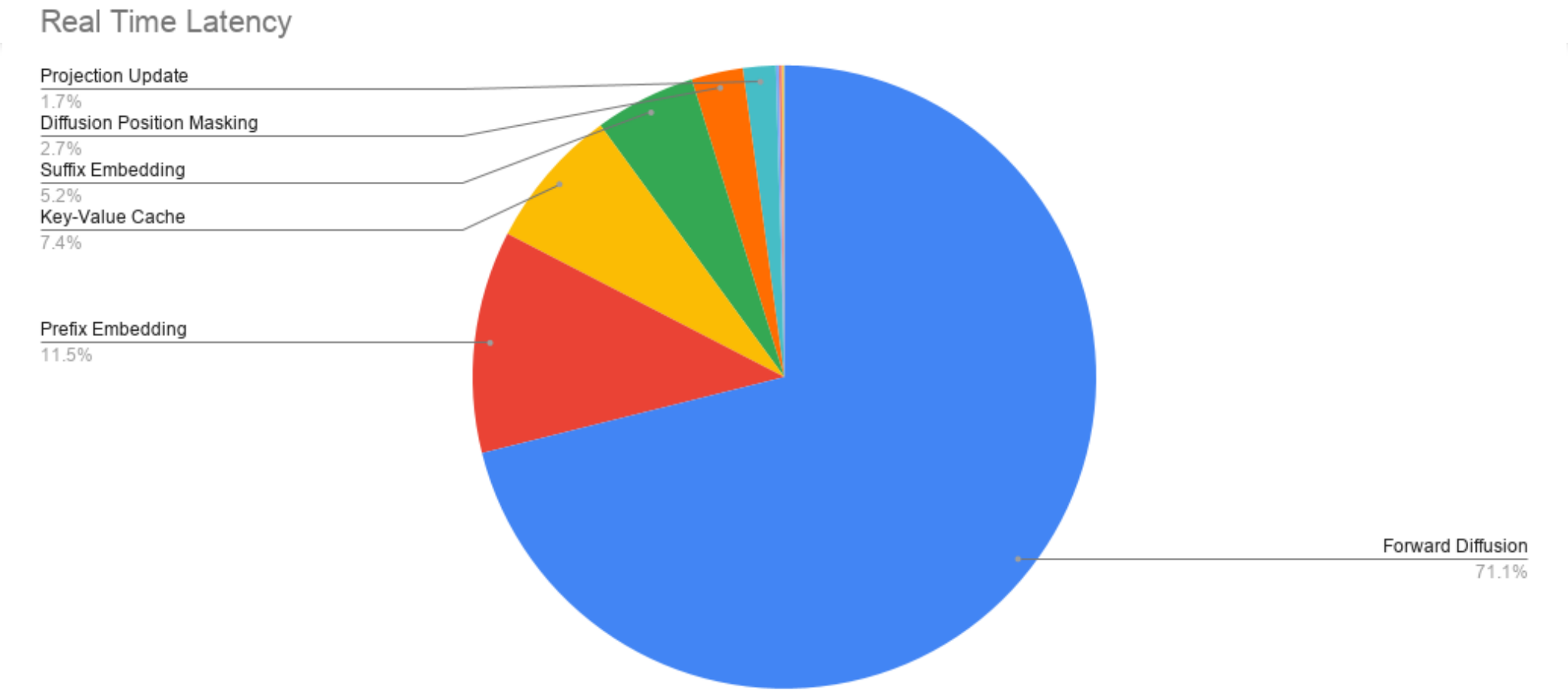
Pi0.5 Model Pipeline



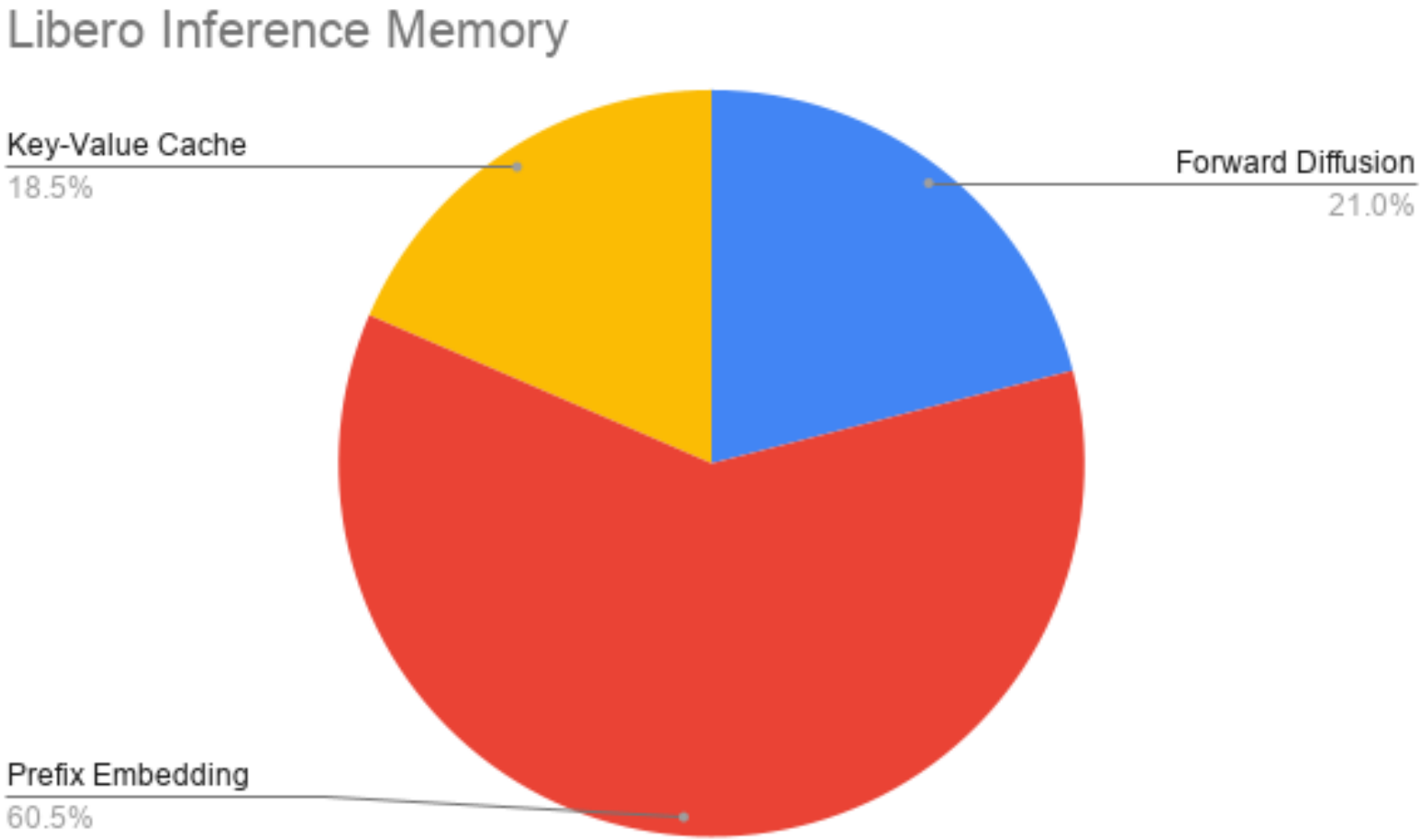
Latency Results – Libero



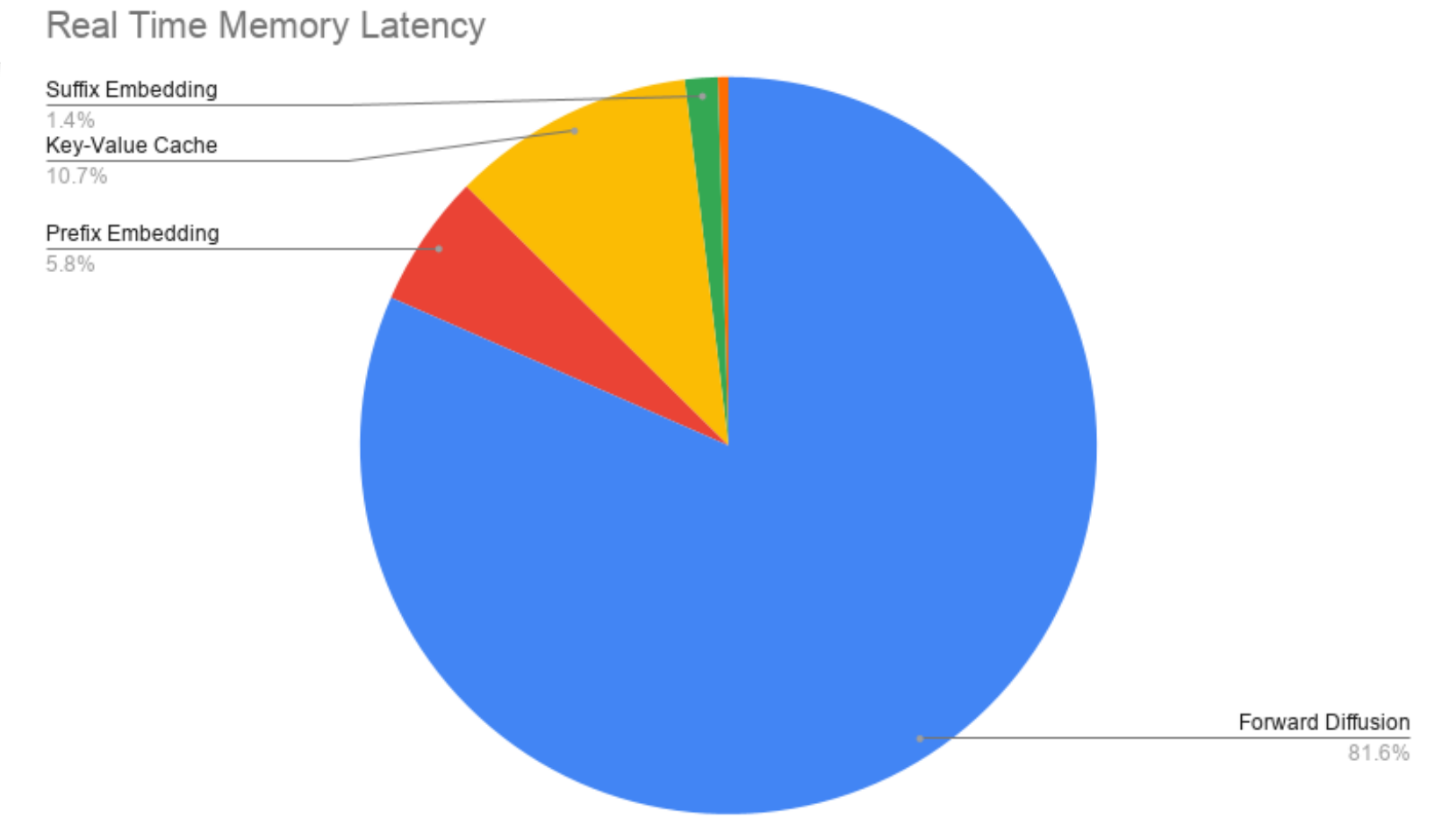
Latency Results – Real Time



Memory Results – Libero



Memory Results – Real Time



Compression Ideas

- Mainly compress forward diffusion process
- Several methods to try (recommended by Dr. Shang)
  - “One Step Diffusion with Distribution Matching Distillation”
  - Improve using methods in “Improved Distribution Matching Distillation for Fast Image Synthesis”
- Will also do further research to find more things to try

Next Steps

- Work on compression with Libero checkpoint
- Label data and re-train
- Collect new data if necessary