Week 9 Presentation

Slow-Fast Weights for Long-Term Video Understanding
Rajat Modi, Dr. Rawat

David Gardner, July 24 2025

Project Overview

Motivation:

- Many long video applications (CCTV, sports)

Limitations:

- High computational cost
- Memory bank compression loss





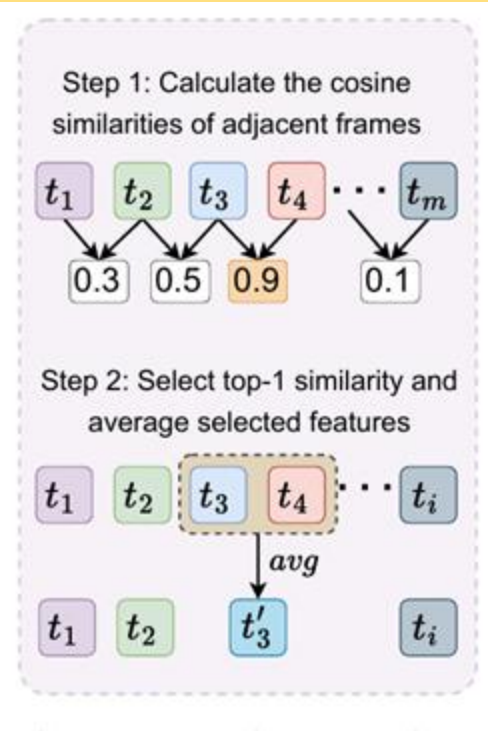
Question: What color is man with the number 7 in the video?

Answer: Red

Project Overview

Our Approach:

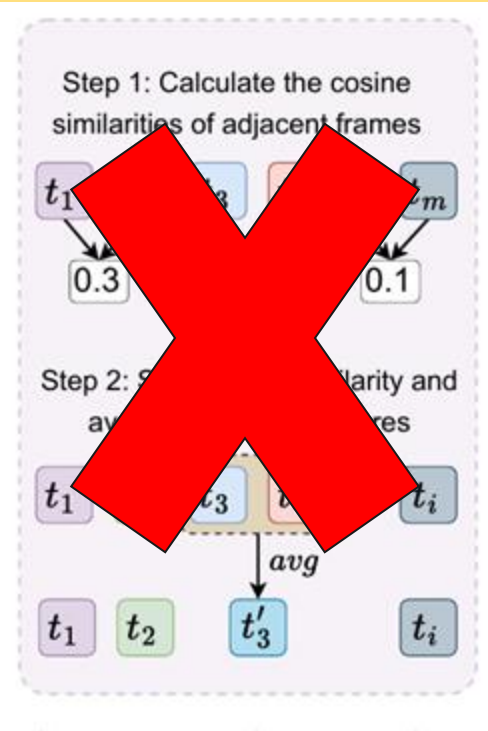
- Store memory in weights
- Remove need for compression rule
- Increase storage capacity



Project Overview

Our Approach:

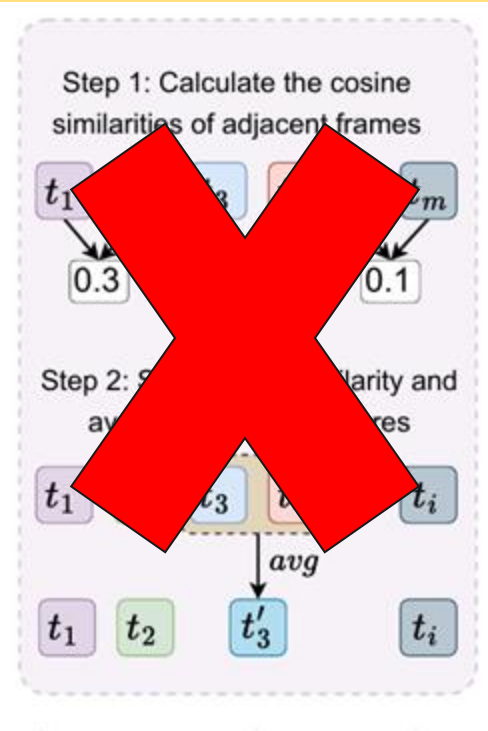
- Store memory in weights
- Remove need for compression rule
- Increase storage capacity



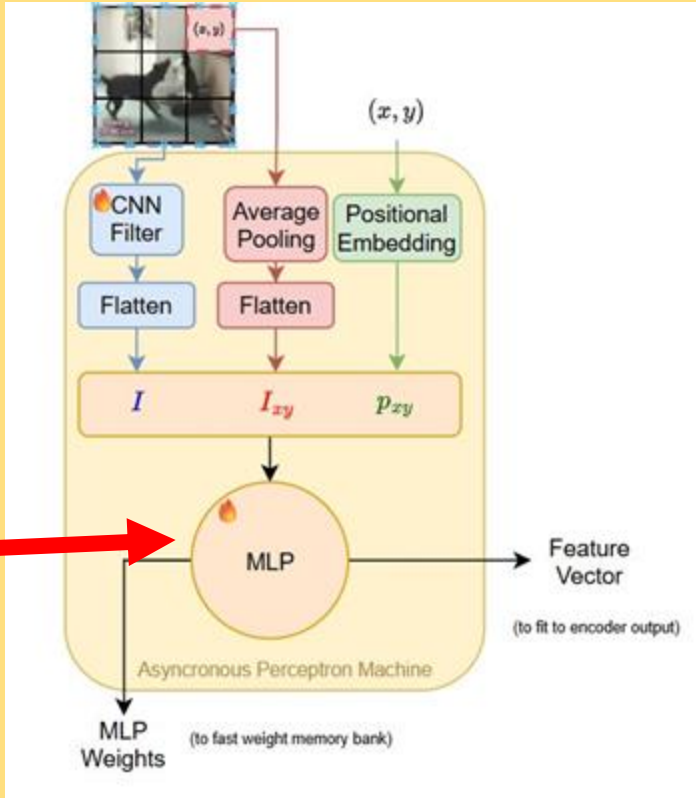
Project Overview

Our Approach:

- Store memory in weights
- Remove need for compression rule
- Increase storage capacity



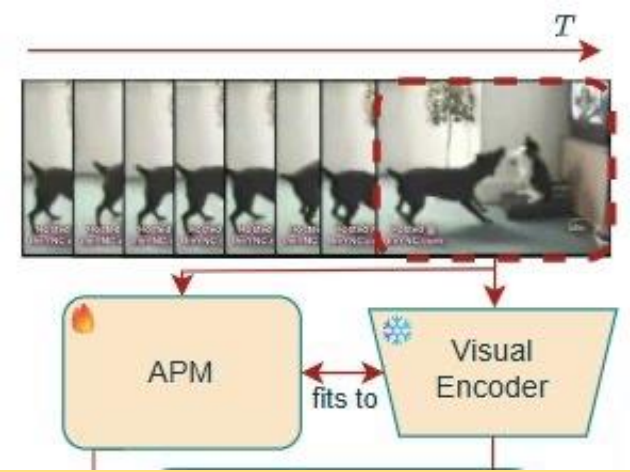
Memory in weights



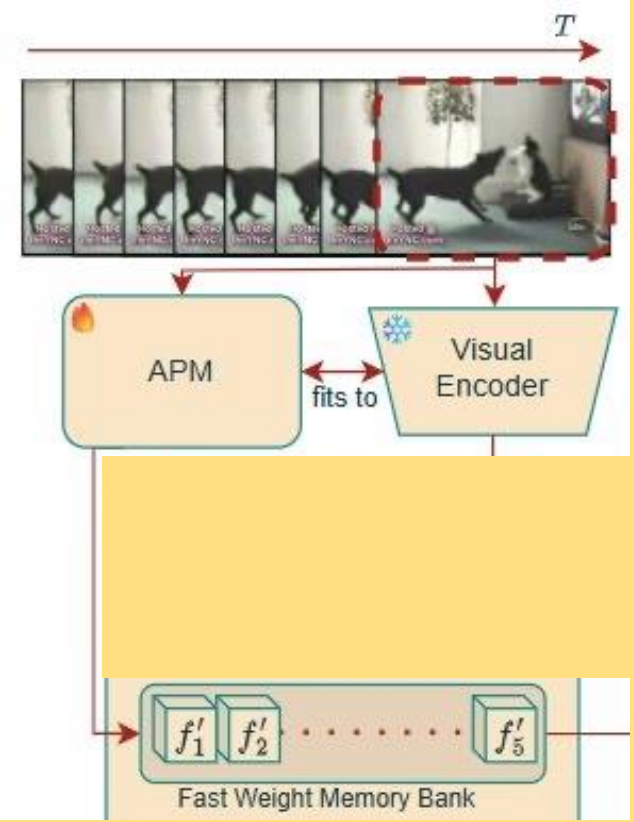
Our Architecture Flow



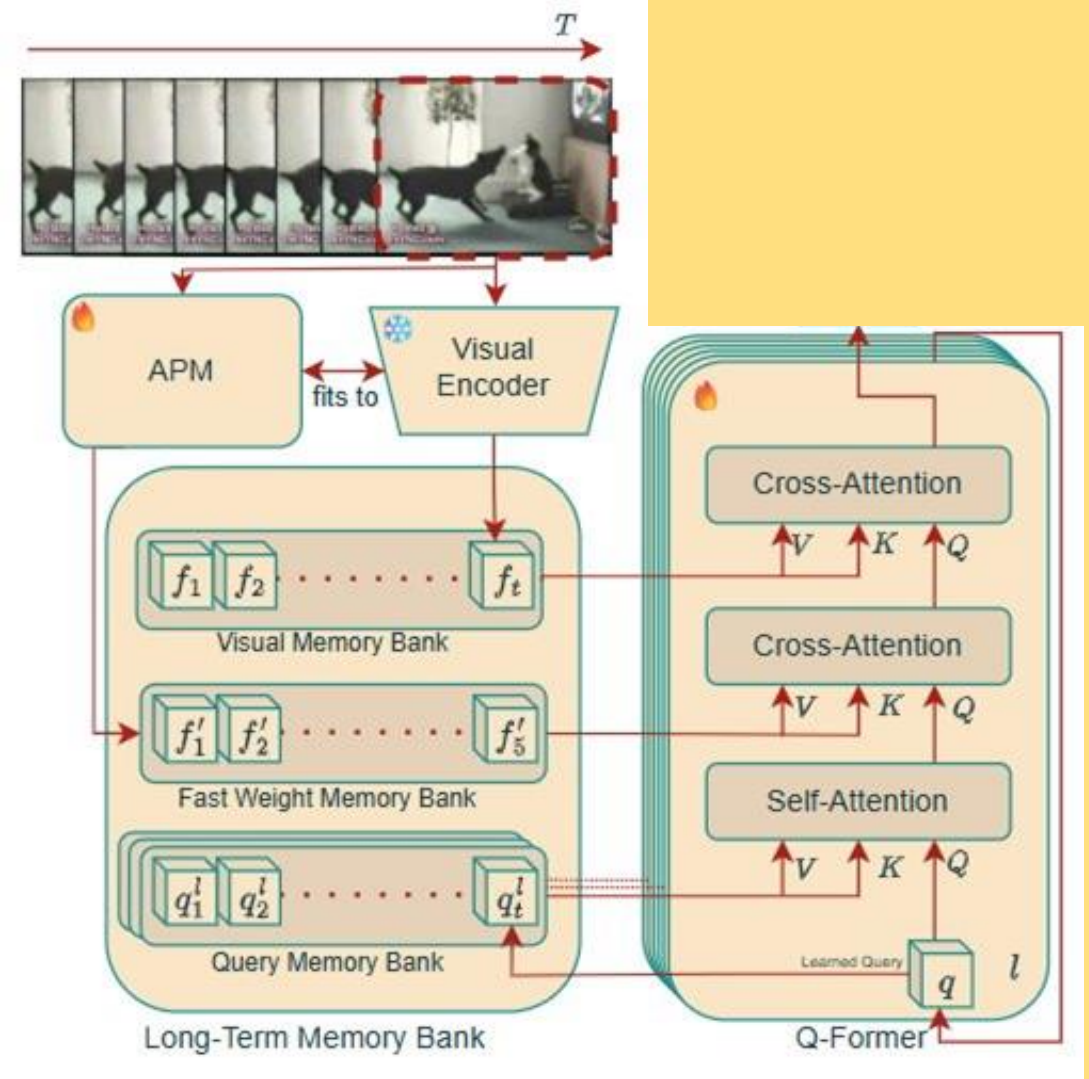
Our Architecture Flow



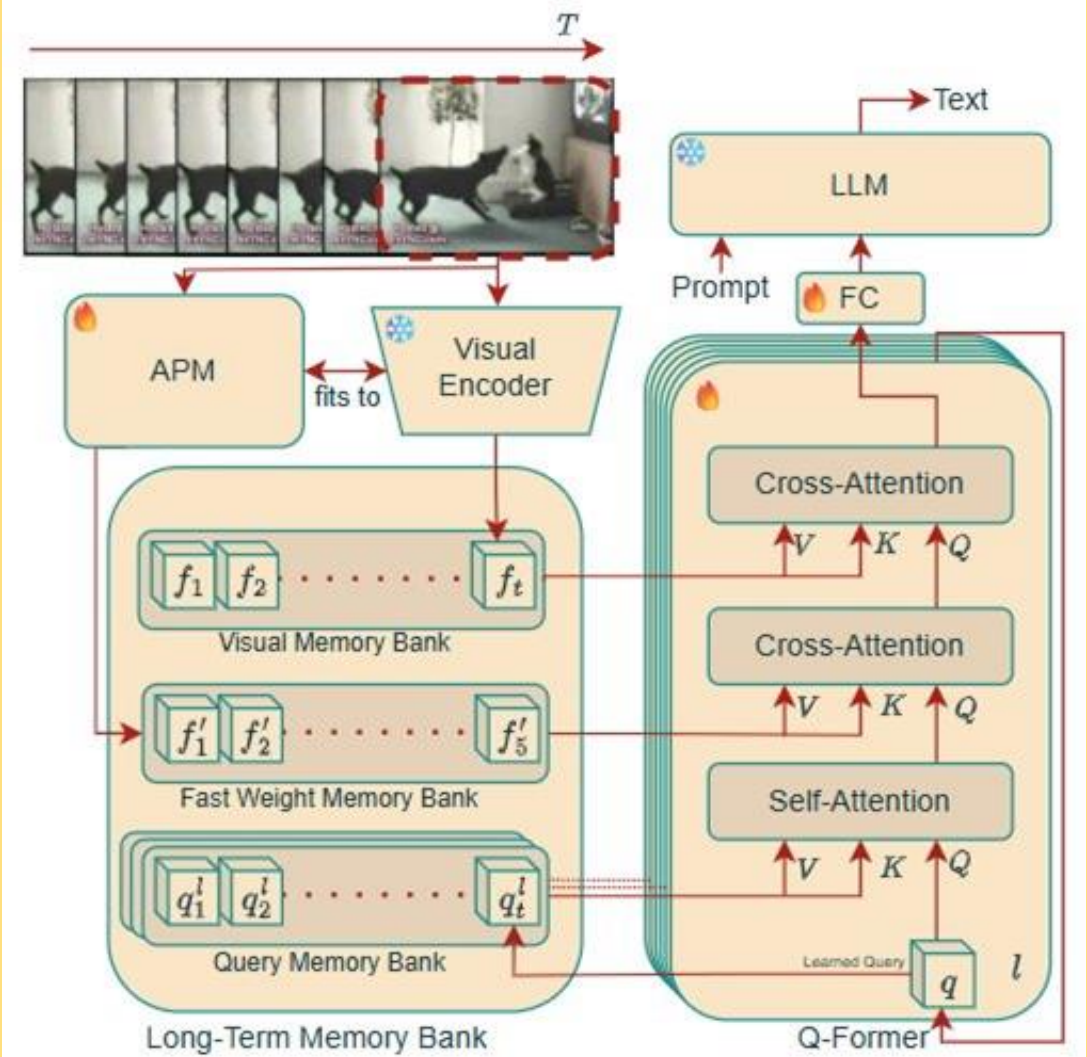
Our Architecture Flow



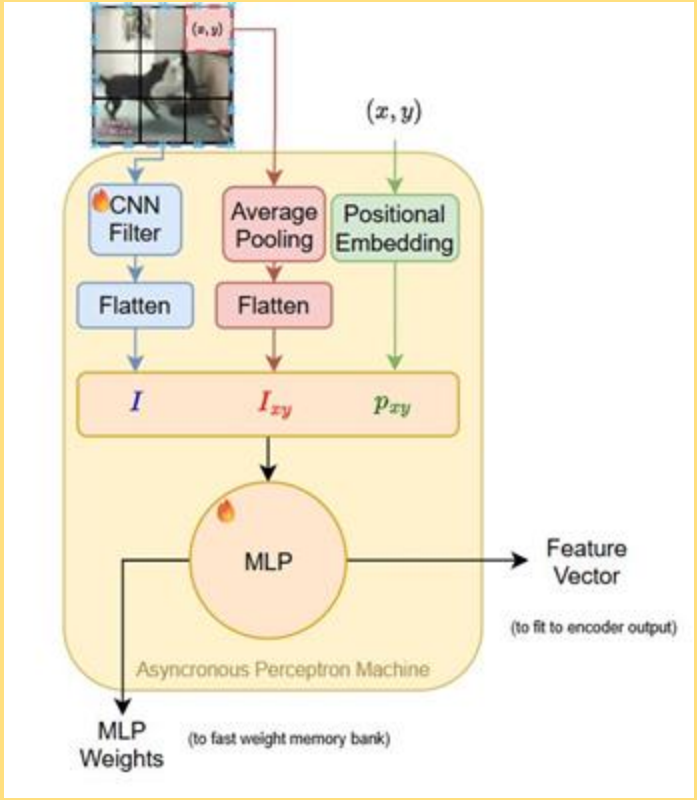
Our Architecture Flow



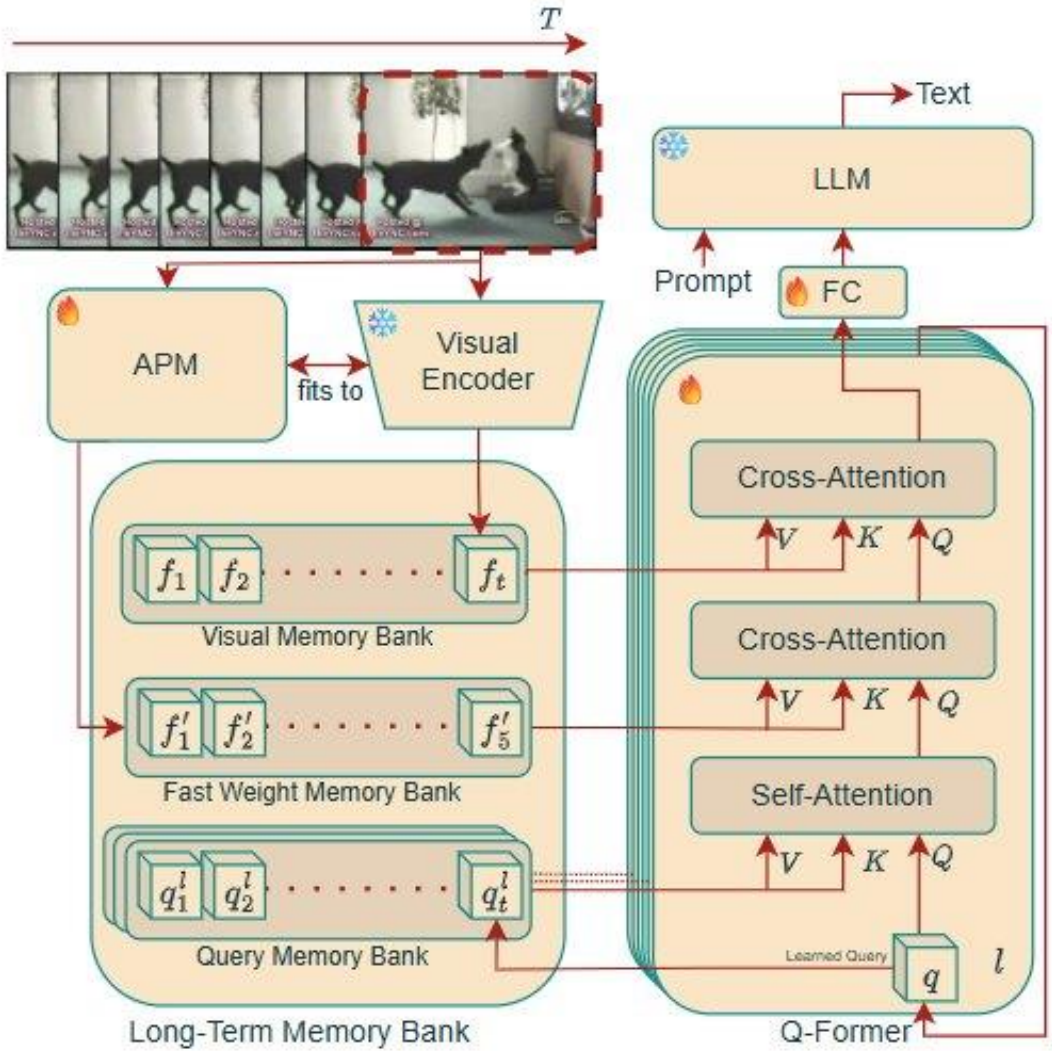
Our Architecture Flow



APM MA-LMM Architecture

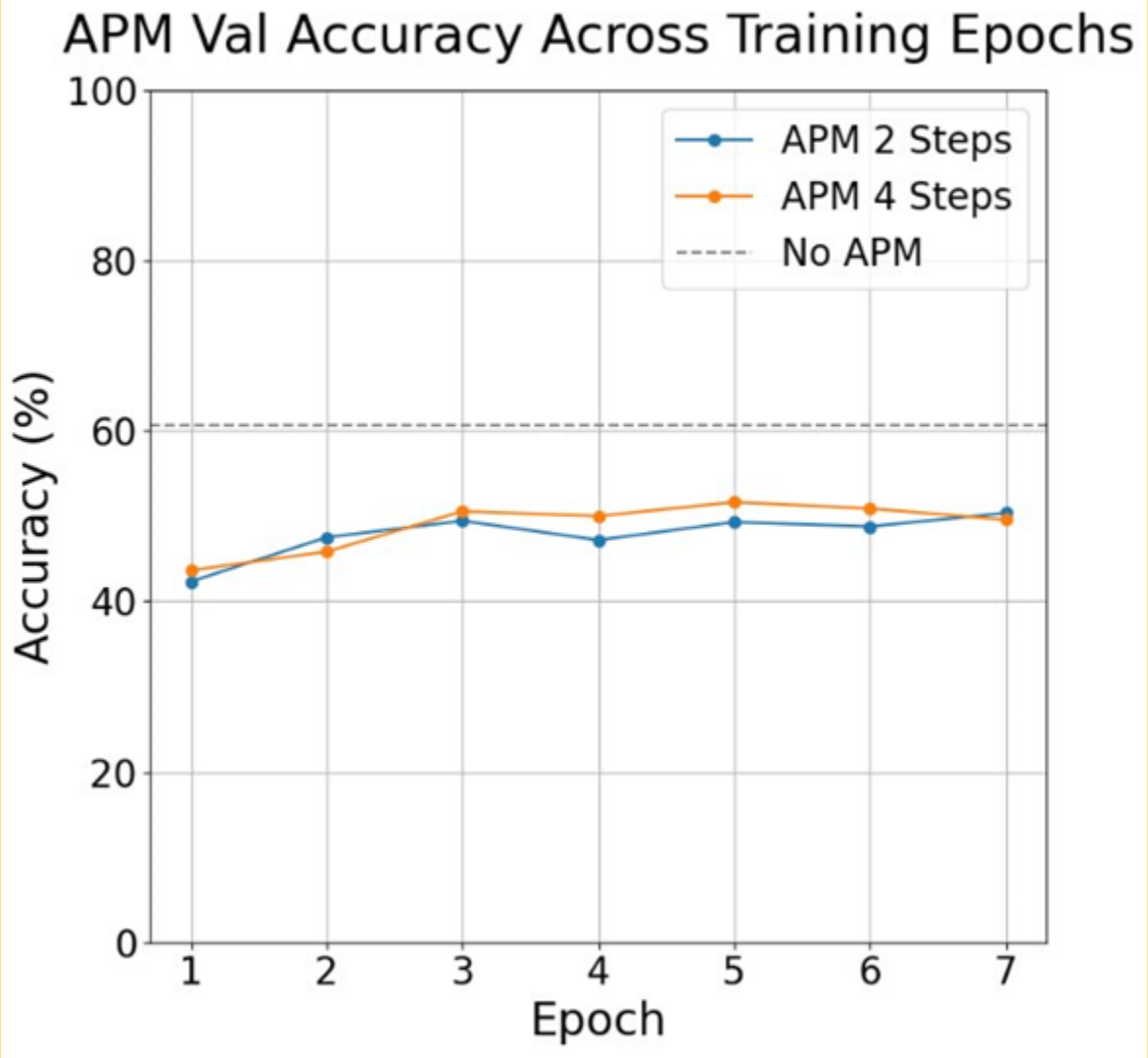


APM



This Week

1. Trained APM cross attention, compared with original
    a. Msvd dataset qa, whole original training dataset
    b. Did 2 and 4 training steps for APM
2. Made poster, started report



1. APM did worse than no APM
2. Not significant difference between 2 and 4 APM steps



Question: how many boys splashed in the little pool?

❌ **Ours:** 2
✅ **MA-LMM:** 3



Question: what did the officer pull a gun on?

✅ **Ours:** car
❌ **MA-LMM:** motorcyclist

Next Steps

1. Run 3 more experiments
    a. Only FWMB, no VMB
    b. Separate Q-former for FAWMB, VMB
    c. Increase FWMB size to 20
2. Finish report

