Week 5 Presentation

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

David Gardner, June 26 2025

Last Week

1. Finished paper reproduction
2. Almost finished implementing APM into MA-LMM video model

CVPR 2024

MA-LMM: Memory-Augmented Large Multimodal Model for Long-Term Video Understanding

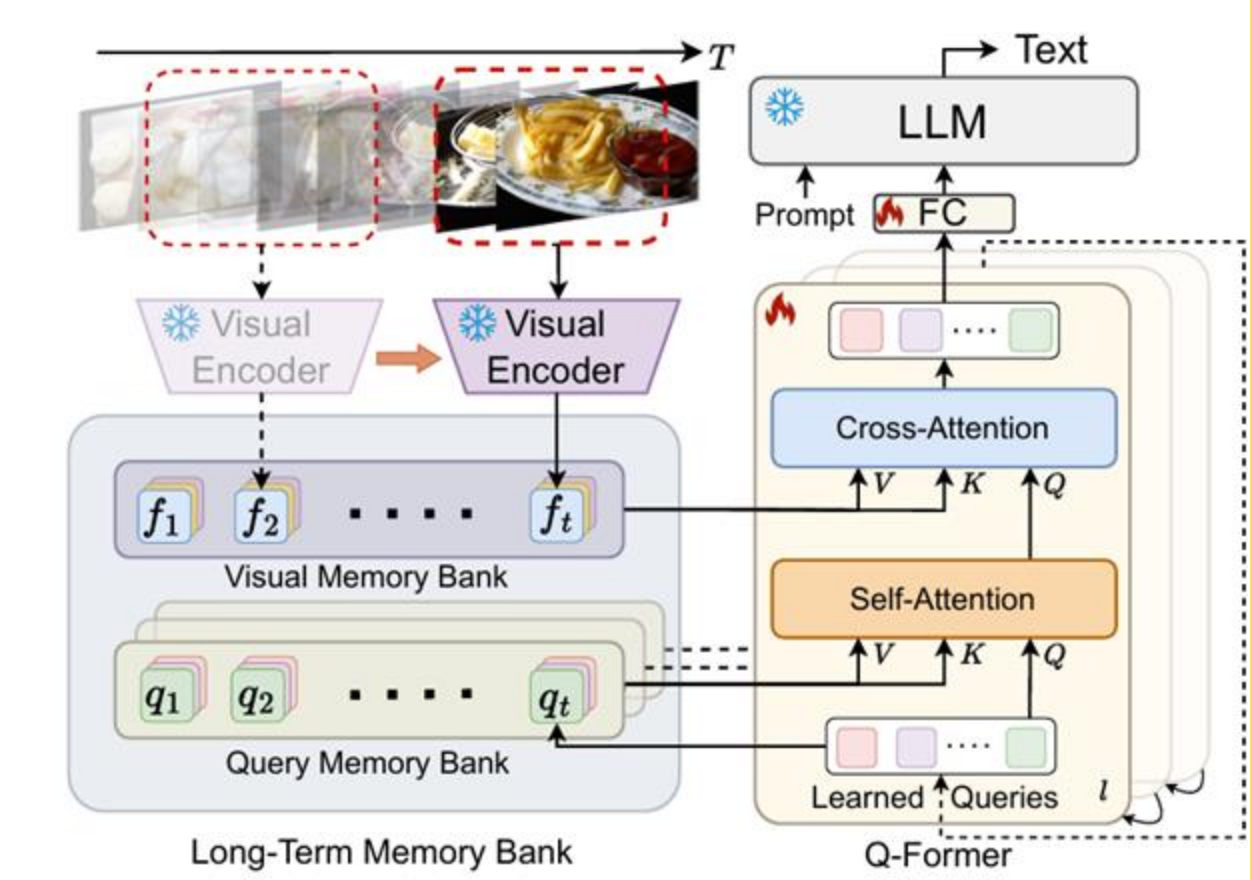
Bo He[1,2*] Hengduo Li[2] Young Kyun Jang[2] Menglin Jia[2] Xuefei Cao[2]
Ashish Shah[2] Abhinav Shrivastava[1] Ser-Nam Lim[3]

- we use this paper as a baseline
- tasks are the same as this paper

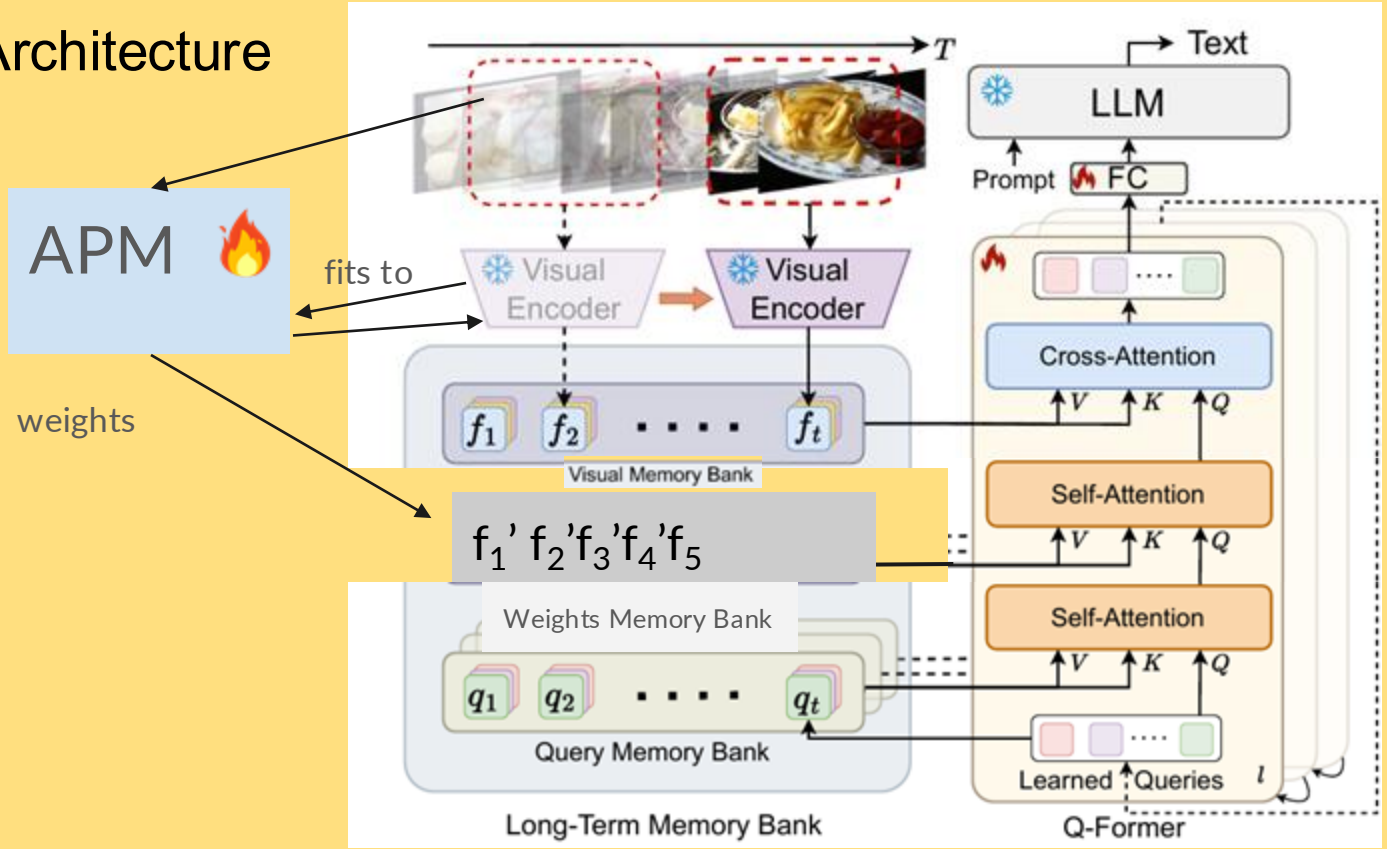
Reproduced Results

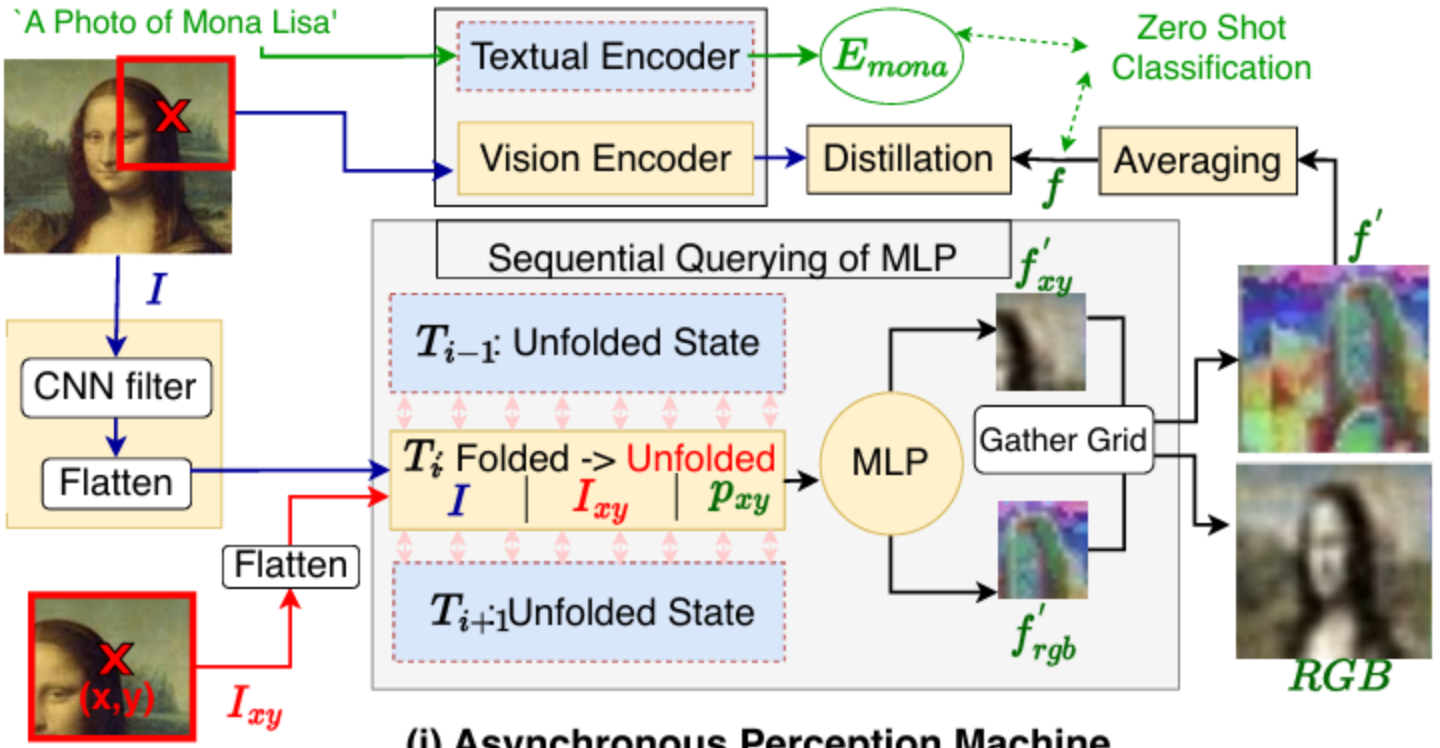
C	D	E	F	G	H	I	J
	Classification	Video question answering	Video question answering	Captioning	Captioning	Captioning	Captioning
	Breakfast	MSVD (Q/A)	MSRVTT (Q/A)	MSRVTT (caption)	MSRVTT (caption)	MSVD (captioning)	MSVD (captioning)
				METEOR	CIDEr	METEOR	CIDEr
Paper results, top 1	93	60.6	48.5	33.4	74.6	51	179.1
Reproduction, top 1 test	92.96	60.56	48.78	33.5	74.7	50.8	178.3
Reproduction, top 1 val	92.96	61.78	48.59	33.6	72.4	50.5	194.5
%error	0.04%	0.07%	0.58%	0.30%	0.13%	0.39%	0.45%

MA LMM Architecture



Learned Memory Architecture





(i) Asynchronous Perception Machine