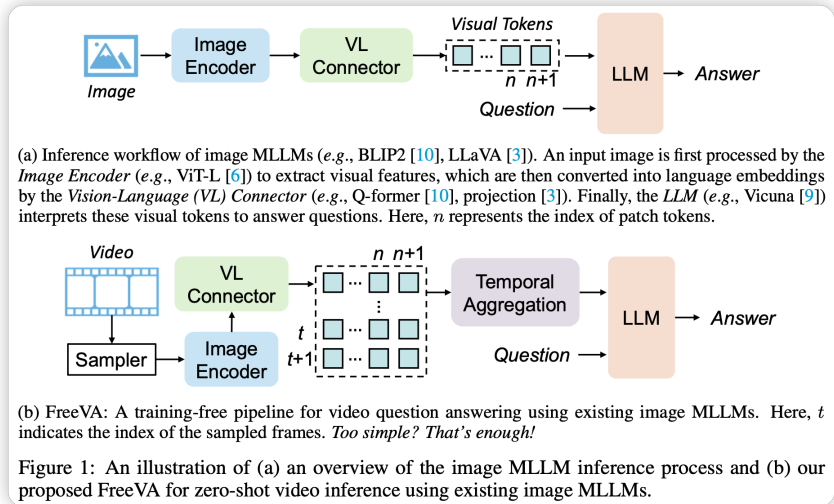




home
papers
study
life
About me

FreeVA

- FreeVA: Offline MLLM as Training-Free Video Assistant
- <https://arxiv.org/abs/2405.07798>
- arxiv



training-free MLLM zero-shot
MLLM MLLM
VideoInstruct-100k LLaVA LLaVA
MLLM MLLM
zero-shot GPT API
MLLM
embedding
embedding
token token token token
VideoChatGPT LLaVA

Table 7: Comparison with SOTA methods on zero-shot video QA datasets. We recognize that the default version of GPT-3.5-Turbo periodically, significantly impacting performance evaluation. Thus, we also report the possible GPT-3.5 versions used for evaluation. “MAR” denotes GPT-3.5-Turbo-0301 version, “JUN” denotes GPT-3.5-Turbo-0613 version, and “JAN” denotes the latest GPT-3.5-Turbo-0125 version. Unless specific, we use * to signify the likely version based on the paper’s timing. † denotes the results reported in [11]. To ensure fair comparison and support the growth of the community, we hope future work will specify the GPT-3.5 version.

Method	LLM Size	GPT-3.5 Version	Video-Text Train Data	Train Free	MSVD-QA Acc	MSVD-QA Score	MSRVTT-QA Acc	MSRVTT-QA Score	ANet-QA Acc	ANet-QA Score
FrozenBiLM [41]†	7B	MAR*	10M	✗	32.2	—	16.8	—	24.7	—
Video-LLaMA [36]†	7B	MAR*	2M+11K	✗	51.6	2.5	29.6	1.8	12.4	1.1
VideoChat [35]†	7B	MAR*	10M+11K	✗	56.3	2.8	45.0	2.5	26.5	2.2
Video-ChatGPT [11]†	7B	MAR*	100K	✗	64.9	3.3	49.3	2.8	35.2	2.7
Valley [13]	7B	MAR*	702K+73K	✗	65.4	3.4	45.7	2.5	42.9	3.0
VaQuitA [14]	7B	MAR	100K	✗	74.6	3.7	68.6	3.3	48.8	3.3
Chat-UniVi [42]	7B	JUN*	100K	✗	65.0	3.6	54.6	3.1	45.8	3.2
BT-Adapter [43]	7B	JUN*	100K	✗	67.5	3.7	57.0	3.2	45.7	3.2
LLaMA-VID [12]	7B	JUN*	232K+107K	✗	69.7	3.7	57.7	3.2	47.4	3.3
LLaMA-VID [12]	13B	JUN*	232K+107K	✗	70.0	3.7	58.9	3.3	47.5	3.3
Video-LLaVA [37]	7B	JUN*	702K+100K	✗	70.7	3.9	59.2	3.5	45.3	3.3
MovieChat [38]	7B	JAN*	2M+11K	✗	75.2	3.8	52.7	2.6	45.7	3.4
FreeVA w/ LLaVA-1.5[34]	7B	MAR	N/A	✓	81.5	4.0	72.9	3.5	58.3	3.5
	13B				82.4	4.0	73.6	3.6	59.1	3.6
FreeVA w/ LLaVA-1.5[34]	7B	JUN	N/A	✓	71.2	3.8	58.5	3.2	53.7	3.5
	13B				71.8	3.8	59.2	3.3	54.5	3.5
FreeVA w/ LLaVA-1.5[34]	7B	JAN	N/A	✓	73.8	4.1	60.0	3.5	51.2	3.5
	13B				74.4	4.1	61.1	3.6	51.6	3.5
FreeVA w/ InstructBLIP[5]	13B	JAN	N/A	✓	57.8	3.5	34.6	2.7	23.0	2.4
FreeVA w/ InternVL[39]	20B	JAN	N/A	✓	73.8	4.0	54.8	3.3	58.4	3.6
FreeVA w/ DenseConnector[40]	13B	JAN	N/A	✓	75.1	4.1	60.8	3.5	52.6	3.5
	34B				77.4	4.2	62.1	3.6	55.8	3.6

-
- MSVD-QA, ActivityNet-QA, MSRVTT-QA VideoChatGPT
- 1 A100(40G)
- <https://github.com/whwu95/FreeVA>

Newer

Older

2025-01-08
IG-VLM

2025-01-06
VTimeLLM

leicheng © 2022-2025

[Archive](#) [RSS feed](#) [GitHub](#) [Email](#) [QR Code](#)

Made with [Montaigne](#) and [bigmission](#) 