

The Labyrinth of Links: Navigating the Associative Maze of Multi-modal LLMs

Hong Li, Nanxi Li, Yuanjie Chen, Jianbin Zhu, Qinlu Guo, Cewu Lu, Yong-Lu Li



**SHANGHAI JIAO TONG
UNIVERSITY**



RHOS 
mvig-rhos.com

What is the Association?



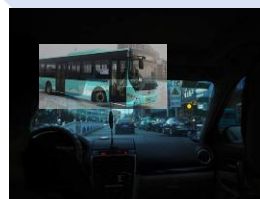
7 am:



Get off the plane



9 am:



Commuting by bus

Noon:



Observation

Afternoon Chat

Memory



Association: The **bus** we saw? Or the **airplane** this morning?

Deduction: They are all **plentiful**?

Question

Look over there, isn't the decoration more plentiful than what we see today?

Motivation: How about MLLMs in the Human-like Association?



RHOS
mvig-rhos.com



ICLR

Association Task

Association step

(Which candidate may be linked with the query?)



Instruction: {memory}. {Association Question Instruction}

Query image



Image1



Image2



Image2. (✓ asso_step += 1)

Deduction step

(What is the labyrinth of association? Same attribute?)



Instruction: {Evidence Question Instruction}

Image1



Image2



Metal.

Association step



Instruction: {memory}. {Association Question Instruction}

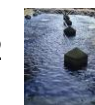
Query image



Image1



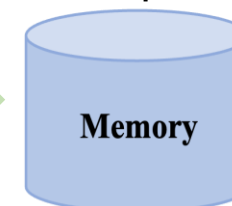
Image2



...

Sequential association until error

Update
memory

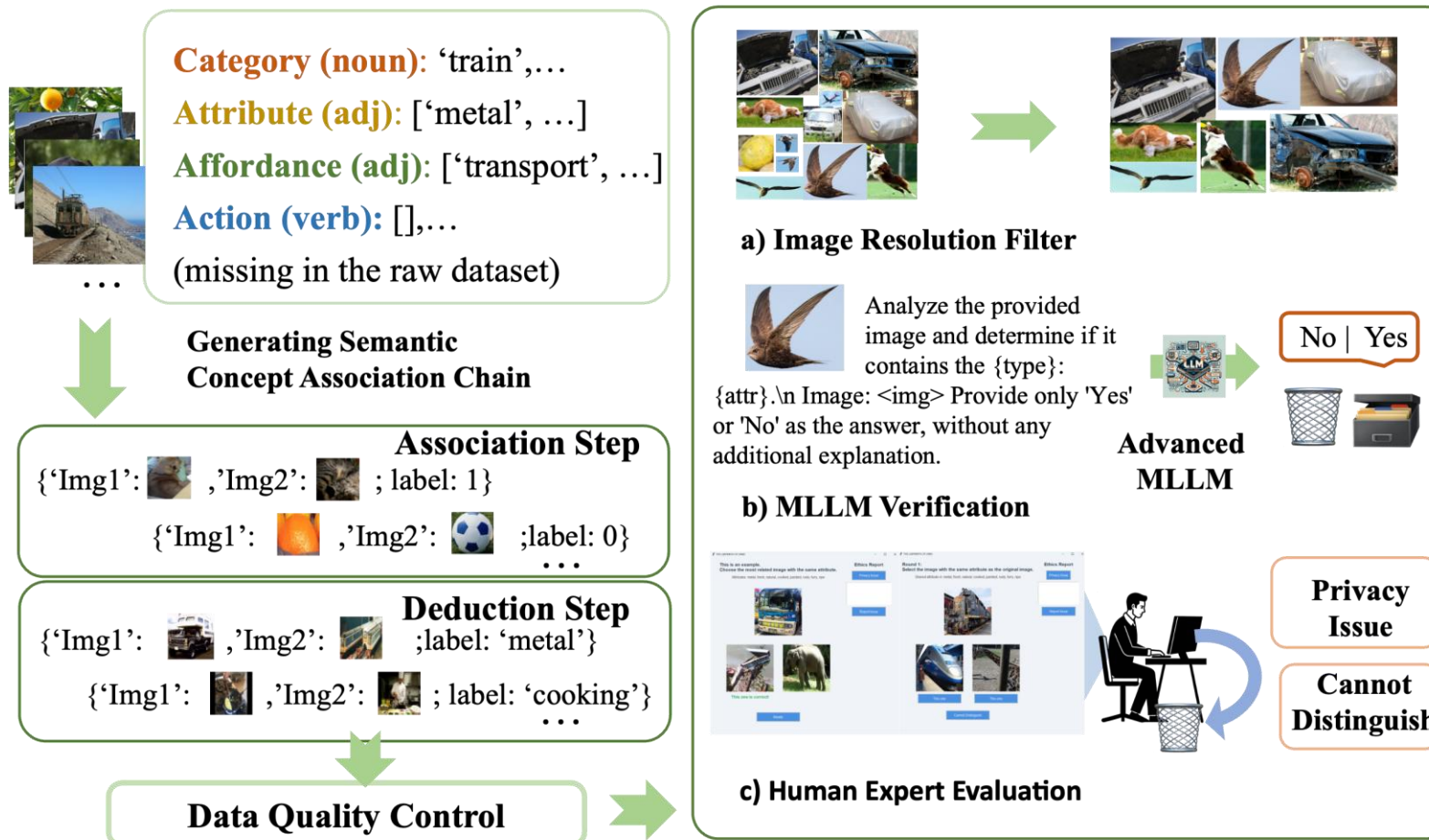


Current memory: rusty

Metal: ↑ Rusty: ↓

Updated memory: rusty; metal

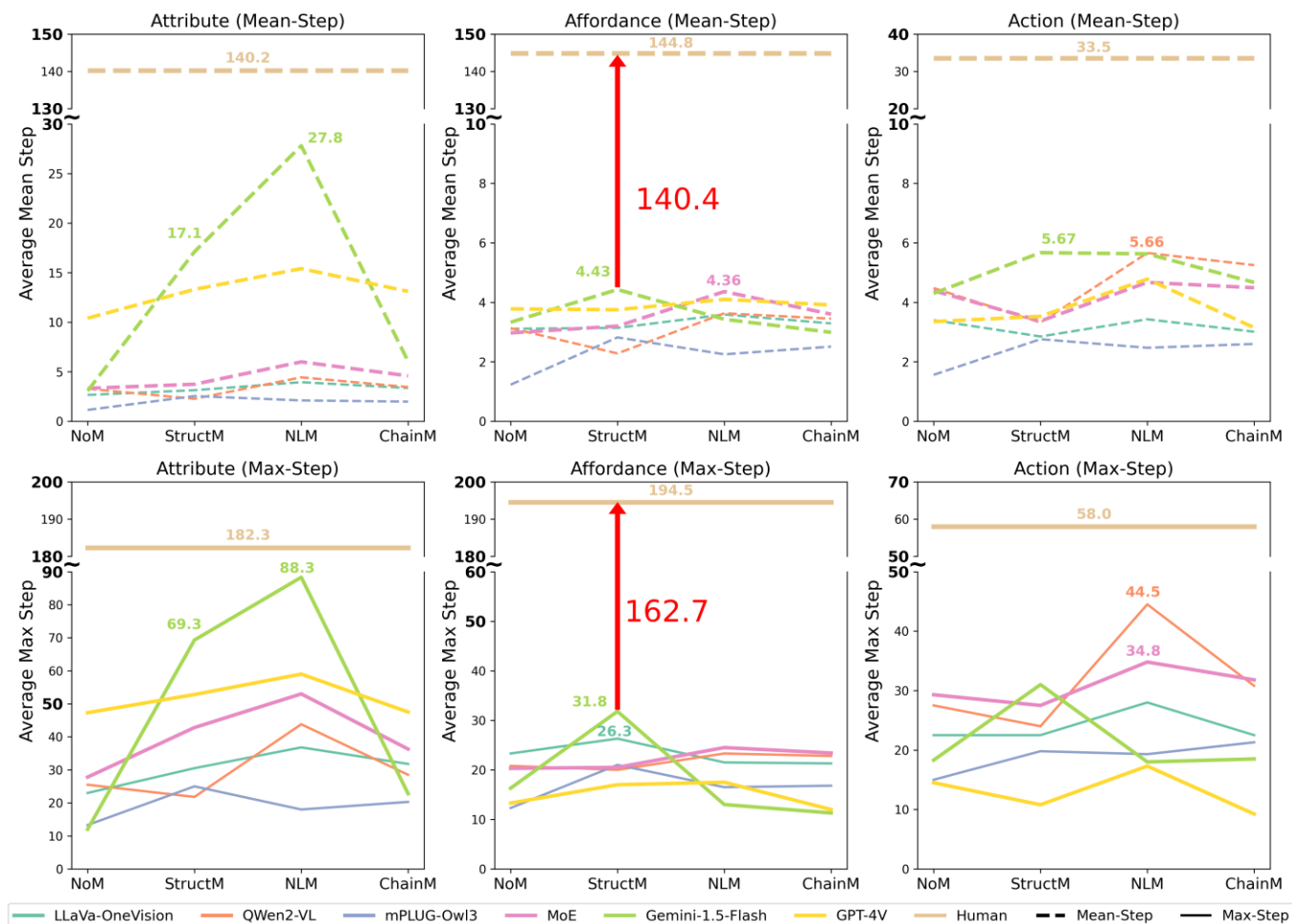
Data Curation and Quality Control



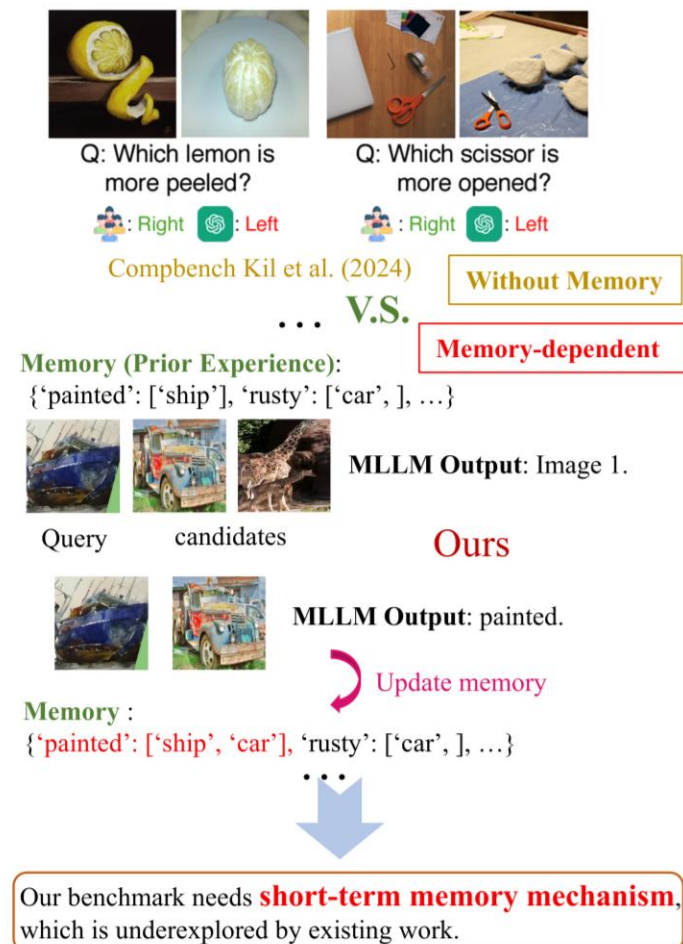
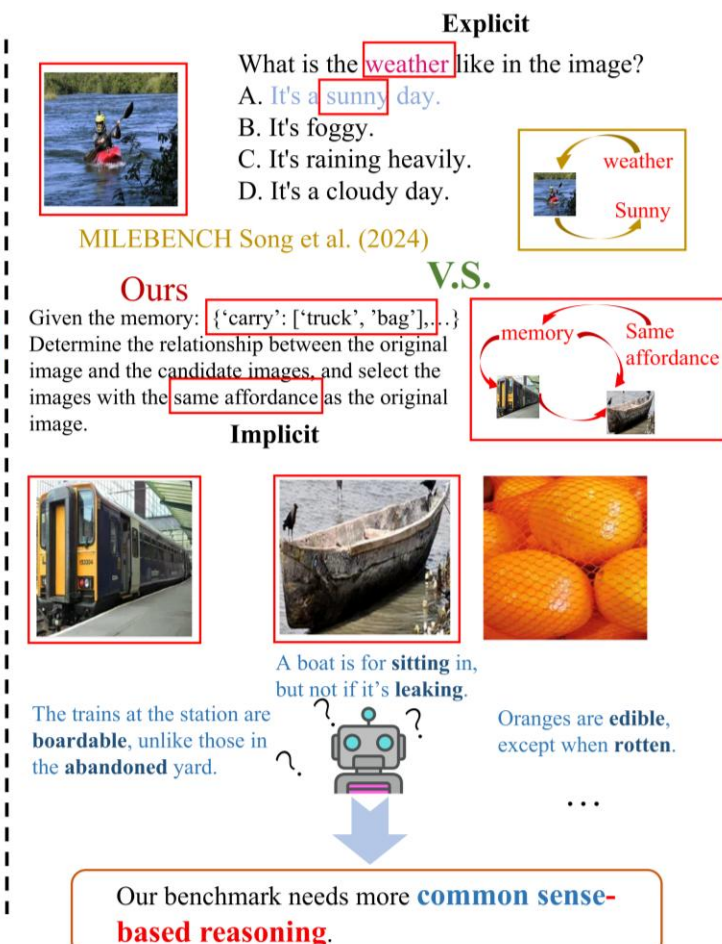
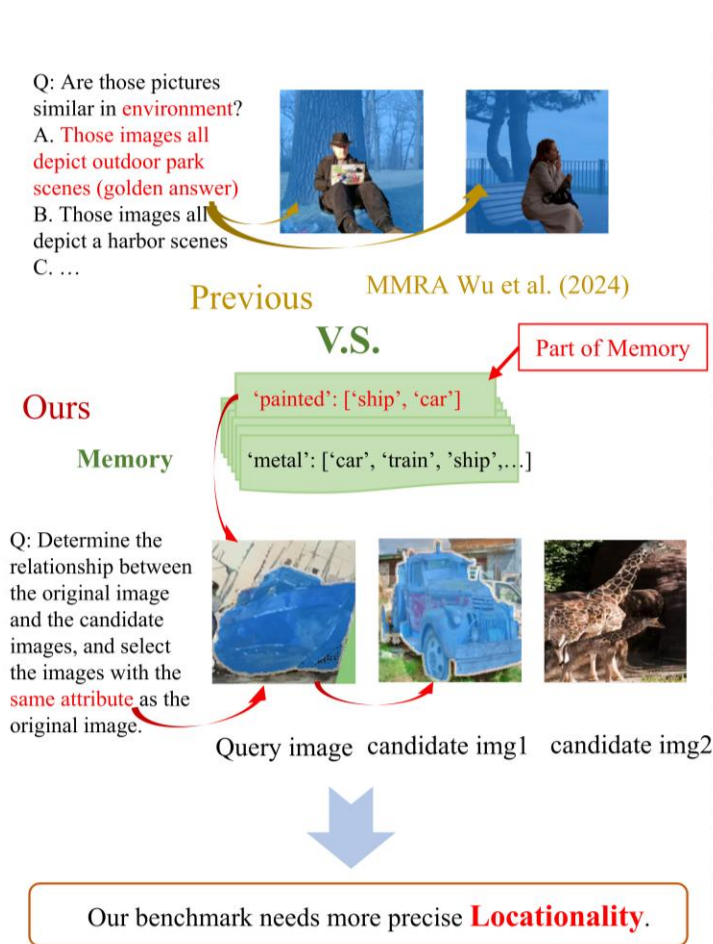


Results & Analysis

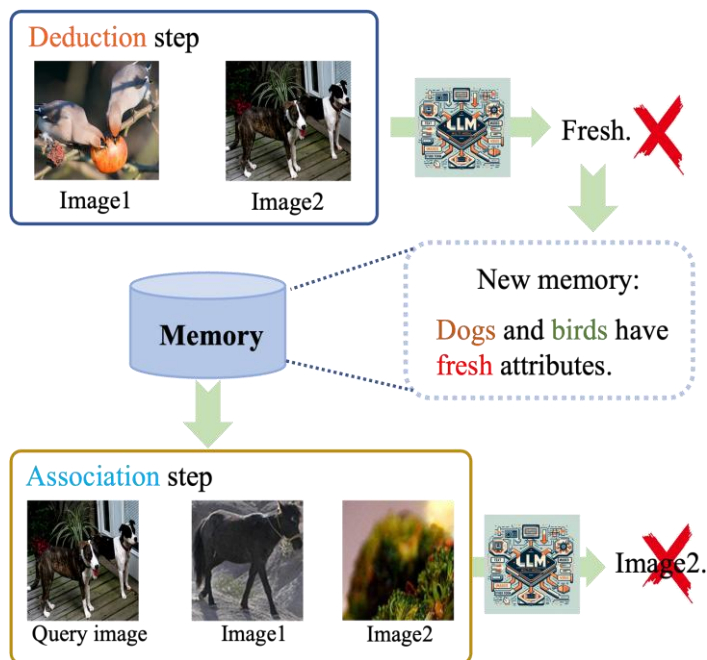
Asynchronous association with paired categories across different concepts



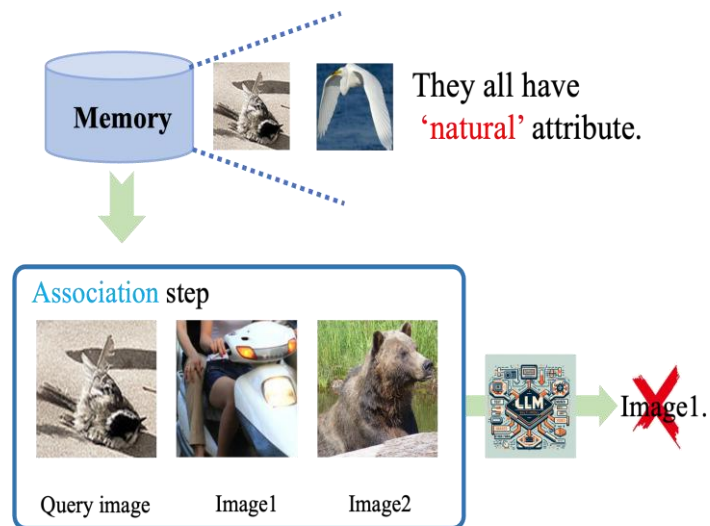
Results & Analysis



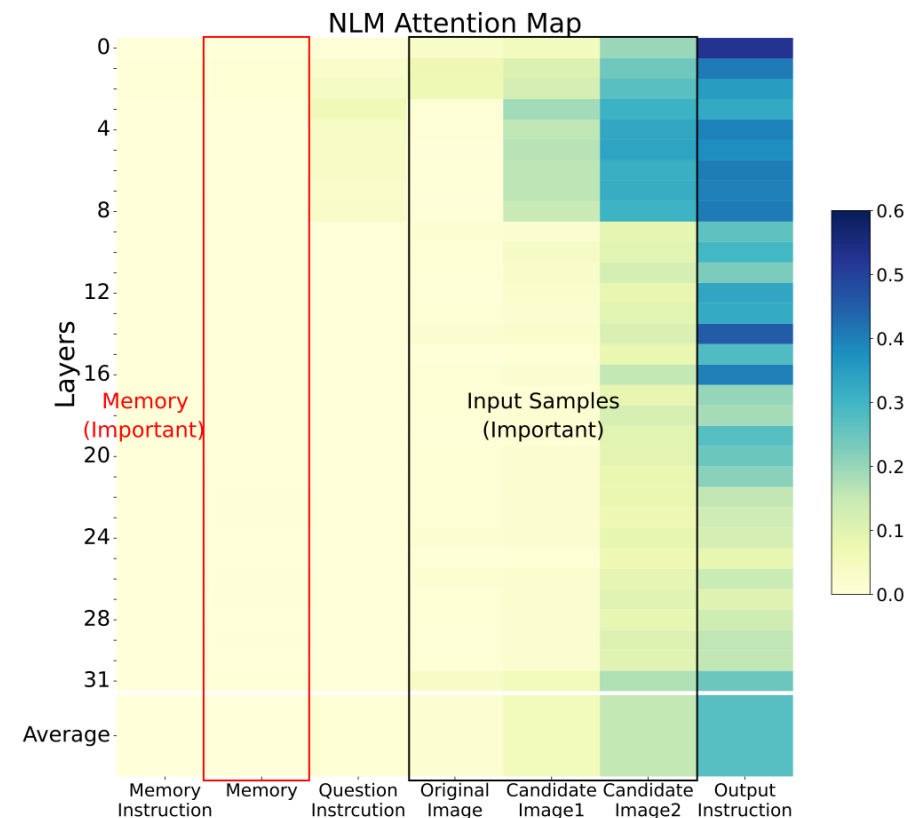
Results & Analysis



Failure case 1: Error Deduction



Failure case 2: Error Association





ICLR

Thanks for Listening!

https://mvig-rhos.com/llm_inception



**SHANGHAI JIAO TONG
UNIVERSITY**



RHOS 
mvig-rhos.com