

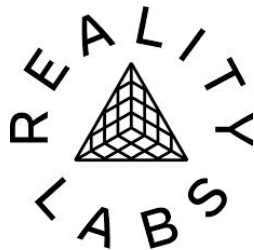


NEURAL INFORMATION  
PROCESSING SYSTEMS

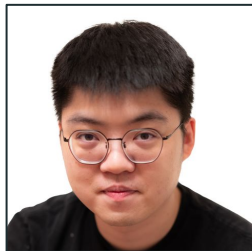
# VisualLens: Personalization through Task-Agnostic Visual History

*Presenter:*  
**Wang Bill Zhu**  
[wangzhu@usc.edu](mailto:wangzhu@usc.edu)

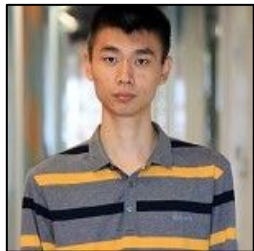
 **Meta**



# Collaborators



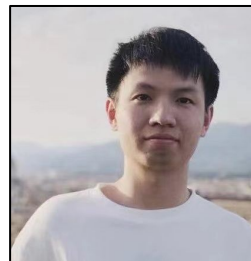
*Deqing Fu*



*Kai Sun*



*Yi Lu*



*Zhaojiang Lin*



*Seungwhan Moon*



*Kanika Narang*



*Mustafa Canim*



*Yue Liu*



*Anuj Kumar*

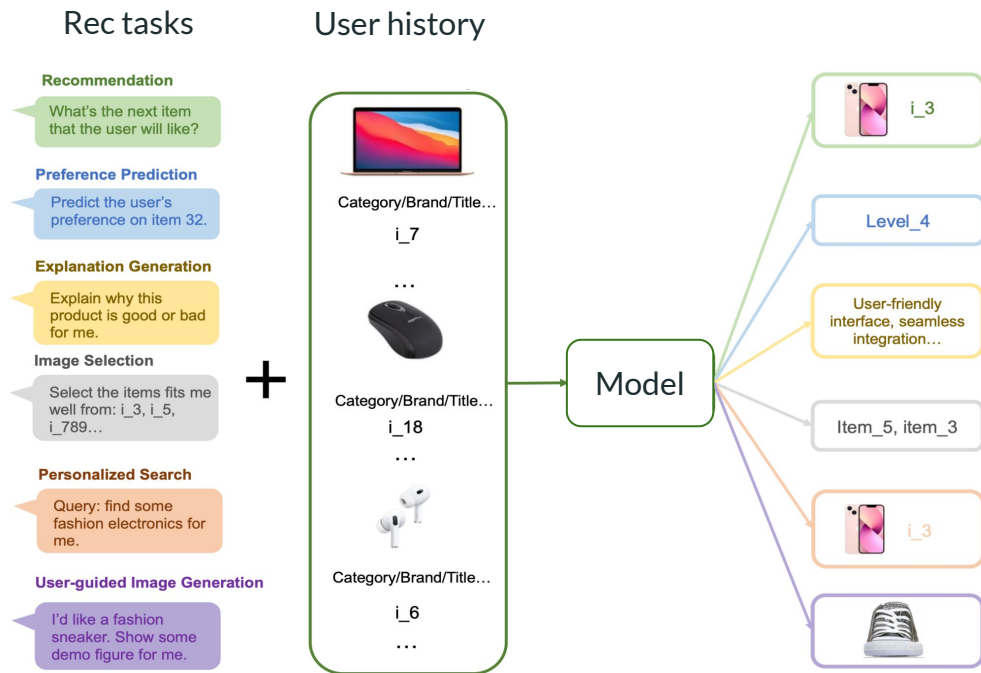


*Xin Luna Dong*

# Traditional RecSys rely on domain-specific user history

- Online shopping
  - Amazon baby, clothes, etc.
- Watching movies / books
  - MovieLens, Netflix
- Online social connection
- Ordering takeouts

Easy to get record as the history



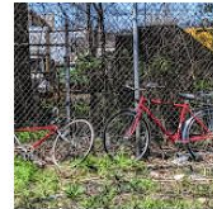
(Wei et al. ICLR 2024)

# Visual record as user history

- Bob comes to San Diego for the first time and would like to visit museums
- Nancy is starving and would like to search for food nearby.

**Without past museum or restaurant history, what can we recommend?**

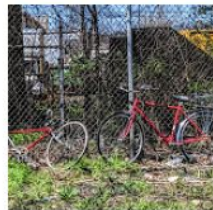
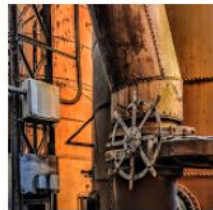
**Your personal photos reveal your taste!**



# Challenges: noisy history and new evaluation



Existence of visual record as user history



Noisy history:

- no **explicit** item ID in history
- category **diversity** in history
- no guarantee on **positive** feedback



No existing **benchmarks** for evaluation

# Key idea: retrieval with spectrum user profiling



Noisy history:

- no **explicit** item ID in history
- category **diversity** in history
- no guarantee on **positive** feedback



No existing **benchmarks** for evaluation



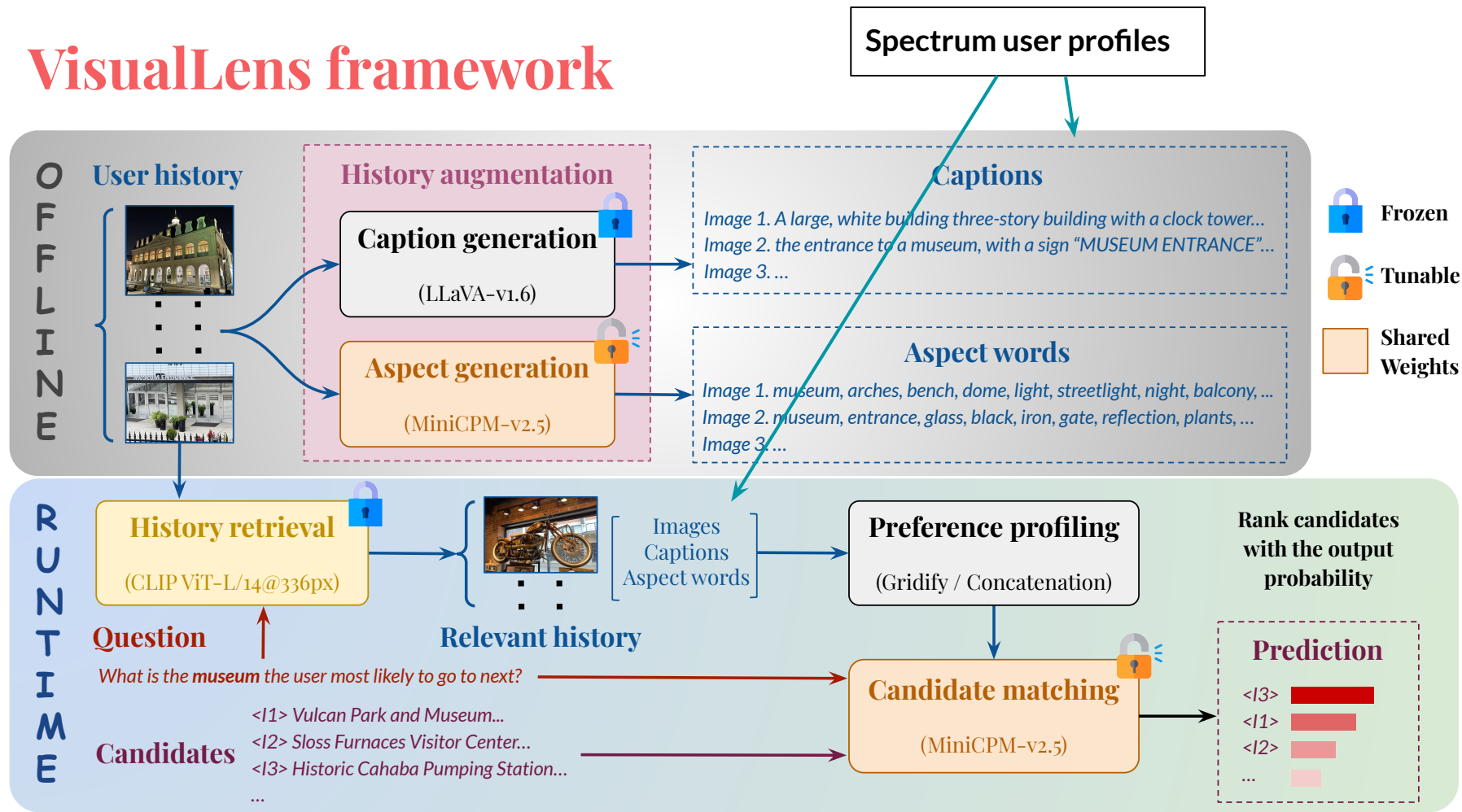
Effective retrieval:

- based on question **category**
- with **spectrum** user profile augmentation



New **benchmark creation**

# VisualLens framework



# Dataset creation: Google-Review-V and Yelp-V

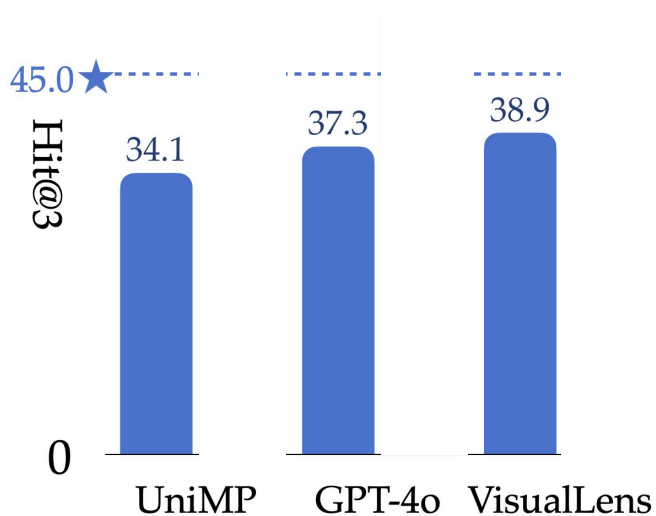
We create two benchmarks, **Google-Review-V** and **Yelp-V**, leveraging publicly available data from **Google Local Review** and **Yelp**.





# VisualLens outperforms prior SotA and frontier VLMs

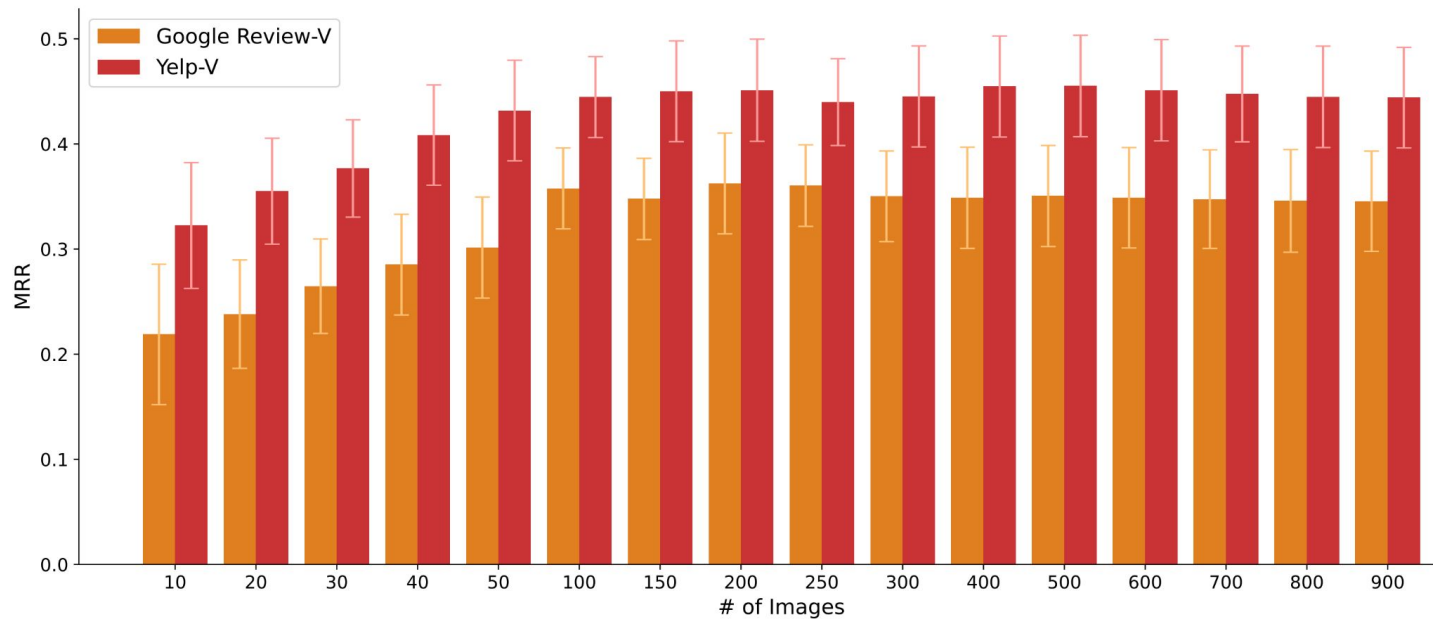
■ GoogleReview-V ■ Yelp-V ★ Human Oracle



# Retrieval and spectrum user profiling are crucial

Representation			Ret.	Google Review-V				Yelp-V			
Asp.	Cap.	Img.		Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10	MRR
✓	✓	✓	✓	15.7	35.2	75.4	32.5	26.9	57.5	88.2	42.9

# VisualLens is effective with limited visual history



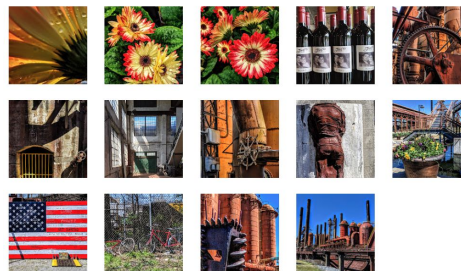
# Takeaways

- We proposed a novel **VisualLens** framework as a first step towards harnessing **users' visual records for recommendation**.
- We find a **spectrum** user profile is crucial for task-agnostic visual recommendation.
- We created **Google-Review-V** and **Yelp-V** as **benchmarks** for evaluating task-agnostic visual recommendation.

*Scan for paper*



## User history



## Question

*What are the museums the user most likely to go to next?*