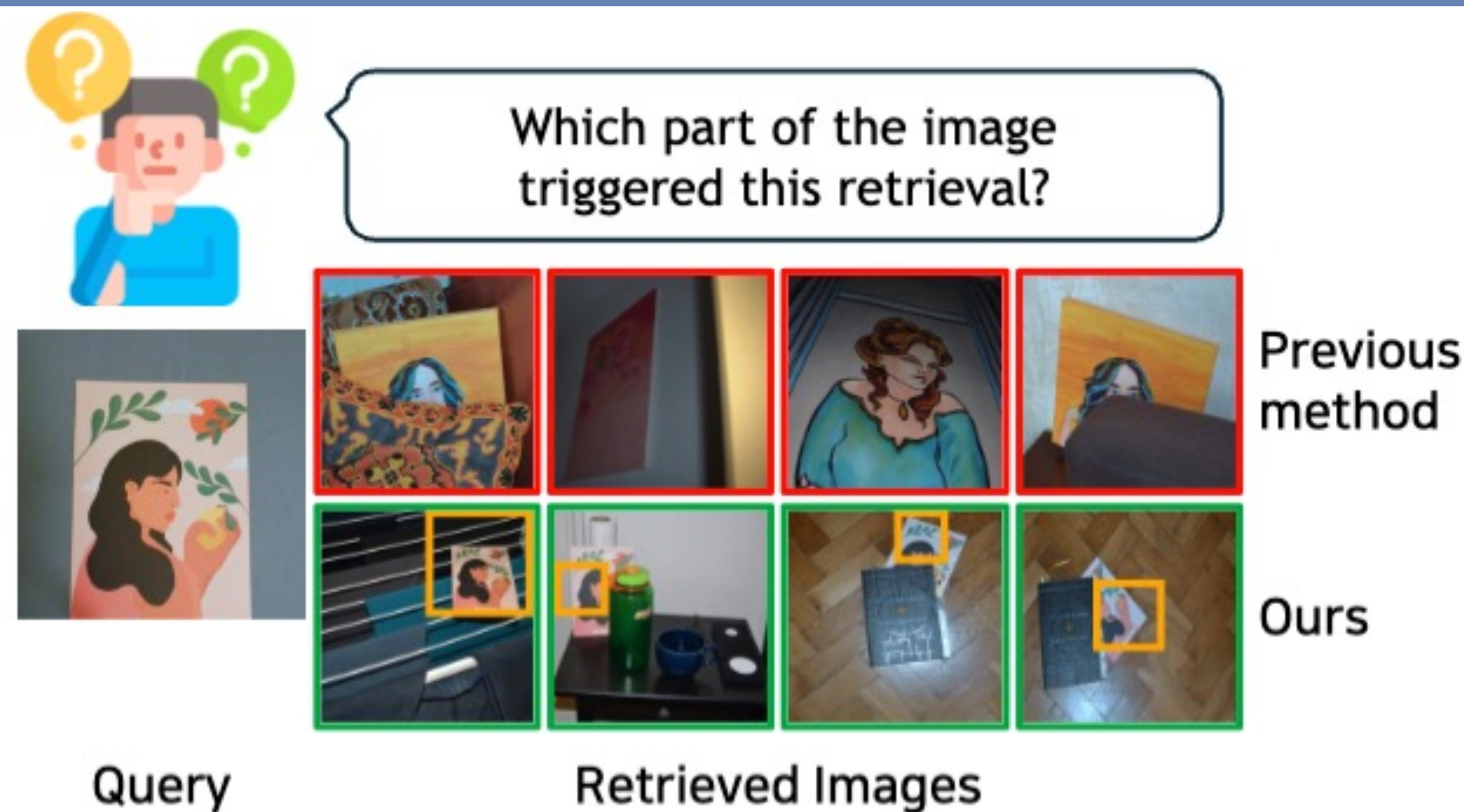
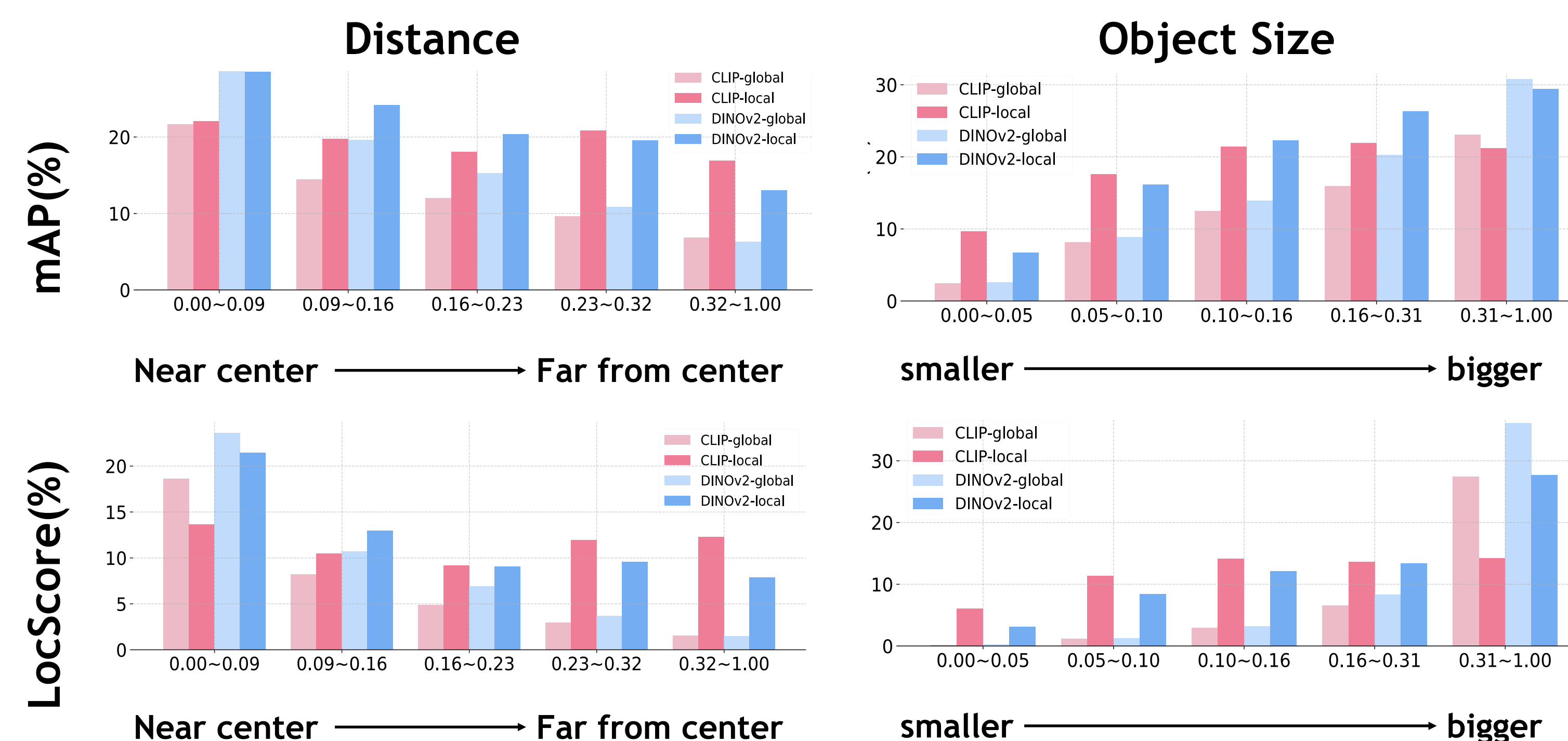


Motivation



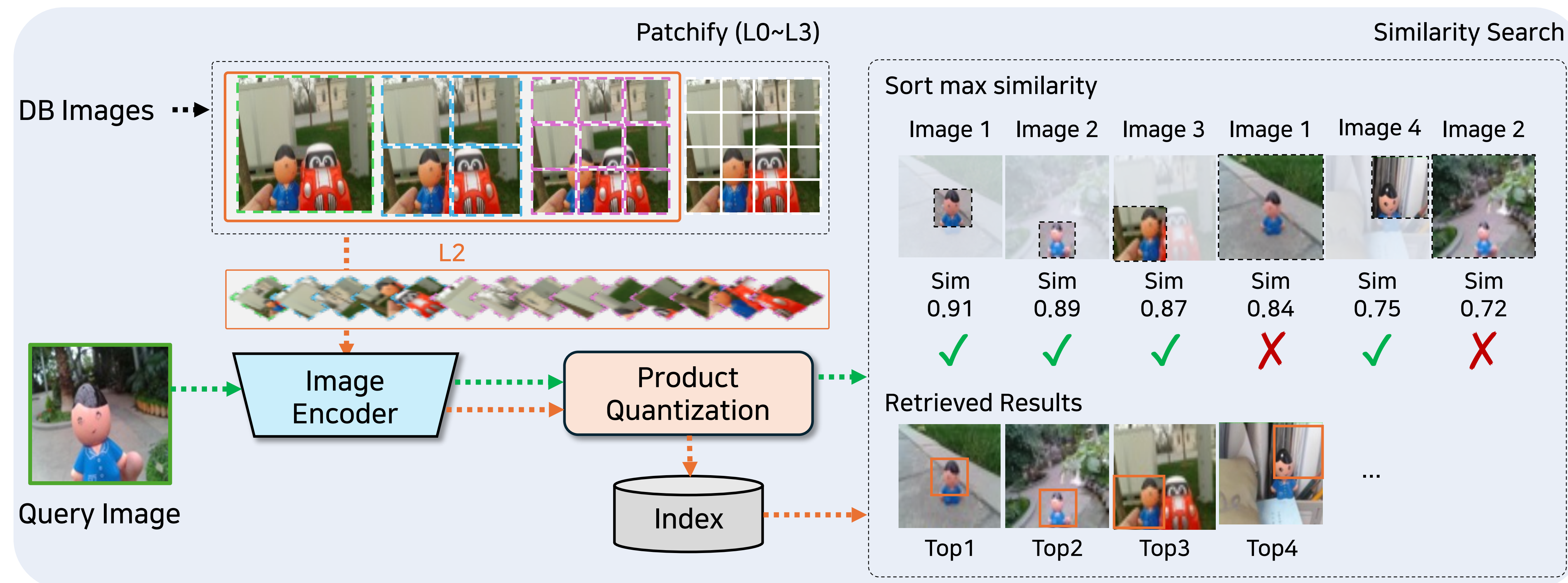
- **Interpretability:** Explicitly shows *where* the match occurs.
- **Performance:** Higher accuracy via local cues.
- **Scalability:** Plug & Play module and efficient large-scale retrieval with PQ.

Preliminary experiment



- Global representations blur localized evidence, failing to capture small or off-center objects accurately, and more distracted.
- Patch-based(local patch matching) retrieval preserves spatial cues, enabling more robust localization and consistent performance.

Patchify



- Patchify divides each image into multi-scale patches and encodes them individually.
- During retrieval, the system compares query patches to all database patches and ranks images by the maximum similarity.
- This enables **fine-grained, explainable instance retrieval with spatial localization by design.**
- As a **simple, training-free** framework, Patchify can leverage any new powerful backbone without retraining.

LocScore: Localization-aware Metrics

(a) Continuous LocScore

$$LocScore = \frac{1}{N} \cdot \frac{1}{I_n} \sum_{n=1}^N \sum_{i=1}^{I_n} \frac{h_{n,i}^{n,i}}{r_{n,i}^{n,i}} \cdot IoU(B_{gt}^{n,i}, B_{pred}^{n,i})$$

(b) Discrete LocScore

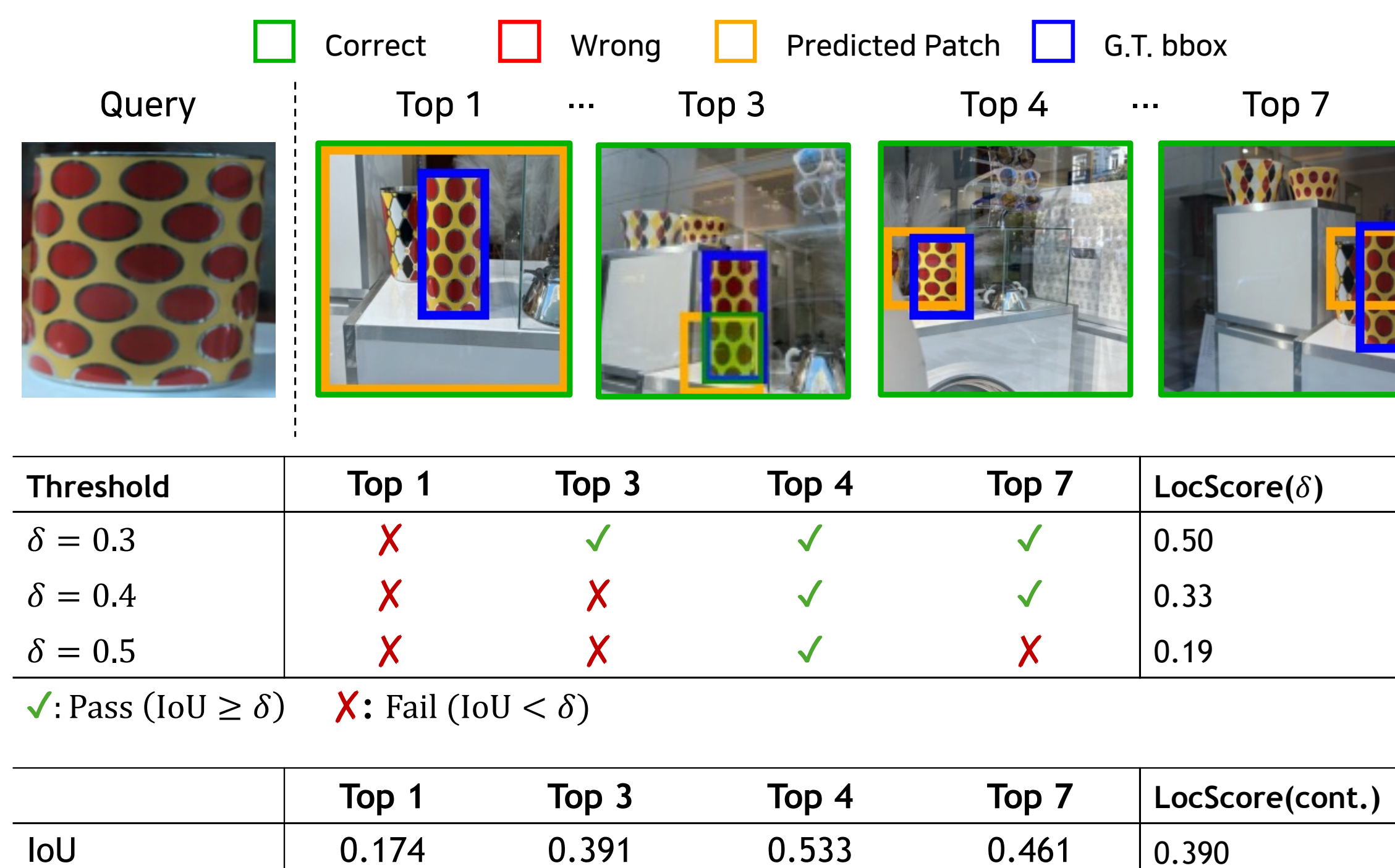
$$LocScore(\delta) = \frac{1}{N} \cdot \frac{1}{I_n} \sum_{n=1}^N \sum_{i=1}^{I_n} \frac{h_{n,i}^{n,i}}{r_{n,i}^{n,i}} \cdot \mathbb{I}[IoU(B_{gt}^{n,i}, B_{pred}^{n,i}) \geq \delta]$$

I_n : # of GT positives for query n

$r_{n,i}$: Rank position of the retrieved GT images

$h_{n,i}$: # of true positives retrieved within top $r_{n,i}$

- (a) Measures **retrieval quality** with **spatial awareness**.
- (b) further enables user-controllability with the notion of “good localization”.



Experimental results

Different region selection strategies

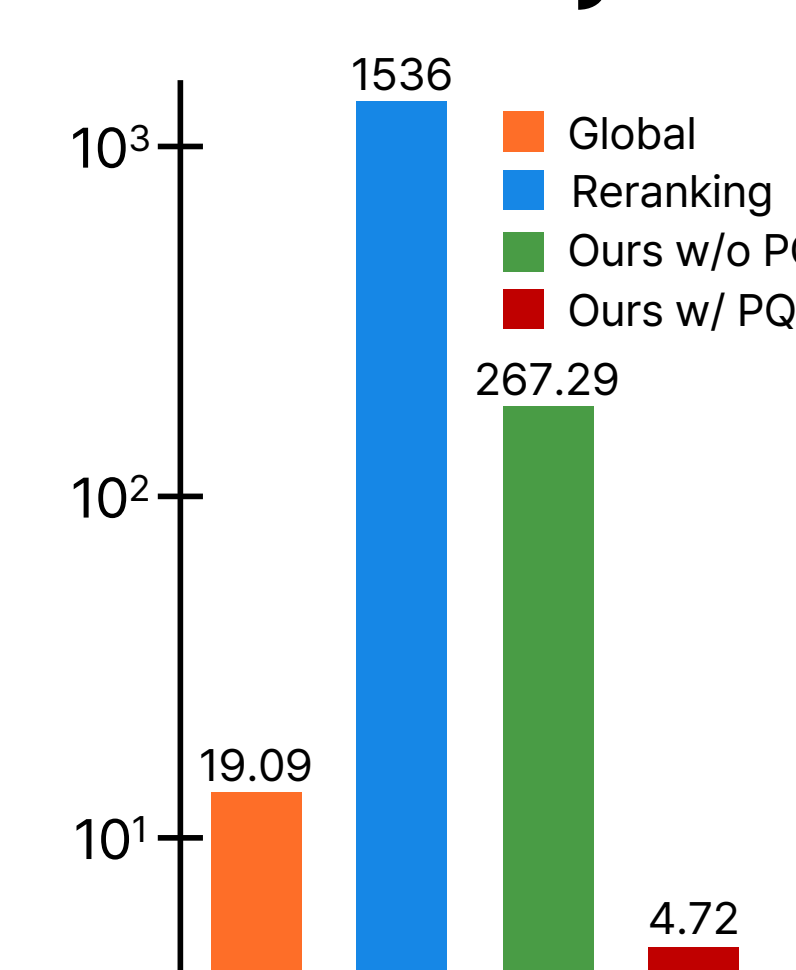
Method	INSTRE		ILIAS	
	mAP	LocScore	mAP	LocScore (%)
Global	78.48	18.92	55.03	14.31
Patchify	87.01	24.29	65.16	19.75
Sliding Window (0.5)	89.26	27.50	68.24	22.36
Sliding Window (0.25)	90.62	29.87	70.02	24.57
Region Proposal	92.96	71.44	84.19	68.23

Comparison with SOTA Methods (on Mini-ILIAS)

Method	mAP@1k	Method	mAP@1k
Global feature		Reranking (AMES)	
DINOv2 [†]	18.80	DINOv2 [†]	26.50
SigLIP2 [†]	37.30	OpenCLIP [†]	32.90
DINOv3	21.80	DINOv3	29.86
+ Patchify (L3)	42.98	+ Patchify (L3)	50.72
SigLIP	20.41	SigLIP	26.99
+ Patchify (L2)	50.27	+ Patchify (L2)	56.54
SigLIP [†]	33.86	SigLIP [†]	40.91
+ Patchify (L3)	40.48	+ Patchify (L3)	48.21

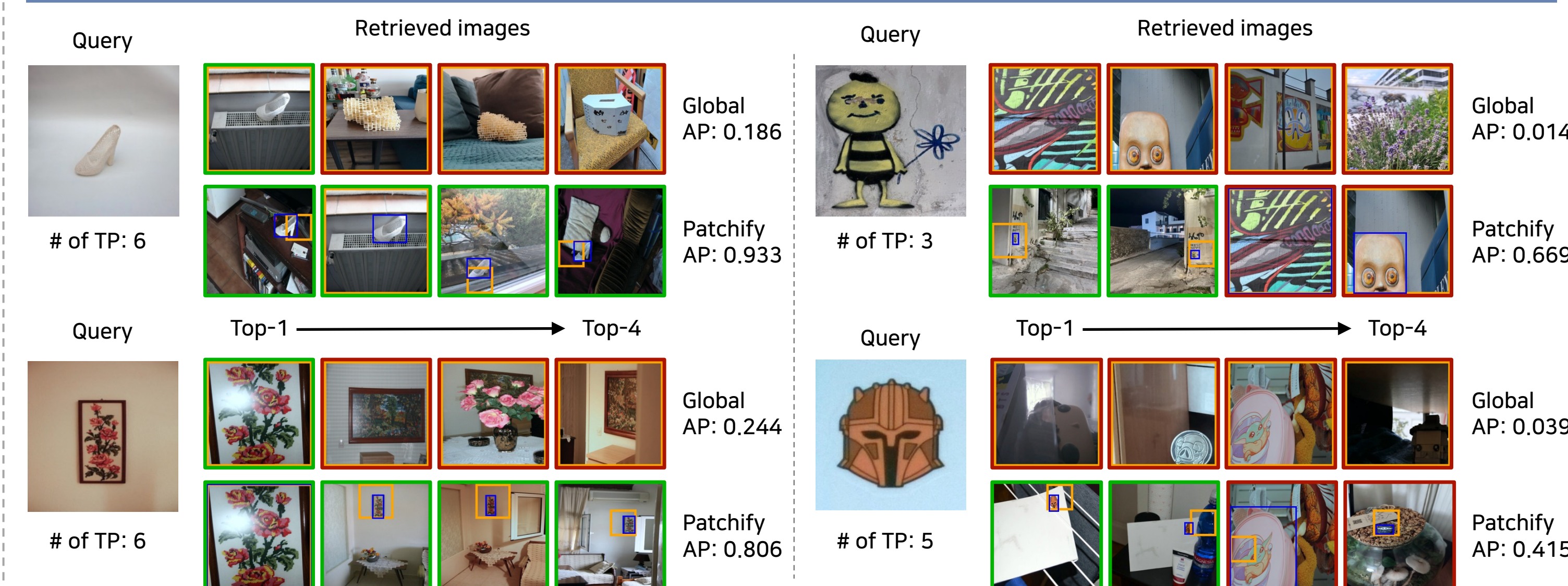
Performance scales with the granularity of region selection.

Memory



Patchify outperforms both global and local methods with less memory!

Qualitative results



Takeaways

- **Interpretable instance retrieval with built-in spatial localization.**
- **Strong performance and scalability with a simple, training-free design.**