# Patch-wise Retrieval: A Bag of Practical Techniques for Instance-Level Image Matching

Wonseok Choi[1] Sohwi Lim[2] Nam Hyeon-Woo[1] Moon Ye-Bin[1] Dong-ju Jeong[3] Jinyoung Hwang[3] Tae-Hyun Oh[2]

[1]POSTECH [2]KAIST [3]Samsung Research

WACV 2026 TUCSON, AZ 3/6 – 3/10

## Contributions

- **A comprehensive study on the effectiveness of localized retrieval for instance-level matching**
- LocScore enables evaluation beyond AP by measuring spatial alignment.
- Structured patch-based retrieval provides a strong baseline for studying localized retrieval.
- Experiments uncover practical design factors such as region granularity, window size, and feature budget.
- Strong performance can be achieved with only a handful of structured patch features under realistic memory constraints.

## LocScore: Localization-aware Metrics

(a) Continous LocScore

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

(b) Discrete LocScore

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

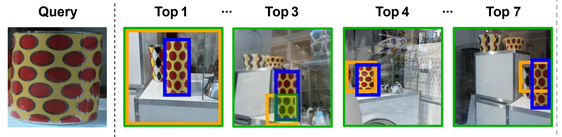
$I_n$: *# of GT positives for query n*

$r_{n,i}$: *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".

### Effect of IoU Threshold on Discrete LocScore



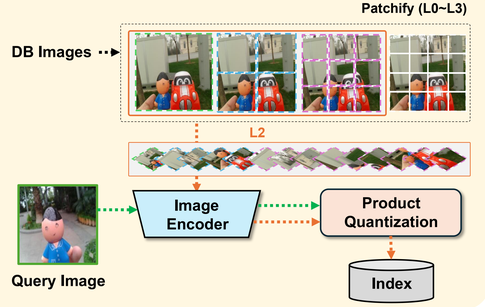

| Threshold | Top 1 | Top 3 | Top 4 | Top 7 | LocScore($\delta$) |
|---|---|---|---|---|---|
| $\delta = 0.3$ | ✗ | ✓ | ✓ | ✓ | 0.50 |
| $\delta = 0.4$ | ✗ | ✗ | ✓ | ✓ | 0.33 |
| $\delta = 0.5$ | ✗ | ✗ | ✓ | ✗ | 0.19 |

✓: Pass (IoU $\geq \delta$) ✗: Fail (IoU $< \delta$)

| | Top 1 | Top 3 | Top 4 | Top 7 | LocScore(cont.) |
|---|---|---|---|---|---|
| IoU | 0.174 | 0.391 | 0.533 | 0.461 | 0.390 |

Correct Wrong Predicted Patch G.T. bbox

## Patchify: Localized Retrieval Baseline

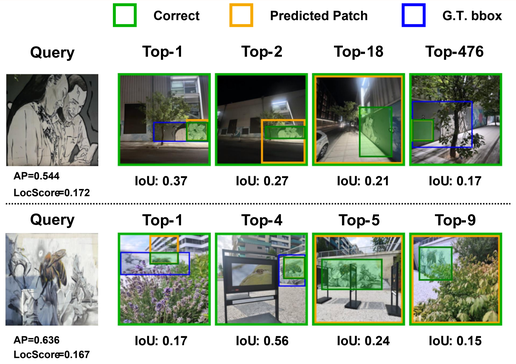



Directly applicable to strong pretrained backbones without fine tuning !

- Structured multi-scale patch representation (L0-L3).
- Patch-level matching with max-similarity scoring for image ranking.

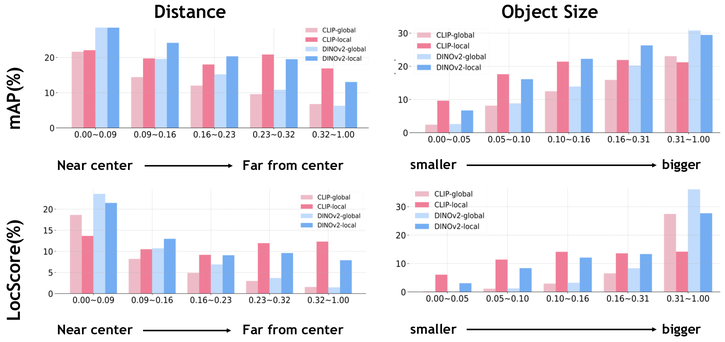
## A Bag of Practical Findings

### High AP Does Not Guarantee Correct Localization




- Both examples achieve perfect ranking performance.
- However, their predicted regions show very different spatial overlap with the target.
- AP measures which images are retrieved, while **LocScore further evaluates how well they are spatially aligned.**

### Local Features Help for Small and Off-Center Objects




- **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.**

### Spatial Granularity Alone Is Not Enough

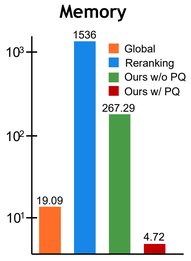
| Method (L3 setting) | INSTRE mAP | INSTRE LocScore(%) | ILIAS mAP | ILIAS LocScore |
|---|---|---|---|---|
| Global | 78.48 | 18.92 | 55.03 | 14.31 |
| Patchify | 87.01 | 24.29 | 65.16 | 19.75 |
| Sliding Window (0.5) | 89.26 | 27.50 | 68.24 | 22.36 |
| Sliding Window (0.25) | 90.62 | 29.87 | 70.02 | 24.57 |
| Region Proposal | 92.96 | 71.44 | 84.19 | 68.23 |

- Retrieval performance increases with finer region granularity.
- However, gains diminish with overly dense sampling.
- Sliding window results suggest that increasing spatial coverage alone is insufficient.
- Region proposal results further indicate that **both window size and region granularity are critical for optimal performance.**

### Competitive Localized Retrieval with a Handful of Features

(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 |




- Patchify provides strong first-stage retrieval performance.
- It integrates effectively with reranking pipelines using only a handful of features.
- With Product Quantization, memory usage remains comparable to global search and significantly lower than dense local reranking methods.
- **Patchify strikes a practical balance between accuracy, scalability, and efficiency.**
- Patchify provides a strong and practical starting baseline for localized retrieval !