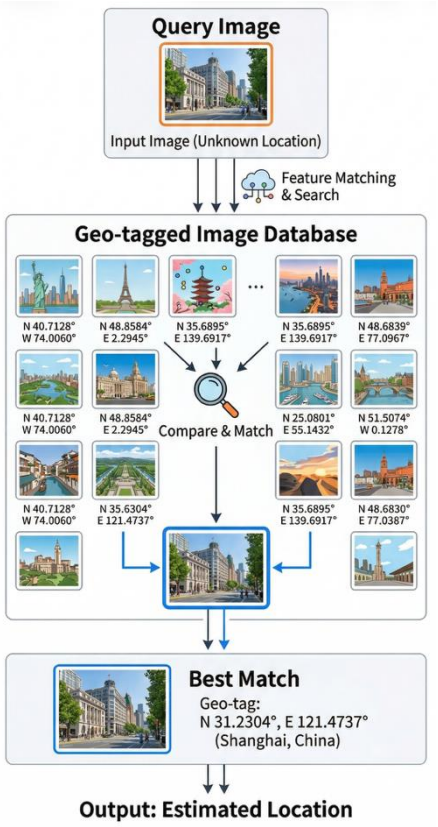
SAGE: Spatial-visual Adaptive Graph Exploration for Efficient Visual Place Recognition

Shunpeng Chen[1], Changwei Wang[2], Rongtao Xu[3], Xingtian Pei[1], Yukun Song[1]
Jinzhou Lin[1], Wenhao Xu[1], Jingyi Zhang[1], Li Guo[1], Shibiao Xu[1*]

[1] School of Artificial Intelligence, Beijing University of Posts and Telecommunications
[2] Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center, Qilu University of Technology [3] Spatialtemporal AI
shunpengchen@bupt.edu.cn, shibiaoxu@bupt.edu.cn



* Shibiao Xu is the corresponding author (shibiaoxu@bupt.edu.cn).

ICLR 2026
April 23-27, 2026
Rio de Janeiro, Brazil

SAGE: Spatial-visual Adaptive Graph Exploration for Efficient Visual Place Recognition

ICLR
International Conference On Learning Representations

Contributions:

- **SoftP Feature Interaction.** We introduce SoftP, a lightweight module that uses residual weighting to amplify discriminative local patches, alongside an InteractHead that models cross-image associations to ensure descriptor coherence across varying views.

- **Dynamic Geo-Visual Graph Mining.** We introduce SoftP, a lightweight module that uses residual weighting to amplify discriminative local patches, alongside an InteractHead that models cross-image associations to ensure descriptor coherence across varying views.

- **Weighted Greedy Clique Expansion.** This weight-guided algorithm seeds batches from high-affinity anchors and iteratively expands into challenging neighborhoods, focusing the model's learning on fine-grained spatial and visual distinctions.

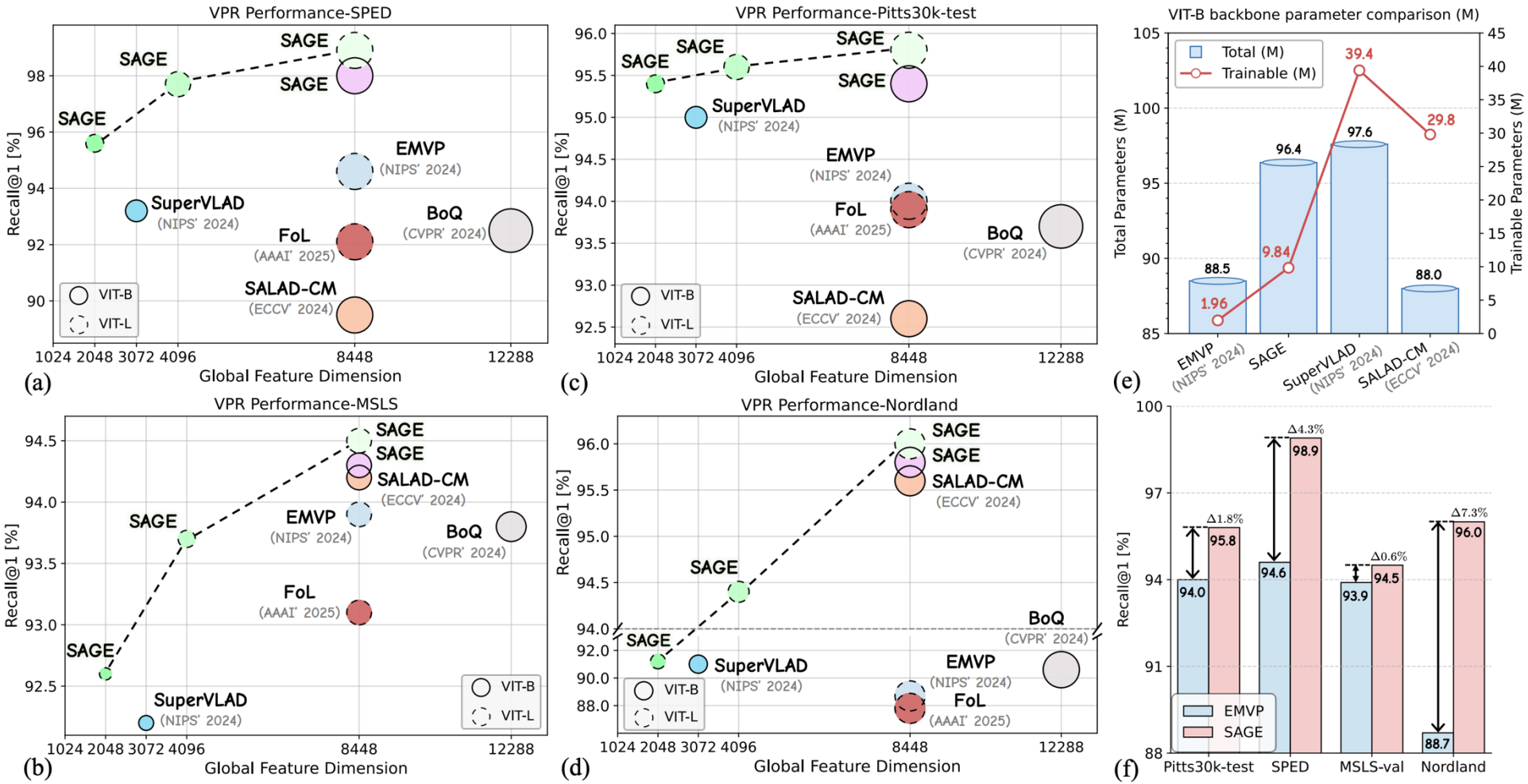
- **Efficient SOTA Accuracy.** By leveraging a fully frozen DINOv2 backbone with parameter-efficient fine-tuning, SAGE achieves state-of-the-art performance across eight challenging VPR benchmarks, maintaining robust retrieval accuracy even with highly compact descriptors.

SAGE: Spatial-visual Adaptive Graph Exploration for Efficient Visual Place Recognition

ICLR International Conference On Learning Representations

Ultra-Parameter-Efficient:

Requires only 9.84M **trainable parameters** by utilizing a frozen DINOv2 backbone.



Performance & Efficiency Overview

SAGE: Spatial-visual Adaptive Graph Exploration for Efficient Visual Place Recognition

ICLR
International Conference On Learning Representations

Motivations:

Static Sampling vs. Dynamic Evolution.

- Prior visual place recognition (VPR) methods rely on static, pre-defined sampling strategies that fail to adapt as the model's embedding space evolves during training. This "think-once, act-always" approach feeds stale examples to the model, limiting its full discriminative potential.

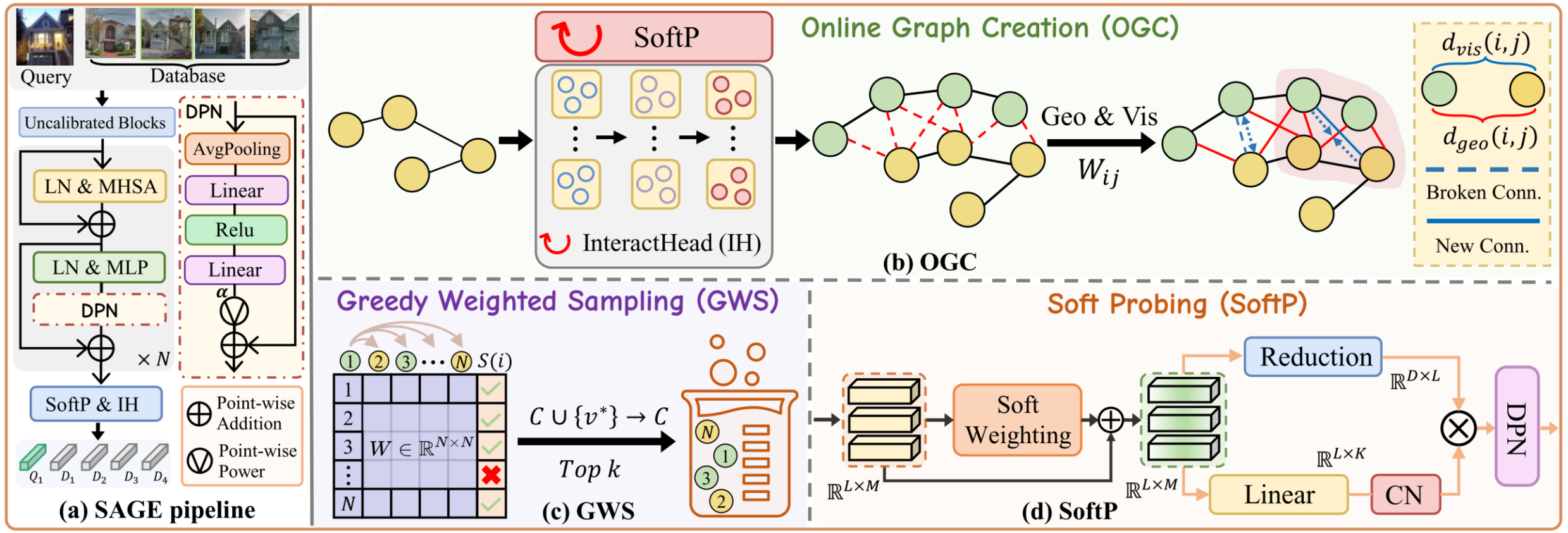
Neglect of Fine-Grained Local Cues.

- Existing global feature aggregation methods often treat all local spatial descriptors with equal weighting. This uniform treatment underemphasizes subtle yet highly discriminative local visual cues that are crucial for distinguishing visually similar places.

Fully Frozen Backbone for Extreme Efficiency.

- Unlike prior approaches that fine-tune portions of the Transformer encoder (e.g., the last few layers) and incur significant training overhead , there is a strong motivation to keep the massive visual foundation model's backbone completely frozen. The goal is to achieve state-of-the-art accuracy exclusively through ultra-lightweight, parameter-efficient fine-tuning (PEFT) modules.

SAGE: Spatial-visual Adaptive Graph Exploration for Efficient Visual Place Recognition



Overall Framework

SAGE: Spatial-visual Adaptive Graph Exploration for Efficient Visual Place Recognition

ICLR International Conference On Learning Representations

Dataset	Description	Number	
		Database	Queries
Pitts30k-test	urban, panorama	10,000	6,816
MSLS-val	urban, suburban	18,871	740
Nordland	natural, seasonal	27,592	27,592
SPED	various scenes	607	607
Tokyo24/7	urban, time-varying	75,984	315
AmsterTime	urban, time-related	1,231	1,231
Pitts250k-test	urban, panorama	83,952	8,280
Eynsham	rural, historical	23,935	23,935

Dataset Diversity

Method	Total ↓	Trainable ↓	Adapter
SALAD CVPR' 2024	**88.0**	29.8	✗
SelaVPR ICLR' 2024	102.8	16.2	✓ 14.2
CricaVPR CVPR' 2024	95.7 (+11.0)	9.15 (+11.0)	✓ 9.2
SALAD-CM ECCV' 2024	**88.0**	29.8	✗
SuperVLAD NIPS' 2024	86.6 (+11.0)	28.4 (+11.0)	✗
EMVP NIPS' 2024	88.5	**1.96**	✗
SAGE (Ours)	88.5 (+7.88)	**1.96** (+7.88)	✗

Adapter-Free Efficiency

Table 2: Comparison to SoTA Methods on VPR Benchmark Datasets. The best and second best metrics are shown in **red bold** and **blue bold**, respectively. Two-stage methods are denoted by †.

Method	Dim	SPED			Pitts30k-test			MSLS-val			Nordland		
		R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10
NetVLAD CVPR' 2016	32768	70.2	84.5	89.5	81.9	91.2	93.7	53.1	66.5	71.1	6.4	10.1	12.5
SFRS ECCV' 2020	4096	80.2	92.6	95.4	89.4	94.7	95.9	69.2	80.3	83.1	16.1	23.9	28.4
CosPlace CVPR' 2022	512	75.5	87.0	89.6	88.4	94.5	95.7	82.8	89.7	92.0	58.5	73.7	79.4
MixVPR WACV' 2023	4096	84.7	92.3	94.4	91.5	95.5	96.3	88.0	92.7	94.6	76.2	86.9	90.3
R2Former CVPR' 2023 †	/	67.5	75.8	77.8	91.1	95.2	96.3	89.7	95.0	96.2	77.0	89.0	91.9
EigenPlaces ICCV' 2023	2048	70.2	83.5	87.5	92.5	96.8	97.6	89.1	93.8	95.0	71.2	83.8	88.1
SelaVPR ICLR' 2024	1024	83.5	92.6	94.6	90.2	96.1	97.1	87.7	95.8	96.6	72.3	89.4	94.4
SelaVPR ICLR' 2024 †	/	88.6	95.1	97.2	92.8	96.8	97.7	90.8	96.4	97.2	87.3	93.8	95.6
CricaVPR CVPR' 2024	4096	91.3	95.2	96.2	94.9	97.3	98.2	90.0	95.4	96.4	90.7	96.3	97.6
SALAD CVPR' 2024	8448	92.1	96.2	96.5	92.5	96.4	97.5	92.2	96.4	97.0	89.7	95.5	97.0
EDTformer TCSVT' 2025	4096	92.4	95.9	96.9	93.4	97.0	97.9	92.0	96.6	97.2	88.3	95.3	97.0
BoQ CVPR' 2024	12288	92.5	95.9	96.7	93.7	97.1	97.9	93.8	96.8	97.0	90.6	96.0	97.5
SALAD-CM ECCV' 2024	8448	89.5	94.9	96.1	92.6	96.8	97.8	**94.2**	97.2	97.4	**95.6**	**98.6**	99.1
SuperVLAD NIPS' 2024	3072	93.2	97.0	98.0	95.0	97.4	98.2	92.2	96.6	97.4	91.0	96.4	97.7
EMVP NIPS' 2024	8448	94.6	97.5	98.4	94.0	97.5	98.2	93.9	**97.3**	97.6	88.7	97.3	**99.3**
FoL AAAI' 2025	8448	92.1	96.5	98.0	93.9	97.2	98.1	93.1	96.9	97.4	87.8	94.5	96.4
FoL AAAI' 2025 †	/	92.6	96.5	97.4	94.5	97.4	98.2	93.5	96.9	97.6	92.6	96.7	97.8
SAGE (Ours)	2048	95.6	99.2	**99.7**	95.4	97.4	97.9	92.6	96.9	**97.7**	91.2	96.6	97.8
	4096	**97.7**	**99.8**	**100**	**95.6**	**97.7**	**98.3**	93.7	**97.3**	**97.8**	94.4	98.2	99.0
	8448	**98.9**	**99.7**	**100**	**95.8**	**97.8**	**98.4**	**94.5**	**97.4**	**97.8**	**96.0**	**98.9**	**99.4**

Comprehensive SOTA

SAGE: Spatial-visual Adaptive Graph Exploration for Efficient Visual Place Recognition

ICLR
International Conference On Learning Representations

Table 5: Ablation of SAGE components. All experiments use ViT-B; results reproduced in this work are marked ‡. OGC denotes Online Graph Creation and GWS denotes Greedy Weighted Sampling.

Method	Components			SPED			Pitts30k-test			MSLS-val			Nordland		
	Aggregation	OGC	GWS	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10
EMVP-B	CFP			91.8	96.5	97.4	93.1‡	96.8‡	97.6‡	93.2	96.9	97.2	80.8‡	90.4‡	93.5‡
SAGE-B	SoftP	✓		96.8	98.2	98.7	94.6	97.2	97.9	93.6	96.8	97.1	95.2	98.4	98.7
	SoftP		✓	96.5	97.8	98.3	93.8	96.5	97.2	92.5	96.6	96.9	94.2	97.4	97.9
	CFP	✓	✓	97.5	98.4	98.9	94.9	97.3	98.0	93.9	97.1	97.4	95.4	98.5	98.8
	SoftP	✓	✓	**98.0**	**98.7**	**99.2**	**95.4**	**97.6**	**98.3**	**94.3**	**97.2**	**97.6**	**95.8**	**98.7**	**99.2**

Component Synergy:

The combination of SoftP, OGC, and GWS is crucial for maximum performance gains.

SAGE: Spatial-visual Adaptive Graph Exploration for Efficient Visual Place Recognition

ICLR
International Conference On Learning Representations

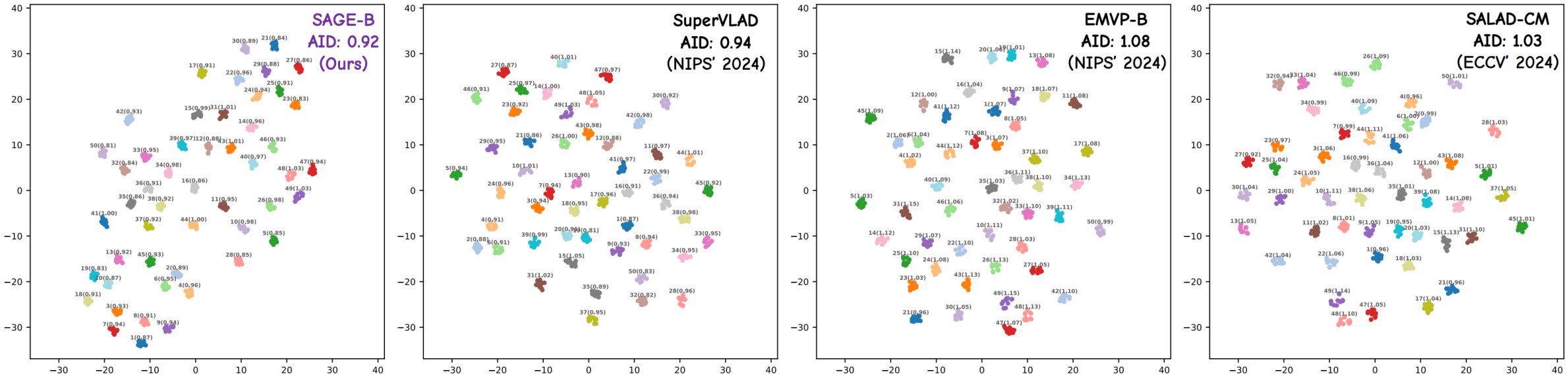


Figure 3: Visualization of spatial feature clustering using t-SNE for four methods and comparison of Average Intra-class Distance (AID). Numbers next to each class indicate intra-class distance (ID).

High Discriminability: Effectively pulls features from the same location tightly together, seamlessly handling intra-class variations.

SAGE: Spatial-visual Adaptive Graph Exploration for Efficient Visual Place Recognition

ICLR
International Conference On Learning Representations

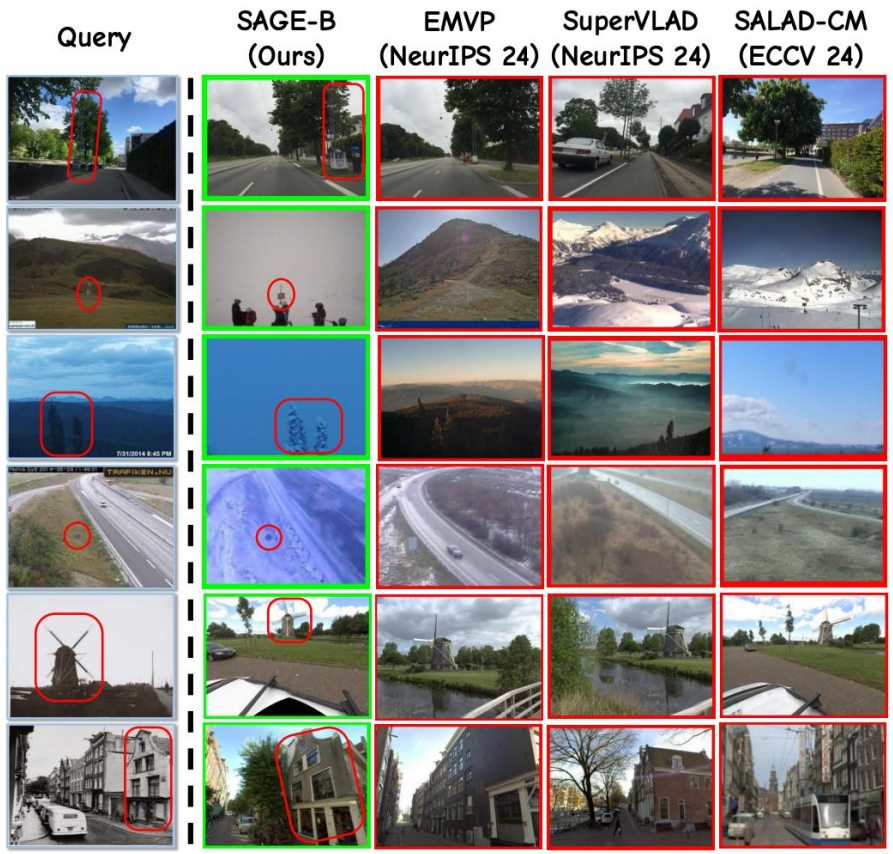


Figure 4: Qualitative results. SAGE consistently retrieves correct database images under severe challenges.

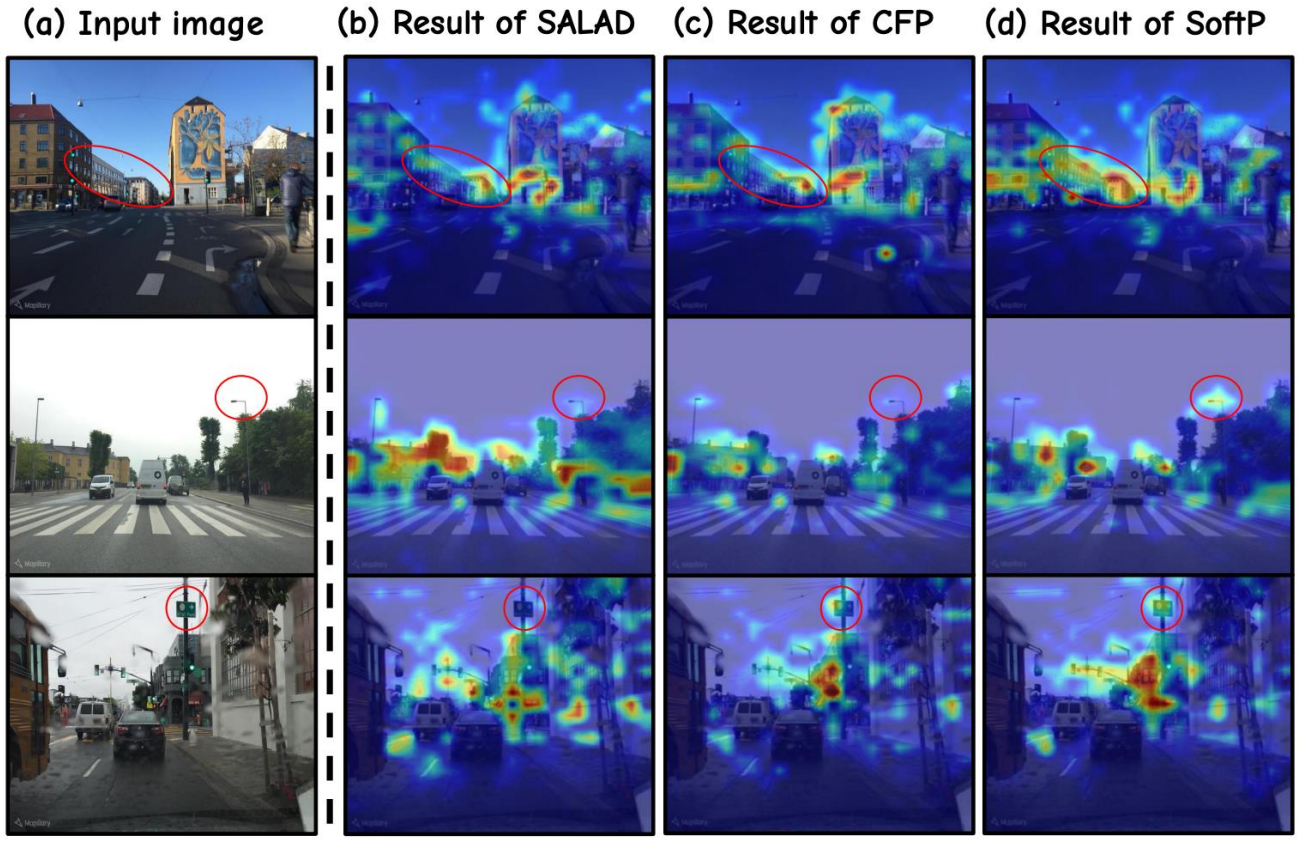


Figure 5: Visual comparison of importance heatmaps. SoftP shows a stronger focus on fine grained regions with high discriminative value than other methods overall.

SAGE: Spatial-visual Adaptive Graph Exploration for Efficient Visual Place Recognition

ICLR International Conference On Learning Representations

Online vs. Offline:

Dynamic graph creation adds minimal overhead but yields substantial accuracy improvements.

Table 6: Comparison of Online and Offline Graph Creation Strategies. Runtimes are reported per epoch for the online method. For the offline strategy, mining is a one-time cost.

Strategy	Method	Mining (min)	Train (min)	SPED			MSLS-val		
				R@1	R@5	R@10	R@1	R@5	R@10
Online	Cliquemining	4.3	25.1	90.0	95.4	96.2	94.3	96.9	97.6
	SAGE (w/o GWS)	6.1	28.4	96.8	98.2	98.7	93.6	96.8	97.1
	SAGE	6.2	28.4	**98.9**	**99.7**	**100**	**94.5**	**97.4**	**97.8**
Offline	Cliquemining	21.6	25.1	89.5	94.9	96.1	94.2	97.2	97.4
	SAGE (w/o GWS)	30.7	28.4	96.7	98.2	98.9	93.5	96.6	97.1
	SAGE	30.9	28.4	**98.5**	**99.3**	**99.5**	**94.2**	**97.3**	**97.7**

Faster Convergence:

Dynamic sampling enables superior learning efficiency in early training epochs.

Table 7: Convergence analysis on MSLS-val. SAGE's dynamic sampling leads to superior performance in early training epochs.

Epoch	SAGE (w/ CM)			SAGE (Ours)		
	R@1	R@5	R@10	R@1	R@5	R@10
2	92.3	96.1	96.6	92.5	96.5	97.1
4	92.7	96.6	97.0	**93.4**	**96.9**	**97.4**

Table 8: Ablation on InteractHead module. We vary the model dimension (d_{model}) and feed-forward dimension (d_{ff}).

$(d_{\text{model}}, d_{\text{ff}})$	SPED			Pitts30k		
	R@1	R@5	R@10	R@1	R@5	R@10
(512, 1024)	98.8	99.5	99.7	95.5	97.6	98.2
(768, 1536)	98.6	99.3	99.7	**96.0**	**97.9**	**98.4**
(768, 1024)	**98.9**	**99.7**	**100**	95.8	97.8	**98.4**

SAGE: Spatial-visual Adaptive Graph Exploration for Efficient Visual Place Recognition

Shunpeng Chen[1], Changwei Wang[2], Rongtao Xu[3], Xingtian Pei[1], Yukun Song[1]
Jinzhou Lin[1], Wenhao Xu[1], Jingyi Zhang[1], Li Guo[1], Shibiao Xu[1*]

Thanks !

ICLR 2026
April 23-27, 2026
Rio de Janeiro, Brazil