StructVPR: Distill Structural Knowledge by Weighting Samples for Visual Place Recognition

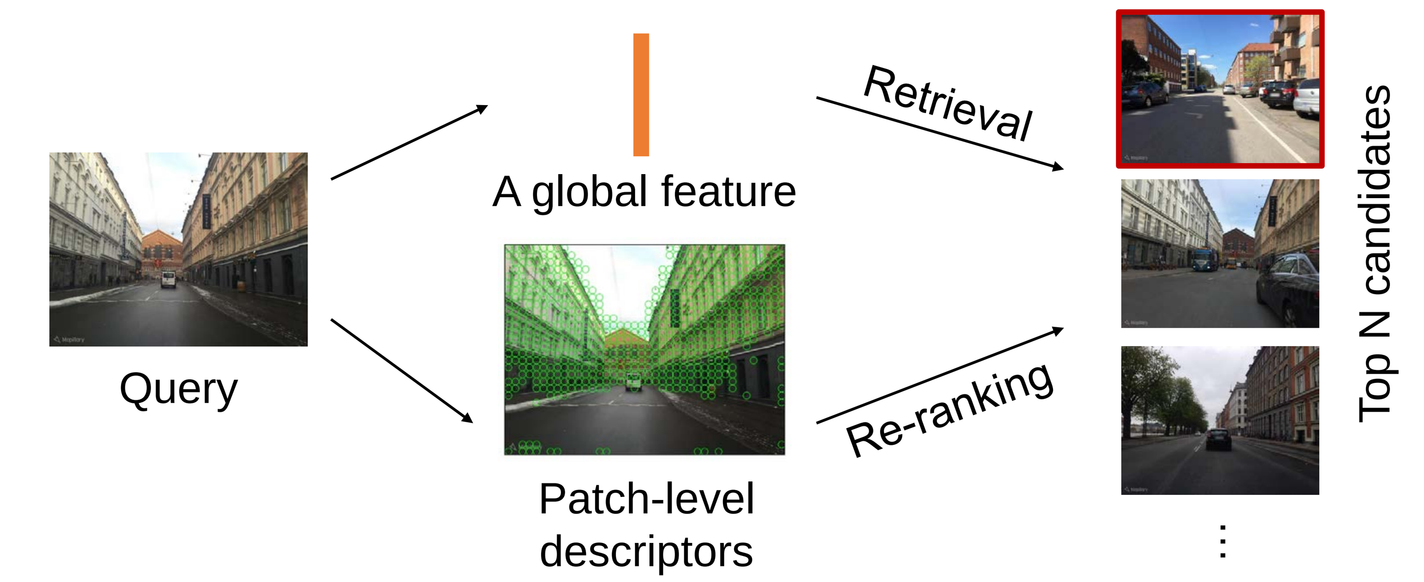
Yanqing Shen, Sanping Zhou, Jingwen Fu, Ruotong Wang, Shitao Chen, Nanning Zheng†
National Key Laboratory of Human-Machine Hybrid Augmented Intelligence, National Engineering Research Center for Visual Information and Applications,
Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Introduction

Background

- Visual place recognition (VPR): a large-scale image retrieval problem.
- Image representations for the VPR pipeline:



Insight

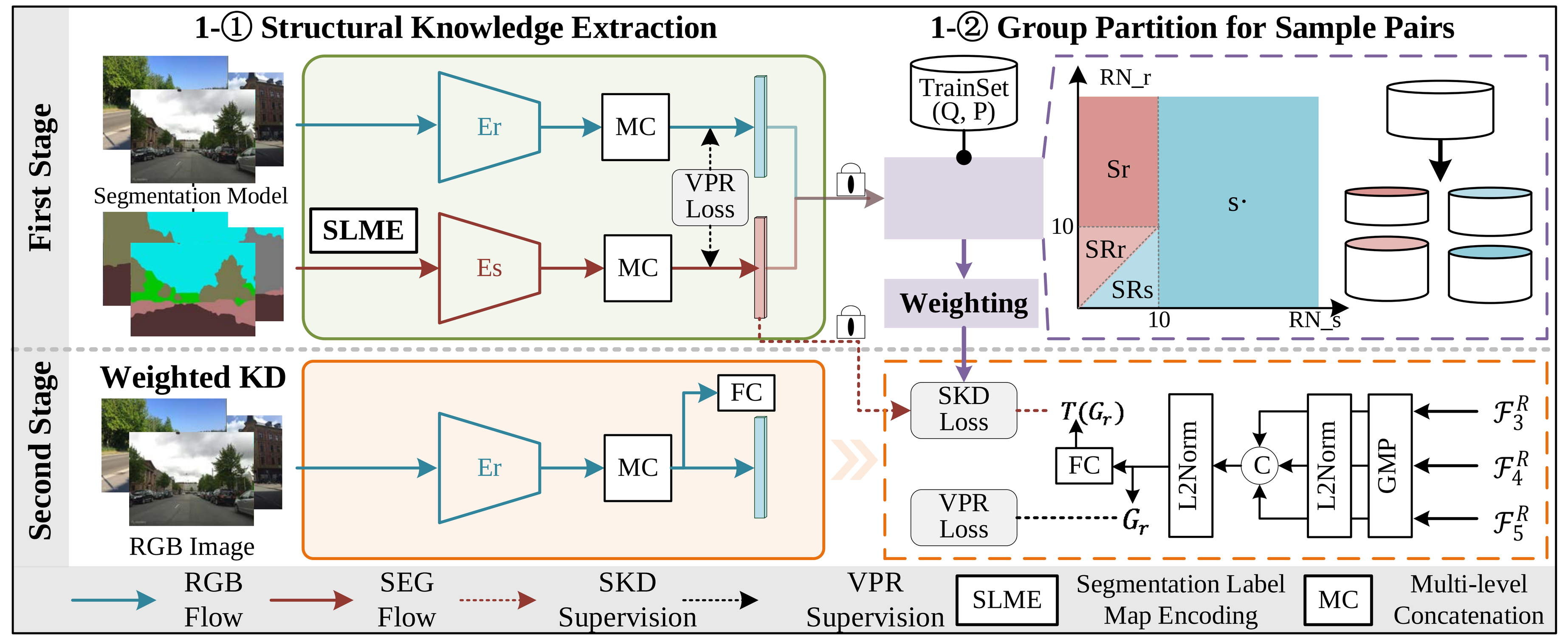
- Segmentation images have rich structural knowledge, which is essential for VPR.
- The left one shows the scene with illumination variation, and segmentation images are more recognizable; The right one shows the scene with changing perspectives, and RGB images are more recognizable.

Contributions

We propose **StructVPR,** distilling the high-quality knowledge from the SEG modality to the RGB modality and avoiding the computation and inference of segmentation during testing.

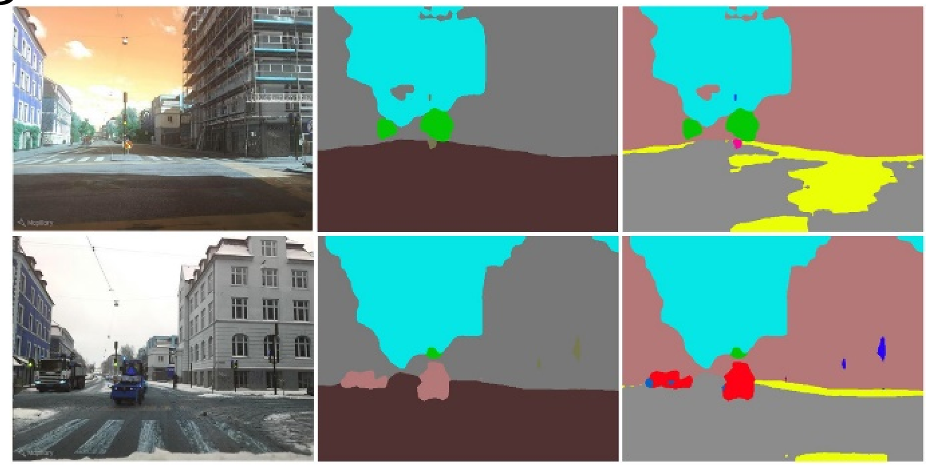
- Segmentation images are pre-encoded into weighted one-hot label maps to extract structural information for VPR.
- StructVPR forges a connection between sample partition and weighted knowledge distillation for each sample.
- *Low* computational time and memory requirements for real-world applications

Methodology



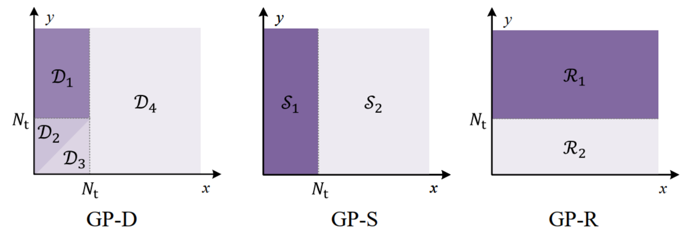
Segmentation Label Map Encoding

- SLME including three steps: formatting, clustering, and weighting.
- Too fine-grained segmentation will interfere with VPR like noises.
- Each semantic class plays a different role in VPR task.



Group Partition

- Not all samples contain high-quality and helpful teacher knowledge for the student, and even some will hurt the student's performance.
- we let the two pre-trained branches participate in seeking a more accurate partition.



$$\mathcal{D}_1 = \{(q,p) | x \leq N_t, y > N_t\},$$
$$\mathcal{D}_2 = \{(q,p) | x \leq y \leq N_t\},$$
$$\mathcal{D}_3 = \{(q,p) | y < x \leq N_t\},$$
$$\mathcal{D}_4 = \{(q,p) | x > N_t\},$$

Weighted Knowledge Distillation

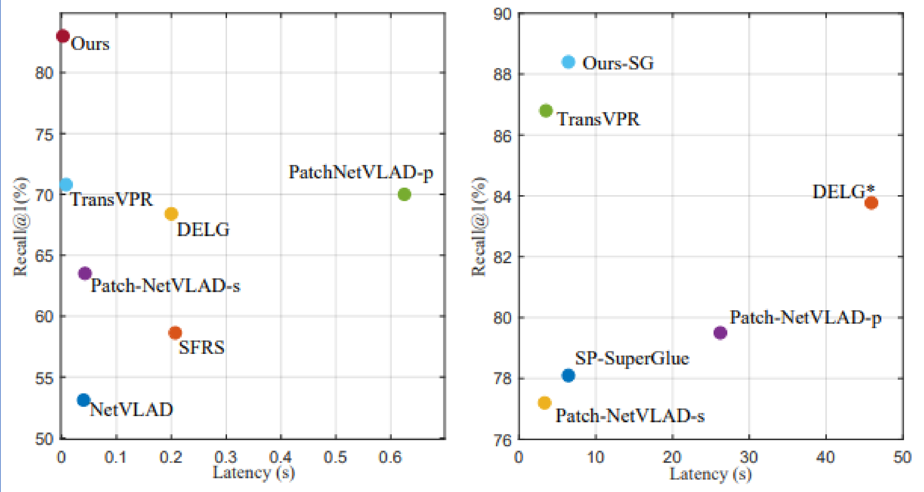
- We define a function to refine the weights on samples from two perspectives. One is the knowledge levels of the teacher on each sample; *the higher the knowledge level, the greater the weight.* Another is the knowledge gap between the teacher and the student; *the greater the gap, the greater the weight.*
- We adopt feature-based distillation loss in second-stage training.

Results

Comparison with SOTA

	Method	Venue	MSLS val			MSLS challenge			Nordland test			Pittsburgh30k test		
			R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10
Global retrieval	NetVLAD [1]	CVPR'16	53.1	66.5	71.1	28.6	38.3	42.9	11.5	17.6	21.9	81.9	91.2	93.7
	SFRS [15]	ECCV'20	58.8	68.2	71.8	30.7	39.3	42.9	14.3	23.1	27.3	71.1	81.0	84.9
	DELG [6]	ECCV'20	68.4	78.9	83.1	37.6	50.5	54.6	27.0	43.3	50.0	79.0	89.0	92.7
	Patch-NetVLAD-s [20]	CVPR'21	63.5	76.5	80.1	36.1	49.9	55.0	18.1	33.2	41.1	81.3	91.1	93.4
	Patch-NetVLAD-p [20]	CVPR'21	70.0	80.4	83.8	38.1	51.2	55.3	24.8	39.4	48.0	83.7	91.8	94.0
	TransVPR [51]	CVPR'22	70.8	85.1	89.6	48.0	67.1	73.6	31.3	53.6	64.8	73.8	88.1	91.9
	(A) Ours		**83.0**	**91.0**	**92.6**	**64.5**	**80.4**	**83.9**	**56.1**	**75.5**	**82.9**	**85.1**	**92.3**	**94.3**
Re-ranking	SP-SuperGlue [10,43]	CVPR'20	78.1	81.9	84.3	50.6	56.9	58.3	37.9	41.2	42.6	87.2	94.8	96.4
	DELG [6]	ECCV'20	83.9	89.2	90.1	56.5	65.7	68.3	64.4	70.8	72.7	89.9	95.4	96.7
	Patch-NetVLAD-s [20]	CVPR'21	77.2	85.4	87.3	48.1	59.4	62.3	50.9	62.7	66.5	88.0	94.5	95.6
	Patch-NetVLAD-p [20]	CVPR'21	79.5	86.2	87.7	51.2	60.3	63.9	62.7	71.0	73.5	88.7	94.5	95.9
	TransVPR [51]	CVPR'22	86.8	91.2	92.4	63.9	74.0	77.5	77.8	86.8	89.3	89.0	94.9	96.2
	(B) Ours-SP-RANSAC		87.3	91.4	92.8	65.5	76.3	81.3	76.8	86.3	90.1	89.4	95.2	96.5
	(B) Ours-SP-SuperGlue		**88.4**	**94.3**	**95.0**	**69.4**	**81.5**	**85.6**	**83.5**	**93.0**	**95.0**	**90.3**	**96.0**	**97.3**

Latency & Memory



(a) Global methods (b) Two-stage methods

Ablation study of distillation

Group for distillation	Sample ratio	MSLS val			MSLS challenge		
		R@1	R@5	R@10	R@1	R@5	R@10
None	0%	75.8	85.3	87.3	55.1	71.9	76.4
All	100%	78.4	87.4	90.1	59.1	73.5	79.3
$\mathcal{D}_1$	4.11%	78.8	89.2	91.1	**62.3**	76.8	80.9
$\mathcal{D}_2$	50.67%	80.0	88.8	90.6	59.1	75.5	79.3
$\mathcal{D}_3$	14.77%	77.7	87.4	89.2	57.8	74.4	79.1
$\mathcal{S}_1$	69.55%	81.6	88.9	91.2	62.2	77.2	82.0
$\mathcal{S}_2(\mathcal{D}_4)$	30.45%	71.6	83.7	85.1	48.2	63.9	68.7
$\mathcal{R}_1$	73.24%	79.3	87.6	88.9	57.0	74.1	78.2
$\mathcal{R}_2$	26.76%	73.8	82.7	85.6	51.0	67.0	72.3
Ours	69.55%	**83.0**	**91.0**	**92.6**	61.7	**79.3**	**83.3**

Visualizations of VPR results