Software for a better world

Maynooth University
National University of Ireland Maynooth

A World Leading SFI Research Centre

Science Foundation Ireland For what's next

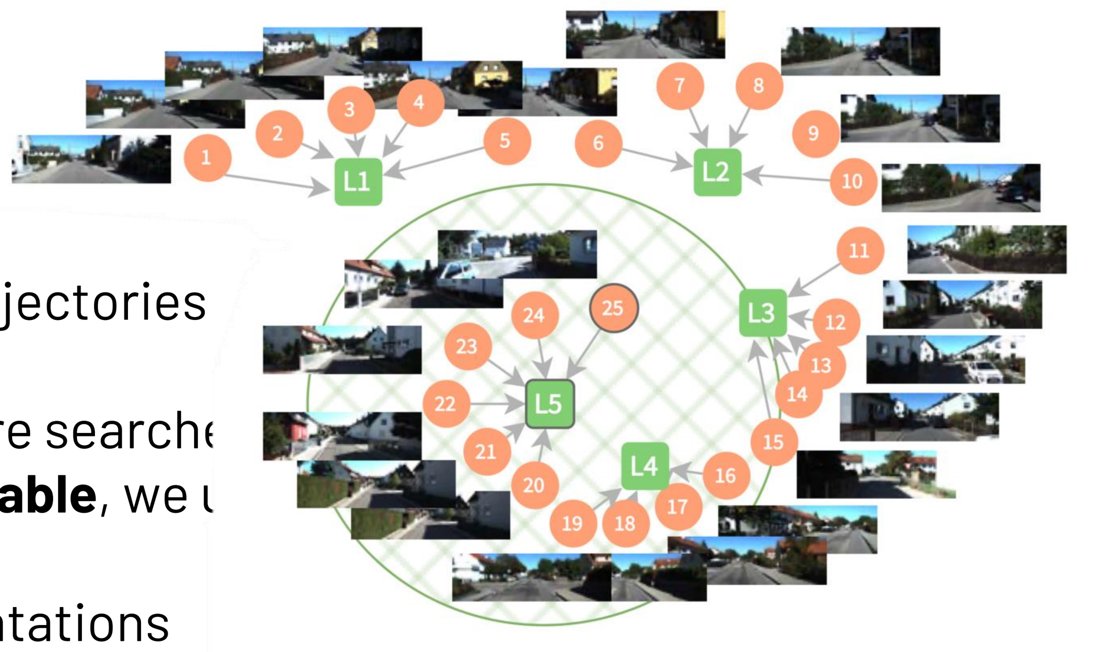
Scalable and Efficient Hierarchical Visual Topological Mapping

Saravanabalagi Ramachandran, Jonathan Horgan, Ganesh Sistu, John McDonald

1 INTRODUCTION:

Compared to metric m
topological maps
- Are simpler and mor
- Occupy less space
- Can map very long trajectories

To make the loop closure searche
more **efficient** and **scalable**, we u
- Indexing techniques
- Hierarchical representations



Hierarchical Mapping and Localization in the global descriptor embedding space
Orange Circle: Images, Outlined: Current
Green Square: Locations, Outlined: Current
Hatched marking: Search region for current image

We propose:

- The use of have **learned global descriptors** to group similar images into locations
- The **characteristics** of an ideal global descriptor for use in a hierarchical setting
- The use of compact learned global descriptors that excel in **continuity** and **distinctiveness** characteristics, as an efficient and scalable means for hierarchical topological mapping.

S. Ramachandran, J. Horgan, G. Sistu and J. McDonald, "Scalable and Efficient Hierarchical Visual Topological Mapping," *2023 21st International Conference on Advanced Robotics (ICAR)*, 2023, pp. 113-120, doi: 10.1109/ICAR58858.2023.10406394.

2 METHODOLOGY:

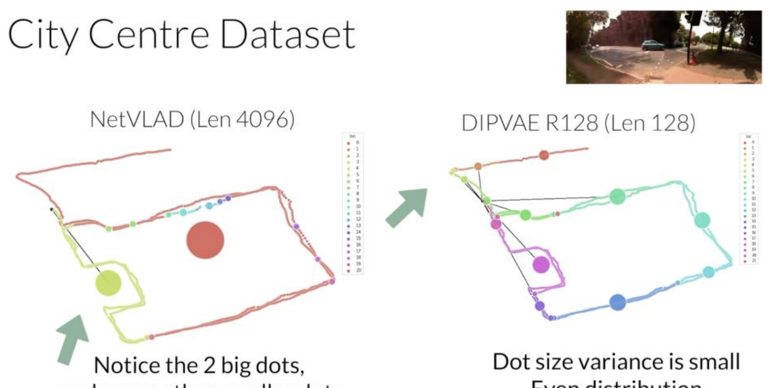
- We build on top of the existing HTMap technique: Garcia-Fidalgo *et al.* 2018
- We add optimization & improvements for **accurate** and **efficient** localization
- We adapt the framework for **learned global descriptors**

Through empirical analysis, we identify and define the **characteristics** of an **ideal** global descriptor for scalable and efficient visual localization

We perform extensive **analysis** on the **impact** of the global descriptor

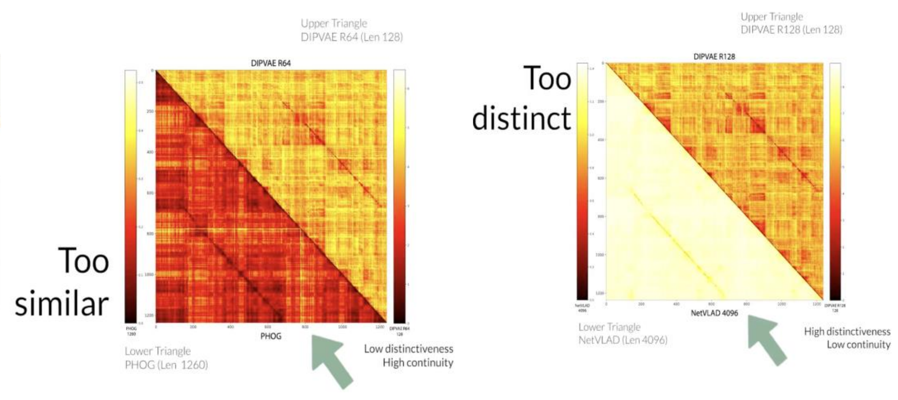
We present a methodology for **quantifying** and **contrasting** these characteristics

City Centre Dataset



Notice the 2 big dots, and many other smaller dots
Uneven distribution

Dot size variance is small
Even distribution



Distance matrices of different global descriptors for *City Centre* dataset with 20-25 locations shown as lower and upper triangular distance matrices along with corresponding distance scales on the sides.

Descriptor similarity decreases gradually as the robot/vehicle moves

Distinctiveness

Descriptor distance between images from different regions should be significantly larger than that to images from similar regions

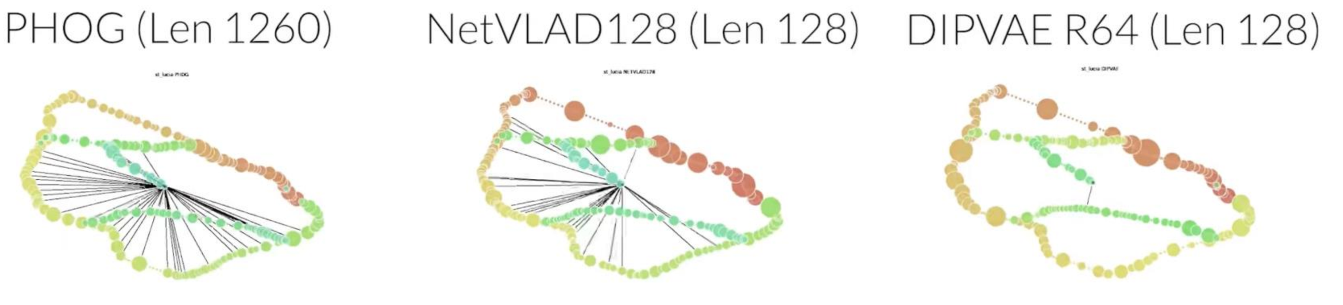
3 EVALUATION:

We **compare** hierarchical topological mapping technique with state-of-the-art global descriptors
- Hand-crafted: **PHOG**
- Learned:
  - Contrastive Supervised: **NetVLAD**
  - Semantic Supervised: **LoST**
  - VAE Unsupervised: **DIPVAE**

and perform extensive analysis on the **impact** of the global descriptor used

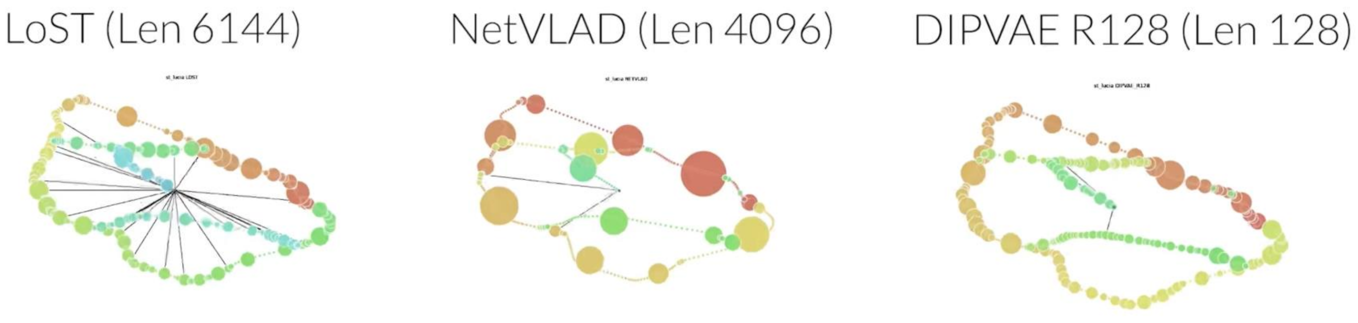
References:
PHOG: Bosch *et al.* 2007
NetVLAD: Arandjelovic *et al.* 2016
LoST: Garg *et al.* 2018
DIPVAE: Ramachandran *et al.* 2022

PHOG (Len 1260) NetVLAD128 (Len 128) DIPVAE R64 (Len 128)



St Lucia Dataset (17.6 km, 21,815 images)

LoST (Len 6144) NetVLAD (Len 4096) DIPVAE R128 (Len 128)



Hierarchical Topological Map using 6 different global descriptors.
Large dots: Locations, size proportional to number of images.
Small dots: Images with same color as the location they belong to

4 RESULTS:

Unsupervised learned VAE-based descriptor DIPVAE (both R64 and R128) variants excels in
- Distinctiveness
- Continuity

Characteristics and achieve:
- Significantly less total false positive locations
- Significantly lower compute time
- Significantly lower total mapping runtime
  - upto* **2.3x faster** than NetVLAD, less than half runtime
  - upto* **9.5x faster** than PHOG, close to one-tenth runtime
  - *maximized gains when trajectory is **longer** and consists of more locations
- While maintaining the same recall performance

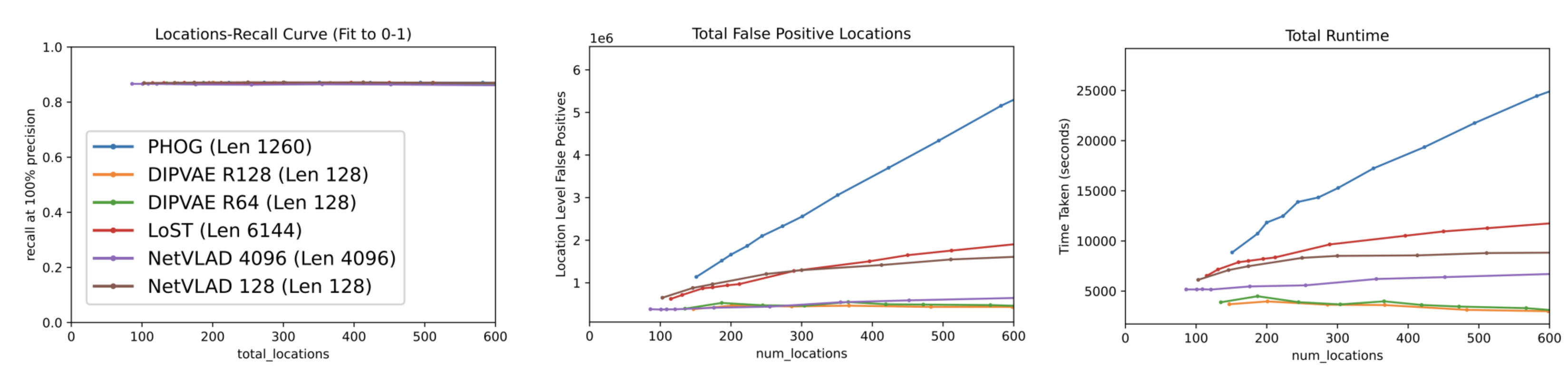
Watch Video

GDescriptor	n(Loc)	Runtime (s) ↓
PHOG	721	27939.47
LoST	654	12041.02
NetVLAD	624	6753.36
NetVLAD Cropped	698	8878.51
DIPVAE R128	665	**2967.95**
DIPVAE R64	609	3070.28

Runtime on *St Lucia* for various descriptors with the number of locations produced corresponding to that run.

GDescriptor	Length ↓	Type	Device	BSize ↑	Compute Time (s) ↓
PHOG	1260	Handcrafted	CPU	16	0.005455 | 0.007601
LoST	6144	Supervised	GPU	22	0.181668 | 0.231041
NetVLAD	4096	Supervised	GPU	22	0.027397 | 0.048229
NetVLAD Cropped	**128**	Supervised	GPU	22	0.027907 | 0.049699
DIPVAE R128	**128**	Unsupervised	GPU	1500	0.000008 | 0.000179
DIPVAE R64	**128**	Unsupervised	GPU	**6000**	**0.000002 | 0.000040**

Global Descriptors considered for evaluation. **Compute times** (per image) are reported for the specified max batch size *Bsize*, and for a batch size of 1 separated by a vertical bar.



Each dot represents one run with the line showing a series of runs of the corresponding global descriptor on *St Lucia*
*Results on other tracks (New College City Centre, KITTI 00, 05, and 06) are in the paper
**Total Mapping Runtime reported does not include descriptor compute time

HOST INSTITUTION
UNIVERSITY OF LIMERICK
OLLSCOIL LUIMNIGH

PARTNER INSTITUTIONS

FUNDED BY: