University of Essex

Efficiency Aware Multi-Technique Visual Place Recognition

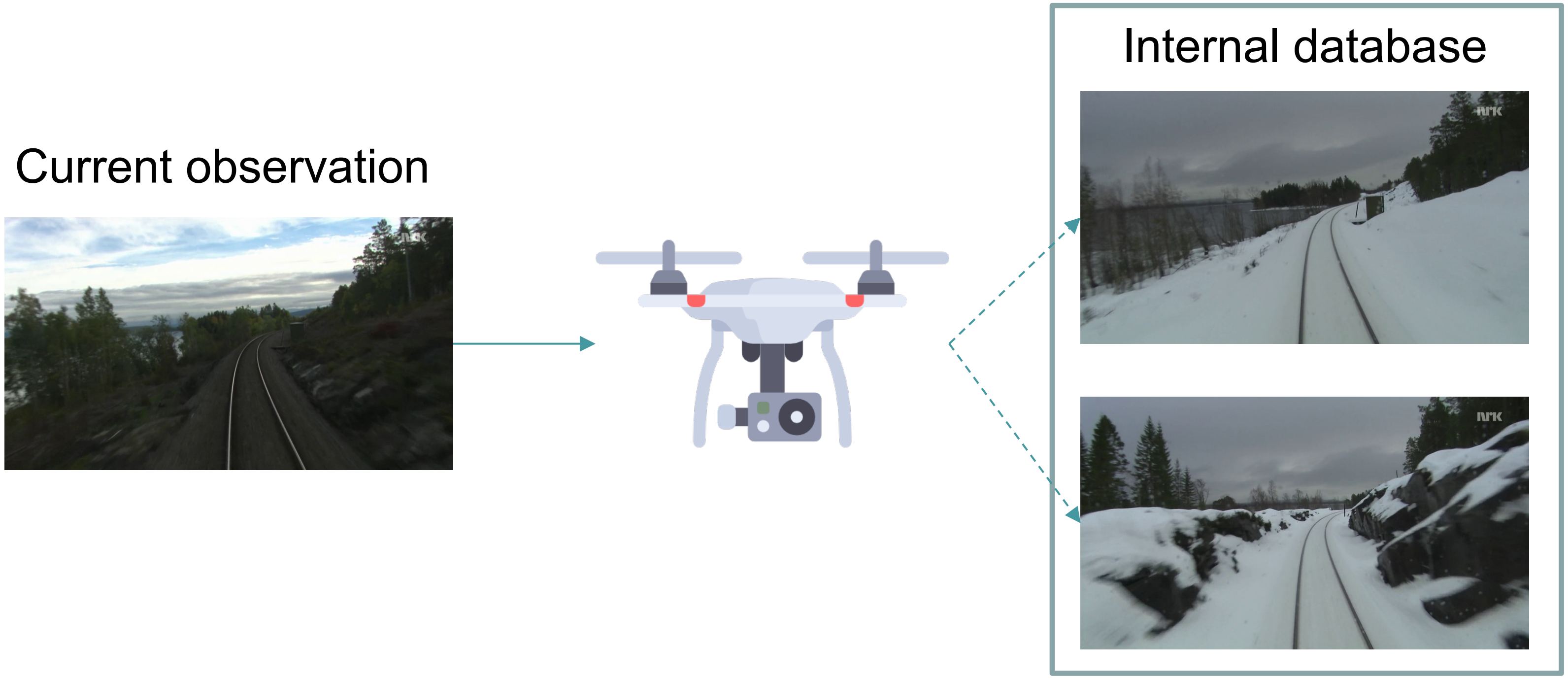
Presented By **Bruno Arcanjo**

Supervised By
Dr. Shoaib Ehsan
Prof. Klaus Mcdonald-Maier

26/04/2023

What Is VPR?

- **Definition: the ability for a system to recognize a previously seen place using only image information**



VPR Draws From Computer Vision and Robotics

Computer Vision

- Larger datasets
- High recall is preferred
- Focus on 3D model reconstruction

VPR

Robotics

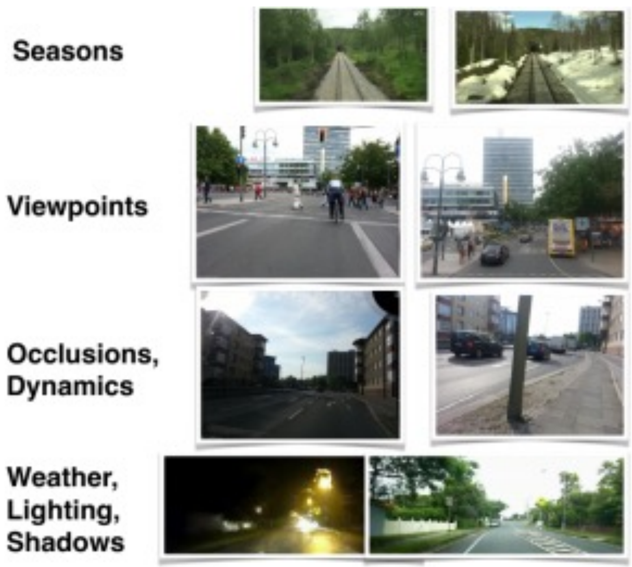
- Smaller datasets
- High precision is preferred
- Focus on loop-closure detection
- Efficiency is paramount

Why Is VPR Challenging?

Visual changes

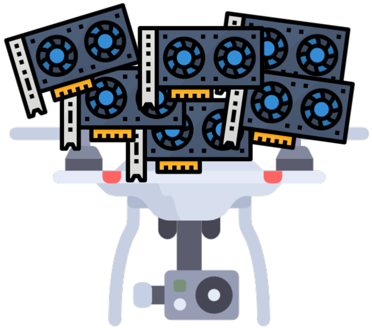
- Same place can look very different
  - Day/night cycles
  - Seasonal changes
  - Different viewpoints
  - Dynamic elements

- Two places can look the same
  - Especially in the same environment



Computational requirements

- Mobile robotics often operate with low-end hardware
- How to efficiently represent an image for matching?
- How to search an ever growing internal image database?



Several Specialist Techniques

	Pros	Cons
Technique 1	- Viewpoint - Efficient	- Appearance
Technique 2	- Appearance	- Viewpoint
Technique 3	- Viewpoint - Appearance	- Very expensive

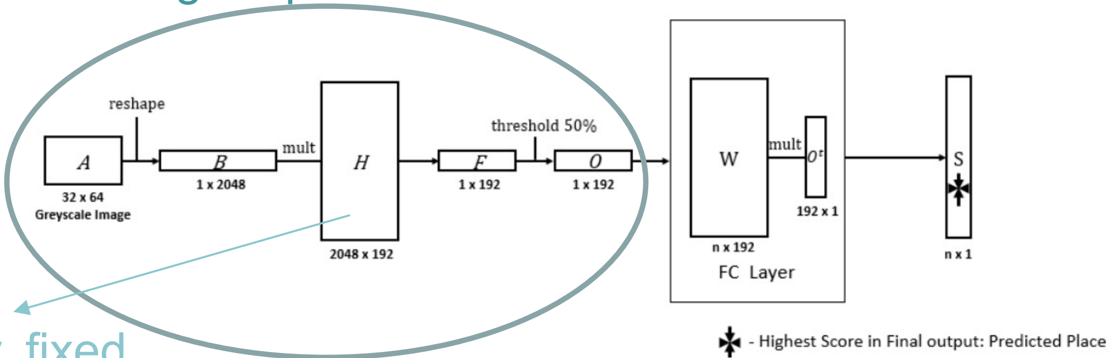
And thus, combining techniques was born

Combining Fly-Inspired Units By Voting

Baseline: Drosonet

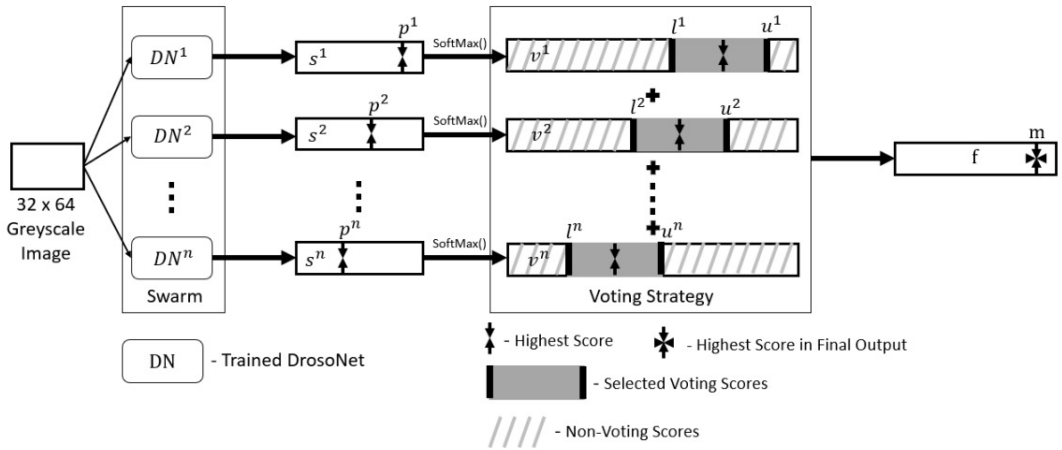
Sources of variation:
- Random H matrix
- Random FC inits.

Image representation



Binary, fixed

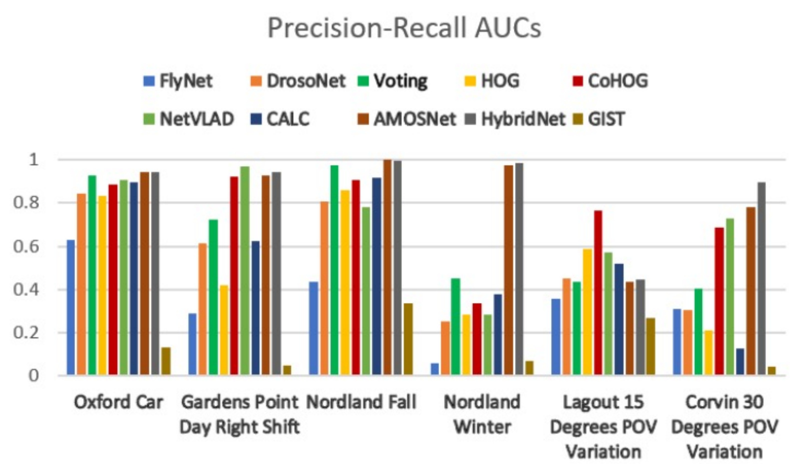
Combination Method: Voting



Voting Results

VPR Performance

Note: VPR performance is usually evaluated in terms of area-under-the precision recall curve (PR-AUC)



Computational Performance

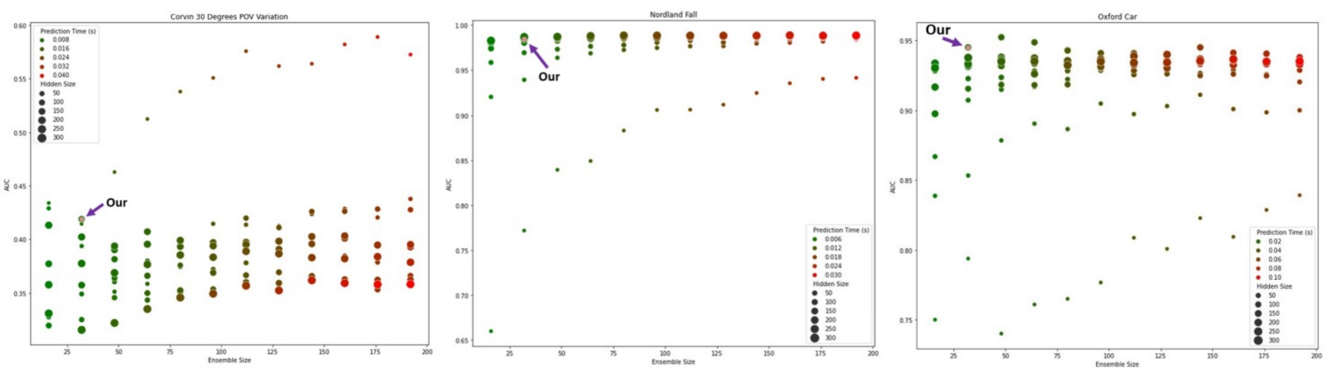
PREDICTION TIMES AND MEMORY USAGE COMPARISON

Model	Prediction time (ms)	FPS	Size (MBs)
HybridNet	1143.92	0.87	61.44
AMOSNet	1138.97	0.88	61.44
CoHOG	3627.18	0.28	123.01
CALC	73.62	13.58	4.26
GIST	225.04	4.44	4.53
HOG	208.71	4.79	142.88
NetVLAD	1435.11	0.70	16.38
FlyNet	1.00	1000	0.26
DrosoNet	1.00	1000	0.19
Voting	18.43	54.27	6.19

Voting Drawbacks

- Not so obvious hyperparameters…

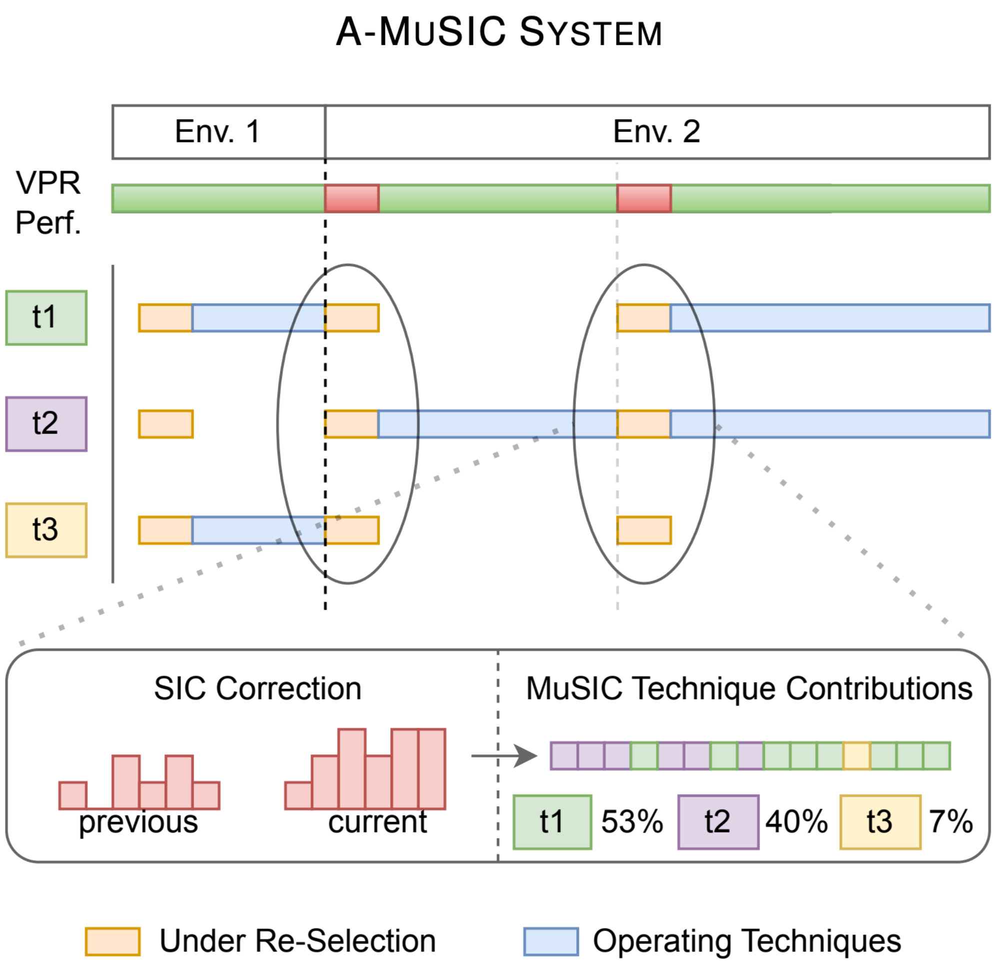
ABLATION STUDIES



(a) (b) (c)

- Huge performance disparities depending on random initialization of Drosonets
  - Requiring large number of Drosonets to guarantee performance (we used 32 baseline models in our work)

A-MuSIC: An Adaptive VPR Ensemble System

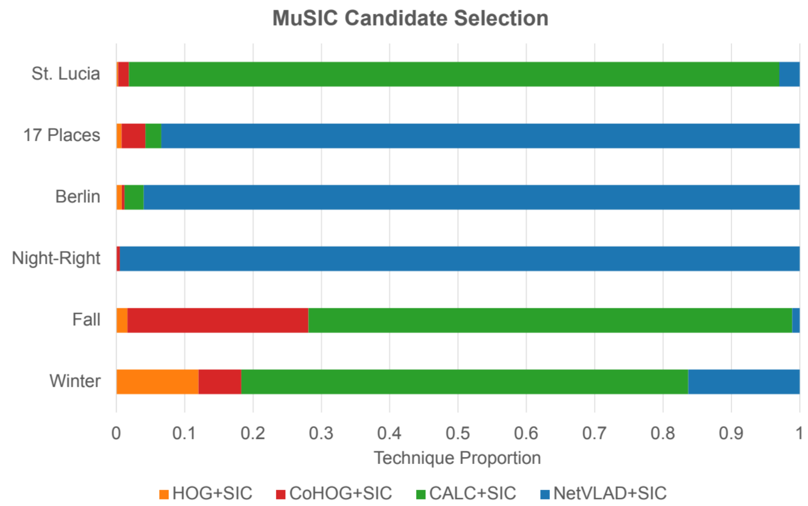


Self Identification and Correction (SIC)

Multi-SIC (MuSIC)

Adaptive-MuSIC (A-MuSIC)

A-MuSIC Results And Selection Pattern



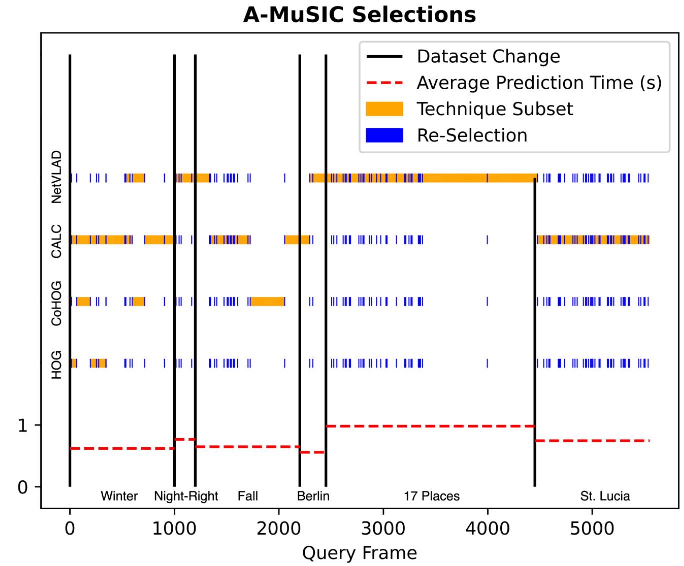
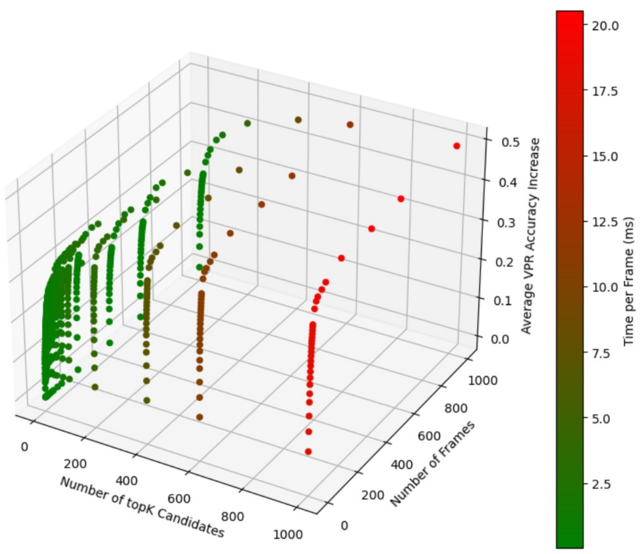


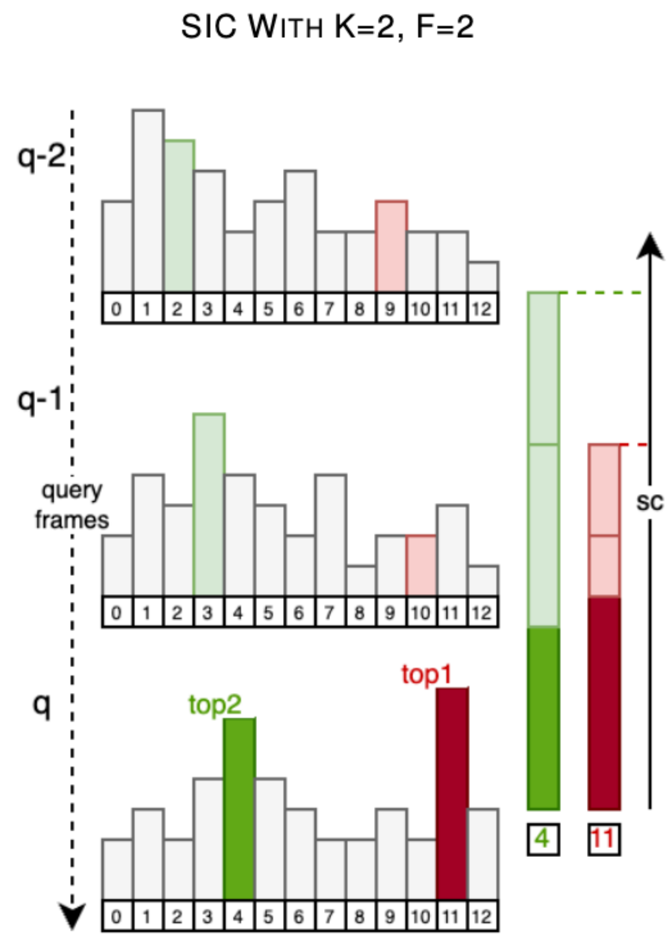
TABLE II: VPR Performance (AUC) and Prediction Time (ms)

	Winter		Fall		Berlin		Night-Right		17 Places		St. Lucia		Average	
	AUC	ms	AUC	ms	AUC	ms	AUC	ms	AUC	ms	AUC	ms	AUC	ms
HOG	0.29	49	0.84	48	0.03	16	0.03	15	0.34	96	0.56	47	0.35	45
CALC	0.30	168	0.88	168	0.06	157	0.13	161	0.38	170	0.48	165	0.37	165
CoHOG	0.23	795	0.85	785	0.28	225	0.45	185	0.31	1525	0.37	729	0.42	707
NetVLAD	0.28	723	0.68	719	0.81	722	0.54	717	0.48	727	0.35	721	0.52	722
HOG+SIC	0.76	54	0.98	55	0.02	20	0.40	17	0.46	104	0.84	52	0.58	50
CALC+SIC	0.86	172	0.99	175	0.56	163	0.71	163	0.67	177	0.92	170	0.79	170
CoHOG+SIC	0.67	747	0.99	776	0.80	218	0.87	184	0.76	1518	0.84	718	0.82	694
NetVLAD+SIC	0.78	747	0.94	743	0.96	736	0.97	742	0.82	745	0.85	717	0.89	738
MuSIC	0.95	1790	1.00	1748	0.95	1164	0.97	1094	0.86	2544	0.92	1937	0.94	1713
A-MuSIC	0.90	621	0.97	649	0.95	558	0.98	766	0.85	982	0.92	747	0.93	721

A-MuSIC Limitations

- Dependance on sequential imagery
- Still quite a lot of re-selection false positives
- Again, not so obvious hyperparameters





Future Work

- Extend DrosoNet and Voting to achieve better individual performance
  - DrosoNet currently works as a classifier, adapting it to work as a local descriptor could be useful for viewpoint resilience
  - A new fusion/voting/combination schema on this DrosoNet descriptor can be developed for better performance at still a low comp. cost

- Extend A-MuSIC selection decision to take computational cost into account
  - If deciding between two equally performing techniques, the less expensive one should be preferred
  - If deciding between two differently performing techniques, the tradeoff of performance-cost ratio should be a major factor