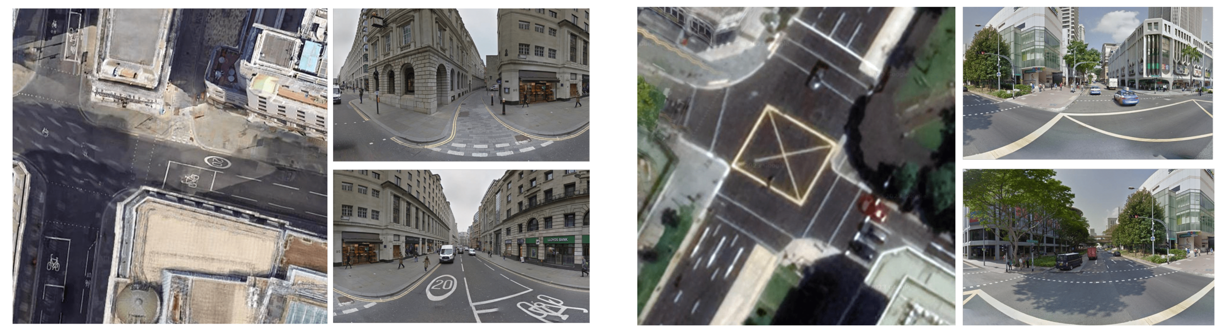
SpaGBOL: Spatial-Graph-Based Orientated Localisation

Tavis Shore [1] Oscar Mendez [2] Simon Hadfield [1]

[1]University of Surrey [2]Locus Robotics

Cross-View Geo-Localisation

CVGL - the task of identifying the geographic location of an ground-level image by matching it to a corresponding geo-referenced satellite image.



This is a challenging task due to significant variation in image features across viewpoints. To enhance the similarity of representations from corresponding images, datasets include paired samples globally. Techniques with varying panoramic Fields-of-View (FOV) have been developed to balance feasibility and performance.

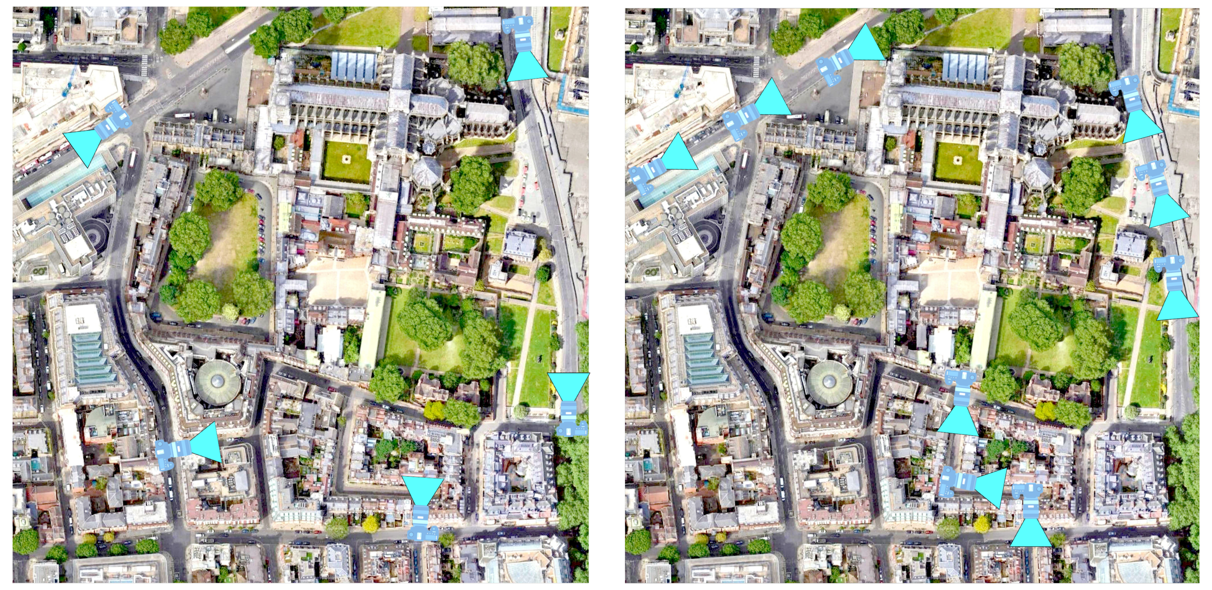


Figure 1. Previous CVGL Data Configurations - Sparse & Sequential

These datasets can be categorised into two types:

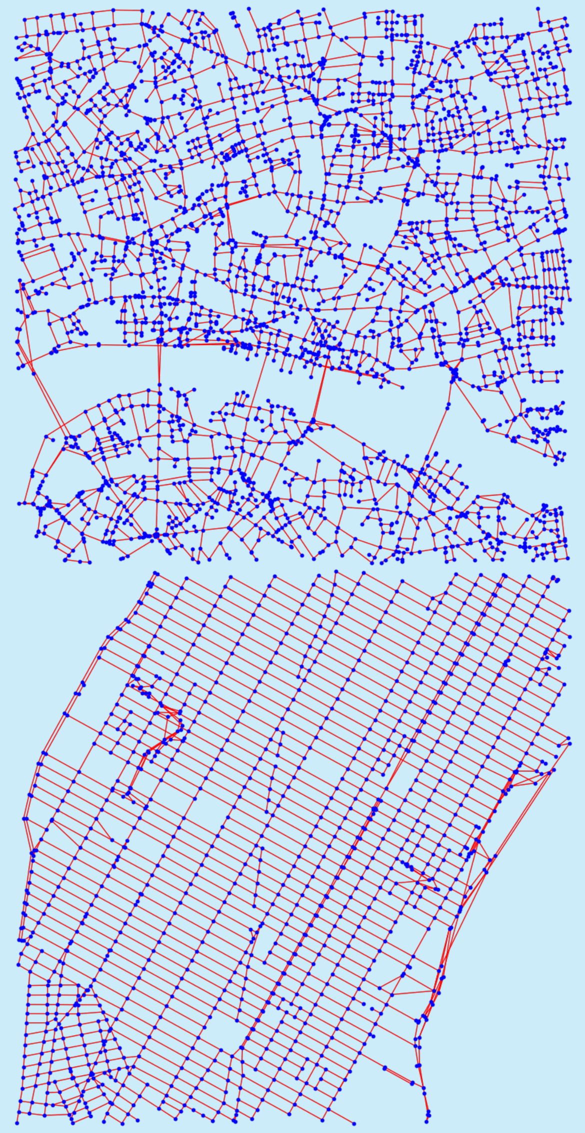
Sparsely Sampled

Image pairs are collected at scales ranging from city-wide to nationwide, with the only inherent relationship being individual streetview-satellite pairings.

Sequentially Sampled

Image sequences, sampled from videos, cover smaller areas than sparse datasets. The data has chronological order but no spatial relationships between samples.

SpaGBOL Graph-Based Representation



Global Distribution - Nodes and Edges

London	Tokyo	New York	Brussels	Singapore
3,155	4,815	1,103	2,190	1,043
4,124	7,942	1,983	3,403	1,567
Philadelphia	**Chicago**	**Hong Kong**	**Guildford**	**Boston**
2,272	1,159	995	1,472	1,567
3,782	1,935	1,440	1,773	2,403

Geo-Spatially Structured Data

Graph networks are based on city road networks, simplified to nodes at junctions:

- Nodes represent junctions; edges are connecting roads.
- Satellite images at nodes have a 0.22m/pixel resolution.
- Five panoramas per node capture varying seasons, lighting, and weather for enhanced learning.

Depth-First Graph Walk Sampling

Graph networks emulate vehicle movement via depth-first sampling, generating walks of configurable lengths. Longer walks improve precision but raise computational complexity. At inference, the database stores all reference walk embeddings.

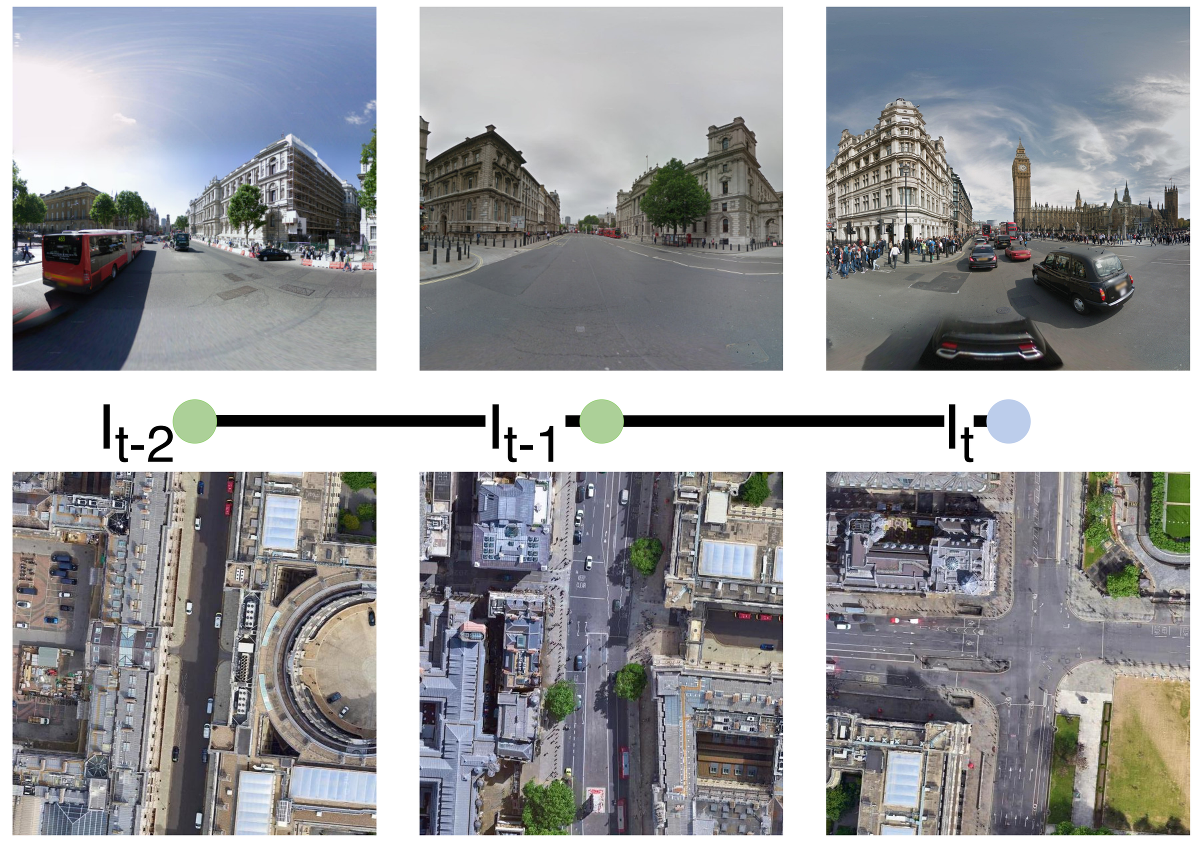


Figure 2. Depth-first walk across the City of London with length 3.

Bearing Vector Matching (BVM)

Orientations to neighbouring nodes are calculated and matched at query time to improve localisation. Road positions in panoramas are quantised into binary encoding for cross-view bearing matching.

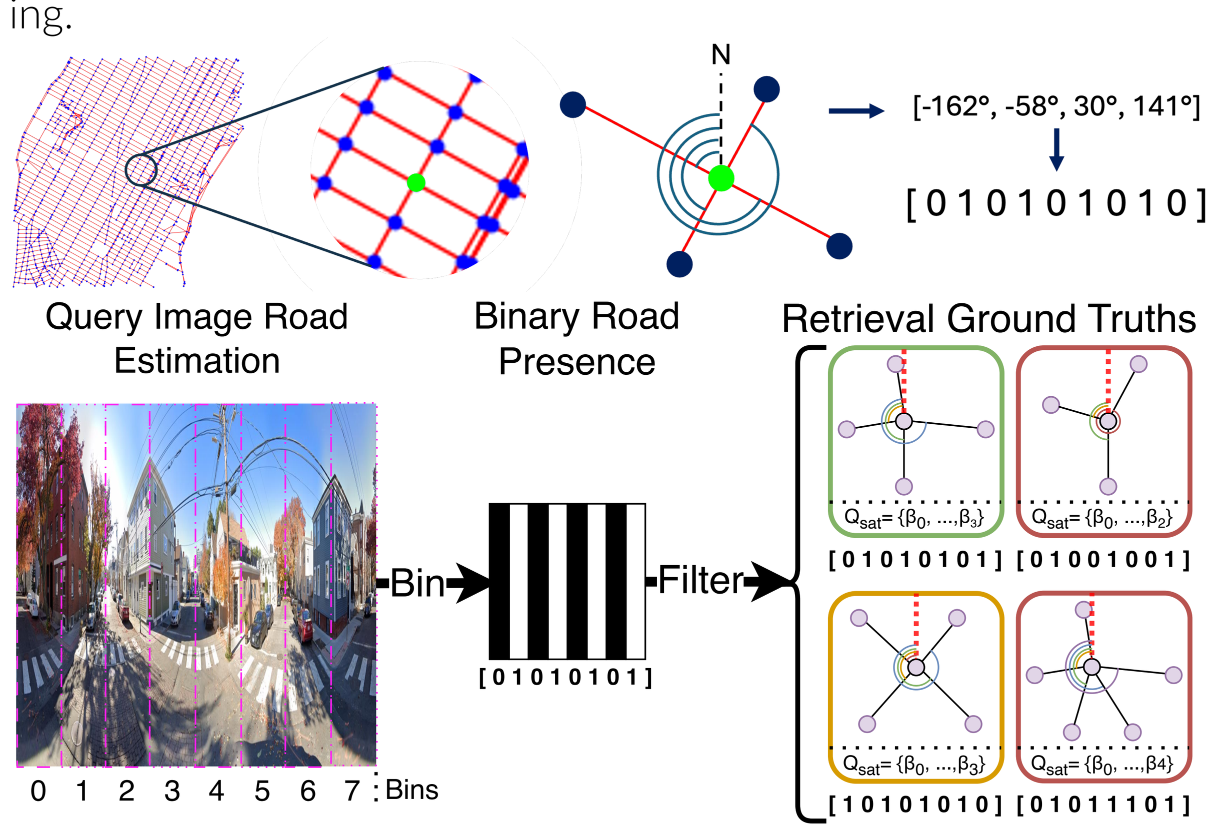


Figure 3. Calculating and estimating neighbour bearing → quantising into a binary encoding.

SpaGBOL CVGL Node Embedding Network

To leverage graph representations for CVGL, we introduce the first GNN-based two-branch network:

1. Extract image features from each walk image.
2. Assign features as *node features*.
3. Process the walk with the GNN to obtain refined, low-dimensional embeddings.

Training employs a triplet loss: the streetview walk as anchor, its corresponding satellite walk as positive, and a random satellite walk as negative.

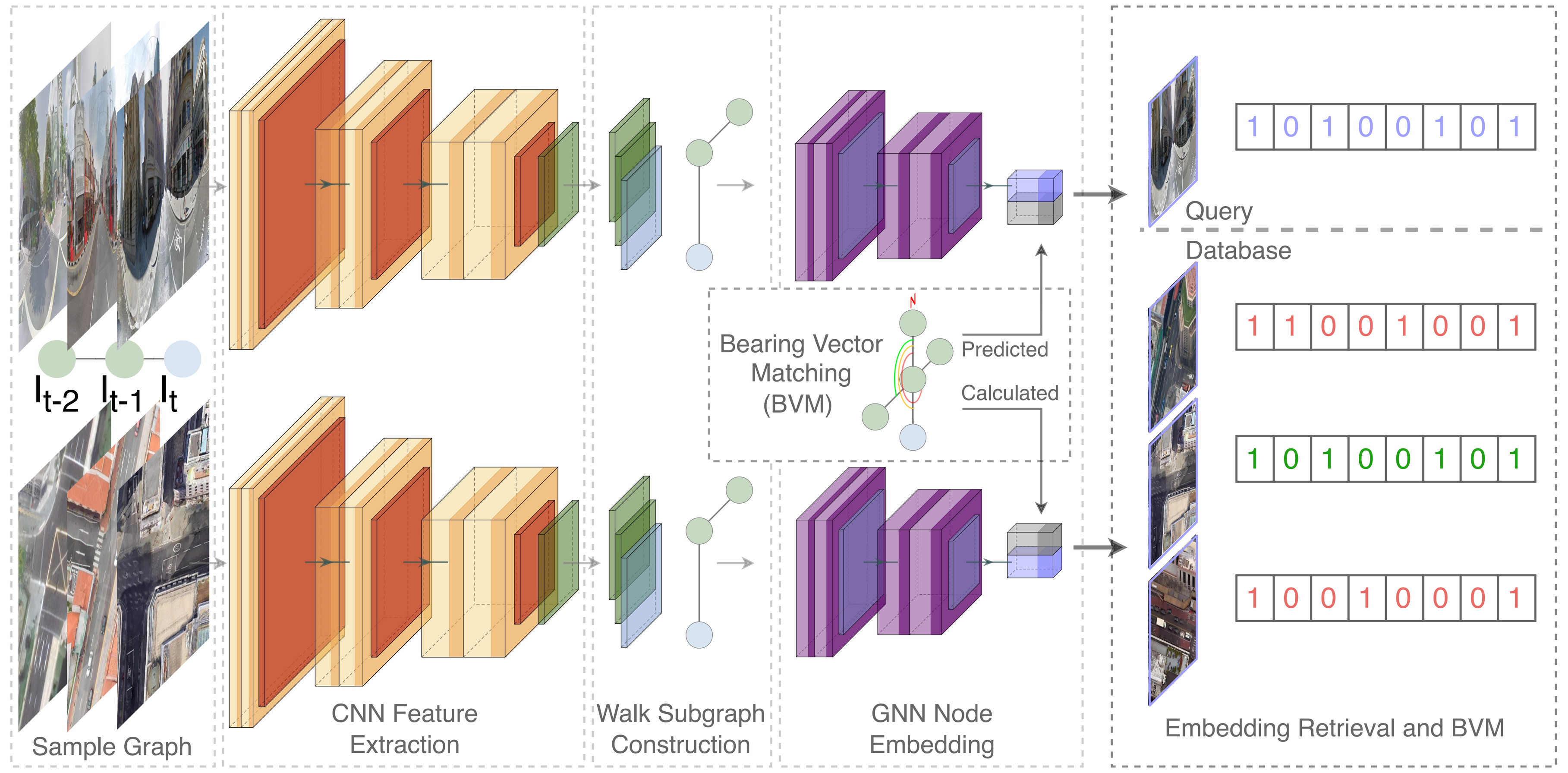


Figure 4. SpaGBOL Network: 2-branch CVGL network with no weight sharing

Evaluation - Top-K Recall Accuracy

- SpaGBOL achieves state-of-the-art performance with spatially strong node embeddings, improving Top-1 Recall by 11%.
- With BVM, performance rises by a further 26%.
- Adding a compass boosts this by 50%.

	360°				180°			
Model	Top-1	Top-5	Top-10	Top-1%	Top-1	Top-5	Top-10	Top-1%
L2LTR [1]	11.23	31.27	42.50	49.52	5.94	18.32	28.53	35.23
GeoDTR+ [2]	17.49	40.27	52.01	59.41	9.06	25.46	35.67	43.33
SAIG-D [3]	25.65	51.44	62.29	68.22	15.12	35.55	45.63	53.10
Sample4Geo [4]	50.80	74.22	79.96	82.32	37.52	**64.52**	71.92	76.39
SpaGBOL	**56.48**	**77.47**	**83.85**	**87.24**	**40.88**	63.79	**72.88**	**78.28**
SpaGBOL+B	64.01	86.54	92.09	94.64	52.01	82.20	89.47	93.62
SpaGBOL+YB	76.13	95.21	97.96	98.98	66.82	92.69	96.38	97.30

Table 1. Recall accuracies where edge-aligned streetview FOV $\theta \in \{360°, 180°\}$

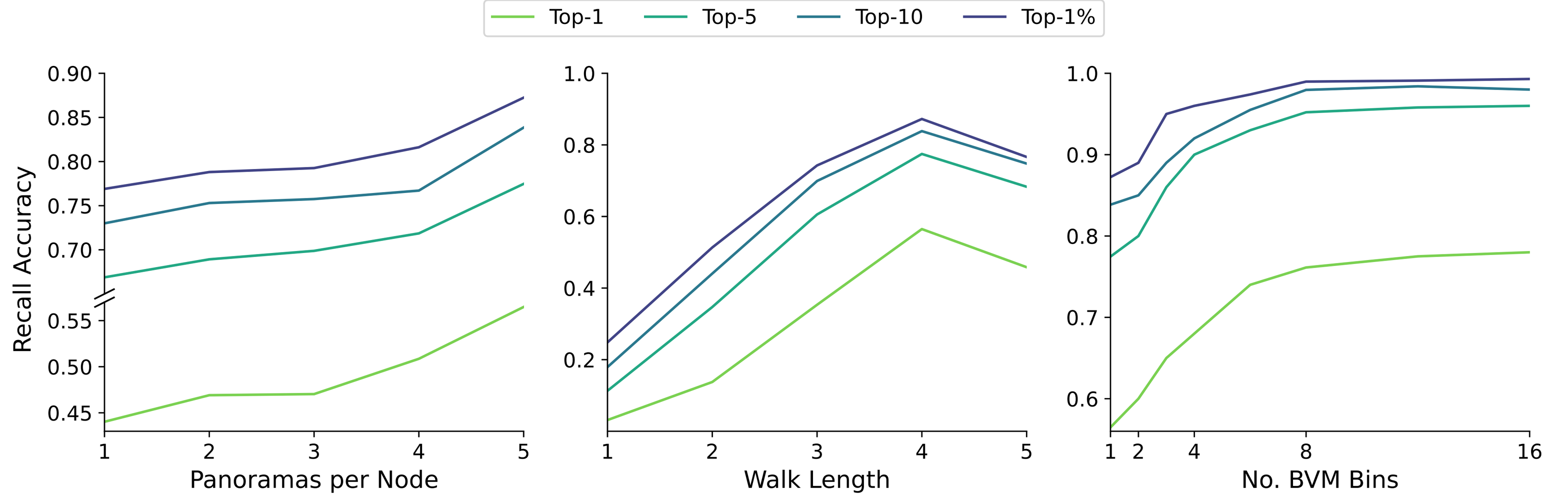


Figure 5. Impact of varying No. Panoramas, Walk length, and BVM granularity.

References

[1] H Yang, X Lu, and Y Zhu. Cross-view geo-localization with layer-to-layer transformer. In *NeurIPS*, 2021.

[2] X Zhang, X Li, W Sultani, C Chen, and S Wshah. Geodtr+: Toward generic cross-view geolocalization via geometric disentanglement. *IEEE TPAMI*, pages 1–19, 2024.

[3] Y Zhu, H Yang, Y Lu, and Q Huang. Simple, effective and general: A new backbone for cross-view image geo-localization, 2023.

[4] Fabian Deuser, Konrad Habel, and Norbert Oswald. Sample4geo: Hard negative sampling for cross-view geo-localisation. In *ICCV*, pages 16801–16810, 2023.

t.shore@surrey.ac.uk