Self-Supervised Landmark Discovery for Terrain-Relative Navigation

Caltech

Connor Lee, Esmir Mesic, and Soon-Jo Chung

ARCLab Autonomous Robotics and Control Lab

Background

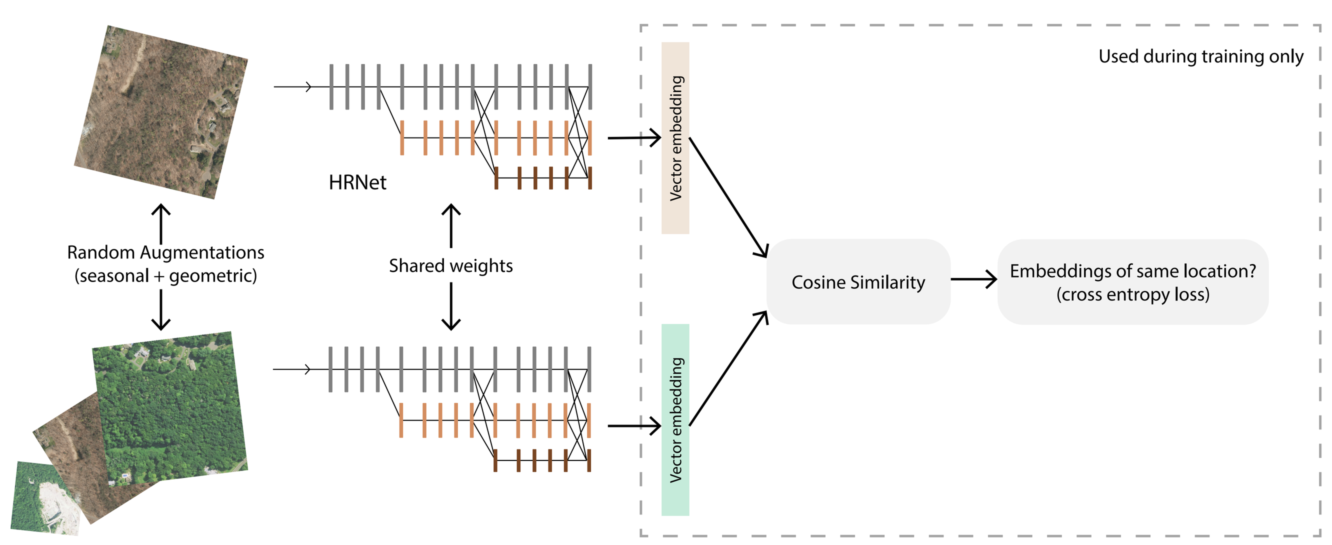
- Current visual terrain-relative navigation (VTRN) methods incur large **onboard storage costs** because they localize using onboard orthorectified imagery or dense encodings of a visual map [1, 2].
- Methods that use specific landmarks like craters for lunar exploration **require expert guidance**, limiting fast adoption in new areas [3, 4, 5].

Approach

Our method automatically discovers and re-identifies sparse, useful navigation landmarks via self-supervised contrastive learning (SSCL).

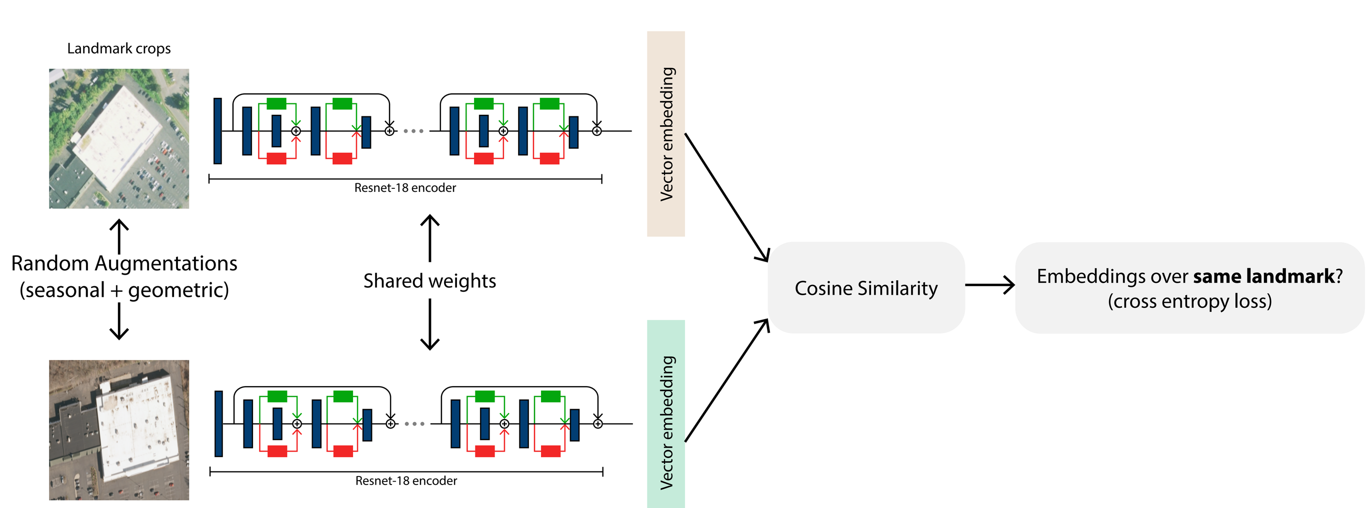
Step 1: Landmark Discovery

- **Training:** An HRNet network is trained on matching (same location) and non-matching image pairs using SSCL.
- **Inference:** Threshold network activations to find landmarks.



Step 2: Landmark Encoding

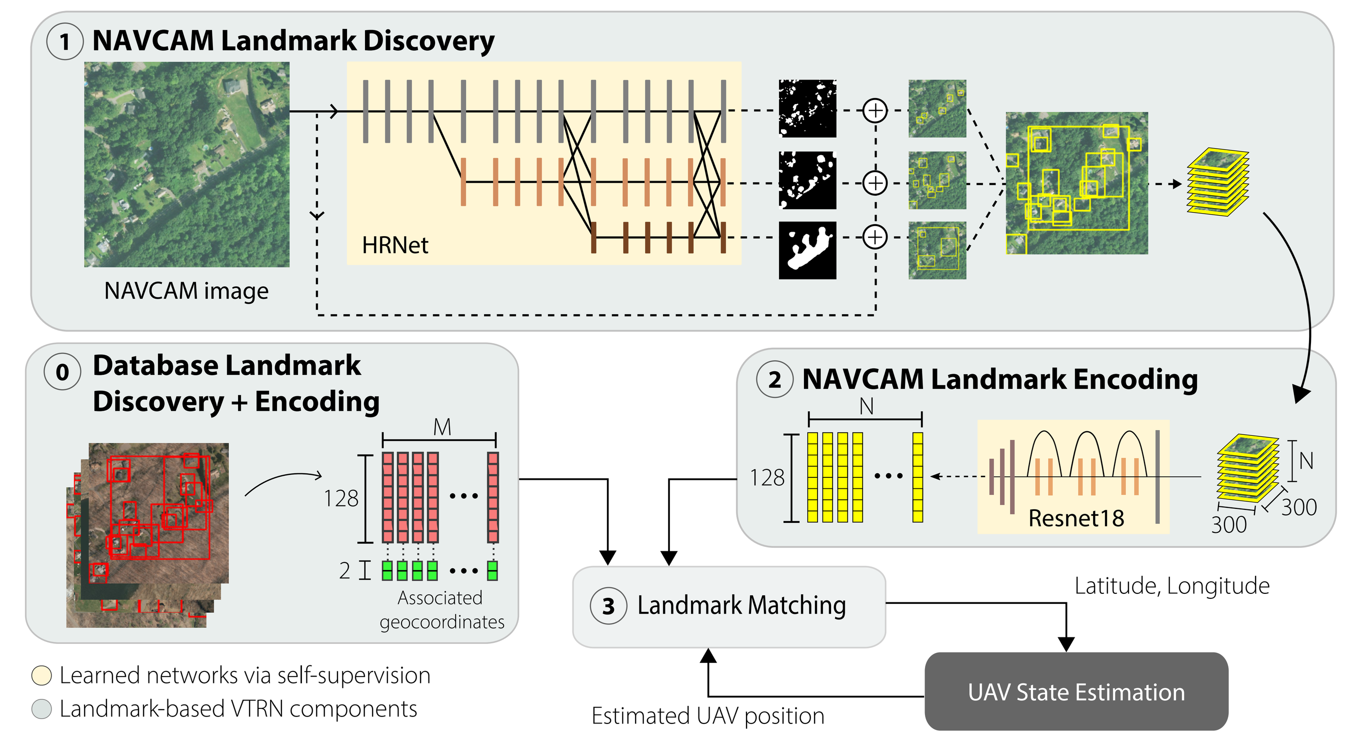
- **Training:** ResNet-18 encoder is trained using the same SSCL scheme on cropped, landmark pairs (discovered in previous step).
- **Inference:** Network encodes found landmarks as 128-d vectors for lightweight caching and fast matching.



Step 3: Landmark Matching

- **Training:** None
- **Inference:** A landmark pair is a match if: (1) it has the maximum cosine similarity among all possible pairs and (2) its similarity exceeds a certain threshold τ.

Combined VTRN Architecture



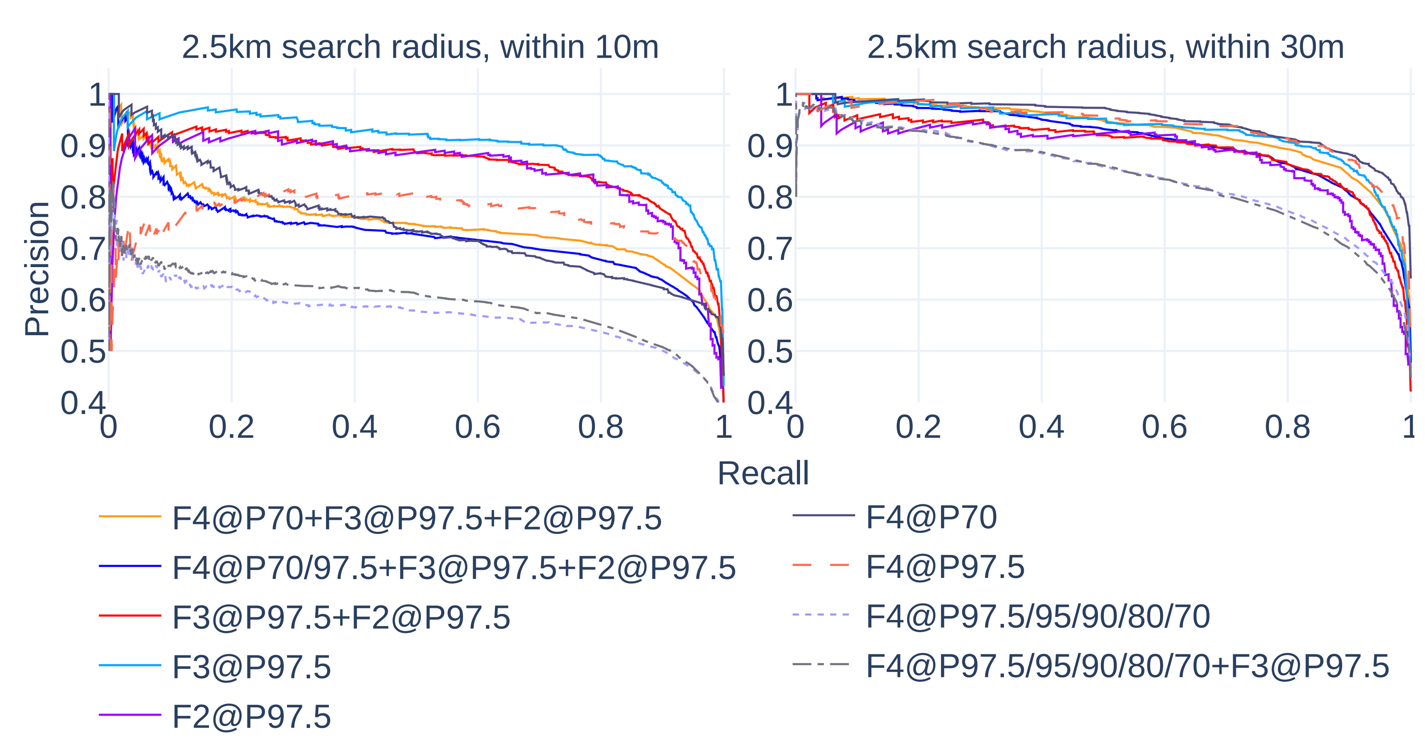
Dataset

- 3639 coregistered image pairs over Connecticut, USA from Spring/Summer 2016.
- Image: 1270×1270 pixels at 0.6 m/pixel
- "Leaf-on vs. leaf-off" seasonal variation
- Contains "obvious" landmarks (houses, buildings, roads, etc...) for easy algorithm validation

Results

Evaluation procedure:

- Evaluate matching performance of VTRN pipeline via precision-recall analysis and use database search radius of 2.5 km.
- Proposed match = landmark pair with max. cosine similarity that exceeds τ.
- Consider ground truth match for landmarks within 10 and 30 m.

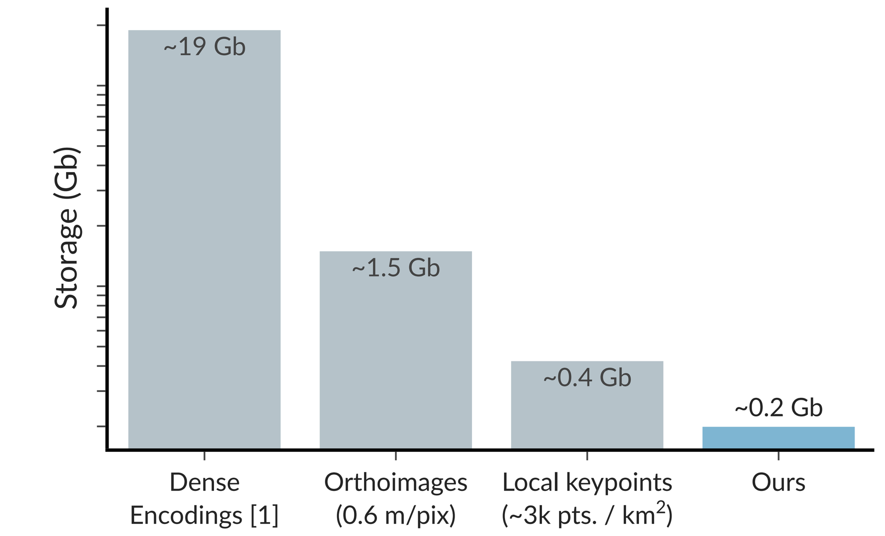


Findings:

- Precise localization (< 10m) with small landmarks (F2/F3@P97.5) is best.
- Large landmarks (F4@P70/97.5) are useful when fine-grained location is not needed (< 30m)

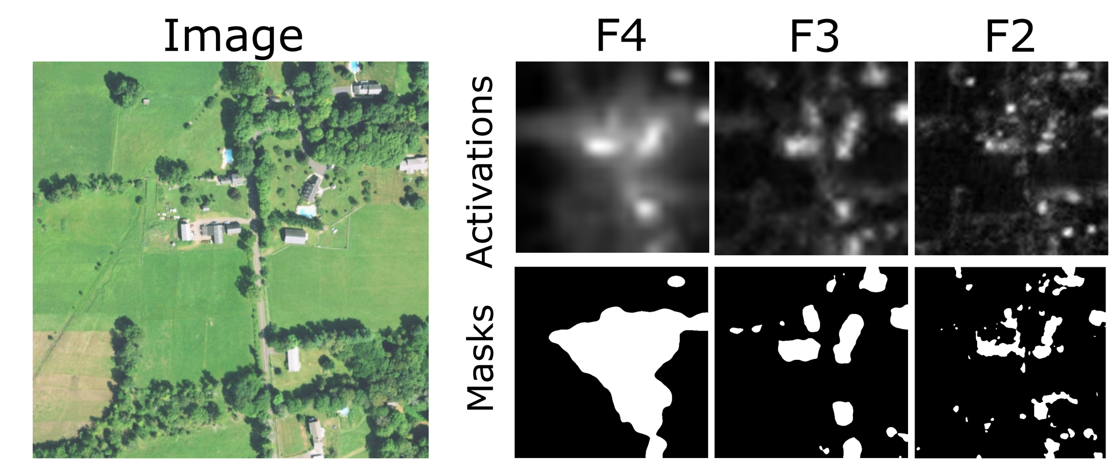
Onboard storage comparison over Salisbury, Connecticut

- 155 km^2 of suburbs, farmland, and forests
- Our method finds useful landmarks and ignores useless ones over uniform areas like forests.

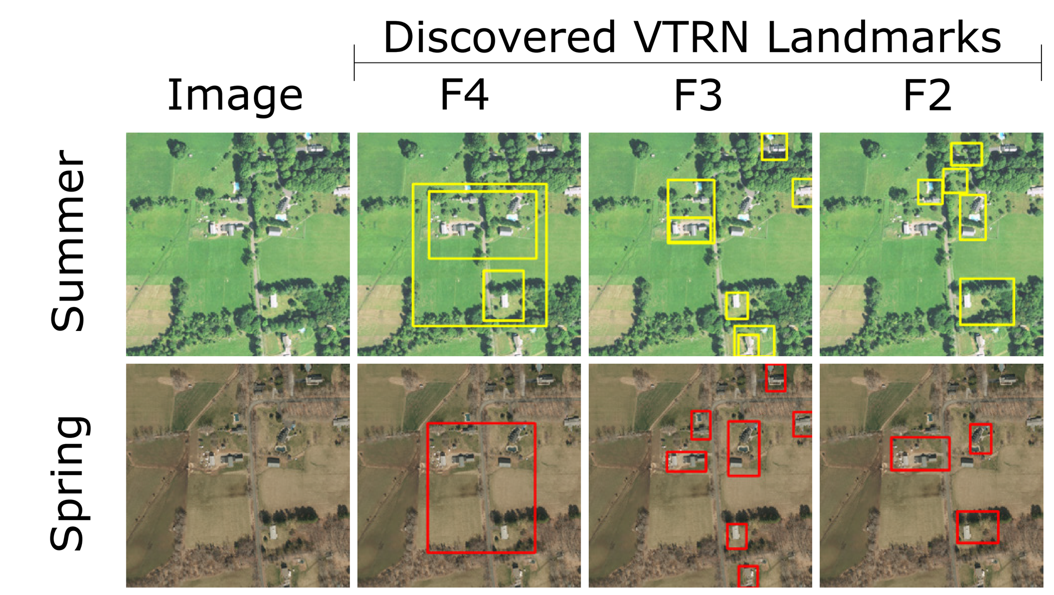


Additional Visuals

Raw HRNet activations w/ binarized masks



Landmarks discovered via contour detection on binary masks



- Landmarks shown w/o non-max-suppression.
- Same landmarks may be found in different resolution streams.

Conclusions

- Self-supervised contrastive learning can help find and encode optimal landmarks for aerial localization without human guidance
- Requires less storage compared to VTRN methods that densely encode images by discretizing map.
- *Future work:* fine-grained localization precision, extend to rugged terrain, winter scenes, and leverage position variances from state estimation

References

[1] M. Bianchi and T. D. Barfoot. Uav localization using autoencoded satellite images. *IEEE Robot. Autom. Lett.*, 6(2):1761–1768, 2021.

[2] S. Chen, X. Wu, M. W. Mueller, and K. Sreenath. Real-time geo-localization using satellite imagery and topography for unmanned aerial vehicles. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots and Syst.*, pages 2275–2281, 2021.

[3] L. Downes, T. J. Steiner, and J. P. How. Deep learning crater detection for lunar terrain relative navigation. In *AIAA SciTech 2020 Forum*, page 1838, 2020.

[4] A. Nassar, K. Amer, R. ElHakim, and M. ElHelw. A deep cnn-based framework for enhanced aerial imagery registration with applications to uav geolocalization. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog. workshops*, 2018.

[5] T. Wang, Y. Zhao, J. Wang, A. K. Somani, and C. Sun. Attention-based road registration for gps-denied uas navigation. *IEEE Trans. Neural Netw. Learn. Syst.*, 32(4):1788–1800, 2021. doi:10.1109/TNNLS.2020.3015660.

ICRA LONDON · 2023