

MEMORY PROXY MAPS FOR VISUAL NAVIGATION

Anonymous authors

Paper under double-blind review

Related Work	No RL	No Sim	No Graph	No Odo.	No Metric Map	Training Cost	Sensors	View
RNR-Map	✓	✓	✓	✗	✗	-	RGBD	Narrow
ZSEL	✗	✗	✓	✓	✓	500M images	RGB	Narrow
OVRL	✗	✗	✓	✗	✓	53 GPU days	RGBD+GPS+Comp.	Narrow
TSGM	✗	✗	✗	✓	✓	-	RGBD	Panoramic
RL	✗	✗	✓	✗	✓	3.5M+ imgs/100M steps	RGBD	Narrow
BC	✓	✓	✓	✗	✓	-	RGBD	Narrow
NRNS	✓	✓	✗	✗	✓	3.5M imgs/10M steps	RGBD	Narrow
FeudalNav (Ours)	✓	✓	✓	✓	✓	37K img/3M steps	RGBD	Narrow

Table 1: Key differences of training and architecture characteristics between SOTA and our method. A check mark ✓ means that the method has the corresponding quality and an ✗ indicates that it does not. We provide a more lightweight approach that uses no reinforcement learning, no graph, no odometry (odo), doesn’t train in a simulator (no sim), and learns no metric map.

1 RELATED WORK

We provide a comparison of the training and model statistics for our method and other SOTA work in Table 1. We compare to RNR-Map Kwon et al. (2023), ZSEL Al-Halah et al. (2022), OVRL Yadav et al. (2022), TSGM Kim et al. (2023), RL/DDPO Wijmans et al. (2019), and NRNS and its behavior cloning baseline Hahn et al. (2021). Note that ours is by far the most lightweight approach that doesn’t use RL, doesn’t train in a simulator, uses no graph or odometry, and learns no metric map. Additionally, we achieve SOTA performance on a significantly smaller fraction of the data used by other methods. We choose to compare our work to NRNS because it is the most similar to our approach.

2 MODEL TRAINING AND TESTING DATA

For training, we use the following Gibson Xia et al. (2018) environments: Ackermanville, Adrian, Airport, Albertville, Aldrich, Alfred, Allensville, Almena, Almota, Aloha, Alstown, American, Anaheim, Anchor, Andover, Angiola, Annawan, Annona, Anthoston, Apache, Applewold, Arbutus, Archer, Arkansaw, Arona, Artois, Ashport, Assinippi, Athens, Auburn, Aulander, Avonia, Azusa, Ballantine, Ballou, Baneberry, Barboursville, Barranquitas, Bautista, Beach, Beechwood, Bellwood, Benevolence, Bethlehem, Bettendorf, Biltmore, Blackstone, Bohemia, Bolton, Bonesteel, Bonnie, Booth, Bountiful, Bowlus, Bowmore, Branford, Braxton, Bremerton, Brevort, Brewton, Broseley, Brown, Browntown, Burién, Bushong, Byers, Callicoon, Calmar, Capistrano, Carpendale, Carpio, Castor, Castorville, Channel, Checotah, Chesaning, Chesterbrook, Chilhowie, Chiloquin, Chireno, Chuchton, Circleville, Cisne, Clive, Cobalt, Codell, Coffeen, Collierville, Convoy, Cooperstown, Copemish, Cottonport, Country, Crandon, Crookston, Culbertson, Cullison, Cutlerville, Dansville, Darden, Darnsetown, Deatsville, Dedham, Deemston, Delton, Dryville, Duarte, Duluth, Eagan, Eagerville, Elton, Emmaus, Euharlee, Everton, and Ewell.

For test, we evaluate our model in Cantwell, Eastville, Elmira, Greigsville, Pablo, Sands, Sisters, Denmark, Edgemere, Eudora, Mosquito, Ribera, Scioto, and Swarmville using the test trajectories provided by NRNS Hahn et al. (2021) here (<https://meerahahn.github.io/nrns/data>).

3 FEUDALNAV TEST PERFORMANCE EXAMPLES

We provide videos showing FeudalNav performance on each category from the test set specified in NRNS Hahn et al. (2021) in file SOTAtrials.mp4.

REFERENCES

- Ziad Al-Halah, Santhosh Kumar Ramakrishnan, and Kristen Grauman. Zero experience required: Plug & play modular transfer learning for semantic visual navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17031–17041, 2022.
- Meera Hahn, Devendra Singh Chaplot, Shubham Tulsiani, Mustafa Mukadam, James M Rehg, and Abhinav Gupta. No rl, no simulation: Learning to navigate without navigating. *Advances in Neural Information Processing Systems*, 34:26661–26673, 2021.
- Nuri Kim, Obin Kwon, Hwiyeon Yoo, Yunho Choi, Jeongho Park, and Songhwai Oh. Topological Semantic Graph Memory for Image-Goal Navigation. In *Conference on Robot Learning*, pp. 393–402. PMLR, 2023.
- Obin Kwon, Jeongho Park, and Songhwai Oh. Renderable Neural Radiance Map for Visual Navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9099–9108, 2023.
- Erik Wijmans, Abhishek Kadian, Ari Morcos, Stefan Lee, Irfan Essa, Devi Parikh, Manolis Savva, and Dhruv Batra. Dd-ppo: Learning near-perfect pointgoal navigators from 2.5 billion frames. *International Conference on Learning Representations*, 2019.
- Fei Xia, Amir R. Zamir, Zhi-Yang He, Alexander Sax, Jitendra Malik, and Silvio Savarese. Gibson env: real-world perception for embodied agents. In *Computer Vision and Pattern Recognition (CVPR), 2018 IEEE Conference on*. IEEE, 2018.
- Karmesh Yadav, Ram Ramrakhya, Arjun Majumdar, Vincent-Pierre Berges, Sachit Kuhar, Dhruv Batra, Alexei Baevski, and Oleksandr Maksymets. Offline visual representation learning for embodied navigation. *arXiv preprint arXiv:2204.13226*, 2022.