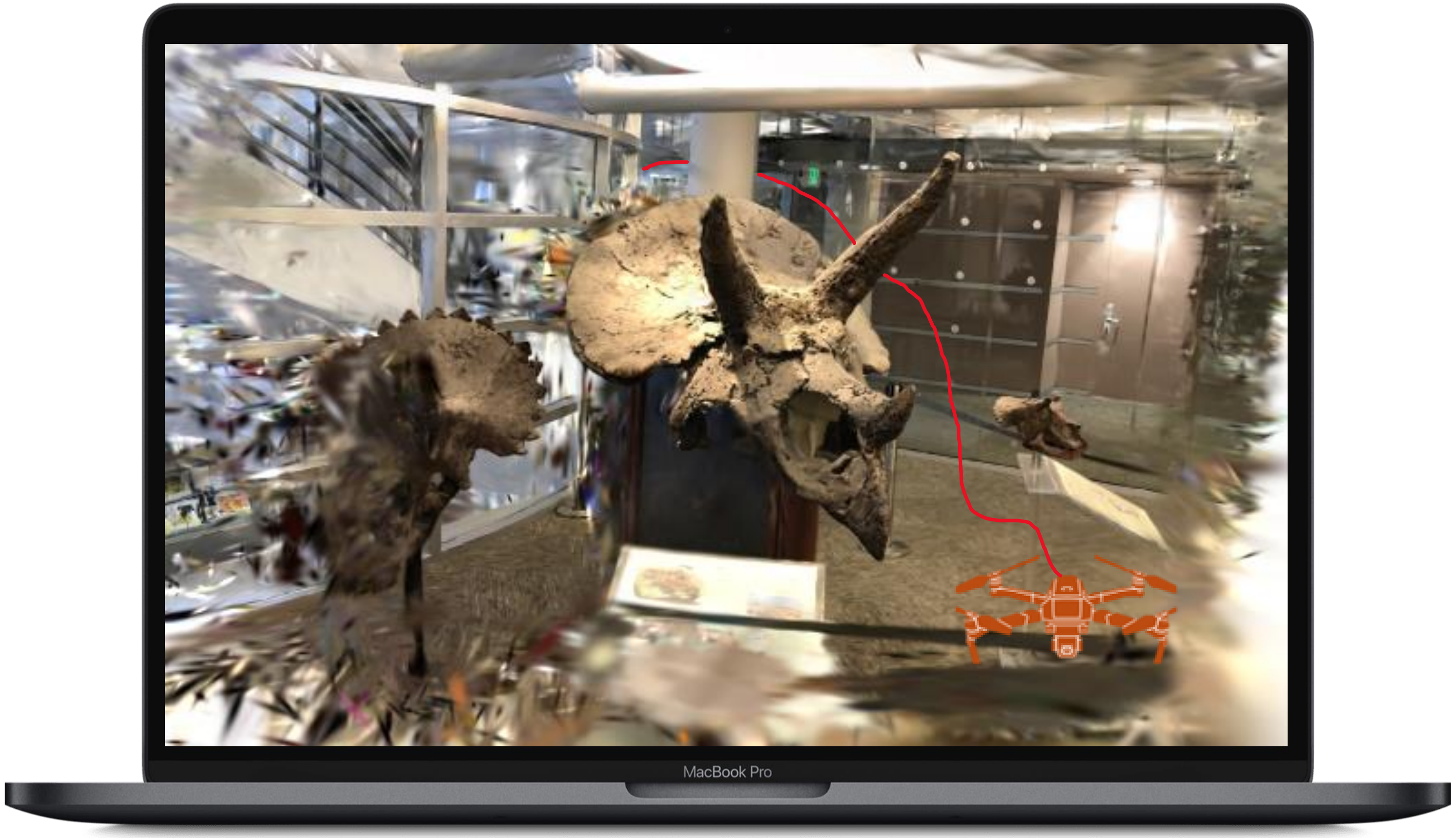
Robot Navigation in Gaussian Splatting Scenes



Luca Mattes Wiehe

IN2107: Robot Perception and Intelligence
Supervisor: Jiaxin Wei
Date: 02 December 2024

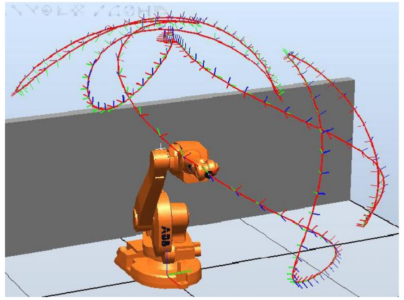
Motivation: 3D Visual Navigation

1. 3D adds crucial information
2. 3D Gaussians as efficient way to model 3D scenes
3. Don't only reconstruct, navigate through scene
4. 3D Navigation needs to be safe and accurate

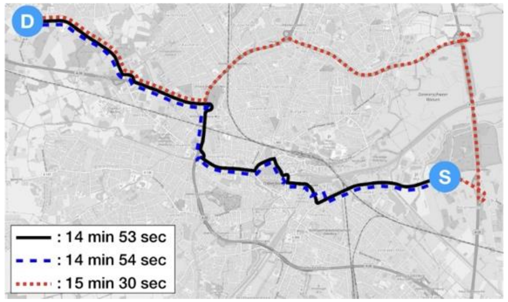
3D Visual Navigation is a difficult problem!

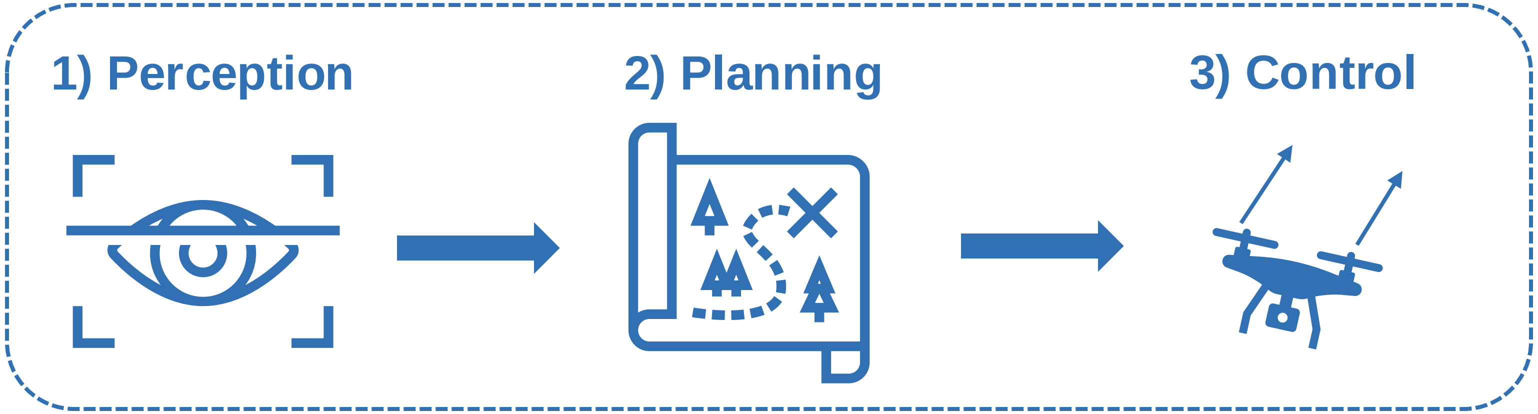
Autonomous Exploration

Robotics



Efficient Path Planning





Visual Navigation Pipeline

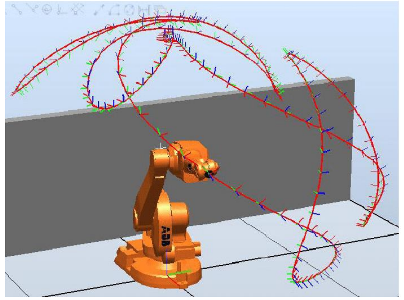
Motivation: 3D Visual Navigation

1. 3D adds crucial information
2. 3D Gaussians as efficient way to model 3D scenes
3. Don’t only reconstruct, navigate through scene
4. 3D Navigation needs to be safe and accurate

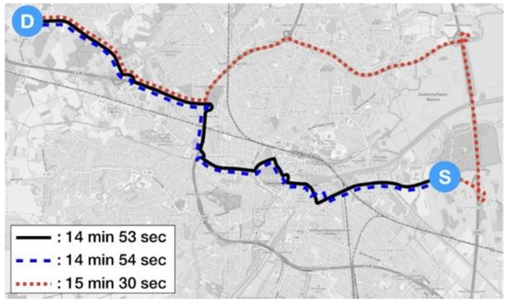
3D Visual Navigation is a difficult problem!

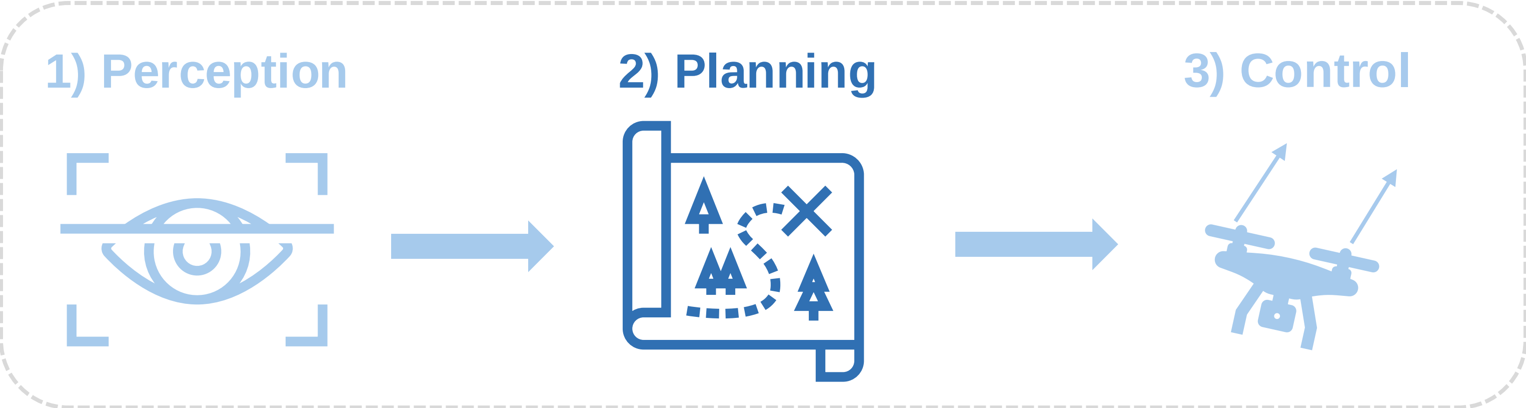
Autonomous Exploration

Robotics



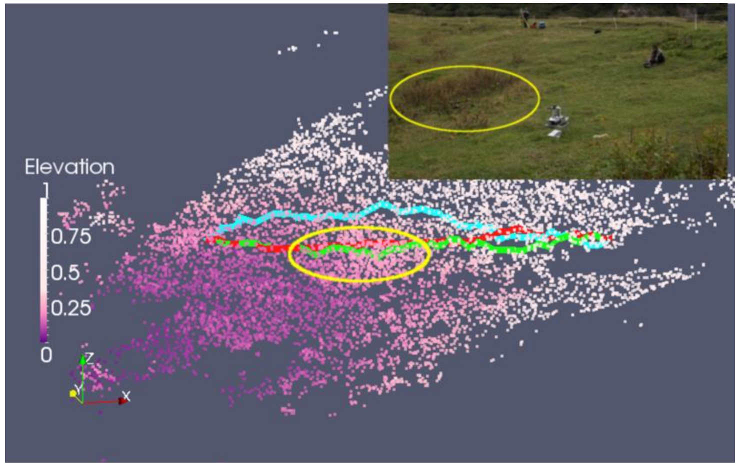
Efficient Path Planning



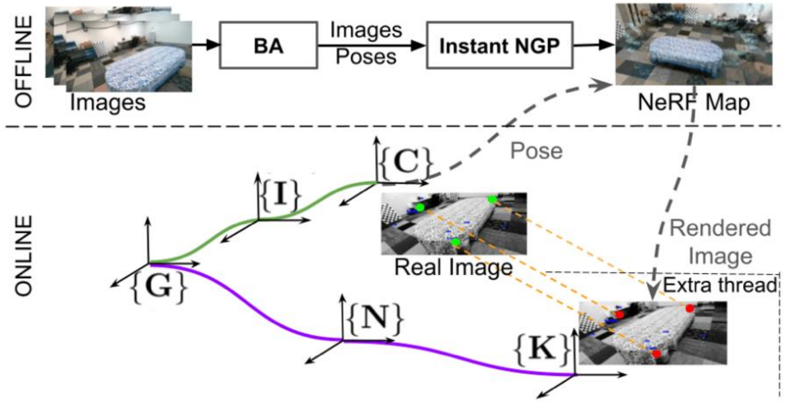


Visual Navigation Pipeline

Related Work



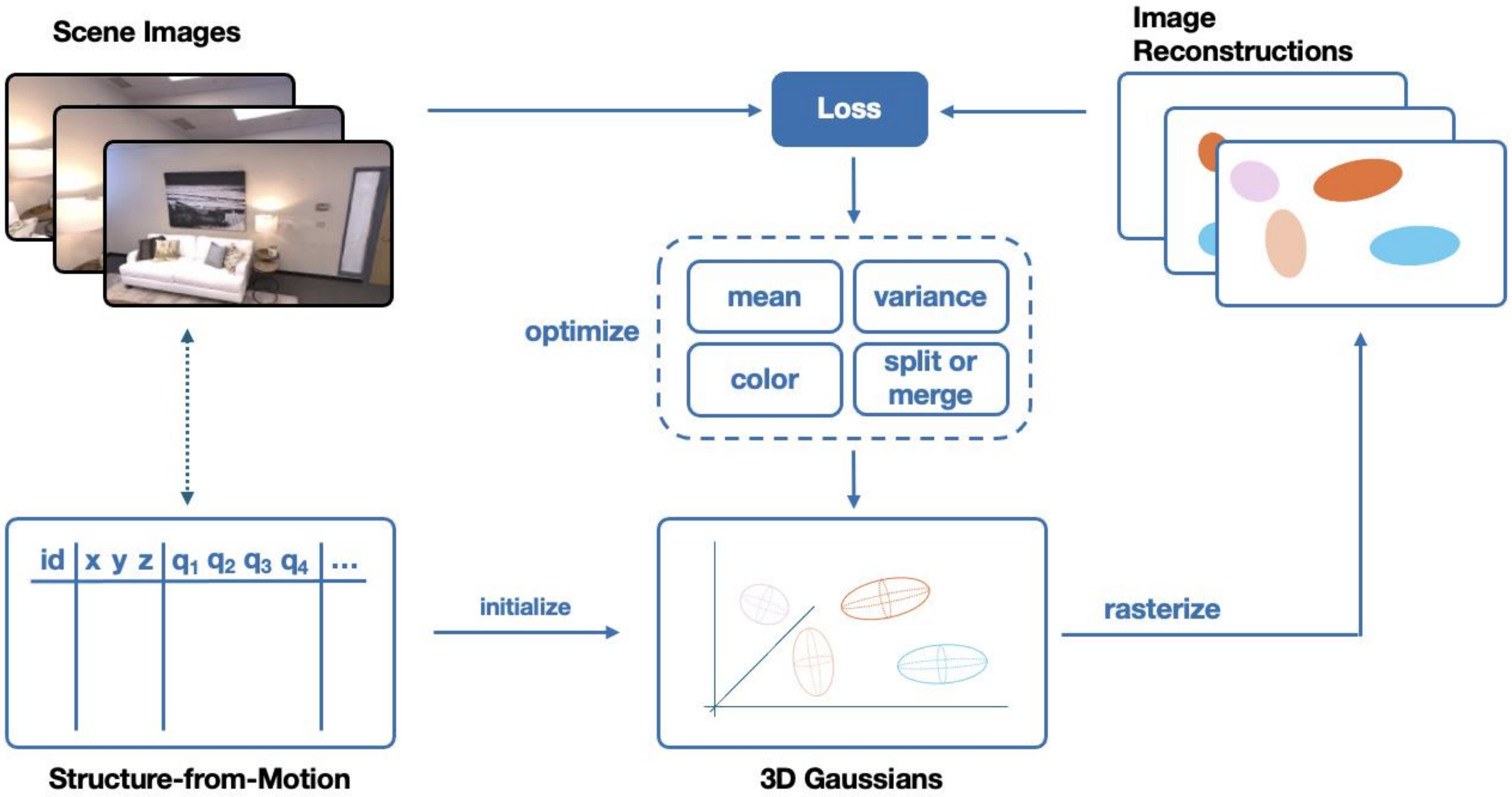
Navigation in Point Clouds[8]



NeRF-based Navigation[9]

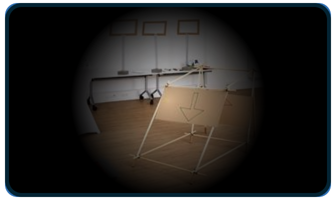
Voxel-based Navigation[10]

3D Gaussian Splatting[1]



Idea: Use 3D reconstruction for visual navigation

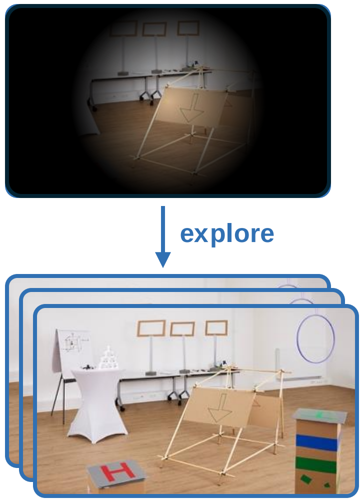
Method 1: GaussNav[2]



Steps

1. Scene Exploration

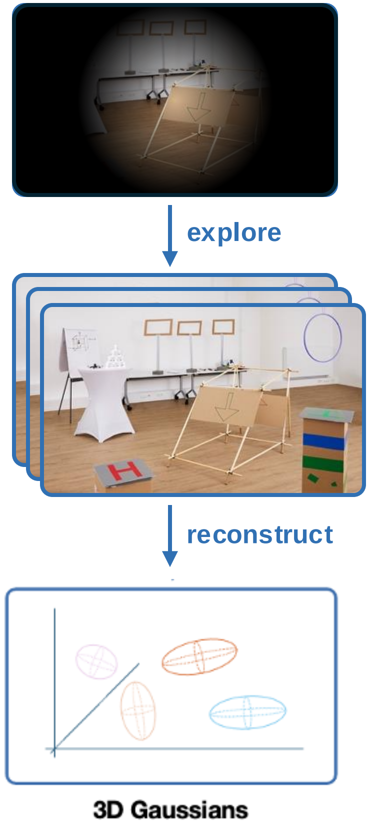
Method 1: GaussNav[2]



Steps

1. Scene Exploration

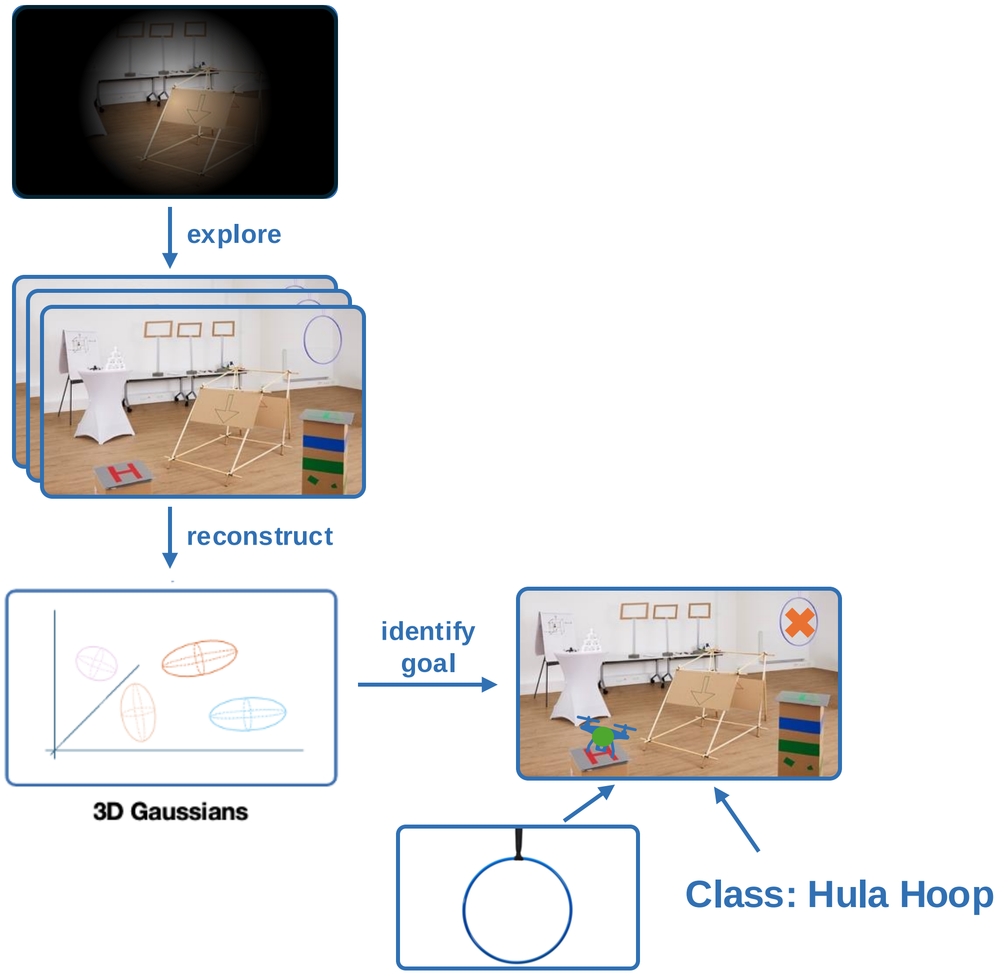
Method 1: GaussNav[2]



Steps

1. Scene Exploration
2. 3D Gaussian Reconstruction

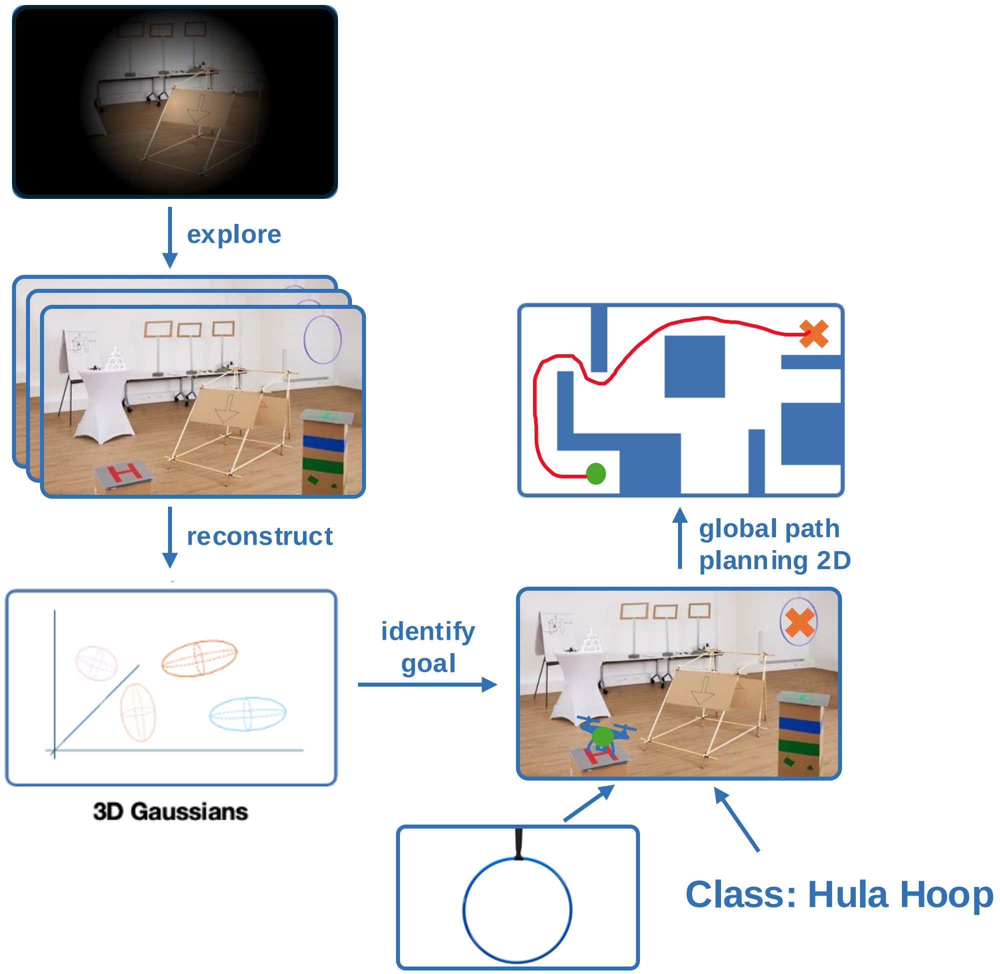
Method 1: GaussNav[2]



Steps

1. Scene Exploration
2. 3D Gaussian Reconstruction
3. Goal Identification through class prompt and reference Image

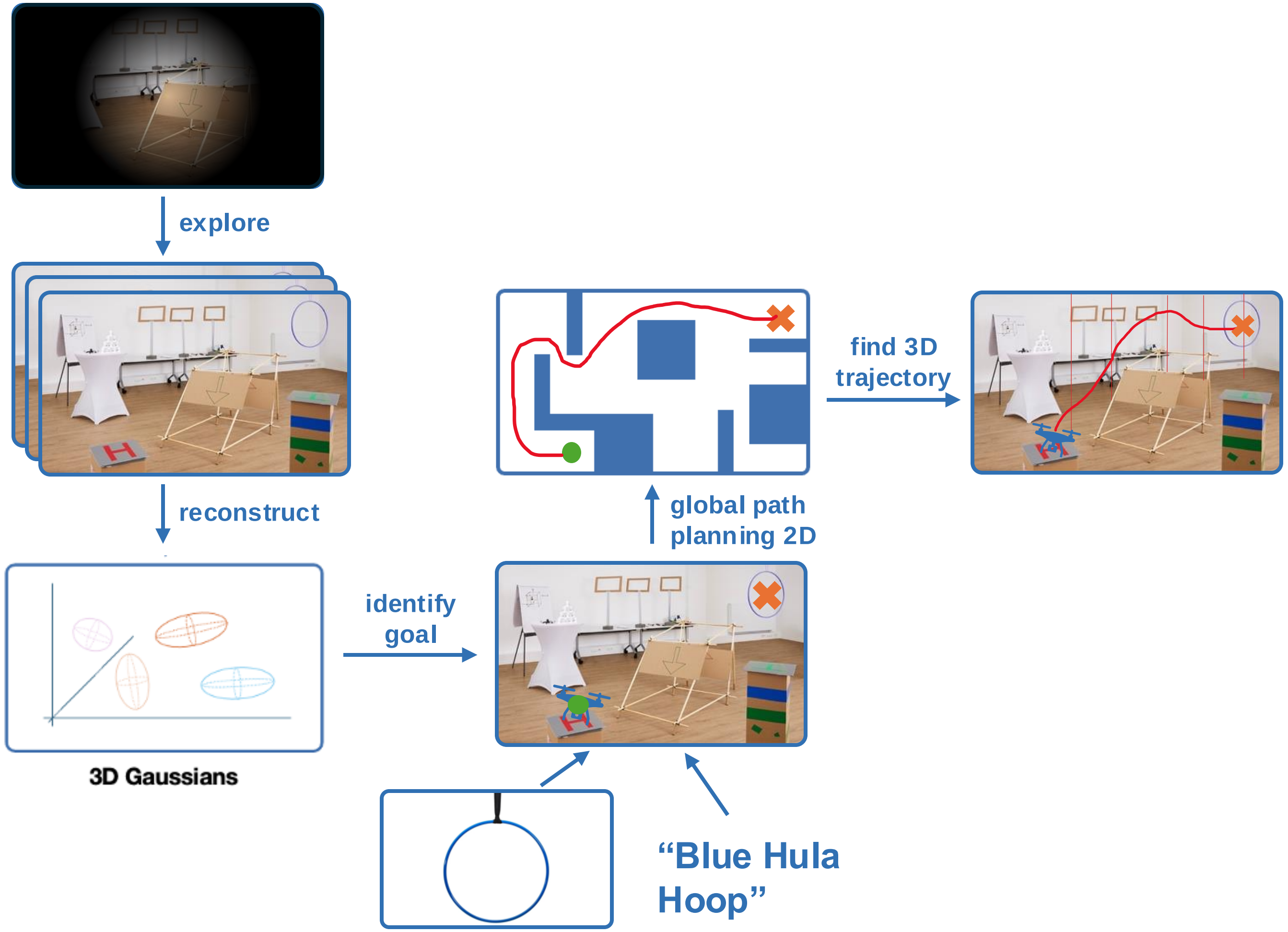
Method 1: GaussNav[2]



Steps

1. Scene Exploration
2. 3D Gaussian Reconstruction
3. Goal Identification through class prompt and reference Image
4. Conversion to 2D BEV map for global planning

Method 1: GaussNav[2]



Steps

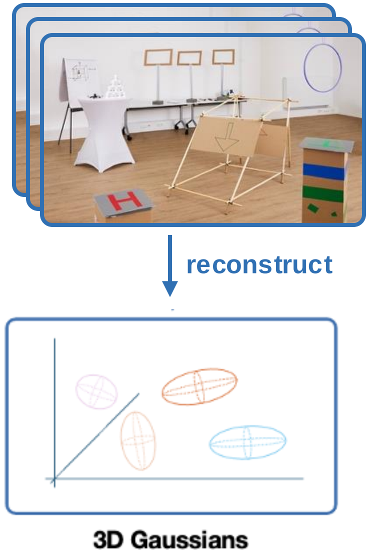
1. Scene Exploration
2. 3D Gaussian Reconstruction
3. Goal Identification through class prompt and reference Image
4. Conversion to 2D BEV map for global planning
5. Conversion of Gaussian Splatting map to Voxel Grid for 3D Navigation

Method 2: Splat-Nav[3]

Steps

1 Start with set of images

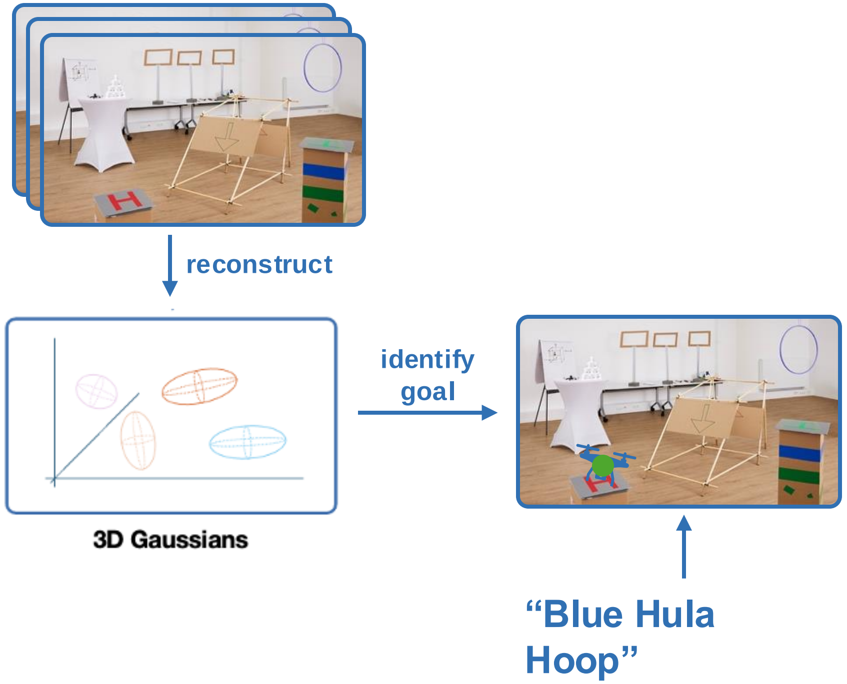
Method 2: Splat-Nav[3]



Steps

1. Start with set of images
2. 3D Gaussian Reconstruction

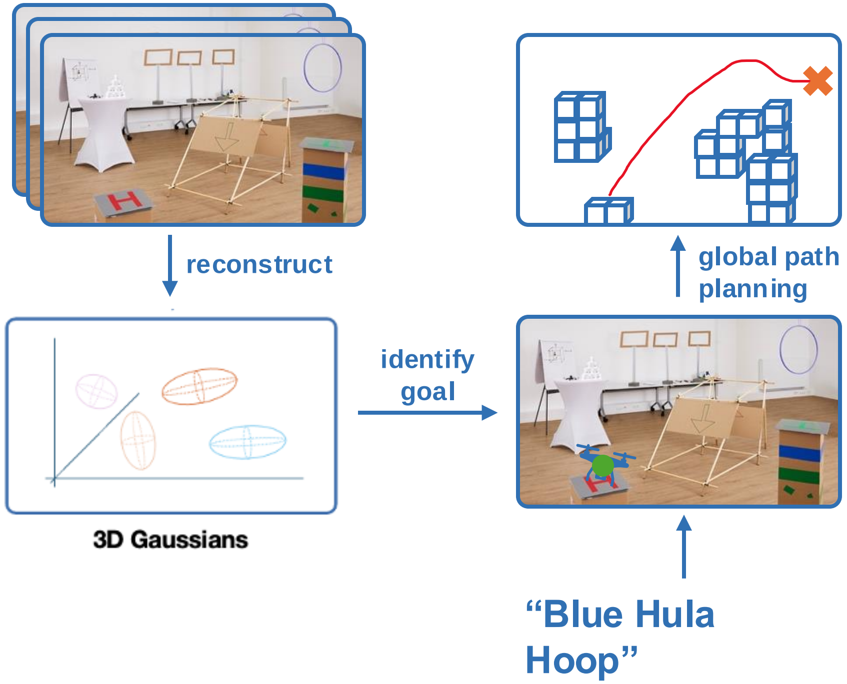
Method 2: Splat-Nav[3]



Steps

1. Start with set of images
2. 3D Gaussian Reconstruction
3. Goal Identification through open-vocabulary prompt

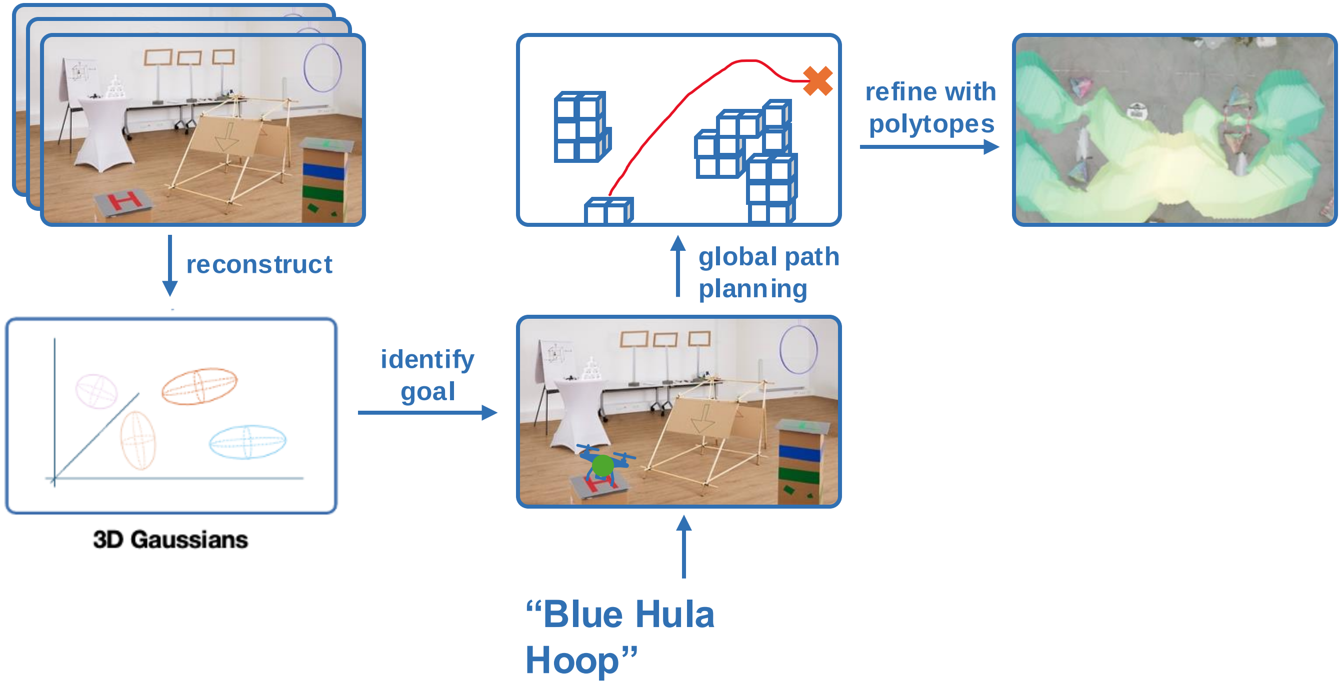
Method 2: Splat-Nav[3]



Steps

1. Start with set of images
2. 3D Gaussian Reconstruction
3. Goal Identification through open-vocabulary prompt
4. 3D Occupancy Grid for Global Path Planning

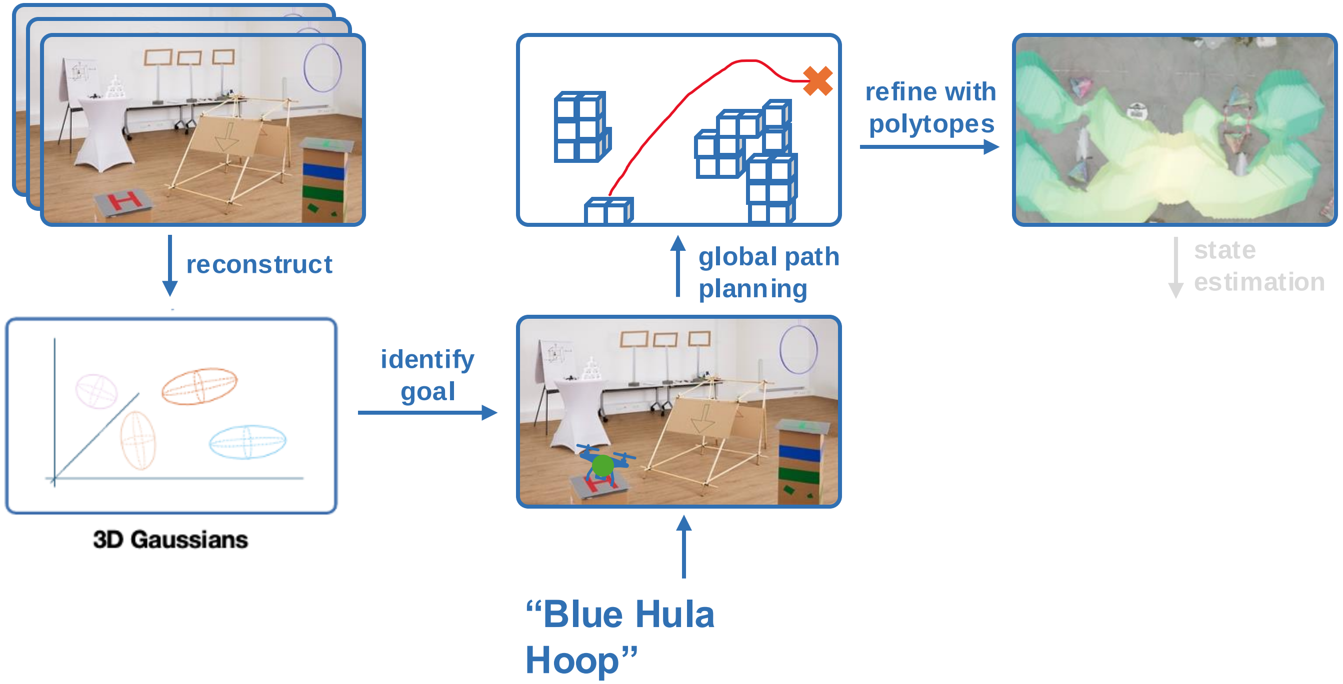
Method 2: Splat-Nav[3]



Steps

1. Start with set of images
2. 3D Gaussian Reconstruction
3. Goal Identification through open-vocabulary prompt
4. 3D Occupancy Grid for Global Path Planning
5. Ellipsoid intersection for computation of safe polytopes

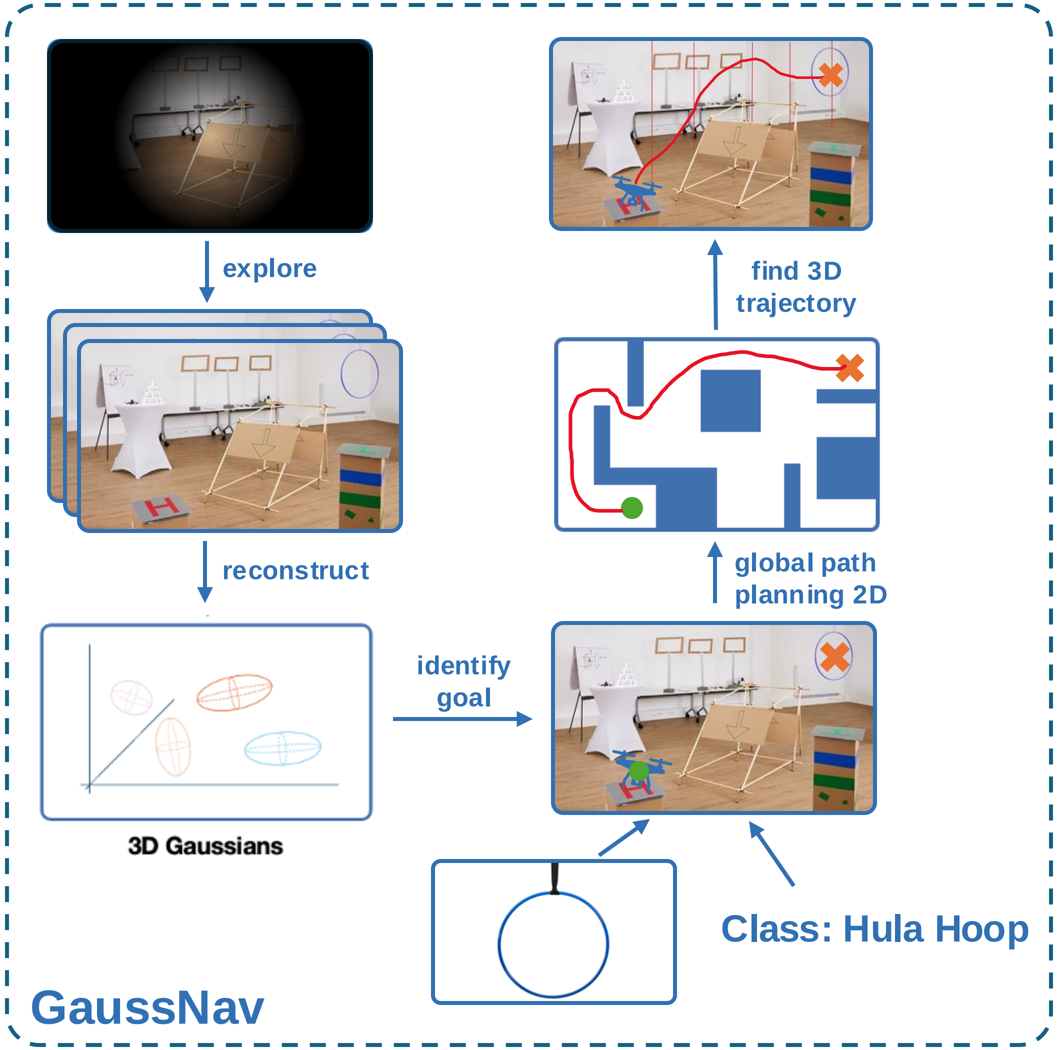
Method 2: Splat-Nav[3]

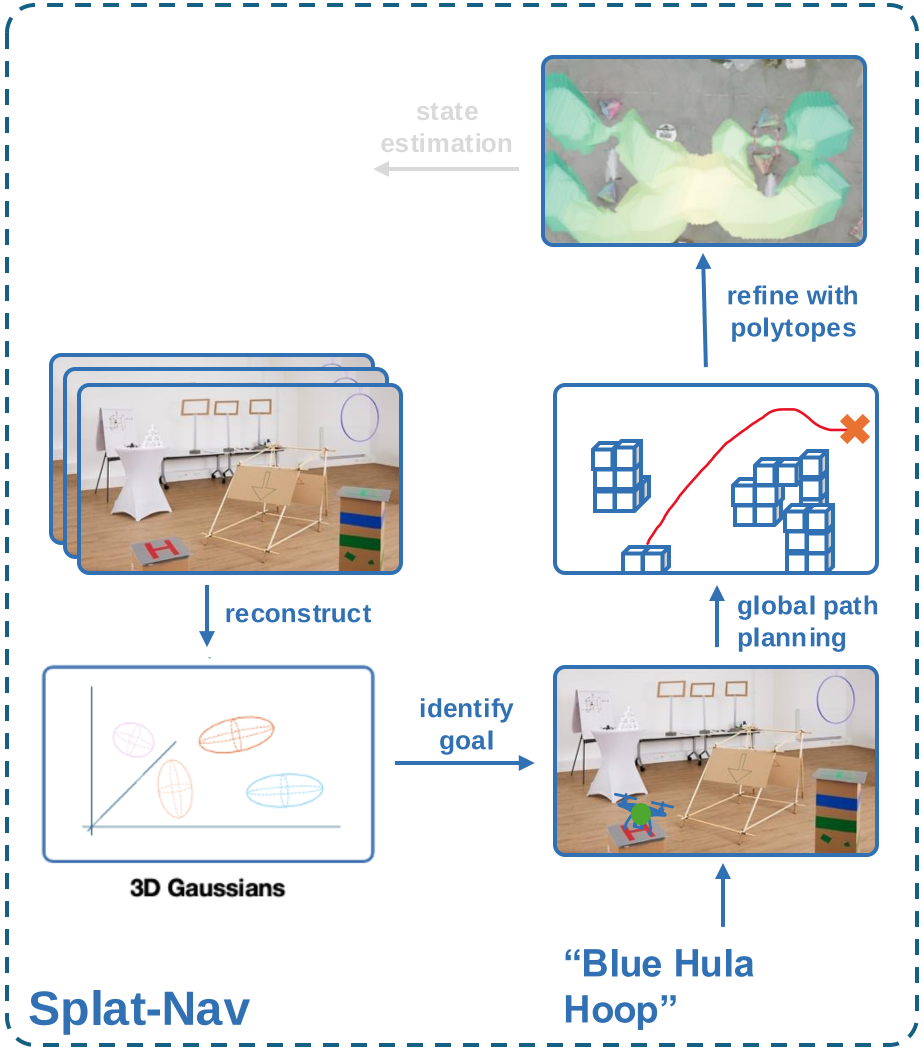


Steps

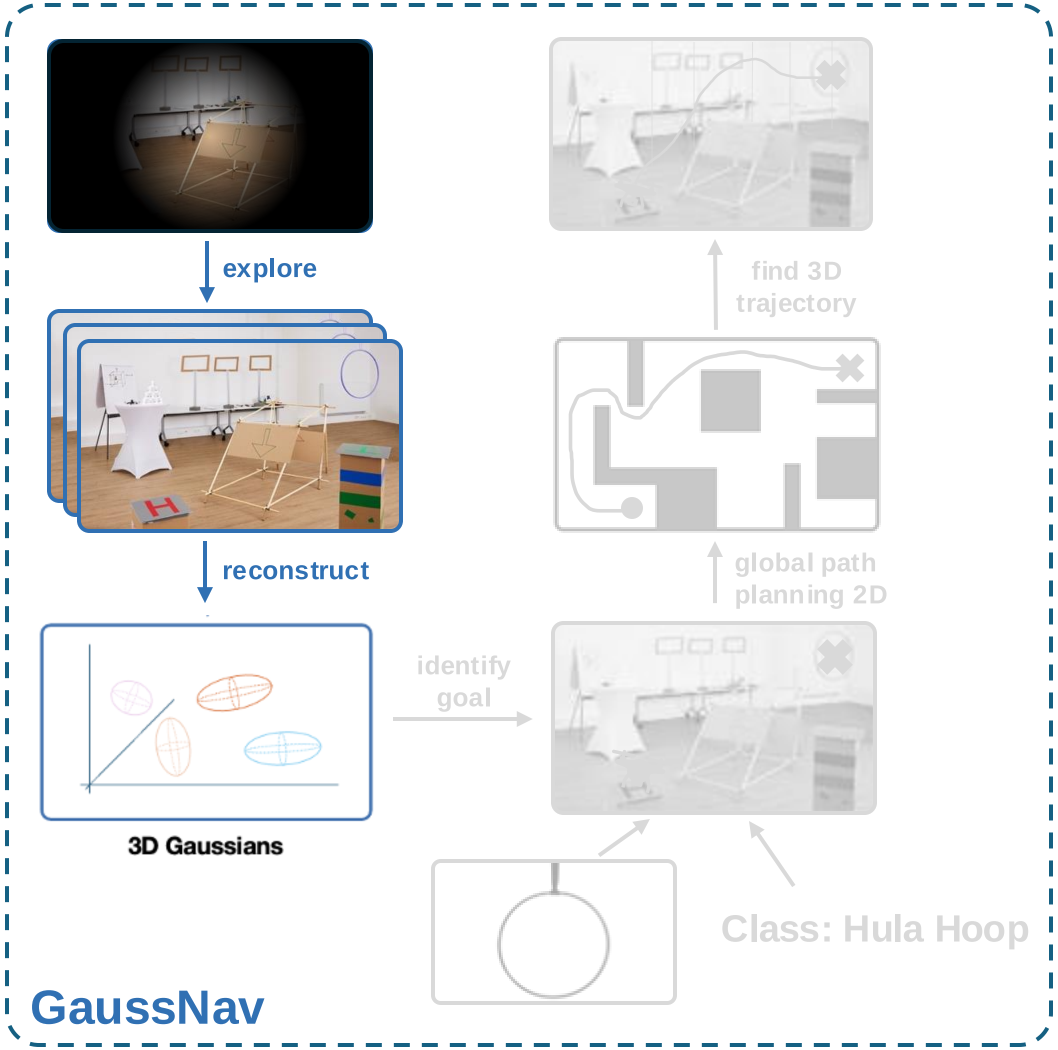
1. Start with set of images
2. 3D Gaussian Reconstruction
3. Goal Identification through open-vocabulary prompt
4. 3D Occupancy Grid for Global Path Planning
5. Ellipsoid intersection for computation of safe polytopes

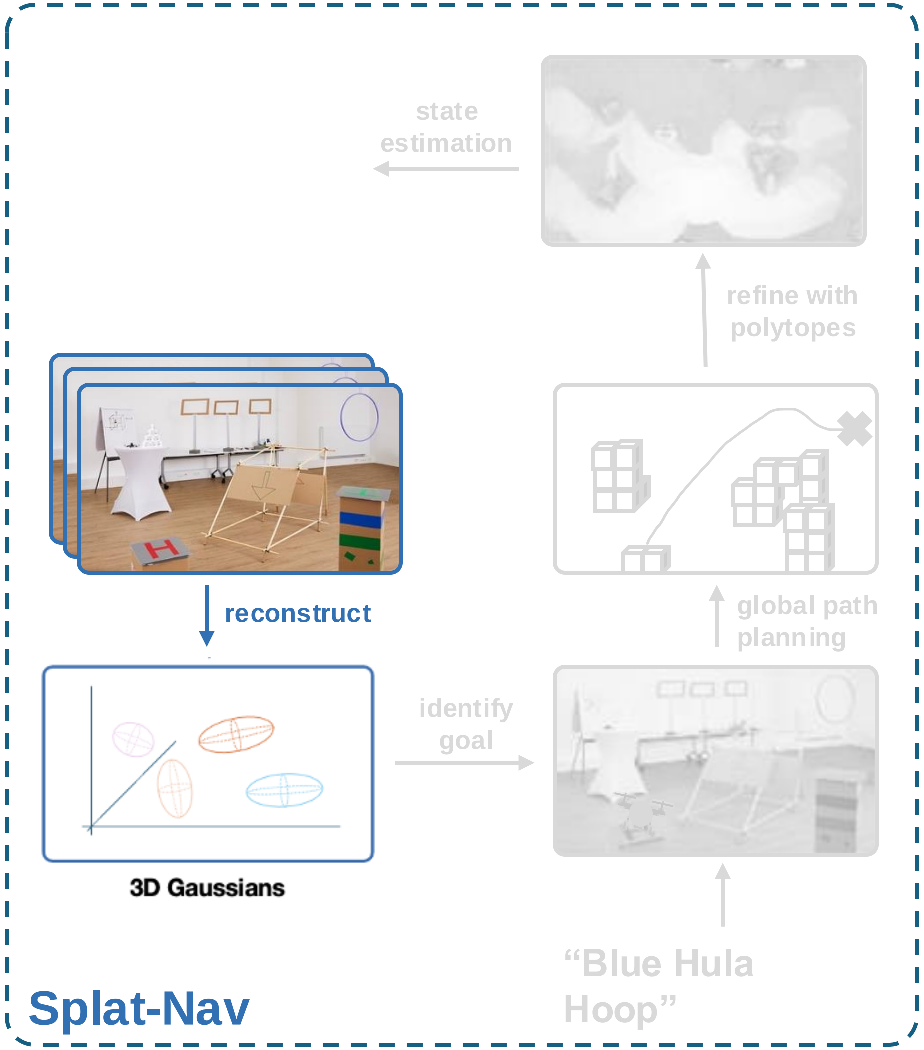
GaussNav vs. Splat-Nav



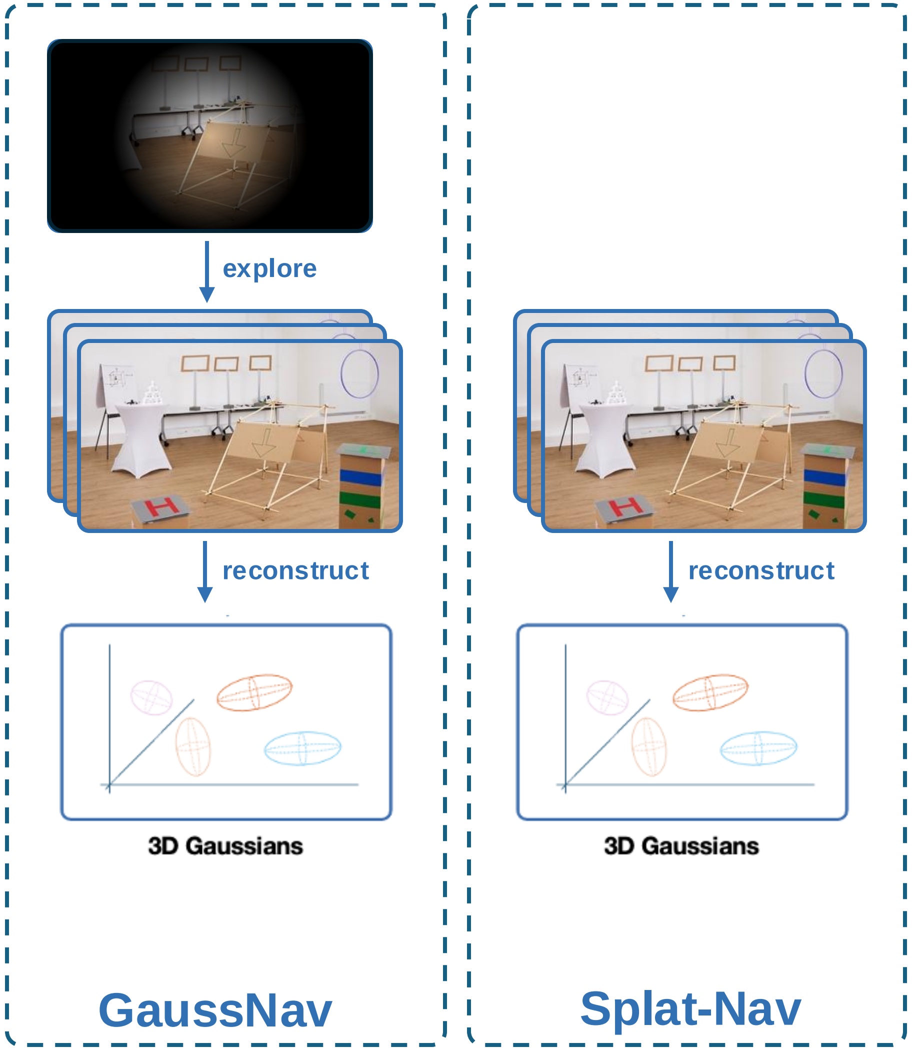


GaussNav vs. Splat-Nav





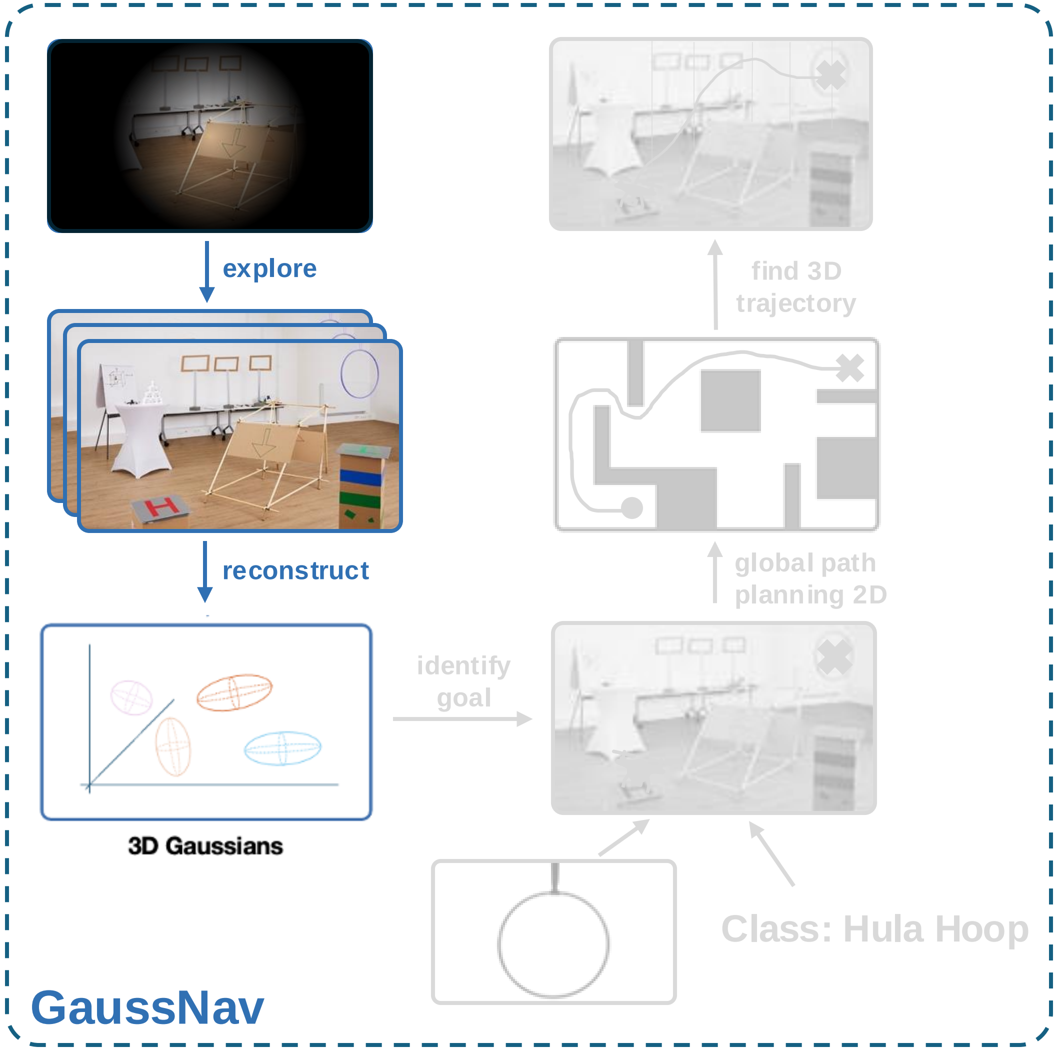
GaussNav vs. Splat-Nav

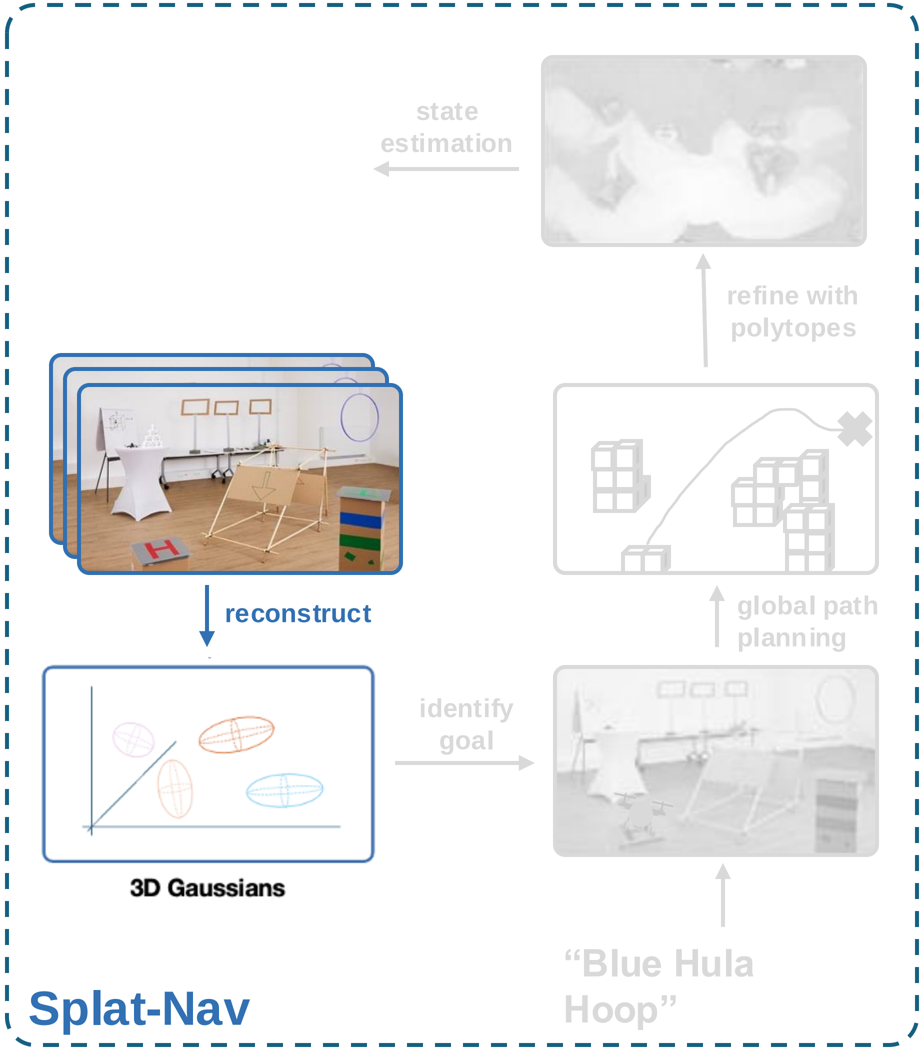


1. Splat-Nav does not use exploration for images
2. GaussNav uses obstacle and exploration maps
3. Both approaches use 3D Gaussian reconstruction
4. GaussNav assumes Gaussians to be isotropic

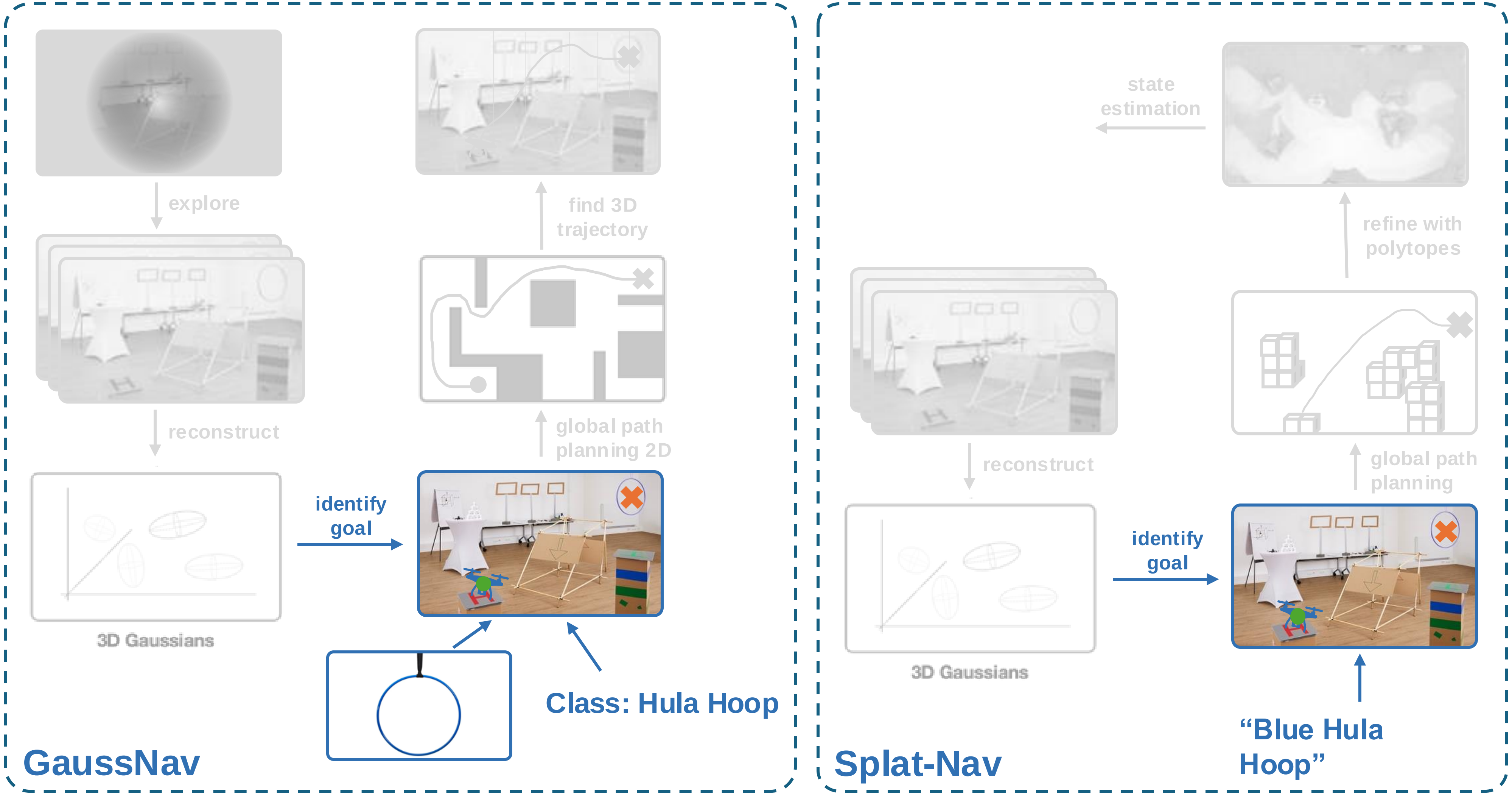
Both approaches rely on simplifying assumptions

GaussNav vs. Splat-Nav

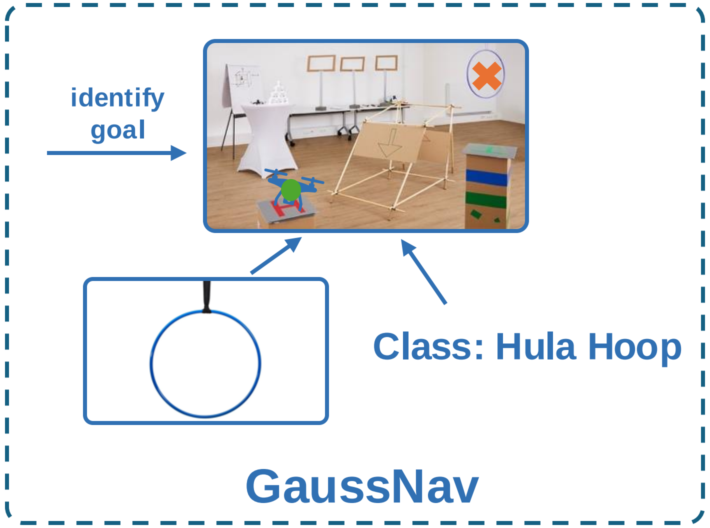


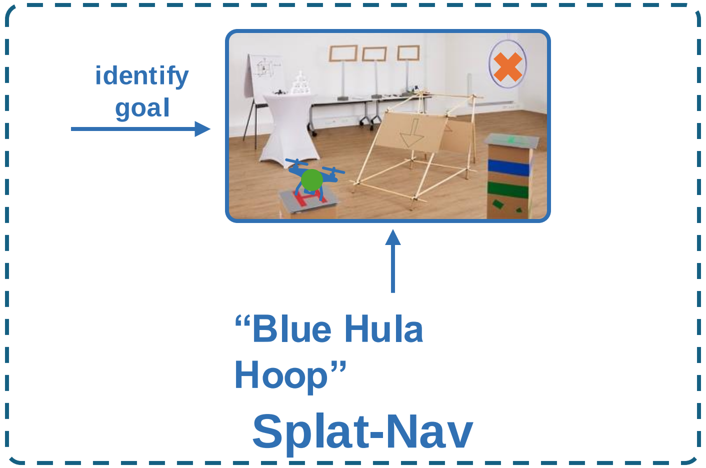


GaussNav vs. Splat-Nav



GaussNav vs. Splat-Nav

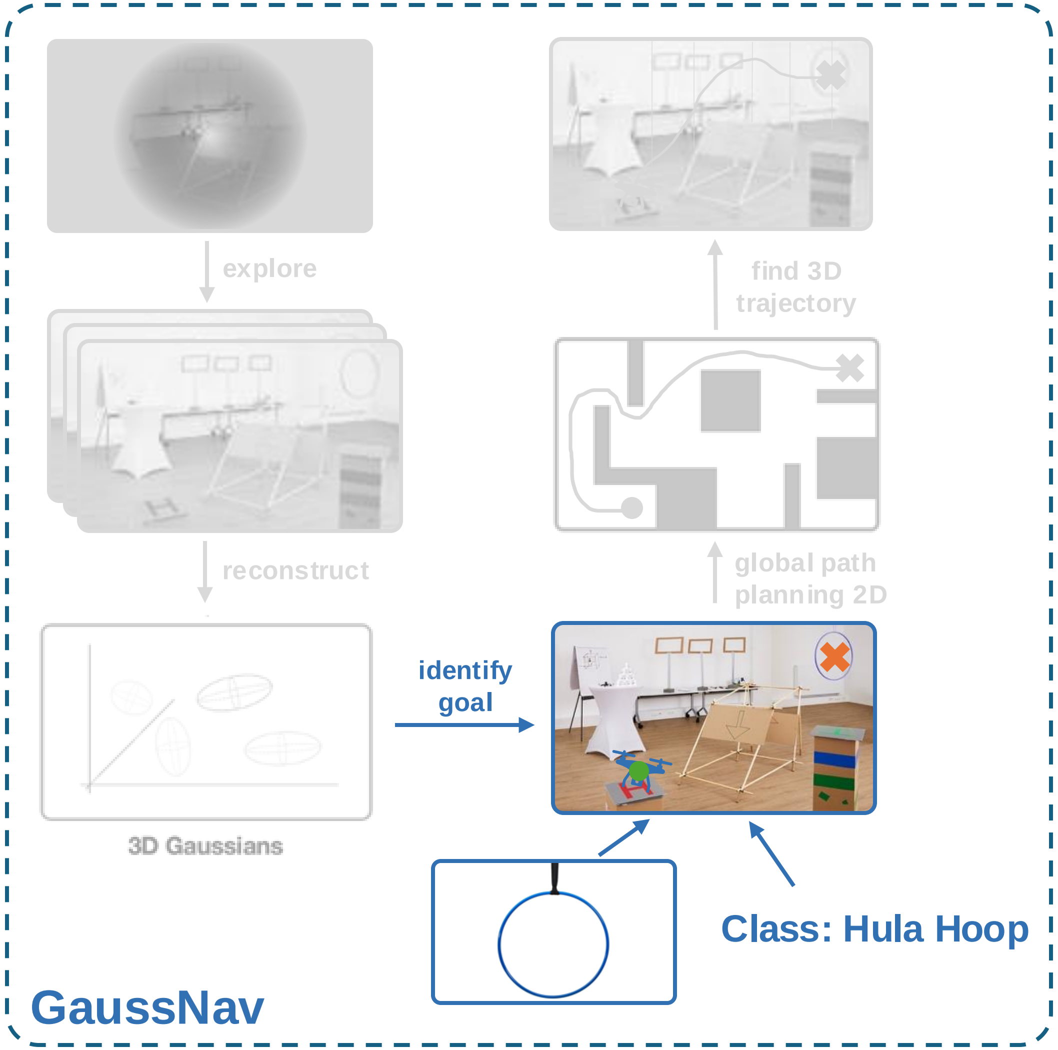


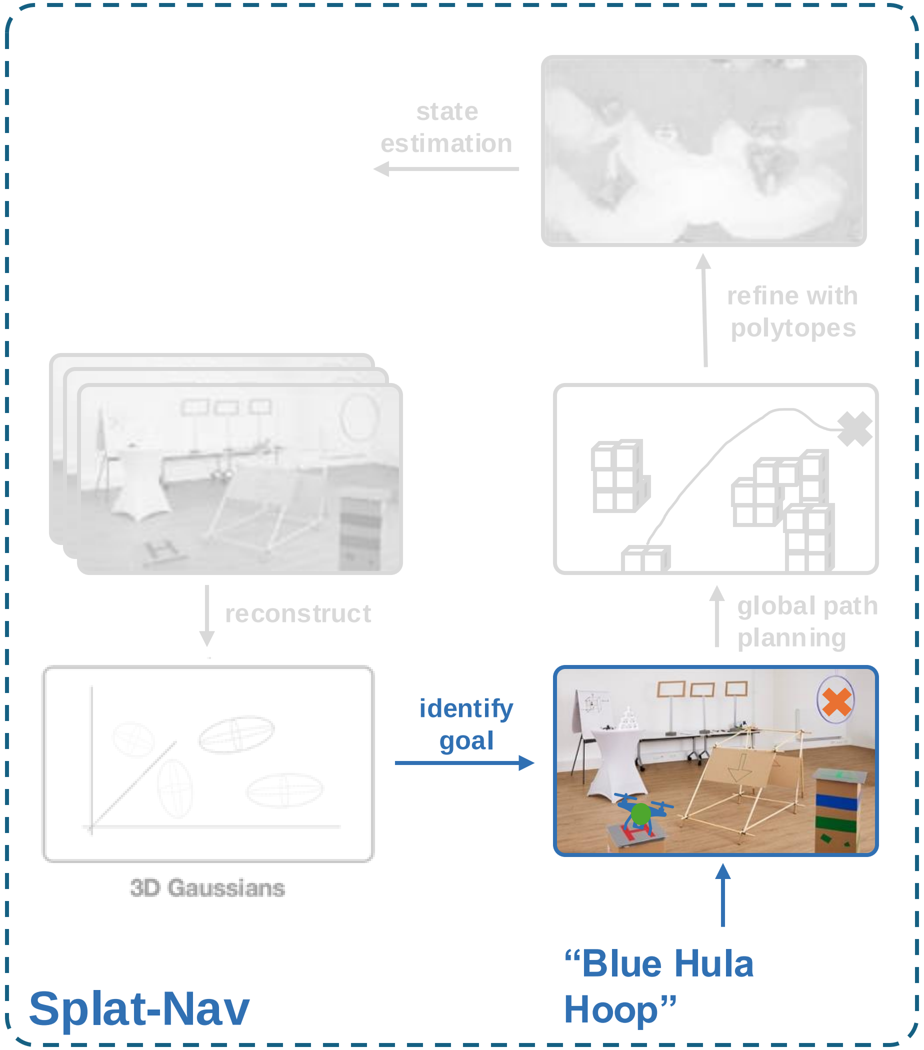


1. GaussNav identifies goal based on Image + Class
2. GaussNav chooses matching frame using keypoint overlap
3. Splat-Nav attaches CLIP[4] features to each Gaussian
4. These CLIP features can be used for direct language prompts

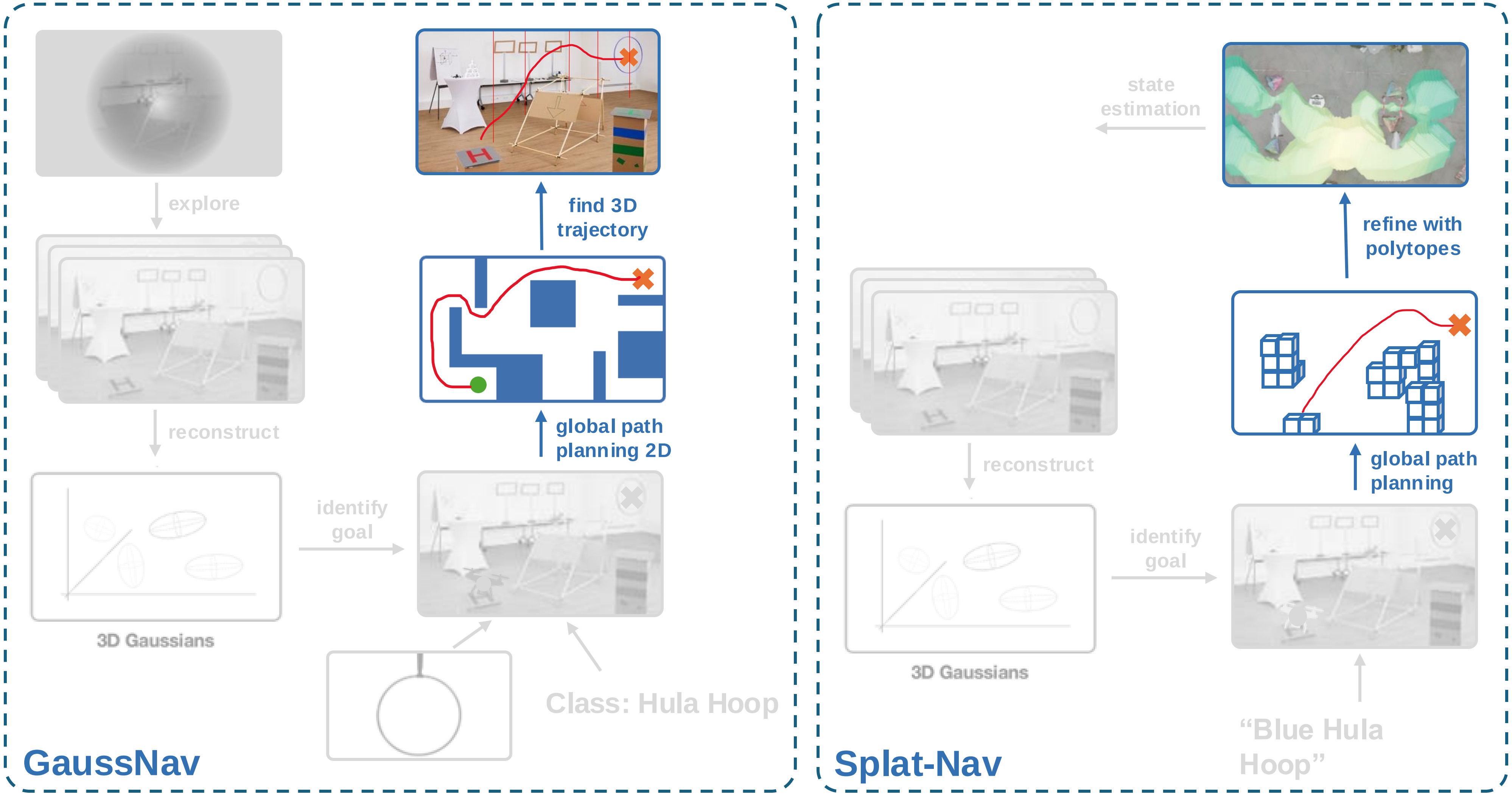
Different approaches for goal identification

GaussNav vs. Splat-Nav

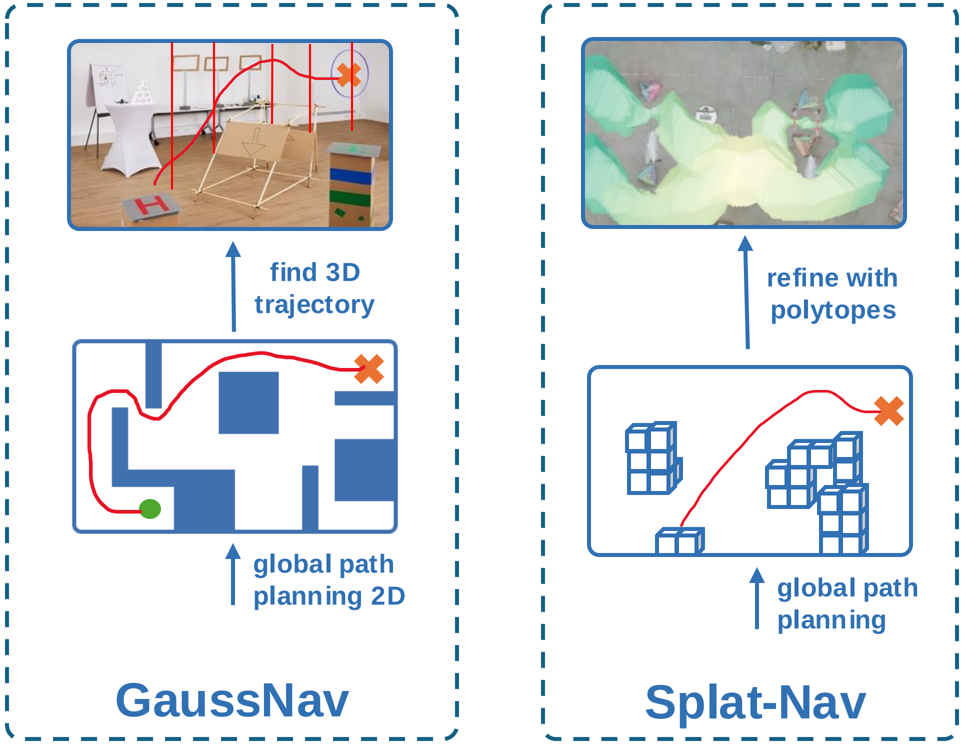




GaussNav vs. Splat-Nav



GaussNav vs. Splat-Nav



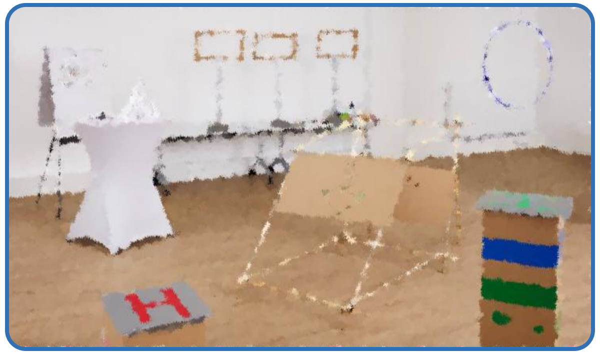
1. GaussNav plans global path using BEV 2D Grid
2. GaussNav refines locally by following way points
3. Splat-Nav starts with A* on a global voxel map
4. Splat-Nav reapplies A* in the current polytope to follow the global path

Division of path-planning into global and local

GaussNav vs. Splat-Nav

Sensed Image

Image estimate (Gaussian Splat)



$\psi(I_t)$ $h(x_t)$

$$\underset{x_t,\, x_{t-1}}{\arg\min} \; \|x_t - f(x_{t-1}, u_{t-1})\|^2 + \|\Psi(I_t) - h(x_t)\|^2 + \|x_{t-1} - \mu_{t-1}\|^2$$

Next State: x_t

State Prediction: $f(x_{t-1}, u_{t-1})$

x_{t-1}: Previous State

μ_{t-1}: Previous Estimate

1. Optimization of previous state and next state
2. Based on image estimate and state estimates
3. Additional constraints to consider polytope safety

Splat-Nav additionally covers state estimation

Experiments

GaussNav[2]

1. Tests visual navigation pipeline on Habitat-Matterport 3D (HM3D) dataset
2. Mainly tests success rate of reaching a specified goal
3. Compares performance to baselines including ViT- and CNN-based architectures

Splat-Nav[3]

1. Tests visual navigation pipeline on Gaussian Splatting scenes and using a real-world drone
2. Includes experiments for reaching goal but also pose estimation and safety of navigation pipeline
3. No direct comparisons to baseline planning algorithms

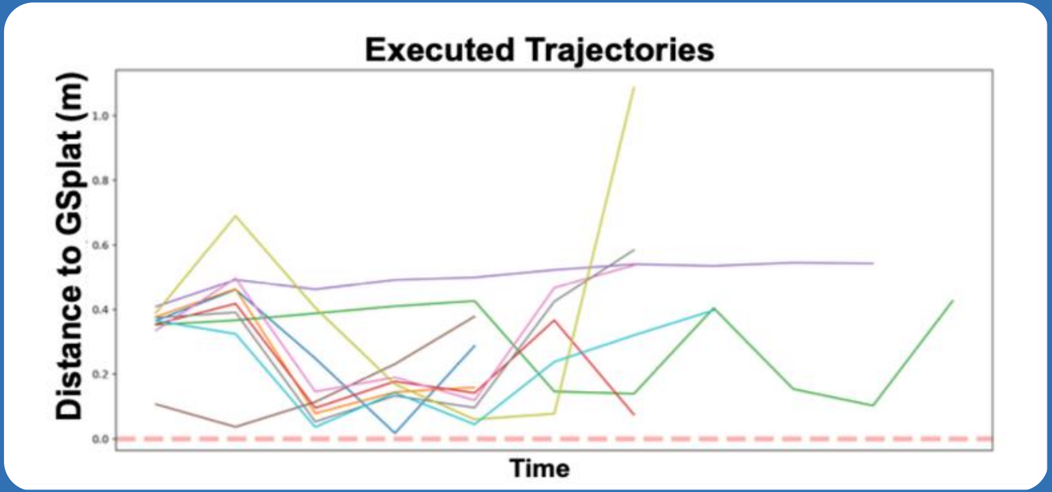
Results

GaussNav[2]

measures path efficiency

	Success	SPL
OvRL-v2 IIN[5]	0.248	0.118
Mod-IIN[6]	0.561	0.233
Ieve InternImage[7]	0.702	0.252
GaussNav	0.752	0.578

Splat-Nav[3]



	Success
GaussNav	0.946
Splat-Nav	1.000

GaussNav assumes known goal position but still not tested on the same dataset

Splat-Nav Polytopes

Limitations

GaussNav[2]

1. 3D Voxel grids requires $O(n^3)$ memory (n is number of voxels per dimension)
2. 2D projection for global path planning is an optimality bottleneck for future research
3. Keypoint matching with n keypoints of goal image and scene images introduces $O(n^2)$ computational complexity

Splat-Nav[3]

1. 3D occupancy grids require $O(n^3)$ memory (n is number of voxels per dimension)
2. Assuming that all images are given ignores exploration subproblem
3. $O(mn)$ computational complexity through pairwise comparisons with m robot and n total Gaussians
4. Can only find locally optimal paths because we only consider collisions with Gaussians in proximity

Future Work

Goal Identification Improvements

1. Major benefit of 3DGS: Photorealistic (more information preserved)
2. None of the approaches bases goal identification directly on Gaussians
3. Integrate 3DGS semantic segmentation (e.g. Langsplat[11])

Path Planning Improvements

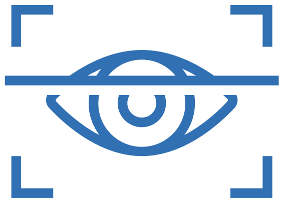
1. Both approaches use simplified geometry for global planning
2. Paths need to be refined locally to avoid collisions
3. It is desirable to avoid this representation conversion for path planning

Geometrical Improvements

1. Splat-Nav computes pairwise ellipsoid intersections
2. Many Gaussians collectively represent an object
3. They could be summarized to simpler geometric shapes

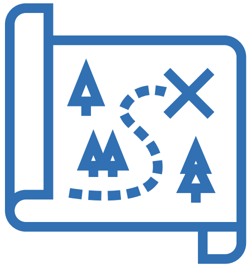
Summary

1) Perception



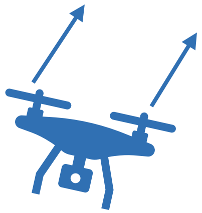
- Exploration
- 3D Reconstruction

2) Planning



- Goal Identification
- Obstacle Avoidance
- Optimal Path Planning

3) Control



- Pose Estimation
- Low-level pose control
- Low-level trajectory control

Key Takeaways

1. Visual Navigation combines several subproblems, papers usually tackle one specific subproblem
2. 3DGS reconstructions preserve more semantic meaning than traditional representations
3. GS requires MANY Gaussians => pairwise comparisons of ellipsoids are expensive
4. Navigating GS scenes in real-time will remain a challenge for a while

References

1. Kerbl, B., Kopanas, G., Leimkühler, T. and Drettakis, G., 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Trans. Graph.*, *42*(4), pp.139-1.
2. Lei, X., Wang, M., Zhou, W. and Li, H., 2024. GaussNav: Gaussian Splatting for Visual Navigation. *arXiv preprint arXiv:2403.11625*.
3. Chen, T., Shorinwa, O., Bruno, J., Yu, J., Zeng, W., Nagami, K., Dames, P. and Schwager, M., 2024. Splat-nav: Safe real-time robot navigation in gaussian splatting maps. *arXiv preprint arXiv:2403.02751*.
4. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J. and Krueger, G., 2021, July. Learning transferable visual models from natural language supervision. In International conference on machine learning (pp. 8748-8763). PMLR.
5. Yadav, K., Majumdar, A., Ramrakhya, R., Yokoyama, N., Baevski, A., Kira, Z., Maksymets, O. and Batra, D., 2023. Ovrl-v2: A simple state-of-art baseline for imagenav and objectnav. arXiv preprint arXiv:2303.07798.
6. Krantz, J., Gervet, T., Yadav, K., Wang, A., Paxton, C., Mottaghi, R., Batra, D., Malik, J., Lee, S. and Chaplot, D.S., 2023. Navigating to objects specified by images. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 10916-10925).
7. Lei, X., Wang, M., Zhou, W., Li, L. and Li, H., 2024. Instance-aware Exploration-Verification-Exploitation for Instance ImageGoal Navigation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 16329-16339).
8. Liu, M. and Siegwart, R., 2014, May. Navigation on point-cloud—A Riemannian metric approach. In 2014 IEEE International Conference on Robotics and Automation (ICRA) (pp. 4088-4093). IEEE.
9. Katragadda, S., Lee, W., Peng, Y., Geneva, P., Chen, C., Guo, C., Li, M. and Huang, G., 2024, May. Nerf-vins: A real-time neural radiance field map-based visual-inertial navigation system. In 2024 IEEE International Conference on Robotics and Automation (ICRA) (pp. 10230-10237). IEEE.
10. Gorte, B., Zlatanova, S. and Fadli, F., 2019. Navigation in indoor voxel models.
11. Qin, M., Li, W., Zhou, J., Wang, H. and Pfister, H., 2024. Langsplat: 3d language gaussian splatting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 20051-20060).