EE 367 Final Project Proposal: Diffusion Priors for Sparse and Noisy Robotic Mapping

Aviad Golan Peretz

Abstract

This project explores diffusion / score-based generative models as learned priors for solving inverse problems in robotic perception and mapping. Concretely, the goal is to reconstruct a dense 2D occupancy grid map (and optionally dense depth) from sparse, noisy range observations by combining a measurement model with a diffusion prior. The core hypothesis is that diffusion priors can enforce structural realism in maps (e.g., walls, corners, free-space continuity) more effectively than classical handcrafted regularizers (e.g., TV) and prior plug-and-play denoisers, improving map fidelity under aggressive sparsity and noise.

Background and Motivation Robotic mapping can be written as an inverse problem:

$$y = A(x) + \varepsilon, \tag{1}$$

where x is a latent map (occupancy grid), y are sensor observations (e.g., 2D LiDAR beams or sparse depth), $A(\cdot)$ is a forward projection / raycasting operator, and ε is noise. Classical occupancy grid mapping builds probabilistic updates from beams [1], and modern SLAM systems rely heavily on scan matching and graph optimization [2, 3]. However, in challenging regimes (limited field-of-view, sparse beams, occlusions, noisy sensors), the mapping problem is severely ill-posed and benefits from strong priors. These can be Diffusion models which provide a principled learned prior by approximating the score $\nabla_x \log p(x)$ [4, 5]. This enables MAP-like reconstruction or posterior sampling conditioned on measurements [6], bridging modern generative modeling with classical inverse problem formulation.

Objectives The goals are to train or adapt a diffusion / score model over 2D occupancy grid maps (and optionally depth maps) derived from simulated robot environments. Preference to train a small network, but will depend on time, maybe only fine tune an existing model using my data. Then, I will use it to solve a measurement-conditioned reconstruction problem using the diffusion prior, producing dense maps

from sparse/noisy observations. Lastly, I will compare against existing baselines such as least squares and TV-regularized least-square. I will compare using standard metrics such as SSIM.

Technical Approach

For Data and Map Representation, required to generate ground-truth occupancy grids from simulated embodied environments using Habitat [7]. A robot trajectory is sampled; at each pose, synthetic range measurements are produced (e.g., a subset of LiDAR beams, limited FOV, or subsampled depth pixels). Ground-truth maps are rasterized at a fixed resolution (e.g., 256×256 or 128×128). The goal is for these measurements to be sparse (low number of beams).

The **Forward Model** can be represented by the forward operator $A_t(x)$ which returns predicted measurements (e.g., expected range-to-hit along rays). Aggregating over T poses yields:

$$y = A(x) + \varepsilon, \quad A(x) := \{A_t(x)\}_{t=1}^{T}. \tag{2}$$

For feasibility within the course timeframe, we will implement a differentiable approximation of raycasting (e.g., discrete ray marching with soft occupancy) or use a linearized proxy operator in the occupancy image domain (e.g., sparse projection / masking) while keeping the evaluation grounded in a robotics measurement process.

Diffusion Prior and Conditioning We will use a diffusion model trained on maps to estimate either a denoising model $\epsilon_\theta(x_t, t)$ (DDPM-style) [5], or a score model $s_\theta(x, \sigma) \approx \nabla_x \log p_\sigma(x)$ (SDE/score-based) [4].

To incorporate measurements, we will implement **diffusion posterior sampling (DPS)** [6] or an equivalent score-based update:

$$x_{k+1} = x_k - \eta \nabla_x \mathcal{L}_{\text{data}}(x_k) + \eta\lambda\, s_\theta(x_k, \sigma_k) + \sqrt{2\eta}\, \xi_k, \tag{3}$$

where $\mathcal{L}_{\text{data}}(x) = \|A(x) - y\|^2$ (or a likelihood-consistent term), σ_k follows the diffusion noise schedule, and $\xi_k \sim \mathcal{N}(0, I)$.

Resources and Compute

Training a 2D diffusion model on 128^2–256^2 grids is feasible on a single GPU. If compute is constrained, smaller grids are to be used, fewer diffusion steps, or fine-tune a lightweight UNet.

References

[1] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," *Computer*, vol. 22, no. 6, pp. 46–57, 1989.

[2] W. Hess, D. Kohler, H. Rapp, and D. Andor, "Real-time loop closure in 2d lidar slam," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016.

[3] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT Press, 2005.

[4] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," in *International Conference on Learning Representations (ICLR)*, 2021.

[5] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[6] H. Chung, J. Kim, M. T. McCann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," in *International Conference on Learning Representations (ICLR)*, 2023.

[7] M. Savva, A. Kadian, O. Maksymets, *et al.*, "Habitat: A platform for embodied ai research," in *IEEE International Conference on Computer Vision (ICCV)*, 2019.