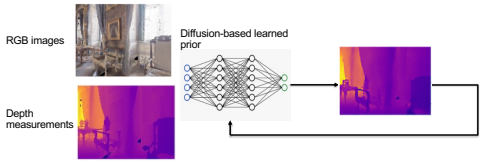
Sparse Depth Reconstruction Using Diffusion Priors Guided by RGB Observations

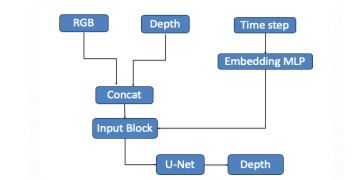
Aviad Golan Peretz
Stanford University

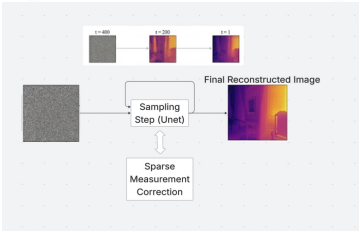
Motivation

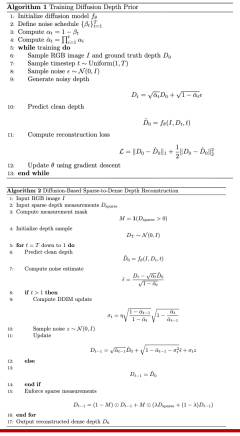
- Robots and autonomous systems require dense 3D maps to understand and interact with their environments
- LiDAR sensors provide accurate depth, but sparse and expensive. RGB measurements alone are ambiguous and ill-posed
- Diffusion models can act as learned geometric priors



Diffusion + U-Net Learned Prior Architecture







Related Work

- Monocular depth estimation networks, struggle with depth ambiguity [1]
- Sparse-to-dense depth completion methods fuse LiDAR with RGB images using CNNs, may over-smooth geometry and fail to model uncertainty [2]
- Diffusion models have recently demonstrated strong performance used as generative priors for inverse problems, such as image reconstruction, image inpainting [3,4]

References

[1] D. Eigen, C. Puhrsch, and R. Fergus. Depth Map Prediction from a Single Image using a Multi-Scale Deep Network. Advances in Neural Information Processing Systems (NuerIPS), 2014.
[2] F. Ma and S. Karaman. Sparse-to-Dense: Depth Prediction from Sparse Depth Samples and a Single Image. IEEE International Conference on Robotics and Automation (ICRA), 2018.
[3] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole. Score-Based Generative Modeling through Stochastic Differential Equations. International Conference on Learning Representations (ICLR), 2021.
[4] B. Kawar, M. Elad, S. Ermon, and J. Song. Denoising Diffusion Restoration Models. Advances in Neural Information Processing System, 2022.

Experimental Results

Model	MAE	RMSE	PSNR	SSIM
U-Net Baseline	0.0278	0.0537	28.16	0.823
Diffusion (ours)	**0.0141**	**0.0352**	**33.45**	**0.949**

- The diffusion-based architecture outperforms the base U-Net with Sharper edge preservation as can be seen visually, and numerically by the SSIM metric
- Overall, quantitatively higher signal quality is shown by the diffusion-based approach (PSNR), as well as lower reconstruction error (MAE / RMSE)