BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2026
http://excel.fit.vutbr.cz

Beyond Dense Grids: Sparse SDF Diffusion for Efficient 3D Shape Completion

Bc. Tomáš Rajsigl*

Abstract
Real-world 3D scans often suffer from missing geometry, requiring automated shape completion. However, current volumetric models rely on dense grids, creating a severe memory bottleneck that limits resolution. To overcome this, we propose a dual-branch diffusion framework that operates on a sparse Signed Distance Field representation. The model is trained entirely in a self-supervised manner by dynamically simulating occlusions on complete 3D shapes. We validate our solution against a dense state-of-the-art baseline. Our least aggressive variant matches baseline performance while reducing active voxels by 36%, and pushing this reduction to 54% yields only a modest 5.05 points drop in *m*-IoU. Ultimately, this efficiency enables us to double spatial resolution to 64^3, capturing finer details under the same VRAM budget.

*xrajsi01@stud.fit.vutbr.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

Modern 3D scanning is highly accessible, driving rapid adoption across robotics, medicine, and other fields. However, these tools are inherently imperfect, often suffering from missing geometry that necessitates repair before downstream use. Because manual repair of broken meshes is exceedingly slow and expensive at scale, there is a critical need for automated reconstruction pipelines that can reliably infer missing structure.

The core task of shape completion is to automatically generate missing 3D geometry from a partial scan. The central challenge of developing a practical solution to this problem is twofold:

1. **Geometric Plausibility:** The method must fill missing regions while remaining structurally consistent and plausible in the global context.
2. **Computational Efficiency:** Since 3D data scales cubically, the solution must keep memory usage low enough to support high-resolution details.

Current state-of-the-art methods often compromise between these two goals. Volumetric models such as DiffComplete [1] produce high-quality results but use dense voxel grids, causing a severe resolution bottleneck. Approaches like ShapeFormer [2] and SC-Diff [3] mitigate this by compressing 3D data into a latent space, but their reconstruction quality depends strongly on a pre-trained autoencoder's capacity. Alternatively, Yariv *et al.* [4] recently introduced Mosaic-SDF, a localized representation for unconditional generation, which, although targeting a different task, inspires the sparse representation in our completion pipeline.

The proposed solution is a dual-branch diffusion framework operating exclusively on a sparse approximation of a *Signed Distance Field* representation. To overcome the difficulty of acquiring real-world paired scans, we train the model in a self-supervised manner. By simulating occlusions on complete 3D models, the framework is able to generate infinite synthetic training pairs.

The contributions of this work are as follows. We successfully adapted the modern EDM diffusion formulation by Karras *et al.* [5] to operate natively on sparse 3D data using sparse convolutions [6], significantly improving sampling efficiency. This optimization further enables our framework to **double** the spatial resolution to 64^3 compared to the baseline derived from DiffComplete. This allows for the reconstruction of much higher detail on the **same hardware**.

2. Surface-Centric SDF Representation

To process 3D shapes with a neural network, we first need to mathematically represent them. We utilize a *Signed Distance Function* (SDF), which simply measures the shortest distance from any given point in space to the object's surface, as defined in Equation 1.

However, to feed this continuous implicit representation into a deep learning model, we must discretize it into a grid. As illustrated by Figure 1, there are two approaches to this discretization. The standard volumetric method seen in Figure 1a evaluates the SDF uniformly across a dense bounding box, wasting massive amounts of compute on empty air.

Our solution is to discretize the space selectively. By implementing a sparse grid Figure 1b, we compute values strictly within a narrow surface boundary. This eliminates redundant processing, reducing the memory footprint drastically and enabling the network to scale to higher resolutions.

3. Simulate Holes to Repair Them

Training deep learning models requires large datasets. For shape completion, this means needing a complete 3D model paired with its broken counterpart. Because such datasets are incredibly difficult to capture in the real world, we sidestepped this limitation using a self-supervised approach.

We begin with a curated subset of roughly 6,000 complete furniture models from the Objaverse [7] dataset. As illustrated in Figure 2, we dynamically subtract chunks from these shapes using geometric primitives such as spheres or boxes, additionally perturbed by Perlin noise. This specific noise addition is critical because it more closely mimics the irregular patterns of real-world sensor occlusions.

By algorithmically damaging complete shapes on the fly, we generate the exact input-output pairs the network needs. This also yields a far greater variety of inputs than traditional completion datasets [8] and state-of-the-art approaches [1, 2, 9] that precompute a fixed set of damaged versions per object.

4. Dual-Branch Diffusion Framework

As can be seen in Figure 3, the proposed neural architecture builds upon the foundations of DiffComplete by incorporating a ControlNet-style [10] Condition Branch for separate partial input processing.

However, utilizing sparse convolutions introduces a unique challenge – the network cannot create geometry outside of its predefined active voxel coordinates. To address this, we use an interactive user-guided mechanism. The user draws a rough bounding area indicating the general location of the missing region. This bounding volume, combined with the partial input scan, defines the network's so called *operational space*.

After the initial preprocessing convolutions (denoted as $\psi_{\mathbf{x}}$ and $\psi_{\mathbf{c}}$ in Figure 3), the Condition Branch encodes known geometry from the partial scan, while the Main Branch synthesizes missing structure within the defined area. After the diffusion process iteratively refines the completed sparse SDF, we use Marching Cubes to extract the repaired mesh.

5. Qualitative and Quantitative Results

Although qualitative 64^3 results can be seen throughout the poster, our initial experiments first validated the sparse approach at the baseline resolution of 32^3.

In Table 1, *m*-IoU measures volumetric overlap between a prediction and ground truth in the missing region. Chamfer Distance (CD) measures similarity between two sampled point clouds, and *F*-Score is the harmonic mean of surface precision (accuracy of generated geometry) and recall (recovered true geometry).

To assess the memory–quality trade-off, we varied the cutoff distance d_c, which sets the width of the sparse surface boundary. The active voxel percentage shows the fraction of the grid stored in memory relative to a fully dense volume. As shown in Table 1, our **Accurate** variant matches the DiffComplete baseline while discarding 36% of the grid computations. Pushing this efficiency to the limit, our **Efficient** variant increases memory savings to 54% while remaining competitive.

6. Conclusions

This work proposes a diffusion-based approach for 3D shape completion, directly addressing the memory bottleneck of dense grids by operating on sparse SDFs.

Beyond integrating modern literature advances and a more diverse data approach, this architectural shift lets us double the resolution of the 2023 state-of-the-art DiffComplete baseline within the same VRAM budget.

6.1 Future Work

Although the proposed solution shows respectable results, there is still room for improvement. Sampling efficiency could be increased by adopting the novel Flow Matching [11] paradigm. It would also be valuable to evaluate the method on different data, such as cranial implants or other medical data, a direction currently under discussion with the supervisor.

Acknowledgements

I would like to thank my supervisor, doc. Ing. Michal Španěl, Ph.D., for providing help and guidance during my master's thesis on which this submission is based.

References

[1] Ruihang Chu, Enze Xie, Shentong Mo, Zhenguo Li, Matthias Nießner, Chi-Wing Fu, and Jiaya Jia. Diffcomplete: Diffusion-based generative 3d shape completion. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine, editors, *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023.

[2] Xingguang Yan, Liqiang Lin, Niloy J. Mitra, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. Shapeformer: Transformer-based shape completion via sparse representation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pages 6229–6239. IEEE, 2022.

[3] Juan D. Galvis, Xingxing Zuo, Simon Schaefer, and Stefan Leutenegger. Sc-diff: 3d shape completion with latent diffusion models. *CoRR*, abs/2403.12470, 2024.

[4] Lior Yariv, Omri Puny, Oran Gafni, and Yaron Lipman. Mosaic-sdf for 3d generative models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16-22, 2024*, pages 4630–4639. IEEE, 2024.

[5] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022.

[6] Spconv Contributors. Spconv: Spatially sparse convolution library. https://github.com/traveller59/spconv, 2022.

[7] Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3d objects. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*, pages 13142–13153. IEEE, 2023.

[8] Wentao Yuan, Tejas Khot, David Held, Christoph Mertz, and Martial Hebert. PCN: point completion network. In *2018 International Conference on 3D Vision, 3DV 2018, Verona, Italy, September 5-8, 2018*, pages 728–737. IEEE Computer Society, 2018.

[9] Paritosh Mittal, Yen-Chi Cheng, Maneesh Singh, and Shubham Tulsiani. Autosdf: Shape priors for 3d completion, reconstruction and generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pages 306–315. IEEE, 2022.

[10] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*, pages 3813–3824. IEEE, 2023.

[11] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023.