Hybrid Inpainting for 3D Meshes

EEE 515 Midterm Update

Shantanu Patne
EEE (AME)
1229496959
sapatne@asu.edu

Amogh Ravindra Rao
DSE (CDA)
1233433996
aravin17@asu.edu

Samudyata Sudarshan Jagirdar
CEN (EE)
1233568442
sjagird1@asu.edu

Surgit Madhavhen
DSE (EE)
1233313538
smadhavh@asu.edu

Kristy McWilliams
RAS-EE
1219163686
kamcwil1@asu.edu

***Abstract*—Applications spanning virtual reality, robotics, medical imaging, and archaeology depend on the restoration of partial 3D objects—known as 3D mesh inpainting. Particularly with complicated surface geometries and changing data losses, current reconstruction methods suffer in precisely reconstructing missing or degraded areas. This work suggests a hybrid strategy based on diffusion-based generative models improved by 3D latent diffusion techniques together with neural radiance fields (NeRFs). Data preparation from the Objaverse 1.0 dataset is done using a random masking technique. Using 3D reconstruction techniques, the model seeks to generate early structures directing successful inpainting. This work offers potential improvements in domains that depend mostly on correct digital restoration and visualization, so greatly advancing the dependability and generalizing powers of 3D reconstruction.**

I. Introduction

Real-world objects—damaged sculptures, incomplete medical scans, or weathered artifacts—like a vintage jigsaw puzzle with missing pieces can sometimes be partially recreated from context. Using technologies from geometry processing, computer vision, and deep learning, 3D reconstruction seeks to automate this process.

Main uses are cultural preservation, medical imaging, robotics, and virtual reality. Though they may struggle with occlusions, sparse coverage, or low-texture areas, classical techniques like Multi-View Stereo (MVS) [1] and Structure-from-Motion (SfM) [2] recreate 3D geometry from photographs.

Gaps in 3D meshes—caused by occlusion, limited sensor range, or reflective surfaces—can impair downstream activities like rendering or simulation. While they may try to recover, surface- and volume-based hole-filling techniques have difficulty with uneven or big gaps [3], usually striking a compromise between accuracy and efficiency [4].

In this project, we propose a hybrid model that combines traditional geometric approaches with learning-based approaches to generate structurally sound 3D in-painted models. This paper is laid out as follows: Sec. II explains the existing research into this domain. Sec. III outlines the experimental setup, including the data preparation and the system design. Sec. IV explains the chosen methods in detail with Sec. V showcasing the results from these methods.

II. Related Work

Over the past decade, the field of 3D shape completion and mesh inpainting has evolved from traditional geometric heuristics to powerful learning-based techniques. Early methods were largely categorized into surface-based and volume-based approaches. Surface-based techniques locally interpolate geometry to fill gaps, while volume-based methods voxelize point clouds and perform global remeshing. Although foundational, these classical techniques often struggle with large holes, complex topologies, and fine geometric details.

The emergence of deep learning redefined the paradigm for 3D inpainting. Initial efforts, such as those by Hernández-Bautista and Melero [5], projected 3D curvature data into 2D image space for convolutional inpainting before reconstructing the modified shape in 3D. While effective on simpler geometries, such approaches are limited to disk-homeomorphic regions and falter on arbitrary mesh topologies. To address this, more advanced architectures—including 3D convolutional GANs, autoencoders, and recurrent convolutional networks—have been introduced. For instance, Wang et al. [6] utilize recurrent structures to enhance spatial and temporal consistency, reflecting a broader trend toward coarse-to-fine strategies for modeling global shape structures.

Point- and graph-based completion methods also contribute valuable perspectives. The Point Completion Network (PCN) [7] adopts a coarse-to-fine decoder to recover dense point clouds from partial inputs. Graph-based models like GASCN [8] incorporate attention and hierarchical reasoning to infer complex topologies.

Recently, diffusion-based generative models have gained prominence for high-fidelity 3D reconstruction. RenderDiffusion [9] employs a triplane latent space to generate 3D shapes from image data, while Bayesian Diffusion Models

(BDM) [10] integrate top-down generative priors with bottom-up likelihoods for probabilistic mesh generation. These models offer improved detail synthesis and controllable structure, pushing the boundaries of inpainting fidelity.

In parallel, Neural Radiance Field (NeRF)-based techniques have significantly advanced photorealistic 3D reconstruction from sparse multi-view inputs. Methods like MVIP-NeRF [11] and NeRFiller [12] apply Score Distillation Sampling and diffusion-informed priors to ensure multi-view geometric and photometric consistency. Though primarily designed for dynamic scene completion, they showcase how multi-view priors can inform structural reconstruction. However, such methods often depend on calibrated cameras or high-resolution RGB inputs, limiting applicability in sparse or single-view settings.

Finally, mesh-specific and self-supervised approaches have explored inpainting through direct manipulation of mesh topology. Hattori et al. [13] utilize spectral graph convolutional networks (SGCNs) and multi-scale GCNs (MGCNs) to guide vertex displacement using learned structural priors. Although effective on unlabeled data, such models often lack mechanisms to enforce global consistency or integrate multi-view cues.

III. Experimental Setup

A. Data Preparation

Creating a library of hole meshes—which are needed to mimic real-world situations where 3D scans could be partial or damaged—is the first stage in this initiative. Starting with the Objaverse 1.0 dataset [14], a large-scale collection of over 800,000 3D mesh objects gathered from various online sites. Although Objaverse is a great source, its combination of single-object meshes and complete 3D sceneries is not always appropriate for object-centric learning activities.

For every mesh, we randomly choose two vertices separated by at least geodesic distance d. This restriction guarantees that the chosen vertices are not obviously near and lets us create spatially scattered holes. We provide a local area with surface area m around each chosen vertex; this area is then removed from the mesh. Local surface patch extraction followed by face deletion of the related patches removes the mesh, hence producing realistic missing areas.

B. System Design

We specify the general system for the project as indicated in Fig. 1 once we have established our assessment criteria and input data. We begin with input masked data, in which portions of the object models are purposely masked or erased to mimic missing geometry. Training and assessment of inpainting algorithms depend on this stage since it enables regulated generation of partial three-dimensional data.

The masked objects produced after masking are sent to two distinct paths: a collection of baseline models and our hybrid model. Serving as a standard for comparison, the baseline models reflect current or conventional methods for the inpainting work. Conversely, the hybrid model being developed aims to enhance the quality and durability of inpainting by using several techniques—maybe integrating geometric, learning-based, or generative approaches.

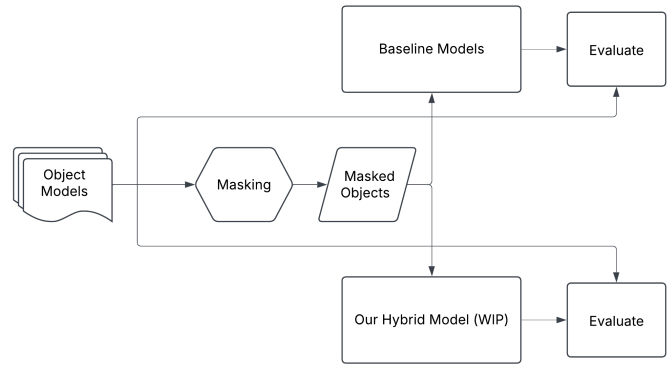

Fig. 1. System Design

The hybrid model and the baseline models both inpaint on the masked objects and generate finished copies of the 3D data. The evaluation part then receives these results to evaluate the inpainting quality. This two-path assessment guarantees the team's hybrid model outperforms or at least matches the performance of current techniques and lets them iteratively enhance it. We use PSNR and SSIM as evaluation metrics for quantitative analysis.

IV. Methodology

A. Geometric (Classical) Approaches

1) MeshFix: The first baseline we apply for the hole-filling task is the readily accessible MeshFix [15] algorithm, along with its Python port `pymeshfix`. Designed to fix gaps and flaws in 3D triangle meshes, MeshFix is a geometric and topological method. Originally meant to generate watertight meshes from erroneous or partial 3D scans, it has since developed into a standard geometric mesh preprocessing tool. The program aims to reconstruct a valid surface using higher-order geometric features of the surrounding mesh, concentrating on locating non-manifold edges and topological anomalies.

2) Poisson Surface Reconstruction: The second baseline we employ is a classical reconstruction approach based on Poisson Surface Reconstruction [16], a widely used algorithm for creating watertight meshes from oriented point clouds. This technique assumes a smooth indicator function whose gradient matches the input point cloud normals, solving a Poisson equation to reconstruct the surface.

To build this pipeline, we first convert each masked mesh into a point cloud using uniform surface sampling from the visible mesh. We then estimate surface normals and apply Poisson reconstruction with depth-adaptive filtering and density-based outlier removal. To mitigate boundary artifacts and floating structures, we prune out low-density regions and clip vertices that lie far outside the original mesh bounds. We also apply post-processing filters, including Laplacian smoothing and degenerate triangle removal, to ensure geometric plausibility.

B. Learning-Based Approaches

1) PCN: The Point Completion Network (PCN) follows a coarse-to-fine approach to reconstruct a dense point cloud from inputs with holes. The initial implementation of the Point Completion Network (PCN) began with the PCN-PyTorch repository [17] but was hindered by compiler errors and incompatible dependencies. As a fallback, the original TensorFlow-based implementation by the CVPR 2018 authors [7] was adopted, which, although stable, was not modular and was outdated.

To improve generalization and reconstruction quality, a custom hybrid PCN architecture was developed. The encoder used three 1D convolutional layers with ReLU activation and global max pooling to produce a compact shape descriptor. A coarse decoder generated 1024 initial points through two fully connected layers. The fine decoder, inspired by FoldingNet [18], refined these using a 2D grid and skip-connected 1D convolutional layers, enabling detail preservation (Fig. 2). The model was trained using a composite loss—Chamfer Distance, repulsive loss, and structure loss—ensuring both geometric accuracy and surface consistency.

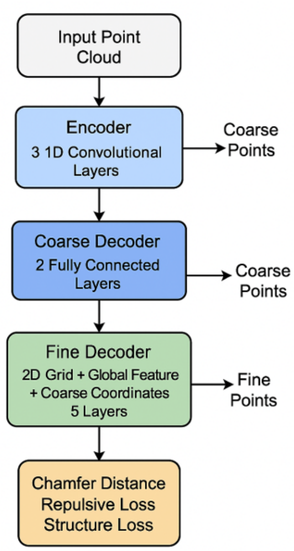

Fig. 2. PCN System Design

2) NeRFiller: NeRFiller is a 3D inpainting framework that uses off-the-shelf 2D inpainting models for 3D generative inpainting [12]. The model uses a trained base NeRF model, usually the Nerfacto available in NerfStudio [19], and fine-tunes the NeRF renders using 2D inpainting techniques. The model also uses a grid-based approach, known as *Grid-Prior*, for multi-view consistent inpainting. The renders are arranged in a grid of 40 images and then inpainted using the chosen inpainter. For our project, we employ a simple RGB npainter using Stable Diffusion [20]. We also inpaint the depth of the missing region using the ZoeDepth [21] metric depth estimator. The model iteratively generates the NeRF renders,

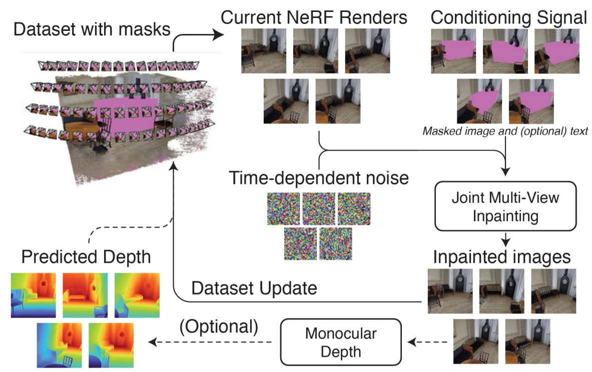

Fig. 3. NeRFiller model [12]

inpaints the RGB and depth, and updates the dataset for further NeRF training. This creates a fine-tuned NeRF representation that has filled in the occluded regions in the original structure.

The base Nerfacto is trained for $30k$ iterations, following which we iteratively inpaint the dataset every 1000 iterations for a total of 30 times. The perceptual similarity metric, LPIPS [22], is used as a loss metric for the training. We lower the number of training rays in the inpainting stage to match the GPU memory constraints of our machines.

C. Hybrid Approach

To overcome the structural limitations and occasionally incomplete geometry produced by NeRF-based outputs, we designed a hybrid pipeline that combines the best of both neural rendering and classical geometric reconstruction. As outlined in the flowchart (Fig. 4), the process begins by transforming NeRF-rendered RGB and depth images into a dense point cloud using Open3D utilities. The NeRFiller model we used was trained for approximately 10,000 iterations, and it provides photometrically consistent outputs from masked inputs. Simultaneously, the known geometry captured in the form of a partial mesh is uniformly sampled into another point cloud containing 3,000 surface points.

These two sources, one rich in visual detail and the other grounded in actual geometry, are then spatially aligned using centroid-based translation. The combined point cloud is cleaned with voxel-based downsampling and further refined by estimating surface normals. We then apply the Ball Pivoting Algorithm (BPA) for surface reconstruction, using an adaptive pivoting radius derived from the mean nearest neighbor distance. This results in a watertight mesh that not only reflects the texture fidelity from the NeRF but also respects the underlying structure of the original partial geometry.

This hybrid approach allows us to preserve sharp edges, recover small surface features, and maintain consistency with original geometry, something that is often difficult to achieve with NeRF or traditional methods alone. By fusing these complementary strengths, our model aims to strike a meaningful balance between data-driven visual quality and structural realism.

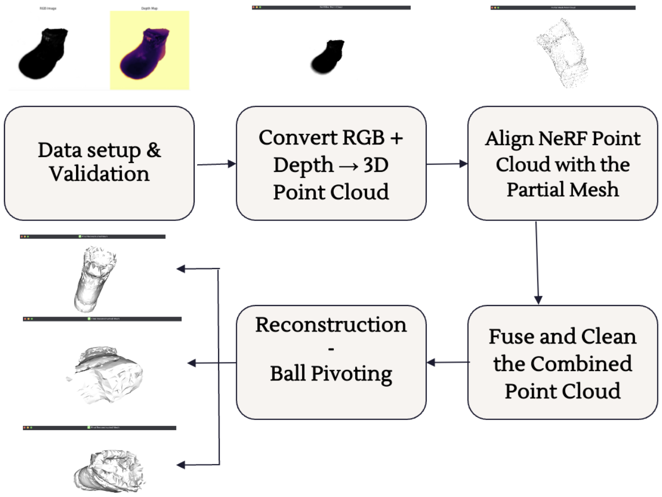

Fig. 4. Hybrid Model System Design

V. Results & Discussion

A. Geometric Approaches

The features of MeshFix's performance are shown in Fig. 5. The mesh on the left seems waterproof and physically complete, which is a benefit of MeshFix's geometric repair approach. The related error map, however, shows significant discrepancies, especially in areas needing more subtle geometric or semantic interpretation.

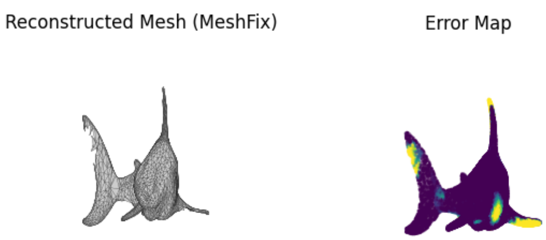

Fig. 5. Hole Filling using MeshFix and Error maps

The Poisson-based reconstruction technique shows promise for correcting small geometric flaws, as seen by the precise restoration of the shark's smaller missing area in Fig. 6. Its efficacy, though, declines with bigger or more complex holes. Reconstruction of the shark, for instance, with a more significant hole (around the tail area) exposes surface flow discontinuities ascribed to the lack of support from adjacent normals and inadequate structural gradients.

B. PCN

The results of the PCN approach are shown in Figs. 7 and 8. These outputs demonstrate that PCN effectively recovers the overall shape for simpler geometries, maintaining a coherent structure while filling gaps. However, surface distortions are

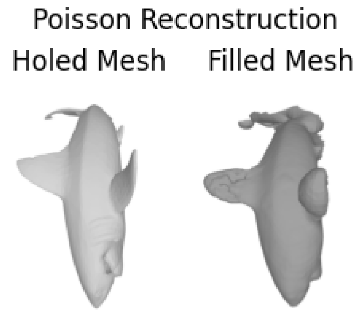

Fig. 6. Hole Filling using Poisson Reconstruction

prevalent, which indicates a loss of fine detail. For more complex geometries, PCN struggles to retain intricate features, leading to a completed but imprecise mesh with significant deviations from the expected structure. Particularly around sharp edges and detailed regions, the fine decoder fails to preserve object-specific detail, resulting in a noisy output that lacks accuracy.

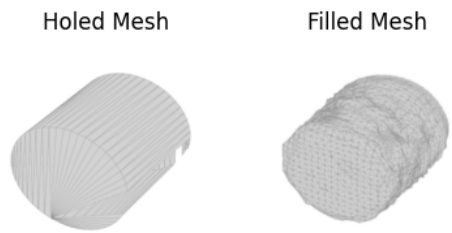

Fig. 7. Hole Filling using PCN Sample 1

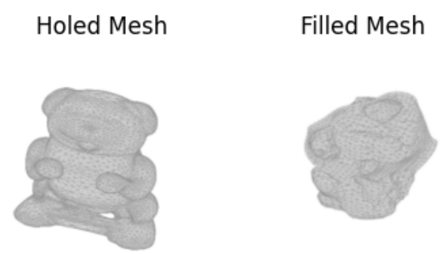

Fig. 8. Hole Filling using PCN Sample 2

C. NeRFiller

The results of the NeRFiller are shown in Fig. 9. We use the *chair* model provided by the original authors of NeRFiller along with the prepared *box* and *trawler* models from Objaverse. We can observe from the images that the base Nerfacto fails to properly fill the holes and reconstruct the missing regions, while the *Grid-Prior* method generates a well-defined structure with inpainted textures.

We can compare this quantitatively, as in Tab. I. The Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity

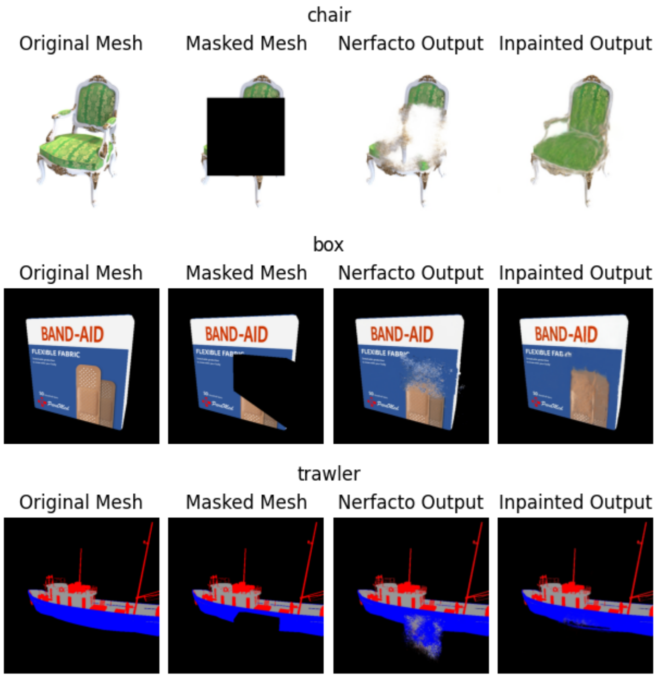

Fig. 9. NeRFiller outputs for mesh inpainting

Index Measure (SSIM) are used as evaluation metrics for this approach. The *Grid-Prior* approach shows very good results when compared to the Nerfacto. With a PSNR jump of 15 dB for the chair model and almost 10 dB for the box and trawler models, the inpainted models show good reconstruction.

TABLE I
EVALUATION OF NERFILLER METHODS

Mesh	Method	PSNR	SSIM
Chair	Nerfacto	6.95	0.733
	Grid-Prior	21.42	0.879
Box	Nerfacto	17.95	0.907
	Grid-Prior	31.57	0.974
Trawler	Nerfacto	22.59	0.912
	Grid-Prior	33.81	0.978

While the authors propose the grid-based approach, they also provide individual inpainting frameworks. However, none of these are able to generate a NeRF representation even though the 2D inpainted views are accurate.

D. Hybrid model

The final output from our hybrid pipeline genuinely captures the essence of what we aimed to achieve. As shown in Fig. 10, the reconstructed mesh manages to retain the shape and surface features of the original boot quite faithfully. It was especially exciting to see how the NeRF-generated point clouds, when fused with the partial geometry, brought out intricate details like the folds and edges around the ankle, details that are often easy to overlook. The Ball Pivoting step really tied it all together by forming a watertight mesh that felt natural and continuous. That said, not everything is perfect. The front toe region still appears under-reconstructed, which reflects some of the known limitations with visibility and sampling. But even with these challenges, the result is a promising step toward better geometry completion in real-world settings.

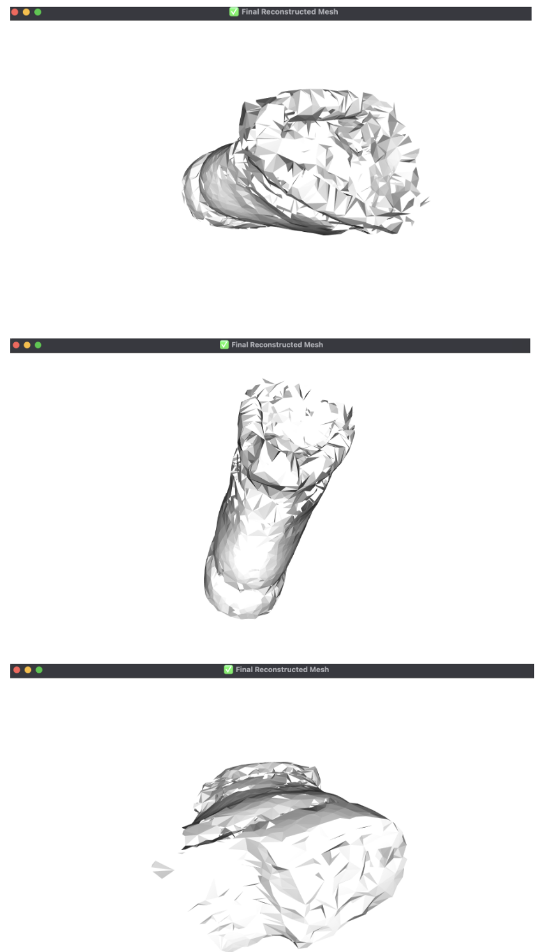

Fig. 10. Hybrid Model outputs

Across all evaluated methods, the results highlight the inherent trade-offs between traditional geometric, learning-based, and hybrid inpainting strategies. While MeshFix and Poisson reconstruction offered rapid mesh repair with some success on smaller gaps, they struggled with structural realism in more complex regions. The PCN-based learning approach performed better on overall shape recovery but lacked detail fidelity in intricate regions. NeRFiller significantly improved textural reconstruction through multi-view inpainting, especially when combined with depth-aware inputs, yet fell short in fully recovering structural integrity. The hybrid model effectively leveraged the strengths of both NeRF-generated visual cues and geometric priors from partial meshes, yielding the most balanced reconstructions in terms of both completeness and detail preservation. These results validate the promise of multi-modal fusion techniques in advancing the state of 3D mesh inpainting.

VI. Future Objectives & Conclusion

Our current hybrid inpainting pipeline, which fuses geometry-aware models with radiance field-driven predictions, has shown promising results in restoring missing mesh regions with improved visual realism and structural coherence. However, the approach remains a work in progress. We intend to expand its capabilities by incorporating semantic guidance to better inform the inpainting process, especially in scenarios where the missing regions are functionally or contextually significant. Furthermore, we plan to refine the fusion methodology between NeRF-derived point clouds and partial geometric data, potentially leveraging learned alignment techniques that can more intelligently resolve inconsistencies in scale, orientation, or density. A broader evaluation on a larger sample of objects from Objaverse is also under consideration to better validate the generalization potential of our pipeline.

In conclusion, this work presents a hybrid mesh inpainting framework that balances the structural reliability of geometric methods with the expressiveness of neural rendering. Through careful integration of classical surface reconstruction with NeRF-based visual inference, our method achieves more watertight and visually improved reconstructions than conventional approaches alone. While certain limitations exist, the results highlight the feasibility of combining data-driven priors with explicit surface information. This project lays the foundation for future research in 3D restoration pipelines that are both visually coherent and structurally sound, offering potential applications in cultural heritage preservation, autonomous perception, and immersive virtual environments.

References

[1] Steven M. Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 519–528, 2006.

[2] Johannes L. Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. *arXiv preprint arXiv:1604.03550*, 2016.

[3] Xiaoyuan Guo, Jun Xiao, and Ying Wang. A survey on algorithms of hole filling in 3d surface reconstruction. *Unpublished manuscript*, 2021.

[4] Emiliano Pérez, Santiago Salamanca, Pilar Merchán, and Antonio Adán. A comparison of hole-filling methods in 3d. *International Journal of Applied Mathematics and Computer Science*, 26(4):885–903, 2016.

[5] Marina Hernández-Bautista and FJ Melero. 3d hole filling using deep learning inpainting. *arXiv preprint arXiv:2407.17896*, 2024.

[6] Weiyue Wang, Qiangui Huang, Suya You, Chao Yang, and Ulrich Neumann. Shape inpainting using 3d generative adversarial network and recurrent convolutional networks, 2017.

[7] Wentao Yuan, Tejas Khot, David Held, Christoph Mertz, and Martial Hebert. Pcn: Point completion network. In *3D Vision (3DV), 2018 International Conference on*, 2018.

[8] Chao Zhang, Tao He, Yifang Liu, Yanzhi Liang, and Minglun Gong. Gascn: Graph attention shape completion network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2022.

[9] Titas Anciukevicius, Zexiang Xu, Matthew Fisher, Paul Henderson, Hakan Bilen, Niloy J Mitra, and Paul Guerrero. Renderdiffusion: Image diffusion for 3d reconstruction, inpainting and generation. In *CVPR*, 2023.

[10] Haiyang Xu, Yu Lei, Zeyuan Chen, Xiang Zhang, Yue Zhao, Yilin Wang, and Zhuowen Tu. Bayesian diffusion models for 3d shape reconstruction. *arXiv preprint arXiv:2403.06973*, 2024.

[11] Honghua Chen, Chen Change Loy, and Xingang Pan. Mvip-nerf: Multi-view 3d inpainting on nerf scenes via diffusion prior. In *CVPR*, 2024.

[12] Ethan Weber, Aleksander Hołyński, Varun Jampani, Saurabh Saxena, Noah Snavely, Abhishek Kar, and Angjoo Kanazawa. Nerfiller: Completing scenes via generative 3d inpainting, 2023.

[13] Shota Hattori, Tatsuya Yatagawa, Yutaka Ohtake, and Hiromasa Suzuki. Learning self-prior for mesh inpainting using self-supervised graph convolutional networks. *IEEE Transactions on Visualization and Computer Graphics*, 31(2):1448–1464, February 2025.

[14] 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. *arXiv preprint arXiv:2212.08051*, 2022.

[15] Marco Attene. A lightweight approach to repairing digitized polygon meshes. *The visual computer*, 26:1393–1406, 2010.

[16] Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. In *Proceedings of the Fourth Eurographics Symposium on Geometry Processing*, SGP '06, page 61–70, Goslar, DEU, 2006. Eurographics Association.

[17] Qingle Wang. Pcn-pytorch: Point completion network in pytorch. https://github.com/qinglew/PCN-PyTorch, 2021. Accessed: 2025-05-09.

[18] Yaoqing Yang, Chen Feng, Yiru Shen, and Dong Tian. Foldingnet: Point cloud auto-encoder via deep grid deformation, 2018.

[19] Matthew Tancik, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Justin Kerr, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, David McAllister, and Angjoo Kanazawa. Nerfstudio: A modular framework for neural radiance field development. In *ACM SIGGRAPH 2023 Conference Proceedings*, SIGGRAPH '23, 2023.

[20] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, June 2022.

[21] Shariq Farooq Bhat, Reiner Birkl, Diana Wofk, Peter Wonka, and Matthias Müller. Zoedepth: Zero-shot transfer by combining relative and metric depth, 2023.

[22] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018.