Slice3D: Multi-Slice, Occlusion-Revealing, Single View 3D Reconstruction

Supplementary Material

6. Rendering Details

Given a 3D model, we first normalize the length of its body diagonal to 1. Then, we put the 3D model in front of the camera with a distance of 1.2, i.e., the camera is pointed to the centre of the bounding box of the object with a distance of 1.2. Next, we randomly give a rotation and a scaling to the object. The rotation is done by introducing a pair of elevation and azimuth randomly sampled from the ranges of [-10°, 40°] and [0°, 360°], respectively. When introducing scaling, we restrict the 3D model to be always inside $[-0.5, 0.5]^3$, from where the query points are sampled in both training and inference.

This rendering rule is implemented in both ShapeNet [5] and Objaverse [10]. Note that positions of an object and the camera are relative. If the object is assumed to be fixed, we can tailor the rendering specifics by adjusting the camera through rotation.

7. Visualization of Hole-Filled Slices

As our ablation studies already showed (Tab. 3), operating on slice images *without* the holes filled provides more information about both the inside and outside structures of an object, yielding better reconstruction quality. In Fig. 12, we visually compared slice images with (left) vs. without holes filled.

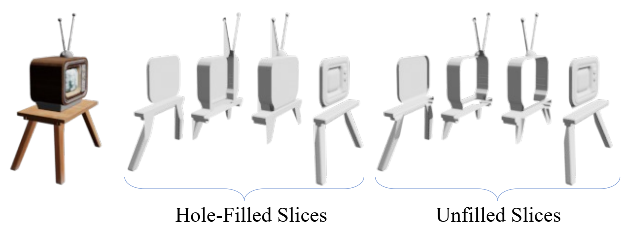

Figure 12. Slicing with and without filling holes. The demonstrated images are sliced along axis X of the object.

8. Details on Two Slicing Directions

As mentioned in Section 4, we utilize camera-aligned slicing for Objaverse dataset where many 3D models feature arbitrary orientations, i.e., they do not have clear canonical poses. In other cases, e.g., for ShapeNet shapes, we rely on the canonical poses of the objects and perform slicing along the X, Y, and Z directions with respect to those canonical poses. Fig. 13 illustrates and contrasts these two choices for the slicing directions, using a chair example from ShapeNet dataset.

The 3D models in ShapeNet possess canonical poses (e.g., chairs consistently have a front orientation) and they are aligned along the default X, Y, and Z axes. Consequently, objects are always sliced along their default X, Y, and Z axes regardless of the rendered views.

For 3D models in Objaverse, we use camera-aligned slicing. As described in Sec. 6, the camera is fixed and an object will be randomly rotated and scaled. We slice the bounding boxes of the object along the axes in the camera world.

9. Whether to Estimate Camera Poses

For our method, camera pose estimation is necessary only when we want to output 3D shapes in their canonical poses. In ShapeNet, we follow most existing methods [6, 37, 43, 67] to produce 3D shapes in their canonical poses, which involves camera pose estimation. We follow the strategy in DISN [67] to estimate the camera poses. First, assuming a fixed set of intrinsic parameters, we only need to predict a translation vector $\Theta_t \in \mathbb{R}^3$ and a rotation matrix $\Theta_r \in \mathbb{R}^{3\times3}$. Then, a CNN, e.g., VGG-16 [50], is trained to estimate $\hat{\Theta}_t$ and $\hat{\Theta}_r$ from input view I. Afterwards, we sample a point cloud $P \in \mathbb{R}^{N_p\times3}$ from the object along with its camera-aligned version P'. Finally, the loss function of the CNN is to align P' with P using $\hat{\Theta}_t$ and $\hat{\Theta}_r$, i.e., $\mathcal{L}_{cam} = \frac{1}{N_p}||P - (P'\Theta_r + \Theta_t)||_2^2$.

However, in Objaverse, where many objects do not have canonical poses and exhibit random orientations, we avoid estimating the rotations and translations of objects, and reconstruct the 3D shapes as they are in the camera world.

10. Training Cost and Inference Speed

Tab. 4 makes a detailed comparison of training cost and inference speed among single-view- reconstruction methods. Our method does not rely on big pre-trained model like Stable Diffusion [46]. Compared to multi-view based methods, our method runs much faster in the process of producing 3D meshes from slice images because we employ neural signed distance filed instead of NeRFs whose optimization is time-consuming.

11. Concatenation of Slice Images in Diffusion

We perform DDPM [21] on the entirety of slice images that can be stacked either on the color dimension or a spatial dimension. Note that concatenating along a spatial dimension significantly increases complexity because of the self attentions operated on the spatial dimensions in a diffusion network. By default, we stack them on the color channel

Method	Pre-trained model	Training data	Training GPU & time	Infer. speed
DISN [67]	VGG [50]	SPN-Chair	1 × A40 for 1 day	<5s
AutoSDF [37]	ResNet-18 [20]	SPN-Chair	unknown	<10s
NeRF-Img [43]	VGG-16 [50]	SPN-Chair	4 × A100	<30s
SSDNeRF [6]	N/A	SPN-Chair	2× 3090 for 6 days	45s-1min
Ours	VGG-16 [50]	SPN-Chair	1 × A40 for 2 days	<10s(R)/<20s(G)
Real-Fusion [34]	SD [46]	Per-case opt.	N/A	90min
One-2-3-45 [29]	Zero-1-to-3 [30]	Objv-40k*	2× A10 for 6 days	≈45s
Ours	VGG-16 [50]	Objv-40k	1× A40 for 3 days	<10s(R)/<20s(G)

Table 4. Training cost and inference speed of single-view-reconstruction methods. "SPN-Chair" denotes ShapeNet Chair dataset. "opt." denotes optimization. "Objv-40k" denotes a subset from Objaverse with around 40k 3D models. 'R' and and 'G' denote regression-based and generation-based slicing, respectively. Note that the first stage of One-2-3-45 (i.e., Zero-1-to-3) is trained with nearly the whole Objaverse dataset. In the second stage, it is trained with a subset in the scale of 40k 3D models. The inference speed is tested on a single Nvidia-A40-GPU for all methods.

Method	CD↓	F1↑	HD↓
Ours (G-C)	25.0	1.51	16.4
Ours (G-S)	**20.0**	1.51	**14.1**

Table 5. Quantitative results of single-view 3D reconstruction on the Objaverse dataset. 'G-C' and 'G-S' denote concatenating the slice images along the color and a spatial dimension, respectively.

to reduce the training and inference time. Tab. 5 provides a quantitative comparison of these two concatenation methods, revealing that concatenating along a spatial dimension achieves better performance than the color dimension. This outcome is logical as the former can model the spatial correspondence of different slice images throughout the diffusion network. Given sufficient computational resources, prioritizing concatenation along a spatial dimension is recommended.

12. View Inconsistency Problem

As mentioned in our main paper, recent methods [31, 32, 44, 54, 73] aim to enhance the consistency of synthesized views by performing spatial attention across different views. In Fig. 14, a comparison is made between Slice3D and SyncDreamer [31] from these works. The findings indicate that despite the utilization of expensive spatial attentions, the challenge of maintaining view consistency persists. This further substantiates the advantages of employing multi-slices over multi-views.

13. Visualization of Predicted Slice Images

The predicted slice images for the examples in Fig. 7 and 8 can be found in Fig. 15 and Fig. 16, respectively. Notably, our slice3D can produce slice images with a high level of consistency.

14. More Visual Results

More visual results and comparisons are provided in Fig. 17 and 18 for ShapeNet, Fig. 19 for Objaverse, and Fig. 20 for Google Scanned Objects (GSO) [12]. As apparent, our results respect the geometric details better than the other techniques while they do not suffer from unwanted artifacts or noise. Also compared to other techniques such as AutoSDF, it better respects the input image and does not retrieve a model that might look clean and noise-free but it is far from the input image (e.g., Fig.17; first two rows).

15. Image Resolution

Due to limited computing resources, the resolution of our input images and slice images is set to only 128. We plan to increase the resolution to 256 or 512 in the future and produce 3D meshes with better quality and details.

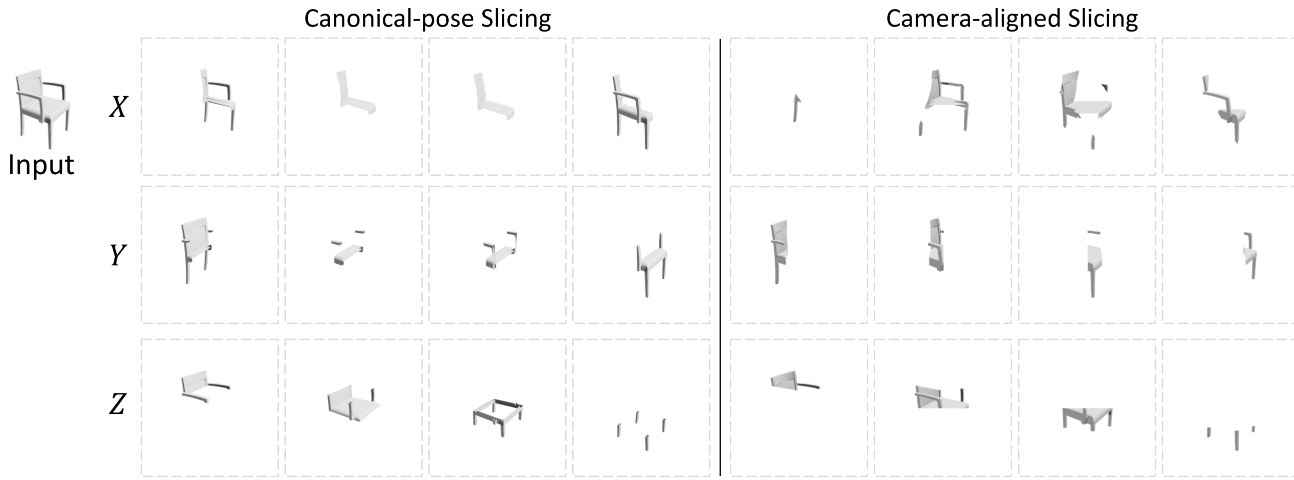

Figure 13. Canonical-pose slicing v.s. camera-aligned slicing. In canonical-pose slicing, the slicing directions are determined by the canonical pose of the object. In camera-aligned slicing, the slicing directions are determined by the orientation of the camera.

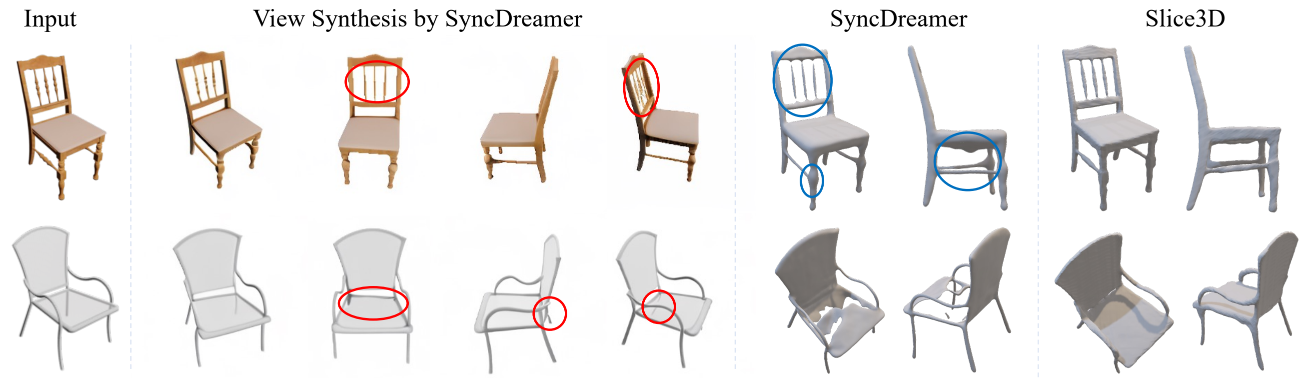

Figure 14. Visual comparison against SyncDreamer [31], which aims to enhance the consistency by performing spatial attention across different views. The red circles highlight the inconsistency across different synthesized views. For example, the pillars in the first chair and the rear legs in the second chair. The blue circles highlight the artifacts in the 3D mesh.

Figure 15. Visualization of the predicted (regressed) slice images from the input views.

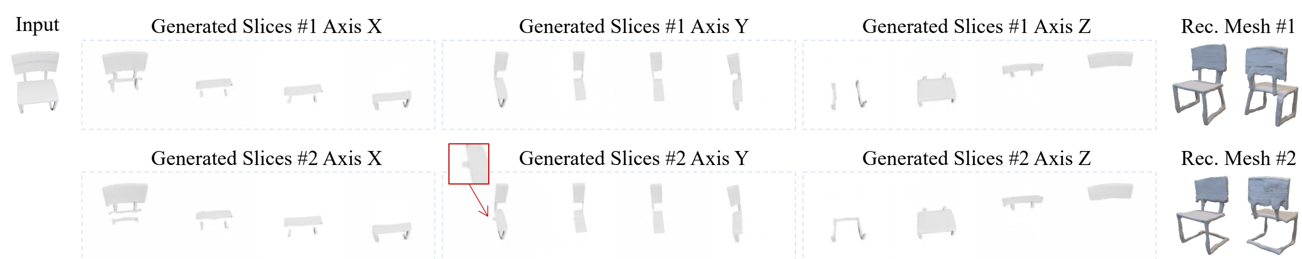

Figure 16. Multi-slice generation results and the meshes resulted from them.

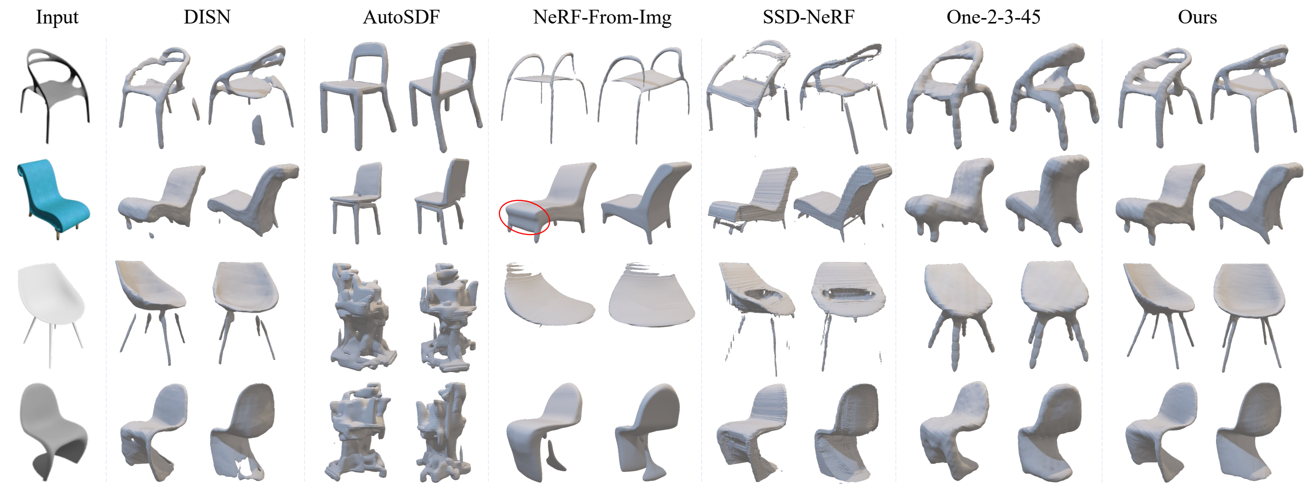

Figure 17. More visual comparison between single-view 3D reconstruction methods on ShapeNet chairs. DISN and our method (based on regressive slicing) utilize the same estimated camera parameters. Two different views are displayed to remove view bias.

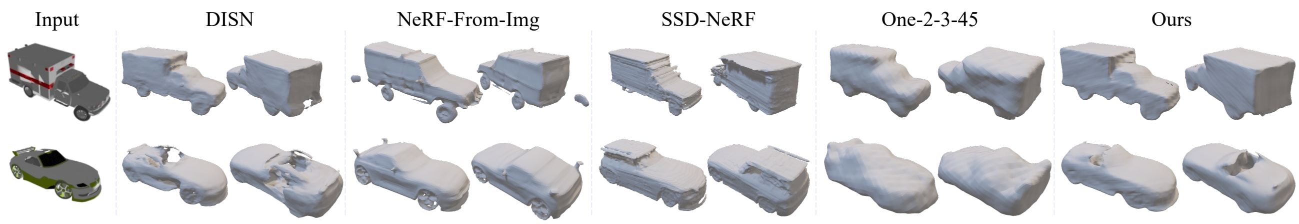

Figure 18. Visual comparison between single-view 3D reconstruction methods on two ShapeNet cars.

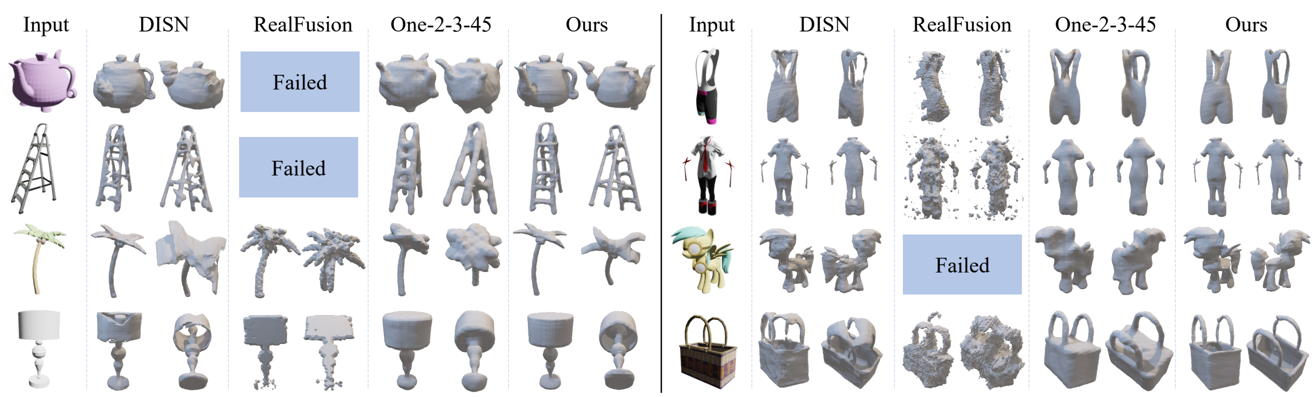

Figure 19. More results on Objaverse. “Failed” denotes no meaningful results after several optimizations of RealFusion.

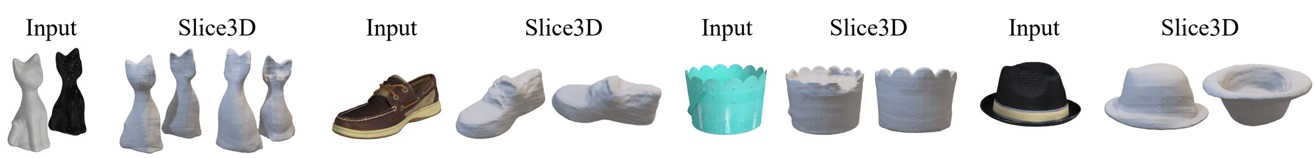

Figure 20. More results on GSO [12] dataset.

References

[1] Bisect - blender manual. https://docs.blender.org/manual/en/2.80/modeling/meshes/editing/subdividing/bisect.html. Accessed: 2023-11-15. 5

[2] Kfir Aberman, Oren Katzir, Qiang Zhou, Zegang Luo, Andrei Sharf, Chen Greif, Baoquan Chen, and Daniel Cohen-Or. Dip transform for 3D shape reconstruction. *ACM Transactions on Graphics*, 36(4):Article 79, 2017. 4

[3] Eric R Chan, Connor Z Lin, Matthew A Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J Guibas, Jonathan Tremblay, Sameh Khamis, et al. Efficient geometry-aware 3d generative adversarial networks. In *CVPR*, pages 16123–16133, 2022. 3

[4] Eric R Chan, Koki Nagano, Matthew A Chan, Alexander W Bergman, Jeong Joon Park, Axel Levy, Miika Aittala, Shalini De Mello, Tero Karras, and Gordon Wetzstein. Generative novel view synthesis with 3d-aware diffusion models. *arXiv preprint arXiv:2304.02602*, 2023. 3

[5] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 3, 5, 9

[6] Hansheng Chen, Jiatao Gu, Anpei Chen, Wei Tian, Zhuowen Tu, Lingjie Liu, and Hao Su. Single-stage diffusion NeRF: A unified approach to 3d generation and reconstruction. *arXiv preprint arXiv:2304.06714*, 2023. 2, 3, 6, 9, 10

[7] Wenzheng Chen, Jun Gao, Huan Ling, Edward J. Smith, Jaakko Lehtinen, Alec Jacobson, and Sanja Fidler. Learning to predict 3d objects with an interpolation-based differentiable renderer. *Advances in Neural Information Processing Systems*, 2019. 2

[8] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *CVPR*, pages 5939–5948, 2019. 3

[9] Zhiqin Chen, Andrea Tagliasacchi, and Hao Zhang. Bsp-net: Generating compact meshes via binary space partitioning. In *CVPR*, 2020. 1

[10] 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 *CVPR*, pages 13142–13153, 2023. 3, 5, 8, 9

[11] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794, 2021. 3

[12] Laura Downs, Anthony Francis, Nate Koenig, Brandon Kinman, Ryan Hickman, Krista Reymann, Thomas B McHugh, and Vincent Vanhoucke. Google scanned objects: A high-quality dataset of 3d scanned household items. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 2553–2560. IEEE, 2022. 6, 7, 10, 12

[13] John Flynn, Michael Broxton, Paul Debevec, Matthew DuVall, Graham Fyffe, Ryan Overbeck, Noah Snavely, and Richard Tucker. Deepview: View synthesis with learned gradient descent. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2367–2376, 2019. 4

[14] Yasutaka Furukawa, Carlos Hernández, et al. Multi-view stereo: A tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 9(1-2):1–148, 2015. 3

[15] Kyle Gao, Yina Gao, Hongjie He, Dening Lu, Linlin Xu, and Jonathan Li. Nerf: Neural radiance field in 3d vision, a comprehensive review. *arXiv preprint arXiv:2210.00379*, 2022. 3

[16] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 3

[17] Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan Russell, and Mathieu Aubry. Atlasnet: A papier-mâché approach to learning 3d surface generation. In *CVPR*, 2018. 1

[18] Jiatao Gu, Alex Trevithick, Kai-En Lin, Joshua M Susskind, Christian Theobalt, Lingjie Liu, and Ravi Ramamoorthi. Nerfdiff: Single-image view synthesis with nerf-guided distillation from 3d-aware diffusion. In *International Conference on Machine Learning*, pages 11808–11826. PMLR, 2023. 3

[19] Yuxuan Han, Ruicheng Wang, and Jiaolong Yang. Single-view view synthesis in the wild with learned adaptive multiplane images. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–8, 2022. 4

[20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 10

[21] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 3, 5, 9

[22] Chenbo Jiang, Jie Yang, Shwai He, Yu-Kun Lai, and Lin Gao. Neuralslice: Neural 3d triangle mesh reconstruction via slicing 4d tetrahedral meshes. 2023. 4

[23] Heewoo Jun and Alex Nichol. Shap-e: Generating conditional 3d implicit functions. *arXiv preprint arXiv:2305.02463*, 2023. 3

[24] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, pages 4401–4410, 2019. 3

[25] Shay Kels and Nira Dyn. Reconstruction of 3d objects from 2d cross-sections with the 4-point subdivision scheme adapted to sets. *Computers & Graphics*, 35:741–746, 2011. 4

[26] Manyi Li and Hao Zhang. D^2IM-Net: Learning detail disentangled implicit fields from single images. In *CVPR*, 2021. 1, 3

[27] Wallace Lira, Johannes Merz, Daniel Ritchies, Daniel Cohen-Or, and Hao Zhang. GANHopper: Multi-hop gan for unsupervised image-to-image translation. In *ECCV*, 2020. 3

[28] Lu Liu, Chandrajit Bajaj, Joseph O. Deasy, Daniel A. Low, and Tao Ju. Surface reconstruction from non-parallel curve networks. *Computer Graphics Forum*, 27:155–163, 2008. 4

[29] Minghua Liu, Chao Xu, Haian Jin, Linghao Chen, Zexiang Xu, Hao Su, et al. One-2-3-45: Any single image to 3d mesh

in 45 seconds without per-shape optimization. *arXiv preprint arXiv:2306.16928*, 2023. 1, 2, 3, 6, 8, 10

[30] Ruoshi Liu, Rundi Wu, Basile Van Hoorick, Pavel Tokmakov, Sergey Zakharov, and Carl Vondrick. Zero-1-to-3: Zero-shot one image to 3d object. *arXiv preprint arXiv:2303.11328*, 2023. 1, 3, 5, 8, 10

[31] Yuan Liu, Cheng Lin, Zijiao Zeng, Xiaoxiao Long, Lingjie Liu, Taku Komura, and Wenping Wang. Syncdreamer: Generating multiview-consistent images from a single-view image, 2023. 3, 4, 10, 11

[32] Xiaoxiao Long, Yuan-Chen Guo, Cheng Lin, Yuan Liu, Zhiyang Dou, Lingjie Liu, Yuexin Ma, Song-Hai Zhang, Marc Habermann, Christian Theobalt, et al. Wonder3d: Single image to 3d using cross-domain diffusion. *arXiv preprint arXiv:2310.15008*, 2023. 4, 10

[33] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Seminal graphics: pioneering efforts that shaped the field*, pages 347–353. 1998. 5

[34] Luke Melas-Kyriazi, Iro Laina, Christian Rupprecht, and Andrea Vedaldi. Realfusion: 360deg reconstruction of any object from a single image. In *CVPR*, pages 8446–8455, 2023. 3, 6, 8, 10

[35] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *CVPR*, pages 4460–4470, 2019. 3

[36] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 3

[37] Paritosh Mittal, Yen-Chi Cheng, Maneesh Singh, and Shubham Tulsiani. Autosdf: Shape priors for 3d completion, reconstruction and generation. In *CVPR*, pages 306–315, 2022. 3, 6, 9, 10

[38] Alex Nichol, Heewoo Jun, Prafulla Dhariwal, Pamela Mishkin, and Mark Chen. Point-e: A system for generating 3d point clouds from complex prompts. *arXiv preprint arXiv:2212.08751*, 2022. 3

[39] M. Niemeyer, L. Mescheder, M. Oechsle, and A. Geiger. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *CVPR*, 2020. 2

[40] Michael Oechsle, Songyou Peng, and Andreas Geiger. Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In *ICCV*, pages 5589–5599, 2021. 3

[41] Azimkhon Ostonov. Cut-and-approximate: 3d shape reconstruction from planar cross-sections with deep reinforcement learning. *arXiv preprint arXiv:2210.12509*, 2022. 4

[42] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *CVPR*, pages 165–174, 2019. 3

[43] Dario Pavllo, David Joseph Tan, Marie-Julie Rakotosaona, and Federico Tombari. Shape, pose, and appearance from a single image via bootstrapped radiance field inversion. In *CVPR*, pages 4391–4401, 2023. 2, 3, 6, 9, 10

[44] Guocheng Qian, Jinjie Mai, Abdullah Hamdi, Jian Ren, Aliaksandr Siarohin, Bing Li, Hsin-Ying Lee, Ivan Skorokhodov, Peter Wonka, Sergey Tulyakov, et al. Magic123: One image to high-quality 3d object generation using both 2d and 3d diffusion priors. *arXiv preprint arXiv:2306.17843*, 2023. 3, 4, 10

[45] Stephan R. Richter and Stefan Roth. Matryoshka networks: Predicting 3d geometry via nested shape layers. In *CVPR*, 2018. 1

[46] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, pages 10684–10695, 2022. 3, 6, 7, 9, 10

[47] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 4

[48] Manolis Savva, Fisher Yu, Hao Su, M Aono, B Chen, D Cohen-Or, W Deng, Hang Su, Song Bai, Xiang Bai, et al. Shrec16 track: largescale 3d shape retrieval from shapenet core55. In *Proceedings of the eurographics workshop on 3D object retrieval*, 2016. 5

[49] Haim Sawdayee, Amir Vaxman, and Amit H. Bermano. Orex: Object reconstruction from planar cross-sections using neural fields. In *CVPR*, pages 20854–20862, 2023. 4

[50] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2015. 9, 10

[51] Fun Shing Sin, Daniel Schroeder, and Jernej Barbič. Vega: non-linear fem deformable object simulator. In *Computer Graphics Forum*, pages 36–48. Wiley Online Library, 2013. 6

[52] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265. PMLR, 2015. 3

[53] Junshu Tang, Tengfei Wang, Bo Zhang, Ting Zhang, Ran Yi, Lizhuang Ma, and Dong Chen. Make-it-3d: High-fidelity 3d creation from a single image with diffusion prior. *arXiv preprint arXiv:2303.14184*, 2023. 3

[54] Shitao Tang, Fuyang Zhang, Jiacheng Chen, Peng Wang, and Yasutaka Furukawa. Mvdiffusion: Enabling holistic multi-view image generation with correspondence-aware diffusion. *arXiv preprint arXiv:2307.01097*, 2023. 4, 10

[55] Maxim Tatarchenko, Alexey Dosovitskiy, and Thomas Brox. Octree generating networks: Efficient convolutional architectures for high-resolution 3d outputs. In *ICCV*, 2017. 1

[56] Maxim Tatarchenko, Stephan R Richter, René Ranftl, Zhuwen Li, Vladlen Koltun, and Thomas Brox. What do single-view 3d reconstruction networks learn? In *CVPR*, pages 3405–3414, 2019. 6

[57] Maxim Tatarchenko, Stephan R. Richter, René Ranftl, Zhuwen Li, Vladlen Koltun, and Thomas Brox. What do single-view 3d reconstruction networks learn? In *CVPR*, 2019. 2

[58] Richard Tucker and Noah Snavely. Single-view view synthesis with multiplane images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 551–560, 2020. 4

[59] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017. 3

[60] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 2, 5

[61] Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. Pixel2mesh: Generating 3d mesh models from single rgb images. In *ECCV*, pages 52–67, 2018. 3

[62] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689*, 2021. 3

[63] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 6

[64] Daniel Watson, William Chan, Ricardo Martin-Brualla, Jonathan Ho, Andrea Tagliasacchi, and Mohammad Norouzi. Novel view synthesis with diffusion models. *arXiv preprint arXiv:2210.04628*, 2022. 3

[65] Yiheng Xie, Towaki Takikawa, Shunsuke Saito, Or Litany, Shiqin Yan, Numair Khan, Federico Tombari, James Tompkin, Vincent Sitzmann, and Srinath Sridhar. Neural fields in visual computing and beyond. *Comput. Graph. Forum*, 2022. 3

[66] Dejia Xu, Yifan Jiang, Peihao Wang, Zhiwen Fan, Yi Wang, and Zhangyang Wang. Neurallift-360: Lifting an in-the-wild 2d photo to a 3d object with 360deg views. In *CVPR*, pages 4479–4489, 2023. 3

[67] Qiangeng Xu, Weiyue Wang, Duygu Ceylan, Radomir Mech, and Ulrich Neumann. Disn: Deep implicit surface network for high-quality single-view 3d reconstruction. *Advances in neural information processing systems*, 32, 2019. 1, 3, 5, 6, 9, 10

[68] Xingguang Yan, Liqiang Lin, Niloy J Mitra, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. Shapeformer: Transformer-based shape completion via sparse representation. In *CVPR*, pages 6239–6249, 2022. 3

[69] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. *Advances in Neural Information Processing Systems*, 34:4805–4815, 2021. 3

[70] Biao Zhang, Jiapeng Tang, Matthias Niessner, and Peter Wonka. 3dshape2vecset: A 3d shape representation for neural fields and generative diffusion models. *arXiv preprint arXiv:2301.11445*, 2023. 3

[71] Mingfang Zhang, Jinglu Wang, Xiao Li, Yifei Huang, Yoichi Sato, and Yan Lu. Structural multiplane image: Bridging neural view synthesis and 3d reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16707–16716, 2023. 4

[72] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pages 586–595, 2018. 4, 6

[73] Minda Zhao, Chaoyi Zhao, Xinyue Liang, Lincheng Li, Zeng Zhao, Zhipeng Hu, Changjie Fan, and Xin Yu. Efficientdreamer: High-fidelity and robust 3d creation via orthogonal-view diffusion prior. *arXiv preprint arXiv:2308.13223*, 2023. 4, 10

[74] Tinghui Zhou, Richard Tucker, John Flynn, Graham Fyffe, and Noah Snavely. Stereo magnification: Learning view synthesis using multiplane images. *arXiv preprint arXiv:1805.09817*, 2018. 4