SparseRecon: Neural Implicit Surface Reconstruction from Sparse Views with Feature and Depth Consistencies

Liang Han[1], Xu Zhang[3], Haichuan Song[2*], Kanle Shi[4], Yu-Shen Liu[1*], Zhizhong Han[5]
[1]School of Software, Tsinghua University, Beijing, China
[2]Computer Science and Technology, East China Normal University, Shanghai, China
[3]China Telecom [4]Kuaishou Technology, Beijing, China
[5]Department of Computer Science, Wayne State University, Detroit, USA
hanl23@mails.tsinghua.edu.cn, zhangxu@chinatelecom.cn, hcsong@cs.ecnu.edu.cn
shikanle@kuaishou.com, liuyushen@tsinghua.edu.cn, h312h@wayne.edu

1. Dataset Details

We selected the scenes similar to those in S-VolSDF [15] in both the DTU [7] dataset and the BlendedMVS [16] dataset. For the DTU [7] dataset, the scenes include scans 21, 24, 34, 37, 38, 40, 82, 106, 110, 114, and 118. We evaluated our method with both large-overlapping views and small-overlapping views in each scene. For the BlendedMVS [16] dataset, the scenes and sparse view indices are as follows: Doll: 9, 10, 55; Egg: 9, 52, 59; Head: 22, 26, 27; Angel: 11, 39, 53; Bull: 32, 42, 47; Robot: 28, 34, 57; Dog: 2, 5, 25; Bread: 16, 21, 33; Camera: 10, 16, 60.

(a) Large-overlapping views

(b) Small-overlapping views

Figure 1. Visual comparison between large-overlapping views and small-overlapping views.

Figure 1 presents a comparison between small-overlapping views and large-overlapping views. Small-overlapping views exhibit greater viewpoint variations, while large-overlapping views have smaller angular differences. Therefore, sparse view reconstruction is more challenging with small-overlapping views.

*Corresponding author: Yu-Shen Liu and Haichuan Song.

2. More Experiments

2.1. Additional Results on DTU

We also conduct experiments on the DTU dataset with large-overlapping views and perform both quantitative and visual comparisons with other overfitting-based, generalization-based, and several dense-view reconstruction methods. The quantitative results of the Chamfer Distance are presented in Table 1, which demonstrates that our method also achieves superior performance on large-overlapping views.

Figure 3 presents a visual comparison of more reconstruction results on DTU dataset with small-overlapping views.

2.2. Additional Results on BlendedMVS

Fig. 4 shows the additional reconstruction results on the BlendedMVS [16] dataset with small-overlapping views. Other methods produce rough or incomplete results, while our method can reconstruct more complete and detailed meshes. In the scene of **bread**, although our method can reconstruct the side, the large angles among the three views make it impossible to use COLMAP [13] to obtain a sparse point cloud for depth calibration. Therefore, the reconstruction of the side of the bread is inaccurate due to the lack of depth supervision.

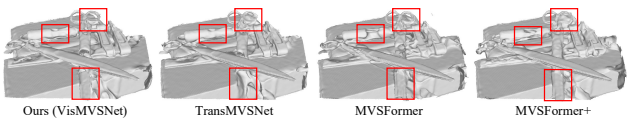

Figure 2. Reconstruction results by different feature extractors.

Methods	21	24	34	37	38	40	82	106	110	114	118	Mean
NeuS [14]	2.46	2.82	1.05	6.90	1.22	6.95	1.71	1.13	3.35	0.72	2.72	2.82
NeuralWarp [3]	1.91	0.67	0.75	1.71	0.73	1.37	1.63	1.17	0.86	0.57	1.02	1.13
MonoSDF [18]	2.44	1.77	1.48	4.14	1.64	2.11	2.77	5.38	3.89	0.83	3.43	2.72
Vis-MVSNet [19]	1.68	0.97	0.55	2.69	0.79	1.63	1.28	0.97	**0.43**	0.47	0.87	1.12
MVSDF [20]	2.36	1.22	0.76	3.95	0.95	1.98	1.61	0.94	0.60	0.48	0.90	1.43
SparseNeuS$_{ft}$ [9]	2.02	1.13	0.87	3.00	1.21	2.53	1.39	1.17	0.79	0.58	1.16	1.44
VolRecon [12]	1.55	1.27	0.81	2.63	0.99	1.65	1.44	1.20	1.37	0.74	1.23	1.35
ReTR [8]	1.47	1.05	0.69	2.31	0.83	1.44	1.32	1.09	0.77	0.59	1.06	1.15
UFORecon [10]	**1.33**	0.78	0.62	2.04	0.78	1.35	1.21	0.87	0.60	0.57	0.90	1.00
S-VolSDF [15]	1.81	0.90	0.79	2.28	1.04	1.61	1.80	1.01	0.73	0.71	1.21	1.26
NeuSurf [5]	3.01	0.78	0.99	2.35	0.85	1.55	1.14	0.74	0.49	**0.39**	**0.75**	1.19
SparseCraft [17]	1.80	1.17	0.68	1.74	0.90	1.80	1.37	0.80	0.56	0.44	0.77	1.09
FatesGS [6]	2.87	**0.64**	1.53	1.94	1.52	**1.17**	**1.06**	0.75	0.48	0.41	0.78	1.20
Ours	1.44	0.66	**0.49**	**1.68**	**0.71**	1.28	1.19	**0.66**	0.45	**0.39**	**0.75**	**0.88**

Table 1. Quantitative results on DTU dataset with *3 large-overlapping* images. The methods are divided into three categories, from top to bottom: (1) dense-view reconstruction methods related to ours, (2) generalization-based sparse-view reconstruction methods, and (3) overfitting-based sparse-view reconstruction methods. *Bold* results are the best score.

2.3. Efficiency Comparison

we compare the efficiency of all methods specifically designed for sparse view reconstruction in the baseline, including both generalization-based methods and overfitting-based methods. All methods are tested on a single NVIDIA RTX 3090 GPU, with the tests including training time and GPU memory usage, as detailed in Table 2. Although generalization-based methods can obtain reconstruction results in a few minutes, they typically require several days for pretraining and consume more GPU memory. More importantly, their generalization performance on new scenes is unsatisfactory, especially when input views have small overlaps. Among the various overfitting-based methods, our method achieves a comparable time consumption to SVolSDF [15]. Although SparseCraft [17] and FatesGS [6] have less time occupation, our method delivers higher-quality reconstruction results.

2.4. Additional Ablation Study

To validate the effectiveness of feature extractors from recent multi-view stereo methods within our framework, we conduct experiments by replacing the Vis-MVSNet feature extractor used in our method with those from TransMVSNet [4], MVSFormer [1], and MVS-Former++ [2], respectively. Table 3 presents the quantitative results on the DTU dataset under small-overlapping view setting. Figure 2 shows the visual comparison. Although these more recent methods reported better performance than the one we used on standard multi-view stereo benchmarks, we do not observe significant differences among them in our experiments.

Methods	Training Time		GPU Mem. Usage
	Pre-Training	Per-scene	
SparseNeuS$_{ft}$ [9]	$\sim$ 2.5 days	20 mins	7 GB
VolRecon [12]	$\sim$ 2 days	-	17 GB
GenS$_{ft}$ [11]	$\sim$ 3 days	25 mins	17 GB
ReTR [8]	$\sim$ 3 days	-	22 GB
UFORecon [10]	$\sim$ 10 days	-	23 GB
MonoSDF [18]	-	6 hours	14 GB
S-VolSDF [15]	-	3 hours	4 GB
NeuSurf [5]	-	14 hours	8 GB
SparseCraft [17]	-	**10 mins**	10 GB
FatesGS [6]	-	14 mins	4 GB
Ours	-	3.5 hours	8 GB

Table 2. Efficiency Comparison of sparse view construction methods. The GPU memory usage is obtained during training.

References

[1] Chenjie Cao, Xinlin Ren, and Yanwei Fu. MVSFormer: Multi-view stereo by learning robust image features and temperature-based depth. *Transactions of Machine Learning Research*, 2023. 2, 3

[2] Xinlin Ren Chenjie Cao and Yanwei Fu. Mvsformer++: Revealing the devil in transformer's details for multi-view stereo. In *International Conference on Learning Representations*, 2024. 2, 3

[3] François Darmon, Bénédicte Bascle, Jean-Clément Devaux, Pascal Monasse, and Mathieu Aubry. Improving neural implicit surfaces geometry with patch warping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6260–6269, 2022. 2

[4] Yikang Ding, Wentao Yuan, Qingtian Zhu, Haotian Zhang, Xiangyue Liu, Yuanjiang Wang, and Xiao Liu. TransMVSNet: Global context-aware multi-view stereo network with

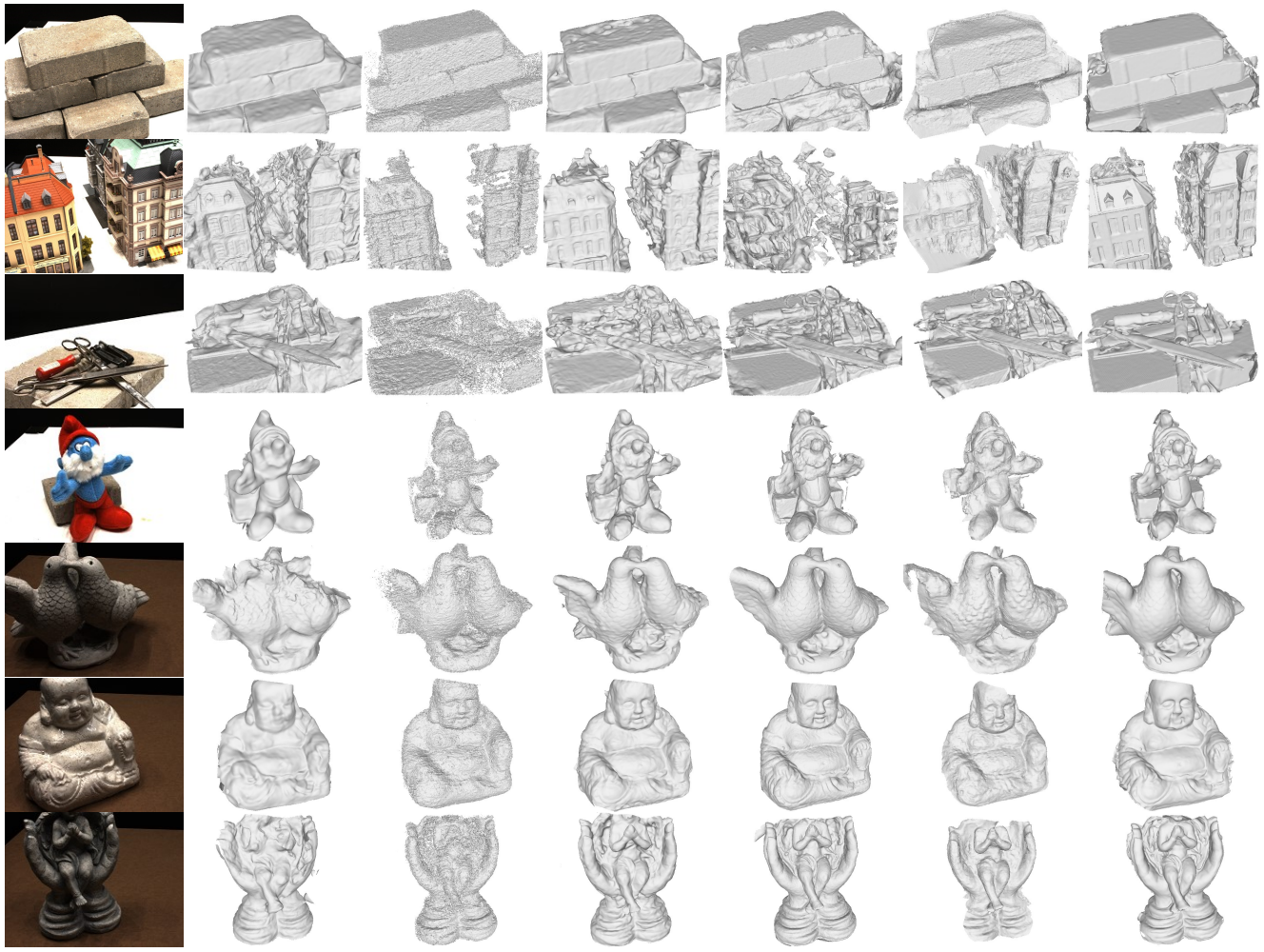

Figure 3. Visual results comparison on DTU dataset with *3 small-overlapping* images.

Feat. Extractors	21	24	34	37	38	40	82	106	110	114	118	Mean
TransMVSNet [4]	2.19	1.31	**0.71**	1.68	**0.84**	1.68	**1.26**	0.95	**0.56**	0.47	**0.83**	1.13
MVSFormer [1]	2.69	1.57	0.84	1.99	0.91	1.71	1.57	1.05	0.96	0.46	0.98	1.34
MVSFormer++ [2]	2.28	1.48	0.79	2.00	0.94	1.69	1.52	1.00	0.83	0.45	0.89	1.26
VisMVSNet (Ours) [19]	**2.14**	**1.26**	0.72	**1.46**	0.86	**1.39**	1.37	**0.94**	0.77	**0.44**	**0.83**	**1.11**

Table 3. Quantitative comparison of different feature extractors on the DTU dataset with *3 small-overlapping* images. *Bold* results are the best score.

transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8585–8594, 2022. 2, 3

[5] Han Huang, Yulun Wu, Junsheng Zhou, Ge Gao, Ming Gu, and Yu-Shen Liu. NeuSurf: On-surface priors for neural surface reconstruction from sparse input views. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2312–2320, 2024. 2

[6] Han Huang, Yulun Wu, Chao Deng, Ge Gao, Ming Gu, and Yu-Shen Liu. FatesGS: Fast and accurate sparse-view surface reconstruction using gaussian splatting with depth-feature consistency. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025. 2

[7] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engin Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In *Proceedings of the IEEE/CVF Conference on Com-*

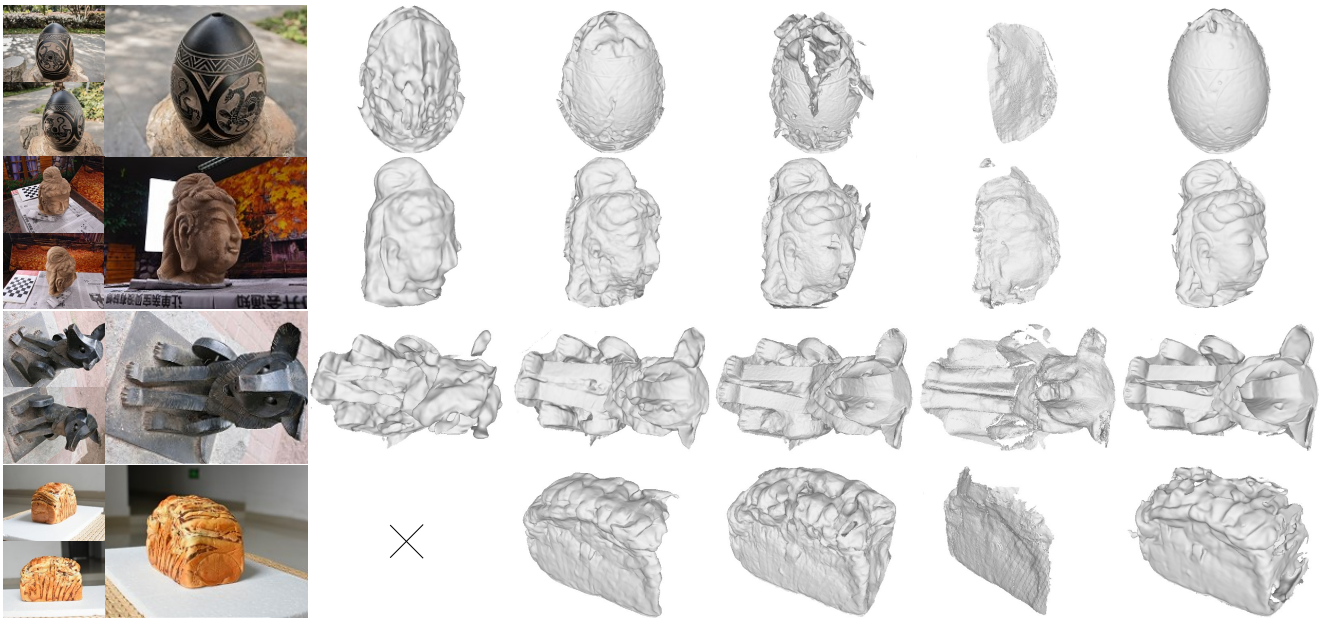

Figure 4. Visual results comparison on BlendedMVS dataset. '×' indicates reconstruction failure.

puter Vision and Pattern Recognition, pages 406–413, 2014. 1

[8] Yixun Liang, Hao He, and Yingcong Chen. ReTR: Modeling rendering via transformer for generalizable neural surface reconstruction. *Advances in Neural Information Processing Systems*, 36, 2024. 2

[9] Xiaoxiao Long, Cheng Lin, Peng Wang, Taku Komura, and Wenping Wang. SparseNeuS: Fast generalizable neural surface reconstruction from sparse views. In *European Conference on Computer Vision*, pages 210–227. Springer, 2022. 2

[10] Youngju Na, Woo Jae Kim, Kyu Beom Han, Suhyeon Ha, and Sung-Eui Yoon. UFORecon: Generalizable sparse-view surface reconstruction from arbitrary and unfavorable sets. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5094–5104, 2024. 2

[11] Rui Peng, Xiaodong Gu, Luyang Tang, Shihe Shen, Fanqi Yu, and Ronggang Wang. GenS: Generalizable neural surface reconstruction from multi-view images. In *Advances in Neural Information Processing Systems*, pages 56932–56945, 2023. 2

[12] Yufan Ren, Tong Zhang, Marc Pollefeys, Sabine Süsstrunk, and Fangjinhua Wang. VolRecon: Volume rendering of signed ray distance functions for generalizable multi-view reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16685–16695, 2023. 2

[13] Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision*, 2016. 1

[14] 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. *Advances in Neural Information Processing Systems*, 2021. 2

[15] Haoyu Wu, Alexandros Graikos, and Dimitris Samaras. S-VolSDF: Sparse multi-view stereo regularization of neural implicit surfaces. *International Conference on Computer Vision*, 2023. 1, 2

[16] Yao Yao, Zixin Luo, Shiwei Li, Jingyang Zhang, Yufan Ren, Lei Zhou, Tian Fang, and Long Quan. BlendedMVS: A large-scale dataset for generalized multi-view stereo networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1790–1799, 2020. 1

[17] Mae Younes, Amine Ouasfi, and Adnane Boukhayma. SparseCraft: Few-shot neural reconstruction through stereopsis guided geometric linearization. In *European Conference on Computer Vision*, pages 37–56. Springer, 2024. 2

[18] Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, and Andreas Geiger. MonoSDF: Exploring monocular geometric cues for neural implicit surface reconstruction. *Advances in Neural Information Processing Systems*, 35:25018–25032, 2022. 2

[19] Jingyang Zhang, Yao Yao, Shiwei Li, Zixin Luo, and Tian Fang. Visibility-aware multi-view stereo network. *The British Machine Vision Conference*, 2020. 2, 3

[20] Jingyang Zhang, Yao Yao, and Long Quan. Learning signed distance field for multi-view surface reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6525–6534, 2021. 2