FreBIS: Frequency-Based Stratification for Neural Implicit Surface Representations
—Supplementary Material—

Naoko Sawada[1,2] Pedro Miraldo[1] Suhas Lohit[1] Tim K. Marks[1] Moitreya Chatterjee[1]
[1]Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA
[2]Information Technology R&D Center, Mitsubishi Electric Corporation, Kanagawa, Japan
Sawada.Naoko@df.MitsubishiElectric.co.jp,{miraldo,slohit,tmarks,chatterjee}@merl.com

The following summarizes the supplementary materials we present:

1. Ablation study of the *redundancy-aware weighting* module.
2. Comparative study of the number of frequency levels.
3. Comparative study of encoder architecture variants.

1. Ablation study of the redundancy-aware weighting module

A key innovation of FreBIS is the *rendundancy-aware weighting* module which combines the complementary information from the different encoders by promoting mutual dissimilarity. Table 1 shows quantitative comparison results of the FreBIS with and without this module. The results show that our model with this module outperforms the variant without it, where a simple averaging of the encoder features is performed, clearly bringing out its effectiveness.

2. Comparative study of the number of frequency levels

We conduct experiments to study the effect of the choice of frequency levels N for both FreBIS and Scaled-up VolSDF [1]. As shown in Table 2 and Fig. 1, the Scaled-up VolSDF is sensitive to the choice of frequency levels and has particular difficulty in dealing with higher frequency encodings. In particular, the Scaled-up VolSDF with $N = 9$ results in a reconstructed mesh with too many bumps, while that with $N = 12$ results in a mesh that is hard to interpret. On the other hand, FreBIS is capable of processing higher–frequency information without sacrificing information gleaned from the low–frequency bands. Fig. 2 and 3 show the qualitative comparisons of rendered images and reconstructed meshes with $N = 6, 9, 12$ using FreBIS on the *Doll*, *Bull*, and *Robot* scenes.

3. Comparative study of encoder architecture variants

In order to design the encoders of FreBIS optimally, we study the effect of varying the number of layers of each of the three encoders of FreBIS and compare their performances. As seen from the results in Table 3 as well as Fig. 4, and Fig. 5. FreBIS performs comparably irrespective of the choice of encoder architecture, maintaining a good performance throughout. Based on this analysis and in order to stay consistent with the baseline VolSDF [1] architecture, we choose the 6–layer architecture for each encoder, with each layer having 256 dimensions.

References

[1] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume Rendering of Neural Implicit Surfaces. In *Proceedings of The Conference on Neural Information Processing Systems*. arXiv, 2021. arXiv:2106.12052 [cs]. 1

	(a) Scaled-up VolSDF	(b) Ours w/o redundancy-aware weighting	(c) Ours
PSNR (↑)	28.32	28.31	**28.56**
SSIM (↑)	0.949	0.950	**0.952**
LPIPS (↓)	0.028	0.027	**0.026**

Table 1. Ablation of the redundancy-aware weighting module: We show quantitative results for the *Dog* scene using the Scaled-up VolSDF, Ours without redundancy-aware weighting, and Ours.

Method	Frequency level (N)	Doll	Egg	Head	Angel	Bull	Robot	Dog	Bread	Camera	**Mean**
Scaled-up VolSDF	6	26.07	27.15	26.62	30.37	26.08	25.07	28.32	29.44	23.02	26.90
Ours	6	26.22	27.48	27.29	30.52	26.33	26.69	28.56	30.22	23.08	27.38
Scaled-up VolSDF	9	25.69	26.66	26.94	28.59	26.02	22.67	26.78	32.62	23.45	26.60
Ours	9	26.10	27.47	27.24	30.56	25.78	26.85	28.88	30.08	23.28	27.36
Scaled-up VolSDF	12	–	–	–	–	–	–	24.86	–	19.59	–
Ours	12	26.02	27.54	25.81	30.56	26.89	26.66	28.62	30.18	30.26	27.21

Table 2. Comparison of viewpoint-based rendering performance with a varying number of frequencies, as measured by PSNR. – denotes that the method failed to construct a mesh during training.

$N_{\mathrm{L}}, N_{\mathrm{M}}, N_{\mathrm{H}}$	Doll	Egg	Head	Angel	Bull	Robot	Dog	Bread	Camera	**Mean**
$6, 6, 6$	26.22	27.48	27.29	30.52	26.33	26.69	28.56	30.22	23.08	27.38
$5, 5, 5$	26.18	27.47	27.14	30.42	26.37	26.62	28.55	30.20	23.10	27.34
$4, 4, 4$	26.25	27.51	26.96	30.49	26.37	26.51	28.18	31.12	23.17	27.39
$4, 5, 6$	26.18	27.45	27.13	30.50	26.38	26.64	28.60	30.16	23.19	27.36
$2, 4, 6$	26.26	27.47	24.45	30.44	25.95	26.67	28.74	31.60	23.21	27.19

Table 3. Performance comparison of variants of FreBIS with varying number of encoder layers, as measured by PSNR.

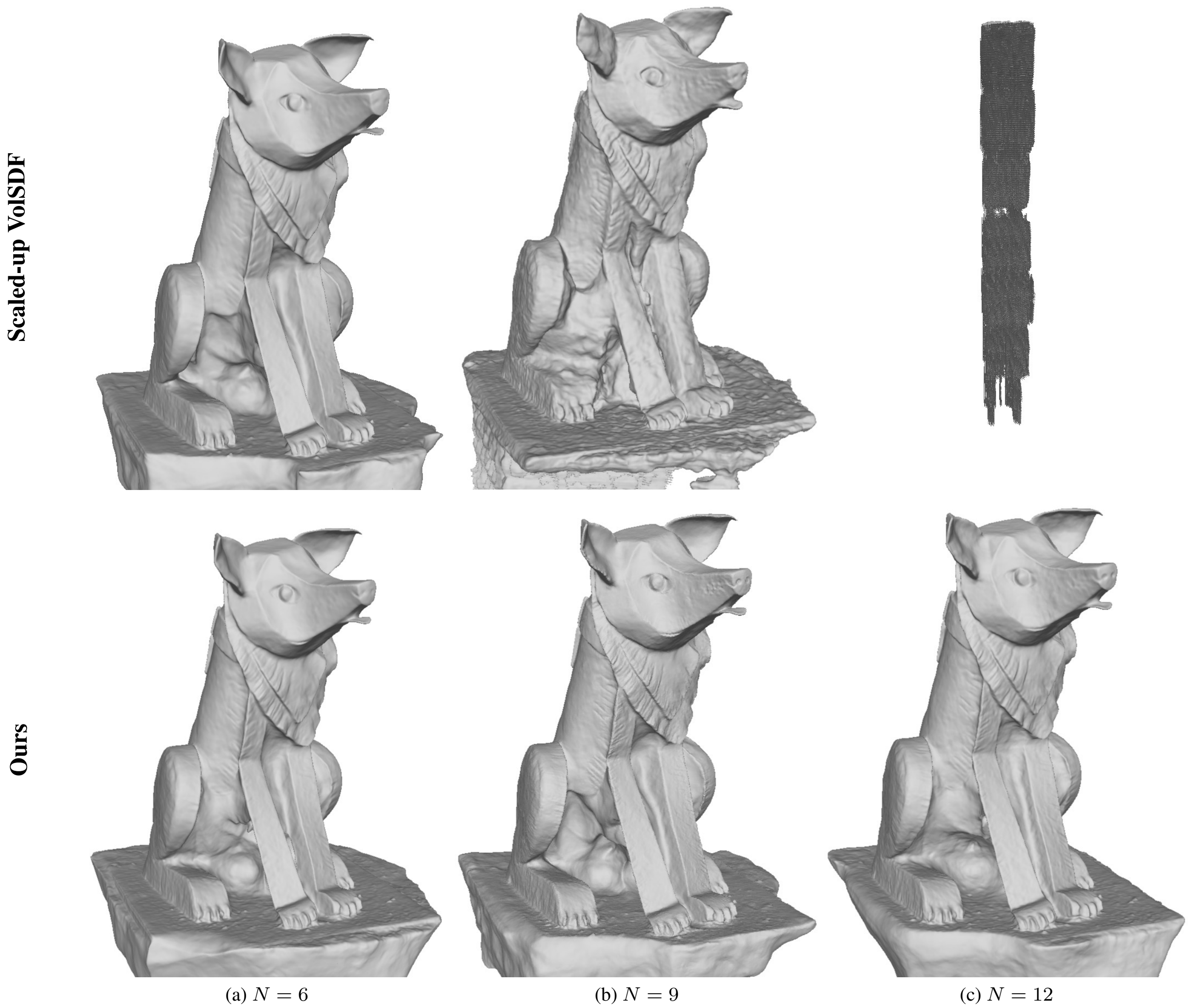
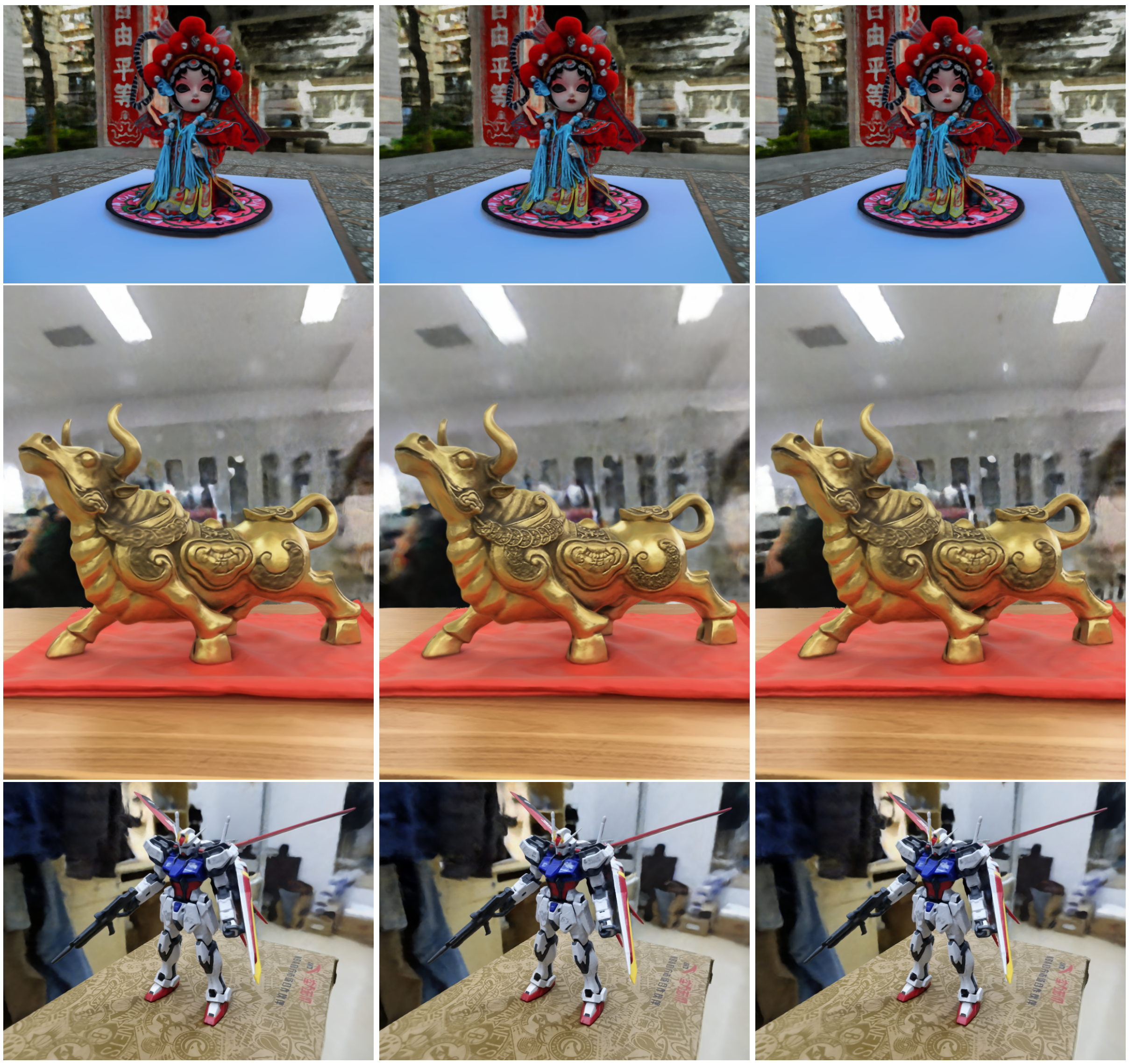


Figure 1. Qualitative comparison on the capability to deal with higher frequencies.

Doll

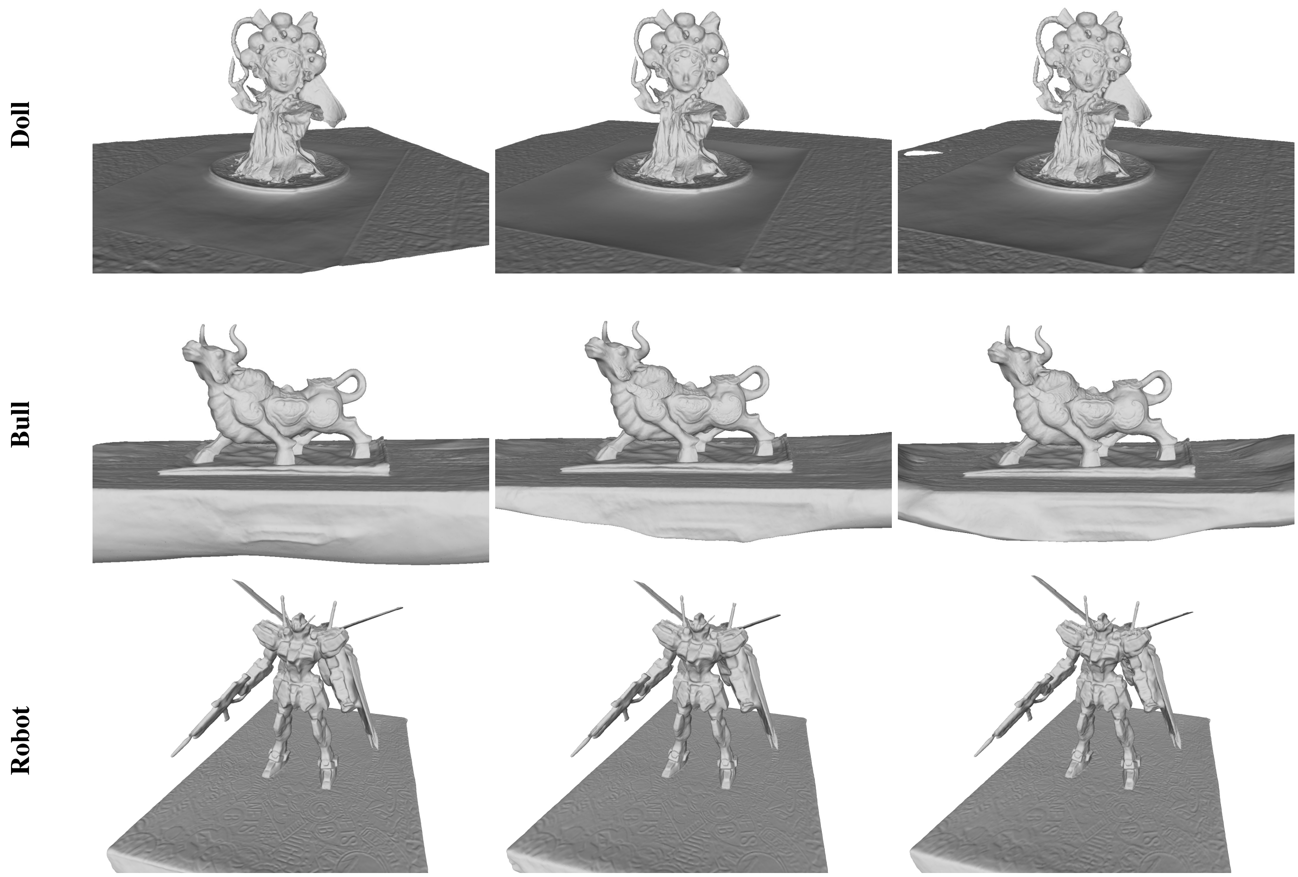
Bull

Robot



(a) $N = 6$ (b) $N = 9$ (c) $N = 12$

Figure 2. Qualitative comparison of viewpoint-based scene rendering with varying number of frequencies.



(a) $N = 6$ (b) $N = 9$ (c) $N = 12$

Figure 3. Qualitative comparison on surface reconstruction with a different number of frequencies.

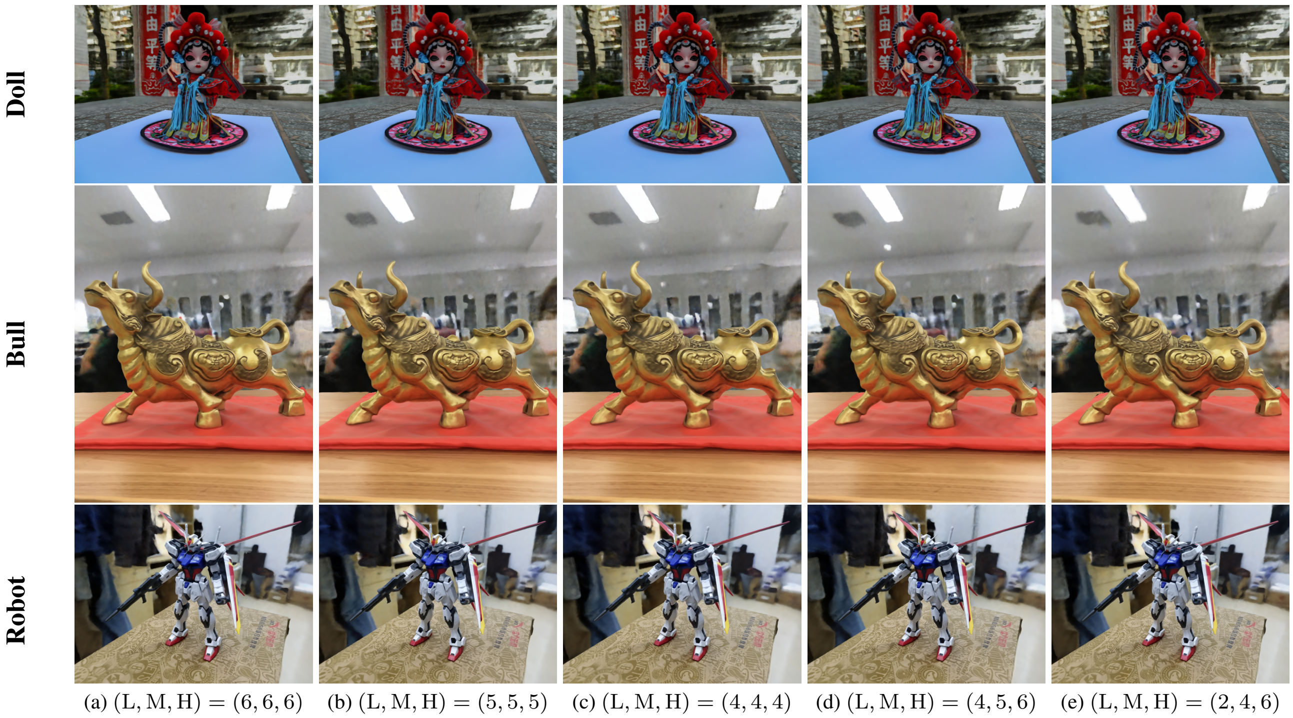


Figure 4. Qualitative comparison on viewpoint-based scene rendering using FreBIS, obtained by varying the number of encoder layers.

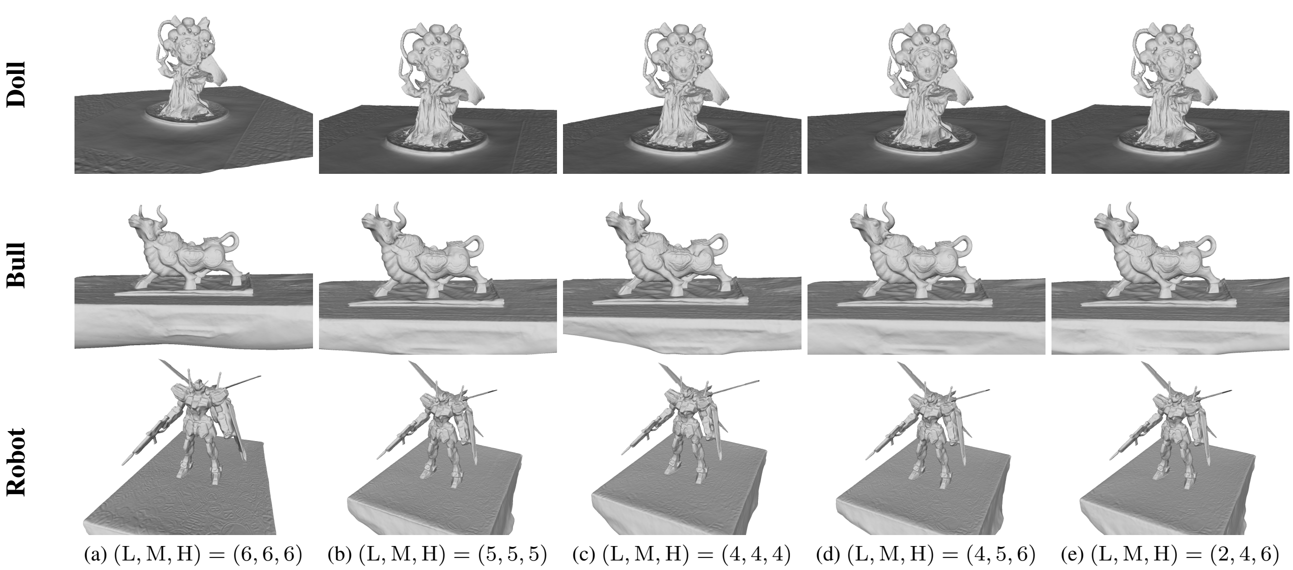


Figure 5. Qualitative comparison based on 3D surface reconstruction using FreBIS, obtained by varying the number of encoder layers.