Sphere-Guided Training of Neural Implicit Surfaces
Supplementary Material

Andreea Dogaru[1,2] Andrei Timotei Ardelean[1,2] Savva Ignatyev[1]
Egor Zakharov[1] Evgeny Burnaev[1,3]

[1]Skolkovo Institute of Science and Technology [2]Friedrich-Alexander-Universität Erlangen-Nürnberg
[3]Artificial Intelligence Research Institute

1. Implementation details

Sphere resampling. We employ sphere resampling to ensure all the primitives model the area of interest. To identify the spheres that fail to reach the surface due to local minimum regions of the implicit geometry field, we uniformly sample $K = 1000$ points in each sphere and evaluate the implicit function. Then, we mark the spheres that contain points either with all values lower, or all values higher than the surface level. Each marked sphere is replaced as follows: we randomly pick a sphere that is not marked and center a Gaussian distribution with a standard deviation of $\sigma = 2 * r_{min}$ around its origin, we sample the new origin of the sphere from this distribution; finally, we reinitialize the optimizer state for the considered primitive. A similar resampling strategy is applied periodically after every 5000 iterations for the spheres that are pushed by the repulsion outside of the scene bounds.

We found the resampling of empty spheres to bring the most benefits to UNISURF [5] model which uses an occupancy field for the implicit surface. This representation is more susceptible to local minimums than the SDF which has an additional Eikonal regularization [2]. Also, the number of spheres that require resampling for NeuralWarp [1] / VolSDF [7] is around 10 times higher in the initial steps than for NeuS [6]. We believe this is because the former imposes the Eikonal penalty only on two points on each ray, while the latter constrains the gradients of all points sampled.

Sphere-guided ray sampling. Compared to base methods, the sphere-guided models have the relevant areas of the volume explicitly defined. We can exploit the sphere bounds into sampling more informative rays during training as follows: we randomly sample a point inside each sphere and project the points to the training views. We exclude the points that project outside of the image bounds and randomly sample from the remaining points a batch of corresponding camera rays which are used for a training step.

Sphere-guided ray marching. The proposed approach can be applied to the considered 3D reconstruction methods without altering their formulation and training process. However, the introduced sphere-guided volume rendering enhances the point sampling along the ray procedures of the base methods as described in **Algorithm 1**, which imposes the following modifications:

- UNISURF [5] - the root-finding procedure is adjusted by only searching for the surface within the volume covered by the spheres. More precisely, we sample points uniformly inside the intervals (as found at step 5 of the algorithm) to find the first sign change. We then apply the secant method on this segment as in the original method. Reducing the search to the area around the surface enables the algorithm to better estimate the ray-surface intersection. The rest of the sampling procedure follows the original model.
- NeuS [6] - the points sampled within the ray-sphere intersections are used to compute a coarse probability estimation along the ray. If the ray intersects multiple surfaces, the set computed at step 5 of the algorithm can have more than one interval. As the region between two such intervals is outside the sphere cloud, we do not want to include it in the importance sampling. Therefore, we set the probabilities of these regions to zero, ensuring that the added points through importance sampling will belong to the set of intervals. Similarly, we set the weights of the midpoints used in color computation that fall outside the sphere bounds to zero. In the experiments that do not have segmentation masks as input, NeuS samples a set of points outside the bounds of the scene for background modeling; we do not interfere with these samples.
- VolSDF [7] / NeuralWarp [1] - We perform similar modifications as for the NeuS model. We set the uncertainty estimation of the ray segments between intervals to zero, so that the points added during the upsampling

stage are contained within the intersections of the ray with the sphere cloud. Additionally, we consider the estimated opacity of the previously mentioned segments as zero before performing inverse transform sampling to ensure that the final set of points lies within the intervals computed at step 5 of the algorithm. We do not interfere with the points sampled in the base method for background modeling outside the scene bounds.

2. Additional results

Further ablation study. We perform additional experiments to highlight the benefits of the proposed sphere-guided training. We compare the baseline NeuS model and our improved model in three different setups by varying the number of points sampled per ray: 128 points (default setting), 64 points (32 linearly spaced + 32 importance sampled), and 32 points (16 linearly spaced + 16 importance sampled). The results indicate that our approach can better handle the diminished number of input points, as can be seen in Figure 2.

We further evaluate independently the effect of the proposed ray-sampling and ray-marching procedures based on the optimized sphere cloud. We consider both the base setting with 128 points per ray and also using 64 points per ray. We show in Table 2 that both components contribute to our final results. Sphere-based ray sampling has a larger and more consistent effect by itself in the default setting, whereas the ray-marching procedure becomes more important when the number of samples is low. This is because it is harder to approximate the integral with fewer points per ray, and an improved sampling strategy has more obvious benefits. The effect is especially visible in scenes with thin structures, such as Ficus (Figure 2) and Ship. We note that the results on Drums are not always representative because the scene contains semi-transparent surfaces, which are, by design, not handled well by the base methods.

We also include a visual comparison between our complete model and three ablations in Figure 3. For this comparison, we use NeuS as the base system and train all versions without mask supervision.

DTU dataset experiments. We show additional quantitative (Figure 1) and qualitative (Figures 5-8) evaluation of our method on the DTU [3] dataset. In Figure 1, we investigate more in-depth the comparative performance between our method and NeuralWarp. To this end, we randomly picked five scenes and trained four models with different random seeds. We add these data points to the ones reported in the main paper and present them in this figure. We conclude that, on average, our method performs better than raw NeuralWarp, when the stochasticity of the base method is accounted for. Our method on average achieves a mean Chamfer distance of 0.73 against the NeuralWarp's 0.76.

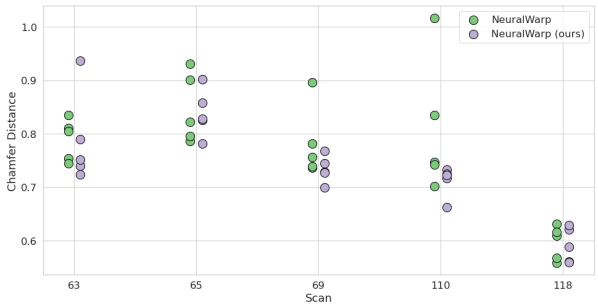

Figure 1. Quantitative results for the subset of five scenes of the DTU [3] dataset. We separately train each method for each scene five times with different starting random seeds. The results for our approach are reported in green, while for the base method they are in purple. The x-axis represents the scene ID, and the y-axis shows the obtained Chamfer distances. This comparison utilizes a masked Chamfer distance, following the same setting as in Table 1 in the experiments section. Our approach achieves a noticeably reduced variance of the results compared to the base system, and outperforms it on average, obtaining mean Chamfer distance of 0.73, averaged across five seeds and five scans, against the NeuralWarp's 0.76. We also have better worst- and best-case performance, with our method having mean best-case metrics of 0.69 against NeuralWarp's 0.71 and mean worst-case of 0.79 against 0.86.

We also show additional qualitative results for eight more scenes of DTU dataset in Figures 5-8.

Realistic Synthetic 360 dataset [4] experiments. This dataset has more complex geometries (with multiple objects per scene and fine details) compared with the DTU dataset, making it more challenging for the networks to accurately reconstruct them. We found that the default Eikonal regularization weight of 0.1 discourages the NeuS-based models without mask supervision from reconstructing disconnected objects (such as the chair in the drums scene). Therefore, the weight was divided by 10 for all experiments on this dataset, including the results in the main paper and the ablations. Before evaluating the Chamfer distance of the reconstructed meshes we filter the geometries using the ground truth segmentation masks dilated with a radius of 12 as in the DTU evaluation.

Next, we include qualitative results for additional scenes of the Realistic Synthetic 360 dataset [4] in Figures 9-11, as well as renders in Figures 12-13. We observe that our method not only achieves superior quality of reconstruction but also renders, which is confirmed by the image metrics presented in Table 1. We report the PSNR, SSIM, and LPIPS [8] metrics evaluated on the 200 test views provided in the dataset (not seen during training).

Additionally, in Figure 4 we present the learning curves for both our and the NeuS base model when trained for more iterations on one of the scenes. This experiment confirms

PSNR ↑									
Method	Chair	Drums	Ficus	Hotdog	Lego	Mats	Mic	Ship	Mean
NeRF [4]	33.00	25.01	30.13	36.18	32.54	29.62	32.91	28.65	31.01
NeuS [6]	30.89	20.89	27.44	36.04	30.45	30.21	31.13	27.08	29.26
NeuS (ours)	**32.21**	**21.81**	**30.40**	**36.76**	**31.80**	**31.08**	**31.44**	**28.63**	**30.52**
NeuralWarp [1]	29.29	18.41	24.50	32.32	27.90	**27.45**	**29.15**	24.07	26.64
NeuralWarp (ours)	**30.17**	**18.82**	**26.79**	**32.73**	**28.81**	24.76	28.67	**25.22**	**27.00**

SSIM ↑									
Method	Chair	Drums	Ficus	Hotdog	Lego	Mats	Mic	Ship	Mean
NeRF [4]	0.967	0.925	0.964	0.974	0.961	0.949	0.980	0.856	0.947
NeuS [6]	0.948	0.897	0.957	0.973	0.952	0.960	0.971	0.864	0.940
NeuS (ours)	**0.960**	**0.909**	**0.973**	**0.978**	**0.962**	**0.965**	**0.974**	**0.883**	**0.951**
NeuralWarp [1]	0.936	0.872	0.933	0.960	0.928	0.938	**0.962**	0.817	0.918
NeuralWarp (ours)	**0.947**	**0.878**	**0.950**	**0.964**	**0.939**	**0.941**	0.960	**0.828**	**0.926**

LPIPS ↓									
Method	Chair	Drums	Ficus	Hotdog	Lego	Mats	Mic	Ship	Mean
NeRF [4]	0.046	0.091	0.044	0.121	0.050	0.063	0.028	0.206	0.081
NeuS [6]	0.064	0.118	0.047	0.045	0.060	0.053	0.030	0.184	0.075
NeuS (ours)	**0.050**	**0.103**	**0.029**	**0.037**	**0.047**	**0.048**	**0.026**	**0.165**	**0.063**
NeuralWarp [1]	0.076	0.161	0.072	0.068	0.091	**0.074**	**0.043**	0.243	0.104
NeuralWarp (ours)	**0.065**	**0.142**	**0.053**	**0.061**	**0.076**	0.076	0.047	**0.228**	**0.094**

Table 1. Quantitative image results on the Realistic Synthetic 360 dataset [4]. We evaluate PSNR, SSIM (higher is better), and LPIPS [8] (lower is better). The proposed approach improves the rendering quality of the NeuS and NeuralWarp models.

	Scene name							
Method	Chair	Drums	Ficus	Hotdog	Lego	Mats	Mic	Ship
NeuS	0.38	1.88	0.51	**0.52**	0.68	0.40	**0.60**	0.60
w/o **S**-guided ray sampling	**0.36**	1.79	0.49	0.55	0.69	0.43	0.72	0.79
w/o **S**-guided ray marching	0.38	**0.81**	**0.40**	0.54	**0.61**	**0.29**	0.69	0.67
NeuS (ours)	0.39	1.20	**0.40**	0.57	**0.61**	0.31	0.67	**0.54**
NeuS 64	**0.35**	2.28	0.61	**0.52**	0.73	0.35	0.57	0.84
w/o S-guided ray marching 64	0.38	1.64	0.52	0.55	0.65	**0.28**	0.58	0.75
NeuS (ours) 64	0.40	**0.79**	**0.45**	0.59	**0.60**	**0.28**	**0.56**	**0.69**

Table 2. Ablation study on the importance of sphere-based ray sampling and marching. We evaluate the base setting with 128 points per ray and also using 64 points per ray (as fewer points emphasize the effect of improved ray-marching).

that our approach indeed achieves a better local optima, and not just increases the convergence speed.

Ellipse-guided training. We additionally attempted to further increase the sampling efficiency by using a cloud of ellipses instead of spheres to encapsulate the learned surface. We set the ellipses to have their two major axes equal to the scheduled radius and be aligned with the tangent plane, while the remaining minor axis to be aligned with the surface normal. This would allow us to increase the number of samples along the ray with nonzero opacity even further, compared to the sphere-based sampling. However, after running the preliminary experiments with different scaling factors which define the length of the minor axis, we observed no further improvements compared to the sphere-based training. We argue it to be caused by the high efficiency of the combination of our proposed sphere-based training and importance sampling, which is part of the base method.

References

[1] Francois Darmon, Bénédicte Bascle, Jean-Clément Devaux, Pascal Monasse, and Mathieu Aubry. Improving neural implicit surfaces geometry with patch warping. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6260–6269, 2022. 1, 3, 9, 12, 14

[2] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron

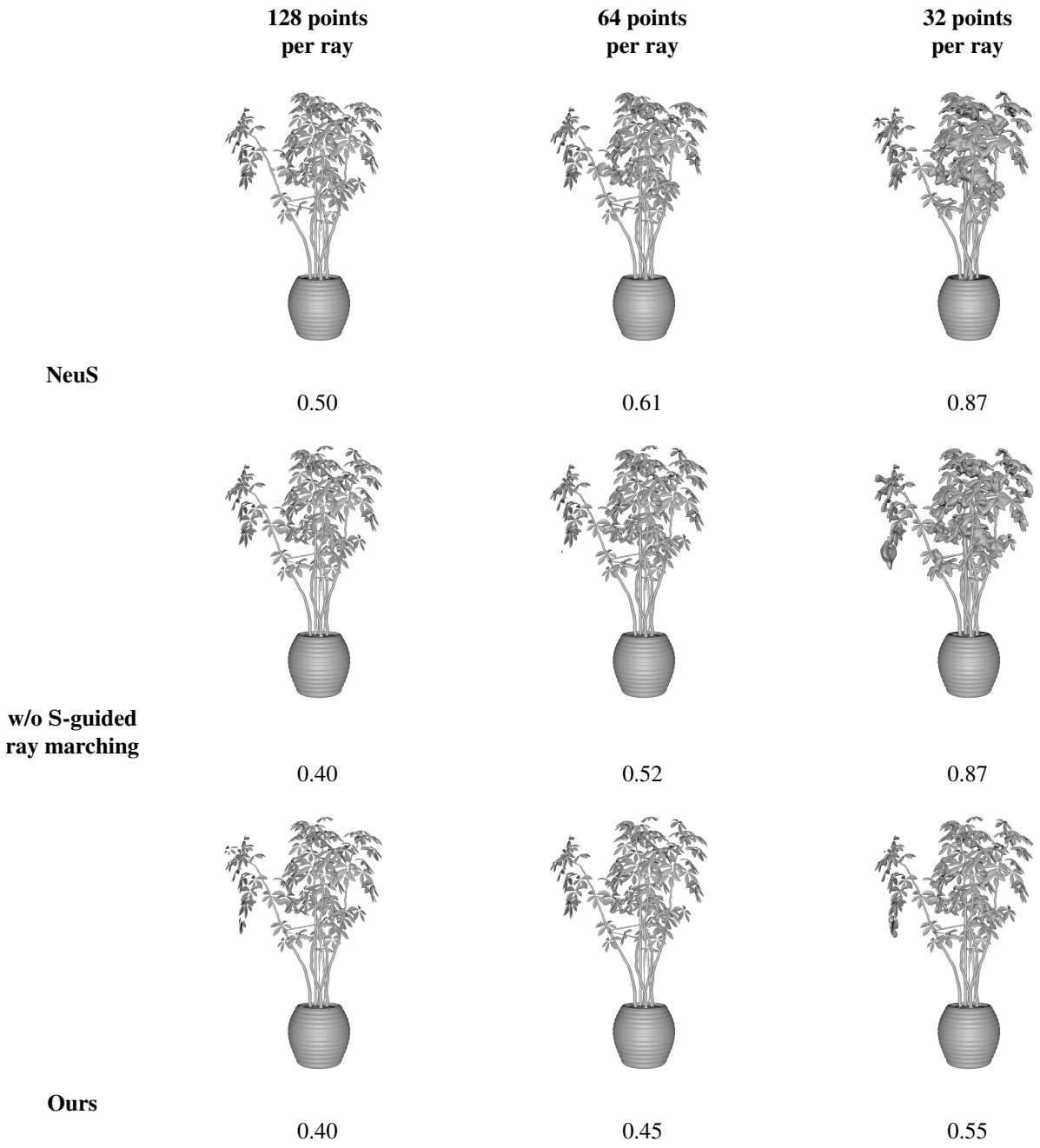

Figure 2. Performance comparison when training with a reduced number of points per ray. For each reconstruction, Chamfer distances are reported under the rendered meshes. Our model achieves larger improvements when the number of points per ray is lower.

Lipman. Implicit geometric regularization for learning shapes. In *ICML*, 2020. 1

[3] Rasmus Ramsbol Jensen, A. Dahl, George Vogiatzis, Engil Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 406–413, 2014. 2, 6, 7, 8, 9

[4] 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. In *European Conference on Computer Vision (ECCV)*, 2020. 2, 3, 10, 11, 12, 13, 14

[5] Michael Oechsle, Songyou Peng, and Andreas Geiger. Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5569–5579, 2021. 1, 8

[6] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku

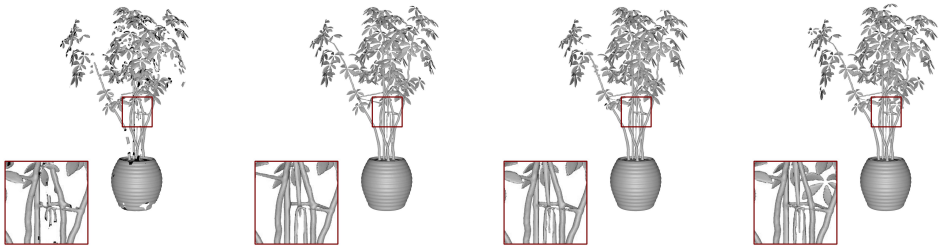

Figure 3. Ablation study. We have used the NeuS base system, trained without mask supervision, as a base model. We show both the resulting qualitative results, as well as obtained Chamfer distances for the shown scene. As reference, the baseline NeuS model obtains a Chamfer distance of 0.51.

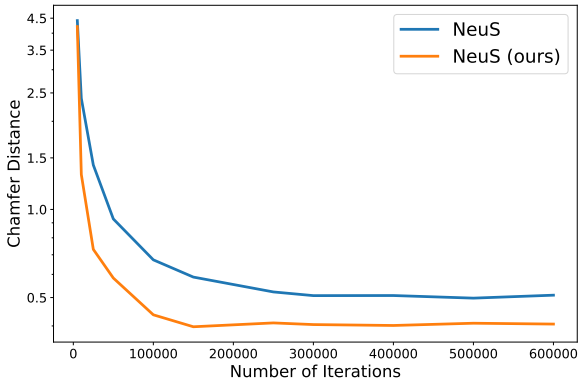

Figure 4. The learning curves for the ficus scene of the Realistic Synthetic 360 dataset showcase that our method converges to a better local optima than the base system, while using the same set of hyperparameters.

Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *Advances in Neural Information Processing Systems*, 34:27171–27183, 2021. 1, 3, 6, 7, 10, 11, 13

[7] 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. 1

[8] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 586–595, 2018. 2, 3

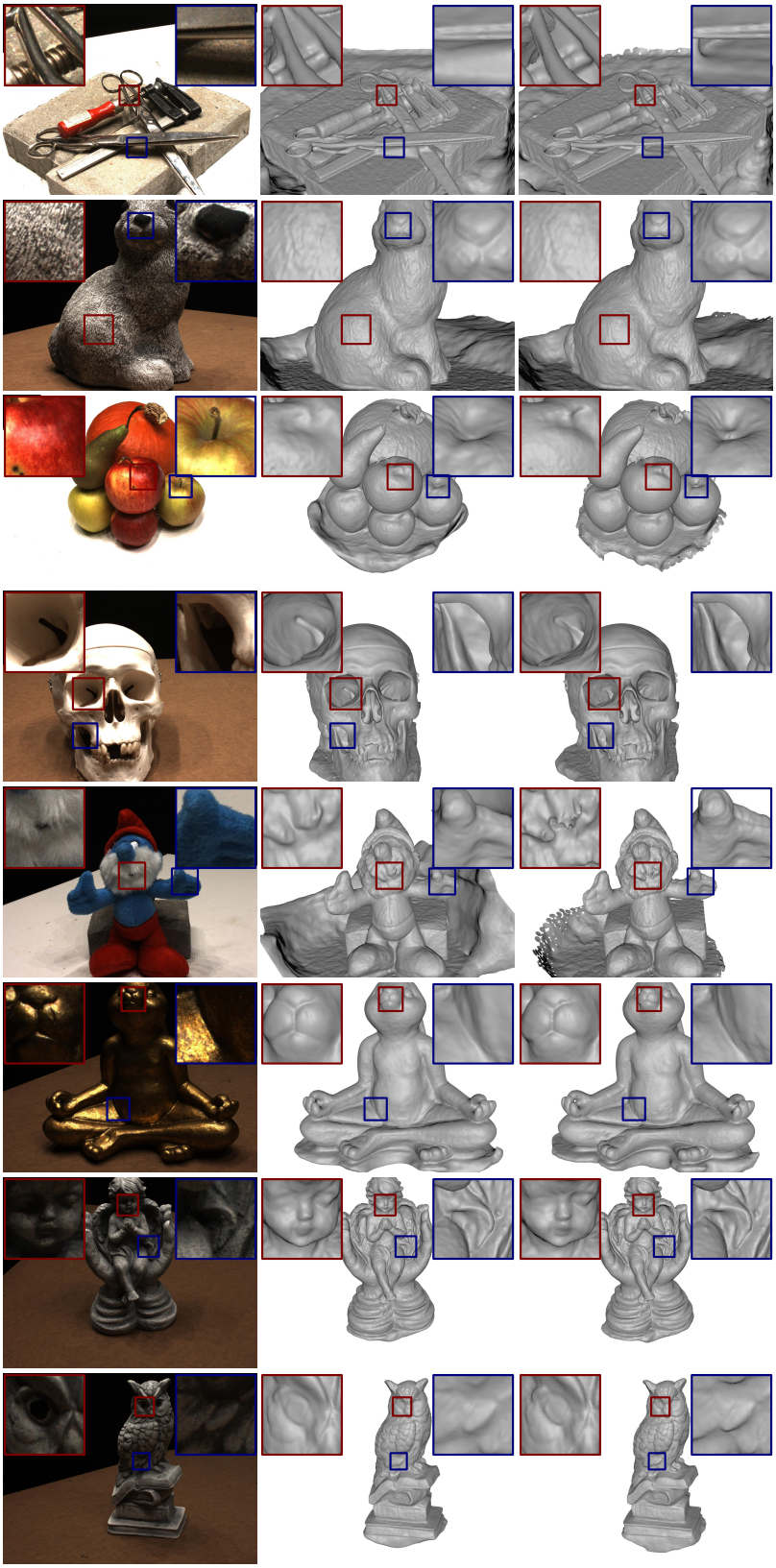

Figure 5. Additional qualitative results on the DTU [3] dataset for NeuS [6] method.

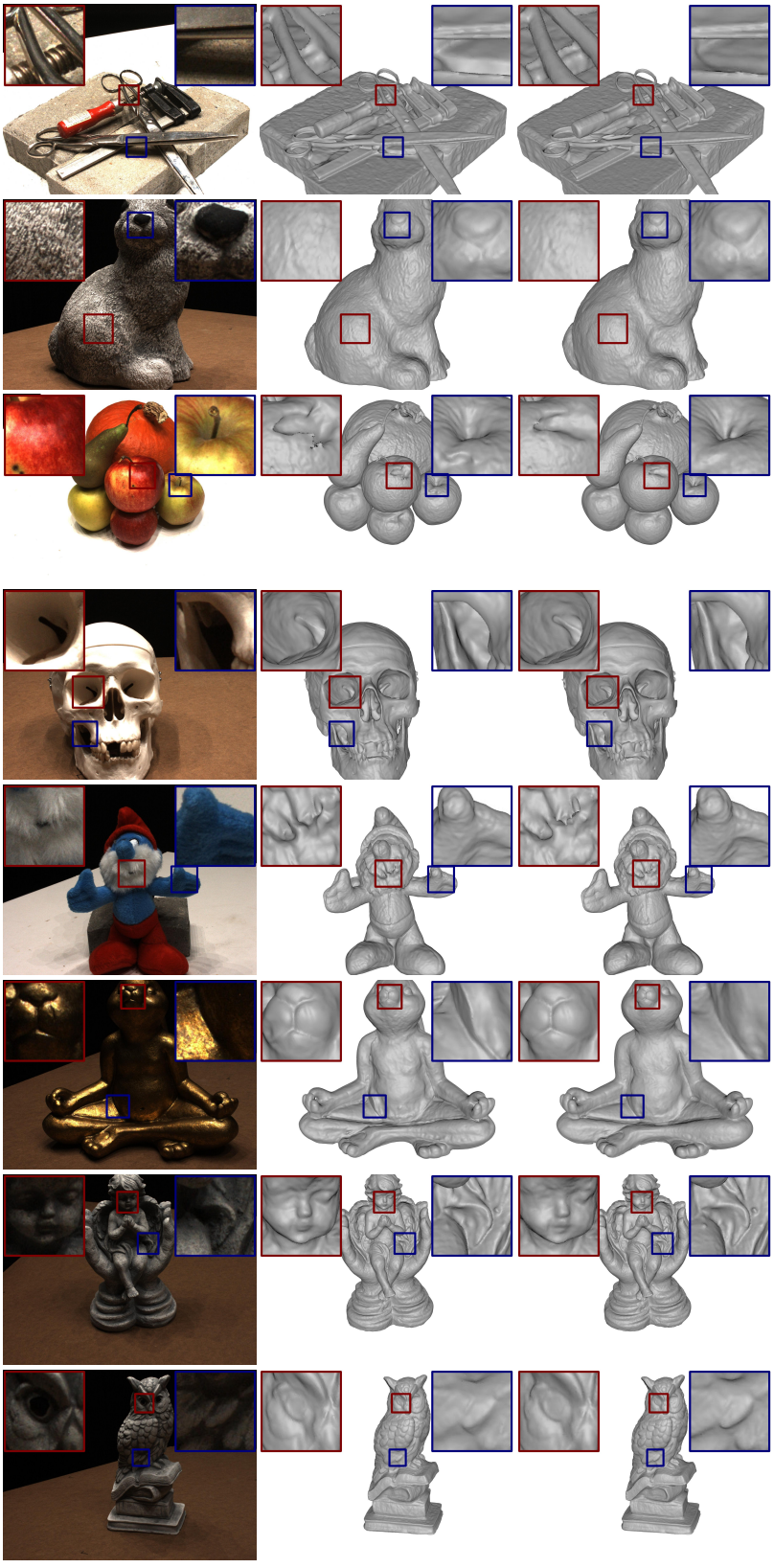

Figure 6. Additional qualitative results on the DTU [3] dataset for NeuS [6] method trained with masks supervision.

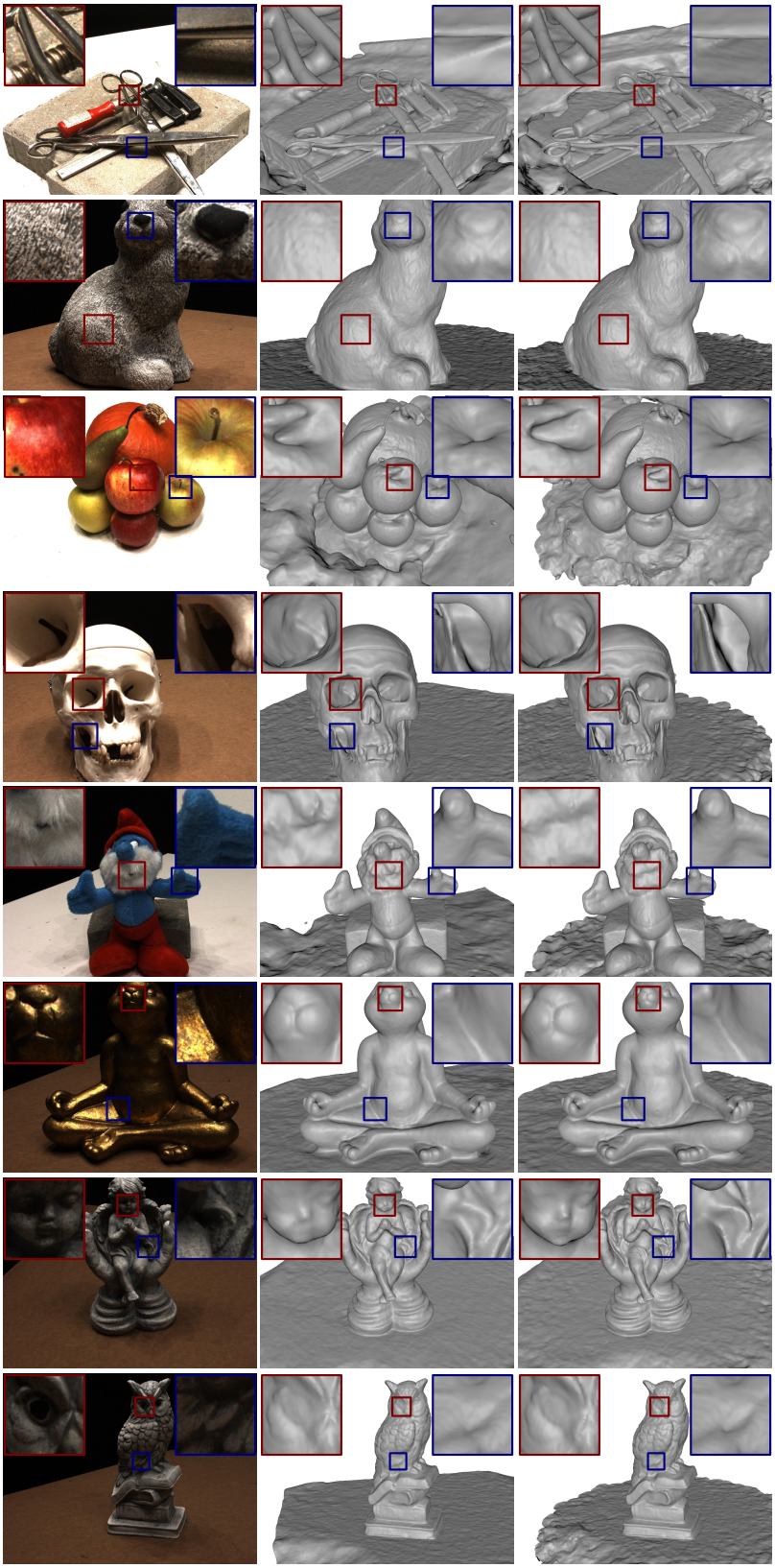

Figure 7. Additional qualitative results on the DTU [3] dataset for UNISURF [5] method.

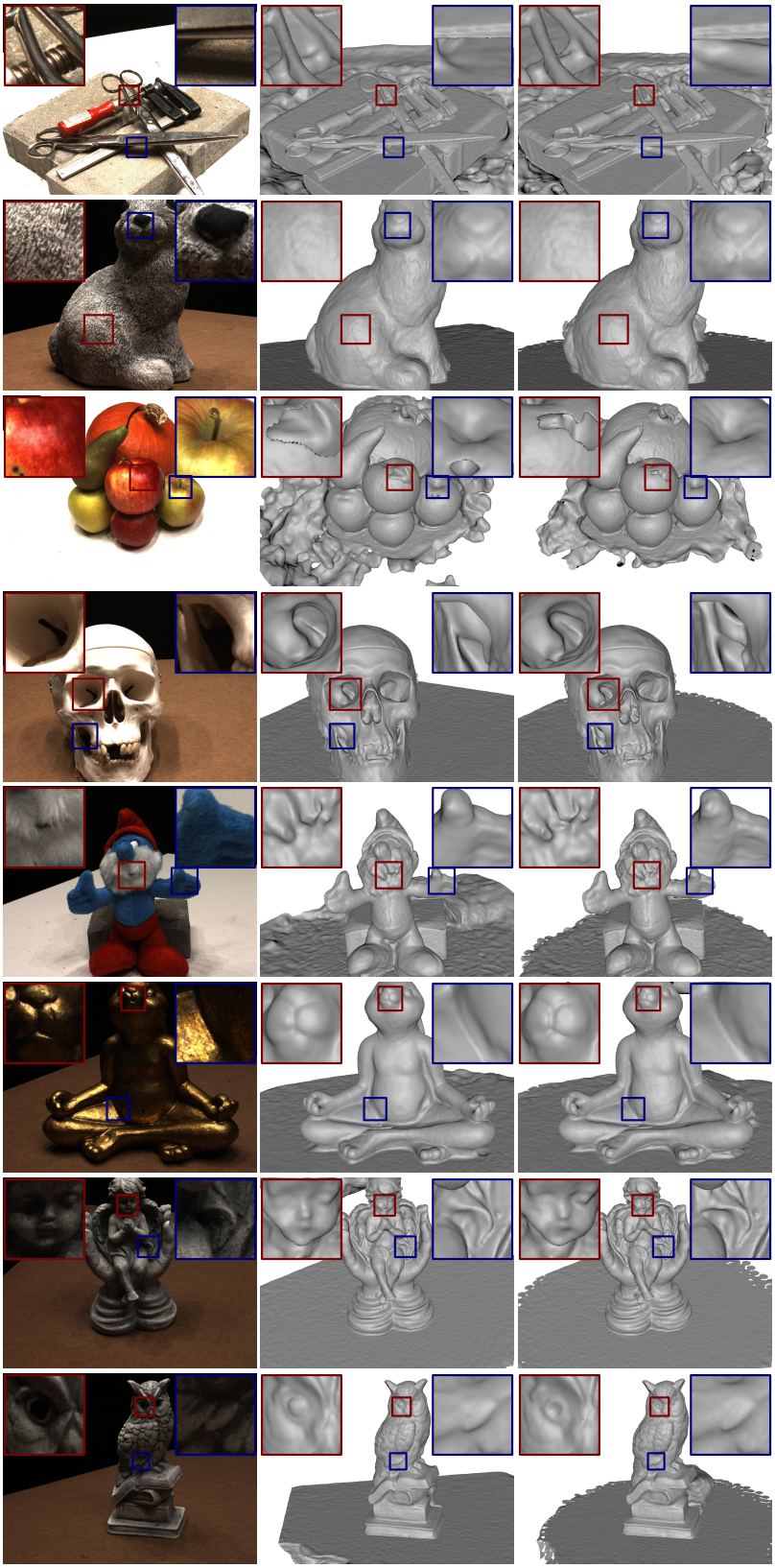

Figure 8. Additional qualitative results on the DTU [3] dataset for NeuralWarp [1] method.

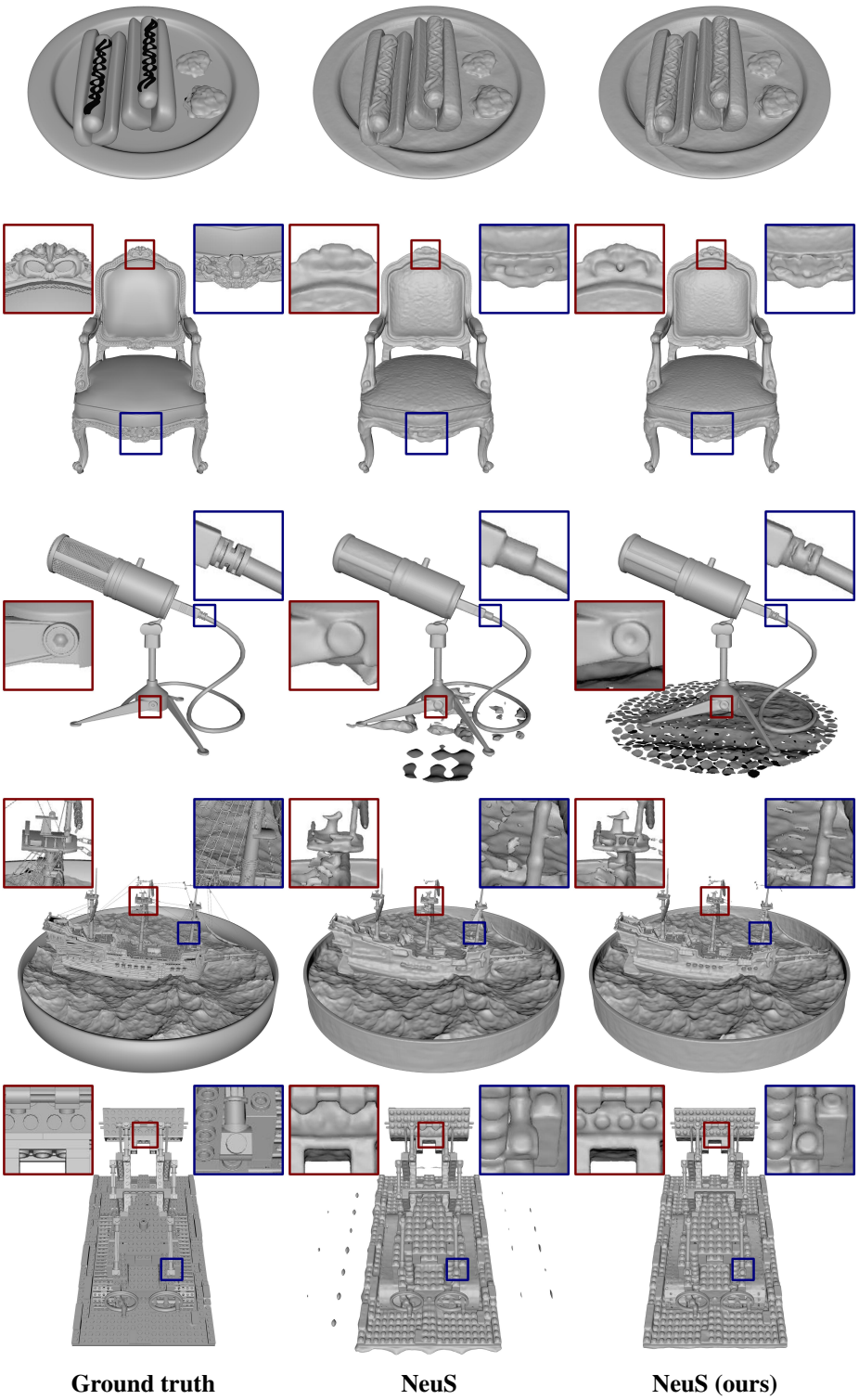

Figure 9. Additional qualitative results on the Realistic Synthetic 360 dataset [4] for NeuS [6] method.

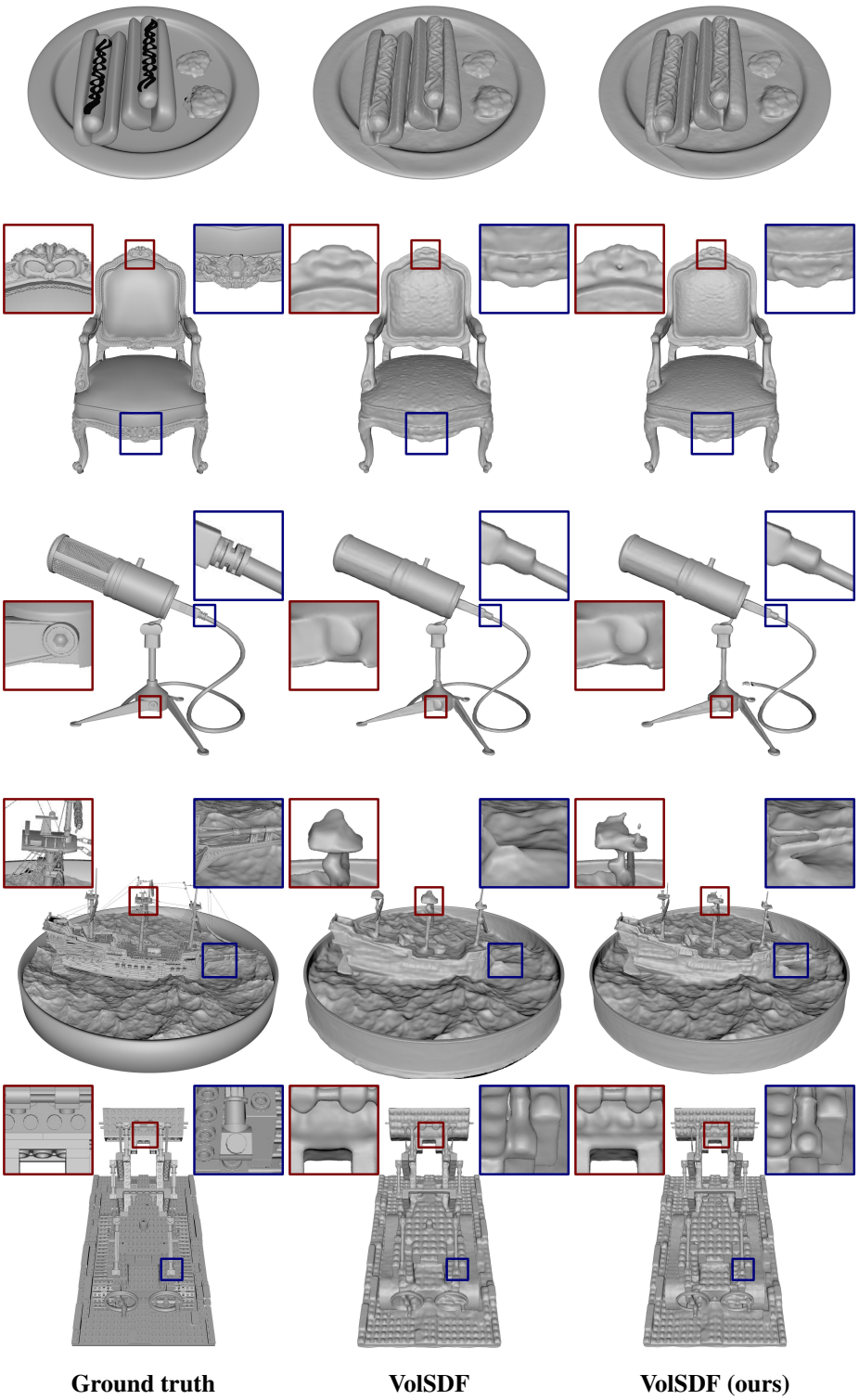

Figure 10. Qualitative results on the Realistic Synthetic 360 dataset [4] for the unofficial implementation of VolSDF [6] method.

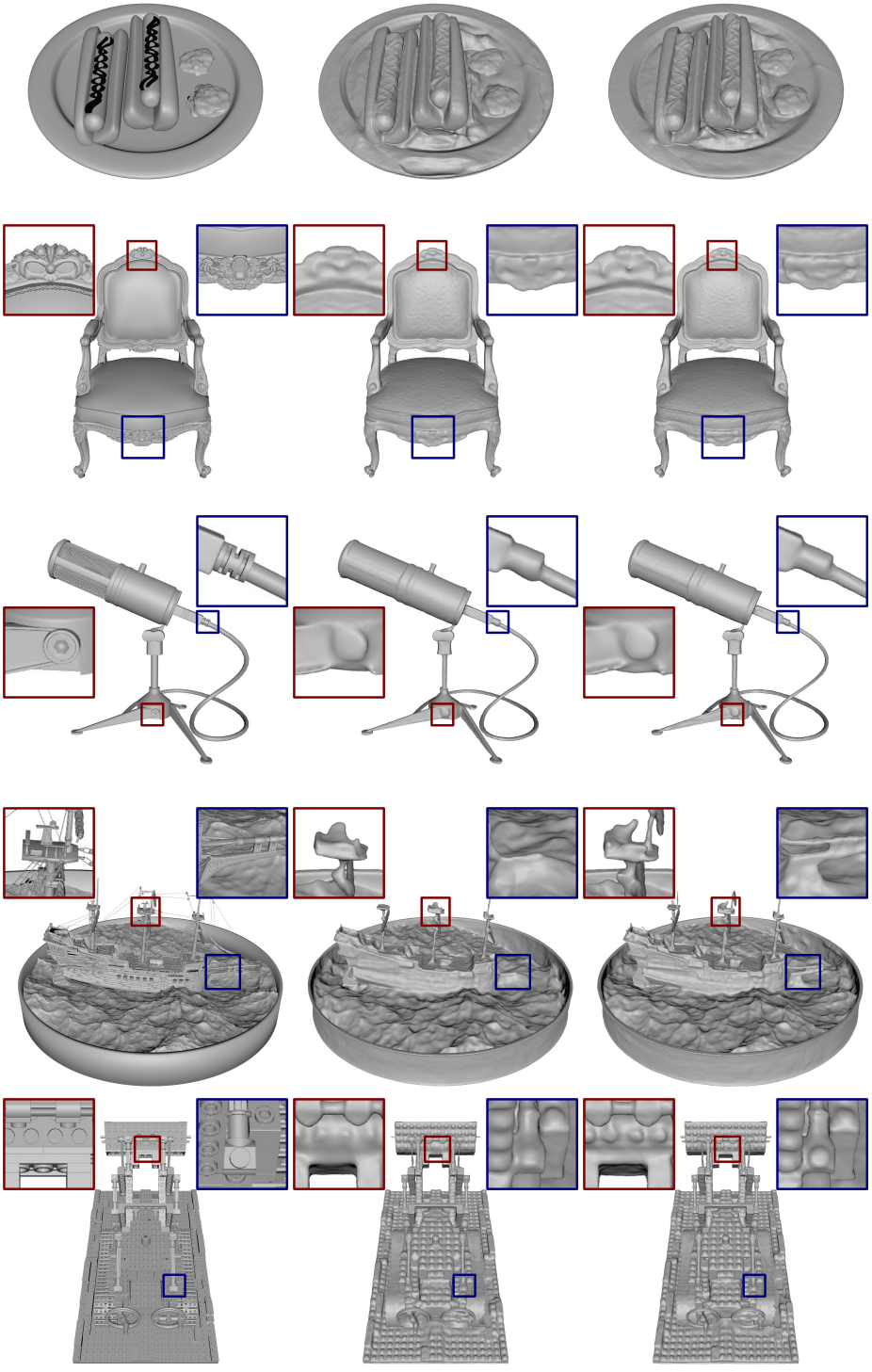

Figure 11. Additional qualitative results on the Realistic Synthetic 360 dataset [4] for NeuralWarp [1] method.

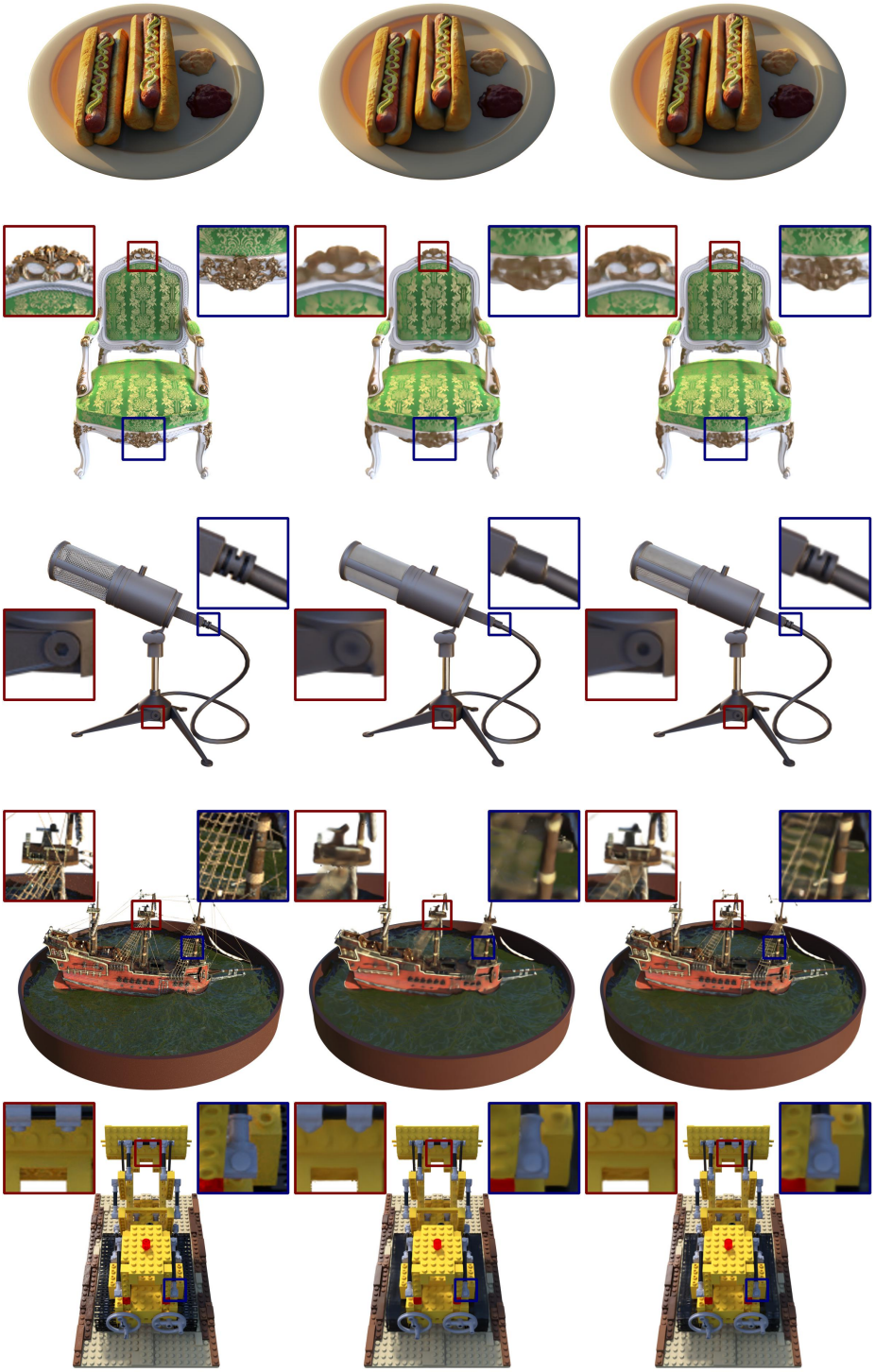

Figure 12. Rendering results on the Realistic Synthetic 360 dataset [4] for NeuS [6] method.

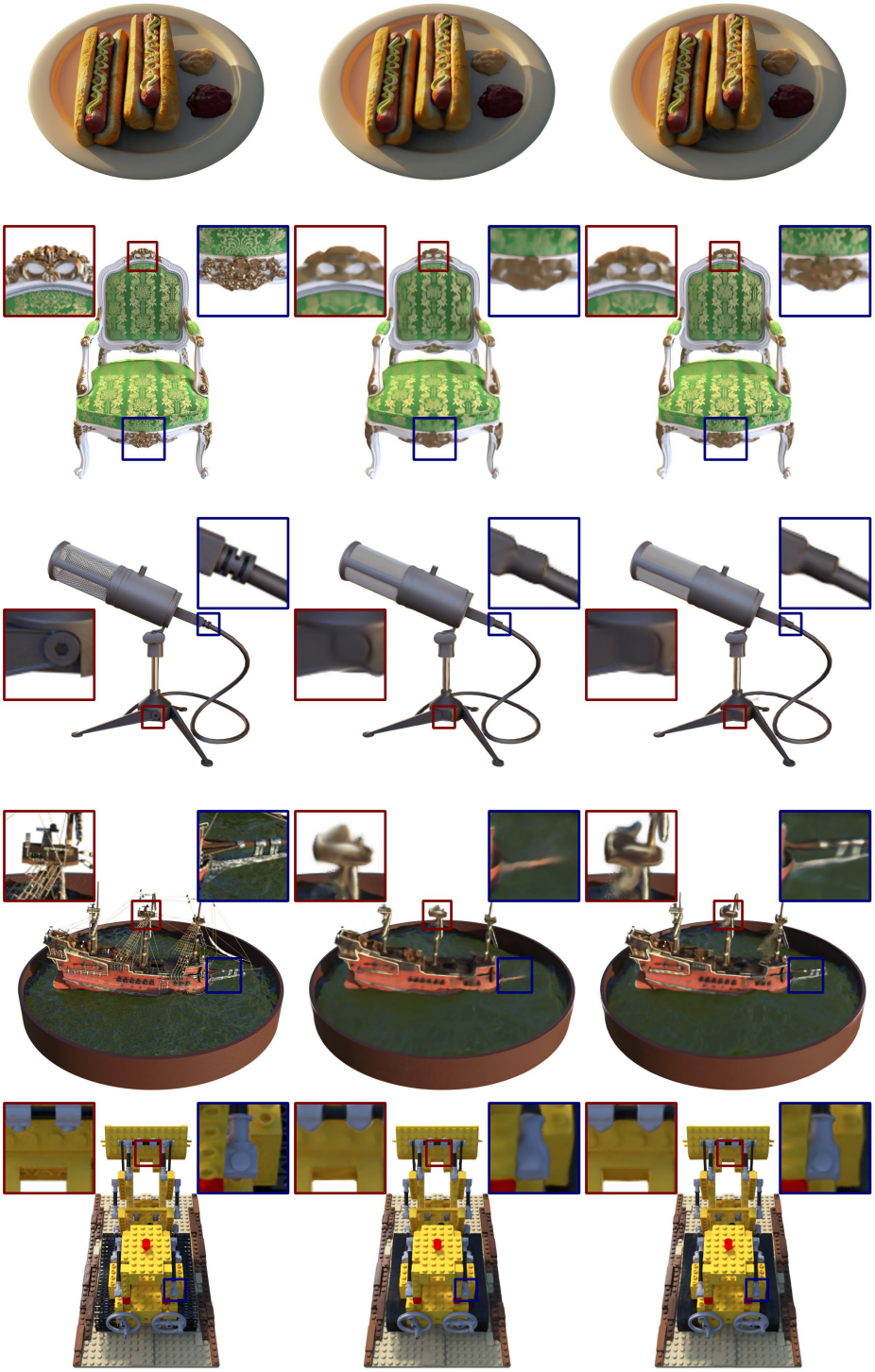

Figure 13. Rendering results on the Realistic Synthetic 360 dataset [4] for NeuralWarp [1] method.