Supplementary material for "MVCPS-NeuS: Multi-view Constrained Photometric Stereo for Neural Surface Reconstruction"

Hiroaki Santo Fumio Okura Yasuyuki Matsushita
Graduate School of Information Science and Technology, Osaka University
{santo.hiroaki, okura, yasumat}@ist.osaka-u.ac.jp

In this supplementary material,

A. we provide a comprehensive visualization of the estimated normal maps,
B. we show the comparison of the decomposition accuracy with and without HO-GSVD, and
C. we show the further comparison of DNN-based UPS [2] and MVCPS.

A. Visualization of Normal Maps

We provide a comprehensive visualization of the estimated normal maps from views used in training (referred to as "training views") and those not used in training ("test views"), as shown in Figs. S1 to S20. We present the results from all four training views and selected four test views. For the DiLiGenT-MV dataset, we include all five scenes. The results consistently agree with the results shown in the main paper and demonstrate the accuracy of the proposed method.

B. Decomposition Accuracy with and without HO-GSVD

To evaluate the decomposition accuracy with and without HO-GSVD, we use our synthetic scenes with the Lambertian surface, which are rendered under 24 viewpoints and 37 light directions. We compute the mean angular errors of the decomposed surface normals and light directions to measure the accuracy of the decomposition, which are disambiguated using the ground truth under the assumption of perfect disambiguation. We vary the number of light directions, using 3, 5, and 8, which are sampled from the 37 available light directions. We conduct 20 trials by randomly sampling light directions and present the mean of these trials. Table S1 shows the evaluation results. These results further demonstrate the advantage of our method with HO-GSVD, with which multi-view observations are jointly considered rather than independently.

C. Comparison of DNN-based UPS and MVCPS

We further investigate the estimation accuracy of DNN-based UPS [2] and the proposed method, MVCPS. We use the same synthetic dataset as in the main paper, increasing the number of light directions to 37. We conduct 20 trials, each randomly sampling 3 and 8 light directions, and present the mean of these trials. While the synthetic experiments in the main paper use diffuse images for MVCPS to align with the Lambertian assumption, we here use both diffuse and diffuse+specular (referred to as "specular") images as input for both methods. In this experiment, we assume that the ambiguity in MVCPS can be perfectly resolved through subsequent neural surface optimization, and we use the ground truth for disambiguation.

Table S2 shows the estimation accuracy of DNN-based UPS and MVCPS. In general, diffuse observations are suitable for photometric stereo; however, in the context of UPS, specular observations are useful for disambiguation. Indeed, the Lambertian observations significantly increase the estimation error due to the difficulties in disambiguation, as discussed in [1, 2]. The limited number of light sources also presents a challenge. Under these challenging settings, our MVCPS, which employs a factorization-based approach, demonstrates superior performance.

We would like to note that MVCPS inherently produces ambiguous estimations, and the presented mean angular errors are computed after disambiguation using the ground truth. However, as shown in the main paper, we can effectively resolve the ambiguity and achieve accurate shape estimation by incorporating these ambiguous surface normals into neural surface reconstruction.

References

[1] P.N. Belhumeur, D.J. Kriegman, and A.L. Yuille. The bas-relief ambiguity. In *Computer Vision and Pattern Recognition (CVPR)*, 1997. 1

Table S1. Comparison of the decomposition accuracy with and without HO-GSVD. We show the mean angular errors of the surface normal and light directions in degrees with disambiguation by the ground truth.

	3 lights				5 lights				8 lights			
	Ours		w/o HO-GSVD		Ours		w/o HO-GSVD		Ours		w/o HO-GSVD	
	[normal]	[light]	[normal]	[light]	[normal]	[light]	[normal]	[light]	[normal]	[light]	[normal]	[light]
BLOBBY	**9.32**	**8.19**	9.39	9.27	**7.42**	**7.11**	7.70	8.34	**5.97**	**6.22**	6.11	7.07
BUNNY	**7.91**	**5.37**	8.40	7.42	**9.59**	**5.37**	10.08	7.79	**9.45**	**5.23**	9.83	6.98

Table S2. Comparison of MVCPS (Ours) and DNN-based UPS ("DNN-UPS") [2]. We use two scenes, BLOBBY and BUNNY, with two different materials: diffuse and specular. We show the mean angular errors of the surface normal and light directions in degrees. MVCPS uses the ground truth for disambiguation.

	3 lights				8 lights			
	MVCPS (Ours)		DNN-UPS		MVCPS (Ours)		DNN-UPS	
	[normal]	[light]	[normal]	[light]	[normal]	[light]	[normal]	[light]
BLOBBY (diffuse)	**11.7**	**9.6**	30.2	11.9	**8.2**	**5.1**	20.9	8.6
BLOBBY (specular)	**13.2**	9.0	19.1	**8.0**	**10.1**	**3.7**	11.2	6.5
BUNNY (diffuse)	**14.5**	**10.8**	27.8	25.5	**11.6**	**5.6**	16.0	14.2
BUNNY (specular)	**16.0**	**10.0**	19.5	12.8	12.8	**5.7**	**12.0**	8.3

[2] Guanying Chen, Michael Waechter, Boxin Shi, Kwan-Yee K. Wong, and Yasuyuki Matsushita. What is learned in deep uncalibrated photometric stereo? In *European Conference on Computer Vision (ECCV)*, 2020. 1, 2

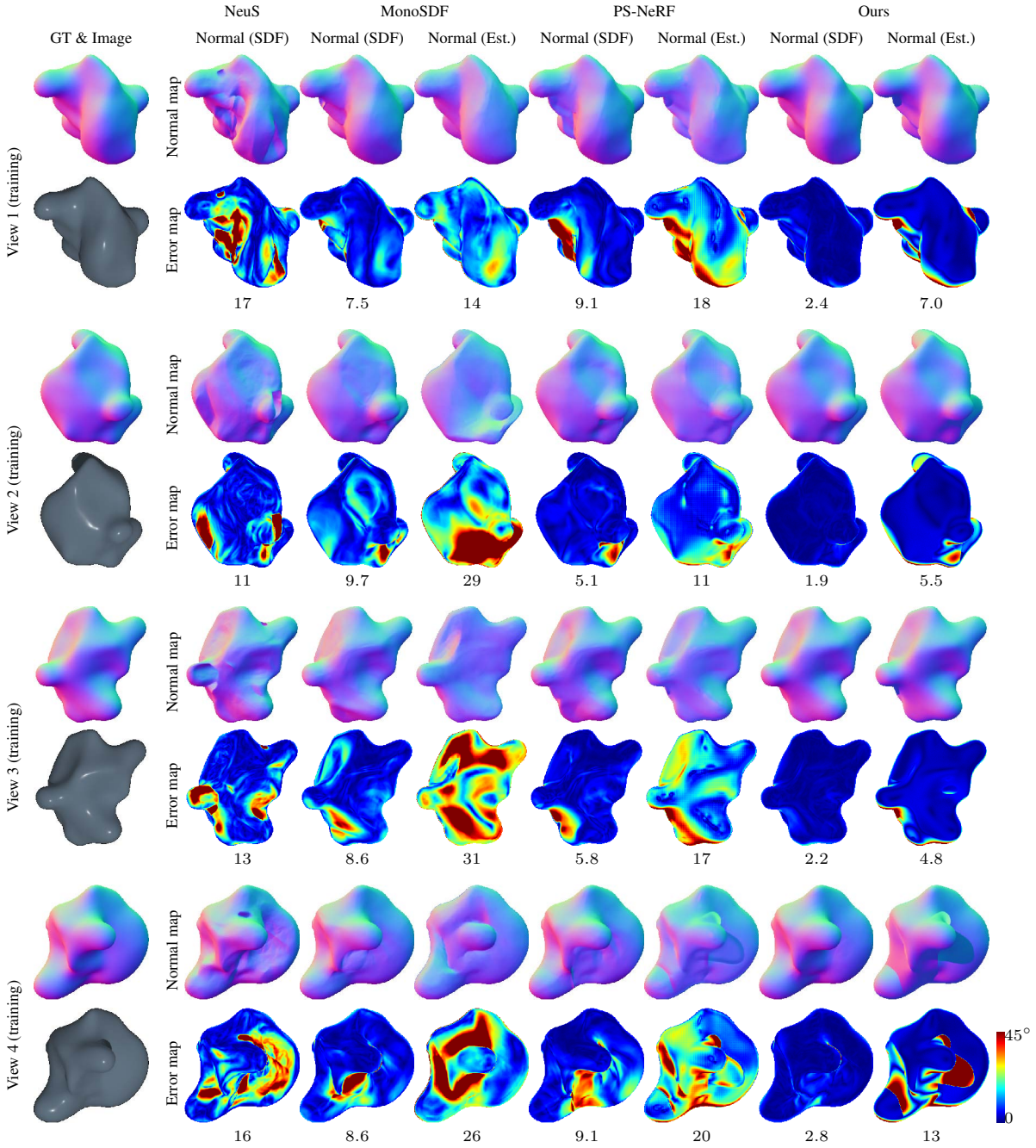

Figure S1. Estimated normal maps for the BLOBBY scene from training views. For each view and method, we present the rendered normal map of the SDF, the estimated normal map fed to the optimization, and corresponding error maps. The numbers under the error maps represent mean angular errors in degrees.

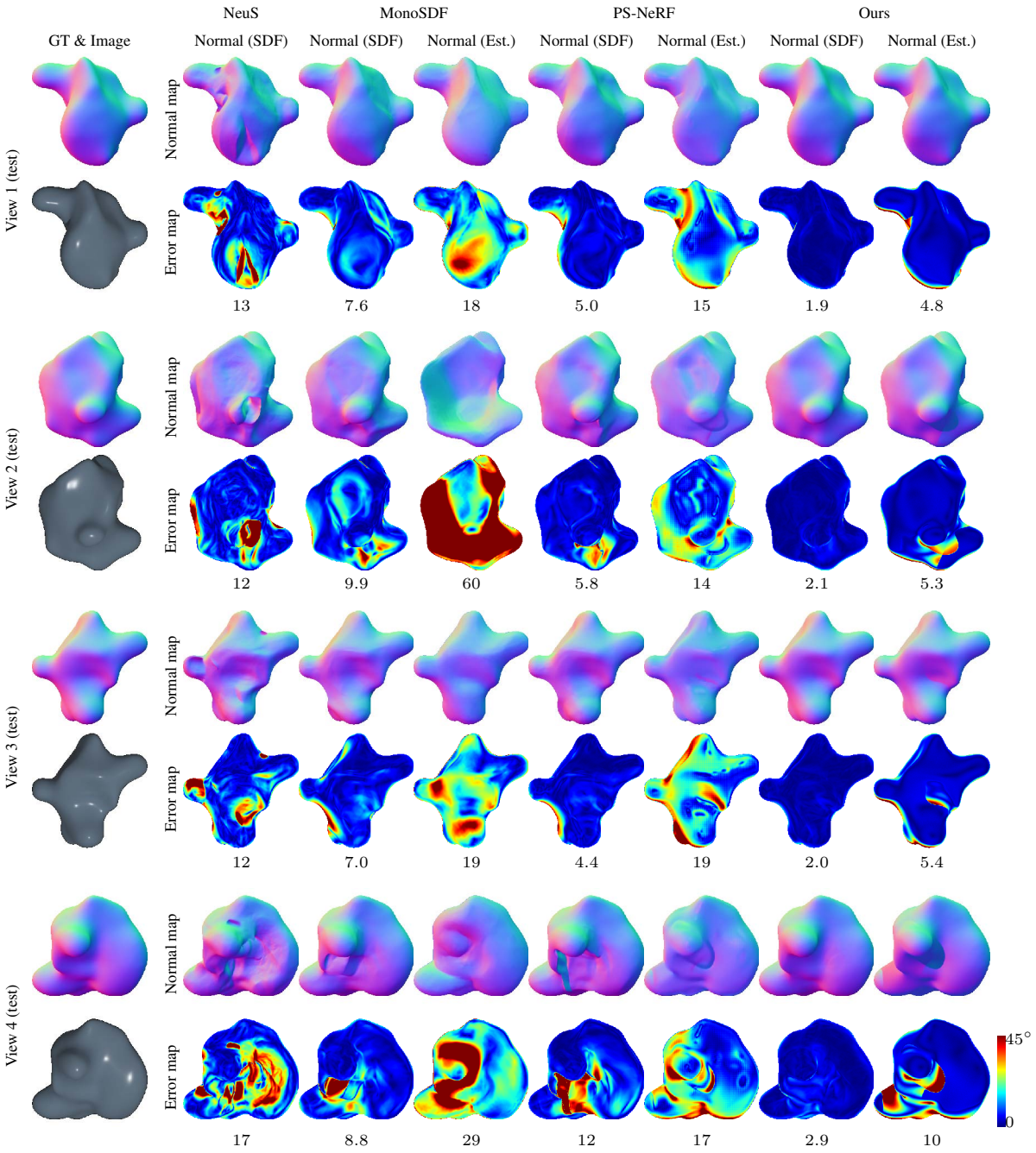

Figure S2. Estimated normal maps for the BLOBBY scene from test views.

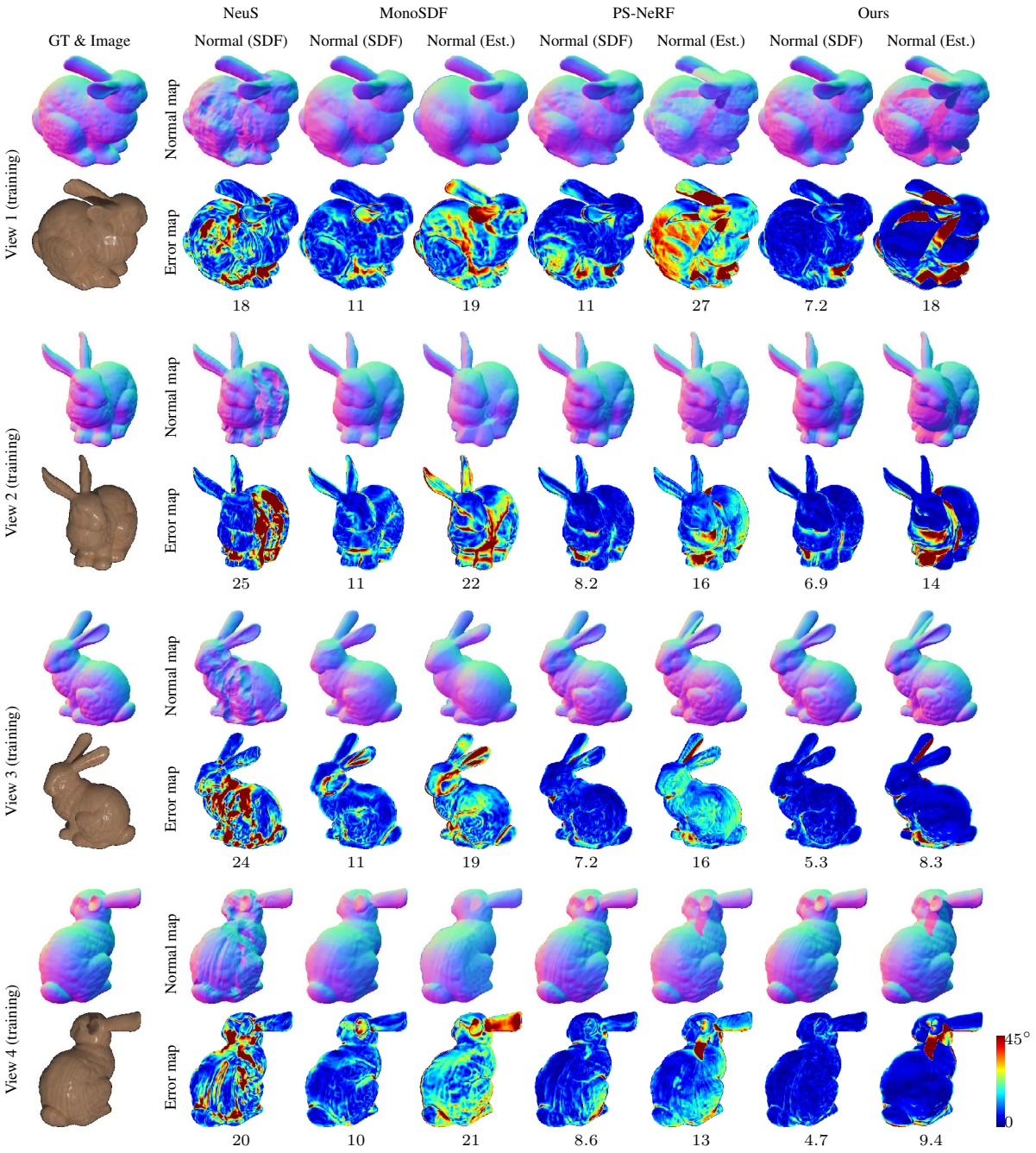

Figure S3. Estimated normal maps for the BUNNY scene from training views.

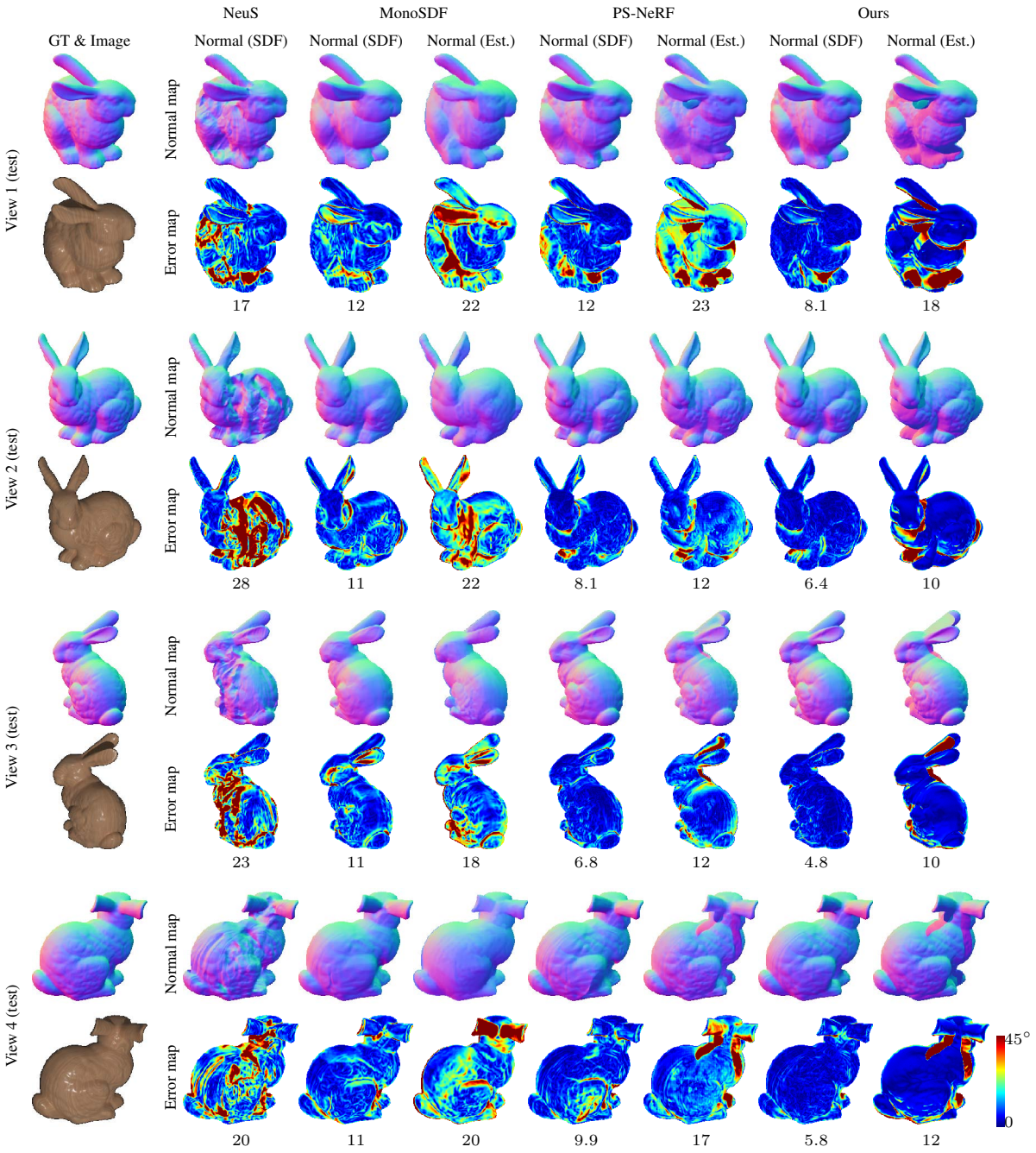

Figure S4. Estimated normal maps for the BUNNY scene from test views.

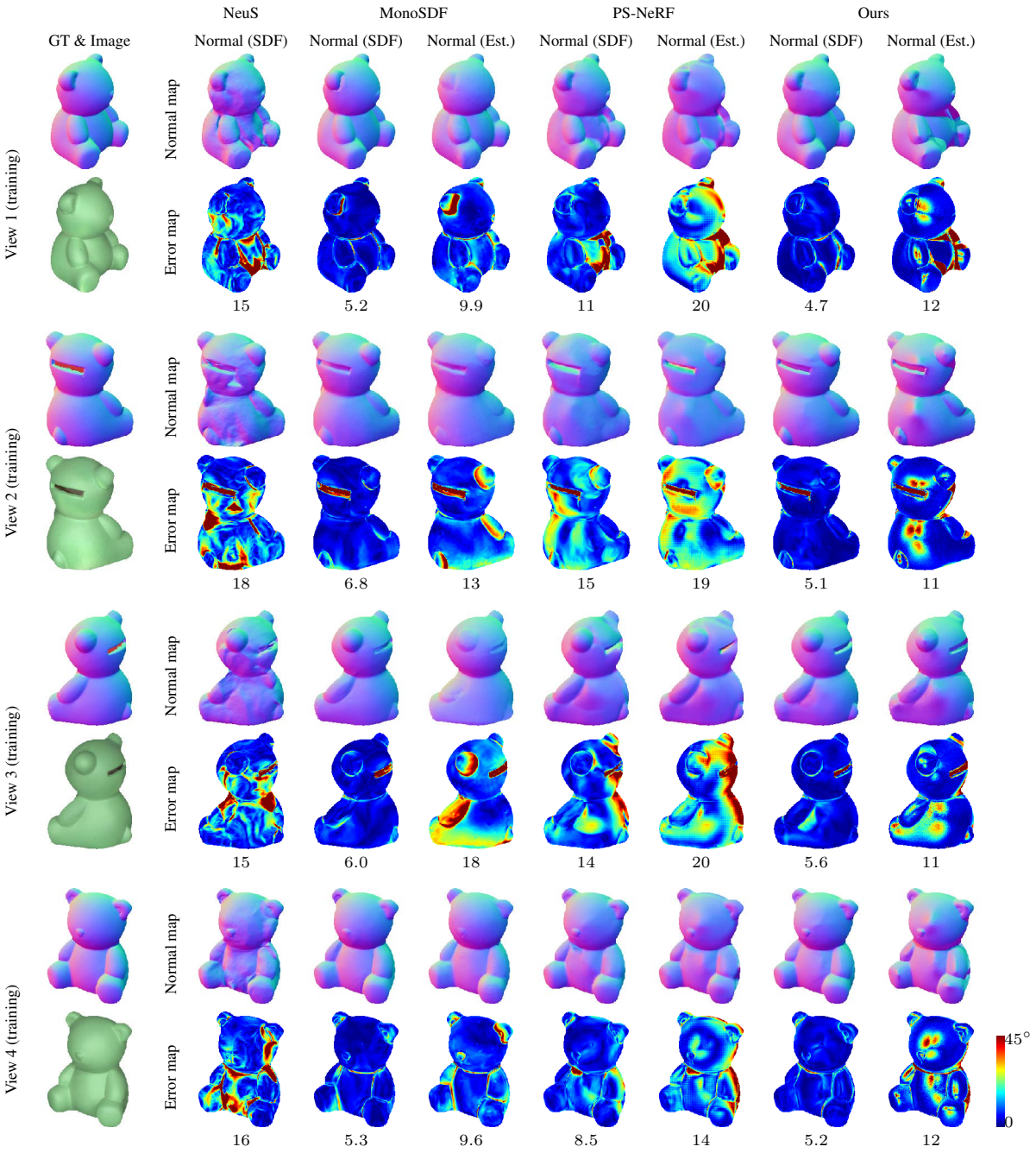

Figure S5. Evaluation of the normal maps for BEAR from training views.

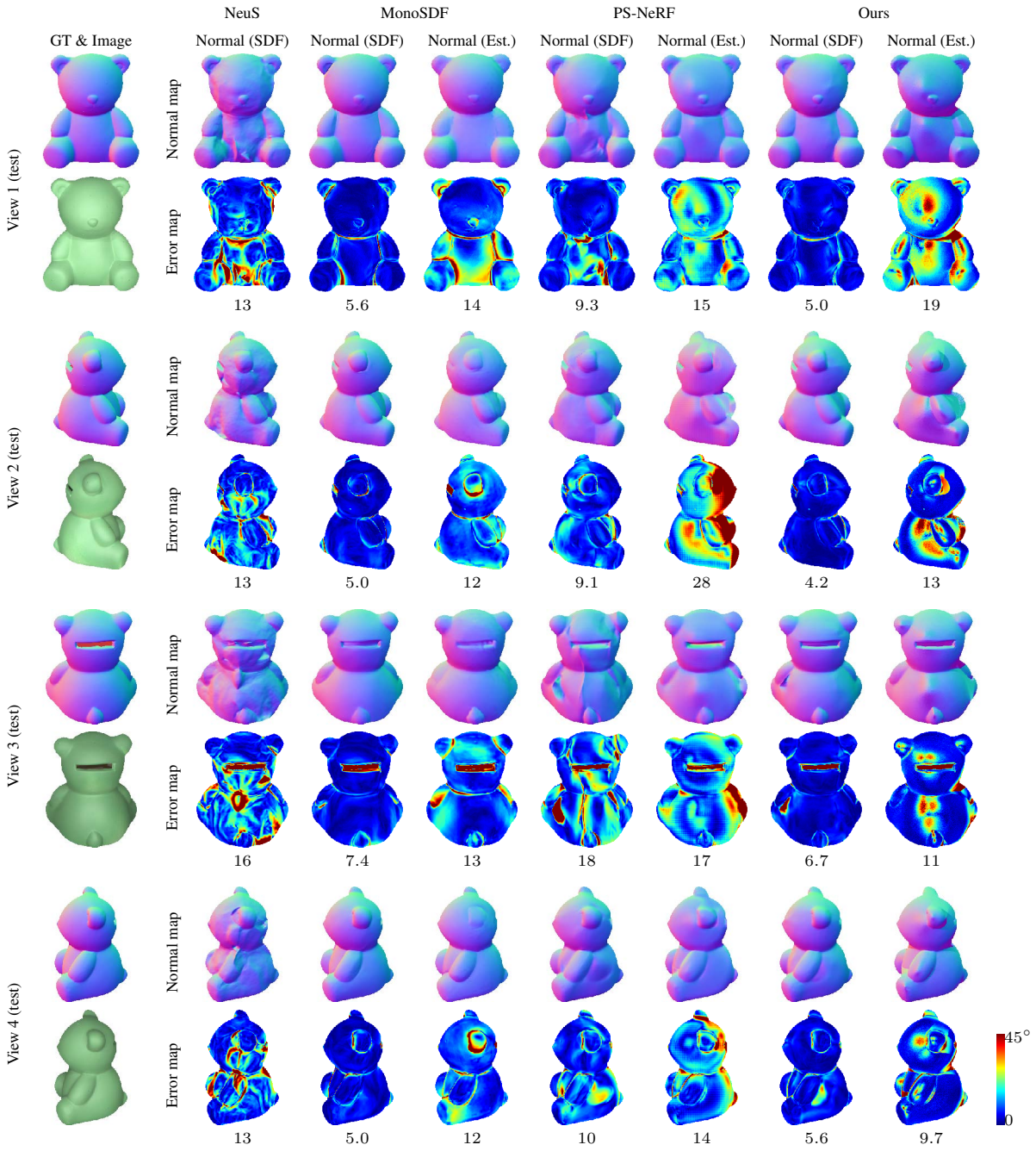

Figure S6. Evaluation of the normal maps for BEAR from test views.

Figure S7. Estimated normal maps for the BUDDHA scene from training views.

Figure S8. Estimated normal maps for the BUDDHA scene from test views.

Figure S9. Estimated normal maps for the POT2 scene from training views.

Figure S10. Estimated normal maps for the POT2 scene from test views.

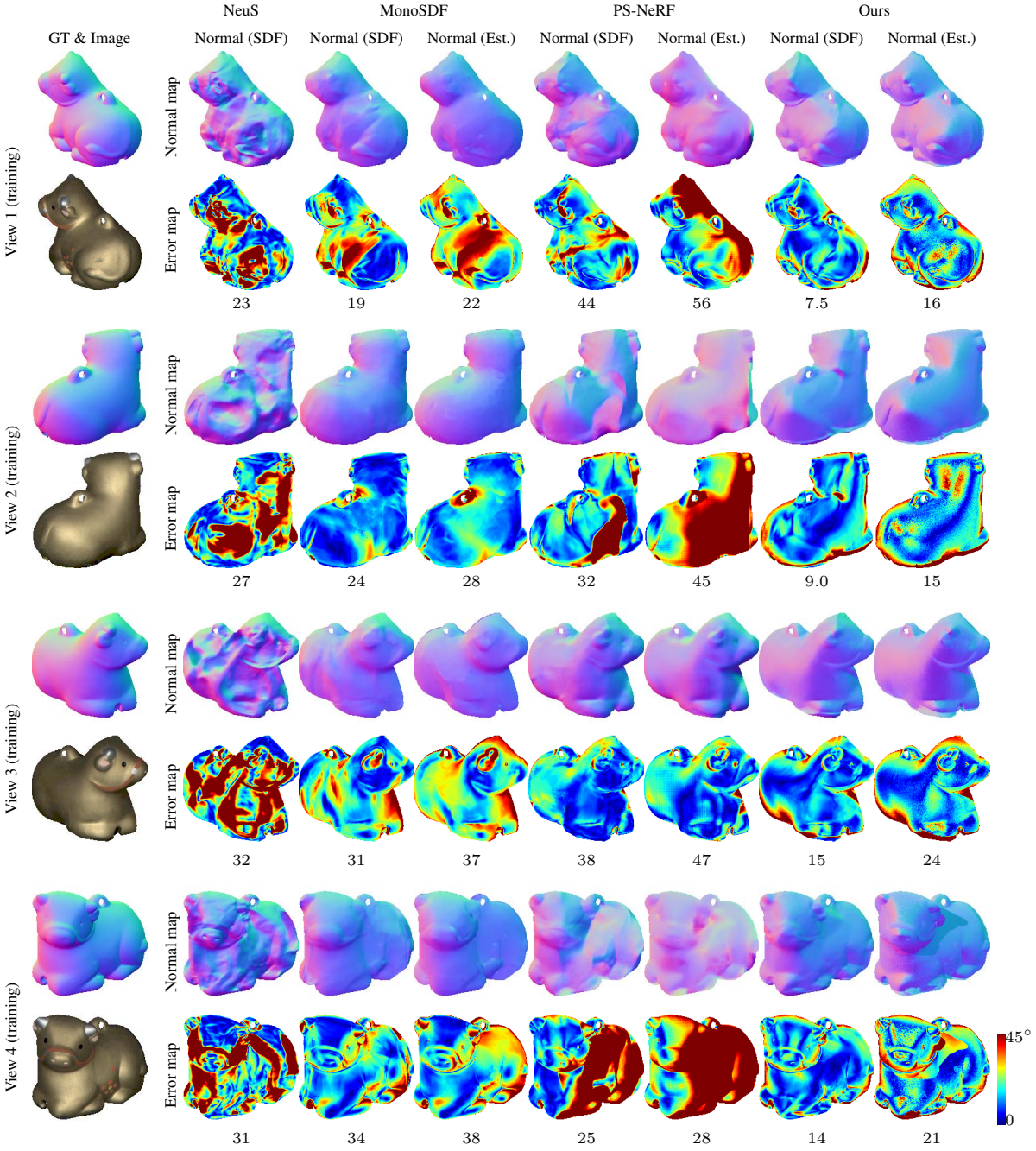

Figure S11. Estimated normal maps for the COW scene from training views.

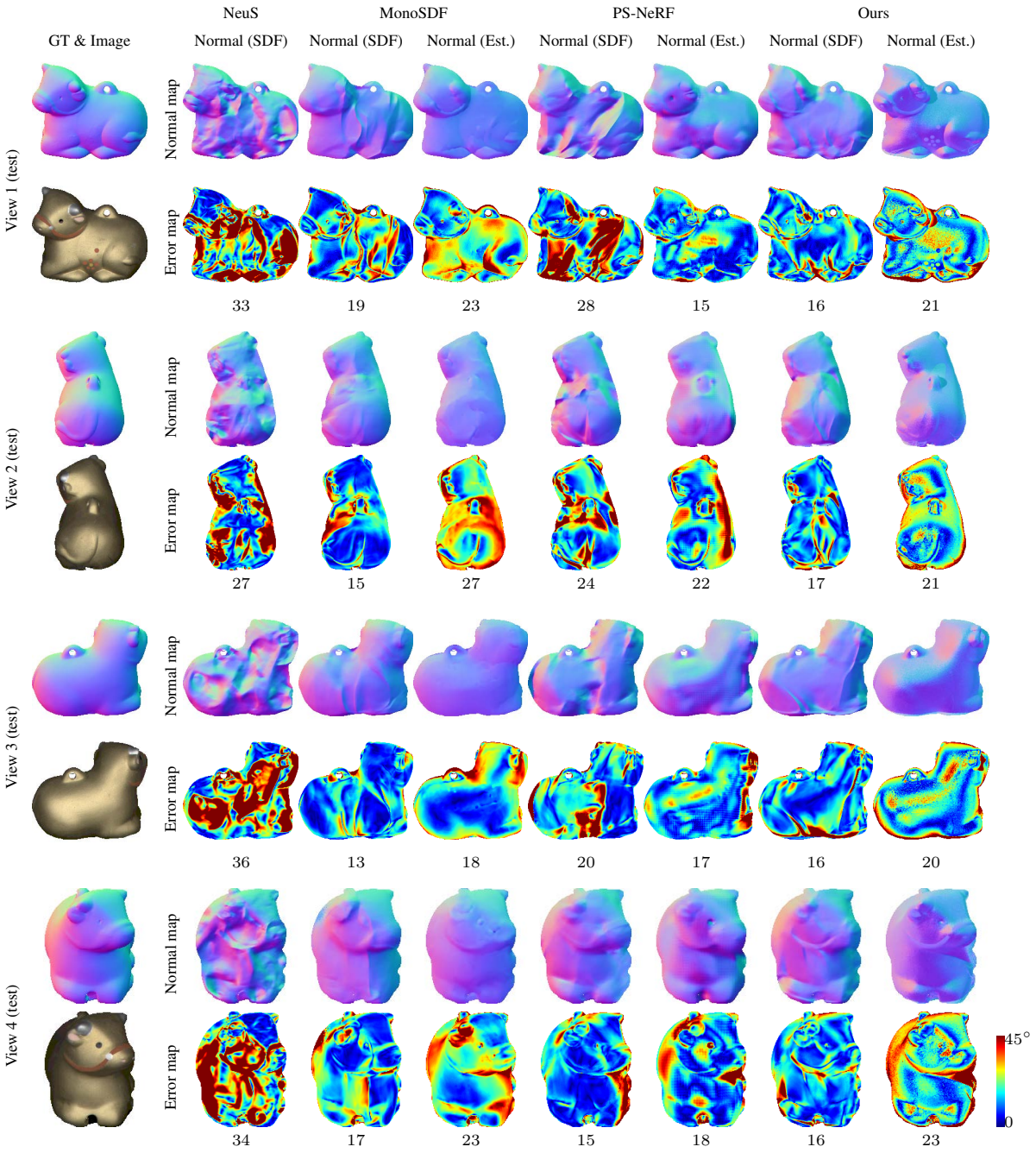

Figure S12. Estimated normal maps for the COW scene from test views.

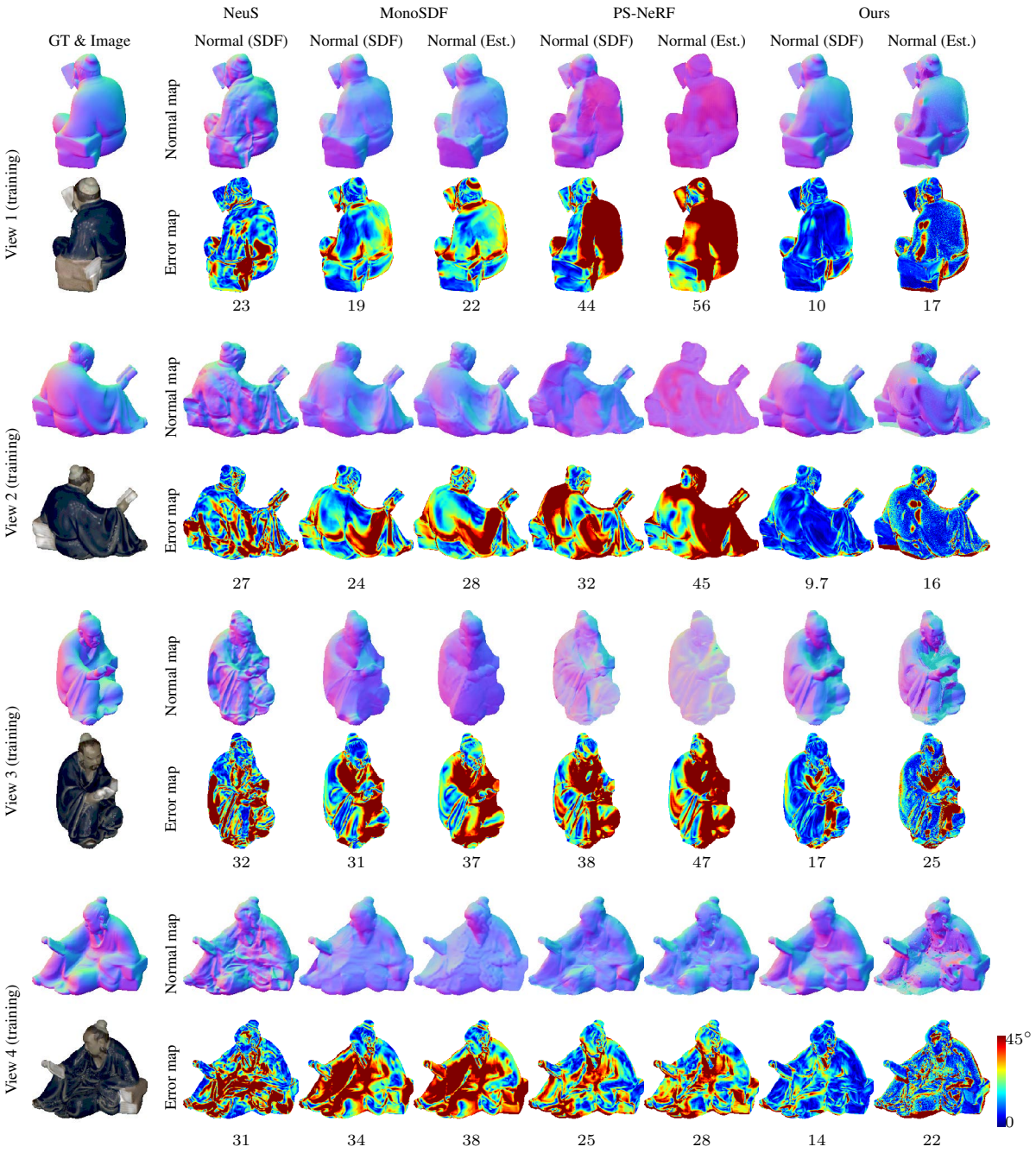

Figure S13. Estimated normal maps for the READING scene from training views.

Figure S14. Estimated normal maps for the READING scene from test views.

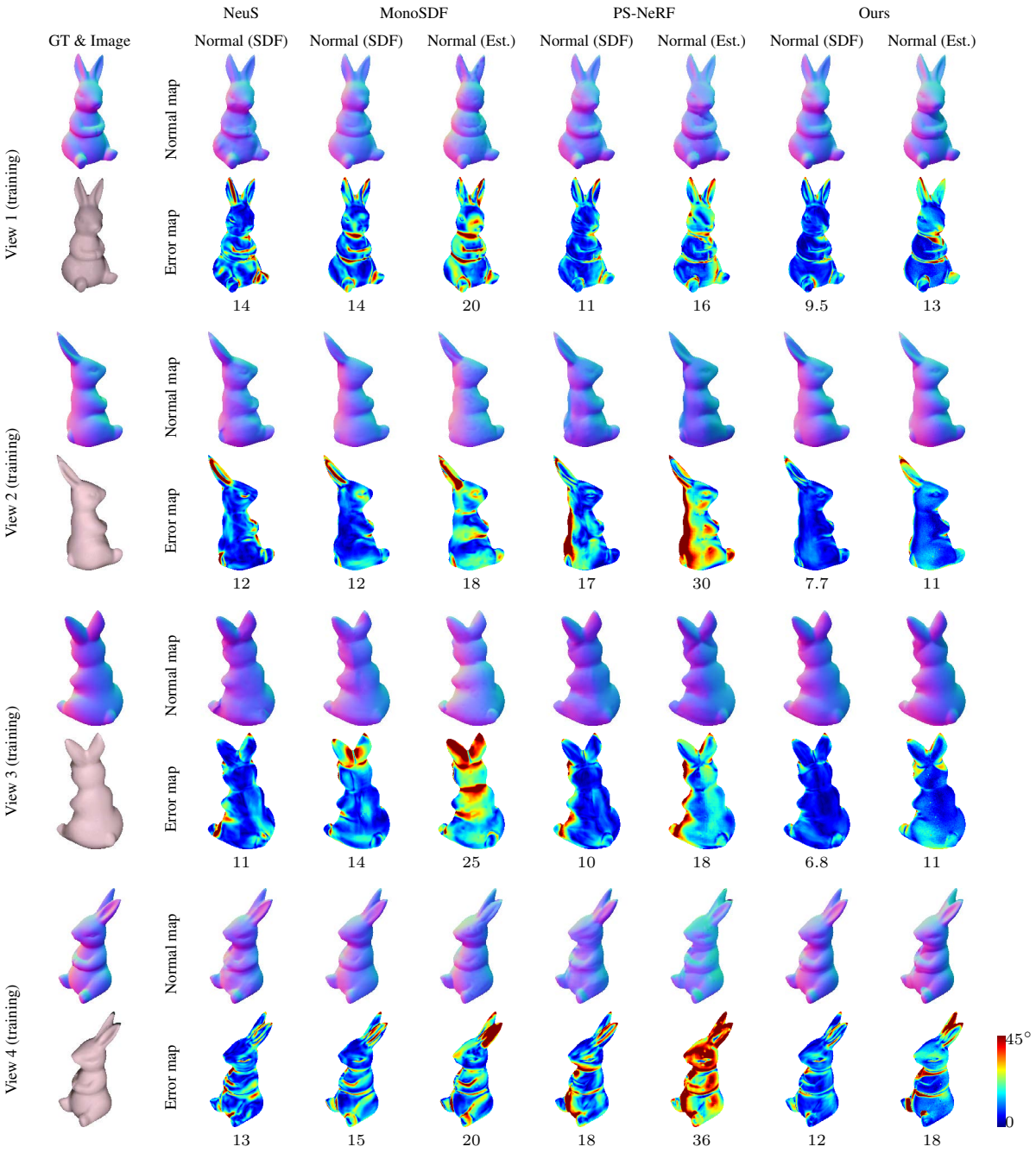

Figure S15. Estimated normal maps for the RABBIT scene from training views.

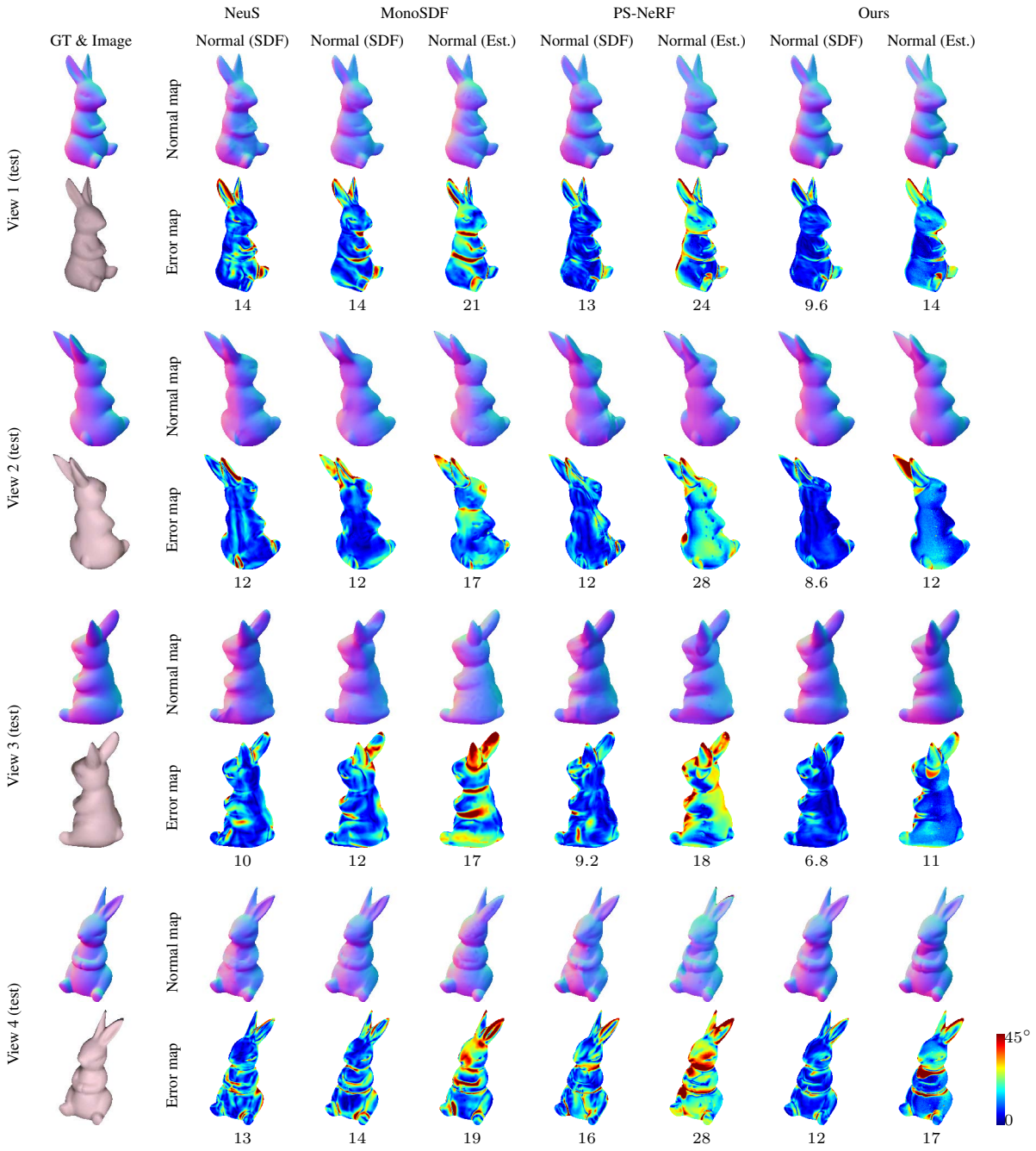

Figure S16. Estimated normal maps for the RABBIT scene from test views.

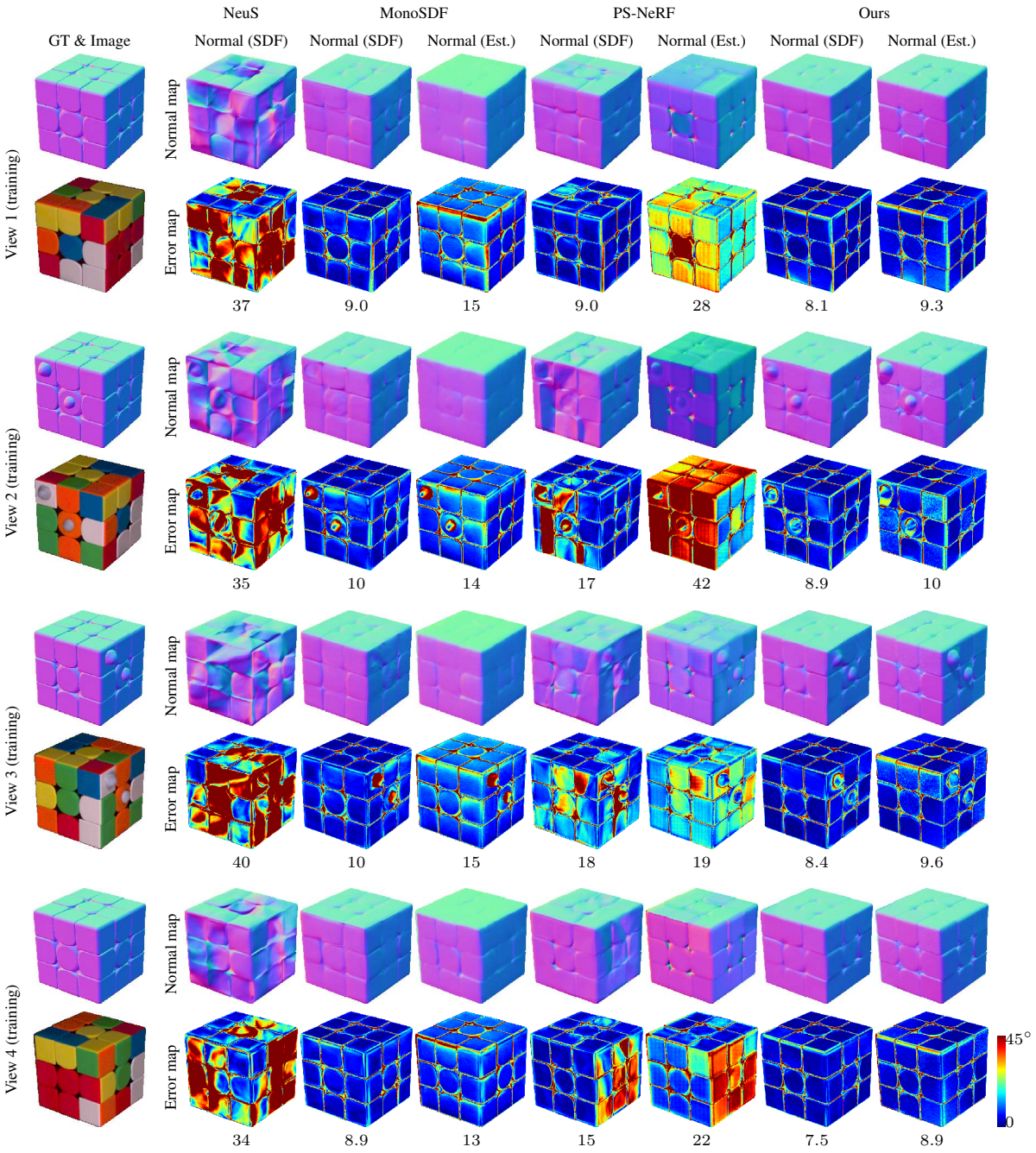

Figure S17. Estimated normal maps for the CUBE scene from training views.

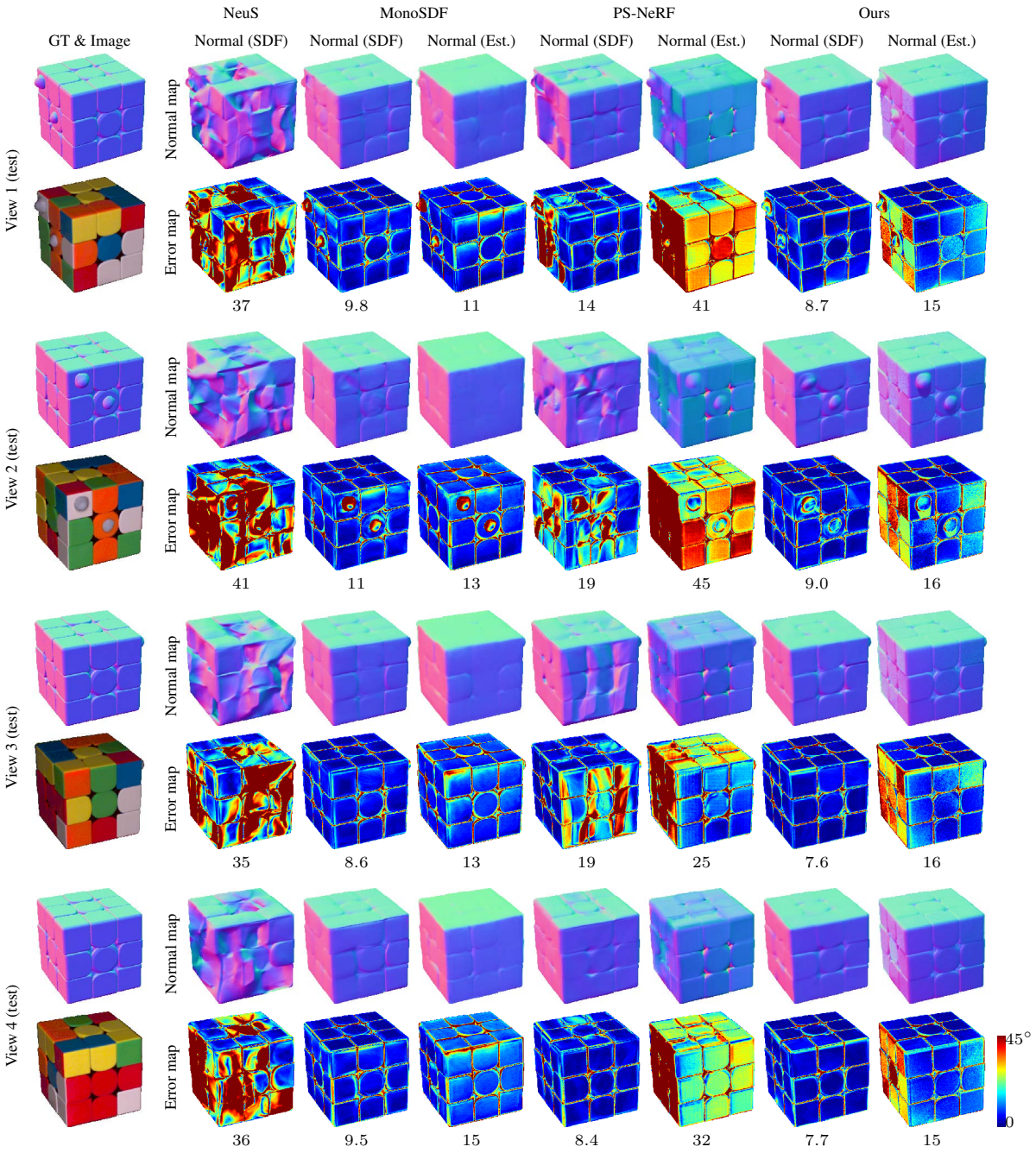

Figure S18. Estimated normal maps for the CUBE scene from test views.

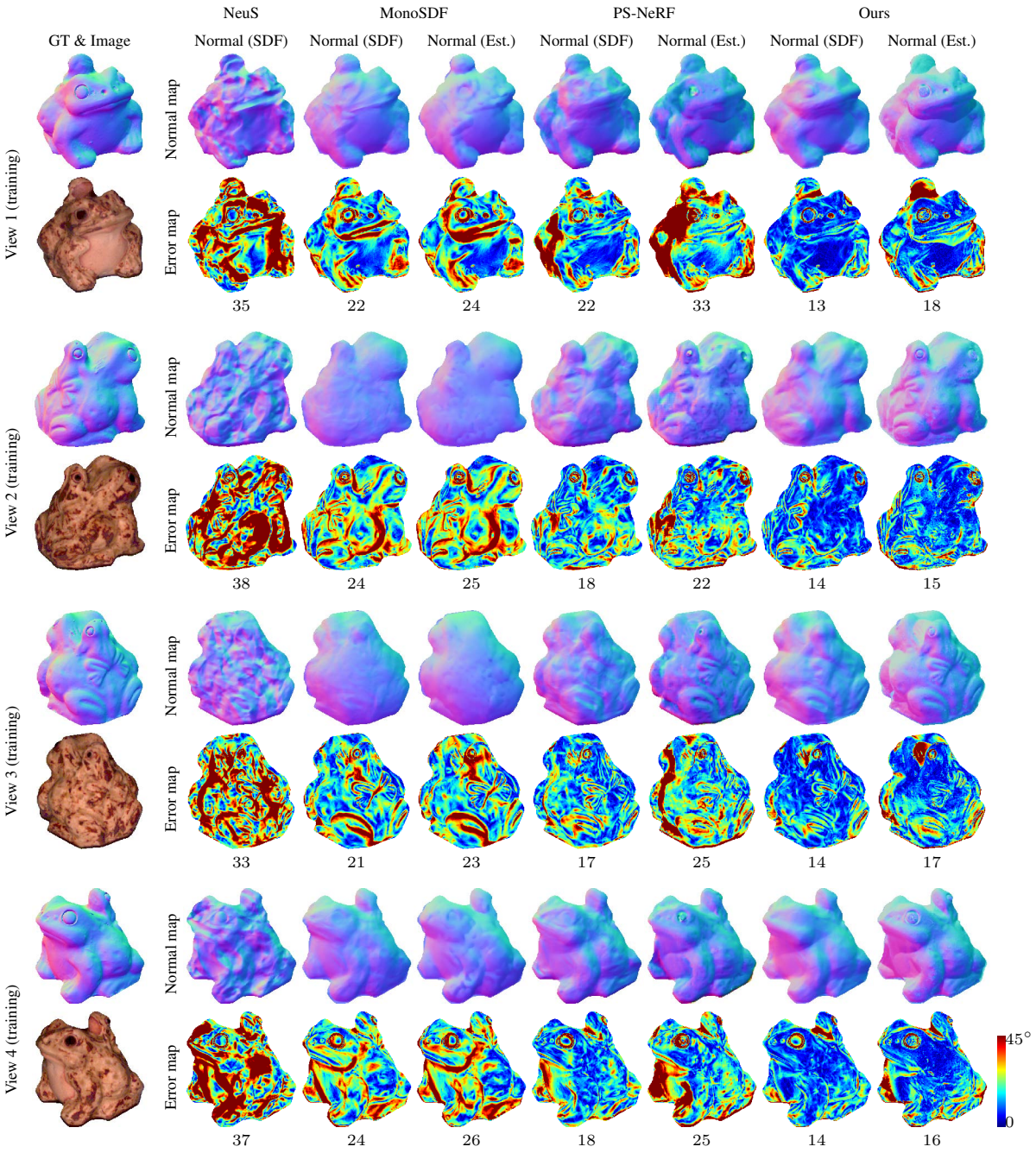

Figure S19. Estimated normal maps for the FROG scene from training views.

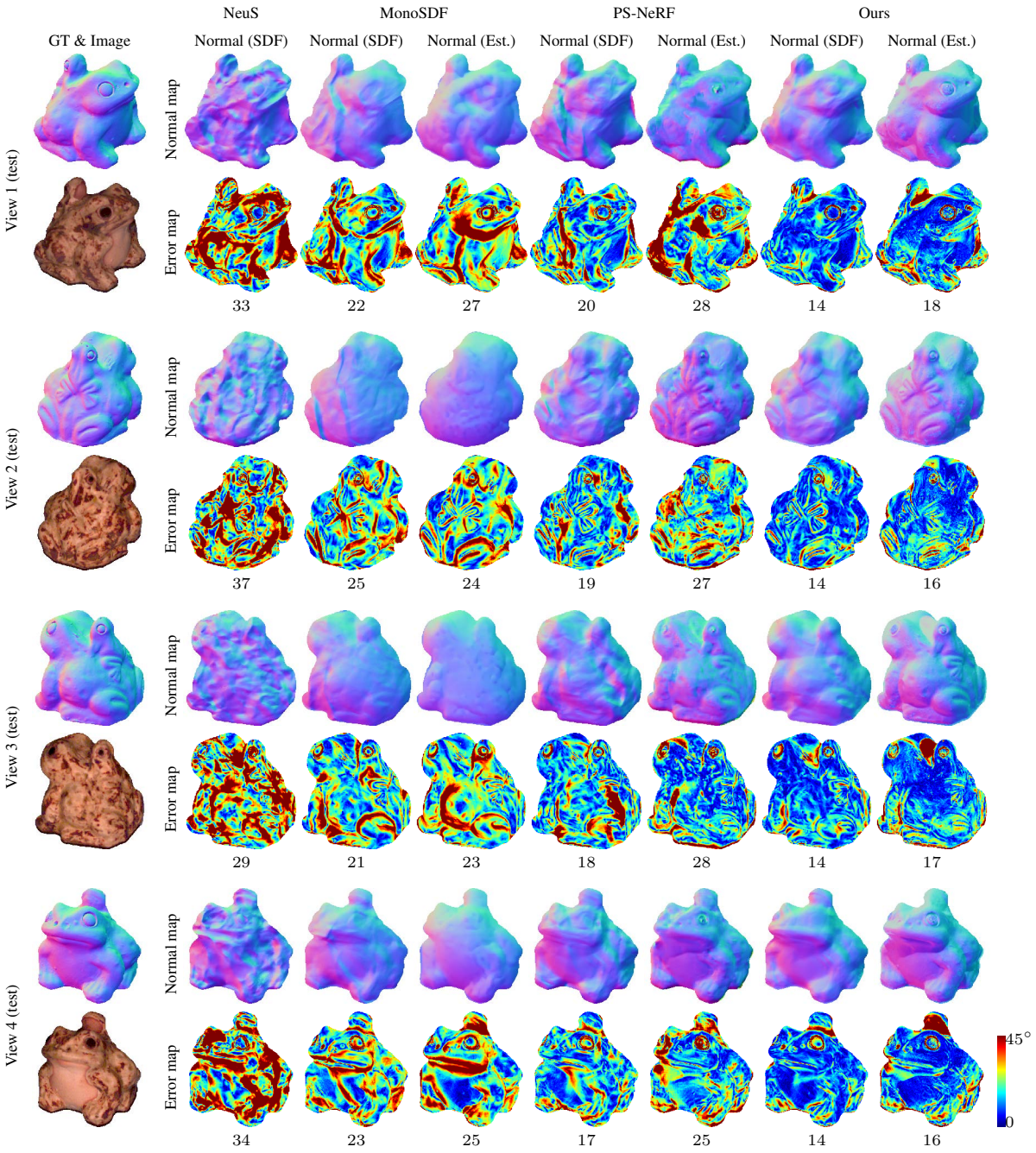

Figure S20. Estimated normal maps for the FROG scene from test views.