PAGaS: Pixel-Aligned 1DoF Gaussian Splatting for Depth Refinement

Supplementary Material

A. Visual comparison with the ground truth

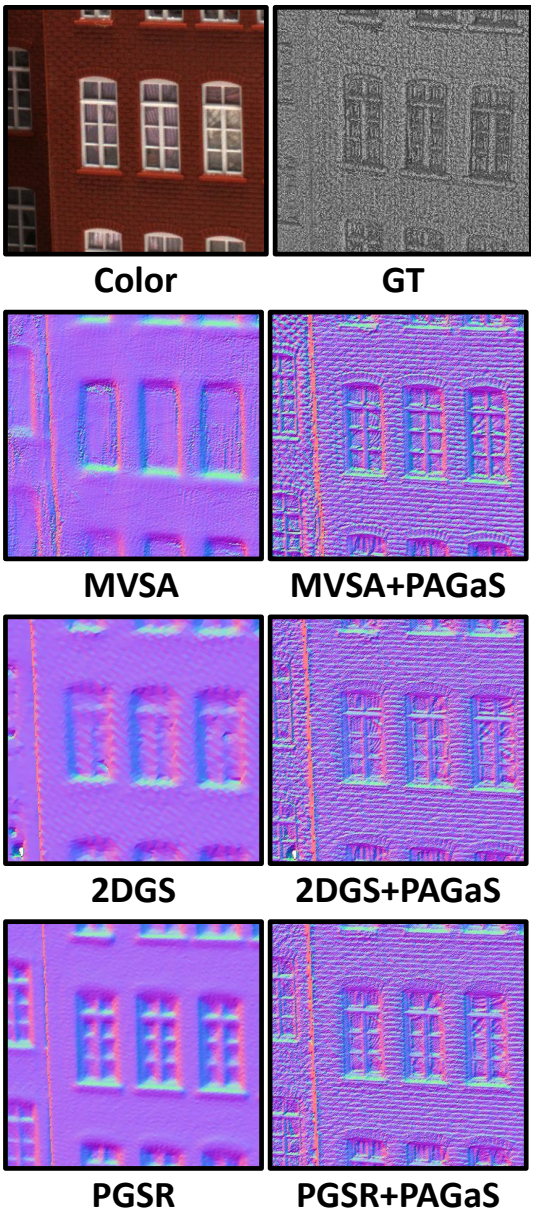

Figure SA.1. **Comparison of the normals** from depth for the baselines MVSAnywhere, 2DGS, and PGSR **before and after refinement** with our PAGaS, and the color **and the ground truth point cloud** in DTU scan24. See how PAGaS adds state-of-the-art high-frequency surface details independently of the quality of the initialization method. Notice that the individual building blocks are present in both the color and the ground truth, and PAGaS is the only method currently capable of capturing this level of detail.

Fig. SA.1 extends the teaser in Fig. 1 by providing results for the baselines (MVSAnywhere, 2DGS, and PGSR) and additionally visualizing the color and ground truth point cloud for DTU scan24. PAGaS consistently restores high-frequency geometric details that are lost in the baselines. It is the only method that accurately recovers the fine building-block structure visible in both the color and the ground truth. Full images and meshes can be found in Fig. SE.1. There, it can be observed that the fine details added in the per-view refined depth maps are preserved after fusion into the global 3D meshes, demonstrating that our method does not introduce locally inconsistent details but instead produces globally consistent ones.

B. TnT inconsistent camera exposure problem

Figure SB.1. **Inconsistent camera exposure problem on TnT**. Each image was captured with a different and unknown camera exposure. This misleads photometric optimization methods that assume pixel color consistency across stereo views to triangulate 3D points, including 2DGS and our PAGaS.

The DTU dataset was captured in a controlled studio environment with stable camera exposures. In contrast, TnT contains frames with unknown and varying exposures, as shown in Fig. SB.1. Purely photometric optimization methods whose only input is the set of color images, and which rely on no external priors to constrain the problem, are therefore highly sensitive to the quality of these images. Large exposure shifts between adjacent views make the downstream reconstruction task more difficult, leading to degraded results. PGSR explicitly optimizes a per-image illumination and exposure model, which reduces the impact of these variations. Motivated by this observation, in Tab. SB.1 we evaluate the effect of using a similar exposure model and a multi-view stereo consistency check during our PAGaS refinements on top of all baselines. Since neither 2DGS nor MVSAnywhere compensate for exposure in the input images, we also avoid compensating for them when refining their depths with PAGaS in the main experiments in

Scene	Barn	Caterpillar	Courthouse	Ignatius	Meetingroom	Truck	Mean
MVSAnywhere + PAGaS w/o exposure	0.29	0.12	0.09	0.25	0.15	0.21	0.18
MVSAnywhere + PAGaS w/ exposure	0.32	0.14	0.09	0.28	0.16	0.22	**0.20**
2DGS + PAGaS w/o exposure	0.42	0.27	0.17	0.57	0.05	0.19	0.28
2DGS + PAGaS w/ exposure	0.47	0.33	0.22	0.68	0.04	0.19	**0.32**
PGSR + PAGaS w/o exposure	0.49	0.40	0.21	0.75	0.13	0.23	0.37
PGSR + PAGaS w/ exposure	0.52	0.45	0.24	0.81	0.14	0.23	**0.40**

Table SB.1. **Quantitative TnT evaluation with and without camera exposure compensation** showing the F1-score ↑ for baseline methods before and after our PAGaS refinement.

Tab. 3. This allows us to demonstrate, under identical input conditions, that PAGaS consistently improves their reconstructions.

C. Ablations

2DGS	2DGS + PAGaS	2DGS	2DGS + PAGaS
w/o Depth from mesh	0.564	200, 100 Steps	0.468
Random depth	10,79	100, 200 Steps	0.458
1 Context view	0.465	w/o Fix opacity	0.457
2 Context views	0.464	w/o Scale from depth	0.479
3 Context views	0.464	Learning rate 1e-4	0.563
4 Context views	0.463	3 Radius threshold	0.461
5 Context views	0.462	4 Depth Slices	0.460
6 Context views	0.461	w/o Occlusion-Aware Rasterizer	0.464
7 Context views	0.461	w/o Smoothness	0.457
8 Context views	0.459	View-dependent color	0.454
9 Context views	0.458	0,8 λ_c	0.456
w/o Scales	0.460	w/ Exposure compensation	0.458
4, 2, 1 Scales	0.495	w/ Stereo consistency	0.471
8, 4, 2, 1 Scales	0.630	Full model	0.456

Table SC.1. **Ablation study** of the characteristic parameters of our PAGaS. Showing Chamfer distance in mm (lower better) for 2DGS after PAGaS refinement in scan24 of DTU dataset.

Tab. SC.1 summarizes the ablations of the key parameters of PAGaS.

If instead of using the depth extracted from the baseline mesh, we directly input the baseline's estimated depth for that view, this leads to poorer convergence. PAGaS is a refinement method: it adds small-scale surface detail to the input depth without changing its global structure. The more stereo-consistent the input depth is, the better PAGaS performs, because the details refined in one view align with those captured in the others.

If no initialization depth is provided (i.e., we start from random depth bounded by the cameras' pose range), PAGaS fails to converge. Its role is to move the Gaussians attached to each pixel's depth slightly along the ray, not to correct large errors or traverse long distances.

Accuracy improves as more context views are used to constrain the optimization. Notably, even a small number of views already approaches the full-model performance, demonstrating the robustness and stability of our method. This also means the number of context views can be reduced to save GPU memory and increase speed with little loss in quality. Our full model with 10 context views at full resolution uses 8 GB of VRAM, whereas using a single context view requires only 2 GB and runs about 4× faster.

We use a two-scale pyramid in all experiments, since the baseline depth is typically produced at half the image resolution. Using more scales (i.e., more aggressive initial downscaling) makes the starting depth too coarse, whereas PAGaS benefits from an initialization that is already a close approximation to the true scene geometry.

Another experiment to highlight is modeling per-Gaussian view-dependent color using standard 3rd-order spherical harmonics. This increases the per-Gaussian parameter count from 1 (depth) to 48 (3 RGB channels×16 SH coefficients). Accuracy remains essentially unchanged, but GPU VRAM usage grows substantially (it no longer fits on an 11 GB GPU) and the optimization takes roughly twice as long.

As explained in Appendix B, the DTU dataset has consistent camera exposure across all images. Consequently, applying per-view exposure compensation or a multi-view stereo consistency check on DTU does not provide any benefit, as the photometric optimization remains stable across views.

D. Depth vs normal visualization

Fig. SD.1 illustrates the effect of our refinement by comparing the depth and corresponding normal maps on DTU scan24 before and after applying PAGaS. While depth maps alone tend to obscure fine-scale geometry due to the small magnitude of relative depth variations, the normal maps clearly reveal the real quality of the baselines and the improvements introduced by our method. As shown, PAGaS enhances high-frequency surface structure that is largely absent in the original MVSAnywhere reconstruction.

E. Additional qualitative evaluations

Extensive additional qualitative evaluations of the baselines MVSAnywhere, 2DGS and PGSR before and after our PAGaS refinement are provided for DTU in scans 24, 106, 122, 69 and 37 at Figs. SE.1 to SE.5; for TnT in Barn, Caterpil-

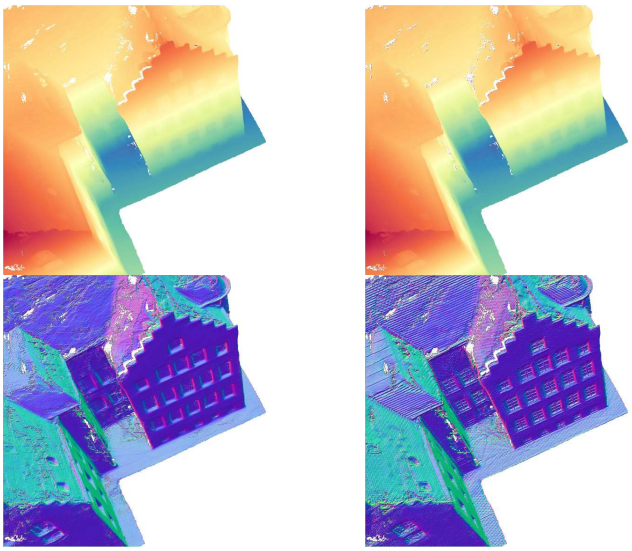

Figure SD.1. **Depth and normal comparison on DTU scan24.** Left: depth and normals from the original MVSAnywhere mesh. Right: results after applying our PAGaS refinement. Although depth maps may appear similar at first glance, the normals highlight the substantial improvement in fine-scale surface geometry.

lar, Courthouse, Ignatius and Truck at Figs. SE.6 to SE.10; and of Colmap for BlendedMVS at Fig. SE.11.

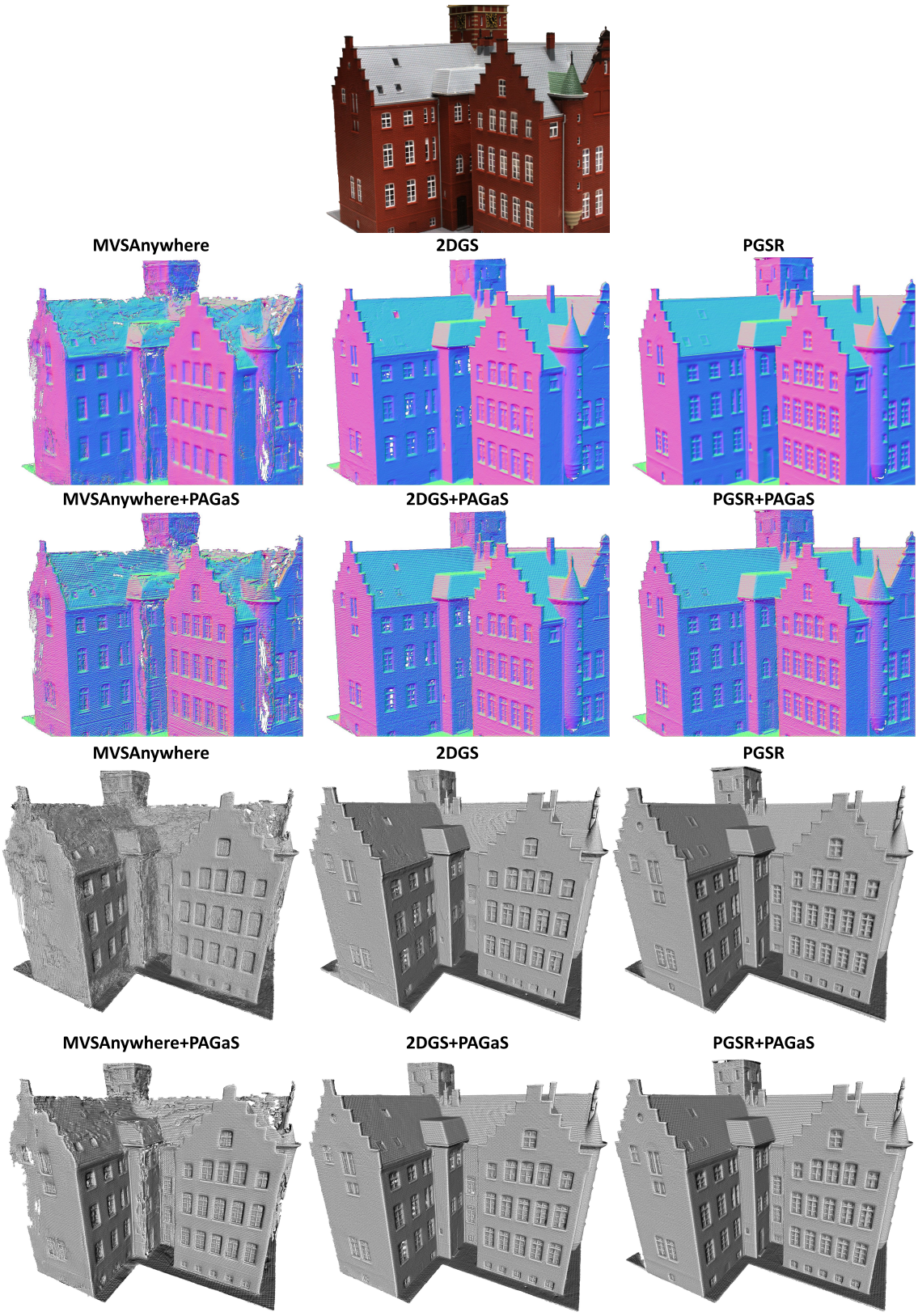

Figure SE.1. **Qualitative evaluation of the normals from depth and the 3D meshes in DTU scan24** for the three main baselines before and after refinement with our PAGaS. Zoom in to appreciate the added pixel-wise details.

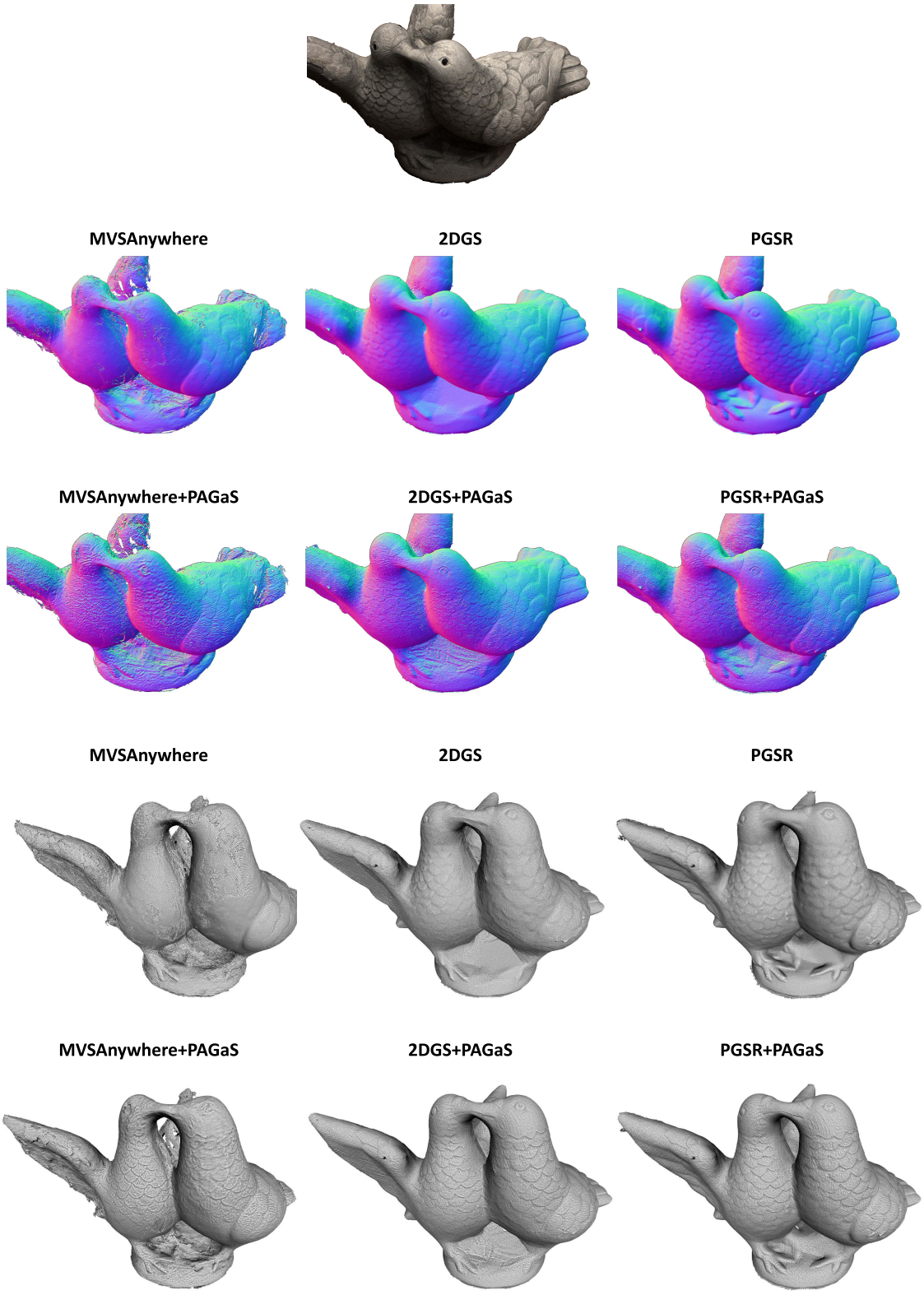

Figure SE.2. **Qualitative evaluation of the normals from depth and the 3D meshes in DTU scan106** for the three main baselines before and after refinement with our PAGaS. Zoom in to appreciate the added pixel-wise details.

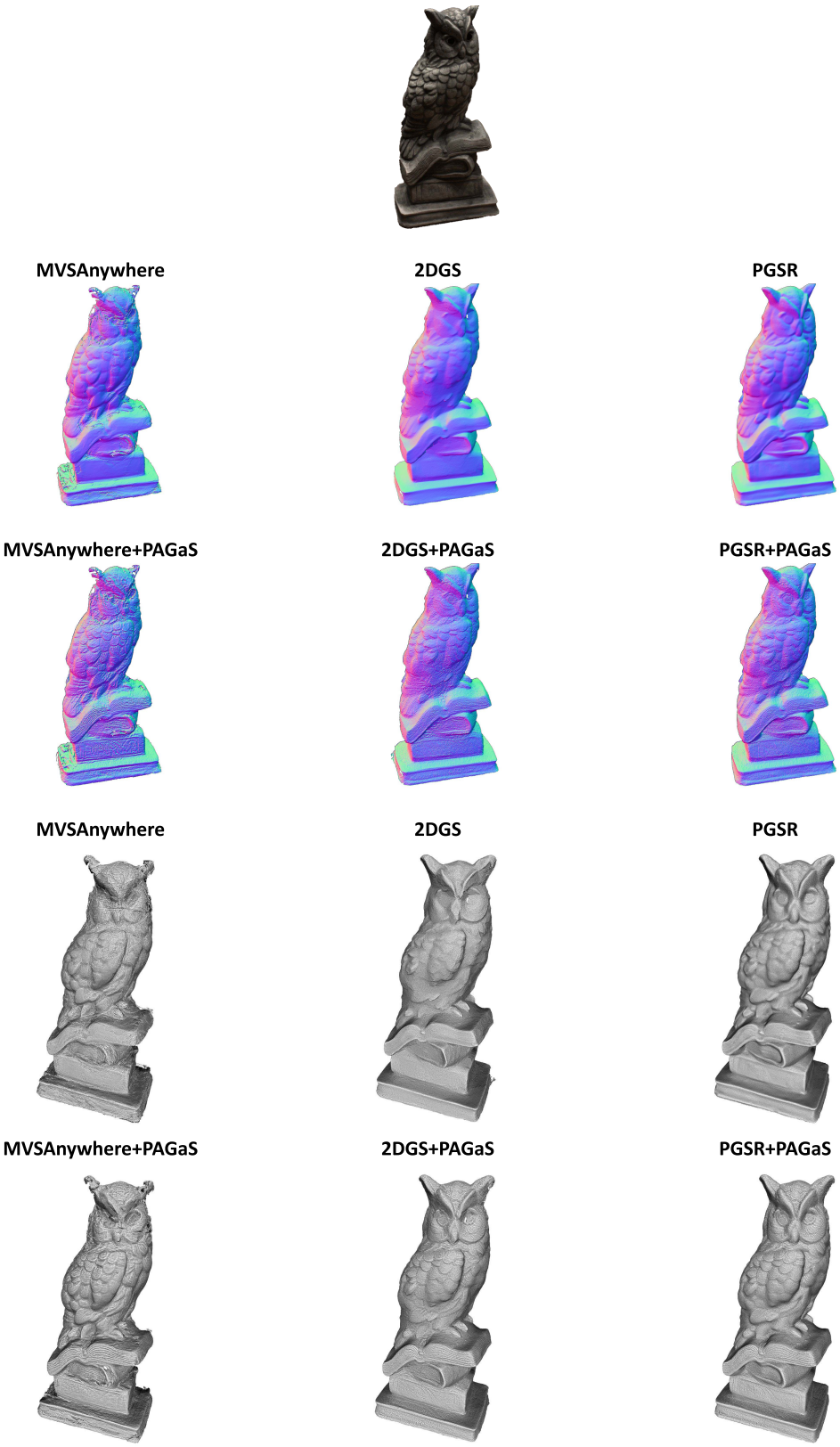

Figure SE.3. **Qualitative evaluation of the normals from depth and the 3D meshes in DTU scan122** for the three main baselines before and after refinement with our PAGaS. Zoom in to appreciate the added pixel-wise details.

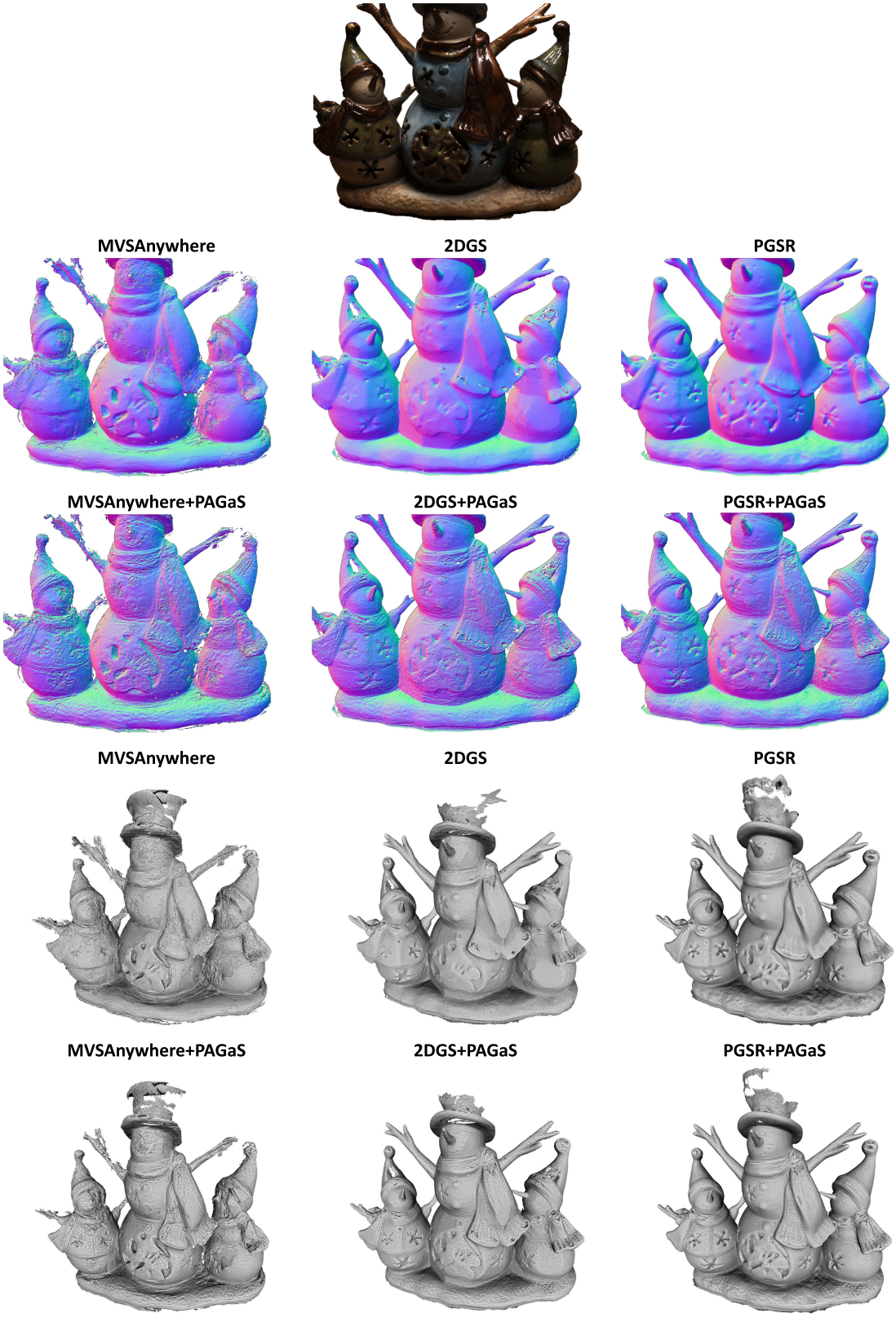

Figure SE.4. **Qualitative evaluation of the normals from depth and the 3D meshes in DTU scan69** for the three main baselines before and after refinement with our PAGaS. Zoom in to appreciate the added pixel-wise details.

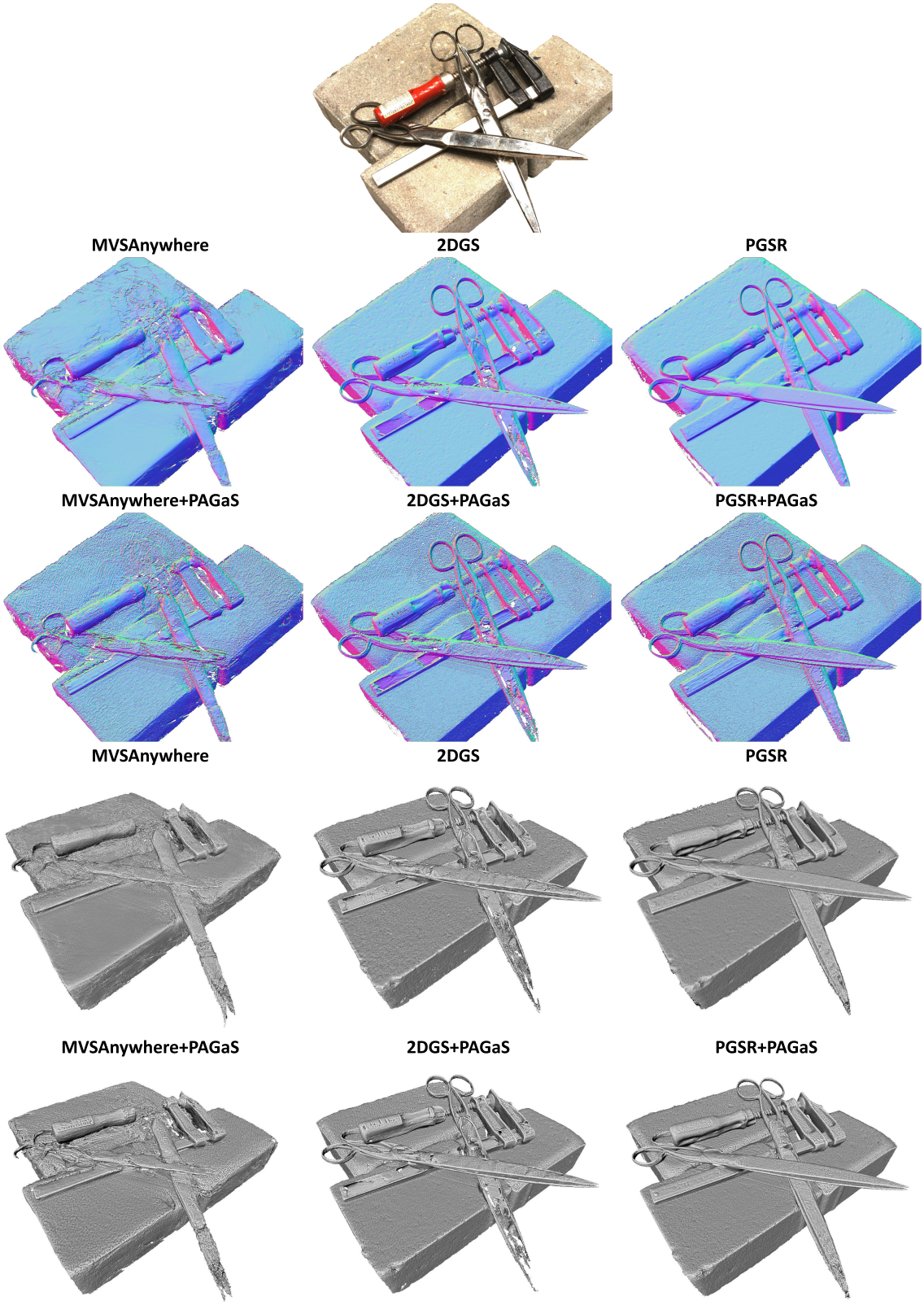

Figure SE.5. **Qualitative evaluation of the normals from depth and the 3D meshes in DTU scan37** for the three main baselines before and after refinement with our PAGaS. Zoom in to appreciate the added pixel-wise details.

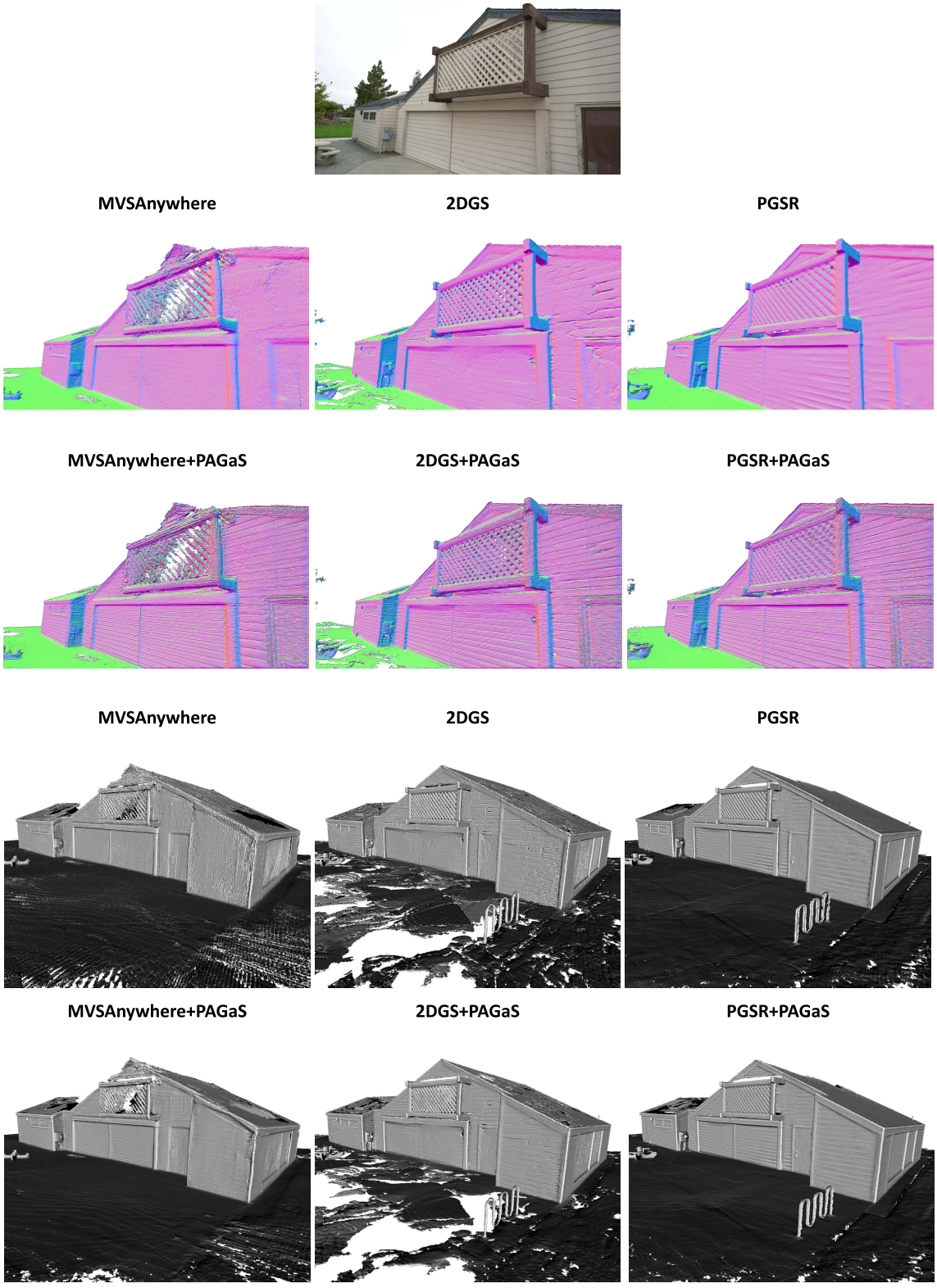

Figure SE.6. **Qualitative evaluation of the normals from depth and the 3D meshes in TNT Barn** for the three main baselines before and after refinement with our PAGaS. Zoom in to appreciate the added pixel-wise details.

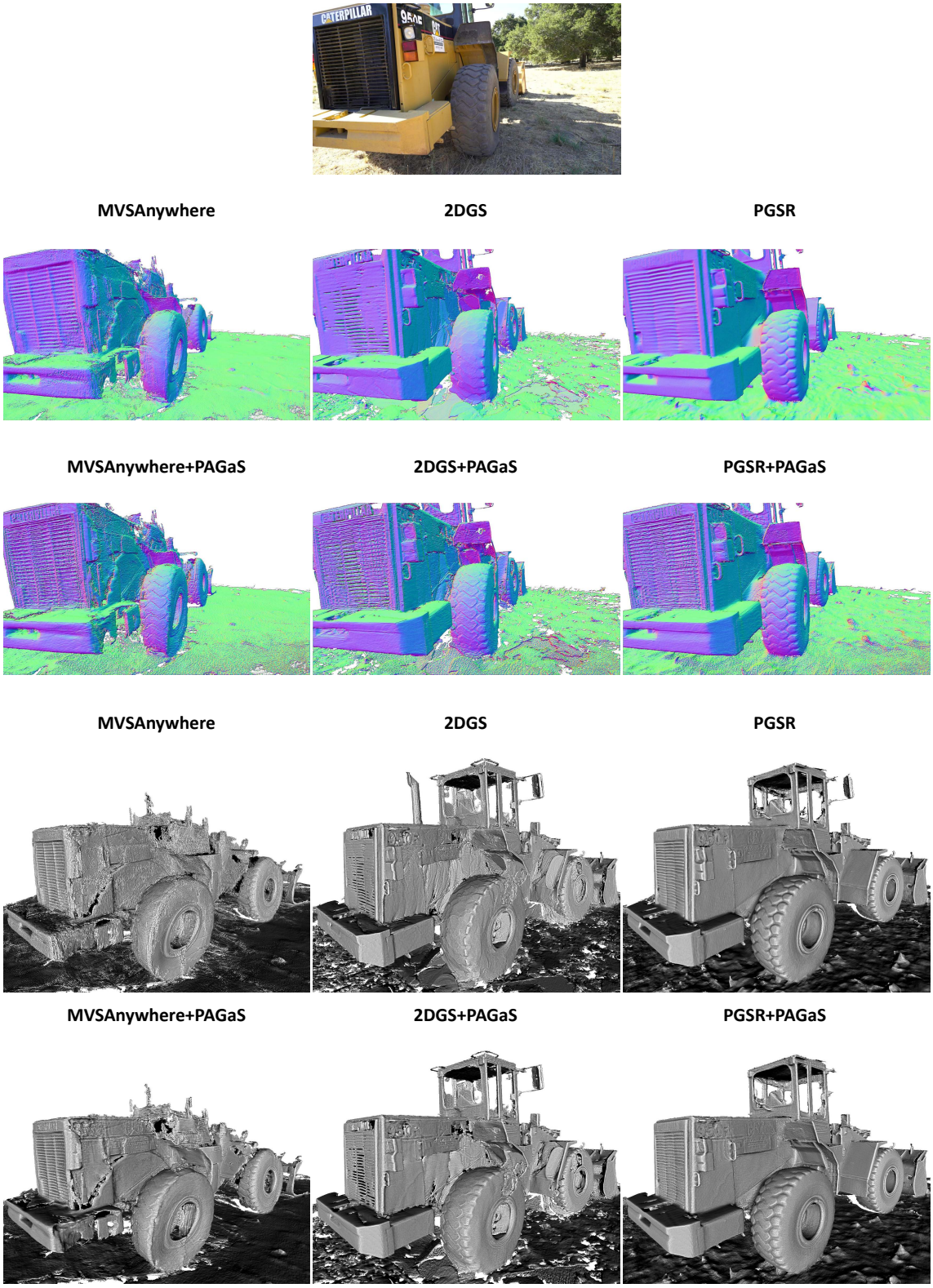

Figure SE.7. **Qualitative evaluation of the normals from depth and the 3D meshes in TNT Caterpillar** for the three main baselines before and after refinement with our PAGaS. Zoom in to appreciate the added pixel-wise details.

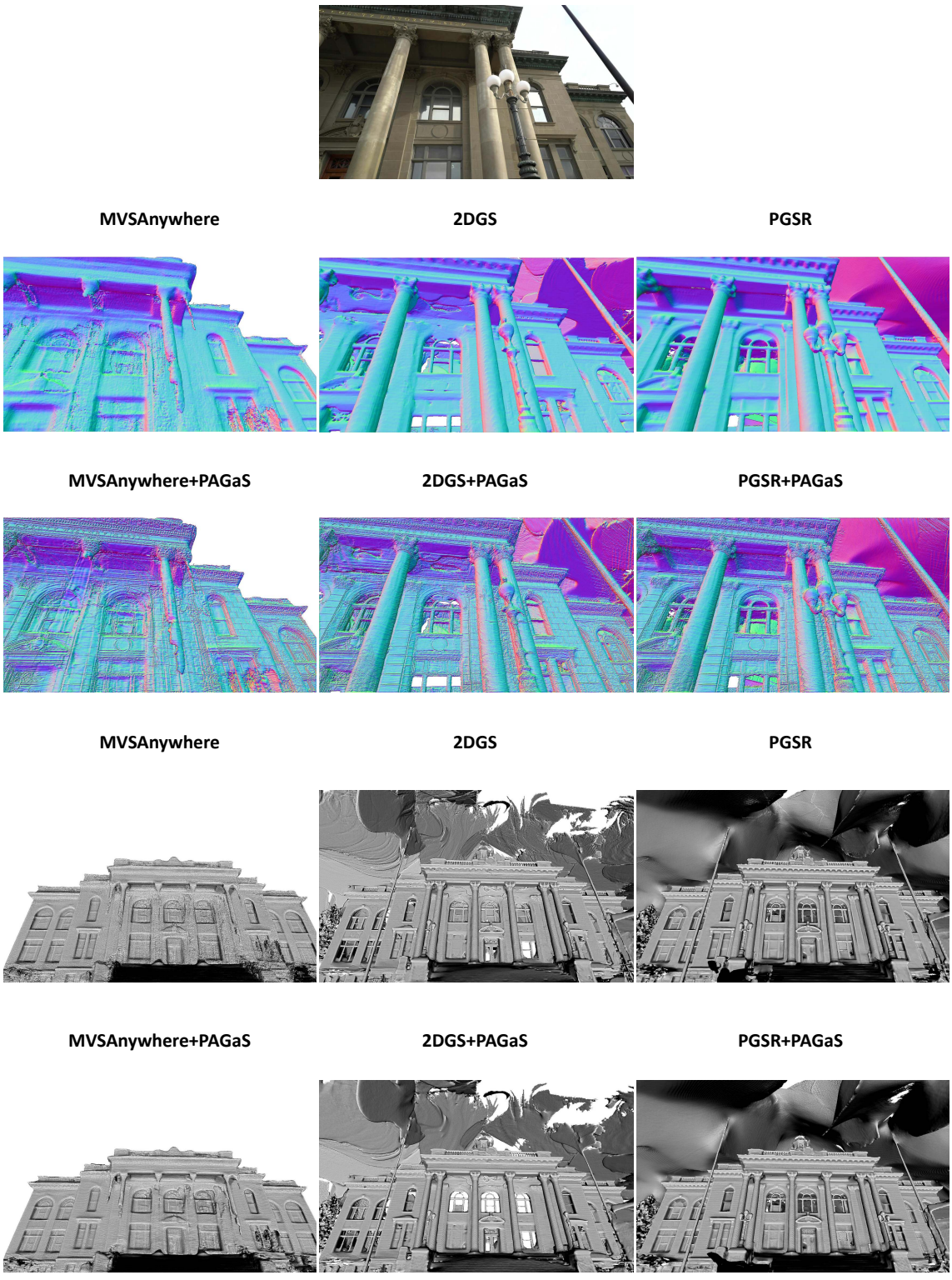

Figure SE.8. **Qualitative evaluation of the normals from depth and the 3D meshes in TNT Courthouse** for the three main baselines before and after refinement with our PAGaS. Zoom in to appreciate the added pixel-wise details.

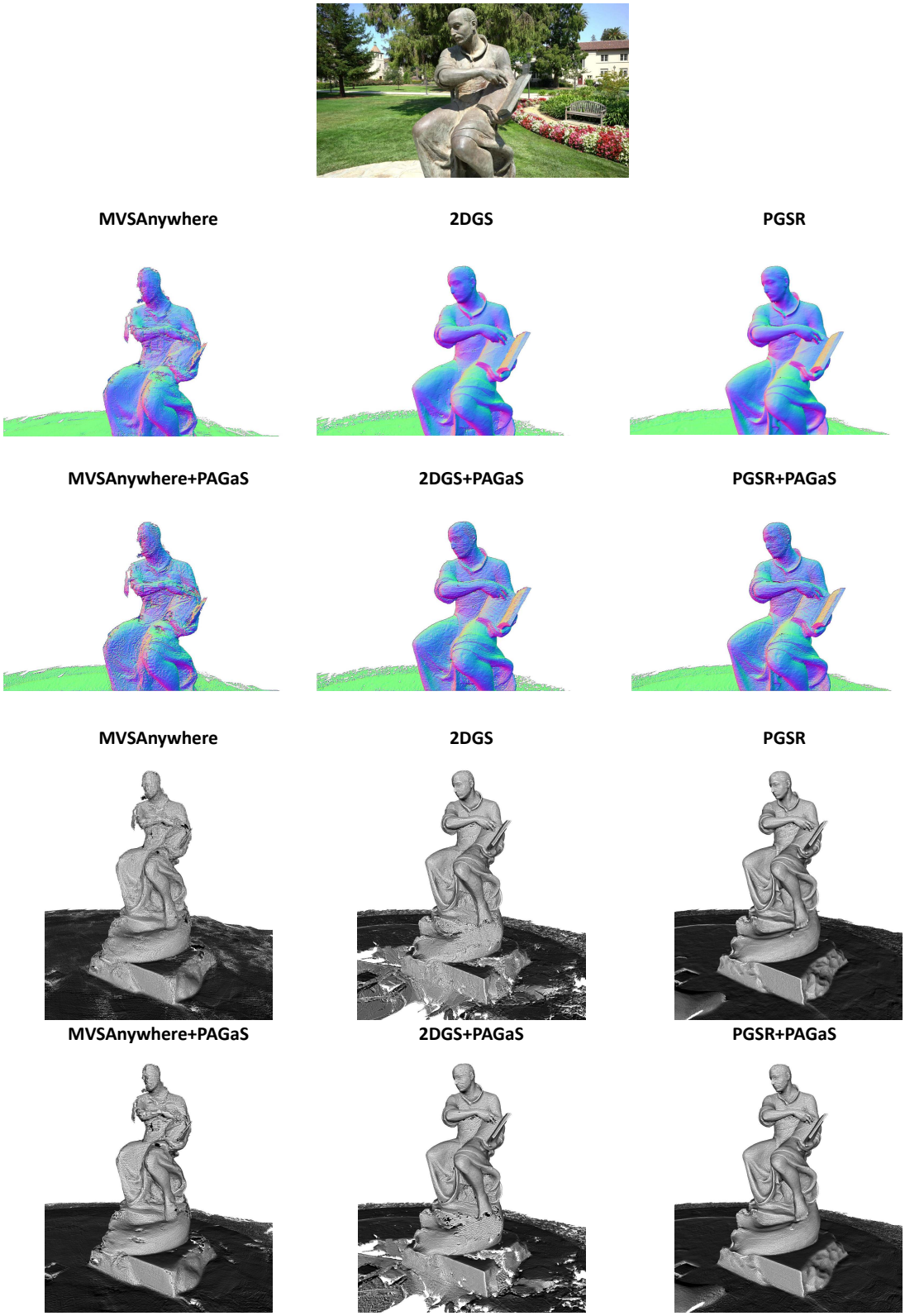

Figure SE.9. **Qualitative evaluation of the normals from depth and the 3D meshes in TNT Ignatius** for the three main baselines before and after refinement with our PAGaS. Zoom in to appreciate the added pixel-wise details.

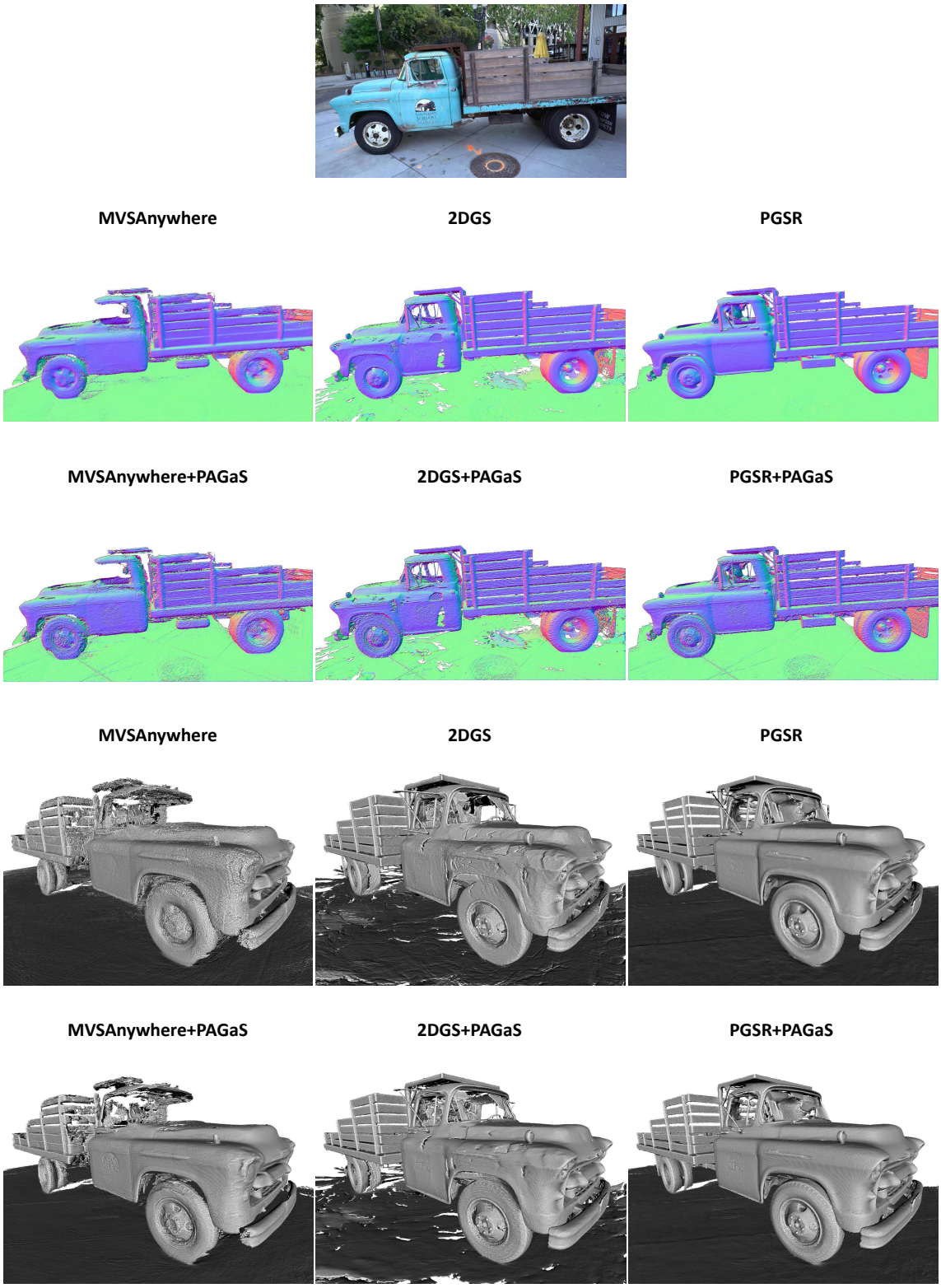

Figure SE.10. **Qualitative evaluation of the normals from depth and the 3D meshes in TNT Truck** for the three main baselines before and after refinement with our PAGaS. Zoom in to appreciate the added pixel-wise details.

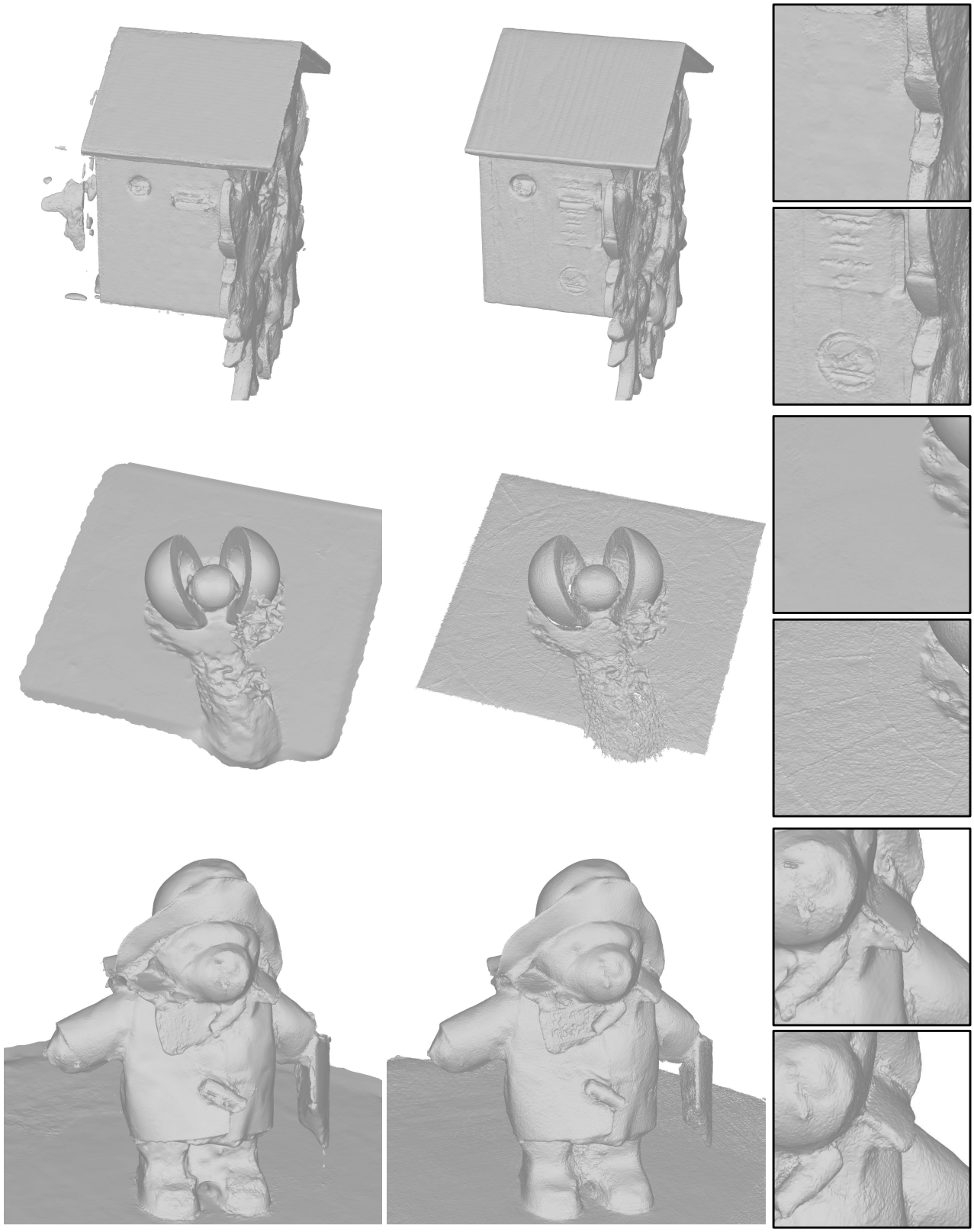

Figure SE.11. **BlendedMVS meshes** by COLMAP's multi-view stereo before (left) and after (right) applying our PAGaS with some close-up areas. From top to bottom: clock, stone, bear. Zoom in to observe the fine-grained details refined by our PAGaS.