Denoising Monte Carlo Renders with Diffusion Models

Supplementary Material

6. Additional Evaluation

We evaluate our method on the noisebase dataset (https://balint.io/noisebase/datasets/index.html), which consists of 1024 sequences of 64 frames each, 256-res. Each training and test sample has accompanying feature buffers (depth, normal, and albedo) as well as temporal information like camera parameters and motion vectors that we do not consider here. Each pixel contains per-sample information up to 32spp, which we average to pixel-space at the appropriate sampling rate (as our method operates on pixels and not samples). We then compare our method to other pixel-space methods, [1], [49]. The test sets consist of up to 9 scenes with 40-160 full HD frames in each. In this experiment, we train one model to denoise the full spectrum of available sampling rates (1-32spp) for our method and AFGSA [49]. We use OIDN's pretrained model as additional evaluation [1]. In our qualitative evaluation, we present several examples showing that our method excels at producing reasonable textures particularly in undersampled regions. Other methods shift the color or hallucinate details very different from the input.

Table 2. **Quantitatively, our method is competitive with SOTA**. We compare error metrics across different spp settings (2, 4, 8, 32) for various methods on the noisebase test set, consisting of several scenes at full HD. Across all sampling rates, we are the best method for most metrics, which evaluate both pixel-wise and perceptual quality. DINO [12, 29] and CLIP measure cosine similarity of the generated and GT global image feature vector.

Method	**L1** ↓	**PSNR** ↑	**LPIPS** ↓	**DINO** ↑ [29]	**CLIP** ↑ [31]	**FliP** ↓ [3]	**FoVVDP** ↑ [25]
AFGSA [49]	0.163	24.8	0.436	0.946	0.887	0.167	6.34
OIDN [1]	**0.0990**	**26.8**	0.421	0.974	0.924	0.148	**7.07**
Ours	0.114	26.5	**0.401**	**0.975**	**0.939**	**0.147**	6.94

2spp

Method	**L1** ↓	**PSNR** ↑	**LPIPS** ↓	**DINO** ↑ [29]	**CLIP** ↑ [31]	**FliP** ↓ [3]	**FoVVDP** ↑ [25]
AFGSA [49]	0.0992	27.3	0.394	0.963	0.914	0.125	7.07
OIDN [1]	0.0769	28.2	0.392	0.980	0.936	0.123	**7.54**
Ours	**0.0748**	**28.5**	**0.368**	**0.982**	**0.950**	**0.114**	7.50

4spp

Method	**L1** ↓	**PSNR** ↑	**LPIPS** ↓	**DINO** ↑ [29]	**CLIP** ↑ [31]	**FliP** ↓ [3]	**FoVVDP** ↑ [25]
AFGSA [49]	0.0834	28.8	0.361	0.975	0.938	0.104	7.61
OIDN [1]	0.0611	29.7	0.363	0.986	0.947	0.101	**8.00**
Ours	**0.0558**	**30.3**	**0.339**	**0.988**	**0.959**	**0.0901**	**8.00**

8spp

Method	**L1** ↓	**PSNR** ↑	**LPIPS** ↓	**DINO** ↑ [29]	**CLIP** ↑ [31]	**FliP** ↓ [3]	**FoVVDP** ↑ [25]
AFGSA [49]	0.0616	31.7	0.310	0.988	0.970	0.0760	8.42
OIDN [1]	0.0387	**33.4**	0.318	0.994	0.965	0.0634	**8.93**
Ours	**0.0371**	**33.4**	**0.296**	**0.995**	**0.973**	**0.0589**	8.89

32spp

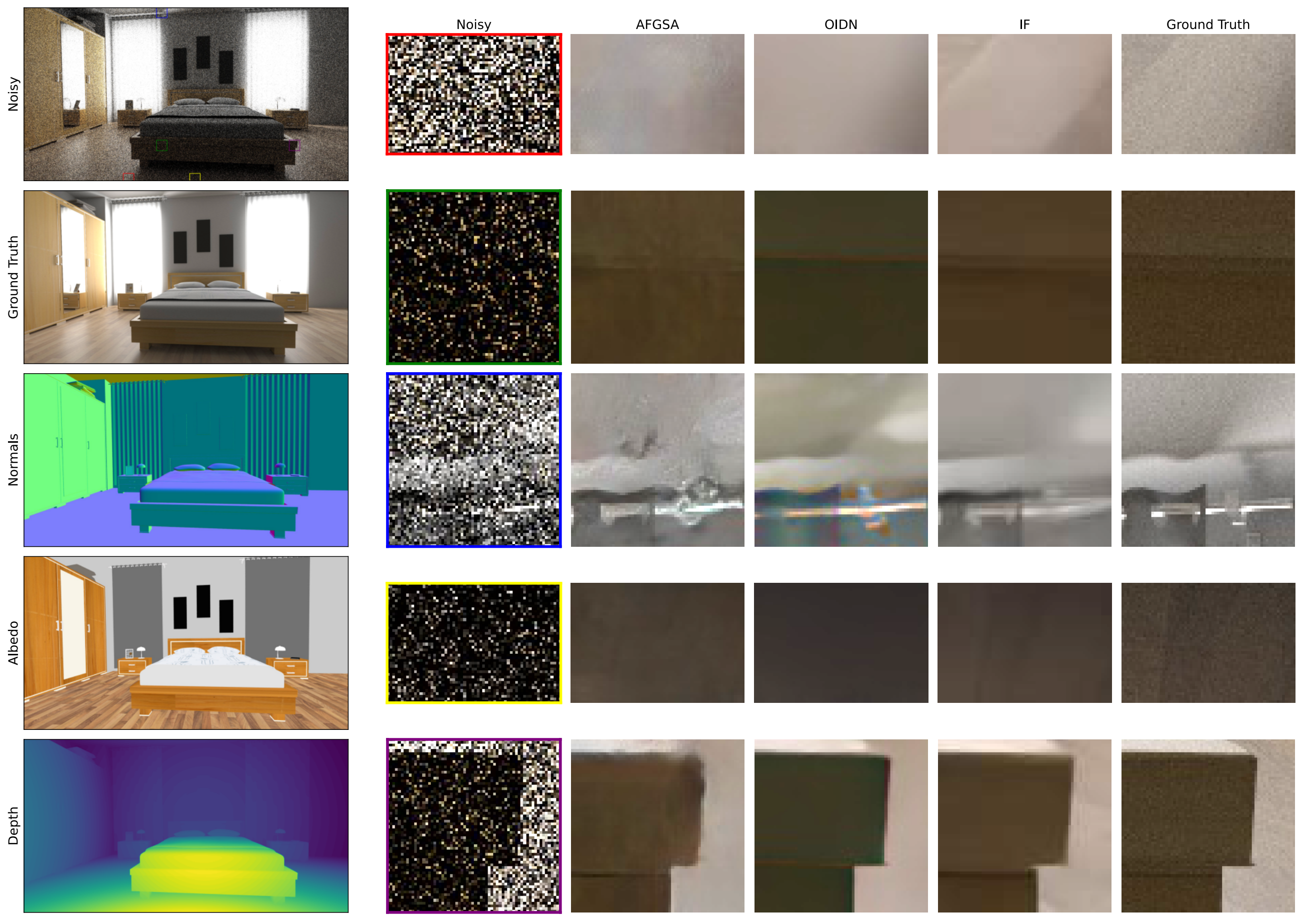

Figure 6. Additional qualitative results on the noisebase dataset. The first column shows the noisy radiance at 8 spp, reference, and auxiliary buffers. The noisy radiance has color-coded boxes of interest. **Columns 3 and 4** show results for competing methods. Observe how AFGSA generates noise because it uses an adversarial loss and the Ground Truth is a bit noisy. OIDN occasionally fails to reproduce color correctly. Our method (fifth column, **IF**) consistently produces reasonable results especially in undersampled regions.

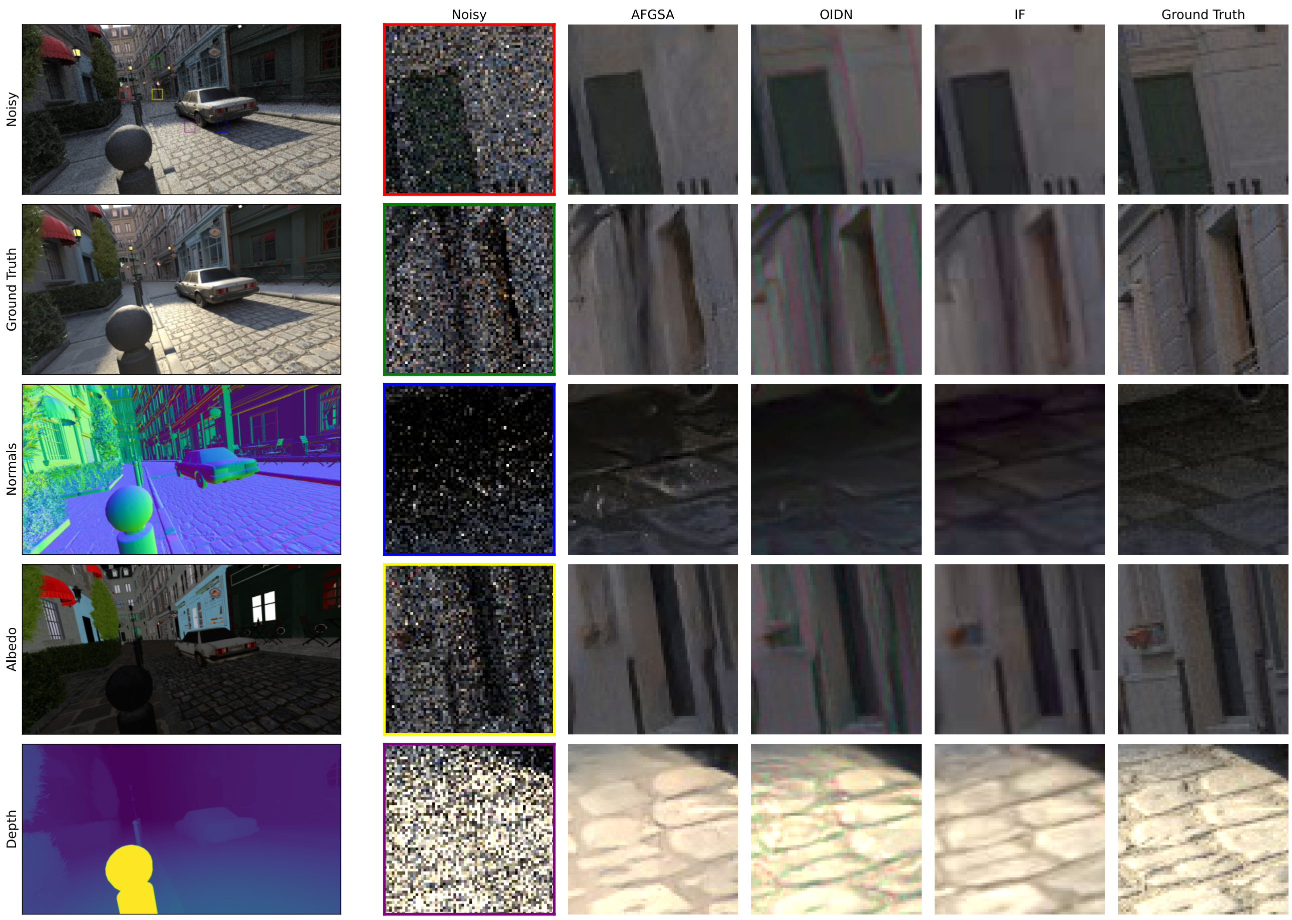

Figure 7. Additional qualitative results at 8 spp.

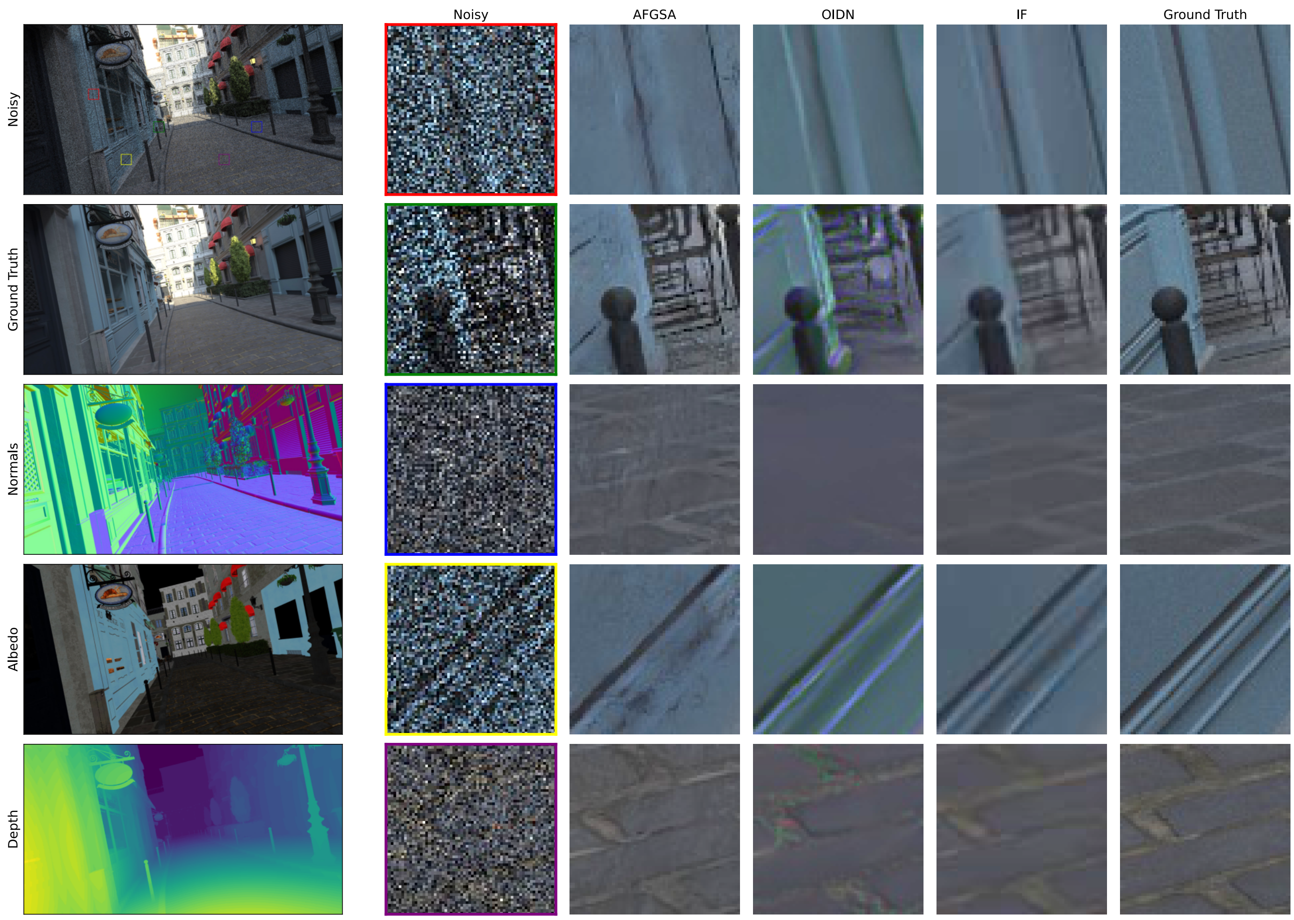

Figure 8. Additional qualitative results at 8 spp.

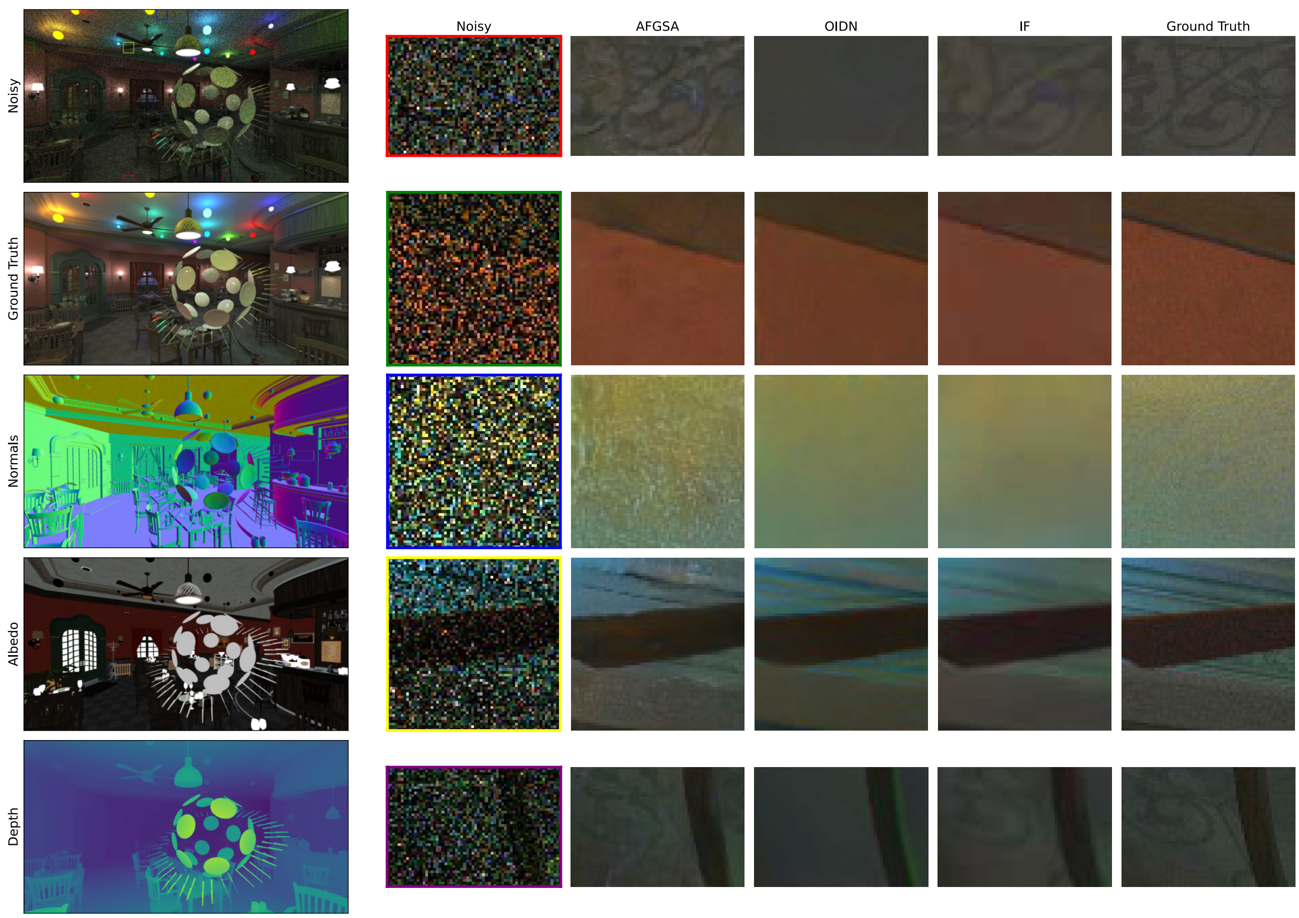

Figure 9. Additional qualitative results at 8 spp.

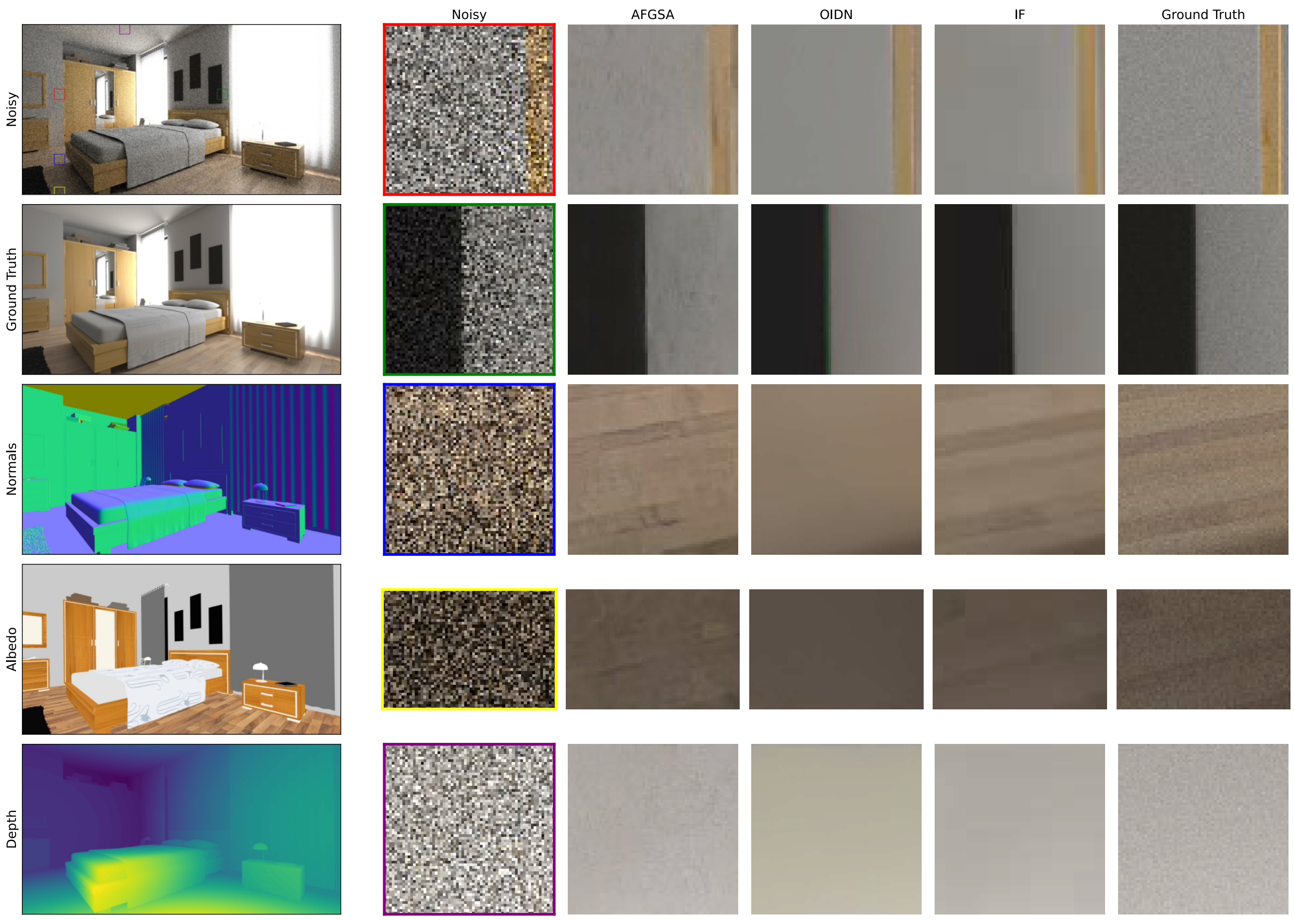

Figure 10. Additional qualitative results at 32 spp.

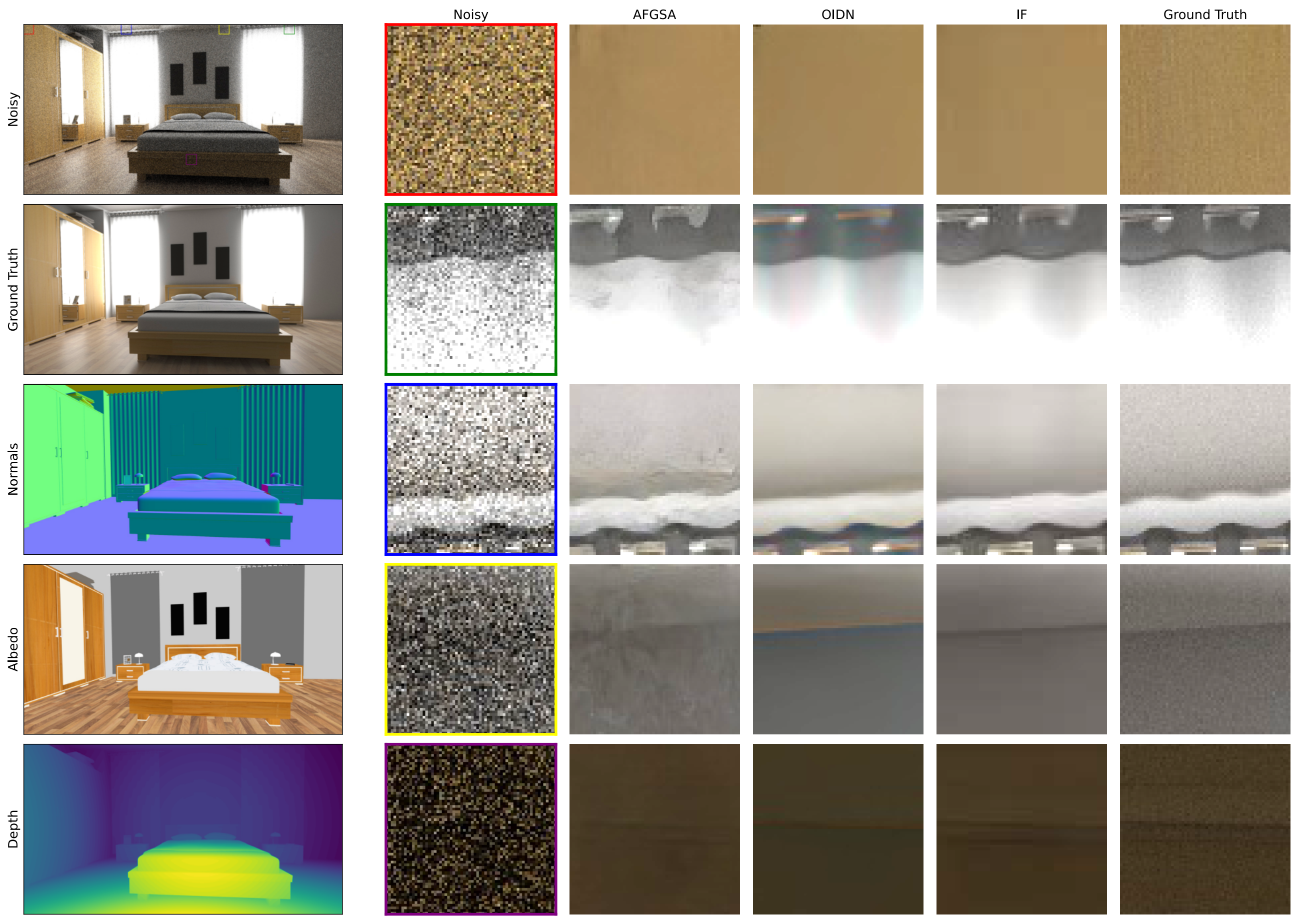

Figure 11. Additional qualitative results at 32 spp.

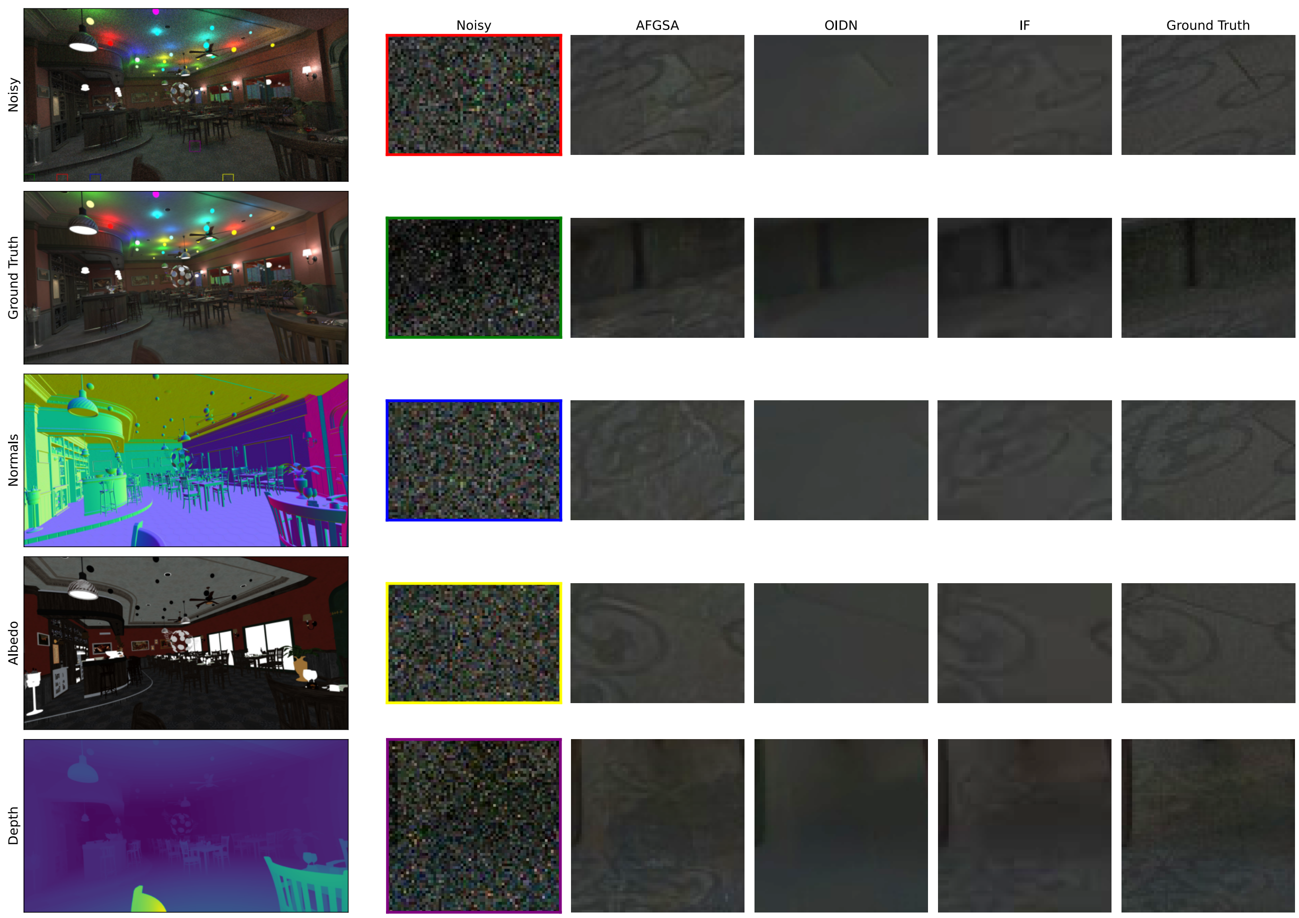

Figure 12. Additional qualitative results at 32 spp.

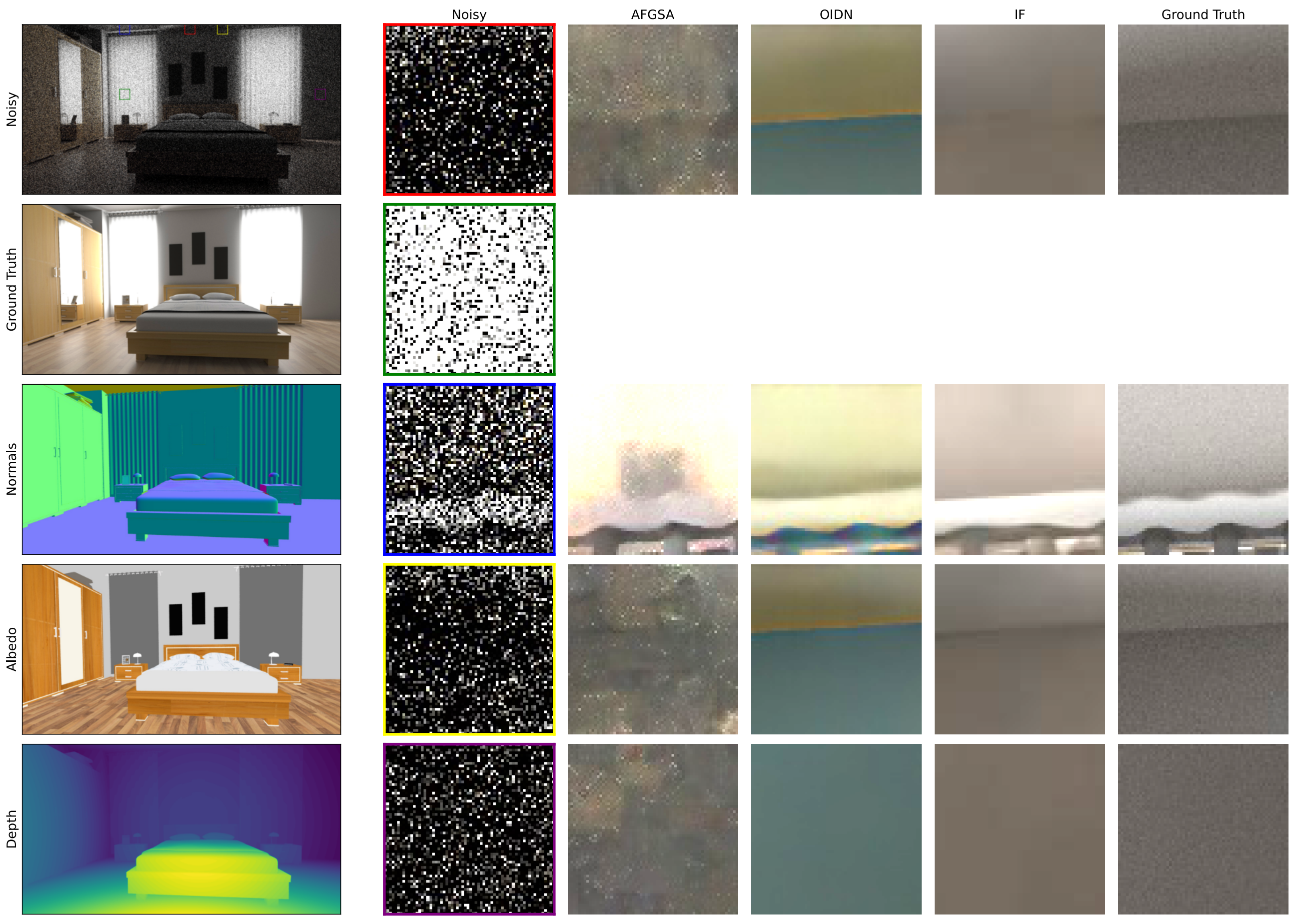

Figure 13. Additional qualitative results at 2 spp.

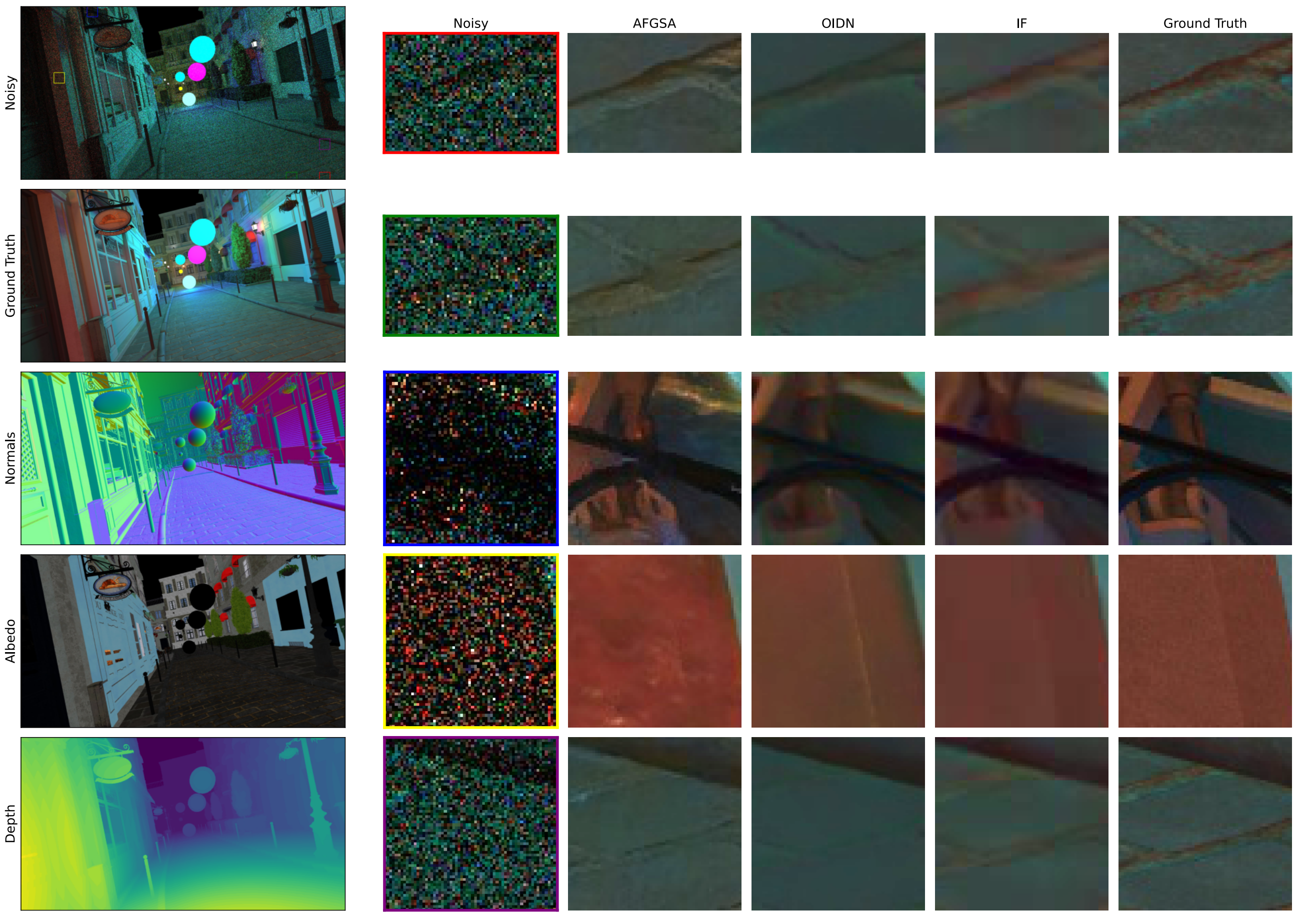

Figure 14. Additional qualitative results at 2 spp.

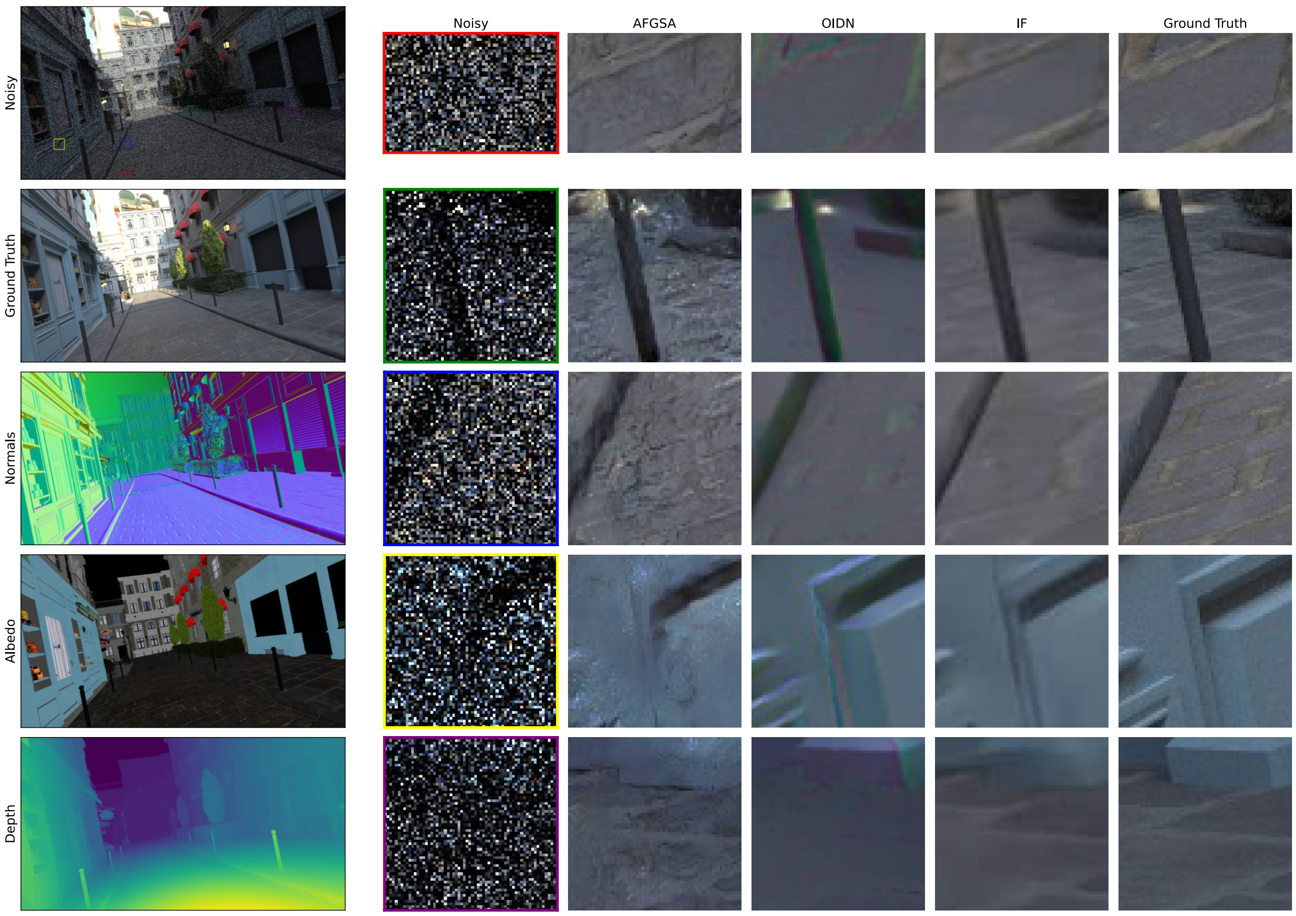

Figure 15. Additional qualitative results at 2 spp.

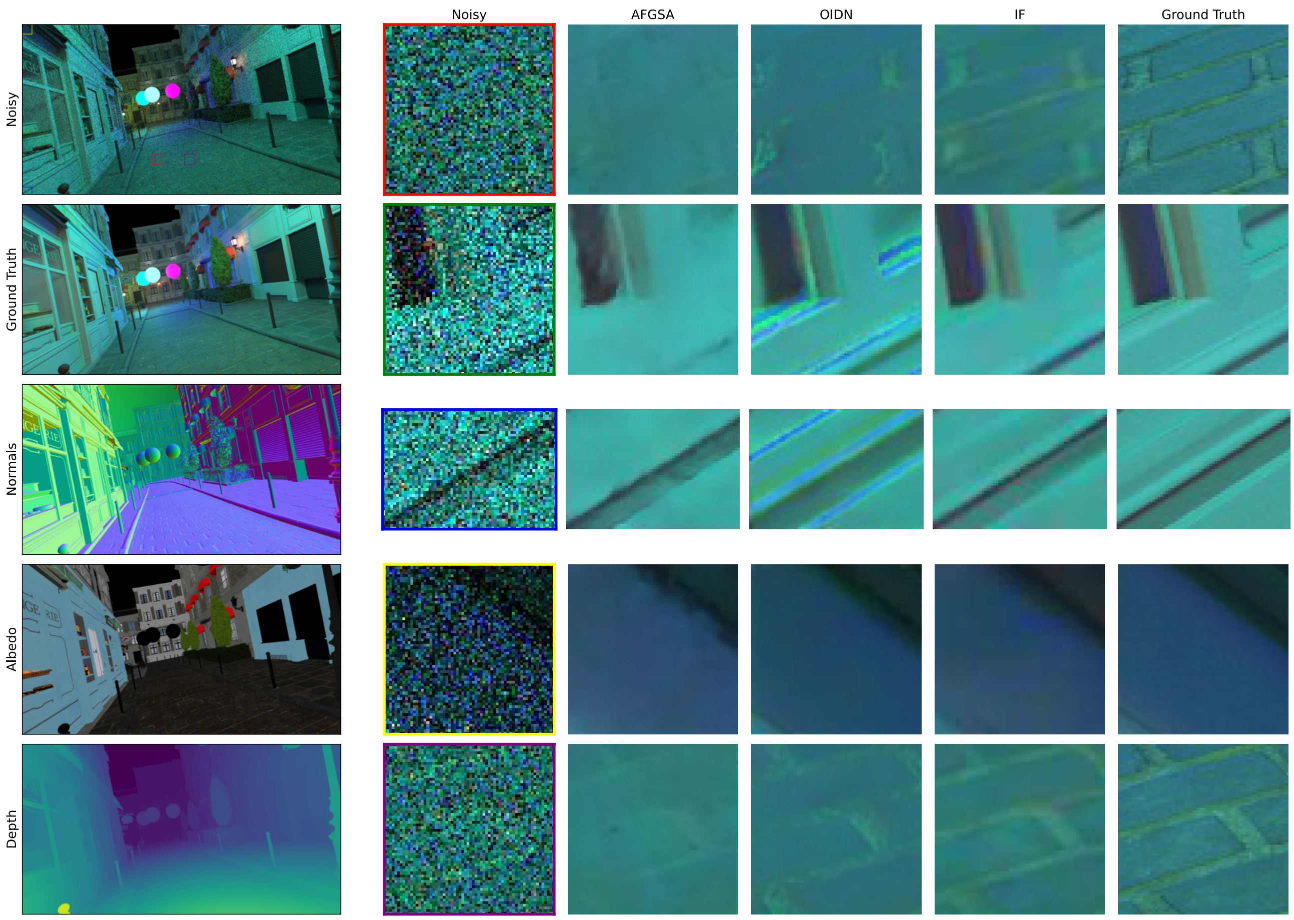

Figure 16. Additional qualitative results at 4 spp.

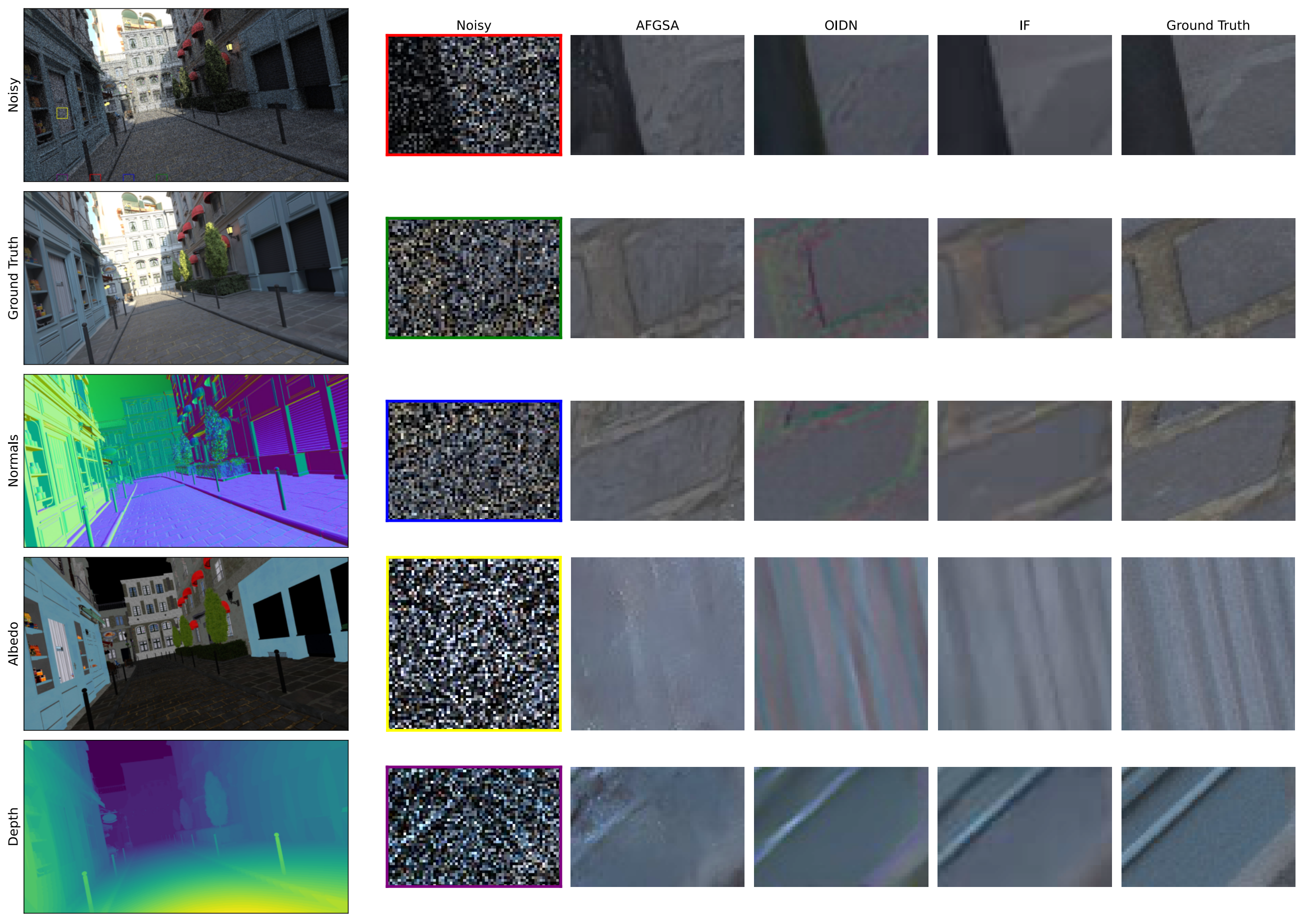

Figure 17. Additional qualitative results at 4 spp.

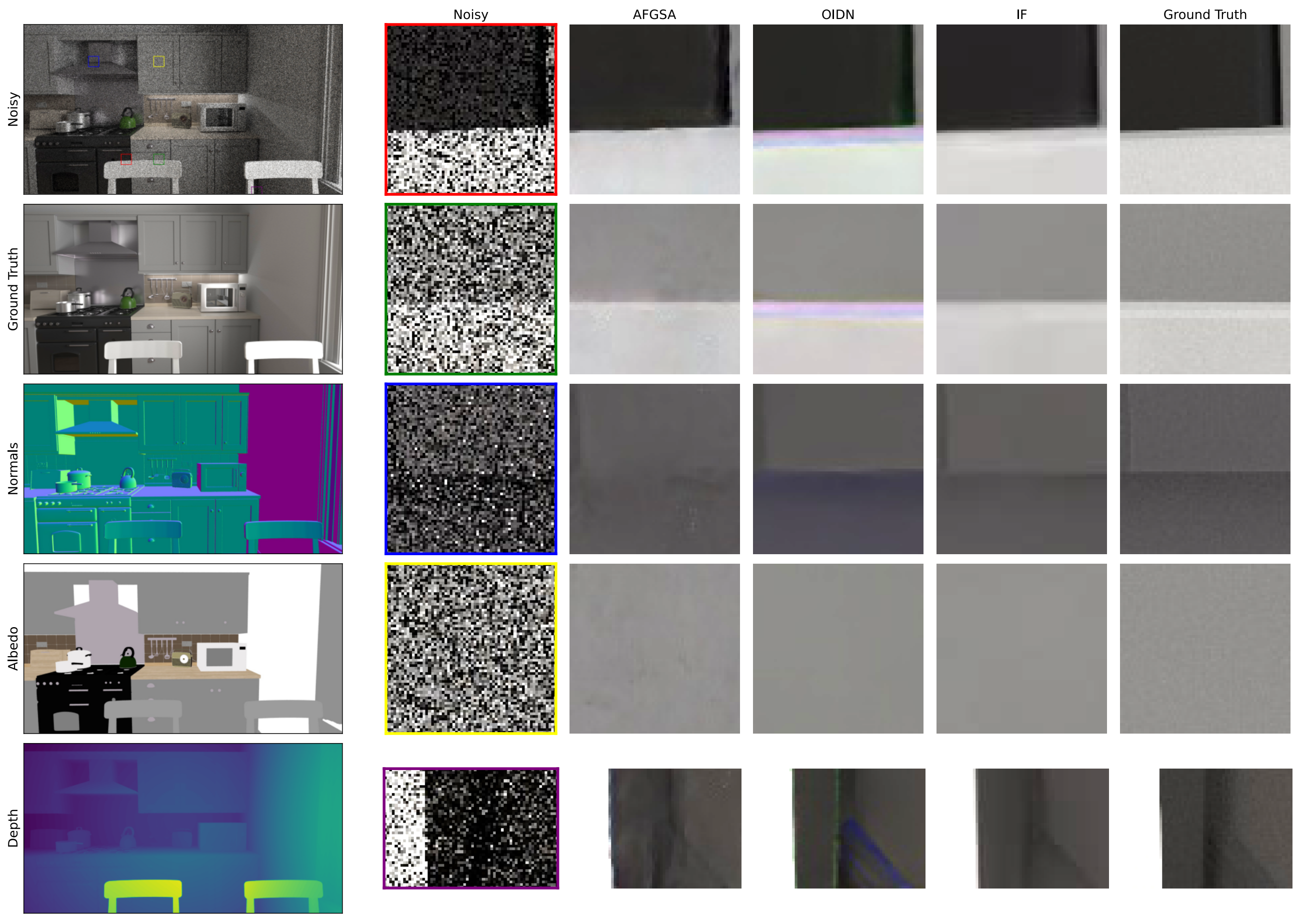

Figure 18. Additional qualitative results at 4 spp.