EUROGRAPHICS 2020 / U. Assarsson and D. Panozzo
(Guest Editors)

Volume 39 (2020), Number 2

Neural Temporal Adaptive Sampling and Denoising – Supplemental material

J. Hasselgren, J. Munkberg, M. Salvi, A. Patney and A. Lefohn

NVIDIA

1 Network Scaling

We have experimented with a smaller versions of the networks detailed in the paper, targeted at real-time applications with a runtime of about 13 ms per frame at 1920×1080. In our experience, image quality scales gracefully with network size. It is hard to claim an optimal design, and it rather depends on run-time constraints or a subjective notion of how much quality can be sacrificed for performance.

In particular, we note that the sampler network can be aggressively downscaled without noticeable impact on image quality. We believe this is due to our use of reprojected denoised history as a guide, which simplifies the sample map estimation problem. In constrast, DASR has to learn sample map estimation from much noisier data.

Our small sampler is shown in Figure 1 and the small denoiser in Figure 2. The networks are very similar to the ones described in the paper. The small denoiser has the same number of level, but with reduced feature counts for all layers of the U-net. The small adaptive sampler is a tiny U-net with just two hierarchical levels using eight features in all layers. Additionally, we switch from traditional concatenated skip connections [RFB15] to residual skip connections [HZRS15], to reduce data amplification and feature counts.

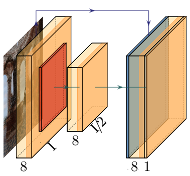

Figure 1: *Small sampler network.*

In Figure 3 we show PSNR and tPSNR scores for our small networks. Even though they are roughly 3× faster than the larger versions presented in the paper, they still have consistent high quality, and outperforms all previous work. This shows that our network approach is tuneable to multiple performance/cost ratios.

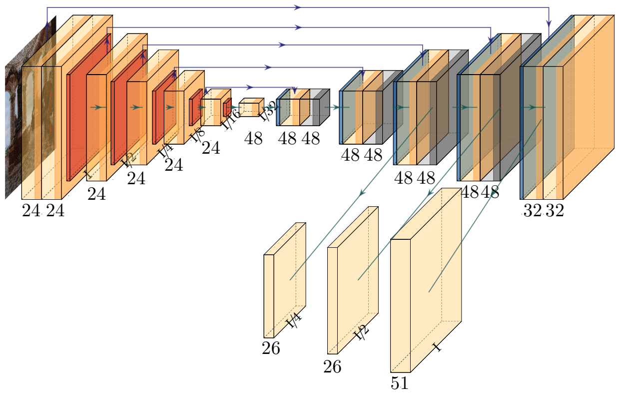

Figure 2: *Small denoiser network.*

2 SVGF

The SVGF algorithm [SKW*17] is designed for 1 spp using rasterized primary visibility. We generate input for SVGF using a rasterized geometry buffer with 1 spp and then traced four rays at the primary intersection. Therefore, we don't directly get spatial anti-aliasing. To study the effect of anti-aliasing on the error metrics, we also generate a version with the input at 1 spp with a single path at 2× resolution, applied SVGF at the large resolution and subsequently downscaled the denoised results with a 2×2 box filter. This improves anti-aliasing and the PSNR scores with roughly one decibel, but comes at a 4× higher runtime cost. We show PSNR and tPSNR charts over the SunTemple animation in Figure 5 and a visual comparison in Figure 4.

We show PSNR and tPSNR scores our algorithm, DASR and SVGF for the Bathroom animation in Figure 6 and for the Classroom animation in Figure 7.

3 Holdout Generalization Experiment

Figure 8 shows some example crops from the holdout generalization study described in Section 3.1 of the main paper. The loss is slightly lower for the specialized networks, but in the SunTemple

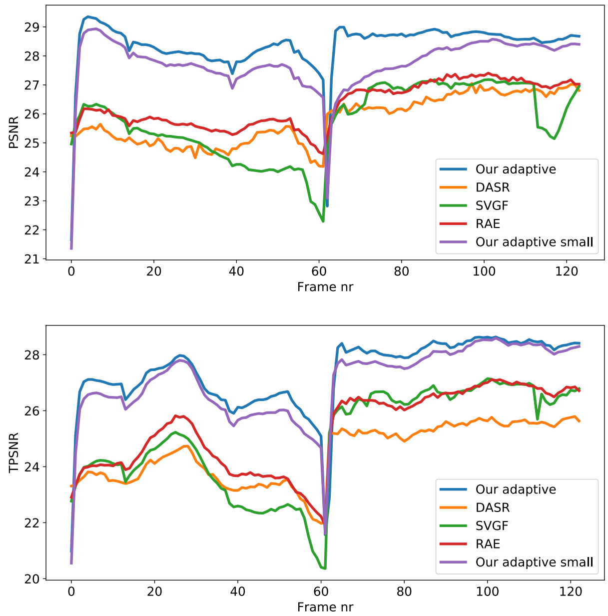

Figure 3: *PSNR and tPSNR scores for the* SunTemple *scene for two versions of our adaptive algorithm, SVGF, RAE and DASR.*

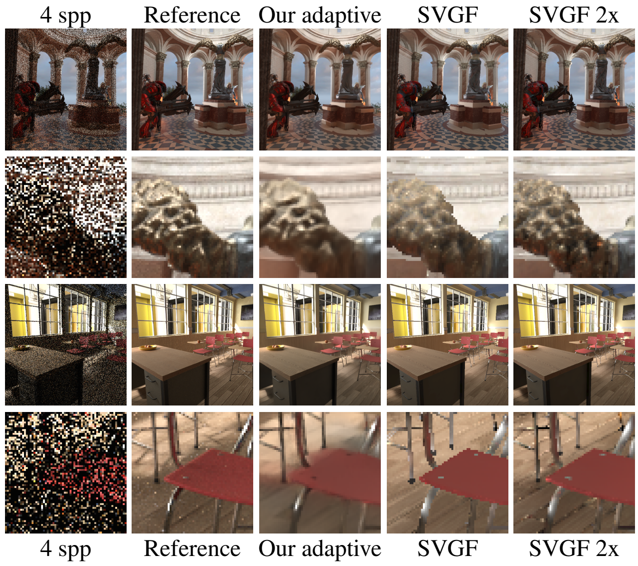

Figure 4: *Denoising quality from an average of 4 spp. These scenes were not part of the training set. We show comparison with SVGF (spatiotemporal variance guided filtering), and SVGF running at twice the resolution with one path per pixel, then downsampled (to get spatial anti-aliasing).*

and BistroIndoor scenes it is not clear if the specialized versions objectively look better. They tend to capture additional specular highlights or details, but at the cost of extra ghosting artifacts.

For the PinkRoom scene it is clear that the network not trained on the scene has a harder time to denoise the specular reflections, and leaves more noise in the final image. This is also consistent with the larger quality reduction for this scene. This is our only scene with smooth specular reflections, which may explain why the training is particularly sensitive to removing this scene.

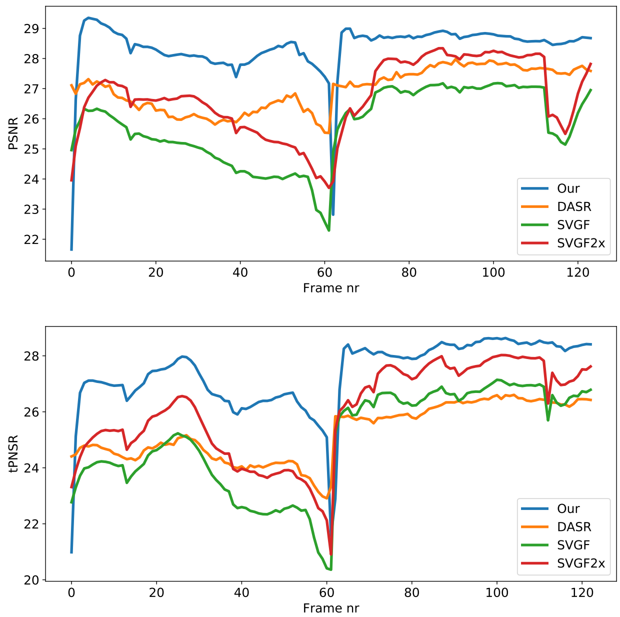

Figure 5: *PSNR and tPSNR scores for the* SunTemple *scene for our algorithm, SVGF and DASR.*

References

[HZRS15] He K., Zhang X., Ren S., Sun J.: Deep residual learning for image recognition. *CoRR abs/1512.03385* (2015). 1

[RFB15] Ronneberger O., Fischer P., Brox T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (2015), vol. 9351, pp. 234–241. 1

[SKW*17] Schied C., Kaplanyan A., Wyman C., Patney A., Chaitanya C. R. A., Burgess J., Liu S., Dachsbacher C., Lefohn A., Salvi M.: Spatiotemporal Variance-guided Filtering: Real-time Reconstruction for Path-traced Global Illumination. In *Proceedings of High Performance Graphics* (2017), HPG '17, pp. 2:1–2:12. 1

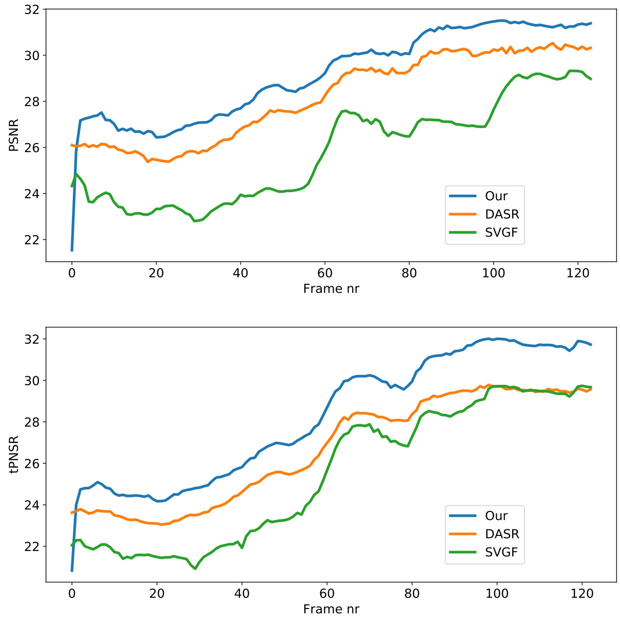

Figure 6: *PSNR and tPSNR scores for the* BATHROOM *scene for our algorithm, SVGF and DASR.*

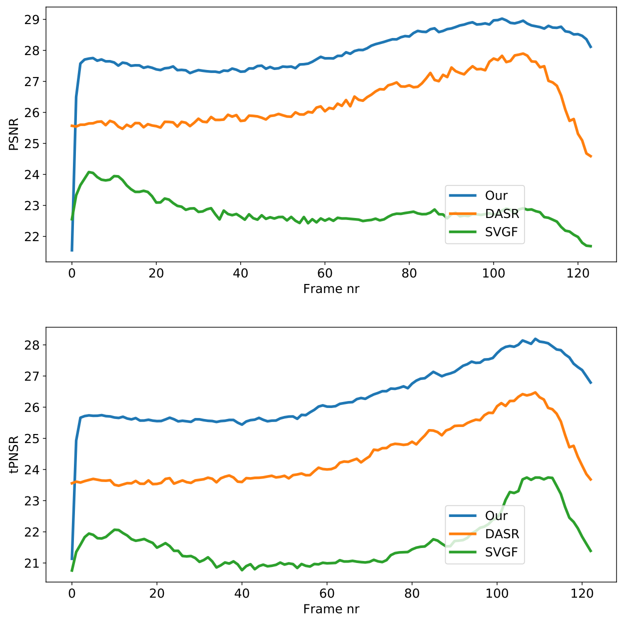

Figure 7: *PSNR and tPSNR scores for the* CLASSROOM *scene for our algorithm, SVGF and DASR.*

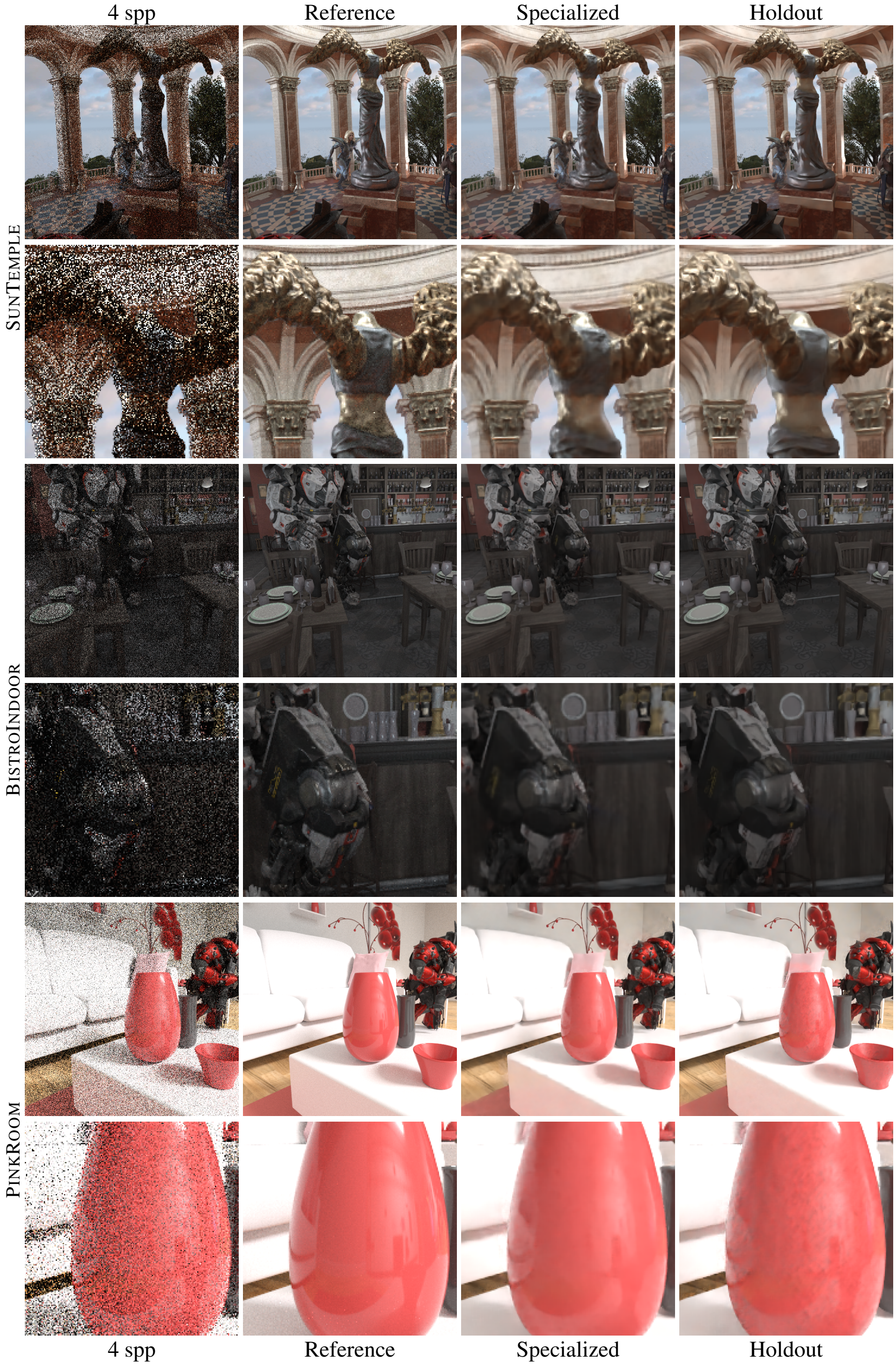

Figure 8: *Visual comparison for the holdout study. Specialized denotes a network only trained on the same scene. Holdout is a network trained on nine scenes, not including the scene showed in this comparison.*