

Log-linear Kernel Cluster Distillation for Lightweight Neural Monte Carlo Denoising: Supplementary Material

YE YUAN, ZIJUAN YUAN, JUNYUAN HE, ZIREN WANG, CHANGYUN ZHI, CHENXI ZHOU, BOSHAN LIU, YANWEN GUO, and JIE GUO^{✉*}, State Key Lab for Novel Software Technology, Nanjing University, China

ACM Reference Format:

Ye Yuan, Zijuan Yuan, Junyuan He, Ziren Wang, Changyun Zhi, Chenxi Zhou, Boshan Liu, Yanwen Guo, and Jie Guo. 2026. Log-linear Kernel Cluster Distillation for Lightweight Neural Monte Carlo Denoising: Supplementary Material. In *SIGGRAPH Asia 2026 Conference Papers (SA Conference Papers '26)*, December 01–04, 2026, Kuala Lumpur, Malaysia. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3829340.3842233>

A Network Backbone

Our student and the matched baseline NPPD (Small) share nearly identical backbones and parameter counts (8.8M vs. 8.9M): we obtain NPPD (Small) by replacing the U-Net backbone of NPPD [Balint et al. 2023] with the reduced-capacity version shown below, and the student reuses the same backbone and training schedule. The only difference is the prediction head: NPPD (Small) predicts dense pyramid-kernel coefficients, whereas our student predicts compact mixture weights over archetypes and residual offsets. The shared backbone is

c64c64d → c64c96d → c96c128d → c128c192d → c192c256d
→ c512c512c256u → c256c192u → c192c128u → c128c96u
→ c96c64u → c64c64.

B Archetypal-Analysis Details

We fit archetypal analysis independently for the spatial, temporal, and upsampling kernel families. For each fit, we initialize archetypes with farthest-first selected teacher kernels and then alternate between projected-gradient updates of sample mixture coefficients on the probability simplex and Frank–Wolfe updates of archetype coefficients. The latter keeps every archetype in the convex hull of the input samples for the corresponding fit (teacher kernels in Stage 1 and residual samples in Stage 2). Optimization stops when the relative residual-sum-of-squares improvement is below 10^{-6} or the iteration limit is reached.

We collect teacher kernels from 128 training scenes (16 frames per scene at 256p), yielding 320M spatial, 80M temporal, and 250M

*Corresponding author.

Authors' Contact Information: Ye Yuan, 502024330061@smail.nju.edu.cn; Zijuan Yuan, 522025320206@smail.nju.edu.cn; Junyuan He, 221240065@smail.nju.edu.cn; Ziren Wang, 221220149@smail.nju.edu.cn; Changyun Zhi, 241220044@smail.nju.edu.cn; Chenxi Zhou, chenxizhou@smail.nju.edu.cn; Boshan Liu, 231220035@smail.nju.edu.cn; Yanwen Guo, ywguo@nju.edu.cn; Jie Guo, guojie@nju.edu.cn, State Key Lab for Novel Software Technology, Nanjing University, Nanjing, China.



This work is licensed under a Creative Commons Attribution 4.0 International License. SA Conference Papers '26, Kuala Lumpur, Malaysia
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2842-6/2026/12
<https://doi.org/10.1145/3829340.3842233>

upsampling samples. Samples from all pyramid levels of the same family are pooled for the corresponding fit.

We use (8, 4) archetype/residual counts for spatial kernels and (4, 4) for both temporal and upsampling kernels. The student first keeps these AA-generated components frozen for 10 epochs, then fine-tunes them jointly with the backbone at one tenth of the backbone learning rate. After this fine-tuning stage, the archetypes and residual offsets no longer satisfy the convex-hull constraint; the result is an AA-initialized learned dictionary.

C Generalization Across Sample Counts

To verify the generalization of our representation to different input sample counts, we additionally conduct experiments beyond the 1-spp evaluation reported in the main paper: we retrain all models at 2 spp and 4 spp using the same architectures, training schedule, and losses as their 1-spp counterparts, and evaluate them on the same five scenes. Table 1 reports the per-scene PSNR and SSIM at 2 spp, and Table 2 reports the same metrics at 4 spp; in both cases our student remains close to the 15M teacher and is comparable to or better than the matched NPPD (Small) baseline.

D Analysis of the Compact Parameterization

We investigate the hyperparameters of the compact kernel representation, in particular the numbers of archetypes and residual offsets. All results are averaged over the five test scenes used in the main paper. *Ours* denotes the default 8M student with 8 spatial archetypes and 4 spatial residual offsets.

Spatial-kernel capacity. Table 3 changes only the numbers of spatial-kernel archetypes and residual offsets. The configurations are chosen to separate the effects of coarse archetype capacity and residual capacity.

Table 1. Scene-wise quality on 2-spp inputs. Learning-based models are trained at 2 spp under the same protocol as in the main paper; OIDN and OptiX are pre-trained production denoisers included as external references. **Best** and **second-best** values per scene and metric are highlighted.

PSNR (dB)						
Scene	Ours	Teacher	NPPD (Small)	FAN	OIDN	OptiX
Bistro	30.0665	30.3236	29.6555	28.8715	28.22	29.81
Bedroom	29.1623	29.2365	29.0789	27.7678	28.48	28.86
Zeroday	25.1668	25.4657	25.1010	24.3117	24.61	24.44
Dining Room	31.0617	31.4531	30.0370	29.6817	30.76	29.62
Kitchen	29.5032	29.8748	29.3471	28.2674	29.19	29.94
SSIM						
Scene	Ours	Teacher	NPPD (Small)	FAN	OIDN	OptiX
Bistro	0.8724	0.8769	0.8613	0.8413	0.8219	0.8618
Bedroom	0.8104	0.8114	0.8106	0.7984	0.8027	0.8086
Zeroday	0.7699	0.7758	0.7722	0.7511	0.7525	0.7584
Dining Room	0.9483	0.9491	0.9491	0.9437	0.9460	0.9536
Kitchen	0.8560	0.8597	0.8557	0.8459	0.8470	0.8619

Table 2. Scene-wise quality on 4-spp inputs. Learning-based models are trained at 4 spp under the same protocol as in the main paper; OIDN and OptiX are pre-trained production denoisers included as external references. **Best** and **second-best** values per scene and metric are highlighted.

PSNR (dB)						
Scene	Ours	Teacher	NPPD (Small)	FAN	OIDN	OptiX
Bistro	30.52	31.04	29.87	30.59	29.28	30.58
Bedroom	30.09	30.00	29.47	29.80	29.66	29.73
Zeroday	26.04	26.12	25.72	25.77	25.66	25.53
Dining Room	32.23	32.39	31.83	32.13	32.19	30.37
Kitchen	30.50	30.64	29.79	30.11	30.49	30.99
SSIM						
Scene	Ours	Teacher	NPPD (Small)	FAN	OIDN	OptiX
Bistro	0.8776	0.8896	0.8612	0.8779	0.8471	0.8794
Bedroom	0.8173	0.8173	0.8066	0.8116	0.8128	0.8172
Zeroday	0.7888	0.7890	0.7714	0.7784	0.7746	0.7859
Dining Room	0.9547	0.9550	0.9504	0.9491	0.9547	0.9590
Kitchen	0.8629	0.8651	0.8542	0.8526	0.8595	0.8717

Table 3. Spatial-kernel capacity and matched component ablations. Here, M_a and M_δ are the spatial archetype and residual counts, and “Spatial ch.” is the corresponding number of per-pixel spatial output channels. Only the spatial counts change in the archetype/residual-capacity variation; all other architecture and training settings are fixed. Results are averaged across the five test scenes.

Variant	M_a	M_δ	Spatial ch.	PSNR	SSIM
<i>Archetype/residual capacity</i>					
$c1/f8$	1	8	9	27.00	0.823
$c2/f8$	2	8	10	27.16	0.825
$c4/f4$	4	4	8	27.14	0.825
$c4/f8$	4	8	12	28.08	0.843
$c8/f8$	8	8	16	28.26	0.843
$c8/f4$ (Ours)	8	4	12	28.25	0.843
<i>Component and output controls</i>					
Residual offsets only	0	12	12	27.46	0.836
NPPD Small + same distillation	–	–	25	27.88	0.839

The capacity results show that increasing archetype capacity improves quality at the low end, while increasing residual capacity from (4, 4) to (4, 8) further improves quality once a moderate coarse basis is available. We use (8, 4), which is within 0.01 dB of (8, 8) at a smaller spatial mixture dimension.

Role of archetypes and residual offsets. At a fixed 12-channel spatial budget, the residual-only variant reaches 27.46 dB and 0.836 SSIM, indicating that residual offsets cannot replace the coarse archetype basis. The complementary no-residual ablation, including its SSIM, is reported in the main paper.

Representation vs. distillation. The matched NPPD (Small) control—identical backbone, supervision, and training schedule, differing only in the output parameterization—falls short of our student at 27.88 dB and 0.839 SSIM. The gain is attributable to the compact parameterization rather than to distillation.

Dictionary initialization. All variants share the student backbone, per-family archetype/residual-offset counts, output dimension, distillation losses, and two-phase training schedule; only the offline construction of the archetype and residual-offset dictionaries differs.

Table 4. Dictionary-initialization comparison under the unified training protocol. Values are average PSNR and SSIM over the five test scenes. **Best** and **second-best** values are highlighted.

Initialization	Avg. PSNR	Avg. SSIM
NMF initialization	26.90	0.8141
VQ-codebook initialization	27.64	0.8294
AA initialization (Ours)	28.25	0.843

Two-stage AA is the strongest of the tested constructions, consistent with AA preserving the simplex structure during offline fitting; the main paper provides an additional qualitative k -means comparison.

E Temporal Stability

PSNR and SSIM are computed per frame and therefore do not penalize temporal artifacts such as flicker or ghosting. We additionally measure temporal stability with FovVideoVDP [Mantiuk et al. 2021], which scores a full sequence against its reference under a display and viewing model (JOD scale; 10 denotes no visible difference). All methods are evaluated on the animated test sequences after a 16-frame warm-up, with identical settings against the same 4096-spp references. Table 5 reports the scene-wise scores: on average, our student improves over the matched NPPD (Small) baseline and remains close to the NPPD teacher.

References

- Martin Balint, Krzysztof Wolski, Karol Myszkowski, Hans-Peter Seidel, and Radosław Mantiuk. 2023. Neural Partitioning Pyramids for Denoising Monte Carlo Renderings. In *ACM SIGGRAPH 2023 Conference Proceedings*. ACM, New York, NY, USA, Article 60, 11 pages. doi:10.1145/3588432.3591562
- Rafal K. Mantiuk, Gyorgy Denes, Alexandre Chapiro, Anton Kaplanyan, Gizem Rufo, Romain Bachy, Trisha Lian, and Anjul Patney. 2021. FovVideoVDP: A Visible Difference Predictor for Wide Field-of-View Video. *ACM Trans. Graph.* 40, 4 (jul 2021), 19 pages.

Table 5. Scene-wise temporal quality measured by FovVideoVDP on the JOD scale. Higher is better. The Average row is the arithmetic mean of the five scene scores.

Scene	Ours	NPPD (Small)	FAN	OIDN	NPPD (Teacher)
Bistro	7.5030	7.3725	7.4640	6.8020	7.5596
Zeroday	7.2186	7.0911	7.1570	6.8280	7.2094
Dining Room	8.0223	8.1562	7.6100	8.1770	8.3170
Bedroom	7.9486	7.9166	8.0920	7.5840	7.9856
Kitchen	7.7028	7.6663	8.0090	7.3900	7.7086
Average	7.6791	7.6405	7.6664	7.3562	7.7560