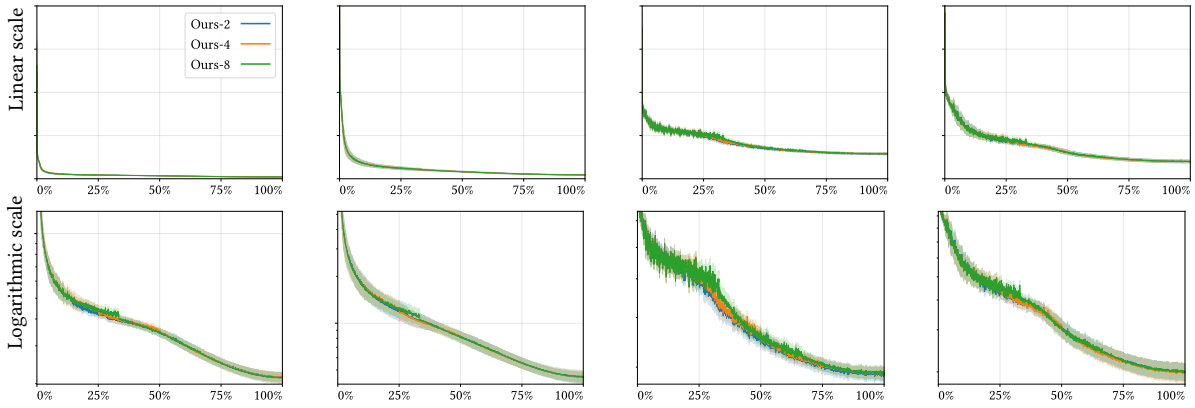
Taming optimization variance in compact neural shading networks: Supplemental Material

BENEDIKT BITTERLI, NVIDIA, USA
PETRIK CLARBERG, NVIDIA, Sweden
CHRIS CUMMINGS, NVIDIA, United Kingdom
AARON LEFOHN, NVIDIA, USA
STEVE MARSCHNER, Cornell University, USA and NVIDIA, USA
JAN NOVÁK, NVIDIA, Czech Republic
FABRICE ROUSSELLE, NVIDIA, Switzerland
ANDREA WEIDLICH, NVIDIA, Canada
TIZIAN ZELTNER, NVIDIA, Switzerland

A Choice of pruning schedule

Figure 1 shows results using pruning factors 2×, 4× and 8×. The pruning factor has negligible impact in our tests.

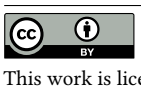
	Num. of instances K	Batch size B
Ours-2	64 → 32 → 16 → 8 → 4 → 2 → 1	1k → 2k → 4k → 8k → 16k → 32k → 64k
Ours-4	64 → 16 → 4 → 1	1k → 4k → 16k → 64k
Ours-8	64 → 8 → 1	1k → 8k → 64k



(a) Oxydized metal (b) Brushed brass (c) Gold and ceramic (d) Glazed ceramic

Fig. 1. Comparison of pruning factors 2×, 4× (used for all paper results) and 8× in a setup identical to Figure 5 in the paper.

Authors' Contact Information: Benedikt Bitterli, NVIDIA, USA; Petrik Clarberg, NVIDIA, Sweden; Chris Cummings, NVIDIA, United Kingdom; Aaron Lefohn, NVIDIA, USA; Steve Marschner, Cornell University, USA and NVIDIA, USA; Jan Novák, NVIDIA, Czech Republic; Fabrice Rousselle, NVIDIA, Switzerland; Andrea Weidlich, NVIDIA, Canada; Tizian Zeltner, NVIDIA, Switzerland.



This work is licensed under a Creative Commons Attribution 4.0 International License.

B Shared data generation

In a multi-instance training pipeline, one can choose to share training batches between instances or to have independent batches for all instances. Sharing batches offers two benefits: it reduces the compute and memory overhead of data generation, and it facilitates the ranking of instances as their training losses are correlated. On the other hand, independent batches could lead to better performance as they aid exploration. To evaluate the impact of sharing data, which we used in all experiments in the main paper, we ran an additional experiment without data sharing. The results are shown in Table 1 and Table 2. As expected, the ranking is slightly less accurate when data is not shared (72.4% without sharing, 79.1% with sharing), however, the best loss is similar in both cases (0.081 with and without data sharing), suggesting that there is no significant downside to sharing data between instances.

Table 1. Relative frequency that the surviving instance (selected by culling) is among the best performing instances. The first column shows the probability that it is the best instance (rank 1), the next columns show the probability that the surviving instance is among the top 3 (top 5%), top 6 (top 10%), and top 12 (top 20%) of all instances.

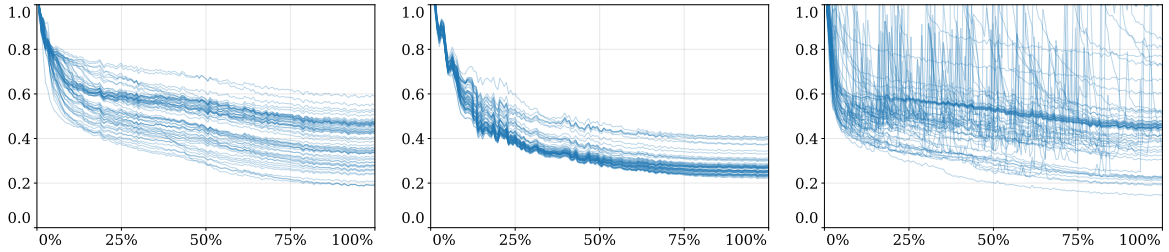
Material	Best (rank 1)	Top 5% (top 3)	Top 10% (top 6)	Top 20% (top 12)
da Vinci Workshop materials				
Birdcage	86.7%	100.0%	100.0%	100.0%
Chandelier	53.3%	93.3%	93.3%	100.0%
Hammer	80.0%	93.3%	100.0%	100.0%
Lantern	86.7%	100.0%	100.0%	100.0%
Mirror	60.0%	93.3%	100.0%	100.0%
Palette	73.3%	93.3%	100.0%	100.0%
Chair	73.3%	86.7%	93.3%	100.0%
Scales	93.3%	100.0%	100.0%	100.0%
Table	53.3%	86.7%	100.0%	100.0%
Layered materials				
Scratched steel	73.3%	100.0%	100.0%	100.0%
Bumpy plastic	86.7%	86.7%	100.0%	100.0%
Oxydized metal	73.3%	93.3%	100.0%	100.0%
Gold and ceramic	46.7%	93.3%	100.0%	100.0%
Brushed brass	66.7%	93.3%	100.0%	100.0%
Glazed ceramic	80.0%	100.0%	100.0%	100.0%
Average	72.4%	94.2%	99.1%	100.0%

Table 2. We disable culling, and merely simulate it to identify the instance that would survive. We report the final loss of the best instance across all 64 instances (**2nd column**) and the loss of *our* surviving instance (normalized by the value of the best) (**3rd column**). The last column shows the same ratio but for a *randomly* picked instance.

Material	Best loss $\mathcal{L}_b$	Our inst. $\mathcal{L}_o/\mathcal{L}_b$	Random inst. $\mathcal{L}_r/\mathcal{L}_b$
DA VINCI WORKSHOP MATERIALS			
Birdcage	0.055 ± 0.004	1.004 ± 0.011	1.832 ± 0.132
Chandelier	0.039 ± 0.002	1.019 ± 0.031	1.765 ± 0.103
Hammer	0.032 ± 0.003	1.007 ± 0.015	1.669 ± 0.133
Lantern	0.022 ± 0.002	1.009 ± 0.031	1.807 ± 0.162
Mirror	0.081 ± 0.004	1.007 ± 0.012	1.826 ± 0.107
Palette	0.257 ± 0.011	1.009 ± 0.019	1.601 ± 0.078
Chair	0.045 ± 0.007	1.038 ± 0.124	3.243 ± 0.704
Scales	0.048 ± 0.002	1.001 ± 0.006	1.649 ± 0.061
Table	0.019 ± 0.001	1.066 ± 0.082	1.569 ± 0.078
LAYERED MATERIALS			
Scratched steel	0.056 ± 0.007	1.006 ± 0.016	2.513 ± 0.317
Bumpy plastic	0.018 ± 0.001	1.005 ± 0.014	1.360 ± 0.069
Oxydized metal	0.021 ± 0.001	1.003 ± 0.007	1.445 ± 0.082
Gold and ceramic	0.281 ± 0.007	1.015 ± 0.018	1.458 ± 0.065
Brushed brass	0.045 ± 0.004	1.021 ± 0.038	3.651 ± 0.242
Glazed ceramic	0.197 ± 0.015	1.010 ± 0.031	1.911 ± 0.109
Average	0.081 ± 0.088	1.015 ± 0.017	1.953 ± 0.668

C Optimizers

In Figure 2, we show additional results obtained from training with different optimizers. We train the same configuration as Fig. 1 (c) in the main paper, using three different optimizers: Adam [Kingma and Ba 2014], Muon [Jordan et al. 2024] and SOAP [Vyas et al. 2025]. Although switching optimizers has an impact on the convergence rate, the spread of loss trajectories remains similar. All three optimizers would benefit from our multi-instance training scheme. We use Adam for all results in the main paper.



(a) Adam (b) Muon (c) SOAP

Fig. 2. Comparison of optimizers on neural material appearance fitting. Each column shows training loss of 64 differently initialized networks. All three optimizers show significant variance in training outcomes.

D Additional improvements for neural materials

We present a quantitative evaluation of the three neural material model improvements from the main paper. We measure how each one affects reconstruction quality across the material set. Each table group isolates one design decision while keeping all other settings fixed. Table 3 compares loss function mappings. Table 4 and Table 5 evaluate activation functions. Table 6 and Table 7 examine decoder input parameterizations.

We report HDR ꟻLIP errors [Andersson et al. 2021] to assess final reconstruction quality and training loss to reflect optimization behavior. The activation and parameterization comparisons report both metrics, while the loss-mapping comparison in Table 3 reports only the HDR ꟻLIP error. The underlying training losses differ by design and are not directly comparable. We compute ꟻLIP on a shared test scene with a material test sphere and a plane. All values are averaged over five training runs with different initializations. The best-performing configuration in each row is highlighted in bold.

Table 3. HDR ꟻLIP error mean and standard deviation for the discussed loss mappings.

Material	Logarithmic mapping	Power mapping
da Vinci workshop materials		
Birdcage	0.0541 ± 0.0097	**0.0370** ± 0.0029
Chandelier	0.0537 ± 0.0045	**0.0345** ± 0.0013
Hammer	0.0862 ± 0.0128	**0.0412** ± 0.0024
Lantern	0.0608 ± 0.0016	**0.0346** ± 0.0007
Mirror	0.1196 ± 0.0121	**0.0810** ± 0.0039
Palette	0.1134 ± 0.0135	**0.0991** ± 0.0144
Chair	0.1265 ± 0.0096	**0.0776** ± 0.0134
Scales	0.1379 ± 0.0154	**0.0652** ± 0.0012
Table	0.0707 ± 0.0059	**0.0425** ± 0.0024
Layered materials		
Glazed ceramic	0.1224 ± 0.0138	**0.1063** ± 0.0081
Brushed brass	0.0541 ± 0.0014	**0.0404** ± 0.0028
Scratched steel	0.0468 ± 0.0011	**0.0353** ± 0.0012
Bumpy plastic	0.0524 ± 0.0025	**0.0391** ± 0.0006
Oxydized metal	0.0536 ± 0.0038	**0.0400** ± 0.0012
Gold and ceramic	0.1177 ± 0.0113	**0.1014** ± 0.0100
Average	0.0847 ± 0.0340	**0.0583** ± 0.0273

Table 4. HDR ꟻLIP error mean and standard deviation for the discussed activation functions.

Material	ReLU	LeakyReLU	LeakySmeLU
da Vinci workshop materials			
Birdcage	0.0417 ± 0.0041	0.0407 ± 0.0026	**0.0370** ± 0.0029
Chandelier	0.0364 ± 0.0017	0.0376 ± 0.0034	**0.0345** ± 0.0013
Hammer	0.0421 ± 0.0026	0.0419 ± 0.0016	**0.0412** ± 0.0024
Lantern	0.0379 ± 0.0027	0.0361 ± 0.0010	**0.0346** ± 0.0007
Mirror	0.0847 ± 0.0012	**0.0803** ± 0.0026	0.0810 ± 0.0039
Palette	0.0990 ± 0.0210	**0.0929** ± 0.0080	0.0991 ± 0.0144
Chair	0.0865 ± 0.0056	**0.0756** ± 0.0127	0.0776 ± 0.0134
Scales	0.0672 ± 0.0023	0.0671 ± 0.0033	**0.0652** ± 0.0012
Table	0.0441 ± 0.0027	0.0438 ± 0.0007	**0.0425** ± 0.0024
Layered materials			
Glazed ceramic	0.0865 ± 0.0037	**0.0832** ± 0.0027	0.1063 ± 0.0081
Brushed brass	0.0424 ± 0.0034	0.0415 ± 0.0031	**0.0404** ± 0.0028
Scratched steel	0.0373 ± 0.0018	0.0373 ± 0.0019	**0.0353** ± 0.0012
Bumpy plastic	0.0398 ± 0.0004	0.0400 ± 0.0009	**0.0391** ± 0.0006
Oxydized metal	0.0438 ± 0.0032	0.0432 ± 0.0017	**0.0400** ± 0.0012
Gold and ceramic	0.1123 ± 0.0022	0.1110 ± 0.0040	**0.1014** ± 0.0100
Average	0.0601 ± 0.0264	**0.0582** ± 0.0246	0.0583 ± 0.0273

Table 5. Training loss mean and standard deviation for the discussed activation functions.

Material	ReLU	LeakyReLU	LeakySmeLU
da Vinci workshop materials			
Birdcage	0.0591 ± 0.0014	0.0602 ± 0.0027	**0.0528** ± 0.0039
Chandelier	0.0407 ± 0.0039	0.0410 ± 0.0024	**0.0326** ± 0.0015
Hammer	0.0389 ± 0.0026	0.0377 ± 0.0017	**0.0298** ± 0.0007
Lantern	0.0276 ± 0.0006	0.0272 ± 0.0010	**0.0195** ± 0.0006
Mirror	**0.0690** ± 0.0032	0.0703 ± 0.0019	0.0738 ± 0.0065
Palette	0.2932 ± 0.0130	**0.2918** ± 0.0067	0.3017 ± 0.0122
Chair	0.0590 ± 0.0016	0.0516 ± 0.0064	**0.0512** ± 0.0090
Scales	0.0545 ± 0.0068	0.0502 ± 0.0023	**0.0448** ± 0.0017
Table	0.0278 ± 0.0020	0.0255 ± 0.0020	**0.0190** ± 0.0013
Layered materials			
Glazed ceramic	0.1450 ± 0.0063	**0.1352** ± 0.0058	0.2141 ± 0.0137
Brushed brass	0.0432 ± 0.0020	0.0440 ± 0.0017	**0.0388** ± 0.0037
Scratched steel	0.0533 ± 0.0029	0.0524 ± 0.0031	**0.0455** ± 0.0036
Bumpy plastic	0.0206 ± 0.0004	0.0208 ± 0.0008	**0.0159** ± 0.0011
Oxydized metal	0.0247 ± 0.0014	0.0242 ± 0.0010	**0.0209** ± 0.0008
Gold and ceramic	0.1856 ± 0.0176	**0.1824** ± 0.0171	0.1925 ± 0.0111
Average	0.0761 ± 0.0754	**0.0743** ± 0.0746	0.0769 ± 0.0867

Table 6. HDR ꟻLIP error mean and standard deviation for the discussed input parameterizations.

Material	ω_i, ω_o	ω_h, ω_d	$\omega_h, \omega_d{}'$	$\omega_i, \omega_o, \omega_h, \omega_d{}'$
da Vinci workshop materials				
Birdcage	0.0410 ± 0.0018	0.0381 ± 0.0027	0.0434 ± 0.0026	**0.0370** ± 0.0029
Chandelier	0.0358 ± 0.0019	0.0346 ± 0.0010	0.0388 ± 0.0034	**0.0345** ± 0.0013
Hammer	**0.0411** ± 0.0014	0.0444 ± 0.0021	0.0476 ± 0.0017	0.0412 ± 0.0024
Lantern	0.0367 ± 0.0010	0.0383 ± 0.0015	0.0396 ± 0.0024	**0.0346** ± 0.0007
Mirror	0.0986 ± 0.0030	0.0865 ± 0.0034	0.0911 ± 0.0025	**0.0810** ± 0.0039
Palette	0.0839 ± 0.0046	**0.0837** ± 0.0116	0.0854 ± 0.0051	0.0991 ± 0.0144
Chair	0.0739 ± 0.0083	**0.0691** ± 0.0064	0.0832 ± 0.0160	0.0776 ± 0.0134
Scales	0.0686 ± 0.0043	0.0674 ± 0.0014	0.0703 ± 0.0015	**0.0652** ± 0.0012
Table	0.0434 ± 0.0032	0.0439 ± 0.0026	0.0504 ± 0.0029	**0.0425** ± 0.0024
Layered materials				
Glazed ceramic	0.1357 ± 0.0046	0.1105 ± 0.0090	0.1205 ± 0.0082	**0.1063** ± 0.0081
Brushed brass	0.0485 ± 0.0041	0.0456 ± 0.0041	0.0490 ± 0.0042	**0.0404** ± 0.0028
Scratched steel	0.0399 ± 0.0009	0.0416 ± 0.0013	0.0395 ± 0.0019	**0.0353** ± 0.0012
Bumpy plastic	0.0403 ± 0.0003	0.0456 ± 0.0015	0.0433 ± 0.0024	**0.0391** ± 0.0006
Oxydized metal	0.0409 ± 0.0009	0.0588 ± 0.0081	0.0532 ± 0.0056	**0.0400** ± 0.0012
Gold and ceramic	0.1298 ± 0.0063	0.1096 ± 0.0069	0.1096 ± 0.0039	**0.1014** ± 0.0100
Average	0.0639 ± 0.0339	0.0612 ± 0.0257	0.0643 ± 0.0272	**0.0583** ± 0.0273

Table 7. Training loss mean and standard deviation for the discussed input parameterizations.

Material	ω_i, ω_o	ω_h, ω_d	$\omega_h, \omega_d{}'$	$\omega_i, \omega_o, \omega_h, \omega_d{}'$
da Vinci workshop materials				
Birdcage	0.0947 ± 0.0028	0.0720 ± 0.0024	0.0708 ± 0.0046	**0.0528** ± 0.0039
Chandelier	0.0548 ± 0.0017	0.0471 ± 0.0009	0.0502 ± 0.0049	**0.0326** ± 0.0015
Hammer	0.0491 ± 0.0022	0.0569 ± 0.0024	0.0506 ± 0.0024	**0.0298** ± 0.0007
Lantern	0.0328 ± 0.0011	0.0351 ± 0.0016	0.0374 ± 0.0021	**0.0195** ± 0.0006
Mirror	0.1272 ± 0.0032	0.1017 ± 0.0072	0.1030 ± 0.0044	**0.0738** ± 0.0065
Palette	0.3282 ± 0.0049	0.3233 ± 0.0117	0.3247 ± 0.0100	**0.3017** ± 0.0122
Chair	0.0593 ± 0.0029	0.0735 ± 0.0033	0.0667 ± 0.0056	**0.0512** ± 0.0090
Scales	0.0725 ± 0.0031	0.0621 ± 0.0027	0.0641 ± 0.0030	**0.0448** ± 0.0017
Table	0.0314 ± 0.0016	0.0369 ± 0.0011	0.0335 ± 0.0035	**0.0190** ± 0.0013
Layered materials				
Glazed ceramic	0.4221 ± 0.0129	0.2193 ± 0.0104	0.2371 ± 0.0156	**0.2141** ± 0.0137
Brushed brass	0.0945 ± 0.0022	0.0507 ± 0.0038	0.0560 ± 0.0044	**0.0388** ± 0.0037
Scratched steel	0.0997 ± 0.0072	0.0973 ± 0.0041	0.0774 ± 0.0043	**0.0455** ± 0.0036
Bumpy plastic	0.0274 ± 0.0013	0.0559 ± 0.0023	0.0292 ± 0.0030	**0.0159** ± 0.0011
Oxydized metal	0.0289 ± 0.0005	0.0542 ± 0.0021	0.0318 ± 0.0035	**0.0209** ± 0.0008
Gold and ceramic	0.3852 ± 0.0328	0.2608 ± 0.0125	0.2473 ± 0.0181	**0.1925** ± 0.0111
Average	0.1272 ± 0.1345	0.1031 ± 0.0895	0.0986 ± 0.0924	**0.0769** ± 0.0867

References

Pontus Andersson, Jim Nilsson, Peter Shirley, and Tomas Akenine-Möller. 2021. Visualizing Errors in Rendered High Dynamic Range Images. In *Eurographics Short Papers*. 25–28. doi:10.2312/egs.20211015

Keller Jordan, Yuchen Jin, Vlado Boza, You Jiacheng, Franz Cesista, Laker Newhouse, and Jeremy Bernstein. 2024. Muon: An optimizer for hidden layers in neural networks, 2024. *URL https://kellerjordan. github. io/posts/muon* 6, 3 (2024), 4.

Diederik Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations* (12 2014).

Nikhil Vyas, Depen Morwani, Rosie Zhao, Mujin Kwun, Itai Shapira, David Brandfonbrener, Lucas Janson, and Sham Kakade. 2025. SOAP: Improving and Stabilizing Shampoo using Adam. arXiv:2409.11321 [cs.LG] https://arxiv.org/abs/2409.11321