

Fitting Compact Heat Kernel Textures

Zhicheng Du Lan Ma*

Tsinghua Shenzhen International Graduate School, Tsinghua University

duzc21@mails.tsinghua.edu.cn

malan@sz.tsinghua.edu.cn

Abstract

Can a smaller heat-kernel texture render better because it is compact, or because it receives more fitting? Matching optimization updates makes the advantage of parent-initialized compact fitting small and capacity-dependent. We investigate this distinction through the conditional colour problem exposed by fixed kernel geometry. Independent samples, controlled updates and response interventions distinguish incomplete fitting from limitations that colour recovery does not remove. Corrections help selected warm starts but add little after the tested continuous direct fits. At deployment, a streaming evaluator yields a median $1.26\times$ render speedup on 15 accuracy-qualified fields. A separate two-seed heat-feature comparison gains 0.277 dB mean rendered PSNR on 19 complete objects out of 22 declared. Readout and early-routing controls show why extra branches and fewer responses need not improve deployment. Our contribution is a controlled account of compact fitting and scoped computational improvements, with explicit evidence for when additional computation is useful.

1. Introduction

A compact texture can render better than the field from which it was compressed. This observation appears to promise lower storage and higher quality at once. Yet compression is followed by additional fitting. Before attributing the gain to a better representation, we ask: *what happens when direct fitting receives the same number of updates?*

Heat Kernel Textures (HKTex) [14] make this question concrete. Their intrinsic kernels have adaptable positions, shapes and colours on a triangular mesh. Removing kernels and refining the result can improve rendered quality over the parent. In our expanded Objaverse comparison, however, giving direct fitting the same total updates changes the mean advantage to $-0.060/ +0.074$ dB at 4,000/3,000 kernels (Fig. 1). The parent comparison alone therefore does not identify the source of the gain. Equal updates also do not

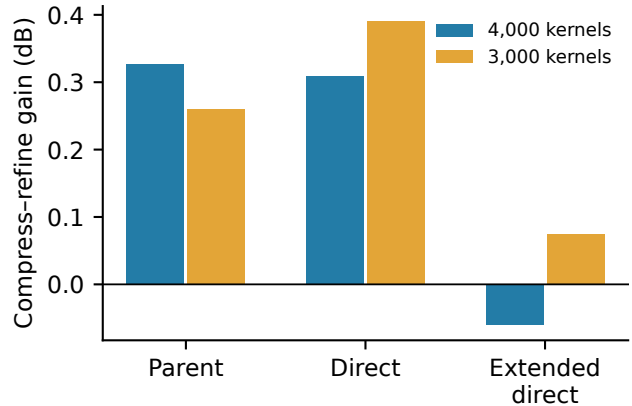


Figure 1. The control changes the interpretation. Compress-refine minus three comparison routes on the new Objaverse cohort, with 23/26 finite paired objects at the two capacities. Reference and ordinary direct fitting use 5,000 updates; compress-refine and extended direct use 10,000. Editing and colour initialization cost extra.

imply equal elapsed cost: editing and colour initialization require additional work.

We investigate what computation can improve the remaining error. HKTex provides a useful conditional structure: once kernel geometry and query support are fixed, the unclipped colour readout is linear. Colour fitting can test what those responses explain on specified samples. An incomplete solve, unstable sampling and limited response structure require different interventions. Even a useful direction can be too large: actual colour interpolation across three render resolutions tests this distinction (Appendix H). These are not additive image-error terms.

The distinction is practical. On four warm-start development objects, safeguarded colour corrections improve rendered PSNR by 0.586 dB over matched optimizer-reset controls. On 19 reserved objects after continuous direct fitting, the improvement is only 0.019 dB. A useful correction at one initialization is not automatically a useful default. Conversely, large losses remain after colour recovery when directional shape or local weighting is removed from the fit-

*Corresponding author.

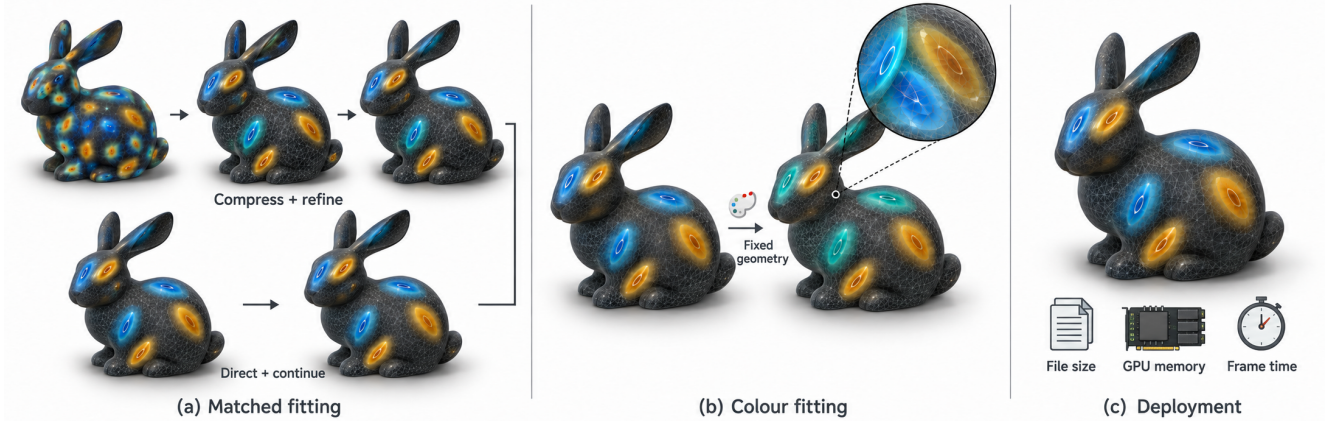


Figure 2. A controlled study of compact texture fitting. (a) Compare editing and refinement with longer direct fitting at matched update counts, charging editing separately. (b) Fix kernel centres, shapes and support to isolate what changing colours can repair. (c) Validate the resulting appearance and separately measure stored files, GPU-resident state and frame time. The conceptual footprints illustrate the controlled variables; quantitative results are reported separately.

ted fields. Some of HKTex’s response structure cannot be replaced by recolouring those configurations.

Deployment requires the same care. Kernel count, file size, resident spectral state and frame time are different resources. A small field may still require a large spectral basis. A query may use few responses, but finding them can cost more than evaluating all candidates. We implement and test conditional solving, streaming evaluation and fixed-time heat-feature alignment, while retaining UV, readout and routing controls that reveal when a seemingly attractive change is ineffective.

Our contributions are threefold. First, matched-effort comparisons change the interpretation of compact-fitting gains. Second, conditional experiments identify recoverable fitting error and response structure that recovery should preserve. Third, scoped solver, evaluator and feature improvements are evaluated against their numerical and resource costs. Figure 2 summarizes the study. These components are not presented as a jointly validated automatic optimizer.

2. Related work

Intrinsic appearance. HKTex supplies the intrinsic representation, surface-constrained optimization, adaptive density control and renderer used here [14]. Its original study already compares UV, vertex and neural textures. Our question concerns fitting and deploying this representation, rather than establishing a new state of the art across all texture families. LocalHK replaces spectral evaluation with local unfolding and analytic responses; ThermalRF constructs randomized thermal features without mesh-sized eigensystems [18]. These alternatives motivate separating represen-

tation and evaluator costs. Our local prototype implements only part of LocalHK.

Budgeted representations and coding. Budgeted Gaussian representations and precision allocation motivate count-matched and byte-matched controls [10, 19, 28]. TexF distinguishes coded bitstreams from GPU-resident compression [27]. Random-access neural texture compression and RTXNTC’s on-load/on-sample decoding similarly distinguish storage from sampling [22, 26]; differentiable block compression targets a hardware-supported format [32]. These systems are related deployment designs, not reproduced baselines in this study. Recent thread-efficient neural texture decoding groups materials around shared decoders to reduce GPU divergence [20]. Hypernetwork-based compression instead amortizes per-material fitting by predicting latent textures and decoder parameters [3]. These address different costs: efficient execution of a fitted texture and the work needed to obtain one.

Conditional optimization and computation. Eliminating linear coefficients while optimizing nonlinear variables is classical variable projection [16]; adaptive-basis training already combines least squares and gradient descent [7]. Sampling stability is a separate question [5]. Majorization and projected updates [1, 2] diagnose this conditional problem. Sparse mixtures of experts separate capacity from active computation [24]; KiloNeRF illustrates the importance of efficient execution for local small models [23]. Neither analogy makes kernel routing free. AudioCache provides an indirect motivation for separating computational reuse from validation [25]; we do not transfer its temporal audio-caching rule or speed measurements.

Evaluation protocols and downstream interfaces. Du et al. audit training-free shape retrieval with Diffused Geodesic Moments, showing that field construction and aggregation can change the ranking attributed to a descriptor [13]. Our matched-fitting study asks the analogous attribution question for surface appearance, with optimization effort and the renderer as additional variables. In a different domain, decoder-interface analysis distinguishes latent error from the stability of decoded tokens [12]. Here the corresponding distinction is between surface colour error and rendered appearance. Neither descriptor retrieval nor language generation is a texture baseline; these studies motivate checking the complete system through which a representation is judged.

3. The fitting and deployment problem

3.1. A field with discrete support

Our primary comparison uses the released direct-RGB read-out, without an optional nonlinear output network or colour post-normalization:

$$f_\theta(x) = \text{clip}_{[0,1]} \left(b + \frac{\sum_{i \in S_\theta(x)} w_i(x) c_i}{\max(1, \sum_{i \in S_\theta(x)} w_i(x))} \right). \quad (1)$$

Here b is the background, $c_i \in [-1, 1]^3$ are signed coefficients and $w_i \geq 0$ are filtered responses. The evaluator searches up to 50 intrinsic candidates and retains the 30 largest filtered responses. Support and weights depend on non-colour parameters θ . The denominator floor follows the released implementation; removing it changes the field. The output is not a convex mixture of RGB colours.

This distinction constrains simplification. Writing $a_i = w_i / \max(1, \sum_j w_j)$ and $\beta = \sum_i a_i \leq 1$, the unclipped field can also be written as

$$(1 - \beta)b + \sum_i a_i(b + c_i). \quad (2)$$

A probabilistic mixture with bounded RGB experts would impose a different feasible set: the effective values $b + c_i$ may leave the RGB cube. Simple absorption via $c_i \leftarrow b + c_i$ is valid when $\beta = 1$ and the shifted coefficients remain admissible. Normalizing every row to one or replacing signed colours with bounded experts can therefore remove useful functions before any optimization is attempted. The background-removal control in Appendix C illustrates this distinction. Such changes require a new fitting comparison, rather than being treated as algebraically free reductions.

The original geometry and intrinsic kernels remain central. We alter fitting routes or numerical evaluation under explicit contracts. A decoded field must be evaluated with its actual rebuilt support; a cached deletion score does not certify that support after editing.

3.2. Routes and resources

A reference fit uses the local adaptive-density recipe for 5,000 updates. A direct fit uses the same recipe with a smaller count cap. Compress-refine edits the reference, distils its colours, then spends another 5,000 source-supervised updates at fixed compact count. Extended direct instead continues the direct fit for 5,000 updates. All continuations reset their optimizers and learning-rate schedules. The reference may already be below a requested cap; such an object is not a compression case at that cap.

We distinguish four measurements: complete encoded field bytes, decoded coefficient storage, shared spectral/geometry state, and measured execution time. Field-file savings exclude the shared mesh and spectral basis. Fixed-time feature construction, editing, fitting, export and evaluation are charged when their corresponding total-cost comparison is made. Different protocols are never combined into a synthetic quality-speed result.

A deployment budget specifies a workload. Let e denote an evaluator and q an encoding of a fitted field f . For fixed cameras, lighting and shared-asset scope, the practical problem can be written as

$$\begin{aligned} \min_{f,q,e} \quad & D_{\text{img}}(R_e(\text{decode}_q(f)), I) \\ \text{s.t.} \quad & B_{\text{file}}(f, q, e) \leq B, \\ & M_{\text{resident}}(f, q, e) \leq M, \\ & T_{\text{setup}}(f, q, e) + NT_{\text{frame}}(f, q, e) \leq T. \end{aligned} \quad (3)$$

Here N is the intended number of frames, and setup includes the preparation, fitting and conversion charged to that workload. Router parameters and evaluator caches are counted in the resource they occupy. This formulation organizes our comparisons; we do not claim to solve its joint discrete-continuous optimization. A better byte-quality point need not satisfy a memory or latency constraint. Likewise, a more expensive initialization can be worthwhile for a repeatedly rendered asset without being attractive when assets change frequently.

4. Conditional diagnosis and scoped improvements

4.1. What a colour solve can establish

Fixing θ , background, query coordinates and tie-breaking fixes a normalized response matrix A . Recolouring does not change its entries or selected indices. For background-subtracted targets Y , the unrestricted, unclipped conditional residual is

$$\min_C \|AC - Y\|_F^2 = \|(I - AA^+)Y\|_F^2. \quad (4)$$

This classical projection identity is a lower bound for the same unclipped problem with bounded coefficients. It is not a lower bound for clipped rendered PSNR. Moving kernels changes A and need not enlarge its column space monotonically.

For the convex box problem $F(C) = \frac{1}{2}\|AC - Y\|_F^2$, a computable first-order gap is

$$\begin{aligned} \gamma(C) &= \langle \nabla F(C), C \rangle + \|\nabla F(C)\|_1, \\ F(C) - \gamma(C) &\leq F^* \leq F(C). \end{aligned} \quad (5)$$

The box is $[-1, 1]$ for each coefficient. If a target loss is above the upper bound it is already feasible; if it is below the lower bound, further solving cannot reach it with this matrix and colour box. The intermediate case remains unresolved. This separates solving from changing support on the fitted samples, without certifying sample coverage or image quality. Appendix B gives the conditions.

Three questions before moving a kernel. The conditional problem gives an operational order of diagnosis. First, inspect the optimization gap on the chosen fitting samples: a wide gap leaves the current response family unresolved. Second, evaluate the proposed correction on separate samples. A training gain that disappears there points to sampling or update selection, rather than demonstrating a need for more geometric flexibility. Third, after substantial colour recovery, compare controlled response interventions. A persistent loss can then identify information that recolouring those responses does not restore. This order prevents a poor conditional solve from being mistaken for a limitation of the representation. It does not turn finite-sample tests into a certificate over the surface.

A practical conditional solver. Nonnegativity provides a separable curvature bound. Let $H = A^T A$, $d = H\mathbf{1}$ and $D = \text{diag}(d)$. Since

$$v^T(D - H)v = \frac{1}{2} \sum_{ij} H_{ij}(v_i - v_j)^2 \geq 0, \quad (6)$$

the update $C^+ = \text{clip}_{[-1,1]}(C - D^{-1}\nabla F(C))$ does not increase F in exact arithmetic. Zero-response columns remain unchanged. We implement sparse GPU products and safeguard an accelerated proposal by checking the true quadratic, falling back to this update. The construction is classical majorization [1, 2]; its value here is a measurable solver for decoded response matrices.

A correction is more than its endpoint. A solve can improve its fitting samples while its full update harms fresh samples. We therefore test sample coverage and validation-selected partial steps before subsequent geometry updates.

The reserved correction study uses 65,536 fitting points, 8,192 separate guard points from the training pool, 100 coordinate sweeps, and step fractions $\{0, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1\}$. Scheduled reset-only controls perform the same colour-moment resets, including after rejected fits. Guard points are not the independent test set. Rejection still incurs computation; validation is not a guarantee of later rendered improvement.

The candidate update is $C_\alpha = (1 - \alpha)C + \alpha C^*$, with the zero step included. Selection is therefore allowed to retain the incoming field. The matched reset-only arm distinguishes a benefit from recolouring from a benefit caused by clearing optimizer state. A further geometry-frozen arm asks whether the gain requires later changes to centres and shapes. These controls answer different questions: conditional endpoint quality, subsequent optimization, and the interaction between the two. We report their contrasts separately.

4.2. Preserving responses while reducing execution cost

For a frozen field, source-dependent spectral decay can be cached. The streaming evaluator fuses query interpolation and contraction without materializing the query-candidate-spectrum tensor. It keeps the global basis and the original search, filtering and readout. Thus it targets temporary work and latency, not elimination of the eigensystem. We compare against a control that additionally compiles whole-query gathering and diffusion, with matched numerical checks.

For a query block of size Q , K candidates and spectral width r , a materialized candidate-spectrum intermediate has QKr entries. Streaming contracts this dimension as it loads it, retaining compact query outputs and source-side decay terms. It does not remove the underlying spectral arithmetic or shared basis. This is why we measure incremental allocation separately from total resident state, and frame time separately from an isolated kernel's timing. Compilation of the control matters: otherwise an apparent algorithmic gain may largely reflect a different execution boundary.

Top- k support can amplify small numerical changes. At a positive cutoff tie, exchanging responses of weight w with different colours changes the unclipped output by $w(c_i - c_j)/d$ when the denominator is unchanged. A small response error need not imply a small colour error. For a fixed outer candidate set, a useful local condition is available. Let $\delta = w_{(k)} - w_{(k+1)}$ be the exact top- k cutoff margin. If every response changes by at most ϵ and $\delta > 2\epsilon$, selection is unchanged: every selected-unselected pair remains ordered by at least $\delta - 2\epsilon > 0$. At a tie this certificate disappears, even if response errors are small. Consequently, compiler settings, actual support and held-out query agreement accompany every acceleration claim.

4.3. Aligning features at a fixed diffusion time

LocalHK’s thermal-feature analysis already identifies the issue of aligning a basis while leaving its diagonal diffusion weights unchanged [18]. For fixed time t , define $Z_t = \Phi \exp(-t\Lambda/2)$. Any orthogonal alignment satisfies

$$(Z_t Q)(Z_t Q)^T = Z_t Z_t^T. \quad (7)$$

We align these pre-diffused features before interpolation over anisotropy and angle. The grid-node identity is elementary; off-grid behavior and fitting benefit require measurement. Both compared variants use verified FP64 alignment factors, stored in FP32. The construction retains spectral precomputation and fixes diffusion time; changing time requires rebuilding the features. It is not a drop-in conversion of trained fields.

The order of these operations matters. Aligning raw eigenvectors and then applying the old diagonal weights gives $\Phi Q \exp(-t\Lambda) Q^T \Phi^T$. This equals the original fixed-time kernel only under an appropriate commutation condition. By incorporating diffusion before alignment, Eq. (7) preserves the grid-node Gram matrix for any orthogonal Q . Interpolation between aligned feature grids still changes the family of off-grid responses. We therefore compare separately fitted fields with identical raw parameter arrays, and explicitly avoid claiming that they start from identical functions or that every interpolated kernel becomes more accurate.

5. Experiments

5.1. Protocols and coverage

The expanded fitting-route study locks 40 new Objaverse objects [9], 18 public LocalHK assets (ten main and eight low-poly), eight Google Scanned Objects [11] and five official HKTex tuning examples. The 71 admitted objects include failures; 59 produce finite seed-zero reference fields. Newly selected data are conditional on fixed geometry, size and licence screening, not uniform collection samples. Objects are not replaced after fitting failure. The tuning examples are descriptive replications. The conditional, feature and deployment studies use separate cohorts and are not added to this count as independent samples.

Final decoded fields use common 65,536-point area-uniform surface tests and five 256×256 , four-sample PBR views per object. Rendering randomness is fixed; each frame is independently denoised. Primary quality is mean per-view sRGB PSNR including background, then an equal-weight mean over finite paired objects. Pooled-image PSNR, SSIM and LPIPS are supplementary. Exact-match views can yield infinite scores; undefined differences are retained in coverage and never counted as wins. Seeds and views are repetitions within objects, not additional indepen-

Table 1. Compress–refine minus comparison PBR PSNR (dB). Extended direct matches the total update count. Parent/direct comparisons include additional fitting; editing costs remain extra.

Cohort / cap	n	Parent	Direct	Extended
New Obj. / 4k	23	+ .326	+ .310	− .060
New Obj. / 3k	26	+ .260	+ .390	+ .074
Public main / 4k	8	+ .381	+ .452	+ .062
Public main / 3k	9	+ .182	+ .444	+ .090
Public low-poly / 4k	7	+ .162	+ .216	+ .061
Public low-poly / 3k	7	+ .031	+ .236	+ .023
GSO / 4k	7	+ .303	+ .309	− .021
GSO / 3k	6	+ .179	+ .328	− .027

dent samples. We study source-texture reconstruction, not the original multiview inverse-rendering task.

5.2. Does compact initialization beat more fitting?

Table 1 gives the decisive contrast. At the two capacities, compress–refine gains 0.326/0.260 dB over its parent and saves 17.4/33.9% of mean field-file bytes on the new Objaverse panel. Its gain over ordinary direct fitting is 0.310/0.390 dB. After extending direct fitting to the same total updates, the mean differences become −0.060/ +0.074 dB, with medians −0.032/ +0.015 dB. The smaller-budget bootstrap interval crosses zero. On the 23 objects common to both budgets the differences are −0.060/ +0.081 dB; the sign change is not solely due to denominators. It is also not evidence of a general capacity-transition law.

The same scalar codec preserves this heterogeneous result. Refined minus extended-direct quality becomes −0.062/ +0.065 dB, while the refined bitstreams average 0.63/0.45% larger. Equal counts and precisions do not ensure equal bytes. Three-seed repeats on a locked subset include negative outcomes; the earlier all-positive historical repeats do not establish general seed robustness. Appendix A retains all coverage and the distinction between primary and diagnostic evidence.

5.3. When is conditional fitting worth executing?

Twelve events on four historical fields expose a useful distinction: full refits improve their own samples but reduce fresh-point PSNR by 0.713 dB on average, despite negative directional derivatives at the current field. A useful direction can have an excessive endpoint. Larger fitting sets and selected partial steps repair much of this behavior. Subsequent matched continuations test whether the repair survives geometry updates.

The development gain cannot be explained by moment resets alone. With geometry frozen, the correction advantage is 0.146 dB; the joint-minus-frozen interaction is positive on all four objects. This supports a useful interac-

Table 2. Correction minus matched reset-only control. Cohorts and training states differ; the rows are not a pooled effect.

Setting	Objects	Δ PBR	Positive
Warm-start study	4	+.586	4/4
Continuous direct fit	19/24	+.019	15/19
Partial local evaluator	4	+.006	2/4

tion with subsequent geometry changes, without establishing causal mediation. SmallLimo contributes about 70% of the summed rendering differences while its surface PSNR falls 0.045 dB against that reset control: even a positive image result need not improve both metrics.

The reserved comparison is more limiting. Nineteen of 24 objects support the fixed configuration and complete three seeds and three arms: 171 runs. Three objects fail cache construction and two are below the fixed rank. After 5,000 continuous updates, the mean correction gain is 0.019 dB (median 0.004; object-bootstrap 95% interval [0.006, 0.037]). Only 123/228 attempted corrections are accepted. Nine runs reject every correction and have exactly the reset control’s final parameter arrays. This small gain does not establish useful speed. The 36-run partial local-evaluator comparison is mixed (Table 2), not a complete LocalHK reproduction.

What should the fitting preserve? On four fixed compact fields, all arms receive the same 1,000-sweep bounded colour recovery. Minimum anisotropy, removal of distance weighting, uniform positive weights and winner-only response lose 3.625, 5.725, 6.391 and 11.369 dB of surface PSNR, respectively. For these response interventions, sampled unclipped lower bounds exceed the original arm’s feasible loss on every object. Unfinished colour solving therefore cannot fully explain those fitting losses. These are conditional interventions, not full retraining ablations.

5.4. Can feature consistency improve fitted quality?

A separate comparison locks 22 previously geometry-screened objects, held out from development of the fixed-time feature construction. Both variants use 3,000 kernels, 5,000 updates, two seeds, the same raw initial parameter arrays, native targets and batch stream. Equal arrays need not define equal initial functions. Two tiny meshes cannot support the fixed configuration and one object loses a seed to a timeout. Nineteen complete objects form the primary aggregate (Table 3); no replacement is made.

The largest gain, Doormate at 3.505 dB, raises the mean. Omitting it post hoc leaves 0.098 dB on 18 objects; this is a sensitivity check, not a replacement estimate. Native PSNR gains 0.418 dB, but overall SSIM and LPIPS intervals cross zero. Objaverse mean LPIPS worsens. The result supports

Table 3. Feature alignment minus verified basis/diagonal alignment. Each object averages two seeds. The interval resamples objects.

Cohort	Objects	Mean Δ PBR	Positive
Objaverse	11	+.399	9/11
GSO	8	+.109	6/8
All complete	19	+.277	15/19

Median: +.096 dB; 95% interval: [+ .051, +.668] dB.

Table 4. Separate two-mesh replay: feature-minus-control quality and accounted total process cost. Costs are measured stage sums.

Object	Seed	Δ PBR	Control s	Feature s
Barricade	0	+.377	462.2	462.9
Barricade	1	+.672	458.4	462.6
Wrench	0	+1.009	439.8	443.2
Wrench	1	+.820	436.8	445.1

a modest, heterogeneous quality benefit. Shared-resource timings do not establish broader speed or cost parity.

A separate two-mesh, two-seed cost replay rebuilds both caches. All four pairs improve PBR quality, averaging 0.720 dB, with accounted total cost increasing 0.15–1.90%. This is a quality gain at similar measured cost, not a speedup. The stage totals include preparation, fitting, export and the prescribed evaluation; they are not uninterrupted cold-start measurements. These two development objects remain separate from the broader quality result.

Figure 3 shows the recorded quality–cost trajectories. At the smaller common endpoint budget, the last eligible checkpoints are the 4,750-update feature fit and the 5,000-update control in all four pairs. Their quality differences remain positive: 0.344, 0.698, 0.974 and 0.811 dB. Thus the endpoint signal is not explained solely by the feature variant’s slightly larger accounted cost. Checkpoints are discrete, however; connecting them does not establish an exact time-to-quality crossing. Nor can these diagnostic pairs substitute for timing the larger confirmation cohort.

5.5. Which savings survive actual deployment?

Solving and streaming. The eight-object deployment pilot has two fitted fields per object. The conditional GPU solver reaches the stated gap on all 16 fields; the conservative global-step accelerated control does so on six. On those mutually converged fields, the median solve-time ratio is $6.26\times$, with response construction and preparation separate. This is not a comparison against every tuned least-squares solver.

On fitted four-bit fields, streaming renders a median $1.26\times$ faster than the compiler-matched control on 15 mutually accuracy-qualified fields. All 16 streaming fields pass

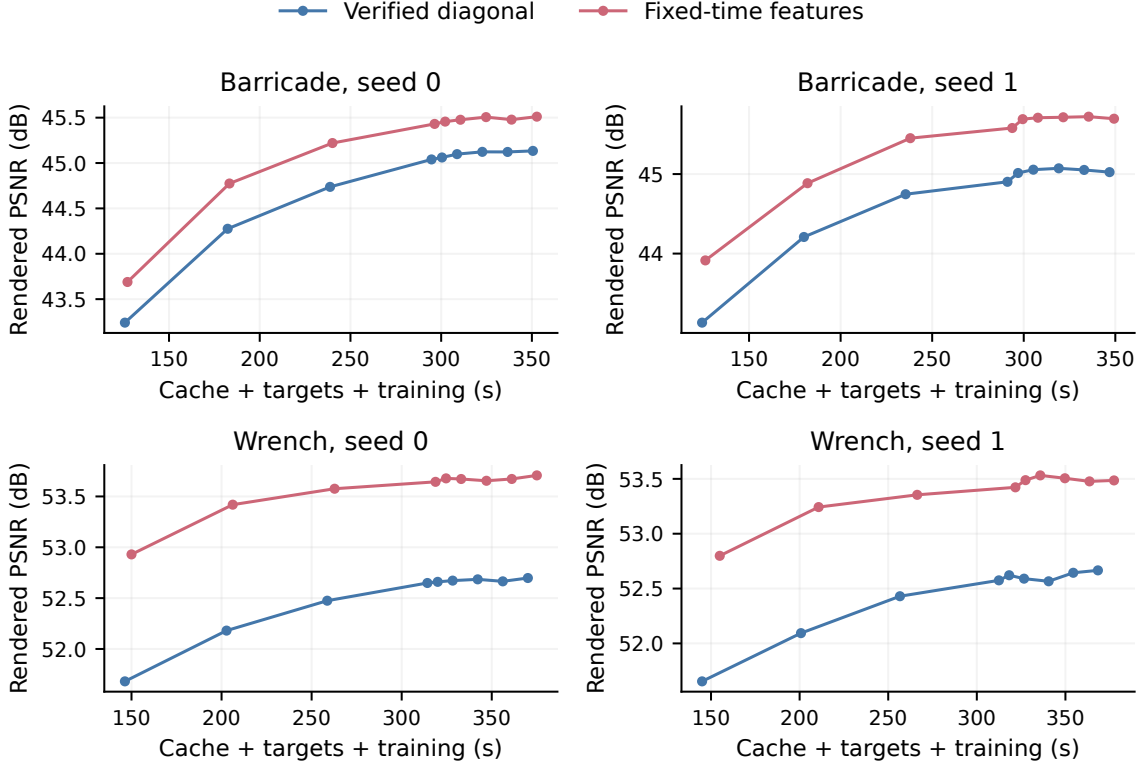


Figure 3. Quality against measured fitting cost on two diagnostic meshes and two seeds. Points are recorded checkpoints, with no interpolation used to select a comparison. The horizontal axis includes fresh cache construction, target preparation and training; export and offline evaluation are added in Table 4; stage definitions are in Appendix D. These four pairs remain separate from the 19-object feature confirmation.

that query check. Incremental PyTorch temporary allocation is a median 4.50% of the control, with 1.116–3.072 MB of extra decay cache; total GPU memory does not shrink by that factor. The global basis remains. Appendix E gives coding and precision controls. Figure 4 separates solving, execution and coding: these improvements use related fields, but their denominators and resource measurements differ.

Fewer responses can cost more. We next fit a small geometric scorer that ranks candidates before spectral evaluation. A calibration-only rule chooses the smallest budget losing at most 0.05 dB of surface PSNR. Three historical fields retain about 32–45/50 candidates; Orange falls back to all 50. Scoring and dispatch outweigh the avoided evaluations. The median full-stream/routed frame-time ratio is 0.415, below one. Fusion roughly halves routing query time, but the three non-fallback fields remain about $1.8\times$ slower than full streaming. The router adds about 61 KB, comparable to the compact field file.

The idealized after-response selector reveals a second limitation. It already knows all 50 responses, so its 10–19 retained terms are not saved spectral evaluations. On

SmallLimo, fixed-24 selection improves surface PSNR by 0.067 dB yet loses 0.477/0.459 dB on the original and shifted view sets. Its adaptive counterpart loses only 0.026 dB on surface test but 0.201/0.315 dB in rendering. This is a concrete failure of transferring the tested surface criterion to the image target.

The discrepancy has a simple local explanation. For a fixed visibility configuration and a differentiable rendering map, a small perturbation Δf gives $\Delta I \approx J_R \Delta f$. Its image error is then approximately $\Delta f^T J_R^T J_R \Delta f$, whereas uniform surface sampling weights errors by area. The two quadratic forms need not order candidates equally. Visibility, projected size and shading change the sensitivity of different regions. This is an explanatory local approximation, not a bound through arbitrary visibility changes or denoising. The SmallLimo counterexample shows why an image-level check remains necessary even when a surface guard accepts the policy.

All four primary objects complete both five-view sets. Three fusion follow-ups fail strict repeated-baseline uint8 image equality and are excluded from qualified complete-render claims. Hummingbird completes and remains slower

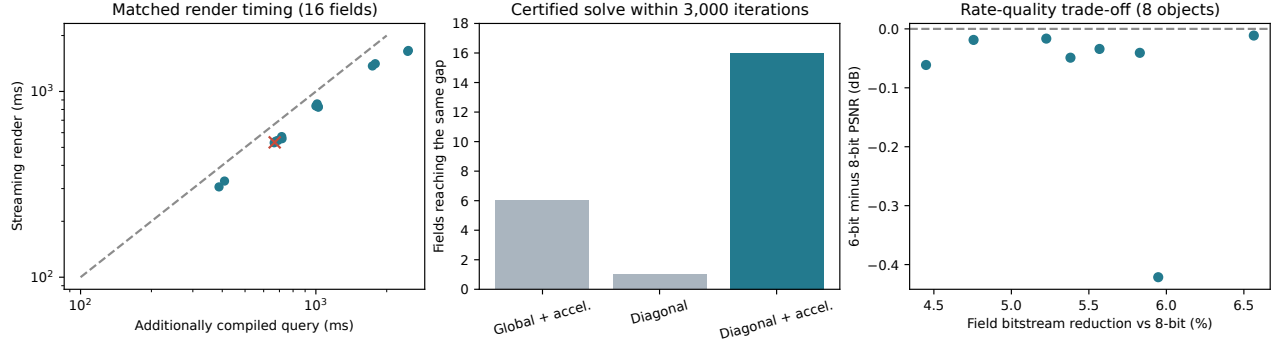


Figure 4. Three deployment questions on the eight-object pilot. Left: frame time on fitted four-bit fields; the failed compiled control is marked and excluded from the qualified speed ratio. Centre: fields reaching a common conditional gap within the iteration budget. Right: a separate six-bit/eight-bit file-size comparison. Reduced file bytes do not imply reduced resident precision.

in both view sets. Orange has sparse larger raw-image differences; these are not all dismissed as rounding or attributed to HKTex theory. The unchanged original evaluator also changes. The predeclared speed criterion fails, so no new-object expansion occurs.

Simpler alternatives remain necessary. Native UV is strong when an atlas works; atlas waste alone does not determine failure. All 20 historical fixed-support residual hybrids are dominated by the corresponding byte-matched finer-UV controls. Later nonlinear single-branch controls exceed the learned gate on six of nine finite objects at smaller payloads; Piggybank retains a gate advantage. These experiments delimit those tested hybrids, not all possible combinations or a global Pareto frontier.

6. Discussion and limitations

The central lesson is to spend computation on a measured limitation. Equal updates test the attribution of compact-fitting gains. Conditional fits test the current responses on defined samples. Actual decoded evaluation tests whether savings survive support selection, hardware execution and rendering. None of these measurements substitutes for the others.

The study is narrower than a new representation or a universal optimizer. The full original multiview task is not rerun. Geometry screening, configuration failures and limited seed coverage constrain generalization. The deployment studies reuse objects and use RTX 3090 GPUs; no second-hardware claim is made. The broader heat-feature study does not establish a speedup. A partial local implementation cannot establish superiority over LocalHK. The recent routing failures also limit interpretation of tiny rendered differences.

What the evidence recommends. For a new compact-fit comparison, extending direct training is an essential control before introducing an initialization mechanism. For a fixed field with a substantial conditional gap, colour fitting provides a measurable intervention; after a well-developed direct fit, its small average gain makes unconditional insertion difficult to justify. For a frozen renderer, reducing intermediate storage is useful when measured frame time improves without changing accepted output. A routing policy faces the stronger requirement of paying for its own selection and preserving the task’s image target. These are decisions supported by particular controls, rather than an automatic policy validated across the joint budget in Eq. (3).

Appendix G tests whether geometry predicts these gains. Exploratory clusters and leave-one-object-out prediction do not establish a reliable shape category. A high-gain thin-mesh group is driven by one object. The evidence therefore supports diagnosis of fitting state and execution cost more directly than a shape-based selection rule.

7. Conclusion

A smaller field rendering better than its parent does not by itself establish a better compression strategy. Matching fitting effort changes that comparison. Conditional colour fitting then helps identify what is recoverable and which response structure should be preserved. Streaming evaluation and fixed-time feature alignment provide scoped computational and quality improvements, while readout and routing controls expose the limits of proxy budgets. HKTex supplies the intrinsic representation; this study makes its compact fitting and deployment choices more explicit, measurable and reproducible.

References

- [1] Amir Beck and Marc Teboulle. [A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems](#). *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.
- [2] Amir Beck and Marc Teboulle. [Fast Gradient-Based Algorithms for Constrained Total Variation Image Denoising and Deblurring Problems](#). *IEEE Transactions on Image Processing*, 18(11):2419–2434, 2009.
- [3] Laurent Belcour. [Neural Texture Compression using Hypernetworks](#). arXiv:2606.26913, 2026.
- [4] Yoojin Choi, Mostafa El-Khamy, and Jungwon Lee. [Variable Rate Deep Image Compression With a Conditional Autoencoder](#). arXiv:1909.04802, 2019.
- [5] Albert Cohen and Giovanni Miglioni. [Optimal weighted least-squares methods](#). *The SMAI Journal of Computational Mathematics*, 3:181–203, 2017.
- [6] Jorge Condor, Nicolas Moenne-Loccoz, Merlin Nimier-David, Piotr Didyk, Zan Gojcic, and Qi Wu. [Neural Harmonic Textures for High-Quality Primitive Based Neural Reconstruction](#). arXiv:2604.01204, 2026.
- [7] Eric C. Cyr, Mamikon A. Gulian, Ravi G. Patel, Mauro Perego, and Nathaniel A. Trask. [Robust Training and Initialization of Deep Neural Networks: An Adaptive Basis Viewpoint](#). In *Mathematical and Scientific Machine Learning*, PMLR 107:512–536, 2020.
- [8] Chinmay Datar, Taniya Kapoor, Abhishek Chandra, Qing Sun, Erik Lien Bolager, Iryna Burak, Anna Veselovska, Massimo Fornasier, and Felix Dietrich. [Fast training of accurate physics-informed neural networks without gradient descent](#). In *International Conference on Learning Representations*, 2026.
- [9] Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. [Objaverse: A Universe of Annotated 3D Objects](#). In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13142–13153, 2023.
- [10] Francesco Di Sario, Riccardo Renzulli, Marco Grangetto, Akihiro Sugimoto, and Enzo Tartaglione. [GoDe: Gaussians on Demand for Progressive Level of Detail and Scalable Compression](#). arXiv:2501.13558, 2025.
- [11] Laura Downs, Anthony Francis, Nate Koenig, Brandon Kinman, Ryan Hickman, Krista Reymann, Thomas B. McHugh, and Vincent Vanhoucke. [Google Scanned Objects: A High-Quality Dataset of 3D Scanned Household Items](#). In *IEEE International Conference on Robotics and Automation*, pages 2553–2560, 2022.
- [12] Zhicheng Du and Lan Ma. [Continuous Language Diffusion as a Decoder-Interface Problem](#). arXiv:2606.08810, 2026.
- [13] Zhicheng Du, Changyue Liu, Wenji Xi, Zhaotian Xie, Zhuo Deng, Ziheng Zhang, Yang Liu, and Lan Ma. [Auditing Training-Free 3D Shape Retrieval with Diffused Geodesic Moments](#). arXiv:2605.29004, 2026.
- [14] Simone Foti, Caner Korkmaz, Stefanos Zafeiriou, and Tolga Birdal. [Heat Kernel Textures: the Geodesic Gaussians That Do Not Splat](#). In *European Conference on Computer Vision*, pages 306–323, 2026.
- [15] Elias Frantar, Saleh Ashkboos, Torsten Hoefer, and Dan Alistarh. [GPTQ: Accurate Post-Training Quantization for Generative Pre-trained Transformers](#). arXiv:2210.17323, 2022.
- [16] Gene H. Golub and Victor Pereyra. [The Differentiation of Pseudo-Inverses and Nonlinear Least Squares Problems Whose Variables Separate](#). *SIAM Journal on Numerical Analysis*, 10(2):413–432, 1973.
- [17] Jingwen He, Chao Dong, and Yu Qiao. [Interactive Multi-Dimension Modulation with Dynamic Controllable Residual Learning for Image Restoration](#). arXiv:1912.05293, 2019.
- [18] Zhewen He, Junyi Hu, and Yi Fang. [Accelerating HKTex without Mesh Eigensystems: Local Unfolding and Randomized Thermal Features](#). arXiv:2609.14105, 2026.
- [19] Saswat Subhajyoti Mallick, Rahul Goel, Bernhard Kerbl, Markus Steinberger, Francisco Vicente Carrasco, and Fernando De La Torre. [Taming 3DGS: High-Quality Radiance Fields with Limited Resources](#). In *SIGGRAPH Asia Conference Papers*, pages 1–11, 2024.
- [20] Janarbek Matai, Sho Ikeda, Lukasz Lipski, and Takahiro Harada. [Thread-Efficient Decoding for Neural Texture Compression](#). arXiv:2608.27888, 2026.
- [21] Markus Nagel, Rana Ali Amjad, Mart van Baalen, Christos Louizos, and Tijmen Blankevoort. [Up or Down? Adaptive Rounding for Post-Training Quantization](#). arXiv:2004.10568, 2020.
- [22] NVIDIA. RTX Neural Texture Compression SDK. <https://github.com/NVIDIA-RTX/RTXNTC>. Accessed October 14, 2026.
- [23] Christian Reiser, Songyou Peng, Yiyi Liao, and Andreas Geiger. [KiloNeRF: Speeding up Neural Radiance Fields with Thousands of Tiny MLPs](#). In *IEEE/CVF International Conference on Computer Vision*, 2021.
- [24] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarczyk, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. [Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer](#). In *International Conference on Learning Representations*, 2017. arXiv:1701.06538.
- [25] Qingyang Shi, Zhicheng Du, Jiasheng Lu, Yingshan Liang, Xinyu Zhang, Yiran Wang, Jing Peng, and Kehong Yuan. [AudioCache: Accelerate Audio Generation With Training-Free Layer Caching](#). In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2025.
- [26] Karthik Vaidyanathan, Marco Salvi, Bartłomiej Wronski, Tomas Akenine-Moller, Pontus Ebelin, and Aaron Lefohn. [Random-Access Neural Compression of Material Textures](#). *ACM Transactions on Graphics*, 42(4):1–25, 2023.
- [27] Jianqiang Wang, Junhui Hou, Siyu Ren, Weiyao Lin, and Wenping Wang. [Beyond UV Mapping: Mesh Texture Compression via Surface-Aligned Texture Fields](#). arXiv:2609.23606, 2026.
- [28] Shuzhao Xie, Jiahang Liu, Weixiang Zhang, Shijia Ge, Sicheng Pan, Chen Tang, Yunpeng Bai, Cong Zhang, Xiaoyi Fan, and Zhi Wang. [SizeGS: Size-aware Compression of 3D Gaussian Splatting via Mixed Integer Programming](#). In *ACM International Conference on Multimedia*, pages 8214–8223, 2025.
- [29] Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. [Mip-Splatting: Alias-Free 3D Gaussian](#)

- [Splatting](#). In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19447–19456, 2024.
- [30] A. L. Yuille and Anand Rangarajan. [The Concave-Convex Procedure](#). *Neural Computation*, 15(4):915–936, 2003.
- [31] Fengdi Zhang, Haocheng Ren, Qing Luo, Yaqing Li, Jibing Lou, and Hongwei Li. [Compact Representation of Mipmapped SVBRDFs via Shared Gaussians](#). arXiv:2607.27943, 2026.
- [32] Tao Zhuang, Wentao Liu, and Ligang Liu. [Differentiable Block Compression for Neural Texture](#). *Eurographics Symposium on Rendering*, 2025.

A. Comparison contracts and coverage

A.1. What is held fixed

The expanded fitting-route comparison uses source-UV supervision and the configured direct-RGB field. A reference fit has adaptive density control for 5,000 updates. A compact direct fit has the same recipe with a smaller maximum count. The compress-refine initializer ranks deletions once using the cached change to the reference output on 4,096 calibration queries, retains the selected kernels and refits colours. It then uses another 5,000 source-supervised updates at fixed count. The initializer distils the reference, whereas refinement revisits the source texture. A rebuilt-support post-hoc refit is a separate control, not this initializer.

Extended direct and full-capacity continuation spend the extra updates without compression. They preserve the attained direct count or the reference count, respectively. All continuations reset optimizers and schedules. A direct fit can finish below its cap. This is an explicitly reported route comparison, not a claim that every endpoint has an identical number of active parameters.

Table 5. Coverage of the locked expanded cohort. Valid denotes a finite exported field and completed decoded evaluation, not a finite value for every image metric.

Cohort	Plan	Ref. valid	Eligible 4k/3k	Refined 4k/3k
New Objaverse	40	29	23/26	23/26
Public main	10	10	8/9	8/9
Public low-poly	8	8	8/8	8/8
GSO	8	8	8/8	7/6
Tuning examples	5	4	4/4	4/4

A.2. Recruitment and failure handling

The new Objaverse frame contains 320 candidates after licence, size and author exclusions. Fifty-seven pass the geometric predicates; the first 40 in locked order enter. For GSO, 24 fixed candidates give 17 geometry successes; the first eight from distinct named categories enter. Conversion preserves face counts, corner geometry and UV coordinates within recorded tolerances and exact decoded texture pixels, without remeshing or resizing. Spectral or fitting failures do not cause replacement.

The historical ten-object intake changed during execution and replaced one failed object. It is retained as exploratory evidence and is not pooled with the locked expansion. Some historical editors were corrected; the new expanded comparison uses the frozen corrected recipe, without relabelling earlier runs.

The original HKTex paper reports 313 UV-fitting, 162 multiview-fitting and 92 ablation objects [14]; these need

not be disjoint and their full membership has not been verified here. The 18 public LocalHK assets improve object overlap but do not make the present training, rendering and timing protocols identical to LocalHK. This study does not replace either original evaluation at its full task scale.

Three-seed repeats select the first two public main objects, first two low-poly objects, first four new Objaverse objects and first two GSO objects in frozen order. Unavailable objects are not replaced. Their object identities, per-seed measurements and terminal failures are included in the released evidence. The primary route results use seed zero; repeated seeds do not multiply the independent sample size.

A.3. Metrics and quantitative examples

Training retains the pinned native geodesic correction and source-colour initialization, including the original high/TF32 matrix-product setting. Evaluation disables TF32 matrix multiplication. Every route for an object uses the same spectral cache. Primary image scores average five per-view PSNRs and then average finite paired objects equally. The average of logarithms differs from the logarithm of pooled MSE; one is not substituted for the other on individual objects. The background is included in image scores.

The common-23-object control gives $-0.060/ +0.081$ dB for compress-refine minus extended direct. Full-capacity continuation instead exceeds refinement by $0.044/0.095$ dB on the two eligible Objaverse panels, at larger field size. Refinement recovers $1.39/2.70$ dB over the edited-and-recoloured references, but this includes additional source supervision and optimization.

Whole-frame examples in Fig. 6 use a recorded quartile selection rule. They illustrate actual outputs; they do not demonstrate that every small signed PSNR difference is perceptually visible. Historical repeated renders under nominally identical settings differed by up to 0.0077 dB. This observed diagnostic is not a calibrated uncertainty threshold for every later experiment.

A.4. Expanded results and measurement accounting

This appendix preserves the distinction between the declared image score and supplementary aggregation. Pooled PSNR is computed from the MSE over all five images before taking the logarithm. It is applied uniformly, including objects with some exact-match views. Bootstrap intervals are descriptive object resamples within each finite paired cohort; no sequential significance stopping is used. The expanded Objaverse/public collection has eleven full spectral-cache failures (five SVD nonconvergence and six embedding-rank assertions) and one further reference-initialization failure caused by texture-array dimensionality. Ten of the cache failures and the initialization failure

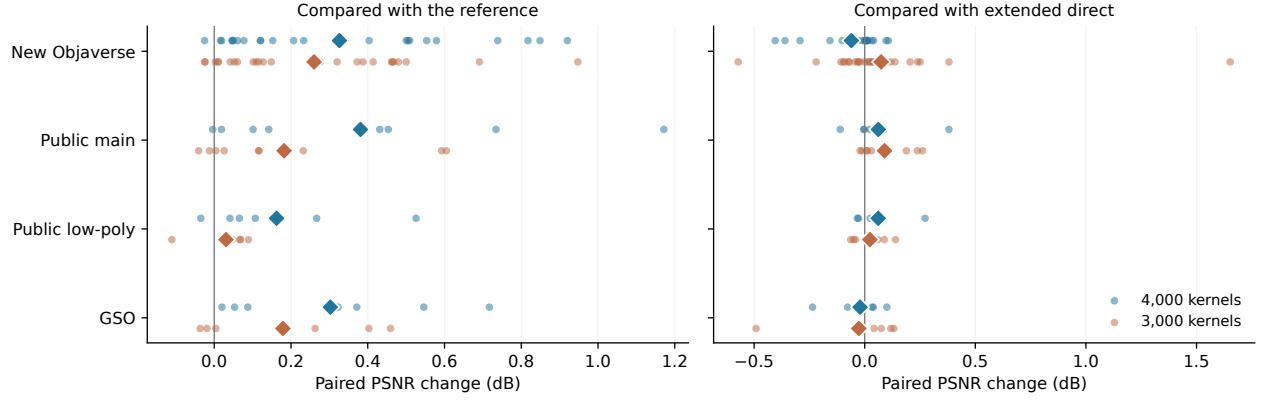


Figure 5. Object-level expanded route comparison. Small points are finite paired objects; diamonds are cohort means. Cohorts are kept separate and missing outcomes remain in the ledger. The result should not be read as a single uniform advantage from compression.

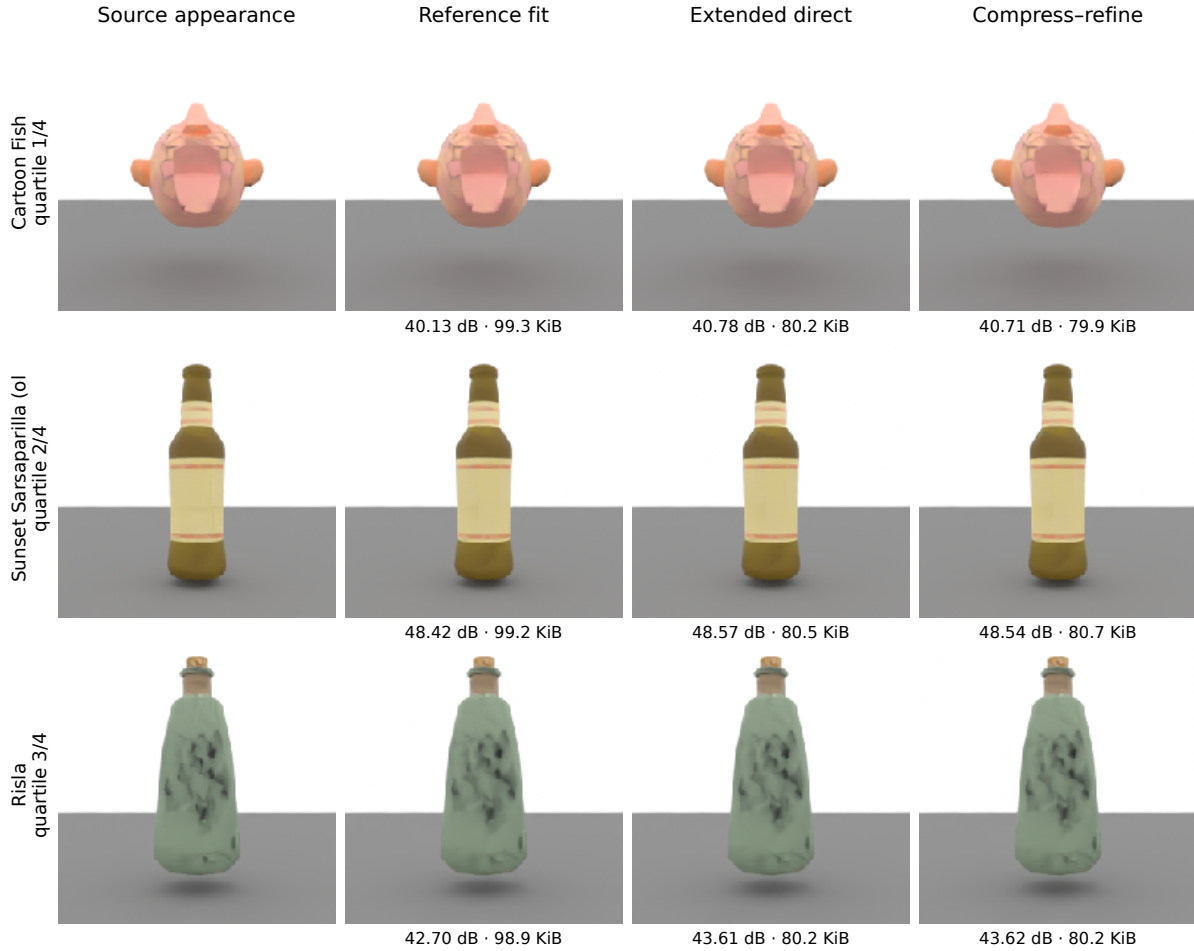


Figure 6. Recorded reconstruction examples from the expanded comparison. Images are actual shared-renderer outputs selected by the stored quartile rule. Figure 2 instead gives a conceptual overview. Source assets (CC BY metadata): [Cartoon Fish](#) by DwaalAnker; [Sunset Sarsaparilla \(old\)](#) by biliam; and [Rislá](#) by Asylum Nox. Views are rendered from these assets and the fitted fields.

Cohort	Budget	View-mean Δ	Pooled Δ	LPIPS Δ	SSIM Δ
New Objaverse	4,000	+0.326 (23)	+0.321 (23)	-0.00155	+0.00156
New Objaverse	3,000	+0.260 (26)	+0.261 (26)	-0.00096	+0.00096
Public main	4,000	+0.381 (8)	+0.398 (8)	-0.00109	+0.00076
Public main	3,000	+0.182 (9)	+0.201 (9)	-0.00029	+0.00022
Public low-poly	4,000	+0.162 (7)	+0.112 (8)	-0.00098	+0.00028
Public low-poly	3,000	+0.031 (7)	-0.012 (8)	-0.00038	+0.00009
GSO	4,000	+0.303 (7)	+0.334 (7)	-0.00094	+0.00091
GSO	3,000	+0.179 (6)	+0.230 (6)	-0.00083	+0.00071
Tuning examples	4,000	+0.385 (4)	+0.469 (4)	-0.00370	+0.00289
Tuning examples	3,000	+0.165 (4)	+0.241 (4)	-0.00127	+0.00080

Table 6. Metric sensitivity for compress–refine versus the reference. PSNR cells give finite paired counts; smaller LPIPS and larger SSIM are favourable. Neither a positive mean PSNR nor one aggregation convention implies universal or perceptual improvement.

Cohort	Budget	Reference: mean [95% interval]	Extended direct: mean [95% interval]
New Objaverse	4,000	+0.326 [+0.208, +0.452]	-0.060 [-0.117, -0.012]
New Objaverse	3,000	+0.260 [+0.171, +0.356]	+0.074 [-0.041, +0.234]
Public main	4,000	+0.381 [+0.137, +0.663]	+0.062 [-0.018, +0.163]
Public main	3,000	+0.182 [+0.039, +0.342]	+0.090 [+0.025, +0.159]
Public low-poly	4,000	+0.162 [+0.052, +0.302]	+0.061 [+0.002, +0.140]
Public low-poly	3,000	+0.031 [-0.021, +0.067]	+0.023 [-0.029, +0.076]
GSO	4,000	+0.303 [+0.129, +0.492]	-0.021 [-0.102, +0.044]
GSO	3,000	+0.179 [+0.026, +0.342]	-0.027 [-0.219, +0.100]
Tuning examples	4,000	+0.385 [+0.338, +0.425]	-0.017 [-0.058, +0.015]
Tuning examples	3,000	+0.165 [+0.027, +0.348]	+0.116 [+0.033, +0.199]

Table 7. Descriptive bootstrap intervals for the mean-view PSNR difference, using 10,000 object resamples. These convenience cohorts do not justify population-wide confidence guarantees.

belong to the new Objaverse panel; the remaining cache failure is an official tuning example. On GSO, the missing smaller-budget Spider-Man fit fails CUDA initialization before training begins; it is an environment failure, not evidence of an optimization breakdown. The booster-seat case instead has a nonfinite larger-budget export and a failed smaller-budget optimization. No selected-case repair changes these confirmation endpoints. The machine-readable ledger distinguishes these causes and records every planned training arm.

A.5. Does the comparison survive coding?

This completion evaluates the original seed-zero fields from the frozen expanded cohort. All 71 admitted objects remain in the coverage ledger; 59 have a valid reference. Across those objects, the available fields comprise 59 references, 118 extended-direct fields and 103 refined fields. Missing or invalid refinements remain missing or invalid. No model is retrained, repaired or selected using its coded quality.

Fixed coding and evaluation. We apply three existing settings of the same scalar codec to every available field, for 840 prescribed coded endpoints. Colour coefficients

use eight bits in all settings. Barycentric coordinates and each response parameter use 12/8, 14/10 and 16/12 bits in the standard, higher and highest settings, respectively. The standard setting is the previously reported codec and is the primary comparison; the other two test sensitivity to precision. Per-field ranges, transformations and container compression follow the same implementation for all routes. Complete bitstreams, including metadata and indices, are charged. The completed ledger contains 840/840 audited coded endpoints: 162 reused and 678 newly evaluated. Evaluation decodes those bitstreams and retains the stored surface queries, five-view rendering protocol and metrics. Existing standard-setting reference/refined results are reused only after source fields, query coordinates, decoded arrays and evaluation settings agree. Every new evaluator process also renders the original uncoded reference as a calibration check. These checks do not turn concurrent evaluation times into controlled latency measurements.

Paired fitting routes. Refinement and extended direct each receive two stages of 5,000 source updates; editing remains an additional cost for refinement. Their attained counts match on all 103 available paired comparisons. The

Route	Valid	Env.	Fit fail	Invalid	Ineligible	Blocked
Reference	75	0	1	0	0	15
Direct 4k	75	0	0	0	0	16
Direct 3k	75	0	0	0	0	16
Refined 4k	64	0	0	1	10	16
Refined 3k	67	1	1	0	6	16
Full continuation	75	0	0	0	0	16
Extended direct 4k	75	0	0	0	0	16
Extended direct 3k	75	0	0	0	0	16

Table 8. Training-arm accounting over all 91 prescribed object–seed slots. Valid means a finite field with completed evaluation, including exact-match image views. Blocked arms lack a required cache, reference or initializer; they are not additional independent failures. Fit failures include initialization failures. The original environment and numerical failures remain missing outcomes.

Repeated object	Ref. 4k	Extended 4k	Ref. 3k	Extended 3k
Calabash - Bottle Gourd	+0.002 (3/3)	-0.004 (3/3)	+0.002 (3/3)	-0.008 (3/3)
Table	– (0/3)	– (0/3)	– (0/3)	– (0/3)
PS1 Chicken	-0.016 (3/3)	-0.009 (3/3)	-0.050 (3/3)	-0.021 (3/3)
Grosz z 1534 roku	+0.356 (3/3)	+0.069 (3/3)	+0.009 (3/3)	-0.027 (3/3)
Shovel 1	– (0/3)	– (0/3)	– (0/3)	– (0/3)
Cartoon Fish	+0.521 (3/3)	+0.004 (3/3)	+0.375 (3/3)	-0.015 (3/3)
Pipe	– (0/3)	– (0/3)	– (0/3)	– (0/3)
Pipe 01	+0.031 (3/3)	+0.002 (3/3)	-0.018 (3/3)	+0.014 (3/3)
Reebok shoe	+0.010 (3/3)	+0.013 (3/3)	-0.027 (3/3)	-0.013 (3/3)
Garden Swing	+0.086 (3/3)	+0.029 (3/3)	-0.023 (3/3)	+0.108 (3/3)

Table 9. Object means over the three prescribed seeds. A mean is reported only when all three paired differences are finite; parentheses show valid coverage. Missing or budget-ineligible cases are not replaced. Seed0 cohort averages and these object-level repeat means answer different questions.

corresponding new-Objaverse panel contains 23/26 objects at 4,000/3,000 kernels. Any subsequent coding or evaluation failures remain in the endpoint ledger and reduce the measured denominators rather than trigger replacement. Counts and coded bytes are distinct: even the same count and precision can produce different file sizes. Table 12 therefore reports quality and actual bytes separately. All settings and cohorts are retained. Object bootstrap intervals are descriptive and do not establish population-level significance. As a post-hoc distribution check, the largest positive difference in the standard-coded, 3,000-kernel new-Objaverse panel is +1.644 dB. The remaining 25 objects average +0.002 dB. This sensitivity check removes no object from the reported means or intervals; it cautions against interpreting the small cohort mean as a consistent advantage.

A finite candidate envelope. For each standard-coded refined field, we also ask what quality the measured direct candidates reach without exceeding its actual file size. Candidates are the two direct count caps at all three fixed precisions. We require all six candidates to have been evaluated, then minimize pooled image MSE among those fitting the byte budget. This is a retrospective envelope of measured

points, not a learned selection rule, an interpolated operating point or a global Pareto frontier. No feasible candidate and an incomplete candidate set are distinct outcomes. Zero MSE is retained rather than dropped as infinite PSNR.

Across the standard-coded targets, 0 candidate sets are incomplete. Such cases are excluded from quality comparisons and never counted as wins. The public object-level records also report pooled MSE/PSNR, SSIM, LPIPS, surface scores, quantization changes from uncoded fields and qualification outcomes. They preserve the distinction between a finite mean-view PSNR pair and a valid exact-match measurement.

Repeated-reference calibration. The unchanged parent is rendered in each object–precision task and compared with its original measurement. Table 14 reports the observed absolute PSNR changes. These are diagnostics of repeated evaluation, not statistical error bars or a universal noise threshold. We do not subtract the drift from coded scores. The standard setting combines verified historical reference/refined measurements with newly evaluated direct fields; it is not a same-process remeasurement of every candidate. Tiny signed differences should be interpreted with

Cohort	Budget	Edited reference	Rebuilt-support refit
New Objaverse	4,000	+1.387 (23)	+1.384 (23)
New Objaverse	3,000	+2.701 (26)	+2.368 (26)
Public main	4,000	+2.241 (8)	+2.465 (8)
Public main	3,000	+3.979 (9)	+3.768 (9)
Public low-poly	4,000	+0.861 (7)	+0.910 (7)
Public low-poly	3,000	+2.557 (7)	+1.975 (7)
GSO	4,000	+1.305 (7)	+1.327 (7)
GSO	3,000	+3.835 (6)	+3.556 (6)
Tuning examples	4,000	+2.362 (4)	+2.403 (4)
Tuning examples	3,000	+3.977 (4)	+3.826 (4)

Table 10. Quality recovered by compress–refine over the two post-hoc colour-fitting controls, in mean-view PSNR dB (finite paired counts). Post-hoc fitting distills the reference; refinement revisits source appearance and uses additional optimization, so these margins do not isolate a geometry effect.

Cohort	Budget	Objects	Coded reference KiB	Coded refined KiB	PSNR Δ
New Objaverse	4,000	23	51.37	42.51	+0.317 (23)
New Objaverse	3,000	26	51.11	32.36	+0.253 (26)
Public main	4,000	8	54.23	44.38	+0.391 (8)
Public main	3,000	9	53.72	33.62	+0.200 (9)
Public low-poly	4,000	8	50.91	41.80	+0.111 (8)
Public low-poly	3,000	8	50.91	31.89	-0.045 (8)
GSO	4,000	7	53.87	44.73	+0.331 (7)
GSO	3,000	6	53.85	33.76	+0.223 (6)
Tuning examples	4,000	4	52.99	43.60	+0.466 (4)
Tuning examples	3,000	4	52.99	33.09	+0.239 (4)

Table 11. The same scalar codec applied to paired reference and refined fields. Sizes are medians of actual bitstreams on the indicated common objects. Quality uses supplementary pooled PSNR; its finite count is shown in parentheses. This tests count reduction alongside coding, not superiority over every coded training route.

this scope in mind. Ground-truth images are bitwise identical in 177/177 repetitions. Parent images need not be: the maximum observed UInt8 component difference is 10. Small changes in an average image score do not imply pixelwise identity.

Table 12. Refined minus equally extended direct after decoding. n gives finite mean-view PSNR pairs/all valid pairs; wins count positive differences among finite pairs. Byte change is the mean of per-object relative bitstream changes on all valid pairs, independently of finite PSNR. Intervals are object-bootstrap 95% intervals. Negative byte change means the refined field is smaller.

Budget	Precision	n	Mean dB	Median dB	95% interval	Wins	Bytes %
<i>New Objaverse</i>							
4,000	Standard	23/23	-0.062	-0.024	[-0.120, -0.013]	7/23	+0.63
4,000	Higher	23/23	-0.063	-0.029	[-0.119, -0.016]	8/23	+0.60
4,000	Highest	23/23	-0.065	-0.025	[-0.120, -0.017]	8/23	+0.69
3,000	Standard	26/26	+0.065	+0.007	[-0.050, +0.218]	15/26	+0.45
3,000	Higher	26/26	+0.069	+0.008	[-0.046, +0.222]	15/26	+0.52
3,000	Highest	26/26	+0.068	+0.007	[-0.047, +0.221]	15/26	+0.73
<i>Public main</i>							
4,000	Standard	8/8	+0.059	+0.026	[-0.017, +0.158]	7/8	+0.58
4,000	Higher	8/8	+0.059	+0.024	[-0.017, +0.154]	6/8	+0.59
4,000	Highest	8/8	+0.056	+0.024	[-0.019, +0.153]	5/8	+1.05
3,000	Standard	9/9	+0.088	+0.032	[+0.018, +0.165]	6/9	+0.35
3,000	Higher	9/9	+0.089	+0.053	[+0.024, +0.160]	7/9	+0.40
3,000	Highest	9/9	+0.091	+0.048	[+0.025, +0.162]	7/9	+0.62
<i>Public low-poly</i>							
4,000	Standard	7/8	+0.052	+0.026	[-0.010, +0.137]	5/7	+0.95
4,000	Higher	7/8	+0.039	+0.024	[-0.035, +0.124]	5/7	+0.91
4,000	Highest	7/8	+0.041	+0.026	[-0.035, +0.126]	5/7	+1.02
3,000	Standard	7/8	+0.019	+0.049	[-0.029, +0.066]	4/7	+0.33
3,000	Higher	7/8	+0.022	+0.055	[-0.027, +0.069]	4/7	+0.26
3,000	Highest	7/8	+0.029	+0.055	[-0.019, +0.077]	4/7	+0.51
<i>GSO</i>							
4,000	Standard	7/7	-0.017	+0.008	[-0.097, +0.043]	4/7	+0.62
4,000	Higher	7/7	-0.019	-0.003	[-0.101, +0.043]	3/7	+0.65
4,000	Highest	7/7	-0.019	-0.004	[-0.100, +0.042]	3/7	+0.86
3,000	Standard	6/6	-0.030	+0.048	[-0.226, +0.103]	4/6	+0.59
3,000	Higher	6/6	-0.032	+0.051	[-0.227, +0.098]	4/6	+0.74
3,000	Highest	6/6	-0.030	+0.050	[-0.225, +0.100]	4/6	+1.06
<i>Tuning examples</i>							
4,000	Standard	4/4	-0.015	-0.005	[-0.053, +0.014]	2/4	+0.54
4,000	Higher	4/4	-0.016	-0.006	[-0.056, +0.015]	2/4	+0.53
4,000	Highest	4/4	-0.016	-0.008	[-0.054, +0.015]	2/4	+0.68
3,000	Standard	4/4	+0.118	+0.114	[+0.038, +0.198]	4/4	+0.31
3,000	Higher	4/4	+0.118	+0.112	[+0.039, +0.198]	4/4	+0.44
3,000	Highest	4/4	+0.119	+0.113	[+0.039, +0.199]	4/4	+0.69

Cohort	Budget	Compared	Refined better	Equal	Direct better	No fit
New Objaverse	4,000	23	9	0	14	0
New Objaverse	3,000	21	12	0	9	5
Public main	4,000	8	4	0	4	0
Public main	3,000	7	5	0	2	2
Public low-poly	4,000	8	6	0	2	0
Public low-poly	3,000	5	4	0	1	3
GSO	4,000	7	3	0	4	0
GSO	3,000	5	3	0	2	1
Tuning examples	4,000	4	2	0	2	0
Tuning examples	3,000	4	3	0	1	0

Table 13. Quality at no greater measured bytes, using pooled image MSE. “No fit” means none of the six tested direct bitstreams fits the refined field’s byte budget; it is not a quality win. The comparison covers only these discrete candidates.

Cohort	Finite repetitions	Mean absolute dB	Maximum absolute dB
New Objaverse	87	0.0002	0.0043
Public main	30	0.0003	0.0015
Public low-poly	21	0.0001	0.0010
GSO	24	0.0004	0.0029
Tuning examples	12	0.0002	0.0006

Table 14. Repeated uncoded-reference measurements across all three settings. Repetitions of an object are not independent objects. Exact-view cases remain in the coverage and pooled-MSE records even when a PSNR difference is undefined.

B. Conditional identities and their limits

B.1. Projection and a bounded empirical gap

With fixed support and background, AA^+ is the orthogonal projector onto the range of A . Decomposing each target channel into that range and its orthogonal complement proves Eq. (4). Bounding the coefficients restricts the feasible set and cannot lower the unrestricted minimum. If the background is optimized as an unrestricted variable, append a constant column and use non-subtracted targets. If the background has its own bounds, its linear minimization must use those bounds instead of the colour box.

For differentiable convex F , convexity gives $F(U) \geq F(C) + \langle \nabla F(C), U - C \rangle$. Minimizing the affine expression over $U \in [-1, 1]^{G \times 3}$ gives $F^* \geq F(C) - \gamma(C)$ with the gap in Eq. (5). The feasible current point gives $F^* \leq F(C)$. Numerical gaps here use floating-point arithmetic, not interval certificates. A small gap establishes near-optimality only for the specified empirical convex objective.

The separable update follows because $D - H$ is positive semidefinite. The resulting quadratic upper model touches F at the current point; its box minimizer decreases that model and hence cannot increase F in exact arithmetic. A zero entry of d implies the corresponding nonnegative response column is zero. Our acceleration safeguard checks the actual quadratic before accepting a proposal. These claims concern an unclipped loss; clipping makes the target problem different and the projection bound need not transfer.

B.2. Why a good direction can have a bad endpoint

Even a convex one-dimensional loss can decrease initially and rise after an excessive step. In a nonconvex texture fit, the conditional direction is additionally estimated from particular samples. For an optimizer such as Adam, the first bias-corrected scalar update is approximately $-\eta g / (|g| + \epsilon)$: a small gradient magnitude alone does not ensure a small parameter displacement. This motivates measuring changes to the field, but does not determine an optimal learning rate.

Changing sampling probabilities without correcting the loss weights changes the fitted objective. Inverse-probability weighting preserves the conditional expected

gradient for the stated target measure under its support assumptions. It does not make Adam trajectories or rendered quality unbiased. Consequently, changes in sample identities, sample counts, target filtering and regularization must be identified separately rather than described as one universal sampling improvement.

B.3. Truncation with a changed normalizer

Let $u_i, v_i \geq 0$ be original and deployed readout weights padded to a common candidate set. They include each readout's selection: unselected terms are zero. Set

$$d = \max(1, \sum_i u_i), \quad d' = \max(1, \sum_i v_i),$$

$$N = \sum_i u_i c_i, \quad N' = \sum_i v_i c_i.$$

With the same background and clipping, adding and subtracting N'/d and using nonexpansiveness of clipping gives

$$\|f - f'\|_\infty \leq \frac{\|N - N'\|_\infty}{d} + \frac{\|N'\|_\infty |d - d'|}{dd'}. \quad (8)$$

If $E = \sum_i |u_i - v_i|$ and $C_{\max} = \max_i \|c_i\|_\infty$, then $\|N - N'\|_\infty \leq C_{\max} E$, $|d - d'| \leq E$ and $\|N'\|_\infty / d' \leq C_{\max}$. Thus $\|f - f'\|_\infty \leq 2C_{\max} E / d$. If both raw sums are at most one, the sharper numerator-only bound $C_{\max} E$ applies. Keeping the original denominator is another, different contract. An early predictor has no certified bound on E in this study; its predicted tail is an empirical heuristic.

On one fixed distribution, let $L = \mathbb{E}\|f - Y\|_2^2/3$ and $D = \mathbb{E}\|f' - f\|_2^2/3$. Expansion and Cauchy-Schwarz give

$$L' - L \leq 2\sqrt{LD} + D. \quad (9)$$

The exact cross term can have either sign. For $L > 0$, limiting PSNR loss to η requires $L'/L \leq 10^{\eta/10}$ on that same distribution. An absolute RGB tolerance is not this relative-error condition. Visibility, illumination, filtering, sRGB conversion and denoising also intervene before the measured image. No global stability theorem for that entire renderer is asserted.

B.4. Conditional optimization and stability details

The following facts are classical optimization and perturbation arguments. They specify conditional diagnostics, feasibility assumptions and stability limits for the tested field.

A computable conditional optimization gap. Include a background column in A , and let $\mathcal{B} = [\ell, u]$ impose the deployed colour and background bounds. In this paragraph, Y denotes the original targets before background subtraction, and the augmented C includes the background coefficients. On a fixed sample matrix define $f(C) = \|AC - Y\|_F^2 / (2md)$, with $d = 3$ channels, and $G = A^\top (AC - Y) / (md)$. Minimize the linear form $\langle G, V \rangle$ over $V \in \mathcal{B}$ by choosing each lower endpoint when its gradient is nonnegative and its upper endpoint otherwise. For a feasible $C \in \mathcal{B}$, convexity gives

$$0 \leq f(C) - f^* \leq g(C) := \langle G, C - V \rangle, \quad (10)$$

$$f^* \geq \max\{0, f(C) - g(C)\}.$$

Indeed $f(C^*) \geq f(C) + \langle G, C^* - C \rangle$ and the endpoint choice minimizes the last inner product. This is the usual conditional-gradient gap; it requires neither a nonsingular Gram matrix nor an exact linear solve. For a feasible quantized export $\hat{C}$, the same lower bound L also gives $f(\hat{C}) - f^* \leq f(\hat{C}) - L$. Thus one can separate solver error from serialization error without claiming that rounded coefficients solve the discrete quantization problem. Floating-point evaluations of these expressions are numerical diagnostics, not interval-certified proofs.

Column equilibration, $D_{ii} = \max(\|A_{:,i}\|_2, 10^{-8})$ and $Z = DC$, changes the design matrix to AD^{-1} and the bounds to $D\ell \leq Z \leq Du$. It preserves the feasible predictions in real arithmetic. Our experiment combines this change with objective rescaling, different stopping settings and a larger iteration cap; it does not isolate their individual contributions. A small conditional gap says nothing about unsampled surface points, clipped objectives or final image metrics.

Why clipping changes the question. For background $b = 1/2$, normalized responses $A = (1/2, 1)^\top$ and original targets $(1, 1)^\top$, the background-subtracted target is $Y = (1/2, 1/2)^\top$. The unrestricted projection residual is $1/20$, while the feasible colour $C = 1$ gives zero error after clipping $b + AC$ to $[0, 1]$. Thus Eq. (4) cannot lower-bound clipped colour fitting. Nor does improving a fixed training matrix guarantee improvement on a validation matrix. Optimizing geometry while eliminating linear colour variables is a form of variable projection [16]; its smooth local formulation needs appropriate rank and support regularity. Top- k switches and changes in active colour bounds require sepa-

rate treatment. We have not implemented a globally convergent geometry optimizer of this form.

A global upper model for the clipped loss. The preceding gap bounds suboptimality of an *unclipped* convex quadratic. For the clipped objective we instead use a touching upper model, an application of classical majorization and CCCP [30]. It certifies empirical descent, not distance to a clipped global optimum.

Proposition 1 (Clipped-square upper model). *For $y \in [0, 1]$, let $\ell_y(z) = (\text{clip}_{[0,1]}(z) - y)^2$ and $q_y(z) = (\text{clip}_{[0,1]}(z) - y)\mathbf{1}_{[0,1]}(z)$. For every real z and δ ,*

$$\ell_y(z + \delta) \leq \ell_y(z) + 2q_y(z)\delta + \delta^2. \quad (11)$$

The bound is global, including steps that cross clipping boundaries.

Proof. The continuous function $h_y(z) = \ell_y(z) - z^2$ has derivative $-2z$ below zero, $-2y$ inside $(0, 1)$, and $-2z$ above one. Its derivative is nonincreasing, including downward jumps at zero and one since $0 \leq y \leq 1$. Thus h_y is concave. The value $2q_y(z) - 2z$ is a valid supergradient, also at both boundaries. Apply its supporting-line inequality and expand $(z + \delta)^2$. $\square$

Let A include the background column, $z_0 = AC_0$, $q_0 = (\text{clip}(z_0) - Y)\mathbf{1}_{[0,1]}(z_0)$ and $T = z_0 - q_0$. Using full MSE $F(C) = \text{mean}[(\text{clip}(AC) - Y)^2]$ over query-channel entries, Eq. (11) gives

$$F(C) \leq U(C | C_0) = F(C_0) + \text{mean}[(AC - T)^2 - q_0^2]. \quad (12)$$

This touches at C_0 . Any box-feasible nonincrease of $\|AC - T\|_F^2$ therefore cannot increase F ; an exact surrogate optimum is unnecessary. We retain the full response matrix, including saturated rows, and recompute T at each outer iteration. Strictly saturated entries ($z_0 < 0$ or $z_0 > 1$) have pseudo-target z_0 , which must not be clipped; at the exact boundaries, $T = Y$. Inner boxed coordinate updates start at the current field.

For a feasible endpoint C_1 , set $d = A(C_1 - C_0)$, $a = 2\text{mean}(q_0 d)$ and $b = \text{mean}(d^2)$. The same bound becomes $F(C_0 + \alpha(C_1 - C_0)) \leq F(C_0) + a\alpha + b\alpha^2$. For $b > 0$ its minimum on $[0, 1]$ is $\alpha = \text{clip}_{[0,1]}(-a/(2b))$; use zero when $b = 0$. At clipping boundaries, the supporting linear coefficient a need not equal the one-sided directional derivative in Eq. (17). Applied to validation responses, this rule protects only that validation sample, not automatically the fitting sample as well. We recheck the claimed empirical objective after FP32 export and reject increases. Neither argument guarantees fresh-test or rendered improvement, monotonicity after geometry changes, or a global clipped optimum: a strictly saturated scalar start can have $q_0 = 0$ and stall despite a better feasible colour.

Residual benefit and allocation. For a fixed base B , target Y and additive residual R , set $E = Y - B$. The reduction in unclipped squared error is exactly

$$\Delta = 2\langle E, R \rangle - \|R\|_F^2. \quad (13)$$

For a single scalar gate $\alpha \in [0, 1]$ and $R \neq 0$, its best value on these samples is $\text{clip}_{[0,1]}(\langle E, R \rangle / \|R\|_F^2)$. If $B, Y \in [0, 1]$, clipping the final sum cannot increase its distance to Y , so $\Delta > 0$ is sufficient for a clipped improvement on these same samples. It does not imply a rendered or byte-matched improvement.

With unrestricted coefficients, adding one fixed column a to a fixed linear dictionary A improves its best fit by $\|q^\top E\|_2^2 / \|q\|_2^2$, where $q = (I - AA^\top)a$ and $E = (I - AA^\top)Y$, provided $q \neq 0$ (zero otherwise). This follows by minimizing $\|E - qc\|_F^2$ over a row vector c . It motivates accounting for residual alignment and redundancy, rather than base error alone. Bounded colours, normalization and support reselection can invalidate this fixed-dictionary formula; our hybrid therefore rebuilds its actual responses after selecting support. A proposed gain-per-byte rule must also beat spending those bytes on the base, as the negative hybrid control shows.

Two discrete selection layers. An outer nearest-candidate set and an inner filtered top- k set both need stability. A distance gap larger than twice a uniform distance-error bound preserves the outer cutoff; the analogous weight gap preserves the inner cutoff. These statements assume a fixed candidate universe and consistent tie-breaking. An inner margin alone cannot certify batch-compacted outer search. For nonnegative selected weight vectors u, v , let $D_u = \max(1, \|u\|_1)$ and similarly for v . If $|c_i| \leq 1$, then

$$\left| \frac{u^\top c}{D_u} - \frac{v^\top c}{D_v} \right| \leq 2\|u - v\|_1. \quad (14)$$

The numerator difference contributes at most $\|u - v\|_1 / D_u$; the denominator contribution is at most $|D_u - D_v| / D_u$, using $|v^\top c| \leq D_v$. Both denominators are at least one. With identical selected indices and per-weight error ϵ , this yields $2k\epsilon$. Across a cutoff switch, small raw response error need not give small selected-vector error: weights $(1/2 + \epsilon, 1/2)$ and $(1/2, 1/2 + \epsilon)$, colours $(1, -1)$, background $1/2$ and $k = 1$ produce nearly unit RGB difference after clipping as $\epsilon \rightarrow 0$.

What a heat-operator approximation would need to preserve. In mass-normalized coordinates, let $L, \tilde{L}$ be symmetric positive semidefinite discrete generators. Duhamel's identity and contraction of both semigroups give

$$\|e^{-tL} - e^{-t\tilde{L}}\|_2 \leq t\|L - \tilde{L}\|_2, \quad t \geq 0. \quad (15)$$

This bound concerns raw thermal operators, not the complete texture evaluator. If an algorithm returns an approximation $\hat{H}_t$ rather than $e^{-t\tilde{L}}$, the triangle inequality additionally charges $\|e^{-t\tilde{L}} - \hat{H}_t\|_2$. Random-feature, low-rank and polynomial approximation errors therefore need their own bounds or measurements; the generator bound alone does not certify them. Self-normalization requires a positive lower bound on its divisor; filtering can amplify response errors; the two selection margins above are then needed. For example, if both divisors exceed $a > 0$ and the reference numerator has magnitude at most H , division contributes at most $\delta_h/a + H\delta_d/a^2$ for numerator/divisor errors δ_h, δ_d . These quantities are not bounded globally by our data.

An orthogonal basis rotation $\Phi' = \Phi Q$ must transform the represented generator to $Q^\top \Lambda Q$ to preserve $\Phi e^{-t\Lambda} \Phi^\top$; retaining diagonal Λ works only when the rotation commutes with it. The aligned-basis issue is already discussed by Accelerating HKTex [18]. Also, a convex mixture of thermal responses is generally not the semigroup of one interpolated generator, even when each component is valid. Neither observation proves that the learned texture representation is invalid. They identify contracts for future local or reduced-order evaluators. Such an evaluator must be evaluated jointly with training, memory, and support stability before claiming composition with our budgeted fields.

Source normalization is not a global maximum guarantee. For a positive-semidefinite feature kernel K_t and $K_t(p, p) > 0$, Cauchy-Schwarz gives

$$\left| \frac{K_t(q, p)}{K_t(p, p)} \right| \leq \sqrt{\frac{K_t(q, q)}{K_t(p, p)}}. \quad (16)$$

Thus source-self normalization sets the source response to one (ignoring a numerical stabilization), but alone does not establish a global unit maximum. A 12-vertex, 10-triangle rectangular strip with standard linear-FEM stiffness, lumped mass and the full heat spectrum at $t = 0.1$ gives a ratio 1.0977 away from the source; the construction is retained in the artifact. Symmetric normalization by $\sqrt{K_t(q, q)K_t(p, p)}$ bounds the absolute response by one when both diagonals are positive. It changes the kernel, does not ensure entrywise positivity for a truncated feature approximation, and does not preserve the learned texture automatically. Our frozen-field check finds above-one source-normalized responses on all nine objects; substituting the symmetric form without retraining lowers native-surface PSNR on all nine. Neither observation invalidates the fitted texture or establishes that retraining with a different normalization is unhelpful.

C. Correction, continuation and response structure

C.1. Follow-up protocols

The four warm-start development objects are historical diagnostic fields, not fresh confirmations. Twenty-four continuation controls compare colour-only, non-colour-only and joint updates from matched starts. After 1,000 updates, the guarded quadratic correction with moment resets gains 0.605 dB over ordinary joint continuation. Matched resets alone change quality by 0.020 dB, leaving a 0.586 dB correction-minus-reset contrast. With geometry frozen this contrast is 0.146 dB. The interaction supports subsequent geometry adaptation in these cases, without a claim of causal mediation. Only three objects qualify for the separate common-budget timing comparison; its 1.058 dB gain is not a four-object cold-pipeline result.

The reserved comparison instead begins with continuous direct fitting. The four corrections occur at updates 4,000, 4,250, 4,500 and 4,750. All three arms use the same one-million-point native training pool, 4,096-point batches, and matched starting parameters. Guard points are disjoint from fitting points but belong to that training pool. Each admitted object runs three seeds and three arms. The 19 admitted objects comprise 12 Objaverse and seven GSO objects; dataset mean correction gains are 0.022 and 0.015 dB. Five configuration/preparation failures remain in the 24-object declaration.

The local-response comparison has four objects, three seeds and three arms, for 36 qualified runs. It omits parts of author-system adaptive density control and proxy approximation. Twenty-one original failed attempts remain separate from the uniform validity rerun. Its roughly 0.006 dB mean gain with two gains and two losses does not establish a useful LocalHK composition.

C.2. Intervene, permit compensation, then inspect the solve

Orange, SmallLimo, Hummingbird and Piggybank provide four fixed 3,000-kernel fields. Counts, centres, thresholds and sharpness remain fixed. Separate interventions minimize anisotropy, remove biharmonic post-diffusion weighting while retaining intrinsic search, redistribute mass uniformly within positive support, keep only the largest response at the same row mass, or remove the background. All arms receive 1,000 bounded colour sweeps and the same calibrated FP16 step menu $\{0, \frac{1}{4}, \frac{1}{2}, 1\}$.

All four objects lose held-out surface quality under each intervention. Mean losses for the four response changes are 3.625, 5.725, 6.391 and 11.369 dB. For each, the unclipped empirical lower bound exceeds the original’s feasible fitting loss on all objects. This limits an unfinished-solve explana-

tion for that sampled objective. It does not prove that the alternative would fail after full geometric retraining.

The original HKTex ablation already reports nearly neutral removal of pruning or colour-formation KNN in its recipe [14]. Removing error-based insertion reduces both quality and final kernel count. Our conditional experiment asks a different question and should not be restated as a rerun of that ablation.

Background is not always redundant. Write $a_i = w_i / \max(1, \sum_j w_j)$ and $\beta = \sum_i a_i \leq 1$. Before clipping,

$$b + \sum_i a_i c_i = (1 - \beta)b + \sum_i a_i (b + c_i).$$

If $\beta = 1$ and coefficients are unrestricted, the background can be absorbed. Otherwise the deficit term remains, and the shifted coefficients may violate their box. On Orange, unrestricted absorption loses 4.414 dB despite only 1.77% of sampled raw masses being below one. Near-unit average coverage is not sufficient to remove the term. The effective values $b + c_i$ can lie outside the RGB cube, so this identity does not turn the original field into a bounded probabilistic MoE.

C.3. Historical conditional fits and stronger controls

These studies reuse the historical objects and precede the reserved comparisons. Their sample counts, fitting states and render protocols are kept distinct.

How much conditional colour optimization remains?

We fit colours and background on every corrected 3,000-kernel deletion support, copying every non-colour array exactly. The normalized response matrix uses 16,384 new area-uniform training queries; 4,096 disjoint validation queries select among the initializer, unrestricted fitting followed by clamping, bounded linear fitting, and clipped-objective refinement. Selection occurs after rounding to the export dtypes. Native source texture supplies targets; the stored evaluation queries are disjoint. Final PBR evaluation uses the main eager renderer. This experiment has different samples, importance weighting and compute from joint refinement and is not a matched causal geometry ablation.

Column equilibration, objective rescaling, and a larger iteration budget substantially tighten the bounded linear solver’s conditional-gradient gap (Appendix B.4); this is not an isolated ablation of equilibration. The gap bounds suboptimality of the continuous, unclipped, finite-query box objective. Reusing its lower bound also bounds the exported linear solution’s error; it does not certify the validation-selected candidate or the clipped objective. Across nine objects, the pre-export relative gap has median 0.034% and

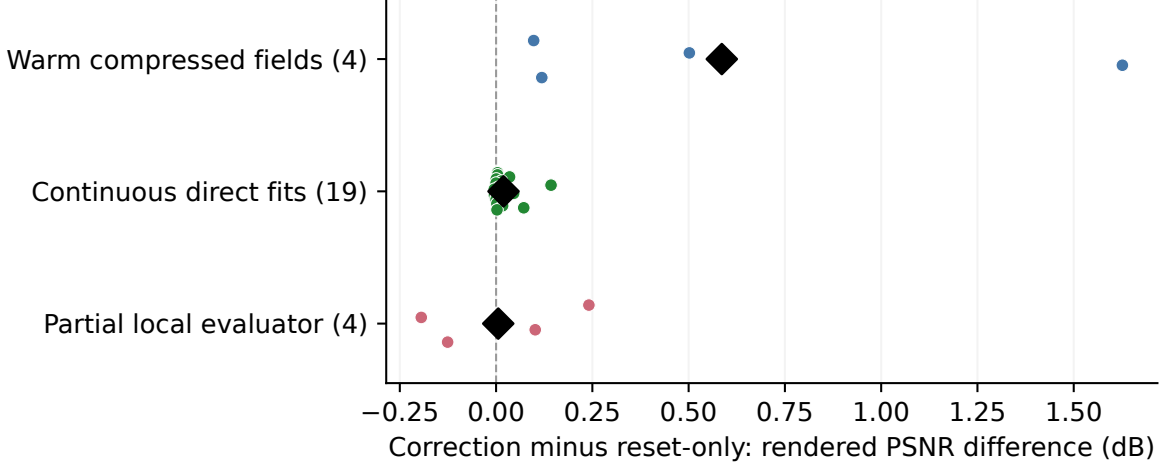


Figure 7. Completed correction follow-up. The reserved direct-training cohort and the four-object partial local-evaluator study have distinct scopes. The small endpoint effects do not establish an acceleration.

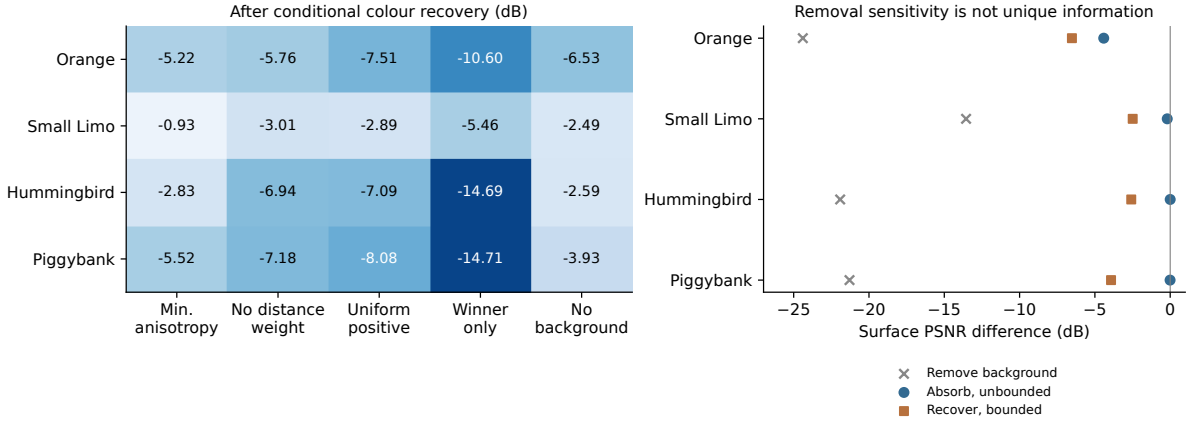


Figure 8. Conditional interventions after equal colour-recovery budgets. These are fixed-field surface tests, not original-paper retraining ablations or PBR measurements.

maximum 0.347%; the exported box solution’s relative sub-optimality bound is at most 0.348%. These are floating-point diagnostics, not interval certificates. The box solution improves validation MSE on 8/9 objects; validation retains the initializer on 1/9. Under this 16k protocol, corrected refinement exceeds the selected colour fit on 9/9 rendered objects, by 2.47 dB on average. The stronger sampling control below changes the per-object comparison. The result separates numerical solution quality from generalization and rendering quality. Better conditioning removes one solver confound; it does not prove a global placement bottleneck. The exported models, stopping status, unscaled solver results, gap values and validation decisions are retained.

A stronger sampling control changes the conclusion. We additionally fit the same frozen deletion supports us-

ing a fresh nested 16,384/65,536-point training split and a shared 4,096-point validation set. For each sample count, validation selects among the initializer and four bounded Tikhonov fits with $\lambda \in \{0, 0.001, 0.01, 0.1\}$, minimizing $\|AC' - Y\|_F^2 + \lambda \|D(C - C_0)\|_F^2$ with training-column norms in D . All non-colour arrays remain byte-identical; selection uses exported coefficients. This is an adaptive extension from four diagnostic objects to all nine, with additional samples and compute, not an independent confirmation cohort or a matched geometry ablation. Solver termination records are retained; no optimality certificate is asserted for this grid. The larger-sample recipe improves rendered PSNR on 9/9 objects, by 0.65 dB on average. It exceeds corrected refinement on SmallLimo and TShirt; joint refinement retains a 1.57 dB mean advantage and wins 7/9 objects. Thus a tight finite-training-set solver gap does not identify the remaining

Cohort	Objects	Mean Δ PBR	Median	Positive
Objaverse	12/16	0.02171	0.00386	8/12
GSO	7/8	0.01517	0.00641	7/7
Combined qualified	19/24	0.01930	0.00395	15/19

Table 15. Completed correction-minus-reset comparison after continuous direct fitting, in dB. Failed preparation/configuration objects remain in coverage denominators; no gains are imputed for them. Positive signs do not imply perceptually meaningful improvements.

Public object	Δ PBR	Δ native	Δ LPIPS
Main A	0.24098	0.21078	-0.000412
Main B	-0.19425	-0.11725	0.000576
Low-poly A	0.10165	0.47630	-0.000592
Low-poly B	-0.12584	-0.13121	-0.000347
Mean	0.00564	0.10965	-0.000194

Table 16. Correction minus reset-only under the partial local evaluator, averaging three seeds per object. Main A/B and low-poly A/B follow the fixed order of the corresponding public cohorts. These mixed results do not establish stable composition gains or author-system speedups.

population error as geometric. Sample coverage, regularization, and readout optimization can explain part of the warm benefit. The earlier 9/9 comparison is specific to its 16k fitting protocol, not a lower bound on all colour-only repairs. This distinction is central to interpreting the projection argument.

Subsequent short-training and sampling controls are reported in Appendix C.4.

C.4. Initialization and sampling follow-up

A final exploratory extension separates readout initialization, continuation step size, and target sampling. On Orange, PoisonDartFrog, SmallLimo and TShirt, we compare 1,000 fixed-count joint updates from the corrected deletion and from the 65k colour fit of Table 18. The parent remains seed 0; calibration data and computation are additional costs. The original trainer’s UV sampling targets and the native bitmap used by our stronger colour fits and PBR differ in filtering and texel-coordinate conventions. We cross both initializers with both target conventions, retaining the same pool geometry and starting RNG states; later priority feedback can diverge. The stronger initialization improves all four recorded 1,000-step PBR results under each convention, by 0.178/0.297 dB on average. This does not establish a universal initializer: SmallLimo and TShirt still lose quality relative to their own strong starting fields under the default restart. Reducing all optimizer learning rates to one tenth changes that comparison. Table 19 gives the stronger comparison: the learning rate is chosen independently for each initializer by native validation MSE. The additional SmallLimo seed changes continuation randomness, not the parent or calibrated initializer. These adaptive diagnostics do not establish a population effect or an optimal learning-

rate rule. At the shared small rate, deletion reaches only 40.49/40.53 dB on SmallLimo; those weaker baselines are not used in the selected comparison above. An additional SmallLimo continuation seed with the colour fit and small rate reaches 48.30/48.35 dB under Orig/Native targets. Its matched small-rate deletion controls are retained, but this extra seed does not repeat a fully tuned grid.

An elementary explanation is that Adam’s bias-corrected first scalar step is $-\eta g/(|g| + \epsilon)$: small residuals do not ensure small parameter steps. For $f(x) = h(x - x^*)^2/2$, a fixed sign step decreases the loss only when $0 < \eta < 2|x - x^*|$. This is an illustrative mechanism, not a causal proof for our nonlinear trajectories. In fact, SmallLimo’s default continuation lowers uniform native validation MSE from about 0.002246 to 0.002043 while reducing PBR PSNR. Surface-average descent and rendered improvement are different criteria; this case cannot be diagnosed as scalar-objective overshoot alone. The two-object 5,000-step and simple variance/ affine-calibration controls are retained with all outcomes in the artifact.

We also compare 16,384-draw sketches of the fixed 65,536-point colour-fitting pool on Orange, PoisonDartFrog and SmallLimo, with two sampling seeds. Both adaptive distributions mix half uniform mass and apply $1/(Np_i)$ correction: one uses initial gradient norms in equilibrated coordinates, the other uses a target-independent row-energy coverage proxy. The latter is not an exact leverage score. Both improve PBR in all six comparisons with uniform draws; mean gains are 0.363/0.445 dB. The same gradient-biased draws without correction lose 2.116 dB on average and lose all six comparisons. The full target pool and scan are charged, so this is a solver-sketch experiment, not reduced texture acquisition. Numerical polishing bounds the

Table 17. Conditional fitting at 3,000 kernels. “Gap” bounds the exported box solution’s excess unclipped training loss as a percentage of that loss; “Box val. ratio” is its clipped validation MSE divided by the initializer’s. PBR uses the validation-selected candidate, which can differ from the box solution. Refined denotes the corrected replication.

Object	Selected	Gap (%)	Box val. ratio	Fit dB	Refined dB
Orange	clipped	0.348	0.521	39.81	42.33
PoisonDartFrog	clipped	0.001	0.973	29.76	34.42
PinkOmbreHair	initial	0.026	1.093	45.98	47.24
SmallLimo	clipped	0.034	0.464	45.19	47.11
TShirt	clipped	0.094	0.345	48.71	48.90
Hummingbird	clipped	0.332	0.397	47.52	47.70
CE0824OLLA	clipped	0.107	0.914	40.86	45.68
Apple	box	0.001	0.816	40.72	46.36
Piggybank	clipped	0.017	0.163	46.95	47.96

Table 18. Stronger colour-only controls on corrected 3,000-kernel deletion supports. Both fitting columns use the same validation-selected regularization grid; PSNR is measured with the main eager PBR renderer. Refined is the corrected replication. Different selected regularizers and numerical convergence mean this is not an isolated sampling-rate ablation.

Object	Fit 16k (dB)	Fit 65k (dB)	Refined (dB)
Orange	39.63	40.24	42.33
PoisonDartFrog	29.88	30.30	34.42
PinkOmbreHair	46.42	46.89	47.24
SmallLimo	45.60	47.48	47.11
TShirt	48.71	48.92	48.90
Hummingbird	47.41	47.68	47.70
CE0824OLLA	42.27	42.91	45.68
Apple	40.82	41.66	46.36
Piggybank	46.95	47.49	47.96

exported solutions’ sampled-objective suboptimality by at most 0.411%; two of 24 continuous solves miss the requested 0.1% gap. No such gap certifies validation or rendering. The result motivates separating solver, sampling and update errors; it does not introduce a new importance-sampling theorem.

C.5. Does an initializer survive continuation?

We complete the initialization study with 41 new 1,000-step runs on Orange, PoisonDartFrog and SmallLimo, reusing four previous joint endpoints. All retain 3,000 kernels, the seed-0 parent, reset optimizers and native bitmap targets. One additional original run failed the training-pool matching check and was repeated; its output is preserved and excluded from the paired summary. This is adaptive diagnosis on existing objects, not a new confirmation set.

Which parameters need to move? Table 20 crosses update blocks with two learning-rate multipliers. Colour includes the background; non-colour includes positions, orientations, anisotropies, thresholds and sharpness. Frozen FP32 tensors are verified unchanged. Colour-only exports preserve the source non-colour arrays; the stock export’s ex-

tra position rounding and its scores are retained separately. All cells are reported without selection by test PBR.

Sampling must be tested after training. We continue both uniform and correctly weighted coverage initializers from Appendix C.4, using joint updates, native targets and a common $0.1\times$ rate. The two sketch draws retain the same parent and continuation seed. Table 21 reports every paired difference. The full-pool scan remains an initialization cost. These runs retain the original priority trainer; the correction in the initializer does not make that subsequent trainer unbiased.

Preserving the target during continuation. Nine further runs use the strong 65k initializer and joint $0.1\times$ updates: uniform batches, a half-uniform priority mixture with $1/(Np_i)$ weighting, and the same mixture rule without weighting (Table 22). Priority feedback uses unweighted errors. The sampling paths can diverge after their losses diverge. Weights are bounded by two. Conditional on previous updates, the weighted stochastic gradient targets the uniform finite-pool loss; neither Adam’s final parameters nor PBR quality inherit an unbiasedness guarantee.

Table 19. Validation-selected 1,000-step continuations (PBR dB). Orig/Native are training target conventions. Each initializer independently selects its learning-rate multiplier from $\{1, 0.1\}$ by existing native validation MSE. Init is the 65k colour fit before joint continuation. Del/Fit start from deletion/the colour fit. This is an adaptive diagnostic on 3,000-kernel fields.

Object	Target	Init	Del LR	Del dB	Fit LR	Fit dB
SmallLimo	Orig	47.48	1	46.24	0.1	48.24
SmallLimo	Native	47.48	1	45.93	0.1	48.42
TShirt	Orig	48.92	0.1	48.83	0.1	48.89
TShirt	Native	48.92	0.1	48.85	0.1	48.91

Table 20. Strong colour initialization followed by 1,000 updates (PBR dB). Rates multiply all unfrozen optimizer groups. The same initializer precedes all six endpoints of each object.

Object	LR	Initial	Colour	Non-colour	Joint
Orange	1	40.24	40.48	41.88	42.08
Orange	0.1	40.24	40.52	41.23	41.87
PoisonDartFrog	1	30.30	30.47	32.82	33.30
PoisonDartFrog	0.1	30.30	30.47	33.13	33.79
SmallLimo	1	47.48	47.06	47.20	46.61
SmallLimo	0.1	47.48	47.37	48.03	48.42

A priority-buffer confound. The inherited configuration initializes every priority to 10^5 with EMA factor $\beta = 0.99$. Since clipped RGB channel-sum squared loss lies in $[0, 3]$, after m visits its real-arithmetic value obeys

$$10^5 \beta^m \leq e_m \leq 10^5 \beta^m + 3(1 - \beta^m) + 10^{-4},$$

where the last term bounds the implementation’s 10^{-6} additive update. At 20 visits the sentinel still contributes about 81,791 while losses contribute at most 0.547. Stored buffers of all ten historical parents average about 81,500 after training. In this configuration, visitation history can overwhelm actual difficulty; this is not a claim about every upstream configuration. We therefore add six runs that initialize priority from an actual one-million-point error scan, retaining the uniform warmup and charging the scan separately. Table 23 gives all outcomes. This intervention is not rerun for the main parent/capped cohorts and is not a new headline method.

A local interpretation, not a new optimizer. Write a weighted local residual as $r + B\Delta C + J\Delta\theta$, where B includes the colour/background design and J differentiates non-colour parameters. Both incorporate fixed sample weights and, when clipping is linearized, the same fixed differentiable clipping branch. With unrestricted tangent colour updates and $P = BB^+$,

$$\min_{\Delta C} \|r + B\Delta C + J\Delta\theta\|^2 = \|(I - P)(r + J\Delta\theta)\|^2.$$

The identity follows by orthogonally projecting onto the range of B . At an unconstrained or interior stationary

colour solution, $B^\top r = 0$: only the component of the non-colour response outside the existing colour span creates a new local fitting direction. This is the classical variable-projection viewpoint [16], not an implemented reduced optimizer. Active bounds need feasible tangent directions; clipping transitions and top- k switches limit the smooth local model. A learning rate should therefore be assessed through the induced field change, not merely the magnitude of its parameters.

For a differentiable renderer R , the local output-change metric is $\|J_R \Delta z\|^2 = \Delta z^\top J_R^\top J_R \Delta z$. We retain diagnostics on 16,384 fresh area-uniform points and eight new non-equatorial camera directions, together with response-support changes on 1,024 points. These cameras are used for neither training nor selection. Their primary-ray albedo error is only a visibility proxy: it omits shading, indirect illumination and denoising, and must not be identified with the renderer Jacobian. The artifact reports all 60 model diagnostics, including the no-update fields. SmallLimo illustrates the proxy’s limit: its default joint continuation improves fresh area MSE and the independent visible-albedo MSE relative to the initializer, yet lowers the original five-view PBR PSNR. Different camera measures, shading and image processing remain confounded; this observation does not isolate a physical cause or validate a render-aware objective. No automatic scale, support-update or stopping rule is claimed from these three objects.

Table 21. Coverage minus uniform PBR (dB), before and after 1,000 updates. Both use 16,384 fitting draws from the same 65k pool.

Object	Sketch seed	Before	After
Orange	26001	+0.336	+0.124
Orange	26002	+0.365	+0.158
PoisonDartFrog	26001	+0.217	+0.101
PoisonDartFrog	26002	+0.071	+0.018
SmallLimo	26001	+1.270	+0.250
SmallLimo	26002	+0.414	+0.187

Table 22. Online sampling controls (PBR dB). Original is the existing priority sampler; mixture columns differ in loss weighting.

Object	Original	Uniform	Weighted	Unweighted
Orange	41.866	41.852	41.847	41.846
PoisonDartFrog	33.791	33.774	33.788	33.786
SmallLimo	48.422	48.551	48.530	48.519

C.6. Colour fitting, configuration updates and accounted cost

This completed follow-up uses four historical diagnostic objects and one training seed. It is separate from the expanded comparison. All six arms receive 1,000 continuation updates at 3,000 kernels. Five start from the same newly reconstructed pruned-and-refitted field; the sixth continues the corresponding direct fit. Both historical parent and direct fits have 5,000 updates. Optimizers are restarted. During continuation, checks verify identical native-texture training targets and minibatch index streams across all arms.

The main factorial crosses an additional conditional colour fit with joint versus colour-only continuation. Joint updates include positions, orientations, anisotropy and response parameters; this is not a position-only intervention. Colour-only controls freeze every non-colour parameter, verified in the exported arrays. The additional fit constructs the current response at 16,384 fixed training-pool samples and runs 100 box-coordinate sweeps. Rounded FP32 coefficients are accepted only if the clipped sample loss does not increase; accepted fits reset colour optimizer moments. This safeguard is in-sample. The candidate half-MSE gaps for the unclipped quadratic range from 1.28×10^{-5} to 1.00×10^{-4} ; these capped fits are not treated as exact conditional optima. A periodic arm repeats the fit before updates 1, 251, 501 and 751. Test renders do not select the fit or its acceptance.

At equal updates, an additional colour fit improves joint continuation by 0.247 dB on average, with three of four differences positive. Its effect is heterogeneous: the per-object differences are 0.072, 0.312, 0.656 and -0.051 dB. After this fit, permitting configuration updates instead of only colour updates adds 1.618, 2.790, 0.006 and 0.064 dB, respectively. Small Limo therefore has almost no additional configuration benefit in this protocol, whereas Poison Dart

Frog still needs substantial adaptation. On Small Limo, configuration updates without the additional fit lose 0.450 dB relative to colour-only continuation. A good colour fit can alter the usefulness of subsequent updates; it does not establish a universal benefit from either freezing or moving kernels. Periodic fitting beats one initial additional fit on only one of four objects, so we do not adopt it as a generally better recipe. Indeed, all twelve intermediate refits across these four objects lower the clipped sample MSE yet immediately lower rendered PSNR, by 0.552 dB on average (range 0.235–1.433). These correlated events are not twelve independent objects. They demonstrate a mismatch between the sampled acceptance objective and rendered quality, without isolating the contributions of finite sampling, visibility, filtering or an inexact solve. A conditional fit is a diagnostic and initialization tool, not an unconditional rule to refit after every update. The additional frozen-field controls in Appendix C.7 subsequently isolate sampled transfer and step length under the same native target convention.

Accounting for preparation and elapsed time. Figure 9 charges the historical measured spectral precomputation and parent/direct fitting times, the new calibration and pruning/refitting process, and current continuation wall time up to each saved field. Continuation includes response construction, transfer, the first solver invocation and snapshot writes. Offline checkpoint evaluation is separately recorded; it is measurement work, not part of generating the displayed field. These are reconstructed component costs across runs and dates, not one contemporaneous cold-start measurement. Each arm has one timed training run; we do not estimate cross-run timing uncertainty from these four objects.

For a descriptive common-budget comparison, use the earliest final recorded pipeline time among the four joint routes on each object, and take each route’s most recent

Table 23. Measured-error priority initialization, with the same half-uniform mixture and 1,000 joint updates at $0.1 \times$ (PBR dB).

Object	Uniform	Calibrated weighted	Calibrated unweighted
Orange	41.852	41.852	41.824
PoisonDartFrog	33.774	33.790	32.987
SmallLimo	48.551	48.513	43.736

Continuation	Orange	Frog	Limo	Hair
Joint	42.018	32.942	46.215	47.025
Additional colour fit + joint	42.090	33.254	46.870	46.974
Periodic colour fitting + joint	42.042	33.428	46.694	46.952
Additional fit + colour-only	40.472	30.464	46.865	46.910
Colour-only	40.422	30.419	46.665	46.917
Direct fit + joint	41.951	33.011	45.903	46.974

Table 24. Rendered PSNR after 1,000 continuation updates. Frog, Limo and Hair refer to Poison Dart Frog, Small Limo and Pink Ombre Hair. All arms are reported; these are historical diagnostics, not new independent confirmation objects.

saved field within that budget. This neither interpolates quality nor selects the best test checkpoint. The budgets are approximately 768, 458, 422 and 429 seconds. At those budgets, additional colour fitting beats ordinary joint continuation on two of four objects, rather than the three of four equal-update outcomes. Against extended direct fitting, it also wins two of four. The quality–cost curves thus retain the central conclusion: extra fitting can help, but its benefit depends on the object and the cost being matched. Sparse checkpoint spacing limits temporal resolution; these curves do not establish a globally optimal time allocation.

Launch granularity does not give another GPU speedup.

On the eight existing refined fields from the deployment pilot, we additionally batch one, four or eight query–kernel pairs per thread block in the streaming contraction. Candidate selection, the representation and top- k readout are retained. Seven of eight fields pass the 2×10^{-5} query-error check for all variants. Within these seven paired render comparisons, no tiled variant beats the original streaming implementation; median speed ratios are 0.988, 0.994 and 0.988, respectively, where a ratio above one would indicate a gain. This negative result is distinct from the positive CPU implementation check in Appendix E.6.

The Pink Bow field fails the query gate and is excluded from speed summaries. Garden Swing passes sampled queries but reaches a maximum linear-render difference of about 0.249 for both the original stream and the four-pair variant, with maximum denoised eight-bit difference three. Query-qualified speed measurements therefore do not certify render equivalence. These outcomes support checking final rendering and retaining failures; they provide no new GPU acceleration claim or validated LocalHK composition.

C.7. Separating sample coverage, solve accuracy and step length

Fixed design and target fidelity. We revisit twelve frozen-field events: updates 250, 500 and 750 on Orange, Poison Dart Frog, Small Limo and Pink Ombre Hair from Appendix C.6. These are correlated observations on four reused development objects. Kernel configuration remains fixed throughout each event. We reconstruct each original million-point training pool and match the archived faces, barycentric coordinates and targets exactly. The conditional fit used only its first 16,384 points; a further 65,536 pool points test transfer beyond that fit, although they were available to SGD. Separately generated native-target queries provide 65,536 fitting points, 8,192 validation points and 8,192 test points. The smaller fresh fit is a nested 16,384-point prefix. Validation alone chooses step length; neither test points nor renders choose an endpoint.

Native texture coordinates retain the original interpolation precision; recomputed targets equal the stored training targets exactly. The largest matrix-versus-live prediction discrepancy is 1.94×10^{-7} . Non-colour exports remain identical. These checks support interpreting the following differences as fitting effects under the stated contract.

Sample fit is not population fit. The original full colour update improves its own fitted-point PSNR by 1.550 dB on average, but reduces PSNR on both the excluded training-pool points and fresh test points in all twelve events, by 0.760 and 0.713 dB, respectively. Thus rendering alone is not responsible for its failure to transfer. Merely replacing the 16,384 fitting points also fails in all twelve events (mean fresh-test change -0.801 dB). Increasing this nested fit to 65,536 points, with the same 100 coordinate sweeps, improves test PSNR relative to the smaller fresh fit in every

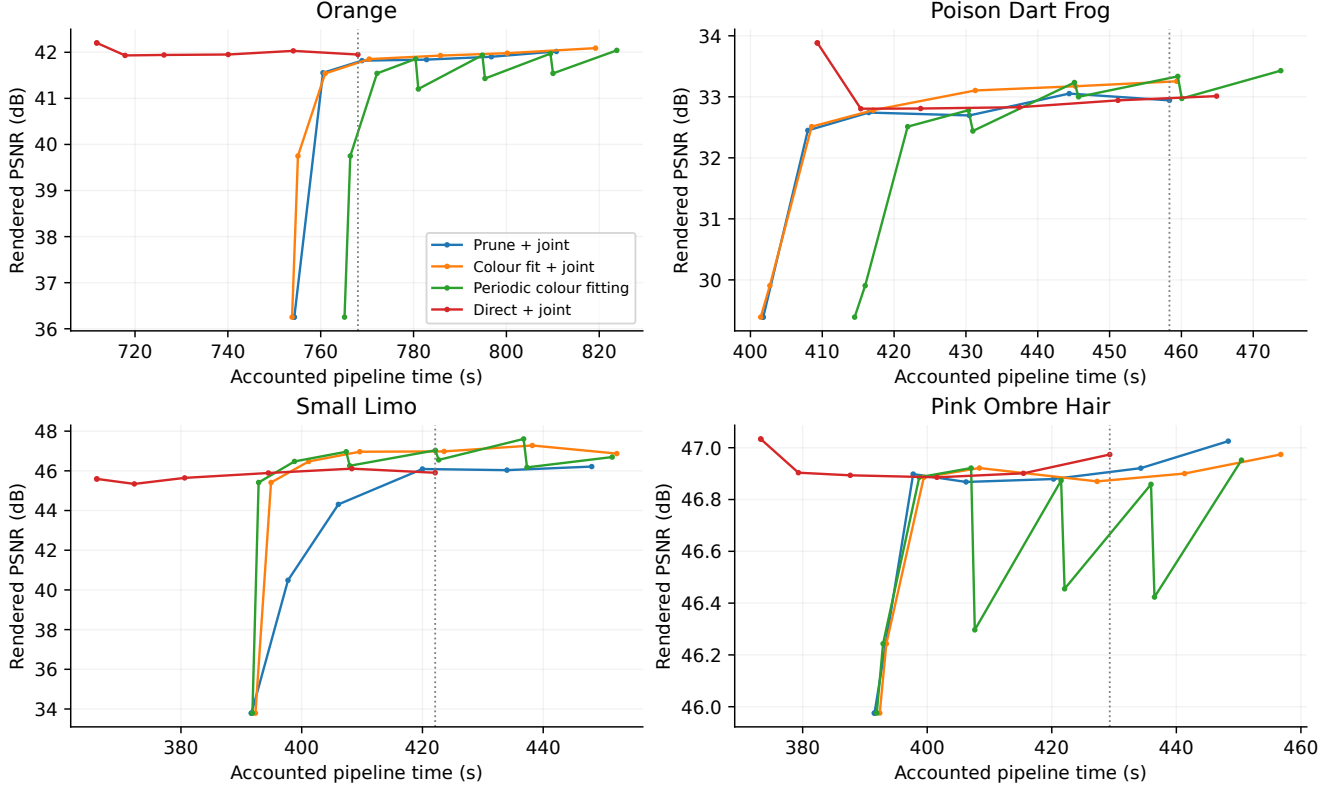


Figure 9. Quality against accounted pipeline cost on four historical objects. Dots are actual exported checkpoints; connecting lines only guide the eye. The vertical line is the descriptive common budget defined in the text. Historical preparation/fitting and new process times are both charged, with the cross-run accounting limitation stated explicitly.

event, by 0.862 dB on average. Its full step nevertheless still loses to the starting field in three events.

Solving longer does not remove this failure. At the four earliest events, increasing the original fit from 100 to 1,000 sweeps changes mean test PSNR by -0.032 dB relative to the shorter solve; all four remain below their starting fields. Clipping also matters: matched L-BFGS-B controls with the same 100-iteration cap change only the quadratic versus clipped fitting loss. At 65,536 points the clipped loss improves mean full-step test PSNR by 0.131 dB relative to the quadratic, with ten positive differences and two ties. These controls distinguish finite-sample transfer, solver effort and the implemented readout.

A useful direction can have a harmful endpoint. Include the background as a constant column of A , write the pre-clipped prediction as $z = AC_0$, and let $h = A(C_1 - C_0)$. For clipped sample MSE, the one-sided derivative along the colour update is

$$L'(0^+) = 2 \text{ mean}[(\text{clip}(z) - y) h m(z, h)], \quad (17)$$

where $m = 1$ inside $(0, 1)$ and at an endpoint when h points inward, and $m = 0$ otherwise, applied channelwise. This el-

ementary directional calculation is not a new optimization theorem. The original direction has negative derivative on fresh test points in all twelve events, despite its harmful full step. The evidence therefore supports excessive displacement along these sampled directions, rather than absence of any usable colour direction. It supplies neither a universal safe step nor a population guarantee.

We test the ordinary interpolation $C_\alpha = C_0 + \alpha(C_1 - C_0)$ and choose $\alpha \in \{0, 1/8, 1/4, 1/2, 1\}$ by fresh validation MSE. For the original direction this yields mean test improvement 0.067 dB: eleven gains and one unchanged field. The larger quadratic fit yields 0.184 dB, but one test case still regresses by 0.253 dB. The larger clipped fit yields 0.184 dB with twelve positive differences. Validation shrinkage is a conservative repair tested here, not a novel solver or a guarantee against overfitting validation data.

A conditional majorization update. Proposition 1 gives a global clipped-loss upper model, yielding a pseudo-target boxed solve and an analytic validation step. This applies classical MM/CCCP to the fixed readout; it is distinct from the unclipped optimality gap. After inspecting the preceding diagnostics, we test ten outer iterations of ten coordi-

nate sweeps on the same 65,536-point fits. Full steps improve mean fresh-test PSNR by 0.186 dB (11/12 positive); the analytic validation step gives 0.261 dB (12/12), versus 0.242 dB (11/12) with the same validation grid along the same directions. Twenty outer iterations of five sweeps give a consistent 0.251 dB with analytic steps; neither schedule is selected per object. All 360 outer updates satisfy the upper-bound checks within floating-point roundoff, despite changes in clipping state. Total recorded CPU solve times are 30.8/32.9 s for the two schedules, including their extra response products and checks. A subsequent one-outer-step control with 1/10/100 sweeps gives 0.160/0.231/0.247 dB, each with 11/12 gains; it is another development diagnostic, not a selected replacement schedule. The following render audit predates MM and does not evaluate its fields; none of these tests establishes fresh-object generalization.

A fixed rendering audit. At the earliest checkpoint we render the starting field, the original full update and the three validation-selected updates on all four objects. Candidate families and this checkpoint are fixed before rendering; surface validation determines the step lengths. Table 25 reports every object under the same five-view PBR protocol. The larger quadratic and clipped fits improve both mean-view sRGB PSNR and raw linear image PSNR on all four objects. Mean linear-PSNR changes are -0.197 , $+0.018$, $+0.082$ and $+0.060$ dB for the four table columns, respectively. Raw images use unit peak and no denoising, gamma transform or additional clipping; their metric scale differs from the display-image metric. The original shrunken fit slightly reduces linear PSNR on Orange and leaves Small Limo unchanged. This audit supports transfer of the larger conservative updates to these views, with no render-based method selection.

Do the repairs survive joint continuation? After these diagnostics we freeze six arms on the same four objects: joint continuation, larger quadratic fits with full, half or safeguarded steps, and direct clipped-loss L-BFGS-B fits with full or safeguarded steps. All 24 runs complete 1,000 actual updates from identical prepared fields, with matched native targets and minibatch indices. Fits at updates 0, 250, 500 and 750 use 65,536 newly permuted training-pool points. Quadratic and clipped solves retain their 100-sweep and 100-iteration caps. Guarded arms choose the same finite step grid on the next 8,192 points, then reject any sampled-training-loss increase. These acceptance points are disjoint from the current fit, but have been available to SGD; they are not independent generalization validation. Nonzero accepted updates reset colour Adam moments. Test queries and renders do not choose updates. The intervention combines the colour change with that reset; the separate reset-only controls below distinguish their effects.

The safeguarded quadratic’s per-object PBR gains are 0.029, 0.645, 1.697 and 0.050 dB in table 25’s order; Small Limo’s native surface difference versus joint continuation remains negative (-0.022 dB). Against reset-only continuation it is -0.045 dB, as in the main paper. The clipped controls’ weaker continuation outcomes counter the inference that a better frozen-field objective necessarily yields better joint optimization.

Does resetting the optimizer explain the gain? Four follow-up joint-training controls reset only colour and background Adam state at the same four boundaries. All 16 safeguarded corrections in the reference runs were accepted, so reset schedules agree exactly. The controls preserve non-colour optimizer state, learning rates and schedulers; sample and targets match the earlier runs. All available snapshots through the first 250 updates also match the untreated reference exactly, before the first nontrivial moment reset. The reference endpoints are rerendered in each control’s evaluation invocation. Resetting alone gives PBR differences of -0.067 , $+0.143$, $+0.071$ and -0.068 dB (mean $+0.020$). Adding the safeguarded colour fit gives $+0.098$, $+0.502$, $+1.626$ and $+0.119$ dB over these reset controls (mean $+0.586$), with positive PBR differences on all four fields. Native-surface PSNR improves by 0.035 dB on average, but still falls on Small Limo (-0.045 dB); SSIM and LPIPS changes are also mixed across objects. The PBR gain is therefore not explained by resets alone in this development cohort. This does not establish new-object generalization, cross-metric dominance, or mediation through subsequent configuration changes.

Eight further controls hold all non-colour parameters fixed for the entire 1,000-update continuation, comparing colour optimization with scheduled resets against the same process with conditional corrections. Every frozen run verifies non-colour invariance after all 1,000 updates. Together with the four reset-only joint controls, all twelve follow-up runs pass their initialization, target, batch, optimizer-state and endpoint checks. Table 27 compares the correction effects under the two update policies. Allowing centre-and-shape updates increases the correction’s PBR advantage on every object, with mean interaction 0.440 dB. The largest interaction again occurs on Small Limo. This factorial contrast supports coupling between colour correction and subsequent kernel configuration updates; it is not a mediated-effect estimate, a percentage of the gain explained by geometry, or an isolated time comparison.

Twenty-three runs pass GPU-occupancy timing checks; external occupancy invalidates the clipped safeguarded Hair timing, so common-budget comparisons use only the other three objects. The budget is each object’s minimum final snapshot time across six arms; each arm contributes its last actual snapshot within that budget, without interpolation

Object	Original full	Original partial	Larger quadratic	Larger clipped
Orange	−0.653	+0.039	+0.179	+0.162
Poison Dart Frog	−0.340	+0.213	+0.504	+0.290
Small Limo	−0.708	0.000	+0.922	+0.087
Pink Ombre Hair	−0.624	+0.007	+0.084	+0.038
Mean	−0.581	+0.065	+0.422	+0.144

Table 25. Change in mean-view PBR PSNR (dB) from the same starting field. Original fits use 16,384 points; larger fits use 65,536 fresh points. All columns except the full update use validation-selected step lengths. These four checkpoints are a subset of the twelve surface diagnostics, not independent confirmation.

Conditional correction	PBR Δ	Native surface Δ	PBR gains
Quadratic, full	+0.370	−0.038	4/4
Quadratic, half	+0.314	+0.012	3/4
Quadratic, safeguarded	+0.605	+0.066	4/4
Clipped L-BFGS-B, full	−0.005	+0.000	1/4
Clipped L-BFGS-B, safeguarded	+0.023	−0.019	1/4

Table 26. Object-weighted PSNR differences (dB) from joint continuation after 1,000 updates. All four development objects contribute; these are not MM continuations or new confirmation samples.

or test-quality selection. The safeguarded quadratic gains 1.058 dB at these three budgets. Construction, solves, selection, transfer and reset costs are charged; shared historical preparation is a reconstructed offset, and offline evaluation is separate. One target-consistency failure was retained before an unchanged-code replay: its 86.6 s cost leaves the corresponding clipped arm without a valid Small Limo checkpoint under the failure-inclusive budget. These observations are neither a four-object timing claim nor a contemporaneous cold-pipeline speedup.

A separately prescribed MM continuation now completes on all four objects. It uses ten outer iterations of ten coordinate sweeps and the analytic acceptance step, with all 16 corrections accepted and all 160 majorizer checks passing. At the same 1,000 updates, its mean PBR and native-surface gains over joint continuation are only 0.052 and 0.051 dB, respectively; two of four PBR differences are positive. It trails the safeguarded quadratic by 0.553 dB in mean PBR PSNR. Thus a valid empirical-descent guarantee and better frozen-field results do not establish a better continuation algorithm. These post-diagnostic runs retain their separate protocol and do not redefine the original six-arm budget. Only two MM comparisons meet the unchanged primary timing contract; no four-object MM speed advantage is claimed.

Qualification of a local moving-kernel prototype. An independent partial LocalHK-inspired implementation [18] runs 32 updates on one 94-vertex mesh with 3,000 transferred kernels and shared minibatches. Surface PSNR changes from 17.924 dB to 19.631 for colour-only, 20.973

for centre-and-shape, and 22.102 for joint updates. Native constrained centre steps and momentum transport pass planar/folded crossing checks. Support rebuilding costs 145.4/141.2 s in the moving arms, versus 0.48/0.48 s for optimization; caching alone gives no material improvement. This establishes partial training execution, not acceleration or complete LocalHK reproduction: density control, the original schedule, general orientation transport, independent initialization and common rendering were not established by that 32-update qualification. Appendix C separately reports the later multi-object, 1,000-update study with common rendering, while retaining its transferred-initializer and partial-implementation limits.

A subsequent implementation check replaces the prototype’s global minimum-distance-bucket schedule with a separate frontier for each kernel, while imposing the same canonical kernel–face order in both implementations. On the same mesh, both execute 32 real joint updates with support rebuilt after every geometry/radius change. All 32 same-input support checks pass, and all eight parameter tensors and current centre-face IDs remain bitwise identical after every update. Initial colour predictions and all raw/projected parameter gradients also pass the prespecified tolerance checks. This qualifies the implementation change on one moving trajectory, rather than showing that a fixed support cache remains valid after movement. The shared-GPU run is excluded from speed claims. LocalHK already describes parallel per-kernel expansion; this implementation check is not evidence that the authors’ unavailable implementation has the prototype’s cross-kernel scheduling overhead.

Object	Reset only	Fit, joint	Fit, fixed	Interaction
Orange	-0.067	+0.098	+0.071	+0.027
Poison Dart Frog	+0.143	+0.502	+0.071	+0.431
Small Limo	+0.071	+1.626	+0.429	+1.197
Pink Ombre Hair	-0.068	+0.119	+0.013	+0.106
Mean	+0.020	+0.586	+0.146	+0.440

Table 27. PBR PSNR contrasts in dB after 1,000 updates on four reused development objects. Reset only compares joint training with and without scheduled colour-state resets. Fit, joint and Fit, fixed compare adding conditional fitting under matched resets, with non-colour parameters respectively updated or fixed. Interaction is the difference of these two correction effects. Reference fields are rerendered in the control evaluations. These are post-diagnostic, single-seed controls.

The reason this rescheduling is possible is elementary: with geometry and kernel parameters fixed during a support build, each kernel has its own visited set, frontier and route priorities. Advancing one kernel does not change another kernel’s state. These updates therefore commute in the abstract algorithm, provided the within-kernel priority and tie rules are preserved. The implementation retains duplicate resolution, radius and edge tests; it changes which independent frontiers may advance in the same GPU iteration. Floating-point transforms and downstream ordering require separate checks, as above. This is a dependency argument, not a new heat-kernel approximation or a proof of equivalence to the authors’ complete method.

An isolated component replay on two reused meshes passes exact support-key, within-kernel visit-sequence and edge-rejection checks. Transform discrepancies are reported separately at FP64 roundoff. Median complete-build times change from 4.429 to 0.098 s on 184 faces and from 4.872 to 0.283 s on 2,109 faces, each with 3,000 kernels and three alternating repeats. The immutable mesh cache costs 0.015/0.013 s to construct; a cache-only control still takes 4.394/4.809 s. The component therefore isolates a scheduling bottleneck in this prototype. One-second process sampling records no foreign occupancy during this replay, but cannot exclude between-sample contention. These are support-construction times, not end-to-end fitting/render speedups or a comparison against LocalHK author code. A pre-benchmark import failure and its 3.0 s cost are retained separately from the successful replay.

A separate isolated replay includes all 32 optimization steps and every support rebuild on the 94-vertex mesh. These operations take 157.433 s in the reference prototype and 4.471 s with per-kernel frontiers, a $35.2\times$ reduction. All same-input support checks pass and the complete parameter trajectory remains bitwise identical. Per-arm initialization and initial readout cost a further 4.701/0.199 s; the paired validation process, including extra shadow checks, evaluations and artifact writes, takes 179.761 s. One-second occupancy records contain no foreign GPU process. This is a single-object, 32-step implementation benchmark, not a full

training curve, a new representation, or a speed comparison with LocalHK author code.

C.8. Do richer local models improve fitting?

This development study uses four previously inspected objects at the 500-update checkpoint. It is separate from the reserved-object confirmation. A first probe additionally uses updates 250 and 750 on the same objects; these twelve checkpoints are not twelve independent objects. Fitting uses 16,384 fresh native-target points, with disjoint 8,192-point validation and test sets. All choices are committed on validation before test or rendered evaluation.

Function anchoring. We compare no penalty, coefficient ridge, a diagonal anchor Gram and a full anchor Gram, using the same bounded coordinate solver. The latter two use identical additional unlabeled queries; their anchor targets are the initial field, not texture labels. Penalty weights and step length are selected from identical fixed grids. Mean held-out surface gains over the unchanged field are 0.0851, 0.0895, 0.0975, 0.0918 dB, respectively. The full Gram beats the diagonal on only four of twelve checkpoints and has no object-averaged advantage. Its finite-query drift bound therefore does not justify the extra structure empirically.

Colour-compensable shape directions. At the full conditional colour fit, we form four bounded negative-gradient directions for angle, anisotropy, threshold and sharpness; source positions remain fixed. Finite differences approximate their field changes. A regularized Schur elimination removes the changes explainable by free colour coefficients. Saturated channels are masked and active colour bounds are conservatively frozen. This is a restricted local subspace, not full variable-projection training. The fraction of direction energy removed ranges from 41.8% to 83.3%.

Both ordinary damped Gauss–Newton and the reduced step receive the same recovered colour compensation and validation step-length search, including zero. The reduced step changes mean rendered PSNR by -0.0009 dB relative

to this matched Gauss–Newton control. Large compensable tangent energy does not produce a useful gain here. Both start from the full colour fit, whose generalization may already be worse than a safeguarded partial fit.

Normalization is a representation choice. Removing the denominator while keeping kernel configuration and colour bounds changes the field class. At the matched 100-sweep solve budget, the additive substitution loses 9.770 dB in mean rendered PSNR relative to the safeguarded normalized fit, on all four objects. By contrast, the safeguarded fit gains 0.343 dB over the initial field. An adaptive follow-up runs 1,000 and 5,000 coordinate sweeps for both readouts from the same colours. Test quality remains lower for the additive arm; its residual quadratic gaps are recorded. This does not establish inferiority of a separately initialized, fully trained additive representation.

Separating the conditional optimum from generalization. We recompute convex first-order gaps from the feasible coefficients. For the finite, unclipped quadratic objective, a feasible value is an upper bound on the optimum and subtracting its gap gives a lower bound. In Orange, Poison Dart Frog and Pink Ombre Hair, the additive lower bound exceeds the normalized upper bound; an incomplete solve alone cannot reverse that ordering. Small Limo has the opposite certified ordering, yet its additive test quality remains worse. Thus conditional representational limits and sampling generalization produce different failures. These are numerical bounds for fixed configurations and colour boxes, not formal interval-arithmetic proofs, clipped-loss bounds or conclusions about fully trained representation families.

Coordinate and implementation errors differ. An initial wrapper check compared saved barycentric points with points reconstructed through Cartesian coordinates. On the initial 256-query Small Limo audit, the round trip changes barycentric coordinates by up to 1.74×10^{-4} and can change top- k selection. At identical reconstructed queries, all seven tested fields agree with their matrix responses within 2×10^{-7} on that audit. We retain the original failed sensitivity checks and complete rendering with a same-query check on 4,096 points. Resource-only changes to callback batching and constant-tensor residence are separately qualified within 2×10^{-5} ; completed original renders and completion renders remain distinguishable. Shared-GPU timings are not speed evidence.

Implication. These probes support using conditional structure to diagnose a fit, but do not establish a more elaborate optimizer. The classical eliminated system, full Gram penalty and additive readout each require evidence beyond

their mathematical plausibility. We retain their negative results to delimit the simpler fitting controls rather than count them as additional contributions.

C.9. Does a smaller update scale transfer?

The selected numerical failure in the selected numerical diagnostic does not establish that a smaller learning rate is generally preferable. We test this question in an exploratory follow-up on eight previously evaluated objects: four from the expanded Objaverse cohort and four from GSO. Before scoring this follow-up, we select the first eligible object in the frozen order of each cohort and physical GPU, requiring a valid reference with more than 4,000 kernels. Selection follows observation of the earlier experiments; these are not independently held-out confirmation objects.

Controlled continuation. All runs use 3,000 kernels, seed 0 and 5,000 additional updates. Three routes start from the same deleted and recoloured reference: the default update scale, a constant $0.1 \times$ scale, and the adaptive rule below. A fourth route continues the budget-limited direct fit at the default scale. Both reference and direct initial fits received 5,000 updates. All four routes reset their optimizers and use uniform sampling from a fixed pool, retaining the inherited source-UV sampler’s colour targets. This changes the original priority sampling and does not use the native-bitmap targets of the preceding initialization study. All 32 final exports are finite and complete the prescribed updates; evaluation uses the same five-view PBR protocol. No checkpoint or per-object scale is selected using test quality, and these results do not replace any endpoint in the main confirmation experiment.

A first-order scale rule. Let L_t be the current batch loss, Δ_t the next scheduled Adam and geodesic-momentum parameter proposal, and $d_t = \nabla L_t^\top \Delta_t$. We apply the common multiplier

$$s_t = \begin{cases} \min\{1, 0.1L_t/(-d_t)\}, & d_t < 0, \\ 0.1, & d_t \geq 0. \end{cases}$$

The proposal is computed before parameter constraints and geodesic tracing. Optimizer moments evolve normally; the multiplier is removed before the ordinary scheduler update. Nonfinite quantities fail explicitly. This is neither a line search nor a trust-region guarantee on the actual field: constraints, tracing and support changes can invalidate the predicted loss change.

Interpretation. The fixed small scale loses quality on all eight objects, while the adaptive rule stays close to the default (Table 29). Across its 40,000 updates, the adaptive rule

Object	Initial	Safeguarded	Full colour	GN	Reduced	Additive
Orange	46.762	47.034	45.317	45.535	45.538	38.998
Poison Dart Frog	33.686	33.886	33.070	33.078	33.079	29.440
Small Limo	49.417	50.010	48.314	48.158	48.151	36.065
Pink Ombre Hair	52.350	52.655	50.838	50.838	50.838	40.003

Table 28. Rendered PSNR on four reused development objects. Safeguarded and full colour use the original normalized readout. GN and reduced updates begin at the same full colour fit. Additive removes normalization and independently refits colours at fixed kernel configuration. These are local development probes, not independent full-training method rankings.

Table 29. Exploratory scale control at 3,000 kernels. Entries are mean paired five-view PBR PSNR differences in dB; parentheses count positive differences. Each cohort contains four previously evaluated objects.

Cohort	Adaptive – default	One tenth – default	Warmup – default
Objaverse	+0.0217 (3/4)	−0.4625 (0/4)	+0.0150 (2/4)
GSO	+0.0052 (3/4)	−0.3891 (0/4)	+0.0085 (3/4)

reduces the scale on only 810 (2.025%); every object’s median scale is 1, and all scales after the first 1,000 updates are 1. No non-descent fallback occurs. The end of scale reduction coincides with the first scheduled learning-rate decay, so these traces alone cannot attribute it to an initialization that has settled. For fixed loss, gradients and moment states, halving the scheduled rates halves d_t and doubles the unclamped descent-branch multiplier. The schedule can therefore make the controller inactive without certifying a better fitted field.

A fixed warmup control. After observing the pilot, we prescribe one additional rule for all eight objects: multiply every learning rate by $0.1 + 0.9(t - 1)/99$ at updates $1 \leq t \leq 100$, and by 1 thereafter. All other continuation settings and the original physical GPU per object are retained. Initial parameters, random states and first-batch losses match the reused controls exactly. All eight additional 5,000-update fits are finite and complete evaluation. The duration and endpoint are fixed before these new quality measurements; the study nevertheless remains a post-hoc follow-up on previously observed objects.

Warmup improves the two cohort means by 0.0150/0.0085 dB over default (Table 29). Adaptive minus warmup is +0.0066/−0.0033 dB, or +0.0017 dB averaged across all eight objects. These results do not establish statistical equivalence or a transferable controller advantage. They do close the simpler-control gap: within this small, single-seed experiment, a fixed warmup reaches a similar mean endpoint without using the predicted loss change. The successful small-scale repair of one failed run cannot justify reducing all continuation steps. No further scale tuning is included.

D. Fixed-time feature confirmation

D.1. What the algebra preserves

For orthogonal Q , $QQ^T = I$, which proves Eq. (7). By contrast, rotating Φ but retaining diagonal weights generally gives $\Phi Q \exp(-t\Lambda) Q^T \Phi^T$; equality requires the appropriate commutation or a corresponding weight transformation. This concern is already present in LocalHK/ThermalRF [18]. Our construction and evidence concern a pinned public implementation, not an error in HK-TeX’s underlying geometric theory.

The fixed-time feature construction recomputes raw eigensolutions before alignment. Rescaling an already mixed legacy cache cannot recover the original operator in general. Both compared constructions use checked FP64 or orthogonal factors and identical sequential reference choices; stored arrays are FP32. The isotropic search embedding is unchanged. Off-grid interpolation and variable-time behavior do not follow from the grid-node identity. This version keeps time fixed and does not reduce spectral-cache dimensions.

D.2. Confirmation and retained failures

The 22 declared objects exclude the two diagnostic meshes. They were previously geometry-screened and some had unrelated earlier fitting results; they are held out from this method’s development, not an entirely untouched recruitment sample. Object identities, sources, protocol and aggregation were frozen before these feature outcomes.

Both variants use the same raw learned initialization arrays, 5,000 target batches and native colour targets. Each 4,096-point batch is accumulated in two 2,048-point chunks before one update. Both use the same memory adapter, including checkpointed gathers and CPU placement of large constant bases. This arithmetic need not reproduce the un-

Table 30. All 22 declared feature-confirmation objects. Gains are feature minus control in rendered PSNR. Two-seed means enter the primary result; incomplete or unsupported objects do not. Names are shortened; released records retain full names and IDs.

Object	Source	Pairs	Δ dB
Shell	Obj.	2	+0.178
Banana	Obj.	2	+0.004
Game Boy Color	GSO	2	+0.318
Seated figures	Obj.	2	-0.109
Butterfly pencil case	GSO	2	-0.033
SelfieTennis ball	Obj.	2	+0.097
Football	Obj.	2	-0.177
Cascadia shoe	GSO	2	+0.083
wrench	Obj.	2	+0.073
Kraken headset	GSO	2	+0.133
Lp Banana Uvw	Obj.	2	+0.164
Naga mouse	GSO	2	+0.253
Mad Gab game	GSO	2	+0.096
MRAC Golden Freddy’s Hat	Obj.	2	+0.318
Seahorse	Obj.	2	+0.319
Whey protein tub	GSO	2	+0.068
Seashell Fossil	Obj.	2	+0.016
CC0 - Doormate	Obj.	2	+3.505
Tropical hat	GSO	2	-0.044
Product box (8 vertices)	Obj.	0	—
Borsalino	Obj.	1	single
Dore 1 Low	Obj.	0	—

modified full-batch trajectory. There are no extra colour corrections or warm-up steps.

All 44 seed pairs are terminal. Two meshes with eight and sixteen vertices do not support the fixed configuration. Borsalino completes seed one but seed zero times out; its completed seed gains 0.072 dB and is descriptive only. The 19 complete objects supply 38 primary pairs and 76 fields. Including Borsalino’s single pair gives 39 evaluated pairs and 78 fields in the ledger.

All 78 completed training cells required export-only recovery after an initial export failure. Recovery reused the final weights without retraining. Paired initialization, batches and targets were verified, and exported parameters matched the final training state. The timeout remains excluded from the primary paired comparison.

The primary mean gain is 0.276890 dB, median 0.096378, with 15/19 positive objects and 95% object-bootstrap interval [0.050909, 0.668333]. There are 29 positive primary seed contrasts out of 38, but objects remain the resampling unit. The pooled-image contrast is 0.274967 dB with 14/19 positive objects. Overall SSIM changes by +0.000330 and LPIPS by −0.000280; both intervals cross zero. Objaverse mean LPIPS worsens by 0.000142, while GSO improves by 0.000862.

D.3. Separate complete-cost replay

The two diagnostic meshes are Barricade and Wrench. Each uses two seeds, with variant order reversed for seed one and the same physical RTX 3090 within a pair. Both caches are freshly rebuilt. Native targets, batches and initialization arrays match within each pair. One-second occupancy monitoring records no foreign GPU process. This does not establish exclusive host resources or a cold filesystem.

Table 4 in the main paper reports the four stage totals.

The totals charge per-variant precomputation, shared target preparation equally, training, export, preflight and nine-checkpoint evaluation. The curve in Fig. 3 instead includes fresh cache, target preparation and elapsed training through each checkpoint; offline export/evaluation and startup outside that clock are excluded. The two cost definitions are intentionally distinguished.

Using the smaller of the two endpoint checkpoint budgets and selecting the last recorded checkpoint within it chooses the 4,750-update feature fit and 5,000-update control in all pairs. Differences remain positive at 0.344, 0.698, 0.974 and 0.811 dB. Coarse checkpoint spacing and clock variability preclude an exact time-to-quality or global Pareto claim. These two objects do not establish cost parity for the separate 19-object confirmation.

E. Deployment and alternative allocations

E.1. Coding, solving and execution

The deployment pilot reuses eight objects with two compact routes each. The GPU conditional solver compares the same decoded response matrix, GPU, initialization and target gap. It reaches the target on all 16 fields; the conservative global-step accelerated control reaches it on six within the iteration cap. The 6.26× ratio is restricted to mutually converged fields and excludes response construction and device preparation.

Six-bit fitted colours gain 0.135 dB over nearest six-bit rounding. Relative to eight-bit rounding they save 5.47% of complete field bytes with mean loss 0.082 dB; the worst field loses 0.450 dB. Decoded coefficients remain FP32. This is a file-size trade-off, not a matching reduction in resident coefficient memory. The expanded codec controls contain 840 evaluated endpoints across three prescribed precisions; these are repeated configurations of the existing objects, not 840 independent objects.

The compiler-matched four-bit benchmark has 15 mutually qualified fields, median render speedup 1.26× and median temporary allocation ratio 4.50%. A measured positive-cutoff tie creates an RGB jump of 0.00889 despite agreement in outer candidates and spectral gathering. In a separate six-bit composition with compilation disabled, raw streaming passes only 14/16 held-out query checks. A calibration-only selector still fails one held-out field. These

failures are not repaired by changing the declared thresholds.

Using all 50 already-computed responses speeds all 16 fields, but a low-poly object loses up to 0.482 dB after refitting. A later source-assisted readout selection keeps inner selection on that object and chooses all-response readout elsewhere. It is a separate calibration experiment, not a universal numerical guarantee. Precision selection and query-wise fallback retain additional quality or execution failures in the public records.

E.2. UV and nonlinear readout controls

Historical hybrid experiments fit a kernel residual on a UV base while keeping kernel support fixed. Residual gain on samples is $2\langle Y - B, R \rangle - \|R\|^2$ before clipping, but a positive gain does not make the hybrid a better allocation of bytes. All 20 tested fixed-support hybrids are dominated by their byte-matched finer-UV comparators. Atlas waste alone is not a valid classifier for native UV failure.

The later learned gate is compared with scalar mixing and nonlinear single-branch readouts. Nine of ten object comparisons are finite. A pure-HK readout beats the gate on six, with a smaller payload; Piggybank retains a roughly 0.912 dB gate advantage over the tested HK-network control. These comparisons do not exclude other jointly trained representations. A gate whose inputs include both expert outputs also pays both evaluation costs.

E.3. Editing and initialization contract

The expanded compress-refine route uses the cached deletion-and-refitting initializer described here. The rebuilt-support refit is a separate post-hoc control; it does not initialize those continuations.

Static deletion scores. On 4,096 stored calibration queries, we cache all 50 outer candidates and their filtered responses from the reference field. The original 30-response readout supplies the distillation target, including its denominator floor, background and output clipping. For each kernel, we simulate its deletion alone, allow the next cached response to enter, and sum the squared RGB change on affected queries. We compute these scores once and stably sort them in ascending order. For target count K , we delete the first $N - K$ entries; scores are not recomputed after earlier deletions. The resulting readout reselects up to 30 surviving responses within the cached universe, assigning zero mass to any deleted padding entries. It does not rebuild the outer search.

Cached colour refit. Let W contain the surviving, unnormalized filtered responses, $D_{jj} = \max(1, \sum_i W_{ji})$, C_0 the retained colours, b the unchanged background, and Y

the clipped reference targets. For each channel, a damped sparse least-squares solve approximately minimizes

$$\|W\Delta C - [D(Y - b) - WC_0]\|_F^2 + 0.2^2 \|\Delta C\|_F^2.$$

Both LSMR stopping tolerances are 10^{-4} , with at most 24 iterations. A finite line search uses six scales:

$$C_\alpha = \text{clip}_{[-1,1]}(C_0 + \alpha \Delta C), \quad \alpha \in \{0, 1/8, 1/4, 1/2, 3/4, 1\}.$$

We select the candidate with the lowest clipped cached-field MSE. Thus the solve uses a numerator-space objective, while the finite line search checks the deployed-form readout. Including $\alpha = 0$ prevents a cached loss increase before export rounding; it is not a guarantee after rounding or rebuilding the evaluator. Retained non-colour arrays and the background are copied, and colours are cast to the original export dtype. This exported field initializes joint refinement.

The separate rebuilt control recomputes the actual edited search and normalized response matrix on the same queries, targeting the reference evaluator’s outputs. It tests damping $\{0, 0.01, 0.1, 0.2\}$, LSMR tolerances 10^{-7} and at most 400 iterations, using the same positive line-search scales and the unchanged initializer. Selection minimizes clipped calibration MSE after export rounding. These are finite-budget distillation procedures, not exact colour optima or source-supervised controls.

E.4. Deployment experiments and resource accounting

The hybrid and response-table studies below belong to the historical diagnostic cohort and retain its recorded fields. They are not reruns of the expanded comparison.

E.4.1. Fixed-support residuals against a byte-matched UV control

We use Orange, PoisonDartFrog, SmallLimo and Apple. Each object has UV long edges 64/128 and residual counts 128/512, plus one gated version of the 64-edge, 512-kernel configuration: five tested hybrids per object. The native bitmap and signed residual are added before final clipping. Training and validation use 16,384 and 4,096 newly sampled surface points, disjoint from the stored evaluation queries. The parent supplies fixed kernel geometry. We compare support ranked by response mass or residual correlation, then rebuild each selected support’s normalized response matrix. Colours use LSMR with damping $\{0, 0.01, 0.1, 0.3\}$, tolerances 10^{-7} and at most 400 iterations; bounded colours are scaled by $\{1/4, 1/2, 1\}$ and rounded to the export dtype. Validation selects support, damping and scale, including a zero residual candidate. No rendered test score selects these choices.

For the gated case, training residual benefit ranks faces, and validation selects the retained fraction from

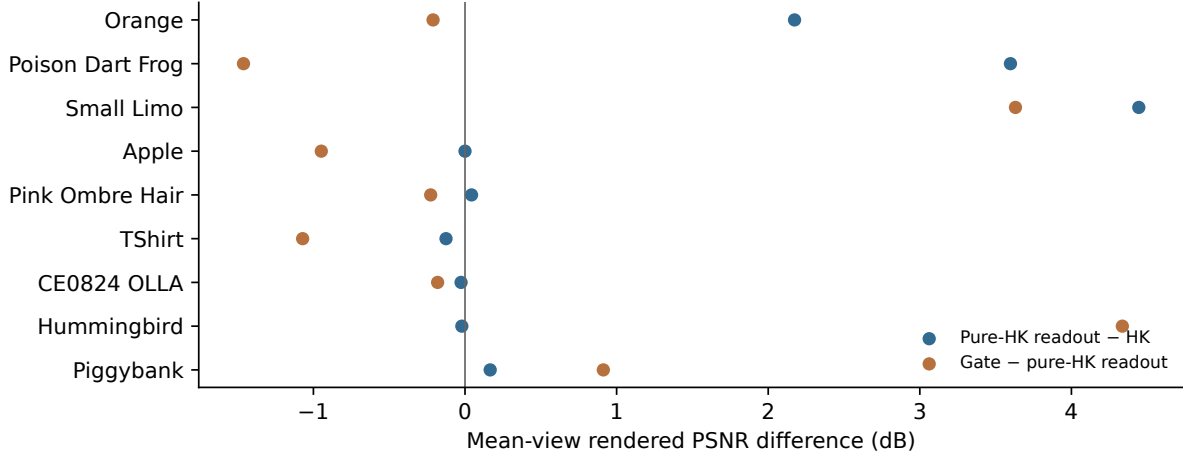


Figure 10. Readout controls distinguish an additional representation from additional nonlinear colour capacity. Historical objects and finite-pair coverage are retained; this is not a new-object benchmark.

$\{1/4, 1/2, 3/4, 1\}$. The corresponding vertex gate is interpolated at queries. Gated image scores use the actual pre-gated renderer: batch compaction can change outer candidates, so multiplying a full-batch response by the gate is not generally the same experiment.

Residual parameters and gate arrays are losslessly encoded. For each hybrid, we enumerate UV long edges 32, 36, ..., 256 and choose the largest whose actual PNG fits the hybrid’s PNG-plus-residual-plus-gate bytes. Selection uses bytes only. Both alternatives use identical, bit-preserved UV streams, so the comparison holds whether those streams are shared or fully charged.

E.4.2. Can a small controller make fusion worthwhile?

An additional development study on the same four objects separates the ability to mix two experts from the value of storing both. We combine native UV long edges 64/128 with either the recorded parent or a 512-kernel field whose fixed support is reused from the earlier study and whose absolute RGB colours are refitted. Each object has 32,768 new fitting points, 8,192 validation points and 32,768 test points. We compare a fixed half mixture, a fitted scalar, and validation-selected local UV-cell gates. All mixtures evaluate both experts; they make no gate-induced query-speed claim.

An exploratory follow-up conditions a 16-hidden-unit network on the two RGB predictions, optionally with periodic UV coordinates. Validation selects the input variant and checkpoint after FP16 weight storage. Across the eight compact-field/resolution configurations, this 129–193-parameter gate improves five-view PSNR over the scalar by 0.508 dB (8/8 positive), and over the local lookup gate by 0.339 dB (7/8 positive). These are repeated configurations on four development objects, not eight independent objects. The actual weight containers occupy 1,070–1,213

bytes. Spending the complete base, model and gate payload on pure UV still gives higher PSNR in all eight cases, by 2.724 dB on average. This control searches long edges 32, 36, ..., 1024 by encoded bytes only; neither surface test errors nor rendered views select a deployed gate.

Thus, useful error-dependent mixing need not improve the deployment frontier. A globally zero gate can be eliminated at export and reduces to pure UV, without paying for an unused HK branch. This result does not rule out jointly trained experts or other spatial bit allocations. A budget-conditioned policy must also distinguish shared from transmitted UV coordinates and count the HK evaluator’s runtime dependencies; a small mixing network alone establishes neither lower residency nor faster queries.

E.4.3. Which decisions should adapt to the target?

Conditional networks already control compression operating points and residual weights in image restoration [4, 17]. We use this design pattern to distinguish three questions: choosing quality under a byte budget, choosing bytes for a quality target, and adapting the local mixture to content. This is a development study, not a new general controller contribution.

We split the historical ten objects into six controller-training objects, two validation objects, and two controller-held-out objects, Hummingbird and Piggybank. Each has 8,192 new fitting points, 4,096 calibration points and 16,384 test points. All objects were used in earlier studies. Candidate fields are already trained; the experiment cannot establish savings in their original fitting cost. The menu contains native UV, three recorded HK capacities, constant colour and a fitted scalar mixture. We additionally audit full-source UV and an RGB8 constant, preserving the original capped-menu results.

Object	UV64 residual gain	UV128 residual gain	Matched UV gain
Orange	+0.70	+0.12	+0.14—+2.94
PoisonDartFrog	+1.19	+0.25	+0.41—+1.11
SmallLimo	−0.05	+0.08	+1.59—+9.01
Apple	+0.37	+0.20	+0.27—+1.66

Table 31. Historical hybrid controls, PBR PSNR differences in dB. The first two columns compare the 512-kernel hybrid with its same-resolution UV base. The last column ranges over five hybrids per object and reports byte-matched pure UV minus hybrid. All 20 matched UV controls have higher PSNR and no larger files. This tests the stated fixed-support family.

Construction decisions. A small quality predictor uses image statistics, candidate metadata and 256-point error probes of one UV and one HK anchor. An exact byte constraint filters proposals, and calibration checks up to three candidates. A quality request that fails these checks triggers a scan of the remaining menu. The request conditions this selection procedure, not the predictor’s neural input. Validation prefers ridge regression to a 32-unit MLP, but both predictors miss good budget choices on Piggybank. Direct 256-point probing of every candidate, followed by calibration of three proposals, matches the exhaustive selector’s test MSE on all 28 budget requests across the two held-out objects and both UV accounting regimes. Its 20 quality requests also meet their test targets and match exhaustive calibration’s selected bytes. These are repeated requests on two objects, not 28 or 20 independent samples. Reduced query counts are simulated cache accesses; online wall-time savings remain unmeasured.

Local adaptation. We combine UV long edge 128 with a recorded 3,000-kernel field. A 161-parameter gate reads the two RGB predictions and two object-level error summaries. Shared training, per-object fitting, and 512-point adaptation are evaluated separately, with FP16 weights and calibration-selected checkpoints. The shared gate loses 1.577 dB to the fitted scalar on Hummingbird. Per-object fitting improves both held-out objects, but their deployment outcomes differ (Table 32). The per-object fits use target supervision; they are not zero-shot transfer. Both branches are evaluated at every query.

Piggybank’s gate occupies 63,572 bytes when UV coordinates are shared, or 77,846 bytes when they are charged, versus 74,620 bytes for the pure 4,000-kernel field. Thus only the shared-coordinate comparison improves both encoded bytes and PSNR; LPIPS is worse than the 4,000-kernel control. The gate’s serialized weights and eight bytes of object context are charged. UV matching enumerates long edges 32, 36, . . . , 1024, the source image, and 2,048/4,096 when no larger than the source, selecting by bytes only. The opposite outcomes on the two objects rule out a universal preference for either fusion or UV within this tested family.

To test a simpler explanation, we refit only the 3,000-kernel field’s colours using the same fitting samples, 400 bounded coordinate sweeps and exported step sizes $\{0, 1/4, 1/2, 1\}$. Calibration selects zero update on both objects. Thus this conditional-colour control does not recover Piggybank’s gate gain; it does not exclude stronger regularization or other pure-field corrections. Original-field and target images are bitwise identical across these paired render runs. Historical scores from other mesh wrappers are not subtracted from these native-geometry comparisons.

What must be estimated? For $e_U = U - Y$ and $e_H = H - Y$, ordinary squared-error minimization gives

$$g^* = \text{clip}_{[0,1]} \frac{\mathbb{E}\|e_U\|^2 - \mathbb{E}\langle e_U, e_H \rangle}{\mathbb{E}\|e_H - e_U\|^2},$$

when the denominator is nonzero; otherwise the experts coincide almost surely. Individual expert MSEs alone do not determine this optimum: their joint error also matters. Local conditioning estimates these moments within feature groups. This classical identity motivates target calibration without claiming that it proves the cause of the observed transfer failure. Resource selection remains discrete: an interior gate pays for both experts, while an exactly constant endpoint may eliminate a branch.

E.4.4. Frozen response-table evaluation

The paired study uses nine historical objects, each with its reference, refined 4,000-kernel, refined 3,000-kernel and coded refined 3,000-kernel fields. Both evaluators retain candidate search, windowing, filtering and top- k composition; the alternative precomputes vertex-kernel responses. It still requires spectral preparation. On an RTX3090 with TF32 matrix multiplication disabled, query timing evaluates the same 65,536 points at batch size 512, using seven synchronized repeats after warmup and alternating evaluator order. Rendering uses five views and two warmed repeats per view, randomized method order, and synchronization around rendering and denoising. The per-field render time is the mean of the five per-view medians; scene construction, model loading, image saving and scoring are excluded. The medians of the 36 paired speed ratios are $1.48\times$ for queries

Object	Gate—scalar	Gate—4,000 HK	Gate—matched UV
Hummingbird	+1.104	+4.235	−3.416
Piggybank	+1.084	+0.550	+5.738

Table 32. Per-object gate gains in five-view PBR PSNR (dB). The gate uses a 3,000-kernel expert. The UV control has no larger PNG-plus-field-plus-gate payload and shares the same UV dependency. The 4,000-kernel column compares recorded capacities, not equal total resources. These are two development objects, with no latency or population-generalization claim.

and $1.34\times$ for rendering, including numerically failing configurations.

The numerical gate is maximum absolute RGB difference at most 2×10^{-5} over those 65,536 queries. It passes 31/36 fields. Failures are the CE0824OLLA reference, SmallLimo reference and coded refined 3,000 field, and the PinkOmbreHair and Hummingbird refined 4,000 fields. The largest observed RGB difference is 0.01650; the largest absolute mean-view PSNR change is 0.0050 dB. The extra FP32 tables occupy 3.7–953.6 MiB. Both paths retain the original spectral state in this benchmark. Consequently these timings are neither uniformly pointwise-accurate speedups nor evidence of reduced peak memory.

E.5. Does fusion require a second representation?

A nonlinear gate may help because it supplies additional colour-correction capacity, rather than because two representations are necessary. We test this alternative on all ten historical objects, reusing the preceding 8,192/4,096/16,384 fitting/calibration/test points. This is a development follow-up, not ten new confirmation objects.

The controls are a quadratic RGB residual polynomial, pure-HK and pure-UV residual networks with 16 or 24 hidden units, and a joint RGB residual network with 16 hidden units. The networks have 115/171 parameters for a single branch and 163 for two branches; the gate has 161. All new networks use the same 1,000 full-batch updates, a zero-initialized output layer, and calibration of exported FP16 checkpoints and identity. Width selection uses calibration only. The polynomial fits 30 coefficients with a fixed ridge/step grid. These are matched-sample family comparisons, not exactly matched capacity or fitting time. Serialized metadata and all required branches are charged. HK-TeX already permits output networks; these controls do not introduce a new neural texture architecture. The recorded HK experts are historical compact fits, not newly retrained original-paper baselines.

Piggybank remains a useful finite-design result: pure-HK correction gains 0.166 dB over its original field, while the gate exceeds that corrected field by 0.912 dB. In contrast, Hummingbird’s pure-UV correction is only 0.064 dB below the gate, and its matched-byte native UV is still better. Across nine finite pairs the pure-HK readout gains a mean 1.139 dB but a median of only 0.043 dB, with five

gains, three losses and one tie. The mean is concentrated in three objects. The gate beats its scalar control on seven of nine, but loses to pure-HK correction on six. Piggybank and Poison Dart Frog exceed the tested matched-byte UV in mean-view PSNR; Small Limo’s matched UV reproduces the reference images exactly and is retained as an adverse MSE comparison. None of these observations establishes a universal gate, perceptual dominance or a global Pareto frontier.

Additional information versus additional capacity. For square-integrable target Y and frozen RGB outputs H, U , let $f_H = \mathbb{E}[Y \mid H]$ and $f_{HU} = \mathbb{E}[Y \mid H, U]$. Classical conditional projection gives

$$\begin{aligned} \inf_f \mathbb{E} \|Y - f(H)\|^2 - \inf_g \mathbb{E} \|Y - g(H, U)\|^2 \\ = \mathbb{E} \|f_{HU} - f_H\|^2 \geq 0. \end{aligned}$$

Indeed, $Y - f_{HU}$ is orthogonal to every square-integrable function of (H, U) , including $f_{HU} - f_H$, and the tower property gives the result. The infima range over unrestricted measurable readouts, not the finite networks above. If U is determined by H , it adds no information in this sense; finite architectures can nevertheless differ in optimization and inductive bias. Our experiment tests simpler alternatives without claiming to estimate these conditional expectations or prove that UV is necessary.

Surface acceptance is also not image acceptance. For example, TShirt’s selected HK correction loses 0.125 dB in PBR despite passing surface calibration. The query distribution and rendering objective must remain part of a target-conditioned decision. Pure-HK correction does not query UV, and pure-UV correction does not query HK. Same-query bridge checks and repeated reference images qualify implementation; shared-card timings do not establish an acceleration. All ten objects, checkpoint choices, costs and metric failures remain in the reproducibility ledger.

E.6. Conditional solving and deployment follow-up

These completed pilots reuse existing checkpoints and are separate from the locked training comparison. They test source-free field distillation, conditional solving, execution and small readout extensions. Earlier training outcomes and initializers are unchanged.

Quantity	Accounting and measurement scope
Field or codec bytes	Complete stored field or decoded-and-evaluated bitstream, including its metadata; excludes shared mesh, runtime libraries and spectral state.
UV bytes	PNG plus UV coordinates and texture indices for the expanded comparison; hybrid controls additionally check an already-shared UV stream.
Spectral and response caches	Required spectral state is separate from field bytes; the historical response-table experiment measures additional table storage.
GPU memory	No per-method total resident or peak-memory reduction is established; the paired evaluator benchmark keeps both paths resident.
Fitting and preprocessing	Update counts and recorded stage costs describe the recipes; concurrent expanded runs are not isolated wall-time comparisons.
Query and render latency	Historical paired measurements above; table construction, spectral preparation and loading are separate setup costs.

Table 33. Resources are reported separately. None of the field-byte comparisons is a total-memory, setup-time or latency frontier.

E.6.1. Selection, samples and measurement

We take the first two objects in the fixed identifier order with both a refined and an extended-direct 3,000-kernel field from each of the new Objaverse, GSO, public-main and low-poly cohorts. The objects are Pink Bow, Cartoon Fish, Lactoferrin, Garden Swing, and public assets Jack-O-Lantern, Calabash - Bottle Gourd /dried /punctured, PS1 Chicken and Saul Goodman. Both routes attain 3,000 kernels in this pilot. There are eight objects and 16 correlated fields. Each receives 73,728 newly sampled area-uniform queries (seed 36003), split into 65,536 calibration and 8,192 held-out queries. Quality summaries first average the two routes within each object; field-level timing summaries explicitly count those routes.

Field fitting targets the original unclipped field with its fixed background removed. Source appearance is used only in evaluation. Geometry uses the same twelve-bit barycentric and eight-bit response codec; colour precision varies. We preserve each original per-channel quantizer range, encode the actual fitted integer codes, and evaluate the decoded result. Headers and entropy-coded payloads are charged, so equal bit depth does not guarantee equal final file size. Shared mesh, spectral state and runtime libraries remain external. Six-bit colours decode to FP32 in the measured execution path.

Four RTX3090 GPUs run independent paired jobs, with per-GPU locks. The environment is PyTorch 2.8.0+cu129 and Triton 3.4.0. PBR uses the same five 256^2 views, four samples per pixel and seed 119. Streaming, readout and detail timing use one warm and two measured renders per view; the six-bit coding-only follow-up has one measured render and supplies no timing claim. Scene setup, fitting and image scoring are outside renderer timing. Timing repetition is limited and does not establish cross-device confidence intervals.

E.6.2. Matched-accuracy GPU colour solving

The solver evaluates the calibration quadratic in FP64. Sparse forward and transpose products and the diagonal majorizer are prepared separately. Accelerated proposals are accepted only after checking the true objective; otherwise a majorized step from the current iterate is used. The majorization argument in Section 4 concerns exact-arithmetic descent; numerical checks allow the explicitly recorded tolerance and are not interval arithmetic.

All three controls use the same initialization, matrix, target and GPU. They check $\gamma/(3N) \leq 10^{-8}$ every ten iterations and stop after at most 3,000. This bounds calibration MSE optimization error by 2×10^{-8} . Each field has two timing repeats with reversed method order. Repeats are not independent samples. The global accelerated control uses the valid row-sum curvature bound $L = \max_j (A_d^T A_d \mathbf{1})_j$.

Against the global accelerated control, the six mutually converged fields give median speedup $6.264 \times$ (range 4.550–8.113). The plain diagonal control has only one mutually qualified comparison. No conclusion about the best possible global step, GPU quasi-Newton solver or general optimization algorithm follows. The finite-query gap says nothing about held-out appearance, discrete optimality or simultaneous geometry training.

CPU execution without changing the iteration. A separate four-object implementation check uses the historical Orange, Poison Dart Frog, Small Limo and Pink Ombre Hair response matrices (69,632 samples and 3,001 coefficients including the background). Caching column squared norms and sharing sparse reads across RGB channels preserves the coordinate order, box constraints and stopping rule. On an Intel Xeon Silver 4210R, five alternating repeats of 300 sweeps give a median object speedup of $2.02 \times$ (range 1.93–2.11) over the original CPU loop. Coefficients, residuals and per-sweep histories match exactly in FP64 on

Solver	Qualified fields	Median iterations	Median solve time (s)
Global majorizer, accelerated	6/16	3,000	5.680
Diagonal majorizer	1/16	3,000	6.174
Diagonal majorizer, accelerated	16/16	785	1.471

Table 34. Medians across all fields include unsuccessful capped runs; these are not times to accuracy for the failing controls. Only mutually converged fields enter speed ratios. Preparation costs are recorded separately.

all four matrices. Each solve uses one CPU thread; the host is shared and object jobs can overlap. First-call compilation is logged separately. This check measures warm fixed-matrix solving, excluding response construction and training; it does not establish a pipeline or renderer speedup.

E.6.3. Response-aware colour coding

With fixed channel range l, s and integer code q , a coordinate change Δ changes the squared error by exactly $2\Delta\langle a_j, R \rangle + \Delta^2\|a_j\|_2^2$. The nearest feasible grid point to the unconstrained coordinate minimizer minimizes that coordinate quadratic. We accept strictly improving moves for at most 30 sweeps. A finite grid ensures termination if run to a coordinate fixed point, not necessarily within this cap and not at a global optimum. This specializes established task-aware quantization ideas [15, 21].

Six-bit fitting follows the observed four-bit losses and is therefore adaptive exploration. It improves on nearest six-bit rounding for 15/16 fields and 8/8 object means; the worst loss against eight-bit rounding is 0.450 dB. At four bits, fitting with decoded rather than original responses changes mean PSNR by -0.0017 dB. A continuous majorized initialization before grid fitting changes the mean by only $+0.0012$ dB and improves 3/8 object means. Neither theoretical consistency with the decoder nor a continuous warm start establishes a practical quality advantage in these comparisons.

E.6.4. Streaming evaluation and compiler controls

For source i and spectral index e , query interpolation gives

$$\phi_i(x, e) = \sum_r w_{ir} \sum_v b_v(x) \Phi_r(v, e).$$

The normalized heat response contracts this feature against $\exp(-\lambda_{ie} t_i) \phi_i(i, e)$ and the original source-self divisor and spatial weighting. The fused operator evaluates the interpolation and contraction together, caching only the source-dependent decay in GE floats. It retains the original FAISS search, response filter, inner top- k and colour readout, together with the full spectral state.

This compiler-matched benchmark uses the fitted four-bit fields. The historical output label “Eager” in this pilot retains checkpoint-enabled component compilation; it is not a wholly uncompiled evaluator. The stronger “Compiled” control additionally compiles whole query gathering

and diffusion. All 16 streaming fields pass maximum RGB error 2×10^{-5} on the 73,728 queries against the component-compiled reference, while 15/16 additionally compiled controls pass. On those 15 mutually qualified fields, median speed ratios are 1.256 for full rendering and 1.376 for 512-batch queries. The full-render range is 1.181–1.494. Incremental peak PyTorch temporary allocation has median ratio 0.04497 at query batch size 512. That statistic excludes total resident state and allocations outside PyTorch. The additional decay cache occupies 1.116–3.072 MB. Maximum absolute PBR PSNR change against the component-compiled reference is below 0.0019 dB.

The later six-bit composition disables compilation in all reference components. Its larger speed ratios cannot be substituted for this compiler-matched comparison. It passes only 14/16 query checks. A backend selector uses only the calibration part to choose streaming when maximum error is at most 2×10^{-5} , otherwise reference execution. It passes 15/16 held-out checks: Pink Bow’s direct field passes calibration at 1.597×10^{-5} but fails testing at 4.196×10^{-5} . The threshold is not tuned after this outcome. The policy was designed after inspecting the raw failures, so it is not a blinded confirmation. Its image/time effects reuse measured routes rather than new timing. Backend metadata is serialized and charged; the failed package is explicitly labelled unvalidated.

E.6.5. Observed support switches and the cost of removing one

With a fixed outer candidate list, suppose the k th and $(k + 1)$ th filtered weights meet at $\alpha > 0$ with different colours. The two limiting selected sums have the same denominator D , so their preclip difference is

$$\Delta z = \alpha(c_j - c_i)/D.$$

It need not vanish with the response perturbation. Clipping can suppress some channels; an isolated zero-weight tie has zero jump. This supplies a direct mechanism behind the margin conditions in Appendix B.4.

Replay of Cartoon Fish’s direct compiled field finds two differing queries, identical outer candidates and spectral gather, and a zero inner cutoff gap. The error is localized to the compiled diffusion step. Among 30 unit-weight responses, a swap between colours $(-1, -1, -1)/15$ and

Paired comparison	Mean PBR difference (dB)	Mean field-byte ratio
Fitted four-bit / nearest four-bit	+1.304	1.0079
Fitted four-bit / nearest eight-bit	−1.231	0.8859
Fitted six-bit / nearest six-bit	+0.135	1.0012
Fitted six-bit / nearest eight-bit	−0.082	0.9453

Table 35. Equal-weight object means after averaging the two training routes. Ratios charge complete field bitstreams. These are source-render metrics, although fitting distils the original field.

$(3, -5, -1)/15$ predicts the surviving 0.00888884 RGB jump. This diagnoses the support mechanism without assigning it to one unisolated floating-point instruction.

Using all 50 already-computed responses removes this inner selection. Its normalized readout is continuous in nonnegative weights, with the fixed-support bound from Appendix B.4; outer neighbour switches remain. We refit six-bit colours against the stored decoded-geometry inner-selection response, while rendered comparisons use the streaming inner-selection path. These are different readout families; no inclusion of one function class in the other is assumed.

The first two objects supply the discovery experiment; six more use the locked, unchanged recipe, with no tuning on failures. All 16 fields render faster (1.020 – $1.064\times$); the paired-object median is $1.042\times$. Mean PSNR loss is 0.046 dB after fitting versus 0.139 dB without it. The six expansion objects lose 0.060 dB on average, and low-poly asset Saul Goodman loses 0.187/0.482 dB for its two routes. All are previously explored objects. An explicit readout-mode envelope adds five bytes to each compared field payload; all 48 experimental envelopes round-trip exactly. This saving in selection work does not establish a uniformly preferable representation.

E.6.6. Historical detail extension

A separate replay uses the four available historical objects Apple, Orange, Piggybank and TShirt. A local directional detail augments a fixed parent response by a signed coefficient times $\tanh((x - p_i)^T s_i)$; setting all detail coefficients to zero recovers that model’s own base exactly. The fitted base is not necessarily the original colour-only model. Sixty-four gain-selected units improve mean PBR PSNR by 0.075 dB over colour-only fitting and 0.065 dB over random placement, with 3.23% more bytes and 1.61% more rendering time. This is not an equal-cost comparison. Fresh same-GPU response reconstruction agrees with direct decoding within 1.9×10^{-7} ; historical response-cache discrepancies up to about 2.1×10^{-4} remain separately recorded. Query-only timing was not collected. The modest gains neither replace the negative UV-hybrid result nor establish a new dominant field representation.

E.7. Selecting deployment costs and testing local detail

This final exploratory follow-up asks whether explicit choices can retain the preceding gains. It reuses the same eight objects and 16 fields; the detail control separately reuses four historical objects. Fresh surface samples do not make these new object cohorts. Protocols and thresholds are fixed before execution, with a subsequent equal-sweep control explicitly identified below. All comparisons disable compilation and retain the five diagnostic PBR views. They do not replace the compiler-matched speed comparison in Appendix E.6.

E.7.1. Source-assisted precision and readout choices

We draw 32,768 area-uniform calibration queries and 65,536 separate test queries (seeds 44001 and 44002). Candidate colours were fitted earlier; this stage selects among their decoded outputs using native-source surface MSE. For precision, choose the smallest complete four-, six- or eight-bit field whose calibration MSE is at most $10^{0.05/10}$ times the eight-bit MSE. Eight-bit execution is always eligible. For readout, choose the fitted all-response field only if its calibration MSE is at most the same factor times inner-selection MSE; otherwise retain inner selection. Choices are written before test predictions. These are two separate source-assisted policies, not source-free coding or a jointly validated configuration.

After averaging routes within objects, precision selection saves 2.882% of complete field bytes at a mean PBR loss of 0.021 dB. Two refined fields, Lactoferrin and low-poly asset 531a8489, lose about 0.083 dB despite passing the surface criterion. Readout selection retains inner selection on both routes of that low-poly asset. It gives mean full-render speed ratio $1.036\times$ and mean PBR loss 0.004 dB; the worst selected field loses 0.034 dB. All 16 held-out surface and diagnostic PBR differences are within 0.05 dB in this finite pilot. This improves the observed readout trade-off, but it is not independent-object validation or a uniform error certificate.

Every candidate and reference pays the same five-byte mode header in addition to its original field payload. All 64 deployed policy envelopes recover the exact selected payload and mode. Geometry and backgrounds are identical

Policy	Selected fields	Surface within 0.05 dB	PBR within 0.05 dB
Precision	9 six / 7 eight	16/16	14/16
Readout	14 all / 2 inner	16/16	16/16

Table 36. Each policy uses its own reference: eight-bit coding for precision, inner-selection streaming for readout. Counts are fields, not independent objects; surface queries are held out and PBR views are existing diagnostics. Four-bit coding is never selected.

across precision/readout variants. Shared mesh and spectral state remain external, and decoded coefficients remain FP32. The policies do not turn bitstream savings into equal resident-memory savings.

E.7.2. Why query-wise fallback still fails

A separate evaluator keeps the original outer candidates and falls back to reference gathering and diffusion on queries whose filtered inner cutoff margin is at most 2×10^{-5} and whose cutoff weight exceeds 10^{-5} . These are heuristic thresholds, not certified response-error bounds. We retain the original top- k call shape and never rerun neighbour search on the compacted fallback queries. Tests include both fresh query sets above and the previous 73,728-query stress set, with maximum RGB tolerance unchanged at 2×10^{-5} .

The raw streaming path passes 14/16 fresh held-out checks; query-wise fallback passes 15/16, but only 14/16 fields pass all three query sets. Both Pink Bow routes fail at least one set. Fallback covers about 10.0% of test queries on average. Its mean render speed ratio relative to raw streaming is only $0.583\times$, so detecting and compacting queries also has a substantial cost. It is not adopted as a reliable default.

Post-outcome replay finds identical ordered outer IDs at the worst Pink Bow queries. Nevertheless, reference evaluation on the compacted subset differs from reference evaluation in the original batch: the refined calibration case reaches 1.111×10^{-4} RGB error; the direct test case reaches 2.241×10^{-5} . Thus preserving candidate identity is insufficient to reproduce the original numerical path. The replay isolates the compacted reference path, not one specific floating-point instruction. No threshold is changed after these failures.

The distinction between surface and image error also has a local mathematical interpretation. For coefficient perturbation δ , let $B\delta$ be the stacked field-colour change on a fixed query/support set. At a differentiable rendering configuration with image Jacobian J , the squared rendered change is $\delta^T B^T J^T J B \delta + o(\|\delta\|^2)$, whereas unweighted surface fitting uses $\delta^T B^T B \delta$. The metrics need not agree. This identifies a possible role for rendering-sensitive calibration; no such Jacobian-weighted solver or guarantee across visibility, clipping or support changes is claimed here.

E.7.3. Smooth details versus extra fitting and hard splits

Apple, Orange, Piggybank and TShirt receive fresh source-supervised train, validation and test queries: 32,768/8,192/32,768, with seeds 45001/45002/45003. A fixed-geometry base receives 300 coordinate sweeps. Starting from it, 32- and 64-unit variants receive another 300 sweeps, using either a smooth tanh gate or a hard sign gate. Both families derive centres and two principal directions from training responses, independently rank one-coordinate residual gains, and use the same child-colour box constraints. All fields are losslessly serialized and decoded before evaluation; the mode header is charged. This is a fixed-iteration comparison, not a converged representation-capacity bound.

Within the smooth 64-unit field’s actual byte ceiling, validation MSE selects among the base and hard 32/64-unit variants. It chooses 32 hard units for Apple, Orange and Piggybank, and the base for TShirt. The hard 64-unit fields exceed the byte ceiling on the latter two objects; they are still reported separately as equal-count controls. The candidate set does not exhaust all possible allocations or byte budgets.

Because the initial base had fewer fitting sweeps, a post-outcome fairness audit continues that base for another 300 sweeps on the same training data. It changes no detail parameters or locked validation choices and preserves the original results. The base and detail fields now each have 600 total sweeps, though their arithmetic work per sweep differs. The base’s final half-MSE gap ranges from 1.32×10^{-7} to 2.05×10^{-6} ; this control is not an exact conditional optimum.

The mean gain over the equally continued base is 0.119 dB, with 2.647% more field bytes and 1.622% more rendering time. All four PBR differences and all four fresh test-surface differences are positive. Continuing the base alone changes mean PBR PSNR by about -0.001 dB. Thus the measured detail gain survives this extra-fitting control. Its advantage over hard details is modest, timing has only two measured repetitions, and the objects are historical selections. These results support a small optional allocation, not a new dominant representation or a global Pareto frontier. All 24 detail/control envelopes recover the exact serialized payload and mode; geometry matches the fixed runtime initializer.

Object	Smooth 64-unit PBR advantage (dB)		
	600-sweep base	Hard 64-unit	Selected base/hard
Piggybank	+0.112	+0.027	+0.077
Apple	+0.103	+0.045	+0.065
TShirt	+0.045	+0.013	+0.042
Orange	+0.213	+0.006	+0.054
Mean	+0.119	+0.023	+0.060

Table 37. Four historical objects; all differences are paired within their respective runs. Equal-count hard fields may have different encoded bytes. The last column uses the locked validation-selected, byte-feasible control; it does not claim the best possible competing method.

F. Early routing: a failed speed hypothesis

F.1. A frozen pilot with a stopping rule

Four historical 3,000-kernel fields use 8,192 fitting, 4,096 calibration and 16,384 test queries. No source-field colours or geometry are retrained. Intrinsic search still finds 50 candidates. A distance rule or a per-source quadratic score ranks them before expensive spectral evaluation. The score has ten polynomial terms in normalized relative position, ridge 0.1, and targets $\log(\max(w, 10^{-6}))$. FP16 coefficients and scales are decoded to FP32 for calibration and execution.

Fixed budgets are $\{8, 16, 24, 32, 40, 50\}$. An adaptive predicted-tail rule uses eight tolerances from 0.000125 to 0.016 and reserves $\{0, 8, 16\}$, then rounds up to a budget. Each family selects the smallest average count with calibration loss at most 0.05 dB, breaking ties on calibration MSE. Full evaluation is an admissible fallback. The predictor is not a certified error bound. The retained responses use their actual partial normalizer.

The after-response controls first evaluate all candidates, then select among 8/16/24/30 terms using true responses; adaptive choices also check actual field distortion. They measure a conditional opportunity, not an accelerated method or the exact minimum support. The surface operating point is not selected on rendered outcomes.

Original views use azimuth $180 + 72j$ degrees, zero elevation and seed 119; shifted views use $216 + 72j$, elevation 15 degrees and seed 120. Both have five frames and the same 512-query bridge. The shift changes both cameras and randomness, so it is not a causal camera-only intervention. Four RTX 3090s each run one object with a shared host. Compiler settings are matched within this pilot, with whole-query compilation disabled and Triton kernels active.

Expansion requires every object to lose at most 0.05 dB on both view sets and median complete-render speed ratio at least 1.20, with no nonfinite values. The speed criterion fails and the pilot stops. This is an engineering stopping rule, not a publication acceptance criterion.

Table 38. Primary early-route outcomes. Time ratio is full streaming divided by routing; below one means routing is slower. Orange is an unchanged full fallback, not a routing success.

Object	Mean K /50	Δ PBR (dB)		Time ratio
		Original	Shifted	
Orange	50.00	+0.0043	+0.0004	1.012
SmallLimo	32.16	−0.0402	−0.0453	.430
Hummingbird	45.17	+0.0029	−0.0004	.401
Piggybank	43.51	−0.0002	+0.0000	.399

F.2. The practical gap between selection and sparsity

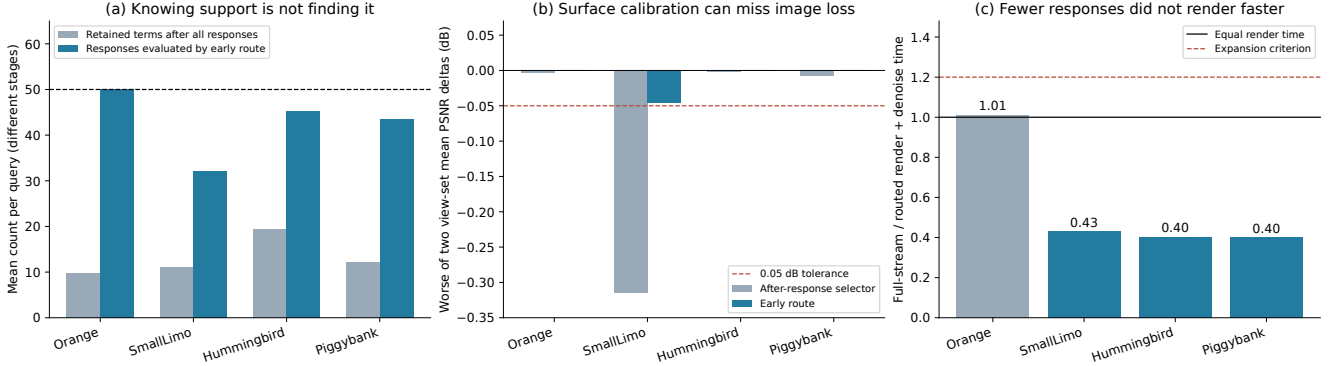
Distance selection retains 40 candidates for Orange and falls back elsewhere. Fixed quadratic selection falls back on all four objects. Adaptive routing reduces response work on three but increases complete-render latency. It also adds about 61 KB of file bytes and 132 KB of decoded score tensors, retaining the global basis. The fallback object need not deploy its rejected router.

After-response adaptive counts are 9.75, 11.13, 19.37 and 12.22 for Orange, SmallLimo, Hummingbird and Piggybank. These are counts after computing the responses, not early evaluation savings. SmallLimo’s adaptive surface loss is only 0.026 dB but its rendering losses are 0.201 and 0.315 dB. The fixed-24 control even improves surface PSNR by 0.067 dB while losing 0.477 and 0.459 dB in images. Thus these points do not define a matched-render-quality oracle frontier.

F.3. Fusion, failure preservation and cost scope

An implementation-only follow-up keeps policies and thresholds fixed. A masked streaming kernel skips inactive spectral loads, and a second variant fuses the geometric score. All native predictions match bucketed routing within 2.98×10^{-8} . Fusion roughly halves query overhead on the three non-fallback fields, but still costs about $1.8 \times$ full streaming in the same follow-up.

Three objects fail the strict cross-run uint8 baseline replay check. SmallLimo and Piggybank change at most one



Four historical fields; no source-field retraining. Orange uses the unchanged full fallback. After-response selection computes all 50 responses first. Timings show the complete primary run; failed fusion replays are excluded.

Figure 11. Early-routing diagnosis. The after-response control already pays for all responses. Surface calibration can miss rendering loss, and the tested early routing fails to improve latency. The plotted timings use the complete primary run; failed fusion replays do not enter a qualified complete-render aggregate.

uint8 level; Orange reaches four for the original evaluator and seven for full streaming. Its sparse raw-linear differences reach 0.187 and 0.713, respectively. All available repeated native-UV reference images are bitwise identical. The original evaluator also changes, so no attribution to routing or the original theory is established. Hummingbird completes; its full/routed render ratios are 0.634 and 0.647. The three failed follow-ups remain excluded from qualified complete render claims. A uint8 equality check does not imply bitwise raw-linear equality.

An interrupted evaluation resumed from unchanged cached responses and fitted routing weights. No policy was refitted during recovery. The recovery wall time excludes the original attempt; both records and the complete observed interval are retained. Query and steady-state frame times exclude startup. Cache generation in this diagnostic also includes calibration/test response collection, so it is not advertised as a minimal production preprocessing recipe. All-method PyTorch peaks do not include Mitsuba or FAISS allocations and are not total per-method GPU memory.

For a measured workload, let extra preparation cost be $P \geq 0$, baseline frame time T_f , routed frame time T_r , and reuse n . Total savings require $n(T_f - T_r) > P$. If routing is slower in steady state, reuse alone cannot make it faster in this workload. No claim is made that every possible router must fail.

G. Where do the interventions help?

G.1. Fitting state and object category are different hypotheses

The same object can benefit from one intervention and not another. A conditional colour correction changes op-

timization at a fixed response matrix; a feature construction changes that matrix; a streaming evaluator changes its evaluation cost. These interventions need not share an applicability class. Table 39 summarizes observed conditions without treating different experiments as one combined method or pooling their objects into a larger independent sample.

The warm-start and reserved direct-fit cohorts differ in both membership and protocol. Their effect difference alone does not causally identify optimization maturity. Within-object continuation controls, empirical fitting gaps and independent-sample checks provide the more direct evidence. Likewise, a high surface residual is insufficient for a useful residual branch: it must align with available responses, survive rendering and justify its storage and evaluation cost.

G.2. An exploratory geometry-only grouping

We test whether observable geometry explains which objects benefit. This is a post-hoc analysis of completed experiments: aggregate effects were already known. Descriptor definitions, two clusters and the prediction rule were fixed before computing associations, but this does not make the analysis an independent confirmation.

We keep two cohorts separate: 19 complete heat-feature objects with two seeds per object, and 19 complete conditional-correction objects with three seeds. Effects are averaged over the prescribed seeds before analysis. The cohorts share 18 objects and are not 38 independent observations. In the second cohort the contrast is correction minus reset-only. Failures retain their original status; absence of a measurement is not a zero effect.

Five descriptors use the actual converted meshes: log vertex count, log of one plus the coefficient of variation of triangle area, two surface-covariance shape ratios, and mean tri-

Table 39. Observed applicability and limits. These are experimental conditions, not a validated classifier of unseen objects. Effects refer to the specific controls and costs stated in the corresponding appendix.

Intervention	Evidence of usefulness	Where usefulness is not established
Conditional colour fitting	Four historical warm starts: +0.586 dB versus matched optimizer resets with joint continuation; +0.146 dB with geometry fixed.	Nineteen reserved continuous direct fits: only +0.019 dB on average. Lower empirical fitting loss does not ensure improved rendering.
Subsequent geometry updates	After colour fitting, Orange and Frog gain 1.618 and 2.790 dB over colour-only continuation in the historical factorial.	Small Limo gains 0.006 dB and Hair 0.064 dB in that same comparison. Benefit depends on the starting field and update protocol.
Fixed-time heat features	Nineteen complete objects: mean +0.277 dB, median +0.096 dB, 15 positive objects.	Four objects lose quality; one large gain strongly affects the mean. No reliable geometry-based selection rule is established below.
Streaming evaluation	Median 1.26× render speedup on 15 accuracy-qualified fields; smaller query-dependent temporary storage.	Accuracy qualification excludes a failed field; shared spectral storage remains. Extra launch tiling does not improve the tested streaming implementation.
UV and kernel combinations	A small learned-mixture case is competitive when an already resident UV base is treated as shared.	Fixed-support residuals lose to spending bytes on finer UV in all 20 tested settings. Strong pure-branch controls and charged UV storage remove broader superiority claims.
Smooth detail allocation	Four historical objects gain 0.119 dB with 64 smooth detail units under the matched 600-sweep comparison.	Field bytes rise 2.647% and render time 1.622%. This is a small quality–resource trade-off, not a free speedup or a validated universal allocation rule.

angle quality. If $\lambda_1 \geq \lambda_2 \geq \lambda_3$ are the covariance eigenvalues of the uniform surface measure, the ratios are $\sqrt{\lambda_3/\lambda_1}$ and $\sqrt{\lambda_2/\lambda_1}$. Surface moments are integrated over triangles, rather than estimated from unweighted vertices. Triangle quality is the area-weighted mean of $4\sqrt{3}a/(l_1^2 + l_2^2 + l_3^2)$, where a and l_i denote area and edge lengths. These descriptors capture discretization and global shape, but not geodesic complexity, visibility or response support completely.

Within each cohort, standardized descriptors enter two-cluster k -means with 30 initializations and a fixed random seed. Cluster fitting uses no quality outcomes. We number the lower-vertex-count centroid first. The cohorts have separately fitted clusters, so equal cluster numbers do not identify the same geometric population. Figure 12 shows all outcomes against vertex count, one of the five inputs.

The first heat-feature cluster contains Banana, wrench and Doormate. It has coarser and thinner meshes on these descriptors and a mean gain of 1.194 dB, initially suggesting a favorable geometric regime. However, its median is only 0.073 dB. Doormate alone gains 3.505 dB; without it, the other two average 0.038 dB, below the second cluster’s 0.105 dB. The full 19-object mean remains primary. The 18-object sensitivity mean is 0.098 dB, close to the full-cohort median. This distinction prevents an outlier-driven subgroup narrative.

For conditional correction, cluster means are 0.022 and 0.016 dB. A two-sided permutation test of the cluster mean contrast, shuffling effects within each dataset, gives $p =$

0.186 for heat features and $p = 0.791$ for correction (10,000 permutations, plus-one correction). Cluster stability is also limited: the minimum adjusted Rand index when refitting after leaving out one object is 0.027 and 0.081. Corresponding medians are 1.000 and 0.777, and silhouette scores are 0.364 and 0.238. Stable assignments for most deletions do not establish a natural taxonomy.

G.3. Associations and held-out prediction

We additionally measure RGB variance and mean squared RGB difference to the nearest Euclidean neighbor on the first 16,384 points of each object’s existing seed-zero training pool. This is a limited texture descriptor, not an intrinsic surface gradient: nearby sheets can be Euclidean neighbors. Baseline held-out surface PSNR is a separate difficulty descriptor requiring a fitted control and target access; it is not a geometry-only deployment input.

For each cohort we test rank associations between the effect and five geometry descriptors, two texture descriptors and baseline surface PSNR. Together with the two cluster contrasts these form one family of 18 tests. Two-sided permutation tests and Benjamini–Hochberg adjustment give no adjusted value below 0.83. Vertex count versus heat-feature gain has Spearman correlation -0.400 ; flatness versus correction gain has correlation -0.381 . These weak trends are hypotheses for new objects, not confirmed selection criteria. Small samples, restricted recruitment and unmeasured response properties limit what negative results rule out.

Prediction uses a fixed ridge penalty of 10 without hyperparameter search. Standardization and fitting use only the

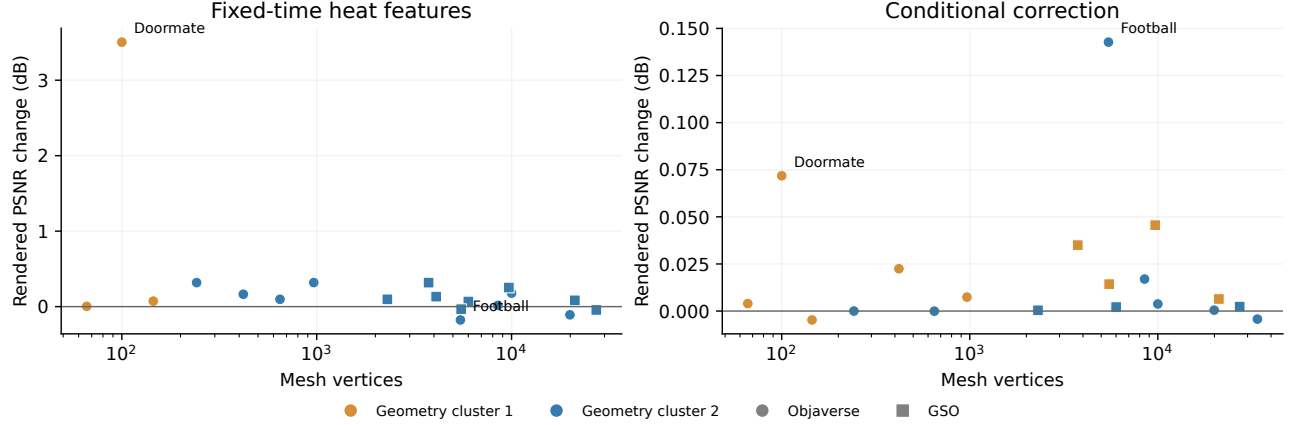


Figure 12. Exploratory geometry grouping and object-level effects. Points average the prescribed seeds. Colours denote clusters fitted independently within each cohort; marker shapes distinguish datasets. All effects are retained, including Doormate’s large heat-feature gain. Football loses quality under the heat-feature change but benefits from conditional correction. Vertex count alone is not the clustering rule.

Table 40. Geometry grouping and sensitivity. Gains are rendered PSNR changes in dB. Removing the largest gain is a descriptive sensitivity check only, not a replacement for the complete primary result.

Cohort	Group	Objects	Positive	Mean	Median
Heat features	Cluster 1	3	3	1.194	0.073
	Cluster 2	16	12	0.105	0.097
	Cluster 1 without Doormate	2	2	0.038	0.038
	All objects	19	15	0.277	0.096
	All except Doormate	18	14	0.098	0.090
Conditional correction	Cluster 1	9	8	0.022	0.014
	Cluster 2	10	7	0.016	0.001
	All objects	19	15	0.019	0.004

18 training objects in each leave-one-object-out fold. The baseline predicts their mean effect. Geometry-only prediction has mean absolute error 0.440 dB versus 0.372 for the baseline on heat features, and 0.0250 versus 0.0246 dB on correction. Adding the texture descriptors raises these errors to 0.482 and 0.0260 dB. Baseline surface PSNR alone also does not improve the mean baseline. These controls do not justify a learned router from this sample.

Across the 18 shared objects, the interventions’ gains have Spearman correlation 0.123 (exploratory permutation $p = 0.619$, separate from the preceding family). Football changes by -0.177 dB under the feature intervention and $+0.143$ dB under correction; wrench changes by $+0.073$ and -0.005 dB. These are different experimental contrasts, not direct head-to-head rankings. Their disagreement reinforces the need to diagnose the intervention rather than infer a universal category from object names.

G.4. Practical scope

Current evidence supports organizing applicability by the limiting operation: remaining conditional colour error, failure to generalize from fitting samples, a response limitation after colour compensation, or evaluation cost. Geometry can influence each without being a sufficient predictor itself. We do not recommend replacing per-field validation with an animal/vehicle, smooth/detailed or thin/volumetric classification; neither do we establish that shape is irrelevant. A future selection rule should fix indicators and thresholds before a new cohort, charge the cost of obtaining them, and test whether selection improves the stated deployment objective.

H. What a better conditional fit does and does not establish

This analysis asks a narrower question than designing another texture model: *what prevents a more accurate conditional solve from improving the deployed image?* We

separate finite-sample fitting, the size of the induced field change, and the interpolation and approximation conventions of the evaluator. The identities below are classical algebra and risk arguments. Our evidence concerns their consequences for the tested HKTex fields, not the invention of least squares, importance sampling, or thermal-response interpolation.

H.1. Solver accuracy and sampling are distinct

For a fixed feasible function family $\mathcal{F}$, let R and $\hat{R}$ be the population and empirical versions of the *same* squared risk. If $\hat{R}(f) \leq \inf_{g \in \mathcal{F}} \hat{R}(g) + \gamma$, then

$$R(f) - \inf_{g \in \mathcal{F}} R(g) \leq \gamma + 2 \sup_{g \in \mathcal{F}} |R(g) - \hat{R}(g)|. \quad (18)$$

To see this, add and subtract empirical risk at f and at a population minimizer, and apply the empirical optimization inequality; approximate minimizers give the same result when the infimum is not attained. Thus a small empirical gap does not bound the sampling term. Neither term in this statement is automatically a bound on clipped or rendered error. The familiar orthogonal projection decomposition additionally requires an unconstrained linear space; box-constrained colour fitting is not that projection.

The adaptive-basis viewpoint already motivates solving linear output variables while adapting nonlinear features [7]. Stable weighted least squares also depends on the response space and sampling measure [5]. Frozen-PINN offers a recent example of separating basis construction, linear fitting and numerical evolution in a different task [8]. Its PDE guarantees and timings do not transfer to this texture problem.

Observed sample transfer. We inspect all four historical development objects at updates 250, 500 and 750. For each event we retain the initial field and fit the same bounded colour problem using the original 16,384 queries, a fresh set of 16,384, or its nested 65,536-query extension. Native surface validation and test queries remain separate. Table 41 reports full 100-sweep updates. The mean independent-test/fitting energy ratio of the colour-induced field change falls from 1.232 to 1.055 with the larger fresh pool; the fraction of colour entries on a box bound also falls. These measurements describe sampling sensitivity, rather than establishing a new predictor of object difficulty.

A fixed 128-dimensional Rademacher projection provides a further diagnostic. We whiten its response covariance using a separate, target-free 16k anchor and a ridge of 10^{-6} times the mean eigenvalue. The mean spectral norm of the projected fitting/test Gram difference is 0.491 for fresh 16k and 0.450 for nested 65k. The latter pool includes the anchor: this is not an independent whitening comparison. Projection can miss weak localized directions, and these numbers do not certify the full 3,001-column problem.

Fit samples	Test Δ	Energy ratio	Active (%)
Original 16k	-0.713	1.232	19.5
Fresh 16k	-0.801	1.232	19.5
Nested 65k	+0.061	1.055	17.0

Table 41. Twelve correlated events on four objects, 100-sweep full colour updates. Test Δ is surface PSNR in dB. Energy ratio is independent-test over fitting mean squared field change. Active is the fraction of non-background colour entries at a box bound.

Solve	Sampling	Full	Selected
100 sweeps	Uniform	-1.061	+0.070
100 sweeps	Projected leverage	-1.140	+0.049
Gap stopped	Uniform	-1.099	+0.071
Gap stopped	Projected leverage	-1.178	+0.049

Table 42. Surface-test PSNR change (dB), four objects and three sampling seeds. Both replacement samplers estimate the same finite-pool quadratic risk. Selected steps use only surface validation; projection, weighting and selection costs are additional.

Queries, checkpoints and seeds are not counted as independent objects.

An objective-preserving sampling intervention. We next use a common 65,536-row finite pool at update 500. For three fixed seeds per object, both arms draw 16,384 rows with replacement. Uniform sampling is compared with a half-uniform, half-projected-ridge-leverage mixture p_i . Each sampled row and target are multiplied by $(Np_i)^{-1/2}$. For any fixed coefficients, this preserves the expected finite-pool quadratic loss:

$$\mathbb{E}_{i \sim p} \frac{\|a_i C - y_i\|^2}{Np_i} = \frac{1}{N} \sum_i \|a_i C - y_i\|^2. \quad (19)$$

The uniform component bounds the inverse-probability weight by two. This is leverage only in the projected space, not exact full-space leverage. The validation step grid is $\{0, 1/8, 1/4, 1/2, 1\}$; test targets are used only for scoring. Replacement sampling is distinct from the earlier without-replacement fresh 16k subset.

The 100-sweep fits have mean half-MSE box gaps of roughly $3 \cdot 10^{-5}$, so solver error must not be relabelled sampling error. We therefore repeat the update-500 comparisons on the same sampled matrices with a 10^{-7} empirical gap target and a 4,000-sweep cap. All 36/36 fits meet the target; the largest achieved gap is $9.91 \cdot 10^{-8}$. The gap-stopped fresh 16k and nested 65k full steps give -0.793 and $+0.117$ dB on the four update-500 fields, respectively. These are four fields, not the twelve-event aggregate above. For replacement sampling, projected leverage

minus uniform after validation selection is -0.022 dB, positive in 2/12 paired seeds. This higher-accuracy follow-up preserves the negative conclusion for the tested sampler. It was motivated by the initial outcomes and is an exploratory robustness check, not a new preregistered confirmation.

The weighted intervention does not produce a useful gain under this protocol. That result constrains this particular projected sampler, not the established theory of optimal sampling. It also illustrates why changing sampling without preserving the risk would confound sample allocation with a new objective. Projection, probability construction, weighted fitting and validation must all be charged; the additional projection is not free at deployment.

H.2. A useful image direction can be too large

Let I be the initial rendered image, Y its reference, and $I + \delta$ a new rendered image. With mean-normalized inner products,

$$L_1 - L_0 = 2\langle I - Y, \delta \rangle + \|\delta\|^2. \quad (20)$$

For nonzero norms, define $\rho = -\langle I - Y, \delta \rangle / (\|I - Y\| \|\delta\|)$ and $u = \|\delta\| / \|I - Y\|$. Improvement is equivalent to $u < 2\rho$. Positive correlation with the needed image change is insufficient if its magnitude is too large.

In the four saved full colour replacements, ρ ranges from 0.287 to 0.392, but $u > 2\rho$ in every case; the prior guarded updates improve. This holds in both sRGB and raw linear output. These are post-hoc image identities, not causal identification or a deployable image-line oracle. In particular, blending two output images is not the same operation as interpolating their model coefficients and rendering again.

Actual parameter intervention. We therefore render the actual path $C(\alpha) = C_0 + \alpha(C_1 - C_0)$, where C_1 is the previously saved full conditional solution, while keeping all geometric and response parameters fixed. We use five pre-declared fractions and render each at 128^2 , 256^2 and 512^2 with the original five cameras, lighting, materials, four samples per pixel and fixed seed. The denoiser is rebuilt at the actual resolution, with no resizing or cross-frame state. The colour fraction is selected on the surface validation set *before* examining these images, yielding 1/4 for all four objects.

The selected quarter-step improves all twelve object-resolution conditions, with object means of $+0.516$, $+0.344$ and $+0.193$ dB respectively. Full replacements give -0.912 , -1.169 and -1.224 dB. The mean pooled raw-linear changes at the quarter-step are also positive: $+0.148$, $+0.124$ and $+0.097$ dB. The raw-linear and mean-view sRGB metrics are different summaries and are not substituted for one another. Half-steps can improve the denoised sRGB metric while reducing the mean raw-linear metric.

Thus neither a single coefficient fraction nor one output metric describes the whole deployment response.

Both query-wrapper and batch-size agreement gates are below $2 \cdot 10^{-5}$. At 256^2 , the native reference images exactly match the prior renders; replayed field images differ by at most two UInt8 levels, consistent with retained numerical rendering differences. All new comparisons use their own paired initial render. These are instantaneous frozen-field changes, not the continuation gains reported in the main paper. They use four repeatedly studied development objects, not a new generalization cohort. The result supports guarded updates in these examples; it does not identify aliasing as the cause or establish a universal fraction. Sampling-rate studies such as Mip-Splatting [29] motivate this check without supplying a texture-specific guarantee.

H.3. Local redundancy is mass times colour contrast

For a fixed candidate set, partition responses into retained K and removed T . Write $S_K = \sum_K w_i$, $S_T = \sum_T w_i$, $S = S_K + S_T$, and $N_K = \sum_K w_i c_i$, $N_T = \sum_T w_i c_i$. The exact unclipped change, with a shared unchanged background, is

$$f_K - f = N_K \left(\frac{1}{\max(1, S_K)} - \frac{1}{\max(1, S)} \right) - \frac{N_T}{\max(1, S)}. \quad (21)$$

If $S > 1$, $S_K \geq 1$ and $S_T > 0$, this becomes

$$f_K - f = -\frac{S_T}{S} (\bar{c}_T - \bar{c}_K). \quad (22)$$

The proof is expansion of the two normalized weighted means. RGB clipping is nonexpansive, so the output L^∞ error is at most $(S_T/S) \|\bar{c}_T - \bar{c}_K\|_\infty$. The coarse bound $2C_{\max} S_T/S$ ignores this local colour contrast.

For a fixed top-30 to top-24 deletion on the four saved fields, the normalized regime covers 64.6–96.0% of queries. On applicable queries, the ratio of mean contrast bound to mean coarse bound is 0.125–0.265, a 3.8–8.0-fold tightening. All 32,768 saved fitting/validation/test queries per field are pooled for this descriptive audit; none select a routing policy. The identities and clipping bounds pass numerical checks. Other denominator-floor regimes use Eq. (21).

This explains why small response mass alone is an incomplete error indicator: similar retained and removed colours can make deletion benign. However, evaluating the tail mean itself costs work. The identity is an *a posteriori* diagnostic, not a free early-selection certificate, and does not reverse our measured routing slowdown.

A fixed-readout non-nesting example. Even kernel count need not order fixed-response approximation power when normalization and bounded coefficients interact. On two queries, let $W = I_2$, background $1/2$, and coefficients

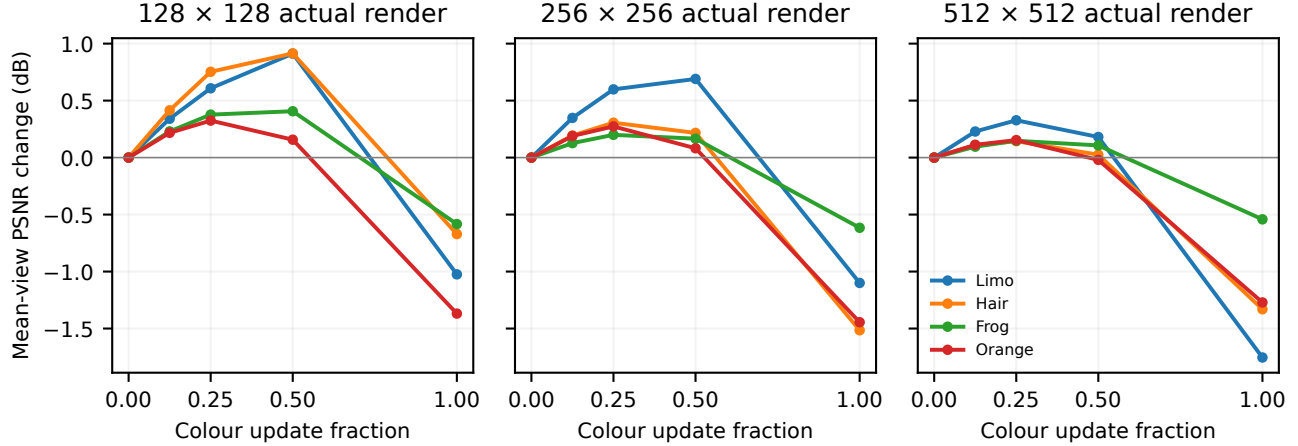


Figure 13. A smaller actual colour update improves the rendered image where the full update fails. Geometry, support and reference material are fixed; each resolution is rendered directly. Curves are descriptive and are not used to select the reported surface-validation fraction. The optimal image fraction varies with object and resolution.

Object	128 ²		256 ²		512 ²	
	Partial	Full	Partial	Full	Partial	Full
Limo	+0.608	-1.025	+0.599	-1.100	+0.327	-1.754
Hair	+0.753	-0.671	+0.306	-1.515	+0.144	-1.331
Frog	+0.377	-0.583	+0.200	-0.616	+0.148	-0.542
Orange	+0.325	-1.369	+0.273	-1.445	+0.154	-1.271

Table 43. Actual colour-path renders: change in mean-view PSNR (dB) relative to the same-resolution initial field. Partial means $\alpha = 1/4$, selected on surface validation for all four objects before rendering; full means $\alpha = 1$. All three resolutions use five original cameras. These are four reused development objects, not twelve independent objects.

$(1/2, -1/2)$; the output is $(1, 0)$. Add three unit columns shared by both queries. Row sums become four, and any coefficients in $[-1, 1]$ yield an inter-query difference at most $1/2$, which clipping cannot increase. The closest output to $(1, 0)$ is $(3/4, 1/4)$, attaining MSE $1/16$ instead of zero. This is a readout counterexample, not proof that these matrices arise from a particular heat-kernel geometry or that fully retrained larger HKTex models are weaker. If extra kernels can be switched off without changing routing, nesting can be restored. Equal-update experiments, rather than this algebraic example, support our compression attribution claim.

H.4. Identical nodes do not define identical interpolation

Let nodal heat factors be F_j , with convex interpolation weights w_j . Factor interpolation uses $K_F = \bar{F}\bar{F}^\top$, where $\bar{F} = \sum_j w_j F_j$; response interpolation uses $K_R = \sum_j w_j F_j F_j^\top$. Their difference is exactly

$$K_R - K_F = \sum_j w_j (F_j - \bar{F})(F_j - \bar{F})^\top \succeq 0. \quad (23)$$

Expansion proves this familiar covariance identity. It quantifies a spectral-energy difference, not entrywise dominance or a texture-quality ordering. Orthogonal changes $F_j \mapsto F_j Q_j$ preserve each nodal kernel but generally change K_F between nodes. A shared Q preserves both constructions; independently chosen node gauges need not. An alignment convention therefore helps define the interpolated model.

We isolate this issue using the exact same saved FP64 heat factors on Barricade and Wrench, 128 fixed vertex probes, and nine off-grid parameter points per mesh. A fresh eigensolution with the same spectral rank supplies the off-grid comparison, not an infinite-rank heat-kernel ground truth. Response interpolation lowers relative discrepancy at 9/9 points on Barricade but only 4/9 on Wrench (Table 44). Factor/response trace ratios average 0.915 and 0.950. Finite differences in angle and log-anisotropy also record different parameter derivatives. Deterministic sign changes leave every nodal kernel unchanged while substantially changing off-grid factor interpolation, as the identity predicts. This sign intervention is deliberately synthetic; it is not evidence that HKTex’s alignment routine produces arbitrary sign errors.

Mesh	Factor error	Response error	Trace ratio
Barricade	0.143	0.132	0.915
Wrench	0.082	0.087	0.950

Table 44. Same nodal heat factors, nine off-grid parameter points per mesh. Error is relative Frobenius discrepancy from a fresh same-rank off-grid eigensolution on 128 fixed vertices. Trace ratio is factor over response interpolation. These operator metrics are not rendered texture quality.

LocalHK/ThermalRF already distinguishes aligned spectral interpolation from heat-semigroup consistency and uses response mixing [18]. We do not claim that distinction as new. The mixed off-grid results here prevent attributing our 19-object texture gains to a universally more accurate continuous operator. Node consistency, interpolation behaviour, optimization conditioning and the learned image objective remain separate.

H.5. Implications and remaining questions

These controls support a specific interpretation: compact fitting should be judged by a correction’s transfer to the required field and image objective, not just its empirical optimum. Smaller solver gaps, more samples, more kernels and more accurate operators are different interventions; none alone guarantees a better image at lower total cost. The strongest new positive result is the actual parameter-path experiment, while projected sampling and off-grid response mixing provide useful limits.

Neural Harmonic Textures separates local spatial support from appearance capacity in another representation [6]; it motivates asking which response restrictions matter without making its benchmarks comparable to ours. Sharing spatial support across mip levels and material channels, as in Gaussian Texture Compression [31], suggests a further question: can one intrinsic support serve several deployment scales or channels? Our RGB assets and current point-colour fit do not establish that result. Answering it requires matched material channels, explicit pixel-footprint targets, and charging all per-level coefficients and evaluator state. This remains a future experiment, rather than an untested addition to the method or a claim implied by the present resolution sweep.