WHEN DO COORDINATES HELP PROJECTION-TO-VOLUME LEARNING?

Anonymous authors
Paper under double-blind review

ABSTRACT

Positional coordinates can resolve spatial ambiguity in learned 3D inverse problems, but their benefit can reverse as measurement coverage increases. We study this in 3D lung segmentation from simulated X-ray projections, separating position from calibrated lifting, attention, and decoder processing. On 24 held-out LCTSC patients under a prospectively locked protocol, coordinates increase source-acquisition Dice by 30.8 percentage points at one view but reduce it by 6.3 at twelve across three initializations. Two-view source acquisitions show a 19.5-point benefit when clustered and a 3.6-point penalty when spread. In our pipeline, sparse-view gains persist without learned attention or the separate geometry-conditioning branch, but accuracy drops substantially without both coordinates and learned spatial slots. Across two development cohorts, along-ray depth exceeds detector-plane coordinates by 9.7 and 3.1 source-acquisition Dice points across three initializations, with advantages persisting under held-out rotations, consistent with difficulty localizing along projection rays. Synthetic coordinate benefits diminish or disappear for uniformly translated populations and depend on sampled frequencies. In a matched context experiment, wider decoder context recovers nearly half of the single-view coordinate advantage without positional input. A scalar acquisition-dependent gate preserves prespecified sparse and broad endpoints in both development cohorts; broader training coverage only partly repairs its intermediate-angle deficit. Coordinate utility therefore depends jointly on measurement ambiguity, population alignment, and decoder processing; neither view count nor geometry alone predicts its benefit.

1 INTRODUCTION

A calibrated X-ray projection associates a 3D query with a detector location, but does not directly reveal where anatomy lies along the ray. Learned projection-to-volume models combine measurements with anatomical regularities to resolve this ambiguity (Arridge et al., 2019). Coordinates give a decoder direct access to location (Liu et al., 2018), which can be informative in aligned anatomies. As angular coverage increases, however, the measurements provide different spatial evidence. When does the same positional input help, and when does it reduce accuracy?

We find that the same positional input reverses its effect with measurement coverage. With 36 training cases, a prospectively locked replication on 24 held-out LCTSC patients with manual lung contours shows a 30.8-point gain in source one-view Dice and a 6.3-point loss at twelve views. Both signs hold across three initializations. The sign also flips at a fixed view count: two nearby views gain 19.5 points, while two widely separated views lose 3.6. View count alone does not explain the effect (Figure 1); Table 3 reports the absolute scores.

Explaining this behavior requires separating spatial inputs that often occur together. CoordConv adds explicit coordinates, while padding can supply implicit position (Liu et al., 2018; Islam et al., 2020; Kayhan & van Gemert, 2020). Calibrated lifting maps image features to 3D queries (Kyung et al., 2023; Philion & Fidler, 2020; Wang et al., 2022); geometry-aware attention also uses camera relations to structure feature interactions (Miyato et al., 2024). An accuracy gain in a model combining these components cannot, by itself, identify which spatial information was useful. In our comparisons, calibration is retained even when attention, separate geometry conditioning, or positional inputs are removed.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

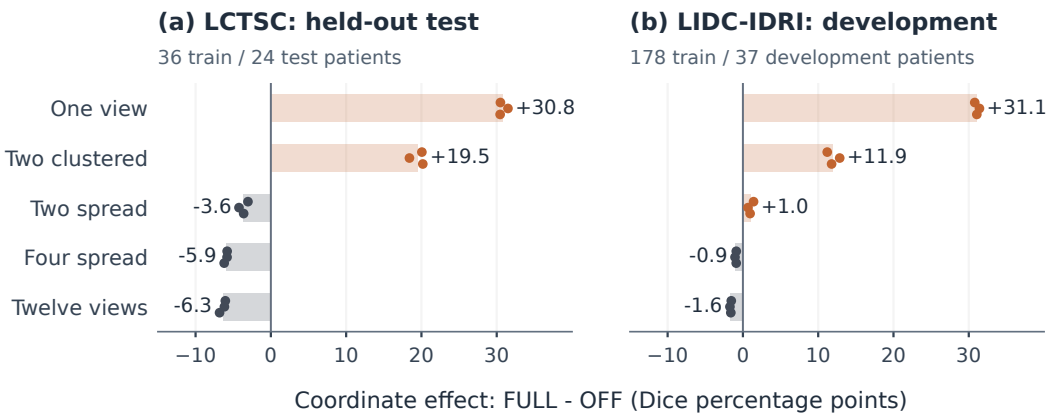


Figure 1: **The coordinate benefit reverses with coverage on held-out and development cohorts.** Source acquisitions on (a) 24 prospectively held-out LCTSC patients and (b) 37 LIDC-IDRI development patients. Bars show mean FULL minus OFF in Dice points; dots are three paired initialization effects, not patients or confidence intervals. The same-count sign flip is specific to LCTSC: two spread views remain slightly positive on LIDC. Absolute scores and patient intervals appear in Table 3; full development results are in Appendix D.

Positional inputs also appear in neural fields and multi-camera 3D perception (Mildenhall et al., 2020; Liu et al., 2022). Their benefit at one operating point need not justify retaining them as measurement coverage changes. We investigate this design question through controlled projection-to-volume experiments: directional inputs test which position is useful, synthetic populations separate alignment from measured content, and decoder-context interventions test whether wider spatial processing reduces positional dependence.

The practical question is whether one model can retain the sparse benefit without paying the broad-coverage cost. We test a scalar coordinate gate derived from acquisition directions (Figure 2). It retains prespecified sparse and broad endpoints in both development cohorts, but fails at intermediate two-view spans. Broader training coverage reduces that deficit without closing it. We report the full acquisition range so that endpoint success does not hide this boundary.

Our contributions are:

- A coverage-dependent reversal confirmed on a locked held-out cohort, including opposite coordinate effects at the same view count.
- Controlled attribution within one pipeline: spatial pathways preserve sparse-view accuracy without learned attention, and along-ray depth consistently outperforms detector-plane position.
- Synthetic and decoder-context interventions that identify population alignment, measured frequencies, and spatial processing as factors that change coordinate benefit.
- An acquisition-dependent gate that retains prespecified sparse and broad endpoints in two cohorts, with intermediate-layout failures partly reduced by broader training coverage.

2 RELATED WORK AND THE QUESTION UNDER STUDY

Explicit and implicit position. CoordConv supplies coordinate channels (Liu et al., 2018), whereas position probes reveal that padding and boundary effects can encode absolute location without them (Islam et al., 2020; Kayhan & van Gemert, 2020). Islam et al. (2021) further show that position information can help or hurt downstream performance. Coordinate representations also underpin neural fields (Mildenhall et al., 2020); Fourier feature mappings change the frequencies an MLP can learn (Tancik et al., 2020). We study an additive coordinate pathway in a lifted feature volume, rather than a coordinate-only scene representation. Removing that pathway does not make the entire CNN position-free.

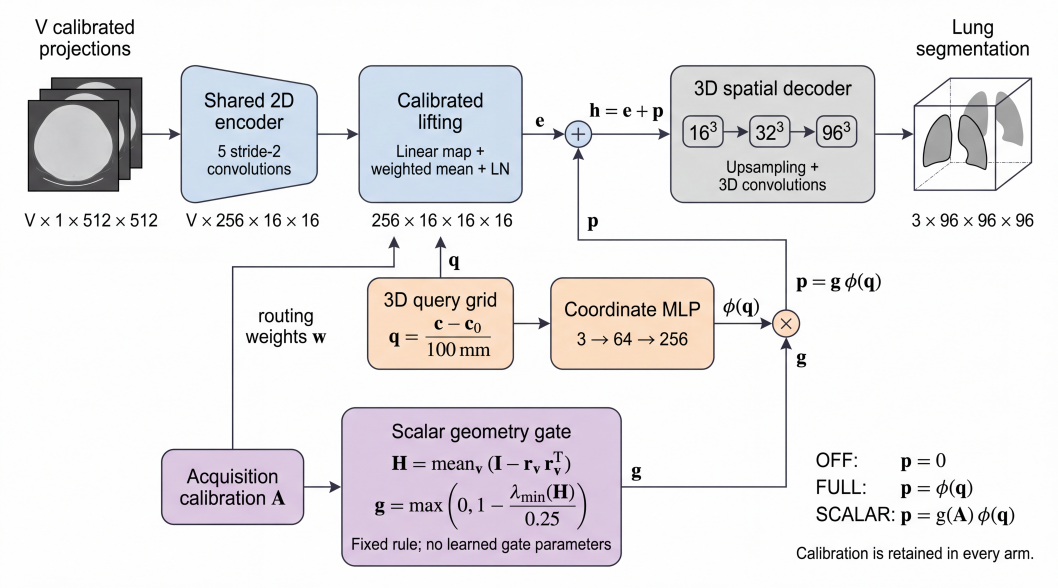


Figure 2: **Shared fixed-lifting architecture and scalar coordinate intervention.** Calibrated projections supply evidence on a 16^3 query grid. A coordinate MLP supplies an additive positional tensor before spatial decoding. SCALAR multiplies the MLP's output, including biases, by one acquisition-wide gate; OFF bypasses that output. Calibration is retained in every arm. Tensor shapes omit the batch dimension; input and output icons are schematic. The earlier learned-slot/attention factorial is a separate architectural comparison.

Calibrated lifting and geometric attention. Lift-Splat-Shoot lifts image features into camera frustums before aggregation (Philion & Fidler, 2020). DETR3D retrieves image features using calibrated 3D queries (Wang et al., 2022), PETR embeds 3D position into image features (Liu et al., 2022), and GTA transforms attention features between camera frames (Miyato et al., 2024). These are distinct uses of geometry. Our controls separate calibrated transport from explicit position and content-dependent aggregation within one pipeline. They do not identify the cause of gains in those other systems.

Learned inverse problems and projection-to-volume prediction. Data-driven inversion combines physical operators with learned priors in several ways (Arridge et al., 2019): a CNN can refine a direct reconstruction (Jin et al., 2017), learned primal-dual updates can incorporate the forward operator (Adler & Öktem, 2018), and RAM uses operator-derived Krylov features (Terris et al., 2026). Network structure itself can also provide an image prior (Ulyanov et al., 2018). For biplanar X-ray prediction, X2CT-GAN learns a volumetric reconstruction (Ying et al., 2019), while PerX2CT uses perspective-projected features (Kyung et al., 2023). Our segmentation ablations ask which spatial inputs a learned decoder uses; they are not a reconstruction benchmark against these methods.

Equivariance and symmetry breaking. Group-equivariant CNNs constrain how features transform (Cohen & Welling, 2016; Cohen et al., 2019). Frame averaging and learned canonicalization offer alternative constructions (Puny et al., 2022; Kaba et al., 2023). An equivariant map cannot break a symmetry of its input (Kaba & Ravanbakhsh, 2023); probabilistic symmetry breaking addresses this restriction (Lawrence et al., 2025). We use the established restriction to design synthetic controls. Our finite cone-beam pipeline does not satisfy the exact assumptions, and we do not claim a new equivariance theorem.

Spatial context and learned cue selection. Dilated convolutions enlarge spatial context (Yu & Koltun, 2016), although effective receptive fields need not fill their theoretical support (Luo et al., 2016). Cue interventions also expose texture bias, shortcuts, and simplicity bias (Geirhos et al., 2019; 2020; Shah et al., 2020). In language models, DroPE studies positional embeddings that aid

training but constrain length generalization (Gelberg et al., 2026); that result concerns a different architecture and shift. We vary coordinate direction, acquisition, population alignment, and decoder context. These interventions test when position is useful without assuming that every observed penalty is a shortcut.

3 SETUP: EVIDENCE, POSITION, AND ASSESSMENT

Let y_v be a projection at calibrated acquisition v, and let $c_j \in \mathbb{R}^3$ be a query on a regular 16^3 grid. An image encoder produces tokens V_u, where u indexes both view and detector token. The fixed lifting baseline computes

$$e_j = \mathrm{LN}\left(\frac{\sum_u w_{ju} V_u}{\sum_u w_{ju}}\right), \qquad h_j = e_j + p_j, \qquad \hat{m} = D_\theta(h). \tag{1}$$

The nonnegative weights w_{ju} are obtained from calibrated projection geometry; unsupported tokens receive zero weight. The denominator is clamped for empty support. D_θ is a spatial decoder producing 96^3 lung/background logits. The positional term is either zero, an MLP of coordinates, or a location-indexed learned embedding. Learned aggregation instead combines a content score with the same geometric support. We distinguish the calibration used for routing from a separate learned geometry-conditioning branch. Turning the latter off does not make the model geometry-free.

Clinical data. We use CT-derived simulated projections from LIDC-IDRI (Armato III et al., 2011), TotalSegmentator (Wasserthal et al., 2023), and LCTSC (Yang et al., 2017; 2018). LIDC has 178 training/37 development patients; TotalSegmentator has 214/34. LIDC lung targets are TotalSegmentator-generated merged lobes, not expert nodule annotations. TotalSegmentator uses its supplied masks. The new LCTSC replication uses manual contours, the published 36 training cases, and the pooled 12 off-site plus 12 live-test cases. Inputs are 24 rendered views on a 7.5° half-circle grid with a common detector scale per case. No measured radiographs are used.

Matched comparisons. Within each experiment, arms share patient ordering, acquisition schedules, optimizer settings, and initialization of shared weights. Single-view directional experiments use six training phases, $0^\circ, 30^\circ, \ldots, 150^\circ$. We separately report those phases, held rotations at 45° and 135°, and 7.5° offsets of the training phases. The mixed-view study gives each training patient equal numbers of one-, two-, four-, eight-, and twelve-view presentations. Clustered views use nearby angles; spread views cover the half circle. We fix the final checkpoint at epoch 250. Details and differences between experiment families are given in Appendix A.

Evidence and uncertainty. The primary metric averages left/right lung Dice within patient, then patients and seeds. LIDC and TotalSegmentator assessments were reused during adaptive development; campaign locks do not make them fresh tests. LCTSC was previously unused in this project: its protocol was fixed prospectively, and all six final models were frozen before test prediction (Appendix H). Seed SD and paired patient intervals measure different uncertainty; angles and seeds do not add independent patients. Synthetic assessment objects are generated independently. Appendix I discloses historical holdout exposure.

4 WHAT CARRIES THE SPARSE-VIEW GAIN?

Two initial factorials separate aggregation, geometry conditioning, and spatial inputs (Appendix A.5). Fixed aggregation without conditioning reaches 0.824 single-view Dice, compared with 0.818 for learned aggregation with conditioning. These single-seed results motivate a replicated test with geometry conditioning off in both arms.

Adding two initializations confirms that fixed aggregation retains sparse performance (Table 1). Three-seed source means are 0.822 versus 0.819 at one view and 0.833 versus 0.831 at two clustered views. Fixed aggregation is within the locked 0.005 retention margin in all four sparse source/offset groups. Both arms retain slots and coordinates. This establishes the sufficiency of fixed aggregation in this pipeline; the full conditioning interaction remains single-seed.

Table 1: **Fixed aggregation retains sparse-view performance across initializations.** Source-acquisition lung Dice on 37 LIDC-IDRI development patients, mean ± seed SD over three initializations. Both arms retain calibrated routing, slots, and coordinates; geometry conditioning is off. Differences are fixed minus learned in Dice points. The full single-seed factorials are in Appendix A.5.

Acquisition	Fixed aggregation	Learned aggregation	Difference
One view	0.8218 ± 0.0024	0.8193 ± 0.0009	$+0.25$
Two clustered	0.8331 ± 0.0022	0.8313 ± 0.0010	$+0.17$
Four spread	0.9000 ± 0.0025	0.8980 ± 0.0007	$+0.20$
Twelve views	0.9081 ± 0.0022	0.9062 ± 0.0003	$+0.19$

Table 2: **Directional contrast across cohorts and rotations.** Ordinary Dice, mean ± sample SD over three initializations; differences are Dice percentage points. Acquisition groups are averaged within case. Cohorts contain 37 and 34 development cases, respectively.

Cohort	Acquisition	Depth	Detector plane	Difference
LIDC-IDRI	Source	0.8229 ± 0.0008	0.7261 ± 0.0282	$+9.68$
LIDC-IDRI	Held rotations	0.8143 ± 0.0020	0.7255 ± 0.0244	$+8.88$
LIDC-IDRI	Offset	0.8172 ± 0.0013	0.7239 ± 0.0279	$+9.34$
TotalSegmentator	Source	0.7168 ± 0.0033	0.6853 ± 0.0119	$+3.15$
TotalSegmentator	Held rotations	0.7152 ± 0.0055	0.6814 ± 0.0158	$+3.38$
TotalSegmentator	Offset	0.7130 ± 0.0020	0.6827 ± 0.0132	$+3.02$

In the spatial-pathway factorial, coordinates alone and slots alone preserve most single-view performance: 0.817 and 0.821. Removing both drops Dice to 0.480, while improving four-spread Dice from 0.899 to 0.913. Either pathway is sufficient for most of the sparse-view gain. Slot removal also removes 1,048,576 active parameters, so spatial information and capacity are not completely separated. We next vary the *direction* of position in a shared architecture.

5 Which position helps, and under what conditions?

5.1 Along-ray depth is more useful than detector-plane position

For a source s, optical axis d, detector axes u, v, and box center c_0, define $w(c) = (c - s)^\top d$ and $w_0 = w(c_0)$. Our acquisition-frame coordinates are

$$q_A(c) = \left[w(c) - w_0, \quad \frac{w_0 (c - s)^\top u}{w(c)} - (c_0 - s)^\top u, \quad \frac{w_0 (c - s)^\top v}{w(c)} - (c_0 - s)^\top v \right]. \quad (2)$$

The first entry varies along a ray; the last two are centered detector-plane coordinates and remain constant along that actual divergent ray. This projective frame avoids treating a cone beam as parallel. The depth arm retains the first entry and zeros the others; the plane arm does the converse. Each active descriptor receives the same geometry-only total-RMS normalization, then the same $3 \to 64 \to 256$ MLP. No target or image statistic enters coordinate normalization.

Depth outperforms detector-plane position in both cohorts and all three initializations (Table 2). On LIDC-IDRI, the source-acquisition difference is 9.68 Dice points; under held rotations it is 8.88 points. At those rotations, depth wins on 36/37, 37/37, and 36/37 patients across the three seeds. The independent-cohort source difference is 3.15 points. The direction of the effect replicates more consistently than its magnitude: the detector-plane model is more sensitive to initialization.

Full/no-coordinate arms were not repeated in this directional replication. Equal registered parameter counts do not ensure equal effective capacity when one versus two input dimensions are active.

Table 3: **The coverage-dependent reversal replicates on held-out LCTSC patients.** Ordinary lung Dice, mean ± seed SD over three initializations; 36 training and 24 test patients. Differences are FULL minus OFF in Dice points, with institution-stratified paired-patient 95% intervals conditional on the fitted seeds. All models were frozen before test prediction.

Source acquisition	OFF	FULL	Difference	95% interval
One view	0.3637 ± 0.0046	0.6719 ± 0.0018	$+30.81$	$[+28.96, +32.38]$
Two clustered	0.4837 ± 0.0124	0.6792 ± 0.0025	$+19.55$	$[+17.88, +21.07]$
Two spread	0.7605 ± 0.0048	0.7242 ± 0.0066	-3.63	$[-5.15, -2.13]$
Four spread	0.7940 ± 0.0029	0.7346 ± 0.0049	-5.94	$[-7.64, -4.31]$
Twelve views	0.8033 ± 0.0015	0.7398 ± 0.0056	-6.34	$[-7.86, -4.90]$

5.2 Coverage changes the sign of the coordinate benefit

In a fresh mixed-view comparison, we train the same fixed-lifting architecture with its coordinate pathway off or on. The three-seed result in Figure 1 reproduces the sparse/broad contrast: the coordinate benefit is 31.09 Dice points with one view and 11.95 points with two clustered views, but -0.94 points with four spread views and -1.62 points with twelve views. Two spread views differ substantially from two clustered views despite having the same count.

Locked held-out replication. Six scratch-trained LCTSC models reproduce the reversal on 24 published test patients (Table 3). FULL minus OFF is +30.81 Dice points at one view (paired-patient 95% interval [28.96, 32.38]) and -6.34 at twelve views [-7.86, -4.90], with both signs retained in every seed. The one-minus-twelve interaction is +37.16 points [35.64, 38.67]. Two clustered views benefit while two spread views lose accuracy, despite equal count. Rotated and offset conditions retain the pattern (Appendix H).

The penalty includes poorer training fit. At twelve views, FULL also fits the LCTSC training cases worse than OFF (0.8381 versus 0.8585 Dice; Appendix H). The penalty therefore includes a fitting difference, not only a held-out generalization gap. This is consistent with training interactions, although final fitting scores cannot isolate an optimization mechanism. We next test alignment and decoder processing.

5.3 Alignment and measured frequencies constrain the effect

We use two synthetic inverse problems to separate acquisition ambiguity from the population's spatial distribution. The same 128^2 complex-valued object generator supplies canonically aligned objects and uniformly circularly translated objects. A residual CNN receives a lifted measurement and either no coordinates, coordinates along the ambiguous axis, or coordinates along the orthogonal axis. Local and wider-context models have the same parameter count; dilation changes their receptive fields. Across both operators, the campaign contains 264 fixed-budget fits, with three-seed replication on the prespecified key conditions. Each condition uses 512 fitting, 128 validation, and 256 independent assessment objects.

For Cartesian Fourier sampling, the baseline mask retains every fourth line, including DC. We replace selected lines with central or noncentral lines while keeping 32 sampled lines. For parallel-beam projection, two measurements have a separation of $0°$, $5°$, $15°$, $45°$, or $90°$; the $0°$ case duplicates one view. Figure 3 displays the replicated conditions.

With exact-comb Fourier sampling, the ambiguous-axis coordinate reduces NMSE by approximately 0.40 for aligned objects under both decoder contexts. The gain is absent for uniformly translated objects. Replacing eight lines by noncentral lines preserves a gain of approximately 0.40 even though exact repetition is broken. Replacing eight by central lines instead produces small, mixed gains: from -0.0022 to $+0.0146$ across seeds and contexts. Thus, merely breaking an exact symmetry is insufficient to predict useful positional information; the content supplied by the changed measurements matters.

Parallel-beam results also depend on alignment and context. At a $15°$ separation, the aligned coordinate benefit is 0.304–0.322 NMSE with local context and 0.087–0.128 with wider context.

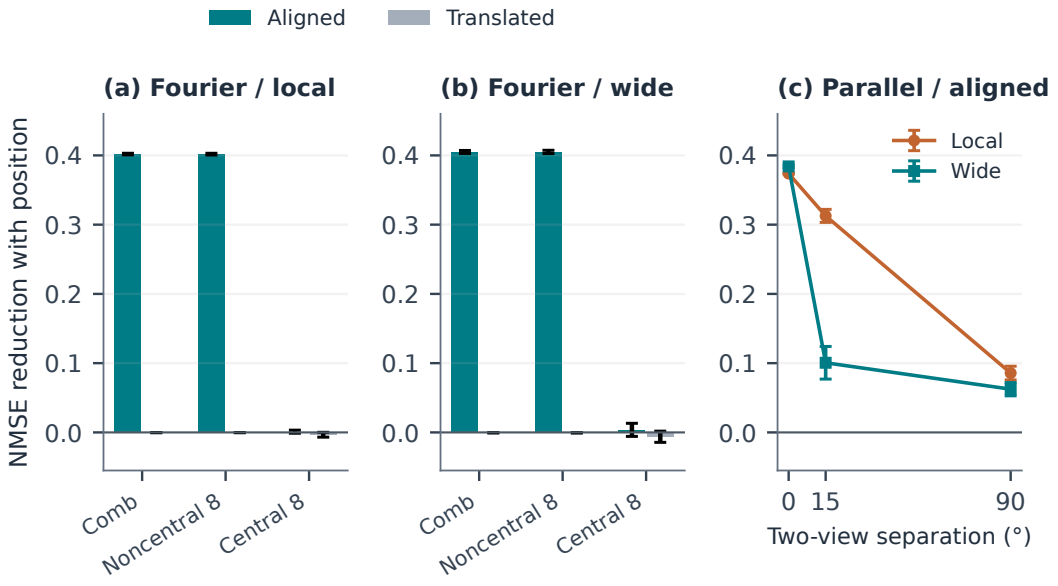


Figure 3: **Position helps under some ambiguities, populations, and decoders.** Values are NMSE without coordinates minus NMSE with ambiguous-axis coordinates; higher is better. Fourier panels keep the number of measured lines fixed. Central and noncentral replacements have different effects despite both breaking exact-comb repetition. The parallel-beam panel shows aligned objects. Bars indicate sample SD across three model initializations, not object confidence intervals.

Translated-object effects are much smaller and often negative. These results reject a rule based only on continuous versus discrete ambiguity, or only on view count.

5.4 Wider spatial processing reduces positional dependence

We next change two decoder convolutions from dilation $(1, 1)$ to $(2, 3)$, with matching padding and no change in parameter count. The coordinate branch remains unchanged. This four-arm context-by-coordinate comparison uses one matched initialization and the same clinical schedule.

At one view, expanded context raises coordinate-free Dice from 0.489 to 0.639, while raising coordinate-enabled Dice only from 0.813 to 0.824 (Table 4). The coordinate advantage decreases by 13.83 points, with a paired-patient bootstrap interval of [12.84, 14.84]: a 42.8% reduction of the local decoder's positional advantage. The coordinate-free expanded model still trails the local coordinate model by 17.36 points. At twelve views, wider context does not remove the coordinate penalty.

The result identifies spatial processing as a contributor to positional dependence, without isolating a single receptive-field mechanism: padding and GroupNorm remain. A separate three-seed coordinate-free comparison supports the context gain. Acquisition-adaptive isotropic context raises source one-view Dice by 14.6 points, while twelve-view Dice is 0.930 versus 0.929. Stretching context along the weak acquisition direction performs worse than isotropic context. This replicates the context gain; the original coordinate-by-context factorial remains single-seed (Appendix F).

6 A bounded intervention: acquisition-dependent coordinate strength

The preceding experiments motivate conditioning position on the acquisition rather than supplying it at fixed strength. We test a deliberately small intervention without modifying the encoder, lifting,

or decoder. For source positions s_v and box center c_0, define

$$r_v = \frac{c_0 - s_v}{\|c_0 - s_v\|_2}, \qquad H(A) = \frac{1}{V}\sum_{v=1}^{V}(I - r_v r_v^\top), \qquad g(A) = \max\left\{0, 1 - \frac{\lambda_{\min}(H(A))}{0.25}\right\}. \tag{3}$$

The scalar arm uses $p_j = g(A)\phi(q_j)$, where $q_j = (c_j - c_0)/(100\,\mathrm{mm})$ is the existing world-coordinate input. The cutoff 0.25 was fixed before the twelve-run campaign. It is a chosen hyperparameter; "parameter-free" here means that the gate has no learned weights. When $g = 0$, the entire positional output, including MLP biases, is zero. H is a central-ray directional-diversity descriptor, not the normal operator of the volumetric inverse problem or an exact null-space estimator.

We compare OFF, FULL, SCALAR, and DIRECTIONAL across three LIDC initializations. DIRECTIONAL projects q_j onto the weakest eigenvector of H, restores its total RMS, and applies the same gate and MLP. This uses one acquisition-wide direction, unlike Equation 2. We then train OFF/FULL/SCALAR from scratch on TotalSegmentator, with three initializations and the unchanged architecture and cutoff.

Prespecified endpoints improve. SCALAR passes the locked sparse/broad endpoint screen (Appendix A), including retention within 0.005 of the corresponding reference. Its average broad gain over FULL is 1.63 Dice points, with a paired-patient interval of [1.08, 2.34]. DIRECTIONAL fails sparse retention and does not demonstrate an additional directional benefit.

Second-cohort replication. On TotalSegmentator, SCALAR also passes all relative endpoint criteria; the LIDC-specific absolute thresholds are not applied. Source one-view Dice is 0.741 versus FULL's 0.719; twelve-view Dice rounds to 0.874 for both SCALAR and OFF (Figure 4). The average broad gain over FULL is 4.07 Dice points, with a paired-case interval of [2.87, 5.52], and is positive in every seed. This is separate cohort-specific training, not zero-shot transfer; all 34 assessment cases are reused development cases.

Intermediate spans expose a training-coverage limitation. For the held $0°/60°$ pair on LIDC, SCALAR reaches 0.761 against FULL's 0.853. TotalSegmentator also fails all-layout retention: its largest loss is 6.55 points at $7.5°/67.5°$. Both cohorts lose at other two-view spans (Figure 4). These failures combine coordinate removal with acquisition-distribution shift: two-view training contains only $15°$ and $90°$ separations. They cannot identify a gate-only mechanism.

Retraining both arms across three seeds with five two-view separations, withholding $52.5°$–$67.5°$, improves source-$60°$ SCALAR Dice by 3.83 points to 0.800; matched FULL reaches 0.859. Budget and per-patient positive/zero-gate counts remain fixed. The offset gain is 4.05 points. Training exposure explains part of the deficit, but the remedy screen still fails (Appendix F.2).

At broad coverage, SCALAR's positional tensor is zero, yet its predictions differ from OFF because their training differs. Zeroing one branch does not equate separately trained models or restore exact equivariance. The result supports a useful endpoint baseline, not the necessity of exact zeros or a common cause for all failed remedies. Full layouts and subsequent context/fusion controls are retained in the appendix.

7 DISCUSSION AND LIMITATIONS

Calibration, explicit position, and decoder processing serve different roles in this pipeline. Alignment can make location predictive; measured content and spatial context can reduce its value. Neither alignment nor geometry alone explains all outcomes. The symmetry result in Appendix B is a limiting case, not a complete clinical mechanism or an explanation of other architectures' gains.

The LCTSC test confirms the OFF/FULL coverage interaction on a previously unused cohort with manual contours. It does not validate the gate, directional contrast, or context intervention there; those findings still rely on reused development cases. LCTSC training fits also differ between arms, and the smaller cohort receives fewer updates at the same epoch budget. Absolute accuracy remains limited. Projections are simulated, LIDC targets are automated, and all tasks are coarse lung segmentation. The full conditioning/slot and original context-by-coordinate factorials remain single-seed. No clinical-use or reconstruction-performance claim follows.

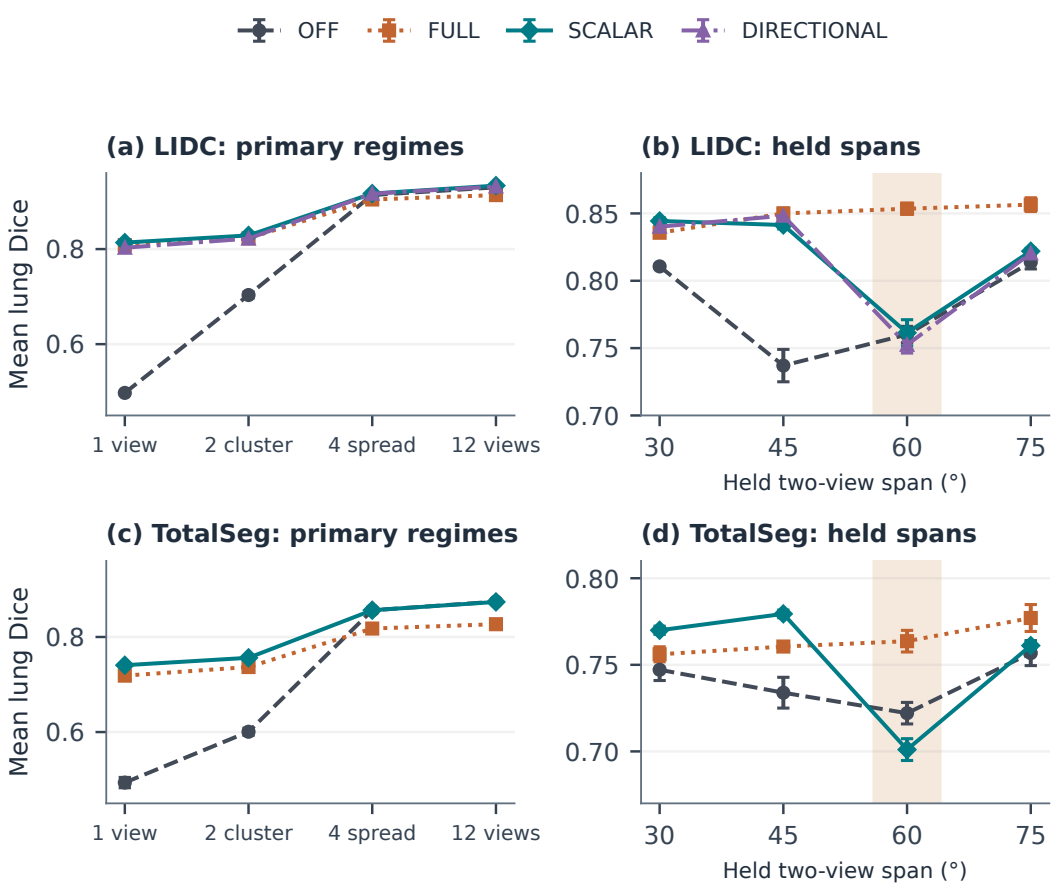


Figure 4: **Endpoint gains and intermediate-span failures replicate across cohorts.** Three-seed development means with seed SD; LIDC-IDRI has 37 patients and TotalSegmentator 34 cases. Models are trained separately per cohort and used across acquisitions. (a,c) Primary source regimes. (b,d) Held $0°/\theta$ pairs expose the failure around 60°. All offsets and other layouts remain in the appendix; DIRECTIONAL was tested only on LIDC.

In practice, coordinate ablations should vary both assessment geometry and training coverage. The scalar baseline retains favorable endpoints, but coverage and fusion interventions leave an intermediate-span gap. Endpoint success alone does not establish a solved tradeoff.

8 CONCLUSION

Coordinate utility depends on acquisition, alignment, and spatial processing. A locked third-cohort test confirms that coordinates help sparse prediction but hurt at broad coverage. Fixed aggregation retains sparse performance; depth outperforms detector-plane position in the development cohorts. Scalar gating preserves endpoints there, but broader training only partly repairs its intermediate-angle deficit. Calibrated transport, position, and decoder context warrant separate controls.

REPRODUCIBILITY STATEMENT

Sections 3–6 specify the model and interventions. Appendix A gives preprocessing, architecture, optimization, acquisition schedules, and statistical units; Appendix B states the assumptions and proof of the symmetry observation. Appendices D–G retain full acquisition results, follow-up protocols, and failed screens. Appendix H documents the prospectively locked LCTSC split, preparation amendments, checkpoint rule, and bootstrap. Appendix I distinguishes reused development data from the new test assessment. Numerical tables are generated from saved measurements and checked against experiment summaries. Source/configuration locks, paired initialization checks, checkpoint hashes, and prediction-reload checks are retained in the experiment record. The accompanying manuscript-source archive rebuilds the paper and figures; it is not a release of the training implementation, checkpoints, or patient data.

AI USE STATEMENT

We used OpenAI ChatGPT/Codex assistants for literature discovery and synthesis, proposing and critiquing hypotheses, developing conceptual explanations and mathematical arguments, designing experiments, implementing and debugging methods, checking protocols and numerical summaries, interpreting results, and drafting and revising manuscript text and mathematical exposition. Assistance also covered preprocessing and code for procedural synthetic objects and CT-derived projection simulation. These datasets were computed by the specified generators and physical rendering code, rather than supplied as image-generator outputs. Gemini Nano Banana Pro generated the architecture schematic from a supplied specification; its icons are illustrative, not patient images or model predictions. Quantitative plots and tables are computed from saved experimental measurements. Verification procedures and the limits of the resulting evidence are described in the appendices; automated checks do not replace author review. The authors retain responsibility for the final text, claims, code, and artifacts, including AI-assisted material.

REFERENCES

Jonas Adler and Ozan Öktem. Learned primal-dual reconstruction. *IEEE Transactions on Medical Imaging*, 2018. doi: 10.1109/TMI.2018.2799231.

Samuel G. Armato III, Geoffrey McLennan, Luc Bidaut, Michael F. McNitt-Gray, Charles R. Meyer, Anthony P. Reeves, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans. *Medical Physics*, 38(2):915–931, 2011. doi: 10.1118/1.3528204.

Simon Arridge, Peter Maass, Ozan Öktem, and Carola-Bibiane Schönlieb. Solving inverse problems using data-driven models. *Acta Numerica*, 28:1–174, 2019. doi: 10.1017/S0962492919000059.

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016. URL https://arxiv.org/abs/1607.06450.

Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, Lawrence Tarbox, and Fred Prior. The cancer imaging archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging*, 26(6):1045–1057, 2013. doi: 10.1007/s10278-013-9622-7. URL https://doi.org/10.1007/s10278-013-9622-7.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 2990–2999, 2016. URL https://proceedings.mlr.press/v48/cohenc16.html.

Taco S. Cohen, Mario Geiger, and Maurice Weiler. A general theory of equivariant CNNs on homogeneous spaces. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://arxiv.org/abs/1811.02017.

Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, and Wieland Brendel. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bygh9j09KX.

Robert Geirhos, Jörn-Henrik Jacobsen, Claudio Michaelis, Richard Zemel, Wieland Brendel, Matthias Bethge, and Felix A. Wichmann. Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2020. doi: 10.1038/s42256-020-00257-z.

Yoav Gelberg, Koshi Eguchi, Takuya Akiba, and Edoardo Cetin. Extending the context of pretrained LLMs by dropping their positional embedding. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/62133b95efaa6f10a5395eb10ca087cd-Abstract-Conference.html.

Md Amirul Islam, Sen Jia, and Neil D. B. Bruce. How much position information do convolutional neural networks encode? In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=rJeB36NKvB.

Md Amirul Islam, Matthew Kowal, Sen Jia, Konstantinos G. Derpanis, and Neil D. B. Bruce. Position, padding and predictions: A deeper look at position information in CNNs. *arXiv preprint arXiv:2101.12322*, 2021. URL https://arxiv.org/abs/2101.12322.

Kyong Hwan Jin, Michael T. McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9): 4509–4522, 2017. doi: 10.1109/TIP.2017.2713099.

Sékou-Oumar Kaba and Siamak Ravanbakhsh. Symmetry breaking and equivariant neural networks. *arXiv preprint arXiv:2312.09016*, 2023. URL https://arxiv.org/abs/2312.09016.

Sékou-Oumar Kaba, Arnab Kumar Mondal, Yan Zhang, Yoshua Bengio, and Siamak Ravanbakhsh. Equivariance with learned canonicalization functions. In *International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 15546–15566, 2023. URL https://proceedings.mlr.press/v202/kaba23a.html.

Osman Semih Kayhan and Jan C. van Gemert. On translation invariance in CNNs: Convolutional layers can exploit absolute spatial location. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. URL https://arxiv.org/abs/2003.07064.

Daeun Kyung, Kyungmin Jo, Jaegul Choo, Joonseok Lee, and Edward Choi. Perspective projection-based 3D CT reconstruction from biplanar X-rays. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1–5, 2023. URL https://arxiv.org/abs/2303.05297.

Hannah Lawrence, Vasco Portilheiro, Yan Zhang, and Sékou-Oumar Kaba. Improving equivariant networks with probabilistic symmetry breaking. In *International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2503.21985.

Rosanne Liu, Joel Lehman, Piero Molino, Felipe Petroski Such, Eric Frank, Alex Sergeev, and Jason Yosinski. An intriguing failing of convolutional neural networks and the CoordConv solution. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://arxiv.org/abs/1807.03247.

Yingfei Liu, Tiancai Wang, Xiangyu Zhang, and Jian Sun. PETR: Position embedding transformation for multi-view 3D object detection. In *European Conference on Computer Vision*, 2022. URL https://arxiv.org/abs/2203.05625.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://arxiv.org/abs/1711.05101.

Wenjie Luo, Yujia Li, Raquel Urtasun, and Richard Zemel. Understanding the effective receptive field in deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, volume 29, 2016. URL https://arxiv.org/abs/1701.04128.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision*, 2020. URL https://arxiv.org/abs/2003.08934.

Takeru Miyato, Bernhard Jaeger, Max Welling, and Andreas Geiger. GTA: A geometry-aware attention mechanism for multi-view transformers. In *International Conference on Learning Representations*, 2024. URL https://arxiv.org/abs/2310.10375.

Jonah Philion and Sanja Fidler. Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3D. In *European Conference on Computer Vision*, 2020. URL https://arxiv.org/abs/2008.05711.

Omri Puny, Matan Atzmon, Heli Ben-Hamu, Ishan Misra, Aditya Grover, Edward J. Smith, and Yaron Lipman. Frame averaging for invariant and equivariant network design. In *International Conference on Learning Representations*, 2022. URL https://arxiv.org/abs/2110.03336.

Harshay Shah, Kaustav Tamuly, Aditi Raghunathan, Prateek Jain, and Praneeth Netrapalli. The pitfalls of simplicity bias in neural networks. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://arxiv.org/abs/2006.07710.

Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://arxiv.org/abs/2006.10739.

Matthieu Terris, Samuel Hurault, Maxime Song, and Julián Tachella. Reconstruct anything model: A lightweight general model for computational imaging. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=Ks9zNS6OsU.

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9446–9454, 2018. URL https://arxiv.org/abs/1711.10925.

Yue Wang, Vitor Campagnolo Guizilini, Tianyuan Zhang, Yilun Wang, Hang Zhao, and Justin Solomon. DETR3D: 3D object detection from multi-view images via 3D-to-2D queries. In *Proceedings of the 5th Conference on Robot Learning*, volume 164 of *Proceedings of Machine Learning Research*, pp. 180–191, 2022. URL https://proceedings.mlr.press/v164/wang22b.html.

Jakob Wasserthal, Hanns-Christian Breit, Manfred T. Meyer, Maurice Pradella, Daniel Hinck, Alexander W. Sauter, Tobias Heye, Daniel T. Boll, Joshy Cyriac, Shan Yang, Michael Bach, and Martin Segeroth. TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images. *Radiology: Artificial Intelligence*, 5(5):e230024, 2023. doi: 10.1148/ryai.230024.

Yuxin Wu and Kaiming He. Group normalization. In *European Conference on Computer Vision*, pp. 3–19, 2018. URL https://arxiv.org/abs/1803.08494.

J. Yang, G. Sharp, H. Veeraraghavan, W. van Elmpt, A. Dekker, T. Lustberg, and M. Gooding. Data from Lung CT Segmentation Challenge (LCTSC) (version 3), 2017. URL https://doi.org/10.7937/K9/TCIA.2017.3R3FVZ08. Data set.

Jinzhong Yang, Harini Veeraraghavan, Samuel G. Armato III, Keyvan Farahani, Justin S. Kirby, Jayashree Kalpathy-Cramer, Wouter van Elmpt, Andre Dekker, et al. Autosegmentation for thoracic radiation treatment planning: A grand challenge at AAPM 2017. *Medical Physics*, 45 (10):4568–4581, 2018. doi: 10.1002/mp.13141. URL https://doi.org/10.1002/mp.13141.

Xingde Ying, Heng Guo, Kai Ma, Jian Wu, Zhengxin Weng, and Yefeng Zheng. X2CT-GAN: Reconstructing CT from biplanar X-rays with generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. URL https://arxiv.org/abs/1905.06902.

Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. In *International Conference on Learning Representations*, 2016. URL https://arxiv.org/abs/1511.07122.

A PROTOCOLS, ARCHITECTURE, AND STATISTICAL UNITS

A.1 CLINICAL IMAGE FORMATION AND TARGETS

The clinical experiments operate on rendered projections rather than measured radiographs. Each case supplies a CT volume, physical image geometry, and a 96^3 target in the patient box. A single detector scale is used across the 24 angles of a case, so acquisition rotation is not confounded with independent per-view resizing. The source/detector geometry is retained by all lifting arms. Projection images are 512^2; the encoder uses $\log(1 + y)$ followed by per-view standardization. These choices fix the appearance model for the reported comparisons, but do not reproduce clinical scatter, beam hardening, or scanner-specific noise.

LIDC-IDRI supplies the CT cohort, not expert whole-lung targets for this study. Its lung lobes are segmented by TotalSegmentator and merged into left and right lungs. TotalSegmentator-cohort experiments use the supplied anatomical labels. In the lung comparisons, heart and other non-lung structures are supervised as background. Ordinary left/right lung Dice is the primary reported metric; a heart-excluding evaluation is a secondary diagnostic and is not substituted for it in the main tables.

A.2 SHARED ARCHITECTURE

The image encoder contains five stride-two 3×3 convolutions with widths 32, 64, 128, 192, and 256, GroupNorm (Wu & He, 2018), and SiLU, followed by a 1×1 convolution. Thus a 512^2 projection produces 16^2 tokens with 256 channels. A learned linear projection precedes normalized calibrated lifting to 16^3 locations. LayerNorm (Ba et al., 2016) acts on the lifted feature channels.

The 3D decoder first projects 256 channels to 128. Trilinear upsampling to 32^3 is followed by two 3^3 convolutions with 64 channels, GroupNorm, and SiLU. A second trilinear upsampling to 96^3 precedes a 32-channel 3^3 convolution, normalization, activation, and a 1^3 three-class head. Interpolation uses `align_corners=False`. The coordinate MLP is Linear$(3, 64)$–GELU–Linear$(64, 256)$, with 16,896 parameters. The directional and mixed-view fixed-lifting experiments register 1,327,555 parameters. Disabled coordinates are registered but frozen and bypassed, leaving 1,310,659 active parameters.

The initial attribution factorial is implemented in the original slot architecture and has different active counts, listed in Table 5. Its results must not be pooled as additional seeds of the later fixed-lifting coordinate model. Both slot-table and coordinate removal are explicit interventions; all other shared weights are initialized identically within the corresponding experiment.

A.3 TRAINING AND ACQUISITIONS

Clinical arms use FP32, AdamW (Loshchilov & Hutter, 2019) with learning rate 3×10^{-4} and weight decay 10^{-4}, gradient clipping at 1, cross-entropy plus soft lung Dice, and gradient accumulation over four patient presentations. The final partial accumulation at each epoch is retained. The inherited routing schedule uses temperature 4 to 1 over the fixed presentation horizon. Primary directional and scalar comparisons use epoch 250, not a retrospectively selected development checkpoint.

LIDC directional runs contain 44,500 patient presentations and 11,250 optimizer updates; TotalSegmentator directional runs contain 53,500 presentations and 13,500 updates. Each case appears at one of six phases according to its sorted index and epoch, giving 41 or 42 exposures per phase. The two non-training rotation phases are 45° and 135°. "Held rotation" means excluded from that run's training schedule, not an angle never inspected during the research program.

In the scalar comparison, each training case in either cohort has exactly 50 presentations at each of 1, 2, 4, 8, and 12 views. For intermediate counts, clustered and spread layouts are balanced. Source layouts and 7.5° offsets are assessed separately. The assessment comprises 42 unique acquisitions; duplicate descriptions of the one-view and twelve-view layouts do not add independent observations. TotalSegmentator uses 214 training cases, 53,500 presentations and 13,500 updates per model; all nine models train from scratch. The architecture, cutoff 0.25, loss and optimizer are unchanged. Fourteen held acquisition sets are absent from its training schedule. The screen and final-epoch rule were locked before the replication. The complete results are in Appendices D and E.

The LIDC scalar endpoint screen requires source one-view Dice at least 0.80, twelve-view Dice at least 0.925, every primary sparse condition within 0.005 of FULL, and every primary broad condition within 0.005 of OFF. The average broad gain over FULL must exceed 0.010 and be positive in every seed. Sparse conditions are one view and two clustered views; broad conditions are four spread and twelve views, each at source and offset acquisitions. TotalSegmentator uses the relative criteria without transferring the LIDC-specific absolute thresholds.

The attention replication adds fixed/learned, geometry-off models at seeds 1 and 2 and reuses the original seed-zero checkpoints after exact score replay. Every run uses the original factorial's fixed patient/acquisition schedule, 178 training patients, 44,500 presentations, and 11,250 updates. This schedule differs from the later balanced-view scalar protocol. Shared weights are initialized identically within each seed; inactive query/key parameters in fixed aggregation leave unequal active counts. All 28 original layouts are assessed. The practical screen requires fixed-minus-learned mean Dice of at least -0.005 in each primary sparse source/offset group. It is a retention screen, not a formal equivalence test.

The context intervention changes the dilations of the two decoder convolutions at 32^3 from $(1, 1)$ to $(2, 3)$ with matching padding. Initial values and parameter counts are unchanged. This intervention changes spatial processing without claiming that the entire network has a strictly local receptive field: GroupNorm and finite boundaries remain.

Table 4: **Context changes the value of coordinates.** LIDC-IDRI development Dice, one matched initialization. The interaction is the local coordinate advantage minus the expanded-context coordinate advantage, in percentage points.

	Local context		Expanded context		
Acquisition	Off	On	Off	On	Interaction
One view	0.4893	0.8127	0.6391	0.8242	+13.83
Two clustered	0.6950	0.8275	0.7613	0.8390	+5.48
Two spread	0.8612	0.8752	0.8857	0.8860	+1.36
Twelve views	0.9247	0.9194	0.9270	0.9211	+0.06

A.4 Aggregation, variability, and reproducibility

For each clinical case, we first average the left/right Dice values, then the distinct angles belonging to a condition. Means then give equal weight to cases and to model seeds. Sample SD over three initialization-level means quantifies seed variation; it is not a patient confidence interval. For a paired aggregate interval, we average each case's paired difference over seeds, then bootstrap the same case identities. The scalar comparison uses 10,000 resamples. Such intervals describe sampling variability on the reused cohort; they do not correct adaptive experiment selection.

Numerical tables are checked against saved summaries; directional, context, and gate tables are generated directly from their machine-readable records. The experiment records preserve source/configuration locks, patient schedules, initialization checks, final-checkpoint hashes, and prediction-reload checks. Implementations were qualified with geometry, gradient, disabled-path, and reload tests before the corresponding runs. These checks establish implementation consistency, not scientific generality. We do not report a count of evaluation records as a sample size.

Figure preparation. Figure 2 is a schematic generated with Gemini Nano Banana Pro from an author-specified architecture description and checked against the implementation. Its image icons are illustrations, not clinical samples or model predictions. All quantitative plots are generated programmatically from saved numerical summaries.

A.5 Full attribution factorials

Table 5 preserves the two original one-seed factorials. The fixed/learned, geometry-off rows subsequently receive two additional initializations, reported in Table 1. Neither that replication nor its small mean differences extend to the unreplicated conditioning interaction or the slot-removal comparison.

Table 5: **Single-seed component attribution.** Ordinary lung Dice on 37 LIDC-IDRI development patients. Fixed lifting retains calibration. The slot/coordinate factorial uses fixed aggregation and conditioning off. Registered but inactive parameters are excluded from active counts. The repeated fixed-off/both-on row is the same checkpoint, not an independent run.

Aggregation	Conditioning	One view	Two clustered	Four spread	Twelve
Fixed	Off	.8239	.8349	.8991	.9073
Fixed	On	.8214	.8324	.8994	.9135
Learned	Off	.8201	.8324	.8985	.9066
Learned	On	.8177	.8311	.8987	.9100
Spatial slots / coordinates	Active parameters				
Both on	2,441,923	.8239	.8349	.8991	.9073
Coordinates only	1,393,347	.8170	.8300	.9000	.9084
Slots only	2,425,027	.8207	.8304	.9011	.9088
Both off	1,376,451	.4799	.6811	.9134	.9235

B A SYMMETRY OBSERVATION AND ITS LIMITS

The following is a standard consequence of equivariance, included to state the assumptions behind the synthetic diagnostic. It is not a new theorem about cone-beam reconstruction (Kaba & Ravanbakhsh, 2023; Lawrence et al., 2025).

Observation 1. *Let T act on an input representation and its output, and let f satisfy $f(Tz) = Tf(z)$. If $Tz = z$, then $Tf(z) = f(z)$. More generally, for a finite group $\mathcal{H}$ acting unitarily on output space, let $P_{\mathcal{H}} = |\mathcal{H}|^{-1} \sum_{T \in \mathcal{H}} T$. If every $T \in \mathcal{H}$ fixes z and f is equivariant to $\mathcal{H}$, then for any target x,*

$$\|f(z) - x\|_2^2 \;\geq\; \|(I - P_{\mathcal{H}})x\|_2^2. \tag{4}$$

Proof. The first statement follows by substitution. The group average $P_{\mathcal{H}}$ is the orthogonal projector onto the invariant subspace. Since $f(z)$ belongs to that subspace, its residual against $(I - P_{\mathcal{H}})x$ is orthogonal to $f(z) - P_{\mathcal{H}}x$. The squared-error decomposition gives the bound. □

For a 128-pixel periodic signal with every fourth Fourier line retained, the zero-filled reconstruction repeats under translations by 32 pixels along the undersampled axis. The circularly padded, stride-one coordinate-free synthetic decoder preserves those translations. An absolute phase coordinate can remove that architectural restriction. Whether learning benefits from the additional input still depends on the data distribution: the coordinate is predictive for a canonically aligned population but need not select the correct translate for a uniformly translated one.

For an ideal single parallel-beam view, backprojected detector features are constant along the ray, away from finite-support effects. Real cone-beam rays are not lattice translations, and clinical decoders use padding and normalization. We therefore do not apply the exact bound to their Dice scores. Nor is $\lambda_{\min}(H)$ in Equation 3 the dimension of an inverse problem's null space. Exact equivariance is a mathematical property; the size of a learned coordinate effect is an empirical quantity, not a binary consequence of that property.

C SYNTHETIC INTERVENTIONS

The generator produces complex-valued 128^2 objects. Fitting, validation, and assessment use independent generator seeds, with 512, 128, and 256 objects, respectively. The translated condition applies a uniform circular translation before rendering; it is a matched population change, not an unseen translation applied only at assessment. Cartesian Fourier experiments balance the two sampling axes in training and average them within assessment object. Two coordinate channels are registered in every arm; inactive channels are zero. Active coordinates use unit total RMS.

The synthetic network has 24 channels, four residual blocks with two convolutions each, and circular padding. It has no normalization, pooling, or strided layers. Local dilation gives a nominal receptive field of 21 pixels; block dilations 1, 2, 4, and 8 give 65 pixels. Both contexts register

42,986 parameters. Every run uses 5,000 Adam updates at learning rate 10^{-3}, batch size 12, and complex squared error, with paired initialization and object order. Assessment uses the locked final checkpoint without extension.

Fourier masks. Each mask has 32 lines and includes DC. The comb uses multiples of four. Central variants replace 2, 8, or 16 lines; the noncentral control replaces 8 with lines at large frequency magnitude. Central-8 and noncentral-8 remove the same original lines. This separates the number of samples from their content, though it does not hold information content or reconstruction difficulty constant. That distinction is the intervention's purpose.

Parallel beam. We use pixel-driven linear splatting and its matched discrete adjoint, with 192 detector bins at spacing $2/128$. The lifted input is normalized as $A^\top y/(A^\top A\mathbf{1})$ on valid support. The two measurements have angles 0° and θ; $\theta = 0$ is an exact duplicate measurement. Real and imaginary channels are treated separately. Interpolation makes this a discrete approximation to parallel-beam tomography, not an exact continuous model.

Replication and boundaries. The first initialization covers all 120 combinations of operator, sampling condition, alignment, context, and coordinate input. Two further initializations replicate Fourier comb, central-8, and noncentral-8, and parallel separations 0°, 15°, and 90°, adding 144 fits. These follow-up conditions were specified before that campaign. The larger Fourier benefit under noncentral replacement is important: loss of exact repetition alone does not remove the benefit. Likewise, the central-8 results are mixed rather than a uniformly positive smaller effect. We retain these outcomes rather than collapse the campaign into a universal symmetry rule.

D ALL SCALAR-COMPARISON REGIMES AND HELD LAYOUTS

Table 7 reports primary and non-primary conditions from the same twelve models. In ideal coplanar geometry, two central rays separated by $\theta \leq 90^\circ$ give $\lambda_{\min}(H) = (1 - \cos\theta)/2$. The chosen gate therefore becomes zero at 60°, as well as for orthogonal pairs and symmetric broad layouts. It assigns the same zero to acquisitions that differ in measured content and decoder difficulty. This limitation belongs to the tested descriptor, not to every observability-based gate.

Table 6: **Primary regimes and intermediate spans.** Three-seed means, source acquisitions. These values accompany Figure 4; all offsets and seed SDs follow in Table 7.

Acquisition	OFF	FULL	SCALAR	DIRECTIONAL
1 view	0.4974	0.8084	0.8136	0.8029
2 cluster	0.7033	0.8228	0.8287	0.8220
2 spread	0.8610	0.8711	0.8550	0.8622
4 cluster	0.8285	0.8520	0.8590	0.8546
4 spread	0.9137	0.9043	0.9168	0.9165
12 views	0.9298	0.9136	0.9335	0.9316
Two views, $0^\circ/30^\circ$	0.8107	0.8357	0.8444	0.8402
Two views, $0^\circ/45^\circ$	0.7370	0.8500	0.8414	0.8483
Two views, $0^\circ/60^\circ$	0.7601	0.8535	0.7614	0.7524
Two views, $0^\circ/75^\circ$	0.8137	0.8565	0.8220	0.8203

RMS matching controls descriptor magnitude. Nevertheless, DIRECTIONAL's held-layout difference from SCALAR is -0.00213 Dice (paired-patient interval $[-0.00393, -0.00038]$). The replicated single-view depth contrast therefore does not establish a benefit for mixed-view weak-direction projection.

Table 7: **Complete scalar-comparison condition summaries.** Mean ± sample SD across three initializations, 37 LIDC-IDRI development patients. "cl." denotes clustered and "sp." spread. Offsets and explicit angles are in degrees. Identical one-view/spread and twelve-view/clustered summaries are omitted only to avoid duplicate rows; no distinct evaluated acquisition is removed.

Condition	OFF	FULL	SCALAR	DIRECTIONAL
1 cl.; offset 0	0.4974 ± 0.0025	0.8084 ± 0.0025	0.8136 ± 0.0025	0.8029 ± 0.0030
1 cl.; offset 7.5	0.4897 ± 0.0078	0.8085 ± 0.0013	0.8140 ± 0.0028	0.8043 ± 0.0033
2 cl.; offset 0	0.7033 ± 0.0061	0.8228 ± 0.0024	0.8287 ± 0.0022	0.8220 ± 0.0024
2 cl.; offset 7.5	0.6987 ± 0.0029	0.8251 ± 0.0018	0.8304 ± 0.0026	0.8250 ± 0.0028
2 sp.; offset 0	0.8610 ± 0.0044	0.8711 ± 0.0012	0.8550 ± 0.0032	0.8622 ± 0.0010
2 sp.; offset 7.5	0.8548 ± 0.0047	0.8676 ± 0.0015	0.8502 ± 0.0005	0.8580 ± 0.0019
4 cl.; offset 0	0.8285 ± 0.0019	0.8520 ± 0.0019	0.8590 ± 0.0018	0.8546 ± 0.0009
4 cl.; offset 7.5	0.8250 ± 0.0025	0.8522 ± 0.0021	0.8572 ± 0.0019	0.8532 ± 0.0010
4 sp.; offset 0	0.9137 ± 0.0011	0.9043 ± 0.0019	0.9168 ± 0.0007	0.9165 ± 0.0012
4 sp.; offset 7.5	0.9123 ± 0.0015	0.9008 ± 0.0030	0.9138 ± 0.0018	0.9124 ± 0.0003
8 cl.; offset 0	0.9048 ± 0.0006	0.8908 ± 0.0007	0.9040 ± 0.0016	0.9018 ± 0.0031
8 cl.; offset 7.5	0.9041 ± 0.0011	0.8897 ± 0.0009	0.9047 ± 0.0009	0.9019 ± 0.0031
8 sp.; offset 0	0.9244 ± 0.0012	0.9097 ± 0.0018	0.9273 ± 0.0008	0.9260 ± 0.0011
8 sp.; offset 7.5	0.9223 ± 0.0013	0.9083 ± 0.0019	0.9255 ± 0.0012	0.9241 ± 0.0010
12 sp.; offset 0	0.9298 ± 0.0016	0.9136 ± 0.0024	0.9335 ± 0.0009	0.9316 ± 0.0014
12 sp.; offset 7.5	0.9276 ± 0.0016	0.9124 ± 0.0022	0.9322 ± 0.0008	0.9302 ± 0.0013
Held: 0, 15, 30, 45, 60, 75, 90, 120	0.9114 ± 0.0008	0.8977 ± 0.0001	0.9144 ± 0.0015	0.9107 ± 0.0014
Held: 0, 15, 45, 60	0.8460 ± 0.0026	0.8600 ± 0.0036	0.8634 ± 0.0031	0.8639 ± 0.0022
Held: 0, 30	0.8107 ± 0.0015	0.8357 ± 0.0025	0.8444 ± 0.0024	0.8402 ± 0.0017
Held: 0, 30, 60, 90	0.8958 ± 0.0021	0.8906 ± 0.0021	0.8965 ± 0.0027	0.8953 ± 0.0043
Held: 0, 45	0.7370 ± 0.0120	0.8500 ± 0.0041	0.8414 ± 0.0030	0.8483 ± 0.0011
Held: 0, 60	0.7601 ± 0.0061	0.8535 ± 0.0041	0.7614 ± 0.0097	0.7524 ± 0.0061
Held: 0, 75	0.8137 ± 0.0050	0.8565 ± 0.0051	0.8220 ± 0.0021	0.8203 ± 0.0026
Held: 7.5, 22.5, 37.5, 52.5, 67.5, 82.5, 97.5, 127.5	0.9132 ± 0.0008	0.8961 ± 0.0015	0.9144 ± 0.0023	0.9108 ± 0.0023
Held: 7.5, 22.5, 52.5, 67.5	0.8441 ± 0.0057	0.8596 ± 0.0045	0.8605 ± 0.0029	0.8615 ± 0.0031
Held: 7.5, 37.5	0.7936 ± 0.0070	0.8405 ± 0.0023	0.8464 ± 0.0038	0.8414 ± 0.0021
Held: 7.5, 37.5, 67.5, 97.5	0.8946 ± 0.0010	0.8888 ± 0.0022	0.8947 ± 0.0022	0.8918 ± 0.0038
Held: 7.5, 52.5	0.7414 ± 0.0116	0.8492 ± 0.0022	0.8372 ± 0.0013	0.8430 ± 0.0025
Held: 7.5, 67.5	0.7585 ± 0.0059	0.8484 ± 0.0050	0.7662 ± 0.0046	0.7561 ± 0.0064
Held: 7.5, 82.5	0.8322 ± 0.0004	0.8624 ± 0.0033	0.8451 ± 0.0017	0.8424 ± 0.0028

E SECOND-COHORT SCALAR AND AGGREGATION REPLICATIONS

All thirteen new runs completed 250 epochs under their locked schedules: nine TotalSegmentator models and four LIDC aggregation models. The original two seed-zero aggregation models were reused. Every new final checkpoint passed hash, exact prediction-reload, and unchanged-evaluation-state checks; the saved histories match the locked patient/acquisition schedules. No final-test data enter these results. The 34 TotalSegmentator and 37 LIDC development cases remain the independent assessment units, not the numbers of acquisition records.

Scalar screen and failure boundary. TotalSegmentator passes each relative criterion: within 0.005 of FULL in each primary sparse group, within 0.005 of OFF in each primary broad group, and average broad gain over FULL of at least 0.010, positive in every seed. Broad gains are 0.04342, 0.03454, and 0.04413 Dice for seeds 0, 1, and 2. Their mean is 0.04070, sample seed SD 0.00534, and paired-case 95% interval [0.02874, 0.05519]. The absolute LIDC cutoffs were not transferred. Source one-view OFF/FULL/SCALAR means are 0.4937/0.7187/0.7406; two-clustered means are 0.6006/0.7368/0.7562. Four-spread means are 0.8562/0.8178/0.8563; twelve-view means are 0.8744/0.8269/0.8737.

All-layout retention fails. At the source $0^\circ/60^\circ$ pair, SCALAR reaches 0.7011 versus FULL's 0.7637. The worst difference from the better OFF/FULL reference is -0.06552 at $7.5^\circ/67.5^\circ$. Two-spread and eight-clustered conditions also expose losses outside the primary screen. Table 8 includes all 42 unique layouts, with seed SD, rather than only the passing subset.

Table 8: **Complete TotalSegmentator scalar replication.** Ordinary lung Dice on 34 reused development cases, mean ± sample SD over three initializations. Angles are degrees; each row is a unique acquisition. All arms use the final epoch 250.

Angles	OFF	FULL	SCALAR
0	0.5193 ± 0.0080	0.7164 ± 0.0045	0.7339 ± 0.0029
7.5	0.5236 ± 0.0099	0.7234 ± 0.0078	0.7391 ± 0.0018
45	0.4680 ± 0.0138	0.7210 ± 0.0046	0.7473 ± 0.0056
52.5	0.4370 ± 0.0149	0.7144 ± 0.0061	0.7466 ± 0.0054
0, 15	0.6487 ± 0.0082	0.7364 ± 0.0050	0.7513 ± 0.0025
0, 30	0.7471 ± 0.0062	0.7561 ± 0.0043	0.7699 ± 0.0024
0, 45	0.7339 ± 0.0089	0.7606 ± 0.0015	0.7794 ± 0.0022
0, 60	0.7221 ± 0.0062	0.7637 ± 0.0062	0.7011 ± 0.0063
0, 75	0.7568 ± 0.0072	0.7770 ± 0.0078	0.7611 ± 0.0026
0, 90	0.7921 ± 0.0030	0.7911 ± 0.0055	0.7875 ± 0.0027
7.5, 22.5	0.6463 ± 0.0067	0.7428 ± 0.0054	0.7534 ± 0.0008
7.5, 37.5	0.7323 ± 0.0043	0.7581 ± 0.0038	0.7734 ± 0.0007
7.5, 52.5	0.7221 ± 0.0036	0.7624 ± 0.0021	0.7740 ± 0.0027
7.5, 67.5	0.7227 ± 0.0088	0.7690 ± 0.0057	0.7035 ± 0.0055
7.5, 82.5	0.7850 ± 0.0032	0.7903 ± 0.0081	0.7778 ± 0.0061
7.5, 97.5	0.7891 ± 0.0058	0.7903 ± 0.0074	0.7745 ± 0.0027
45, 60	0.5526 ± 0.0107	0.7371 ± 0.0035	0.7611 ± 0.0053
45, 135	0.7676 ± 0.0110	0.7792 ± 0.0024	0.7613 ± 0.0043
52.5, 67.5	0.5464 ± 0.0194	0.7363 ± 0.0022	0.7604 ± 0.0057
52.5, 142.5	0.7593 ± 0.0115	0.7796 ± 0.0025	0.7470 ± 0.0026
0, 15, 30, 45	0.7647 ± 0.0062	0.7636 ± 0.0039	0.7799 ± 0.0002
0, 15, 45, 60	0.7815 ± 0.0015	0.7724 ± 0.0037	0.7933 ± 0.0029
0, 30, 60, 90	0.8368 ± 0.0042	0.8052 ± 0.0054	0.8268 ± 0.0028
0, 45, 90, 135	0.8562 ± 0.0060	0.8178 ± 0.0053	0.8563 ± 0.0024
7.5, 22.5, 37.5, 52.5	0.7511 ± 0.0059	0.7655 ± 0.0029	0.7816 ± 0.0011
7.5, 22.5, 52.5, 67.5	0.7729 ± 0.0019	0.7755 ± 0.0031	0.7875 ± 0.0013
7.5, 37.5, 67.5, 97.5	0.8326 ± 0.0059	0.8043 ± 0.0068	0.8203 ± 0.0033
7.5, 52.5, 97.5, 142.5	0.8557 ± 0.0047	0.8178 ± 0.0055	0.8510 ± 0.0024
45, 60, 75, 90	0.7107 ± 0.0145	0.7733 ± 0.0029	0.7913 ± 0.0054
52.5, 67.5, 82.5, 97.5	0.7041 ± 0.0207	0.7735 ± 0.0048	0.7902 ± 0.0061
0, 15, 30, 45, 60, 75, 90, 105	0.8455 ± 0.0051	0.8097 ± 0.0055	0.8391 ± 0.0029
0, 15, 30, 45, 60, 75, 90, 120	0.8538 ± 0.0053	0.8124 ± 0.0063	0.8480 ± 0.0039

Continued on next page

Table 8 continued

Angles	OFF	FULL	SCALAR
0, 15, 45, 75, 90, 105, 135, 165	0.8692 ± 0.0044	0.8245 ± 0.0058	0.8683 ± 0.0026
0, 30, 45, 60, 90, 120, 135, 150	0.8668 ± 0.0057	0.8230 ± 0.0063	0.8672 ± 0.0027
7.5, 22.5, 37.5, 52.5, 67.5, 82.5, 97.5, 112.5	0.8434 ± 0.0047	0.8108 ± 0.0062	0.8319 ± 0.0048
7.5, 22.5, 37.5, 52.5, 67.5, 82.5, 97.5, 127.5	0.8531 ± 0.0039	0.8147 ± 0.0061	0.8441 ± 0.0041
7.5, 22.5, 52.5, 82.5, 97.5, 112.5, 142.5, 172.5	0.8693 ± 0.0025	0.8258 ± 0.0066	0.8639 ± 0.0030
7.5, 37.5, 52.5, 67.5, 97.5, 127.5, 142.5, 157.5	0.8657 ± 0.0048	0.8223 ± 0.0058	0.8626 ± 0.0027
45, 60, 75, 90, 105, 120, 135, 150	0.8259 ± 0.0077	0.8084 ± 0.0034	0.8106 ± 0.0033
52.5, 67.5, 82.5, 97.5, 112.5, 127.5, 142.5, 157.5	0.8247 ± 0.0071	0.8087 ± 0.0042	0.8088 ± 0.0039
0, 15, 30, 45, 60, 75, 90, 105, 120, 135, 150, 165	0.8744 ± 0.0048	0.8269 ± 0.0067	0.8737 ± 0.0027
7.5, 22.5, 37.5, 52.5, 67.5, 82.5, 97.5, 112.5, 127.5, 142.5, 157.5, 172.5	0.8742 ± 0.0029	0.8275 ± 0.0069	0.8716 ± 0.0024

Attention replication. Fixed aggregation meets the 0.005 sparse retention margin at all four primary source/offset groups. Fixed-minus-learned source one-view Dice is 0.00247, with paired-patient interval [0.00069, 0.00430]; the offset contrast is 0.00187 [−0.00057, 0.00422]. Source two-clustered Dice differs by 0.00174 [−0.00031, 0.00385], and offset two-clustered by 0.00129 [−0.00107, 0.00367]. These intervals condition on the three fitted initializations. We use the prespecified retention margin, not nonsignificance as evidence of equivalence. Table 9 retains all 28 unique layouts, including the broader acquisitions.

Table 9: **Complete fixed/learned aggregation replication.** Ordinary lung Dice on 37 reused LIDC development patients, mean ± sample seed SD. Geometry conditioning is off in both arms. Angles are degrees.

Angles	Fixed aggregation	Learned aggregation
0	0.8131 ± 0.0010	0.8121 ± 0.0008
7.5	0.8176 ± 0.0033	0.8159 ± 0.0014
45	0.8304 ± 0.0037	0.8266 ± 0.0021
52.5	0.8248 ± 0.0029	0.8227 ± 0.0034
0, 15	0.8277 ± 0.0021	0.8258 ± 0.0004
0, 90	0.8780 ± 0.0009	0.8741 ± 0.0003
7.5, 22.5	0.8326 ± 0.0028	0.8305 ± 0.0007
7.5, 97.5	0.8762 ± 0.0003	0.8726 ± 0.0006
45, 60	0.8384 ± 0.0026	0.8368 ± 0.0017
45, 135	0.8707 ± 0.0025	0.8716 ± 0.0023
52.5, 67.5	0.8360 ± 0.0010	0.8356 ± 0.0010
52.5, 142.5	0.8672 ± 0.0029	0.8683 ± 0.0020
0, 15, 30, 45	0.8574 ± 0.0025	0.8556 ± 0.0023
0, 45, 90, 135	0.9000 ± 0.0025	0.8980 ± 0.0007
7.5, 22.5, 37.5, 52.5	0.8578 ± 0.0033	0.8569 ± 0.0014
7.5, 52.5, 97.5, 142.5	0.8986 ± 0.0018	0.8965 ± 0.0002
45, 60, 75, 90	0.8585 ± 0.0027	0.8584 ± 0.0014
52.5, 67.5, 82.5, 97.5	0.8558 ± 0.0031	0.8569 ± 0.0022
0, 15, 30, 45, 60, 75, 90, 105	0.8931 ± 0.0022	0.8933 ± 0.0008
0, 15, 45, 75, 90, 105, 135, 165	0.9062 ± 0.0024	0.9042 ± 0.0005
0, 30, 45, 60, 90, 120, 135, 150	0.9043 ± 0.0021	0.9026 ± 0.0007
7.5, 22.5, 37.5, 52.5, 67.5, 82.5, 97.5, 112.5	0.8916 ± 0.0023	0.8923 ± 0.0007
7.5, 22.5, 52.5, 82.5, 97.5, 112.5, 142.5, 172.5	0.9050 ± 0.0022	0.9028 ± 0.0008
7.5, 37.5, 52.5, 67.5, 97.5, 127.5, 142.5, 157.5	0.9045 ± 0.0018	0.9030 ± 0.0010

Continued on next page

Table 9 continued

Angles	Fixed aggregation	Learned aggregation
45, 60, 75, 90, 105, 120, 135, 150	0.8867 ± 0.0021	0.8871 ± 0.0010
52.5, 67.5, 82.5, 97.5, 112.5, 127.5, 142.5, 157.5	0.8868 ± 0.0023	0.8868 ± 0.0007
0, 15, 30, 45, 60, 75, 90, 105, 120, 135, 150, 165	0.9081 ± 0.0022	0.9062 ± 0.0003
7.5, 22.5, 37.5, 52.5, 67.5, 82.5, 97.5, 112.5, 127.5, 142.5, 157.5, 172.5	0.9077 ± 0.0022	0.9056 ± 0.0005

F REPLICATED CONTEXT AND TRAINING-COVERAGE CONTROLS

F.1 ISOTROPIC VERSUS ACQUISITION-ALIGNED CONTEXT

Fifteen fresh fits compare five coordinate-free decoders at three paired initializations. Each uses 178 training patients, the original balanced mixed-view schedule, 250 epochs, and the final checkpoint. The arms share active parameter counts and initial values. All modify only the two convolutions at 32^3, leaving the encoder, lifting, normalization, activations, and later decoding unchanged.

ISO-FIXED uses isotropic dilations 2 and 3. For the other context arms, let $\lambda_0 \le \lambda_1 \le \lambda_2$ be the eigenvalues of H in Equation 3. The adaptive strength is $s = (\lambda_1 - \lambda_0)/(\lambda_1 + \lambda_0)$, clamped to $[0, 1]$ and set to zero when the relative eigenvalue gap is at most 10^{-6}. ISO-ADAPTIVE scales kernel offsets isotropically by $1 + (d_{\max} - 1)s$. ALIGNED instead stretches only along the weakest eigenvector u, using $I + (d_{\max} - 1)suu^\top$. PERPENDICULAR uses the normalized cross product of detector-up and u. Physical voxel spacing is included in the transform; fractional offsets use trilinear sampling with zero padding. Zero extension takes the original convolution path. These operations enlarge context without adding an absolute coordinate input.

Table 10: **Coordinate-free spatial context, three initializations.** Source-acquisition ordinary lung Dice, mean $\pm$ seed SD, on 37 reused development patients. These are fresh matched models, not additional seeds of the earlier four-arm context-by-coordinate factorial.

Acquisition	Off	Iso-fixed	Iso-adaptive	Aligned	Perpendicular
One	0.4945 ± 0.0129	0.6343 ± 0.0146	0.6409 ± 0.0139	0.6012 ± 0.0051	0.5751 ± 0.0107
Two clustered	0.6926 ± 0.0070	0.7564 ± 0.0061	0.7580 ± 0.0022	0.7332 ± 0.0051	0.7325 ± 0.0011
Two spread	0.8582 ± 0.0096	0.8842 ± 0.0032	0.8599 ± 0.0015	0.8574 ± 0.0038	0.8561 ± 0.0028
Four spread	0.9127 ± 0.0022	0.9167 ± 0.0024	0.9153 ± 0.0010	0.9145 ± 0.0024	0.9155 ± 0.0030
Twelve	0.9287 ± 0.0012	0.9271 ± 0.0010	0.9297 ± 0.0002	0.9294 ± 0.0023	0.9278 ± 0.0024

Isotropic context improves sparse prediction (Table 10), but the directional hypothesis fails. On the prespecified sparse average over one and two-clustered views at source and offset phases, ALIGNED trails ISO-ADAPTIVE by 3.04 Dice points (paired-patient 95% interval $[-3.38, -2.68]$) and exceeds PERPENDICULAR by only 1.41 points. The proposed method had to exceed both by at least one point and be positive in every seed; it does not. Its broad-retention screen passes, while sparse and all-layout retention fail. The result does not show that all forms of directional processing must fail.

A separate three-seed combination of adaptive isotropic context and SCALAR reaches 0.8168 one-view and 0.9331 twelve-view source Dice. At $0^\circ/60^\circ$ it reaches only 0.7739, compared with SCALAR's 0.7614; the offset score is 0.7805. Both miss the locked 0.80 intermediate-band target. Favorable context and gating results measured separately therefore do not guarantee that their combination closes the gap. Neither comparison licenses an additive causal decomposition of the coordinate benefit.

F.2 A MATCHED CHANGE TO TWO-VIEW TRAINING COVERAGE

The original mixed-view schedule uses two-view separations of 15° and 90° only. To distinguish this limited exposure from the gate itself, we train fresh FULL and SCALAR models at three paired seeds. Only the two-view schedule changes: each patient's 25 formerly clustered presentations are balanced over $15^\circ/30^\circ/45^\circ$, and its 25 formerly spread presentations over $75^\circ/90^\circ$. The source

phases, non-two-view presentations, initialization, optimizer, loss, and budget remain fixed. Each patient still receives 25 positive-gate and 25 zero-gate two-view inputs, although positive gate magnitudes change. This is a discrete five-span policy, not unrestricted angle randomization.

The inclusive interval $[52.5^\circ, 67.5^\circ]$ is withheld from training. We assess its three separations at phases 0° and 7.5°, averaging the six layouts within patient before averaging patients and seeds. Original models are re-evaluated on these layouts; all historical scores replay exactly. This controls model-training exposure, not research-program exposure: the 60° cases and all 37 patients were already inspected during development.

Table 11: **Broader training reduces, but does not resolve, the held-span deficit.** Mean ± seed SD across three initializations. Other than the two explicitly marked offset/interval rows, acquisitions are source layouts. Original and broader columns differ only in the two-view training policy.

	FULL		SCALAR	
Acquisition	Original	Broader	Original	Broader
Source 60°	0.8535 ± 0.0041	0.8586 ± 0.0025	0.7614 ± 0.0097	0.7996 ± 0.0078
Offset 60°	0.8484 ± 0.0050	0.8568 ± 0.0011	0.7662 ± 0.0046	0.8067 ± 0.0059
Withheld interval	0.8514 ± 0.0042	0.8582 ± 0.0018	0.7880 ± 0.0018	0.8228 ± 0.0034
One view	0.8084 ± 0.0025	0.8124 ± 0.0031	0.8136 ± 0.0025	0.8085 ± 0.0032
Two spread	0.8711 ± 0.0012	0.8706 ± 0.0010	0.8550 ± 0.0032	0.8481 ± 0.0048
Four spread	0.9043 ± 0.0019	0.9061 ± 0.0021	0.9168 ± 0.0007	0.9168 ± 0.0005
Twelve views	0.9136 ± 0.0024	0.9154 ± 0.0026	0.9335 ± 0.0009	0.9324 ± 0.0006

SCALAR's source-60° gain is 3.83 points, with a paired-patient interval of [3.30, 4.32]; the offset gain is 4.05 [3.65, 4.43]. The six-layout interval gain is 3.47 [3.13, 3.78], positive in all three seeds. The corresponding interval gain for FULL is smaller: the difference-in-differences is 2.79 points [2.46, 3.08]. These intervals average paired seed effects before resampling patient identities; they are not uncertainty over new training seeds.

The locked remedy screen nevertheless fails: source-60° Dice remains below 0.82, and three of ten primary retention contrasts fall below -0.005. Their differences are -0.00516 at one-view source, -0.00692 at two-orthogonal source, and -0.00511 at two-orthogonal offset. Close misses remain misses. The coverage intervention supports a contribution from training exposure, without proving that it explains the entire failure or that gating imposes an unavoidable limit.

G EXPLORATORY FUSION CONTROLS

These controls test whether view aggregation discards information needed in the intermediate band. They are subsequent adaptive-development experiments, not a replacement for the coordinate analysis. Every arm uses the unchanged SCALAR gate and broader acquisition schedule from Appendix F.2. Let z_{vj} be the per-view lifted feature at query j, α_{vj} its normalized routing mass, and $\mu_j = \sum_v \alpha_{vj} z_{vj}$. Each residual is added to the historical pooled feature before LayerNorm and coordinate addition. Its final projection starts at zero.

MEAN transforms μ_j through a bias-free $256 \to 96 \to 256$ MLP. STATS supplies the mean, weighted feature variance, and mean axial ray geometry to a width-64 MLP. A generic per-view RAY-MLP transforms features and axial geometry before weighted pooling. The candidate T instead retains their centered association:

$$T_j = \sum_v \alpha_{vj} \big(Bz_{vj} - \overline{Bz}_j\big)\big(q_{vj} - \bar{q}_j\big)^\top, \tag{5}$$

where B compresses features to 32 channels and q has five trace-free components of the unit-ray outer product $rr^\top$. A GELU and linear projection map this statistic back to 256 channels. It is invariant to joint view permutation and ray sign, and vanishes at one view. These algebraic properties do not imply equivariance of the whole model or equal one-view predictions after training. Added counts are 49,152, 49,472, 49,632, and 49,152 parameters for MEAN, STATS, RAY-MLP, and T, respectively.

The four-arm seed-zero screen required T to exceed each control by at least 0.010 Dice over the withheld interval. It failed. A separately locked follow-up added seeds 1 and 2 for T and STATS, and

three seeds each for two combinations: STATS+T and STATS+MEAN. Both combinations add exactly 98,624 parameters and start from matched shared weights. Their active one-view computation differs because T is skipped at one view. Each new fit uses 250 epochs, 44,500 patient presentations, and 11,250 updates. Final checkpoints alone are assessed on the same 37 development patients and 46 layouts.

Table 12: **Fusion follow-ups do not establish a complete remedy.** Ordinary Dice, mean $\pm$ seed SD over three initializations. The first five rows are source acquisitions; the interval averages six held-span layouts. STATS+MEAN is the matched-parameter control for STATS+T.

Acquisition	T	Stats	Stats+T	Stats+Mean
One view	0.8134 ± 0.0020	0.8148 ± 0.0015	0.8151 ± 0.0009	0.8120 ± 0.0030
Two clustered	0.8281 ± 0.0025	0.8393 ± 0.0023	0.8397 ± 0.0014	0.8375 ± 0.0030
Two spread	0.8689 ± 0.0016	0.8674 ± 0.0045	0.8684 ± 0.0022	0.8681 ± 0.0046
Four spread	0.9200 ± 0.0020	0.9187 ± 0.0024	0.9219 ± 0.0010	0.9210 ± 0.0010
Twelve views	0.9339 ± 0.0021	0.9319 ± 0.0021	0.9351 ± 0.0013	0.9341 ± 0.0016
Source 60°	0.8119 ± 0.0103	0.8152 ± 0.0109	0.8031 ± 0.0148	0.8153 ± 0.0043
Offset 60°	0.8280 ± 0.0039	0.8123 ± 0.0058	0.8175 ± 0.0116	0.8128 ± 0.0068
Withheld interval	0.8373 ± 0.0054	0.8308 ± 0.0060	0.8340 ± 0.0098	0.8331 ± 0.0031

T exceeds STATS on the interval by 0.656 Dice points, positive in all three seeds, with a paired-patient interval of [0.315, 1.006] points. This is a modest replicated difference, not a pass of the historical one-point screen or evidence of superiority on every layout. The combination improves over its matched control by only 0.090 points, with interval $[-0.208, 0.398]$; one of three seed effects is negative. It retains all ten primary groups relative to broader SCALAR, but fails the interval-gain and source/offset-60° requirements. Broader FULL still scores 0.8582 on the interval, above all four variants in Table 12. We retain the negative combination result and do not promote T as a resolved or uniquely effective method.

H PROSPECTIVELY LOCKED LCTSC REPLICATION

H.1 COHORT, TARGETS, AND PREPARATION

LCTSC contains 60 thoracic CT cases from three institutions, 20 per institution, with expert lung and heart RTSTRUCT contours (Yang et al., 2017; 2018). We obtained the public release through TCIA (Clark et al., 2013), under its CC BY 3.0 license. We preserve its 36 training cases and pool the 12 off-site and 12 live-test cases into a 24-case assessment, with eight test cases per institution. No case is excluded or replaced. This is the first use of LCTSC in this project, according to project records and author confirmation. Released case identifiers define subjects; exact-pixel duplicate checks do not establish cross-dataset person-level nonoverlap.

The replication uses manual Lung_L, Lung_R, and Heart contours, not automated labels. We verify CT/RTSTRUCT frame references, slice geometry, contour types, nonempty bilateral labels, and native left/right correspondence. Targets occupy the full original scan box, without a label-defined crop. The unchanged renderer uses 24 views at 7.5° increments, 512^2 detectors, 256 samples per ray, and one detector scale per case; stored projections are float16. Occupancy targets use a 96^3 grid with supersampling factor two. Heart is treated as background by the lung loss. We report ordinary contour Dice, not the original challenge metric with challenge-specific ignored regions.

Two preparation attempts stopped before model training because strict left/right overlap checks failed. A training-only audit found small intersections in the supplied contours: at most 189 native voxels, less than 0.008% of the lung union across all 36 training cases. Before converting the test cases, a documented amendment fixed an overlap guard of 0.1%, retained the original deterministic label priority (right over left lung, then heart over lung), and required every overlap to be recorded. All 60 cases passed the final conversion. The failed attempts and amendment remain in the audit record. Test-volume access for deterministic conversion and geometry QA occurred before prediction assessment; it was not model selection.

H.2 TRAINING LOCK AND PREDECLARED CONTRASTS

OFF and FULL use the same fixed-lifting architecture as the mixed-view comparisons, with world-centered coordinates as the sole arm difference. Six models train from scratch: two arms at each of three seeds, with paired shared-weight initialization, patient order, and acquisition schedule. Both register 1,327,555 parameters; OFF freezes and bypasses the 16,896-parameter coordinate MLP. Each fit has exactly 250 epochs, 9,000 patient presentations, and 2,250 optimizer updates. Each training case receives 50 presentations at each of 1, 2, 4, 8, and 12 views; the intermediate counts split equally between clustered and spread layouts. Training phases are $0^\circ, 30^\circ, \ldots, 150^\circ$. Loss, optimizer, accumulation, clipping, and temperature schedule are those in Appendix A; training uses FP32 with TF32 disabled. Equal epochs across cohorts do not mean equal update counts.

Training-only qualification checks finite outputs and active gradients, loss decrease, disabled-path behavior, and checkpoint/optimizer reload. Qualification updates are discarded. There is no validation-based checkpoint selection, early stopping, budget extension, or test-set tuning. Every model uses its final epoch-250 checkpoint, and all six final checkpoints are frozen before the first test prediction. The locked protocol precedes this cohort's model fitting and assessment; it is a prospective replication motivated by earlier adaptive research, not a preregistration of the entire project.

Let Δ_N be FULL minus OFF Dice at a source layout. The primary predictions require $\Delta_1 > 0$ and $\Delta_1 - \Delta_{12} > 0$ in all three seeds and positive lower paired-patient 95% bootstrap bounds. A dense reversal is a separate stronger finding requiring the upper bound for Δ_{12} to be negative. All four primary checks pass; the dense reversal also meets its criterion. The one-minus-twelve interaction is +37.16 points [35.64, 38.67]. The prespecified two-view clustered-minus-spread interaction is +23.18 points [21.59, 24.68]; it supplements rather than replaces the endpoint hypotheses.

For each contrast we average the three paired seed effects within patient, then perform 10,000 patient bootstrap resamples, stratified by institution with eight draws per institution (random seed 2026092401). Intervals are conditional on these fitted initializations and this cohort. Model-seed SD is reported separately. Neither the 2,160 patient/acquisition/model records nor the repeated phase descriptions increase the sample size beyond 24 patients. Secondary intervals are descriptive and are not adjusted for multiple comparisons.

H.3 COMPLETE ACQUISITION AND SECONDARY RESULTS

The five base layouts are $[0]$, $[0, 15]$, $[0, 90]$, $[0, 45, 90, 135]$, and $[0, 15, \ldots, 165]$ degrees. Phases 0°, 45°, 135°, and 7.5° are applied modulo 180°. This defines 20 named groups but only 15 distinct angle sets: the two rotated two-spread descriptions coincide, and the source/45°/135° four-spread and twelve-view sets coincide. Table 13 preserves those group names for auditability, not as independent replication. Individual angles can overlap training even when the global phase is held out. The offset conditions use held 7.5°-shifted angles.

Table 13: **All locked LCTSC acquisition groups.** Ordinary Dice, mean ± seed SD; difference and paired-patient interval in Dice points. Twenty named groups share 15 unique angle sets.

Layout	Phase	OFF	FULL	Difference	95% interval
One view	0°	0.3637 ± 0.0046	0.6719 ± 0.0018	$+30.81$	$[+28.96, +32.38]$
One view	45°	0.2341 ± 0.0213	0.6527 ± 0.0045	$+41.85$	$[+40.00, +43.71]$
One view	135°	0.2094 ± 0.0285	0.6527 ± 0.0098	$+44.33$	$[+41.98, +46.52]$
One view	7.5°	0.3640 ± 0.0048	0.6735 ± 0.0030	$+30.95$	$[+29.06, +32.54]$
Two clustered	0°	0.4837 ± 0.0124	0.6792 ± 0.0025	$+19.55$	$[+17.88, +21.07]$
Two clustered	45°	0.3980 ± 0.0218	0.6648 ± 0.0060	$+26.67$	$[+24.34, +29.02]$
Two clustered	135°	0.4022 ± 0.0038	0.6708 ± 0.0094	$+26.86$	$[+24.75, +28.77]$
Two clustered	7.5°	0.4779 ± 0.0130	0.6762 ± 0.0047	$+19.83$	$[+18.15, +21.40]$
Two spread	0°	0.7605 ± 0.0048	0.7242 ± 0.0066	-3.63	$[-5.15, -2.13]$
Two spread	45°	0.7238 ± 0.0082	0.6967 ± 0.0030	-2.71	$[-5.00, -0.50]$
Two spread	135°	0.7238 ± 0.0082	0.6967 ± 0.0030	-2.71	$[-5.00, -0.50]$
Two spread	7.5°	0.7583 ± 0.0027	0.7228 ± 0.0060	-3.56	$[-5.22, -2.03]$

Continued on next page

Table 13 continued

Layout	Phase	OFF	FULL	Difference	95% interval
Four spread	0°	0.7940 ± 0.0029	0.7346 ± 0.0049	-5.94	$[-7.64, -4.31]$
Four spread	45°	0.7940 ± 0.0029	0.7346 ± 0.0049	-5.94	$[-7.64, -4.31]$
Four spread	135°	0.7940 ± 0.0029	0.7346 ± 0.0049	-5.94	$[-7.64, -4.31]$
Four spread	7.5°	0.7944 ± 0.0010	0.7367 ± 0.0058	-5.77	$[-7.24, -4.39]$
Twelve views	0°	0.8033 ± 0.0015	0.7398 ± 0.0056	-6.34	$[-7.86, -4.90]$
Twelve views	45°	0.8033 ± 0.0015	0.7398 ± 0.0056	-6.34	$[-7.86, -4.90]$
Twelve views	135°	0.8033 ± 0.0015	0.7398 ± 0.0056	-6.34	$[-7.86, -4.90]$
Twelve views	7.5°	0.8020 ± 0.0009	0.7393 ± 0.0056	-6.27	$[-7.76, -4.86]$

Table 14: **LCTSC secondary checks.** FULL minus OFF in Dice points. Heart-excluding effects have paired-patient 95% intervals; institution columns use ordinary Dice, with eight test patients per institution. These descriptive analyses are not independent cohorts or adjusted multiple-comparison tests.

Layout	Phase	Heart-excl.	95% interval	Site 1	Site 2	Site 3
One view	0°	$+31.32$	$[+29.40, +32.92]$	$+30.16$	$+28.54$	$+33.73$
One view	45°	$+42.48$	$[+40.66, +44.33]$	$+42.36$	$+39.27$	$+43.93$
One view	135°	$+45.08$	$[+42.72, +47.21]$	$+42.04$	$+43.40$	$+47.56$
One view	7.5°	$+31.43$	$[+29.46, +33.05]$	$+30.74$	$+28.75$	$+33.36$
Two clustered	0°	$+19.72$	$[+17.97, +21.30]$	$+19.66$	$+17.80$	$+21.19$
Two clustered	45°	$+26.92$	$[+24.56, +29.29]$	$+26.65$	$+25.73$	$+27.64$
Two clustered	135°	$+27.20$	$[+25.07, +29.12]$	$+25.02$	$+27.18$	$+28.38$
Two clustered	7.5°	$+19.92$	$[+18.19, +21.54]$	$+20.05$	$+18.61$	$+20.83$
Two spread	0°	-3.55	$[-5.03, -2.11]$	-3.64	-3.08	-4.17
Two spread	45°	-2.69	$[-4.97, -0.46]$	-4.46	-2.78	-0.89
Two spread	135°	-2.69	$[-4.97, -0.46]$	-4.46	-2.78	-0.89
Two spread	7.5°	-3.39	$[-5.00, -1.93]$	-2.86	-2.48	-5.33
Four spread	0°	-5.81	$[-7.46, -4.25]$	-6.46	-5.98	-5.39
Four spread	45°	-5.81	$[-7.46, -4.25]$	-6.46	-5.98	-5.39
Four spread	135°	-5.81	$[-7.46, -4.25]$	-6.46	-5.98	-5.39
Four spread	7.5°	-5.66	$[-7.06, -4.34]$	-5.42	-5.64	-6.25
Twelve views	0°	-6.23	$[-7.67, -4.84]$	-6.35	-6.44	-6.25
Twelve views	45°	-6.23	$[-7.67, -4.84]$	-6.35	-6.44	-6.25
Twelve views	135°	-6.23	$[-7.67, -4.84]$	-6.35	-6.44	-6.25
Twelve views	7.5°	-6.18	$[-7.59, -4.84]$	-6.30	-6.46	-6.05

H.4 Fitting, verification, and scope

Table 15: **LCTSC final-checkpoint fitting versus held-out accuracy.** Ordinary source Dice, mean $\pm$ seed SD. Training uses all 36 fitting patients; test uses the published 24 held-out patients. These are endpoint diagnostics, not checkpoint-selection scores.

	Training		Test	
Views	OFF	FULL	OFF	FULL
1	0.3743 ± 0.0019	0.7298 ± 0.0077	0.3637 ± 0.0046	0.6719 ± 0.0018
12	0.8585 ± 0.0024	0.8381 ± 0.0131	0.8033 ± 0.0015	0.7398 ± 0.0056

FULL also underperforms OFF on twelve-view training fits (Table 15); the held-out penalty cannot be assigned solely to generalization. Final fitting scores do not establish optimization convergence. Lower absolute scores than on the development cohorts need not have a single cause: data, manual-label conventions, full scan extents, and updates per fixed epoch budget differ. This experiment tests the coordinate contrast within LCTSC, not cross-cohort equality of absolute accuracy.

Six final-checkpoint hashes, 54,000 training-history rows, 2,160 assessment records, split/acquisition membership, and source/data locks passed verification. An independent recomputation reproduced all group means, seed SDs, paired intervals, and secondary statistics. Data

checks covered 9,713 raw files and 120 series. Sixteen implementation tests and two training-only GPU smoke tests passed. Sequential training and fitting checks took 172.9 minutes on one A100 80GB; preparation through final assessment took about 3 hours 42 minutes. These are implementation timings, not efficiency comparisons.

Fixed examples use the first assessment case from each institution, at source one and twelve views, for every seed and arm. Saved predictions reproduce their full-volume Dice; visual inspection shows coarse lung shapes with remaining boundary and extent errors. No unfavorable example was replaced. The replication does not evaluate SCALAR, depth/plane inputs, context changes, or zero-shot transfer on LCTSC. It neither validates measured-radiograph use nor establishes near-SOTA segmentation or reconstruction. The earlier development results and failed remedy decisions remain unchanged.

I SCOPE, UNSUCCESSFUL INTERVENTIONS, AND REMAINING CONFIRMATION

What the comparisons establish. The clinical directional contrast and scalar endpoint result are each replicated across three initializations in LIDC and TotalSegmentator. The locked LCTSC test adds a third-cohort replication of OFF/FULL coverage dependence only. The fixed/learned aggregation contrast, coordinate-free context comparison, coverage control, and selected fusion follow-ups use three LIDC initializations. The full attention/conditioning and slot/coordinate factorials and original clinical context-by-coordinate interaction use one each. These are not interchangeable sources of replication. The scalar gate's intermediate-layout losses replicate and are partly reduced by broader training coverage; they do not establish an unavoidable gating failure.

Prior assessment exposure. The 37 LIDC and 34 TotalSegmentator cases have been reused throughout development. The historical 46-case LIDC holdout also has earlier evaluation records; it cannot be described as untouched. No new evaluation of that partition is included here. The historical TotalSegmentator holdout is likewise excluded. LCTSC is the separate, previously unused cohort assessed under the frozen plan in Appendix H; its test result does not retroactively make the earlier development comparisons confirmatory. Exact-file and disjoint-case checks do not replace original identity linkage, so subject accounting follows dataset documentation rather than independent re-identification.

Other attempted remedies. Earlier experiments with learned count, angular-spread, and geometry gates, coordinate dropout, scaling, and restricted coordinate inputs did not establish a generally resolved sparse/broad tradeoff under their respective protocols. A protected-adaptation experiment retained broad-view accuracy but did not recover sufficient sparse accuracy. These are different training interventions, not a matched leaderboard from which one can infer a single cause of failure. The full research record motivated the current comparison; it was not itself preregistered as one confirmatory study.

Follow-up count-protected and collision-based gates improved intermediate two-view performance but did not satisfy their all-layout retention screen. The collision gate did not demonstrate an advantage over the count control. Altering the scalar gate only at inference also failed its continuation criterion. These results do not justify claiming that a collision score orders every layout, that training and inference have incompatible optimal gates in general, or that a remedy is complete.

Checks versus scientific evidence. Integrity checks support reproducibility, not a hypothesis whose prediction fails. Endpoint passes do not erase held-span losses. All reported runs are complete. The locked LCTSC assessment is the only new clinical test evaluation added in this revision; no follow-up tuning on it is reported.

Use and risk. The models are not validated for diagnosis, planning, or dose reduction. Population-consistent masks may fail on atypical anatomy, and simulated projections do not establish robustness to measured radiographs. Public research CT data are used under their applicable access conditions. No deployment recommendation follows.