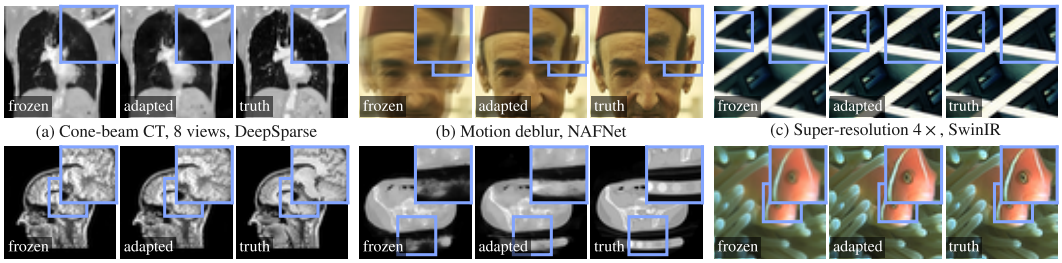
DATA CONSISTENCY REACHES THE NULL SPACE

Anonymous authors
Paper under double-blind review

ABSTRACT

Sparse-view reconstruction is hard because most of the error lives where the measurements cannot see it, the null space of the forward operator. A measurement-consistency loss minimised over the image cannot move that component. We ask whether the same loss reaches the null space when it is minimised through a network instead, and what limits it there. Routed data consistency (RDC) frees one internal activation of a released reconstructor and fits the measurements through the frozen decoder behind it, with every weight fixed. How much it removes from the null space is bounded by one alignment, the cosine $\cos\theta$ between its null-space update and the null-space error. For a fixed operator that alignment is set by the pretrained network and its target, not by the scan, and it caps the removable fraction at $1-\sin\theta$. With the relative length r of the update it forms an identity that ranks candidate variables from their decompositions alone, and a label-free test on withheld views defers to the weights where the frozen mapping is wrong. On eight-view cone-beam CT, forty steps finish 1.72 dB above the range oracle, the bound that no fit of the image can exceed, and surpass a budget-matched supervised finetune given forty-eight labelled volumes. On 18 clinical scans RDC leads weight-space adaptation in 16. Across six networks on four operators the identity predicts the better variable in 110 of 115 comparisons.

1 INTRODUCTION



(a) Cone-beam CT, 8 views, DeepSparse (b) Motion deblur, NAFNet (c) Super-resolution 4×, SwinIR

(d) Cartesian MRI 4×, U-Net (e) Cone-beam CT, pelvis, DeepSparse (f) Super-resolution 4×, EDSR

Figure 1: **Frozen output, adapted output and ground truth on four operators and five networks.** Boxed window enlarged in the corner. The examples are among the largest gains of their conditions. The means are in the tables, and Figure 6 shows the mean-gain scan.

A released reconstructor arrives at a new site with a prior learned from other data, and the only resource the site can add is its own measurements. Our primary setting is eight-view cone-beam CT, which reduces a full orbit by more than an order of magnitude. Released reconstructors for this setting are pretrained on large CT corpora (Lin et al., 2024; 2026; Lee et al., 2026), but they reach their published accuracy on a new dataset only after supervised finetuning on 600 to 800 ground-truth volumes (Lin et al., 2026, Tab. I), which a site that needs eight views cannot collect (Appendix A). Test-time adaptation instead minimises a measurement-consistency loss at inference and needs no ground truth (Darestani et al., 2022; Yaman et al., 2022). Such a loss compares a reconstruction only with what was measured, and under sparse acquisition most of the error lies in what was not, so the central question is how much of that error it can remove.

If the image itself is the optimisation variable, the answer is none. The loss is invariant to every component in the null space of the forward operator, so gradient descent leaves the null-space component wherever the initialisation put it. Equivariant imaging (Chen et al., 2021) circumvents

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.

this impossibility by adding a group invariance. In practice, however, every test-time method optimises through a network, whose decoder couples the range space and the null space (Heckel & Soltanolkotabi, 2020). Weight-space adaptation therefore already moves the null-space component. It does so slowly, and how far it can go has never been analysed.

We answer the question. How far a measurement-only loss moves the null-space component is bounded by the alignment $\cos\theta$ between its null-space update and the null-space error, and we prove it. That alignment is set by the pretraining and the target together, not by the scan, and we measure it across scans. The claim holds under three conditions, a fixed operator, a step budget, and a feed-forward reconstructor whose range-space component is imperfect. The alignment lives in the frozen layers between the variable and the output. Where those layers fit the target, an activation behind them is the better variable. Where only the weights can repair the mapping, they are, and a label-free test tells the two cases apart (Sections 4.4 and 5.5).

The answer gives the method. Every weight stays frozen and the per-scan objective has a single free variable, an internal activation with the frozen decoder behind it. On the cone-beam backbone that variable is either the internal 3D feature volume through which the whole volumetric pathway passes or the input projections. We call this routed data consistency (RDC, Figure 2). The method transfers unchanged to Cartesian MRI, super-resolution, deblurring and inpainting, through seven further reconstructors including a released EDSR (Lim et al., 2017), SwinIR (Liang et al., 2021) and NAFNet (Chen et al., 2022) (Figure 1 and Appendices V to Y).

Contributions. (1) The limit of that loss for a fixed operator at a given step budget. Whether the null-space component moves at all is decided by the variable's tangent kernel relative to the operator's range (Proposition 1). How much of the null-space error it can remove is bounded by $1-\sin\theta$ (Proposition 2). (2) An identity that ranks candidate variables from their decompositions alone (Corollary 3). Evaluated on ground-truth decompositions, it predicts the PSNR ordering on 110 of 115 comparisons across four operators. A selection procedure settles every remaining choice. It is label-free apart from a one-time rate sweep, and it was fixed before each new network was tried. (3) Routed data consistency, the method the limit points to (Section 4). On eight-view cone-beam CT, forty steps finish 1.72 dB above the range oracle, the bound that no fit of the image can exceed, and no weight-space method exceeds them on any of nine settings or on 18 clinical scans, where RDC leads the weights in 16.

2 Related work

Null-space limits and null-space priors. Chen et al. (2021) state the limitation within which this paper works. For a fixed operator with a non-trivial null space it is "fundamentally impossible to learn the signal model beyond the range space of $A^\top$ using only compressed samples", and Tachella et al. (2022; 2023) prove it. That result concerns learning a signal model from measurements. Here the model is already contained in the frozen reconstructor, and the measurements only select one reconstruction from it. Jin et al. (2026) ask which tangent directions of a generative prior an operator observes. Our alignment asks the same of the update produced through a frozen decoder. For an untrained network Heckel & Soltanolkotabi (2020) prove what content a prior places in the null space, the smoothest completion (Ulyanov et al., 2018; Heckel & Hand, 2019). Our procedure chooses the variable through which a pretrained decoder fills it. A second family fills the null-space component by design. It includes plug-and-play methods (Venkatakrishnan et al., 2013; Romano et al., 2017), null-space networks (Schwab et al., 2019; Angermann et al., 2023; Jácome et al., 2025), DDNM (Wang et al., 2023), unrolled networks that learn only that component (Chen et al., 2024), intermediate-layer optimisation (Daras et al., 2021), equivariant imaging (Chen et al., 2021; Schut et al., 2026) and self-supervised sparse-view CT (Gruber et al., 2024). None chooses the variable that a measurement-only loss should move, or measures how far it moves.

Test-time adaptation for reconstruction. Darestani et al. (2022) close most of the distribution-shift gap in accelerated MRI by minimising the data-consistency loss at inference, Yaman et al. (2022) split each scan's measurements into three subsets, and Zhang et al. (2025) adapt per patient through an implicit representation. All optimise the weights and need a stopping rule on held-out measurements. On sparse-view CT, Song et al. (2023) adapt a black-box model through its input and Zhao et al. (2024) through added adaptive layers. Both choose a variable, neither decomposes

the error, and they correspond to the input column of Table 1 and to Appendix G. Darestani et al. (2022) report "minimal to no" improvement under purely supervised pretraining and move data consistency into pretraining, which a downstream user cannot do. Outside reconstruction, test-time training (Sun et al., 2020) and TENT (Wang et al., 2021) adapt weights or normalisation statistics without a forward operator.

Adapting a restricted variable. Yu et al. (2023) reconstruct inside a subspace fixed in advance from physics. Our reachable set is the nonlinear image of a pretrained decoder. Fitting the input of a frozen generator is compressed sensing under a generative model (Bora et al., 2017), and intermediate-layer optimisation (Daras et al., 2021; Braure et al., 2025) descends through an unconditional prior, so that the optimisation is the whole reconstruction. Our variable is the internal volume of a conditional reconstructor that has already processed the measurements. Adapting an activation of a frozen model is established outside reconstruction (Chen et al., 2025; Luo et al., 2025; Wang et al., 2026), as is adapting one with labels (Lester et al., 2021; Li & Liang, 2021; Jia et al., 2022; Hu et al., 2022). We apply a measurement loss to such an activation, with no labels.

3 Background

Sparse-view reconstruction and the decomposition into range and null space. A scan measures $y = Ax$, with x the volume to be reconstructed and A a known linear operator, for cone-beam CT the X-ray transform of the scanner geometry over M views. Under sparse sampling A has a large null space. Let $P_{\mathcal{R}}$ and $P_{\mathcal{N}}$ denote the exact orthogonal projectors onto the range of $A^\top$ and the null space of A. Every volume then splits as $x = P_{\mathcal{R}}x + P_{\mathcal{N}}x$ into a range-space component, which the measurements constrain, and a null-space component, which they do not. We write $\hat{x}_0$ for the frozen reconstruction, $\hat{x}$ for the adapted one and x_{gt} for the ground truth. The *range oracle* $P_{\mathcal{R}}x_{\text{gt}} + P_{\mathcal{N}}\hat{x}_0$ combines the ground truth's range-space component with the frozen output's null-space component.

When the image itself is the optimisation variable, a free image, a measurement loss is blind to the null-space component. Writing $L(x) = \frac{1}{2}\|Ax - y\|_2^2$, we have for any n with $An = 0$

$$L(x+n) = L(x), \qquad \nabla_x L = A^\top(Ax - y) \in \text{range}(A^\top). \tag{1}$$

Gradient descent started at x_0, or any method whose iterates stay in the span of the gradients it has evaluated, therefore satisfies $P_{\mathcal{N}}(x_k - x_0) = 0$ at every step.

Lemma 1 (the range oracle is an upper bound on PSNR)**.** *Let $\mathcal{C} = \{\hat{x} : P_{\mathcal{N}}\hat{x} = P_{\mathcal{N}}\hat{x}_0\}$ be the set of reconstructions that keep the frozen output's null-space component. By equation 1 and the span argument that follows it, $\mathcal{C}$ contains every reconstruction obtained by measurement-loss descent of a free image from $\hat{x}_0$. Because the projectors are orthogonal, every $\hat{x} \in \mathcal{C}$ has $\|\hat{x} - x_{\text{gt}}\|^2 = \|P_{\mathcal{R}}(\hat{x} - x_{\text{gt}})\|^2 + \|P_{\mathcal{N}}(\hat{x}_0 - x_{\text{gt}})\|^2 \geq \|P_{\mathcal{N}}(\hat{x}_0 - x_{\text{gt}})\|^2$, with equality only at the range oracle, the unique PSNR maximiser over $\mathcal{C}$. A higher PSNR therefore requires moving the null-space component.*

For $y = Ax_{\text{gt}}$ the oracle is label-free, $\hat{x}_0 - A^+(A\hat{x}_0 - y)$, the projection of $\hat{x}_0$ onto $\{Ax = y\}$ (Schwab et al., 2019; Angermann et al., 2023) where free-image descent ends (Table 2).

Test-time data consistency through a frozen reconstructor. Test-time training for reconstruction (Darestani et al., 2022; Yaman et al., 2022) minimises the loss of equation 1 over the network weights at inference. Our primary backbone is DeepSparse (Lin et al., 2026), a released foundation model for sparse-view cone-beam CT. Its weights θ stay frozen. Its encoder E_θ produces the *fused 3D feature volume* $F_0 = E_\theta(y)$ of 128×32^3 entries together with pixel-aligned 2D features G queried from the projections. Its point decoder D_θ is the only path from the volumetric pathway to the output, so the reconstruction is $x(F) = D_\theta(F, G, \cdot)$ over the voxel grid. We adapt eight reconstructors in total, namely DeepSparse and a second cone-beam backbone (Liu et al., 2025), a U-Net on MRI, two U-Nets on inpainting with a pixel mask, EDSR, SwinIR and NAFNet. EDSR, SwinIR and the whole-image inpainting U-Net already fit their measurements at their training operator, so these three are adapted under an operator shift (Appendices W and X).

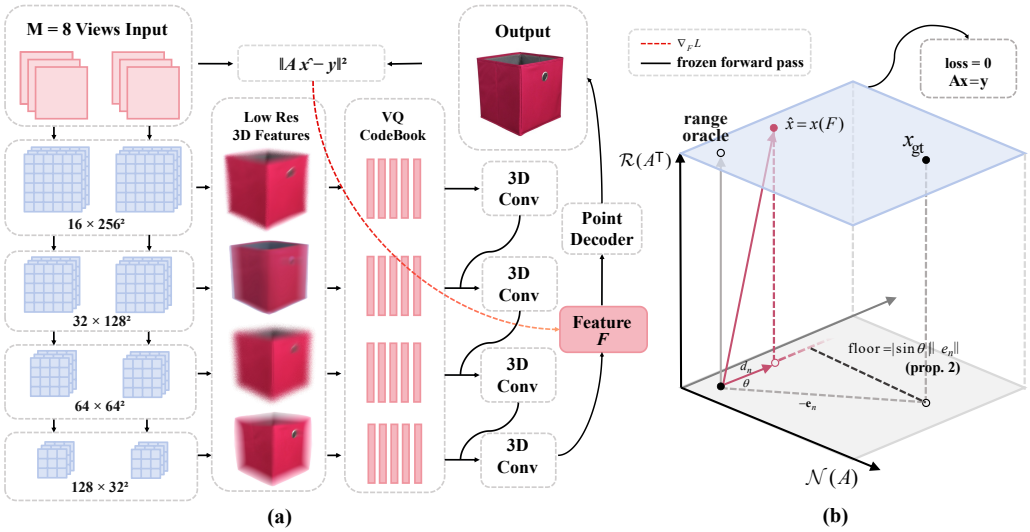


Figure 2: **Routed data consistency, and where its update lands.** (a) On the cone-beam backbone the measurement loss $\|A\hat{x}-y\|^2$ is backpropagated through the frozen point decoder to the feature volume F, the only free variable (Section 4.3). (b) A free image stops at the range oracle (Lemma 1). The activation update has a null-space component d_n at angle θ to the null-space error $-e_n$, and leaves at least $|\sin\theta|\,\|e_n\|$ of it (Proposition 2).

4 METHOD

Setting and assumptions. A site has three things. It has a released feed-forward reconstructor, whose weights stay frozen and whose training data the method never touches. It has the measurements y of one scan. And it has the forward operator A with its geometric calibration. A is linear with a non-trivial null space, the same for every scan, and the method takes it as exact. The reconstructor leaves a range-space residual on the new scans. That residual is all the loss has to work with, and the free variable reaches the image through the frozen decoder. The site has no ground truth for the target domain, at training, at test time or for early stopping. The one exception is the one-time step-size sweep of P4 (Section 5.1). The goal is a better reconstruction of that one scan within a fixed step budget, forty steps unless stated. The main results use noiseless measurements, $y = Ax_{\text{gt}}$, under which the range oracle is label-free (Lemma 1). Noise is treated separately (Appendices H and T). Unrolled networks with data-consistency cascades already fit their measurements, so they leave no residual for the loss (Appendix V). Measured clinical projections are treated in Section 5.2.

4.1 THE REACHABLE SET AND THE TANGENT KERNEL

The reconstruction is the output of a frozen network, $x = D_\theta(F, G)$, so a per-scan optimisation can only reach volumes that this network can produce from the variable it is allowed to change. This reachable set is the image of the frozen decoder, a curved set of dimension at most the variable's inside the S^3-dimensional space of volumes, and its shape was determined by pretraining rather than by A. There are three possible variables, the volume itself, an internal activation and the weights. For the volume the reachable set is the whole space and equation 1 holds as written. The other two reach the image through the frozen decoder. Write $x(v)$ for the reconstruction as a function of the free variable v. To first order one gradient step on v displaces the image by $\delta x = -\eta\, JJ^\top g$ with $J = \partial x(v)/\partial v$ and $g = \nabla_x L$.

Proposition 1 (null-space movement is determined by the frozen network)**.** *Let $K = JJ^\top$ be the tangent kernel of the free variable. Then $P_{\mathcal{N}}\,\delta x = 0$ for every residual if and only if K maps $\mathcal{R}(A^\top)$ into itself. A free image corresponds to $K = I$, which satisfies this condition and recovers equation 1. A variable that passes through a frozen decoder inherits a K fixed by pretraining, which need not.*

Proof. As the residual ranges over the measurement space, $g = A^\top(Ax-y)$ ranges over $\mathcal{R}(A^\top)$. If K maps $\mathcal{R}(A^\top)$ into itself, then $\delta x = -\eta Kg \in \mathcal{R}(A^\top) = \mathcal{N}(A)^\perp$, so $P_{\mathcal{N}}\,\delta x = 0$. Conversely, $P_{\mathcal{N}} Kg = 0$ for every $g \in \mathcal{R}(A^\top)$ means $Kg \in \mathcal{N}(A)^\perp = \mathcal{R}(A^\top)$. □

4.2 Alignment, relative length and the error identity

The null-space components of the update $d = \hat{x} - \hat{x}_0$ and of the error it is meant to remove are related by two scalars, an alignment and a relative length, and the fraction of null-space error that remains is a function of these two alone.

$$e_n = P_{\mathcal{N}}(\hat{x}_0 - x_{\mathrm{gt}}), \qquad d_n = P_{\mathcal{N}}(\hat{x} - \hat{x}_0), \qquad \cos\theta = \frac{\langle d_n, -e_n \rangle}{\|d_n\|\,\|e_n\|}, \qquad r = \frac{\|d_n\|}{\|e_n\|},$$
$$\rho^2 = \frac{\|e_n + d_n\|^2}{\|e_n\|^2} \;=\; 1 - 2r\cos\theta + r^2 \;=\; \sin^2\theta + (r - \cos\theta)^2. \tag{2}$$

The pair $(r, \cos\theta)$ measured for one variable on one setting is its decomposition, and Section 5.3 measures it (Figure 2b). Completing the square yields a bound.

Proposition 2 (an upper bound independent of the step length)**.** $\rho \geq |\sin\theta|$*, with equality at* $r = \cos\theta$ *when* $\cos\theta \geq 0$*. When* $\cos\theta < 0$*, no step length reduces the error at all. For a given alignment, the fraction of null-space error that a variable can remove is therefore at most* $1 - |\sin\theta|$*.*

Corollary 3 (the identity as a criterion)**.** ρ *is a function of the two scalars alone, so two candidate variables are ordered from their decompositions by* $v^\star = \arg\min_v \rho_v$*. Whether this null-space ordering also predicts the PSNR ordering is tested in Section 5.4.*

4.3 Routed data consistency

Proposition 2 makes the alignment the quantity to maximise, and Proposition 1 places it in the frozen layers between the variable and the image. Section 5.5 finds that it grows with the depth of those layers. We therefore leave θ alone, free one internal activation v of the frozen network, the deepest whose entries do not exceed the voxel count, and minimise the same loss per scan,

$$\hat{v} \;=\; \arg\min_{v} \; \big\|A\,x(v) - y\big\|_2^2, \qquad \hat{x} = x(\hat{v}), \tag{3}$$

starting from the value the encoder produces. The objective is the data-consistency loss of Section 3. The only change is the variable with respect to which it is differentiated. We call this routed data consistency (RDC, Figure 2a). On the cone-beam backbone, v is either the fused feature volume F or the input projections (Appendix G). Table 1 reports both, and F is the variable of the mechanism study. For the other networks the selection procedure of Section 4.4 picks the input. Freeing a second activation behind further frozen layers changes the result (Table 1, RDC$^+$, Appendix B).

4.4 Optimisation, budget and the selection procedure

A backward pass through all S^3 query points is infeasible, so the loss is backpropagated through a point-subsampled surrogate. At each step 2×10^5 query points are drawn afresh with the residual held fixed, which gives an unbiased gradient up to a scalar factor that the learning rate absorbs (Appendix B). We run 40 Adam steps at learning rate 0.02, which takes 202 s per scan on a single RTX 4090. On the noiseless grid, PSNR and SSIM rise monotonically to a plateau (Figure 7). Under the backbone's own noise level the curve peaks near step 50 and then declines, so there forty steps is the stopping rule (Appendix T). The per-scan optima can be distilled into one shared operator W at unchanged inference speed (Appendix B).

The label-free selection procedure. Given a released reconstructor and one scan, five choices remain. All but the rate are made without labels. *Variable.* Free the deepest activation whose free entries do not exceed the voxel count. On the cone-beam backbone this is the input projections, within 0.28 dB of the fused feature volume. The two have alignment 0.59 and 0.61 (Section 5.2), so the rule trades that margin for a variable that needs no access to the network's interior. On every image network it is the input. *Relative length.* Compute $r = \|P_{\mathcal{N}} d\| / \|e_n\|$ with $\|e_n\| = \|e_r\| \sqrt{s/(1-s)}$. The range-space error $e_r = P_{\mathcal{R}}(\hat{x}_0 - x_{\mathrm{gt}})$ has norm $\|A^+(A\hat{x}_0 - y)\|$ for $y = Ax_{\mathrm{gt}}$, so it is read from the residual, and $P_{\mathcal{N}}$ follows from the geometry. The share s of the frozen error's energy that lies in the null space is calibrated once on the backbone's own pretraining distribution, $s = 0.81$ on 10 AbdomenAtlas scans, and transferred. On the target domains it measures 78 to 81% (Appendix R). An r above 0.75 exceeds every alignment measured on the cone-beam backbones and is therefore past

the identity's minimum at $r = \cos\theta$. The variable is then narrowed to a coarser grid until it is not (Appendix U). *Budget.* Forty Adam steps and the final iterate, unchanged across backbones. *Rate.* One learning rate per backbone, set once on a labelled sweep (P4) and never revisited. *Activation or weights.* Fit both variants to all but $M/4$ of the views and keep the one with the smaller residual on the withheld views. This is correct on 18 of 18 scan-setting pairs here, including both reversals, the settings in which the weights end above every activation variable (Appendix R). It is a diagnostic pass. The chosen variant is then rerun on all M views, and that run is the one reported. Appendix S backtests the procedure on three backbones and applies it blind to two more.

What each label-free test covers. The procedure uses two label-free quantities, and each answers one question. The relative length r sizes a single variable. It detects one that has overshot the minimum of the identity (Appendix U). It does not detect a reversal, because every reversal we observe is decided by the alignment, and no measurement of the scan contains the alignment. Withheld views choose between two variants. They are the only label-free test here that catches the reversals at four views and under the tooth prior, a checkpoint whose mapping is wrong for the target (Appendix P). There the fitted residual picks the wrong variant on every scan (Table 22). They do not size a variable, because the withheld residual falls monotonically as a variable narrows.

5 Experiments

We ask three things of the experiments. Does RDC beat every label-free alternative at a fixed budget, on simulated and on clinical scans, and how far does it get without labels (Section 5.2)? Where does the gain lie, is the bound of Proposition 2 tight, and does the identity predict which variable wins (Sections 5.3 and 5.4)? And when both variants run to convergence on other operators, which ends higher, and why (Section 5.5)? Section 5.1 fixes the protocol.

5.1 Setup

Protocol. Five rules hold throughout. (P1) All numbers are our own unless typeset as *literature values*. Volumes are 256^3, projections are digitally reconstructed radiographs over a half circle in each dataset's geometry, and $M \in \{4, 6, 8, 10\}$. The clinical row of Table 1 uses measured projections (Section 5.2). (P2) M=8 cells average every held-out scan of the dataset, the other rows the first ten (Appendix B). Compared variants are scored on the scans both completed. (P3) Compared variants differ only in the free variable (Section 4) and report the final iterate, at forty steps unless a table says otherwise. (P4) Each variant runs at a single global step size, set once by PSNR gain on ten labelled lung scans and confirmed on twenty more. The same rate is then applied unchanged to pelvis, abdomen and every other setting, so those rows use no target label at any stage. A $4\times$ change in rate moves the result by 0.11 dB (Appendix B). (P5) Ground truth enters only the one-time rate setting of P4 and the diagnostics (Section 4). Three domains are evaluated against one frozen backbone pretrained on AbdomenAtlas (Li et al., 2024), namely **lung** (LUNA16 (Setio et al., 2017)), **pelvis** (PENGWIN (Sang et al., 2026)) and **abdomen**, its own pretraining distribution.

Baselines. The central baseline is test-time training (TTT), in which Darestani et al. (2022) optimise exactly our objective over the network weights. The text calls it the weights. We run it under our protocol at its own step size, swept on the same ten scans as ours (Appendix B). Its published form holds two views out for early stopping and is 0.03 to 0.56 dB lower. Equivariant imaging (EI) (Chen et al., 2021) adds an invariance of the imaged distribution and is applied to the weights, its invariance weight swept on the same scans (Appendix B). Running the same objective on a randomly initialised backbone (Ulyanov et al., 2018), the untrained column, isolates the contribution of pretraining. RDC is ahead of the per-scan implicit field (Zha et al., 2022), the competing method at our label budget, on 65 of 65 lung scans and by 3.60 dB. The field reaches 24.12 dB, -0.77 dB relative to the frozen model, and DeepSparse's supervised finetune bounds what labels can achieve (Appendix B).

Generative priors. RDC finishes 4.30 dB above a generative prior trained on a disjoint corpus (51 M parameters, sampled as Chung et al. (2023b) do at 840 operator applications) over 30 scans. The prior itself ends -1.52 dB relative to the frozen model. DDNM (Wang et al., 2023) with the same prior agrees to within +0.01 dB (Appendix F).

			Optimises weights		Optimises activations			
Setting	n	Frozen	EI	TTT	untrained	**RDC** (in)	**RDC** (F)	RDC$^+$
No column uses target-domain ground truth in the adaptation. Each rate is set once (P4).								
Domain transfer at M=8								
Lung (LUNA16)	66	24.89 | 80.7	26.93 | 82.8	27.14 | 83.2	22.64 | 70.7	27.55 | 82.2	**27.72** | **83.7**	28.02 | 84.7
Pelvis (PENGWIN)	30	24.29 | 78.3	26.28 | 80.9	26.57 | 81.4	23.21 | 70.4	27.08 | 80.2	**27.17** | **82.1**	27.50 | 83.1
Abdomen (source)	21	26.56 | 76.3	28.04 | 80.4	28.31 | 80.8	20.91 | 49.8	**29.24** | **81.6**	28.97 | 80.8	29.15 | 81.3
View count, lung								
M=4	10	21.12 | 69.0	24.25 | 77.9	24.57 | 76.5	20.88 | 69.3	**24.65** | **72.3**	24.09 | 75.1	24.32 | 76.1
M=6	10	23.95 | 78.4	25.98 | 80.9	26.35 | 81.9	22.30 | 71.9	**26.77** | **80.2**	26.68 | 81.8	26.90 | 82.9
M=10	10	25.60 | 83.0	27.33 | 84.3	27.56 | 84.9	23.38 | 74.2	28.07 | 84.5	**28.54** | **86.3**	28.89 | 87.3
Jittered gantry phase, M=8								
Lung	10	25.42 | 82.8	27.34 | 84.7	27.52 | 84.9	23.07 | 73.8	28.08 | 84.2	**28.22** | **85.7**	28.47 | 86.5
Pelvis	10	24.42 | 78.0	26.30 | 80.5	26.98 | 81.3	23.19 | 70.0	**27.37** | **80.5**	27.33 | 81.7	27.64 | 82.9
Abdomen	10	26.93 | 76.7	28.25 | 80.8	28.65 | 81.3	21.00 | 50.2	**29.43** | **81.7**	29.25 | 81.3	29.32 | 81.6
Mean gain over frozen (dB)		–	+1.95	+2.27	-2.51	+2.78	+2.75	+3.00
Real clinical scans, M=8, inside the field of view								
Pelvis (COBRA2026)	18	21.56	21.48	22.20	9.78	22.31	**22.63**	22.26

Table 1: **Label-free adaptation across nine acquisition settings and on real clinical scans.** PSNR | SSIM$\times 10^{-2}$, best per row bold and second underlined. RDC frees the input projections (in) or the feature volume F. RDC$^+$ frees a second tensor whose entries exceed the voxel count and is outside the ranking. The input projections' rate was swept on one lung scan (Appendix G). The mean row is over the nine DRR settings. The clinical row is PSNR inside the field of view, every variant with a per-view gain in the loss (Section 5.2).

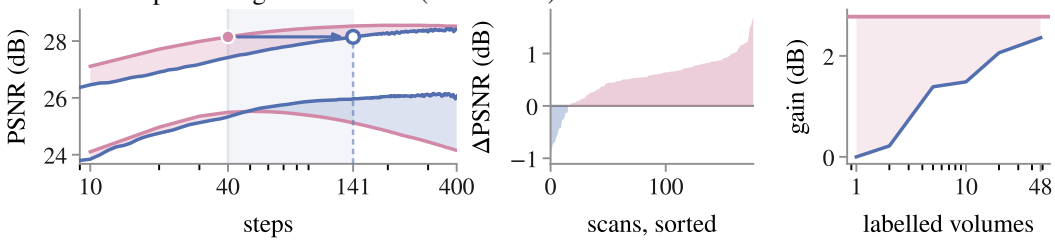


Figure 3: **The forty-step comparison in context.** Rose is RDC, blue the weights. The left panel follows both to 400 steps on six lung volumes at M=8, noiseless (upper) and at the backbone's noise level (lower), with ticks at our budget and at the step the weights need to match it. The middle panel shows the per-scan difference at forty steps over Table 1, positive on 161 of 177 scans, median +0.59 dB. The right panel is a budget-matched supervised sweep.

5.2 Cone-beam CT

Main result. RDC has the highest PSNR on every one of the nine settings. One of its two variables tops each row. That includes the model's own pretraining distribution. A finetune of the same backbone at the same per-scan optimisation budget gains +2.36 dB with 48 labelled volumes, below the +2.78 that the feature volume gains with none. Without labels the feature volume closes 47% of the gap to the fully supervised in-domain finetune at 30.92 dB (Appendices D and E). The one reversal on this backbone's grid is at four views, where the feature volume alone is below the weights. Section 5.4 predicts its sign, and RDC's input projections match the weights there (Appendix G). The procedure's own pick, the input projections, removes 19% of the null-space error at alignment 0.59 on the same scans, within a point of the feature volume's 20% at 0.61 (Appendix G). Per scan, the feature volume beats equivariant imaging on 63 of 66 scans (Wilcoxon $p < 0.001$) and improves on the frozen model on every scan (error maps in Appendix C).

Downstream tasks, robustness and budget. A frozen nodule classifier and a frozen segmentor, both trained on ground truth, are applied to every reconstruction. RDC recovers 55% of the AUC that the frozen model loses on LUNA16 and 27% of the Dice it loses on PENGWIN, four times as much as either weight-space variant (Appendices K and L). The gain varies by about half a decibel across view counts and gantry phases, keeps its margin over the range oracle up to 20 views (Appendix I), holds across the backbone's four anatomy checkpoints, and does not fit Poisson noise within the

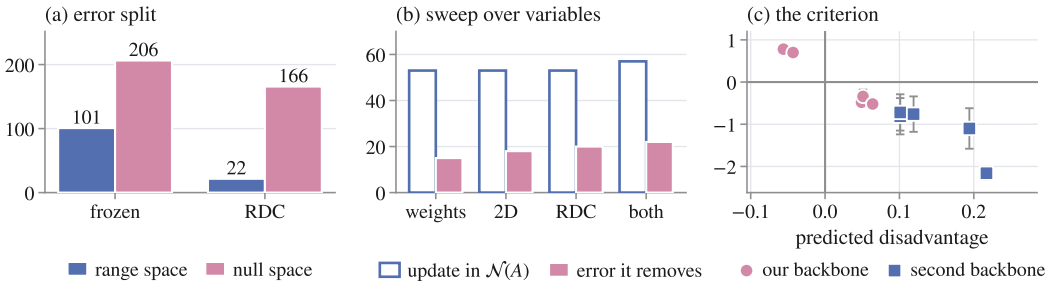


Figure 4: **The three measurements of this section.** The axes are error norm, per cent, and margin over the oracle in dB. (b) compares four parameterisations of one loss. (c) applies the criterion of Corollary 3, negative where the weights are the better variable. Exact values in Appendix N.

budget (Appendix H). The weights need 5.7× the wall clock of RDC to catch up (Figure 3), and RDC never diverges over the range of step sizes in which they do (Appendix T).

Real clinical scans. On 18 clinical pelvic scans of COBRA2026 (Thummerer et al., 2026), the last row of Table 1 orders the adaptation methods as the DRR rows do. RDC gains +1.06 dB inside the field of view, against +0.64 for the weights. It improves on the frozen model in 17 of them, leads the weights in 16 (Wilcoxon $p = 0.0002$) and equivariant imaging in all 18. Real projections carry a per-view intensity error that the operator does not model, so every variant fits them with a label-free per-view gain in the loss (Appendix J.2).

5.3 Where the gain lies

Movement in the null space. We apply the projectors exactly, by conjugate gradient, and compare against the range oracle (Lemma 1). Adapting the feature volume removes 20% of the null-space error over 30 scans and ends 1.72 dB *above* that bound (Figure 4a).

The sweep over variables differs in direction, not size. Figure 4b varies only the variable. Every one places 53 to 57% of its update energy in the null space, yet the null-space error it removes ranges from 15 to 22%, and every variant exceeds the range oracle (Appendix N). This is why variables are ranked by the direction term $\cos\theta$ of equation 2 rather than by step length (Proposition 2).

Alignment and relative length. Over 6 scans the null-space update has relative length $r = 0.51$ and alignment $\cos\theta = 0.61$. The residual ratio ρ measured directly is 0.80, the value the identity predicts from them. The alignment is a property of the setting, not of the scan. It is set by the pretraining and the target together, 0.61 on lung and 0.60 on pelvis for this backbone, and 0.71 under the tooth prior on the same lung scans (Table 15). Across scans its standard deviation is 0.02 here, 0.04 at four views and 0.08 over 30 pelvis scans, and the ordering of the two variables by alignment holds scan by scan, on 58 of 62 paired scans over six settings (Table 15). The bound of Proposition 2 is tight. Figure 9 (Appendix N) places all 174 decompositions in the $(r, \cos\theta)$ plane. In every one the measured ρ lies at or above $|\sin\theta|$ and below 1, agreeing with the identity to 8×10^{-5}, and the runs reach a median 98% of the bound, 97% on these six scans (Table 14). The length term is quadratic, so it stays negligible unless the variable is badly sized.

5.4 The identity as a criterion

Criterion. Evaluated on ground-truth decompositions, the error identity predicts which variable wins on every one of the 7 settings decomposed with both variants. These reach beyond Table 1 and include both outcomes, settings where RDC wins and reversals where the weights win (Table 16). Figure 4c applies Corollary 3 to all 10 rows, the 7 settings with the second backbone at each of its four step budgets, and gives the correct sign on every one. Ordering by $\cos\theta$ alone also gets every row right, so no reversal is caused by the step length (Table 14).

On deblurring through a released NAFNet, ρ ranks the weights first on 4 of 5 conditions, and there they are indeed the better variable (Appendix Y). Pooled over the six networks with a non-zero alignment, the identity is correct on 110 of 115 pairs (Table 24).

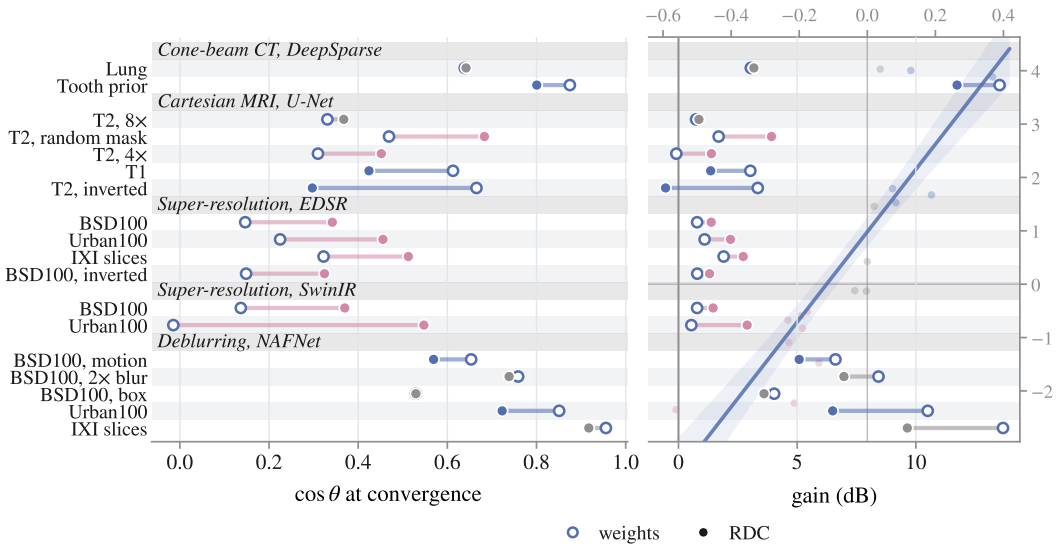


Figure 5: **Where the variants converge to the same alignment, and where the weights depart.** Every condition at 400 steps. Hollow is the weights, filled is RDC, coloured by the verdict (meet within 0.05). The right panel plots the same rows as gain gap (right scale) against alignment gap (top scale), with a least-squares fit and one standard error, Pearson 0.84 (Table 25).

5.5 Convergence across operators

Which variant ends higher. Across the four operators of Figure 5, the kind of shift decides. Where the frozen network's mapping fits the target, weight-space descent spends the residual on restoring consistency inside the range space, while the frozen layers behind the RDC variable carry the update into the null space. The weights then converge to RDC's alignment or below it. Where the shift changes what the network has to map, a contrast it never saw or a prior trained on the wrong anatomy, only the weights can change the mapping, and they end above RDC at every budget.

The alignment grows with frozen depth. On the U-Net, EDSR, SwinIR and NAFNet the alignment rises with the depth of frozen layers behind the variable, the input highest. On DeepSparse the two RDC variables are within 0.02 of each other and both above the weights (Table 24). An activation with only a head, an upsampler or a bottleneck's skip connections behind it does not move the null-space error, and on the pixel mask, where the range-space and null-space components share no support in the image domain, no variable has a non-zero alignment (Appendix X).

6 Conclusion and limitations

A measurement-only loss is blind to the null space only when its variable is the image. Routed through the frozen layers of a released reconstructor, it moves the null-space component as far as an alignment fixed by the pretraining and the target allows. A site can therefore adapt a released reconstructor to its own scans, clinical ones included, with no ground truth in the adaptation itself. Only a one-time rate sweep per backbone uses labels (P4, Section 4.4).

Limitations. Every claim is made for a fixed operator, at a given step budget, through a feed-forward reconstructor whose range-space component is imperfect (Section 4). At the backbone's own noise level the forty-step advantage is +0.16 dB and -1.90 at 400 steps, where RDC must stop at forty and the weights need not (Appendix T). The operators after the first are two-dimensional and simulated, and the withheld-view selection is tested on one backbone (Appendix Z).

Reproducibility statement

The supplementary material contains the full pipeline, the patches applied to the two released backbones, the per-dataset preprocessing, the drivers for every experiment including the sweeps and controls, and the scripts that produce every table and figure. Every released backbone is used as

published. The networks trained here are trained from public data with their training scripts in the supplement, and all datasets are public. Numbers taken from the literature are typeset *like this*. The per-scan adaptation is deterministic given the scan, and Appendix B reports the sensitivity to the step size, the sampling seed and the number of sampled points, together with the compute cost of every study.

ETHICS STATEMENT

All experiments use public, de-identified datasets under their respective licences, and no new patient data was collected. The method reconstructs from an order of magnitude fewer projections, and lower dose is its motivation, but the content of the null-space component is supplied entirely by a learned prior and not by the measurements. The result is a synthesis constrained to a pretrained manifold, and a prior far from the target still fills that component confidently (Appendix P). The downstream studies quantify when this matters at the task level and where the grid resolution cannot express the question (Appendix M). They do not justify clinical use, which would require prospective validation of exactly these failure modes. This work is a research method and not a cleared reconstruction product.

AI USE STATEMENT

Of the tasks for which the ICLR 2027 AI policy requires disclosure, we used generative AI tools for one, writing code, and only for the scripts that render the tables and figures from stored results. The text of this paper was written by the authors in English. We did not use generative AI tools for any other task on that list. The theoretical framework, the alignment identity and its derivation, the hypotheses, the design of every experiment and control, and the interpretation of every result are the authors' own. No dataset here was produced by a generative model. The simulated measurements are forward projections of public volumes under the operators of Section 5. Qualitative and thematic data analysis is not applicable to this work. Additionally, we used a large language model for one task on the recommended-disclosure list, namely editing author-written English to improve readability, at the level of grammar, wording and sentence structure.

We have reviewed all AI-assisted work. An author read every edited sentence against the original and against the result it reports, and rewrote those that had drifted. The generated rendering scripts were read by an author before they were run, and every table and figure was checked against the stored results it renders. Every number and every figure in this paper is produced by a complete run of the pipeline on the data described, so no result reported here rests on code that was not executed end to end. The supplementary material contains that pipeline, and the reproducibility statement above says where each number is generated. We take responsibility for the final content of this work, including its text, claims and artifacts.

REFERENCES

Jonas Adler and Ozan Öktem. Learned primal-dual reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1322–1332, 2018.

Anders H Andersen and Avinash C Kak. Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. *Ultrasonic Imaging*, 6(1):81–94, 1984.

Christoph Angermann, Simon Göppel, and Markus Haltmeier. Uncertainty-aware null space networks for data-consistent image reconstruction. *arXiv:2304.06955*, 2023.

Biomedical Image Analysis Group, Imperial College London. IXI dataset. `https://brain-development.org/ixi-dataset/`, 2008. CC BY-SA 3.0.

Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G. Dimakis. Compressed sensing using generative models. In *International Conference on Machine Learning (ICML)*, 2017.

Leise Borg, Jürgen Frikel, Jakob Sauer Jørgensen, and Eric Todd Quinto. Analyzing reconstruction artifacts from arbitrary incomplete X-ray CT data. *SIAM Journal on Imaging Sciences*, 11(4): 2786–2814, 2018.

Thomas Braure, Delphine Lazaro, David Hateau, Vincent Brandon, and Kévin Ginsburger. Conditioning generative latent optimization for sparse-view CT image reconstruction. *Journal of Medical Imaging*, 12(2):024004, 2025.

Tatiana A Bubba, Gitta Kutyniok, Matti Lassas, Maximilian März, Wojciech Samek, Samuli Siltanen, and Vignesh Srinivasan. Learning the invisible: a hybrid deep learning-shearlet framework for limited angle computed tomography. *Inverse Problems*, 35(6):064002, 2019.

Dongdong Chen, Julián Tachella, and Mike E. Davies. Equivariant imaging: Learning beyond the range space. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

Jiacheng Chen, Jiawei Jiang, Fei Wu, and Jianwei Zheng. Null space matters: Range-null decomposition for consistent multi-contrast MRI reconstruction. In *AAAI*, 2024.

Kecheng Chen, Xinyu Luo, Tiexin Qin, Jie Liu, Hui Liu, Victor Ho Fun Lee, Hong Yan, and Haoliang Li. Test-time adaptation for foundation medical segmentation model without parametric updates. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025.

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European Conference on Computer Vision (ECCV)*, 2022.

Hyungjin Chung, Jeongsol Kim, Michael T McCann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations (ICLR)*, 2023a.

Hyungjin Chung, Dohoon Ryu, Michael T McCann, Marc L Klasky, and Jong Chul Ye. Solving 3d inverse problems using pre-trained 2d diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023b.

Giannis Daras, Joseph Dean, Ajil Jalal, and Alexandros G. Dimakis. Intermediate layer optimization for inverse problems using deep generative models. In *International Conference on Machine Learning (ICML)*, 2021.

Mohammad Zalbagi Darestani, Jiayu Liu, and Reinhard Heckel. Test-time training can close the natural distribution shift performance gap in deep learning based compressed sensing. In *International Conference on Machine Learning (ICML)*, 2022.

L. A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone-beam algorithm. *Journal of the Optical Society of America A*, 1(6):612–619, 1984.

Nadja Gruber, Johannes Schwab, Elke Gizewski, and Markus Haltmeier. Sparse2inverse: Self-supervised inversion of sparse-view CT data. In *International Conference on Image Formation in X-Ray Computed Tomography (CT Meeting)*, 2024. arXiv:2402.16921.

Reinhard Heckel and Paul Hand. Deep decoder: Concise image representations from untrained non-convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2019.

Reinhard Heckel and Mahdi Soltanolkotabi. Compressive sensing with un-trained neural networks: Gradient descent finds a smooth approximation. In *ICML*, 2020.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations (ICLR)*, 2022.

Román Jácome, Romario Gualdrón-Hurtado, Leon Suarez, and Henry Arguello. NPN: Non-linear projections of the null-space for imaging inverse problems. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. Visual prompt tuning. In *European Conference on Computer Vision (ECCV)*, 2022.

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.

Pengfei Jin, Na Li, and Quanzheng Li. Measurement geometry and design for trustworthy generative inverse problems. *arXiv:2606.02309*, 2026.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Seungryong Lee, Woojeong Baek, Joosang Lee, and Eunbyung Park. ILV: Iterative latent volumes for fast and accurate sparse-view CT reconstruction. *arXiv:2603.14915*, 2026.

Brian Lester, Rami Al-Rfou, and Noah Constant. The power of scale for parameter-efficient prompt tuning. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 3045–3059, 2021.

Haodong Li, Shuo Han, Haiyang Mao, Yu Shi, Changsheng Fang, Jianjia Zhang, Weiwen Wu, and Hengyong Yu. Cross-distribution diffusion priors-driven iterative reconstruction for sparse-view CT. *IEEE Transactions on Medical Imaging*, 45(7):3878–3894, 2026.

Wenxuan Li, Chongyu Qu, Xiaoxi Chen, Pedro R. A. S. Bassi, et al. AbdomenAtlas: A large-scale, detailed-annotated, & multi-center dataset for efficient transfer learning and open algorithmic benchmarking. *Medical Image Analysis*, 2024.

Xiang Lisa Li and Percy Liang. Prefix-tuning: Optimizing continuous prompts for generation. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 4582–4597, 2021.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using swin transformer. In *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 1833–1844, 2021.

Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPR Workshops*, 2017.

Yiqun Lin, Jiewen Yang, Hualiang Wang, Xinpeng Ding, Wei Zhao, and Xiaomeng Li. C^2RV: Cross-regional and cross-view learning for sparse-view CBCT reconstruction. In *CVPR*, 2024.

Yiqun Lin, Jixiang Chen, Hualiang Wang, Jiewen Yang, Jiarong Guo, Yi Zhang, and Xiaomeng Li. Deepsparse: A foundation model for sparse-view CBCT reconstruction. *IEEE Transactions on Medical Imaging*, 2026. arXiv:2505.02628.

Zhentao Liu, Yu Fang, Changjian Li, Hao Wu, Yuan Liu, Dinggang Shen, and Zhiming Cui. Geometry-aware attenuation learning for sparse-view CBCT reconstruction. *IEEE Transactions on Medical Imaging*, 44(2):1083–1097, 2025.

Xinyu Luo, Jie Liu, Kecheng Chen, Junyi Yang, Bo Ding, Arindam Basu, and Haoliang Li. Efficient test-time adaptation through latent subspace coefficients search. *arXiv:2510.11068*, 2025.

Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (RED). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017.

Yudi Sang, Yanzhen Liu, Sutuke Yibulayimu, et al. Benchmark of segmentation techniques for pelvic fracture in CT and X-ray: Summary of the PENGWIN 2024 challenge. *IEEE Transactions on Medical Imaging*, 45(5):2212–2228, 2026.

Dirk Elias Schut, Adriaan Graas, Robert van Liere, and Tristan van Leeuwen. Equivariance2Inverse: A practical self-supervised CT reconstruction method benchmarked on real, limited-angle, and blurred data. *IEEE Transactions on Computational Imaging*, 12:800–811, 2026.

Johannes Schwab, Stephan Antholzer, and Markus Haltmeier. Deep null space learning for inverse problems: convergence analysis and rates. *Inverse Problems*, 35(2):025008, 2019.

Arnaud Arindra Adiyoso Setio, Alberto Traverso, Thomas de Bel, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *Medical Image Analysis*, 42:1–13, 2017.

Liyue Shen, John Pauly, and Lei Xing. NeRP: Implicit neural representation learning with prior embedding for sparsely sampled image reconstruction. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

Eugene Siow. super-image: EDSR base 4× model card. `https://huggingface.co/eugenesiow/edsr-base`, 2021.

Bowen Song, Liyue Shen, and Lei Xing. PINER: Prior-informed implicit neural representation learning for test-time adaptation in sparse-view CT reconstruction. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1928–1938, 2023.

Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *International Conference on Learning Representations (ICLR)*, 2022.

Anuroop Sriram, Jure Žbontar, Tullie Murrell, Aaron Defazio, C. Lawrence Zitnick, Nafissa Yakubova, Florian Knoll, and Patricia Johnson. End-to-end variational networks for accelerated MRI reconstruction. In *MICCAI*, 2020.

Yu Sun, Xiaolong Wang, Zhuang Liu, John Miller, Alexei A Efros, and Moritz Hardt. Test-time training with self-supervision for generalization under distribution shifts. In *International Conference on Machine Learning (ICML)*, 2020.

Julián Tachella, Dongdong Chen, and Mike Davies. Unsupervised learning from incomplete measurements for inverse problems. In *NeurIPS*, 2022.

Julián Tachella, Dongdong Chen, and Mike Davies. Sensing theorems for unsupervised learning in linear inverse problems. *Journal of Machine Learning Research*, 24, 2023.

Adrian Thummerer, Simon Rit, Florian Kamp, Matteo Maspero, Martijn P. W. Intven, Thomas G. Boné, Christopher Kurz, Guillaume Landry, Thomas Baudier, Mustafa Kadhim, Julius Arnold, Michael Rauter, Barbara Knäusl, and Lukas Zimmermann. COBRA2026: a large-scale multicenter pelvic cone-beam computed tomography projection dataset. *arXiv preprint arXiv:2607.20037*, 2026.

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *CVPR*, 2018.

Singanallur V. Venkatakrishnan, Charles A. Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2013.

Chen Wang, Hao Tan, Wang Yifan, Zhiqin Chen, Yuheng Liu, Kalyan Sunkavalli, Sai Bi, Lingjie Liu, and Yiwei Hu. tttLRM: Test-time training for long context and autoregressive 3D reconstruction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations (ICLR)*, 2021.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *International Conference on Learning Representations (ICLR)*, 2023.

Burhaneddin Yaman, Seyed Amir Hossein Hosseini, and Mehmet Akçakaya. Zero-shot self-supervised learning for MRI reconstruction. In *International Conference on Learning Representations (ICLR)*, 2022.

Heng Yu, Yamin Arefeen, and Berkin Bilgic. SubZero: Subspace zero-shot MRI reconstruction. *arXiv:2311.17251*, 2023.

Jure Žbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J. Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, et al. fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv preprint arXiv:1811.08839*, 2018.

Ruyi Zha, Yanhao Zhang, and Hongdong Li. NAF: Neural attenuation fields for sparse-view CBCT reconstruction. In *MICCAI*, 2022.

Lipei Zhang, Rui Sun, Zhongying Deng, Yanqi Cheng, Carola-Bibiane Schönlieb, and Angelica I. Aviles-Rivero. D2SA: Dual-stage distribution and slice adaptation for efficient test-time adaptation in MRI reconstruction. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Zhenkai Zhang, Markus Hiller, Krista A. Ehinger, and Tom Drummond. From sparse x-rays to 3D CT: Training-free reconstruction with diffusion priors. *arXiv:2606.20763*, 2026.

Yutian Zhao, Tianjing Zhang, and Hui Ji. Test-time model adaptation for image reconstruction using self-supervised adaptive layers. In *European Conference on Computer Vision (ECCV)*, 2024.

A THE TWO WAYS TO CUT DOSE, AND THE RELEASED ROUTE TO A NEW DATASET

Views against flux. Dose can be reduced along two axes: by acquiring fewer projections or by lowering the flux of each projection. The two reductions fail in different ways. Low flux leaves the sampling intact and adds noise; because this noise enters through y, it is confined to the range space. Fewer projections shrink the range space itself. The streaks this produces are missing information rather than corrupted information, so no denoiser can restore them. Sparse acquisition is also what interventional and image-guided settings require: a full orbit takes time that the procedure does not have, and the scan is repeated over a course of treatment (Lin et al., 2026). We therefore work at $M \in \{4, 6, 8, 10\}$, which is the regime in which $P_{\mathcal{N}}e$ dominates the error and in which the mechanism studied in this paper has an effect. Section 5 also perturbs the other axis, by adding Poisson statistics to the projections.

The released finetuning procedure. The procedure that DeepSparse prescribes for a new dataset is a two-step *supervised* finetune. Step 1 adapts all parameters on the target set, and step 2 adjusts for the view count and trains a denoising layer. Both steps optimise a reconstruction loss against ground-truth CT and use 600–800 target volumes for 200 epochs each (Lin et al., 2026, Tab. I). Without this finetune, the released foundation model reaches 24.89 dB on LUNA16 at M=8, against *31.14* for the full two-step procedure.

B IMPLEMENTATION DETAILS

The second activation. RDC$^+$ optimises a second activation tensor in addition to F. It adds +0.25 dB on average over the feature volume alone and is never worse on any of the 9 rows (+0.08 to +0.35). Its number of free entries exceeds the voxel count, which is why Table 1 keeps it outside the ranking.

The upper bound on the distilled adapter is set by how scan-specific the per-scan optima are. A student network whose only input is the frozen feature volume recovers barely a quarter of the teacher's gain, regardless of how large it is made. Adding the backprojected measurement residual $A^\top(Ax_0 - y)$ as a second input doubles that share while using a fraction of the capacity. The pairwise cosine similarity between per-scan feature updates averages 0.17, which indicates that most of what the optimisation finds is specific to the scan and cannot be predicted from any single forward pass.

The distilled operator of Section 4 is fitted by least squares over K unlabelled target scans,

$$\min_W \sum_{i=1}^{K} \left\| W\left(F_0^{(i)}\right) - \left(\hat{F}^{(i)} - F_0^{(i)}\right) \right\|^2, \qquad F \leftarrow F_0 + W(F_0), \tag{4}$$

where $\hat{F}^{(i)}$ is the per-scan optimum. Wherever the distilled adapter is quoted, its value is a leave-one-out estimate over eight scans.

Data. Volumes are resampled to 1.6 mm isotropic and centre-cropped or padded to 256^3. Intensities are clipped to $[-1000, 1024]$ HU and mapped linearly to $[0,1]$. This is the convention under which the released backbone was trained, and it is the reason error maps are converted to HU by a factor of 2024. Projections are DRRs generated with TIGRE over a half circle using each dataset's scanner geometry. LUNA16 uses the 738/50/100 split of Lin et al. (2026). We preprocess the seven subsets served by one of its two Zenodo records, which limits the pool of volumes available to us. Within each row of every table, all columns are averaged over the same scans. The headline M=8 rows use every preprocessed test volume of the corresponding dataset: sixty-six for lung, thirty for pelvis and twenty-one for the abdomen subset. The remaining rows, which vary the view count or jitter the gantry phase, are averaged over the first ten test volumes of their dataset. Sample sizes therefore differ between rows but never within a row. PENGWIN and the AbdomenAtlas subset are preprocessed identically.

The surrogate. A backward pass through all S^3 query points is infeasible, so we use the standard two-pass surrogate. Pass A evaluates the full volume without gradients and forms g, normalised by its mean absolute value. Pass B backpropagates a point-subsampled surrogate,

$$\widetilde{L}(F) = \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} g_i \, x_i(F), \qquad \mathbb{E}\big[\nabla_F \widetilde{L}\big] = S^{-3} \nabla_F L, \tag{5}$$

over a uniformly drawn subset $\mathcal{B}$ of the query points, with g held constant. The estimate is therefore unbiased up to a scalar factor, which the learning rate absorbs. The noise of the estimate does not move the null-space component (Appendix U).

RDC. We run Adam on the fused feature volume alone for 40 steps at learning rate 0.02, with $|\mathcal{B}| = 2 \times 10^5$ query points resampled at every step and the gradient g recomputed at every step from a full-volume forward pass. The chunk size of that forward pass has no effect on the result, both by construction and in measurement: 40 steps take 202, 203 and 204 s at chunk sizes of 2, 4 and 8×10^5 points, and the resulting PSNR agrees to four decimals. Every run is deterministic given the scan.

Sensitivity. All sensitivity runs use the same ten scans. Quadrupling the step size lowers PSNR by 0.11 dB, and quartering it lowers PSNR by 0.77 (both sweeps are given in Table 28). Three seeds of the random point subsets agree to 0.003 dB. The surrogate gradient is close to converged in the number of sampled points: with 50k points the PSNR is 0.23 dB lower, whereas 800k points, four times our setting, yield only a further 0.10.

Baselines. The weight-space variant optimises the identical objective over all backbone parameters, and equivariant imaging adds its invariance term, with rotations up to 30° at unit weight, on the same parameters. That weight is swept over 0.1, 0.3, 1 and 3 on the 10 scans the learning rate was chosen on. It is not at its optimum: α=0.1 reaches 27.37 dB against 27.13 at unit weight, +0.24 dB and better on 10 of 10 scans, while α=3 is -0.31. The sweep is monotone over the range, so the reported cells sit on its falling side. At its own best weight the arm reaches 27.37 against 28.12 for RDC on those scans, +0.75 dB on 10 of 10, so the ordering is the one the table reports. Both baselines run under our exact protocol: every measured view enters the loss, the same 40 steps are taken, and the final iterate is scored. Their learning rate is 3×10^{-5}, which was swept under that protocol over $\{10^{-5}, 3\times10^{-5}, 10^{-4}\}$ and lies in the interior of that range. Batch normalisation runs in inference mode in every variant. The weight-space variant therefore updates the affine parameters and leaves the running statistics at their pretraining values; re-estimating those statistics on the target data, as TENT does (Wang et al., 2021), is not run here. The published form of these baselines instead holds two of the M measured views out of the loss and selects the stopping step on them. Under that form the two baselines reach 27.08 and 26.89 on the main lung row, against 27.14 and 26.93 under the matched protocol. Matching the step budget therefore improves the weight-space variant by +0.03 to +0.56 dB across the nine rows, and equivariant imaging by -0.20 to +0.21. Only the robustness table keeps the published protocol, because its perturbation conditions were run under it. At four views the optimum of the matched weight-space variant remains 3×10^{-5}, whereas RDC's moves to 0.04. At both optima the weight-space variant is better by 0.40 ± 0.10 dB on 9 of ten scans, so the reversal at four views is not an artefact of the learning rate. The untrained control is our objective on the same variable with the backbone's weights randomly initialised. The supervised reference is the authors' released in-domain checkpoint run in inference only. A NAF-style hash-grid field (Zha et al., 2022) is fitted per scan for 3000 steps on our grid, geometry and view budget, on every scan of the lung row. The renderer's convention is searched jointly over all M views before each fit and

the run is refused if any single view disagrees with the measured projections by more than 3%; the accepted convention is the same on every scan and its worst view has a median error of 2.0%. Every fit converges, the median loss falling by a factor of 61×10^3, at 314 s per scan. It reaches 24.12 dB over 65 scans, -0.77 dB relative to the frozen model and 3.60 dB below RDC, which is ahead on 65 of 65 scans. Its authors report *20.51* at their own geometry and view budget. One scan (48) is refused by the convention check at 3.4% on one of its eight views; at a relaxed tolerance it reaches 23.58 dB, ranking 46 of 66.

Mechanism. The projector $P_{\mathcal{R}} = A^{+}A$ is applied with 40 conjugate-gradient iterations on $A^{\top}A$ rather than by forming any matrix. That iteration is stable over a window and not beyond it, and the window is measured: over 20 to 80 iterations the margin over the range oracle moves by 0.01 dB on 3 scans, so the reported count sits inside it. Past the window it does not converge but diverge. Conjugate gradient on $AA^{\top}$ assumes a symmetric positive definite operator, which requires A and $A^{\top}$ to be an exact adjoint pair; the projector library's are matched but not exact, and at 120 iterations one scan's unmeasured error grows from 149 to 1422, with the frozen model's own unmeasured error – a quantity no iteration count may touch – already drifting by 139 at 80. The walnut study met the same failure at 120 against 300 and answered it with a projector whose adjoint is exact by construction (Appendix J); the second backbone's decomposition used that construction from the start, and its margin moves by 0.01 dB between 120 and 300 iterations over 4 scans (Appendix U). The range oracle is $P_{\mathcal{R}}x_{\mathrm{gt}} + P_{\mathcal{N}}\hat{x}_0$; it is therefore given the ground-truth range-space component and keeps the null-space component of the frozen model.

Compute. Every run uses one RTX 4090. A per-scan adaptation takes 202 s including model loading and scoring. Under the same protocol the weight-space variant takes 5.4 min per scan and equivariant imaging takes 8.9 min, so twenty of our steps take about a third of the time of the weight-space variant's forty. The full grid of runs behind Table 1 takes roughly 340 GPU-hours.

How Figure 6 was selected. A qualitative figure is the easiest place in a paper to make a favourable selection. Both selections were therefore made from the ground truth alone and fixed before any reconstruction was inspected. The scan shown is the one whose gain is closest to the mean over the test set, rather than the scan with the largest gain. Each slice shown is the one with the most ground-truth detail within the central half of its axis. All panels share one display window taken from the ground-truth volume and one error scale, so that no panel is autoscaled to appear better than its neighbours. Each view is cropped to the patient and shown at its own aspect ratio rather than stretched.

Departures from the protocol. The following entries depart from the protocol. The weight-space column of the robustness table uses the published protocol. The weight-space variants in the downstream tables hold two views out of the loss. The weight-space variant on the second backbone is given its best learning rate per scan. The learning rate of the input projections in Table 1 was swept on one lung scan rather than ten (Appendix G). Every other cell in the paper follows the five rules of Section 5.

Variable size. F has 4.19M entries, whereas the image has 16.8M voxels. Optimising the 2D features as well gives 19.92M entries, which is more than the image has. The limiting factor is therefore the reachable set rather than a parameter budget; Appendix O sweeps the variable size over three orders of magnitude.

C RECONSTRUCTIONS AND ERROR MAPS

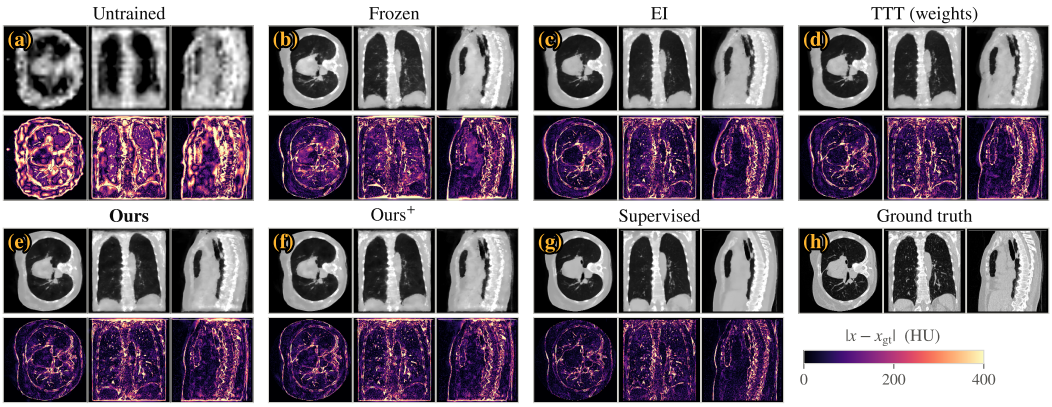


Figure 6: **Reconstructions and error maps in three planes.** LUNA16 at M=8, ordered from worst to best. Each block: axial, coronal and sagittal slices above the error map $|x - x_{\text{gt}}|$, one shared display window and HU scale. Selection protocol in Appendix B.

The scenes of Figure 1. Each panel shows one scan or image under a condition reported in the appendices, adapted with the activation variable used in the corresponding section. (a) LUNA16 at M=8, with the feature volume as the variable. (b) BSD100 under the doubled motion blur of Table 37 through the released NAFNet, with the input as the variable, at the variant and learning rate of that table; this image ends above its own range oracle. (c) Urban100 at 4× under a Gaussian kernel through the released SwinIR, with the input as the variable. (d) IXI T1 at 4×, with the U-Net bottleneck as the variable. (e) A walnut scan at M=8, with the feature volume as the variable. (f) BSD100 at 4× through the released EDSR, with an early feature map as the variable. The examples were chosen for visibility and are among the largest gains under their conditions; the means over every one of those conditions are reported in the tables.

D THE BUDGET VIEW

Table 2 sets out the organising principle of Section 5 explicitly: every method is placed according to what it needs from the target dataset and what it requires per scan. The four rows that also appear in Table 1 are repeated here so that the quoted numbers have a reference frame.

The range oracle as a method. On the DRR grid the measurements are consistent, $y = Ax_{\text{gt}}$, so $A^+y = P_{\mathcal{R}}x_{\text{gt}}$ for the pseudoinverse A^+ and the range oracle of Lemma 1 equals $\hat{x}_0 - A^+(A\hat{x}_0 - y)$, the frozen output projected onto $\{Ax = y\}$. It needs no ground truth; it is the point that measurement-loss descent of a free image from $\hat{x}_0$ converges to; and it is the projection that null-space networks build into the architecture (Schwab et al., 2019; Angermann et al., 2023). The frozen + projection row computes it from the measurements alone, $\hat{x}_0 - A^+(A\hat{x}_0 - y)$ by the forty conjugate-gradient iterations on $AA^\top$ that Table 12 uses for its projectors, on every one of the 66 scans of the lung row. On the same scans the oracle computed with x_{gt} is 26.01 dB, and the two differ by at most 0.005 dB on any scan, because the stored projections agree with the projector of the loss to 0.38% of their norm. The row is +1.11 dB above the frozen model and +1.71 dB below RDC; the least-squares row starts from zero and does not reach the frozen model. Under noise the projection adds $A^+\varepsilon$ to the range-space component, so the oracle, which keeps the noise-free one, is again only a bound (Appendix H). It bounds PSNR only: SSIM is not bounded by it.

The same source reports its self-supervised arms at six and ten views as well. The best of them at ten views is NeRP at 26.12 dB; the adaptation reaches 26.77 dB at six, +0.65 dB above it on four fewer views, and +1.89 above the best at eight. The two protocols agree to 0.22 dB where they can be compared directly, on the released checkpoint in the last row of the table.

Method	Cost/scan	PSNR
Needs nothing from the target set		
Frozen model	seconds	24.89
SART (Andersen & Kak, 1984)	minutes	*22.80*
Least squares (CG, 840 ops)	minutes	24.24
Frozen + projection onto $\{Ax{=}y\}$	seconds	26.01
NAF (Zha et al., 2022)	minutes	24.12 / *20.51*
NeRP (Shen et al., 2022)	minutes	*25.83*
TTT on weights	minutes	27.14
RDC	$\sim$3.5 min	**27.72**
Needs unlabelled scans, once		
Distilled adapter	seconds	26.57
Needs 600–800 *GT volumes*		
DeepSparse finetune (Lin et al., 2026)	seconds	30.92 / *31.14*

Table 2: **The budget view.** LUNA16 at $M{=}8$. Italics: values reported by Lin et al. (2026) in its Table II, LUNA16 at 256^3 and eight views, so the four quoted rows are comparable with each other and with our runs; others: our runs. Last row: one checkpoint scored both ways, differing by 0.22 dB. Our rows share 66 scans; distilled adapter: leave-one-out estimate over eight scans. Least squares is the measurement-only iteration at 840 operator applications from a zero initialisation, over 66 scans; frozen + projection is the frozen output projected onto $\{Ax{=}y\}$ by conjugate gradient from the scan's own measurements, over the same 66 scans: the range oracle of Lemma 1, which it matches to 0.005 dB on every scan (Appendix D); NAF is our own run over 65 scans beside the value its authors report, and is the only row where the two are given together (Appendix B).

E What the labelled alternative requires

Label efficiency. In a budget-matched sweep that varies only the number of labels, supervised finetuning of the same backbone reaches +2.36 dB with 48 labels (Table 3). This is below the +2.78 that the feature volume reaches with no labels. With seventy-three labelled volumes, however, the backbone's own finetuning procedure outperforms RDC.

Our method is intended for the case in which the target dataset has no ground truth. The relevant question is therefore not whether labels can outperform it, which they can, but how many labels are needed. The backbone's authors measured exactly this. At six views, their pretrained and finetuned model reaches *28.35* dB with seventy-three labelled volumes. It reaches *29.70* with a hundred and forty-seven and *30.22* with the full seven hundred and thirty-eight, so the curve is already flat at half the split. Our label-free adaptation of the same frozen model reaches 26.68 at that view count. A site that can label seventy-three volumes and train for two hundred epochs per stage should do so; our claim is restricted to the case in which it cannot.

Table 3 is the controlled version of the same question, and it is the version we can run end to end. It holds the backbone, the test scans and the per-scan budget fixed and varies only the number of labels. The authors' measurement quoted above cannot provide this control, because it differs from our setting in test split, view count and optimisation length at the same time. The table is a budget-matched comparison and nothing more: a finetune is run once and its cost amortises over every later scan, whereas our adaptation is run for every scan.

Two entries in the table need careful reading. At $K{=}1$ the supervised gain is exactly zero because the validation loss never improved on the frozen initialisation, and the early-stopping procedure therefore returned the initialisation unchanged. The entry thus means that one labelled volume was not enough for this protocol to move away from the initialisation, rather than that the finetune yielded 0.00 dB. The granularity of early stopping is the reason the validation checks are tied to fixed visit counts rather than to epochs. With a fixed visit budget an epoch is $400/K$ visits long. Validating once per epoch would therefore give $K{=}1$ four hundred opportunities to stop and $K{=}48$ only eight, and a stopping procedure whose granularity varies fifty-fold with the quantity under study does not give a controlled comparison. A sweep without that control, and without a schedule or gradient clipping, diverges over its budget and produces a curve that decreases between $K{=}10$ and $K{=}20$. With both in place the curve increases monotonically, as a label-efficiency curve must. The supervised variant then uses its budget properly and still falls short of our label-free adaptation.

Every K in the table below receives 400 visits and eight validation checks, and none is tuned to convergence, so each entry is a lower bound on the gain its labels can yield.

Labelled target volumes K	1	2	5	10	20	48	**0 (RDC)**
Gain over frozen (dB)	+0.00	+0.22	+1.39	+1.49	+2.06	+2.36	**+2.78**

Table 3: **Supervised finetuning on K labelled target volumes**, same backbone, matched per-scan budget, against our label-free adaptation. LUNA16 at M=8, first ten test volumes; our column reads +2.78 here and +2.83 in Table 1.

F A DIFFUSION PRIOR UNDER THE SAME BUDGET

Sparse-view CBCT. Analytic and algebraic reconstruction methods (Feldkamp et al., 1984; Andersen & Kak, 1984) degrade below a few tens of views. Learned methods post-process a direct inversion (Jin et al., 2017), unroll the operator (Adler & Öktem, 2018) or reconstruct in a feed-forward manner (Lin et al., 2024; 2026; Lee et al., 2026); they are trained and adapted with labels. Per-scan implicit fields (Zha et al., 2022; Shen et al., 2022) need neither labels nor a prior, and they are the competing methods under the same label budget as ours. At our geometry and view budget one of them reaches 24.12 dB against 24.89 for the frozen model, so at eight views a field fitted to the measurements alone does not reach a pretrained reconstructor that was never adapted (Appendix B). Score-based priors (Song et al., 2022; Kawar et al., 2022; Chung et al., 2023a) and the frozen 3D prior of Zhang et al. (2026), which is the closest published setting to ours, are assessed against the label budget below.

Diffusion priors under the label budget. Score-based priors solve sparse-view CT by posterior sampling (Song et al., 2022; Kawar et al., 2022; Chung et al., 2023a). They are trained on ground-truth volumes from the target distribution, which is the resource we assume to be unavailable; Li et al. (2026) train one prior across datasets. Zhang et al. (2026) guide a frozen 3D prior with measurement consistency and update no weights, which is the closest published setting to ours. Their prior, however, is trained on LIDC-IDRI, from which LUNA16 is drawn, so on our lung rows it has seen the target distribution. A prior that does satisfy the label budget is trained and reported below; it does not exceed the frozen model.

A prior that satisfies the label budget. A generative prior determines the null-space component directly, so we measure here what such a prior requires and what it yields. The prior is a 51 M-parameter 2D denoising diffusion model with a cosine schedule, trained from scratch for 120,000 steps on 105,435 axial TotalSegmentator slices at 256^2. It sees no volume from LUNA16, PENGWIN or the AbdomenAtlas split at training or at test time, so it uses no target ground truth. It has seen lung, pelvis and abdomen, whereas our backbone has only ever seen abdomen.

A 2D prior cannot be applied slice by slice here. Cone-beam rays cross axial planes, so a single slice is not measured independently, and the two-dimensional construction of Song et al. (2022) does not apply; that construction relies on a fan-beam geometry, in which each slice is measured independently. We therefore use the alternating scheme of Chung et al. (2023b): we denoise each axial slice with the score model, take conjugate-gradient steps on $\min \|Ax-y\|^2$ over the full volume, and re-noise along the DDIM path. The operator, the measurements, the geometry and the metric are those used in Section 5.

The sampler is given 40 warm-start consistency steps, followed by 100 reverse steps with 8 consistency steps each. This amounts to 840 operator applications in all and 14 minutes per scan, compared with the forty gradient steps of RDC. Over 30 scans it reaches 23.55 dB, which is -1.52 dB relative to the frozen model and 4.30 dB below RDC on the same scans through the same operator.

A second sampler. We also replace the normal-equation steps of the scheme above with the range-space replacement of Wang et al. (2023), which sets the range-space component to the data and assigns the null-space component to the prior, and we change nothing else. This gives 23.70 dB over 6 of these scans, against 23.69 for the scheme above; it is -1.65 dB relative to the frozen model and 4.45 below RDC. The per-scan gap between the two samplers never changes sign and never exceeds +0.01 dB, so the two published sampling schemes for this prior give essentially the same result.

The sampler configuration was chosen on a single scan. Starting the reverse process from pure noise gives 16.38 dB, because a slice-wise prior has no reason to make neighbouring slices agree and only the consistency step couples them. A conjugate-gradient warm start with the reverse process switched off entirely gives 21.22. Warm starting and then entering the schedule at t=0.3 gives 22.48, and entering it at t=0.6 gives 22.68. Adding total variation along the axis that the slice-wise prior cannot see gives 23.15, which is the configuration reported above.

The prior against the same iteration. Those configurations differ in operator budget as well as in the prior: the warm start costs about forty operator applications and the reported configuration costs 840. The control that isolates the prior runs the identical loop with the score model removed — same warm start, same 100 steps of 8 conjugate-gradient iterations, same total-variation weight. Over 30 scans it reaches 24.48 dB without the total-variation step and 24.44 with it, against 23.55 with the prior. The learned prior is worth -0.89 dB and is behind on 30 of 30 scans; the axial total variation is worth -0.04 once the iterations are there.

At the point where the sampler enters the schedule, t=0.6T, the score model recovers a noised axial slice of a test volume from 8.15 to 28.06 dB, against 30.16 on slices of the corpus it was trained on, over 8 slices each. The sampler hands it a noised least-squares reconstruction rather than a noised volume: from that input, whose own accuracy is 22.05 dB, its prediction of the clean term reaches 14.27 dB.

The same iteration over the whole lung row reaches 24.24 dB on 66 scans, -0.66 dB from the frozen model, with RDC ahead by +3.48 dB on 66 of 66 (Table 2). A slice-wise prior under a cone-beam operator is the mismatch the alternating scheme exists to bridge, and here it subtracts rather than adds, while the two published samplers agree to +0.01 dB. The claim the measurement supports is about this construction: a 2D prior trained within our budget, sampled two published ways, reaches neither a pretrained reconstructor nor the least-squares iteration it is built on.

G THE INPUT PROJECTIONS AND AN ADAPTIVE LAYER AS THE VARIABLE

The result on the second operator, where the input image rather than an internal feature is the variable that performs best, raises the question of whether the primary backbone admits the same variable. It does. Its input is the projection stack, and with every weight frozen the stack can be optimised as a free variable whose update passes through the whole frozen network, namely the 2D encoder, the unprojection, the codebook and the 3D decoders, exactly as the update of the input image passes through the U-Net. We run this variant under the matched protocol on the main lung row. Its learning rate is fragile, whereas that of the feature volume is not. On the single sweep scan, the gains at learning rates of 10^{-4}, $3\cdot10^{-4}$, 10^{-3}, $2\cdot10^{-3}$, $3\cdot10^{-3}$ and 10^{-2} are +1.88, +2.07, +1.70, +0.86, -0.20 and -6.28 dB respectively. A learning rate three times above the optimum therefore loses the entire gain, and at 10^{-2} the volume is six decibels worse than the frozen one after nine steps, because the dense, sign-like update on every detector pixel acts on the 2D encoder in the same way as an adversarial perturbation. At its best learning rate, over 66 scans it gains +2.66 dB, compared with +2.83 for the feature volume and +2.24 for the weight-space variant on the same scans; it is better than the weight-space variant on 66 of 66 scans and below the feature volume on 53 of 66. Decomposed over 6 of those scans, it removes 19% of the null-space error at an alignment of 0.59, whereas the feature volume removes 20% at 0.61 and the weight-space variant removes 15% at 0.53. On this backbone the input projections therefore lie between the two other variants. Their null-space update has nearly the same alignment and relative length as that of the feature volume, and it removes nearly the same fraction of the null-space error; what they give up within the step budget is range-space error. They are better than the weight-space variant on every scan and below the feature volume on all but one, and their learning rate has to be found by a sweep.

On every setting. The input projections are the variable through which PINER adapts a black-box model (Song et al., 2023); here they are optimised inside the same frozen network as the feature volume. We run them at that learning rate on every setting of Table 1, 9 settings in all. The input projections outperform the matched weight-space variant on 175 of 177 scans, that is on 99% of them, and they outperform every weight-space method on every row; their smallest setting mean is +2.47 dB over the frozen model, and their worst scan is +1.25. Compared with the feature volume, they are better on 5 rows and lower on 4. They are better on the backbone's own abdominal dis-

tribution, with +2.68 against +2.41 dB and +2.55 against +2.31 under the jittered gantry. They are also better at the sparsest sampling, with +3.52 against +2.97 at four views; there they match the weight-space variant in PSNR (+3.44) and are below it by 4.2 SSIM points. They are lower than the feature volume on lung and pelvis at eight views, with +2.66 against +2.83 and +2.79 against +2.88, and at ten views, with +2.47 against +2.94. Decomposed on ten scans each, these differences are differences of alignment. On the abdomen the null-space update of the input projections has alignment $\cos\theta = 0.64$, against 0.54 for the feature volume and 0.50 for the weight-space variant, and it removes 23% of the null-space error against 16 and 14. At four views its alignment is 0.66, against 0.60 for the feature volume and 0.66 for the weight-space variant; the input projections thus reach the same alignment as the weight-space variant. This is why, at four views, the alignment bound on the removable fraction of null-space error holds the feature volume below the weight-space variant but does not hold the input projections below it. The variable with the whole frozen network between it and the output moves the null-space component furthest where the frozen network fits the target best, which is the interpretation given in Section 4. The claim of Section 5.2 that no weight-space method exceeds an activation variable on any setting rests on the two activation variables taken together. Which of the two is more effective on a given row depends on the backbone and on the row; the corresponding result for the U-Net is given in Appendix V.

The adaptive layer. The other variable chosen in the sparse-view literature is an added adaptive layer (Zhao et al., 2024). Its analogue here is a 1^3 convolution on the fused feature volume, initialised to the identity. Every network weight and the volume itself are frozen, so that only the 16,512 weights and the bias of this layer are free, and the frozen decoders between it and the output are the same as for the feature volume. Its learning rate was chosen in the same way as for the other variants, on lung scans 0 to 4, and has an interior optimum at 1e-3. Table 4 reports it on every setting of Table 1. It gains +0.59 to +1.55 dB over the frozen model and is positive on every scan (worst +0.20). It is better than the matched weight-space variant on 0 of 177 scans, and it is below that variant by more than a decibel in every setting mean. Decomposed on the six lung scans of Figure 4, its null-space update has alignment $\cos\theta = 0.30$, against 0.61 for the feature volume with the same frozen decoders in both cases, and it removes 5% of the null-space error against 20. The alignment is determined by the frozen layers between the variable and the output and by the directions in which the variable can move. A channel-mixing layer can move the volume only within the span of its own channels, just as the narrow grids of Appendix W could move the early feature only within a coarse grid. The alignment of the frozen network is attained by an activation optimised at its native width, which is the variable that the label-free selection procedure of Appendix S selects.

Setting	n	Frozen	Adaptive layer	Weights	Input	Feature volume
Lung (LUNA16), M=8	66	24.89	25.90	27.14	27.55	27.72
Pelvis (PENGWIN), M=8	30	24.29	25.37	26.57	27.08	27.17
Abdomen (source), M=8	21	26.56	27.18	28.31	29.24	28.97
Lung, M=4	10	21.12	22.67	24.57	24.65	24.09
Lung, M=6	10	23.95	24.99	26.35	26.77	26.68
Lung, M=10	10	25.60	26.44	27.56	28.07	28.54
Lung, jittered	10	25.42	26.50	27.52	28.08	28.22
Pelvis, jittered	10	24.50	25.49	26.98	27.37	27.33
Abdomen, jittered	10	26.93	27.52	28.65	29.43	29.25

Table 4: **The adaptive layer on every setting.** Mean PSNR on the scans shared by all four variants, forty steps at each variant's global learning rate; the weights column is the matched weight-space variant of Table 1.

H Robustness to measurement perturbations

Perturbations. The lower two blocks of Table 1 vary the view count and jitter the gantry phase, and across them the gain varies within a range of roughly half a decibel. When the backbone's four anatomy checkpoints are exchanged with the target fixed at lung, the margin over the range oracle stays positive for every checkpoint, from +1.14 to +2.35 dB. The tooth-trained checkpoint, which has the largest domain gap in the paper, is the one for which the criterion selects the weight-space variant (Appendix P). At forty steps the objective does not fit the noise. Under Poisson noise that lowers the frozen model by 3.4 dB the gain grows, and it survives a miscalibrated projector and four views (Appendix H); beyond that step budget the objective does fit the noise (Appendix T).

The gain grows under noise, from +2.78 to +3.46. Under $I_0{=}10^5$ the weight-space variant falls 1.97 dB from its unperturbed value and RDC ends +0.08 dB ahead of it; beyond the forty-step budget it overfits (Appendix T). Miscalibrating the projector used by the loss, so that the adaptation minimises a residual computed with the wrong geometry, lowers the result by 0.17 dB at half a degree and by 0.54 dB at a full degree. The weight-space variant is the less sensitive of the two to this perturbation, losing 0.04 and 0.16 dB, and stays +0.66 and +0.41 dB below. At four views, where the frozen model has fallen to 21.12 dB, the method gives +2.97, so the method does not depend on the backbone already being nearly correct.

The comparison under the matched protocol, with every measured view in the loss, is given in Appendix T.

LUNA16		Frozen	TTT	**RDC**
No perturbation		25.34	27.33	**28.12**
Poisson noise	$I_0{=}10^6$	24.85	27.09	**27.56**
	$I_0{=}10^5$	21.98	25.36	**25.44**
Miscalibrated	0.5°	25.34	27.29	**27.95**
	1.0°	25.34	27.17	**27.58**
Fewer views	$M{=}4$	21.12	24.01	**24.09**

Table 5: **Robustness** (PSNR, dB), first ten lung volumes, so the unperturbed row exceeds the sixty-six-scan row of Table 1. Poisson noise perturbs the projections (network input and loss); miscalibration perturbs only the loss's projector.

The supervised line in the figure below is computed over the same 10 scans, so it lies at 5.18 dB rather than at the 6.03 of Table 1.

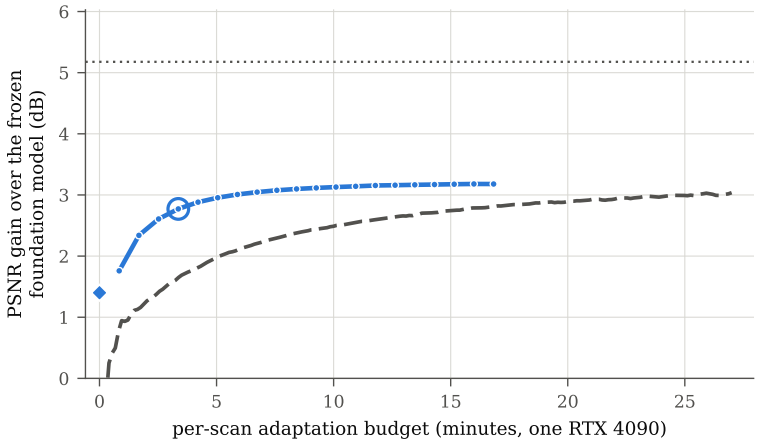


Figure 7: **Gain against per-scan adaptation budget.** Mean gain over the frozen model against wall time on one RTX 4090, LUNA16 at $M{=}8$, 10 scans. Solid: RDC, 200 steps in 17 min; the circle marks the reported operating point of 40 steps (3 min, 2.77 dB). Dashed: the matched weight-space variant, the same loss on the weights, against its own wall time (3.04 dB at 200 steps, 27 min). Dotted: a supervised finetune on 600–800 labelled volumes, at 5.18 dB on these scans. Diamond at zero minutes: the distilled adapter, which adds no per-scan cost.

I Does the effect survive a better-sampled scan?

The mechanism section reports a single operating point, and a method demonstrated at eight views raises the question of whether the effect is an artefact of that point, or of extreme sparsity in general. Table 6 varies the view count with everything else held fixed. The margin over the range oracle stays between +1.63 and +1.90 dB from 4 views to 20, and the fraction of null-space error removed improves rather than decays as the scan becomes better sampled, from 19% to 24%.

The last column rules out a confound. The second backbone of Appendix U is evaluated over a full circle, where only 28% of the frozen error lies in the null space, against 81% here. Because the

range oracle is given the entire range-space component, it lies much higher there. At twenty views the null-space share drops to 70%, yet the margin does not fall with it. The difference between the two settings is therefore not the size of the null space.

Views	Frozen	RDC	vs. oracle	Null err.	Unmeasurable
4	21.04	23.99	+1.63	-19.2%	78%
6	23.90	26.72	+1.77	-20.1%	81%
8*	25.07	27.85	+1.72	-19.7%	81%
10	25.59	28.57	+1.90	-21.7%	80%
20	26.68	30.26	+1.88	-24.0%	70%

Table 6: **View count against the null-space margin.** LUNA16, RDC, 6 scans per row except *, the 30-scan row of Table 12. Last column: fraction of the frozen error in the null space of the operator.

J THE PRIMARY BACKBONE ON REAL SCANS

J.1 WALNUT MICRO-CT

The walnut study in the preceding appendix uses the second backbone, so every number reported so far for the primary backbone of this paper was obtained on DRRs. The same five scans store the full calibrated circle at 0.3° per frame. The half-circle acquisition that this backbone assumes corresponds to frames 0–599 of the same data, and the eight views are taken at $0, 22.5, \ldots, 157.5^\circ$. The laboratory geometry is rescaled uniformly to the 1.6 mm voxel convention, which changes only the unit of the line integral. The frozen model reaches 19.07 dB, which lies between the tooth and pelvis configurations of Appendix P, and 78% of its error lies in the null space, the same operator regime as lung at eight views. Both variants are swept at the converged budget of 400 steps. Both optima are interior to the sweep over 9 learning rates, at 0.002 for RDC and 3e-05 for the weight-space variant.

Results. Our variant gains +1.12 dB over the frozen model and lowers its null-space error by 8%, with 54% of its update energy in the null space. The null-space coupling stated in the title therefore holds on real projections for this backbone as well. Structure invented in the null-space component would raise that error, whereas on every one of the 5 scans the error falls. The random point subsets that drive the update change the result by 0.003 dB across three seeds (Appendix B). Our variant does not exceed the range oracle here: it is -0.18 dB relative to the oracle and exceeds it on 0 of 5 scans. The weight-space variant exceeds the oracle by +0.33 on every scan, and the paired difference, RDC minus the weight-space variant, is -0.52 ± 0.04 dB. The criterion of Section 5.4 predicts this reversal in advance, and Table 16 shows what is distinctive about it. Both variants lie almost exactly at the minimum of the error identity; RDC reaches 99% of its own upper bound on the removable null-space error, so nothing is lost to the $r - \cos\theta$ term. Our r also stays inside the observed range of $\cos\theta$, which makes this the one reversal that the one-sided label-free procedure cannot detect.

Forward model. Replacing the measurements with Ax_{gt} computed through the same rays, and changing nothing else, moves our margin over the range oracle from -0.18 to +1.56 dB and the change in null-space error from -8% to -18%. These are the values that the DRR grid gives on lung. The frozen model itself is affected before any adaptation: it reaches 20.87 dB under consistent measurements, against 19.07 on the real ones. The operator of the second backbone is built from the scanner's per-view calibration, and the same control moved its margin by 0.07 dB. The primary backbone reconstructs through an idealised circular model of the same orbit, whose deviation from the real rays is measured at 2.5% of projection scale, and the 1.75 dB difference between the two controls is the cost of that idealisation. The only measurable difference between this study and the DRR grid is therefore the fidelity of the reconstruction operator.

Budget on real projections. On DRRs the PSNR curve over steps never decreases (Figure 7). Here, every learning rate above the optimum peaks by step 40–200 and then declines, so for the first time the protocol without a stopping rule has a cost. At the forty-step budget of the main grid RDC gains +0.67 dB over the frozen model, against +1.12 at the converged budget. At the converged optimum the tail of the curve is flat, with -0.01 dB over the last fifty steps. A stopping rule exists

to prevent fitting a measurement that the operator cannot reproduce. Here the learning-rate sweep absorbs most of that effect, because the optimum moves down to 0.002, a tenth of the optimum on the DRR grid.

Projector. The decompositions in this section use an exact-adjoint reimplementation of the projection geometry. TIGRE's forward projector and backprojector are a matched pair but not an adjoint pair, and under this geometry conjugate gradients on their composition diverges in double precision as well as in single precision. The replacement samples the same cone geometry differentiably and obtains $A^\top$ by automatic differentiation, exactly as the decomposition for the second backbone does. It agrees with the pipeline's projector to a correlation of 0.9999 on the scans' own projections, and it is idempotent to 10^{-3}. Its margins at 120 and 300 CG iterations differ by less than 0.01 dB, a check that the mismatched pair fails by a margin of the order of decibels.

	Internal volume (RDC)	Network weights
Gain over frozen (dB)	+1.12	+1.64
vs. range oracle (dB)	-0.18	+0.33
Null-space error	-8%	-14%
Synthetic Ax_{gt}, vs. oracle	+1.56	–
Synthetic, null-space error	-18%	–

Table 7: **Real scans on the primary backbone.** 5 scans, half circle at eight views, 400 steps, each variant at its interior optimum of 9 swept rates, fixed-range metric. Synthetic rows: measurements replaced by Ax_{gt} through the same rays.

J.2 Clinical pelvic scans

COBRA2026 (Thummerer et al., 2026) releases raw clinical cone-beam acquisitions from pelvic radiotherapy at six centres. Each case carries the measured detector signal, the acquisition geometry, the vendor calibration files and a planning CT deformed to the anatomy of the day. We use the Halcyon scans of one centre. They are acquired with a centred detector over an arc of about 210°, whereas the other centres mostly offset the detector, which the backbone cannot read. Of 22 cases downloaded, 18 are kept. The rule was fixed before any adaptation was run. A case is kept if, at every one of the eight views, its measured projections correlate above 0.8 with the projections of its reference. The 4 cases that fail it do so through poor agreement between reference and measurement, not through the geometry.

Geometry. The backbone expects eight views from 0 to 157.5°, the first of them lateral. The Halcyon arc holds a window of that width only when the first view is posterior. We therefore rotate the patient frame by 90° about the scanner axis, for the volume and the reference alike, and take the measured frame nearest to each required angle at its true angle, which lies within 0.3° of it. The same rotation applied to the first 5 PENGWIN test scans moves the frozen model from 23.71 to 24.05 dB and the gain of RDC from +2.63 to +2.54, so the prior does not depend on that orientation. The panel is binned to 256×128 elements, and all lengths are rescaled so that a 2.4 mm voxel is declared as the backbone's 1.6 mm, as for the walnut scans (Appendix J.1). The detector signal becomes a line integral through the nearest air scan, scaled by tube current and pulse length as in the dataset's own pipeline.

Reference. The reference is valid only inside the field of view of the clinical CBCT, which covers 6 to 19% of the cube, and every PSNR in this appendix is computed inside it. Across the kept cases its projections correlate with the measured ones at 0.86 to 0.99. The rest is registration error, scatter and beam hardening. It limits how much of a real gain the reference can show.

The loss on real projections. The raw detector signal carries a per-view intensity error that the idealised operator cannot reproduce. Fitted without a correction for it, both variants fall below the frozen model (Table 8). The same case with its measurements replaced by projections of the reference behaves as on the DRR grid, so the pipeline is sound and the difference lies in the measurement. The dataset's own scatter correction does not change the outcome. A per-view gain in the loss does. At each step the predicted projections p_v of view v are scaled by the least-squares factor

$a_v = \langle y_v, p_v \rangle / \|p_v\|^2$, and the loss is $\sum_v \|a_v p_v - y_v\|^2$. It uses no label and is applied to every variant alike. Its form matters. Fitting the measurement to the prediction instead, $a_v y_v + b_v \approx p_v$, admits a flat prediction as an exact minimiser, and the weights collapse onto it. The per-view gain has no such minimiser, because $a_v p_v$ matches y_v only when p_v is proportional to it. With it, RDC gains +1.43 dB on the first case, against +1.40 when the same case is simulated.

Results. Over the 18 cases the frozen model reaches 21.56 dB (Table 9). RDC on the feature volume gains +1.06 dB and improves on the frozen model in 17 of 18 cases (Wilcoxon $p < 10^{-4}$). The weights gain +0.64 dB and improve on it in 13. RDC is ahead of the weights by +0.42 dB, in 16 of 18 cases ($p = 0.0002$). Equivariant imaging changes the frozen model by -0.08 dB and RDC on the input projections gains +0.75. RDC on the feature volume is ahead of the first in 18 cases and of the second in 17 ($p < 10^{-4}$), the ordering of Table 1. The gain grows with the quality of the reference, with a rank correlation of 0.69 between the two. On the half of the cases whose projections agree best with their reference, at 0.93 to 0.99, RDC gains +1.53 and the weights +1.14. On the other half they gain +0.59 and +0.15. The lead of RDC is +0.40 and +0.45 dB in the two halves, so it does not come from the cases whose reference is best.

Case	Projections and loss	Frozen	RDC (F)	TTT
F003	Synthetic Ax_{ref}, no correction	24.54	+1.40	+0.98
	Measured, no correction	21.89	-0.32	-1.05
	Measured, scatter-corrected	21.95	-0.48	-0.52
	Measured, per-view affine in the loss	21.89	+1.49	-12.40
	Measured, per-view gain in the loss	21.89	+1.43	+1.07
F025	Measured, no correction	21.10	-0.09	-0.11
	Measured, scatter-corrected	21.75	-0.53	-0.36
	Measured, per-view affine in the loss	21.10	+1.70	-1.83
	Measured, per-view gain in the loss	21.10	+2.03	+1.77

Table 8: **Which loss real projections need.** Change in PSNR inside the field of view over the frozen model, forty steps. The synthetic row replaces the measurements by projections of the reference through the same rays.

			Optimises weights		Optimises activations	
Case	corr.	Frozen	EI	TTT	RDC (in)	RDC (F)
F003	0.93	21.89	-0.84	+1.09	+0.95	+1.43
F009	0.86	22.26	-0.81	-0.02	-0.41	+0.00
F011	0.93	22.41	-0.55	-0.30	-0.12	+0.52
F013	0.87	21.38	-0.78	+0.40	+0.11	+0.51
F015	0.87	22.64	-0.34	+0.15	+0.67	+0.69
F017	0.88	21.23	-0.58	-0.23	+0.30	+0.60
F020	0.94	21.66	+0.05	+0.51	+0.37	+0.76
F022	0.89	21.54	-0.67	-0.51	-0.24	+0.23
F025	0.99	21.10	+0.74	+1.80	+1.72	+2.03
F034	0.99	21.07	+0.99	+1.87	+1.79	+2.01
F035	0.96	21.82	+0.42	-0.53	+0.51	+0.79
F039	0.99	20.46	+0.67	+1.02	+1.25	+1.43
F040	0.90	23.22	-0.39	+0.03	-0.42	-0.08
F042	0.99	21.30	+0.00	+0.70	+0.53	+1.04
F044	0.92	21.94	-0.20	+0.12	+0.79	+1.26
F046	0.99	17.58	+1.97	+4.98	+5.08	+4.80
F048	0.94	22.39	-0.53	+0.17	+0.12	+0.43
F049	0.90	22.24	-0.70	+0.28	+0.54	+0.68
Mean		21.56	-0.08	+0.64	+0.75	+1.06
Above frozen			7/18	13/18	14/18	17/18
RDC (F) ahead			18/18	16/18	17/18	–

Table 9: **Real clinical scans.** 18 Halcyon pelvic scans of COBRA2026, eight views, forty steps, per-view gain in the loss for every variant. Frozen is PSNR inside the field of view and the other columns its change. Corr. is the mean correlation between measured projections and projections of the reference at the eight views.

K THE SEGMENTATION STUDY, IN FULL

Both tasks. A small 3D classifier is trained on ground-truth volumes alone, frozen, and applied to every reconstruction at identical patch coordinates. Over 45 volumes it separates 88 annotated nodules from four times as many lung locations with an AUC of 0.974 on ground truth. The frozen model loses 0.944 of that AUC, and the adaptation recovers 55% of the loss, reaching 0.961 against 0.956 for the weight-space variant (Appendix L). On PENGWIN a segmentor trained on ground truth recovers 27% of the Dice that the frozen model loses, four times the recovery of either weight-space variant. The rest of this section describes that study.

PENGWIN provides per-fragment labels for pelvic fractures, and segmenting the three bones is a task that the 1.6 mm grid can express. A reference model is trained on ground-truth volumes and their labels, and never on a reconstruction. It reaches 0.969 Dice on the ground truth, and none of the 90 scan-and-class cells falls below the gate. Applied to the frozen backbone's reconstructions, its Dice falls to 0.709. Our adaptation recovers 27% of that loss (Table 10). All three variants improve PSNR on these scans, and all three improve Dice, but the two gains are not proportional. The two weight-space variants gain +0.016 and +0.018 Dice, whereas the feature volume gains +0.069, so four times as much of the improvement reaches the task. In terms a surgeon would use, the labels contain 1.06 pieces of bone per class, and the frozen reconstruction fragments them into 2.77. Our adaptation removes about half of the spurious pieces, giving 1.94 against 2.09 and 2.16 for the baselines.

PSNR and Dice reward different properties of a reconstruction. PSNR improves as the total error energy decreases, and blurring reduces that energy; Dice is decided at the cortical boundary, which blurring destroys. SSIM is closer to Dice in this respect. Over the same 30 scans RDC gains +3.8 SSIM points, against +3.1 for the weight-space variant and +2.6 for equivariant imaging. This margin is smaller than the margin in Dice, because most of the task-level gap is not visible in either image metric.

Reconstruction read by the segmentor	Dice	vs. frozen	Bone pieces
Ground truth	0.969	—	1.14
Frozen backbone	0.709	—	2.77
RDC	**0.778**	**+0.069**	**1.94**
Weights	0.725	+0.016	2.09
EI	0.726	+0.018	2.16

Table 10: **Ground-truth-trained segmentor on each reconstruction.** PENGWIN at M=8, 90 cells over 30 volumes; the segmentor reaches 0.968 Dice on held-out ground truth. Bone pieces: connected components, against 1.06 in the labels.

L THE NODULE DETECTOR, IN FULL

The whole range from the frozen model to ground truth spans only 0.030 AUC, so at 1.6 mm this task is barely sensitive to reconstruction quality. Figure 8 shows, lesion by lesion, where the ordering between the methods holds and where it breaks down.

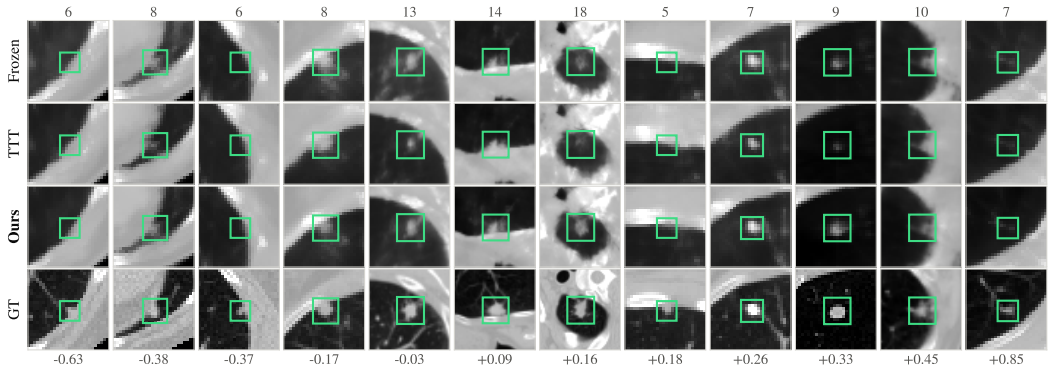


Figure 8: **Every retained lesion, four reconstructions each.** Ordered by detector-score change from frozen to RDC; diameter above each lesion, score change below. All 12 of the 88 annotated lesions retained in the preprocessed pool and detected on ground truth.

Reconstruction	AUC	Sens. at 1 FP/vol	AUC, < 6 mm
Ground truth	0.974	0.977	0.966
Frozen foundation model	0.944	0.966	0.938
DC on weights	0.956	0.966	0.949
RDC: DC on the feature volume	0.961	0.966	0.953

Table 11: **Ground-truth-trained detector on each reconstruction.** LUNA16 at M=8, 45 volumes, identical patch coordinates for every method.

M A TASK-LEVEL QUESTION THE GRID CANNOT EXPRESS

The per-fragment labels of PENGWIN make it possible to test whether the dark line between two bone fragments survives sparse-view reconstruction. The line does not survive even in the *ground truth*. At 1.6 mm the gap between fragments is a partial-volume edge rather than a resolved feature, and on the ground-truth grid the intensity in the gap lies only 2–17% of the way from bone to air. Our adaptation changes that figure by ≈ 0.01, which is the expected outcome when the target quantity is absent from the data. This is a limitation of the preprocessing resolution, and of any backbone trained at that resolution, rather than evidence about the method. Answering the question would require finer voxels and a model that accepts them. The lung-nodule study of Section 5 asks the same class of question at a resolution that can express the answer, because a 6.5 mm lesion spans four voxels.

N EXACT VALUES BEHIND FIGURE 4

	PSNR	$\|P_{\mathcal{R}}e\|$	$\|P_{\mathcal{N}}e\|$	null share	range oracle
Frozen foundation model	25.07	100.8	206.4	80.8%	26.13
RDC (data consistency on F)	**27.85**	21.6	**165.7**	98.3%	27.93
same optimisation, untrained net	22.85	67.5	289.3	94.8%	21.37
update energy $\hat{x}-\hat{x}_0$: 47% range, **53% null space** (n=30 scans)					

Table 12: **Where the gain is located.** Exact projectors (CG-40), LUNA16 at M=8, the 30 scans this decomposition covers (hence slight differences from Table 1). Upper bound: the range oracle of the *frozen* row.

Free variable	PSNR	above oracle	$\|P_{\mathcal{N}}e\|$ drop	null share of update
Frozen	25.07	–	–	–
Network weights (Darestani et al., 2022)	27.16	+1.03	15%	53%
2D per-view features	27.61	+1.48	18%	53%
Fused 3D volume (RDC)	27.85	+1.72	20%	53%
Both feature tensors	28.14	+2.01	22%	57%

Table 13: **A sweep over parameterisations, measured in the null space.** Loss, step count, the 30 scans and the frozen starting reconstruction are the same in every row; only the free variable changes.

Sweep over variables. Fixing the objective and the initial reconstruction and varying only the variable being optimised gives the sweep of Table 13 and Figure 4b. Every configuration places nearly the same share of its update energy in the null space, 53 to 57%, yet the fraction of null-space error that this share removes ranges from 15 to 22%. Neither the amount of null-space movement nor the parameter count separates the variables; the weights are by far the largest variable and recover the least. At this step budget the variables differ in the direction of their null-space update. Every variant exceeds the range oracle, so the null-space component moves regardless of which variable the loss is differentiated with respect to.

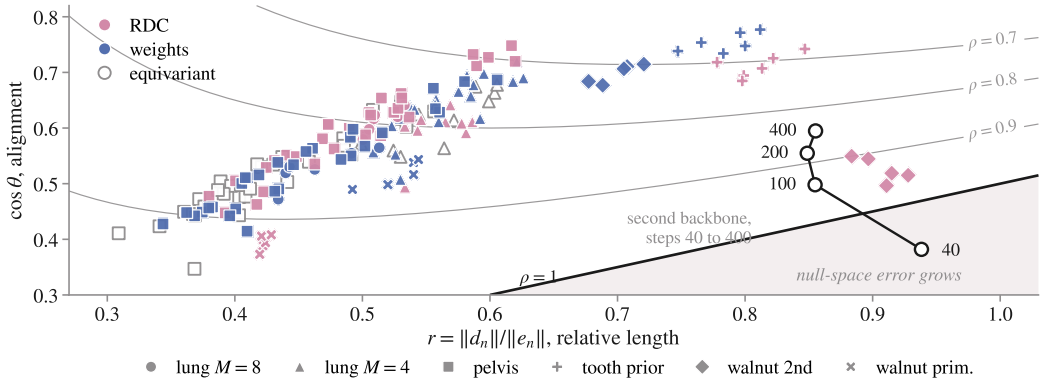


Figure 9: **The error identity on every decomposition.** Contours: predicted reduction; heavy line $\cos\theta = r/2$: below it the null-space error grows (shaded). One marker per scan (174 scans); trajectory: second backbone at four step budgets.

The step-length term. The $r - \cos\theta$ term of the identity is large in only one place: RDC on the second backbone at its released width. There, a variable with four times as many entries as the image has a null-space update that is longer than the null-space error it corrects, and narrowing the variable moves both terms in the right direction (Appendix U). Our own variable never reaches that regime over a factor of 512 in the number of entries (Appendix O). The four-view row of Table 1 is likewise a matter of the alignment bound, because fitting all four views raises the weight-space variant's $\cos\theta$ to 0.66, against 0.60 for RDC, over 10 scans. At the sparsest sampling, where 81% of the error lies in the null space, the null-space direction chosen by adapting the weights is better aligned than the direction that the frozen decoder provides.

The upper bounds in the table below describe where each variable has arrived at its step budget rather than a fixed property of the variable. The variable with the higher upper bound is better in every row, with no reference to the relative step length, and on the lung row the two variants converge to the same alignment by 400 steps (Appendix T). The $r - \cos\theta$ term accounts for more than 0.8 points only on the second backbone at its released width, where RDC overshoots by 0.26 or more; elsewhere the overshoot is at most 0.11. That overshoot is the largest effect in the table and the only one that changes a sign: at forty steps on the second backbone at its released width the variant reaches -122% of its own upper bound, which is to say it increases the null-space error rather than reducing it, because its relative step length is more than twice its alignment. This is the row that the width rule of Section 4.4 is for, and narrowing the variable returns it to its bound (Appendix U).

Each row of the table below is measured at its own step budget, and four of the rows use forty steps.

		Internal volume (RDC)			Network weights	
Setting	n	$r - \cos\theta$	bound	removed	bound	removed
Lung, M=8	6	-0.10	21.0%	20.4%	15.1%	14.9%
Lung, M=4	10	-0.04	19.8%	19.6%	25.1%	24.7%
Pelvis, M=8	30	-0.11	20.5%	19.8%	15.8%	15.4%
Tooth prior on lung	6	+0.10	29.8%	29.1%	34.3%	34.2%
Second backbone, 400 steps	6	+0.26	19.9%	15.5%	26.3%	26.1%
Second backbone, 200 steps	6	+0.29	17.0%	11.7%	22.5%	22.4%
Second backbone, 100 steps	6	+0.36	13.5%	6.2%	18.7%	18.6%
Second backbone, 40 steps	6	+0.56	7.8%	-7.8%	12.5%	11.8%
Walnut, second backbone	5	+0.38	14.9%	6.7%	28.5%	28.5%
Walnut, primary backbone	5	+0.03	8.1%	8.0%	14.4%	14.4%

Table 14: **The error identity as an upper bound.** Every decomposition with both variants on the same scans. Since $\rho^2 = \sin^2\theta + (r - \cos\theta)^2$, removable null-space error is bounded by $1 - \sin\theta$ (*bound*); the $r - \cos\theta$ term is the gap to *removed*.

The alignment across scans. Section 5.3 claims that, for a fixed operator, the alignment is set by the pretraining and the target and not by the scan. Table 15 is the evidence: within every setting the standard deviation of $\cos\theta$ across scans is 0.01–0.08, and the ordering of the two variables by alignment, which decides the sign of the reversal (Table 16), holds on 58 of 62 paired scans over the six settings. The four exceptions are pelvis scans, the setting with the widest spread. Pooled over the 174 decompositions of Figure 9, the measured ρ is never below $|\sin\theta|$, agrees with the identity to 8×10^{-5}, and reaches a median 98% of the bound, 169 of the 174 runs above 90%; the five below it are the second backbone's feature volume on the walnut scans, whose relative length overshoots its alignment (the row with the largest $r - \cos\theta$ in Table 14).

		Feature volume, $\cos\theta$		Weights, $\cos\theta$		feat. > weights
Setting	n	mean $\pm$ sd	range	mean $\pm$ sd	range	on scans
Lung, M=8	6	0.61 ± 0.02	[0.58, 0.64]	0.53 ± 0.03	[0.47, 0.56]	6 of 6
Lung, M=4	10	0.60 ± 0.04	[0.49, 0.64]	0.66 ± 0.04	[0.61, 0.72]	0 of 10
Pelvis, M=8	30	0.60 ± 0.08	[0.45, 0.75]	0.53 ± 0.08	[0.41, 0.69]	26 of 30
Tooth prior on lung	6	0.71 ± 0.02	[0.68, 0.74]	0.75 ± 0.02	[0.73, 0.78]	0 of 6
Walnut, primary	5	0.39 ± 0.01	[0.37, 0.41]	0.52 ± 0.02	[0.49, 0.54]	0 of 5
Walnut, second	5	0.52 ± 0.02	[0.50, 0.55]	0.70 ± 0.02	[0.68, 0.71]	0 of 5

Table 15: **The alignment across scans.** Per setting, $\cos\theta$ of each variable at its step budget, mean $\pm$ standard deviation and range over the scans, and the number of paired scans on which the feature volume's alignment is above the weights'. The settings of Table 14 other than the second backbone's step-budget sweep.

The table below includes the second backbone at each of its four step budgets and the real scans on both backbones.

Free variable	Entries	vs. oracle	Null err.
R=4 grid	8k	+0.39	-9.1%
R=8 grid	66k	+1.01	-14.1%
R=16 grid	524k	+1.47	-17.8%
Native grid (RDC)	**4.19M**	**+1.72**	**-19.7%**

Table 17: **Size of the free variable over three orders of magnitude.** LUNA16 at M=8, the same 30 scans as Table 13, free volume restricted to an R^3 grid upsampled into the decoder's native input.

		$r - \cos\theta$		predicted	measured
Setting	n	RDC	weights	difference	difference
Lung, M=8	6	-0.099	-0.064	-0.056	$+0.78 \pm 0.09$
Pelvis, M=8	30	-0.105	-0.079	-0.043	$+0.70 \pm 0.07$
Lung, M=4	10	-0.043	-0.075	$+0.049$	-0.48 ± 0.10
Tooth prior on lung	6	$+0.097$	$+0.030$	$+0.051$	-0.34 ± 0.16
Walnut, primary backbone	5	$+0.029$	$+0.010$	$+0.064$	-0.52 ± 0.04
Second backbone, 400 steps	6	$+0.260$	-0.029	$+0.101$	-0.81 ± 0.43
Second backbone, 200 steps	6	$+0.294$	-0.024	$+0.101$	-0.72 ± 0.43
Second backbone, 100 steps	6	$+0.357$	$+0.003$	$+0.119$	-0.76 ± 0.42
Second backbone, 40 steps	6	$+0.556$	$+0.074$	$+0.194$	-1.10 ± 0.48
Walnut, second backbone	5	$+0.382$	$+0.001$	$+0.217$	-2.16 ± 0.15

Table 16: **Which variable to free, predicted and measured.** Every decomposition with both variants on the same scans, ordered by Corollary 3. Last column: paired mean over n scans $\pm$ one standard error; negative means the weight-space variant is better.

O THE WIDTH OF THE FREE VARIABLE

The size of the free variable can also be swept continuously. Restricting the free volume to a coarse grid and upsampling it into the decoder changes only the number of free entries. Over a factor of 512 in the number of entries, every configuration exceeds the range oracle, including one with 2048 times fewer entries than the image has voxels. The margin over the oracle rises monotonically with the number of entries, from +0.39 to +1.72 dB (Table 17). What a larger variable yields is therefore a larger set of anatomies that it can express, rather than more freedom in the fit.

P DOES THE MECHANISM NEED A PRIOR THAT SUITS THE TARGET?

The mechanism depends on the frozen decoder moving the null-space component in a direction that the anatomy supports. The most direct way to break it is therefore to apply the decoder to an anatomy it was not trained on. Our backbone has released checkpoints trained with the same training procedure on four anatomies. Fixing the target at lung and swapping the checkpoint therefore changes the distance between prior and target while the architecture, the operator, the protocol and the scans stay fixed. No other comparison in this paper separates these two factors, because a second model differs in architecture and in domain gap at the same time. Table 18 reports this sweep over 6 scans.

The margin over the range oracle does not close. It ranges from +1.14 to +2.35 dB across a span of 16.7 dB in frozen quality. The configuration that should break the mechanism, if the mechanism depended on a well-matched prior, is the one that does not break it. A model trained on teeth and applied to lungs starts at 13.52 dB, which is a larger domain gap than any other in this paper, and it still exceeds its own range oracle by +1.53 dB. The null-space error moves in the opposite direction to this prediction: it falls by 29% in that configuration, against 14% in domain. A prior that suits the target less therefore recovers more of the null-space error rather than less.

Two features of the last configuration deserve comment. First, the share of the update that lies in the null space collapses from roughly three quarters to a tenth. This is the expected behaviour of a badly matched prior, because almost all of the error is then range-space error and that is where the objective acts. Even so, the null-space error still falls by more than in any other configuration.

Second, this is the only configuration in the table where the weight-space variant outperforms RDC, which is the behaviour to expect once the premise that the decoder is worth keeping fixed has itself become weak.

Convergence with the tooth prior. The tooth prior is the cone-beam instance of a decoder whose learned mapping is wrong for the target. Before the run, we therefore predicted that this would be the condition in which the weight-space variant converges to a different alignment from the feature volume. Both variants were run to 400 steps under the matched protocol on the same 6 scans. Both start at 13.52 dB. At forty steps the feature volume gains +9.96 dB and the weight-space variant +10.39. At 400 steps the weight-space variant is better on 6 of 6 scans, with a gain of +13.53 against +11.74. The alignment $\cos\theta$ of the null-space update with the null-space error is 0.87 for the weight-space variant against 0.80 for RDC, and the weight-space variant removes 51% of the null-space error against 40. The alignment gap of +0.07 exceeds the threshold fixed in advance, in the upward direction, whereas on lung the same two variants end -0.00 apart (Appendix T). Table 25 contains both rows, labelled above and meet respectively.

A backbone without a prior. The other end of the sweep over priors is a backbone with no prior at all. Running the identical optimisation on a randomly initialised backbone, with the same variable, loss and number of steps, reaches only 22.64 dB over the 66 test scans of the main comparison (Table 1). Its update places 11% of its energy in the null space, against 53% for RDC. Its null-space error before adaptation is already larger than that of the pretrained model after adaptation (Table 12).

		Internal volume (RDC)		Network weights	
Prior trained on	Frozen	vs. oracle	Null err.	vs. oracle	Null err.
Lung, in domain	30.22	+1.14	-13.5%	+0.30	-6.8%
Abdomen	28.13	+1.65	-18.4%	+0.73	-11.9%
Pelvis	25.15	+2.35	-25.1%	+0.80	-13.0%
Teeth	13.52	+1.53	-29.1%	+2.06	-34.6%

Table 18: **One architecture, four priors, one target.** Our backbone's checkpoints for four anatomies, lung at M=8, the same 6 scans, $P_{\mathcal{R}}$ by forty CG iterations. Null err.: change in $\|e_{\text{null}}\|$ from frozen (negative is a recovery).

Q THE UPPER BOUND OF THIS CLASS OF METHOD

Once the adaptation is complete, the data-consistency loss has nothing further to act on. The measurement residual has fallen by two orders of magnitude, and the remaining error is almost entirely in the null space (Table 12). Adding a further consistency-driven stage on top, whether with free or with geometry-anchored Gaussians, therefore yields $+0.00$ dB on every scan. Any further improvement requires information that the measurements do not carry.

Reachable against removed. The upper bound $1-\sin\theta$ given by the error identity is a bound along the direction that the loss actually took. A second bound asks whether the null-space component of the target lies inside the reachable set of the variable at all. We measure it by fitting the activation variable in image space, with no operator in the loss, to $\hat{x}_0 + P_{\mathcal{N}}(x_{\text{gt}} - \hat{x}_0)$, which is the complement of the range oracle, for 300 steps at the learning rate of that variant. The same fit is run towards a random null-space vector of the same size and towards the ground truth (Appendix U). Table 20 reports the three residuals on every operator next to the share of the null-space error that forty steps of data consistency removed. On the cone-beam backbone the fit towards the true null-space direction leaves a residual of 0.45 of the offset on lung, 0.40 with the tooth prior and 0.32 at four views. A random null-space direction is not followed at all, with residuals of 0.999 to 0.999. On the same scans, data consistency removes 20, 29 and 19% of the null-space error. The two conditions in which the weight-space variant is better, the tooth prior and four views, are therefore not conditions in which the activation variable can move the null-space component less far. What separates them is the direction that the loss chooses. On the U-Net, adapting the input image leaves a residual of between 0.13 and 0.34 towards the true direction; the residual is largest under the inverted contrast, where the mapping learned by the frozen network is wrong. The LR input of EDSR, which is sixteen times smaller than the image, leaves a residual of 0.73. On the deblurring network, whose null

space is a set of lines in frequency space and whose input has the size of the image, the input leaves a residual of 0.33 to 0.43 towards the true direction and a similar residual of 0.27 to 0.50 towards a random one. The selectivity of the cone-beam and super-resolution decoders is therefore not universal. On the pixel mask the true content of the hole is reachable, with residuals of 0.54 to 0.59, whereas random hole content is not, with residuals of 0.86 to 0.96. Data consistency moves the null-space error there by between -2 and 2%. The zero alignment reported in Appendix X is therefore a property of the loss rather than a limit on how far the variable can move the null-space component. The headroom, defined as the reachable share minus the removed share, ranges from 0.20 on EDSR, Gaussian kernel to 0.82. On lung at convergence, where 23% of the null-space error is removed, the headroom is still a third of the null-space error. This is the part that a source of information beyond the scan could recover, and Appendix R quantifies what measurements alone cannot provide.

A corrector trained outside the scan. The headroom above is the part of the null-space error that a source of information beyond the scan could recover, and the nearest such source is the distribution the backbone was pretrained on. We train a corrector on that distribution and apply it after the adaptation, so that it reads $\hat{x}_1$ and returns a null-space increment: c_{hi} corrects the high band alone, c_{full} the whole null-space component, and c_0 is the same architecture applied to the frozen output in place of the adaptation. Six predictions were registered with numerical thresholds before the runs. Three hold and three fail, and Table 19 reports every arm on both sets. The ordering holds: c_{full} at 28.23 is above c_{hi} at 27.97, which is above the adaptation at 27.85, and the post-processor alone reaches 25.30, which is -2.56 dB below the adaptation it replaces, so the gain is not something a learned post-process obtains from the frozen output. What fails is the size of the correction. Against a registered bar of $+0.5$ dB the full corrector adds +0.37 on lung and +0.77 on the source held-out set; the additions are consistent, positive on 30 of 30 and 4 of 4 scans, but they are a fraction of the adaptation's own +2.78. In null-space terms the corrector removes 4% of the error the adaptation leaves on lung and 9% on the source held-out set, against registered bars of 15 and 40%, and in the band above 0.5 it removes 2% at an alignment of 0.20, below the 4% and 0.28 that the adaptation itself reaches there. A corrector trained on the source distribution and run for 14000 iterations therefore recovers little of the headroom, on the domain it was trained on and on the one it was not. The ceiling of this class of method is set by the prior in the frozen decoder, and a second network trained on the same distribution does not raise it.

	Lung (LUNA16), 30 scans						Source held-out, 4	
Reconstruction	PSNR	null-space error removed (%), total and by band					PSNR	removed (%)
		total	<0.1	0.1–0.25	0.25–0.5	>0.5		
frozen $\hat{x}_0$	25.07	0	0	0	0	0	24.35	0
$\hat{x}_0 + c_0$, source post-processor	25.30	2	5	2	1	1	25.04	6
RDC $\hat{x}_1$	27.85	20	46	28	11	4	26.65	15
$\hat{x}_1 + P_{\mathcal{N}} H c_{\text{hi}}$	27.97	21	46	28	13	6	26.83	17
$\hat{x}_1 + P_{\mathcal{N}} c_{\text{full}}$	28.23	23	52	33	14	5	27.42	22
range oracle	25.98	0	0	0	0	0	25.28	0
alignment $\cos\theta$ of the null-space update, by band, lung								
RDC			0.85	0.70	0.46	0.28		
c_{hi}			0.02	0.09	0.22	0.20		
c_{full}			0.46	0.40	0.24	0.15		
c_0			0.33	0.24	0.18	0.15		

Table 19: **A source-trained corrector after the adaptation.** Null-space error removed is measured against the frozen output and split by $|k|$ band. c_{hi} and c_{full} are applied to the adapted volume $\hat{x}_1$ and projected onto the null space; c_0 is applied to $\hat{x}_0$ instead. The range oracle removes none by construction. Lung is held out from the corrector's training; the source held-out set is not.

Alignment by frequency and direction. The global $\cos\theta$ is a single angle between two vectors. Table 21 resolves it per region of the null space: every frequency of a 32^3 patch is assigned to a radial band and to a bin of its angle to the nearest sampled plane through the patch. Frequencies are placed per patch because the direction of a cone-beam ray changes with the voxel by up to the half-cone angle, which at eight views exceeds half the view spacing. Two observations follow. First, the alignment is a low-pass property. On lung at forty steps, the alignment of the feature volume falls from 0.84 below a tenth of Nyquist to 0.30 above half of Nyquist, and that of the weight-space

variant falls from 0.77 to 0.20. The removed share falls with it, from 45 to 4%. Both quantities fall with frequency in every condition and variant we decomposed, including the tooth prior, the noise run and four views. The highest band holds 24% of the null-space error, and almost nothing is removed there. This is the smoothest-completion behaviour that Heckel & Soltanolkotabi (2020) prove for an untrained generator; it survives pretraining as a preferred alignment, and it belongs to the point decoder through which every activation variable passes. On the same scans, the per-view 2D features and both tensors together fall from 0.84 and 0.87 in the lowest band to 0.22 and 0.30 in the highest. The four variables keep the same order in every band, namely both tensors, then the volume, then the 2D features, then the weights. None of them exceeds an alignment of 0.30 in the highest band, which holds 24% of the null-space error. Second, the direction relative to the sampled views is the weaker coordinate. Within the three lower bands the alignment rises from the frequencies nearest a sampled plane to the farthest, from 0.66 to 0.77, and only in the highest band does it fall, from 0.35 to 0.27. The energy of the null-space error is nearly flat across the distance bins. Microlocal analysis of which singularities a data set can see (Borg et al., 2018; Bubba et al., 2019) organises the null space by direction, whereas this decoder fills it by frequency. The deficit of the weight-space variant at forty steps, which is 0.07 to 0.14 per band and largest in the 0.25–0.5 band, is spread across the directions. At 400 steps the two variants agree per band to within 0.09. On MRI, where each unsampled line has a well-defined distance to the nearest sampled line, the alignment does fall with that distance, from 0.62 to 0.26 at 8×, and it also falls with frequency. On super-resolution and motion deblurring the alignment falls with frequency as it does on the cone-beam projector, from 0.51 to 0.18 on EDSR and from 0.78 to 0.18 on NAFNet over Urban100. On the box blur it is flat, 0.48 to 0.43, and on the pixel mask it stays near zero in every shell of distance from the observed pixels. Data consistency through a frozen decoder therefore leaves the high-frequency part of the null-space error in place, whatever its direction, and Table 20 shows that this part is within the reachable set of the variable.

				image-space fit, residual / offset			
Condition	weights	$\cos\theta$	removed (%)	true null	random null	truth	headroom
CBCT lung M=8	meet	0.61	20	0.45	0.999	0.37	+0.35
CBCT tooth prior on lung	above	0.71	29	0.40	0.999	0.13	+0.31
CBCT lung M=4	weights lead	0.59	19	0.32	0.999	0.24	+0.48
MRI T2 8×	meet	0.35	6	0.13	0.624	0.06	+0.82
MRI T2 4×	below	0.42	7	0.22	0.574	0.10	+0.71
MRI T1 4×	above	0.38	2	0.26	0.678	0.12	+0.72
MRI T2 inverted	above	0.38	4	0.34	0.705	0.24	+0.62
MRI random mask	below	0.63	21	0.20	0.597	0.07	+0.58
EDSR, Gaussian kernel	below	0.35	7	0.73	0.996	0.65	+0.20
SwinIR, Gaussian kernel	below	0.36	7	0.67	0.998	0.58	+0.26
NAFNet motion, BSD100	above	0.42	9	0.33	0.271	0.25	+0.57
NAFNet motion, Urban100	above	0.54	19	0.41	0.448	0.27	+0.39
NAFNet box, BSD100	meet	0.40	13	0.43	0.502	0.28	+0.43
Inpainting, hole loss, rectangles	no alignment	-0.05	-2	0.54	0.961	0.56	+0.48
Inpainting, full loss, IXI thin	small alignment	0.22	2	0.58	0.881	0.57	+0.40
Inpainting, full loss, BSD100 thin	no alignment	0.02	-1	0.59	0.856	0.58	+0.42

Table 20: **Reachable against removed.** Per condition and activation variable: $\cos\theta$ and removed share of null-space error at forty steps; residual of a 300-step image-space fit of that variable to three targets, relative to the target's offset from frozen; headroom, one minus true-null residual minus removed share. Cone-beam rows: scans 0–5; 2D rows: held-out images. Weights column: verdict of Table 25 at 400 steps.

	$\cos\theta$ by radial band (Nyquist)				removed (%)				
Variant	< 0.1	0.1–0.25	0.25–0.5	> 0.5	< 0.1	0.1–0.25	0.25–0.5	> 0.5	$\cos\theta$
Lung, M=8*, forty steps*, share of $\|e_n\|^2$ by band 15 / 34 / 28 / 24%									
feature volume	0.84	0.74	0.55	0.30	45	32	16	4	0.61
per-view 2D features	0.84	0.72	0.48	0.22	44	30	12	2	0.58
both tensors	0.87	0.77	0.58	0.30	50	35	18	4	0.64
weights	0.77	0.64	0.41	0.20	36	22	9	2	0.53
Lung, M=8*,* 400 *steps*, share of $\|e_n\|^2$ by band 15 / 34 / 28 / 24%									
feature volume	0.85	0.73	0.58	0.37	47	31	19	7	0.64
weights	0.88	0.74	0.54	0.28	53	33	16	4	0.64
Lung, I_0=10^5*,* 400 *steps*, share of $\|e_n\|^2$ by band 23 / 38 / 21 / 18%									
feature volume	0.74	0.56	0.34	0.17	33	15	2	-14	0.49
weights	0.93	0.81	0.55	0.29	62	41	16	-3	0.70
Tooth prior on lung, forty steps, share of $\|e_n\|^2$ by band 31 / 40 / 18 / 11%									
feature volume	0.83	0.72	0.59	0.50	44	30	18	12	0.71
weights	0.86	0.78	0.65	0.54	49	37	24	15	0.75
Lung, M=4*, forty steps*, share of $\|e_n\|^2$ by band 32 / 39 / 17 / 11%									
feature volume	0.67	0.59	0.41	0.23	25	19	9	3	0.59
weights	0.77	0.66	0.44	0.25	35	24	10	3	0.66

Table 21: **Alignment by radial band.** $\cos\theta$ and removed share of null-space error per band of $|k|$ (Nyquist units), pooled over scans 0–5, each frequency placed within a 32^3 patch. Last column: global $\cos\theta$ of the same volumes.

R Applying the criterion without labels

Only one quantity in Corollary 3 is unavailable at deployment. Let s denote the share of the frozen error's energy that lies in the null space, $s = \|e_n\|^2/(\|e_n\|^2 + \|e_r\|^2)$. Then $\|e_n\| = \|e_r\|\sqrt{s/(1-s)}$ holds exactly, and both $\|d_n\|$ and $\|e_r\|$ are available without labels: the first is the null-space projection of the update, the second follows from the measurement residual, and the projector itself follows from the acquisition geometry. The relative length r is therefore known up to s. We calibrate s once on the backbone's own pretraining distribution, $s = 0.81$ over 10 AbdomenAtlas scans, and transfer it. Wherever the frozen model already fits its measurements, s is a property of the *operator*: it measures 78–81% across two anatomies, two view counts and the real scans in this paper, so the transfer changes r by under 5%. With the tooth prior it collapses to 19%, but that case is easy to recognise, because its measurement residual is 4 times larger than in any other condition here. The quantity that remains unavailable is $\cos\theta$, because it requires the direction of an error that cannot be observed. Across every variant and setting on the two cone-beam backbones it lies in $[0.38, 0.75]$, and on the other operators it ranges from near zero on the pixel mask to 0.85 on deblurring. The observed range of $\cos\theta$ is therefore a property of the operator and the backbone, and it is read off their own decompositions. A label-free application of Corollary 3 is therefore one-sided but usable. A value of r above the observed range of $\cos\theta$ is past the minimum of the error identity for *any* $\cos\theta$ in that range, and 3 of the 5 reversals have such an r (Appendix J). What this test detects is a variable that has overshot its own upper bound. For that reason it can size a variable without ground truth: at the width it selects, the RDC is better than the weight-space variant on the spine split and worse on the real scans, where the advantage of the weight-space variant is one of alignment (Appendix U). The test does not detect the reversal itself, because every one of the 5 reversals is decided by the alignment, and with the tooth prior the check triggers while the variant is still within 98% of its own upper bound. The share of the update that lies in the null space is not a proxy for its direction: it is flat across the sweep over variables, at 53 to 57%, while the error the update removes ranges from 15 to 22% (Figure 4b). Withheld views do provide such a proxy, and they form the second half of the label-free selection procedure.

Choosing the variant with withheld views. We fit both variants to $M - k$ of the views, with $k = M/4$ and the same views for each, and rank them by the residual on the k views that neither variant saw. A withheld view measures part of the component that the fitted views do not constrain, so unlike the fitted residual this quantity is not blind to the null-space component by construction. It orders the variants correctly on 3 of 3 settings and on 18 of 18 scan–setting pairs, including the

two reversals (Table 22). The control is the same comparison made on the fitted views, which is the quantity that both variants are minimising. It is correct on 1 of 3 settings and 6 of 18 scan–setting pairs, and it misses both reversals on every scan. Withholding views reduces the PSNR of the RDC by up to 0.55 dB relative to its run on all views, and that of the weight-space variant by up to 0.18. The withheld-view run is therefore a diagnostic run rather than the run whose output is kept. The withheld residual does not size a variable. On the sweep over widths it falls monotonically as the variable narrows, and it selects the narrowest width offered (Appendix U); the widths in that sweep differ in how tightly they fit the views rather than in alignment. The two label-free quantities therefore answer different questions: r sizes a single variable, and withheld views choose between two variants. Neither is used outside the question it answers.

Table 22: **Choosing the variant without labels.** Both variants fit the same $M-k$ views, $k=M/4$, matched protocol; ranked by mean residual on the k withheld views. PSNR columns judge the ranking and enter no decision; fitted residual: the control. Six scans per setting; last two settings: the reversals (weight-space variant better).

	PSNR (dB)		held-out residual		fitted residual		
Setting	volume	weights	volume	weights	volume	weights	picked
Lung, M=8	**27.84**	27.39	2.45e-06	2.87e-06	4.22e-07	2.38e-06	volume
Tooth prior on lung	23.08	**23.89**	2.95e-05	1.55e-05	8.78e-06	1.42e-05	weights
Lung, M=4	23.44	**24.30**	2.33e-05	9.92e-06	4.63e-07	4.07e-06	weights

What can be learned from measurements alone. The selection procedure is one-sided because $\cos\theta$ requires the direction of an error that no measurement of the scan contains. The sensing theorems of Tachella et al. (2023) state how far learning from measurements alone could go instead. Under a single incomplete operator the signal set is not identifiable inside $\mathrm{null}(A)$, however many scans are observed. This is the fixed protocol of Table 1, in which every scan shares one operator. With $|\mathcal{G}|$ distinct operators, identification requires $m \geq n/|\mathcal{G}|$ measurements per signal, and it holds for almost every set of operators once $m > k + n/|\mathcal{G}|$, where k is the dimension of the signal set. Here $n = 16.8M$ voxels and $m = 524k$ at eight views, so $n/m = 32$. The jittered protocol provides 25 distinct gantry phases on LUNA16 and 6 on PENGWIN, which is below the 32 that the necessary condition requires; at four views the required number is 64. Learning the null-space component from the measurements in this paper alone is therefore ruled out on both protocols by this count, before any algorithm is chosen. The remaining measurement-only route is a group invariance, which is what equivariant imaging assumes and what its column in Table 1 yields at this step budget. Where labels or a second modality exist, the count does not apply, and Appendix Q measures how much of the null-space error then remains to be learned.

S THE SELECTION PROCEDURE, BACKTESTED AND THEN APPLIED BLIND

The measurements on the first three backbones and operators leave a question that a method paper has to answer: given a released reconstructor and one measurement, which tensor should be freed, for how many steps and at what learning rate, without a sweep over every candidate. This section states the label-free selection procedure that those measurements support, backtests it on the same three backbones, and then applies it, with the choice written down before each run, to the networks and operators that were added afterwards.

The procedure. *Variable.* Free the deepest activation whose number of free entries does not exceed the number of image voxels. The deepest activation is chosen because the alignment is set by the frozen layers between the variable and the output, and a variable followed by only two convolutions does not move the null-space error at all. The size limit is imposed because a variable with more free entries than the image overshoots the minimum of the error identity (Appendix U). Where the mapping learned by the frozen decoder is wrong for the target, for example a contrast the decoder never saw or a prior trained on another anatomy, adapt the weights instead (Table 25). *Budget.* Use forty steps, which is where every noise-free activation trajectory we have measured has flattened, and under noise forty steps is the stopping rule; where the frozen residual is already below the noise level, the loss is not used at all. *Rate.* Run one learning-rate sweep on eight samples per backbone and apply the result unchanged to the remaining samples. The activation-space variants lose their

entire gain at a rate a factor of three above their optimum, so the sweep never extends beyond that point.

Alignment by depth. The variable clause rests on Table 24, which lists the forty-step alignment of every activation we ran against the frozen layers between that activation and the output, for one reference condition per network.

Depth and kind are separable here, and separating them is a control the weight-space arm has not had: it frees all 6.04M parameters, the encoder and codebook among them, and those lie before the fused feature volume rather than between it and the image. Freeing only the 5.44M that do lie between them, at its own swept step size of 3e-5, reaches 27.40 dB against 27.39 for the full arm on the same 5 scans, a difference of +0.005 dB, better on 3. What the weights buy is therefore bought entirely behind the feature volume; the parameters in front of it contribute nothing. Freeing the point decoder alone, 0.18M with almost nothing frozen between it and the output, reaches 26.07 at its own best rate, -1.326 dB and better on 0 of 5. That subset is both the shallowest and the smallest, so it does not separate depth from capacity on its own; the 5.44M subset does, and it holds +0.005 dB of the full arm at a tenth of the distance from the output. An activation followed only by a head, by an upsampler or by the skip connections of a bottleneck does not move the null-space error at all. On the pixel mask no variable has a non-zero alignment, and the check of the procedure triggers in that case. Applied to a new architecture, the procedure replaces the search itself, because it names one activation from the structure of the network alone, as it did for the transformer.

The backtest. Table 23 lists every variable we ran, together with the frozen layers between it and the output, its size relative to the image, its gain at forty steps at its chosen rate, the factor over which its rate can vary while staying within 0.1 dB of the best, and the step at which its 400-step trajectory peaks. The variable clause picks the measured best variable on 3 of 5 backbones. On the cone-beam backbone it picks the input projections, which is the deeper of the two variables that fit within the size of the image. On the ten sweep scans of the lung row the fused volume is 0.28 dB better than that pick, and across the nine settings of Table 1 the pick is better on 5 rows and the fused volume on 4 (Appendix G). The rate-tolerance column explains why the rate clause is a sweep rather than a constant. The variables that tolerate a wide range of rates are the ones that achieve the least, and every variable that moves the null-space component has a rate that must be found once.

Applied blind. The fourth network is the released SwinIR (Liang et al., 2021) for 4× super-resolution, a Swin transformer with six residual groups. The fourth operator is inpainting through a U-Net trained on DIV2K with rectangular holes; its pixel mask has a null space that is a spatial region rather than a band of frequencies. For both, the pick of the procedure, namely the input tensor, was written down before any run. Three predictions were recorded with it: that the late feature does not move the null-space component, that the early feature at native width overshoots and a narrower grid improves on it, and that the weight-space variant converges to a lower alignment than the activation feature on a network trained to reproduce its measurements. On SwinIR the pick of the procedure is the best variable on every held-out image but one. At forty steps on BSD100 it gains +1.37 dB, against +0.86 for the weight-space variant and +0.66 for the early feature, and the weight-space variant converges to a lower alignment (labelled below in Table 25). On this transformer the early feature has zero alignment both at native width and when narrowed, so the two predictions about it do not hold. On the pixel mask no variable has a non-zero alignment (Appendix X); the error identity then indicates that no adaptation should be performed, and the pick of the procedure is moot.

A fifth operator tested the procedure a third time. Deblurring through the released NAFNet leaves a residual of +10.99 dB, so the adaptation is run, and the procedure picks the blurred input, as it did on the other natural-image networks. On all 5 conditions the weight-space variant is better at forty steps: it gains +2.13 against +2.01 dB on the rate-sweep condition and +6.97 against +5.47 on the IXI slices. The pick of the procedure is better than the other activation variables on four of the 5 conditions, and the error identity ranks the weight-space variant above it, which settles that comparison (Appendix Y). The procedure therefore sizes a variable and sets its step budget and rate, while the alignment decides between an activation variable and the weight-space variant. Table 23 contains all three tests.

Variable	Frozen after it	Entries/image	Rate	Gain at 40	Tolerance	Peak
Cone beam (DeepSparse)						
weights	everything	–	3e-05	+2.07	1×	367
input projections⋆	encoder to decoders	0.03	0.0003	+2.49	1×	–
fused feature volume	decoders	0.25	0.02	+2.78	1×	252
MRI (U-Net), 8×						
weights	everything	37.87	3e-05	+0.34	3×	400
bottleneck	decoder half	1.00	0.1	+0.43	10×	389
input image⋆	whole network	1.00	0.003	+0.90	1×	339
SR (EDSR), Gaussian kernel						
weights	everything	7.72	0.0001	+0.75	3×	114
late feature	upsampler	1.33	0.01	+0.59	30×	–
early feature	body, upsampler	1.33	0.003	+1.23	1×	71
LR input⋆	whole network	0.06	0.01	+1.31	10×	226
SR (SwinIR), Gaussian kernel						
weights	everything	60.53	0.0001	+0.75	1×	49
late feature	upsampler	3.75	0.001	+0.58	3×	–
early feature	body, upsampler	3.75	0.003	+0.63	3×	54
LR input⋆	whole network	0.06	0.01	+1.39	1×	173
Inpainting (U-Net), rectangles						
weights	everything	18.81	1e-05	+2.72	1×	51
bottleneck	decoder half	0.33	0.03	+0.00	30×	–
early feature	encoder, decoder	10.67	0.001	+0.25	1×	52
masked input⋆	whole network	1.00	0.0003	+0.43	1×	91

Table 23: **Backtest of the selection procedure.** Every variable run at its chosen global rate: frozen layers to the output, free entries relative to the image, forty-step gain, rate tolerance (factor within 0.1 dB of the best) and mean step of the 400-step peak. Starred: procedure's pick.

T BOTH VARIANTS AT CONVERGENCE

	$\cos\theta$ by the frozen layers between the variable and the output				Pairs
Network, condition	Weights	Deepest activation	Early activation	Input	ordered
DeepSparse, lung M=8	0.53	0.61 (feature volume)	–	0.59	10 of 10
U-Net, T2 4×	0.27	0.29 (bottleneck)	–	0.43	11 of 12
EDSR, Gaussian kernel	0.17	0.01 (late feature)	0.30	0.34	12 of 12
SwinIR, Gaussian kernel	0.19	–	0.16	0.36	6 of 6
NAFNet, motion blur	0.46	0.13 (middle block)	0.41	0.42	71 of 75
Inpainting U-Net, rectangles	0.04	-0.02 (bottleneck)	0.06	-0.03	–
Total					**110 of 115**

Table 24: **Alignment by depth at forty steps.** $\cos\theta$ per variant on one reference condition per network; columns ordered by frozen depth between variable and output, least to most. Last column: pairs of variants, over every condition of that network, whose ordering by ρ has the sign of their PSNR ordering; inpainting excluded (every alignment zero). Values from Tables 32, 33, 35 and 37.

Every comparison in the main text is made at the matched forty-step budget. In this section we run both variants until they stop improving, both on the DRR grid and under the noise level the backbone was trained with, and we decompose the volume that each variant converges to. The result shows what the forty-step budget yields and what it gives up.

Budget curve on DRRs. Figure 7 plots reconstruction quality against wall time. We stop at 40 steps, which reach 87% of the quality that 200 steps reach at four times the compute. After 20 of its 40 steps RDC already reaches 27.24 dB on the main lung row, whereas the weight-space variant reaches 27.14 after all forty; RDC is better on 79% of the 66 scans. The weight-space variant needs 141 steps and 5.7× the wall time to reach our forty-step quality. At 400 steps the gap narrows to +0.11 dB, with the same null-space error for both variants (see below).

Convergence on DRRs. Table 26 reports the first 6 scans of the main lung row, with both variants run for 400 steps under the matched protocol (every view in the loss, the final iterate, and the global

		$\cos\theta$				Null drop (%)		Gain (dB)		
Condition	Shift	Weights	Activ.	Gap	Verdict	Weights	Activ.	Weights	Activ.	n
Cone-beam CT, DeepSparse; activation variable: the feature volume										
Lung	Anatomy	0.64	0.64	+0.00	meet	23	23	+3.05	+3.18	6
Tooth prior	Prior map	0.87	0.80	+0.07	above	51	40	+13.53	+11.74	6
Cartesian MRI, U-Net; activation variable: the input										
T2, 8×	Mask	0.33	0.37	-0.04	meet	6	6	+0.74	+0.86	24
T2, random mask	Mask	0.47	0.68	-0.21	below	11	27	+1.69	+3.92	24
T2, 4×	None	0.31	0.45	-0.14	below	-6	7	-0.09	+1.39	24
T1	Contrast	0.61	0.42	+0.19	above	21	2	+3.03	+1.36	24
T2, inverted	Contrast	0.67	0.30	+0.37	above	25	-21	+3.34	-0.54	24
Super-resolution, EDSR; activation variable: the input										
BSD100	Kernel	0.15	0.34	-0.20	below	1	7	+0.79	+1.38	92
Urban100	Kernel, content	0.22	0.46	-0.23	below	3	14	+1.10	+2.19	92
IXI slices	Kernel, content	0.32	0.51	-0.19	below	5	14	+1.90	+2.73	32
BSD100, inverted	Kernel, contrast	0.15	0.32	-0.18	below	1	7	+0.79	+1.31	92
Super-resolution, SwinIR; activation variable: the input										
BSD100	Kernel	0.14	0.37	-0.23	below	1	8	+0.78	+1.46	52
Urban100	Kernel, content	-0.01	0.55	-0.56	below	-4	20	+0.54	+2.89	52
Deblurring, NAFNet; activation variable: the input										
BSD100, motion	Kernel	0.65	0.57	+0.08	above	26	18	+6.61	+5.09	52
BSD100, 2× blur	Kernel	0.76	0.74	+0.02	meet	40	36	+8.43	+6.97	52
BSD100, box	Kernel	0.53	0.53	+0.00	meet	15	17	+4.03	+3.61	52
Urban100	Kernel, content	0.85	0.72	+0.13	above	53	34	+10.50	+6.50	52
IXI slices	Kernel, content	0.96	0.92	+0.04	meet	71	60	+13.67	+9.64	32

Table 25: **Alignment of both variants at convergence.** Both run for 400 steps. Gap: weight-space $\cos\theta$ minus activation-space $\cos\theta$. Within 0.05 (fixed in advance): meet; otherwise above or below gives the side on which the weight-space variant ends. Verdicts: 5 meet, all within 0.04; 5 above, by 0.07 or more; 8 below, by 0.14 or more.

step sizes of Table 1). Our variant peaks near step 253 and is flat afterwards, whereas the weight-space variant is still gaining +0.10 dB per hundred steps at step 400. Our margin falls from +0.74 dB at forty steps to +0.11 at 400 steps, where RDC is better on 5 of 6. The weight-space variant first reaches our forty-step quality at step 141. At 8.1 s per step, against our 5.05, this corresponds to 1142 s against 202, or 5.7× the wall time. Over ten scans run to 200 steps the same margin is +0.74 to +0.15 dB (better on 8 of 10), and the margin over equivariant imaging is +0.46. Figure 7 plots both curves against their own wall time.

Steps	40	100	200	300	400
RDC (feature volume)	28.15	28.47	28.55	28.54	28.52
Weights	27.42	27.99	28.28	28.32	28.42
Lead	+0.73	+0.48	+0.27	+0.22	+0.11

Steps	40	100	200	300	400
RDC (feature volume)	25.49	25.34	24.83	24.45	24.15
Weights	25.33	25.91	26.04	26.10	26.05
Lead	+0.16	-0.57	-1.22	-1.65	-1.90

Table 26: **Both variants run to** 400 **steps.** PSNR (dB), LUNA16 at M=8, scans 0–5, matched protocol; frozen model 25.35. Left: DRRs. Right: Poisson noise at I_0=10^5 on the projections (network input and loss); frozen model 21.94.

Alignment at convergence. Table 27 decomposes the 400-step volumes with the exact projectors. On DRRs the two variants remove the same fraction of the null-space error, 23% against 23%. They do so at the same alignment between the null-space update and the null-space error, with $\cos\theta$ of 0.64 against 0.64. The two variants therefore share the upper bound of Section 5.3, and both attain it; the error identity holds for both (0.766 and 0.770). What separates the variants at 400 steps is the *range-space* component, 14.5 against 27.4 from a frozen value of 95.8, and this accounts for

the whole of the remaining +0.11 dB. The forty-step sweep of Figure 4b, in which the weight-space variant recovers less, is therefore a statement about how far each variable has progressed at forty steps. The alignment that the loss can reach is set by the tangent kernel of the frozen decoder, whichever tensor is free; the choice of variable sets the relative length of the null-space update and the number of steps needed to reach it.

	Free variable	$\|P_{\mathcal{R}}e\|$	$\|P_{\mathcal{N}}e\|$	null err.	$\cos\theta$	r	vs. oracle
DRR	**RDC**	96→14.5	200→153.7	−23%	0.64	0.65	+2.15
	weights	96→27.4	200→154.3	−23%	0.64	0.59	+2.02
I_0=10^5	**RDC**	182→58.1	273→248.7	−9%	0.49	0.75	+0.37
	weights	182→61.1	273→196.5	−28%	0.69	0.71	+2.26

Table 27: **The decomposition at** 400 **steps**, same scans and projectors as Table 12. Norms before→after; null err. is the change in $\|P_{\mathcal{N}}e\|$; vs. oracle is the margin over the frozen model's range oracle.

Under noise. The right half of Table 26 repeats the run with Poisson noise at I_0=10^5, the level under which the backbone was pretrained. At forty steps the two variants are level, at 25.49 against 25.33. Our variant peaks at step 50 (25.53) and then begins to fit the noise, falling by 1.38 dB by step 400; the weight-space variant, at its own step size, rises to 26.05 and stays there. The decomposition explains this in the terms of the error identity. At 400 steps the null-space update of RDC is too long for its alignment: its relative length r is 0.75 against a $\cos\theta$ of 0.49, and it removes 9% of the null-space error. The weight-space variant has r of 0.71 against $\cos\theta$ of 0.69 and removes 28%. The predicted values of ρ are 0.909 against 0.718, and their ordering has the same sign as the outcome. This is the case in which the step-length term $(r-\cos\theta)$ of the error identity (Section 5.3) dominates. Here it is produced by running past the step budget rather than by widening the variable, and it is the one setting in the paper where the feature volume needs a stopping rule and the weight-space variant, within 400 steps, does not. The forty-step budget we report lies before this peak. The robustness table (Appendix H) keeps the published protocol for its weight-space column, so its noise rows do not contain this comparison; at the matched forty-step budget the margin under this noise is +0.16 dB.

Rate sensitivity. Table 28 reports the two step-size sweeps that fixed the global step sizes, on the same ten scans at forty steps. Over a 16× range of step sizes RDC never diverges, and its worst configuration is 0.87 dB below its best. The weight-space variant diverges on the first step at 10× its optimal step size, where the stopping rule of the published protocol returns the frozen model. A deployment site cannot sweep a step size for every scanner. This is the practical form of the step-length term of the error identity: a variable whose update is a bounded activation tolerates a step size that the network's weights do not.

RDC, rate	0.005	0.01	0.02	0.04	0.08
gain (dB)	+2.01	+2.48	+2.78	+2.88	+2.66
Weights, rate	1×10^{-5}	3×10^{-5}	1×10^{-4}	3×10^{-4}	1×10^{-3}
gain (dB)	+1.67	+2.00	+1.44	+0.00	+0.00

Table 28: **Step-size sweeps**: gain over the frozen model at forty steps on ten lung scans. The weight-space variant runs under its published protocol; +0.00 means the stopping rule returned step 0 after divergence.

U A SECOND BACKBONE

Every other number in this paper comes from one released model, so the mechanism argument is stated for a class of networks but tested on a single member. This section runs the same comparison on a second reconstructor. We chose it because it differs from DeepSparse in the two respects that would matter if the effect were an artefact of DeepSparse's design rather than a consequence of the choice of variable.

The model of Liu et al. (2025) comes from a different research group and reverses both design choices. Its internal volume holds 67.1 M entries against the image's 16.8 M voxels, that is, four

times as many, whereas the internal volume of DeepSparse holds a quarter as many. The compression that might be credited with the effect is therefore absent. Its decoder is a 3D convolutional generator rather than a point-queried MLP through a vector-quantised codebook, so the mapping from the internal volume to the image belongs to a different family. It has 27.5 M parameters, against roughly six for DeepSparse.

We take its released checkpoint and use its own forward operator, its own data-consistency loss taken from its training code, its own normalisation and its own metric; we change only which tensor is free. The checkpoint was trained on dental scans. On the spine split its frozen reconstruction starts at 18.72 dB, against 27.17 in domain, which is a gap of 8.45 dB.

At the forty-step budget of the main grid, and under this backbone's own metric, RDC gains +4.13 dB over the frozen model, against +0.22 for the weight-space variant at a per-scan step size chosen with ground truth; RDC is better on 6 of 6 scans. This result is superseded below, because forty steps is not a converged budget on this backbone, and the converged decomposition reverses the ordering.

Null-space movement on this backbone. Against its own operator, both the weights and the internal volume of this model exceed the range oracle and reduce their null-space error, by 26% and 15% respectively. A measurement-domain loss that moves the null-space component is therefore a property of the setting rather than of the backbone we chose. The rest of this appendix places this backbone on the same curve of margin against variable size as our own backbone, and narrows its variable.

Height of the range oracle. This model is evaluated under its own protocol of ten views over a full circle, where 28% of the frozen error lies in the null space. Our protocol uses eight views over a half circle, where the corresponding fraction is 81%. The range oracle is given the whole range-space component, so it removes about three quarters of the error energy under this model's protocol and about a fifth under ours. A margin over the oracle therefore does not transfer between the two protocols. This is a difference in the acquisition rather than in the model, and it must be held fixed before anything below is read as a statement about architectures.

Transfer of the ordering. Running the decomposition of Section 5.3 against this model's own operator gives Table 29. Both parameterisations exceed the range oracle and both reduce their null-space error, RDC by 15% and the weight-space variant by 26%. A measurement-domain loss that moves the null-space component is therefore not a property of DeepSparse: the effect persists under a different research group, a different decoder family and a different anatomy. At the released configuration, however, the weight-space variant is the more effective variable on this backbone, with +1.83 against our +1.02 over the oracle. This is the reverse of the ordering on our own backbone.

The error identity of Section 5.3 indicates where to look. The identity is minimised at $r = \cos\theta$, so $\cos\theta$ bounds what any trajectory can reach and r measures how much of that bound is used. Our variant has $r = 0.855$ against $\cos\theta = 0.595$, which overshoots its own optimum by more than forty percent, whereas the weight-space variant has r of 0.640 against $\cos\theta$ of 0.668 and is essentially at its optimum. In both settings where the weight-space variant is better on this backbone, our null-space update is also too long. The free variable is the obvious cause, because this backbone optimises 67.1 M entries, four times the number of image voxels, whereas DeepSparse optimises a quarter of that number. The relative step length does not decide either comparison, since RDC placed exactly at its own upper bound would still be below the weight-space variant here (Section 5.3). It is, however, the part of the deficit that a choice of variable width can recover, and the rest of this section recovers it.

Overshoot and dimension. This model lies on the far side of the curve, at $r = 0.855$ against $\cos\theta = 0.595$, and narrowing its variable reverses the ordering: RDC reaches +2.55 over the oracle with 64× fewer entries. The two backbones therefore lie on one curve, measured from opposite ends. We restrict the free variable to an R^3 grid, by the same construction as the sweep of Section 5.3, with the decoder, the loss, the step size, the step budget and the scans all fixed. This moves r down monotonically across a 512× span of entry counts (Table 30), as we predicted before the sweep was run. At the same time $\cos\theta$ rises from 0.595 to 0.708. Constraining the variable therefore makes the null-space update both shorter and better aligned, which is why the ordering reverses. At

$R = 16$ RDC reaches +2.55 over the oracle, and its null-space error falls by 29%; it outperforms the weight-space variant on both measures.

The optimum is interior, so the sweep shows a peak rather than a preference for small variables. The released configuration is 64× above the optimum in entry count, and this accounts for the whole of the deficit with which this section opened. It also explains why the sweep in Section 5.3 shows only a rising limb: the native variable of our own backbone is a quarter of the image and never reaches the far side of the peak, so compressing it can only reduce quality. The two sweeps should be read together as one curve measured from opposite ends.

Label-free width selection. We report the ordering at the released configuration as well as at the optimum, because R in Table 30 is chosen with ground truth, exactly as the per-scan step size of the weight-space variant is throughout this section. The obvious label-free alternative is to withhold two of the ten views from the loss and to rank the configurations by the residual on those two views. That residual falls monotonically as the variable is narrowed, 0.88, 1.10, 1.26 and 1.37 in units of 10^{-4} from R=8 to R=64. It therefore selects the narrowest configuration of whatever grid it is offered, R=8 here, whereas the reconstruction is best at R=16. What separates the configurations lies in the null space, which these measurements cannot observe: the narrower configuration fits the range-space component slightly better and recovers less of the null-space error. Between two variables, rather than two widths, the same held-out residual does work, and Appendix R uses it for that purpose. This is the premise of the paper, encountered from the other direction.

The one-sided criterion derived from the error identity (Section 5.3) gives a label-free selection procedure that does not need the withheld views. The relative length r can be computed without labels up to the null-space share s of the error. This backbone has no source-domain decomposition, so here, and only here, s is calibrated leave-one-scan-out on the target scans, which changes r by at most 7%. An r above the observed range of $\cos\theta$ is past the minimum of the identity for any $\cos\theta$ in that range. Taking the widest configuration whose label-free r does not exceed the upper edge of that range, 0.75, selects R=32. At this width RDC reaches +2.29 dB over the range oracle, against +1.83 for the weight-space variant, and is better on 4 of 6 scans, with no ground truth used in the choice. Selecting per scan instead gives +2.30, better on 5 of 6. The selected width is an interior configuration rather than an edge of the grid, and the choice is not sensitive to the two constants it depends on: any threshold from 0.60 to 0.85 and any s within ten percent of the calibrated value select a configuration that outperforms the weight-space variant, and only a threshold that admits the native r=0.855 falls back to the released width. The procedure does not find the optimum, which is +2.55 at R=16, where the feature volume outperforms the weight-space variant on every scan. It avoids the overshooting configuration, which is what a one-sided test can do, and stops short of the peak, which would require $\cos\theta$. We report the width as chosen with ground truth in the tables and as chosen by this procedure here, and we do not mix the two.

The selection procedure on the real scans. We run the same sweep on the real projections reported below, where the margin of the weight-space variant is larger. Narrowing the variable moves both terms in the right direction there as well. From R=64 to R=8, r falls from 0.88 to 0.64 and $\cos\theta$ rises from 0.52 to 0.61. The margin over the range oracle rises from +0.48 to +1.73 dB and is still rising at the narrowest configuration. The procedure selects R=16, at +1.29 dB. This is an interior configuration that outperforms the released width but is below the weight-space variant's +2.55 on every one of the 5 scans. The selection moves with the threshold: it is one configuration narrower at 0.70 and one wider at 0.80. What separates the narrowest configuration from the weight-space variant here is the alignment, 0.61 against 0.70, which the procedure cannot observe. On this split the procedure sizes the variable, and the remaining gap to the weight-space variant is the one that the alignment dictates (Table 14).

Three protocol mismatches had to be removed before this comparison measured the parameterisation rather than the domain transfer. Each is settled by a property of the run that can be checked without computing a margin at all, which is what distinguishes correcting a protocol from tuning until a number comes out favourably.

First, RDC fits every measured view here, as it does on our own backbone, where only the baselines hold two views out for a stopping rule. Holding two views out of our loss and then decomposing against all ten had scored RDC against an oracle that was given the ground truth of two views the

loss never saw; this is visible from the protocol alone. Second, both variants are decomposed at their ground-truth-best step size. This is more than either could have in deployment, but it is the only arrangement that isolates the parameterisation from the tuning. A sweep puts the optimum of this backbone at 0.01 rather than at the 0.02 our main results use, on all six scans, and stepping past an optimum places movement where the loss cannot observe it.

The third mismatch is the step budget. On DeepSparse the cost curve is flat by forty steps. Here, in contrast, every one of the six scans is still gaining half a decibel over its last nine steps at step forty, and the fit loss is still falling by a third. A decomposition taken at forty steps therefore measures the budget rather than the backbone, and the trajectory shows this without reference to any margin. At 400 steps the sign reverses on all six scans, and it does so monotonically across every intermediate budget we ran, which distinguishes a convergence effect from a coincidence. The projector is converged independently of this: the same volumes decomposed with 300 conjugate-gradient iterations rather than 120 agree to within 0.01 dB on 4 scans.

Two further controls bear on the mechanism itself. The first concerns the prior. On this anatomy the checkpoint is 8.45 dB outside its training domain, which is far outside anything in Section 5. The mechanism depends on the frozen decoder moving the null-space component in a direction that the anatomy supports, and a dental prior plausibly does not do this for a spine. Appendix P varies the prior directly over a 16.7 dB span of frozen quality, and the margin stays positive at every configuration, so the effect does not require the prior to be good.

The second control concerns the size of the variable, and it is the sharper of the two. We optimise each internal volume in image space, with no measurement operator in the loss, towards a random direction in null(A) scaled to the frozen model's own null-space error. Neither parameterisation moves: the residual stays at 0.9997 of the offset on this backbone and at 0.9990 on ours. Having four times as many free entries as image voxels yields no additional freedom, because the image of the decoder is not surjective merely because its domain is large. The same optimisation towards the ground truth, a target that the prior does support, moves to 0.55 on this backbone and to 0.37 on ours. On our backbone, the fit towards the null-space component of the ground truth alone, which is the complement of the range oracle, moves to 0.45 over 6 scans (0.37 to 0.55). The null-space component of the target therefore lies inside the reachable set, whereas the random direction does not. Both parameterisations are thus selective rather than merely constrained. This is the property that the mechanism requires, and neither the entry count nor the decoder family predicts it. Why the weight space is the better variable on this backbone while the internal volume is better on ours does not follow from any measurement we have, and we leave it open.

At the forty steps used by our main results, before the cost curve of this backbone had flattened, RDC scored -1.64, with its null-space error *growing* by 8%. The table below is therefore taken at 400 steps.

Free variable	vs. range oracle (dB)	Null-space error
Network weights	+1.83	-26%
Internal volume (RDC)	+1.02	-15%
RDC on DeepSparse (Sec. 5.3)	$+1.72$	-20%

Table 29: **The decomposition on the second backbone**, 6 scans, both variants at their ground-truth-best step size and 400 steps; $P_{\mathcal{R}}$ by 120 conjugate-gradient iterations.

The weight-space variant has no counterpart to the width restriction; its values, +1.83 and -26%, are the same throughout the sweep below.

Real projections. Every other measurement in this paper uses DRRs, so we ran this backbone on the one sparse-view cone-beam dataset whose projections are real rather than simulated: a walnut scanned on a laboratory micro-CT. There RDC exceeds the range oracle by +0.40 dB on 5 of 5 scans, and its null-space error falls by 7%. The mechanism this paper studies is therefore not an artefact of simulated projections. The weight-space variant is again better, at +2.55 and 28%, which is the same ordering this backbone gives on spine and the opposite of the ordering on our own backbone.

The control experiment is the most important result of this section. Replacing the measurements with Ax_{gt}, rendered through the same rays of the same real geometry, and changing nothing else, moves the margin by 0.07 dB. Scatter, beam hardening and calibration error, that is, everything a

R	8	16	32	64
Free entries (M)	0.13	1.05	8.39	67.11
Fraction of image voxels	0.008	0.062	0.500	4.000
$\cos\theta$	0.699	0.708	0.685	0.595
$r = \|d_n\|/\|e_n\|$	0.675	0.720	0.744	0.855
Null-space error	-28	-29	-27	-15
vs. range oracle (dB)	+2.34	+2.55	+2.29	+1.02

Table 30: **Width sweep of the free variable on the second backbone.** Our variant with the free variable restricted to an R^3 grid, 6 scans, 400 steps, everything else fixed; R=64 is the released configuration.

DRR omits, together account for less than a tenth of a decibel here. This number is a property of the operator of this backbone, which includes the scanner's per-view calibration. Appendix J processes the same scans through the idealised circular model of the primary backbone, where the same control moves the margin by 1.75 dB.

Two protocol choices matter here. First, RDC fits every view, as it does everywhere else. Holding two of the eight views out for a stopping rule, which is the split the baselines need, removes one direction, because over a full circle the withheld pair is separated by 180°; this costs 0.83 dB. Second, the comparison is made at the converged budget. At forty steps RDC is negative at all seven step sizes swept, and it turns positive on every scan by step 400. Both variants are therefore swept at convergence, where both optima are interior, RDC's at 0.005 and the weight-space variant's at 3e-05, and neither is pinned to the edge of the grid. The backbone's own metric rescales each volume independently, and the oracle comparison is not defined under it; both metrics are given in Table 31, and they agree on the ordering.

The scans cover a full circle, and each variant runs at its interior optimum over 7 swept step sizes. Two remarks concern how the table below should be read. The first row uses the metric of the forty-step result above, and the fixed-range rows are the only ones under which the oracle comparison is defined. The synthetic row changes nothing but the measurements, so the difference between it and the row above it is due to scatter, beam hardening and calibration error.

	Internal volume (RDC)	Network weights
Gain over frozen (dB)	+1.14	+3.27
vs. range oracle (dB)	+0.40	+2.55
Null-space error	-7%	-28%
Synthetic Ax_{gt}, vs. oracle	+0.47	–
Synthetic, null-space error	-7%	–

Table 31: **Real cone-beam projections and the forward-model control.** Walnut micro-CT, 5 scans at eight views, 400 steps. First row: the backbone's self-normalising metric; remaining rows: fixed-range metric. Synthetic rows use Ax_{gt} as the measurements.

One further caveat applies. The released checkpoint is not the one reported in its paper: its authors state that the original weights were lost and that the model was re-trained briefly. It should therefore be read as an instance of the architecture rather than as that paper's result.

We also explain why this backbone is not run under our own protocol. Its encoder was trained with views covering a full circle. On a half circle its frozen reconstruction falls to 7.46 dB and does not recover as views are added (7.46, 7.50 and 7.48 at eight, ten and twenty views). This is a violated assumption rather than a domain shift, so we report the backbone where it is defined rather than forcing it onto our grid.

Surrogate noise. The point-subsampled gradient surrogate of Appendix B is stochastic, so the null-space displacement could in principle be due to its variance. A random walk would give $\|d_n\| \sim t^{1/2}$. Over four step budgets on this backbone the measured exponent is -0.04, and $\cos\theta$ *rises* from 0.38 to 0.59, which a random walk cannot produce.

V A SECOND OPERATOR: CARTESIAN MRI

All experiments in the main body use a single forward operator, the three-dimensional cone-beam projector, and a single family of backbones. This section asks whether the mechanism depends on that operator or follows from the argument of Section 4. To answer this, we repeat the measurements on an operator that shares nothing with the projector, namely single-coil Cartesian MRI. Here $A = MF$, where F is the two-dimensional Fourier transform and M is a line mask, so that the range projector $P_{\mathcal{R}} = F^H MF$ is exact and no conjugate gradient solve is needed. All predictions were written down before any adaptation was run.

Setting. The data are IXI brain volumes (Biomedical Image Analysis Group, Imperial College London, 2008): 80 T2 subjects and 40 T1 subjects, with sixteen axial slices per subject taken around the volume centre at 320^2. The split is by subject, so that no test subject appears in training. The T2 test set contains 16 subjects (256 slices) and the T1 test set contains 8 subjects (128 slices). The measurement model assigns each magnitude image a smooth random phase, takes its Fourier transform, and keeps an equispaced set of lines together with a fully sampled centre band of 8% at $4\times$ acceleration and 4% at $8\times$. The frozen reconstructor is a feed-forward image-domain U-Net of the fastMRI baseline family (Žbontar et al., 2018) with 7.8M weights and four pooling levels. It was trained on the T2 training slices at $4\times$ with an ℓ_1 loss for forty epochs and reaches a validation PSNR of 35.97 dB. It plays the role that DeepSparse plays in the main body: a released prior that takes the zero-filled image and returns a reconstruction. There are four test conditions: T2 at $8\times$ (a shift of the mask), T1 at $4\times$ (a shift of the contrast), T2 at $4\times$ (in distribution), and T2 at $4\times$ with Gaussian k-space noise at 5% of the k-space rms.

Variables and protocol. We optimise three tensors, one per variant, under the same per-slice data-consistency loss $\|MF\hat{x} - y\|^2$ and the same optimiser. The first is the full set of weights. The second is the U-Net bottleneck, an additive tensor of 205k entries at its native width, which we also restrict to 10^2 and 5^2 grids as in Appendix O. The third is the zero-filled complex input image itself, with 205k entries, the same count as the bottleneck; the whole U-Net lies between the input image and the output, whereas only the decoding half lies between the bottleneck and the output. The protocol is that of Section 5: every measured line enters the loss, the budget is forty steps, the final iterate is kept, and one global learning rate per variant is chosen on eight sweep slices by the pooled mean gain over the two clean T2 conditions. The chosen rates are then applied unchanged to 248 held-out slices and to the two conditions that the sweep never saw. Each cell of the results table is decomposed with the exact projectors and compared against the frozen model's range oracle, as in Section 5.3. The weight-space and bottleneck variants and the input image are also run to 400 steps.

The variational network. The first backbone we trained for this operator was an end-to-end variational network with cascaded data-consistency layers (Sriram et al., 2020), which is the standard architecture for Cartesian MRI. Its frozen output already fits the measurements: the range oracle is +0.02 dB above the frozen model at $8\times$ and +0.20 at $4\times$, compared with +0.24 and +0.67 for the U-Net. Consequently no variant gains on it (the best gain is +0.05 dB at $8\times$). A test-time data-consistency loss has no residual to act on when the reconstructor has already fitted the measurements in its forward pass. The mechanism therefore requires a reconstructor whose range-space component is imperfect, which is the case for the feed-forward reconstructors used in the main body. The tangent-kernel path into the null space exists in both architectures, but only a residual can drive it.

In the table below, the forty-step rows of the four main conditions are averages over 248 T2 slices (120 slices for T1). The 400-step rows and the inverted and random-mask conditions are averages over 24 slices.

For the 400-step rows, the last column gives the mean step at which the trajectory peaks and the loss of the final iterate relative to that peak.

Results. Table 32 reports the full grid. Under the mask shift, every activation variable that passes through the decoder lowers the null-space error, and every variant exceeds the frozen model's range oracle, which no fit of the measurements alone can achieve. The bottleneck exceeds the oracle by +0.09 dB and the input image by +0.60. When run to 400 steps on that condition, the three variants converge to nearly the same gain, +0.74, +0.74 and +0.86 dB, with alignments $\cos\theta$ of 0.33, 0.35

Variable	Gain (dB)	Null drop (%)	vs. oracle (dB)	$\cos\theta$	r	Wins / peak, loss
T2, 8× *(mask shift)*, frozen 26.11 dB, range oracle +0.24						
Weights	+0.39	+2.7	+0.15	0.23	0.21	–
Bottleneck, native	+0.33	+2.7	+0.09	0.24	0.18	56/248
Bottleneck, 10^2	+0.17	+1.4	-0.07	0.18	0.10	6/248
Bottleneck, 5^2	+0.12	+1.0	-0.12	0.17	0.08	2/248
Input image	+0.84	+6.2	+0.60	0.36	0.47	244/248
Weights, 400 steps	+0.74	+5.7	+0.50	0.33	0.31	400, +0.00
Bottleneck, 400 steps	+0.74	+6.2	+0.50	0.35	0.40	389, -0.00
Input image, 400 steps	+0.86	+6.3	+0.62	0.37	0.49	339, -0.00
T1, 4× *(contrast shift)*, frozen 26.40 dB, range oracle +1.05						
Weights	+2.26	+15.4	+1.21	0.54	0.46	–
Bottleneck, native	+1.00	+4.9	-0.05	0.32	0.39	0/120
Bottleneck, 10^2	+0.56	+3.3	-0.49	0.25	0.23	0/120
Bottleneck, 5^2	+0.35	+2.0	-0.70	0.20	0.19	0/120
Input image	+1.31	+1.9	+0.26	0.38	0.70	1/120
Weights, 400 steps	+3.03	+21.0	+2.04	0.61	0.60	400, +0.00
Bottleneck, 400 steps	+2.00	+12.0	+1.02	0.50	0.68	400, +0.00
Input image, 400 steps	+1.36	+2.4	+0.37	0.42	0.79	142, -0.07
T2, 4× *(in distribution)*, frozen 35.81 dB, range oracle +0.67						
Weights	+0.54	+1.4	-0.13	0.27	0.49	–
Bottleneck, native	+0.75	+4.5	+0.08	0.29	0.28	205/248
Bottleneck, 10^2	+0.35	+2.2	-0.32	0.22	0.14	77/248
Bottleneck, 5^2	+0.22	+1.4	-0.45	0.18	0.11	43/248
Input image	+1.34	+7.3	+0.67	0.43	0.65	242/248
Weights, 400 steps	-0.09	-5.9	-0.81	0.31	0.78	55, -0.78
Bottleneck, 400 steps	+0.25	-3.1	-0.47	0.30	0.68	131, -0.75
Input image, 400 steps	+1.39	+7.4	+0.66	0.45	0.70	141, -0.08
T2, 4×*, k-space noise*, frozen 35.68 dB, range oracle +0.70						
Weights	+0.53	+1.4	-0.17	0.27	0.48	–
Bottleneck, native	+0.72	+4.4	+0.02	0.29	0.28	202/248
Bottleneck, 10^2	+0.34	+2.2	-0.36	0.22	0.14	74/248
Bottleneck, 5^2	+0.21	+1.3	-0.49	0.17	0.11	42/248
Input image	+1.15	+6.0	+0.45	0.42	0.66	236/248
Weights, 400 steps	-0.57	-11.7	-1.33	0.26	0.83	42, -1.22
Bottleneck, 400 steps	+0.21	-3.2	-0.55	0.29	0.68	118, -0.74
Input image, 400 steps	+0.54	-1.1	-0.22	0.40	0.82	38, -0.65
T2, 4×*, inverted contrast (P7a)*, frozen 23.54 dB, range oracle +0.94						
Weights	+1.92	+14.6	+0.97	0.52	0.46	–
Bottleneck, native	+1.01	+7.7	+0.07	0.39	0.30	0/24
Input image	+1.19	+2.8	+0.25	0.39	0.69	0/24
Weights, 400 steps	+3.34	+25.0	+2.39	0.67	0.60	364, -0.02
Bottleneck, 400 steps	+1.15	+5.4	+0.21	0.41	0.63	236, -0.27
Input image, 400 steps	-0.54	-20.6	-1.49	0.30	1.03	25, -1.80
T2, 4×*, random line mask (P7b)*, frozen 31.26 dB, range oracle +0.75						
Weights	+1.02	+5.3	+0.27	0.34	0.45	–
Bottleneck, native	+0.95	+6.7	+0.20	0.37	0.29	12/24
Input image	+3.48	+23.5	+2.73	0.65	0.71	24/24
Weights, 400 steps	+1.69	+10.6	+0.94	0.47	0.62	359, -0.03
Bottleneck, 400 steps	+1.99	+12.9	+1.24	0.50	0.62	352, -0.03
Input image, 400 steps	+3.92	+26.8	+3.17	0.68	0.74	398, -0.00

Table 32: **The second operator.** Cartesian MRI through the U-Net, held-out slices; forty steps at each variant's global rate, then 400. Null drop: reduction of null-space error; vs. oracle: margin over the range oracle; Wins: slices where the variant's PSNR exceeds the weight-space variant's at the same budget.

and 0.37 and null-space error reductions that agree to within one percentage point. This reproduces the convergence result of Appendix T on an operator that shares nothing with the projector. In distribution, the weight-space variant is the one that cannot be run indefinitely. After its peak at step 55 its null-space error grows, and it ends 0.78 dB below its peak. The bottleneck peaks at step 131, and the input image is still within 0.08 dB of its peak at 400 steps. Narrowing the bottleneck reduces the gain monotonically on every condition, so the width selector of Appendix U has nothing to select here; the native width is the best width for that variable. The error identity holds to within 8e-8 on every cell. Read as the criterion of Section 5.4, it predicts the sign of the cell-mean PSNR difference on 11 of 12 pairs of variants. One of these pairs, the weight-space variant against the bottleneck at 8×, is decided by a mean ρ difference below 0.001 and a PSNR difference of 0.06 dB. Slice by slice, the criterion agrees with the PSNR ordering on 2303 of 2592 pairs; the remaining pairs are decided by the range-space component, which ρ does not measure.

Input image, contrast shift and noise. The input image, the widest of the variables, recovers the largest fraction of null-space error: 6.2% against 2.7% for the bottleneck at 8×, and 7.3% against 4.5% at 4×. At forty steps it outperforms the weight-space variant on 244 of 248 held-out slices at 8× and on 242 of 248 at 4×. The input image has the encoder between it and the output, which the bottleneck lacks: the whole U-Net lies between the input and the image, whereas only the decoding half lies between the bottleneck and the image. Since the prior resides in the frozen layers between the variable and the image (Section 4), it is the depth of the frozen network between the variable and the output that orders the variants, and not the width of the variable. A second reversal relative to the cone-beam results occurs under the contrast shift. There the weight-space variant is better at every step budget, with +2.26 dB against +1.00 for the bottleneck at forty steps and +3.03 against +2.00 at 400 steps. The variants do not converge to the same alignment; their values of $\cos\theta$ remain at 0.61, 0.50 and 0.42. A frozen decoder trained on T2 maps to intensities that the T1 anatomy does not have, and only the weights can change that mapping. The activation variables that pass through the decoder inherit its alignment, and under the contrast shift that alignment is wrong. The third condition is the noise. On the projector the weight-space variant did not overfit within 400 steps, but here it overfits first: it peaks at step 42 and loses 1.22 dB by step 400. The input image peaks at step 38 and the bottleneck last, at step 118. The step budget of forty steps is therefore the stopping rule for every variable on this operator, and the bottleneck is the variable that depends on it least.

Further conditions and the full test set. The contrast reversal above was observed on a single condition, so we ran two further conditions with their predictions fixed in advance. Two variants are said to converge to the same alignment (labelled meet) when their gap in $\cos\theta$ at 400 steps is within 0.05 (Table 32, last two blocks). Inverting the foreground intensities of the T2 slices changes nothing in the anatomy or the operator and only changes what the decoder must map; this reproduces the T1 result. The weight-space variant converges to an alignment 0.26 above that of the bottleneck and reaches +3.34 dB against +1.15. A random line mask at 4× is a sampling pattern the model never saw, while the mapping learned by the decoder remains valid. It gives a gap of 0.03 between the weight-space variant and the bottleneck, so these two variants converge to the same alignment. Under that mask the input image is far better than both, at +3.92 dB after 400 steps with 26.8% of the null-space error removed, because the variable that carries the aliasing pattern is the one that can undo a change in it. On the full test set, the forty-step cell means of the four main conditions are within 0.09 dB of the twenty-four-slice means. The only ordering that differs is that of the bottleneck against the weight-space variant at 8×: the two are tied on the subset, whereas on the full set the weight-space variant is better, at +0.39 against +0.33 dB.

Alignment on the U-Net. A measurement-only loss moves the null-space component through the frozen network, and the alignment $\cos\theta$ between the null-space update and the null-space error is determined by the frozen layers between the variable and the output. On the U-Net this ordering is directly visible. When the mapping learned by the decoder remains valid, the input, which has the whole network between it and the output, reaches a higher alignment than the bottleneck, which has only the decoding half: 0.45 against 0.30 at 4× and 0.68 against 0.50 under the random mask. Under the contrast shift the order reverses, 0.42 against 0.50, because the frozen network between the input and the output is precisely the mapping that is wrong; the weights, which alone can change it, reach a higher alignment than both. Table 25 reports the input, which is the variable that the label-free selection procedure of Appendix S picks on this backbone. For the bottleneck the random-mask

gap is -0.03 and the inverted gap is +0.26, which are the convergence to the same alignment (meet) and the higher weight-space alignment (above) predicted in advance. For the input the contrast rows remain above, while the random-mask and 4× rows fall below, with gaps of -0.21 and -0.14, because the weights converge to an alignment lower than that of a variable with the whole network between it and the output. Which activation variable is the most effective depends on the backbone. On DeepSparse it is the fused volume, with the input projections slightly below it and better than the weight-space variant (Appendix G); on the U-Net it is the input image. Where the weight-space variant converges above, its alignment is already higher at forty steps, 0.54 against 0.32 on T1 and 0.52 against 0.39 under inversion, so the crossover precedes the step budget rather than following it. On every MRI condition, the sign of the criterion at forty steps equals its sign at 400 steps for the weight-space variant against either activation variable.

W A THIRD OPERATOR: SUPER-RESOLUTION OF NATURAL IMAGES

The first two operators are medical, and their reconstructors were trained either by us or by a group we cite. This section takes a released reconstructor from a field with no medical connection, single-image super-resolution, and repeats the measurements on natural photographs. The operator is 4× downsampling, $A = SK$, where K is a blur and S is the subsampling. The operator is separable, so its adjoint and its range projector are exact and available in closed form without a conjugate gradient solve. The null space accounts for fifteen sixteenths of the image. The frozen prior is the released EDSR baseline (Lim et al., 2017; Siow, 2021), a feed-forward network with 1.5M weights trained on DIV2K at 4× with the antialiased bicubic kernel. All predictions were written down before any run.

Scope. At its own kernel, the released network is already consistent with its measurements. On BSD100, Urban100, DIV2K, the IXI slices and inverted BSD100, the range oracle is +0.01, +0.01, +0.03, +0.03 and +0.01 dB above the frozen output. A data-consistency loss therefore has no residual to act on, whatever the content, and no variant moves (Table 34, last block). This is the same situation as for the variational network in Appendix V, reached here without any data-consistency layer, by a feed-forward network trained on the exact operator it is tested with. The study is therefore run where the frozen output is not consistent: under a Gaussian blur of two pixels applied before the subsampling, which is the kernel-mismatch case of blind super-resolution. This condition leaves +0.48 dB to the oracle, and both the loss and the network's input use the true kernel. The second preliminary result concerns the choice of variable. The feature read by the upsampler has only two convolutions and two pixel shuffles between it and the output, and that tangent kernel maps the range space to the range space. At every learning rate this feature therefore gains in the range-space component and leaves the null-space error unchanged (-0.1% at forty steps). The feature after the head, which has the sixteen residual blocks between it and the output, is the variable that moves the null-space component here. It is the feature variant that the rest of this section reports, while Table 25 reports the input, which is the variable the selection procedure picks and which moves the null-space component further still. The prior resides in the frozen layers between the variable and the image, and the late feature has almost none.

The Gaussian-kernel conditions of Tables 33 and 34 are averages over 92 images, or 32 for the IXI slices.

Results. Table 33 reports the grid. Under the kernel shift, every variable with the network body between it and the output lowers the null-space error and exceeds the frozen model's range oracle on every content, natural or medical. The early feature exceeds the oracle by +0.48, +0.97 and +0.80 dB on BSD100, Urban100 and the IXI slices. The LR input exceeds it by +0.64, +1.31 and +1.25, and the weight-space variant by +0.09, +0.36 and +0.32. The LR input, which is sixteen times smaller than the image and has the whole network between it and the output, is the most effective variable at forty steps on every condition, as it was on the U-Net. Narrowing the early feature to 16^2 or 8^2 removes the entire gain. All three variants plateau by the eightieth step, so the 400-step rows represent convergence, and there the data add the third case of Table 25. The weight-space variant does not converge to the same alignment as the activation feature; it converges below it, at 0.15 against 0.30 on BSD100, 0.22 against 0.39 on Urban100 and 0.32 against 0.48 on the IXI slices. A network trained to be consistent with its operator restores consistency, once the operator changes,

Variable	Gain (dB)	Null drop (%)	vs. oracle (dB)	$\cos\theta$	r	Wins / peak, loss
BSD100, Gaussian kernel (in-distribution content), frozen 24.26 dB, range oracle +0.69, $n = 92$						
Weights	+0.78	+1.6	+0.09	0.17	0.25	–
Late feature	+0.67	-0.1	-0.02	0.01	0.04	34/92
Early feature	+1.17	+5.4	+0.48	0.30	0.37	89/92
Early feature, 16^2	+0.01	-2.0	-0.68	0.07	0.25	0/92
Early feature, 8^2	-0.09	-3.0	-0.79	0.06	0.26	0/92
LR input	+1.33	+7.1	+0.64	0.34	0.34	92/92
Weights, 400 steps	+0.79	+1.1	+0.09	0.15	0.24	114, -0.05
Early feature, 400 steps	+1.19	+5.4	+0.49	0.30	0.37	71, -0.01
LR input, 400 steps	+1.38	+7.4	+0.69	0.34	0.33	226, -0.02
Urban100, Gaussian kernel (content shift), frozen 21.65 dB, range oracle +0.78, $n = 92$						
Weights	+1.14	+4.2	+0.36	0.26	0.32	–
Late feature	+0.75	+0.0	-0.02	0.02	0.05	26/92
Early feature	+1.75	+9.7	+0.97	0.39	0.51	77/92
Early feature, 16^2	-0.12	-4.5	-0.90	0.09	0.35	2/92
Early feature, 8^2	-0.32	-7.0	-1.10	0.06	0.38	1/92
LR input	+2.09	+13.2	+1.31	0.44	0.45	91/92
Weights, 400 steps	+1.10	+3.4	+0.32	0.22	0.29	75, -0.26
Early feature, 400 steps	+1.78	+9.9	+1.01	0.39	0.51	72, -0.03
LR input, 400 steps	+2.19	+13.9	+1.41	0.46	0.44	183, -0.03
IXI slices, Gaussian kernel (far content shift), frozen 24.52 dB, range oracle +1.46, $n = 32$						
Weights	+1.78	+4.8	+0.32	0.31	0.34	–
Late feature	+1.40	+0.1	-0.06	0.05	0.07	7/32
Early feature	+2.26	+9.4	+0.80	0.48	0.71	27/32
Early feature, 16^2	-0.57	-15.3	-2.03	-0.01	0.55	0/32
Early feature, 8^2	-0.55	-13.6	-2.01	-0.04	0.50	0/32
LR input	+2.71	+14.2	+1.25	0.52	0.59	32/32
Weights, 400 steps	+1.90	+5.4	+0.44	0.32	0.31	153, -0.03
Early feature, 400 steps	+2.31	+9.7	+0.85	0.48	0.71	90, -0.03
LR input, 400 steps	+2.73	+14.1	+1.27	0.51	0.55	157, -0.11
BSD100, Gaussian kernel, inverted, frozen 24.26 dB, range oracle +0.69, $n = 92$						
Weights	+0.80	+1.8	+0.11	0.18	0.25	–
Late feature	+0.67	-0.0	-0.02	0.01	0.04	41/92
Early feature	+1.13	+4.9	+0.44	0.29	0.37	86/92
Early feature, 16^2	+0.03	-1.7	-0.66	0.08	0.24	0/92
Early feature, 8^2	-0.05	-2.2	-0.74	0.06	0.25	0/92
LR input	+1.28	+6.5	+0.59	0.32	0.33	92/92
Weights, 400 steps	+0.79	+1.2	+0.10	0.15	0.24	116, -0.05
Early feature, 400 steps	+1.14	+5.0	+0.45	0.29	0.37	67, -0.01
LR input, 400 steps	+1.31	+6.7	+0.62	0.32	0.33	216, -0.01
BSD100, Gaussian kernel, LR noise at 5%, frozen 22.97 dB, range oracle +1.67, $n = 92$						
Weights	-2.41	-26.3	-4.08	-0.02	0.72	–
Late feature	-1.53	-0.8	-3.20	-0.01	0.10	92/92
Early feature	-1.68	-5.4	-3.35	0.01	0.33	92/92
Early feature, 16^2	-0.79	-14.5	-2.46	0.01	0.54	92/92
Early feature, 8^2	-0.92	-17.0	-2.59	0.01	0.60	90/92
LR input	-1.71	-5.4	-3.38	0.01	0.32	92/92
Weights, 400 steps	-2.14	-18.1	-3.81	-0.01	0.58	14, -1.32
Early feature, 400 steps	-1.71	-5.4	-3.38	0.01	0.33	27, -0.13
LR input, 400 steps	-1.70	-5.3	-3.37	0.01	0.32	66, -0.03

Table 33: **The third operator.** 4× super-resolution through the released EDSR, held-out images; forty steps at each variant's global rate, then 400. Columns as in Table 32. Late feature: read by the upsampler; early feature: after the head.

Variable	Gain (dB)	Null drop (%)	vs. oracle (dB)	$\cos\theta$	r	Wins / peak, loss
BSD100, Gaussian kernel, LR noise at 1% *(follow-up)*, frozen 24.22 dB, range oracle +0.73, $n = 92$						
Weights	+0.45	-1.0	-0.29	0.12	0.31	–
Early feature	+0.73	+3.1	-0.00	0.24	0.34	86/92
LR input	+0.85	+4.4	+0.11	0.26	0.29	92/92
Weights, 400 steps	+0.40	-1.1	-0.34	0.10	0.28	62, -0.20
Early feature, 400 steps	+0.74	+3.1	+0.00	0.24	0.34	55, -0.01
LR input, 400 steps	+0.86	+4.5	+0.13	0.26	0.29	258, -0.01
BSD100, bicubic kernel (no residual), frozen 25.64 dB, range oracle +0.01, $n = 92$						
Weights	-0.15	-1.9	-0.16	0.01	0.19	–
Late feature	+0.01	-0.0	-0.00	-0.00	0.01	90/92
Early feature	-0.09	-1.2	-0.10	0.02	0.17	48/92
Early feature, 16^2	-0.06	-0.8	-0.07	0.02	0.14	88/92
Early feature, 8^2	-0.06	-0.8	-0.07	0.03	0.15	85/92
LR input	+0.02	+0.0	+0.00	0.03	0.06	88/92

Table 34: **The third operator, continued.** Follow-up condition with one percent noise, and the no-residual control at the released network's own kernel; protocol and columns as in Table 33.

through the weight change that fits the range-space component, and that change barely enters the null space. The activation variables have no such option, because everything between them and the output is frozen, and it is the frozen body that carries the residual into the null space. Inversion is not a shift of the mapping for this network: its frozen output reaches 24.26 dB on the inverted images against 24.26 on the upright ones. The prediction that the weight-space variant would be better under inversion therefore had no basis, and the variants behave as they do on the upright images. Noise at five percent of the LR rms degrades every variant from the tenth step onwards, the weight-space variant most (-2.14 dB at 400 steps) and the activation variants least (-1.71 and -1.70). With fifteen sixteenths of the image in the null space, a plain data-consistency loss writes the noise into the reconstruction immediately, and a step budget cannot stop at a peak that is never reached. At one percent noise, a follow-up condition (Table 34), the weight-space variant reaches +0.40 dB at 400 steps against +0.74 for the early feature and +0.86 for the input, with peaks at steps 62, 55 and 258. At the bicubic kernel, the no-residual control, the feature variants do not move, and the weight-space variant loses -0.15 dB by moving the null-space component without any residual in the measurements to justify it.

Per image. A forty-step adaptation of the LR input takes 0.21 s per image at the median over the three Gaussian-kernel conditions, and 213 of 216 of the held-out images finish in under one second. The remaining images are the first image of each run, which incurs the CUDA warm-up of about 9 s; no later image exceeds 0.48 s. The weight-space variant takes 0.48 s at the median. Counted per image, the LR input exceeds the frozen model's range oracle on 212 of 216 of the held-out images and the weight-space variant on 143 of 216. On BSD100 the counts are 92 of 92 against 54 of 92, on Urban100 88 of 92 against 64 of 92, and on the IXI slices 32 of 32 against 25 of 32. The margin of the LR input over the weight-space variant on the same images is +0.55, +0.95 and +0.93 dB.

Criterion. Read as the criterion of Section 5.4, the error identity predicts the sign of the cell-mean PSNR difference on 12 of 12 pairs of variants over the four Gaussian-kernel conditions, and image by image it agrees with the PSNR ordering on 921 of 924 pairs. The identity itself holds to within 2e-13 on every cell.

Alignment across backbones. The alignment is determined by the frozen layers between the variable and the image, and on this operator the depth of those frozen layers decides whether the variable moves the null-space component at all. The weight-space variant is not tied to the activation alignment in either direction. On the projector under an anatomy shift it converges to the same alignment; under a contrast shift on MRI it converges to a higher alignment; and on a network consistent with its own operator, or when compared with an activation variable that has the whole network between it and the output, it converges to a lower alignment. Table 25 contains all three cases.

X A FOURTH OPERATOR: INPAINTING, WHERE NO VARIABLE HAS A NON-ZERO ALIGNMENT

The first three operators have null spaces defined in the frequency domain: the high frequencies under the downsampler, the unsampled lines of k-space under the mask, the frequencies annihilated by a blur, and the unsampled directions under the projector. A pixel mask instead places the null space in a spatial region of the image, the hole, which is disjoint from every measured pixel. The operator is $A = M$, its adjoint is M, and its range projector is M itself, which is exact and diagonal. The reconstructor is a U-Net with 3.7M weights that we trained on DIV2K crops with one to three rectangular holes covering between a fifth and a third of the image; its input is the masked image together with the mask, and its output is the full image. Predictions were written down before training finished.

Two networks. When trained with an ℓ_1 loss on the whole image, the U-Net copies the observed pixels through its skip connections regardless of the input it is shown. The range oracle is +0.06, +0.10, +0.01 and +0.07 dB above the frozen output on BSD100, Urban100, the IXI slices and inverted BSD100. It is +0.28 above under brush strokes the network never saw, and +0.07 above under strokes covering 44% of the image. No condition leaves a residual of the size the study requires, so on this network the study is run only under the mask shifts of Table 36. This is the fourth reconstructor in this paper that is consistent with its operator by training, after the variational network, EDSR and SwinIR. A second network with the same architecture, data and masks is trained with the loss restricted to the hole, so that it has never been trained on the observed region. It leaves +5.56 to +9.88 dB to the oracle, and the study is run on it.

Results. Table 35 reports the grid. Every variant fits the observed pixels, and the weight-space variant gains +2.62 dB on BSD100 by doing so. Every variant also moves the null-space component, the early feature by more than the magnitude of the error itself, and every one of them moves it in the wrong direction. The null-space error grows by -7.0% under the weight-space variant, by -74.4% under the early feature and by -1.7% under the input, with alignments $\cos\theta$ of 0.04, 0.06 and -0.03. The bottleneck, with only the decoder between it and the output, does not move at all. This is not a matter of learning rate, because at the smallest rate in the sweep the alignment is already zero. At 400 steps only the relative length r has changed: the weight-space variant ends -4.64 dB below the frozen output with -243.7% more null-space error. The full-image network under the same variants and rates, which serves as the no-residual control, moves by -0.05, -0.10 and +0.04 dB. On this operator the measured residual is located on pixels disjoint from the hole, and the decoder's tangent kernel carries it into the hole in directions unrelated to the ground truth. The error identity describes this situation correctly. With $\cos\theta$ equal to zero, its minimum is at $r = 0$, so the correct step budget is zero steps. What the label-free selection procedure observes is an r far above the observed range of $\cos\theta$, which is the condition under which the check triggers for a variable that has overshot the minimum of the identity; here that check describes the entire outcome.

Operator or residual. Two explanations are consistent with the grid above, and they are different claims. Under the first, the zero alignment comes from the geometry of the operator: the measured residual is located on pixels disjoint from the hole, and the decoder carries it across without any alignment to the error. Under the second, the zero alignment comes from the origin of the residual. The hole-only network was never trained on the observed region, where its output reaches 16.8 dB per pixel against 40.7 dB for the full-image network. The residual it leaves is therefore unrelated to its error in the hole; across images the two correlate at 0.19 to 0.36, against 0.56 to 0.81 on the super-resolution conditions, where the mechanism works. Under the second explanation, a residual left by a consistent network under a genuine shift could still produce a non-zero alignment. A further test, with its decision rule fixed in advance, separated the two explanations. The full-image network, which copies the observed pixels under every mask of its training family, was shown masks it had never seen, chosen so that every hole pixel lies within a few pixels of observed ones: random dropout of 30% of the pixels, dense strokes three to six pixels wide on BSD100, Urban100 and the IXI slices, and the brush strokes of the survey above on Urban100. These residuals originate from prior mismatch, and they are large enough for the study, with the range oracle +1.40, +0.73, +0.66, +0.37 and +0.34 dB above the frozen output.

Variable	Gain (dB)	Null drop (%)	vs. oracle (dB)	$\cos\theta$	r	Wins / peak, loss
BSD100, rectangles (the training family), hole 25%, frozen 17.00 dB, range oracle +5.56, $n = 92$						
Weights	+2.62	-7.0	-2.94	0.04	0.40	–
Bottleneck	-0.00	-0.3	-5.57	-0.02	0.06	0/92
Early feature	+0.58	-74.4	-4.99	0.06	1.44	12/92
Early feature, 64^2	-1.96	-139.4	-7.53	0.07	2.22	0/92
Early feature, 32^2	-2.70	-162.2	-8.27	0.08	2.48	0/92
Masked input	+0.44	-1.7	-5.12	-0.03	0.16	0/92
Weights, 400 steps	-4.64	-243.7	-10.21	0.10	3.37	51, -7.44
Early feature, 400 steps	+0.09	-97.8	-5.47	0.07	1.74	52, -1.68
Masked input, 400 steps	-1.92	-136.0	-7.49	0.04	2.15	91, -2.66
Urban100, rectangles (content shift), hole 25%, frozen 15.75 dB, range oracle +5.16, $n = 92$						
Weights	+1.98	-7.8	-3.17	0.04	0.41	–
Bottleneck	-0.00	-0.2	-5.16	-0.01	0.05	0/92
Early feature	+0.09	-73.1	-5.07	0.07	1.41	19/92
Early feature, 64^2	-1.96	-125.6	-7.12	0.08	2.06	0/92
Early feature, 32^2	-2.62	-144.0	-7.78	0.09	2.28	0/92
Masked input	+0.36	-1.6	-4.79	-0.02	0.14	1/92
Weights, 400 steps	-3.23	-174.5	-8.39	0.08	2.60	46, -5.36
Early feature, 400 steps	-0.45	-103.9	-5.61	0.08	1.80	77, -1.81
Masked input, 400 steps	-1.77	-119.1	-6.93	0.07	1.97	100, -2.47
IXI slices, rectangles (far content shift), hole 26%, frozen 18.56 dB, range oracle +1.90, $n = 32$						
Weights	+1.00	+0.1	-0.90	0.12	0.17	–
Bottleneck	-0.01	-0.2	-1.91	-0.07	0.03	0/32
Early feature	+0.93	-7.9	-0.97	0.13	0.55	11/32
Early feature, 64^2	-1.06	-37.0	-2.96	0.20	1.15	2/32
Early feature, 32^2	-2.78	-68.9	-4.68	0.22	1.59	0/32
Masked input	+0.37	+0.1	-1.53	0.06	0.09	1/32
Weights, 400 steps	-4.31	-109.1	-6.21	0.18	2.00	45, -5.36
Early feature, 400 steps	+1.03	-10.4	-0.87	0.18	0.67	126, -0.30
Masked input, 400 steps	-0.56	-28.6	-2.46	0.01	0.82	102, -1.21
BSD100, brush strokes (mask-type shift), hole 14%, frozen 17.64 dB, range oracle +9.88, $n = 92$						
Weights	+3.98	-10.3	-5.90	-0.01	0.43	–
Bottleneck	+0.01	-0.1	-9.87	-0.00	0.03	0/92
Early feature	+4.09	-54.8	-5.78	0.04	1.18	63/92
Early feature, 64^2	+2.48	-93.9	-7.40	0.03	1.66	10/92
Early feature, 32^2	+2.52	-84.5	-7.35	0.03	1.55	7/92
Masked input	+0.63	-0.9	-9.25	0.09	0.24	0/92
Weights, 400 steps	-1.12	-247.0	-11.00	-0.03	3.28	66, -5.45
Early feature, 400 steps	+4.81	-75.0	-5.07	0.08	1.49	242, -0.23
Masked input, 400 steps	+1.88	-105.8	-8.00	0.05	1.82	290, -0.37
BSD100, rectangles, inverted, hole 25%, frozen 16.08 dB, range oracle +6.33, $n = 92$						
Weights	+2.95	-9.3	-3.38	0.04	0.45	–
Bottleneck	+0.00	-0.4	-6.33	-0.03	0.06	0/92
Early feature	-0.95	-128.3	-7.28	0.11	2.14	3/92
Early feature, 64^2	-2.89	-191.9	-9.22	0.11	2.83	0/92
Early feature, 32^2	-3.52	-215.0	-9.85	0.11	3.08	0/92
Masked input	+0.49	-0.9	-5.84	0.04	0.17	0/92
Weights, 400 steps	-2.54	-185.2	-8.87	0.03	2.68	50, -5.63
Early feature, 400 steps	-1.88	-170.6	-8.21	0.12	2.61	16, -2.94
Masked input, 400 steps	-3.42	-212.5	-9.75	0.12	3.06	72, -4.18

Table 35: **The fourth operator.** Inpainting through the hole-only U-Net, held-out images; forty steps at each variant's global rate, then 400. Columns as in Table 32. Early feature: after the first block; bottleneck: the deepest activation.

Variable	Gain (dB)	Null drop (%)	vs. oracle (dB)	$\cos\theta$	r
BSD100, random dropout, hole 30%, frozen 18.42 dB, range oracle +1.49, $n = 92$					
Weights	+0.80	-6.0	-0.69	+0.00	0.34
Early feature	-0.01	-18.3	-1.50	-0.54	0.29
Masked input	+0.12	-3.2	-1.37	-0.46	0.07
Weights, 400 steps	-1.19	-37.5	-2.67	-0.37	0.62
Early feature, 400 steps	+0.51	-12.0	-0.98	-0.24	0.32
Masked input, 400 steps	+0.01	-17.3	-1.48	-0.39	0.34
BSD100, thin strokes, hole 24%, frozen 24.36 dB, range oracle +0.87, $n = 92$					
Weights	-2.19	-40.0	-3.06	-0.09	0.86
Early feature	+0.09	-9.6	-0.78	+0.03	0.46
Masked input	+0.17	-1.3	-0.70	+0.02	0.18
Weights, 400 steps	-3.07	-58.6	-3.94	-0.00	1.23
Early feature, 400 steps	+0.31	-7.0	-0.55	+0.08	0.46
Masked input, 400 steps	-1.44	-31.5	-2.31	+0.02	0.86
Urban100, thin strokes, hole 24%, frozen 22.57 dB, range oracle +0.80, $n = 92$					
Weights	-2.28	-41.2	-3.07	-0.11	0.86
Early feature	+0.04	-8.8	-0.76	+0.00	0.41
Masked input	+0.12	-1.2	-0.68	-0.01	0.15
Weights, 400 steps	-3.28	-61.6	-4.08	-0.05	1.21
Early feature, 400 steps	+0.24	-7.0	-0.56	+0.06	0.43
Masked input, 400 steps	-1.17	-25.5	-1.96	+0.00	0.74
IXI slices, thin strokes, hole 24%, frozen 25.60 dB, range oracle +0.40, $n = 32$					
Weights	-1.03	-15.5	-1.43	-0.01	0.57
Early feature	+0.58	+2.2	+0.18	+0.23	0.32
Masked input	+0.32	+2.1	-0.08	+0.22	0.15
Weights, 400 steps	-1.65	-26.4	-2.05	+0.12	0.90
Early feature, 400 steps	+0.66	+2.9	+0.26	+0.25	0.31
Masked input, 400 steps	+0.01	-4.4	-0.39	+0.21	0.58
Urban100, brush strokes, hole 14%, frozen 24.34 dB, range oracle +0.49, $n = 92$					
Weights	-2.67	-43.9	-3.16	-0.10	0.90
Early feature	-0.70	-14.7	-1.20	-0.05	0.47
Masked input	-0.03	-1.9	-0.53	-0.03	0.15
Weights, 400 steps	-6.44	-129.6	-6.93	-0.07	1.98
Early feature, 400 steps	-0.47	-12.4	-0.96	-0.02	0.46
Masked input, 400 steps	-2.69	-47.3	-3.19	-0.06	0.98

Table 36: **The full-image U-Net under mask shifts.** Trained on rectangles, tested on mask types it never saw, held-out images; forty steps at each variant's global rate (pooled mean over the two rate-sweep conditions), then 400. Columns as in Table 35.

Outcome. Table 36 reports the grid. On the photographs, under dropout, thin strokes and brush strokes, the alignment stays at zero for every variable: it lies between -0.54 and +0.03 over the 12 pairs of variant and condition at forty steps, and the null-space error grows in every one of them. Across the rate sweeps the alignment stays at or below zero at every rate, between -0.41 and +0.02 for the weight-space variant, between -0.29 and -0.23 for the early feature, and between -0.24 and -0.13 for the input. On the IXI slices under the same strokes, the trigger of the decision rule fixed in advance does occur. The early feature reaches an alignment of +0.23 with the null-space error falling by 2.2%, and it ends +0.18 dB above the range oracle; the input reaches +0.22. The zero alignment holds on the other four conditions. The decision rule had assigned a trigger to the residual-origin explanation, but the photographs contradict that explanation, because the same network, the same masks and the same origin of residual produce no alignment there. A shell probe then tested the graded form of the geometric explanation, namely that the alignment decreases with the distance from the nearest observed pixel. On IXI the alignment is flat across shells one to four pixels deep, at +0.13, +0.25, +0.24 and +0.26 for the early feature. On BSD100 and Urban100 it is zero in every shell, from -0.14 to +0.01 and from -0.11 to -0.03. What separates the conditions is the image content. The decoder carries a residual into the null space along the directions about which the residual is informative. Where the range-space and null-space components overlap in the image domain, a local residual is informative about the local null-space component, and a kernel shift alone produces the alignments of Table 25. Where the two components are disjoint, a residual created by the mask itself, on content the prior has learned, is uninformative about the hole. Only content far

from the prior yields an alignment, and that alignment is small, about 0.23, and the same at every distance from the measurements.

Alignment on a pixel mask. The alignment is a joint property of the decoder, the operator and the residual, and not of any single variable. On the three operators whose range-space and null-space components overlap in the image domain, the frozen prior converts a measured residual into an update with a non-zero alignment to the null-space error. On the pixel mask, where the two components are disjoint, it converts a residual created by the mask itself into an alignment of zero, and a residual from a far content shift into an alignment of about 0.23. Across the 10 mask conditions of the two networks, the two activation variables agree on this alignment to within 0.10. The error identity describes every cell correctly, because with $\cos\theta$ equal to zero its minimum is at $r = 0$, and the label-free check triggers on every variable. The three conditions of Section 4 yield the null-space coupling, and the operator determines how much of it there is.

Y A FIFTH OPERATOR: DEBLURRING, WHERE WEIGHT-SPACE ADAPTATION IS BETTER AND THE CRITERION PREDICTS IT

The fifth operator is a blur: A is a circular convolution with a box kernel of side L. When L divides the grid size, the transfer function has exact zeros, so the null space is exact and its projector is the indicator of those frequencies. This projector is diagonal in the Fourier domain, just as the other four projectors are exact in their own bases. A horizontal motion blur with L=16 annihilates 5.9% of the frequencies, and a box blur of the same side annihilates 11.4%. The reconstructor is the released NAFNet trained on GoPro motion deblurring (Chen et al., 2022), with 17.1M weights, all of which are frozen. Its training operator is real camera motion, so a box kernel is an operator shift of the same kind as the Gaussian kernel of the third operator. The variables are the weights, the middle-block output (131k entries, with the decoder between it and the output), the feature after the first convolution (2.1M entries, with the whole encoder and decoder between it and the output) and the blurred input image (197k entries, with the whole network between it and the output). Predictions were written down before any adaptation run, including the blind choice of the selection procedure, which is the input.

Alignment. A blur defines its null-space component in the frequency domain, so the range-space and null-space parts of an image overlap everywhere in space, and a measured residual can be carried into the null space. The alignment is accordingly large for every variable that moves. At forty steps the input reaches a $\cos\theta$ between 0.37 and 0.81 over the 5 conditions and removes 8.9 to 41.1% of the null-space error, and the weight-space variant reaches 0.39 to 0.85. No other variable in this paper reaches 0.8. The bottleneck is the exception. It moves the image by $r = 0.01$ at forty steps, for a gain of +0.03 dB; it is the second deepest activation in the paper, after the bottleneck of the inpainting U-Net, that the skip connections leave with nothing to contribute.

Criterion. This is the first operator in the paper on which the weights are the better variable at the forty-step budget. The weight-space variant outperforms the variable chosen by the selection procedure on all 5 conditions, by +2.13 against +2.01 dB on the sweep condition and by +6.97 against +5.47 on the IXI slices. Per image, the input is better than the weight-space variant on 24 of 52 images on the sweep condition and on 0 of 32 on the IXI slices. The selection procedure orders the activation variables correctly here, since its choice outperforms the other two on four of the five conditions, and the error identity settles the choice between the activation variables and the weights (Appendix S). Read as the criterion of Corollary 3 on the held-out cells, which involves no PSNR, ρ ranks the weight-space variant above the input on 4 of 5 conditions. The exception is BSD100 under the box kernel, where the two differ by +1.92 against +1.88 dB. Over every pair of variants on every condition, the sign of the cell-mean ρ difference matches the sign of the PSNR difference on 71 of 75 pairs, and the pairs it misses are at most 0.19 dB apart. The identity holds to within 7e-13.

Convergence. When run to 400 steps, the weight-space variant converges to the same alignment as the activation variable on three conditions and to a higher alignment on two, by +0.08 on the rate-sweep condition and by +0.13 on Urban100 (Table 25). The case of a released restorer under a kernel shift, which on EDSR and SwinIR placed the weight-space variant below, does not carry

Variable	Gain (dB)	Null drop (%)	vs. oracle (dB)	$\cos\theta$	r	Wins / peak, loss
BSD100, motion blur (the sweep condition), null space 5.9%, frozen 25.84 dB, range oracle +10.99, $n = 52$						
Weights	+2.13	+12.3	-8.86	0.46	0.42	–
Bottleneck	+0.03	+0.1	-10.96	0.13	0.01	0/52
Early feature	+1.78	+9.1	-9.21	0.41	0.49	7/52
Early feature, 64^2	+1.14	+4.5	-9.84	0.29	0.39	0/52
Early feature, 32^2	+0.93	+4.2	-10.06	0.27	0.33	0/52
Blurred input	+2.01	+10.0	-8.98	0.42	0.49	24/52
Weights, 400 steps	+6.61	+25.6	-4.37	0.65	0.78	400, +0.00
Bottleneck, 400 steps	+0.19	+0.7	-10.80	0.19	0.04	400, +0.00
Early feature, 400 steps	+3.66	+15.1	-7.33	0.53	0.70	400, +0.00
Blurred input, 400 steps	+5.09	+17.8	-5.90	0.57	0.72	400, +0.00
BSD100, twice the blur, null space 12.1%, frozen 19.96 dB, range oracle +9.26, $n = 52$						
Weights	+5.01	+28.3	-4.25	0.65	0.55	–
Bottleneck	+0.07	+0.1	-9.19	0.13	0.01	0/52
Early feature	+4.68	+26.1	-4.57	0.64	0.62	9/52
Early feature, 64^2	+3.13	+17.0	-6.12	0.50	0.47	5/52
Early feature, 32^2	+2.73	+14.8	-6.52	0.46	0.43	3/52
Blurred input	+4.84	+27.4	-4.42	0.65	0.61	12/52
Weights, 400 steps	+8.43	+40.2	-0.83	0.76	0.75	400, +0.00
Bottleneck, 400 steps	+0.29	+0.8	-8.96	0.16	0.07	400, +0.00
Early feature, 400 steps	+6.72	+34.3	-2.53	0.73	0.76	400, +0.00
Blurred input, 400 steps	+6.97	+36.0	-2.29	0.74	0.74	400, +0.00
BSD100, box blur (kernel family), null space 11.4%, frozen 20.72 dB, range oracle +10.39, $n = 52$						
Weights	+1.92	+7.6	-8.47	0.39	0.42	–
Bottleneck	+0.02	+0.0	-10.38	0.06	0.01	0/52
Early feature	+1.88	+7.8	-8.51	0.38	0.45	15/52
Early feature, 64^2	+1.01	+2.4	-9.39	0.24	0.32	3/52
Early feature, 32^2	+0.81	+2.8	-9.58	0.22	0.25	3/52
Blurred input	+1.88	+8.9	-8.51	0.37	0.42	13/52
Weights, 400 steps	+4.03	+15.2	-6.36	0.53	0.63	384, -0.01
Bottleneck, 400 steps	+0.17	+0.2	-10.23	0.11	0.04	400, +0.00
Early feature, 400 steps	+3.77	+17.4	-6.62	0.54	0.62	400, +0.00
Blurred input, 400 steps	+3.61	+17.1	-6.79	0.53	0.60	392, -0.02
Urban100 (content shift), null space 5.9%, frozen 21.35 dB, range oracle +11.64, $n = 52$						
Weights	+4.69	+29.7	-6.95	0.67	0.60	–
Bottleneck	+0.08	+0.2	-11.56	0.18	0.01	0/52
Early feature	+3.44	+21.7	-8.20	0.58	0.55	1/52
Early feature, 64^2	+2.26	+12.6	-9.38	0.45	0.43	1/52
Early feature, 32^2	+1.87	+10.4	-9.77	0.41	0.38	0/52
Blurred input	+3.51	+22.0	-8.13	0.58	0.56	3/52
Weights, 400 steps	+10.50	+52.8	-1.14	0.85	0.86	400, +0.00
Bottleneck, 400 steps	+0.43	+2.0	-11.21	0.26	0.08	400, +0.00
Early feature, 400 steps	+5.77	+33.5	-5.87	0.71	0.72	400, +0.00
Blurred input, 400 steps	+6.50	+34.5	-5.14	0.72	0.74	400, -0.00
IXI slices (far content shift), null space 5.9%, frozen 24.84 dB, range oracle +11.95, $n = 32$						
Weights	+6.97	+47.5	-4.98	0.85	0.77	–
Bottleneck	+0.04	+0.3	-11.91	0.34	0.01	0/32
Early feature	+5.76	+42.5	-6.20	0.82	0.76	0/32
Early feature, 64^2	+4.50	+33.7	-7.45	0.75	0.63	0/32
Early feature, 32^2	+4.05	+31.8	-7.90	0.73	0.62	0/32
Blurred input	+5.47	+41.1	-6.48	0.81	0.77	0/32
Weights, 400 steps	+13.67	+70.8	+1.72	0.96	0.94	400, +0.00
Bottleneck, 400 steps	+0.39	+2.7	-11.56	0.40	0.07	400, +0.00
Early feature, 400 steps	+9.41	+60.5	-2.54	0.92	0.89	400, +0.00
Blurred input, 400 steps	+9.64	+60.4	-2.31	0.92	0.90	399, -0.00

Table 37: **The fifth operator.** Deblurring through the released NAFNet, held-out images; forty steps at each variant's global rate, then 400. Columns as in Table 32. Predictions are evaluated on the null-space columns, since most of the error is range-space.

over to a restorer that is this far from consistent with the shifted operator. The frozen NAFNet leaves +10.99 dB to the range oracle, so weight-space descent has a mapping to change rather than a consistency to restore.

Z LIMITATIONS, IN FULL

The advantage of RDC at forty steps is a property of the step budget and of the speed of the variable. The same loss applied to the weights reaches our forty-step quality after 141 steps; on DRRs it matches RDC at its own plateau, and under the backbone's noise level it surpasses RDC there (Appendix T). The mechanism is decomposed on six to thirty scans, and on the real scans the primary backbone reconstructs through an idealised circular orbit at a cost of 1.75 dB (Appendix J). The operators after the first are two-dimensional and simulated from images, and every prediction made for them is reported together with its outcome (Appendices V, W and Y). On a pixel mask the alignment is zero for every variable under a shift of the mask and about 0.23 under a far shift of content (Appendix X). Labels remain better where they exist: DeepSparse fine-tuned on 600–800 labelled volumes reaches *31.14* dB and training from scratch reaches *30.44*, so our claim is made for sites without target ground truth. The criterion requires ground truth for $\cos\theta$. Its one-sided label-free form selects a width without labels that outperforms the weight-space variant on the spine split and is below it on the real scans, and it falls short of the best configuration on both (Appendix U). Withheld views recover the better variant without labels on 3 of 3 settings of that backbone, but carrying this procedure to the other operators is untested (Appendix R). Adaptation requires 3.5 min per scan rather than seconds; the distilled adapter removes this cost at half the gain. The per-scan implicit field (Zha et al., 2022) is run here on every scan of the lung row, but at one step budget (3000 steps) and on one of the five operators, so its 3.60 dB margin is a statement about this operator at that budget (Appendix B). The real cone-beam data are 5 walnut scans on a micro-CT, whose ground truth cannot represent the fracture line at 1.6 mm (Appendix M), and 18 clinical pelvic scans from one centre. On the clinical scans the reference is a deformed planning CT, valid only inside the field of view (Appendix J.2). The weight-space variant keeps its normalisation statistics at their pretraining values throughout, so the approach taken by TENT is not among the baselines (Appendix B).