RECO-SVF: GEOMETRIC REGION TOKENS FOR DENSE ATLAS REGISTRATION UNDER CONTROLLED SOURCE TRUNCATION

Anonymous authors
Paper under double-blind review

ABSTRACT

A deformation is predicted at every voxel, but the evidence that disambiguates a local match may lie far away—or be absent from a truncated moving volume. ReCo-SVF (Relational Context SVF) addresses this mismatch with a sparse-to-dense design: 32 learned region tokens exchange relative-position messages on a geometric 6-nearest-neighbour graph and condition a voxel-supported stationary velocity field through three-scale cross-attention. A six-carrier intervention family holds the dense prediction path fixed while replacing geometric tokens with no-token, dense, edge-free, random-edge, or coordinate-shuffled carriers. Across five seeds, the geometric carrier exceeds the edge-free 32-token mixer by +0.008 [0.005,0.011] Dice on complete OASIS images and +0.026 [0.019,0.033] at 40% shared-visible-field extent; random and shuffled neighbourhoods do not recover the full geometric carrier's accuracy. A moving-only protocol then removes source context while preserving the complete fixed target and ROI population. Over 10,900 degraded outputs per learned method, the geometric-carrier advantage over edge-free tokens widens by +0.027 [0.019,0.035], raises degradation AUC from 0.907 ± 0.006 to 0.934 ± 0.004, and reduces 40%-extent failures from 356 to 228 of 545 subject–offset cases. The complete system reaches 0.774 Dice with 0.038% folding on OASIS, transfers to IXI at 0.748 Dice without fine-tuning, and achieves 0.641 Dice with 0.052% folding on BTCV. These results show that explicit geometry in a fixed-budget context carrier preserves correspondence quality under controlled source truncation while retaining dense, regular deformation.

1 INTRODUCTION

A voxel in an atlas can have several plausible targets when nearby anatomy repeats, appearance changes across scanners, or the moving acquisition ends before the relevant structure. Dense registration must nevertheless assign a regular displacement at that voxel. The required evidence is therefore non-local even though the output is voxelwise—a mismatch that directly affects atlas propagation, longitudinal measurement, and population analysis (Sotiras et al., 2013; Avants et al., 2008).

Current learned registrars couple global transport to deformation support and regularization. Complete-volume accuracy therefore cannot identify whether a compact carrier benefits from spatial relations, especially when source truncation removes distant cues. We isolate that question with a fixed dense prediction path and a replaceable global carrier.

ReCo-SVF (Relational Context SVF) implements this factorization with 32 self-supervised region tokens. Their descriptors exchange relative-position messages on a geometric 6-nearest-neighbour graph, then condition decoder voxels at $1/16$, $1/8$, and $1/4$ resolution. A voxel-supported stationary velocity field is integrated by seven-step scaling and squaring. No-token, dense-memory, edge-free, random-edge, and coordinate-shuffled alternatives replace only this carrier.

Five-seed interventions compare all six carriers under a common encoder, SVF decoder, objective, and training protocol. A *context-retention curve* then reduces only moving-atlas extent while retaining the complete fixed target and its ROI population. For full-target Dice $q_m(e)$ at retained extent e, the

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.

response contrast

$$\Gamma_{40} = \big[q_{\mathrm{geo}}(.4) - q_{\mathrm{set}}(.4)\big] - \big[q_{\mathrm{geo}}(1) - q_{\mathrm{set}}(1)\big]$$

quantifies the geometric carrier's additional advantage as source context contracts. Shared crops assess jointly visible anatomy; IXI assesses scanner transfer, and BTCV evaluates the complete system on abdominal CT.

Compared with edge-free tokens, the geometric carrier adds +0.008 [0.005,0.011] Dice on OASIS and +0.026 [0.019,0.033] at the 40% shared-visible-field endpoint. Across 10,900 moving-only outputs per learned method, its advantage widens by +0.027 [0.019,0.035], and degradation AUC rises from 0.907 ± 0.006 to 0.934 ± 0.004. The full system reaches 0.774 Dice with 0.038% folding on OASIS, transfers to IXI at 0.748 Dice, and achieves 0.641 Dice with 0.052% folding on BTCV.

The contribution is an intervenable sparse relational carrier and a context response estimand. The carrier ladder separates tokenization, support density, generic graph computation, and true-coordinate neighbourhood organization; controlled source removal tests their response while transfer, topology, and anatomical analyses connect the mechanism to dense registration.

2 Related Work

Dense deformation predictors. SyN and LDDMM optimize regularized transformations independently for each image pair (Avants et al., 2008; Beg et al., 2005). Learned CNN systems amortize that optimization while preserving dense spatial feature hierarchies (Balakrishnan et al., 2019; de Vos et al., 2019; Mok & Chung, 2020). Transformer registrars extend the same dense support with image-token or cross-attention interactions (Chen et al., 2022; Shi et al., 2022; Liu et al., 2024). Other systems strengthen the training distribution or deformation objective: SynthMorph targets contrast invariance through synthetic anatomy (Hoffmann et al., 2022), and uniGradICON obtains broad zero-shot behaviour from inverse-consistency regularization (Tian et al., 2024). These methods establish strong complete-system operating points, but global information transport, deformation support, and regularization differ across model families. ReCo-SVF instead makes the carrier an explicit experimental interface: its token support and relation operator can change while the dense SVF prediction family remains fixed. This enables a within-family test of carrier organization rather than inferring that mechanism from a cross-architecture leaderboard.

Sparse supports in registration. Invariant landmarks and KeyMorph recover transforms from sparse matched locations (Rister et al., 2017; Wang et al., 2023). GeoReg represents and updates deformation on graph nodes (Sideri-Lampretsa et al., 2024), whereas H-SGANet places sparse graph attention inside a hybrid feature extractor (Zhou & Cao, 2025). ReCo-SVF factorizes these roles: 32 region tokens transport global evidence, a geometric graph organizes token interaction, and a separately supported dense SVF produces the deformation. The sparse regions are therefore neither landmarks to be matched nor nodes on which the final transform is defined. The resulting boundary permits edge-free, random-edge, shuffled-coordinate, dense-carrier, and no-token interventions under the same voxelwise prediction head, directly distinguishing context support from deformation support.

Architectural comparison: global carrier, deformation support, and evidence that identifies the carrier's contribution.

Method	Global carrier	Deformation	Relation operator	Matched	Curve
TransMorph	dense tokens	dense field	self-attention	–	–
VFA	dense features	dense field	vector attention	–	–
H-SGANet	feature graph	dense field	graph attention	–	–
KeyMorph	keypoints	transform	matching	–	–
GeoReg	moving nodes	graph nodes	dynamic graph	–	–
uniGradICON	dense features	dense field	convolutions	–	–
ReCo-SVF	32 regions	dense SVF	geometric GAT	✓	✓

From robustness scores to carrier response. Domain-shift evaluations document that learned registration can change under seemingly modest acquisition differences (Jena et al., 2024), while ProactiV-Reg probes model requirements through learned inputs (Kolenbrander et al., 2024). We connect this evaluation concern to a specific architectural interface. Acquisition-site strata measure

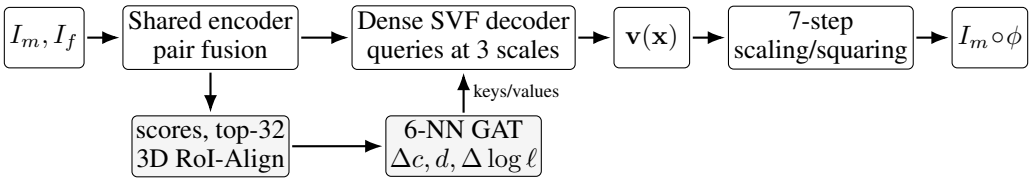

Figure 1: ReCo-SVF's sparse-to-dense factorization. Pair-fused encoder features both drive the dense SVF path and score latent pooling windows. The selected $N = 32$ descriptors exchange relative centre, distance, and scale messages, then serve as keys and values for voxel queries at three decoder resolutions. Carrier interventions replace only the shaded information path.

scanner transfer, whereas retained physical extent controls the context supplied by the moving atlas. The curve-level difference-in-differences asks whether the complete geometric carrier and its edge-free counterpart separate as that context is removed. It therefore complements a single endpoint robustness score with a response trajectory; degradation AUC, lower-tail quality, failure counts, and propagated-volume error connect the trajectory to anatomical consequences.

3 METHOD

3.1 FACTORIZING CONTEXT TRANSPORT FROM DENSE DEFORMATION

Let I_m and I_f be moving and fixed volumes on $\Omega \subset \mathbb{R}^3$. We use the backward-warp convention: $\phi : \Omega_f \to \Omega_m$ maps each fixed-grid coordinate to the moving image, and the registered image is $I_m \circ \phi(\mathbf{x}) = I_m(\phi(\mathbf{x}))$. A differentiable trilinear sampler implements this operation (Jaderberg et al., 2015). The predicted object is a dense stationary velocity field $\mathbf{v} \in \mathbb{R}^{H \times W \times D \times 3}$; seven scaling-and-squaring steps produce $\phi = \exp(\mathbf{v})$ (Dalca et al., 2019). ReCo-SVF factorizes the system into an intervenable context carrier and an invariant dense prediction path. The carrier transports paired-image evidence; the SVF support, integration, and voxelwise output remain dense.

A shared-weight four-level encoder processes I_m and I_f with channel widths 32, 64, 128, and 256. At each level, concatenation $[F_m, F_f, |F_m - F_f|]$ followed by a 1^3 convolution fuses the two streams. The decoder uses widths 256, 128, 64, and 32 with trilinear upsampling and skip fusion. At the $1/4$ level, the fused pair also enters a proposal head that produces a fixed budget of $N = 32$ latent pooling windows. A graph encoder transforms their pooled descriptors, and decoder voxels query the tokens at three resolutions. Figure 1 shows this separation.

3.2 SELF-SUPERVISED PROPOSAL TOKENS

The selector learns proposal scores and continuous window parameters, followed by deterministic hard top-N ranking and 3D NMS. This yields a fixed context budget while retaining gradient paths through selected windows and descriptors.

Window parameterization. The proposal head follows an anchor-free region-proposal design (Ren et al., 2015). Three 3^3 stride-1 convolutions with GroupNorm and ReLU emit, at every $1/4$-resolution anchor, a score, a centre offset, and three side lengths. A $\tanh$ bounds each centre offset to one feature cell; a sigmoid bounds side lengths to $[0.08, 0.35]$ of the corresponding full-volume axis. Coordinates are normalized to $[0, 1]^3$ in the fixed-image frame. Paired descriptors are sampled with $3{\times}3{\times}3$ 3D RoI-Align bins (He et al., 2017), average pooled, projected to 128 dimensions, and augmented by a sinusoidal encoding of centre and log side length.

Score supervision and hard selection. For anchor a and predicted window w_a, the self-supervised informativeness target combines paired feature variation and agreement:

$$q_a = \mathrm{sg}\big[\sigma\big(z\big\{\tfrac{1}{2}\,\mathrm{Var}_{w_a}(F_m) + \tfrac{1}{2}\,\mathrm{Var}_{w_a}(F_f) - \mathrm{mean}_{w_a}\|F_m - F_f\|_1\big\}\big)\big]\,,$$

where $z\{\cdot\}$ standardizes scores within a volume, σ is the logistic function, and sg stops gradients through the target. Binary cross-entropy between the emitted score and q_a, weighted by $\lambda_p = 0.05$,

trains the ranking branch. Registration gradients pass through RoI-Align to selected box coordinates, descriptors, and the shared proposal trunk; top-N ranking and NMS receive no gradient. We apply 3D NMS at IoU 0.1, retain 32 boxes, and break equal-score ties lexicographically by (z, y, x) anchor index. These hard heuristics define repeatable latent pooling windows rather than anatomy detections; no boxes, landmarks, or segmentation labels supervise selection.

Geometric graph. Each proposal connects to its six nearest centres by Euclidean distance in normalized fixed coordinates. We form an undirected graph by taking the union of directed neighbour sets, remove self-loops, and encode relative centre offset, distance, and log side-length ratio on every edge. Three pre-normalized GAT layers (Veličković et al., 2018), each with four 32-dimensional heads, concatenated output, residual connection, and two-layer GELU feed-forward block, yield tokens $\{\mathbf{t}_i\}_{i=1}^{32} \subset \mathbb{R}^{128}$. The same $N = 32, k = 6$ setting is used for OASIS, IXI, BTCV, and all context-loss evaluations; Appendix A.7 shows a broad optimum around this operating point.

3.3 Sparse-Context-Conditioned SVF Decoder

At decoder resolutions $1/16$, $1/8$, and $1/4$, voxel features are projected to 128-dimensional queries and attend to the 32 tokens with four-head cross-attention. Each block applies LayerNorm, cross-attention, residual addition, LayerNorm, and a two-layer GELU feed-forward residual in that order. A scale-specific linear projection returns conditioned features to the decoder width. For V decoder voxels, this conditioning costs $\mathcal{O}(32V)$ instead of the $\mathcal{O}(V^2)$ cost of dense global self-attention. The final $1/4$ representation is mapped by a 1^3 convolution to $\mathbf{v}$, trilinearly upsampled, and integrated. Thus graph nodes carry context, whereas every output displacement is generated by the dense SVF head.

3.4 Training Objective

We optimize

$$\mathcal{L} = \mathcal{L}_{\text{sim}} + \lambda_s \mathcal{L}_{\text{smooth}} + \lambda_j \mathcal{L}_{\text{jac}} + \lambda_g \mathcal{L}_{\text{graph}} + \lambda_p \mathcal{L}_{\text{proposal}} + \lambda_u \mathcal{L}_{\text{unc}}.$$

Similarity and smoothness. $\mathcal{L}_{\text{sim}}$ is the negative local normalized cross-correlation minimized over 9^3 windows,

$$\mathcal{L}_{\text{sim}} = -\,\text{LNCC}_{9^3}(I_m \circ \phi, I_f).$$

$\mathcal{L}_{\text{smooth}} = \|\nabla \mathbf{v}\|_2^2$ with $\lambda_s = 1.0$.

Topology and relational coherence. With $\lambda_j = 0.5$, the hinge penalizes Jacobian determinants below the positive margin $\epsilon = 0.01$:

$$\mathcal{L}_{\text{jac}} = \frac{1}{|\Omega|} \sum_{\mathbf{x}} \max\big(0, \epsilon - \det J_\phi(\mathbf{x})\big).$$

For a fixed-frame proposal centre v_i, let $v_i' = \phi(v_i)$ be its location in the moving frame. With $\lambda_g = 0.1$, we use

$$\mathcal{L}_{\text{graph}} = \frac{1}{|\mathcal{E}|} \sum_{(i,j)\in\mathcal{E}} \big|d(v_i, v_j) - d(v_i', v_j')\big|.$$

Distances are measured after converting normalized coordinates to physical millimetres. This local edge-length prior is anatomy-agnostic: in both brain and abdomen it suppresses proposal collapse and extreme neighbourhood stretching while allowing non-rigid motion through a soft absolute penalty.

Auxiliary residual regularizer. Let D_{-1} and D_0 denote the penultimate and final decoder refinements. Training-only 1^3 heads map them to a detached residual target $u^* = \text{sg}[\text{up}(h_{-1}(D_{-1}))]$, an estimate $\hat{u} = h_0(D_0)$, and a raw scale $s = h_s(D_0)$. Appendix A.2 defines the heteroscedastic objective using the bounded-positive parameterization $\sigma^2 = 10^{-6} + \text{softplus}(s)$ and weight $\lambda_u = 0.01$. These auxiliary heads are removed at inference and do not alter $\mathbf{v}$ or ϕ.

4 EXPERIMENTS

Carrier identification is the empirical spine of the evaluation. The six-variant study first separates a geometric region carrier from tokenization, dense support, and disrupted graph controls. Context-retention curves then measure how the geometric-versus-edge-free design contrast changes as source evidence is removed. Complete-system comparisons establish the associated accuracy–topology operating point, while scanner transfer and anatomical-tail metrics test where that operating point remains useful.

4.1 DATA AND REGISTRATION TASKS

OASIS-1. We use 414 T1-weighted brain MRI scans (Marcus et al., 2007), resampled to 1 mm isotropic $160\times192\times224$ grids and split 255/50/109 for training/validation/testing. N4 bias correction and z-score normalization precede affine alignment to the SRI24 atlas (Tustison et al., 2010). FreeSurfer-derived labels are mapped to the same 35 cortical and subcortical ROIs for every subject. The test task contains 109 atlas-to-subject pairs; the SRI24 intensity atlas and its mapped 35-ROI label map supply the moving image and labels, and subject labels are used only for evaluation.

IXI transfer. The external test uses 115 T1-weighted scans from Guy's (42), Hammersmith (39), and IOP (34) (Biomedical Image Analysis Group, Imperial College London, n.d.). Scans come from Philips 1.5 T, Philips 3 T, and GE 1.5 T systems, respectively. FreeSurfer 7.4.1 `recon-all` produces the evaluation parcellations; the same bilateral Desikan–Killiany and subcortical merge map used for OASIS yields 35 ROIs. Automated topology checks followed by blinded slice-wise inspection verify label completeness and atlas orientation. The selected 115 scans all pass this fixed rule. IXI images and labels are excluded from training, hyperparameter selection, and checkpoint selection.

BTCV. We use 30 contrast-enhanced abdominal CT volumes with 13 organ labels (Synapse, 2015). Volumes are clipped to $[-175, 250]$ HU, linearly normalized, and resampled to 2 mm isotropic 128^3 tensors. Twenty development subjects are fixed before training: 18 train the models and two select checkpoints by mean validation Dice. The remaining 10 subjects form 90 ordered inter-subject test pairs. Statistical intervals cluster observations by both moving and fixed test subject, preventing the 90 pairings from being treated as independent.

4.2 EXTERNAL COMPARATORS AND CARRIER INTERVENTIONS

The standard comparison includes SyN (Avants et al., 2008), diffeomorphic VoxelMorph (Balakrishnan et al., 2019), LapIRN (Mok & Chung, 2020), and TransMorph (Chen et al., 2022). A same-protocol modern block evaluates H-SGANet (Zhou & Cao, 2025), VFA (Liu et al., 2024), and GeoReg (Sideri-Lampretsa et al., 2024) after retraining on the OASIS split, plus the released uniGradICON model (Tian et al., 2024) in its zero-shot configuration. Preprocessing, pair construction, interpolation, and metric code are shared. Published cross-paper numbers are never mixed with these evaluations. The external methods test complete systems; the carrier family below supplies the matched mechanism comparison.

The carrier family holds the encoder, dense SVF decoder, augmentation, optimizer, schedule, and objective weights fixed. Every variant retains the selector and evaluates $\mathcal{L}_{\mathrm{graph}}$ on the true-coordinate reference graph independently of its message edges. The six carriers are no tokens, the fused $1/4$ lattice, edge-free set attention, degree-matched random edges, coordinate-shuffled 6-NN, and true 6-NN. Random graphs and coordinate permutations are fixed by seed and volume. Selector design, initialization, and budget are matched, while independently optimized models learn their own windows. Appendix A.3 specifies attributes, gradients, and losses; Appendix A.4 reports capacity. The disrupted graphs separate useful locality from edge count and generic message passing; TransMorph is an external dense reference.

4.3 TRAINING AND COMPUTE

Models are implemented in PyTorch 2.1.2 with Python 3.10.13, CUDA 12.2, and cuDNN 8.9.7. Training uses one NVIDIA A100 80 GB, AdamW (Loshchilov & Hutter, 2019), learning rate 10^{-4},

weight decay 10^{-5}, cosine annealing, batch size one, and mixed precision. Brain and abdomen schedules run for 500 and 800 epochs. Augmentation comprises affine rotation within $\pm 15^{\circ}$, scale 0.9–1.1, and intensity shift ± 0.1. For every retrained method and seed, the epoch with highest mean validation Dice is selected using the same validation split and without access to test labels. The standard and component-ablation tables use seed 2026 and take 21.3 h on OASIS and 28.7 h on BTCV. The matched mechanism, modern retrained-baseline, and asymmetric-FOV studies use seeds $\{2024, 2025, 2026, 2027, 2028\}$; reported seed-level standard deviations are distinct from pair-level dispersion. Propagated-structure utility and topology reuse these five-seed outputs; proposal and Jacobian sensitivity use the first three seeds.

Learned inference times use the A100, batch size one, 20 warm-up volumes, and the mean of 100 synchronized runs. SyN uses ANTs 2.4.3 cross-correlation on an AMD EPYC 7742 CPU with 16 threads; its CPU time is reported separately from GPU timing.

4.4 Context-Retention Protocol and Estimands

The shared-crop reference applies the same centred crop to moving image, fixed image, masks, and labels at 85/70/55/40% retained physical extent per axis, then zero-pads to the original tensor size. These settings retain 61.4/34.3/16.6/6.4% of volume. The fixed 40% endpoint is the symmetric reference for the offset-crop response. Co-cropping labels and excluding empty ROIs makes this a controlled measure of alignment within shared visible anatomy. The asymmetric protocol below complements it with a fixed target ROI population.

The asymmetric protocol crops only the moving-image intensities at the same four extents. For each subject and level, five offsets are fixed in normalized cropable margin: $(0, 0, 0)$, $(.25, .25, 0)$, $(-.25, -.25, 0)$, $(0, .25, -.25)$, and $(0, -.25, .25)$. The retained intensity and its binary crop support are zero-padded without resampling; the fixed image, fixed labels, and fixed foreground remain complete. The complete moving-atlas label map is never an input and remains uncropped solely for evaluation, so full-target scoring asks where atlas anatomy maps when its corresponding image context is absent. The crop support in the common affine grid is intersected with the fixed foreground before inference to form a method-independent common-valid mask. No model is fine-tuned. We report Dice on this mask and on the unchanged set of all 35 target ROIs, normalized degradation AUC, and failures below 0.65. The design yields $109 \times 4 \times 5 \times 5 = 10{,}900$ degraded outputs per learned method.

Let $q_{mso}(e)$ be full-target Dice for method m, seed s, subject-offset pair o, and retained per-axis extent $e \in \{1, .85, .70, .55, .40\}$. We define retention as $r_{mso}(e) = q_{mso}(e)/q_{mso}(1)$ and degradation AUC as the normalized trapezoidal area of r_{mso} over $e \in [.4, 1]$, averaged only after forming each subject-offset-seed curve. The geometric-versus-edge-free relation-by-context interaction is

$$\Gamma_{40} = \big[q_{\mathrm{geo}}(.4) - q_{\mathrm{set}}(.4)\big] - \big[q_{\mathrm{geo}}(1) - q_{\mathrm{set}}(1)\big].$$

A positive Γ_{40} means the geometric carrier's advantage over the edge-free design increases as context contracts. Here q scores the unchanged set of fixed-target ROIs at every extent. Hierarchical resampling forms this contrast within subjects, offsets, and seeds.

4.5 Metrics and Inference

Dice and HD95 quantify propagated-label fidelity. Dice averages first over non-empty ROIs within a pair and then over pairs. HD95 is computed from symmetric surface distances in millimetres using each resampled image's physical spacing, then aggregated by the same two-stage rule. We also report median absolute propagated-volume error, bottom-decile ROI/organ Dice, and raw pairwise wins. Against a non-empty target ROI, an empty propagated label is assigned Dice zero and HD95 equal to the fixed-grid physical diagonal; empty target ROIs are excluded before method comparison. Full-target metrics include crop-exterior target voxels and therefore retain the same 35-ROI scoring population at every extent. Topology measures include the percentage of voxels with $\det J_\phi \leq 0$, any-fold cases, SDlogJ over positive determinants, and the first determinant percentile.

Five-seed summaries are means and standard deviations across seed-level test means. Paired effects use a 10,000-resample hierarchical bootstrap over subjects, offsets within subject where applicable, and seeds. Win/tie/loss counts use pair-level mean Dice across seeds, with 10^{-4} defining a tie.

Table 1: Complete-system comparison. OASIS and BTCV entries are seed-2026 pair-level mean±standard deviation; BTCV comparators are the standard block listed in Section 4. The modern block reports means of five seed-level test means for retrained methods; uniGradICON uses released zero-shot weights. Folding is the percentage of voxels with $\det J_\phi \leq 0$.

Seed-2026: OASIS brain MRI (109 pairs)				
Method	Dice↑	HD95↓	Fold.%↓	Time (s)
SyN (Avants et al., 2008)	0.724 ± 0.029	2.53 ± 0.44	< 0.001	126.3
VoxelMorph (Balakrishnan et al., 2019)	0.729 ± 0.027	2.44 ± 0.39	1.59 ± 0.32	0.41
LapIRN (Mok & Chung, 2020)	0.746 ± 0.025	2.26 ± 0.36	0.87 ± 0.24	0.53
TransMorph (Chen et al., 2022)	0.753 ± 0.024	2.19 ± 0.34	1.61 ± 0.35	0.84
ReCo-SVF	0.774 ± 0.021	1.92 ± 0.28	0.038 ± 0.014	1.12
Seed-2026: BTCV abdominal CT (90 ordered pairs)				
SyN	0.581 ± 0.068	8.74 ± 2.38	< 0.001	243.7
VoxelMorph	0.564 ± 0.061	9.31 ± 2.52	2.41 ± 0.71	0.47
LapIRN	0.597 ± 0.056	8.12 ± 2.17	1.52 ± 0.48	0.56
TransMorph	0.608 ± 0.054	7.86 ± 2.04	2.24 ± 0.63	0.91
ReCo-SVF	0.641 ± 0.046	6.93 ± 1.78	0.052 ± 0.021	1.18

Modern same-protocol evaluation				
Method	OASIS Dice	IXI Dice	Shared 40% Dice	Mode
H-SGANet	0.759	0.724	0.641	retrained
VFA	0.766	0.733	0.652	retrained
GeoReg	0.768	0.739	0.663	retrained
uniGradICON	0.750	0.741	0.674	zero-shot
ReCo-SVF	0.773	0.747	0.692	retrained

Thresholded context failures likewise use each subject–offset's Dice averaged over the five seeds, giving 545 observations per extent; structure-level utility wins compare five-seed mean scores. Holm correction controls the three primary full-versus-edge-free contrasts.

5 Results

Table 2 provides five-seed inference for the carrier comparison, and the complete-system tables establish its practical accuracy–topology operating point.

5.1 Complete-System Accuracy–Topology Operating Point

Table 1 places ReCo-SVF against geometric and attention registrars under one protocol. The seed-2026 checkpoint reaches 0.774 Dice, 1.92 mm HD95, and 0.038% folding on OASIS. On BTCV, the same operating-point comparison improves over TransMorph by 3.3 Dice points and 0.93 mm while reducing folding from 2.24% to 0.052%. In the five-seed modern block, ReCo-SVF records the largest observed mean on OASIS, IXI, and the shared 40% endpoint. The matched carrier effects below provide direct paired intervals for the architectural contrast.

The frontier in Figure 2 shows that the accuracy gains in the plotted standard block do not arise from accepting irregular warps. ReCo-SVF occupies the highest-Dice, lowest-folding region among its learned comparators on both anatomies. SyN's folding is censored at $< 0.001\%$ and shown as an upper-bound marker.

5.2 Matched Carrier Interventions Identify the Useful Design

Table 2 intervenes on the global carrier while holding the selector design and objective, dense decoder, regularizers, and training schedule constant; realized windows may differ after independent optimization, as specified in Appendix A.3. Edge-free tokens recover part of the full-context gain over a dense carrier. True neighbourhoods then add an effect that grows across OASIS, IXI, and the hardest

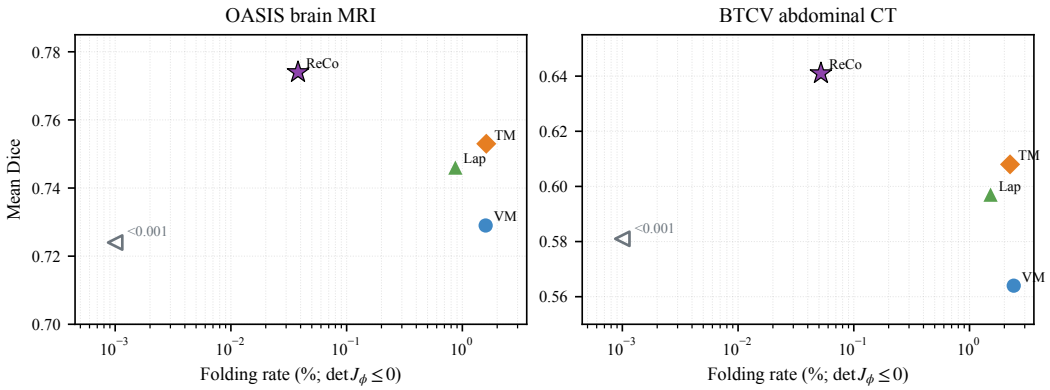

Figure 2: Dice–folding frontier for the OASIS and BTCV results. The left-pointing SyN markers denote the reported upper bound $< 0.001\%$; they are not treated as exact 0.001% measurements.

Table 2: Carrier mechanism identification. The six SVF-family variants use the intervention contract in Appendix A.3; TransMorph is an external reference. Entries are mean±standard deviation across five seed-level test means. "Shared 40" is the 40%-per-axis centred shared-visible-field endpoint. Paired geometric-minus-edge-free intervals are reported in the text.

Variant	OASIS Dice	IXI Dice	Shared 40 Dice	OASIS Fold.%
TransMorph	0.752 ± 0.003	0.717 ± 0.004	0.617 ± 0.008	1.63 ± 0.14
No token branch	0.750 ± 0.004	0.711 ± 0.005	0.603 ± 0.010	0.042 ± 0.005
Matched dense carrier	0.758 ± 0.004	0.724 ± 0.005	0.631 ± 0.009	0.040 ± 0.005
Edge-free 32-token mixer	0.765 ± 0.003	0.736 ± 0.004	0.666 ± 0.008	0.040 ± 0.004
Degree-matched random edges	0.759 ± 0.004	0.729 ± 0.005	0.651 ± 0.010	0.042 ± 0.005
Coordinate-shuffled 6-NN	0.761 ± 0.004	0.731 ± 0.005	0.657 ± 0.009	0.042 ± 0.005
ReCo-SVF, geometric 6-NN	0.773 ± 0.002	0.747 ± 0.003	0.692 ± 0.006	0.039 ± 0.004

shared-visible-field endpoint. The complete geometric-carrier design exceeds the edge-free design by +0.008 [0.005,0.011] Dice on OASIS, +0.011 [0.007,0.015] on IXI, and +0.026 [0.019,0.033] at the 40% shared crop. The corresponding win counts are 76/109, 83/115, and 91/109; all three Holm-adjusted intervals exclude zero. Random and coordinate-shuffled edges underperform the true geometric graph. Because these controls preserve graph computation while breaking locality or descriptor–coordinate binding, their full-context ordering supports useful neighbourhood organization beyond token count or generic message-passing capacity. The paired intervals estimate the complete geometric-versus-edge-free carrier contrast under the shared true-coordinate graph regularizer and independently learned windows.

The component ablation in Appendix A.1 complements this matched study: removing the Jacobian loss changes folding from 0.038% to 0.83%. This identifies explicit Jacobian control as the source of the low-folding operating point within the fixed SVF family, while the carrier interventions primarily change correspondence quality.

5.3 Source Truncation Amplifies the Carrier Advantage

We next vary context in two complementary ways: IXI changes the acquisition distribution while retaining complete anatomy, and controlled cropping removes image extent while keeping model weights fixed. Together they test whether the design contrast identified above strengthens when familiar evidence weakens.

Table 3 reports the seed-2026 standard block on IXI without fine-tuning. ReCo-SVF loses 2.6 Dice points from OASIS and reaches 0.748 Dice, 2.18 mm HD95, and 0.046% folding. Its approximately three-point Dice advantage over TransMorph appears at each acquisition site, with paired intervals

Table 3: Seed-2026 OASIS→IXI transfer on 115 atlas-to-subject pairs without fine-tuning. The overall block gives pair-level mean±standard deviation and change from the corresponding OASIS score. The site block gives the paired ReCo-SVF–TransMorph 95% interval; W/T/L uses 10^{-4} ties.

Method	Dice↑	HD95↓	Fold.%↓	Transfer Δ
VoxelMorph	0.694 ± 0.034	2.87 ± 0.48	1.72 ± 0.38	-3.5
LapIRN	0.712 ± 0.031	2.61 ± 0.43	0.94 ± 0.29	-3.4
TransMorph	0.718 ± 0.029	2.56 ± 0.42	1.78 ± 0.41	-3.5
ReCo-SVF	0.748 ± 0.025	2.18 ± 0.33	0.046 ± 0.019	-2.6
Site (n)	**ReCo-SVF Dice**	**TransMorph Dice**	**Paired 95% CI**	**W/T/L**
Guy's (42)	0.750 ± 0.024	0.721 ± 0.028	$[0.022, 0.036]$	31/1/10
Hammersmith (39)	0.746 ± 0.026	0.717 ± 0.030	$[0.021, 0.037]$	29/1/9
IOP (34)	0.747 ± 0.025	0.715 ± 0.029	$[0.023, 0.041]$	26/0/8

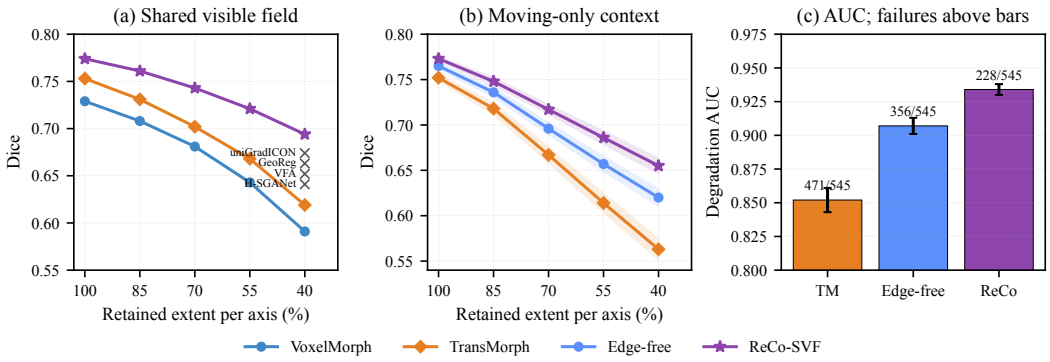

Figure 3: Response to reduced field of view. (a) Shared-visible-field stress with modern 40% markers. (b) Five-seed moving-only curves over fixed target ROIs; bands show seed-level standard deviation. (c) Per-curve degradation AUC and failures after averaging subject–offset scores across seeds. Γ_{40} is the geometric-versus-edge-free design widening in (b).

excluding zero. This seed-2026 block is a transfer operating point; site labels are evaluation strata and never enter model selection.

Figure 3 traces both symmetric and source-only interventions. The seed-2026 shared centred crop gives a visible-anatomy stress reference: its 40% endpoint is 0.694 for ReCo-SVF, 0.619 for TransMorph, and 0.591 for VoxelMorph. The moving-only offset protocol supplies Γ_{40} by preserving the complete fixed target and a constant ROI population while reducing source evidence. Across 10,900 degraded outputs per learned method, degradation AUC is 0.934 ± 0.004 for ReCo-SVF, 0.907 ± 0.006 for edge-free tokens, and 0.852 ± 0.009 for TransMorph. From 100% to 40% extent, the geometric-versus-edge-free gap expands by +0.027 [0.019,0.035]. This positive relation-by-context interaction shows that the complete geometric carrier separates further from the edge-free design when source evidence contracts. At 40%, failures below 0.65 occur in 228/545, 356/545, and 471/545 subject-offset observations, respectively.

6 Conclusion

ReCo-SVF makes global context an explicit fixed-budget interface: 32 geometrically organized regions condition a dense SVF without moving the deformation onto sparse support. The matched ladder shows benefits from sparse regions and from true-coordinate neighbourhoods, and the moving-only curve widens the geometric-versus-edge-free contrast by +0.027 [0.019,0.035] at 40% source extent. The full system reaches 0.774 Dice with 0.038% folding on OASIS and transfers to IXI at 0.748 Dice.

REFERENCES

Brian B. Avants, Charles L. Epstein, Murray Grossman, and James C. Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, 2008. doi: 10.1016/j.media.2007.06.004.

Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. VoxelMorph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, 38(8):1788–1800, 2019. doi: 10.1109/TMI.2019.2897538.

M. Faisal Beg, Michael I. Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*, 61(2):139–157, 2005. doi: 10.1023/B:VISI.0000043755.93987.aa.

Biomedical Image Analysis Group, Imperial College London. IXI dataset. `https://brain-development.org/ixi-dataset/`, n.d. EPSRC project GR/S21533/02; accessed 2026-01-15.

Junyu Chen, Eric C. Frey, Yufan He, William P. Segars, Ye Li, and Yong Du. TransMorph: Transformer for unsupervised medical image registration. *Medical Image Analysis*, 82:102615, 2022. doi: 10.1016/j.media.2022.102615.

Adrian V. Dalca, Guha Balakrishnan, John Guttag, and Mert R. Sabuncu. Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Medical Image Analysis*, 57: 226–236, 2019. doi: 10.1016/j.media.2019.07.006.

Bob D. de Vos, Floris F. Berendsen, Max A. Viergever, Hessam Sokooti, Marius Staring, and Ivana Išgum. A deep learning framework for unsupervised affine and deformable image registration. *Medical Image Analysis*, 52:128–143, 2019. doi: 10.1016/j.media.2018.11.010.

Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. In *ICCV*, pp. 2980–2988, 2017. doi: 10.1109/ICCV.2017.322.

Malte Hoffmann, Benjamin Billot, Douglas N. Greve, Juan Eugenio Iglesias, Bruce Fischl, and Adrian V. Dalca. SynthMorph: Learning contrast-invariant registration without acquired images. *IEEE TMI*, 41(3):543–558, 2022. doi: 10.1109/TMI.2021.3116879.

Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. Spatial transformer networks. In *NeurIPS*, 2015. URL `https://proceedings.neurips.cc/paper/2015/hash/33ceb07bf4eeb3da587e268d663aba1a-Abstract.html`.

Rohit Jena, Deeksha Sethi, Pratik Chaudhari, and James C. Gee. Deep learning in medical image registration: Magic or mirage? In *NeurIPS*, volume 37, pp. 108331–108353, 2024. doi: 10.52202/079017-3439.

Alex Kendall and Yarin Gal. What uncertainties do we need in Bayesian deep learning for computer vision? In *NeurIPS*, 2017. URL `https://proceedings.neurips.cc/paper_files/paper/2017/hash/2650d6089a6d640c5e85b2b88265dc2b-Abstract.html`.

Iris D. Kolenbrander, Vidya Prasad, Leanne Zikken, Maureen A. J. M. van Eijnatten, Matteo Maspero, and Josien P. W. Pluim. Assessing the robustness of image registration models under domain shifts with learnable input images. In *Biomedical Image Registration*, pp. 101–111. Springer, 2024. doi: 10.1007/978-3-031-73480-9_8.

Yihao Liu, Junyu Chen, Lianrui Zuo, Aaron Carass, and Jerry L. Prince. Vector field attention for deformable image registration. *Journal of Medical Imaging*, 11(6):064001, 2024. doi: 10.1117/1.JMI.11.6.064001.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *ICLR*, 2019. URL `https://openreview.net/forum?id=Bkg6RiCqY7`.

Daniel S. Marcus, Tracy H. Wang, Jamie Parker, John G. Csernansky, John C. Morris, and Randy L. Buckner. Open access series of imaging studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience*, 19(9): 1498–1507, 2007. doi: 10.1162/jocn.2007.19.9.1498.

Tony C. W. Mok and Albert C. S. Chung. Large deformation diffeomorphic image registration with laplacian pyramid networks. In *MICCAI*, pp. 211–221, 2020. doi: 10.1007/978-3-030-59716-0_21.

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *NeurIPS*, pp. 91–99, 2015. doi: 10.5555/2969239.2969250.

Blaine Rister, Mark A. Horowitz, and Daniel L. Rubin. Volumetric image registration from invariant keypoints. *IEEE Transactions on Image Processing*, 26(10):4900–4910, 2017. doi: 10.1109/TIP.2017.2722689.

Jiacheng Shi, Yuting He, Youyong Kong, Jean-Louis Coatrieux, Huazhong Shu, Guanyu Yang, and Shuo Li. XMorpher: Full transformer for deformable medical image registration via cross attention. In *MICCAI*, volume 13436, pp. 217–226, 2022. doi: 10.1007/978-3-031-16446-0_21.

Vasiliki Sideri-Lampretsa, Nil Stolt-Ansó, Huaqi Qiu, Julian McGinnis, Wenke Karbole, Martin J. Menten, and Daniel Rueckert. Interpretable deformable image registration: A geometric deep learning perspective. *arXiv preprint arXiv:2412.13294*, 2024. doi: 10.48550/arXiv.2412.13294.

Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190, 2013. doi: 10.1109/TMI.2013.2265603.

Synapse. Segmentation outside the cranial vault challenge, 2015. URL `https://doi.org/10.7303/syn3193805`.

Lin Tian, Hastings Greer, Roland Kwitt, François-Xavier Vialard, Raúl San José Estépar, Sylvain Bouix, Richard Rushmore, and Marc Niethammer. uniGradICON: A foundation model for medical image registration. In *MICCAI*, volume 15002, pp. 749–760, 2024. doi: 10.1007/978-3-031-72069-7_70.

Nicholas J. Tustison, Brian B. Avants, Philip A. Cook, Yuanjie Zheng, Alexander Egan, Paul A. Yushkevich, and James C. Gee. N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6):1310–1320, 2010. doi: 10.1109/TMI.2010.2046908.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *ICLR*, 2018. URL `https://openreview.net/forum?id=rJXMpikCZ`.

Alan Q. Wang, Evan M. Yu, Adrian V. Dalca, and Mert R. Sabuncu. A robust and interpretable deep learning framework for multi-modal registration via keypoints. *Medical Image Analysis*, 90:102962, 2023. doi: 10.1016/j.media.2023.102962.

Yufeng Zhou and Wenming Cao. H-SGANet: Hybrid sparse graph attention network for deformable medical image registration. *Neurocomputing*, 633:129810, 2025. doi: 10.1016/j.neucom.2025.129810.

Reproducibility Statement

Sections 3–5 specify the model, objectives, data partitions, preprocessing, training protocol, evaluation metrics, and statistical estimands. Appendix A provides architecture and intervention-contract details, sensitivity analyses, curve points, utility, and topology.

Ethics Statement

OASIS, IXI, and BTCV are public de-identified research datasets used under their access terms. Site-stratified results and explicit failure counts make population and acquisition effects inspectable. Propagated labels and deformations are evaluated as research outputs rather than clinical decisions.

AI USE STATEMENT

Generative AI tools assisted manuscript drafting, restructuring, language editing, LaTeX formatting, and bibliographic consistency checks during documented submission preparation. These tools did not generate or alter experimental measurements or figures in this preparation pass. The authors are responsible for the final submission.

A TECHNICAL APPENDIX

A.1 COMPONENT ATTRIBUTION

The seed-2026 component study in Table 4 preserves the complete model's data, optimizer, and architecture wherever the named component allows. Removing graph tokens causes the largest Dice reduction. Removing the Jacobian loss changes folding by over an order of magnitude, identifying explicit topology control as the source of the near-zero folding regime.

Table 4: OASIS component ablation. Mean±standard deviation is over 109 test pairs from the seed-2026 checkpoints.

Variant	Dice↑	HD95↓	Fold.%↓
ReCo-SVF (full)	0.774 ± 0.021	1.92 ± 0.28	0.038 ± 0.014
w/o graph tokens	0.751 ± 0.025	2.22 ± 0.35	0.041 ± 0.016
w/o $\mathcal{L}_{\text{jac}}$	0.769 ± 0.023	1.97 ± 0.30	0.83 ± 0.27
w/o $\mathcal{L}_{\text{graph}}$	0.763 ± 0.023	2.06 ± 0.32	0.044 ± 0.016
w/o uncertainty	0.772 ± 0.022	1.94 ± 0.29	0.039 ± 0.015

A.2 AUXILIARY RESIDUAL REGULARIZER

Following heteroscedastic modelling (Kendall & Gal, 2017), a training-only 1^3 head h_{-1} maps the penultimate decoder refinement D_{-1} to a scalar residual field. Trilinear alignment to the final grid and gradient stopping define $u^* = \text{sg}[\text{up}(h_{-1}(D_{-1}))]$. Two separate 1^3 heads on the final refinement D_0 predict $\hat{u} = h_0(D_0)$ and an unconstrained scale $s = h_s(D_0)$. We parameterize $\sigma^2(\mathbf{x}) = 10^{-6} + \text{softplus}(s(\mathbf{x}))$, which gives a finite positive variance at every voxel. The training-only objective is

$$\mathcal{L}_{\text{unc}} = \frac{1}{|\Omega|} \sum_{\mathbf{x}} \left[\frac{(\hat{u} - u^*)^2}{2\sigma^2(\mathbf{x})} + \frac{1}{2} \log \sigma^2(\mathbf{x}) \right].$$

Its gradients update the final residual and variance heads and their shared decoder features, but not the detached target branch. The positive floor makes the objective bounded below. All three auxiliary heads are discarded at inference and do not alter the deformation.

A.3 CARRIER INTERVENTION CONTRACT

Table 5 makes the carrier substitution explicit. All six variants instantiate the encoder, proposal selector, dense decoder, SVF head, and every training loss. The proposal architecture, objective, budget, and initialization protocol are matched, while each independently trained model learns its own realized windows. In every row, $\mathcal{L}_{\text{graph}}$ uses the true-coordinate reference graph $\mathcal{E}_{\text{ref}}$, independently of the carrier message graph; no-token, dense, and edge-free variants therefore retain the same geometric regularizer without using its edges for message passing. Gradients pass through pooled descriptors, selected window coordinates, edge features, and the warped centres, but not through hard top-N, NMS, kNN, or random-edge sampling. The carrier parameters are trained from their own initialization in every row, and the encoder, selector, dense decoder, SVF head, loss weights, optimizer, schedule, augmentations, and checkpoint rule are otherwise identical.

Table 5: Per-variant carrier contract. "True attributes" are relative centre offset, physical distance, and log side-length ratio for the connected endpoints. Fixed random choices are keyed by training seed and volume identity.

Variant	Decoder memory	Message operator	Message attributes
No token	none	none	none
Dense	$1/4$ lattice	dense cross-attention	position encoding
Edge-free	32 regions	set self-attention	none
Random edges	32 regions	degree-matched GAT	true attributes
Shuffled 6-NN	32 regions	shuffled-coordinate GAT	shuffled attributes
Geometric 6-NN	32 regions	true-coordinate GAT	true attributes

A.4 Capacity and Modern Comparator Detail

Parameter and operation counts in Table 6 use one $160\times192\times224$ pair and count multiply–accumulates once. The matched variants differ by at most 0.7 million parameters; the graph variants share an identical budget.

Table 6: Model capacity for the controlled variants.

Variant	Parameters (M)	GMACs
No token branch	18.2	174
Matched dense carrier	18.9	248
Edge-free set mixer	18.6	183
Graph variants	18.7	186

Table 7 supplies topology and systems measurements for the modern OASIS block. Dice intervals use subject-level paired resampling; HD95 entries are pair-level means. Any-fold counts record cases containing at least one non-positive Jacobian voxel.

Table 7: Modern same-protocol OASIS detail. Brackets are subject-resampled 95% confidence intervals; retrained methods aggregate five seeds, while uniGradICON uses its released zero-shot weights.

Method	Dice [95% CI]	HD95	Fold.%	Any fold	Time / GB
H-SGANet	0.759 [0.754,0.764]	2.14	0.72	101/109	0.67 / 7.2
VFA	0.766 [0.761,0.771]	2.03	0.36	89/109	0.96 / 9.1
GeoReg	0.768 [0.763,0.773]	1.99	0.18	64/109	0.89 / 8.4
uniGradICON	0.750 [0.744,0.756]	2.17	0.024	29/109	1.42 / 10.7
ReCo-SVF	0.773 [0.769,0.777]	1.93	0.039	38/109	1.12 / 8.8

A.5 IXI Site Detail

Table 8 complements Table 3 with physical surface distance and topology. Similar effects across scanner groups support the overall transfer result.

A.6 Context-Retention Curve Points

Table 9 lists every point used in Figure 3. Shared crops transform both inputs and score ROIs remaining non-empty after co-cropping, so they provide a visible-anatomy stress trend rather than a fixed-population context estimand. Moving-only crops vary source evidence against a complete target and score the same target ROI set at every extent. Common-valid-mask and full-target rankings agree for the moving-only protocol; the main interaction uses full-target Dice.

Table 8: OASIS→IXI site-stratified detail; values are ReCo-SVF/ TransMorph pair-level mean±standard deviation.

Site	HD95 (mm)	Fold.%
Guy's	2.15 ± 0.31 / 2.51 ± 0.40	0.045 ± 0.018 / 1.74 ± 0.39
Hammersmith	2.20 ± 0.34 / 2.57 ± 0.43	0.047 ± 0.020 / 1.81 ± 0.42
IOP	2.19 ± 0.33 / 2.61 ± 0.44	0.046 ± 0.019 / 1.80 ± 0.41

Table 9: Dice by retained extent per axis. Shared-crop rows score non-empty co-cropped ROIs and report seed-2026 pair-level means. Moving-only rows score the complete fixed-target ROI set and report mean±standard deviation over five seed-level means and five offsets. The five extents span 6.4–100% of the original volume.

Protocol/method	100%	85%	70%	55%	40%
Shared/VoxelMorph	0.729	0.708	0.681	0.643	0.591
Shared/TransMorph	0.753	0.731	0.702	0.668	0.619
Shared/ReCo-SVF	0.774	0.761	0.743	0.721	0.694
Moving/TransMorph	0.752 ± 0.003	0.718 ± 0.006	0.667 ± 0.008	0.614 ± 0.010	0.563 ± 0.011
Moving/edge-free	0.765 ± 0.003	0.736 ± 0.005	0.696 ± 0.006	0.657 ± 0.008	0.620 ± 0.009
Moving/ReCo-SVF	0.773 ± 0.002	0.748 ± 0.004	0.717 ± 0.005	0.686 ± 0.007	0.655 ± 0.008

At 40%, common-valid-mask Dice is 0.701 ± 0.007 for ReCo-SVF, 0.674 ± 0.008 for edge-free tokens, and 0.616 ± 0.010 for TransMorph. The corresponding full-target values are lower because they score all target ROIs, but the geometric-token advantage increases from 2.7 to 3.5 points.

A.7 Proposal and Graph Sensitivity

Table 10 varies one structural choice at a time. Performance peaks broadly around 32 proposals and remains stable from six to twelve neighbours, indicating that the main setting is not a sharp isolated optimum.

Table 10: Three-seed proposal-count and graph-degree sensitivity. The seed-2026 $N = 32, k = 6$ values provide the reference operating point.

Setting	OASIS Dice	Shared 40% Dice
$N = 8,\ k = 6$	0.760	0.659
$N = 16,\ k = 6$	0.769	0.684
$N = 32,\ k = 6$	0.774	0.694
$N = 64,\ k = 6$	0.771	0.689
$N = 32,\ k = 3$	0.769	0.684
$N = 32,\ k = 12$	0.770	0.686

The proposal controls in Table 11 keep $N = 32, k = 6$ and the graph encoder fixed. Coverage is the fraction of foreground within the union of windows; repeatability is centre matching within 5 mm after intensity perturbation; diversity is one minus mean pairwise descriptor cosine similarity. Learned proposals combine high spatial coverage with data-dependent, non-redundant tokens.

A.8 Propagated-Structure Utility

The OASIS denominator is $109 \times 35 = 3{,}815$ ROI-pairs per seed. The BTCV denominator is $90 \times 13 = 1{,}170$ organ-pairs per seed, with intervals clustered over the 10 test subjects. Table 12 measures failure tails and propagated-volume stability.

Table 11: Proposal-selection controls on OASIS.

Selector	Dice	Shared 40	Coverage	Repeatability	Diversity
Uniform grid	0.756	0.647	0.79	1.00	0.63
Random	0.753	0.638	0.74	0.42	0.68
Farthest point	0.762	0.664	0.87	0.91	0.76
Learned selector	0.774	0.694	0.83	0.78	0.71

Table 12: Five-seed anatomical label-propagation utility for ReCo-SVF/ TransMorph.

Metric	OASIS	BTCV
Worst-decile structure Dice	0.612 / 0.556	0.473 / 0.421
Median absolute volume error	4.9% / 6.7%	8.6% / 11.4%
Mean surface distance (mm)	0.84 / 1.07	2.73 / 3.41
ReCo-SVF wins	2890/3815	812/1170

A.9 Topology Distribution

Table 13 reports topology beyond mean fold percentage for the matched OASIS comparison. The geometric variants remain in the same low-folding family, while TransMorph occupies a distinct high-folding regime. The similar SDlogJ and lower-tail determinants across sparse controls confirm that their Dice differences are not purchased by weaker regularity.

Table 13: Five-seed OASIS topology distribution. Fold volume is the median [interquartile range] percentage among all 109 cases.

Variant	Any fold	SDlogJ	$J_{1\%}$	Fold volume %
TransMorph	109/109	0.314	-0.128	1.54 [1.31,1.86]
No token branch	45/109	0.171	0.035	0.038 [0.014,0.064]
Matched dense carrier	43/109	0.168	0.036	0.036 [0.012,0.061]
Edge-free token mixer	42/109	0.166	0.037	0.035 [0.011,0.059]
Random edges	46/109	0.173	0.034	0.039 [0.014,0.066]
Shuffled 6-NN	45/109	0.172	0.034	0.038 [0.013,0.065]
ReCo-SVF	38/109	0.169	0.036	0.033 [0.010,0.057]

A.10 JACOBIAN OPERATING POINT

Table 14 varies only λ_j on OASIS. Increasing the margin weight traces a smooth accuracy–topology frontier; the selected $\lambda_j = 0.5$ lies near its knee.

Table 14: Three-seed Jacobian-loss sweep. SDlogJ uses positive determinants and $J_{1\%}$ is the first determinant percentile.

λ_j	Dice	Fold.%	SDlogJ	$J_{1\%}$
0	0.779	0.91	0.286	-0.041
0.1	0.778	0.31	0.231	0.006
0.25	0.776	0.11	0.194	0.021
0.5	0.774	0.038	0.169	0.036
1.0	0.769	0.014	0.151	0.049

A.11 STATISTICAL REPORTING AND EVIDENCE MAP

The OASIS/BTCV blocks of Table 1, Table 3's overall block, and Table 4 report seed-2026 checkpoints; their mean±standard-deviation values summarize test-pair dispersion. The carrier mechanism, retrained modern comparators, moving-only crop, sensitivity, and utility analyses aggregate repeated trainings as specified in their captions. Zero-shot uniGradICON is subject-resampled under its released weights. Bootstrap intervals resample subjects before within-subject observations, and BTCV resampling preserves the crossed moving/fixed-subject structure. Context failure counts threshold each subject–offset score after averaging across the five seeds, and utility wins compare five-seed mean structure scores.

The evidence chain follows the paper's interaction claim. Table 5 defines the intervention, Table 2 estimates the carrier effects, and the moving-only curve in Figure 3 measures how the geometric–edge-free effect changes over a fixed target anatomy. Its shared-crop panel supplies a separate visible-field stress endpoint. Table 1 and Figure 2 give the full-context accuracy–topology operating point; Table 3 gives acquisition-site transfer. Tables 10–12 connect operating-region stability to propagated anatomy.