Learn the Metric, Not the Deformation: A Small Contrastive Encoder for Cross-Modal Histology-MRI Registration
Supplementary Material

Jigneshkumar Mistry, Keerthi Ram, Supriti Mulay, Sivathanu Neelakantan, and Mohanasankar Sivaprakasam

Sudha Gopalakrishnan Brain Centre, Indian Institute of Technology Madras, India
jigneshkumar@htic.iitm.ac.in

S1 Data, ethics and availability

All imaging data used in this work are publicly released datasets: BigBrain histology and the matching donor MRI [2], the ICBM152 2009 templates, the AFID landmark annotations [4], the Allen MRI–histology pairs of ds003590 [3], and the AHEAD ex-vivo brains [1]. No new human data were acquired for this study and no identifiable participant information was accessed, so no additional ethics approval or consent procedure was required; the original consent and de-identification arrangements are those of the respective data providers.

We will release the encoder and decoder weights, the per-pair optimization implementation, the AFID protocol with per-fiducial breakdowns, and the scripts that reproduce every table and figure in the main paper.

S2 Encoder receptive-field sweep

Table 1. PatchNCE encoder receptive-field sweep, identical training recipe and per-pair evaluation. The 3-block default (RF $\approx$ 11) is the architectural sweet spot on Allen (n=30) and BigBrain (n=30) test pairs.

			Allen		BigBrain	
Variant	RF (px)	Params	NMI ↑	MIND ↓	NMI ↑	MIND ↓
small (2 conv blocks)	5	40k	1.264	0.097	1.223	0.151
ours (3 blocks)	**11**	**≈76k**	**1.274**	**0.091**	**1.242**	**0.139**
large (5 blocks)	33	150k	1.269	0.100	1.221	0.150
multi-scale (concat 3/9/33)	multi	155k	1.270	0.099	1.220	0.150

S3 Rotation envelope of the decoder warm-start

Freezing E and training only D to convergence does not beat the joint warm-start on Allen or BigBrain: what matters is the joint adaptation of metric and warm-start, which lands v_0 in the basin of the same cost surface E evaluates. Over the rotation envelope $\theta \in \{0, 10, 20, 30, 40\}^\circ$, D raises the recovery rate from $\sim$75% to $\geq$91%; without it, zero-velocity init fails on $\sim$25% of slices at $\theta \geq 20^\circ$.

Table 2. Rotation envelope of the decoder warm-start. The moving slice is rotated by θ, then registered back. Mean post-registration values against the unrotated reference; recovery is near-perfect up to $\pm 20^\circ$ and graceful at 40°. The warm-start does most of the work up to 20°.

	Allen (n=30)		BigBrain (n=30)	
θ (°)	NMI ↑	MIND ↓	NMI ↑	MIND ↓
0	**1.278**	**0.090**	**1.243**	**0.135**
20	1.224	0.113	1.179	0.166
40	1.153	0.136	1.150	0.174

S4 Jacobian determinants and velocity magnitudes

Reporting "0% folding" alone does not say how much headroom the integration has, and in a discrete implementation positive Jacobians are not unconditional. Table 3 therefore gives the measured distribution of $\det J$ and of the velocity field over all 28 BigBrain AFID slices at the 192×192 working resolution (1,032,192 pixels in total). Two quantities matter. The minimum $\det J$ over every pixel of every slice is 0.54, so the warps are not marginally fold-free but keep roughly a factor of two of headroom. And the per-step displacement of the squaring recursion, $\max\|v\|/2^{n_{\text{steps}}}$ with n_{steps}=7, peaks at 0.034 px: the scaling-and-squaring approximation to $\exp(v)$ is accurate only while this stays well below one pixel, and it is this quantity, not the bending energy, that keeps the integration well conditioned. A bending penalty does not by itself bound $\|v\|$, since a constant or affine velocity field has zero second difference at arbitrary magnitude; we therefore measure the velocity instead of assuming it is small.

S5 Seed variability

The per-pair optimization is stochastic only through initialization order and GPU non-determinism, but the headline TRE should not depend on a lucky run. Repeating the complete per-pair optimization over 50 random seeds and averaging per fiducial gives a set mean of 1.468 ± 0.016 mm on the 28-fiducial

Table 3. Jacobian determinant and velocity statistics, pooled over 28 BigBrain AFID slices on the 192×192 registration grid. Folding is the fraction of pixels with $\det J \leq 0$. Displacements are in pixels of that grid.

Quantity	Value
$\det J$ minimum (all pixels, all slices)	0.540
$\det J$ percentile 0.01 / 0.1 / 1	0.617 / 0.654 / 0.772
$\det J$ median	1.003
$\det J$ percentile 99 / maximum	1.257 / 2.749
folding ($\det J \leq 0$)	**0.000%**
$\|v\|$ mean / maximum	0.463 / 4.329 px
$\max\|v\|/2^7$ (per squaring step)	**0.034 px**
$\|\phi\|$ mean / maximum	0.459 / 4.042 px

subset (range 1.447–1.493) and 1.704 ± 0.012 mm on the 30-fiducial set (range 1.685–1.721). Seed-to-seed spread is therefore roughly two orders of magnitude smaller than the between-landmark spread (SD 0.90 mm and 1.27 mm respectively), i.e. the uncertainty in these numbers is dominated by which anatomical landmarks are used, not by the optimizer.

S6 Decoder warm-start: training objective

The encoder E and the velocity decoder D are trained jointly, not sequentially. Each step draws a co-aligned MRI–histology pair $(x_{\text{fix}}, x_{\text{mov}})$ and perturbs the moving image synthetically: $\tilde{x}_{\text{mov}}$ is x_{mov} under a random similarity pose composed with a random elastic warp whose amplitude is ramped from 2 to 12 px over the course of training, an easy-to-hard curriculum. With $v = D(E(x_{\text{fix}}), E(\tilde{x}_{\text{mov}}))$ and $\phi_v = \text{VecInt}(v)$, the training loss is

$$\begin{aligned}\mathcal{L}_{\text{train}} = & \underbrace{\big(1 - \cos\big(E(\phi_v \circ \tilde{x}_{\text{mov}}),\, E(x_{\text{fix}})\big)\big)}_{\text{warm-start recovers the synthetic warp}} \\ & + \lambda_{\text{NCE}}\, \mathcal{L}_{\text{PatchNCE}}\big(E(x_{\text{mov}}), E(x_{\text{fix}})\big) + \lambda_S\, \text{bend}(v),\end{aligned} \tag{1}$$

with $\lambda_{\text{NCE}} = 1$ and $\lambda_S = 0.5$. The PatchNCE term is computed on the *unperturbed* pair, so it keeps the metric anchored to real cross-modal correspondence while the first term teaches D to point in the right direction; the bending term keeps v_0 smooth.

Two consequences are worth stating plainly. First, D *is* trained against deformations, and the paper's claim is therefore about inference, not training: at test time both networks are frozen and D contributes only the initial condition v_0, with the deformation itself solved per pair. Second, because the supervision is synthetic warps of real pairs rather than ground-truth inter-subject correspondence, the warm-start is not fitted to the anatomical deformation statistics of

any particular cohort, which is consistent with the same weights transferring to AHEAD unchanged.

S7 Training and inference protocol

Table 4. Training and inference settings. All values are those used for the checkpoints reported in the main paper.

Setting	Value
Working resolution	192×192
Encoder width / embedding dim	64 / C=32, ℓ_2-normalized per pixel
Encoder / decoder parameters	76,576 / 147,842
PatchNCE temperature (τ)	$\tau = 0.07$
PatchNCE samples	256 per image
Joint training optimizer	AdamW, weight decay 10^{-5}, cosine schedule
Learning rates	encoder 1.5×10^{-4}, decoder 10^{-3}
Training steps	10,000
Loss weights	$\lambda_{\mathrm{NCE}} = 1.0$, $\lambda_S = 0.5$
Synthetic warp curriculum	random similarity pose + elastic, amplitude $2 \rightarrow 12$ px
Model selection	best validation NMI, network-only (no test-time optimization)
Inference optimizer	Adam, lr 0.1, velocity only
Coarse-to-fine schedule	downsample $\{8, 4, 2, 1\}$, iterations $\{120, 120, 60, 60\}$
VecInt steps / velocity scale	$n_{\mathrm{steps}} = 7$ / $\alpha = 0.05$
Inference loss weights	bending 10, containment 6, MIND 1
Runtime / hardware	$\approx$0.9 s per pair, one NVIDIA A100 (80 GB)
Peak GPU memory	122 MiB allocated (156 MiB reserved) per pair

S8 Per-dataset inference metrics

Table 5 gives the NMI and MIND distance of the frozen pipeline on all four evaluation sets. AHEAD R_1 and R_2^* are out-of-distribution: neither the brains nor the MRI contrasts appear in the encoder training set, and no weights or hyper-parameters were changed for them. Note that NMI and MIND are image-similarity surrogates; AHEAD has no fiducial protocol, so these rows are evidence of cross-contrast transfer of the metric, not of validated anatomical correspondence.

S9 Significance tests

Table 6 gives the tests summarized in the main paper. Comparisons are paired over slices against each baseline and against the unregistered pair, with Holm

Table 5. Model inference across four datasets. NMI higher is better; MIND distance lower is better. AHEAD R_1/R_2^* are out-of-distribution (not in the encoder training set); Allen and BigBrain are training-distribution.

Dataset	n	NMI ↑	MIND Distance ↓	Distribution
Allen	30	1.281 ± 0.072	0.050 ± 0.025	held-out
BigBrain	30	1.216 ± 0.018	0.127 ± 0.033	held-out
AHEAD R_1	32	1.247 ± 0.058	0.077 ± 0.032	OOD
AHEAD R_2^*	32	1.240 ± 0.053	0.078 ± 0.029	OOD

correction across the family of comparisons within each dataset. The unit of analysis is the slice; slices from a single brain are not independent observations, so these tests support within-subject rather than population-level conclusions.

Table 6. Paired Wilcoxon signed-rank tests (Holm-corrected) against learned baselines (TransMorph, SynthByReg) and the unregistered baseline, plus the Friedman omnibus test, on NMI and MIND.

Dataset	n (slices)	Wilcoxon p	Friedman χ^2
Allen	30	$< 10^{-7}$	> 79 ($p < 10^{-16}$)
BigBrain	30	$< 10^{-7}$	> 79 ($p < 10^{-16}$)
AHEAD R_1	32	$< 10^{-6}$	> 79 ($p < 10^{-16}$)
AHEAD R_2^*	32	$< 10^{-6}$	> 79 ($p < 10^{-16}$)

References

1. Alkemade, A., Mulder, M.J., et al.: The Amsterdam Ultra-high field adult lifespan database (AHEAD): A freely available multimodal 7 Tesla submillimeter magnetic resonance imaging database. NeuroImage **221**, 117200 (2020)
2. Amunts, K., Lepage, C., Borgeat, L., et al.: BigBrain: an ultrahigh-resolution 3D human brain model. Science **340**(6139), 1472–1475 (2013)
3. Casamitjana, A., Lorenzi, M., Ferraris, S., Peter, L., Modat, M., Stevens, A., Fischl, B., Vercauteren, T., Iglesias, J.E.: Robust joint registration of multiple stains and MRI for multimodal 3D histology reconstruction: Application to the Allen human brain atlas. Medical Image Analysis **75**, 102265 (2022). `https://doi.org/10.1016/j.media.2021.102265`
4. Lau, J.C., Xiao, Y., Haast, R.A., Gilmore, G., Uludag, K., MacDougall, K.W., Menon, R.S., Parrent, A.G., Peters, T.M., Khan, A.R.: A reproducible evaluation of ANTs similarity metric performance in brain image registration via anatomical fiducials (AFIDs). NeuroImage **185**, 349–360 (2019)