

# INST-Align: Implicit Neural Alignment for Spatial Transcriptomics via Canonical Expression Fields

Bonian Han, Cong Qi, Przemyslaw Musialski, and Zhi Wei

New Jersey Institute of Technology, Newark, NJ, USA  
{bh348, cq5, przemyslaw.musialski, zhi.wei}@njit.edu

**Abstract.** Spatial transcriptomics (ST) measures mRNA expression while preserving spatial organization, but multi-slice analysis faces two coupled difficulties: large non-rigid deformations across slices and inter-slice batch effects when alignment and integration are treated independently. We present INST-Align, an unsupervised pairwise framework that couples a coordinate-based deformation network with a shared Canonical Expression Field, an implicit neural representation mapping spatial coordinates to expression embeddings, for joint alignment and reconstruction. A two-phase training strategy first establishes a stable canonical embedding space and then jointly optimizes deformation and spatial-feature matching, enabling mutually constrained alignment and representation learning. Cross-slice parameter sharing of the canonical field regularizes ambiguous correspondences and absorbs batch variation. Across nine datasets, INST-Align achieves state-of-the-art mean OT Accuracy (0.702), NN Accuracy (0.719), and Chamfer distance, with Chamfer reductions of up to 94.9% on large-deformation sections relative to the strongest baseline. The framework also yields biologically meaningful spatial embeddings and coherent 3D tissue reconstruction. The code will be released after review phase.

**Keywords:** Spatial Transcriptomics · Implicit Neural Representation · Spatial Registration · Cross-slice Integration · Unsupervised Learning

## 1 Introduction

Spatial transcriptomics (ST) enables gene expression measurement with spatial context [1, 2], spanning imaging-based approaches such as multiplexed FISH and in situ sequencing [2, 3] and sequencing-based spatial barcoding platforms [1, 4]. These technologies have enabled molecular mapping across diverse biological systems [5–8], from the human dorsolateral prefrontal cortex [9] to whole mouse embryos [10], making multi-slice tissue analysis increasingly practical. However, reconstructing coherent 3D tissue volumes from serial sections remains challenging [11]: slices undergo complex non-rigid distortions during sample preparation (tearing, folding, compression), and arbitrary inter-slice rotations further compound the geometric ambiguity. Biological complexity makes ground-truth correspondences largely unattainable, rendering alignment inherently unsupervised.

Without a common coordinate frame, cross-slice expression integration is confined to individual 2D sections, leaving the collective three-dimensional context unexploited.

Existing alignment methods fall into two categories [11]. *Integration-based* methods such as STAligner [12], GraphST [13], SPACEL [14], and SANTO [15] learn joint spatial embeddings to correct coordinates, but their alignment quality is bounded by representation fidelity. *Geometric* methods align coordinates directly via optimal transport [16, 17], Gaussian processes [18], or diffeomorphic mappings [19, 20], but treat alignment and integration as independent steps. A common limitation is that these approaches decouple spatial alignment from expression integration: integration-based methods risk over-correcting and diminishing genuine biological diversity [11], while geometric methods lack expression-level feedback. Since raw expression inherently contains inter-slice batch effects, we argue that alignment and integration should be solved jointly.

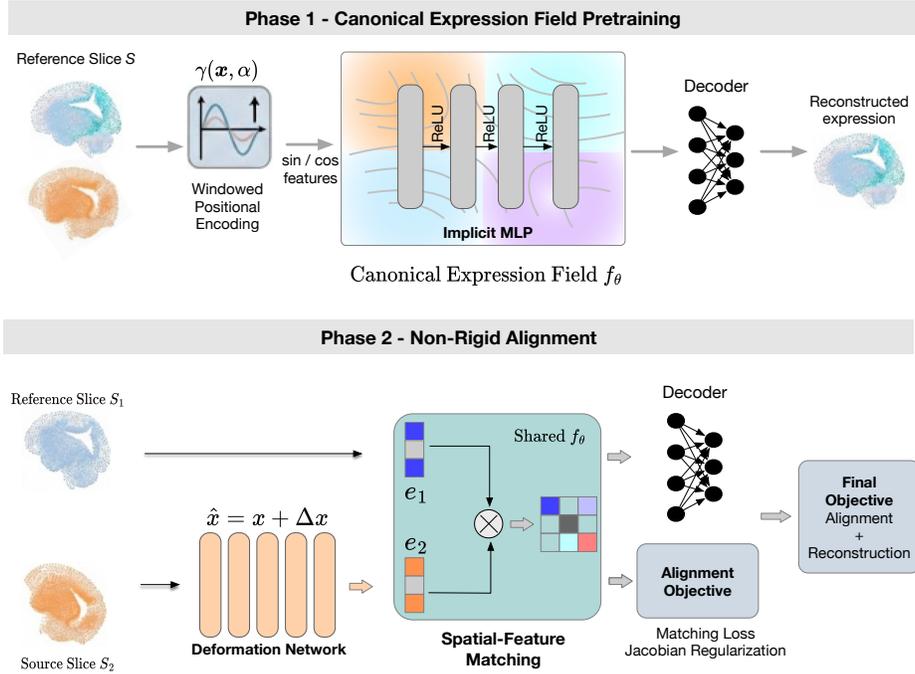
Implicit Neural Representations (INRs) model signals as continuous coordinate-conditioned functions  $f_\theta : \mathbf{x} \mapsto \mathbf{y}$ , enabling resolution-independent querying with compact parameterizations [21–23]. Although INRs are naturally suited to this setting—jointly learning deformation and reconstruction from coordinates, as demonstrated by Nerfies [30] in novel-view synthesis—and implicit deformation models have been explored in medical image registration [26, 27], existing INR-based ST methods such as STINR [24] and SUICA [25] only model within-slice expression fields and do not estimate deformation fields for multi-slice alignment.

In this paper, we propose INST-Align, a framework that couples spatial alignment and expression integration through a shared Canonical Expression Field parameterized as an INR. Requiring all slices to be explained by a *single continuous field* regularizes the ill-posed registration problem, turning alignment into a mutually constrained reconstruction task. Our key contributions are:

- **Joint alignment and integration via a shared canonical field.** A shared Canonical Expression Field coupled with a deformation network jointly solves spatial alignment and expression integration, achieving state-of-the-art mean alignment performance across nine datasets.
- **Spatial-expression alignment loss.** A unified cost combining geometric proximity and canonical-field feature similarity with adaptive soft assignment drives unsupervised non-rigid registration without ground-truth correspondences.
- **Biologically meaningful embeddings.** The canonical field produces spatial embeddings that support downstream analysis (e.g., clustering) without requiring dedicated graph-based embedding methods.

## 2 Method

Fig. 1 illustrates the overall pipeline of INST-Align. Given a pair of ST slices, we first perform rigid alignment via adaptive ICP, then jointly optimize a Canonical Expression Field and a Deformation Network through two-phase training. In Phase 1, we pretrain slice-specific expression fields with a shared decoder



**Fig. 1.** Overview of the INST-Align pipeline. Phase 1 learns a shared canonical expression field  $f_\theta$  as an implicit neural representation that maps spatial coordinates to embedding vectors and reconstructs gene expression profiles. Phase 2 jointly optimizes a non-rigid deformation network and feature-guided spatial matching within the shared field, enabling mutually reinforcing alignment and representation learning under smoothness regularization.

to establish a stable and biologically meaningful embedding space. In Phase 2, we switch to a single shared canonical field and jointly optimize deformation and alignment losses, enabling mutually constrained non-rigid registration and representation learning.

## 2.1 Problem Formulation

Let  $\mathcal{S}_1 = \{(\mathbf{x}_i^{(1)}, \mathbf{g}_i^{(1)})\}_{i=1}^{N_1}$  and  $\mathcal{S}_2 = \{(\mathbf{x}_j^{(2)}, \mathbf{g}_j^{(2)})\}_{j=1}^{N_2}$  denote two ST slices, where  $\mathbf{x} \in \mathbb{R}^2$  are spatial coordinates and  $\mathbf{g} \in \mathbb{R}^G$  are gene expression vectors. We designate  $\mathcal{S}_1$  as the reference slice and  $\mathcal{S}_2$  as the source slice.

Our goal is to learn: (1) a deformation function  $\phi_\psi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  that maps source coordinates into the reference frame, and (2) a canonical expression field  $f_{\theta_1} : \mathbb{R}^2 \rightarrow \mathbb{R}^d$  that maps any spatial location to a  $d$ -dimensional embedding capturing the biological identity at that position. The coordinate pairs are z-score normalized to a common scale, and a rigid pre-alignment is applied using adaptive ICP [28, 29] with expression-guided rotation selection.

## 2.2 Canonical Expression Field

The Canonical Expression Field is parameterized as an implicit neural representation (ExprINR)  $f_{\theta_1}$  defined on the reference frame. Given a 2D coordinate  $\mathbf{x}$ , we first apply a windowed positional encoding [30, 31]:

$$\gamma(\mathbf{x}, \alpha) = \left[ \mathbf{x}, w_k(\alpha) \sin(2\pi\sigma_k\mathbf{x}), w_k(\alpha) \cos(2\pi\sigma_k\mathbf{x}) \right]_{k=0}^{L-1} \quad (1)$$

where  $\sigma_k = 2^{k \cdot l_{\max}/(L-1)}$  are log-linearly spaced frequencies and  $w_k(\alpha)$  is the coarse-to-fine window of [30]. The window parameter  $\alpha \in [0, L]$  increases linearly during training, progressively enabling higher frequencies to prevent overfitting to noise in early epochs.

The encoded coordinates are processed by a 4-layer MLP with LayerNorm and ReLU activations to produce a  $d$ -dimensional bottleneck embedding ( $d = 64$ ):

$$\mathbf{e} = f_{\theta}(\mathbf{x}, \alpha) = \text{MLP}_{\theta}(\gamma(\mathbf{x}, \alpha)) \in \mathbb{R}^d \quad (2)$$

An expression decoder  $h_{\phi} : \mathbb{R}^d \rightarrow \mathbb{R}^{G'}$  reconstructs the highly variable gene (HVG) expression from the embedding:

$$\hat{\mathbf{g}} = h_{\phi}(\mathbf{e}) \quad (3)$$

A key design choice spans both training phases. In Phase 1, two independent ExprINRs ( $f_{\theta_1}$  for the reference,  $f_{\theta_2}$  for the source) share a single decoder, forcing their embedding spaces to align. In Phase 2, the source INR ( $f_{\theta_2}$ ) is discarded: since the Deformation Network maps source coordinates into the reference frame, a single canonical field  $f_{\theta_1}$  can encode both slices. This progressive design—from paired to canonical—ensures the canonical field is well-initialized before joint optimization begins, providing a stable biological prior that guides deformation.

## 2.3 Deformation Network

The Deformation Network  $\phi_{\psi}$  maps source coordinates to a displacement field. Given rigidly aligned source coordinates  $\mathbf{x}^{(2)}$ , it predicts:

$$\hat{\mathbf{x}}^{(2)} = \phi_{\psi}(\mathbf{x}^{(2)}, \alpha) = \mathbf{x}^{(2)} + \Delta\mathbf{x}(\mathbf{x}^{(2)}) \quad (4)$$

where  $\Delta\mathbf{x}$  is the learned displacement. The network consists of a 6-layer trunk MLP ( $d_{\text{hidden}} = 128$ ) followed by a lightweight spatial head ( $d_{\text{hidden}} = 64$ , 1 layer) that outputs the 2D displacement. The spatial head is initialized near zero so that  $\phi_{\psi}$  starts as an identity mapping, ensuring stable early-stage optimization.

To ensure spatially smooth and physically plausible deformations, we regularize the Jacobian of  $\phi_{\psi}$ . For a sampled subset of points, we compute the  $2 \times 2$  Jacobian matrix  $\mathbf{J} = \partial\hat{\mathbf{x}}/\partial\mathbf{x}$  via automatic differentiation and extract its singular values  $\sigma_1, \sigma_2$  via SVD. The regularization loss penalizes deviations from volume-preserving behavior with an **asymmetric** weighting:

$$\mathcal{L}_{\text{jac}} = \frac{1}{|\mathcal{B}|} \sum_{\mathbf{x} \in \mathcal{B}} \sum_{k=1}^2 w_k \cdot (\log \sigma_k)^2, \quad w_k = \begin{cases} 5.0 & \text{if } \log \sigma_k < 0 \text{ (compression)} \\ 1.0 & \text{otherwise (expansion)} \end{cases} \quad (5)$$

This asymmetry reflects the observation that compression (cells crowding together) is more pathological than moderate expansion in tissue alignment [32].

## 2.4 Joint Optimization

*Spatial-Expression Alignment Loss (ours).* Our core registration objective is an original *spatial-expression alignment loss* that establishes correspondences between deformed source points  $\hat{\mathbf{x}}_j^{(2)}$  and reference points  $\mathbf{x}_i^{(1)}$  using a unified geometry–feature cost derived from the canonical field. For each deformed point, we find its  $K$  nearest spatial neighbors in the reference and compute:

$$c_{ij} = \frac{\|\hat{\mathbf{x}}_j^{(2)} - \mathbf{x}_i^{(1)}\|^2}{s_{\text{sp}}} + \lambda_f \cdot \frac{1 - \langle \bar{\mathbf{e}}_j^{(2)}, \bar{\mathbf{e}}_i^{(1)} \rangle}{s_{\text{ft}}} \quad (6)$$

where  $\bar{\mathbf{e}} = \mathbf{e}/\|\mathbf{e}\|$  are L2-normalized ExprINR embeddings,  $s_{\text{sp}}$  and  $s_{\text{ft}}$  are running EMA scale factors, and  $\lambda_f$  balances the two terms. Soft assignment weights are obtained as  $p_{ij} = \text{softmax}(-c_{ij}/\tau)$ , where  $\tau$  is an adaptive temperature updated by EMA based on the mean weighted cost.

The embeddings entering the cost matrix are computed within a `torch.no_grad()` context, keeping the canonical embeddings as a frozen prior during correspondence estimation and preventing degenerate feature collapse.

The matching loss is bidirectional: the forward loss pulls each deformed source point toward its weighted target centroid, while the reverse loss ensures each reference point is explained by at least one source point, preventing many-to-one collapse.

*Reconstruction Loss.* Following SUICA [25], the expression decoder reconstructs HVG profiles from INR embeddings using a composite loss: masked MSE on nonzero entries, L1 on all entries, and a Dice loss on the zero/nonzero pattern:

$$\mathcal{L}_{\text{recon}} = \text{MSE}_{\text{masked}} + \text{L1} + 0.01 \cdot \mathcal{L}_{\text{dice}} \quad (7)$$

*Two-Phase Training. Phase 1* (300 epochs): Two independent ExprINRs ( $f_{\theta_1}$  for the reference,  $f_{\theta_2}$  for the source) share a single decoder and are jointly trained via  $\mathcal{L}_{\text{recon}}$  on both slices, aligning their embedding spaces. Concurrently, the DeformationNet is independently pre-trained with PCA-based matching. The coarse-to-fine  $\alpha$  window opens progressively over the first third of training. The best reconstruction checkpoint is restored before Phase 2.

*Phase 2* (400 epochs):  $f_{\theta_2}$  is discarded; all remaining components ( $f_{\theta_1}$ ,  $h_\phi$ ,  $\phi_\psi$ , and the matcher) are jointly optimized with the total loss:

$$\mathcal{L} = \mathcal{L}_{\text{match}} + \lambda_r \mathcal{L}_{\text{recon}} + \lambda_j \mathcal{L}_{\text{jac}} \quad (8)$$

where  $\lambda_r = 0.1$  (reduced to prevent reconstruction from dominating alignment) and  $\lambda_j$  controls Jacobian regularization. A full-coverage reverse matching step is performed each epoch to address global collapse missed by mini-batch training. Learning rate is scheduled via ReduceLROnPlateau.

**Table 1.** D1–D3: DLPFC samples; MB2/MB7/MB11: MERFISH Brain sections; ST: STARMap; MH: MERFISH Hypothalamus; ME: MouseEmbryo. Spatial alignment results across nine datasets. Best in **bold**, second-best underlined. “–” indicates method not available or no alignment.

Metric	Method	DLPFC			MERFISH Brain			ST	MH	ME	Mean
		D1	D2	D3	MB2	MB7	MB11				
OT Acc $\uparrow$	Raw	.745	.887	.673	.013	.050	.209	.860	.051	.116	.400
	PASTE	.859	<u>.904</u>	.806	.267	.726	<u>.349</u>	<b>.882</b>	.674	.031	.611
	STalign	<b>.866</b>	<b>.905</b>	.808	.022	.046	.143	<u>.861</u>	.051	.112	.424
	Spateo	<u>.864</u>	.884	.818	<b>.343</b>	<b>.750</b>	.298	.860	.681	<u>.723</u>	<u>.691</u>
	ICP	.829	.888	<u>.829</u>	.251	.709	.347	.852	<u>.689</u>	.710	.678
	<b>Ours</b>	.863	.899	<b>.839</b>	<u>.318</u>	<u>.746</u>	<b>.357</b>	.857	<b>.690</b>	<b>.746</b>	<b>.702</b>
NN Acc $\uparrow$	Raw	.765	.913	.729	.130	.064	.276	<u>.890</u>	.167	.172	.456
	PASTE	<u>.859</u>	.904	.804	.244	.731	<u>.360</u>	.883	.680	.032	.611
	STalign	<b>.875</b>	.894	.788	.136	.065	.181	.872	.167	.166	.460
	Spateo	.845	<b>.925</b>	.823	<b>.378</b>	<b>.784</b>	.215	.889	<b>.745</b>	<u>.759</u>	<u>.707</u>
	ICP	.851	<u>.925</u>	<u>.831</u>	.228	.723	.280	.889	<u>.701</u>	.733	.685
	<b>Ours</b>	.853	.920	<b>.849</b>	<u>.322</u>	<u>.780</u>	<b>.365</b>	<b>.899</b>	.693	<b>.788</b>	<b>.719</b>
Chamfer $\downarrow$	Raw	<u>0.1</u>	<b>0.1</b>	0.5	–	–	–	136.5	–	–	–
	PASTE	0.3	0.3	<b>0.2</b>	173.1	<b>5.8</b>	<u>51.7</u>	<b>15.3</b>	<b>4.4</b>	235.5	54.1
	STalign	0.6	0.6	0.7	–	–	–	140.6	–	–	–
	Spateo	0.6	0.7	0.7	166.2	16.8	103.3	142.7	12.2	<u>12.0</u>	50.6
	ICP	<b>0.0</b>	<u>0.1</u>	<u>0.4</u>	<u>77.9</u>	18.7	107.0	135.7	<u>11.0</u>	15.3	<u>40.7</u>
	<b>Ours</b>	0.7	0.6	0.6	<b>12.6</b>	<u>13.9</u>	<b>17.1</b>	<u>135.0</u>	<u>11.0</u>	<b>11.9</b>	<b>22.6</b>

### 3 Experiments

#### 3.1 Datasets and Baselines

We evaluate nine ST datasets spanning grid-based and continuous-coordinate technologies: DLPFC [9] (3 samples, the first adjacent section pair per sample,  $\sim 3.4$ – $4.8$ k spots), STARMap [3] ( $\sim 1$ k spots), MERFISH Brain [34] (3 samples,  $\sim 5.5$ – $24.8$ k cells) and MERFISH Hypothalamus [35] (1 sample,  $\sim 17$ – $20$ k cells), and MouseEmbryo [10] ( $\sim 17$ – $20$ k cells). Alignment baselines are PASTE [16], STalign [19], Spateo [20], plus Raw (unaligned) and ICP (rigid-only); integration-focused methods are evaluated separately in Table 2 to avoid task mismatch. We report OT Accuracy [16], NN Accuracy, and Chamfer distance [33, 11] for alignment, and ARI/NMI (mclust) for embedding quality.

**Table 2.** Embedding quality (ARI / NMI) via mclust clustering. Each is from the target slice. Baselines include STAligner [12], GraphST [13], SPIRAL [36], and Seurat [37].

Method	DLPFC Samp. 1		DLPFC Samp. 2		DLPFC Samp. 3		MouseEmbryo		Mean	
	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI
PCA	.298	.385	.205	.364	.208	.340	.128	.215	.210	.326
Seurat	.237	.372	.285	.287	.253	.397	.078	.220	.213	.319
STAligner	.273	.474	<u>.463</u>	<u>.556</u>	.293	<u>.491</u>	<b>.356</b>	<b>.552</b>	.346	<u>.518</u>
GraphST	<u>.431</u>	<u>.641</u>	<b>.490</b>	<b>.589</b>	<b>.618</b>	<b>.724</b>	.266	<u>.516</u>	<b>.451</b>	<b>.618</b>
SPIRAL	<b>.558</b>	<b>.652</b>	.359	.484	<u>.356</u>	.478	.192	.355	<u>.366</u>	.492
INST-Align (Ours)	.397	.490	<u>.463</u>	.446	.246	.393	<u>.341</u>	.512	.362	.460

### 3.2 Spatial Alignment Results

Table 1 summarizes alignment across nine datasets. INST-Align achieves the best mean OT Accuracy (0.702), NN Accuracy (0.719), and Chamfer distance (22.6), compared with Spateo (0.691/0.707/50.6). On near-aligned DLPFC (D1–D3), all methods are competitive after rigid initialization; our slightly higher Chamfer (0.6–0.7 vs. PASTE/ICP) reflects limited room for non-rigid improvement on already well-aligned grid data.

On large-deformation datasets the gap widens: Chamfer drops from 173.1 to 12.6 on MB2 (vs. PASTE), from 51.7 to 17.1 on MB11, and from 235.5 to 11.9 on MouseEmbryo, while OT/NN scores remain comparable. STalign produces excessively large deformations on MERFISH Brain and MouseEmbryo, making its metrics non-comparable (marked “-”). Overall, the advantage of INST-Align grows with dataset size and deformation severity, suggesting that the canonical field provides stronger regularization as geometric ambiguity increases.

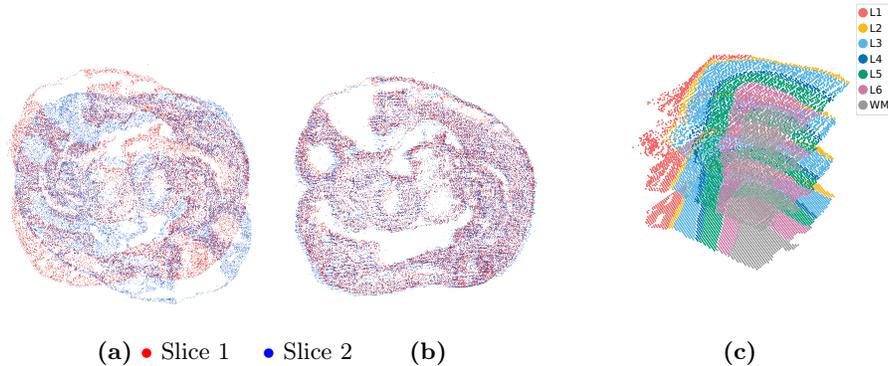
### 3.3 Downstream Analysis

Table 2 reports embeddings from the canonical expression field. Because INST-Align is optimized for alignment, embedding quality is a byproduct rather than the primary target, yet results remain competitive (e.g., ARI 0.463 on DLPFC Sample 2, matching STAligner; ARI/NMI 0.341/0.512 on MouseEmbryo). GraphST achieves the highest mean ARI/NMI with dedicated spatial graph networks. That a coordinate-conditioned INR designed for alignment produces biologically meaningful embeddings without graph construction supports the view that a well-regularized canonical field captures tissue structure as an emergent property.

Fig. 2 shows that INST-Align resolves large non-rigid mismatches on MouseEmbryo (a→b) and produces coherent layer structure in 3D reconstruction (c).

### 3.4 Ablation Studies

Table 3 shows that pretraining is critical for embedding quality (ARI drops from 0.334 to 0.045 without it), joint optimization improves geometric alignment (Chamfer 0.36 vs. 0.56), and Jacobian regularization benefits both.



**Fig. 2.** Qualitative results. (a) MouseEmbryo after rigid ICP pre-alignment; (b) INST-Align non-rigid alignment on the same pair; (c) 3D reconstruction of DLPPFC Sample 3 from consecutively aligned slices, colored by cortical layer.

**Table 3.** Ablation study on DLPPFC Sample 1. Each row removes one component from the full model.

Variant	Alignment		Embedding	
	OT Acc $\uparrow$	Chamfer $\downarrow$	ARI $\uparrow$	NMI $\uparrow$
Full (Ours)	.810	<b>0.36</b>	<b>.334</b>	<b>.450</b>
w/o Phase 1 (pretrain)	<b>.813</b>	0.47	.045	.079
w/o Phase 2 (joint opt.)	.807	0.56	.295	.434
w/o Jacobian reg.	.805	0.49	.287	.400

## 4 Conclusion

We presented INST-Align, which couples a deformation network with a shared Canonical Expression Field for pairwise ST alignment and reconstruction. The shared field provides biological regularization for ambiguous correspondences under sparsity, batch effects, and large deformations. Across nine datasets, INST-Align achieves state-of-the-art mean OT Accuracy, NN Accuracy, and Chamfer distance, with consistent performance across both near-aligned grid data and large-deformation sections, yielding Chamfer reductions of up to 94.9% relative to PASTE. Ablations confirm contributions from Phase 1 pretraining, Phase 2 joint optimization, and Jacobian regularization. The current framework operates in a pairwise setting and lacks a global multi-slice consistency objective. In addition, biological validation is limited to clustering-based metrics (ARI/NMI), and the learned embeddings remain below specialized embedding methods. Future work will extend the model toward globally consistent multi-slice 3D alignment with enhanced biological evaluation.

## References

1. Ståhl, P.L., et al.: Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**(6294), 78–82 (2016)
2. Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., Zhuang, X.: Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**(6233), aaa6090 (2015)
3. Wang, X., et al.: Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**(6400), eaat5691 (2018)
4. Stickels, R.R., et al.: Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nature Biotechnology* **39**(3), 313–319 (2021)
5. Zhang, B., et al.: A human embryonic limb cell atlas resolved in space and time. *Nature* **635**(8039), 668–678 (2024)
6. Allen, W.E., et al.: Molecular and spatial signatures of mouse brain aging at single-cell resolution. *Cell* **186**(1), 194–208 (2023)
7. Khaliq, A.M., et al.: Spatial transcriptomic analysis of primary and metastatic pancreatic cancers highlights tumor microenvironmental heterogeneity. *Nature Genetics* **56**(11), 2455–2465 (2024)
8. Longo, S.K., Guo, M.G., Ji, A.L., Khavari, P.A.: Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nature Reviews Genetics* **22**(10), 627–644 (2021)
9. Maynard, K.R., et al.: Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature Neuroscience* **24**(3), 425–436 (2021)
10. Chen, A., et al.: Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**(10), 1777–1792 (2022)
11. Dong, K., et al.: Benchmarking multi-slice integration and downstream applications in spatial transcriptomics data analysis. *Genome Biology* **26**, 318 (2025)
12. Zhou, X., Dong, K., Zhang, S.: Integrating spatial transcriptomics data across different conditions, technologies and developmental stages. *Nature Computational Science* **3**(10), 894–906 (2023)
13. Long, Y., et al.: Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nature Communications* **14**(1), 1155 (2023)
14. Xu, H., et al.: SPACEL: deep learning-based characterization of spatial transcriptome architectures. *Nature Communications* **14**(1), 7603 (2023)
15. Li, H., et al.: SANTO: a coarse-to-fine alignment and stitching method for spatial omics. *Nature Communications* **15**, 6048 (2024)
16. Zeira, R., Land, M., Strzalkowski, A., Raphael, B.J.: Alignment and integration of spatial transcriptomics data. *Nature Methods* **19**(5), 567–575 (2022)
17. Liu, X., Zeira, R., Raphael, B.J.: Partial alignment of multislice spatially resolved transcriptomics data. *Genome Research* **33**(7), 1124–1132 (2023)
18. Jones, A., Townes, F.W., Li, D., Engelhardt, B.E.: Alignment of spatial genomics data using deep Gaussian processes. *Nature Methods* **20**(9), 1379–1387 (2023)
19. Clifton, K., et al.: STalign: Alignment of spatial transcriptomics data using diffeomorphic metric mapping. *Nature Communications* **14**(1), 8123 (2023)
20. Qiu, X., et al.: Spatiotemporal modeling of molecular holograms. *Cell* **187**(26), 7351–7373 (2024)
21. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: Representing scenes as neural radiance fields for view synthesis. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) *ECCV 2020*. LNCS, vol. 12346, pp. 405–421. Springer, Cham (2020)

22. Sitzmann, V., Martel, J.N.P., Bergman, A.W., Lindell, D.B., Wetzstein, G.: Implicit neural representations with periodic activation functions. In: NeurIPS, pp. 7462–7473 (2020)
23. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.* **41**(4), 102:1–102:15 (2022)
24. Luo, Y., Zhao, X., Ye, K., Meng, D.: STINR: Deciphering spatial transcriptomics via implicit neural representation. In: CVPR, pp. 25930–25939 (2025)
25. Zhu, Q., Zheng, Y., Sang, Y., Zhan, Y., Zhu, Z., Ding, J., Zheng, Y.: SUICA: Learning super-high dimensional sparse implicit neural representations for spatial transcriptomics. In: ICML. PMLR, vol. 267, pp. 80448–80462 (2025)
26. Wolterink, J.M., Zwienenberg, J.C., Brune, C.: Implicit neural representations for deformable image registration. In: MIDL. PMLR, vol. 172, pp. 1349–1359 (2022)
27. Jena, R., Chaudhari, P., Gee, J.C.: Deep implicit optimization enables robust learnable features for deformable image registration. *Medical Image Analysis* **103**, 103577 (2025)
28. Besl, P.J., McKay, N.D.: Method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(2), 239–256 (1992)
29. Musialski, P.: Point cloud to model registration. Technical Report TR-VRVIS-009-2009, VRVis Research Center, Vienna (2009)
30. Park, K., et al.: Nerfies: Deformable neural radiance fields. In: ICCV, pp. 5845–5854 (2021)
31. Tancik, M., et al.: Fourier features let networks learn high frequency functions in low dimensional domains. In: NeurIPS, pp. 7537–7547 (2020)
32. Chun, S.Y., Fessler, J.A.: A simple regularizer for B-spline nonrigid image registration that encourages local invertibility. *IEEE Journal of Selected Topics in Signal Processing* **3**(1), 159–169 (2009)
33. Fan, H., Su, H., Guibas, L.J.: A point set generation network for 3D object reconstruction from a single image. In: CVPR, pp. 2463–2471 (2017)
34. Zhang, M., et al.: Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* **598**(7879), 137–143 (2021)
35. Moffitt, J.R., et al.: Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**(6416), eaau5324 (2018)
36. Guo, T., et al.: SPIRAL: integrating and aligning spatially resolved transcriptomics data across different experiments, conditions, and technologies. *Genome Biology* **24**, 241 (2023)
37. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., Satija, R.: Comprehensive integration of single-cell data. *Cell* **177**(7), 1888–1902 (2019)