STTT: TEST-TIME TRAINING FOR CROSS-DOMAIN SPATIAL TRANSCRIPTOMICS PREDICTION FROM HISTOLOGY

Anonymous authors
Paper under double-blind review

ABSTRACT

High costs and limited availability of spatial transcriptomics data motivate predicting spatial gene expression directly from routine H&E histology. However, variations in staining, scanners, cohorts, platforms, and organs create substantial domain shifts. Most histology-to-spatial-transcriptomics predictors rely on static inference, leaving unlabeled target slides unable to guide adaptation at prediction time. We propose Spatial Test-Time Training (STTT), a framework for spatially structured adaptation at inference time. STTT combines masked-self reconstruction and bidirectional center–context reconstruction to adapt using unlabeled target-slide morphology and spatial neighborhoods. Low-rank query/value updates, representation damping, and slide-wise reset constrain this label-free adaptation. Across four cross-domain settings, STTT consistently improves mean PCC over source-trained SpaFactor and achieves the highest mean four-task PCC among seven compared TTT and TTA baselines.

1 INTRODUCTION

Spatial transcriptomics (ST) preserves spatial information within tissues while measuring gene expression, thereby establishing a direct link between molecular profiles and tissue organization (Ståhl et al., 2016). Despite its immense value in biomedical research, ST experiments remain costly and require specialized workflows, which limits the availability of samples with matched histology and gene expression data (Wang et al., 2025). In contrast, routine hematoxylin and eosin (H&E) stained slides are far more accessible. This gap in data availability has driven the development of histology-to-spatial-transcriptomics (HE2ST) methods, which predict spatial gene expression directly from tissue morphology using paired H&E and ST data (He et al., 2020; Pang et al., 2021). However, because paired data remain scarce, collecting new labeled samples and retraining a predictive model for every target cohort or domain is often impractical.

Substantial domain variation further complicates the deployment of HE2ST. The visual appearance of H&E slides can change depending on staining conditions, scanners, and image acquisition protocols. Meanwhile, ST measurements and the underlying relationships between morphology and gene expression can vary across cohorts, datasets, sequencing platforms, and organs. These data shifts may affect image representations and the associations between morphology and expression learned from the source data (Tellez et al., 2019; Wang et al., 2025). Consequently, a target slide may differ from the source domain not only in its visual appearance but also in its molecular distribution and spatial organization.

Previous HE2ST studies have explored a wide array of model designs, including regression models, contrastive learning, Transformer and graph architectures, multiresolution aggregation, and generative modeling (He et al., 2020; Pang et al., 2021; Zeng et al., 2022; Xie et al., 2023; Jia et al., 2024; Zhang et al., 2024; Chung et al., 2024; Zhu et al., 2025; Wang et al., 2025). However, most existing methods adhere to a static deployment paradigm: a predictive model is trained on source data and subsequently applied to target slides without updating its parameters. While such models may still generalize to unseen data, their deployment performance relies entirely on the knowledge acquired from the source domain. They cannot explicitly leverage the morphology and spatial structure of the unlabeled target slide to adapt to variations specific to the slide prior to making predictions.

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.

Test-time training (TTT) offers a natural mechanism to introduce this adaptability. TTT updates a model trained on source data using an auxiliary self-supervised objective on unlabeled test inputs before making predictions for the primary task (Sun et al., 2020). In the HE2ST task, a target slide provides both histological features and spatial coordinates, making it feasible to construct adaptation signals without relying on target gene expression. However, generic TTT and test-time adaptation (TTA) objectives do not explicitly exploit the spatial neighborhood structure of ST spots. Spots within a tissue section are not isolated samples; their neighborhoods contain information about local tissue organization. Treating spots as isolated feature vectors may therefore leave spatial relationships underutilized. This motivates a test-time objective for HE2ST that integrates center-spot morphology and spatial neighborhood structure.

To this end, STTT extends HE2ST from static inference to adaptive test-time inference. It combines masked self-reconstruction of center-spot morphology with bidirectional reconstruction between the center and its context. The masked self-reconstruction objective provides a slide-specific signal from each center spot's morphology, while the center-from-context objective uses neighboring spots to reconstruct the center representation. The reverse context-from-center objective promotes consistency between the center spot and its surrounding context. Together, these objectives use both individual spot morphology and spatial relationships within the target slide. During source training, a differentiable inner update encourages the predictor to benefit from this self-supervised signal; at target time, adaptation uses only the unlabeled slide's image features and coordinates.

To constrain target-time adaptation, STTT applies low-rank adaptation of the query and value projections (Q/V LoRA). Representation damping limits deviations from the source representations, while slide-wise reset prevents adaptation state from accumulating across target slides. We instantiate STTT on SpaFactor as the base HE2ST predictor; during target-time adaptation, SpaFactor remains frozen and only the Q/V LoRA fast state is updated.

Our contributions are threefold:

- We introduce STTT, an HE2ST-specific test-time training framework that extends static deployment toward adaptive inference using unlabeled target slides.
- We develop a spatial, label-free objective combining masked-self and bidirectional center–context reconstruction, paired with constrained Q/V LoRA adaptation and representation damping.
- Using SpaFactor as the base HE2ST predictor, we evaluate STTT across four cross-domain settings and against seven TTT/TTA methods. STTT consistently improves mean PCC across the four tasks and achieves the highest mean PCC among the compared methods; ablation and sensitivity analyses further characterize its adaptation behavior.

2 Related Work

2.1 Histology-to-Spatial-Transcriptomics Prediction

HE2ST research has developed through several complementary model families. Direct regression and convolutional methods, including ST-Net and DeepSpaCE, learn mappings from histology to expression. Transformer, graph, and hybrid predictors such as HisToGene, Hist2ST, and THItoGene model image features, spots, and their spatial relations in different ways (He et al., 2020; Monjo et al., 2022; Pang et al., 2021; Zeng et al., 2022; Jia et al., 2024). More recent methods incorporate spatial context (DeepSpot), image–expression contrastive learning (BLEEP), and multiresolution fusion (TRIPLEX) (Nonchev et al., 2025; Xie et al., 2023; Chung et al., 2024). Other directions include super-resolution expression prediction, generative modeling, single-cell resolution, spot-level prediction, and context-conditioned generation (Zhang et al., 2024; Zhu et al., 2025; Huang et al., 2025; Fu et al., 2025; Yin et al., 2026; Fang et al., 2026; Xiang et al., 2026; Liu et al., 2026).

Spatial context and gene program modeling provide another relevant direction. Gene program models represent coordinated expression through lower dimensional latent programs. SpaFactor is a lightweight predictor that combines this representation with spatial context (Ruan et al., 2026); we use it as the base HE2ST predictor for STTT.

Despite their differences in architecture and representation learning, most HE2ST predictors follow a static source training and target inference workflow: after training on source data, the predictor

is applied to a target slide without updating on the unlabeled target distribution. This leaves target morphology and spatial structure unused as adaptation signals. STTT studies how those observations can update the predictor before expression is estimated.

2.2 Test-Time Training and Test-Time Adaptation

Test-time adaptation (TTA) updates a model or its representations using unlabeled target inputs. Existing methods use different signals: Tent minimizes prediction entropy, CoTTA addresses continual adaptation, and SAR uses sharpness aware updates (Wang et al., 2021; 2022; Niu et al., 2023). BUFR restores features under measurement shift, ActMAD matches activations, while SSA, RSD, and PEA use subspace or embedding criteria (Eastwood et al., 2022; Mirza et al., 2023; Adachi et al., 2025; Chen et al., 2021; Ma et al., 2026). These methods address different settings and objectives; their terminology and protocols are not fully interchangeable.

Test-time training (TTT) places particular emphasis on an auxiliary self-supervised task during source training, which is reused to update the model on current unlabeled test data. The original TTT formulation and masked autoencoding TTT illustrate this source to test connection (Sun et al., 2020; Gandelsman et al., 2022). STTT follows this design: its differentiable source inner update prepares the HE2ST predictor to benefit from the same adaptation objective when processing a target slide.

Existing TTT and TTA objectives were generally developed for classification, generic feature adaptation, or regression rather than the spatial structure of HE2ST prediction. HE2ST predicts continuous gene expression, and spots on a tissue section are linked by local morphology instead of being independent samples. At adaptation time, target expression is unavailable, while histology features, coordinates, and spatial neighborhoods remain observable. These properties motivate a label-free objective that combines center-spot morphology with bidirectional center–context reconstruction.

3 Method

3.1 Problem Formulation

Let $\mathcal{S}$ be the labeled source slides. For spot i on slide $s \in \mathcal{S}$, $x_i^{(s)} \in \mathbb{R}^{1536}$ is its histology feature, $r_i^{(s)} \in \mathbb{R}^2$ its coordinate, and $y_i^{(s)} \in \mathbb{R}^G$ its expression target. A target slide supplies $X_t = \{x_i^{(t)}\}_{i=1}^{n_t}$ and $R_t = \{r_i^{(t)}\}_{i=1}^{n_t}$ for predictions $\hat{y}_i^{(t)} \in \mathbb{R}^G$; the evaluator accesses target labels only after predictions are finalized.

Each target slide is one transductive adaptation unit: its spots jointly provide features and coordinates for spatial neighborhoods and a label-free update. Adaptation uses no target expression and does not process spots independently online.

3.2 Frozen Histology Representation and Base HE2ST Predictor

STTT uses precomputed 1536-dimensional GigaPath features x_i from H&E images (Xu et al., 2024); the pathology encoder stays frozen. Coordinates define local ($k = 4$) and regional ($k = 16$) neighborhoods. A shared ST2 Transformer with two layers encodes each scale, producing $h_{i,\mathrm{loc}}$ and $h_{i,\mathrm{reg}}$.

The SpaFactor gene-program predictor (Ruan et al., 2026) takes the 4608-dimensional concatenation of the original feature and two spatial representations:

$$\hat{y}_i = \mathrm{SpaFactor}([x_i, h_{i,\mathrm{loc}}, h_{i,\mathrm{reg}}]) \in \mathbb{R}^G. \tag{1}$$

3.3 Spatial Self-Supervised Adaptation Objective

For each center, the mean local and regional ST2 output is its relation representation. The shared key projection W_k and fixed value projection W_v map to a 256-dimensional relation space. Source objective-head pretraining fits W_k with the relation heads; it is then frozen for expression meta-training and target adaptation, while W_v stays fixed throughout. For relation vectors a, b, the reconstruction distance is $d_{\mathrm{rel}}(a, b) = 1 - \cos(a, b) + 0.1\,\mathrm{MSE}(a, b)$, with MSE averaged over

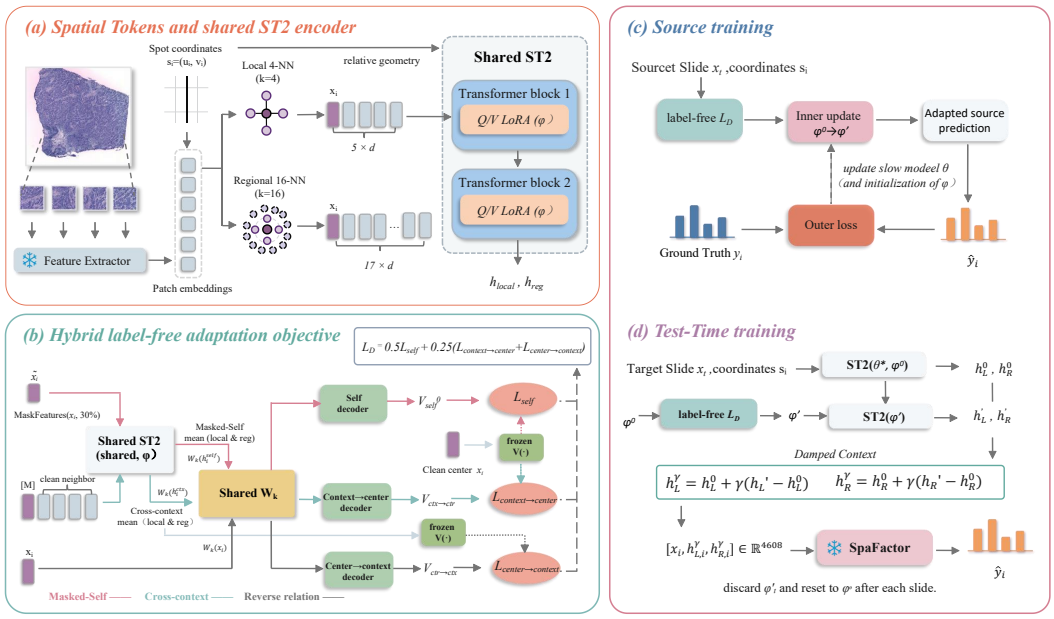

Figure 1: STTT architecture and training workflow. (a) Histology features and spot coordinates form local ($k = 4$) and regional ($k = 16$) tokens for a shared ST2 encoder with Q/V LoRA. (b) The label-free objective combines masked self-reconstruction with bidirectional center–context reconstruction. (c) Differentiable source training updates the model initialization through an inner step. (d) Target-time training adapts the Q/V fast state, applies residual damping, predicts with SpaFactor, and resets the fast state after each slide.

dimensions. Masked self-reconstruction and two directional center–context terms use only histology features and coordinates.

3.3.1 Masked Self-Reconstruction

A random mask with ratio 0.30 replaces selected entries of the clean feature x_i with a learned token, producing $\tilde{x}_i$. With no valid neighbors at either scale, the mean local and regional output h_i^{self} depends only on the corrupted center. The key projection and self-reconstruction head ψ_{self} predict the fixed value projection of clean x_i:

$$\mathcal{L}_{\text{self}} = d_{\text{rel}}\big(\psi_{\text{self}}(W_k(h_i^{\text{self}})),\ \text{sg}(W_v(x_i))\big)\,. \tag{2}$$

Here sg stops gradients. The reconstruction target is a stop-gradient projection of the clean histology embedding, rather than gene expression.

3.3.2 Bidirectional Center–Context Reconstruction

For the spatial relation path, the center token is masked but local and regional neighbors remain. Shared ST2 self-attention yields the neighbor-conditioned h_i^{ctx}, the mean of both center outputs, without a separate cross-attention module. The shared key projection and two heads define

$$\begin{aligned}\mathcal{L}_{\text{context}\to\text{center}} &= d_{\text{rel}}\big(\psi_{\text{context}\to\text{center}}(W_k(h_i^{\text{ctx}})),\ \text{sg}(W_v(x_i))\big)\,,\\ \mathcal{L}_{\text{center}\to\text{context}} &= d_{\text{rel}}\big(\psi_{\text{center}\to\text{context}}(W_k(x_i)),\ W_v(h_i^{\text{ctx}})\big)\,.\end{aligned} \tag{3}$$

The first term predicts the clean center value from context; the reverse predicts context value from the center key. Although W_v is fixed, its context input still depends on ST2.

3.3.3 Hybrid Objective

Averaging over centers, the self term has weight 0.5 and each spatial direction 0.25:

$$\mathcal{L}_D = 0.5\,\mathcal{L}_{\text{self}} + 0.25\,\mathcal{L}_{\text{context}\to\text{center}} + 0.25\,\mathcal{L}_{\text{center}\to\text{context}}\,. \tag{4}$$

3.4 Learning to Adapt During Source Training

STTT first fits W_k, the relation heads, and mask token on source features and coordinates without expression targets, with the source-trained ST2 predictor and W_v fixed. All objective parameters are then frozen for expression meta-training and target adaptation.

Expression meta-training initializes ST2 and SpaFactor from a source-supervised checkpoint and the fast state from learnable Q/V LoRA factors. On each labeled source slide, $\mathcal{L}_D$ from that slide drives one differentiable inner update; the resulting representations predict its expression. The outer source-expression loss differentiates through the update to train ST2 base parameters, SpaFactor, and initial LoRA factors. The objective heads, W_k, and W_v remain fixed.

3.5 Constrained Target-Time Adaptation

For both ST2 Transformer blocks, STTT represents the query and value weights with functional low-rank updates

$$W'_{\ell,p} = W_{\ell,p} + \frac{\alpha_{\mathrm{LoRA}}}{r} B_{\ell,p} A_{\ell,p}, \qquad p \in \{Q, V\}. \tag{5}$$

Both rank r and LoRA alpha are four, fixed before the supplemental target-domain rank analysis. Base weights $W_{\ell,p}$ stay unchanged. The fast state starts from source-learned factors, and target gradient descent updates only these functional Q/V factors:

$$\phi^{(k+1)} = \phi^{(k)} - \eta_t \nabla_{\phi^{(k)}} \mathcal{L}_D(X_t, R_t; \phi^{(k)}), \qquad k = 0, \ldots, K_t - 1. \tag{6}$$

GigaPath, ST2 base weights, SpaFactor, objective heads, and W_v remain fixed, restricting adaptation to the Q/V fast state.

After adaptation, separate residual damping is applied to the local and regional representations:

$$h^{\gamma}_{i,\mathrm{loc}} = h^{0}_{i,\mathrm{loc}} + \gamma\big(h'_{i,\mathrm{loc}} - h^{0}_{i,\mathrm{loc}}\big), \qquad h^{\gamma}_{i,\mathrm{reg}} = h^{0}_{i,\mathrm{reg}} + \gamma\big(h'_{i,\mathrm{reg}} - h^{0}_{i,\mathrm{reg}}\big). \tag{7}$$

Superscript 0 marks the initial source representation; the prime marks its adapted version. The scale γ controls their contribution to the unchanged SpaFactor predictor without target labels: $\gamma = 0$ recovers the source representation and $\gamma = 1$ uses the adapted one.

3.6 Slide-Level Adaptation and Slide-Wise Reset

For each target slide, STTT resets the Q/V fast state to its source-trained initialization, updates it with $\mathcal{L}_D$ from that slide, applies residual damping, and predicts expression with unchanged SpaFactor. The fast state is discarded after prediction.

All spots on a slide may contribute jointly to the label-free objective; no target expression or adaptation state carries across slides.

4 Experiments

4.1 Datasets and Cross-Domain Settings

We evaluate four transfers: Dataset–Heart, STimage Visium Heart to HEST Visium Heart (dataset shift with organ and platform fixed); Organ–Breast, STimage Visium Liver to STimage Visium Breast (organ shift); Platform–Heart, HEST Visium Heart to HEST ST Heart (platform shift); and Compound–Breast, STimage Visium Breast to HEST ST Breast (cohort and platform shift). Appendix A gives dataset statistics and task configurations.

4.2 Preprocessing and Evaluation Protocol

Each spot uses a frozen 1536-dimensional GigaPath feature and two-dimensional coordinate; coordinates define $k = 4$ local and $k = 16$ regional neighbors.

Source training, source validation, and target slides use fixed task splits. Adaptation uses only target image features and coordinates, resets for each slide, and finalizes predictions before the evaluator accesses target expression.

For each task and seed, a source-supervised SpaFactor checkpoint initializes STTT. Source objective-head pretraining and adaptation-aware expression meta-training then produce the STTT checkpoint described in Section 3.4. During that source stage, the label-free $\mathcal{L}_D$ drives an inner TTT update, and source expression supervises the outer update. At deployment, STTT applies a new label-free update to each target slide while SpaFactor remains frozen. The SpaFactor (Source-only) reference predicts directly from its source-supervised checkpoint, without either STTT source-stage training or target-time adaptation. Thus, the main comparison evaluates complete training and deployment pipelines, rather than a target update from the same checkpoint.

4.3 Test-Time Adaptation Baselines

We compare STTT against seven representative test-time training or test-time adaptation methods implemented within the same prediction framework: SSA (Adachi et al., 2025), TTT-MAE-style (Gandelsman et al., 2022), SAR-Reg (Niu et al., 2023), ActMAD (Mirza et al., 2023), BUFR (Eastwood et al., 2022), RSD-TTA (Chen et al., 2021), and PEA (Ma et al., 2026).

All methods start from the corresponding SpaFactor predictor trained on source data and share GigaPath features, coordinates, target spots, gene panels, preprocessing, and evaluation. Source-only SpaFactor is the no-adaptation reference; none uses target expression during adaptation.

4.4 Metrics and Implementation Details

We report Pearson correlation coefficient (PCC), mean absolute error (MAE), and mean squared error (MSE). PCC is computed across the gene panel for each spot, averaged within slides, then aggregated by the same task protocol for all methods.

Matched seed scores are summarized by mean $\pm$ sample standard deviation (ddof = 1). PCC is reported as $100 \times \mathrm{PCC}$ (%); MAE and MSE use log1p expression.

The source-only versus STTT comparison uses twelve matched seeds, the adaptation-method comparison a common seven-seed subset, and the mechanism ablation six matched seeds. Each comparison keeps the seed and evaluation protocol fixed.

5 Results

5.1 Main Cross-Domain Performance

We first compare the source-only SpaFactor pipeline with the complete STTT–SpaFactor pipeline under domain shift. Table 1 reports their results across four transfer settings using the same twelve matched seeds.

STTT improves mean PCC in all four settings (Table 1). The gain is largest on Organ–Breast, rising from 30.76% to 33.53%, followed by Platform–Heart, which rises from 39.12% to 40.92%. The error metrics separate these correlation gains from changes in prediction magnitude. On Dataset–Heart and Compound–Breast, both MAE and MSE fall after adaptation. On Organ–Breast, MSE falls from 0.3440 to 0.3198 while MAE rises from 0.2910 to 0.3518; on Platform–Heart, both errors rise despite the PCC gain. Thus, adaptation improves correlation across all four shifts, but does not consistently improve the absolute expression estimates.

5.2 Comparison with Test-Time Adaptation Methods

We next investigate whether the observed improvement can be reproduced by generic test-time training or test-time adaptation strategies. All methods are evaluated using the same common seven-seed subset across the four transfer tasks. For each seed, we first compute an equal-weight macro average across the four tasks and subsequently aggregate the resulting seed-level macro scores.

STTT has the highest observed four-task macro PCC, at 45.31%, compared with 43.93% for source-only SpaFactor (Table 2). PEA reaches a similar PCC of 45.22%, but its MAE and MSE are higher than the source-only values. STTT instead lowers both macro errors, from 0.2606 to 0.2501 for MAE and from 0.2657 to 0.2080 for MSE. TTT-MAE-style and BUFR produce smaller PCC gains, while

Table 1: Cross-domain performance over 12 matched seeds. Mean ± std; best mean per metric within each task is bolded.

Task	Method	PCC (%) ↑	MAE ↓	MSE ↓
Dataset–Heart	SpaFactor (Source-only)	57.16±1.14	0.1102±0.0057	0.0654±0.0028
	STTT–SpaFactor (Ours)	**57.71**±1.02	**0.0908**±0.0047	**0.0643**±0.0046
Organ–Breast	SpaFactor (Source-only)	30.76±2.71	**0.2910**±0.0024	0.3440±0.0074
	STTT–SpaFactor (Ours)	**33.53**±2.68	0.3518±0.0119	**0.3198**±0.0294
Platform–Heart	SpaFactor (Source-only)	39.12±1.67	**0.1010**±0.0064	**0.0694**±0.0019
	STTT–SpaFactor (Ours)	**40.92**±1.40	0.1180±0.0165	0.0725±0.0067
Compound–Breast	SpaFactor (Source-only)	48.70±1.81	0.5217±0.0525	0.5562±0.0973
	STTT–SpaFactor (Ours)	**49.32**±1.49	**0.4489**±0.0340	**0.3843**±0.0409

Table 2: Four-task macro performance of TTT/TTA methods over seven matched seeds. Mean ± std; best mean per metric is bolded.

Method	PCC (%) ↑	MAE ↓	MSE ↓
SpaFactor (Source-only)	43.93±0.93	0.2606±0.0152	0.2657±0.0281
SSA	43.89±0.93	0.2633±0.0127	0.2707±0.0223
TTT-MAE-style	44.55±0.85	0.2717±0.0200	0.2750±0.0358
SAR-Reg	43.85±0.94	0.2633±0.0129	0.2709±0.0226
ActMAD	43.96±0.93	0.2606±0.0151	0.2655±0.0278
BUFR	44.38±0.85	0.2603±0.0130	0.2638±0.0245
RSD-TTA	43.95±0.93	0.2608±0.0152	0.2658±0.0281
PEA	45.22±0.25	0.2754±0.0100	0.2865±0.0191
STTT–SpaFactor (Ours)	**45.31**±0.79	**0.2501**±0.0091	**0.2080**±0.0097

SSA and SAR-Reg remain near the source-only reference. The comparison shows that a gain in correlation alone does not imply an error reduction; STTT combines the strongest observed macro correlation with the lowest macro errors in this set of methods.

5.3 Mechanism Ablation

We analyze the adaptation components on Compound–Breast using six matched seeds and the same target-time protocol (Table 3). Removing masked-self reconstruction reduces mean PCC from 49.68% to 48.22%, close to the 48.29% No update result. This pattern identifies masked-self reconstruction as the clearest contributor to correlation in this ablation. Removing center–context reconstruction produces a smaller PCC change, to 49.50%, and yields the lowest MAE and MSE in the table. The spatial reconstruction term therefore changes the balance between correlation and absolute error on this task, rather than improving every metric. Removing damping raises PCC to 49.96%, and updating the full ST2 reaches 51.03%; both variants have higher MAE and MSE than formal STTT. The formal Q/V LoRA update with damping restricts how much the representation changes, but the ablation does not identify it as the best configuration for every metric.

5.4 Q/V LoRA Rank Sensitivity

On Compound–Breast, Q/V LoRA ranks 1, 2, and 4 have close mean PCC values of 49.37%, 49.19%, and 49.33%, respectively (Table 4). Rank 1 also has the lowest mean MAE and MSE in this three-seed comparison. Increasing the rank to 8 lowers mean PCC to 48.65% without improving either error metric relative to rank 1. The extra adaptation capacity therefore provides no consistent benefit in this setting. The seed-level values in Appendix F show variation within each rank, so the small differences among ranks 1–4 should be read as a sensitivity pattern rather than a stable ordering.

Table 3: Mechanism ablation on Compound–Breast over six matched seeds. Mean ± std; best mean per metric is bolded.

Variant	PCC (%) ↑	MAE ↓	MSE ↓
No update	48.29±2.01	0.4413±0.0373	0.3894±0.0423
No center–context	49.50±2.41	**0.4361**±0.0362	**0.3513**±0.0597
No masked-self	48.22±1.95	0.4441±0.0429	0.3798±0.0623
No damping	49.96±1.80	0.4698±0.0445	0.4223±0.0641
Full ST2 update	**51.03**±0.90	0.4614±0.0372	0.4027±0.0710
STTT (Q/V LoRA; Ours)	49.68±1.83	0.4467±0.0363	0.3836±0.0375

Table 4: Compound–Breast Q/V LoRA rank sensitivity over three matched seeds.

Rank	PCC (%) ↑	MAE ↓	MSE ↓
1	**49.37**±0.26	**0.4349**±0.0187	**0.3615**±0.0414
2	49.19±1.15	0.4718±0.0255	0.3947±0.0067
4	49.33±2.04	0.4533±0.0384	0.3874±0.0427
8	48.65±1.35	0.4494±0.0185	0.3923±0.0039

5.5 Adaptation Step and Learning Rate Sensitivity

The single-seed Compound–Breast grid reveals a tradeoff between correlation and absolute error (Table 5). With two steps and learning rate 30, PCC reaches 51.03% and MSE falls to the table minimum of 0.3673, compared with 48.45% and 0.4000 without adaptation. At three steps and learning rate 30, PCC rises further to its maximum of 51.85%, but MAE and MSE increase to 0.4935 and 0.4995. The lowest MAE, 0.4163, occurs at three steps and learning rate 60, where PCC falls to 47.60%. Larger learning rates at two steps also increase the errors despite high PCC. More steps or a larger learning rate therefore do not improve all metrics together, and the preferred setting depends on which metric is prioritized.

Table 5: Step and learning-rate sensitivity on Compound–Breast (seed 42). Values are pooled (micro-aggregated) over all target spots and are therefore not directly comparable with the slide-level, group-equal-weighted metrics used in the formal evaluation. PCC is in percent; MAE/MSE are on log1p expression. Best per metric is bolded.

Steps	LR	PCC (%) ↑	MAE ↓	MSE ↓
0	0	48.45	0.4414	0.4000
1	30	48.66	0.4400	0.3979
1	60	48.99	0.4371	0.3937
1	120	49.87	0.4273	0.3790
2	30	51.03	0.4192	**0.3673**
2	60	51.77	0.4849	0.4848
2	120	51.12	0.5506	0.6320
3	30	**51.85**	0.4935	0.4995
3	60	47.60	**0.4163**	0.4186
3	120	49.30	0.5249	0.5631
4	30	49.72	0.4457	0.4215
4	60	47.53	0.5365	0.5932
4	120	49.40	0.5201	0.5555

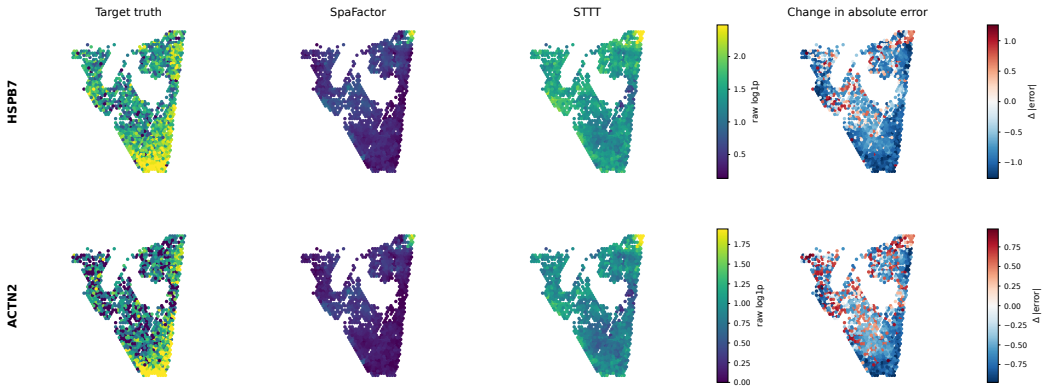

Figure 2: Spatial effects of target-time adaptation on Dataset–Heart slide MISC101 for *HSPB7* and *ACTN2*. Columns show target expression, source-only SpaFactor, STTT, and change in absolute error (blue: lower after adaptation; red: higher).

5.6 Spatial Effects of Target-Time Adaptation

Figure 2 compares target expression with source-only SpaFactor and STTT predictions for *HSPB7* and *ACTN2* on the Dataset–Heart slide MISC101. The source-only maps show lower expression and weaker regional contrast than the measured maps. STTT increases predicted expression across parts of the tissue and follows some high-expression regions more closely, although its maps still differ from target truth.

The rightmost panels show the spot-wise change in absolute error,

$$\Delta|e| = |\hat{y}_{\mathrm{STTT}} - y| - |\hat{y}_{\mathrm{Source}} - y|.$$

Negative (blue) values indicate lower error after adaptation; positive (red) values indicate higher error. For both genes, broad blue regions appear toward the lower part of the tissue where STTT raises expression relative to the source-only maps. Scattered red spots elsewhere show that this change is spatially uneven: the same target-time update improves some regions while increasing error in others.

6 Limitations

STTT is evaluated with SpaFactor as the base HE2ST predictor across four cross-domain transfer settings. Whether the combined source-stage meta-training and target-time adaptation generalize to other HE2ST backbones remains untested. Although PCC improves consistently at the task-mean level, absolute-error metrics remain domain dependent. The target-time update also introduces additional per-slide computation, and several sensitivity analyses are conducted on Compound–Breast only. Broader evaluation across predictors, datasets, and adaptation regimes is needed.

7 Conclusion

STTT uses unlabeled target-slide morphology and spatial neighborhoods to adapt a source-trained HE2ST predictor through masked-self and bidirectional center–context reconstruction. A differentiable source-stage inner update prepares the model for slide-level Q/V LoRA adaptation, while damping and reset constrain representation change. Across four cross-domain transfers, STTT–SpaFactor improves mean PCC over source-only SpaFactor and attains the highest four-task macro PCC among seven compared TTT/TTA methods on matched seeds. Ablation and spatial maps show that the effect depends on the adaptation objective and tissue region, while absolute-error responses remain task-dependent. These findings support spatial self-supervision for HE2ST test-time training in the evaluated settings.

AI USE STATEMENT

Generative AI tools, including OpenAI Codex and ChatGPT, were used to assist with manuscript drafting and editing, Chinese–English translation, literature organization and citation formatting, preparation of tables and figure descriptions, code development and debugging, and the organization and interpretation of completed experimental results. Generative AI was not used to generate experimental measurements or replace the execution of the reported experiments. All reported numerical results were obtained from the implemented experimental pipeline and were checked against the corresponding evaluation records. AI-assisted code, manuscript text, citations, tables, and figure descriptions were manually reviewed and verified by the authors. The authors take full responsibility for the final content of the manuscript and all reported results.

ETHICS STATEMENT

This study uses previously released spatial transcriptomics and histology datasets and does not involve the collection of new human-subject data. No patient-level identifying information is used in our experiments, and we do not attempt to identify or re-identify individual participants. The datasets are used in accordance with their respective access conditions and intended research purposes.

REPRODUCIBILITY STATEMENT

The manuscript describes the STTT architecture, source-stage training procedure, and slide-level target-time adaptation protocol used in the reported experiments. Appendix A summarizes the dataset and task configurations, while Appendix B provides the main implementation and computational setup. An anonymous repository accompanying the submission contains the core implementation of the proposed STTT method.

REFERENCES

Kazuki Adachi, Shin'ya Yamaguchi, Atsutoshi Kumagai, and Tomoki Hamagami. Test-time adaptation for regression by subspace alignment. In *International Conference on Learning Representations*, 2025.

Xinyang Chen, Sinan Wang, Jianmin Wang, and Mingsheng Long. Representation subspace distance for domain adaptation regression. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 1749–1759, 2021.

Youngmin Chung et al. Accurate spatial gene expression prediction by integrating multi-resolution features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11591–11600, 2024.

Cian Eastwood, Ian Mason, Christopher K. I. Williams, and Bernhard Schölkopf. Source-free adaptation to measurement shift via bottom-up feature restoration. In *International Conference on Learning Representations*, 2022.

Donghai Fang, Yongheng Li, Zhen Wang, Yuansong Zeng, and Wenwen Min. Adapting a pre-trained single-cell foundation model to spatial gene expression generation from histology images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5720–5729, 2026.

Xiaohang Fu et al. Spatial gene expression at single-cell resolution from histology using deep learning with GHIST. *Nature Methods*, 22:1900–1910, 2025. doi: 10.1038/s41592-025-02795-z.

Yossi Gandelsman, Yu Sun, Xinlei Chen, and Alexei A. Efros. Test-time training with masked autoencoders. In *Advances in Neural Information Processing Systems*, 2022.

Bryan He et al. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nature Biomedical Engineering*, 4:827–834, 2020. doi: 10.1038/s41551-020-0578-x.

Tinglin Huang, Tianyu Liu, Mehrtash Babadi, Wengong Jin, and Zhitao Ying. Scalable generation of spatial transcriptomics from histology images via whole-slide flow matching. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 25550–25565. PMLR, 2025.

Yuran Jia et al. THItoGene: a deep learning method for predicting spatial transcriptomics from histological images. *Briefings in Bioinformatics*, 25(1):bbad464, 2024. doi: 10.1093/bib/bbad464.

Ziang Liu et al. HistoGPA: A context-conditioned gene-prior attention framework for histology-based spatial gene expression prediction. *arXiv preprint arXiv:2607.24364*, 2026.

Xiao Ma, Young D. Kwon, Pan Zhou, and Dong Ma. Architecture-agnostic test-time adaptation via backprop-free embedding alignment. In *International Conference on Learning Representations*, 2026.

Muhammad Jehanzeb Mirza, Pol Jané Soneira, Wei Lin, Mateusz Kozinski, Horst Possegger, and Horst Bischof. ActMAD: Activation matching to align distributions for test-time-training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 24152–24161, 2023.

Taku Monjo et al. Efficient prediction of a spatial transcriptomics profile better characterizes breast cancer tissue sections without costly experimentation. *Scientific Reports*, 12:4133, 2022. doi: 10.1038/s41598-022-07685-4.

Shuaicheng Niu et al. Towards stable test-time adaptation in dynamic wild world. In *International Conference on Learning Representations*, 2023.

Kalin Nonchev et al. DeepSpot: Leveraging spatial context for enhanced spatial transcriptomics prediction from H&E images. *medRxiv*, 2025. doi: 10.1101/2025.02.09.25321567. Preprint.

Minxing Pang, Kenong Su, and Mingyao Li. Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors. *bioRxiv*, 2021. doi: 10.1101/2021.11.28.470212. Preprint.

Shiting Ruan, Xitong Ling, Qiming He, Ziyou Yan, Huaitian Yuan, Tian Guan, Ying Xiao, Xu Guan, and Yonghong He. SpaFactor: Lightweight spatial context-aware gene program modeling for histology-to-transcriptomics inference. *arXiv preprint arXiv:2609.28563*, 2026.

Patrik L. Ståhl et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294):78–82, 2016. doi: 10.1126/science.aaf2403.

Yu Sun et al. Test-time training with self-supervision for generalization under distribution shifts. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 9229–9248, 2020.

David Tellez et al. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Medical Image Analysis*, 58:101544, 2019. doi: 10.1016/j.media.2019.101544.

Chuhan Wang et al. Benchmarking the translational potential of spatial gene expression prediction from histology. *Nature Communications*, 16:1544, 2025. doi: 10.1038/s41467-025-56618-y.

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*, 2021.

Qin Wang, Olga Fink, Luc Van Gool, and Dengxin Dai. Continual test-time domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7201–7211, 2022. doi: 10.1109/CVPR52688.2022.00706.

Jinxi Xiang et al. A multimodal foundation model of spatial transcriptomics and histology for biological discovery and clinical prediction. *arXiv preprint arXiv:2604.03630*, 2026.

Ronald Xie et al. Spatially resolved gene expression prediction from H&E histology images via bi-modal contrastive learning. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-3095.

Hanwen Xu et al. A whole-slide foundation model for digital pathology from real-world data. *Nature*, 630:181–188, 2024. doi: 10.1038/s41586-024-07441-w.

Wang Yin et al. HESpotEx: A dual-stream deep learning framework for spot-level gene expression prediction from histological images. *Nature Computational Science*, 6:842–855, 2026. doi: 10.1038/s43588-026-00992-0.

Yuansong Zeng et al. Spatial transcriptomics prediction from histology jointly through transformer and graph neural networks. *Briefings in Bioinformatics*, 23(5):bbac297, 2022. doi: 10.1093/bib/bbac297.

Daiwei Zhang et al. Inferring super-resolution tissue architecture by integrating spatial transcriptomics with histology. *Nature Biotechnology*, 42(9):1372–1377, 2024. doi: 10.1038/s41587-023-02019-9.

Sichen Zhu et al. Diffusion generative modeling for spatially resolved gene expression inference from histology images. In *International Conference on Learning Representations*, 2025.

A Dataset and Task Details

Table A.1 summarizes the dataset and cross-domain task configurations used in the experiments. Source train/validation counts, target slide counts, and gene panels follow the formal task records.

Table A.1: Cross-domain dataset and task configuration.

Task	Source	Target	Shift	Source train / val.	Target slides	Genes
Dataset–Heart	STimage Visium Heart	HEST Visium Heart	Dataset	23/4	42	17,587
Organ–Breast	STimage Visium Liver	STimage Visium Breast	Organ	31/8	87	13,851
Platform–Heart	HEST Visium Heart	HEST ST Heart	Platform	32/5	19	16,287
Compound–Breast	STimage Visium Breast	HEST ST Breast	Compound	85/2	108	1,403

B Implementation and Computational Setup

B.1 Computational Environment

Experiments were run with an NVIDIA GeForce RTX 5090 using Python 3.10.20, PyTorch 2.13.0+cu130, and CUDA 13.0. Computation used FP32 precision.

B.2 Target-Time Adaptation Protocol

The GigaPath feature extractor is frozen. During target-time adaptation, only the functional Q/V LoRA state is updated; the ST2 base weights, SpaFactor predictor, and objective heads remain fixed. Target slides are processed independently, and the fast Q/V LoRA state is reset before each slide.

Target expression is not used during target-time adaptation. Predictions are finalized before the evaluator accesses target expression.

C Main Cross-Domain Comparison Results

Each table reports the 12 matched seed-level results for source-only SpaFactor and STTT. PCC is reported as a percentage, and summary values are mean ± std across matched seeds. Complete seed-level results are also provided in the supplementary data files.

Dataset–Heart

Table C.1: Dataset–Heart seed-level results for source-only SpaFactor and STTT.

Seed	PCC (%) ↑		MAE ↓		MSE ↓	
	SpaFactor	STTT	SpaFactor	STTT	SpaFactor	STTT
42	56.63	57.57	0.1113	0.0858	0.0666	0.0669
123	55.35	57.12	0.1127	0.0829	0.0695	0.0747
3407	56.89	57.34	0.1071	0.0985	0.0641	0.0601
2026	58.98	59.18	0.1106	0.0909	0.0620	0.0682
777	58.11	58.89	0.1030	0.0877	0.0624	0.0609
1337	57.98	58.89	0.1039	0.0926	0.0622	0.0588
999	57.15	58.35	0.1167	0.0916	0.0676	0.0619
512	57.84	58.45	0.1231	0.0938	0.0702	0.0613
1024	56.04	57.06	0.1072	0.0894	0.0663	0.0635
8888	55.81	56.14	0.1088	0.0988	0.0647	0.0610
8080	58.56	57.22	0.1128	0.0895	0.0657	0.0675
31415	56.62	56.36	0.1058	0.0881	0.0632	0.0667
Mean	57.16	**57.71**	0.1102	**0.0908**	0.0654	**0.0643**
Sample SD	1.14	1.02	0.0057	0.0047	0.0028	0.0046

Organ–Breast

Table C.2: Organ–Breast seed-level results for source-only SpaFactor and STTT.

Seed	PCC (%) ↑		MAE ↓		MSE ↓	
	SpaFactor	STTT	SpaFactor	STTT	SpaFactor	STTT
42	28.87	33.00	0.2896	0.3552	0.3521	0.3008
123	29.98	33.48	0.2907	0.3452	0.3486	0.3040
3407	30.82	32.77	0.2910	0.3492	0.3430	0.3113
2026	29.92	27.82	0.2892	0.3603	0.3497	0.3151
777	32.61	35.99	0.2887	0.3467	0.3426	0.3336
1337	27.72	31.87	0.2926	0.3544	0.3476	0.3127
999	31.87	32.35	0.2964	0.3413	0.3279	0.3241
512	28.78	35.71	0.2941	0.3543	0.3421	0.2896
1024	26.49	31.88	0.2892	0.3447	0.3549	0.3335
8888	32.75	38.17	0.2919	0.3424	0.3406	0.2791
8080	33.01	33.27	0.2894	0.3434	0.3434	0.3410
31415	36.29	36.03	0.2891	0.3845	0.3353	0.3929
Mean	30.76	**33.53**	**0.2910**	0.3518	0.3440	**0.3198**
Sample SD	2.71	2.68	0.0024	0.0119	0.0074	0.0294

Platform–Heart

Table C.3: Platform–Heart seed-level results for source-only SpaFactor and STTT.

Seed	PCC (%) ↑		MAE ↓		MSE ↓	
	SpaFactor	STTT	SpaFactor	STTT	SpaFactor	STTT
42	40.13	41.09	0.1090	0.0941	0.0693	0.0686
123	38.98	40.82	0.1055	0.1219	0.0680	0.0745
3407	39.05	41.46	0.0947	0.1122	0.0700	0.0673
2026	40.32	41.71	0.0953	0.1072	0.0690	0.0667
777	38.49	42.20	0.1023	0.1179	0.0691	0.0713
1337	41.33	41.49	0.1025	0.1390	0.0668	0.0837
999	35.99	36.83	0.0952	0.0924	0.0731	0.0743
512	37.83	40.27	0.0922	0.1239	0.0717	0.0675
1024	37.29	41.41	0.0937	0.1503	0.0711	0.0870
8888	40.93	40.52	0.1077	0.1247	0.0673	0.0719
8080	41.02	41.45	0.1037	0.1125	0.0671	0.0660
31415	38.07	41.77	0.1099	0.1197	0.0703	0.0707
Mean	39.12	**40.92**	**0.1010**	0.1180	**0.0694**	0.0725
Sample SD	1.67	1.40	0.0064	0.0165	0.0019	0.0067

COMPOUND–BREAST

Table C.4: Compound–Breast seed-level results for source-only SpaFactor and STTT.

Seed	PCC (%) ↑		MAE ↓		MSE ↓	
	SpaFactor	STTT	SpaFactor	STTT	SpaFactor	STTT
42	50.69	51.03	0.4582	0.4200	0.4218	0.3511
123	44.25	47.08	0.5495	0.4953	0.6064	0.4344
3407	48.83	49.89	0.5427	0.4445	0.5840	0.3768
2026	50.96	51.23	0.5519	0.4820	0.6024	0.4068
777	49.67	50.42	0.5108	0.4014	0.5142	0.3239
1337	49.12	47.92	0.5220	0.4115	0.5543	0.3653
999	49.26	50.85	0.6548	0.4810	0.8045	0.4068
512	48.00	48.66	0.4886	0.4757	0.5038	0.4158
1024	49.51	50.37	0.5004	0.4486	0.4691	0.3760
8888	47.30	47.67	0.4540	0.4484	0.4792	0.3790
8080	47.29	47.64	0.5086	0.4782	0.5867	0.4534
31415	49.52	49.12	0.5193	0.4006	0.5475	0.3220
Mean	48.70	**49.32**	0.5217	**0.4489**	0.5562	**0.3843**
Sample SD	1.81	1.49	0.0525	0.0340	0.0973	0.0409

D TEST-TIME ADAPTATION COMPARISON RESULTS

Each task uses the same common seven-seed subset. Summary values are reported as mean ± std across matched seeds. Main-text Table 2 instead averages the four tasks equally within each seed before aggregating seed-level macros. Complete seed-level results are also provided in the supplementary data files.

DATASET–HEART

Table D.1a: Dataset–Heart TTT/TTA comparison over seven matched seeds.

Method	PCC (%) ↑	MAE ↓	MSE ↓
SpaFactor (Source-only)	57.30±1.18	0.1093±0.0049	0.0649±0.0030
SSA	57.34±1.17	0.1092±0.0049	0.0648±0.0030
TTT-MAE-style	57.47±1.19	0.1149±0.0048	**0.0627**±0.0031
SAR-Reg	57.30±1.18	0.1093±0.0049	0.0649±0.0030
ActMAD	57.37±1.17	0.1091±0.0049	0.0648±0.0030
BUFR	**58.21**±0.86	0.1074±0.0048	0.0635±0.0028
RSD-TTA	57.30±1.17	0.1094±0.0049	0.0649±0.0030
PEA	57.72±0.82	0.1157±0.0061	0.0659±0.0037
STTT–SpaFactor (Ours)	58.19±0.84	**0.0900**±0.0051	0.0645±0.0057

Table D.1b: Dataset–Heart seed-level TTT/TTA results.

Seed	Method	PCC (%) ↑	MAE ↓	MSE ↓
42	SpaFactor (Source-only)	56.63	0.1113	0.0666
42	SSA	56.68	0.1112	0.0665
42	TTT-MAE-style	55.03	0.1142	0.0661
42	SAR-Reg	56.63	0.1113	0.0666
42	ActMAD	56.71	0.1111	0.0664
42	BUFR	57.68	0.1084	0.0646
42	RSD-TTA	56.64	0.1113	0.0666
42	PEA	58.01	0.1129	0.0653
42	STTT–SpaFactor (Ours)	57.57	0.0858	0.0669
123	SpaFactor (Source-only)	55.35	0.1127	0.0695
123	SSA	55.42	0.1125	0.0694
123	TTT-MAE-style	58.58	0.1159	0.0617
123	SAR-Reg	55.35	0.1127	0.0695
123	ActMAD	55.44	0.1126	0.0694
123	BUFR	56.74	0.1123	0.0685
123	RSD-TTA	55.37	0.1128	0.0695
123	PEA	56.48	0.1261	0.0733
123	STTT–SpaFactor (Ours)	57.12	0.0829	0.0747
3407	SpaFactor (Source-only)	56.89	0.1071	0.0641
3407	SSA	56.93	0.1069	0.0640
3407	TTT-MAE-style	56.97	0.1132	0.0610
3407	SAR-Reg	56.89	0.1071	0.0641
3407	ActMAD	56.96	0.1068	0.0639
3407	BUFR	57.94	0.1041	0.0623
3407	RSD-TTA	56.87	0.1072	0.0641
3407	PEA	57.12	0.1120	0.0654
3407	STTT–SpaFactor (Ours)	57.34	0.0985	0.0601
2026	SpaFactor (Source-only)	58.98	0.1106	0.0620
2026	SSA	59.00	0.1106	0.0620
2026	TTT-MAE-style	57.71	0.1139	0.0606
2026	SAR-Reg	58.98	0.1106	0.0620
2026	ActMAD	59.00	0.1105	0.0620
2026	BUFR	59.26	0.1096	0.0613
2026	RSD-TTA	58.97	0.1106	0.0621
2026	PEA	57.72	0.1150	0.0639
2026	STTT–SpaFactor (Ours)	59.18	0.0909	0.0682
777	SpaFactor (Source-only)	58.11	0.1030	0.0624
777	SSA	58.15	0.1029	0.0623
777	TTT-MAE-style	58.04	0.1139	0.0635
777	SAR-Reg	58.11	0.1030	0.0624
777	ActMAD	58.18	0.1028	0.0623
777	BUFR	58.87	0.1019	0.0615
777	RSD-TTA	58.11	0.1031	0.0624

Seed	Method	PCC (%) ↑	MAE ↓	MSE ↓
777	PEA	59.11	0.1164	0.0645
777	STTT–SpaFactor (Ours)	58.89	0.0877	0.0609
1337	SpaFactor (Source-only)	57.98	0.1039	0.0622
1337	SSA	58.02	0.1038	0.0621
1337	TTT-MAE-style	58.27	0.1089	0.0589
1337	SAR-Reg	57.98	0.1039	0.0622
1337	ActMAD	58.05	0.1037	0.0621
1337	BUFR	58.85	0.1022	0.0608
1337	RSD-TTA	57.96	0.1040	0.0622
1337	PEA	58.05	0.1072	0.0616
1337	STTT–SpaFactor (Ours)	58.89	0.0926	0.0588
999	SpaFactor (Source-only)	57.15	0.1167	0.0676
999	SSA	57.21	0.1165	0.0674
999	TTT-MAE-style	57.70	0.1246	0.0673
999	SAR-Reg	57.15	0.1167	0.0676
999	ActMAD	57.23	0.1164	0.0674
999	BUFR	58.15	0.1137	0.0655
999	RSD-TTA	57.17	0.1168	0.0676
999	PEA	57.58	0.1202	0.0674
999	STTT–SpaFactor (Ours)	58.35	0.0916	0.0619

ORGAN–BREAST

Table D.2a: Organ–Breast TTT/TTA comparison over seven matched seeds.

Method	PCC (%) ↑	MAE ↓	MSE ↓
SpaFactor (Source-only)	30.26±1.68	**0.2912**±0.0026	0.3445±0.0081
SSA	30.29±1.66	0.2912±0.0026	0.3442±0.0080
TTT-MAE-style	31.79±1.34	0.3011±0.0027	0.3267±0.0051
SAR-Reg	30.26±1.68	**0.2912**±0.0026	0.3445±0.0081
ActMAD	30.23±1.66	0.2912±0.0026	0.3444±0.0080
BUFR	29.98±1.48	0.2918±0.0026	0.3440±0.0072
RSD-TTA	30.30±1.66	0.2912±0.0026	0.3442±0.0080
PEA	31.52±1.14	0.2973±0.0015	0.3392±0.0051
STTT–SpaFactor (Ours)	**32.47**±2.44	0.3503±0.0066	**0.3145**±0.0113

Table D.2b: Organ–Breast seed-level TTT/TTA results.

Seed	Method	PCC (%) ↑	MAE ↓	MSE ↓
42	SpaFactor (Source-only)	28.87	0.2896	0.3521
42	SSA	28.95	0.2897	0.3516
42	TTT-MAE-style	32.58	0.2986	0.3254
42	SAR-Reg	28.87	0.2896	0.3521
42	ActMAD	28.88	0.2896	0.3518
42	BUFR	28.85	0.2905	0.3502
42	RSD-TTA	28.94	0.2896	0.3517
42	PEA	30.15	0.2986	0.3464
42	STTT–SpaFactor (Ours)	33.00	0.3552	0.3008
123	SpaFactor (Source-only)	29.98	0.2907	0.3486
123	SSA	30.04	0.2908	0.3483
123	TTT-MAE-style	32.75	0.2984	0.3215
123	SAR-Reg	29.98	0.2907	0.3486
123	ActMAD	30.00	0.2907	0.3485
123	BUFR	30.15	0.2912	0.3473
123	RSD-TTA	30.04	0.2907	0.3483
123	PEA	32.70	0.2980	0.3386
123	STTT–SpaFactor (Ours)	33.48	0.3452	0.3040

Seed	Method	PCC (%) ↑	MAE ↓	MSE ↓
3407	SpaFactor (Source-only)	30.82	0.2910	0.3430
3407	SSA	30.85	0.2910	0.3427
3407	TTT-MAE-style	31.04	0.2993	0.3288
3407	SAR-Reg	30.82	0.2910	0.3430
3407	ActMAD	30.79	0.2910	0.3430
3407	BUFR	30.36	0.2914	0.3430
3407	RSD-TTA	30.87	0.2910	0.3427
3407	PEA	31.57	0.2956	0.3368
3407	STTT–SpaFactor (Ours)	32.77	0.3492	0.3113
2026	SpaFactor (Source-only)	29.92	0.2892	0.3497
2026	SSA	29.97	0.2893	0.3493
2026	TTT-MAE-style	32.80	0.3011	0.3259
2026	SAR-Reg	29.92	0.2892	0.3497
2026	ActMAD	29.94	0.2893	0.3494
2026	BUFR	29.91	0.2901	0.3480
2026	RSD-TTA	29.98	0.2893	0.3494
2026	PEA	31.52	0.2978	0.3423
2026	STTT–SpaFactor (Ours)	27.82	0.3603	0.3151
777	SpaFactor (Source-only)	32.61	0.2887	0.3426
777	SSA	32.58	0.2888	0.3422
777	TTT-MAE-style	33.01	0.3041	0.3211
777	SAR-Reg	32.61	0.2887	0.3426
777	ActMAD	32.53	0.2888	0.3426
777	BUFR	31.79	0.2894	0.3425
777	RSD-TTA	32.61	0.2887	0.3423
777	PEA	32.00	0.2965	0.3388
777	STTT–SpaFactor (Ours)	35.99	0.3467	0.3336
1337	SpaFactor (Source-only)	27.72	0.2926	0.3476
1337	SSA	27.76	0.2926	0.3474
1337	TTT-MAE-style	29.50	0.3007	0.3360
1337	SAR-Reg	27.72	0.2926	0.3476
1337	ActMAD	27.69	0.2926	0.3476
1337	BUFR	27.43	0.2928	0.3479
1337	RSD-TTA	27.77	0.2926	0.3474
1337	PEA	29.90	0.2952	0.3416
1337	STTT–SpaFactor (Ours)	31.87	0.3544	0.3127
999	SpaFactor (Source-only)	31.87	0.2964	0.3279
999	SSA	31.88	0.2964	0.3279
999	TTT-MAE-style	30.87	0.3051	0.3281
999	SAR-Reg	31.87	0.2964	0.3279
999	ActMAD	31.82	0.2964	0.3280
999	BUFR	31.35	0.2971	0.3288
999	RSD-TTA	31.88	0.2964	0.3279
999	PEA	32.80	0.2991	0.3301
999	STTT–SpaFactor (Ours)	32.35	0.3413	0.3241

PLATFORM–HEART

Table D.3a: Platform–Heart TTT/TTA comparison over seven matched seeds.

Method	PCC (%) ↑	MAE ↓	MSE ↓
SpaFactor (Source-only)	39.18±1.71	0.1006±0.0057	0.0693±0.0020
SSA	39.23±1.70	0.1008±0.0057	0.0693±0.0019
TTT-MAE-style	40.85±1.10	0.1049±0.0059	**0.0677**±0.0011
SAR-Reg	39.18±1.71	0.1006±0.0057	0.0693±0.0020
ActMAD	39.24±1.70	**0.1006**±0.0056	0.0693±0.0020
BUFR	39.95±1.46	0.1012±0.0055	0.0686±0.0018
RSD-TTA	39.20±1.70	0.1007±0.0057	0.0693±0.0020
PEA	**41.38**±0.48	0.1126±0.0062	0.0712±0.0025
STTT–SpaFactor (Ours)	40.80±1.81	0.1121±0.0163	0.0724±0.0059

Table D.3b: Platform–Heart seed-level TTT/TTA results.

Seed	Method	PCC (%) ↑	MAE ↓	MSE ↓
42	SpaFactor (Source-only)	40.13	0.1090	0.0693
42	SSA	40.17	0.1092	0.0694
42	TTT-MAE-style	41.42	0.1033	0.0668
42	SAR-Reg	40.13	0.1090	0.0693
42	ActMAD	40.17	0.1087	0.0692
42	BUFR	40.75	0.1052	0.0681
42	RSD-TTA	40.15	0.1091	0.0694
42	PEA	41.41	0.1087	0.0704
42	STTT–SpaFactor (Ours)	41.09	0.0941	0.0686
123	SpaFactor (Source-only)	38.98	0.1055	0.0680
123	SSA	39.00	0.1057	0.0680
123	TTT-MAE-style	40.61	0.1091	0.0672
123	SAR-Reg	38.98	0.1055	0.0680
123	ActMAD	39.03	0.1056	0.0680
123	BUFR	39.71	0.1066	0.0676
123	RSD-TTA	38.99	0.1056	0.0680
123	PEA	41.20	0.1138	0.0700
123	STTT–SpaFactor (Ours)	40.82	0.1219	0.0745
3407	SpaFactor (Source-only)	39.05	0.0947	0.0700
3407	SSA	39.09	0.0949	0.0700
3407	TTT-MAE-style	40.45	0.1034	0.0675
3407	SAR-Reg	39.05	0.0947	0.0700
3407	ActMAD	39.10	0.0947	0.0700
3407	BUFR	39.87	0.0952	0.0693
3407	RSD-TTA	39.06	0.0947	0.0700
3407	PEA	41.30	0.1038	0.0684
3407	STTT–SpaFactor (Ours)	41.46	0.1122	0.0673
2026	SpaFactor (Source-only)	40.32	0.0953	0.0690
2026	SSA	40.35	0.0955	0.0690
2026	TTT-MAE-style	41.28	0.1015	0.0674
2026	SAR-Reg	40.32	0.0953	0.0690
2026	ActMAD	40.35	0.0954	0.0690
2026	BUFR	40.77	0.0975	0.0682
2026	RSD-TTA	40.34	0.0954	0.0690
2026	PEA	41.71	0.1126	0.0699
2026	STTT–SpaFactor (Ours)	41.71	0.1072	0.0667
777	SpaFactor (Source-only)	38.49	0.1023	0.0691
777	SSA	38.54	0.1025	0.0690
777	TTT-MAE-style	41.38	0.1100	0.0672
777	SAR-Reg	38.49	0.1023	0.0691
777	ActMAD	38.56	0.1024	0.0690
777	BUFR	39.64	0.1047	0.0683
777	RSD-TTA	38.50	0.1024	0.0691
777	PEA	41.82	0.1238	0.0750
777	STTT–SpaFactor (Ours)	42.20	0.1179	0.0713
1337	SpaFactor (Source-only)	41.33	0.1025	0.0668
1337	SSA	41.36	0.1028	0.0668
1337	TTT-MAE-style	42.12	0.1120	0.0679
1337	SAR-Reg	41.33	0.1025	0.0668
1337	ActMAD	41.37	0.1026	0.0668
1337	BUFR	41.81	0.1054	0.0667
1337	RSD-TTA	41.33	0.1025	0.0668
1337	PEA	41.78	0.1148	0.0703
1337	STTT–SpaFactor (Ours)	41.49	0.1390	0.0837
999	SpaFactor (Source-only)	35.99	0.0952	0.0731
999	SSA	36.07	0.0953	0.0730
999	TTT-MAE-style	38.71	0.0951	0.0701
999	SAR-Reg	35.99	0.0952	0.0731
999	ActMAD	36.06	0.0950	0.0731
999	BUFR	37.13	0.0936	0.0723
999	RSD-TTA	36.03	0.0953	0.0731
999	PEA	40.44	0.1105	0.0745
999	STTT–SpaFactor (Ours)	36.83	0.0924	0.0743

COMPOUND–BREAST

Table D.4a: Compound–Breast TTT/TTA comparison over seven matched seeds.

Method	PCC (%) ↑	MAE ↓	MSE ↓
SpaFactor (Source-only)	48.97±2.23	0.5414±0.0595	0.5840±0.1166
SSA	48.72±2.09	0.5521±0.0470	0.6043±0.0920
TTT-MAE-style	48.08±2.33	0.5657±0.0784	0.6429±0.1451
SAR-Reg	48.67±2.10	0.5519±0.0477	0.6048±0.0935
ActMAD	49.00±2.24	0.5414±0.0589	0.5834±0.1154
BUFR	49.37±2.42	0.5409±0.0513	0.5791±0.1019
RSD-TTA	48.99±2.23	0.5420±0.0595	0.5847±0.1167
PEA	**50.27**±0.70	0.5760±0.0396	0.6694±0.0788
STTT–SpaFactor (Ours)	49.77±1.63	**0.4480**±0.0383	**0.3807**±0.0379

Table D.4b: Compound–Breast seed-level TTT/TTA results.

Seed	Method	PCC (%) ↑	MAE ↓	MSE ↓
42	SpaFactor (Source-only)	50.69	0.4582	0.4218
42	SSA	48.67	0.5327	0.5686
42	TTT-MAE-style	49.83	0.5067	0.5540
42	SAR-Reg	48.61	0.5319	0.5679
42	ActMAD	50.77	0.4583	0.4213
42	BUFR	51.63	0.4611	0.4196
42	RSD-TTA	50.72	0.4588	0.4224
42	PEA	51.50	0.5239	0.5384
42	STTT–SpaFactor (Ours)	51.03	0.4200	0.3511
123	SpaFactor (Source-only)	44.25	0.5495	0.6064
123	SSA	44.29	0.5486	0.6042
123	TTT-MAE-style	44.39	0.5459	0.6022
123	SAR-Reg	44.25	0.5495	0.6064
123	ActMAD	44.24	0.5488	0.6052
123	BUFR	44.23	0.5424	0.5947
123	RSD-TTA	44.27	0.5500	0.6071
123	PEA	49.50	0.5294	0.6052
123	STTT–SpaFactor (Ours)	47.08	0.4953	0.4344
3407	SpaFactor (Source-only)	48.83	0.5427	0.5840
3407	SSA	48.89	0.5434	0.5843
3407	TTT-MAE-style	46.59	0.4680	0.4389
3407	SAR-Reg	48.83	0.5427	0.5840
3407	ActMAD	48.87	0.5424	0.5831
3407	BUFR	49.31	0.5378	0.5730
3407	RSD-TTA	48.86	0.5431	0.5845
3407	PEA	49.96	0.5795	0.6708
3407	STTT–SpaFactor (Ours)	49.89	0.4445	0.3768
2026	SpaFactor (Source-only)	50.96	0.5519	0.6024
2026	SSA	50.99	0.5525	0.6025
2026	TTT-MAE-style	50.39	0.6913	0.8817
2026	SAR-Reg	50.96	0.5519	0.6024
2026	ActMAD	50.96	0.5518	0.6020
2026	BUFR	51.01	0.5492	0.5957
2026	RSD-TTA	50.97	0.5526	0.6033
2026	PEA	49.90	0.5829	0.7133
2026	STTT–SpaFactor (Ours)	51.23	0.4820	0.4068
777	SpaFactor (Source-only)	49.67	0.5108	0.5142
777	SSA	49.69	0.5119	0.5159
777	TTT-MAE-style	50.75	0.6450	0.7748
777	SAR-Reg	49.67	0.5108	0.5142
777	ActMAD	49.71	0.5123	0.5163
777	BUFR	50.24	0.5302	0.5427
777	RSD-TTA	49.69	0.5114	0.5151

Seed	Method	PCC (%) ↑	MAE ↓	MSE ↓
777	PEA	49.96	0.6269	0.7435
777	STTT–SpaFactor (Ours)	50.42	0.4014	0.3239
1337	SpaFactor (Source-only)	49.12	0.5220	0.5543
1337	SSA	49.18	0.5222	0.5538
1337	TTT-MAE-style	47.35	0.5291	0.6194
1337	SAR-Reg	49.12	0.5220	0.5543
1337	ActMAD	49.17	0.5226	0.5548
1337	BUFR	49.81	0.5294	0.5623
1337	RSD-TTA	49.14	0.5225	0.5547
1337	PEA	50.97	0.6187	0.7608
1337	STTT–SpaFactor (Ours)	47.92	0.4115	0.3653
999	SpaFactor (Source-only)	49.26	0.6548	0.8045
999	SSA	49.30	0.6535	0.8010
999	TTT-MAE-style	47.29	0.5742	0.6291
999	SAR-Reg	49.26	0.6548	0.8045
999	ActMAD	49.28	0.6533	0.8013
999	BUFR	49.40	0.6361	0.7655
999	RSD-TTA	49.29	0.6554	0.8055
999	PEA	50.09	0.5707	0.6541
999	STTT–SpaFactor (Ours)	50.85	0.4810	0.4068

E Mechanism Ablation Results

The mechanism ablation is limited to Compound–Breast and uses the six matched seeds in the formal evaluation. Summary values are reported as mean ± std across matched seeds; individual seed-level values follow. Complete seed-level results are also provided in the supplementary data files.

Table E.1: Compound–Breast mechanism ablation over six matched seeds.

Variant	PCC (%) ↑	MAE ↓	MSE ↓
No update	48.29±2.01	0.4413±0.0373	0.3894±0.0423
No center–context	49.50±2.41	**0.4361**±0.0362	**0.3513**±0.0597
No masked-self	48.22±1.95	0.4441±0.0429	0.3798±0.0623
No damping	49.96±1.80	0.4698±0.0445	0.4223±0.0641
Full ST2 update	**51.03**±0.90	0.4614±0.0372	0.4027±0.0710
STTT (Q/V LoRA; Ours)	49.68±1.83	0.4467±0.0363	0.3836±0.0375

Table E.2: Compound–Breast mechanism ablation seed-level results.

Seed	Variant	PCC (%) ↑	MAE ↓	MSE ↓
42	No update	48.61	0.4361	0.3913
42	No center–context	44.82	0.4379	0.3703
42	No masked-self	47.40	0.5119	0.4592
42	No damping	51.87	0.4967	0.5049
42	Full ST2 update	49.99	0.4973	0.4781
42	STTT (Q/V LoRA; Ours)	51.91	0.4511	0.4021
123	No update	44.25	0.4846	0.4361
123	No center–context	49.80	0.4644	0.3954
123	No masked-self	46.77	0.4731	0.4358
123	No damping	49.06	0.5365	0.4976
123	Full ST2 update	49.96	0.4425	0.3476
123	STTT (Q/V LoRA; Ours)	47.11	0.4949	0.4327
3407	No update	48.83	0.4533	0.4091
3407	No center–context	49.77	0.4514	0.3887
3407	No masked-self	45.90	0.4346	0.3589
3407	No damping	50.42	0.4392	0.3599
3407	Full ST2 update	51.47	0.4039	0.3271

Seed	Variant	PCC (%) ↑	MAE ↓	MSE ↓
3407	STTT (Q/V LoRA; Ours)	49.87	0.4439	0.3760
777	No update	49.67	0.3861	0.3177
777	No center–context	51.78	0.4041	0.2804
777	No masked-self	48.50	0.4346	0.3844
777	No damping	50.56	0.4656	0.4083
777	Full ST2 update	50.93	0.4506	0.3804
777	STTT (Q/V LoRA; Ours)	50.44	0.4013	0.3236
1337	No update	49.12	0.4136	0.3663
1337	No center–context	50.55	0.4765	0.4024
1337	No masked-self	49.46	0.4221	0.3550
1337	No damping	46.80	0.4083	0.3638
1337	Full ST2 update	51.81	0.5031	0.5020
1337	STTT (Q/V LoRA; Ours)	47.89	0.4117	0.3659
999	No update	49.26	0.4744	0.4160
999	No center–context	50.25	0.3826	0.2707
999	No masked-self	51.27	0.3884	0.2855
999	No damping	51.03	0.4725	0.3994
999	Full ST2 update	52.04	0.4711	0.3813
999	STTT (Q/V LoRA; Ours)	50.86	0.4774	0.4010

F Target-Time Adaptation Sensitivity Results

Main-text Table 4 reports the three-seed rank summary. This appendix lists the individual rank results on Compound–Breast; rank 4 reuses the formal result, while ranks 1, 2, and 8 are supplementary runs.

Table F.1: Compound–Breast Q/V LoRA rank seed-level results.

Rank	Seed	PCC (%) ↑	MAE ↓	MSE ↓
1	42	49.17	0.4407	0.3731
1	123	49.28	0.4140	0.3155
1	3407	49.66	0.4501	0.3959
2	42	47.89	0.4615	0.3906
2	123	49.63	0.5009	0.3910
2	3407	50.06	0.4530	0.4024
4	42	51.03	0.4200	0.3511
4	123	47.08	0.4953	0.4344
4	3407	49.89	0.4445	0.3768
8	42	49.10	0.4514	0.3968
8	123	47.14	0.4669	0.3899
8	3407	49.73	0.4300	0.3902