

GigaPath-RoMa: Pathology-Adapted Dense Correspondence for Gigapixel Whole-Slide Image Registration

Yi Wang*

Zhejiang Key Laboratory of Multi-Omics in Infection and Immunity,
Center for Infectious Disease Research, School of Medicine, Westlake University,
No. 18 Shilongshan Street, Zhuantang Subdistrict, Xihu District,
Hangzhou, Zhejiang Province 310024, China
wangyi32@westlake.edu.cn

Abstract

Registering whole-slide images (WSIs) is essential for integrating complementary information across histological stains, yet remains challenging because of gigapixel resolution, pronounced cross-stain appearance discrepancies, nonlinear deformation, tissue tearing, and partial tissue loss. Although general-purpose registration methods have advanced substantially, they are not tailored to histopathological images and remain difficult to apply directly at gigapixel scale. In this work, we first present GigaPath-RoMa, a pathology-adapted dense correspondence framework for WSI registration. We adapt RoMa, a robust dense feature matching model, to the pathology domain through synthetic dense transformations and quality-aware certainty supervision. Furthermore, to scale dense matching to gigapixel WSIs, we introduce a global-to-local strategy that estimates a global correspondence field from a downsampled 4096-pixel WSI view and refines it across increasing WSI resolutions. Experiments on two histological registration datasets demonstrate that GigaPath-RoMa consistently outperforms representative classical and learning-based methods, with particularly strong improvements in robustness and tail registration errors. These results establish pathology-adapted dense correspondence as an effective basis for accurate gigapixel WSI registration.

1 Introduction

Histological stains reveal complementary aspects of tissue biology. Hematoxylin and eosin (H&E) depicts general morphology, whereas immunohistochemistry (IHC) and other special stains expose molecular markers and tissue-specific structures (Wodzinski et al. 2024b; Roy et al. 2023). Joint analysis of these signals supports annotation transfer, spatial biomarker analysis, three-dimensional reconstruction, and the integration of morphological and molecular evidence (Gatenbee et al. 2023). Such analyses require accurate registration: even a small local misalignment can associate a cell, gland, or lesion with an incorrect molecular measurement. The need for automatic and reproducible alignment has therefore motivated dedicated histological registration benchmarks such as ANHIR (Borovec et al. 2020) and a

growing family of whole-slide image (WSI) registration systems (Wodzinski and Müller 2021; Gatenbee et al. 2023; Wodzinski et al. 2024b).

WSI registration is substantially more difficult than ordinary image alignment. A single slide may contain hundreds of millions to billions of pixels, precluding direct full-resolution inference with modern dense matching networks. Serial sectioning and slide preparation introduce nonlinear deformation, folds, tears, fragmented tissue, and regions with no valid counterpart (Figure 1). Cross-stain registration further violates brightness and color constancy because corresponding structures may have markedly different appearances or may only be visible under one stain (Borovec et al. 2020; Weitz et al. 2024; Elhaminia et al. 2025). Slides can additionally differ by large translation or discrete rotation. These effects occur jointly and at different spatial scales: global orientation and displacement must be recovered while fine tissue boundaries require localized nonrigid correction.

Existing histological registration methods commonly follow a coarse-to-fine pipeline. They first estimate a rigid or affine alignment and subsequently optimize a nonrigid transformation at higher resolution (Wodzinski and Müller 2021; Gatenbee et al. 2023; Wodzinski et al. 2024b). This design has produced strong practical systems, but local refinement remains dependent on a sufficiently reliable initialization. A direct tiled extension presents a related problem: a target tile and a source tile extracted at identical coordinates need not depict corresponding tissue after large displacement, rotation, deformation, or fragmentation. Once an invalid tile pair is passed to a local matcher, an apparently confident local transformation can corrupt an otherwise plausible global alignment. Thus, scaling WSI registration requires not only an accurate local matcher, but also a reliable mechanism for constructing and fusing local correspondences.

Dense correspondence models offer a promising basis for this mechanism. Recent approaches have progressively improved coverage and robustness through learned dense confidence, coarse-to-fine refinement, and foundation-model features (Edstedt et al. 2023, 2024). In particular, RoMa produces pixel-dense warps together with certainty estimates and remains robust under substantial appearance and geometric changes (Edstedt et al. 2024). Nevertheless, directly transferring existing image registration methods to WSI registration faces two fundamental gaps. First, models trained

*Corresponding author.

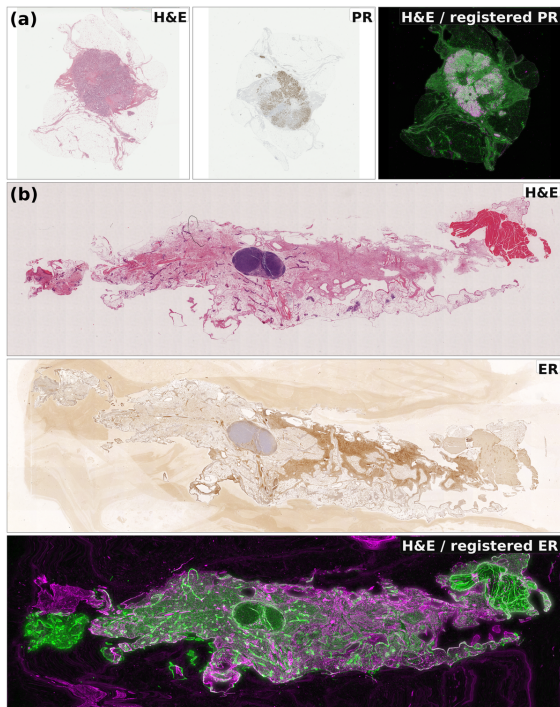


Figure 1: Examples of cross-stain WSI pairs before and after registration: (a) H&E and progesterone receptor (PR) IHC; (b) H&E and estrogen receptor (ER) IHC.

on natural or general medical images are not adapted to stain-dependent appearance, repetitive histological texture, and partially missing tissue. Second, these methods are designed for much lower image resolutions: globally downsampling a gigapixel WSI discards fine tissue detail, whereas local processing requires corresponding regions to be identified and their predictions to be fused into a consistent high-resolution mapping. Achieving accurate registration at gigapixel scale therefore remains a central challenge.

In this work, we present *GigaPath-RoMa*, a pathology-adapted dense correspondence framework for gigapixel WSI registration. We separate pathology-domain correspondence learning from gigapixel inference. First, we fine-tune RoMa with pathology-derived synthetic dense transformations and introduce quality-aware certainty supervision, so that certainty reflects both correspondence validity and localization accuracy. Second, we introduce a multiscale strategy for gigapixel WSI registration. A high-detail global correspondence field first establishes whole-slide alignment and then guides two levels of bounded local residual refinement. The resulting confidence-weighted residuals are fused on a smooth control grid to produce a continuous full-resolution mapping. Extensive benchmarking on ANHIR and HyReCo shows that *GigaPath-RoMa* outperforms all evaluated full-coverage baselines.

We summarize the main contributions as follows:

- We adapt a general-purpose dense correspondence model to the pathology domain using synthetic dense transformations and quality-aware certainty supervision, provid-

ing a reliability signal suited to filtering and weighting histological correspondences.

- We introduce a global-to-local framework for gigapixel dense registration that progressively constructs mapping-guided local correspondences and integrates their residuals into a continuous full-resolution mapping.
- We conduct comprehensive experiments on ANHIR and HyReCo, demonstrating the best performance among the evaluated full-coverage methods.

2 Related Work

2.1 Histological and Whole-Slide Image Registration

Histological registration must handle stain shifts, gigapixel images, and structural discontinuities (Elhaminia et al. 2025). ANHIR evaluates nonrigid landmark accuracy across tissues and stains, while ACROBAT studies clinical H&E-IHC breast slides; both expose failures under large displacement, damage, and partial overlap (Borovec et al. 2020; Weitz et al. 2024).

Conventional pipelines combine low-resolution initialization with patch-based or region-focused refinement (Lotz et al. 2016; Rossetti et al. 2017; Paknezhad et al. 2020; Jiang et al. 2019). Multiresolution intensity optimization, as in Elastix, can be brittle across stains (Klein et al. 2010); VALIS instead combines tissue masking, feature-based rigid alignment, and nonrigid micro-registration (Gatenbee et al. 2023).

Learning-based methods learn features, synthesize shared appearance, or predict deformation (Wodzinski and Müller 2021; Ge et al. 2022; Awan et al. 2023; Hoque et al. 2022; Roy et al. 2023). RegWSI and DeeperHistReg combine learned initialization with WSI-scale refinement infrastructure (Wodzinski et al. 2024b,a); SGTTA adapts style and structure at test time, and SAR2Net retrieves spatially anchored regions (Zhou, Xiong, and Wu 2024; Shen, Yan, and Zhang 2026). Wang, He, and Yan (2026) applied foundation-model matching to heterogeneous spatial omics. Unlike methods requiring coarse overlap or fixed local pairing, ours constructs local pairs from an evolving dense mapping.

2.2 Image Registration and Correspondence Methods

Intensity-based methods optimize similarity under spline, diffusion, or diffeomorphic models (Rueckert et al. 1999; Thirion 1998; Avants et al. 2008). VoxelMorph and TransMorph instead predict dense deformation with convolutional and transformer models (Balakrishnan et al. 2018; Chen et al. 2022), but cross-stain shifts are harder than their typical same-modality settings.

Sparse methods range from SIFT to learned detection and assignment with SuperPoint, SuperGlue, and LightGlue (Lowe 2004; DeTone, Malisiewicz, and Rabinovich 2018; Sarlin et al. 2020; Lindenberger, Sarlin, and Pollefeys 2023), but uneven coverage limits local deformation accuracy. Detector-free methods such as LoFTR, DKM, and RoMa provide broader correspondence coverage (Sun et al.

2021; Edstedt et al. 2023, 2024). Pyramids, tiling, and distributed frameworks address large-image memory costs (Jena et al. 2026).

RoMa is particularly suitable for WSI registration: its pixel-dense warp captures spatially varying deformation when landmarks are sparse, while certainty suppresses unreliable matches in background, damaged, missing, or non-overlapping regions. Yet natural-image training and ordinary-resolution inference do not address pathology-specific appearance or gigapixel pair construction and fusion. GigaPath-RoMa closes these gaps through pathology adaptation and global-to-local multiscale refinement, in which an evolving dense mapping guides local residual matching across WSI resolutions.

3 Methodology

3.1 Problem Formulation and Framework Overview

Let $F : \Omega_F \rightarrow \mathbb{R}^3$ and $I : \Omega_I \rightarrow \mathbb{R}^3$ denote a target (reference) whole-slide image (WSI) and a source WSI, respectively. Our goal is to estimate a dense backward mapping

$$M : \Omega_F \rightarrow \Omega_I, \quad (1)$$

such that a location $p \in \Omega_F$ corresponds to $M(p)$ in the source image. The source image registered on the target-image grid is consequently obtained as $I_{\text{warp}}(p) = I(M(p))$. This backward formulation is consistent with dense image resampling and allows all estimated correspondences to be represented in the coordinate system of the target WSI.

Directly applying a dense correspondence network to a WSI is infeasible because the slide can contain hundreds of millions to billions of pixels, whereas existing matchers operate on ordinary-resolution images. Independently matching tiles at identical spatial coordinates is also unreliable: rotation, large displacement, nonlinear tissue deformation, tearing, and missing tissue can cause two equally positioned tiles to contain unrelated structures. GigaPath-RoMa addresses these two issues by separating pathology-domain correspondence learning from gigapixel inference.

The framework consists of two stages (Figure 2). First, we adapt a general-purpose dense correspondence model to histopathology using synthetic dense transformations and quality-aware certainty supervision. Second, at inference time, we perform a discrete rotation search and estimate a high-detail global correspondence field from aspect-ratio-preserving WSI views. This field serves as a spatial anchor for a two-level residual pyramid. At each level, uniformly distributed overlapping target tiles are paired with source search windows located by the evolving mapping. Each local matcher predicts a bounded residual correction rather than an independent absolute transformation. Reliable residual observations are finally integrated into a continuous field on a smooth control grid.

Starting from the global mapping M_0 , the multiscale update is written as

$$M_{s+1}(p) = M_s(p) + \Delta_s(p), \quad (2)$$

where Δ_s is the fused residual field estimated at refinement scale s .

3.2 Pathology-Adapted Dense Correspondence

Synthetic Pathology Correspondence Generation

General-purpose dense matchers are trained predominantly on natural images and therefore do not fully capture the stain appearance, repeated texture, weakly textured regions, and tissue boundaries encountered in histology. We adapt RoMa to this domain using H&E tiles sampled from TCGA-UT (Komura et al. 2022). TCGA-UT contains approximately 1.6 million multiresolution H&E tiles extracted from pathologist-selected uniform tumor regions across 32 TCGA cancer types. Given a pathology tile X , we sample a synthetic spatial transformation T and generate a paired image $X' = \mathcal{W}(X, T)$. The sampled transformations combine global geometric perturbations, including translation, rotation, scale, anisotropic scale, and shear, with smooth local elastic deformation. Cropping and out-of-bound sampling additionally produce partially valid correspondence regions that resemble tissue loss and non-overlap.

Because T is known, every valid pixel has a dense ground-truth correspondence. This provides substantially denser supervision than sparse landmark annotations and permits transformations to be resampled online during training. We retain RoMa’s dense correspondence objective (Edstedt et al. 2024) and fine-tune the model with these pathology-derived synthetic pairs.

Quality-Aware Certainty Supervision A binary validity target does not distinguish an accurate correspondence from a poorly localized prediction when both are valid. This distinction is particularly important for WSI registration because certainty is subsequently used to filter and weight local residuals. For a pixel p , let $y_{\text{valid}}(p) \in \{0, 1\}$ indicate whether its synthetic correspondence is valid, and let $e(p)$ denote the endpoint error between the predicted correspondence and the synthetic ground truth, measured in pixels. We define the quality-aware certainty target as

$$y_{\text{cert}}(p) = y_{\text{valid}}(p) \exp\left(-\frac{e(p)}{\tau}\right), \quad (3)$$

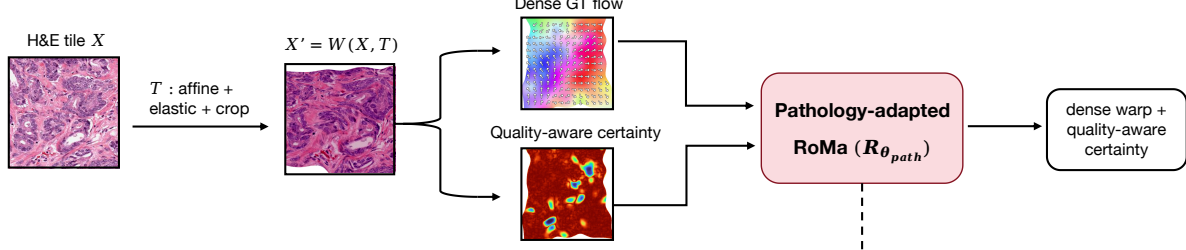
where τ controls the rate at which certainty decreases with localization error. Thus, invalid correspondences receive zero target certainty, while valid predictions receive high certainty only when their endpoint error is small. The certainty head is optimized against this soft target together with the dense correspondence objective, with gradients stopped through the target. We denote the resulting pathology-adapted matcher by $\mathcal{R}_{\theta_{\text{path}}}$.

This objective encourages certainty to encode correspondence quality rather than validity alone. It does not assume that the output is a perfectly calibrated probability on every unseen stain pair; instead, it provides a more informative reliability signal for the quality-controlled WSI inference described below.

3.3 Automatic Rotation Alignment and Global Correspondence

Histological sections may be stored under different discrete orientations. We therefore align their coarse orientation be-

(a) Pathology-domain dense correspondence adaptation



(b) Mapping-guided gigapixel WSI registration

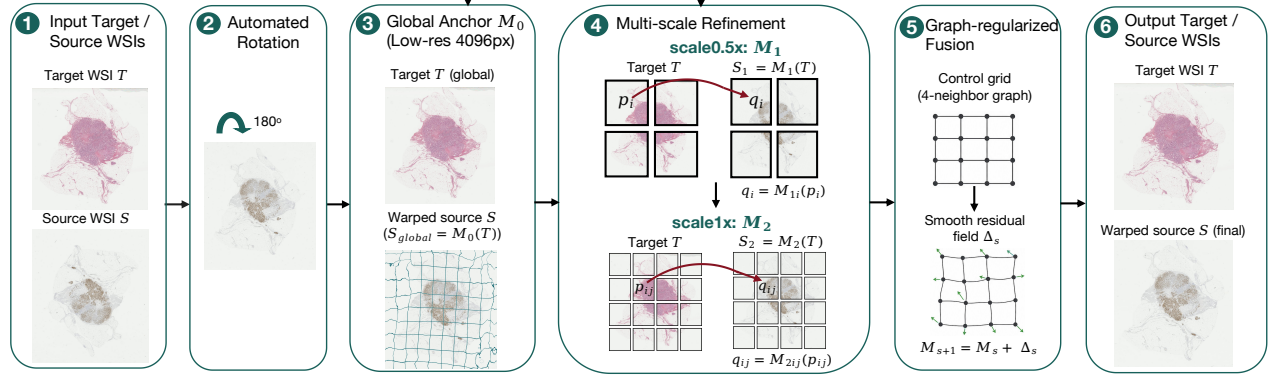


Figure 2: Pipeline of GigaPath-RoMa. (a) Synthetic pathology pairs adapt RoMa using dense correspondence and quality-aware certainty supervision; (b) global alignment guides multiscale local refinement and control-grid fusion.

fore estimating the final mapping. Low-resolution views of the two WSIs are first constructed while preserving their aspect ratios. For each candidate rotation

$$r \in \mathcal{O} = \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}, \quad (4)$$

we rotate the source-image view by r , apply the pathology-adapted matcher, and compute a correspondence score S_r from its dense matching certainty. The selected orientation is

$$\hat{r} = \arg \max_{r \in \mathcal{O}} S_r. \quad (5)$$

All subsequent matching is performed against the oriented source image $\tilde{I} = \mathcal{T}_{\hat{r}}(I)$. The final correspondences are transformed back to the coordinate system of the original source WSI, so the orientation operation does not alter the definition of M .

After rotation alignment, we construct a second pair of aspect-ratio-preserving WSI views with a maximum side length of 4096 pixels. The pathology-adapted matcher retains its 560-pixel coarse reasoning stage while enabling high-resolution refinement and prediction upsampling at the WSI-view resolution. We apply $\mathcal{R}_{\theta_{\text{path}}}$ to obtain the initial global field

$$\tilde{M}_0 = \mathcal{R}_{\theta_{\text{path}}} \left(F^{(0)}, \tilde{I}^{(0)} \right). \quad (6)$$

The predicted coordinates are rescaled from these WSI views to the original slide coordinate systems, yielding M_0 . The 4096-pixel prediction retains substantially more spatial detail

than the orientation-search view while remaining tractable for whole-image inference. Nevertheless, it is used as a global anchor rather than treated as a full-resolution solution: it captures orientation, large displacement, and coarse-to-intermediate tissue correspondence, while local stages correct the remaining error.

3.4 Mapping-Guided Multiscale Residual Matching

We refine M_0 at two strictly increasing image scales. The same refinement algorithm is used for every dataset, but the pyramid levels are selected from the native WSI dimensions so that a level contains multiple 2048×2048 tiles without producing an intractable tile count. We choose the finest computationally feasible level and use half of that resolution as the preceding level. Thus, the medium-resolution ANHIR images use $0.5\times$ and $1.0\times$, whereas the much larger native-resolution HyReCo slides use $0.0625\times$ and $0.125\times$. At each level, fixed-size overlapping tiles are distributed uniformly across the target domain. Uniform spacing avoids a narrow, atypical edge tile while preserving complete coverage and approximately 50% overlap.

At refinement scale s , let p_i be the center of the target tile i . Instead of extracting a source tile at the same spatial coordinate, we query the current correspondence field:

$$q_i = M_s(p_i). \quad (7)$$

A source-image search window is then centered at q_i . This mapping-guided construction conditions local matching on

the alignment accumulated from all coarser stages. It is essential when the corresponding tissue is displaced from the target tile coordinate or when only part of the tile overlaps the source section.

The target tile and its predicted source search region are passed to the pathology-adapted matcher. For a sampled target location p_{ij} in tile i , the local prediction, converted to WSI coordinates, is denoted by $M_{s,i}^{\text{local}}(p_{ij})$. We formulate its contribution as a residual observation

$$d_{ij} = M_{s,i}^{\text{local}}(p_{ij}) - M_s(p_{ij}). \quad (8)$$

This formulation prevents a local tile prediction from directly replacing a globally plausible mapping. It also supplies a natural zero-residual prior: where local evidence is absent or unreliable, the current mapping is retained.

Refinement proceeds from coarser to finer WSI resolutions. The tile size is fixed in level pixels, so its physical footprint decreases as resolution increases and progressively finer tissue structure becomes available. To prevent a late local mismatch from damaging the global solution, the norm of each fused level update is clipped by a scale-dependent trust bound,

$$\|\Delta_s(p)\|_2 \leq \rho_s, \quad \rho_{s+1} < \rho_s, \quad (9)$$

with full-resolution bounds of 8 and 4 pixels from coarse to fine in the selected two-level configuration. After residuals at scale s are fused and bounded, the updated field M_{s+1} is used to relocate every source search window at the next scale. Consequently, local correspondence construction and mapping estimation form a closed coarse-to-fine process: a better mapping produces better-aligned search regions, which in turn provide more reliable residual evidence.

Before fusion, we filter unreliable residuals from background, missing tissue, or incorrect search regions. Correspondences with low certainty or excessive displacement are removed, and tiles must contain sufficient tissue and a sufficient number of retained samples. Robust within-tile statistics suppress residual outliers and reject tiles with excessive median absolute deviation (MAD). The remaining observations are weighted by correspondence certainty and spatial proximity. Finally, we retain a tile update only when it improves grayscale normalized mutual information (NMI) between the target and locally warped source; otherwise, the current mapping is preserved.

3.5 Graph-Regularized Control-Grid Fusion

The retained residual observations are interpolated into a continuous update on a regular control grid. Indexing all retained residuals by i , let $G_j \in \mathbb{R}^2$ denote the displacement at control node j , and let $B_i G$ denote the bilinearly interpolated grid displacement at observation i . We minimize

$$\begin{aligned} \mathcal{E}(G) = & \sum_i w_i \|B_i G - d_i\|_2^2 + \lambda_{\text{prior}} \sum_j \|G_j\|_2^2 \\ & + \lambda_{\text{smooth}} \sum_{(j,k) \in \mathcal{E}_G} \|G_j - G_k\|_2^2, \end{aligned} \quad (10)$$

where w_i is the residual reliability weight, and $\mathcal{E}_G$ contains four-neighbor control-grid edges. The three terms fit the retained observations, penalize drift from the global anchor,

and encourage neighboring control nodes to vary smoothly. Robust reweighting limits the influence of inconsistent observations. Bilinear interpolation of the optimized grid gives $\Delta_s(p)$, which updates the mapping through Eq. (2) before the next refinement level.

4 Experiments

4.1 Datasets and Metrics

For domain adaptation, synthetic transformations pair H&E tiles from TCGA-UT and provide dense correspondence and certainty supervision without manual annotations or images from either evaluation dataset (Komura et al. 2022).

We evaluate on two cross-stain datasets. ANHIR spans tissues, stains, orientations, and resolutions; our fixed manifest has 222 pairs and 19,967 landmark correspondences (Borovec et al. 2020). HyReCo contains high-resolution H&E slides paired with slides stained for CD8, CD45, Ki-67, or PHH3; excluding seven corrupted archive images leaves 27 pairs and 428 landmarks (Wodzinski et al. 2024b). Every method uses the same manifests.

For a target landmark p_i , its corresponding source landmark q_i , and an estimated backward mapping M , we compute the target registration error (TRE) as

$$e_i = \|M(p_i) - q_i\|_2. \quad (11)$$

ANHIR follows the benchmark convention for relative TRE, $\text{rTRE}_i = e_i/D_I$, where D_I is the source-image diagonal (Borovec et al. 2020); HyReCo TRE is converted to μm using the source slide’s axis-specific pixel spacing. We report pooled mean, median, and 95th-percentile (P95) errors, plus the median per-pair median for HyReCo. Metrics in the two units are reported separately.

4.2 Baseline Methods

We compare three reproducible WSI pipelines: VALIS, which combines tissue preprocessing, feature-based rigid alignment, and nonrigid micro-registration (Gatenbee et al. 2023); DeeperHistReg, a deep-initialized multistage framework whose outputs we convert from padded, downsampled coordinates back to native coordinates (Wodzinski et al. 2024a); and LDDMM–Demons, a smooth intensity-based baseline with automatic orientation handling (Thirion 1998; Avants et al. 2008). We include only locally verified implementations that export dense transformations and run on identical manifests.

4.3 Implementation

Pathology adaptation uses seed 0 for Python, NumPy, and PyTorch, 560×560 inputs, batch size 2, and 5,000 mixed-precision AdamW steps at 10^{-5} with a frozen encoder. The certainty loss weight is 0.01 and its base τ is 4 pixels, scaled at coarser prediction levels. Inference parameters are shared across datasets except for pyramid scales. The global view has a maximum side length of 4096 pixels; RoMa matches coarsely at 560×560 and upsamples its prediction (Edstedt et al. 2024). Uniform local grids use 2048×2048 tiles, 1024-pixel overlap, 512-pixel source margin, 256-pixel sampling stride, and 512-pixel control spacing.

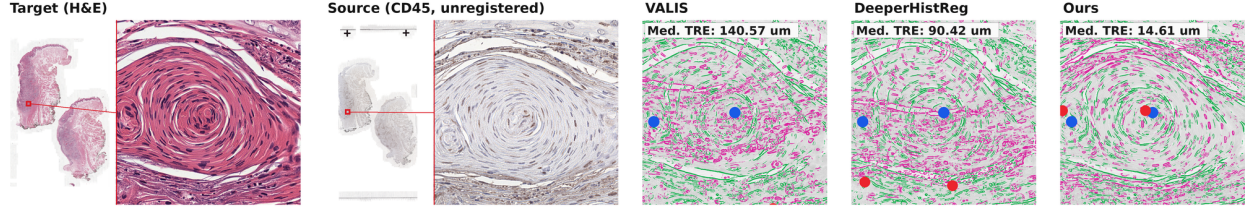
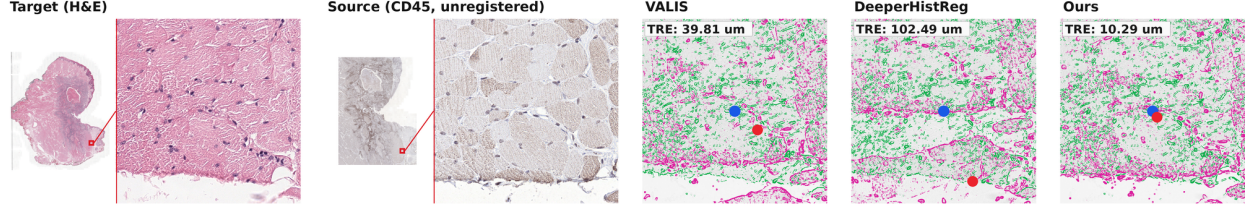
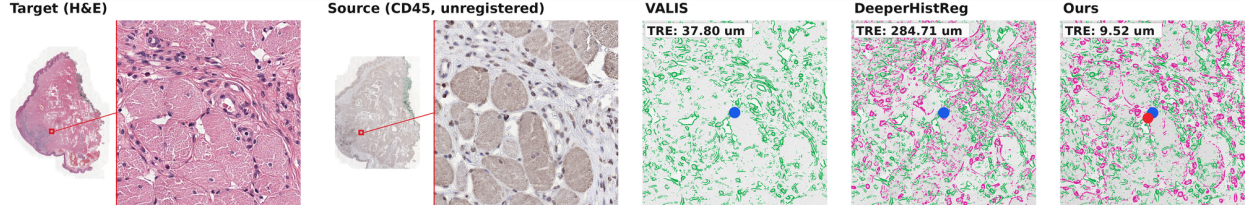
HyReCo 108: H&E to CD45**HyReCo 533: H&E to CD45****HyReCo 611: H&E to CD45**

Figure 3: Qualitative comparison on three HyReCo H&E-CD45 pairs. Each displayed crop covers a 1024×1024 pixel region in the native WSI coordinate system, located by the red box in the corresponding whole-slide thumbnail. Green and magenta contours show target and registered-source edges, respectively; blue dots are target landmarks, and red dots are the corresponding transformed source landmarks. Red landmarks are absent from the VALIS panels when their transformed locations fall outside the displayed region because of large registration errors. Lower TRE indicates better registration.

Tiles require 2% tissue, certainty ≥ 0.10 , and 12 retained samples. Fusion uses prior and smoothness weights of 0.5 and 0.05, respectively, with two robust iterations; updates are accepted only when NMI is at least 1.0 and exceeds that of the current mapping. The selected scales are $0.0625\times$ and $0.125\times$ for HyReCo (median native long side of 220,572 pixels), and $0.5\times$ and $1.0\times$ for ANHIR (10,987 pixels). At the coarse and fine levels, residuals above 256 and 128 native pixels, respectively, are rejected. The corresponding MAD thresholds are 64 and 32 pixels, and the update bounds are 8 and 4 pixels. Scales are fixed within each dataset.

The three-level ablation prepends $0.03125\times$ on HyReCo or $0.25\times$ on ANHIR, with residual, MAD, and update limits of 512, 128, and 16 native pixels. Baselines use aspect-ratio-preserving views of at most 4096 pixels; mappings are evaluated in native WSI coordinates. Experiments used two 40-GB NVIDIA A100-SXM4 GPUs and an AMD EPYC 7763 CPU. GigaPath-RoMa and DeeperHistReg used parallel single-GPU shards; VALIS and LDDMM-Demons used parallel 14-thread CPU shards. Each configuration was evaluated once on its complete fixed manifest; failed jobs were retried without ensembling, and tables report these single-run evaluations. The supplementary code and data package provides all preprocessing, experiment, and analysis code with step-level comments, complete parameters, and hardware/software versions; it will accompany the public code release. Generative AI tools were used for language polish-

ing and programming assistance; all resulting content was verified by the authors, who take full responsibility for the manuscript.

4.4 Quantitative Comparison

Tables 1 and 2 summarize the final controlled reruns, including both the complete two-level GigaPath-RoMa and its Global-4096 variant without local refinement. On HyReCo, the Global-4096 variant already reduces mean, median, and P95 TRE by 8.9%, 9.7%, and 4.0%, respectively, relative to VALIS, the strongest complete baseline. Local multiscale refinement further reduces these errors by 1.6%, 7.5%, and 1.4%, yielding total reductions of 10.4%, 16.4%, and 5.4% relative to VALIS. DeeperHistReg has a competitive median but a much larger mean and P95, revealing severe outliers, while LDDMM-Demons produces substantially larger errors and a heavier error tail.

On ANHIR, the Global-4096 variant and full GigaPath-RoMa perform similarly: local refinement slightly improves mean and P95 rTRE while the Global-4096 median is marginally lower. Both substantially outperform DeeperHistReg, the strongest baseline that completes the entire manifest; the full model reduces mean, median, and P95 rTRE by 37.2%, 20.2%, and 27.2%, respectively. VALIS succeeds on only 196 of 222 pairs despite a retry; we report its successful-pair statistics for completeness but exclude it from the strict full-manifest ranking. All other methods com-

Table 1: Quantitative comparison on ANHIR using the mean, median, and 95th percentile (P95) of rTRE. Reported values are multiplied by 10^3 ; lower is better. VALIS[†] is evaluated on 196 successfully registered pairs, with failures on the remaining 26 pairs.

Method	Mean ↓	Median ↓	P95 ↓
GigaPath-RoMa	3.145	1.404	11.433
Global-4096	3.156	1.397	11.501
DeeperHistReg	5.005	1.758	15.708
VALIS [†]	107.861	3.427	485.245
LDDMM	58.431	8.235	380.949

Table 2: Quantitative comparison on HyReCo using the mean, median, P95, and pair-level median TRE. TRE is reported in μm ; lower is better. Pair-med. denotes the median of the per-pair median TRE.

Method	Mean ↓	Median ↓	P95 ↓	Pair-med. ↓
GigaPath-RoMa	18.619	5.919	45.888	6.164
Global-4096	18.921	6.400	46.559	7.170
VALIS	20.769	7.084	48.502	7.258
DeeperHistReg	66.495	8.019	345.926	8.092
LDDMM	460.898	73.507	984.527	62.158

plete both manifests. Table 1 reports rTRE multiplied by 10^3 ; Table 2 reports TRE in μm , with Pair-med. denoting the median per-pair median. Lower is better throughout.

Figure 3 compares three challenging HyReCo H&E-CD45 pairs. GigaPath-RoMa aligns landmarks and local tissue boundaries more closely than VALIS and DeeperHistReg. Missing red landmarks in the VALIS panels lie outside the displayed crops because of large registration errors.

4.5 Ablation Studies

We first study the effect of pathology-domain fine-tuning by comparing the original RoMa checkpoint with the pathology-adapted checkpoint using identical Global-4096 initialization and two-level Pyr2 refinement. Both variants completed all 222 ANHIR and 27 HyReCo pairs. As shown in Table 3, pathology adaptation reduces the mean, median, and P95 errors on ANHIR by 20.2%, 15.7%, and 13.0%, respectively. The improvements on HyReCo are smaller but consistent across all three metrics.

Table 3: Ablation study of pathology-domain fine-tuning.

Dataset	Matcher	Path FT	Mean ↓	Median ↓	P95 ↓
ANHIR	Original RoMa	–	3.942	1.665	13.145
	Path-adapted	✓	3.145	1.404	11.433
HyReCo	Original RoMa	–	18.965	5.929	46.173
	Path-adapted	✓	18.619	5.919	45.888

We next evaluate the number of residual-refinement levels using independently executed Global-4096, Pyr2, and Pyr3 branches initialized by numerically identical Global-4096 mappings. This is an online ablation rather than an offline replay: each finer-level tile location depends on the mapping produced by the preceding level. Table 4 reports complete results on all 222 ANHIR and 27 HyReCo pairs.

Table 4: Ablation study of multiscale refinement configurations. M, F, and C denote mid, fine, and coarse levels.

Dataset	Variant	Levels	Mean ↓	Median ↓	P95 ↓
ANHIR	Global-4096	None	3.156	1.397	11.501
	Pyr2	M+F	3.145	1.404	11.433
	Pyr3	C+M+F	3.146	1.404	11.441
HyReCo	Global-4096	None	18.921	6.400	46.559
	Pyr2	M+F	18.619	5.919	45.888
	Pyr3	C+M+F	18.582	5.893	46.054

HyReCo benefits most from multiscale refinement. Compared with the Global-4096 variant, Pyr2 reduces mean, median, and P95 TRE by 1.6%, 7.5%, and 1.4%, respectively. In contrast, the changes on ANHIR are small and mixed across the three metrics. This difference is likely because HyReCo contains substantially higher-resolution WSIs than ANHIR (median native long side of 220,572 versus 10,987 pixels), making local fine-grained correspondence more important for recovering details that are lost in the global view.

Adding a third refinement level provides little additional benefit. On HyReCo, Pyr3 lowers mean and median TRE by 0.2% and 0.4% but increases P95 by 0.4% relative to Pyr2. On ANHIR, the differences between Pyr2 and Pyr3 remain below 0.1% for all three metrics, and the pairwise comparisons are mixed. The added coarse level therefore provides no consistent accuracy or tail-error benefit across datasets. We select the two-level configuration for the final model because it matches the three-level variant while avoiding an additional full tiling and fusion stage.

5 Conclusion

In this paper, we present GigaPath-RoMa, a pathology-adapted dense correspondence framework for gigapixel cross-stain WSI registration. Synthetic pathology correspondences and quality-aware certainty supervision adapt RoMa to histology, while a global-to-local multiscale strategy enables gigapixel inference. Results on ANHIR and HyReCo show that GigaPath-RoMa outperforms VALIS, DeeperHistReg, and LDDMM-Demons. Global-4096 provides strong alignment, while local refinement is most beneficial for very high-resolution slides, allowing the inference strategy to be selected according to image scale and accuracy requirements. Current limitations include manual quality-control thresholds and the computational cost of multiscale dense matching. Future work will explore adaptive thresholds and schedules, uncertainty-guided tile selection, model compression, topology-preserving constraints, and downstream effects on biomarker analysis and 3D reconstruction.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2024YFA1306600). This work was also supported by the Yuhang District Postdoctoral Research Funding Program, Hangzhou, Zhejiang Province, China.

References

- Avants, B. B.; Epstein, C. L.; Grossman, M.; and Gee, J. C. 2008. Symmetric Diffeomorphic Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly and Neurodegenerative Brain. *Medical Image Analysis*, 12(1): 26–41.
- Awan, R.; Raza, S. E. A.; Lotz, J.; Weiss, N.; and Rajpoot, N. M. 2023. Deep Feature Based Cross-Slide Registration. *Computerized Medical Imaging and Graphics*, 104: 102162.
- Balakrishnan, G.; Zhao, A.; Sabuncu, M. R.; Guttag, J.; and Dalca, A. V. 2018. An Unsupervised Learning Model for Deformable Medical Image Registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 9252–9260.
- Borovec, J.; Kybic, J.; Arganda-Carreras, I.; Sorokin, D. V.; Bueno, G.; Khvostikov, A. V.; Bakas, S.; et al. 2020. AN-HIR: Automatic Non-Rigid Histological Image Registration Challenge. *IEEE Transactions on Medical Imaging*, 39(10): 3042–3052.
- Chen, J.; Frey, E. C.; He, Y.; Segars, W. P.; Li, Y.; and Du, Y. 2022. TransMorph: Transformer for Unsupervised Medical Image Registration. *Medical Image Analysis*, 82: 102615.
- DeTone, D.; Malisiewicz, T.; and Rabinovich, A. 2018. SuperPoint: Self-Supervised Interest Point Detection and Description. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 224–236.
- Edstedt, J.; Athanasiadis, I.; Wadenbäck, M.; and Felsberg, M. 2023. DKM: Dense Kernelized Feature Matching for Geometry Estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17765–17775.
- Edstedt, J.; Sun, Q.; Bökman, G.; Wadenbäck, M.; and Felsberg, M. 2024. RoMa: Robust Dense Feature Matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19790–19800.
- Elhaminia, B.; Alsalemi, A.; Nasir, E.; Jahanifar, M.; Awan, R.; Young, L. S.; Rajpoot, N. M.; Minhas, F.; and Raza, S. E. A. 2025. From Traditional to Deep Learning Approaches in Whole Slide Image Registration: A Methodological Review. *Journal of Pathology Informatics*, 19: 100512.
- Gatenbee, C. D.; Baker, A.-M.; Prabhakaran, S.; Swinyard, O.; Slebos, R. J. C.; et al. 2023. Virtual Alignment of Pathology Image Series for Multi-Gigapixel Whole Slide Images. *Nature Communications*, 14: 4502.
- Ge, L.; Wei, X.; Hao, Y.; Luo, J.; and Xu, Y. 2022. Unsupervised Histological Image Registration Using Structural Feature Guided Convolutional Neural Network. *IEEE Transactions on Medical Imaging*, 41(9): 2414–2431.
- Hoque, M. Z.; Keskinarkaus, A.; Nyberg, P.; Mattila, T.; and Seppänen, T. 2022. Whole Slide Image Registration via Multi-Stained Feature Matching. *Computers in Biology and Medicine*, 144: 105301.
- Jena, R.; Zope, V.; Chaudhari, P.; and Gee, J. 2026. A Scalable Distributed Framework for Multimodal GigaVoxel Image Registration. In *International Conference on Learning Representations*.
- Jiang, J.; Larson, N. B.; Prodduturi, N.; Flotte, T. J.; and Hart, S. N. 2019. Robust Hierarchical Density Estimation and Regression for Re-Stained Histological Whole Slide Image Co-Registration. *PLOS ONE*, 14(7): e0220074.
- Klein, S.; Staring, M.; Murphy, K.; Viergever, M. A.; and Pluim, J. P. W. 2010. Elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Transactions on Medical Imaging*, 29(1): 196–205.
- Komura, D.; Kawabe, A.; Fukuta, K.; Sano, K.; Umezaki, T.; Koda, H.; Suzuki, R.; Tominaga, K.; Ochi, M.; Konishi, H.; Masakado, F.; Saito, N.; Sato, Y.; Onoyama, T.; Nishida, S.; Furuya, G.; Katoh, H.; Yamashita, H.; Kakimi, K.; Seto, Y.; Ushiku, T.; Fukayama, M.; and Ishikawa, S. 2022. Universal Encoding of Pan-Cancer Histology by Deep Texture Representations. *Cell Reports*, 38(9): 110424.
- Lindenberger, P.; Sarlin, P.-E.; and Pollefeys, M. 2023. Light-Glue: Local Feature Matching at Light Speed. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 17627–17638.
- Lotz, J.; Olesch, J.; Müller, B.; Polzin, T.; Galuschka, P.; Lotz, J. M.; Heldmann, S.; Laue, H.; González-Vallinas, M.; Warth, A.; Lahrmann, B.; Grabe, N.; Sedlacek, O.; Breuhahn, K.; and Modersitzki, J. 2016. Patch-Based Nonlinear Image Registration for Gigapixel Whole Slide Images. *IEEE Transactions on Biomedical Engineering*, 63(9): 1812–1819.
- Lowe, D. G. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2): 91–110.
- Paknezhad, M.; Loh, S. Y. M.; Choudhury, Y.; Koh, V. K. C.; Yong, T. T. K.; Tan, H. S.; Kanesvaran, R.; Tan, P. H.; Peng, J. Y. S.; Yu, W.; et al. 2020. Regional Registration of Whole Slide Image Stacks Containing Major Histological Artifacts. *BMC Bioinformatics*, 21: 558.
- Rossetti, B. J.; Wang, F.; Zhang, P.; Teodoro, G.; Brat, D. J.; and Kong, J. 2017. Dynamic Registration for Gigapixel Serial Whole Slide Images. In *2017 IEEE 14th International Symposium on Biomedical Imaging*, 424–428.
- Roy, M.; Wang, F.; Teodoro, G.; Bhattarai, S.; Bhargava, M.; Rekha, T. S.; Aneja, R.; and Kong, J. 2023. Deep Learning Based Registration of Serial Whole-Slide Histopathology Images in Different Stains. *Journal of Pathology Informatics*, 14: 100311.
- Rueckert, D.; Sonoda, L. I.; Hayes, C.; Hill, D. L. G.; Leach, M. O.; and Hawkes, D. J. 1999. Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images. *IEEE Transactions on Medical Imaging*, 18(8): 712–721.
- Sarlin, P.-E.; DeTone, D.; Malisiewicz, T.; and Rabinovich, A. 2020. SuperGlue: Learning Feature Matching with Graph

Neural Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4938–4947.

Shen, T.; Yan, F.; and Zhang, X. 2026. SAR2Net: Learning Spatially Anchored Representations for Retrieval-Guided Cross-Stain Alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12544–12553.

Sun, J.; Shen, Z.; Wang, Y.; Bao, H.; and Zhou, X. 2021. LoFTR: Detector-Free Local Feature Matching with Transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8922–8931.

Thirion, J.-P. 1998. Image Matching as a Diffusion Process: An Analogy with Maxwell’s Demons. *Medical Image Analysis*, 2(3): 243–260.

Wang, Y.; He, Z.; and Yan, Y. 2026. SPOMiAlign: A Modality-Agnostic Computational Framework for Multi-modal Spatial Omics Alignment Enabled by a Feature Matching Foundation Model. *Briefings in Bioinformatics*, 27(3): bbag331.

Weitz, P.; Valkonen, M.; Solorzano, L.; Carr, C.; Kartasalo, K.; Boissin, C.; Koivukoski, S.; Kuusela, A.; Rasic, D.; Feng, Y.; et al. 2024. The ACROBAT 2022 Challenge: Automatic Registration of Breast Cancer Tissue. *Medical Image Analysis*, 97: 103257.

Wodzinski, M.; Marini, N.; Atzori, M.; and Müller, H. 2024a. DeeperHistReg: Robust Whole Slide Images Registration Framework. *arXiv:2404.14434*.

Wodzinski, M.; Marini, N.; Atzori, M.; and Müller, H. 2024b. RegWSI: Whole Slide Image Registration Using Combined Deep Feature- and Intensity-Based Methods: Winner of the ACROBAT 2023 Challenge. *Computer Methods and Programs in Biomedicine*, 250: 108187.

Wodzinski, M.; and Müller, H. 2021. DeepHistReg: Unsupervised Deep Learning Registration Framework for Differently Stained Histology Samples. *Computer Methods and Programs in Biomedicine*, 198: 105799.

Zhou, S.; Xiong, Z.; and Wu, F. 2024. Test-Time Adaptation via Style and Structure Guidance for Histological Image Registration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 7677–7685.