Test-Time Adaptation via Style and Structure Guidance for Histological Image Registration

Shenglong Zhou[1], **Zhiwei Xiong**[1, 2*], **Feng Wu**[1, 2]

[1] University of Science and Technology of China
[2] Institute of Artificial Intelligence, Hefei Comprehensive National Science Center
slzhou96@mail.ustc.edu.cn, zwxiong@ustc.edu.cn, fengwu@ustc.edu.cn

Abstract

Image registration plays a crucial role in histological image analysis, encompassing tasks like multi-modality fusion and disease grading. Traditional registration methods optimize objective functions for each image pair, yielding reliable accuracy but demanding heavy inference burdens. Recently, learning-based registration methods utilize networks to learn the optimization process during training and apply a one-step forward process during testing. While these methods offer promising registration performance with reduced inference time, they remain sensitive to appearance variances and local structure changes commonly encountered in histological image registration scenarios. In this paper, for the first time, we propose a novel test-time adaptation method for histological image registration, aiming to improve the generalization ability of learning-based methods. Specifically, we design two operations, style guidance and shape guidance, for the test-time adaptation process. The former leverages style representations encoded by feature statistics to address the issue of appearance variances, while the latter incorporates shape representations encoded by HOG features to improve registration accuracy in regions with structural changes. Furthermore, we consider the continuity of the model during the test-time adaptation process. Different from the previous methods initialized by a given trained model, we introduce a smoothing strategy to leverage historical models for better generalization. We conduct experiments with several representative learning-based backbones on the public histological dataset, demonstrating the superior registration performance of our test-time adaptation method.

Introduction

Image registration is an important task in computer vision, particularly within the realm of medical image analysis (Chakravarty et al. 2006; Li et al. 2022; Chen et al. 2023). The goal of image registration is to establish a transformation that aligns a pair of images (i.e., a fixed image and a moving image), thereby enabling a wide range of clinical applications. In the field of histological image analysis, the utilization of various stains during histology sample preparation can offer valuable information, and each stain reveals distinct tissue properties. Their fusion can benefit tasks such as grading, classification, and 3D reconstruction. However, different preparation processes and the use of consecutive tissue slices introduce complex and inevitable deformations. Therefore, non-rigid registration becomes essential to facilitate further processing.

Traditional methods solve the registration task by formulating it as an optimization problem for each image pair. Though traditional methods provide reliable registration accuracy, one obvious limitation is that the optimization can be computationally expensive. Recently, learning-based methods (Dalca et al. 2018; Mok and Chung 2020; Hu et al. 2022a; Zhou et al. 2023; Liu et al. 2023) utilize networks to learn the optimization process from the training image pairs, thus regarding the registration task as a mapping from an image pair to a deformation field during testing. Along with the development of network structures from direct designs such as U-Net (Dalca et al. 2018; Zhou et al. 2019), to progressive designs such as Pyramid (Hu et al. 2020; Mok and Chung 2020) and Cascade (Zhao et al. 2019a; Hu et al. 2022b), learning-based methods achieve a promising registration performance with a reduced inference burden.

In the field of histological image registration, learning-based methods also attract research attention. As mentioned in (Borovec et al. 2020), TUB proposes a supervised convolutional neural network that relies on manually labeled key points. DeepHistReg (Wodzinski and Müller 2021) proposes an unsupervised method by designing a pyramid-based non-rigid registration network, which does not demand any manual annotations. It is worth mentioning that, different from biomedical datasets such as MRI or CT, the registration of histological images usually suffers from the following challenges as shown in Fig. 1. First, there are appearance variances between histological images due to multiple stains. These appearance variances not only exist between the fixed image and the moving image, but also between training images and test images, even with pre-processes such as gray translation or color normalization. Such appearance variances can significantly affect the robustness and generalization of learning-based registration methods. Second, there are local structure changes such as repetitive textures and missing sections in histological images. Therefore, learning-based methods are hard to learn and may perform poorly when encountering changed structures in test images. How to design robust and generalizable learning-based registra-

*Corresponding Author

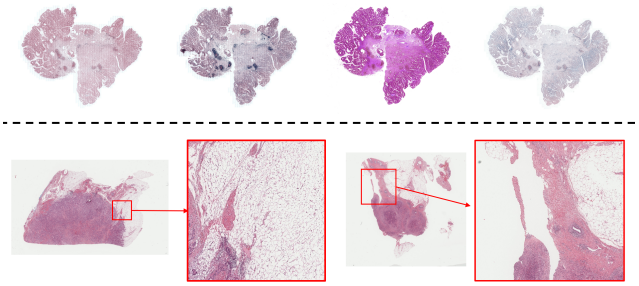

Figure 1: Challenges of the histological image registration. Above are the appearance variances due to multiple staining, and below are the local structure changes.

tion methods to solve the above challenges is an important problem. SFG (Ge et al. 2022) proposes to introduce dense SIFT features and automatic key points for histological image registration, aiming to handle the challenges of structural features. But this method still focuses on the training time and ignores the model's potential during the test time. Different from previous methods, we explore the test-time adaptation for histological image registration, aiming to improve the generalization and robustness.

Test-time adaptation is a useful approach to improving the model's generalization ability, which has been explored in several computer vision tasks including classification (Sun et al. 2020; Wang et al. 2020) and segmentation (Hu et al. 2021). In terms of medical image registration, test-time adaptation further tunes the given trained model on each test image pair, which can be regarded as seeking a middle ground (Zhu et al. 2021) between traditional optimization-based methods and pure learning-based methods. SSMSR (Zhu et al. 2021) introduces the test-time adaptation for MRI/echocardiogram registration with a straightforward multi-scale design. But this method does not consider the challenges of histological image registration, and thus cannot handle appearance variances and local structure changes well. Meanwhile, SSMSR ignores the continuity of models during the test-time adaptation process, thus restricting further improvement of registration performance.

In this paper, for the first time, we propose a novel test-time adaptation method for histological image registration, named SGTTA. We design two operations, style guidance and structure guidance, for solving challenges in histological images. Style guidance aims to handle appearance variances, and the core idea is to transfer the style representation from training images to test images, which can help narrow the style gap between them. Specially, we extract the features from the encoder branch of the trained model, then calculate the statistics (i.e., mean and standard deviation values) of features by instance normalization (IN), and regard them as the style representation. When conducting the test-time adaptation for image pairs, style guidance combines the style representation from training images with the test images' features by adaptive instance adaptation (AdaIN (Huang and Belongie 2017)). It is worth mentioning that, feature statistics from training images do not leak the complete information of images (raw images cannot be recovered from the statistics), which can protect the privacy of biomedical datasets. Structure guidance aims to handle local structure changes, and the core idea is to introduce the structural constraints when conducting the test-time adaptation. We utilize the HOG (Dalal and Triggs 2005) descriptors as the structure representation for each test image pair. Then structure guidance constrains the similarity of HOG descriptors between the fixed image and the warped moving image.

Furthermore, we consider the continuity of models during the test-time adaptation process. Different from the previous methods initialized by a given trained model for each test image pair, we introduce a smoothing strategy to leverage historical models. The smoothing strategy combines the model from the last test image pair and the given trained model. The model from the last test image pair contains the learned parameters in the test domain, and the given trained model provides a strong registration ability, which can decrease error accumulation and catastrophic forgetting. Therefore, the combination of them can obtain better generalization performance. To evaluate the effectiveness of SGTTA, we conduct comprehensive experiments on the public histological dataset with representative learning-based backbones including U-Net, Pyramid, and Cascade. Both quantitative and qualitative results demonstrate the superior performance of our SGTTA. We summarize the main contributions as follows:

- We explore the test-time adaptation for histological image registration for the first time, aiming to improve the generalization of learning-based methods.
- We propose a novel test-time adaptation method, by designing style guidance and structure guidance to handle appearance variances and local structure changes.
- We introduce a smoothing strategy to leverage historical models, considering the continuity of models during the test-time adaptation process.
- We conduct comprehensive experiments on the public histological dataset with representative backbones (U-Net, Pyramid, and Cascade), demonstrating better registration performance of our SGTTA.

Related Work

Biomedical Image Registration

Traditional methods solve the registration task by formulating it as an optimization problem for each image pair. Numerous traditional methods have been developed for non-rigid image registration, including B-spline deformation-based methods (Song et al. 2013), elastic deformation-based model (Du Bois d'Aische et al. 2005), large deformation diffeomorphic metric image matching algorithm (Ceritoglu et al. 2010), and greedy diffeomorphic algorithm (Venet et al. 2021). ANHIR (Borovec et al. 2020) also describes many traditional methods for histological image registration.

Recently, deep neural networks have been applied to biomedical image registration (Hu et al. 2022a). VoxelMorph (Balakrishnan et al. 2018) adopts U-Net to generate the deformation field directly, which saves considerable

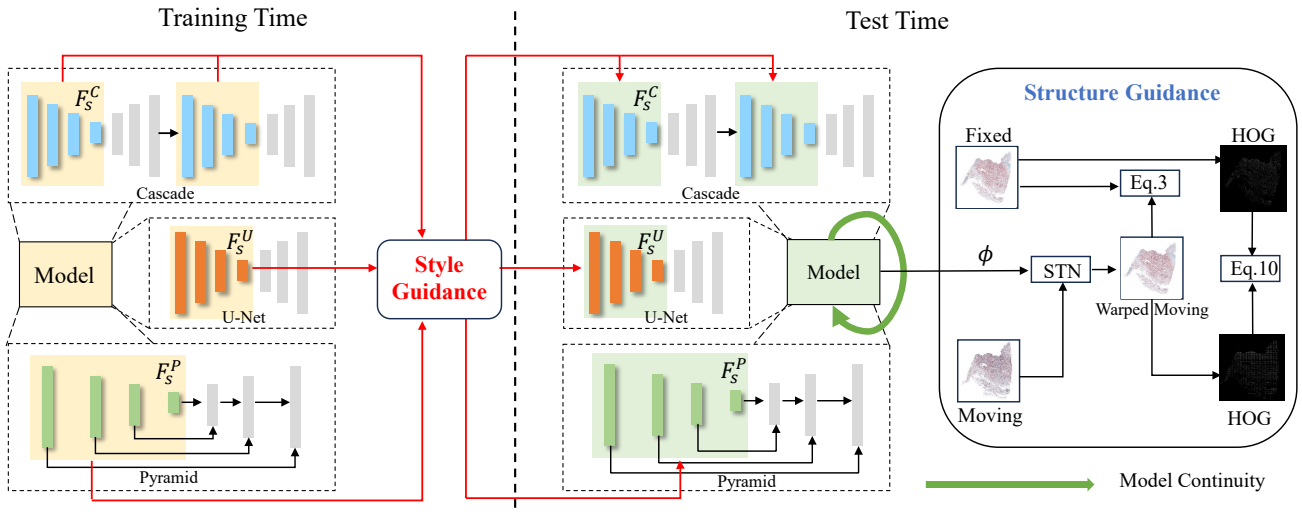

Figure 2: Overview of our SGTTA, which consists of style guidance, structure guidance, and model continuity.

inference time compared with traditional methods. Dual-PRNet (Hu et al. 2019) proposes a dual-stream pyramid structure to generate the deformation field in a coarse-to-fine manner, and LapIRN (Mok and Chung 2020) adopts an image Laplacian pyramid to generate and refine deformation fields. Recursive cascaded network (Zhao et al. 2019a) takes U-shape networks as its sub-networks, and analyzes the effect in different cascading stages. For histological image registration, Pyramid structures (Ge et al. 2022; Wodzinski and Müller 2021) and Cascade structures (Borovec et al. 2020) have also been adopted. But these methods focus on the training time, and we explore the test-time adaptation for histological image registration for the first time.

Test-time Adaptation

Test-time adaptation methods can update a model with the distributional information provided by a single or batch of test data. TTT (Sun et al. 2020) adapts the feature extractor at test time by leveraging an auxiliary self-supervised task of rotation prediction. TTT++ (Liu et al. 2021) improves TTT by further aligning the first- and second-order statistics of the training and test data. Tent (Wang et al. 2020) proposes to adapt the affine parameters in batch normalization layers at test time by minimizing the entropy of model predictions. T3A (Iwasawa and Matsuo 2021) adjusts the classifier of a trained source model by computing a pseudo-prototype representation of different classes using unlabeled test data. In the field of image registration, SSMSR (Zhu et al. 2021) introduces the test-time adaptation for MRI/echocardiogram registration with a multi-scale design. Differently, we propose a novel test-time adaptation method for solving challenges in histological image registration.

Methodology

Preliminaries and Notations

Given a pair of histological images as the fixed image $\boldsymbol{I}_f$ and the moving image $\boldsymbol{I}_m$, nonrigid registration aims to obtain the deformation field ϕ. The warped moving image $\boldsymbol{I}_m(\phi)$ is aligned to $\boldsymbol{I}_f$ by the deformation field ϕ. According to (Balakrishnan et al. 2019), the image registration problem can be formulated as the minimization of differences between $\boldsymbol{I}_m(\phi)$ and $\boldsymbol{I}_f$, which is subject to a smoothness constraint on the deformation field. The specific formula is shown as

$$\hat{\phi} = \arg\min_{\phi} \mathcal{L}_S(\boldsymbol{I}, \boldsymbol{I}_m(\phi)) + \lambda \mathcal{L}_R(\phi), \tag{1}$$

where $\mathcal{L}_S$ is a reconstruction loss measuring the dissimilarity between two images, $\mathcal{L}_R$ constrains the smoothness of the deformation field, and λ is a regularization parameter balancing the trade-off between the reconstruction and smoothness losses. The smoothness term is defined as

$$\mathcal{L}_R(\phi) = \sum_{\mathbf{p}\in\Omega} \sum_{i=1}^{n} \|\nabla \phi_i(\mathbf{p})\|^2, \tag{2}$$

where $\mathbf{p}$ is the coordinate, n is the number of pixels, ∇ is the spatial gradients, and Ω is the neighbouring region.

Learning-Based Registration

Following common learning-based registration methods, we model the deformation field ϕ through a network $\mathcal{N}$ with learnable parameters $\boldsymbol{\theta}$, which receives $\boldsymbol{I}_f$ and $\boldsymbol{I}_m$ as input and generates ϕ as output. The whole process can be formulated as $\phi = \mathcal{N}(\boldsymbol{I}_f, \boldsymbol{I}_m; \boldsymbol{\theta})$. Therefore, the determination of the deformation field is treated as a learning problem, seeking to identify the optimal parameters $\boldsymbol{\theta}$ that minimize the loss function presented in Eq. (1). By the way, many metrics can be used to measure the dissimilarity $\mathcal{L}_S$ in Eq. (1), and we choose the negative normalized local cross-correlation as the reconstruction loss in this paper.

We choose several advanced registration backbones as the network $\mathcal{N}$, including U-Net, Pyramid, and Cascade. For the U-Net backbone, we follow the design in VoxelMorph (Dalca et al. 2018). U-Net has the encoder part and the decoder part, where the encoder part generates four features

$\{\boldsymbol{F}_s^u\}_{s=1}^4$ according to the spatial scale. For the Pyramid backbone, we follow the design in DeepHistReg (Wodzinski and Müller 2021) and RDN (Hu et al. 2022b) roughly. Pyramid also has the encoder part and the decoder part, while the decoder part generates the deformation fields in a coarse-to-fine manner. For consistency, we design the encoder in Pyramid to generate four features $\{\boldsymbol{F}_s^p\}_{s=1}^4$ according to the spatial scale. For the Cascade backbone, we follow the design in RDN (Hu et al. 2022b) and RCN (Zhao et al. 2019a) roughly, which generates the deformation field in a recursive manner. We denote the features in encoder of Cascade as $\{\boldsymbol{F}_s^c\}_{s=1}^4$, and omit different subnetworks for convenience.

Test-Time Adaptation for Registration

Given the training dataset including image pairs $\{\boldsymbol{I}_f^{tr}, \boldsymbol{I}_m^{tr}\}$, we first learn the model's parameters by minimizing the loss function Eq. 1 and obtain the given trained model $\boldsymbol{\theta}^{tr}$. In the common paradigm, we can directly apply $\boldsymbol{\theta}^{tr}$ on the test dataset with a one-step forward process. The benefit of this paradigm is efficient inference but at the cost of performance drop. The main reason for this is the inherent characteristics of each image pair, posing challenges for learned models to generalize effectively to new test images. Particularly, when dealing with histological images with appearance variances and structural changes, a notable performance gap between the training and test datasets becomes obvious.

We introduce test-time adaptation to improve the generalization of learning-based registration. Under this paradigm, given the test image pair $\{\boldsymbol{I}_f^{te}, \boldsymbol{I}_m^{te}\}$, the network parameters $\boldsymbol{\theta}^{te}$ are initialized by $\boldsymbol{\theta}^{tr}$ and further optimized as

$$\boldsymbol{\theta}^{te} = \arg\min_{\boldsymbol{\theta}} \mathbb{E}\left[\mathcal{L}_S(\boldsymbol{I}_f^{te}, \boldsymbol{I}_m^{te}(\boldsymbol{\phi}); \boldsymbol{\theta}) + \lambda \mathcal{L}_R(\boldsymbol{\phi}; \boldsymbol{\theta})\right]. \quad (3)$$

Test-time adaptation not only alleviates the drawbacks in traditional registration methods including high cost in optimization, long running time, and poor performance due to local optimality (KingmaandJ 2015), but also improves the performance of pure learning-based methods by further adapting on test image pairs.

Style Guidance for Test-Time Adaptation

Appearance variances usually exist in histological image registration, so we propose style guidance to solve this challenge during the test-time adaptation process. The core idea is to transfer the style representation from training images to test images, which can help narrow the style gap between images. How to define the style representation is the first question. In the field of style transfer, it is well-known that convolutional feature statistics can represent the style information of an image, such as channel-wise mean and variance (Gatys, Ecker, and Bethge 2016). Following (Ulyanov, Vedaldi, and Lempitsky 2017), image style can be removed by instance normalization (IN). For an image $\boldsymbol{I}$, the feature of $\boldsymbol{I}$ can be defined as $\boldsymbol{F} \in \mathbb{R}^{C\times H\times W}$, where H and W are spatial dimensions, and C is the number of channels. Therefore, IN can be formulated as:

$$\mathrm{IN}(\boldsymbol{F}) = \gamma \frac{\boldsymbol{F} - \boldsymbol{\mu}}{\boldsymbol{\sigma}} + \beta \quad (4)$$

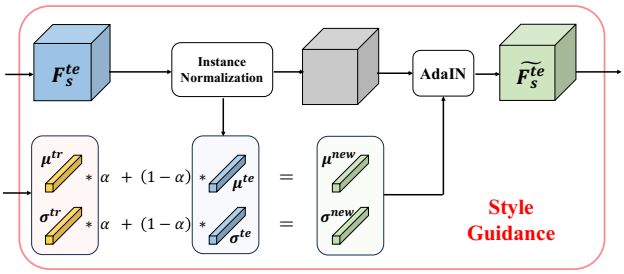

Figure 3: The illustration of style guidance.

where γ, β are learnable affine transformation parameters, and $\boldsymbol{\mu}, \boldsymbol{\sigma} \in \mathbb{R}^C$ are the channel-wise mean and standard deviation of feature map calculated as

$$\boldsymbol{\mu} = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\boldsymbol{F}_{chw}, \quad (5)$$

$$\boldsymbol{\sigma} = \sqrt{\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}(\boldsymbol{F}_{chw} - \boldsymbol{\mu})^2 + \epsilon}, \quad (6)$$

where ϵ is a constant for numerical stability. Inspired by the above style transfer designs, we choose feature statistics (mean and standard deviation) as the style representation. Moreover, AdaIN is proposed to convert one image style to another one, which replaces the affine parameters by the specific style statistics $(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\sigma}})$. The formulation of AdaIN is defined as

$$\mathrm{AdaIN}(\boldsymbol{F}, (\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\sigma}})) = \tilde{\boldsymbol{\sigma}}\frac{\boldsymbol{F} - \boldsymbol{\mu}}{\boldsymbol{\sigma}} + \tilde{\boldsymbol{\mu}}. \quad (7)$$

Considering the simplicity, we choose the AdaIN to transfer the style representation from training images to test images.

Specically, given the trained model $\boldsymbol{\theta}^{tr}$, we use it to extract the features of the encoder part on the training dataset. In fact, there are different types of encoded features based on different backbones, $\boldsymbol{F}_s^{tr(u)}/\boldsymbol{F}_s^{tr(p)}/\boldsymbol{F}_s^{tr(c)}$, and we denote them as $\boldsymbol{F}_s^{tr}$ for convenience. Then we calculate the feature statistics (mean and standard deviation) of features following IN, and average them as the style representation $(\boldsymbol{\mu}^{tr}, \boldsymbol{\sigma}^{tr})$. During the test-time adaptation process, given one test image pair, we extract the encoded features $\boldsymbol{F}_s^{te}$ and calculate the style representation $(\boldsymbol{\mu}^{te}, \boldsymbol{\sigma}^{te})$ in a similar way. Style guidance combines the test style representation with ones from the training dataset, aiming to transfer the style representations from training images. The combination is shown as

$$\boldsymbol{\mu}^{new} = \alpha\boldsymbol{\mu}^{tr} + (1-\alpha)\boldsymbol{\mu}^{te}, \quad (8)$$

$$\boldsymbol{\sigma}^{new} = \alpha\boldsymbol{\sigma}^{tr} + (1-\alpha)\boldsymbol{\sigma}^{te}, \quad (9)$$

where α is the hyperparameter to control the style transfer. Finally, we utilize the AdaIN to obtain the new encoded features $\tilde{\boldsymbol{F}}_s^{te} = \mathrm{AdaIN}\left(\boldsymbol{F}^{te}, (\boldsymbol{\mu}^{new}, \boldsymbol{\sigma}^{new})\right)$, as shown in figure 3. After that, new encoded features are passed into the next encoder block or decoder part.

Structure Guidance for Test-Time Adaptation

Local structure changes usually exist in histological image registration, so we propose structure guidance to solve this challenge during the test-time adaptation process. As discussed in (Ge et al. 2022), structural representations are robust to staining images and retaining the anatomical structures is helpful for diagnosis in histological images (Miranda et al. 2012). The core idea of structure guidance is to introduce the structural constraints when conducting the test-time adaptation. The first question is to determine the structure representations. In fact, there are many structure-based descriptions such as SIFT (Lowe 1999), shape context (Belongie, Malik, and Puzicha 2002), and HOG (Dalal and Triggs 2005). Considering the efficiency and lightweight during the test-time adaptation process, we choose the HOG as the structure representation.

Specifically, given the test image pair $\{\boldsymbol{I}_f^{te}, \boldsymbol{I}_m^{te}\}$, we can obtain the deformation field $\boldsymbol{\phi}$ by the given model. According to the previous statements, we use Eq. 3 to find the optimal parameters. Here, we add another constraint based on the structure representations. We extract structure representations of the fixed image $\boldsymbol{I}_f^{te}$ as $\boldsymbol{H}_f^{te}$, along with the representations of the warped moving image $\boldsymbol{I}_m^{te}(\boldsymbol{\phi})$ as $\boldsymbol{H}_m^{te}(\boldsymbol{\phi})$. Furthermore, considering the complicated local structures, we provide multi-scale structure representations for better performance. To obtain the multi-scale representations, we downsample the images into several scales by bilinear interpolation and extract HOG descriptors in each scale. We can regard the multi-scale structure representations of the fixed image $\{\boldsymbol{H}_{f(s)}^{te}\}_{s=1}^{4}$ and the warped moving image $\{\boldsymbol{H}_{m(s)}^{te}(\boldsymbol{\phi})\}_{s=1}^{4}$ as the structure guidance for the constraint during the test-time adaptation process. The constraint can be formulated as

$$\mathcal{L}_{sg} = - \frac{\left(\left(\boldsymbol{H}_{f(s)}^{te} - \overline{\boldsymbol{H}_{f(s)}^{te}}\right) \cdot \left(\boldsymbol{H}_{m(s)}^{te}(\boldsymbol{\phi}) - \overline{\boldsymbol{H}_{m(s)}^{te}(\boldsymbol{\phi})}\right)\right)^2}{\left(\boldsymbol{H}_{f(s)}^{te} - \overline{\boldsymbol{H}_{f(s)}^{te}}\right)^2 \cdot \left(\boldsymbol{H}_{m(s)}^{te}(\boldsymbol{\phi}) - \overline{\boldsymbol{H}_{m(s)}^{te}(\boldsymbol{\phi})}\right)^2}, \tag{10}$$

where $\bar{*}$ means the local means operation.

Model Continuity during Test-time Adaptation

We further consider the continuity of the model during the test-time adaptation process. Given the test image pair $\{\boldsymbol{I}_f^{te}, \boldsymbol{I}_m^{te}\}$ at time point t (it is reasonable to assume that pairs of images appear sequentially over time), we define three models to present our smoothing strategy. First, we define $\boldsymbol{\theta}_t^{'}$ as the historical model, then we define $\boldsymbol{\theta}_t^{te*}$ as the initialization model, finally we define $\boldsymbol{\theta}_t^{te}$ as the obtained model after the test-time adaptation process. To consider the model's continuity, we propose a smoothing strategy to leverage historical models as

$$\boldsymbol{\theta}_{t+1}^{'} = w\boldsymbol{\theta}_t^{'} + (1-w)\boldsymbol{\theta}_t^{te}, \tag{11}$$

where w is a smoothing factor. For the next time point $t+1$, the previous method just applies the given trained model $\boldsymbol{\theta}^{tr}$ as the initialized model $\boldsymbol{\theta}_{t+1}^{te*} = \boldsymbol{\theta}^{tr}$. Differently, we initialize the model with the combination of the given trained model and the continuous model as $\boldsymbol{\theta}_{t+1}^{te*} = k\boldsymbol{\theta}^{tr} + (1-k)\boldsymbol{\theta}_{t+1}^{'}$. Then, starting from the $\boldsymbol{\theta}_{t+1}^{te*}$, the model $\boldsymbol{\theta}_{t+1}^{te}$ can be obtained after test-time adaptation process.

Experiments

Datasets and Metrics. We conduct experiments on the public histological dataset ANHIR for comparison. SGTTA focuses on the test-time stage and is complementary to learning-based methods in the training stage, so it is feasible for us to choose the dataset for evaluation. For fair comparison and efficient experiments, we use the images containing public landmarks in the raw ANHIR dataset as our dataset, and we split them into 115 pairs for training and 115 pairs for testing. The public landmarks represent obvious structures in the images. Based on landmarks, we use the relative target registration error of landmarks (rTRE) for each pair of images as the evaluation metric. We calculate the median, average, and maximum of all rTRE values in an image pair. At the case level, there is aggregation by the median or the average. The metrics are median-median rTRE (MMrTRE), average-median rTRE (AMrTRE), median-average rTRE (MArTRE), average-average rTRE (AArTRE), median-maximum rTRE (MMxrTRE) and average-maximum rTRE (AMxrTRE). Robustness is evaluated by the relative number of successfully registered landmarks.

Baseline Methods. First, we implement six traditional methods as our main comparison methods. Following the guidance from ANHIR, we implement bUnwarpJ (Arganda-Carreras et al. 2006), RVSS (Arganda-Carreras et al. 2006), NiftyReg (Rueckert et al. 1999), Elastix (Klein et al. 2009), ANTs (Avants et al. 2008), and DROP (Glocker et al. 2011). Second, considering that SGTTA is complementary to learning-based methods, we implement three advanced registration backbones (U-Net, Pyramid, and Cascade) as the comparison methods. Specifically, as mentioned before, we denote the U-Net backbone as HistRegU. For the Pyramid backbone, we follow the design in DeepHistReg (Wodzinski and Müller 2021) and RDN (Hu et al. 2022b) roughly and denote it as HistRegP. For the Cascade backbone, we follow the design in VTN (Zhao et al. 2019b) and RCN (Zhao et al. 2019a) roughly and denote it as HistRegC.

Implementation. All the learning-based methods are implemented on PyTorch on 4 cards of NVIDIA TITAN XP. We apply the same pre-processing step and follow (Wodzinski and Müller 2021) to conduct the rotation prediction and affine registration, so we focus on the nonrigid registration problem. For a fair comparison, we do not use any pre-trained models and apply the same training schedule for all learning-based methods. Specifically, during the training stage, we set the batch size as 1 and the number of epochs as 100. We set the regularization parameter λ as 30 following (Wodzinski and Müller 2021). For other hyperparameters, we set α as 0.2, w as 0.99, and k as 0.8 empirically.

Method	Average rTRE		Median rTRE		Max rTRE		Robustness		Time [min]
	Average	Median	Average	Median	Average	Median	Average	Median	Average
bUnwarpJ	0.0472	0.0193	0.0463	0.0192	0.1035	0.0524	0.7866	0.9525	10.57
RVSS	0.0269	0.0107	0.0278	0.0087	0.0648	0.0394	0.8455	1.0000	5.25
NiftyReg	0.0433	0.0243	0.0434	0.0237	0.1097	0.0502	0.7624	0.8974	0.14
Elastix	0.0411	0.0144	0.0374	0.0094	0.0898	0.0418	0.8698	0.9866	3.50
ANTs	0.0397	0.0132	0.0391	0.0096	0.0872	0.0422	0.7998	0.9882	48.24
DROP	0.0325	0.0092	0.0327	0.0057	0.0804	0.0387	0.8971	1.0000	3.99
HistRegU	0.0254	0.0084	0.0257	0.0062	0.0784	0.0327	0.9534	0.9948	0.01
HistRegU + SGTTA	**0.0196**	**0.0051**	**0.0189**	**0.0037**	**0.0509**	**0.0237**	**0.9820**	**1.0000**	0.12
HistRegC	0.0223	0.0072	0.0214	0.0041	0.0611	0.0254	0.9647	**1.0000**	0.03
HistRegC + SGTTA	**0.0174**	**0.0041**	**0.0155**	**0.0026**	**0.0484**	**0.0203**	**0.9812**	**1.0000**	0.67
HistRegP	0.0207	0.0067	0.0223	0.0039	0.0545	0.0258	0.9748	**1.0000**	0.02
HistRegP + SGTTA	**0.0161**	**0.0032**	**0.0149**	**0.0022**	**0.0467**	**0.0217**	**0.9823**	**1.0000**	0.39

Table 1: Comparison results with the traditional methods, U-Net learning-based method (HistRegU), Pyramid learning-based method (HistRegP), Cascade learning-based method (HistRegC), and our SGTTA.

Results

Comparison with Baseline Methods

SGTTA is complementary to learning-based methods, so we apply our method for all three learning-based methods. We do not modify any training details about them, and tune the model on the test datasets using our method. According to the results in Table 1, SGTTA consistently boosts the performance of learning-based methods. Specifically, comparing HistRegU+SGTTA with HistRegU, all the metrics are improved, indicating the better registration performance of SGTTA. In terms of robustness, SGTTA improves both average and median robustness which verifies SGTTA improves the generalization ability of the learning-based method.

Compared with HistRegU, HistRegP and HistRegC achieves better registration performance due to the inherent decomposition (Hu et al. 2022b). Even though HistRegP and HistRegC have a higher start point, SGTTA still improves them with an obvious gain. Specifically, SGTTA improves HistRegC from 0.0214 to 0.0155 and HistRegP from 0.023 to 0.0149 in terms of AMrTRE. Meanwhile, AmaxrTRE and MMaxrTRE are improved obviously for both HistRegC and HistRegP through SGTTA. Though HistRegC and HistRegP have achieved a 1.0 median robustness, SGTTA still improves the average robustness from 0.9647 to 0.9812 or from 0.9748 to 0.9823, which is promising for clinic scenarios. By the way, the learning-based method equipped with SGTTA shows a large advantage in registration accuracy compared with all traditional methods. Due to the further optimization of the network parameters, the inference time of SGTTA is relatively longer. But in fact, SGTTA is faster than most traditional methods, while the performance is obviously better. Also, compared with pure learning-based methods, the increased inference time is acceptable considering the improvement of registration performance. Moreover, We conduct the statistical significance test, demonstrating the significant improvements by SGTTAs, where HistRegU+SGTTA outperforms HistRegU with p-values below 5e-4 for AArTRE and 5e-3 for MArTRE.

Components			Metrics	
Style	Structure	Continuity	AArTRE	MArTRE
			0.0254	0.0084
✓			0.0231	0.0071
✓	✓		0.0209	0.0059
✓	✓	✓	0.0196	0.0051

Table 2: Ablation studies of SGTTA about network components. Style is style guidance, Structure is structure guidance, and continuity is model continuity.

Visualization Comparison

We take different images as examples to show the visualization quality of SGTTA in Fig 4. In the image comparisons, we depict the distances between landmarks, and we observed that SGTTA results in better landmark alignment in most regions. This finding indicates that SGTTA can further enhance the accuracy of image structure registration, thereby demonstrating its effectiveness in improving the generalization capability of learning-based methods.

Ablation Studies

In this section, we delve deep into the effect of our design in SGTTA. We take HistRegU as an example where the impacts are consistent with the other two backbones, and we use AArTRE and MArTRE as the metrics.

Effect of Components. We conduct ablation experiments to verify the effectiveness of our design in SGTTA, including style guidance, structure guidance, and model continuity. The detailed results are shown in Table 2. All three designs improve registration accuracy. Specifically, style guid-

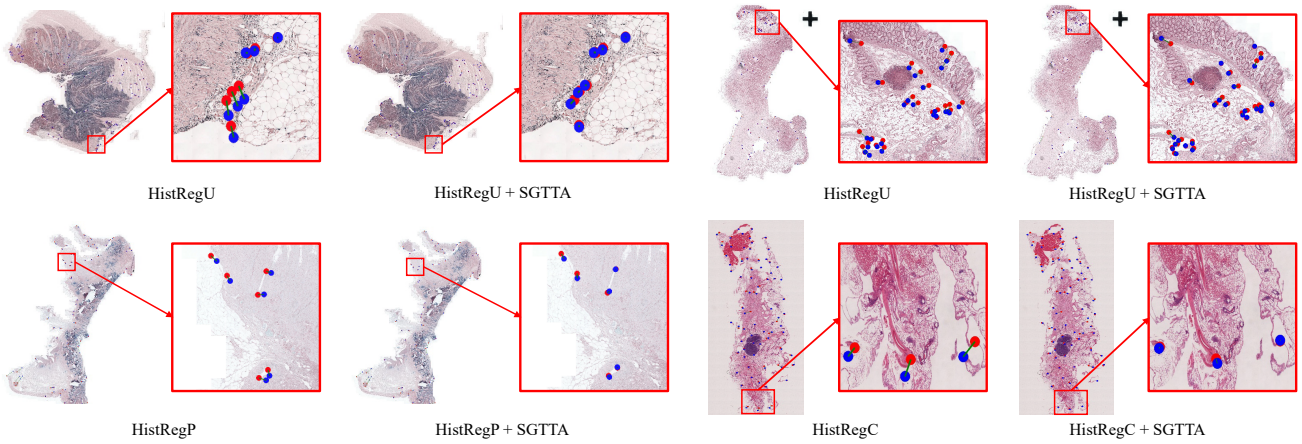

Figure 4: Visualization results of our SGTTA with pure learning-based methods in ANHIR dataset.

Position				Metric	
1st	2nd	3rd	4th	AArTRE	MArTRE
✓				0.0207	0.0062
✓	✓			0.0203	0.0057
✓	✓	✓		0.0198	0.0053
✓	✓	✓	✓	0.0196	0.0051

Table 3: Ablation studies of SGTTA about encoded features' positions in style guidance

Scale				Metric	
1	1/2	1/4	1/8	AArTRE	MArTRE
✓				0.0219	0.0071
✓	✓			0.0206	0.0059
✓	✓	✓		0.0199	0.0055
✓	✓	✓	✓	0.0196	0.0051

Table 4: Ablation studies of SGTTA about HOG descriptors' scale in structure guidance

ance improves AArTRE from 0.0254 to 0.0231, indicating that transferring the style representations between the training dataset and the test dataset is helpful for generalization. And structure guidance further boosts the test-time adaptation performance, which is consistent with the claims in (Ge et al. 2022). Finally, model continuity based on a smoothing strategy improves registration accuracy. Though model continuity does not give as significant an improvement as the other two, it is easy to implement and does not have much extra burden, so it is meaningful for the test-time adaptation process of registration.

Effect of Feature Position in Style Guidance. In style guidance, we combine the style representations from the training images with features from the test images by AdaIN, and we can choose the position of the combined features. As mentioned before, each backbone has four encoded features according to the spatial scale, we involve these four features in the style guidance in default. Here, we analyze the effect of encoded features' position, and the results are shown in Table 3. The results show that even with the first encoded feature lonely, the performance is improved, indicating the effectiveness of the style guidance. Along with introducing more encoded features, the registration performance is improved gradually. However, the latter the features' position is, the less impact on the registration performance. We think the reason is early features usually capture low-level information such as style while later features usually encode high-level information such as semantic content. So early features bring in more impact on the style guidance, thus influencing the performance.

Effect of HOG Scale in Structure Guidance. In structure guidance, we utilize the HOG descriptors as the structure representations to constrain the fixed image and warped moving image. In the default setting, we use multiple HOG descriptors with 4 spatial scales including $\{1, 1/2, 1/4, 1/8\}$. Here, we analyze the impact of different scales and show the results in Table 4. From the results, we can determine that multi-scale HOG is better than single-scale HOG, and the reason is that rough and fine structure representations are both important in the registration problem as mentioned in (Ge et al. 2022).

Conclusion

In this paper, for the first time, we propose a novel test-time adaptation method for histological image registration, named SGTTA. We design two operations, style guidance and structure guidance, for solving the challenges of appearance variances and local structure changes in histological images. Furthermore, we consider the continuity of the model and propose a smoothing strategy to leverage historical models. We conduct experiments on the public histological dataset with representative backbones, such as U-Net, Pyramid, and Cascade, demonstrating the superior performance of SGTTA.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62021001.

References

Arganda-Carreras, I.; Sorzano, C. O.; Marabini, R.; Carazo, J. M.; Ortiz-de Solorzano, C.; and Kybic, J. 2006. Consistent and elastic registration of histological sections using vector-spline regularization. In *Computer Vision Approaches to Medical Image Analysis: Second International ECCV Workshop, CVAMIA 2006 Graz, Austria, May 12, 2006 Revised Papers 2*, 85–95. Springer.

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.

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.

Balakrishnan, G.; Zhao, A.; Sabuncu, M. R.; Guttag, J.; and Dalca, A. V. 2019. VoxelMorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8): 1788–1800.

Belongie, S.; Malik, J.; and Puzicha, J. 2002. Shape matching and object recognition using shape contexts. *IEEE transactions on pattern analysis and machine intelligence*, 24(4): 509–522.

Borovec, J.; Kybic, J.; Arganda-Carreras, I.; Sorokin, D. V.; Bueno, G.; Khvostikov, A. V.; Bakas, S.; Eric, I.; Chang, C.; Heldmann, S.; et al. 2020. ANHIR: automatic non-rigid histological image registration challenge. *IEEE transactions on medical imaging*, 39(10): 3042–3052.

Ceritoglu, C.; Wang, L.; Selemon, L. D.; Csernansky, J. G.; Miller, M. I.; and Ratnanather, J. T. 2010. Large deformation diffeomorphic metric mapping registration of reconstructed 3D histological section images and in vivo MR images. *Frontiers in human neuroscience*, 4: 895.

Chakravarty, M. M.; Bertrand, G.; Hodge, C. P.; Sadikot, A. F.; and Collins, D. L. 2006. The creation of a brain atlas for image guided neurosurgery using serial histological data. *Neuroimage*, 30(2): 359–376.

Chen, Y.; Huang, W.; Zhou, S.; Chen, Q.; and Xiong, Z. 2023. Self-supervised neuron segmentation with multi-agent reinforcement learning. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, 609–617.

Dalal, N.; and Triggs, B. 2005. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, 886–893. Ieee.

Dalca, A. V.; Balakrishnan, G.; Guttag, J.; and Sabuncu, M. R. 2018. Unsupervised learning for fast probabilistic diffeomorphic registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 729–738. Springer.

Du Bois d'Aische, A.; De Craene, M.; Geets, X.; Gregoire, V.; Macq, B.; and Warfield, S. K. 2005. Efficient multi-modal dense field non-rigid registration: alignment of histological and section images. *Medical image analysis*, 9(6): 538–546.

Gatys, L. A.; Ecker, A. S.; and Bethge, M. 2016. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2414–2423.

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.

Glocker, B.; Sotiras, A.; Komodakis, N.; and Paragios, N. 2011. Deformable medical image registration: setting the state of the art with discrete methods. *Annual review of biomedical engineering*, 13: 219–244.

Hu, B.; Zhou, S.; Xiong, Z.; and Wu, F. 2020. Self-recursive Contextual Network for Unsupervised 3D Medical Image Registration. In *International Workshop on Machine Learning in Medical Imaging*, 60–69. Springer.

Hu, B.; Zhou, S.; Xiong, Z.; and Wu, F. 2022a. Cross-Resolution Distillation for Efficient 3D Medical Image Registration. *IEEE Transactions on Circuits and Systems for Video Technology*.

Hu, B.; Zhou, S.; Xiong, Z.; and Wu, F. 2022b. Recursive Decomposition Network for Deformable Image Registration. *IEEE Journal of Biomedical and Health Informatics*, 26(10): 5130–5141.

Hu, M.; Song, T.; Gu, Y.; Luo, X.; Chen, J.; Chen, Y.; Zhang, Y.; and Zhang, S. 2021. Fully test-time adaptation for image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24*, 251–260. Springer.

Hu, X.; Kang, M.; Huang, W.; Scott, M. R.; Wiest, R.; and Reyes, M. 2019. Dual-stream pyramid registration network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 382–390. Springer.

Huang, X.; and Belongie, S. 2017. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision*, 1501–1510.

Iwasawa, Y.; and Matsuo, Y. 2021. Test-time classifier adjustment module for model-agnostic domain generalization. *Advances in Neural Information Processing Systems*, 34: 2427–2440.

KingmaandJ, D. 2015. L. Ba,"ADAM: Amethodforstochasticoptimization,". In *Proc. 3rd Int. Conf. Learn. Representations*, 1–15.

Klein, S.; Staring, M.; Murphy, K.; Viergever, M. A.; and Pluim, J. P. 2009. Elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging*, 29(1): 196–205.

Li, M.; Zhou, S.; Chen, C.; Zhang, Y.; Liu, D.; and Xiong, Z. 2022. Retinal Vessel Segmentation with Pixel-Wise Adaptive Filters. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, 1–5. IEEE.

Liu, X.; Zhang, Y.; Zhou, S.; Xiong, Z.; and Sun, X. 2023. Electron Microscopy Image Registration Using Correlation Volume. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, 1–5. IEEE.

Liu, Y.; Kothari, P.; Van Delft, B.; Bellot-Gurlet, B.; Mordan, T.; and Alahi, A. 2021. Ttt++: When does self-supervised test-time training fail or thrive? *Advances in Neural Information Processing Systems*, 34: 21808–21820.

Lowe, D. G. 1999. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, 1150–1157. Ieee.

Miranda, G. H. B.; Barrera, J.; Soares, E. G.; and Felipe, J. C. 2012. Structural analysis of histological images to aid diagnosis of cervical cancer. In *2012 25th SIBGRAPI Conference on Graphics, Patterns and Images*, 316–323. IEEE.

Mok, T. C.; and Chung, A. C. 2020. Large Deformation Diffeomorphic Image Registration with Laplacian Pyramid Networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 211–221. Springer.

Rueckert, D.; Sonoda, L. I.; Hayes, C.; Hill, D. L.; 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.

Song, Y.; Treanor, D.; Bulpitt, A. J.; and Magee, D. R. 2013. 3D reconstruction of multiple stained histology images. *Journal of pathology informatics*, 4(2): 7.

Sun, Y.; Wang, X.; Liu, Z.; Miller, J.; Efros, A.; and Hardt, M. 2020. Test-time training with self-supervision for generalization under distribution shifts. In *International conference on machine learning*, 9229–9248. PMLR.

Ulyanov, D.; Vedaldi, A.; and Lempitsky, V. 2017. Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6924–6932.

Venet, L.; Pati, S.; Feldman, M. D.; Nasrallah, M. P.; Yushkevich, P.; and Bakas, S. 2021. Accurate and robust alignment of differently stained histologic images based on Greedy diffeomorphic registration. *Applied Sciences*, 11(4): 1892.

Wang, D.; Shelhamer, E.; Liu, S.; Olshausen, B.; and Darrell, T. 2020. Tent: Fully test-time adaptation by entropy minimization. *arXiv preprint arXiv:2006.10726*.

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.

Zhao, S.; Dong, Y.; Chang, E. I.; Xu, Y.; et al. 2019a. Recursive cascaded networks for unsupervised medical image registration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10600–10610.

Zhao, S.; Lau, T.; Luo, J.; Eric, I.; Chang, C.; and Xu, Y. 2019b. Unsupervised 3D end-to-end medical image registration with volume tweening network. *IEEE journal of biomedical and health informatics*, 24(5): 1394–1404.

Zhou, S.; Hu, B.; Xiong, Z.; and Wu, F. 2023. Self-Distilled Hierarchical Network for Unsupervised Deformable Image Registration. *IEEE Transactions on Medical Imaging*.

Zhou, S.; Xiong, Z.; Chen, C.; Chen, X.; Liu, D.; Zhang, Y.; Zha, Z.-J.; and Wu, F. 2019. Fast and accurate electron microscopy image registration with 3D convolution. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 478–486. Springer.

Zhu, W.; Huang, Y.; Xu, D.; Qian, Z.; Fan, W.; and Xie, X. 2021. Test-time training for deformable multi-scale image registration. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 13618–13625. IEEE.