Deep Learning Based Registration of Serial Whole-slide Histopathology Images in Different Stains

Mousumi Roy[1,*], Fusheng Wang[1,2,*], George Teodoro[3], Ritu Aneja[4]other paper in another context., and Jun Kong[5,6,*]

[1] Department of Computer Science, and Biomedical Informatics, Stony Brook University, NY, 11794, USA
[2] Department of Biomedical Informatics, Stony Brook University, NY, 11794, USA
[3] Department of Computer Science, Federal University of Minas Gerais, Belo Horizonte, 31270-901, Brazil
[4] Department of Biology, Georgia State University, Atlanta, GA, 30303, USA
[5] Department of Mathematics and Statistics and Computer Science, Georgia State University, Atlanta, GA, 30303, USA
[6] Department of Computer Science and Winship Cancer Institute, Emory University, Atlanta, GA, 30322, USA
[*]E-mail: mousumi.roy@stonybrook.edu and fusheng.wang@stonybrook.edu and jkong@gsu.edu

Author Declaration There is no conflict of interest to report.

Abstract. Whole-slide image (WSI) analysis has been largely performed in a 2D tissue space to support routine pathology diagnosis and imaging based biomedical research. For a more definitive representation and characterization of the tissue spatial space, it is critical to extend such tissue based investigations to a 3D space by spatially aligning 2D serial sections, which are often stained differently, such as Hematoxylin and Eosin (H&E) and Immunohistochemistry (IHC) stains. However, registration of whole slide images is challenged by the overwhelmingly scale of images, the complexity of local histology structure changes across slides, and significant variations of tissue appearance between staining methods. We propose a novel translation based registration network CycGANRegNet using deep learning for serial WSI images in different stains, which requires no prior deformation field information for deep model training. We first generate synthetic IHC slides from H&E slides through a robust image synthesis algorithm. The synthetic IHC images and the real IHC images are then registered through a Fully Convolutional Network with multi-scale based deformable vector fields and a joint loss optimization for enhancing image alignment. We perform the registration at original image resolution with a patch-wide approach, thus tissue details at the highest resolution are retained in the results. CycGANRegNet outperforms both the state-of-the-art conventional and deep learning-based registration methods based on the evaluation using a serial WSI image dataset in H&E stain and IHC stain with two biomarkers from 76 breast cancer patients. The experimental and comparison results demonstrate that CycGANRegNet can produce promising registration results with

serial WSIs in different stains, suggesting its potential for integrative 3D tissue-based biomedical investigations.

Keywords: H&E · IHC · Pathology Image · CycleGAN · WSI.

1 Introduction

Histopathology Whole-Slide Images (WSIs) of tissue sections provide high resolution tissue details critical for disease prognosis and study. However, such high resolution WSIs have been largely analyzed in a 2D tissue space by far. As each 2D WSI only captures information from the tissue cutting plane, it is inevitably subject to information loss and even distortion. Therefore, it is important and necessary to extend to a 3D tissue space for studies requiring a definitive tissue characterization. Additionally, there is an increasing demand of integrative tissue analyses involving both histology tissue phenotype descriptions by the clinically used Hematoxylin and Eosin (H&E) stain and underlying disease molecular underpinnings by highlighted Immunohistochemistry (IHC) biomarkers, which urges a 3D tissue space based approach. Apparently, an accurate registration of serial tissue WSIs is a prerequisite to such 3D tissue analyses and integration. However, WSI registration is technically challenging and existing methods only achieve a limited success due to the overwhelmingly large WSI scale, complex local histology structure change across adjacent slides, and significant tissue appearance difference by different staining methods [1].

With the emergence of deep learning methods, it becomes feasible to take advantage of the image-to-image translation through the latent feature disentanglement for registering images in different appearances [4–6]. Specifically, translation-based approaches use Generative Adversarial Network(GAN) to translate images from one modality into another, simplifying the difficult multi-stained image registration task into a much easier analysis with images in a single stain. By this strategy, GAN can be used to convert H&E to synthetic IHC images for a multi-stain image integrative analysis. Although much simplified, such an analysis still presents two significant challenges. First, H&E and IHC WSIs are unpaired data as each pathology tissue slide is stained only once in most clinical practice. Second, it is time and financially costly to have good quality annotations on landmark pairs from serial WSIs for registration.

Recently, CycleGAN with the cycle consistent loss has been developed to learn an image-to-image mapping between two domains from unpaired data [8]. The cycle consistency loss for the adversarial training process forces the generator to find an accurate mapping between two different domains with unpaired data. With this approach, synthetic slice-wise computed tomography (CT) data has been produced from magnetic resonance (MR) head images [9]. In another study, a mono-modal image registration with CycleGAN-produced synthetic images presents a comparable performance to the multi-modal deformable registration with paired image data of thoracic and abdominal organs [4]. These studies leverage the standard CycleGAN [8] for image synthesis not customized

for pathology data. Furthermore, a deformation field is used for registration in a dual-stream fashion with CycleGAN-translated MR and original CT image pairs [10]. MIND based [10] loss term added to CycleGAN loss describes the local image structure. It is computed from gray-scale images, and thus cannot be applied to multi-stained pathology image data. A CycleGAN based image generation method is reported to generate IHC pathology microscopic images from H&E images without any annotation [11]. As this method with class-related information is used as an additional input image patch channel, it is not appropriate for our study.

In this paper, we present a new translation based deep learning registration approach (CycGANRegNet) for serial whole-slide histopathology images in different stains. The developed unsupervised registration approach requires no prior deformation field information for deep model training. It consists of an image translation and an image registration module. The image translation module produces synthetic IHC slides (i.e. synIHC images) from H&E slides through a robust image synthesis algorithm. With Fully Convolutional Network (FCN) model [12] as a building block, the synIHC and the real IHC image pairs are registered through a multi-scale based FCN registration model. The major contributions of this paper are summarized in multiple folds:

- We develop a modified CycleGAN method to generate synthetic IHC pathology images (i.e. synIHC) from unpaired H&E pathology slides. To enhance the image stain translation ability, we propose to adopt a perceptual loss in the CycleGAN loss function, resulting in a better image mapping from H&E to synthetic IHC images. Such an image translation enables a better registration between synIHC and real IHC images.
- We extend the original FCN model to a multi-scaled architecture. Our proposed multi-scale FCN uses a coarse-to-fine multi-scale deformable image registration strategy that combines the Deformable Vector Fields (DVFs) at multiple resolutions for better image alignment.
- To overcome the overwhelming scale of WSI images, instead of resizing images to a lower resolution [1], we recover the WSI registration results with patch-based image registration results at the highest image resolution. In this way, high resolution tissue details captured by WSIs are retained.

2 Methods

We develop a deep learning based model CycGANRegNet to register serial IHC to H&E histopathology images for pathology hallmark and biomarker integration. It has an end-to-end deep learning process in two stages. First, we develop an image translation module with a modified CycleGAN to translate real reference H&E to synthetic reference synIHC image patches. Next, we develop a multi-scale FCN in the image registration module to estimate the spatial mapping from the moving real IHC to the synthetically produced synIHC image patches. Finally, the moving real IHC image is transformed to the reference H&E

image space via Spatial Transformation Network (STN) [13]. Individual registered image patches are spatially assembled to recover registered WSI blocks. We present the overall schema of CycGANRegNet in Figure 1.

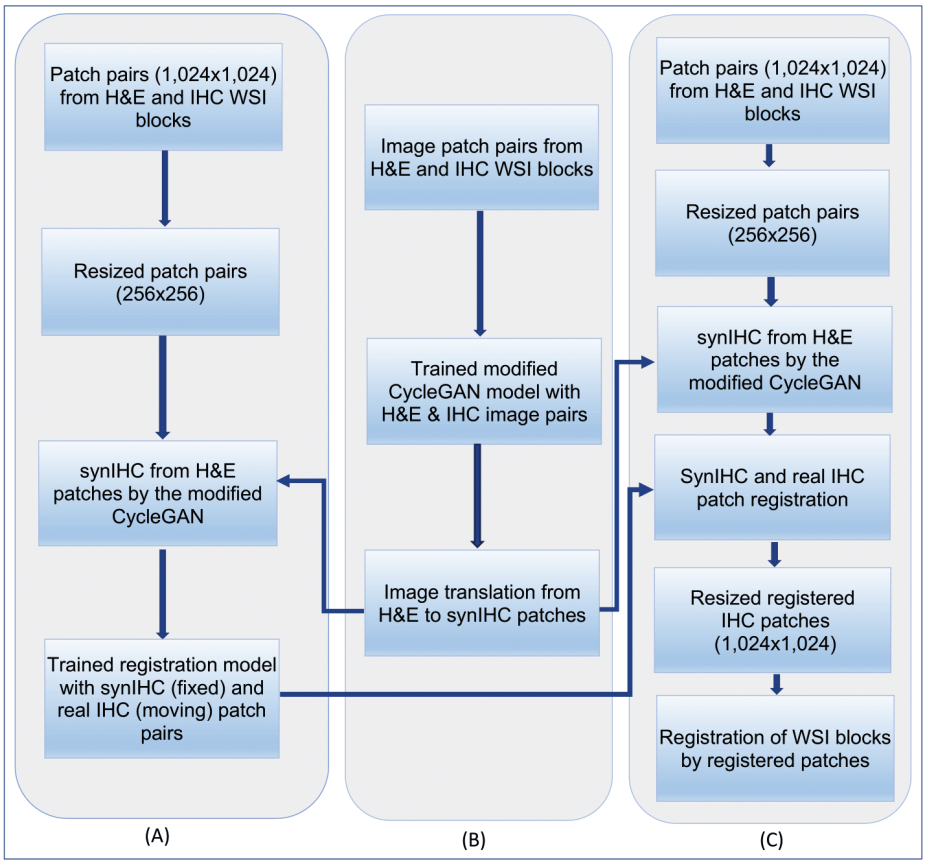

Fig. 1. The overall schema of the CycGANRegNet model. (A) Patch-based H&E-IHC image registration; (B) Translation from H&E to synIHC images; and (C) WSI block registration.

2.1 Unpaired Image Translation

Although serial slides in different stains look similar at the global tissue level (Supplemental Figure 1), they are unpaired at the pixel level. Although CycleGAN can be applied to unpaired image-to-image translation, it can be tailored to pathology image translation for an enhanced performance [11, 8, 14]. Therefore, we propose a modified CycleGAN for an enhanced H&E-IHC image translation.

Illustrated in Figure 2, the modified CycleGAN consists of an encoder and a decoder module. Both share the same network structure that includes a generator and a discriminator. The generator translates an image between stain domains and the discriminator assesses the generated image quality. The modified CycleGAN model consists of two generators G_{HE} and G_{IHC}. The generator G_{HE} translates IHC to synHE image, while G_{IHC} translates synHE to synIHC image(i.e. the red arrows). Similarly, the reverse translation goes from H&E to synIHC and then from synIHC to synHE (i.e. the black arrows). Each generator module consists of two-dimensional fully convolutional networks with nine residual blocks and two fractionally strided convolution layers as used in ResNet [32]. Additionally, the model has two discriminators D_{HE} and D_{IHC} for distinguishing between translated H&E (i.e. synHE) and real H&E images, and between translated synIHC and real IHC images, respectively. Each discriminator has a fully convolutional architecture to predict if overlapping image patches of size 70×70 by pixels are real or synthetic [28](PatchGAN). The leaky ReLU activation function is used with factor 0.2. All data are normalized with instance normalization. The detailed architecture of generator is adopted from [15].

The CycleGAN training loss includes adversarial loss (i.e. $L_{D_{HE}}$ and $L_{D_{IHC}}$) from two discriminators and the cycle-consistent loss (L_{cyc}). Although the cycle-consistent loss plays an important role of prohibiting the generators from generating images not related to the inputs [8], this loss is not sufficient to enforce feature or structural similarity between translated and real images. To address this problem, we adopt the vgg-16 based perceptual loss function as an additional constraint in the CycleGAN loss function to regularize the tissue content and the stain style discrepancies, the key contributing to an enhanced histopathology image translation. Such a perceptual loss can measure high-level perceptual and semantic differences between each image pair [15]. The perceptual loss function is implemented as a deep convolutional neural network denoted as ϕ [15]. The loss network ϕ is a 16-layer VGG network [16] pre-trained on the ImageNet dataset [17]. The transformed network output image $\hat{y}$ and the input real image y are enforced to have similar feature representations from the loss network ϕ. Let $\phi_j(x)$ be the activation of the j-th layer of the network ϕ with the input image x. The feature reconstruction loss $\mathcal{L}_{feat}$ is computed as:

$$\mathcal{L}_{feat} = l_{feat}^{\phi,j}(\hat{y}, y) = \frac{1}{C_j H_j W_j} \|\phi_j(\hat{y}) - \phi_j(y)\|_2^2 \tag{1}$$

where $\phi_j(x)$ is an activation map of size $C_j \times H_j \times W_j$. Note the image transformation network trained by the feature reconstruction loss encourages the output image $\hat{y}$ to be perceptually similar to the target image y. However, it does not force them to match each other exactly.

The total loss $\mathcal{L}$ of our modified CycleGAN is defined as:

$$\mathcal{L} = \mathcal{L}_{D_{HE}} + \mathcal{L}_{D_{IHC}} + \lambda_{cyc}\mathcal{L}_{cyc} + \lambda_{feat}\mathcal{L}_{feat} \tag{2}$$

where λ_{cyc} and λ_{feat} are weights for different loss terms.

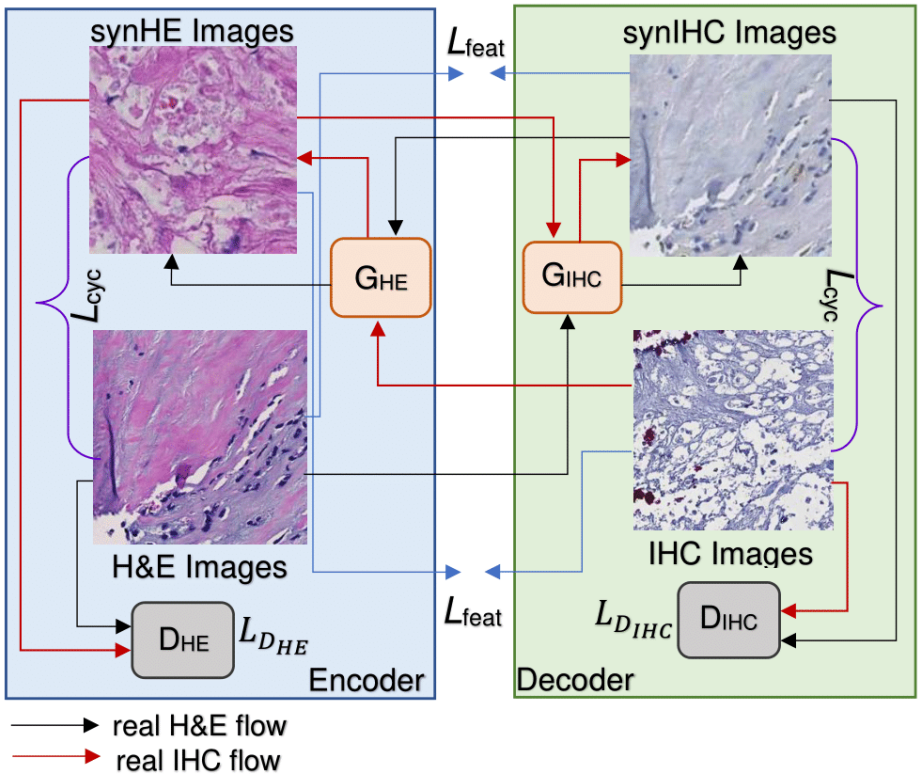

Fig. 2. The overall schema of the modified CycleGAN with the forward and backward translation information flow.

2.2 Multi-scaled Image Patch Registration

The FCN model is a state-of-the-art deep learning approach with known promising performance for histopathology image registration [12]. Multiple prior studies on flow estimation have shown the effectiveness of multi-scale strategy [19, 2]. Using FCN and the multi-scale strategy as building blocks, we create a multi-scale FCN model with multi-scale Displacement Vector Fields (DVFs) to enable a coarse-to-fine multi-scale deformable image registration. The developed multi-scale registration framework consists of three DVF estimation models and is demonstrated in Figure 3.

Fixed I_F and moving I_M image pairs are inputs to the multi-scale FCN model. For each pair, the real moving and fixed image are concatenated and provided to the first DVF estimation model to estimate the DVF V_1 at scale-1. V_1 has three different components (i.e. V_{11}, V_{12} and V_{13}) generated from two different regression layers and the final layer of the FCN model. The resulting warped moving images are $V_{11} \circ I_M$, $V_{12} \circ I_M$, and $V_{13} \circ I_M$ respectively. Next,

the fixed I_F and moving I_M image pairs are concatenated and down-sampled by a factor of four. The resulting image pairs are provided to the second DVF estimation model for DVF estimation at scale-2. The resulting DVF is up-sampled to match the original input image size and denoted as V_2. Similarly, V_2 is applied to the input moving image I_M to generate the warped moving image $V_2 \circ I_M$. In the next step, the warped moving image $V_2 \circ I_M$ at scale-2 and the input fixed image I_F are concatenated and down-sampled by a factor of two. The resulting image pairs are provided to the third DVF estimation model for the residual DVF estimation. The resulting DVF is up-sampled to the original input data size and denoted as V_3 at scale-3. Finally, V_3 is used to deform the moving image I_M for the warped moving image $V_3 \circ I_M$. Instead of training each model separately, we train all three DVF estimation models to minimize the joint loss at multiple scale levels, achieving an overall end-to-end optimal performance. The total loss function is defined as:

$$\begin{aligned} \mathcal{L} = \sum_{i=1}^{3} \sigma_{1i} \{ \mathcal{L}_{sim}(I_F, I_M \circ V_{1i}) + R(V_{1i}) \} \\ + \sigma_2 \mathcal{L}_{sim}(I_F, I_M \circ V_2) + \sigma_3 \mathcal{L}_{sim}(I_F, I_M \circ V_3) \end{aligned} \tag{3}$$

where $\mathcal{L}_{sim}$ is the similarity loss measured as negative Normalized Cross-Correlation (NCC) [7], penalizing the differences in appearance between fixed and moving images. Parameter σ_{11}, σ_{12} and σ_{13} are weights of the similarity loss metrics at scale level 1 and σ_2 and σ_3 are weights at scale levels 2 and 3 respectively. $R(V)$ is a total variation based regularizer that makes the transformation spatially smooth and physically plausible [3].

After the weight initialization, all weights are updated by the joint training of three DVF estimation models in an end-to-end manner for the harmonic minimization of the composite loss. With displacement vectors between the fixed and moving image pairs, we use Spatial Transformer Network (STN) [13] to deform the moving image with the dense deformation field V [12]. To make the resulting registered images retain more tissue details, we adopt the Enhanced SRGAN (ESRGAN) model [31] in the post-processing step. Figure 4 demonstrates the registered images by our proposed multi-scale FCN model with and without the post-processing. There is a noticeable difference in image details with and without such post-processing. After post-processing, the registration performance is improved by all performance metrics as presented in Table 2.

2.3 WSI Block Registration

Due to the limited GPU memory size, deep learning methods cannot process giga-pixel WSIs at the full histopathology image resolution. Therefore, the tissue region in each WSI is first partitioned into image blocks of size $8,000 \times 8,000$ pixels for tissue pre-alignment. Each H&E block is next translated to synIHC block by the developed modified CycleGAN. Real IHC and synIHC image blocks are then divided into image patches of size $1,024 \times 1,024$ to retain sufficient tissue

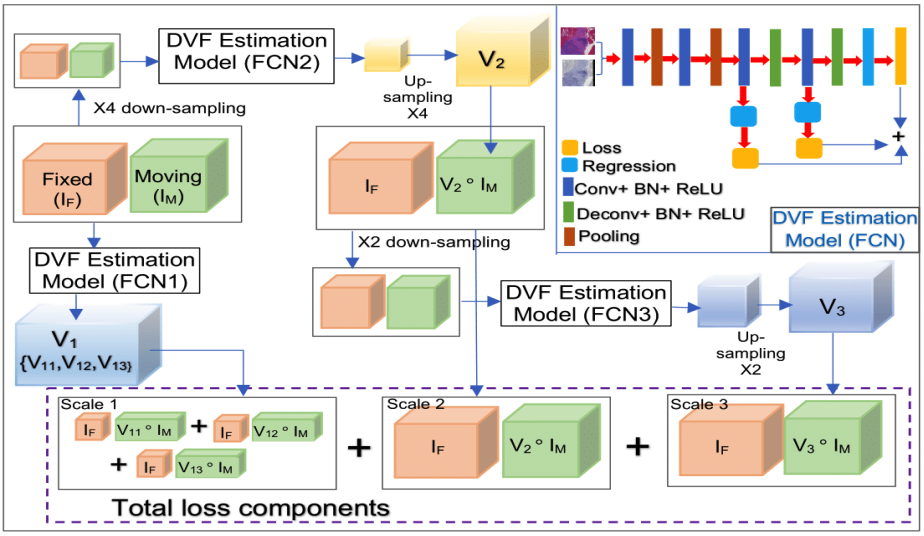

Fig. 3. The overall architecture of the developed multi-scale FCN model is presented with detailed illustrations of FCN layers for multi-sscale DVF estimation and multiple loss function components in the total loss.

information for registration. The resulting synIHC and real IHC patch pairs are further resized to 256×256 pixels for the deep learning model training and prediction. After registration, the registered real IHC image patches are resized back to $1,024 \times 1,024$ size and spatially combined [20].

3 Experimental Result

3.1 Dataset and Implementation

We assess our model on 228 WSIs of tumor tissue sections of 76 Neoadjuvant Chemotherapy (NAC) treated Triple Negative Breast Cancer (TNBC) patients from Dekalb Medical Center in Emory University Healthcare. Formalin-fixed paraffin-embedded serial section samples are collected before neoadjuvant therapy. The serial sections are H&E and immunohistochemically stained with Ki67 biomarker for cell proliferation and Phosphohistone H3 (PHH3) for mitotic activity. After the image pre-alignment by the global affine spatial transformation at a low image resolution, the resulting transformation is mapped to the full image resolution level. The pre-aligned tissue regions at the full image resolution level are next partitioned into 1,023 WSI blocks of size $8,000 \times 8,000$ pixels by each stain. The pre-aligned WSI blocks are further partitioned into non-overlapping image patches of size $1,024 \times 1,024$ pixels, followed by resizing to 256×256 to make the image size appropriate for deep learning models. Patches containing more than 30% background pixels are excluded from further analyses. This

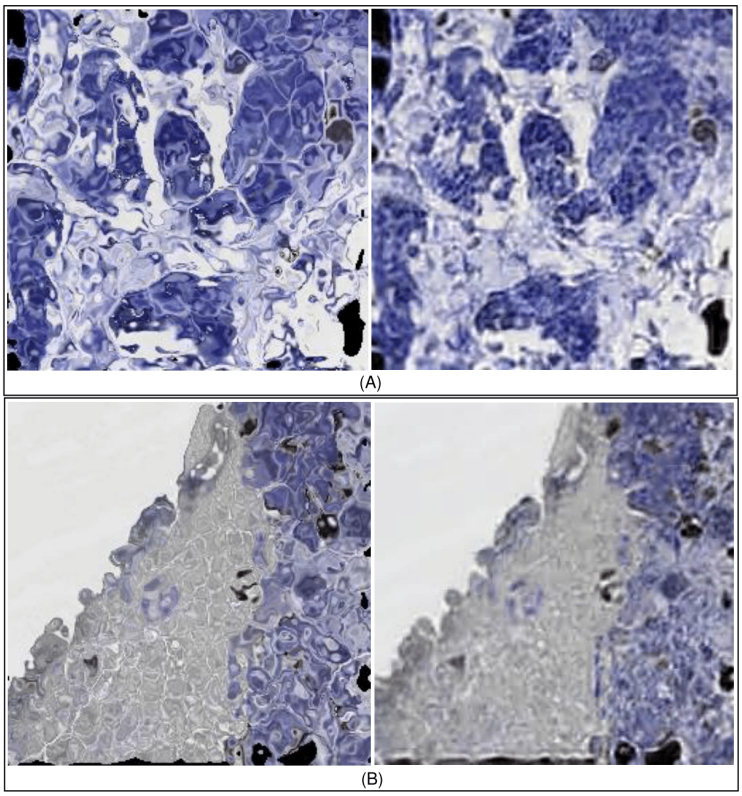

Fig. 4. Registration results from our proposed multi-scale FCN model with and without the ESRGAN model based post-processing. A registered image patch from (A) testing, and (B) validation dataset is demonstrated (Left) with and (Right) without the post-processing.

results in 60,000 image patches that are randomly divided into training, validation and testing cohorts by the 80:10:10 ratio. Our model is first tested with H&E-Ki67 slide registration, followed by H&E-PHH3 registration for additional validation. We compare our model with multiple state-of-the-art methods using 'real' and 'synthetic' datasets. The 'real' dataset includes pairs of H&E and real IHC images, while a 'synthetic' dataset consists of real IHC and synIHC image pairs with synKi67 for testing and synPHH3 for validation, respectively. Note the 'synthetic' data with synIHC images generated from the CycleGAN is labeled as 'syn-1' dataset, whereas the 'syn-2' dataset includes synIHC images from our modified CycleGAN model. The developed CycGANRegNet is implemented with the open-source deep learning library Tensorflow [21]. The experiment is carried out on Tesla K80 and V100 GPUs with CUDA 9.1. Adam optimization algorithm [22] with learning rate 0.0001 is used to train both image translation and

image registration models. For the modified CycleGAN training, H&E and IHC images are partitioned into 256×256 image patches. The modified CycleGAN is trained for up to 2,00,000 iterations. Loss weights λ_{cyc} and λ_{feat} are set to 1. All other parameter settings are suggested by the original CycleGAN work [8]. The values of registration loss weights σ_{11}, σ_{12}, σ_{13}, σ_2 and σ_3 are set to 0.9, 0.6, 0.3, 0.05 and 0.05 respectively.

3.2 Evaluation of Image Translation Module

Image translation performance of the CycleGAN and our modified CycleGAN is evaluated and compared both at the patch and WSI block level. Representative translated image patches and WSI blocks are demonstrated in Figure 5 and Figure 6, respectively. In addition to the qualitative assessment, we quantitatively evaluate the translated IHC image quality by Root Mean Square Error (RMSE), Structural Similarity Index Measure (SSIM), and Peak Signal-to-Noise Ratio (PSNR) [23, 24, 10]. Specifically, the trained generator G_{IHC} and G_{HE} are used to translate a H&E to a synIHC image, and then back from the synIHC to the synHE (i.e. black arrows in Figure 2), in turn. The similarity between the real H&E and resulting synHE images is quantitatively evaluated and presented in Table 1. Both forward (i.e. H&E to synIHC and synIHC to synHE) and reversed (IHC to synHE and synHE to synIHC) translation performances by the original and our modified CycleGAN model are presented. Note the image translation between H&E and Ki67 is for testing, while the H&E-PHH3 translation is for validation. Our proposed modified CycleGAN presents consistent superior performance to the original CycleGAN by all evaluation metrics in both forward and reversed translation directions. Additionally, the translated synIHC image patches are spatially combined to generate the translated synIHC WSI blocks as depicted in Figure 6. Both qualitative and quantitative experiments with the testing and validation dataset suggest an enhanced image translation by our modified CycleGAN.

Table 1. Quantitative image translation performance comparison between the original CycleGAN and our modified CycleGAN model with testing and validation dataset.

Testing Data				Validation Data			
	RMSE	SSIM	PSNR		RMSE	SSIM	PSNR
CycleGAN (H&E → Ki67)	15.200	0.934	29.928	CycleGAN (H&E → PHH3)	17.641	0.905	28.528
Our (H&E → Ki67)	**14.414**	**0.935**	**30.387**	Our (H&E → PHH3)	**17.210**	**0.911**	**28.754**
CycleGAN (H&E ← Ki67)	13.088	0.955	31.215	CycleGAN (H&E ← PHH3)	15.114	0.928	29.956
Our (H&E ← Ki67)	**12.123**	**0.959**	**31.867**	Our (H&E ← PHH3)	**14.430**	**0.935**	**30.282**

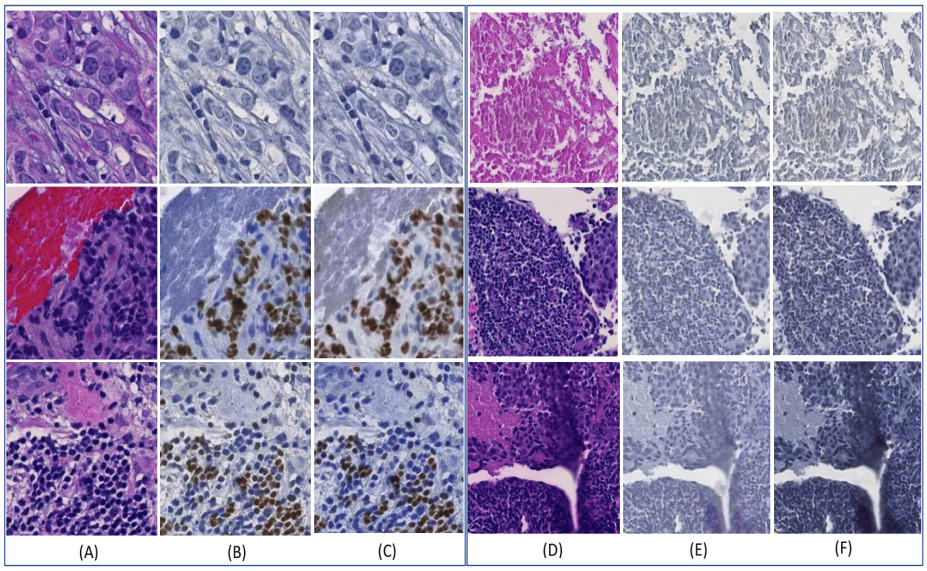

Fig. 5. Representative image patch translation results. (A) Real H&E image patch; (B) synKi67 patch by CycleGAN; (C) synKi67 patch by our modified CycleGAN; (D) Real H&E patch; (E) synPHH3 patch by CycleGAN; (F) synPHH3 patch by our modified CycleGAN.

3.3 Evaluation of Patch-based Image Registration Module

After the image translation, the resulting synIHC and real IHC WSI blocks are first pre-aligned by a global intensity-guided rigid transformation [18]. Next, CycGANRegNet takes rigidly registered synIHC and real IHC WSI blocks as inputs and fine tunes image alignment by our proposed multi-scale FCN model. To evaluate the registration performance, we apply our method to the 'real', 'syn-1' and 'syn-2' dataset from both the testing and validation data. Our proposed multi-scale FCN model is compared with multiple state-of-the-art deep learning based registration methods, including DirNet [25], FCN [12], Unet VoxelMorph [26] and the conventional registration method SimpleElastix [27].

Note the H&E and IHC WSI blocks are pre-aligned by the global affine transformation before registration methods for comparison, the registration effect on the image patch level can be not salient in some cases. To manifest the method efficacy, we first manually deform the moving images by both affine and elastic transformations, resulting in multiple synthetically deformed moving images. Diverse shear transformations with rotation angle range $[-40, 30]$ and elastic deformations are applied to moving images. A total of $20,000$ synthetically deformed image patches of size 256×256 are derived from 1,023 WSI blocks for training purpose. Typical registration results from the shear transformation and the elastic deformation by state-of-the-art deep learning-based registration mod-

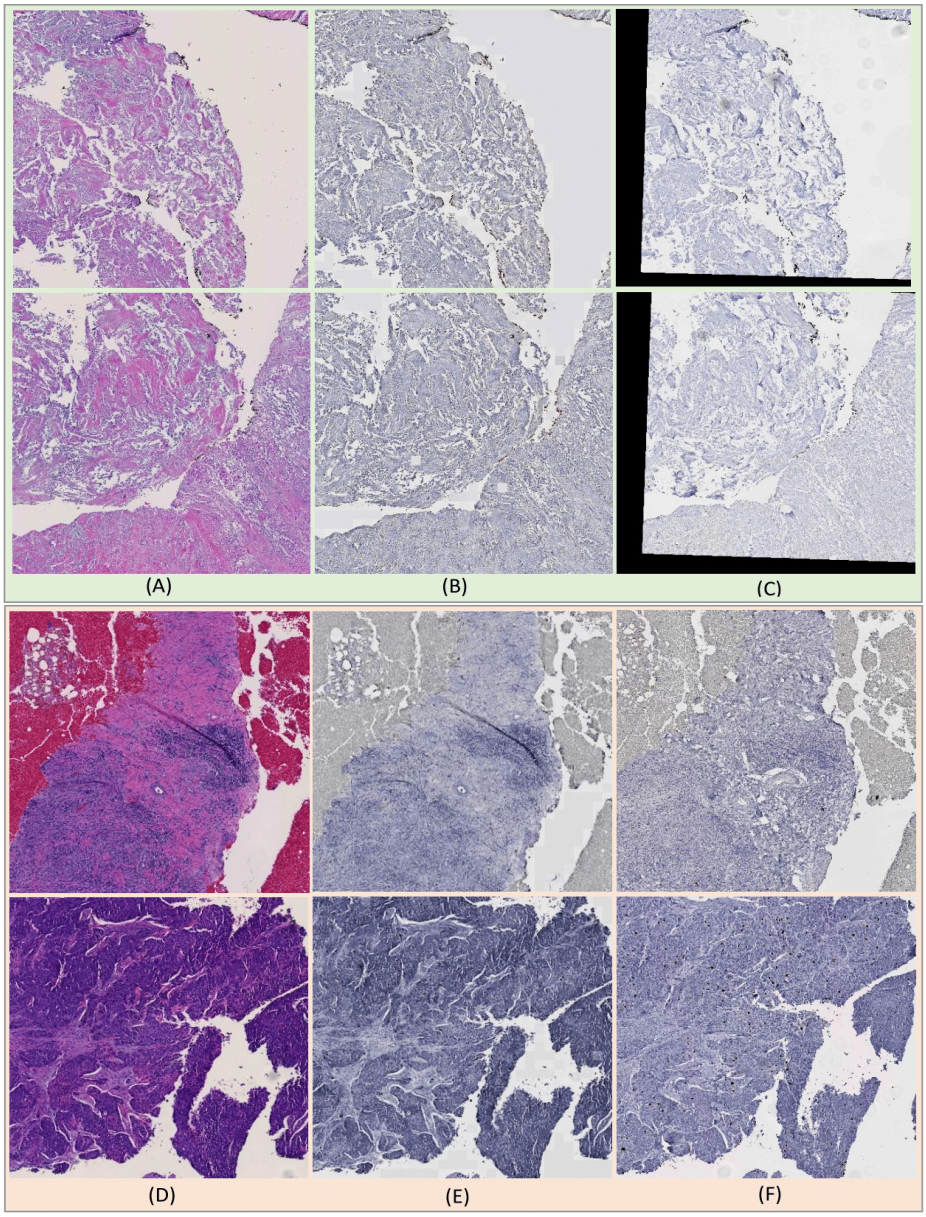

Fig. 6. Representative WSI block translation results with the test (top) and validation dataset (bottom). (A) Real H&E WSI blocks; (B) synKi67 WSI blocks by our modified CycleGAN; (C) Real Ki67 WSI blocks; (D) Real H&E WSI blocks; (E) synPHH3 WSI blocks by our modified CycleGAN; (F) Real PHH3 WSI blocks.

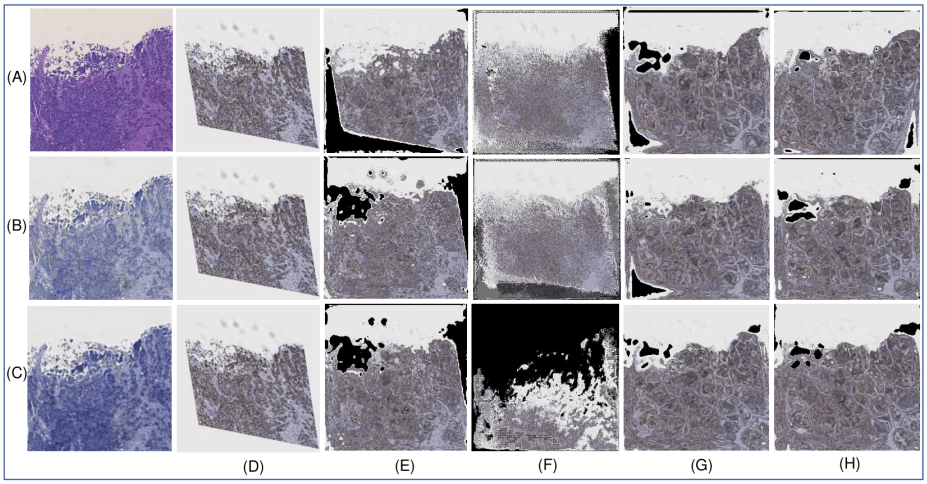

Fig. 7. Patch based registration performance by the synthetic shear transformation. (A)Fixed real H&E image; (B)Fixed synIHC image by CycleGAN; (C)Fixed synIHC image by our modified CycleGAN); (D) Manually rotated Moving image; Registration results by (E)DirNet, (F)VoxelMorph Unet, (G)FCN, and (H)Proposed multi-scale FCN.

els for comparison are presented in Figure 7 and Figure 8, respectively. By visual assessments, FCN and our proposed multi-scale FCN model present significantly better registration results than other deep learning methods.

After method evaluation with manually deformed images, we next apply registration methods to 'real', 'syn-1' and 'syn-2' datasets from the testing (i.e. H&E and Ki67 image pairs) and validation (i.e. H&E and PHH3 image pairs) data. We present the registration results in Figure 9 and Figure 10, respectively. By visual comparisons, baseline FCN and our proposed multi-scale FCN demonstrate a superior registration performance to other methods when 'syn-2' dataset is used. Additionally, we present quantitative performance evaluation results in Table 2 where Normalized Cross Correlation (NCC), SSIM and Normalized Mutual Information (NMI) are used to report the registration accuracy. Note our developed multi-scale FCN with the 'syn-2' dataset from both testing and validation data achieves the best performance by NCC, and the second best by SSIM. By contrast, the performance of the conventional registration method SimpleElastix is limited due to its over-deformed image outputs (Supplemental Figure 2). Additionally, all deep learning-based models outperform with the 'syn-2' than the 'syn-1' or 'real' dataset, suggesting the efficacy of the enhanced image translation quality by our modified CycleGAN.

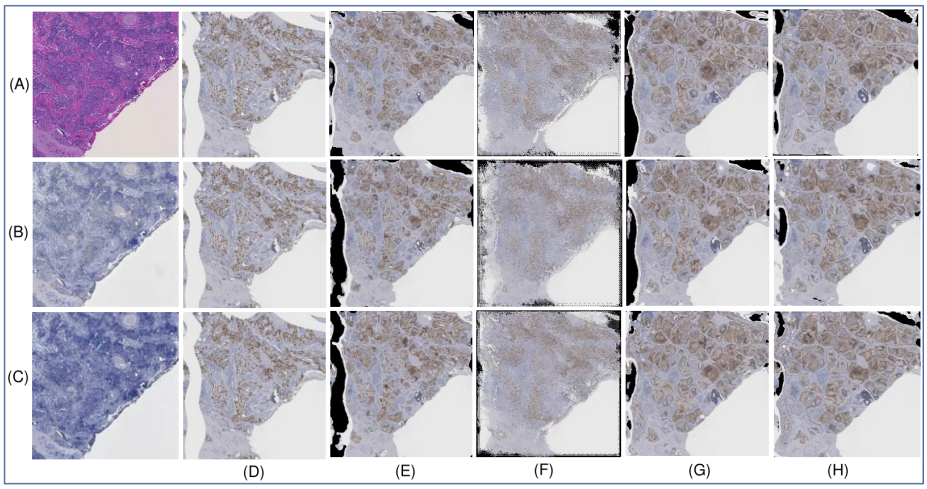

Fig. 8. Patch based registration performance by the synthetic elastic deformation. (A)Fixed real H&E image; (B)Fixed synIHC image by CycleGAN; (C)Fixed synIHC image by our modified CycleGAN); (D) Manually rotated Moving image; Registration results by (E)DirNet, (F)VoxelMorph Unet, (G)FCN, and (H)Proposed multi-scale FCN.

3.4 Evaluation of WSI Block Registration

We further evaluate the image registration with WSI blocks. As each WSI block has a size of $8,000 \times 8,000$ pixels, it is partitioned into small image patches for registration. After individual image patch registration, they are spatially combined to generate the registered WSI blocks. In this study, we adopt Dice Similarity Coefficient (DSC) [29], Hausdorff Distance (HD) [30], SSIM and NCC as the similarity metrics for WSI block registration accuracy evaluation.

Eight Region of Interest (ROI) pairs are manually annotated from the testing and validation WSI blocks before ROI pairs are registered. The complete evaluation process is presented in Supplemental Figure 3. The quantitative evaluation results are reported in Table 3. Our proposed multi-scale FCN with 'syn-2' dataset exhibits the second best performance by DSC and HD both for the testing and validation dataset, respectively. Note our proposed multi-scale FCN with 'syn-2' dataset produces the best performance by NCC and the third best by SSIM value in both testing and validation dataset. The best performance by SSIM is achieved by DirNet with the testing 'syn-2' dataset and FCN with the validation 'syn-1' dataset, respectively.

We present the WSI block derived ROI registration results with the testing and validation dataset in Figure 11 and Figure 12, respectively. By visual assessments, the registered ROIs by our proposed CycGANRegNet model are better aligned with the fixed images than results from other models for comparison. Thus, both quantitative and qualitative results suggest that our proposed

Table 2. Image patch registration performance with the testing and validation data.

Method Name	dataset	Testing Data			Validation Data		
		NCC	SSIM	NMI	NCC	SSIM	NMI
DirNet	real	0.2045	0.2307	1.0358	0.2030	0.2413	1.0344
	syn-1	0.2644	0.3320	1.0356	0.2613	0.3161	1.0334
	syn-2	0.2977	**0.3505**	1.0374	0.2553	**0.3443**	1.0324
FCN	real	0.2165	0.2168	1.0373	0.1742	0.2266	1.0362
	syn-1	0.2508	0.3119	1.0363	**0.2688**	0.3050	1.0339
	syn-2	0.3057	0.3356	1.0392	0.2683	0.3356	1.0340
Our Proposed CycGANRegNet	syn-2	**0.3161**	0.3403	1.0387	0.2625	0.3375	1.0343
Voxel Morph UNet	real	0.0231	0.1922	1.0329	0.0346	0.2078	1.0324
	syn-1	0.0755	0.3164	1.0314	0.0910	0.2828	1.0321
	syn-2	0.1063	0.3384	1.0347	0.1306	0.3243	1.0260
Simple Elastix	real	0.0968	0.1379	**1.0472**	0.0880	0.1453	**1.0474**

CycGANRegNet exhibits promising registration performance. Additionally, all registration results with WSIs in different stains suggest that the image translation from H&E to synIHC images help better align moving to fixed images.

Table 3. Registration performance with WSI blocks from the testing and validation data.

Dataset	Method Name	Testing Data				Validation Data			
		DSC	HD	SSIM	NCC	DSC	HD	SSIM	NCC
real	DirNet	0.9516	925.6837	0.7914	0.7983	0.9072	1290.1709	0.8153	0.7447
	FCN	0.9566	606.5160	0.7887	0.7833	0.9088	1302.1997	0.8145	0.7323
	Simple Elastix	0.9423	1240.1000	0.7348	0.6372	0.8881	1214.1727	0.7534	0.5276
syn-1	DirNet	0.9510	272.2740	0.8362	0.7958	0.9004	1300.1804	0.8659	0.7241
	FCN	0.9572	272.1319	0.8347	0.7988	**0.9161**	984.5001	**0.8709**	0.7639
syn-2	DirNet	0.9554	264.6888	**0.8369**	0.8075	0.9073	**486.4495**	0.8641	0.7576
	FCN	**0.9631**	**244.2928**	0.8345	0.8101	0.9149	1216.2255	0.8637	0.7614
	Our Proposed Multi-scale FCN	0.9612	257.2176	0.8347	**0.8175**	0.9152	724.4312	0.8654	**0.7683**

4 Discussion

In this study, we propose CycGANRegNet for registration of serial WSI images in different stains. By both qualitative and quantitative experiments, CycGANRegNet demonstrates a comparable performance to the standard FCN-based registration and a superior performance to other state-of-the-art registration

methods for comparison. Specifically, our proposed multi-scale FCN, i.e. the registration component in CycGANRegNet, follows a coarse-to-fine multi-scale deformable image registration strategy and optimizes the joint loss at multi-scale levels, leading to a more accurate DVF estimation critical for registration.

In Table 2, performance values for the patch-based registration are relatively low as some image patches are extracted from some poorly matched WSI block pairs after the pre-alignment step (Supplement Figure S4). When such registration performance is evaluated with well aligned WSI ROI pairs after pre-alignment step, performance values are much improved (c.f. Supplement Figure S5). To test the CycGANRegNet efficacy, we synthetically deform moving image patches by different affine transformations. In such controlled experiments, they can be well aligned to the fixed images. As real patch pairs are pre-aligned by a global rigid registration, the registration effect with real patch pairs from both testing and validation data is relatively subtle. To manifest the registration effect at the WSI block level, we zoom into specific tissue regions. Both registration results at the patch- and WSI block level suggest the necessity of image translation from one stain to another. The resulting synIHC images by the image translation help better estimate DVF for registration. In future, we will extend this work by better integrating spatial transformations from the H&E-IHC and synIHC-IHC pipeline for registration.

5 Conclusion

In this study, we propose a fully unsupervised translation based network CycGANRegNet for H&E and IHC pathology image registration. The resulting synthetic IHC images from the image translation module can be used for a better alignment with the real IHC images, with a multi-resolution approach to preserve histology structures at the original resolution. CycGANRegNet can be efficiently trained without any ground truth image deformation information. Experiments results at both high resolution image patches level and WSI blocks level demonstrate high effectiveness of the registration method with real world dataset.

Acknowledgements

This research was supported by National Institutes of Health [U01CA242936, R01EY028450, R01CA214928, R01CA239120], National Science Foundation [ACI 1443054, IIS 1350885], CNPq and FAPEMIG agencies.

References

1. Borovec, Jiří, Jan Kybic, Ignacio Arganda-Carreras, Dmitry V. Sorokin, Gloria Bueno, Alexander V. Khvostikov, Spyridon Bakas et al. "ANHIR: automatic non-rigid histological image registration challenge." IEEE transactions on medical imaging 39, no. 10 (2020): 3042-3052.

2. de Vos, Bob D., Floris F. Berendsen, Max A. Viergever, Hessam Sokooti, Marius Staring, and Ivana Išgum. "A deep learning framework for unsupervised affine and deformable image registration." Medical image analysis 52 (2019): 128-143.
3. Vishnevskiy, Valery, Tobias Gass, Gabor Szekely, Christine Tanner, and Orcun Goksel. "Isotropic total variation regularization of displacements in parametric image registration." IEEE transactions on medical imaging 36, no. 2 (2016): 385-395.
4. Tanner Christine, Firat Ozdemir, Romy Profanter, Valeriy Vishnevsky, Ender Konukoglu, and Orcun Goksel. "Generative adversarial networks for mr-ct deformable image registration." arXiv preprint arXiv:1807.07349 (2018).
5. Wei, Dongming, Sahar Ahmad, Jiayu Huo, Wen Peng, Yunhao Ge, Zhong Xue, Pew-Thian Yap, Wentao Li, Dinggang Shen, and Qian Wang. "Synthesis and inpainting-based mr-ct registration for image-guided thermal ablation of liver tumors." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 512-520. Springer, Cham, 2019.
6. Qin, Chen, Bibo Shi, Rui Liao, Tommaso Mansi, Daniel Rueckert, and Ali Kamen. "Unsupervised deformable registration for multi-modal images via disentangled representations." In International Conference on Information Processing in Medical Imaging, pp. 249-261. Springer, Cham, 2019.
7. Briechle, Kai, and Uwe D. Hanebeck. "Template matching using fast normalized cross correlation." In Optical Pattern Recognition XII, vol. 4387, pp. 95-102. International Society for Optics and Photonics, 2001.
8. Zhu, Jun-Yan, Taesung Park, Phillip Isola, and Alexei A. Efros. "Unpaired image-to-image translation using cycle-consistent adversarial networks." In Proceedings of the IEEE international conference on computer vision, pp. 2223-2232. 2017.
9. Wolterink, Jelmer M., Anna M. Dinkla, Mark HF Savenije, Peter R. Seevinck, Cornelis AT van den Berg, and Ivana Išgum. "Deep MR to CT synthesis using unpaired data." In International workshop on simulation and synthesis in medical imaging, pp. 14-23. Springer, Cham, 2017.
10. Xu, Zhe, Jie Luo, Jiangpeng Yan, Ritvik Pulya, Xiu Li, William Wells, and Jayender Jagadeesan. "Adversarial uni-and multi-modal stream networks for multimodal image registration." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 222-232. Springer, Cham, 2020.
11. Xu, Zidui, Xi Li, Xihan Zhu, Luyang Chen, Yonghong He, and Yupeng Chen. "Effective Immunohistochemistry Pathology Microscopy Image Generation Using CycleGAN." Frontiers in Molecular Biosciences 7 (2020).
12. Li, Hongming, and Yong Fan. "Non-rigid image registration using self-supervised fully convolutional networks without training data." In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 1075-1078. IEEE, 2018.
13. Jaderberg, Max, Karen Simonyan, and Andrew Zisserman. "Spatial transformer networks." Advances in neural information processing systems 28 (2015): 2017-2025.
14. Hiasa, Yuta, Yoshito Otake, Masaki Takao, Takumi Matsuoka, Kazuma Takashima, Aaron Carass, Jerry L. Prince, Nobuhiko Sugano, and Yoshinobu Sato. "Cross-modality image synthesis from unpaired data using CycleGAN." In International workshop on simulation and synthesis in medical imaging, pp. 31-41. Springer, Cham, 2018.
15. Johnson, Justin, Alexandre Alahi, and Li Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution." In European conference on computer vision, pp. 694-711. Springer, Cham, 2016.

16. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).
17. Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang et al. "Imagenet large scale visual recognition challenge." International journal of computer vision 115, no. 3 (2015): 211-252.
18. Knaan, Dotan, and Leo Joskowicz. "Effective intensity-based 2D/3D rigid registration between fluoroscopic X-ray and CT." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 351-358. Springer, Berlin, Heidelberg, 2003.
19. Sokooti, Hessam, Bob De Vos, Floris Berendsen, Boudewijn PF Lelieveldt, Ivana Išgum, and Marius Staring. "Nonrigid image registration using multi-scale 3D convolutional neural networks." In International conference on medical image computing and computer-assisted intervention, pp. 232-239. Springer, Cham, 2017.
20. `https://blogs.mathworks.com/steve/2013/08/07/tiff-bigtiff-and-blockproc/`
21. Abadi, Martín, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin et al. "Tensorflow: A system for large-scale machine learning." In 12th USENIX symposium on operating systems design and implementation (OSDI 16), pp. 265-283. 2016.
22. Kingma, D. P., & Ba, J.: Adam: A method for stochastic optimization. In: arXiv preprint arXiv:1412.6980 (2014)
23. Wang, Zhou, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. "Image quality assessment: from error visibility to structural similarity." IEEE transactions on image processing 13, no. 4 (2004): 600-612.
24. Hore, A., and D. Ziou. "Image quality metrics: PSNR vs. SSIM, in '2010 20th International Conference on Pattern Recognition'." IEEE (2010).
25. de Vos, Bob D., Floris F. Berendsen, Max A. Viergever, Marius Staring, and Ivana Išgum. "End-to-end unsupervised deformable image registration with a convolutional neural network." In Deep learning in medical image analysis and multimodal learning for clinical decision support, pp. 204-212. Springer, Cham, 2017.
26. Balakrishnan, Guha, Amy Zhao, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. "Voxelmorph: a learning framework for deformable medical image registration." IEEE transactions on medical imaging 38, no. 8 (2019): 1788-1800.
27. Marstal, Kasper, Floris Berendsen, Marius Staring, and Stefan Klein. "SimpleElastix: A user-friendly, multi-lingual library for medical image registration." In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 134-142. 2016.
28. Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. "Image-to-image translation with conditional adversarial networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1125-1134. 2017.
29. Thada, Vikas, and Vivek Jaglan. "Comparison of jaccard, dice, cosine similarity coefficient to find best fitness value for web retrieved documents using genetic algorithm." International Journal of Innovations in Engineering and Technology 2, no. 4 (2013): 202-205.
30. Huttenlocher, Daniel P., Gregory A. Klanderman, and William J. Rucklidge. "Comparing images using the Hausdorff distance." IEEE Transactions on pattern analysis and machine intelligence 15, no. 9 (1993): 850-863.
31. Wang, Xintao, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. "Esrgan: Enhanced super-resolution generative adversarial networks." In Proceedings of the European conference on computer vision (ECCV) workshops, pp. 0-0. 2018.

32. He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.

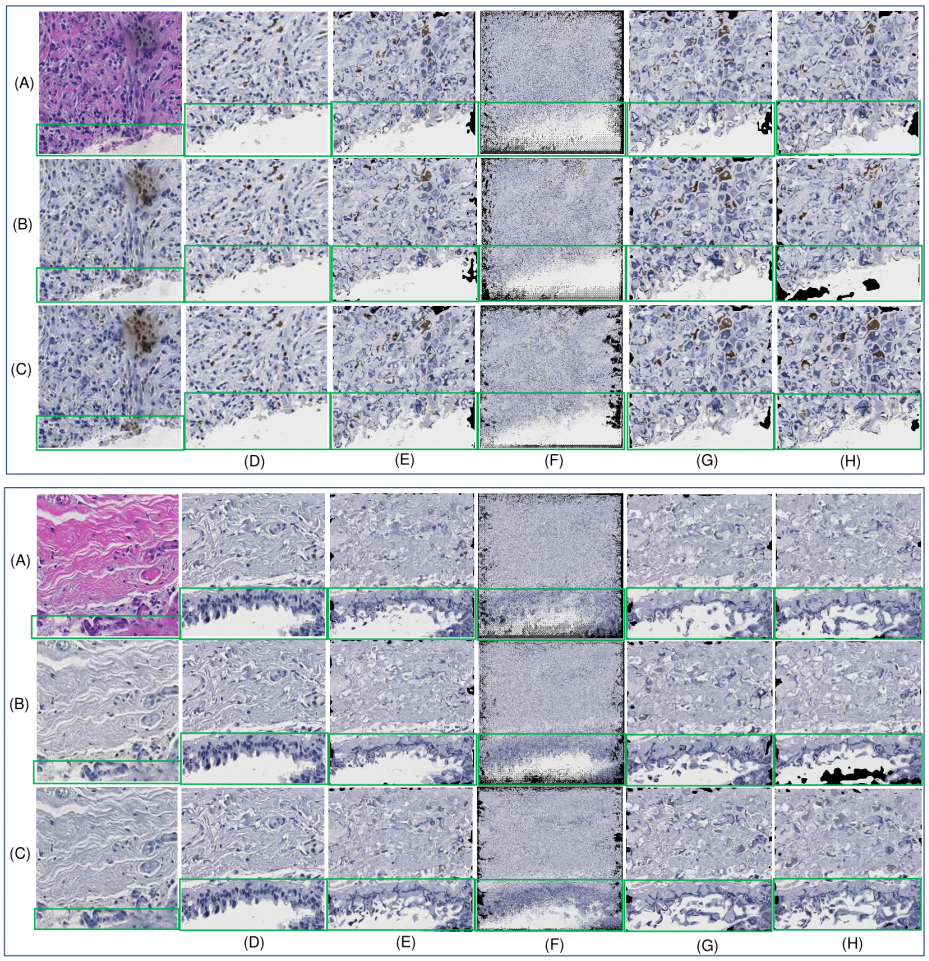

Fig. 9. Patch based registration performance with two typical image regions. (A)Fixed real H&E image; (B)Fixed synKi67 image by CycleGAN; (C)Fixed synKi67 image by the modified CycleGAN; (D)Real Ki67 moving image; Registration results by (E)DirNet, (F)VoxelMorph Unet, (G)FCN, and (H)Our proposed multi-scale FCN. A green box is used to highlight the registration result.

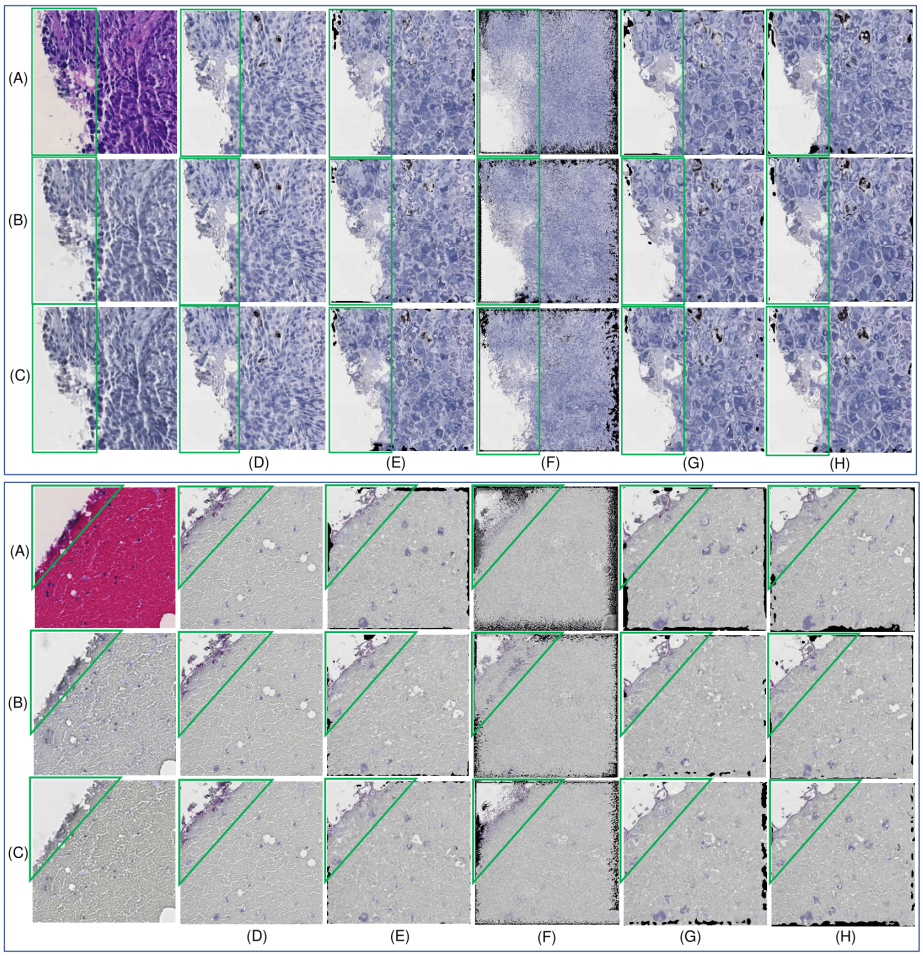

Fig. 10. Patch based registration performance with two typical image regions. (A)Fixed real H&E image; (B)Fixed synPHH3 image by CycleGAN; (C)Fixed synPHH3 image by the modified CycleGAN; (D)Real PHH3 moving image; Registration results by (E)DirNet, (F)VoxelMorph Unet, (G)FCN, and (H)Our proposed multi-scale FCN. A green box is used to highlight the registration result.

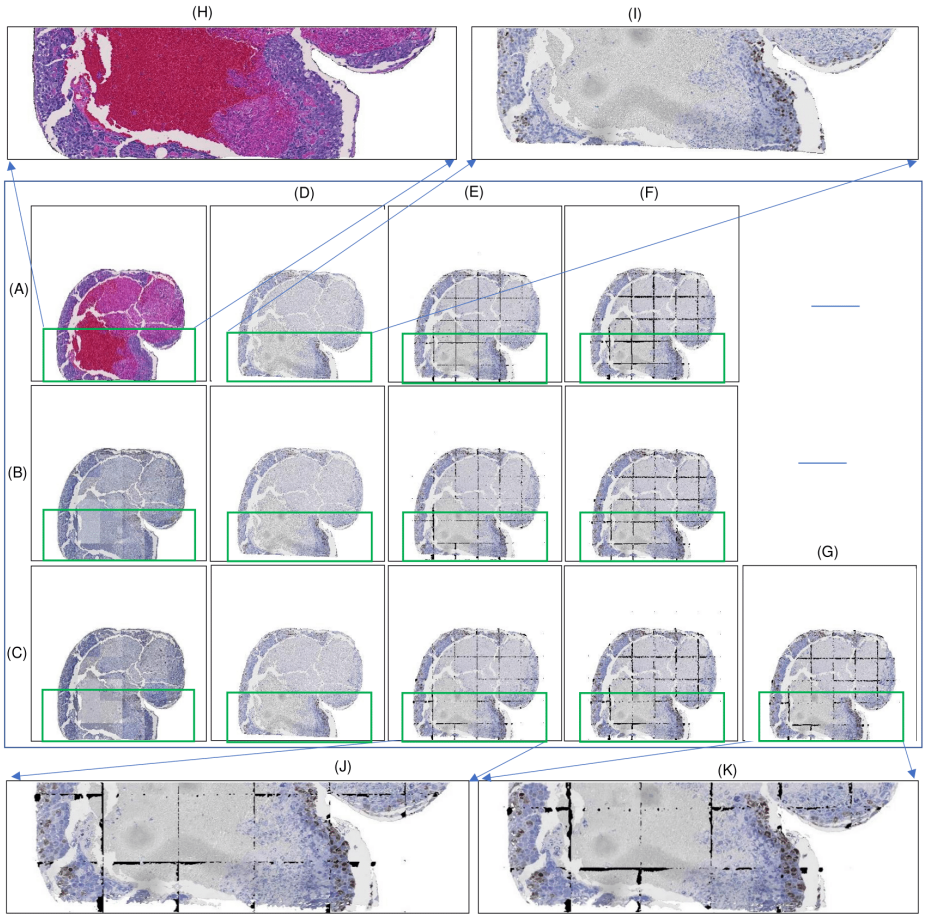

Fig. 11. Registration of ROIs from WSI blocks of H&E stain and Ki67 IHC biomarker. A green box is used to highlight the registration result.(A)Fixed real H&E image; (B)Fixed synKi67 image by CycleGAN; (C)Fixed synKi67 image by the modified CycleGAN; (D)Real Ki67 moving image; Registration results by (E)DirNet, (F)FCN, and (G)Proposed CycGANRegNet; (H)Zoomed view of green boxed region of Fixed image (Original H&E); Close-up views of (I) the moving image, (J) the registered image by DirNet with the 'syn-2' dataset, and (K) the registered image by CycGANRegNet.

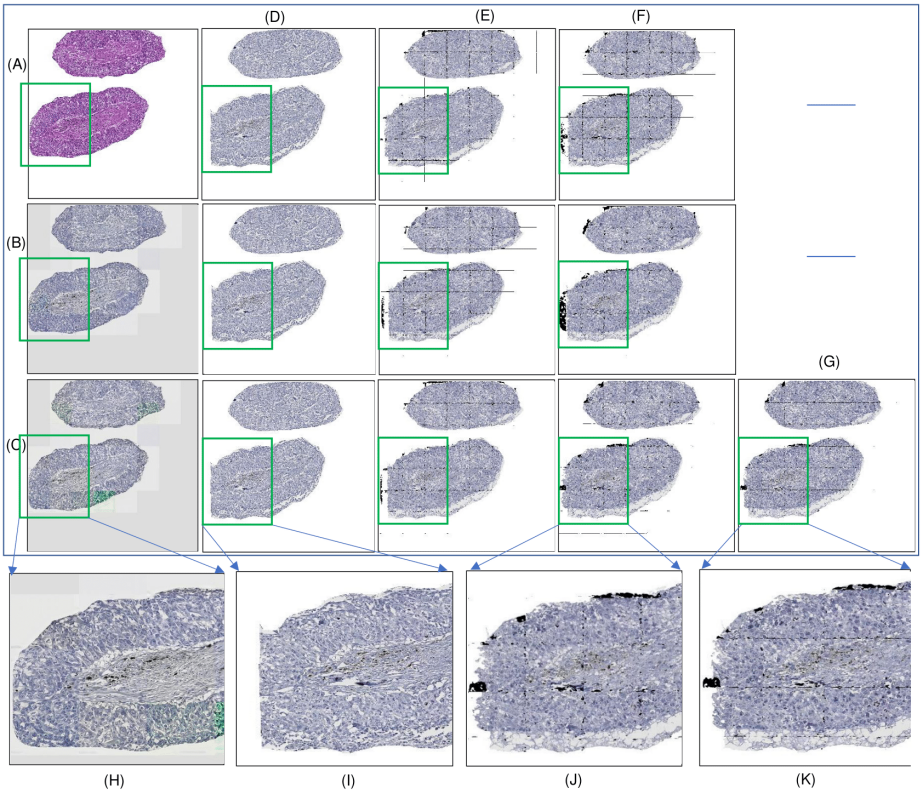

Fig. 12. Registration of ROIs from WSI blocks of H&E stain and PHH3 IHC biomarker. A green box is used to highlight the registration result.(A)Fixed real H&E image; (B)Fixed synPHH3 image by CycleGAN; (C)Fixed synPHH3 image by the modified CycleGAN; (D)Real PHH3 moving image; Registration results by (E)DirNet, (F)FCN, and (G)Proposed CycGANRegNet; (H)Zoomed view of green boxed region of Fixed image (Original H&E); Close-up views of (I) the moving image, (J) the registered image by DirNet with the 'syn-2' dataset, and (K) the registered image by CycGANRegNet.