Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.049640

ARTICLE

GAN-DIRNet: A Novel Deformable Image Registration Approach for Multimodal Histological Images

Haiyue Li[1], Jing Xie[2], Jing Ke[3], Ye Yuan[1], Xiaoyong Pan[1], Hongyi Xin[4] and Hongbin Shen[1,*]

[1]Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai, 200240, China

[2]Department of Pathology, Ruijin Hospital, Shanghai Jiao Tong University, School of Medicine, Shanghai, 200025, China

[3]Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China

[4]Global Institute of Future Technology, Shanghai Jiao Tong University, Shanghai, 200240, China

*Corresponding Author: Hongbin Shen. Email: hbshen@sjtu.edu.cn

Received: 13 January 2024 Accepted: 26 April 2024 Published: 18 July 2024

ABSTRACT

Multi-modal histological image registration tasks pose significant challenges due to tissue staining operations causing partial loss and folding of tissue. Convolutional neural network (CNN) and generative adversarial network (GAN) are pivotal in medical image registration. However, existing methods often struggle with severe interference and deformation, as seen in histological images of conditions like Cushing's disease. We argue that the failure of current approaches lies in underutilizing the feature extraction capability of the discriminator in GAN. In this study, we propose a novel multi-modal registration approach GAN-DIRNet based on GAN for deformable histological image registration. To begin with, the discriminators of two GANs are embedded as a new dual parallel feature extraction module into the unsupervised registration networks, characterized by implicitly extracting feature descriptors of specific modalities. Additionally, modal feature description layers and registration layers collaborate in unsupervised optimization, facilitating faster convergence and more precise results. Lastly, experiments and evaluations were conducted on the registration of the Mixed National Institute of Standards and Technology database (MNIST), eight publicly available datasets of histological sections and the Clustering-Registration-Classification-Segmentation (CRCS) dataset on the Cushing's disease. Experimental results demonstrate that our proposed GAN-DIRNet method surpasses existing approaches like DIRNet in terms of both registration accuracy and time efficiency, while also exhibiting robustness across different image types.

KEYWORDS

Histological images registration; deformable registration; generative adversarial network; cushing's disease; machine learning; computer vision

1 Introduction

Deep learning based multi-modal histological image registration is indispensable in medical image processing [1–4]. Intraoperative navigation, 3D modeling, and fusion analysis of information from multiple sources all necessitate registration as a prerequisite step [5–8]. Existing registration methods based on deep learning are generally tailored for the registration between the computed tomography (CT) and magnetic resonance (MR) images [6,9–12]. In contrast, histological image registration presents distinct challenges compared with CT-MR registration. While CT-MR registration deals with images of the same section but different modalities, histological images require registration between different modalities of adjacent sections, significantly complicating the task of identifying corresponding anatomical relationships [13,14]. Moreover, dyeing operations inevitably introduce local fuzziness and noise interference in histological images, distorting or blurring the original features. Current CT-MR registration methods lack optimization for these challenges or integration of a priori knowledge of such deformations into their models. Furthermore, the annotation process for histological images demands more manpower and resources compared to CT/MR images [15,16]. Consequently, existing methods struggle in histological image registration. These challenges have spurred researchers to develop a novel unsupervised deep learning method for histological image registration with severe interference and deformation.

As a model of deep learning based unsupervised registration, the deformable image registration network (DIRNet) [17] can achieve results comparable to traditional registration methods in a relatively short timeframe, without requiring manual annotation of the dataset. However, DIRNet is limited to registering images of the same modality and does not account for modal differences when registering cross-modal images. Improved variants such as EDIRNet [18], MS-DIRNet [19] and Sang's work [20] have addressed some of these limitations. Among them: EDIRNet utilizes attention mechanism to enhance the influence of edge features on registration. However, in histological images of conditions like Cushing's disease, tissue defects and overlaps caused by operations often occur at the edges, which can limit the effectiveness of edge-based methods like EDIRNet. MS-DIRNet [19] trains a combination of global and local networks, enabling the model to capture larger abdominal movements. Sang et al. [20] implicitly model the deformation prior of the spatial changes and incorporate it into the image registration framework. However, the deformation priors caused by motion that they considered are limited by the specific situations they address. For example, the abdominal movement studied by Lei et al. [19] is mainly composed of longitudinal and transverse stretching, which may not adequately address the complex deformations present in histological images. The magnetic resonance and CT images studied by Sang et al. [20] do not exhibit significant angle deviation and partial loss as in the histological images of our study. Additionally, they do not consider the extraction of modal features from multi-modal images, further limiting their applicability to multi-modal histological image registration tasks. In Table 1, we provide detailed characteristics of the aforementioned registration methods and explain why they may not be suitable for addressing challenges encountered in such situations.

Generative Adversarial Network (GAN) [21] is a generative model for deep learning trained in the form of confrontation. The generator converts random vectors into synthetic images resembling the training images, while the discriminator distinguishes between synthetic and real images. By pitting the discriminator against the generator, GAN self-supervises the training. Building upon GAN, cyclic GAN [22] has the advantage of not requiring image pairs, but forcing the deformation field to be consistent. Mahapatra et al. [23] use GAN to generate the warped image directly and prevent unrealistic registration by leveraging the consistent forced deformation of cyclic GAN. Lei et al. [19] integrate GAN into deformable image registration (DIR) to punish unrealistically warped images with additional dense displacement vector field (DVF) regularization. However, these approaches often underutilize the feature extraction capability of the GAN discriminator.

Table 1: Drawbacks of existing methods compared to our proposed method

Existing method	Characteristic	Drawback
DIRNet [17]	Assuming that the modalities of the images are the same	Difficulty in solving multi-modal registration problems
EDIRNet [18]	Enhance the edge features by attention mechanism	Difficult to handle images with edge distortion
MS-DIRNet [19]	Directly predicting deformation vector fields using the generator	No suitable for images with overlapping and missing tissues
Sang's work [20]	Incorporate implicit modeling deformation prior into DIRNet	No suitable for images with significant displacement

In summary, existing methods have struggled to accomplish deformable multi-modal registration tasks on images with significant modality differences, severe noise interference, and deformation, as evidenced by datasets like Cushing's disease [24]. To tackle these challenges, in this study, we propose a novel multimodal registration approach GAN-DIRNet, which adapts generative adversarial networks for deformable histological image registration.

Compared with previously published studies, the major contributions of this work are three-fold:

1. We designed a novel deformable multimodal image registration network based on unsupervised deep learning, which can be used for histological image registration tasks with severe noise and deformation such as the images of Cushing's disease.
2. We proposed a GAN-DIRNet embedded with discriminators for auxiliary feature extraction.
3. We proposed a novel modal-specific feature descriptor for cross-modal histological image registration.

2 Related Work

Traditional histological registration methods are mainly divided into three categories: Feature-point-based [25], gray-scale-based [26], metric-based [27] and transformation-based methods [12]. The registration method based on feature points [25,28,29] offers the advantages of rapid execution and low computational complexity. However, it requires manual setting of feature points, automatic search by algorithms, or a combination of both [30].

These systems often struggle to effectively manage significant disparities between modalities in pairs of histological images. They commonly face challenges in processing images of separate modalities into feature pairs. For instance, in the CRCS dataset [24], the reference images are stained to highlight adreno-cortico-tropic-hormone (ACTH) in the cytoplasm, whereas the moving images are stained to emphasize the transcription of ACTH-secreting tumor cells (t-pit) in the nucleus. Automatically identified feature points tend to focus on the nucleus of the moving image, while the corresponding positions in the fixed image are ignored due to staining defects. In essence, feature-point-based methods either necessitate meticulous manual intervention or exhibit limited reliability in handling multi-modal images.

The transformation-based approach initially involves converting one image modality to the other [31], or both modalities to a shared latent space [32]. Subsequently, registration occurs within this common feature space. Specifically, DTR-GAN [12] utilizes a bidirectional translation network to convert multi-modal images into the same modality before performing single-modal registration. Similarly, Deform-GAN [33] proposed by Zhang et al. employs the generator in GAN to transform multi-modal similarity into single-modal similarity, thereby enhancing precision. However, our images for registration employ different staining methods (one staining the nucleus and the other staining the cytoplasm), making direct conversion challenging due to substantial differences. Metric-based methods sidestep these challenges by selecting metrics less sensitive to modal variations for optimization, such as mutual information (MI) [27], regional MI, mean square error (MSE), and normalized cross correlation (NCC) [34,35].

To alleviate the need for annotation workload, unsupervised or weakly-supervised deep learning models have been deployed in multi-modal image registration. For instance, DIRNet [17], introduced by Vos et al., an unsupervised deep learning model for deformable image registration, and ADMIR [36], designed for end-to-end unsupervised affine and deformable registration of brain magnetic resonance images. A semi-supervised model proposed by Liu et al. [30] offers a method to progressively increase the number of landmarks from a limited set of annotations. However, the fundus images they analyze are relatively easy to annotate manually, unlike the histological images in our study. Among all the unsupervised models, DIRNet [17] bears the closest resemblance to our approach. Although DIRNet is adaptable to multi-modal histological image registration, it lacks distinct modules for handling moving and fixed images separately. Essentially, it overlooks the modalities' disparities and treats them as a single-modal image registration task. While DIRNet performs adequately on the MNIST [37] dataset, our experiments reveal its subpar performance in multi-modal histological image registration. As a remedy, we propose integrating an additional feature extraction module at the outset. Subsequent experiments in later sections demonstrate that this multi-modal feature extraction module significantly enhances the accuracy of both single-modal and multi-modal image registration.

3 Methodology

3.1 Experimental Materials

To showcase the effectiveness of our proposed method, GAN-DIRNet, we conducted evaluations on multiple datasets, including MNIST, the CRCS dataset, and the Automatic Non-rigid Histological Image Registration (ANHIR) dataset, comprising eight subsets such as breast, lung lobes, gastric, kidney, etc.

In CRCS-dataset, histological sections with adreno-cortico-tropic-hormone (ACTH) and t-pit (the transcription of ACTH-secreting tumor cells) from 35 Cushing's disease patients was retrospectively collected. Each pair of images slated for registration had at least 5 pairs of landmarks implanted to calculate the target registration error and facilitate registration evaluation.

The ANHIR-dataset, introduced by Borovec et al. [15], encompasses eight subsets of histological images namely breast, chronic obstructive air way disease (COAD), gastric, kidney, lung lesion, lung lobes, mammary-gland, and mice-kidney respectively. It comprises a total of 481 cross-modal image pairs sourced from 18 different tissue staining operations. On average, 86 landmarks were placed on each image by 9 annotators, which can be publicly accessed together with the images.

3.2 The Proposed Method

Overall, we innovatively designed the discriminators of two GANs as embedded pre-trained parallel networks and integrated them into the input part of DIRNet as the modal feature extraction layer to enhance cross-modal registration outcomes.

The original version of DIRNet comprises three components: A convolutional neural network, a spatial transformation network, and a resampling network. Convolutional neural networks translate the input fixed and moving images into local deformation parameters. Subsequently, the spatial transformation network receives these parameters and generates a dense displacement vector field. Finally, the resampling network distorts the moving image into the warped image based on the dense displacement vector field. However, the input layers for receiving moving images and fixed images in DIRNet are identical, indicating a lack of consideration for modal differences in cross-modal image registration.

Our proposed method GAN-DIRNet comprises of three modules: GAN training, rigid registration, and deep-learning based non-rigid registration, as illustrated in Fig. 1. During the initial stage of GAN training, the training sets of moving images and fixed images are individually inputted into two identical generative adversarial networks for model training. Following adversarial training, two implicit feature extraction modules corresponding to each training set are derived from the discriminators of the two GAN networks, respectively.

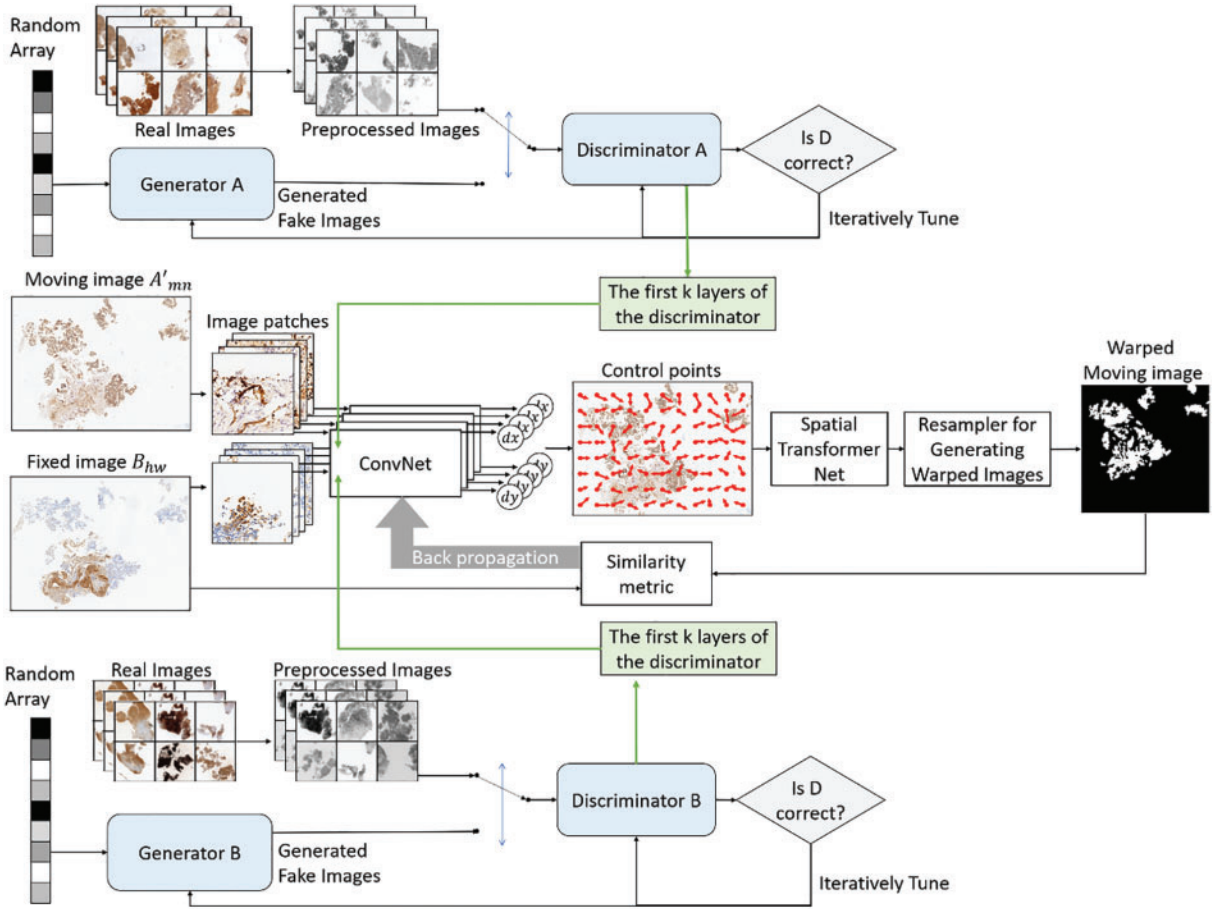


Figure 1: Schematic diagram of the principle of our proposed method

The second stage involves implementing a traditional iterative rigid registration method for the image pairs intended for registration, aiming to acquire a transformation matrix containing explicit spatial information such as translation and rotation. However, for handwritten dataset registration, this second step can be omitted since there are no rotation or translation transformations involved.

In the third stage, the first few layers of the discriminators in GAN pre-trained by moving and fixed images are embedded into the input layers of moving and fixed images in DIRNet, respectively. Subsequent to inputting the resultant image of rigid registration from the second stage, a displacement vector field (DVF) generated by GAN-DIRNet performs non-rigid deformation on the moving image via an interpolation network, with the process recorded as *Warp*. The similarity between the warped image and fixed image is then backpropagated to GAN-DIRNet for unsupervised training, aiming to generate a more suitable DVF. Finally, the moving image and the trained DVF generate new non-rigidly deformed images through interpolation networks to participate in the subsequent iteration.

1) Firstly, the moving and fixed images, preprocessed from color to grayscale, are separately inputted into GAN training. Generator A and Generator B generate fake images from random sequences that closely resemble their respective input images, while Discriminator A and Discriminator B distinguish between real images and the fake images generated by their corresponding generator. During adversarial training, the feature descriptors of the multi-modals (moving and fixed images) are extracted by the discriminators in two GANs separately. When dealing with single-modal images from the MNIST dataset, parameter k is set to 5, whereas for histological images, parameter k is set to 3. 2) The modal feature descriptors are then connected to the feature input layers of DIRNet respectively. 3) For incomplete slice images, adaptive threshold filling is conducted on the border part to mitigate the impact of feature points created by cutting and edges.

3.3 GAN Training

During GAN training, the training sets of moving and fixed images are individually inputted into two identical GAN networks for training, aiming to acquire the implicit feature representation of the respective image modalities. The GAN within GAN-DIRNet comprises a generator and a discriminator, both of which are constructed using convolutional neural networks (CNN), as depicted in Fig. 1.

In implementation, the generator consists of five transposed convolutional layers with a kernel size of 4 × 4, as illustrated in Fig. 2. Batch normalization [38] and rectified linear activation function (ReLU) are applied throughout the generator, except for the last layer which utilizes a tanh activation function. Conversely, the discriminator comprises five convolutional layers with the same kernel size as the generator. The final layer of the discriminator adopts a sigmoid activation function, while the other layers employ leaky ReLU. Similarly, the outputs of the middle three layers of the discriminator undergo batch normalization.

3.4 Rigid Registration

The rigid registration module utilized in this article adopts the registration method outlined in the CRCS pipeline [24]. Its specific procedure is as follows. Assuming a moving image with length m and width n is named A_{mn}. Correspondingly, the fixed image is named B_{hw}, and the resultant image processed by the rigid transformation is represented as $A'_{mn} = R(A_{mn}, B_{hw})$, where R refers to the rigid registration module.

The rigid registration module R in this article leverages a two-stage traditional iterative method from coarse to fine, as shown in Fig. 3a. R follows the pyramid hierarchy concept akin to Scale Invariant Feature Transform (SIFT) [28] and performs preliminary registration R_{coarse} on images with lower resolutions. The output, containing coarse rotation angle and displacement information, is used as the initial matrix to participate in the downstream fine registration R_{fine}. The combined process of R_{coarse} followed by R_{fine} constitutes the entirety of the rigid registration process named R.

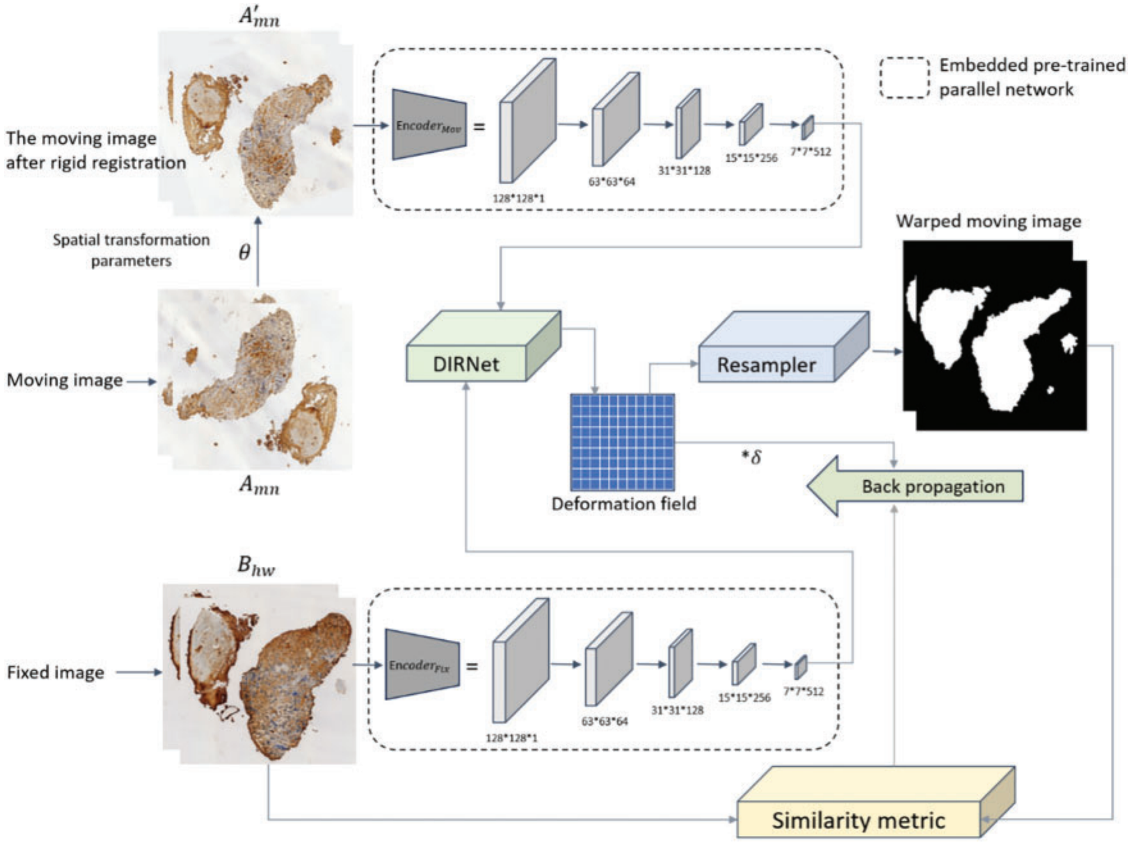
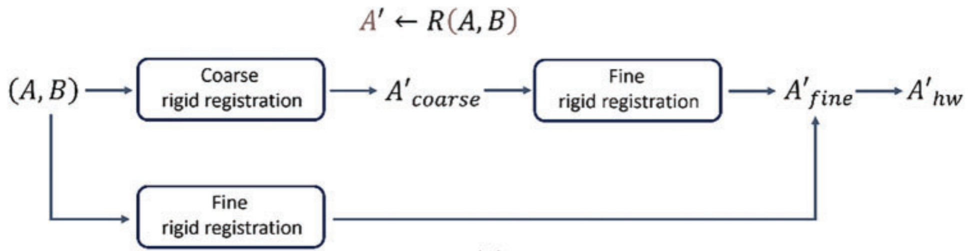


Figure 2: Schematic diagram illustrating the process of embedding the modal feature extraction module into DIRNet for model training. θ represents the initial rigid registration matrix of the moving image B. δ is the parameter for adjusting the weights of the similarity term and the smoothing term. The two sections highlighted by dashed lines depict the embedded pre-trained parallel networks



(a)

Figure 3: (Continued)

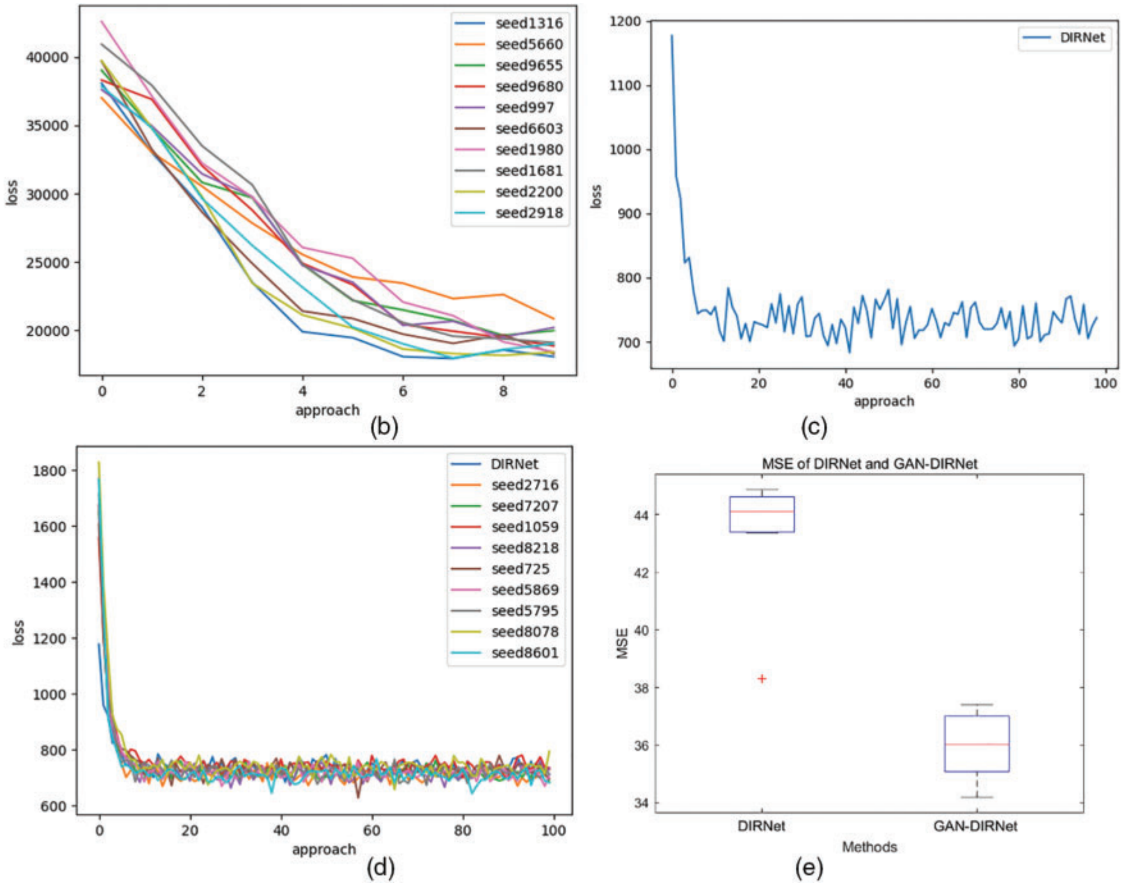


Figure 3: (a) Schematic diagram of the rigid registration module and its connection with upstream and downstream tasks. (b): The descent curves of loss during the optimization processes of GAN-DIRNet. The horizontal axis represents the number of iteration steps, while the vertical axis represents the value of the loss function, detailed in Section 3.5.2. (c): The descent curve of loss during the optimization process of DIRNet. (c) is a descent curve reproduced from the original version of the DIRNet code. In order to eliminate the impact of randomness on the results, we conducted tests on GAN-DIRNet using 10 different random seeds. Thus, there are ten descent curves in (b), whereas (c) only presents one. For comparative purposes, we retested DIRNet with 9 additional random seeds and displayed their descent curves alongside the original curve in (d). The loss function records the total loss considering the gradient accumulation and the regularization term. Its numerical value is also influenced by the image size (16 ∗ 16 pixels for the original DIRNet, while 64 ∗ 64 pixels for the proposed GAN-DIRNet). Therefore, the schematic diagrams of the loss function (b–d) only show their convergence rates; their values cannot be horizontally compared as in (e). (e): The boxplots represent the MSE errors on the MNIST dataset using the DIRNet and GAN-DIRNet registration models from left to right, respectively. There is a standardized measurement rule for the results of different methods before calculation, enabling horizontal comparison of their values. The sole outlier represents the sample from Vos' work [17]. Among the ten samples from each of the two groups, even the worst performing point in the proposed model exhibits a smaller error than the best outlier in the original DIRNet

Two optimization algorithms are employed separately in the different branches of the proposed integration method: Gradient descent and one plus one evolution. Among them, mean square error (MSE) is combined with gradient descent optimization as the similarity metric, while mutual information is utilized in conjunction with the other optimization algorithm.

The specific optimization objectives for the two stages are articulated as follows:

$$A'_{coarse} = \underset{A'_{coarse}=R_{coarse}(A_{mn})}{\operatorname{argmin}} \frac{1}{h*w} \sum\nolimits_{A',B} (a'-b)^2 \; a' \epsilon A'_{coarse}, b \epsilon B_{hw} \tag{1}$$

$$A'_{fine} = \underset{A'_{fine}=R_{fine}(A'_{coarse})}{\operatorname{argmin}} \sum\nolimits_{A',B} p\left(A'_{fine}, B_{hw}\right) \log\left(\frac{p\left(A'_{fine}, B_{hw}\right)}{p\left(A'_{fine}\right) p\left(B_{hw}\right)}\right) \tag{2}$$

$$A'_{hw} = \underset{A'_{hw}=R(A_{mn})}{\operatorname{argmin}} \sum\nolimits_{A',B} (a'-b)^2 \; a' \epsilon A'_{hw}, b \epsilon B_{hw} \tag{3}$$

where $p\left(A'_{fine}, B_{hw}\right)$ denotes the joint probability density function, while $p\left(A'_{fine}\right)$ and $p\left(B_{hw}\right)$ represent the marginal probability density functions.

Furthermore, during the registration process, various parameter settings are applied to each input pair of images to be registered. The optimal result from all the calculation branches is counted as the global optimal result A'_{hw}, with the corresponding optimal rigid transformation matrix denoted as θ.

θ is a 9 × 9 matrix for rigid transformation. In the specific computation, θ transforms the pixel at coordinate (m, n) in $A_{mn}$to the position at coordinate (h, w) in A'_{hw}.

$$\begin{bmatrix} h \\ w \\ 1 \end{bmatrix} = \begin{bmatrix} cos\gamma & sin\gamma & x \\ -sin\gamma & cos\gamma & y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} m \\ n \\ 1 \end{bmatrix} \tag{4}$$

where x and y respectively denote the translation amount of the pixel in both horizontal and vertical directions, and γ represents the rigid rotation angle of the entire image.

3.5 GAN-DIRNet

3.5.1 Embedded Feature Encoder Module

The proposed GAN-DIRNet augments the DIRNet network architecture with a modal feature extraction module utilizing GAN's discriminator. Unlike traditional DIRNet which treats moving and fixed images equally, our method pre-extracts modal features from both types of images and integrates them into the input layer of DIRNet. The input layers of the moving and fixed images share a similar structure derived from GAN's discriminators.

Conceptually speaking, our network consists of encoders used to extract feature maps from a given input image. The module for extracting features of moving image is called $Encoder_{Mov}$, while that of the fixed image is called $Encoder_{Fix}$, as illustrated in Fig. 2.

The simplified description of the entire process is as follows:

$$\begin{cases} A'_{hw} \leftarrow R\left(A_{mn}, B_{hw}\right) \\ Encoder_{mov} \leftarrow GAN\left(A_{mn}\right) \\ Encoder_{fix} \leftarrow GAN\left(B_{hw}\right) \\ Layers_{mov} \leftarrow Encoder_{Mov}(A'_{hw}) \\ Layers_{fix} \leftarrow Encoder_{Fix}(B_{hw}) \\ DVF \leftarrow DIRNet\left(Layers_{mov}, Layers_{fix}\right) \\ A''_{hw} \leftarrow Warp(A'_{hw}, DVF) \end{cases} \tag{5}$$

where $A'_{hw} \leftarrow R\left(A_{mn}, B_{hw}\right)$ represents the process of obtaining A'_{hw} after rigid registration between the moving image A_{mn} and fixed image B_{hw}. R transforms the registered A_{mn} to the same size as B_{hw} with the result denoted as A'_{hw}. $Encoder_{Mov}$ and $Encoder_{Fix}$ denote respectively the feature extraction layers trained by the moving and fixed image datasets through GAN. They are embedded in DIRNet in parallel to concatenate with two input layers, and their parameters also participate in the iterative optimization of the neural network. $Layers_{mov}$ and $Layers_{fix}$ are the feature descriptors extracted by $Encoder_{Mov}$ and $Encoder_{Fix}$ on A'_{hw} and B_{hw}. DIRNet takes on the results ($Layers_{mov}$ and $Layers_{fix}$) from the feature extraction layers and finally generates the DVF. $Warp(A'_{hw}, DVF)$ represents the process in which the moving image undergoes non-rigid deformation of the DVF generated by GAN-DIRNet through an interpolation network. A''_{hw} denotes the final result of A_{mn} after rigid and non-rigid registration. In an ideal state, the anatomical position of A''_{hw} should be consistent with B_{hw}.

3.5.2 Mutually Reinforcing Registration and Encoder Module

During the training phase, modal feature extraction and deformable image registration are distinct processes. The modal feature modules of moving and fixed images are trained independently by two GANs with identical structures. Conversely, the image registration network is trained by simultaneously inputting images from both modalities. This setup entails optimizing the modal feature extraction model twice during the training of the image registration network.

In backpropagating through GAN-DIRNet, it is crucial to balance considerations for both similarity measures and the smoothness of the transformation field. As a result, our loss function is defined as follows:

$$L_{GAN-DIRNet} = \sum\nolimits_{A''_{hw}, B_{hw} \in batch} \left\| \nabla A''_{hw} - \nabla B_{hw} \right\|_2 + \delta \left\| DVF \right\|_2 \tag{6}$$

In the defined loss function, the first term represents the similarity measure, while the second term serves as a smoothing component. The parameter δ adjusts the weight of the smoothing term. Given that rigid registration precedes the process, the anticipated deformation vector field (DVF) matrix should exhibit sparsity. Therefore, we assign a relatively high weight of 500 to δ, ensuring that both the similarity and smoothing terms possess comparable magnitudes.

3.6 Performance Metrics for Image Registration

The quantitative indicator utilized in this article is the relative target registration error (rTRE) evaluation system [15], which is assessed following the evaluation system proposed in the Automatic

Non-rigid Histological Image Registration Challenge (ANHIR). The calculation method is outlined as follows:

In principle, the image pair (A, B) undergoing registration should exhibit identical anatomical structures. Manually placed landmarks on the two images, denoted as m^A and m^B, respectively, should correspond to the same anatomical locations. Specifically, 5 pairs of landmarks are positioned on each image pair, denoted as $(m_k^A, m_k^B)_{k=1-5}$. The number of landmarks may vary across datasets, and the collection of landmarks in the same image pair (A, B) is denoted as $K_{A,B}$.

The warped image is denoted as A'' with the landmarks on it transformed to $\tilde{m}_k^A$. The Euclidean distance between $\tilde{m}_k^A$ and m_k^B reflects the accuracy of the registration. For a single landmark, the relative Target Registration Error (rTRE) is defined as Eq. (7) [15].

$$rTRE_k^{AB} = \frac{\left\| \tilde{m}_k^A - m_k^B \right\|_2}{d_B} \tag{7}$$

where d_B is the diagonal length of B. The response of an individual landmark to the overall configuration effect may not be comprehensive. The set of all image pairs to be registered is denoted as I. To provide a more comprehensive evaluation of the overall registration performance of each image pair $(A, B) \in I$, six different measurement standards have been extended based on $rTRE$ [15] namely $AMrTRE$, $MMrTRE$, $AMxrTRE$, $MMxrTRE$, $AArTRE$ and $MArTRE$, respectively. The specific calculation methods for these 6 indicators are paired with each other, as shown from Eqs. (8) to (13) [15].

$$AMrTRE\left(T_I\right) = \underset{(A,B)\in I}{\text{mean}} \mu^{A,B}{}_{med}\left(A, B\right) \tag{8}$$

$$MMrTRE\left(T_I\right) = \underset{(A,B)\in I}{\text{median}} \mu^{A,B}{}_{med}\left(A, B\right) \tag{9}$$

$$where\ \mu^{A,B}{}_{med}\left(A, B\right) = \underset{k\in K_{A,B}}{\text{median}}\ rTRE_k^{AB}\left(A, B\right) \tag{10}$$

where T_I represents the registration method T applied on dataset I. The top two indicators first calculate the median of $rTRE$ in a single image, and then calculate the median or average of μ values in all images, which have the advantage of strong anti-interference ability. The middle two indicators, $AMxrTRE$ and $MMxrTRE$, first determine the maximum value of $rTRE$ in a single image, and then repeat the same calculation as Eqs. (8) or (9) above, which have the advantage of fully exposing registration errors.

$$AMxrTRE\left(T_I\right) = \underset{(A,B)\in I}{\text{mean}} \mu^{A,B}{}_{max}\left(A, B\right) \tag{11}$$

$$MMxrTRE\left(T_I\right) = \underset{(A,B)\in I}{\text{median}} \mu^{A,B}{}_{max}\left(A, B\right) \tag{12}$$

$$where\ \mu^{A,B}{}_{max}\left(A, B\right) = \underset{k\in K_{A,B}}{\max} rTRE_k^{AB}\left(A, B\right) \tag{13}$$

Similarly, the last two indicators replace the initial step of calculating the $rTRE$ of a single image with the average value, which compensates for the first two indicators.

4 Results and Discussion

4.1 Results on MNIST Dataset

We initially benchmarked GAN-DIRNet on the MNIST dataset. Following the original setting in DIRNet [17], we randomly sampled the training and testing datasets from MNIST. For reproducibility and consistency, we benchmarked both DIRNet and our in-house GAN-DIRNet over ten preset random seeds, as shown in Figs. 3b–3e. Among them, Figs. 3b–3d display the comparison of the loss updates during training of DIRNet and GAN-DIRNet. Fig. 3e presents the boxplots representing the MSE errors on the MNIST dataset using the DIRNet registration model and GAN-DIRNet registration model from left to right, respectively. The solitary outlier denotes the sample code from the paper of DIRNet [17]. The red "+" outlier in Fig. 3e represents the reproduction of the original DIRNet code, which is constrained to use an initial seed of 0. We replaced 9 additional random seeds during the reproduction of DIRNet in comparison with the 10 random-seed-results of GAN-DIRNet.

The box diagram provides an intuitive observation: Out of the ten initial seeds randomly selected from each of the two groups, even the worst performance in GAN-DIRNet is on par with the best performance of the original DIRNet, amounting to a 10% improvement on average. In addition, GAN-DIRNet is 22% faster than DIRNet (as shown in Table 2), and converges much more rapidly during training. GAN-DIRNet achieves a similar performance to DIRNet after just 100 iterations compared to the 1000 iterations required by DIRNet. The experimental results are listed in Table 3.

Table 2: Comparison of the running time and iteration times between the proposed method and baseline testing. All experiments were conducted on the same computer, with a processor model of 11th Gen Intel(R) Core(TM) i7-1165G7 @ 2.80 GHz 2.80 GHz

Method	Dataset	Run time	Iteration
DIRNet [17]	MNIST	22:25:22 min	10000
GAN-DIRNet	**MNIST**	**17:09:98 min**	**1000**

Table 3: Registration results of different methods on ten datasets. The bold face indicates the results of our proposed method. There are no manually placed landmarks in the MNIST dataset, therefore six rTRE-based metrics are not available for this dataset

Dataset	Method	rTRE						MSE
		AMrTRE	AMxrTRE	AArTRE	MMrTRE	MMxrTRE	MArTRE	
MNIST	DIRNet							0.150212118
	GAN-DIRNet							**0.137553356**
CRCS	DIRNet	0.230399	0.319151	0.232325	0.228503	0.388766	0.252475	0.238530677
	Rigid-DIRNet	0.142981	0.200086	0.148708	0.055322	0.093152	0.063625	0.178690915
	Rigid	0.170412	0.234098	0.175051	0.06397	0.13253	0.071767	0.102450668
	GAN-DIRNet	**0.143314**	**0.199770**	**0.148539**	**0.055081**	**0.093724**	**0.063144**	**0.099281044**
	UPENN	0.352534	0.511205	0.365368	0.347191	0.508108	0.401134	0.53519053
	demon	0.280942	0.433673	0.28846	0.29884	0.475925	0.311924	0.350789355
Breast	DIRNet	0.33326	0.487292	0.328663	0.390716	0.583903	0.378626	0.170121783
	GAN-DIRNet	**0.334126**	**0.488629**	**0.329621**	**0.392351**	**0.585489**	**0.379907**	**0.091640242**
COAD	DIRNet	0.057767	0.101866	0.057124	0.046024	0.078643	0.045257	0.206254468
	GAN-DIRNet	**0.058091**	**0.101874**	**0.057299**	**0.046388**	**0.077991**	**0.045648**	**0.109428424**

(Continued)

Table 3 (continued)

Dataset	Method	rTRE						MSE
		AMrTRE	AMxrTRE	AArTRE	MMrTRE	MMxrTRE	MArTRE	
Gastric	DIRNet	0.205734	0.359673	0.208762	0.073893	0.139795	0.075592	0.212041292
	GAN-DIRNet	**0.20643**	**0.360078**	**0.209013**	**0.076766**	**0.141945**	**0.076726**	**0.124060458**
Kidney	DIRNet	0.017235	0.026701	0.017196	0.00749	0.018309	0.008225	0.082842371
	GAN-DIRNet	**0.017132**	**0.027029**	**0.017703**	**0.007629**	**0.018162**	**0.008184**	**0.057610677**
Lung-lesion	DIRNet	0.03664	0.056628	0.03674	0.029954	0.054866	0.03169	0.310328106
	GAN-DIRNet	**0.03605**	**0.055896**	**0.036084**	**0.029573**	**0.055565**	**0.030856**	**0.241889585**
Lung-lobes	DIRNet	0.018092	0.044062	0.019225	0.015913	0.032931	0.017035	0.21709511
	GAN-DIRNet	**0.018208**	**0.043443**	**0.019295**	**0.015427**	**0.03269**	**0.016975**	**0.120533902**
Mammary-gland	DIRNet	0.029392	0.062695	0.030781	0.0269	0.057971	0.03052	0.100423656
	GAN-DIRNet	**0.02914**	**0.062318**	**0.030535**	**0.02737**	**0.059176**	**0.029962**	**0.059707542**
Mice-kidney	DIRNet	0.069546	0.169787	0.073018	0.074814	0.20081	0.08207	0.166053922
	GAN-DIRNet	**0.070252**	**0.170983**	**0.073512**	**0.075418**	**0.20354**	**0.082382**	**0.083607792**

4.2 Results on Histological Image Dataset

For histological images, we firstly benchmarked both DIRNet, its variant Ridge-DIRNet, and our adaptation: GAN-DIRNet over the CRCS dataset to validate the advantages of GAN-DIRNet compared to DIRNet. Additionally, we conducted comparison with other 2 existing methods: UPENN [15] and demon [39] on CRCS, to ascertain the advantages of GAN-DIRNet compared to the most commonly used methods currently available. Specifically, UPENN is an unsupervised method that integrates multiple tools for histological image registration and achieved the second place in the ANHIR competition. While demon performs well in medical image registration such as magnetic resonance images, which also does not require manual annotation. These two methods exhibit strengths in different scenarios. As comparative experiments, they serve to confirm the unique advantages of GAN-DIRNet on histological image datasets such as CRCS from various perspectives.

Fig. 4 illustrates the registration results of the respective methods over the CRCS images. As Fig. 4 depicted, Rigid-GAN-DIRNet (simplified as GAN-DIRNet in this paper) corrected the local deformation between the registered images on the basis of rigid registration, whereas DIRNet and its variant failed to do so. Furthermore, it is evident that the latter two methods struggled to identify accurate rotation angles and translation components.

In addition to visual assessment, we also conducted quantitatively analysis using two types of evaluation metrics: The landmark-based metrics (six indicators based on rTRE mentioned above) and the region-based metric, as shown in Table 3. We adopt MSE as the region-based metric by calculating the average of the sum of the squared errors obtained by subtracting a fixed image from a warped image, as illustrated in Eq. (14). The landmark-based metrics reflect the registration effect from a local perspective, while the region-based metric provides a global assessment.

$$MSE = \frac{1}{K \times h \times w} \sum_{i=1}^{K} \left(\hat{A}_i - B_i\right)^2 \tag{14}$$

where K represents all pairs of images in the dataset with each pair indexed as i. $\hat{A}_i$ represents the warped image, B_i represents the fixed image. h and w are the height and width of the image, respectively (Both of the images are normalized to 64 × 64 pixels before being subtracted pixel by pixel).

In order to avoid inter-modal interference, the landmarks on image pairs of the CRCS dataset are manually placed at edges, as depicted in Fig. 5, which may not comprehensively reflect the registration effect. To compensate for the shortcomings of rTRE, we also introduce a universal evaluation index named mean square error (MSE) for regional area. From Table 3, it is evident that the proposed scheme demonstrates the best comprehensive performance across the ten datasets. Rigid-GAN-DIRNet achieves nearly identical performance measured by landmark-based evaluation metrics compared to Rigid-DIRNet (a method embedding only the proposed rigid registration module in DIRNet without the feature extraction module), but its advantage in the evaluation index based on regional area is obvious.

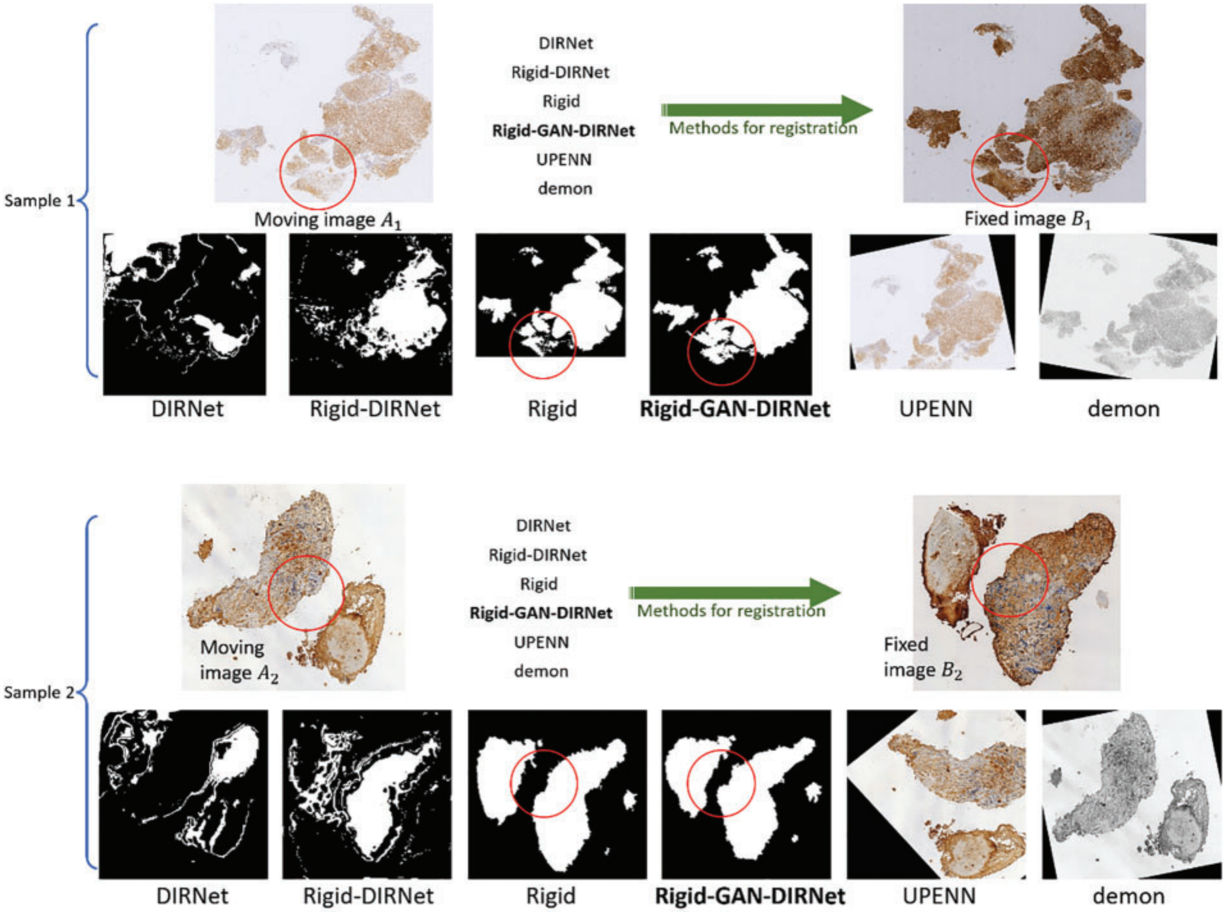


Figure 4: (Continued)

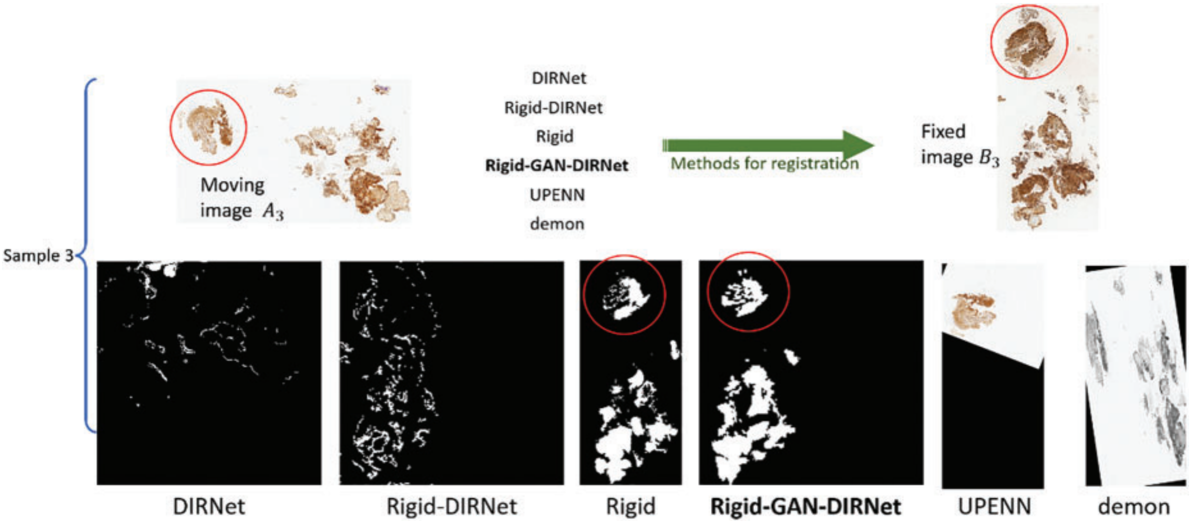


Figure 4: Comparison of the visualization results of the various registration methods. For simplicity, A represents a collective term for A_1, A_2, A_3, and B for B_1, B_2, B_3. The first two images in each group are the moving image A and the fixed image B. The second row of each group displays the resulting images of the moving images distorted by DVF. From left to right are: 1) the direct registration results of the original version of DIRNet, 2) the registration results of rigid registration concatenate with DIRNet, namely Rigid-DIRNet, 3) the registration results of rigid registration A', 4) the registration results A'' of rigid registration followed by GAN-DIRNet, namely Rigid-GAN-DIRNet, 5) the registration results of UPENN, 6) the registration results of demon. The red circles indicate the non-rigid deformation between the image pairs, which is not within the processing range of the rigid registration. However, the proposed GAN-DIRNet made revisions that are more in line with the ground truth based on rigid registration

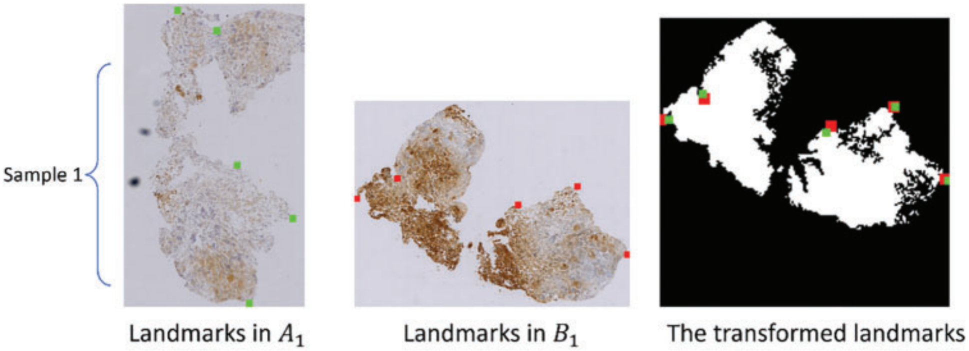


Figure 5: (Continued)

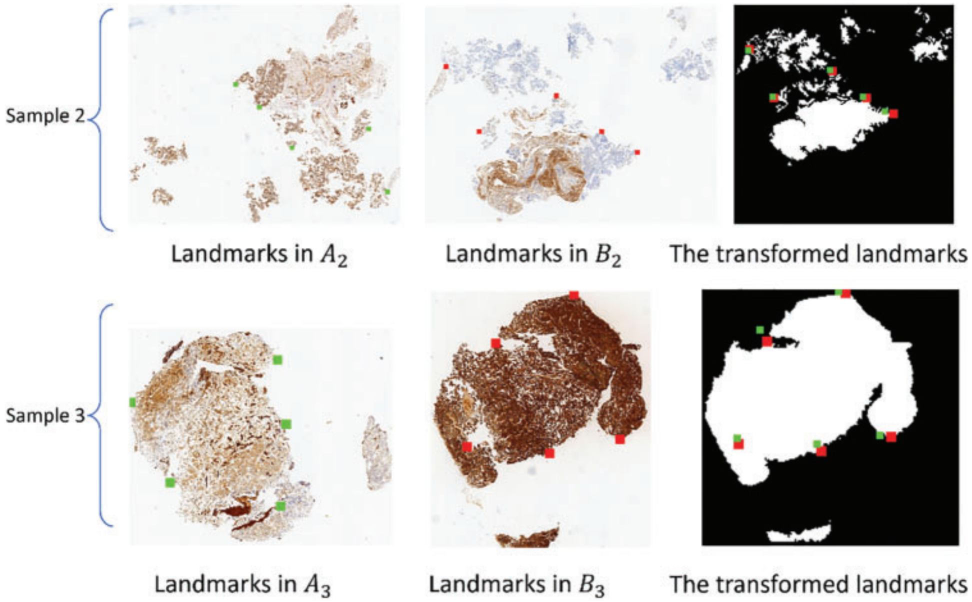


Figure 5: The first column (A_1, A_2, A_3) and the second column (B_1, B_2, B_3) respectively display the original moving/fixed images of each sample pair (A_1, A_2, A_3 denote the moving images and B_1, B_2, B_3 denote the fixed ones) with the manually annotated landmarks on them. The third column shows the transformed landmarks of A_1, A_2, A_3 (the green square dots) and the landmarks of B_1, B_2, B_3 (the red square dots). For simplicity, A represents a collective term for A_1, A_2, A_3, and B for B_1, B_2, B_3. Due to the input requirements of DIRNet, both A and B have a unified size of 180 × 180 before being sent into the network. Specifically, after A and B are proportionally scaled equally, adaptive threshold filling is performed based on their backgrounds to achieve the uniform size. The output image is generated by the DVF transformation of the same size as the input image. Therefore, the size of each third-column-image is also 180 × 180, which is different from A

4.3 Discussion

In this study, we propose a novel method GAN-DIRNet to integrate the modal-specificity information and rigid registration modules into the DIRNet for deformable image registration. The results demonstrate that GAN-DIRNet achieves more accurate and faster performance than existing methods. To investigate the superiority of GAN-DIRNet, we visualize the descriptors extracted by GAN-DIRNet through t-SNE (Fig. 6). Among them, Figs. 6a and 6b represent visualizations of the test subset of the MNIST dataset and its trained descriptors through GAN-DIRNet (i.e., the output of the GAN module and the input of the DIRNet module), respectively. By comparison, it can be seen that the latter exhibits better aggregation than the former. Despite MNIST being a single-modal image dataset, the discrimination between the ten categories of the feature descriptors output by the embedded modal feature extraction layer of GAN-DIRNet is higher than that of the unprocessed original image data, which proves the powerful feature extraction capability of GAN-DIRNet.

Fig. 6c illustrates a visualization of the descriptors trained on GAN-DIRNet for the CRCS dataset, where the first label represents all t-pit nuclear-staining moving images, and the second label represents all ACTH cytoplasm-staining fixed images. The difference in the distribution of the two reflects variance between the modal features of the two differently stained images. Quantitative and qualitative analysis confirms that the final registration effect of our proposed GAN-DIRNet is superior to that of the original DIRNet, which may be attributed to the function of the embedded modal feature extraction layer for differentiated processing of multi-modal images.

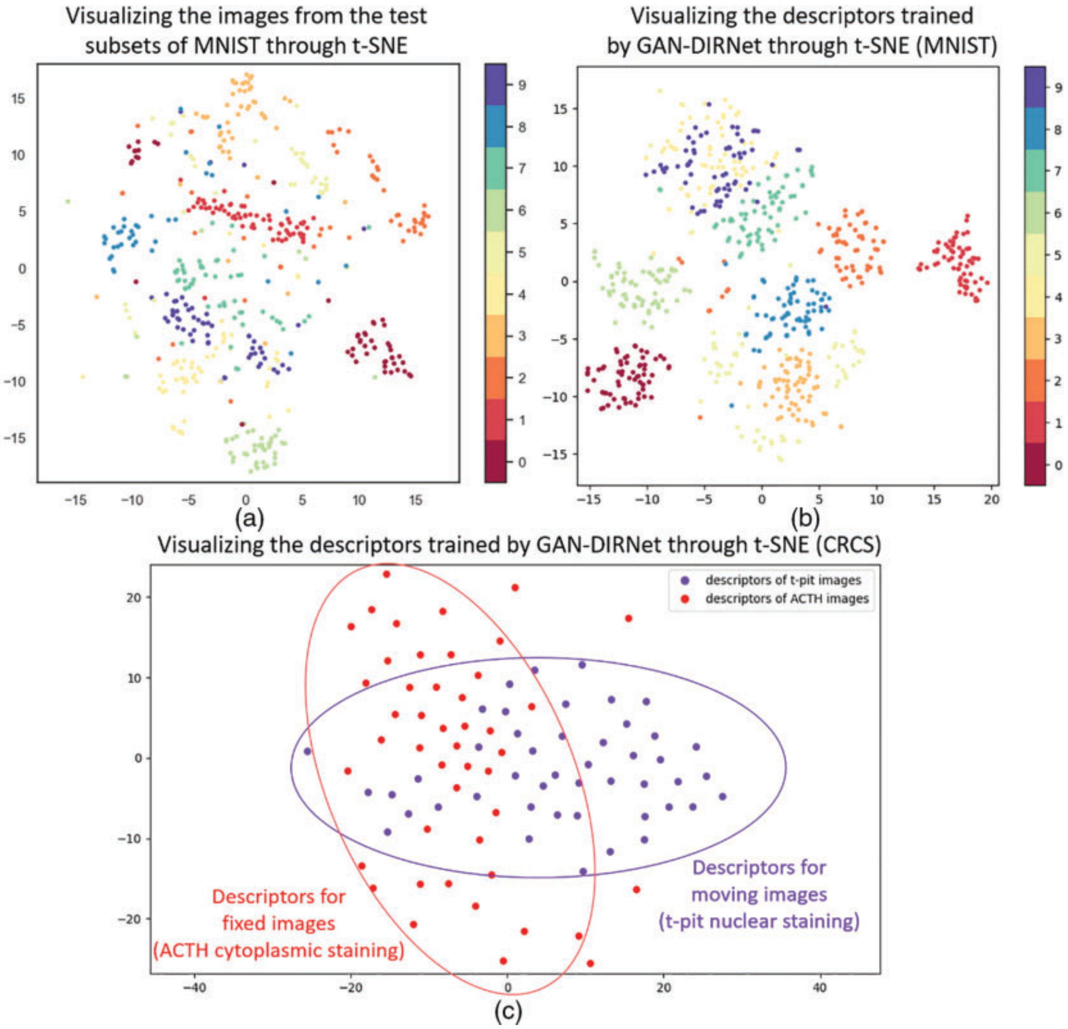


Figure 6: Visualizations through t-SNE. (a) and (b) are visualizations of the test subset of the MNIST dataset and its trained descriptors through GAN-DIRNet (i.e., the output of the GAN module and the input of the DIRNet module), respectively. By comparison, it can be seen that the latter shows better aggregation than the former. (c) is a visualization of the descriptors trained on GAN-DIRNet for the CRCS dataset

5 Conclusions

Multi-modal histological image registration poses one of the most challenging tasks in medical image registration. Due to the irreversibility of dyeing operations, there is inevitably non-rigid deformation and severe noise between adjacent slices, significantly affecting the registration outcome. Existing non-rigid multi-modal registration methods often struggle with histological images exhibiting substantial modality differences. In this article, we innovatively designed the discriminators of two

GANs as the embedded parallel networks and integrated them into the input part of DIRNet as the modal feature extraction layer. Our proposed model, named "GAN-DIRNet", integrates the modal-specificity information and rigid registration modules into the original DIRNet to achieve more accurate and faster performance. We conducted experimental validation on nine public datasets and a non-public clinical dataset for a rare disease to demonstrate the effectiveness of GAN-DIRNet. Finally, we performed dimensionality reduction analysis on the "intermediate results" output from the network embedding part, and analyzed the reasons why it can assist in achieving better registration results.

The advantages of GAN-DIRNet are: 1) It can adapt to non-rigid registration tasks in multimodal histological images with severe noise and artifacts; 2) It can compensate for missing tissues with non-rigid deformation; 3) It can complete convergence with fewer iterations. However, this study has limitations in the following two aspects: 1) Only 5 landmarks were placed in each image of the non-public dataset CRCS, resulting in metrics based on feature points not fully reflecting the registration effect on this dataset. 2) The rigid registration part in GAN-DIRNet adopts an integrated traditional intensity-based registration algorithm, which is computationally intensive for large-sized histological images.

The innovation of this study lies in three key points. Firstly, we establish a novel multi-modal unsupervised deep-learning based DIR for histopathology images of Cushing's disease. Secondly, we integrate the discriminator of GAN into DIRNet to assist with image registration. Finally, we propose a novel modal-specific feature descriptor for cross-modal histological image registration.

The idea of leveraging the features of each modality to aid registration in this study may offer additional perspectives for future multi-modal registration endeavors. Importantly, both the feature extraction and rigid registration process are automated. Therefore, applying our method to the registration task between any two modal images will not incur additional workload.

Acknowledgement: Thank Mr. Yu-Xuan Chen for providing guidance on paper writing.

Funding Statement: None.

Author Contributions: Study conception and design: Haiyue Li; data collection, Haiyue Li and Jing Xie; analysis and interpretation of results: Haiyue Li, Jing Ke, Ye Yuan and Hongbin Shen; draft manuscript preparation: Haiyue Li, Xiaoyong Pan and Hongyi Xin. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: We have made our code freely available at: github.com/lhy2024/GAN-DIRNet (accessed on 20/06/2024).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] Y. Ding *et al.*, "Deep-learning based fast and accurate 3D CT deformable image registration in lung cancer," *Med. Phys.*, vol. 50, no. 11, pp. 6864–6880, Nov. 2023. doi: 10.1002/mp.16548.

[2] X. Deng, E. Liu, S. Li, Y. Duan, and M. Xu, "Interpretable multi-modal image registration network based on disentangled convolutional sparse coding," *IEEE Trans. Image Process.*, vol. 32, no. 78, pp. 1078–1091, Jan. 2023. doi: 10.1109/TIP.2023.3240024.

[3] Z. Zhu, Y. Ji, and Y. Wei, "Multi-resolution medical image registration with dynamic convolution," presented at the 2022 IEEE Biomed. Circ. Syst. Conf. (BioCAS), Taipei, Taiwan, Oct. 13–15, 2022, pp. 645–649.

[4] S. Rai, J. S. Bhatt, and S. K. Patra, "Deep learning in medical image analysis: Recent models and explainability," in *The Explainable AI in Healthcare*, 1st ed. New York, USA: Chapman and Hall/CRC, 2024, pp. 23–49. Accessed: Jul. 17, 2023. [Online]. Available: https://www.taylorfrancis.com/books/edit/10.1201/9781003333425

[5] J. Hu, L. Li, Y. Fu, M. Zou, J. Zhou and S. Sun, "Optical flow with learning feature for deformable medical image registration," *Comput. Mater. Contin.*, vol. 71, no. 2, pp. 2773–2788, Dec. 2022. doi: 10.32604/cmc.2022.017916.

[6] M. Zou, *et al.*, "Rigid medical image registration using learning-based interest points and features," *Comput. Mater. Contin.*, vol. 60, no. 2, pp. 511–525, Aug. 2019. doi: 10.32604/cmc.2019.05912.

[7] S. Divakaran, "An explainable method for image registration with applications in medical imaging," in *The Explainable AI in Healthcare*, 1st ed. New York, USA: Chapman and Hall/CRC, 2024, pp. 68–82. Accessed: Jul. 17, 2023. [Online]. Available: https://www.taylorfrancis.com/books/edit/10.1201/9781003333425

[8] X. Wang, Z. Song, J. Zhu, and Z. Li, "Correlation attention registration based on deep learning from histopathology to MRI of prostate," *Critical Revi. TM in Biomed. Eng.*, vol. 52, no. 2, pp. 39–50, 2024. doi: 10.1615/CritRevBiomedEng.2023050566.

[9] M. A. Azam, K. B. Khan, M. Ahmad, and M. Mazzara, "Multimodal medical image registration and fusion for quality enhancement," *Comput. Mater. Contin.*, vol. 68, no. 1, pp. 821–840, Mar. 2021. doi: 10.32604/cmc.2021.016131.

[10] R. Han *et al.*, "Deformable MR-CT image registration using an unsupervised, dual-channel network for neurosurgical guidance," *Med. Image Anal.*, vol. 75, no. 31, pp. 102292, Jan. 2022. doi: 10.1016/j.media.2021.102292.

[11] J. Q. Zheng, Z. Wang, B. Huang, N. H. Lim, and B. W. Papież, "Residual aligner-based network (RAN): Motion-separable structure for coarse-to-fine discontinuous deformable registration," *Med. Image Anal.*, vol. 91, no. 42, pp. 103038, Jan. 2024. doi: 10.1016/j.media.2023.103038.

[12] A. Yang, T. Yang, X. Zhao, X. Zhang, Y. Yan and C. Jiao, "DTR-GAN: An unsupervised bidirectional translation generative adversarial network for MRI-CT registration," *Appl. Sci.*, vol. 14, no. 1, pp. 95, Dec. 2023. doi: 10.3390/app14010095.

[13] B. Zitova and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003. doi: 10.1016/S0262-8856(03)00137-9.

[14] M. Roy *et al.*, "Deep learning based registration of serial whole-slide histopathology images in different stains," *J. Pathol. Inf.*, vol. 14, no. 35, pp. 100311, Oct. 2023. doi: 10.1016/j.jpi.2023.100311.

[15] J. Borovec *et al.*, "ANHIR: Automatic non-rigid histological image registration challenge," *IEEE Trans. Med. Imaging*, vol. 39, no. 10, pp. 3042–3052, Oct. 2020. doi: 10.1109/TMI.2020.2986331.

[16] Y. Zhang, J. Chen, X. Ma, G. Wang, U. A. Bhatti and M. Huang, "Interactive medical image annotation using improved Attention U-net with compound geodesic distance," *Expert. Syst. Appl.*, vol. 237, no. PA, pp. 121282, Mar. 2024. doi: 10.1016/j.eswa.2023.121282.

[17] B. D. D. Vos, F. F. Berendsen, M. A. Viergever, M. Staring, and I. Išgum, "End-to-end unsupervised deformable image registration with a convolutional neural network," presented at the Deep Learn. Med. Image Anal. Multimod. Learn. Clinic. Deci. Supp., Québec City, QC, Canada, Sep. 14, 2017, pp. 204–212.

[18] Q. Zeng, C. Yang, Q. Gan, Q. Wang, and S. Wang, "EDIRNet: An unsupervised deformable registration model for X-ray and neutron images," *Appl. Opt.*, vol. 62, no. 29, pp. 7611–7620, Oct. 2023. doi: 10.1364/AO.500442.

[19] Y. Lei *et al.*, "4D-CT deformable image registration using multiscale unsupervised deep learning," *Phy. Med. & Biol.*, vol. 65, no. 8, pp. 085003, Apr. 2020. doi: 10.1088/1361-6560/ab79c4.

[20] Y. Sang, X. Xing, Y. Wu, and D. Ruan, "Imposing implicit feasibility constraints on deformable image registration using a statistical generative model," *J. Med. Imaging*, vol. 7, no. 6, pp. 064005–064005, Nov. 2020. doi: 10.1117/1.JMI.7.6.064005.

[21] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 53–65, Jan. 2018. doi: 10.1109/MSP.2017.2765202.

[22] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," presented at the Proc. IEEE Int. Conf. Comput. Vis., Venice, Italy, Oct. 22–29, 2017.

[23] D. Mahapatra, B. Antony, S. Sedai, and R. Garnavi, "Deformable medical image registration using generative adversarial networks," presented at the 2018 IEEE 15th Int. Symp. Biomed. Imaging (ISBI 2018), Washington, DC, USA, Apr. 4–7, 2018, pp. 1449–1453.

[24] H. Li *et al.*, "CRCS: An automatic image processing pipeline for hormone level analysis of Cushing's disease," *Methods*, vol. 222, no. 11, pp. 28–40, Feb. 2024. doi: 10.1016/j.ymeth.2023.12.003.

[25] J. M. Morel and G. Yu, "ASIFT: A new framework for fully affine invariant image comparison," *SIAM J. Imaging Sci.*, vol. 2, no. 2, pp. 438–469, 2009. doi: 10.1137/080732730.

[26] S. Bharati, M. Mondal, P. Podder, and V. Prasath, "Deep learning for medical image registration: A comprehensive review," *Int. J. Comput. Inform. Syst. & Industrial Manag. Appl.*, vol. 14, no. 8, pp. 173–190, Apr. 2022. doi: 10.48550/arXiv.2204.11341.

[27] D. Sengupta, P. Gupta, and A. Biswas, "A survey on mutual information based medical image registration algorithms," *Neurocomputing*, vol. 486, no. 14, pp. 174–188, May 2022. doi: 10.1016/j.neucom.2021.11.023.

[28] G. Lippolis, A. Edsjö, L. Helczynski, A. Bjartell, and N. C. Overgaard, "Automatic registration of multimodal microscopy images for integrative analysis of prostate tissue sections," *BMC Cancer*, vol. 13, no. 1, pp. 1–11, Sep. 2013. doi: 10.1186/1471-2407-13-408.

[29] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features," presented at the 9th Eur. Conf. Comput. Vis., Graz, Austria, May 7–13, 2006, pp. 404–417.

[30] J. Liu, X. Li, Q. Wei, J. Xu, and D. Ding, "Semi-supervised keypoint detector and descriptor for retinal image matching," presented at the Eur. Conf. Comput. Vis., Tel Aviv, Israel, Oct. 23–27, 2022, pp. 593–609.

[31] T. Cao, C. Zach, S. Modla, D. Powell, K. Czymmek and M. Niethammer, "Multi-modal registration for correlative microscopy using image analogies," *Med. Image Anal.*, vol. 18, no. 6, pp. 914–926, Aug. 2014. doi: 10.1016/j.media.2013.12.005.

[32] T. C. Mok and A. Chung, "Fast symmetric diffeomorphic image registration with convolutional neural networks," presented at the Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Seattle, WA, USA, Jun. 13–19, 2020, pp. 4644–4653.

[33] X. Zhang, W. Jian, Y. Chen, and S. Yang, "Deform-GAN: An unsupervised learning model for deformable registration," arXiv preprint arXiv:2002.11430, Feb. 2020. doi: 10.48550/arXiv.2002.11430.

[34] S. Liu, B. Yang, Y. Wang, J. Tian, L. Yin and W. Zheng, "2D/3D multimode medical image registration based on normalized cross-correlation," *Appl. Sci.*, vol. 12, no. 6, pp. 2828, Mar. 2022. doi: 10.3390/app12062828.

[35] Y. R. Rao, N. Prathapani, and E. Nagabhooshanam, "Application of normalized cross correlation to image registration," *Int. J. Res. Eng. Technol.*, vol. 3, no. 5, pp. 12–16, May 2014. doi: 10.15623/ijret.2014.0317003.

[36] K. Tang, Z. Li, L. Tian, L. Wang, and Y. Zhu, "ADMIR-affine and deformable medical image registration for drug-addicted brain images," *IEEE Access*, vol. 8, pp. 70960–70968, Apr. 2020. doi: 10.1109/ACCESS.2020.2986829.

[37] L. Yann, "The MNIST database of handwritten digits," 1998. Accessed: Oct. 1, 2010. [Online]. Available: http://yann.lecun.com/exdb/mnist/

[38] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," presented at the Int. Conf. Mach. Learn., Lille, France, Jul. 6–11, 2015, pp. 448–456.

[39] D. J. Kroon and C. H. Slump, "MRI modalitiy transformation in demon registration," presented at the 2009 IEEE Int. Symp. Biomed. Imaging: From Nano to Macro, Boston, MA, USA, Jun. 28–Jul. 1, 2009, pp. 963–966.