Neurocomputing 500 (2022) 799–808

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

CoCycleReg: Collaborative cycle-consistency method for multi-modal medical image registration

Chenyu Lian [a], Xiaomeng Li [b], Lingke Kong [c], Jiacheng Wang [a], Wei Zhang [c], Xiaoyang Huang [a,*], Liansheng Wang [a,*]

[a] Department of Computer Science at School of Informatics, Xiamen University, Xiamen 361005, China
[b] Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, China
[c] Manteia Technologies Co., Ltd, Xiamen 361005, China

ARTICLE INFO

Article history:
Received 30 November 2021
Revised 20 March 2022
Accepted 28 May 2022
Available online 2 June 2022
Communicated by Zidong Wang

Keywords:
Medical image analysis
Multi-modal image registration
Cycle-consistency
Image-to-image translation

ABSTRACT

Multi-modal image registration is an essential step for many medical image analysis applications. Recent advances in multi-modal image registration rely on image-to-image translation to achieve good performance. However, the performance is still limited owing to the poor use of complementary regularization between image registration and translation, which is able to simultaneously enhance both parts' accuracy. To this end, we propose CoCycleReg, a novel method that formulates image registration and translation in a ***Co****llaborative* ***Cycle****-consistency* manner. Instead of dividing into two discrete stages, we unify the image registration and translation via cycle-consistency in an end-to-end training process, such that each part can benefit from the other one. To ensure the deformation fields' reversibility in the cycle, we extensively introduce a novel dual-head registration network, consisting of one single backbone to extract the features and two heads to respectively predict the deformation fields. The experiments on T1-T2(MRI) and CT-MRI datasets validate that the proposed CoCycleReg surpasses the other state-of-the-art conventional and deep learning approaches comprehensively considering the speed, accuracy, and regularity of deformation fields. In the ablation analysis, a method that sets the cycle-consistency Corresponding authors at: Department of Computer Science at School of Informatics, Xiamen University, Xiamen 361005, Chinaconstraints of registration and image-to-image translation separately is compared, and the results demonstrate the effectiveness of collaborative cycle-consistency. In addition, the improvement of image-to-image translation is also verified in further analysis. The code is publicly available at https://github.com/DopamineLcy/cocycle-reg/.

1. Introduction

Medical images from different modalities such as Computed Tomography(CT) and Magnetic Resonance Imaging(MRI) provide complementary information, which can significantly aid in the early detection of tumors or other diseases and help improve diagnostic accuracy [1,2]. However, multi-modal images usually have inevitable misalignment issues due to patient motion and variations in anatomical structures. Rigid registration can perform well in structures that are not susceptible to elastic changes (e.g., bone). But for soft tissues, many factors, including tissue abnormalities, respiratory movements, and muscle contractions, can cause elastic deformation. For this situation, deformable registration is more suitable and accurate. Deformable image registration has been a fundamental component of many medical image analysis applications, such as monitoring diseases' progression and quantifying treatment mechanisms' effectiveness [3–6]. The goal of deformable image registration is to achieve a high speed, high accuracy and guarantee deformation fields to be realistic.

Previous works on multi-modal image registration mainly include conventional iterative optimization-based methods [7–10], metric-based deep learning methods [11,12] and image-to-image translation-based deep learning methods [13–15].

The conventional iterative optimization-based methods estimate the deformation fields by optimizing certain objective functions like Mutual Information (MI) [7,9,10] and Modality Independent Neighbourhood Descriptor (MIND) [8]. The most severe limitation of this sort of methods is that the optimization process is very computationally expensive and time-consuming. Besides the computational disadvantage, designing accurate met-

* Corresponding authors.
E-mail addresses: xyhuang@xmu.edu.cn (X. Huang), lswang@xmu.edu.cn (L. Wang).

https://doi.org/10.1016/j.neucom.2022.05.113
0925-2312/

rics to evaluate the similarity of images from different modalities is challenging.

With the advances of deep neural networks, researchers began to investigate the deep learning-based methods for mono-modal image registration [16,17]. The deep learning methods optimize a spatial transform network (STN) [18] by comparing the warped image and the target one using similarity metrics like Mean Squared Error (MSE) and Normalized Cross-Correlation (NCC). Furthermore, the mono-modal image registration methods have been extended to multi-modal image registration and these extended methods can be broadly classified into metric-based deep learning methods [11,12] and image-to-image translation-based deep learning methods [13–15]. The common idea of the metric-based methods is to find a metric to evaluate the similarity of images from different modalities and solve multi-modal registration problem based on mono-modal registration methods. To achieve this, MI, Structural Similarity (SSIM) [19] and MIND [8] are utilized. However, statistic metrics like MI and SSIM introduce inaccuracy while the upper limit of handcrafted metrics like MIND is obvious since it's arguably impossible to design a metric to suit all kinds of modalities. Instead of directly measuring the similarity of images from different modalities, image-to-image translation-based deep learning methods translate the moving images to the modality of target images and use simple mono-modal metrics like MSE to measure the similarity [13,14]. This kind of methods discards complicated handcrafted metrics completely and the performance benefits from the development of image-to-image translation.

In recent years, cycle-consistency has become prevalent since Zhu et al. proposed the cycle-consistency of image-to-image translation and made a great success in CycleGAN [20]. For the image registration task, the cycle-consistency is used to improve the invertibility of image registration [21]. More details about related cycle-consistency can be found in 2.3. Besides mentioned in the last paragraph that image-to-image translation-based deep learning methods benefit a lot from the image translation, the translation-registration collaboration is also a key point. However, to the best of our knowledge, the complementary regularization between image registration and translation is still underexplored. Inspired by cycle-consistency and the integrated translation and registration framework in [14], we propose CoCycleReg, delving into the cycle-consistency of the integrated translation and registration framework to improve the performance of multi-modal image registration.

In the present work, our main contributions are:

- we introduce a *collaborative cycle-consistency* framework for multi-modal image registration, where the image registration and translation part can regularize each other during the training process. The regularization enhances the accuracy and regularity of image registration and the consistency of geometry shape during image-to-image translation;
- we propose dual-head deformation fields generating network to generate bi-directional deformation fields with a single network. Compared to inverse the deformation field of one direction directly to obtain the other one, the proposed dual-head network generates bi-directional deformation fields with better invertibility. In the meantime, training two networks to generate bi-directional deformation fields is avoided, which reduces the network parameters and makes training easier;
- the entire framework is end-to-end and image-to-image translation in 3D volumes is achieved directly, instead of doing translation in 2D slice by slice and concatenating, which makes *supervisory* information be aware by the generators. The end-to-end framework improves the performance of image-to-image translation and thus promotes the image registration process.

We validate the effectiveness of our method with the example of pairwise multi-modal registration of 3D CT and MRI scans. Specifically, we evaluate the model performance on a well-aligned T1-T2 (MRI) dataset with manual deformations and a clinical CT-MRI dataset. Experiment results demonstrate our method outperforms other state-of-the-art approaches comprehensively considering the speed, accuracy, and regularity of deformation fields and has some positive effects on the image-to-image translation process. Further ablation analysis validates the effectiveness of the proposed collaborative cycle-consistency manner.

2. Related work

2.1. Deep Learning-based Medical Image Registration

VoxelMorph [17] has been the most prevalent method of medical image registration for giving a generic unsupervised learning pattern. In recent years, most of the proposed methods, both mono-modal and multi-modal registration, are based on the pattern to optimize a spatial transform network (STN) [18] by comparing the warped image and the target one using similarity metrics. In our study, the VoxelMorph pattern is integrated in the proposed collaborative cycle-consistency manner.

2.2. Image-to-image Translation

Our work is an image-to-image translation-based method and image-to-image translation was proposed for 2D natural images synthesis originally. Phillip et al. proposed supervised image-to-image translation method Pix2Pix [22] and then Zhu *et al.* proposed unsupervised image-to-image translation method CycleGAN [20]. After that, a lot of medical image synthesis methods were proposed to translate medical images from one modality to another, for example, from MRI to CT or from CBCT to CT [23–25]. The image-to-image translation process is crucial in our method for making it possible to evaluate the similarity between images from different modalities.

2.3. Cycle-consistency

Zhu et al. proposed the cycle-consistency of image-to-image translation in CycleGAN [20] and Kim *et al.* adopted the cycle-consistency to improve the invertibility of image registration [21]. The proposed method unified the cycle-consistency of image registration and translation collaboratively.

The cycle-consistency forms of CycleGAN, the cycle-consistency of registration and proposed CoCycleReg are illustrated in Fig. 1. Specifically, CoCycleReg ensures the cycle-consistency during image translation, which helps to keep the geometry shape consistent during translation. And CoCycleReg also guarantees the cycle-consistency during image registration, which helps to keep the invertibility of bi-directional deformation fields.

2.4. Image-to-image Translation Based Multi-modal Image Registration

With the development of image-to-image translation, it is possible to convert multi-modal image registration to mono-modal image registration. For example, Wei et al. [13] used CycleGAN [20] with the mutual information constraint to generate synthesized CT image from the corresponding MR image slice by slice and then concatenate 2D slices into 3D volumes, converting multi-modal registration into a mono-modal problem. This type of methods made the image-to-image translation and image regis-

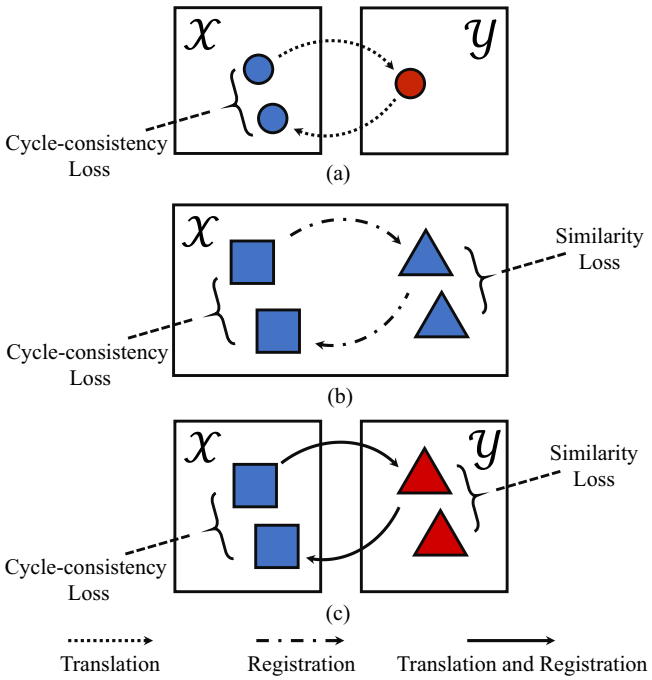


Fig. 1. Cycle-consistency forms of (a) CycleGAN, (b) the cycle-consistency of registration and (c) proposed CoCycleReg, respectively. CycleGAN uses cycle-consistency to achieve image-to-image translation between two modalities, and the cycle-consistency of registration works on the image registration process in the single modality. Differently, the proposed CocycleReg uses the collaborative cycle-consistency to help both the image registration and translation processes in two different modalities. Only the flow from $\mathscr{X}$ to $\mathscr{Y}$ is given here and there is the similar one from $\mathscr{Y}$ to $\mathscr{X}$ actually. $\mathscr{X}$ and $\mathscr{Y}$ represent two image modalities indicated by red and blue, respectively.

tration processes *gradient discontinuous*, resulting in that these two stages cannot regulate each other [13,15].

Most related to our work, Arar et al. [14] introduced a multi-modal registration method in 2D based on geometry preserving image-to-image translation. They integrate the translation and registration process and the translation and registration networks are optimized simultaneously. However, this work focuses on 2D natural image registration and the cycle-consistency of image-to-image translation and image registration are not addressed. It should be noted that the regularity and reversibility of deformation fields are significantly important for medical image registration, which is fully addressed in the proposed method.

3. Methods

Given a set of multi-modal image pairs $\{(x_i, y_i)\}_{i=1}^{n}$, where $x \in \mathscr{X}$ and $y \in \mathscr{Y}$, where $\mathscr{X}$ and $\mathscr{Y}$ denote two image modalities. For simplicity, we denote a pair of multi-modal images as (x, y) instead of (x_i, y_i). As our task bi-directional multi-modal image registration, for a given input image pair (x, y), our goal is to estimate bi-directional deformation fields (ϕ_{x2y}, ϕ_{y2x}).

The pipelines of the proposed method from x to y and from y to x are completely symmetrical and here we take the cycle process from x to y as an example, as shown in Fig. 2(b). First, the deformation fields generating network R_Φ generates bi-directional deformation fields (ϕ_{x2y}, ϕ_{y2x}), as shown in Fig. 2(a). Second, the image x is translated and warped through forward translation and registration flow into $\hat{y}$, where $\hat{y} = T_{x2y}(x) \circ \phi_{x2y}$. The process is called *forward flow* of the collaborative cycle process. Finally, the obtained $\hat{y}$ is translated and warped through backward translation registration flow back into $\hat{x}$, where $\hat{x} = T_{y2x}(\hat{y}) \circ \phi_{y2x}$. The process is called *backward flow* of the collaborative cycle process. The forward and backward processes constitute the collaborative cycle-consistency of translation and registration flow.

The whole network is end-to-end and optimized by a combination of the similarity losses, the cycle-consistency losses and the GAN losses. The similarity losses minimize the differences between $\hat{y}$ and y to train registration network and provide supervisory information to image-to-image translation networks; the cycle-consistency loss minimizes the differences between $\hat{x}$ and x to help translation networks to keep the geometry shape consistent during translation and help to keep the invertibility of bi-directional deformation fields; the GAN loss trains the image translation networks to translate images from the source domain to the target domain.

3.1. Image Registration Network

Image registration network $R = (R_\Phi, R_S)$ is a spatial transformation network (STN) [18] composed of dual-head deformation fields generating network R_Φ and resampling layer R_S. The dual-head deformation fields generating network R_Φ generates bi-directional deformation fields (ϕ_{x2y}, ϕ_{y2x}) and resampling layer R_S warps the moving images with corresponding deformation fields that produced by R_Φ, as shown in Fig. 2(a).

We know that bi-directional deformation fields are required for bi-directional registration. And to improve the invertibility of bi-directional deformation fields, previous approaches inverse the deformation field of one direction directly to obtain the other one [16] or adopt two registration networks and utilize cycle-consistency loss [26]. The former brings strict invertibility constraint that is an over-strict requirement and not training-friendly while the latter doubles the network parameters and increases the difficulty of training as well. In contrast to these methods, we adopt dual-head deformation fields generating network (Fig. 3) and and regard the invertibility as a training target during the collaborative cycle-consistency process, which simplifies the training process and reduces about half of the parameters of the registration network.

Next we will introduce how the registration network works. For a given moving image x, the value of the warped image $x \circ \phi_{x2y} = R_S(x, \phi_{x2y})$ at voxel $\mathbf{v} = (i, j, k)$ is given by:

$$x \circ \phi_{x2y}[v] = x[\mathbf{v} + \phi_{x2y}(\mathbf{v})], \tag{1}$$

where $\phi_{x2y}(\mathbf{v}) = (\Delta z, \Delta y, \Delta x)$ is the deformation at voxel $\mathbf{v} = (i, j, k)$. Because image values are only defined at integer locations, we apply tri-linear interpolate to do an approximation.

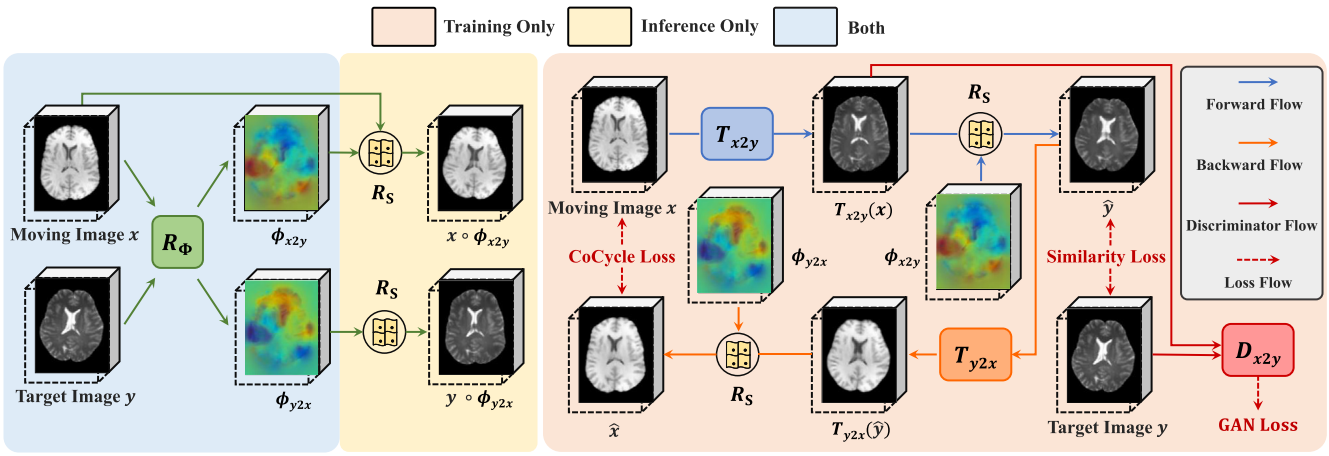
Specifically, in our method R_S works on the collaborative cycle and the grads can be passed to R_Φ for backward propagation. To avoid generating non-smooth and not physically realistic deformation fields, we adopt smoothness regularization. We follow [16] and encourage a smooth displacement ϕ using a regularizer on the spatial gradients of displacement $\mathbf{u}$:

$$\mathscr{L}_{regular}(\phi) = \sum_{\mathbf{v} \in \Omega} \|\nabla \mathbf{u}(\mathbf{v})\|^2, \tag{2}$$

where $\phi = Identity + \mathbf{u}$ and $\mathbf{v}$ is a voxel of image. In practice, we approximate the spatial gradients by differences between neighboring voxels. In summary, the smoothness loss in our method is given by:

$$\mathscr{L}_{smooth}(R) = \mathscr{L}_{regular}(\phi_{x2y}) + \mathscr{L}_{regular}(\phi_{y2x}), \tag{3}$$

where ϕ_{x2y} and ϕ_{y2x} are bi-directional deformation fields.



(a) Image registration. R_s warps images with corresponding deformation fields produced by R_Φ.

(b) Pipeline of the proposed method. The framework is bi-directional, only the direction from x to y is shown here.

Fig. 2. (a) The image registration network. Given the input (x, y), R_Φ generates bi-directional deformation fields (ϕ_{x2y}, ϕ_{y2x}). R_S is a tri-linear re-sampler layer, which warps the moving images with corresponding deformation fields. (b) The collaborative cycle pipeline of the proposed CoCyleReg. We showcase the example from $\mathscr{X}$ to $\mathscr{Y}$, and x is the moving image from domain $\mathscr{X}$ and y is the fixed image from domain $\mathscr{Y}$. T_{x2y} and T_{y2x} are image-to-image translation networks from domain $\mathscr{X}$ to $\mathscr{Y}$ and from $\mathscr{Y}$ to $\mathscr{X}$, respectively. D_{x2y} distinguishes whether an image from domain $\mathscr{Y}$ is real or not and used for calculating GAN loss.

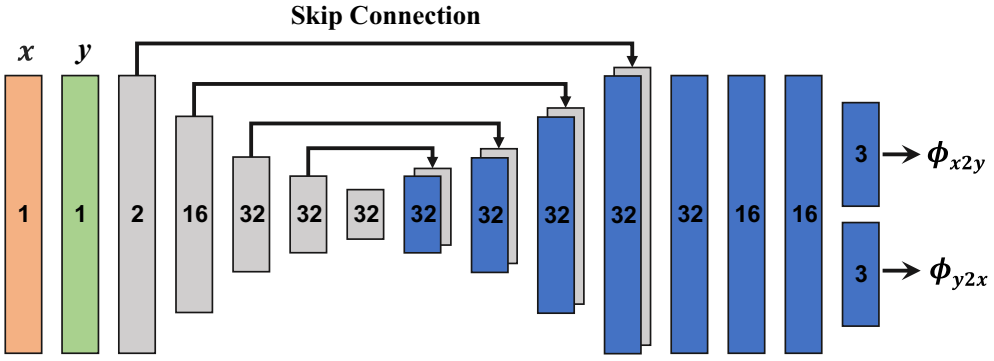


Fig. 3. The network structure of dual-head deformation fields generating network R_Φ. The whole network except the final dual-head layer is UNet and the same to the registration network in VoxelMorph[17]. Each head is a 3D convolutional layer with 16 channels input and 3 channels output following the last layer of the UNet backbone.

3.2. Collaborative Cycle-consistency Network

The framework of our method is bi-directional and completely symmetrical, we only describe the process from domain $\mathscr{X}$ to domain $\mathscr{Y}$ as an example for simplicity and there is a similar process from $\mathscr{Y}$ to $\mathscr{X}$ actually.

3.2.1. Translation and Registration Flow

The translation and registration flow first apply an image-to-image translation on x to get $T_{x2y}(x)$, and then a spatial transformation on $T_{x2y}(x)$ to generate $\hat{y}$, where $\hat{y} = T_{x2y}(x) \circ \phi_{x2y}$. To ensure the global fidelity of the translated image, the GAN loss is applied. We use PatchGAN [22] discriminator network D_{x2y} to classifies y as real and $\hat{y}$ as fake. The GAN loss $\mathscr{L}^{GAN}_{x2y}$ is given by Eq. 4.

$$\mathscr{L}^{GAN}_{x2y}(T_{x2y}) = \mathbb{E}[D^2_{x2y}(y)] + \mathbb{E}[(1 - D_{x2y}(\hat{y}))^2] \tag{4}$$

The ideal translation and registration flow should generate an output that is very close to the target image, i.e., $\hat{y} \approx y$. To achieve this goal, we use the similarity loss to maximize the similarity between $\hat{y}$ and y. The similarity loss is defined as:

$$\mathscr{L}^{sim}_{x2y}(R, T_{x2y}) = \|\hat{y} - y\|_1 \tag{5}$$

The translation and registration flow is indicated by blue arrows in Fig. 2(b). During the translation and registration flow, the registration network R is trained with the similarity loss. The translation network T_{x2y} is not only trained with the GAN loss but also with similarity loss, which shouldn't be possible without well-aligned images. However, the registration network corrects the misalignment and introduces supervisory information to assist the training of image-to-image translation networks, promoting the performance of image-to-image translation.

3.2.2. Collaborative Cycle-consistency Regularization

The translation and registration flow in the above section is called *forward flow* in the whole pipeline. In forward flow, the network generates an output $\hat{y}$ that is very close to the target image y. However, we still need cycle consistency of image-to-image translation to keep the geometry shape consistent during translation and cycle consistency of image registration to keep the invertibility of deformation fields. So here we take the reversed translation and registration process as the *backward flow* to be the regularization of image registration and translation, which is indicated by orange arrows in (Fig. 2(b)). Hence, when we transform $\hat{y}$ to the original modality $\mathscr{X}$ through the translation and registration process, the

generated output $\hat{x}$ should be very close to the input image x. We formulate the whole training process in a collaborative cycle-consistency way with the forward flow and the backward flow. Specifically, after getting the output $\hat{y}$ through forward flow, network T_{y2x} translate $\hat{y}$ back to $T_{y2x}(\hat{y})$ and then warp it back with deformation field ϕ_{y2x}. After the backward flow, we get x, where $\hat{x} = T_{y2x}(\hat{y}) \circ \phi_{y2x}$. Formally, the CoCycle loss is given by:

$$\mathscr{L}_{x2y}^{CoCycle}(R, T_{x2y}, T_{y2x}) = \|\hat{x} - x\|_1 \tag{6}$$

3.2.3. Overall Loss Function

As mentioned above, the data flow of the proposed method is bi-directional and Fig. 2 (b) only shows the data flow from domain $\mathscr{X}$ to domain $\mathscr{Y}$. In the data flow of the other direction, y is translated and warped to $\hat{x}\prime$ during forward flow, where $\hat{x}\prime = T_{y2x}(y) \circ \phi_{y2x}$. And the obtained $\hat{x}\prime$ is fed into the translation network T_{x2y} and warped by ϕ_{x2y} to get the output $\hat{y}\prime$ during backward flow, where $\hat{y}\prime = T_{x2y}(\hat{x}\prime \circ \phi_{x2y})$. The adversarial loss, similarity loss, and CoCycle loss of the direction from domain $\mathscr{X}$ to domain $\mathscr{Y}$ are given by Eq. (7)–(9), respectively.

$$\mathscr{L}_{y2x}^{GAN}(T_{y2x}) = \mathbb{E}[D_{y2x}^2(x)] + \mathbb{E}[(1 - D_{y2x}(\hat{x}\prime))^2] \tag{7}$$

$$\mathscr{L}_{y2x}^{sim}(R, T_{y2x}) = \|\hat{x}\prime - x\|_1 \tag{8}$$

$$\mathscr{L}_{y2x}^{CoCycle}(R, T_{y2x}, T_{x2y}) = \|\hat{y}\prime - y\|_1 \tag{9}$$

In summary, the overall loss of image translation and registration is given by:

$$\begin{aligned}\mathscr{L}(R, T_{x2y}, T_{y2x}) = {} & \mathscr{L}_{x2y}^{sim}(R, T_{x2y}) + \mathscr{L}_{y2x}^{sim}(R, T_{y2x}) + \lambda_{GAN} \\ & \cdot [\arg\max_{D_{x2y}} \mathscr{L}_{x2y}^{GAN}(T_{x2y}, D_{x2y}) \\ & + \arg\max_{D_{y2x}} \mathscr{L}_{y2x}^{GAN}(T_{y2x}, D_{y2x})] + \lambda_{CoCycle} \\ & \cdot [\mathscr{L}_{x2y}^{CoCycle}(R, T_{x2y}, T_{y2x}) \\ & + \mathscr{L}_{y2x}^{CoCycle}(R, T_{x2y}, T_{y2x})] + \lambda_{smooth} \\ & \cdot \mathscr{L}_{smooth}(R).\end{aligned} \tag{10}$$

The goal of the optimization is to find R^*, T_{x2y}^* and T_{y2x}^* such that

$$R^*, T_{x2y}^*, T_{y2x}^* = \arg\min_{R, T_{x2y}, T_{y2x}} \mathscr{L}(R, T_{x2y}, T_{y2x}). \tag{11}$$

In addition, $\lambda_{GAN} = 1, \lambda_{CoCycle} = 1$, and $\lambda_{smooth} = 1$ in the experiments. However, these weights can be adjusted in the practical application according to the accuracy-regularity trade-off.

3.2.4. Networks Details

The backbone of deformation fields generating network R_Φ is based on UNet [27]. The only modification is that we add two heads after the final feature layer to generate bi-directional deformation fields and each head is a 3D convolutional layer with 16 channels input and 3 channels output, as shown in Fig. 3. In our experiments, the input is of size $2 \times 80 \times 144 \times 112$ for T1-T2 (MRI) dataset and $2 \times 64 \times 144 \times 112$ for CT-MRI dataset, but because the batch size is 1, any depth, height, width that satisfies a multiple of 16 is acceptable to the network. And R_S is a resample layer based on tri-linear interpolation. The image translation network T_{x2y} and T_{y2x} are Resnet [28]-based generators and the discriminators D_{x2y} and D_{y2x} are PatchGAN [22] classifiers, following [20,22]. The input size of generators and discriminators is the same to the input size of R_Φ but with only 1 channel.

4. Experiments

4.1. Experimental Design

We evaluate the present approach on two datasets with precise manual segmentation. During the registration, the segmentation of the moving image was warped simultaneously with the image, and the accuracy was measured by the degree of overlap between the fixed and warped segmentation. Dice Similarity Coefficient (DSC) [29] and Hausdorff Distance-95 (HD95) are computed between masks of fixed and warped images to measure the registration accuracy. In addition, the number of voxels with a non-positive Jacobian determinant is used to evaluate the regularity of the deformation fields. The Jacobian determinant $J_\phi(\mathbf{v}) = \nabla\phi(\mathbf{v}) \in \mathscr{R}^{3\times3}$ reflect the local properties of deformation field ϕ around voxel $\mathbf{v}$, and $|J_\phi(\mathbf{v}| \leq 0$ indicates the deformation in voxel $\mathbf{v}$ is not diffeomorphic [30]. We validate the effectiveness of the proposed approach by setting up comparison experiments with mainstream methods, and we set up an ablation analysis to validate the superiority of the proposed collaborative cycle-consistency. Moreover, the performances of image-to-image translation are compared between CycleGAN[20] and other comparative image-to-image translation-based deep learning methods with the proposed CoCycleReg to validate the effectiveness of collaborative cycle-consistency manner furtherly.

4.2. Datasets and Preprocessing

We employ public **BraTS dataset** [31,32] and private **neck and head dataset** to evaluate our method. This study was approved by the institutional review boards at the hospital, and informed consent was waived. The **BraTS dataset** is a dataset of multi-modal MRI scans (T1, T2, T2-FLAIR, and T1CE) with precise manual segmentation of tumors, including 285 cases. The dataset was collected from 285 patients with glioblastomas from 19 to 86 years old, both male and female. Only the T1 and T2-weighted images were utilized in our experiments. Some images in the dataset are very blurry. And the anatomy, which is very important for image registration, is almost unrecognizable. Therefore, we asked professional radiologists to help us exclude these low-quality images (56 cases). In addition, we noticed that even in the same modality, the contrast of the obtained MRI-T1 images varied considerably. In order to more accurately compare the effectiveness of different approaches and reduce the impact caused by the differences of data, we excluded the images with too high or too low contrast. Specifically, we calculated the standard deviation of the grayscale of each voxel for each image and sorted the images from smallest to largest by the standard deviation. The largest and smallest 25% of each were excluded, and 114 cases were kept finally. The images had been standardized into 3D volumes in size of $155 \times 240 \times 240$ with 1 mm isotropic resolution and we cropped and resized all cases into $80 \times 144 \times 112$. As the provided T1-T2 (MRI) were collected at almost the exact moment and already aligned, we followed [33] to use random elastic deformation on control points followed by Gaussian smoothing. The dataset was divided into training, validation, and testing set by the ratio of 8:1:1. The **Neck and head dataset** is a clinical CT-MRI dataset of head and neck collected from 151 patients receiving a head and neck tumor diagnosis from 27 to 81 years old, both male and female. The two images of a CT-MRI pair were collected at intervals ranging from a week to a month. The dataset consists of 131 pairs without segmentation as training set and 20 pairs with precise manual segmentation of parotid gland, which is divided into validation set (10 cases) and testing set (10 cases). Affine alignment is carried out during preprocessing and the clinical image pairs exist inevitable misalign-

ments, so manual elastic deformation is not required. All cases were standardized into 3D volumes with $3mm \times 1mm \times 1mm$ isotropic resolution and cropped and resized into $64 \times 144 \times 112$.

4.3. Experimental Results and Analysis

4.3.1. Comparisons with Mainstream Methods

To show the effectiveness of the proposed approach, we compare with mainstream methods including conventional iterative optimization-based method **Elastix** [10], metric-based deep learning method **VMIND** [12], which is based on VoxelMorph with similarity loss MIND and regarded as the state-of-the-art metric-based deep learning method. Besides, in order to show the superiority of our method over the latest image-to-image translation-based deep learning method, we modified [14], which is the state-of-the-art method published in 2020, from 2D to 3D and tuned the parameters carefully as a comparison, denoted as **NeMAR** (the official name given by the authors) [14].

Examples of qualitative registration results are shown in Fig. 4, where red, green and orange contours represent the moving, ground-truth and warped boundaries, respectively. Red circles mark regions where our method outperforms other methods. We can see that our **CoCycleReg** is closer to the ground-truth boundary than other methods. In order to measure the results more accurately, we summarize the quantitative registration results in Table 1. We use Dice Similarity Coefficient (DSC) [29] and Hausdorff Distance-95 (HD95) to evaluate the accuracy and use the number of voxels with a non-positive Jacobian determinant ($|J_\phi|$) to evaluate the regularity of deformation fields.

As shown in Table 1, our method surpasses all the other mainstream methods in registration accuracy. Conventional iterative optimization-based method **Elastix** has an outstanding performance in the regularity of deformation fields but with inferior accuracy. The most serious problem is that it is much slower than other deep-learning methods, which is very disadvantageous to clinical practice. Registering a pair of images takes close to twenty seconds, which is considerable for tasks with high real-time requirements, such as surgical navigation, automatic planning of radiotherapy, etc. The image-to-image translation-based method **NeMAR** has the closest performance in terms of accuracy of registration to our method, but the regularity of deformation fields is much inferior to other methods. We can summarize from the table that our method, as a learning-based method, has the advantage of high speed compared to the conventional iterative optimization-based methods like **Elastix**. Besides, the proposed method has a better performance of accuracy and regularity of deformation fields than other comparative learning-based methods.

4.3.2. Ablation Analysis

To the best of our knowledge, we are the first to integrate the cycle-consistency of image registration and translation. However, in order to validate the effectiveness of the proposed collaborative cycle-consistency method, we compare our method with **SepCycleReg**, where two cycle-consistency constraints are set **Sep**arately. The comparison of cycle-consistency manners between our **CoCycleReg** and the method **SepCycleReg** for ablation analysis is illustrated in Fig. 6. Formally, here we replace Eq. 6, 9 with Eq. 12, 13:

$$\mathscr{L}_{x2y}^{SepCycle}(R, T_{x2y}, T_{y2x}) = \left\|(x \circ \phi_{x2y}) \circ \phi_{y2x} - x\right\|_1 + \left\|T_{y2x}(T_{x2y}(x)) - x\right\|_1 \tag{12}$$

$$\mathscr{L}_{y2x}^{SepCycle}(R, T_{y2x}, T_{x2y}) = \left\|(y \circ \phi_{y2x}) \circ \phi_{x2y} - y\right\|_1 + \left\|T_{x2y}(T_{y2x}(y)) - y\right\|_1 \tag{13}$$

To be consistent with comparative experiments, we use DSC and HD95 to evaluate registration accuracy and use the number of voxels with a non-positive Jacobian determinant to evaluate the regularity of deformation fields. The qualitative and quantitative results

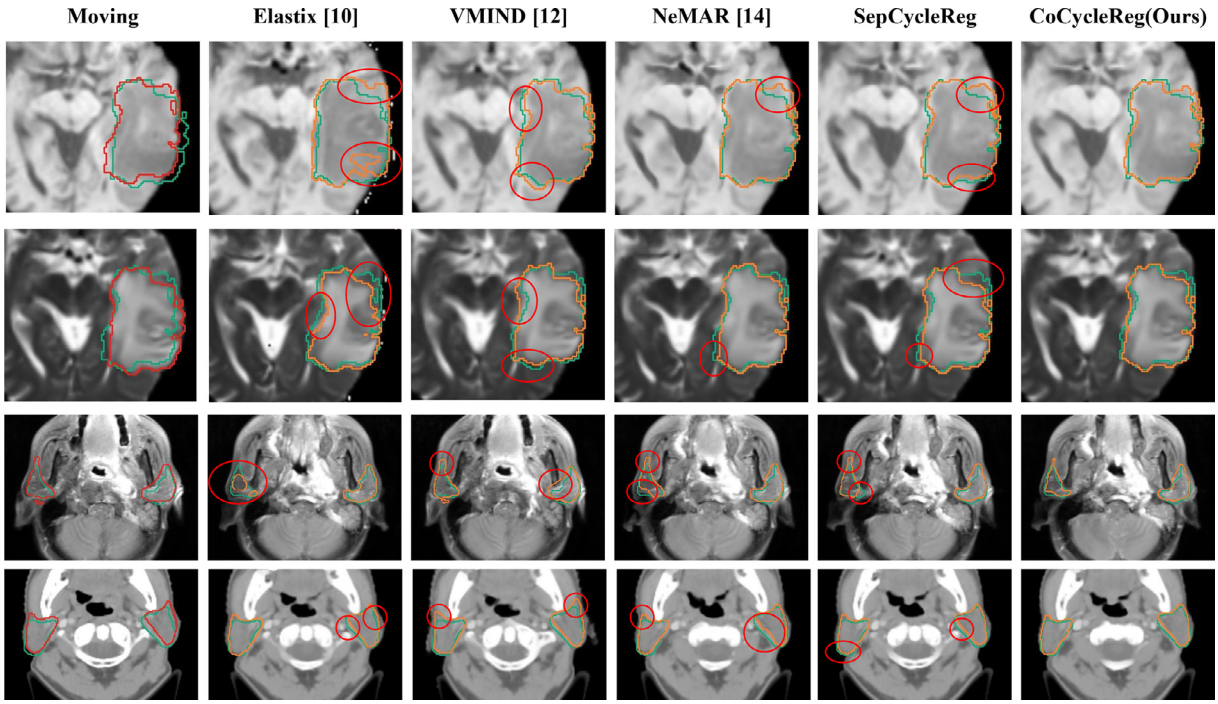


Fig. 4. Visualization of the registration results. The red, green and orange contours represent the moving, ground-truth and warped boundaries, respectively. Red circles mark regions where our method outperforms other methods. The four rows show examples of T1→T2, T2→T1, CT→MRI and MRI→CT respectively from top to bottom. Best seen when zoomed in.

Table 1
Quantitative registration results on T1-T2 (MRI) dataset and CT-MRI dataset. DSC and HD95 measure the registration accuracy and the number of voxels with a non-positive Jacobian determinant ($|J_\phi|$) is used to evaluate the regularity of the deformation fields. Running time on CPU and GPU measured in seconds shows the speed of different methods.

		Affine	Elastix [10]	VMIND [12]	NeMAR [14]	SepCycleReg	Ours
DSC(%)↑	T1→T2	80.22(4.88)	89.61(2.68)	87.53(2.85)	89.45(3.16)	89.17(3.01)	**90.0(2.53)**
	T2→T1	80.22(4.88)	88.71(1.89)	86.59(2.77)	89.27(3.32)	88.62(2.85)	**89.72(2.21)**
	MRI→CT	60.35(10.56)	73.67(3.34)	65.76(7.22)	71.77(4.14)	71.98(7.1)	**74.27(5.27)**
	CT→MRI	60.35(10.56)	71.65(5.03)	63.65(7.57)	72.5(4.89)	72.12(6.89)	**74.36(5.17)**
HD95↓	T1→T2	7.2(1.74)	**4.63(1.3)**	5.61(0.82)	5.17(1.28)	5.07(0.87)	4.84(0.92)
	T2→T1	7.2(1.74)	4.86(1.01)	5.59(0.67)	4.9(0.74)	5.15(0.86)	**4.74(0.85)**
	MRI→CT	7.49(2.02)	6.47(1.69)	7.13(1.67)	6.74(1.92)	6.61(1.96)	**6.36(1.64)**
	CT→MRI	7.49(2.02)	6.93(1.55)	7.81(1.86)	6.52(1.67)	6.6(1.76)	**6.45(1.64)**
$\lvert J_\phi\rvert \le 0$ ↓	T1→T2	/	**0.0(0.0)**	0.0(0.0)	107.25(76.14)	21.08(47.57)	8.92(21.78)
	T2→T1	/	**0.0(0.0)**	0.17(0.55)	49.67(36.71)	0.0(0.0)	0.5(1.38)
	MRI→CT	/	**0.0(0.0)**	441.0(293.35)	1330.4(1396.56)	102.4(145.11)	76.5(42.94)
	CT→MRI	/	**0.0(0.0)**	772.3(371.42)	4900.5(5731.96)	15.3(41.44)	132.1(115.75)
CPU sec↓	/	/	17.87(0.43)	**0.68(0.07)**	**0.68(0.07)**	**0.68(0.07)**	**0.68(0.07)**
GPU sec↓	/	/	/	0.03(0.01)	0.03(0.01)	0.03(0.01)	0.03(0.01)

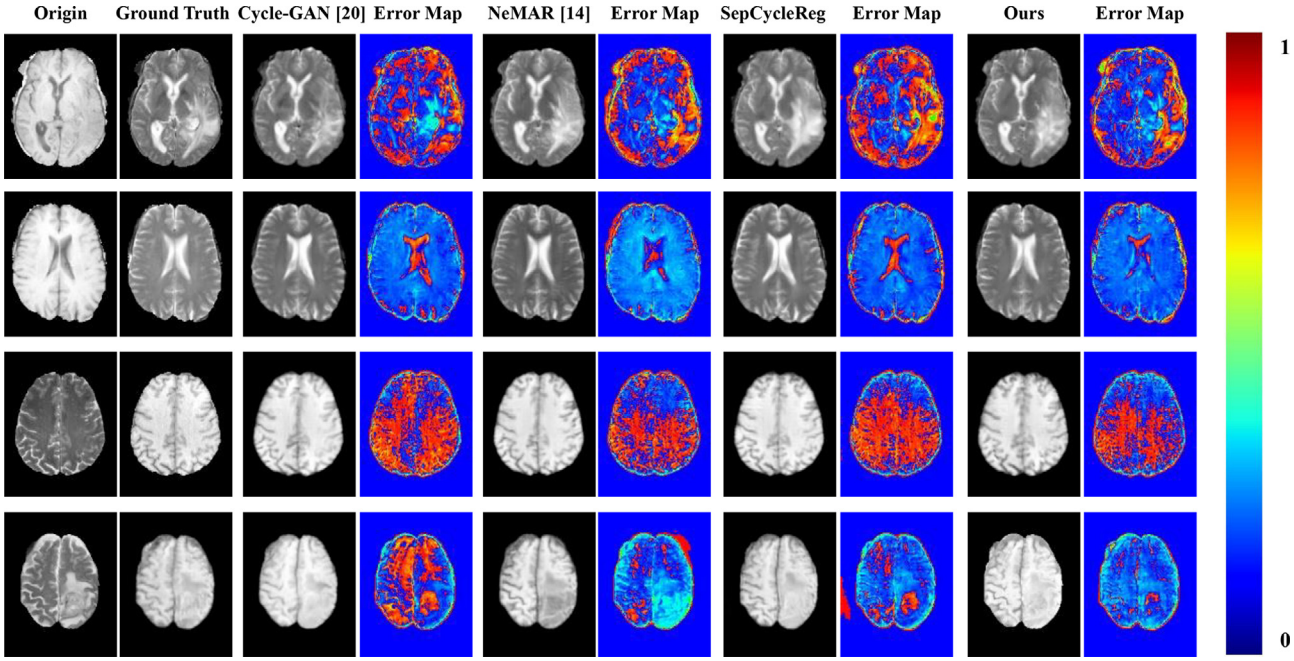


Fig. 5. Visualization of the image-to-image translation results. Each error map shows the absolute differences between translated images and the ground truth. The warmer color shows larger differences and pure blue represents zero difference.

of **SepCycleReg** are shown in the penultimate column of Fig. 4 and Table 1, respectively. Experimental results shows that the proposed method **CoCycleReg** outperforms **SepCycleReg** in registration accuracy and the regularity of deformation fields is in the same range.

4.3.3. Extensive Analysis of Image-to-image Translation Performance

The proposed multi-modal image registration method is an image-to-image translation-based deep learning method. Most of the time the registration performance of this kind of method depends heavily on the performance of image-to-image translation. In order to analyse the influence of the proposed collaborative cycle-consistency on image-to-image translation, we compare the results of widely used image-to-image translation method **CycleGAN** [20], the comparative method **NeMAR**[14] and the method in ablation analysis **SepCycleReg** with the proposed **CoCycleReg**. Root Mean Squared Error (RMSE), Peak Signal to Noise Ratio (PSNR) [34] and Structural SIMilarity (SSIM) [19] are utilized to measure the quality of translated images. Quantitative results in Table 2 shows that the proposed **CoCycleReg** outperforms other methods in image-to-image translation. The result shows that the collaborative cycle-consistency framework can not only promote image registration but also shows the same or, in some cases, slightly better quality in image-to-image translation than baseline methods. Visualization of the image-to-image translation results is shown in Fig. 5, showing the translated images and their error maps. Each error map shows the absolute differences between translated image and the ground truth. The warmer color shows larger differences and pure blue represents zero difference.

4.3.4. Analysis on all 285 Cases of the BraTS dataset

And mentioned in Section 4.2, we have preprocessed the **BraTS dataset** and only used 114 cases of the dataset. To validate the fairness of the data cleaning, we supplemented experiments of the proposed **CoCycleReg**, conventional iterative optimization-based method **Elastix** and **NeMAR** (the best one of the comparative learning-based methods) on all 285 cases of the **BraTS** dataset. The experimental results (Table 3) showed that the performances of each method decreased slightly, but the conclusion of the com-

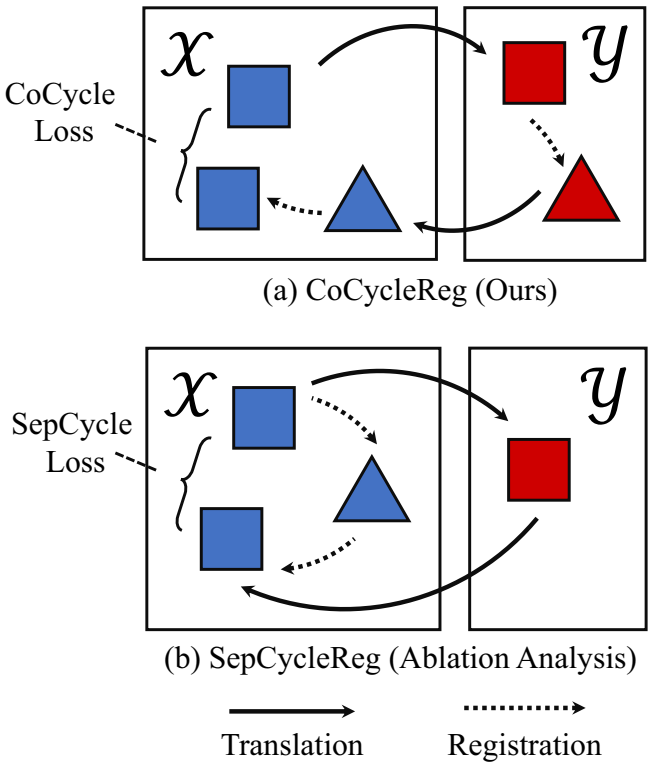


Fig. 6. Illustration of comparison between (a) CoCycle loss in our method and (b) SepCycle loss in the method for ablation analysis.

parative experiments still held for all the 285 cases. Notably, the results in the supplementary experiments have more considerable variances than the original experiments (especially notable for HD95), which shows that the data cleaning process reduces the impact caused by the difference of data.

4.4. Implementation Details

Our code is implemented using PyTorch 1.9.0 [35] and the experiments were conducted on a single GeForce RTX 3090 GPU. The training time is 45 h for 800 epochs using a single NVIDIA GeForce RTX 3090 GPU, and 20 GB memory is required. We only tested on NVIDIA GeForce RTX 3090 GPU, but any other cards for deep learning with 20 GB memory are expected to meet computational power requirements. We train the network with batch size 1, and use Adam Optimizer with parameters $lr = 1 \times e^{-4}, \beta_1 = 0.5$ and $\beta_2 = 0.999$. All networks were initialized by the Kaiming [36] initialization method. For a fair comparison, the parameters are the same and we train from scratch in all comparative experiments and ablation studies. What's more, the network backbone of registration networks, image-to-image translation networks and discriminator networks are the same in all experiments.

Table 3
Quantitative registration results on all 285 cases of the BraTS dataset.

		Elastix	NeMAR	Ours
DSC(%)↑	T1→T2	89.23(2.95)	89.25(3.2)	**89.58(3.63)**
	T2→T1	88.93(3.3)	88.21(2.55)	**89.25(3.48)**
HD95↓	T1→T2	5.2(1.51)	5.22(1.57)	**5.01(1.35)**
	T2→T1	5.09(1.43)	5.32(1.12)	**5.08(1.41)**
$\lvert J_\phi \rvert \leq 0$ ↓	T1→T2	**0.0**	252.36(299.9)	5.25(24.81)
	T2→T1	**0.0**	165.40(191.9)	2.75(6.27)

5. Discussion and Conclusions

In this paper, we have proposed a novel deep learning framework CoCycleReg for multi-modal medical image registration, which focuses on the deep relationship between image registration and image-to-image translation. CoCycleReg outperforms other state-of-the-art approaches comprehensively considering the speed, accuracy, and regularity of deformation fields.

We performed a set of comparative experiments validating that CoCycleReg outperforms state-of-the-art methods of conventional iterative optimization-based methods, metric-based deep learning methods and image-to-image translation-based deep learning methods. The ablation analysis validated that the collaborative cycle-consistency was better than setting cycle-consistency of image registration and translation separately, as Table 1 shows. Further analysis of image-to-image translation performance (Table 2) showed the proposed collaborative cycle-consistency could not only promote the image registration process but also had some positive effects on the image-to-image translation process, which is very reasonable for the image-to-image translation-based method.

CoCycleReg is a generic learning model for multi-modal image registration. We didn't design a particularly elaborate network structure and the similarity loss function is just MAE, which is very simple and can be replaced by other simple metrics if necessary. Any extra constraints like MI, SSIM or MIND are not required in the image registration or translation process. But the proposed approach, like other deep learning-based approaches, is still modality-related. i.e., a trained model can only work for two modalities. And if there are many modalities, a lot of models will be needed. We would like to extend the method to be more generalized for different modalities using domain adaptation or domain generalization, etc.Admittedly, our method is currently unable to escape from the limitations of deep learning methods that are highly data-dependent, while conventional iterative optimization-based methods like Elastix do not need training and are suitable for tasks without large databases. However, our method is entirely unsupervised, and the training set does not require any manual annotations, which is not too difficult to obtain. Besides, we all believe that the development of science is gradual, and we proposed an effective method in improving the accuracy of multi-modal medical image registration in this work. We believe more than ever that more high-quality data sets will gradually emerge with the continuous improvement of the intelli-

Table 2
Quantitative results of the image-to-image translation.

		CycleGAN [20]	NeMAR[14]	SepCycleReg	Ours
RMSE↓	T1→T2	0.2367(0.0435)	0.2413(0.0377)	0.2299(0.0519)	**0.2263(0.0345)**
	T2→T1	0.157(0.0226)	0.1553(0.0257)	0.151(0.0293)	**0.1495(0.026)**
PSNR↑	T1→T2	23.15(1.16)	22.96(1.45)	23.45(1.56)	**23.5(1.18)**
	T2→T1	22.5(1.09)	22.63(1.38)	22.92(1.54)	**22.97(1.3)**
SSIM↑	T1→T2	0.8435(0.0145)	0.8469(0.012)	0.8461(0.0184)	**0.8473(0.0172)**
	T2→T1	0.8464(0.0154)	0.8482(0.0171)	**0.8515(0.0175)**	0.8485(0.0161)

gent medical system. The development prospects of deep learning methods are promising, while conventional methods like Elastix have relatively smaller development space.

In conclusion, the proposed CoCycleReg provides a simple, steady and good performing training framework for the image-to-image translation-based multi-modal image registration and we hope the collaborative cycle-consistency will be useful to push this frontier.

CRediT authorship contribution statement

Chenyu Lian: Conceptualization, Methodology, Writing - original draft. **Xiaomeng Li:** Investigation, Writing - review & editing. **Lingke Kong:** Investigation, Software. **Jiacheng Wang:** Visualization, Software. **Wei Zhang:** Investigation, Validation. **Xiaoyang Huang:** Validation, Formal analysis. **Liansheng Wang:** Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Fundamental Research Funds for the Central Universities (Grant No. 20720190012, 20720210121).

References

[1] C.S. Kidwell, J.A. Chalela, J.L. Saver, S. Starkman, M.D. Hill, A.M. Demchuk, J.A. Butman, N. Patronas, J.R. Alger, L.L. Latour, et al., Comparison of mri and ct for detection of acute intracerebral hemorrhage, Jama 292 (15) (2004) 1823–1830.
[2] K. Sandrasegaran, A. Rajesh, D.A. Rushing, J. Rydberg, F.M. Akisik, J.D. Henley, Gastrointestinal stromal tumors: Ct and mri findings, European radiology 15 (7) (2005) 1407–1414.
[3] I. Bankman, Handbook of medical image processing and analysis, Elsevier, 2008.
[4] S. Oh, S. Kim, Deformable image registration in radiation therapy, Radiation oncology journal 35 (2) (2017) 101.
[5] M.A. Schmidt, G.S. Payne, Radiotherapy planning using mri, Physics in Medicine & Biology 60 (22) (2015) R323.
[6] F. Liu, J. Cai, Y. Huo, C.-T. Cheng, A. Raju, D. Jin, J. Xiao, A. Yuille, L. Lu, C. Liao, et al., Jssr: A joint synthesis, segmentation, and registration system for 3d multi-modal image alignment of large-scale pathological ct scans, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16, Springer, 2020, pp. 257–274..
[7] B.B. Avants, C.L. Epstein, M. Grossman, J.C. Gee, Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain, Medical image analysis 12 (1) (2008) 26–41.
[8] M.P. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F.V. Gleeson, M. Brady, J.A. Schnabel, Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration, Medical image analysis 16 (7) (2012) 1423–1435.
[9] B. Zitova, J. Flusser, Image registration methods: a survey, Image and vision computing 21 (11) (2003) 977–1000.
[10] K. Marstal, F. Berendsen, M. Staring, S. 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, 2016, pp. 134–142..
[11] X. Cao, J. Yang, L. Wang, Z. Xue, Q. Wang, D. Shen, Deep learning based inter-modality image registration supervised by intra-modality similarity, in: International workshop on machine learning in medical imaging, Springer, 2018, pp. 55–63.
[12] C.K. Guo, Multi-modal image registration with unsupervised deep learning, Ph. D. thesis, Massachusetts Institute of Technology (2019)..
[13] D. Wei, S. Ahmad, J. Huo, W. Peng, Y. Ge, Z. Xue, P.-T. Yap, W. Li, D. Shen, Q. 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, Springer, 2019, pp. 512–520.
[14] M. Arar, Y. Ginger, D. Danon, A.H. Bermano, D. Cohen-Or, Unsupervised multi-modal image registration via geometry preserving image-to-image translation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2020, pp. 13410–13419..
[15] Z. Xu, J. Luo, J. Yan, R. Pulya, X. Li, W. Wells, J. Jagadeesan, Adversarial uni-and multi-modal stream networks for multimodal image registration, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 222–232.
[16] G. Balakrishnan, A. Zhao, M.R. Sabuncu, J. Guttag, A.V. Dalca, An unsupervised learning model for deformable medical image registration, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 9252–9260..
[17] G. Balakrishnan, A. Zhao, M.R. Sabuncu, J. Guttag, A.V. Dalca, Voxelmorph: a learning framework for deformable medical image registration, IEEE transactions on medical imaging 38 (8) (2019) 1788–1800.
[18] M. Jaderberg, K. Simonyan, A. Zisserman, et al., Spatial transformer networks, Advances in neural information processing systems 28 (2015) 2017–2025.
[19] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE transactions on image processing 13 (4) (2004) 600–612.
[20] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 2223–2232.
[21] B. Kim, J. Kim, J.-G. Lee, D.H. Kim, S.H. Park, J.C. Ye, Unsupervised deformable image registration using cycle-consistent cnn, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 166–174.
[22] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1125–1134..
[23] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, D. Shen, Medical image synthesis with context-aware generative adversarial networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017, pp. 417–425.
[24] X. Liang, L. Chen, D. Nguyen, Z. Zhou, X. Gu, M. Yang, J. Wang, S. Jiang, Generating synthesized computed tomography (ct) from cone-beam computed tomography (cbct) using cyclegan for adaptive radiation therapy, Physics in Medicine & Biology 64 (12) (2019) 125002.
[25] L. Kong, C. Lian, D. Huang, Y. Hu, Q. Zhou, et al., Breaking the dilemma of medical image-to-image translation, Advances in Neural Information Processing Systems 34..
[26] D. Mahapatra, B. Antony, S. Sedai, R. Garnavi, Deformable medical image registration using generative adversarial networks, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, 2018, pp. 1449–1453.
[27] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
[28] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778..
[29] L.R. Dice, Measures of the amount of ecologic association between species, Ecology 26 (3) (1945) 297–302.
[30] J. Ashburner, A fast diffeomorphic image registration algorithm, Neuroimage 38 (1) (2007) 95–113.
[31] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, et al., The multimodal brain tumor image segmentation benchmark (brats), IEEE transactions on medical imaging 34 (10) (2014) 1993–2024.
[32] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, J.B. Freymann, K. Farahani, C. Davatzikos, Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features, Scientific data 4 (1) (2017) 1–13.
[33] C. Qin, B. Shi, R. Liao, T. Mansi, D. Rueckert, A. Kamen, Unsupervised deformable registration for multi-modal images via disentangled representations, in: International Conference on Information Processing in Medical Imaging, Springer, 2019, pp. 249–261.
[34] A. Hore, D. Ziou, Image quality metrics: Psnr vs. ssim, in: 2010 20th international conference on pattern recognition, IEEE, 2010, pp. 2366–2369.
[35] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., Pytorch: An imperative style, high-performance deep learning library, Advances in neural information processing systems 32 (2019) 8026–8037.
[36] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, in: Proceedings of the IEEE international conference on computer vision, 2015, pp. 1026–1034..

Chenyu Lian received the B.S. degree from Xiamen University in 2021 and now is a master student in the Department of Computer Science, Xiamen University, Xiamen, China. His main research interests include medical image analysis and machine learning.

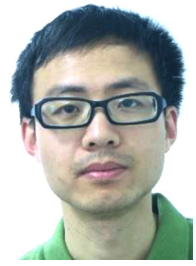
Dr. Xiaomeng Li is an Assistant Professor at the Department of Electronic and Computer Engineering at The Hong Kong University of Science and Technology. Before joining HKUST, she was a Postdoctoral Research Fellow at Stanford University. She obtained my Ph.D. degree from The Chinese University of Hong Kong. Her research lies in the interdisciplinary areas of artificial intelligence and medical image analysis, aiming at advancing healthcare with machine intelligence.

Lingke Kong received the M.S. degree from HuaQiao University in 2020 and now is an algorithm engineer in the department of Scientific Research from Manteia Technologies Co.,Ltd., Xiamen, China. His main research interests include medical image registration and machine learning.

Jiacheng Wang received the B.S. degree from Xiamen University in 2018 and now is a Ph.D. student in the Department of Computer Science from Xiamen University, Xiamen, China. His main research interests include medical image processing and machine learning.

Wei Zhang, researcher at Manteia Technologies Co.,Ltd., located in Xiamen, China, got master degree of engineering at Nanjing University of Science and Technology in 2018. His current research interests focus on deep learning-based applications in the field of radiation therapy, including medical image analysis, automated planning.

Xiaoyang Huang is currently an Assistant Professor in the Department of Computer Science, Xiamen University, Xiamen, China. His research interests include medical image processing.



Liansheng Wang received the Ph.D. degree in Computer Science from the Chinese University of Hong Kong in 2012. He is currently an Associate Professor in the Department of Computer Science, Xiamen University, Xiamen, China. His research interests include medical image processing and analysis, machine learning, big medical data.