A Novel CNN-based Model for Medical Image Registration

Hui GAO, Mingliang LIANG*

Zhengzhou Railway Vocational and Technical College, Henan, 451460, China

Abstract—The registration of the deformable image is applied widely to image diagnosis, the monitoring of the disease, and the navigation of the surgery with the aim of learning the correspondence of the anatomist among an image of motion and an image of static. The procedure of the registration of an image mainly includes three steps: the creation of a model of the deformation, a function design for the mensuration of the similarity, and the step of learning for the optimization of the parameter. In the current article, 2-stream architecture is designed, which has the ability to sequentially estimate the fields of the registration of the multi-level by a couple of the pyramids of the feature. In this paper, a 3D network of the encoder-decoder with the 2-stream is designed, which calculates 2 pyramids of the feature of the convolutional as separately by 2 volumes of the input. Also, the registration of the pyramid of the sequential is proposed, which in it, a trail of the modules of the pyramid registration (PR) for the prediction of the fields of the registration of the multi-level is designed as straight by the pyramids of the feature of the decoding. In addition, the modules of PR can be augmented with the computation of the 3D correlations of the local among the pyramids of the feature, which this work leads to the further improvement of the presented approach. Thus, it is capable of collecting the detailed anatomical structure of the brain. The proposed method is tested in three criterion datasets about the registration of MRI of the brain. The evaluation outcomes display that the presented approach outperforms the advanced approaches with a big value.

Keywords—*Image registration; convolutional neural network; Pyramid Registration (PR); encoder-decoder*

I. INTRODUCTION

Today, various imaging methods such as MRI, CT, PET, SPECT, and ultrasound imaging are used for the identification of the anatomical structure and the physiological performance of the human body so that each one of them provides specific information to doctors [1], [2]. The medical images usually do not match each other, and to use their information together, it is necessary to apply the methods of the registration of the image [3], [4], [5], [6], [7]. The registration of the image is the procedure of the orientation of the spatial of 2 images from the same scene in a way that their features can be easily related to each other. These images may be created by different sensors or one sensor at different times [3].

Due to the used images, the registration of medical images can be divided into two types: (1) the registration of the same type; (2) the registration of the different types [3]. In the registration of the same type, the used images are prepared by the same type of sensor. For example, to check the drug performance on cancer tumors, the MRI image is recorded

from the patient at certain time intervals. With the registration of these images, it is possible to check the progress of the disease or the progress of its improvement [8]. However, in the registration of the different types, the goal is the registration of the recorded images by two different types of sensors. For example, it can refer to the registration of image of MR anatomical by functional (physiological) PET image and with the registration of these two types of images, can achieve the anatomical features and the functional features in an image [9].

The registration of medical images is divided into two categories in terms of the application: 1) The image registration of various persons. 2) The registration of 2 images from the identical person. The registration of images from the different persons is used for their physiological anatomical comparison and also in the preparation of the medical atlases. Also, the image registration of one person is more used to integrate the information of the different images or to evaluate the treatment by observing the changes in the recorded images at different times [3]. In the image registration process, the motion image $I_M(x)$ ($I_M: \Omega \to R$ which in it, Ω displays the image scope of I) is transformed for the matching on the fixed image $I_F(x)$ ($I_F: \Omega \to R$) which in it, x = (x. y. z) $\in \Omega$ is the coordinates of spatial of the pixels. Namely, in the image's registration, the purpose is the finding of the transformation $y = T(x)$ ($such\ that$ T: $\Omega \to \Omega$) in the way that the points of the corresponding on 2 images $I_F(x)$ and $I_M(y)$ are coincided with each other. For this purpose, different transformations can be used, and these transformations differ from each other in terms of the computational complexity and the flexibility [4].

The transformations used in the image registration are distributed into 2 general groups: the non-hard transformations and the hard transformations. The hard transformations (such as the common geometric transformations including the Affine, the rigid body, the translation, the rotation, the rescaling, etc.) can be parametrically defined, and they are the same for all image pixels. These types of transformations are usually used in the preparation of the medical atlases and in the general registration of the medical images [5], [6], [7], [9], [10], [11], [12], [13], [14]. On the other hand, the non-hard transformations (in comparison to the hard transformations) have more flexibility for image registration. In this type of transformation, a vector field is transferred for the description of the changes between the static image and the motion image so that with the use of it, the corresponding point with each pixel of the static image can be identified in the motion image. Usually, to achieve a proper registration between 2 images, it is essential to use the hard transformation and the non-hard transformation together. In this way, first, by finding the

appropriate hard transformation, the images are generally matched with each other, and then, by optimizing the appropriate non-hard transformation, the remaining local differences between them are compensated [15].

In this paper, a non-hard transformation is used for the image registration. In the current article, a 2 -stream architecture is designed, which has the ability to sequentially estimate the fields of the registration of the multi-level by a couple of the pyramids of the feature. In this paper, a 3D network of the encoder-decoder with the 2-stream is designed, which calculates 2 pyramids of the feature of the convolutional as separately by 2 volumes of the input. Also, the registration of the pyramid of the sequential is proposed, which in it, a trail of the modules of the pyramid registration (PR) for the prediction of the fields of the registration of the multi-level is designed as straight by the pyramids of the feature of the decoding. In addition, the modules of PR can be augmented with the computation of the 3D correlations of the local among the pyramids of the feature, which this work leads to the further improvement of the presented approach. Thus, it is capable of collecting the detailed anatomical structure of the brain.

This article is introduced with the aim of improving the registration of successive pyramids. This is done with advanced PR modules that increase performance. The overall contributions can be summarized as follows: (1) a two-stream 3D encoder-decoder network is designed to compute two convolution feature pyramids separately from two input volumes, and generate more robust deep features for deformation estimation. (2) Sequential pyramid registration is proposed in which a sequence of registration fields is estimated by a set of designed pyramid registration modules. The estimated registration fields perform successive sweeps on the decoding layers, which gradually refine the feature pyramids from coarse to fine. It equips the model with a strong ability to handle large deformations. (3) The PR module can be enhanced by computing local 3D correlations (between two feature pyramids) followed by multiple residual convolutions, which gather richer local details of the anatomical structure for better estimation of deformation fields. It leads to the improvement of previous methods. In addition, 3D correlations with more complex layers in the advanced PR module can enlarge the receiver field, which further increases the ability to handle large deformations. The continuation of the current article is as follows. Section II presents an overview of related works. Section III, the proposed two-stream architecture is presented. Section IV, the used datasets, the designed experiments, and the obtained outcomes are provided. Eventually, Section V presents the general conclusions and suggestions for future research.

II. RELATED WORKS

So far, the researchers have used the non-hard transformation and the hard transformation for the registration of the images of the medical on various works. In the current part, in summary, the articles in this field are provided. For example, Horm and Schunk [16], for the first time, have used the optical flux model for the non-hard registration of the images. This model is proposed by adapting the motion of the relative of objects and also the motion of the relative of the

viewer. In this method, the authors suppose the brightness intensity of pixels corresponding to the static images and the motion images do not differ as significantly. In this way, by writing the Taylor series on the brightness intensity function of the motion image, the light flux relation can be obtained for the estimation and for the evolution of the transfer vector field (the non-hard transformation) [16]:

$$\nabla I. u + I_t = 0 \tag{1}$$

where, $\nabla = [\frac{\partial}{\partial x}. \frac{\partial}{\partial y}. \frac{\partial}{\partial z}]^T$ is the gradient operator. Also, $I(x.t)$ ($x = [x.y.z]^T$ displays the coordinates, and t is the time) represents a series of consecutive frames and $u(x) = [u_1(x).u_2(x).u_3(x)]^T$ ($u: \Omega \rightarrow R^3$) is the displacement vector that should be optimally adjusted. It should be noted that here $u_1(x)$, $u_2(x)$ and $u_3(x)$ represent the factors from the displacement vector on the axis of x, the axis of y, and the axis of z (and in coordinates of x). In the application of the optical flux method for the image registration, it is assumed that I_F and I_M are two consecutive temporal frames from I (namely, $I(x.0) = I_F(x)$ or $I(x.1) = I_M(x)$). In this case, the temporal derivative $I_t = \partial I/\partial t$ (at $t = 0$) after the discretization will be equivalent to the disagreement between the image of static and the image of motion (namely, $I_t(x.0) \approx I_M(x) = I_F(x)$). It is obvious that by solving the above equation under the latter conditions, the vector field u will be optimized to maximize the likeness among the corresponding cases in I_F and I_M.

The optical flux model has been used widely in the non-hard registration of the images of the medical, which, for example, can refer to the research of Palos et al. [15] After the general registration of the images based on the hard Affine transformation, they used the speed function of the grayscales difference in the optical flux model for the adaption of the details. In study [17], the authors have provided a novel registration method between the MRI images and the images of the ultrasound. To decrease the speckle noise in the procedure of registration, they have used the semi-automatic segmentation method for the magnetic resonance images and for the ultrasound images by using the active contour model. Then, they built a strong optical flux model between the segmented ultrasound images and the magnetic resonance images, and they estimated the flow field vector by using the Gaussian pyramid. Cooper and Ritter [18] have used the optical flux model for the evaluation of registration in 2D images of medical and 3D images of medical and also for the calculation of disagreements in images of static and images of motion. Also, Cao et al.[19] have used the symmetric optical flux model for the registration of 4-dimensional images of CT of the chest to deal with the image registration problems (which are caused by the variation of the local illumination intensity and the large displacements).

In the other article, namely in [20], the authors have applied the model of the elastic for the registration of the images of the brain of the human:

$$\mu \Delta u + (\lambda + \mu)\nabla(\nabla \cdot u) + f = 0 \tag{2}$$

μ and λ are the constants of the elasticity, and f is an external force. The above model is extended based on the relation of the linear elasticity of Navier for the field of the

displacement u. It describes the deformation of a material of the elastic of the deformable beneath the impression of the outside force f.

The model of the elastic has also been applied widely for the registration of medical images. For the example, it can refer to the work of Marami et al. [21]. They have estimated the prostate deformation, which is made from the crab of the prostate, by using the method based on the finite element and the model of the elastic. For this goal, the T2-weighted image of MRI from the prostate tumor before the treatment was compared with the T2-weighted image of MRI during the treatment. Mahapatra and Sun [22] have used the elastic model along with the field of the random of Markov and the integration of the information of the segmentation to improve medical image registration. In study [23], the authors also used the elastic model to align the perfusion sequence of the images of MRI of the cardiac. According to the assumption that states in the aligned time curve should have the property of sparseness, they have presented a framework for the alignment of the time-dependent physical phenomena. The introduced method had satisfactory results in the elastic deformation, and the efficiency of the approach had an important betterment in comparison to previous methods. Khallaghi et al. [24] have presented a dynamic three-dimensional registration algorithm based on the elastic model, and they have shown its performance by using the similarity method based on intensity and the volume overlap on three different clinical datasets (including the artery, the liver, and the kidney). Their feature-based algorithm had the proper time and the proper accuracy.

In another work, Christensen [25] provided the model of the fluid of the viscous for the registration images by using the relation of Nessavier-Stokes, which is according to the following relation:

$$\mu\Delta v + (\lambda + \mu)\nabla(\nabla \cdot v) + f = 0 \qquad (3)$$

Where the velocity field $v(x) = [v_1(x). v_2(x). v_3(x)]^T$ $(V: \Omega \to R^3)$ is described as a derivative of the temporal field of displacement u according to the below relation:

$$v = \frac{du}{dt} = \frac{\partial u}{\partial t} + v_1 \frac{\partial u}{\partial x} + v_2 \frac{\partial u}{\partial y} + v_3 \frac{\partial u}{\partial z} \qquad (4)$$

In this method, in order to achieve the large transformations (while maintaining the image continuity), the field of the velocity is applied rather than a field of the displacement. In general, in elastic models and in viscous fluid models, the force of the outside changes the form of the motion image in the direction of the registration with the static image. This force is elected due to the usage of the image and the kind of the image, and it can be the similarity criterion gradient, the disagreement of the grayscale levels, or the interval among the curves of the corresponding 2 images. Another example, it can refer to the work of Agostino et al [26]. They have used the common information criterion gradient as a force of the outside on the model of the fluid of the viscous for registration of images of the non-homogeneous brain.

In 2009 [27], the authors provided a standard database to evaluate the registration accuracy in the non-hard transformation-based methods. They have identified a set of index points in the four-dimensional CT images from the chest of five patients, on the condition of the most aspiration and on the condition of the most exhalation, with high accuracy. Obviously, if the registration of two images from the above set is more accurate, next, the points of the index of the corresponding 2 images will be closer to each other. Next, they introduced the MLS method for the registration of the images. In this method, the index points are first matched based on the maximum correlation criterion, and then the optimal Affine transformation is obtained by using the least squares error method. The above database has been used in several pieces of research to evaluate the precision of registration of the image. For example, in another work, which is called 4DLTM, the same group estimated the movement direction of each index pixel as a polynomial function with the registration of the images of a period from the inhalation and the exhalation. For this purpose, they have used the assumption of the optical flux model that states that the brightness of the pixels is constant during the displacement [28]. In another research, with the presentation of the ALK method, they have improved the performance of the previous algorithm by using the polynomial estimation of the image brightness intensity and the Tikhonov regularizer [29]. Also, in the CCLG method, instead of the optical flux, the mass conservation equation has been used for the modeling of the movement of the voxels [30]. In the LFC method, the movement of the voxel is modeled by using the compressible flux based on the sum of the non-linear squares [31]. Due to the use of temporal information, the last three algorithms can be considered as the four-dimensional methods.

Also, the Demons method uses the static image as a local force for the movement of the voxels of the motion image in order to match the static image [32]. The methods below are all improved versions of the Demons algorithm for image registration based on their grayscale level difference: PF, EPF, AF, DF, ADF, and IC [33]. By using the frame registration method and also by using the improved version of the squared error sum criterion, Li et al. [34] have presented an algorithm called BM for the transformable registration of the images.

III. THE PRESENTED APPROACH

In the current part, the details of the presented approach are presented. The proposed method in this paper consists of three main components, which are as follows: (a) a network of the encoder-decoder of 2-stream for the calculation of the pyramids of the feature, (b) the registration of the pyramid of the sequential, and (c) the bettered modules of PR.

A. Basic Topics

The purpose of the registration of 3D images of the medical is the estimation of the field of the deformation Φ that it can warp the volume of the motion $M \subset R^{H \times W \times D}$ to the volume of the static $F \subset R^{H \times W \times D}$ such which the volume of the warped $W = M \circ \Phi \subset R^{H \times W \times D}$ can be precisely equaled with constant F. $M \circ \Phi$ is used for the determination of the implementation of the field of the deformation Φ on the volume of motion by using the operation of the warping. The registration of the image is described as a problem of the optimization:

$$\hat{\Phi} = arg\, min_\Phi\, \zeta(F. M. \Phi) \qquad (5)$$

$$\zeta(F.M.\Phi) = \zeta_{sim}(F.M \circ \Phi) + \lambda\zeta_{smooth}(\Phi) \qquad (6)$$

ζ_{sim} displays the function which measures the likeness among the image of the warped ($M \circ \Phi$) and the image of the static (F). ζ_{smooth} displays a regular limitation in the field of the deformation (Φ) that enforces spatial uniformity. ζ_{sim} and ζ_{smooth} can be described in different shapes. The last endeavors have been assigned to the development of a strong method for computation of the field of the deformation Φ.

B. The Proposed 2-Stream Model

The proposed method is based on an architecture of the encoder-decoder that is improved with the introduction of a two-stream design. The proposed architecture is displayed in Fig. 1. It can be seen from Fig. 1 that the base of the proposed method includes an encoder-decoder of 2-stream along with the collective factors. An encoder with a similar framework to U-Net [35] is used, which consists of 5 blocks of the convolutional. For the blocks, the bating for 1-th block of the convolution, every block has a layer of the 3D convolution for the down-sampling by the stride equal to 2, which this layer is coupled with an operation of ReLU. Therefore, the encoder decreases the resolution of the spatial from the volumes of the input with a factor equal to 16, as displayed in Fig. 2. In the step of the decoding, the connections of the skip are applied on the maps of the convolution of the corresponding procedure of the encoding and the procedure of the decoding. The maps of the convolutional with the low resolution are up-sampled. Then, they are joined to the maps with higher resolution. Next, a convolutional layer with a size equal to $3 \times 3 \times 3$ and an operation of ReLU is applied, as displayed in Fig. 1.

Eventually, 2 pyramids of the feature by the features of the convolutional with the multi-resolution are achieved, which are calculated separately by the volume of the motion and by the volume of static.

The presented 2-stream model lets us calculate the pyramids of the feature separately by 2 volumes of the input, and next, it allows us to predict the deformable fields from the more robust distinct learned features of the convolutional, that this point is the guidance for the bettered efficiency. The mentioned design differs from the existing networks with 1-stream, like [36]. These existing networks calculate the features of the convolutional by 2 integrated volumes, and they appraise the fields of the deformation with the use of the filters of the convolution with 1-stream. In addition, the presented 2-stream framework can calculate 2 pairwise pyramids of the feature, which in it, the fields of the layered deformation can be sequentially appraised in the various levels. The mentioned point permits the method to create a trail of the fields of the deformation with the designing of a novel approach for the registration of the sequential pyramid.

Advert, in the presented approach, the applied base in [37] is modified with the increment of the convolution blocks from 4 to 5. Also, for 5 layers, the channel number from 32 channels per layer to [8.16.16.32.32] is reduced. In addition, in the proposed method, the modified units in [37] have been removed for the maintenance of a lightweight, effective model. These changes lead to a significant reduction in the parameters of the model ($410K \rightarrow 175K$), while the model maintains a similar performance.
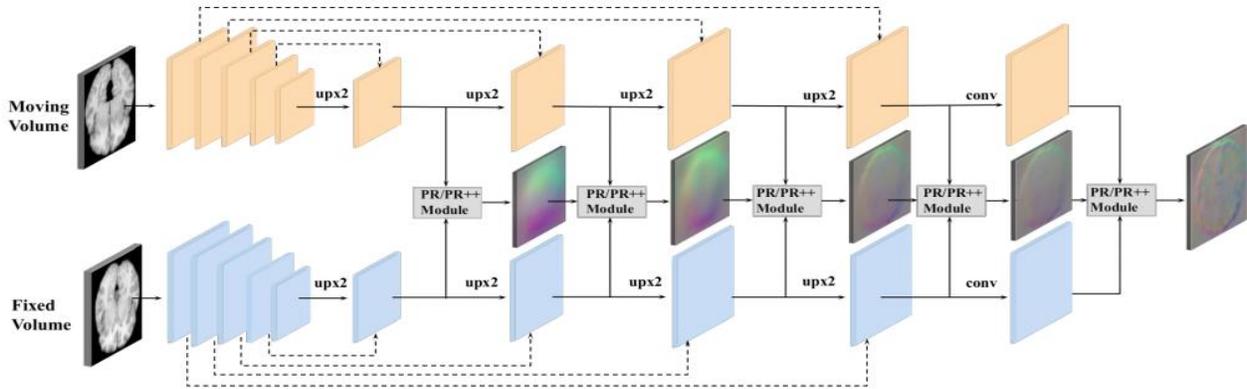

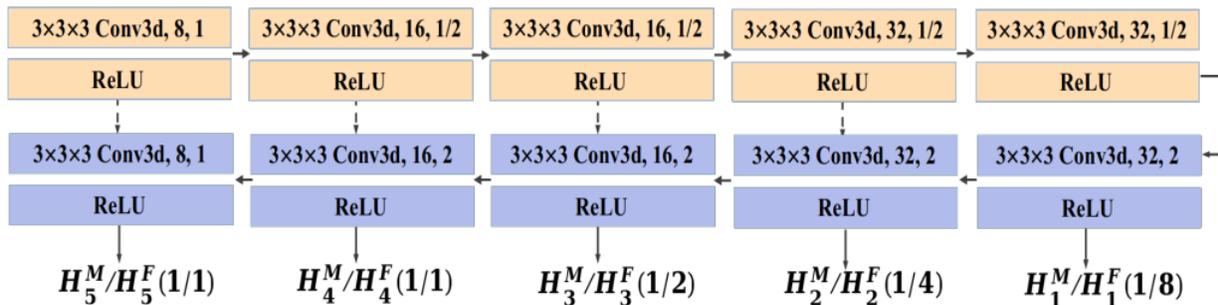Fig. 1. The proposed method for registration of the medical image.

Fig. 2. The basis of our presented approach, which includes one encoder (with the color of yellow) and one decoder (with the color of blue). H_l^M and H_l^F show the computed maps of the feature by volume of motion and volume of static, respectively.

C. The Sequential Pyramid Registration

In the presented method in the current paper, a novel registration of the pyramid is provided with the designing of a collection of the modules of PR that are sequentially performed in every layer of the decoding. This work permits the method to estimate the fields of the multiscale deformation by enhancing resolution and also allows the model to create a trail from the fields of the deformation of the pyramid, as displayed in Fig. 1.

Each module of PR apprises a field of the deformation in every layer of the decoding. As the input, the module of PR applies a couple of the features of the convolutional along with a computed field of the deformation by the prior layer (the batting for the 1-th layer of the decoding in which the field of the deformation is not accessible). As the output, the module of PR obtains an appraised field of the deformation in a level of resolution, and this field is applied to the subsequent level of the pyramid. The module of PR consists of a trail from the warping operation, the stacking operation, and the convolution operation (as displayed in Fig. 3(a)). These operations are repeatedly performed on the layers of the decoding.

In the case of sequential operations, in particular, the 1-th field of the deformation (Φ_1) is calculated in 1-th layer of the decoding. First, 2 calculated features of the convolutional in 1-th layer of the decoding are superimposed, and next, one convolution of 3D by the size equal to $3 \times 3 \times 3$ is applied for the estimation of the field of the deformation. The field of the deformation (Φ_1) displays the maps in 3D with the identical form along with the corresponding maps of the feature of the convolution. It can exploit the information of the background with the coarse level, like the structure of the anatomical with top level from the brain, that this information is encoded in the computed features of the convolutional in the subsequent layer of the decoding through the warping of the feature: (1) the top field of the deformation are up-sampled with the use of the interpolation of the bilinear by the coefficient equal to 2, which it is denoted by $u(\Phi_1)$; (2) next, it is used for the warping of the maps of the convolutional volume of the motion on subsequent layers with the use of an operation of the network sampling, as displayed in Fig. 3(a). Next, the warped maps of the convolutional are superimposed anew by the corresponding features of the convolutional, which are generated from the constant volume. In the following, a convolution operation is placed for the estimation of the novel field of the deformation. The mentioned procedure is recurred in every layer of the decoding. It is described as follows:

$$\Phi_l = C_l^{3\times3\times3}(H_l^M \circ u(\Phi_{l-1}).H_l^F) \tag{7}$$

where, $l = 1.2.\cdots.N$ displays the number of the layers of the decoding. $C_l^{3\times3\times3}$ represents a three-dimensional convolution in the layer of the decoding l. The functor $\circ$ represents the operation of the warping, which this operation maps H_l^M to H_l^F by using $u(\Phi_{l-1})$. H_l^M and H_l^F are the convolutional feature pyramids, which are calculated by the volume of motion and by the volume of static in the layer of the decoding l.

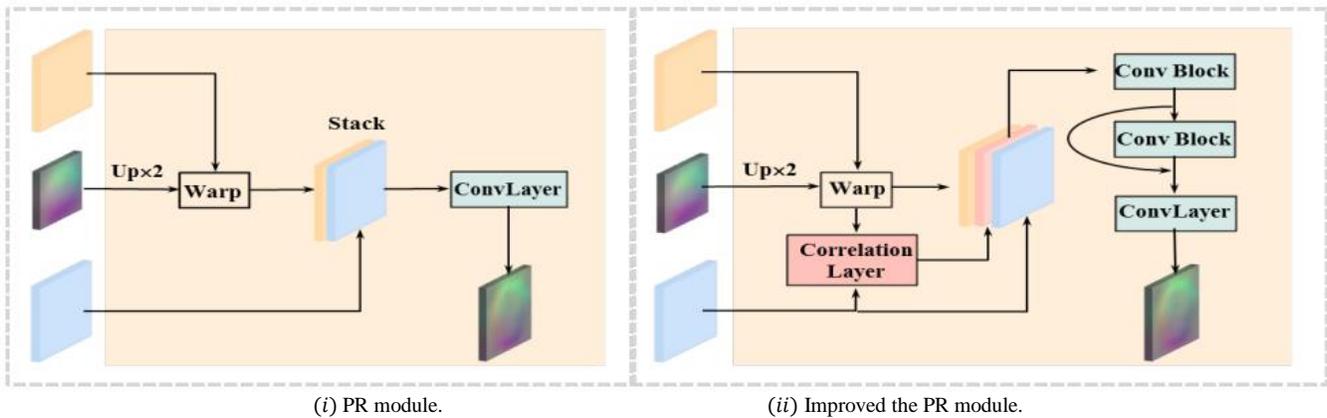(i) PR module.   (ii) Improved the PR module.

Fig. 3. The presented module of (i) PR and (ii) the improved PR. The module of the improved PR betters the module of PR with the computation of the features of the correlation via the increment of the convolutions of the residual.

Regarding the 3D layer of the correlation, on the module of the improved PR, a 3D layer of the correlation is defined for calculation of the correlations of the local among 2 volumes of the input on the space of the feature of the convolution. It permits us to collect the related features that are not addressed directly in the main module of PR. However, these features can emphasize the details of the local representation of the deep. In particular, suppose P_i^W, P_j^F represent the voxels of focal of the blocks of 3D (by the size equal to $(2k + 1)^3$), which are sampled by the maps of the feature of the warped motion volume and the maps of the feature of the static volume. The correlation among 2 blocks of the sampled three-dimensional can be calculated as follows:

$$C(w_i.f_i) = \frac{1}{(2k+1)^3} \sum_{n_w.n_f \in [-k.k]^3} P_{i+n_w}^W \times P_{j+n_f}^F \tag{8}$$

$n \in [-k.k]^3$ displays n iterations in a three-dimensional neighbouring equal to $[-k.k] \times [-k.k] \times [-k.k]$ from P_i^W or P_j^F. According to a block of 3D of the local in the maps of the feature of the motion volume, the computation of the correlations of the dense on total blocks of 3D sampled by the maps of the feature of the static volume is time-consuming. Thus, according to a block of 3D along by P_i^W, the correlations of the local with the sampling from a collection of P_j^F is calculated in a 3D neighborhood with a size equal to $d \times d \times d$. It can be performed as the convolutions of 3D. A stride

equal to $s_w = 1$ is applied to the sample as densely P_i^W by the maps of the features of the warped. Then, the neighboring of the correlation is adjusted by $d = 3$ in the maps of the feature of the static, which in it, P_j^F with the stride equal to $s_f = 1$ is sampled. Every block of the sampled has an identical size equal to $[-k.k] \times [-k.k] \times [-k.k]$. Also, the straight correlation among 2 blocks of the sampled is calculated with the use of the relation in Eq. (8). It produces the three-dimensional correlation maps (P^C) with the shape $[2 \times FL(d/s_f) + 1]^3 \times (H/s_w) \times (W/s_w) \times (D/s_w)$, where $[2 \times FL(d/s_f) + 1]^3 = 27$ displays the channel's number. FL represents a calculation of the floor. The created maps of the correlation have a similar form of 3D with the maps of the feature of the motion volumes and the static volumes. It ensures which 3 maps can be combined for more processing.

In the case of the convolutional enhancement, the proposed 2-stream framework calculates 2 pyramids of the feature separated by 2 volumes of the input. Nevertheless, the funds of the work of image registration are the learning of the correspondence of the anatomical of the powerful among 2 volumes on the space of the feature, which inspired us to model a novel approach for the more aggregation of the features of the calculated pyramid. The main subordinate of the presented module of the improved PR is to present a strong method for the learning of the details of the local of the richer by 2 features, that it ensures the further precise estimation of the fields of the deformation in the different scales. To richen the features of the learned, the calculated maps of the correlation are accumulated by 2 features of the pyramid in every decoding layer: the features of the warped by the volume of the motion and the features of the pyramid from the volume of the static (as displayed on Fig. 3(b)). The maps of the correlation have 27 channels in total layers of the decoding whenever the channel number of 2 features of the pyramid changes in the various layers as below: [8.16.16.32.32].

For this purpose, as shown in Fig. 3(b), 2 blocks of the convolution of 3D are applied to process the features of the stacked. Each convolution block contains two convolution layers with a size equal to $3 \times 3 \times 3$, which in the following, an operation of ReLU is done. 1-th block of the convolution significantly decreases the stacked feature channels from [43.59.59.91.91] to [8.16.16.32.32], which this reduction is stable by the used channels number on the module of PR for the computational performance. Additionally, a connection of the residual is used on a 2-th block of the convolutional in the attempt to preserve more background information and also in the attempt to extract the distinctive features of 2 volumes. Eventually, a layer of convolution is applied for the estimation of the field of the deformation. The novel module of the improved PR is used in the proposed framework of the 2-stream registration, which this work leads to an improved version of the proposed method.

D. The Field of the Final Deformation

The presented method creates five consecutive fields of the deformation with increasing resolution, which these fields are shown as $[\Phi_1.\Phi_2.\Phi_3.\Phi_4.\Phi_5]$. To calculate the field of the final deformation, a field of the deformation of the estimated is up-sampled with a coefficient equal to 2. Next, it is warped with the field of the below deformation that this field is estimated. This up-sampling operation and this warping operation are performed as frequently and as ordinally to create the final field of the deformation (see Fig. 4) that encodes the rich information of the background of the multi-level by the deformations of the multiscale. It permits the model to propagate the robust information of the background among the layers of the decoding of the hierarchy, in which the fields of the deformation of the predicted are progressively filtered in a manner of the coarse-to-fine. Thus, it collects information on the background of the top level and the features of the low level. The information of the background of the top-level equips the proposed method to the capability for the function by the deformations of the large-scale, whenever the features of the low-scale allow it to model the detailed information of the structure of the anatomical. The modules PR and modules of the improved PR are integrated into the proposed 2-stream framework, which results in a model of the trainable end-to-end. A correlation of the cross of the local of the negative is applied as the function of the loss, and it is combined by a regularization of the smooth. For example, it can mention a diffusion regularizer, which calculates the proximate gradients of the spatial with the use of the difference among the voxels of the neighboring.
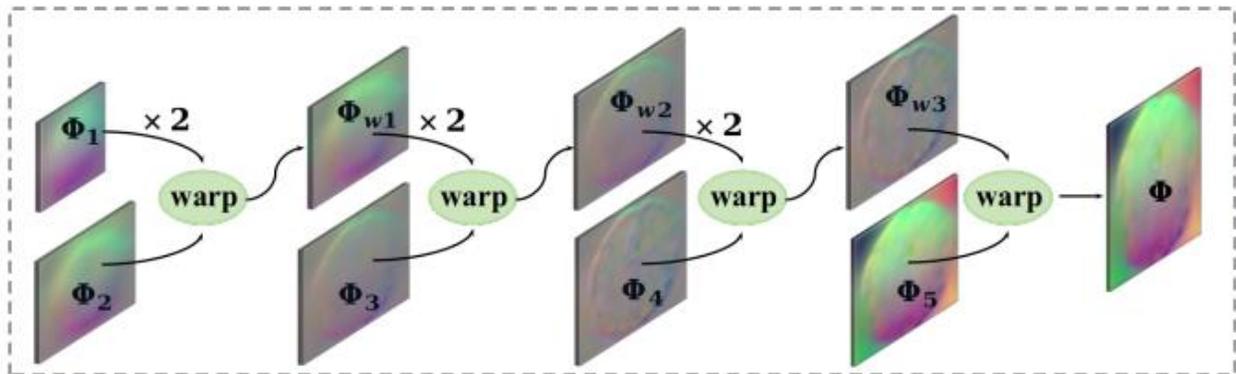

Fig. 4. The final field of the deformation is calculated from the sequential warping of the late field and the prior field.

IV. THE TESTS AND THE ANALYSIS OF THE RESULTS

In the current part, the experiments and the obtained outcomes by the proposed method of the registration of the brain MRI images are shown. In these experiments, first, a series from the atlas-based registration experiments is presented, where a field of the registration is calculated among one atlas or one volume of the reference. An atlas displays one reference or one volume of the average. Generally, it is made by frequently co-aligning the dataset from the volumes of MRI of the brain and averaging them together [40]. In this paper, a calculated atlas is used by using an external dataset [39]–[41]. Fig. 5 displays an instance from the image coupled with the use of the identical static atlas for total samples. Then, the result of the size of the dataset of the training in the image registration is briefly reviewed. Also, the results are provided in a dataset that includes the segmentations of the manual. Finally, the proposed method is trained by using a random pair of the subjects of the training as the input and the registration of the test among a pair of the subjects of the unseen. The language of the programming of Python has been used to implement these experiments. The presented method is implemented on a computer with 8G RAM and Core(TM) i7 CPU 3.0 GHz Intel(R). The network of the convolutional is performed on GPU, and the card used for the graphics on the current approach is GEFORCE 840M from NVIDIA.

A. The Used Datasets and the Evaluation Criteria

In this paper, the multi-study large-scale datasets of the scan of MRI of the brain of the T1-weighted are used from 3 publicly accessible datasets: ABIDE [42], OASIS [43], and FreeSurfer Buckner40 [39]. The details of the acquisition, the scopes of age of the subject, and the conditions of the health are various for every dataset. Total scans are down-sampled into a network with a size equal to $256 \times 256 \times 256$ and with the isotropic voxels with a length equal to $1\ mm$. The standard stages of the preprocessing (the normalization of the spatial and the extraction of the brain for every scan) were performed with the use of FreeSurfer [39], and the obtained images were cropped into $160 \times 192 \times 224$. Total MRI images were segmented as anatomically by FreeSurfer. Also, the control of the quality was applied with the use of the inspection of the visual for the detection of the errors of the gross in the results of the segmentation and in the alignment of the affine. The datasets are divided by a ratio equal to 85, 7.5, and 7.5 for validation, testing, and training, respectively. Additionally, it should be noted that the Buckner40 dataset was only used for the test step by using manual segmentation.

The obtaining of the registration of the real dense images for this data is not described as well because the multiple fields of the registration can provide analogous images of the warped. Here, first, the proposed approach is evaluated with the use of the volumetric overlay of the segmentations of the anatomical. If the field of the registration φ displays an exact match, then the areas f, $m°\varphi$, which correspond with the identical structure of the anatomical, must be overlapped as well (see Fig. 5). The overlap of the volume among structures is quantified with the use of the score of Dice.

In addition, symmetric normalization (SyN) [42] is used and the registration algorithm with superior performance [44] in a comparative study. To implement the SyN, the software package of the generally accessible Advanced Normalization Tools (i.e., ANTs) [45] is used, which comes with a cross-correlation similarity criterion. During the experiments, a step size of SyN equal to 0.25 is used and the Gaussian parameters ($9\ and\ 0.2$) in 3 scales by a maximum epoch equal to 201. The package of NiftyReg is used as the 2-th base. the 2D layer of the transformer of the spatial of the linear interpolation to $n - D$ is developed. Now, $n = 3$ is applied. Also, the ADAM optimizer [46] by the rate of learning equal to 0.0001 is used. The implementation includes a default with the iterations equal to 150000.
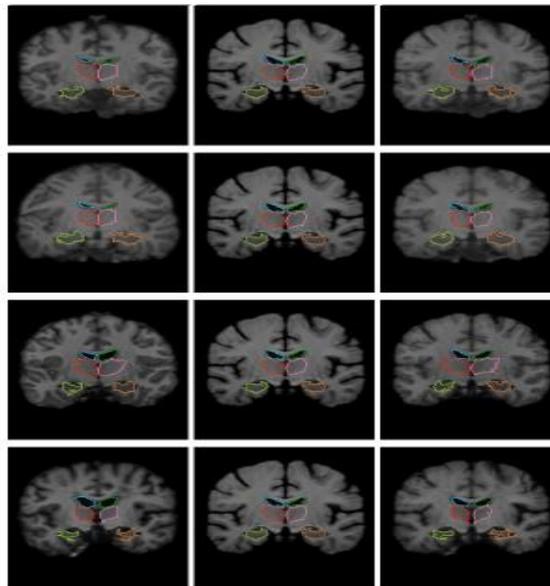

Fig. 5. The example of the extracted MRI crown crops from the couples of the input (columns $1 - 2$) and the obtained $M°\varphi$ from the presented approach (third column). Our presented approach can handle the different changes in the structure form, for example, the expansion / shrinking of ventricles in row 2 and row 3 or the stretching of the hippocampus in row 4.

B. The Experiments and the Obtained Results

As mentioned, first the proposed method is trained for the atlas-based registration. First the presented network is trained. Next, the results of the Dice score on the dataset of the test are reported. Table I shows the mean of the calculated scores of Dice for total subjects in structures of the baseline method (the Affine alignment), the ANTs method, the approach of NiftyReg, and the presented approach. Due to the Dice scores, the presented approach has comparable efficiency with the ANTs method and the approach of NiftyReg, and it is significantly superior to the alignment of Affine. The examples of the outcomes of the visual images of the warped from the proposed method are displayed in Fig. 5 and Fig. 6. The proposed method is capable of performing the changes of the considerable form for the different structures. Additionally, Fig. 7 shows the scores of Dice for every method in the box form. The presented approach obtains comparable scores of Dice with the approach of ANTs and the approach of NiftyReg for all structures. It performs better in several structures, like the ventricles of the lateral, and it also performs worse in other structures, such as the hippocampus.

Table I contains the number of voxels, and in these voxels, the determinant of Jacobian is the non-positive. It is discovered which whole approaches lead to the deformations by the little islands from these voxels, yet they are different in the majority of the vast of the voxels. Fig. 6 and Fig. 8 show several examples of the deformation fields of the presented approach. The presented approach does not have any explicit restrictions for the diffeomorphic deformations. The ANTs method and the NiftyReg method include implementations that can forcefully countenance the diffeomorphic transformations. However, during the parameter searches, these cases negatively affect the time of the run or the outcomes. In the current article, the base implementations are performed by the configurations that have obtained the foremost scores of Dice. Thus, it is found that this work creates good deformation regularization.

Next, the result of the size of the dataset of the training in the precision and the relationship among the depreciated optimization and the optimization of the sample-specific is evaluated. The proposed method is trained on the subsets that are different from the dataset of the training, and then the scores of Dice are reported. The test dataset is run with the fine-tuning of the obtained displacements by the proposed method for 100 iterations per the test pair. Fig. 9 shows the obtained outcomes. The small size of the dataset of the training, equal to 10 scans, makes the scores of Dice for the training and for the test slightly lower in comparison with the bigger training dataset size. Nevertheless, there is an important disagreement on the scores of Dice when the network is trained by 100 scans or by the total dataset.

TABLE I. THE MEAN SCORES OF DICE FOR THE ALIGNMENT OF AFFINE, THE ANTs, THE NIFTYREG, AND THE PROPOSED APPROACH. THE VOXELS NUMBER AND THE PERCENT OF THE VOXELS BY A NON-POSITIVE DETERMINANT OF JACOBIAN FOR EVERY APPROACH ARE DISPLAYED

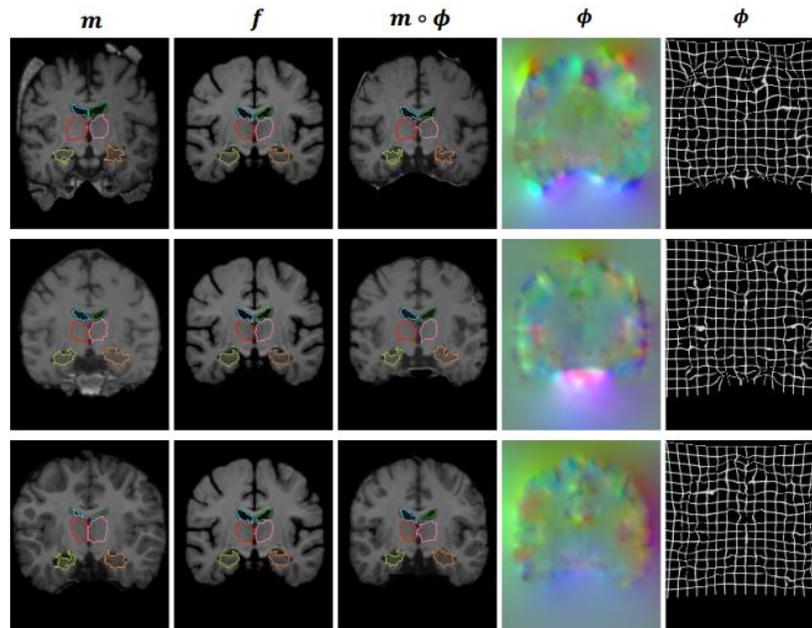| Method	Dice	$|J_\varphi| \leq 0$	% of $|J_\varphi| \leq 0$
Affine	0.584 (0.157)	0	0
ANTs	0.749 (0.136)	9662 (6258)	0.140 (0.091)
NiftyReg	0.755 (0.143)	41251 (14336)	0.600 (0.208)
Proposed Method	0.754 (0.144)	191352 (5985)	0.374 (0.118)

Fig. 6. An example of the fields of the deformation φ (columns $4-5$), which are exploited with the registration of a motion image (column 1) to a static image (column 2). The warped volume m°φ is displayed in column 3.
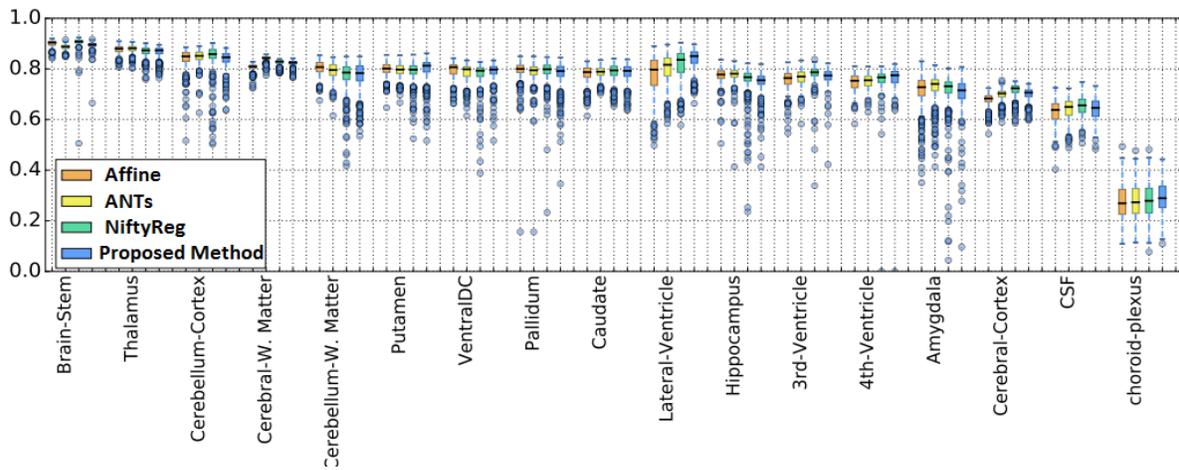
Fig. 7. The plots of the box of the scores of Dice for the various structures of the anatomical for the alignment of Affine, the ANTs, the NiftyReg, and the proposed method.
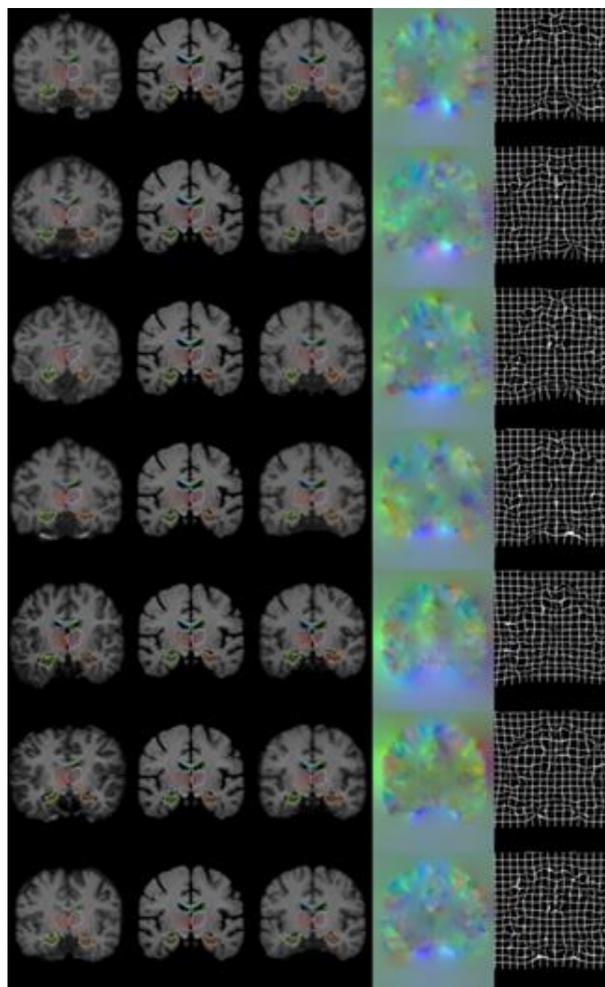

Fig. 8. An example of the fields of the deformation φ (columns $4-5$), which are exploited with the registration of a motion image (column 1) to a static image (column 2). The warped volume m°φ is displayed in column 3.
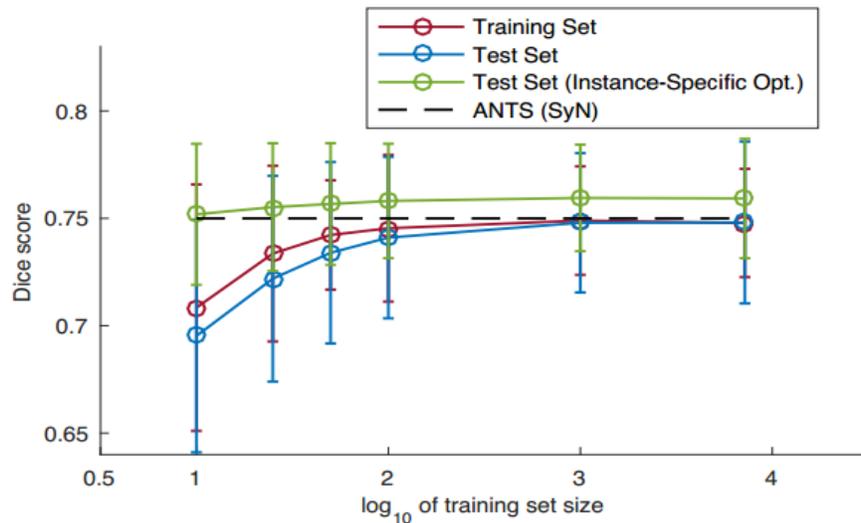
Fig. 9. Result of the size of the dataset of the training in the score of Dice.

TABLE II. THE RESULTS OF THE MANUAL ANNOTATION EXPERIMENT FOR THE AFFINE ALIGNMENT, THE ANTs, THE NIFTYREG, THE PRESENTED APPROACH, AND THE PRESENTED APPROACH BY OPTIMIZATION OF THE SAMPLE-SPECIFIC

Method	Dice
Affine	0.608 (0.175)
ANTs	0.776 (0.130)
NiftyReg	0.776 (0.132)
Proposed Method	0.773 (0.134)
Proposed Method with Instance-Specific Optimization	0.784 (0.132)

Since manual segmentation is not accessible for the uttermost datasets, the accessibility of the FreeSurfer segmentation enables a vast scope of the above experiments. In the current experiment, the registration test on the Buckner40 dataset (the us-seen) is used, which has 39 scans. The dataset of Buckner40 includes the specialist hand-lines from the identical used structures of the anatomical on the prior tests that here from them are used for the evaluation. The outcomes of the score are displayed in Table II. These outcomes display that our presented approach has similar behavior with the approach of ANTs and the approach of NiftyReg, which is stable with the 1-th test. The presented approach further improves the outcomes with the optimization of the sample-specific. In this dataset, the obtained results by the proposed method obtain slightly lower scores. However, they are bettered with optimization of the sample-specific to be comparable by the approach of ANTs and the approach of NiftyReg.

In the next experiment, the presented approach is trained for the registration of the subject-to-subject. In respect, there is a further variation on every registration. Thus, the feature number for every layer of the network is doubled. The efficiency of the presented approach is calculated by optimization of the sample-specific. Table III displays the mean scores of Dice of the test step in 250 test pairs, which are randomly selected for registration. The obtained scores of Dice from the presented approach (with the doubled number of features) are comparable with the approach of ANTs and the approach of NiftyReg. Also, the scores of Dice of the presented

approach with the sample-specific optimization are comparable with both baseline methods.

C. Discussion

The proposed method in this paper, with the unsupervised loss, has the similar performance to the state-of-the-art methods ANT and NiftyReg, in terms of Dice score, while it reduces the computation time from hours to minutes on a CPU and less than a second on a GPU. The proposed method is flexible and it handles partial observed or coarsely specified auxiliary information during training, which this can lead to improved Dice score while maintaining improved runtime. Also, the proposed method performs the degenerate optimization and it learns general performance parameters that are optimal for the entire training dataset. The sample-specific optimization improves the performance of the proposed method by one dice point. This is a small increase and it indicates that the depreciated optimization can lead to near-optimal registration.

The performance gain varies based on the quality and number of anatomical sections available. Given a labeled anatomical structure during training, subjects' registration accuracy for that label increases without negatively affecting other anatomy. If half or all labels are observed, or even a coarser segmentation is provided in training, the registration accuracy for all labels will improve during testing. While with one type of auxiliary data, the experiment was done in this study, but the proposed method can use other auxiliary data such as different methods or anatomical key points.

TABLE III.   THE RELATED RESULTS TO THE SUBJECT-TO-SUBJECT REGISTRATION BY USING THE AFFINE ALIGNMENT, THE ANTs, THE NIFTYREG, THE PRESENTED APPROACH, AND THE PRESENTED APPROACH BY OPTIMIZATION OF THE SAMPLE-SPECIFIC

Method	Dice
Affine	0.579 (0.173)
ANTs	0.761 (0.117)
NiftyReg	0.772 (0.117)
Proposed Method	0.763 (0.052)
Proposed Method with Instance-Specific Optimization	0.773 (0.119)

V.   CONCLUSIONS AND SUGGESTIONS

In this paper, a two-stream pyramid registration network with the improved PR module is presented for unsupervised 3D medical image registration. The presented approach, due to the design, has a 2-stream architecture that permits it to calculate 2 pyramids of the feature of the convolutional as separately by 2 volumes of the input. Next, the registration of the pyramid of the sequential by a collection of the modules of PR is provided for the estimation of a trail from the fields of the registration that these fields can filter the learned features of the pyramid incrementally in a manner of the coarse-to-fine through the warping of the sequential. The module of PR is augmented with the residual complexity enhancement and with the computation of the local correlation features. The proposed approach has a competitive performance by the approach of ANTs and the approach of NiftyReg according to the score of Dice. The obtained results from the tests in 3 datasets show that the proposed approach is flexible, and it controls the partially observed information of the auxiliary during the training; this point can result in improved scores of Dice. The proposed method performs the depreciated optimization, and it learns the general performance parameters that are optimal for the full dataset of the training. Also, the outcomes have displayed which optimization of the sample-specific betters the efficiency of the proposed approach.

The proposed method can use the other data of the auxiliary, like the various methods or the anatomical key points, which can be addressed in the subsequent works. Additionally, the proposed approach is a generic mechanism of learning, and it is not restricted to a specific kind of image or anatomy. Its effectiveness can be checked on other applications of the registration of images of the medical, like the scans of MRI of the cardiac or the images of CT of the lung. Also, by providing a proper function of the loss, like the information of the mutual, the proposed approach can do the registration of the multimodal. Thus, it is suggested to be considered in the subsequent works.

ACKNOWLEDGMENT

REFERENCES

[1] M. Saadatmand-Tarzjan and H. Ghassemian, "On analytical study of self-affine maps," Journal of Iranian Association of Electrical and Electronics Engineers, vol. 12, no. 3, pp. 77–92, 2016.

[2] M. Saadatmand-Tarzjan and H. Ghassemian, "On analytical study of self-affine maps," Journal of Iranian Association of Electrical and Electronics Engineers, vol. 12, no. 3, pp. 77–92, 2016.

[3] P. Parsa and R. Safabakhsh, "A New Method for Image Segmentation based on Multi-Objective Differential Evolution Fuzzy Clustering," Journal of Iranian Association of Electrical and Electronics Engineers, vol. 13, no. 2, pp. 103–114, 2016.

[4] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," Phys Med Biol, vol. 46, no. 3, p. R1, 2001.

[5] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "Elastix: a toolbox for intensity-based medical image registration," IEEE Trans Med Imaging, vol. 29, no. 1, pp. 196–205, 2009.

[6] J. Flusser and T. Suk, "A moment-based approach to registration of images with affine geometric distortion," IEEE transactions on Geoscience and remote sensing, vol. 32, no. 2, pp. 382–387, 1994.

[7] D. Li and Y. Zhang, "A novel approach for the registration of weak affine images," Pattern Recognit Lett, vol. 33, no. 12, pp. 1647–1655, 2012.

[8] P. Riyamongkol, W. Zhao, Y. Liu, L. Belayev, R. Busto, and M. D. Ginsberg, "Automated registration of laser Doppler perfusion images by an adaptive correlation approach: application to focal cerebral ischemia in the rat," J Neurosci Methods, vol. 122, no. 1, pp. 79–90, 2002.

[9] J. L. Boes et al., "Image registration for quantitative parametric response mapping of cancer treatment response," Transl Oncol, vol. 7, no. 1, pp. 101–110, 2014.

[10] P. Riyamongkol and W. Zhao, "The Hopfield neural network model for solving affine transformation parameters in the correlation method," in 2006 IEEE Region 5 Conference, IEEE, 2006, pp. 249–253.

[11] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," Int J Comput Vis, vol. 24, no. 2, pp. 137–154, 1997.

[12] Y. Chenoune, Y. Bouaoune, E. Delechelle, E. Petit, J. Garot, and A. Rahmouni, "MR/CT multimodal registration of short-axis slices in CT volumes," in 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2007, pp. 4496–4499.

[13] C. Niu, "Medical image registration based on mutual information using kriging probability density estimation," in 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, IEEE, 2006, pp. 3097–3099.

[14] J. Lin, Z. Gao, B. Xu, Y. Cao, and Z. Yingjian, "The effect of grey levels on mutual information based medical image registration," in The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2004, pp. 1747–1750.

[15] S. Samant, P. K. Nanda, and S. Sahoo, "Multimodal image registration based on weighted feature and mutual information," in 2011 IEEE Recent Advances in Intelligent Computational Systems, IEEE, 2011, pp. 420–424.

[16] G. Palos, N. Betrouni, M. Coulanges, M. Vermandel, V. Devlaminck, and J. Rousseau, "Multimodal matching by maximisation of mutual information and optical flow technique," in The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2004, pp. 1679–1682.

[17] B. K. P. Horn and B. G. Schunck, "Determining optical flow," Artif Intell, vol. 17, no. 1–3, pp. 185–203, 1981.

[18] M. Hou, C. Chen, D. Tang, S. Luo, F. Yang, and N. Gu, "Magnetic microbubble-mediated ultrasound-MRI registration based on robust optical flow model," Biomed Eng Online, vol. 14, pp. 1–15, 2015.

[19] J. Cooper, "Optical flow for validating medical image registration," in Proceedings of the 9th IASTED International Conference on Signal and Image Processing, ACTA Press/IASTED, 2003, pp. 502–506.

[20] Z. Cao, E. Dong, Q. Zheng, W. Sun, and Z. Li, "Accurate inverse-consistent symmetric optical flow for 4D CT lung registration," Biomed Signal Process Control, vol. 24, pp. 25–33, 2016.

[21] R. Bajcsy and S. Kovačič, "Multiresolution elastic matching," Comput Vis Graph Image Process, vol. 46, no. 1, pp. 1–21, 1989.

[22] B. Marami et al., "Elastic registration of prostate MR images based on estimation of deformation states," Med Image Anal, vol. 21, no. 1, pp. 87–103, 2015.

[23] D. Mahapatra and Y. Sun, "Integrating segmentation information for improved MRF-based elastic image registration," IEEE Transactions on Image Processing, vol. 21, no. 1, pp. 170–183, 2011.

[24] L. Cordero-Grande, S. Merino-Caviedes, S. Aja-Fernandez, and C. Alberola-Lopez, "Groupwise elastic registration by a new sparsity-promoting metric: Application to the alignment of cardiac magnetic resonance perfusion images," IEEE Trans Pattern Anal Mach Intell, vol. 35, no. 11, pp. 2638–2650, 2013.

[25] S. Khallaghi, C. G. M. Leung, K. Hastrudi‐Zaad, P. Foroughi, C. Nguan, and P. Abolmaesumi, "Experimental validation of an intrasubject elastic registration algorithm for dynamic‐3D ultrasound images," Med Phys, vol. 39, no. 9, pp. 5488–5497, 2012.

[26] G. E. Christensen, R. D. Rabbitt, and M. I. Miller, "Deformable templates using large deformation kinematics," IEEE transactions on image processing, vol. 5, no. 10, pp. 1435–1447, 1996.

[27] E. D'agostino, F. Maes, D. Vandermeulen, and P. Suetens, "A viscous fluid model for multimodal non-rigid image registration using mutual information," Med Image Anal, vol. 7, no. 4, pp. 565–575, 2003.

[28] R. Castillo et al., "A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets," Phys Med Biol, vol. 54, no. 7, p. 1849, 2009.

[29] E. Castillo, R. Castillo, J. Martinez, M. Shenoy, and T. Guerrero, "Four-dimensional deformable image registration using trajectory modeling," Phys Med Biol, vol. 55, no. 1, p. 305, 2009.

[30] C. B. Hoog Antink, T. Singh, P. Singla, and M. Podgorsak, "Evaluation of advanced Lukas–Kanade optical flow on thoracic 4D-CT," J Clin Monit Comput, vol. 27, pp. 433–441, 2013.

[31] E. Castillo, R. Castillo, B. White, J. Rojo, and T. Guerrero, "Least median of squares filtering of locally optimal point matches for compressible flow image registration," Phys Med Biol, vol. 57, no. 15, p. 4827, 2012.

[32] E. Castillo, R. Castillo, B. White, J. Rojo, and T. Guerrero, "Least median of squares filtering of locally optimal point matches for compressible flow image registration," Phys Med Biol, vol. 57, no. 15, p. 4827, 2012.

[33] J.-P. Thirion, "Image matching as a diffusion process: an analogy with Maxwell's demons," Med Image Anal, vol. 2, no. 3, pp. 243–260, 1998.

[34] X. Gu et al., "Implementation and evaluation of various demons deformable image registration algorithms on a GPU," Phys Med Biol, vol. 55, no. 1, p. 207, 2009.

[35] M. Li, Z. Xiang, L. Xiao, E. Castillo, R. Castillo, and T. Guerrero, "GPU-accelerated block matching algorithm for deformable registration of lung CT images," in 2015 IEEE International Conference on Progress in Informatics and Computing (PIC), IEEE, 2015, pp. 292–295.

[36] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer, 2015, pp. 234–241.

[37] D. Kuang and T. Schmah, "Faim–a convnet method for unsupervised 3d medical image registration," in Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 10, Springer, 2019, pp. 646–654.

[38] Y. Hu, E. Gibson, D. C. Barratt, M. Emberton, J. A. Noble, and T. Vercauteren, "Conditional segmentation in lieu of image registration," in Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22, Springer, 2019, pp. 401–409.

[39] M. Kang, X. Hu, W. Huang, M. R. Scott, and M. Reyes, "Dual-stream pyramid registration network," Med Image Anal, vol. 78, p. 102379, 2022.

[40] B. Fischl, "FreeSurfer," Neuroimage, vol. 62, no. 2, pp. 774–781, 2012.

[41] M. De Craene, A. du Bois d'Aische, B. Macq, and S. K. Warfield, "Multi-subject registration for unbiased statistical atlas construction," in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2004: 7th International Conference, Saint-Malo, France, September 26-29, 2004. Proceedings, Part I 7, Springer, 2004, pp. 655–662.

[42] R. Sridharan et al., "Quantification and analysis of large multimodal clinical image studies: Application to stroke," in Multimodal Brain Image Analysis: Third International Workshop, MBIA 2013, Held in Conjunction with MICCAI 2013, Nagoya, Japan, September 22, 2013, Proceedings 3, Springer, 2013, pp. 18–30.

[43] A. Di Martino et al., "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism," Mol Psychiatry, vol. 19, no. 6, pp. 659–667, 2014.

[44] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," J Cogn Neurosci, vol. 19, no. 9, pp. 1498–1507, 2007.

[45] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," Med Image Anal, vol. 12, no. 1, pp. 26–41, 2008.

[46] A. Klein et al., "Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration," Neuroimage, vol. 46, no. 3, pp. 786–802, 2009.

[47] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, "A reproducible evaluation of ANTs similarity metric performance in brain image registration," Neuroimage, vol. 54, no. 3, pp. 2033–2044, 2011.