2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS) | 978-1-6654-4121-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/CBMS52027.2021.00101

FIRE: Unsupervised bi-directional inter- and intra-modality registration using deep networks

Chengjia Wang
Edinburgh Imaging Facility QMRI,
Centre for Cardiovascular Science,
University of Edinburgh,

Edinburgh EH16 4TJ, UK

chengjia.wang@ed.ac.uk

Guang Yang
Faculty of Medicine, National Heart &
Lung Institute, Imperial College
London

London, SW3 6LY, UK

g.yang@imperial.ac.uk

Giorgos Papanastasiou*
School of Computer Science and
Electronic Engineering, University of
Essex

Colchester CO4 3SQ, UK

g.papanastasiou@essex.ac.uk
* Corresponding author

***Abstract*. Magnetic resonance imaging (MRI) benefits from the acquisition of multiple sequences (thereafter, referred to as "modalities") under a single imaging session. Each modality offers different complementary spatial and functional information in the clinical setting. Inter- and intra (across MR sequence slices)-modality image registration is an important pre-processing step across multiple applications in routine clinical workflows, such as when visual or quantitative imaging biomarkers need to be assessed across multi-sequence/multi-slice MRI data. This paper presents an unsupervised deep learning-based registration network that can learn affine and non-rigid transformations, simultaneously. Inverse-consistency is an important property that is commonly ignored in recent deep learning-based inter-modality registration algorithms. We address this issue through our proposed multi-task, cross-domain image synthesis architecture, in which we incorporated a new comprehensive transformation network. The proposed model learns a modality-independent latent representation to perform cycle-consistent cross-modality synthesis and uses an inverse-consistency loss to learn paired transformations, to align the synthesized with the target image. We name this proposed framework as "FIRE" due to the shape of its structure and we focus on interpreting model components to enhance model interpretability for clinical MR applications. Our method shows comparable and better performances against a well-established baseline method in experiments on multi-sequence brain MR data and intra-modality 4D cardiac Cine-MR data.**

Keywords: Inter- and intra-modality registration, inverse-consistency loss, deep learning interpretability

1. INTRODUCTION

Modern diagnosis from magnetic resonance imaging (MRI) data benefits from extracting complementary information by multiple MR sequences (thereafter, referred to as "modalities") acquired under a single imaging session, using non-ionising radiation. Visual and quantitative imaging biomarkers are derived and cross-assessed from multiple MR modalities, across different areas of the organ under investigation. Thus, image registration is necessary when analysing paired images acquired from different slice orientations, at different times, and/or using different modalities. This makes inter- and intra-modality image registration a critical pre-processing task within routine MR clinical workflows [1]. Previous deep learning-based image registration methods typically model the registration problem as an iterative optimization process in the setting of maximising an image similarity metric between the moving image and the fixed image, thus they are often computationally expensive [2].

As discussed in [2], in the last decade, a variety of deep learning-based methods have been designed to predict the geometric correspondence between a pair of images. Among these, the most robust methods were based on supervised learning, which requires manually generated ground truths, such as pre-aligned image pairs, simulated transformation fields or segmentation labels [3][4][5]. On the other hand, modern unsupervised methods [6] have been mostly examined on limited subsets of 3D volumes or 2D slices with only small misalignments, whilst commonly require affine registrations in the pre-processing. To perform both affine and non-rigid registrations, deep learning models commonly require the involvement of two independent models in the analysis pipeline, to address both types of transformation [7]. In this study, we present a novel unsupervised deep inter-modality registration network that can learn optimal affine and non-rigid transformations, simultaneously.

Our method solves n-D image registration problems through cross-modality image synthesis and inverse-consistent transformations [8]. The cycle consistency adversarial loss has been widely used within this type of methods. Moreover, inverse-consistency (or bi-directional) transformation has been a favourable property towards maintaining the topology and anatomy of organs. However, most previous studies have focused on estimating asymmetric transformations and therefore have failed to address topology and organ anatomy maintenance [2]. Two previous inverse-consistent models presented in previous studies [9][10] are close to our proposed method. However, [10] was developed for intra-modality registration, and [9] has only been examined for 2D non-rigid registration.

We named our proposed model as "FIRE" because of its architecture which reflects the shape of the character "火" (which represents "fire" in the Chinese language), as shown in Fig. 1. We present experiments demonstrating that our method achieves state-of-the-art performances in the setting of registering multi-sequence brain MR data with aggressive simulated deformations and intra-modality 4D (temporal resolution corresponds to the 4th dimension) cardiac MR data. To sum up, contributions of this paper include: (1) the "火"-shape FIRE architecture for inverse-consistent inter-modality registration; (2) simultaneous learning for affine and non-rigid transformation; (3) new regularization for non-rigid registration using the predicted affine transformation; and (4) model interpretability in the context of clinical MR applications examined (i.e. brain and cardiac MRI).

2372-9198/21/$31.00 ©2021 IEEE
DOI 10.1109/CBMS52027.2021.00101

2. METHOD

With two images x^{A} and x^{B}, the proposed FIRE model predicts two transformations $\varphi^{A\rightarrow B}$ and $\varphi^{B\rightarrow A}$ to warp the images into $x^{A} \circ \varphi^{A\rightarrow B}$ and $x^{B} \circ \varphi^{B\rightarrow A}$. Transformation fields are obtained by minimizing a loss $L(x^{A}, x^{B}, \varphi^{A\rightarrow B}, \varphi^{B\rightarrow A})$. Computations described in this section are based on input data normalized to the range [−1, 1].

2.1 Architecture

The FIRE model consists of five sub-networks (Fig. 1): a synthesis encoder, G, that extracts modality-independent features $G(x^{A})$ and $G(x^{B})$; two synthesis decoders, $F^{A\rightarrow B}$ and $F^{B\rightarrow A}$, that map the features extracted by G to synthesized images $\hat{x}^{B} = F^{A\rightarrow B}(G(x^{A}))$ and $\hat{x}^{A} = F^{B\rightarrow A}(G(x^{B}))$; and two transformation networks, $T^{A\rightarrow B}$ and $T^{B\rightarrow A}$, that predict the transformation fields $\varphi^{A\rightarrow B} = T^{A\rightarrow B}(G(x^{A}), G(x^{B}))$ and $\varphi^{B\rightarrow A} = T^{B\rightarrow A}(G(x^{B}), G(x^{A}))$. In the training stage, $G(x^{A})$ and $G(x^{B})$ are also warped into $G(x^{A}) \circ \varphi^{A\rightarrow B}$ and $G(x^{B}) \circ \varphi^{B\rightarrow A}$, then used to generate synthesized images, $\hat{x}_{T}^{B} = F^{A\rightarrow B}(G(x^{A}) \circ \varphi^{A\rightarrow B})$ and $\hat{x}_{T}^{A} = F^{B\rightarrow A}(G(x^{B}) \circ \varphi^{B\rightarrow A})$.

Encoder and Decoder for Synthesis

Fig. 2 outlines details of the architecture component referring to the network used for image synthesis. The encoder G contains an input convolutional layer, two downsampling convolutional layers and four Resnet blocks. The decoder network starts with four Resnet blocks, followed by two upsampling convolutional layers, followed by the output convolutional layers. All convolutional layers use a kernel size of 3, with an instance normalization layer.

Transformation Network

A transformation network $T^{\cdot\rightarrow\cdot}$ learns both an affine transformation φ_{af} and a non-rigid transformation φ_{nr} given $G(x^{A})$ and $G(x^{B})$. Fig. 3 shows the architecture of the transformation networks, $T^{A\rightarrow B}$ and $T^{B\rightarrow A}$. The affine transformation sub-network T_{af} has a similar structure to the original spatial transformation networks (STN). A global average pooling layer is used to resample conv features into a fixed size feature vector. Affine transformation is calculated using two fully connected layers. The non-rigid transformation sub-net $T_{nr}^{A\rightarrow B}$ receives $G(x^{A}) \circ T_{nr}^{A\rightarrow B}$ and $G(x^{B})$ as inputs and processes them as parallel layers. The extracted features are then concatenated to produce the non-rigid deformation $\varphi_{nr}^{A\rightarrow B}$. The Tanh layer is implemented on a normalized coordinate system where a coordinate $p \in [-1, 1]^{n}$ for a n-D image.

2.2 Loss Functions and Training process

The A→ B synthesis generates two synthesized images $\hat{x}^{B}$ and $\hat{x}_{T}^{B}$, where $\hat{x}^{B}$ is aligned with $\hat{x}^{A}$ and $\hat{x}_{T}^{B}$ is identical to the target image x^{B}. The backward synthesis and registration B→A are performed through the same pipeline using the corresponding "B→A" networks. The losses used for training the FIRE model include a synthesis loss, $\mathcal{L}syn$, and a registration loss, $\mathcal{L}reg$. A new regularization $\mathcal{R}$ is also involved for topology-preserving deformations. The loss function of the FIRE model is defined as:

$$\mathcal{L} = \mathcal{L}syn + \mathcal{L}reg + \mathcal{R} \qquad (1)$$

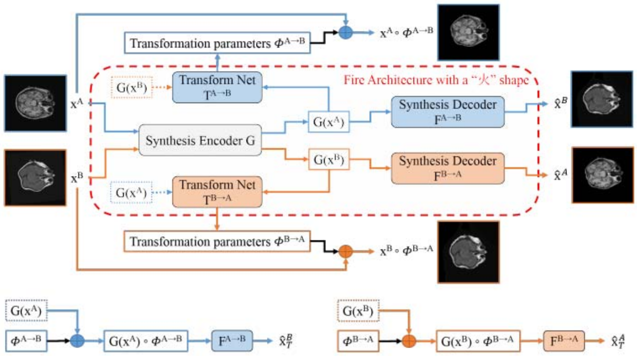
Transformation parameters $\Phi^{A\rightarrow B}$
$x^{A} \circ \Phi^{A\rightarrow B}$
Fire Architecture with a "火" shape
$G(x^{B})$
Transform Net $T^{A\rightarrow B}$
x^{A}
Synthesis Decoder $F^{A\rightarrow B}$
$\hat{x}^{B}$
$G(x^{A})$
Synthesis Encoder G
$G(x^{B})$
Synthesis Decoder $F^{B\rightarrow A}$
$\hat{x}^{A}$
x^{B}
$G(x^{A})$
Transform Net $T^{B\rightarrow A}$
Transformation parameters $\Phi^{B\rightarrow A}$
$x^{B} \circ \Phi^{B\rightarrow A}$
$G(x^{A})$
$\Phi^{A\rightarrow B}$
$G(x^{A}) \circ \Phi^{A\rightarrow B}$
$F^{A\rightarrow B}$
$\hat{x}_{T}^{B}$
$G(x^{B})$
$\Phi^{B\rightarrow A}$
$G(x^{B}) \circ \Phi^{B\rightarrow A}$
$F^{B\rightarrow A}$
$\hat{x}_{T}^{A}$

Fig. 1. Architecture of the FIRE model: a synthesis encoder G, which extracts modality-independent features; two synthesis decoders $F^{A\rightarrow B}$ and $F^{B\rightarrow A}$ which map the features extracted by G to the synthesized images; and two transformation networks $T^{A\rightarrow B}$ and $T^{B\rightarrow A}$, which predict the transformation fields.

Synthesis Loss

The synthesis loss consists of four different elements. First, to develop and maintain accurate cross-domain synthesis, we defined a synthesis accuracy loss by implementing the root-mean-square (RMS) error, $\mathcal{L}syn, acc$ = RMS($\hat{x}_{T}^{B}$, x^{B}) + RMS($\hat{x}_{T}^{A}$, x^{A}). Second, G is expected to extract modality-independent features so that features extracted from the aligned image pairs should be identical regardless of their modalities. To represent this, we defined another feature loss, $\mathcal{L}syn, fea$ = RMS($G(x^{A})$+ $G(x^{B}) \circ \varphi^{B\rightarrow A}$) + RMS($G(x^{B})$+ $G(x^{A}) \circ \varphi^{A\rightarrow B}$). The third cycle-consistency loss was defined as $\mathcal{L}syn, cyc$ = RMS($F^{B\rightarrow A}$ ($G(\hat{x}^{B})$), x^{A}), + RMS($F^{A\rightarrow B}$ ($G(\hat{x}^{A})$), x^{B}) for robust cross-modality synthesis. Finally, to align $x^{\cdot}$ *and* $\hat{x}$ $^{\cdot}$ we set an alignment loss $\mathcal{L}syn, align$ = RMS($G(x^{A})$, G ($\hat{x}^{B}$)) + RMS($G(x^{B})$, $G(\hat{x}^{A})$).

To sum up, the entire FIRE synthesis loss is:

$$\mathcal{L}syn = \mathcal{L}syn, acc + \mathcal{L}syn, fea + \mathcal{L}syn, cyc + \mathcal{L}syn, align \qquad (2)$$

Registration Loss

To perform synthesis, features extracted by G were transformed and registration was achieved by applying the following transformations to the input images $\varphi^{\bullet\rightarrow\bullet} = \varphi_{af}^{\bullet\rightarrow \text{bigcdot}} \circ \varphi_{nr}^{\bullet\rightarrow\bullet}$. Here, a registration accuracy loss was defined: $\mathcal{L}reg, acc$ = RMS($F^{A\rightarrow B}$ ($G(x^{A}) \circ \varphi^{A\rightarrow B}$)), x^{B}) + RMS($F^{B\rightarrow A}$ ($G(x^{B}) \circ \varphi^{B\rightarrow A}$))). For mutually inversed transformations $\varphi^{A\rightarrow B}$ and $\varphi^{B\rightarrow A}$, we set an inverse consistency loss $\mathcal{L}reg, ic$ = RMS(x^{A}, $x^{A} \circ \varphi^{A\rightarrow B} \circ \varphi^{B\rightarrow A}$) + RMS($x^{B}$, $x^{B} \circ \varphi^{B\rightarrow A} \circ \varphi^{A\rightarrow B}$). The entire registration loss was computed as:

$$\mathcal{L}reg = \mathcal{L}reg, acc + \mathcal{L}reg, ic \qquad (3)$$

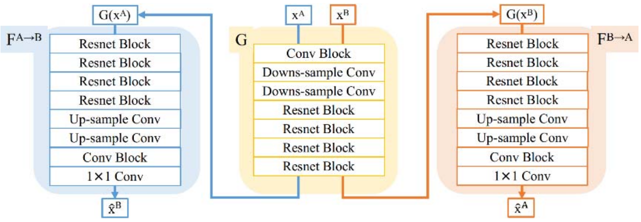

Fig. 2. Architecture of the encoder/ decoders used for synthesis.

Regularisation

Previous studies regularize the non-rigid transformation fields by introducing smoothness regularization $\mathcal{R}smooth = \| \nabla^2 \varphi_{nr}{}^{A\to B} \|^2 + \| \nabla^2 \varphi_{nr}{}^{B\to A} \|^2$, where ∇ is the Laplacian operator. In this work, the estimated affine transformations designed to keep the non-rigid transformations to the minimum. In the synthesis process, the affine transformed features $G(x^A) \circ \varphi_{af}{}^{A\to B}$ and $G(x^B) \circ \varphi_{af}{}^{B\to A}$, can be used as inputs into the synthesis decoders to obtain $F^{A\to B}(G(x^A) \circ \varphi_{af}{}^{A\to B})$ and $F^{B\to A}(G(x^B) \circ \varphi_{af}{}^{B\to A})$. The regularization of the synthesis is then computed as $\mathcal{R}syn = RMS(x^B, F^{A\to B}(G(x^A) \circ \varphi_{af}{}^{A\to B})) + RMS(x^A, F^{B\to A}(G(x^B) \circ \varphi_{af}{}^{B\to A}))$. Similarly, a regularization of registration, $\mathcal{R}reg$, is computed as: $\mathcal{R}reg = RMS(x^B, F^{A\to B}(G(x^A \circ \varphi_{af}{}^{A\to B}))) + RMS(x^A, F^{B\to A}(G(x^B \circ \varphi_{af}{}^{B\to A})))$.

To summarise, the regularization of the FIRE model is:

$$\mathcal{R} = \mathcal{R}syn + \mathcal{R}reg + \lambda \mathcal{R}smooth \qquad (4)$$

where λ is a scaling parameter for $\mathcal{R}smooth$. Empirically, when registering n-D images, $\lambda = 2^{2n}/10N$, where N represents number of points in an input image.

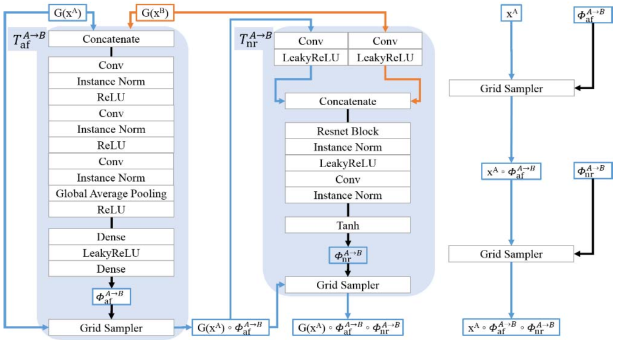

Fig. 3. Architecture of the transformation networks.

Optimization

We used three Adam optimizers to update the parameters $T_{af}{}^{\bullet\to\bullet}$, $T_{nr}{}^{\bullet\to\bullet}$ as well as all the other networks separately, across each of the three consecutive iterations. Learning rates for training $T_{af}{}^{\bullet\to\bullet}$ and $T_{nr}{}^{\bullet\to\bullet}$ as well as G and $F^{\bullet\to\bullet}$ were set to 5×10^{-5} and 10^{-4}, respectively.

3. Experiments

MRBrainS We use a dataset consisted of 3T multi-sequence brain MR data, by fusing the training data from the MRBrains18 (https://mrbrains18.isi.uu.nl/) and the MRBrains13 (http://mrbrains13.isi.uu.nl/) Challenges.

The dataset contains co-registered 3D T1-weighted, inversion recovery (IR) and T2-FLAIR data acquired from 12 subjects. Each T1, IR and T2-FLAIR patient data set contained 192, 192 and 48 slices, respectively. All scans had a voxel size of 0.958× 0.958 × 3.0mm3. We used manual segmentations of 3 anatomical structures to evaluate the performance of the registration algorithms. Multi-slice MR data from 8, 1 and 3 patients were used for training, validation and testing, respectively. For both 3D and 2D registration, we resampled all data to 1.28mm^3 per voxel. We performed 2D and 3D registration between T1 and FLAIR data, and 2D registration between IR and FLAIR data. In the training stage, randomly generated affine and non-rigid transformations were applied to the moving image.

ACDC For intra-modality registration, we used 4D cine-MR data from the 2017 ACDC (https://www.creatis.insa-lyon.fr/Challenge/acdc) Challenge. The training dataset includes data from 100 patients with a variety of pathologies. The in-plane resolution is between 1.37 and 1.68mm^2/pixel, and each 4D image has 28 to 40 phases that cover completely or partially the cardiac cycle. Manual segmentation of 2 phases are provided for each of the 4D patient data. We used all phases for training and the two segmented phases for testing. We used 40, 10 and 50 patients for training, validation and testing, respectively.

Evaluation Metrics and Baselines

We evaluated our method using the overlap of the segmented objects, as measured by the Dice metric. Higher Dice scores indicate better registration performances. Previous comparison studies show that Symmetric Normalization (SyN) [11] implemented using the ANTs toolbox (http://stnava.github.io/ANTs/) has outstanding non-rigid registration performances. We therefore compare our FIRE model against SyN [11], to assess non-rigid registrations. Affine registration performances were compared against the mutual information (MI), implemented in the ANTs toolbox.

3.1 Results and Discussion

Table 1 summarizes the Dice scores obtained from the registration process of the T1 and FLAIR data from the MRBrainS data. Fig. 4 demonstrates further representative results, for the previous registration process. Our FIRE model consistently achieved higher scores versus the SyN method, across all the brain structures examined: the segmented cerebellum (Ce), the brain stem (BS), and the white matter (WHM) (Table 1).

The above results were consistent in the 2D and 3D registration experiments. Fig. 4 presents a visual assessment of the T1-FLAIR registration, in which FIRE achieves improved alignment between the outer contours of the cerebrospinal fluid in the extracerebral space (shown in blue), against the Syn method.

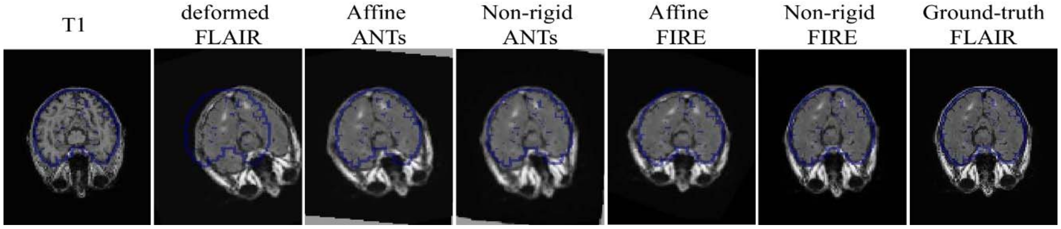

Fig. 4. Representative results from the MRBrainS T1 and FLAIR data. The outer contour outlines the cerebrospinal fluid in the extracerebral space, and it was segmented on T1 anatomical images (shown in blue).

Table 1. Results obtained from the 2D and 3D T1-FLAIR registration, on the MRBrainS data. Dice scores are calculated on cerebellum (Ce), white matter (WHM), brain stem (BS). Standard deviations are included within the parenthesis.

Data	Object	unaligned	ANTs-affine	FIRE-affine	ANTs-SyN	FIRE
2D	BS	11.62 (6.1)	61.25 (3.7)	62.90 (4.1)	78.73 (7.3)	**80.68 (7.7)**
	CE	7.17 (4.4)	63.32 (3.2)	64.36 (4.0)	75.72 (8.1)	**76.96 (7.3)**
	WHM	14.29 (7.5)	59.12 (4.5)	59.97 (4.4)	81.36 (6.0)	**84.18 (3.7)**
3D	BS	27.15 (9.2)	67.15 (3.1)	69.81 (4.1)	79.77 (6.7)	**81.08 (7.0)**
	CE	28.38 (10.3)	68.38 (3.6)	70.62 (3.7)	86.00 (6.9)	**86.13 (7.2)**
	WHM	20.27 (9.3)	60.27 (3.8)	60.61 (3.8)	72.33 (7.4)	**72.56 (7.1)**

However, the registration process of IR and FLAIR images is generally challenging. Using the SyN method, following a grid search on the ANTs toolbox setup, we failed to produce an accurate (visual) alignment. As an example, using the SyN method, the average dice score obtained on the Ce structures was below 0.40. Despite this, our method outperformed the Syn method and achieved a 0.69 Dice score for IR-FLAIR registration. Representative results are presented in Fig. 5.

Table 2 and Fig. 6 show the results of the intra-modality registration, performed on the left ventricle and myocardial cardiac structures, from the ACDC data. On the left ventricle structures, the data show small local displacements across frames. Hence, our FIRE model and the Syn method showed comparable results and considerably high Dice scores (90.08 and 90.81, respectively). Similar results were obtained for the myocardial structures (Table 2).

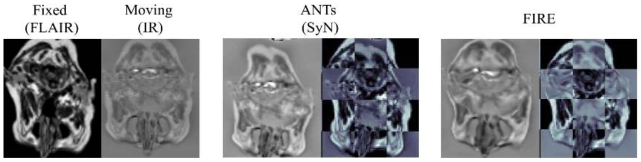

Fig. 5. Example results from the registration of the IR to the FLAIR data.

Table 2. Results on ACDC data. Dice scores computed on left ventricular endocardium (LVe) and myocardium (Myo). Standard deviations are included within the parenthesis.

Object	unaligned	ANTs-SyN	FIRE
LVe	65.75 (16.25)	**90.81 (4.3)**	90.08 (5.5)
Myo	51.97 (14.50)	70.71 (5.6)	**71.66 (6.3)**

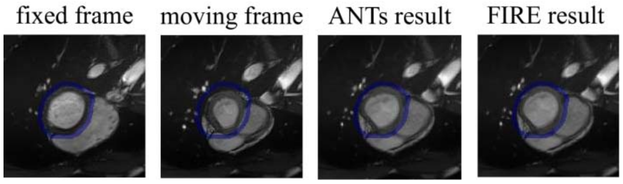

Fig. 6. Representative results from the registration experiments, on the ACDC data. Outer contours of myocardium are shown in blue.

Model interpretability

Inter-modality registration has not yet been widely investigated in deep learning. One of the main reasons is the lack of model interpretability in terms of why the model can or cannot model geometric and semantic topologies, across different types of inter-modality registration. Thus, model interpretability for inter-modality registration can support further deep learning investigations across multiple medical imaging data.

We showed that the FIRE model can depict geometric similarities across modalities, therefore demonstrating improved performance for the registration of multi-sequence MRI data. The inclusion of transformation networks in conjunction with the encoder/decoders used for the synthesis process, played an important role in the overall model performance. A thorough understanding of our model architecture versus the MR data used for data analysis in this work can explain why our FIRE model outperformed the Syn method, in the context of inter-modality registration. This interpretation is also important to understand why the T1-FLAIR registration showed higher performance, versus the FLAIR-IR registration experiments.

Although T1, FLAIR and IR are all standard MR sequences, they have important differences in terms of the MR physics involved and are designed to provide different information in the clinical setting: T1 is primarily used for anatomical assessments, whilst FLAIR and IR are mainly used for functional assessments. Hence, T1 data contain more enriched anatomical information, compared to FLAIR and IR data. This means that when T1 is involved (as a fixed image) in the registration process, our encoder/decoders component could more efficiently predict geometric and semantic similarities between T1 and FLAIR data, by combining anatomical information from the T1 data, and by mapping it to complementary MR information from the FLAIR data (during cross-domain image synthesis). This anatomical and complementary information in the T1-FLAIR registration experiment was not present in the FLAIR-IR data registration, in which the FIRE model did not have enough anatomical information to process from the FLAIR and IR data alone.

Involving a semi-supervised learning algorithm to perform a segmentation task in our FIRE model, could be one approach towards enriching anatomical information in the input data hence, adapting our technique for non-anatomical MR sequences. Despite this, we demonstrate that our FIRE model consistently outperformed the reference standard Syn method and is a promising tool in the clinical setting of inter-modality MR registrations.

4. Conclusions

We propose a deep learning model which solves both large and small displacements in inter- and intra-modality image registration problems respectively, through cross-domain image synthesis and the involvement of spatial transformation networks.

Our FIRE model has a new "火"-shape architecture formed by five sub-networks. Our experiments show that the FIRE model outperformed the reference standard Syn method, for both 3D and 2D registration tasks (MRBrainS data). When local displacements across data were small, our FIRE model showed comparable performance versus the Syn method in 4D registration tasks (ACDC data).

The new spatial transformation network in conjunction with our encode/decoders component and the associated loss functions allow to optimally predict both affine as well as topology preserving non-rigid transformations. Inverse-consistency is addressed through our FIRE architecture, by involving the aforementioned comprehensive spatial transformation network. We interpret our FIRE model in the context of multi-sequence MR data processing, therefore demonstrating that it is a promising tool for clinical MR applications. Our future work involves further developments and investigations of our FIRE model using multi-sequence and multi-organ MR data.

References

1. Rueckert, D., Schnabel, J.A.: Medical image registration. In: Biomedical image processing. Springer (2010) 131–154
2. Haskins, G., Kruger, U., Yan, P.: Deep learning in medical image registration: A survey. arXiv preprint arXiv:1903.02026 (2019)
3. Cao, X., Yang, J., Zhang, J., Nie, D., Kim, M., Wang, Q., Shen, D.: Deformable image registration based on similarity-steered cnn regression. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2017) 300–308
4. Krebs, J., Mansi, T., Delingette, H., Zhang, L., Ghesu, F.C., Miao, S., Maier, A.K., Ayache, N., Liao, R., Kamen, A.: Robust non-rigid registration through agentbased action learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2017) 344–352
5. Roh´e, M.M., Datar, M., Heimann, T., Sermesant, M., Pennec, X.: Svf-net: learning deformable image registration using shape matching. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2017) 266–274
6. Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. In: Advances in neural information processing systems. (2015) 2017–2025
7. de Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., I˜sgum, I.: A deep learning framework for unsupervised affine and deformable image registration. Medical image analysis 52 (2019) 128–143
8. Christensen, G.E., Johnson, H.J.: Consistent image registration. IEEE transactions on medical imaging 20(7) (2001) 568–582
9. Qin, C., Shi, B., Liao, R., Mansi, T., Rueckert, D., Kamen, A.: Unsupervised deformable registration for multi-modal images via disentangled representations. arXiv preprint arXiv:1903.09331 (2019)
10. Zhang, J.: Inverse-consistent deep networks for unsupervised deformable image registration. arXiv preprint arXiv:1809.03443 (2018)
11. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical image analysis 12(1) (2008) 26–41 disentangled representations. arXiv preprint arXiv:1903.09331 (2019)