2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Multi-modality Large Deformation Diffeomorphic Metric Mapping Driven by Single-modality Images

Jiong Wu[1,*], Shuang Zhou[2], Qi Yang[3], Yue Zhang[4,5] and Xiaoying Tang[4,*]

Abstract— **Multi-modality magnetic resonance image (MRI) registration is an essential step in various MRI analysis tasks. However, it is challenging to have all required modalities in clinical practice, and thus the application of multi-modality registration is limited. This paper tackles such problem by proposing a novel unsupervised deep learning based multi-modality large deformation diffeomorphic metric mapping (LDDMM) framework which is capable of performing multi-modality registration only using single-modality MRIs. Specifically, an unsupervised image-to-image translation model is trained and used to synthesize the missing modality MRIs from the available ones. Multi-modality LDDMM is then performed in a multi-channel manner. Experimental results obtained on one publicly-accessible datasets confirm the superior performance of the proposed approach.**

Clinical relevance—**This work provides a tool for multi-modality MRI registration with solely single-modality images, which addresses the very common issue of missing modalities in clinical practice.**

I. INTRODUCTION

Multi-modality magnetic resonance image (MRI) registration plays an important role in a variety of tasks such as atlas alignment [1], image fusion [2] and distortion correction [3]. In addition, utilizing multi-modal registration to incorporate information from MRIs of different modalities can improve the performance of various subsequent MRI analysis tasks such as brain segmentation and surgical planning [4]. However, MRIs of multiple modalities are kind of rare in clinical practice, and thus the application of multi-modal registration has been limited.

One way to deal with such problem is to discard the missing modality MRIs and directly utilize the available ones for registration. This kind of approaches can be mainly divided into three categories, including information theory based approaches, modality reduction approaches and feature-based approaches. Information theory based approaches generally use information theory measures to evaluate the misalignment between images. The most popular measures are mutual information (MI) and normalized mutual information (NMI) [5]. However, unlike intra-modality similarity measures such as the sum of squared difference (SSD), MI and NMI cannot directly nor efficiently quantify local anatomical similarity [6]. Modality reduction approaches convert multiple modalities to a completely new one [7] or one of the existing modalities [8] before registration. Although this conversion simplifies the alignment process, losing anatomical information may reduce the registration accuracy. Different from the above two kinds of methods, feature-based approaches extract features such as morphological features [9] and sparse keypoint features [10] from multi-modal MRIs for registration. Extracting modality-independent features is time-consuming and laborious, therefore its application is limited.

To avoid the aforementioned problems, image synthesis based multi-modality registration approaches have been proposed. In these methods, proxy MRIs of the missing modalities are firstly generated using synthesis approaches. Then multi-modality registration is performed on the generated proxy MRIs and existing single-modality ones via multi-channel registration. For instance, Roy et al. [11] proposed a MR-CT registration approach using intensity patches within an expectation maximization framework to synthesize CT images from T1-weighted (T1w) MRIs. Chen et al. [6] proposed a multi-modality registration algorithm using a trained regression forest to create proxy images.

With the advent of generative adversarial network (GAN) [12], using it to synthesize images becomes a hot research topic. It has already been widely used in medical image synthesis and started to be applied to multi-modal registration. For instance, Tang et al. [4] proposed a multi-modality registration framework using CycleGAN to synthesize multi-modality atlases from T1w images. Qin et al. [13] proposed an unsupervised deformable registration algorithm for multi-modality atlases using latent shape representation. In this paper, we propose a new diffeomorphic framework for multi-modality MRI registration using unsupervised image-to-image translation. A deep learning based multi-modality unsupervised image-to-image translation synthesizer (MUTS) is introduced and combined with large deformation diffeomorphic metric mapping (LDDMM) [14] to convert T1w-only registration into T1w and T2-weighted (T2w) combined multi-modality registration. This proposed framework can be easily extended to other modalities.

*This work was supported by the Scientific Research Project of Hunan University of Arts and Science (20ZD01), the National Natural Science Foundation of China (62071210), the Shenzhen Basic Research Program (JCYJ20190809120205578), the National Key R&D Program of China (2017YFC0112404), and the High-level University Fund (G02236002).

*Correspondence at wujiong@huas.edu.cn and tangxy@sustech.edu.cn

[1] School of Computer and Electrical Engineering, Hunan University of Arts and Science, Hunan, China.

[2] Furong College, Hunan University of Arts and Science, Hunan, China.

[3] School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, China

[4] Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China

[5] Department of Electrical and Electronic Engineering, The University of Hong Kong, Hongkong, China

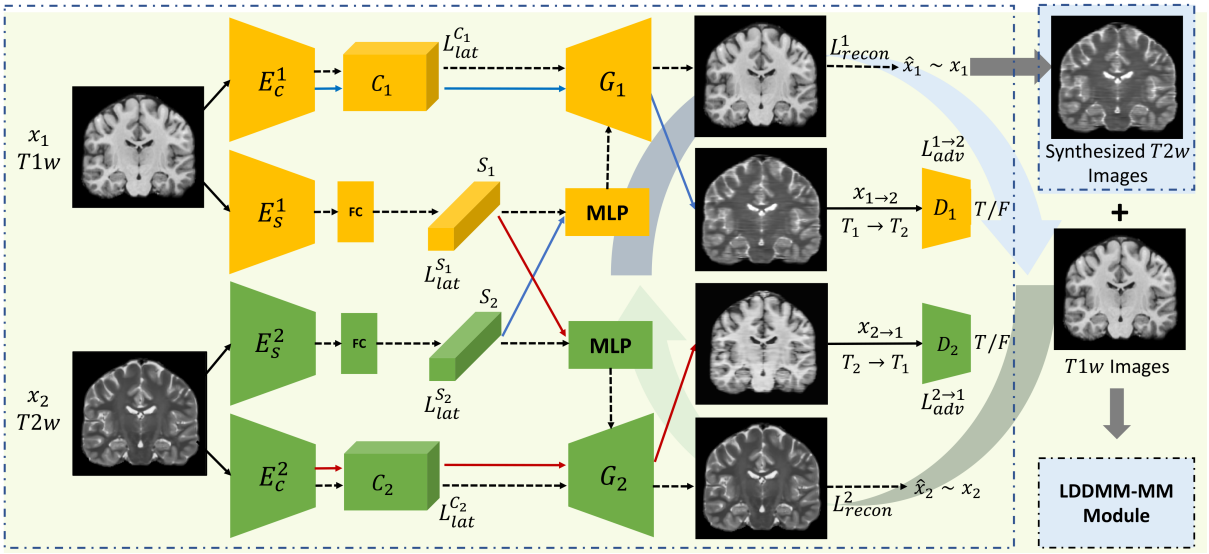

Fig. 1. The proposed framework includes a MUTS module and a multi-modality LDDMM registration (LDDMM-MM) module. $\hat{x}_1$ and $\hat{x}_2$ are translated and reconstructed images. For two to-be-aligned T1w MRIs, the trained MUTS module is firstly adopted to synthesize the corresponding T2w MRIs. Next, the LDDMM-MM module is adopted to perform registration using these four images in a multi-channel manner. Please note, when training the MUTS module, LDDMM-MM is also performed after each epoch using the original T1w images and the synthesized T2w images to find the optimal MUTS for LDDMM-MM.

II. Method

The proposed unsupervised image-to-image translation based multi-modality LDDMM registration framework, as shown in Fig. 1, consists of two modules including a MUTS module and a multi-modality LDDMM (LDDMM-MM) module. The MUTS module is used as a synthesizer to generate T2w MRIs from T1w ones. To obtain the synthesizer, a training dataset including both T1w and T2w MRIs is used. After inputting two to-be-aligned T1w MRIs, the LDDMM-MM module is adopted to perform registration by assigning one channel with the original T1w images and the other channel the corresponding synthesized T2w images. Different from the aforementioned synthesizer based multi-atlas registration methods, we perform LDDMM-MM registration after each epoch in training MUTS to obtain an optimal synthesizer for LDDMM-MM.

A. MUTS

The MUTS module was originally proposed by Huang et al. [15] based on the assumption that unpaired images of different modalities can be embedded into a domain-invariant attribute (content) space and a domain-specific attribute (style) space. For both modalities, as shown in Fig 1, two encoders for encoding the attributes of these two spaces are used, followed by a decoder to generate the corresponding proxy images.

Let $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$ denote unpaired images from two different imaging modalities. As shown in Fig. 1, image x_i $(i = 1, 2)$ is disentangled into content code C_i and style code S_i, where E_c^i and E_s^i respectively encode x_i to C_i and S_i. The generator G_i generates images conditioned on both content and style vectors. Image-to-image translation is performed by swapping the style vectors across modalities. For instance, the generator G_1 acts on C_1 and S_2 so that x_1 is translated to the target modality of x_2. To train the image-to-image translation framework, the overall loss function is defined as a weighted sum of three components including the in-domain reconstruction loss $\mathcal{L}_{rec}$, cross-domain translation loss $\mathcal{L}_{adv}$ and latent space reconstruction loss $\mathcal{L}_{lat}$, i.e.,

$$\mathcal{L}_{total} = \alpha\mathcal{L}_{rec} + \beta\mathcal{L}_{adv} + \gamma\mathcal{L}_{lat}, \tag{1}$$

where $\mathcal{L}_{rec} = \mathcal{L}_{rec}^1 + \mathcal{L}_{rec}^2$, $\mathcal{L}_{adv} = \mathcal{L}_{adv}^{1\to2} + \mathcal{L}_{adv}^{2\to1}$ and $\mathcal{L}_{lat} = \mathcal{L}_{lat}^{c_1} + \mathcal{L}_{lat}^{s_1} + \mathcal{L}_{lat}^{c_2} + \mathcal{L}_{lat}^{s_2}$. $\mathcal{L}_{rec}^i$ is calculated as

$$\mathcal{L}_{rec}^i = \mathbb{E}_{x_i\sim\mathcal{X}_i} \left\| G_i\left(E_{c_i}\left(x_i\right), E_{s_i}\left(x_i\right)\right) - x_i \right\|_1 \tag{2}$$

to evaluate the dissimilarity between the synthesized proxy image and original image, $\mathcal{L}_{adv}^{1\to2}$ is calculated as

$$\begin{aligned}\mathcal{L}_{adv}^{1\to2} = \mathbb{E}_{c_1\sim p(c_1), s_2\sim p(s_2)}\left[log\left(1 - D_2(x_{1\to2})\right)\right] \\ + \mathbb{E}_{x_2\sim\mathcal{X}_2}\left[log\left(D_2(x_2)\right)\right]\end{aligned} \tag{3}$$

to match the distribution of the translated image of x_1 to the image distribution in the domain of $\mathcal{X}_2$, The $\mathcal{L}_{lat}^{c_1}$ and $\mathcal{L}_{lat}^{s_2}$ are respectively defined as

$$\mathcal{L}_{lat}^{c_1} = \left\| E_c^2\left(G_2(c_1, s_2)\right) - c_1 \right\|_1 \tag{4}$$

and

$$\mathcal{L}_{lat}^{s_2} = \left\| E_s^2\left(G_2(c_1, s_2)\right) - s_2 \right\|_1. \tag{5}$$

B. LDDMM-MM

After employing MUTS to synthesize the T2w images, LDDMM-MM is performed. Given two real-valued functions $I_{T_i}^0$ and $I_{T_i}^1$ $(i = 1, 2)$ defined on the background space $\Omega \in R^3$, they respectively represent a 3D grayscale moving image and a 3D grayscale target image of the i-th modality. LDDMM-MM tries to find a diffeomorphism $\varphi : \Omega \to \Omega$ such that $I_{T_i}^0 \circ \varphi^{-1}$ is well aligned to $I_{T_i}^1$. The diffeomorphism

$\varphi = \phi_1$ is calculated by integrating time-varying velocity fields $v_t : \Omega \times t \rightarrow R^3$ from $t = 0$ to $t = 1$ with the following ordinary differential equation

$$\varphi = id + \int_0^1 v_t(\phi_t)dt, \tag{6}$$

where $id : \Omega \rightarrow \Omega$ is the identity mapping such that $id(x) = x, x \in \Omega$ and $\phi_t = id + \int_0^t v_\tau(\phi_\tau)d\tau$, $t \in [0, 1]$.

Let $J^t_{T_i} = I^0_{T_i} \circ \phi_{t0}$ be the deformed template image at time t, where $\phi_{st} = \phi_t \cdot \phi_s^{-1}$, $s \in [0, 1]$ represents a diffeomorphic coordinate transformation from time s to time t, LDDMM-MM finds the optimal time-varying velocity vector fields by minimizing the following energy function,

$$E(v_t) = \frac{1}{2}\int_0^1 \|Lv_t\|^2_{L^2}dt + \frac{1}{2}\sum_{i=1}^{2}\frac{1}{\sigma_i^2}M(J^1_{T_i}, I^1_{T_i}), \tag{7}$$

where $M(J^1_{T_i}, I^1_{T_i})$ denotes the matching cost function used to evaluate the misalignment between the target image $I^1_{T_i}$ and the deformed template image $J^1_{T_i}$, $\|\cdot\|_{L^2}$ denotes the L^2 norm of square-integrable function, L denotes a differential operator smoothing the velocity vector fields, and σ_i determines the weight of the matching cost function relative to the regularization term of the i-th modality.

According to the work of [14], the derivative of $E(v_t)$ can be calculated as

$$\nabla_v E_t = v_t + K\left(\sum_{i=1}^{2}\rho(t)\nabla J^t_{T_i}\right), \tag{8}$$

$$\rho(t) = -\frac{1}{2\sigma^2}(\partial_{I_{T_i}} M(J^1_{T_i}, I^1_{T_i}) \circ \phi_{t1}|D\phi_{t1}|), \tag{9}$$

where $K = (L^\dagger L)^{-1}$, $L^\dagger$ denotes the adjoint of L, $\nabla J^t_{T_i}$ denotes the gradient of $J^t_{T_i}$, $\partial_{I_{T_i}} M(J^1_{T_i}, I^1_{T_i})$ denotes the Gateaux derivative of $M(J^1_{T_i}, I^1_{T_i})$ and $|D\phi_{t1}|$ denotes the Jacobian determinant of ϕ_{t1}.

C. Implementation Detail

The MUTS module is trained using the default setting presented in [15]. In our implementation, we pre-train MUTS using paired T1w and T2w images, and the LDDMM-MM registration is performed after each training epoch. In the loss function, $\alpha = 25$, $\beta = 10$ and $\gamma = 0.1$. The training process is terminated when the mean value of the Dice similarity coefficients between the deformed segmentations of the moving images and the corresponding manual segmentations of the target images on the validation dataset reaches the maximum. For LDDMM-MM, a three-stage cascading strategy is adopted. We set the number of time-varying velocity vector fields to be 2, the matching cost function to be SSD, and the weights of the SSD to be $\sigma_1 = 0.01$ and $\sigma_2 = 0.001$ for T1w and T2w. All other parameters are the same as those in [14] and [16].

III. Experimental Results and Evaluation

A. Dataset and Evaluation Metric

We evaluate the proposed method using two datasets. The first dataset consists of 16 subjects[1] and for each one both T1w and T2w 3D-volume MPRAGE images were collected (image size: $190 \times 230 \times 180$ mm^3). For each subject, a total of 14 brain regions have been manually delineated. The second dataset is publicly available, known as LPBA40[2], consisting of 40 T1w brain images (image size: $181 \times 217 \times 181$ mm^3). This dataset was created by the Laboratory of Neuro Imaging at the University of Southern California. For each of the 40 images, 56 brain regions were manually labeled.

For LPBA40, we followed [17] to merge 56 labels into 7 larger ones. These two datasets were affinely aligned (12 parameters) to the MNI152 space [18] and center-cropped or center-padded into a size of $192 \times 224 \times 192$ followed by intensity normalization. We used the first dataset to train the MUTS module and performed LDDMM-MM using a leave-one-out scheme on 10 images selected from the second dataset. The Dice similarity coefficient (DSC) between the manual segmentation of a testing target image and the deformed manual segmentation of a moving image was adopted to quantify the registration accuracy. Student's t-tests were performed to quantify the significance of all group comparison differences.

B. Results and Discussion

To demonstrate the performance of our proposed framework, we compared its registration results with the conventional SSD based single-modality LDDMM registration accuracy [14]. The comparison experiments were performed on the original T1w images, namely LDDMM (T1w), as well as the synthesized T2w images, namely LDDMM (T2w). All group comparison results are listed in Tabel I. Evidently, the MUTS based LDDMM-MM is superior to the conventional SSD based single-modality LDDMM for either modality (superior to T1w for five structures and superior to T2w for four structures).

Specifically, for the frontal, parietal, temporal, cingulate and hippocampus, the DSCs of MUTS based LDDMM-MM approach are significantly higher than those of LDDMM (T1w) with p-values of 1.20×10^{-24}, 1.78×10^{-16}, 4.20×10^{-18}, 2.40×10^{-17} and 5.98×10^{-15}. Also, for the parietal, temporal, putamen and hippocampus, the DSCs of MUTS based LDDMM-MM approach are significantly higher than those of LDDMM (T2w) with p-values of 4.24×10^{-4}, 7.38×10^{-3}, 6.24×10^{-20} and 4.89×10^{-17}. In addition, the mean DSC computed across all seven structures of MUTS based LDDMM-MM is significantly higher than those of conventional SSD based single-modality LDDMM on both T1w and T2w with p-values of 1.43×10^{-24} and 3.08×10^{-12}. Comparing the results of LDDMM (T1w) with those of LDDMM (T2w), the mean DSCs of all seven structures and the DSCs of four structures including the frontal, parietal, occipital and cingulate of LDDMM (T2w) are higher than those of LDDMM (T1w), suggesting the effectiveness of MUTS in synthesizing T2w images in our framework.

[1] https://www.predict-hd.net/

[2] http://www.loni.usc.edu/atlases/

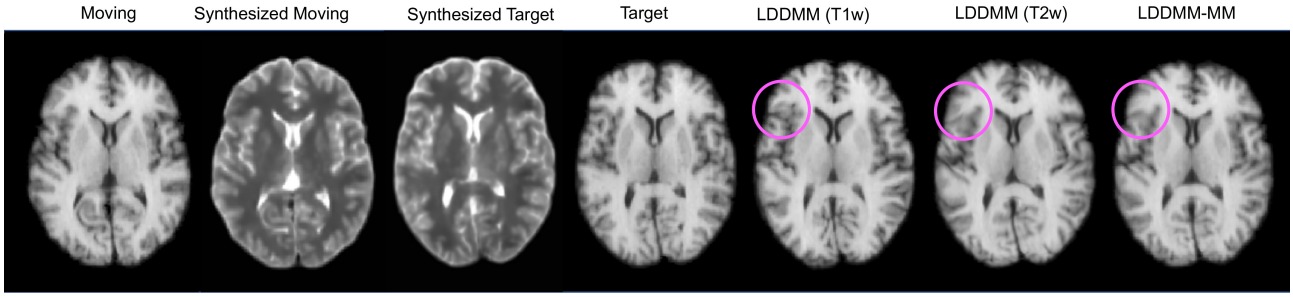

Fig. 2. Visualization results of the proposed LDDMM-MM framework against baseline methods.

TABLE I

THE MEAN DSC FOR EACH OF THE SEVEN STRUCTURES AS WELL AS THE MEAN DSC COMPUTED ACROSS ALL SEVEN STRUCTURES, OBTAINED FROM LDDMM (T1W), LDDMM (T2W) AND MUTS BASED LDDMM-MM. BOLD FONT INDICATES STATISTICALLY SIGNIFICANT GROUP DIFFERENCE.

	LDDMM (T1w)	LDDMM (T2w)	LDDMM-MM
Frontal	0.904 (0.008)	0.910 (0.007)	0.909 (0.008)
Parietal	0.751 (0.024)	0.756 (0.024)	**0.761 (0.025)**
Occipital	0.768 (0.024)	**0.775 (0.024)**	0.768 (0.024)
Temporal	0.855 (0.011)	0.857 (0.012)	**0.859 (0.011)**
Cingulate	0.715 (0.030)	0.721 (0.032)	0.722 (0.031)
Putamen	0.760 (0.027)	0.747 (0.034)	**0.761 (0.031)**
Hippocampus	0.750 (0.028)	0.741 (0.033)	**0.757 (0.027)**
Mean	0.786 (0.011)	0.787 (0.011)	**0.791 (0.012)**

Visual comparisons of the registration results obtained from the three registration methods on one representative image, as well as the moving images, the target images and the corresponding synthesized images are demonstrated in Fig. 2. Clearly, the registration result of the proposed MUTS based LDDMM-MM is the closest to the target image. In addition, from Table I, although we can clearly see that the DSC values of LDDMM (T2w) for the frontal and occipital are the highest, especially for the occipital, with p-values of 2.63×10^{-6} and 1.62×10^{-9} compared to LDDMM (T1w) and LDDMM-MM, the overall registration accuracy of LDDMM-MM is better than the other two methods. A potential reason is that by applying MUTS to synthesize the T2w images, additional anatomical information has been introduced into the LDDMM-MM framework and thus resulted in higher registration accuracy. For future work, we will further validate the efficiency and robustness of the proposed framework and test it on more clinical datasets. Its application to MRIs of other types of modalities is also one of our future endeavors.

IV. CONCLUSION

In this paper, we proposed a novel unsupervised multi-modal image-to-image translation based LDDMM framework for registering brain MRIs. Experimental results demonstrated the superiority of our proposed framework over other conventional approaches in terms of registration accuracy.

REFERENCES

[1] A. W. Toga, P. M. Thompson, *et al.*, "Towards multimodal atlases of the human brain," *Nature Reviews Neuroscience*, vol. 7, no. 12, pp. 952–966, 2006.

[2] T. Nishioka, T. Shiga, *et al.*, "Image fusion between 18fdg-pet and mri/ct for radiotherapy planning of oropharyngeal and nasopharyngeal carcinomas," *International Journal of Radiation Oncology Biology Physics*, vol. 53, no. 4, pp. 1051–1057, 2002.

[3] J. Kybic, P. Thévenaz, *et al.*, "Unwarping of unidirectionally distorted epi images," *IEEE transactions on medical imaging*, vol. 19, no. 2, pp. 80–93, 2000.

[4] Z. Tang, P.-T. Yap, and D. Shen, "A new multi-atlas registration framework for multimodal pathological images using conventional monomodal normal atlases," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2293–2304, 2018.

[5] N. D. Cahill, J. A. Schnabel, *et al.*, "Revisiting overlap invariance in medical image alignment," in *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 2008, pp. 1–8.

[6] M. Chen, A. Carass, *et al.*, "Cross contrast multi-channel image registration using image synthesis for mr brain images," *Medical image analysis*, vol. 36, pp. 2–14, 2017.

[7] C. Wachinger and N. Navab, "Entropy and laplacian images: Structural representations for multi-modal registration," *Medical image analysis*, vol. 16, no. 1, pp. 1–17, 2012.

[8] J. A. Bogovic, P. Hanslovsky, *et al.*, "Robust registration of calcium images by learned contrast synthesis," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2016, pp. 1123–1126.

[9] M. Droske and M. Rumpf, "A variational approach to nonrigid morphological image registration," *SIAM Journal on Applied Mathematics*, vol. 64, no. 2, pp. 668–687, 2004.

[10] J. Chen and J. Tian, "Real-time multi-modal rigid registration based on a novel symmetric-sift descriptor," *Progress in Natural Science*, vol. 19, no. 5, pp. 643–651, 2009.

[11] S. Roy, A. Carass, *et al.*, "Mr to ct registration of brains using image synthesis," in *Medical Imaging 2014: Image Processing*, vol. 9034. International Society for Optics and Photonics, 2014, p. 903419.

[12] A. Creswell, T. White, *et al.*, "Generative adversarial networks: An overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, 2018.

[13] C. Qin, B. Shi, *et al.*, "Unsupervised deformable registration for multi-modal images via disentangled representations," in *International Conference on Information Processing in Medical Imaging*. Springer, 2019, pp. 249–261.

[14] J. Wu and X. Tang, "A large deformation diffeomorphic framework for fast brain image registration via parallel computing and optimization," *Neuroinformatics*, pp. 1–16, 2019.

[15] X. Huang, M.-Y. Liu, *et al.*, "Multimodal unsupervised image-to-image translation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 172–189.

[16] C. Ceritoglu, X. Tang, *et al.*, "Computational analysis of lddmm for brain mapping," *Frontiers in neuroscience*, vol. 7, 2013.

[17] D. Kuang and T. Schmah, "Faim–a convnet method for unsupervised 3d medical image registration," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2019, pp. 646–654.

[18] M. Jenkinson, C. F. Beckmann, *et al.*, "Fsl," *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012.