MRI Imputation based on Fused Index- and Intensity-Registration

Jiyoon Shin[1]   Jungwoo Lee[1,2]

[1]Seoul National University   [2]HodooAI Lab

Abstract

3D MRI imaging is based on a number of imaging sequences such as T1, T2, T1ce, and Flair, and each of them is performed by a group of two-dimensional scans. In practical MRI, some scans are often missing while many medical applications require a full set of scans. An MRI imputation method is presented, which synthesizes such missing scans. Key components in this method are the index registration and the intensity registration. The index registration models anatomical differences between two different scans in the same imaging sequence, and the intensity registration reflects the image contrast differences between two different scans of the same index. Two registration fields are learned to be invariant, and accordingly, allow two estimates of a missing scan, one within corresponding imaging sequence and another along scan index; the two estimates are combined to yield the final synthesized scan. Experimental results highlight that the proposed method improves prevalent limitations existing in previous synthesis methods, blending both structural and contrast aspects and capturing subtle parts of the brain. Quantitative results also show the superiority in various data sets, transitions, and measures.

1. Introduction

Among various diagnoses in neurology and neurosurgery, Magnetic Resonance Imaging (MRI) is one of the most popular due to its safety and information abundance. MRI 3D volumes are generally taken through several imaging sequences (also referred to as pulse sequences or modalities); for example, T1-weighted (T1), T2-weighted (T2), T1 contrast-enhanced (T1ce), and FLuid-Attenuated Inversion Recovery (Flair)[1], which differ by TR (repetition time) and TE (echo time). Each of them is performed and visualized through a group of two-dimensional scans, from one of three anatomical planes: axial, coronal, or sagittal (shown in Fig. 1). Accurate diagnosis and image analysis through

MRI require all T1, T2, and Flair (or even more) sequence scans [14, 25]. Unfortunately, some scans are often missing due to practical limitations such as scanning cost, prolonged scanning time, and motion artifacts.

To be provided as a simple fix, studies for synthesizing a translated modality from a given modality, known as cross-modality translation [11], exist. The process of replacing missing data with substituted ones, generally referred to as imputation [10, 25, 43], contributes to the real-world medical imaging problem. Deformable registration [4], which aims to align similar pairs of images by using geometric correspondences, is also involved in a variety of data augmentation, synthesis, and imputation tasks. However, although an extensive field of work was suggested to generate missing or incomplete medical images, limitations are shown as most methods require closely related or similar data for the process. Also, constraints remain in generating high-quality substitute images, losing either structural or visual parts. Clinically important parts are often small in terms of size and mainly focused on particular regions, which may be difficult to be reconstructed by existing methods. The reconstruction may also include erroneous parts that contain some noisy components. Some methods do show high quantitative measures - however when visualized, images remain insufficient to be used in actual clinical scenarios.

To deal with these limitations and provide a "clinically useful" fill-in, a 3D MRI imputation method based on registration is presented. The motivation lies in the fact that (1) two-dimensional scans within the same imaging sequence show similarities in appearance and contrast while showing different anatomical structures, and (2) different scans with the same scan index (in diverse imaging sequences) show similar anatomies but exhibit differences in appearance and contrast. Two registration fields are introduced: the *index registration field* Φ based on observation (1), and the *intensity registration field* Ψ based on (2). Φ models anatomical differences between two scans in the same imaging sequence, and Ψ reflects the image contrast and appearance differences between two scans of the same index. Both fields are trained as *invariant registration fields*[2] so that the

---

[1]T1-weighted, T2-weighted, T1 contrast-enhanced, and FLuid-Attenuated Inversion Recovery are abbreviated as T1, T2, T1ce, and Flair, respectively, for the rest of the paper.

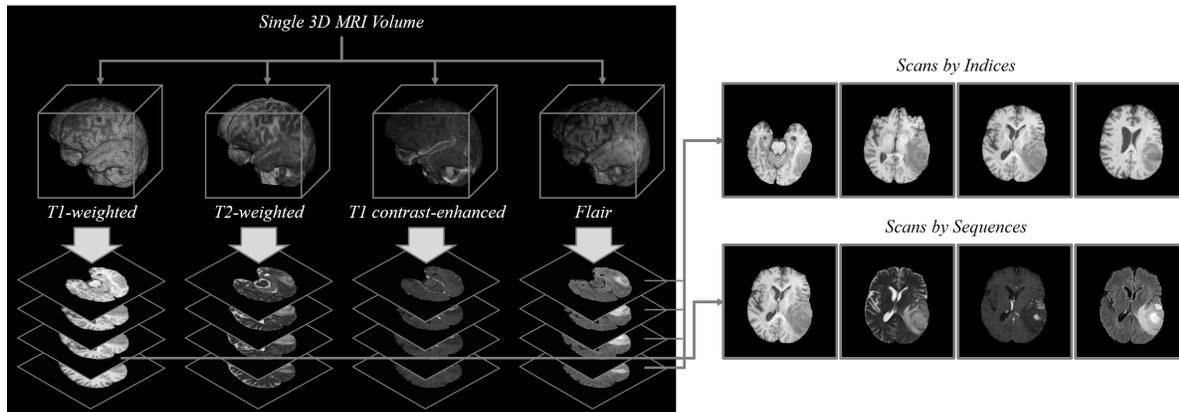[2]A novel term explained in details in Section 3.3 and Fig. 3.

Figure 1. **MRI imaging sequences and two-dimensional scans.** A single 3D MRI volume can be visualized through a variety of imaging sequences. Each sequence volume produces a number of two-dimensional scans in a particular plane (axial plane in figure).

same Φ can be applied in any imaging sequence and the same Ψ can be applied in any scan index. Consequently, the two registration fields allow two estimates of a missing scan, one within corresponding imaging sequence and another along scan index. The two estimates are combined to yield the final synthesized scan, available of preserving both anatomical structure and contrast aspects. An overview of the proposed method is visualized in Fig. 2. Experiments show improvements by our method, which results in synthesized images with higher quality and richer detail, losing neither structural nor visual features. Parts in center of brain which usually contain clinically important data specifically show structural resemblance, while these parts have been kept as a limitation in previous methods. Quantitative measures also show the superiority of the proposed method in various data sets, sequence transitions, and metrics.

The main contributions are summarized as follows.

- Two registration fields for matching anatomical differences and contrast differences are considered. Especially, a new concept of *intensity registration* is introduced for a straightforward sequence translation.
- A novel *invariant registration field* that can be shared along parallel directions of alignment is proposed.
- "Clinically useful" results are shown; synthesized MRI images successfully capture both structural and contrast details. Center parts of brain notably visualize structural resemblance compared to baseline methods.

2. Related Works

2.1. Medical Image Registration

Image registration, also known as image alignment, is a process of aligning two or more anatomically related images based on their spatial appearances. Medical image registration, in particular, has been extensively studied.
Traditional Algorithms : Popular models for traditional

algorithms include elastic [24, 32, 41], b-spline [19, 33, 45], viscous fluid-flow [5, 8], optical flow [6, 7, 30], and diffeomorphism [1, 2, 9, 18, 23, 34, 40].
Deep Learning Image Registration (DLIR) Algorithms : Although not explicitly introduced for registration, the Spatial Transformer Network (STN) [17] is one of the first methods that employ deep learning for image alignment. Simple end-to-end unsupervised methods [12, 37] integrate STN to demonstrate warping of registration fields to images. VoxelMorph [3, 4, 44] uses a structure similar to STN, adopting an encoder-decoder structure for the localization net and producing transformation parameters.

2.2. MRI Cross-Modality Translation

The diversity of imaging sequences enables MRI to generate distinct contrasts while imaging the same anatomy. However, complete multi-modal MR images are not easy to collect due to practical limitations. Synthesizing a translated modality from a given modality without actual acquisition, cross-modality translation, has been studied [38].
Supervised Methods : A number of regression-based synthetic methods [20, 21, 28] have been proposed. Replica [20] employs a supervised random forest that learns a nonlinear regression to predict intensities of alternate contrasts. LSDN [28] integrates image intensity features and spatial information through the location-sensitive deep network.
Unsupervised Methods : Cross-modality translation has also been treated through a general unsupervised approach [35]. For instance, GAN [15] based methods [11, 38] have been proposed to fill in missing contrasts. Two implementations are shown in [11], each of them useful whether the reference images in two contrasts are paired or not.

2.3. Missing Data Imputation

The absence of some data in a particular data set causes substantial bias, making the analysis of data less efficient.
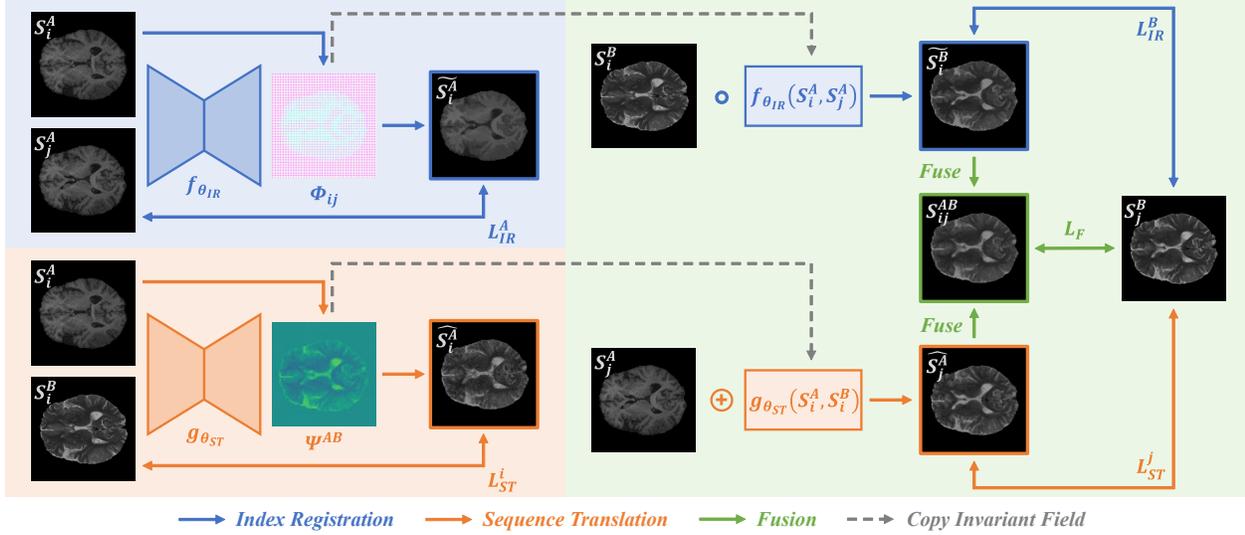
Figure 2. **Overview of proposed method.** Blue, orange, green, and gray dashed arrows represent the procedure of index registration, sequence translation, fusion, and copying invariant registration fields, respectively.

Given as a solution, imputation is the process of replacing missing data with substituted values.

GAN Based Imputation : Imputation models [16, 25, 39, 42] through GAN [15] have been introduced. CollaGAN [25] estimates a single missing data from remaining multiple data, handling the imputation problem through a multi-domain images-to-image translation.

MRI Imputation with Transformers : With the release of vision transformers (ViT) [13], a Cambrian explosion of using transformers in various vision tasks has occurred recently. Particularly in medical tasks [10, 26, 43], transformers are utilized to deal with missing images. A novel aggregated residual transformer (ART) block is introduced in ResViT [10] to preserve localization and context of the missing. PTNet [43] adopts transformer layers along with skip connections and multi-scale pyramid representations for a high-resolution synthesis.

3. Method

A single 3D MRI volume is expressed by a number of imaging sequences such as T1-weighted (T1), T2-weighted (T2), T1 contrast-enhanced (T1ce), and FLuid-Attenuated Inversion Recovery (Flair), which differ by contrasts while showing the same anatomy. As shown in Fig. 1, each sequence volume is visualized by two-dimensional scans in one of axial, coronal, or sagittal planes. Consequently, each scan is associated with an imaging sequence and scan index. Here, "scan index" is defined as the index of a scan when visualized in a particular direction. For example, a volume of resolution $240 \times 240 \times 155$ (by convention, x, y, z refers to sagittal, coronal, and axial axis, respectively) visualized in the axial direction should have a scan index range of $0-154$.

Unfortunately in practical MRI, some scans may be missing while full scans are required for accurate diagnosis, tumor segmentation, and so on. Our approach of imputing such missing scans is briefly illustrated in Fig. 2. Let S_i^A and S_j^A be i-th and j-th indexed scans of imaging sequence A. Similarly, let S_i^B and S_j^B be i-th and j-th indexed scans of imaging sequence B. In imaging sequence B, we assume that S_i^B is available, but S_j^B is missing.

An encoder-decoder structure (similar to Unet) is built - which receives S_i^A and S_j^A at its input and yields an index registration field Φ_{ij} at its output. The warping of S_i^B and Φ_{ij}, denoted by $S_i^B \circ \Phi_{ij}$, yields the first estimate of S_j^B; we call this process index registration. Another encoder-decoder structure is also built, which receives S_i^A and S_i^B and yields an intensity registration field Ψ^{AB}. The second estimate of S_j^B is obtained if S_j^A is concatenated with Ψ^{AB}, or $S_j^A + \Psi^{AB}$; this process shall be called sequence translation. Lastly, the two estimates are fused for the final estimate. Key components in this imputing process are the index registration field Φ and the intensity registration field Ψ. As illustrated in Fig. 3, a single field Φ is learned to be applicable in different imaging sequences, and a single Ψ is learned to be applicable in different scan indices. Two registration fields are called invariant for the salient features.

3.1. Index Registration

Two-dimensional scans within the same imaging sequence share similarities in appearances and contrasts, while showing different anatomical structures (see "Scans by Indices" of Fig. 1). To capture anatomical differences among those scans, a smooth index registration field Φ is adopted. It parameterizes a displacement function u, which

models the change of pixel positions between the moving and fixed images[3]:

$$\Phi(p) = p + u(p). \tag{1}$$

Let $M(p)$ and $\tilde{M}(p)$ correspond to the intensity value of pixel p in moving image M and moved image $\tilde{M}$, respectively. Then,

$$\tilde{M}(p) = M(p') = M(\Phi(p)), \tag{2}$$

where p' is the transferred pixel location. This can also be represented by

$$\tilde{M} = M \circ \Phi \tag{3}$$

where M is said to be "warped" to registration field Φ. To model the distribution of anatomical differences, we compute the registration field that warps moving image M to fixed image F using

$$\Phi = f_{\theta_{IR}}(F, M) \tag{4}$$

where $f_{\theta_{IR}}(\cdot, \cdot)$ is a parametric function with learnable parameters θ_{IR} that we describe in Section 3.4.

3.2. Sequence Translation by Intensity Registration

Two-dimensional scans with the same scan index of different imaging sequences share similar anatomical structures, while showing different appearances and contrasts (see "Scans by Sequences" of Fig. 1). For sequence translation, a new concept of ***intensity registration*** is introduced for nonlinear mapping of contrasts between two scans. Similar to index registration field Φ, intensity registration field Ψ parameterizes function v demonstrating the change of pixel intensity values between moving and fixed images. For an intensity registration of a specific intensity I, one can denote a transferred intensity as:

$$\Psi(I) = I + v(I). \tag{5}$$

Without any movement of pixel location p in the moving image M, a substitution of pixel intensity values with the transferred values gives

$$\hat{M}(p) = \Psi(M(p)) = M(p) + v(M(p)) \tag{6}$$

where $\hat{M}$ refers to the moved image. Nonlinear mappings of contrasts between moving image M and fixed image F are modeled by

$$\Psi = g_{\theta_{ST}}(F, M) \tag{7}$$

where $g_{\theta_{ST}}(,)$ is a parametric function with learnable parameters θ_{ST} described further in Section 3.4.

3.3. Imputation Procedure

Let $\{S_x\}$ be a set of two-dimensional scans with scan index x, and let $\{S^Y\}$ represent a set of scans taken by imaging sequence Y. S_x^Y denotes a single scan with scan index x and imaging sequence Y. Non-negative integers in increasing order, $\{0, 1, 2, 3, ...\}$ is assumed for x; Y is an accessi-

---

[3]In medical image registration, moving image is the image to be matched to the fixed image. Moved image, also called registered image, is the moving image warped to the registration field. Fixed and moved images must be alike with a well-made registration field.
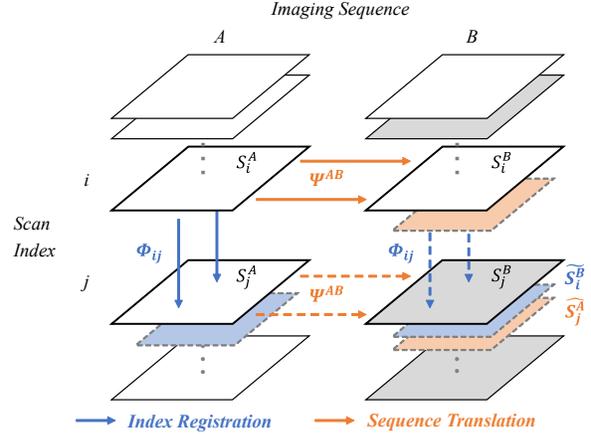

Figure 3. **Visualized explanation of invariant registration fields.** All squares indicate MRI image scans, horizontally sharing the same scan index and vertically sharing the same imaging sequence. White squares are available scans, while gray squares are missing scans. Blue bold and orange bold arrows visualize each registration field obtained by available scans of index registration and sequence translation, and dashed arrows visualize each copied field. Colored squares with dashed boundaries show estimated scans using each registration field.

ble imaging sequence, e.g., $Y \in \{$T1, T2 , T1ce, Flair, ...$\}$. We assume that the scan index is uniquely determined by the anatomical position that is represented by a scan. For example, if sequence A contains scans of index $\{1, 2, 3, 4\}$ and sequence B contains scans of index $\{1, 3\}$, two scans indexed 1 and 3 originate from the same position.

For simplicity of presentation, let us assume four scans from imaging sequences A and B. S_j^B is assumed to be missing while scans S_i^A, S_j^A, and S_i^B are available. An index registration field from S_i^A to S_j^A is acquired by specifying the scans as moving and fixed image input of $f_{\theta_{IR}}$. Similarly, an intensity registration field for sequence translation from S_i^A to S_i^B can be acquired by specifying the scans as moving and fixed image input of $g_{\theta_{ST}}$. The two registration fields are given by

$$f_{\theta_{IR}}(S_i^A, S_j^A) = \Phi_{ij}^A, \qquad g_{\theta_{ST}}(S_i^A, S_i^B) = \Psi_i^{AB}. \tag{8}$$

Both fields Φ and Ψ are trained as invariant registration fields with the assumption that scans across diverse imaging sequences with the same scan index reveal the same anatomy. The assumption practically holds while it is not exactly true as distinct sequences are taken at different phases. ***Invariant registration field*** is a novel concept defined as a registration field that can be copied along the parallel direction of alignment to match other pairs of images. As visualized in details in Fig. 3, blue and orange arrows represent index registration field Φ and intensity registration field Ψ, respectively. Dashed arrows show fields copied from bold arrows from the parallel direction - blue dashed arrows are copied from blue bold arrows on the left, and

orange dashed arrows are copied from orange bold arrows above. Two invariant fields can be formulated as following:

$$f_{\theta_{IR}}(S_i^A, S_j^A) = \Phi_{ij}^A = \Phi_{ij} = f_{\theta_{IR}}(S_i^Y, S_j^Y)$$
$$\text{for all } Y \in \{\text{T1, T2, T1ce, Flair,...}\},$$
(9)

$$g_{\theta_{ST}}(S_i^A, S_i^B) = \Psi_i^{AB} = \Psi^{AB} = g_{\theta_{ST}}(S_x^A, S_x^B)$$
$$\text{for all } x \in \{0, 1, 2, 3, ...\}.$$
(10)

Consequently, in order to impute the missing scan S_j^B, we directly warp the index registration field Φ_{ij} to S_i^B and concatenate intensity registration field Ψ^{AB} to S_j^A as

$$\widetilde{S_i^B} = S_i^B \circ \Phi_{ij}, \qquad \widehat{S_j^A} = S_j^A + \Psi^{AB}.$$
(11)

$\widetilde{S_i^B}$ shows structural changes while preserving contrast details of sequence B, and $\widehat{S_j^A}$ shows contrast changes while preserving anatomical structures of index j. To show both sides of aspects (anatomy and contrast) blended in the final substitute, a fusion process is carried out as below:

$$S_{ij}^{AB} = h(\widetilde{S_i^B}, \widehat{S_j^A})$$
(12)

where $h(,)$ is a fusion function. Image averaging is used in experiments.

3.4. Learning

Functions $f_{\theta_{IR}}$ and $g_{\theta_{ST}}$ are obtained using an encoder-decoder structure similar to Unet [31]. The moving and fixed images are concatenated into a 3D volume for an input of each network. From two networks, we aim to capture the distributions of anatomical differences via index registration field Φ, and nonlinear mappings of contrasts between imaging sequences via intensity registration field Ψ.

Both fields are trained to be invariant, so that index registration fields can be applied regardless of imaging sequence and intensity registration fields can be applied regardless of scan index. Unsupervised losses that evaluate the two networks using only input scans and generated registration fields are given by:

$$L_{IR}^A = L_{sim}(S_j^A, \widehat{S_i^A}) = L_{sim}(S_j^A, S_i^A \circ \Phi_{ij}),$$
(13)

$$L_{ST}^i = L_{sim}(S_i^B, \widehat{S_i^A}) = L_{sim}(S_i^B, S_i^A + \Psi^{AB}).$$
(14)

Losses to train the two registration fields as invariant are

$$L_{IR}^B = L_{sim}(S_j^B, \widetilde{S_i^B}) = L_{sim}(S_j^B, S_i^B \circ \Phi_{ij}),$$
(15)

$$L_{ST}^j = L_{sim}(S_j^B, \widehat{S_j^A}) = L_{sim}(S_j^B, S_j^A + \Psi^{AB}),$$
(16)

to guarantee that re-using obtained registration fields without a new acquisition procedure is practical. An auxiliary fusion loss that matches the final fused image with the desired one is given by

$$L_F = L_{sim}(S_j^B, S_{ij}^{AB}) = L_{sim}(S_j^B, h(\widetilde{S_i^B}, \widehat{S_j^A})).$$
(17)

L_{sim} is an image-wise similarity loss that is used to evaluate the similarity of predicted synthetic image and real im-

age, and is defined by

$$L_{sim}(F, M') = \frac{1}{\Omega} \sum_{p \in \Omega} [F(p) - M'(p)]^2$$
(18)

where F and M' denote the fixed and moved image, respectively, defined over spatial domain $\Omega \subset \mathbb{R}^n$. This is applicable when fixed and moved images have similar image intensity distributions and local contrast. Our total loss function is now given by

$$L = \sigma_{IR}(L_{IR}^A + L_{IR}^B) + (L_{ST}^i + L_{ST}^j) + L_F$$
(19)

where hyperparameter σ_{IR} is the weight for losses of index registration. Loss functions are also illustrated in Fig. 2.

4. Experiments

4.1. Datasets

We use the publicly available BraTS and iSeg-2017 data set. Both consist of MRI brain scans for multiple subjects, where all 3D volumes are preprocessed by skull stripping. Voxel intensities are normalized into $[0, 1]$ to guarantee comparable ranges across subjects.

BraTS [27] : Four types of imaging sequences are associated with 494 patients in BraTS: T1 , T2, T1ce, and Flair. Some imaging sequences are paired for our imputation: (T1, T2), (T2, T1), (T1, T1ce), (T1, Flair), and (T2, Flair), where the first and second components each correspond to sequence A and B, respectively. From 3D volumes of resolution $240 \times 240 \times 155$, we use 8 center axial scans out of 155 to construct the data set for experiments. This leads to ${}_8C_2 = 28$ pairs of scan index combinations and correspondingly, 28 groups (each group composed of 4 scans $S_i^A, S_j^A, S_i^B, S_j^B$) per every patient and sequence pair. The overall data are all split into 3 sets: 10332 for training, 1750 for validation, and 1750 for test.

iSeg-2017 [36] : Two types of imaging sequences are associated with 22 infants in iSeg-2017: T1 and T2. Imaging sequences are paired in bi-direction: (T1, T2), (T2, T1). From 3D volumes of resolution $192 \times 144 \times 256$, 16 center axial scans out of 256 are used. This leads to ${}_{16}C_2 = 120$ pairs of scan index combinations and thus, 120 groups per every patient and sequence pair. The overall data are all split into 3 sets: 2040 for training, 240 for validation, and 360 for test.

4.2. Baselines

The proposed method is compared to two versions of VoxelMorph, two implementations of pGAN-cGAN, CollaGAN, ResViT, and PTNet. All training procedures are carried out without any changes to each setting that is set in the original paper to guarantee their best performance. For data sets, the same number of center scans in the axial plane is used for fair comparison.

VoxelMorph [4] : This is the state-of-the-art learning-based method for medical image registration. Two ver-
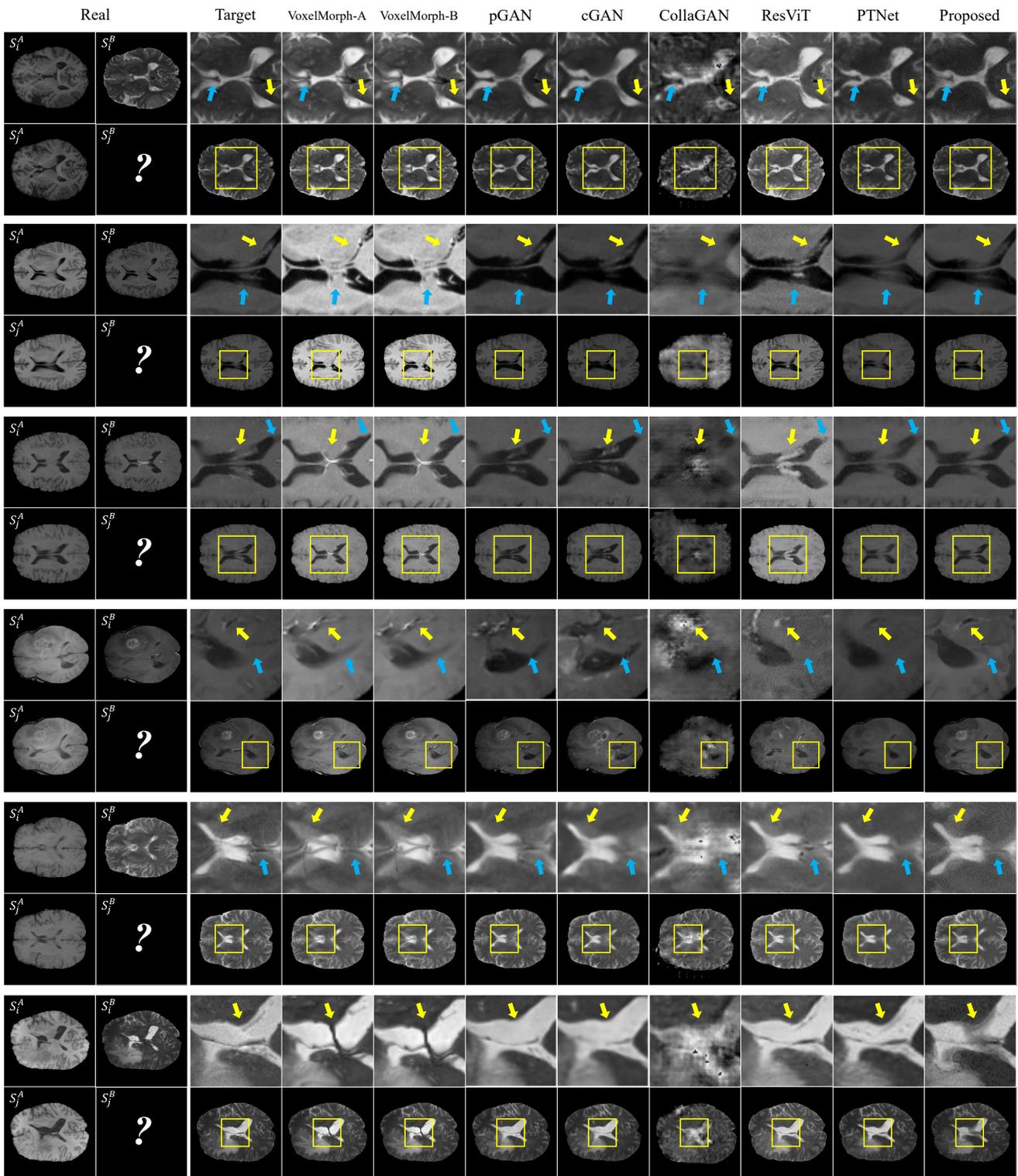
Figure 4. **Visual comparison of methods.** Synthesized scans from seven baseline methods (VoxelMorph-A, VoxelMorph-B, pGAN, cGAN, CollaGAN, ResViT, PTNet) and proposed method, compared to Target. Magnified parts of the center of brain are shown above each image. Yellow and blue arrows point out remarkable parts of the results. Zoom in to see details.

Datasets		BraTs					iSeg-2017	
Transitions		T1 → T2	T2 → T1	T1 → T1ce	T1 → Flair	T2 → Flair	T1 → T2	T2 → T1
VoxelMorph-A [4]	SSIM ↑	0.902	<u>0.918</u>	0.892	0.875	0.904	0.747	0.710
	NMSE ↓	0.109	<u>0.063</u>	0.093	0.084	**0.074**	0.149	0.170
	PSNR ↑	28.823	<u>29.361</u>	29.953	28.673	<u>29.946</u>	23.873	22.603
VoxelMorph-B [4]	SSIM ↑	<u>0.921</u>	0.899	0.899	0.883	0.883	0.779	0.695
	NMSE ↓	<u>0.084</u>	0.057	<u>0.081</u>	<u>0.083</u>	0.083	<u>0.116</u>	0.171
	PSNR ↑	<u>30.459</u>	28.499	<u>30.424</u>	<u>28.883</u>	28.883	25.231	22.337
pGAN [11]	SSIM ↑	0.908	0.915	<u>0.928</u>	0.895	0.894	0.863	0.839
	NMSE ↓	0.598	0.273	0.310	0.319	0.345	0.121	0.169
	PSNR ↑	23.778	21.297	25.555	23.258	22.821	<u>25.271</u>	22.503
cGAN [11]	SSIM ↑	0.917	0.915	0.917	<u>0.896</u>	<u>0.907</u>	0.833	0.840
	NMSE ↓	0.208	0.159	0.184	0.231	0.177	0.155	0.212
	PSNR ↑	25.290	24.625	24.950	22.944	26.607	23.825	23.596
CollaGAN [25]	SSIM ↑	0.740	0.712	0.683	0.719	0.802	0.451	0.457
	NMSE ↓	1.806	0.867	3.702	1.649	0.775	0.858	0.488
	PSNR ↑	19.035	19.993	18.478	19.278	22.109	17.277	15.384
ResViT [10]	SSIM ↑	0.910	0.908	0.915	0.884	0.888	0.864	<u>0.842</u>
	NMSE ↓	0.379	0.159	0.275	0.203	0.319	0.167	<u>0.153</u>
	PSNR ↑	23.426	22.008	24.315	23.007	22.406	25.701	<u>24.233</u>
PTNet [43]	SSIM ↑	<u>0.921</u>	<u>0.918</u>	0.911	0.886	0.902	<u>0.865</u>	0.759
	NMSE ↓	0.725	0.264	0.305	0.509	0.358	0.127	0.489
	PSNR ↑	23.338	21.457	25.868	22.634	22.477	24.670	19.028
Proposed	SSIM ↑	**0.927**	**0.920**	**0.961**	**0.920**	**0.912**	**0.869**	**0.852**
	NMSE ↓	**0.074**	**0.052**	**0.058**	**0.077**	**0.079**	**0.103**	**0.138**
	PSNR ↑	**30.775**	**29.746**	**32.734**	**30.797**	**30.313**	**27.132**	**24.725**

Table 1. Assessment of proposed method against VoxelMorph-A, VoxelMorph-B, pGAN, cGAN, CollaGAN, ResViT, and PTNet. The best results are in **bold**, and the second best results are <u>underlined</u>.

sions of VoxelMorph model are built: VoxelMorph-A and VoxelMorph-B. Each is trained by scans of sequence A and B, respectively, building registration fields for mapping pairs of scans in each sequence.

pGAN-cGAN [11] : Two implementations for MRI sequence translation - pGAN and cGAN - are provided for use. The pGAN model is useful when two images of different contrast for the same anatomy are available. It contains one pair of generator and discriminator, which is trained with pixel-wise loss and perceptual loss. The cGAN model can be used when two images of different contrast cannot be paired for the same anatomy. Two pairs of generator and discriminator are contained: one for synthesizing one contrast from the other (e.g., T2 from T1) and the other for taking care of synthesis in the opposite direction. The two pairs are combined to synthesize the missing image while trained by a cycle loss function.

CollaGAN [25] : CollaGAN is a general GAN based method for imputing a single missing data within a particular set of closely related images.

ResViT [10] : ResViT is a transformer-based generative adversarial model for multi-modal medical image synthesis. Within scans of the same anatomical structure, imputing a single missing sequence from either a group of available sequences or a single available sequence is possible.

PTNet [43] : It employs transformer layers in the bottleneck of a Unet structure for a high resolution MRI scan synthesis task. As in cross-modality translation models, it shows synthesis of a missing sequence scan from an obtained sequence scan with the same anatomy.

4.3. Settings and Evaluation Metrics

Settings : The proposed method is implemented using PyTorch [29] library, and a single NVIDIA Tesla V100 32GB GPU is used along with CUDA 11.2. All models are optimized using the ADAM [22] optimizer, and are trained with a learning rate of 0.0001 and minibatch size of 4.

Evaluation Metrics : Three standard evaluation metrics are chosen to quantitatively compare and assess different methods: structural similarity index measure (SSIM), normalized mean square error (NSME), and peak signal to noise ratio (PSNR). High scores for SSIM and PSNR, and a low score for NMSE correspond to a well-performed method.

5. Results and Discussion

5.1. Qualitative Comparisons

Missing scan S_j^B is synthesized by seven baseline methods as well as the proposed method, and six test cases are visualized in Fig. 4. The target scan is provided to visually assess each synthesized image. Center part of brains indicated by yellow boxes is magnified and shown in separate images; yellow and blue arrows point out parts that are enhanced more competitively by the proposed method.

Synthesized images from VoxelMorph and pGAN-cGAN both show missing or unnecessary anatomical parts. Moreover, some images from VoxelMorph fail to capture colors or contrasts, and pGAN-cGAN shows blurry results in some cases. These limitations are due to the fact that both rely on a single scan for synthesis, which may not be sufficient enough to capture many variables that affect the struc-

L_F	Datasets	BraTs					iSeg-2017	
	Transitions	T1 → T2	T2 → T1	T1 → T1ce	T1 → Flair	T2 → Flair	T1 → T2	T2 → T1
Without	SSIM ↑	0.918	0.873	0.929	0.861	0.887	0.779	0.824
	NMSE ↓	0.115	0.121	0.094	0.108	0.140	0.138	0.194
	PSNR ↑	29.971	27.160	30.589	28.366	28.809	25.878	24.149
With	SSIM ↑	**0.923**	**0.903**	**0.960**	**0.911**	**0.895**	**0.811**	**0.842**
	NMSE ↓	**0.086**	**0.060**	**0.060**	**0.092**	**0.134**	**0.106**	**0.142**
	PSNR ↑	**30.774**	**29.362**	**32.556**	**30.253**	**29.372**	**26.714**	**24.606**

Table 2. Ablation study of proposed method with and without auxiliary loss L_F. The best results are in **bold**.

σ_{IR}	Datasets	BraTs					iSeg-2017	
	Transitions	T1 → T2	T2 → T1	T1 → T1ce	T1 → Flair	T2 → Flair	T1 → T2	T2 → T1
1	SSIM ↑	<u>0.923</u>	<u>0.903</u>	0.960	<u>0.911</u>	<u>0.895</u>	<u>0.831</u>	<u>0.842</u>
	NMSE ↓	<u>0.086</u>	<u>0.060</u>	<u>0.060</u>	<u>0.092</u>	<u>0.104</u>	<u>0.106</u>	<u>0.142</u>
	PSNR ↑	<u>30.774</u>	29.362	32.556	<u>30.253</u>	29.372	26.714	24.606
2	SSIM ↑	**0.927**	**0.920**	**0.961**	**0.920**	**0.912**	**0.869**	**0.852**
	NMSE ↓	**0.074**	**0.052**	**0.058**	**0.077**	**0.079**	**0.103**	**0.138**
	PSNR ↑	**30.775**	**29.746**	**32.734**	**30.797**	**30.313**	**27.132**	**24.725**
4	SSIM ↑	0.898	0.885	0.959	0.911	0.889	0.809	0.824
	NMSE ↓	0.121	0.103	0.063	0.105	0.128	0.130	0.144
	PSNR ↑	29.583	28.387	32.521	30.118	29.261	26.395	24.287

Table 3. Effect of loss weight σ_{IR}. The best results are in **bold**, and the second best results are <u>underlined</u>.

tures and appearances of MRI images. Particularly, Voxel-Morph learns to capture only spatial displacements, which makes synthesizing new structures unseen in the input image and eliminating unnecessary anatomical parts difficult. The results of CollaGAN are also unreliable, showing incomplete anatomical structures and pixels with low resolution. Originally in the paper, CollaGAN attempts to impute a single missing data from remaining data within a group of closely related images. Specifically, if three images with similar anatomical structures are provided, CollaGAN synthesizes a new one that also resembles three others. This assumption does not fit the MRI imputation that we discuss. ResViT synthesizes images that visualize wrong colors or contrasts, and PTNet displays structurally missing or poor images. These pitfalls are similar to those of VoxelMorph and pGAN-cGAN. As both methods perform synthesis by using other imaging sequences, structures and contrasts from those inputs are insufficiently adjusted.

As discussed above, visualization of the results by baselines shows failure in either structural or contrast aspects; or even both. On the other hand, the images from the proposed method are very close to the targets, improving both structural and contrast details. Furthermore, it captures even small anatomical parts located in the center of brain. Unlike general images, clinically important data in medical images are often subtle and focused on small regions; in the case of brains, the center - where most registration, translation, and imputation methods fail to reconstruct or enhance. With a visual inspection, we find that the proposed method provides an actual clinically useful substitute.

5.2. Quantitative Comparisons

SSIM, NMSE, and PSNR scores are calculated between synthesized images and the targets as shown in Table 1. For every data set and sequence transition, the proposed method shows superiority by all measures - achieving the second best result for only one and the best result for the rest. Notably for T1 → T1ce transitions of BraTS, it drastically outperforms competing methods, achieving SSIM of 0.961, NMSE of 0.058, and PSNR of 32.734. All measures for transitions of BraTS are shown by SSIM of above 0.9, NMSE of below 0.1, and PSNR of around 30. Measures for transitions of iSeg-2017 are shown by SSIM of above 0.85, NMSE of slightly above 0.1, and PSNR of around 25.

5.3. Ablation Study and Choice of Hyperparameter

Effectiveness of Fusion Loss : We evaluate the effectiveness of the auxiliary fusion loss L_F by training models with and without the loss. Table 2 summarizes the results. Results show that for all data sets, transitions, and measures, models including the auxiliary loss to match the final fused image to the desired real image show improvements.

Weight of Index Registration : Between various attempts of changing σ_{IR}, a portion of the results are shown in Table 3. For all kinds of data sets, transitions, and measures, models with σ_{IR} set to 2 show the best results, and σ_{IR} set to 1 show the second best results. Due to the hardness of spatial adjustments, σ_{IR}, which is a weight for training the index registration field, heavily affects the learning model.

6. Conclusion

An MRI imputation method using novel invariant registration fields was presented. The proposed method provides clinically useful substitute results where anatomical and visual aspects are preserved and subtle details are captured.

References

[1] J. Ashburner. A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1):95–113, 2007.

[2] 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. *Medical Image Analysis*, 12(1):26–41, 2008.

[3] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. V. Guttag, and A. V. Dalca. An unsupervised learning model for deformable medical image registration. In *Proc. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 9252–9260, 2018.

[4] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. V. Guttag, and A. V. Dalca. Voxelmorph: A learning framework for deformable medical image registration. *IEEE Trans. on Medical Imaging*, 38(8):1788–1800, 2019.

[5] M. Bro-Nielsen and C. Gramkow. Fast fluid registration of medical images. In *Proc. Int. Conf. on Visualization in Biomedical Computing*, pages 265–276, 1996.

[6] T. Brox and J. Malik. Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 33(3):500–513, 2010.

[7] Z. Chen, H. Jin, Z. Lin, S. Cohen, and Y. Wu. Large displacement optical flow from nearest neighbor fields. In *Proc. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 2443–2450, 2013.

[8] E. D'agostino, F. Maes, D. Vandermeulen, and P. Suetens. A viscous fluid model for multimodal non-rigid image registration using mutual information. *Medical Image Analysis*, 7(4):565–575, 2003.

[9] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu. Unsupervised learning for fast probabilistic diffeomorphic registration. In *Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 729–738, 2018.

[10] O. Dalmaz, M. Yurt, and T. Çukur. Resvit: Residual vision transformers for multi-modal medical image synthesis. *CoRR*, abs/2106.16031, 2021.

[11] S. U. H. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, and T. Çukur. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *CoRR*, abs/1802.01221, 2018.

[12] B. D. de Vos, F. F. Berendsen, M. A. Viergever, M. Staring, and I. Išgum. End-to-end unsupervised deformable image registration with a convolutional neural network. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 204–212. Springer, 2017.

[13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *Proc. Int. Conf. on Learning Representations (ICLR)*, 2021.

[14] A. Drevelegas and N. Papanikolaou. Imaging modalities in brain tumors. In *Imaging of brain tumors with histological correlations*, pages 13–33. Springer, 2011.

[15] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio. Generative adversarial networks. *arXiv preprint arXiv:1406.2661*, abs/1406.2661, 2014.

[16] U. Hwang, D. Jung, and S. Yoon. Hexagan: Generative adversarial nets for real world classification. In *Proc. Int. Conf. on Machine Learning (ICML)*, pages 2921–2930, 2019.

[17] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu. Spatial transformer networks. In *Proc. Advances in Neural Information Processing Systems (NIPS)*, pages 2017–2025, 2015.

[18] G. Janssens, L. Jacques, J. O. de Xivry, X. Geets, and B. Macq. Diffeomorphic registration of images with variable contrast enhancement. *Int. J. of Biomedical Imaging*, 2011:891585:1–891585:16, 2011.

[19] P. Jiang and J. A. Shackleford. B-spline registration of neuroimaging modalites with map-reduce framework. In *Proc. Int. Conf. on Brain Informatics and Health*, pages 285–294, 2015.

[20] A. Jog, A. Carass, S. Roy, D. L. Pham, and J. L. Prince. Random forest regression for magnetic resonance image synthesis. *Medical Image Analysis*, 35:475–488, 2017.

[21] A. Jog, S. Roy, A. Carass, and J. L. Prince. Magnetic resonance image synthesis through patch regression. In *Proc. IEEE Int. Symposium on Biomedical Imaging (ISBI)*, pages 350–353, 2013.

[22] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *Proc. Int. Conf. on Learning Representations (ICLR)*, 2015.

[23] J. Krebs, H. Delingette, B. Mailhé, N. Ayache, and T. Mansi. Learning a probabilistic model for diffeomorphic registration. *IEEE Trans. on Medical Imaging*, 38(9):2165–2176, 2019.

[24] J. Kybic. Elastic image registration using parametric deformation models. Technical report, EPFL, 2001.

[25] D. Lee, J. Kim, W. Moon, and J. Ye. CollaGAN: Collaborative GAN for missing image data imputation. In *Proc. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 2487–2496, 2019.

[26] J. Liu, S. Pasumarthi, B. A. Duffy, E. Gong, G. Zaharchuk, and K. Datta. One model to synthesize them all: Multi-contrast multi-scale transformer for missing data imputation. *CoRR*, abs/2204.13738, 2022.

[27] B. H. Menze, A. Jakab, S. Bauer, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. on Medical Imaging*, 34(10):1993–2024, 2015.

[28] H. V. Nguyen, S. K. Zhou, and R. Vemulapalli. Cross-domain synthesis of medical images using efficient location-sensitive deep network. In *Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9349, pages 677–684, 2015.

[29] A. Paszke, S. Gross, F. Massa, et al. Pytorch: An imperative style, high-performance deep learning library. In *Proc. Advances in Neural Information Processing Systems (NIPS)*, pages 8024–8035, 2019.

[30] A. Ranjan and M. J. Black. Optical flow estimation using a spatial pyramid network. In *Proc. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 2720–2729, 2017.

[31] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351, pages 234–241, 2015.

[32] D. Shen and C. Davatzikos. Hammer: hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. on Medical Imaging*, 21(11):1421–1439, 2002.

[33] C. O. S. Sorzano, P. Thévenaz, and M. Unser. Elastic registration of biological images using vector-spline regularization. *IEEE Trans. on Biomedical Engineering*, 52(4):652–663, 2005.

[34] J.-P. Thirion. Image matching as a diffusion process: an analogy with maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998.

[35] R. Vemulapalli, H. V. Nguyen, and S. K. Zhou. Unsupervised cross-modal synthesis of subject-specific scans. In *Proc. Int. Conf. on Computer Vision (ICCV)*, pages 630–638, 2015.

[36] L. Wang, D. Nie, G. Li, É. Puybareau, J. Dolz, Q. Zhang, F. Wang, J. Xia, Z. Wu, J. Chen, et al. Benchmark on automatic six-month-old infant brain segmentation algorithms: the iseg-2017 challenge. *IEEE Trans. on Medical Imaging*, 38(9):2219–2230, 2019.

[37] Z. Xu, J. Luo, J. Yan, R. Pulya, X. Li, W. Wells, and J. Jagadeesan. Adversarial uni- and multi-modal stream networks for multimodal image registration. In *Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 12263, pages 222–232, 2020.

[38] Q. Yang, N. Li, Z. Zhao, X. Fan, E. I. Chang, and Y. Xu. MRI image-to-image translation for cross-modality image registration and segmentation. *arXiv preprint arXiv:1801.06940*, abs/1801.06940, 2018.

[39] J. Yoon, J. Jordon, and M. van der Schaar. GAIN: missing data imputation using generative adversarial nets. In *Proc. Int. Conf. on Machine Learning (ICML)*, volume 80, pages 5675–5684, 2018.

[40] J. Zhang. Inverse-consistent deep networks for unsupervised deformable image registration. *CoRR*, abs/1809.03443, 2018.

[41] J. Zhang, J. Wang, X. Wang, and D. Feng. The adaptive fem elastic model for medical image registration. *Physics in Medicine & Biology*, 59(1):97, 2013.

[42] L. Zhang, M. Pereañez, C. Bowles, S. K. Piechnik, S. Neubauer, S. E. Petersen, and A. F. Frangi. Missing slice imputation in population CMR imaging via conditional generative adversarial nets. In *Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 11765, pages 651–659, 2019.

[43] X. Zhang, X. He, J. Guo, N. Ettehadi, N. Aw, D. Semanek, J. Posner, A. Laine, and Y. Wang. Ptnet: A high-resolution infant MRI synthesizer based on transformer. *CoRR*, abs/2105.13993, 2021.

[44] A. Zhao, G. Balakrishnan, F. Durand, J. V. Guttag, and A. V. Dalca. Data augmentation using learned transformations for one-shot medical image segmentation. In *Proc. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 8543–8553, 2019.

[45] W. Zufeng, L. Tian, W. Jiang, D. Yi, and Q. Zhiguang. Medical image registration using b-spline transform. *Int. J. of Simulation Systems, Science & Technology*, 17(48), 2016.