



HAL
open science

Deformable image registration with deep network priors: a study on longitudinal PET images

Constance Fourcade, Ludovic Ferrer, Noémie Moreau, Gianmarco Santini,
Aislinn Brennan, Caroline Rousseau, Marie Lacombe, Vincent Fleury,
Mathilde Colombié, Pascal Jézéquel, et al.

► To cite this version:

Constance Fourcade, Ludovic Ferrer, Noémie Moreau, Gianmarco Santini, Aislinn Brennan, et al..
Deformable image registration with deep network priors: a study on longitudinal PET images. *Physics
in Medicine and Biology*, 2022, 67 (15), pp.155011. 10.1088/1361-6560/ac7e17 . hal-03584128v2

HAL Id: hal-03584128

<https://hal.science/hal-03584128v2>

Submitted on 15 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Deformable image registration with deep network priors: a study on longitudinal PET images

Constance Fourcade^{1,2} Ludovic Ferrer^{4,5} *PhD*
Noémie Moreau² Gianmarco Santini² *PhD*
Aislinn Brennan² Caroline Rousseau^{3,5} *MD PhD*
Marie Lacombe⁵ *MD* Vincent Fleury⁵ *MD*
Mathilde Colombié⁵ *MD* Pascal Jézéquel^{4,5} *PhD*
Mario Campone^{3,5} *MD, PhD* Mathieu Rubeaux² *PhD*
Diana Mateus¹ *PhD*

1. Ecole Centrale de Nantes, LS2N, UMR CNRS 6004, Nantes, France
2. Keosys Medical Imaging, Saint Herblain, France
3. University of Nantes, CRCINA, INSERM UMR1307, CNRS-ERL6075, Nantes, France
4. University of Angers, CRCINA, INSERM UMR1307, CNRS-ERL6075, Angers, France
5. ICO Cancer Center, Nantes - Angers, France

E-mail: constance.fourcade@ec-nantes.fr

15/03/2022

Abstract.

This paper proposes a novel approach for the longitudinal registration of PET imaging acquired for the monitoring of patients with metastatic breast cancer. Unlike with other image analysis tasks, the use of deep learning has not significantly improved the performance of image registration. With this work, we propose a new registration approach to bridge the performance gap between conventional and deep learning-based methods: Medical Image Registration method Regularized By Architecture (MIRRBA). MIRRBA is a subject-specific deformable registration method which relies on a deep pyramidal architecture to parametrize the deformation field. Diverging from the usual deep-learning paradigms, MIRRBA does not require a learning database, but only a pair of images to be registered that is used to optimize the network's parameters. We applied MIRRBA on a private dataset of 110 whole-body PET images of patients with metastatic breast cancer. We used different architecture configurations to produce the deformation field and studied the results obtained. We also compared our method to several standard registration approaches: two conventional iterative registration methods (ANTs and Elastix) and two supervised deep learning-based models (LapIRN and Voxelmorph). Registration accuracy at a global and local level was evaluated using the detection rate and the Dice score respectively, while the realism of the registration obtained was evaluated using Jacobian's determinant. The ability of the different methods to shrink disappearing lesions was also computed with the disappearing rate. MIRRBA significantly improved the organ and lesion Dice scores of Voxelmorph by 6% and 52% respectively, and of LapIRN by 5% and 65%. Regarding the disappearing rate, MIRRBA more than doubled the score of the best performing conventional approach

ANTs. In this paper, we also demonstrate the regularizing power of deep architectures and present new elements to understand the role of the architecture in deep learning methods used for registration.

Keywords: Image registration, PET, Breast cancer, Deep Image Prior

1. Introduction

Medical image registration is the precise overlaying of a fixed and moving image. It is notably used to create patient models (inter-patient monomodal registration), exploit information from different modalities acquired for a single patient (intra-patient multimodal registration), and to monitor tumor evolution (longitudinal registration).

However, the accuracy of image registration is a long-standing fundamental problem in medical image analysis ([Maurer and Fitzpatrick, 1993, Fu et al., 2020]). In the past 30 years, registration methods have considered registration as an optimization problem: the goal is to minimize a dissimilarity term between moving and fixed images. It is usually measured with mean square error (MSE), mutual information (MI) or normalized cross-correlation (NCC). In addition, a term enforcing smooth and realistic transformations is often used as a regularizing term.

Most conventional methods manage registration by parameterizing the transformation between the fixed and moving image. They use for instance discrete cosine transforms ([Friston et al., 1995]), 3D Fourier series ([Christensen et al., 2007]), cubic B-splines ([Klein et al., 2010]) or optimized velocity fields ([Avants et al., 2009]). The parameters of the model are typically optimized through efficient second-order minimization ([Vercauteren et al., 2007b]) or stochastic gradient descent (SGD) ([Klein et al., 2009a]), within a single- or multi-level optimization strategy. The minimized cost function combines dissimilarity and regularization terms.

Recent deep learning (DL) supervised registration methods ([Fu et al., 2020]) also rely on dissimilarity measurements and regularization terms that can be optimized with variants of SGD. However, there are two key differences compared to conventional approaches. DL methods require a training stage with a training database that gives the model prior knowledge regarding deformations. They also use a different type of transformation model, with convolutional neural networks (CNN) used as an over-parameterized but structured model of the deformation field.

Despite recent developments in DL-based registration ([Chen et al., 2021]), conventional methods still perform better and obtain more accurate results in most applications ([Heinrich and Hansen, 2020]). In addition, even though trained networks are faster than conventional methods, training requires large databases which often restricts their use to a specific therapeutic area.

With this work, we propose a new registration method formalized as a conventional registration approach, with a deformation field modeled by an untrained deep pyramidal

network. We named our approach MIRRBA for *Medical Image Registration method Regularized By Architecture*.

We applied MIRRBA on longitudinal PET images acquired for the evaluation of treatment response in patients with metastatic breast cancer and compared the results to registrations obtained with standard approaches. We also studied the impact of different architecture configurations (14 configurations) on the deformation field.

The contributions of this paper are i) the proposition of a novel registration method regularized by architecture MIRRBA, ii) an extensive comparative study of the effects of different network components on the deformation field, and iii) a solution to register whole body PET images both at a global and local level without the need for prior registration to facilitate the simultaneous monitoring of multiple lesions.

2. Related work

Automatic longitudinal lesion monitoring Automatic approaches for lesion segmentation and/or image registration have been developed over the years to help with cancer monitoring and treatment response assessments.

[Necib et al., 2011] used affine registration to align baseline and follow-up PET images, before subtracting them to identify tumor voxels showing significant changes between the two scans. However, since affine registration is a global method, it mainly performs well for single localized tumors. In a situation where there were multiple lesions to assess, [Hsu, 2020] first segmented all lesions on liver-centered CT images, then used a longitudinal correspondence module to find matching pairs of lesions from the segmentation maps, and finally computed the lesions' evolution. Even if promising, multi-staged methods suffer from error propagation between stages. [Chassagnon et al., 2020] removed the need for a prior segmentation step and relied on a conventional registration algorithm between longitudinal CT pairs to obtain the deformation fields and their Jacobian determinants. This information was then used to train a classifier network assessing systemic sclerosis interstitial lung disease.

Image registration can be a key step in lesion monitoring. As our long-term goal is to use registration results to develop tools for the monitoring of metastatic breast cancer, this work focuses on registration and on how to improve its accuracy. We present hereafter the most used registration approaches and the good practices which contribute to make them successful.

Established registration tools There is a large body of literature on conventional registration methods ([Klein et al., 2009b, Sotiras et al., 2013]). Some of today's most used registration tools include the Demons method ([Vercauteren et al., 2007a]) using non-parametric diffeomorphic displacement fields, the Elastix toolbox ([Klein et al., 2010]) based on cubic B-splines, and the Advanced Normalization Tools (ANTs) ([Avants et al., 2009]) parametrizing the velocity field and relying on bi-directional diffeomorphisms. We compared our proposed approach to the last two methods, as they

perform well with different datasets ([Klein et al., 2009b]) and their pyramidal coarse-to-fine optimization has inspired recent works on DL-based registration which is presented hereafter.

DL methods In recent years, different types of DL-based registration approaches have been proposed ([Chen et al., 2021]). With the monitoring of metastatic breast cancer, deformation fields are not available and are difficult to obtain. We therefore focused on unsupervised registration methods. We categorize as unsupervised methods which do not require ground truth deformation fields, segmentations, or landmarks. Typically, they learn by minimizing a dissimilarity term between the fixed and the wrapped moving image. Among the unsupervised registration methods, various CNN architectures have been proposed. The widely used U-Net ([Ronneberger et al., 2015]) inspired the reference Voxelmorph network first trained on brain MRIs ([Balakrishnan et al., 2018]). [Stergios et al., 2018] added dilated convolutions to the encoder path of Voxelmorph to capture a wider field of view from lung MRI images. Later, [de Vos et al., 2019] proposed a U-Net-based cascade network applied to cardiac cine MRI and chest CT data to perform affine and deformable registrations in stages, at the cost of an increase in complexity. On 3D PET images, [Li et al., 2021] also proposed an iterative DL-based registration method to reconstruct motion compensated PET images. While still using a U-Net-shaped architecture, [Eppenhof et al., 2019] proposed a coarse-to-fine training mimicking the best performing conventional approaches. With a similar training strategy, [Mok and Chung, 2020] won the Learn2Reg 2021 MICCAI challenge[‡] using a pyramidal network named LapIRN. Contrary to the above cited methods, [Mok and Chung, 2020] did not only impose a smoothness constraint on the deformation field through its gradient, but also enforced diffeomorphic transformations using stationary vector fields under the Log-Euclidean framework, as in [Dalca et al., 2019].

Despite recent efforts, when using DL-based registration techniques no major gain in registration performance was reported. Conventional iterative optimization methods still yield better accuracy in many tasks such as inter-patient alignment or intra-patient lung motion registration ([Heinrich and Hansen, 2020]). Large databases would be needed to learn the network parameters and produce accurate deformation fields. Moreover, we can argue that the generalization ability of the trained network is questionable when the deformation patterns do not repeat consistently across the dataset.

Deep Image Prior registration To alleviate the dependency to a dataset and the need for repeated deformation patterns, we adapted the Deep Image Prior (DIP) framework ([Ulyanov et al., 2020]) to propose a learning-free method for deformable medical image registration. Contrary to standard DL-based approaches, DIP does not learn from a database but relies on a single image. It uses a deep architecture not to summarize the

[‡] <https://learn2reg.grand-challenge.org/>

information across samples but as a prior. Effectively, the architecture plays the role of a parametric model in an optimization problem restricting the solution space. While DIP was initially designed for denoising and inpainting tasks ([Ulyanov et al., 2020]), we adapted it here to medical image registration.

[Laves et al., 2019] suggested image registration as a potential application of DIP, with preliminary results in the context of 2D brain MR images. DIP has also been used on medical data for the reconstruction of CT and PET images ([Gong et al., 2019, Bagger et al., 2020]). To perform these reconstructions, an untrained deep network was used as a denoiser and iteratively optimized by a conventional algorithm. Apart from the use of an initial reconstruction and classical regularization, images from other modalities were also given as input to condition the network output.

With this work, we establish a link between conventional, deep learning and DIP-based approaches. In particular, we focus on the role of the untrained network in parameterizing the displacement field, showing that each architecture induces an implicit regularization when used within a conventional optimization scheme. Similar observations have been made in the context of DL for inverse problems [Lucas et al., 2018] and [Dittmer et al., 2020]. As a result, well-structured architectures (e.g. [Mok and Chung, 2020]) provide better over-parameterizations for the deformation fields, both in the supervised and in the untrained cases. We also investigate the role of the input (random, moving or moving and fixed images) and potential interactions of the untrained network with conventional supervised approaches.

3. Method

3.1. Image registration

Given a pair of fixed and moving (F , M) images, deformable image registration aims to estimate a dense deformation field Φ such that the wrapped image $W = M(x + \Phi(x))$ is aligned with F for each voxel x . Given a metric $\mathcal{S}(\cdot, \cdot)$ measuring the dissimilarity between two images, conventional registration methods choose a parametrization of transformation and optimize it through a cost function of the form:

$$\arg \min_{\Phi} \mathcal{S}(F, W) \quad (1)$$

However, the problem stated Eq. 1 is ill-posed: it belongs to the nonlinear transformation class ([Myronenko and Song, 2009]). By explicitly constraining the transformation, parametric approaches can become well-posed ([Rueckert et al., 1999]), at the cost of limiting the range of admissible transformations. Constraining an over-parameterized or complex deformable transformation is effectively done by adding a regularization term $\mathcal{R}_{smooth}(\Phi)$ to the objective function Eq. 1. Regularization terms incorporate priors, which if correctly chosen, guide the optimization towards better estimates of the deformation fields. Weighted by λ_{smooth} , the most common regularization term enforces

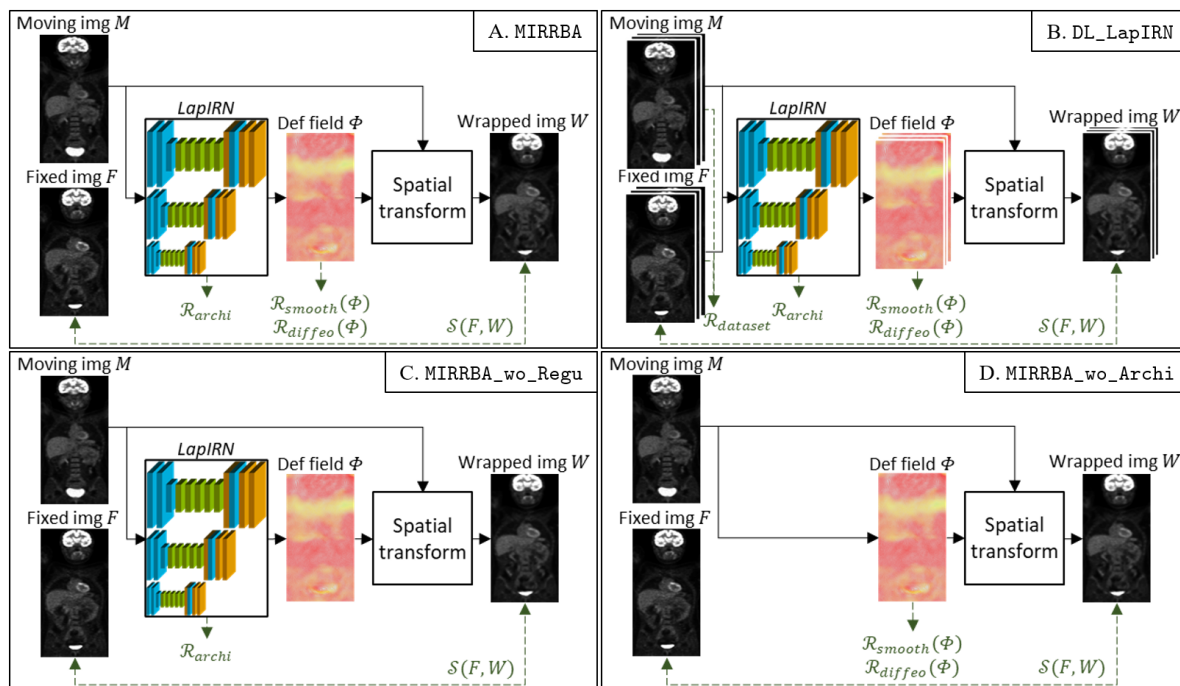


Figure 1. Overview of A. MIRRBA, B. DL_LapIRN and C. MIRRBA_wo_Regu and D. MIRRBA_wo_Archi methods. The LapIRN architecture is visible with the encoder path (blue), the residual blocks (green) and the decoder path (orange with a blue layer from the encoder path). Best viewed in color.

smoothness onto the displacement field by penalizing the spatial derivatives of Φ .

$$\arg \min_{\Phi} \{ \mathcal{S}(F, W) + \lambda_{smooth} \mathcal{R}_{smooth}(\Phi) \} \quad (2)$$

DL-based registration approaches ([Fu et al., 2020]) adapt the cost function in Eq. 2 by modelling Φ to be the output of a CNN trained on a dataset (See Eq. 3). The inputs of this network are the (F, M) pair, and the warping operation to get W is performed using a spatial transformer layer ([Jaderberg et al., 2015]), handling the transformation, the sampling and the interpolation. Regarding the regularization, DL-registration based methods continue relying on a term enforcing the smoothness of the transformation \mathcal{R}_{smooth} , as done in conventional approaches. Smoothness is often completed by \mathcal{R}_{diffeo} , which enforces the diffeomorphism of the transformation by penalizing the determinant of the Jacobian negative values. Regularization terms are weighted respectively by λ_{smooth} and λ_{diffeo} . Moreover, we state that two additional regularizing priors are also implicitly added in the DL setup and influence the predicted deformation field Φ : the first one $\mathcal{R}_{dataset}$ is induced by training on a domain-specific dataset, while the second one \mathcal{R}_{archi} is entailed by the network architecture choice (see Fig. 1.B). Due to the dataset dependency, one limitation of DL-based registration methods is the difficulty to generalize across organs or modalities.

$$\arg \min_{\Phi(\mathcal{R}_{dataset}, \mathcal{R}_{archi})} \{ \mathcal{S}(F, W) + \lambda_{smooth} \mathcal{R}_{smooth}(\Phi) + \lambda_{diffeo} \mathcal{R}_{diffeo}(\Phi) \} \quad (3)$$

Next, we recall the deep image prior concept, which removes the dataset dependency in Section 3.2, and present our method for registration based on an untrained network in Section 3.3.

3.2. Deep Image Prior

The DIP method proposed in [Ulyanov et al., 2020] uses a deep architecture to denoise images using a network, but without any prior learning. Supposing X_0 a distorted image and X the network output, the fitting process is characterized by Eq. 4, with \mathcal{R}_{archi} the implicit prior captured by the network architecture and \mathcal{S}_{DIP} a reconstruction function on a single image.

$$\arg \min_{X(\mathcal{R}_{archi})} \{\mathcal{S}_{DIP}(X_0, X)\} \quad (4)$$

DIP reconstructs a noisy image (e.g. with JPEG compression noise) from a white noise image by training a generator architecture to fit the noisy image. A denoised image is obtained by stopping the training before completely fitting the noise. We propose to adapt this idea to the registration of a pair of images, where we modify the moving image to match the fixed one. Contrary to the original DIP, we are interested in fitting all the way up to the finest deformations.

3.3. MIRRBA (Medical Image Registration method Regularized by Architecture)

In this paper, we argue that deep architectures, as parametric models with high capacity, are powerful representations for deformation fields. They can thus be exploited as implicit priors for image registration in iterative optimization schemes without a training stage ([Heckel and Soltanolkotabi, 2020, Dittmer et al., 2020, Lucas et al., 2018]). Although DIP priors have been explored in the context of image reconstruction, there is no prior in depth study of such architecture priors in the context of image registration.

Since there is no learning step, the training dataset no longer influences the transformation. Thus, we transform Eq. 3 into Eq. 5 to directly optimize a patient-specific CNN for the pair of interest (F , M), as traditionally done with iterative optimization methods (see Fig. 1.A).

$$\arg \min_{\Phi(\mathcal{R}_{archi})} \{\mathcal{S}(F, W) + \lambda_{smooth} \mathcal{R}_{smooth}(\Phi) + \lambda_{diffeo} \mathcal{R}_{diffeo}(\Phi)\} \quad (5)$$

By optimizing Eq. 5, we find the best warped image allowed by the over-parametrization of the architecture. Carefully designing the architecture to be used, it is possible to integrate several of the commonly used tricks in conventional approaches. In this paper, we rely on the LapIRN architecture ([Mok and Chung, 2020]), which incorporates filtering through convolutions, pyramidal coarse-to-fine refinement, and interpolation steps with down- and transpose convolutions. Eq. 5 is optimized with an SGD for an input pair of images. Our idea applies to other architectures and optimization algorithms.

4. Experimental setup

4.1. Dataset description

We ran our experiments on images from a private dataset, acquired in the context of the ongoing prospective multicentric XXX study for metastatic breast cancer monitoring. Patients underwent between two and three PET/CT acquisitions, corresponding to pre-, early- (after a month) and mid-treatment time points. A total number of 110 pairs of images were obtained, a pair being composed of a pre-treatment and either an early- or a mid-treatment image (58 and 52 images respectively). Images were acquired at two different centers. The 54 pairs of images from center A were obtained using a Philips Vereos or a GE Discovery PET/CT imaging systems, while the 56 pairs of images from center B were acquired on two different dual-slice Siemens Biograph PET/CT scanners.

Since we are interested in lesion monitoring in the context of metastatic breast cancer, we worked only on PET images, as shown useful in previous studies [Carlier and Bailly, 2015] and [Avril et al., 2016]. Expert physicians manually segmented all lesions. The brain and the bladder were also delineated, since they can be useful to mask irrelevant regions for patient response assessment. All PET images were normalized by the standardized uptake value (SUV) ([Kim et al., 1994]) and resampled to obtain an isotropic resolution of $1 \times 1 \times 1 \text{ mm}^3$ with 200 pixels for each side. Besides, no prior registration of any kind was performed.

This prognostic study was approved by XXX and a written informed consent was obtained from all patients.

4.2. Architectural implementation details

Our MIRRBA method relies on the LapIRN network architecture proposed in [Mok and Chung, 2020]. LapIRN is a pyramidal network with $N = 3$ depth levels, each level being composed of a feature encoder, a set of residual blocks and a feature decoder, as shown in Fig. 1. For each level $L_{i \in \{1,2,3\}}$, input images are downsampled by a factor $0.5^{(N-1)}$ using a trilinear interpolation. Hence, for the coarsest level L_1 the image resolution is divided by 4, while for the finest L_3 it remains identical. Moreover, a scaling and squaring module ([Dalca et al., 2019]) enforces diffeomorphic deformations.

Network levels were trained in a coarse-to-fine manner, meaning the coarsest level L_1 is first trained alone, and then higher levels are progressively trained to refine the registration. To avoid unstable starts when training levels $L_{i>1}$, lower levels weights were frozen for a fixed number of epochs. Regarding the optimization process, the learning rate was set to $10e-4$, and the Adam optimizer was used for 1000 iterations on the two lower levels and 2000 iterations on the finest. As a dissimilarity metric, we used the NCC, regularized by smooth and diffeomorphic terms, for which λ_{smooth} and λ_{diffeo} were determined with a grid search to 0.1 and 1.0 respectively to minimize the overall loss function over a test set (see Eq. 5), as described below for DL-based methods training. Indeed, since we aim at using registration-based features to monitor metastatic breast

cancer in future work, we need to reach precise global and local registration. More details can be found in [Mok and Chung, 2020].

Starting from the architecture in Fig. 1, we performed an ablation study to measure the influence of each component. Hence, we i) changed the depth of the network (network with 1, 2, 3 or 4 depth levels), ii) computed the results after training each level during the coarse-to-fine registration i.e. the coarsest, the intermediate and the finest, iii) deleted the residual connections of residual blocks to transform them into simple convolutional blocks, iv) replaced the down- and up-convolutions to respectively max-pooling and upsampling operations, v) used deformable convolutions ([Dai et al., 2017]) in the finest level, and vi) used a Gaussian noise image $\mathcal{N}(0, 0.001)$ (as in [Laves et al., 2019]) or concatenated fixed and moving images as network inputs. Moreover, we also vii) set both λ_{smooth} and λ_{diffeo} to 0 to remove the registration-specific regularization terms (see Fig. 1.C). All these implementations are detailed in Table 1.

4.3. Reference methods implementation details

Regarding the methods used as reference, we used a grid search on a test set (as described below for DL-based methods training) to find hyperparameters yielding best performance running the ANTs pipeline ([Avants et al., 2009]), i.e. with a three resolutions coarse-to-fine optimization, a gradient step of 0.2 and a symmetry transformation penalty. We used both MI and NCC as similarity metrics, leading to **SyN** and **SyNCC** methods respectively. We also ran the Elastix pipeline ([Klein et al., 2010]) to perform successive rigid, affine and deformable image registration with four resolutions optimized with an adaptive SGD minimizing the NCC similarity term for 1000 iterations (**Elastix**). As for our method, we used a penalty on the bending energy as regularization to enforce smooth deformations.

To highlight the regularizing power of the architecture in MIRRBA, we ran the pipeline without the network, optimizing directly the deformation field (initialized as a Gaussian noise $\mathcal{N}(0, 0.001)$) with the Adam optimizer, **MIRRBA_wo_Archi** (see Table 1 and Fig 1.D).

Regarding deep learning-based methods, DL-LapIRN ([Mok and Chung, 2020]) was ran with a similar regularization balance as MIRRBA, i.e. with $\lambda_{smooth} = 0.1$, while all other recommended settings were used. For DL-Voxelmorph, we used the latest diffeomorphic version to date ([Dalca et al., 2019]), with the NCC loss and recommended settings. For methods relying on a training stage, we split our dataset into five folds, paying attention to balance data from different acquisition centers among folds. For each fold, we refer to \mathcal{D}_{train} and \mathcal{D}_{test} as the train and test dataset, respectively. We trained the DL-based approaches on \mathcal{D}_{train} before testing them on \mathcal{D}_{test} for the five folds.

All architectures were implemented with PyTorch ([Paszke et al., 2017]) and trained from scratch on a Nvidia V100 32GB SXM2 GPU.

Table 1. Loss terms and structural setups of the methods presented in the paper. “Pyr. net.” stands for “Pyraminal network”, “Sym. diffeo. transfo.” for “Symmetric diffeomorphic transformation”, “Max.” for “Maxpooling”, “Up.” for “Upsampling”, and “Def. conv.” for “Deformable convolutions”. An absence of information in the *Other* column indicates the use of setups described in Sections 4.2 and 4.3.

	Loss terms				Structural choices				
	$\mathcal{R}_{dataset}$	\mathcal{R}_{smooth}	\mathcal{R}_{diffeo}	\mathcal{R}_{archi}	Model	Depth	Trained levels	Input images	Other
MIRRBA		x	x	x	Pyr. net.	3	3	Moving	-
DL_LapIRN	x	x	x	x	Pyr. net.	3	3	Fixed & moving	-
DL_Voxelmorph	x	x	x	x	U-shaped network	1	1	Fixed & moving	-
Elastix		x			Cubic B-splines	4	4	Moving	-
SyN		x			Sym. diffeo. transfo.	3	3	Moving	-
SyNCC		x			Sym. diffeo. transfo.	3	3	Moving	-
Combined		x	x	x	Sym. diffeo. transfo. & Pyr. net.	3 - 4	3 - 4	Fixed & moving	-
MIRRBA_wo_Regu				x	Pyr. net.	3	3	Moving	-
MIRRBA_wo_Archi		x	x		Pyr. net.	3	3	Moving	-
MIRRBA_Depth_1		x	x	x	Pyr. net.	1	1	Moving	-
MIRRBA_Depth_2		x	x	x	Pyr. net.	2	2	Moving	-
MIRRBA_Depth_4		x	x	x	Pyr. net.	4	4	Moving	-
MIRRBA_Level_1		x	x	x	Pyr. net.	3	1	Moving	-
MIRRBA_Level_2		x	x	x	Pyr. net.	3	2	Moving	-
MIRRBA_Max		x	x	x	Pyr. net.	3	3	Moving	Max.
MIRRBA_Up		x	x	x	Pyr. net.	3	3	Moving	Up.
MIRRBA_DefConv		x	x	x	Pyr. net.	3	3	Moving	Def. conv.
MIRRBA_NoiseImg		x	x	x	Pyr. net.	3	3	White noise	-
MIRRBA_FixImg		x	x	x	Pyr. net.	3	3	Fixed & moving	-
MIRRBA_Best		x	x	x	Pyr. net.	4	4	Fixed & moving	Max. Up. Def. conv.

4.4. Evaluation metrics

The first criterion we used to evaluate the registration accuracy is the *detection rate*, defined for an individual as the percentage of lesions presenting an overlap greater than

50% between the wrapped and fixed segmentations ([Moreau et al., 2020]). A higher percentage refers to a better detection.

To evaluate the registration more locally, we computed *Dice scores* between fixed and wrapped i) brain and bladder and ii) lesions segmentation masks. A Dice score close to 1 means a precise local registration, while a Dice close to 0 is unsatisfactory. Since the detection rate represents the percentage of lesions correctly detected, it is positively correlated with the Dice score of the lesions. However, since some lesions (155) are cured over time and disappear on PET images, we removed them from the Dice score and detection rate computations to avoid erroneous null values. Instead, we evaluated the capacity of a method to effectively make lesions disappear by computing its *disappearing rate*, or percentage of volume reduction of a lesion induced by the registration, where a complete disappearance would mean a rate of 100%.

Registration smoothness was evaluated by measuring, for every deformation field, the standard deviation of its Jacobian determinant *SDJDet*. Null values indicate an identity transformation and high ones disorganized and incoherent displacements. Although an optimal value is difficult to define, we sought to obtain small positive values, characterizing smooth deformations ([Mok and Chung, 2020]).

Finally, we evaluated the approximate running time of each approach, using a CPU for SyN, SyNCC and Elastix, and a GPU for all other methods.

4.5. Statistical analysis

To evaluate the statistical significance of our results, we first studied their distribution. According to the Shapiro-Wilk test (testing the null hypothesis that a sample comes from a normal distribution), we cannot reject the null hypothesis. Hence, we ran a paired sample t-test on our results and considered them of statistical significance if $p < 0.05$.

5. Results

5.1. Regularization terms

In the first experiment, we looked at the influence of the different regularization terms from Eq. 3, i.e. \mathcal{R}_{smooth} , \mathcal{R}_{diffeo} , $\mathcal{R}_{dataset}$ and \mathcal{R}_{archi} . To this end, we compared the results to our MIRRBA approach optimized directly (without learning) on the \mathcal{D}_{test} pairs, using \mathcal{R}_{smooth} , \mathcal{R}_{diffeo} and \mathcal{R}_{archi} terms (see Table 1, Fig. 1.A and Eq. 5).

Dataset regularization We trained the Voxelmorph (DL_VoxelMorph) ([Dalca et al., 2019]) and LapIRN (DL_LapIRN) ([Mok and Chung, 2020]) with the four regularization terms from Eq. 3 (see Table 1 and Fig. 1.B). Predictions were made on \mathcal{D}_{test} images. Even if both DL-based methods reached similar numerical results (see Table 2) and coherent visual ones (see Fig. 2 and supplementary material), MIRRBA performed better than both of them globally and locally. Indeed, from Table 2, we note that running a patient-specific

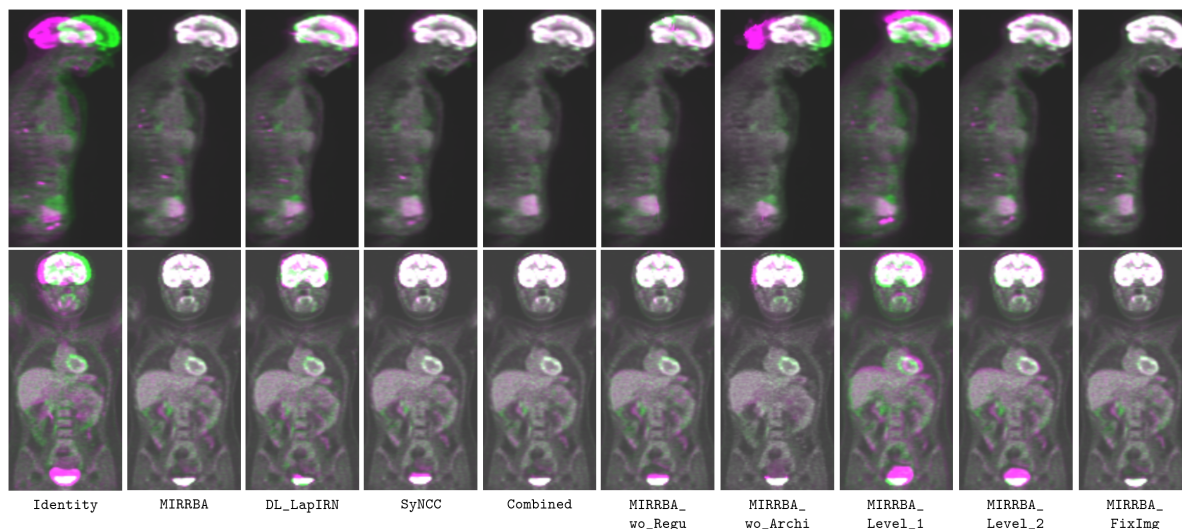


Figure 2. Overlay of the fixed (green) and warped (pink) images on two different patients after performing Identity registration, MIRRBA (corresponding to level 3), DL_LapIRN, SyNCC, Combined, MIRRBA_wo_Regu, MIRRBA_wo_Archi, MIRRBA_Level_1, MIRRBA_Level_2 and MIRRBA_FixImg. Grayscale color indicates good overlapping. It can be noticed that MIRRBA registration is better than DL_LapIRN and SyNCC around the bladder, while Combined obtains nicely registered images. Even without registration-specific regularization terms, MIRRBA_wo_regu wrapped images are realistic looking, unlike MIRRBA_wo_archi. We can also note that the coarsest level (MIRRBA_Level_1) performs global registration, while finest ones refine it (MIRRBA_Level_2) and achieve more local registration (MIRRBA). MIRRBA_FixImg reaches visually accurate registration. Best viewed in color.

optimization, i.e. removing $\mathcal{R}_{dataset}$ and optimizing an untrained network MIRRBA compared to DL_Voxelmorph and DL_LapIRN, helps improve the results especially for the lesions. The organs’ Dice score improved by 6% and 5% while the Dice of the lesions improved by 52% and 65% for DL_Voxelmorph and DL_LapIRN respectively. Moreover, MIRRBA presented lower SDJDet values, hence produced smoother deformations than training-based approaches. On the other hand, DL_Voxelmorph and DL_LapIRN had a higher disappearing rate.

Registration-specific regularization To understand the influence of the registration-specific regularization terms \mathcal{R}_{smooth} and \mathcal{R}_{diffeo} , we removed them from Eq. 5 in MIRRBA_wo_Regu (see Table 1) and Fig. 1.C). As quantitatively show in Table 3, MIRRBA presented Dice scores for the organs and lesions which are respectively 6% and 4% higher compared to MIRRBA_wo_Regu. Yet, the detection and disappearing rates were increased for MIRRBA_wo_Regu over MIRRBA.

Architecture regularization Finally, to also understand the impact of the regularization power of the network architecture on the registration, we looked at the results of MIRRBA_wo_Archi, which directly optimizes the deformation field without including any

Table 2. Comparison of MIRRBA to DL-based methods (Section 5.1) – Detection rate, Dice score of the organs and lesions (Dice_organ and Dice_lesions resp.), SDJDet, disappearing rate, and approximate computational time. Both training and inference computational times are indicated for DL-based methods. All pipelines were computed on GPU. Statistically significant improvement of our MIRRBA method over the others with $p < 0.05$ is indicated with *. Best results are marked in bold, and second best ones in bold-italic, except for SDJDet since no ideal value is defined.

	Detec. rate (%) \uparrow	Dice organs \uparrow	Dice lesions \uparrow	SDJDet \downarrow	Disap. rate (%) \uparrow	Time (min) \downarrow
Identity	5.00	0.626 \pm 0.138*	0.090 \pm 0.115*	0.000 \pm 0.000	0.00*	0
MIRRBA	33.04	0.918 \pm 0.126	0.425 \pm 0.207	0.124 \pm 0.988	9.36	55
DL_LapIRN	11.76	<i>0.878 \pm 0.076*</i>	0.258 \pm 0.198*	0.464 \pm 1.367	19.15	1450 – 3
DL_Voxelmorph	<i>14.13</i>	0.865 \pm 0.077*	<i>0.279 \pm 0.192*</i>	0.224 \pm 0.186	5.67	1200 – 2

Table 3. Ablation study on the regularization terms, architectural choices and inputs (Sections 5.1 & 5.2) – Detection rate, Dice score of the organs and lesions, SDJDet, disappearing rate, and approximate computational time. All pipelines were computed on GPU. Statistically significant improvement of our MIRRBA method over the others with $p < 0.05$ is indicated with *. Best results are marked in bold, and second best ones in bold-italic, except for SDJDet since no ideal value is defined.

	Detec. rate (%) \uparrow	Dice organs \uparrow	Dice lesions \uparrow	SDJDet \downarrow	Disap. rate (%) \uparrow	Time (min) \downarrow
Identity	5.00	0.626 \pm 0.138*	0.090 \pm 0.115*	0.000 \pm 0.000	0.00*	0
MIRRBA	33.04	0.918 \pm 0.126	0.425 \pm 0.207	0.124 \pm 0.988	9.36	55
MIRRBA_wo_Regu	33.70	0.868 \pm 0.199*	0.407 \pm 0.219*	6.655 \pm 30.709	16.84	55
MIRRBA_wo_Archi	20.54	0.753 \pm 0.144*	0.239 \pm 0.223*	1.247 \pm 0.475	59.95	20
MIRRBA_Depth_1	11.74	0.748 \pm 0.169*	0.239 \pm 0.211*	0.016 \pm 0.047	4.82	30
MIRRBA_Depth_2	23.26	0.873 \pm 0.135*	0.364 \pm 0.211*	0.038 \pm 0.174	7.56	45
MIRRBA_Depth_4	<i>40.22</i>	<i>0.945 \pm 0.012</i>	<i>0.466 \pm 0.199</i>	0.057 \pm 0.058	15.12	60
MIRRBA_Level_1	4.35	0.722 \pm 0.091*	0.221 \pm 0.157*	0.013 \pm 0.010	0.00*	2
MIRRBA_Level_2	19.78	0.869 \pm 0.114*	0.371 \pm 0.198*	0.035 \pm 0.1737	0.00*	10
MIRRBA_Max	33.26	0.922 \pm 0.098	0.426 \pm 0.213	0.316 \pm 2.934	8.34	55
MIRRBA_Up	33.04	0.922 \pm 0.112	0.439 \pm 0.204*	0.056 \pm 0.309	8.84	55
MIRRBA_DefConv	35.22	0.487 \pm 0.463*	0.257 \pm 0.275*	0.027 \pm 0.020	12.23	130
MIRRBA_NoiseImg	22.44	0.892 \pm 0.081*	0.381 \pm 0.194*	0.007 \pm 0.009	0.00*	55
MIRRBA_FixImg	36.03	0.941 \pm 0.017	0.451 \pm 0.197	0.071 \pm 0.095	20.52	60
MIRRBA_Best	40.94	0.947 \pm 0.010	0.467 \pm 0.202	0.080 \pm 0.085	19.48	60

specific architecture, hence not including \mathcal{R}_{archi} (see Table 1 and Fig. 1.D). According to Table 3, MIRRBA_wo_Archi was outperformed by MIRRBA for both organ and lesion segmentations by 22% and 78% respectively. Although MIRRBA_wo_Archi disappearing rate was higher than MIRRBA’s, the method without architectural regularization presented a low detection rate confirming the structural bias of convolutional generators.

5.2. Architectural choices

To study the regularization effect of a network architecture on registration, we compared the results of various architectural choices built from the pyramidal network presented in [Mok and Chung, 2020] (see Table 1). Quantitative results are presented in Table 3, while qualitative ones are visible Fig. 2 and in supplementary materials.

Depth of the pyramidal network First, we modified the architecture to optimize a simple U-Net-shaped network (`MIRRBA_Depth_1`), and pyramidal ones with two (`MIRRBA_Depth_2`), three (`MIRRBA`), and four (`MIRRBA_Depth_4`) resolutions. While `MIRRBA_Depth_1` was optimized on full resolution images, all other architectures were trained using a coarse-to-fine strategy.

Results show that increasing the network depth improves the Dice results, as well as the detection and disappearing rates. Indeed, `MIRRBA_Depth_4` presented the second highest Dice scores for both the organs and the lesions among all `MIRRBA` setups. Regarding the `SDJDet` values, they were higher when using a depth 3 or 4, than when the network was trained on less resolutions.

Trained network level To understand the amount of information brought by each network level during the coarse-to-fine training strategy, we computed the registration after optimizing only the lower level (`MIRRBA_Level_1`) on coarse resolution images, both lower levels (`MIRRBA_Level_2`) on coarse and medium resolution images, and the whole network (`MIRRBA`) with the complete coarse-to-fine training strategy.

According to Table 3, training on all three levels of the network improved the registration accuracy over training only on low resolution images. Moreover, the disappearing rate was null when the training only occurred on low resolution images.

Max-pooling and upsampling operations For each network level, we replaced the down-convolution (convolution with stride 2) by a max-pooling operation in `MIRRBA_Max`, and the transpose convolution by an upsampling in `MIRRBA_Up`. As presented Table 3, using these learning-free operations improved the registration accuracy globally and locally. The disappearing rate was however slightly reduced.

Residual blocks We removed the residual connections of the residual blocks (`MIRRBA_wo_RB`) to understand their influence in our architecture. Without them, we obtained null Dice scores and detection rate, as well as very high `SDJDet` value. Besides, wrapped images did not look realistic (see Fig. 2 in the supplementary materials).

Deformable convolutions To adapt the receptive field of the convolutions to the local scale of objects to be registered, we replaced those of the highest resolution level (i.e. level 3) by deformable convolutions (`MIRRBA_DefConv`). While global registration results were improved over `MIRRBA`, local ones were not. Moreover, the `SDJDet` was low.

Table 4. Comparison of MIRRBA to conventional methods (Section 5.3) – Detection rate, Dice score of the organs and lesions (Dice_organ and Dice_lesions resp.), SDJDet, disappearing rate, and approximate computational time. While ANTs and Elastix pipeline were computed on CPU, MIRRBA ran on GPU. Statistically significant improvement of our MIRRBA method over the others with $p < 0.05$ is indicated with *. Best results are marked in bold, and second best ones in bold-italic, except for SDJDet since no ideal value is defined.

	Detec. rate (%) \uparrow	Dice organ \uparrow	Dice lesions \uparrow	SDJDet \downarrow	Disap. rate (%) \uparrow	Time (min) \downarrow
Identity	5.00	0.626 \pm 0.138*	0.090 \pm 0.115*	0.000 \pm 0.000	0.00*	0
MIRRBA	33.04	0.918 \pm 0.126	0.425 \pm 0.207	0.124 \pm 0.988	9.36	55
Elastix	20.54	0.868 \pm 0.124*	0.350 \pm 0.191*	0.096 \pm 0.044	9.59	25
SyN	24.57	0.936 \pm 0.023	0.386 \pm 0.210*	0.016 \pm 0.018	0.00	5
SyNCC	39.57	0.944 \pm 0.014	0.477 \pm 0.211	0.073 \pm 0.066	4.26*	60
Combined	44.71	0.945 \pm 0.012	0.481 \pm 0.197	0.077 \pm 0.072	25.11	115

Input images Regarding the network inputs, instead of conditioning the network with the moving image, we fed it with a Gaussian noise, as in [Laves et al., 2019]. In Table 3, MIRRBA_NoiseImg shows worse results than MIRRBA, except for the SDJDet. We also provided more information to the network by concatenating the fixed to the moving image in MIRRBA_FixImg, which significantly improved all MIRRBA results, especially the disappearing rate.

Combining best practices Finally, we combined the best architectural variations presented above to perform registration with four resolutions, both fixed and moving images as input, max-pooling and upsampling operations, as well as residual blocks: MIRRBA_Best.

MIRRBA_Depth_4 Dice scores were slightly improved by the use of max-pooling and upsampling operations. On the other hand, adding a fourth depth to MIRRBA_FixImg, MIRRBA_Max and MIRRBA_Up statistically improved their results, leading to the best performing MIRRBA-based method in terms of local registration precision MIRRBA_Best. The SDJDet of MIRRBA_Best was smaller or similar to either one of the three other methods, while the disappearing rate lied between their values. Images of lesions registered with MIRRBA_Best are visible Fig. 3.

5.3. Comparison to conventional methods

Conventional registration According to Table 4, even after rigid and affine pre-registration, locally precise deformable registration is challenging on whole-body images for the conventional Elastix ([Klein et al., 2010]) pipeline (see Fig. 1 in supplementary material).

On the other hand, both SyN and SyNCC ([Avants et al., 2009]) statistically

performed better than MIRRBA (which used the NCC metric) on organ segmentation, while the NCC similarity metric allowed SyNCC to also reach a better accuracy on lesion segmentation. Regarding the disappearing rate, MIRRBA performed better than both SyN-based methods, whereas their SDJDet was lower than ours.

Combination of DIP and conventional registration Finally, we pushed the analysis by combining our best MIRRBA-based method with the best performing conventional SyNCC approach. To do so, we optimized MIRRBA_Best, using as input the deformation fields and already registered images obtained by SyNCC. Our assumption was that the conventional pre-registration would be improved by our method.

Results of this Combined pipeline are shown in Table 4. In terms of global and local accuracy, the combined approach outperformed both SyNCC and MIRRBA_Best. In addition, the disappearing rate was significantly improved for both methods.

6. Discussion

6.1. Impact of training on a database

When comparing MIRRBA to the DL-based approaches (DL_LapIRN and DL_VoxelMorph), we show that not learning registration patterns from a dataset helps to obtain precise segmentations, especially at a lesion level. Indeed, while the size of organs and their locations are relatively consistent across patients, this is not the case with the lesions. Hence, there are fewer deformation patterns that can be learned from a database. Since performing locally precise registration with a DL-based method is very challenging in this situation, our method adapts to each subject. Moreover, according to its lower SDJDet value, MIRRBA produces smoother deformations than the DL-based approaches.

6.2. Architectural choices

Regularization terms As presented in Section 3.1, removing the registration-specific regularization terms is equivalent to solving an ill-posed problem in conventional registration methods. In our case, since we use a network to perform registration, \mathcal{R}_{archi} is present when running MIRRBA_wo_Regu. For similar Dice scores and detection rate, MIRRBA_wo_Regu has higher SDJDet than MIRRBA, indicating \mathcal{R}_{smooth} and \mathcal{R}_{diffeo} help smooth the deformation field, even though they have less impact than the regularization of the architecture.

On the opposite configuration MIRRBA_wo_Archi, where \mathcal{R}_{archi} is not used but \mathcal{R}_{smooth} and \mathcal{R}_{diffeo} are, quantitative results show that the lack of architecture negatively impacts the registration smoothness, as well as its local and global precision, making questionable the convergence of the method without architecture. Hence, we confirm that the architecture has a regularization effect on the registration, which helps to find an admissible solution.

Coarse to fine strategies and image resolution Considering the overall network, [Dittmer et al., 2020] affirms that an architecture ran in a DIP pipeline acts as a low-pass filter in the beginning of the optimization, allowing higher frequencies to pass only after lower ones. We observe similar results over the coarse-to-fine training strategy (see Fig. 2), where low frequencies are registered first (global registration), followed later by higher frequencies (local registration). In the same way, looking at the intermediate results during the 4-depth pyramidal optimization of `Elastix` (see Fig. 1 in supplementary materials), we observe a tendency to register global features before local ones.

Modifying the number of coarse resolution levels, i.e. the depth of the pyramidal architecture, we studied the impact of global structural choices to optimize the whole network. The higher `SDJDet` values obtained when more resolutions are used could be explained by the generation of more local transformations, hence a globally less regular deformation field. Moreover, with four resolutions, the receptive field of the coarsest level of `MIRRBA_Depth_4` captures the whole image (i.e. $200 \times 200 \times 200$), explaining the high Dice scores, as well as detection and disappearing rates. Indeed, successful conventional pipelines as `Elastix` ([Klein et al., 2010]) or ANTs ([Avants et al., 2009]) also uses this type of pyramidal strategy.

Network conditioning with input images Contrary to [Ulyanov et al., 2020] and [Laves et al., 2019], who used Gaussian noise as input for their architectures, [Gong et al., 2019] and [Bager et al., 2020] respectively improved CT and PET DIP-based image reconstructions by providing acquisitions from other modalities to their networks. We made similar observations, as `MIRRBA_NoiseImg` reached less accurate results than when we conditioned the model with PET images, as in `MIRRBA`. Moreover, feeding the network with more patient information, using the fixed image as additional input to the moving image in `MIRRBA_FixImg`, improves the results by increasing the network conditioning on a single patient.

Other architectural choices Moving inside the architecture, residual blocks can be related to diffeomorphic registration according to [Rousseau et al., 2020]. Indeed, stacking residual blocks in ResNets ([He et al., 2016]) aims to incrementally map the embedding space to a new unknown space, each block being defined as $y = F(x) + x$, with x and y the respective input and output of the residual blocks, and F the residual mapping to be learned. Similarly, diffeomorphic registration models ([Beg et al., 2005, Sotiras et al., 2013]) address the registration issue by piling up incremental diffeomorphic mappings. Making the link between ResNets and registration, the function F can be seen as a parametrization of an elementary deformation flow, and training a series of residual blocks as learning continuous and integral diffeomorphic operator. With our deep architecture, `MIRRBA_wo_RB` results indicate that the registration without residual blocks fails to converge. Indeed, as explained above, residual blocks allow incremental diffeomorphic mappings, and removing them leads to

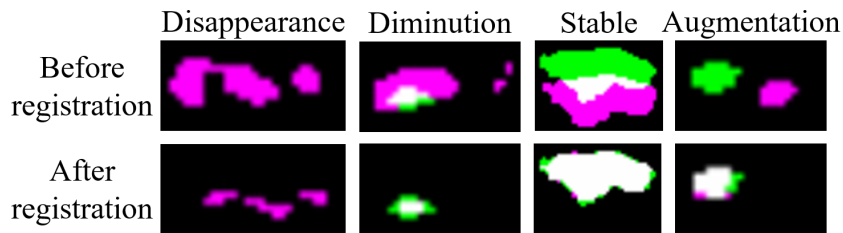


Figure 3. Overlay of lesions disappearing (1st column), reducing (2nd column), stable (3rd column) and growing (last column) before and after performing *MIRRBA_Best*. Fixed lesions can be viewed in green, while moving and wrapped ones in pink. White color indicates overlapping. Best viewed in color.

gradient vanishing for our patient-specific method.

Regarding the convolutions, [Heinrich, 2019] suggested deformable convolutions ([Dai et al., 2017]) to capture larger deformations. These convolutions add 2D or 3D offsets to the regular grid sampling of standard operations. If these offsets are set to zero, deformable convolutions become standard convolutions, otherwise they modify the receptive field. Since these offsets are learnable, deformable convolutions are trained to adapt their receptive fields, in order to focus on objects of interest in classification problems for instance. According to the high detection rate, this is performed globally by *MIRRBA_DefConv*. Yet, the resulting low Dice scores show that *MIRRBA_DefConv* did not achieve precise local registration, which can be explained by the low *SDJDet* value, showing that only smooth and regular deformations occurred instead of locally irregular ones. Indeed, deformable convolutions might need a more complex integration in the architecture ([Liu et al., 2020]). Moreover, the additional learnable parameters and adaptive receptive field of deformable convolutions are to some extent redundant with the pyramidal structure, and make it harder to train.

Considering the amount of parameters to optimize in a network, we looked at the results of *MIRRBA_Max* and *MIRRBA_Up* compared to *MIRRBA*. Even if the learnable down- and up-convolutions are now common to respectively increase or decrease image dimensions within a network, max-pooling and upsampling operations were originally used. As in conventional registration methods, these operations are not learnable. Hence, their results do not depend on optimization parameters, and they help to control the overfitting in a traditional trainable setup. Therefore, the good results obtained using max-pooling and upsampling operations in *MIRRBA_Max*, *MIRRBA_Up* and *MIRRBA_Depth_4_Max_Up* may be explained by the fewer number of parameters to fit (see Fig. 2 in supplementary materials).

6.3. Lesion segmentation

Regarding lesion registration, Fig. 3 shows different lesion evolution scenarios. We can see that the moving masks adapt to the fixed ones when the lesion does not disappear. Although the disappearing lesion from the first column was not deleted, it was reduced

by the registration algorithm. Indeed, lesion disappearance implies deformations which are not smooth nor diffeomorphic, hence the good disappearing rate of `MIRRBBA_wo_Regu` and the probable need to adapt our registration strategy to the specific problem of lesion disappearance in future work.

7. Conclusion

In this paper, we propose an alternative method to perform image registration using a neural network without the typical learning stage on a database. We formalize the registration problem by following the conventional approaches relying on image-based similarities and regularization terms, but also explicitly consider the dataset and architecture bias. Indeed, our study was motivated by recent work on DIP, implying that neural networks create an inductive bias when learning from a database, but also an intrinsic structural bias induced by the architecture [Heckel and Soltanolkotabi, 2020]. These biases are shown sufficient to solve certain image processing tasks. Our proposition also resonates with [Dittmer et al., 2020], who suggested that deep convolutional neural networks process low-frequency information first, to later focus on the finer deformations, both desirable properties for a registration algorithm. We integrate the LapIRN network from [Mok and Chung, 2020], who developed an effective pyramidal architecture, tested in the DL-based set-up. Here, we further demonstrate that beyond any prior coming from the dataset learning step, the architecture design has an important effect on the registration results, acting as an implicit regularizer. Our study also shows the impact of some of the architecture components, particularly the residual blocks, and we justify this behavior by making a link with findings from [Rousseau et al., 2020]. Moreover, we found that for our application, a pyramidal architecture capturing the whole image with a limited amount of parameters to optimize, as conventional registration methods, provides precise registration results.

The architectural prior seems to be a better option than learning from data in cases where there are no positional consistency, which is the case with metastatic breast cancer lesions, which arbitrarily vary in position, size and number. Indeed, finding a set of network parameters allowing precise registration for a whole dataset is a challenging task. Our approach enables to correctly register active organs such as the brain and the bladder, which could be used to automatically propagate annotations masking regions irrelevant for patient response evaluation. Although the Dice scores are relatively low for the lesions, we obtain good detection values and improve the disappearance rate.

Our approach makes a step in bridging conventional and DL-based methods for image registration, and provides several suitable approaches for the challenging 3D full-body longitudinal registration problem. We demonstrated the possibility to perform both global and local registration on whole body medical images using a network but without suffering from dataset bias. In future work, we would study the feasibility of extracting registration-based feature from our method to monitor lesion evolution

without depending on manually performed segmentations.

Acknowledgments

This work is partially financed through xxxx.

References

- [Avants et al., 2009] Avants, B. B., Tustison, N., and Johnson, H. (2009). Advanced Normalization Tools (ANTs). *Insight Journal*, 2(365):1–35.
- [Avril et al., 2016] Avril, S., Muzic, R. F., Plecha, D., Traughber, B. J., Vinayak, S., and Avril, N. (2016). 18F-FDG PET/CT for Monitoring of Treatment Response in Breast Cancer. *Journal of Nuclear Medicine*, 57(Supplement 1):34S–39S.
- [Baguer et al., 2020] Baguer, D. O., Leuschner, J., and Schmidt, M. (2020). Computed tomography reconstruction using deep image prior and learned reconstruction methods. *Inverse Problems*, 36.
- [Balakrishnan et al., 2018] Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., and Dalca, A. V. (2018). VoxelMorph: A Learning Framework for Deformable Medical Image Registration. *IEEE Transactions on Medical Imaging*, 38(8):1–13.
- [Beg et al., 2005] Beg, M. F., Miller, M. I., Trounev, A., and Younes, L. (2005). Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*, 61(2):139–157.
- [Carlier and Bailly, 2015] Carlier, T. and Bailly, C. (2015). State-Of-The-Art and Recent Advances in Quantification for Therapeutic Follow-Up in Oncology Using PET. *Frontiers in Medicine*, 2.
- [Chassagnon et al., 2020] Chassagnon, G., Vakalopoulou, M., Régent, A., Sahasrabudhe, M., Marini, R., Hoang-Thi, T.-N., Dinh-Xuan, A.-T., Dunogué, B., Mouthon, L., Paragios, N., and Revel, M.-P. (2020). Elastic Registration – driven Deep Learning for Longitudinal Assessment of Systemic Sclerosis. *Radiology*, 298(1):189–198.
- [Chen et al., 2021] Chen, X., Diaz-pinto, A., Ravikumar, N., and Frangi, A. F. (2021). Deep learning in medical image registration. *Progress in Biomedical Engineering*, 3(1).
- [Christensen et al., 2007] Christensen, G. E., Song, J. H., Lu, W., Naqa, I. E., and Low, D. A. (2007). Tracking lung tissue motion and expansion/compression with inverse consistent image registration and spirometry. *Medical Physics*, 34(6):2155–2163.
- [Dai et al., 2017] Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., and Wei, Y. (2017). Deformable Convolutional Networks. *Proceedings of the IEEE International Conference on Computer Vision*, pages 764–773.
- [Dalca et al., 2019] Dalca, A. V., Balakrishnan, G., Guttag, J., and Sabuncu, M. R. (2019). Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Medical Image Analysis*, 57:226–236.
- [de Vos et al., 2019] de Vos, B. D., Berendsen, F. F., Viergever, M. A., Sokooti, H., Staring, M., and Išgum, I. (2019). A deep learning framework for unsupervised affine and deformable image registration. *Medical Image Analysis*, 52:128–143.
- [Dittmer et al., 2020] Dittmer, S., Kluth, T., Maass, P., and Baguer, D. O. (2020). Regularization by architecture : A deep prior approach for inverse problems. *Journal of Mathematical Imaging and Vision*, 62(3):456–470.
- [Eppenhof et al., 2019] Eppenhof, K. A., Lafarge, M. W., Veta, M., and Pluim, J. P. (2019). Progressively Trained Convolutional Neural Networks for Deformable Image Registration. *IEEE Transactions on Medical Imaging*, 39(5):1594–1604.
- [Friston et al., 1995] Friston, K. J., Ashburner, J., Frith, C. D., Poline, J. B., Heather, J. D., and Frackowiak, R. S. (1995). Spatial registration and normalization of images. *Human Brain Mapping*, 3(3):165–189.

- [Fu et al., 2020] Fu, Y., Lei, Y., Wang, T., Curran, W. J., Liu, T., and Yang, X. (2020). Deep learning in medical image registration: a review. *Physics in Medicine and Biology*, 60(20):1–32.
- [Gong et al., 2019] Gong, K., Catana, C., Qi, J., and Li, Q. (2019). PET Image Reconstruction Using Deep Image Prior. *IEEE Transactions on Medical Imaging*, 38(7):1655–1665.
- [He et al., 2016] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Identity Mappings in Deep Residual Networks. *European conference on computer vision*, pages 630–645.
- [Heckel and Soltanolkotabi, 2020] Heckel, R. and Soltanolkotabi, M. (2020). Compressive sensing with un-trained neural networks: Gradient descent finds the smoothest approximation. *International Conference on Machine Learning*.
- [Heinrich, 2019] Heinrich, M. P. (2019). Closing the Gap Between Deep and Conventional Image Registration Using Probabilistic Dense Displacement Networks. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, October:50–58.
- [Heinrich and Hansen, 2020] Heinrich, M. P. and Hansen, L. (2020). Highly accurate and memory efficient unsupervised learning-based discrete CT registration using 2.5 D displacement search. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 190–200.
- [Hsu, 2020] Hsu, T.-M. H. (2020). *Automatic Longitudinal Assessment of Tumor Responses*. PhD thesis, Massachusetts Institute of Technology.
- [Jaderberg et al., 2015] Jaderberg, M., Simonyan, K., Zisserman, A., and Kavukcuoglu, K. (2015). Spatial Transformer Networks. *Advances in neural information processing systems*, 28:2017–2025.
- [Kim et al., 1994] Kim, C. K., Gupta, N. C., Chandramouli, B., and Alavi, A. (1994). Standardized uptake values of FDG: Body surface area correction is preferable to body weight correction. *Journal of Nuclear Medicine*, 35(1):164–167.
- [Klein et al., 2009a] Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M. C., Christensen, G. E., Collins, D. L., Gee, J., Hellier, P., Song, J. H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R. P., Mann, J. J., and Parsey, R. V. (2009a). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46(3):786–802.
- [Klein et al., 2009b] Klein, S., Pluim, J. P., Staring, M., and Viergever, M. A. (2009b). Adaptive stochastic gradient descent optimisation for image registration. *International Journal of Computer Vision*, 81(3):227–239.
- [Klein et al., 2010] Klein, S., Staring, M., Murphy, K., Viergever, M., and Pluim, J. (2010). elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205.
- [Laves et al., 2019] Laves, M. H., Ihler, S., and Ortmaier, T. (2019). Deformable medical image registration using a randomly-initialized CNN as regularization prior. *arXiv*, 1:1–6.
- [Li et al., 2021] Li, T., Zhang, M., Qi, W., Asma, E., and Qi, J. (2021). Motion Correction of Respiratory-Gated PET Images Using Deep Learning Based Image Registration Framework. *Physics in Medicine and Biology*, 65(15):155003.
- [Liu et al., 2020] Liu, F., Liu, D., Tian, J., Xie, X., Yang, X., and Wang, K. (2020). Cascaded one-shot deformable convolutional neural networks: Developing a deep learning model for respiratory motion estimation in ultrasound sequences. *Medical Image Analysis*, 65.
- [Lucas et al., 2018] Lucas, A., Iliadis, M., and Molina, R. (2018). Using Deep Neural Networks for Inverse Problems in Imaging: Beyond Analytical Methods. *IEEE Signal Processing Magazine*, 35(1):20–36.
- [Maurer and Fitzpatrick, 1993] Maurer, C. and Fitzpatrick, J. (1993). A Review of Medical Image Registration. *Interactive Image-Guided Neurosurgery*, pages 17–44.
- [Mok and Chung, 2020] Mok, T. C. W. and Chung, A. C. S. (2020). Large Deformation Diffeomorphic Image Registration with Laplacian Pyramid Networks. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 211–221.
- [Moreau et al., 2020] Moreau, N., Rousseau, C., Fourcade, C., Santini, G., Ferrer, L., Lacombe, M.,

- Guillerminet, C., Campone, M., Colombie, M., Rubeaux, M., and Normand, N. (2020). Deep learning approaches for bone and bone lesion segmentation on 18FDG PET/CT imaging in the context of metastatic breast cancer. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2020-July:1532–1535.
- [Myronenko and Song, 2009] Myronenko, A. and Song, X. (2009). Adaptive Regularization of Ill-Posed Problems: Application to Non-rigid Image Registration. *arXiv preprint arXiv:0906.3323*, pages 1–10.
- [Necib et al., 2011] Necib, H., Garcia, C., Wagner, A., Vanderlinden, B., Emonts, P., Hendlitz, A., Flamen, P., and Buvat, I. (2011). Detection and Characterization of Tumor Changes in 18F-FDG PET Patient Monitoring Using Parametric Imaging. *Journal of Nuclear Medicine*, 52(3):354–361.
- [Paszke et al., 2017] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. (2017). Automatic differentiation in PyTorch. *31st Conference on Neural Information Processing System*.
- [Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 9351:234–241.
- [Rousseau et al., 2020] Rousseau, F., Drumetz, L., and Fablet, R. (2020). Residual Networks as Flows of Diffeomorphisms. *Journal of Mathematical Imaging and Vision*, 62(3):365–375.
- [Rueckert et al., 1999] Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O., and Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: Application to breast mr images. *IEEE Transactions on Medical Imaging*, 18(8):712–721.
- [Sotiras et al., 2013] Sotiras, A., Davatzikos, C., and Paragios, N. (2013). Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190.
- [Stergios et al., 2018] Stergios, C., Mihir, S., Maria, V., Guillaume, C., Marie-Pierre, R., Stavroula, M., and Nikos, P. (2018). Linear and deformable image registration with 3D convolutional neural networks. *Image Analysis for Moving Organ, Breast, and Thoracic Images*, pages 13–22.
- [Ulyanov et al., 2020] Ulyanov, D., Vedaldi, A., and Lempitsky, V. (2020). Deep Image Prior. *International Journal of Computer Vision*, 128(7):1867–1888.
- [Vercauteren et al., 2007a] Vercauteren, T., Pennec, X., Malis, E., Perchant, A., and Ayach, N. (2007a). Insight into efficient image registration techniques and the demons algorithm. *Biennial International Conference on Information Processing in Medical Imaging*, July:495–506.
- [Vercauteren et al., 2007b] Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2007b). Non-parametric diffeomorphic image registration with the demons algorithm. *Medical Image Computing and Computer-Assisted Intervention. International Conference on Medical Image Computing and Computer-Assisted Intervention*, 792:319–326.

Supplementary Material

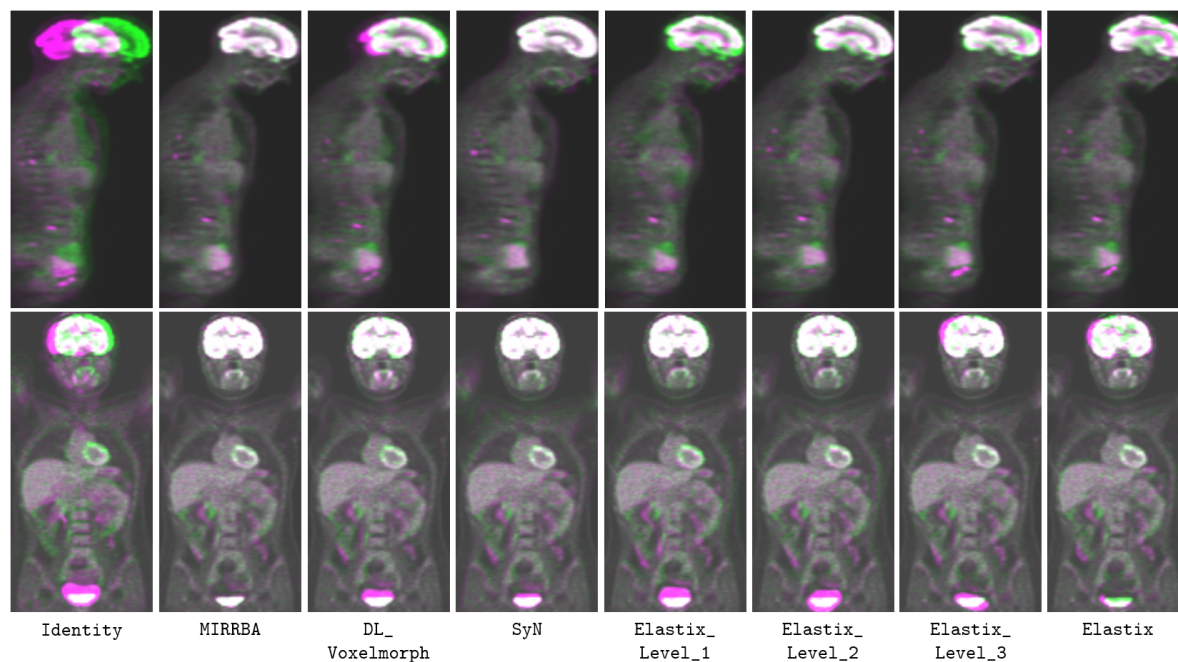


Figure 1. Comparison to reference methods – Overlay of the fixed (green) and warped (pink) images on two different patients after performing *Identity* registration, *MIRRBA*, *DL_Voxelmorph*, *SyN*, *Elastix_Level_1*, *Elastix_Level_2*, *Elastix_Level_3* and *Elastix* (corresponding to the level 4). Grayscale color indicates good overlapping. DL-based method has difficulties to register the bladder because of its important deformation. *SyN* wrapped images look coherent, even if missing a bit of precision around the bladder. The pyramidal optimization of *Elastix* acts as a progressive registration: global features are registered before more local and precise ones. Best viewed in color.

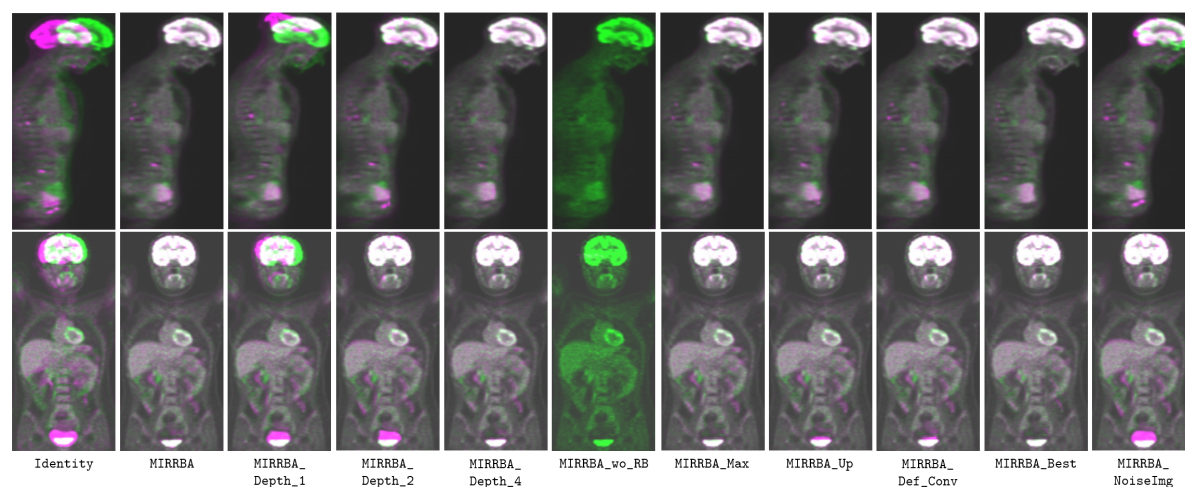


Figure 2. Ablation study – Overlay of the fixed (green) and warped (pink) images on two different patients after performing `Identity` registration, `MIRRBA` (corresponding to depth 3), `MIRRBA_Depth_1`, `MIRRBA_Depth_2`, `MIRRBA_Depth_4`, `MIRRBA_wo_RB`, `MIRRBA_Max`, `MIRRBA_Up`, `MIRRBA_DefConv`, `MIRRBA_Best` and `MIRRBA_NoiseImg`. Grayscale color indicates good overlapping. We can see that the higher the depth, the more precise the registration. A simple U-Net-shaped architecture as `MIRRBA_Depth_1` produces transformations of very low accuracy, while the more resolutions are used, the more precise the registration. The all green `MIRRBA_wo_RB` image is due to the non-convergence of the registration algorithm and a wrapped image not registered to the fixed one. All four other approaches produce realistic looking and coherent transformations, even if `MIRRBA_DefConv` lacks a bit of precision around the bladder. `MIRRBA_NoiseImg` does not reach local precise registration. Best viewed in color.