Deformable Image Registration with Deep Network Priors: a Study on Longitudinal PET Images

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

Medical Image Analysis (2021)

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Deformable Image Registration with Deep Network Priors: a Study on Longitudinal PET Images

Constance Fourcade[a,b,*], Ludovic Ferrer[d,e], Noemie Moreau[b], Gianmarco Santini[b], Aishlinn Brennan[b], Caroline Rousseau[c,e], Marie Lacombe[e], Vincent Fleury[e], Mathilde Colombié[e], Pascal Jézéquel[d,e], Mario Campone[c,e], Mathieu Rubeaux[b], Diana Mateus[a]

[a] *Ecole Centrale de Nantes, LS2N, UMR CNRS 6004, Nantes, France*
[b] *Keosys Medical Imaging, Saint Herblain, France*
[c] *University of Nantes, CRCINA, INSERM UMR1232, CNRS-ERL6001, Nantes, France*
[d] *University of Angers, CRCINA, INSERM UMR1232, CNRS-ERL6001, Angers, France*
[e] *ICO Cancer Center, Nantes - Angers, France*

ARTICLE INFO

Article history:
Received xxx
Received in final form xxx
Accepted xxx
Available online xxx

Communicated by xxx

2000 MSC: 41A05, 41A10, 65D05, 65D17

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

ABSTRACT

Longitudinal image registration is challenging, and unlike other medical image analysis tasks, has not yet benefited from major performance improvements thanks to deep-learning. Inspired by Deep Image Prior (DIP), this paper introduces a different use of deep architectures as regularizers to tackle the image registration question. We propose a subject-specific deformable registration method called `MIRRBA`, which relies on a deep pyramidal architecture to be the prior parametric model constraining the deformation field. Diverging from the supervised learning paradigm, `MIRRBA` does not require a learning database, but only the pair of images to be registered in order to optimize the network's parameters and provide a deformation field. With a series of experiments, we demonstrate the regularizing power of deep architectures and present new elements to understand the role of the architecture in deep learning methods for registration. Hence, to study the impact of the network parameters, we ran our method with 12 different architectural configurations on a private dataset of 110 metastatic breast cancer full-body PET images with manual segmentations of the brain, bladder and metastatic lesions. We compared it against several baselines: conventional iterative registration approaches (ANTs and Elastix) and supervised deep learning-based models (LapIRN and Voxelmorph). Global and local registration accuracies were evaluated using the detection rate and the Dice score respectively, while registration realism was evaluated using the Jacobian's determinant. Moreover, we computed the ability of the different methods to shrink vanishing lesions with the disappearing rate. `MIRRBA` significantly improves the organ and lesion Dice scores of Voxelmorph by 6% and 52%, and of LapIRN by 5% and 65% respectively. Regarding the disappearing rate, `MIRRBA` more than doubles the best performing conventional approach `SyNCC` score. Our work therefore proposes an alternative way to bridge the performance gap between conventional and deep learning-based methods, and demonstrates the regularizing power of deep architectures.

1. Introduction

Image registration is a long-standing fundamental problem in medical analysis (Maurer and Fitzpatrick (1993), Chen et al.

*Corresponding author:
e-mail: `constance.fourcade@ec-nantes.fr` (Constance Fourcade)

(2021)). The precise overlaying of a fixed and a moving image can contribute to create anatomical or functional patient models (*inter-patient monomodal registration*), better exploit information coming from several modalities for a single patient (*intra-patient multimodal registration*), as well as follow a disease evolution (*intra-patient monomodal registration*). In this paper, we are interested in intra-patient monomodal PET registration for the evaluation of treatment response in metastatic breast cancer. While the low spatial resolution of PET acquisitions is usually addressed by jointly registering CT and PET images in multimodal pipelines (Fu et al. (2020)), we chose to register PET images alone as metastatic lesions are not visible on non contrasted CT images taken at the same time as PET. Although the scale difference between a patient's whole body and local lesions is challenging, we aim to precisely register lesions of various locations and sizes.

After nearly 30 years of research, there seems to be a convergence towards methods treating registration as an optimization problem aiming to minimize a dissimilarity term between moving and fixed images. This dissimilarity term can be based on mean square error (MSE), mutual information (MI) or normalized cross-correlation (NCC) for instance. A term enforcing smooth and realistic transformations is often used as a regularizing term. In this paper, we refer to such methods as "conventional", as opposed to deep learning-based (DL) approaches. Conventional methods cope with the ill-posed problem by parametrizing 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)). After defining the model, its parameters are typically optimized though efficient second-order minimization (Vercauteren et al. (2007a)) or stochastic gradient descent (SGD) (Klein et al. (2009b)), within a single- or multi-level optimization strategy. The minimized cost function combines dissimilarity and regularization terms. Recent deep-learning supervised methods for registration, e.g. Balakrishnan et al. (2018), de Vos et al. (2019), rely on the same dissimilarity measures and regularization terms as conventional methods, and can also be optimized with variants of SGD. The main difference with conventional approaches are the training stage which brings prior knowledge about the deformations from training databases, and the type of transformation model, with CNNs being an over-parameterized but structured model of the deformation field.

Despite recent developments in DL-based registration (Chen et al. (2021)), conventional methods still reach more accurate results in most applications (Heinrich and Hansen (2020)). Though trained networks are faster than conventional methods, training requires large databases and often restricts their use to a specific domain. To alleviate the database dependency and understand the architecture's influence and role as a prior, we propose an unsupervised approach to perform registration based on the recent Deep Image Prior work (DIP) (Ulyanov et al. (2020)). 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 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.

In this work, we formalize the registration problem as a conventional registration approach with a deformation field modeled by a deep pyramidal network. Hence, we name our approach *MIRRBA* for *Medical Image Registration method Regularized By Architecture*. Contrary to Ulyanov et al. (2020), who fed the network with white noise, we gave the moving image as input, conditioning the network's weights to parametrize the deformation field and thus favor interpretability.

Since `MIRRBA` optimizes a network while not being biased by the training dataset, we are able to isolate the role of different network components to accurately predict deformation fields. We show that some settings, such as the network's global shape or capture range, significantly impact results. Regarding reference methods, `MIRRBA` outperforms state-of-the-art DL-based approaches, despite our optimization not competing with their inference time. `MIRRBA` does not however yet surpass the best conventional approaches. In fact, our best results were obtained by relying on a conventional method refined through `MIRRBA`. Aside `MIRRBA`'s comparable performance to state-of-the-art methods, and despite its computational limitations, the focus of our work is in the paradigm change bridging the gap between conventional and DL-based methods.

In summary, 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 globally and locally without any prior registration to facilitate the simultaneous monitoring of multiple lesions.

2. Related work

Automatic longitudinal lesion monitoring. To monitor cancer evolution and assess response after completion of therapy, several automatic approaches based on segmentation and/or registration have been developed. Necib et al. (2011) used affine registration to align initial and follow-up PET images, before subtracting them to identify tumor voxels with significant changes between the two scans. Yet, since affine registration is a global method, it mainly performs well for single localized tumors. To deal with multiple lesion cases, 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 the stages. Removing the need of a prior segmentation step, Chassagnon et al. (2020) relies on a conventional registration algorithm between longitudinal CT pairs to obtain the deformation fields and their Jacobian determinants. This information is then used to train a classifier network assessing systemic sclerosis interstitial lung disease. In the same way, we aim at using registration-based features to

monitor metastatic breast cancer. Hence, we present hereafter the most commonly used registration tools and good practices which contribute to make these approaches successful.

Established registration tools. There is a large body of literature on conventional registration methods (Klein et al. (2009a); Sotiras et al. (2013)). Some of today's most commonly used registration tools include the *Demons* pipeline (Vercauteren et al. (2007b)) 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. For our comparisons, we focused on the last two methods, since they performed well among different datasets (Klein et al. (2009a)) and their pyramidal coarse-to-fine optimization inspired recent works on DL-based registration, presented hereafter.

DL methods. In recent years, different types of DL-based registration approaches have been proposed (Chen et al. (2021)). Since for the monitoring of metastatic breast cancer deformation fields are not available and difficult to obtain, especially in the case of multiple lesions, we focused on unsupervised registration methods. We categorize as unsupervised those methods which require neither deformation fields, nor segmentations or landmarks. Among them, various network 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 to perform affine and deformable registrations in stages, at the cost of an increase in complexity. 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[1] 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, no major performance gain has been reached with DL-based registration techniques so far, and conventional iterative optimization methods still yield better accuracy in tasks such as inter-patient alignment or intra-patient lung motion registration (Heinrich and Hansen (2020)). Indeed, large databases are needed to learn the network parameters and produce accurate deformation fields. Moreover, we argue that the generalization ability of the trained network is questionable when the deformation patterns do not repeat consistently across the dataset.

[1] https://learn2reg.grand-challenge.org/

DIP registration. To alleviate the dataset dependency and the need of repeated deformation patterns for DL-based approaches, we adapted the DIP framework (Ulyanov et al. (2020)) and propose a learning-free method for deformable medical image registration. Our method uses the structure of an untrained convolutional network as regularization prior.

DIP has been used on medical data as a prior for the reconstruction of CT and PET images (Gong et al. (2019); Baguer et al. (2020)). In both cases, images from other modalities have been given as input to condition the network output. Similar to our idea, a preliminary study by Laves et al. (2019) combined registration and DIP, generating a deformation field to register 2D brain MR images from a white noise image using Ulyanov et al. (2020)'s pipeline and network architecture. Instead, we relied on a top performing pyramidal architecture (Mok and Chung (2020)), learnt weights conditional to the moving image, and dealt with the registration of longitudinal full-body 3D PET images. More importantly, we derived the idea by making a parallel between conventional and DL approaches. We also interpreted the architecture as an over-parameterization of the transformation, this structure imposing an implicit regularization through the network shape alone. To support this argument, we pointed to recent works in DL for inverse problems Lucas et al. (2018) and Dittmer et al. (2020), and provided an in-depth experimental study of the key architecture and loss components. We also reported comparisons to two well-established conventional and DL approaches and proposed a strategy that brings together the top-performing methods from the two worlds. To the best of our knowledge, we are the first to perform deformable registration on 3D whole-body PET images.

3. Method

3.1. Image registration

Given a fixed and moving (F, M) image pair, 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 $\mathcal{S}(.,.)$ measuring the dissimilarity between two images, conventional registration methods choose a parametrization of transformation and optimize it through a cost function of the form:

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

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. In this way, complex deformable transformations include a regularization term $\mathcal{R}_{smooth}(\Phi)$ in the objective function Eq. 1. This prior, correctly chosen, restricts the admissible solutions by avoiding meaningless transformations and shaping the type of deformation fields. Weighted by λ_{smooth}, common regularization terms enforce smoothness onto the displacement field by penalizing

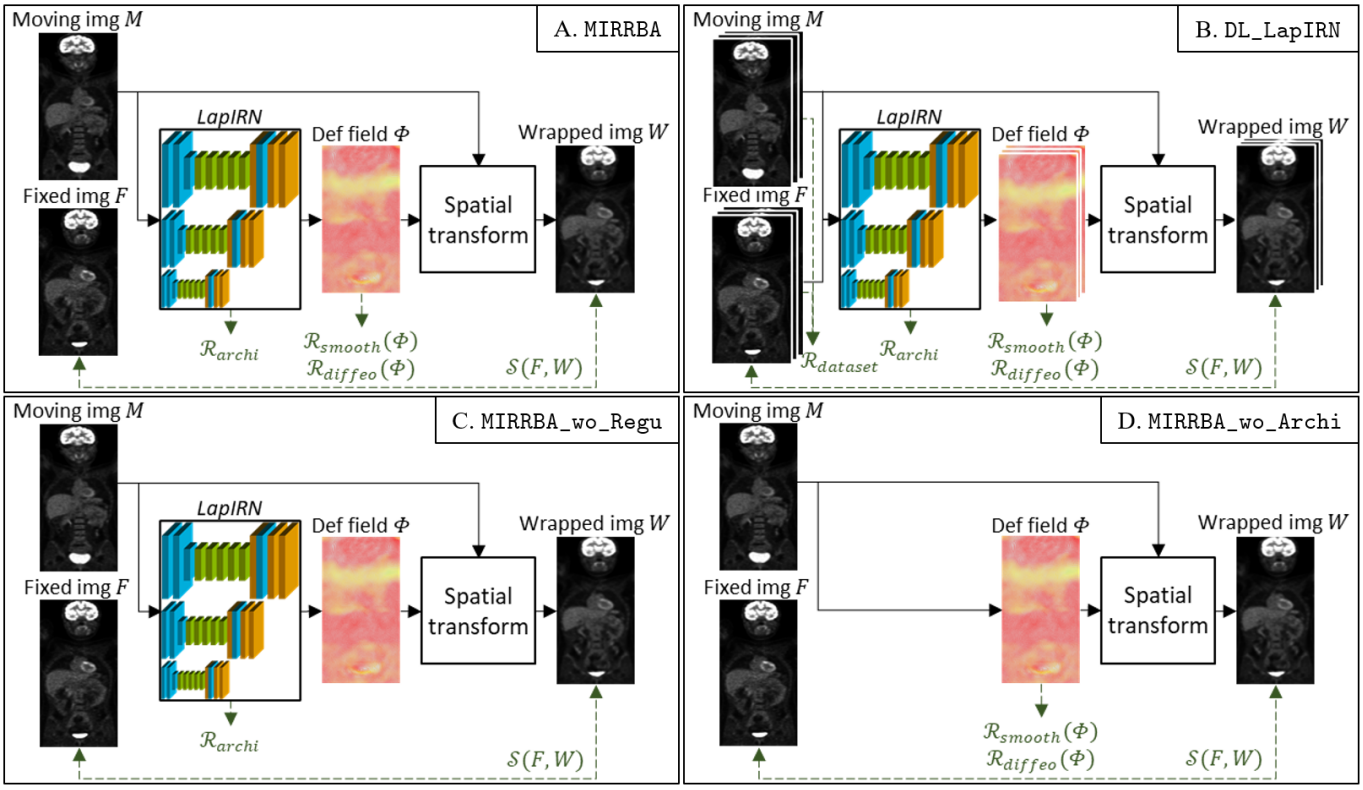

Fig. 1. Overview of A. `MIRRBA`, B. `DL_LapIRN` and C. `MIRRBA_wo_Regul` and D. `MIRRBA_wo_Archi` methods. The LapIRN architecture is visible with the encoder path (blue), the residual blocks (green) and the decoder path (orange with a blue layer from the encoder path). Best viewed in color.

the spatial derivatives of Φ.

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

When using DL-based registration approaches, the cost function Eq. 3, adapted from Eq. 2, is used to train a CNN to predict a dense deformation field Φ. 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)), acting in our case as a bilinear interpolator. Regarding the regularization, it is common to continue relying on a term enforcing the smoothness of the transformation $\mathcal{R}_{smooth}$, as done in conventional approaches. It 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 regularization priors are also implicitly added in the DL setup: 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).

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

We then revised the deep image prior concept in Section 3.2 and adapted it to the image registration problem 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.

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

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`

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 regularization priors for image registration in iterative optimization schemes without a training stage.

In a trainable setup, the fixed and moving pair (F, M) is given as input to the network (Eq. 3). Following the DIP principle (Ulyanov et al. (2020)), we propose to let the deformation field

parametrization depends only upon the moving image M, and indirectly through the loss function upon the fixed image F. 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).

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

By optimizing Eq. 5, we find the best warped image allowed by the over-parametrization of the architecture. By 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.

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 $\text{EPICURE}_{\text{seinmeta}}$ study (NCT03958136) for metastatic breast cancer monitoring (Colombié et al. (2021)). Patients underwent between two and three PET/CT acquisitions, corresponding to pre-, mid- and post-treatment time-steps. A total number of 110 pairs of images were obtained, a pair being composed of a pre-treatment and either a mid- or post-treatment image (58 and 52 images respectively). Images were acquired at two different centers. The 54 pairs of images from Angers center were obtained using a Philips Vereos or a GE Discovery PET/CT imaging systems, while the 56 pairs of images from Nantes were acquired on two different dual-slice Siemens Biograph PET/CT scanners.

Since we are interested in lesions monitoring for 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 $1 \times 1 \times 1$ mm^3 with 200 pixels for each side. Besides, no prior registration of any kind was performed.

This prognostic study was approved by the French Agence Nationale de Sécurité du Médicament et des produits de santé (ANSM, #2018-A00959-46) and the Comité de Protection des Personnes (CPP) IDF I, Paris, France (#CPPIDF1-2018-ND40-cat.1). 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 (see Eq. 5). 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, and v) used deformable convolutions (Dai et al. (2017)) in the finest level. Moreover, we also vi) set both λ_{smooth} and λ_{diffeo} to 0 to remove the registration-specific regularization terms (see Fig 1.C).

4.3. Reference methods implementation details

Regarding the methods used as reference, we ran the ANTs pipeline with the settings recommended in Avants et al. (2009), i.e. with a three resolutions coarse-to-fine optimization 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.

Moreover, 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 Fig 1.D).

With respect to the methods based on deep learning, `DL_LapIRN` (Mok and Chung (2020)) was ran with the same regularization balance as `MIRRBA`, i.e. with $\lambda_{smooth} = 0.1$ and $\lambda_{diffeo} = 1.0$, while all other recommended settings were used.

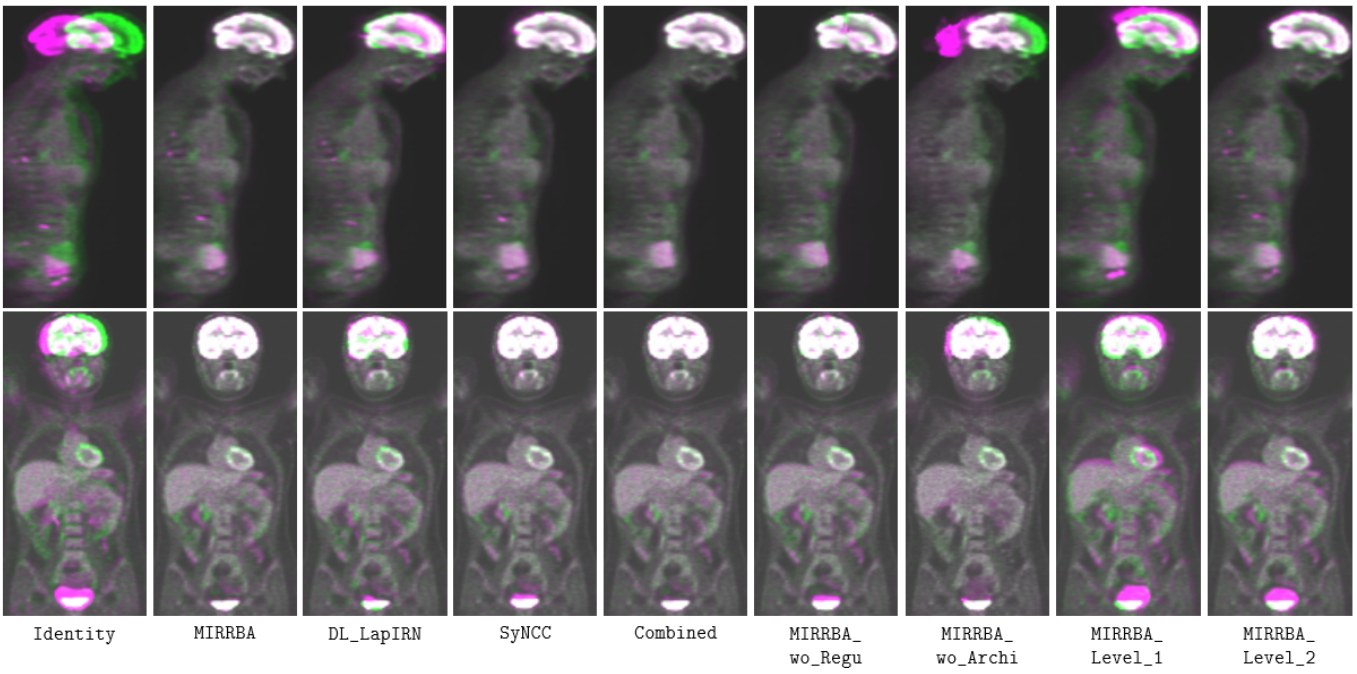

Fig. 2. Overlay of the fixed (green) and warped (pink) images on two different patients after performing `Identity` registration, `MIRRBA` (corresponding to level 3), `DL_LapIRN`, `SyNCC`, `Combined`, `MIRRBA_wo_Regu`, `MIRRBA_wo_Archi`, `MIRRBA_Level_1` and `MIRRBA_Level_2`. Grayscale color indicates 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`). Best viewed in color.

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 $\mathcal{D}_{\text{train}}$ and $\mathcal{D}_{\text{test}}$ as the train and test dataset, respectively. We trained the DL-based approaches on $\mathcal{D}_{\text{train}}$ before testing them on $\mathcal{D}_{\text{test}}$ 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.

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 the fixed images (Moreau et al. (2020)).

Moreover, 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 an accurate segmentation, 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 computation to avoid erroneous null values. Instead, we evaluated the capacity of a method to effectively make lesion disappear by computing its *disappearing rate*, or percentage of lesion shrinkage induced by the registration. This score was computed for converging methods only.

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

Finally, we presented the approximate running time of each approach, using a CPU for `SyN`, `SyNCC` and `Elastix` approaches, 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 can not 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: $\mathcal{R}_{smooth}$, $\mathcal{R}_{diffeo}$, $\mathcal{R}_{dataset}$ and $\mathcal{R}_{archi}$. To this end, we compared the results to our `MIRRBA` approach optimized directly (without learning) on the

Table 1. Comparison to reference methods (Section 5.3) – Detection rate, Dice score of the organs and lesions (Dice_organs and Dice_lesions resp.), SDJDet, disappearing rate, and approximate computational time. Both training and inference computational times are indicated for DL-based methods. While ANTs and Elastix pipeline were computed on CPU, all the other 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.

	Detection rate (%) ↑	Dice organs ↑	Dice lesions ↑	SDJDet ↓	Disap. rate (%) ↑	Comput. time (min) ↓
Identity	5.00	0.626 ± 0.138 *	0.090 ± 0.115 *	0.000 ± 0.000	0.00 *	0
MIRRBA	33.04	0.918 ± 0.126	0.425 ± 0.207	0.124 ± 0.988	9.36	***55***
DL_LapIRN	11.76	0.878 ± 0.076 *	0.258 ± 0.198 *	0.464 ± 1.367	***19.15***	1450 – 3
DL_Voxelmorph	14.13	0.865 ± 0.077 *	0.279 ± 0.192 *	0.224 ± 0.186	5.67	1200 – 2
Elastix	20.54	0.868 ± 0.124 *	0.350 ± 0.191 *	0.096 ± 0.044	9.59	25
SyN	24.57	0.936 ± 0.023	0.386 ± 0.210 *	0.016 ± 0.018	0.00	**5**
SyNNC	***39.57***	***0.944 ± 0.014***	***0.477 ± 0.211***	0.073 ± 0.066	4.26 *	60
Combined	**43.36**	**0.945 ± 0.012**	**0.477 ± 0.202**	0.067 ± 0.076	**19.88**	115

Table 2. Ablation study (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.

	Detection rate (%) ↑	Dice organs ↑	Dice lesions ↑	SDJDet ↓	Disap. rate (%) ↑	Comput. time (min) ↓
Identity	5.00	0.626 ± 0.138 *	0.090 ± 0.115 *	0.000 ± 0.000	0.00 *	0
MIRRBA	33.04	0.918 ± 0.126	0.425 ± 0.207	0.124 ± 0.988	9.36	55
MIRRBA_wo_Regu	33.70	0.868 ± 0.199 *	0.407 ± 0.219 *	6.655 ± 30.709	***16.84***	55
MIRRBA_wo_Archi	20.54	0.753 ± 0.144 *	0.239 ± 0.223 *	1.247 ± 0.475	**59.95**	20
MIRRBA_Depth_1	11.74	0.748 ± 0.169 *	0.239 ± 0.211 *	0.016 ± 0.047	4.82	30
MIRRBA_Depth_2	23.26	0.873 ± 0.135 *	0.364 ± 0.211 *	0.038 ± 0.174	7.56	45
MIRRBA_Depth_4	**40.22**	***0.934 ± 0.092***	***0.456 ± 0.209***	0.104 ± 0.223	15.12	60
MIRRBA_Level_1	4.35	0.722 ± 0.091 *	0.221 ± 0.157 *	0.013 ± 0.010	0.00 *	**2**
MIRRBA_Level_2	19.78	0.869 ± 0.114 *	0.371 ± 0.198 *	0.035 ± 0.1737	0.00 *	***10***
MIRRBA_Max	33.26	0.922 ± 0.098	0.426 ± 0.213	0.316 ± 2.934	8.34	55
MIRRBA_Up	33.04	0.922 ± 0.112	0.439 ± 0.204 *	0.056 ± 0.309	8.84	55
MIRRBA_Def_Conv	35.22	0.487 ± 0.463 *	0.257 ± 0.275 *	0.027 ± 0.020	12.23	130
MIRRBA_Depth_4_Max_Up	***37.83***	**0.935 ± 0.091**	**0.459 ± 0.201**	0.088 ± 0.380	13.05	60

$\mathcal{D}_{\text{test}}$ pairs, using $\mathcal{R}_{smooth}$, $\mathcal{R}_{diffeo}$ and $\mathcal{R}_{archi}$ terms (see Fig 1.A and Eq. 5).

Hence, we trained the DL_LapIRN method on $\mathcal{D}_{\text{train}}$, with the four regularization terms from Eq. 3 (see Fig 1.B). Predictions were made on $\mathcal{D}_{\text{test}}$ images. From Table 1, we note that removing $\mathcal{R}_{dataset}$ in MIRRBA compared to DL_LapIRN helps improve the results, especially for the lesions. Indeed, the organs' Dice score was improved by 5% while the Dice of the lesions improved by 65% when we removed the dependency to the training dataset ($\mathcal{R}_{dataset}$). Moreover, MIRRBA presented lower SDJDet values, hence produced smoother deformations. On the other hand, DL_LapIRN had a higher disappearing rate.

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 Fig 1.C). As quantitatively shown Table 2, MIRRBA presented Dice scores for the organs and lesions which are respectively 6% and 4% lower compared to MIRRBA_wo_Regu. Yet, the detection and disappearing rates were improved for MIRRBA_wo_Regu over MIRRBA. Indeed, though MIRRBA_wo_Regu presents the highest SDJDet score, the realistic appearance of images Fig. 2 makes us consider that the registration process converged.

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 specific architecture, hence not including $\mathcal{R}_{archi}$ (see Fig 1.D). According to Table 2, 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. Moreover, its high SDJDet value combined with the visually imprecise registration (see Fig. 2) makes the convergence of the method questionable.

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). Quantitative results are presented in Table 2, 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. Then, 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 2, 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 only 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 2, 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 were not realistic-looking (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_Def_Conv`). While global registration results were improved over `MIRRBA`, local ones were not. Moreover, the SDJDet was low.

Combining best practices. Finally, we combined the best architectural variations presented above to perform registration with four resolutions, max-pooling and upsampling operations, as well as residual blocks: `MIRRBA_Depth_4_Max_Up`.

`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_Max` and `MIRRBA_Up` statistically improved their results, leading to the best performing MIRRBA-based method in terms of local registration precision. Moreover, the SDJDet of `MIRRBA_Depth_4_Max_Up` was smaller or similar to the one of the three other methods, while the disappearing rate lied between their values. Images of lesions registered with `MIRRBA_Depth_4_Max_Up` are visible Fig. 3.

5.3. Comparison to DL-based and conventional methods

DL-based approaches. As references, we ran the Voxelmorph (`DL_Voxelmorph`) (Dalca et al. (2019)) and LapIRN (`DL_LapIRN`) (Mok and Chung (2020)) networks with a prior training stage on $\mathcal{D}_{\text{train}}$.

Even if both DL-based methods reached similar numerical results (see Table 1) and coherent visual ones (see Fig. 2 and supplementary material), `MIRRBA` performed better than both of them globally and locally. Besides, SDJDet was lower for `MIRRBA` than for the training-based approaches.

Conventional approaches. According to Table 1, 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)) 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_Depth_4_Max_Up`, 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 were shown Table 1. In term of global and local accuracy, the combined approach outperformed both `SyNCC` and `MIRRBA_Depth_4_Max_Up`. In addition, the disappearing rate was significantly improved for both methods.

6. Discussion

6.1. Impact of training on a database

According to the results presented in Subsections 5.1 and 5.3 comparing `MIRRBA` to the DL-based approaches `DL_LapIRN`

and DL_Voxelmorph, not learning registration patterns from a dataset helps to obtain precise segmentations, especially for lesions. Indeed, we argue that while organs' sizes and relative localizations are coherent across patients, lesions' positions and volumes are not. Hence, there are fewer deformation patterns to 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 solve 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 smoothing the deformation field, although having 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.

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. Both qualitative and quantitative MIRRBA_wo_RB results indicate that the registration without the residual blocks fails to converge. Indeed, as explained above, residual blocks allow incremental diffeomorphic mappings, and removing them leads to gradient vanishing. For this architecture, residual connections seem essential to produce realistic deformations.

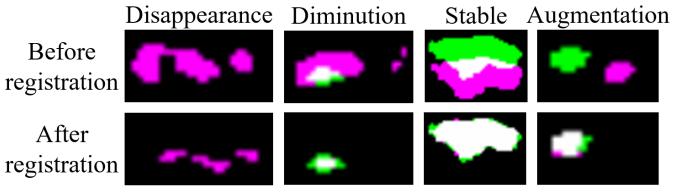

Fig. 3. Overlay of lesions disappearing (1st column), reducing (2nd column), stable (3rd column) and growing (last column) before and after performing MIRRBA_Depth_4_Max_Up. Fixed lesions can be viewed in green, while moving and wrapped ones in pink. White color indicates overlapping. Best viewed in color.

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_Def_Conv. Yet, the resulting low Dice scores show that MIRRBA_Def_Conv 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 train-

able 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 the lesion registration specifically, 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 MIRRBA_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 proposed 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 explicitly considering the dataset and architecture bias. Indeed, our work is motivated by recent work on DIP, implying that neural networks create an inductive bias which is 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 certain 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 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.

For the application at hand, we demonstrate several suitable approaches for the challenging 3D full-body longitudinal registration problem. 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 the dataset bias. In a future work, we would like to study the feasibility of using our registration method to monitor lesion evolution without depending on manually performed segmentations.

Acknowledgments

This work is partially financed through "Programme opérationnel regional FEDER-FSE Pays de la Loire 2014-2020" n°PL0015129 (EPICURE); and by Lilly and AstraZeneca for the clinical trial support.

References

Avants, B.B., Tustison, N., Johnson, H., 2009. Advanced Normalization Tools (ANTS). Insight Journal 2, 1–35.

Avril, S., Muzic, R.F., Plecha, D., Traughber, B.J., Vinayak, S., Avril, N., 2016. 18F-FDG PET/CT for Monitoring of Treatment Response in Breast Cancer. Journal of Nuclear Medicine 57, 34S–39S. doi:10.2967/jnumed.115.157875.

Baguer, D.O., Leuschner, J., Schmidt, M., 2020. Computed tomography reconstruction using deep image prior and learned reconstruction methods. Inverse Problems 36. doi:10.1088/1361-6420/aba415.

Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V., 2018. VoxelMorph: A Learning Framework for Deformable Medical Image Registration. IEEE Transactions on Medical Imaging 38, 1–13. doi:10.1109/TMI.2019.2897538.

Beg, M.F., Miller, M.I., Trouvé, A., Younes, L., 2005. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International Journal of Computer Vision 61, 139–157. doi:10.1023/B:VISI.0000043755.93987.aa.

Carlier, T., Bailly, C., 2015. State-Of-The-Art and Recent Advances in Quantification for Therapeutic Follow-Up in Oncology Using PET. Frontiers in Medicine 2. doi:10.3389/fmed.2015.00018.

Chassagnon, G., Vakalopoulou, M., Régent, A., Sahasrabudhe, M., Marini, R., Hoang-Thi, T.N., Dinh-Xuan, A.T., Dunogué, B., Mouthon, L., Paragios, N., Revel, M.P., 2020. Elastic Registration – driven Deep Learning for Longitudinal Assessment of Systemic Sclerosis. Radiology 298, 189–198. doi:10.1148/radiol.2020200319.

Chen, X., Diaz-pinto, A., Ravikumar, N., Frangi, A.F., 2021. Deep learning in medical image registration. Progress in Biomedical Engineering 3. doi:10.1088/2516-1091/abd37c.

Christensen, G.E., Song, J.H., Lu, W., Naqa, I.E., Low, D.A., 2007. Tracking lung tissue motion and expansion/compression with inverse consistent image registration and spirometry. Medical Physics 34, 2155–2163. doi:10.1118/1.2731029.

Colombié, M., Jézéquel, P., Rubeaux, M., Frenel, J.S., Bigot, F., Seegers, V., Campone, M., 2021. The EPICURE study: a pilot prospective cohort study of heterogeneous and massive data integration in metastatic breast cancer patients. BMC Cancer 21, 1–9. doi:10.1186/s12885-021-08060-8.

Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., Wei, Y., 2017. Deformable Convolutional Networks. Proceedings of the IEEE International Conference on Computer Vision , 764–773doi:10.1109/ICCV.2017.89.

Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R., 2019. Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. Medical Image Analysis 57, 226–236. doi:10.1016/j.media.2019.07.006.

Dittmer, S., Kluth, T., Maass, P., Baguer, D.O., 2020. Regularization by architecture : A deep prior approach for inverse problems. Journal of Mathematical Imaging and Vision 62, 456–470. doi:10.1007/s10851-019-00923-x.

Eppenhof, K.A., Lafarge, M.W., Veta, M., Pluim, J.P., 2019. Progressively Trained Convolutional Neural Networks for Deformable Image Registration. IEEE Transactions on Medical Imaging 39, 1594–1604. doi:10.1109/TMI.2019.2953788.

Friston, K.J., Ashburner, J., Frith, C.D., Poline, J.B., Heather, J.D., Frackowiak, R.S., 1995. Spatial registration and normalization of images. Human Brain Mapping 3, 165–189. doi:10.1002/hbm.460030303.

Fu, Y., Lei, Y., Wang, T., Curran, W.J., Liu, T., Yang, X., 2020. Deep learning in medical image registration: a review. Physics in Medicine and Biology 60, 1–32. doi:10.1088/1361-6560/ab843e.

Gong, K., Catana, C., Qi, J., Li, Q., 2019. PET Image Reconstruction Using Deep Image Prior. IEEE Transactions on Medical Imaging 38, 1655–1665. doi:10.1109/TMI.2018.2888491.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Identity Mappings in Deep Residual Networks. European conference on computer vision , 630–645doi:10.1007/978-3-319-46493-0.

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. doi:10.1007/978-3-030-32226-7_6.

Heinrich, M.P., 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 , 190–200doi:10.1007/978-3-030-59716-0_19.

Hsu, T.M.H., 2020. Automatic Longitudinal Assessment of Tumor Responses. Ph.D. thesis. Massachusetts Institute of Technology.

Jaderberg, M., Simonyan, K., Zisserman, A., Kavukcuoglu, K., 2015. Spatial Transformer Networks. Advances in neural information processing systems 28, 2017–2025. doi:10.5555/2969442.2969465.

Kim, C.K., Gupta, N.C., Chandramouli, B., Alavi, A., 1994. Standardized uptake values of FDG: Body surface area correction is preferable to body weight correction. Journal of Nuclear Medicine 35, 164–167.

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., Parsey, R.V., 2009a. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage 46, 786–802. doi:10.1016/j.neuroimage.2008.12.037.

Klein, S., Pluim, J.P., Staring, M., Viergever, M.A., 2009b. Adaptive stochastic gradient descent optimisation for image registration. International Journal of Computer Vision 81, 227–239. doi:10.1007/s11263-008-0168-y.

Klein, S., Staring, M., Murphy, K., Viergever, M., Pluim, J., 2010. elastix: A Toolbox for Intensity-Based Medical Image Registration. IEEE Transactions on Medical Imaging 29, 196–205. doi:10.1109/TMI.2009.2035616.

Laves, M.H., Ihler, S., Ortmaier, T., 2019. Deformable medical image registration using a randomly-initialized CNN as regularization prior. arXiv , 1–6arXiv:1908.00788.

Liu, F., Liu, D., Tian, J., Xie, X., Yang, X., 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. doi:10.1016/j.media.2020.101793.

Lucas, A., Iliadis, M., Molina, R., 2018. Using Deep Neural Networks for Inverse Problems in Imaging: Beyond Analytical Methods. IEEE Signal Processing Magazine 35, 20–36. doi:10.1109/msp.2017.2760358.

Maurer, C., Fitzpatrick, J., 1993. A Review of Medical Image Registration. Interactive Image-Guided Neurosurgery , 17–44.

Mok, T.C.W., Chung, A.C.S., 2020. Large Deformation Diffeomorphic Image Registration with Laplacian Pyramid Networks. International Conference on Medical Image Computing and Computer-Assisted Intervention , 211–221doi:10.1007/978-3-030-59716-0_21.

Moreau, N., Rousseau, C., Fourcade, C., Santini, G., Ferrer, L., Lacombe, M., Guillerminet, C., Campone, M., Colombie, M., Rubeaux, M., 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. doi:10.1109/EMBC44109.2020.9175904.

Myronenko, A., Song, X., 2009. Adaptive Regularization of Ill-Posed Problems: Application to Non-rigid Image Registration. arXiv preprint arXiv:0906.3323 , 1–10URL: http://arxiv.org/abs/0906.3323, arXiv:0906.3323.

Necib, H., Garcia, C., Wagner, A., Vanderlinden, B., Emonts, P., Hendlisz, A., Flamen, P., Buvat, I., 2011. Detection and Characterization of Tumor Changes in 18F-FDG PET Patient Monitoring Using Parametric Imaging. Journal of Nuclear Medicine 52, 354–361. doi:10.2967/jnumed.110.080150.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A., 2017. Automatic differentiation in PyTorch. 31st Conference on Neural Information Processing System .

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention 9351, 234–241. doi:10.1007/978-3-319-24574-4_28.

Rousseau, F., Drumetz, L., Fablet, R., 2020. Residual Networks as Flows of Diffeomorphisms. Journal of Mathematical Imaging and Vision 62, 365–375. doi:10.1007/s10851-019-00890-3.

Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J., 1999. Nonrigid registration using free-form deformations: Application to breast mr images. IEEE Transactions on Medical Imaging 18, 712–721. doi:10.1109/42.796284.

Sotiras, A., Davatzikos, C., Paragios, N., 2013. Deformable medical image registration: A survey. IEEE Transactions on Medical Imaging 32, 1153–1190. doi:10.1109/TMI.2013.2265603.

Stergios, C., Mihir, S., Maria, V., Guillaume, C., Marie-Pierre, R., Stavroula, M., Nikos, P., 2018. Linear and deformable image registration with 3D convolutional neural networks. Image Analysis for Moving Organ, Breast, and Thoracic Images , 13–22doi:10.1007/978-3-030-00946-5_2.

Ulyanov, D., Vedaldi, A., Lempitsky, V., 2020. Deep Image Prior. International Journal of Computer Vision 128, 1867–1888. doi:10.1007/s11263-020-01303-4.

Vercauteren, T., Pennec, X., Malis, E., Perchant, A., 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. doi:10.1007/978-3-540-73273-0_41.

Vercauteren, T., Pennec, X., Perchant, A., 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. doi:10.1007/978-3-540-75759-7_39.

de Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I., 2019. A deep learning framework for unsupervised affine and deformable image registration. Medical Image Analysis 52, 128–143. doi:10.1016/j.media.2018.11.010.

Supplementary Material

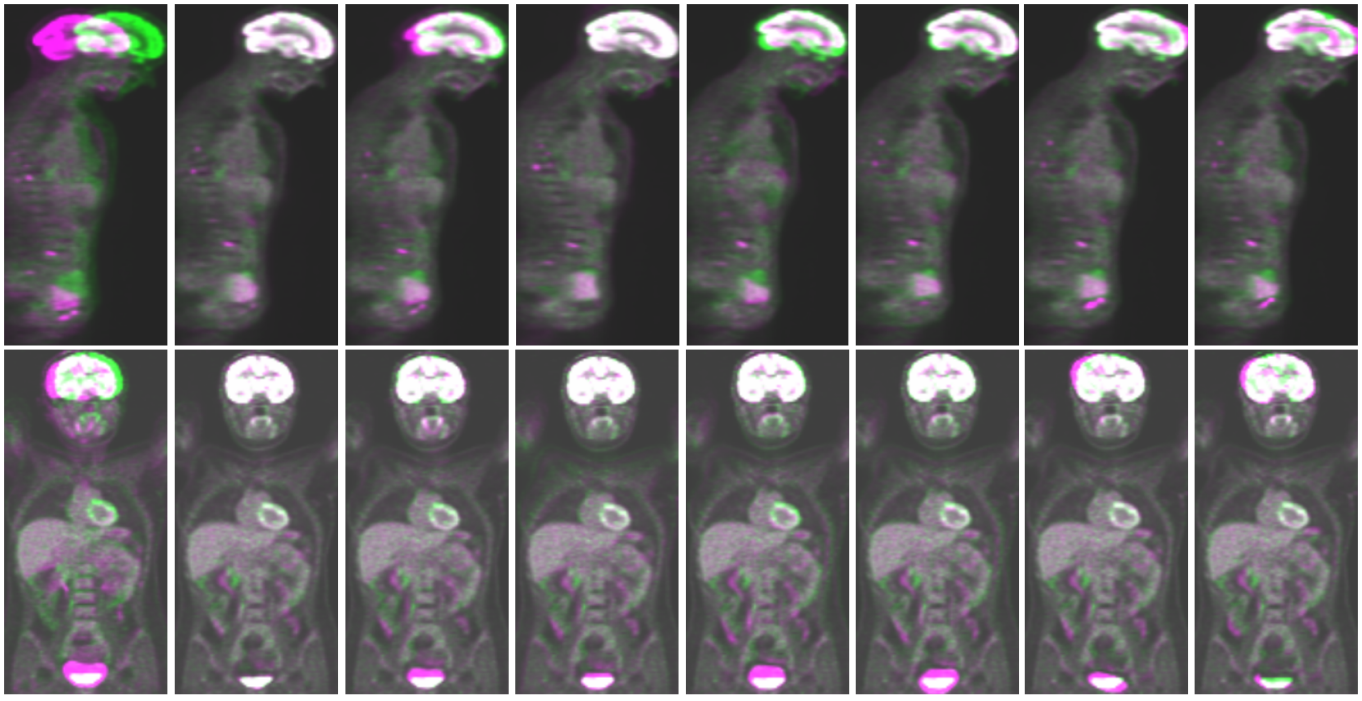

Fig. 1. Comparison to reference methods – Overlay of the fixed (green) and warped (pink) images on two different patients after performing `Identity` registration, `MIRRBA`, `DL_Voxelmorph`, `SyN`, `Elastix_Level_1`, `Elastix_Level_2`, `Elastix_Level_3` and `Elastix` (corresponding to the level 4). Grayscale 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.

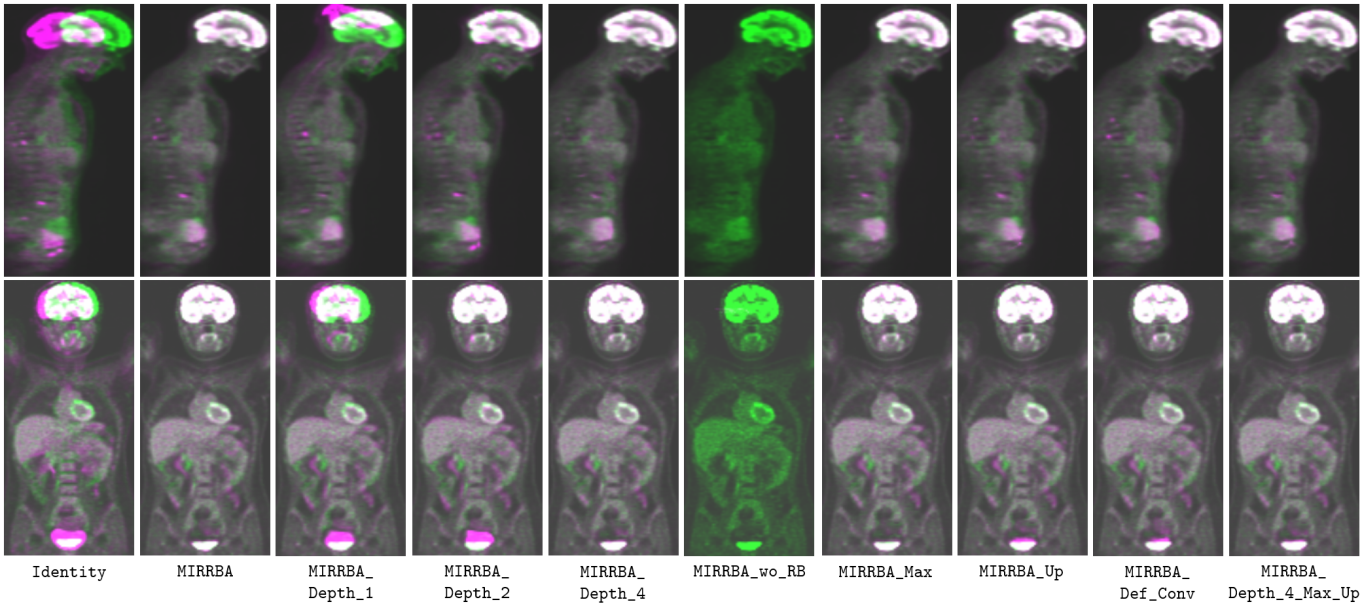

Fig. 2. Ablation study – Overlay of the fixed (green) and warped (pink) images on two different patients after performing `Identity` registration, `MIRRBA` (corresponding to depth 3), `MIRRBA_Depth_1`, `MIRRBA_Depth_2`, `MIRRBA_Depth_4`, `MIRRBA_wo_RB`, `MIRRBA_Max`, `MIRRBA_Up`, `MIRRBA_Def_Conv` and `MIRRBA_Depth_4_Max_Up`. 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_Def_Conv` lacks a bit of precision around the bladder. Best viewed in color.