Contrast-agnostic deep learning–based registration pipeline: Validation in spinal cord multimodal MRI data

E. Béal[a], J. Cohen-Adad[a,b,c]

[a] *NeuroPoly Lab, Institute of Biomedical Engineering, Polytechnique Montreal, Montreal, QC, Canada*
[b] *Functional Neuroimaging Unit, CRIUGM, University of Montreal, Montreal, QC, Canada*
[c] *Mila—Quebec AI Institute, Montreal, QC, Canada*

ABSTRACT

Medical image registration can be challenging, in that optimal solutions depend on the application domain (unimodal, multimodal, intra-subject and inter-subject), anatomical sites (e.g., brain, spinal cord (SC) and lungs), dimensionality of the data (2D, 3D and 4D), deformation constraints (rigid, affine and nonlinear) and computational time. Solutions that could accommodate a large variety of applications while producing satisfactory results are needed. SynthMorph was recently introduced as an unsupervised deep learning–based registration method. A particularly interesting feature is that training is performed on synthetic data so that registration becomes agnostic to image contrast and anatomy. However, SynthMorph is particularly sensitive to the initial closeness of the images. In this work, we extend the SynthMorph method by developing a cascaded pipeline of two models that can accommodate large and fine deformations, respectively. We also validate this pipeline for the registration of intra-subject multimodal and inter-subject uni/multimodal MRI data of the SC. This task is known to be particularly difficult due to the vicinity of multiple tissue types whose morphometrics can vary substantially across subjects and contrasts. Evaluation of the method was conducted on a publicly available dataset (spine-generic, 267 subjects) and was compared with a state-of-the-art benchmark: Spinal Cord Toolbox and Advanced Normalization Tools. Results demonstrate better registration accuracy compared with the benchmark and about 24–30 times faster on CPUs depending on the image size. This proposed pipeline provides an easy-to-use, accurate and fast solution for multimodal 3D registration. The code and trained models are freely available at https://github.com/ivadomed/multimodal-registration.

INTRODUCTION

Registering patient images taken at different times, within and across sessions, is a critical step for automatizing the monitoring of disease evolution, for leveraging multiple contrasts (e.g., T1w, T2w and FLAIR) and for performing atlas-based analysis. Traditionally, the registration process has been formulated as a pairwise optimization problem using a similarity metric, with the aim of aligning a moving/source image with a fixed/reference/target image. Many algorithms and methods have been developed in this sense with cost functions specific to the task. Some of the most popular registration methods and toolkits include the Advanced Normalization Tools (ANTs, (1)) and elastix (2). A traditional pairwise image registration workflow for iterative methods is shown in Figure 1.

The drawback of traditional methods is that they often require substantial computational time, up to tens of minutes to even hours on a CPU (3) for deformable registration of large 3D MRI data. The expensive optimization solved for each pair of test images in traditional methods can be replaced with a global optimization of the registration function during a learning stage using deep learning. Supervised deep-learning methods use ground truth deformations to evaluate registration performance during the training process. To overcome the lack of annotated data, Cao et al. (4) proposed to use the deformation fields obtained by iterative registration methods as ground truths for the training dataset. On the other hand, Sokootie et al. (5) used artificially generated displacement vector fields to train their supervised registration model.
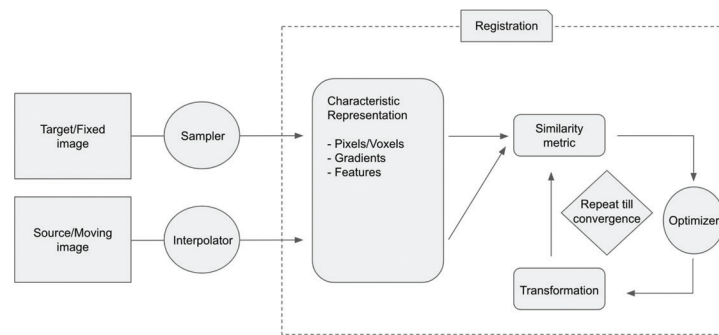
Fig. 1. Representative workflow of iterative image registration. The interpolator building block is used to evaluate moving image voxel intensities at non-grid positions, while the sampler building block defines the subspace of voxels on which the similarity metric is computed. The transformation can be rigid, affine or deformable, for example.

Another possibility is to overcome ground truth deformations during the learning process by using unsupervised deep-learning methods, which have been increasingly developed in recent years for registration (6). The difficulty with unsupervised learning lies in defining an appropriate loss function without using ground truth deformations. With the development of the spatial transformer network (7), loss functions based on image similarity could be used for training neural networks that produce dense deformation fields.

VoxelMorph (3) is one of the unsupervised methods, which has been developed for 3D brain MRI registration. Using a U-Net architecture (8), VoxelMorph takes a pair of images as input and produces a dense deformation field to warp one image onto another. The authors showed that performance comparable to ANTs was achieved while being 150 times faster at registering new image pairs. However, as for the majority of deep-learning registration models, the registration function learned by the models achieves accurate registration under the condition that the new image pair comes from the same distribution (contrasts, geometric content) as the training dataset. This is a major drawback in the development of a registration pipeline for multimodal MRI data. To overcome the issue of MRI contrast generalization, Hoffmann et al. (9) introduced SynthMorph, in which training data are generated from a patch-based noise distribution. This approach is completely unsupervised and has been shown to generalize well to multiple MRI contrasts (and possibly other imaging modalities).

Several works have used deep learning to develop end-to-end medical image registration pipelines (10–13), yet none of them have been developed for spinal cord (SC) data. Zhao et al. (10) proposed cascading registration subnetworks and evaluated the performance of their approach on brain MRI and liver CT data. They showed a significant performance gain using successive registration subnetworks. Only McKenzie et al. (14) studied SC registration using deep-learning models, with generative adversarial networks. They showed that transforming the multimodal problem into a unimodal one led to registration that is more accurate.

This work extends the original SynthMorph approach by implementing a registration method consisting of two cascaded models to accommodate images with large displacements. We evaluate this registration method in the context of SC registration, which, contrary to the brain, includes additional challenges related to the highly deformable structure of the spine. Original aspects of this work include the following:

1. An end-to-end registration framework compatible with any 3D imaging data organized according to the Brain Imaging Data Structure (BIDS) convention (15). The framework is distributed as a standalone open-source Python library.
2. The registration module consists of two successive registration models: a coarse registration model to accommodate images that are far apart and a finer model to ensure high accuracy of registration results.
3. The registration module can accommodate images of any size. If RAM is a limitation (in the case of large 3D images), the pipeline includes an algorithm to register sub-volumes and concatenate all sub-warping fields in space.
4. The post-registration analysis modules allow us to quickly and easily get an idea of the registration accuracy from a qualitative and quantitative point of view. This analysis is tailored to SC images.

MATERIALS AND METHODS

SynthMorph

SynthMorph (9) is a strategy developed to train deep-learning models for deformable (nonlinear) registration. The method is based on unsupervised learning and consists of three main blocks, illustrated in Figure 2, which are involved in each iteration of the learning process: label map generation from noise distribution, grayscale image generation and the Convolutional Neural Network (CNN) with its contrast-agnostic loss used for registration.
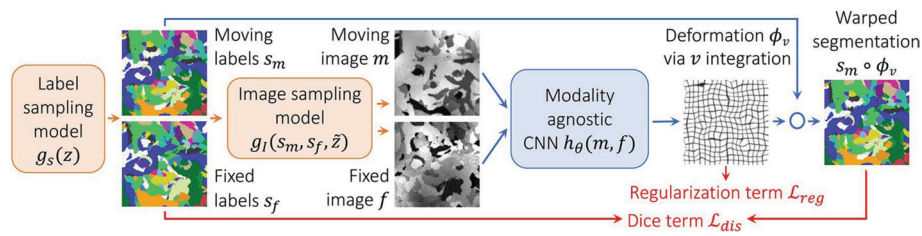
Fig. 2. From SynthMorph: Learning contrast-invariant registration without acquired images (9). Representation of the SynthMorph learning strategy to generate data and train a contrast-agnostic registration model.
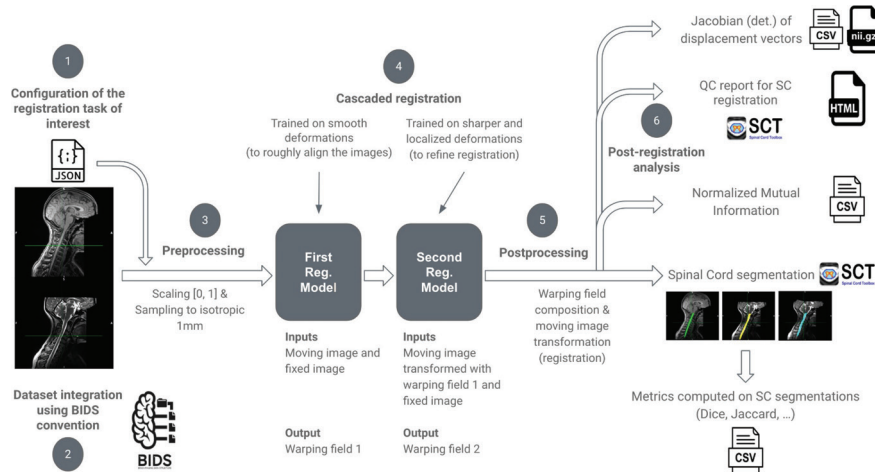
Fig. 3. Overview of the deep learning–based contrast-agnostic registration pipeline and its evaluation tools. Using parameters specified in the configuration file (no. 1), each pair of images in a Brain Imaging Data Structure (BIDS) dataset (no. 2) is preprocessed (no. 3) and registered using two cascaded models (no. 4). Post-registration analyses (nos. 5–6) are performed to gain insight into the accuracy of registration, focusing primarily on the spinal cord.

The label map generation process allows the creation of the training dataset for the registration models (125 label maps for each model). From a smoothly varying noise image, acquired at low resolution and then up-sampled, a random deformation field is applied to create deformations and curves in the image. This process is repeated N times, to generate N different images and create a label map by assigning to each voxel in the label map the label {1, 2, ..., N} corresponding to the image that has the highest intensity for that voxel. These label maps are then used in the SynthMorph learning strategy to produce unregistered pairs of synthetic multi-contrast grayscale images at each iteration of the learning process. The final goal of this method is to obtain a contrast-agnostic registration model that takes as input a pair of images of any MRI contrasts and produces a warping field to register the moving image to the fixed one.

The SynthMorph authors demonstrated for brain MRI data that their registration model matches the state-of-the-art performance of iterative methods, such as ANTs (1) and NiftyReg (16), and deep-learning registration method, VoxelMorph (3), within contrasts and improves registration accuracy on multimodal data.

The SynthMorph method was chosen to obtain deep-learning registration models for the pipeline and was implemented and adapted from the publicly available code at https://github.com/voxelmorph/voxelmorph.

Pipeline

The developed registration pipeline provides an easy-to-use, accurate and fast solution for multimodal 3D registration. This is achieved in a (almost) fully automated environment consisting of six main parts, illustrated in Figure 3.

1. Configuration file

A JSON configuration file, specific to the pipeline, is used to specify parameters that can be modified in the pipeline (interpolation and registration on sub-volumes or the entire volumes). The user has also to specify at the beginning of the pipeline (shell script) the contrasts of the images to co-register, which post-registration analyses to perform and how to organize the registration results.

2. Dataset integration

Any dataset organized according to the BIDS (15) convention is compatible with the registration pipeline without requiring manual interventions.

3. Preprocessing

The preprocessing steps, performed automatically within the pipeline and on all images, include scaling the voxel
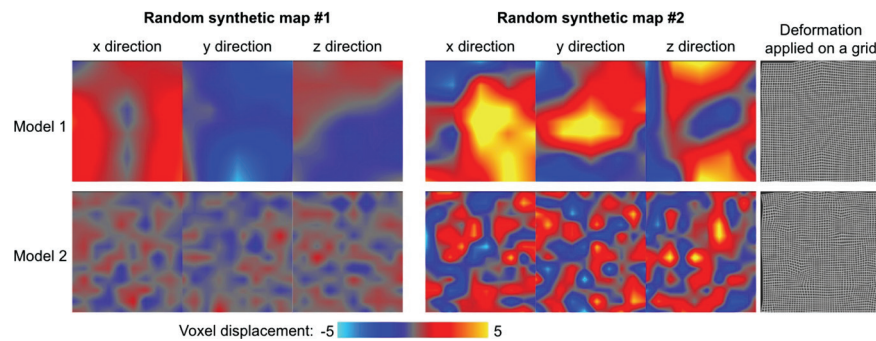
Fig. 4. Example of deformation fields used to generate unregistered image pairs during the registration models learning process. Displacement intensity (in voxels) in the x, y and z directions is shown for each voxel within a specific slice. The deformations applied in model 2 are more localized.

values between 0 and 1 (min–max scaling), resampling the 3D images to 1 mm isotropic resolution, placing the moving image in the fixed image space and zero padding each dimension to the nearest multiple of 16.

4. Cascaded registration

Two registration models, trained with the SynthMorph method (9), are used in succession. The cascaded architecture was developed to allow the registration of widely displaced images (1–2 cm) while ensuring high-registration accuracy. This is done without further deteriorating the image resolution or increasing the interpolation error, as the warping fields obtained from the two models are composed before transforming the moving image. The objective of the first deformable registration model is to roughly align the images, while the second model refines the registration. Therefore, these registration models were trained on different training datasets to learn different types of transformations. The difference in the training datasets relies on the relative resolutions at which noise is sampled to generate warping fields and deform the label maps to create an unregistered pair of images. The first model was trained on pairs of images that were unregistered using smooth deformation fields. The second model was trained on deformation fields with sharper and more localized deformations. An example of deformation fields specific to the two different models is shown in Figure 4, where heat maps representing the translation intensity (in voxels) applied to each voxel to deform the training data and obtain unregistered images can be observed.

5. Postprocessing

Using the composed 3D warping field, the moving image is registered to the fixed image. For each voxel of the moving image, a new voxel location is computed to determine the voxel value in the registered image. However, since the image values are only defined at integer locations, the values must be interpolated. The user can choose between trilinear or nearest neighbor interpolation to obtain the moved image. The warping field and registered volume are also transformed back into the

native space of the moving image space, in addition to the results obtained in the fixed image space. This offers the possibility to bring the fixed image into the space of the moving image (instead of the other way around), which could be useful for some processing methods.

6. Post-registration analyses

The pipeline offers a choice of different analyses that are directly executed at the end of the registration performed by the deep-learning models and can be selected independently depending on the study of interest. Both quantitative and qualitative validation tools are incorporated, focusing primarily on the SC. These tools cannot fully replace a visual assessment of the registration, looking at the whole fixed, moving and moved images. However, they provide complementary information that should be used to easily identify whether or not the registration produced the expected results for individual subjects, thus saving time and providing a useful general overview of the registration accuracy.

Registration Models

The deep-learning registration models used in the pipeline were trained with the SynthMorph method (9), and the parameters were specified in Table S1. The networks were implemented using TensorFlow/Keras (17). The training uses the Adam optimizer (18) and was done on an NVIDIA RTX A600 GPU. The training time of 1 epoch (100 steps, mini-batch size of 1) was approximately 150 seconds.
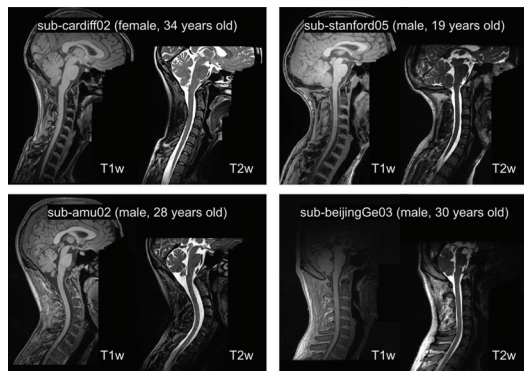
Evaluation

Data

The pipeline was validated using a publicly available multi-subject dataset[1] (19) acquired at 42 centers following the spine-generic protocol (20). The dataset

---

[1] https://github.com/spine-generic/data-multi-subject#spine-generic-public-database-multi-subject

Fig. 5. Example of T1w and T2w 3D MRI data of the spine-generic dataset (sagittal view).

(spine-generic dataset) is composed of MRI data from 267 healthy subjects obtained at multiple sites and from multiple MRI manufacturers (GE, Philips and Siemens) and models. For each subject in the dataset, multimodal (multi-contrast) MRI data of the SC were acquired (e.g., 3D sagittal T1w and 3D sagittal T2w) and resampled to an isotropic resolution (e.g., 1 mm for T1w and 0.8 mm for T2w), which is assumed to be the effective resolution. In this study, the experiments were performed on the sagittal T1w and T2w contrasts. The spine-generic dataset is organized according to the BIDS convention (Figure 5).

For some experiments, affine (rotation, scaling, translation) transformations were applied on specific MRI contrast of the dataset.

Dice similarity coefficient on SC segmentations

Using the Spinal Cord Toolbox (SCT) (21) and the *sct_deepseg_sc* framework developed by Gros et al. (22), SC segmentations are computed independently, within the pipeline, for the moving, fixed and registered images. As shown in Gros et al. (22), the developed framework for SC segmentations handles the heterogeneity of image acquisition features well, providing accurate SC segmentations on a multicenter, multiresolution, multi-contrast dataset. Moreover, the quality of the automatic segmentation applied to the spine-generic dataset was already assessed in previous studies by Bautin et al. (23), and only minor corrections were required in a few subjects. Therefore, SC segmentations of the moving, fixed and registered images are reliable enough to calculate volumetric overlap metrics, such as the Dice similarity coefficient, to evaluate registration in a quantitative approach, focusing on SC.

Baselines

Classical registration was tested using ANTs (1) (symmetric nonlinear normalization (SyN) and rigid registration (Rig)) and axial slice-by-slice regularized (SliceReg) (24) registration through the *sct_register_multimodal* command line of the SCT (21). Different registration strategies were designed to allow a broad comparison with the developed approach. For each registration method tested, the fixed (T1w) and moving (T1w or T2w) images were resampled to 1 mm isotropic resolution to apply the same preprocessing as in the registration pipeline. Preliminary investigations were conducted to select the parameters of the different methods.

SyN

ANTs deformable registration. Mutual information is used as a similarity metric to optimize the registration problem through the symmetric nonlinear normalization (SyN) algorithm. Images are downsampled by a factor of 2 to allow for larger deformations and faster computations, and a smoothing factor of 0.5 mm is used. Registration is performed for 30 iterations with a gradient step size of 0.5.

RigSlicereg

ANTs rigid registration & SliceReg registration. This approach is better suited than conventional affine or rigid methods for the SC registration due to the articulated nature of the spine. The mutual information similarity metric is used for the rigid registration and the SliceReg algorithm. Both registration steps are performed for 30 iterations with a gradient step size of 0.5.

RigSliceregSyn

ANTs rigid registration & SliceReg & ANTs deformable registration. Parameters similar to SyN and RigSlicereg are used for the deformable, rigid and slice-by-slice regularized registration.

RigSliceregSynMask

ANTs rigid registration & SliceReg & ANTs deformable registration focusing on the SC using a mask to optimize the similarity metric. The mask is obtained by dilating the SC segmentation using a mathematical morphology operation with a ball-shaped structuring element of radius 8 voxels. Parameters similar to SyN and RigSlicereg are used for the deformable, rigid and slice-by-slice regularized registration.

SCSegReg

Registration is based on SC segmentation (upper baseline for SC registration). SC alignment is maximized by iterative algorithms using a cost function computed over the SC segmentations. Therefore, this process is expected to provide the most accurate registration results based on SC overlap. It consists of a first affine (translation, rotation and scaling) registration step followed by an axial slice-by-slice regularized registration and a deformable registration using mean squares similarity metric. Each registration step is performed for 30 iterations with a gradient step size of 0.5, and the deformable registration is performed on images downsampled by a factor of 2.

smshapes

The publicly available deep-learning registration model,[2] smshapes, obtained by the authors of the SynthMorph method (9) is used as a benchmark to evaluate the improvement (or deterioration) of the proposed approach, which uses two cascaded models.

Various experiments were conducted, representing realistic scenarios in terms of proximity of the images to co-register and the variety of contrasts to validate the registration pipeline.

Experiment 1: Intra-subject, multimodal, small movements between images

The T2w image was registered to the T1w image for each of the 267 subjects in the spine-generic dataset. Each pair of images in the dataset was included in the analysis, as it was visually assessed that subjects moved very little between T1w and T2w scans in the condition used to acquire the spine-generic dataset. Registration is performed using two successive deformable deep-learning registration models (cascaded models) and using each of the baselines. Registration accuracy, based on the Dice score (DSC) of the SC segmentations, is compared between the methods. Using paired two-sided t-tests, the statistical significance of the mean difference between the methods is determined. A visualization of some fixed, moving and moved images is provided to qualitatively assess the differences between the registration methods.

Experiment 2: Intra-subject, multimodal, large movements between images

Similar analyses than in Experiment 1 are performed but using a dataset where the T2w images have been affine transformed (rotated, scaled and translated) to study how the different models/methods generalize on largely linearly shifted data. Therefore, random affine transformations were applied to the T2w data of each pair of images before registration. The transformations range from −5 to 5 degrees for rotation, from 0.95 to 1.05 for scale factor and from −5% to 5% of the field of view in each direction for translation (e.g., for a field of view of 100 mm, the displacement range would be −5 to 5 mm). The magnitude of the transformations is chosen randomly and independently for each T2w image, the goal being to study the registration accuracy over a wider spectrum of data by adding these linear transformations. The impact of the transformations on the registration accuracy (DSC of SC segmentations) is studied, and some deformation fields are visualized using the Jacobian determinant.

Experiment 3: Inter-subject, unimodal

The image pairs to be registered are formed by using the T1w MRI image of each subject in the dataset as a fixed image and randomly selecting a T1w MRI image of

any other subject in the dataset as a moving image. This leads to pairs of images that have very few features in common, with different orientations, different positions and different relative voxel intensities. The resulting registration task is too complex to apply deep-learning models directly without prior data preprocessing. Therefore, in this case, additional steps were included before the registration of the cascaded models to account for the different scenarios that may be encountered. First, using existing labels of the intervertebral disc position, the discs between the C2/C3 and C7/T1 vertebrae in fixed and moving images are aligned through XYZ translation to approximate the SC registration. The position of the pontomedullary junction (automatically detected using SCT) in both images is used to translate the moving image along the right-left and posterior-anterior axes to coarsely align the brain structure. Finally, the SC segmentations are computed to perform an axial slice-by-slice alignment of the SC center of mass. Deep-learning registration models then use these images to register the entire volume and refine the SC registration. This process, described in Figure 6, is more complex than the intra-subject registration task because of the extreme variations that can occur in pairs composed of randomly selected subjects. However, it does not involve any time-consuming iterative process based on voxel intensities or image features. No baseline comparison is performed in Experiment 3, as the task is too complex to perform direct image registration without any preprocessing for both the classical and the deep-learning registration methods tested in Experiments 1 and 2.

Nine subjects from the spine-generic dataset were removed for this analysis because the labels of their intervertebral discs were not present in the dataset. Therefore, results were computed on 258 pairs. The DSC obtained
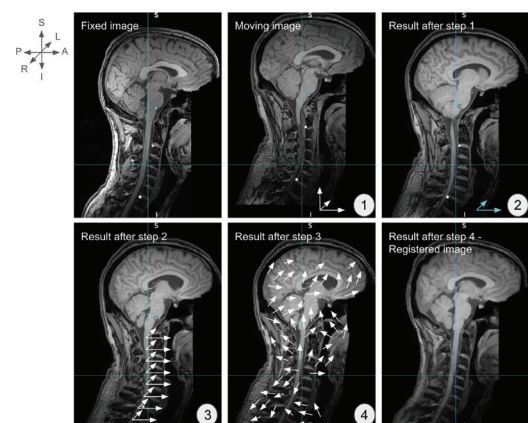
Fig. 6. Inter-subject registration process. The orientation of the 3D images is indicated by S (superior), I (inferior), R (right), L (left), P (posterior) and A (anterior). (1) Translation in the S-I, P-A and R-L directions to align the positions of the C2/C3 and C7/T1 intervertebral discs (white dots, existing labels). (2) Translation in the R-L and P-A directions to align the pontomedullary junction of fixed and moving images (blue dot, automatically detected with Spinal Cord Toolbox (SCT)). (3) The spinal cord segmentations of the fixed and moving images are used to perform a slice-by-slice axial alignment of the spinal cord using the center of mass. (4) Deep-learning deformable registration using cascaded models.

after the different steps of the process is studied to determine the accuracy of the SC registration. Typical registration results are displayed to observe how the different structures of the moving image are registered to the fixed after each stage of the inter-subject registration process. The effect of the deep-learning models on the SC registration is investigated qualitatively by examining quality control reports.
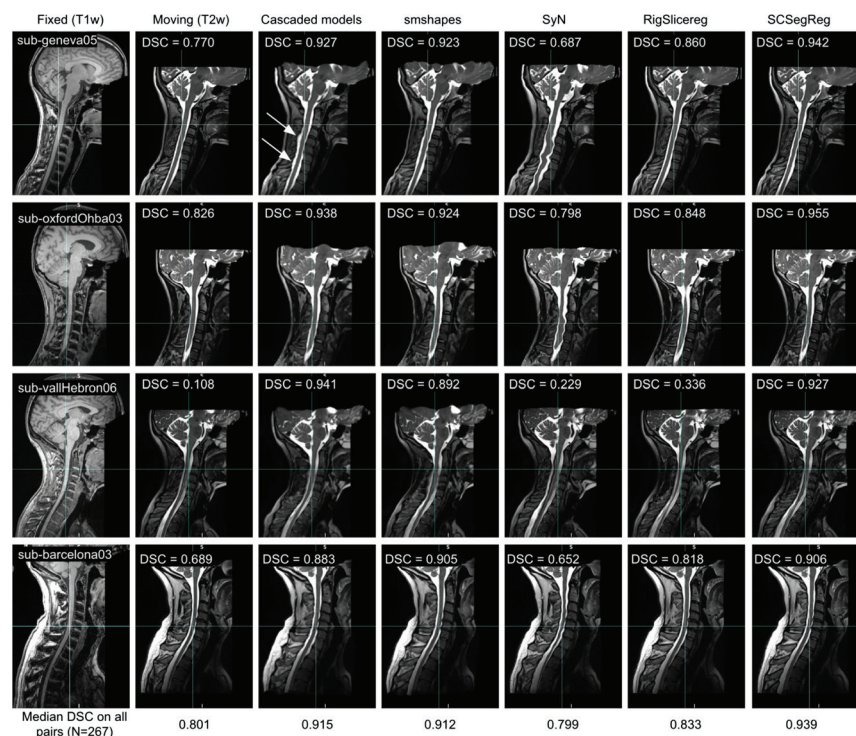
RESULTS

Experiment 1

Figure 7 illustrates typical results (sub-geneva05, sub-oxfordOhba03, sub-vallHebron06 and sub-barcelona03) for the different registration methods. A single slice is shown in the sagittal view for each 3D image. The cascaded and smshapes models demonstrate overall good registration results on the subset of four subjects. For the cascaded models, some overly strong deformations are applied on the cerebrospinal fluid of sub-geneva05, creating an undesirable ripple effect in the fluid surrounding the SC (white arrows in the figure). For the smshapes model, the registration of the intervertebral discs of sub-vallHebron06 is poorly accurate. Apart from these observations, both deep-learning methods show accurate registration, close to that obtained with SCSegReg,

the upper baseline for the SC registration. For classical registration methods, the results are extremely variable. Even methods used specifically for SC registration (i.e., including SliceReg) do not always lead to satisfactory image alignment. For instance, the results of RigSlicereg for sub-vallHebron03 show that the iterative registration algorithms probably reached a local minimum of the cost function (i.e., similarity metric) that is far from the global minimum that would lead to a registration similar to that obtained with SCSegReg.

Figure 8 compares the registration accuracy of the cascaded registration approach with all baselines, based on the DSC of the SC segmentations. These results indicate that the cascaded approach obtains the best accuracy for the SC registration among all registration methods (except for SCSegReg, the upper baseline, which performs registration by maximizing the overlap of SC segmentations). It outperforms classical methods ($p < 1.3 \times 10^{-24}$) and improves the original SynthMorph model (smshapes, $p < 0.009$). Although preliminary research was conducted to select the parameters of the classical methods for this specific registration task, results show that none of these methods improve the registration accuracy of the original image pairs ($p < 0.6$ for RigSlicereg, the classical method with the most accurate SC registration results). The upper baseline, using the SC segmentations to register the SC, achieves a median DSC of 0.939, while the cascaded and the smshapes

Fixed (T1w)	Moving (T2w)	Cascaded models	smshapes	SyN	RigSlicereg	SCSegReg
sub-geneva05	DSC = 0.770	DSC = 0.927	DSC = 0.923	DSC = 0.687	DSC = 0.860	DSC = 0.942
sub-oxfordOhba03	DSC = 0.826	DSC = 0.938	DSC = 0.924	DSC = 0.798	DSC = 0.848	DSC = 0.955
sub-vallHebron06	DSC = 0.108	DSC = 0.941	DSC = 0.892	DSC = 0.229	DSC = 0.336	DSC = 0.927
sub-barcelona03	DSC = 0.689	DSC = 0.883	DSC = 0.905	DSC = 0.652	DSC = 0.818	DSC = 0.906
Median DSC on all pairs (N=267)	0.801	0.915	0.912	0.799	0.833	0.939

Fig. 7. Representative intra-subject multimodal registration results for the deep-learning models and classical methods. Each row shows a pair of fixed (T1w MRI contrast) and moving (T2w MRI contrast) images followed by the warped image obtained using the different methods. The results obtained with the RigSliceregSyn and RigSliceregSynMask methods are not displayed to better visualize the results of the other methods. The Dice score (DSC) representative of the spinal cord segmentations overlap between the fixed 3D image and the moving or moved 3D image is displayed. The median DSC was obtained on the 267 subjects of the dataset. White arrows highlight the ripple effect created by the deep-learning registration on sub-geneva05 data.
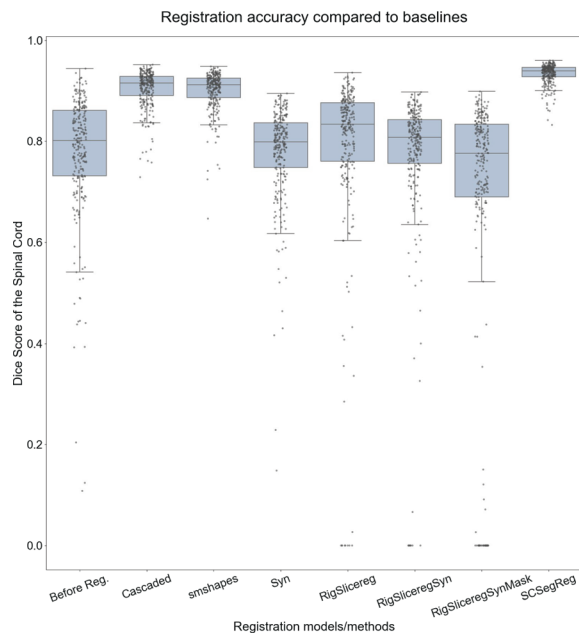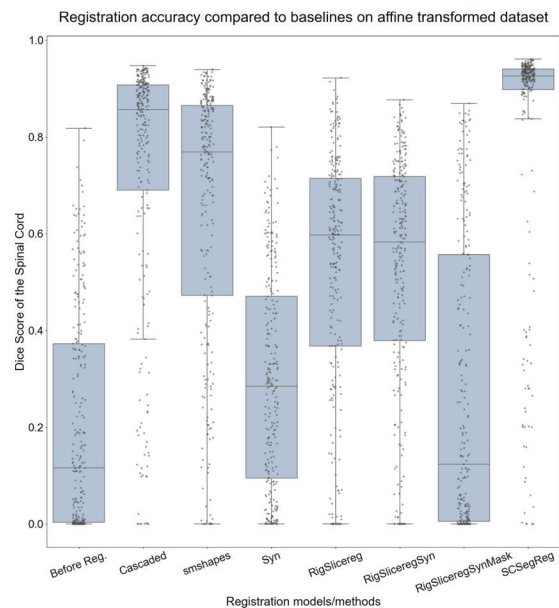
Fig. 8. Registration accuracy (intra-subject, multimodal) of the different models/methods compared, focusing on the spinal cord. Accuracy is evaluated using the Dice score (DSC) representative of the volume overlap of spinal cord segmentations. Each box summarizes the registration results obtained on the 267 subjects of the dataset for the registration of the T2w modality to the T1w. The dots represent the DSC obtained for each subject.


Fig. 9. Registration accuracy (intra-subject, multimodal) of the different models/methods compared, on the affine-transformed dataset. Accuracy is evaluated using the Dice score (DSC) representative of the volume overlap of spinal cord segmentations.

models achieve a median DSC of 0.915 and 0.912, respectively. These results were computed on images with mostly small motion and variations between the two scans (T1w and T2w), as illustrated in Figure 8 by the high DSCs (median of 0.801) of the image pairs before any registration method was used. A DSC of 0 is obtained for 8 and 29 image pairs for RigSlicereg/RigSliceregSyn and RigSliceregSynMask, respectively. These results are obtained because the joint probability distribution function of the two images summed to zero during the axial slice-by-slice regularized registration step. Therefore, the mutual information similarity metric could not be computed and the iterative registration process could not be optimized properly.

Experiment 2

Differences in the registration results between methods are increased when compared with data with more distant initial positions. Figure 9 shows the results of the DSC analysis. In contrast to Figure 8, the DSCs are lower and more scattered for each method. The SCs are initially less well aligned in each pair of multimodal images (median DSC of 0.116) because of the random affine transformations applied to the moving image (T2w contrast). The cascaded registration approach achieves the most accurate registration results based on SC overlap, with a median DSC of 0.857. It outperforms smshapes, the other deep-learning registration model (median DSC of 0.769, $p < 9.4 \times 10^{-13}$) with a 95% confidence interval of 0.079 ±

0.02 for the pairwise DSC difference. The classical methods fail to achieve satisfactory registration results while improving (statistically significantly, $p < 9.5 \times 10^{-38}$ for the most accurate method) the registration compared with the initial positions. The best results for the image-based iterative registration methods are again obtained with rigid registration followed by axial slice-by-slice regularized translation (RigSlicereg) with a median DSC of 0.597. The best overall results are obtained with SCSegReg, the method using SC segmentations to optimize registration.

Figure 10 qualitatively compares the registration results obtained by the different methods. The results highlight the impact of specific motion on the registration accuracy. For sub-geneva05, the T2w data were transformed before registration with −4.9 degrees of rotation along the right-left and inferior-superior axes, 3.8% shrinkage and 1, −1 and 4 voxel translation for the right-left, posterior-anterior and inferior-superior axes, respectively. These transformations did not lead to a significant decrease in the SC registration accuracy of the deep-learning models, as it can be visually observed by comparing the first two rows, but a less accurate registration of the intervertebral discs and cerebrospinal fluid is observed for the smshapes model. RigSlicereg exhibits a significant loss in the registration accuracy due to the transformation constraints of the method (global translation, slice-wise translation and rotation, no scaling). With the registration results obtained on sub-barcelona03, we have an idea on how the models/methods process the translated data, as the T2w image of this subject was transformed with −1, 13 and 11 voxels translation for the right-left, posterior-anterior and inferior-superior axes, 1.3% shrinkage and less than
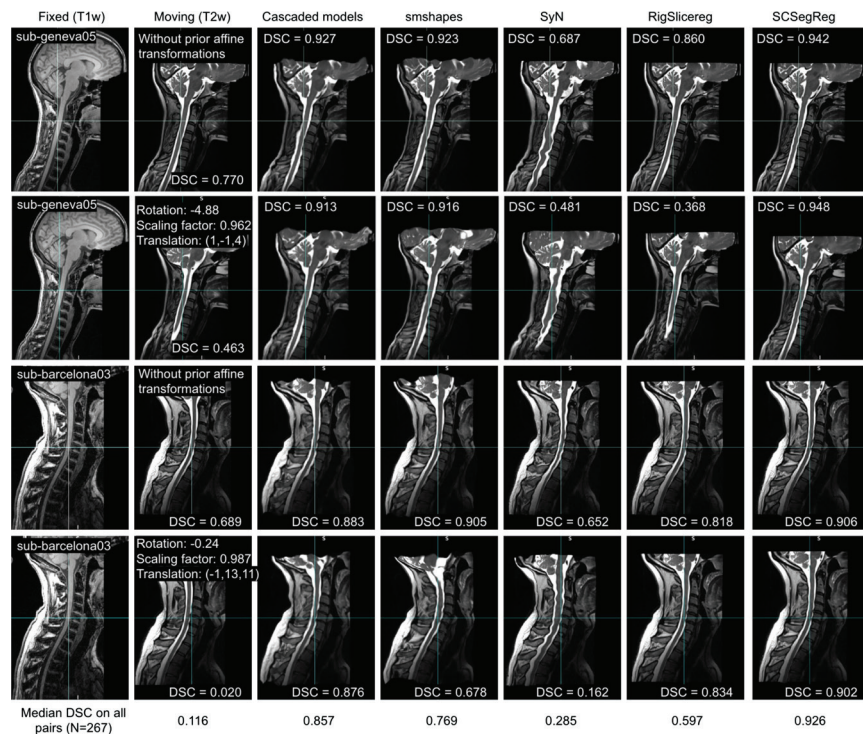
Fig. 10. Example of registration results on affine-transformed data compared with results obtained when no transformation is applied before the registration task. The first and third rows show the results for two pairs of fixed (T1w MRI contrast) and moving (T2w MRI contrast) images of the normal dataset. The second and fourth rows show the registration results for the same subjects, but with the moving image that has been transformed by rotation, scaling and translation before registration. The Dice score (DSC) representative of the overlap of the spinal cord segmentations between the fixed image and the moving or moved image is displayed. The median DSC was obtained on the 267 subjects in the affine-transformed dataset. The results obtained with the RigSliceregSyn and RigSliceregSynMask methods are not displayed to better visualize the results of the other methods.

0.3 degrees rotation. Considering the deep-learning models, this transformation is best handled with the cascaded approach, which leads to a registered image similar to the one obtained without prior transformation. In contrast, smshapes shows a significant decrease in registration accuracy when registering the transformed data. Interestingly, this transformation has no effect on RigSlicereg, certainly due to the translational nature of the deformation.

The comparison of the warping field obtained with the cascaded models to that resulting from the smshapes registration gives information about the specificities of the two approaches. The 3D image of the Jacobian determinant, automatically computed by the pipeline, is used to better visualize the warping fields and the transformations applied to register the moving image. Figure 11 compares the warping fields on the two subjects studied in Figure 10. For the results of sub-geneva05, the deformations applied in the two approaches appear to be quite different, even in the SC area where both methods lead to a high DSC, as observed in Figure 10. This result highlights the ill-conditioned property of the deformable registration problem. The results on sub-barcelona03 show the limitations of smshapes with a lot of local folding
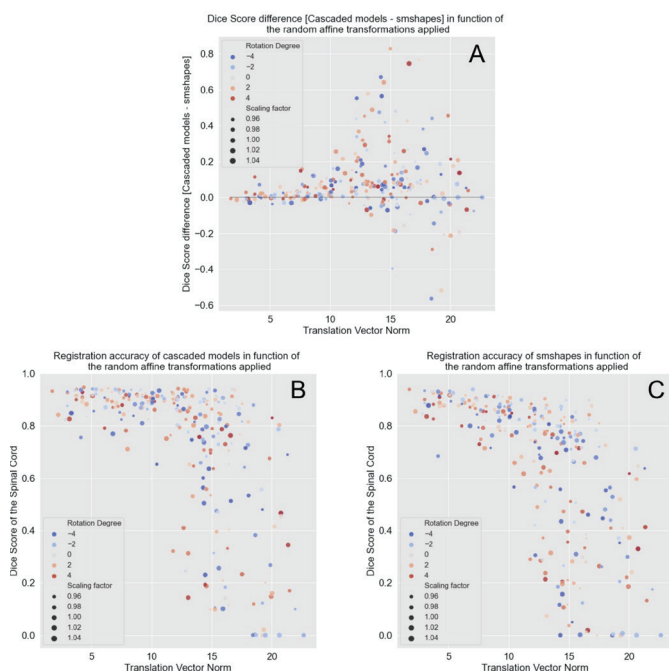
(black voxels, negative Jacobian determinant). In contrast, the transformations applied with the cascaded models are smoother and include fewer folding voxels. Therefore, this analysis of the Jacobian determinant of the deformation fields indicates that, when the data are initially further apart, the cascaded models apply more realistic transformations to register the images than smshapes.

Figure 12 illustrates the impact of the combination of affine transformations applied before registration on the accuracy of the SC registration using the cascaded or smshapes models. Figure 12A highlights the benefits of using two successive registration models (cascaded approach), with more accurate SC registration on this transformed dataset. However, as shown in Figure 12B, the registration accuracy decreases sharply when the transformations applied before registration are too large, even with this approach. Figure 12B and C shows a clear trend indicating that the transformation with the greatest (detrimental) impact on registration performance is translation, for both the cascaded and smshapes models. The effect of each transformation (rotation, scaling and translation) on the registration accuracy of both approaches is studied in the results section of Experiment S2 (Figure S4).
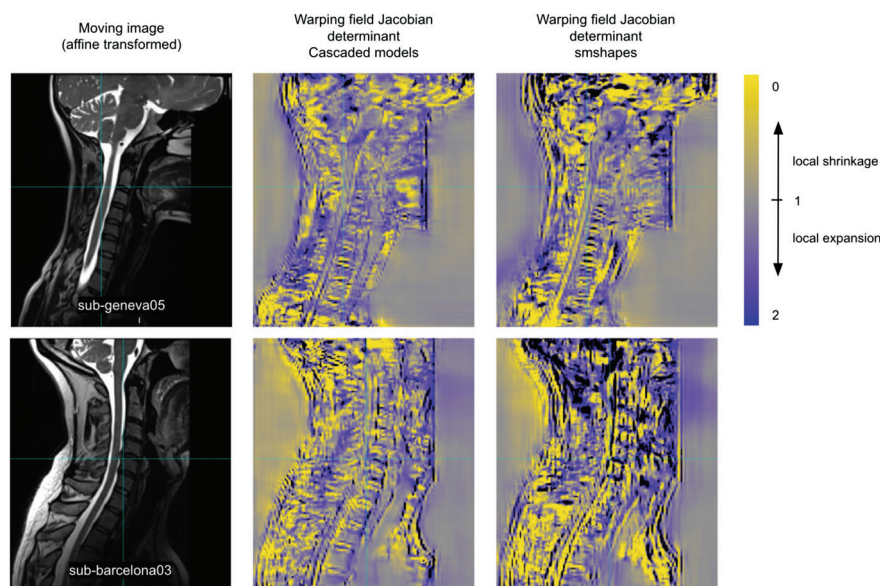
Fig. 11. Jacobian determinant of the warping field obtained for registering data that have undergone an affine transformation before the registration task. The Jacobian determinant gives information about the local transformation applied to each voxel. The yellow color indicates a local shrinkage, the blue color indicates a local expansion and the black voxels are representatives of a folding (negative Jacobian determinant).


Fig. 12. Dice score (DSC) of spinal cord segmentations as a function of applied transformations before registration. (A) Comparison of the DSC obtained in the two deep-learning registration approaches. (B) and (C). DSC obtained after registration by the cascaded and smshapes models, respectively, as a function of the translation (x axis), degree of rotation (color) and scale factor (point size) applied to transform the moving image before registration.

Experiment 3

Figure 13 illustrates how the moving image is transformed to be registered with the fixed one. The image resulting from each step of the inter-subject registration process is shown for four different pairs of images. The results for the first and second pairs are really accurate, with the final registered 3D image (last column) being very similar to the fixed image (first column). The differences observed between the last two columns show the role of cascaded deep-learning models in this registration process. The models refine the registration of the SC by scaling it to the correct volume and smoothing the entire anatomical structure. In addition, they register all other structures (e.g., intervertebral discs, white and gray matter of the brain) present in the 3D MRI images and reduce the sharp transitions that appear due to the previous registration task, which applies translation only on the axial slices of the SC. The results for the composite pair of sub-cardiff04 and sub-tokyo750w07 MRI data (third row) are also very accurate even though the task is more complicated because there is a difference in the details that can be observed in the two images with the moving image showing anatomical structures with small differences in voxel intensity, in contrast to the fixed image. The results of the sub-unf05 T1w to sub-beijingGE03 T1w registration show a displacement of the intervertebral discs between the thoracic vertebrae that can be observed in the inferior part of the fixed and registered sagittal images. The brain registration is poorly accurate either, with the deep-learning models applying too much compression on the superior part. The edge-shading artifact observed in the fixed image may account for this decrease in registration accuracy compared with the other image pairs.

To better observe the effects of cascaded deep-learning models on SC registration, we can use the quality control (QC) reports that are created in the registration pipeline. The QC reports provide insight into the
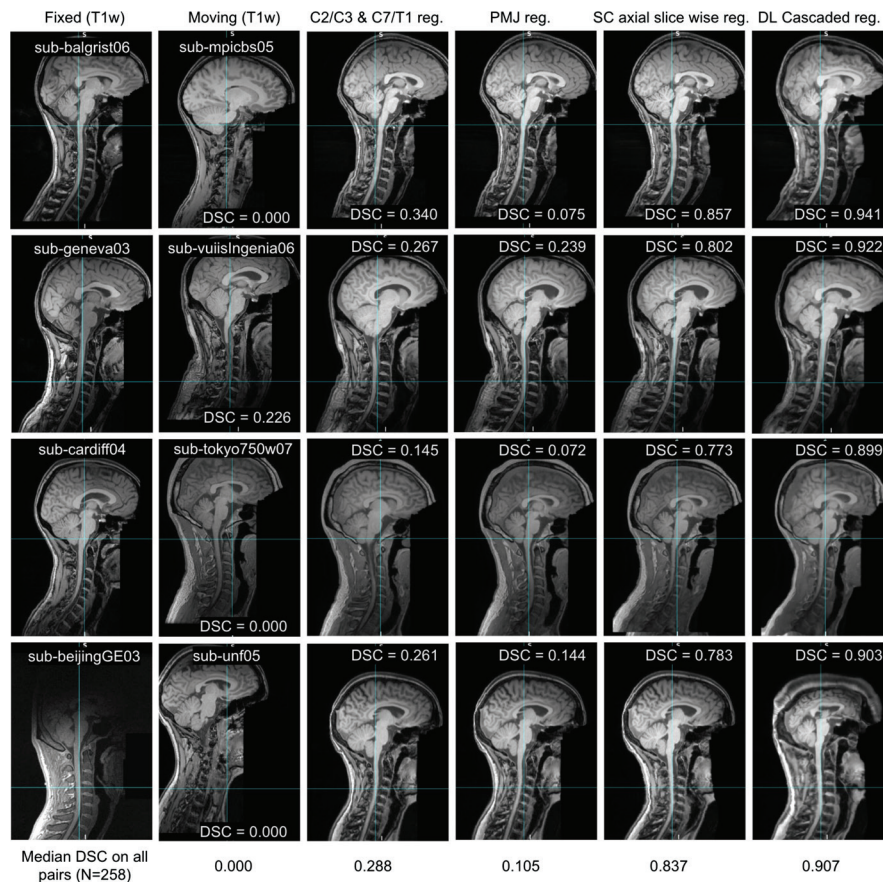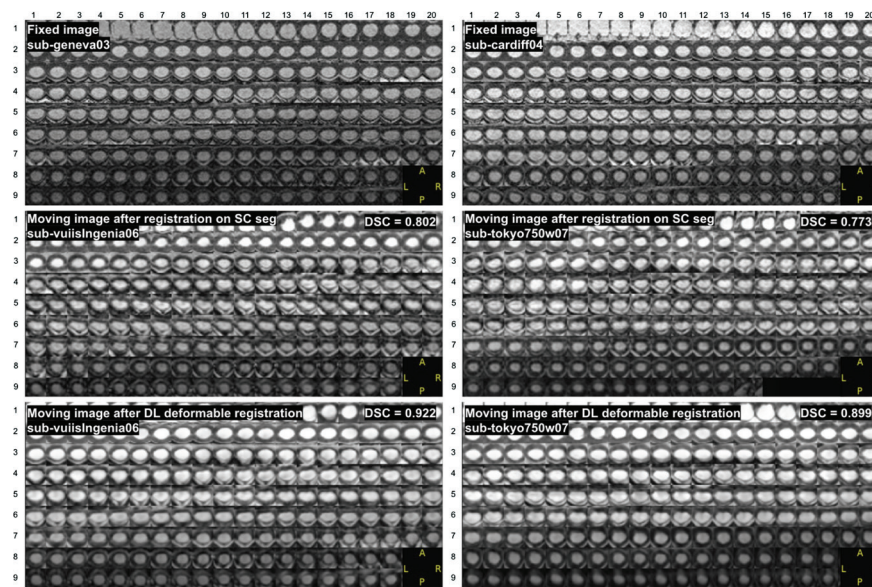
Fig. 13. Results of unimodal inter-subject registration at different stages of the registration process. Each row shows a pair of fixed (T1w MRI contrast of one subject) and moving images (T1w MRI contrast of a different subject) followed by the transformed moving image after each step. The Dice score (DSC) representative of the overlap of the spinal cord segmentations between the fixed and moving 3D image is displayed. The median DSC was obtained from 258 pairs of images.

registration of the entire SC by qualitatively examining the overlap of each axial slice. Figure 14 shows the results for the sub-geneva03/sub-vuiisIngenia06 and sub-cardiff04/sub-tokyo750w07 pairs. The comparison of the second and third groups of images in each column highlight the role of deep-learning models for SC registration. In these two registration examples, we can observe an overall expansion of the SC to match the volume of the fixed image. In addition, in rows 5 and 6 of the axial slices of subvuiisIngenia06, we can see that the deep-learning models perform some sort of rotational transformation to better register the SC. The QC reports also allow us to identify where the registration is not extremely accurate. For example, the slices in row 9 and the last columns in row 8 for sub-tokyo750w07 are not accurately registered to the T1w data of sub-cardiff04. The individual slices appear not to be large enough, suggesting that the expansion applied by the deep-learning models was not strong enough on the inferior part of the SC.

Figure 15 shows the DSC representative of the SC overlap and thus the accuracy of the registration at different stages of the process. Before any registration, these scores are very low or zero for the majority of randomly formed image pairs, due to the inherent differences

between subjects and their positions during the scans, as well as the relative space of the data. The first step in the process, aiming to approximate the alignment of the SC, leads to still very low DSCs (median value of 0.288) due to the articulated nature of the spine, which results in a nonlinear relationship of this anatomical part between subjects. A translation along the right-left and posterior-anterior axes based on the relative positions of the pontomedullary junction decreases the DSCs and accuracy of the SC registration to better align the head and brain structures of the two images and improve the overall registration. The next step in the process significantly improves the accuracy of the SC registration, as the segmentations of this structure are used to perform a slice-by-slice axial alignment. After this step, the median DSC is 0.837. This score is not higher because there is no registration along the inferior-superior axis (transformations are performed only slice by slice) and the transformation for each slice is limited to a simple center-of-mass alignment (e.g., no rotation or scaling). Finally, the application of the cascaded deep-learning registration models leads to very high DSCs (median value of 0.907), which shows the advantages of the deformable registration step even for the accuracy of SC registration. Overall, the DSCs

Fig. 14. Results of the inter-subject unimodal registration quality control report. Axial sections of the spinal cord are displayed to control its registration. The first group of images represents spinal cord axial slices of fixed images while the second and third groups of images represent spinal cord axial slices of moving images at different stages of the inter-subject registration process. Comparison of second and third groups of images shows the effects of the cascaded deep-learning models on spinal cord registration. The Dice score (DSC) representatives of the spinal cord overlap between the fixed and moving/registered images are shown.
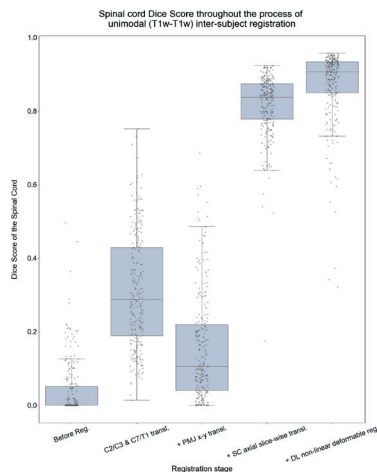
Fig. 15. Accuracy of spinal cord registration (Dice score (DSC)) after the different steps of the inter-subject (unimodal) registration process. Results were obtained on 258 pairs of images. The T1w modality of each subject in the spine-generic dataset was used as the fixed image. The moving image (T1w) was randomly selected for each fixed image.

obtained at the end of the process reflect a robust and accurate registration pipeline, especially considering the difficulty of this task, illustrated by the low DSCs at the beginning of the process.

DISCUSSION

We propose a comprehensive framework for 3D MRI image registration based on deep-learning models trained with the SynthMorph strategy. We extend this method by cascading the registration models to account

for large deformations and evaluate the registration results on SC data for multimodal/unimodal intra-/inter-subject tasks.

The following sections discuss the performance and limitations of the cascaded registration models compared with the benchmarks, as well as potential directions for improvement and generalization of the study after an initial review of the pipeline options and the registration validation strategy.

Registration Pipeline and Validation Strategy

Building on SynthMorph, we have developed a publicly available framework that can adapt to different scenarios with minimal user intervention to provide an easy-to-use, deep learning–based end-to-end registration pipeline that is both faster and more accurate than traditional registration techniques.

A wide spectrum of options is included in the pipeline to allow the user to choose which interpolation methods to use, which MRI modalities to co-register and to provide the ability to process multi-sessions data or perform sub-volume registration. All results presented in this study were obtained directly from the post-registration evaluation tools built into the pipeline. The variety of tools provides an indication of the registration accuracy achieved by the deep-learning models, both quantitatively and qualitatively. Computed SC metrics are a straightforward method to determine the accuracy of SC registration over a large number of registered image pairs and allow comparison of different models/methods based on the registration performance obtained on

large datasets. Visualization of the Jacobian determinant of the warping fields is important in specific registration tasks to determine whether the transformations applied to register the moving image are realistic and relevant from a biological/physical perspective. Finally, the QC reports show the effects of the registration on the entire SC with the representation of all axial sections of the SC for the fixed, moving and registered images, thus qualitatively determining the registration accuracy.

Baseline Comparison

Figures 7 and 8 illustrate the complexity of using traditional registration methods on multimodal SC MRI data. Since there is no specific intensity-based relationship between modalities, the difficulty of quantifying the similarity between the images used to guide the iterative process is increased. Mutual information, which measures the mutual dependence between images, was used in this work because it is considered the gold standard similarity measure in multimodal registration tasks and has been shown to scale to different intensities between modalities (25, 26). However, this metric has some
limitations, as it does not account for spatial information and its performance decreases in the presence of local intensity variations (27). The results obtained with traditional methods tend to confirm the limitations of the mutual information similarity metric with performance that cannot match that of learning-based registration features for the multimodal registration task. Furthermore, even though the parameters used in these methods were selected based on preliminary results on a subset of the data, they may not be optimal for the registration task on that dataset (e.g., smoothing factor, shrinkage factor, gradient step, number of iterations, degree-of-freedom/ deformation constraints). A drawback of these classical registration methods is hence the need to optimize the parameters for each new pair of images to be registered, which increases the difficulty and uncertainty of the registration task. With deep-learning models, parameter optimization is only performed during the network-training phase. Therefore, using the trained models provided in the registration pipeline, any user could directly register their data and obtain consistent (less scattered, as observed in Figure 8) results over the entire dataset without having to change the parameters.
The registration module, consisting of two successive registration models, has been developed to accommodate larger deformations than the original SynthMorph model (smshapes). However, this approach should not lead to a reduction of the registration accuracy on data requiring only minor and localized deformations. The results obtained in Experiment 1 showed that on image pairs consisting of successive MRI scans of different modalities (small movements of the subjects between scans), the cascaded deep-learning models obtained similar

and even slightly better results than smshapes. These results, therefore, tend to confirm the effectiveness of the second model in its role of refining the registration results to ensure high accuracy on localized deformations. In Experiment 2, the random translations, rotations and scaling applied to the moving image before registration complexified the task with initially more distant data. Once again, traditional methods did not provide satisfactory registration, with low median DSCs and large discrepancies in results. For Experiments 1 and 2, highly accurate SC registration results are obtained with the SCSegReg method. This method differs from traditional image-based registration methods, as it relies on SC segmentations to register data, which explains the extremely good registration results evaluated using overlapping SC segmentations. The disadvantage of this method compared with the cascaded deep-learning approach is its need for SC segmentations to optimize the registration process. In addition, this method only registers the SC, but not the other structures or tissues. Therefore, for global registration, deep-learning approaches should be preferred.
The comparison of the results obtained with smshapes and the cascaded approach reflects the advantages of using cascaded models for deep-learning registration. The latter leads to a higher median DSC and fewer misregistered image pairs. It shows that the first model, trained using smoother deformation fields, is useful for coarse image registration before the second model refines the registration of smaller structures. However, it appears that the registration is somewhat less accurate (low DSCs, Figure 9) for some of the 267 image pairs registered, which also shows some limitations of this method when both images to co-register are initially too far apart.

Limits on Multimodal Intra-Subject Registration

The results observed in Experiment 3 suggest that the limits of cascaded models appear mainly when the data are heavily translated. The study of the independent impact of the different transformations (rotation, scaling and translation) on the accuracy of the intra-subject multimodal registration, performed in Experiment S2, confirms the detrimental effect of translation on the registration accuracy. The cascaded models support translation up to 12 to 14 voxels (1.2 to 1.4 cm), but larger differences in the data to be registered result in a significant decrease in registration performance (Figure S4C). However, the results show a significant improvement over smshapes (single model) for which SC registration accuracy is already reduced after translations with norm vectors of 6–8 voxels (0.6–0.8 cm).
Different methods could be considered to improve the registration of highly translated data. The first model of the cascaded approach could be trained on even

stronger deformations than was done by increasing the parameter that determines the maximum standard deviation for drawing the Gaussian distribution of noise in the unregistration process of the label map pairs (*vel-std*). This could allow the former model to handle larger deformations (especially translations) but potentially at the cost of reduced registration performance when only small transformations are required. Zhao et al. (10) showed that, with their subnetworks registration strategy, the most accurate registration results from multimodal brain and liver data were obtained using three successive subnetworks. Therefore, similar to what was observed in this study, we could expect an increase in registration accuracy by cascading more than two deep-learning models. Due to the composition of the warping fields, this process would not increase the interpolation error.

Inter-Subject Registration

To counteract the limitations observed when the data have very little in common, additional steps were added to the pipeline for the inter-subject registration task. These steps require labeling of the pontomedullary junction and two intervertebral discs to perform a coarse alignment of the brain and spine structures of the two volumes and to ensure that the deep-learning models achieve accurate registration. Therefore, the inter-subject registration task is somewhat more restrictive than the intra-subject one. However, the required labels do not need to be precise, meaning that either they could be quickly defined by the user or an algorithm could be developed to automatically detect these areas. In addition, they allow for fast registration results, as they do not require an iterative image-based process and are not dependent on the modality of the image. That is why this approach was preferred over using affine or rigid image-based preregistration (which is time-consuming and not always accurate for multimodal data), or the direct use of cascaded deep-learning registration models (which can lead to inaccurate registration if the data are too far apart).

Experiment 3 showed very good registration results for unimodal data (Figures 13 and 15) with significant improvement in global and SC registration after using the cascaded deep-learning models. With multimodal data, the positive effect of the deformable registration step is less pronounced, with SC registration deteriorating in some examples (Figures S2 and S3). The inter-subject multimodal registration process is extremely complex because the appearance of the anatomical structures to be registered differ in the two 3D images, due to the image acquisition process, the inherent properties of the subjects and the highly deformable structure of the spine. Taking these aspects into account, the registration results are satisfactory, although better results are obtained by limiting the registration to one modality (inter-subject, unimodal) or to one subject (intra-subject, multimodal).

Perspectives

We extended the work done by Hoffmann et al. (9) in SynthMorph by cascading the registration models and evaluating the performance of SC data. The results obtained in the different experiments showed that SynthMorph is a suitable deep-learning registration strategy for SC images. Considering that SynthMorph was initially tested on brain and cine-cardiac data by Hoffmann et al., the registration pipeline and cascaded models strategy should work for registration of images of different anatomical structures. An extension of this study would therefore be to test the pipeline and the cascaded registration models on the registration of different body parts. This would include the integration of additional assessment tools to qualitatively and quantitatively evaluate the results of the registration, as has been done for the SC.

The inter-subject registration task was evaluated by forming random pairs of subjects from the dataset and obtained generally accurate registration results. This is encouraging for another interesting registration task that was not explored in this work: registration to template. Registering data from different subjects in a common reference space allows calculation of morphometry, functional analyses and is useful for reproducibility and for performing multicenter studies. For SC data, registration can be performed on the PAM50 template (28). Accurate template-based registration is essential to transpose the available masks and labels (e.g., white/gray matter mask, pointwise vertebral body labels, pointwise intervertebral disc labels) to the data of interest and perform additional analyses and comparisons between different subjects. It would therefore be interesting to extend the pipeline and provide the ability to perform registration to template for a complete dataset.

We evaluated multimodal and unimodal registration using T1w and T2w MRI modalities. However, SynthMorph has been developed to be contrast agnostic and to obtain accurate registration results regardless of the modality used. Therefore, it would be interesting to use the provided pipeline to register pairs of images from different MRI modalities than T1w and T2w and to evaluate the similarities or discrepancies in the resulting registration. If additional modalities are present in a dataset, they can be used in the pipeline by simply changing the contrasts to co-register at the beginning of the script.

In this study, we used a heterogeneous dataset consisting of multicenter data. The distribution and number of data points (267 subjects) were therefore quite large and tend to confirm that the observed results are not distribution specific, a characteristic of the SynthMorph method. However, it would be interesting to evaluate and compare the registration results on other datasets to better assess the generalizability of the cascaded models performance. Ideally, a reference dataset should be used so that the registration results obtained in this work can be compared with the results obtained in other studies.

The developed pipeline can be used with any dataset that follows the BIDS convention. Because the registration models are not specific to a training dataset, these can be applied to other structures, such as brain, heart or liver, although the performance of the registration was only assessed for SC data. In addition, the present evaluation was only performed in healthy individuals; hence, it would be important to extend the evaluation in patients with SC pathologies, given that multimodal MRI can help improve the diagnosis and prognosis of some SC conditions.

CONCLUSION

Multimodal SC registration is a challenging problem due to the ill-conditioned nature of deformable registration, differences in appearance of anatomical structures caused by multimodal image acquisition, and the articulated geometry of the SC. Our study establishes an end-to-end solution based on deep-learning models for the registration of 3D multimodal MRI data, using a cascaded model strategy to account for larger deformations and expand the range of registration possibilities. We demonstrate that this solution provides accurate registration in a variety of scenarios: multimodal intra-subject registration, unimodal inter-subject registration and multimodal inter-subject registration. The developed pipeline is freely accessible as a standalone Python package and will hopefully be useful for application to other datasets and for baseline comparison with modified versions of the architecture.

REFERENCES

1. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage. 2011 Feb 1;54(3):2033–44.
2. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix: a toolbox for intensity-based medical image registration. IEEE Trans Med Imaging. 2010 Jan;29(1):196–205.
3. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. VoxelMorph: a learning framework for deformable medical image registration. IEEE Trans Med Imaging. 2019 Aug;38(8):1788–800.
4. Cao X, Yang J, Zhang J, Nie D, Kim M, Wang Q, et al. Deformable image registration based on similarity-steered CNN regression. Med Image Comput Comput Assist Interv. 2017 Sep;10433:300–8.
5. Sokooti H, de Vos B, Berendsen F, Lelieveldt BPF, Išgum I, Staring M. Nonrigid image registration using multi-scale 3D convolutional neural networks. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S, editors. Medical Image Computing and Computer Assisted Intervention − MICCAI 2017. Cham: Springer International Publishing; 2017. p. 232–9.
6. Fu Y, Lei Y, Wang T, Curran WJ, Liu T, Yang X. Deep learning in medical image registration: a review. Phys Med Biol. 2020 Oct 22;65(20):20TR01.
7. Jaderberg M, Simonyan K, Zisserman A, Kavukcuoglu K. Spatial transformer networks. arXiv:150602025 [cs] [Internet]. 2016 Feb 4 [cited 2022 Mar 15]. Available from: http://arxiv.org/abs/1506.02025
8. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. arXiv:150504597 [cs] [Internet]. 2015 May 18 [cited 2022 Mar 15]. Available from: http://arxiv.org/abs/1505.04597
9. Hoffmann M, Billot B, Greve DN, Iglesias JE, Fischl B, Dalca AV. SynthMorph: learning contrast-invariant registration without acquired images. IEEE Trans Med Imaging. 2022 Mar;41(3):543–58.
10. Zhao S, Lau T, Luo J, Chang EI-C, Xu Y. Unsupervised 3D end-to-end medical image registration with Volume Tweening Network. IEEE J Biomed Health Inform. 2020 May;24(5):1394–404.
11. Islam KT, Wijewickrema S, O'Leary S. A deep learning based framework for the registration of three dimensional multi-modal medical images of the head. Sci Rep. 2021 Jan 21;11(1):1860.
12. Liu X, Jiang D, Wang M, Song Z. Image synthesis-based multi-modal image registration framework by using deep fully convolutional networks. Med Biol Eng Comput. 2019 May 1;57(5):1037–48.
13. Shao W, Banh L, Kunder CA, Fan RE, Soerensen SJC, Wang JB, et al. ProsRegNet: a deep learning framework for registration of MRI and histopathology images of the prostate. Med Image Anal. 2021 Feb;68:101919.
14. McKenzie EM, Santhanam A, Ruan D, O'Connor D, Cao M, Sheng K. Multimodality image registration in the head-and-neck using a deep learning derived synthetic CT as a bridge. Med Phys. 2020 Mar;47(3):1094–104.
15. Gorgolewski KJ, Auer T, Calhoun VD, Craddock RC, Das S, Duff EP, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci Data. 2016 Jun 21;3(1):160044.
16. Modat M, Ridgway GR, Taylor ZA, Lehmann M, Barnes J, Hawkes DJ, et al. Fast free-form deformation using graphics processing units. Comput Methods Programs Biomed. 2010 Jun;98(3):278–84.
17. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: a system for large-scale machine learning. In: Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation. USA: USENIX Association; 2016. p. 265–83. (OSDI'16).
18. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv:14126980 [cs] [Internet]. 2017 Jan 29 [cited 2022 Mar 15]. Available from: http://arxiv.org/abs/1412.6980
19. Cohen-Adad J, Alonso-Ortiz E, Abramovic M, Arneitz C, Atcheson N, Barlow L, et al. Open-access quantitative MRI data of the spinal cord and reproducibility across participants, sites and manufacturers. Sci Data. 2021 Aug 16;8(1):219.
20. Cohen-Adad J, Alonso-Ortiz E, Abramovic M, Arneitz C, Atcheson N, Barlow L, et al. Generic acquisition protocol for quantitative MRI of the spinal cord. Nat Protoc. 2021 Oct;16(10):4611–32.
21. De Leener B, Lévy S, Dupont SM, Fonov VS, Stikov N, Louis Collins D, et al. SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. NeuroImage. 2017 Jan 15;145:24–43.
22. Gros C, De Leener B, Badji A, Maranzano J, Eden D, Dupont SM, et al. Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. NeuroImage. 2019 Jan 1;184:901–15.
23. Bautin P, Cohen-Adad J. Minimum detectable spinal cord atrophy with automatic segmentation: investigations using an open-access dataset of healthy participants. Neuroimage Clin. 2021 Oct 4;32:102849.
24. Cohen-Adad J, Lévy S, Avants B. (ISMRM 2015) Slice-by-slice regularized registration for spinal cord MRI: SliceReg [Internet]. [cited 2022 Mar 15]. Available from: https://archive.ismrm.org/2015/4428.html
25. Viola P, Wells III WM. Alignment by maximization of mutual information. Int J Comput Vis. 1997 Sep 1;24(2):137–54.
26. Collignon A, Maes F, Delaere D, Vandermeulen D, Suetens P, Marchal G. Automated multi-modality image registration based on information theory. In: Bizais. 1995.
27. Woo J, Stone M, Prince JL. Multimodal registration via mutual information incorporating geometric and spatial context. IEEE Trans Image Process. 2015 Feb;24(2):757–69.
28. De Leener B, Fonov VS, Collins DL, Callot V, Stikov N, Cohen-Adad J. PAM50: unbiased multimodal template of the brainstem and spinal cord aligned with the ICBM152 space. NeuroImage. 2018 Jan 15;165:170–9.
29. Studholme C, Hill DLG, Hawkes DJ. An overlap invariant entropy measure of 3D medical image alignment. Pattern Recog. 1999 Jan 1;32(1):71–86.

Supplementary Methods

Registration on sub-volumes

To reduce the required computer resources (amount of RAM) and to accommodate large images, it is possible to perform the registration on sub-volumes. Therefore, if the choice of using sub-volumes is made via the configuration file of the registration pipeline, an algorithm is executed to create sub-volumes, register them and concatenate the results to obtain a deformation field in the original dimension. The abrupt transitions (discontinuities) resulting from concatenation are limited by a weighted average applied to the displacement vectors of the overlapping areas.

Post-registration analysis tools

Normalized mutual information

The normalized mutual information (NMI) (29) before and after registration is reported in a CSV file aggregating this information for each subject of the dataset.

$$NMI(F, M) = \frac{H(F) + H(M)}{H(F, M)}$$

where $H(F)$ and $H(M)$ represent the marginal entropies of the fixed and moving (or moved) images and $H(F, M)$ the joint entropy.

Spinal cord segmentation

Using the SCT (21) and the *sct_deepseg_sc* framework developed by Gros et al. (22), SC segmentations are computed for the moving, fixed and registered images. The segmentations are saved in separate files and can be used to qualitatively or quantitatively evaluate the registration.

Metrics on spinal cord segmentation

As shown by Gros et al. (22), the developed framework for SC segmentations handles the heterogeneity of image acquisition features well, providing accurate SC segmentations on a multicenter, multiresolution, multi-contrast dataset. Therefore, SC segmentations of the moving, fixed and registered images are reliable and can be used to calculate volumetric overlap metrics to evaluate registration in a quantitative approach, focusing on SC. Different metrics such as the Dice similarity coefficient (DSC, DSC, F1) or the Jaccard index (intersection over union) are computed to indicate the extent to which the spinal cords overlap before and after registration. The measurements obtained for each subject are attached in a CSV file aggregating this quantitative analysis of the registration for the entire dataset.

$$DSC(Fseg, Mseg) = \frac{2 \, | \, Fseg \cap Mseg \, |}{| \, Fseg \, | + | \, Mseg \, |}$$

Where $|Fseg|$ and $|Mseg|$ represent the number of voxels in the fixed and moving (or moved) image segmentation. And $|Fseg \cap Mseg|$ represents the number of voxels in common in both segmentations.

$$Jacc(Fseg, Mseg) = \frac{| \, Fseg \cap Mseg \, |}{| \, Fseg \, | \cup | \, Mseg \, |}$$

Where $|Fseg \cup Mseg|$ represents the union of both segmentations.

Quality control report

Using the functionality developed in SCT (21), a QC report (HTML format) is produced to qualitatively assess the registration by visualizing the SC at each axial slice of the 3D images. These reports allow switching between the fixed and moving images and between the fixed and registered images. Therefore, they provide an easy way to evaluate the registration and transformation that has been applied along the SC. In addition, a QC report is created to observe the SC segmentations and ensure that the segmentations are accurate, indicating that the quantitative analysis of the registration performed on these segmentations is meaningful.

Jacobian determinant

The regularity of the warping fields is evaluated using the Jacobian matrix, which captures the local properties of a field around a voxel. The number and percentage of folding voxels (negative Jacobian determinant), where the deformation is not diffeomorphic, are reported in a CSV file. A 3D volume representative of the Jacobian determinant is also saved to easily observe how the moving volume has been transformed to be registered with the fixed volume.

Data

In this study, the experiments were performed on the T1w and T2w contrasts. T1w 3D images ranged in size from 60 to 192 voxels (60–192 mm) in the right-left direction, 170 to 320 voxels (170–320 mm) in the posterior-anterior direction and were fixed at 320 voxels (320 mm) in the inferior-superior direction. T2w images ranged from 56 to 64 voxels (45–51 mm) in the right-left direction and were fixed at 320 voxels (256 mm) in the posterior-anterior and inferior-superior directions.

Experiment S1: Inter-subject, multimodal

Using the same process as in Experiment 3, pairs of images, consisting of a T1w MRI image as a fixed image and a T2w MRI image of another randomly selected subject as a moving image, are registered. A qualitative analysis showing typical results and a quantitative analysis based on the DSC of SC segmentations are performed.

The results were computed on 223 pairs of images. From the spine-generic dataset, 267 image pairs were randomly formed, but 44 of them were not used because of the absence of intervertebral labels or the non-detection of the pontomedullary junction in one or both images forming the pair.

Experiment S2: Limits of deep-learning models, multimodal, intra-subject

In this experiment, the registration accuracy is evaluated as a function of gradually affine-transformed (rotation, scaling or translation) MRI to measure the robustness of the cascaded approach on affinely displaced data. The registration results obtained using the two models in succession are compared with those obtained using only smshapes by looking at the difference in DSC as a function of the amplitude of the applied transformation before registration.

The T2w transformed image is registered to the T1w image for each subject in the spine-generic dataset. The transformation applied to the T2w image ranges from −10 to 10 degrees for the study of the effect of rotation. For the study of the scale effect, the scale factor ranges from 0.9 to 1.1 and for the study of translation, the T2w image is translated from −10% to 10% of the field of view in each of the three dimensions independently.

Experiment S3: Computational time comparison

The computational time to register new image pairs is compared between the deep-learning registration method (cascaded models) and the classical iterative registration methods, RigSlicereg and RigSliceregSyn. The parameters of the classical iterative registration methods are the same as those specified in the Baselines section. The comparison is performed on the registration part (no preprocessing or postprocessing is included as these parts are common to the different approaches). A subset of 6 T1w and T2w MRI data pairs (sub-vallHebron06, sub-geneva06, sub-strasbourg01, sub-ubc01, sub-sherbrooke05, sub-vuiisIngenia04) from the spine-generic dataset is used for this study. Three pairs are formed with images of size $192 \times 256 \times 320$ and the remaining pairs are formed with images of size $192 \times 320 \times 320$. The comparison of computation time is performed on the intra-subject multimodal registration task and on a 64-core CPU machine. The range of computation time for each method is reported.

Supplementary Results

Experiment S1

Inter-subject multimodal registration is a truly complex process with differences in the anatomical structure of the images to be registered caused both by the image acquisition process and by the inherent properties of the subjects. Figure S2 shows the registration results, at different stages of the process, obtained on a subset of four pairs. The transformations performed by the deep-learning registration models can be observed by comparing the last two columns. In this multimodal registration task, we can see that the benefits of applying the deep-learning models are mainly observed on the brain and the intervertebral discs. Registration of the

SC and surrounding cerebrospinal fluid is not always enhanced by deep-learning models, which can create a ripple effect in the SC area, as observed with sub-oxfordFmrib09/sub-juntendo750w04 and sub-cardiff04/sub-sherbrooke07 pairs, compared with the registration results observed after axial slice-by-slice SC registration. Although the results do not show a perfect match between the fixed image and the final registered image, they are pretty satisfactory considering the complexity of the task.

Figure S3 shows some similarities with the results obtained for unimodal inter-subject registration (Figure 15, Experiment 3). The main difference appears for the DSCs obtained after the registration step with cascaded deep-learning models. For unimodal registration, a significant improvement in the SC registration accuracy was observed with the DSCs. With multimodal data, the sparse distribution of DSCs, after deep-learning registration, indicates that this step does not always improve SC registration, as also seen in Figure S2. The median DSC at the end of the inter-subject multimodal registration process is 0.849 (0.907 for unimodal registration).

Experiment S2

Figure S4 compares registration accuracy for the cascaded and smshapes models as a function of the linear transformation applied to the moving image before registration. Each sub-graph represents the pairwise difference between the DSC obtained on the normal data with only small movements between scans (Experiment 1) and the DSC obtained when registering the moving image transformed before registration. The affine transformation applied to the moving image is indicated on the x axis.

Considering the rotational transformation (Figure S4A), the registration performance starts to decrease when the moving images are rotated by an absolute value of 5 degrees or more before the registration. The effect of rotation is better handled with the cascaded approach with a decrease in the DSC that is less pronounced than for the smshapes method. The difference between the methods is statistically significant for degrees of rotation ranging from −10 to −6 degrees ($p < 2 \times 10^{-4}$), from −6 to −4 degrees ($p < 0.02$) and from 8 to 10 degrees ($p < 7 \times 10^{-5}$).

Expanding or shrinking the data from 0% to 10% before registration results in only a small decrease in the registration performance for both methods (maximum median DSC decrease of 0.010 and 0.012 for cascaded and smshapes models, respectively) (Figure S4B). The difference between the methods is only significant for an expansion ranging from 8% to 10% ($p < 0.04$).

It was observed in Experiment 2 that translation is the transformation with which the models fail to generalize well. This observation is confirmed with Figure S4C, which shows an important decrease in DSC with increasing norm of translation. While cascaded networks delay the decrease in the registration performance compared

with smshapes, they also show some limits for registering images that were transformed with translation vectors with a norm greater than 14.5 voxels before registration. The difference between the cascaded and smshapes models is statistically significant for translations from 5 to 16.5 voxels ($p < 0.03$).

Experiment S3

The computational time was compared between three methods (cascaded models, RigSlicereg and RigSliceregSyn) for the intra-subject multimodal registration task. For the cascaded deep-learning models, registration for $192 \times 256 \times 320$ size images took 78.3 to 82.4 seconds and for $192 \times 320 \times 320$ size images, 97.1 to 103.4 seconds. The RigSlicereg iterative method, which achieved the most accurate registration results among the traditional registration methods in Experiments 1 and 2, is 23.5 to 24.2 (1893 to 1934 seconds) and 21.3 to 26.3 (2141 to 2554 seconds) slower than the cascaded models for the $192 \times 256 \times 320$ and $192 \times 320 \times 320$ size images, respectively. Adding a deformable registration step with SyN (RigSliceregSyn) increased the computation time from 256 to 310 seconds, when this step is performed on a volume downsampled by a factor of 2, or from 2023 to 4101 seconds when no prior downsampling of the 3D image is performed. Therefore, the comparison between cascaded deep-learning models and a traditional method including deformable registration (RigSliceregSyn) computed on a downsampled volume shows that deep-learning models register images 23.8 to 30 times faster on CPUs than a classical registration method based on iterative optimization.

Supplementary Discussion

Sub-volumes feature

The ability to perform registration on sub-volumes before spatially concatenating the resulting sub-warping fields is a critically important feature. This is because the registration of large 3D images using complex deep-learning models requires a large amount of RAM (sometimes over 32 GB), which can be a limitation for standalone laptops. Using sub-volumes significantly reduces the amount of RAM required for registration. For example, registering a pair of volumes of size $192 \times 256 \times 320$ requires about 21–22 GB of RAM when registering the entire volume directly while it is reduced to about 12, 4 and 3 GB of RAM using sub-volumes of size $160 \times 160 \times 192$, $80 \times 80 \times 96$ and $64 \times 64 \times 64$, respectively. Therefore, this feature should be used if computer resources are limited. Otherwise, it is recommended to use whole volumes as inputs to the models to avoid a potential decrease in registration accuracy due to the smoothing effect in the overlapping areas of the sub-volumes caused by concatenating the sub-warping fields using a weighted average of the displacement vectors.

Supplementary Figures and Tables
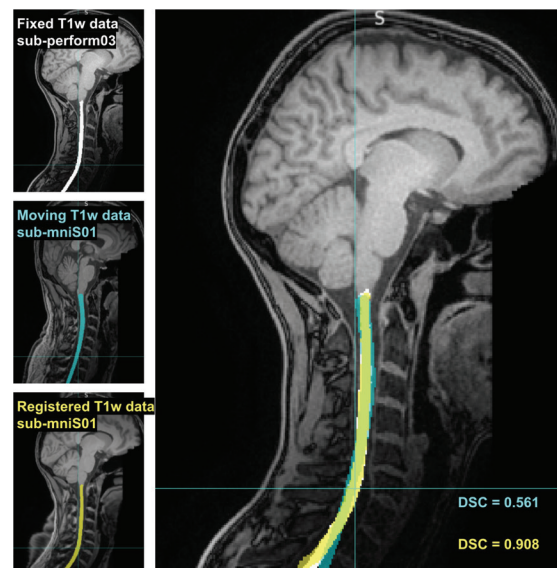
Figure S1. Example of unimodal inter-subject registration with spinal cord segmentations and their associated Dice score (DSC). DSCs are computed between the fixed image (white spinal cord segmentation) and the moving or registered images (blue and yellow spinal cord segmentations, respectively). Before registration (in blue), the DSC is 0.561. This score increases to 0.908 after registration by the cascaded deep-learning models (in yellow).

Table S1. Parameters used to train the deep-learning registration models, using SynthMorph.

| Hyperparameter	Description	Involved in	Final value
in_shape	Dimension of label maps produced. Dimension of the images on which the model is trained.	Label maps generation	160 × 160 × 192
num_labels	Number of different labels in the label maps.	Label maps generation	26
num_maps	Number produced of different label maps	Label maps generation	125
im_scales	Relative resolutions at which noise is sampled in the first step of the label maps generation.	Label maps generation	[16, 32, 64]
def_scales	Relative resolutions at which noise is sampled to generate warping fields and deform the label map.	Label maps generation	[8, 16, 32]
def_max_std	Maximum standard deviation for the Gaussian distribution of noise to generate warping fields and deform the label map.	Label maps generation	3
same_subj	Generate image pairs from the same label map.	Label maps generation	True
vel_res	Relative resolutions at which noise is sampled to generate warping fields and deform the label maps to create an unregistered pair.	Label maps generation	[32, 64] for model 1 16 for model 2
vel_std	Maximum standard deviation for the Gaussian distribution of noise to generate warping fields and deform the label maps to create an unregistered pair.	Label maps generation	3
blur_std	Maximum blurring standard deviation applied on the grayscale images.	Grayscale images generation	1
gamma	Gamma augmentation strength (standard deviation).	Grayscale images generation	0.25
bias_std	Bias field range (standard deviation).	Grayscale images generation	0.3
bias_res	Bias field relative resolution.	Grayscale images generation	40
Epochs	Number of training epochs.	Training process	2000
batch_size	Mini-batch size.	Training process	1
train_frac	Fraction of the label maps that are included in the training dataset. The other label maps are included in the validation dataset.	Training process	0.8
batch_size_val	Mini-batch size for validation dataset.	Training process	1
reg_param	Regularization weight (lambda).	Training process	1
Lr	Learning rate.	Training process	5×10^{-5}
int_steps	Number of integration steps.	Network architecture	5
int_res	Relative resolution of the flow field during vector integration.	Network architecture	2
svf_res	Relative resolution of the predicted stationary velocity field.	Network architecture	2
Enc	U-Net number of convolutional filters per encoder block.	Network architecture	[256, 256, 256, 256]
Dec	U-Net number of convolutional filters per decoder block.	Network architecture	[256, 256, 256, 256, 256, 256]
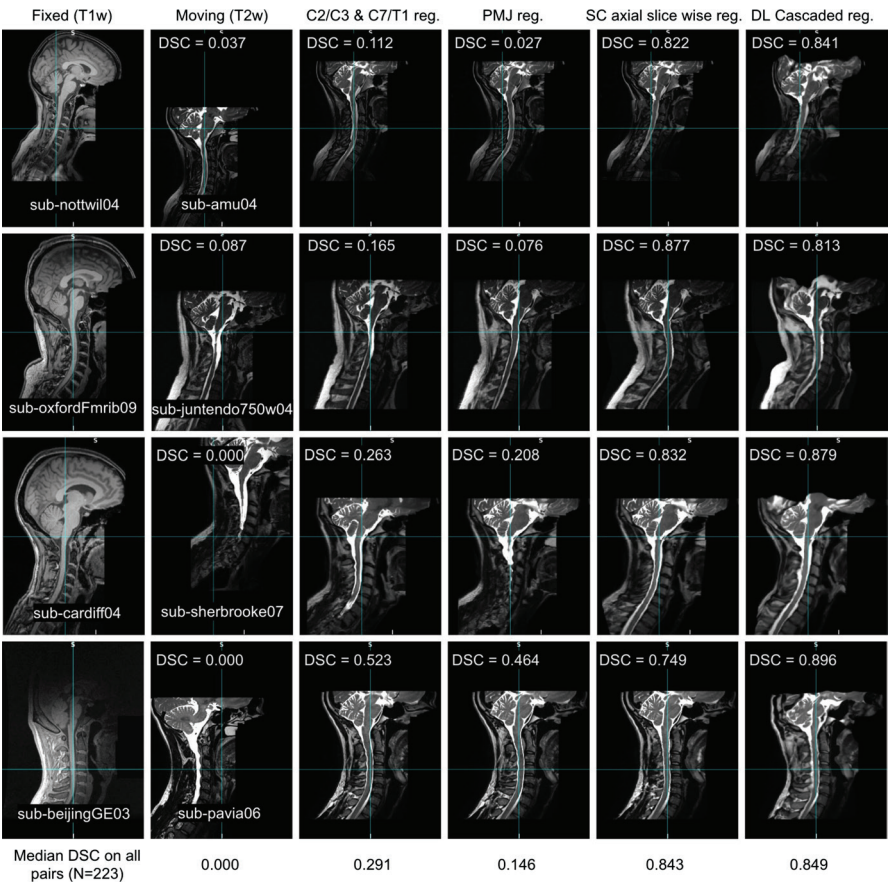
Figure S2. Results of inter-subject multimodal registration at different stages of the registration process. Each row shows a pair of fixed (T1w MRI contrast of one subject) and moving images (T2w MRI contrast of a different subject), followed by the transformed moving image after each step. The median Dice score (DSC) was obtained from 223 image pairs.
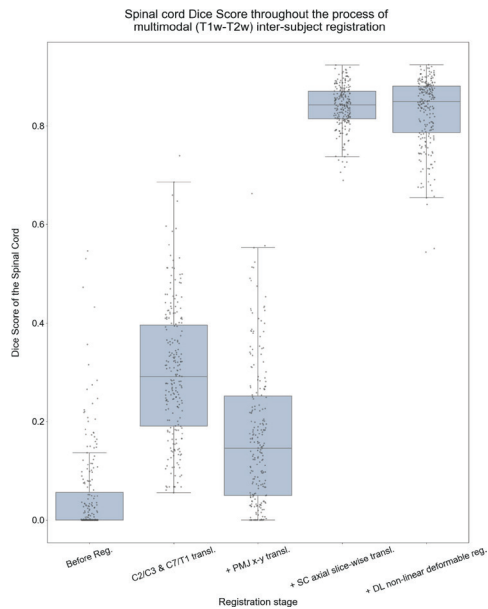
Figure S3. Accuracy of spinal cord registration (Dice score (DSC)) after the different steps of the inter-subject (multimodal) registration process. Results were obtained on 223 pairs of images.
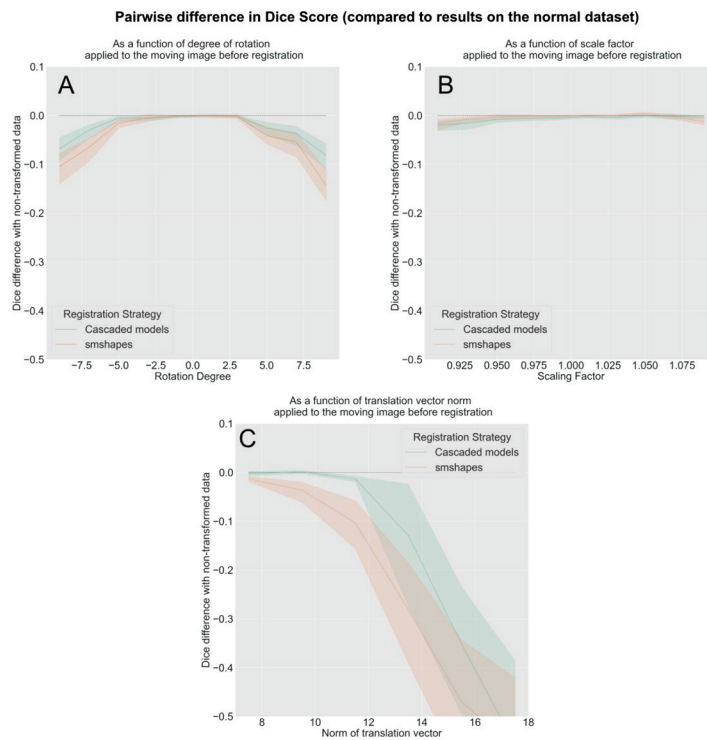
Figure S4. Pairwise difference in Dice score (DSC) (compared with results on the normal dataset) as a function of (A) degree of rotation, (B) scale factor and (C) translation vector norm, applied to the moving image before registration. The results are aggregated on bins of (A) 2 degrees, (B) 2% of scaling and (C) 2 voxels, to represent the mean and the 95% confidence interval (computed with bootstrapping) of the DSC change as a function of transformation intensity. Results are shown in green for the cascaded models and orange for smshapes.