SVF-Net: Learning Deformable Image Registration Using Shape Matching

Marc-Michel Rohé[1(✉)], Manasi Datar[2], Tobias Heimann[2], Maxime Sermesant[1], and Xavier Pennec[1]

[1] Université Côte d'Azur, Inria, Sophia-Antipolis, France
[2] Medical Imaging Technologies, Siemens Healthcare Technology Center, Erlangen, Germany

Abstract. In this paper, we propose an innovative approach for registration based on the deterministic prediction of the parameters from both images instead of the optimization of a energy criteria. The method relies on a fully convolutional network whose architecture consists of contracting layers to detect relevant features and a symmetric expanding path that matches them together and outputs the transformation parametrization. Whereas convolutional networks have seen a widespread expansion and have been already applied to many medical imaging problems such as segmentation and classification, its application to registration has so far faced the challenge of defining ground truth data on which to train the algorithm. Here, we present a novel training strategy to build reference deformations which relies on the registration of segmented regions of interest. We apply this methodology to the problem of inter-patient heart registration and show an important improvement over a state of the art optimization based algorithm. Not only our method is more accurate but it is also faster - registration of two $3D$-images taking less than 30 ms second on a GPU - and more robust to outliers.

1 Introduction

Non-linear registration - the process of finding voxel correspondence between pair of images - is a key instrument in computational anatomy and has gained an increasing importance in the past years. Traditional methods to find the optimal deformation field mapping two images rely on the optimization of a matching criteria controlling the local correspondence of the voxel intensities. These methods usually have several drawbacks: their high computational cost (time and memory) as they often requires many iterations and evaluations of the energy function [3] and also the possibility of the optimization to remain stuck in a local minimum because of the non-convexity of the objective function.

New approaches to predict registration parameters based on machine learning have been proposed. In particular, Convolutional Neural Networks have set new standards where there is a need to predict highly non-linear function. Whereas these methods have gained large popularity for medical image segmentation and classification, they are still underrepresented in the context of image registration

M. Descoteaux et al. (Eds.): MICCAI 2017, Part I, LNCS 10433, pp. 266–274, 2017.
DOI: 10.1007/978-3-319-66182-7_31

due to the difficulty to provide a ground truth mapping between pairs of images. While it is possible for a human to classify an image or to draw the contours of the segmentation, the task of defining pairwise dense voxel matching is very difficult, especially when the correspondances have to be found in 3D images. Therefore, to train learning-based methods, we have to find alternative ways to define ground truth. One way is to compute synthetic images deformed with known transformations, but it is hard for these images to account for the inherent noise and artifacts present between pairs of medical images, often leading to over-simplistic synthetic examples which would not be as challenging as real images. In [11], the prediction is trained based on ground truth provided by a registration algorithm previously run on pair of images. However, the problems seen in the algorithm used for computing the ground truth will likely be learned by the learning method if one does not add additional information.

We chose to take another approach and train on reference deformations defined from the registration of previously segmented region of interests instead of the result of a registration algorithm on the images. Therefore, our approach does not try to replicate a classical registration algorithm matching voxel intensities, but learn to align more global and structural information of the images. This choice is supported by the fact that a good matching of intensities is not always correlated with physiologically plausible deformations. In a classical optimization approach, it is difficult to avoid minimizing a matching criteria based on the difference of voxel intensities. However, learning-based methods give us the opportunity to learn on different type of information than intensities of the images. For example, in the context of inter-patients cardiac registration, one is mostly interested in getting a very accurate matching of the contours of both myocardiums rather than a good intensity matching so it seems natural to use this information to define our references deformations. These deformations are then used to train a fully convolutional network building our registration method.

The article is organized as follows. We first present the method to derive ground truth deformations between pairs of images using segmented shapes. These deformations are encoded by dense Stationary Velocity Fields (SVF) [1]. We then present our SVF-Net architecture, a fully convolutional network adapted to the task of registration trained on the previously computed SVFs. Finally, we validate our method by comparing with a state of the art registration algorithm [8] on a large database of 187 cardiac images from multiple clinical centers. We show that, not only the accuracy is increased, but also it is more robust and faster.

The contributions of this paper include:

- A method for computing reference transformations between pair of images, using mesh segmentations which are registered in the space of currents.
- As the shapes can be defined or corrected manually, our method provides an efficient way to introduce manual input in the computation of the deformations to train a learning-based registration algorithm. This is not the case with approaches that rely on synthetic images or the result of a registration algorithm

on the images. This approach comes nearer to the classical definition of a ground truth defined manually.
- A fully convolutional neural network for 3D registration prediction. Our architecture is able to detect global features and deformations that could not be detected with a sliding-window approach (for ex. [11]). It also proves to be faster at testing time as only one pass of the whole image is required.
- A more robust and faster registration method validated on a large database of 187 segmented cardiac images.

2 Methods: Learning Image Deformations

Modeling References Deformations from Shapes. In this section, we detail our methodology to derive a reference deformation mapping two shapes together. We place ourselves in the context of shapes defined by surfaces, as this is a traditional output of segmentation algorithms, but the method defined here is generic and could also be applied to other types of data structures such as point clouds, landmarks, curves and volumes.

From a database of N images I_n where the region of interests have been segmented, we consider the segmentations as surfaces S_n and we register these surfaces to a common template T (chosen as one of the segmented shape of our dataset) giving us a deformation φ_n mapping the template to each of the segmented shape. To do so, the framework of currents [5] provides an elegant mathematical method in order to treat the problem as true surface matching without the point correspondence issue. Each point p_k of the template T can then be though as a landmark which is mapped with the deformation $\varphi_n(p_k)$ into each of the surface meshes S_n of our database. Then, for all pair of images (I_i, I_j) the pair $(\varphi_i(p_k), \varphi_j(p_k))$ defines a point correspondence.

The point correspondence between pair of images gives us a displacement field defined for the set of landmarks. To interpolate it to something defined on the whole image grid an elastic body spline interpolation is used. The elastic body spline is a 3-D spline that is based on a physical model (the Navier equations) of an elastic material [4]. This interpolation is driven by a physical model, making it a natural choice for regions where no landmarks are found. We obtain a displacement field $u_{i,j}$ defined on the whole image grid which parametrizes a transformation that maps the landmarks $(\varphi_i(p_k), \varphi_j(p_k))$.

One of the limitations of the parametrization with displacement fields is the lack of constraints to ensure that the transformations computed are invertible and smooth. In order to illustrate the generecity of the method, we propose to change the parametrization into a diffeomorphic one. We use Stationary Velocity Fields (SVF) [1] but the method could also be adapted to other choices of diffeomorphic parametrization such as time-varying velocity fields [2] and B-splines. The SVFs are extracted by using an iterative log-approximation scheme with the scaling and squaring approach [1] starting with the displacement field u. An example of the resulting SVF can be seen in Fig. 1.

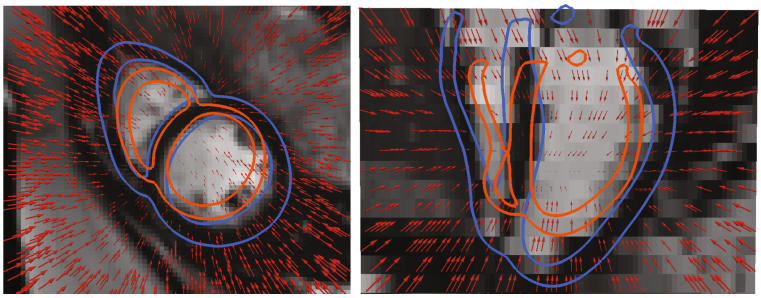

Fig. 1. Example of a reference deformation SVF (red vector field scaled at 0.3) computed from two segmented surfaces. The moving image is shown with the segmentation of the myocardium of the fixed (orange) and moving (blue) images. (Left): Short-axis view. (Right): Longitudinal view.

SVF-Net: Fully Convolutional Neural Network Architecture. Convolutional Neural Networks (CNN) provide a very efficient way to learn highly non-linear functions. Recent methods to apply CNN to the task of registration tackle the problem in a patch-based approach [11], which are easy and fast to train. A side effect is that we are looking only locally at each patch and therefore we might miss global information about the transformation. For image segmentation, fully convolutional networks [7] have been developed in order to process the whole image in a stream, therefore having the advantage to also look at global features, instead of looking only locally at each patch. It also has the benefit to be faster at test time as there is only one prediction to perform for the whole image instead of predicting each patch individually in a sliding-window approach. An important contribution of our work is to adapt the fully convolutional architecture to the task of registration prediction by training on the dense ground truth SVF previously computed.

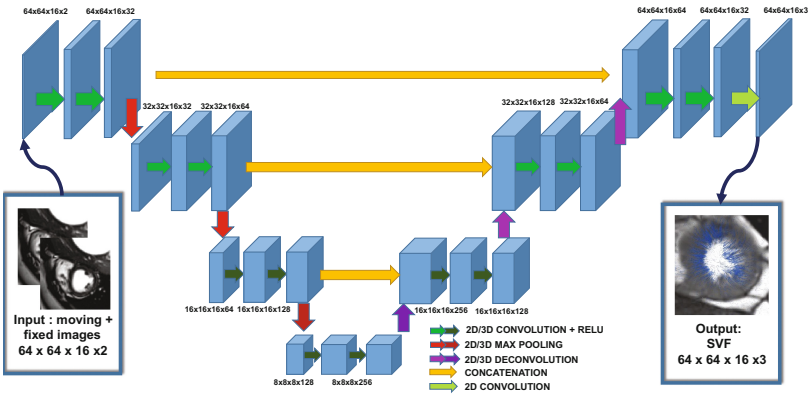

Fig. 2. Fully convolutional neural networks architecture for $3D$ registration.

Figure 2 illustrates our proposed network architecture. Similar to the standard U-Net architecture [9]. The input of the network is both images (moving and fixed) stacked together. In our application, we study cardiac MRI short-axis images, which are acquired with a non-isotropic resolution in the Z axis (slice spacing ranges from 6 to 10 mm whereas in-plane spacing ranges from 1.5 to 2.5 mm). To account from this discrepancy, the first two layers are 2D layers, then our features map is isotropic in all directions and we apply 3D layers for the 3rd and 4th layers. Finally, in the last layer, a simple 2D convolution builds the 3 layers corresponding to the SVF parametrization in the X, Y, Z axis at the same grid as the initial input images.

3 Validation on a Cardiac Image Database

We test our method on the problem of inter-patients registration on a large dataset of segmented 3D MRI cardiac images. Our dataset consists of $N = 187$ short-axis acquisitions of end-diastolic cardiac images acquired in multiple clinical centers (University College London Hospitals, Ospedale Pediatrico Bambino Gesú Roma and Deutsches Herzzentrum Berlin). For each of these images, the myocardium was segmented based on a data-driven approach combining the methods of [6,10] and quality controlled by experts. As a preprocessing step, and to study all the data in a common space, we first rigidly align the images (using the information from the ground truth segmentations) and resample them to have a consistent image size of $64 \times 64 \times 16$ through all the dataset.

Training. We divide our dataset into training ($80\% = 150$ images) and testing ($20\% = 37$ images) sets. On the training set we compute the reference SVFs for all the pair-wise registrations based on the framework we described. The computation of the surface matching of each of the 150 segmented surface to the template with the framework of currents took 6 hours on a single core CPU (a cluster of CPU was used). Once the surface are mapped to the template, the process to compute the SVF parametrization took 4 min for each of the pair of images. Because our method already gives us a large database of ground truth data, we only use small translations in the X and Y axis for data augmentation (this also improves the robustness of the learned network over slight rigid misalignment of both images). For the loss function, we used the sum of squared difference between the predicted SVF parametrization and the ground truth. We implement the network using Tensorflow[1] and we train it on a NVIDIA TitanX GPU with $100,000$ iterations using the ADAM solver which took approximately 20 h.

Evaluation. We compare the results with the registration algorithm *LCC Log-Demons* [8] for which we use a set of parameters optimized in a trial and error

[1] www.tensorflow.org.

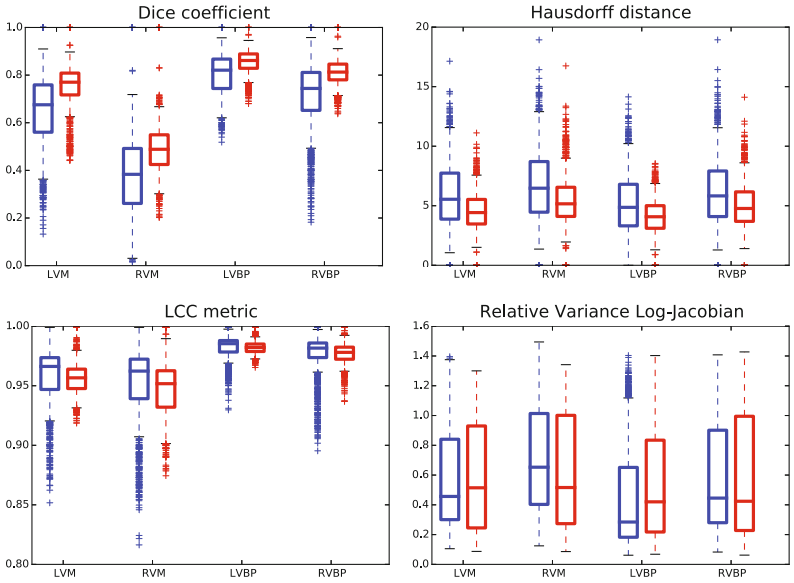

Fig. 3. Boxplot of the similarity score between the ROIs. (Red): proposed method, (Blue): Log-Demons with LCC metric. The ROIs we look at are: Left Ventricle Blood Pool (LVBP), Left Ventricle Myocardium (LVM), Right Ventricle Blood Pool (RVBP) and Right Ventricle Myocardium (RVM).

approach on a subset of the training set. To evaluate the accuracy of the registration, we compute the registration of all pair of images in the test set for a total of $37^2 = 1369$ registrations. On average, one registration with the *LCC Log-Demons* algorithm took approx. 4 min with a Intel Core i7-4600U CPU 2.10 GHz whereas our SVF-Net algorithm took 6 s on the same CPU and less than 30 ms on a NVidia TitanX GPU, increasing the speed by $\times 40/\times 8000$ with the CPU/GPU implementation.

We compare both methods on 4 ROIs (Fig. 3) using three different similarity metrics: Dice coefficient, the Hausdorff distance, the Local Correlation Coefficient (LCC) metric (which is the metric optimized with the Log-Demons algorithm) and one metric measuring the smoothness of the deformations. Our proposed approach performed better with respect to the Dice coefficient and the Hausdorff distance in approx 75% of the cases. In particular, the difference in accuracy is larger for the RV than for the LV. It is not surprising since the texture for the RV is usually less consistent between different patients, therefore a traditional registration method can have difficulty to match voxel intensities without shape a-priori. As for the LCC metric, which measures voxel intensity matching, although on average it is better for the LCC Log-Demons algorithm (which optimizes this metric), there is many outliers for which our method performs better as well for these cases, probably because the optimization algorithm gets stuck in a local minimum of the function. Finally, we compare the smoothness of

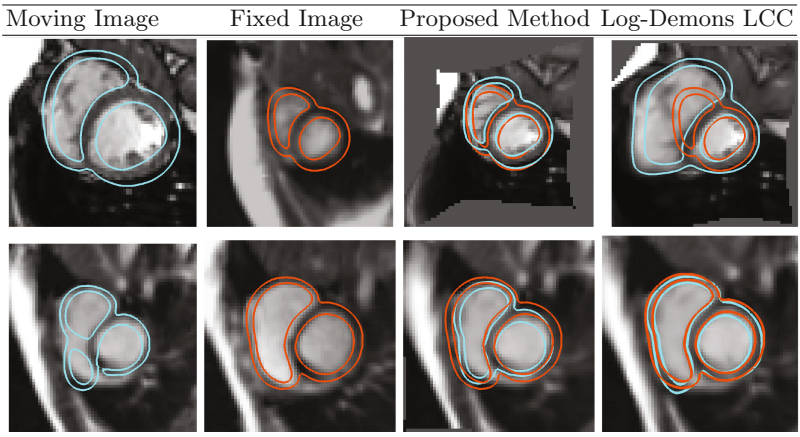

Fig. 4. Two examples of the results of our registration method versus optimization approach. (Column 1–2): the moving (resp. fixed) image with the segmentation. (3rd column): our proposed registration with the deformed myocardium segmentation of the moving image in cyan and the target segmentation in orange. (Right column): the optimization comparison (*Log-Demons LCC*).

the deformations. The difference of shapes seen in the images can be important, therefore Jacobians tend to be large even for regular deformations. We are more interested in evaluating how variable Jacobians are inside each region and we show the variance of the Log-Jacobians normalized by its value in Fig. 3 (bottom right). Statistically significant differences (p-value of Welch's test) can only be seen for the RVM (our proposed method has lower variance) and the LVBP (our method has higher variance). Overall, both methods output deformations with similar regularity.

Finally, Fig. 4 shows two different cases and the matching given by both methods. First row sees a typical case where the LCC-Log Demons fails because of the large difference of shapes seen in both images: it gets stuck in a local minimum and does not manage to retrieve the fixed image shape. On the second row, we show one of the case where the optimization performs the best with respect to our method. Although, one can see that, our matching is still acceptable.

4 Conclusions

In this article, we propose a novel methodology to build ground truth deformation from pair of segmented images and to train a predictive algorithm with them. Compared to other methods, our method has the benefit not to rely on either synthetic images or on the previous computation from an optimization method. We instead take advantage of the learning approach and chose to learn on different type of information not related to voxel intensity correspondence between images. A possible extension of this work would be to consider texture

information as well, for example by using this method as an initialization for a image registration algorithm optimizing a voxel intensity criteria.

We show that our method outperforms a state-of-the-art optimization method. We also highlighted multiple advantages of our method: (i) it is faster with a speed-up of $\times 40/\times 8000$ with the CPU/GPU implementation, (ii) it does not require the choice of any parameter at test time making it easy to use for a non-experienced user, (iii) it is more robust to outliers. These qualities are more important than just raw accuracy as they represent the main problems currently holding back registration methods to be used for large scale database.

One possible perspective is the multi-atlas mesh segmentation with a very large number of templates (here 187). In addition to the most probable segmentation, it also gives us a quantification of the uncertainty of the segmentation at each vertex. Enclosed is the result of one such segmentation computed using our database colored by the estimation of the variability.

Ackowledgements. The authors acknowledge the partial funding by the EU FP7-funded project MD-Paedigree (Grant Agreement 600932).

References

1. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A log-euclidean framework for statistics on diffeomorphisms. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 924–931. Springer, Heidelberg (2006). doi:10.1007/11866565_113
2. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Comput. Vis. **61**(2), 139–157 (2005)
3. Ceritoglu, C., Tang, X., Chow, M., Hadjiabadi, D., Shah, D., Brown, T., Burhanullah, M.H., Trinh, H., Hsu, J., Ament, K.A., et al.: Computational analysis of LDDMM for brain mapping. Front. Neurosci. **7**, 151 (2013)
4. Davis, M.H., Khotanzad, A., Flamig, D.P., Harms, S.E.: A physics-based coordinate transformation for 3-D image matching. IEEE Med. Imaging **16**, 317–328 (1997)
5. Durrleman, S., Prastawa, M., Charon, N., Korenberg, J.R., Joshi, S., Gerig, G., Trouvé, A.: Morphometry of anatomical shape complexes with dense deformations and sparse parameters. NeuroImage **101**, 35–49 (2014)
6. Jolly, M.-P., Guetter, C., Lu, X., Xue, H., Guehring, J.: Automatic segmentation of the myocardium in cine MR images using deformable registration. In: Camara, O., Konukoglu, E., Pop, M., Rhode, K., Sermesant, M., Young, A. (eds.) STACOM 2011. LNCS, vol. 7085, pp. 98–108. Springer, Heidelberg (2012). doi:10.1007/978-3-642-28326-0_10
7. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
8. Lorenzi, M., Ayache, N., Frisoni, G.B., Pennec, X.: LCC-Demons: a robust and accurate symmetric diffeomorphic registration algorithm. NeuroImage **81**, 470–483 (2013)

9. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). doi:10.1007/978-3-319-24574-4_28
10. Wang, Y., Georgescu, B., Chen, T., Wu, W., Wang, P., Lu, X., Ionasec, R., Zheng, Y., Comaniciu, D.: Learning-based detection and tracking in medical imaging: a probabilistic approach. In: Hidalgo, M.G., Torres, A.M., Gómez, J.V. (eds.) Deformation Models. Lecture Notes in Computational Vision and Biomechanics, pp. 209–235. Springer, Netherlands (2013)
11. Yang, X., Kwitt, R., Niethammer, M.: Fast predictive image registration. In: Carneiro, G., Mateus, D., Peter, L., Bradley, A., Tavares, J.M.R.S., Belagiannis, V., Papa, J.P., Nascimento, J.C., Loog, M., Lu, Z., Cardoso, J.S., Cornebise, J. (eds.) LABELS/DLMIA -2016. LNCS, vol. 10008, pp. 48–57. Springer, Cham (2016). doi:10.1007/978-3-319-46976-8_6