Deformable Medical Image Registration Using a Randomly-Initialized CNN as Regularization Prior

Max-Heinrich Laves
laves@imes.uni-hannover.de
Sontje Ihler
ihler@imes.uni-hannover.de
Tobias Ortmaier
ortmaier@imes.uni-hannover.de

Institute of Mechatronic Systems, Leibniz Universität Hannover, Germany

Editors: Under Review for MIDL 2019

Abstract

We present deformable unsupervised medical image registration using a randomly-initialized deep convolutional neural network (CNN) as regularization prior. Conventional registration methods predict a transformation by minimizing dissimilarities between an image pair. The minimization is usually regularized with manually engineered priors, which limits the potential of the registration. By learning transformation priors from a large dataset, CNNs have achieved great success in deformable registration. However, learnt methods are restricted to domain-specific data and the required amounts of medical data are difficult to obtain. Our approach uses the idea of deep image priors to combine convolutional networks with conventional registration methods based on manually engineered priors. The proposed method is applied to retinal optical coherence tomography scans showing different disease conditions. We show that our approach registers image pairs with state-of-the-art accuracy by providing dense, pixel-wise correspondence maps. It does not rely on prior training and is therefore not limited to a specific image domain.

Keywords: deformable registration, convolutional network, optical coherence tomography

1. Introduction

Deformable registration is a major challenge in medical image processing. The result is a dense mapping showing pixel-wise non-linear correspondences between a pair of images that best aligns the input image $I$ onto the target image $T$ by means of some similarity definition $\mathcal{L}$. Deformable registration is applied in the analysis of patient-specific temporal or anatomical changes, e.g. from pre-operative to post-operative state, or to show inter-patient variances (Sotiras et al., 2013). Deformable registration is also performed in atlas-based segmentation, where an input image is matched onto a target image with known segmentation (Cabezas et al., 2011).

Existing registration methods can be separated into two categories. The first category is based on non-learning methods which estimate a registration $w$ by optimizing a cost function of the form

$$\arg \min_w \left\{ \mathcal{L}(T, w \circ I) + \lambda R(w) \right\},$$

where $w \circ I$ denotes $I$ warped by $w$. A common assumption of $w$ is a displacement vector field $\mathbf{u}$ mapping every pixel coordinate to other pixel coordinates. The first term is referred to as data term, which is typically chosen to be a pixel intensity error measure. Optimization
Deformable Medical Image Registration Using a Randomly-Initialized CNN as Prior

Figure 1: Overview of our method. The randomly-initialized generator network $f_\theta$ acts as parameterization of the registration field $u$. The parameters $\theta$ are optimized for every image pair individually by gradient descent.

of the data term alone is considered ill-posed. The second term $R$, weighted by trade-off factor $\lambda$, is a regularizer that shapes the registration by any chosen prior, which helps solving the ill-posed problem. Common regularization is done by enforcing smoothness onto the displacement vector field by penalizing first or higher order spatial derivatives of $u$ (Werlberger et al., 2010). The result of the registration algorithm heavily depends on the cost function and therefore on the chosen prior of $R$.

The second category implicitly learns the regularization prior by training a convolutional network on a large database of domain-specific images. Early approaches rely on ground truth registrations (Sokooti et al., 2017), which are hard to obtain especially in medical imaging. More recent methods (Balakrishnan et al., 2019) propose unsupervised registration using the spatial transformer function (Jaderberg et al., 2015). However, these methods either only support small displacements or require segmentation maps of the image pairs during training to assist the convergence (Hu et al., 2018). Additionally, the trained networks are limited to register images from the training domain (e.g. CT or MRI).

Inspired by the idea of deep image priors (Lempitsky et al., 2018), we subsequently propose our learning-free method for deformable medical image registration using the structure of an untrained convolutional network as regularization prior.

2. Methods

Lempitsky et al. have recently shown that the excellent performance of CNNs for inverse image problems, such as denoising, is not only based on their ability to learn image priors from data, but is also based on the structure of a convolutional image generator itself (Lempitsky et al., 2018). They gave evidence that the structure of a network alone is sufficient to capture enough image statistics to provide state-of-the-art performance in inverse image tasks.

Leveraged by this idea, we reformulate the task of deformable image registration by using the structure of a convolutional network as regularizer (see Fig. 1). An image generator network $u = f_\theta(z)$ with randomly-initialized parameters $\theta$ is interpreted as parameterization of the dense correspondence mapping $u \in \mathbb{R}^{2 \times H \times W}$ between an input image $I \in \mathbb{R}^{C \times H \times W}$.
and a target image \( T \in \mathbb{R}^{C \times H \times W} \). The input \( z \in \mathbb{R}^{C' \times H \times W} \sim \mathcal{N}(0, 0.1) \) has the same spatial dimensions as \( u \) and is sampled from a random normal distribution in every iteration. This leads to the following optimization problem

$$
\arg\min_{\theta} \left\{ \mathcal{L}(T, f_{\theta}(z) \circ I) \right\},
$$

where \( f_{\theta}(z) \circ I \) denotes the differentiable spatial transformer function (Jaderberg et al., 2015). Eq. (2) is optimized for every image pair \( \{I, T\} \) using the Adam gradient descent optimizer (Kingma and Ba, 2014). As data term, we chose pixel-wise mean absolute error \( \mathcal{L}(T, f_{\theta} \circ I) = |(f_{\theta} \circ I) - T| \). The architecture of the image generator network \( f_{\theta} \) is chosen according to (Lempitsky et al., 2018). It has an encoder-decoder structure with skip connections between the encoding and decoding part.

3. Results & Conclusion

We demonstrate our approach on the task of retinal optical coherence tomography (OCT) registration. The data used in this work contain 100 pairs of randomly chosen OCT scans from (Kermany et al., 2018) showing pathological and healthy retinas. We use the aforementioned dissimilarity measurement \( \mathcal{L} \) and the structural similarity index (SSIM) (Wang et al., 2004) between \( f_{\theta} \circ I \) and \( T \) as evaluation metric. We compare our method to state-of-the-art methods from the Insight ToolKit (ITK) registration framework by combining an initial affine registration and a subsequent deformable displacement field registration (Avants et al., 2012). Results for exemplary image pairs and boxplots of results for all image pairs are shown in Fig. 2. Additional results including registration fields are shown in appendix A.

The results reveal that the structure of a convolutional network can act as regularization in deformable medical image registration with state-of-the-art performance. This connects traditional non-learning methods and learning-based methods by using randomly-initialized convolutional networks as prior.
Acknowledgments
This research has received funding from the European Union as being part of the EFRE OPhonLas project.

References


Appendix A. Results

<table>
<thead>
<tr>
<th></th>
<th>$\mathcal{L}$</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITK</td>
<td>$0.064 \pm 0.024$</td>
<td>$0.713 \pm 0.102$</td>
</tr>
<tr>
<td>ours</td>
<td><strong>$0.019 \pm 0.006$</strong></td>
<td><strong>$0.912 \pm 0.056$</strong></td>
</tr>
</tbody>
</table>

Table 1: Mean results compared to a state-of-the-art method from ITK. Bold values denote best results.
Figure 3: Additional results.