Supervised learning in medical image registration

Sokooti, H.

1

Introduction

1.1 Medical image registration

Image registration is the process of aligning images by finding the spatial relation between the images. Assuming two images called fixed and moving images are taken at different time, different spatial location, or via a different imaging technique, the aim of image registration is to find an optimal transformation that aligns the fixed and the moving images.

Image registration has many applications in medical image analysis [1]. By aligning images from different modalities, the information can be fused together and provide complementary insight to a medical expert. For instance, in head and neck radiation therapy, Magnetic Resonance Imaging (MRI) provides higher contrast in soft-tissue. However, Computerized Tomography (CT) images commonly have better spatial resolution and provide electron density information. By aligning the MR and CT images, it is possible to exploit the advantage of each modality [2]. In fundus photography, image registration is also utilized to compose several images taken from different angles generating a single image with larger field of view (FOV). Fig. 1.1 illustrates an example of registration in fluorescein angiography (FA) retinal images [3]. Another application of the registration is in Alzheimer's disease classification with Brain MR images. The Jacobian of the transformation indicates the local volume change within the brain, which is an informative feature to detect Alzheimer's disease [4].

Given a fixed $\mathrm{I_F}(\boldsymbol{x})$ and a moving image $\mathrm{I_M}(\boldsymbol{x})$, the aim of a pair-wise registration is to find a displacement $u(\boldsymbol{x})$ that makes $\mathrm{I_M}\big(\boldsymbol{x} + u(\boldsymbol{x})\big)$ spatially aligned to $\mathrm{I_F}(\boldsymbol{x})$. The *transformation* is usually referred to as $\boldsymbol{T}(x) = \boldsymbol{x} + u(\boldsymbol{x})$. Here, we defined the direction of this mapping from the fixed image to the moving image as illustrated in Figure 1.1. In *parametric* registration, the transformation is defined by a model, like thin-plate

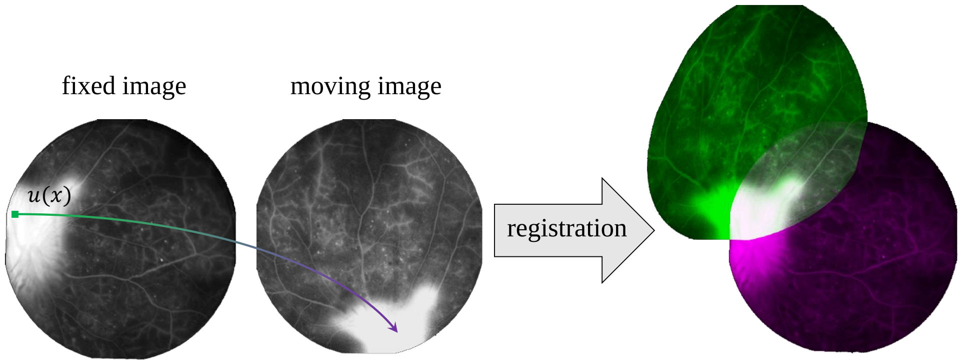


Figure 1.1: An example of two-dimensional image registration applied to fluorescein angiography retinal images. By aligning images taken from different angles, a single image with larger field of view (FOV) is composed. [3].

splines [5] or B-splines [6], while in *non-parametric* registration, the degree of freedom of the transformation is equal to the total number of voxels in the image. After finding the transformation, the moving image will be resampled in the fixed image domain by an interpolation technique.

Conventionally, image registration is cast to an optimization problem. A loss function is defined based on a dissimilarity measure. For instance, a simple loss can be defined as a mean squared difference (MSD) of the intensity values of the overlapping region between the fixed and the moving images. This optimization can be solved with an iterative approach like stochastic gradient descent [7]. Finally, the optimal transformation will be found with respect to the loss function.

Fine-tuning an image registration algorithm is a time-consuming task. Both the dissimilarity metric and the transformation model need to be selected and tuned in order to achieve high quality registration performance. Another drawback of conventional image registrations is that their inference time is rather slow. Since most of them use an iterative optimization method, it is not trivial to run the optimization in parallel. Fast image registration is required in several medical tasks such as registering the follow-up scans in adaptive radiotherapy [8] and image-guided surgery [9].

1.2 Learning-based image registration

Learning-based registration techniques are becoming more popular [10]. One of the applications is to learn a dissimilarity metric instead of tuning over handcrafted dissimilarity metrics. The advantage of learning a dissimilarity metric is more prominent in multi-modal image registration like ultrasound (US)/MR, [11, 12]. It is reported that the learned metric outperforms well-known multimodal metrics such as mutual information (MI) [13] and the modality independent neighbourhood

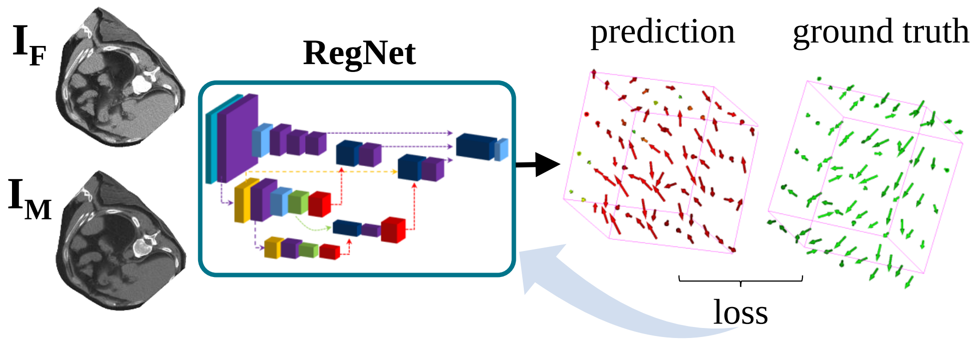
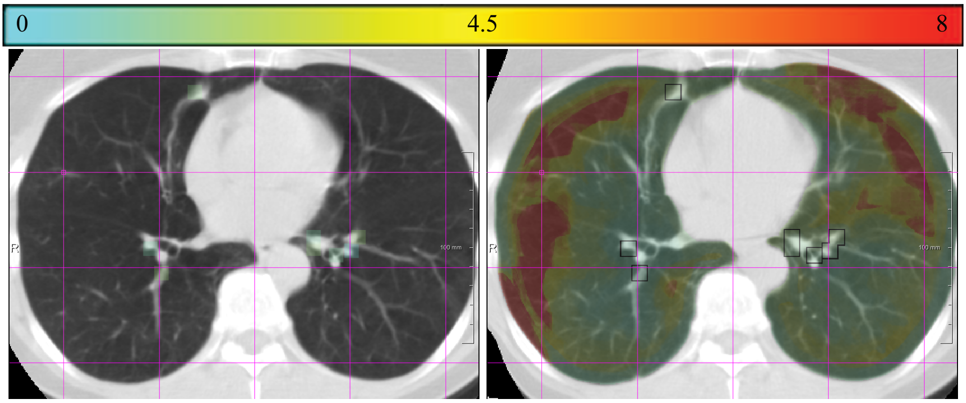


Figure 1.2: A schematic view of convolutional neural network based registration with the supervised transformation approach. The loss function is computed based on the dissimilarity between a ground truth transformation and the predicted one from the network.

descriptor (MIND) [14].

Several methods have been proposed to learn the entire image registration pipeline using deep learning (DL). Most methods were introduced after starting this PhD theses. In reinforcement learning (RL) approaches [15], instead of a conventional optimization, a trained agent is used to perform the registration [16]. However, the RL approach can still be time-consuming. In principle, both optimization and dissimilarity metrics can be learned simultaneously. Thus, at the inference time, the registration is usually performed in one shot. In the supervised transformation approach, a known transformation is used during the training [17, 18, 19, 20, 21]. The loss function then can be defined based on the difference between the the known transformation and the predicted transformation. Fig. 1.2 illustrates a schematic design of this approach. To train the network, three inputs are needed, the fixed image, the moving image and the known ground truth transformation between the fixed and the deformed moving image. In the unsupervised transformation approach, an indirect loss is utilized to guide the transformation. Several examples of this indirect loss are the mutual information, cross correlation ([22, 23, 24]), the Dice overlap of known segmentation maps [25], normalized gradient field distance measure [26], or even using generative adversarial networks (GAN [27]) to learn a new indirect loss [8, 28]. Recent papers showed that utilizing a proper regularization such as bending energy [18], volume change penalty [26] or a graph regularization network on a keypoint-based registration [29] can improve the performance of registration.



(a) Fixed Image (b) Deformed moving image

Figure 1.3: An example of registration error map, which is overlaid on the deformed moving image in a chest CT scan pair [33]. The color bar indicates the estimated registration error in mm.

1.3 Uncertainty and error in image registration

In most registration methods, no assessment of the registration quality is provided, and simply the result is returned. Many medical pipelines are based on registered images and it is important to know the uncertainty of registration before continuing to the next phase in order to prevent the accumulation of errors. For example, in online adaptive radiotherapy daily contouring of the tumor and organs-at-risk can be performed with the help of image registration and therefore quality assessment (QA) is mandatory to ensure patient safety [30]. Visualizing the registration error can also be directly helpful in various medical applications. For instance, an error map of registering a pre-operative scan and an atlas could provide more insight about the localization error during a surgical procedure [31]. Refinement of registration is another important application of error prediction [32]. An example of a registration error map is given if Fig. 1.3. The color bar indicates the estimated registration error in mm. For instance, the registration probably should be improved in the red regions. This error map provide insight about the registration quality. Hence, a medical expert can consider the local uncertainty of the subsequent analysis on the aligned images. Currently, the quality assessment of registration is usually performed manually, which is a time-consuming task and prone to human fatigue.

Defining the registration error is not trivial as image registration is an ill-posed problem. The registration error can be better explained on the corresponding distinctive landmark locations. However, computing the registration error over homogeneous regions is more challenging. Registration uncertainty can be counted as a measure of confidence in the registration output. Probabilistic image registration (PIR) methods usually can provide transformation and uncertainty at once [34]. In registration methods using continuous optimization, one way to estimate the uncertainty is to perturb the initial state [35]. The uncertainty sometimes is used as a surrogate for registration error. It should be noted that high uncertainty does not always means high registration error and vice versa [36].

Several naive intensity-based and registration-based features were proposed as a surrogate for registration misalignment, such as local normalized mutual information (NMI) [37] and the local gradient of the loss function [38]. Simply, the smaller value of NMI or the gradient indicates smaller registration misalignment. More advanced learning techniques are also utilized in predicting the registration error such as learning over landmark locations [39] and learning over artificially generated image pairs [40]. However, the accuracy of naive methods are not promising and the inference of the advanced methods is time-consuming.

1.4 Outline of the thesis

The aim of this thesis is to develop a learning-based image registration method as a much faster alternative to conventional methods without requiring hyper-parameter tuning. We also aimed to improve accuracy as well as inference time of registration misalignment detection methods, via a fully automatic solution. Although all the proposed methods in this thesis are generic, all the experiments are performed on chest CT scans.

Chapter 2 presents a novel method to solve nonrigid image registration through a learning approach, instead of via iterative optimization of a predefined dissimilarity metric. We design a Convolutional Neural Network (CNN) architecture that, in contrast to all other work, directly estimates the displacement vector field (DVF) from a pair of input images. This is one of the first methods proposed in literature to solve nonrigid image registration via deep learning.

Chapter 3 extends chapter 2 into a practical pipeline based on efficient supervised learning from artificial deformations. The proposed architectures are embedded in a multi-stage approach to increase the capture range of the networks in order to more accurately predict larger displacements.The proposed method achieved the best result on the DIR-Lab 4DCT study among all published DL-based registration methods up to the publication date.

Chapter 4 proposes a new automatic method to predict the registration error in a quantitative manner and is applied to chest CT scans. A random regression forest is utilized to predict the registration error locally. The forest is built with features related to the transformation model and features related to the dissimilarity after registration. Several of the proposed features are novel and unique as well.

Chapter 5 presents a supervised method to predict registration misalignment using convolutional neural networks (CNNs). This task is casted to a classification problem with multiple classes of misalignment: “correct” 0-3 mm, “poor” 3-6 mm and “wrong” over 6 mm. Rather than a direct prediction, we propose a hierarchical approach, where the prediction is gradually refined from coarse to fine. Our solution is based on a convolutional Long Short-Term Memory (LSTM), using hierarchical misalignment predictions on three resolutions of the image pair, leveraging the intrinsic strengths of an LSTM for this problem.

Chapter 6 summarizes and discusses the overall achievements of this thesis.