Supervised learning in medical image registration

Sokooti, H.

6

Summary and Future Work

Image registration is a crucial task in medical image processing. Performing an automatic fast image registration with less manual finetuning can speed up numerous medical image processing procedures. In addition, an automatic quality assessment of registration can speed up this time-consuming task. In this thesis, we developed a learning-based image registration technique called RegNet. Moreover, we proposed two quality assessment mechanisms using random forests (RF) and convolutional long short term memory (ConvLSTM), in which the latter performs faster and more accurate. In this chapter, we summarize the previous chapters and discuss potential directions of future research.

6.1 Summary

In the first chapter, we provided general information about image registration and quality assessment of registration. In Chapter 2, we propose a convolutional neural network architecture to solve nonrigid image registration through a learning approach. The proposed RegNet is trained using a set of random artificially generated DVFs with a maximum deformation of 8 mm in each direction. In Chapter 3, we substantially improve the proposed RegNet by utilizing a multi-stage approach and improving the artificial data generation procedure. A quantitative error prediction of medical image registration is proposed in Chapter 4 using regression forests. The forest is built with features related to the transformation model and features related to the dissimilarity after registration on distinctive landmark locations. In Chapter 5, a hierarchical prediction of registration misalignment using a convolutional LSTM with application to chest CT scans is proposed. The proposed method is substantially faster than methods involving multiple registrations.

Chapter 2 In this chapter, we propose a 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 chapter is one of the first proposed methods in nonrigid DL-based image registrations. The proposed RegNet is trained using a set of random artificially generated DVFs with a maximum deformation of 8 mm in each direction.The proposed method does not explicitly define a dissimilarity metric, and integrates image content at multiple scales to equip the network with contextual information. At testing time nonrigid registration is performed in a single shot, in contrast to current iterative methods. We tested RegNet on an in-house chest CT study called SPREAD. It achieved an average target registration error (TRE) of 1.66 mm over the test cases. The results show that the accuracy of RegNet is on par with a conventional B-spline registration, for anatomy within the capture range, i.e. less than 8 mm.

In **Chapter 3**, we substantially enhance the initial RegNet method introduced in Chapter 2. The newly proposed method utilizes a multi-stage approach, which significantly enlarges the capture range. The artificial data generation was improved by including more generic deformations as well as more realistic deformations like respiratory motion. We experimented with various network architectures and the proposed "U-Net-advanced" design achieved better performance in the validation set. This design was similar to a U-Net, but with addition of dilated convolutional layers. The proposed method, RegNet, is evaluated on multiple databases of chest CT scans and achieved a target registration error of 2.32 ± 5.33 mm and 1.86 ± 2.12 mm on SPREAD and DIR-Lab-4DCT studies, respectively. Consequently, the enhanced RegNet achieved the best result on the DIR-Lab 4DCT study among all published DL-based registration methods. The average inference time of RegNet with two stages is about 2.2 s.

Chapter 4 presents a quantitative error prediction of medical image registration using regression forests. A new automatic method is proposed 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. The feature set consists of the variation of displacement vector field, the coefficient of variation of joint histograms, determinant of the Jacobian, the modality independent neighborhood descriptor (MIND), and the local normalized mutual information. The forest is trained and tested using manually annotated corresponding points between pairs of chest CT scans in two experiments: SPREAD (trained and tested on SPREAD) and inter-database (including three databases SPREAD, DIR-Lab-4DCT and DIR-Lab-COPDgene). The results show that the mean absolute errors of regression are 1.07 ± 1.86 and 1.76 ± 2.59 mm for the SPREAD and inter-database experiment, respectively. The overall accuracy of classification in three classes (correct, poor, and wrong registration) is 90.7% and 75.4%, for SPREAD and inter-database respectively. The good performance of the proposed method enables important applications such as automatic quality control in large-scale image analysis.

In **Chapter 5**, a hierarchical prediction of registration misalignment using a convolutional LSTM with application to chest CT scans is proposed. The proposed method is substantially faster than methods involving multiple registrations. 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. The convolutional LSTM is trained on a set of artificially generated image pairs obtained from artificial displacement vector fields (DVFs). Results on chest CT scans show that incorporating multi-resolution information, and the hierarchical use via an LSTM for this, leads to overall better F1 scores, with fewer misclassifications in a well-tuned registration setup. The final system yields an accuracy of 87.1%, and an average F1 score of 66.4% aggregated in two independent chest CT scan studies.

6.2 Discussion and Future Work

The work presented in this thesis was aimed at developing methods to perform image registration as well as quality assessment of image registrations.

In the proposed RegNet in Chapters 2 and 3, we utilized a deep convolutional neural network approach. Although deep learning methods in segmentation applications achieved promising results, several challenges still exist in the registration applications. Finding the optimal solution in conventional segmentation methods like level set and min cost (minimum of the cost function) are usually iterative similar to the conventional image registration techniques. However, in conventional image segmentations, the main image is constant in all iterations and the segmentation is updated in each iteration. On the contrary, in the conventional iterative image registration, an implicit (or explicit) resampling of the deformed moving image is performed at each iteration. Apparently, predicting the final transformation in one shot is still challenging in DL-based methods. We noticed significant improvement when using the multi-stage approach, in which a resampling was also performed. As reported in Table 2.2 the average TRE was improved from 3.80 mm to 1.57 mm. It is worth noting that the registration quality may still be improved by sequentially employing multiple RegNets in the original resolution. Potentially, a simple stopping criterion like the difference of variation between subsequent transformations or a more complex approach such as reinforcement learning can be used.

We utilized a supervised transformation approach in Chapter 2 and 3 of this thesis. Quite a few articles proposed an unsupervised transformation approach [23, 24, 8]. One of the advantages of the unsupervised transformation method is that the training data can be fully realistic including fully realistic ground truth transformations. Usually the training is performed with a simple dissimilarity metric like mutual information. Thus, a potential disadvantage of unsupervised methods is that the ground truth transformation is not known, and the trained network may not necessarily be better than a conventional iterative registration with the same dissimilarity metric. On the other hand, in the supervised transformation approach, the ground truth transformation is accurate (not necessarily unique as registration is usually an ill-posed problem). On the contrary, the transformation and usually one of the images (the deformed moving image) may not be completely realistic. Although artificial data generation could be a potential way to get higher performance than human experts, this could be too idealistic at this moment, where current deep learning based registrations could be much further enhanced. All in all, in order to improve the supervised approaches, the necessity for a large medical dataset providing ground truth for transformations can be strongly felt. This can be done by annotating distinctive landmarks as well as region segmentations.

In Chapter 4, we proposed a regression approach to predict the registration error and in Chapter 5, we simplified the task into a classification approach. Each of these approaches have their owns pros and cons. It should be highlighted that in a regression approach, the acceptable margin of the regression error correlates with the ground truth value. Thus, a normalized loss could help the training to converge better. In the classification approach, this issue is eliminated when considering a class with large values like $[6, \infty)$ mm. However, it should be pointed out that when defining the labels as correct, $[0, 3)$ mm, poor $[3, 6)$ mm, and wrong $[6, \infty)$ mm, the labels are not ordinal anymore but more similar to ordered labels. This means that a mis-classification between the correct and wrong label is worse than a mis-classification between two adjacent labels such as correct and poor. This criterion does not naturally exist in classification approaches but can be imposed using hierarchical classification. In general, the classification is more difficult for values close to the border. For instance, it is not trivial to classify a value of 2.99 in either the correct or the poor class. A solution might be to utilize soft ground truth labels, for example for the value 2.99, the ground truth can be set to [0.60, 0.40, 0] for classes correct, poor, and wrong respectively. In the hyperspherical prototype approach [131] the one hot encoded ground truth will be mapped to an output space. This way we can provide *a priori* information about the labels and utilize that in the organization of the output space. For instance, the wrong and the poor labels can be close to each other, while the wrong and the correct labels can have more distance. Another interesting application in the hyperspherical prototype is to simultaneously perform the regression and classification even in the same output space.

In Chapters 2, 3, and 5, we utilize artificial data generations in order to train a convolutional neural network to learn registration or registration error. We proposed "single frequency", "mixed frequency", and "respiratory motion" approaches to artificially generate displacement vector fields. One of the limitations of the aforementioned generations is that they are expected to be sensitive to anatomical changes like tumor growth. This limitation may potentially be addressed by adding a new type of deformation to the artificial training data strategy, which mimics such anatomical changes. In general, the artificial generation can be further enhanced by adding more realistic and complex simulations. For instance, if a rib segmentation is available in chest CT scans, it is possible to perform nonrigid deformation outside of the rib and rigid deformations inside the rib. The network potentially can learn the relation between organs and the rigidity of the deformations. Other realistic deformations like sliding motion of the lungs can also be added to the training images.

Although all experiments in this thesis are performed in chest CT scans, all proposed methods are generic and potentially can be applied to other modalities and anatomical sites as well. In a similar study on intrasubject magnetic resonance brain images registration, RegNet was trained on brain MR images and showed promising results [21]. However, utilizing artificial data generation in multi-modality images need to be investigated in the future, as potentially an intensity mapping approach [132] might be needed.

6.3 General conclusions

In conclusion, this thesis proposes learning-based methods for medical image registration and for quality assessment of image registration. All proposed methods are fully automatic and do not require human interactions. The proposed RegNet architecture was tested on registration chest CT scan pairs and achieved on par results with a conventional B-spline registration method. The hierarchical classification framework to detect registration misalignment using long short term memory convolutional neural networks (ConvLSTM) obtained promising results. All deep learning methods described in this thesis have a runtime in the order of seconds, substantially improving over conventional methods.