HHS Public Access
Author manuscript
Comput Med Imaging Graph. Author manuscript; available in PMC 2021 September 01.

Published in final edited form as:
Comput Med Imaging Graph. 2020 September ; 84: 101769. doi:10.1016/j.compmedimag.2020.101769.

Deep Adaptive Registration of Multi-modal Prostate Images

Hengtao Guo[a], **Melanie Kruger**[b], **Sheng Xu**[c], **Bradford J. Wood**[c], **Pingkun Yan**[a,*]

[a]Department of Biomedical Engineering and the Center for Biotechnology and Interdisciplinary Studies at Rensselaer Polytechnic Institute, Troy, NY 12180, USA.

[b]Shenendehowa High School, Clifton Park, NY 12065, USA

[c]National Institutes of Health, Center for Interventional Oncology, Radiology & Imaging Sciences, Bethesda, MD 20892, USA

Abstract

Artificial intelligence, especially the deep learning paradigm, has posed a considerably impact on cancer imaging and interpretation. For instance, fusing transrectal ultrasound (TRUS) and magnetic resonance (MR) images to guide prostate cancer biopsy can significantly improve the diagnosis. However, multi-modal image registration is still challenging, even with the latest deep learning technology, as it requires large amounts of labeled transformations for network training. This paper aims to address this problem from two angles: (i) a new method of generating large amount of transformations following a targeted distribution to improve the network training and (ii) a coarse-to-fine multi-stage method to gradually map the distribution from source to target. We evaluate both innovations based on a multi-modal prostate image registration task, where a T2-weighted MR volume and a reconstructed 3D ultrasound volume are to be aligned. Our results demonstrate that the use of data generation can significantly reduce the registration error by up to 62%. Moreover, the multi-stage coarse-to-fine registration technique results in a mean surface registration error (SRE) of 3.66mm (with the initial mean SRE of 9.42mm), which is found to be significantly better than the one-step registration with a mean SRE of 4.08mm.

Keywords

Convolutional Neural Networks; Prostate Cancer; Image Registration; Multi-modal Image Fusion

---

[*]Corresponding author yanp2@rpi.edu (Pingkun Yan).

Hengtao Guo: Methodology, Data curation, Software, Visualization, Writing- Original draft preparation.
Melanie Kruger: Formal analysis, Writing- Reviewing and Editing.
Sheng Xu: Investigation, Data curation, Resources, Writing- Reviewing and Editing.
Bradford J. Wood: Investigation, Resources, Writing- Reviewing and Editing.
Pingkun Yan: Conceptualization, Software, Writing- Reviewing and Editing, Supervision, Project administration.

Conflict of Interest
All authors have no conflict of interest to report.

Declaration of interests
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

1. Introduction

For prostate cancer, a main cause of death for men in the United States [1], the fusion of transrectal ultrasound (TRUS) and magnetic resonance (MR) images for guiding the biopsy aids the clinical diagnosis [2, 3]. However, the registration of 3D image volumes with different modalities has shown to be a challenging task. Conventional approaches have tried to optimize the registration field according to an intensity-based similarity metric. The most commonly used technique, mutual information, studies the pixel intensity correlation between different modalities for multi-modal image registration [4]. However, the image mapping is known to create poor results because of insufficient descriptions of the structural image features. Sun *et al.*[5] proposed the use of the modality independent neighborhood descriptor (MIND) [6]. Although it showed an improved performance, its descriptors require manually crafted features and can fail when the initialization has a significant distance from the target transformation. Similar shortcomings appear in other multi-modality image registration methods based on the intensity metric [7, 8, 9].

Over the past few years, Deep Learning (DL) has rapidly become an alternative image processing tool, including medical image registration [10]. Particularly, deep convolutional neural networks (DNNs), a key component of DL, has become an important tool for automatic feature extraction [11]. Inspired by the established iterative registration methods, Haskins *et al.* [12] proposed to train a deep similarity network that serves as the metric for iterative registration optimization, which remains to be computationally complex. In addition, de Vos *et al.* [13] proposed an unsupervised image registration method by maximizing the normalized cross correlation between source image and target image. Several other methods adopted similar ideas by optimizing other similarity metrics to align medical images. For example, Eppenhof *et al.* [14] proposed to register 3D thoracic CT images by minimizing the squared error between the network output and artificially generated displacement field. Balakrishnan *et al.* [15] used mean squared voxelwise difference and local cross-correlation as the image similarity term to train a network for brain 3D MRI registration. A recent work by de Vos *et al.* [16] used mutual information for unsupervised MR image registration. However, these methods cannot be directly applied to MR-TRUS image registration, as none of the traditional simple image similarity metric can reliably evaluate the alignment of MRI and TRUS due to the significant image appearance difference between the two modalities.

It has been found that typical auxiliary information, such as the atlas-based segmentation, can be used to improve the registration performance [17]. For example, Krebs *et al.* [18] generated a large amount of artificial deformation fields and used prostate segmentation to facilitate non-rigid registration. However, this method requires accurate image segmentation, which may cause the registration to fail when the quality of segmentation is unreliable. Hu *et al.* [19] proposed a weakly-supervised CNN for prostate US and MR registration. They also reported an adversarial network used for deformation field regularization in [20]. However, both works rely on extensive manually labeled corresponding landmarks for generating weak labels during the training phase. The POINT2 method, proposed by Liao *et al.* [21], extracts features from digitally reconstructed radiography. These features are used as convolutional kernels to find matching points in X-ray images for multimodal registration.

This method, however, can only be applied to multi-view scenarios and requires a large dataset for reliable feature learning. Other methods, including the work by Sokooti *et al.* [22], use DNNs with multi-scale input to predict deformation vector fields between two cropped patches from the moving and fixed images.

Following the preceding discussion, DL for image registration focuses, predominantly, on one-step registrations, in which the network predicts a final target transformation from a pair of input images and an initial transformation. However, upon completing the training of such a one-step network, further improvements of the registration accuracy are difficult. One solution is to adjust the hyper-parameters, redesign the network structure and then retrain the model, which is massively time-consuming and requires extensive human efforts. More importantly, there is no guarantee that retraining networks with different structures yields significant improvements.

Another possible solution is to repeatedly use the same trained one-step-registration network multiple times, in which the output of the previous step acts as the input to the next. Regardless, it should be noted that the distribution (the range of the registration error) of the input and output transformations are very different for each iteration. As and example, Fig. 1 shows histograms of surface registration errors (SREs) for an iterative image registration based on the same network. The values listed above each figure indicate the mean surface registration error (SRE) ± the standard deviation. One should note that the distributions in Fig. 1(a) and (b) are very different from each other. This explains the increased registration error when using the same network repeatedly, as its source domain has drastically changed after the first iteration. We therefore argue that networks for fine-adjusting results need to be trained independently from their source domains. For example, the 1-stage network trained on the initialization can map the testing SRE from A to B, while recursively applying this network will lead to an increase in the registration error. On the other hand, a 2-stage network which is separately trained on the second phase distribution (generated from the statistics of B) can further reduce the registration error from B to C. More details will be revisited in Section 3.3. A similar idea was proposed by de Vos *et al.* [23] to train convolutional networks for hierarchical multi-resolution and multi-level image registration. However, this method is limited to single modality image registration such as CT-CT or MRI-MRI.

This paper introduces a coarse-to-fine multi-stage registration (MSReg) framework, which consists of N consecutive networks. A summary is as follows.

1. We propose an error scaling method to sample training data that can have any target distribution based on the registration error. Specifically, this data generation method creates different training data distributions, making the networks perform well even on different error levels.

2. We introduce a coarse-to-fine multi-stage registration framework with a network more adaptive to its source distribution. Since each sub-network is trained using data following distributions from a given stage, there are improvements in the registration accuracy after each iteration. Additionally, this proposed multi-stage

framework is highly flexible and can be easily extended to other image registration tasks.

3. Based on a well-sized MR-TRUS registration dataset, we show the workings of both innovations and confirm that multi-stage registration achieves significant improvements comparing to the conventional methods on this challenging task.

2. Methods

This section presents details of multi-stage registration, starting with the workflow of DNN regression to update the transformation matrix. We then introduce the surface registration error (SRE), followed by the data, generation method built on top of the SRE scaling. Finally, we discuss the training strategy of the coarse-to-fine multi-stage registration.

2.1. Registration Workflow

This work is developed on the basis of rigid transformation to estimate the 6 degrees of freedom $\theta = \{\Delta t_x, \Delta t_y, \Delta t_z, \Delta \alpha_x, \Delta \alpha_y, \Delta \alpha_z\}$. Here, Δt and $\Delta \alpha$ refer to the translation in millimeters and the rotation in degrees along three directions, respectively, relative to the center of a moving image. A DNN estimates the amount of adjusting necessary of the moving image to improve its alignment with the fixed image. The rationale of using rigid registration is because it is widely used in the medical image registration field and can achieve clinically significant results. We propose extending the concept introduced here to incorporate deformable image registration to future work.

DNNs have shown their potential for various image processing tasks [11]. The core component of the 3D ResNeXt network [24] is used in this work as a building block that repeats itself multiple times, which Fig. 2 shows. The topology of the multiple pathways enables the extraction of important features, which, in turn, allows for the training of larger numbers of network layers (yielding deeper networks). The pre-operative MR image serves as the fixed image and the TRUS image as the moving image, since the axial view of MR is larger than TRUS so we can better observe the registration improvements over the stages (see Sect. 3.4). It should be noted that such fixed-moving relationship can be reversed during real-time MR-TRUS registration since the transformation rationale is the same.

To train the entire network, we use a 4 × 4 rigid transformation matrix (composed of a 3 × 3 rotation matrix and 3 translation parameters) as an initial transformation to roughly overlap a pair of MR and TRUS volumes to initialize the algorithm. During the image preprocessing, every MR and TRUS image pair is resampled to the same resolution according to the spacing information. After applying an initial transformation, the fixed image and moving image overlap with each other. We then determine a bounding box of the prostate in MRI based on the previously performed prostate segmentation. We use the same ROI bound to crop the TRUS image as well. Finally, the cropped ROIs are both resampled to a size of 32 × 96 × 96 voxels. The intensities of these two ROI volumes are scaled between [0, 1] and their concatenation will serve as the input to our multi-stage registration network.

The entire network is constructed from 3D convolutional layers to extract hybrid structural features between MR and TRUS image pairs. Residual blocks, with cardinality equalling to

16, compose the main part of the network. For network training, we use shortcut connections to avoid gradient vanishing. Also, the extracted features (from the convolutional layers) are applied to the dense layers to regress the parameters Δt_x, Δt_y, Δt_z, $\Delta \alpha_x$, $\Delta \alpha_y$ and $\Delta \alpha_z$. After the training, the network estimates the 6 degrees of freedom to adjust the position of the moving TRUS images.

During the network training process, the initial transformation matrices, which are used to initialize the entire workflow, are generated by adding 6 random perturbations to the ground-truth transformation matrices. These randomly generated perturbations are stored as labels to compute the mean squared error (MSE) loss L_{MSE} to update the network's parameters as:

$$L_{MSE} = \frac{1}{6} \sum_{d=1}^{6} (\theta_d^{gt} - \theta_d)^2 \tag{1}$$

where θ_d^{gt} and θ_d are the dth degrees of freedom of the ground-truth label and network-predicted result, respectively.

2.2. Training Data Generation

SRE is a commonly used criteria for evaluating the registration accuracy. For network training, we save the optimal network based on the mean SRE of the validation set, at the end of each training epoch. To calculate this, we use M points located on the surface of a prostate segmentation. By transforming the M surface points with the ground-truth and estimated transformation matrices, we can compute the mean Euclidean distance between each corresponding point as:

$$SRE(T) = \frac{1}{M} \sum_{i=1}^{M} \| T_{gt}(P_i) - T(P_i) \|_2 \tag{2}$$

where $T_{gt}(P_i)$ and $T(P_i)$ are the 3D coordinates of the ith surface point P_i transformed by the ground-truth matrix T_{gt} and testing matrix T, respectively. The moving and fixed images are sampled at 1 mm spacing, so SRE is measured in millimeters. Since SRE can be assumed to be linearly influenced by the translation parameters as well as by the random rotation perturbations, which only vary within a very small range, we can scale the SRE to a target value by scaling the 6 degrees of freedom with the same ratio.

For one-step registrations, we can calculate the SRE before and after the registration to evaluate the model performance for each case. Registration results with lower SRE values indicate that the final transformation matrix is closer to the ground-truth transformation matrix.

Although a fixed initialization matrix for each case yields rapid convergence during network training, it can also result in over-fitting. To address this, we generate a random transformation matrix for each case in the training set at each epoch. In other words, we introduce small random perturbations for the 6 degrees of freedom θ, under certain constraints. These perturbations can be added to the ground-truth transformation matrix T_{gt}

Author Manuscript
Author Manuscript
Author Manuscript

by $-\theta$ to generate a random initialization matrix T_0. Thus, the corresponding label of T_0 is the generated random perturbations θ, meaning that: given an initial transformation matrix T_0, by taking additional adjustments by the amount of θ to the moving image along 6 degrees of freedom, the moving image can be well aligned to the fixed image.

To generate a training set whose SRE values are drawn from the target distribution, we introduce the approach described in Algorithm 1. Take the 1- stage registration network training as an example: (1) First, we sample random perturbations θ' in a small range [−5, 5] at each degree of freedom; (2) we add the perturbations $-\theta'$ to the groundtruth matrix T_{gt} to get a random matrix T' and calculate its SRE; (3) we then draw a random value from a target distribution (e.g. uniform distribution) and assign it as the target SRE; (4) we calculate the scaling ratio r between the target SRE and T''s SRE; (5) we scale the random perturbation θ' to θ using the ratio r and then add the perturbations $-\theta$ to T_{gt} to generate the initial matrix T_0 for training the network. Since we use the ratio r, which is calculated between target SRE and T'' SRE, to scale the 6 degrees of freedom θ', the generated T_0's SRE is approximately equal to the target SRE. By doing this, we can generate large amounts of training data that produce SRE values drawn from any target distribution. Such a generated dataset, with theoretically infinite numbers of training samples, can significantly improve the network's robustness and reduce the registration error (see Sect. 3.2).

Algorithm 1 Training data generation by error scaling

Input: Ground-truth transformation matrix T_{gt} of a training case
Output: Initial transformation matrix T_0 and label θ used for training
procedure GENERATE_TRAINING_DATA(T_{gt})
  $\{\Delta t_x, \Delta t_y, \Delta t_z, \Delta\alpha_x, \Delta\alpha_y, \Delta\alpha_z\} \leftarrow Random\_Value()$
  $\triangleright$ Sample random perturbations
  $\theta' \leftarrow \{\Delta t_x, \Delta t_y, \Delta t_z, \Delta\alpha_x, \Delta\alpha_y, \Delta\alpha_z\}$
  $T' \leftarrow Update\_Mat(T_{gt}, -\theta')$ $\triangleright$ Add $-\theta'$ to T_{gt}
  $Current\_SRE \leftarrow SRE(T')$ $\triangleright$ Compute current SRE
  $Target\_SRE \leftarrow X \sim U(a, b)$
  $\triangleright$ Randomly sample a value from a target distribution $U(a, b)$
  $r \leftarrow Target\_SRE \;/\; Current\_SRE$ $\triangleright$ Compute the scaling ratio r
  $\theta \leftarrow \theta' \times r$ $\triangleright$ Scale $Current\_SRE$ by scaling θ'
  $T_0 \leftarrow Update\_Mat(T_{gt}, -\theta)$
  return T_0, θ $\triangleright$ Initialization T_0 and its label θ

2.3. Multi-Stage Registration

Essentially, deep learning-based image registrations map an initial distribution to a target distribution. Since these two distributions are different, applying the same pre-trained network to iteratively compute the registration is impractical (see Sect. 1). With this in mind, we introduce the coarse-to-fine multi-stage registration (MSReg) to improve the accuracy of MR-TRUS image fusion, trainable in two ways: (1) the network at each stage can be trained separately on a target distribution and then stacked together during the testing phase, as in Fig. 3; (2) the entire framework can be trained in an end-to-end fashion, as in Fig. 4.

With reference to Fig. 3, we limit the number of stages N to 3, as this has been shown to be sufficient in Sect. 3.2. When training the network of the ith stage, we first calculate the mean SRE value μ_i of its input data. Then we generate a training set whose SRE is draw from a uniform distribution on [0, 2μ_i mm] for training the stage-specified network. More precisely, the first of these three stages is trained on the uniform-distribution U_1 for [0, 20mm]. The second stage network is trained on the uniform distribution U_2 [0, 8mm]. Finally, the third network is trained on the uniform distribution U_3 [0, 7mm]. Using other types of distributions are also feasible, but here we choose to use a uniform distribution rather than a Gaussian distribution out of two reasons: (1) Setting the mean value and standard deviation of Gaussian distribution is a heuristic progress and may yield significant bias for the cases far away from the peak; (2) Uniform distribution can produce an even diffusion of values along the search space, making the network capable of producing valid registration for extreme cases.

Fig. 4 shows the steps of end-to-end training of the multi-stage registration, with N networks trained simultaneously. We generate the initialized transformations for the first network according to a target distribution. The outputs of the first network, i.e. the 6 degrees of freedom, are used to update the transformation. The updated transformation is then used to sample a new image pair which serves as the input to the next stage network. During this end-to-end training, a sequence of N networks can be trained together and the parameters in each network are updated through their respective MSE loss. When it comes to the scenario where the testing cases do not have ground-truth transformation matrix, we can assign one initial transformation matrix which roughly overlays the two volumes and then the proposed registration method can automatically optimize the positioning parameters for a better alignment.

3. Experiments

This section introduces the dataset and provides the details of network training. This is followed by presenting the experimental results. We also show the visualizations of several test cases at different registration stages using MSReg.

3.1. Materials

For the experiments performed in this study, a total of 679 sets of data from the Nation Institute of Health (NIH) were acquired from IRB-approved clinical trail MR-TRUS fusion-guided biopsy procedures. Each set contains a T2-weighted MR volume and a 3D ultrasound volume. The MR volume has a size of 512 × 512 × 26 voxels with the resolution of 0.3mm ×0.3mm×0.3mm. The TRUS volume is reconstructed from a freehand 2D ultrasound sweep of the prostate, through either electro-magnetic tracked positioning [2] or deep learning based ultrasound frame alignment [25]. The TRUS volumes have varying sizes and resolutions which are determined by the ultrasound scanning parameters associated with the reconstruction algorithm. The dataset is split into three groups: training, validation and testing sets, each consisting of 539, 70 and 70 cases, respectively.

The networks are trained on a NVIDIA DGX-1 deep learning server for 300 epochs with batch size of 16. We use Adam [26] optimizer with an initial learning rate of 5×10^{-5}, which decays by 0.9 after 5 epochs each time. The proposed architecture is implemented in Python using the open source PyTorch library [27].

3.2. Registration Results

After training the multi-stage registration framework, we get a series of (up to) 3 consecutive networks, capable of registering MR and TRUS image pairs from coarse to fine. The testing results are presented as follows.

First, we compare our results with other baseline methods, including the mutual information [4], MIND [6] and Deep Similarity Metric [12], which are all based on the similarity energy optimization. We set two experimental groups, each with a fixed SRE starting point as a constraint for every testing case. Specifically, for each testing case in the first experimental group, 6 degrees of freedom are randomly sampled and then scaled to make sure that the initial SRE equals to 8mm. Table 1 summarizes the results of our comparisons. Our best result, i.e. MSReg (2-stage End2end), consists of two consecutive networks which are trained in an end-to-end fashion on the generated dataset U_1. Comparing to the "Deep Metric (mp)" [12], which yields the best performance among energy-based methods, our proposed MSReg can significantly reduce the registration error at the confidence level of 0.05 (validated by paired *t*-test). In addition to the registration accuracy, our proposed method is also superior in computational speed as a single-forward method, as the other energy-based iterative methods are considerably more time consuming. For example, Deep Metric takes more than a minute to register a pair of 3D images, while our proposed MSReg can perform the same task with an average of 70.7 milliseconds per testing case, including all the preprocessing steps.

Next, we validate the effectiveness of the multi-stage registration strategy. For this, we generate 100 random transformations for each testing case without any SRE constraints, which yields a total of 7,000 testing cases. Such an enormous testing set, containing much more cases with varied starting SRE, can better test the robustness of our multi-stage training strategy. Following from the central limit theorem, the random sampling of 6 degrees of freedom in θ results in the SRE values to be normally distributed. Based on this initial transformation set, we did the multi-stage registration using the 3 consecutive networks, as discussed in Sec. 2.3. Fig. 5 shows the MSReg results during each registration stage, with the mean SRE ± the standard deviation of each stage provided below the bar plots measured in *mm*. To test for improvements, the *p*-values between two consecutive stages were calculated for comparison. In Fig. 5, four stars between the initialization and 1-stage registration indicated that the improvement was significant at the confidence level of 1×10^{-4}. In addition to that, there was also a significant improvement at the second stage registration. However, the difference between the 2nd-stage and the 3rd-stage was not significant. We therefore concluded that including additional networks is not required here.

3.3. Ablation Study

This section presents the ablation studies with different experimental settings such as the generated training set, the number of trainable parameters, and the multi-stage training strategy.

Table 2 summarizes the ablation study results, where testing data followed the same protocal as in Table 1. “Fixed Initialization” refers to the experiment where each subject in the training set only has one fixed initial transformation matrix. “1-stage” composes one network trained on U_1; “2-stage” composes two networks trained separately on U_1 and U_2, while the second network takes the first network’s outputs as inputs during the testing phase; “3-stage” works in the same manner. “1-stage (MixedDistr)” is a single network trained on a mixed distribution which is the combination of U_1 and U_2. “2-stage (End2end)” is the same as “MSReg (2-stage End2end)” discussed in Table 1. All the networks above use RsNeXt-50 backbone architecture, while “1-stage (Res-101)” is built based on ResNeXt-101.

It is important to note that the different results between “Fixed Initialization” and “1-stage” confirm the impact of the first contribution of this work, *i.e.* augmenting the dataset using generated transformations. With such errorscaling data augmentation, we can generate theoretically infinite numbers of training samples in any favorable distribution. Also, the network trained on this dataset performs more robustly and yields better results. By applying data generation to augment the training set in both experimental groups, the mean testing registration errors reduced from 6.76mm to 4.04mm (40.2% reduction) and from 13.86mm to 5.26mm (62.0% reduction) under 8mm and 16mm initializations, respectively.

To demonstrate the impact of the second contribution here, it is worth noting that a 1-stage registration is not as good as the Deep Metric method, which performed best among the baseline techniques. Nevertheless, applying the second-stage registration network showed significant improvements from the previous stage and outperformed each of the comparison methods.

To further demonstrate that our method benefits from the multi-stage coarse-to-fine registration strategy, not from the augmentation of training data distribution, we trained one network with a mixed distribution which merges two uniform distributions with ranges of [0, 20mm] and [0, 8mm] as “1-stage (Mixed-Distr)”. We can observe that such a network with mixed distribution cannot outperform the multi-stage networks which are trained separately on different distributions. The explanation is that the mixed distribution with multiple peaks distracts the network’s attention.

To intuitively visualize the performance improvement brought by the multi-stage registration, we plot the results of recursive registration and multi-stage registration in Fig. 6. Both methods use the same initializations. While they achieved the same performance at Stage1, the network trained in multi-stage can further reduce the registration error from B to C. In contrast, when we recursively apply the network trained with initializations only, the registration error started increasing as shown by the red curve. This indicates that our multi-stage registration strategy can adapt to the error distribution at each stage and make progressive improvements.

As state, in Sect. 2.3, we explored the influence of applying different strategies on the multi-stage networks training: networks at different stages trained separately and networks trained jointly in an end-to-end fashion. Again, the motivation of multi-stage registration is that the SRE distribution before and after one-step registration are very different from each other. In the separate-training strategy, we are creating the distribution using hyper-parameters (the mean SRE of the validation set). In contrast, end-to-end training can automatically create the training distribution for the second network, which is reasonably more accurate to the real distribution. While there seems to be a marginal improvement using end-to-end training, no significant difference is found between these two training strategies. However, end-to-end training is more straight-forward and easier to implement than the separate-training, as it does not require predetermined training data distribution.

Finally, it is worth noting that although we have achieved significant improvements comparing to the baseline methods in both registration error and running time, the framework's capacity will be doubled or even tripled due to the multi-stage registration strategy. One may argue that the marginal improvements are probably coming from the increased network parameters. However, while we continue to increase the network parameters by adding more layers, *e.g.*switching the network from ResNeXt50 to ResNeXt101 by increasing the number of parameters by 38.6%, there is no significant improvement in the registration performance.

3.4. Registration Visualizations

Fig. 7 shows four image pairs when registered using the multi-stage registration strategy introduced in this paper. The TRUS images are drawn in a color map for visual purposes. The different colors in the TRUS image indicate the organ's structural features. From left to right, the gradually decreasing SRE indicates that the registration accuracy is improving over time. Furthermore, the boundaries of the prostate in the MR and TRUS volumes are gradually aligning. Fig. 8 presents another image pair and displays three views at each registration stage, which are obtained by applying the same multi-stage registration strategy.

4. Conclusions and Future Work

This paper has introduced two innovations to the deep learning based 3D image registration framework. The first is a new strategy to to improve network training by generating augmented datasets according to the data distributions at each stage. The second contribution is the development of a coarse-to-fine multi-stage registration framework in which each network is trained on the data distribution of its corresponding stage. While currently the presented coarse-to-fine network is based on rigid registration, an extension to incorporate deformable registration is straightforward, which we will consider for future work.

The application of both innovations has yielded an improved accuracy when compared to existing techniques. For future work, we propose to incorporate different network architectures for different stages. We hypothesize that the 1-stage registration network should be more complex and applied to a larger input ROI for a better coarse registration. The 2nd-stage network should be less complex and applied to a smaller-sized input ROI to focus on local regions. In addition to that, we propose to explore the incorporation of deep similarity metrics [12] to directly evaluate the registration performance of each stage, which can provide direct feedback to the registration stages.

Acknowledgments

This work was partially supported by National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health (NIH) under awards R21EB028001 and R01EB027898, and through an NIH Bench-to-Bedside award made possible by the National Cancer Institute.

References

[1]. Siegel RL, Miller KD, Jemal A, Cancer statistics, 2019, CA: a cancer journal for clinicians 69 (1) (2019) 7–34. [PubMed: 30620402]

[2]. Pinto PA, Chung PH, Rastinehad AR, Baccala AA, Kruecker J, Benjamin CJ, Xu S, Yan P, Kadoury S, Chua C, et al., Magnetic resonance imaging/ultrasound fusion guided prostate biopsy improves cancer detection following transrectal ultrasound biopsy and correlates with multiparametric magnetic resonance imaging, The Journal of urology 186 (4) (2011) 1281–1285. [PubMed: 21849184]

[3]. Azizi S, Bayat S, Yan P, Tahmasebi A, Nir G, Kwak JT, Xu S, Wilson S, Iczkowski KA, Lucia MS, et al., Detection and grading of prostate cancer using temporal enhanced ultrasound: combining deep neural networks and tissue mimicking simulations, International journal of computer assisted radiology and surgery 12 (8) (2017) 1293–1305. [PubMed: 28634789]

[4]. Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P, Multi-modality image registration by maximization of mutual information, IEEE TMI 16 (2) (1997) 187–198.

[5]. Sun Y, Yuan J, Rajchl M, Qiu W, Romagnoli C, Fenster A, Efficient convex optimization approach to 3d non-rigid mr-trus registration, in: International Conference on MICCAI, Springer, 2013, pp. 195–202.

[6]. Heinrich MP, Jenkinson M, Bhushan M, et al., Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration, Medical image analysis 16 (7) (2012) 1423–1435. [PubMed: 22722056]

[7]. Khallaghi S, Sánchez CA, Rasoulian A, Sun Y, et al., Biomechanically constrained surface registration: Application to mr-trus fusion for prostate interventions, IEEE TMI 34 (11) (2015) 2404–2414.

[8]. Sparks R, Bloch BN, Feleppa E, Barratt D, Madabhushi A, Fully automated prostate magnetic resonance imaging and transrectal ultrasound fusion via a probabilistic registration metric, in: Medical Imaging 2013: Image-Guided Procedures, Robotic Interventions, and Modeling, Vol. 8671, International Society for Optics and Photonics, 2013, p. 86710A.

[9]. Zettinig O, Shah A, Hennersperger C, et al., Multimodal image-guided prostate fusion biopsy based on automatic deformable registration, IJCARS 10 (12) (2015) 1997–2007.

[10]. Haskins G, Kruger U, Yan P, Deep learning in medical image registration: a survey, Machine Vision and Applications 31 (1) (2020) 8.

[11]. Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS, Deep learning for visual understanding: A review, Neurocomputing 187 (2016) 27–48.

[12]. Haskins G, Kruecker J, Kruger U, Xu S, Pinto PA, Wood BJ, Yan P, Learning deep similarity metric for 3d mr–trus image registration, International journal of computer assisted radiology and surgery 14 (3) (2019) 417–425. [PubMed: 30382457]

[13]. de Vos BD, Berendsen FF, Viergever MA, Staring M, Išgum I, End-to-end unsupervised deformable image registration with a convolutional neural network, in: Deep Learning in Medical Image Analysis and Multi-modal Learning for Clinical Decision Support, Springer, 2017, pp. 204–212.

[14]. Eppenhof KA, Lafarge MW, Moeskops P, Veta M, Pluim JP, Deformable image registration using convolutional neural networks, in: Medical Imaging 2018: Image Processing, Vol. 10574, International Society for Optics and Photonics, 2018, p. 105740S.

[15]. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV, Voxelmorph: a learning framework for deformable medical image registration, IEEE transactions on medical imaging 38 (8) (2019) 1788–1800.

[16]. de Vos BD, van der Velden BH, Sander J, Gilhuijs KG, Staring M, Išgum I, Mutual information for unsupervised deep learning image registration, in: Medical Imaging 2020: Image Processing, Vol. 11313, International Society for Optics and Photonics, 2020, p. 113130R.

[17]. Rohé M-M, Datar M, Heimann T, Sermesant M, Pennec X, Svf-net: Learning deformable image registration using shape matching, in: International Conference on MICCAI, Springer, 2017, pp. 266–274.

[18]. Krebs J, Mansi T, Delingette H, et al., Robust non-rigid registration through agent-based action learning, in: International Conference on MICCAI, Springer, 2017, pp. 344–352.

[19]. Hu Y, Modat M, Gibson E, Li W, Ghavami N, Bonmati E, Wang G, Bandula S, Moore CM, Emberton M, et al., Weakly-supervised convolutional neural networks for multimodal image registration, Medical image analysis 49 (2018) 1–13. [PubMed: 30007253]

[20]. Hu Y, Gibson E, Ghavami N, Bonmati E, Moore CM, Emberton M, Vercauteren T, Noble JA, Barratt DC, Adversarial deformation regularization for training image registration neural networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 774–782.

[21]. Liao H, Lin W-A, Zhang J, et al., Multiview 2d/3d rigid registration via a point-of-interest network for tracking and triangulation, in: Proceedings of the IEEE Conference on CVPR, 2019, pp. 12638–12647.

[22]. Sokooti H, de Vos B, Berendsen F, Lelieveldt BP, Išgum I, et al., Non-rigid image registration using multi-scale 3d convolutional neural networks, in: International Conference on MICCAI, Springer, 2017, pp. 232–239.

[23]. de Vos BD, Berendsen FF, Viergever MA, Sokooti H, Staring M, Išgum I, A deep learning framework for unsupervised affine and deformable image registration, Medical image analysis 52 (2019) 128–143. [PubMed: 30579222]

[24]. Xie S, Girshick R, Dollár P, Tu Z, He K, Aggregated residual transformations for deep neural networks, in: Proceedings of the IEEE conference on CVPR, 2017, pp. 1492–1500.

[25]. Guo H, Xu S, Wood B, Yan P, Sensorless freehand 3d ultrasound reconstruction via deep contextual learning, arXiv:2006.07694.

[26]. Kingma DP, Ba J, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980.

[27]. Paszke A, Gross S, Chintala S, et al., Automatic differentiation in pytorch, in: NIPS 2017 Workshop Autodiff, 2017, pp. 1–4.

Highlights

- Fusing transrectal ultrasound (TRUS) and magnetic resonance (MR) images can improve prostate cancer diagnosis.
- Multi-modal image registration, especially MR-TRUS registration, is very challenging.
- Deep learning as a subspecies of AI has been used for multi-model image registration, which, however, requires large amounts of labeled transformations for network training.
- This paper proposes a new method of generating large amount of transformations following a targeted distribution to improve the network training.
- A coarse-to-fine multi-stage method to better map the distribution from source to target has been developed to further improve the performance.

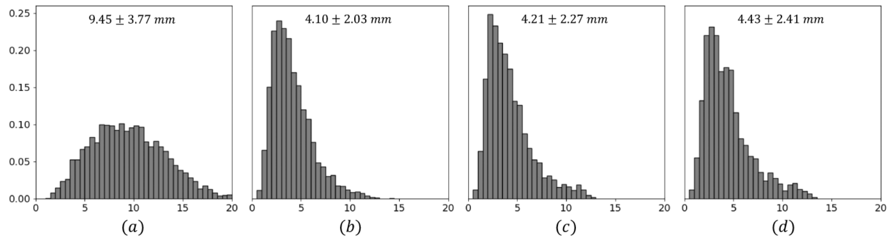

Figure 1:

(a) The initial registration error, (b) one-step iterative registration, (c) two-step iterative registration, and (d) three-step iterative registration.

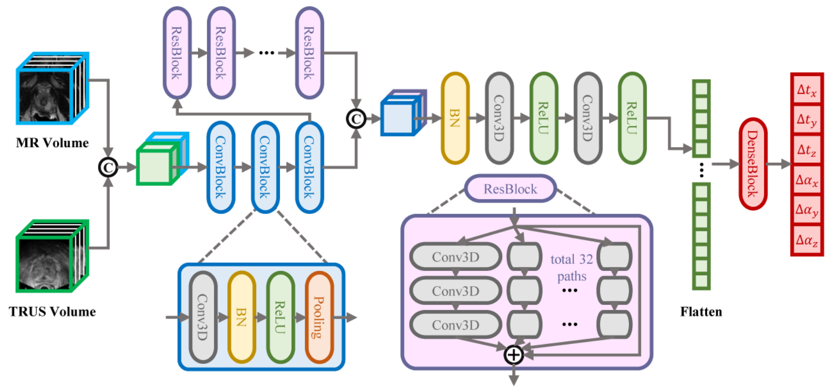

Figure 2:
The proposed network structure using ResNeXt as its backbone architecture.

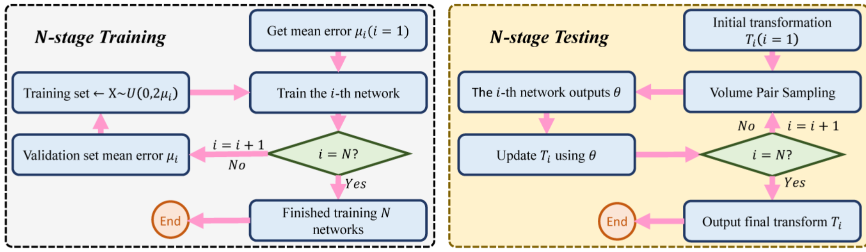

Figure 3:

Flowcharts of the multi-stage training and testing processes. During the training phase, the network at each stage is trained separately on its corresponding source distribution.

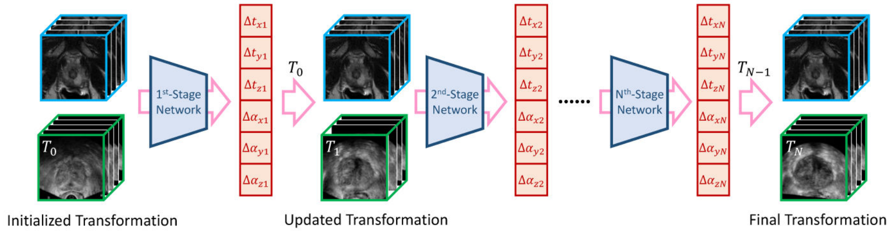

Figure 4:
Multi-stage end-to-end registration framework. A sequence of *N* networks can be trained in an end-to-end fashion.

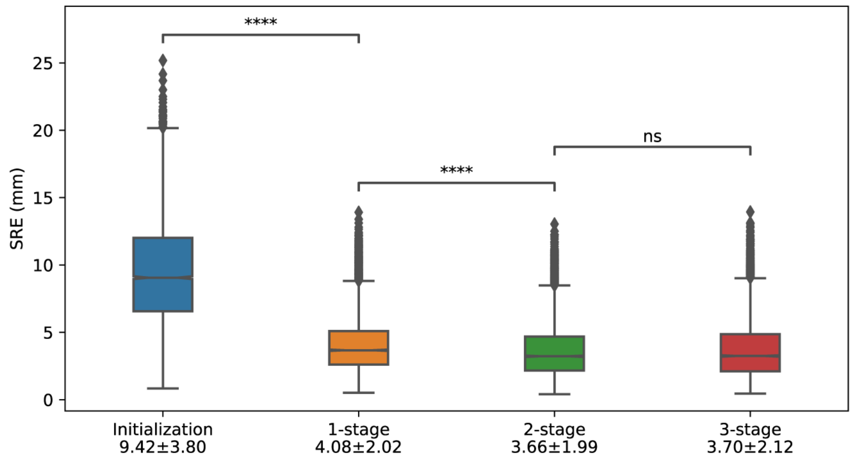

Figure 5:

Boxplots showing the SRE distribution of the testing set at each registration stage, including 7,000 registration samples. From left to right showing the SRE distribution of: (1) the initialization, (2) 1-stage registration, (3) 2-stage registration and (4) 3-stage registration.

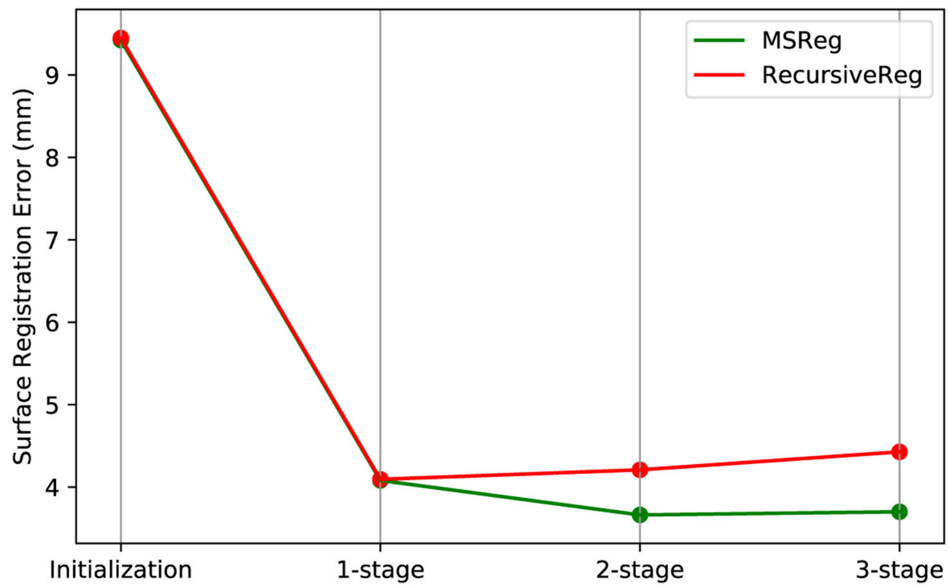

Figure 6:
Comparison of recursive registration and multi-stage registration with same initializations.

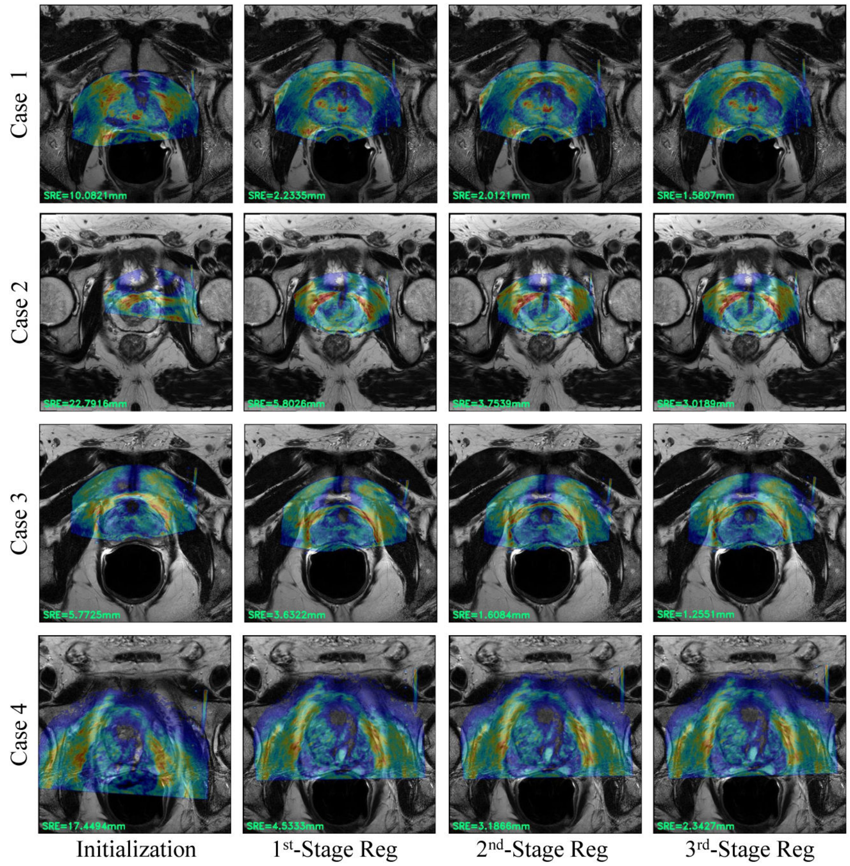

Figure 7:

Coarse-to-fine multi-stage registration examples, each row showing a different case. The columns from left to right indicate the registration stages from 1 to 3, respectively.

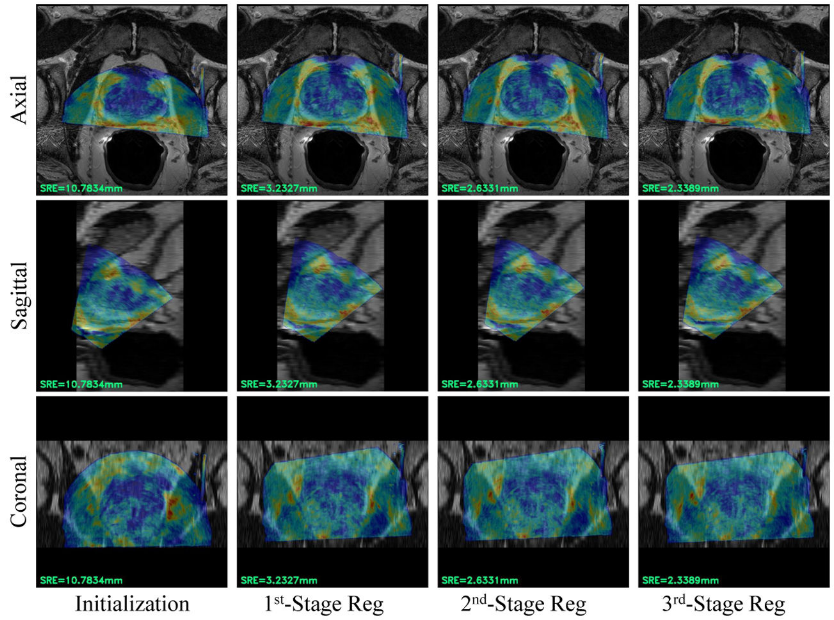

Figure 8:
Multi-stage registration visualization of a case from different views. From left to right, the registration error gradually decreases.

Table 1:

Comparison of the SREs using our MSReg method and other similarity metric-based registration methods.

Registration Method	Initial	Final mean±std (mm)	[min, max] (mm)
Mutual Information [4]	8mm	8.96±1.28	[5.50, 13.45]
SSD MIND [6]		6.42±2.86	[1.75, 10.64]
Deep Metric (sp) [12]		3.97±1.67	[0.77, 8.51]
Deep Metric (mp) [12]		3.82±1.63	[0.65, 8.80]
MSReg (2-stage End2end)		**3.57±2.03**	[0.92, 9.08]
Mutual Information [4]	16mm	10.07±1.40	[8.82, 14.09]
SSD MIND [6]		6.62±2.96	[1.58, 13.63]
Deep Metric (sp) [12]		4.21±1.64	[1.08, 8.68]
Deep Metric (mp) [12]		3.94±1.47	[1.53, 9.37]
MSReg (2-stage End2end)		**3.65±1.97**	[1.03, 10.37]

Table 2:

Results comparison using our MSReg method with different experimental settings.

Registration Method	**Initial**	**Final mean±std (mm)**	**[min, max] (mm)**	**Parameters**
Fixed Initialization	8mm	*6.76±1.29*	[2.81, 9.86]	11,614,198
1-stage		4.04±1.80	[1.53, 11.89]	11,614,198
1-stage (Res-101)		3.99±1.82	[1.64, 11.21]	16,095,756
1-stage (MixedDistr)		5.78±1.36	[1.89, 10.22]	11,614,198
2-stage		3.63±1.78	[0.96, 9.28]	23,228,396
2-stage (End2end)		**3.57±2.03**	[0.92, 9.08]	23,228,396
3-stage		3.70±2.09	[1.02, 10.78]	34,842,594
Fixed Initialization	16mm	*13.86±2.68*	[7.26, 19.92]	11,614,198
1-stage		5.26±2.15	[0.92, 11.75]	11,614,198
1-stage (Res-101)		5.14±2.06	[0.95, 10.32]	16,095,756
1-stage (MixedDistr)		10.34±2.47	[5.84, 14.73]	11,614,198
2-stage		3.75±2.03	[1.01, 11.04]	23,228,396
2-stage (End2end)		**3.65±1.97**	[1.03, 10.37]	23,228,396
3-stage		3.74±2.11	[0.98, 10.88]	34,842,594