Meta-Learning Initializations for Interactive Medical Image Registration

Zachary M. C. Baum,

Yipeng Hu,

Dean C. Barratt

Wellcome/EPSRC Centre for Interventional and Surgical Sciences, University College London, London W1W 7TS, U.K.,; UCL Centre for Medical Image Computing, University College London, London W1W 7TS, U.K.

Zachary M. C. Baum: zachary.baum.19@ucl.ac.uk; Yipeng Hu: yipeng.hu@ucl.ac.uk; Dean C. Barratt: d.barratt@ucl.ac.uk

Abstract

We present a meta-learning framework for interactive medical image registration. Our proposed framework comprises three components: a learning-based medical image registration algorithm, a form of user interaction that refines registration at inference, and a meta-learning protocol that learns a rapidly adaptable network initialization. This paper describes a specific algorithm that implements the registration, interaction and meta-learning protocol for our exemplar clinical application: registration of magnetic resonance (MR) imaging to interactively acquired, sparsely-sampled transrectal ultrasound (TRUS) images. Our approach obtains comparable registration error (4.26 mm) to the best-performing non-interactive learning-based 3D-to-3D method (3.97 mm) while requiring only a fraction of the data, and occurring in real-time during acquisition. Applying sparsely sampled data to non-interactive methods yields higher registration errors (6.26 mm), demonstrating the effectiveness of interactive MR-TRUS registration, which may be applied intraoperatively given the real-time nature of the adaptation process.

Index Terms

Medical image registration; meta-learning; interactive machine learning; prostate cancer

I Introduction

A Medical Image Registration

IMAGE registration is a fundamental task in medical imaging research whereby correspondence is established between anatomical structures in paired images. Using methodologies from "classical" iterative registration algorithms, learning-based methods have been proposed. Learning-based methods have used different architectures, such as convolutional neural networks [1–2] and vision transformers [3], different training strategies, such as generative adversarial networks [4, 5], supervised [1, 4], unsupervised [2, 6–8] or reinforcement learning [9–11], or different transformation constraints, based on parametric splines [6], diffeomorphism [12] and biomechanics [13]. Semi-supervised learning [14],

few-shot- and meta-learning [15–16], unsupervised contrastive learning [17], inference-time augmentation [16, 18], and amortized hyperparameter learning [19] methodologies have also been used to improve data efficiency and generalizability. For further discussion on these learning-based registration methods, readers are referred to recent systematic surveys [20–22].

B Interactive Machine Learning

For many clinical applications, registration errors may be identified or corrected by users, and integration of interactions into machine learning frameworks may assist in predicting more accurate solutions [23–26]. This integration is referred to as interactive machine learning (IML) [27]. Recently, IML-based methods for medical image analysis have focused on error correction in image segmentation. Most existing methods use simple interactions, such as user-defined bounding boxes as a guide for initial predictions [24] or 'scribbles' to indicate areas that should or should not be considered during refinement [23–24, 26]. Other methods may train multiple networks in tandem; one for segmentation and another for refinement [25].

Though the authors were unable to find existing IML-based methods for medical image registration in the literature, interaction has been utilized in classical methods. To improve the alignment of patient images, anatomical landmarks may be interactively selected [28] or acquired with spatially tracked intra-operative surgical instruments [29]. These methods are widely used and considered among the gold standard for medical image analysis methods.

C Gradient-Based Meta-Learning

Meta-learning [30–31] formalizes the commonly-applied fine-tuning intuition by iteratively learning to improve future performance on related tasks over multiple learning episodes. In particular, 'gradient-based' meta-learning approaches, such as model agnostic meta-learning (MAML) [32] and Reptile [33], learn adaptable initializations from gradients observed during learning episodes. Such algorithms are simple, learn quickly, and generalize well at test time with limited examples, as evidenced by their application in the medical imaging domain [15–16, 34–37]. Gradient-based methods have been used for domain-agnostic generalization and subsequent rapid-adaptation in image registration [15–16] and segmentation [34–37] on datasets of limited size from new domains.

Unlike most aforementioned examples, this work focuses on improving performance for individual tasks, formed by data that are varied by interactions. The simplicity in incorporating new data, efficiency in adaptation, and effectiveness in various computer vision and medical imaging applications are particularly desirable and motivate the use of meta-learning in formulating interactive registration. Other meta-learning methodologies [30] should be tested in future development.

D Contributions

We define a framework to meta-learn network initializations for interactive image registration. This framework consists of three components: a learning-based medical image registration algorithm, a form of user interaction, which is easily simulated in training, to

refine predictions at inference, and a gradient-based meta-learning protocol that learns a rapidly adaptable network initialization, by considering data variability due to interaction in individual patients as separate tasks.

To investigate the application of such a framework to clinical data, we register 3D magnetic resonance (MR) imaging volumes to a series of interactively-acquired sparse 2D transrectal ultrasound (TRUS) images for use in targeted prostate biopsy guidance. This exemplar application illustrates a clinical scenario in which real-time, interventional imaging, such as TRUS, is acquired interactively to iteratively refine the registration throughout a single acquisition of interventional imaging modality as it traverses the target anatomy. This work compares the accuracy of our proposed interactive registration method with alternative learning-based methods. We outline the key contributions in this work as follows:

- We provide a detailed description of our interactive meta-learning framework for medical image registration and describe how it may enable a range of useful applications.

- We introduce and describe the registration, interaction, and meta-learning strategy for our exemplar clinical application; volume-to-sparse registration of prostate MR to TRUS.

- We present rigorous validation experiments, comparing our method to various learning-based methods for prostate MR-TRUS registration, including variations to meta-learning parameters to assess their effects on the registration process.

II Learning-Based Interactive Image Registration

A Learning-based Image Registration

Learning-based registration may be categorized from an application perspective; network inputs may be unimodal, multimodal, inter-patient, or intra-patient – with each image bearing its own dimensionality [20], requiring different loss functions based on image similarity [6], label similarity [1], or some combination of the two [2]. Each image pair may encompass any number of anatomical sites of clinical interest, requiring a registration method to utilize different deformation models, commonly, rigid, affine, or deformable [20].

Given N pairs of training source and target images, $\{x_n^{source}\}$ and $\{x_n^{target}\}$, and accompanying source and target labels, $\{l_n^{source}\}$ and $\{l_n^{target}\}$, respectively, where $n = 1,\ldots,N$, existing approaches predict the voxel correspondence or transformation $u_n^{\phi} = f^{\phi}(x_n^{source}, x_n^{target})$ using a registration network f^{ϕ} with network parameters or weights ϕ The training goal thus is minimizing an image and/or label loss function $\mathscr{L}_{sim}$ over N training pairs, to obtain the optimal $\hat{\phi}$:

$$\hat{\phi} = \arg \min_{\phi} \sum_{n=1}^{N} \left[\mathscr{L}_{sim}(\phi) + \alpha^{def} \mathscr{L}_{def}(\phi) \right], \tag{1}$$

where $\mathscr{L}_{def}(\phi \mid x_n^{source}, x_n^{target}) = \mathscr{L}_{def}(f^{\phi}(x_n^{source}, x_n^{target}))$ provides regularization on the deformation smoothness u_n^{ϕ}, weighted by α^{def}. In general, the similarity-based loss can further combine a negative unsupervised image similarity function $\mathscr{L}_{sim}^{image}(x_n^{source}(u_n^{\phi}), x_n^{target})$, between the

transformation-warped images $\boldsymbol{x}_n^{source}(\boldsymbol{u}_n^\phi)$ and the target images $\boldsymbol{x}_n^{target}$, and a negative weak-supervision loss based on label similarity $\mathscr{L}_{sim}^{label}(\boldsymbol{l}_n^{source}(\boldsymbol{u}_n^\phi), \boldsymbol{l}_n^{target})$, between the warped source labels $\boldsymbol{l}_n^{source}(\boldsymbol{u}_n^\phi)$ and the target labels $\boldsymbol{l}_n^{target}$:

$$
\begin{aligned}
\mathscr{L}_{sim}\,(\phi \mid \boldsymbol{x}_n^{source}, \boldsymbol{x}_n^{target}, \boldsymbol{l}_n^{source}, \boldsymbol{l}_n^{target}) = \\
\alpha^{image}\,\mathscr{L}_{sim}^{image}(\boldsymbol{x}_n^{source}(\boldsymbol{u}_n^\phi), \boldsymbol{x}_n^{target}) + \\
\alpha^{label}\,\mathscr{L}_{sim}^{label}(\boldsymbol{l}_n^{source}(\boldsymbol{u}_n^\phi), \boldsymbol{l}_n^{target}),
\end{aligned} \tag{2}
$$

where the general form contains hyperparameters α^{image} and α^{label} which may be set to zero to represent weakly supervised and unsupervised algorithms, respectively.

B Interaction for Image Registration

In general, the performance improvement seen in other interactive applications, such as the above-discussed interactive segmentation [34–37], may be expected from interactive registration. Other benefits, such as those related to expandability, and owing to the human-in-the-loop of machine learning models for registration applications are also important, but are considered out of the scope for this work.

To adapt existing learning-based registration methods to accept interactions, we must first define interactions that may be learned in training, and are feasible at test-time. We consider interaction to be any action taken by the user. Depending on application-specific needs, a combination of sequential interactions may best improve the registration.

This user-to-computer interaction may entail image reacquisition or annotation of poorly aligned areas. Image reacquisition may be local (i.e. one, or a few images) or global (i.e. entire image volume) when, for example, image quality is poor, or there has been patient motion. Local re-acquisition is pertinent when using real-time imaging modalities, such as ultrasound, that can be rapidly acquired.

We propose formulating image reacquisition and annotation additional labelled data, where the quantity and availability of labels or images may vary per application. In practice, interactions may be application-specific. The determination of use-cases for each interaction is considered out of scope for this work, though we provide a description and evaluation of the additional data interaction for ultrasound-guided prostate biopsy to illustrate a possible use of the proposed framework.

C Meta-Learning Interactive Initializations

In this work, we train registration networks in the inner loop of a meta-learning optimization to accept newly labelled data provided by interaction, while the network adaptability across subjects is optimized in an outer loop. For a given test subject, this enables the interactively-acquired data to adapt the trained registration network efficiently, before being used in inference.

III Methods

A Images and Annotations as Interaction

We denote possible pairs of interactions sampled from the source and target images as $\{i_{mn}^{source}\}$ and $\{i_{mn}^{target}\}$, from n training data. Each nth pair is also associated with M_n interactions that are possible on image pair n, $m = 1,\ldots,M_n$. These time-agnostic interactions are represented as sets of interactively obtained images $\{x_{mn}^{source}\}$ and $\{x_{mn}^{target}\}$ and annotations, in the form of segmentation labels, ($\{\ell_{mn}^{source}\}$ and $\{\ell_{mn}^{target}\}$), i.e. $i_{mn}^{source} = \left[(x_{mn}^{source})^T, (\ell_{mn}^{source})^T\right]^T$ and $i_{mn}^{target} = [(x_{mn}^{target})^T, (\ell_{mn}^{target})^T]^T$. For notational brevity, both images and annotations can include the previously available annotated data, for individual subject, therefore the interactions $\{i_{mn}^{source}\}$ and $\{i_{mn}^{target}\}$ are interchangeably used with interaction-updated source and target, respectively. A sequence of interactions may benefit from explicit sequential modelling; however this is considered out of scope of this work, where only a few steps of interaction is considered feasible in the application of interest.

This formulation does not distinguish between registrations which may have different initial image and annotation data from one without such initial registration, as they can be consistently represented by both the non-interactive registration formulation, described in Section II.A, and the interactive adaptation, described in Section III.B.

We note that not all the interactive image or annotation data need to be available or varying for a given interaction. We describe a sample of scenarios which demonstrate the versatility of interactive registration. Additionally, active learning methodologies [38] may appear similar in nature, and may be able to utilize similar scenarios for interactive learning in practice. Our application is developed and validated with respect to Scenario 4, a special case of Scenario 3. Though not tested, other scenarios are included for discussion purposes.

1. Successive user-defined image annotations improve the registration over multiple interactions, i.e. variable labels $\ell_{m=a,n}^{source} \neq \ell_{m=b,n}^{source}$ and $\ell_{m=a,n}^{target} \neq \ell_{m=b,n}^{target}$ but fixed images $x_{m=a,n}^{source} = x_{m=b,n}^{source}$ and $x_{m=a,n}^{target} = x_{m=b,n}^{target}$, when $a \neq b$.

2. An unsupervised learning algorithm, without initial labels $\ell_{m=0,n}^{source}$ and $\ell_{m=0,n}^{target}$, receives successive user-defined annotations to improve alignment, which requires the simulation of $\ell_{m>0,n}^{source}$ and $\ell_{m>0,n}^{target}$ during training.

3. An image-guidance application may have a fixed pre-operative image x_{mn}^{source}, but adds new intra-operative images, i.e. $x_{m=a,n}^{source} \neq x_{m=b,n}^{source}$ and $x_{m=a,n}^{target} = x_{m=b,n}^{target}$, when $a \neq b$. This application may use user-defined annotations on the pre- and intra-operative images, as in Scenario 1.

4. An ultrasound-guided prostate cancer application, such as that used in this work; similar to Scenario 3, but does not require new annotations on the source images, however, additional annotations on the target images may be acquired automatically using a segmentation network, i.e. using the generation of labelled intra-operative ultrasound images as the interaction, $x_{m=a,n}^{source} = x_{m=b,n}^{source}$, $\ell_{m=a,n}^{source} = \ell_{m=b,n}^{source}$, $x_{m=a,n}^{target} \neq x_{m=b,n}^{target}$ and $\ell_{m=a,n}^{target} \neq \ell_{m=b,n}^{target}$, when $a \neq b$.

B   Meta-Learning for Interactive Registration

As the interaction data $\{i_{mn}^{source}\}$ and $\{i_{mn}^{target}\}$ are defined as images and annotations – in Section III.A – $\{x_{mn}^{source}\}$, $\{x_{mn}^{target}\}$, $\{\ell_{mn}^{source}\}$ and $\{\ell_{mn}^{target}\}$, they are consistent with the data used in the non-interactive registration formulation – in Section II.A – $\{x_n^{source}\}$, $\{x_n^{target}\}$, $\{l_n^{source}\}$ and $\{l_n^{target}\}$. We propose to formulate the training of an interactive registration network $f^{\tilde{\phi}}$ by adapting the optimization in Eq. (1) to a bi-level optimization [30, 38], therefore learning the interactive image registration becomes a meta-learning problem:

$$\tilde{\phi} = \arg\min_{\phi} \sum_{n=1}^{N} \sum_{m=1}^{M_n} \left[\mathcal{L}_{sim}^{*}\left(\phi^{*}(\phi)\right) + \right.$$
$$\left. \alpha^{def} \mathcal{L}_{def}^{*}\left(\phi^{*}(\phi)\right) \right] \tag{3}$$

$$s\,.\,t\,.\,\phi^{*} = \arg\min_{\phi} \sum_{n=1}^{N} \sum_{m=1}^{M_n} \left[\mathcal{L}_{sim}^{*}(\phi) + \alpha^{def} \mathcal{L}_{def}^{*}(\phi) \right], \tag{4}$$

where $\mathcal{L}_{sim}^{*}$ is obtained by substituting interaction in Eq. (2):

$$\mathcal{L}_{sim}^{*}(\phi) = \mathcal{L}_{sim}\left(\phi \mid x_{mn}^{source}, x_{mn}^{target}, \ell_{mn}^{source}, \ell_{mn}^{target}\right) \tag{5}$$

similarly, $\mathcal{L}_{def}^{*}(\phi) = \mathcal{L}_{def}\left(\phi \mid x_{mn}^{source}, x_{mn}^{target}\right)$. $\mathcal{L}_{sim}^{*}(\phi^{*}(\phi))$ and $\mathcal{L}_{def}^{*}(\phi^{*}(\phi))$ denote the optimized functions of ϕ, by optimized ϕ^{*} at the inner-level. ϕ^{*} hereinafter used for brevity.

It is noteworthy that, unlike the training defined in Eq. (1) which minimizes the expected loss over the N pairs of training images, the task-specific inner-level Eq. (4) aims to minimize the expected loss over the M_n samples of interactions. At the outer-level, Eq. (3), different N pairs of images and annotation are usually sampled to learn the optimal network parameters, such that, at inference, the network $f^{\tilde{\phi}}$ can be adapted to new pairs of interactions $\{i_{m,test}^{source}\}$ and $\{i_{m,test}^{target}\}$, where $m = 1, …, M_{test}$ and be generalized to this new *test* task, i.e. we define the training meta-tasks be N different cases that need registration, rather than M_n steps of interactions.

Such a meta-learning framework learns an initialization of network parameters $\tilde{\phi}$ which enables data-efficient adaptation to a new task at inference. The efficient adaptation means that registering a new pair of images x_{test}^{source} and x_{test}^{target} may only require a few M_{test} steps of interaction, often constrained by human effort and time-critical applications.

C   Gradient-Based Meta-Learning Algorithms for Network Initialization

Gradient-based meta-learning algorithms are applicable for training the proposed interactive registration and are comprised of the meta-learning and the meta-test phases. To start meta-training, the registration model is initialized with random weights. During each iteration of the outer-level loop, one task $\left(i_{mn}^{source}, i_{mn}^{target}\right)_n$ is randomly sampled from the task set $\{\left(i_{mn}^{source}, i_{mn}^{target}\right)_{n=1,…,N}\}$ containing all possible tasks, with a set of k interactions $\{\left(i_{mn}^{source}, i_{mn}^{target}\right)_{n,m=1,…,k}\}$ randomly sampled from this given task, to form an episode (Fig. 1). Each sampled task corresponds to a task-specific loss in Eq. (4). We define our meta-learning task as a pair of source and target images with their associated source and target

annotations, from each subject. During each episode we undergo 'task-level learning' using stochastic gradient descent (SGD) or its variants, for k SGD steps, the task-specific gradient $g_n^m(\phi)$ can be computed to update the network weights ϕ:

$$\phi_m^* \leftarrow \phi - \beta^{task} \cdot g_n^m(\phi), \qquad (6)$$

where

$$g_n^m(\phi) = \frac{\partial}{\partial \phi}\big[\mathcal{L}_{sim}^*(\phi) + \alpha^{def}\mathcal{L}_{def}^*(\phi)\big], \qquad (7)$$

and β^{task} is the learning rate. After an episode of k steps, a cross-task gradient $g_n(\phi^*)$ is used to update the network weights at the outer-level loop, corresponding to Eq. (3):

$$\phi_n \leftarrow \phi - \beta^{meta} \cdot g_n(\phi^*), \qquad (8)$$

where

$$g_n(\phi^*) = \frac{\partial}{\partial \phi}\big[\mathcal{L}_{sim}^*(\phi) + \alpha^{def}\mathcal{L}_{def}^*(\phi)\big](\phi^*), \qquad (9)$$

and β^{meta} is the meta-learning rate. With gradient-based meta-learning methods, such as MAML [32], the cross-task meta-gradient $g_n(\phi^*)$ is computed directly to obtain the Jacobian for updating parameters, at the inner-loop-optimized weight values ϕ^*. However, estimating the Jacobian involves computationally problematic second derivates; First-Order MAML [32] and Reptile [33] have been proposed to approximate this meta update step, and we adapt such approximations to train the interactive registration network.

In the meta-test phase, parameters $\tilde{\phi}$ are adapted to the test task through few-shot learning. During meta-testing, a few interactions $\{i_{m,\,\text{test}}^{source}\}$ and $\{i_{m,\,test}^{target}\}$ are acquired from the test task to compute a few steps of test-task-specific gradients and update the model, using Eq. (6), before predicting the transformation using images x_{test}^{source} and x_{test}^{target} (Fig. 2).

D Exemplar Clinical Application: Volume-to-Sparse Weakly-Supervised Multimodal Image Registration

In this section, we discuss our proposed methods for interactive registration to a real-world clinical application, in which only sparse TRUS images are available to be registered to preoperative MR images, using an interactive weakly-supervised multimodal image registration.

Prostate MR-TRUS image registration leverages MR imaging to aid tumour-targeted needle biopsy [40–47] and focal treatments [48–49] for suspected clinically significant prostate cancer. Image registration allows the presentation of MR-visible information, such as tumour size and location, for guiding surgical instruments or therapeutic energy placement. Often, the MR-derived lesion and tumour information are superimposed on the TRUS images as a visual aid.

A weakly-supervised methodology used to train an interactive registration network with a label-driven loss can be considered as a meta-learning problem, as described in Eq. (3) and Eq. (4), with $a^{image} = 0$, without using explicit intensity-based similarity measures which have been considered less effective [1]. To accommodate sparse ultrasound images, readily available as interactions in this application, we develop a volume-to-sparse registration algorithm, where the training target images being a set of TRUS slices $\{x_{mn}^{target}\}$ and annotations of anatomical structures identified on these slices $\{l_{mn}^{target}\}$, with source MR images $\{x_{mn}^{source}\}$ and the corresponding MR annotations $\{l_{mn}^{source}\}$. These annotations can contain multiple types of anatomical structures [1], though this notation is omitted for brevity. We discuss the detailed representation of the interactive data in Section III.E and the need for TRUS slice localization information in the Discussion.

Our implementation utilizes LocalNet, a recent method for weakly-supervised image registration [1]. LocalNet's encoder-decoder structure comprises down- and up-sampling blocks and can predict a DDF that is summed over multiple resolutions. LocalNet is similar to the UNet [50] architecture found in VoxelMorph [2] – often used for unsupervised and weakly-supervised image registration. Compared to VoxelMorph, LocalNet has a smaller memory requirement and is more densely connected, with multiple types of residual shortcuts and summation-based skip layers to allow deeper supervision [1].

E Interactive Acquisition of Labelled TRUS Images

This study investigates an MR-TRUS registration where volume-to-sparse registration continually re-occurs throughout acquisition, as opposed to discreate registration to reconstructed 3D TRUS volumes. The continuous 2D TRUS images in such registration are considered the addition of new data, with or without the automatically acquired prostate gland segmentation [51], as interactions. At inference, this continuous stream of interactively acquired data provides additional context and a constantly up to date registration.

Here, interaction stems from the continual acquisition of frames by moving TRUS probe. Therefore, during few-shot learning, new frames are incorporated into the input of the model. This requires knowledge of the spatial relationship between each frame, so that the new frame may be inserted into the correct location within the TRUS volume. To provide initial spatial information for the network, the first interaction comprises two frames, and subsequent interactions require at least one new frame. Given the current clinical workflow for tumour-targeted needle biopsies, this interaction is unlikely to introduce any delay or modifications to existing protocols.

To simulate interactions in training, we select one pair of target interactions i_{mn}^{target} by randomly selecting a series of TRUS images in a clinically feasible manner, whilst the target "interaction" is the fixed MR images and their annotation i_{mn}^{source}, as described in Scenario D in Section III.A. The label pair ℓ_{mn}^{source} and ℓ_{mn}^{target} may define either the prostate boundary, the apex and base of the prostate, or any other patient-specific landmarks; such as zonal structures, water-filled cysts, and calcifications [52]. The binary mask is generated to randomly include some number of frames F, where $F \in \mathbb{N} : F \in [F_{min}, F_{max}]$, which defines the image

slices within the TRUS volume $\boldsymbol{x}_{mn}^{target}$. Once generated, sections of the input image $\boldsymbol{x}_{mn}^{target}$, and corresponding label $\boldsymbol{\ell}_{mn}^{target}$ are masked-out, leaving only TRUS slices and corresponding labels from the simulated acquisition.

F   Meta-Learning an Initialization with Reptile

We adopt Reptile [33] as our gradient-based meta-learning strategy. Reptile provides a computationally efficient optimization of the gradient-based update procedure to approximate Eq. (8) and Eq. (9) by:

$$\phi_n \leftarrow \phi - \beta^{meta} \cdot \frac{1}{k} \sum_{m=1}^{k} (\phi - \phi_m^*),$$
(10)

where ϕ_m^* can be estimated using Eq. (6).

It is of note that, given that the complete prostate (and other patient-specific landmarks) labels are available, a stronger form of supervision is used to compute the loss during meta-training, such that the entire label similarity is computed rather than a partial similarity on sparse labels. This allows the initialization to be learned from complete data, illustrating how interactive labels and images used in computing training losses may differ from those seen in meta-testing in order to better guide learning.

During the meta-test phase, for evaluation, few-shot learning occurs with F gradient updates on interactions $\boldsymbol{x}_{mn}^{target}$ and $\boldsymbol{\ell}_{mn}^{target}$ from the test task. This fine-tunes the model to obtain adapted parameters ϕ' which can perform accurate registrations on the test patient. Unlike the random generation of interactions during the meta-training phase, $\boldsymbol{x}_{mn}^{target}$ and $\boldsymbol{\ell}_{mn}^{target}$ define a continuous, single-sweep TRUS acquisition. Therefore, the first few-shot learning gradient update contains F_{min} images and subsequent updates add an image, until the final update with $F_{max} - 1$ images. This ensures that the inference step is computed on an input with F_{max} images. During the meta-test phase, we only use the label which defines the prostate boundary. This is done to emulate the labels which may be available (via automatic segmentation) in practice with the application of our method. A visual summary of the meta-learning phases for our application is shown in Fig. 3.

G   Loss Functions

Two loss functions are used in training. In weakly-supervised registration, we seek to maximize the expected label similarity using a multiscale soft probabilistic Dice [1], which has shown effectiveness, especially when small foreground labels do not overlap initially. Using interactively acquired TRUS labels $\boldsymbol{\ell}_{mn}^{target}$ and pre-operative MR labels $\boldsymbol{\ell}_{mn}^{source}$. we obtain:

$$\mathcal{L}_{sim}^*(\phi) = \frac{1}{Z} \sum_{\sigma} \mathcal{S}_{Dice}\left(f_\sigma(\boldsymbol{\ell}_{mn}^{target}), f_\sigma(\boldsymbol{\ell}_{mn}^{source}(\boldsymbol{u}_n^\phi))\right),$$
(11)

where S_{Dice} is the soft probabilistic Dice [53], f_σ is a 3D Gaussian filter with an isotropic standard deviation $\sigma \in \{0, 1, 2, 4, 8, 16, 32\}$ in mm, and $Z = |\sigma|$. We additionally use bending energy [54] to regularize deformation $\mathcal{L}_{def}^*(\phi)$ on $\boldsymbol{u}_n^\phi$ in tandem with $\mathcal{L}_{sim}^*(\phi)$ as in Eq. (3) and Eq. (4).

H Data

We used 108 pairs of pre-operative T2-weighted MR and intraoperative TRUS images from 76 patients, acquired during the SmartTarget clinical trials [52], a study approved by the London-Dulwich Research Ethics Committee (REF 14/LO/0830) and conducted at University College London Hospital (UCLH). Images were split into training and test sets comprising 88 and 20 images, respectively. No patient appears in both sets. Images were normalized and resampled to an isotropic voxel size of $0.8 \times 0.8 \times 0.8$ mm^3. MR segmentations were acquired as part of the SmartTarget clinical trial protocols [52]. TRUS prostate gland segmentations were acquired automatically [51], and landmarks were manually segmented.

I Baseline Model Implementation and Training

The framework was implemented in TensorFlow [55] and Keras [56]. The weakly-supervised registration framework and loss functions were adapted from DeepReg [57]. Hyper-parameters are as described in [1] unless otherwise specified. A random affine transformation, without flipping, was applied to each image-label pair for data augmentation.

The Baseline interactive registration model was trained for 250000 iterations with the Adam optimizer [58], a minibatch size of 4, and an initial learning rate, β^{task}, of 1×10^{-5}. In the meta-training phase, the value of k for task-level learning was 10, and the initial meta-learning rate, β^{meta}, was set to 0.5, with a linear decay to 1×10^{-5} at the final training iteration. Loss weights γ and a were both set to 1.0. We let $F_{min} = 2$ and $F_{max} = 10$. Training took approximately 120 hours using one Tesla V100 GPU. We note that the number of iterations comprises each episode of task-level training, but does not include the meta-update; such is to say that we perform 25000 episodes of task-level learning, where each episode of task-level learning encompasses k gradient updates.

J Comparison with Meta-Learning Variants

Without extensively searching all hyper-parameters, which may misrepresent generalizability, we provide experimental results and validation on variants to the Baseline. First, we modify the number of gradient updates performed in task-level learning, k, to 1 and 100. Notably, when $k = 1$, a single step of SGD on the expected loss is equivalent to jointly training on a mixture of all tasks [33]. Though k is often defined as 10 in other meta-learning applications [33], we also demonstrate training with a higher value. Due to the changes introduced to the training process (for $k = 1$), and the deviation of the gradients from those which would normally be encountered in a non-meta-learning-based training protocol (for $k = 100$), these variants will likely underperform relative to the baseline. Second, we modify the initial meta-learning rate, β^{meta}, to 0.25 and 1.0. The linear decay remains unchanged. To prevent arbitrary selection, we choose values corresponding closely to those presented in [33]. Finally, we vary the maximum number of frames used in training, F_{max}, to 5 and 15. We expect a higher and lower F_{max} would result in better and worse performance, respectively. Though if the increase in performance gained per frame diminishes as F_{max} increases, training with a smaller F_{max} may be beneficial. Conversely, if

the increase in performance per frame does not significantly diminish, training with a higher F_{max} and acquiring additional frames in practice may be prudent.

K Comparison with State-of-the-Art Approaches

We compare the proposed Baseline to the application of 'registration' without alignment, and of a simple initialization whereby the prostate gland centroids are aligned. Furthermore, we compare to two weakly-supervised state-of-the-art approaches for deformable pairwise medical image registration; LocalNet [1], and VoxelMorph [2]. In all comparisons, we use complete 3D volumes for source and target input images – unlike our interactive meta-learning approach which provides a sparse target input. Hyper-parameters are all kept at defaults as described in [1] and [2], and we set loss weights γ and a to 1.0.

L Comparison with Non-Meta Learning Approaches

We emulate the sparse 2D target input of our interactive meta-learning approach with instances of LocalNet and VoxelMorph with 5 or 10 randomly sampled 2D target input images in training. We demonstrate the effects of few-shot learning on these models trained without meta-optimization and our meta-learning Baseline by performing inference with and without any few-shot learning. To illustrate the effectiveness of the meta-learned initialization, we randomly initialize LocalNet and VoxelMorph models and apply few-shot learning to the networks. While the impact of sparse data was not investigated in [1] or [2], and therefore, may adversely impact their performance, this assessment provides a benchmark to which we may compare the performance of our approach to learning-based methods with comparable amounts of input data.

M Evaluation of Registration Methods

To compare the Baseline to all aforementioned methods, we test interactions which represent a clinically realistic scenario on our real-world, clinical test data. Through the continuous acquisition of frames from a right-to-left sweep through the prostate, we obtain a series of sagittal images which are uniformly distributed through the prostate (Fig. 4). As noted in Section III.F, we initially acquire two images (as $F_{min} = 2$) to provide spatial context of the frames in this first acquisition.

Registration accuracy was quantified using the Dice similarity coefficient (DSC) and target registration error (TRE). Two-tailed paired t-tests, at significance level $a = 0.05$, are used to compare each method to the Baseline. DSC is computed between the warped MR label and the entire ground-truth TRUS label. TRE is defined as the root-mean-square distance between the geometric centroids of the registered landmark pairs. In our dataset, landmarks consisted of 309 pairs of manually identified points, including the apex and base of the prostate, and various patient-specific landmarks including zonal boundaries, water-filled cysts, and calcifications [1, 52, 59]. Notably, such landmarks have been previously utilized yield an overall spatial distribution which is representative of the full TRE distribution in this application [1, 4–5, 7–8, 10–11, 13, 60–70], therefore providing an evaluation of registration accuracy and an estimate of registration errors, such as those associated with tumour localization. We also report the computational time per few-shot learning gradient update and subsequent registration in the meta-test phase for our approach.

IV Results

A Baseline Performance

One gradient update and inference for the Baseline model requires 0.67 ± 0.07s and 0.37 ± 0.05s, respectively. Therefore, during adaptation, which may occur during image acquisition, approximately 6s is needed to perform fine-tuning and inference, considerably faster than the 2-4 mins required for acquisition, contouring, and registration in conventional image-fusion targeted biopsies, such as those reported in the SmartTarget clinical trials [52].

After few-shot learning, we achieved a median TRE of 4.26 mm and a mean DSC of 0.85 with 10 input TRUS frames. This is within range of previously defined clinically significant thresholds of 2.97 mm [71] and 5.00 mm [62]. A detailed summary of TRE and DSC is given in Table 1. Example slices of input MR and TRUS image pairs and registered MR images are provided in Fig. 5 for qualitative visual assessment for the Baseline through each few-shot step in the meta-test phase.

B Performance of Baseline Variants

After few-shot learning, the $k = 1$ variant had a median TRE of 4.48 mm and mean DSC of 0.83, whereas the $k = 100$ variant had a median TRE of 4.58 mm and mean DSC of 0.85. In both, no significant difference was found between TRE or DSC relative to the Baseline. The effects of k in training on TRE are illustrated in Fig. 6 and summarized in Table 2.

After few-shot learning, the $\beta^{meta} = 0.25$ variant had a median TRE of 4.33 mm and a mean DSC of 0.84, whereas the $\beta^{meta} = 1.0$ had a median TRE of 3.29 mm and a mean DSC of 0.87, no significant difference to the Baseline, as summarized in Table 3 and Fig. 6 with varying β^{meta}.

After $F_{max} - F_{min}$ gradient updates of few-shot learning, the $F_{max} = 5$ variant had a median TRE of 4.50 mm and mean DSC of 0.85, whereas the $F_{max} = 15$ variant had a median TRE of 3.58 mm and mean DSC of 0.84, no significant difference to the Baseline. The effects of F_{max} during training on TRE are illustrated in Fig. 6, and given for TRE and DSC in Table 4.

We note that the $F_{max} = 5$ variant performs better than the $F_{max} = 15$ variant for all values of $F \ 5$. This is likely due to the distribution of the input images in the presented clinical scenario, whereby one continuous sweep of the prostate occurs, as presented in Fig 4. For example, when $F = 5$, while the input frames of the $F_{max} = 5$ variant will be evenly distributed across the entire prostate, while the 5 input frames of the $F_{max} = 15$ variant will be condensed into the right-most third of the prostate, resulting in less spatial information being presented about the remaining prostate volume.

Example slices of input MR and TRUS image pairs and registered MR images are provided in Fig. 7 for qualitative visual assessment of the results for each variant.

C   Performance of State-of-the-Art Approaches

With no initial registration or alignment, a median TRE of 32.4 mm and mean DSC of 0.66 are obtained. Further, a median TRE of 18.4 mm and mean DSC of 0.77 are obtained if only prostate gland centroid alignment is performed.

The performance of the Baseline model was not significantly different than LocalNet for TRE and DSC, where a median TRE and mean DSC of 3.97 mm and 0.87 are obtained. The performance of the Baseline model was not significantly different than voxelMorph, for TRE and DSC, where a median TRE and mean DSC of 4.32 mm and 0.84 are obtained.

A summary of TRE for the Baseline and the non-meta-learning-based methods is given in Fig. 8. Example slices of input MR and TRUS image pairs and the registered MR images are provided in Fig. 9 for qualitative visual assessment of the registration results for each approach. It is important to note that these methods use complete 3D volumes for source and target input images, and only achieve comparable performance to our method, which uses between two and ten frames of the target image in training and at inference. This represents between 1.6% and 8.5% of the complete 3D volume, which contains 118 image slices.

D   Performance of Non-Meta-Learning Approaches

When emulating sparse input on LocalNet, a median TRE of 7.51 mm and a mean DSC of 0.76 are obtained with 5 input images. A median TRE of 6.26 mm and mean DSC of 0.79 are obtained with 10 input images. Fine-tuned Baseline model performance is significantly different than when providing 5 ($p < 0.01$) and 10 ($p = 0.04$) input images for TRE. No significant difference is observed for DSC. When using VoxelMorph, a median TRE of 7.36 mm and a mean DSC of 0.78 are obtained with 5 input images. A median TRE of 5.86 mm and mean DSC of 0.81 are obtained with 10 input images. Performance of the fine-tuned Baseline model is significantly different than when providing 5 input images ($p < 0.01$), but not 10 images, for TRE. No significant difference is observed for DSC.

Applying few-shot learning to these same models at inference, LocalNet obtains a median TRE of 7.64 mm and mean DSC of 0.76 are obtained with 5 input images. A median TRE of 7.23 mm and mean DSC of 0.73 are obtained with 10 input images. VoxelMorph obtains a median TRE of 7.30 mm and mean DSC of 0.79 are obtained with 5 input images. A median TRE of 5.81 mm and mean DSC of 0.81 are obtained with 10 input images. This suggests that few-shot learning has little effect when applied to conventionally trained registration networks, without the meta-trained network initialization. Using the Baseline, without few-shot learning, a higher median TRE of 4.57 mm and a lower mean DSC of 0.82 is obtained without detected significance, compared to the Baseline when using few-shot learning. Applying few-shot learning to an untrained model, where the weights are initialized randomly, results in a median TRE of 19.4 mm and a mean DSC of 0.76 for LocalNet, and a median TRE of 20.1 mm and a mean DSC of 0.77 for VoxelMorph.

Detailed results summarizing the TRE of the Baseline and the non-meta-learning-based methods are illustrated in Fig. 8. Example slices of input MR and TRUS image pairs and the registered MR images are provided in Fig. 10 for qualitative visual assessment of the registration results for each approach.

V   Discussion

This work presents a deep learning framework for interactive medical image registration using meta-learning. As illustrated in Fig. 9, the performance of our Baseline network for volume-to-sparse registration provides accuracy that is comparable from recent 3D-to-3D methods, while using a fraction of the data. Further, it yields significantly improved metrics compared to other tested volume-to-sparse methods and indicates that our method is not sensitive to meta-learning hyper-parameters, demonstrating flexibility and generalizability. This motivates use for other registration applications.

of importance for multimodal image registration, the lack of robust voxel-level similarity between image pairs necessitates the tested weakly-supervised registration algorithms, which require labelled structures in training, but not at inference. As we utilize few-shot learning in the meta-test phase, real-time prostate segmentations may be required on 2D TRUS images. High DSC and rapid inference times have been reported for this task [13, 51], as such, the need for segmentation must be considered, but should not be considered prohibitive to the real-time implementation of interactive registration in practice given that the addition of these additional segmentation inference steps would add, at most, several seconds to the total time required to compute the registration.

All employed volume-to-sparse methods require positional information for the TRUS images relative to a fixed reference. In practice, this may be obtained using positional, mechanical, or electromagnetic/optical tracking. Assessing our method's suitability for un-tracked TRUS images, however, is considered out of the scope of this work.

VI   Conclusion

This paper presents a novel interactive image registration approach, using an exemplar application of partial registration of MR to sparsely acquired intra-operative TRUS images. We obtain similar registration accuracies to state-of-the-art 3D image registration methods which require complete image volumes. our method significantly outperforms alternative methods when applied to the same challenging partial data problem. This work demonstrates the effectiveness and efficiency of our real-time interactive image registration method, which may be applied during intraoperative procedures, such as prostate biopsy.

Acknowledgments

References

[1]. Hu Y, et al. Weakly-supervised convolutional neural networks for multimodal image registration. Med Image Anal. 2018; 49: 1–13. [PubMed: 30007253]

[2]. Balakrishnan G, Zhoa A, Sabuncu M, Guttag J, Dalca A. voxelMorph: A learning framework for deformable medical image registration. IEEE Trans on Med Imaging. 2019; 38 (8) 1788–1800.

[3]. Chen J, He Y, Frey E, Li Y, Du Y. vitv-net: vision transformer for unsupervised volumetric medical image registration. arXiv preprint. 2021. arXiv:2104.06468

[4]. Yan, P; Xu, S; Rastinehad, AR; Wood, BJ. Adversarial image registration with application for MR and TRUS image fusion; International Conference on Med. Image Computing and Computer Assisted Interventions Mach. Learning in Med. Imaging Workshop; 2018. 197–204.

[5]. Xu, Z; , et al. Adversarial uni-and multi-modal stream networks for multimodal image registration; International Conference on Med Image Computing and Computer Assisted Interventions; 2020. 222–232.

[6]. de vos B, Berendsen F, Viergever M, Sokooti H, Staring M, Isgum I. A deep learning framework for unsupervised affine and deformable image registration. Med Image Anal. 2019; 52: 128–143. [PubMed: 30579222]

[7]. Baum, ZMC; Hu, Y; Barratt, DC. Multimodality biomedical image registration using free point transformer networks; International Conference on Med. Image Computing and Computer Assisted Interventions Advances in Simplifying Med. UltraSound Workshop; 2020. 116–125.

[8]. Baum ZMC, Hu Y, Barratt DC. Real-time multimodal image registration with partial intraoperative point-set data. Med Image Anal. 2021; 74 102231 [PubMed: 34583240]

[9]. Krebs, J; , et al. Robust non-rigid registration through agent-based action learning; International Conference on Med. Image Computing and Computer Assisted Interventions; 2017. 344–52.

[10]. Sun, S; , et al. Robust multimodal image registration using deep recurrent reinforcement learning; Asian Conference on Computer Vision; 2018. 511–526.

[11]. Hu J, et al. End-to-end multimodal image registration via reinforcement learning. Med Image Anal. 2021; 68 101878 [PubMed: 33197714]

[12]. Wang, J; Zhang, M. DeepFLASH: an efficient network for learning-based medical image registration; IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2020.

[13]. Fu Y, et al. Biomechanically constrained non-rigid MR-TRUS prostate registration using deep learning based 3D point cloud matching. Med Image Anal. 2021; 67 101845 [PubMed: 33129147]

[14]. Xu, Z; Niethammer, M. Deepatlas: joint semi-supervised learning of image registration and segmentation; International Conference on Med. Image Computing and Computer Assisted Interventions; 2019. 420–429.

[15]. Park H, et al. A meta-learning approach for medical image registration. arXiv preprint. 2021. arXiv:2104.10447

[16]. Hu J, et al. Towards accurate and robust multi-modal medical image registration using contrastive metric learning. IEEE Access. 2019; 7: 132816–132827.

[17]. Baum ZMC, Hu Y, Barratt DC. Meta-Registration: learning test-time optimization for single-pair image registration. arXiv preprint. 2022. arXiv:2207.10996

[18]. Zhu W, Huang Y, Zu D, Qian Z, Fan W, Xie X. Test-time training for deformable multi-scale image registration. arXiv preprint. 2021. arXiv:2103.13578

[19]. Hoopes, A; Hoffman, M; Fishcl, B; Guttag, J; Dalca, A. HyperMorph: Amortized Hyperparameter Learning for Image Registration; International Conference on Information Processing in Medical Imaging; 2021. 3–17.

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

[21]. Fu Y, Lei Y, Wang T, Curran W, Liu T, Yang X. Deep learning in medical image registration: a review. Physics in Medicine and Biology. 2020; 65 (20)

[22]. Boveiri H, Khayami R, Javidan R, Mehdizadeh A. Medical image registration using deep neural networks: a comprehensive review. Computers Electrical Engineering. 2020; 87

[23]. Amrehn, M; , et al. UI-Net: interactive artificial neural networks for iterative image segmentation based on a user model; Eurographics Workshop on Visual Computing for Biology and Medicine; 2017.

[24]. Wang G, et al. Interactive medical image segmentation using deep learning with image-specific fine-tuning. IEEE Trans on Med Imaging. 2018; 37 (7) 1562–1573.

[25]. Wang G, et al. DeepIGeoS: A deep interactive geodesic framework for medical image segmentation. IEEE Trans on Pattern Anal and Machine Intelligence. 2019; 41 (7) 1559–1572.

[26]. Boers TGW, et al. Interactive 3D U-net for the segmentation of the pancreas in computed tomography scans. Physics and Medicine in Biology. 2020; 65 (6) 065002

[27]. Fails, J; Olsen, D. Interactive machine learning; Proceedings of the 8th International Conference on Intelligent User Interfaces; 2003. 39–45.

[28]. Hill D, et al. Registration of MR and CT images for skull base surgery using point-like anatomical features. The British Journal of Radiology. 1991; 64 (767) 983–B106. [PubMed: 1742597]

[29]. Maurer C, Fitzpatrick J, Wang M, Galloway R, Maciunas R, Allen G. Registration of head volume images using implantable fiducial markers. IEEE Trans on Med Imaging. 1997; 16 (4) 447–462.

[30]. Vilalta R, Drissi Y. A perspective view and survey of meta-learning. Artificial Intelligence Review. 2002; 18: 77–95.

[31]. Hospedales T, Antoniou A, Micaelli P, Storkey A. Meta-learning in neural networks: a survey. arXiv preprint. 2020. arXiv:2004.05439

[32]. Finn, C; Abbeel, P; Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks; International Conference on Machine Learning; 2017. 1126–1135.

[33]. Nichol A, Achiam J, Schulman J. on first-order meta-learning algorithms. arXiv preprint. 2018. arXiv:1803.02999

[34]. Liu, Q; Dou, Q; Heng, P. Shape-aware meta-learning for generalizing prostate mri segmentation to unseen domains; International Conference on Med. Image Computing and Computer Assisted Interventions; 2020. 475–485.

[35]. Khandelwal, P; Yushkevich, P. Domain generalizer: a few-shot meta learning framework for domain generalization in medical imaging; International Conference on Med Image Computing and Computer Assisted Interventions Workshop on Domain Adaptation and Representation Transfer; 2020. 73–84.

[36]. Khadga R, et al. Few-shot segmentation of medical images based on meta-learning with implicit gradients. 2021. arXiv:2106.03223

[37]. Zhang P, Li J, Wang Y, Pan J. Domain adaptation for medical image segmentation: a meta-learning method. J of Imaging. 2021; 7 (2) 31. [PubMed: 34460630]

[38]. Budd S, Robinson E, Kainz B. A survey on active learning and human-in-the-loop deep learning for medical image analysis. Med Image Anal. 2021; 71 102062 [PubMed: 33901992]

[39]. Sinha A, Malo P, Deb K. A review on bilevel optimization: from classical to evolutionary approaches and applications. IEEE Trans on Evolutionary Computation. 2017; 22 (2) 276–95.

[40]. Kaplan I, Oldenburg N, Meskell P, Blake M, Church P, Holupka E. Real time MRI-ultrasound image guided stereotactic prostate biopsy. Magn Reason Imaging. 2002; 20 (3) 295–299.

[41]. Singh A, et al. Initial clinical experience with real-time transrectal ultrasonography-magnetic resonance imaging fusion-guided prostate biopsy. BJU Int. 2008; 101 (7) 841–845. [PubMed: 18070196]

[42]. Ukimura O, et al. Technique for a hybrid system of real-time transrectal ultrasound with preoperative magnetic resonance imaging in the guidance of targeted prostate biopsy. Int J Urol. 2010; 17 (10) 890–893. [PubMed: 20812938]

[43]. Miyagawa T, et al. Real-time virtual sonography for navigation during targeted prostate biopsy using magnetic resonance imaging data. Int J Urol. 2010; 17 (10) 855–860. [PubMed: 20807266]

[44]. Pinto P, et al. Magnetic resonance imaging/ultrasound fusion guided prostate biopsy improves cancer detection following transrectal ultrasound biopsy and correlates with multiparametric magnetic resonance imaging. J Urol. 2011; 186 (4) 1281–1285. [PubMed: 21849184]

[45]. Ukimura O, et al. 3-Dimensional elastic registration system of prostate biopsy location by real-time 3-dimensional transrectal ultrasound guidance with magnetic resonance/transrectal ultrasound image fusion. J Urol. 2012; 187 (3) 1080–1086. [PubMed: 22266005]

[46]. Sonn G, et al. Targeted biopsy in the detection of prostate cancer using an office based magnetic resonance ultrasound fusion device. J Urol. 2013; 189 (1) 86–91. [PubMed: 23158413]

[47]. Marks L, Young S, Natarajan S. MRI-ultrasound fusion for guidance of targeted prostate biopsy. Curr Opin in Urol. 2013; 23 (1) 43–50.

[48]. Reynier C, et al. MRI-TRUS data fusion for prostate brachytherapy. Preliminary results. Med Phys. 2004; 31 (6) 1568–1575. [PubMed: 15259661]

[49]. Dickinson L, et al. Image-directed, tissue-preserving focal therapy of prostate cancer: a feasibility study of a novel deformable magnetic resonance-ultrasound (MR-US) registration system. BJU International. 2013; 112 (5) 594–601. [PubMed: 23819525]

[50]. Ronneberger, O; Fischer, P; Brox, T. U-Net: convolutional networks for biomedical image segmentation; International Conference on Med. Image Computing and Computer Assisted Interventions; 2015. 234–241.

[51]. Ghavami N, et al. Integration of spatial information in convolutional neural networks for automatic segmentation of intraoperative transrectal ultrasound images. J of Med Imaging. 2018; 6 (1) 011003

[52]. Hamid S, et al. The SmartTarget biopsy trial: a prospective, within-person randomised, blinded trial comparing the accuracy of visual-registration and magnetic resonance imaging/ultrasound image-fusion targeted biopsies for prostate cancer risk stratification. Eur Urol. 2019; 75 (5) 733–740. [PubMed: 30527787]

[53]. Milletari, F; Navab, N; Ahmadi, S. V-Net: Fully convolutional neural networks for volumetric medical image segmentation; International Conf on 3D Vision; 2016. 565–571.

[54]. Rueckert D, Sonoda L, Hayes C, Hill D, Leach M, Hawkes DJ. Nonrigid registration using free-form deformations: application to breast mr images. IEEE Trans on Med Imaging. 1999; 18: 712–721.

[55]. Abadi M, et al. TensorFlow: Large-scale machine learning on heterogeneous systems. 2015.

[56]. Chollet F. Keras. 2015.

[57]. Fu Y. DeepReg: a deep learning toolkit for medical image registration. J of Open Source Software. 2020; 5 (55) 2705

[58]. Kingma, DP; Ba, J. Adam: a method for stochastic optimization; The International Conference for Learning Representations; 2015.

[59]. Ghavami N, et al. Automatic segmentation of prostate MRI using convolutional neural networks: Investigating the impact of network architecture on the accuracy of volume measurement and MRI-ultrasound registration. Med Image Anal. 2019; 58 101558 [PubMed: 31526965]

[60]. Hu, Y; , et al. A statistical motion model based on biomechanical simulations for data fusion during image-guided prostate interventions; International Conference on Med Image Computing and Computer-Assisted Interventions; 2008. 737–744.

[61]. Hahn DA, Daum V, Hornegger J. Automatic parameter selection for multimodal image registration. IEEE Trans on Med Imaging. 2010; 29 (5) 1140–1155.

[62]. Karnik VV, et al. Assessment of image registration accuracy in three-dimensional transrectal ultrasound guided prostate biopsy. Med Phys. 2010; 37 (2) 802–813. [PubMed: 20229890]

[63]. Hu Y, et al. Modelling prostate motion for data fusion during image-guided interventions. IEEE Trans on Med Imaging. 2011; 30 (11) 1887–1900.

[64]. Hu Y, et al. MR to ultrasound registration for image-guided prostate interventions. Med Image Anal. 2012; 16 (3) 687–703. [PubMed: 21216180]

[65]. De Silva T, et al. 2D-3D rigid registration to compensate for prostate motion during 3D TRUS-guided biopsy. Med Phys. 2012; 40 (2) 022904

[66]. Hu Y, Gibson E, Ahmed HU, Moore CM, Emberton M, Barratt DC. Population-based prediction of subject-specific prostate deformation for MR-to-ultrasound image registration. Med Image Anal. 2015; 26: 332–344. [PubMed: 26606458]

[67]. Sun Y, Yuan J, Qiu W, Rajchl M, Romagnoli C, Fenster A. Three-dimensional nonrigid MR-TRUS registration using dual optimization. IEEE Trans on Med Imaging. 2015; 34 (5) 1085–95.

[68]. Zettinig O, et al. Multimodal image-guided prostate fusion biopsy based on automatic deformable registration. International J of Computer Assisted Radiology and Surg. 2015; 10: 1997–2007.

[69]. Wang Y, et al. Towards personalized statistical deformable and hybrid point matching for robust MR-TRUS registration. IEEE Trans on Med Imaging. 2016; 35 (2) 589–604.

[70]. Onofrey JA, et al. Learning non-rigid deformations for robust, constrained point-based registration in image-guided MR-TRUS prostate intervention. Med Image Anal. 2017; 39: 29–43. [PubMed: 28431275]

[71]. van de Ven, W; Litjens, G; Barentsz, J; Hambrock, T; Huisman, H. Required accuracy of MR-US registration for prostate biopsies; Prostate Cancer Imaging: Image Analysis and Image-Guided Interventions; 2011. 92–99.
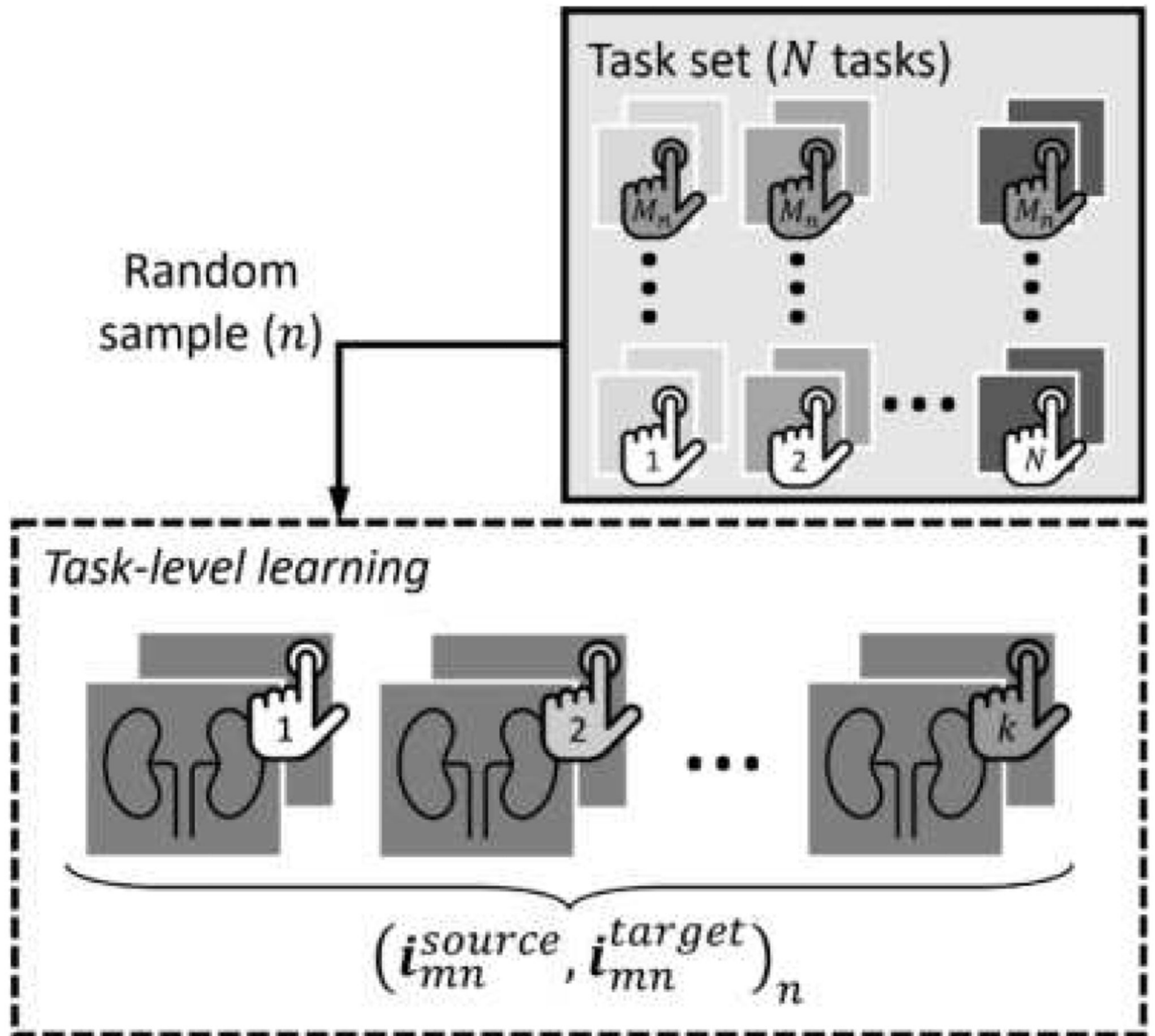
Fig. 1.
Schematic of one episode of task-level learning. For each of the *k* tasks in the sampled task set, the image pair is coupled with an associated annotation.
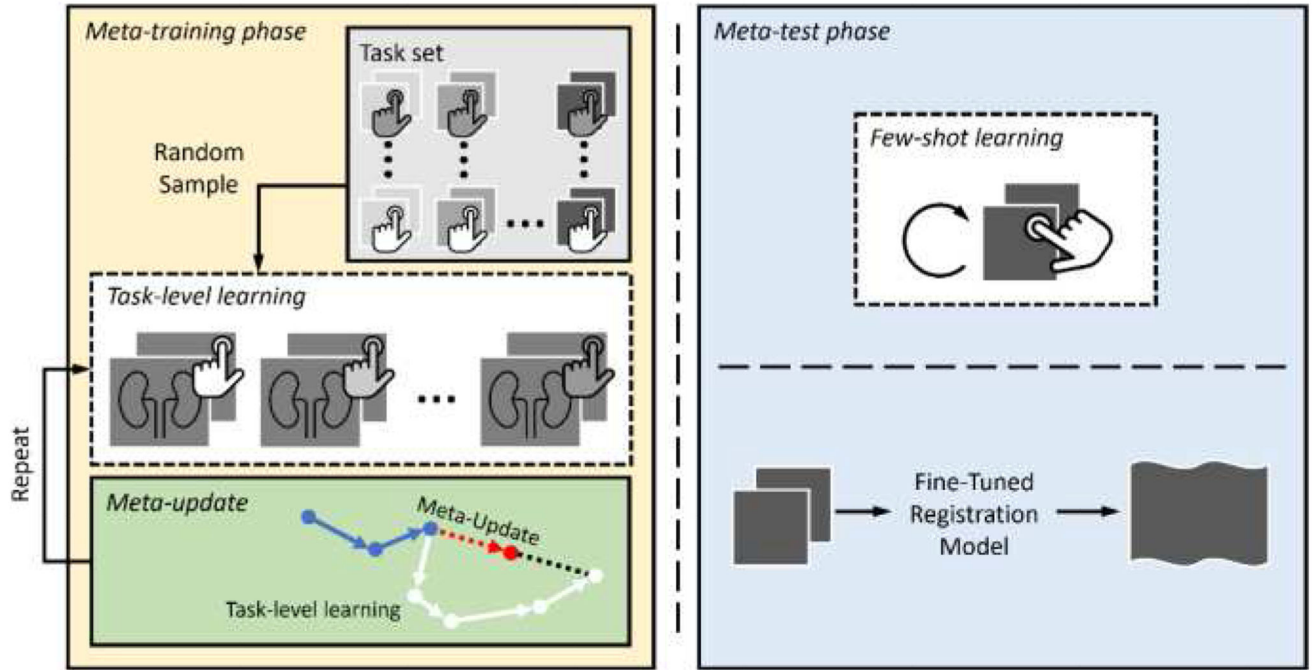
Fig. 2.
An interactive meta-learning medical image registration framework. A learning-based registration model is trained over multiple episodes in meta-training (left) to learn an initialization for adaptation at inference. In each task-level learning episode, a task is sampled to train the model. Then, the meta-update (red arrow) updates the model based on the direction (black dashed line) of the task-level learning gradients (white arrows), continued from previously learned gradients (blue arrows). Later, the model is fine-tuned in the meta-test phase (right) with few-shot learning, coupled with user-defined interactions.
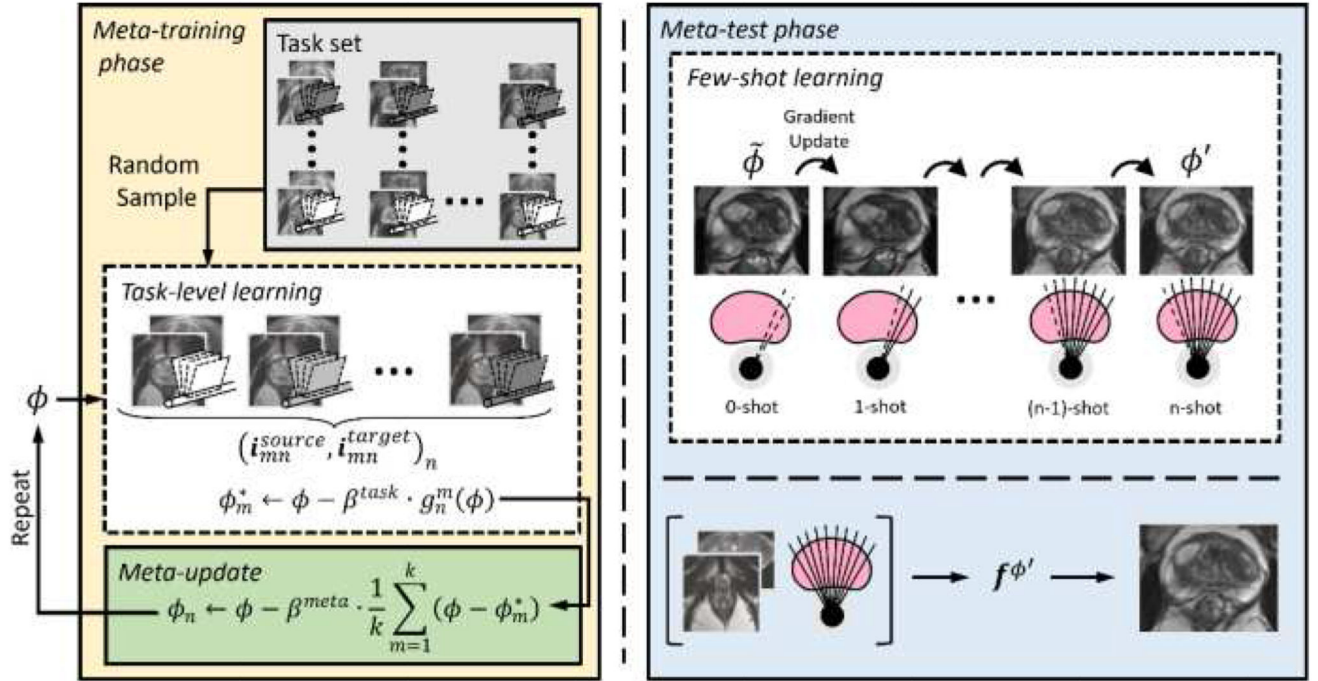
Fig. 3.
Proposed framework for interactive medical image registration with meta-learning, as applied to weakly-supervised volume-to-sparse prostate MR-TRUS registration. The learner is trained over multiple episodes in metatraining (left) to learn an initialization for adaptation at inference. In each tasklevel learning episode, a set of images, labels, and some number of frames F is sampled and trained on. After each episode, the meta-update updates the learner using the Reptile algorithm based on the task-level learning gradients. Once training is complete, the learner is optimized in the meta-test phase (right). Here, interactively-acquired data is coupled with few-shot learning to fine-tune a registration model in real-time as the TRUS image acquisition occurs.
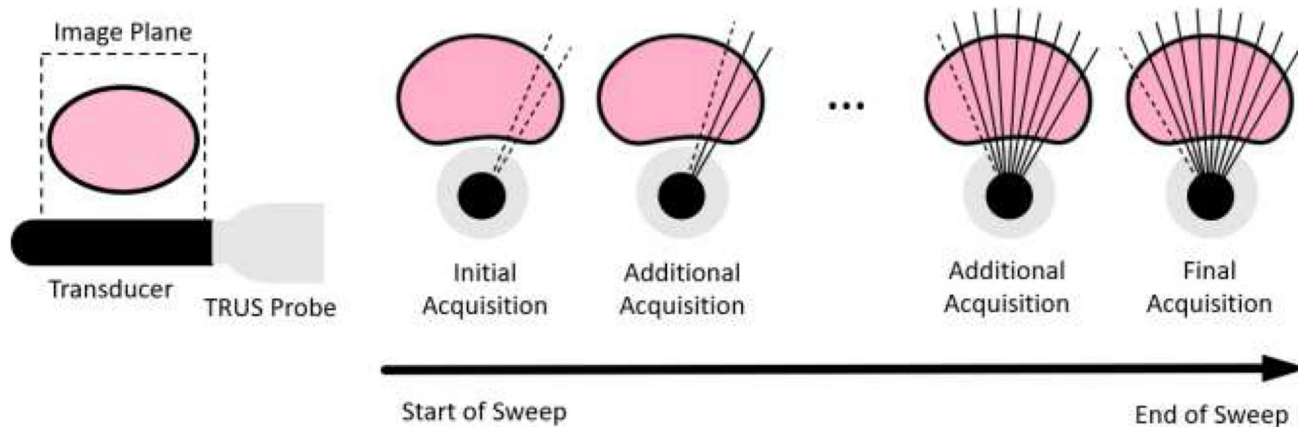
Fig. 4.
Illustration of TRUS images acquired in the presented clinical scenario. Acquired images (dashed lines) are captured in the sagittal plane (left) and shown with previously acquired images (solid lines) through one continuous 'sweep' of the prostate with the TRUS probe until full coverage is obtained.
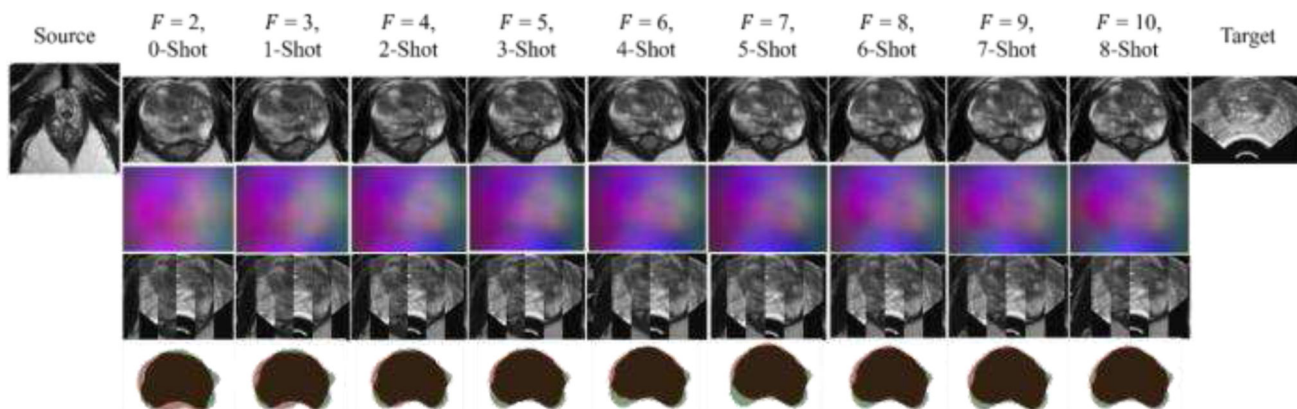
Fig. 5.
Example image slice from one test case. The left-most column contains an image slice from source MR volume. The right-most column contains the corresponding target TRUS image slice. Other columns present the warped source MR image, resulting DDF, alternating vertical slices of the warped MR and target TRUS image, and warped MR prostate gland contour (Red) overlaid on the target TRUS prostate gland contour (Green), using the Baseline at a given shot of training during few-shot learning, with *F* frames.
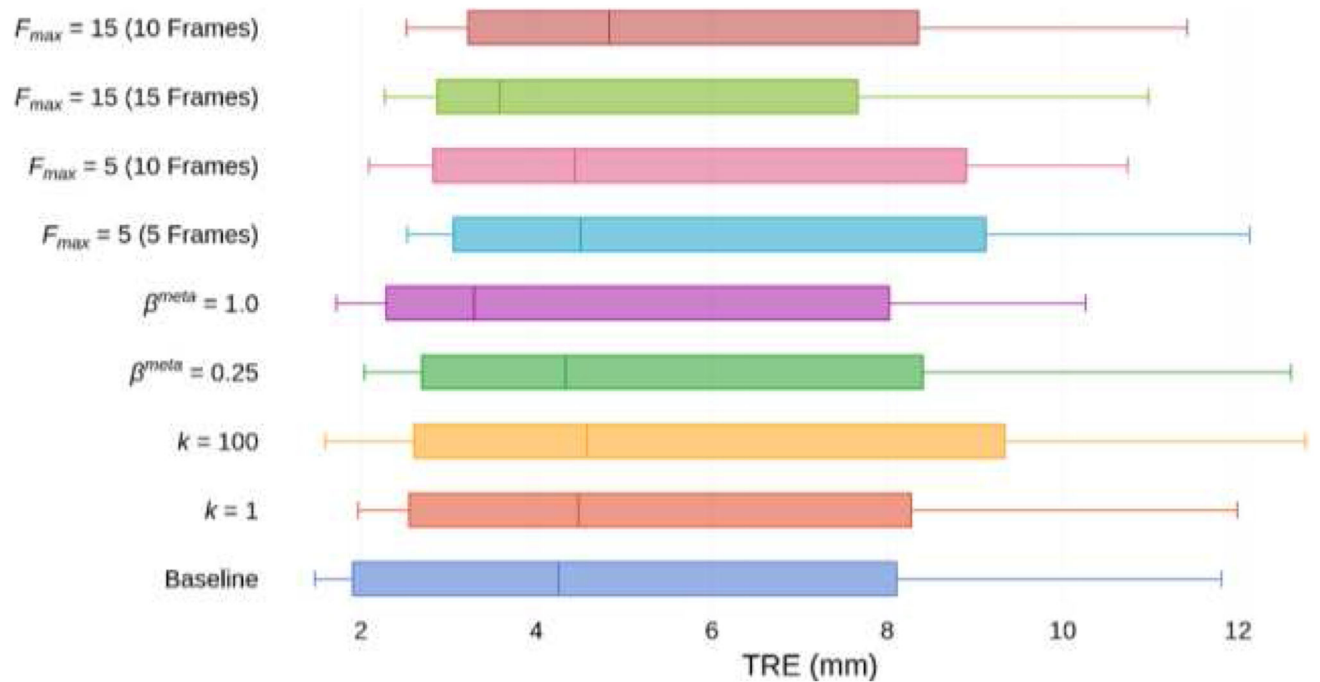
Fig. 6.
Tukey's boxplots of TRE for the Baseline and all variants in the MR-TRUS registration experiment. Whiskers indicate 10th and 90th percentiles. Results are presented for registrations with 10 frames unless otherwise indicated. For F_{max} variants, we additionally present results at 10 frames for direct comparison to the Baseline and other variants.
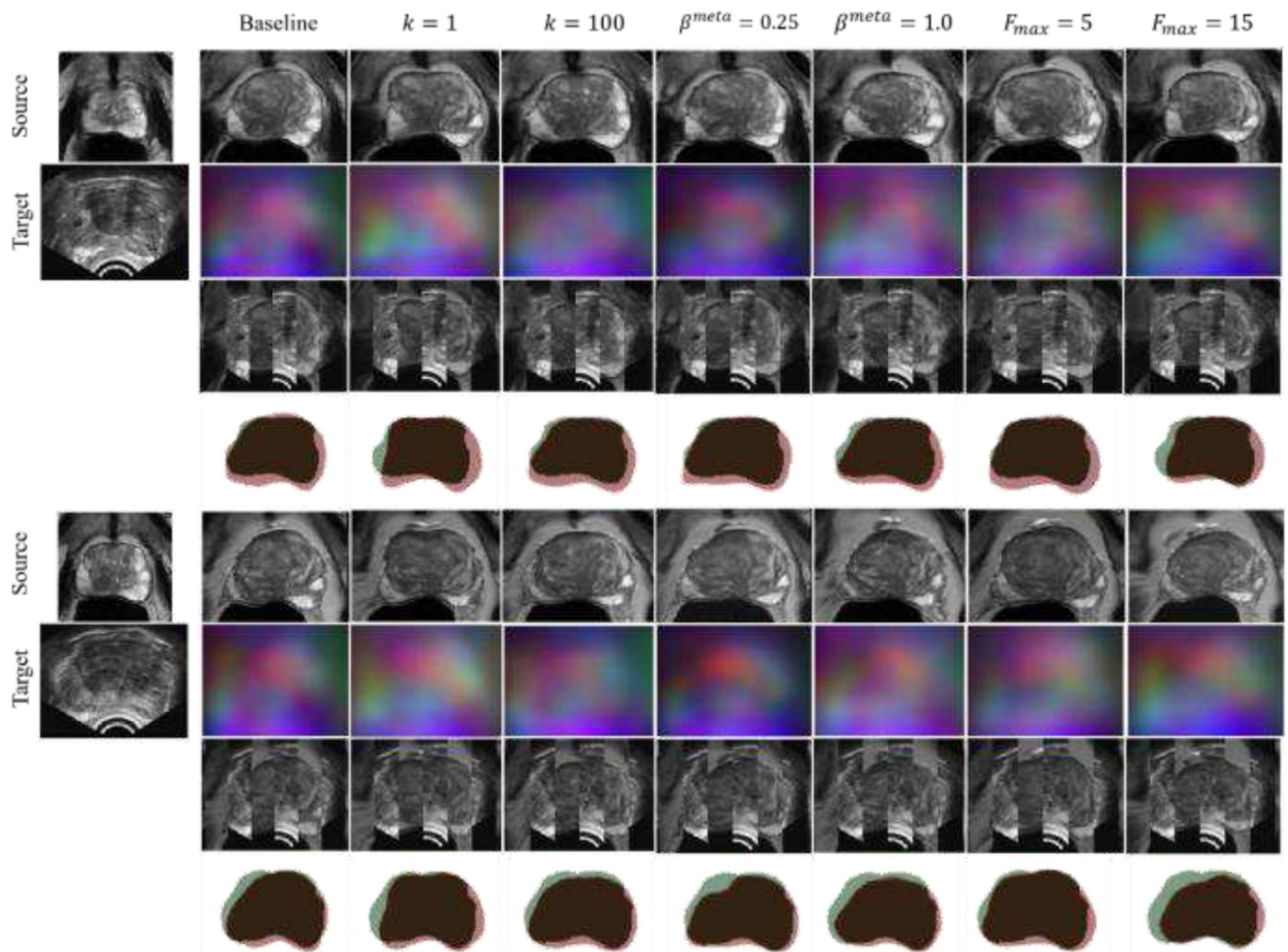
Fig. 7.
Example image slices from one test case. The left-most column contains image slices from source MR volume and corresponding target TRUS image slice. other columns present the warped source MR image, resulting DDF, alternating vertical slices of the warped MR and target TRUS image, and warped MR prostate gland contour (Red) overlaid on the target TRUS prostate gland contour (Green), using the above-labelled network.

Fig. 8.
Tukey's boxplots of TRE for Baseline, state-of-the-art, and all non-meta-learning methods in the MR-TRUS registration experiment. Whiskers indicate 10th and 90th percentiles. Baseline results presented for registrations with 10 frames, with input size indicated explicitly for all other methods.

Fig. 9.
Example image slices from two test cases. The left-most column contains image slices from source MR volume and the corresponding target TRUS image slice. other columns present the warped source MR image, resulting DDF, alternating vertical slices of the warped MR and target TRUS image, and warped MR prostate gland contour (Red) overlaid on the target TRUS prostate gland contour (Green), using the above-labelled network.
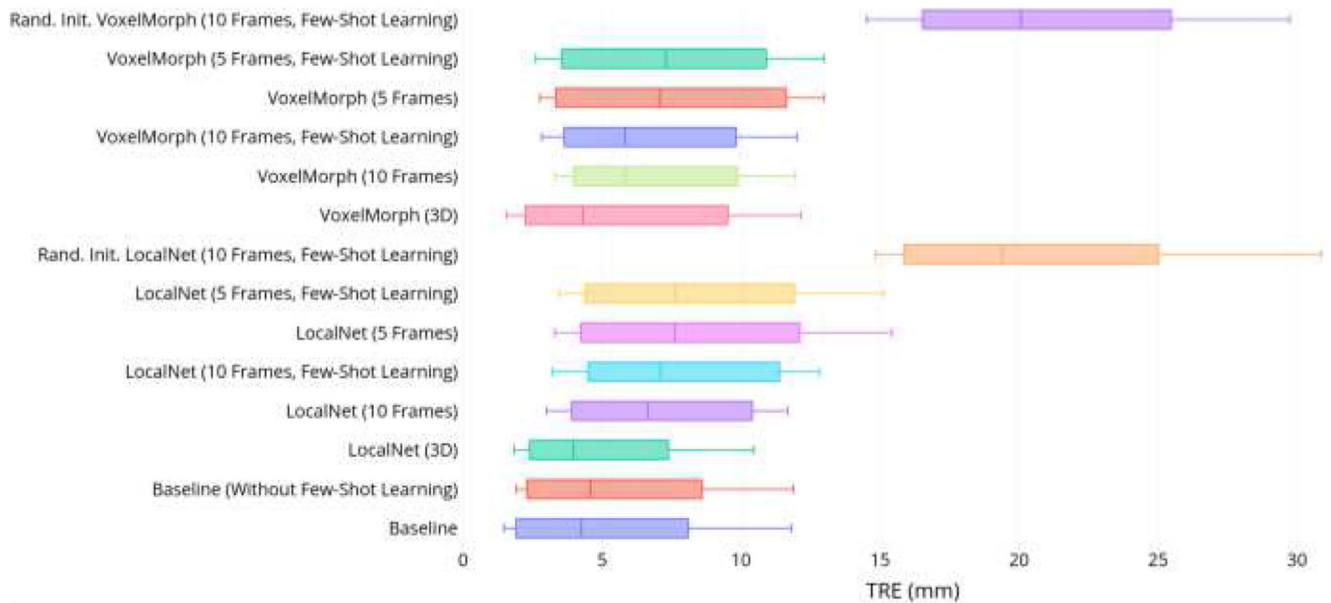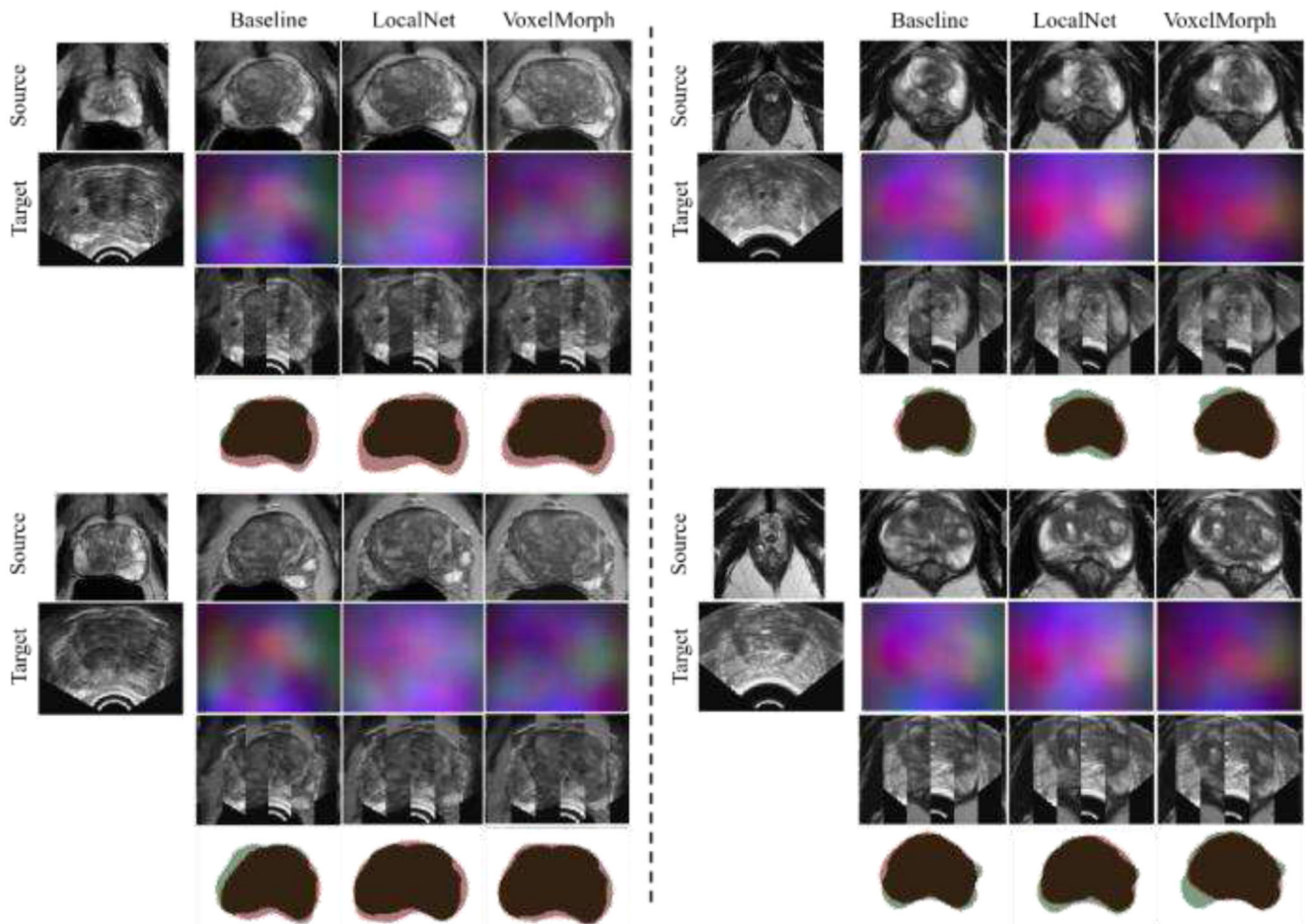
Fig. 10.
Example image slices from one test case. The left-most column contains image slices from source MR volume and corresponding target TRUS image slice. other columns present the warped source MR image, resulting DDF, alternating vertical slices of the warped MR and target TRUS image, and warped MR prostate gland contour (Red) overlaid on the target TRUS prostate gland contour (Green), using the above-labelled network.
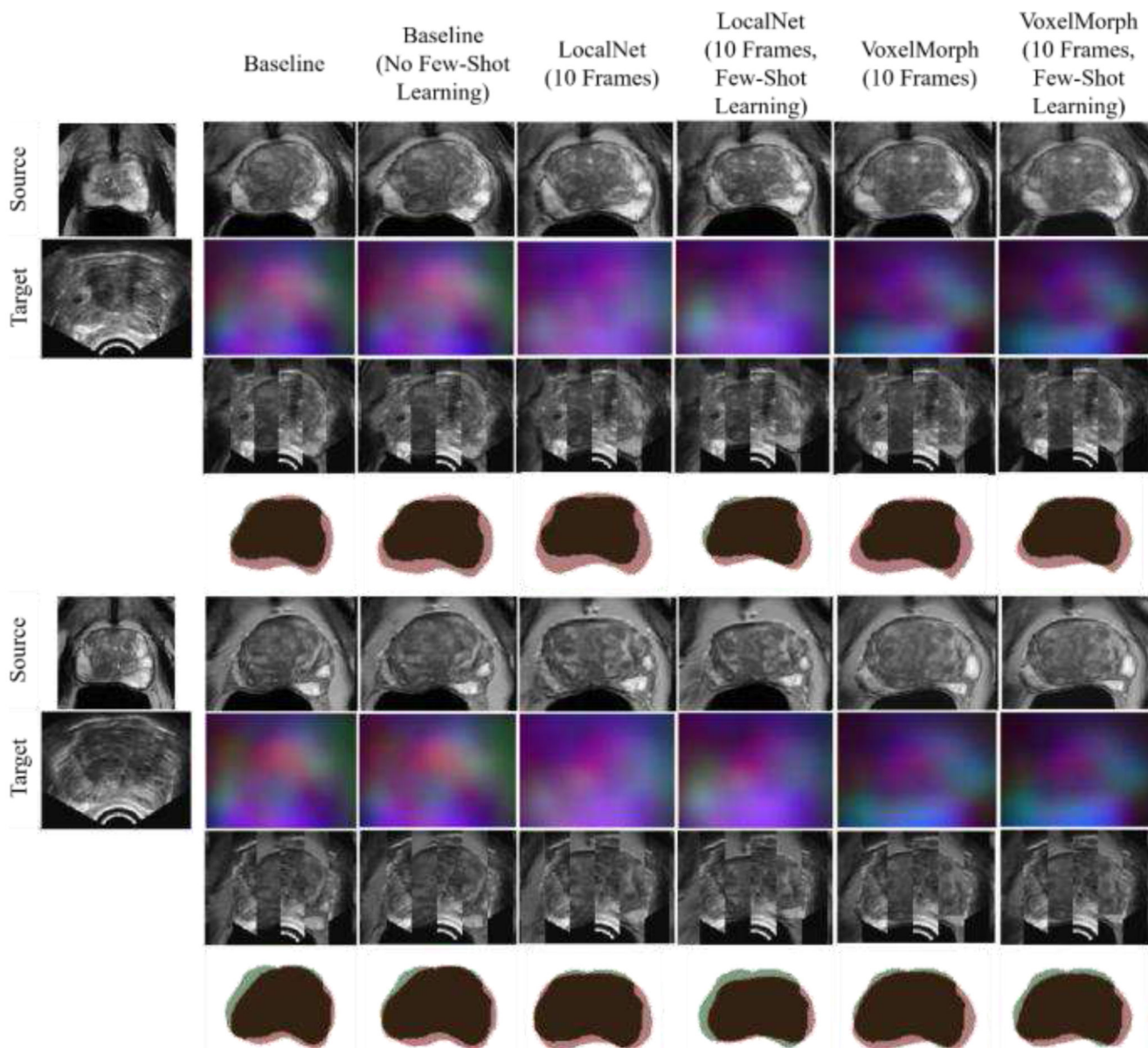
Table I

Summary TRE And DSC For The Baseline Network At Each Step Of Few-Shot Learning. Values Are Presented ± SD. TRE in MM.

F	Grad. Updates	Median TRE	Mean DSC
2	0	7.02 ± 4.08	0.77 ± 0.06
3	1	6.98 ± 3.98	0.79 ± 0.06
4	2	6.02 ± 4.17	0.81 ± 0.06
5	3	5.61 ± 4.11	0.82 ± 0.07
6	4	5.34 ± 4.16	0.82 ± 0.07
7	5	5.27 ± 4.08	0.83 ± 0.07
8	6	4.34 ± 4.12	0.84 ± 0.06
9	7	4.37 ± 4.13	0.84 ± 0.06
10	8	4.26 ± 4.19	0.85 ± 0.06

Table II

Summary TRE and DSC for the *k* = 1, 100 variants at each step of Few-Shot Learning. Values Are Presented ± SD. TRE in MM.

k	*F*	Grad. Updates	Median TRE	Mean DSC
	2	0	7.79 ± 3.75	0.76 ± 0.06
	3	1	7.55 ± 3.88	0.78 ± 0.06
	4	2	7.03 ± 3.97	0.79 ± 0.06
	5	3	5.79 ± 3.90	0.80 ± 0.06
1	6	4	5.59 ± 3.90	0.80 ±0.06
	7	5	5.49 ± 4.12	0.80 ± 0.06
	8	6	4.93 ± 4.07	0.81 ± 0.06
	9	7	4.43 ± 4.01	0.83 ± 0.06
	10	8	4.48 ± 3.96	0.83 ± 0.05
	2	0	7.83 ± 3.86	0.76 ± 0.06
	3	1	7.05 ± 4.00	0.78 ± 0.06
	4	2	6.49 ± 4.12	0.79 ± 0.06
	5	3	5.88 ± 4.20	0.81 ± 0.06
100	6	4	6.03 ± 4.32	0.82 ± 0.06
	7	5	5.64 ± 4.43	0.83 ± 0.06
	8	6	5.18 ± 4.44	0.84 ± 0.05
	9	7	4.78 ± 4.48	0.84 ± 0.05
	10	8	4.58 ± 4.48	0.85 ± 0.04

Table III

Summary TRE and DSC for the $\beta^{meta} = 0.25, 1.0$ variants at each step of few-Shot learning. values are presented $\pm$ SD. TRE in MM.

β^{meta}	F	Grad. Updates	Median TRE	Mean DSC
	2	0	7.06 ± 4.00	0.75 ± 0.07
	3	1	6.95 ± 3.95	0.76 ± 0.07
	4	2	6.44 ± 4.03	0.78 ± 0.06
	5	3	5.70 ± 3.88	0.80 ± 0.05
0.25	6	4	5.35 ± 4.00	0.80 ± 0.05
	7	5	5.26 ± 4.04	0.81 ± 0.05
	8	6	4.62 ± 4.06	0.82 ± 0.05
	9	7	4.31 ± 4.11	0.83 ± 0.05
	10	8	4.33 ± 4.11	0.84 ± 0.05
	2	0	7.54 ± 3.76	0.79 ± 0.05
	3	1	7.17 ± 3.77	0.81 ± 0.05
	4	2	6.62 ± 3.73	0.83 ± 0.05
	5	3	5.05 ± 3.64	0.84 ± 0.05
1.0	6	4	4.41 ± 3.64	0.84 ± 0.04
	7	5	4.22 ± 3.71	0.85 ± 0.04
	8	6	3.64 ± 3.84	0.86 ± 0.04
	9	7	3.22 ± 3.93	0.87 ± 0.04
	10	8	3.29 ± 3.97	0.87 ± 0.04

Table IV

Summary TRE and DSC for the F_{max} = 5, 15 variants at each step of few-Shot learning. values are presented ± SD. TRE in MM.

F_{max}	F	Grad. Updates	Median TRE	Mean DSC
5	2	0	6.49 ± 3.80	0.79 ± 0.06
	3	1	5.67 ± 3.94	0.82 ± 0.06
	4	2	4.58 ± 3.91	0.84 ± 0.05
	5	3	4.50 ± 3.93	0.85 ± 0.04
15	2	0	7.19 ± 4.01	0.76 ± 0.04
	3	1	6.82 ± 4.08	0.77 ± 0.05
	4	2	6.53 ± 4.19	0.78 ± 0.05
	5	3	6.33 ± 4.27	0.79 ± 0.05
	6	4	5.71 ± 4.16	0.80 ± 0.05
	7	5	5.51 ± 4.14	0.81 ± 0.06
	8	6	5.36 ± 4.10	0.81 ± 0.06
	9	7	5.44 ± 4.06	0.81 ± 0.06
	10	8	4.83 ± 4.03	0.81 ± 0.06
	11	9	4.37 ± 3.99	0.82 ± 0.06
	12	10	4.03 ± 3.96	0.83 ± 0.06
	13	11	3.86 ± 3.95	0.84 ± 0.06
	14	12	3.64 ± 4.01	0.84 ±0.06
	15	13	3.58 ± 4.00	0.84 ± 0.06