Cross-Modality Image Registration using a Training-Time Privileged Third Modality

Qianye Yang,David Atkinson,Yunguan Fu,Tom Syer,Wen Yan,Shonit Punwani,Matthew J. Clarkson,Dean C. Barratt,Tom Vercauteren,Yipeng Hu

***Abstract*— In this work, we consider the task of pairwise cross-modality image registration, which may benefit from exploiting additional images available only at training time from an additional modality that is different to those being registered. As an example, we focus on aligning intra-subject multiparametric Magnetic Resonance (mpMR) images, between T2-weighted (T2w) scans and diffusion-weighted scans with high b-value ($\text{DWI}_{\text{high-b}}$). For the application of localising tumours in mpMR images, diffusion scans with zero b-value ($\text{DWI}_{\text{b=0}}$) are considered easier to register to T2w due to the availability of corresponding features. We propose a learning from privileged modality algorithm, using a training-only imaging modality $\text{DWI}_{\text{b=0}}$, to support the challenging multi-modality registration problems. We present experimental results based on 369 sets of 3D multiparametric MRI images from 356 prostate cancer patients and report, with statistical significance, a lowered median target registration error of 4.34 mm, when registering the holdout $\text{DWI}_{\text{high-b}}$ and T2w image pairs, compared with that of 7.96 mm before registration. Results also show that the proposed learning-based registration networks enabled efficient registration with comparable or better accuracy, compared with a classical iterative algorithm and other tested learning-based methods with/without the additional modality. These compared algorithms also failed to produce any significantly improved alignment between $\text{DWI}_{\text{high-b}}$ and T2w in this challenging application.**

Q. Yang, M. J. Clarkson, D. C. Barratt and Y. Hu are with the UCL Centre for Medical Image Computing, Department of Medical Physics and Biomedical Engineering, University College London, London WC1E 6BT, U.K., and also with the Wellcome/EPSRC Centre for Interventional and Surgical Sciences University College London, London WC1E 6BT, U.K. (e-mail: qianye.yang.19@ucl.ac.uk; m.clarkson@ucl.ac.uk; d.barratt@ucl.ac.uk; yipeng.hu@ucl.ac.uk).

D. Atkinson is with the Centre for Medical Imaging, University College London, London W1W 7TS, U.K. (e-mail:d.atkinson@ucl.ac.uk)

Y. Fu is with the UCL Centre for Medical Image Computing, Department of Medical Physics and Biomedical Engineering, University College London, London WC1E 6BT, U.K., and also with the InstaDeep Co. (e-mail:yunguan.fu.18@ucl.ac.uk)

T. Syer is with the Centre for Medical Imaging, Division of Medicine, University College London, London WC1E 6BT, U.K. (e-mail:t.syer@ucl.ac.uk)

W. Yan is with City University of Hong Kong, department of Electrical Engineering, Hong Kong, China, and with the UCL Centre for Medical Image Computing, Department of Medical Physics and Biomedical Engineering, University College London, London WC1E 6BT, U.K. (e-mail:wenyan6-c@my.cityu.edu.hk)

S. Punwani is with the Centre for Medical Imaging, Division of Medicine, University College London, London WC1E 6BT, U.K. (e-mail:s.punwani@ucl.ac.uk)

T. Vercauteren is with the School of Biomedical Engineering & Imaging Sciences, King's College London, London, U.K. (e-mail:tom.vercauteren@kcl.ac.uk)

***Index Terms*— Medical image registration, Privileged learning, Deep learning, Multi-parametric MRI**

I. Introduction

MULTIPARAMETRIC Magnetic Resonance (mpMR) imaging is now recommended by international guidelines for the initial detection of prostate cancer for men with suspected disease [1]–[3]. Most subtypes of prostate cancer diagnosed on mpMR manifest themselves as low signal on T2-weighted MRIs, apparent diffusion coefficient (ADC) images, and high signals on high b-value diffusion MRI (DWI). As shown in both recent radiological and technical studies [4]–[6], mpMR images can lead to more accurate results on prostate cancer detection and staging, compared to only using single modality MR imaging [7]–[11], in particular, the T2-weighted and the diffusion-weighted scans have been recommended as two necessary modalities to include in any mpMR examination [12]. Jointly assessing mpMR scans can usually be expedited by accurate alignment [7], [13], especially when localisation of the pathological regions has become increasingly important for followup monitoring, diagnosis and treatment. In real-world clinical data, spatial differences exist between mpMR scans which are usually caused by patient movement during image acquisition, internal organ movement and distortions due to imperfect magnetic fields during image acquisition. However, registering multimodal mpMR images that are designed to provide complementary information is challenging. As these factors are difficult to decouple between different scans, scanner coordinates are often the only geometric reference after acquisition-time magnetic-field correction [14], [15].

For instance, echo-planar imaging using a high diffusion weighting (DWI_{high-b}) is considered sensitive for detecting prostate lesions in both peripheral and central gland, but it also can spatially differ from the T2-weighted (T2w) scans, the latter of which provides not only spatial reference for localising tumour of interest but also significant diagnostic value [16]. In many cases, it is visibly evident that registration between the two is required due to the coupled distortion and unknown patient/organ motion. The low signal-to-noise (SNR) in DWI_{high-b} and lack of spatially-corresponding features between the two challenges this registration task for both classical algorithms and recent deep-learning-based methods. Feature-based or semi-automated registration methods have been proposed for this task [17], [18]. In Section IV, we

provide quantitative evidence to demonstrate the need and the difficulty in direct registration between DWI_{high-b} and T2w scans.

DWI scans with low b-value (DWI_{low-b}), on the other hand, are less frequently used directly for diagnosis due to its diminished added clinical benefit with the presence of both DWI_{high-b} and T2w scans. For example, a time-critical imaging protocol for high-throughput application, e.g. [19], may suggest excluding ADC maps, which normally requires DWI_{low-b} to calculate. However, DWI_{low-b} scans are in general of higher SNR than DWI_{high-b} and has better tissue contrast, closer to T2w scans, as shown in Fig. 1. Meanwhile, DWI_{low-b} and DWI_{high-b} scans share similar distortion patterns and we have also observed a smaller spatial difference between DWI scans with different b-values, compared to the difference between DWI and T2w scans [20]. Quantitative results for supporting this observation are reported in Section IV. This is probably because DWI_{low-b} is less prone to artifacts and distortion. In this study we use $\text{DWI}_{b=0}$ as an example of DWI_{low-b} to facilitate the registration between T2w and DWI_{high-b} images. Although for some simplified imaging protocols, $\text{DWI}_{b=0}$ may be omitted, they can still be acquired readily for study purposes, such as neural network training. In addition, DWIs with b-values within a range of 0-100 sec/mm^2 can be used as alternatives for $\text{DWI}_{b=0}$ images, as suggested by [12]. In fact, in many existing mpMR imaging protocols for prostate cancer, $\text{DWI}_{b=0}$ data have been available for model training purposes. In this study, the $\text{DWI}_{b=0}$ data are used as the privileged information for training registration models, which are not required at the inference stage. The possibility and the potential of using other diffusion scans with low b-values may also be interesting under different clinical context, but will not be discussed further in this work.

This work has thus been motivated by a) the above-described clinical scenarios that can take advantage of DWI_{high-b} and T2w, bi-parametric imaging; and b) the hypothesized benefits of using DWI_{low-b} in aiding the cross-modality registration. We investigate deep learning algorithms that incorporate the $\text{DWI}_{b=0}$ scans in training registration networks that, once trained, take only DWI_{high-b} and T2w images as network input to register them - a case of learning using privileged information [21], [22]. In Section II, we describe a training strategy to facilitate the use of such a privileged third modality; then compare its performance to the alternative learning-based and non-learning methods; we report in Sect. IV experimental results using independent landmarks identified on holdout image pairs of registration interest in this work, i.e. DWI_{high-b} and T2w scans.

The aim of this work is to develop new learning methodologies and test their feasibility in improving the registration performance by incorporating extra imaging modality only in training. Learning-based registration methods have been proposed [23]–[33], especially, taking advantages of highly efficient deep registration networks during inference, with or without graphic processing units (GPUs). Learning-based registration, due to their being formulated as a machine learning task, can readily accommodate other observed latent variables to model additional information, such as a privileged third modality that is of interest in this study.

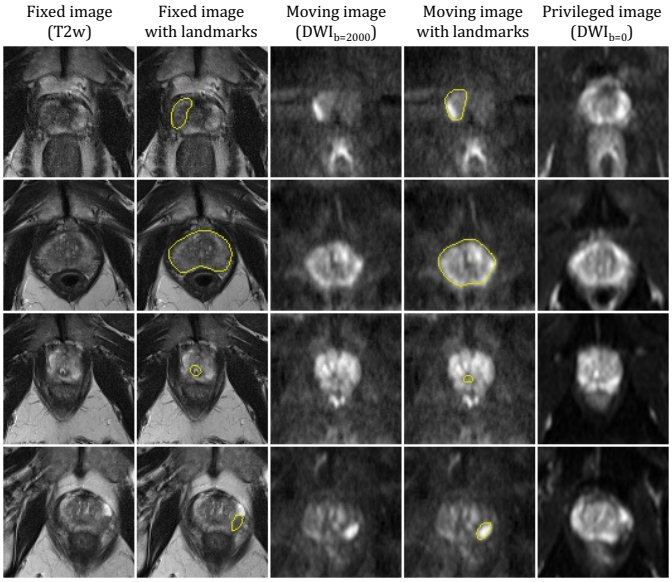

Fig. 1. Four example cases of the T2w, $\text{DWI}_{\boldsymbol{b=2000}}$, and $\text{DWI}_{\boldsymbol{b=0}}$ images used in this study. It shows that, compared with $\text{DWI}_{\boldsymbol{b=2000}}$, the $\text{DWI}_{\boldsymbol{b=0}}$ images in general have richer contrast between different structures and higher signal-to-noise ratio. The misalignment between $\text{DWI}_{\boldsymbol{b=2000}}$ and $\text{DWI}_{\boldsymbol{b=0}}$ is also smaller than that between $\text{DWI}_{\boldsymbol{b=2000}}$ and T2w images. The yellow contours indicate the annotated anatomical landmarks for validation purposes, including tumors, urethra, prostate glands and its zonal structures.

The work aims to show quantitative registration results on real clinical data and also highlight that the proposed methods utilising privileged images in this prostate cancer imaging application. However, we also envisage that this type of algorithms may be of wider applicability to other medical image registration problems. For example, longitudinal image registration when training data are available at more time points from retrospective subjects than those that need registration, or an interventional image registration task that with a missing reference image that is easier to register due to larger field-of-view or better image quality. The experiments presented in this work is focusing on unsupervised registration [34]–[38], due to the challenges in identifying substantial number of corresponding regions of interest (ROIs) labels for weak supervision [23]–[25]. Other approaches using deep feature for multi-modal registration [39], also requires anatomical annotations to learn registration-useful representations. However, when such labels are available, they may further aid cross-modality registration with the additional modality, but may be considered outside of the scope of this work.

We summarise the contributions in this work: 1) we propose to use additional images only in training to assist a challenging mpMR image registration task; 2) we propose and compare registration network training strategies using the privileged images; 3) we present experimental results using clinical imaging data from 356 prostate cancer patients; and 4) we provide quantitative results comparing the proposed methods with other learning-based registration methods with and without using the privileged third modality, in addition to a comparison to a non-learning algorithm, and report improved

or non-inferior registration performance from the proposed registration algorithm.

II. Methods

In this section, we describe a training strategy to train a registration network $f_\theta^{M\to F}(\mathbf{X}^M, \mathbf{X}^F)$ with network parameters θ and to input moving and fixed image pair $(\mathbf{X}^M, \mathbf{X}^F)$, given a set of training image trios $\{(x_n^M, x_n^F, x_n^P), n = 1, 2, ...N\}$, where N is the total number of MR studies. x_n^M, x_n^F and x_n^P are moving, fixed, and privileged images available during training, respectively. The registration network f_θ takes only two images as input and predicts the transformation, e.g. $\mu_n^{M\leftarrow F} = f_\theta^{M\to F}(x_n^M, x_n^F)$, where $\mu_n^{M\leftarrow F}$ is a dense displacement field (DDF) that can be used to obtain the warped moving image $x_n^M \circ \mu_n^{M\leftarrow F}$, where $\circ$ represents the resampling operation.

A. Learning from privileged supervision

First, we describe a formulation that enables training registration networks using the third modality that is not required during inference, as the network does not take the third image modality x_n^P as input, but rather is considered as a special type of supervision.

This is conceptually similar to weak supervision [23], where the segmentation labels of regions of interest have been proposed for weakly supervising registration networks. As illustrated in Fig. 2, the proposed registration network $f_\theta^{M\to F}$ accepts the same input image pairs, x_n^M and x_n^F, but is trained by maximising the image similarity between warped privileged images $x_n^P \circ \mu_n^{M\leftarrow F}$ and fixed images x_n^F, as opposed to the similarity measure used in an unsupervised approach between warped moving images $x_n^M \circ \mu_n^{M\leftarrow F}$ and the fixed images.

This formulation can also be considered as using the privileged-image-generated DDFs $\hat{\mu}_n^{P\leftarrow F}$ as the noisy labels for $\mu_n^{M\leftarrow F}$. It is important to highlight that the necessary condition for an unbiased estimate of $\mu_n^{M\leftarrow F}$ is $\mathbb{E}[\mu_n^{M\leftarrow F}] = \mathbb{E}[\hat{\mu}_n^{P\leftarrow F}]$, rather than the stringent sufficient condition $\mu_n^{M\leftarrow F} = \hat{\mu}_n^{P\leftarrow F}, \forall n$, since the method aims to provide a good estimate of the expected (average) of the deformation among all image pairs, rather than precise estimation for individual training image pairs.

B. Monte-Carlo resampling for bias reduction

Second, we develop a simple yet effective numerical resampling procedure to maximise the benefits of the privileged images during registration network training described in Section II-A.

To reduce the bias $\mathbb{E}[\mu_n^{M\leftarrow F} - \hat{\mu}_n^{P\leftarrow F}] = \mathbb{E}[\mu_n^{M\leftarrow F}] - \mathbb{E}[\hat{\mu}_n^{P\leftarrow F}]$, it is sufficient to spatially align the moving and the privileged images, x_n^M and x_n^P, which results in algorithms that are similar to the joint training (Section II-D.1). As argued earlier, aligning x_n^M and x_n^P is itself a multimodal image registration that can be challenging or unreliable in practice. We propose a simple Monte Carlo update step [40] to reduce the upper-bound of this bias, using affine-transformed privileged images $\tilde{x}_n^P$ before being warped by the network-generated DDFs $\mu_n^{M\leftarrow F}$. Here, $\tilde{x}_n^P = \text{argmax}_{x\in\{x_n^P \circ A_i\}}[MI(x_n^M, x)]$, where $\{A_i\}_{i=1}^I$ is a set of $I = 5$ randomly generated affine transformations. A proof of this bias upper-bound is provided in the following section. The "standard" unsupervised loss is used here between the warped privileged images and the fixed images, with a weighted deformation regularisation term $\mathcal{C}$.

$$J(\theta) = -\alpha \cdot MI(x_n^F, \tilde{x}_n^P \circ \mu_n^{M\leftarrow F}) + \beta \cdot \mathcal{C}(\mu_n^{M\leftarrow F}) \quad (1)$$

C. Effectiveness of surrogate supervision

Here, we provide an analysis to show that the Monte-Carlo procedure, described in Section II-A, is effective to reduce the bias between the warped privileged image and the ground-truth image without accessing to the ground-truth.

Denote a set of training image trios $(x^M, x^F, x^P) \in \mathcal{X}^3$, representing the moving, fixed and privileged images, respectively. $\mathcal{X}$ is the vector space for images. Given a dense displacement field $\mu^{M\leftarrow F}$, denote $T : \mathcal{X} \to \mathcal{X}$ that maps x^M to $T(x^M) = x^M \circ \mu^{M\leftarrow F}$ and $\circ$ represents resampling.

The registration task is thereby to minimize the difference between the warped moving image and the fixed image:

$$J = d(x^M \circ \mu^{M\leftarrow F}, x^F), \quad (2)$$

where $d : \mathcal{X}^2 \to [0, +\infty)$ is a metric defined on $\mathcal{X}$.

The proposed method uses an affine transformed privileged image $\tilde{x}^P$ as the surrogate of x^M, therefore it minimizes a different objective $J_{\text{surrogate}}$:

$$J_{\text{surrogate}} = d(\tilde{x}^P \circ \mu^{M\leftarrow F}, x^F), \quad (3)$$

Using triangulation inequality [41], the difference between the two objectives has the following upper-bound:

$$\begin{aligned} J_{\text{surrogate}} - J = \; & d(\tilde{x}^P \circ \mu^{M\leftarrow F}, x^F) - d(x^M \circ \mu^{M\leftarrow F}, x^F) \\ \leq \; & d(\tilde{x}^P \circ \mu^{M\leftarrow F}, x^M \circ \mu^{M\leftarrow F}) \\ = \; & d(T(\tilde{x}^P),\ T(x^M)) \end{aligned}$$

If T is Lipschitz continuous, then there exists a constant K such that

$$J_{\text{surrogate}} - J \leq \; Kd(\tilde{x}^P, x^M)$$

Thus, the surrogate objective $J_{\text{surrogate}}$ approximates the target objective J, when $d(\tilde{x}^P, x^M)$ is minimised. This justifies the proposed update step, in which multiple random affine transformations are applied on the privileged image and the closest one to the moving image is then selected. However, in this application, the adopted mutual information based distance $d(x_1, x_2) = -MI(x_1; x_2)$ is not strictly a *metric* on $\mathcal{X}$. Complications of the use of MI may warrant further investigation, but in practice, the above-described Monte-Carlo procedure almost always found a resampled images that lower the MI to the moving image with as few as 5-10 samples.

D. Alternative methods for utilising the third modality

Last but not least, it is important to test other, arguably simpler, approaches for training registration networks that can utilise the privileged information from the latent third modality. We describe two such alternatives as below, in which the third images are used only in training and are not required during inference.

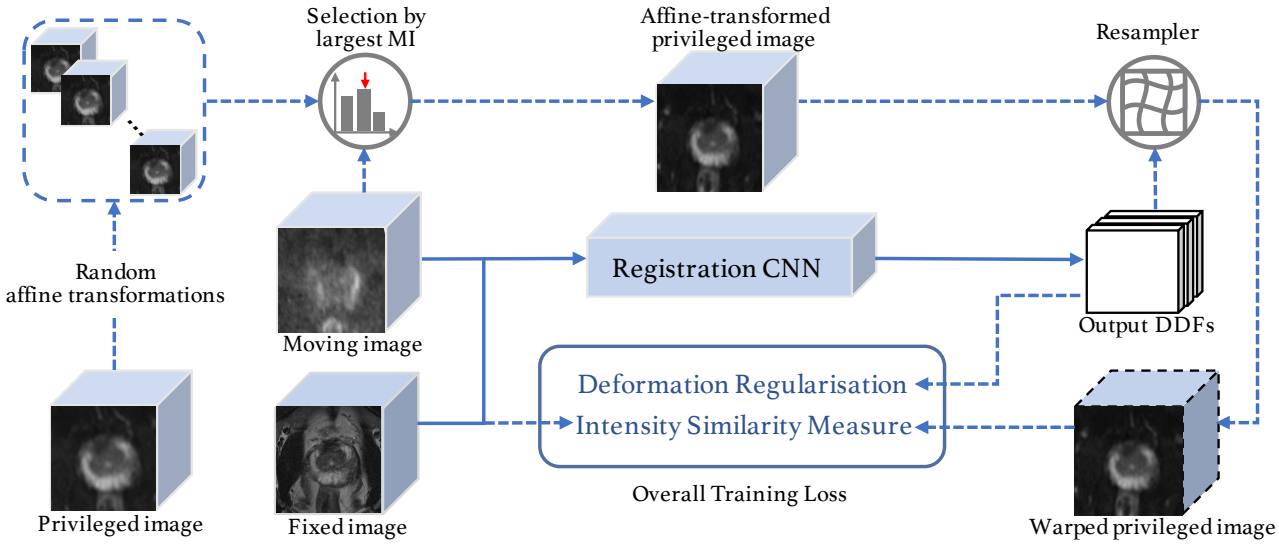

Fig. 2. The proposed privileged supervision for training a registration network in Section II-A. The dotted lines indicate the data flow only used in training.

1) Joint training: One approach to utilise the privileged images x_n^P is to estimate ground-truth DDFs $\hat{\mu}_n^{M\leftarrow F}$, by composing two intermediate transformation, $\hat{\mu}_n^{M\leftarrow F} = \hat{\mu}_n^{M\leftarrow P} \circ \hat{\mu}_n^{P\leftarrow F}$. While either classical algorithms or learning-based registration networks can be used to estimate $\hat{\mu}_n^{M\leftarrow P}$ and $\hat{\mu}_n^{P\leftarrow F}$ independent of training $f_\theta^{M\rightarrow F}$, we discuss two joint training algorithms.

The first algorithm trains three registration networks, $f_\theta^{M\rightarrow F}$, $f_{\phi_1}^{M\rightarrow P}$ and $f_{\phi_2}^{P\rightarrow F}$, to simultaneously estimate $\hat{\mu}_n^{M\leftarrow F}$, $\hat{\mu}_n^{M\leftarrow P}$ and $\hat{\mu}_n^{P\leftarrow F}$, respectively. A mean-square difference (MSD) can be used to minimise the difference between the network-predicted $\mu_n^{M\leftarrow F}$ and estimated ground-truth $\hat{\mu}_n^{M\leftarrow P} \circ \hat{\mu}_n^{P\leftarrow F}$. To train the latter two networks, an image dissimilarity loss, such as mutual information (MI), can be used between $(x_n^P, x_n^M \circ \mu_n^{M\leftarrow P})$ and between $(x_n^F, x_n^P \circ \mu_n^{P\leftarrow F})$.

With feature-rich moving images, a variant of the joint training can be implemented by maximising the image similarity between $x_n^M \circ \hat{\mu}_n^{M\leftarrow P} \circ \hat{\mu}_n^{P\leftarrow F}$ and $x_n^M \circ \hat{\mu}_n^{M\leftarrow F}$, without explicitly minimising the loss on the DDF difference. As the alignment between the two transformed moving images can be effectively measured by MSD. Results presented in this work are based on the following joint training loss in its general form:

$$J(\theta) = J^{M\rightarrow F} + J^{M\rightarrow P} + J^{P\rightarrow F} + MSD(x_n^M \circ \mu_n^{M\leftarrow F}, x_n^M \circ \mu_n^{M\leftarrow P} \circ \mu_n^{P\leftarrow F}) \quad (4)$$

where,

$$J^{A\rightarrow B}(\theta) = -\alpha \cdot MI(x_n^B, x_n^A \circ \mu_n^{A\leftarrow B}) + \beta \cdot \mathcal{C}(\mu_n^{A\leftarrow B}) \quad (5)$$

where, the image similarity and a deformation regularisation term $\mathcal{C}(\mu_n^{A\leftarrow B})$ are weighted by α and β, respectively, with shared values between the terms in Eq.1. L^2-norm on DDF gradient is used in this work: $\mathcal{C}(\mu_n^{A\leftarrow B}) = ||\nabla \mu_n^{A\leftarrow B}||_2$.

2) Mixed sampling: Rather than using the x_n^P as an intermediate imaging modality as in Section II-D.1, we consider to learn a shared registration network to predict the DDFs from both pairs of images, (x_n^M, x_n^F) and (x_n^P, x_n^F). An unsupervised registration network can be trained by sampling moving and fixed image pairs from the mixed set $\{(x_n^M, x_n^F)\} \cup \{(x_n^P, x_n^F)\}$. The loss function is given by:

$$J(\theta) = -\alpha(MI(x_n^F, x_n^M \circ \mu_n^{M\leftarrow F}) + MI(x_n^F, x_n^P \circ \mu_n^{P\leftarrow F})) + \beta \cdot \mathcal{C}(\mu_n^{(\theta)}) \quad (6)$$

where, hyper-parameters α and β specify the weights on the intensity dissimilarity and deformation regularisation, respectively.

This unsupervised approach utilises x_n^P during training, but still uses an image dissimilarity measure between transformed moving images and the fixed images. The lack of reliable and robust similarity measure between the two has not been addressed directly. While methods such as domain adaptation and semi-supervised learning make use of similarity between modalities x_n^P, x_n^M and x_n^F, such that the registration network predict reasonable DDFs without robust measure between x_n^M and x_n^F. These remain interesting future research, although it might also be further complicated by the distribution shift between training set $\{(x_n^M, x_n^F)\} \cup \{(x_n^P, x_n^F)\}$ and testing set $\{(x_n^M, x_n^F)\}$. Nevertheless, the described mixed sampling presents a reference performance from a single registration network, with quantitative results reported in Section IV.

E. Evaluation

All the registration networks described in Section II aim to register the moving and fixed images, without using the privileged images at test time. The anatomical and pathological landmarks are manually identified including patient-specific tumors, urethra, prostate glands and zonal structures,

and labelled volumetrically as binary masks. The root-mean-square distance was computed as target registration errors (TREs), between the centers of the mass of the corresponding landmarks independently defined on the fixed and network-warped moving images on holdout data set.

Experiment details are described in Section III, in which the intra-subject DWI_{high-b}, T2w and $\text{DWI}_{b=0}$ are used as the moving, fixed and privileged images, respectively. When appropriate, MI is also reported, which may be less relevant to the quality of registration, compared to the TREs on independent landmarks, but provides a quantitative measure how the optimisation during training and generalisation during inference perform. When comparisons are made, p-values are reported from paired two-sided t-tests at a significance level of $\alpha = 0.05$.

III. Experiments

A. Data and preprocessing

369 mpMR image studies were acquired from 356 prostate cancer patients at University College London Hospitals. One or two studies of mpMR images were available for each patient. The mpMRIs were acquired from 1.5T SIEMENS MR scanners, with original voxel resolution of $0.625 \times 0.625 \times 1.0\ mm^3$ and $1.0 \times 1.0 \times 5.0\ mm^3$ for the T2w and DWIs, respectively. All the image volumes were resampled to voxel dimension of $1.0 \times 1.0 \times 1.0\ mm^3$ and got a center-cropped volume of $104 \times 104 \times 92$ voxels, with a normalised intensity range of $[0, 1]$. In order to validate the registration performance, 35 pairs of mpMRIs from 35 patients with obviously large initial misalignment were selected as the holdout set. The rest of the data set was split into 302 and 32 MRI studies, from 289 and 32 patients, for training and validation sets, respectively. Up to three pairs of landmarks were identified for each study and a total of 50 pairs of landmarks were labelled for the holdout set. The annotation of the landmarks was performed by two biomedical imaging researchers, who have completed a BAUS-accredited MRI course on prostate cancer. The landmarks were labelled by one observer before being checked by the other. To investigate the intra-observer variance, the holdout test set was annotated again, two-months after, and blind to, the first annotation. An intra-observer landmark localization error of 1.08 ± 0.54mm is achieved.

Two additional data set were used for external validation. Data Set A was acquired from a different hospital, with an approved Institutional Review Board protocol designed at the University College London Hospital (UCLH). The original voxel resolution was $0.625 \times 0.625 \times 1.0\ mm^3$ and $2.0 \times 2.0 \times 5.0\ mm^3$ for the T2w and diffusion-weighted images, respectively. Data Set B was obtained from the Cancer Imaging Archive [42], with the original voxel resolutions of $0.27 \times 0.27 \times 3.0\ mm^3$ and $0.7 \times 0.7 \times 4.0\ mm^3$, for the T2w and DWIs, respectively. The mpMRIs were acquired from a 3T GE MR scanner, with endorectal coil. In this public data sets, we only have access to the DWI_{high-b} with b=1400 sec/mm^2 in this data set. The same image prepossessing and the landmark annotation were used, as on the UCLH data set. A total of 30 patients with 42 pairs of landmarks and 20 patients with 21 pairs of landmarks are used in the Data Sets A and B, respectively, for assessing the registration performance on external data sets.

The MI was adopted for the similarity measure, suggested by a previous study [43]. The MI was also used as the validation metric for hyperparameter search, specifying the weightings of loss terms α and β to 0.5 and 1×10^3, respectively. Fine-tuning of these hyper-parameters by, for example, systematic or automated hyperparameter search should benefit and is a subject of future studies.

B. Network training

An encoder-decoder registration network [23] was used for DDF prediction in all the models in this work. Random affine transformations were added to the input of the network, both for data augmentation and the Monte-Carlo resampling. The method of the random affine transformation is adapted from the open-source code DeepReg [44], which is generated by randomly resampling the image corners from a uniform distribution, in order to keep minimal sampling outside the original image. The image warping method is implemented using a standard grid sampling method with trilinear interpolation and zero-padding [44]. The network training was implemented with PyTorch [45] and made open-source https://github.com/QianyeYang/mpmrireg. The Adam optimizer with an initial learning rate of 10^{-5} was used. The "privileged supervision" networks described in Section 2 were trained on Nvidia Tesla V100 GPUs with a minibatch of 4 sets of image data, each containing a trio of intra-subject $\text{DWI}_{b=2000}$, T2w and $\text{DWI}_{b=0}$ images. Each network was run for 600,000 iterations, approximately 50 hours. All registration networks were trained using the same training strategy unless otherwise specified.

C. Other learning-based registration

The "joint training" and the "mixed sampling" networks that implemented methods described in Section II-D.1 and Section II-D.2, respectively, were also trained to test these alternative approaches to incorporate the third image modality. Like in evaluating the privileged supervision network, these two networks were trained with the trio of intra-subjective images, but only took T2w and $\text{DWI}_{b=2000}$ images as input, during test stage using the holdout set.

In addition, a learning-based registration methods were compared for directly aligning T2w and DWI scans with b values being 2000, $\text{DWI}_{b=2000}$. The registration network was trained using the unsupervised learning algorithm, similar to the one used in Section II-D.2, but with only T2w and $\text{DWI}_{b=2000}$ sampled in training without $\text{DWI}_{b=0}$. This is referred to as the "Direct" method. For further understanding the role of $\text{DWI}_{b=0}$ scans and the potential benefits in adding the bias-reducing Monte-Carlo resampling, described in Section. II-B, another unsupervised registration network was trained using only T2w and $\text{DWI}_{b=0}$ in training without $\text{DWI}_{b=2000}$. These two registration networks were both tested on registering the T2w and $\text{DWI}_{b=2000}$ images on the holdout set.

Weakly supervised registration [23], [24] methods have also been proved to be effective for the multi-modal registration problems. However, the gland masks of the $\text{DWI}_{b=2000}$ in this study are not available and arguably much more difficult to annotate accurately. For example, rectal gas is known for generating magnetic distortion around posterior regions of the prostate glands, which complicates in determining capsule boundaries; and DWI_{high-b} has high sensitivity in certain types of pathology but lacks contrast in gland itself. This study is to investigate how much improvement is feasible for using unsupervised learning methods with unlabelled image data, which are in practice more feasible to obtain.

D. Non-learning registration

Learning-based registration methods in general provide superior efficiency, compared with the alternative classical registration algorithms based on iterative optimisation, especially for large 3D volumetric medical images [46]. However, it is useful to report the performance using the classical methods which have been developed for registering multimodal image registration in similar applications [47]–[51], for last two decades.

For its fast GPU implementation, the NiftyReg package was used to compare a non-rigid B-spline based free-form deformation algorithm with the above learning-based methods, by directly registering the T2w and $\text{DWI}_{b=2000}$. The NiftyReg Package was used as an example of non-learning algorithms, with normalised mutual information and other parameter values followed a previous prostate MR registration study [52]. In our experiment, MI was used as the similarity measure for comparison purpose with a bending energy weight of 0.005, among other default configurations. These parameters may not be directly comparable to those used in the learning-based algorithms due to difference between the pairwise optimisation and stochastic-gradient-based learning process, in addition to varying implementation choices. The MI values before and after respective algorithms are reported in Sect IV.

The aim for reporting results from the non-learning registration is not intended to compare their registration accuracy, as substantially more comprehensive experiments shall be required to draw a convincing conclusion that may also be dependent on the application and the experimental data used. Rather, this provides a reference of the registration performance with a readily-available, non-learning registration algorithm that does not require the third modality in this specific prostate cancer application.

IV. Results

A. Registration performance on holdout set

The TRE and MI results on the holdout set are summarised in Table. I. The proposed privileged supervision increased the median MI from 0.06 to 0.20 and lowered the median TRE from 7.96 mm to 4.34 mm, improved from those before registration, both with statistical significance (p-values$<$0.001). All the other tested methods showed improved TREs in this application with statistical significance as well (NiftyReg:p-value=0.03, others:p-value$<$0.001). The Jacobian determinants of each predicted DDF was also computed and, from all proposed methods in this study, no negative values were found.

Results from two groups of cases are also summarised in Table. II, with those that have the largest initial misalignment, measured by landmark distance before registration, and the most improvement by registration, measured by TRE. It is noteworthy that the latter group was selected by the improvement after the registration results were obtained, which were not be available prior to registration, therefore only provides a selective reference for measuring the potential contribution. Together with the cases with largest misalignemnt, these two subgroup results represent the comparison on those cases that need the registration the most.

The registration results have been visually assessed and examples are provided in Fig. 3. In the first three cases, the morphology and the location of the tumor are more consistent with the fixed T2w images after registration. The fourth and the fifth cases are challenging cases with larger initial misalignment. Case 4 shows a visible improvement in morphology of the central gland. The registration compensated the distorted region near the rectum. Meanwhile, the increased hyper-intensity area on the top of the warped privileged image indicates an improved alignment of the bladder. In case 5, although the registration could be further improved, the registered location of the prostate gland and the tumor indicate the predicted contributed to visibly reduce the misalignment.

Figure 4 provides further examples that demonstrate the potential benefits from the third modality during training, $\text{DWI}_{b=0}$ in this case. Case 1 is an example with minor misalignment. A suspected tumour was found in the central gland in the zoomed-in ROI, with the contoured tumor being aligned visually better after registration from all methods. Case 2 presents a relatively severe misalignment of the whole prostate gland, with the gland center being aligned after registration. Case 3 shows an example of a well-aligned local area of the urethra using the proposed method. Case 4 demonstrates a highly severe misalignment, for both of the prostate gland and the contoured tumor. From Case 1 to 4, it is visually recognisable that the Privileged method outperforms the others tested. For case 4, although a minor misalignment still exists after registration, the Privileged method transformed the tumor closer to the target location and shape, with respect to the reference bounding-box ROI, while the others are absent to varying degrees. Case 5 shows an example of the distortion in the gland posterior region, which was reduced by the registration.

To further investigate the performance, a set of Bland-Altman plots, from each proposed method, are provided to show the differences in MI and TREs before and after registration in Fig. 5. Each point represents a pair of landmarks in the holdout set, where the x axis represents the TRE before registration and the y axis represents the difference in TRE after registration. For each of the proposed method, improvements are observed both on MIs and TREs. The Privileged method outperforms the others with improvements of 3.88mm and 0.14, in TREs and MIs, respectively.

For the inference time, our proposed method got 0.12s while the NiftyReg got 10.19s for registering each pair of 3D image,

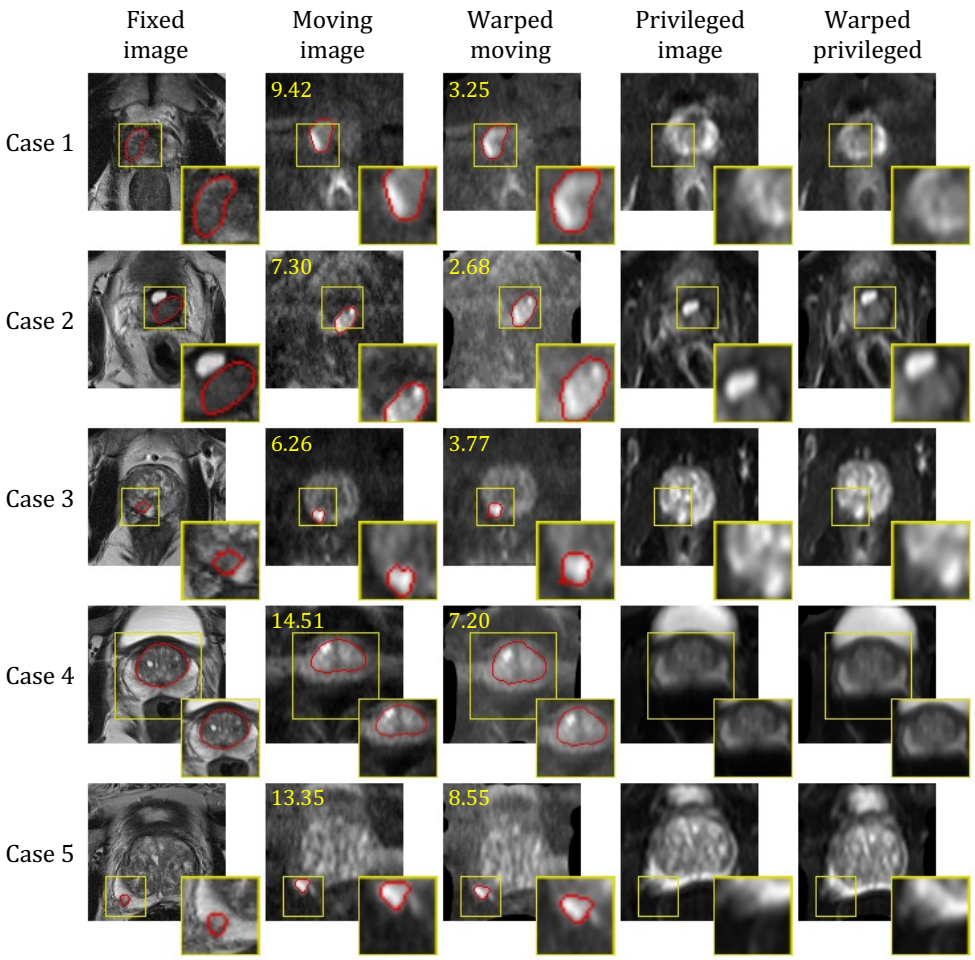

Fig. 3. Five example registered cases using the proposed privileged supervision method with yellow bounding boxes, zoomed in at the same reference positions, and the red landmark contours that indicate ROIs for assessing registration. Annotation on privileged images, in the last two columns, are challenging and not available during test time, with images being presented for comparison. For each case, the bounding boxes are placed at the same reference locations in order to assessing the registration. The the TREs (mm) before and after registration are provided in yellow, on the moving and the warped moving images, respectively.

both with GPU acceleration.

B. The need for registering T2w and DWI$_{b=2000}$

The MI and TREs on the test data set are computed to indicate the original difference between the two images without registration. All registration methods have made positive contributions to align images based on the increased MI values. All of the tested registration methods reduced TREs. This is an indication that registration in general would help align the T2w and DWI$_{b=2000}$ scans in this application.

Table II provide results from the 10% and 20% cases with the largest initial misalignment and 10% and 20% cases with the most improvement observed after registration. The results from both these subgroups showed a larger initial misalignment and arguably more substantial improvements from the registration. For example, for 20% cases with largest initial misalignment, the proposed privileged supervision network improved the mean TREs from 12.93 mm to 6.77 mm.

We also report a set of selective results *only* for inspecting the extent of the registration error, from the 10% and 20% cases with the most improvement by registration, measured by TRE. However, identifying either of these scans that need registration the most remains an interesting open research question, as it may be that, based on the clinical data set used in this work, the proposed method would be of increased clinical value when applied to this subset of patient studies.

C. Comparison to other privileged learning methods

From both Table I and Table II. The proposed methods outperformed the alternative joint training and mixed sampling methods, in terms of TREs. The advantage is both consistent and statistically significant (p-values$<$0.001). This set of results demonstrate the effectiveness of the proposed privileged learning method to align the T2w and DWI$_{b=2000}$, in this application. These results concludes that adding images from a different modality to training may not be trivial and, without appropriate adaptation, may reduce the registration performance. In addition, the same network was trained with the Privileged method with 10 random affine transformations in the Monte-Carlo resampling (i.e., $I = 10$), described in

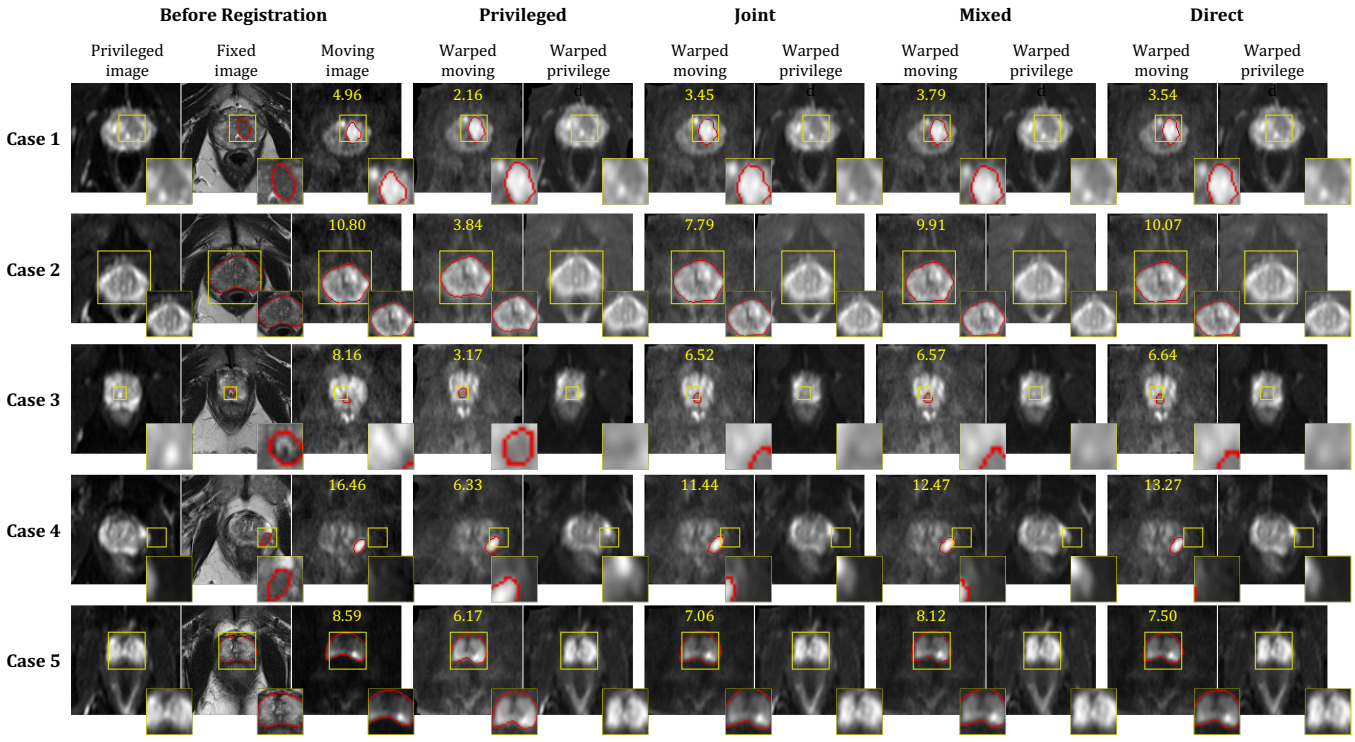

Fig. 4. Five example registered cases from different deep learning methods. The triplet in the first column indicates the privileged images, fixed images, and moving images, from left to right. The following four columns contain the results of the registered moving and privileged images, from different registration methods. The TREs(mm) from each methods are indicated in yellow, before and after registration, on the top of the moving images and the warped moving images, respectively. The yellow bounding boxes and the red validation anatomical landmarks together with their zoomed-in versions indicate ROIs for assessing registration. For each case, the bounding boxes are palced at the same reference spatial locations.

TABLE I

HOLDOUT SET PERFORMANCE FOR REGISTERING T2W-DWI$_{b=2000}$

Methods	Training input	MI: Mean, Median, 90^{th} Pctl.	TRE: Mean, Median, 90^{th} Pctl. (mm)
w/o registration	-	0.063±0.034, 0.060, 0.096	8.331±3.016, 7.964, 11.874
NiftyReg	non-learning method	0.076±0.051, 0.068, 0.099	8.128±2.935, 7.862, 11.121
Direct	T2w-DWI$_{b=2000}$	0.196±0.062, 0.200, 0.289	7.471±3.067, 6.895, 11.969
Mixed	T2w-DWI$_{b=0,b=2000}$	0.190±0.063, 0.200, 0.281	7.386±2.801, 7.132, 12.110
Joint	T2w-DWI$_{b=0,b=2000}$	0.178±0.070, 0.191, 0.269	6.405±2.356, 6.138, 9.730
Privileged Sup.	T2w-DWI$_{b=0,b=2000}$	**0.201±0.055, 0.204, 0.277**	**4.456±2.055, 4.339, 6.860**

TABLE II

HOLDOUT SET PERFORMANCE IN 10% AND 20% SAMPLES WITH THE LARGEST INITIAL MISALIGNMENT (PRE-REGISTRATION INDEPENDENT STRATIFICATION) AND THOSE WITH THE MOST IMPROVEMENT (SELECTIVE RESULTS FOR REFERENCE). AS A REFERENCE, FOR THE REGISTRATION RESULTS WITH THE LARGEST INITIAL MISALIGNMENT, THE RESULTS BEFORE REGISTRATION ARE MI:0.02±0.01, TREs:14.94±1.71MM (TOP 10%) AND MI:0.03±0.01, TREs:12.93±2.37MM (TOP 20%), RESPECTIVELY.

	10% with largest initial misalignment		20% with largest initial misalignment		10% with most improvement (selective)				20% with most improvement (selective)			
					Before		After		Before		After	
	MI	TREs(mm)	MI	TREs(mm)	MI	TREs(mm)	MI	TREs(mm)	MI	TREs(mm)	MI	TREs(mm)
NiftyReg	0.03±0.01	14.77±1.95	0.03±0.02	12.17±3.14	0.07±0.03	11.34±1.16	0.15±0.08	9.42±2.02	0.06±0.03	10.27±3.05	0.12±0.07	9.15±3.17
Direct	0.17±0.05	13.44±1.13	0.16±0.04	12.24±1.66	0.07±0.03	11.06±3.80	0.28±0.03	8.22±3.68	0.08±0.03	10.12±4.14	**0.28±0.03**	7.68±3.94
Mixed	0.16±0.05	12.59±0.25	0.16±0.05	11.66±1.20	0.07±0.03	12.63±4.45	0.28±0.03	9.32±3.98	0.08±0.03	10.12±4.14	0.27±0.02	7.55±3.46
Joint	0.13±0.07	10.99±1.15	0.13±0.06	9.49±1.82	0.07±0.03	13.63±3.20	0.27±0.03	8.57±2.93	0.08±0.03	11.13±3.71	0.27±0.02	6.99±3.08
Privileged	**0.18±0.04**	**7.45±0.88**	**0.17±0.03**	**6.77±1.36**	0.07±0.03	13.59±3.29	**0.28±0.03**	**5.16±2.11**	0.08±0.03	11.39±3.26	0.27±0.02	**4.07±2.09**

TABLE III
T2W-DWI$_{high-b}$ REGISTRATION PERFORMANCE ON EXTERNAL VALIDATION DATA SETS.

	Data set A		Data set B	
	MI	TREs(mm)	MI	TREs(mm)
w/o registration	0.06±0.03	11.01±4.06	0.04±0.02	6.89±2.14
NiftyReg	0.13±0.06	8.67±4.71	**0.16±0.05**	5.34±3.42
Direct	0.24±0.08	9.99±4.33	0.07±0.03	5.45±2.81
Mixed	0.24±0.08	8.43±4.61	0.07±0.03	5.26±2.84
Joint	0.22±0.09	9.19±4.20	0.06±0.02	5.98±2.42
Privileged	**0.25±0.07**	**6.53±4.20**	0.08±0.03	**4.58±2.72**

Sec. II-B. The mean TRE increased from 4.456±2.055mm to 4.347±1.845mm. Improvement was observed but without statistical significance (p-value=0.34). Considering the computation-associated feasibility of the propose algorithm, we chose to report results based on $I = 5$, as the number for the Monte-Carlo resampling.

D. Comparison to other learning-based registration

Direct registration marginally lowered the mean TRE, although with significance (p-value$<$0.001). The proposed privileged learning method obtained a lower mean TRE, compared to the direct registration method with statistical significance (p-value$<$0.001). It is consistent with the observations from the qualitative results in 4, which indicates the privileged learning method showed effective registration itself (Sect IV-A) whilst the Direct method did not. Interestingly, the Direct method produced a relatively high MI. This may be expected as the direct algorithm was trained to maximise MI directly, but the optimization was influenced by the heavy noise can lead to inferior TREs without the potential benefits from the added DWI$_{b=0}$ images, as discussed in Sect IV-A. It is also interesting to report that, using T2w and DWI$_{b=0}$ as network input both in training and testing (Section III-C), the warped DWI$_{b=2000}$ also led to a mean TREs of 4.59±2.01mm, outperforms the Direct, Joint, and the Mixed methods (p-value$<$0.001). These results summarise that the non-trivial difficulties in direct registering T2w and DWI$_{b=2000}$.

E. Comparison to non-learning registration

NiftyReg also improved the mean with statistical significance achieved(p-values=0.03), but the improvement is very limited. The mean and median TRE from privileged supervision is improved over those from NiftyReg results (p-value$<$0.001). Results from the two subgroups are summarised in Table II. It may be interesting to report that, the selective group (the lower two columns in Table II), on which registration provided most improvement, has a larger misalignment with the privileged modality, compared to those with NiftyReg. This perhaps indicates the potential utilisation of the extra anatomical and pathological information retained in the privileged images.

F. Interpretation of the external validations

Table III summarises the registration performance from each method on two external validation data sets. On both data sets, our proposed Privileged method outperforms the results before registration and from the other methods (all p-values$\leq$0.01). Compared with the original data set, the Data Set A is with larger initial misalignment, with a mean TRE of 11.01mm. Although Data Set B is with smaller initial misalignment, it was acquired with larger difference in acquisition protocols. For example, the MRIs from Data Set B were taken with endorectal coil and the b-value of the DWI$_{high-b}$ is only 1400 sec/mm^2. The Direct method and the Joint method show smaller improvements on TREs for both data sets, compared with the proposed method, albeit arguably smaller difference in the optimised MI. The Mixed method achieved relatively competitive performance on both data sets, second to the Privileged method. It probably because that the mixed sampling introduced more training data and thus increased its generalisability. It is also interesting to report that, the NiftyReg achieved lower TREs than the Direct and Joint methods in the external validation, although statistical significance was not found in these cases (p-values=0.83 and 0.27, respectively).

V. DISCUSSION

The proposed use of the third modality was not only evident in helping many cases in our application, but also provide an interesting new mechanism beyond improving registration performance, by bringing in a potentially more intuitive and radiologically-interpretable modality, for future registration studies, such as investigating registration error distribution, local loss function design and evaluation methodologies.

This work examined a particularly challenging cross-modality task in registering T2w and high b-value diffusion scans, from prostate cancer patients. During the investigation, we summarise the difficulties as follows: 1) high variance exist in both imaging and validation landmark annotating, in particular, clinical data contain variable and unknown misalignment from different patients; 2) the lack of consistent and robust similarity measures as a loss function between the two complementary imaging modalities.

Largely motivated by the high efficiency from the recent learning-based registration methods, we developed and compared registration networks and their associated training strategies. More interestingly, we proposed to use a third modality image that is arguably ”closer” to both images to register to help the training procedure. In experimental results, we show that such addition could indeed help the registration in a number of scenarios, with consistent and statistically significant advantages with the moderately-sized multimodal image data set from clinical practice.

In summary, the presented experimental results confirmed that the proposed registration network training method can benefit from an additional modality during training. The improvement over other learning-based method, with different ways to make use of the “privileged modality” or without using it at all, is effective and consistent, especially for a subset of these patient cases that with largest misalignment, therefore needing the registration the most.

We have demonstrated the proposed registration method using a privileged modality with the specific prostate cancer

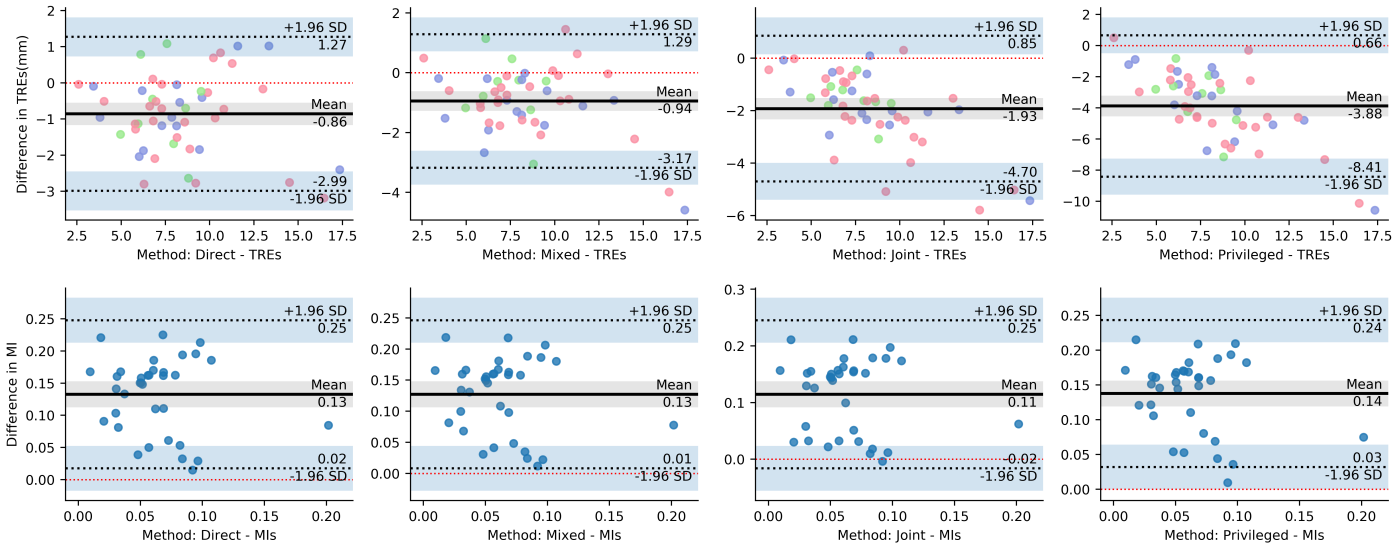

Fig. 5. Bland-Altman plots of the TRE and MI differences following the proposed privileged supervision registration algorithm. Each point represents a pair of landmarks in the holdout set, where the x axis represents the TREs (1st row) and the MIs (2nd row) before registration and the y axis represents the differences of the TREs and MIs after registration. The colors of the points in the first row of figures indicates the types of the landmarks (Blue: tumors; Yellow: urethra; Red: zonal structures).

imaging application. While this method has potentials for training registration networks using other types of available images in wider clinical applications, including and beyond those potential applications discussed in Section I, these require further investigation and validation.

VI. Conclusion

We have proposed strategies for the third modality images to aid the training of bi-modality image registration networks. The competitive registration accuracy has been experimentally demonstrated on mpMR data from prostate cancer patients. The proposed novel methodology may be generally applicable to a wide range of clinical image registration tasks.

Acknowledgment

This work was supported by the International Alliance for Cancer Early Detection, a partnership between Cancer Research UK [C28070/A30912; C73666/A31378], Canary Center at Stanford University, the University of Cambridge, OHSU Knight Cancer Institute, University College London and the University of Manchester. This work was also supported by the Wellcome/EPSRC Centre for Interventional and Surgical Sciences [203145Z/16/Z], the Wellcome/EPSRC Centre for Medical Engineering 203148/Z/16/Z; NS/A000049/1] (TV), the EPSRC CDT in i4health [EP/S021930/1], an MRC Clinical Research Training Fellowship [MR/S005897/1] (VS), a Royal Academy of Engineering / Medtronic Research Chair [RCSRF1819\7\734] (TV). For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

[1] N. Mottet, R. C. van den Bergh, E. Briers, T. Van den Broeck, M. G. Cumberbatch, M. De Santis, S. Fanti, N. Fossati, G. Gandaglia, S. Gillessen *et al.*, "Eau-eanm-estro-esur-siog guidelines on prostate cancer—2020 update. Part 1: screening, diagnosis, and local treatment with curative intent," *European Urology*, vol. 79, no. 2, pp. 243–262, 2021.

[2] M. A. Bjurlin, P. R. Carroll, S. Eggener, P. F. Fulgham, D. J. Margolis, P. A. Pinto, A. B. Rosenkrantz, J. N. Rubenstein, D. B. Rukstalis, S. S. Taneja *et al.*, "Update of the standard operating procedure on the use of multiparametric magnetic resonance imaging for the diagnosis, staging and management of prostate cancer," *The Journal of Urology*, vol. 203, no. 4, pp. 706–712, 2020.

[3] P. F. Fulgham, D. B. Rukstalis, I. B. Turkbey, J. N. Rubenstein, S. Taneja, P. R. Carroll, P. A. Pinto, M. A. Bjurlin, and S. Eggener, "Aua policy statement on the use of multiparametric magnetic resonance imaging in the diagnosis, staging and management of prostate cancer," *The Journal of Urology*, vol. 198, no. 4, pp. 832–838, 2017.

[4] N. B. Delongchamps, M. Rouanne, T. Flam, F. Beuvon, M. Liberatore, M. Zerbib, and F. Cornud, "Multiparametric magnetic resonance imaging for the detection and localization of prostate cancer: combination of T2-weighted, dynamic contrast-enhanced and diffusion-weighted imaging," *BJU International*, vol. 107, no. 9, pp. 1411–1418, 2011.

[5] C. K. Kim, B. K. Park, H. M. Lee, and G. Y. Kwon, "Value of diffusion-weighted imaging for the prediction of prostate cancer location at 3T using a phased-array coil: preliminary results," *Investigative Radiology*, vol. 42, no. 12, pp. 842–847, 2007.

[6] H. A. Vargas, O. Akin, T. Franiel, Y. Mazaheri, J. Zheng, C. Moskowitz, K. Udo, J. Eastham, and H. Hricak, "Diffusion-weighted endorectal MR imaging at 3T for prostate cancer: tumor detection and assessment of aggressiveness," *Radiology*, vol. 259, no. 3, pp. 775–784, 2011.

[7] X. Yang, C. Liu, Z. Wang, J. Yang, H. Le Min, L. Wang, and K.-T. T. Cheng, "Co-trained convolutional neural networks for automated detection of prostate cancer in multi-parametric MRI," *Medical Image Analysis*, vol. 42, pp. 212–227, 2017.

[8] X. Yang, Z. Wang, C. Liu, H. M. Le, J. Chen, K.-T. T. Cheng, and L. Wang, "Joint detection and diagnosis of prostate cancer in multi-parametric MRI based on multimodal convolutional neural networks," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, vol. 10435. Springer, 2017, pp. 426–434.

[9] N. Aldoj, S. Lukas, M. Dewey, and T. Penzkofer, "Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network," *European Radiology*, vol. 30, no. 2, pp. 1243–1253, 2020.

[10] R. Cao, A. M. Bajgiran, S. A. Mirak, S. Shakeri, X. Zhong, D. Enzmann, S. Raman, and K. Sung, "Joint prostate cancer detection and gleason score prediction in mp-MRI via Focalnet," *IEEE Transactions on Medical Imaging*, vol. 38, no. 11, pp. 2496–2506, 2019.

[11] M. H. Le, J. Chen, L. Wang, Z. Wang, W. Liu, K.-T. T. Cheng, and X. Yang, "Automated diagnosis of prostate cancer in multi-parametric MRI based on multimodal convolutional neural networks," *Physics in Medicine & Biology*, vol. 62, no. 16, p. 6497, 2017.

[12] B. Turkbey, A. B. Rosenkrantz, M. A. Haider, A. R. Padhani, G. Villeirs, K. J. Macura, C. M. Tempany, P. L. Choyke, F. Cornud, D. J. Margolis *et al.*, "Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2," *European Urology*, vol. 76, no. 3, pp. 340–351, 2019.

[13] V. Giannini, S. Mazzetti, A. Vignati, F. Russo, E. Bollito, F. Porpiglia, M. Stasi, and D. Regge, "A fully automatic computer aided diagnosis system for peripheral zone prostate cancer detection using multi-parametric magnetic resonance imaging," *Computerized Medical Imaging and Graphics*, vol. 46, pp. 219–226, 2015.

[14] R. A. Rakow-Penner, N. S. White, D. J. Margolis, J. K. Parsons, N. Schenker-Ahmed, J. M. Kuperman, H. Bartsch, H. W. Choi, W. G. Bradley, A. Shabaik *et al.*, "Prostate diffusion imaging with distortion correction," *Magnetic Resonance Imaging*, vol. 33, no. 9, pp. 1178–1181, 2015.

[15] K. V. Embleton, H. A. Haroon, D. M. Morris, M. A. L. Ralph, and G. J. Parker, "Distortion correction for diffusion-weighted MRI tractography and fMRI in the temporal lobes," *Human Brain Mapping*, vol. 31, no. 10, pp. 1570–1587, 2010.

[16] A. B. Rosenkrantz, S. Kim, R. P. Lim, N. Hindman, F.-M. Deng, J. S. Babb, and S. S. Taneja, "Prostate cancer localization using multiparametric MR imaging: comparison of prostate imaging reporting and data system (PI-RADS) and likert scales," *Radiology*, vol. 269, no. 2, pp. 482–492, 2013.

[17] M. De Luca, V. Giannini, A. Vignati, S. Mazzetti, C. Bracco, M. Stasi, E. Armando, F. Russo, E. Bollito, F. Porpiglia *et al.*, "A fully automatic method to register the prostate gland on T2-weighted and EPI-DWI images," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011, pp. 8029–8032.

[18] Y. Fu, T. Wang, Y. Lei, P. Patel, A. B. Jani, W. J. Curran, T. Liu, and X. Yang, "Deformable MR-CBCT prostate registration using biomechanically constrained deep learning networks," *Medical Physics*, vol. 48, no. 1, pp. 253–263, 2021.

[19] "Reimagine prostate cancer risk - full text view." [Online]. Available: https://clinicaltrials.gov/ct2/show/NCT04060589

[20] T. de Perrot, M. Scheffler, J. Boto, B. M. Delattre, C. Combescure, M. Pusztaszeri, J.-C. Tille, C. Iselin, and J.-P. Vallée, "Diffusion in prostate cancer detection on a 3T scanner: How many b-values are needed?" *Journal of Magnetic Resonance Imaging*, vol. 44, no. 3, pp. 601–609, 2016.

[21] V. Vapnik and R. Izmailov, "Learning using privileged information: similarity control and knowledge transfer." *Journal of Machine Learning Research*, vol. 16, no. 1, pp. 2023–2049, 2015.

[22] X. Yang, M. Wang, and D. Tao, "Person re-identification with metric learning using privileged information," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 791–805, 2018.

[23] Y. Hu, M. Modat, E. Gibson, W. Li, N. Ghavami, E. Bonmati, G. Wang, S. Bandula, C. M. Moore, M. Emberton *et al.*, "Weakly-supervised convolutional neural networks for multimodal image registration," *Medical image analysis*, vol. 49, pp. 1–13, 2018.

[24] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "Voxelmorph: a learning framework for deformable medical image registration," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

[25] A. Hering, S. Kuckertz, S. Heldmann, and M. P. Heinrich, "Enhancing label-driven deep deformable image registration with local distance metrics for state-of-the-art cardiac motion tracking," in *Bildverarbeitung für die Medizin 2019*. Springer, 2019, pp. 309–314.

[26] Y. Sun, A. Moelker, W. J. Niessen, and T. van Walsum, "Towards robust CT-Ultrasound registration using deep learning methods," in *Understanding and Interpreting Machine Learning in Medical Image Computing Applications*. Springer, 2018, pp. 43–51.

[27] S. Sun, J. Hu, M. Yao, J. Hu, X. Yang, Q. Song, and X. Wu, "Robust multimodal image registration using deep recurrent reinforcement learning," in *Asian Conference on Computer Vision*. Springer, 2018, pp. 511–526.

[28] L. Sun and S. Zhang, "Deformable MRI-ultrasound registration using 3D convolutional neural network," in *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation*. Springer, 2018, pp. 152–158.

[29] C. Stergios, S. Mihir, V. Maria, C. Guillaume, R. Marie-Pierre, M. Stavroula, and P. Nikos, "Linear and deformable image registration with 3D convolutional neural networks," in *Image Analysis for Moving Organ, Breast, and Thoracic Images*. Springer, 2018, pp. 13–22.

[30] H. Sokooti, B. De Vos, F. Berendsen, B. P. Lelieveldt, I. Išgum, and M. Staring, "Nonrigid image registration using multi-scale 3D convolutional neural networks," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, vol. 10433. Springer, 2017, pp. 232–239.

[31] R. W. So and A. C. Chung, "A novel learning-based dissimilarity metric for rigid and non-rigid medical image registration by using bhattacharyya distances," *Pattern Recognition*, vol. 62, pp. 161–174, 2017.

[32] J. M. Sloan, K. A. Goatman, and J. P. Siebert, "Learning rigid image registration - utilizing convolutional neural networks for medical image registration," in *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies - BIOIMAGING,*, INSTICC. SciTePress, 2018, pp. 89–99.

[33] M. Simonovsky, B. Gutiérrez-Becker, D. Mateus, N. Navab, and N. Komodakis, "A deep metric for multimodal registration," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, vol. 9902. Springer, 2016, pp. 10–18.

[34] C. Qin, B. Shi, R. Liao, T. Mansi, D. Rueckert, and A. Kamen, "Unsupervised deformable registration for multi-modal images via disentangled representations," in *International Conference on Information Processing in Medical Imaging*. Springer, 2019, pp. 249–261.

[35] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "An unsupervised learning model for deformable medical image registration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9252–9260.

[36] J. Krebs, T. Mansi, B. Mailhé, N. Ayache, and H. Delingette, "Unsupervised probabilistic deformation modeling for robust diffeomorphic registration," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2018, vol. 11045, pp. 101–109.

[37] C. Qin, W. Bai, J. Schlemper, S. E. Petersen, S. K. Piechnik, S. Neubauer, and D. Rueckert, "Joint learning of motion estimation and segmentation for cardiac MR image sequences," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, vol. 11071. Springer, 2018, pp. 472–480.

[38] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, "Unsupervised learning for fast probabilistic diffeomorphic registration," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, vol. 11070. Springer, 2018, pp. 729–738.

[39] M. C. Lee, O. Oktay, A. Schuh, M. Schaap, and B. Glocker, "Image-and-spatial transformer networks for structure-guided image registration," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, vol. 11765. Springer, 2019, pp. 337–345.

[40] J. S. Liu, R. Chen, and T. Logvinenko, "A theoretical framework for sequential importance sampling with resampling," in *Sequential Monte Carlo Methods in Practice*. Springer, 2001, pp. 225–246.

[41] R. Fitzpatrick, *Euclid's Elements of Geometry*. Euclidis Elementa, 2007.

[42] A. Fedorov, M. G. Vangel, C. M. Tempany, and F. M. Fennessy, "Multiparametric magnetic resonance imaging of the prostate: repeatability of volume and apparent diffusion coefficient quantification," *Investigative Radiology*, vol. 52, no. 9, p. 538, 2017.

[43] J. Chappelow, B. N. Bloch, N. Rofsky, E. Genega, R. Lenkinski, W. DeWolf, and A. Madabhushi, "Elastic registration of multimodal prostate MRI and histology via multiattribute combined mutual information," *Medical Physics*, vol. 38, no. 4, pp. 2005–2018, 2011.

[44] Y. Fu, N. M. Brown, S. U. Saeed, A. Casamitjana, Z. M. C. Baum, R. Delaunay, Q. Yang, A. Grimwood, Z. Min, S. B. Blumberg, J. E. Iglesias, D. C. Barratt, E. Bonmati, D. C. Alexander, M. J. Clarkson, T. Vercauteren, and Y. Hu, "Deepreg: a deep learning toolkit for medical image registration," *Journal of Open Source Software*, vol. 5, no. 55, p. 2705, 2020. [Online]. Available: https://doi.org/10.21105/joss.02705

[45] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library." in *NeurIPS*, vol. 32, 2019.

[46] A. Nazib, C. Fookes, and D. Perrin, "A comparative analysis of registration tools: Traditional vs deep learning approach on high resolution tissue cleared data," *arXiv preprint arXiv:1810.08315*, 2018.

[47] J. P. Pluim, J. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, 2003.
[48] T. Gaens, F. Maes, D. Vandermeulen, and P. Suetens, "Non-rigid multimodal image registration using mutual information," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, vol. 1496. Springer, 1998, pp. 1099–1106.
[49] X. Lu, S. Zhang, H. Su, and Y. Chen, "Mutual information-based multimodal image registration using a novel joint histogram estimation," *Computerized Medical Imaging and Graphics*, vol. 32, no. 3, pp. 202–209, 2008.
[50] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, 1997.
[51] J. Sun, C. Liu, C. Li, Z. Lu, M. He, L. Gao, T. Lin, J. Sui, K. Xie, and X. Ni, "Crossmodalnet: exploiting quality preoperative images for multimodal image registration," *Physics in Medicine & Biology*, vol. 66, no. 17, p. 175002, 2021.
[52] Q. Yang, Y. Fu, F. Giganti, N. Ghavami, Q. Chen, J. A. Noble, T. Vercauteren, D. Barratt, and Y. Hu, "Longitudinal image registration with temporal-order and subject-specificity discrimination," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, vol. 12263. Springer, 2020, pp. 243–252.