Deep Feature based Cross-slide Registration

*Ruqayya Awan, Shan E Ahmed Raza, Johannes Lotz, Nick Weiss and Nasir Rajpoot

***Abstract*— Registration of multiple sections in a tissue block is an important pre-requisite task before any cross-slide image analysis. Non-rigid registration methods are capable of finding correspondence by locally transforming a moving image. These methods often rely on an initial guess to roughly align an image pair linearly and globally. This is essential to prevent convergence to a non-optimal minimum. We explored a deep feature based registration (DFBR) method which utilises data-driven descriptors to estimate the global transformation. We adopted a multi-stage strategy for improving the quality of registration. We also developed a visualisation tool to view registered pairs of WSIs at different magnifications. With the help of this tool, one can apply a transformation on the fly without the need to generate a transformed moving WSI in a pyramidal form. We compared the performance on our dataset of data-driven descriptors with that of hand-crafted descriptors. Our approach can align the images with only small registration errors. We also evaluated the efficacy of our proposed method for a subsequent non-rigid registration step. To this end, we replaced the first two steps of the ANHIR winner's framework with DFBR to register image pairs provided by the challenge. The modified framework produced comparable results to those of the challenge winning team.**

***Index Terms*— ANHIR, Deep Learning, Non-rigid Registration, Rigid Registration, WSI Visualisation Tool**

I. Introduction

REGISTRATION refers to aligning a pair of images where one of the images is referred to as the *reference image* while the other image is referred as the *moving image*. The moving image is spatially transformed so that it aligns with the reference image. Registration often serves as an essential pre-processing step for many medical image analysis tasks.

Broadly speaking, histology image registration has three main applications: cross slide image analysis, multi-modal image fusion and 3D reconstruction from serial sections. Cross-slide image analysis provides additional information by analysing the expression of different biomarkers as compared to a single slide analysis, as in [1] for instance. Slides stained with different biomarkers are analysed side by side which may reveal unknown relationships between the different biomarkers. For instance, phosphohistone H3 (PHH3) stained WSIs show nuclear expression for mitotic figures and can be used to perform nucleus detection and classification [2] on registered H&E stained WSIs [3]. Multi-modal image fusion refers to the co-registration and the fusion of images obtained using different imaging modalities. In [4], fusion of a histology image is performed with the magnetic resonance (MR) image to characterize lung inflammation. 3D reconstruction from serial sections involves constructing a 3D model of an organ or any part of the body system. An example is the reconstruction of microvasculature presented in [5], [6]. In our work, emphasis has been given to the registration of a stack of consecutive multi-stain histology images. In another study [7], these multi-stain registered images are used to study mismatch repair (MMR) and microsatellite instability (MSI). In this study, the patch-level ground truth labels are generated by computing H-score values from the corresponding patches in MMR stained images.

Cross-slide analysis requires the alignment of WSIs of serial sections. During slide preparation, tissue sections cut from the same tissue block will not retain their continuity in the z-axis. Therefore, registering these images is an important step prior to any automated cross-slide analysis. Currently, registration by pathologists is done manually, which is time-consuming to align with sufficient accuracy at higher magnifications in addition to the large number of sections taken from a single tissue block. The introduction of an automated alignment mechanism can permit high-throughput cross-slide analysis in both manual and automated frameworks. However, automated cross-slide registration of histology images is a challenging task due to: changing structure between the sections, missing parts, tissue folds and broken tissue. The overall morphology of the tissue could change due to its fragility. There is a need for a registration approach that can perform well under these conditions and is able to align pairs of images in a reasonable time to facilitate the downstream analytical or diagnostic pipeline.

There are two main methods for automatic image registration: intensity-based registration and feature-based registration [8], [9]. As the name suggests, intensity-based techniques utilise the pixel intensity information in an image pair and optimise a similarity metric to find the correspondence. The correspondence is found by transforming a moving image such that it maximises the similarity measure between the reference and transformed moving images. On the other hand, feature-based methods would first identify the key features from the images and then a transformation is estimated by a matching system utilising the matching feature points between

Manuscript received May 7, 2022. The asterisk indicates the corresponding author.

Ruqayya Awan, Shan E Ahmed Raza and Nasir Rajpoot are with the Department of Computer Science, University of Warwick, CV4 7AL Coventry, U.K. (e-mail: awanruqayya2@gmail.com, shan.raza@warwick.ac.uk, n.m.rajpoot@warwick.ac.uk)

Johannes Lotz and Nick Weiss are with the Fraunhofer Institute for Digital Medicine MEVIS, Lübeck, Germany. (e-mail: johannes.lotz@mevis.fraunhofer.de, nick.weiss@mevis.fraunhofer.de).

Nasir Rajpoot is also with the Department of Pathology, University Hospitals Coventry, Warwickshire, UK and The Alan Turing Institute, London, UK.

two images. The choice between the two methods depends on the nature of the images.

Cross-section slides mostly have non-linear deformations that cannot be tackled with a global transformation alone. On the other hand, non-rigid registration methods are capable of finding correspondence by locally transforming a moving image. However, registration with Newton-type optimization relies on an initial guess that is close to the optimum to avoid convergence to a poor quality local minimum and also obtains fast convergence. The focus of this study is to explore the data-driven features for the estimation of the global transformation and demonstrate its robustness for aligning multi-stained tissue images. We refer to this approach as "deep feature based registration" (DFBR). We employed an existing method for non-rigid registration [10], [11] which has been successfully applied to multi-stain histology images. We present a comparative performance analysis of deep features compared with hand-crafted features. Experimentation was done on the COMET dataset and an additional multi-IHC histology dataset [12] which was released by the organisers of a recent challenge contest on non-linear registration.

II. Literature Review

CNN features have been shown to outperform handcrafted features by a large margin in several different tasks, across the computer vision and medical imaging domains. Several studies have attempted to use deep learning models as feature descriptors in a matching task for medical images. In [13], [14], the authors proposed an unsupervised approach for learning discriminating features. These are later integrated into existing registration tools for the prediction of a dense deformation field to register MR brain images. In line with the above two studies, deep features were used for predicting the rigid transformation for histology images [15]. In that work, an autoencoder was trained to generate an output similar to the input to learn a feature representation. Features extracted from the encoder part of a trained autoencoder were used to find the best transformation using gradient descent.

There is a significant amount of work done with handcrafted features for an alignment task — [1], [16]–[22] to list a few. There have been very few CNN based studies predicting the transformation parameters for a highly deformed pair of images. This is because a known correspondence between two images is needed for training a CNN, and this is seldom available. Also, the trained CNN may not perform comparably well on an unseen dataset. These limitations can be addressed by using an unsupervised approach to some extent and are yet to achieve a significant improvement in terms of registration accuracy. In 2015, Jaderberg *et al.* [23] proposed a learnable module for applying the spatial transformation to an image, referred to as the 'spatial transformer'. Since this module is differentiable, it can be added to any network for end-to-end training. This was not primarily designed for registration purposes; instead, the aim was to enable CNN to learn features that are invariant to spatial transformations. After the introduction of the spatial transformer, deep learning gained momentum in designing neural network architectures suitable for registration in an unsupervised learning manner. It has now become a core component of most of deep learning-based registration methods. Chang *et al.* [24] proposed a multi-scale iterative framework for registering microscopic images of serial sections of a mouse brain. A CNN model with a spatial transformer as one of the layers was used to predict the affine transformation. The model was trained to minimise the mean square error between the reference and warped moving images. Shengyu *et al.* [25] proposed a deep learning architecture for 3D registration of CT images of the liver and MR images of the brain. The proposed architecture consisted of two sub-networks: one for predicting the affine transformation and the other for predicting the non-linear deformation field. This approach was also applied to the multi-stain histology dataset provided as a part of the ANHIR challenge [12] and was observed to be the fastest performing method. However, it was not close to the best-performing methods in terms of registration error due to its limited generalisability. It was ranked 6 out of 10 teams who submitted the results. In another study [26], the authors trained a CNN model with good generalisability for predicting the affine transformation in an unsupervised manner. They compared their results with that of SIFT, SURF and Elastix tool. Their proposed approach didn't outperform other methods in terms of accuracy; however, considering the success rate, the authors claimed that the reported accuracy had been sufficient to perform non-linear alignment successfully. Dwarikanath [27] integrated segmentation maps to aid in performing non-linear registration using a self-supervised deep learning-based approach. Segmentation maps were generated by applying k-means clustering to concatenated multi-scale feature maps, extracted from a pre-trained segmentation model. The author employed VoxelMorph architecture [28] for registration. In [28], manual segmentation maps are used to improve registration during training while in [27] manual segmentation maps are replaced with fine-grained feature maps. Similar to other non-linear registration methods, this method also requires two images to be linearly aligned before its application.

There has been a significant amount of work on non-linear registration. However, the focus has been on monomodal and multi-modal registration of X-ray, CT and MRI images and very few methods have been proposed for histology images. Wodzinski and Muller [29] proposed a deep learning based non-rigid registration method, performing comparably to the winning team of the ANHIR challenge contest and is significantly faster than other iterative methods. Their proposed approach employs UNET like architecture, trained in a multi-level unsupervised manner using negative normalised cross-correlation (NCC) as an objective function. However, data-driven approaches which learn from the data, most of them require a large number of training samples to perform highly accurate registration.

Our proposed approach for registering multi-stain images comes under the same umbrella of using CNN as a feature descriptor. Our work is inspired by the work in [30] on registering multi-temporal remote sensing images using a CNN, whereby the authors used multi-scale deep features for the detection of matching feature points between an image

pair. These matching feature points are then used to solve thin-plate spline (TPS) interpolation for image alignment. In our work, we have followed a similar approach for feature description and the correspondence between the feature points of two images was found by computing the Euclidean distance measure. Instead of TPS formulation, we used these matching feature keypoints for estimating the affine transformation.

III. The Proposed Approach

A registration step for any downstream co-localisation analysis workflow should be able to allow a significant spatial overlap between the two images such that the location of corresponding tissue structures can be determined. To this end, we propose a pipeline comprising three main steps: pre-processing, and estimation of rigid alignment using the DFBR method, followed by a non-linear registration. During the pre-processing step, we generate a tissue mask for an image pair and modify the input images such that they appear spatially similar. After an image is registered using our DFBR, we observe a slight offset in some cases. To fix this offset, we add a translative transform module which is followed by an existing non-linear registration method. The overall proposed pipeline for registration is shown in Figure 1.

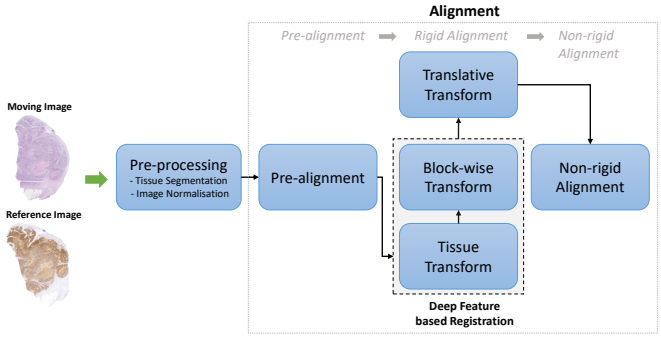

Fig. 1. Overall pipeline for cross-slide image registration. The pipeline comprises the deep feature based registration (DFBR) method, followed by an existing non-rigid registration method.

A. Pre-processing

1) Tissue Segmentation: We perform tissue segmentation to exclude features from the non-tissue regions. To this end, we generate tissue segmentation (TS) masks, considering all tissue and non-tissue regions as foreground and background respectively. In Figure 2, a downsampled version of WSI is shown along with its tissue masks. Note, in our experiments, fatty tissue is considered as a background since it does not contain adequate texture and it is very likely to generate poorly discriminatory features. Therefore, we exclude fatty tissue from the tissue mask to avoid getting incorrect matching points and refer to this mask as TSEF. The details of the tissue segmentation pipeline are given in Supplementary Figure 1. We train a CNN to generate these masks so that registration could be carried out using matching points from the active or discriminatory tissue area only (while excluding the ones heavily surrounded by the fatty tissue). Similarly to [19], control tissue is excluded while estimating the transformation parameters.

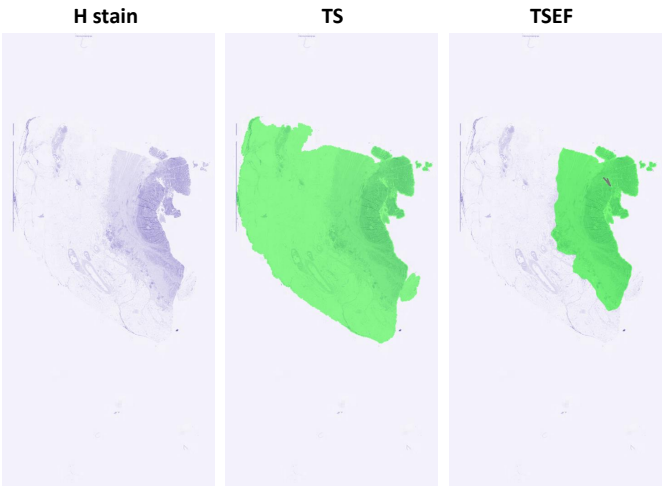

Fig. 2. An example image with the two types of tissue masks. In TS, all the tissue including the fatty region is included whereas, in TSEF, the fatty region is excluded due to its negative impact on our DFBR registration method. TSEF is used in our DFBR registration method.

2) Image Normalisation: The images are converted to greyscale, with histogram matching performed as a normalisation step to unify the appearance of an image pair. In histogram matching, the histogram of an image is modified to be similar to that of another image. Similarly to [19], the image with higher entropy is taken to be the reference image and the histogram of an image with lower entropy is matched to the reference image. Since most pre-trained CNN models accept input images with 3 channels, greyscale images were stacked as the colour channel.

B. Alignment

Broadly speaking, the tissue alignment is performed in three main steps: pre-alignment, rigid alignment and non-rigid alignment. The output registered image generated in each step is given as input to the next step along with the reference image. All these steps are discussed in detail in the following sections.

1) Pre-alignment: In this step, rough estimates of translational and angular offsets are computed. Since CNN features are not rotation invariant, this step is key to performing deep feature matching successfully. First, we estimate the translation offset by finding a centre of mass (COM) for an image pair. The COM is a vector of x and y coordinates and is computed from the inverted greyscale intensity values of the tissue region only. The difference between the COM values of a pair of images is used to obtain a translation matrix. This matrix transforms the moving image such that its COM is at the same position in the coordinate system as that of a reference image. Next, we compute a rotation matrix with different angles ranging between 1 and 360 with a step of 10. We then select a rotation matrix resulting in a maximum

DICE overlap between the tissue masks of an image pair. Once a moving image is roughly aligned with a reference image, we crop the tissue regions from an image pair using their tissue masks. We determine a bounding box that includes the tissue region of both images. In the following steps, we only use tissue regions instead of the whole image for registration.

2) Deep Feature based Registration: The objective of our feature based alignment step is to refine the alignment between reference and pre-registered moving images by registering their feature points. We present data-driven features extracted using a pre-trained VGG-16 model [31]. The partial architecture of VGG-16 that we use for feature extraction is shown in Figure 3. Features are extracted from three different layers of VGG-16 for an image pair and are referred to as multi-scale features in this paper. The features are then processed to find matching pairs by considering the feature points at a small feature distance.

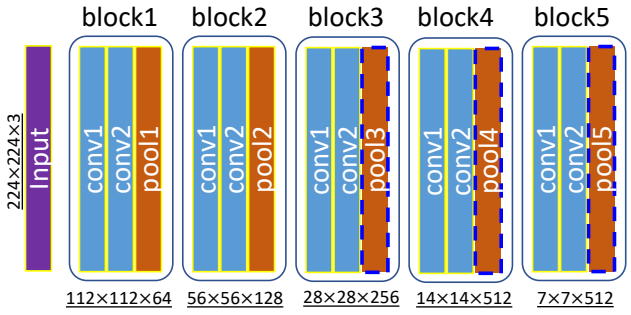

Fig. 3. Partial architecture of the VGG16 model used for the DFBR module. Layers highlighted with dashed blue borders are used for feature extraction in DFBR.

A. Feature Descriptors: Feature descriptors describe the properties of regions centred at the distinctive feature points and are used to find correspondence between two images in a feature matching step. In a classical feature based method for registration, the first step is to detect the feature points, followed by a feature extraction step which involves the computation of descriptors. Whereas, in our feature based method for alignment, the feature detection step doesn't exist. Instead, an image is divided into a grid and a feature descriptor is computed for every grid cell in a sliding window fashion. Similar to [30], our feature descriptor is formed by deep features extracted from three different layers of a VGG-16 model, pre-trained for ImageNet classification. Since the bottom fully connected layers are removed, the partial model can take input images of any size (spatial dimensions multiples of 32), with larger images increasing the computation time. Our experiments are conducted with an input size of 224×224 pixels. Features extracted from the bottom three pooling layers (pool3, pool4 and pool5) are used to build descriptors. Each of these layers corresponds to receptive fields [32] of different sizes as shown in Figure 4. The pool4 and pool5 feature descriptors are mapped to those of pool3 such that they have the same spatial dimensions. Following are some annotations that we use in this section: F_i^j refers to the feature descriptor of an image j extracted from a pooling layer i. For example, pool3 feature descriptor of a reference image R is denoted by F_3^R.

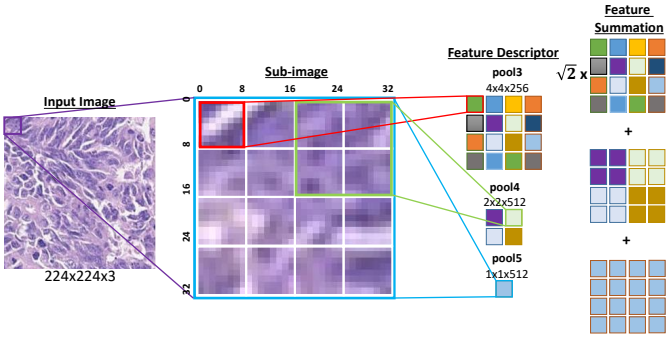

Fig. 4. Demonstration of feature summation procedure. Feature descriptors are shown for a sub-image of size $\mathbf{32} \times \mathbf{32}$ pixels. For pool3, each feature descriptor corresponds to an $\mathbf{8} \times \mathbf{8}$ image region, with the centre of that region considered as a feature point. For pool4 and pool5, each feature descriptor corresponds to a $\mathbf{16} \times \mathbf{16}$ and $\mathbf{32} \times \mathbf{32}$ image region, respectively. pool4 and pool5 descriptors are mapped to that of pool3 by duplicating them for the feature points they are sharing. For instance, the pool5 descriptor is shared by 16 feature points from pool3, hence is duplicated 16 times.

- **pool3** outputs feature descriptor F_3 of dimension $28 \times 28 \times 256$. Each descriptor has a receptive field of size 8×8, dividing the input image into a 28×28 grid. The center of each grid cell is considered a feature point for the respective descriptor.
- **pool4** outputs feature descriptor F_4 of dimension $14 \times 14 \times 512$. Each descriptor has a receptive field of size 16×16 and therefore it is shared by four feature points from pool3.
- **pool5** outputs feature descriptor F_5 of dimension $7 \times 7 \times 512$, each descriptor has a receptive field of size 32×32 and therefore it is shared by 16 feature points from pool3.

Each of these descriptors is normalised to unit variance.

B. Feature Matching: Once feature descriptors are formed for an image pair, the Euclidean distance is computed between all possible pairs of feature points. The distance metric for feature descriptor of layer i is computed as

$$D_i(p^R, p^M) = distance_{euc}(F_i^R, F_i^M)$$

where $i \in [3, 4, 5]$. F^R and F^M refer to feature descriptors computed for reference and moving images, respectively. Each value in a feature distance matrix for pool3 relates to an individual feature point, which is not the case with distance matrices for pool4 and pool5 feature descriptors. Each distance value in D_4 and D_5 corresponds to 4 and 16 feature points, respectively. Therefore, we replicate each distance value in D_4 4 times. Similarly each distance value in D_5 is replicated 16 times. D_3 has a smaller number of distance values and this is because pool3 outputs 256 feature channels whereas pool4 and pool5 output 512 feature channels. To compensate this difference, weight is given to D_3, followed by summation of D_3, D_4 and D_5. Feature distance between two feature points is computed as

Distance Matrix *D*

Image2 \ Image1	Feat1	Feat2	Feat3	Feat4	Feat5	Feat6	Feat7	Feat8	min_1	min_2	'Quality' min_2/min_1
Feat1	21.68	40.56	56.06	57.40	64.16	78.38	75.45	76.54	21.68	40.56	1.87
Feat2	38.71	23.49	42.56	44.44	52.00	75.16	73.38	79.80	23.49	38.71	1.65
Feat3	52.27	37.77	26.28	28.62	37.23	66.34	66.52	73.58	26.28	28.62	1.09
Feat4	52.58	37.48	25.28	27.75	36.64	66.31	66.53	73.63	25.28	27.75	1.10
Feat5	59.88	44.73	27.95	30.50	26.37	65.86	67.58	74.61	26.37	27.95	1.06
Feat6	21.68	40.56	56.06	57.40	64.16	38.02	58.74	66.78	21.68	38.02	1.75
Feat7	38.71	23.49	42.56	44.44	52.00	44.42	17.21	30.29	17.21	23.49	1.36
Feat8	52.27	37.77	26.28	28.62	37.23	58.56	30.86	14.72	14.72	26.28	1.79

Fig. 5. A toy example of a distance matrix to demonstrate how the quality of matched feature points is determined. Feature points with the smallest distance $\boldsymbol{min_1}$ are considered matched feature points. $\boldsymbol{min_2}$ is the second smallest distance after $\boldsymbol{min_1}$. The quality of matched feature points is determined by the difference between $\boldsymbol{min_1}$ and $\boldsymbol{min_2}$.

$$D(p^R, p^M) = \sqrt{2}D_3(p^R, p^M) + replicate(D_4(p^R, p^M), 4) + replicate(D_5(p^R, p^M), 16)$$

p^R and p^M refer to feature points of reference and moving images, respectively. The best candidate match for a feature point p_i^R is found by locating a feature point in p^M with the smallest Euclidean distance. However, this will generate incorrect matched pairs as well and would need discarding pairs that do not have any good match. For this purpose, we utilise Lowe's ratio test [33] and determine the matching quality of a matched pair of feature points by the Euclidean distance difference between the best candidate and the second best candidate. For a correct match, the distance of the best match should be significantly smaller than the second best match. While, for a false match, the distance of the best and second best match would be similar. Let's say that p_1^R is matching to p_8^M with the smallest distance between their descriptors while another pair of feature points with the second smallest distance for p_1^R is p_5^M, the quality is computed by taking $D(p_1^R, p_5^M)/D(p_1^R, p_8^M)$. Hence, the greater the difference, the better the quality. An example illustration for this is shown in Figure 5.

There are two conditions for matching point p^R to point p^M

1) $D(p^R, p^M) < D(:, p^M)$; which means there shouldn't be any other feature distance smaller than $D(p^R, p^M)$.
2) The quality of matched feature points should be greater than a threshold value which is computed automatically for each image pair. A threshold value is set such that U pairs of matching points are selected.

Using the above steps, we find matched feature points for any two images. Similarly to [30], we select U = 128 pairs of matched points. We then use matched feature points as control points for estimating the transformation parameters. We use a least-squared approach to find a rigid transformation that best aligns these matched points. To use this method for histology image registration, we apply this step twice. First, we apply it to the whole tissue region to find the best matched points globally. The resulting transformation is referred to as the 'tissue transform'. Secondly, we divide the tissue regions into four parts and apply the feature matching method to each part individually. This is referred to as the 'block-wise transform'. This procedure results in a larger number of matching feature points and is likely to further improve the alignment. The block-wise feature matching step can be applied in a parallel fashion to speed up the process. Both tissue and block-wise transforms are applied to the moving image only if they tend to improve the DICE overlap. The proposed DFBR approach is integrated into an existing Computational Pathology Toolbox named TIAtoolbox [34]. An example of registering a pair of images is shown at `https://tia-toolbox.readthedocs.io/en/latest/_notebooks/jnb/10-wsi-registration.html`.

3) Translative Transform: On visualising a registered image alongside its corresponding reference image at both low and high resolution, we observe global and local translation offsets, respectively. The global offset can reasonably be estimated from the offset of patches. On viewing an image pair at a higher resolution with a minor global offset, a local translation offset between reference and registered images can be observed. This is likely to persist due to non-linear deformations in some parts of the images. Supplementary Figure 2 shows an image overlay of reference and registered moving tissue sections along with their corresponding landmark points. This figure shows local offsets between the two images. To fix a global offset, we employ a phase correlation method after applying DFBR to determine a shift between two images at a magnification of 0.3125×. This method can be applied to the greyscale or H stain images or a tissue mask as in [1]. We apply it to tissue masks highlighting glandular structures and found that it performed better, perhaps because glands are the prevalent histological structures. We also integrate this method in our visualisation tool (section IV) to fix the local shift. During visualisation, the user can fix the offset by clicking on the 'Fix Offset' button. An example of an image pair before and after local refinement is shown in Figure 6.

4) Non-Rigid Alignment: Once an image pair is registered using rigid transformation, it is often the case that some of the tissue areas are not accurately aligned to that of a reference image. This is due to the fact that the tissue slices are so thin and fragile that the slide preparation step is likely to introduce non-linear deformations and artefacts such as tissue folds, tissue stretching and compression and even torn/missing tissue parts [35]. The presence of non-linear deformations makes the registration process more challenging. Since these artefacts change the morphology of the tissue, none of the rigid registration methods can tackle such deformities. Therefore, a non-rigid registration approach is applicable in such scenarios.

There are many existing non-rigid registration methods in the literature that have been used for histology image alignment. In the ANHIR challenge, several non-rigid methods were evaluated using a benchmark dataset. We apply a non-rigid algorithm [10] proposed by the winner of the ANHIR challenge (MEVIS group) on top of our DFBR method to further improve the registration accuracy. Their proposed method as used in the challenge comprises three steps: pre-alignment, parametric registration and non-parametric registration. The efficacy of non-parametric registration is highly dependent on

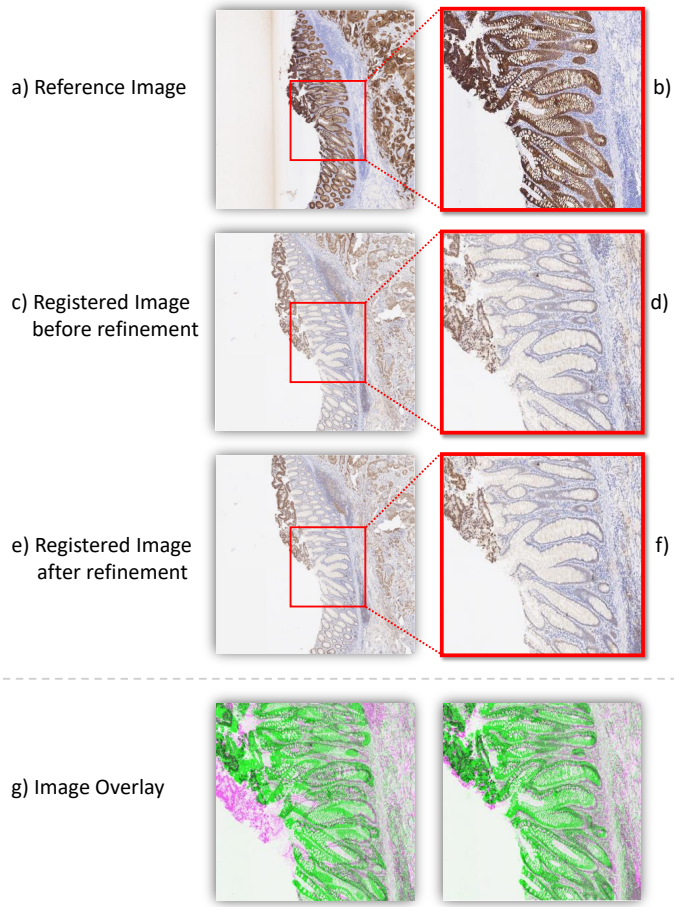

Fig. 6. Example images of the local phase-based refinement method integrated into our visualisation tool (see section IV) to fix translation offset. In bottom row g) we show overlaid false colour images before (left) and after refinement (right). Reference and registered moving images are shown in green and purple colours, respectively.

the accuracy of the initialiser transform. We experimented with their method after replacing their first two steps with our DFBR method. NGF distance measure [36] is used as a similarity metric in their framework whereas we used DICE overlap of tissue mask. Their pre-alignment step is similar to our pre-alignment step except that they do not perform tissue segmentation and exclusion of control tissue. While in the parametric step, the affine transformation is estimated based on the greyscale intensity values using a Gauss-Newton optimization, for which we employ the feature matching method. Both intensity and feature matching methods have their own limitations, however feature-based methods are often considered better in terms of robustness for handling transformations [37], [38].

IV. Visualisation Tool

To visualise registered images while being able to zoom in and out and pan across a WSI, we develop a web-based tool with a split screen, the left panel for displaying the reference image and the right panel for the registered moving image. On each split screen, a dot pointer with a different colour is shown. This changes its position with the mouse movement. The regions indicated by these points on the two screens are extremely helpful in visually estimating the performance of the registration method. A screenshot of the interface is shown in Supplementary Figure 3 while a short video of the interface is also provided in the Supplementary Materials.

This interface is developed to display the results of our deep feature based method. The input to this tool comprises three directory paths: to reference and moving images and the pre-computed affine transformation parameters. Registration is applied to the tiles on the fly as they are viewed. Therefore, there is no need to generate a transformed WSI in a pyramidal format. Since registration is performed at the lowest resolution, we observe a translation offset between image pairs which may also be due to a non-linear transformation. In this case, a part of the tissue has been distorted so that no global rigid transformation can fix the whole tissue. We deal with such misalignment by using the phase correlation method. We add a button to the interface for the user to fix the offset. Once the offset is computed, it is applied to every FOV as the user zooms or pans through the slide. The code for this tool is available online: `https://github.com/ruqayya/reg_visualization_tool`. The implementation is carried out in Python and JavaScript. We used OpenSeadragon, an open-source viewer for this tool.

V. Datasets & Performance Measures

We evaluate the performance of the DFBR method using two multi-stain datasets: the COMET and ANHIR datasets. A detailed description of the datasets and evaluation metrics is presented in the following subsections.

A. Datasets

The COMET dataset is obtained from the University Hospitals Coventry and Warwickshire (UHCW) NHS Trust in Coventry, UK. This dataset comprises WSIs of 86 cases, taken from different patients. There are 16 slides per case, each scanned using the Omnyx VL120 scanner at 0.275 microns/pixel. These slides are stained with different stains. The exact sequence of staining is CK8/18, Ki67, p53, Vimentin, CK8/18, MLH1, MSH2, MSH6 and PMS2, CK8/18, Ecadherin, EpCAM, PTEN, CK8/18, Negative control and H&E. We select a set of 7 cases and present the quantitative evaluation of registration methods using them. We consider six slides per case, involving MMR prediction from H&E and CK8/18 images for the end purpose of MSI prediction. Therefore, we perform registration of MMR slides with CK8/18 and H&E slides. The alignment of H&E wrt CK8/18 is challenging as there are 10 sections in between them, including the MMR markers' sections. CK8/18 and MMR biomarker slides are consecutive sections sliced in the given order: CK8/18, MLH1, MSH2, MSH6 and PMS2. These IHC stained slides tend to be highly correlated in terms of tissue structures. This is not the case with H&E slides which are around 50µm away from the neatest IHC slides. Therefore, substantial variability in the tissue structures exists among the H&E and IHC stained slides, as shown in Figure 7. For evaluation, we consider 15 pairs

of sections per case for registration: aligning MMR markers w.r.t CK8/18 (4 pairs), aligning all IHC slides w.r.t H&E (5 pairs) and all possible combinations of MMR biomarker slides (6 pairs). This results in 105 pairs in total against 7 cases. To evaluate the performance of registration methods, we manually defined landmarks on all the images selected for evaluating registration performance. The ImageJ tool is used for annotating significant tissue structures with landmarks.

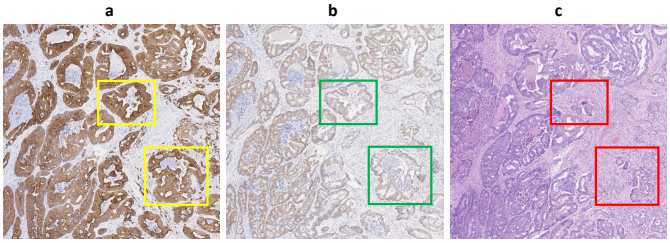

Fig. 7. Similarities and dissimilarities in the tissue architecture among spatially corresponding visual fields extracted from a) CK8/18, b) MLH1 and c) H&E stained images of the COMET dataset. a and b are taken from consecutive slides (5µm apart), and hence have more similarities compared to a and c or b and c (more than 50µm apart).

The ANHIR dataset is a public dataset, made available by organisers of the ANHIR challenge. It comprises 8 different tissue types stained with 18 different stains, hence making it a challenging dataset. There are 230 training and 251 testing pairs for registration. For more details on this dataset, readers are referred to the challenge paper [12]. Landmarks are provided for the training set while landmarks of reference images in the test image pairs are kept private by the organisers. The evaluation of a registration method can be performed by uploading the results on the challenge portal.

B. Performance Measures

To evaluate the performance of registration methods, we compute the target registration error (TRE) for image pairs in the test set $\mathcal{T}$. This registration metric was used by the ANHIR challenge organisers for evaluating the submissions [12]. For an image pair j, we compute the distance-based error measure using the following formula

$$TRE(R_j, M_j^{'}) = ||R_{land_j}, M_{land_j}^{'}||_2$$

where R and $M^{'}$ belong to an image pair j and represent reference and transformed moving images, respectively. R_{land} and $M_{land}^{'}$ denote landmarks of reference and transformed moving images, respectively. The registration error is normalised by the length of the reference image diagonal.

$$rTRE(R_j, M_j^{'}) = TRE(R_j, M_j^{'})/hypot(R_j)$$

where $hypot(R_j) = \sqrt{w^2 + h^2}$ and w and h denote the width and height of the reference image, respectively. The above equation for computing rTRE generates a list of values for an image pair j which we aggregated by taking their average. The aggregation results using the median are presented in the Supplementary material document. Overall registration error for $\mathcal{T}$ is computed by either taking average or median of the aggregated rTRE, namely as the average of average rTRE (AArTRE) and the median of average rTRE (MArTRE). We also report the average of the maximum rTRE (AMaxrTRE).

We also test the robustness of registration results by comparing the transformed landmarks with the initial landmarks before any alignment. We compute the robustness in a similar way defined by the ANHIR challenge organisers which is the relative number of successfully transformed landmarks. Any given landmark pair R_{land^i} and $M_{land^i}^{'}$ is considered to have been registered successfully only if the distance between them is smaller than the difference between R_{land^i} and landmark of moving image before registration M_{land^i}. Robustness for an image is computed by counting the number of successfully registered landmarks divided by the total number of landmarks for that image. Robustness over the whole dataset is computed by taking the mean over all the image pairs' robustness.

VI. Experimental Results

A. Experimental Setup

We consider different scales of information in our registration pipeline. For the COMET dataset, downsampled WSIs at magnification $0.15625\times$, $0.3125\times$ and $0.625\times$ are used for pre-alignment, DFBR including translative transform and non-rigid alignment, respectively. For the ANHIR dataset, downsampled WSIs are provided rather than WSIs with varying downsampling rates for different datasets. We rescale the images to 5% and perform our DFBR registration, followed by a non-rigid registration. For the non-rigid registration, we use the same parametric values as the challenge winner. The readers are referred to [12] for the non-rigid registration parameters.

B. Result Summary

1) COMET: Our DFBR method performs better when the transformation matrix is estimated using matching points from the discriminatory tissue area. This is accomplished using a TSEF mask as discussed in section III-A.1. The quantitative and qualitative comparison of DFBR using TS and TSEF masks are shown in Table I and Figure 8. Table II shows results obtained with different versions of input images: original RGB images, greyscale images and H stain images. As our initial experiments show greyscale images to perform better, greyscale images are used in all our experiments.

TABLE I
COMPARATIVE RESULTS OF THE DFBR METHOD USING AVERAGE AGGREGATION FOR TWO DIFFERENT TISSUE SEGMENTATION MASKS. THE EXCLUSION OF FATTY AREAS SIGNIFICANTLY DECREASED THE REGISTRATION ERROR. THESE RESULTS ARE GENERATED USING THE COMET DATASET. RESULTS USING MEDIAN AGGREGATION ARE PRESENTED IN SUPPLEMENTARY TABLE 1.

	AArTRE	**MArTRE**	**AMaxrTRE**
TS	0.0090	0.0044	0.0209
TSEF	**0.0073**	**0.0042**	**0.0177**

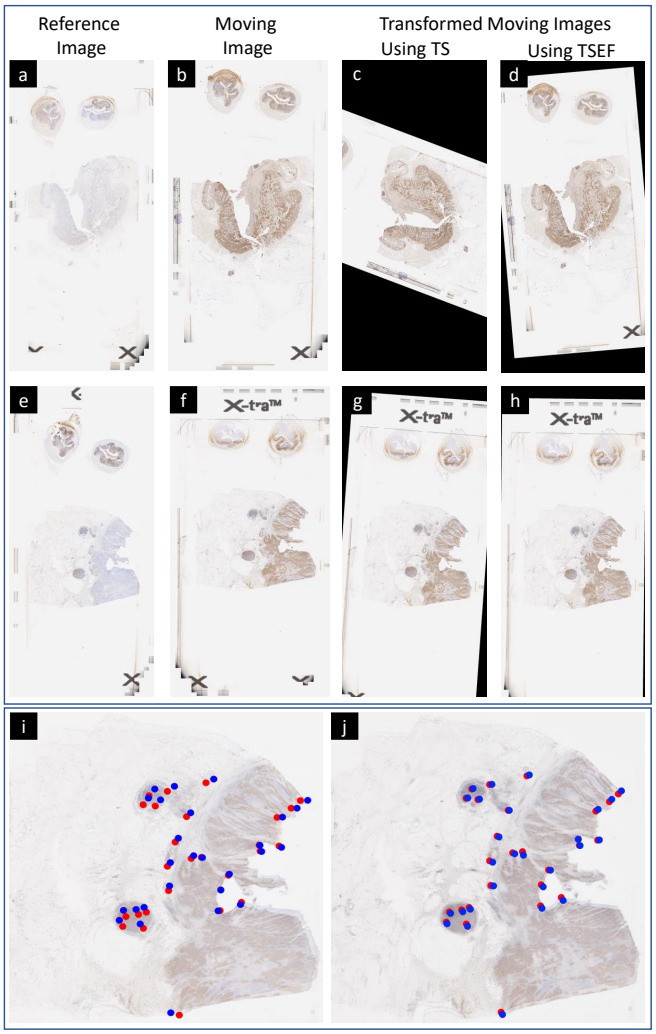

Fig. 8. A qualitative comparison of the registration accuracy of transformations estimated using segmentation TS and TSEF. a) and e) reference images, b) and f) moving images, c) and g) transformed moving images using TS while d) and h) transformed moving images using TFSER. i) shows the overlay of e) and g) and j) shows the overlay of e) and h). Landmarks are overlaid for the purpose of visualising the offset between the two images.

Each step in our pipeline is adopted to improve the alignment. The quantitative results in Table III demonstrate that our first three steps result in the improvement of the rigid alignment between the two images while the last step has been able to tackle the non-linear deformation. Box plots shown in Figure 9 demonstrate the reduction of average rTRE after each step in our pipeline. The pre-alignment step has significantly improved the average rTRE. Figure 10 shows overlay false colour images of reference and registered moving images along with the zoomed-in visual fields. In overlay images, the extent of the green colour indicates the extent of misalignment. It can be seen that the misalignment is improved after each step.

We conducted experiments with the MEVIS original framework on the COMET dataset and did not find reasonable results. This is due to the fact that this framework performs registration using overall image content. While, for the COMET dataset, tissue segmentation is a mandatory step to exclude tissue artefacts and most importantly the control tissue. Since the MEVIS framework does not take into account the presence of control tissue, it is likely to find a transformation that best aligns both control and main tissue, hence not resulting in an optimal alignment of the main tissue which we are only interested in. The output of the MEVIS framework on an example image pair from the COMET dataset is shown in Supplementary Figure 4.

TABLE II

DEMONSTRATION OF THE EFFICACY OF OUR DFBR APPROACH USING AVERAGE AGGREGATION FOR DIFFERENT VERSIONS OF INPUT IMAGE PAIRS. GREYSCALE IMAGES ARE SHOWN TO OUTPERFORM OTHER PRE-PROCESSED INPUT IMAGES. THESE RESULTS ARE GENERATED USING THE COMET DATASET. RESULTS USING MEDIAN AGGREGATION ARE PRESENTED IN SUPPLEMENTARY TABLE 2.

	AArTRE	**MArTRE**	**AMaxrTRE**
Blue Ratio	0.0147	0.0061	0.0288
RGB	0.0088	0.0047	0.0192
H stain	0.0079	0.0044	0.0187
Greyscale	**0.0073**	**0.0042**	**0.0177**

TABLE III

QUANTITATIVE RESULTS GENERATED USING AVERAGE AGGREGATION ON THE COMET DATASET. THESE RESULTS DEMONSTRATE THAT THE AVERAGE ERROR IS REDUCED WITH EACH ALIGNMENT STEP IN OUR PIPELINE. RESULTS USING MEDIAN AGGREGATION ARE PRESENTED IN SUPPLEMENTARY TABLE 3.

	AArTRE	**MArTRE**	**AMaxrTRE**
Initial	0.1800	0.1045	0.2334
Pre-alignment	0.0117	0.0088	0.0220
Tissue Transform	0.0083	0.0051	0.0187
Block-wise Transform	0.0073	0.0042	0.0177
Non-rigid Transform	0.0050	0.0031	0.0188

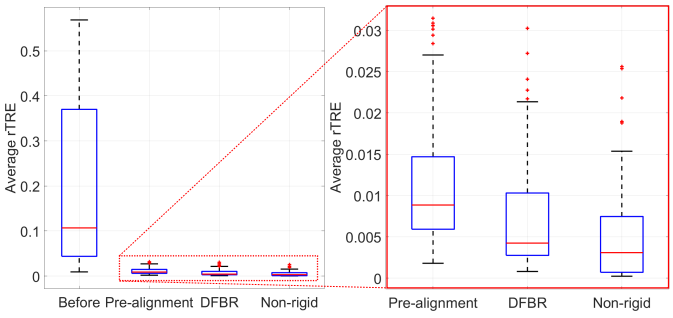

Fig. 9. Demonstration of the average rTRE before and after applying each registration module using a box plot. The average rTRE is computed for the COMET dataset. The box plots for the median rTRE are shown in Supplementary Figure 5.

2) ANHIR: Similar to the COMET dataset, we evaluate each step of our registration pipeline on the ANHIR dataset. As the organisers made landmarks of the training set publicly available, we first evaluate our pipeline on the training set only.

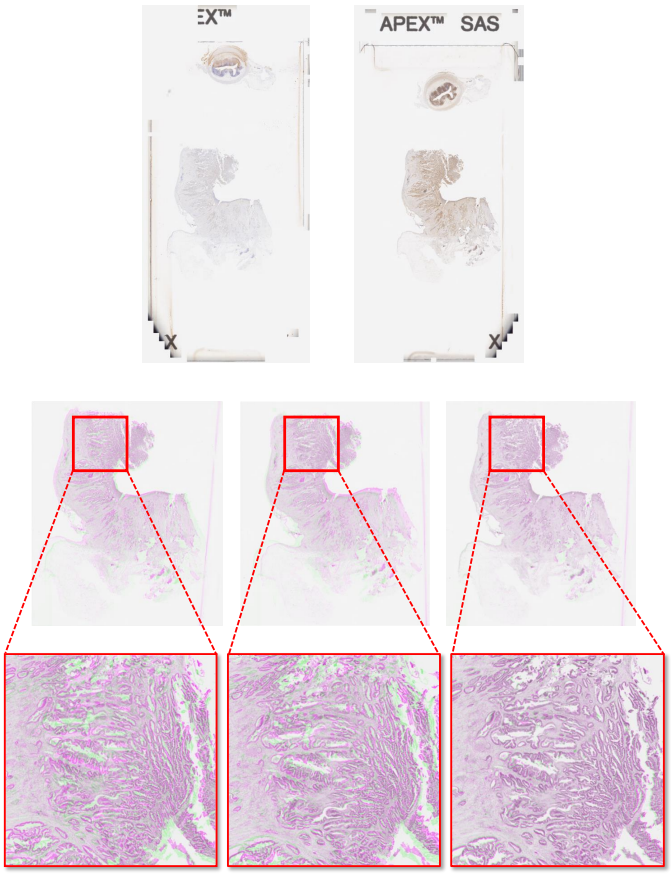

Fig. 10. Demonstration of the efficacy of our proposed rigid registration in a step-wise manner using the overlay of reference and registered moving images. The top row shows reference and pre-aligned moving images. The middle row shows the overlay of corresponding cropped tissue regions. The left overlay image shows the output of the pre-alignment step, the middle overlay image shows the output of the tissue based alignment step and the right overlay image shows the output of the block-wise alignment. The bottom row shows the zoomed-in overlay patches. Reference and registered moving images are shown in green and purple colours, respectively.

The improvement in average rTRE (training set only) after each step of our method is shown in Figure 11. In Table IV, we show values of different metrics computed for the MEVIS and our proposed methods. Our results compare favourably in terms of aggregation average rTRE, median rTRE, maximum rTRE and robustness for the training set. We compare our results to those of the MEVIS team to show the effectiveness of the proposed method.

C. Comparative Results

In Table V, we compare our DFBR method with case-wide registration, proposed in [19]. Trahearn *et al* [19] proposed a novel approach based on MSER features for the alignment of multi-IHC CRC sections. This approach extracts features for each MSER detected from the preprocessed H stain channel. These features are utilised for finding the best pair of MSERs and are used as corresponding control points for estimating rigid transformation. The authors employed this approach for finding the optimal order of the sections for estimating the

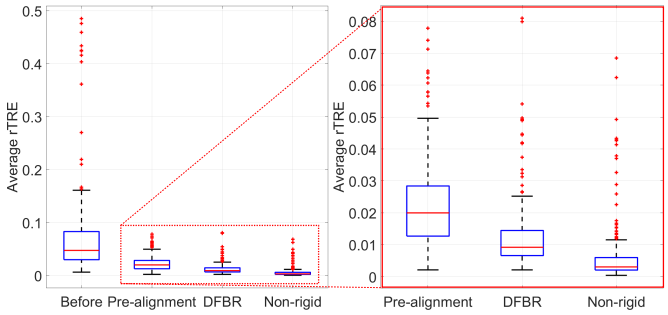

Fig. 11. Demonstration of the average rTRE before and after each registration module using box plots. The average rTRE is computed for the ANHIR training set only. The box plots for the median rTRE are shown in Supplementary Figure 6.

TABLE IV
QUANTITATIVE RESULTS OF THE MEVIS GROUP AND OUR PIPELINE ON THE ANHIR DATASET. THE FIRST ROW IN THE HEADER REPRESENTS THE AGGREGATION METHOD FOR AN IMAGE PAIR AND THE SECOND ROW REPRESENTS THE AGGREGATION METHOD FOR ALL PAIRS IN A SET.

Method	Average rTRE		Median rTRE		Max rTRE		Average
	Average	Median	Average	Median	Average	Median	Robustness
MEVIS_Train	**0.0061**	0.0030	0.0049	0.0019	**0.0271**	**0.0183**	**0.9806**
Ours_Train	0.0065	0.0030	0.0049	0.0019	0.0300	0.0207	0.9688
MEVIS_Eval	**0.0044**	**0.0027**	**0.0029**	0.0018	**0.0251**	**0.0188**	**0.9880**
Ours_Eval	0.0046	0.0028	0.0031	**0.0017**	0.0252	0.0197	0.9842
MEVIS_All	**0.0052**	0.0029	**0.0039**	0.0018	**0.0261**	**0.0186**	**0.9845**
Ours_All	0.0055	0.0029	0.0040	0.0018	0.0275	0.0203	0.9768

transformation of the whole stack. We present comparative results for three different settings: 1) IHC vs IHC, 2) H&E vs IHC and 3) the combination of the first two. Typically, tissue sections of around 3-5 microns in thickness are sliced from the tissue block. The exact thickness of the tissue section could not be retrieved for the COMET sections. However, if we consider it to be 5 microns then the spatial distance between the H&E and IHC slides would range between 30-50 microns whereas for IHC images it would be 5 to 15 microns, making the registration of the H&E and IHC slides challenging not just because of staining differences but also due to the morphological differences between them. Our DFBR approach outperforms the case-wide approach [19] in both settings. Most importantly, the results demonstrate that our DFBR approach can align images even when the tissue structures vary significantly between two images.

In terms of processing time, the DFBR approach is comparable to the MSER feature based approach on registering pair of images rather than the whole stack. For any image pair, the DFBR approach takes a consistent amount of time with a mean processing time of 14 seconds within 1 standard deviation across 105 pairs of images. The consistency is because there are a fixed number of feature points against which feature matching is performed. Whereas, the processing time for the pairwise MSER approach is dependent on the number of MSERs detected. In our experiments with 105 pairs of images, it took 17$\pm$13 seconds. Similarly, processing time varies for

case-wide registration; on average it takes around 5 minutes for transforming the whole stack of images [19].

TABLE V
COMPARATIVE RESULTS OF REGISTRATION USING AVERAGE AGGREGATION FOR CASE-WIDE [19] AND DFBR APPROACHES. THIS COMPARATIVE ANALYSIS IS PRESENTED FOR TWO DIFFERENT SETS: IHC VS IHC AND H&E VS IHC. RESULTS USING MEDIAN AGGREGATION ARE PRESENTED IN SUPPLEMENTARY TABLE 4.

		Average(rTRE)		Max(rTRE)	Robustness
		Average	Median	Average	
IHC vs IHC	Case-wide	0.0057	0.0036	0.0114	0.9801
	DFBR	**0.0041**	**0.0031**	**0.0092**	**0.9947**
H&E vs IHC	Case-wide	0.0152	**0.0115**	0.0381	0.9529
	DFBR	**0.0138**	0.0132	**0.0353**	**1**
All	Case-wide	0.0087	0.0054	0.0200	0.9715
	DFBR	**0.0073**	**0.0042**	**0.0177**	**0.9965**

VII. CONCLUSION

In this paper, we presented a deep feature matching approach which is shown to outperform the hand-crafted feature based approach. In the DFBR framework, we introduced a pre-alignment step which produces a roughly aligned image pair. Tissue segmentation is an essential pre-processing step to exclude feature points from the texture sparse region. Our experiments with deep features showed that a good alignment is difficult to obtain in situations when the slide has more fatty tissue. Therefore, a good tissue segmentation which considers the fatty region as a background is required.

In a digital stack of tissue sections, it is likely that we will observe differences in the tissue structures across the whole stack. This is due to the thickness of each section in relation to the size of the tissue structures. Smaller structures like cells are unlikely to be seen between the consecutive sections and as the distance between the two sections increases, even some bigger tissue structures (such as glands) will not persist. A registration method should be able to perform alignment in the presence of architectural differences between the two sections. Our deep feature method has been shown to outperform an existing MSER based case-wide method [19] when the distance between image pairs is around 50 microns.

ACKNOWLEDGEMENT

We would like to thank and acknowledge David Epstein and Muhammad Shaban for their input in the early stage of this project and David Epstein for proofreading the manuscript.

COMPETING INTEREST STATEMENT

R.A. was funded by the Chancellor's International Scholarship of the University of Warwick. S.R. and N.R. were partly funded by the PathLAKE programme. N.R. is the co-founder of Histofy Ltd, a computational pathology startup company. All the other authors declare no conflict of interest.

REFERENCES

[1] N. Trahearn, D. Epstein, I. Cree, D. Snead, and N. Rajpoot, "Hyper-stain inspector: a framework for robust registration and localised co-expression analysis of multiple whole-slide images of serial histology sections," *Scientific Reports*, vol. 7, no. 1, pp. 1–13, 2017.

[2] T. Hassan, S. Javed, A. Mahmood, T. Qaiser, N. Werghi, and N. Rajpoot, "Nucleus classification in histology images using message passing network," *Medical Image Analysis*, p. 102480, 2022.

[3] D. Tellez, M. Balkenhol, I. Otte-Höller, R. van de Loo, R. Vogels, P. Bult, C. Wauters, W. Vreuls, S. Mol, N. Karssemeijer, *et al.*, "Whole-slide mitosis detection in h&e breast histology using phh3 as a reference to train distilled stain-invariant convolutional networks," *IEEE transactions on medical imaging*, vol. 37, no. 9, pp. 2126–2136, 2018.

[4] M. Rusu, H. Wang, T. Golden, A. Gow, and A. Madabhushi, "Multiscale multimodal fusion of histological and mri volumes for characterization of lung inflammation," in *Medical Imaging 2013: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 8672, pp. 223–231, SPIE, 2013.

[5] Y. Song, D. Treanor, A. J. Bulpitt, and D. R. Magee, "3d reconstruction of multiple stained histology images," *Journal of pathology informatics*, vol. 4, no. Suppl, 2013.

[6] Y. Xu, J. G. Pickering, Z. Nong, E. Gibson, J.-M. Arpino, H. Yin, and A. D. Ward, "A method for 3d histopathology reconstruction supporting mouse microvasculature analysis," *PloS one*, vol. 10, no. 5, p. e0126817, 2015.

[7] R. Awan, M. Nimir, S. E. A. Raza, J. Lotz, D. Snead, A. Robison, and N. M. Rajpoot, "Deep learning based prediction of msi in colorectal cancer via prediction of the status of mmr markers," *arXiv preprint arXiv:2203.00449*, 2022.

[8] G. Bartoli, "Image registration techniques: a comprehensive survey," *Visual Information Processing and Protection Group*, pp. 1–54, 2007.

[9] M. Abdel-Basset, A. E. Fakhry, I. El-Henawy, T. Qiu, and A. K. Sangaiah, "Feature and intensity based medical image registration using particle swarm optimization," *Journal of medical systems*, vol. 41, no. 12, pp. 1–15, 2017.

[10] J. Lotz, N. Weiss, and S. Heldmann, "Robust, fast and accurate: a 3-step method for automatic histological image registration," *arXiv preprint arXiv:1903.12063*, 2019.

[11] J. Lotz, N. Weiss, J. van der Laak, *et al.*, "High-resolution image registration of consecutive and re-stained sections in histopathology," *arXiv preprint arXiv:2106.13150*, 2021.

[12] J. Borovec, J. Kybic, I. Arganda-Carreras, D. V. Sorokin, G. Bueno, A. V. Khvostikov, S. Bakas, I. Eric, C. Chang, S. Heldmann, *et al.*, "Anhir: automatic non-rigid histological image registration challenge," *IEEE transactions on medical imaging*, vol. 39, no. 10, pp. 3042–3052, 2020.

[13] G. Wu, M. Kim, Q. Wang, Y. Gao, S. Liao, and D. Shen, "Unsupervised deep feature learning for deformable registration of mr brain images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 649–656, Springer, 2013.

[14] G. Wu, M. Kim, Q. Wang, B. C. Munsell, and D. Shen, "Scalable high-performance image registration framework by unsupervised deep feature representations learning," *IEEE transactions on biomedical engineering*, vol. 63, no. 7, pp. 1505–1516, 2015.

[15] R. Awan and N. Rajpoot, "Deep autoencoder features for registration of histology images," in *Annual Conference on Medical Image Understanding and Analysis*, pp. 371–378, Springer, 2018.

[16] M. Schwier, T. Böhler, H. K. Hahn, U. Dahmen, and O. Dirsch, "Registration of histological whole slide images guided by vessel structures," *Journal of pathology informatics*, vol. 4, no. Suppl, 2013.

[17] C.-W. Wang, S.-M. Ka, and A. Chen, "Robust image registration of biological microscopic images," *Scientific reports*, vol. 4, no. 1, pp. 1–12, 2014.

[18] N. Trahearn, D. Epstein, D. Snead, I. Cree, and N. Rajpoot, "A fast method for approximate registration of whole-slide images of serial sections using local curvature," in *Medical Imaging 2014: Digital Pathology*, vol. 9041, p. 90410E, International Society for Optics and Photonics, 2014.

[19] "Registration and multi-immunohistochemical analysis of whole slide images of serial tissue sections." url=http://webcat.warwick.ac.uk/record=b3071272 S15. Accessed: 2021-10-20.

[20] L. Solorzano, G. M. Almeida, B. Mesquita, D. Martins, C. Oliveira, and C. Wählby, "Whole slide image registration for the study of tumor heterogeneity," in *Computational pathology and ophthalmic medical image analysis*, pp. 95–102, Springer, 2018.

[21] J. Borovec, A. Munoz-Barrutia, and J. Kybic, "Benchmarking of image registration methods for differently stained histological slides," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 3368–3372, IEEE, 2018.
[22] M. Wodzinski and A. Skalski, "Multistep, automatic and nonrigid image registration method for histology samples acquired using multiple stains," *Physics in Medicine & Biology*, vol. 66, no. 2, p. 025006, 2021.
[23] M. Jaderberg, K. Simonyan, A. Zisserman, *et al.*, "Spatial transformer networks," *Advances in neural information processing systems*, vol. 28, pp. 2017–2025, 2015.
[24] C. Shu, X. Chen, Q. Xie, and H. Han, "An unsupervised network for fast microscopic image registration," in *Medical Imaging 2018: Digital Pathology*, vol. 10581, p. 105811D, International Society for Optics and Photonics, 2018.
[25] S. Zhao, T. Lau, J. Luo, I. Eric, C. Chang, and Y. Xu, "Unsupervised 3d end-to-end medical image registration with volume tweening network," *IEEE journal of biomedical and health informatics*, vol. 24, no. 5, pp. 1394–1404, 2019.
[26] M. Wodzinski and H. Müller, "Learning-based affine registration of histological images," in *International Workshop on Biomedical Image Registration*, pp. 12–22, Springer, 2020.
[27] D. Mahapatra, "Registration of histopathogy images using structural information from fine grained feature maps," *arXiv preprint arXiv:2007.02078*, 2020.
[28] 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.
[29] M. Wodzinski and H. Müller, "Unsupervised learning-based nonrigid registration of high resolution histology images," in *International Workshop on Machine Learning in Medical Imaging*, pp. 484–493, Springer, 2020.
[30] Z. Yang, T. Dan, and Y. Yang, "Multi-temporal remote sensing image registration using deep convolutional features," *Ieee Access*, vol. 6, pp. 38544–38555, 2018.
[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[32] A. Araujo, W. Norris, and J. Sim, "Computing receptive fields of convolutional neural networks," *Distill*, vol. 4, no. 11, p. e21, 2019.
[33] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
[34] J. Pocock, S. Graham, Q. D. Vu, M. Jahanifar, S. Deshpande, G. Hadjigeorghiou, A. Shephard, R. M. S. Bashir, M. Bilal, W. Lu, *et al.*, "Tiatoolbox as an end-to-end library for advanced tissue image analytics," *Communications medicine*, vol. 2, no. 1, pp. 1–14, 2022.
[35] J. Pichat, J. E. Iglesias, T. Yousry, S. Ourselin, and M. Modat, "A survey of methods for 3d histology reconstruction," *Medical image analysis*, vol. 46, pp. 73–105, 2018.
[36] E. Haber and J. Modersitzki, "Intensity gradient based registration and fusion of multi-modal images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 726–733, Springer, 2006.
[37] Y. Jianchao, "Image registration based on both feature and intensity matching," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)*, vol. 3, pp. 1693–1696, IEEE, 2001.
[38] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *International Journal of Computer Vision*, vol. 129, no. 1, pp. 23–79, 2021.