DeepHistReg: Unsupervised Deep Learning Registration Framework for Differently Stained Histology Samples

Marek Wodzinski*

AGH University of Science and Technology
Department of Measurement and Electronics
Kraków, Poland
wodzinski@agh.edu.pl

Henning Müller

University of Applied Sciences Western Switzerland (HES-SO Valais)
Information Systems Institute
Sierre, Switzerland
henning.mueller@hevs.ch

Abstract

Background and objective

The use of several stains during histology sample preparation can be useful for fusing complementary information about different tissue structures. It reveals distinct tissue properties that combined may be useful for grading, classification, or 3-D reconstruction. Nevertheless, since the slide preparation is different for each stain and the procedure uses consecutive slices, the tissue undergoes complex and possibly large deformations. Therefore, a nonrigid registration is required before further processing. The nonrigid registration of differently stained histology images is a challenging task because: (i) the registration must be fully automatic, (ii) the histology images are extremely high-resolution, (iii) the registration should be as fast as possible, (iv) there are significant differences in the tissue appearance, and (v) there are not many unique features due to a repetitive texture.

Methods

In this article, we propose a deep learning-based solution to the histology registration. We describe a registration framework dedicated to high-resolution histology images that can perform the registration in real-time. The framework consists of an automatic background segmentation, iterative initial rotation search and learning-based affine/nonrigid registration.

*Corresponding author

Preprint submitted to Computer Methods and Programs in Biomedicine September 22, 2021

Results

We evaluate our approach using an open dataset provided for the *Automatic Non-rigid Histological Image Registration* (ANHIR) challenge organized jointly with the IEEE ISBI 2019 conference. We compare our solution to the challenge participants using a server-side evaluation tool provided by the challenge organizers. Following the challenge evaluation criteria, we use the target registration error (TRE) as the evaluation metric. Our algorithm provides registration accuracy close to the best scoring teams (median rTRE 0.19% of the image diagonal) while being significantly faster (the average registration time is about 2 seconds).

Conclusions

The proposed framework provides results, in terms of the TRE, comparable to the best-performing state-of-the-art methods. However, it is significantly faster, thus potentially more useful in clinical practice where a large number of histology images are being processed. The proposed method is of particular interest to researchers requiring an accurate, real-time, nonrigid registration of high-resolution histology images for whom the processing time of traditional, iterative methods in unacceptable. We provide free access to the software implementation of the method, including training and inference code, as well as pretrained models. Since the ANHIR dataset is open, this makes the results fully and easily reproducible.

Keywords: ANHIR, Histology, Image Registration, Deep Learning

1. Introduction

The registration of histology images acquired using several stains is an important problem in a period where pathology departments in hospitals are becoming increasingly digital. Fusing information available in structures revealed by different dyes from consecutive slices may be useful for segmentation, classification, grading, or 3-D reconstruction. However, the registration of differently stained histology images is challenging, because of: (i) the necessity of fully automatic registration, (ii) the extremely large size of histopathology images, (iii) real-time requirements in many clinical situations, (iv) the problem of missing data due to differences in the tissue appearance, and (v) the lack of unique features due to a repetitive texture in these large images [1].

The first two challenges are particularly difficult for classical, iterative, nonrigid image registration methods [2]. These algorithms optimize an enormous number of parameters, especially considering the high-resolution of histology images. This results in long analysis time, thus lowering the clinical usefulness. This can be efficiently addressed by deep learning-based registration [3]. In deep learning registration, most of the computational time is being spent on training, which can be done offline using a server dedicated to machine learning. As a

result, during inference, deep learning enables a real-time nonrigid registration, which is crucial to clinical practice.

The importance of the histology registration was a motivation to organize an open challenge called *Automatic Non-rigid Histological Image Registration* (ANHIR) [1, 4, 5]. It was organized jointly with the IEEE ISBI 2019 conference (International Symposium on Biomedical Imaging) and more than 250 participants from more than 30 countries registered for the competition. Surprisingly, almost all of the best scoring methods were based on the classical, iterative image registration resulting in the long time required for the analysis [1]. Even though the registration accuracy of the proposed methods is close to the level of the human annotation, the computational time is relatively high, thus limiting their clinical usefulness. We assume that the majority of the challenge participants did not decide to use a deep learning approach because of the high-resolution of histology images, making them difficult to register due to the GPU memory constraints (most GPUs have a maximum of 16-32 GB of RAM). This motivated us to propose a deep learning-based registration framework addressing the challenges and enabling a real-time nonrigid registration of histology samples.

1.1. Related Work

Medical image registration is a mature field with hundreds of important contributions [2, 6]. Nevertheless, due to the challenges related to the registration of histology images, the general state-of-the-art algorithms often fail and have limited usefulness. This observation was confirmed by the ANHIR organizers [5]. They evaluated algorithms like bUnwarpJ [7], NiftyReg [8], RVSS [7], ANTs [9], DROP [10] or Elastix [11] and showed rather poor results comparing to the dedicated algorithms [1].

Nonetheless, there are contributions focused directly on the histology registration. In [12] an interesting concept about intensity-based registration driven by unsupervised classification of the structural similarity of histology images was proposed. Unfortunately, the authors did not take part in the ANHIR challenge. Another successful contribution was proposed in [13] by Fraunhofer MEVIS where the authors introduced a registration framework consisting of initial rotation search, iterative affine, and B-Spline-based nonrigid registration using the normalized gradient fields (NGF) as the similarity metric [14]. The method was the winner of the ANHIR challenge and compared to other classical registration approaches was amazingly fast due to a well-optimized, commercial implementation. Other methods dedicated to histology registration were introduced by researchers from the University of Pennsylvania (UPENN) [15] and the AGH University (AGH) [16]. These teams achieved the second and third best scores in the ANHIR challenge. The first approach is based on a proper preprocessing consisting of background removal by deconvolution, followed by random initial alignment, affine registration, and diffeomorphic registration by the "Greedy" tool [17, 18]. The AGH team proposed a method similar to the MEVIS team in terms of a multistep approach, with the main difference related to the nonrigid registration. Instead of B-Splines NGF-based registration they

proposed a Demons-based solution using the modality independent neighborhood descriptor (MIND) as the cost function [19, 20]. The main disadvantage of the method was the long time required for the registration due to the unoptimized CPU implementation. Only one of the best scoring teams decided to use a deep learning-based solution. The TUB team [21] used a modified volume tweening network. First, they resampled the images to relatively low resolution and second, they tuned the network by using manually defined landmarks resulting in a significant difference between the training and evaluation sets [1]. Moreover, their method could not be considered as fully automatic since in practice it would require many manual annotations, making it difficult to use in clinical practice. Nevertheless, considering the relatively low inference time of their method, it inspired us to propose a deep learning-based framework. The goal was to propose a fully automatic method that accurately registers the histology images in real-time.

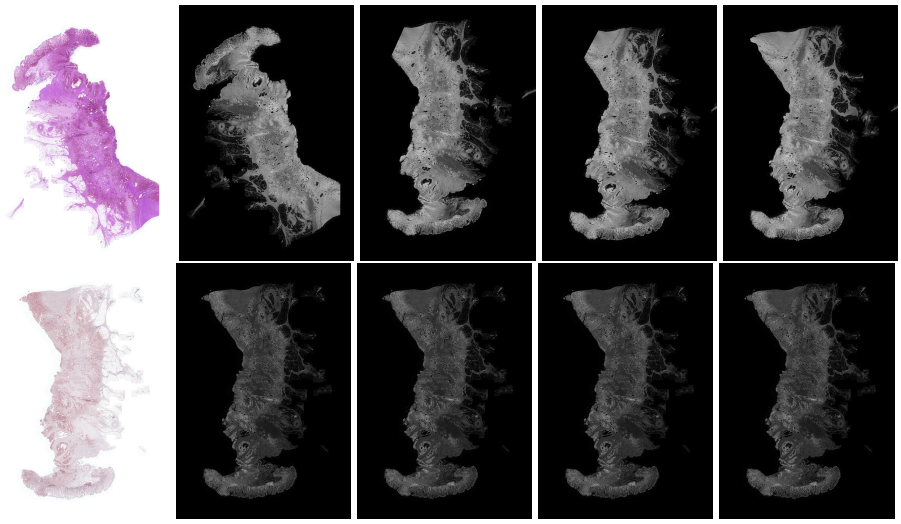
The deep learning-based medical image registration is a relatively new field [3, 22]. It immediately turned out to be useful because of the low time required during the model inference enabling real-time nonrigid registration. This is crucial, e.g. for registration during surgical interventions. The deep registration approaches can be divided into three main categories based on the training scheme: (i) supervised [23, 24], (ii) unsupervised [25, 26, 27], and (iii) adversarial registration [28, 29]. The supervised registration requires ground-truth deformation fields or pre-aligned images that are often impossible to obtain. The adversarial approach, which is based on generator and discriminator networks, suffers from similar limitations. Moreover, the adversarial networks are not trivial to train [30]. To train the discriminator, registered image pairs representing the ground-truth alignment are necessary. For some tasks it can be achieved [28], but it is usually costly and time-consuming. On the other hand, unsupervised methods do not require any ground-truth. They are based on minimization of a given cost function and the registration accuracy mostly depends on: (i) a proper choice of the similarity measure, (ii) a regularization term enforcing plausible deformations, (iii) the ability to converge during training, and (iv) a generalization ability. The unsupervised registration can be seen as a way to speed-up the classical, iterative image registration. Since the methods proposed by the ANHIR challenge participants (e.g. MEVIS, UPENN, AGH) achieved results close to the human annotation error [1], we decided to focus on the unsupervised registration, addressing the difficulties connected with the high-resolution of histology images and the required robustness while providing the real-time alignment.

The main challenge with the learning-based registration of histology images is connected with the high resolution of these images, coupled with large and complex deformations. The deep learning methods suffer from large GPU memory utilization. The higher the image resolution, the larger the necessary receptive field and the required GPU memory. The simplest solution is to downsample the images. This approach was used by the TUB team to apply the volume tweening network during the ANHIR challenge [1, 21]. However, downsampling the images reduces the registration quality and makes it harder

to register fine details, e.g. the TUB team downsampled the images from resolution above 15000x15000 to resolution below 1000x1000 to use learning-based approach [1]. In the work by de Vos *et al.* [25] the transformation was parameterized by the B-Splines transformation model to reduce the decoder memory footprint. It decreases the memory consumption and still, it is not applicable to histology images. In the work by Fan *et al.* [28] a simple patch-based approach was applied. In this approach, the images are divided into patches to reduce the number of parameters and the required GPU memory. This approach reduces the maximum recoverable deformations to a fraction of the patch size and does not solve the problem of high-resolution histology images because, even after unfolding, the images do not fit into the GPU memory. Another interesting approach was proposed by Heinrich and Hansen [31] for unsupervised learning-based registration of CT volumes. They iteratively subdivided the 3-D transformation space into orthogonal planes resulting in 2.5-D displacement search. The method not only significantly improved the registration accuracy but also decreased the GPU memory consumption and inference time. In the proposed method, we apply a pyramid, patch-based approach together with a sequential transfer of the batches to the GPU memory. This approach not only makes it possible to process images of any resolution, but also addresses the limitation of patch-based approaches related to maximum recoverable deformations. What is more, training a network with a low number of parameters on numerous small patches reduces the risk of overfitting.

1.2. Contribution

In this work, we propose an unsupervised deep learning-based registration framework dedicated to histology images acquired using different stains. The method is robust to different dyes and tissue types within the evaluated dataset, fully automatic and provides results comparable to the state-of-the-art methods while being significantly faster, which is important in clinical practice. The main technical novelty is related to the ability to recover large, complex, deformations between high-resolution images (even above 15000x15000) in real-time. The method is built on our previous contributions [32, 33], however, the methods are greatly extended and improved. An additional learning-based background segmentation was introduced, significantly improving the initial alignment. Moreover, the framework user can now decide which affine registration is suitable for the given problem. Finally, the nonrigid registration is improved by changing the order of the the unfold/fold operations, the velocity field composition, and the image transformation. These changes eliminated the problem of inconsistency at patch boundaries. We provide free access to the framework source code [34], including training/inference scripts and pre-trained models. This, together with open access to the ANHIR dataset, makes the results fully and easily reproducible. Moreover, the framework was evaluated using the independent ANHIR evaluation platform. The results can easily be verified and compared using the challenge website [4]. The proposed framework is easily extendable making it possible to further improve the registration by other researchers.



a) Original b) Preprocessed c) Prealigned d) Affinely Registered e) Final result

Figure 1: The source and target at different registration stages (top and bottom row respectively, pair 178). The target is replicated for presentation clarity. High quality picture, best viewed zoomed in electronic format.

2. Methods

2.1. Overview

We propose an unsupervised deep learning-based registration framework. The framework pipeline consists of data loading, transferring to GPU, preprocessing, initial alignment, affine registration, and finally nonrigid registration. All the steps are described in detail in the following sections. The negative normalized cross-correlation (NCC) is used as the cost function during all registration steps. The global NCC is used for initial alignment and affine registration, while the patch-based NCC is used for the nonrigid registration. The curvature (CURV) is used as the regularization term for the nonrigid registration:

$$S(M, F, u) = -\text{NCC}(M, F) + \alpha \text{CURV}(u) \rightarrow min, \tag{1}$$

where NCC is the normalized cross-correlation (global or local version, depending on context, CURV denotes the curvature regularization (only for the nonrigid registration), α is the regularization parameter controlling the deformation smoothness, and M, F, u are respectively the warped moving patches, target patches and the displacement fields.

The detailed framework structure and pipeline is presented in Figure 2. The visualization of an example registration pair after each registration step is shown in Figure 1.

2.2. Preprocessing

The preprocessing consists of offline and online stages. In the offline stage, all the image pairs are padded and parsed from .jpg/.png into uncompressed .mha format to speed-up the data loading. This is quite important because it decreases the time needed for the data loading during the training and inference.

The pipeline starts from the image pair loading, converting both images to grayscale and transferring them to GPU memory. After this, the images are downsampled (preceded by smoothing by Gaussian filtering) to a relatively low resolution (512 pixels in the smaller dimension). The downsampled image pairs are used during the background segmentation, initial alignment, and affine registration. For these steps, the use of a higher resolution is not mandatory and lowering the resolution decreases the registration time.

Then, the tissues are segmented from the background by a U-Net-based [35] network. This is not a demanding task but significantly improves the registration results for image pairs with background artifacts (e.g. mammary glands in the used dataset). This step could be done differently (e.g. by color deconvolution as in [15]) but we decided that deep segmentation is fast, robust, and easily convertible to other histology datasets. The visualization of an example source-target pair after the preprocessing is shown in Figure 1b. The background segmentation is negligible in terms of the computational time (a few milliseconds).

2.3. Initial Alignment

The goal of the initial alignment step is to perform very fast but not to have an accurate rigid registration. The initial alignment presented here is not deep learning-based since it is optimizing just one parameter, the rotation angle. Still, it is implemented using the GPU to lower the registration time. The initial alignment is not mandatory for all cases since pairs with smaller global deformations can be handled well by the following affine registration. However, for cases misaligned by e.g. 180 degree rotation this step is crucial.

The initial alignment begins with the calculation of the source and target centroids. Then, the source is translated by the translation vector between the centroids. This is followed by an exhaustive rotation angle search with a predefined step. The source is being warped for each angle and the NCC is being calculated. The angle with the lowest negative NCC is chosen. The final rigid transformation is a composition of the centroid translation vector and the rotation matrix. The visualization of images after the initial alignment is shown in Figure 1c.

One can argue that this step could also be handled by the deep network. It is true but the entire procedure optimizes only a single parameter and enforcing deep solutions to this simple problem is an exaggeration. Moreover, the proposed procedure is much faster than the following nonrigid registration. Proposing a dedicated deep network would not speed-up the registration significantly (especially considering the large receptive field that would be required by the initial alignment deep network).

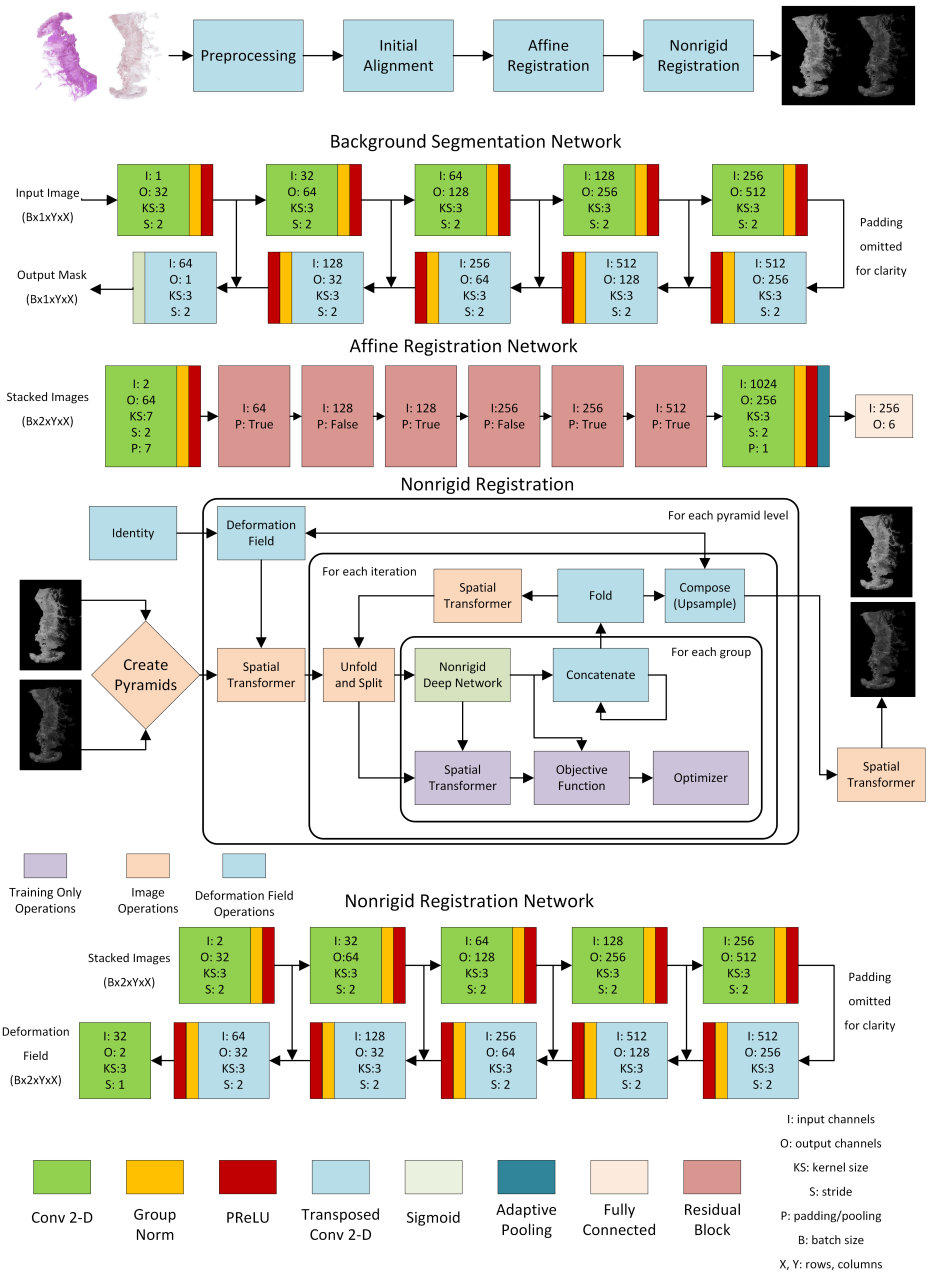


Figure 2: Visualization of the proposed deep histology registration framework.

2.4. Affine Registration

The affine registration is done by a relatively simple ResNet-like [36] convolutional neural network, as in Figure 2. The resolution of the input images is the same as in the initial alignment step. The network output is an affine transformation matrix (2x3) that is then converted to the transformation grid used in the spatial transformer. The network was trained using negative NCC as the cost function. The dataset was augmented by random affine transformations applied randomly to both the source and target images. Alternatively to the proposed network, the framework offers also the affine registration network described in [33] that registers images at a higher resolution. The visualization of images after the affine registration is shown in Figure 1d.

2.5. Nonrigid Registration

The nonrigid registration is the most difficult step in the histology registration. It is impossible to achieve accurate registration preserving the fine details using the simple, single-shot networks because the parameter gradients do not fit into the GPU memory due to the high resolution. Moreover, even a common patch-based approach to reduce the problem into smaller patches that are then combined in the batch dimension is not enough since the batches would not fit in the GPU memory too.

Therefore, we propose a pyramid-based, patch-based, group-based, and iterative deep registration solution [32]. The pyramid-based approach means that the images are registered at different resolutions, starting at the coarsest level. Then, after a given pyramid level, the calculated deformation fields are upsampled to the next resolution, as in the iterative methods. Patch-based means that at the given resolution the images are unfolded into smaller patches that can be handled by a relatively small deep network. Group-based means that only small groups of patches are propagated by the network at once due to the GPU memory constraints and the loss function is being evaluated and optimized at the group level, not at the image level. Finally, the method is iterative since at each pyramid level the images are propagated through the network several times, progressively composing the calculated velocity fields.

The nonrigid registration procedure can be summarized as follows. To start with, resolution pyramids are built for both the source and target images. Then, starting at the coarsest resolution, for a given number of iterations the images are being registered. First, the source image is warped using the current deformation field, starting with the identity transformation. Then, in each iteration, the images are unfolded into overlapping patches that are then split into groups with predefined size (limited by the GPU memory). The stride of overlapping patches is half of the patch size. This slightly increases the registration time. However, it mitigates the problem of deformation field discontinuities at the patch boundaries. Each corresponding group is propagated through the registration network, calculating the current velocity field. During training, the group is transformed and the cost function is calculated. The cost function is the sum of the negative NCC and curvature regularization term [37]. The

calculated velocity fields for each group are concatenated. After all the groups are processed, the concatenated velocity fields are folded back into the velocity field with the same shape as the current deformation field. The current deformation field is composed of the velocity field and used for the next iteration. This makes the interpolation error negligible since the source image is never interpolated more than once. After composing, the current level deformation field is upsampled to the next resolution. The deformation field after the highest resolution becomes the final deformation field. The detailed visualization of the nonrigid registration procedure is shown in Figure 2. An example of nonrigid registration outcome is shown in Figure 1e.

The nonrigid method has several parameters: (i) the patch size, (ii) stride, (iii) group size, (iv) number of pyramid levels, (v) number of iterations per level, and (vi) the regularization parameter. The patch size and the stride are responsible for lowering the number of network parameters and the required receptive field and thus the required GPU memory. The lower the patch size and stride, the larger the available group size, at the cost of decreasing the maximum magnitude of deformations. The number of pyramid levels is responsible for ensuring that the given patch size is able to capture sufficiently large deformations. Since the patch size is constant for all pyramid levels, the lower resolution is able to capture larger deformations, as in the traditional, iterative techniques. The number of iterations per level defines how many times the patches are passed through the network at each resolution. This increases the registration accuracy at the cost of increasing the registration time. Finally, the regularization parameter is responsible for controlling the deformation smoothness.

2.6. Technical Details

The framework is implemented using PyTorch [38]. All the registration steps, excluding data loading, are implemented on the GPU. The network was trained using GTX RTX 2080 Ti. The Adam optimizer was used for all registration stages together with exponentially decaying learning rate schedulers. The framework structure is easily extendable. One can easily replace a given registration or preprocessing step by another method. This makes it possible to e.g. propose an alternative nonrigid registration network, introduce different similarity measures or regularization terms, without worrying about the preprocessing and initial affine registration.

We freely release the framework software, including data parsing scripts, training/inference source code, and pretrained models [34]. We also attach the transformed landmarks (after each registration step) used for evaluation as the supplementary material, making the results verifiable. A guide on how to run the scripts is available in the framework repository.

2.7. Dataset and Experimental Setup

We used the open ANHIR dataset to evaluate the proposed framework [1, 39, 40, 41, 42] and promote research transparency, openness, and reproducibility. The dataset consists of 481 image pairs split into 251 evaluation and 230

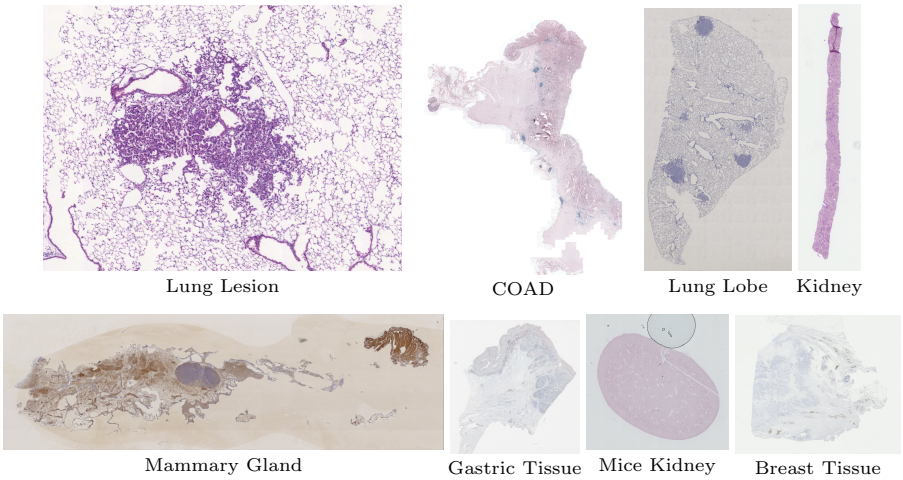


Figure 3: Visualization of different tissue types stained using various dyes [1, 4]. It presents the robustness and generalization ability required by the deep registration framework (best viewed zoomed in electronic format).

training pairs. There are 8 tissue types: (i) mammary glands, (ii) the colon adenocarcinomas (COADs), (iii) gastric mucosa and adenocarcinomas, (iv) breast, (v) mice kidney, (vi) human kidney, (vii) lung lesions, and (viii) lung lobes. The consecutive slices were stained by: (i) prosurfactant protein C, (ii) antigen KI-67, (iii) clara cell 10 protein, (iv) human epidermal growth factor receptor 2, (v) progesterone receptor, (vi) estrogen receptor, (vii) platelet endothelial cell adhesion molecule, (viii) cytokeratin, (ix) hematoxylin and eosin, (x) podocin. Visualization of different tissues stained using distinct dyes is shown in Figure 3. The dataset providers resampled the images to approximately 25% of the full original resolution, resulting in larger size varying from 6k to 17k pixels in one dimension (from 4369x6930 to 17179x15042, the resolution is different for each image). The images are provided as .jpg and .png files without the metadata. However, during the dataset parsing they are converted to the .mha format and during preprocessing to grayscale images. A more detailed description of the dataset and the staining procedure is available in [4].

The dataset was annotated by providing corresponding landmarks. In total, 9 qualified annotators chose on average 86 landmarks per image. The average error between the landmarks chosen by two annotators is 0.05% of the image diagonal. This can be used as the indicator of the human-level accuracy and a threshold below which the registration methods become indistinguishable [1, 4]. The corresponding landmarks are openly available only for the training images. For the evaluation set, only the source image landmarks are released and the evaluation must be done using the server-side evaluation platform developed by the ANHIR organizers with a very limited number of available submissions. This makes the results reliable and trustworthy since the proposed methods cannot

be tuned to the evaluation set. During training, we used only the training pairs without utilizing any information about the manually selected landmarks.

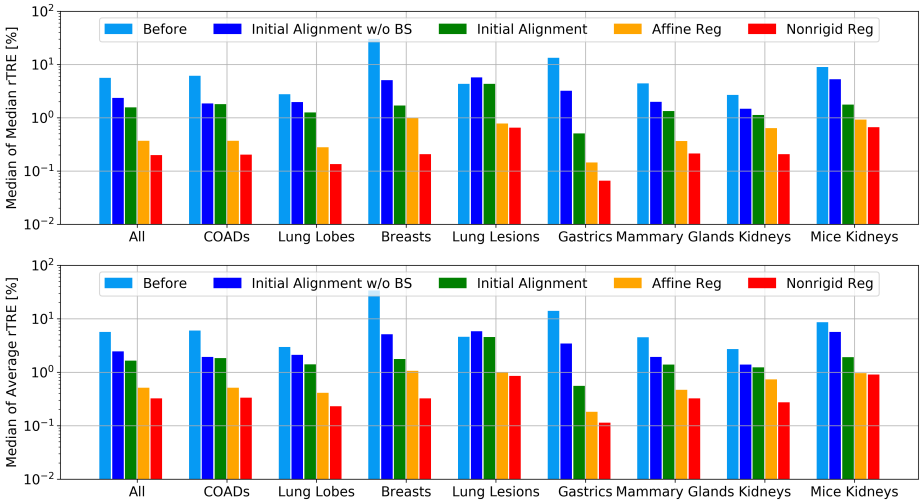


Figure 4: The Median of Median rTRE and Median of Average rTRE reported in % of image diagonal for all registered image pairs. Please note that the tissue types are not equipotent, COADs have the highest influence on the averaged results (logarithmic scale used for presentation clarity). The results originate from the ANHIR evaluation system [4] and are as of 05.08.2020. W/o BS means "without background segmentation".

2.8. Evaluation Criteria

The registration accuracy is measured by the target registration error (TRE) measuring the Euclidean distance between the annotated and transformed landmarks. To make the error comparable between image pairs with different resolutions, the TRE is normalized by the image diagonal:

$$rTRE = \frac{TRE}{\sqrt{w^2 + h^2}}, \tag{2}$$

where TRE denotes the target registration error, w is the image width and h is the image height. There are different rTRE-based metrics: (i) average of median rTRE, (ii) median of median rTRE, (iii) median of average rTRE, (iv) average of average rTRE. For clarity, the average of median rTRE shows the average at the case level of the median rTRE at the landmarks level. The median of median/average rTRE is good for describing the quality of the registration, while the average of median/average can be used to verify potential outliers (e.g. due to the initial alignment fail).

Apart from the rTRE, the robustness and normalized processing time are evaluated. The robustness is defined as the fraction of landmarks for which the rTRE decreased after registration to the total number of landmarks. The processing time reports the total time required for data loading, preprocessing,

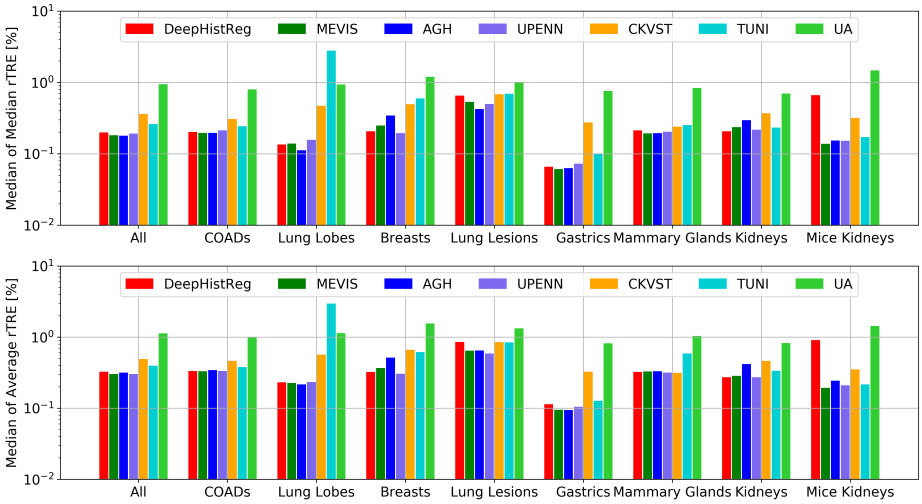


Figure 5: Comparison of the proposed framework to the state-of-the-art nonrigid histology registration methods in terms of Median of Median rTRE and Median of Average rTRE reported in % of the image diagonal, evaluated for all image pairs. The results originate from the ANHIR evaluation system [4]. The results for TUB method are not reported since they used the manually annotated landmarks for training and the results are not comparable (rTRE artificially close to 0% for half of the image pairs). The results are as of 05.08.2020.

and registration (without saving the outcomes). All the values are reported by an independent, server-side evaluation tool.

3. Results

We present the outcomes from subsequent registration steps and compare the final results to the state-of-the-art algorithms. We decided to use the ANHIR evaluation platform and to compare only to the methods submitted there. The reasons for this are: (i) independent, reliable comparison, and (ii) mitigation of the evaluation bias resulting from incorrect parameter tuning. As a result, there is no possibility to artificially improve the outcomes since all method parameters were tuned by their authors or the ANHIR organizers. Therefore, one can assume that everyone did their best to get as good results as possible. Noteworthy, the evaluation landmarks are unavailable and one can perform only a single submission per day, making it difficult to tune the parameters with respect to the evaluation set. We encourage everyone interested in the histology registration to submit their method using the submission system.

In Figure 4, we show the landmark-based evaluation of the subsequent registration stages: (i) initial, (ii) after the initial alignment, (iii), after the affine registration, and finally (iv) after the nonrigid registration. The results are presented together and separately for all the tissues and show the median of median rTRE and median of average rTRE. Please note that the tissue types

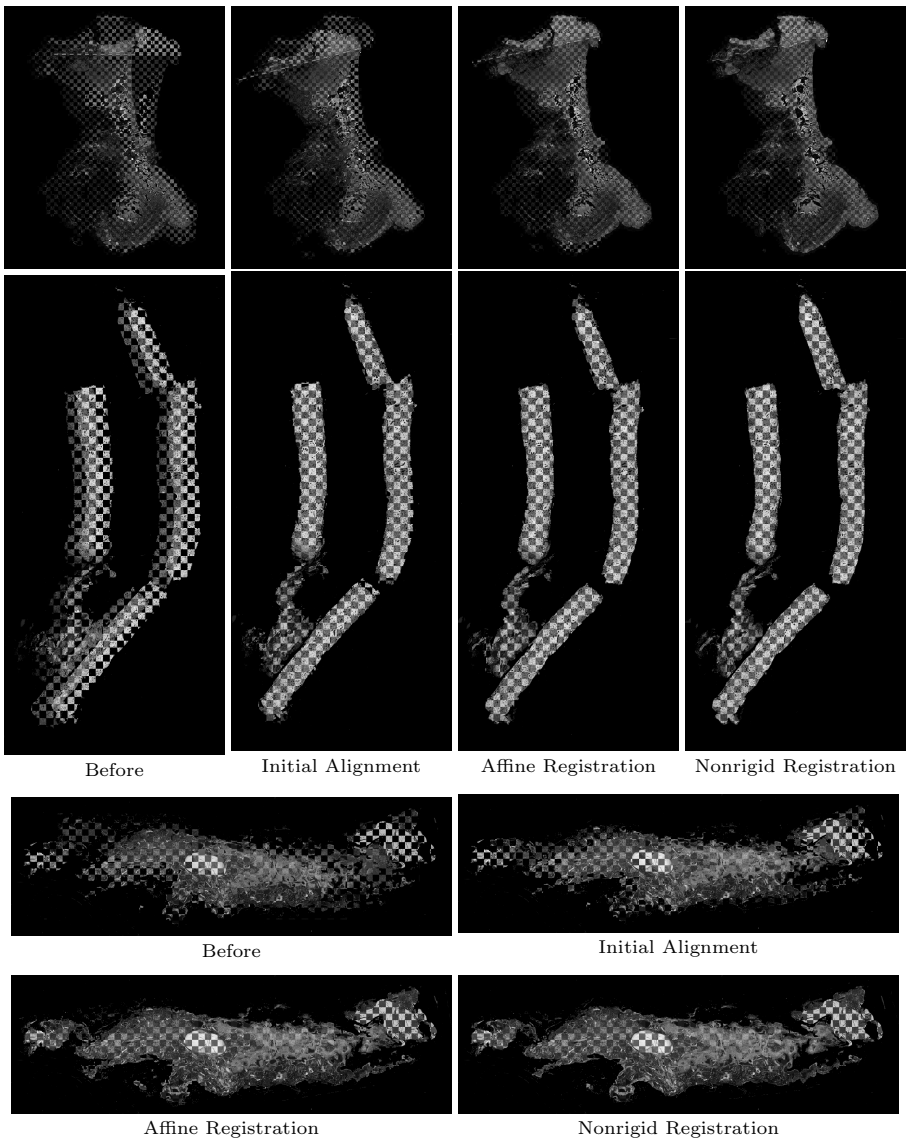
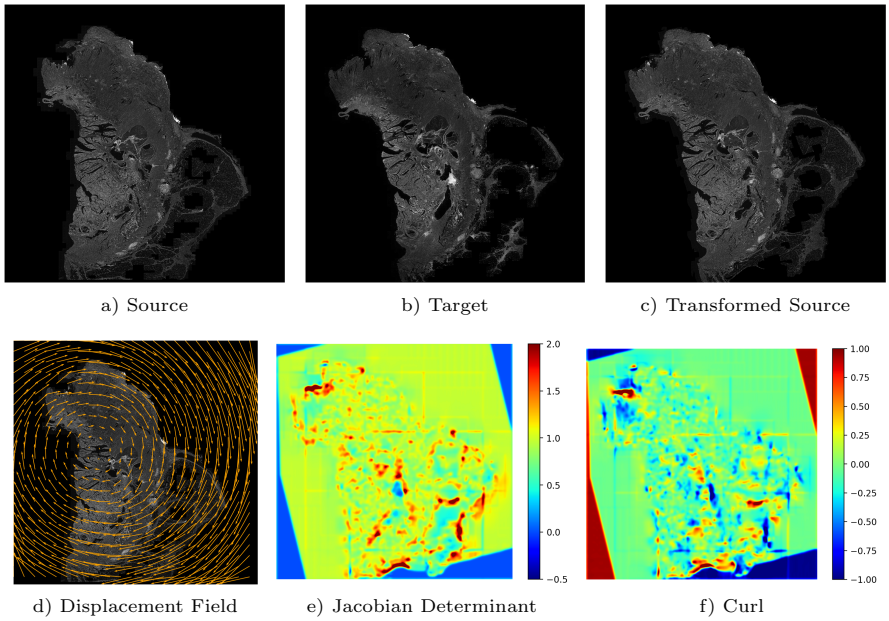


Figure 6: Example checkerboards at different registration stages for pairs no. 13, 317, 430 (COAD, kidney, mammary gland). High quality pictures, best viewed zoomed in electronic format.



a) Source b) Target c) Transformed Source

d) Displacement Field e) Jacobian Determinant f) Curl

Figure 7: An exemplary visualization of the registration results, together with the deformation field, its Jacobian determinant and Curl.

Table 1: Quantitative results summary based on the ANHIR evaluation website comparing our method to the state-of-the-art algorithms [4]. The table presents results only for the evaluation set. The method abbreviations denote the team names, for detailed information about the team members we refer to [1]. The results are as of 05.08.2020.

	Average rTRE		Median rTRE		Max rTRE		Robustness		*Average*
method	Average	Median	Average	Median	Average	Median	Average	Median	*time [min]*
DeepHistReg	*0.0061*	*0.0033*	*0.0047*	*0.0019*	*0.0276*	*0.0224*	*0.9799*	*1.0000*	*0.03*
MEVIS	0.0043	0.0028	0.0028	0.0018	0.0251	0.0188	0.9880	1.0000	*0.14*
AGH	0.0073	0.0032	0.0036	0.0017	0.0290	0.0214	0.9795	1.0000	*8.60*
UPENN	0.0041	0.0029	0.0029	0.0019	0.0238	0.0190	0.9898	1.0000	*1.45*
CKVST	0.0042	0.0027	0.0026	0.0023	0.0239	0.0189	0.9883	1.0000	*7.13*
TUB	0.0089	0.0029	0.0077	0.0021	0.0280	0.0178	0.9845	1.0000	-
TUNI	0.0063	0.0031	0.0048	0.0021	0.0287	0.0204	0.9822	1.0000	*10.32*
UA	0.0536	0.0100	0.0506	0.0082	0.1124	0.0353	0.8209	0.9852	*1.47*

are not equipotent, therefore one tissue type can have a larger influence on the averaged results. We show the results for initial alignment with and without the background removal to show its influence on the registration results. The affine registration and the nonrigid registration are reported only with the background removal. In Figure 6, we present several example checkerboards to visually verify the registration results. In Figure 7, we show an example registration result together with the calculated displacement field, its Jacobian determinant and Curl. Unfortunately, we do not have access to the deformation fields of other methods evaluated on the ANHIR dataset. Thus, quantitative comparison of the deformation complexity (e.g. using standard deviation of Jacobian determinant) is not possible.

To compare the outcomes to the state-of-the-art algorithms, we present an exhaustive comparison in Figure 5. It shows the median of median rTRE and median of average rTRE after the nonrigid registration for all the tissues, compared to the other histology registration methods. All the method statistics originate from the ANHIR evaluation system [4] (as of 05.08.2020). Moreover, we present all the evaluation metrics in Table 1. We must emphasize, that the table is an extended version of the table presented in the ANHIR summary article [1] but with results obtained directly from the evaluation system without any additional outlier rejection and without the rTRE ranking at the case level since it is not being evaluated anymore.

4. Discussion

The presented results show that the proposed method is comparable in terms of the rTRE to the best state-of-the-art algorithms. The results are slightly worse (by about 0.002% of the image diagonal) than the three best state-of-the-art algorithms. However, this is somehow expected and can be justified. In unsupervised deep registration, the network is learning how to minimize a given objective function. The objective function is usually the same or very similar compared to the classical, iterative registration. Nonetheless, in the iterative registration, the objective function is minimized separately for each new case. In deep learning, it could be compared to tuning the network to each new evaluation case. This does not make sense in practice because the main goal of using the unsupervised deep registration is to strongly speed-up the analysis. We hypothesize that for larger datasets the difference between the results would be even smaller. Moreover, the adversarial registration could also be a solution because there would be no need to define the similarity measure and the objective function would be learned. However, ground-truth registrations are necessary and for histology such registrations are very hard to obtain.

The results reported in Figure 4 show that all the subsequent steps increase the registration quality. The initial alignment significantly improves the registration for all tissues except the lung lesions that do not require any initial rotation correction. It is crucial for gastric and breast tissues for which the initial alignment alone decreases the rTRE by an order of magnitude. Noteworthy, the necessity of this step depends on the dataset. For other histology

datasets not containing consecutive slices rotated by e.g. 180 degrees, this may be unnecessary. The affine registration decreases the rTRE for all tissue types and the computational time required by this step (on average 4 ms) is negligible. A correct affine registration is a requirement for the following nonrigid registration that further improves the registration. Compared to the initial or affine alignment, the influence on the rTRE is not that important. However, the nonrigid registration is crucial for registering the fine details, as presented in Figure 6.

The results of the proposed framework in terms of rTRE are comparable to the best performing methods for all tissue types except the mice kidneys. This is the case because there is only a single mice kidney tissue in the ANHIR dataset and the nonrigid deep network cannot learn from such a small amount of data. Interestingly, for kidneys the proposed method is the most accurate and for COADs, lung lobes, mammary glands, breast, and gastric tissue there are no significant differences compared to other best-performing methods. Importantly, the different tissues are not equipotent, so it is important to compare the results separately for each tissue type. The results presented in Table 1 confirm that our method has a very good generalization ability since the reported rTRE is lower than in the presented figures (they present the results for all image pairs), showing that the rTRE for the evaluation set is even lower than for the training set.

In can be observed that the registration accuracy depends strongly on the tissue type, both for our as well as other methods (Figure 4, Figure 5). The differences are a consequence of the tissue characteristics, the procedure of the sample preparation and the applied dyes. For example, for lung lesions the rTRE is order of magnitude higher than for gastric tissues or mammary glands. The reasons for this are large amount of missing data between the subsequent slices and difficulties with determining differentiating features of which the similarity could be used to drive the nonrigid registration. This is not the case for mammary glands or gastric tissues that do not suffer from these limitations.

The computational time of the proposed framework is significantly lower than the classical, iterative methods. This could be expected since the registration consists of only an extremely fast model inference and low-resolution GPU-based initial rotation search. Moreover, about half of the total registration time is related to the data loading and initial preprocessing. Thus, the registration part consisting of the initial alignment, affine registration, and nonrigid registration is approximately two times faster than reported. The only state-of-the-art algorithm that is comparable in terms of the computational time is the MEVIS method [13]. Their well-optimized, multi-core, matrix-free implementation is amazingly fast but it is a really optimized commercial tool and not a simple research prototype.

The proposed method has two limitations. First, the training time of the nonrigid network is significantly longer compared to the simple, single-step forward pass registration networks that use the downsampled images. The process of training the networks takes several days (on the ANHIR dataset). This limitation may be addressed by the use of pretrained models since tuning the model

is significantly faster. The second limitation is connected to the generalizbility and the amount of the required training data. The proposed method is unsupervised, it does not require any annotations. However, it requires big enough training data to be successful. It is already visible in the results, e.g. the results for COADs are at the level or even better than the other approaches, on the other hand, results for mice kidneys are not as accurate. However, there is only a single case of the mice kidney tissue in the ANHIR dataset.

There are two strongly connected areas in which the methods can be improved. The first one is related to a better similarity measure than the NCC. Even though the results achieved using the NCC are robust and accurate, state-of-the-art methods show that there is still an area for further improvement. The MEVIS method [13] used the NGF and the AGH method [20] using a MIND-based similarity measure. Both of them achieved a lower median of median rTRE than the UPENN [15] method which used the NCC. However, the use of MIND or NGF in deep registration is problematic. The challenge is connected with similarity metric hyper-parameters and the proper weight of the regularization function. We did initial experiments with MIND and NGF. Due to the necessity of properly tuning the regularization and similarity metric parameters simultaneously, the training was unable to converge to a better solution than the one achieved by NCC. We forecast that further research about adaptive, deep regularization functions is crucial to further improve the method.

Another interesting idea is connected with style transfer using the adversarial networks [43]. It would be interesting to apply style connected with a particular dye from a given slice to another consecutive slice. This could be useful to create ground-truth alignments for adversarial registration networks which may produce even more accurate registration, without the necessity to define a similarity measure [28].

5. Conclusions

To conclude, we propose an unsupervised deep learning-based image registration framework dedicated to histology images acquired using different stains. The proposed framework provides results comparable to the best state-of-the-art methods while being significantly faster. The proposed method is of particular interest to researchers requiring a real-time, accurate, nonrigid registration of high-resolution images. We freely release the framework source code and provide access to pretrained models, making the results fully reproducible. The proposed method can be a useful baseline for other image registration researchers eager to propose novel deep learning-based nonrigid registration algorithms, dedicated to histology images, with common preprocessing and initial alignment steps.

Declaration of Competing Interest

M. Wodzinski is a researcher at the AGH University of Science and Technology, Poland, and is funded by the National Science Centre in Poland. H. Müller

is a professor at University of Applied Sciences Western Switzerland (HES-SO Valais), Information Systems Institute, and University of Geneva, Switzerland. The authors declare no conflict of interest.

Acknowledgments

The work was supported by the National Science Centre in Poland, under the Preludium project UMO-2018/29/N/ST6/00143 and Etiuda project UMO-2019/32/T/ST6/00065.

References

[1] J. Borovec, et al., ANHIR: Automatic Non-rigid Histological Image Registration Challenge, IEEE Transactions on Medical Imaging 39 (2020) 3042–3052.

[2] A. Sotiras, C. Davatzikos, N. Paragios, Deformable medical image registration: A survey, IEEE Transactions on Medical Imaging 32 (2013) 1153–1190.

[3] G. Haskins, U. Kruger, P. Yan, Deep Learning in Medical Image Registration: A Survey, Machine Vision and Applications 31 (2020).

[4] ANHIR Website, `https://anhir.grand-challenge.org`, 2019.

[5] J. Borovec, A. Munoz-Barrutia, J. Kybic, Benchmarking of Image Registration Methods for Differently Stained Histological Slides, IEEE International Conference on Image Processing (2018) 3368–3372.

[6] F. Oliveira, J. Tavares, Medical image registration: A review, Computer Methods in Biomechanics and Biomedical Engineering 17 (2014) 73–93.

[7] I. Arganda-Carreras, et al., Consistent and elastic registration of histological sections using vector-spline regularization, Lecture Notes in Computer Science 4241 LNCS (2006) 85–95.

[8] D. Rueckert, Nonrigid Registration Using Free-Form Deformations: Application to Breast MR images, IEEE Transactions on Medical Imaging 18 (1999) 712–721.

[9] B. Avants, et al., Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain, Medical Image Analysis 12 (2008) 26–41.

[10] B. Glocker, et al., Deformable medical image registration: Setting the state of the art with discrete methods, Annual Review of Biomedical Engineering 13 (2011) 219–244.

[11] S. Klein, et al., Elastix: A toolbox for intensity-based medical image registration, IEEE Transactions on Medical Imaging 29 (2010) 196–205.

[12] Y. Song, et al., Unsupervised content classification based nonrigid registration of differently stained histology images, IEEE Transactions on Biomedical Engineering 61 (2014) 96–108.

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

[14] E. Haber, J. Modersitzki, Intensity Gradient Based Registration and Fusion of Multi-modal Images, MICCAI 2006 (2006) 726–733.

[15] L. Venet, et al., Accurate and Robust Alignment of Variable-stained Histologic Images Using a General-purpose Greedy Diffeomorphic Registration Tool, arXiv:1904.11929 (2019).

[16] M. Wodzinski, A. Skalski, Automatic Nonrigid Histological Image Registration with Adaptive Multistep Algorithm, arXiv:1904.00982 (2019).

[17] S. Joshi, B. Davis, M. Jomier, G. Gerig, Unbiased diffeomorphic atlas construction for computational anatomy, NeuroImage 23 (2004) 151–160.

[18] P. Yushkevich, et al., Fast Automatic Segmentation of Hippocampal Subfields and Medial Temporal Lobe Subregions in 3 Tesla and 7 Tesla T2-Weighted MRI, Alzheimer's & Dementia 12 (2016) 126–127.

[19] J. Thirion, Image matching as a diffusion process: An analogy with Maxwell's demons, Medical Image Analysis 2 (1998) 243–260.

[20] M. Heinrich, et al., MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration, Medical Image Analysis 16 (2012) 1423–1435.

[21] S. Zhao, T. Lau, J. Luo, E. Chang, Y. Xu, Unsupervised 3D End-to-End Medical Image Registration with Volume Tweening Network, IEEE Journal of Biomedical and Health Informatics (2019). (Early Access).

[22] G. Litjens, et al., A survey on deep learning in medical image analysis, Medical Image Analysis 42 (2017) 60–88.

[23] D. DeTone, T. Malisiewicz, A. Rabinovich, Deep Image Homography Estimation, arXiv:1606.03798 (2016).

[24] E. Chee, Z. Wu, AIRNet: Self-Supervised Affine Registration for 3D Medical Images using Neural Networks, arXiv:1810.02583 (2018).

[25] B. de Vos, F. Berendsen, M. Viergever, H. Sokooti, M. Staring, I. Isgum, A deep learning framework for unsupervised affine and deformable image registration, Medical Image Analysis 52 (2019) 128–143.

[26] G. Balakrishnan, A. Zhao, M. Sabuncu, J. Guttag, A. Dalca, VoxelMorph: A Learning Framework for Deformable Medical Image Registration, IEEE Transactions on Medical Imaging 38 (2019) 1788–1800.

[27] A. Dalca, G. Balakrishnan, J. Guttag, M. Sabuncu, Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces, Medical Image Analysis 57 (2019) 226–236.

[28] J. Fan, X. Cao, Q. Wang, P. Yap, D. Shen, Adversarial learning for mono- or multi-modal registration, Medical Image Analysis 58 (2019).

[29] D. Mahapatra, B. Antony, S. Sedai, R. Garnavi, Deformable medical image registration using generative adversarial networks, IEEE ISBI (2018) 1449–1453.

[30] M. Arjovsky, L. Bottou, Towards principled methods for training generative adversarial networks, ICLR 2017 (2017).

[31] M. Heinrich, L. Hansen, Highly accurate and memory efficient unsupervised learning-based discrete CT registration using 2.5D displacement search, MICCAI 2020 1 (2020) 1–11. Preprint.

[32] M. Wodzinski, H. Müller, Unsupervised Learning-based Nonrigid Registration of High Resolution Histology Images, MICCAI-MLMI 2020 (2020) 1–10.

[33] M. Wodzinski, H. Müller, Learning-based Affine Registration of Histological Images, 9th International Workshop on Biomedical Image Registration (2020) 1–10.

[34] Proposed Method Software, `https://github.com/lNefarin/DeepHistReg`, 2020.

[35] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, MICCAI 2015 (2015) 234–241.

[36] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, IEEE CVPR (2016) 770–778.

[37] B. Fischer, J. Modersitzki, Curvature based image registration, Journal of Mathematical Imaging and Vision 18 (2003) 81–85.

[38] A. Paszke, et al., Automatic differentiation in pytorch (2017).

[39] R. Fernandez-Gonzalez, et al., System for combined three-dimensional morphological and molecular analysis of thick tissue specimens, Microscopy Research and Technique 59 (2002) 522–530.

[40] L. Gupta, et al., Stain independent segmentation of whole slide images: A case study in renal histology, Proceedings - International Symposium on Biomedical Imaging (2018) 1360–1364.

[41] I. Mikhailov, N. Danilova, P. Malkov, The immune microenvironment of various histological types of ebv-associated gastric cancer, Virchows Archiv (2018).

[42] G. Bueno, O. Deniz, AIDPATH: Academia and Industry Collaboration for Digital Pathology, `http://aidpath.eu`, 2017.

[43] Z. Xu, M. Wilber, C. Fang, A. Hertzmann, H. Jin, Learning from Multi-Domain Artistic Images for Arbitrary Style Transfer, Proceedings of the 8th ACM/EG Expressive Symposium (2019) 21–31.