

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

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 is 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, non-rigid 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

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

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

37 1.1. Related Work

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

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

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

78 The deep learning-based medical image registration is a relatively new field [3,
79 22]. It immediately turned out to be useful because of the low time required
80 during the model inference enabling real-time nonrigid registration. This is
81 crucial, e.g. for registration during surgical interventions. The deep registra-
82 tion approaches can be divided into three main categories based on the training
83 scheme: (i) supervised [23, 24], (ii) unsupervised [25, 26, 27], and (iii) adver-
84 sarial registration [28, 29]. The supervised registration requires ground-truth
85 deformation fields or pre-aligned images that are often impossible to obtain.
86 The adversarial approach, which is based on generator and discriminator net-
87 works, suffers from similar limitations. Moreover, the adversarial networks are
88 not trivial to train [30]. To train the discriminator, registered image pairs rep-
89 resenting the ground-truth alignment are necessary. For some tasks it can be
90 achieved [28], but it is usually costly and time-consuming. On the other hand,
91 unsupervised methods do not require any ground-truth. They are based on min-
92 imization of a given cost function and the registration accuracy mostly depends
93 on: (i) a proper choice of the similarity measure, (ii) a regularization term en-
94 forcing plausible deformations, (iii) the ability to converge during training, and
95 (iv) a generalization ability. The unsupervised registration can be seen as a
96 way to speed-up the classical, iterative image registration. Since the methods
97 proposed by the ANHIR challenge participants (e.g. MEVIS, UPENN, AGH)
98 achieved results close to the human annotation error [1], we decided to focus
99 on the unsupervised registration, addressing the difficulties connected with the
100 high-resolution of histology images and the required robustness while providing
101 the real-time alignment.

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

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

132 1.2. Contribution

133 In this work, we propose an unsupervised deep learning-based registration
134 framework dedicated to histology images acquired using different stains. The
135 method is robust to different dyes and tissue types within the evaluated dataset,
136 fully automatic and provides results comparable to the state-of-the-art meth-
137 ods while being significantly faster, which is important in clinical practice. The
138 main technical novelty is related to the ability to recover large, complex, defor-
139 mations between high-resolution images (even above 15000x15000) in real-time.
140 The method is built on our previous contributions [32, 33], however, the meth-
141 ods are greatly extended and improved. An additional learning-based back-
142 ground segmentation was introduced, significantly improving the initial align-
143 ment. Moreover, the framework user can now decide which affine registration
144 is suitable for the given problem. Finally, the nonrigid registration is improved
145 by changing the order of the the unfold/fold operations, the velocity field com-
146 position, and the image transformation. These changes eliminated the problem
147 of inconsistency at patch boundaries. We provide free access to the framework
148 source code [34], including training/inference scripts and pre-trained models.
149 This, together with open access to the ANHIR dataset, makes the results fully
150 and easily reproducible. Moreover, the framework was evaluated using the in-
151 dependent ANHIR evaluation platform. The results can easily be verified and
152 compared using the challenge website [4]. The proposed framework is easily
153 extendable making it possible to further improve the registration by other re-
154 searchers.

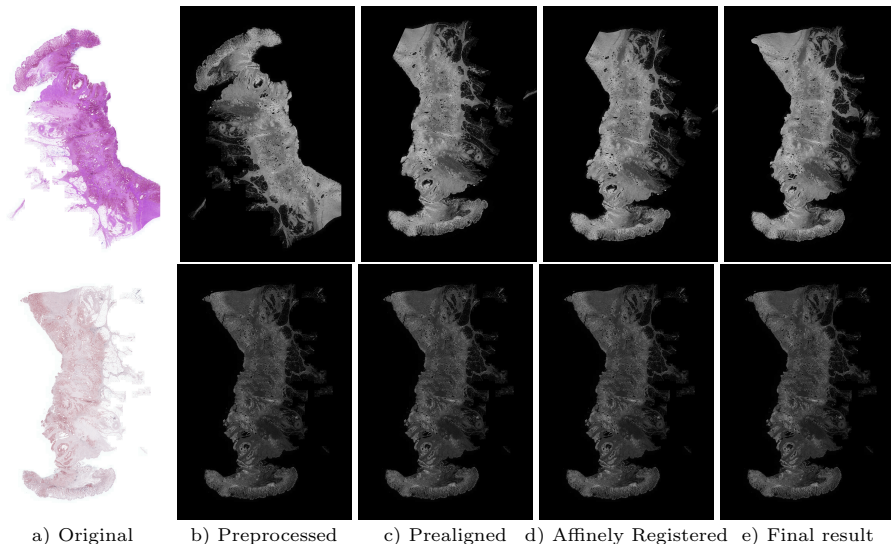


Figure 1: The source and target at different registration stages (top and bottom row respectively, pair 178). The target is replicated for presentation clarity. High quality picture, best viewed zoomed in electronic format.

155 2. Methods

156 2.1. Overview

157 We propose an unsupervised deep learning-based registration framework.
 158 The framework pipeline consists of data loading, transferring to GPU, prepro-
 159 cessing, initial alignment, affine registration, and finally nonrigid registration.
 160 All the steps are described in detail in the following sections. The negative nor-
 161 malized cross-correlation (NCC) is used as the cost function during all registra-
 162 tion steps. The global NCC is used for initial alignment and affine registration,
 163 while the patch-based NCC is used for the nonrigid registration. The curvature
 164 (CURV) is used as the regularization term for the nonrigid registration:

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

165 where NCC is the normalized cross-correlation (global or local version, depend-
 166 ing on context, CURV denotes the curvature regularization (only for the non-
 167 rigid registration), α is the regularization parameter controlling the deformation
 168 smoothness, and M, F, u are respectively the warped moving patches, target
 169 patches and the displacement fields.

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

173 *2.2. Preprocessing*

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

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

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

194 *2.3. Initial Alignment*

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

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

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

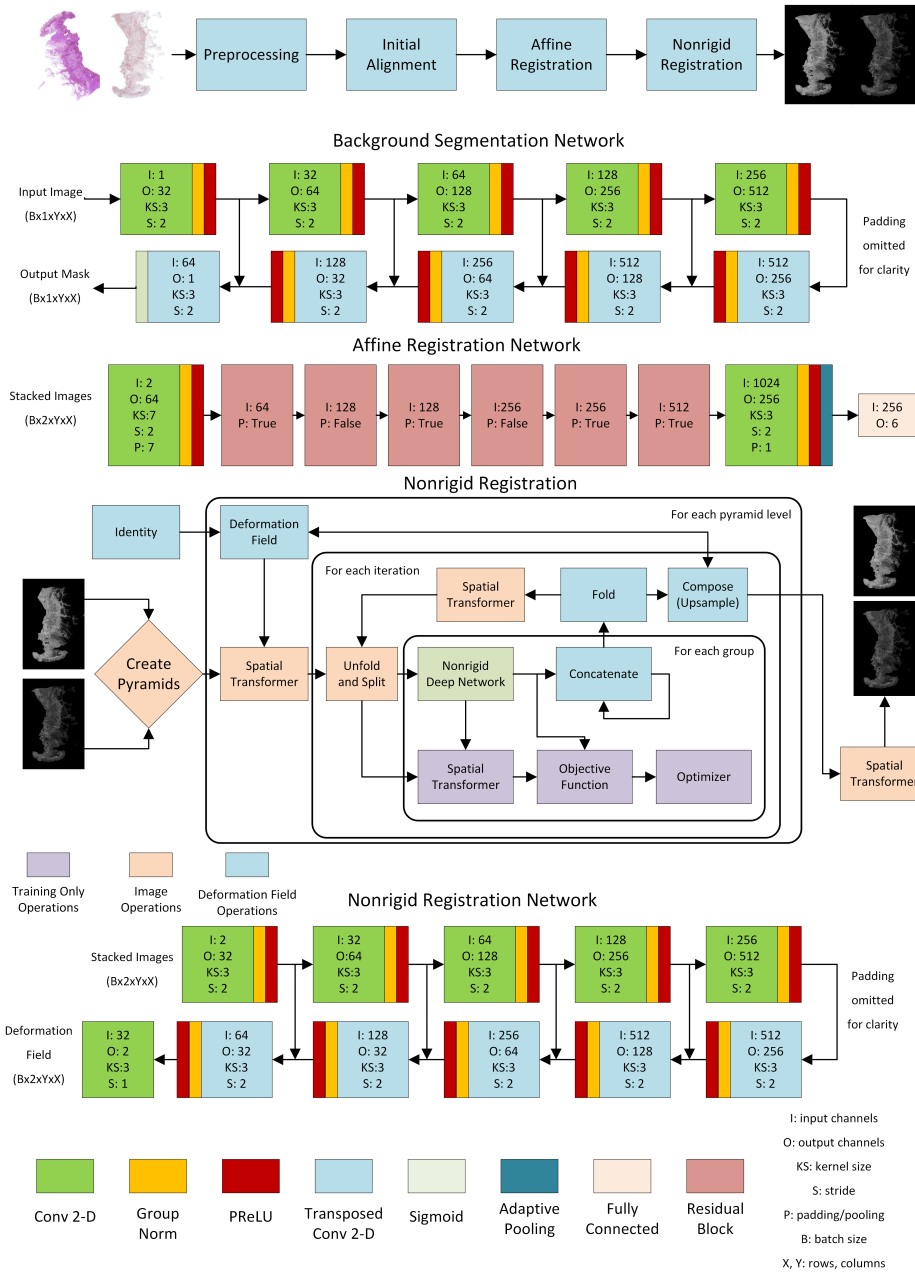


Figure 2: Visualization of the proposed deep histology registration framework.

217 *2.4. Affine Registration*

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

228 *2.5. Nonrigid Registration*

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

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

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

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

271 The nonrigid method has several parameters: (i) the patch size, (ii) stride,
272 (iii) group size, (iv) number of pyramid levels, (v) number of iterations per level,
273 and (vi) the regularization parameter. The patch size and the stride are respon-
274 sible for lowering the number of network parameters and the required receptive
275 field and thus the required GPU memory. The lower the patch size and stride,
276 the larger the available group size, at the cost of decreasing the maximum mag-
277 nitude of deformations. The number of pyramid levels is responsible for ensuring
278 that the given patch size is able to capture sufficiently large deformations. Since
279 the patch size is constant for all pyramid levels, the lower resolution is able to
280 capture larger deformations, as in the traditional, iterative techniques. The
281 number of iterations per level defines how many times the patches are passed
282 through the network at each resolution. This increases the registration accu-
283 racy at the cost of increasing the registration time. Finally, the regularization
284 parameter is responsible for controlling the deformation smoothness.

285 *2.6. Technical Details*

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

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

300 *2.7. Dataset and Experimental Setup*

301 We used the open ANHIR dataset to evaluate the proposed framework [1,
302 39, 40, 41, 42] and promote research transparency, openness, and reproducibil-
303 ity. The dataset consists of 481 image pairs split into 251 evaluation and 230

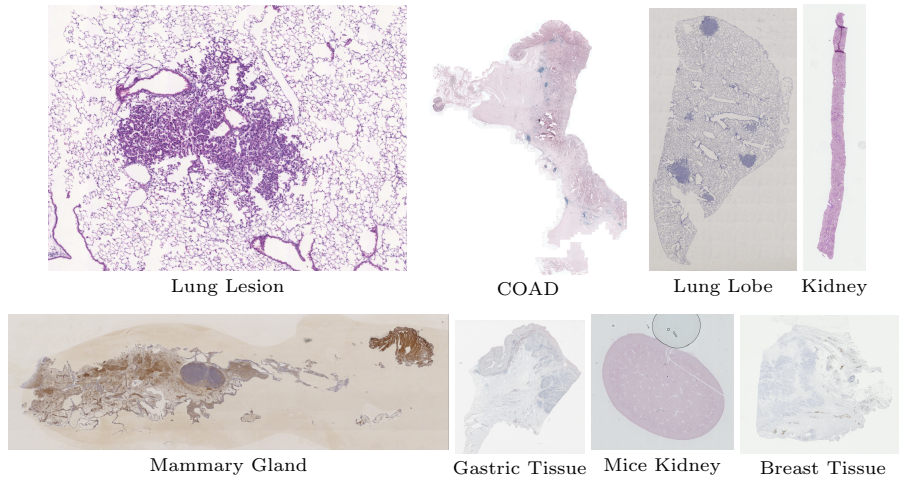


Figure 3: Visualization of different tissue types stained using various dyes [1, 4]. It presents the robustness and generalization ability required by the deep registration framework (best viewed zoomed in electronic format).

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

319 The dataset was annotated by providing corresponding landmarks. In total,
 320 9 qualified annotators chose on average 86 landmarks per image. The average
 321 error between the landmarks chosen by two annotators is 0.05% of the image
 322 diagonal. This can be used as the indicator of the human-level accuracy and a
 323 threshold below which the registration methods become indistinguishable [1, 4].
 324 The corresponding landmarks are openly available only for the training images.
 325 For the evaluation set, only the source image landmarks are released and the
 326 evaluation must be done using the server-side evaluation platform developed by
 327 the ANHIR organizers with a very limited number of available submissions. This
 328 makes the results reliable and trustworthy since the proposed methods cannot

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

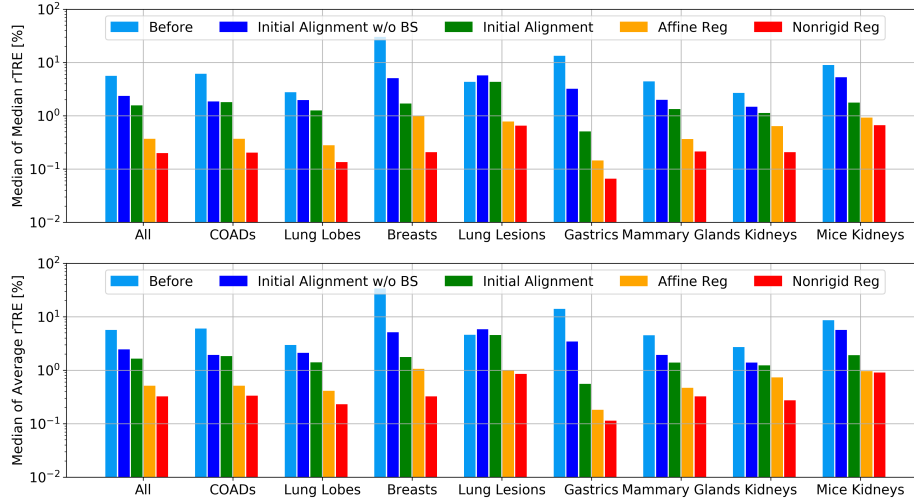


Figure 4: The Median of Median rTRE and Median of Average rTRE reported in % of image diagonal for all registered image pairs. Please note that the tissue types are not equipotent, COADs have the highest influence on the averaged results (logarithmic scale used for presentation clarity). The results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

330

331 2.8. Evaluation Criteria

332 The registration accuracy is measured by the target registration error (TRE)
 333 measuring the Euclidean distance between the annotated and transformed land-
 334 marks. To make the error comparable between image pairs with different resolu-
 335 tions, the TRE is normalized by the image diagonal:

$$rTRE = \frac{TRE}{\sqrt{w^2 + h^2}}, \quad (2)$$

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

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

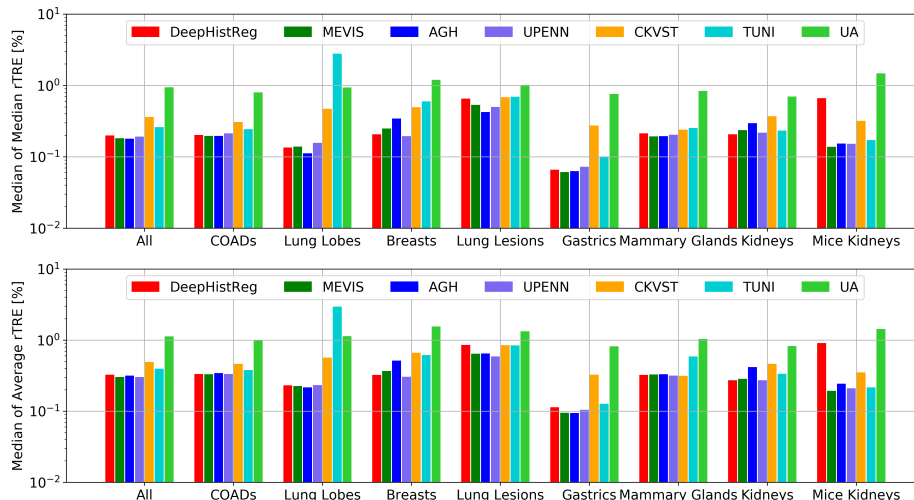


Figure 5: Comparison of the proposed framework to the state-of-the-art nonrigid histology registration methods in terms of Median of Median rTRE and Median of Average rTRE reported in % of the image diagonal, evaluated for all image pairs. The results originate from the ANHIR evaluation system [4]. The 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.

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

350 3. Results

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

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

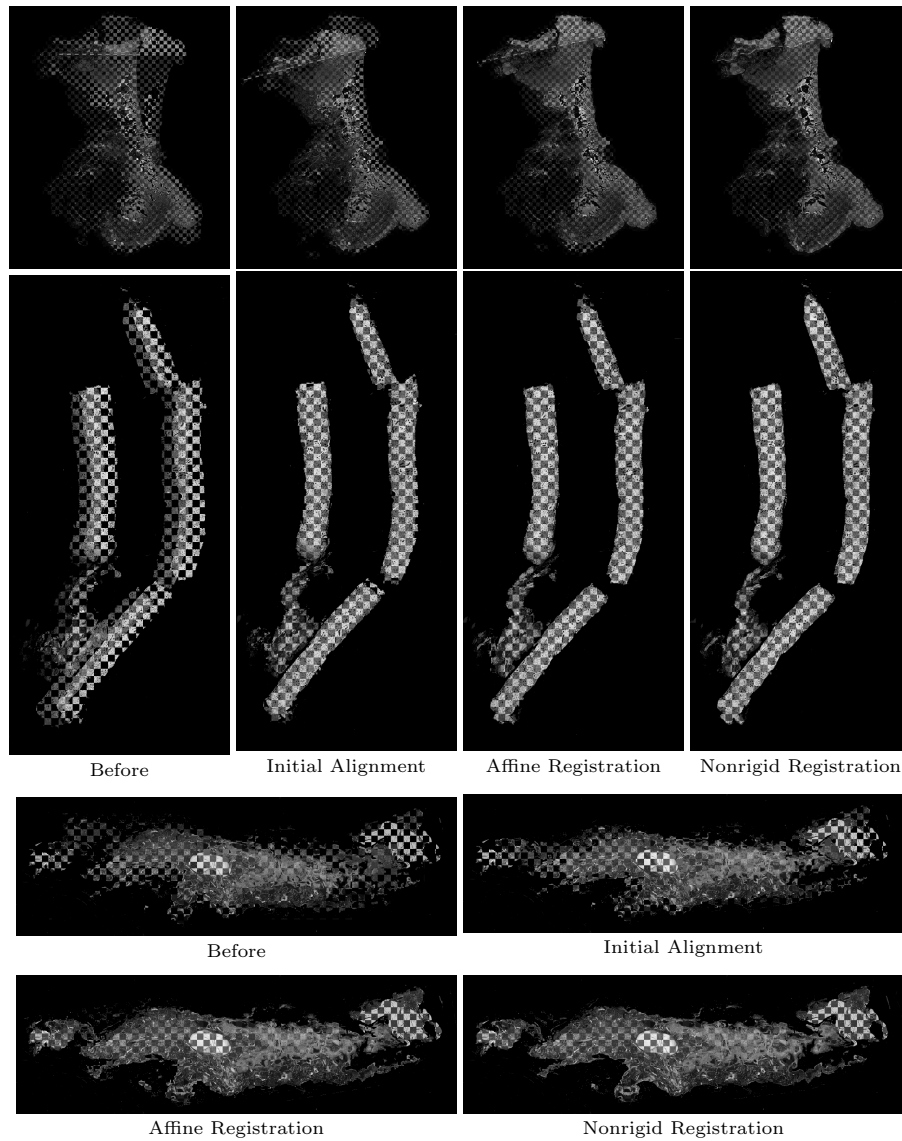


Figure 6: Example checkerboards at different registration stages for pairs no. 13, 317, 430 (COAD, kidney, mammary gland). High quality pictures, best viewed zoomed in electronic format.

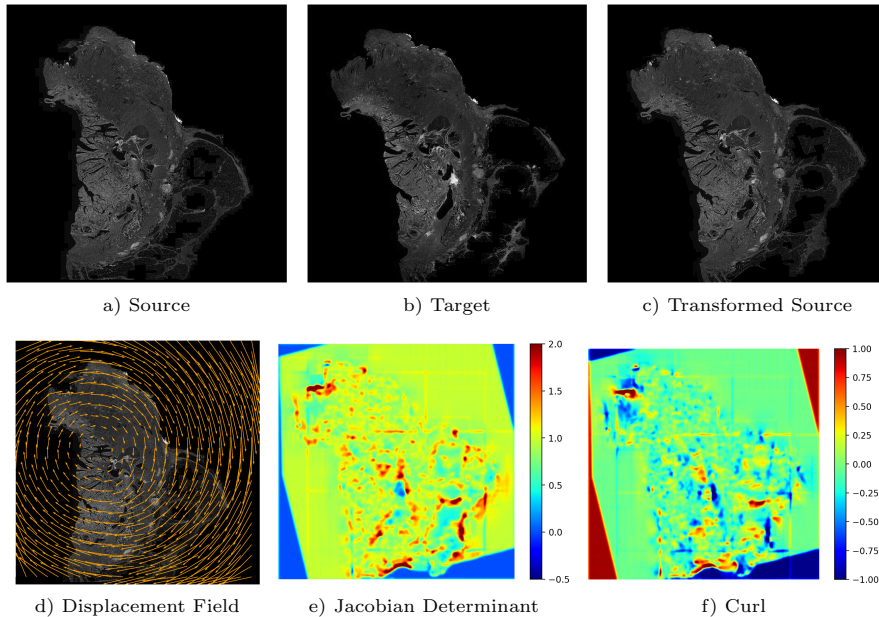


Figure 7: An exemplary visualization of the registration results, together with the deformation field, its Jacobian determinant and Curl.

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.

method	Average rTRE		Median rTRE		Max rTRE		Robustness		Average time [min]
	Average	Median	Average	Median	Average	Median	Average	Median	
<i>DeepHistReg</i>	<i>0.0061</i>	<i>0.0033</i>	<i>0.0047</i>	<i>0.0019</i>	<i>0.0276</i>	<i>0.0224</i>	<i>0.9799</i>	<i>1.0000</i>	<i>0.03</i>
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

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

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

389 4. Discussion

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

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

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

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

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

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

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

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

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

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

485 **5. Conclusions**

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

496 **Declaration of Competing Interest**

497 M. Wodzinski is a researcher at the AGH University of Science and Technol-
498 ogy, Poland, and is funded by the National Science Centre in Poland. H. Müller

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

502 **Acknowledgments**

503 The work was supported by the National Science Centre in Poland, under
504 the Preludium project UMO-2018/29/N/ST6/00143 and Etiuda project UMO-
505 2019/32/T/ST6/00065.

506 **References**

- 507 [1] J. Borovec, et al., ANHIR: Automatic Non-rigid Histological Image Registration
508 Challenge, *IEEE Transactions on Medical Imaging* 39 (2020) 3042–3052.
- 509 [2] A. Sotiras, C. Davatzikos, N. Paragios, Deformable medical image registration:
510 A survey, *IEEE Transactions on Medical Imaging* 32 (2013) 1153–1190.
- 511 [3] G. Haskins, U. Kruger, P. Yan, Deep Learning in Medical Image Registration: A
512 Survey, *Machine Vision and Applications* 31 (2020).
- 513 [4] ANHIR Website, <https://anhir.grand-challenge.org>, 2019.
- 514 [5] J. Borovec, A. Munoz-Barrutia, J. Kybic, Benchmarking of Image Registration
515 Methods for Differently Stained Histological Slides, *IEEE International Confer-
516 ence on Image Processing* (2018) 3368–3372.
- 517 [6] F. Oliveira, J. Tavares, Medical image registration: A review, *Computer Methods
518 in Biomechanics and Biomedical Engineering* 17 (2014) 73–93.
- 519 [7] I. Arganda-Carreras, et al., Consistent and elastic registration of histological
520 sections using vector-spline regularization, *Lecture Notes in Computer Science*
521 4241 LNCS (2006) 85–95.
- 522 [8] D. Rueckert, Nonrigid Registration Using Free-Form Deformations: Application
523 to Breast MR images, *IEEE Transactions on Medical Imaging* 18 (1999) 712–721.
- 524 [9] B. Avants, et al., Symmetric diffeomorphic image registration with cross-
525 correlation: Evaluating automated labeling of elderly and neurodegenerative
526 brain, *Medical Image Analysis* 12 (2008) 26–41.
- 527 [10] B. Glocker, et al., Deformable medical image registration: Setting the state of the
528 art with discrete methods, *Annual Review of Biomedical Engineering* 13 (2011)
529 219–244.
- 530 [11] S. Klein, et al., Elastix: A toolbox for intensity-based medical image registration,
531 *IEEE Transactions on Medical Imaging* 29 (2010) 196–205.
- 532 [12] Y. Song, et al., Unsupervised content classification based nonrigid registration of
533 differently stained histology images, *IEEE Transactions on Biomedical Engineer-
534 ing* 61 (2014) 96–108.

- 535 [13] J. Lotz, N. Weiss, S. Heldmann, Robust, fast and accurate: a 3-step method for
536 automatic histological image registration, arXiv:1903.12063 (2019).
- 537 [14] E. Haber, J. Modersitzki, Intensity Gradient Based Registration and Fusion of
538 Multi-modal Images, MICCAI 2006 (2006) 726–733.
- 539 [15] L. Venet, et al., Accurate and Robust Alignment of Variable-stained Histo-
540 logic Images Using a General-purpose Greedy Diffeomorphic Registration Tool,
541 arXiv:1904.11929 (2019).
- 542 [16] M. Wodzinski, A. Skalski, Automatic Nonrigid Histological Image Registration
543 with Adaptive Multistep Algorithm, arXiv:1904.00982 (2019).
- 544 [17] S. Joshi, B. Davis, M. Jomier, G. Gerig, Unbiased diffeomorphic atlas construction
545 for computational anatomy, *NeuroImage* 23 (2004) 151–160.
- 546 [18] P. Yushkevich, et al., Fast Automatic Segmentation of Hippocampal Subfields
547 and Medial Temporal Lobe Subregions in 3 Tesla and 7 Tesla T2-Weighted MRI,
548 *Alzheimer’s & Dementia* 12 (2016) 126–127.
- 549 [19] J. Thirion, Image matching as a diffusion process: An analogy with Maxwell’s
550 demons, *Medical Image Analysis* 2 (1998) 243–260.
- 551 [20] M. Heinrich, et al., MIND: Modality independent neighbourhood descriptor for
552 multi-modal deformable registration, *Medical Image Analysis* 16 (2012) 1423–
553 1435.
- 554 [21] S. Zhao, T. Lau, J. Luo, E. Chang, Y. Xu, Unsupervised 3D End-to-End Medical
555 Image Registration with Volume Tweening Network, *IEEE Journal of Biomedical
556 and Health Informatics* (2019). (Early Access).
- 557 [22] G. Litjens, et al., A survey on deep learning in medical image analysis, *Medical
558 Image Analysis* 42 (2017) 60–88.
- 559 [23] D. DeTone, T. Malisiewicz, A. Rabinovich, Deep Image Homography Estimation,
560 arXiv:1606.03798 (2016).
- 561 [24] E. Chee, Z. Wu, AIRNet: Self-Supervised Affine Registration for 3D Medical
562 Images using Neural Networks, arXiv:1810.02583 (2018).
- 563 [25] B. de Vos, F. Berendsen, M. Viergever, H. Sokooti, M. Staring, I. Isgum, A deep
564 learning framework for unsupervised affine and deformable image registration,
565 *Medical Image Analysis* 52 (2019) 128–143.
- 566 [26] G. Balakrishnan, A. Zhao, M. Sabuncu, J. Guttag, A. Dalca, VoxelMorph: A
567 Learning Framework for Deformable Medical Image Registration, *IEEE Trans-
568 actions on Medical Imaging* 38 (2019) 1788–1800.
- 569 [27] A. Dalca, G. Balakrishnan, J. Guttag, M. Sabuncu, Unsupervised learning of
570 probabilistic diffeomorphic registration for images and surfaces, *Medical Image
571 Analysis* 57 (2019) 226–236.
- 572 [28] J. Fan, X. Cao, Q. Wang, P. Yap, D. Shen, Adversarial learning for mono- or
573 multi-modal registration, *Medical Image Analysis* 58 (2019).

- 574 [29] D. Mahapatra, B. Antony, S. Sedai, R. Garnavi, Deformable medical image
575 registration using generative adversarial networks, *IEEE ISBI* (2018) 1449–1453.
- 576 [30] M. Arjovsky, L. Bottou, Towards principled methods for training generative
577 adversarial networks, *ICLR 2017* (2017).
- 578 [31] M. Heinrich, L. Hansen, Highly accurate and memory efficient unsupervised
579 learning-based discrete CT registration using 2.5D displacement search, *MICCAI*
580 2020 1 (2020) 1–11. Preprint.
- 581 [32] M. Wodzinski, H. Müller, Unsupervised Learning-based Nonrigid Registration of
582 High Resolution Histology Images, *MICCAI-MLMI 2020* (2020) 1–10.
- 583 [33] M. Wodzinski, H. Müller, Learning-based Affine Registration of Histological Im-
584 ages, 9th International Workshop on Biomedical Image Registration (2020) 1–10.
- 585 [34] Proposed Method Software, <https://github.com/1Nefarin/DeepHistReg>, 2020.
- 586 [35] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomed-
587 ical Image Segmentation, *MICCAI 2015* (2015) 234–241.
- 588 [36] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition,
589 *IEEE CVPR* (2016) 770–778.
- 590 [37] B. Fischer, J. Modersitzki, Curvature based image registration, *Journal of Math-*
591 *ematical Imaging and Vision* 18 (2003) 81–85.
- 592 [38] A. Paszke, et al., Automatic differentiation in pytorch (2017).
- 593 [39] R. Fernandez-Gonzalez, et al., System for combined three-dimensional morpho-
594 logical and molecular analysis of thick tissue specimens, *Microscopy Research*
595 *and Technique* 59 (2002) 522–530.
- 596 [40] L. Gupta, et al., Stain independent segmentation of whole slide images: A case
597 study in renal histology, *Proceedings - International Symposium on Biomedical*
598 *Imaging* (2018) 1360–1364.
- 599 [41] I. Mikhailov, N. Danilova, P. Malkov, The immune microenvironment of various
600 histological types of ebv-associated gastric cancer, *Virchows Archiv* (2018).
- 601 [42] G. Bueno, O. Deniz, AIDPATH: Academia and Industry Collaboration for Digital
602 Pathology, <http://aidpath.eu>, 2017.
- 603 [43] Z. Xu, M. Wilber, C. Fang, A. Hertzmann, H. Jin, Learning from Multi-Domain
604 Artistic Images for Arbitrary Style Transfer, *Proceedings of the 8th ACM/EG*
605 *Expressive Symposium* (2019) 21–31.