Unsupervised Learning-based Nonrigid Registration of High Resolution Histology Images

Marek Wodzinski[1][0000−0002−8076−6246], Henning Müller[2][0000−0001−6800−9878]

[1]AGH University of Science and Technology
Department of Measurement and Electronics, Krakow, Poland
wodzinski@agh.edu.pl
[2]University of Applied Sciences Western Switzerland (HES-SO Valais)
Information Systems Institute, Sierre, Switzerland
henning.mueller@hevs.ch

Abstract. The use of different dyes during histological sample preparation reveals distinct tissue properties and may improve the diagnosis. Nonetheless, the staining process deforms the tissue slides and registration is necessary before further processing. The importance of this problem led to organizing an open challenge named Automatic Non-rigid Histological Image Registration Challenge (ANHIR), organized jointly with the IEEE ISBI 2019 conference. The challenge organizers provided 481 image pairs and a server-side evaluation platform making it possible to reliably compare the proposed algorithms. The majority of the methods proposed for the challenge were based on the classical, iterative image registration, resulting in high computational load and arguable usefulness in clinical practice due to the long analysis time. In this work, we propose a deep learning-based unsupervised nonrigid registration method, that provides results comparable to the solutions of the best scoring teams, while being significantly faster during the inference. We propose a multi-level, patch-based training and inference scheme that makes it possible to register images of almost any size, up to the highest resolution provided by the challenge organizers. The median target registration error is close to 0.2% of the image diagonal while the average registration time, including the data loading and initial alignment, is below 3 seconds. We freely release both the training and inference code making the results fully reproducible.

Keywords: Image registration · Deep learning · Histology · ANHIR

1 Introduction

Registration of histology images acquired using different stains is a difficult and important task that makes it possible to fuse information and improve further processing and diagnosis. The problem is challenging due to: (i) a very high resolution of the images, (ii) complex, large deformations, (iii) difference in the

appearance and partially missing data. A dedicated challenge named Automatic Non-rigid Histological Image Registration Challenge (ANHIR) [1,2,3] was organized in conjunction with the IEEE ISBI 2019 conference to address the problem and compare algorithms developed by different researchers. The challenge organizers provided a high quality and open dataset [1,4,5,6,7], manually annotated by experts and reasonably divided into training and evaluation sets. Moreover, an independent, server-side evaluation platform was developed that made it possible to reliably compare the participant's solutions [3].

The challenge participants proposed several different algorithms, mostly using the classical, iterative approach to the image registration [2]. The three best scoring methods were quite similar. The winner team (MEVIS) [8] proposed a method consisting of brute-force initial alignment followed by affine and B-Splines-based nonrigid registration. The researchers used the normalized gradient field (NGF) similarity metric [9] and strongly optimized code resulting in undoubtedly the best and clinically applicable method. The team with the second-best score (UPENN) [10] proposed an algorithm consisting of background removal by stain deconvolution, random initial alignment, affine registration and diffeomorphic, nonrigid registration based on the Greedy tool [11,12]. The third best team (AGH) [13] developed a method similar to the winners with the differences that instead of the NGF they used the modality independent neighborhood descriptor (MIND) [14], and used the Demons algorithm to directly optimize the dense deformation field replacing the B-Spline deformation model, as in [8]. Interestingly, only a single team proposed a method based on deep learning [15]. However, the team first re-sampled the images to a relatively low resolution and second, they fine-tuned the deep network using the landmarks provided for the evaluation, thus introducing strong bias into the results. Nonetheless, since their method was amazingly fast and therefore potentially the most useful in real-world applications, we feel inspired to propose a method based on deep learning that works for high resolution images without the requirement to fine-tune the network using manually annotated landmarks to achieve results comparable to the best scoring solutions based on the classical approach.

One of the challenges related to deep learning registration is connected to image size. High resolution images cannot be simply propagated through the network because the number of parameters required to provide an appropriate receptive field and accurate registration results would be too high to fit in the GPU memory. There are several approaches that decrease the GPU memory consumption by e.g. using B-Splines transformation model instead of the dense deformation field [16] or using a patch-based approach [17]. However, it is not well-established how to deal with patches that cannot be directly propagated through the network after unfolding due to the GPU memory constraints, which is the case for the high resolution histology images (full images are in the range of 100k x 100x pixels).

In this work, we propose a nonrigid registration method based on deep networks trained in an unsupervised way in a multi-level, patch-based and multi-iteration framework. The proposed approach works for images with any resolu-

tion using a single GPU, both during the training and the inference. The results are comparable to the best scoring teams using a traditional, iterative approach while being significantly faster during the inference. We make the source code freely available that, together with the open access to the data set, makes the results fully reproducible [18].

2 Methods

We assume that the input to the proposed method consists of images initially aligned by an affine registration. In this work, we used the affine registration method described in [19] that accurately aligns the large majority of the image pairs.

The first step is to transfer the source and target images to the GPU memory. The image transfer is being done only if a single GPU is used and both images fit into the memory. Otherwise, if a multi-GPU computing cluster is used and the memory transfer is done later. Second, the images are re-sampled to a pre-defined number of levels, building a classical resolution pyramid. This approach allows the method to calculate significantly larger deformations. Then, starting at the lowest resolution, the images are unfolded into overlapping patches with a given size and stride. The patches overlap because we use only the centers of the calculated displacement fields. The displacement vectors calculated at patch boundaries are not reliable because the real displacement may point outside the given image patch. The unfolded patches are grouped in the tensor batch dimension with a pre-defined size. This parameter controls the GPU memory usage and can be adjusted differently during the training and inference. If a multi-GPU computing cluster is used, the groups are divided between the GPUs. For each group, the patches are passed through the deep network.

We used the encoder/decoder U-Net-like architecture [20] with the batch normalization replaced by the group normalization [21] and max-pooling replaced by strided convolutions. The network architecture is part of Figure 1.

As the next step, the current group is warped with the resulting displacement field and the loss is calculated, backpropagated and the optimizer is updated. In this work, we use the negative NCC as the cost function and the curvature as the displacement field regularization term [15,22]. The objective function can be defined as:

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

where NCC denotes the normalized cross-correlation, CURV is the curvature regularization, α is the regularization parameter controlling the deformation smoothness, and M, F, u are the warped moving patches, target patches and the displacement fields respectively.

The displacement fields calculated for each group are being concatenated. After the groups are processed, the displacement field patches are folded back into a single displacement field, which is then composed with the current deformation field, using the same patch size and stride as during the unfolding.

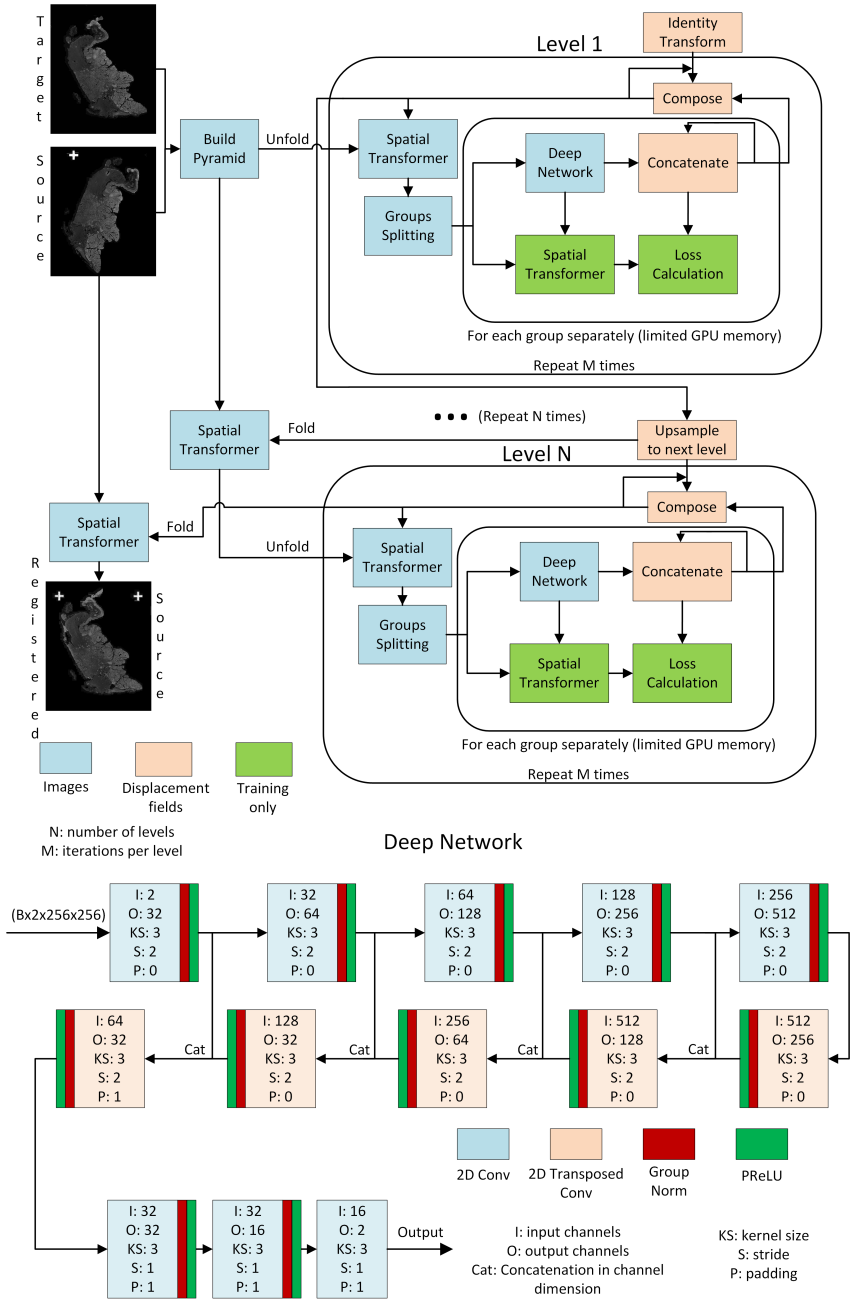


Fig. 1: Visualization of the proposed framework and the deep network architecture.

The whole process is repeated for a pre-defined number of times for each resolution, resulting in a multi-iteration registration. The network weights are shared between iterations at a given level. However, they differ between the pyramid levels. Finally, after each level, the calculated deformation field is up-sampled to the next pyramid level. The source patches are not interpolated more than once. Instead, the deformation fields are being composed together after each iteration and pyramid level. As a result, the interpolation error is negligible. The method is summarized in Algorithm 1 and visualized in Figure 1.

Algorithm 1: Algorithm Summary.

Input : **M** (affinely registered moving image), **F** (fixed image), N (number of pyramid levels), M (iterations per level), P (patch size), S (stride), G (group size)
Output: **u** (deformation field)

$\mathbf{P_M}, \mathbf{P_F}$ = create pyramids using **M**, **F** and N
u = initialize with an identity transform on the coarsest level
for *each resolution in N* **do**
  $\mathbf{F_c}$ = get current level $\mathbf{P_F}$ and unfold using P, S
  if *current resolution > 0* **then**
    $\mathbf{M_c}$ = warp current level $\mathbf{P_M}$ using **u**
  $\mathbf{M_c}$ = unfold $\mathbf{M_c}$ using P, S
  v = initialize with an identity transform and unfold using P, S
  for *each inner iteration in M* **do**
    if *current iteration > 0* **then**
      $\mathbf{M_c}$ = warp $\mathbf{M_c}$ with **v**
    $\mathbf{M_g}, \mathbf{T_g}$ = divide $\mathbf{F_c}, \mathbf{M_c}$ into G-sized batches
    $\mathbf{v_i}$ = initialize with an empty tensor
    for *each group* **do**
      $\mathbf{v_t}$ = model($\mathbf{M_g}, \mathbf{T_g}$)
      if *training* **then**
        $\mathbf{M_w}$ = warp $\mathbf{M_g}$ with $\mathbf{v_t}$
        $\mathbf{S}(\mathbf{M_w}, \mathbf{T_g}, \mathbf{v_t})$ = use equation (1) and update optimizer (free GPU memory for the next group)
      $\mathbf{v_i}$ = concatenate($\mathbf{v_i}$, $\mathbf{v_t}$)
    v = **v** $\circ$ $\mathbf{v_i}$
  v = fold **v** using P, S
  u = **u** $\circ$ **v**
return u

The proposed method has 5 main parameters: (i) number of pyramid levels, (ii) number of iterations per level, (iii) patch size, (iv) stride, and (v) group size. Increasing the number of pyramid levels makes it possible to calculate larger deformations, however, at the cost of increasing the registration time. The number

of iterations per level is important for registering fine details, but similarly to the number of resolutions, increasing the value leads to longer registration time. The patch size is connected with the deep network architecture, its value should be chosen to correctly utilize the network receptive field. The stride defines how much the unfolded patches overlap. Finally, the group size defines the number of patches registered simultaneously. The larger the value, the faster the registration, as well as the GPU memory consumption. The patch size, stride and group size were established by calculating the theoretically required receptive field. The number of pyramid levels and number of iterations per level were tuned by a simple brute force search (the range of reasonable value is low).

The framework was trained using the Adam optimizer, with a fixed number of epochs, without using the early stopping technique. Only the images denoted as training by the challenge organizers were used during training, while the evaluation set was used for the validation. However, no decision was made based on the validation set results. Overfitting was not observed. The training set was augmented by small, random affine and color transformations. Schedulers were used, different for each pyramid level, decreasing the learning rate by a given factor after each epoch. No information about the landmarks was used during training. We make the source code freely available [18].

3 Results

The dataset consists of 481 image pairs, 230 in the training and 251 in the evaluation set. The landmarks are provided only for the training set, the evaluation of the remaining images must be done using the challenge platform, independently of the method authors. There are 8 different tissue types stained using 10 distinct dyes. The images vary from 8k to 16k pixels in one dimension. The full dataset description, including details about the acquisition, landmarks annotation, tissue abbreviations, and image size, is available at the challenge website [3].
The evaluation metric used to compare the participant's method is based on

Table 1: The normalized average processing time (in minutes) calculated by the automatic ANHIR evaluation system. The registration time is reported for RTX 2080 Ti.

Ours	MEVIS	AGH	UPENN	TUNI	CASIA	DROP	CKVST	BTP
0.033	0.141	8.596	1.374	8.977	4.824	3.388	7.488	0.684

the target registration error, divided by the image diagonal, defined as:

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

where TRE denotes the target registration error, w is the image width and h is the image height. We use the median of rTRE to compare the methods, following the original challenge rules. The average difference between two landmarks

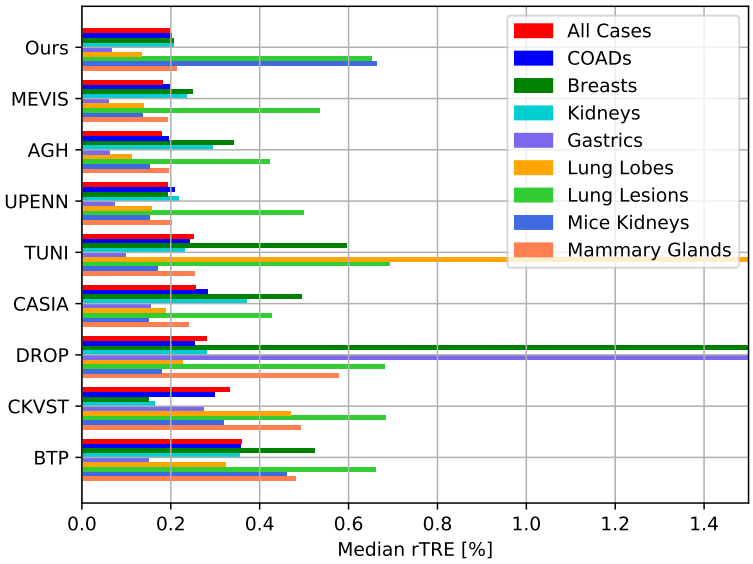


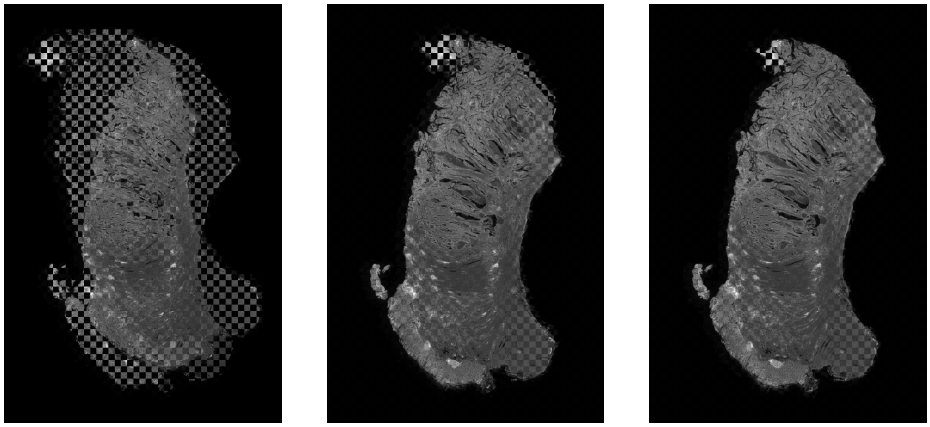
Fig. 2: The median rTRE calculated by the ANHIR evaluation system for all cases and each tissue separately. For team and tissue abbreviations see [2,3].

annotators was 0.05% while the best scoring methods achieve accuracy at the level of 0.2% of the image diagonal [2].

In Figure 2, we compare our method to the best solutions in terms of the median rTRE, for all cases together and for each tissue type separately. We also show the normalized average registration time in Table 1. An exemple checkerboard visualization of the images before the registration, after the affine registration and after the proposed nonrigid method is shown in Figure 3. We decided to compare only the solutions submitted using the automatic, server-side evaluation tool. We chose to not use other classical/deep algorithms and tune the parameters because the results would be unintentionally biased towards our method.

4 Discussion and Conclusion

The proposed framework achieves results comparable to other well-performing methods proposed for the nonrigid histological registration. The results in terms of the median rTRE are comparable to the best-scoring methods. Only the results for mice kidneys are considerably worse. The reason for this is an extremely low number of training samples available for the mice kidneys. In general, it can be observed that the larger the number of training samples for a particular tis-



Source/Target (Initial) Source/Target (Affine) Source/Target (Nonrigid)

Fig. 3: Exemplary checkerboards for visual quality assessment at different registration stages (high quality pictures, best viewed zoomed in electronic format).

sue type, the more accurate the results. Similar challenges are related to the lung lesions. However, there is also a problem connected with the regularization parameter that should be different for this type of tissue. An adaptive regularization would be beneficial.

The proposed solution is significantly faster and thus potentially more useful in practice since registration time above several minutes for a single pair is usually unacceptable. The only method that can be directly compared to our solution in terms of the computational time was proposed by the MEVIS team [8]. They proposed not only the most accurate but also a greatly optimized method. Unfortunately, their solution is commercial and the source code is unavailable. All other methods used rather home-designed software and we think that their computational time could be significantly decreased. However, most probably even the most efficient optimizations and use of GPU would not shorten the processing below time required by just few passes through a deep network.

In future work, we plan to investigate other similarity metrics like MIND [14] or NGF [9] since NCC is less resistant to missing structures and without a proper regularization or diffeomorpishm enforcement may introduce folding in such regions. Nonetheless, both similarity metrics are hard to use in patch-based deep frameworks and require an adaptive regularization. What is more, the noise parameter in NGF needs to be tuned, as well as the radius and the neighborhood type of the MIND descriptor. Moreover, we will investigate the network itself, since recent advances in encoder-decoder architectures may further improve the results. Finally, the curvature regularization is not perfect for tissues like lung lesions or mammary glands. We will look for alternative, adaptive regularization possibilities.

To conclude, we propose a fully automatic deep learning-based nonrigid registration method for high resolution histological images, working independently of the image resolution and achieving results comparable to other well-performing ANHIR methods while being significantly faster, thus potentially more useful in clinical practice.

Acknowledgments

This work was funded by NCN Preludium project no. UMO-2018/29/N/ST6/00143 and NCN Etiuda project no. UMO-2019/32/T/ST6/00065.

References

1. Borovec, J., Munoz-Barrutia, A., Kybic, J.: Benchmarking of Image Registration Methods for Differently Stained Histological Slides. IEEE International Conference on Image Processing (2018) 3368–3372
2. Borovec, J., et al.: ANHIR: Automatic Non-rigid Histological Image Registration Challenge. IEEE Transactions on Medical Imaging (2020)
3. Borovec, J., et al.: ANHIR Website. `https://anhir.grand-challenge.org`
4. Fernandez-Gonzalez, R., et al.: System for combined three-dimensional morphological and molecular analysis of thick tissue specimens. Microscopy Research and Technique **59**(6) (2002) 522–530
5. Gupta, L., Klinkhammer, B., Boor, P., Merhof, D., Gadermayr, M.: Stain independent segmentation of whole slide images: A case study in renal histology. IEEE ISBI (2018) 1360–1364
6. Mikhailov, I., Danilova, N., Malkov, P.: The immune microenvironment of various histological types of ebv-associated gastric cancer. Virchows Archiv (2018)
7. Bueno, G., Deniz, O.: AIDPATH: Academia and Industry Collaboration for Digital Pathology. `http://aidpath.eu`
8. Lotz, J., Weiss, N., Heldmann, S.: Robust, fast and accurate: a 3-step method for automatic histological image registration. arXiv:1903.12063 (2019)
9. Haber, E., Modersitzki, J.: Intensity Gradient Based Registration and Fusion of Multi-modal Images. MICCAI 2006 (2006) 726–733
10. Venet, L., Pati, S., Yushkevich, P., Bakas, S.: Accurate and Robust Alignment of Variable-stained Histologic Images Using a General-purpose Greedy Diffeomorphic Registration Tool. arXiv:1904.11929 (2019)
11. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage **23** (2004) 151–160
12. Yushkevich, P., 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
13. Wodzinski, M., Skalski, A.: Automatic Nonrigid Histological Image Registration with Adaptive Multistep Algorithm. arXiv:1904.00982 (2019)
14. Heinrich, M., et al.: MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. Medical Image Analysis **16**(7) (2012) 1423–1435

15. Zhao, S., Lau, T., Luo, J., Chang, E., Xu, Y.: Unsupervised 3D End-to-End Medical Image Registration with Volume Tweening Network. IEEE Journal of Biomedical and Health Informatics (2019) (Early Access).
16. de Vos, B., Berendsen, F., Viergever, M., Sokooti, H., Staring, M., Isgum, I.: A deep learning framework for unsupervised affine and deformable image registration. Medical Image Analysis **52** (2019) 128–143
17. Fan, J., Cao, X., Wang, Q., Yap, P., Shen, D.: Adversarial learning for mono- or multi-modal registration. Medical Image Analysis **58** (2019)
18. Wodzinski, M.: The Source Code. `https://github.com/lNefarin/DeepHistReg`
19. Wodzinski, M., Müller, H.: Learning-based affine registration of histological images. 9th International Workshop on Biomedical Image Registration (WBIR) (2020)
20. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI 2015 (2015) 234–241
21. Wu, Y., He, K.: Group Normalization. arXiv:1803.084943 (2018)
22. Fischer, B., Modersitzki, J.: Curvature based image registration. Journal of Mathematical Imaging and Vision **18**(1) (2003) 81–85