Implementation of an Unsupervised Deep Learning Registration Framework for Histology Samples with Various Staining

Valentina Pucci, Lorenzo Sciarretta, Mattia Cestari
Politecnico di Milano, Italy
{ valentina1.pucci, lorenzo.sciarretta, mattia.cestari }@mail.polimi.it

Abstract

Image registration, the process of aligning images from different sources or times, is crucial in clinical settings for both diagnosis and therapy guidance. Algorithms designed for this purpose often utilize joint intensity histogram-based methods due to their adaptability. It's important to differentiate between rigid-body and non-rigid transformations, especially for inter-subject and non-rigid intra-subject registrations.In this article, we propose a deep learning-based solution for histology image registration. Our approach includes image preprocessing, initial alignment, and learning-based affine/non-rigid registration. We focus on enhancing preprocessing techniques and developing user-friendly software to facilitate these processes.

I. Introduction

In the context of the expanding digitization of pathology departments in hospitals, aligning histology images from various stains presents a significant challenge. Accurate registration is crucial for comparing different biomarkers and enabling precise quantitative analysis. Traditional methods, including iterative and nonrigid registration approaches, often struggle to address these challenges effectively. Recent advancements in deep learning and hardware technology have revolutionized quantitative medical image analysis, yet implementing these techniques remains challenging due to the complexity and large size of histopathology images.

In this study, we propose the implementation of deep learning-based algorithms developed for the *ANHIR Grand Challenge* [1], which focuses on comparing the accuracy and speed of automatic non-linear registration methods on a set of large images from the same tissue samples but stained with different biomarkers. Specifically, we encountered two primary challenges: the large size of histopathology images, which requires substantial computational resources and time, and the need to develop a fully automated method.

To address these challenges, we leveraged deep learning techniques and GPUs, enabling us to overcome the limitations of standard methods. Our approach includes preprocessing, initial alignment, and learning-based affine/non-rigid registration, ultimately leading to a robust and efficient solution for histology image registration.

A. Related Work and Contributions

Medical image registration has seen numerous advancements over the years, with countless significant contributions. However, despite these strides, the registration of histology images remains challenging, often resulting in limited efficacy of state-of-the-art algorithms. This challenge prompted the creation of the *ANHIR Grand challenge*, where organizers provided a huge *dataset* of High-resolution (up to 40x magnification) whole-slide images of different types of tissue (lesions, lung-lobes, mammary-gland). These images are organized into sets of consecutive tissue slices, with each slice stained by a different dye. The task aimed to develop an automatic nonlinear registration method, a real issue due to non-linear deformations affecting the tissue samples, variations in appearance due to different stains, repetitive textures, and the large size of the whole slide images.

In this work, we propose an unsupervised deep learning-based registration framework dedicated to this dataset, built on the previous work [2] of Wodzinski and Müller. Additionally, we studied other methods used in the challenge. For example, we examined the work of the AGH University team [3], who proposed a multistep approach, utilizing a Demons-based solution with the modality independent neighborhood descriptor (MIND [4]) as the cost function for nonrigid registration. This work was quite efficient but not as precise as the deep learning-based approach. Moreover, it had two main problems: firstly, the framework works only on Linux OS (due to the MIND library), and secondly, it does not exploit the power of GPUs. Other significant work for us was [5] that presented a fast, learning-based framework for deformable image registration.

II. Methods

A. Overview

Our implementation follows the structure proposed by [2]. The framework pipeline consists of methods for data loading, preprocessing, rotation alignment, affine registration, and nonrigid registration. Each method is described in the following subsections. During each registration step we employ the *negative* normalized cross-correlation (NCC) as the cost function, which measures the similarity between two images.

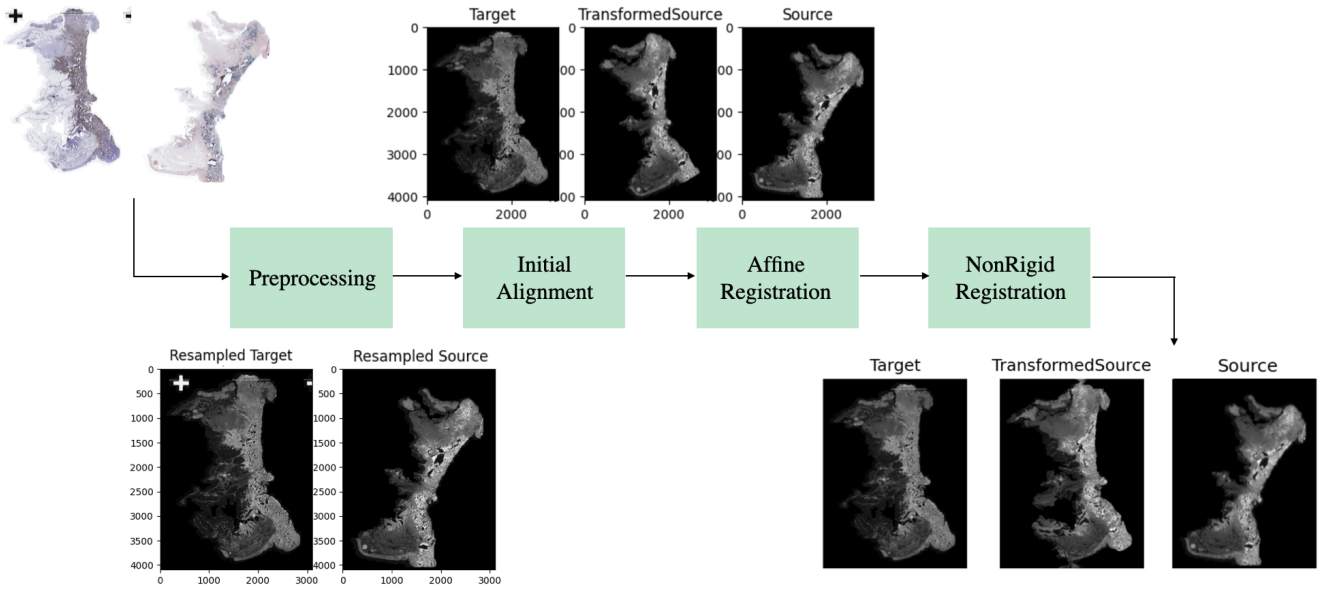

Figure 1. Visual description of our workflow from preprocessing to final results

It quantifies the similarity between images by comparing their intensity distributions while accounting for variations in illumination and contrast. NCC is commonly utilized in image processing tasks such as image registration. For two images I_1 and I_2, with pixels $I_1(x, y)$ and $I_2(x, y)$, the NCC between the two images is given by:

$$\text{NCC}(I_1, I_2) = \frac{\sum_{x,y}(I_1(x,y)-\bar{I}_1)(I_2(x,y)-\bar{I}_2)}{\sqrt{\sum_{x,y}(I_1(x,y)-\bar{I}_1)^2}\cdot\sqrt{\sum_{x,y}(I_2(x,y)-\bar{I}_2)^2}} \quad (1)$$

where $\bar{I}_1$ and $\bar{I}_2$ are the means of pixel intensities in the respective images.

B. Preprocessing

The preprocessing step is crucial for image registration because it prepares the input images in a standardized and enhanced format, optimizing them for accurate alignment. Initially, we convert the images from the dataset formats (jpg/png etc.) into *.mha* (MetaImage Header) format, which offers advantages in preserving data quality, supporting volumetric data, and accelerating data loading. Following this, all images are converted to grayscale, downsampled, and smoothed with a *Gaussian Filter*, aiming to enhance uniformity in resolution and noise levels for improved alignment accuracy.

In the context of image registration, ”source” and ”target” images refer to two sets of images that need to be aligned or registered with each other. The source image set typically consists of images that require transformation or alignment to match the characteristics of the target image set. The target image set serves as the reference or baseline to which the source images are aligned. The goal is to adjust the spatial orientation or appearance of the source images to closely match those of the target images. Therefore, the next preprocessing step involves histogram equalization of both the source and target images based on their image entropy, which measures the complexity or variability of an image. This equalization process aims to enhance image quality by adjusting the distribution of pixel intensity levels, which has been demonstrated to improve the quality of results. After computing the entropies of both source and target images, their mean values are compared. If the mean entropy of the target image surpasses that of the source, histogram equalization is applied to align the source image more closely with the target, and vice versa. This adaptive approach ensures greater coherence between images, facilitating more effective registration processes.

Then, the tissue in the downsampled images is segmented from the background using a U-Net-based network, following the methodology outlined in [2]. This is a fast method that significantly improves the registration results for images with background artifacts. An example of preprocessed source and target is in Fig. 2.

C. Initial Alignment

The initial alignment is the first phase of the registration. Here, it is not not used deep learning, but it is used a standard approach, since we are doing just a rotation and translation. Moreover, a standard approach, using the GPU, is faster, since it has to modify just one parameter, the rotation angle. The process of the alignment begin with finding the centroids of both the source and the target images. Then, the source image is translated to match the position of the target centroid. Next, a search is conducted to find the optimal rotation angle: the source image is rotated at each angle, and the NCC metric is calculated between the source and target images. The angle

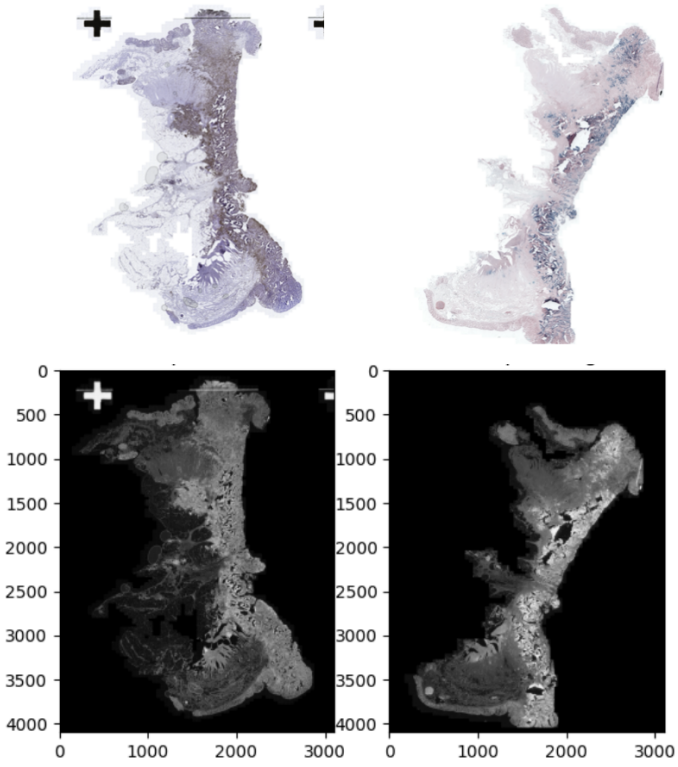

Figure 2. This figure displays the source and target images as they appear before and after the preprocessing steps. The top row shows the original source and target images, while the bottom row illustrates the same images after preprocessing. This preprocessing is crucial for aligning the images accurately in subsequent registration steps, ensuring that both images are better prepared for the affine and non-rigid registration processes

resulting in the lowest negative NCC value is chosen. In conclusion, the complete transformation is done by combining the translation and the rotation of the source image.

D. Affine Registration

The affine registration step employs a convolutional neural network similar to the ResNet architecture. During this stage, the input images are processed at their original resolution. The network generates a 2x3 affine transformation matrix, which is then converted into a spatial transformer grid. For training purposes, the network uses NCC as the loss function.

E. Non-rigid Registration

Nonrigid registration in histology is particularly challenging because maintaining fine detail accuracy is difficult with simple, single-shot networks. These networks are hindered by high-resolution image data that exceeds GPU memory capacities. Even when breaking down the problem into smaller patches and processing them together in batches, the combined size of these batches still exceeds the available GPU memory capacity.

To address this, a pyramid-based, patch-based, group-based, and iterative deep registration method is proposed.

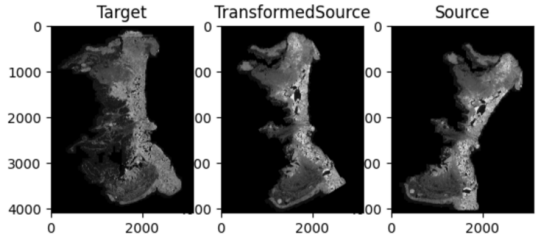

Figure 3. This figure illustrates the results of the affine registration step. The target image serves as the reference that the source image aims to align with. The transformed source image demonstrates the outcome of applying affine transformations to the source image to closely resemble the target image

The pyramid-based approach registers images at different resolutions, starting from the coarsest level. The calculated deformation fields at each level are upsampled to the next finer resolution iteratively. In the patch-based method, images are divided into smaller patches manageable by a smaller network. Group-based processing involves propagating only small groups of patches through the network at once, with the loss function evaluated and optimized at the group level.

The nonrigid registration process begins by creating resolution pyramids for both the source and target images. Registration starts at the coarsest level and proceeds iteratively for a set number of iterations. Initially, the source image is warped using the current deformation field, starting with an identity transformation. During each iteration of the nonrigid registration process, images are partitioned into overlapping patches. These patches are then grouped together and processed through the network to calculate velocity fields. The stride of overlapping patches, set to half the patch size, slightly increases registration time but reduces deformation field discontinuities at patch boundaries.

The loss function combines Normalized Cross-Correlation (NCC) and a curvature regularization term, which penalizes excessive deformations and promotes smoothness in the deformation field. After processing all patch groups, the velocity fields are concatenated and folded back into the deformation field. This deformation field is then upsampled to the next resolution level, with the final deformation field being achieved at the highest resolution.

Several parameters influence the method: patch size, stride, group size, number of pyramid levels, iterations per level, and regularization parameter. Smaller patch sizes and strides reduce network parameters and GPU memory requirements but limit deformation magnitude. The number of pyramid levels ensures patches capture large deformations, while more iterations per level enhance registration accuracy. The regularization parameter controls deformation smoothness.

F. Dataset and Experimental Setup

The framework is implemented using PyTorch, version 2.2.2 . The non-rigid step runs on the T4 GPU provided by Colab, while all other steps run on the CPU of a MacBook Pro

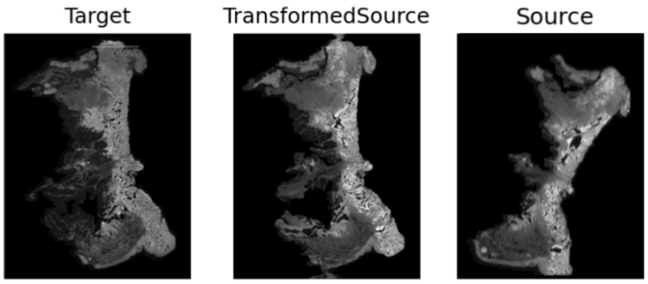

Figure 4. This figure showcases the results of the non-rigid registration process. The target image remains the reference for alignment. The transformed source image illustrates the source image after undergoing non-rigid registration, which includes complex deformations to achieve a more precise match with the target image. An example of this deformation is the dark grey cloud present in the transformed source image that is not present in the original source image. This example highlights how non-rigid registration can adjust local features to better align the images

M1 2022. We utilized the open ANHIR dataset to assess the proposed framework. This dataset comprises 481 image pairs, divided into 251 evaluation pairs and 230 training pairs, and includes eight tissue types: mammary glands, colon adenocarcinomas (COADs), gastric mucosa and adenocarcinomas, breast tissue, mouse kidney, human kidney, lung lesions, and lung lobes.

The dataset's providers resampled the images to about 25% of their original resolution, resulting in sizes ranging from 6,000 to 17,000 pixels in one dimension.

Annotations for the dataset were provided by nine qualified annotators, who selected an average of 86 landmarks per image. The average discrepancy between landmarks chosen by two annotators is 0.05% of the image diagonal, indicating human-level accuracy and providing a benchmark below which registration methods are indistinguishable. Corresponding landmarks are publicly available only for the training images; for the evaluation set, only the source image landmarks are released. For this reason, in our setting we use part of the training pairs for evaluation purposes. In particular, the original dataset was reorganized as follows:

- 200 training pairs, taken both from the original training and original evaluation images.
- 26 test pairs, taken from the original training dataset (both landmarks available)
- 26 validation pairs, taken from the original training dataset (both landmarks available)

The proportions of the different tissue types is maintained across the three groups. Employing regular computing power, in order to allow the completion of the training process in a reasonable amount of time, the size of the dataset was reduced to only 252 images.

Execution time [minutes]	Initial TRE Median	Final TRE Median
1.5656763354937235	0.1400795178125555	0.0176205924930984
1.5673068006833395	0.0824279735978108	0.0203751899297307
1.5838518818219502	0.1472066131869088	0.0190702293953639
1.595568064848582	0.0663763849541152	0.0252215357771753
1.5659751613934836	0.0184342484243306	0.0043046300231535
1.5704150120417275	0.0615828127394805	0.0138029537313664
1.5847459355990092	0.0431323383253378	0.0092946026606601
1.55560671488444	0.0929340169418997	0.0144318544079373
1.5449019511540731	0.0732238874291217	0.0358304848925689
1.5846077005068462	0.0630937374802031	0.0250176421839863
1.58082621494929	0.0341424700451668	0.0376016449920944

Figure 5. Initial and Final TRE with DeepHistReg [2] Approach This figure illustrates the initial and final Target Registration Error (TRE) for a selected pair of images registered using the DeepHistReg approach. The DeepHistReg method, while effective, serves as a benchmark against which improvements can be measured. The reduction in TRE from initial to final values indicates the performance of the registration method on these specific image pairs

Execution time [minutes]	Initial TRE Median	Final TRE Median
1.6104047815004985	0.1400795178125555	0.0146427881609139
1.6294115781784058	0.0824279735978108	0.0104763488517222
1.6339345852533975	0.1472066131869088	0.0172765759596078
1.62514728307724	0.0663763849541152	0.0259598070997587
1.6280463020006817	0.0184342484243306	0.0039454997744215
1.610827620824178	0.0615828127394805	0.0162165048581792
1.6521146138509115	0.0431323383253378	0.0186432164614721
1.6275380730628968	0.0929340169418997	0.0071452466478173
1.629325302441915	0.0732238874291217	0.0211955892065038
1.639236831665039	0.0630937374802031	0.0104335435834585
1.651493036746979	0.0341424700451668	0.0232271835339816

Figure 6. Initial and Final TRE with Our Approach This figure displays the initial and final Target Registration Error (TRE) for a pair of images registered using our newly proposed method. Our approach incorporates enhancements such as histogram equalization based on image entropy, which were not present in the DeepHistReg method. The significant reduction in TRE from the initial to final values demonstrates the effectiveness of our approach in achieving more accurate image registration

G. Results

Our main contribution lies in reorganizing the code from the original DeepHistReg[2] repository, which lacks a comprehensive guide for reproducing results and documentation explaining the various registration steps. We provide a clear and detailed pipeline in the form of a Jupyter Notebook, enabling users to replicate the experiment rapidly and out-of-the-box. Our sequential, step-by-step approach facilitates the easy replacement of individual stages, offering a solid foundation for further experimentation. Leveraging this pipelined structure, we implemented an additional preprocessing stage of histogram equalization based on image entropy, which

was absent in the original work. This enhancement positively impacted the resulting images, reducing the overall median Target Registration Error (TRE), as shown in Figures 5 and 6 (where TRE is the evaluation metric used in the ANHIR Grand Challenge).

Execution time [minutes]	Initial TRE Median	Final TRE Median
1.3686714331309	0.0380121100330073	0.0170565114255214
1.350566577911377	0.0466791303689456	0.0232233778558197
1.6670307477315267	0.0671009355468272	0.0039262974351448
1.6337249318758646	0.0756675388460972	0.0281027418285291
1.5438604354858398	0.0682888441022998	0.0425295772626168
1.363205262025197	0.0424193005371604	0.0138954372642007
1.364901332060496	0.0366452709175384	0.0600866434619966
1.3875412027041116	0.0306407801148482	0.0771815861904488
1.555924133459727	0.0588634062443045	0.0137837338603803
1.382131004333496	0.0457498796484877	0.0070088759829961
1.5543039997418722	0.0687757425936889	0.0127553259418257

Figure 7. TRE Results on Unseen Test Images This figure presents the results obtained with a set of test images that were not used during the model's training phase. The final TRE is notably lower than the initial TRE, showcasing the robustness and generalization capability of our registration method. The consistent reduction in TRE across these unseen images underscores the method's potential for broader applicability in various image registration tasks

H. Conclusions

In conclusion, we enhance the DeepHistReg[2] repository by reorganizing the code and providing a detailed Jupyter Notebook pipeline, which improves reproducibility and facilitates further experimentation. All our work is available open-source on GitHub in our repository [6].
Our introduction of histogram equalization based on image entropy reduces the overall TRE. This framework offers a solid foundation for future research in image registration.

References

[1] ANHIR Grand Challenge, 2019. https://anhir.grand-challenge.org/Data/.

[2] Marek Wodzinski and Henning Müller. Deephistreg: Unsupervised deep learning registration framework for differently stained histology samples. *Computer Methods and Programs in Biomedicine*, 198:105799, 2021. https://www.sciencedirect.com/science/article/pii/S0169260720316321, GitHub: https://github.com/MWod/DeepHistReg.

[3] Philippe Weitz. Repository that contains the software used by the agh team during the anhir challenge organized jointly with the ieee isbi 2019 conference. https://github.com/MWod/ANHIR_MW?tab=readme-ov-file.

[4] Mattias P. Heinrich, Mark Jenkinson, Manav Bhushan, Tahreema Matin, Fergus V. Gleeson, Sir Michael Brady, and Julia A. Schnabel. Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Medical Image Analysis*, 16(7):1423–1435, 2012. Special Issue on the 2011 Conference on Medical Image Computing and Computer Assisted Intervention.

[5] Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. Voxelmorph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, 38(8):1788–1800, 2019.

[6] V. Pucci L. Sciarretta and M. Cestari. Online repository github of our work. 2024. https://tinyurl.com/45h3yxwb .