

NEMESIS: NEural iMplicit rEpresentations for whole Slide Image regiStration

Bruno De Santi

*Multi-Modality Medical Imaging, Technical Medical Centre University of Twente
Enschede, The Netherlands
b.desanti@utwente.nl*

Jelmer M. Wolterink

*Department of Applied Mathematics, Technical Medical Centre University of Twente
Enschede, The Netherlands
j.m.wolterink@utwente.nl*

Abstract—This technical report provides a description of the algorithm proposed by team NEMESIS for the AutomatiC Registration Of Breast cAncer Tissue (ACROBAT) challenge organized at MICCAI 2022.

Index Terms—Image Registration, Neural Implicit Representation, ACROBAT challenge

I. INTRODUCTION

In this technical report, we provide a description of our algorithm for the AutomatiC Registration Of Breast cAncer Tissue (ACROBAT) challenge organized at MICCAI 2022. The algorithm is able to align whole-slide images (WSIs) of breast cancer tissue sections that were stained with immunohistochemistry (IHC) to corresponding tissue regions that were stained with haematoxylin and eosin (H&E) in two steps: i) initial rigid registration based on scale-invariant feature transform (SIFT) features and a modified random sample consensus (RANSAC) algorithm, ii) deformable image registration using coordinate networks, following our recent paper present at the Medical Imaging with Deep Learning (MIDL) 2022 conference [1].

II. METHODS

The algorithm we propose consists of two stages that are jointly implemented in a pipeline:

- 1) Initial rigid registration estimation using SIFT keypoints and RANSAC
- 2) Deformable image registration using implicit neural representations (INRs)

A. Initial rigid registration

1) *Preprocessing*: For each case, the first level of the pyramidal tiff files (scale 10x) of the H&E and the IHC image were loaded. In this first stage of the algorithm, the images were downsampled by a factor equal to 10. Before extracting the SIFT keypoints, the H&E RGB image was converted to the L*a*b* colorspace and the channel L was selected. The preprocessing consisted of four steps: i) a contrast limited adaptive histogram equalization (CLAHE), ii) Gaussian smoothing filtering with $\sigma = 5 \mu\text{m}$, iii) taking the inverse of the image and iv) cropping the image to remove possible artefacts in the image edges (6 % of the image size for both image axes). Regarding the IHC preprocessing, after conversion from RGB to HSV

colorspace, the aforementioned four preprocessing steps were applied to the V channel. The choice of these channels was made based on visual examination on a subset of the training set. These preprocessed channels were used as input for the SIFT algorithm.

2) *SIFT extraction and matching*: The SIFT algorithm was used to extract relevant keypoints in both images [2]. The OpenCV package implementation with default parameters was adopted. For the keypoints matching, the Flann K-Nearest Neighbour matching method implemented in OpenCV was used ($K = 2$). Matches with a Lowe's ratio below 0.95 were discarded as possible false matches.

3) *Tissue segmentation*: In order to improve the performance of the inlier detection we segmented the tissue in the whole slide. In both H&E and IHC images, tissue was segmented using the k-means algorithm ($k = 3$). The cluster with the highest average intensity was classified as the white background, hence, the tissue mask was defined as the union of the other two clusters.

4) *Inlier matches detection using RANSAC*: RANSAC is an iterative method used to estimate a model from data containing outliers [3]. In this work, we used RANSAC to exclude outlier matches from the total set of matches. The transformation model we adopted was a rigid transform. For each iteration, the RANSAC algorithm randomly samples three matches, computes the rigid transform with the least squares method and computes the registration error (euclidean distance) for every match. Matches with a registration error less than 50 pixels were defined as inliers. Additionally, for each iteration, the IHC tissue mask was transformed and the overlap, as Dice score, with the H&E tissue mask was computed. After 120 000 iterations, the best iteration was chosen by maximizing the following objective function:

$$O = \hat{n}_i^2 + \hat{D}_i^2 \quad (1)$$

where n_i and D_i are the number of inlier matches and the Dice score, respectively, at iteration i . The \mathbf{n} and \mathbf{D} arrays were min-max scaled to obtain $\hat{\mathbf{n}}$ and $\hat{\mathbf{D}}$. Also, a scale constraint was applied: iterations with a scale of the rigid transformation outside the range [0.95, 1.05] were discarded.

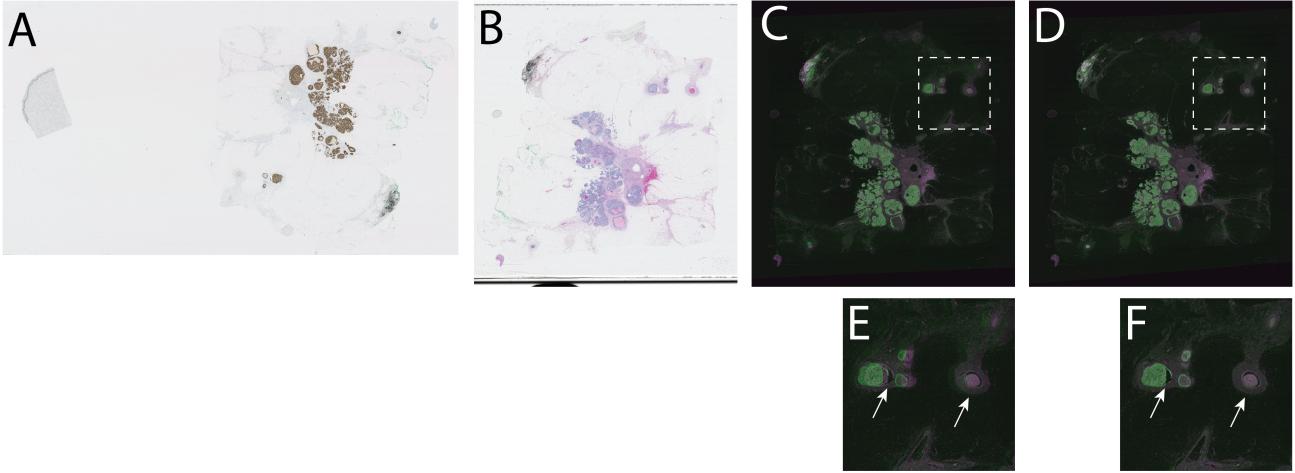


Fig. 1. Results of the registration algorithm for image pair 19 of the validation set. (A) IHC (Estrogen receptor) WSI, (B) H&E WSI, (C) Overlay after the initial rigid registration, (D) Overlay after INR deformable registration, (E) and (F) show a zoom-in on the white dashed regions, white arrows show fine scale misalignments corrected by the INR deformable registration. In the overlay image, the H&E image is encoded in the red channel and the IHC image in the green channel.

Finally, the rigid transformation with the maximum score was applied to the preprocessed IHC V channel to obtain an initial coarsely aligned image.

B. Deformable image registration using INRs

In implicit neural representations (INRs), a function on a spatial domain is implicitly represented by the weights of a neural network that operates on coordinates in the domain. Recently, INRs have been showing great potentials in image synthesis and 3D scene rendering thanks to its ability of representing an image as a function defined on a continuous spatial domain [4]–[6]. Recently, our research group proposed the use of INRs in deformable image registration [1]. A multi-layer perceptron network (MLP) can be optimized to implicitly represent a function ϕ mapping coordinates from a moving to a fixed image. This has several advantages compared to traditional and convolutional neural network based methods. First, the deformation field is continuously defined over the image domain with a fixed number of parameters. Second, the method is not restricted to any particular image resolution and does not require the expensive task of computing discrete spatial gradients. This is particularly appealing in high-resolution whole-slide images. Third, one can control the ability of representing large or small local deformations by modifying the activation functions of the MLP.

In this study, we optimized a MLP to find a transformation $\phi(\mathbf{x}) = u(\mathbf{x}) + \mathbf{x}$ such that coordinate \mathbf{x} in the IHC image corresponds to coordinate $\phi(\mathbf{x})$ in the H&E image. For further details please refer to the original paper [1]. In this work, we used an MLP with three hidden layers, each of which contained 256 units with ReLU activation functions. Normalized cross correlation was used as loss function. An Adam

optimizer was used with a learning rate of $1e^{-5}$ and 25 000 epochs.

The preprocessed (same procedure described in II-A1) H&E L channel and the rigidly aligned IHC V channel at scale 1x were used as the fixed and moving image, respectively. During the MLP optimization, only coordinates belonging to the H&E tissue mask were sampled. In order to include also tissue edge information, the H&E tissue mask was dilated with a square structuring element of 51x51 pixels.

III. RESULTS

The entire pipeline was implemented in PyTorch and is provided in a Jupyter Notebook (Python version 3.9.12). All training and testing was done on a Windows 10 machine with an Intel Core i9-11900K @ 3.5GHz, 128 GB RAM and NVIDIA RTX3090 24GB. The same algorithm and parameters were used for each pair of images. Average time of execution for an image pair was 6.2 ± 1.6 min. The final version of the algorithm which was used for the test set submission is available at github.com/MIAGroupUT/NEMESIS.

Figure 1 shows an example result after the initial rigid registration and the resulting deformed image after deformable registration with the INR. This is a representative case of the validation set. In this case, as in all cases that we inspected, we visually found the INR to improve over the rigid registration result. Note that due to time constraints, we were unable to perform a full quantitative analysis on the validation set, Hence, we eagerly await the quantitative results on the test set as provided by the challenge organizers at the MICCAI 2022 workshop.

REFERENCES

- [1] Jelmer M. Wolterink, Jesse C. Zwienenberg, and Christoph Brune. Implicit neural representations for deformable image registration. In *Medical Imaging with Deep Learning*, 2022.
- [2] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [3] Martin A. Fischler and Robert C. Bolles. Random sample consensus. *Communications of the ACM*, 24(6):381–395, 1981.
- [4] Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *CoRR*, abs/2006.10739, 2020.
- [5] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *CoRR*, abs/2003.08934, 2020.
- [6] Yiheng Xie, Towaki Takikawa, Shunsuke Saito, Or Litany, Shiqin Yan, Numair Khan, Federico Tombari, James Tompkin, Vincent Sitzmann, and Srinath Sridhar. Neural fields in visual computing and beyond. *Computer Graphics Forum*, 2022.