UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola Tècnica Superior d'Enginyeria
de Telecomunicació de Barcelona

Non-Rigid registration of histopathological breast cancer images using Deep Learning

Master Thesis
submitted to the Faculty of the
Escola Tècnica d'Enginyeria de Telecomunicació de Barcelona
Universitat Politècnica de Catalunya
by

Lucia Cavallari

In partial fulfillment
of the requirements for the master in
TELECOMMUNICATION ENGINEERING

Advisor: Veronica Vilaplana Besler
Co-advisor: Josep Ramon Casas
Barcelona, July 2022

Contents

List of Figures

List of Tables

Acronyms

WSI: Whole Slide Image
HE: Hematoxylin and eosin stain
CK19: Cytokeratin 19
PR: Progesterone receptor
GPI: Image and video Processing Group
CT: computed tomography
MR: magnetic resonance
PET: Positron emission tomography
ANHIR: Automatic Non-rigid Histological Image Registration challenge
ACROBAT: AutomatiC Registration Of Breast cAncer Tissue

Abstract

Cancer is one of the leading causes of death in the world, in particular, breast cancer is the most frequent in women. Early detection of this disease can significantly increase the survival rate. However, the diagnosis is difficult and time-consuming. Hence, many artificial intelligence applications have been deployed to speed up this procedure.

In this MSc thesis, we propose an automatic framework that could help pathologists to improve and speed up the first step of the diagnosis of cancer.
It will facilitate the cross-slide analysis of different tissue samples extracted from a selected area where cancer could be present. It will allow either pathologists to easily compare tissue structures to understand the disease's seriousness or the automatic analysis algorithms to work with several stains at once. The proposed method tries to align pairs of high-resolution histological images, curving and stretching part of the tissue by applying a deformation field to one image of the pair.

1 Introduction

1.1 Motivation

Cancer is one of the main cause of death throughout the world. There are over 100 different types of cancers that are categorized on the basis of the affected tissue or organ of the human body. Although each type has its unique features, the basic processes that produce cancer are quite similar in all forms of the disease.
Cancer occurs when the cellular reproduction process goes out of control. Usually normal cells in the body grow and divide as necessary to replace defective or dying cells. Cancer cells are characterized by uncontrolled, uncoordinated and undesirable cell division, they continue to grow and divide, replicating into more and more harmful cells. A clump of cancer cells is known as a tumor and it causes many of the symptoms of cancer by pressuring, crushing and destroying surrounding non-cancerous cells and tissues.

Breast cancer is the most frequent malignancy in women worldwide and is curable in 70–80% patients with early-stage, non-metastatic disease. Advanced breast cancer with distant organ metastases is considered incurable with currently available therapies [2].
The sooner this disease is detected, the more efficient the treatment will be and the most likely the patient will recover. Two important conclusions can be derived from this fact, the disease must be detected and well-classified, and it must be detected as early as possible.
Unfortunately, this analysis is an arduous process that is difficult, time-consuming, and requires in-depth knowledge. For this reason, many artificial intelligence (AI) applications have been deployed to improve and speed up this procedure.

1.2 Histopathological analysis

Before the actual treatment for breast cancer, the most important steps are the diagnosis and the prognosis (the expected development of a disease).

The diagnosis of breast cancer utilizes breast imaging techniques, the most adopted ones are X-ray mammography, echography, magnetic resonance imaging (MRI) and computed axial tomography (CT). Each imaging technique plays its own role corresponding to the stage of tumor, and therefore for many cases a complete diagnosis would involve more than one imaging technique.

For the prognosis, the standard procedure is histopathological tissue analysis which corresponds to the detailed analysis of a biopsy tissue sample performed by a pathologist. Thanks to this last step, it is possible to gather detailed information on the types of cells and on the characteristic of the tumor.

Before conducting the histopathological analysis, there are some steps that need to be executed. [5]

- Tissue extraction: extraction of a tissue sample (specimen) from the area of interest where the tumor is located.
- Tissue fixation: the specimen is placed in a liquid fixing agent (fixative) such as formaldehyde solution (formalin). This will slowly penetrate the tissue causing chemical and physical changes that will harden and preserve the tissue and protect it against subsequent processing steps. [4]
- Specimen transfer to cassettes: specimens are trimmed using a scalpel to enable them to fit into an appropriately labeled tissue cassette.
- Tissue processing: before cutting the sample, it needs to be dehydrated, cleared and embed with a specific agent that make the tissue solid.
- Sectioning: the tissue is cut into sections that are placed on a slide.
- Staining: most cells are transparent and appear almost colorless when unstained. Histochemical stains (typically hematoxylin and eosin) are therefore used to provide contrast to tissue sections, making tissue structures more visible and easier to evaluate. Each stain highlights different information regarding the tumor.

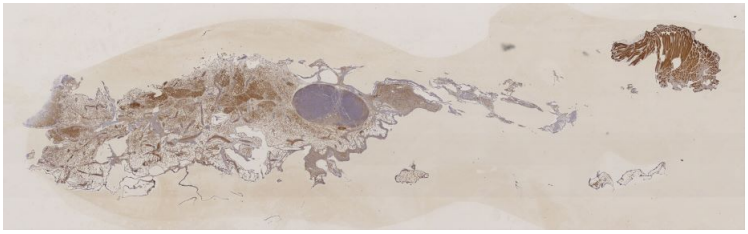
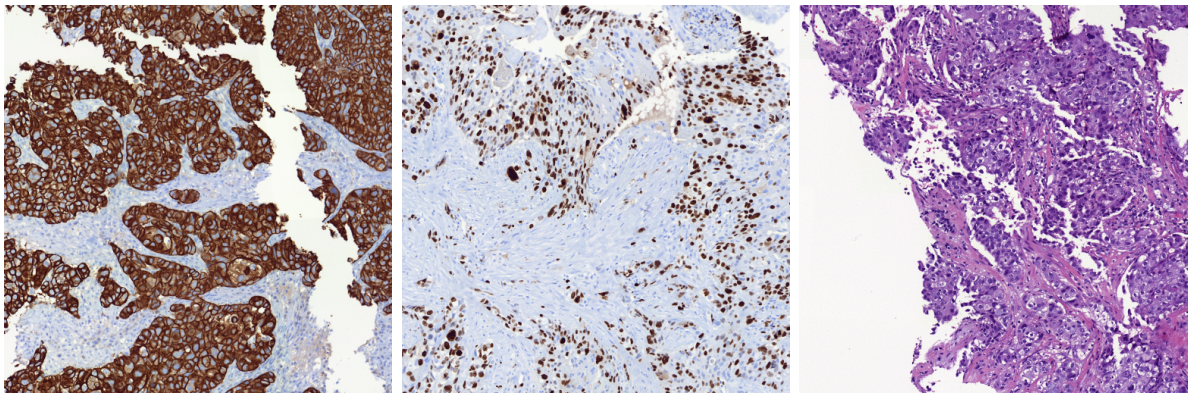


Figure 1: WSI of mammary gland

Finally, each stained slice is scanned and the Whole Slide Images (WSI) (i.e. Fig. 1 [1]) are generated. The WSI are high-resolution digital files that can be efficiently stored, accessed, analyzed, and shared.
Thanks to this process, pathologists can analyze the tissue by navigating through different parts of the image to explore the tissue and to determinate whether tumor cells are present or not. If it is confirmed the presence of these type of cells in the sample, then pathologists need to analyze other slices colored with different stains to obtain more information relevant to the diagnosis and treatment.
Hence, pathologists need to be able to compare tissue slides colored with different stains in order to obtain a good diagnosis and prognosis of the cancer stage and to highlight different structural and/or functional information: they combine visual information obtained from multiple stains from the same tissue sample.
In the figure 2, there are shown different samples of tissue colored with CK19, KI67 and HE stain.



(a) CK19 stain (b) KI67 stain (c) HE stain

Figure 2: Stained images of breast cancer tissues

1.3 Project Objective

The aim of this project is to develop a solution that can perform the alignment of multi-staining slides extracted from the same tissue, this method is called image registration. This task is needed because, in the laboratory during the creation of the slices, they are often stretched, rotated and displaced, causing WSI of the same tissue to be very different among each other.
The disalignments present on slices makes the cross-slide analysis very difficult: the more different the slices are, the harder is for pathologists to compare them and to get a quick diagnosis. As previously mentioned, cross-slide analysis is performed on serial or transverse sections taken consecutively from a tissue block. Each tissue section is stained to highlight particular features of interest in the tissue. In this case, immunohistochemical (IHC) stains are of particular interest, as they indicate the presence and expression of a chosen protein. Expression profiles of the various different markers can usually be compared directly, as each section is only a microscopic distance (typically 3–5 microns) from the neighbouring sections. Thus larger anatomical structures are likely to be present across many sections. Analysis of the expression profiles of multiple markers in a common region of interest, such as a tumour region, can potentially reveal important information about the tumour's molecular composition [8].

In Figure 3, there is displayed an example of two WSIs of a sample tissue extracted from a lung before before any registration method is applied.
Image registration will enable automated analysis algorithms to estimate classes by fusing information from the different stains. Moreover, it will allow pathologists to make more precise and quicker diagnoses from the collected information. For example, when an area of interest is found in a reference stain, the same area has to be retrieved and displayed in the other stain, so the different visual information can be merged and it is possible to obtain an overall view of the cells in the selected area.

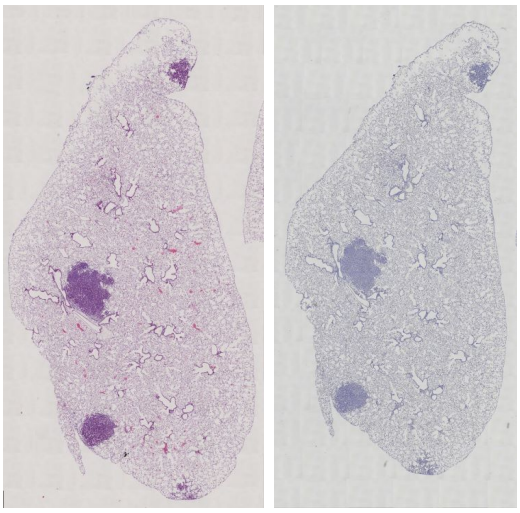


Figure 3: Example of two WSIs before registration [1]

Image registration refers to the process of aligning images so that comparable characteristics can be easily related to one another. During this process, points from one image are mapped to analogous points on another image.

There are two main image registration methods used in fusion biopsy: rigid registration and non-rigid (or elastic) registration.
Rigid registration does not change the images acquired, it preserves the internal architecture of the tissue. Each image set is limited to rotational and translational transformations. On the other side, non-rigid (elastic) registration changes the images acquired by stretching one of the image volumes to match the other.

A rigid registration solution was already developed by another student[7] from the DigiPatics project. In this project, we will combine the rigid registration solution with non-rigid registration approach. The details of the method deployed are going to be explained in the third chapter.

Overall, the main goals of the project are:

1. Understand the medical motivation behind the task.
2. Develop a solution that is able to perform an automatic non-rigid registration, taking as input pairs of tiles where the rigid registration is already applied.
3. Test and evaluate the performance of the solution.
4. Create a framework that combines rigid and non-rigid registration.

1.4 DigiPatics project

This thesis project was done in collaboration with the DigiPatics group at UPC.
DigiPatics is a 4-year project that aims to optimize the anatomopathological diagnosis in the network of hospitals of the Institut Català de la Salut (ICS) through the digitization of the images and the use of artificial intelligence.[3]
It works directly in touch with doctors and researchers, it specifically tries to optimize their resources and improve the quality of the diagnostic process of patients. DigiPatics delivers solutions to pathologists that will allow them to take measurements, create annotations, use image processing tools and apply quantification and computer vision algorithms on images in an easy and fast way.

The Universitat Politècnica de Catalunya, more specifically the Image and Video Processing Group (GPI), is involved in the development of these computer vision algorithms. Other companies are participating in the project such as 3DHistech and Palex.
The overview of the project organization is shown in Figure 4.

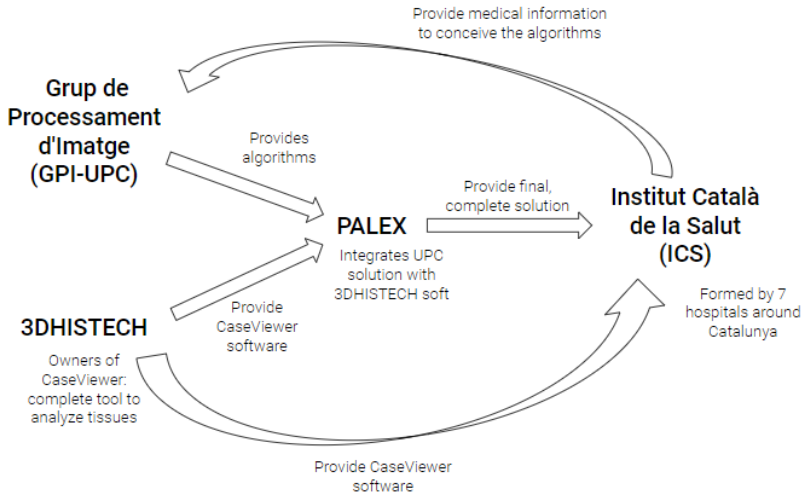


Figure 4: DigiPatics organization flow

The DigiPatics project already explored solutions for image registration on sample of tissues coming from breast cancer only using a rigid approach.
However, in order to improve the results, the requirement of the present project is to develop a solution that could integrate the rigid registration, and to apply also a non-rigid one with the use of neural networks.

1.5 Document Structure

This document is presented in a sequential format and is divided into the following chapters:

- Chapter 2: State of art gives an overview of the previous image registration algorithms deployed in the past with and without the use of deep neural networks. In addition, it explains the main challenges that are currently being faced.
- Chapter 3: Methodology describes the solution and the methods adopted during the development of the thesis.
- Chapter 4: Results starts by the stating the hypothesis being tested and presents the results obtained using different approaches.
- Chapter 6 Conclusions and Future Work states the conclusions drawn from the experiments and presents the challenges met alongside possible projections on the research of image registration.

2 State of art

2.1 Image registration

Image registration, also known as image alignment, is a crucial step in a wide range of applications, including remote sensing, medical imaging, and multi-sensor fusion-based target recognition. It's a fundamental image processing technique that's effective for combining data from various sensors, detecting changes in photos recorded at different times, inferring three-dimensional information from stereo images, and recognizing model-based objects, among other things.

The registration process can be done manually or automatically.
Human operators manually choose corresponding features in the photos to be registered in the first case. To acquire reasonably registration results, an operator must select a significant number of feature pairs over the entire image set, which is not only time-consuming and exhausting, but also prone to inconsistency and restricted precision. As a result, there is a natural need to develop automated techniques that require little or no operator supervision. Over the years, a broad range of automated techniques has been developed for various types of data and problems. Moreover, the growth of neural networks has affected also this field of study, giving the possibility to create more robust solutions.

An important distinction in image registration is the image modality: the reference and aligned image can be uni-modal or multi-modal.
There are many different imaging methods, depending on the physical characteristics needed to study. However, each one of them has also some weaknesses that make image interpretation based on the single image difficult. Therefore, it can be very useful to acquire images using different imaging techniques and then to combine the information gathered from each of them.
Mono-modal registration algorithms are used to register images acquired using the same modality. Multi-modal registration methods concentrate on aligning images originating from different modalities. Such images usually have a totally different appearence[16].

Image registration can be applied not only to 2D images, many applications are able to register 3D images as well. A very interesting multi-modal case is 2D-3D registration: it is performed by generating 2D projections of the 3D volume which are then compared to the 2D image the volume is being registered to. It can be very useful in biomedical applications.

The most relevant distinction is the case of rigid and non-rigid registration.
In the following sections, the main difference between rigid and non-rigid image registration is described. The main techniques used in the biomedical field are explained, with an in depth study of the histological image registration case.

2.2 Rigid vs Non-rigid Registration

Image-registration has traditionally been classified as:

- rigid: where images are assumed to be of objects that can be rotated and translated with respect to one another to achieve correspondence;
- non-rigid: where correspondence between structures in two images cannot be achieved without some localized stretching of the images due to structural differences, image acquisition, or both.

The main difference is that linear transformation functions (rigid registration case) map any straight line to a straight line, whereas nonlinear transformation functions (non-rigid case) maps a straight line to a curve.
Today, rigid registration is frequently extended to incorporate affine registration, which includes scale factors and shears and can partially account for variations in scanner calibration or large scale variances between participants [9].
The main differences between the two types of registration are displayed in figure 5.

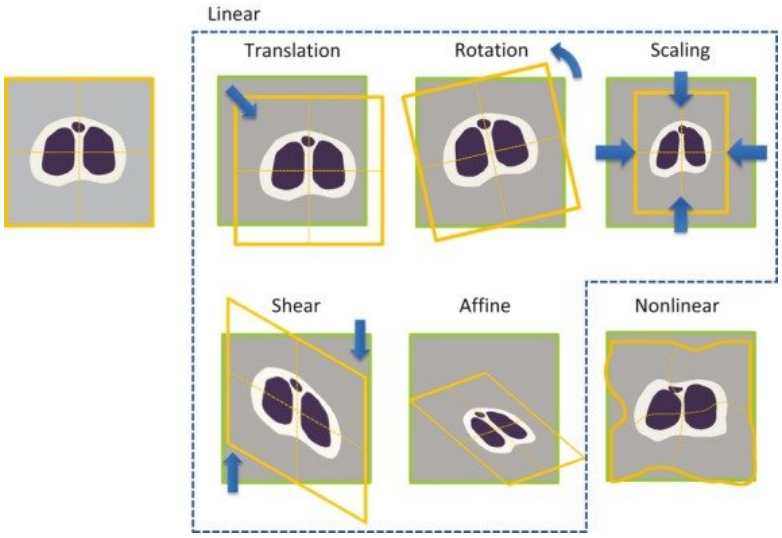


Figure 5: Differences between rigid and non-rigid registration [1]

2.2.1 Rigid Registration

A horizontal and/or vertical shift is the most basic linear transformation function. This shifting transformation is called translation. For picture registration, rotation and scaling (i.e. zoom up/down) are also common linear geometric changes.
Affinity transformation and perspective transformation are two more general geometric transformations. Translation, rotation, scaling, shear, and their arbitrary combinations are all examples of affine transformations. The main property of affine transformation is that parallel lines are still parallel after applying it.

Considering the case where we want to apply a rigid registration among two images A and B, any linear transformation function can be represented as a matrix:

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix} \tag{1}$$

or equivalently:

$$\begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} = \begin{pmatrix} a & b & e \\ c & d & f \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{2}$$

where (x, y) are the coordinates of image A and (X, Y) of image B.
Translation is represented by parameters e and f in x and y direction, respectively.
For rotation:

$$a = d = cos\theta \tag{3}$$
$$-b = c = sin\theta \tag{4}$$

For affine transformation, all six parameters are arbitrary (this explains that affine transformation includes translation, etc., as noted above.)

In rigid registration, the goal is to estimate the parameters of the linear transformation. The estimation criterion is typically described as:

$$\sum_{x,y} ||I_A(x,y) - I_B(W(x,y|a,b,..,f))|| \tag{5}$$

which has to be minimize with respect to the parameters a, b, ..,f.
In the formula, I_A and I_B are the images A and B to register, W() denotes the above geometric transformation function, or the warping function, and maps (x, y) to (X, Y) according to the six parameters a, b, .., f.

It should be noted that, despite its simplicity, the geometric transformation function's estimation problem is not trivial. In truth, there is no direct or analytical method for determining the ideal parameters.

Registration techniques are divided into intensity-based and feature-based methods. Prior to the registration phase, feature-based approaches require the identification or extraction of some features that can be control points, edges, contours, surfaces, prominent features or statistical features. On the other hand, the use of intensity-based approaches does not necessitate the extraction of any features. In this case, raw pixel values are used directly[11].
In most cases, the image registration procedure is carried out in three steps:

1. Feature selection/extraction. Features could be control points, edges, contours, surfaces, salient features, and statistical features[12].
2. Selection of similarity metrics used to determinate the quality of matching between images. The most popular metrics include Mutual Information (MI), Cross Correlation (CC), Sum of Squared Difference (SSD) and absolute difference
3. Definition of a spatial transformation model to determine the positions of corresponding points between images. The transformations involve parameters that may need to be optimized for obtaining the best alignment.

Metric selection is one of the most important decisions in the registration problem. The metric specifies the goal of the process measuring the quality of matching the Target image with the Reference image after application of a transformation on it. The selection of Metric depends on the types of images to be registered and the expected miss-alignment. To obtain the optimum transformation parameters needed to align the images, optimization techniques are used. Good algorithms determine the transformation parameters, hence they determine the quality of the final registration.

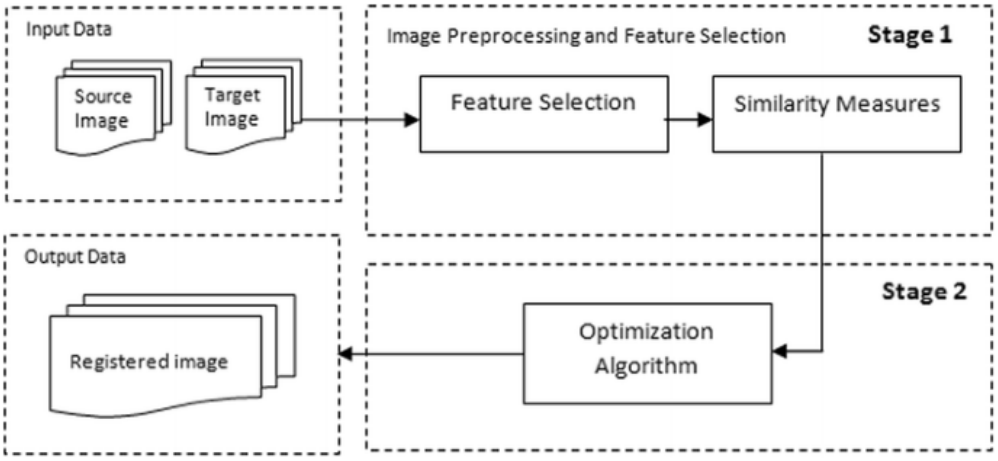


Figure 6: Rigid transformation steps [12]

2.2.2 Non-rigid Registration

Non rigid image registration, realizes more flexible image registration than linear image registration.
The choice of transformation is very specific to the nature of the registration problem, and will depend on the modalities of the imaging systems.
Mathematically, the non-rigid registration problem can be defined as finding the optimal transformation T* such that:

$$T^* = argmin_T C(I, T(J)) \tag{6}$$

Unlike rigid image registration, in which T is restricted to a rigid transformation, for non-rigid image registration there is still no common consensus in the literature regarding how the transformation T should be modeled.
Some models restrict T to be of low degree of freedom, hence, they constrain the transformation to be smooth or elastic.
However, there are cases where the structure of the two images varies significantly, and a transformation of low degree of freedom may not have the flexibility to represent these complex changes. Therefore, any hard constraints on the domain of T should not be imposed.

In equation 6, we are optimizing C(I, T(J)) without adding restrictions on T, in this way, it can map any point in J to any point in I. Thus, it's better to add a penalization function S(T), to penalize transformations that are not smooth.
λ is added as a positive constant that penalizes non-smooth T, it becomes:

$$T^* = argmin_T C(I, T(J)) + \lambda S(T) \tag{7}$$

A non-rigid registration defines a deformation field that gives a translation or mapping for every pixel in the image [13]. This is generally described by the following equation:

$$I_f \circ T(x) = I_f(x - u(x)) = I_r \tag{8}$$

where I_f is the image undergoing the deformation, I_r is the reference image. T denotes the non-rigid transformation which equates to a translation of every pixel x in the floating image by a certain displacement defined by u(x).
The displacement is u: $\mathbb{R}^2 \rightarrow \mathbb{R}^2$, u = ($u_1$, u_2).

A common way to estimate u(x) is through deformation models. They can be roughly split into two groups: [15]

- Deformations derived from physical models, for example elastic body models, fluid flow models and diffusion based models. These models are non-parametric in nature, they allow a per pixel estimation for the deformation.
- Deformation models derived from approximation theory. One popular example can be free form deformation whereby the deformation field is represented using basis functions, such as B-splines at fixed integer grid positions. They are capable of describing a wide range of transformations using a low/limited number of parameters.

In Figure 7, the concept of a displacement field is shown. For every pixel position in the template image, the displacement field gives the direction and the distance it has to move in order to match the reference image. At the end, the field is subsampled.

Many different solutions have been deployed and tested in the last years, however, non-rigid registration continues to be a field of active research and development.

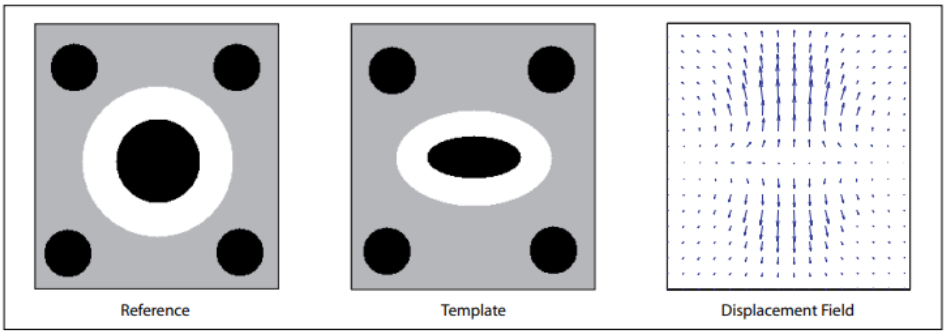


Figure 7: Example of deformation field [16].

2.3 Image Registration in the biomedical field

The expanding variety of medical imaging techniques provides the medical community an increasingly detailed view of functionality and structure of our anatomy. The information provided by the various imaging modalities is often complementary and synergistic (i.e. the combination of information provides useful extra information).
For example, X-ray computed tomography (CT) and magnetic resonance (MR) imaging shows brain anatomy but provide little information on the functional aspect of the brain. Positron emission tomography (PET) and single photon emission computed tomography (SPECT) scans display aspects of brain function and allow metabolic measurements but poorly shows its anatomy. [26] Furthermore, CT and MR images describe complementary morphologic features. For example, bone and calcification are best seen on CT images, while soft-tissue structures are more visible by MR imaging. Clinical diagnosis and therapy planning and evaluation are increasingly based on this complementary image information that can be done using image registration.

That is not the only application in which image registration is adopted: it is utilized in medical image processing for a variety of purposes. Tumor identification, anatomy segmentation, image subtraction for contrast enhanced images and computational model building are just some of the fields in which this technology is used.
Moreover, it plays a vital role in treatment planning. For example, registration is used in the analysis between healthy subjects and ones with brain tumors in order to localize important brain structures to be taken into consideration for surgical planning.
Registration is important even during the surgery itself as it allows for accurate localization of anatomical structures accounting for position shifts induced by surgical operations. In terms of treatment planning, we should also note the impact that registration has in radiotherapy by localizing tumorous cells and thus limiting the destruction of healthy ones. Finally, it has played a crucial role in the analysis of histopathological images.

Image registration, along with image segmentation, is one of the most important problems in the field of image processing. Despite important advances, registration is still considered a challenging problem.

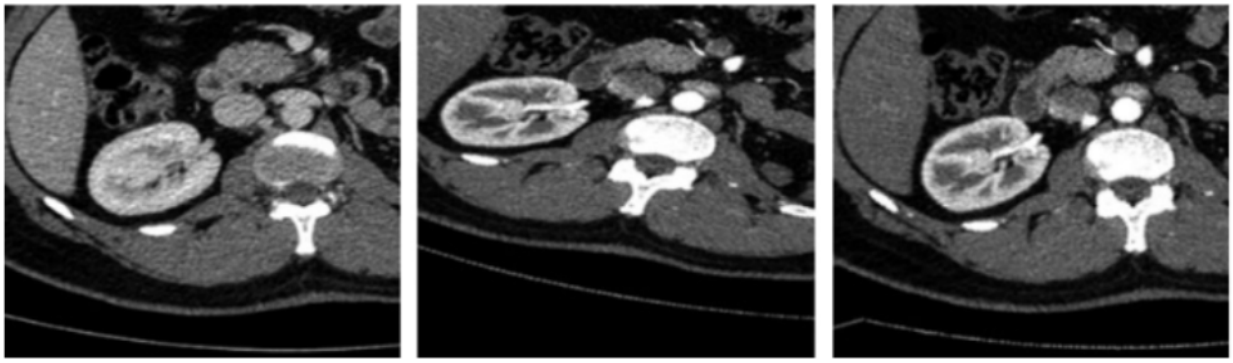


Figure 8: Example of image registration: from left to right – the reference, target, and the registered CT kidney images

In the figure 8, [17] there is an example of uni modal image registration applied to a CT scan of kidney.

2.3.1 Registration of histological images

Most of the solutions realized until now, apply registration on MRI or CT scans images of all our organs, especially brain and lungs. Nowadays, more focus is given to registration of histological images and it is becoming a popular field of research.

Different methods have been proposed to register histological images. Some techniques involve intensity variations [27], and some others are based on the shape of the sections. For some specimens the anatomical structures can be used as landmarks, along with landmark-based registration methods. [28]

The deep learning-based medical image registration is a relatively new field. It is a powerful tool and it is starting to be more employed because of the low time required during the execution, permitting the development of real-time nonrigid registration. This is crucial, e.g. for registration during surgical interventions.
The deep registration approaches can be divided into three main categories based on the training scheme:

- Supervised, which requires ground-truth deformation fields or pre-aligned images that are often impossible to obtain [20].
- Adversarial registration, which is based on generator and discriminator networks, suffers from similar limitations as the supervised category. Moreover, the training stage, it is not trivial, but is usually costly and time-consuming [22].
- Unsupervised, that do not require any ground-truth [21]. They are based on minimization of a given cost function and the registration accuracy mostly depends on:
  - choice of the similarity measure,

- a regularization term enforcing plausible deformations (as previously explained in equation 8),
- the ability to converge during training,
- a generalization ability.

The main challenge with the learning-based registration of histology images is connected with the high resolution of these images, coupled with large and complex deformations. WSI are designed to be accessed through different resolution levels in order to avoid memory problems. They can have different resolution levels, for example, the ones used in the projects have 11 levels: the highest resolution is at level 0, while at 10 there is the lowest resolution.
Each level is associated with a downsampled factor, from level 0 to 10, the factors are 2, 4, .., 512, 1024, respectively. Therefore, the size of the images in each of the levels also varies according to the detail that is required. For example, while the highest resolution image has size 200.000×90.000 (original), the smallest one has 195 × 90 pixels.
Hence, the size of these images is one of the main problems to face in the deployment of registration algorithms. On top of this, deep learning methods suffer from large GPU memory utilization. The higher the image resolution, the larger the necessary receptive field and the required GPU memory. The simplest solution is to downsample the images, however, it reduces the registration quality and makes it harder to register fine details.

To summarize, the main problems related to the registration of histological images are given by: (i) a very high resolution of the images, (ii) complex, large deformations, (iii) difference in the appearance and partially missing data.

3 Methodology

3.1 Previous work: rigid registration

As previously mentioned, a rigid registration algorithm was already implemented by another student from the DigiPatics project. This rigid part has to be applied before the non-rigid registration, it can be seen as a preprocessing step that the input images have to go through.

The algorithm aligns a selected area given as input from two WSI that can be at any resolution level. The image used as a reference (the one in which the area is selected) is called template. The image to be registered called moving image or the aligned image (when the algorithm has already been applied).
The algorithm must first locate the region of interest with respect to the entire moving image. In this way, it will be possible to approximately know in which part of the other staining the region to be registered is.

In particular, the algorithm takes as input two Whole Slide Images from the same patient with different stains (the available stains are CK19, HE, HER2, KI67, RE and RP) and the coordinates of the area of interest.
In order to locate the selected area on the moving image, it creates a image pyramid at different resolutions: from the area of interest, it expands the visual field zooming out so more and more areas can be seen, at the cost of reducing the level of details. Hence, the resolution seen is the lowest.
Then, from the lowest resolution to the higher one, it starts the registration and it begins to aligning all the tissue until it ends up align the region of interest.

The approach can be divided in three stages:

- a prealignment using the shape of all the tissue,
- a global registration using also the shape but of a delimited region,
- a local alignment using internal structure of the tissue.

In each step, a different technique and different evaluation metrics are used.

3.1.1 Prealignment

The objective of the first step is to obtain the binary masks of the tissue pixels of the template and the image to be registered.
At the lowest resolution, it differentiates the tissue from the background pixel and focus on aligning the shape obtained through the binarization. It obtains the translation vector from the center of mass of both masks and the difference between the center point of the template staining and the one to be aligned is calculated.
The mask of the moving image is then translated according to the translation vector and the optimal angle of rotation is found maximizing the Intersection Over Union (IOU) between the two masks.

3.1.2 Global alignment

From the binary mask obtained before, the distance transform of the tissue pixels to the closest background pixel is calculated. From this distance map, the optimal translation vector is searched, the one that maximize the cross-correlation between the two images is chosen.
In order to compute this, the FFT (Fast Fourier Transform) is used to find the best value directly from the time shifting property of the transform.

3.1.3 Local alignment

This last step is applied at the higher resolution. The main problem is that the tissue can have can have a very different appearance in different stains, both in color and shape in the highlighted elements in each of them. For this reason, in both images the hematoxylin channel is extracted. In this way, an attempt is made to extract the nuclei of the cells to highlight these elements and to make the images as similar as possible.
From the hematoxylin channel, as in the previous cases, the optimal translation is also obtained from the crosscorrelation using the FFT.

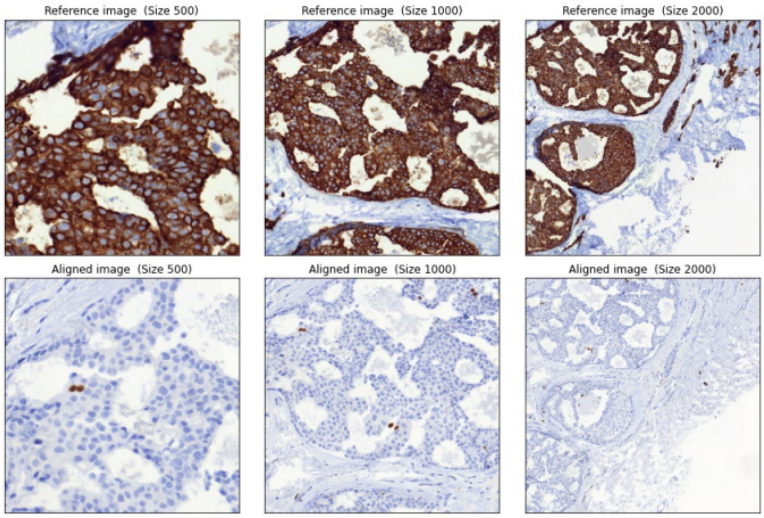


Figure 9: Example of rigid registration of a pair in different region size (500, 1000 and 2000)

3.2 Dataset

3.2.1 ANHIR dataset

The dataset used to train the neural network employed in the project comes from the ANHIR challenge [1]. ANHIR was the first open competition of image registration algorithms on microscopy images.
(ANHIR) Automatic Non-rigid Histological Image Registration challenge was organized to compare the performance of image registration algorithms on several kinds of microscopy histology images in a fair and independent manner.

ANHIR offers a dataset that contains high-resolution (up to 40× magnification) whole slide images. They are organized in sets such that any two images within a set could be meaningfully registered, as they come from spatially close slices. Different stains are used for each image in a set and the local structure often differs.
It consists of 481 image pairs, there are 8 tissue types: (i) mammary glands, (ii) the colon adenocarcinomas, (COADs), (iii) gastric mucosa and adenocarcinomas, (iv) breast, (v) mice kidney, (vi) human kidney, (vii) lung lesions, and (viii) lung lobes. [18]
The consecutive slices were stained by: (i) prosurfactant protein C, (ii) antigen KI-67, (iii) clara cell 10 protein, (iv) human epidermal growth factor receptor 2, (v) progesterone receptor, (vi) estrogen receptor, (vii) platelet endothelial cell adhesion molecule, (viii) cytokeratin, (ix) hematoxylin and eosin, (x) podocin [19]. In total there are 49 sets, each set has on average 5 slices. Each slide is resampled to approximately 25% of the full original resolution, resulting in larger size varying from 6k to 17k pixels in one dimension (from 4369x6930 to 17179x15042). Depending on the image the resolution varies.
The images are provided as .jpg and .png files without the metadata.
This dataset was used for the huge variety of images, belonging to different organs and colored in different stains. In this way, the network is robust to any changes and it can be used for more application.
Figure 10, there are shown some images taken from the dataset. It is possible to see different tissue types stained using various dyes.

3.2.2 DigiPatics images

The images to which the algorithm are applied were provided by the DigiPatics project. For now, the project is focusing on breast cancer histology images, however, it is starting to analyze also lung histology images.
All the images come from the Institut Catala de la Salut (ICS), specifically those from the Vall d'Hebron hospital. The available stains are: CK19, HE, HER2, KI67, RE and RP.
Some examples are shown in Figure 9 and Figure 2.
The images used are not WSIs but selected areas of WSI where regions of interest are present. They are called tiles and they were generated by the rigid-registration previously explained.

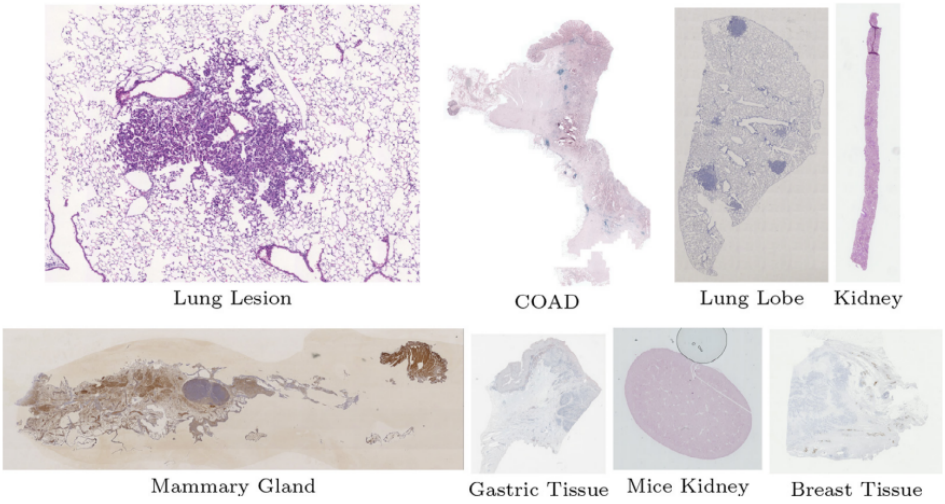


Figure 10: Example of images from the ANHIR dataset[1]

3.3 Project development

In order to decide the most suitable non-rigid registration technology to apply to the Digipatic dataset, the algorithms published in the most recent years were studied and compared.
In the state-of-the-art, all the methods proposed are applying non rigid image registration to WSI images, never to smaller and more detailed tiles. For this reason, the search of a suited algorithm was very challenging.

Consulting the ANHIR challenge (organized jointly with the IEEE ISBI 2019 conference, which is the only open comparison of image registration algorithms on microscopic images), it was seen that almost all of the best scoring methods were based on the classical, iterative image registration approach resulting in the long time required for the analysis[18]. Even though the registration accuracy of the proposed methods is close to the level of the human annotation, the computational time is relatively high, hence, the usefulness of the solution is lower. Probably the majority of the challenge participants didn't adopt a deep learning approach because of the high-resolution of histology images, making them difficult to register due to the GPU memory constraints (most GPUs have a maximum of 16-32 GB of RAM). The only method adopting deep learning was proposed by Tsinghua University [29] and it applies a Structural Feature Guided Convolutional Neural Network for the non-rigid registration. The network is first trained in an unsupervised manner, maximizing an image correlation coefficient, then finetuned using provided landmark positions on the training data. However, it performs much better on training than on testing data, hinting again on possible overfitting.
Other solutions adopting neural network are registering WSI as wholes, while, in this project, the goal is to register tiles which have higher resolution and they are much smaller and more detailed.

For these reason, it was chosen a framework proposed outside of the challenge: Deep-

HistReg: Unsupervised Deep Learning Registration Framework for Differently Stained Histology Samples [19]. It is an unsupervised deep learning-based registration framework. The pipeline consists of data loading, transferring to GPU, preprocessing, initial alignment, affine registration, and, finally, nonrigid registration.
One of the main benefits of this solution is the fact that it registers the WSI iteratively at different resolution included the higher one, making it suitable for this project.
Three different neural networks are implemented in the pipeline:

- Tn the preprocessing, tissues are segmented from the background by a UNet-based [31] network, which is fast, robust, and easily convertible to other histology datasets.
- The affine registration is executed thanks to a ResNet-like convolutional neural network [32]. The output is an affine transformation matrix (2x3) that is then converted to the transformation grid used in the spatial transform.
- For the non-rigid part it is proposed a pyramid-based, patch-based, groupbased, and iterative deep registration solution. The architecture of the network is explained in the next subsection.

As most of the registration methods, it can be divided into two parts: rigid registration and non-rigid registration. This project is based only on the non-rigid part, however, the first part is still trained and tested to compare it with the rigid solution implemented by the DigiPatics project.

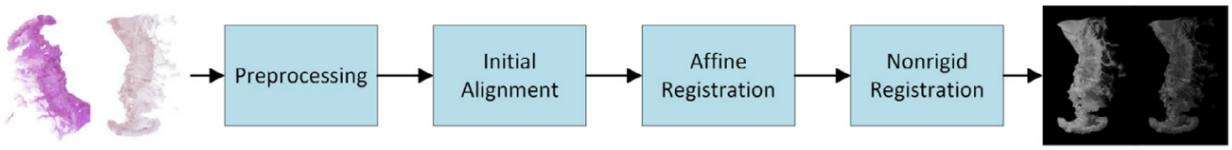


Figure 11: Framework pipeline of DeepHistReg

The nonrigid registration is the most challenging step in histology registration. Only using a simple network, it is almost impossible to achieve accurate registration because the parameter gradients don't fit into the GPU memory. Also, even a common patch-based approach that reduce an image into smaller patches, that are then combined in the batch dimension, is not enough since the batches would not fit in the GPU memory too. The method proposed tries to solve these issues adopting a new state of the art network.

3.4 Requirements

All the code is implemented with Python using the following libraries:

- NumPy and Pandas for general purpose data manipulation.
- PyTorch to create and manage the neural networks.
- SimpleITK, PIL and tifffile to read, manage, process and save the images.
- SciPy to apply interpolation to the data.

On the hardware side, all the code was run on the GPI (Image and video processing group) servers. GPI offers GNU/Linux servers with the following resources for each experiment run: 10 GB of RAM, 10 GB of GPU memory and 16-core CPU with a maximum of 15 threads.
Moreover, to open the different image formats present in the framework, these three specific softwares are used:

- QuPath: to visualize and navigate the WSI [23].
- Matlab, "Read Medical 3D data" library: to visualize .mha format [24].
- GIMP: to visualize .tiff format; also Photoshop can be used [24].

3.5 Preprocessing

To obtain the input tiles of the non-rigid network, first the rigid part has to be executed. The WSIs to align are visualized using a software called QuPath which allows to navigate and visualize the single WSI at different resolutions. The coordinates and the size of the area of interest are used as input to run the rigid algorithm.
The output of the rigid part is the pair of tiles of the selected area rigid registered. The images are saved in .png format.

Before entering the network, each image has to go through a phase of preprocessing.
All the pairs are padded and parsed from .png format into a .mha uncompressed format. MHA files mostly belong to ITK, the graphic data files contain the information regarding the Insight Segmentation and Registration Toolkit (ITK). This format is mostly used in 3D MRI images, but in this case, it was very useful to speed up the data loading on the GPU during the training.
There were some problems encountered with this data format, especially for the visualization of the images without any software. At the end, it was possible to visualize the images using a specific library in Matlab.

Then, all images are converted to greyscale and downsampled. The downsample depends on the size of the tile During the training set this step was necessary because the images are much bigger, however, when tiles are processed, it can be skipped.

Then, all the images are transferred to the GPU memory. The image transfer is being done only if a single GPU is used and both images fit into the memory. If it's used a multi-GPU computing cluster the memory transfer is done later.
At the end, before entering the network, images are re-sampled to a predefined number of levels, building a classical resolution pyramid.

Moreover, it is possible to apply the registration on couples of tiles saved in .png format. It doesn't convert them in .mha, however it applies the same preprocessing steps listed above. The main advantage comes from the fact that a large portion of memory is saved without having to convert pairs in .mha format. In fact, .mha is an uncompressed format and it occupies a large amount of memory (at least three times a .png picture). The main

disadvantage is that the network works a bit slower, however, in this testing stage and with small tiles, it is not a problem.

3.6 Non-rigid Network

In order to create a flexible and robust network, the solution proposed is structured according to the following approaches: pyramid-based, patch-based, group based, and iterative.

- Pyramid-based: the images are registered at different resolutions starting from the coarsest level. From the first level, a deformation field is computed and subsequently it is upsampled to the next resolution.
- Patch-based: at the given resolution the images are unfolded into smaller patches that can be handled by a relatively small deep network.
- Group-based: patches are grouped together and, at once, only a small group of patches is propagated by the network due to GPU memory constraints and the loss function is being evaluated and optimized at the group level, not at the image level.
- Iterative: at each pyramid level, images are propagated through the network several times composing the calculated deformation fields.

It is defined as source, the moving image that has to be aligned to the pair image which is called target.
In this network, the cost function is given by the negative normalized cross-correlation (NCC) and by the curvature (CURV) as the displacement field regularization term.
The objective is:

$$S(F, M, u) = -NCC(M, F) + \alpha CURV(u) \rightarrow min \tag{9}$$

where M, F, u are the warped moving patches, target patches and the displacement fields respectively, α controls the deformation smoothness and it is the regularization parameter that was explained in equation 7; CURV is the curvature regularization.
Normalized cross correlation (NCC) has been commonly used as a metric to evaluate the degree of similarity (or dissimilarity) between two images. It was chosen because it improves the ordinary cross correlation: the main advantage is that it is less sensitive to linear changes in the amplitude of illumination in the two compared images. [33].

Figure 12 shows a scheme on how the network is structured.
It starts after the resolution pyramids are built for both the source and the target image. Then, starting at the coarsest resolution, for a given number of iterations the images are being registered. The deformation field is initialized with the identity transformation and it is warped to the source image.
The images are then unfolded into overlapping patches in each iteration. Patches have a stride of half the patch size and this causes a slightly longer registration time, which is

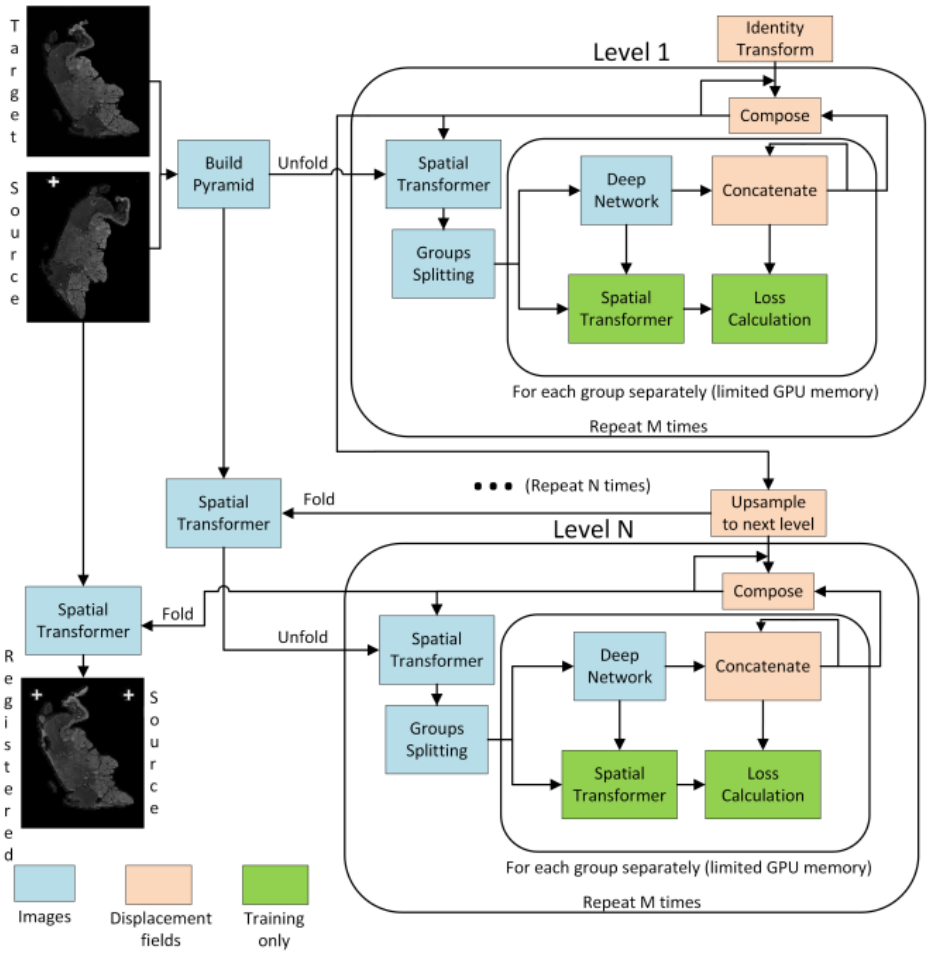


Figure 12: Structure of the non-rigid network [34]

not a problem. But more importantly, it addresses the issue of deformation field discontinuities at patch boundaries.
Then, a fixed number of patches, based on the GPU limitation, is grouped together. With the GPI servers, there were no limitations, however, this constraint has to be considered in case the algorithm is run on other machines.
The current velocity field is calculated by propagating each relevant group via the registration network. The group is converted and the cost function is calculated during training.

The calculated velocity fields for each group are concatenated together, and after all the groups are processed, the concatenated velocity fields are folded back into the velocity field with the same shape as the current deformation field.
The current deformation field is composed of the velocity field and used for the next iteration where the whole process is repeated.
This makes the interpolation error negligible since the source image is never interpolated more than once.
After composing, the current level deformation field is upsampled to the next resolution. The deformation field after the highest resolution becomes the final deformation field

which is the one applied to the source image.

This structure is used for training the network with WSI. But, depending on the size of the pair of tiles given as input, the number of pyramid levels can be decreased: if the tiles is small, it's possible to just consider one level of the pyramid, if it gets to bigger size it can be divided into 2, if it is used the whole slide, the image is divided into three levels. This structure is very robust to any given size of pair of images.

In the following page, there is the pseudocode of the algorithm.

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)

```
P_M, 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
    F_C = get current level P_F and unfold using P, S
    if current resolution > 0 then
        M_C = warp current level P_M using u
    end if
    v = initialize with an identity transform and unfold using P, S
    for each inner iteration in M do
        if current iteration > 0 then
            M_C = warp M_C with v
        end if
        M_g , T_g = divide F_C ,M_C into G-sized batches
        v_i = initialize with an empty tensor
        for each group do
            v_t = model(M_g, T_g)
            if training then
                M_w = warp Mg with v_t
                S(M_w, T_g, v_t) = use equation (9) and update optimizer (free GPU memory for the next group)
            end if
            v_i = concatenate(v_i, v_t)
        end for
        v = v ∘ v_i
    end for
    v = fold v using P, S
    u = u ∘ v
end for
return u
```

3.6.1 Parameters

There are several parameters to set in this network:

1. the patch size,
2. the stride,
3. the group size,
4. number of pyramid leves,
5. number of iterations per level
6. the regularization parameter (α)

Every parameter is important in the functionality and quality of the network. Increasing the number of pyramid levels allows for bigger deformations to be calculated, but at the cost of increased registration time. The number of iterations per level is critical for registering fine details, but raising the value, like increasing the number of resolutions, increases the registration time.
The patch size is linked to the deep network architecture, and its value should be chosen carefully to make the most of the network's receptive field.
The stride determines the amount of overlap between the unfolded patches.
Finally, the group size determines how many patches are registered at the same time. The higher the value, the faster the registration and GPU memory usage will be.
Finally, the regularization parameter is responsible for controlling the deformation smoothness.

3.6.2 Training

As mentioned before, the training was done using the dataset offered by the ANHIR challenge. There are used 481 image pairs split into 251 evaluation (validation set) and 230 training pairs.
The training of the first two neural network took around a few hours each, however, it took three days to complete the training of the non-rigid neural network because of the complicated structure and the big size of the training images. In fact, the bigger the number of levels, the longer is the time needed to train the network and to register images.
In the model the parameters are set as:

1. 256x256 patch size,
2. 128 stride size,
3. 32 patches per group,
4. 3 pyramid leves,
5. 3 iterations per level
6. 0.001 regularization parameter (α)

3.7 Interpolation of the original image

One of the main issues of the postprocessing steps was to apply the deformation field computed by the neural network to the source colored image. In the preprocessing step, the pair of images is converted to mha format, converted to greyscale, downsampled and normalized. This steps convert the pixel values in float numbers between 0 and 1. The deformation field values are float numbers between -1 and 1.
The goal of the project is to apply this deformation to the colored original image, not to the processed one.
The original input is in '.png' format and its pixel values are integer numbers between 0 and 255. Hence, there is the need to apply some invertible transformations to it in order to make it suitable for the spatial transformation that will apply the final deformation field to the image.

In the preprocessing stage the images are mirrored with respect to the vertical axis and rotated by 90°, hence, also the colored image have to go through the same process.
A necessary condition in order to apply the deformation field is to have an image with just one channel (i.e. greyscale images). The png format, by definition, has 4 channels: red, green, blue and opacity. The opacity channel doesn't contain any information in the dataset used, hence it can be discarded. The other three channels need to be considered as single greyscale images.
The next step is to normalize the pixel values of each image and to convert them in float. After applying these operations, it is possible to warp the images with the deformation field.

The spatial transformation used to apply the deformation is the bilinear interpolation. Bilinear interpolation is a method for two-dimensional interpolation. To perform a spatial transformation of the input image, a sampler must take the set of sampling points (deformation field in this case), along with the input image U and produce the sampled output image V.
Each (x_i, y_i), coordinates of the deformation field that belongs to the grid G, defines the spatial location in the input where a sampling kernel is applied to get the value at a particular pixel in the output image. In this case, the spatial kernel used applied is the bilinear interpolation.[35].
The transformation can be defined as:

$$V_i = \sum_{n}^{H} \sum_{m}^{W} U_{nm} k(x_i - m, \Phi_x) k(y_i - n, \Phi_y) \quad \forall i \in [1, .., H, W] \tag{10}$$

where Φ_x and Φ_x are the parameters of the kernel k() which defines the bilinear interpolation, U_{mn} is the value at location (n, m) of input image U. V_i is the output value for pixel i at location (x_i, y_i) of output image. H and W are the height and width of the input image and of the grid that contains the deformation field (in this case they are the same).

Once the three deformed images are obtained, it is necessary to merge them back together to return to a RGB image format.

The main issue is that the pixel values after the transformation are float numbers. The classical image formats (i.e. png, jpg) are constrained to integer pixel value which, in this case, can't be used. In fact, if the floats of each greyscale image are rounded to integer and then merged together, the resulting image will present areas where the colors don't combine and the deformation is not accurate. An example of this effect is shown in Figure 13.

The only image format that supports a three channels, floating pixels image is TIFF (Tagged Image File Format) which is an high-quality graphics format. Therefore, the output of the network has to be saved as .tiff. It can only be open by dedicated software (i.e. Photoshop). In this project it was used GIMP (GNU Image Manipulation Program) which is free and available for any operating system.

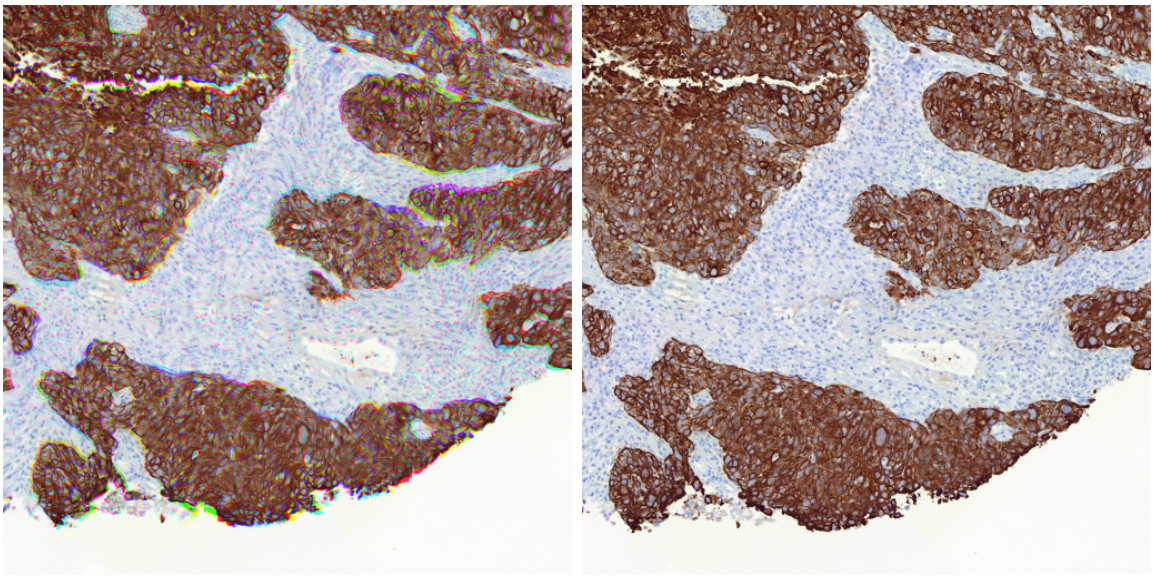


Figure 13: On the left image saved with integer values, on the right same image saved with float values.

4 Results

4.1 Evaluation metric: rTRE

Unfortunately, there are no standard way to evaluate the quality of non-rigid registered pairs of images. Unlike the rigid registration where there are many different metrics available, it is hard to measure the accuracy of a deformation.
In this project, the method employed by the online challenges of histological image registration (ANHIR, ACROBAT) was adopted: the target registration error (TRE) which measures the Euclidean distance between the annotated and transformed landmarks.

In the ANHIR dataset, in each pair of images, the most significant structures are manually annotated with landmarks. On average, around 80 landmarks per image were positioned by pathologists, and their coordinates were saved.
In particular, for each pair of images (i, j), the coordinates of corresponding landmarks x_i^l, y_i^l were determined, where l $\in L_i$, and L_i is a set of landmarks that occurs in both i and j.
When the deformation is applied to the image, also the landmarks are moved according to the deformation field. In this way, it is possible to keep track of the exact movement of each landmark, hence of the deformation of the main structures of the image. The coordinates of the landmarks on the transformed image are defined as $\hat{x}_i^l$.
The relative Target Registration Error (rTRE) is the Euclidean distance between the computed coordinates $\hat{x}_j^l$ and the manually determined (ground truth) coordinates x_j^l and it is computed as:

$$rTRE_l^{ij} = \frac{||\hat{x}_j^l - x_j^l||}{d_j} \tag{11}$$

where d_j is the length of the image diagonal.
The main criterion used to evaluate the registration algorithms is the average of median rTRE:

$$\mu_{i,j}(m) = median\ \ rTRE_l^{ij}(m) \tag{12}$$

Finally, to evaluate the overall performance of the network, average median rTRE (AMrTRE) and median of median rTRE (MMrTRE) are computed:

$$AMrTRE(m) = mean\ \ \mu_{i,j}(m) \tag{13}$$
$$MMrTRE(m) = median\ \ \mu_{i,j}(m) \tag{14}$$

Moreover, to evaluate robustness, individual rTREs for each landmark and the relative Initial Registration Error (rIRE) before registration are compared:

$$rIRE_l^{ij} = \frac{||x_i^l - x_j^l||}{d_j} \tag{15}$$

The registration times t_{ij} in minutes for each registration, including loading input images and writing the output files is also evaluated.

An example of landmarks on the ANHIR dataset is shown in Figure 14.

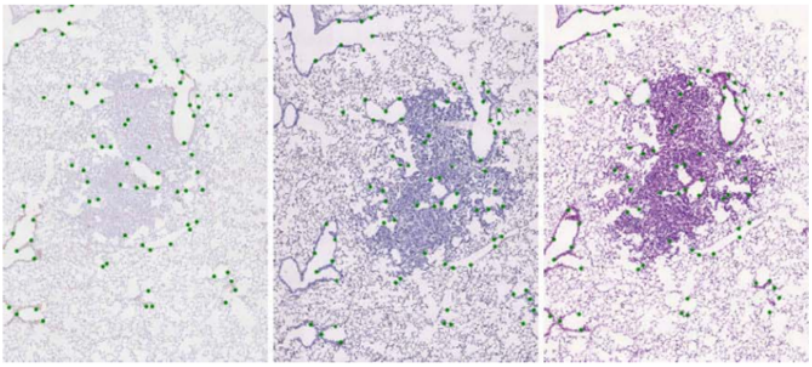


Figure 14: Landmarks on ANHIR dataset

4.2 Landmark positioning

In order to evaluate the output of the algorithm on the breast cancer images from the DigiPatics projects, landmarks were manually positioned as done in other challenges and the evaluation metrics were computed

To generate the landmarks it was used the library OpenCV. At the same time and with the same order, points have to be positioned on the same portion of the tissue on both images. With some particular stains (KI67 for example), it is not possible to identify structures of the tissue, hence, the contrast of the images was increased. Moreover, a grid on top of the images is added to facilitate the process.
All the coordinates are then saved in a file with the same order and sent as input to the code.

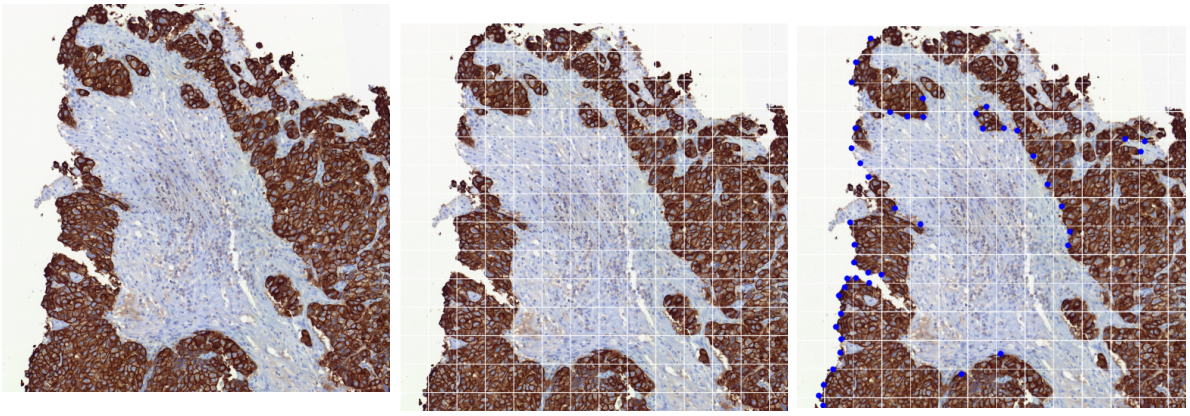


Figure 15: Procedure of positioning landmarks

4.3 Tests and results

The quality of the output of the network depends mainly on:

- the quality of the rigid registration,
- the size of the image given as input and the number of pyramid levels chosen for that particular size.

The network was tested tuning these values until the best score was obtained.

On the images of ANHIR dataset, the output of the non-rigid registration performs well. The results presented on the paper [34] show that the DeepHistReg method is comparable in terms of the rTRE to the best state-of-the-art algorithms. The results are slightly worse (by about 0.002% of the image diagonal) than the three best state-of-the-art algorithms. However, the images from the DigiPatics dataset are different, smaller and more detailed, therefore, the accuracy of the registration decreases.

4.3.1 Case study

To demonstrate the performance of the framework, in this subsection, a case study using a pair coming from the DigiPatics dataset is presented. The target is stained with KI67 and the source with CK19, they are shown in Figure 17. From a quick view, they look very similar, however, the cells have different structures.
As explained before, to evaluate the performance of the registration, landmarks on both images were positioned. In this case, I applied 6 landmarks on well defined contours. (see Figure 15)

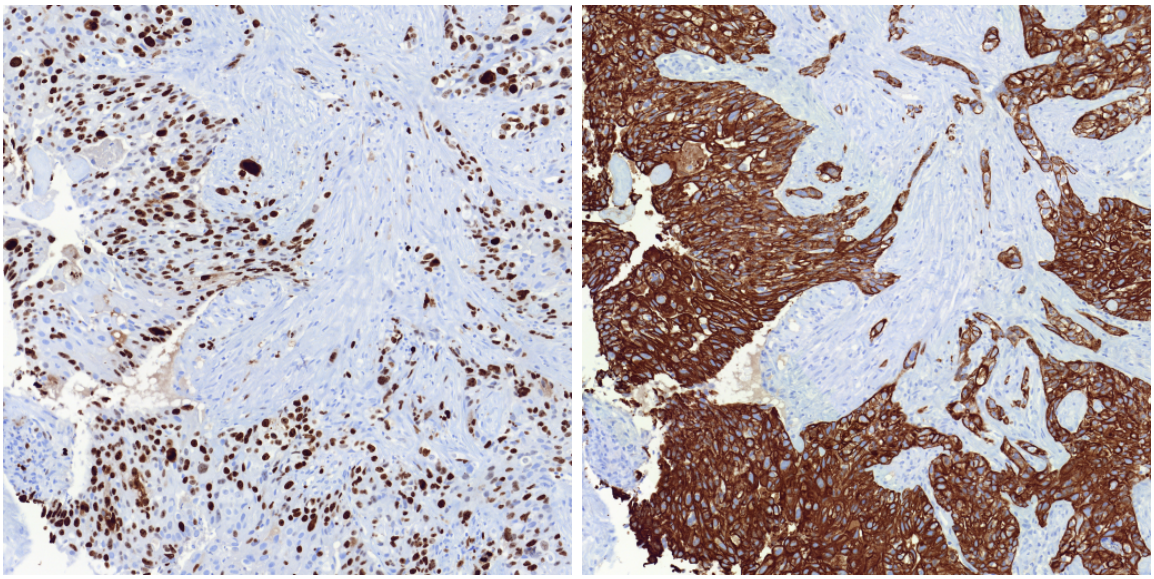


Figure 16: Target and source images from DigiPatics dataset

The network was tested for different values of pyramid levels and iterations. Best results were obtained when using two levels of the pyramid. In fact, the network was trained

with whole slide images that have less resolutions and less details. The tiles registered are 4096x4096 pixels, hence they are much smaller than a WSI.
The number of iterations affects the precision of the deformation that is applied to the tile. In this case, the tiles are very similar and 2 iterations each were applied.
From the result, it can be seen that the tissue was deformed following the structure and contours of the target tile, however, the network is not able to erase or create new tissue. The deformation field can just move it a small amount.
This is the main limitation of the network and it decreases the quality of the registration.

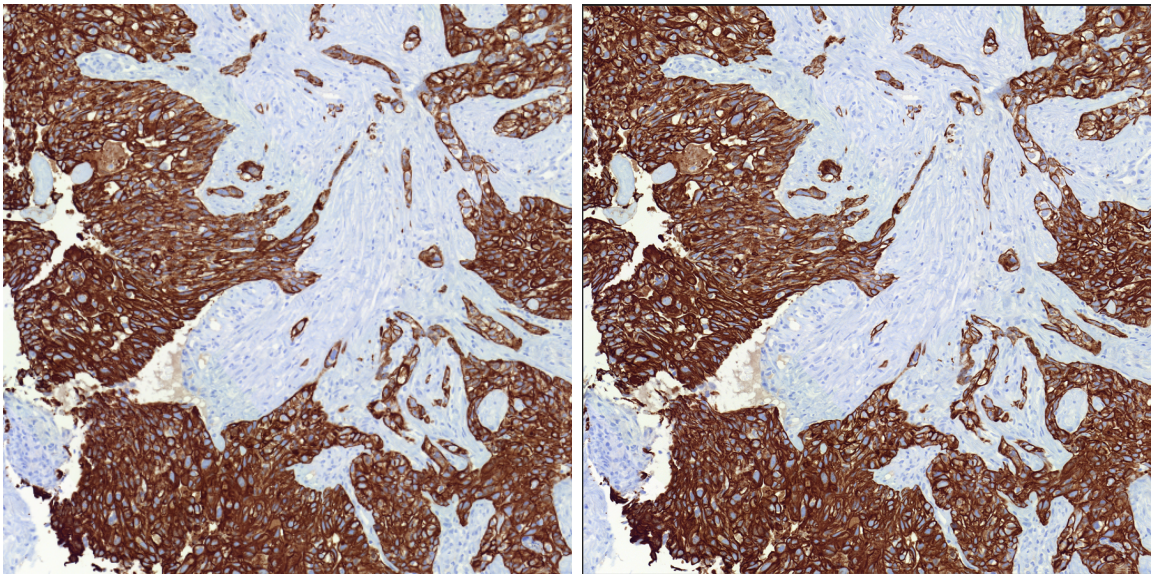


Figure 17: Source and deformed source

From the image, changes are not very visible or distinguishable. However, using the evaluation metric TRE, the improvement is clear:

	Initial	Final
median TRE	0.010751	0.001516
pixel error	62.264	8.780

Table 1: evaluation on pair of tiles studied

As indicated by equation 11, the target registration error is divided by the diagonal of the image. In this case, the two images have size 4096x4096, their diagonal is 5792. It means that in the original pair, the difference of the landmark coordinates is equal to more than 62 pixels, in the second case is less than 10. This is not an absolute criteria considering that landmarks where positioned on the contours of the tissue, but it still gives an idea whether the registration is improving the alignment or not.

The displacement field generated by the network can be divided into the horizontal and

vertical directions and it is possible to visualize it as a greyscale image of the size of the source and target images.
For this case study, the displacement has the following shape:

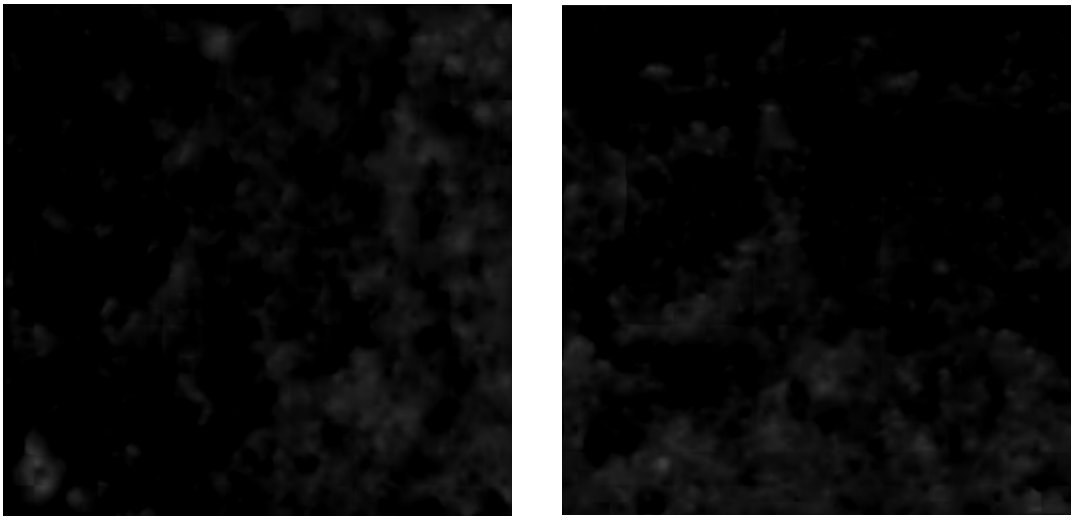


Figure 18: Horizontal and vertical displacement field

Finally, in the following figure, the changes applied by the neural network to the source image are highlighted. In Figure 19, bright colors underlies the main differences among the two images.

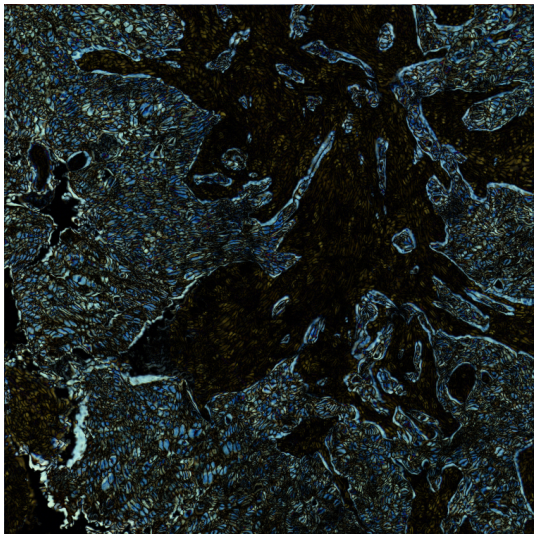
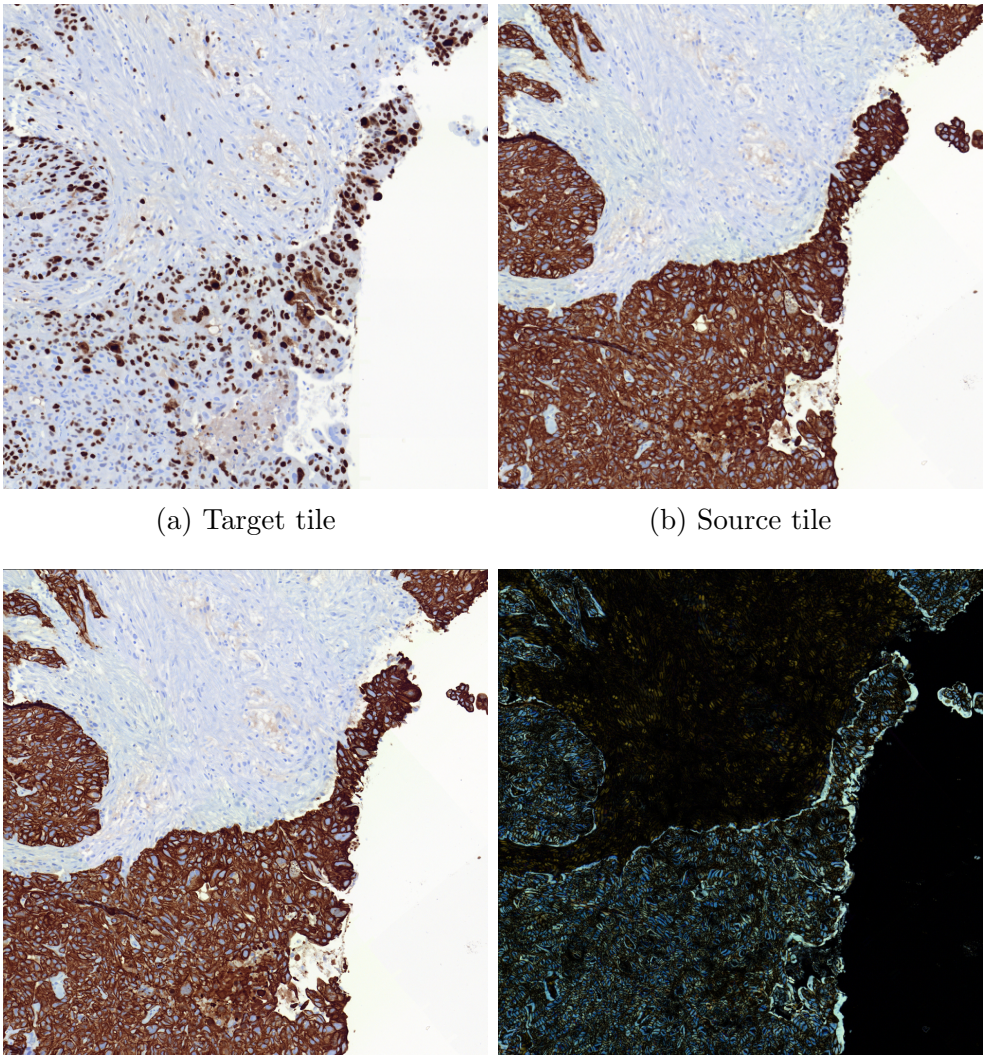


Figure 19: Difference between source and transformed source

As previously mentioned, the main changes appear on the border of the tissue, also other internal structures are modified as well in order to match with the target image.

Another example is given by tiles from another patient, where the KI67 stained is used as target and CK19 as source:

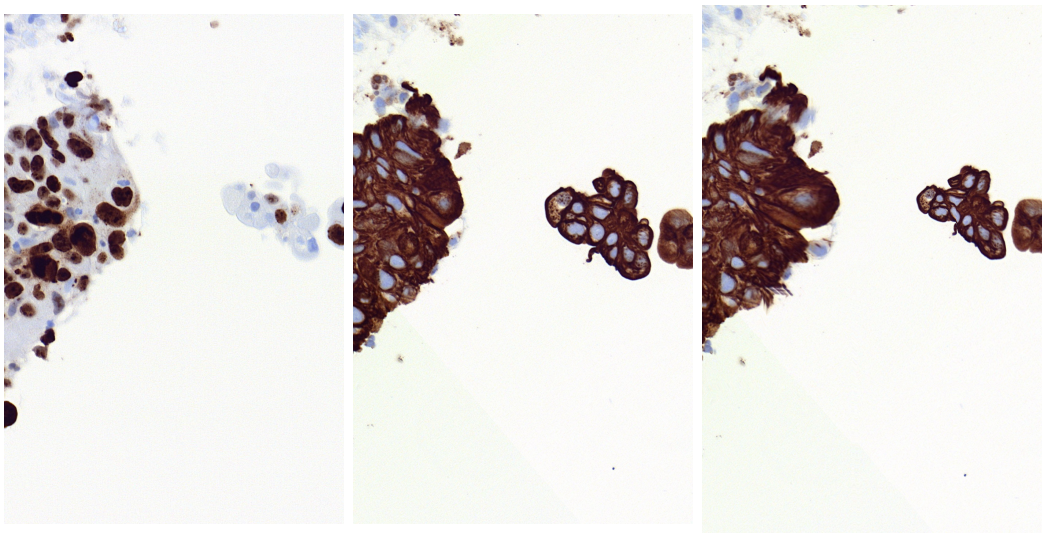


(a) Target tile (b) Source tile

(c) Transformed source tile (d) Difference

Figure 20: Another example of non-rigid registration

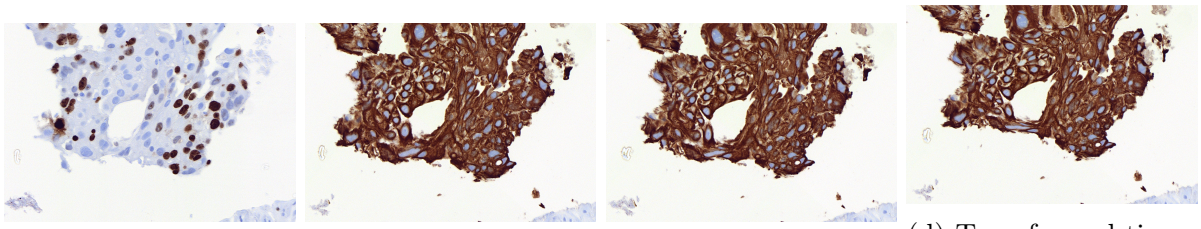
The deformation field succeed in moving the tissue along the correct direction, however, it doesn't align perfectly the two images.
It has to be considered that the two stains are very different: in the target, the structures of the cells are not as defined as the one in the source. Even if the images are processed in greyscale colors, the differences are still very noticeable.
In Figure 22, the deformation field shrinks the tissue of the source to align it to the target in the small portion of cells in the right part of the image. While, in the center, it stretches it.



(a) Target tissue (b) Source tissue (c) Transformed tissue

Figure 21: Details of non rigid registration

A more visible example on how the deformation field works can be seen in the next sequence of images taken from another pair. As before, the tile stained with KI67 is the target and the one stained with CK19 is the source that is transformed.



(a) Target tissue (b) Source tissue (c) Transformed tissue (d) Transformed tissue with 3 pyramid levels

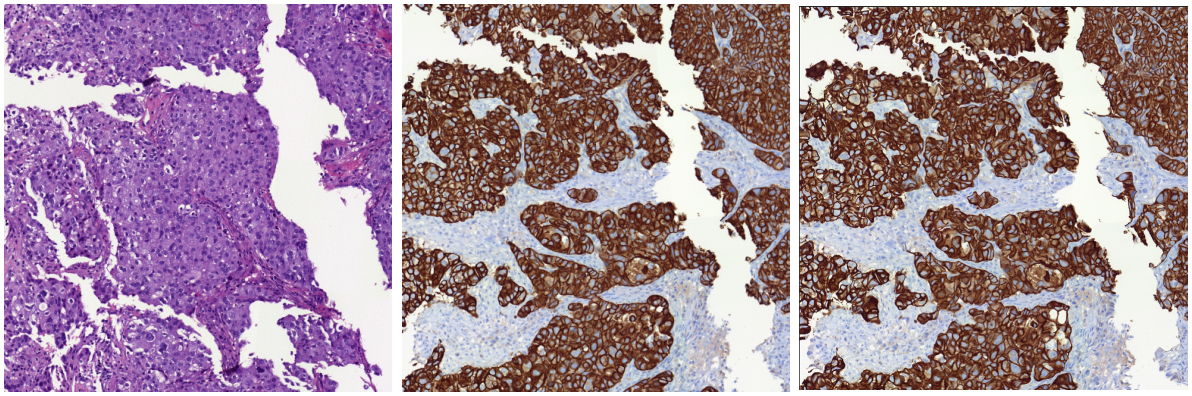
Figure 22: Details of non rigid registration

In the third image, it can be seen the good alignment achieved using 1 level of the pyramid: the white space in the middle is better aligned and the tissue is shrunk. In the fourth image, there were used 3 levels and it is clear that the deformation is too big and it changes the structure of the cell.

The most relevant characteristic of the proposed solution is the speed at which it performs the non-rigid registration. It takes on average 4 seconds to register each pair of 4096x4096 tiles. Moreover, even for bigger images (i.e. WSI used to train the network), it only takes less than 8 seconds to generate the deformation field making this solution very fast.

4.3.2 Limitation of the network

The main limitation of the network is that it can only move the tissue but it can't generate it when it is not present in one of the two tiles. It moves it along to the correct directions but, in some cases, it's not enough to compensate the differences among the pair.
Based on the output of the rigid registration, the non-rigid registration fails or succeeds to align correctly the two tiles.



(a) Target tissue (b) Source tissue (c) Transformed tissue

Figure 23: Bad results on non rigid registration

In this case, the deformation field generated is to small to compensate the differences in the pair: only small deformations are applied, but the main dissimilarity remains untouched.

4.3.3 Tables of results

To test the overall performance of the network, landmarks were positioned on 9 pairs of tiles from the DigiPatics dataset. Initial TRE is compared to the final TRE of each pair and also the time of execution is evaluated.
The images chosen to test the performance have a defined structure of the tissues in both stains in order to position correctly landmarks. On average, there are 60 landmarks per image.
To evaluate the TRE in each pair, it has to be taken into account that the precision of the landmarks is not perfect, there can be an error of a few pixel. Moreover, most of landmarks are positioned on the contour of the tissue and not on internal parts (i.e. Figure 15).

The average execution time consists of the time needed to load the image on the GPU and the time needed to generate and apply the deformation field. On average is around 4 seconds, except for the first pair, which takes a bit longer.
On most cases, the deformation generated by the network improves the target registration error which ranges from 120 pixels error to 30 pixels error. However, there are two cases where the registration doesn't work: on pair 5 and pair 9 there aren't any improvement. The pixels error range from 350 to 320, hence the error is still pretty large. Pair 9 is shown in Figure 23.

Pair	Initial TRE	Final TRE	Execution Time [min]
1	0.010013	0.0048	0.24994
2	0.010108	0.005796	0.082272
3	0.010751	0.007027	0.082006
4	0.00596	0.003754	0.082009
5	0.0641	0.063653	0.082358
6	0.008305	0.00475	0.082238
7	0.012146	0.00645	0.083008
8	0.02164	0.00547	0.082192
9	0.054184	0.05369	0.082415

Table 2: Landmarks evaluation

Overall, when the initial TRE is already low, so when the pair is already good aligned from the rigid registration, the non-rigid improves it, successfully deforming the tissue along the correct direction. But when the disalignment is too big, the network fails to align the pair.

5 Conclusions and future development:

In the state-of-the-art, there are not models that can apply non-rigid registration to high resolution histological images. At the moment, most of researches are focused on aligning Whole Slide Images, which are much larger. The DeepHistReg framework was chosen because of its accuracy in aligning WSI and of its pyramid-level approach, which makes it scalable to different resolutions.
It is a powerful framework that can be still used on different WSI to apply real-time, complete and accurate registrations.

In the case analyzed, the proposed solution has some limitations on the quality of the deformation, especially when the source presents some tissue that doesn't appear in the target or viceversa. At this level of resolution, the deformation field generated is not enough to correct big changes. If the quality of the pairs generated by the rigid registration is not good enough, the non-rigid registration fails to align correctly the images. On the other hand, it corrects small deformations, especially on contours points, minimizing the target registration error.

This limitations might come from the dataset used to train the network: WSIs are multiresolution images, tiles are low resolution images of WSIs. The model of the network can't generate big field to deform correctly some pairs.

Fortunately, this could be improved training the network with pairs of tiles collected by the DigiPatics project. Once the rigid registration algorithm generates a sufficient number of pairs belonging to different stains, it will be possible to use them to train the network. Moreover, based on the size of these images, one, two or three levels of the pyramid can be generated, increasing the final accuracy. This is an idea for a future implementation that can improve the quality of the network.

The main issues of big deformations, however, is that they might change too much the structures of cells, modifying important information needed for doctors to conduct accurate diagnosis. It's a trade off between the deformation required and the regularization term that holds the network from applying deformations to all the pixels in an image and changing its structure [36]. This is another field of research within image registration algorithms and it has to be taken into consideration, especially when working with biomedical images.

Another promising method that could improve the network consists in applying style transfer using the adversarial networks [22] to one image of the pair, to make it more similar to the other one. It consists in applying a style connected with a particular dye from a given slice to another consecutive slice. For example, considering to apply style transfer to two slices stained with HE and CK19. The algorithm will convert HE to the CK19 without losing information. An example of style transfer is displayed in Figure 24. This could be useful to create ground-truth alignments for adversarial registration networks which may produce even more accurate registration, without the necessity to define a similarity measure. Or to directly use these transformed pairs as input of the non-rigid network.
There already exists some state-of-the-art technologies that perform style transfer. More-

over, it is also a subject of research of the DigiPatics project, they are developing an algorithm that applies style transfer to the DigiPatics dataset used in this project. Hence, a future work could be to use those images to train the non-rigid network.

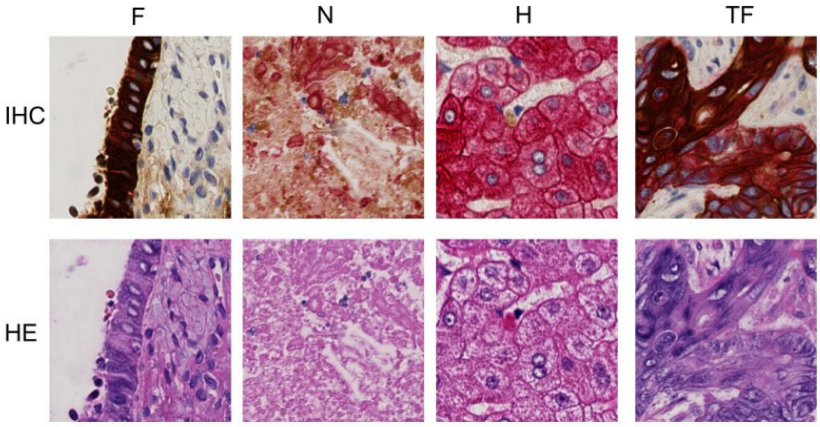


Figure 24: Example of style transfer on histological images[37]

References

[1] Anhir: Grand challenge.

[2] Nadia Harbeck, Frédérique Penault-Llorca, Javier Cortes, Michael Gnant, Nehmat Houssami, Philip Poortmans, Kathryn Ruddy, Janice Tsang, and Fatima Cardoso. Breast cancer. *Nature reviews Disease primers*, 5(1):1–31, 2019.

[3] Institut català de la salut. consulta mercat digipatics. technical report, generalitat de catalunya, 2019.

[4] D Hopwood, JD Bancroft, and A Stevens. Theory and practice of histological techniques. *Theory and practice of histological techniques*, 1996.

[5] How histology slides are prepared.

[6] Lisa Gottesfeld Brown. A survey of image registration techniques. *ACM computing surveys (CSUR)*, 24(4):325–376, 1992.

[7] S. Rabanaque Rodriguez. Alignment of multi-stained digital histological images. *UPC, Data Science Eng. degree, CBI report*, 2021.

[8] Nicholas Trahearn, David Epstein, Ian Cree, David Snead, and Nasir Rajpoot. Hyperstain inspector: a framework for robust registration and localised co-expression analysis of multiple whole-slide images of serial histology sections. *Scientific Reports*, 7(1):1–13, 2017.

[9] William R Crum, Thomas Hartkens, and DLG Hill. Non-rigid image registration: theory and practice. *The British journal of radiology*, 77(suppl_2):S140–S153, 2004.

[10] Seiichi Uchida. Image processing and recognition for biological images. *Development, growth & differentiation*, 55(4):523–549, 2013.

[11] Johan Öfverstedt, Joakim Lindblad, and Nataša Sladoje. Fast and robust symmetric image registration based on distances combining intensity and spatial information. *IEEE Transactions on Image Processing*, 28(7):3584–3597, 2019.

[12] Mohamed Abdel-Basset, Ahmed E Fakhry, Ibrahim El-Henawy, Tie Qiu, and Arun Kumar Sangaiah. Feature and intensity based medical image registration using particle swarm optimization. *Journal of medical systems*, 41(12):1–15, 2017.

[13] Clinton Fookes and Anthony Maeder. Comparison of popular non-rigid image registration techniques and a new hybrid mutual information-based fluid algorithm. In *Proceedings of the 2003 APRS Workshop on Digital Image Computing*, pages 57–62. University of Queensland, 2003.

[14] Clinton Fookes and Mohammed Bennamoun. Rigid medical image registration and its association with mutual information. *International journal of pattern recognition and artificial intelligence*, 17(07):1167–1206, 2003.

[15] Christopher Gilliam and Thierry Blu. Local all-pass geometric deformations. *IEEE Transactions on Image Processing*, 27(2):1010–1025, 2017.

[16] Loren Arthur Schwarz. Non-rigid registration using free-form deformations. *Technische Universität München*, 6:4, 2007.

[17] Fatma El-Zahraa Ahmed El-Gamal, Mohammed Elmogy, and Ahmed Atwan. Current trends in medical image registration and fusion. *Egyptian Informatics Journal*, 17(1):99–124, 2016.

[18] Jiří Borovec, Jan Kybic, Ignacio Arganda-Carreras, Dmitry V Sorokin, Gloria Bueno, Alexander V Khvostikov, Spyridon Bakas, I Eric, Chao Chang, Stefan Heldmann, et al. Anhir: automatic non-rigid histological image registration challenge. *IEEE transactions on medical imaging*, 39(10):3042–3052, 2020.

[19] 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.

[20] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Deep image homography estimation. *arXiv preprint arXiv:1606.03798*, 2016.

[21] Bob D De Vos, Floris F Berendsen, Max A Viergever, Hessam Sokooti, Marius Staring, and Ivana Išgum. A deep learning framework for unsupervised affine and deformable image registration. *Medical image analysis*, 52:128–143, 2019.

[22] Jingfan Fan, Xiaohuan Cao, Qian Wang, Pew-Thian Yap, and Dinggang Shen. Adversarial learning for mono-or multi-modal registration. *Medical image analysis*, 58:101545, 2019.

[23] Qupath.

[24] Matlab, read 3d medical data.

[25] Gimp.

[26] Calvin R Maurer and J Michael Fitzpatrick. A review of medical image registration. *Interactive image-guided neurosurgery*, 1:17–44, 1993.

[27] Eric Bardinet, Sébastien Ourselin, Didier Dormont, Grégoire Malandain, Dominique Tandé, Karine Parain, Nicholas Ayache, and Jérôme Yelnik. Co-registration of histological, optical and mr data of the human brain. In *International conference on medical image computing and computer-assisted intervention*, pages 548–555. Springer, 2002.

[28] Rushin Shojaii, Tigran Karavardanyan, Martin Yaffe, and Anne L Martel. Validation of histology image registration. In *Medical Imaging 2011: Image Processing*, volume 7962, pages 437–443. SPIE, 2011.

[29] Lin Ge, Xingyue Wei, Yayu Hao, Jianwen Luo, and Yan Xu. Unsupervised histological image registration using structural feature guided convolutional neural network. *IEEE Transactions on Medical Imaging*, 2022.

[30] Marek Wodzinski and Andrzej Skalski. Multistep, automatic and nonrigid image registration method for histology samples acquired using multiple stains. *Physics in Medicine & Biology*, 66(2):025006, 2021.

[31] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[32] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[33] Y Raghavender Rao, Nikhil Prathapani, and E Nagabhooshanam. Application of normalized cross correlation to image registration. *International Journal of Research in Engineering and Technology*, 3(5):12–16, 2014.

[34] Marek Wodzinski and Henning Müller. Unsupervised learning-based nonrigid registration of high resolution histology images. In *International Workshop on Machine Learning in Medical Imaging*, pages 484–493. Springer, 2020.

[35] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. *Advances in neural information processing systems*, 28, 2015.

[36] Pascal Cachier and Nicholas Ayache. *Regularization in image non-rigid registration: I. Trade-off between smoothness and intensity similarity.* PhD thesis, INRIA, 2001.

[37] Zhaoyang Xu, Carlos Fernández Moro, Béla Bozóky, and Qianni Zhang. Gan-based virtual re-staining: a promising solution for whole slide image analysis. *arXiv preprint arXiv:1901.04059*, 2019.

Appendices

A Code functions

A.1 Visualize .mha image

Function used to visualize .mha file in matlab using the library "read medical 3D data". Note that the conversion in mha format works only if the depth of the image is 32bit.

```
data1 = mha_read_header(path_to_mha_file);
V1 = mha_read_volume(data1);
figure();
imshow(squeeze(V1(:,:,round(end/2))),[]);
```

A.2 Position landmarks

Run the script twice, opening target and source image at the same time and position landmarks in the same order for both images.

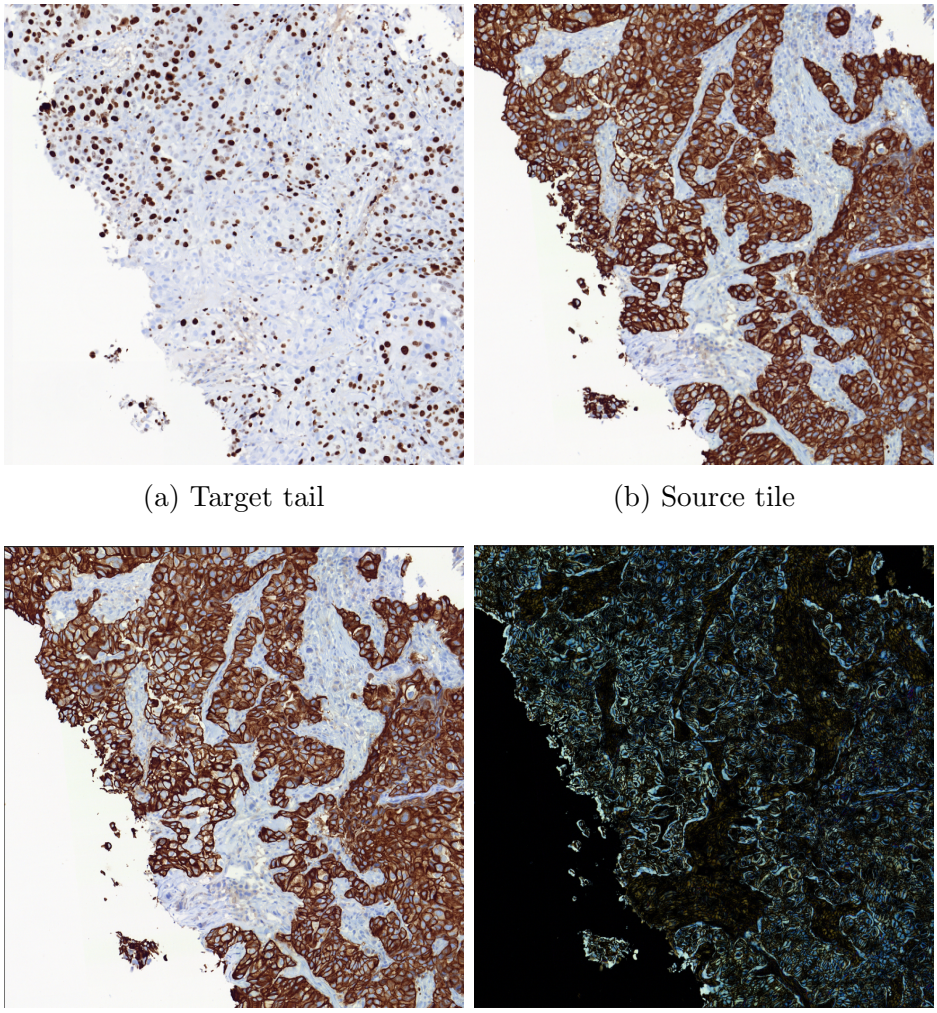
```
import cv2

def draw_circle(event,x,y,flags,param):
    if event == cv2.EVENT_LBUTTONDOWN:
        cv2.circle(img,(x,y),30,(255,0,0),-1)
        #cv2.putText(img, str(x) + ',' + str(y), (x+300,y+300),font,5, 0, 0)
        print(x, ' ', y)
        s=str(x) + ' ' + str(y) + '\n'
        file_object.write(s)

img = cv2.imread("reference.png",1)
file_object = open('reference1.txt', 'a')
cv2.namedWindow("Resized_Window", cv2.WINDOW_NORMAL)
# Using resizeWindow()
cv2.resizeWindow("Resized_Window", 800, 800)
cv2.imshow('Resized_Window', img)
cv2.setMouseCallback('Resized_Window',draw_circle)
while(1):
    cv2.imshow('Resized_Window',img)
    if cv2.waitKey(20) & 0xFF == 27:
        break
cv2.destroyAllWindows()
file_object.close()
```

B Additional results

B.1 Example 1



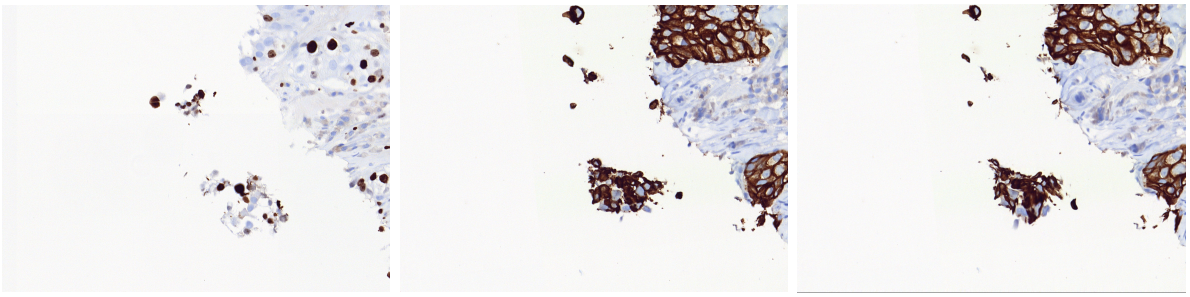
(a) Target tail (b) Source tile

(c) Transformed source tile (d) Difference

Figure 25: Example 1 of non-rigid registration

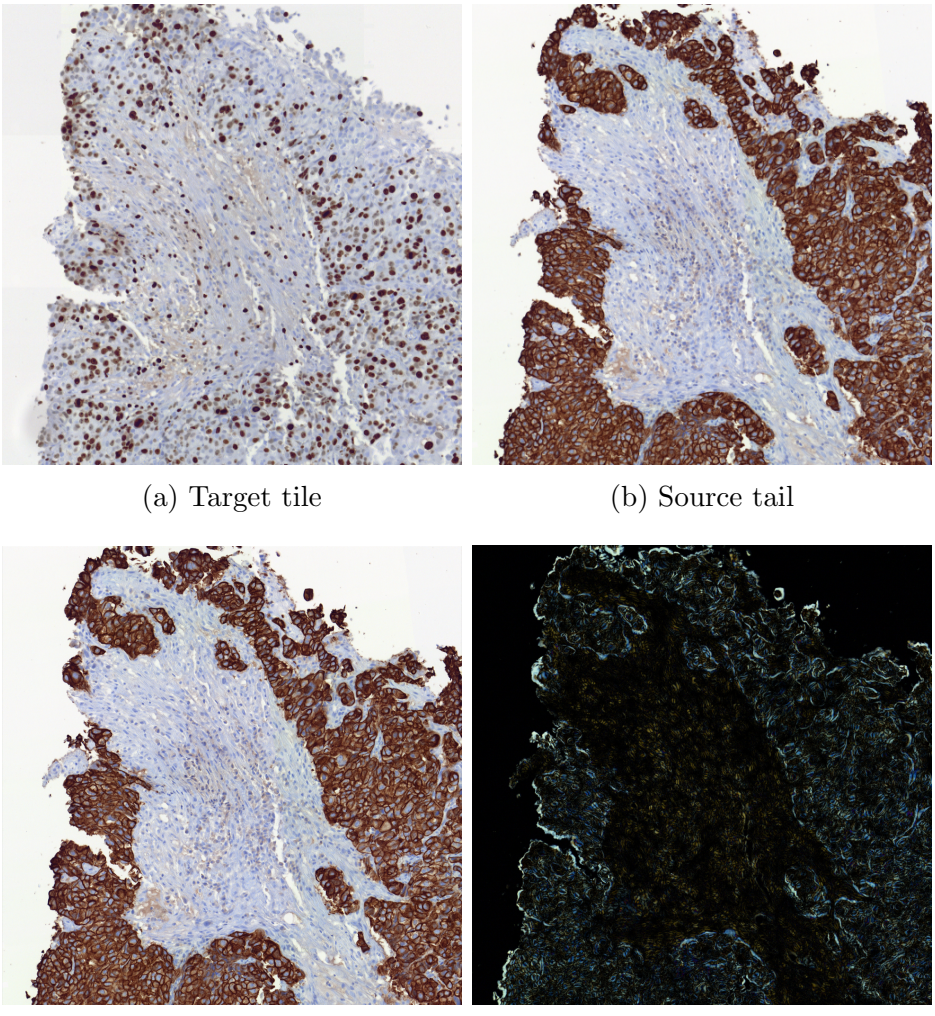
Details:
In Figure 26, there is shown how the deformation field moves the tissue of the source to align it to the target. The main problem is the fact that the deformation can't erase the tissue.

B.2 Example 2



(a) Target tissue (b) Source tissue (c) Transformed tissue

Figure 26: Details of non rigid registration



(a) Target tile (b) Source tail

(c) Transformed source tile (d) Difference

Figure 27: Example 2 of non-rigid registration