UNIVERSITÄT ZU LÜBECK

From the Institute of Mathematics and Image Computing
of the University of Lübeck
Director: Prof. Dr. rer. nat. Jan Modersitzki

Combined Local and Global Image Registration
and its Application to Large-Scale Images in Digital Pathology

Dissertation
for Fulfillment of
Requirements for the Doctoral Degree
of the University of Lübeck

from the Department of Computer Sciences and Technical Engineering

Submitted by

Johannes Lotz
from Berlin

Lübeck, 2019

First referee: Prof. Dr. Jan Modersitzki
Second referee: Prof. Dr. Heinz Handels
Date of oral examination: April 30, 2020
Approved for printing. Lübeck, May 5, 2020

Acknowledgements

I am thankful for the extraordinary support I received from many people during the creation of this thesis. I enjoyed the discussions, the joint work and the fun!

Many thanks to my advisor Jan for his guidance and his advice.

Special thanks to

My family

Judith, Paulina, and Jakob,
Gabi, Thomas, Kerstin, Anneliese, Beate, Matthias, and Irene,

Friends and Colleagues

Alessa, Alex, André, Andrea, Annkristin, Anja, Benjamin, Bernd, Björn, Caterina, Constantin, Dana, Daniel, Frederic, Hari, Henning, Herbert, Jan L., Jan R., Janine, Judith, Kai, Kanglin, Kerstin, Lars K., Lars R., Marc, Marian, Mark, Martin B., Nadine, Nick, Nils, Ole, Roland, Ruben, Sonja, Stefan, Sven, Thomas P., Thomas V., Till, and Yury,

Cooperation Partners

Andreas, Benedikt, Francesco, Geert, Jeroen, Kai, Maschenka, and Yi!

Contents

Summary

Large-scale images that exceed a single computer's random-access memory (RAM) represent a great challenge for nonlinear image registration. One area where these images occur is in digital pathology, where differently stained sections are superimposed to combine their information. Here, registration of large-scale images compensates for the deformation that is caused by a cutting and staining procedure.

In the literature, the decomposition of large-scale, nonlinear problems aims at increasing the available RAM by distributing the problem to multiple computing nodes. Two critical building blocks common among many nonlinear decomposition approaches are the formulation of local nonlinear subproblems and the coupling of the local subproblems to the global problem.

In this thesis, we present the novel **c**ombined fine **l**ocal and coarse **g**lobal (CLG) registration method. Decomposing the large-scale problem into smaller subproblems, CLG adapts the building blocks from domain decomposition for a single computing node. Unlike previously proposed methods, CLG links independent fine local registration problems to a coarse global representation of the global image, which improves the compatibility of the local solutions. By solving local problems independently of each other, only a fraction of the high-resolution image data needs to be loaded into RAM at once. The proposed method can be expressed as a discretization of a global, variational formulation of the registration problem. The downside of the new approach are potential inconsistencies between the solutions of the subproblems, which are compensated by a blending strategy.

To combine the local solutions, we adapt and compare different blending approaches from the literature. The proposed blending scheme computes a globally smooth deformation and preserves the homogeneity of neighboring solutions.

The accuracy and improved compatibility of local deformations are demonstrated in academic examples. We compare the proposed method to a coarse global and to a fine, purely local variant. In these experiments, the new CLG method is always more accurate than a coarse global registration and at least as accurate as a purely local registration in terms of image distance and deformation error. Registering the images sequentially, the CLG method does not compute one but a combination of multiple globally regularized deformations. The new method outperforms a purely local registration in terms of deformation error and irregularity measure in cases where the subdomain boundary region consists of low-contrast image information.

In addition to academic examples, we apply the new method to clinical whole-slide image data—each of multiple gigapixels in size—that comprises four different staining combinations and originates from two independent laboratories. When evaluat-

ing a total of 82 manually placed landmarks, the CLG and fine global registration are of similar accuracy, while CLG uses only a fraction of the RAM.

An exemplary image pair of 24 000 × 54 000 pixels is registered on a single workstation using 6 GB of RAM where—otherwise—more than 32 GB would be needed. Independently of the image size, only a fraction of the RAM is required to solve the same registration problem. While a standard registration method would be faster for small problems, the CLG method enables the registration of large-scale images that could otherwise not be computed.

Brief thesis summary
A large-scale, nonlinear image registration problem can be partitioned into smaller independent subproblems by adding a global, coarsely discretized distance measure. The remaining inconsistencies between subdomains are smaller than without the coarse distance term and can be incorporated into the global solution by a blending method. Reaching a similar accuracy, the new method enables the registration of large-scale images that could otherwise not be computed.

Zusammenfassung

Große, hochaufgelöste Bilder, die den Arbeitsspeicher eines einzelnen Computers übersteigen, stellen eine Herausforderung für die nichtlineare Bildregistrierung dar. Ein Bereich, in dem diese Bilder auftreten, ist die digitale Pathologie. An Schnitten von Gewebeproben treten durch Schneide- und Färbeverfahren verursachte Verformungen auf, die durch Bildregistrierung kompensiert werden. Dadurch können unterschiedlich gefärbte Gewebeschnitte überlagert und die darin enthaltenen Informationen kombiniert werden.

Die in der Literatur als *nonlinear domain decomposition* bezeichnete Zerlegung von großen nichtlinearen Problemen zielt darauf ab, ein nichtlineares Problem auf mehrere Computer zu verteilen und dadurch den verfügbaren Arbeitsspeicher zu erhöhen. Zwei kritische Bausteine, die bei vielen nichtlinearen Ansätzen vorkommen, sind die Formulierung lokaler nichtlinearer Teilprobleme und die Kopplung der lokalen Teilprobleme an das globale Problem.

In dieser Arbeit stellen wir das neuartige kombiniert lokale und globale—**c**ombined **l**ocal and **g**lobal (CLG)—Registrierungsverfahren vor. Das CLG-Verfahren zerlegt das große Problem in kleinere Teilprobleme, indem die Bausteine aus der *domain decomposition* für einen einzelnen Rechenknoten adaptiert werden. Im Gegensatz zu bisher vorgeschlagenen Methoden verknüpft das CLG-Verfahren fein aufgelöste, voneinander unabhängige, lokale Registrierungsprobleme mit einer groben Diskretisierung des globalen Bildes, wodurch die Kompatibilität zwischen den lokalen Lösungen verbessert wird. Durch die unabhängige Lösung lokaler Probleme muss nur ein Bruchteil der hochauflösenden Bilddaten auf einmal in den Arbeitsspeicher geladen werden, wodurch der Speicherbedarf reduziert wird. Die vorgeschlagene Methode kann als Diskretisierung einer globalen, variationellen Formulierung des Registrierungsproblems formuliert werden. Der Nachteil des neuen Ansatzes sind mögliche Inkonsistenzen zwischen den Lösungen der Teilprobleme, die durch eine Mischungsstrategie kompensiert werden.

Um die lokalen Lösungen zu kombinieren, werden verschiedene Mischungsstrategien aus der Literatur angepasst und verglichen. Das vorgeschlagene Mischungsverfahren berechnet eine global glatte Deformation und bewahrt die Homogenität benachbarter Lösungen.

Die Genauigkeit und verbesserte Kompatibilität von lokalen Deformationen werden an akademischen Beispielen demonstriert. Das vorgeschlagene Verfahren wird mit einer grob-globalen und einer fein-lokalen Registrierung verglichen. In diesen Experimenten ist das neue CLG-Verfahren immer genauer als eine grob-globale Registrierung und mindestens so genau wie eine rein fein-lokale Registrierung in Bezug auf

Distanzmaß und Deformationsfehler. Durch die sequentielle Registrierung der Bilder berechnet die CLG-Methode nicht eine, sondern eine Kombination aus mehreren global regularisierten Deformationen. Das neue Verfahren übertrifft eine rein lokale Registrierung hinsichtlich Deformationsfehler und Regularisiererwert in den Fällen, in denen die Grenze der Gebiete aus kontrastarmer Bildinformation besteht.

Zusätzlich zu den akademischen Beispielen wird das neue Verfahren auf klinische Bilddaten—so gennante *Whole Side Images*—angewendet. Jedes dieser Bilder hat eine Größe von mehreren Gigapixeln. Insgesamt werden vier verschiedene Färbekombinationen aus zwei unabhängigen Laboren untersucht. Bei der Auswertung von insgesamt 82 manuell gesetzten Landmarken sind das CLG-Verfahren und das Referenzverfahren—eine feine, global regularisierte Registrierung—von ähnlicher Genauigkeit, während das CLG-Verfahren nur einen Bruchteil des Arbeitsspeichers benötigt.

Ein exemplarisches Bildpaar von 24 000 × 54 000 Pixeln kann mit dem CLG-Verfahren auf einem einzelnen Computer mit 6 GB RAM registriert werden, während ein Standardverfahren mehr als 32 GB benötigen würden. Unabhängig von der Bildgröße wird nur ein Bruchteil des Arbeitsspeichers benötigt, um das gleiche Registrierungsproblem zu lösen. Während ein Standardverfahren für kleine Probleme schneller ist, ermöglicht das CLG-Verfahren die Registrierung von großen Bildern, die sonst nicht berechnet werden könnten.

Kurzzusammenfassung
Ein großes, nichtlineares Bildregistrierungsproblem kann durch einen zusätzlichen globalen, grob diskretizierten Distanzterm in kleine, voneinander unabhängige Teilprobleme zerlegt werden. Die verbleibende Inkonsistenz zwischen den Lösungen der Teilprobleme ist kleiner als bei der Berechnung ohne den zusätzlichen Distanzterm. Die lokalen Lösungen können durch ein Mischungsverfahren zu einer globalen Lösung zusammengeführt werden. Das neue Verfahren ermöglicht die Registrierung von großen Bilddaten, die mit dem Referenzverfahren nicht berechnet werden können und erreicht dabei eine ähnliche Genauigkeit.

1 Introduction to Large-Scale Image Registration in Digital Pathology

Contents

Image Registration of two given images is the process of finding a reasonable transformation that maps elements from the reference image to corresponding elements in the template image [Mod09].

One application for image registration is in digital pathology. To characterize specific cells and growth patterns, histological whole slide images are cut from a block of tissue—similarly to slicing bread—stained to reveal tissue properties, and scanned to obtain a digital image [Muk17]. Image registration is used to compensate for the deformation that is caused by the cutting and staining procedure such that differently stained sections can be superimposed [And17] and their information combined.

Today, the scanned images exceed the available random-access memory (RAM, also referred to as the main memory) of a single computer, which is a challenge for image registration.

In most image processing algorithms, the image data is loaded into the computer's RAM due to its faster access times [Mey03, pp. 1–4]. If the image data exceeds the available RAM, one can either reduce the image data [Car10; Sch13] or resort to parallel computation using domain decomposition [Ron17; Man16] which combines the memory of multiple computing nodes in a supercomputing cluster.

However, in many situations, cluster computing systems are unavailable, and the complete, unreduced registration problem is to be solved on one computer with insufficient RAM. In this case, local solutions can be computed by processing smaller parts of the image subsequently [And17; Mag15; Rob12]. These local solutions are each independent of the rest of the image, and the local deformations are therefore incompatible.

This thesis addresses the *large-scale image registration problem* of registering images that are too large to be loaded into a computer's RAM. Based on concepts originating in domain decomposition, we propose a nonlinear registration scheme that combines the image distance on a coarse global scale and a fine local scale while maintaining a globally smooth deformation. The proposed algorithm uses only a fraction of the RAM that is required to compute the same registration problem globally.

Unlike previously proposed methods, our approach links a local registration problem to a coarse representation of the global image, which improves the compatibility of the local solutions. In addition, the proposed method can be expressed as a discretization of a global, variational formulation of the registration problem. We demonstrate the superiority of the new method compared to a purely coarse and to a purely local method in numerical examples with real and artificial images. We also apply the new method to microscopic whole-slide image data.

The image data for these and many other experiments along with practical registration problems and insights into various applications have been kindly provided by our cooperation partners without whom this work would not have been possible: Dr. Jeroen van der Laak and his team at the Radboud University Medical Centre in Nijmegen, Dr. Kai Breuhahn from the University of Heidelberg, and Dr. Andreas Turzynski from the pathology clinic in Lübeck.

The thesis is structured as follows. In Chapter 1, we define the central problem discussed in this thesis: The image registration of images that exceed the main memory capacity of the computer at hand. One application for this problem is in digital pathology, and we briefly describe the steps from tissue block to digital image that lead to the image registration problem.

In Chapter 2, we formulate registration in a general framework for variational image registration following the book by Modersitzki [Mod09]. In Chapter 3 we give an overview of the existing work in the area of registration of histological whole slide images and compare domain decomposition methods that are used to distribute the computation on multiple computing nodes. In Chapter 4, we propose the new registration method based on the variational framework and discuss its properties based on a set of examples. We apply the new method to differently stained whole slide images from two independent laboratories (Chapter 7) and evaluate the registration accuracy based on landmarks. We conclude this thesis in Chapter 8.

Using only a fraction of the memory of a global registration, the new method solves the large-scale registration problem in model problems and real data, resulting in similar deformations to the ones computed by a global registration.

1.1 Virtual Multistaining in Digital Pathology

The first step in a registration pipeline in digital pathology is image acquisition. Following [Lan13; Mes16], we briefly describe the workflow in histo-pathology with a

Figure 1.1.1: The pathological workflow from tissue specimen to digital image. Top row: A microtome is used to cut the tissue which is embedded in a paraffin block. In the cutting process, the tissue is heavily deformed. After being cut, the sections are placed into a water bath to straighten the deformation. Bottom row: The section is fixated on a glass slide and stained to visualize structure and specific antibodies. The images are digitized using a whole slide scanner. Images 1–3 are provided by Mark Schenk, Fraunhofer MEVIS, Lübeck. The scanner image is provided by ZEISS.

focus on the acquisition and analysis of digitized, large-scale microscopic images.

In conventional pathology, a diagnosis is formed based on tissue properties that the pathologist evaluates by examining a tissue specimen under the microscope. The processing of tissue slides is standardized and depends on the diagnostic task. We will only give a brief overview here. For more details see, e.g., [Lan13]. A tissue sample is extracted from the body either by biopsy, where a needle is used to extract a small amount of tissue or by surgery, where larger tissue samples are extracted. The extracted tissue is then chemically processed, water is removed, and the sample is embedded into paraffin. The resulting block is cut into slices, which are usually between two and five micrometers thick [Sla11]. During the cutting process, the samples are nonlinearly deformed (Figure 1.1.1). The resulting section is placed in a water bath where it regains most of its original shape, and it is then fixated on a glass slide. The tissue is then stained to highlight specific cells or structures of interest and examined

under a microscope.

In digital pathology, the sectioned tissue is scanned and stored digitally. Digitization facilitates the exchange of tissue slides between doctors and enables the use of algorithms to analyze the tissue image [Zar19]. First digital workstations were approved for clinical use in the United States in 2017 [Eva18]. In contrast to radiology, where digitization meant to replace an x-ray-sensitive film by a digital sensor, digitization in pathology requires an additional scanning step in the clinical workflow. As a consequence, a digital pathology workflow has an increased cost due to additional required hardware and data storage. In the light of the shortage of pathologists in most parts of Europe, North America and the African Continent [Pet18; Nel18; Pat18], the additional flexibility for doctors and the possibilities for automation in a digital workflow seems to outweigh these costs and an increased adoption of digital pathology workflows can be observed [Wil18].

Staining is necessary to visualize the structure and specific properties of the tissue. Specific immunohistochemical stains are used to identify certain cells based on antigens or proteins that are expressed in the cell [Ova13, p. 479]. Often, multiple stainings are needed to form a diagnosis. One example is the grading of breast carcinoma, where two stains are used to determine the malignancy. Since both are expressed in brown, using both markers on the same slide is difficult [Mas00; Sta14].

Histological image registration is used to superimpose differently stained adjacent sections [And17; Koo18] in those cases, where the chemical combination of multiple markers is impractical.

1.2 Image Registration for 3D Reconstruction of Tissue Stacks

While this thesis is focused on the registration of two consecutive high-resolution slides, a frequent application of registration in digital pathology is 3D reconstruction, and we will give a brief overview here.

1.2.1 3D Reconstruction with a Reference Volume

Starting back in the 1990s, the first methods were established to reconstruct digitized histological data to 3D volumes, mostly for a better anatomical understanding of specific organs. Many authors formulate the goal to reconstruct 2D histological images to 3D volumes and fuse them to corresponding 3D volumes of another modality, e.g., MRI or PET scans [Meg97], block-face images [Cas17; Sch07; Dau07; Gef03; Bar02b; Kim97] or both [Sch98; Bar02a]. For this aim, every 2D histological slide is aligned with a corresponding slide of the reference volume. As the reference volumes are of limited image resolution, the used resolutions of the histological slide images are also low.

An extensive medical dataset is the BigBrain atlas [Amu13], recently updated with more data by [Moh16] consisting of 7600 slides of 7000 × 6000 pixels each that were registered to an MRI and also aligned slide to slide.

1.2.2 3D Reconstruction without a Reference Volume

Three-dimensional reconstruction without a reference volume has the disadvantage that some information of the original 3D shape of the object is missing from the data [Sch07]. For low-resolution images, affine or rigid registration approaches of two or more consecutive slides are proposed [Hib88; And92; Our01; Bar02b; Mal04; Xu15]. More complex deformation models allow a more accurate alignment. Examples include piecewise or weighted affine deformations [Ars05; Pit06; Hua06], B-spline deformations [Yun07; Feu11; Rob12], thin-plate splines [Cha06], a moving least squares approach applied to SIFT points [Car10] and elastic registration [Sch07].

One application for 3D reconstruction in cancer research can be found in [Bro14]. The authors describe how the presence of small groups of tumor cells (tumor buds) close to the tumor is related to a specific transformation of epithelial cells. Three-dimensional reconstruction is necessary to determine whether the tumor bud is connected to the main tumor. We contributed three-dimensional reconstructions to this article using a methodology similar to the one in [Lot14] (see below). Reconstructed tumor volumes are shown in Figure 1.2.1.

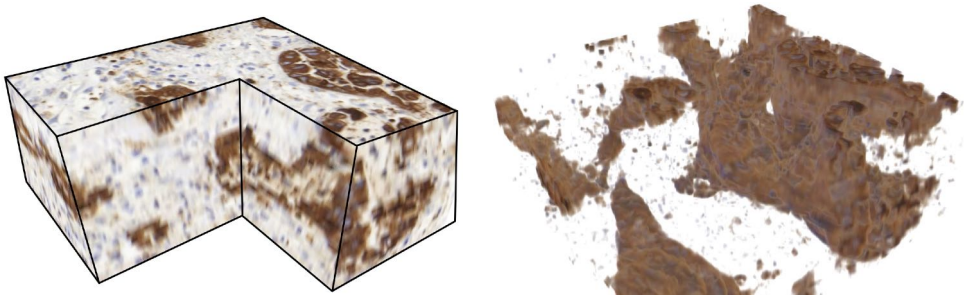
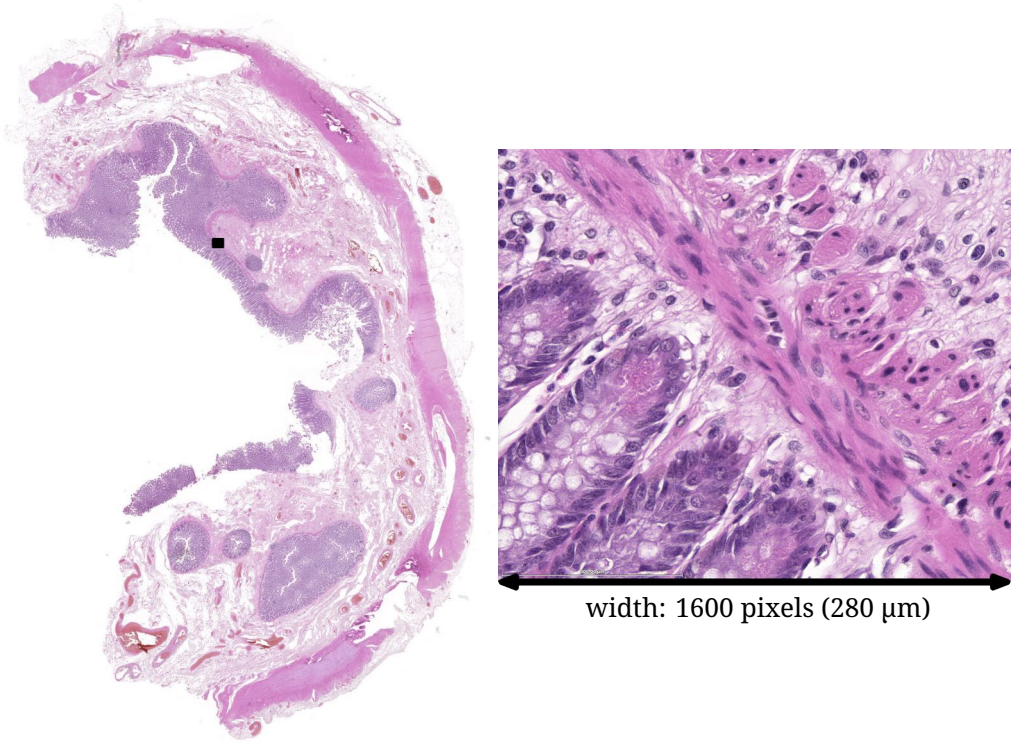


Figure 1.2.1: 3D reconstruction of tumor buds. Slides have been stained with pan-cytokeratin (PCK), tumor cells are brown. Left: 3D reconstruction visualization where a piece of the block is removed. Right: 3D reconstruction with thresholding to only show positively stained cells. Connections between cell groups are visible.

1.3 The Challenge of Large Images

One challenge when dealing with digital histological images in image processing is their size (Figure 1.3). Depending on the scanner's resolution, images often surpass the size of 116 659 × 188 744 pixels of the image shown in Figure 1.3, which corresponds to 66 GB of uncompressed image data. These images usually cannot be loaded

entirely into a computer's RAM. In the context of this thesis, we call images "large" if their uncompressed size exceeds the memory of the computer at hand. We will use the computer that is used to write this thesis as a reference (MacBook Pro (2016), 16 GB RAM, 2.6 GHz quad-core CPU).



a) Whole Slide Image
116 659 × 188 744 pixels
20 mm × 32 mm, 66 GB

b) Subimage
1600 × 1400 pixels
0.28 mm × 0.24 mm, 6.7 MB

Figure 1.3.1: Whole slide image of human colon, Haematoxylin and Eosin (H&E) stain, full resolution: 116 659 × 188 744 pixels. The area marked by the small black rectangle in the left image is shown in the right image. Note that fine structures such as cell nuclei (stained in dark pink) have a diameter of ca. 5 µm (30 pixels) such that they are invisible in the low-resolution overview image on the left. Circa 66 GB of RAM are needed to load the full-resolution image which exceeds the available memory of the computer that this thesis is written on. Image provided by Dr. Andreas Turzynski, Lübeck

The amount of memory that is required for the registration varies depending on the way these images are stored. Usually, images are stored as color images with 3 bytes per pixel. While the pathologist relies on the color of the different stains, most of the structure is still visible in gray-scale images. The images are therefore converted

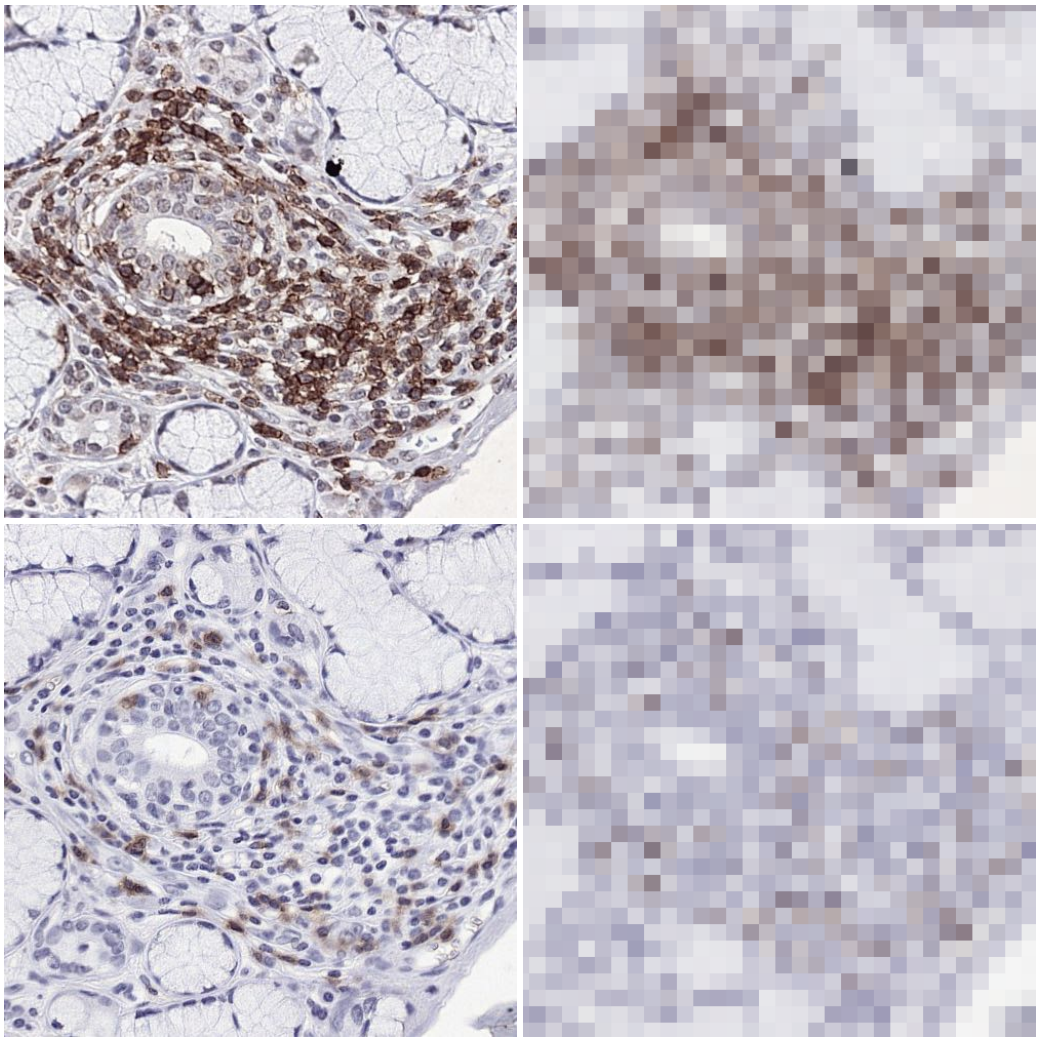


Figure 1.3.2: The same image region in two consecutive slides of human oral tissue, stained for CD8 and CD45, after registration. Left: Resolution 0.445 µm/px (20× magnification). Right: Resolution 7.12 µm/px (1x magnification). Smaller structures such as small vessels or nuclei may not be visible in the low resolution images. The images at 20× magnification occupy 400 times the memory of the images with 1x magnification.

prior to registration, reducing the storage to 1 byte per pixel. During registration, the deformed image is interpolated at sub-pixel positions, and an image derivative is computed. In these cases, we use a data type with higher precision to reduce rounding errors. This increases the memory requirements to up to 8 bytes per pixel (data type double) for the deformed image. Depending on the per-pixel memory requirement, the histological image in Figure 1.3 will need between 22 GB and 176 GB of storage in RAM. On the reference computer, loading two of these images is infeasible, especially if additional memory is required to compute the image registration.

We aim to register these images at a high resolution since the objects—such as individual cells—that need to be identified by a pathologist for a diagnosis measure 5 to 7 µm in diameter (red blood cells, [Ova13, p. 158]). These objects cannot be located clearly in a low image resolution image, and we expect the registration to be most accurate at a high image resolution. We demonstrate the effect of lowering the image resolution in Figure 1.3.2 in an example.

Large Images in Image Registration

We formalize the problem of registering large images following [Mod09] in a variational framework (see Chapter 2 for more details) and define the registration problem as the minimization of the objective function

$$\mathcal{J}(u) = \mathcal{D}(T, R, u) + \mathcal{S}(u) \xrightarrow{u} \min \tag{1.1}$$

with respect to a displacement $u : \Omega \mapsto \mathbb{R}^2$ on the image domain $\Omega \subset \mathbb{R}^2$ where $\mathcal{D}$ is a distance measure, and $\mathcal{S}$ is a regularizer. The distance measure is small if the reference image R and the deformed template image $T(u)$ are similar. The regularizer is small if the deformation implied by u is smooth. The above formulation provides a general description of the registration problem as a combination of the desired result (low image distance) and the necessary restriction in degrees of freedom (regularity).

At this point, no assumptions are made on the relation between the image data as it is measured in the scanner and the image functions R and T. In the following, we discuss different choices for the image model.

Discussion of Different Image Models

The most direct approach to model the image function is to use interpolation (see, e.g., [Thé00, pp. 393 ff.]) of the full pixel data at least in the final steps of the registration. We will use this method as the gold standard as the data model is very close to the acquisition of the image. In the whole slide scanner, image data is acquired by an array of pixels which gather the incoming light from the object plate. As discussed in this chapter, using this raw-data interpolation is not an option when processing the full-sized image.

A straightforward way to reduce the number of interpolation coefficients is to use downsampling, resulting in a coarse representation of the image, which can directly

be used to compute an approximate registration. However, downsampling results in a loss of high-frequency information and, therefore, in a loss of detail in the image (cf. Figure 1.3.2). More sophisticated methods to reduce the size of the image data are used in image compression.

Image compression aims to analyze the image for duplicate information and to store these duplicates only once. For instance, JPEG compression [Wal92] computes a discrete cosine transform for each image block and avoids storing similar block-representations. Some work has been published on registration of Fourier- or wavelet-transformed images, which are used in the JPEG2000 standard [Chr00]. An approach for parametric registration in the wavelet-space is proposed by [Ayd17], which would allow for a low-memory registration in combination with sparse wavelet-coefficients. However, a deformation of the image in position space has to be translated back into the compressed space of the respective basis functions. The author is not aware of an extension of this or similar approaches to a non-parametric deformation model.

Instead of compressing the entire image, surrogate models are used to reduce the image to key features. In [Wei15], we propose a registration scheme where two density functions are generated from the images based on the positions of the cell-nuclei. These density functions are directly used as image functions R and T. Similar approaches like the Coherent Point Drift method [Myr10] or the purely point-based Iterative Closest Points method [Zha94; Bes92] are used once corresponding feature-sets are determined in both images. The difficulty in applying these methods to the registration problem in digital pathology is the construction of the surrogate representation. In histological images, the appearance of the tissue varies in each staining. The automatic detection of corresponding features in different stains has not been solved reliably and is still a subject of ongoing research, see for example [Sir16; Höf18].

When reducing the image information is impractical, domain decomposition can be used to distribute the problem between multiple computers and solve it in parallel (Chapter 3.2). However, cluster computing systems are mostly available at larger research institutions and cloud computing providers and are not found in many pathology laboratories.

While we see much potential in the idea of using a compressed or surrogate image representation in registration, there are still too many practical hurdles to apply these approaches universally to histological images. Therefore, we build upon the standard interpolation approach and aim to find a solution to register memory-exceeding images on a local computer with limited RAM.

In this work, we focus on off-the-shelf workstation computers with limited main memory (also called random-access memory or RAM, see [Pat14, pp. 19 ff.]).

If the underlying image data $\mathtt{R}_N \in \mathbb{R}^N$ and $\mathtt{T}_N \in \mathbb{R}^N$ of the images R_N and T_N in the registration problem (1.1) is so large, that it cannot be stored in the computer's RAM, we call

$$\mathcal{J}(R_N, T_N, u) \xrightarrow{u} \min, \text{ with } 2N > \mathtt{MEM} \tag{1.2}$$

the **large-scale image registration problem**, where MEM is the number of pixels that can be accommodated and processed in the computer's available RAM. This number may vary from system to system and also depends on other processes running at the same time.

Before discussing image registration in more detail in Chapter 2, we will briefly sketch the problem that occurs when the images exceed the available memory.

The distance measure $\mathcal{D}$ in the objective function (Equation 1.1) is composed of the image functions R and T. As discussed above, we base these functions on the interpolation of the image data. As the objects of interest in the image—such as individual cells—are of few pixels in diameter, we want to interpolate the images using a fine grid of data points, which is ultimately equal to the pixels in the image data. To evaluate the distance measure, all pixels have to be taken into account and, therefore, loaded into the RAM. We note that evaluation of the distance measure alone can be done on a subdomain basis, where each subdomain is loaded into memory and evaluated independently. The same does not hold in combination with the regularizer.

The regularizer $\mathcal{S}$ is needed to solve the otherwise ill-posed registration problem (see Chapter 2.3). It is also used to make the registration adhere to a specific deformation model by minimizing an energy term based on a differential operator [Mod04, pp. 84 ff.]. As we point out in Chapter 2.8, computing an update step—for example, in a Gauß–Newton optimization—cannot be done independently on subdomains. Each step of the optimization on a subdomain requires image data from all or at least from neighboring subdomains. Therefore, the registration problem cannot be solved by solving smaller independent subdomain problems, which is challenging if not all data can be loaded at once.

A positive effect of global regularization is that smooth global deformations can be computed even when parts of the image are missing. One example is shown in Figure 1.3.3, where part of the tissue is torn off during preparation. Here, the deformation continues smoothly in the area of non-correspondence based on information from the intact tissue neighborhood.

As the main contribution of this thesis, we solve the dilemma between a globally coupled deformation model and large-scale image data by combining the image distance on a coarse global scale and on a fine local scale while maintaining a globally regularized deformation. This new approach is discussed in Chapters 4 and 5.

An Image Registration Algorithm for Large Images: Requirements

The criteria for a method to register large pathology images are

1. Nonlinear, global deformation.

   The deformation of the tissue in the sectioning process is nonlinear. We therefore want to allow nonlinear deformations, which in turn require the regularization of the registration problem.

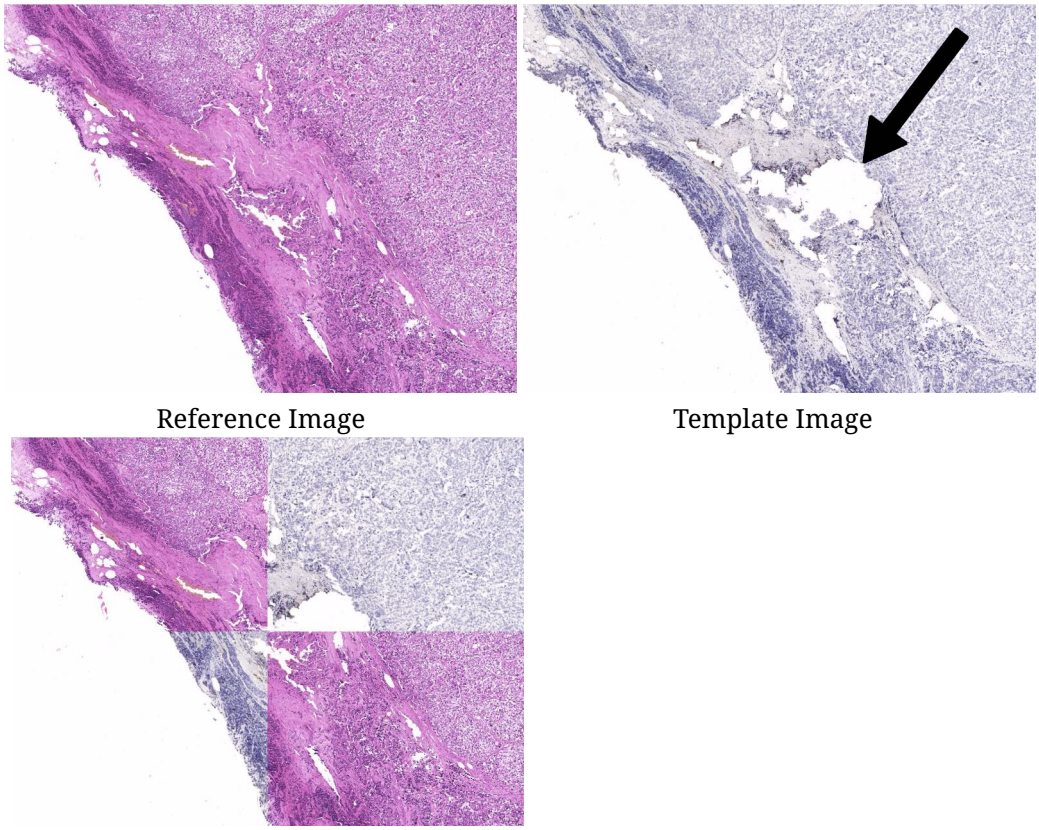


Figure 1.3.3: Two images of mammary gland tissue stained with H&E (top-left) and CK5/14 (top-right) after registration where part of the tissue is torn off during preparation. Local changes in the deformation in the torn region do not affect the distance measure. With regularization, the deformation is smoothly continued based on information from the intact tissue neighborhood. The correct alignment can be verified in the checkerboard image (bottom-left). Images provided by Dr. Andreas Turzynski, Lübeck.

2. High image resolution.

   The structures such as single cells or cell nuclei that need to be correlated by the registration often correspond to only a few pixels in the image. To allow these structures to contribute to the registration, we want to use the available image resolution.

3. No cluster computer needed.

   Supercomputing hardware is usually not available in a pathology lab, and transferring the image data is impractical due to legal reasons and bandwidth limitations.

1.4 Other Scientific Contributions Related to this Thesis

Preliminary results that are part of or linked to this thesis have already been published in [Lot14; Lot16a; Wei15]. Our submission to the Automatic Non-rigid Histological Image Registration (ANHIR) challenge received the first prize for the most accurate registration using the approach described in [Lot19]. We applied registration of histological images to medical and biological research problems and published the results with partners from these fields in [Bro14; Lot16b; Yin16; Bul19].

In [Lot14; Lot16a], we describe methods to register large images at a high image resolution.

The first method [Lot14] focuses on a small area of a specimen. A tissue block is reconstructed from a series of consecutive sections around a user-defined point of interest. At this selected point, a slide-pair is registered with increasingly fine resolution while reducing the size of the registered image area (zooming in). The process is repeated for subsequent slide pairs until the selected area of the tissue block is reconstructed.

The second method [Lot16a] is a subdomain-based registration method that repeats the zooming approach across the whole tissue section. We combine the resulting deformations for each subdomain by interpolation to form one global deformation.

The basis for the subdomain-based methods presented in this thesis is a robust and automatic coarse-resolution image registration. In [Lot19], we describe this method and also introduce a brute-force pre-alignment scheme that is used to obtain an initial guess for subsequent gradient-based optimization steps. Our method won first place in the challenge “Automatic Non-rigid Histological Image Registration” (ANHIR) that was part of the ISBI conference in 2019. N. Weiss and the author of this thesis contributed equally to the participation in this challenge.

Another strategy to register large images is to reduce the information by building a surrogate image model. In [Wei15], cell nuclei were detected automatically, and the registration problem is formulated by measuring the similarity of the density functions of the nuclei. This article is the result of a master’s thesis that was co-supervised by the author of this thesis.

In the following publications, we applied one of the above methods to a research problem from another field.

In [Bro14], we performed a 3D-reconstruction to determine if a tumor island is connected to the main tumor which cannot be told by observing single 2D slides.

Three-dimensional reconstruction is combined with matrix-assisted laser desorption ionization (MALDI) imaging in [Lot16b]. In order to build a combined model of MALDI imaging data and immunohistochemistry, we performed a 3D reconstruction of histological sections, which was also applied to MALDI imaging data. The result is a 3D model of a head and neck cancer that allows evaluating the functional heterogeneity of the tumor tissue.

Different histological stainings were combined in one 3D stack in [Yin16]. To analyze tumor vasculature and nutrient supply, we performed 3D reconstructions to correlate vessel measurements in histological sections with measurements from diffusion MRI.

In [Bul19], we contributed a preliminary subdomain-based version of the registration method proposed in this thesis to register differently stained sections to generate ground-truth for a machine-learning algorithm. A deep neural network is trained with a pair of images stained with a standard Haematoxylin and Eosin (H&E) stain and a and specific immunohistochemical stain that marks epithelium. The trained network can be used to predict epithelial regions in the H&E-stained slides.

Publications of Methods

1. [Lot14] J. Lotz, J. Berger, B. Müller, K. Breuhahn, N. Grabe, S. Heldmann, A. Homeyer, B. Lahrmann, H. Laue, J. Olesch, M. Schwier, O. Sedlaczek, and A. Warth. “Zooming in: High Resolution 3D Reconstruction of Differently Stained Histological Whole Slide Images”. In: *Proc. SPIE 9041*. Medical Imaging 2014: Digital Pathology. Ed. by M. N. Gurcan and A. Madabhushi. San Diego, California, USA, 2014, p. 904104. DOI: 10.1117/12.2043381 (peer-reviewed conference proceedings)
2. [Lot16a] J. Lotz, J. Olesch, B. Müller, T. Polzin, P. Galuschka, J. M. Lotz, S. Heldmann, H. Laue, A. Warth, B. Lahrmann, N. Grabe, O. Sedlaczek, K. Breuhahn, and J. Modersitzki. “Patch-Based Nonlinear Image Registration for Gigapixel Whole Slide Images”. In: *IEEE Transactions on Biomedical Engineering* 63.9 (2016), pp. 1812–1819. DOI: 10.1109/TBME.2015.2503122 (peer-reviewed journal article)
3. [Lot19] J. Lotz, N. Weiss, and S. Heldmann. *Robust, Fast and Accurate: A 3-Step Method for Automatic Histological Image Registration*. arXiv:1903.12063 [cs]. 2019 (preprint, full article in preparation)
4. [Wei15] N. Weiss, J. Lotz, and J. Modersitzki. “Multimodal Image Registration in Digital Pathology Using Cell Nuclei Densities”. In: *Bildverarbeitung Für Die Medizin 2015*. Ed. by H. Handels, T. M. Deserno, H.-P. Meinzer, and T. Tolxdorff.

Berlin, Heidelberg: Springer, 2015, pp. 245–250. DOI: 10.1007/978-3-662-46224-9_43 (peer-reviewed conference proceedings)

Publications of Applications

5. [Bro14] P. Bronsert, K. Enderle-Ammour, M. Bader, S. Timme, M. Kuehs, A. Csanadi, G. Kayser, I. Kohler, D. Bausch, J. Hoeppner, U. Hopt, T. Keck, E. Stickeler, B. Passlick, O. Schilling, C. Reiss, Y. Vashist, T. Brabletz, J. Berger, J. Lotz, J. Olesch, M. Werner, and U. Wellner. "Cancer Cell Invasion and EMT Marker Expression: A Three-Dimensional Study of the Human Cancer-Host Interface: 3D Cancer-Host Interface". In: *The Journal of Pathology* 234.3 (2014), pp. 410–422. DOI: 10.1002/path.4416 (peer-reviewed journal article)
6. [Lot16b] J. M. Lotz, F. Hoffmann, J. Lotz, S. Heldmann, D. Trede, J. Oetjen, M. Becker, G. Ernst, P. Maas, T. Alexandrov, O. Guntinas-Lichius, H. Thiele, and F. von Eggeling. "Integration of 3D Multimodal Imaging Data of a Head and Neck Cancer and Advanced Feature Recognition". In: *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* (2016). DOI: 10.1016/j.bbapap.2016.08.018 (peer-reviewed journal article)
7. [Yin16] Y. Yin, O. Sedlaczek, J. Lotz, J. Olesch, K. Breuhahn, D. Drasdo, and I. E. Vignon-Clementel. "Tumor Microvasculature in Lung Cancer and Diffusion-Weighted MRI: Preliminary Results". In: *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop Proceedings (NSS/MIC/RTSD)*. Strasbourg: IEEE, 2016. DOI: 10.1109/NSSMIC.2016.8069545 (peer-reviewed conference proceedings)
8. [Bul19] W. Bulten, P. Bándi, J. Hoven, R. van de Loo, J. Lotz, N. Weiss, J. van der Laak, B. van Ginneken, C. Hulsbergen-van de Kaa, and G. Litjens. "Epithelium Segmentation Using Deep Learning in H&E-Stained Prostate Specimens with Immunohistochemistry as Reference Standard". In: *Scientific Reports* 9.1 (2019). DOI: 10.1038/s41598-018-37257-4 (peer-reviewed journal article)

2 A Variational Perspective on Nonlinear Image Registration

Contents

Image registration is the process of bringing corresponding structures in two or more images into alignment [Fis08]. In digital pathology, image registration of large whole slide images is used to combine information from different biomarkers [Mue11]. We established criteria for the registration of these images in the previous chapter. The core of this thesis is the discussion of a new subdomain-based image registration method that combines global regularization with local image processing.

Following [Mod09], we formulate registration in a general framework for variational image registration. In this framework, nonlinear registration of two images is an optimization problem based on an objective function consisting of a distance measure and a regularizer. The distance measure will be small if the two images are similar. However, as outlined in Chapter 2.3, an optimization problem that is purely based on image distance is ill-posed, and regularization is needed to find a solution [Fis08]. Different choices are available for both of these components, and we will discuss some of them in the following sections.

We formulate the registration problem in the variational framework because it provides a general description of the registration problem as a combination of the desired result (image distance) and the necessary restriction in degrees of freedom (regularity). This formulation is well suited to describe the challenge in the registration of large images—the combination of local data fit with a global deformation (see Chapter 4.2). To compute a solution, we use the discretize-then-optimize approach proposed by Haber and Modersitzki [Hab04] as a baseline method because it has shown good results in other areas of medical imaging [Rüh17b; Kön18; Ole14; Pol14] in terms of memory footprint, computation time and registration accuracy. Furthermore, standard optimization methods can be used by computing analytic derivatives of the discretized objective function. We build upon this framework, discretize the deformation, and compute a minimum using a Gauß–Newton or L-BFGS iteration (see Chapter 2.5).

To compute a global deformation, an energy term is minimized across the entire image domain. In this chapter, we show that regularization in the sense of the variational framework couples the registration problem such that a solution cannot be computed independently on subdomains. Even if one is only interested in the solution in a subdomain of the image, global image data is needed to solve the registration problem. However, high-resolution image data is not available globally for large-scale images. To the knowledge of the author, all available methods to register large images either loosen the requirement of global regularization [Son14; Yig17] or use a cluster of multiple computing nodes to increase the available amount of memory [Mod99; Amu13; Man16].

In this chapter, we will follow the book by Modersitzki [Mod09] and briefly summarize the variational framework and a straightforward discretization. Extensions to this work that addresses the problem of registering images exceeding the memory capacity of the computer will be discussed in Chapter 4.

2.1 Structure of the Optimization Problem

We understand nonlinear registration of two images as an optimization problem based on an objective function $\mathcal{J}$ with

$$\mathcal{J}(R, T, u) = \mathcal{D}(R, T(y)) + \mathcal{S}(u) \xrightarrow{u} \min \tag{2.1}$$

where $\mathcal{D}$ is a distance measure, and $\mathcal{S}$ is a regularizer. We define images as functions $R : \Omega \to \mathbb{R}$ and $T : \Omega_T \to \mathbb{R}$, $\Omega \subset \mathbb{R}^d$, $\Omega_T \subset \mathbb{R}^d$, that map a gray value to each point of an image domain. In the context of histological images, we will only consider two-dimensional images with $d = 2$ and one color channel. Color images are converted to grayscale by computing a weighted sum of the color channels [Bur09, pp. 109 ff.].

In the following, we will denote $y : \Omega \to \mathbb{R}^2$, $y(x) = u(x) + x$, $x \in \Omega$ as the deformation and $u(x)$ as the displacement. While the displacement $u(x)$ describes how each point $x \in \Omega$ is altered, the transformation $y(x)$ is a direct mapping from a point x in the reference image to its new location in the template image. To keep the notation

concise, we will sometimes drop the argument of the functions u and y throughout this chapter.

In the present approach, one image is regarded as the reference image and remains unchanged. In digital pathology, the H&E stained image is often chosen as the reference because it is often registered to multiple other images of different staining [Mue11]. The reference image is often referred to as the fixed image, while the template image can be deformed by evaluating it in concatenation with the transformation $y(x)$. We refer to the deformed template image as $T(y)$.

Next to the deformed image, the displacement is the main result of the registration, and its plausibility is often more important than a pleasing match of the images' gray values. One intuitive example can be found in lung registration, where the deformation can be used to analyze the ventilation of the lung [Mur12]. Unfortunately, there is no clear definition of a plausible deformation. One rule of thumb is the following: Given two registration results with the same image distance, the one with the smoother deformation—in the sense of a regularity measure—is more plausible [For10]. Especially in nonlinear registration, finding the balance between data fit and smoothness of the deformation is part of the registration problem.

Deformations can be divided into two groups: parametric and non-parametric. In the first group, the degrees of freedom of the deformation y are limited by the number of parameters. For a linear deformation, y is written as a linear function

$$y(x) = Ax + b$$

with $A \in \mathbb{R}^{2\times 2}$ and $b \in \mathbb{R}^2$, which allows global translations and rotations but also scaling and shearing. These deformations are intrinsically smooth and of a global nature with limited degrees of freedom. The same deformation rule is applied to all points in the image domain. In turn, local deformations cannot be compensated by this model. Such rigid or affine registration schemes are often used to obtain an intermediate registration based on a low image resolution. The obtained deformation is then used as an initial guess for a registration that uses a nonlinear deformation model.

In non-parametric registration, no assumptions need to be made on the deformation function such that each point $x \in \Omega$ in the image domain can be moved independently. This means that for every point in the image, a deformation vector needs to be found. With only the image distance as the optimality criterion, the registration problem is ill-posed in the sense of Hadamard (see Chapter 2.3). Regularization is required in order to find a solution.

A regularizer is added to the model to reduce the degrees of freedom in the objective function such that the solution is unique. By the choice of the regularizer, a bias towards specific properties of the deformation is introduced. Popular choices like elasticity [Bro81] or curvature [Fis03b] regularization are based on the norm of differential operators that are applied to the displacement, see [Mod09, Chapter 8].

In the registration of histological images, we assume a nonlinear deformation of the tissue that is caused by the cutting process. However, we do not have evidence that

the deformation follows any specific nonlinear deformation model. Therefore, we focus on a regularized, non-parametric image registration approach, where nonlinear deformations can be accommodated, no explicit selection of a basis function is needed, and the deformation model is only influenced by the choice of the regularizer.

We briefly summarize distance measures and regularizers in the next two chapters.

2.2 Distance Measures

In this thesis, two measures of similarity are used. The sum of squared differences (SSD) distance measure (see, e.g., [Mod04])

$$\mathcal{D}_{\text{SSD}}(R, T, u) = \int_{\Omega} (R(x) - T(y(x)))^2 \, \mathrm{d}x$$

is based on intensity differences in the images and is used for mono-modal image pairs. SSD is one of the simplest distance measures and has no additional parameters to choose such that it is ideal for numerical experiments. We use SSD for the experiments in Chapter 6.1, where both images are based on the same image data.

In digital pathology, images are a result of different staining, and therefore, corresponding structures are often of a different color. Figure 2.2.1 shows an example where the same tissue section has been stained, scanned, and in a second step restained with another stain. This process has been repeated four times, resulting in four images that represent the same object in different colors.

To cope with the multimodality of differently stained image sections, we use the Normalized Gradient Field (NGF) distance measure [Hab07a]. The main idea in NGF is that the changes in intensity in the registered images "occur at the same locations" [Hab07a]. The intensity changes are modeled as normalized gradients to adjust for the different intensity levels of the corresponding structures in different modalities.

The NGF distance measure can be written as

$$\mathcal{D}_{\text{NGF}}(T, R, u) = \int_{\Omega} 1 - \left(\frac{\nabla T(y(x))^T \nabla R(x) + \epsilon^2}{\|\nabla T(y(x))\|_\epsilon \, \|\nabla R(x)\|_\epsilon} \right)^2 \mathrm{d}x \tag{2.2}$$

where $\|x\|_\epsilon^2 := \|x\|_2^2 + \epsilon^2$.

The edge parameter ϵ in the NGF distance measure is used to differentiate noise and edges in the images. It can be set depending on the overall noise level in the image. We implement the automatic choice of the parameter

$$\epsilon = \frac{\eta}{V} \int_{\Omega} |\nabla R(x)| \mathrm{d}x \tag{2.3}$$

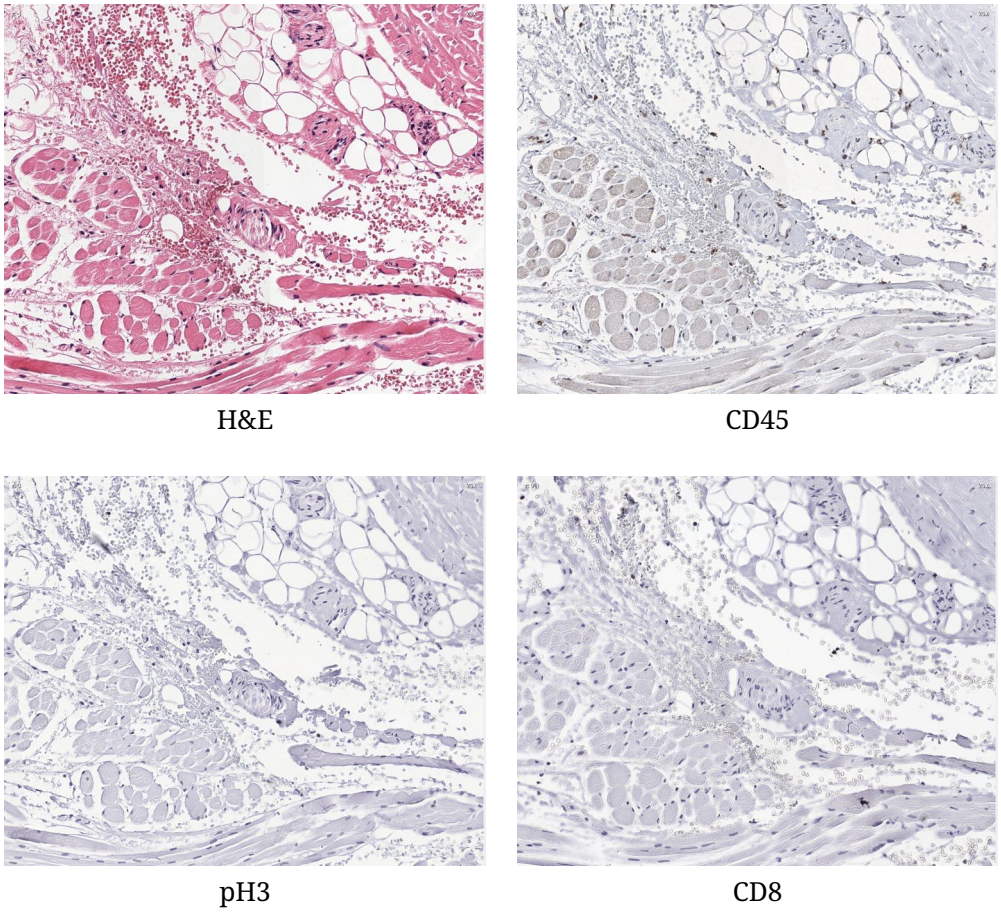


Figure 2.2.1: Restained section with four different stains. Although the staining differences are substantial, common structures are visible throughout all images. Images kindly provided by J. van der Laak and his team, Radboud UMC, Nijmegen

proposed by [Hab07a] where R is an image η is the noise level, which remains to be estimated manually. The advantage of estimating η is that it is less dependent on the image since the average image gradient captures the effect of global intensity change or change in resolution.

Another popular multimodal distance measure that has been successfully applied in the context of medical image registration [Bro17; Plu03] is mutual information (MI) [Vio97]. MI is based on estimating the entropy of joint image histograms, and one of the main advantages of MI is that the gray values in the images do not need to correlate linearly [Plu03]. We choose NGF over MI because the NGF can be computed locally, which is not possible for MI since all image values are required for the histogram calculation.

For other distance measures used in medical image registration, we refer the reader to the survey by Keszei, Berkels, and Deserno [Kes17]. We will not discuss them here because the new method proposed in this thesis is independent of the choice of the distance measure as long as it can be formulated in the variational framework.

2.3 Regularization

An image registration problem that is purely based on image distance is an ill-posed problem in the sense of Hadamard [Had02]. A practical example of a registration problem with multiple solutions can be found in the histological images discussed in Chapter 1 (Figure 1.3.3), where parts of the image consist of background or homogeneous unstained area. A local disturbance in the deformation in these areas can be found such that the distance measure remains unchanged. To find a consistent deformation in these areas, one has to rely on information from their neighborhood.

Regularization is needed to define a unique, stable solution [Mod04, p. 3] of the registration problem. It is also used to introduce a bias towards desired properties such as a low intrinsic energy of the deformation with respect to the respective regularization operator [Mod09, Chapter 8].

While the distance measure can be understood as the external force that promotes image similarity, the regularizer is the inner force that leads to smooth deformations [Fis04]. When partitioning the image domain into subdomains, special care has to be taken to preserve the smoothness across subdomain borders.

A frequently used [Oli14; Sot13] regularization approach is to minimize the norm of a differentiable displacement function u subject to a differential operator B applied to the displacement

$$\mathcal{S}(u) = \int_\Omega (Bu)^2 \, \mathrm{d}x = \int_\Omega \langle Bu, Bu \rangle_2 \, \mathrm{d}x$$

along with the distance measure. First-order methods include diffusive [Fis01] or elastic [Bro81] regularization, a second-order operator is used in curvature [Fis03b] regularization.

The new method presented in this thesis is independent of the choice of the regularizer, and we do not have enough insight into the physical deformation of the tissue during sectioning to define an accurate model. To keep the registration model uncomplicated, we concentrate on curvature regularization as an elementary case of second-order regularization.

We briefly summarize the regularizer following the book by Modersitzki [Mod09]. The curvature regularizer

$$\mathcal{S}_{\text{curv}}(u) = \alpha \int_{\Omega} \|B_{\text{curv}} u(x)\|_2^2 \, \mathrm{d}x$$

uses a second-order operator and yields smoother deformations. Here, the Laplace operator Δ is applied to each component of u and the operator $B_{\text{curv}} : \mathcal{C}^n(\mathbb{R}^2) \to \mathcal{C}^{n-2}(\mathbb{R}^2)$ with

$$B_{\text{curv}} = \begin{pmatrix} \Delta & 0 \\ 0 & \Delta \end{pmatrix}$$

acts on u such that $B_{\text{curv}} u = \begin{pmatrix} \Delta u_1, & \Delta u_2 \end{pmatrix}$.

A solution u to the registration problem in (2.1) has finite regularizer energy. In the case of curvature regularization, it is therefore at least twice differentiable. The kernel of the curvature regularizer includes affine-linear deformations such that these deformations have zero regularizer energy. For the experiments in Chapters 6.1 and 7, we choose curvature regularization because of the higher smoothness of the solution. As an additional advantage, the curvature energy can be used to directly compare the inner energy of multiple deformations independently of an affine preregistration.

2.4 Parameter Search Strategy

To balance image similarity and inner energy of the deformation, a parameter $\alpha \in \mathbb{R}$ is included in the regularizer. The strategy to choose α is not obvious as it balances two different quantities that are independently scaled [Cac01]. Some research has been targeted at determining α in a multilevel context [Hab06], but to the author's knowledge, there is no established method that is robust and automatic. To determine the optimal parameter for a registration experiment, we resort to a parameter search strategy.

To find the optimal parameter we first define a search range $\alpha \in [0.0001, 1000]$. In previous experiments, values outside of this range have not resulted in satisfactory registration results. The range is sampled adaptively and a registration is performed for each candidate parameter. The result is evaluated based on regularizer energy, image distance and a quality measure (Figure 2.4.1). When a ground truth deformation is available, we compute the deformation error $e_{\text{true}} = \|\mathbf{y} - \mathbf{y}_{\text{true}}\|_2$ to the true deformation. In the absence of a ground truth, we use manually placed landmarks

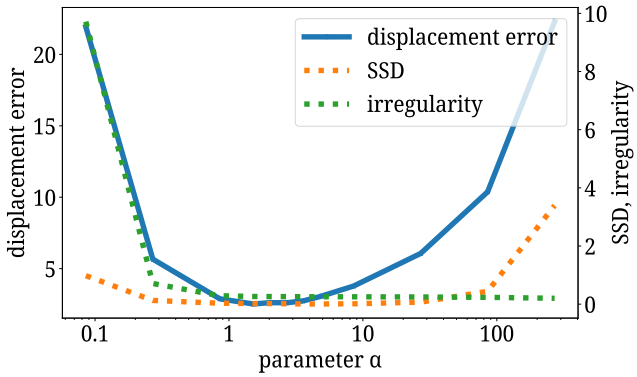


Figure 2.4.1: SSD image distance, irregularity, and deformation error relative to the regularizer parameter α for an example registration. The optimal parameter ($\alpha = 2.7$) is selected by the lowest displacement error (blue).

and compute the average landmark error norm $e_{\text{LM}} = \sum_i \|\mathbf{l}_i^T - y(\mathbf{l}_i^R)\|_2$ of template landmark $\mathbf{l}^T$ and transformed reference landmark $y(\mathbf{l}_i^R)$. The parameter resulting in the lowest error is selected for the respective registration.

The author is unaware of a theoretically funded method to determine the optimal weight for the regularizer. The main disadvantage of this strategy is that some ground truth data is needed. However, if such data is available, the parameter search enables the comparison of different registration methods without introducing a bias caused by the choice of the parameter.

2.5 Discretization

Being acquired as pixel data $\mathtt{T} \in \mathbb{R}^{n_1 \times n_2}$, a continuous representation of the image is obtained by interpolation. We discuss other image models apart from pixel-based interpolation in Chapter 1.3.

2.5.1 Image Interpolation

We denote the interpolation of the template image $T : \mathbb{R}^2 \to \mathbb{R}$ at a point $(x, y) \in \mathbb{R}^2$ given image data $\mathtt{T}$ as

$$T((x,y)) = \text{interpolate}((x,y), \mathtt{T})$$

and will omit the explicit mention of the image data when possible.

Different interpolation functions are used in image registration, see [Thé00] for a comparative overview. We use cubic B-Spline interpolation [dBoo01, pp. 87 ff.] in the numerical experiments in Chapter 6.1 and resort to linear interpolation in Chapter 7 for the registration of large-scale images.

Cubic B-Splines are continuous and twice differentiable. Especially differentiability is needed to maintain the quadratic reduction of the quadrature error when applying the midpoint rule in (2.4).

Linear interpolation results in an interpolant that is continuous but not continuously differentiable at the cell boundaries. This violates the requirement in (2.4) and therefore risks a slower convergence. We use it nevertheless when dealing with large-scale image data for its more straightforward implementation and its faster evaluation, which does not require pre-computation of coefficients. Comparing linear and cubic B-spline evaluation, Thévenaz, Blu, and Unser [Thé00] and also Bankman [Ban00] report a factor of two to four in execution time.

2.5.2 Multilevel Discretization

A solution to the objective function in (2.1) cannot be given in a closed form, and we therefore solve the problem numerically. In the literature, this has been done in two ways. First, by formulating the necessary condition for a minimum analytically and constructing a fixed-point iteration based on this condition. This approach is called optimize-then-discretize [Fis04].

As it allows the use of standard Newton-type optimization methods, we will follow the other approach—the so-called discretize-then-optimize approach [Hab04]—where the problem is first discretized and then solved numerically.

We assume that Ω is rectangular and that all grid cells are quadratic with size $\bar{h} = h \cdot h$.

Following [Mod09, pp. 20 ff.], we discretize the domain using a cell-centered grid. In each dimension of the domain $\Omega = [\omega_{1,1},\ \omega_{1,2}] \times [\omega_{2,1},\ \omega_{2,2}] \subset \mathbb{R}^2$, $\omega_{i,j} \in \mathbb{R}$, we discretize the invervals $[\omega_{1,1},\ \omega_{1,2}]$ and $[\omega_{2,1},\ \omega_{2,2}]$ by vectors

$$\left(\omega_{d,1} + h \cdot \left(i + \frac{1}{2}\right)\right)_{i=1,\dots,m_d} \in \mathbb{R}^{m_d}, d \in \{1, 2\}$$

where h is the length of one quadratic grid cell and m_d is the number of cells in the respective dimension. A two-dimensional grid of size $m_1 \times m_2$ is constructed based on the one-dimensional vectors as shown in Figure 2.5.1.

The grid is a cell-centered grid, where each grid point is in the center of a cell. The displacement evaluated at the grid points in $\mathbf{x}$ will be denoted by $\mathbf{u} \in \mathbb{R}^{2 \cdot m_1 \cdot m_2}$. For convenience, we denote the pair of x- and y-coordinates of the grid by $\mathbf{x}_i := (\mathbf{v}_i, \mathbf{v}_{i+m_1 m_2}) \in \mathbb{R}^2$. This notation is used for $\mathbf{u}_i \in \mathbb{R}^2$ accordingly.

The reference image can be written as a vector that is obtained by evaluating $R(\mathbf{x}_i)$ at the grid points $\mathbf{x}_i,\ i = 1, \dots, m_1 m_2$. Since R is not evaluated at transformed grid points, $R(\mathbf{x}_i)$ is constant during the optimization. The template image T is an interpolation function that can be evaluated at arbitrary grid points. We sometimes write T_i as a short form for the template image evaluated at grid point i, $T(\mathbf{u}_i + \mathbf{x}_i)$.

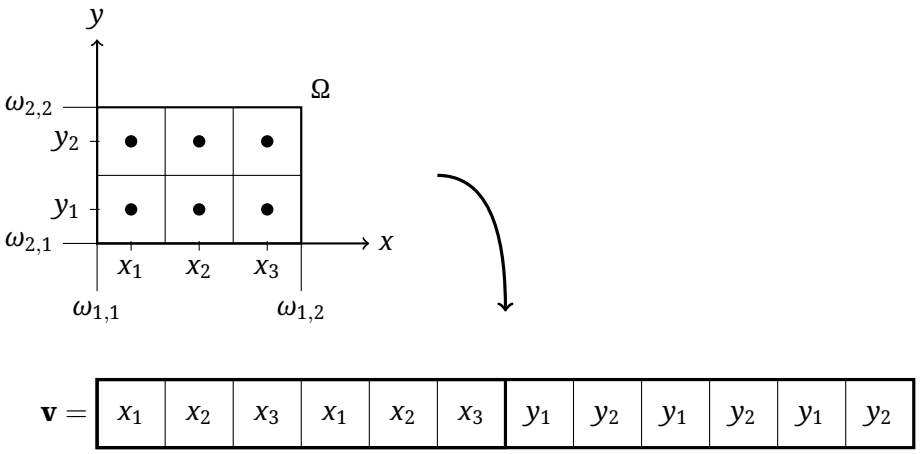


Figure 2.5.1: Arrangement of discretized grid points in the grid vector $\mathbf{x} \in \mathbb{R}^{2m_1 \cdot m_2}$ for $m_1 = 3$ and $m_2 = 2$. The grid coordinate (x_1, y_1) is stored in the components $\mathbf{x}_1 := (\mathbf{v}_1, \mathbf{v}_7)$.

Given the cell-centered grid, the integrals in $\mathcal{D}$ and $\mathcal{S}$ can be evaluated by the two-dimensional midpoint rule

$$\int_\Omega f(x)\,\mathrm{d}x = \sum_{i=1}^{N} f(\mathbf{x}_i) \cdot \bar{h} + \mathcal{O}(\bar{h}^2) \tag{2.4}$$

for any differentiable function $f \in \mathcal{C}^1$, $f : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ with $\mathbf{x}_i$ and $\bar{h} = h \cdot h$ and $N = m_1 \cdot m_2$.

The subsequent interpolation and discretization described above are used to obtain a representation of the images on an arbitrary grid. Based on these representations, we implement a coarse-to-fine multilevel-scheme following [Mod09, pp. 146 ff.].

A Multilevel Pyramid of Discretizations

A Newton-like method converges to a global optimum if the objective function J is convex [Noc06, pp. 8 ff.]. If J is only locally convex, convergence to the global solution can only be shown in its immediate proximity [Noc06, p. 52]. The objective function is often non-convex, and the iteration is likely to converge in a local minimum. To counteract these two issues, the registration problem is embedded in a multilevel-pyramid.

To obtain representations at different scales, the objective function is discretized using different grid spacings. Constructing a pyramid of image interpolation functions, the image data is smoothed and downsampled repeatedly. For each level of the pyramid, corresponding interpolation functions are constructed based on the downsampled image data to avoid data loss caused by under-sampling.

The registration is started at the coarsest resolution, and the result is used as an initial guess on the next finer level. Operations on smaller, low-resolution images require less memory and are faster to compute such that non-local deformations can be computed efficiently on these low-resolution image representations. The image resolution is increased subsequently to account for details in the images. For more details, see, e.g., [Mod09, pp. 40-42].

The low-resolution layers of the image pyramid are represented by a smaller amount of data such that coarse global image representations of large-scale images can be loaded into the RAM. We make use of these representations by combining a local high-resolution layer and a global low-resolution layer in the combined method proposed in Chapter 4.2.

An alternative to the multilevel pyramid of image representations is the multi-scale approach [Mod09, p. 145] where no downsampling but only smoothing is used to convexify the objective function. However, the multi-scale approach does not reduce the size of the input data, and its computation is, therefore, more expensive.

2.5.3 Distance Measures

Each component of the objective function is discretized on the grid shown in Figure 2.5.1. For the distance measures introduced in Section 2.2, discretized distances can be derived based on the grid $\mathbf{x}$ using the mid-point rule. Instead of calligraphic letters $\mathcal{D}$ and $\mathcal{S}$, we will use upright letters for the discretized functions D and S. We define the discrete SSD and NGF distance measures as

$$D_{\mathrm{SSD}}(R, T, \mathbf{u}) = h^2 \cdot \sum_{i=1}^{N} (T(\mathbf{x}_i + \mathbf{u}_i) - R(\mathbf{x}_i))^2$$

and

$$D_{\mathrm{NGF}}(R, T, \mathbf{u}) = h^2 \cdot \sum_{i=1}^{N} 1 - \left(\frac{\nabla T(\mathbf{x}_i + \mathbf{u}_i)^T \nabla R(\mathbf{x}_i) + \epsilon^2}{\|\nabla T(\mathbf{x}_i + \mathbf{u}_i)\|_\epsilon \, \|\nabla R(\mathbf{x}_i)\|_\epsilon} \right)^2,$$

where $\|x\|_\epsilon^2 = \|x\|_2^2 + \epsilon^2$ [Mod09, p. 92; Rüh13]. In NGF, an image gradient is included in the distance. To be able to capture oscillating image functions, we combine the scalar products (and norms) of the short finite differences to both sides of the grid point $\mathbf{x}_i$ as proposed in [Rüh13].

At the image boundary, the prevailing choices are Dirichlet or Neumann boundary conditions. A histological image usually consists of a tissue section on a bright background which is homogeneous to the human observer but which often contains noise and sometimes artifacts. In most cases, the tissue of interest is not located close to the boundary. One can either convert the image such that the background becomes black (gray value = 0) or use the image as is. In the first case, Dirichlet boundary conditions can be used. However, in order to do so, the background has to be modified using some segmentation, which introduces an additional challenge for the registration pipeline. We, therefore, leave the images unaltered and use Neumann conditions at

their boundaries. One drawback of this approach is that artifacts close to the image boundary can distort the registration result at these positions.

The above discretizations can both be written in the form

$$D(R, T, \mathbf{u}) = \sum_{i=1}^{N} \psi(r_i(R, T, \mathbf{u}_i)),$$

where each value $\psi(r_i)$ corresponds to the local image distance at one grid point and can—given its immediate neighborhood—be computed independently of the rest of the image. Therefore, this formulation is easily computed in parallel and also well suited for the localized approach presented in this thesis, where only parts of the image are evaluated at its full resolution.

2.5.4 Regularizers

Following [Mod09], the discretized regularizer can be written in matrix notation. In the two-dimensional case, Kronecker products are used to construct two-dimensional operators from their one-dimensional counterparts.

While the Kronecker product allows a convenient notation that facilitates the formulation of the operator structure, the operators are not actually constructed in this way during the computation. Matrix-free algorithms can be built by analyzing the matrix structure and converting the necessary operations into a cheaper algorithmic description. See [Kön18] for more details on matrix-free implementations in image registration.

We use the midpoint rule to discretize the regularizer from Section 2.3 by

$$\mathcal{S}[u] \approx S(\mathbf{u}) = \alpha\, \mathbf{u} B^T B \mathbf{u} \tag{2.5}$$

where B is a discretized differential operator, that is specific to the regularizer of choice.

Diffusive Regularizer

For diffusive regularization, the nabla operator ∇ can be discretized by finite differences in one dimension with

$$B_m^{\nabla} = \frac{1}{h} \begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix} \in \mathbb{R}^{(m-1)\times m}$$

where m is the number of cells in the respective dimension. Multiplication with B_m^{∇} means that in each row, a short finite difference is computed as the difference between two neighboring grid points divided by h.

We note that in our formulation, the operator B_m^∇ is not quadratic and implicitly changes the grid at which the derivative is approximated. The new grid points are between the original cell centers, producing a staggered grid. When multiplying with the transpose $\left(B_m^\nabla\right)^T$, the grid is mapped back to the original points. However, the combined operator $\left(B_m^\nabla\right)^T B_m^\nabla$ has only rank $m-1$. This can be compensated by adding a boundary condition, usually done by modifying the element $(1,1)$ and $(m-1,m)$. We will keep constant deformations in the null-space of the operator and rely on the distance measure to form a full-rank joint operator.

In two dimensions, the Kronecker product can be used to define the operator

$$B_{m_1\times m_2}^\nabla = \begin{pmatrix} I_{m_2}\otimes B_{m_1}^\nabla & \\ B_{m_2}^\nabla \otimes I_{m_1} & \\ & I_{m_2}\otimes B_{m_1}^\nabla \\ & B_{m_2}^\nabla \otimes I_{m_1} \end{pmatrix}$$

for an equally spaced grid of size $m_1 \times m_2$ which is organized as in Figure (2.5.1). In the upper-left block, $B_{m_1\times m_2}^\nabla$ acts on the first component of u—the displacement in the x-direction. In the two rows, the first two partial derivatives are computed. In the lower-right block, the second component is computed likewise.

Curvature Regularizer

Curvature regularization is based on second-order derivatives. We again avoid explicitly setting boundary conditions by applying the one-dimensional operator B^∇ twice. Starting with the one-dimensional case,

$$B_m^\Delta = B_{m-1}^\nabla B_m^\nabla \in \mathbb{R}^{(m-2)\times m}$$

is a matrix with stencil $\frac{1}{h^2}(1,\ -2,\ 1)$ in each line. In contrast to the first-order operator B_m^∇, the new grid points are again on the same cell-centered grid. However, at the boundary of Ω, at grid points $i=1$ and $i=m$, the second derivative cannot be computed, since the outer neighbors are missing. In [Mod09, pp. 130], Dirichlet boundary conditions are proposed. In the numerical examples in this thesis, we will proceed in analogy to the diffusive operator with an operator $\left(B_m^\Delta\right)^T B_m^\Delta$ of rank $(m-2)$, where both constant and linear functions are in the null-space. Therefore, an affine preregistration does not affect the value of the regularizer, and no explicit treatment of the affine components of the deformations is necessary.

Again, we use a Kronecker block matrix to construct the two-dimensional operator

$$B_{m_1\times m_2}^\Delta = \begin{pmatrix} I_{m_2}\otimes B_{m_1}^\Delta & \\ & B_{m_2}^\Delta \otimes I_{m_1} \end{pmatrix}$$

where only the two diagonal blocks of the two-by-two block matrix are non-zero.

2.6 Optimization

We understand image registration as an optimization problem where the objective function consists of the building blocks *distance measure* and *regularizer* that we discussed in this chapter. Since the solution to the optimization problem cannot be formulated explicitly, a numerical optimization method is needed. We follow the book by Nocedal and Wright [Noc06] where these and other methods for numerical optimization are covered in greater detail.

We formulate a discretized version of the original image registration problem in (2.1) as

$$J(R, T, \mathbf{u}) = D(R, T, \mathbf{u}) + S(\mathbf{u}) \xrightarrow{\mathbf{u}} \min$$

and we will denote the objective function simply by $J(\mathbf{u}) = J(R, T, \mathbf{u})$ throughout this section to simplify notation.

2.6.1 (Inexact) Gauß–Newton Method

In general terms, if $\mathbf{u}^*$ is the minimizer of an objective function $J(\mathbf{u}) : \mathbb{R}^{2m_1 \cdot m_2} \to \mathbb{R}$, the Jacobian $\nabla J(\mathbf{u}^*) = 0$ is zero and the Hessian $\nabla^2 J(\mathbf{u}^*)$ is positive semi-definite [Noc06, p. 14]. Iterative optimization methods aim to find the minimum by computing multiple updates $\mathbf{s}$ based o an initial guess $\mathbf{u}^{(0)}$. A simple method to compute an update $\mathbf{s}$ is gradient descent, where

$$\mathbf{s} = -\nabla J(\mathbf{u}),$$

is the negative gradient of J. While this update is easy to compute, the main disadvantage of the gradient descent method is its slow convergence even if the objective function is well conditioned [Noc06, pp. 42-44]. An alternative that uses the information from the Hessian and has better convergence properties is the Newton method [Noc06, pp. 23].

The Newton method can be derived by a linearization through the Taylor expansion of $\nabla J(\mathbf{u})$

$$\nabla J(\mathbf{u} + \mathbf{s}) = \nabla J(\mathbf{u}) + H(\mathbf{u})\,\mathbf{s} + \mathcal{O}(\|\mathbf{s}\|_2^2) \tag{2.6}$$

where $H(\mathbf{u}) = \nabla^2 J(\mathbf{u})$ is the Hessian of $J(\mathbf{u})$. With the necessary condition for the minimizer at the updated displacement $\nabla J(\mathbf{u} + \mathbf{s}) \stackrel{!}{=} 0$ we obtain a linear system

$$H(\mathbf{u})\,\mathbf{s} = -\nabla J(\mathbf{u}) \tag{2.7}$$

whose solution is the update $\mathbf{s}$. Since the linearization in (2.6) approximates J locally, the updated candidate $\mathbf{u} + \mathbf{s}$ will not be the minimizer and the computation of (2.7) needs to be repeated with the update $\mathbf{u} \leftarrow \mathbf{u} + \mathbf{s}$ until convergence.

The above iteration has two drawbacks. First, the direct solution to the linear system becomes expensive if the number of unknowns is large. The number of operations

required to solve the linear system using a matrix factorization method is in the order of $\mathcal{O}(N^3)$ operations [Den96, p. 51] where $N = 2m_1m_2$ is the number of unknowns in the system. The factorized system is equivalent to the original linear system, and the error of the computed solution is entirely caused by the use of finite precision arithmetic. Approximations to the solution can be computed at lower computational cost by using iterative algorithms such as the multigrid [Tro01; Hen00] or conjugate gradient (CG) [Noc06, pp. 102 ff.] methods.

The system in (2.7) is the result of approximations on the objective function such that its solution can also be only an approximation. Instead of computing an expensive accurate solution to the linear system based on a matrix factorization, the solution to the linear system is approximated using the CG method as proposed in [Mod09, p. 134] for curvature regularization. The resulting method is referred to as the inexact Newton method [Dem82].

The second drawback of the Newton method is its requirement for a convex objective function, which in turn requires a positive definite Hessian $H(\mathbf{u})$ at each iteration. Due to the arbitrary nature of the images, this cannot be guaranteed globally. Different schemes have been proposed to modify the Hessian such that positive definiteness can be guaranteed, see, e.g., [Den96, p. 101] where a scaled identity matrix $I \in \mathbb{R}^{N\times N}$ is added to the Hessian.

Another approach that guarantees positive semi-definiteness without introducing an additional parameter is the Gauß–Newton approximation (see [Den96, pp. 221 ff.]) where the Hessian is approximated by a quadratic term $H(\mathbf{u}) = \nabla J(\mathbf{u})^T \nabla J(\mathbf{u})$ based on the gradient of the objective function.

In the image registration framework described in this chapter, this approximation does only apply to the distance measure, which can be highly nonlinear depending on the nature of the image data. These strong nonlinearities are ignored in the Gauß–Newton approximation, resulting in a positive semi-definite Hessian approximation at the cost of possibly slower convergence [Den96, p. 224]. The regularizer is already written in a quadratic form and is guaranteed to be positive semi-definite.

Following [Mod09], we combine the Gauß–Newton approximation with the inexact solution to the linear system and refer to the resulting method as the inexact Gauß–Newton method which is described in Algorithm 2.6.1.

Algorithm 2.6.1 Inexact Gauß–Newton Method

```
1 for j=1, 2, ...:
2     # solve for s up to a predefined accuracy using an
          iterative linear solver
3     H(u^(j))s = −∇J(u^(j))
4     # update with step length β ∈ ℝ computed by an Armijo line
          search method [Noc06, pp. 31 ff.]
5     u^(j+1) = u^(j) + βs
```

Stopping criteria that determine if a local minimum has been reached can be found in more detail in [Gil97, p. 306] and are based on

- change of the objective function: $J(\mathbf{u}) - J(\mathbf{u}+\mathbf{s}) < \theta_J$
- size of the norm of the update $\|\mathbf{s}\|_2 < \sqrt{\tau}\,(1 + \|\mathbf{u}\|_2)$
- size of the norm of the gradient $\|\nabla J(\mathbf{u}+\mathbf{s})\| < \sqrt[3]{\tau}\,(1 + |J(\mathbf{u}+\mathbf{s})|)$.

If all the above quantities are smaller than predefined thresholds, the iteration is terminated.

In the Gauß–Newton algorithm, an approximation of the Hessian of the distance measure has to be implemented. An alternative that approximates the inverse of the Hessian based on first-order derivatives, which are computed during the iteration, is the L-BFGS method.

2.6.2 L-BFGS Method

The Gauß–Newton method requires solving a linear system in each update step. The BFGS method (named after Broyden, Fletcher, Goldfarb, and Shanno) avoids solving the linear system by directly approximating the inverse of the Hessian using first-order derivatives only. In consequence, the computation of each update is computationally cheaper, but often more iterations are needed until convergence [Noc06, p. 141, pp. 178–180].

The second-order derivative of the objective function is estimated using the differences in the gradients between consecutive iterations [Noc06, pp. 136–141]. Two rank-one updates are computed in each iteration, accumulating to the inverse Hessian approximation.

Instead of storing all these updates, the L-BFGS (limited memory BFGS) method only stores a fixed number of previous updates to save memory. For the experiments with real histological whole slide images in Chapter 7, we use the two-loop algorithm described in [Noc06, p. 178] and use the exact Hessian of the regularizer as initial guess for the approximation.

Both, the Gauß–Newton and the L-BFGS approximation of the Hessian, are implemented as a combination of first-order terms from the image distance with the analytic second-order derivative of the regularizer. L-BFGS uses rank-1 updates to build the approximation over multiple iterations. A benefit when using L-BFGS is that the approximated Hessian does not need to be implemented. However, especially in the first steps of the optimization, the Hessian approximation is of low rank, possibly increasing the total number of iterations. We are not aware of a systematic comparison of both methods for image registration. We choose the Gauß–Newton method for experiments on small-scale images and resort to the L-BFGS method when registering large-scale images to simplify the implementation. A further comparison of both methods is left for future work.

2.7 Derivatives of Distance Measures and Regularizer

We further follow the book by Modersitzki [Mod09] and compute the gradient and the Hessian of the objective function as follows.

Derivative of the Discretized Regularizer

Starting with the regularizers, we can take advantage of the operator formulation introduced earlier in (2.5). The derivatives of both regularizers—diffusive and curvature— can be written in the form

$$\begin{aligned} S(\mathbf{u}) &= \frac{\alpha}{2}\|B\mathbf{u}\|_2^2 = \frac{\alpha}{2}\,\mathbf{u}^T B^T B\mathbf{u}, \\ \nabla S(\mathbf{u}) &= \alpha\, B^T B\mathbf{u} \quad \text{and} \\ \nabla^2 S(\mathbf{u}) &= \alpha\, B^T B \end{aligned}$$

where B is the discretized differential operator associated with the regularizer.

Derivative of the Discretized Distance Measures

By applying the chain rule, the first-order derivative of the SSD distance measures can be written as

$$\begin{aligned} D_{\text{SSD}}(R,T,\mathbf{u}) &= \frac{h^2}{2}\cdot\sum_{i=1}^{M}(T(\mathbf{x}_i+\mathbf{u}_i)-R(\mathbf{x}_i))^2 \\ \frac{\partial D_{\text{SSD}}(R,T,\mathbf{u})}{\partial \mathbf{u}_i} &= h^2(T(\mathbf{x}_i+\mathbf{u}_i)-R(\mathbf{x}_i))\cdot\frac{\partial}{\partial \mathbf{u}_i}T(\mathbf{x}_i+\mathbf{u}_i) \end{aligned}$$

where $M = m_1 \cdot m_2$ and $\frac{\partial D_{\text{SSD}}}{\partial \mathbf{u}_i} \in \mathbb{R}^2$ has one component for each dimension. In order to match the structure of the grid, the gradient ∇D_{SSD} with respect to $\mathbf{u}$ can be written using two components: First, there is the template image derivative

$$\begin{aligned} \nabla T(\mathbf{x}+\mathbf{u}) &= \mathrm{d}T(\mathbf{x}+\mathbf{u}) \\ &= \begin{pmatrix} \mathrm{d}_x T_1 & & & \mathrm{d}_y T_1 & & \\ & \ddots & & & \ddots & \\ & & d_x T_N & & & \mathrm{d}_y T_N \end{pmatrix} \in \mathbb{R}^{N\times 2N} \end{aligned} \tag{2.8}$$

where $(\mathrm{d}_x T_i, \mathrm{d}_y T_i) := \frac{\partial T(\mathbf{x}_i+\mathbf{u}_i)}{\partial \mathbf{u}_i}$ are the two components of the partial derivative with respect to $\mathbf{u}_i$. The second component is the vector of the values of $T(\mathbf{x}_i+\mathbf{u}_i)$ evaluated at the grid points. The product of these two components is the gradient of the distance measure such that

$$\nabla D_{\text{SSD}}(R,T,\mathbf{u}) = h^2\cdot \mathrm{d}T(\mathbf{x}+\mathbf{u})^T \begin{pmatrix} T(\mathbf{x}_1+\mathbf{u}_1)-R(\mathbf{x}_1) \\ \vdots \\ T(\mathbf{x}_N+\mathbf{u}_N)-R(\mathbf{x}_N) \end{pmatrix}.$$

For the Hessian, the Gauß–Newton approximation is used, where the second-order derivatives $\nabla^2 T(\mathbf{x}+\mathbf{u})$ are ignored and the Hessian is approximated by the outer product

$$H_{\mathrm{SSD}}(R,T,\mathbf{u}) = h^2 \cdot \mathrm{d}T(\mathbf{x}+\mathbf{u})^T \mathrm{d}T(\mathbf{x}+\mathbf{u})$$

with $\mathrm{d}T(\mathbf{x}+\mathbf{u})$ as in (2.8) above.

For the NGF distance measure, the additional gradients of T and R have to be considered. We define

$$\mathrm{ng}(\mathbf{u}_i) = \frac{g^{(1)}(\mathbf{u}_i)}{g^{(2)}(\mathbf{u}_i)} = \frac{\nabla T(\mathbf{x}_i+\mathbf{u}_i)^T \nabla R(\mathbf{x}_i) + \epsilon^2}{\|\nabla T(\mathbf{x}_i+\mathbf{u}_i)\|_\epsilon \, \|\nabla R(\mathbf{x}_i)\|_\epsilon}$$

as the local normalized gradient at $\mathbf{x}_i+\mathbf{u}_i$, where ∇T and ∇R are approximated by short finite differences. Based on the above notation, the discretized NGF at $\mathbf{x}_i+\mathbf{u}_i$ can be written as

$$\begin{aligned} D_{\mathrm{NGF}}(R,T,\mathbf{u}) &= \frac{h^2}{2} \cdot \sum_{j=1}^{M} 1 - \mathrm{ng}(\mathbf{u}_j)^2 \\ \frac{\partial D_{\mathrm{NGF}}(R,T,\mathbf{u})}{\partial \mathbf{u}_i} &= \sum_{j=1}^{M} -h^2 \mathrm{ng}(\mathbf{u}_j) \\ &\qquad \cdot \left(\frac{\partial g^{(1)}(\mathbf{u}_j)}{\partial \mathbf{u}_i} \cdot \frac{1}{g^{(2)}(\mathbf{u}_j)} + g^{(1)}(\mathbf{u}_j) \cdot \frac{\partial (g^{(2)}(\mathbf{u}_j)^{-1})}{\partial \mathbf{u}_i} \right) \end{aligned} \tag{2.9}$$

The derivatives of the numerator and the denominator of $\mathrm{ng}(\mathbf{u}_i)$ are

$$\begin{aligned} \frac{\partial g^{(1)}(\mathbf{u}_i)}{\partial \mathbf{u}_i} &= (4r_i - \sum_{k\in\{\pm 1,\pm m\}} R(\mathbf{x}_{i+k})) \frac{\partial T_i}{\partial \mathbf{u}_i} & \\ \frac{\partial g^{(1)}(\mathbf{u}_j)}{\partial \mathbf{u}_i} &= -(R(\mathbf{x}_i) - R(\mathbf{x}_j)) \frac{\partial T_j}{\partial \mathbf{u}_i} & \text{if } |i-j| = 1 \\ & & \text{or } |i-j| = m_1 \\ \frac{\partial g^{(1)}(\mathbf{u}_j)}{\partial \mathbf{u}_i} &= 0 & \text{otherwise} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial g^{(2)}(\mathbf{u}_i)}{\partial \mathbf{u}_i} &= -\frac{(4T_i - \sum_{k\in\{\pm 1,\pm m\}} T_{i+k})}{2\,\|\nabla T(\mathbf{u}_i)\|_\epsilon^3 \|\nabla R(\mathbf{x}_i)\|_\epsilon} \frac{\partial T_i}{\partial \mathbf{u}_i} & \\ \frac{\partial g^{(2)}(\mathbf{u}_j)}{\partial \mathbf{u}_i} &= -\frac{(T_i - T_j)}{2\,\|\nabla T(\mathbf{u}_j)\|_\epsilon^3 \|\nabla R(\mathbf{x}_i)\|_\epsilon} \frac{\partial T_i}{\partial \mathbf{u}_i} & \text{if } |i-j| = 1 \\ & & \text{or } |i-j| = m_1 \\ \frac{\partial g^{(2)}(\mathbf{u}_j)}{\partial \mathbf{u}_i} &= 0 & \text{otherwise.} \end{aligned}$$

The Hessian approximation $H_{\mathrm{NGF}} \in \mathbb{R}^{2N\times 2N}$ is constructed again using the Gauß–Newton approximation and neglecting second-order terms.

2.8 Properties of the Regularization Operator and its Inverse

The solution to the linear system in line 3 in the Gauß–Newton Algorithm 2.6.1 leads to a coupling of the registration problem that we need to take into account when decomposing the registration problem in local subproblems.

As discussed in Chapter 2.3, the regularizer is a measure for the smoothness of the deformation. It consists of a differential operator that acts on the displacement, and that is integrated over the image domain [Fis01; Fis03b; Bur13]. After discretization, the linear system

$$-H(\mathbf{u})\,\mathbf{s} = \nabla J(\mathbf{u}) \tag{2.10}$$

is solved to compute an update step $\mathbf{s}$ where (in the case of SSD distance measure)

$$H(\mathbf{u}) = \alpha B^T B + h^2 \left(dT(\mathbf{x}+\mathbf{u})\right)^T \left(dT(\mathbf{x}+\mathbf{u})\right)$$

as discussed in Chapter 2.6.1 (Equation 2.7). The Hessian H is a band matrix. In the case of the curvature regularizer, it has five non-zero diagonals in the one-dimensional case, and we assume that the data term is made such that H is symmetric and positive definite. Solving for the update step $\mathbf{s}$ in Equation 2.10 smoothes the information from the right-hand side $\nabla J(\mathbf{u})$ across the domain. The smoothing kernel has global support, which can be verified by examining the Cholesky factorization [Gol96, p. 144] of H and performing forward and backward substitution. The inverse H^{-1} is visualized in Figure 2.8.1.

The smoothing is applied in each iteration of the Gauß–Newton method, and the information from the alignment of high-contrast regions is hence distributed into low-contrast regions.

A drawback of this approach is that information from the entire image domain is needed in each iteration. Depending on α, the bandwidth of the inverse of $H(\mathbf{u})$ varies, and further study of the operator could result in ways to benefit from the distribution of information without requiring the solution to the global system. We leave this thought for future work and will instead use a reduced representation of the deformation when partitioning the registration into local subproblems in Chapter 4.2.

2.9 Memory Requirements of an Implementation

When implementing a registration algorithm based on the framework described above, some components have to be stored in memory. We give an estimate of the storage requirements of the most essential of these components in Table 2.1,

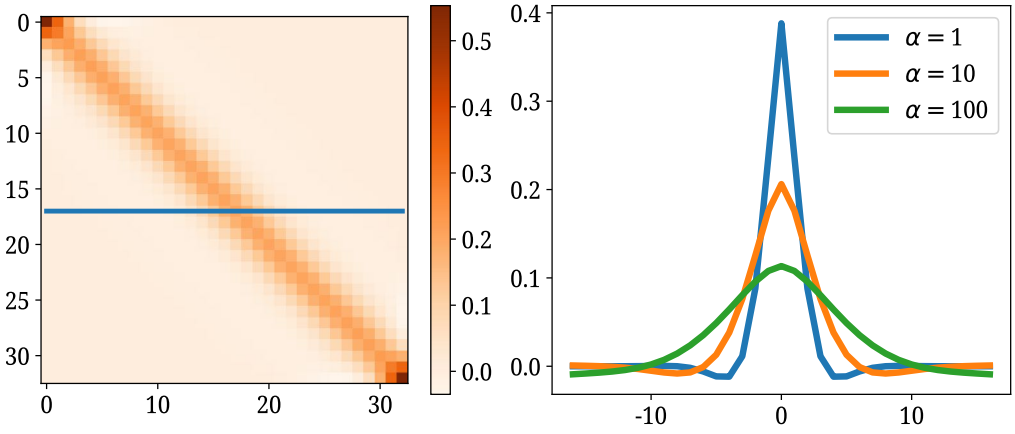


Figure 2.8.1: Left: Inverse H^{-1} of the Hessian in a one-dimensional example (m = 33) using an identity matrix as derivative of the data term $h^2\,(dT(\mathbf{x}+\mathbf{u}))^T(dT(\mathbf{x}+\mathbf{u})) = I$. Right: Center row of H^{-1} for different values of α.

Table 2.1: Memory requirements of the major components of a single level registration with $N = m_1 \cdot m_2$ pixels per image. For the computation of the distance measure, we store the derivative of the image ∇T in a vector of size $2N$, which contains the derivative in the first and second dimensions. We also store the non-zero partial derivatives in the NGF computation (∇D, Equation 2.9) for faster computation. For each pixel in the reference image, five values have to be stored to compute the short finite differences, resulting in $5N$ values.

Component		Memory required (variables)
Distance Measure	R, T	N each
	∇T	$2N$
	∇D	$2N$ (SSD)
		$5N$ (NGF)
Regularizer	$B^TB\mathbf{u}$	$2N$
Optimization	$\mathbf{u}$	$2N$
	$\mathbf{s}$	$2N$
	CG method	$4N$ [Saa03, p. 200]
	L-BFGS buffer	$5 \cdot 2N$ (for buffer size = 5)
Total		$\approx 18N$ (NGF)
		$\approx 15N$ (SSD)

noting that the list is not minimal. Depending on the implementation, some of these components can be re-computed at each occurrence to save memory. The only exception is the image data, which would need to be re-loaded from disk, delaying the algorithm substantially.

Table 2.1 shows the memory requirements for the essential components of a single level registration implementing the discretize-then-optimize framework. In summary, the required memory grows linear with the number of pixels and quadratically with the image resolution in terms of pixels per milimeter.

Embedding the registration in a multilevel scheme as in Chapter 2.5.2 does not increase the memory requirements as each level can be computed separately. The image pyramid itself can be computed before beginning the registration. The only result that needs to be stored after each level is the resulting transformation $\mathbf{u}$, which is used as the initial guess of the registration on the subsequent higher level.

The registration at each level of the pyramid needs a fourth of the memory of the subsequent higher level. In each step, the number of pixels is divided by two in each dimension. The upper bound for the memory that is needed to store the complete image pyramid is $\frac{N}{1-\frac{1}{4}} = \frac{4}{3}N$.

The absolute amount of memory depends on the data type of the stored values. Assuming double precision (64 bits per value), a registration of two $10\,000 \times 10\,000$ pixel images requires about 23 GB of memory. This number is a lower bound for an implementation that stores the components listed in Table 2.1. For two images of size $100\,000 \times 100\,000$ pixels as they occur in practice, the required memory exceeds 2.3 TB.

The variational image registration framework presented above has a solid theoretical background, and the presented implementation is backed by state-of-the-art optimization. It has shown good results in other areas of medical imaging. The memory requirements in Table 2.1 show that the computation becomes infeasible for large images from digital pathology on the reference computer that this thesis is written on.

The main factor for the amount of required memory is the size of the image data that needs to be stored independently of the registration algorithm. Therefore, the problem of registering large images as it is addressed in this thesis goes beyond the particular implementation of the registration but is inherent to the registration problem itself.

3 State of the Art in Large-Scale Image Registration with Focus on Pathology

Contents

In Chapter 1.3, we established criteria for a registration algorithm such that it can be used to register large images on off-the-shelf workstations. The author is not aware of an existing solution that fulfills all the criteria, but certain aspects are discussed in the literature. We will discuss these approaches in this chapter.

First, we address work in the area of the registration of histological serial sections, starting with early methods and focusing on the challenge posed by large images. A few methods have been proposed to use image registration in high-performance computing using domain decomposition.

Our new approach takes up some aspects from the area of domain decomposition. To show these similarities, we discuss domain decomposition techniques for linear problems and their adaptions to nonlinear problems in the second part of this chapter. In particular, the coupling between subdomains is a fundamental idea that we discuss.

We demonstrate the amount of communication needed by these methods in numerical experiments based on a simple image registration problem.

A general overview of the field of image registration in medical imaging can be found in the general surveys in [Sot13; Oli14; Kes17]. An overview of image registration of digital pathology images is given in [Pic18] with a focus on 3D reconstruction of tissue stacks.

3.1 Image Registration in Digital Pathology

A large part of the work in the field of digital pathology for image registration focuses on 3D reconstruction. In most publications, reconstruction is performed by multiple image registrations of consecutive histologic slides using low-resolution images. The registration of slide pairs in order to combine their information [Coo07; Mue11] is reported less often. However, this area gained popularity in the last years in combination with automatic image analysis of whole slide images [Mad16; Har18].

Advanced imaging technology results in higher amounts of data and in a shift of reconstruction tasks. One example is the growing interest in the reconstruction of global or functional entities such as micro-vasculature [Yin16] or immunohistochemical markers [Rob12]. This is likely a result of the emerging field of digital pathology where computers are used to visualize, store, and analyze digitized whole slide images. In these applications, the challenge is to reconstruct and fuse the data on a completely different level: smaller structures like individual cells and the comparison of different functional markers across slides are of increasing interest, resulting in the need of reconstructions ideally on the scale of cell nuclei. Due to the local deformations introduced in the tissue preparation, affine or rigid solutions are not sufficient for this task. Nonlinear deformations that are introduced during the cutting and staining process have to be corrected to achieve satisfying results.

Even though the technology to compute 3D reconstructions has advanced significantly, the high amount of histological data cannot be handled with the established methods on a single computer since these machines are limited in RAM. This even holds true for the data needed to fulfill only a part of the 3D reconstruction task, the registration of two successive slides. There are different approaches to address the challenge of large image dimensions.

In the following, we distinguish between global methods that compute a solution based on extracted features or subsets of the data on the one hand and those methods that implement a divide and conquer approach on the other.

3.1.1 Feature Reduction

Schwier et al. [Sch13] reduce the image data by segmenting vessel structures and use these structures to steer their two-step approach. First, a rigid, iterative best-fit matching of the segmented vessel structures is calculated, which is refined by an elastic registration step on a low-resolution image. The resulting deformation is then applied to the original slide data. Working with low-resolution images while computing the 3D reconstruction reduces the problem size while the resulting deformation can still be applied to high-resolution images. However, structures that are smaller than the coarse image resolution cannot be aligned accurately by a low-resolution approach such that its result is less accurate.

By matching features from the scale-invariant feature transform (SIFT), Cardona et al. [Car10] register images from transmission electron microscopy (TEM). Sub-images

captured from TEM are stitched in plane while they are reconstructed in 3D at the same time. Using SIFT point correspondences as a distance measure, the authors of [Sch06] combine rigid and nonlinear deformation components by a moving least squares approach. While the size of the registration problem can be reduced effectively by identifying relevant corresponding points in both images, the difficulty in such an approach is the robust identification of those critical points.

Weiss, Lotz, and Modersitzki [Wei15] use a global detection of cell nuclei to reduce the tissue data to a density map that can be stored efficiently due to the sparsity of the nuclei. These densities are independent of a particular staining and are used to compute a global deformation of whole slide images. The difficulty here is again in the robust identification of the nuclei in different stainings.

In the following, we limit our focus on methods that are based on image intensity as opposed to feature-based methods. In such methods, a large part of the registration problem is in the identification and matching of corresponding features, which is beyond the scope of this work. We refer the reader to [Pic18, Section 4.1], where further feature-based methods are discussed.

3.1.2 Divide and Conquer Methods

Instead of reducing the image data to selected features and possibly losing information in the process, local methods divide the image into smaller parts and process these parts independently. One advantage of such an approach is that well-established registration methods can be used locally. The main concern is to combine the individually computed results into one final deformation eventually.

Arsigny, Pennec, and Ayache [Ars05] follow the idea of affinely transforming selected image subdomains. The authors compute a global, poly-affine registration by combining multiple affine transformations while maintaining smoothness at the tile borders. Ehrhardt et al. [Ehr19] extend the method to compensate for motion artifacts in the reconstruction of magnetic particle imaging. However, because of the simultaneous computation of multiple regions, the poly-affine method is not meant to work with large images. Cooper et al. [Coo07] combine this approach with a local subdomain-based refinement. Using non-rectangular regions, Pitiot et al. [Pit06] propose a registration framework, where automatically segmented partitions of the images are generated based on tissue structure such as the gyri of the brain. The regions are transformed independently by an affine registration. A global transformation is then found by interpolating the transformation between the registered partitions.

A recent approach using subdomain-based registration to cope with large datasets in histology is the work of Song et al. [Son14] on three-dimensional tissue reconstruction of histological sections that are differently stained. The authors propose a tile-based approach previously published by Roberts et al. [Rob12] that first computes a rough globally rigid transformation, which is then refined by calculating rigid transformations on smaller subdomains of the image with higher resolution. Multi-modal registration between differently stained slides is achieved by an automated content

classification. A global nonlinear deformation is computed by interpolating between rigidly transformed points on individual subdomains using B-spline transformations.

In breast cancer research, image registration has been used to combine different immunohistochemical markers by registering adjacent slides. Andersen et al. [And17] compared a registration-based approach to a traditional physical multistaining approach. To compute the H-Score [Ham10] that is used in grading, human breast cancer cells expressing an estrogen receptor (ER) have to be counted inside a tumor region. One of two slides was stained for ER-positive cells and was combined with the adjacent slide that was stained with an epithelium marker (pan-cytokeratin, PCK). The combined image is used to automatically count only those ER-positive cells that were located in a tumor area. The procedure was repeated with a physically double-stained slide with the same stains. A malignancy score was computed by a machine learning algorithm on the registered slides and manually by visual inspection on the physically double-stained slides. The authors report a perfect correlation based on 45 tissue cores. The analyzed tissue data consists of small circle-shaped tissue cores that are assembled in one slide, a so-called tissue microarrays (TMAs). The registration was done on a per-core basis, such that the size of the registration problem is only moderate, and no specific strategy for large-scale data is necessary.

A similar comparison with different stains has been made by Røge et al. [Røg16]. Both works show the benefit of an accurate registration to automate histological analysis.

In pharmaceutical research, data mining is used to identify patterns that are correlated to a clinical outcome. When using tissue images as a data source, the combination of different biomarkers increases the feature space in which correlations can be discovered. Harder et al. [Har18] use registration of adjacent tissue slides in order to investigate biomarker combinations together with additional data such as sex and age. For the prediction of tumor progression for some prostate cancer patients, the authors report that the discovered combination of features is superior to an established scoring mechanism. The registration is performed on hierarchically defined subdomains, where for each subdomain, a locally rigid transformation is computed and assigned to its center. Transformation landmarks are interpolated between the subdomain centers to obtain a global deformation.

In the studies described above [Pit06; Rob12; Son14; Mag15; Røg16; And17; Har18], the authors employ a divide-and-conquer approach to handle large images. However, none of these methods uses a global, physically motivated deformation model. Furthermore, in many of the methods, the subdomains are computed independently, such that the area surrounding the subdomain cannot contribute and stabilize the registration process.

3.1.3 Locally Nonlinear Deformation Models

Opposed to the approaches mentioned above, nonlinear transformation models offer a global, physically motivated transformation. We used a zooming strategy [Lot14; Pap08] to compute a high-resolution registration of a successively decreasing image

area. We were able to demonstrate a benefit of nonlinear registration over an affine registration on a 2000 x 2000 pixel image region.

For a registration of the entire image domain, we propose a subdomain-based registration framework [Lot16a], which consists of two parts. First, a global, coarse, nonlinear registration is computed on low-resolution data. In the second step, this preliminary deformation is used as the kernel in another registration on overlapping subdomains. On each subdomain, the registration problem is solved independently with the coarse registration as an initial guess. The deformations on the subdomains are combined into a global deformation field by interpolation. We note that, while our method computes a local nonlinear registration on all subdomains, the resulting deformation on the whole image is not global since the regularizer is not evaluated globally.

3.1.4 Domain Decomposition on Multi-Node Clusters

Computers contain a memory hierarchy, where faster access speeds are limited to relatively small amounts of memory [Mey03, pp. 1–4], since larger memory banks need more space and hence the distance to the processor increases. Exploiting the different layers of the memory hierarchy for faster algorithms is an active area of research, especially in the light of distributed computer systems, see [Mar18; Tho18; Mit15] and the references therein.

In the case of large-scale image registration, we are dealing with a situation where images exceed locally available memory. We will thus focus on distributed memory systems that gained popularity in recent years, partly driven by increased use of GPUs [Sha10].

Distributed memory (DM) domain decomposition is realized on clusters of computing nodes where each node is equipped with its own processor and memory. The nodes are connected such that information can be exchanged between them. The available memory and the number of CPUs in those systems can be scaled by adding additional nodes to the cluster. A large dataset can be partitioned between these nodes to avoid the memory limitation of a single workstation. However, communication between the nodes is expensive and has to be explicitly implemented on distributed memory clusters [Sha10].

While this work focusses on off-the-shelf workstation computers with limited RAM, we briefly discuss the general idea of domain decomposition and how it has been applied to image registration. In addition, domain decomposition methods are described in greater detail in Chapter 3.2 and highlight those aspects that are the basis for the new method proposed in Chapter 4.

Schmitt et al. [Sch07] use pairwise elastic image registration for the 3D reconstruction of a human brain and a rat brain based on serial sections and block-face images. The registration is parallelized by using a parallel conjugate gradient (CG) solver on a distributed memory cluster system [Mod99] of 48 computing nodes where each node stores a stripe of the image.

Using a domain decomposition approach, [Amu13; Moh16] compute a 3D reconstruction of the human brain based on 7676 serial section, block-face images and MRI. The computation of the 3D reconstruction is done on a high-performance computer consisting of 1872 computing nodes. The nodes process the image data in parallel such that each node only loads a small part of the whole image data set.

Another parallel approach with an emphasis on an efficient memory-parallel implementation is proposed by Mang, Gholami, and Biros [Man16]. The authors aim for a diffeomorphic deformation, which is computed by composing the deformation of multiple time steps and enforcing a non-zero determinant across all grid cells in each time step. The objective function is minimized using a Gauß–Newton approach [Noc06]. The image data is partitioned spatially, and deformation data is communicated across nodes at each Newton step. By implementing their approach on a DM cluster consisting of 64 nodes, a registration problem with an image size of 512^3 voxels is solved in 33 seconds.

As the above publications show, parallel computing is well suited to process large, prominent datasets as part of long-term research projects. The main downside of these methods is that they require specialized and expensive hardware, which is unavailable in many situations.

Each of the above methods satisfies one or more of our requirements. However, to our knowledge, no method can compute a global nonlinear registration on large images without the need to access the entire image data multiple times.

3.2 Domain Decomposition of Large-Scale Image Registration Problems

Domain decomposition describes a field of methods that aim to solve a large-scale problem arising from linear algebra or partial differential equations by instead solving smaller subproblems assigned to one of several nodes in a computer cluster [Dol15]. By repetitively communicating boundary information between these nodes, a global solution is computed in an iterative process.

An image registration method based on domain decomposition would satisfy all but one item from the list on page 10: efficient computation on a regular workstation. Despite being usually targeted on multi-node infrastructure, these concepts are discussed here since our new method takes up two ideas from domain decomposition: coupling the local problems by global operations (3.2.2) and solving localized nonlinear problems (3.2.3).

The problem of communication cost in the light of high single-core performance is also a question of efficiency of the respective domain decomposition algorithms [Lee17, p. xi]. The cost of communication is illustrated in the following calculation.

Side note: communication vs. computation [Nat16]

In a wire of three meters that connects two computing nodes, a signal needs 10^{-8} seconds to travel from one node to the other at the speed of light. A modern off-the-shelf workstation processor can compute circa 100 GFlops = 10^{11} floating-point operations per second. This means that in the time the signal travels from machine to machine

$$10^{-8}\text{s} \cdot 10^{11} \frac{\text{flop}}{\text{s}} = 1000 \text{ flop}$$

1000 floating-point operations can be computed.

Communication between nodes—especially if synchronous where one process is waiting for information from another one—can be more expensive than many additional computations.

Since the growth of computation speed of processors in terms of the number of instructions per second is slowing down, the trend in hardware development is towards parallel architectures [Dol15, pp. vii] and clusters of commodity computers [Sha10]. Today, communication between computing nodes or processors is, therefore, one of the main concerns in parallel computing.

If one is limited to one compute node, only one subdomain can be loaded into memory at any time, and communication between domains requires re-loading the data from slower memory. For comparison, loading an image region of 15 232 × 15 378 pixels from SSD into memory requires around 9.3 seconds on the reference computer. Thus, if re-loading the data from disk is necessary, domain-decomposition approaches are much more expensive. However, some approaches aim at limiting the communication. These strategies reduce the overall number of global iterations and are the foundation for the new method that we propose in this thesis.

Due to the nonlinearity of the image data, image registration is an intrinsically nonlinear problem. In numerical algorithms, unconstrained nonlinear problems are often solved iteratively, linearizing the problem in each step [Noc06, pp. 18 ff.]. Domain decomposition of nonlinear problems can be addressed in two ways:

- using linear domain decomposition on the linearized problem in each iteration [Cai94; Cai09] or
- solving nonlinear, local sub-problems on each of the subdomains and using a “globalization” step to re-establish the connection between the local sub-problems [Cai02; Kla17].

In this chapter, we address the decomposition based on the overlapping Schwarz method [Sch70] for a finite difference discretization. Other strategies like the family of finite element tearing and interconnect methods (FETI) [Far92; Far01] are used with a finite element discretization and follow the same general ideas (see [Kla14] and the references therein).

The remainder of this chapter is organized as follows. We start by describing linear domain decomposition in Chapter 3.2.2. These methods establish the ground for the

nonlinear approaches, which we discuss afterward in Chapter 3.2.3. We then compare the linear and nonlinear approach based on a model problem (Chapter 3.2.4). Our experiments show that the nonlinear decomposition approach requires less communication than a decomposition of the linear problem. As a conclusion from this chapter, we select the decomposition of the global problem into local nonlinear subproblems as a building block of the new method presented in Chapter 4.

3.2.1 Schwarz Domain Decomposition of Linear Problems

The common goal of linear Schwarz methods is to find a solution $u \in \mathcal{C}^2(\mathbb{R}^2), u : \Omega \subset \mathbb{R}^2 \mapsto \mathbb{R}$ for an elliptic differential equation such as

$$\begin{aligned} \Delta u &= f \quad \text{s.t.} \\ u &= 0 \quad \text{on } \partial\Omega \end{aligned}$$

where $\Delta : \mathcal{C}^2(\mathbb{R}^2) \mapsto \mathcal{C}^0(\mathbb{R}^2)$ is the Laplace operator and $f : \Omega \mapsto \mathbb{R}$ [Dol15, Chapter 1]. For a simpler presentation, we will only discuss the case of two subdomains.

Alternating Schwarz Methods

A domain decomposition method has been proposed even before the onset of computers by Schwarz in 1870 [Sch70]. In the article "Ueber einen Grenzübergang durch alternirendes Verfahren" Schwarz describes how a global solution to the Laplace equation on an irregular domain can be constructed based on the local solutions in two overlapping, regular subdomains $\Omega_1 \subset \Omega$ and $\Omega_2 \subset \Omega$. With the increasing popularity of parallel computing, the method was re-discovered in the 1970s and 1980s, see [Lio88] for a summary.

The core of the alternating Schwarz method is an alternation between two subdomains in which the Laplace equation is solved locally. In each iteration k, a local solution $u_i^{(k)}$, $i = 1, 2$ is computed on Ω_i with a constraint on the boundary that is in common with the other subdomain. This boundary constraint is set such that the local function matches the previous iteration's solution

$$u_i^{(k)}(x) = u_{3-i}^{(k-1)}(x) \text{ for } x \in \partial\Omega_i \cap \Omega_{3-i},\ i = 1, 2$$

on the other domain.

Schwarz showed that a) the series of solutions converges in each subdomain to a solution u_1^* and u_2^* respectively and b) both u_1^* and u_2^* are equal inside the overlap $\Omega_1 \cap \Omega_2$. Therefore the combination of both solutions solves the Laplace equation on Ω.

The concept of the alternating Schwarz method has been applied to discretized linear and nonlinear problems in different ways. For linear problems, the "Restricted

Additive Schwarz" or "Additive Schwarz Method" (ASM) are used as preconditioners [Efs03]. To solve a linear system

$$H\mathbf{s} = \mathbf{f}$$

with $\mathbf{s}, \mathbf{f} \in \mathbb{R}^N$, $H \in \mathbb{R}^{N\times N}$ using Schwarz preconditioning, the operator

$$M^{-1} = \sum_{k=1}^{K} R_k^T H_k^{-1} R_k$$

is defined using the restriction operator $R_k \in \mathbb{R}^{n\times N}, n \ll N$ and a local inverse $H_k^{-1} = (R_k H R_k^T)^{-1} \in \mathbb{R}^{n\times n}$. Note that multiplication with M^{-1} can be computed in parallel for each k. The preconditioned system

$$M^{-1}(H\mathbf{s} - \mathbf{f}) = 0$$

can then be solved using a linear solver such as a Krylov subspace method [Saa03, pp. 157 ff]. In each iteration, one additional multiplication with the preconditioner matrix M^{-1}, which is composed of K local solves is required. This means that ASM can be interpreted as a combination of first solving the problem on the restricted subspaces and then computing a global correction.

ASM is applied to linear problems either directly or as a preconditioner. However, the registration problem at hand is nonlinear due to the nonlinearity of the distance measure. In later extensions of ASM to nonlinear problems, convergence could be shown for several problem classes, see, e.g., [Lio88; Lui01].

In image registration, Newton-type iterations are often used as nonlinear solvers [Mod09; Che18]. At each step of the Newton iteration, a linearized subproblem is solved to obtain an updated approximation to the solution to the nonlinear problem.

A simple nonlinear domain decomposition method can be constructed, if the linear subproblem in each Newton step is computed using a linear domain decomposition method such as ASM. We discuss this and other strategies for domain decomposition methods for nonlinear problems in the following section.

3.2.2 Domain Decomposition for Newton-Type Iterations

In the literature, two approaches to generalize the alternating Schwarz method for Newton-type methods are pre-dominant: First, a Newton-like linearization scheme can be combined with a linear domain decomposition method that solves the linear system at each iteration. Examples are the Newton–Schwarz or Newton–Krylov–Schwarz methods [Cai94; Cai09].

In the second approach, a smaller nonlinear problem is solved separately on each subdomain. A globalization strategy transfers the information across subdomains after each round of nonlinear solves, as proposed in the ASPIN method [Cai02].

We will briefly discuss the first approach but eventually focus on the second: When the decomposition is performed in the linear solver, communication between the

subdomains is needed after each Newton step. If implemented on a single computing node, this would require re-loading the data from slower disk storage in each iteration and would thus not be realistic.

All the three methods use an additive approach where the result of the previous global step is used to couple the local subdomains across their interior boundaries. Multiple variants of the three methods have been published (see [Dry94; Dol16; Mot17] and the references therein), including multiplicative approaches [Liu15], where the subdomain solutions take the result of previously computed subdomains in the current global step into account. Similarly to the transition from a Jacobi to a Gauß-Seidel scheme [Saa03, Chapter 4], multiplicative methods are reported to have better convergence [Cai94]. Their disadvantage is the dependency of the solution on the order in which the subdomains are computed. To avoid this problem and to maintain a concise notation, we will only discuss the additive variants of the above methods.

In all the following methods, we decompose the reference domain Ω into overlapping subdomains $\Omega_k \subset \mathbb{R}^2, k = 1, ..., K$ with a corresponding discretized restriction operator $R_k : \mathbb{R}^{2N} \to \mathbb{R}^{2n}$ such that the number of pixels $n \ll N$ in the subdomain is much smaller than in the original image domain Ω.

Newton–Schwarz (or Schwarz–Alternating)

The simplest of the presented nonlinear domain decomposition methods is a nonlinear Schwarz method. Sometimes the term “Schwarz–Alternating” method is used [Cai09]. For consistency, we call it the Newton–Schwarz method, because the outer iteration is motivated by a Newton-type iteration, and the inner iteration is an additive Schwarz domain decomposition. The method is described in Algorithm 3.2.1.

In the algorithm, each of the overlapping subdomains Ω_k, k=1, ..., K is assigned to one computing node. A partition of the global initial guess $\mathbf{u}^{(0)}$, is copied to each node k as well as the corresponding data, such that $\mathbf{u}_k^{(0)} = R_k \mathbf{u}^{(0)}$.

All the discussed methods couple the subdomains either by an explicit interior boundary condition or implicitly by the differential operator. We formalize the boundary condition in all methods by adding the penalty function

$$b_k(\mathbf{u}_k^{(j)}) = \alpha_b \left\| R_k^{\partial} R_k^T \mathbf{u}_k^{(j)} - R_k^{\partial} \mathbf{u}^{(j-1)} \right\|_2^2$$

to the objective function where $R_k^{\partial} \in \mathbb{R}^{N \times \bar{n}}$ is the projection operator that selects the nodes on the interior boundary $\partial\Omega_k \setminus \partial\Omega$ of Ω_k.

The penalty approach does not guarantee that the boundary condition is fulfilled exactly, and other methods such as the augmented Lagrangian framework [Noc06, Chapter 17] are generally better suited [Noc06, p. 525] for a constraint optimization problem. In our comparison of domain decomposition methods, the previous iterations' solution $\mathbf{u}^{(j-1)}$ is an approximation, and we do not require an exact match.

Furthermore, the penalty method allows us to loosen the boundary requirement to observe the effect of the coupling of the subdomains.

The local gradients are computed as

$$\nabla J_k(\mathbf{u}_k^{(j)}) = R_k \nabla J(R_k^T \mathbf{u}_k^{(j)}) + \nabla b_k(\mathbf{u}_k^{(j)})$$

and the local Hessians are

$$H_k(\mathbf{u}_k^{(j)}) = R_k H(R_k^T \mathbf{u}_k^{(j)}) R_k^T + \nabla^2 b_k(\mathbf{u}_k^{(j)}) \tag{3.1}$$

where H is the approximation of the Hessian of J. We note that $R_k^T \mathbf{u}_k^{(j)} = R_k^T R_k \mathbf{u}^{(j)} \in \mathbb{R}^{2N}$ has the same size as $\mathbf{u}^{(j)}$ but that the elements outside Ω_k are zero. When copying the data to the nodes, it has to be assured that an additional margin around Ω_k is included such that the finite difference stencils in $R_k \nabla J(R_k^T \mathbf{u}_k^{(j)})$ and $R_k H(R_k^T \mathbf{u}_k^{(j)}) R_k^T$ can be computed.

The Newton–Schwarz iteration is computed as in Algorithm 3.2.1.

Algorithm 3.2.1 Newton–Schwarz Method, see [Dol15, p. 13]

```
for j=1, 2, ...:
    for k=1, ..., K:
        # solve the local problem in Ω_k
        H_k s_k = ∇J_k(u_k^(j))
        u_k^(j) = u_k^(j) + s_k
    for k=1, ..., K:
        for all neighboring subdomains Ω_l:
            # overwrite overlap from neighboring nodes
            u_k^(j+1)(x) = u_l^(j)(x)  for  x ∈ (Ω_k ∩ Ω_l)
```

As a final step, the local solutions $\mathbf{u}_k^{(j)}$, $k = 1, \ldots, N$ have to be collected and combined to a global solution $\mathbf{u}$.

This nonlinear analog to the linear additive Schwarz method may converge slowly or not at all [Cai09]. It also does not involve any global step, which distinguishes it from the following methods.

Newton–Krylov–Schwarz Method

The Newton–Krylov–Schwarz method is a preconditioning method tied to a global linear Krylov solver [Cai94; Cai09]. The Schwarz method is used as a linear preconditioner for a Newton–Krylov iteration. For a purely parallel implementation, the global step has to be parallelized separately using a parallel Krylov solver, see [Bal18].

The Newton–Krylov–Schwarz method works by preconditioning the Newton–Krylov iteration

$$H(\mathbf{u}^{(j)})\mathbf{s} = -\nabla J(\mathbf{u}^{(j)})$$
$$\mathbf{u}^{(j+1)} = \mathbf{u}^{(j)} + \mathbf{s}$$

where H is the approximation of the Hessian of J, and the linear system is solved using a Krylov method with a preconditioner

$$M^{-1} = \sum_{k=1}^{K} R_k^T \left(H_k(\mathbf{u}^{(j)}) \right)^{-1} R_k$$

that is composed of locally inverted Jacobians. The local inversion can be computed on subdomains in parallel. The preconditioned system is then solved in each Newton step such that

$$M^{-1}H(\mathbf{u}^{(j)})\,\mathbf{s} = M^{-1}\nabla J(\mathbf{u}^{(j)}) \tag{3.2}$$
$$\mathbf{u}^{(j+1)} = \mathbf{u}^{(j)} + \mathbf{s} \tag{3.3}$$

resulting in Algorithm 3.2.2.

Algorithm 3.2.2 Newton–Krylov–Schwarz Method [Cai94]

```
for j=1, 2, ...:
    for k=1, ..., K:
        # solve the local problem in Ω_k
        H_k s_k = ∇J_k(u_k^(j))
        # combine s_k, k = 1,...,K to s̄
        s̄ = Σ_{k=1}^{K} R_k^T s_k
    # solve the global system
    M^{-1} H s = s̄
    u^(j+1) = u^(j) + s
```

While the preconditioner can be applied completely in parallel, a global system has to be solved in each step. Also, communication between subdomains is needed after each round of local linear solves. The method is therefore not suited for the large-scale registration problem at hand. A combination of the ideas of Newton–Krylov–Schwarz and the solution to a local nonlinear problem is found in the ASPIN method.

3.2.3 Additive Schwarz Preconditioned Inexact Newton

A method building on top of the idea of Newton–Krylov–Schwarz (NKS) is the so-called “Additive Schwarz Preconditioned Inexact Newton” (ASPIN) [Cai02]. In contrast to NKS, the preconditioning is performed by nonlinearly solving a local subproblem. This leads to a reduction of global iterations and thus reduces communication between nodes.

The motivation of localized preconditioning is that the global (inexact) Newton step is limited by local (unbalanced) nonlinearities, which can slow the convergence down. Cai and Keyes [Cai02] propose to partition the problem into smaller nonlinear subdomains on which a separate, local, approximate solution is computed. This intermediate solution is then used as a preconditioner of the full problem. We again start with the original Newton–Krylov iteration to solve the optimization problem

$$J(\mathbf{u}) \longrightarrow \min$$

where a series of updates to $\mathbf{u}^{(j)}$, $j = 1, 2, \ldots$ is computed.

In ASPIN, the preconditioning is computed by first solving the local nonlinear problems

$$J_k(\mathbf{u}_k^{(j)} + \mathbf{s}_k) \xrightarrow{\mathbf{s}_k} \min \tag{3.4}$$

using an iterative method. The local solution $\mathbf{u}^{(j)} + \mathbf{s}_k$ is only updated in the subdomain Ω_k but ghost values of $\mathbf{u}^{(j)}$ from the outside of the subdomain are needed to compute (3.4). These ghost values are not updated in the iteration. The results are combined to $\bar{\mathbf{s}} = \sum_{k=1}^{K} R_k^T \mathbf{s}_k$ and used as the new right-hand side in a global (inexact) Newton step

$$M^{-1} H(\mathbf{u}^{(j)} + \bar{\mathbf{s}})\, \mathbf{s} = \bar{\mathbf{s}}. \tag{3.5}$$

The preconditioner

$$M^{-1} = \sum_{k=1}^{K} R_k^T \left(H_k(\mathbf{u}_k^{(j)} + \mathbf{s}_k) \right)^{-1} R_k \in \mathbb{R}^{2N \times 2N}$$

is based on the locally inverted Hessians. The above steps (3.4) and (3.5) are repeated until convergence in Algorithm 3.2.3.

Algorithm 3.2.3 Extended ASPIN Method, based on [Cai02]

```
for i=1, 2, ...:
    for k=1, ..., K:
        s_{k,1} = 0
        for j=1, ...: # inner local loop
            # solve the local problem in Ω_k
            H_k(u^(j) + s_{k,1}) s_{k,2} = R_k ∇J(u^(j) + s_{k,1})
            s_{k,1} = s_{k,1} + s_{k,2}

    # combine s_k, k = 1, ..., K to s̄
    s̄ = Σ_{k=1}^{K} R_k^T s_k
    for j=1, ..., N_g: # inner global loop
        M^{-1} H s = s̄ # solve for s
        u^(j+1) = u^(j) + s
```

Generalizing the original ASPIN method [Cai02], we allow multiple inner global iterations $N_g > 1$. In the original method, this number is set to $N_g = 1$. In our experiments,

we observe faster convergence if multiple inner global iterations are computed in each outer step.

In its extended form, the ASPIN method is a generalization of the Newton–Schwarz and the Newton–Krylov–Schwarz methods. The Newton–Schwarz method is obtained by stopping after one local iteration, setting $N_g = 0$, and moving the final update step $\mathbf{u}^{(j+1)} = \mathbf{u}^{(j)} + \bar{\mathbf{s}}$ in the outer loop. The NKS method is obtained by computing only one iteration in both the inner local and inner global loop ($N_g = 1$).

Structure of the Preconditioned Operator

In the initial publication, Cai and Keyes [Cai02] apply the ASPIN method to a cavity flow problem and present numerical experiments. Especially in the case of high turbulences, nonlinear preconditioning reduces the number of global iterations that are needed to reach a comparable residual.

The authors show that if started in a neighborhood of the true solution, the ASPIN method computes the same local minimum as the original inexact Newton method. The authors also note, that in the non-overlapping case, the product (in a two-subdomain example)

$$\mathcal{H} = \sum_{k=1}^{2} \left(H_k(\mathbf{u}_k^{(j)}) \right)^{-1} H(\mathbf{u}^{(j)}) \text{ with } H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}$$

has a block structure

$$\mathcal{H} = \begin{bmatrix} I & H_{11}^{-1} H_{12} \\ H_{22}^{-1} H_{21} & I \end{bmatrix}$$

where the main diagonal is one. The difference between a solution computed on subdomains only and a solution computed on the whole domain is in the coupling of these subdomains.

The main disadvantage of the ASPIN method for a large-scale image registration problem is shared with the other domain decomposition methods that we discussed: Multiple iterations, each including the complete image data, are required to find the final solution.

In the CLG registration method proposed in this thesis, we avoid the preconditioned global operator but maintain the two main ideas of ASPIN: the decomposition into local nonlinear problems and the need for a coupling between the subdomains.

In the following chapter, we compare the amount of global communication in the three presented approaches based on a model image registration problem.

3.2.4 Comparison Study

In the decomposition approaches discussed in Chapters 3.2.2 – 3.2.3, a global step requires communication between subdomains to transfer boundary information and—in the case of NKS and ASPIN—to solve a global registration problem. We want to

avoid this kind of inter-subdomain communication to be able to decompose the image registration problem into smaller, independent sub-problems.

In the following experiment, we show that the NKS and ASPIN methods lead to a reduction in global steps in a model registration problem. Especially the decomposition into nonlinear sub-problems in the ASPIN method is effective, lowering the number of global Gauß-Newton iterations to 2 compared to 37 in a standard Gauß-Newton method.

We base the comparison of the algorithms on the number of global and local iterations. Our focus is on the cost of communication such that we pay less attention to computation time and memory requirements. However, the iteration count can give some insight into the computation time. The memory consumption of the three methods is comparable because they are composed of the same building blocks. Only the Newton–Schwarz iteration does not require a global operation and, therefore, requires less memory.

For the different methods, a global iteration consists of

- global only: one global Gauß–Newton step
- Newton–Schwarz: one local Gauß–Newton step on each subdomain
- Newton–Krylov–Schwarz: one local Gauß–Newton step on each subdomain and one global Gauß–Newton step
- ASPIN: multiple local Gauß–Newton steps on each subdomain and N_g global Gauß–Newton steps

For ASPIN, we implement a modified algorithm that allows all components to be expressed as standard Gauß–Newton registrations: Instead of the global ASPIN step from Equation 3.5

$$M^{-1}H(\mathbf{u}^{(j)} + \bar{\mathbf{s}})\,\mathbf{s} = \bar{\mathbf{s}} \tag{3.6}$$

$$\text{with } M^{-1} = \sum_{k=1}^{K} R_k^T H_k(\mathbf{u}^{(j)})^{-1}, \tag{3.7}$$

as the sum of the subspace inverse Hessians, we compute this step as

$$H(\mathbf{u}^{(j)} + \bar{\mathbf{s}})\,\mathbf{s} = \nabla J(\mathbf{u}^{(j)} + \bar{\mathbf{s}}), \tag{3.8}$$

which is only exact in the non-overlapping case. Without overlap of the subdomains, $\sum_{k=1}^{K} R_k^T H_k(\mathbf{u}^{(j)})$ is a block-diagonal matrix, such that the preconditioner can be multiplied with both sides and the right-hand side

$$M\,\bar{\mathbf{s}} = \left(\sum_{k=1}^{K} R_k^T H_k\right)\left(\sum_{k=1}^{K} R_k^T \mathbf{s}_k\right) = \sum_{k=1}^{K} R_k^T H_k \mathbf{s}_k = \sum_{k=1}^{K} R_k^T \nabla J_k = \nabla J$$

is the Jacobian of the objective function. This modification avoids constructing the combined operator $M^{-1}H$ or solving a linear system in every multiplication with $M^{-1}H$ if an iterative linear solver is used. Effectively, the combined local solutions

Table 3.1: Parameters used in the computation of the model registration problem.

parameter	value
image size	120 × 240 pixels
subdomain image size	120 × 124 pixels
overlap between subdomains	8 pixels
distance measure	sum of squared differences (SSD)
regularizer	curvature
regularization parameter α	1.0
boundary penalty weight α_b	0, 0.01, 1, 1000

are used as an initial guess for the global iteration. It allows the ASPIN method to be implemented as two interleaved Gauß–Newton iterations which only rely on the Hessians of J and $R_k J$. In the overlapping case, the result may differ from the original ASPIN iteration, which does not affect the comparison of the linear and nonlinear solutions to the local subproblems.

We compare the methods using the model problem shown in Figure 3.2.1. In the reference image, the registration problem consists of a box and a stripe pattern on a black background. In the template, the same structures have been deformed artificially. The objects are located such that each subdomain contains one object, and their independent solution will differ in the overlap region. The parameters used for this experiment are shown in Table 3.1. We arbitrarily choose $\alpha = 1$. Different boundary penalty weights are used to observe the effect of the coupling of the two subdomains.

In all methods, we perform a global pre-registration on a coarse grid of 9 × 16 pixels as an initial guess (Figure 3.2.1, top).

The fine global Gauß–Newton registration reaches the stopping criterium (see Chapter 2.6.1) after 38 steps and the final objective function value will be referred to as $J(\mathbf{u}_{\text{final}}) = J_{\text{final}}$. The result of this registration is shown in (Figure 3.2.1, bottom). The other methods are stopped once the reference objective function value J_{final} is reached in order to make the results independent of the stopping criteria.

Coupling of Subdomains by Penalty, Global Iterations, and Coarse Space

When altering α_b from 1 to 1000 or 0.1, the iteration numbers differ by one step or less across all methods. Without the boundary conditions ($\alpha_b = 0$), none of the standard methods reaches the target value J_{final} before the maximum number of 100 global or 200 local iterations is reached. However, the modified ASPIN variant with multiple inner global iterations ($N_g < 10$) reaches the target function value after a comparable number of local and global iterations—even without the interior boundary condition.

We conclude that in our experiment, a coupling between the subdomains is required

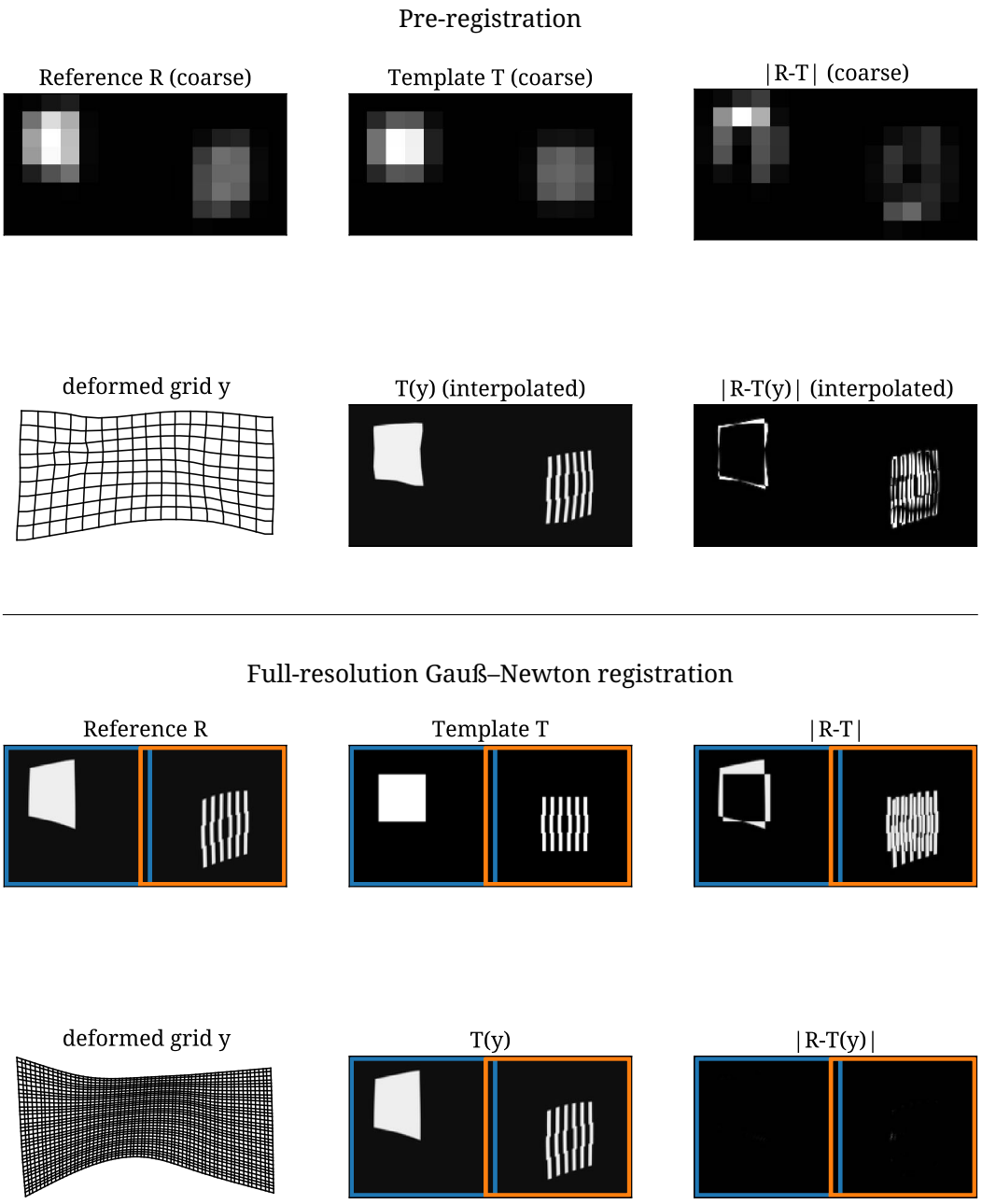


Figure 3.2.1: Image registration model problem after coarse pre-registration and full-resolution Gauß–Newton registration. For each registration, six images are shown. First row: Reference image R, template image T, $|R-T|$. Second row: Coarse-resolution Gauß–Newton result y visualized as a transformed grid. Deformed template image $T(y)$, $|R-T(y)|$. After coarse registration, an offset ist visible in the difference image, which mostly disappears after full-resolution registration. The subdomains that are used in the Newton–Schwarz, Newton–Krylov–Schwarz, and ASPIN methods are marked in blue and orange.

Table 3.2: Number of iterations needed to reach the target objective function value J_{final} that has been determined based on the Gauß–Newton iteration. In all methods, 18 global iterations on a coarse level are not included in the global iteration count. The ASPIN method requires, by far, the lowest number of global iterations.

Method	local iterations	global iterations
Gauß–Newton	0	38
Newton–Schwarz	144	72
Newton–Krylov–Schwarz	38	19
ASPIN $N_g = 1$	56	2
ASPIN $N_g < 10$	50	2

for the local methods to converge. Furthermore, additional inner global iterations in the ASPIN method can have this coupling function.

Another way such a coupling is realized in the literature is by introducing an embedded coarse space [Nic87; Dol15, pp. 102 ff; Mar05; Kei14]. In these methods, an additional coarse global system is solved in each global iteration. Heinlein and Lanser [Hei19] compare different alternatives to combine a nonlinear coarse problem with different variants of the ASPIN method. The authors find a reduction in outer global iterations and also in overall GMRES iterations when using a coarse space. The reduction is stronger for higher numbers ($K \geq 25$) of subdomains.

In the nomenclature used in [Hei19], our coarse pre-registration can be interpreted as an additive coarse space method. However, the coarse correction would need to be computed in each global iteration and not only once at the beginning, which can be relevant when computing three or more outer iterations.

Faster Convergence Due to Multiple Inner Iterations

The number of iterations for the decomposition methods are shown in Table 3.2.

Among the three methods, the ASPIN method requires the lowest number of global iterations in our experiments (Table 3.2). The characteristic attribute of ASPIN among the other methods is the solution to a nonlinear subproblem between global solves, which reduces the number of global iterations. However, in each global iteration, at least one global high-resolution linear system has to be solved, which will exceed the available memory in the large-scale image registration problem.

In Chapter 4, we present a new method that uses these two strategies—locally nonlinear subproblems and an additional coupling by a coarse global objective function—to skip the communication and to compute an approximated solution in only one sweep through the subdomains.

The proposed absence of global iterations can lead to differences between the solutions calculated on the subdomains. In the next chapter, we discuss methods for combining subdomain solutions proposed in the literature.

3.3 Blending of Multiple Deformations

The proposed large-scale image registration method (Chapter 4) skips the communication step between subdomains, which potentially results in contradictory deformations at their interface. In the literature, some concepts to blend differing deformations are discussed. We summarize these concepts in this chapter.

In the following methods, in order to improve the result of a registration problem, multiple solutions for the same objective functions are computed, and the resulting deformations are combined locally to an improved solution.

In *MetaReg*, the method proposed by Muenzing, van Ginneken, and Pluim [Mue12], multiple methods for lung registration are applied to the same image pair. The quality of the results is measured by automatically detected landmarks—mostly at lung vessel branches. The image domain is decomposed into subdomains, in each of which the best registration result is determined. The different results are first linked without modification at the subdomain borders, although this results in folding artifacts at the interfaces. The following step is named *UnfoldReg*: Existing folding artifacts are detected automatically, and the affected regions and their surroundings are removed. B-Spline interpolation [Tus06] is used to fill the empty areas based on the uncorrupted deformation in the vicinity.

The *MetaReg* method provides an automatic approach to evaluate concurrent registration results and eliminates folding artifacts from the blended result. However, the interpolated deformation does not adhere to the original registration problem such that it is difficult to evaluate the quality of the composed deformation. Moreover, all image information in the replaced area is discarded so that local nonlinearities within the area can hardly be restored using this technique.

In the *DIRBoost* method [Mue14], the question of combining multiple global deformations is addressed. First, a registration is computed by applying an established registration method multiple times with different local weights. In each iteration, the deformation is evaluated using a voxel-wise registration error detection method. A *DIRBoost* update is computed by only admitting the trusted regions of the computed deformation to the solution and filling the untrusted region using a smoothing method. Next to the *UnfoldReg* approach, another mentioned smoothing approach is to reposition B-Spline nodes in the corrupted region and to rely on the implicit smoothness of the B-Splines instead [Mod10].

For multi-channel registration, Forsberg et al. [For11] propose a weighted averaging of the displacements of the different channels in diffusion tensor imaging. As weights, the authors propose the normalized norm of the target image's gradient. To

blend multiple rigid deformations on subdomains, Kajihara et al. [Kaj19] use a normalized weight function based on the distance to a central point in the subdomain. The authors apply their method to histological serial sections, where—if multiple objects are placed on one slide—these objects sometimes move independently.

Weighted averaging approaches have the advantage that no image information is required in the blending procedure and that local nonlinearities are preserved. It is also computationally inexpensive. However, artifacts in one of the candidate deformations are likely to affect the blended result as well.

Lempitsky, Roth, and Rother [Lem08] propose a combination of different flow fields from an optical flow problem (see, e.g., [Ban00, pp. 382 ff.]). Different solutions for the optical flow problem are computed using different parameterizations of two optical flow methods. The combination of the solutions is formulated as a discrete optimization problem: For each node of the displacement, one of the pre-computed solutions is selected using a graph cut method [Boy01]. The objective function in this step consists of an image distance term local to each node and a regularization term penalizing the difference between the displacement of neighboring nodes.

An extended method following a similar idea is proposed by [Hei16] for image registration. The authors combine multiple candidate displacements computed on different supervoxel segmentations of the image. The image is partitioned into several layers of supervoxels, where each layer is a different supervoxel representation of the same image. Regularity is obtained by penalizing the differences in the displacement of neighboring supervoxels in a graph structure. For each supervoxel, a cost consisting of image distance and regularity is assigned. The k displacements with the lowest cost are computed for each supervoxel, defining local probability distribution of displacements. The best displacement vector for each pixel is selected based on the probabilities determined on all supervoxels overlapping at its position.

Both of the last two methods evaluate a set of possible deformations based on an image distance and a regularity term, and their solution is therefore conceptually close to the original registration problem. However, the authors solve the problem of combining multiple global deformations, which is different from the problem of blending neighboring deformations that result from a domain decomposition. Also, the approach seems too computationally expensive in time and memory to be applied in the large-scale image registration problem.

We compare different blending approaches in Chapter 5. We include some of the above ideas but exclude those approaches that require image information to avoid reloading of the image data.

4 A Novel Approach for Combined Local and Global Image Registration

Contents

In this chapter, we focus on the core of this thesis: the combined local and global registration to solve the large-scale image registration problem (1.2). We aim to find a registration method to register large images that is accurate in terms of remaining image distance and spatially consistent in terms of a globally regular deformation. In the variational formulation described in Chapter 2, the distance measure can be evaluated using local data only. However, the regularizer couples the registration problem such that information from the entire domain is needed to compute the deformation (Chapter 2.8).

We propose a new registration method that combines a coarse global image registration problem with local, high-resolution image data on subdomains (Chapter 4.2).

The proposed method approximates a global solution by combining global solutions computed on local subdomains. Each subdomain is computed independently of the solutions computed on neighboring subdomains. The price for the independent computation is paid in global regularity. A simultaneous coarse global optimization is used to reduce the impact of the independent computation. This hybrid solution has different properties than coarse global, fine global, and purely local solutions. Using artificial (Chapter 6.1) and realistic (Chapter 7) experiments, we evaluate these solutions in terms of deformation error, image distance and regularity. We observe that, qualitatively, the new method is superior to a coarse global and to a purely subdomain-based registration while it does not reach the same result as a fine global registration.

The new method is based on the nonlinear registration method described in Chapter 2. The discretized image registration problem is written as a minimization of the objective function (2.6)

$$J(R, T, \mathbf{u}) = D(R, T, \mathbf{y}) + S(\mathbf{u}) \xrightarrow{\mathbf{u}} \min,$$

where a minimum is found by iteratively updating the candidate solution $\mathbf{u} \in \mathbb{R}^{2m_1m_2}$ using the gradient ∇J and a Hessian approximation (see Chapter 2.6.1). The deformation $\mathbf{y} = \mathbf{u} + \mathbf{x}$ is computed by adding the identity deformation $\mathbf{x}$ to the displacement $\mathbf{u}$. For this computation, all inputs and intermediate variables are stored in the computer's main memory.

For images that surpass the available RAM, we first discuss two options (Chapter 4.1) to reduce the size of the input data: a) converting the images to a lower image resolution which reduces accuracy and b) solving the problem only for a part of the image domain which reduces global consistency.

Based on these options and given a limited amount of RAM, we propose the combined local and global (CLG) subdomain registration method that combines the two reduction ideas to approximate a global solution to the image registration problem (1.2) using local high-resolution image data.

Including the new CLG method, we compare four different approaches to compute a solution for the large-scale image registration problem (1.2): coarse global registration, fine global registration, purely local subdomain registration, and the CLG subdomain registration.

Fine global registration is a global registration with a fine image resolution. We define this registration as a gold-standard using small image sizes as it cannot be applied to a large-scale image without exceeding the computer's memory.

Coarse registration is a global registration where the images are interpolated based on a downsampled version of the image data such that the registration problem can be solved without exceeding the memory capacity of the computer.

The following two methods are proposed in this thesis and use a spatial partitioning.

Purely local subdomain registration is a registration that uses subdomains with local image information at a high image resolution. The results are computed independently, and the deformations computed on the subdomains are combined in a subsequent blending step (see Chapter 5).

CLG: Combined local and global subdomain registration approximates a fine global registration. The deformation is optimized globally, and the image distance is optimized simultaneously on a coarse global resolution image and a local high-resolution subdomain. The deformations are computed independently for all subdomains, and they are combined in a subsequent blending step (see Chapter 5).

All four registration methods use the registration framework described in Chapter 2.

4.1 Division of the Registration Problem into Smaller Subproblems

The proposed global subdomain registration is a combination of two strategies to address the large-scale registration problem (1.2): Reducing the problem size and splitting the large-scale problem into multiple smaller ones. Both approaches can also be used independently as described below.

4.1.1 Consequences of Low Image Resolution

A multilevel scheme is used in image registration (Chapter 2.5.2) for two reasons: a) The local convexity of the objective function is increased by smoothing the image data and b) the computational cost is reduced by first solving less expensive problems with a lower number of pixels. A low-resolution registration can be obtained by omitting a number of high-resolution layers at the bottom of the image pyramid and stopping the multilevel scheme early.

The inevitable drawback of this strategy is that fine structures are not represented in the smoothed low-resolution images. Therefore, the accuracy of the low-resolution registration is reduced.

Both drawbacks could be avoided by switching to a multi-scale approach [Mod09, p. 145] where no downsampling but only smoothing is used to convexify the objective function. However, without downsampling, the multi-scale approach does not reduce the size of the input data and is therefore not suitable for the large registration problem (1.2).

4.1.2 Preliminary Purely Local Subdomain Registration

To compute a registration on large high-resolution images, we propose a decomposition approach that solves the registration problem on smaller subproblems. We first introduce a preliminary, purely local approach that is extended by a global component at a later point in Chapter 4.2.

We divide the image domain Ω into K overlapping subdomains $\Omega_k = [\omega^k_{1,1},\ \omega^k_{1,2}] \times [\omega^k_{2,1},\ \omega^k_{2,2}]$, $k = 1, \ldots, K$, such that $\cup_k \Omega_k = \Omega$. On each subdomain, overlapping solutions are computed. The overlap is later used to blend the local solutions.

We have proposed a simple subdomain-based registration approach based on this decomposition [Lot16a]. First, a coarse global pre-registration is computed, and the result is used as an initial guess for the following subdomain registrations and to determine the positions of the template subdomains. In a second step, independent registration problems

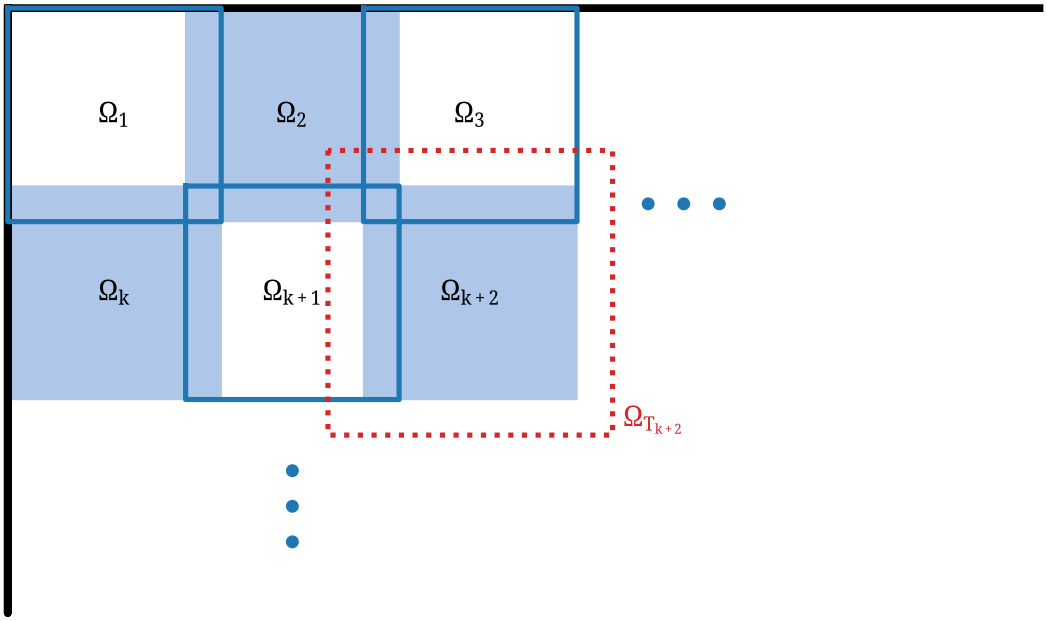


Figure 4.1.1: Decomposition of the domain into overlapping reference subdomains Ω_k (blue). The template image subdomains Ω_{T_k} are based on the reference image subdomains and extended by an additional safety margin (shown in dashed red).

$$J_{\text{loc}}(\mathbf{u}_k) = \Big[D(R,T,\mathbf{y}_k)\Big]_{\Omega_k} + S(\mathbf{u}_k) \longrightarrow \min \tag{4.1}$$

with $\mathbf{y}_k = \mathbf{u}_k + \mathbf{x}$ are solved for all subdomains Ω_k, $k = 1,\ldots,N$. Here, $[D(R,T,\mathbf{y}_k)]_{\Omega_k}$ denotes the evaluation of the distance measure in the local subdomain only. For each subdomain, we compute a solution $\mathbf{u}$ on Ω. Since no image information is used outside of Ω_k, the displacement is extended with minimal regularizer energy.

To compute a solution to Equation (4.1), only the image data inside Ω_k is needed. The subdomain can be chosen such that the required image data can be loaded completely into RAM. For the reference image, the required image data can be obtained by projecting R onto Ω_k. For the deformed template image, the deformation has to be taken into account. We estimate an extension $u_{\max} = \|\mathbf{u}_{\text{hist}}\|_{\max}$ from previously computed deformations $\mathbf{u}_{\text{hist}}$. The template image data is loaded on the extended domain $\Omega_{T_k} = [\omega^k_{1,1} - u_{\max},\ \omega^k_{1,2} - u_{\max}] \times [\omega^k_{2,1} + u_{\max},\ \omega^k_{2,2} + u_{\max}]$ (Figure 4.1.1).

The resulting deformations may differ in the overlap $\Omega_k \cap \Omega_j$ of neighboring pairs k, j

of subdomains. We compute a weighted sum

$$\mathbf{u} = \sum_{k=1}^{N} \mathbf{w}_k \cdot \mathbf{u}_k$$

where $\cdot$ is the element-wise multiplication with a weight vector $\mathbf{w}_k \in \mathbb{R}^{2m_1m_2}, \mathbf{w}_k \geq 0, \sum_k \mathbf{w}_k = (1, ..., 1)$, to blend the deformations in the boundary regions. Blending strategies are discussed in more detail in Chapter 5.

This simple subdomain-based registration can be computed entirely on local data, which reduces the required memory. An estimate of the memory reduction can be obtained by the following calculation, assuming that the required memory depends linearly on the size of the image data. One whole slide image of 100 000 × 100 000 pixels requires approximately 10 GB of memory to store the image only (see Chapter 2.9). We use this number as a lower bound and neglect the additional storage for the multilevel representations and the deformation. In comparison, an image subdomain of 10 000 × 10 000 pixels needs 100 MB of memory, which is only 1 percent of the original memory requirement. A more detailed analysis of the memory requirement of the subdomain-based approach is made in Chapter 4.2.3.

The disadvantage of this approach is its locality: Since only local registration problems are solved, information outside of the subdomain is not considered. Most importantly, subdomains with low contrast or consisting of image background cannot rely on global regularization as they would in a global registration. At the interface between the subdomains, the lack of global information can lead to a mismatch of the deformation. We address these properties in experiments in Chapter 6.1.

4.1.3 Different Resolutions of Deformation and Images

Storing the discretization of the deformation at image resolution would require the same number of variables as in a full-resolution image. Depending on the data type, the required storage space in memory would even be higher. Aiming for a low memory footprint, we choose to represent the deformation by an interpolation function based on a coarse grid of coefficients. Low-resolution grids are used in some registration software packages such as NIFTYREG [Mod10], but some argue that a higher number of coefficients leads to a more accurate registration [Man16].

Throughout the registration, we optimize the objective function with respect to this fixed number of coarse grid coefficients. The disadvantage of such an interpolation strategy is that this kind of deformation can not represent nonlinearities that are on a smaller scale than the distance between two nodes of the coarse grid.

Having a coarse deformation grid combined with a fine image grid can be motivated by an analogy to parametric registration. Here, even fewer transformation parameters are optimized. As long as the deformation model has sufficient degrees of freedom, a higher image resolution leads to a better match in the registration [Zha16].

A registration with respect to a fixed low-resolution deformation grid can be interpreted as a parametric registration, where interpolation is used to derive a global deformation from a small number of parameters.

Another option would be to use an adaptive strategy, such as presented by Haber, Heldmann, and Modersitzki [Hab07b]. The authors propose an adaptive approach that discretizes the image and the deformation in a tree structure, which is refined adaptively during the optimization. Such an adaptive strategy could be an extension of the method that we propose in this thesis.

In the following, we denote the number of points of the deformation grid $\mathbf{x}_D \in \mathbb{R}^{2\cdot(m_1\cdot m_2)}$ by $m_1 \times m_2$ and the number of pixels in the image by $M_1 \times M_2$, with $M_1 \gg m_1$ and $M_2 \gg m_2$, defined on the image grid $\mathbf{x}_I \in \mathbb{R}^{2\cdot(M_1\cdot M_2)}$.

When evaluating the deformed template image, an interpolation of the deformation grid onto the image grid is necessary. We implement this grid change using a linear interpolation operator $P \in \mathbb{R}^{2\cdot m_1\cdot m_2 \times 2\cdot M_1\cdot M_2}$ [Kön18], and write the objective function as

$$J(R, T, \mathbf{u}) = D(R, T, P\mathbf{u}) + S(\mathbf{u}). \tag{4.2}$$

While the image is discretized on a cell-centered grid, the coarser grid consists of nodal grid points defined on the same domain such that there is always an outward deformation node for each image node (Figure 4.1.2). To be able to decompose the domain into subdomains later, we construct the grids such that each cell of the image grid is contained in one cell of the deformation grid (Figure 4.1.2).

Formally, the deformation grid is defined as

$$\begin{aligned} \mathbf{x}_D &= \left(1_{m_2} \otimes \mathbf{x}_{1\mathrm{D},1}^{\mathrm{nodal}},\ \mathbf{x}_{1\mathrm{D},2}^{\mathrm{nodal}} \otimes 1_{m_2}\right) \in \mathbb{R}^{2\cdot(m_1\cdot m_2)} \text{ where} \\ \mathbf{x}_{1\mathrm{D},d}^{\mathrm{nodal}} &= (\omega_{d,1} + h_D \cdot i)_{i=0,\dots,m_d-1} \in \mathbb{R}^{m_d}, d \in \{1, 2\} \end{aligned} \tag{4.3}$$

where h is the width (and height) of a quadratic deformation grid cell and $(\omega_{1,1}, \omega_{2,1})$ is the upper left corner of the reference image domain Ω (see also Chapter 2.5). A matching image grid $\mathbf{x}_I$ for the above deformation grid can be written as

$$\begin{aligned} \mathbf{x}_I &= \left(1_{M_2} \otimes \mathbf{x}_{1\mathrm{D},1}^{\mathrm{cc}},\ \mathbf{x}_{1\mathrm{D},2}^{\mathrm{cc}} \otimes 1_{M_2}\right) \in \mathbb{R}^{2\cdot(M_1\cdot M_2)} \text{ where} \\ \mathbf{x}_{1\mathrm{D},d}^{\mathrm{cc}} &= \left(\omega_{d,1} + h_I \cdot (i + \frac{1}{2}))\right)_{i=1,\dots,m_d} \in \mathbb{R}^{m_d}, d \in \{1, 2\} \end{aligned} \tag{4.4}$$

and the operator P maps from $\mathbf{x}_D$ to $\mathbf{x}_I$. To ensure that the cells of the image grid $\mathbf{x}_I$ do not intersect with the cells of the deformation grid, the image grid is defined on the same domain and the width and height of the image grid cells

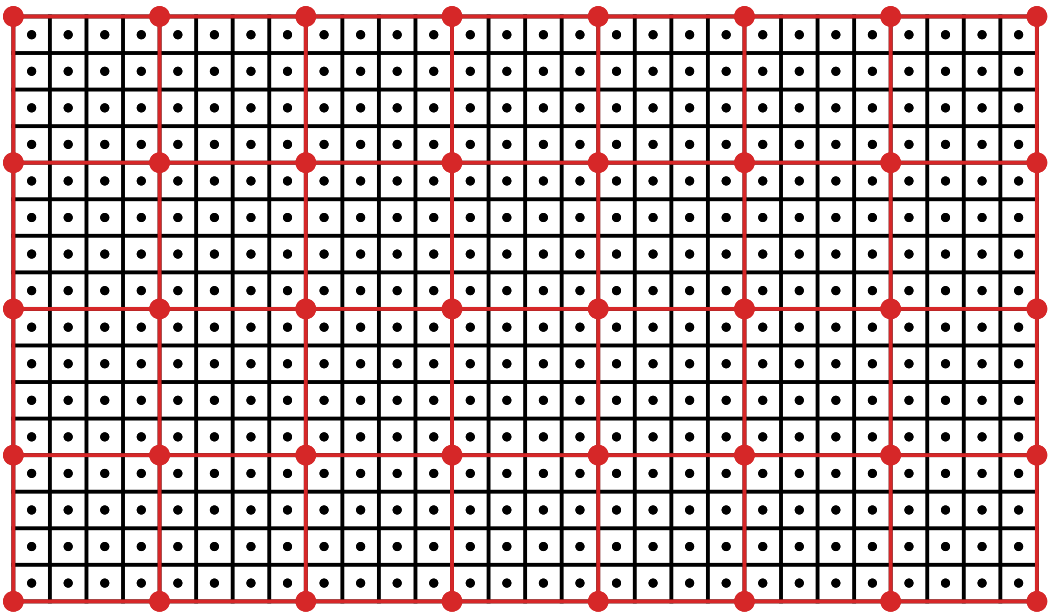


Figure 4.1.2: Coarse nodal grid (red) combined with a fine cell-centered image grid (black). The size of the deformation grid cells is four times the size of the image grid cells in each dimension. $M = (16, 32)$, $m = (5, 9)$.

$$h_I = \frac{h_D}{n},\ n \in \mathbb{N}$$

is set to an integer fraction of h_D. This is important to be able to partition the domain for subdomain-based processing.

When evaluating the template image with respect to a coarse deformation, we make use of the continuous image model. The image T is a representation of the underlying image data, and its values are obtained by interpolation at arbitrary points. In addition, we use a multiresolution approach [Mod09, pp. 40-41], where the underlying pixel data is smoothed and downsampled to match the image grid resolution approximately.

When computing the derivative of the objective function (Equation 4.2), the transpose of the prolongation operator is needed such that

$$\nabla J(R, T, \mathbf{u}) = \nabla D(R, T, P\mathbf{u}) + \nabla S(\mathbf{u}) \tag{4.5}$$

$$= P^T \nabla_{P\mathbf{u}} D(R, T, P\mathbf{u}) + \nabla S(\mathbf{u}), \tag{4.6}$$

where $\nabla_{P\mathbf{u}} D$ is the gradient of D evaluated on the image grid. P^T is the innermost derivative of $\nabla D(R, T, P\mathbf{u})$.

Compared to a low-resolution registration with an equally spaced deformation grid ($h_D = h_I$), higher image resolution is used to compute the gradient of the distance measure. Therefore, finer image structures contribute to the computation without the need to store a full-resolution deformation grid. The new subdomain registration approach proposed in the next chapter uses this combination of low-resolution deformation and high-resolution images to compute a simultaneously local and global image registration.

4.2 Combined Local and Global Registration (CLG)

We have established a first purely subdomain-based registration approach in Chapter 4.1.2 and discussed the advantages of global regularization in Chapter 2.8. In this chapter, we extend the purely local approach with a coarse global-resolution term.

When applying a domain decomposition approach to the registration problem, information needs to be distributed between subdomains after each optimization step (see Chapters 3.2 and 3.1.4). For each step of the Newton (Chapter 2.6.1) or L-BFGS method (Chapter 2.6.2), a linear system has to be solved that depends on the results of the previous iteration. This means that in each step of the iteration, updates are required from outside the current subdomain. In a domain decomposition method, the values of the boundary region are communicated between nodes whenever needed. While this is possible on a multi-node cluster system, it cannot be done on a single machine, where not all subdomains can be processed simultaneously.

This constitutes the dilemma that we address in this thesis: A high-resolution registration is more accurate but not globally feasible on a single workstation. On the other hand, global regularization improves the quality of the deformation, especially in regions, where the image contrast is low. We propose a new method called *combined local and global registration* (CLG). The new method seeks a compromise between local accuracy and global regularization by combining a coarse global and a fine local registration problem in one objective function.

4.2.1 Extension of the Objective Function

The CLG methods computes a set of global deformations that are combined into one final deformation in a second blending step (Chapter 5). Table 4.1 lists the method's in- and output.

The new method can be written formally as the repeated minimization of the local objective function

$$J_{\mathrm{CLG}}(\mathbf{u}_k) = \Big[D(R_k, T_k, \mathbf{u}_k)\Big]_{\Omega_k} + \Big[D(R_0, T_0, \mathbf{u}_k)\Big]_{\Omega\setminus\Omega_k} + S(\mathbf{u}_k) \tag{4.7}$$

Table 4.1: In- and output of the CLG registration method.

Input	
$R_0 : \Omega \mapsto \mathbb{R},\ T_0 : \mathbb{R}^2 \mapsto \mathbb{R},\ T_0, R_0 \in \mathcal{I}$	reference and template image, coarse resolution
$R_k : \Omega_k \mapsto \mathbb{R},\ T_k : \mathbb{R}^2 \mapsto \mathbb{R},$ $T_k, R_k \in \mathcal{I},\ k = 1, \ldots, N$	reference and template subdomain images (N subdomains in total), high-resolution
$\mathcal{I} \subset \mathcal{C}^0(\mathbb{R}^2)$	space of image interpolation functions
$\Omega,\ \Omega_k,\ k = 1, \ldots, N$	reference and subdomain domains
$\mathbf{u}_0 \in \mathbb{R}^{2m_1m_2}$	discrete initial displacement, acquired using a conventional nonlinear registration
$D : \mathcal{I}^2 \times \mathbb{R}^{2m_1m_2} \mapsto \mathbb{R}$	distance measure
$S : \mathbb{R}^{2m_1m_2} \mapsto \mathbb{R}$	regularizer
Output	
$\mathbf{u}_k \in \mathbb{R}^{2m_1m_2},\ k = 1, \ldots, N$	global displacement computed using high image resolution in subdomain k
$\mathbf{u}_{\text{CLG}} \in \mathbb{R}^{2m_1m_2}$	global displacement combined from all $\mathbf{u}_k,\ k = 1, \ldots, N$

for each subdomain $k = 1, \ldots, N$ where

$$T_k(\mathbf{x}_i) = \text{interpolate}(\mathbf{x}_i, \mathtt{T}_k),$$

$\mathtt{T}_0 \in \mathbb{R}^{m_1 \times m_2}$	is low-resolution image data on Ω, *and*
$\mathtt{T}_k \in \mathbb{R}^{n_1 \times n_2}$	is high-resolution image data associated with the subdomain Ω_{T_k}.

The notation is used analogously for the reference images R_0 and R_k, respectively. The Problem 4.7 is solved independently for all subdomains Ω_k, each computation is embedded in a multilevel framework and uses the result of a coarse global registration as an initial guess (Algorithm 4.2.1).

The difference to a coarse global registration is the partition of the distance measure into two discretizations: a coarse global term and an additional local data term $[D(R_k, T_k, \mathbf{u}_k)]_{\Omega_k}$. The remainder of the multilevel registration framework remains unchanged. Compared to the preliminary, purely local approach (Equation 4.1), the additional global distance term $[D(R_0, T_0, \mathbf{u}_k)]_{\Omega \setminus \Omega_k}$ is added such that (coarse) image information from the entire domain is included in the computation. In Chapter 3.2.4, we found that a coupling between the subdomains is necessary for the convergence of a domain decomposition method. The additional global data term in CLG achieves such a coupling between the local subdomain and the global optimization problem.

Algorithm 4.2.1 CLG

```
# solve a coarse registration, obtain u_0
J_0(u_0) = D(R_0, T_0, u_0) + S(u_0) --u_0--> min
for k=1, ..., K:
    # solve the nonlinear registration problem (4.7), using a
      multilevel approach and u_o as initial guess
    J_CLG(u_k) --u_k--> min
# blend local results
u_CLG = blend(u_k, k = 1, ..., N) (Chapter 5)
```

Each local registration problem results in a globally smooth displacement where the respective subdomain region has been registered using a fine image resolution. Since the optimization problems in Step 2 can be solved independently, the `for`-loop could be easily parallelized.

We note that in contrast to the preliminary subdomain-based approach, the CLG objective function can be interpreted as a different discretization of the original variational image registration problem 2.1. Each $\mathbf{u}_k$ is a global solution to a different discretization of the original large-scale image registration problem. The interpolation error of the images at a point $x \in \mathbb{R}^2$ is lowest on the corresponding subdomain $x \in \Omega_k$ and, therefore, the corresponding solution $\mathbf{u}_k$ is selected for this point. To form one global displacement, the individual solutions for each subdomain are blended in a final step. We discuss different blending strategies in Chapter 5, the simplest of which is to copy the local part of each subdomain's displacement into a global displacement without modification. This simple combination is also called "gluing" [Dol15, p. 3] and is well suited to analyze the CLG registration (4.2.4) as it does not alter the locally computed displacements.

4.2.2 Relation to Domain Decomposition Methods

The presented approach combines different ideas discussed in earlier chapters. In domain decomposition, two-level approaches that reduce the necessary communication between nodes by solving an intermediate problem on a coarse space (Chapter 3.2.4). If we interpret the CLG method as a one-step ASPIN [Cai02] (Chapter 3.2.3) variant, the coarse pre-registration is such a coarse-space step.

In the ASPIN method, independent nonlinear problems are solved locally, and the solution is applied as a preconditioner for the larger global problem. The CLG method utilizes the idea of solving local nonlinear problems. Instead of solving the excessively large global problem in a second step, we globalize the local problems by simultaneously optimizing a coarse global distance measure.

In the nonlinear Schwarz methods Newton–Schwarz and NKS (Chapters 3.2.2), intermediate results are communicated between subdomains by repetitively solving

local problems and transferring the information through boundary conditions. If these methods were executed on a single computer, the image data would need to be loaded from disk many times, which is time-consuming. In the CLG method, each local nonlinear problem is only solved once. This shortcut leads to a solution that is different from the original problem and whose properties are examined in Chapter 4.2.4. Most importantly, the local solutions differ at the subdomain boundaries. We compensate for this difference by using a blending method described in Chapter 5.

4.2.3 Memory Requirements and Computation Time

By decomposing the large-scale registration problem into independent smaller subproblems, the required memory is reduced such that the problem can be solved on a single computer. The CLG method has about twice the memory footprint of a conventional low-resolution registration. Assuming an image size of $M \times M$ pixels for both the subdomain images ($R_k,\ T_k$) and the coarse resolution global images ($R_0,\ T_0$) and a deformation grid size of $m \times m$, the occupied memory of each registration component can be found in Table 4.2. The number of pixels in the subdomain images corresponds to the coarse-resolution images plus an overlap. With this choice, the two distance terms that contribute equally in (4.7), represent a similar number of pixels. We assume an overlap of 25% of the subdomains and an additional overlap of 10% of the template subdomain. These values are used in the implementation of the experiments in Chapter 6.1.

Table 4.2: Memory requirements of the major components of a single level CLG registration. The memory growth rate of the algorithm is $\mathcal{O}(M^2 + m^2)$, where coarse and subdomain images have a size of M^2 pixels and m^2 is the number of nodes in the displacement.

Component		Number of Variables Stored
Global Distance Measure $[D(R_0, T_0, \mathbf{u}_k)]_{\Omega\setminus\Omega_k}$	R, T	M^2 each
	∇T	$2\,M^2$
	∇D	$2\,M^2$ (SSD)
		$5\,M^2$ (NGF)
Subdomain Distance Measure $[D(R_k, T_k, \mathbf{u}_k)]_{\Omega_k}$	(same as global + overlap)	$11.0\,M^2$ (SSD) / $16.7\,M^2$ (NGF)
Regularizer	$B^TB\mathbf{u}$	$2\,m^2$
Optimization	$\mathbf{u}$	$2\,m^2$
	$\mathbf{s}$	$2\,m^2$
	CG method	$4\,m^2$[Saa03, p. 200]
Total		$\approx 25.7\,M^2 + 10\,m^2$ (NGF)
		$\approx 17.0\,M^2 + 10\,m^2$ (SSD)

Due to the two distance terms that evaluate similarly sized images, the memory required to store the images and associated variables doubles compared to a coarse-resolution registration. The memory required to store the displacement and its associated variables remains unchanged. The memory growth rate of the algorithm remains $\mathcal{O}(M^2 + m^2)$, where M^2 and m^2 are the number of elements of the images and the displacement, respectively.

A single-level registration with NGF distance measure of two 10 000 × 10 000 pixel images (coarse resolution) and two equally sized subdomain images with a displacement size of 129 × 129 requires approximately 20.6 GB of memory for the components in Table 4.2.

This required memory is independent of the number of subdomains that the image domain is decomposed in. Therefore, the maximum size of the full image in the CLG method is not restricted by the available memory. However, a coarser representation of the image in $\Omega \setminus \Omega_k$ leads to a larger discretization error that will also influence the registration result inside Ω_k. One approach to minimize such an error is to construct a pyramid of nested subdomains with increasingly fine discretizations.

Most of the memory is consumed by storing the images and image-based intermediate values: The image data for images R and T is stored in vectors $\mathtt{T} \in \mathbb{R}^{100 \cdot 10^6}$, $\mathtt{R} \in \mathbb{R}^{100 \cdot 10^6}$ with $10 \cdot 10^6$ values each. For the computation of the distance measure, we store the derivative of the image ∇T in a vector of size $200 \cdot 10^6$, which contains the derivative in the first and second dimensions. We also store the non-zero partial derivatives in the NGF computation (∇D, Equation 2.9) for faster computation. For each pixel in the reference image, five values have to be stored to compute the short finite differences, resulting in $5 \cdot 100 \cdot 10^6$ values. In sum, $900 \cdot 10^6$ values are stored in the computation of the NGF distance measure of the coarse images.

For the subdomain images, the same components are added, but the image size is larger due to the overlap. The reference image is extended by 25%, the template image by an additional 10% to account for possible deformations, resulting in a 37.5% increase. In total, $1668.75 \cdot 10^6$ values are stored to compute the subdomain NGF distance measure. $0.5 \cdot 10^6$ values are used to store the low-resolution displacement grid $\mathbf{u}$, the update $\mathbf{s}$, and intermediate values in the Conjugate Gradient method [Saa03, p. 200]. Together, coarse image, subdomain image, and grid amount to $2569.27 \cdot 10^6$ values for an NGF registration. If we store all values in double precision, each value consumes 8 bytes, resulting in a total memory consumption of 20.6 GB. Without additional subdomains, only 7.2 GB of memory are required.

We note that the derivatives could also be computed on the fly, which would reduce the memory consumption in exchange for additional computations, see [Kön18] for an analysis. Furthermore, not all the values need to be stored in double precision such that memory consumption could be reduced in exchange for numerical precision. Reduced precision could potentially result in slower convergence and increased computation time. However, these measures require a re-implementation of many of the computations and—more importantly—they do not scale if the image size increases further. We therefore focus our attention on the CLG decomposition approach.

The computation time of the algorithm depends on the number of iterations of the optimizer and the duration of each step. Since computation time is not our primary interest, we settle for a rough estimate of the computation time per iteration. To compute a total computation time, we assume a fixed number of Gauß–Newton iterations the coarse registration and for each subdomain. Under these assumptions, the average computation time only depends on the sizes of images and displacement. The increase of the computation for the CLG registration compared to the coarse registration is composed of two parts: increased computation time per subdomain and increased computation time due to the loop over all subdomains.

The computationally intensive steps in each subdomain's registration are a) depending on the image size M for the computation of the derivatives of the distance measures, and b) depending on the displacement size m to compute the derivative of the regularizer and the Gauß–Newton or L-BFGS update step. In the CLG method, an additional distance term is evaluated. In the worst case, the image size M is much larger than the displacement size m, and step b) can be neglected in favor of a). In this case and assuming no additional overlap, the computation time per subdomain would be a bit more than twice as long as in a purely coarse-resolution registration.

If the subdomains are computed sequentially on one computer, the duration will increase linearly with the number of subdomains. A CLG registration with N subdomains will take on average circa $2N$ times the time of a coarse registration that uses the same image sizes.

We conclude with an example: Consider a coarse resolution image with $M \times M$ pixels and a fixed-size grid with $m \times m$ nodes. If we increase the image resolution by a factor of 4 in each direction such that the number of pixels of the fine image is $M_4 \times M_4$, $M_4 = 4M$, the required memory grows quadratically by a factor of 16. If instead the CLG registration is used, the memory requirement is only doubled. The saved memory is paid in computation time.

A precise estimate of the computation time is difficult to asses since it largely depends on the convergence of the CLG objective function. The CLG iteration starts with an initial guess from a coarse registration and only differs in the local subdomain region. In the coarse region, we do not expect large changes in the initial guess.

As an approximation of the computation time, we compare the number of computations required in one iteration. In each iteration of each subdomain, ca. 2.7 times the number of pixels have to be evaluated due to the additional overlap. Since the evaluation of the distance measure is responsible for the largest part of the computation, we assume that one iteration takes 2.7 times longer than one iteration using coarse images. Assuming an equal number of iterations for convergence for coarse and CLG registration, the additional time by iteration has to be multiplied by the number of subdomains. Under these assumptions, a CLG registration will take about $2.7 \cdot 16 = 43.2$ times longer than a coarse registration.

If we continue to assume that the computation time of the registration depends linearly on the number of pixels in the images, a fine global registration takes 16 times the time than the coarse registration in the above setting. Therefore, a lower bound

for the increase in computation time from global fine to CLG registration is 2.7.

Due to the many assumptions of which some will certainly not hold in practice, these numbers can only be a rough estimation of the computation time. In the analysis of the ASPIN method (Chapter 3.2.3), Cai and Keyes [Cai02] argue that subdomain-based approaches can converge more quickly as local nonlinearities can be computed more efficiently. While the CLG registration optimizes a global deformation, the objective function is only aware of high-resolution information in the local subdomain. Similar effects could occur here, possibly reducing the computation time for the CLG method. Since computation time is not our primary interest, we refrain from a further analysis here and refer the reader to the practical results of the experiments in Chapter 7.5.3. However, these show an increase in computation time by a factor of 12 to 27 for CLG registration, which is higher than estimated above.

A possible cause for the slower computation could be in the coarse image distance term, which is optimized along with each subdomain but which we expect to change only little during the subdomain iterations. This could have the opposite effect of the one observed in the isolated ASPIN iteration: small updates in the coarse image regions could negatively impact the line search, and increase the number of Gauß–Newton iterations.

We proposed a fixed overlap of 25% of the size of the subdomain in each dimension. Dryja and Widlund [Dry94] have found that for Schwarz–Alternating algorithms and similar, derived methods, a more extensive overlap leads to a better convergence rate. However, they find that solving the resulting local problems is more expensive if the overlap is larger, see [Dry94], and the references therein. From these results, we expect that a more extensive overlap would lead to a smaller difference of the deformation in the overlap region at the cost of a more expensive computation of each subdomain.

4.2.4 Properties of the Registration Result

The CLG method has different properties than a coarse global, fine global, and purely local subdomain-based method. Locally, the CLG method uses the same image resolution as the fine global registration. Therefore, we expect a similiar registration accuracy. The higher discretization error outside of the local subdomain may lead to an additional error. Furthermore, a mismatch of the deformations at the subdomain borders is to be expected.

We confirm these properties in numerical experiments in academic (Chapter 6.1) and realistic (Chapter 7) images.

High Resolution Accuracy

The displacements $\mathbf{u}_{\mathrm{CLG}}$ that approximates a fine global registration ($\mathbf{u}_{\mathrm{FN}}$) are expected to be superior to coarse registration ($\mathbf{u}_{\mathrm{CRS}}$) in terms of displacement error

$$\|\mathbf{u}_{\mathrm{CLG}} - \mathbf{u}_{\mathrm{FN}}\|_2 \leq \|\mathbf{u}_{\mathrm{CRS}} - \mathbf{u}_{\mathrm{FN}}\|_2$$

and distance measure evaluated on the global, full-resolution images R_F and T_F

$$D(R_F, T_F, \mathbf{u}_{\mathrm{FN}}) \leq D(R_F, T_F, \mathbf{u}_{\mathrm{CLG}}) \leq D(R_F, T_F, \mathbf{u}_{\mathrm{CRS}}).$$

Deformation Mismatch at Subdomain Boundaries

For each subdomain, the CLG registration computes a global solution that exploits the local high-resolution image data to increase local accuracy. Since the displacement for each subdomain is computed independently, a mismatch at the subdomain boundaries can occur. This effect is especially prominent if the image contains low-contrast areas without structuring image information. This kind of mismatch can be observed in a purely local subdomain registration but also—less prominently—in the new method.

Therefore, we expect a lower irregularity measure S

$$S(\mathbf{u}_{\mathrm{CLG}}) \leq S(\mathbf{u}_{\mathrm{PTCH}})$$

in the displacements after CLG registration than after a purely subdomain-based registration ($\mathbf{u}_{\mathrm{PTCH}}$).

The mismatch can be reduced by letting subdomains overlap, as shown in Figure 4.1.1, such that a larger portion of the image data is shared between the subdomains. The remaining mismatch is compensated by using a suitable blending strategy, as described in Chapter 5.

5 Blending of Local Deformations

Contents

During CLG and purely local subdomain image registration, the local registration problems are solved independently, such that the individual solutions may differ in the overlap region. In this chapter, we propose to use a blending method that incorporates all local solutions into one global solution. To avoid re-accessing the image data, we use blending as a pure post-processing step that acts on the deformation only.

Based on the CLG registration scheme, we require a blending method to have the following properties.

1. The blending should preserve the homogeneity of neighboring solutions that occurs if high-contrast image information is available in the overlap.
2. If a mismatch between the deformations occurs at the subdomain boundaries, the method should provide a smooth solution in terms of irregularity.
3. The overall goal for the deformation is to be close to the optimal (fine global) solution.

Previously proposed methods to combine independently computed deformations either rely solely on the pre-computed deformation or also use image data. In the first case, the disagreeing deformations are replaced by smoothly filling the disagreeing region based on values from its vicinity [Mod10; Mue12] or multiple global deformations are combined by averaging [For11]. In the second strategy, the authors compute an optimal deformation with respect to image distance and smoothness [Lem08; Hei16]. We review these approaches in Chapter 3.3.

While the first strategy can be directly applied to the blending problem, the inclusion of image distance requires the image data to be re-loaded from disk, which we want to avoid given the size of the dataset.

We therefore compare two methods: The first approach uses a weighted sum of the overlapping displacements where the weights depend on the distance to the subdomain boundary and can be interpreted as a modification of the averaging approach in [For11] with different weights. The second approach, local regularization, is similar to the *UnfoldReg* approach [Mue12], except that we use the curvature regularization

as smoother instead of resorting to B-Splines. Our regularization approach reuses the regularizer from the original registration problem and is implemented as a constraint optimization problem. The result, therefore, shares the physical smoothness properties with the original registration result, which does not apply when using B-Splines independently of the registration problem.

In addition, we combine the two approaches to balance irregularity and distance to the local solutions. We also propose a new global blending approach that produces results with very low irregularity.

We demonstrate the different blending methods with respect to the above goals using experiments based on a purely local subdomain registration result (Figure 5.1.1, referred to as "Deformation 1") as these show a relatively large mismatch in a low contrast region from a realistic scenario. In Chapter 6.2, experiments with additional deformations (Deformation 2 and 3) that result from CLG registrations are studied. We conclude in Chapter 6.2.4 and recommend the local regularization approach for blending with user input and the weighted sum approach as a fully automatic method.

5.1 Combining Deformations without Blending

Without using a blending approach, a straightforward way of computing a global solution from multiple local solutions is to restrict each local deformation to a non-overlapping subdomain. The global deformation is then just the ensemble of the partitions.

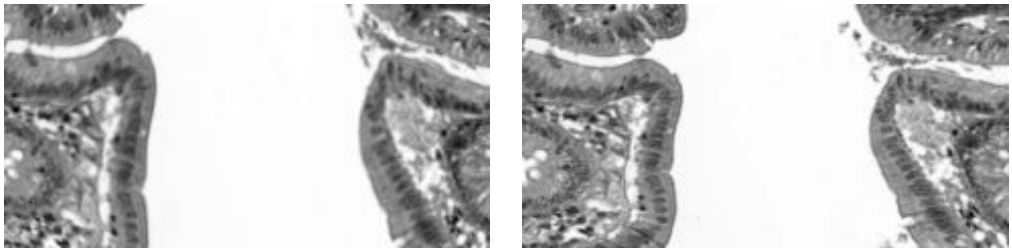
Formally, we write the global combined solution $\mathbf{u}_{\text{global}} \in \mathbb{R}^{2m_1m_2}$ based on the local solutions $\mathbf{u}_k \in \mathbb{R}^{2m_1m_2}$ as

$$\mathbf{u}_{\text{global}} = \sum_{k=1}^{N} \mathbf{w}_k \cdot \mathbf{u}_k, \text{ where } \mathbf{w}_{ki} = \begin{cases} 1 & \text{if } \mathbf{x}_i \in \Omega_k^< \\ 0 & \text{otherwise} \end{cases}$$

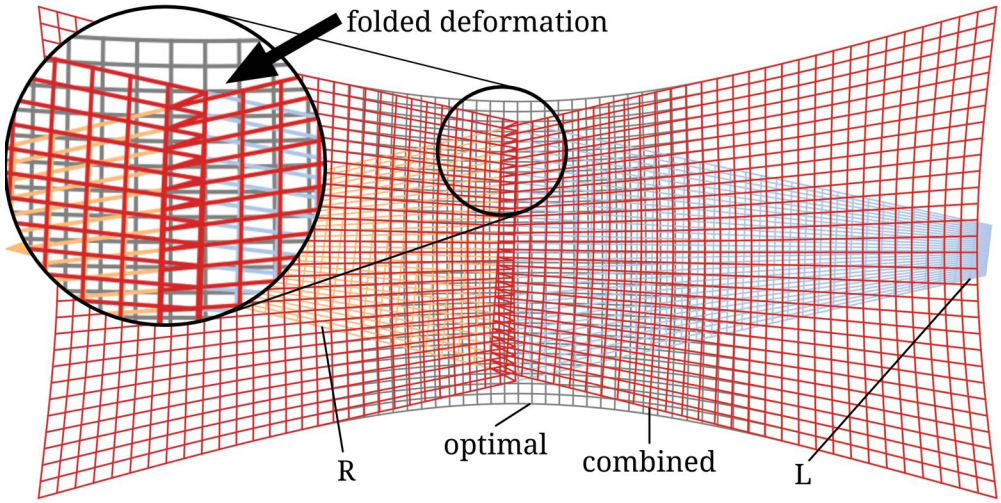
and $\mathbf{w}_{ki}$ is the i-th element of $\mathbf{w}_k \in \mathbb{R}^{2m_1m_2}$. The vector $\mathbf{w}_k$ is the discretized characteristic function of the non-overlapping subdomain $\Omega_k^<$ and "$\cdot$" is the element-wise multiplication of the two vectors. The non-overlapping subdomain $\Omega_k^<$ can be derived by shrinking $\Omega_k = [\omega_k^1, \omega_k^2] \times [\omega_k^3, \omega_k^4]$ to

$$\Omega_k^< = [\omega_k^1 + \epsilon_x, \omega_k^2 - \epsilon_x] \times [\omega_k^3 + \epsilon_y, \omega_k^4 - \epsilon_y]$$

as shown in the bottom of Figure 5.2.1 in the one-dimensional case. The shrinkage parameter ϵ_x, ϵ_y are chosen such that the subdomains $\Omega_k^<,\ k = 1, \ldots, N$ with



Reference R Template T



purely local subdomain registration result, combined without blending

Figure 5.1.1: Deformation 1. Image pair with a low-contrast region after purely local subdomain registration. Top: Reference and template image. Bottom: Combined solution without blending (red) plotted over local subdomain solutions from the left (L, blue) and right (R, orange) subdomain and the optimal solution $\mathbf{y}_{\text{fine}}$ (gray) from a fine global registration. The local and optimal solutions are only shown where different from the combined solution. The difference in the local results (L, R) leads to a folded deformation in the overlap region in the center of the image. Also, a difference between the combined deformation and the optimal registration result can be observed. The goal of the blending approach is to form a smooth deformation by removing the artifacts while maintaining the low displacement error inside the subdomains.

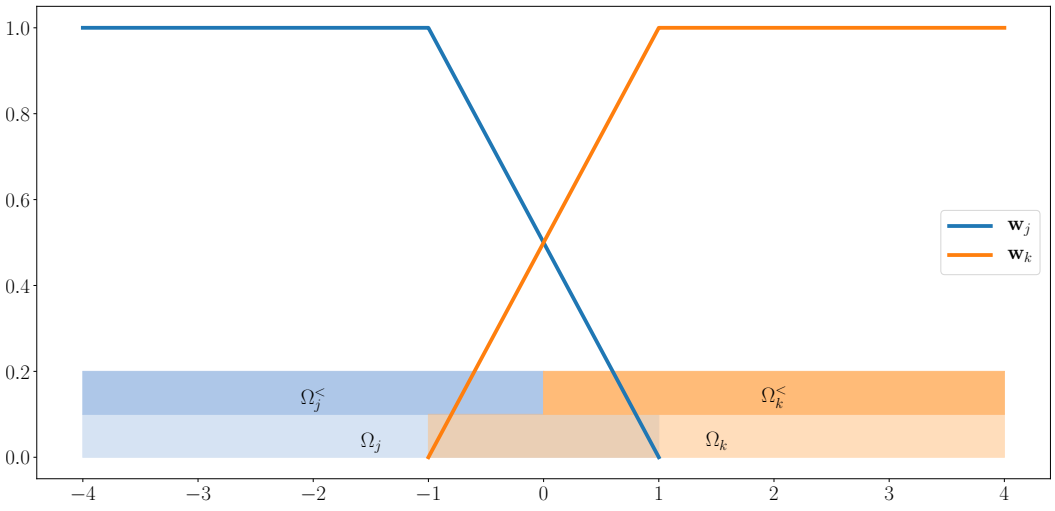


Figure 5.2.1: Blending by weighted sum. One-dimensional projection of the overlap of two subdomain domains $\Omega_j = [-4, 1]$ and $\Omega_k = [-1, 4]$. The contribution to the global solution to each subdomain is determined by the respective weight vectors $\mathbf{w}_j$ and $\mathbf{w}_k$.

$$\int_{\Omega_j^< \cap \Omega_k^<} \mathrm{d}A = 0 \qquad \forall j \neq k,\ 1 \leq j, k \leq N \text{ and}$$

$$\bigcup_{k=1}^{N} \Omega_k^< = \Omega$$

only overlap at the boundary but still cover the whole domain Ω. We use square subdomains and choose the discretization such that subdomain boundaries of Ω_k and $\Omega_K^<$ are always located in between two points of the deformation grid with equal distance to the neighboring grid points.

The disadvantage of this method becomes apparent in the example in Figure 5.1.1, where the results of the purely local subdomain registration are combined. If there is a mismatch of the deformation at the subdomain boundary, the combined deformation is no longer smooth. Letting the subdomains overlap during registration, as in the experiments, can reduce the discrepancy of the deformation. However, the reduction depends on the structural information in the overlapping area.

5.2 Blending by Weighted Sum

The first proposed blending approach computes a weighted sum of the displacements in the overlapping region. The weights are based on the distance to the subdomain border (Figure 5.2.1).

We first consider the one-dimensional problem of fusing the subdomains in one row. The joint deformation $\mathbf{u}_\rho$ of one row of subdomains is computed as the weighted sum

$$\mathbf{u}_\rho = \sum_{k \in \rho} \mathbf{w}_k \cdot \mathbf{u}_k$$

of the deformations $\mathbf{u}_k$ where $k \in \varrho$ are the indices of the respective subdomains. With the subdomains in Figure 4.1.1, the first row consists of subdomains $\Omega_1, \Omega_2, \Omega_3, \ldots$ and $\rho = \{1, 2, 3, \ldots\}$ are this rows' indices.

As introduced in Chapter 2.5.1 (Figure 2.5.1), we use the notation

$$(\mathbf{w})_i = w_i, w_{i+m_1 m_2}$$

to address the x- and y-coordinate of the weight stored in the entries i-th and $i+m_1m_2$ of the vector $\mathbf{w} \in \mathbb{R}^{2m_1m_2}$. The node i of the displacement corresponds to the grid coordinate $\mathbf{x}_i$.

The weight vector $\mathbf{w}_k \in \mathbb{R}^{2m_1m_2}$ for subdomain k is defined as

$$(\mathbf{w}_k)_i = 0 \qquad \text{if } \mathbf{x}_i \notin \Omega_k \tag{5.1}$$

$$(\mathbf{w}_k)_i = 1 \qquad \text{if } \mathbf{x}_i \in \Omega_k \setminus \Omega_j \text{ for all } j \neq k \tag{5.2}$$

$$(\mathbf{w}_k)_i = \frac{\operatorname{dist}(\mathbf{x}_i, \partial\Omega_k)}{d_{k,j}} \qquad \text{if } \mathbf{x}_i \in \Omega_k \cup \Omega_j \text{ for any } j \tag{5.3}$$

where the distance $\frac{\operatorname{dist}(\mathbf{x}_i, \partial\Omega_k)}{d_{k,j}}$ with

$$\operatorname{dist}(\mathbf{x}, M) = \inf\{\|\mathbf{x} - \mathbf{z}\|_2^2,\ \mathbf{z} \in M\} \qquad \text{and} \tag{5.4}$$

$$d_{k,j} = \sup\{\operatorname{dist}(x, \partial\Omega_k),\ x \in \Omega_k \cap \Omega_j\}$$

is the relative distance to the non-overlapping part of Ω_k inside the overlap. All choices for the $\mathbf{w}_k$ have to fulfill the partition of unity [Ban00, p. 401] condition

$$\sum_{k=1}^{K} \mathbf{w}_k = \underline{1},$$

where $\underline{1} \in \mathbb{R}^{2m_1m_2}$ is the vector of ones.

While the above piecewise linear weight function is continuous, its derivative is discontinuous on the boundaries $\partial\Omega_k, k = 1, \ldots, K$. If smoothness of the blending function is required, the distance function (5.4) can be embedded into trigonometric or B-Spline [dBoo01, pp. 87 ff.] basis functions as reported in [Rok99; Oht03] for blending

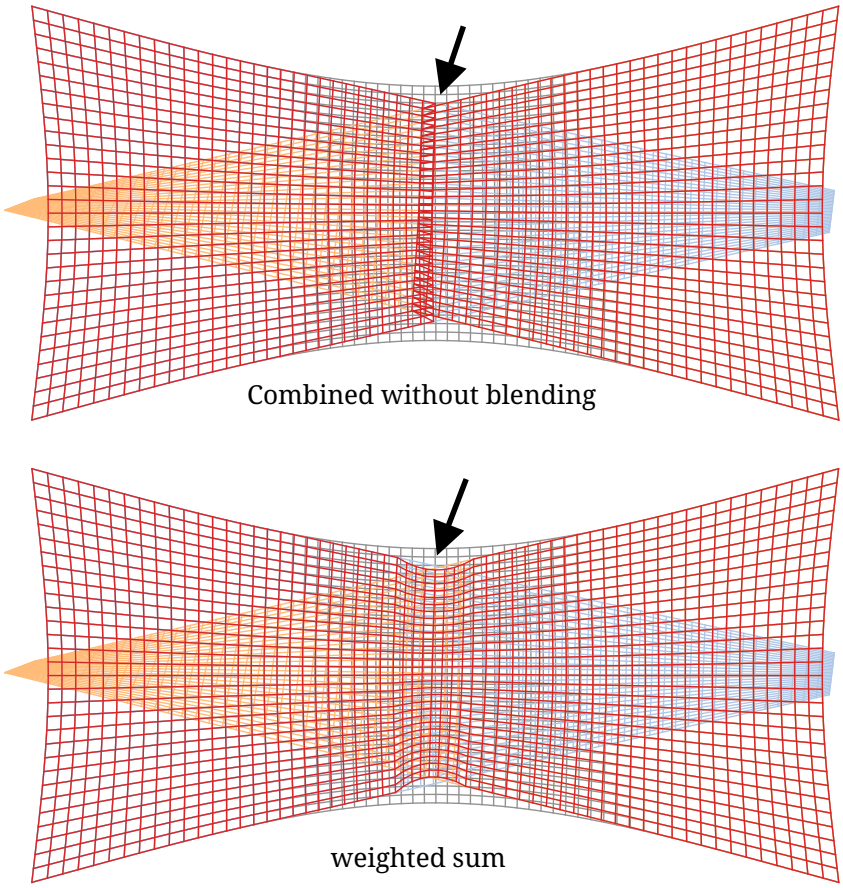


Figure 5.2.2: Deformation 1 from Figure 5.1.1, combined without blending (top) and blending by weighted mean (bottom). Blended solution (red) plotted over local subdomain solutions (blue, orange) and optimal solution $\mathbf{y}_{\text{fine}}$ (gray).

local solutions of decomposed partial differential equations. We use the piecewise linear basis function for its simplicity in our experiments.

The above computation is repeated for all rows of subdomains. To obtain a result in two dimensions, the weighted sum of the overlap between the rows is computed in the same way in a second step. The resulting deformation is visibly smoother than the glued solution and no foldings are detected in the experiment (Figure 5.2.2). However, a discrepancy to the optimal global solution is clearly visible.

We originally proposed this weighted averaging approach to be used with the purely local subdomain registration [Lot16a].

5.3 Blending by Regularization

The goal of a blending strategy is to reduce the irregularity in the final solution. In the computation of the weighted sum approach, a smoother deformation can be computed from two independent deformations, given that both of them are smooth. Extending the work in [Mod10; Mue12], we propose to increase regularity in the resulting global deformation by explicitly formulating the requirement for smoothness in the computation of the final solution.

Global Regularization

As an intermediate step towards a locally regularizing blending method, we first establish a method to blend the deformation using global regularization. Similarly to the variational image registration algorithm described in Chapter 2, the proposed global blending balances irregularity and distance to the local subdomain displacements.

To find the blended solution, we define an objective function G_{global} as

$$G_{\text{global}}(\mathbf{u}) = \mathbf{u}^T B^T B \mathbf{u} + \beta \sum_{k=1}^{N} (\mathbf{u}_k - \mathbf{u})^T I_{\mathbf{w}_k} (\mathbf{u}_k - \mathbf{u}) \rightarrow \min \tag{5.5}$$

where $I_{\mathbf{w}_k} \in \mathbb{R}^{2m_1m_2 \times 2m_1m_2}$ is a diagonal matrix that restricts the influence of $\mathbf{u}_k$ to the domain Ω_k. The diagonal of $I_{\mathbf{w}_k} = \text{diag}(\sqrt{\mathbf{w}}_k)$ can be any of the previous choices for $\mathbf{w}_k$. The parameter β is chosen manually, based on the appearance of the resulting deformed grid and its irregularity. The parameter is used to determine the compromise between regularization and goodness of fit to the local subdomain solutions.

In order to find the optimal global deformation, we compute the first derivative of G_{global} with respect to $\mathbf{u}$

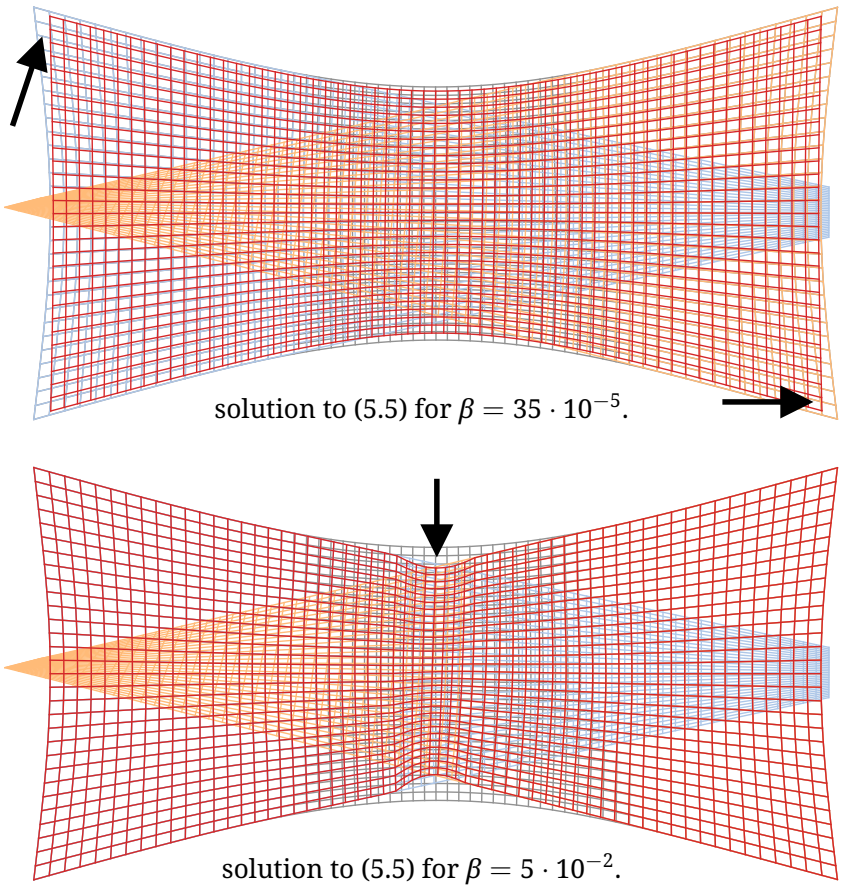


Figure 5.3.1: Deformation 1 from Figure 5.1.1 after blending by global regularization for parameters $\beta = 35 \cdot 10^{-5}$ and $\beta = 5 \cdot 10^{-2}$. Blended solution (red) plotted over local subdomain solutions (blue, orange) and optimal solution $\mathbf{y}_{\text{fine}}$ (gray). Global regularization modifies the deformation outside of the overlap region. Depending on the parameter β, low irregularities can be reached.

$$\nabla G_{\text{global}}(\mathbf{u}) = 2B^T B\mathbf{u} + \beta \sum_{k=1}^{N} I_{\mathbf{w}_k}(\mathbf{u}_k - \mathbf{u})$$

and solve the resulting linear system

$$\left(2B^T B + \beta \sum_{k=1}^{N} I_{\mathbf{w}_k}\right)\mathbf{u} = \beta \sum_{k=1}^{N} I_{\mathbf{w}_k}\mathbf{u}_k.$$

We note that minimizing G is computationally less expensive then solving the registration problem in (4.7) because no large-scale image data is involved in the computation.

For illustration, Figure 5.3.1 shows the blended solutions for two different values of the parameter β. Both have been determined by a parameter search optimizing a) the displacement error and b) balancing displacement error and SSD in the resulting image pair.

The parameter $\beta = 3.5 \cdot 10^{-5}$ yields the smallest displacement error with respect to the optimal solution calculated from a global registration. However, the regularization affects the entire domain, which is undesirable as the confidence in the registration in the central subdomain area is high.

This problem is also manifested in a high SSD value for this case. The second parameter $\beta = 0.2$ results in a global solution that is closer to the subdomain solutions. It was determined by starting with a large value for β and relaxing the weight on the displacement data term until the overall SSD is increased by 10% over the solution of the purely local subdomain registration with blending by weighted sum. We note that these strategies can only be used when an optimal solution is known.

We have used this global regularization approach in combination with a purely local subdomain registration of differently stained whole slide images [Bul19].

Local Regularization

Global regularization as a blending strategy has the drawback that the deformation is also altered outside the overlapping region. Since no image data is used in the blending step, we prefer to preserve the local registration result. When registering subdomains, the confidence in the local solution is higher inside the subdomain since the influence of missing remote subdomains decreases with spatial distance (Chapter 2.8).

The idea of local regularization is similar to earlier approaches [Mue12], where B-Spline interpolation is used to fill regions in which foldings corrupt the deformation.

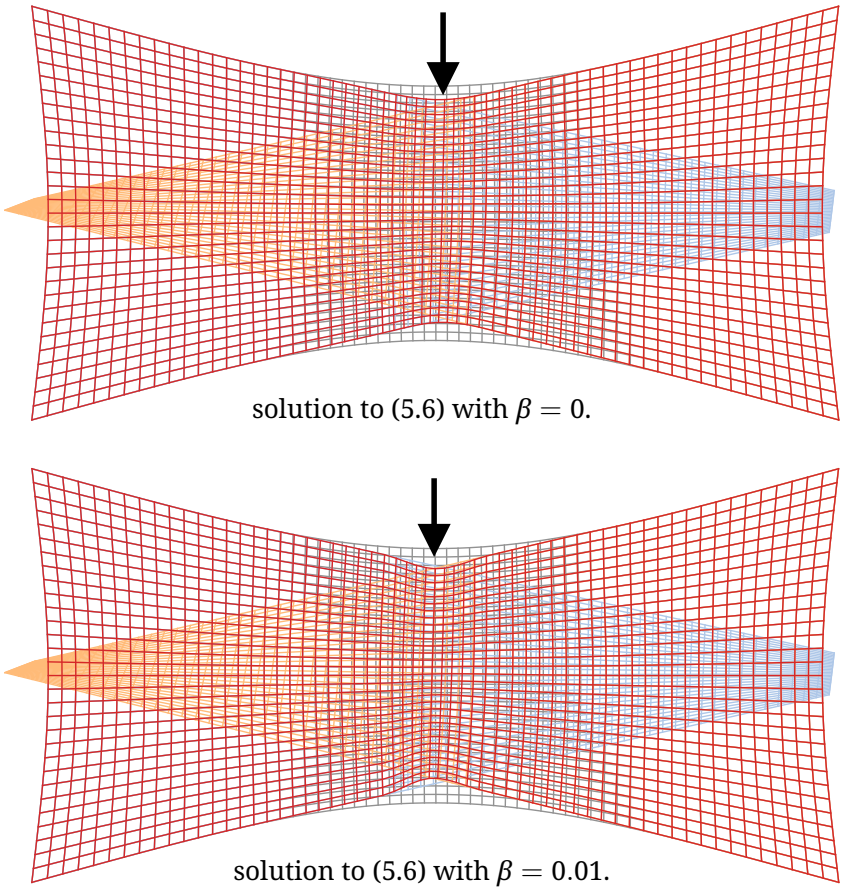


Figure 5.3.2: Deformation 1 from Figure 5.1.1 after blending by local regularization with (top) and without (bottom) local data term. Blended solution (red) plotted over local subdomain solutions (blue, orange) and optimal solution $\mathbf{y}_{\text{fine}}$ (gray).

We extend their approach with the weighted sum from Chapter 5.2. The local objective function

$$\begin{aligned} G_{\text{local}}(\mathbf{u}) =& \mathbf{u}^T B^T B \mathbf{u} \\ &+ \beta \sum_{k=1}^{N} (\mathbf{u} - \mathbf{u}_k)^T I_{\mathbf{w}_k} (\mathbf{u} - \mathbf{u}_k) \rightarrow \min \\ \text{s.t. } \mathbf{u}_i =& (\mathbf{u}_k)_i \quad \text{if } \mathbf{x}_i \in \Omega_k \setminus \Omega_j, \forall\, k, j \text{ with } k \neq j \end{aligned} \tag{5.6}$$

aims to compute a compromise between regularity and goodness of fit to the local subdomain solutions in the overlap $\Omega_k \cap \Omega_j$. The additional constraint fixates the deformation in the inner subdomain region where the subdomains do not overlap. In the overlap region, the parameter β can be adjusted to balance regularity and goodness of fit to the weighted solution from the subdomains.

We compute the solution to (5.6) by re-formulating the equality constraints as Lagrange multipliers [Noc06, pp. 304 ff.] and solving the resulting symmetric system using a Cholesky factorization [Gol96, p. 144] through Julia's backslash polyalgorithm[1] [Bez17].

We re-write the system (5.6) as

$$\begin{aligned} J_{\text{local}}(\mathbf{u}) =& \mathbf{u}^T B^T B \mathbf{u} \\ &+ \beta \sum_{k=1}^{N} (\mathbf{u} - \mathbf{u}_k)^T I_{\mathbf{w}_k} (\mathbf{u} - \mathbf{u}_k) \\ &+ \lambda^T I_\lambda \left(\mathbf{u} - \sum_{k=1}^{N} I_{\mathbf{w}_k} \mathbf{u}_k\right) \end{aligned} \tag{5.7}$$

where $\lambda \in \mathbb{R}^{\hat{m}}$ and $I_\lambda \in \mathbb{R}^{\hat{m} \times 2m_1 m_2}$ is the projector onto the $\frac{\hat{m}}{2}$ grid points associated with the non-overlapping part of the grid. The matrix I_λ contains the rows i and $i + m_1 m_2$ of the identity $I \in \mathbb{R}^{2m_1 m_2 \times 2m_1 m_2}$, for which the corresponding grid point

$$\mathbf{x}_i \in \Omega_k \setminus \Omega_j, \forall\, k, j \text{ with } k \neq j$$

is located in the non-overlapping part of a subdomain $k = 1, \ldots, N$. The solution to (5.6) is computed by solving the resulting linear system

$$\begin{bmatrix} A & I_\lambda^T \\ I_\lambda & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \lambda \end{bmatrix} = \begin{bmatrix} \beta \sum_{k=1}^{N} \mathbf{u}_k \\ I_\lambda \sum_{k=1}^{N} I_{\mathbf{w}_k} \mathbf{u}_k \end{bmatrix}$$

where $A = B^T B + \beta \sum_{k=1}^{N} I_{\mathbf{w}_k}$ for $\mathbf{u}$ and λ.

The choice of the parameter β determines the regularity of the solution. Setting $\beta = 0$ ignores the displacement data term and applies a pure regularization in the overlap

[1] Julia's backslash polyalgorithm can be found in the official source code repository at https://github.com/JuliaLang/julia/blob/v1.2.0/stdlib/SparseArrays/src/linalg.jl#L1358 (accessed 25.11.2019)

region. This approach resembles an earlier method [Mue12] replacing B-Splines with a deformation model matching the registration's regularizer. In the absence of image data, smoothly filling the overlapping area based on the values from its vicinity is the best choice in the example shown in Figure 5.3.2, top. For higher values of β, the solution gets closer to the one obtained by the arithmetic mean. In the second example (Figure 5.3.2, bottom) the solution with an arbitrary $\beta = 0.01$ is shown to illustrate the effect of the additional displacement term.

In the following Chapter 6.2, we compare the presented blending approaches quantitatively in three experiments.

6 Results of Numerical Experiments

Contents

6.1 Numerical Experiments for CLG Registration

The combined local and global registration produces results that are different from fine global, coarse global, and from a purely subdomain-based registration. In the following chapter, we focus on these properties and perform experiments to illustrate these properties. We further apply the new method to large-scale whole slide images as they occur in digital pathology in Chapter 7.

We construct different test images T based on histological whole slide images and artificial image data (Table 6.1). The corresponding second image of an image pair is computed based on a generated true displacement $\mathbf{u}_T$ (Chapter 6.1.1) such that

$$R(\mathbf{x}) = T(\mathbf{u}_T + \mathbf{x}).$$

We then register these data sets using different algorithms and compare the resulting registration results using three criteria: image distance, regularity of the deformation, and difference to the optimal deformation.

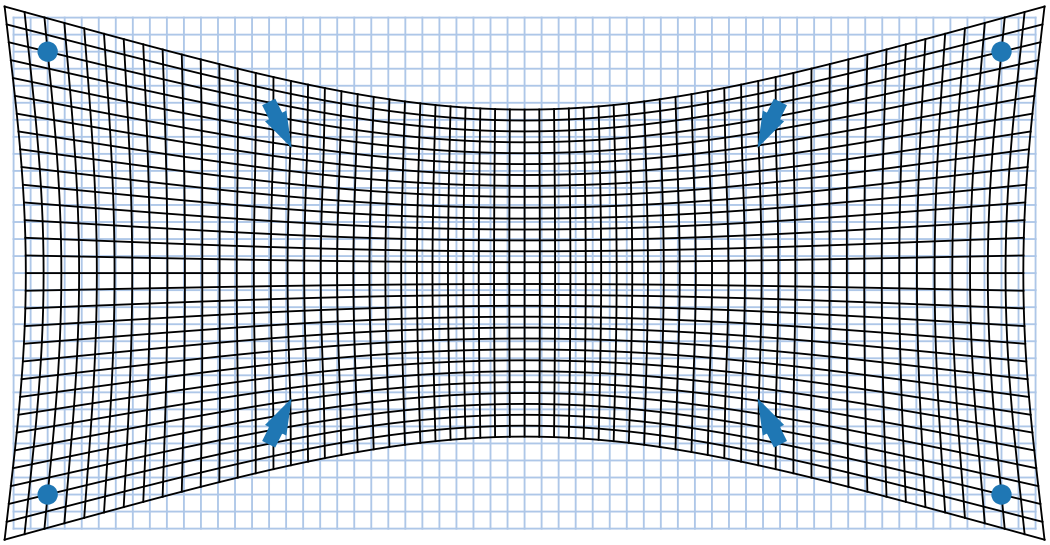


Figure 6.1.1: True deformation applied to the test images in Figure 6.1.2 to form the reference images. The artificial deformation (black) has minimal energy with respect to the regularizer given points with no deformation (circles) and pre-determined displacement (arrows). An undeformed grid (light blue) is shown for comparison.

6.1.1 Mathematical Setup of the Deformation

The artificial true deformation is generated by specifying a desired movement for some points of the displacement grid and computing a smooth interpolation between these points with minimal curvature energy.

Generating images using an artificial displacement has the following advantages: First, a true one-to-one correspondence between template and reference image can be guaranteed, and the optimal image distance between reference and deformed template image is zero. In real-world registration tasks, reference and template images are either made at different points in time or with different modalities such that both images differ even under optimal transformation. Second, the artificial displacement can be used to determine the parameters of the registration algorithm such that its solution is as close as possible to the true displacement.

In the following experiments, we construct a displacement that enlarges the central region in the template image. Therefore, we constrain four outer grid points to remain at their original positions and four inner grid points to move towards the center of the domain (Figure 6.1.1).

A smooth displacement $\mathbf{u}_T \in \mathbb{R}^{2m_1m_2}$ is then computed by formulating an optimization problem with the curvature regularization energy as objective function and requiring the displacement of the points above as constraints following the general idea in [Fis03a; Hab09]. The resulting optimization problem can be written in a La-

grangian formulation [Noc06, pp. 304 ff.]

$$\mathcal{L}(\mathbf{u}_\mathrm{T}, \lambda) = \frac{1}{2}\mathbf{u}_\mathrm{T}^T B^T B \mathbf{u}_\mathrm{T} + \sum_{i=1}^{N} \lambda_i (\mathbf{l}_i^T \mathbf{u}_\mathrm{T} - d_i)$$

where the vector $\mathbf{l}_i \in \mathbb{R}^{2m_1 m_2}$ encodes the position of the i-th constraint coordinate in linear coordinates in a vector

$$\mathbf{l}_i = (0, \quad ..., 0, \quad \underset{\uparrow}{1} \quad , 0, ..., 0)^T$$
$$\xi$$

that is one at exactly one position ξ and zero otherwise.

The displacement corresponding to the position determined in $\mathbf{l}_i$ is stored in d_i such that the expression $\mathbf{l}_i^T \mathbf{u}_\mathrm{T} - d_i$ is zero if and only if the displacement matches the i-th constraint. The first-order optimality conditions of the constraint optimization problem (6.1.1) (KKT conditions, see [Noc06, p. 321])

$$\nabla \mathcal{L}(\mathbf{u}_\mathrm{T}, \lambda) = 0$$
$$\mathbf{l}_i^T \mathbf{u}_\mathrm{T} - d_i = 0$$

result in the linear system

$$L\hat{\mathbf{u}} = \mathbf{r}, \qquad \text{where}$$
$$L = \begin{pmatrix} B^T B & \mathbf{l}_1, ..., \mathbf{l}_N \\ (\mathbf{l}_1, ..., \mathbf{l}_N)^T & 0 \end{pmatrix}$$
$$\hat{\mathbf{u}} = (\mathbf{u}, \lambda_1, \ldots, \lambda_N)$$
$$\mathbf{r} = (0, ..., 0, \mathbf{d})^T$$

and B is the regularizer matrix of the curvature regularizer (Section 2.5.4). The system has a unique solution if the intersection

$$\mathcal{A} \cap \mathcal{B}, \text{ with}$$
$$\mathcal{A} = \{\mathbf{u} \,|\, l_i^T \mathbf{u} - d_i = 0, \, \forall i\}$$
$$\mathcal{B} = \{\mathbf{u} \,|\, \mathbf{u}^T B^T B \mathbf{u} = 0\}$$

of the admissible solutions $\mathcal{A}$ of the constraints and the null space of $B^T B$, $\mathcal{B}$, has one or no element. The null-space of the curvature regularizer used in this experiment contains the linear functions. In other words, a unique solution can be found, if a) the constraints can be fulfilled by exactly one linear function or b) no linear function fulfills the constraints. In the above example, no linear function matches the constraints, and the minimum energy solution with respect to the curvature regularizer is unique.

The resulting deformation $\mathbf{y} = \mathbf{u}_\mathrm{T} + \mathbf{x}$ has minimal curvature energy with respect to the constraints (Figure 6.1.1). This "true" deformation is used to compute the reference image $R = T(\mathbf{y})$ based on a given template image T. Three different template images are used to highlight the different properties of the respective registration methods.

6.1.2 Test Image Data

The test images are taken from a whole slide image of human colon tissue (Figure 6.1.2). An initial guess of the deformation is computed by a coarse preregistration. The image sizes are summarized in Table 6.1.

The image domain Ω of the reference image R is partitioned into two overlapping subdomains, one on the left side of the image and one on the right side.

Table 6.1: Sizes of the image data and the deformation used in the experimental test cases.

Object	Size
reference image data `R`	120 × 240 pixels
template image data `T`	240 × 480 pixels (cropped to 120 × 240)
coarse (downsampled) image data	$\frac{1}{4}$ of the original resolution
discretized true deformation	31 × 61 grid points

Case 1: Realistic Image Snippet From a Stained Histology Image Containing a Low-Contrast Region

The first test image pair (Figure 6.1.3) consists of a cutout of an H&E stained whole slide image of the human colon. The original color image is converted to gray before processing. The lumen in the center is surrounded by mucous tissue in the left and right parts of the image.

The image is chosen to observe the effect of global regularization in a low-contrast image region under realistic conditions.

Case 2: Realistic Image Snippet From a Stained Histology Image Without a Low-Contrast Region

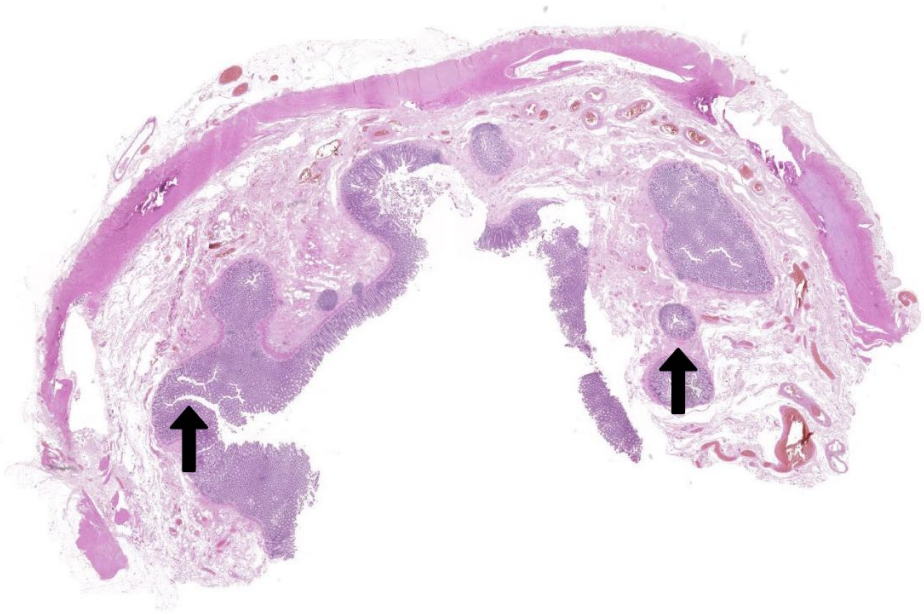
The second test image pair (Figure 6.1.4) shows an area close to the inside of the colon wall from the same slide as the image in Case 1. In contrast to the first case, no low-contrast regions are visible in the selected area.

The image is chosen to observe the dominating effect of the image distance if sufficient contrast is present.

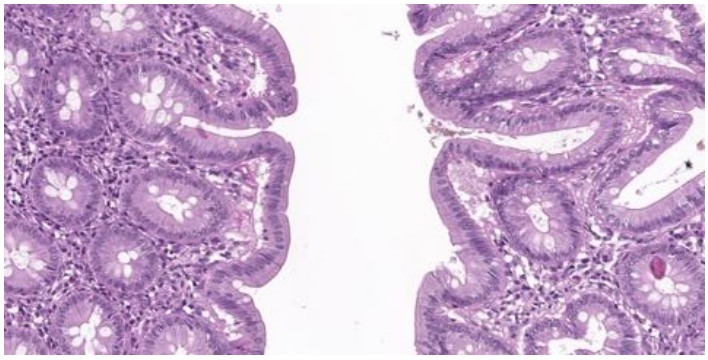
6.1.3 Evaluation Criteria

For each method we compute the deformation error

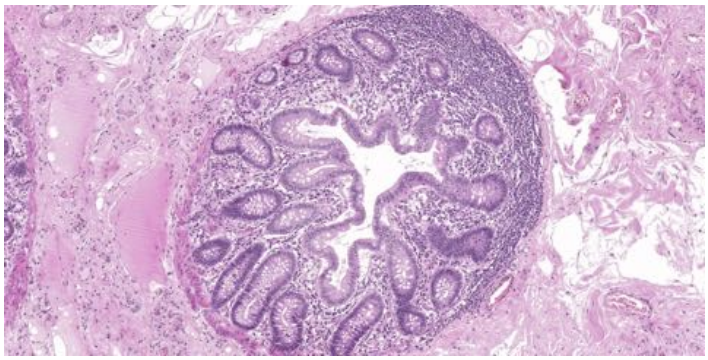
$$e(\mathbf{y}) = \|\mathbf{y} - \mathbf{y}_{\text{fine}}\|_2$$



Full Whole Slide Image



Case 1



Case 2

Figure 6.1.2: The test images are taken from a whole slide image of human colon tissue (top). The positions of the test images are highlighted. Case 1 (left arrow): Part of an H&E stained whole slide image with a low-contrast region. Case 2 (right arrow): Part of an H&E stained whole slide image without low-contrast region. The whole slide image has been kindly provided by Dr. Andreas Turzynski, Lübeck.

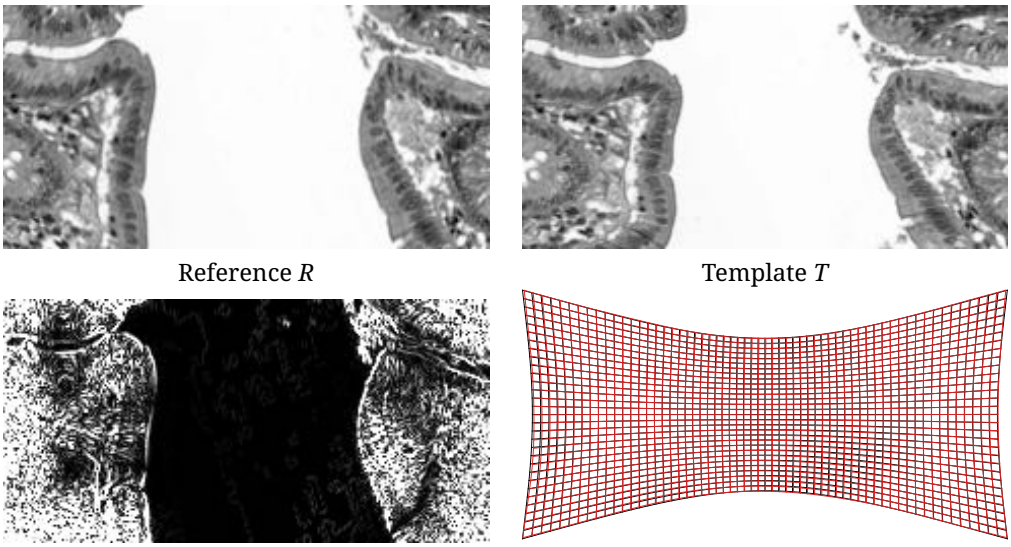


Figure 6.1.3: Case 1: Initial deformation after coarse pre-registration. Top row: original images be for registration. Bottom row: Difference image after coarse registration (left), overlay of deformed grid after coarse and fine registration.

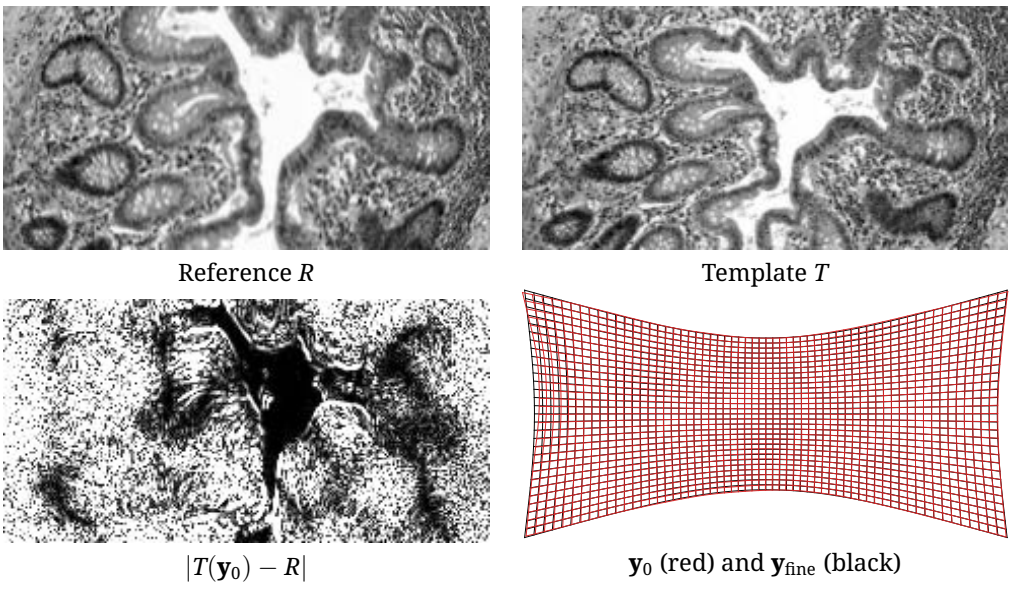


Figure 6.1.4: Case 2: Initial deformation after coarse pre-registration. Top row: original images be for registration. Bottom row: Difference image after coarse registration (left), overlay of deformed grid after coarse and fine registration.

with respect to the fine global registration. As error measure, we evaluate the overall distance to the true deformation in the Euclidean norm. Another choice with a stronger emphasis on localized errors and outliers would be the uniform norm $\|\cdot\|_\infty$. However, in $\|\cdot\|_\infty$, smaller global errors produce relatively smaller values. The computation of a landmark error, as in Chapter 7, can be used to focus on specific locations or if no true deformation is available.

From Chapter 2.5, we compute the irregularity measure

$$s(\mathbf{u}) = \|B\mathbf{u}\|_2^2$$

of the displacement to evaluate its smoothness and the remaining image distance

$$d(\mathbf{y}) = D_{\text{SSD}}(R, T, P\mathbf{y})$$

computed at the high-resolution images to evaluate the remaining image difference available to drive the optimization.

The true deformation that has been used to construct the test images results in an alignment with zero remaining image distance. However, the solution to the registration problem is not identical to the true deformation. While the second image is generated by deforming the first one, not all information from the displacement is stored in the deformed image. Due to the added regularization term, the true deformation will not result in the minimal value of the objective function.

For the evaluation, it cannot be expected that the true deformation is reconstructed exactly since the reconstruction is always based on incomplete information. Consider the extreme case of two black images: Here, the identity-deformation would be optimal with $J(\mathbf{x}) = 0$, independent of the true deformation. The advantages of the used evaluation outweigh its limitations: First, there is a true one-to-one correspondence between template and reference image such that the image distance can be compared to the possibility of a perfect match. Second, the resulting deformation can be compared to the underlying true deformation, and it can be quantified how well the missing information is compensated by the regularization model.

For an example of the disparity between true and optimal deformation, consider the example of experiment Case 1 (Table 6.2). The true deformation results in zero distance $D(\mathbf{y}_{\text{true}}) = 0$ and an irregularity of $S(\mathbf{u}_{\text{true}}) = 0.26$ (with $\mathbf{y} = \mathbf{u} + \mathbf{x}$). Due to the regularization weight of $\alpha = 2.7$, the objective function value of the true deformation is 0.70. The objective function value after fine global registration is

$$\begin{aligned} D_{\text{SSD}}(\mathbf{y}_{\text{fine}}) + \alpha S_{\text{curv}}(\mathbf{u}_{\text{fine}}) &= \\ &= 0.003 + 2.7 \cdot 0.256 = 0.69 < 0.70 \end{aligned}$$

and lower (98.6 %) then the true deformation. Visually, the differences in the two deformations is unnoticeable in the experiments (Figures 6.1.5).

We replicate the true deformation as closely as possible by determining the regularization parameter α accordingly for each method. Since all other methods are just coarser discretizations of the fine global registration, we use the result of the fine registration as the optimal deformation to which the other three methods are then compared to.

6.1.4 Evaluation

Compared Methods

For each of the test cases described, the four registrations methods

- coarse (pre-) registration
- fine global registration
- purely local subdomain registration
- combined local and global (CLG) registration

from Chapter 4 are applied.

In each of the registration methods, the regularization parameter α balances distance measure (D) and regularizer (S). The distance measure is discretized with different image resolutions in each method, potentially altering its value systematically. For each of the examples below, the optimal parameter α is determined by the parameter search described in Chapter 2.4. The parameter is determined independently for each example and each method.

For the CLG and purely local subdomain registration results, the individual subdomain deformations are cut in the middle of the overlap and combined without blending. This simple method is chosen to demonstrate the differences between the methods. More sophisticated blending methods are discussed in Chapter 5.

Case 1: Realistic Image Snippet From a Stained Histology Image

Experiment Case 1 highlights the advantages of the new approach in low-contrast image regions. Comparing the fine global registration to the true deformation (Table 6.2), we note that the fine global solution is close to the true deformation and that the difference in terms of the error-norm is small compared to the other registrations. The other methods are compared to the result of the fine global registration.

As expected, the coarse registration is less accurate than the fine registration in terms of both deformation error and SSD. The coarse registration reaches an acceptable result considering the low image resolution of only 30×60 pixels (Figure 6.1.3).

The results of the subdomain-based registrations can be examined on each subdomain separately. In the purely local subdomain approach (cf. Section 4.1.2) the deformation of the left subdomain is continued into the right subdomain with zero curvature energy which leads to a strong compression of the deformation (Figure 6.1.5, third row). In the CLG registration, this undesired compression is prevented by the additional influence of the low registration image distance (Figure 6.1.5, third row, right).

Both subdomain-based registrations lead to a similar image distance (Figure 6.1.5, second row) but differ in terms of deformation error and irregularity: While the purely local subdomain registration produces a large mismatch at the subdomain

Table 6.2: Quantitative registration results. Case 1 (with low-contrast region): CLG is closest to the fine global registration in terms of displacement error and has a lower irregularity than the purely local registration. Case 2 (without low-contrast region): Due to the presence of image structure in the overlap, the results of fine global, purely local, and CLG registration show only little difference. In both cases, all three high-resolution methods outperform the coarse registration. The parameter α is chosen to minimize $e(\mathbf{y})$.

registration	α (opt)	$d(\mathbf{y})$	$e(\mathbf{y})$	$s(\mathbf{u})$
Case 1				
true deformation		0.000	2.992	0.260
fine global	2.7	0.003	0.000	0.256
coarse global	11.0	2.555	12.179	0.250
purely local	0.6	0.032	86.071	241.382
local & global (CLG)	11.0	0.017	5.462	1.034
Case 2				
true deformation		0.000	0.447	0.260
fine global	7.9	0.005	0.000	0.256
coarse global	2.8	6.753	16.542	0.368
purely local	4.5	0.025	0.808	0.265
local & global (CLG)	6.2	0.022	0.693	0.262

boundary, where the deformation differs by more then two cell lengths, the mismatch in the local & global deformation is only slightly noticeable, and the deformation only differs by a fraction of a grid cell from the fine global registration result (Figure 6.1.5, last row). Both subdomain-based methods lead to a similar final image distance that does not exceed the final distance after fine global registration.

The mismatch at the subdomain boundaries leads to a large irregularity of the purely local subdomain deformation, which exceeds the irregularity of the fine global registration by two orders of magnitude.

Case 2: Realistic Image Without a Low-Contrast Regions

In Case 1, a low contrast image region leads to a small gradient of the distance measure in the respective region. Therefore, regularization is important to obtain a smooth global deformation. In Case 2 (Figure 6.1.4), there is no such low-contrast region and therefore, global regularization is less important in this case.

Regarding the registration results, coarse registration is again less accurate than fine registration. The fine registration methods yield similar results (Table 6.2). While the combined deformations do not differ substantially, the individual subdomain results of the purely local subdomain and CLG registration illustrate the difference between the two subdomain-based methods. Beyond the overlap, the deformation of the in-

dividual subdomains continues with minimal curvature energy in the purely local subdomain registration, while it is close to the result from the coarse registration if CLG is used (Figure 6.1.6). However, due to the presence of image structure in the overlap, no visible mismatch occurs at the subdomain boundary.

We interpret the similarity of the fine registration methods as an effect of the large image contrast in the entire image. The image does not exhibit un-structured regions where the gradient of the image distance is small and where regularization dominates the objective function. When the image distance drives the registration, the coupling of the subdomains in the CLG method has only little effect, and its result is similar to a purely local subdomain registration.

6.1.5 Discussion

We compared the new combined local and global (CLG) method to a coarse global registration and a purely subdomain-based registration. The result of a fine global registration was used as the gold standard. The coarse registration uses a low image resolution, but it is capable of computing a natively global deformation. The simpler, purely subdomain-based method is registering at the full image resolution but lacks global regularization. The CLG registration combines the advantages of both discretizations to compute a global registration with high-resolution image data for each subdomain.

In terms of image distance and deformation error, the new CLG method was at least as accurate as a purely local subdomain registration and always better than a coarse global registration. While CLG combines its global deformation from global, independent subdomain solutions, the new method outperforms the purely local subdomain registration in terms of deformation error and irregularity in cases where the subdomain boundary consists of low-contrast image information.

The mismatch of the deformation at the subdomain boundaries is less expressed in the CLG method, but it can still be noticed in cases where image information is scarce.

In the absence of low-contrast regions, the CLG and the purely local subdomain registration lead to similar results in our experiments. Since the CLG registration is more expensive to compute, a worthwhile extension of the method could be to selectively use CLG or purely local registration based on an analysis of the subdomain image data.

In the above experiments, the individual subdomain deformations were cut in the middle of the overlap and combined without blending. This simple approach was chosen to highlight the effect of mismatching deformations. In order to produce a smooth deformation, a blending method from Chapter 5 should be used.

Figure 6.1.5: Case 1: Registration result with low-contrast region. Differences between purely local and CLG registration are visible in the difference images $|T(\mathbf{y}_{\text{right}}) - R|$ (white arrow) and in the final deformation $\mathbf{y}$ (black circle). The reason for the superior result after CLG registration is in the subdomain results where the deformation after purely local registration deteriorates outside of the subdomain (black arrow).

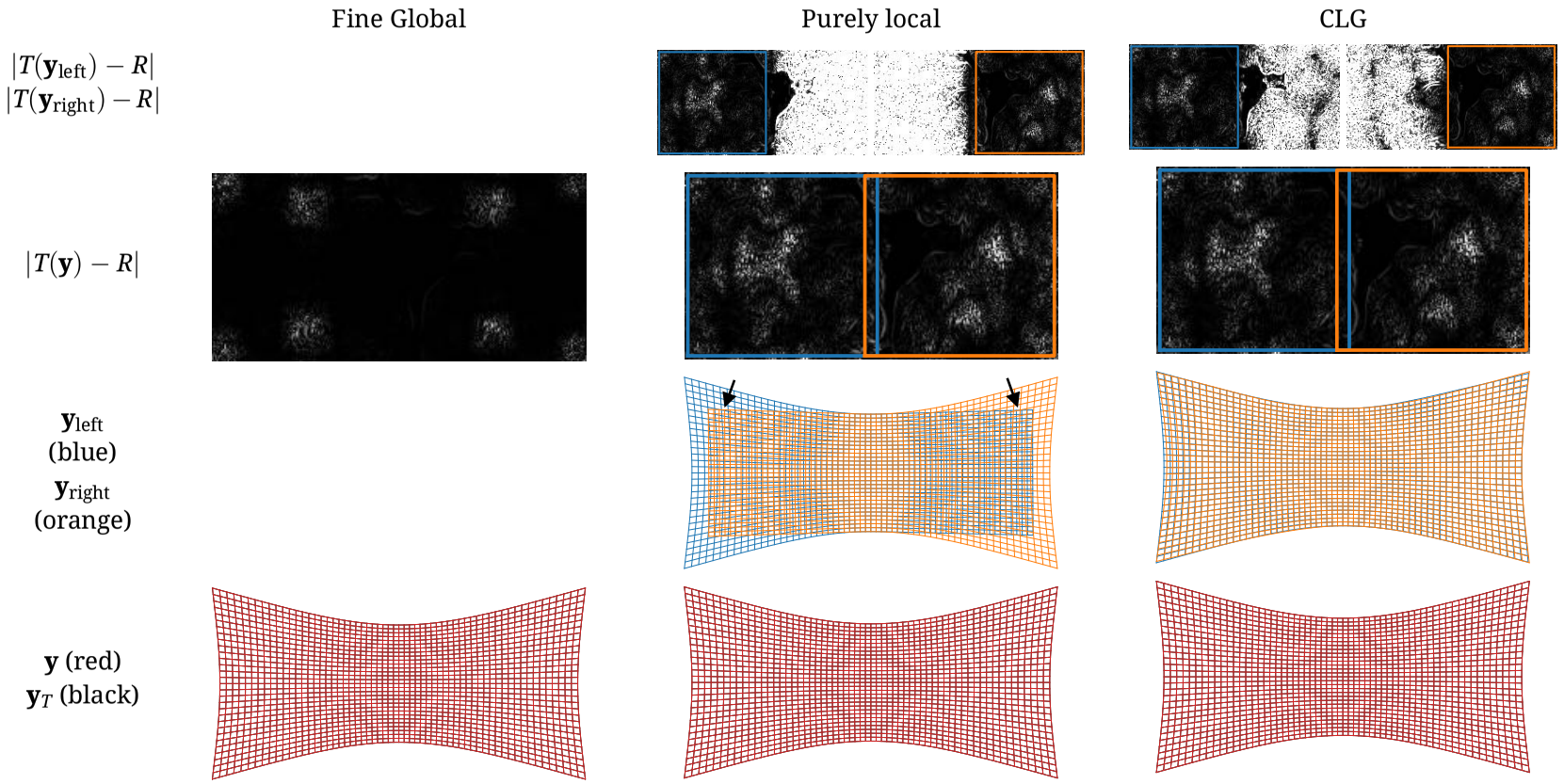


Figure 6.1.6: Case 2: Registration result without low-contrast region. Differences between purely local and CLG registration are not visible in the registration result but in the intermediate subdomain deformations. The deformation proceeds with zero regularizer energy outside of the subdomain (black arrow). This also leads to an increased subdomain image difference $|T(\mathbf{y}_{\text{left}}) - R|$ and $|T(\mathbf{y}_{\text{right}}) - R|$.

Table 6.3: Blending results for Deformation 1 from an image pair with a low-contrast region after purely local registration. All blending methods strongly reduce the irregularity. The overall best result is obtained by local regularization with weight $\beta = 0$ on the distance to the local deformation.

Blending	$D(\mathbf{y})$	$\|\mathbf{y} - \mathbf{y}_{\text{fine}}\|_2$	$S(\mathbf{u})$
true deformation	0.000	2.992	0.260
fine global	0.002	0.000	0.255
without blending (purely local subdomain)	0.006	71.642	**155.180**
weighted sum	0.006	63.106	3.676
local regularization $\beta = 0.01$	0.005	60.889	1.304
local regularization $\beta = 0$	**0.005**	**52.261**	**0.968**
global regularization $\beta = 0.2$	0.006	62.970	1.736
global regularization $\beta = 3.5 \cdot 10^{-5}$	15.411	47.762	0.186

6.2 Numerical Experiments Comparing Blending Approaches

We evaluate the different blending methods based on their effect on the registration results shown in Figures 5.1.1 (Deformation 1), 6.2.1 (Deformation 2), and 6.2.2 (Deformation 3). To cover cases with a large and a small mismatch of the local displacements, the blending methods are applied to the purely local subdomain registration and the CLG registration results. To increase the variation in the example data, Deformation 3 contains an artificially strong nonlinearity, and the blending methods are applied to the modified data.

As evaluation criteria, we compare the deformation error with respect to the optimal deformation computed globally on a fine image resolution but also evaluate SSD and irregularity of the combined solutions. We further expect the blending to preserve the deformation if the mismatch between the subdomain solutions is small.

6.2.1 Deformation 1: Low Contrast Images and Purely Local Subdomain Registration

Using purely local subdomain registration on image subdomains with a large low-contrast region shows the strongest blending effect in our experiments. Most of the results on this example have already been discussed in Chapters 5.1–5.3 individually.

In summary, the results obtained from the weighted sum approach with and without local regularization are in a similar range (Table 6.3) in terms of image distance, deformation error, and irregularity. The use of linear weights without regularization results in a slightly larger irregularity of 0.9 compared to 0.3. Among the local methods, regularization with zero weight on displacement data fit (parameter $\beta = 0$)

Table 6.4: Blending results for Deformation 2 from a CLG registered image pair with a low-contrast region after CLG registration. Even without blending, the irregularity after CLG registration is much lower than after purely local registration. Except for global regularization with a small weight $\beta = 3.5 \cdot 10^{-5}$ on the deformation distance, all methods perform similarly.

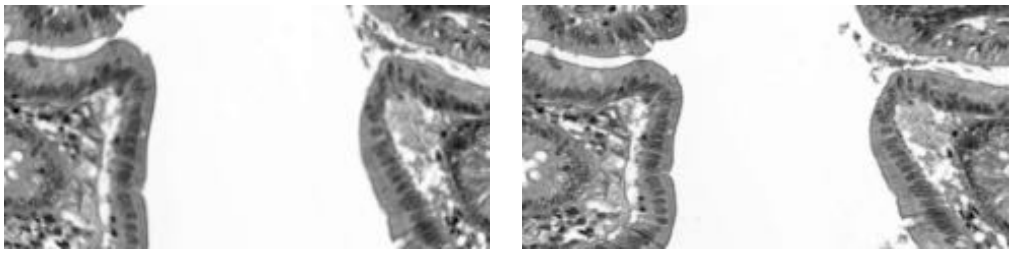
Blending	$D(\mathbf{y})$	$\|\mathbf{y} - \mathbf{y}_{\text{fine}}\|_2$	$S(\mathbf{u})$
true deformation	0.000	2.992	0.260
fine global	0.002	0.000	0.255
without blending (CLG)	0.015	5.003	**0.935**
weighted sum	0.015	4.661	0.264
local regularization $\beta = 0.01$	0.015	4.577	0.256
local regularization $\beta = 0$	0.015	4.290	0.255
global regularization $\beta = 0.2$	0.016	4.656	0.257
global regularization $\beta = 3.5 \cdot 10^{-5}$	19.213	**35.265**	0.138

produces the best result. The small difference between the methods is most likely due to the fact that no image data is present in the overlap region. Therefore, the registration is entirely determined by the regularizer. Purely local regularization computes the deformation in a similar way and therefore leads to a similar result.

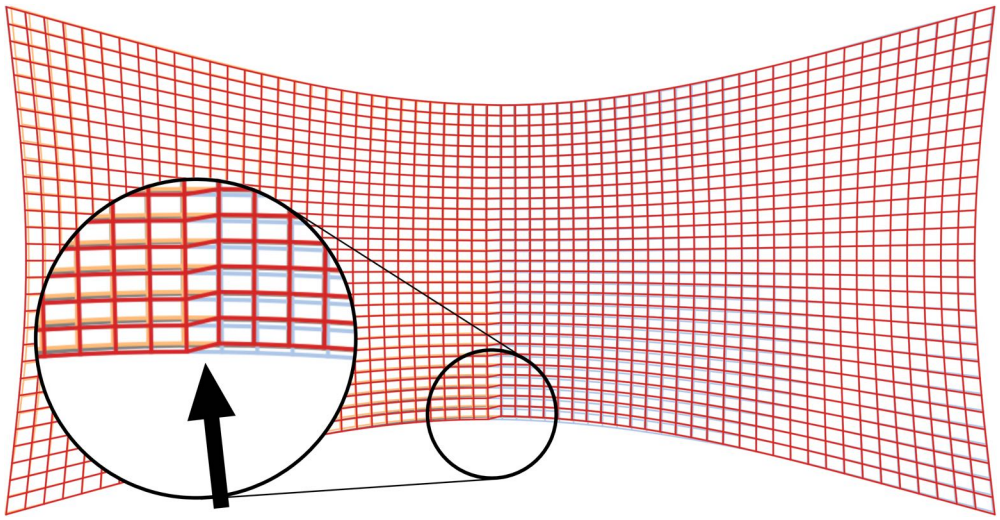
Global regularization with the optimal smoothness parameter (in this case, $\beta = 3.5 \cdot 10^{-5}$) results in the lowest deformation error among the methods examined but leads to a larger overall image distance. This is likely due to the global effect of the regularization outside of the overlap region, where a small displacement error accumulates to a large image distance if compared to the local methods.

6.2.2 Deformation 2: Low Contrast Images and CLG Registration

In the second experiment, the same image pair is registered with CLG registration reducing the mismatch at the subdomain boundaries compared to the purely local subdomain registration. Even when combining the displacements without blending, the result has a lower error then any blending of purely local registration result. The irregularity drops from 155 for purely local subdomain registration to 0.9 for CLG (Table 6.3). This is also apparent when inspecting the deformed grid visually (compare Figure 5.1.1 and 6.2.1). Due to the better initial deformation, the results differ less between the blending methods than after the purely local subdomain registration. The results of the weighted sum, local regularization, and global regularization with $\beta = 1$ are similar due to the small mismatch between the subdomains' deformations. Only in global regularization, the global image distance worsens and displacement errors deteriorate if the parameter β is set too low (Table 6.4).



Reference R Template T



CLG registration result

Figure 6.2.1: Blending results for Deformation 2 from a CLG registered image pair with a low-contrast region after CLG registration. Top: Reference and template image. Bottom: Combined solution without blending (red) plotted over local subdomain solutions from the left (blue) and right (orange) subdomain and the optimal solution $\mathbf{y}_{\text{fine}}$ (gray) from a fine global registration. The local and optimal solutions are only shown where different from the combined solution. After CLG registration, only a small artifact can be observed that is due to the combination of the deformations without blending. The goal of the blending approach is to remove the artifacts while maintaining the low overall displacement error.

Table 6.5: Blending results for Deformation 3 from a CLG registered image pair without low-contrast region and additional strong nonlinearity in the overlap region. An increased deformation error can be observed after the methods with stronger regularization. The result of the weighted sum approach is comparable to or better than all other methods in terms of error and irregularity.

Blending	$D(\mathbf{y})$	$\|\mathbf{y} - \mathbf{y}_{\text{fine}}\|_2$	$S(\mathbf{u})$
true deformation	0.000	1.087	1.784
fine global	0.008	0.000	1.725
without blending (CLG)	0.007	**0.141**	1.728
weighted sum	0.007	0.177	1.729
local regularization $\beta = 0.01$	0.047	1.014	1.700
local regularization $\beta = 0$	0.566	**4.988**	1.660
global regularization $\beta = 0.2$	0.010	0.174	1.722
global regularization $\beta = 3.5 \cdot 10^{-5}$	118.364	**98.899**	0.464

6.2.3 Deformation 3: High Contrast Images with Strongly Nonlinear Deformation

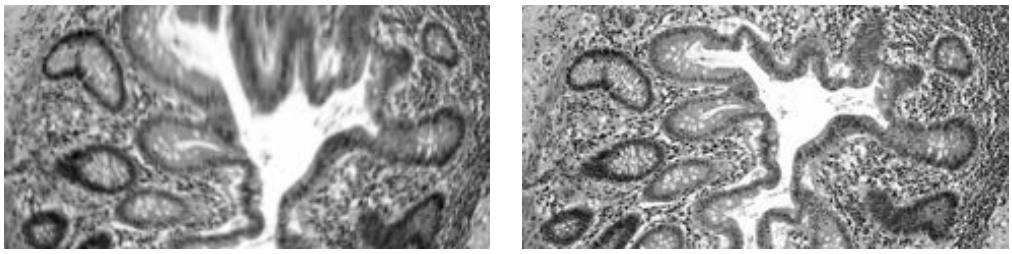
Deformation 3 shows good results in terms of image distance, deformation error and irregularity. To show the drawbacks of additional local or global regularization in the blending, an artificially strong nonlinearity was added to the deformation (Figure 6.2.2). Even with the large nonlinearity in the true deformation, no mismatch is visible at the subdomain boundaries after registration.

Comparing the blending methods applied to Deformation 3, a larger deformation error can be observed after all methods that involve regularization (Table 6.5). This effect is caused by a reduction of the strongly nonlinear components of the deformation, as shown in Figure 6.2.3.

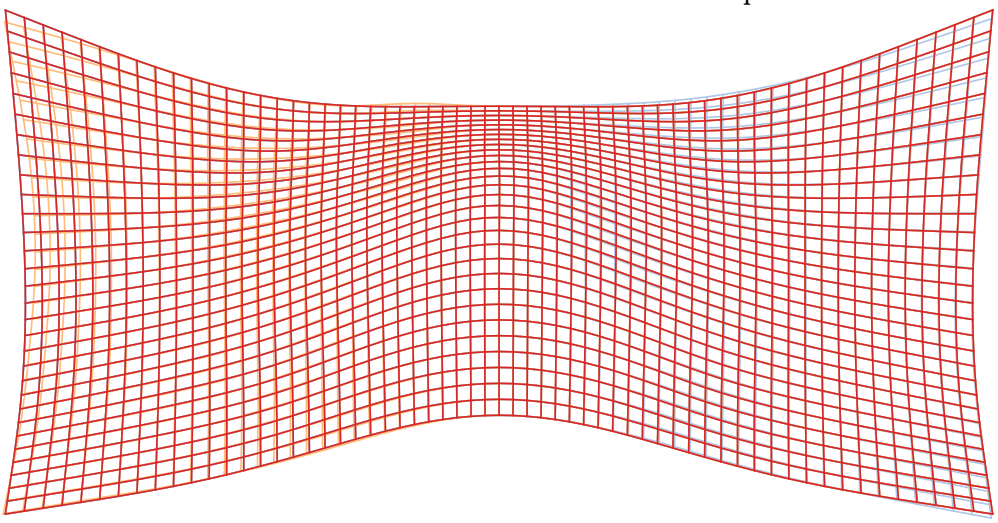
6.2.4 Discussion

We compared local and global blending approaches based on a weighted sum of displacements on overlapping subdomains. Experiments include cases of high and low contrast in the overlap regions and different degrees of nonlinearity of the displacements. Blending is used as a post-processing step to smooth the mismatch of the deformation between neighboring subdomains.

In our experiments of CLG and a purely local subdomain registration, the mismatch after CLG registration is always smaller, and CLG produces better overall results than purely local subdomain registration, before and after blending.



Reference R Template T



CLG registration result

Figure 6.2.2: Combination of Deformation 3 from a CLG registered image pair without low-contrast region and additional strong nonlinearity in the overlap region. The grid shows the combination of the local deformations without blending. Top: Reference and Template image. Bottom: Solution without blending (red) plotted over local subdomain solutions (blue, orange) and optimal solution $\mathbf{y}_{\text{fine}}$ (gray). The local and optimal solutions are only shown where different from the combined solution. Due to the high contrast in the overlap region, even the combination without blending produces visually acceptable results.

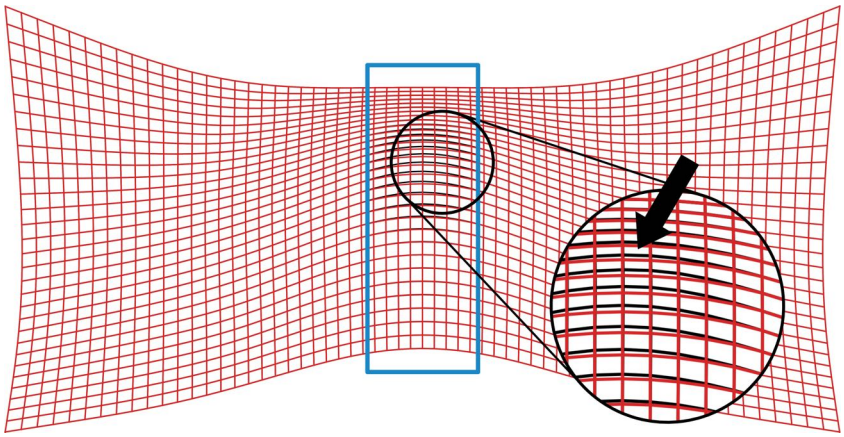


Figure 6.2.3: Blended grid after local regularization (red, $\beta = 0$) of Deformation 3. The box (blue) marks the nodes of the grid that are re-computed by local regularization. The fine global registration result is visible if different (black). The strong nonlinear components in the center are not preserved after blending with regularization.

The disadvantage of the straightforward combination without blending approach is in its high irregularity measure, which is also visible in the images. While this approach is minimally invasive in the sense that the originally computed deformation is not changed, the other extreme is global regularization, where the entire displacement is recomputed.

Regularization approaches are superior after a low-contrast purely local subdomain registration, but only a small benefit can be observed on low-contrast CLG registration. Purely local regularization shows good results in the examples where regularization is the main driver of the registration. If the overlap region consists of a high-contrast part of an image, regularization can degrade an originally good registration. Especially the global regularization approach bears the risk of over-smoothing the deformation that contains a strong nonlinearity.

If the user is willing to manually chose a good parameter β, for example based on the visual appearance of the registration result, blending by local regularization is a good choice as it can compromise between smoothness in low contrast regions and the preservation of deformations in high contrast subdomain overlaps.

For a purely automatic registration, we recommend using the weighted sum approach in combination with CLG registration. The weighted sum approach guarantees to preserve those parts of the deformation where both subdomains are in agreement. Also, no parameter needs to be selected manually.

7 Results of the Application to Large-Scale Images

Contents

In this chapter, we apply the CLG registration to six pairs of serial sections. The dataset comprises four different staining combinations from two different pathology labs (Table 7.1, Figures 7.1.1 – 7.1.2).

We compare the proposed CLG method to a fine global registration with an identical final resolution. In all the six image pairs, both methods result in a similar landmark registration error while requiring less than a fourth of the memory. Compared to a coarse global registration, the new method reaches lower landmark errors in all but the highest image resolutions. The tradeoffs are a higher computation time and an elevated irregularity. In our experiments, the new method is suitable for replacing a fine global registration, especially if the image data would otherwise exceed the available memory.

7.1 Histopathological Image Data

The first four cases are human oral tissue samples, cases five and six were taken from the colon. A gallery of registered images can be found in Figures G.7–G.36, beginning on page 125. All computations in this chapter are performed on a four-core Intel® Core™ i7-7700 CPU with 32 GB of RAM.

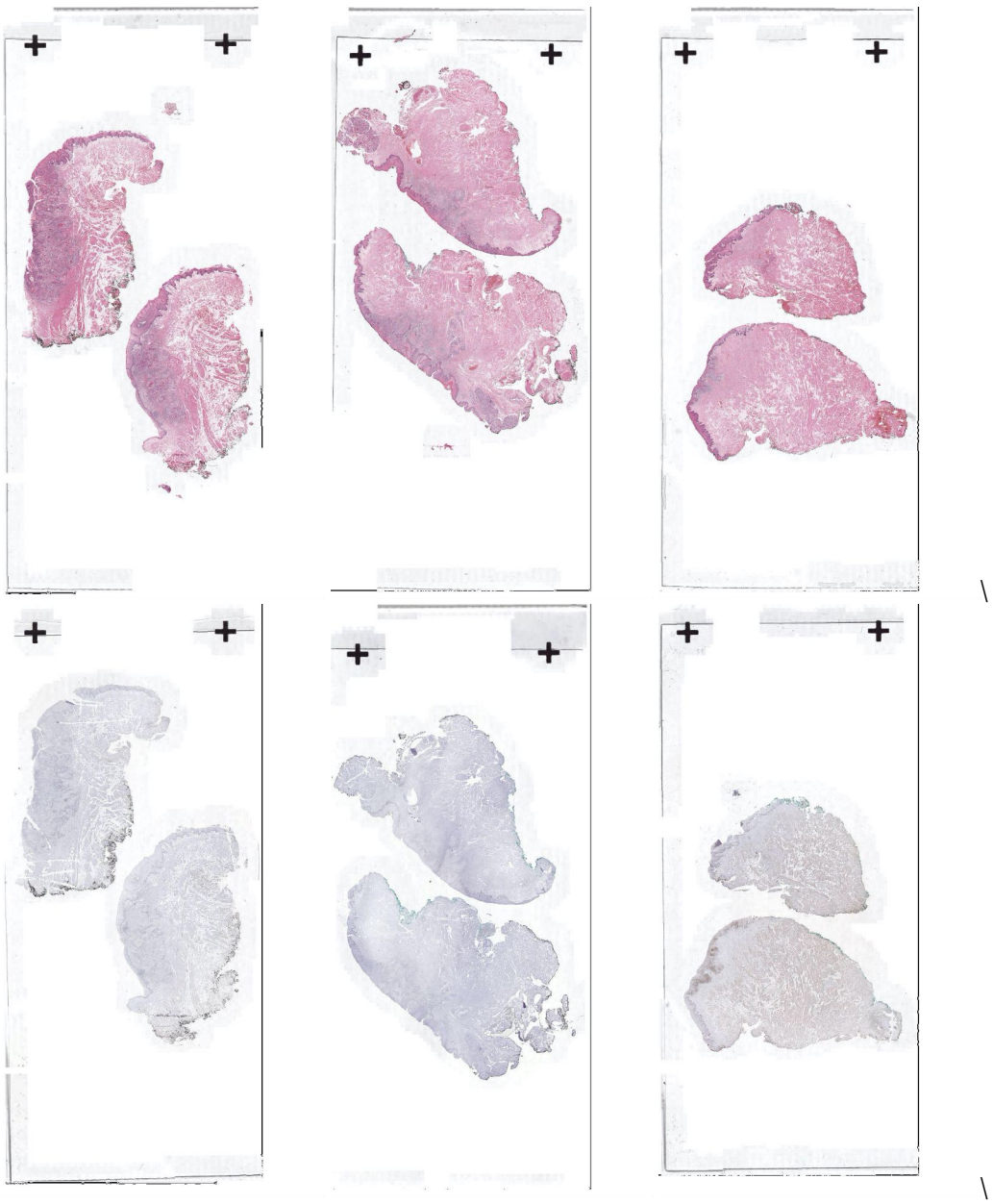


Figure 7.1.1: Image pairs 108, 29, and 361 before registration. Top: reference image, bottom: template image. The size of these images is circa 23 mm in width and up to 55 mm in height. High-resolution crops are shown in the results in Figures G.23 ff.

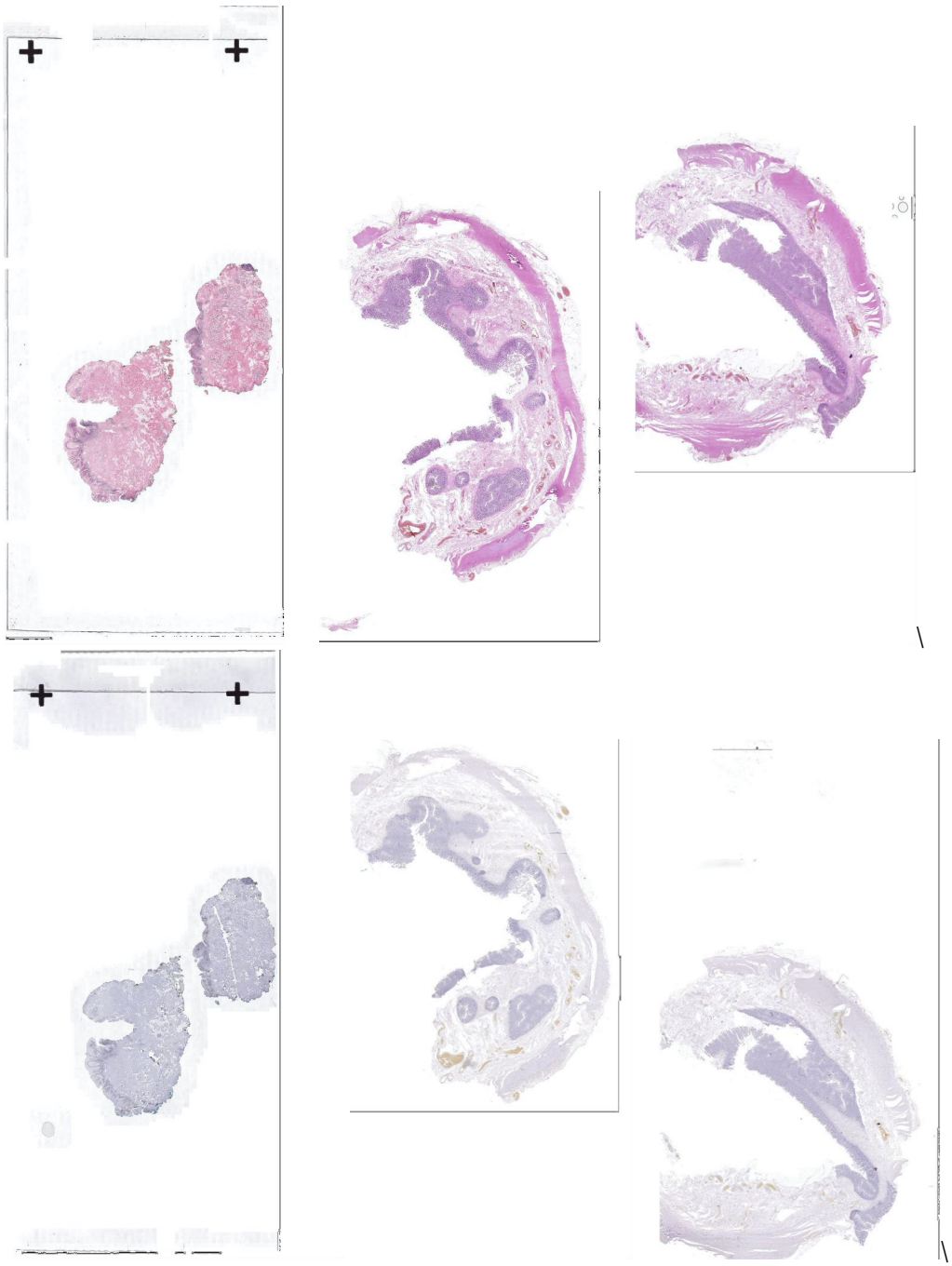


Figure 7.1.2: Image pairs 446, 9515_20, and 9515_21 before registration. Top: reference image, bottom: template image. High-resolution crops are shown in the results in Figures G.23 ff.

Table 7.1: Overview of the medical images used in this chapter. The images are provided by Dr. Jeroen van der Laak and his team at the Diagnostic Image Analysis Group at the Radboud university medical center in Nijmegen (DIAG) and by Dr. Andreas Turzynski at the clinic for pathology in Lübeck (AT). The images show human oral and colon tissue samples. We use manually placed landmarks for evaluation.

Case	Staining-Pair	No. of Landm.	No. of Pixels equival. to <16k	Resol. at <16k	Origin
29	H&E , Ki67	13 pairs	6076 × 13 784 px	3.9 µm/px	DIAG
361	H&E , CD45	8 pairs	6084 × 13 784 px	3.9 µm/px	DIAG
108	H&E , CD8	19 pairs	6068 × 13 660 px	3.9 µm/px	DIAG
446	H&E , Ki67	9 pairs	6096 × 13 660 px	3.9 µm/px	DIAG
9515_20	H&E , Haem.	13 pairs	7292 × 11 800 px	2.8 µm/px	AT
9515_21	H&E , Haem.	20 pairs	7488 × 9296 px	2.8 µm/px	AT

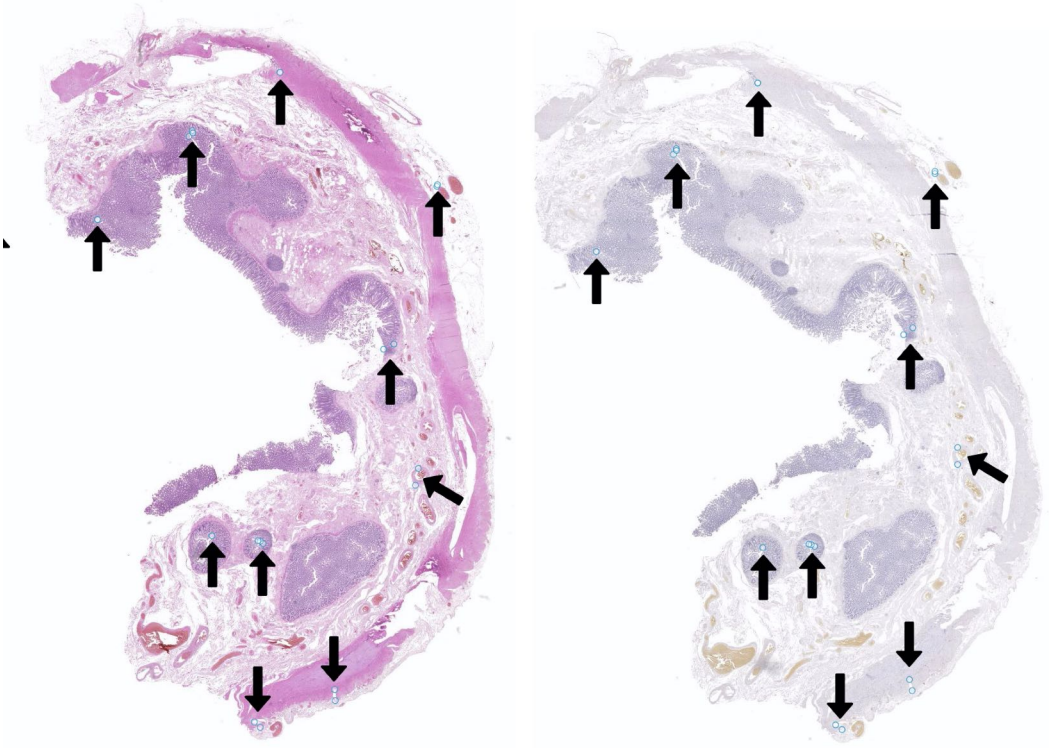


Figure 7.1.3: Example of a pair of colon images (image pair 9515_20) before registration with corresponding landmarks (black arrows). Left: H&E stain, Right: Hematoxilin stain.

The registration of the two stains is needed to combine and predict biomarkers in ongoing cancer research projects. We show an image of one of the slide pairs in Figure 7.1.3.

The first slide in all datasets is stained with Haematoxylin (Haem.) and Eosin (H&E) [Kie15, pp. 137 ff.], one of the principal stains used in most histological analyses. The second image of each pair is stained either using a specific immuno-biomarker (Ki67, CD45, CD8) or using Haematoxylin without the Eosin counterstain.

While Haematoxylin stains nuclei, the counterstain with Eosin also stains other materials such as cytoplasm and erythrocytes. The biomarker Ki67 is a risk factor in the diagnosis of breast cancer [dAza07], and CD45 and CD8 mark specific T-lymphocytes [War10, pp. 139 ff.] which are central to the human immune system.

7.2 Implementation Specifics

The CLG registration extends an existing library for nonlinear image registration [Rüh17a; Kön18] developed by the Institute of Mathematics and Image Computing (University of Lübeck) and Fraunhofer MEVIS with contributions by the author.

We use this implementation in all the following experiments. It is aimed at large data sets and differs from the implementation used on artificial examples shown in Chapter 6.1. First, instead of explicitly constructing matrix operators—such as the matrix B in the regularizer—we use matrix-free operations. This lowers the memory usage of the algorithms and also lowers the computation time because it requires fewer read and write operations. The downside of these optimizations is a more complex codebase, which makes modifications and additions more complicated and time-consuming.

Further implementation differences are the use of linear image interpolation instead of cubic B-Splines and L-BFGS [Noc06, pp. 224 ff.] instead of the Gauß–Newton optimizer. See Chapters 2.5.1 and 2.6.2 for a discussion of the respective methods.

7.3 Data Organisation and Experiment Setup

Most formats for digital whole-slide images store the images along with an image pyramid containing downsampled versions of the image [Sat13]. Depending on the image's size at its finest level, the image size of the low-resolution representations varies. When selecting a level for registration, we define an upper bound for its extent. As an example, the smallest image size is denoted "<1k" which corresponds to the largest image in the pyramid whose dimensions do not exceed 1000 × 1000 pixels. In the images in Table 7.1, <1k corresponds to a resolution of ca. 50–60 µm/pixel. Other sizes used in this chapter are <2k, <4k, <8k, <16k, <32k, and <64k, where each step doubles the number of pixels per dimension.

Table 7.2: Parameters used in the CLG registration pipeline for the H&E-Haematoxylin image pair. Regularizer weight α and NGF noise level η are determined by a parameter search.

Component	Parameter	Value
Affine	image sizes	<1k, <2k, <4k, <8k, <16k
Registration	number of levels	4, 5, 6, 7, 8
	NGF edge parameter ϵ	0.5
Non-Parametric	image sizes	<1k, <2k, <4k, <8k, <16k
Registration	number of levels	4, 5, 6, 7, 8
	regularizer weight α	1
	NGF noise level η	1
	deformation size	129 × 129
CLG	coarse image sizes (R)	<1k, <2k, <4k, <8k, <16k
	subdomain image size (R)	1.2 × coarse image size
	resulting image sizes (R)	<4k, <8k, <16k, <32k, <64k
	number of subdomains	4 × 4
	number of levels	3
	regularizer weight α	1
	NGF noise level η	1
	deformation size	129 × 129
Blending		weighted sum

The alternative of selecting a predefined image resolution in µm/pixel would allow the user to select a desired level of detail. It could also handle situations where the pixel size of both images is very different. However, in the present application, both images are typically acquired using the same scanner such that the image resolution in both images is identical. The selection of the level based on the image size allows for a rough estimation of computation time and memory requirement, which is preferable for our experiments.

Operations on smaller images are computationally less expensive. Besides the multilevel scheme described in Chapter 2.5.2, we compute an initial guess for the CLG registration by the following two steps.

First, we compute an affine registration using a coarse image resolution to obtain a rough initial alignment of the two slides. The second step is a coarse global nonlinear registration that uses the affine registration result as an initial guess. While both steps operate with the same image size, the affine registration has a lower number of degrees of freedom and is, therefore, faster to compute. The result of the coarse nonlinear registration is, in turn, used as an initial guess for the final CLG registration.

To compute the CLG registration, we decompose the image domain into four by four

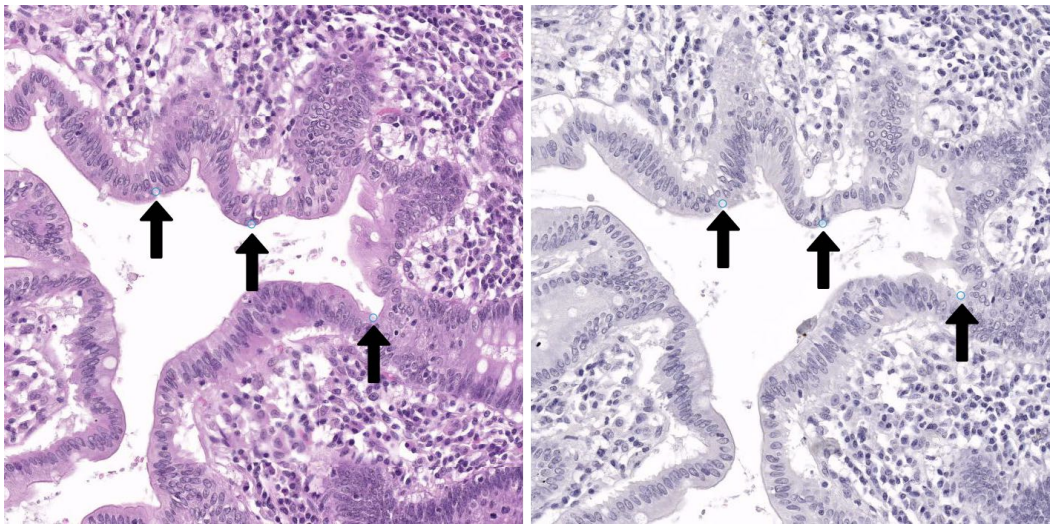


Figure 7.3.1: Three corresponding landmark pairs in the H&E (left) and Hematoxilin (right) image.

subdomains. In each subdomain, we use images with a four times higher resolution than in the low-resolution base images. The four by four subdomain decomposition is selected as a proof of concept as it has proven sufficient for the application to the present whole slide images. A larger number of subdomains could further reduce the memory requirements of the registration but would also increase the overhead caused by the additional overlap. The analysis of larger numbers of subdomains is left for future work.

For each of the following experiments, we set a base image size between <1k and <16k. This image size is then used in the affine, coarse global, and the CLG registration for both the underlying coarse resolution image and the higher resolution subdomain image. The subdomains are extended by a 25% overlap. A base image size of "<nk" results in a final image size of "<4nk" when decomposing the image domain in four by four subdomains.

In the following, we compare a) an affine global registration, b) a nonlinear, high-resolution, global registration, and c) the proposed CLG method. The methods are compared such that their final image size is identical, which means that the CLG registration requires only a fourth of the memory of a corresponding fine global registration (Table 7.3).

In contrast to the experiments in Chapter 6.1, no true deformation is available. Therefore, we use sets of manually placed landmark pairs to measure the registration accuracy (Figure 7.3.1).

Landmarks are placed on corresponding structures in both sections using the maximal magnification. When the sections are not adjacent, some structures are not present in both images. Especially cell nuclei and other small-scale structures are

difficult to correlate on neighboring slides, which makes the placement of accurate landmarks challenging and time-consuming. For this reason, we only provide a small number of landmarks that were created by the author using unregistered images.

The choice of its parameters influences the accuracy of the registration. To determine optimal values, we run a two-dimensional parameter search for η and α as described in Chapter 2.4. As a result of the parameter search we select $\alpha = 1$ and $\eta = 1$ for the registrations in this chapter. An illustration can be found in Figure 7.3.2, where we compare various regularization and noise parameters concerning landmark error. Among the parameters with low landmark error, we prefer larger values for α as the corresponding deformations have a lower irregularity. We note that a moderate change in these parameters does not affect the overall result of the following comparisons. One example can be found in Figure 7.4.1.

An overview of the parameters used in the registration pipeline is listed in Table 7.2.

7.4 Influence of the Deformation Grid Resolution

As described in Chapter 4.1.3, we discretize the deformation with a lower number of grid points than in the images. The advantages of this coarse discretization are a) the reduction of required memory for storing the deformation during registration and b) the reduced problem size in the optimization.

Low-resolution grids are used in some registration software packages such as NIFTYREG [Mod10], and some authors report that a higher number of coefficients leads to a more accurate registration [Man16]. In our experiments (Figure 7.4.1), we see a reduction of the landmark registration error for finer discretizations up to a a grid size of ca. $m \times m = 65 \times 65$. The reduction is independent of the regularization parameter α. We assume that finer discretizations have only little effect because the proportion of the nonlinearity in the deformation is small. We choose the next finer discretization of $m \times m = 129 \times 129$ in our experiments to have some margin for stronger nonlinearities in other images while keeping computation time and memory requirements low.

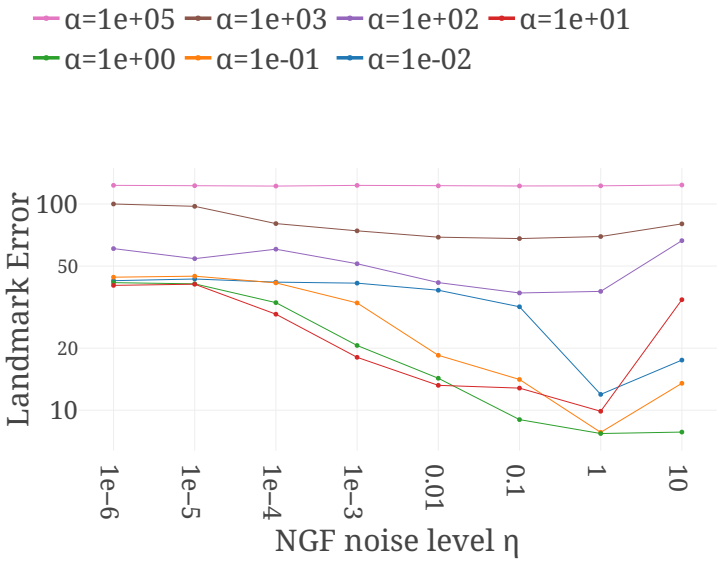
7.5 Comparison of Coarse, CLG, and Fine Registration

We compare the result of the three registrations based on landmark error, irregularity, and memory requirements.

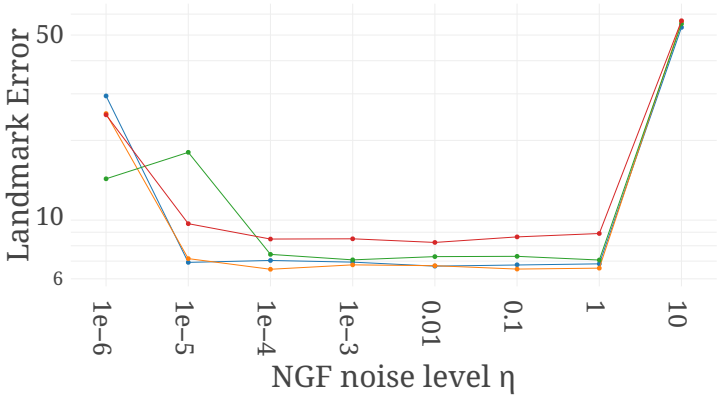
7.5.1 Landmark Error and Image Size

The landmark error is computed as the mean Euclidean distance

$$e(\mathbf{y}) = \frac{1}{N} \sum_{k=1}^{N} \|\mathbf{y}(l_k^R) - l_k^T\|_2^2$$



CLG registration (Image Size: <8k)



Fine Global Registration (Image Size: <8k)

Figure 7.3.2: Example of a parameter search for parameters η and α with respect to landmark error (image pair: 9515_20, s. Table 7.1). In this case, the parameters $\alpha \in \{0.1, 1\}$ result in the lowest error in both cases. The noise level $\eta = 1$ results in the lowest error after CLG registration, after fine global registration $10^{-4} \leq \eta \leq 1$ leads to minimal landmark error.

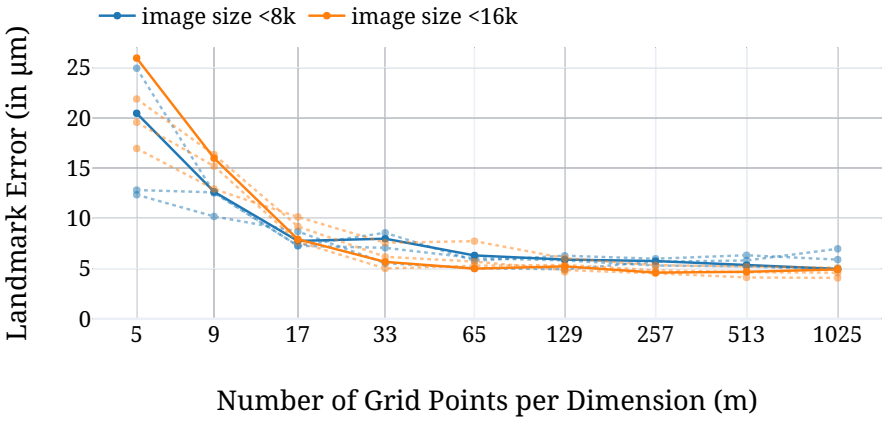


Figure 7.4.1: Influence of the discretization of the displacement u on the landmark error for the image pair 180. Different values of α are shown to account for the influence of the regularization parameter on the landmark error. The solid line shows the error for $\alpha = 1$, the dotted plots show errors for $\alpha \in \{0.01, 0.1, 10, 100\}$ ($\eta = 1$ in all cases). No substantial improvement can be observed for grid sizes above $m \times m = 65 \times 65$ due to the low amount of nonlinearities in the deformation.

of all landmark pairs (l_k^R, l_k^T) in reference and template image.

In our experiments (Figure 7.5.1), CLG registration reaches approximately the same accuracy as non-parametric registration without loading the global high-resolution image into memory. When using a similar amount of memory such as in

- CLG with image size <16k which requires 448 MB of memory and
- fine global with image size <8k which requires 656 MB of memory,

CLG is more accurate than fine global registration in four out of six of the image pairs.

From these experiments, we see that the final image size (and, therefore, image resolution) has the strongest influence on the landmark error. The CLG registration can be used to compute an image registration based on larger images, which can lead to a more accurate registration result.

Overall, the landmark error decreases with increasing image resolution. A limit can be observed in large image sizes over <32k, where the landmark error does decrease little or not at all. We assume that the lack of small-scale correspondences is the determining factor for this effect. If the two sections are too different on a small scale, an image registration method cannot establish correspondences between structures. As a secondary effect, the lack of correspondence also reduces the reliability of the landmark pairs.

These experiments show that in the present dataset, a registration using the highest possible image resolution is not needed since the accuracy does not improve for res-

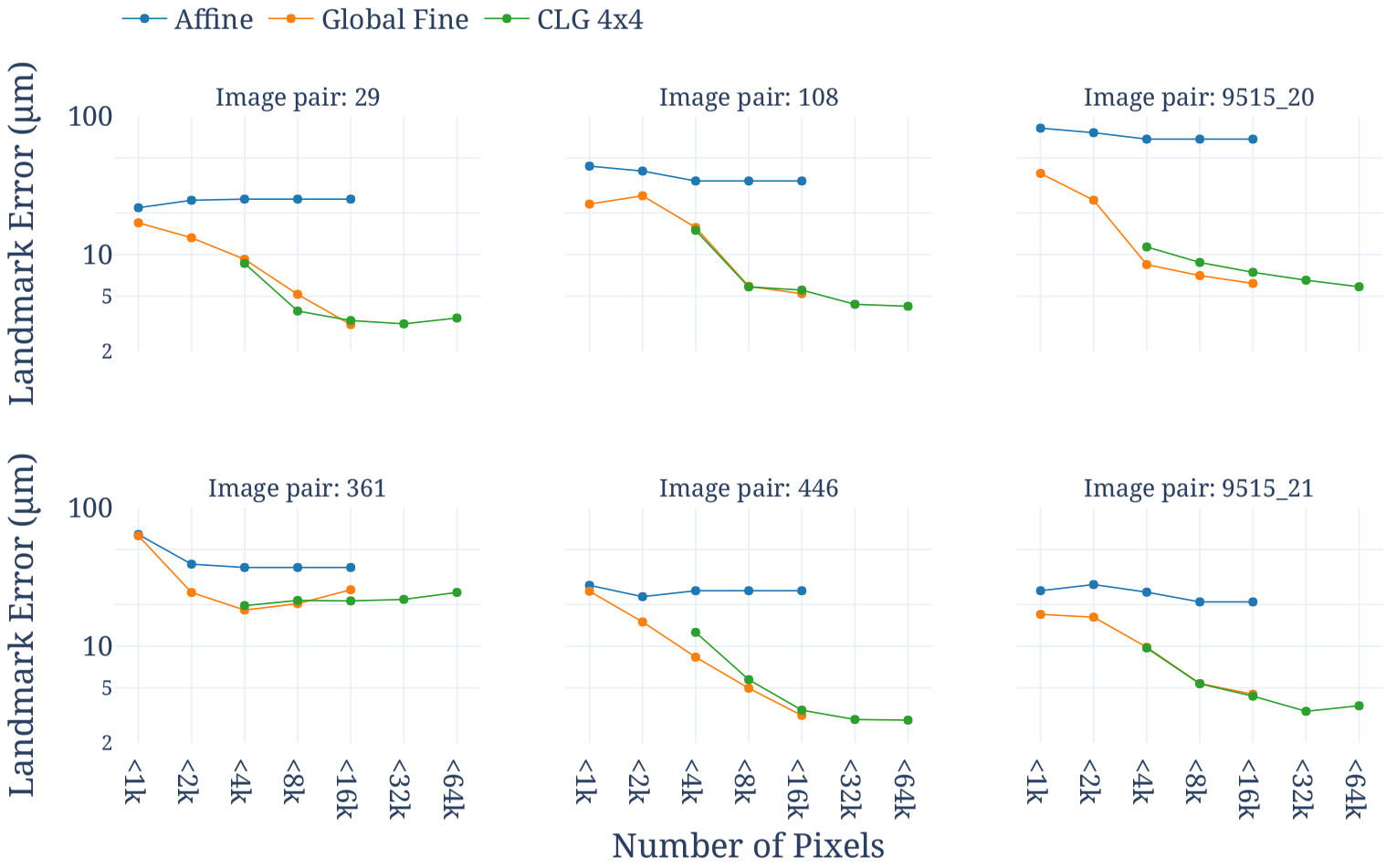


Figure 7.5.1: Landmark error for different image sizes after affine, fine global (non-parametric) and CLG registration with the same final image resolution. CLG registration reaches approximately the same accuracy as non-parametric registration. Affine registration is the least accurate. Fine global registration has not been computed above image sizes of <16k.

olutions over <32k. In general, the required image resolution seems to depend on the similarity of the two sections.

7.5.2 Irregularity

While the landmark error measures the alignment of selected corresponding structures, the regularity is a global indicator for the reliability of the deformation. When comparing CLG and fine global registration (Figure 7.5.2), we observe a higher irregularity after CLG registration, which is caused by a lower irregularity in higher image sizes.

This observation is unexpected at first because higher resolution images contain smaller-scale information, which should lead to more local deformations. However, the projection scheme (Chapter 4.1.3) restricts the gradient of the distance measure to the low-resolution deformation grid. Therefore, the amount of nonlinearity that can be represented is independent of the image resolution. Furthermore, we expect a lower discretization error and less noise from the higher-resolution image gradient, which leads to a higher regularity.

In our experiments, CLG registration results in the same or higher irregularity as after fine global registration. The higher irregularity is an effect of the additional low-resolution image data in the objective function.

In the computation of each subdomain, only the local image distance term uses high-resolution image data. For this reason, in low-contrast image regions, the CLG registration is globally closer to a coarse registration. We compare registration results with low and high irregularity (Figures 7.5.3, 7.5.4) where the similarity of CLG and coarse registration is reflected. The difference to the fine registration occurs mainly in image regions with low contrast. These regions contain no tissue, which is why the deviation from the fine registration does not affect the registration's accuracy.

We observe no artifacts at the subdomain borders, which supports the choice of the weighted sum (Chapter 5.2) to combine the subdomain deformations.

7.5.3 Computation Time

The CLG method reduces the required memory by decomposing the registration problem in multiple smaller sub-problems, which are addressed independently. The objective function in each of these sub-problems includes a second data term, which makes its evaluation more expensive. Furthermore, the fine global registration library is optimized for performance, which does not apply to the same extent for our CLG implementation. Both factors lead to a longer computation time in our experiments.

Comparing fine global and CLG registration with the same final resolution experimentally, CLG is slower by a factor between 27 (image size: <4k) and 12 (image size: <16k) (Figure 7.5.5). The difference in computation time seems to decrease with

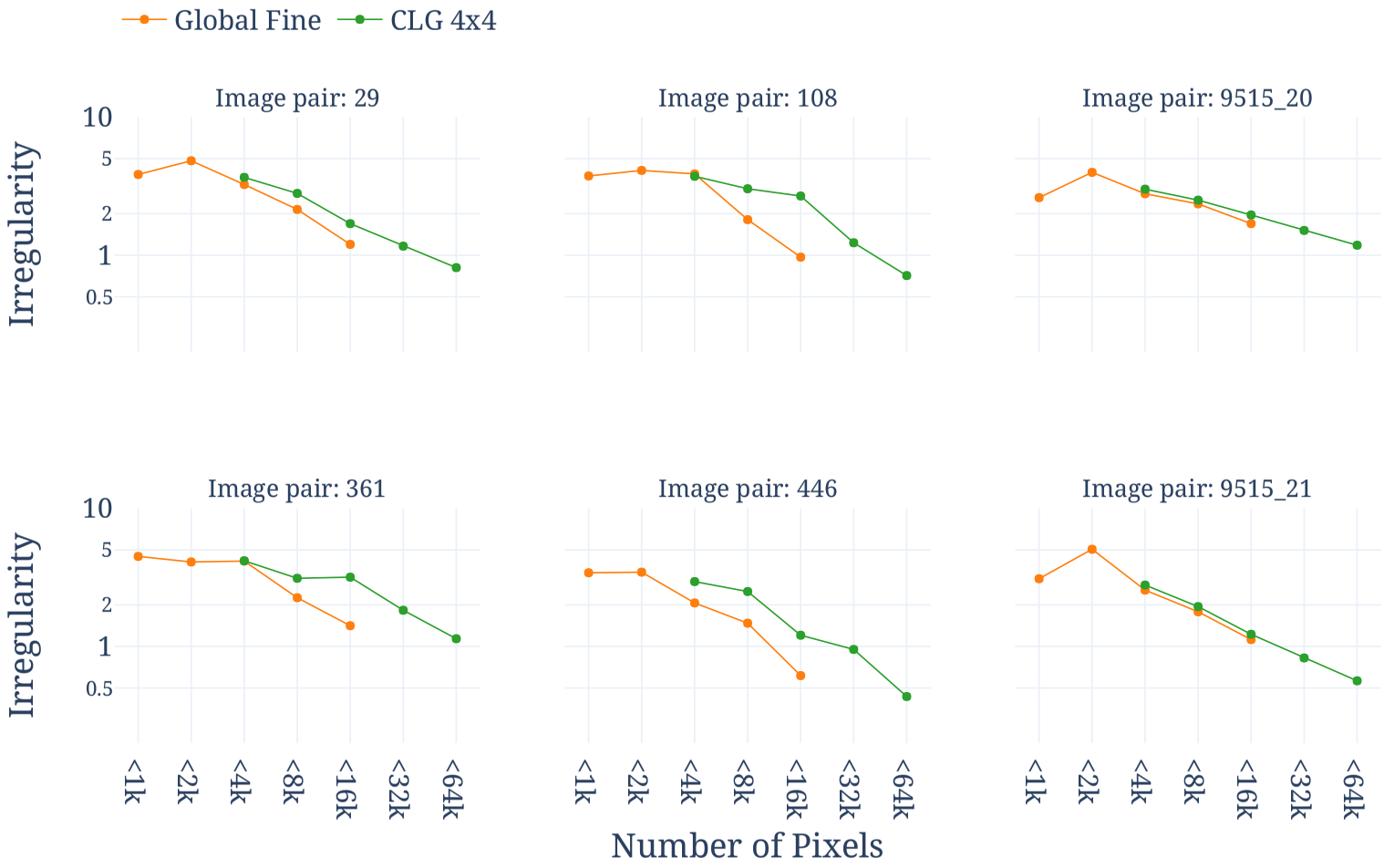


Figure 7.5.2: Irregularity ($S(\mathbf{u})$) for different image sizes after fine global and CLG registration with the same final image resolution. CLG registration results in a higher irregularity in all image pairs. Fine global registration has not been computed above image sizes of <16k.

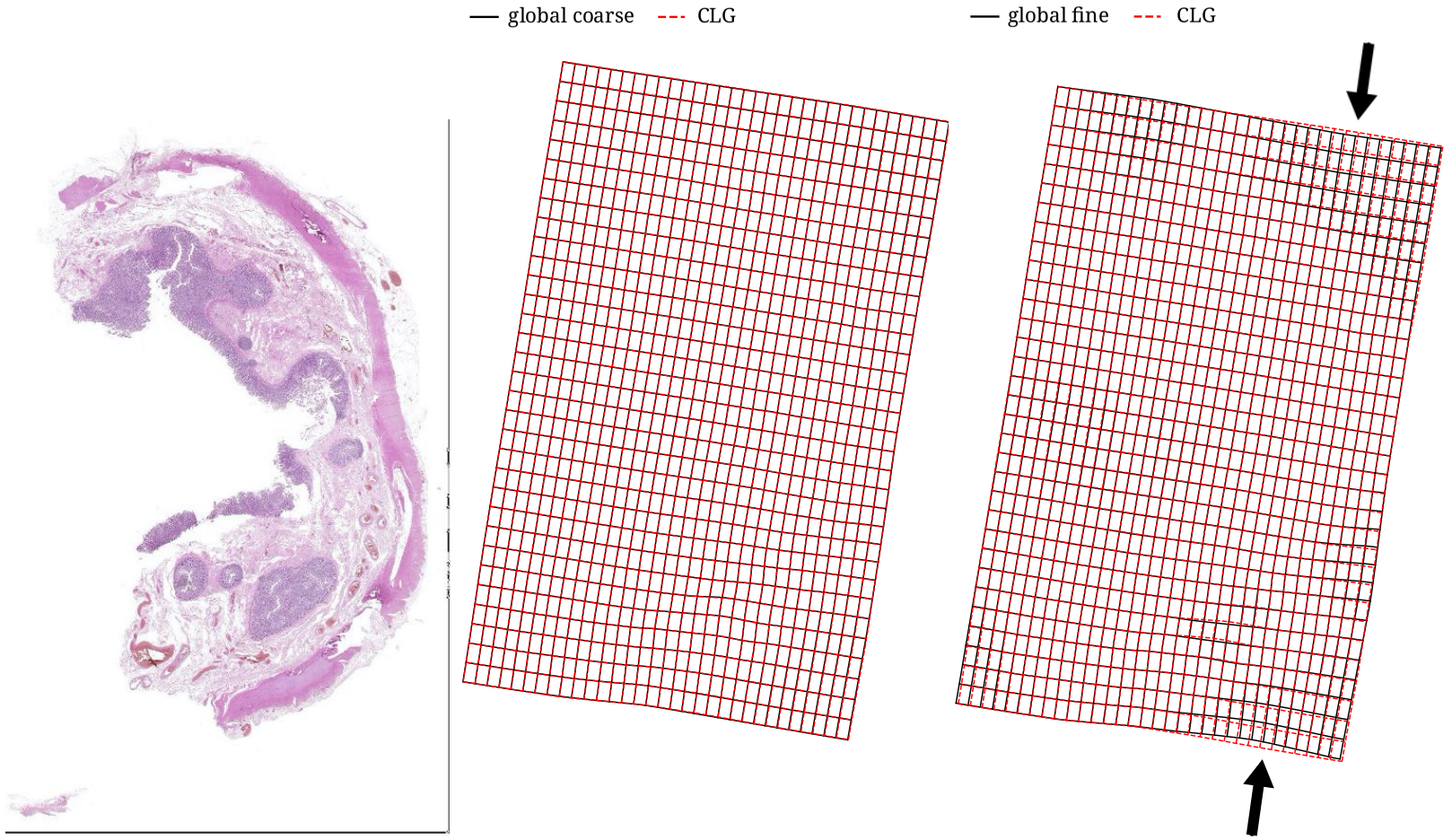


Figure 7.5.3: Reference images (left column) along with overlaid deformations comparing coarse and CLG (center) and fine and CLG (right) registration. Image pair 9515_20 with only small differences.

Figure 7.5.4: Reference images (left column) along with overlaid deformations comparing coarse and CLG (center) and fine and CLG (right) registration. Image pair 446 with a large low-contrast region and larger differences in the deformation. In low-contrast regions, the CLG registration results are similar to the underlying coarse registration.

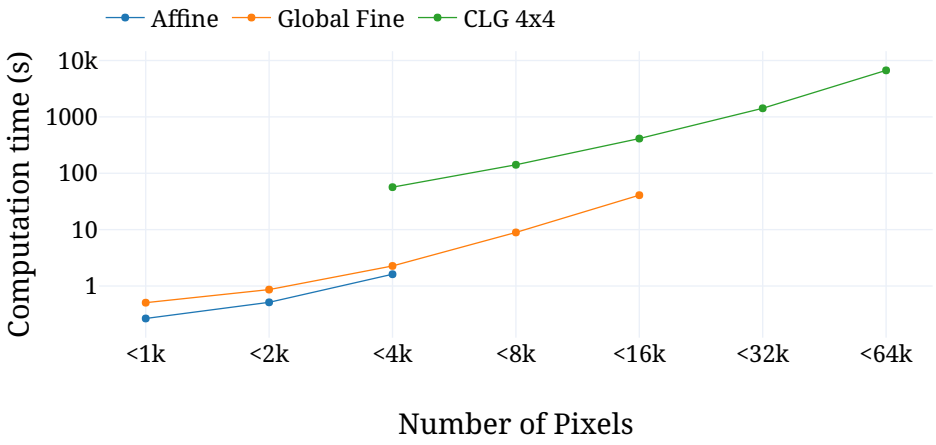


Figure 7.5.5: Mean computation time for all image pairs with respect to registration method and image resolution. Compared to a fine global registration with the same final image size, CLG registration is slower by a factor between 27 (<4k image size) and 12 (<16k image size).

higher image sizes. Still, these factors are considerably higher than the estimated factor of 2.7 in Chapter 4.2.3. Compared to a coarse registration, we observe a larger number of L-BFGS iterations when optimizing the CLG objective function.

As discussed in Chapter 4.2.3, the use of a global step size can slow down the optimization, resulting in a higher number of iterations. A modification of the objective function to include an additional weight of the inner subdomain or an adaptive step size computation could help to improve the convergence speed.

The displacements of the respective subdomains are discretized on a coarse grid. In the case of the weighted sum, we do not require expensive computations for their combination. The time to compute the combination is, therefore, negligible compared to the computation of the registration.

A factor of twelve in computation time is substantial when dealing with large datasets or in time-critical applications. However, in situations where the random access memory is limited, the proposed CLG method enables the user to obtain a more accurate registration, which would not be possible otherwise. Furthermore, the factor could be reduced by the indicated options for optimization.

7.5.4 Memory Usage

We measure the required memory during the execution using the Massif heap profiler. A heap profiler measures the allocated memory of a process along with additional memory that is required for book-keeping by the operating system [Sew08,

Table 7.3: Used memory by the registration of Case 29 measured during execution using the Massif heap profiler. Images of size >64k could not be registered using fine global registration on a computer with 32 GB of RAM. The fine global registration needs 3–5.9 times the memory of the CLG registration.

Size	Non-Parametric	CLG 4 × 4
<1k	51 MB	-
<2k	80 MB	-
<4k	195 MB	65 MB
<8k	656 MB	144 MB
<16k	1856 MB	448 MB
<32k	9868 MB	1669 MB
<64k	-	5567 MB

pp. 89 ff.]. The measured values include memory required by the image loading routines and un-freed memory of pre-alignment, parametric, and non-parametric registration and are therefore higher than the estimations made based on Table 4.2.

Comparing a fine global registration and the CLG registration with the same final image size, the fine global registration is 3 to 5.9 times more expensive in memory than the CLG registration (Table 7.3). We also find that the CLG approach can compute a registration with twice the image size while still using less memory compared to a fine global registration.

7.6 Summary

We compare the CLG registration to a coarse global and a fine global registration using six large-scale image pairs from two independent laboratories. The images show four different stain combinations.

Concerning the landmark error, the difference between CLG and fine global registration is negligible for practical purposes. Both methods outperform a coarse global registration. In some cases, the CLG registration shows an elevated irregularity measure.

Its low memory footprint is the main advantage of the CLG registration. The reduction in memory is made possible by sequentially solving small parts of the registration problem. Processing the images sequentially results in twelve times longer computing time compared to fine global registration. Using the CLG method, the registration of images of 24 000 × 54 000 pixels (<64k) can be solved using less than 6 GB of memory while the fine global reference method exceeds 32 GB of RAM. Using the reference method, the registration of these images was not possible on the given hardware.

8 Discussion, Conclusion, and Outlook

Discussion

In this thesis, we propose a method to solve the large-scale image registration problem using a fraction of the memory that is required by a comparable global registration and apply the new method to images from digital pathology.

The large-scale image registration problem is to superimpose two images that are too large to be stored in random access memory (RAM). In digital pathology, the objects that are visible in the images—like nuclei of cells—are in the range of micrometers such that a stepwise increase in image resolution from ca. 60 µm/pixel to up to 1.4 µm/pixel leads to a successive increase in registration accuracy in the experiments.

The experiments suggest using a resolution of 1.4 to 1.9 µm/pixel, which corresponds to an image size below 32 000 × 32 000 pixels (denoted as "<32k") in our data. Higher image resolutions do not show a reduction in landmark error. However, the required detail depends on the similarity of the two sections, and we recommend adjusting the resolution to new image data.

A standard approach to solve large-scale problems is to resort to parallel domain decomposition methods, where the computational load is distributed to multiple computers and solved in parallel. We discuss concepts from parallel domain decomposition that aim to reduce the communication between computing nodes and apply them to the registration problem at hand. A downside of these methods is the necessity of multiple computing nodes that run in parallel.

The new combined local and global (CLG) method refrains from inter-subdomain communication and decomposes the large-scale problem into independent subproblems. Hence, a large-scale problem such as the registration of images of 24 000 × 54 000 pixels can be solved on one computer with 6 GB of memory while the fine global reference method exceeds the capacity of a computer with 32 GB of RAM. CLG registration leads to results that are comparable with a fine global registration while using less than 20% of the RAM.

While the CLG and fine global registration are not mathematically equivalent, experiments show that their results are comparable in landmark error, which is much lower than after coarse global registration. The new method outperforms a purely local subdomain-based method experimentally in terms of the deformation regularity measure.

We compare different blending approaches to combine the subproblem solutions. The weighted sum approach combines two desirable properties. It guarantees to pre-

serve those parts of the solutions that are in agreement in the overlap between the subproblems, and it does not require a manual parameter selection.

The reduction in memory is paid by a higher computation time: CLG registration takes 12 times longer than a corresponding global registration. However, our focus is not on computation time, and only little effort has been made in its optimization.

In the CLG registration, each subdomain is computed independently, which is an opportunity to save computation time by parallelization. In this work, we focus on the computation on limited hardware. Many methods have been proposed to facilitate the parallel computation of nonlinear problems [Koh07; Dol16; Kla17]. In these methods, communication between computing nodes is a common bottleneck. Due to its independent computation on subdomains, the CLG method avoids communication. If multiple computing nodes are available, the computation can be distributed with little effort because only the initial distribution of the image data and the final combination step requires communication. Parallel computation on 16 nodes would likely reduce the computation time below the time needed for a single node fine global registration. The computation effort of the small-scale blending problem is negligible compared to the image registration problems.

The low memory requirement enables the registration of large-scale images on computers with limited memory, leading to superior accuracy compared to a coarse global registration.

We conduct the experiments by decomposing the image domain into two (Chapter 6.1) and four by four (Chapter 7) overlapping subdomains. A promising next step is the evaluation of a higher number of subdomains. The proportion of the overlap area is independent of the number of subdomains, such that no additional overhead is created. However, the ratio of high-resolution to low-resolution area decreases when the number of subdomains increases, which could affect the optimization. A pyramid scheme of nested subdomains of increasing resolutions could be added to reduce the resolution difference between the local image and its neighborhood.

We successfully apply the new method to practical large-scale image registration problems. Our example data includes six slide pairs from the pathological routine that were acquired in two different laboratories using four different stain combinations. Here, the CLG method outperforms a low-resolution approach and reaches comparable results to a fine global registration. A preliminary version of the method has been applied to a machine learning problem and this work is published together with Bulten et al. [Bul19].

Conclusion and Outlook

The presented CLG method solves the large-scale image registration problem that would otherwise exceed the available memory of the computer.

The CLG method decomposes the problem into multiple local, nonlinear subproblems on subdomains that are solved sequentially. Due to the global component in

the CLG registration, the mismatch between the local solutions is small and can be blended by an inexpensive weighted sum of the deformations. The small mismatch between the subdomains is the main advantage of the CLG method compared to a purely local subdomain registration, where the mismatch is much larger and cannot be compensated by blending the deformations.

The ability to break down the large-scale problem into smaller subproblems enables the parallelization of the registration on the GPU [Owe08]. The GPU architecture is distinguished from the CPU by its limited memory but high data-parallelism, which results in fast computation times and can make it an affordable alternative to a parallel cluster computer [Smi15]. Budelmann et al. [Bud19] report an implementation of a registration on the GPU that reduces the computation time to only 3% of the time of a comparable CPU implementation. First experiments suggest that a similar speedup can be reached for the CLG registration such that it outperforms a CPU-based fine global registration in computation time.

Large-scale image registration in digital pathology is already used to accelerate the development of machine learning in the field by supplying a reliable ground truth that is otherwise difficult to obtain [Jan16]. Combining multiple stainings, registration has been applied successfully in pharmaceutical research to discover prognostic biomarker combinations [Har18]. The presented method has the potential to enable the analysis of high-resolution data on less performant devices.

While this thesis focuses on images from digital pathology, the methods are directly applicable to other 2D datasets and can be extended to higher-dimensional data. Outside of image registration, related large-scale inverse problems occur, for example, in the reconstruction of astronomy images or environmental forecasting and modeling [Sla19; Chu19; Law13].

Massively large datasets of multiple terabytes or even petabytes occur, e.g., in the 3D reconstruction of detailed image stacks in electron microscopy [Saa12] and whole-brain sections in light-microscopy [Moh16]. In these applications, processing the full-resolution data is challenging, even on a cluster computer. An extension of the CLG method that is adapted to these particular problems could increase the accuracy in 3D reconstruction and lead to new insights in medical and non-medical research.

In summary, the presented CLG method solves the large-scale image registration problem that would otherwise exceed the available memory of the computer. In the application to large-scale histopathological research data, it leads to comparable results with the state of the art. With its modest hardware requirements, CLG reduces the need for high-performance hardware. It thus takes a step towards more sustainable computing.

Gallery of Image Registration Results

The following figures show registered image pairs in a diagonal overlay. The lower-left half of the image shows the reference image while the registered template image is shown in the upper-right. At the border of the two images, the quality of the alignment of corresponding structures can be observed.

For each slide, multiple images are shown at increasingly fine resolutions. Each image is a 25% crop of its predecessor, zooming in at a constant position of the slide. A black scale bar of 500 µm width is shown for scale in all images.

The position is chosen such that it contains the intersection of 4 subdomains at its center. As a default, we show the center of the slide. If the center of the slide contains no tissue or is of bad contrast, one of the surrounding subdomain intersections is used.

At the border of the two registered images, it can be observed that corresponding structures are well aligned in most of the examples. In some areas, no corresponding structure exists in the registered image. In these cases, a higher image resolution cannot lead to a better image alignment.

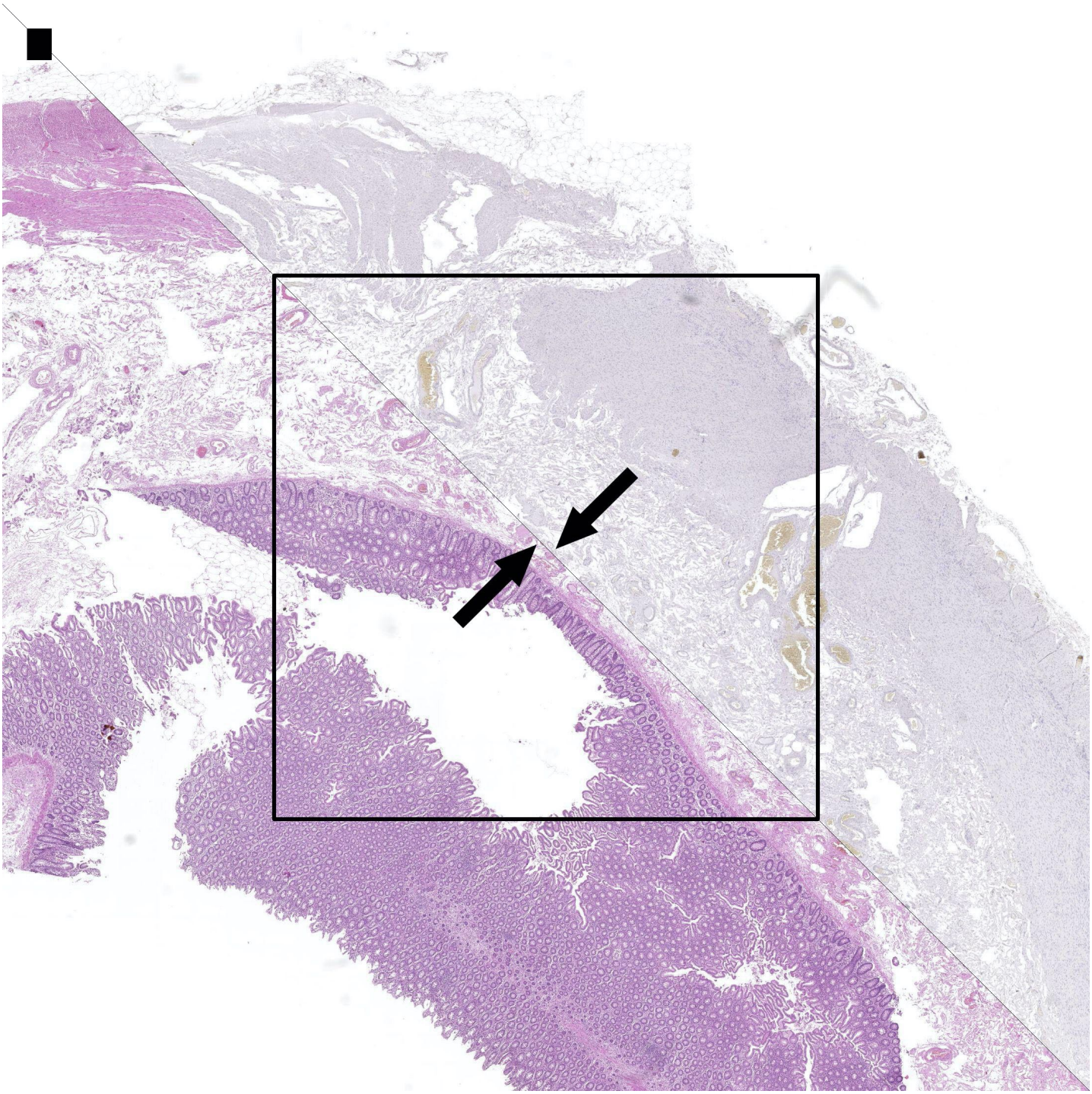


Figure G.1: Overlay of image pair 9515_21 after registration with CLG (registration image size <16k). Zoom level: 5. The black frame marks the image region of the following 25% crop in Figure *G*.2. The arrows indicate a well-aligned structure (left) and a visible difference at the inner boundary (right). Scale bar: 500µm.

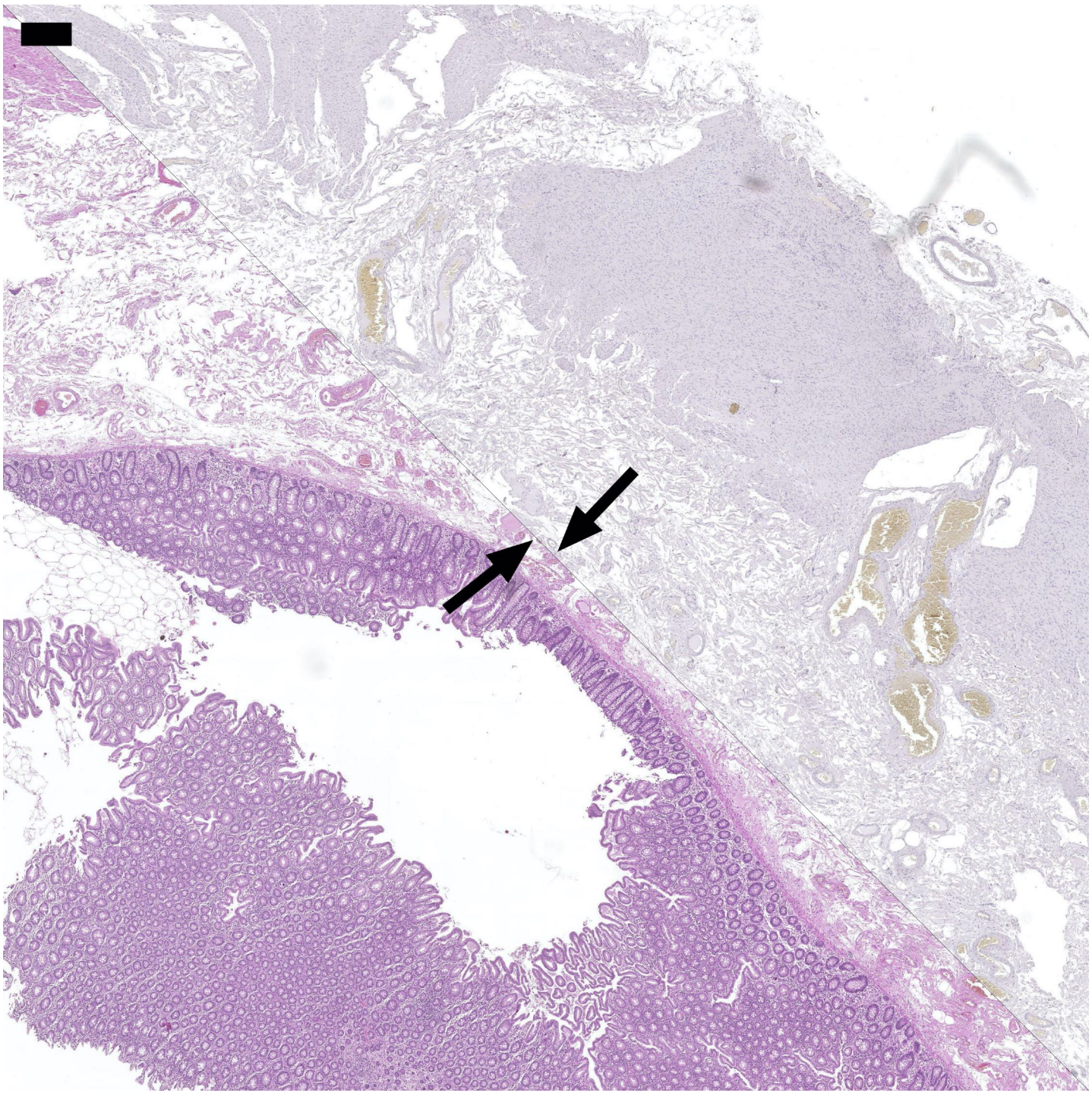


Figure G.2: Overlay of image pair 9515_21 after registration with CLG (registration image size <16k). Zoom level: 4. The arrows indicate a well-aligned structure (left) and a visible difference at the inner boundary (right). Scale bar: 500µm.

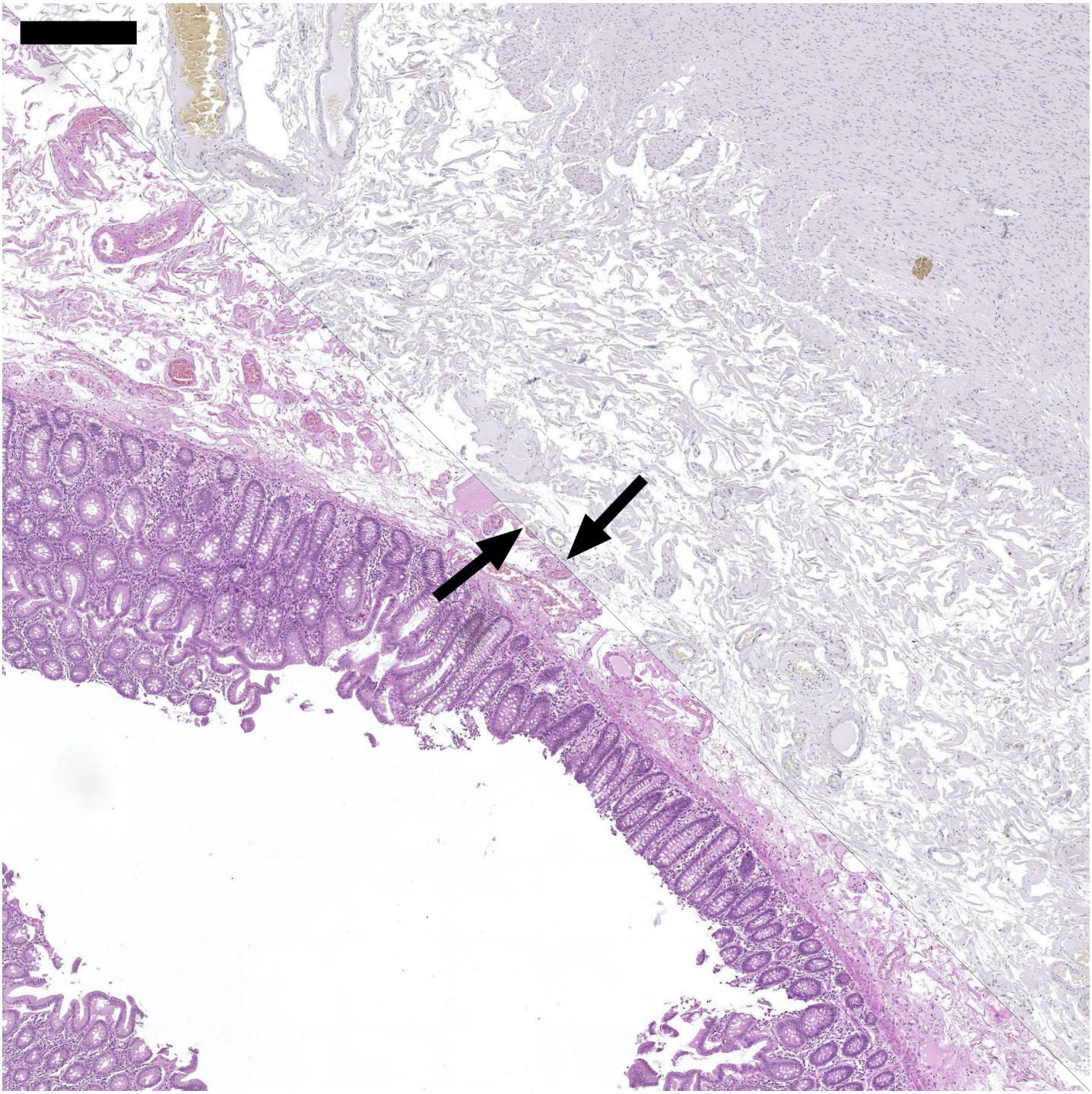


Figure G.3: Overlay of image pair 9515_21 after registration with CLG (registration image size <16k). Zoom level: 3. The arrows indicate a well-aligned structure (left) and a visible difference at the inner boundary (right). Scale bar: 500µm.

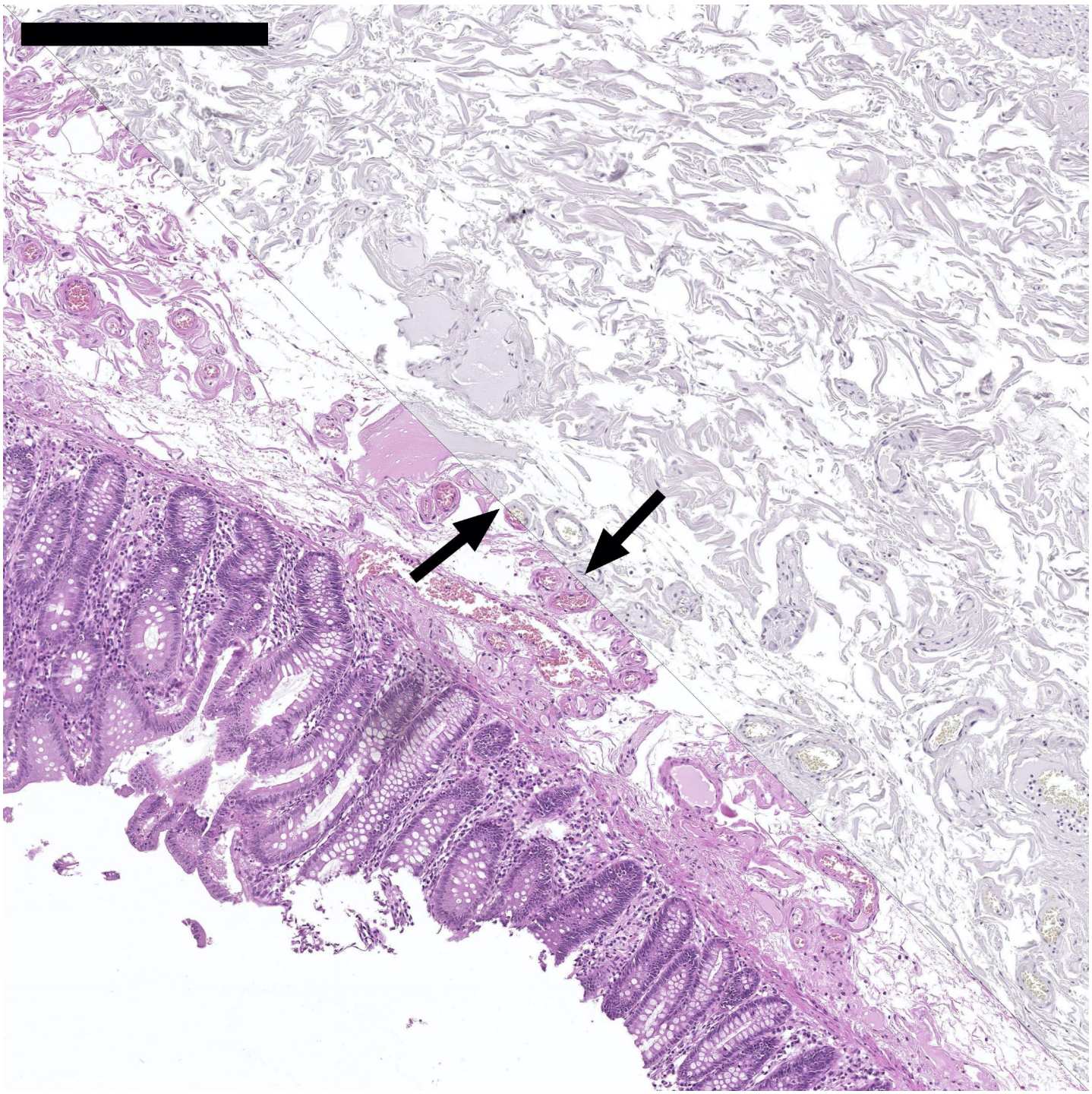


Figure G.4: Overlay of image pair 9515_21 after registration with CLG (registration image size <16k). Zoom level: 2. The arrows indicate a well-aligned structure (left) and a visible difference at the inner boundary (right). Scale bar: 500µm.

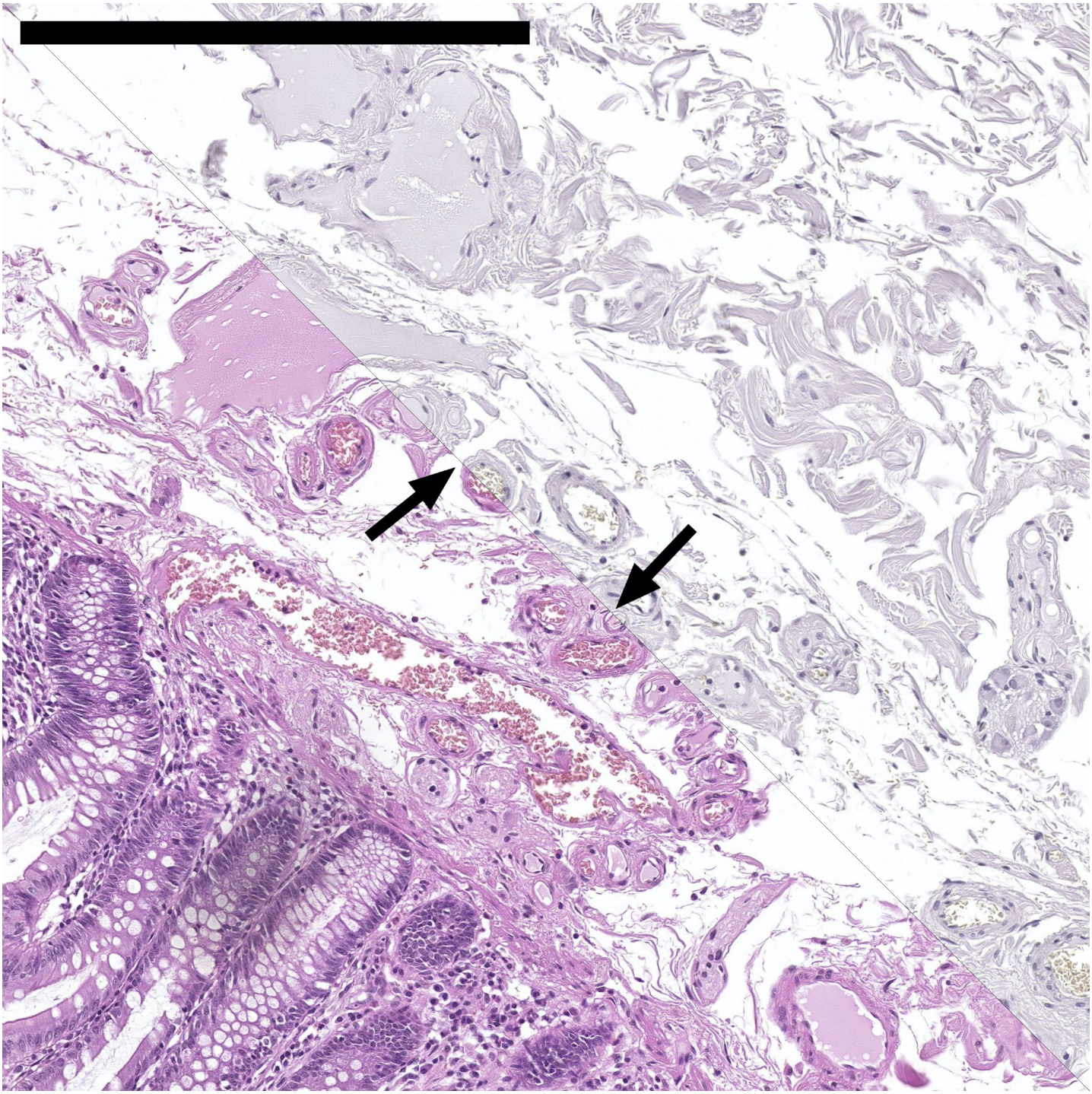


Figure G.5: Overlay of image pair 9515_21 after registration with CLG (registration image size <16k). Zoom level: 1. The arrows indicate a well-aligned structure (left) and a visible difference at the inner boundary (right). Scale bar: 500µm.

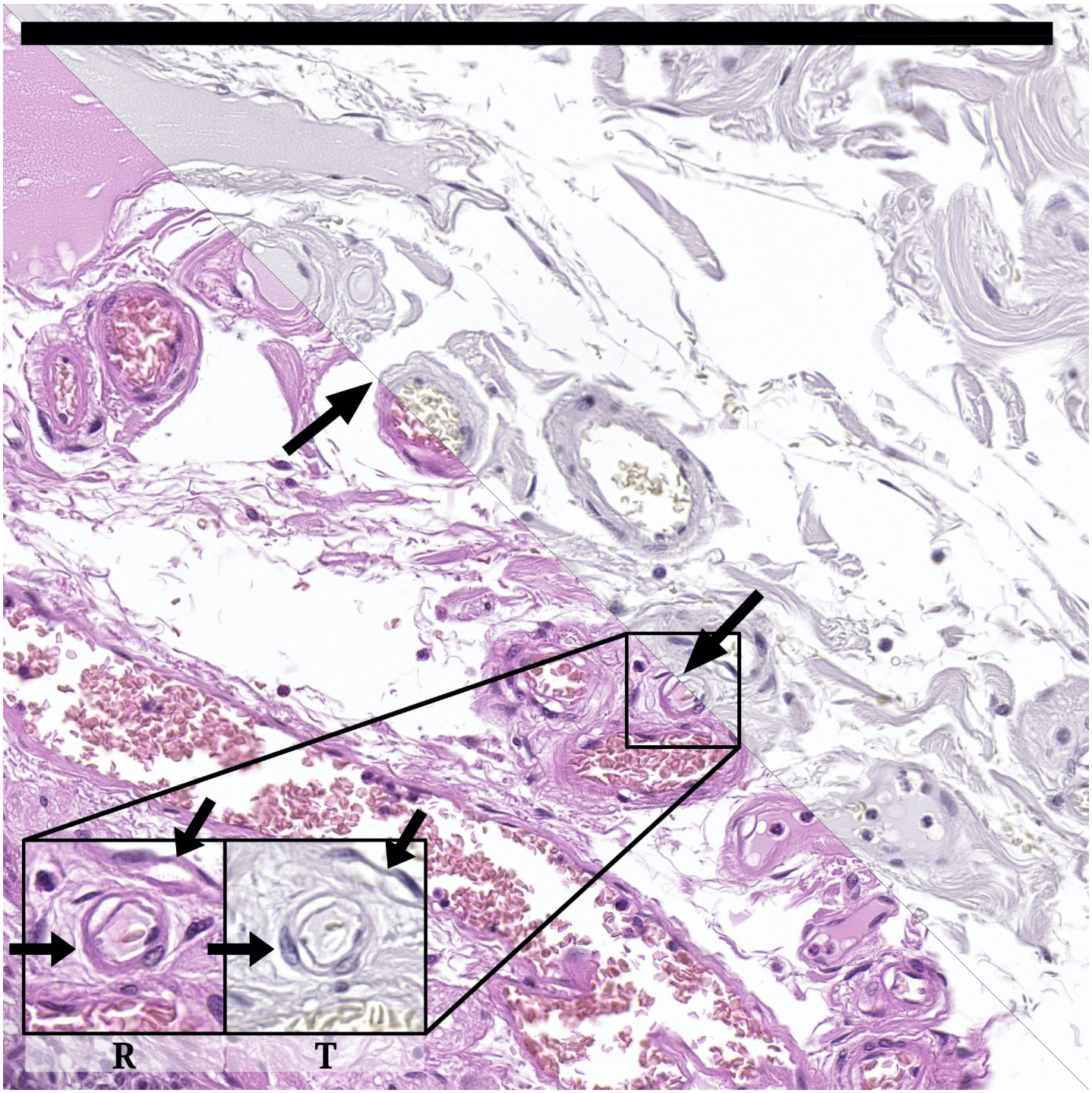


Figure G.6: Overlay of image pair 9515_21 after registration with CLG (registration image size <16k). Zoom level: 0. The large arrows indicate a well-aligned structure (left) and a visible difference at the inner boundary (right). The difference is due to a structural difference in the two sections. The two boxes in the lower left show R and $T(x)$ (undeformed). Small arrows indicate where the images show different cells and structures, which should be preserved by registration. Scale bar: 500µm.

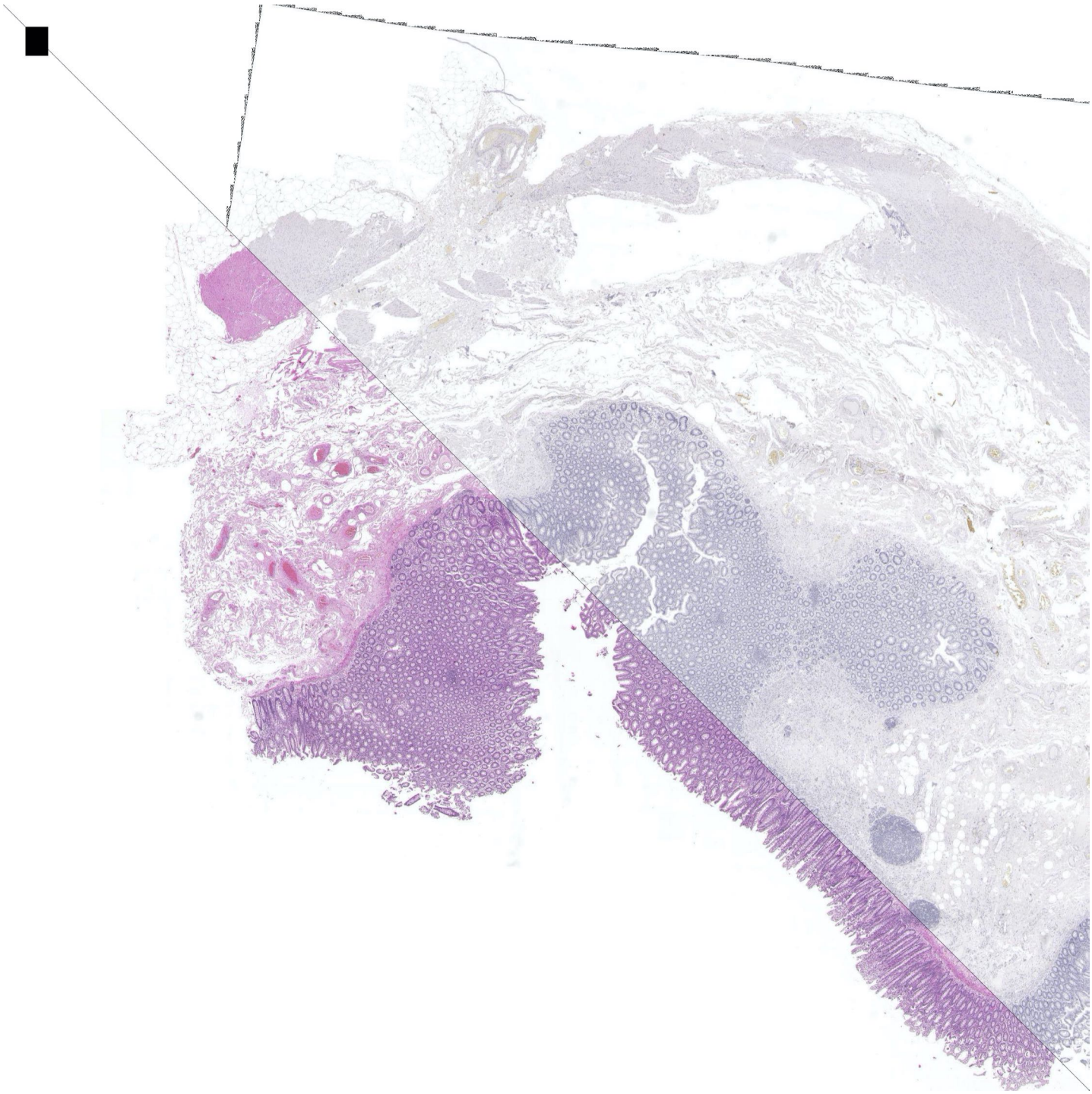


Figure G.7: Overlay of image pair 9515_20 after registration with CLG (registration image size <16k). Zoom level: 5. Scale bar: 500µm.

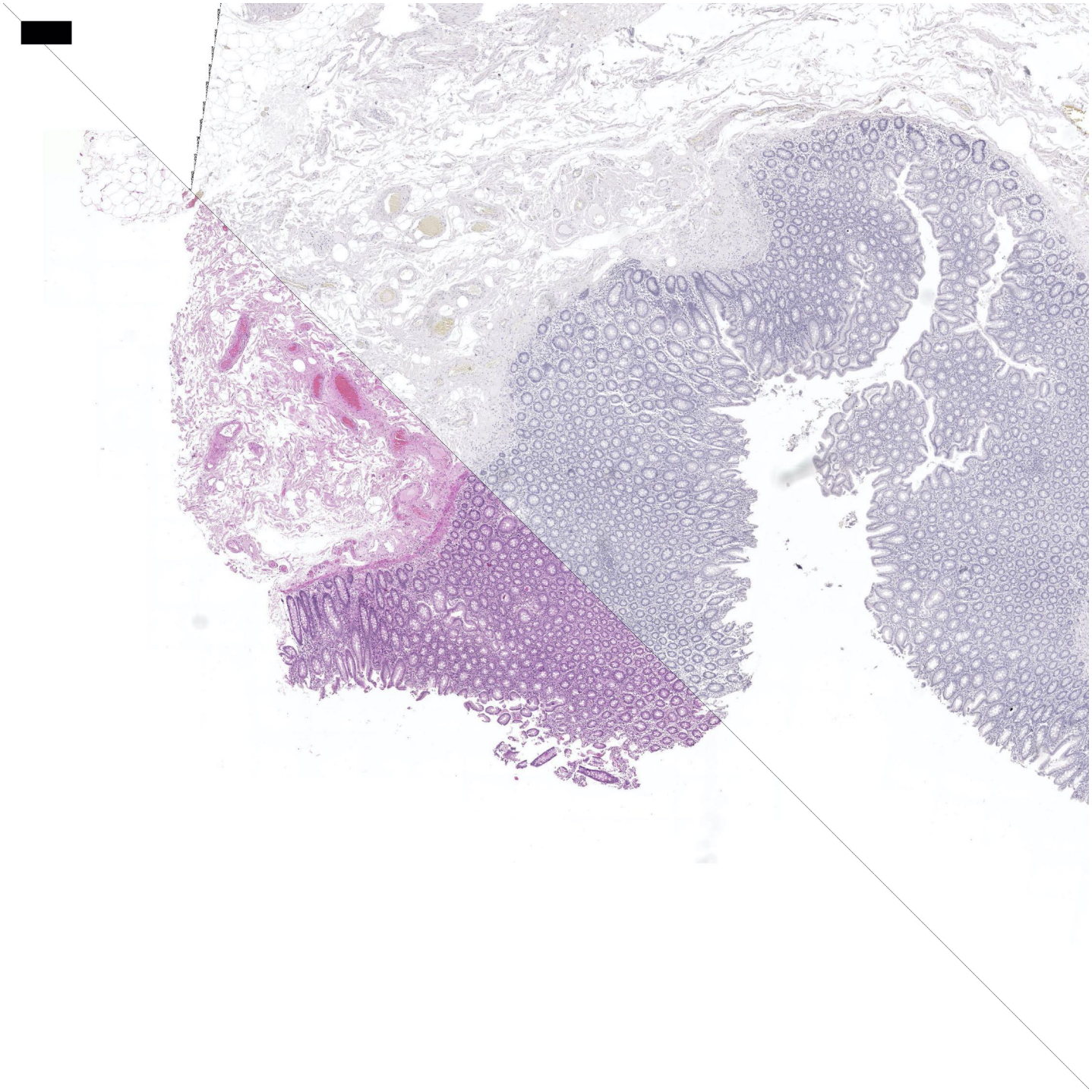


Figure G.8: Overlay of image pair 9515_20 after registration with CLG (registration image size <16k). Zoom level: 4. Scale bar: 500µm.

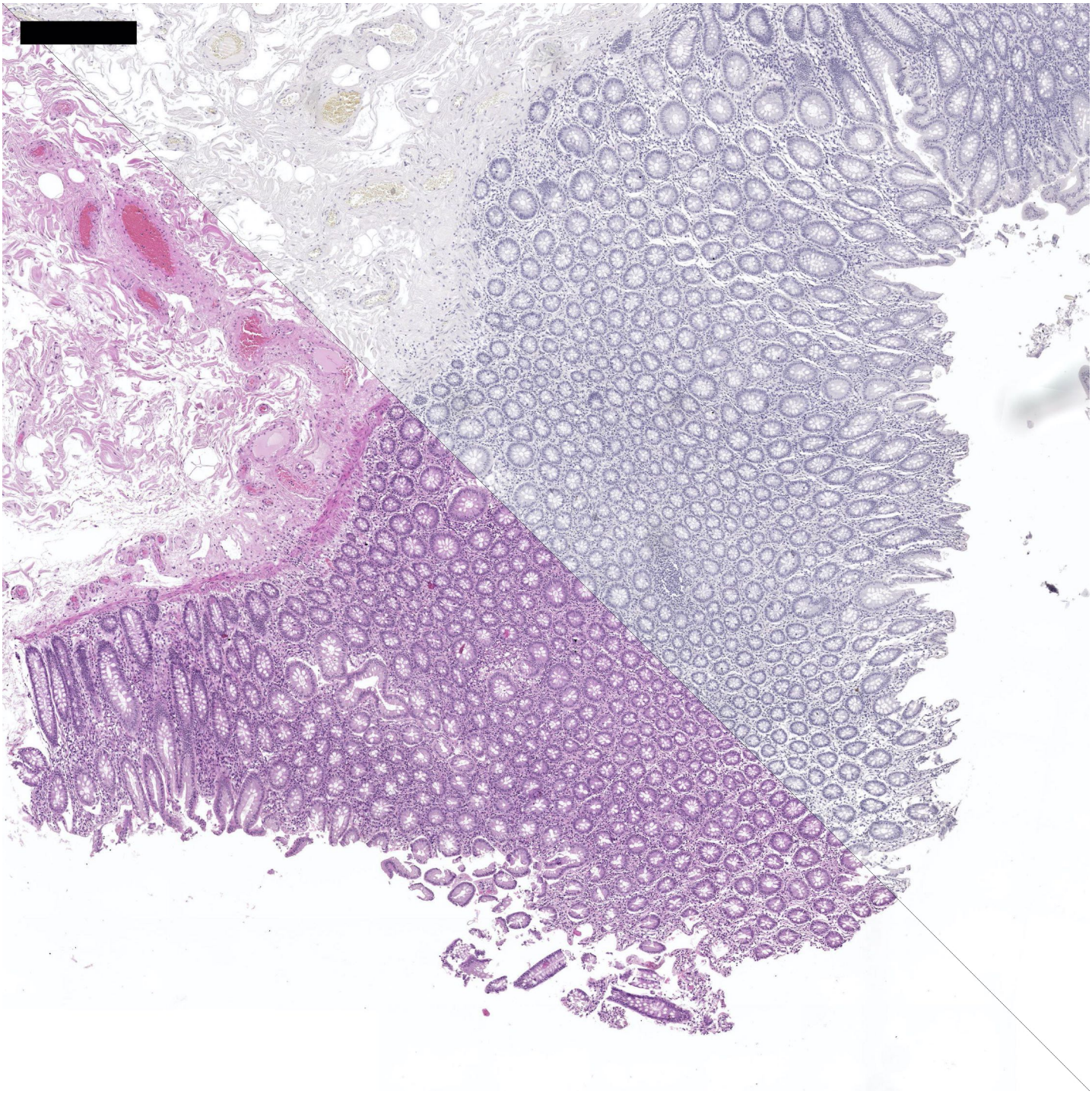


Figure G.9: Overlay of image pair 9515_20 after registration with CLG (registration image size <16k). Zoom level: 3. Scale bar: 500µm.

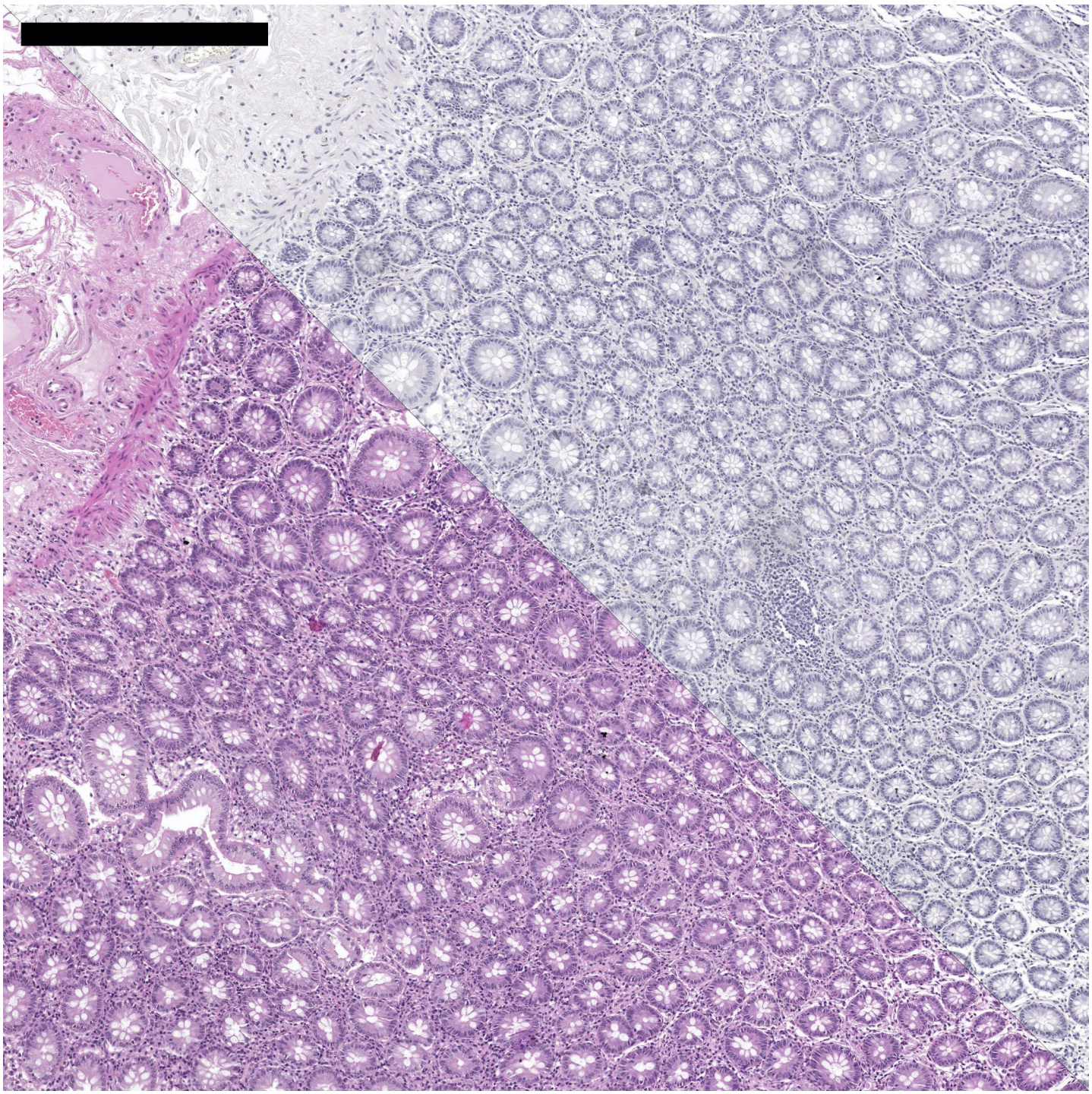


Figure G.10: Overlay of image pair 9515_20 after registration with CLG (registration image size <16k). Zoom level: 2. Scale bar: 500µm.

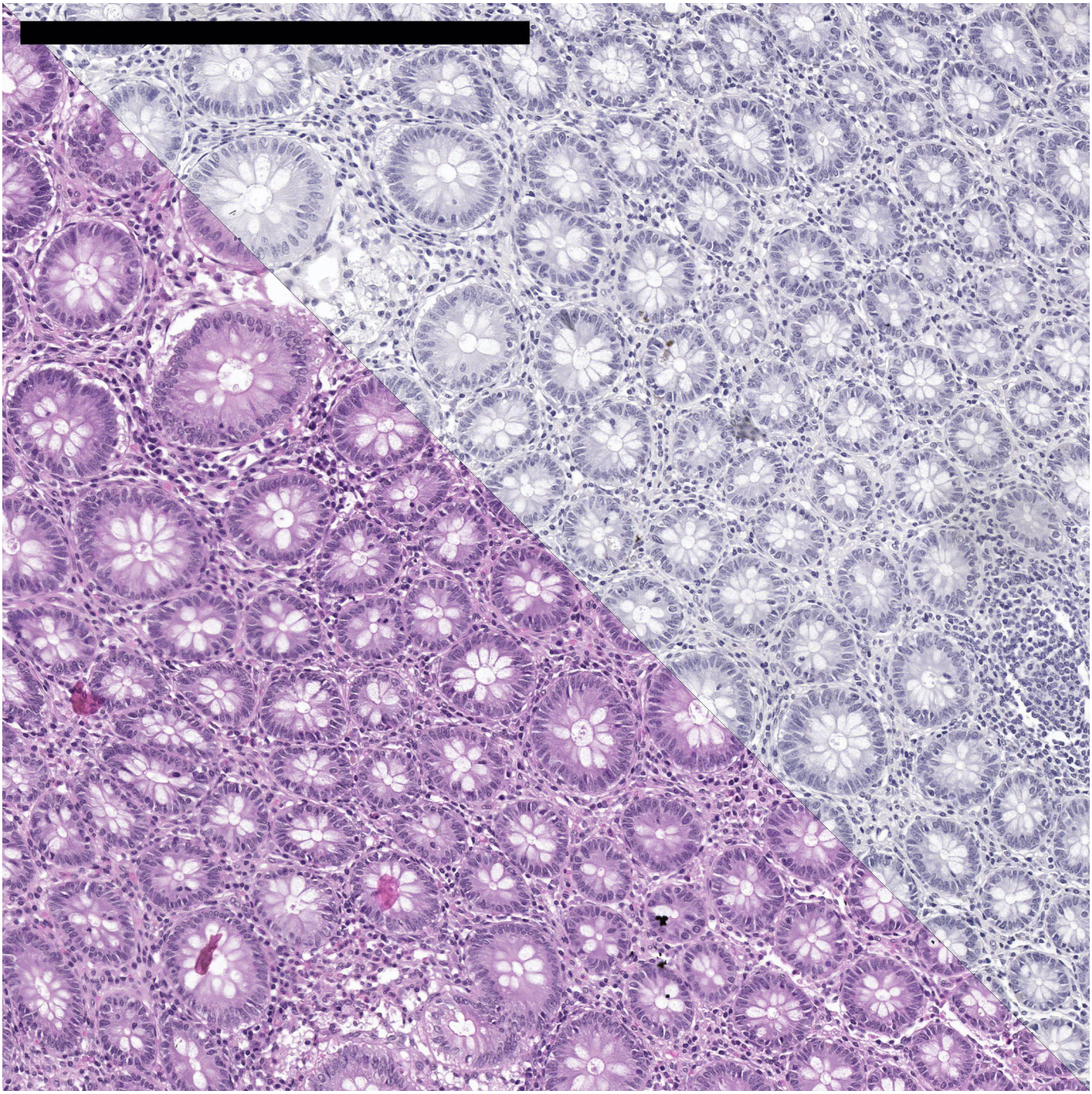


Figure G.11: Overlay of image pair 9515_20 after registration with CLG (registration image size <16k). Zoom level: 1. Scale bar: 500µm.

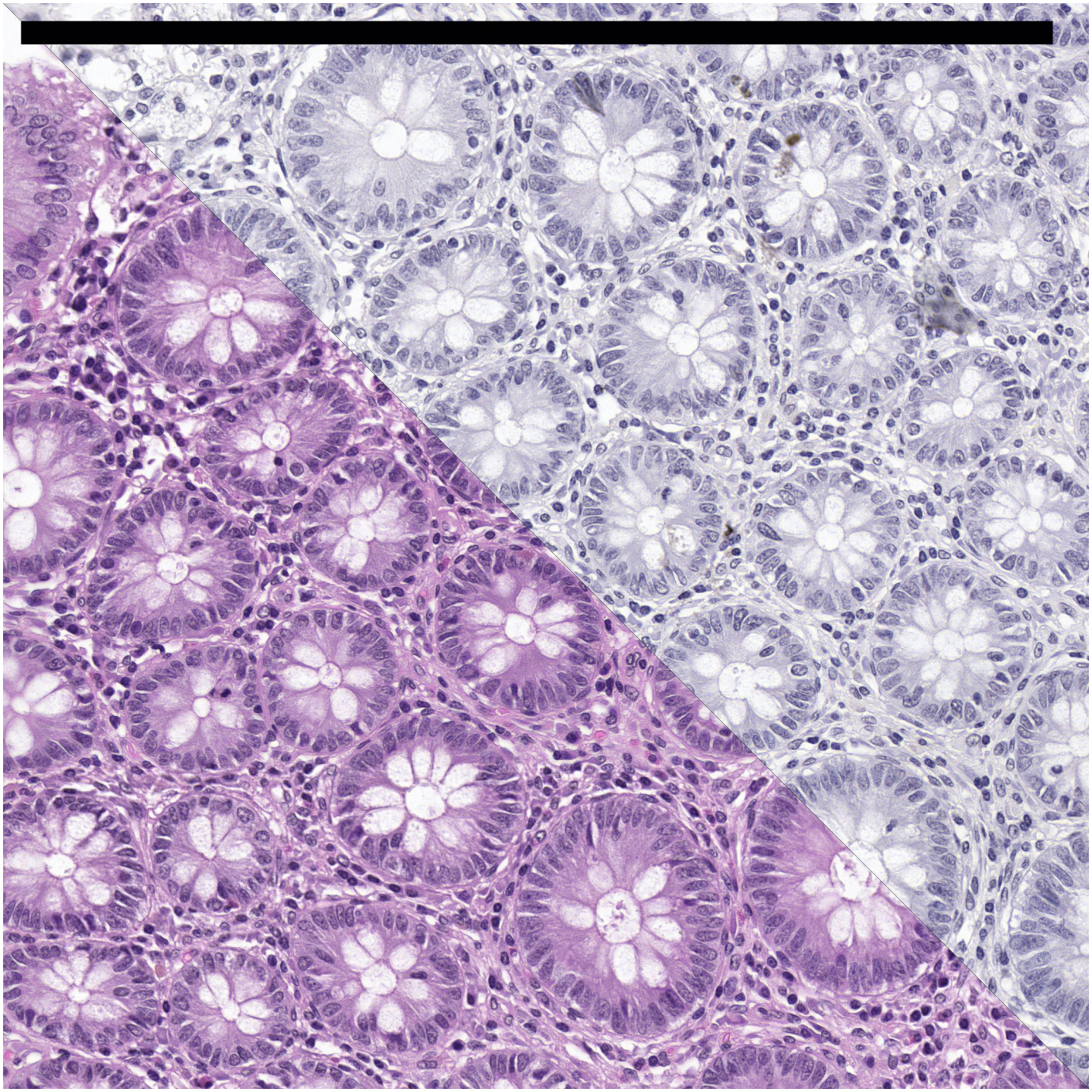


Figure G.12: Overlay of image pair 9515_20 after registration with CLG (registration image size <16k). Zoom level: 0. Scale bar: 500µm.

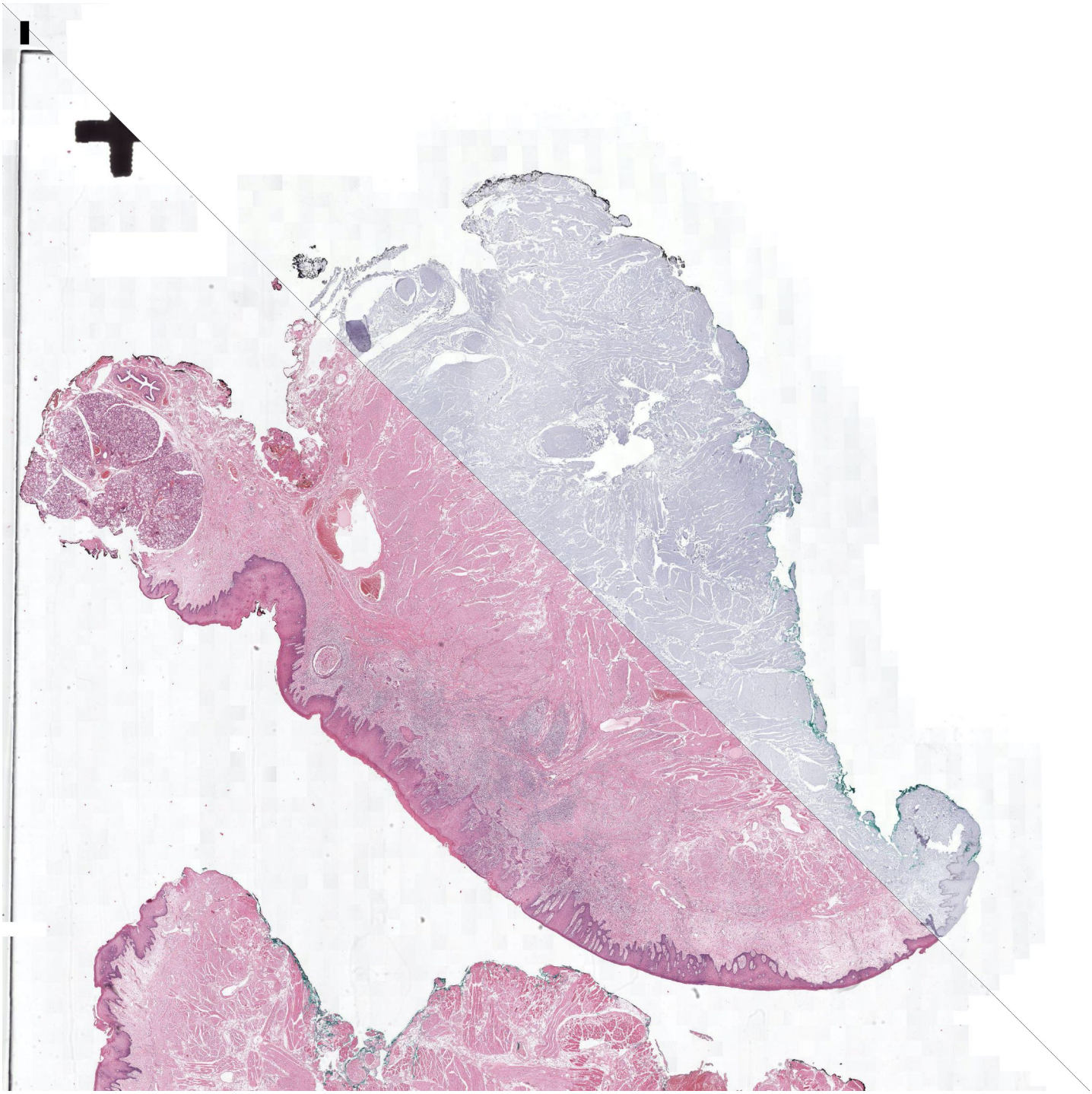


Figure G.13: Overlay of image pair 29 after registration with CLG (registration image size <16k). Zoom level: 5. Scale bar: 500µm.

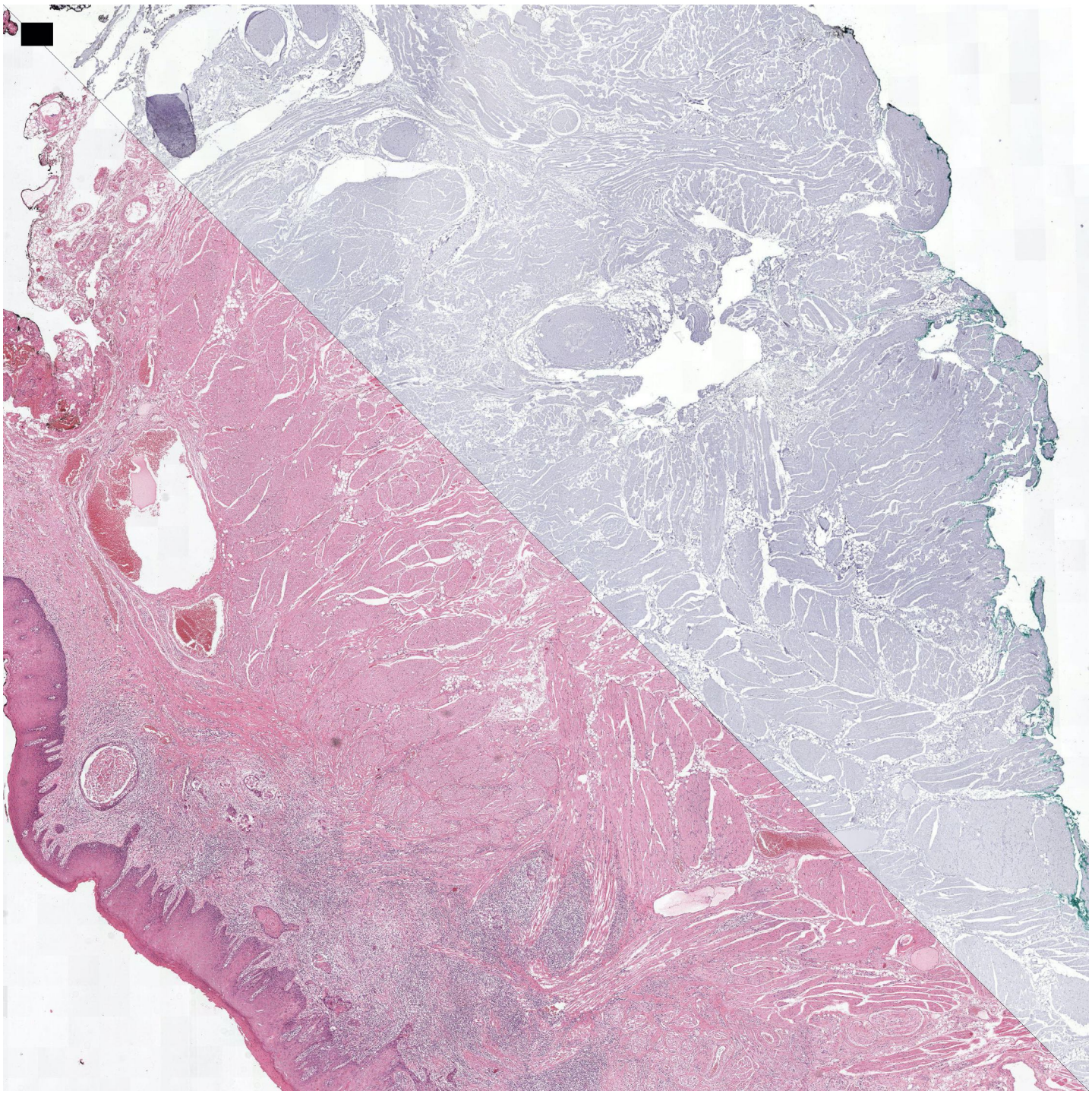


Figure G.14: Overlay of image pair 29 after registration with CLG (registration image size <16k). Zoom level: 4. Scale bar: 500µm.

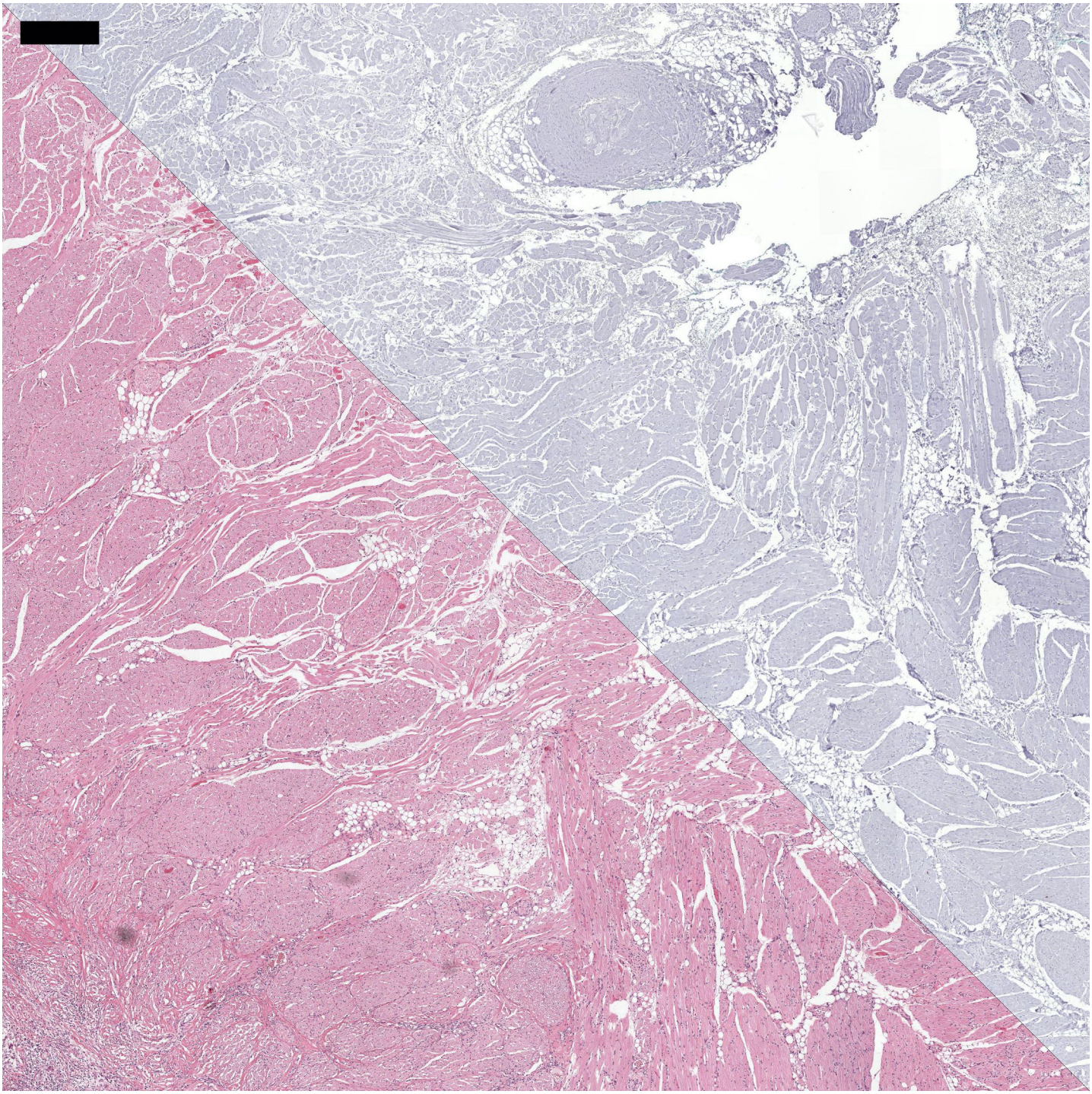


Figure G.15: Overlay of image pair 29 after registration with CLG (registration image size <16k). Zoom level: 3. Scale bar: 500µm.

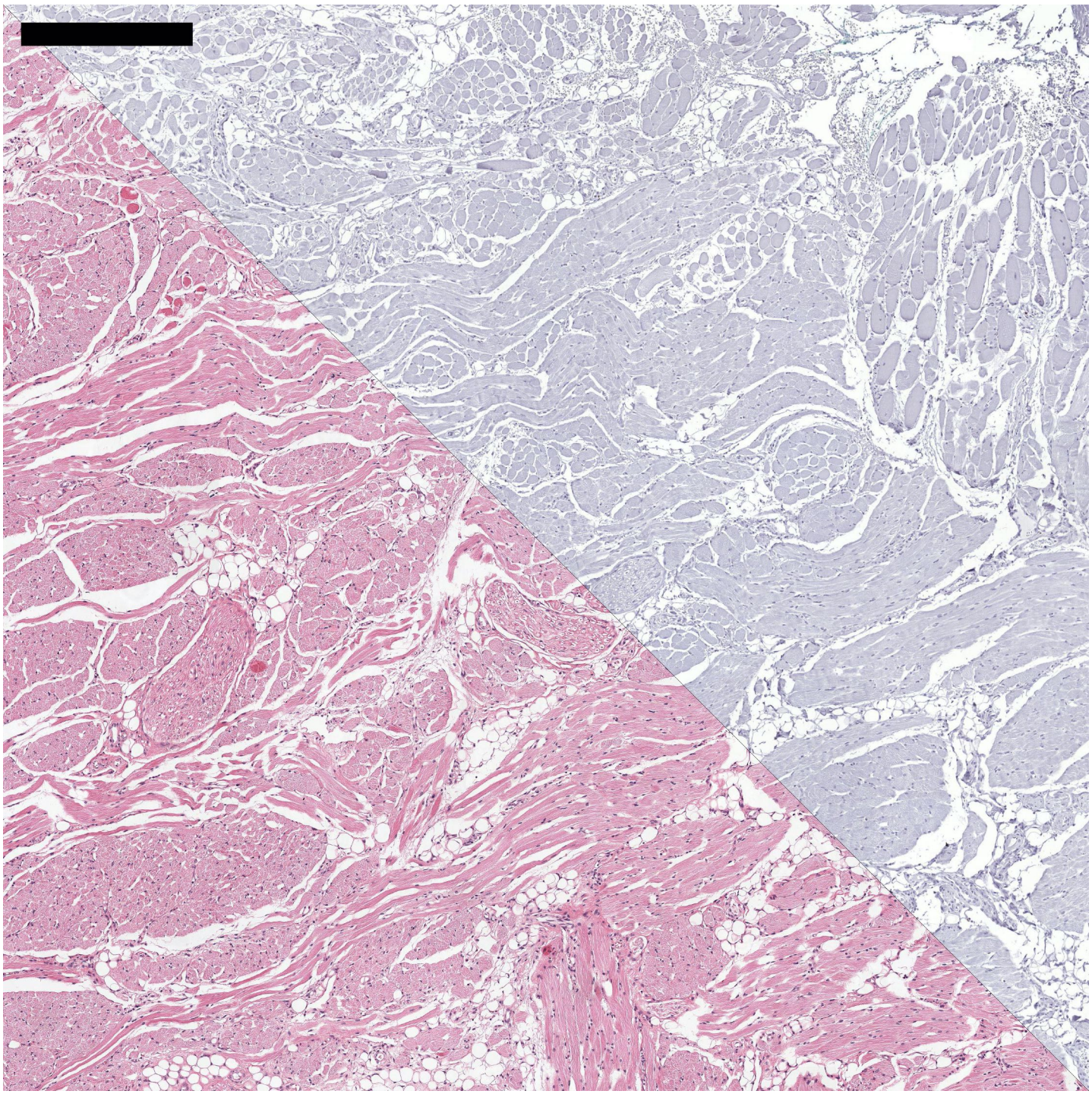


Figure G.16: Overlay of image pair 29 after registration with CLG (registration image size <16k). Zoom level: 2. Scale bar: 500µm.

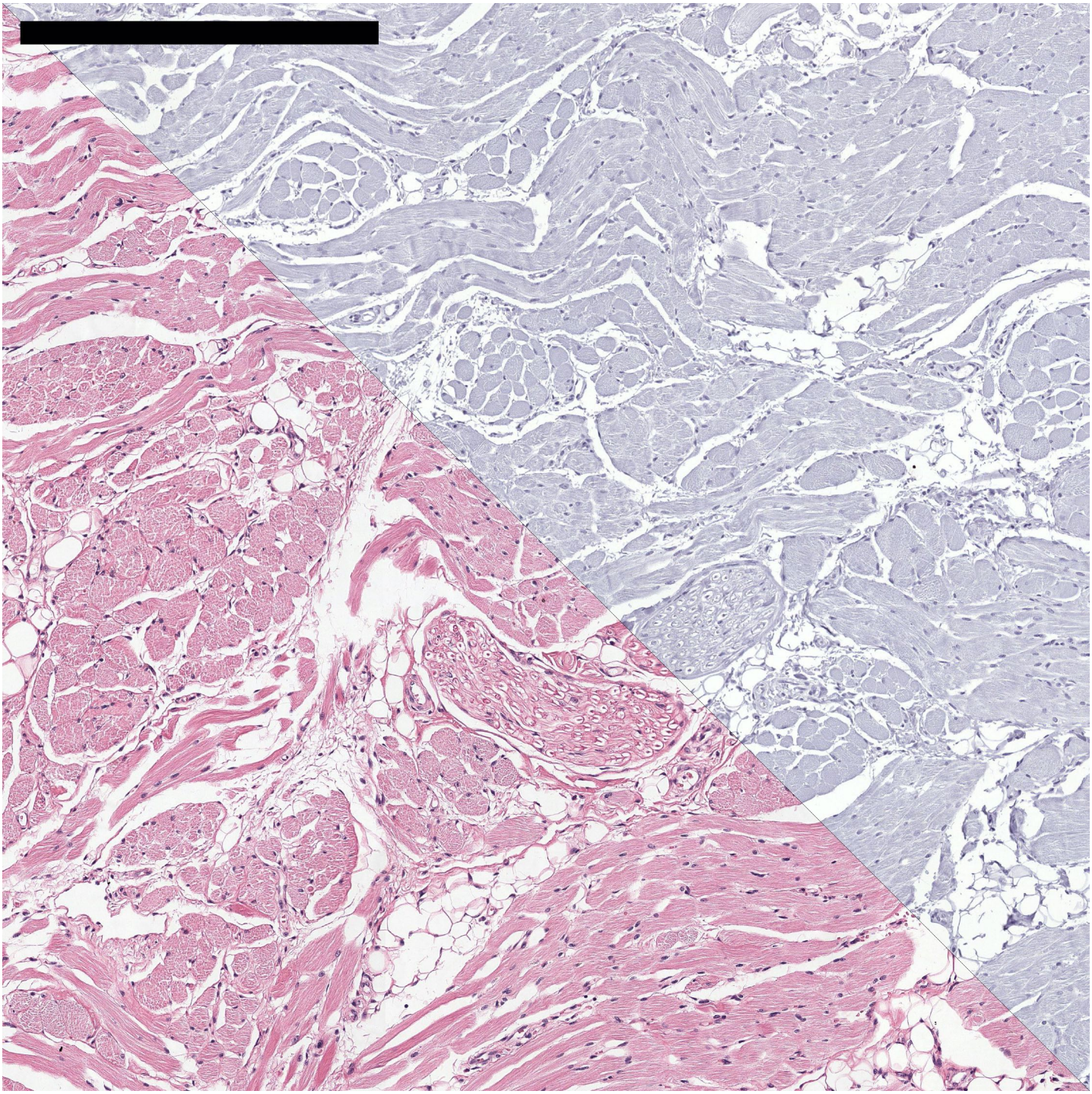


Figure G.17: Overlay of image pair 29 after registration with CLG (registration image size <16k). Zoom level: 1. Scale bar: 500µm.

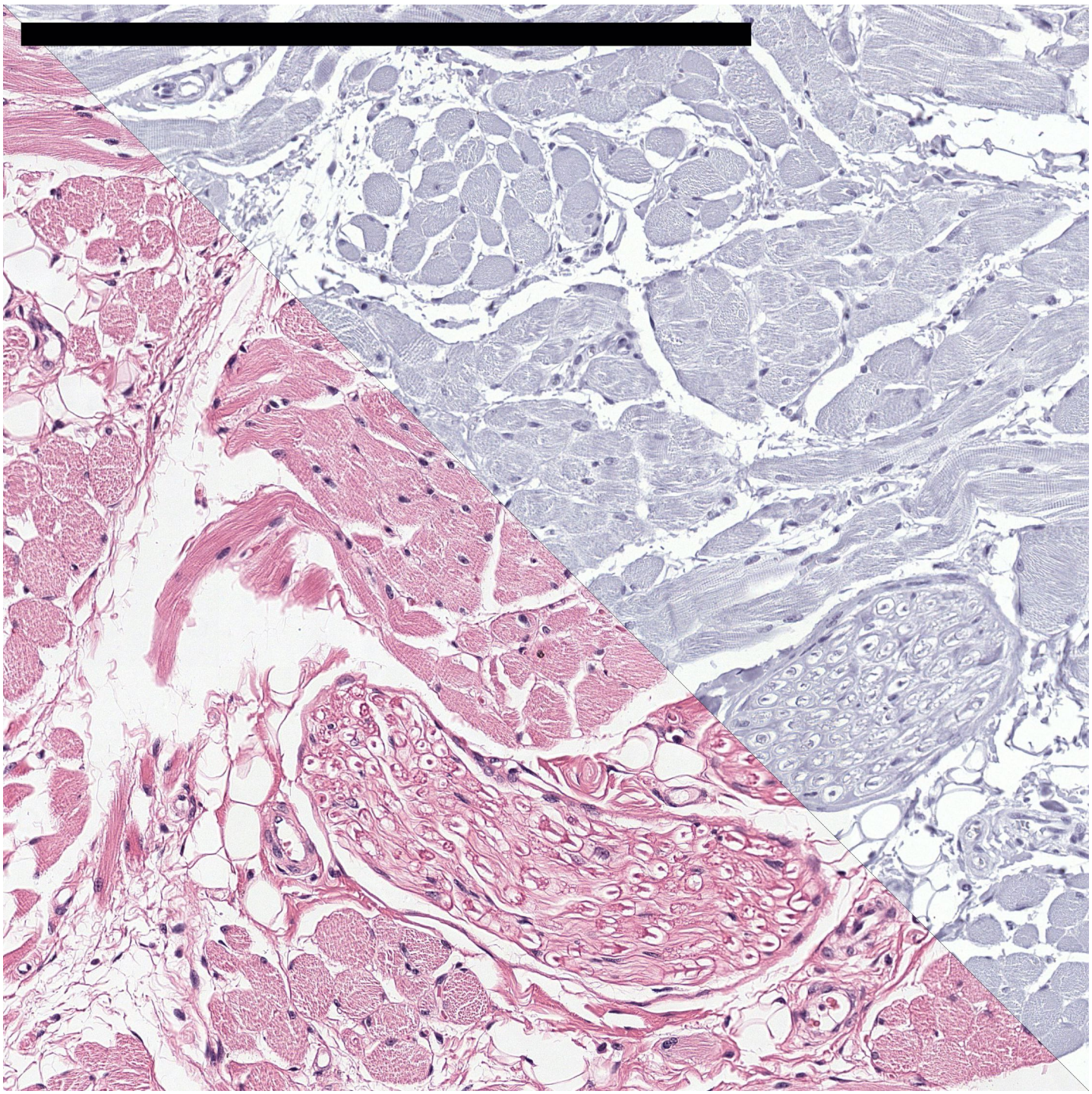


Figure G.18: Overlay of image pair 29 after registration with CLG (registration image size <16k). Zoom level: 0. Scale bar: 500µm.

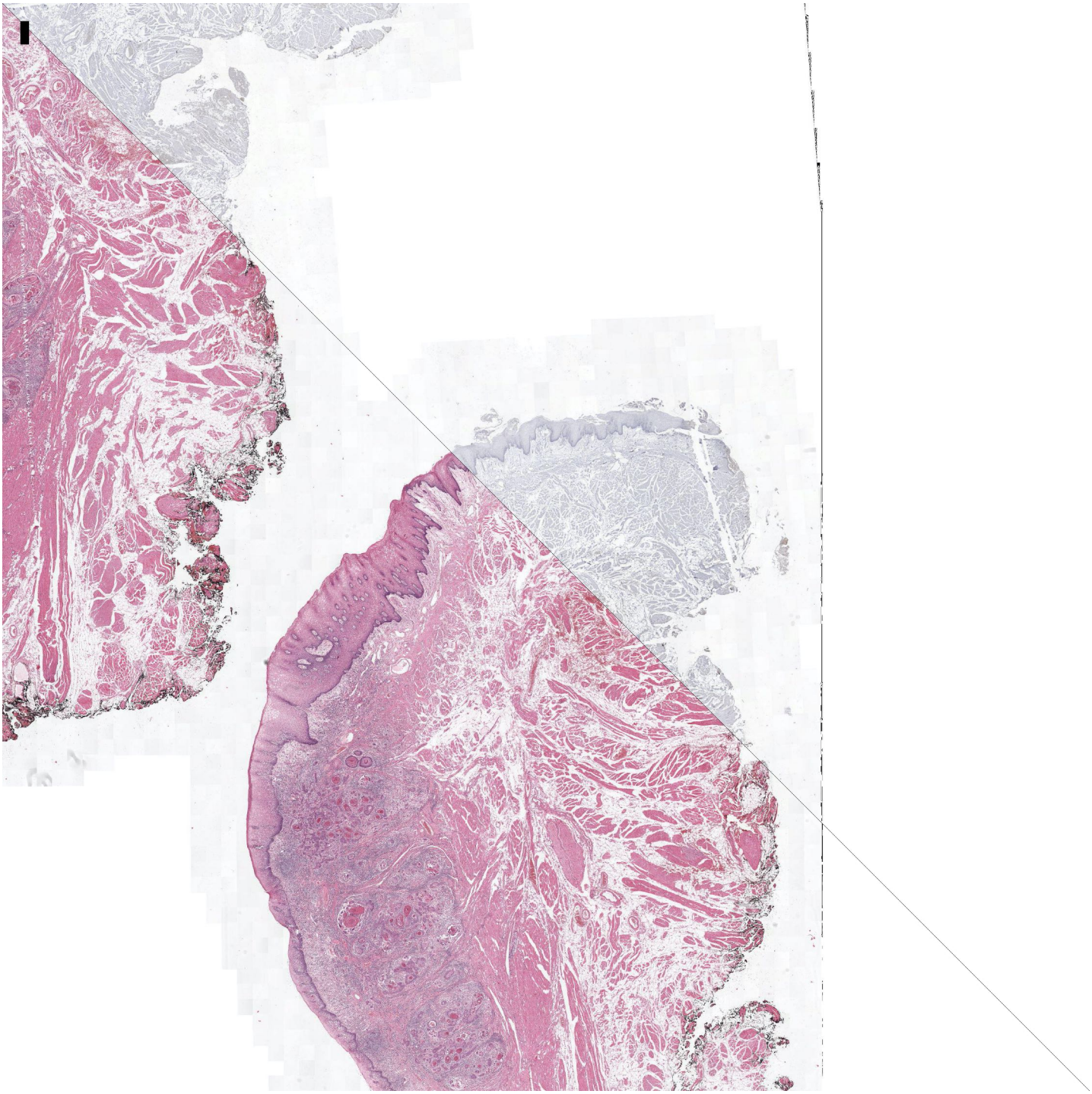


Figure G.19: Overlay of image pair 108 after registration with CLG (registration image size <16k). Zoom level: 5. Scale bar: 500µm.

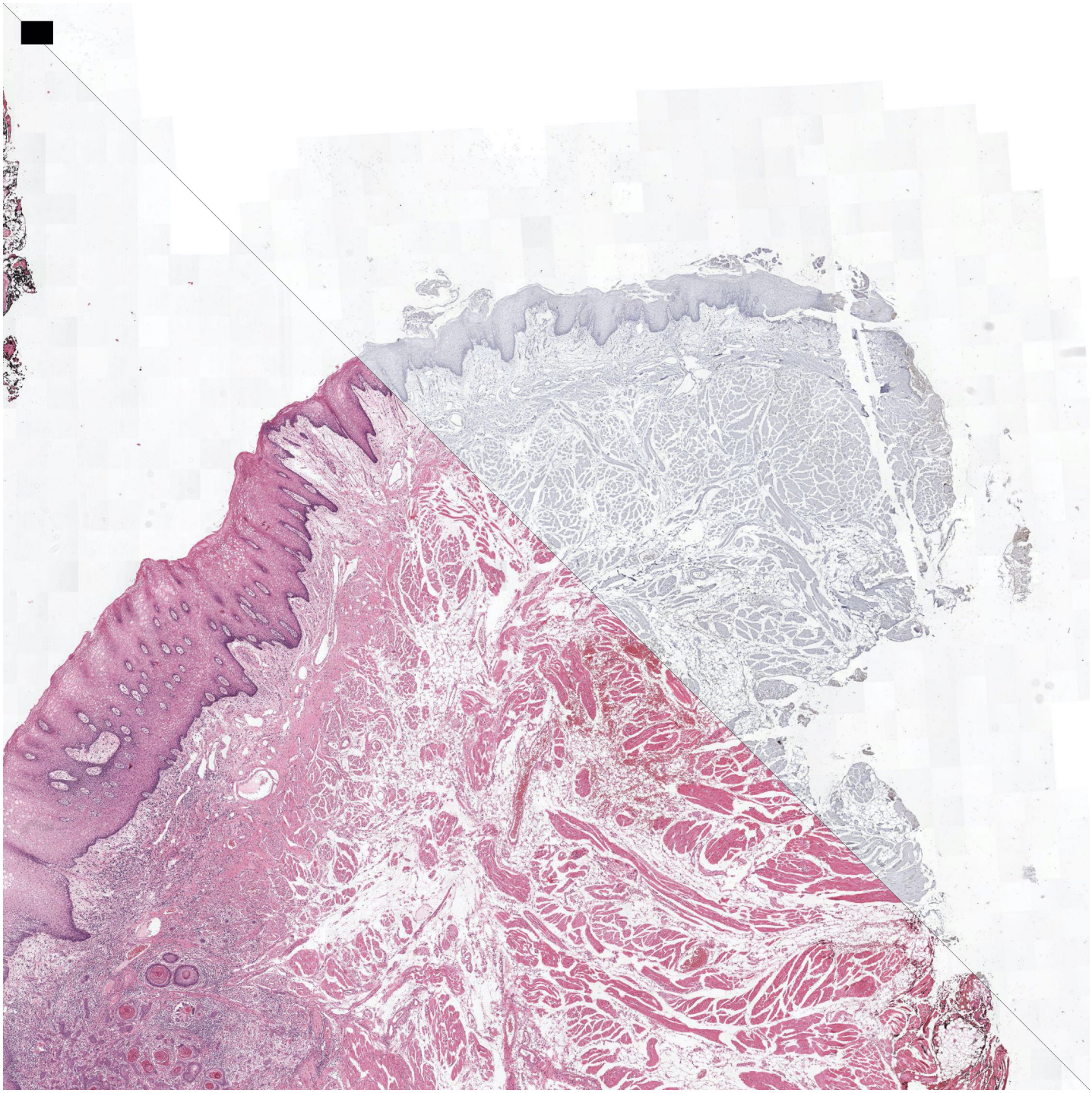


Figure G.20: Overlay of image pair 108 after registration with CLG (registration image size <16k). Zoom level: 4. Scale bar: 500µm.

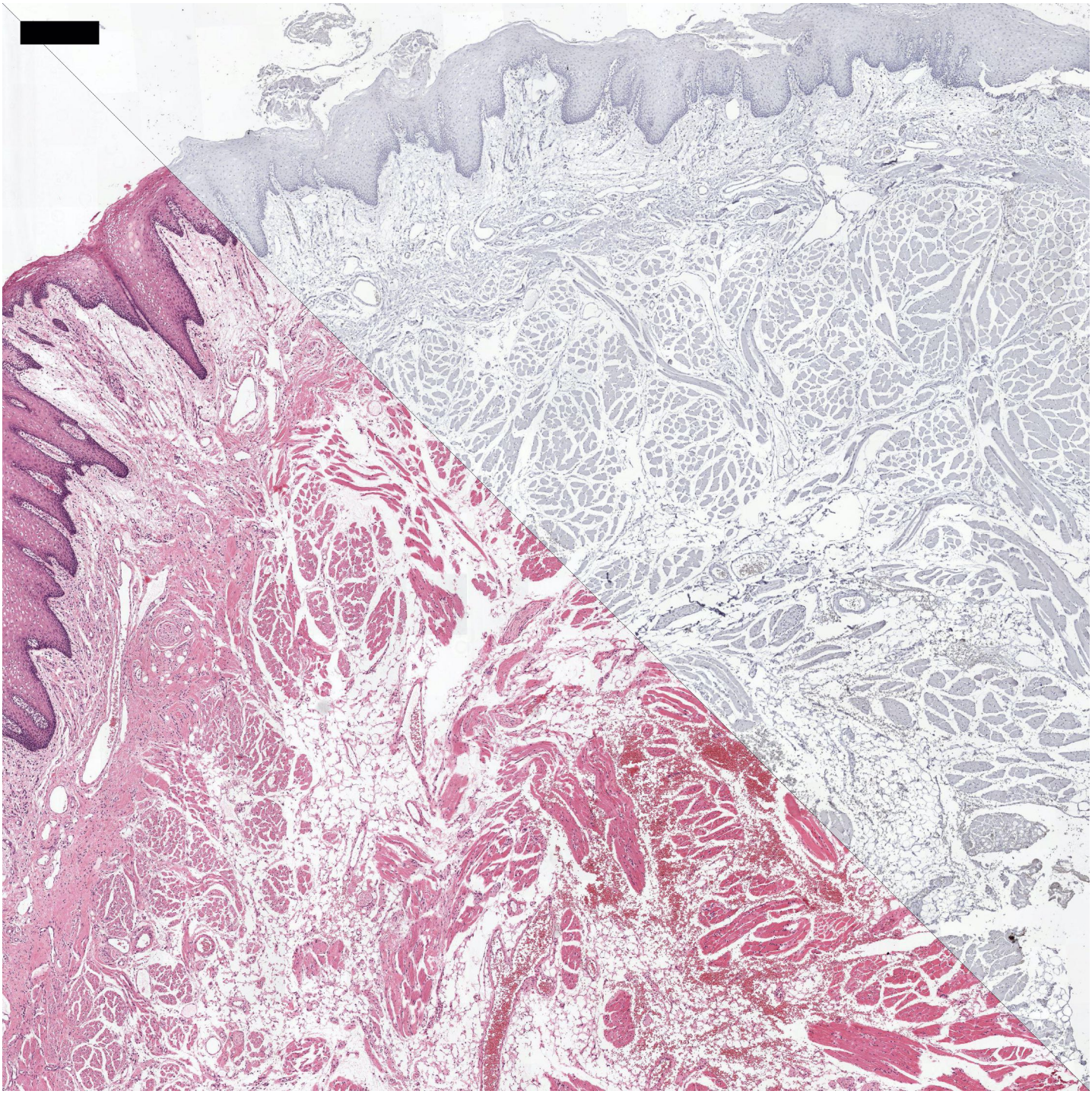


Figure G.21: Overlay of image pair 108 after registration with CLG (registration image size <16k). Zoom level: 3. Scale bar: 500µm.

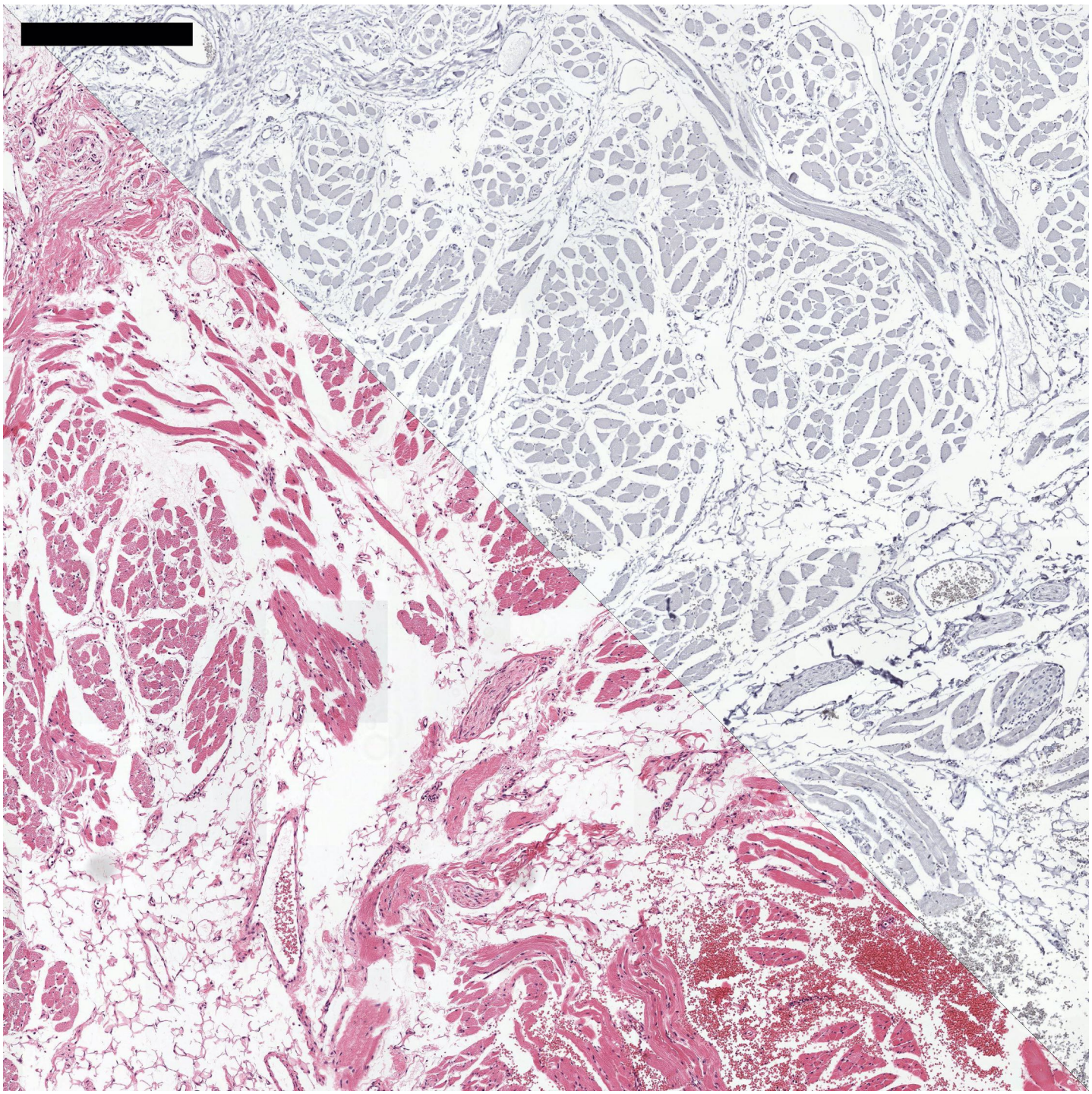


Figure G.22: Overlay of image pair 108 after registration with CLG (registration image size <16k). Zoom level: 2. Scale bar: 500µm.

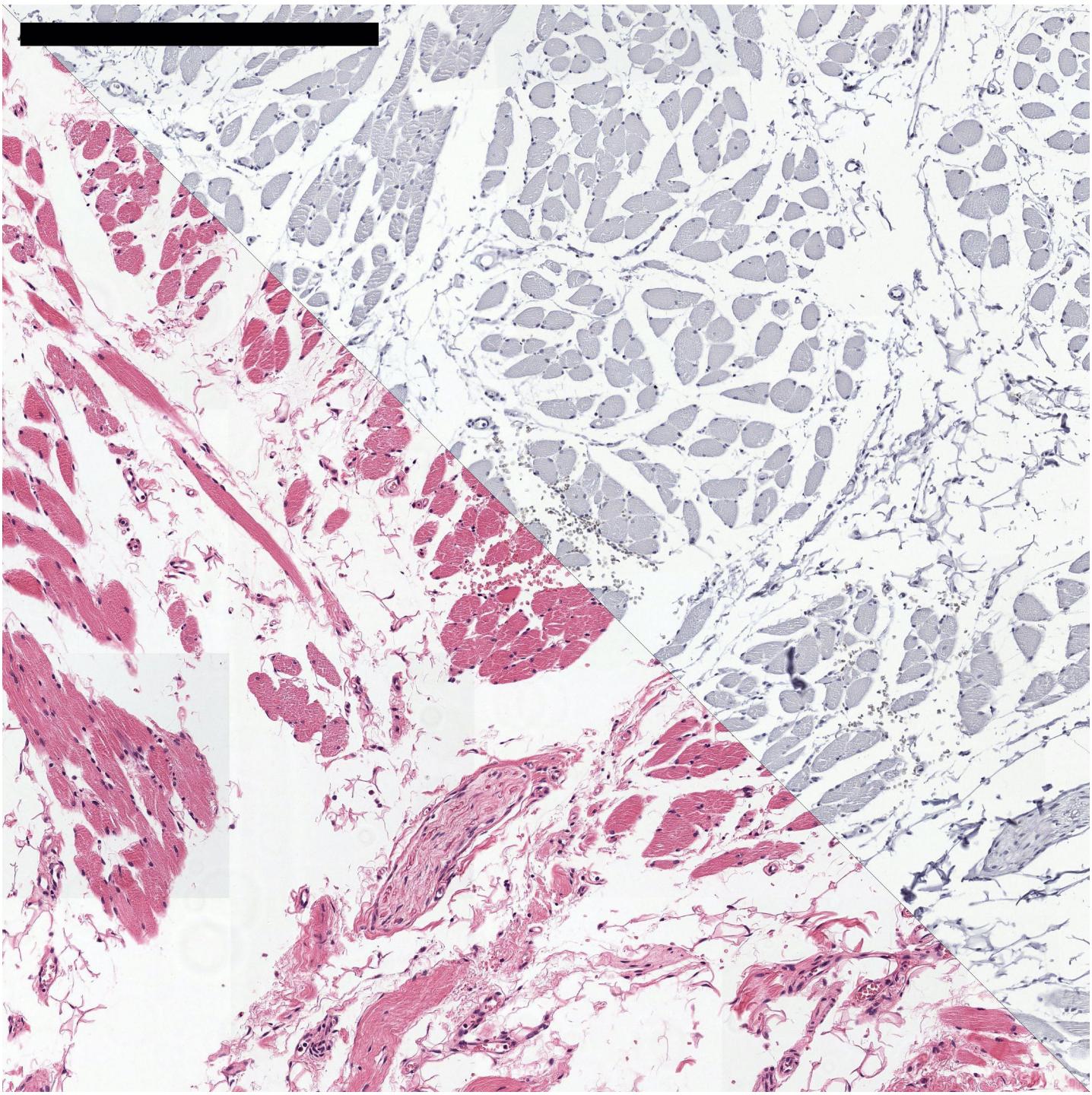


Figure G.23: Overlay of image pair 108 after registration with CLG (registration image size <16k). Zoom level: 1. Scale bar: 500µm.

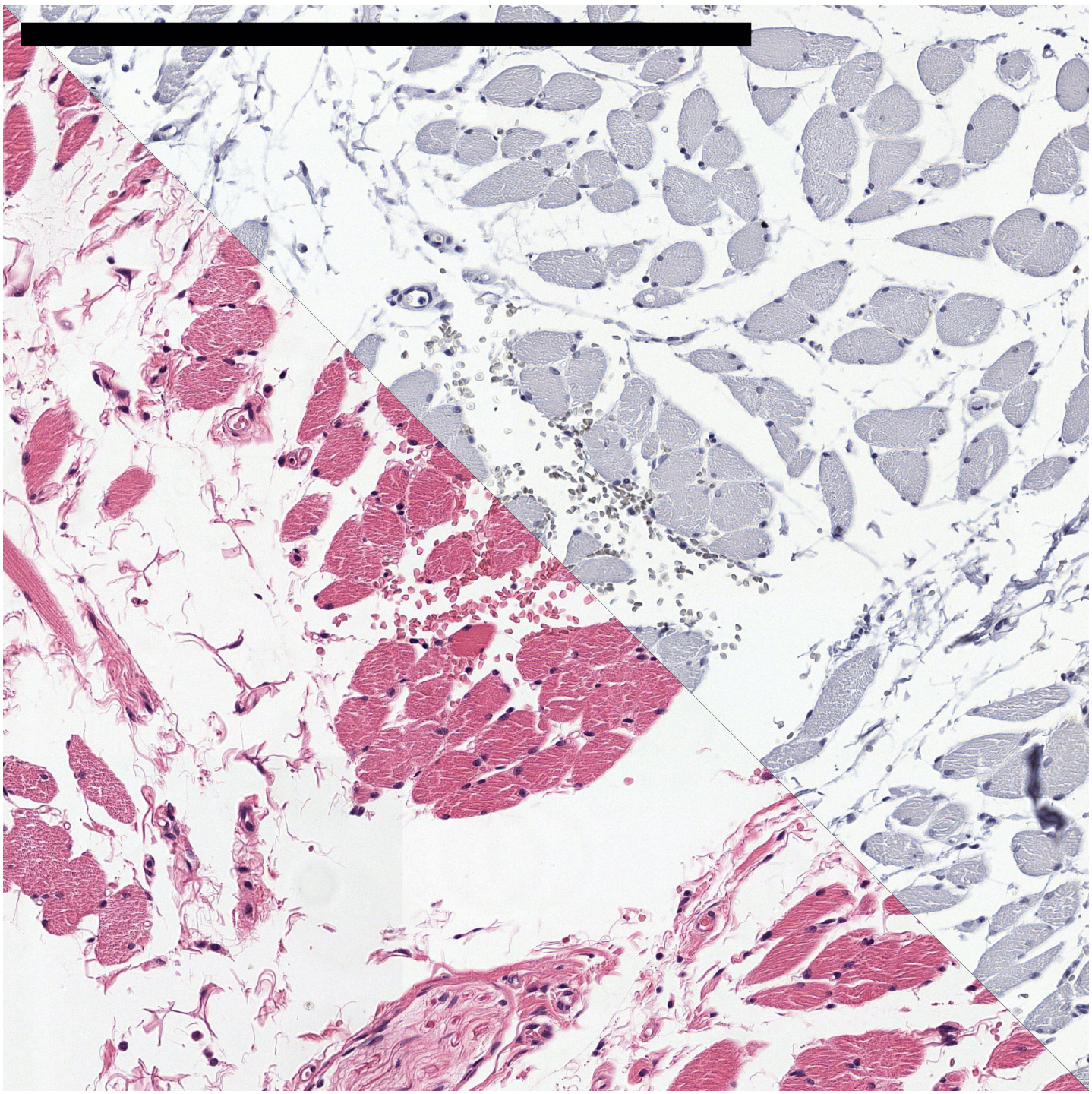


Figure G.24: Overlay of image pair 108 after registration with CLG (registration image size <16k). Zoom level: 0. Scale bar: 500µm.

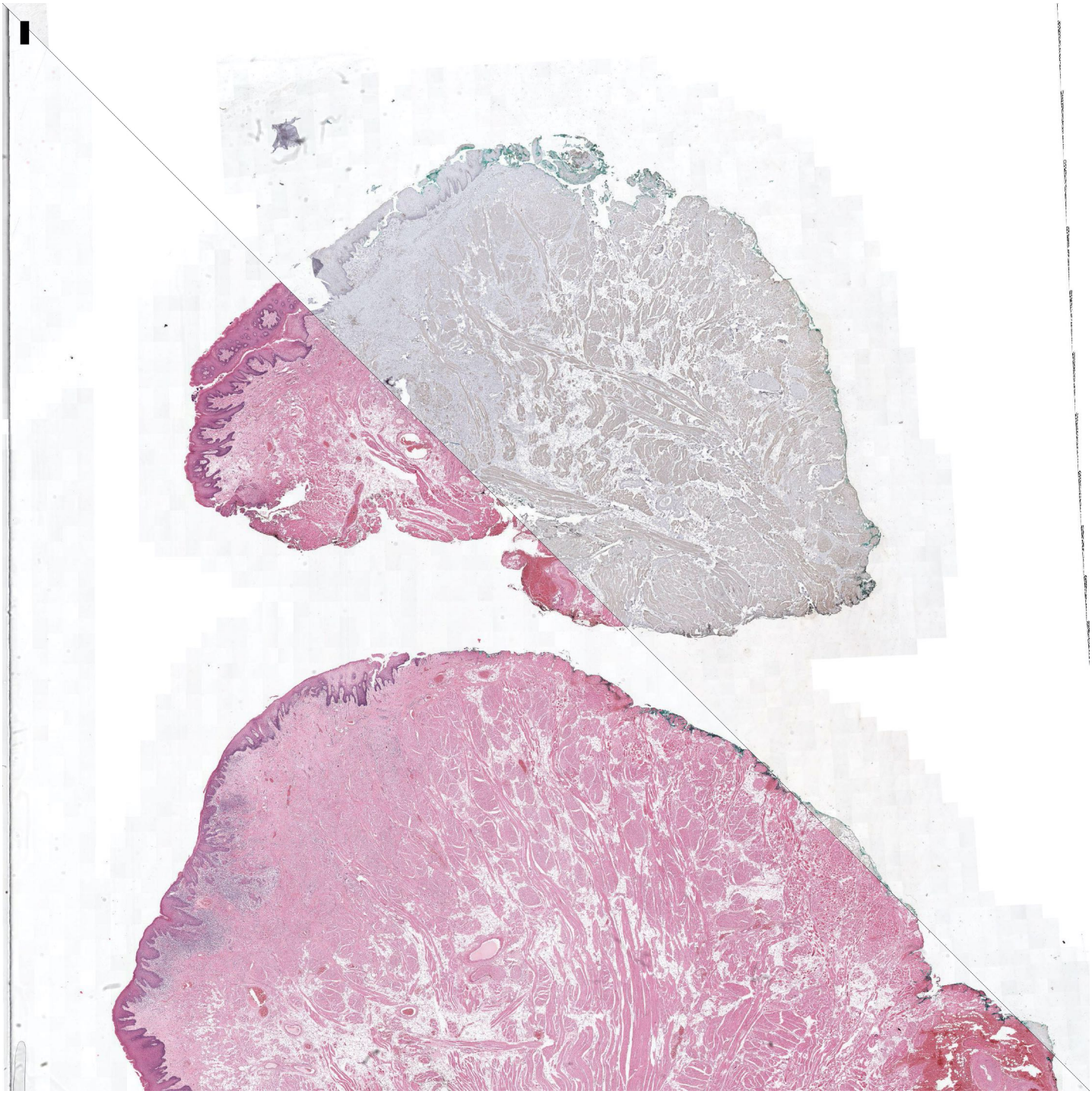


Figure G.25: Overlay of image pair 361 after registration with CLG (registration image size <16k). Zoom level: 5. Scale bar: 500µm.

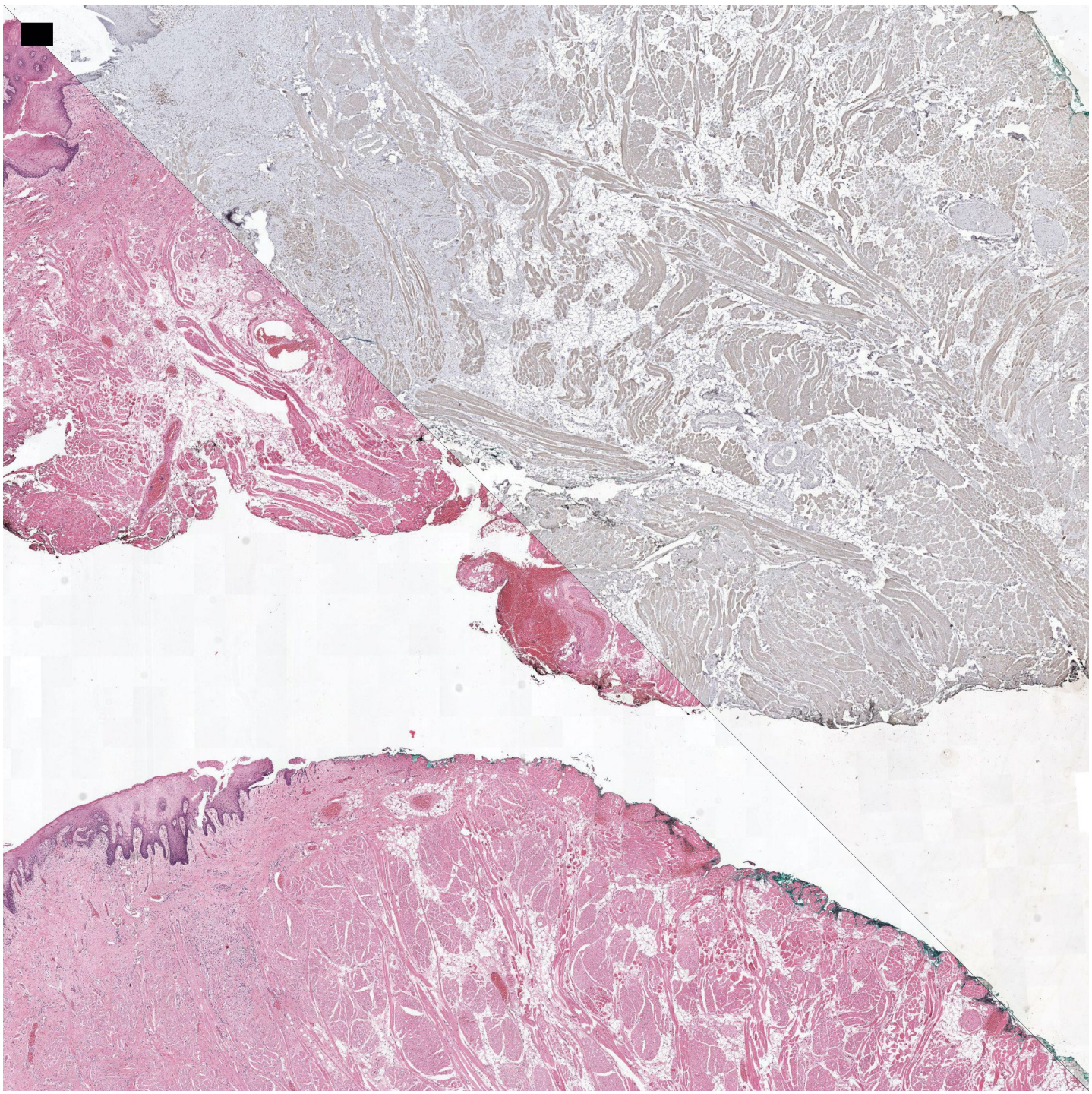


Figure G.26: Overlay of image pair 361 after registration with CLG (registration image size <16k). Zoom level: 4. Scale bar: 500µm.

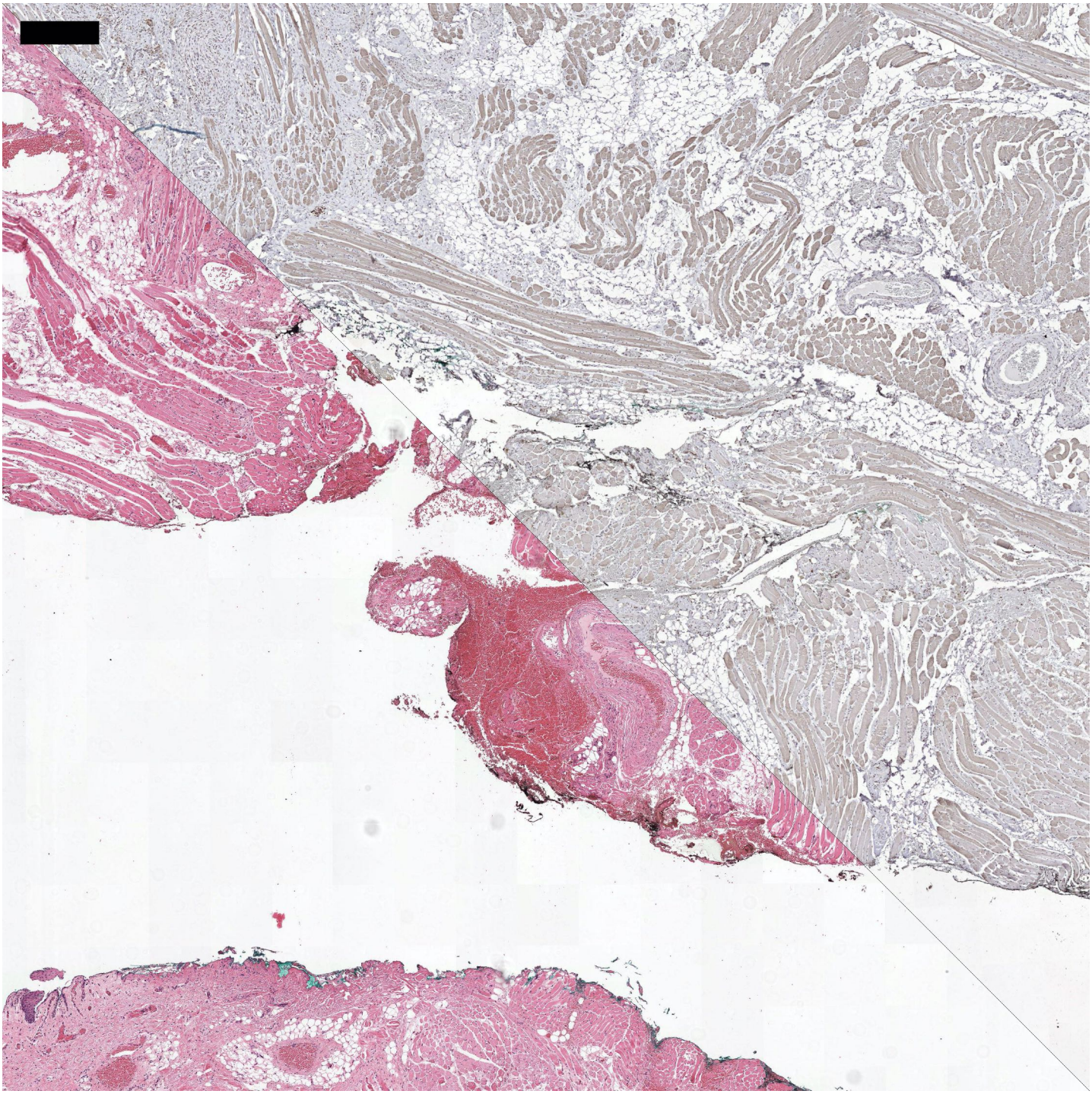


Figure G.27: Overlay of image pair 361 after registration with CLG (registration image size <16k). Zoom level: 3. Scale bar: 500µm.

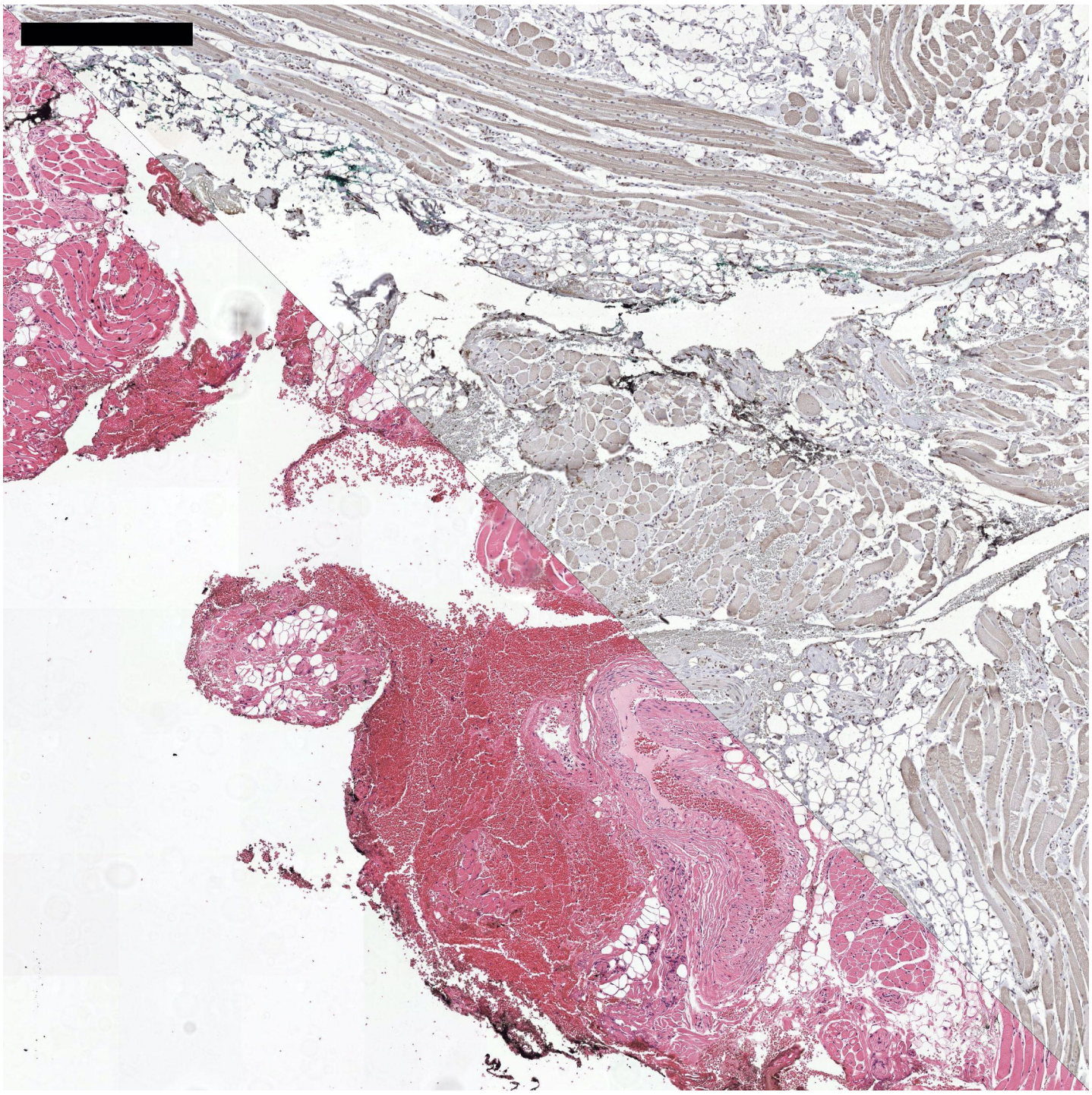


Figure G.28: Overlay of image pair 361 after registration with CLG (registration image size <16k). Zoom level: 2. Scale bar: 500µm.

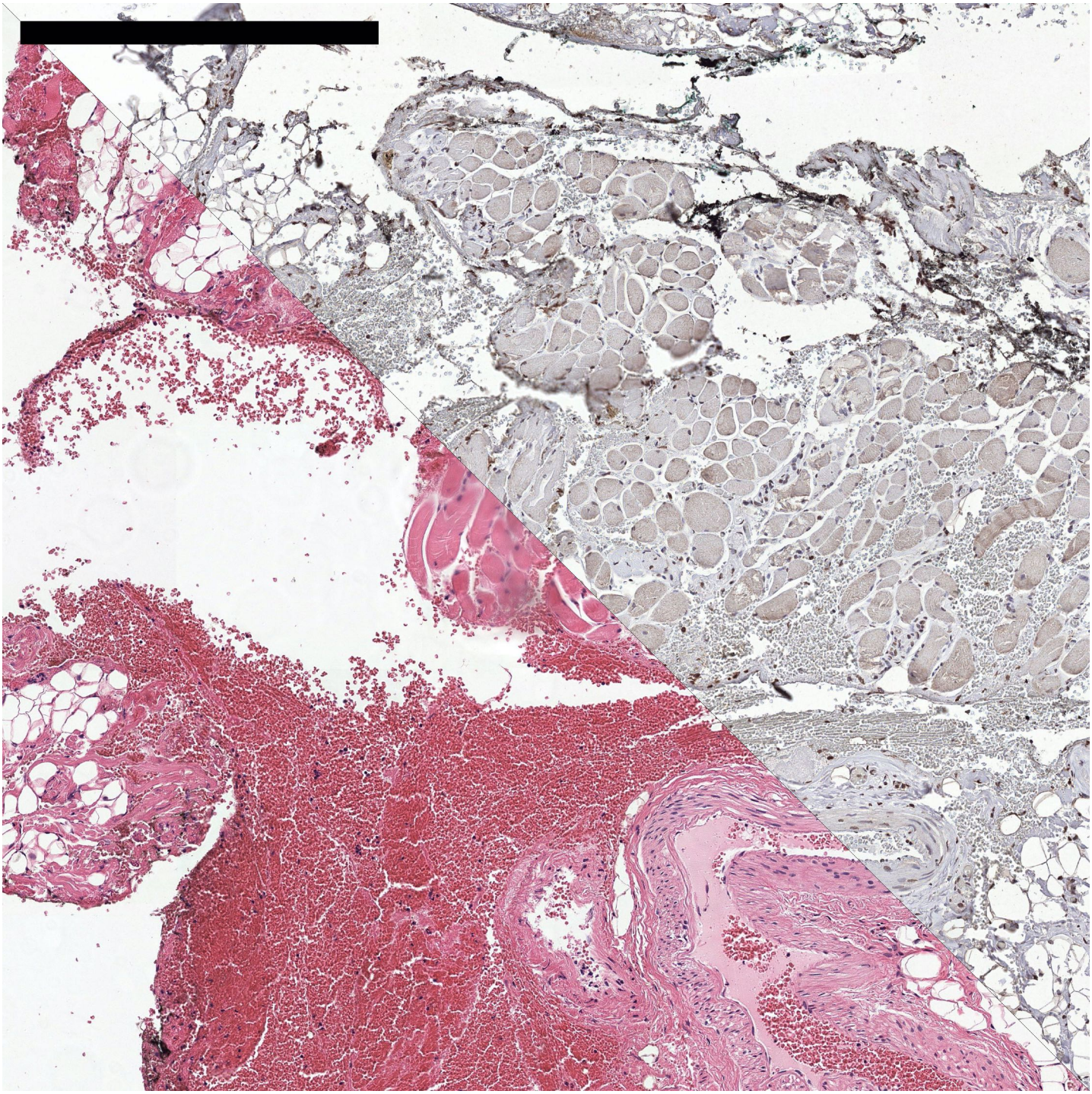


Figure G.29: Overlay of image pair 361 after registration with CLG (registration image size <16k). Zoom level: 1. Scale bar: 500µm.

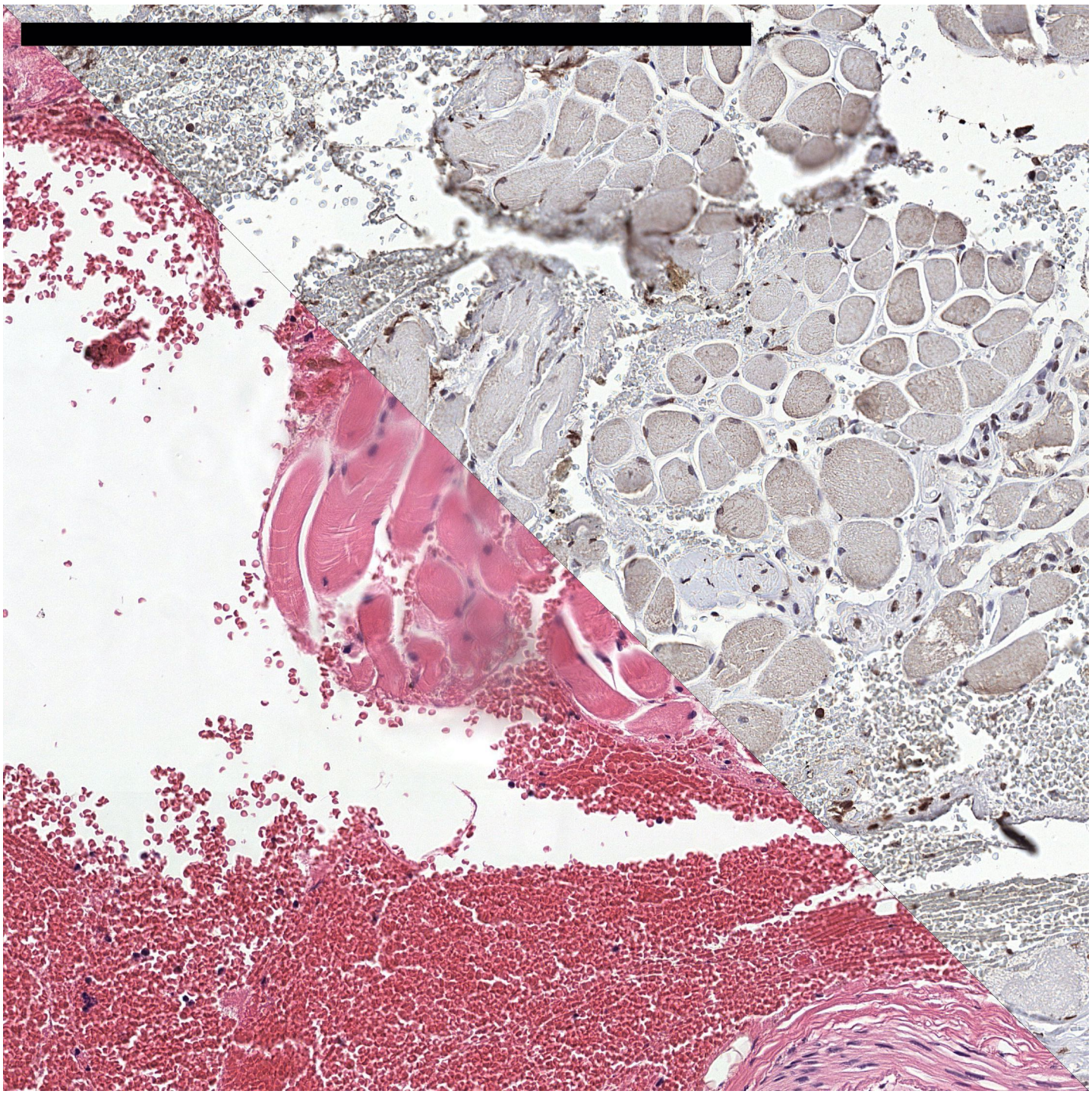


Figure G.30: Overlay of image pair 361 after registration with CLG (registration image size <16k). Zoom level: 0. Scale bar: 500µm.

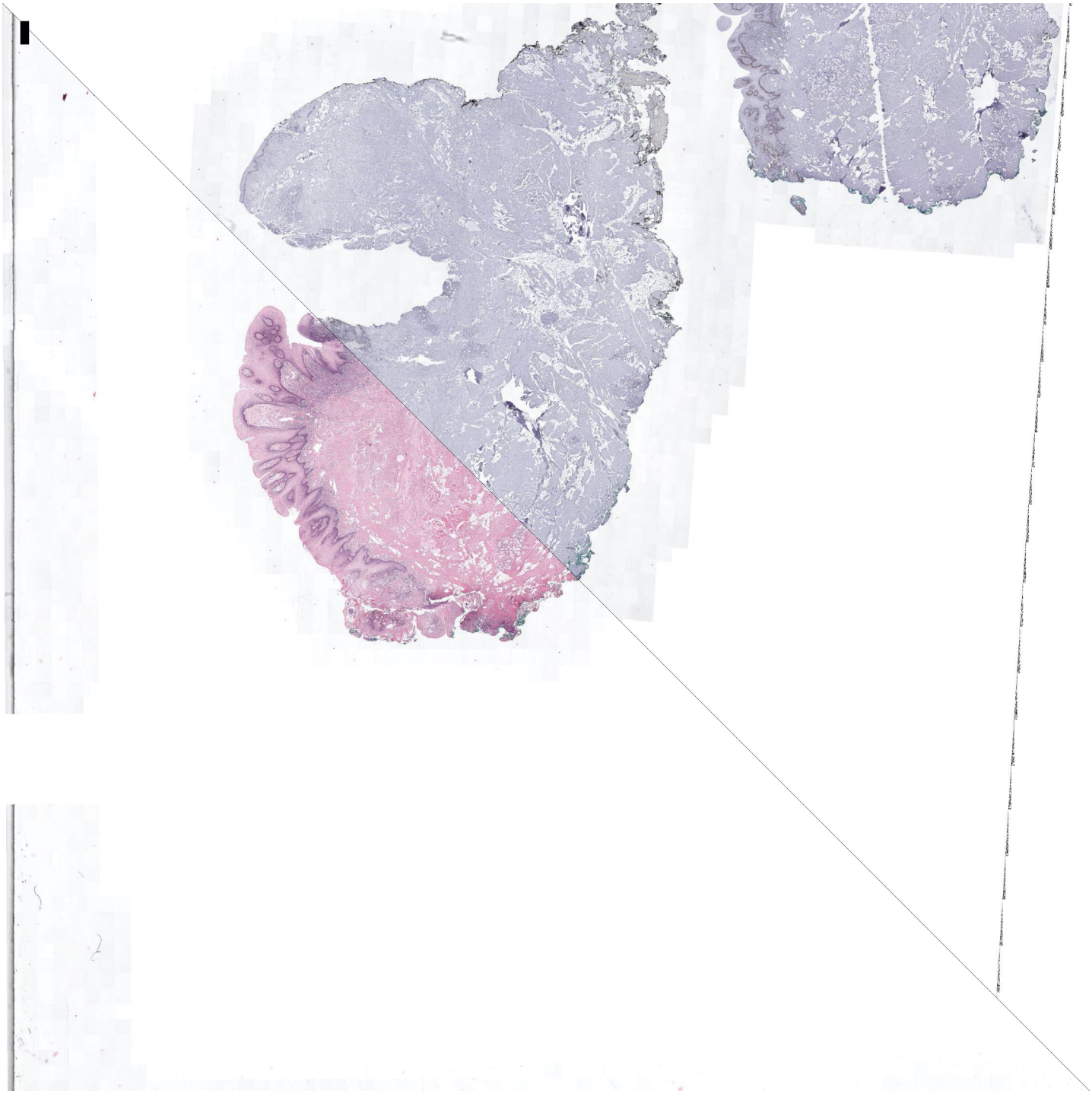


Figure G.31: Overlay of image pair 446 after registration with CLG (registration image size <16k). Zoom level: 5. Scale bar: 500µm.

Figure G.32: Overlay of image pair 446 after registration with CLG (registration image size <16k). Zoom level: 4. Scale bar: 500µm.

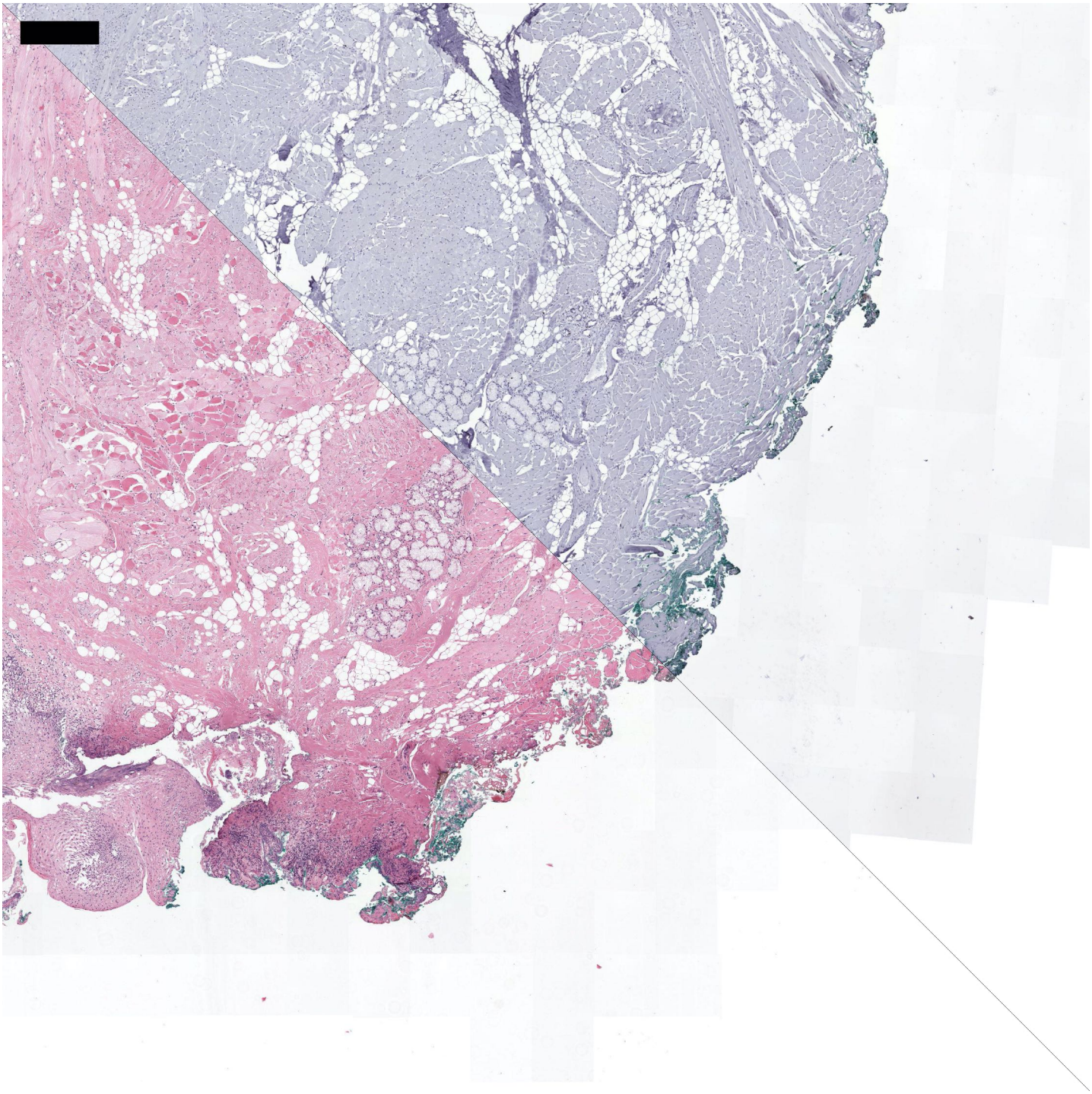


Figure G.33: Overlay of image pair 446 after registration with CLG (registration image size <16k). Zoom level: 3. Scale bar: 500µm.

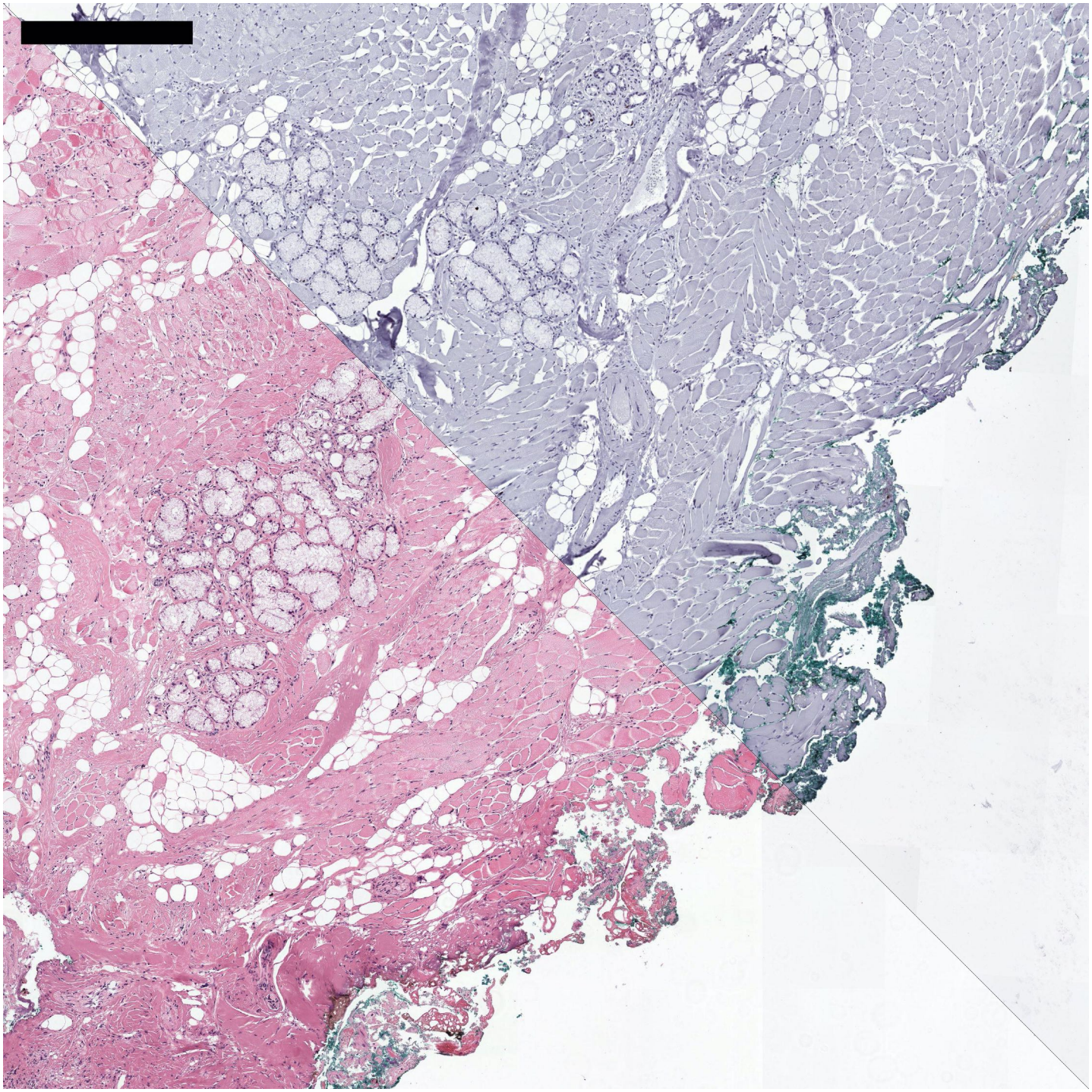


Figure G.34: Overlay of image pair 446 after registration with CLG (registration image size <16k). Zoom level: 2. Scale bar: 500µm.

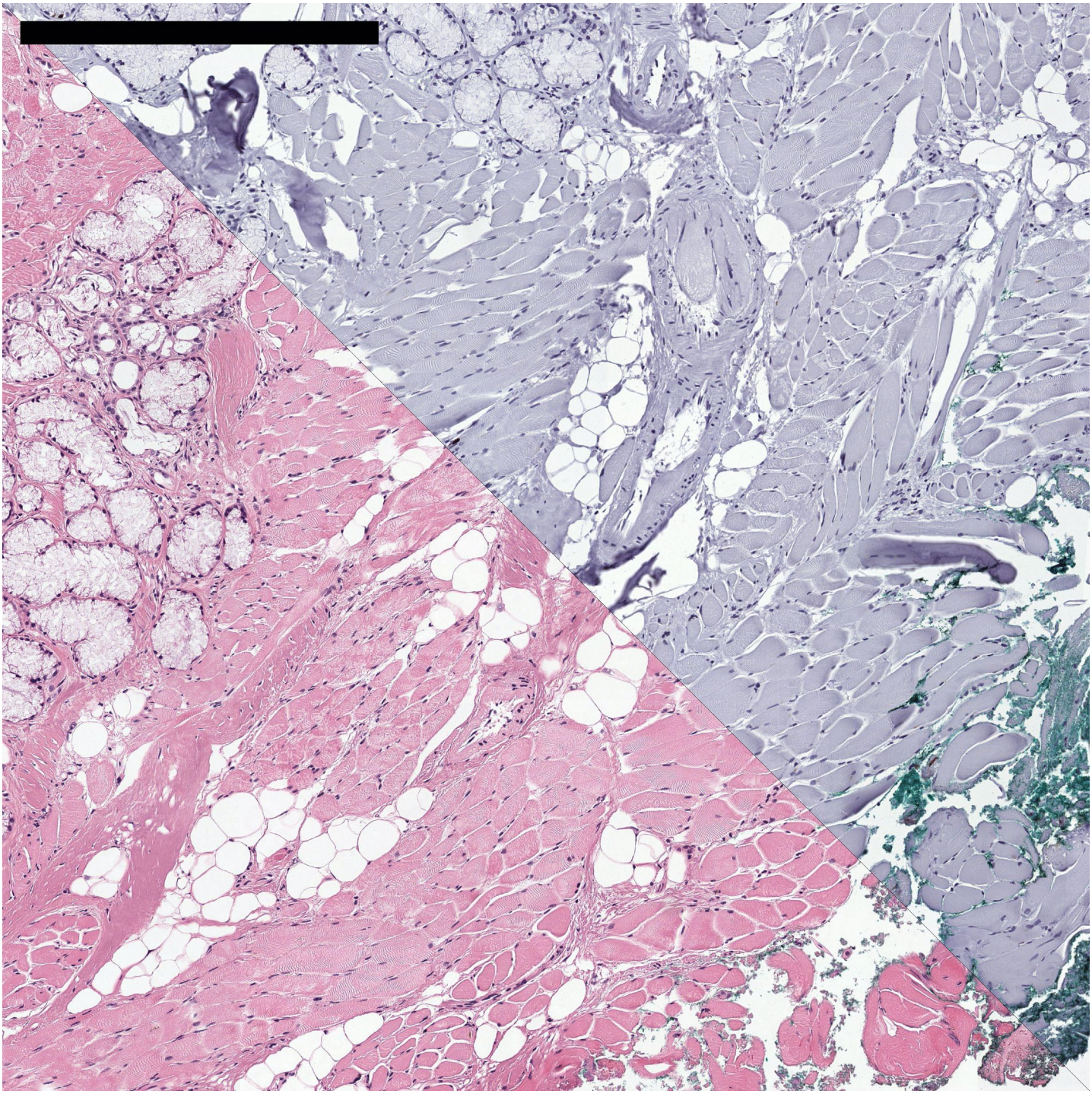


Figure G.35: Overlay of image pair 446 after registration with CLG (registration image size <16k). Zoom level: 1. Scale bar: 500µm.

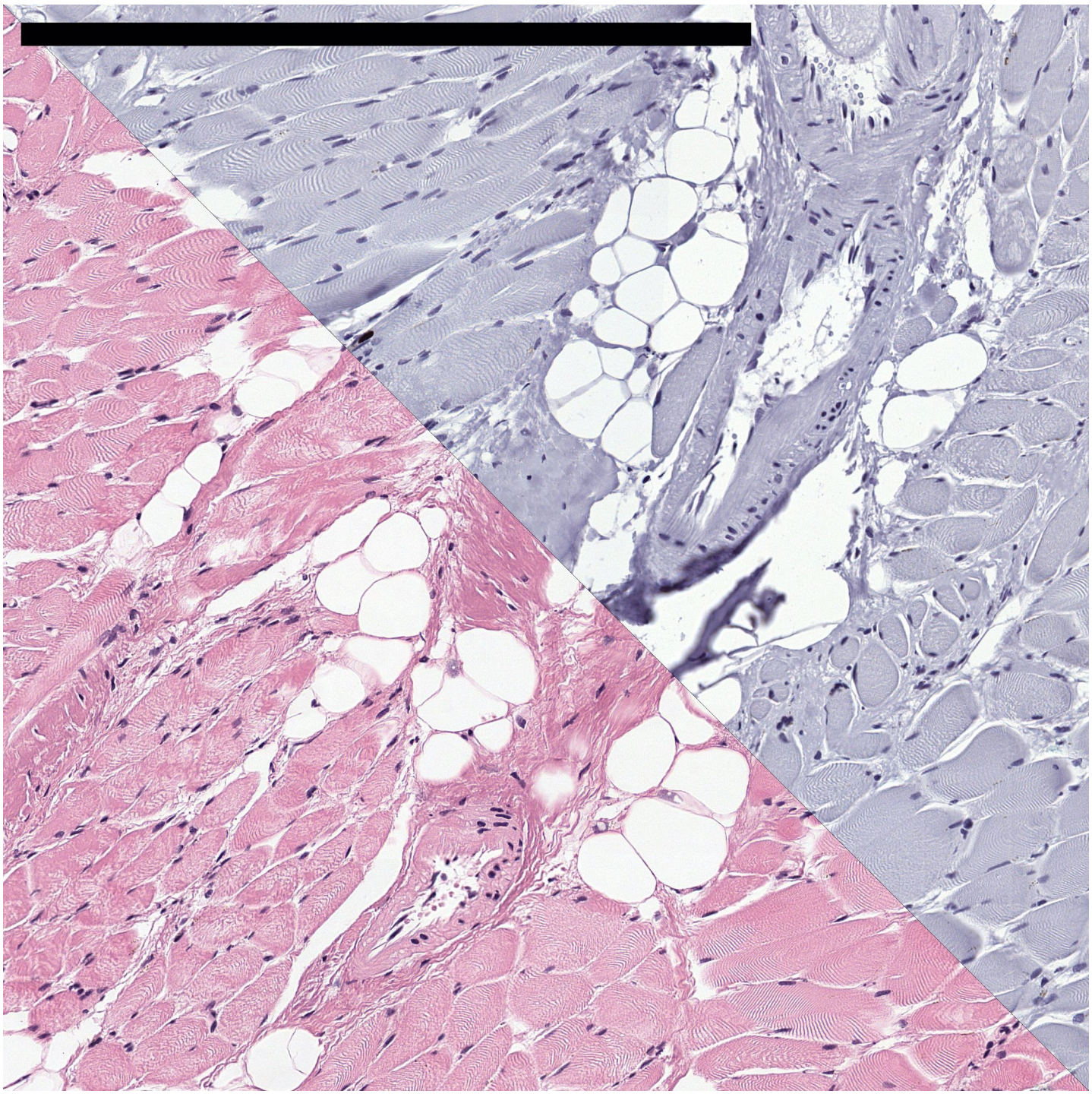


Figure G.36: Overlay of image pair 446 after registration with CLG (registration image size <16k). Zoom level: 0. Scale bar: 500µm.

References

[Amu13] K. Amunts, C. Lepage, L. Borgeat, H. Mohlberg, T. Dickscheid, M.-E. Rousseau, S. Bludau, P.-L. Bazin, L. B. Lewis, A.-M. Oros-Peusquens, N. J. Shah, T. Lippert, K. Zilles, and A. C. Evans. “BigBrain: An Ultrahigh-Resolution 3D Human Brain Model”. In: *Science* 340.6139 (2013), pp. 1472–1475. DOI: 10.1126/science.1235381 (cit. on pp. 5, 16, 42).

[And17] N. L. Andersen, A. Brügmann, G. Lelkaitis, S. Nielsen, M. Friis Lippert, and M. Vyberg. “Virtual Double Staining: A Digital Approach to Immunohistochemical Quantification of Estrogen Receptor Protein in Breast Carcinoma Specimens”. In: *Applied Immunohistochemistry & Molecular Morphology* (2017), pp. 620–626. DOI: 10.1097/PAI.0000000000000502 (cit. on pp. 1, 4, 40).

[And92] A. Andreasen, A. Drewes, J. Assentoft, and N. Larsen. “Computer-Assisted Alignment of Standard Serial Sections without Use of Artificial Landmarks. A Practical Approach to the Utilization of Incomplete Information in 3-D Reconstruction of the Hippocampal Region”. In: *Journal of Neuroscience Methods* 45.3 (1992), pp. 199–207. DOI: 10.1016/0165-0270(92)90077-Q (cit. on p. 5).

[Ars05] V. Arsigny, X. Pennec, and N. Ayache. “Polyrigid and Polyaffine Transformations: A Novel Geometrical Tool to Deal with Non-Rigid Deformations - Application to the Registration of Histological Slices.” In: *Medical Image Analysis* 9.6 (2005), pp. 507–23. DOI: 10.1016/j.media.2005.04.001 (cit. on pp. 5, 39).

[Ayd17] V. A. Aydin and H. Foroosh. “In-Band Sub-Pixel Registration of Wavelet-Encoded Images from Sparse Coefficients”. In: *Signal, Image and Video Processing* 11.8 (2017), pp. 1527–1535. DOI: 10.1007/s11760-017-1116-5 (cit. on p. 9).

[Bal18] S. Balay, S. Abhyankar, M. F. Adams, J. Brown, P. Brune, K. Buschelman, L. Dalcin, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, D. A. May, L. C. McInnes, R. T. Mills, T. Munson, et al. *PETSc Web Page*. 2018. URL: http://www.mcs.anl.gov/petsc (visited on 08/24/2019) (cit. on p. 47).

[Ban00] I. N. Bankman, ed. *Handbook of Medical Imaging: Processing and Analysis*. Academic Press Series in Biomedical Engineering. Academic Press, 2000 (cit. on pp. 23, 56, 77).

[Bar02a] E. Bardinet, S. Ourselin, G. Malandain, D. Tande, K. Parain, N. Ayache, and J. Yelnik. “Three Dimensional Functional Cartography of the Human Basal Ganglia by Registration of Optical and Histological Serial Sections”. In: *Proceedings IEEE International Symposium on Biomedical Imaging*. 2002 IEEE International Symposium on Biomedical Imaging. IEEE, 2002, pp. 329–332. DOI: 10.1109/ISBI.2002.1029260 (cit. on p. 4).

[Bar02b] É. Bardinet, S. Ourselin, D. Dormont, G. Malandain, D. Tandé, K. Parain, N. Ayache, and J. Yelnik. “Co-Registration of Histological, Optical and MR Data of the Human Brain”. In: *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002*. Ed. by T. Dohi and R. Kikinis. Red. by G. Goos, J. Hartmanis, and J. van Leeuwen. Vol. 2488. Springer, 2002, pp. 548–555. DOI: 10.1007/3-540-45786-0_68 (cit. on pp. 4, 5).

[Bes92] P. Besl and N. D. McKay. “A Method for Registration of 3-D Shapes”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14.2 (1992), pp. 239–256. DOI: 10.1109/34.121791 (cit. on p. 9).

[Bez17] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. “Julia: A Fresh Approach to Numerical Computing”. In: *SIAM Review* 59.1 (2017), pp. 65–98. DOI: 10.1137/141000671 (cit. on p. 83).

[Boy01] Y. Boykov, O. Veksler, and R. Zabih. “Fast Approximate Energy Minimization via Graph Cuts”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23.11 (Nov./2001), pp. 1222–1239. DOI: 10.1109/34.969114 (cit. on p. 56).

[Bro14] P. Bronsert, K. Enderle-Ammour, M. Bader, S. Timme, M. Kuehs, A. Csanadi, G. Kayser, I. Kohler, D. Bausch, J. Hoeppner, U. Hopt, T. Keck, E. Stickeler, B. Passlick, O. Schilling, et al. “Cancer Cell Invasion and EMT Marker Expression: A Three-Dimensional Study of the Human Cancer-Host Interface: 3D Cancer-Host Interface”. In: *The Journal of Pathology* 234.3 (2014), pp. 410–422. DOI: 10.1002/path.4416 (cit. on pp. 5, 12–14).

[Bro17] K. K. Brock, S. Mutic, T. R. McNutt, H. Li, and M. L. Kessler. “Use of Image Registration and Fusion Algorithms and Techniques in Radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132”. In: *Medical Physics* 44.7 (2017), e43–e76. DOI: 10.1002/mp.12256 (cit. on p. 20).

[Bro81] C. Broit. “Optimal Registration of Deformed Images”. University of Pennsylvania, 1981 (cit. on pp. 17, 20).

[Bud19] D. Budelmann, L. König, N. Papenberg, and J. Lellmann. “Fully-Deformable 3D Image Registration in Two Seconds”. In: *Bildverarbeitung für die Medizin 2019*. Ed. by H. Handels, T. M. Deserno, A. Maier, K. H. Maier-Hein, C. Palm, and T. Tolxdorff. Springer Fachmedien Wiesbaden, 2019, pp. 302–307. DOI: 10.1007/978-3-658-25326-4_67 (cit. on p. 123).

[Bul19] W. Bulten, P. Bándi, J. Hoven, R. van de Loo, J. Lotz, N. Weiss, J. van der Laak, B. van Ginneken, C. Hulsbergen-van de Kaa, and G. Litjens. “Epithelium Segmentation Using Deep Learning in H&E-Stained Prostate Specimens with Immunohistochemistry as Reference Standard”. In: *Scientific Reports* 9.1 (2019). DOI: 10.1038/s41598-018-37257-4 (cit. on pp. 12–14, 81, 122).

[Bur09] W. Burger and M. Burge. *Principles of Digital Image Processing: Core Algorithms*. Undergraduate Topics in Computer Science. Springer, 2009 (cit. on p. 16).

[Bur13] M. Burger, J. Modersitzki, and L. Ruthotto. “A Hyperelastic Regularization Energy for Image Registration”. In: *SIAM Journal on Scientific Computing* 35.1 (2013), B132–B148. DOI: 10.1137/110835955 (cit. on p. 33).

[Cac01] P. Cachier and N. Ayache. *Regularization in Image Non-Rigid Registration: I. Trade-off between Smoothness and Intensity Similarity*. RR-4188. INRIA, 2001 (cit. on p. 21).

[Cai02] X.-C. Cai and D. E. Keyes. “Nonlinearly Preconditioned Inexact Newton Algorithms”. In: *SIAM Journal on Scientific Computing* 24.1 (2002), pp. 183–200. DOI: 10.1137/S106482750037620X (cit. on pp. 43, 45, 48–50, 66, 70).

[Cai09] X.-C. Cai. “Nonlinear Overlapping Domain Decomposition Methods”. In: *Domain Decomposition Methods in Science and Engineering XVIII*. Ed. by M. Bercovier, M. J. Gander, R. Kornhuber, and O. Widlund. Vol. 70. Springer, 2009, pp. 217–224. DOI: 10.1007/978-3-642-02677-5_23 (cit. on pp. 43, 45–47).

[Cai94] X.-C. Cai, W. D. Gropp, D. E. Keyes, and M. D. Tidriri. “Newton-Krylov-Schwarz Methods in CFD”. In: *Numerical Methods for the Navier-Stokes Equations*. Ed. by F.-K. Hebeker, R. Rannacher, and G. Wittum. Vieweg+Teubner Verlag, 1994, pp. 17–30. DOI: 10.1007/978-3-663-14007-8_3 (cit. on pp. 43, 45–48).

[Car10] A. Cardona, S. Saalfeld, S. Preibisch, B. Schmid, A. Cheng, J. Pulokas, P. Tomancak, and V. Hartenstein. “An Integrated Micro- and Macroarchitectural Analysis of the Drosophila Brain by Computer-Assisted Serial Section Electron Microscopy”. In: *PLoS Biology* 8.10 (2010). Ed. by K. M. Harris. DOI: 10.1371/journal.pbio.1000502 (cit. on pp. 1, 5, 38).

[Cas17] R. Casero, U. Siedlecka, E. S. Jones, L. Gruscheski, M. Gibb, J. E. Schneider, P. Kohl, and V. Grau. “Transformation Diffusion Reconstruction of Three-Dimensional Histology Volumes from Two-Dimensional Image Stacks”. In: *Medical Image Analysis* 38 (2017), pp. 184–204. DOI: 10.1016/j.media.2017.03.004 (cit. on p. 4).

[Cha06] M. M. Chakravarty, G. Bertrand, C. P. Hodge, A. F. Sadikot, and D. L. Collins. “The Creation of a Brain Atlas for Image Guided Neurosurgery Using Serial Histological Data”. In: *NeuroImage* 30.2 (2006), pp. 359–376. DOI: 10.1016/j.neuroimage.2005.09.041 (cit. on p. 5).

[Che18] K. Chen, G. N. Grapiglia, J. Yuan, and D. Zhang. “Improved Optimization Methods for Image Registration Problems”. In: *Numerical Algorithms* (2018). DOI: 10.1007/s11075-018-0486-2 (cit. on p. 45).

[Chr00] C. Christopoulos, A. Skodras, and T. Ebrahimi. “The JPEG2000 Still Image Coding System: An Overview”. In: *IEEE Transactions on Consumer Electronics* 46.4 (Nov./2000), pp. 1103–1127. DOI: 10.1109/30.920468 (cit. on p. 9).

[Chu19] J. Chung, M. Chung, and J. T. Slagel. “Iterative Sampled Methods for Massive and Separable Nonlinear Inverse Problems”. In: *Scale Space and Variational Methods in Computer Vision*. Ed. by J. Lellmann, M. Burger, and J. Modersitzki. Vol. 11603. Springer International Publishing, 2019, pp. 119–130. DOI: 10.1007/978-3-030-22368-7_10 (cit. on p. 123).

[Coo07] L. Cooper, S. Naidu, G. Leone, J. Saltz, and K. Huang. “Registering High Resolution Microscopic Images with Different Histochemical Stainings - A Tool for Mapping Gene Expression with Cellular Structures”. In: *Proceedings of 2nd Workshop on Microsopic Image Analysis with Applications in Biology, Piscataway, NJ, USA*. Ed. by D. Metaxas, J. Rittscher, S. Lockett, and T. Sebastian. 2007 (cit. on pp. 38, 39).

[Dau07] J. Dauguet, T. Delzescaux, F. Condé, J.-F. Mangin, N. Ayache, P. Hantraye, and V. Frouin. “Three-Dimensional Reconstruction of Stained Histological Slices and 3D Non-Linear Registration with in-Vivo MRI for Whole Baboon Brain”. In: *Journal of Neuroscience Methods* 164.1 (2007), pp. 191–204. DOI: 10.1016/j.jneumeth.2007.04.017 (cit. on p. 4).

[dAza07] E. de Azambuja, F. Cardoso, G. de Castro, M. Colozza, M. S. Mano, V. Durbecq, C. Sotiriou, D. Larsimont, M. J. Piccart-Gebhart, and M. Paesmans. “Ki-67 as Prognostic Marker in Early Breast Cancer: A Meta-Analysis of Published Studies Involving 12 155 Patients”. In: *British Journal of Cancer* 96.10 (2007), pp. 1504–1513. DOI: 10.1038/sj.bjc.6603756 (cit. on p. 107).

[dBoo01] C.-W. R. de Boor. *A Practical Guide to Splines*. Rev. ed. Applied Mathematical Sciences v. 27. Springer, 2001 (cit. on pp. 22, 77).

[Dem82] R. S. Dembo, S. C. Eisenstat, and T. Steihaug. “Inexact Newton Methods”. In: *SIAM Journal on Numerical Analysis* 19.2 (1982), pp. 400–408. DOI: 10.1137/0719025 (cit. on p. 29).

[Den96] J. E. Dennis and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Classics in Applied Mathematics 16. Society for Industrial and Applied Mathematics, 1996 (cit. on p. 29).

[Dol15] V. Dolean, P. Jolivet, and F. Nataf. *An Introduction to Domain Decomposition Methods: Algorithms, Theory, and Parallel Implementation*. Society for Industrial and Applied Mathematics, 2015. DOI: 10.1137/1.9781611974065 (cit. on pp. 42–44, 47, 54, 66).

[Dol16] V. Dolean, M. J. Gander, W. Kheriji, F. Kwok, and R. Masson. “Nonlinear Preconditioning: How to Use a Nonlinear Schwarz Method to Precondition Newton’s Method”. In: *SIAM Journal on Scientific Computing* 38.6 (2016), A3357–A3380. DOI: 10.1137/15M102887X (cit. on pp. 46, 122).

[Dry94] M. Dryja and O. B. Widlund. “Domain Decomposition Algorithms with Small Overlap”. In: *SIAM Journal on Scientific Computing* 15.3 (1994), pp. 604–620. DOI: 10.1137/0915040 (cit. on pp. 46, 70).

[Efs03] E. Efstathiou and M. J. Gander. “Why Restricted Additive Schwarz Converges Faster than Additive Schwarz”. In: *BIT Numerical Mathematics* 43.5 (2003), pp. 945–959. DOI: 10.1023/B:BITN.0000014563.33622.1d (cit. on p. 45).

[Ehr19] J. Ehrhardt, M. Ahlborg, H. Uzunova, T. M. Buzug, and H. Handels. “Temporal Polyrigid Registration for Patch-Based MPI Reconstruction of Moving Objects”. In: *International Journal on Magnetic Particle Imaging* Vol 5 (2019). DOI: 10.18416/ijmpi.2019.1908001 (cit. on p. 39).

[Eva18] A. J. Evans, T. W. Bauer, M. M. Bui, T. C. Cornish, H. Duncan, E. F. Glassy, J. Hipp, R. S. McGee, D. Murphy, C. Myers, D. G. O’Neill, A. V. Parwani, B. A. Rampy, M. E. Salama, and L. Pantanowitz. “US Food and Drug Administration Approval of Whole Slide Imaging for Primary Diagnosis: A Key Milestone Is Reached and New Questions Are Raised”. In: *Archives of Pathology & Laboratory Medicine* 142.11 (2018), pp. 1383–1387. DOI: 10.5858/arpa.2017-0496-CP (cit. on p. 4).

[Far01] C. Farhat, M. Lesoinne, P. LeTallec, K. Pierson, and D. Rixen. “FETI-DP: A Dual-Primal Unified FETI Method?Part I: A Faster Alternative to the Two-Level FETI Method”. In: *International Journal for Numerical Methods in Engineering* 50.7 (2001), pp. 1523–1544. DOI: 10.1002/nme.76 (cit. on p. 43).

[Far92] C. Farhat and F.-X. Roux. “An Unconventional Domain Decomposition Method for an Efficient Parallel Solution of Large-Scale Finite Element Systems”. In: *SIAM Journal on Scientific and Statistical Computing* 13.1 (1992), pp. 379–396. DOI: 10.1137/0913020 (cit. on p. 43).

[Feu11] M. Feuerstein, H. Heibel, J. Gardiazabal, N. Navab, and M. Groher. “Reconstruction of 3-D Histology Images by Simultaneous Deformable Registration”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*. Ed. by G. Fichtinger, A. Martel, and T. Peters. Vol. 6892. Springer, 2011, pp. 582–589. DOI: 10.1007/978-3-642-23629-7_71 (cit. on p. 5).

[Fis01] B. Fischer and J. Modersitzki. “Fast Diffusion Registration”. In: *Contemporary Mathematics* 313 (2001), pp. 117–127 (cit. on pp. 20, 33).

[Fis03a] B. Fischer and J. Modersitzki. “Combining Landmark and Intensity Driven Registrations”. In: *PAMM* 3.1 (2003), pp. 32–35. DOI: 10.1002/pamm.200310309 (cit. on p. 86).

[Fis03b] B. Fischer and J. Modersitzki. “Curvature Based Image Registration”. In: *Journal of Mathematical Imaging and Vision* (2003), pp. 81–85. DOI: 10.1023/A:1021897212261 (cit. on pp. 17, 20, 33).

[Fis04] B. Fischer and J. Modersitzki. “A Unified Approach to Fast Image Registration and a New Curvature Based Registration Technique”. In: *Linear Algebra and its Applications* 380 (2004), pp. 107–124. DOI: 10.1016/j.laa.2003.10.021 (cit. on pp. 20, 23).

[Fis08] B. Fischer and J. Modersitzki. “Ill-Posed Medicine—an Introduction to Image Registration”. In: *Inverse Problems* 24.3 (2008), pp. 034008–034008. DOI: 10.1088/0266-5611/24/3/034008 (cit. on p. 15).

[For10] D. Forsberg, M. Andersson, and H. Knutsson. “Adaptive Anisotropic Regularization of Deformation Fields for Non-Rigid Registration Using the Morphon Framework”. In: *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. 2010 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2010, pp. 473–476. DOI: 10.1109/ICASSP.2010.5495704 (cit. on p. 17).

[For11] D. Forsberg, Y. Rathi, S. Bouix, D. Wassermann, H. Knutsson, and C.-F. Westin. “Improving Registration Using Multi-Channel Diffeomorphic Demons Combined with Certainty Maps”. In: *Multimodal Brain Image Analysis*. Ed. by T. Liu, D. Shen, L. Ibanez, and X. Tao. Vol. 7012. Springer, 2011, pp. 19–26. DOI: 10.1007/978-3-642-24446-9_3 (cit. on pp. 55, 73).

[Gef03] S. Gefen, O. Tretiak, and J. Nissanov. “Elastic 3-D Alignment of Rat Brain Histological Images”. In: *IEEE Transactions on Medical Imaging* 22.11 (2003), pp. 1480–1489. DOI: 10.1109/TMI.2003.819280 (cit. on p. 4).

[Gil97] P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization*. Academic Press, 1997 (cit. on p. 30).

[Gol96] G. H. Golub and C. F. Van Loan. *Matrix Computations*. 3rd ed. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 1996 (cit. on pp. 33, 83).

[Hab04] E. Haber and J. Modersitzki. “Numerical Methods for Volume Preserving Image Registration”. In: *Inverse Problems* 20.5 (2004), pp. 1621–1638. DOI: 10.1088/0266-5611/20/5/018 (cit. on pp. 16, 23).

[Hab06] E. Haber and J. Modersitzki. “A Multilevel Method for Image Registration”. In: *SIAM Journal on Scientific Computing* 27.5 (2006), pp. 1594–1607. DOI: 10.1137/040608106 (cit. on p. 21).

[Hab07a] E. Haber and J. Modersitzki. “Intensity Gradient Based Registration and Fusion of Multi-Modal Images”. In: *Methods of Information in Medicine* 46.03 (2007), pp. 292–299. DOI: 10.1160/ME9046 (cit. on pp. 18, 20).

[Hab07b] E. Haber, S. Heldmann, and J. Modersitzki. “An OcTree Method for Parametric Image Registration”. In: *SIAM Journal on Scientific Computing* 29.5 (2007), pp. 2008–2023. DOI: 10.1137/060662605 (cit. on p. 62).

[Hab09] E. Haber, S. Heldmann, and J. Modersitzki. “A Scale-Space Approach to Landmark Constrained Image Registration”. In: *Scale Space and Variational Methods in Computer Vision*. Ed. by X.-C. Tai, K. Mørken, M. Lysaker, and K.-A. Lie. Vol. 5567. Springer, 2009, pp. 612–623. DOI: 10.1007/978-3-642-02256-2_51 (cit. on p. 86).

[Had02] J. Hadamard. “Sur Les Problemes Aux Derive Espartielles et Leur Signification Physique”. In: *Bulletin of Princeton University* 13 (1902), pp. 49–52 (cit. on p. 20).

[Ham10] M. E. H. Hammond, D. F. Hayes, M. Dowsett, D. C. Allred, K. L. Hagerty, S. Badve, P. L. Fitzgibbons, G. Francis, N. S. Goldstein, M. Hayes, D. G. Hicks, S. Lester, R. Love, P. B. Mangu, L. McShane, et al. “American Society of Clinical Oncology/College of American Pathologists Guideline Recommendations for Immunohistochemical Testing of Estrogen and Progesterone Receptors in Breast Cancer (Unabridged Version)”. In: *Archives of Pathology & Laboratory Medicine* 134.7 (2010), e48–e72. DOI: 10.1043/1543-2165-134.7.e48 (cit. on p. 40).

[Har18] N. Harder, M. Athelogou, H. Hessel, N. Brieu, M. Yigitsoy, J. Zimmermann, M. Baatz, A. Buchner, C. G. Stief, T. Kirchner, G. Binnig, G. Schmidt, and R. Huss. “Tissue Phenomics for Prognostic Biomarker Discovery in Low- and Intermediate-Risk Prostate Cancer”. In: *Scientific Reports* 8.1 (2018). DOI: 10.1038/s41598-018-22564-7 (cit. on pp. 38, 40, 123).

[Hei16] M. P. Heinrich, I. J. Simpson, B. W. Papież, S. M. Brady, and J. A. Schnabel. “Deformable Image Registration by Combining Uncertainty Estimates from Supervoxel Belief Propagation”. In: *Medical Image Analysis* 27 (2016), pp. 57–71. DOI: 10.1016/j.media.2015.09.005 (cit. on pp. 56, 73).

[Hei19] A. Heinlein and M. Lanser. *Additive and Hybrid Nonlinear Two-Level Schwarz Methods and Energy Minimizing Coarse Spaces for Unstructured Grids*. Technical report. Universität zu Köln, 2019 (cit. on p. 54).

[Hen00] S. Henn and K. Witsch. “A Multigrid Approach for Minimizing a Nonlinear Functional for Digital Image Matching”. In: *Computing* 64.4 (2000), pp. 339–348. DOI: 10.1007/s006070070029 (cit. on p. 29).

[Hib88] L. S. Hibbard and R. A. Hawkins. “Objective Image Alignment for Three-Dimensional Reconstruction of Digital Autoradiograms”. In: *Journal of Neuroscience Methods* 26.1 (1988), pp. 55–74. DOI: 10.1016/0165-0270(88)90129-X (cit. on p. 5).

[Höf18] H. Höfener, A. Homeyer, N. Weiss, J. Molin, C. F. Lundström, and H. K. Hahn. “Deep Learning Nuclei Detection: A Simple Approach Can Deliver State-of-the-Art Results”. In: *Computerized Medical Imaging and Graphics* 70 (2018), pp. 43–52. DOI: 10.1016/j.compmedimag.2018.08.010 (cit. on p. 9).

[Hua06] K. Huang, L. Cooper, A. Sharma, and T. Pan. “Fast Automatic Registration Algorithm for Large Microscopy Images”. In: *Proceedings of Life Science Systems and Applications Workshop*. IEEE/NLM, 2006. DOI: 10.1109/LSSA.2006.250414 (cit. on p. 5).

[Jan16] A. Janowczyk and A. Madabhushi. “Deep Learning for Digital Pathology Image Analysis: A Comprehensive Tutorial with Selected Use Cases”. In: *Journal of Pathology Informatics* 7.1 (2016), p. 29. DOI: 10.4103/2153-3539.186902 (cit. on p. 123).

[Kaj19] T. Kajihara, T. Funatomi, H. Makishima, T. Aoto, H. Kubo, S. Yamada, and Y. Mukaigawa. “Non-Rigid Registration of Serial Section Images by Blending Transforms for 3D Reconstruction”. In: *Pattern Recognition* 96 (2019), p. 106956. DOI: 10.1016/j.patcog.2019.07.001 (cit. on p. 56).

[Kei14] E. Keilegavlen, J. O. Skogestad, and J. M. Nordbotten. “Domain Decomposition Preconditioning for Non-Linear Elasticity Problems”. In: *Solution Strategies for Nonlinear Conservation Laws*. 2014 (cit. on p. 54).

[Kes17] A. P. Keszei, B. Berkels, and T. M. Deserno. “Survey of Non-Rigid Registration Tools in Medicine”. In: *Journal of Digital Imaging* 30.1 (2017), pp. 102–116. DOI: 10.1007/s10278-016-9915-8 (cit. on pp. 20, 37).

[Kie15] J. A. Kiernan. *Histological and Histochemical Methods: Theory and Practice*. Scion, 2015 (cit. on p. 107).

[Kim97] B. Kim, J. L. Boes, K. A. Frey, and C. R. Meyer. “Mutual Information for Automated Unwarping of Rat Brain Autoradiographs”. In: *NeuroImage* 5.1 (1997), pp. 31–40. DOI: 10.1006/nimg.1996.0251 (cit. on p. 4).

[Kla14] A. Klawonn, M. Lanser, and O. Rheinbach. “Nonlinear FETI-DP and BDDC Methods”. In: *SIAM Journal on Scientific Computing* 36.2 (2014), A737–A765. DOI: 10.1137/130920563 (cit. on p. 43).

[Kla17] A. Klawonn, M. Lanser, O. Rheinbach, and M. Uran. “Nonlinear FETI-DP and BDDC Methods: A Unified Framework and Parallel Results”. In: *SIAM Journal on Scientific Computing* 39.6 (2017), pp. C417–C451. DOI: 10.1137/16M1102495 (cit. on pp. 43, 122).

[Koh07] T. Kohlberger. “Variational Domain Decomposition For Parallel Image Processing”. 2007 (cit. on p. 122).

[Kön18] L. König, J. Rühaak, A. Derksen, and J. Lellmann. “A Matrix-Free Approach to Parallel and Memory-Efficient Deformable Image Registration”. In: *SIAM Journal on Scientific Computing* 40.3 (2018), B858–B888. DOI: 10.1137/17M1125522 (cit. on pp. 16, 26, 62, 68, 107).

[Koo18] T. Koopman, H. J. Buikema, H. Hollema, G. H. de Bock, and B. van der Vegt. “Digital Image Analysis of Ki67 Proliferation Index in Breast Cancer Using Virtual Dual Staining on Whole Tissue Sections: Clinical Validation and Inter-Platform Agreement”. In: *Breast Cancer Research and Treatment* (2018). DOI: 10.1007/s10549-018-4669-2 (cit. on p. 4).

[Lan13] G. Lang. *Histotechnik Praxislehrbuch für die Biomedizinische Analytik*. Springer, 2013 (cit. on pp. 2, 3).

[Law13] A. S. Lawless. "Variational Data Assimilation for Very Large Environmental Problems". In: *Large Scale Inverse Problems*. Ed. by M. Cullen, M. A. Freitag, S. Kindermann, and R. Scheichl. DE GRUYTER, 2013. DOI: 10.1515/9783110282269.55 (cit. on p. 123).

[Lee17] C.-O. Lee, X.-C. Cai, D. E. Keyes, H. H. Kim, A. Klawonn, E.-J. Park, and O. B. Widlund, eds. *Domain Decomposition Methods in Science and Engineering XXIII*. Vol. 116. Lecture Notes in Computational Science and Engineering. Springer International Publishing, 2017. DOI: 10.1007/978-3-319-52389-7 (cit. on p. 42).

[Lem08] V. Lempitsky, S. Roth, and C. Rother. "FusionFlow: Discrete-Continuous Optimization for Optical Flow Estimation". In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008. DOI: 10.1109/CVPR.2008.4587751 (cit. on pp. 56, 73).

[Lio88] P.-L. Lions. "On the Schwarz Alternating Method. I." In: First International Symposium on Domain Decomposition Methods for Partial Differential Equations. 1988 (cit. on pp. 44, 45).

[Liu15] L. Liu and D. E. Keyes. "Field-Split Preconditioned Inexact Newton Algorithms". In: *SIAM Journal on Scientific Computing* 37.3 (2015), A1388–A1409. DOI: 10.1137/140970379 (cit. on p. 46).

[Lot14] J. Lotz, J. Berger, B. Müller, K. Breuhahn, N. Grabe, S. Heldmann, A. Homeyer, B. Lahrmann, H. Laue, J. Olesch, M. Schwier, O. Sedlaczek, and A. Warth. "Zooming in: High Resolution 3D Reconstruction of Differently Stained Histological Whole Slide Images". In: *Proc. SPIE 9041*. Medical Imaging 2014: Digital Pathology. Ed. by M. N. Gurcan and A. Madabhushi. 2014, p. 904104. DOI: 10.1117/12.2043381 (cit. on pp. 5, 12, 13, 40).

[Lot16a] J. Lotz, J. Olesch, B. Müller, T. Polzin, P. Galuschka, J. M. Lotz, S. Heldmann, H. Laue, A. Warth, B. Lahrmann, N. Grabe, O. Sedlaczek, K. Breuhahn, and J. Modersitzki. "Patch-Based Nonlinear Image Registration for Gigapixel Whole Slide Images". In: *IEEE Transactions on Biomedical Engineering* 63.9 (2016), pp. 1812–1819. DOI: 10.1109/TBME.2015.2503122 (cit. on pp. 12, 13, 41, 59, 79).

[Lot16b] J. M. Lotz, F. Hoffmann, J. Lotz, S. Heldmann, D. Trede, J. Oetjen, M. Becker, G. Ernst, P. Maas, T. Alexandrov, O. Guntinas-Lichius, H. Thiele, and F. von Eggeling. "Integration of 3D Multimodal Imaging Data of a Head and Neck Cancer and Advanced Feature Recognition". In: *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* (2016). DOI: 10.1016/j.bbapap.2016.08.018 (cit. on pp. 12–14).

[Lot19] J. Lotz, N. Weiss, and S. Heldmann. *Robust, Fast and Accurate: A 3-Step Method for Automatic Histological Image Registration*. arXiv:1903.12063 [cs]. 2019 (cit. on pp. 12, 13).

[Lui01] S.-H. Lui. “On Monotone and Schwarz Alternating Methods for Nonlinear Elliptic PDEs”. In: *ESAIM: Mathematical Modelling and Numerical Analysis* 35.1 (2001), pp. 1–15. DOI: 10.1051/m2an:2001104 (cit. on p. 45).

[Mad16] A. Madabhushi and G. Lee. “Image Analysis and Machine Learning in Digital Pathology: Challenges and Opportunities”. In: *Medical Image Analysis* 33 (2016), pp. 170–175. DOI: 10.1016/j.media.2016.06.037 (cit. on p. 38).

[Mag15] D. Magee, S. Gilbert, A. Bulpitt, Y. Song, N. Roberts, N. Wijayathunga, R. Wilcox, and D. Treanor. “Histopathology in 3D: From Three-Dimensional Reconstruction to Multi-Stain and Multi-Modal Analysis”. In: *Journal of Pathology Informatics* 6.1 (2015). DOI: 10.4103/2153-3539.151890 (cit. on pp. 1, 40).

[Mal04] G. Malandain, E. Bardinet, K. Nelissen, and W. Vanduffel. “Fusion of Autoradiographs with an MR Volume Using 2-D and 3-D Linear Transformations.” In: *NeuroImage* 23.1 (2004), pp. 111–127. DOI: 10.1016/j.neuroimage.2004.04.038 (cit. on p. 5).

[Man16] A. Mang, A. Gholami, and G. Biros. “Distributed-Memory Large Deformation Diffeomorphic 3D Image Registration”. In: *SC16: International Conference for High Performance Computing, Networking, Storage and Analysis*. IEEE, 2016, pp. 842–853. DOI: 10.1109/SC.2016.71 (cit. on pp. 1, 16, 42, 61, 110).

[Mar05] L. Marcinkowski and X.-C. Cai. “Parallel Performance of Some Two-Level ASPIN Algorithms”. In: *Domain Decomposition Methods in Science and Engineering. Lecture Notes in Computational Science and Engineering*. Ed. by T. J. Barth, M. Griebel, D. E. Keyes, R. M. Nieminen, D. Roose, T. Schlick, R. Kornhuber, R. Hoppe, J. Périaux, O. Pironneau, O. Widlund, and J. Xu. Vol. 40. Springer, 2005, pp. 639–646. DOI: 10.1007/3-540-26825-1_68 (cit. on p. 54).

[Mar18] P. Markthub, M. E. Belviranli, S. Lee, J. S. Vetter, and S. Matsuoka. “DRAGON: Breaking GPU Memory Capacity Limits with Direct NVM Access”. In: *SC18: International Conference for High Performance Computing, Networking, Storage and Analysis*. IEEE, 2018, pp. 414–426. DOI: 10.1109/SC.2018.00035 (cit. on p. 41).

[Mas00] D. Y. Mason, K. Micklem, and M. Jones. “Double Immunofluorescence Labelling of Routinely Processed Paraffin Sections”. In: *The Journal of Pathology* 191.4 (2000), pp. 452–461. DOI: 10.1002/1096-9896(2000)9999:9999<::AID-PATH665>3.0.CO;2-O (cit. on p. 4).

[Meg97] M. S. Mega, S. S. Chen, P. M. Thompson, R. P. Woods, T. J. Karaca, A. Tiwari, H. V. Vinters, G. W. Small, and A. W. Toga. “Mapping Histology to Metabolism: Coregistration of Stained Whole-Brain Sections to Premortem PET in Alzheimer’s Disease”. In: *NeuroImage* 5.2 (1997), pp. 147–153. DOI: 10.1006/nimg.1996.0255 (cit. on p. 4).

[Mes16] A. L. Mescher and L. C. U. Junqueira. *Junqueira’s Basic Histology: Text and Atlas*. Fourteenth edition. McGraw Hill Education, 2016 (cit. on p. 2).

[Mey03] U. Meyer, P. Sanders, and J. Sibeyn. *Algorithms for memory hierarchies: advanced lectures*. Springer, 2003 (cit. on pp. 1, 41).

[Mit15] S. Mittal and J. S. Vetter. "A Survey of CPU-GPU Heterogeneous Computing Techniques". In: *ACM Computing Surveys* 47.4 (2015), pp. 1–35. DOI: 10.1145/2788396 (cit. on p. 41).

[Mod04] J. Modersitzki. *Numerical Methods for Image Registration*. Numerical Mathematics and Scientific Computation. Oxford University Press, 2004 (cit. on pp. 10, 18, 20).

[Mod09] J. Modersitzki. *FAIR: Flexible Algorithms for Image Registration*. Fundamentals of Algorithms FA06. Society for Industrial and Applied Mathematics, 2009 (cit. on pp. 1, 2, 8, 15–17, 20, 21, 23–27, 29, 31, 45, 59, 63).

[Mod10] M. Modat, G. R. Ridgway, Z. a Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin. "Fast Free-Form Deformation Using Graphics Processing Units." In: *Computer Methods and Programs in Biomedicine* 98.3 (2010), pp. 278–84. DOI: 10.1016/j.cmpb.2009.09.002 (cit. on pp. 55, 61, 73, 79, 110).

[Mod99] J. Modersitzki, W. Obelöer, O. Schmitt, and G. Lustig. "Elastic Matching of Very Large Digital Images on High Performance Clusters". In: *High-Performance Computing and Networking*. Ed. by P. Sloot, M. Bubak, A. Hoekstra, and B. Hertzberger. Red. by G. Goos, J. Hartmanis, and J. van Leeuwen. Vol. 1593. Springer, 1999, pp. 141–149. DOI: 10.1007/BFb0100575 (cit. on pp. 16, 41).

[Moh16] H. Mohlberg, B. Tweddell, T. Lippert, and K. Amunts. "Workflows for Ultra-High Resolution 3D Models of the Human Brain on Massively Parallel Supercomputers". In: *Brain-Inspired Computing*. Ed. by K. Amunts, L. Grandinetti, T. Lippert, and N. Petkov. Vol. 10087. Springer International Publishing, 2016, pp. 15–27. DOI: 10.1007/978-3-319-50862-7_2 (cit. on pp. 5, 42, 123).

[Mot17] A. Mota, I. Tezaur, and C. Alleman. "The Schwarz Alternating Method in Solid Mechanics". In: *Computer Methods in Applied Mechanics and Engineering* 319 (2017), pp. 19–51. DOI: 10.1016/j.cma.2017.02.006 (cit. on p. 46).

[Mue11] D. Mueller, D. Vossen, and B. Hulsken. "Real-Time Deformable Registration of Multi-Modal Whole Slides for Digital Pathology." In: *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society* 35.7-8 (2011), pp. 542–56. DOI: 10.1016/j.compmedimag.2011.06.006 (cit. on pp. 15, 17, 38).

[Mue12] S. E. A. Muenzing, B. van Ginneken, and J. P. W. Pluim. "On Combining Algorithms for Deformable Image Registration". In: *Biomedical Image Registration*. Ed. by B. M. Dawant, G. E. Christensen, J. M. Fitzpatrick, and D. Rueckert. Red. by D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, and G. Weikum. Vol. 7359.

Springer, 2012, pp. 256–265. DOI: 10.1007/978-3-642-31340-0_27 (cit. on pp. 55, 73, 79, 81, 84).

[Mue14] S. E. Muenzing, B. van Ginneken, M. A. Viergever, and J. P. Pluim. "DIRBoost–An Algorithm for Boosting Deformable Image Registration: Application to Lung CT Intra-Subject Registration". In: *Medical Image Analysis* 18.3 (2014), pp. 449–459. DOI: 10.1016/j.media.2013.12.006 (cit. on p. 55).

[Muk17] S. Mukhopadhyay, M. D. Feldman, E. Abels, R. Ashfaq, S. Beltaifa, N. G. Cacciabeve, H. P. Cathro, L. Cheng, K. Cooper, G. E. Dickey, R. M. Gill, R. P. Heaton, R. Kerstens, G. M. Lindberg, R. K. Malhotra, et al. "Whole Slide Imaging Versus Microscopy for Primary Diagnosis in Surgical Pathology: A Multicenter Blinded Randomized Noninferiority Study of 1992 Cases (Pivotal Study)". In: *The American Journal of Surgical Pathology* (2017), p. 1. DOI: 10.1097/PAS.0000000000000948 (cit. on p. 1).

[Mur12] K. Murphy, J. P. W. Pluim, E. M. van Rikxoort, P. A. de Jong, B. de Hoop, H. A. Gietema, O. Mets, M. de Bruijne, P. Lo, M. Prokop, and B. van Ginneken. "Toward Automatic Regional Analysis of Pulmonary Function Using Inspiration and Expiration Thoracic CT: Automatic Analysis of Pulmonary Function Using Thoracic CT". In: *Medical Physics* 39.3 (2012), pp. 1650–1662. DOI: 10.1118/1.3687891 (cit. on p. 17).

[Myr10] A. Myronenko and Xubo Song. "Point Set Registration: Coherent Point Drift". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.12 (2010), pp. 2262–2275. DOI: 10.1109/TPAMI.2010.46 (cit. on p. 9).

[Nat16] F. Nataf. *Domain Decomposition, Hybrid Methods, Coarse Space Corrections*. 2016. DOI: 10.24350/CIRM.V.19022003 (cit. on p. 43).

[Nel18] A. M. Nelson, M. Hale, M. I. J.-M. Diomande, Q. Eichbaum, Y. Iliyasu, R. M. Kalengayi, B. Rugwizangoga, and S. Sayed. "Training the Next Generation of African Pathologists". In: *Clinics in Laboratory Medicine* 38.1 (2018), pp. 37–51. DOI: 10.1016/j.cll.2017.10.004 (cit. on p. 4).

[Nic87] R. A. Nicolaides. "Deflation of Conjugate Gradients with Applications to Boundary Value Problems". In: *SIAM Journal on Numerical Analysis* 24.2 (1987), pp. 355–365. DOI: 10.1137/0724027 (cit. on p. 54).

[Noc06] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 2006 (cit. on pp. 24, 28–30, 42, 43, 46, 83, 87, 107).

[Oht03] Y. Ohtake, A. Belyaev, M. Alexa, G. Turk, and H.-P. Seidel. "Multi-Level Partition of Unity Implicits". In: *ACM SIGGRAPH 2003 Papers*. ACM SIGGRAPH 2003 Papers. ACM Press, 2003, p. 463. DOI: 10.1145/1201775.882293 (cit. on p. 77).

[Ole14] J. Olesch. *Bildregistrierung für die navigierte Chirurgie: spezialisierte Ansätze zur Anwendung in der navigierten Leberchirurgie*. Aktuelle Forschung Medizintechnik. Springer Vieweg, 2014 (cit. on p. 16).

[Oli14] F. P. Oliveira and J. M. R. Tavares. “Medical Image Registration: A Review”. In: *Computer Methods in Biomechanics and Biomedical Engineering* 17.2 (2014), pp. 73–93. DOI: 10.1080/10255842.2012.670855 (cit. on pp. 20, 37).

[Our01] S. Ourselin, A. Roche, G. Subsol, X. Pennec, and N. Ayache. “Reconstructing a 3D Structure from Serial Histological Sections”. In: *Image and Vision Computing* 19.1-2 (2001), pp. 25–31. DOI: 10.1016/S0262-8856(00)00052-4 (cit. on p. 5).

[Ova13] W. K. Ovalle, P. C. Nahirney, and F. H. Netter. *Netter’s Essential Histology*. 2nd ed. Elsevier/Saunders, 2013 (cit. on pp. 4, 8).

[Owe08] J. Owens, M. Houston, D. Luebke, S. Green, J. Stone, and J. Phillips. “GPU Computing”. In: *Proceedings of the IEEE* 96.5 (2008), pp. 879–899. DOI: 10.1109/JPROC.2008.917757 (cit. on p. 123).

[Pap08] N. Papenberg, J. Modersitzki, and B. Fischer. “Registrierung im Fokus: Beschleunigung Variationeller Methoden für die Bildregistrierung”. In: *Bildverarbeitung für die Medizin 2008*. Ed. by T. Tolxdorff, J. Braun, T. M. Deserno, A. Horsch, H. Handels, and H.-P. Meinzer. Springer, 2008, pp. 138–142. DOI: 10.1007/978-3-540-78640-5_28 (cit. on p. 40).

[Pat14] D. A. Patterson and J. L. Hennessy. *Computer Organization and Design: The Hardware/Software Interface*. Fifth edition. The Morgan Kaufmann Series in Computer Architecture and Design. Elsevier/Morgan Kaufmann, 2014 (cit. on p. 9).

[Pat18] R. C. of Pathologists. *Meeting Pathology Demand – Histopathology Workforce Census*. 2018 (cit. on p. 4).

[Pet18] A. H. Petriceks and D. Salmi. “Trends in Pathology Graduate Medical Education Programs and Positions, 2001 to 2017”. In: *Academic Pathology* 5 (2018), p. 237428951876545. DOI: 10.1177/2374289518765457 (cit. on p. 4).

[Pic18] J. Pichat, J. E. Iglesias, T. Yousry, S. Ourselin, and M. Modat. “A Survey of Methods for 3D Histology Reconstruction”. In: *Medical Image Analysis* 46 (2018), pp. 73–105. DOI: 10.1016/j.media.2018.02.004 (cit. on pp. 37, 39).

[Pit06] A. Pitiot, E. Bardinet, P. M. Thompson, and G. Malandain. “Piecewise Affine Registration of Biological Images for Volume Reconstruction.” In: *Medical image analysis* 10.3 (2006), pp. 465–83. DOI: 10.1016/j.media.2005.03.008 (cit. on pp. 5, 39, 40).

[Plu03] J. Pluim, J. Maintz, and M. Viergever. “Mutual-Information-Based Registration of Medical Images: A Survey”. In: *IEEE Transactions on Medical Imaging* 22.8 (2003), pp. 986–1004. DOI: 10.1109/TMI.2003.815867 (cit. on p. 20).

[Pol14] T. Polzin, J. Rühaak, R. Werner, H. Handels, and J. Modersitzki. “Lung Registration Using Automatically Detected Landmarks”. In: *Methods of Information in Medicine* 53.04 (2014), pp. 250–256. DOI: 10.3414/ME13-01-0125 (cit. on p. 16).

[Rob12] N. Roberts, D. Magee, Y. Song, K. Brabazon, M. Shires, D. Crellin, N. M. Orsi, R. Quirke, P. Quirke, and D. Treanor. “Toward Routine Use of 3D Histopathology as a Research Tool.” In: *The American journal of pathology* 180.5 (2012), pp. 1835–42. DOI: 10.1016/j.ajpath.2012.01.033 (cit. on pp. 1, 5, 38–40).

[Røg16] R. Røge, R. Riber-Hansen, S. Nielsen, and M. Vyberg. “Validation Of Virtual Double Staining For Estimation Of Ki67 Proliferation Indices In Breast Carcinomas”. In: *Diagnostic Pathology* 1.8 (2016). DOI: 10.17629/www.diagnosticpathology.eu-2016-8:119 (cit. on p. 40).

[Rok99] J. Rokicki and J. Floryan. “Higher-Order Unstructured Domain Decomposition Method for Navier–Stokes Equations”. In: *Computers & Fluids* 28.1 (1999), pp. 87–120. DOI: 10.1016/S0045-7930(98)00007-3 (cit. on p. 77).

[Ron17] A. Ronovsky and A. Vasatova. “Elastic Image Registration Based on Domain Decomposition with Mesh Adaptation”. In: *Advances in Electrical and Electronic Engineering* 15.2 (2017). DOI: 10.15598/aeee.v15i2.2281 (cit. on p. 1).

[Rüh13] J. R. Rühaak, L. Konig, M. Hallmann, N. Papenberg, S. Heldmann, H. Schumacher, and B. Fischer. “A Fully Parallel Algorithm for Multimodal Image Registration Using Normalized Gradient Fields”. In: 2013 IEEE 10th International Symposium on Biomedical Imaging (ISBI 2013). IEEE, 2013, pp. 572–575. DOI: 10.1109/ISBI.2013.6556539 (cit. on p. 25).

[Rüh17a] J. R. Rühaak. “Matrix-Free Techniques for Efficient Image Registration and Their Application to Pulmonary Image Analysis”. PhD Thesis. Jacobs University, 2017 (cit. on p. 107).

[Rüh17b] J. R. Rühaak, T. Polzin, S. Heldmann, I. J. A. Simpson, H. Handels, J. Modersitzki, and M. P. Heinrich. “Estimation of Large Motion in Lung CT by Integrating Regularized Keypoint Correspondences into Dense Deformable Registration”. In: *IEEE Transactions on Medical Imaging* 36.8 (2017), pp. 1746–1757. DOI: 10.1109/TMI.2017.2691259 (cit. on p. 16).

[Saa03] Y. Saad. *Iterative Methods for Sparse Linear Systems*. Second. Society for Industrial and Applied Mathematics, 2003. DOI: 10.1137/1.9780898718003 (cit. on pp. 34, 45, 46, 67, 68).

[Saa12] S. Saalfeld, R. Fetter, A. Cardona, and P. Tomancak. “Elastic Volume Reconstruction from Series of Ultra-Thin Microscopy Sections”. In: *Nature Methods* 9.7 (2012), pp. 717–720. DOI: 10.1038/nmeth.2072 (cit. on p. 123).

[Sat13] M. Satyanarayanan, A. Goode, B. Gilbert, J. Harkes, and D. Jukic. “OpenSlide: A Vendor-Neutral Software Foundation for Digital Pathology”. In: *Journal of Pathology Informatics* 4.1 (2013), p. 27. DOI: 10.4103/2153-3539.119005 (cit. on p. 107).

[Sch06] S. Schaefer, T. McPhail, and J. Warren. “Image Deformation Using Moving Least Squares”. In: ACM SIGGRAPH 2006 Papers. ACM Press, 2006, pp. 533–540. DOI: 10.1145/1179352.1141920 (cit. on p. 39).

[Sch07] O. Schmitt, J. Modersitzki, S. Heldmann, S. Wirtz, and B. Fischer. “Image Registration of Sectioned Brains”. In: *International Journal of Computer Vision* 73.1 (2007), pp. 5–39. DOI: 10.1007/s11263-006-9780-x (cit. on pp. 4, 5, 41).

[Sch13] M. Schwier, T. Böhler, H. Hahn, U. Dahmen, and O. Dirsch. “Registration of Histological Whole Slide Images Guided by Vessel Structures”. In: *Journal of Pathology Informatics* 4.2 (2013), pp. 10–10. DOI: 10.4103/2153-3539.109868 (cit. on pp. 1, 38).

[Sch70] H. A. Schwarz. *Ueber Einen Grenzübergang Durch Alternirendes Verfahren*. Vierteljahrsschrift Der Naturforschenden Gesellschaft in Zürich. Zürcher u. Furrer, 1870 (cit. on pp. 43, 44).

[Sch98] T. Schormann and K. Zilles. “Three-Dimensional Linear and Nonlinear Transformations: An Integration of Light Microscopical and MRI Data”. In: *Human brain mapping* 6.5-6 (1998), pp. 339–347. DOI: 10.1002/(SICI)1097-0193(1998)6:5/6<339::AID-HBM3>3.0.CO;2-Q (cit. on p. 4).

[Sew08] J. Seward, ed. *Valgrind 3.3: Advanced Debugging and Profiling for Gnu/Linux Applications*. 1. print. Network Theory, 2008 (cit. on p. 118).

[Sha10] R. Shams, P. Sadeghi, R. Kennedy, and R. Hartley. “A Survey of Medical Image Registration on Multicore and the GPU”. In: *IEEE Signal Processing Magazine* 27.2 (2010), pp. 50–60. DOI: 10.1109/MSP.2009.935387 (cit. on pp. 41, 43).

[Sir16] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot. “Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images”. In: *IEEE Transactions on Medical Imaging* 35.5 (2016), pp. 1196–1206. DOI: 10.1109/TMI.2016.2525803 (cit. on p. 9).

[Sla11] M. Slaoui and L. Fiette. “Histopathology Procedures: From Tissue Sampling to Histopathological Evaluation”. In: *Drug Safety Evaluation*. Ed. by J.-C. Gautier. Vol. 691. Humana Press, 2011, pp. 69–82. DOI: 10.1007/978-1-60761-849-2_4 (cit. on p. 3).

[Sla19] J. T. Slagel, J. Chung, M. Chung, D. Kozak, and L. Tenorio. “Sampled Tikhonov Regularization for Large Linear Inverse Problems”. In: *Inverse Problems* 35.11 (2019), p. 114008. DOI: 10.1088/1361-6420/ab2787 (cit. on p. 123).

[Smi15] E. Smistad, T. L. Falch, M. Bozorgi, A. C. Elster, and F. Lindseth. “Medical Image Segmentation on GPUs – A Comprehensive Review”. In: *Medical Image Analysis* 20.1 (2015), pp. 1–18. DOI: 10.1016/j.media.2014.10.012 (cit. on p. 123).

[Son14] Y. Song, D. Treanor, A. J. Bulpitt, N. Wijayathunga, N. Roberts, R. Wilcox, and D. R. Magee. “Unsupervised Content Classification Based Nonrigid Registration of Differently Stained Histology Images”. In: *IEEE Transactions on Biomedical Engineering* 61.1 (2014), pp. 96–108. DOI: 10.1109/TBME.2013.2277777 (cit. on pp. 16, 39, 40).

[Sot13] A. Sotiras, C. Davatzikos, and N. Paragios. “Deformable Medical Image Registration: A Survey”. In: *IEEE Transactions on Medical Imaging* 32.7 (2013), pp. 1153–1190. DOI: 10.1109/TMI.2013.2265603 (cit. on pp. 20, 37).

[Sta14] E. C. Stack, C. Wang, K. A. Roman, and C. C. Hoyt. “Multiplexed Immunohistochemistry, Imaging, and Quantitation: A Review, with an Assessment of Tyramide Signal Amplification, Multispectral Imaging and Multiplex Analysis”. In: *Methods* 70.1 (2014), pp. 46–58. DOI: 10.1016/j.ymeth.2014.08.016 (cit. on p. 4).

[Thé00] P. Thévenaz, T. Blu, and M. Unser. “Image Interpolation and Resampling”. In: *Handbook of Medical Imaging: Processing and Analysis*. Ed. by I. N. Bankman. Academic Press Series in Biomedical Engineering. Academic Press, 2000 (cit. on pp. 8, 22, 23).

[Tho18] P. Thoman, K. Dichev, T. Heller, R. Iakymchuk, X. Aguilar, K. Hasanov, P. Gschwandtner, P. Lemarinier, S. Markidis, H. Jordan, T. Fahringer, K. Katrinis, E. Laure, and D. S. Nikolopoulos. “A Taxonomy of Task-Based Parallel Programming Technologies for High-Performance Computing”. In: *The Journal of Supercomputing* 74.4 (2018), pp. 1422–1434. DOI: 10.1007/s11227-018-2238-4 (cit. on p. 41).

[Tro01] U. Trottenberg, C. W. Oosterlee, and A. Schüller. *Multigrid*. Academic Press, 2001 (cit. on p. 29).

[Tus06] N. J. Tustison and J. C. Gee. “Generalized N-D Ck B-Spline Scattered Data Approximation with Confidence Values”. In: *Medical Imaging and Augmented Reality*. Ed. by G.-Z. Yang, T. Jiang, D. Shen, L. Gu, and J. Yang. Red. by D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, and G. Weikum. Vol. 4091. Springer, 2006, pp. 76–83. DOI: 10.1007/11812715_10 (cit. on p. 55).

[Vio97] P. Viola and W. M. Wells III. “Alignment by Maximization of Mutual Information”. In: *International Journal of Computer Vision* 24.2 (1997), pp. 137–154. DOI: 10.1023/A:1007958904918 (cit. on p. 20).

[Wal92] G. Wallace. “The JPEG Still Picture Compression Standard”. In: *IEEE Transactions on Consumer Electronics* 38.1 (1992), pp. xviii–xxxiv. DOI: 10.1109/30.125072 (cit. on p. 9).

[War10] D. A. Warrell, T. M. Cox, and J. D. Firth, eds. *Oxford Textbook of Medicine*. 5th ed. Oxford University Press, 2010 (cit. on p. 107).

[Wei15] N. Weiss, J. Lotz, and J. Modersitzki. “Multimodal Image Registration in Digital Pathology Using Cell Nuclei Densities”. In: *Bildverarbeitung Für Die Medizin 2015*. Ed. by H. Handels, T. M. Deserno, H.-P. Meinzer, and T. Tolxdorff. Springer, 2015, pp. 245–250. DOI: 10.1007/978-3-662-46224-9_43 (cit. on pp. 9, 12, 13, 39).

[Wil18] B. J. Williams, J. Lee, K. A. Oien, and D. Treanor. “Digital Pathology Access and Usage in the UK: Results from a National Survey on Behalf of the National Cancer Research Institute’s CM-Path Initiative”. In: *Journal of Clinical Pathology* 71.5 (2018), pp. 463–466. DOI: 10.1136/jclinpath-2017-204808 (cit. on p. 4).

[Xu15] Y. Xu, J. G. Pickering, Z. Nong, E. Gibson, J.-M. Arpino, H. Yin, and A. D. Ward. “A Method for 3D Histopathology Reconstruction Supporting Mouse Microvasculature Analysis.” In: *PloS one* 10.5 (2015), e0126817–e0126817. DOI: 10.1371/journal.pone.0126817 (cit. on p. 5).

[Yig17] M. Yigitsoy and G. Schmidt. “Hierarchical Patch-Based Co-Registration of Differently Stained Histopathology Slides”. In: ed. by M. N. Gurcan and J. E. Tomaszewski. 2017, p. 1014009. DOI: 10.1117/12.2254266 (cit. on p. 16).

[Yin16] Y. Yin, O. Sedlaczek, J. Lotz, J. Olesch, K. Breuhahn, D. Drasdo, and I. E. Vignon-Clementel. “Tumor Microvasculature in Lung Cancer and Diffusion-Weighted MRI: Preliminary Results”. In: *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop Proceedings (NSS/MIC/RTSD)*. IEEE, 2016. DOI: 10.1109/NSSMIC.2016.8069545 (cit. on pp. 12–14, 38).

[Yun07] T. Yunhao, H. Jing, and D. Ming. “Feature Curve-Guided Volume Reconstruction from 2D Images”. In: *Proceedings of IEEE ISBI*. 2007, pp. 716–719. DOI: 10.1109/ISBI.2007.356952 (cit. on p. 5).

[Zar19] M. D. Zarella, D. Bowman; F. Aeffner, N. Farahani, A. Xthona; S. F. Absar, A. Parwani, M. Bui, and D. J. Hartman. “A Practical Guide to Whole Slide Imaging: A White Paper From the Digital Pathology Association”. In: *Archives of Pathology & Laboratory Medicine* 143.2 (2019), pp. 222–234. DOI: 10.5858/arpa.2018-0343-RA (cit. on p. 4).

[Zha16] C. Zhao, A. Carass, A. Jog, and J. L. Prince. “Effects of Spatial Resolution on Image Registration”. In: SPIE Medical Imaging. Ed. by M. A. Styner and E. D. Angelini. 2016, 97840Y. DOI: 10.1117/12.2217322 (cit. on p. 61).

[Zha94] Z. Zhang. “Iterative Point Matching for Registration of Free-Form Curves and Surfaces”. In: *International Journal of Computer Vision* 13.2 (1994), pp. 119–152. DOI: 10.1007/BF01427149 (cit. on p. 9).