Master's thesis

Faculty of Science and Technology
Multi-Modality Medical Imaging

Towards Longitudinal Three-Dimensional Photoacoustic Imaging:
An Automatic Image Registration Framework

Lucia Kim

Supervisors:
Prof. Dr. S. Manohar
Dr. B. De Santi

External committee member:
Dr. J.M. Wolterink

November 2023

Multi-Modality
M3i
Medical Imaging

UNIVERSITY OF TWENTE.

Abstract

Problem statement:

According to Global Cancer Statistics, breast cancer was the most diagnosed cancer worldwide in 2020, which caused 6.9% of the total cancer deaths, leading to a substantial societal and economic impact.

Neoadjuvant chemotherapy (NAC) is a systemic treatment before tumor surgery and it has made significant progress in the overall survival rate and is becoming standard care in breast cancer. However, not every patient responds to NAC. In such cases, a whole course of NAC can lead to a delayed start of another treatment and unnecessary exposure of the patients to the toxic effects of the drugs. Therefore, a longitudinal imaging protocol is needed to acquire multiple images over time to monitor the effects of NAC and prevent further disease progression due to delays in treatment. Ultrasound and magnetic resonance imaging are common clinical imaging modalities primarily employed for structural imaging., However, these modalities have several drawbacks. Their main limitation is that these modalities detect structural changes, such as a change in tumor size, which occur with a delay to changes in the tumor microstructure and do not correlate with patient outcomes.

Photoacoustic (PA) imaging is one of the most emerging modalities for capturing and quantifying tumor angiogenesis and hypoxia. It can visualize blood vessels in the breast with sub-millimeter resolution at depths of more than 5 cm. However, a valid image registration method is needed to align PA images over time to monitor NAC's effect. Repositioning the breast leads to complex and non-linear deformations of the vasculature and the breast, making quantitative analysis impossible. Current image registration frameworks are not suitable for image registration of PA images due to limitations like requiring a large training data set or not having functionalities to align sparse data, like PA images.

Aim and approach:

This thesis proposes a new robust machine learning framework, MUVINN, which uses a coordinate-based neural network to represent the displacement field of the PA image pair. By using a loss function based on normalized cross-correlation and Frangi vesselness filter at multiple scales, it can align vascular images effectively.

The algorithm is tested on an in vivo data set of breast PA images of a healthy volunteer acquired with the Twente Photoacoustic Mammoscope 3 to validate the framework. First by synthetically deforming the existent images and then by repeating measurements after repositioning the breast under normal conditions and unfavorable conditions, such as using different illumination wavelengths, purposefully mispositioning the volunteer, using a different breast-supporting cup size, and without the use of a cup.

Results:

MUVINN has shown excellent performance in registering synthetically deformed images and repeated images in normal conditions and challenging conditions. It has been shown to be robust to shifts in image intensity and field-of-view inconsistencies.

Conclusions:

MUVINN is a promising tool for quantitatively monitoring disease progression and treatment response in breast cancer using photoacoustic. It has been proven to work for unimodal image registration of PA images with a Twente Photoacoustic Mammoscope 3 imager. However, more research is needed for multimodal image registration for broader applications.

Table of contents

Table of contents

2   Introduction

2.1   Clinical context

According to Global Cancer Statistics, female breast cancer is the most commonly diagnosed cancer worldwide in 2020, with 2.3 million new cases (11.7%) of all diagnosed cancer cases. It caused 6.9% of the total cancer deaths [1], leading to a substantial societal and economic impact. One significant progress in the overall survival rate is using systemic therapy based on breast cancer subtypes. Especially neoadjuvant chemotherapy (NAC), a systemic treatment before tumor surgery, has improved cancer outcomes in breast cancer patients and is progressively becoming the standard care [2].

However, patients' response to NAC varies depending on their breast cancer subtypes. Completing the full course of NAC can last around four months. For non-responders, this can lead to a delayed start of another treatment and unnecessary exposure of the patient to the toxic effects of the drugs used [3]. Therefore, accurate evaluation of the efficacy of NAC is crucial to distinguish responders from non-responders. This will prevent unnecessary exposure of non-responders to the toxicity and costs of ineffective treatment and prevent further disease progression due to postponed surgery. Identifying patients with tumors with a high likelihood of achieving a pathological complete response (pCR) at the first cycles of NAC can significantly impact individualized treatment.

In current clinical practice, structural imaging techniques like magnetic resonance imaging (MRI), ultrasound (US), and mammography are performed to monitor tumor response during NAC. However, structural changes, such as a change in tumor size, occur with a delay to changes in the tumor microstructure and do not correlate with patient outcomes [3–5].

Functional imaging techniques could be a good alternative to structural imaging techniques, as they provide valuable insights into tumor characteristics during chemotherapy. However, some of these methods have limitations for longitudinal studies as they are cost-intensive and time-consuming or require an intravenous injection of contrast agents after each course of NAC [3].

A promising functional imaging technique that can visualize blood vessels is photoacoustic (PA) imaging. Hemoglobin and oxyhemoglobin are widely used absorbers in photoacoustics, which visualize blood vessels in the breast with sub-millimeter resolution at depths of more than 5 cm [6]. Due to its ability to visualize deep blood vessels in high resolution and its non-invasive nature, PA imaging could be a viable option to replace the current standard of care imaging modalities in monitoring NAC in breast cancer. However, for longitudinal imaging with PA imaging, image registration is needed as repositioning of the breast lead to deformations in the vasculature and the breast leading to difficulties in quantitative analysis of the changes that occur in the breast during the treatment.

2.2   Research problem statement

An image registration framework is needed to follow the treatment progress and do a quantitative analysis. Twente Photoacoustic Mammoscope 3 is used as a PA imager for this research, which uses a breast-supporting cup to improve the light distribution and stabilize the breast to prevent motion artifacts. However, that is not enough to make the images reproducible. Due to the ability of breast tissue to change shape, repositioning the breast will introduce repositioning errors and may change the field-of-view. Images should be aligned through image registration to find the potential imaging biomarkers to evaluate the treatment response so that local variations of PA signal intensities can be tracked over time. There is a need for an automatic and accurate image registration framework for tomographic PA images.

2.3 Research objectives and research questions

The objective of the research is to:

- Develop a machine learning-based framework for the registration of PA images.

- Validate whether the framework is suitable for longitudinal imaging to monitor neoadjuvant chemotherapy.

These objectives lead to the overarching research question of this thesis: How can we develop and validate an accurate and robust image registration framework to monitor the effects of neoadjuvant chemotherapy in breast cancer using photoacoustic images?

In the upcoming chapters, these research questions will be addressed to guide the thesis to answer the overarching research question:

Standard care of NAC and the role of PA

1. Which imaging modalities are currently employed in standard NAC monitoring, and do these modalities meet the required criteria for effective monitoring
2. Is PA imaging a viable alternative to existing standard imaging modalities for monitoring neoadjuvant chemotherapy (NAC), and if so, what are its advantages and limitations?

The image registration problem

3. What are the challenges of conducting vascular image registration?
4. What are the existing image registration frameworks, and are they suitable for aligning vascular images?
5. How can the performance of image registration methods be accurately assessed and evaluated?
6. What are the requirements of a robust image registration framework for longitudinal PA imaging?

MUVINN registration framework

7. What is MUVINN, and how does it differ from other existing image registration frameworks?
8. What is the performance of MUVINN, and how is it compared to other image registration frameworks?

Performance analysis of MUVINN

9. What are the appropriate validation methodologies to assess its accuracy?

10. How does MUVINN perform in a realistic scenario and a challenging scenario?

2.4 Significance of the research

Monitoring the treatment progress of NAC is crucial because the response to NAC varies among patients, and delayed start of effective treatment can lead to disease progression and unnecessary exposure to drug toxicity. However, the current standard of care imaging modalities for monitoring NAC have limitations, and functional imaging techniques, such as PA imaging, can be a good alternative.

To find potential imaging biomarkers to evaluate the treatment response, images should be aligned through image registration, and local variations of PA signal intensities should be tracked over time. For this purpose, an automatic image registration for vascular images is needed.

3 Standard of care in monitoring neo-adjuvant chemotherapy

This chapter provides some background information about NAC and standard-of-care imaging modalities for monitoring NAC. First, it explores the necessity of monitoring during NAC and the effectiveness of the currently used imaging modalities. Lastly, it considers the potential role of PA imaging as an alternative to these imaging modalities and evaluates its characteristics for NAC monitoring.

3.1 NAC

The main aim of NAC is to achieve a pathological complete response (pCR), defined as the disappearance of all invasive cancer in the breast after the completion of NAC. When pCR is reached, the risk of metastases and micro-metastases in distant organs, the frequency of the recurrences, and the mortality of patients are decreased [3].

However, patients' response to NAC varies depending on their breast cancer subtypes. Pathologic complete response (pCR) rate ranges from 6% to 33% across different breast cancer subtypes. It has been reported that pCR obtained after NAC is a suitable surrogate endpoint for disease-free survival in patients with luminal B/Human Epidermal growth factor Receptor 2 (HER2) -negative, HER2-positive (non-luminal) and triple-negative(TN) tumors but not for those with luminal B/HER2-positive or luminal A tumors [7].

However, prediction models based on these tumor histopathological characteristics are insufficient. Within these subgroups, responses to chemotherapy also show considerable variation. For instance, HER2-positive patients often receive dual HER2 blockade with trastuzumab and pertuzumab in combination with cytotoxic chemotherapy, which yields a high pCR rate. However, a significant percentage of HER2-positive patients still do not achieve a pCR or near pCR [8].

In such cases, a whole course of NAC, which can last about four months, leads to a delayed start of another treatment and unnecessary exposure of the patient to the toxic effects of the drugs used [3]. Therefore, accurate evaluation of the efficacy of NAC is crucial to distinguish responders from non-responders, to avoid unnecessary exposure of non-responders to the toxicity and costs of ineffective treatment, and to prevent further disease progression due to postponed surgery. The ability to identify patients with tumors that have a high likelihood of achieving a pCR before starting NAC or at the first cycles of NAC can have a major impact on individualized treatment.

3.2 Imaging modalities

In current clinical practice, structural imaging techniques like magnetic resonance imaging (MRI), ultrasound (US), and mammography is performed to monitor tumor response during NAC. These techniques use lesion size and volume changes to predict the NAC response. However, changes in tumor size occur with a delay to changes in the tumor microstructure, and the changes in tumor size often do not correlate with patient outcome [3–5].

Functional imaging techniques provide valuable insights into tumor characteristics during chemotherapy, specifically capturing changes in microstructure, vascularization, and metabolic activity after the initial treatment cycle [3]. These techniques include positron emission tomography (PET), dynamic contrast-enhanced (DCE)-MRI, diffusion-weighted imaging (DWI), color Doppler imaging, elastography, and photoacoustic (PA) imaging, from which color Doppler imaging, elastography, and photoacoustic (PA) imaging are only used in research. However, some of these methods are cost-intensive, time-consuming, or require an intravenous injection of contrast agents after each course of chemotherapy [3].

The following sections will discuss some imaging modalities used for monitoring NAC, highlighting their characteristics, benefits, and limitations.

3.2.1 Mammography

In mammography, signs of tumor lesion calcification disappearance and changes in tumor size and density can be used to predict breast tumor response to NAC. Mammography can show the malignant calcification well, but it is demonstrated that the presence of residual microcalcification after NAC is not always an indication of a residual tumor burden. Instead, residual microcalcification can result from residual tumors and byproducts of necrotic tumor cells post-treatment [9].

In addition, Skarping et al. compared the size of tumors evaluated by mammography and ultrasound after NAC and reported the sensitivity and specificity of mammography as 65% and 81%, respectively. This study also showed that the agreement rate between mammography and pathological assessment in pCR was only 43% [10]. Thus, although mammography is highly specific in detecting tumors, it overestimates the tumor size in about half of the patients. It shows a poor consistency between mammography and pathological results after NAC.

Next to the challenges regarding prediction accuracy, the patient is also exposed to X-ray radiation using mammography, which can be harmful for repetitive measurements. Due to this, most experts consider mammography unsuitable for monitoring NAC in breast cancer [11].

In conclusion, mammography cannot be a gold standard for assessing the efficacy of NAC in breast cancer due to several limitations:

- Residual microcalcification after NAC is often unrelated to the residual tumor.

- It is unable to determine the changes in tumor size accurately.

- It has X-ray radiation, so it cannot be used to examine frequently.

3.2.2 Magnetic resonance imaging

MRI can accurately assess primary lesion dimension, multifocality, multicentricity, and lymph node involvement for tumor response to NAC [12]. The therapeutic effect can be judged mainly by morphology and by measuring the change in the maximum diameter of the lesion in conventional MRI. Lesion size measured by MRI shows a higher correlation with pathological examination than mammography, ultrasound, or clinical palpation [11]. However, MRI tends to overestimate the size of residual cancer due to chemotherapy-induced fibrosis and inflammation [13].

DCE-MRI and DWI-MRI can improve the monitoring of breast tumors by evaluating functional parameter changes during NAC. DCE-MRI can reflect tumor tissue changes based on contrast distribution. DWI-MRI can reflect changes based on apparent diffusion coefficients (ADC), a quantitative measure of the diffusivity of water. ADC provides information related to tumor cellularity and the integrity of cell membranes, which are affected by the intra-tumoral changes induced by chemotherapy [12].

3.2.2.1 Dynamic Contrast Enhanced-MRI

DCE-MRI is a functional MRI technique that requires an intravenous contrast injection of a low molecular weight T1-shortening paramagnetic compound like gadolinium. It can quantify changes in tumor microvasculature, cell density, hypoxia, metabolism, and stiffness to monitor NAC efficacy [14]. Parameters that DCE-MRI can measure to monitor NAC are K_{trans} (contrast agent plasma/interstitial transfer rate

constant), K_{ep} (intravasation rate constant), V_e (extravascular and extracellular volume fraction), and longest tumor diameter [15]. Despite its high sensitivity, DCE-MRI exhibits lower specificity [14]. While it has proven effective in predicting TN or HER2-positive cancers, its accuracy is reduced for ER-positive/HER2-negative breast cancers [8]. Furthermore, DCE-MRI has several limitations which make it unsuitable for repeated use, including its high cost, long duration, and requirement of intravenous contrast agents [16]

3.2.2.2   Diffusion-weighted imaging

DWI-MRI can assess the efficacy of NAC in patients with highly cellular breast cancers. DWI can quantitatively measure the impedance and motion of water in tissue. In the case of tumor, it shows restricted free water molecule movement and lower ADC values. During NAC, tumor density and cellularity decrease, and the ADC values rise, even before the tumor's size change can be detected. However, DWI has some limitations, such as poor spatial resolution, prone to motion artifacts, and difficulty characterizing specific subtypes of breast cancer [4].

To summarize, the main limitations of MRI to assess the efficiency of NAC are:

- Conventional MRI tends to overestimate the size of residual cancer.

- DCE-MRI requires intravenous contrast injection, making it unsuitable for repeated use during treatment due to its high cost and long duration.

- DWI has limitations like poor spatial resolution, susceptibility to motion artifacts, and difficulty characterizing specific breast cancer subtypes.

3.2.3   Ultrasound

Ultrasound (US) can monitor the changes in size, shape, vascularity, and elasticity during NAC safely and inexpensively. The conventional US can describe tumors' size, morphology, and boundary. Color Doppler can describe vascularity, and elastography can describe the elasticity of the tumor [11].

Although the US can be used for both structural and functional imaging, there are several reasons the US could not be the first choice for NAC monitoring. US diagnosis depends on the operator's skills, the scanning time is relatively long, and it cannot accurately distinguish tumor tissue from normal gland tissue [11] and fibrotic scar tissue [17]. Moreover, due to its hand-held nature, imaging and registering the whole breast is challenging, and the reproducibility is low [18].

Baumgartner et al. concluded that conventional US imaging is insufficient to predict pCR due to its low accuracy, with a sensitivity of 60.8% and specificity of 78.0% for the US-predicted residual tumor burden [19]. The lower prediction efficiency might be caused due to the potential limitation of the ability of the US to distinguish viable tumor tissue from fibrotic scar tissue [17] or gland tissue [11].

3.2.3.1   Color Doppler Imaging

Color Doppler imaging can evaluate functional parameters, such as tumor vascular distribution, to assess the response to NAC. Kumar et al. found that Doppler ultrasound has a high sensitivity but low specificity, with a sensitivity and specificity of 91.7% and 38.5%. A biomarker that may reflect the efficiency of NAC is an early decrease or disappearance of tumor vascularity, which can be evaluated with intratumoral flow signals (e.g., RI, PI, and Vmax).

However, Doppler has several limitations. Doppler cannot assess microvascular circulation, while tumor growth depends on microvascular density. Flow velocity below 1 cm/s is challenging to detect. Contrast agents, such as microbubbles, are needed to overcome this limitation. Like conventional ultrasound, Doppler

has a low reproducibility and depends highly on the operator's skills. Small low-pressure vessels can be influenced by probe pressure [20].

3.2.3.2 Elastography

Strain elastography (SE) and shear wave elastography (SWE) are commonly used for breast cancer evaluation. The decrease in tumor stiffness is a good predictor of a pathological response during NAC. SE can measure the tissue softness and hardness to evaluate NAC efficacy. The evaluation compares the elastic score and strain rate ratio before and after NAC. SWE can measure the value of tissue elasticity, which is the absolute value of Young's modulus [11].

According to a review by Kong et al., the sensitivity and specificity of assessing tumor changes by SE after two treatment cycles were 83.3% to 84% and 80% to 85%, while the sensitivity and specificity of determining tumor changes by SWE after two NAC cycles were 72.9% and 85.7% [11].

However, certain factors can affect the results of SE and SWE for monitoring tumor response to NAC. Barr et al. showed that breast thickness and lesion depth are essential factors affecting the quality of elastography images [21]. SE is more affected by the thickness and lesion depth than SWE [22]. Higher hardness caused by fibrosis or hyaline degeneration and uneven internal hardness of the tumor lesions after NAC can affect the measurement results [11]. Additionally, SWE can be influenced by a patient's breathing [17]. The operator's experience and knowledge also significantly impact the measurement results, like the conventional US [11].

To summarize, the US has some limitations for determining the efficiency of NAC:

- Conventional US, color Doppler imaging, and elastography highly depend on the operator's skills.

- It is difficult to distinguish tumor tissue from fibrotic scar tissue or gland tissue with conventional US.

- Color Doppler has a low specificity.

- Elastography can be affected by the thickness and lesion depth.

3.2.4 Positron emission tomography

PET/CT imaging measures NAC-induced changes in ^{18}F-FDG uptake by calculating the change in standardized uptake values (SUV). A study showed that PET/CT has a high sensitivity of 100% but a lower specificity of 63%, leading to delays in definitive treatment and unnecessary biopsies [23]. In addition, there are several other reasons why PET/CT cannot be standardized in assessing NAC efficacy [24]:

- It has low spatial resolution.

- The uptake of radioisotope varies on tumor subtype.

- It cannot reliably detect lesions measuring <1cm to differentiate benign from malignant lesions, which leads to a high false positive rate.

- It uses ionizing radiation.

A summary of the literature research about NAC efficacy studies can be found in *Table 12* in *Appendix A.*

3.3 Photoacoustic Imaging

PA imaging is a promising functional imaging technique that can visualize blood vessels [25].

PA imaging is a hybrid modality that combines optical spectroscopic contrast with ultrasonic resolution, and imaging depth [26]. In PA imaging, certain tissue constituents absorb short pulses of near-infrared (NIR) light and get thermalized. As a reaction, the tissue will expand, creating an initial pressure distribution in the tissue. When the pressure relaxes, acoustic waves will be emitted that propagate to the surface of the tissue [26]. US transducers can then detect acoustic waves to reconstruct an image that shows the map of initial pressure distribution in tissue [27].

Optical imaging provides benefits of light, like high and spectrally selective optical contrast. Acoustic imaging provides the benefits of sound, like high spatial and temporal resolution in deep tissue, due to ultrasound waves propagating in tissues with lower attenuation than optical signals [18]. It also results in fewer tissue scattering artifacts because US scattering in tissue is 2 to 3 orders of magnitude less than optical scattering [17].

Five chromophores in human tissue can absorb deeply penetrating NIR light: hemoglobin, oxyhemoglobin, lipids, melanin, and water [28]. Hemoglobin and oxyhemoglobin are widely used absorbers in photoacoustics. It can visualize blood vessels in the breast with sub-millimeter resolution at depths of more than 5 cm [6]. *Figure 1* shows an example of a photoacoustic image as maximum intensity projections (MIPs).
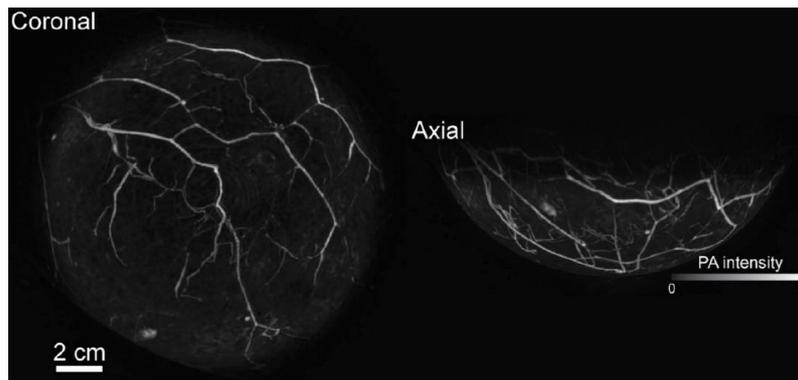
Figure 1: Photoacoustic breast image

3.3.1   Role of PA in NAC monitoring

PA imaging technology is not widely available in clinical settings like other imaging modalities such as mammography, US, and MRI. PA imaging for monitoring NAC is a relatively new field, and its clinical validation is ongoing. However, it could be a valuable alternative due to its ability to image deep breast vessels and provide hemoglobin concentration, which is valuable for detecting changes in vascularization associated with disease progression. . It has favorable characteristics for longitudinal imaging, such as no carcinogenic potential, no necessity for a contrast agent, and non-invasive and painless nature for patients.

PA imaging can help monitor tumor's response to NAC by visualizing changes in tumor vasculature. PA imaging can detect these changes, allowing for early assessment of treatment effectiveness.

A limited amount of research has been conducted on this topic. There are few vascular features that have shown differences between responder and a non-responder. These features could be used to predict the response on NAC, as these features can be visualized with a photoacoustic imager. The reported characteristics are:

- Regions with the highest vessel density often correspond with the tumor region [29].

- Higher baseline vessel density is observed before NAC in responders [30, 31].

- Reduction in vascular volume of the lesion side and vascular discrepancy between the normal side and lesion side is observed in responders [31].

- Increase in tortuosity of vasculature surrounding the tumor was detected before the NAC and decreased perfusion at the tumor core, increased hypoxia, strong contrast at the tumor rim due to increase in hemoglobin concentration caused by the reperfusion of blood vessels that had become blocked due to the treatment [32].

- Decrease in mean tumor sO_2, especially at the periphery compared to the core tumor regions [33].

However, it is not clear if these changes are associated with pCR. More research is needed to find the biomarker related to NAC effectiveness in PA imaging. For a better understanding of changes over time during NAC, the structures will need to be aligned with each other. Although the breast-supporting cup improves the light distribution and stabilizes the breast to prevent motion artifacts, it is not enough to make the images reproducible. Due to the flexibility of the breast tissue, repositioning the breast will introduce repositioning errors and may change the field-of-view. An image registration framework would enable a more comprehensive analysis of the vasculature changes over time.

3.4   Chapter discussion & conclusion

The main aim of this chapter was to (i) give the clinical context of this research, (ii) investigate which methods are currently used to monitor the tumor response to NAC and if these methods meet the requirements for effective monitoring, and (iii) to look into the characteristics of PA imaging to see if it could be a viable alternative for these methods.

The literature examined in this chapter outlines why the mentioned methods are not sufficient for early monitoring of the tumor response to NAC. Specific criteria need to be met to achieve effective monitoring of tumor response to NAC. These criteria have been formulated based on insights from the literature, which revealed that existing methods struggled to meet the following essential requirements for effective early monitoring of tumor response to NAC:

- The imaging modality should be able to demonstrate the changes after the first cycle.

- The imaging modality should be suitable for repetitive use after each cycle.

- The imaging modality should exhibit reproducibility for fair comparison of images over time.

The current standard of care monitoring techniques for NAC in breast cancer present some limitations. These limitations primarily relate to challenges in tracking biomarkers related to NAC efficacy, lack of reproducibility, and the invasive nature of certain imaging technologies for patients.

Functional imaging techniques can detect NAC efficacy in early cycles as changes in tumor microstructures occur earlier than a change in tumor size. Studies have shown that vasculature changes during the course of NAC. In this context, PA imaging emerges as a promising alternative for NAC monitoring. PA imaging has a unique ability to provide high-resolution images of the vasculatures without ionizing radiation or contrast agents.

However, despite its potential, PA imaging is not widely available at clinics. Further research is needed for a deeper understanding of the biomarkers that can accurately assess treatment response to NAC. An image registration framework would be required for reproducibility of the images and comprehensive analysis of changes over time.

4 The image registration problem

Longitudinal imaging requires that the acquired images are comparable to each other over time. Image registration makes the images comparable by matching the same anatomical structures geometrically, which enables a more in-depth analysis of medical images. During NAC, minor changes may occur in the vasculature, which might be missed, and false vasculature changes might be detected because of misalignment. Also, aligning the vasculature makes it easier to detect potential biomarkers for monitoring the tumor response. Therefore, accurate registration of PA images is desired for monitoring the tumor response to NAC.

This chapter will explain the definition of image registration, explore different image registration frameworks and formulate the requirements of a robust image registration framework. Finally, the role and examples of evaluation metrics will be described.

4.1 Image registration

Image registration is the process of aligning two or more images based on image appearances. Generally, there are two inputs for this process: a reference image, also known as a fixed image $F(x)$, and a moving image $M(x)$. The moving image will be deformed to correspond to the fixed image geometrically. In other words, the goal is to find a displacement u(x) that makes $M(x + u(x))$ spatially aligned to $F(x)$.

Registration of two images can be posed as a minimization problem in which the loss function is minimized with regard to desired spatial transformation T, that maps the voxels of M to those of F. The loss function measures the dissimilarity between the fixed and the transformed moving images. By iteratively improving estimates for the desired T, the images are registered such that the defined loss is minimized [34]. The minimization problem is expressed as:

$$\hat{T} = \underset{T}{\operatorname{argmin}} \, L_{data}\left(F, T(M)\right) + \alpha L_{reg}\left(T\right)$$

L_{data} is the loss function to measure the similarity or dissimilarity between the moving and fixed images. Adding a regularization term $L_{reg,}$ to the loss function penalize convergence to unrealistic transformations. It prevents unrealistic deformations and smoothens deformation fields. The α is a tunable parameter that determines the proportionality between the data and regularization terms. The higher the α, the smoother the solution [35]. As finding an optimal transformation is a minimization problem, there is no unique solution to this problem in general.

4.2 Challenges of vascular image registration

There are several reasons why image registration is needed for longitudinal PA imaging. In PA images, repositioning the breast can cause changes in light fluence, field-of-view, and the geometrical positions of the blood vessels. A breast-supporting cup can be used to improve the reproducibility of images [29]. However, the position of the vessels can still change significantly inside the cup due to factors like patient positioning, breast compression during imaging, and respiration due to the deformability of the breast and the operator's experience. These changes can challenge the quantitative evaluation of the temporal vascular changes [29], so image registration is needed to overcome this issue.

Vascular image registration is challenging compared to image registration of other medical images [36]. Firstly, large image regions in vascular images are non-vascular and relatively textureless. The blood vessels are the only predominant feature. Secondly, the widths of blood vessels can be variable. Larger blood vessels require different transformations than smaller structures. Also, a precise registration is necessary as the

vascular structures are relatively small. Thirdly, specific to breast imaging, the breast is quite deformable and repositioning the breast can lead to complex and unknown deformations of the vessels. An additional challenge for PA imaging is that the PA imaging technique is not widely used in clinics. Therefore, a large data set is unavailable, limiting the use of supervised or learned image registration approaches.

4.3   Examples of image registration frameworks

Traditionally, image registration was performed using computer vision and image processing techniques, which relied on manual feature extraction and optimization tasks. However, over the past years, significant growth in the development and use of deep learning-based registration methods has been observed due to the rapid improvements in computational speed with GPUs. Deep learning-based image registration methods use neural networks for image registration tasks. Particularly, convolutional neural networks (CNN) are often used in this field. The benefit of deep learning-based image registration techniques is that the network automatically extracts and learns complex image features and offers end-to-end learning [34].

4.3.1   Elastix

Elastix is an open-source software toolkit developed at the Image Sciences Institute (ISI) in University Medical Center Utrecht, Netherlands. The primary purpose of Elastix is to automate the process of aligning medical images, and it does not rely on a deep learning network.

Elastix uses a parametric approach for image registration, in which the number of possible transformations is limited by introducing a parametrization of the transformation. It requires the specification of a transformation model to align the images. Elastix supports a variety of transformation models, allowing the user to choose the one that best fits a specific registration problem. Some standard transformation models supported by Elastix include rigid, affine, B-spline transformations, etc. [37]. The advantage of using a pre-determined transformation model is that it is computationally efficient, so it is faster and could be more suitable for real-time applications. However, it may be less flexible and prone to errors as the user must know how to define the model.

Elastix does intensity-based registration by optimizing similarity metrics that measure similarity or correspondence between pixel intensities in the image pair, such as mean squared differences, mutual information and normalized correlation coefficient.

Elastix uses a multi-resolution approach to register images. It starts with downsampled, low-resolution images and progressively refines the transformation at higher resolutions [37].

The benefit of Elastix is that it can perform various types of image registration tasks, such as 2D or 3D, from rigid to non-rigid, depending on the customization. Users can select transformation models and adjust similarity metrics to their specific image registration task. However, this requires knowledge to finetune the parameters. Just like deep learning-based image registration framework, it can be challenging to understand why specific registration results were produced.

4.3.2   Implicit neural representation

In recent years, CNNs have been widely used for image registration purposes. Generally, a trained CNN can predict the deformation faster than conventional iterative approaches. However, it comes at the cost of reduced accuracy, and training a CNN network requires large training sets, which may not always be available, especially for medical imaging applications that are not widely used.

To overcome the last mentioned limitation, implicit neural representations (INR) introduced by Wolterink et al. can be used. INR uses a fully connected feedforward neural network, often called a multilayer perceptron (MLP). The network is optimized for each new image pair, so it does not need a training set [38].

INR takes fixed image coordinates as inputs and gives the displacement field as an output. The objective is to find a transformation: $\phi_{(x)} = u(x) + x$, where $u(x)$ is the displacement field and x is the coordinates of the fixed image, so the coordinate x in the fixed image would correspond to the coordinate $\phi_{(x)}$ in the moving image [38].

Unlike CNN, INR operates on continuous coordinates within the spatial domain, so it considers every point in the spatial domain to be a continuous variable. INR implicitly represents deformation vector field (DVF) within the weights of a neural network instead of explicitly representing it as a grid-based representation [39]. Due to INR's ability to represent a function continuously, it is not restricted to any particular grid resolution or image size [38]. By using MLP, INR can numerically compute gradients of the transformation with respect to coordinates so that the derivatives can be easily computed. Thus, techniques such as interpolation, which can introduce errors, are not needed [40].

As mentioned before, another benefit of using INR is that there is no need for a training set. The network is trained for one specific image pair, so training is required for every new image pair. Not requiring a training set is particularly valuable for image registration tasks where large datasets are unavailable [38]. However, the image registration task will take longer as new training is required for every image pair. Additionally, just like other deep learning-based approaches, it is hard to interpret how a network works, making it difficult to fully understand an image registration outcome.

4.4 Requirements of a robust image registration framework

Certain requirements need to be fulfilled to create a robust image registration framework for longitudinal PA imaging. Specific criteria have been defined for this thesis to evaluate if the framework is suitable for facilitating longitudinal PA imaging. These criteria serve as benchmarks to assess the framework's performance, reliability, and suitability for monitoring NAC treatment efficacy over time:

- High accuracy: The primary goal is to minimize misalignment between structures or landmarks in the images. This can be measured using evaluation metrics such as the Dice similarity coefficient (DSC) and target registration error (TRE).

- Efficiency: Although it is not for real-time use, efficiency is still needed. The framework should offer accurate registration results in an acceptable timeframe without sacrificing too much accuracy.

- Ability to correct non-rigid deformations: The deformation of the vasculature of the breast can be complex as the breast is highly deformable due to its flexibility. Therefore, the framework should be able to support non-rigid local registration [41].

- Robustness to noise: The PA image is sparse and contains a limited amount of structures to be aligned, while it contains a large non-vascular background area with potential noise. A robust registration framework should be insensitive to such noise.

- Robustness to intensity variations: PA images can show varying intensities due to illumination inhomogeneity. A robust registration framework should be invariant to such variations [42].

- Using a multiscale approach: The sizes of the vessels can be variable, so a multiscale approach will be beneficial for aligning structures of different sizes. Many frameworks have shown that adopting a

multiscale approach is effective as the framework can handle both global and local misalignments by working at multiple image resolutions [37, 43].

- Robustness to topological changes: During the treatment, topological changes in the vasculature can alter the structure of the blood vessel due to physiological or pathological processes. Some vessels might appear or disappear due to the effect of treatment or disease progression [29, 32, 44, 45]. The diameter of vessels might also vary due to physiological changes [32]. The registration algorithm should be robust to such changes.

If the image registration framework fulfils these criteria, it will indicate that the chosen framework is robust, accurate, and capable of delivering reliable results for longitudinal PA imaging for NAC monitoring.

4.5   Evaluation metrics

Image evaluation metrics provide a systematic and objective way to evaluate the images instead of subjective judgements. They enable comparisons between the image registration qualities.

Each image evaluation metric has its objectives, so it is crucial to select metrics that fit specific research goals. After literature research, the metrics were categorized based on their functions in five different categories. From each category, one metric was picked to assess various aspects of image registration quality:

- Intensity-based metrics: These metrics measure the similarity between the voxel intensities in the images. Peak signal-to-noise ratio (PSNR) and Normalized Cross-Correlation (NCC) are chosen for this category. PSNR measures the similarity between the images based on the voxel intensity. NCC measures the similarity by comparing the voxel intensities of between corresponding patches of the images. Due to normalization, it is less sensitive to intensity variations.

- Perception-based metrics: These metrics measure similarity based on characteristics that align with human visual perception. Structural Similarity Index Measure (SSIM) is chosen for this category. SSIM measures the similarity between the images by assessing changes in luminance, contrast, and structure, which is comparable with the human visual perception.

- Distance-based Metrics: These metrics measure how far apart specific points in two images are. Target Registration Error (TRE) is chosen for this category. It measures the misalignment by quantifying the Euclidean distance between landmark points in the registered image and reference image.

- Overlap-based metrics: These metrics measure the spatial overlap between the images. Dice Similarity Coefficient (DSC) is an often used method in medical imaging and is chosen for this category. DSC analyzes the alignment of segmented structures by comparing the overlap of regions against the average size of these regions.

These metrics offer not only mathematical insights but also perceptual information, ensuring a rounded evaluation. Combining all five mentioned image evaluation metrics can provide a comprehensive view of image registration quality. However, the absolute values of each metric can be difficult to interpret as their values depend from image modality and image registration problem, so to analyze the image registration quality, relative change between before and after image registration is calculated in percentage to show the impact of image registration. This relative is calculated in percentage with this formula:

$$\Delta metric = \big((after - before) \, / \, before\big) \cdot 100$$

The definition and objective of each metric will be described to explain why these metrics were chosen for this research.

4.5.1 Peak-signal-to-noise ratio

The simplest and most widely used quality metric is the mean squared error (MSE), computed by averaging the squared intensity differences of transformed and fixed image voxels.

MSE is defined as:

$$\text{MSE} = \frac{1}{|\Omega_F|}\sum\left[F(x,y,z) - M\big(T(x,y,z)\big)\right]^2 \tag{1}$$

with ΩF the domain of the fixed image F, and |ΩF| the number of voxels.

Peak signal-to-noise ratio (PSNR) is a metric which is defined via the MSE from *Equation (1):*

$$PSNR = 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \tag{2}$$

where MAX_I means the maximum intensity value of the image. It considers differences between the two images as noise or error.

PSNR focuses on the differences in pixel intensity values between the two images. A higher PSNR value indicates a higher similarity between registered images, while the smaller PSNR value indicates poor similarity between images. PSNR is appealing because it is simple to calculate and has a clear physical meaning [46].

4.5.2 Normalized cross-correlation

The NCC is a commonly used intensity-based metric that quantifies how much the pixel intensities or values in two images correlate within the corresponding regions of two images. Instead of focusing on absolute values, the NCC considers the statistical distribution of intensity values, making the NCC more robust to variations in illumination and contrast [42, 47]. NCC is also often used as a similarity metric for the loss term for image registration studies [48].

NCC is defined as:

$$\text{NCC} = \frac{\sum[F(x,y,z) - \bar{F}] \cdot [M(T(x,y,z)) - \bar{M}]}{\sqrt{\sum[F(x,y,z) - \bar{F}]^2\sum[M(T(x,y,z)) - \bar{M}]^2}} \tag{3}$$

Here $F(x,y,z)$ and $M(T(x,y,z)$ are the image intensity at the same position in the fixed image and the transformed moving image. $\bar{F}$ and $\bar{M}$ are the mean pixel values of the fixed image and the transformed moving image. So, the mean intensity values are subtracted, and then the image intensities of the transformed moving image and the fixed image are directly multiplied. The value will always be between +1 and -1 due to normalization. The higher the number, the better the correlation.

- A value of 1 indicates a perfect positive correlation (both signals are identical).

- A value of -1 indicates a perfect negative correlation (one signal is the negative of the other).

- A value close to 0 indicates little to no correlation (the signals are dissimilar).

4.5.3 Structural similarity index

Many intensity-based metrics (e.g. PSNR) rely on quantifying errors between the images by quantifying the difference in the values of each corresponding voxel between the fixed and transformed image. These are widely used due to their clear physical meanings and simplicity to calculate. However, they do not match the perceived visual quality very well.

SSIM perception-based method, which assesses the structural similarity between two images. SSIM metric extracts three critical features from an image and compares the two images based on these three features. The three features are luminance, contrast, and structure. Since luminance and contrast can vary across a scene, local luminance and contrast are used. It tries to replicate the human visual perception, which is highly capable of extracting structural information from a scene. In SSIM, the voxels have strong inter-dependencies, especially when they are spatially close. These dependencies carry information about the structure of the objects [46].

SSIM score is given by:

$$\text{SSIM} (\mathbf{x}, \mathbf{y}) = [l(\mathbf{x}, \mathbf{y})]^{\alpha} \cdot [c(\mathbf{x}, \mathbf{y})]^{\beta} \cdot [s(\mathbf{x}, \mathbf{y})]^{\gamma} \tag{4}$$

Here l stands for luminance, c for contrast and s for structure. $\alpha > 0, \beta > 0$ and $\gamma > 0$ are parameters used to adjust the relative importance of the three components. SSIM value is between 0 and 1. A value of 1 indicates that the two given images are very similar or the same, while a value of 0 means the two given images are very different [46].

In image registration, SSIM can tell how well the framework reserves the structural details of the reference image. SSIM can complement other metrics, such as overlap-based or distance-based metrics. Together, they can provide a more comprehensive view of registration quality, considering both structural similarity and spatial alignment.

4.5.4 Target registration coefficient

TRE is a distance-based metric in image registration assessment that measures the Euclidean distance between a specific reference voxel's position and the same voxel's position after the transformation. TRE validates the image registration performance by showing how accurately the framework aligns anatomical landmarks.

The formula for calculating TRE is:

$$TRE = \left| \sqrt{(x_{GT} - x_T)^2 + (y_{GT} - y_T)^2 + (z_{GT} - z_T)^2} \right| \tag{5}$$

where x, y, z_{GT} stands for the x, y, and z coordinates of the voxel at the ground truth position and x, y, z_T stand for the voxel's x, y, and z coordinates at the transformed position. The mean value and the standard deviation value of multiple voxels from the image pair are used to evaluate the accuracy of the image registration. A low target registration error indicates that the image registration is performed well.

4.5.5 Dice similarity coefficient

DSC is an overlap-based metric that quantifies the similarity or overlap between two images by assessing the degree of voxel-wise overlap or agreement between the fixed and transformed images' segmentations. The benefit of using DSC for assessing image registration results is that it is sensitive to slight differences in overlap and can detect minor misalignments. DSC's sensitivity is even more valuable in vascular image registration because vessels can have a width of only a few voxels. So, minor differences in distance can lead

to minimal to no overlap. A drawback is that the segmentation method and its performance can affect the DSC results.

The Dice score can be calculated by [49]:

$$DSC = \frac{2 * |X \cap Y|}{|X| + |Y|} \tag{6}$$

where X stands for the fixed image and Y stands for the moving image. The Dice score ranges from 0 to 1, where a Dice score of 0 indicates no overlap between the segmented areas, and a Dice score of 1 indicates a complete overlap between the segmented regions, see *Figure 2*.
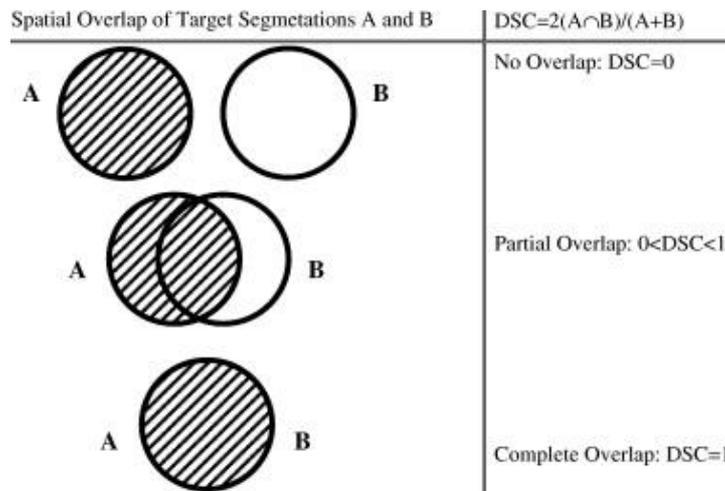
Figure 2: Dice similarity coefficient [49]

4.6 Chapter discussion & conclusion

Image registration plays an essential role in the analysis of medical images by enabling the comparison of heterogeneous data.

There are several challenges for analyzing PA images over time. First of all, there are no large data sets available. Secondly, vascular images contain large non-vascular and relatively textureless regions compared to other medical images, which can be challenging to register. Lastly, blood vessel widths are various, requiring precise registration at different scales. So, an image registration framework is needed that can overcome these challenges. Two image registration frameworks, Elastix and INR, were introduced in this chapter, which could possibly be used.

Another aspect of image registration is the evaluation of the registration results. A list of criteria is formed for this thesis to ensure that the framework is suitable for our purpose, which is longitudinal PA imaging for monitoring NAC. A quantitative performance analysis is needed to assess if these criteria are fulfilled. Evaluation metrics can be used to quantify the accuracy and quality of the alignment achieved by the registration framework. Evaluation metrics will enable objective assessment of the image registration quality, surpassing subjective human judgments.

Each metric has different objectives, so it is necessary to select the metrics that fit the purpose of the research. After literature research, the metrics were categorized based on their purposes and for each category, one or two metrics were selected to examine various aspects of the image registration. The final chosen metrics are PSR, SSIM, DSC, TRE, and NCC.

5 Algorithm: MUVINN

The algorithm MUVINN stands for Multiscale Vesselness-based Image registration using Neural Networks. It is developed by Dr. B. De Santi at the Multi-Modality-Medical Imaging group (M3I) at the University of Twente, Netherlands. The framework is based on INR, developed by Dr. J.M. Wolterink [38].

MUVINN is a robust machine-learning framework optimized for unsupervised automatic image registration of photoacoustic images in a non-rigid method. MUVINN is based on INR, but it has been adapted to suit the application of vascular image registration. MUVINN will be used for this research to register three-dimensional PA images.

5.1 MUVINN architecture

MUVINN uses a coordinate-based MLP with sinusoidal activation functions, which takes the coordinates of the reference image as an input to represent the displacement field, $u(x)$, ensuring that the transformed coordinates $\Phi(x) = x + u(x)$ on the moving image anatomically correspond to coordinates x in the fixed image. MUVINN consists of MLP with a sinusoidal activation function. The network consists of three inputs, six hidden layers, each containing 300 units with a sinusoidal activation function, except for the last layer, and three outputs in 3D image registration. See *Figure 3*.
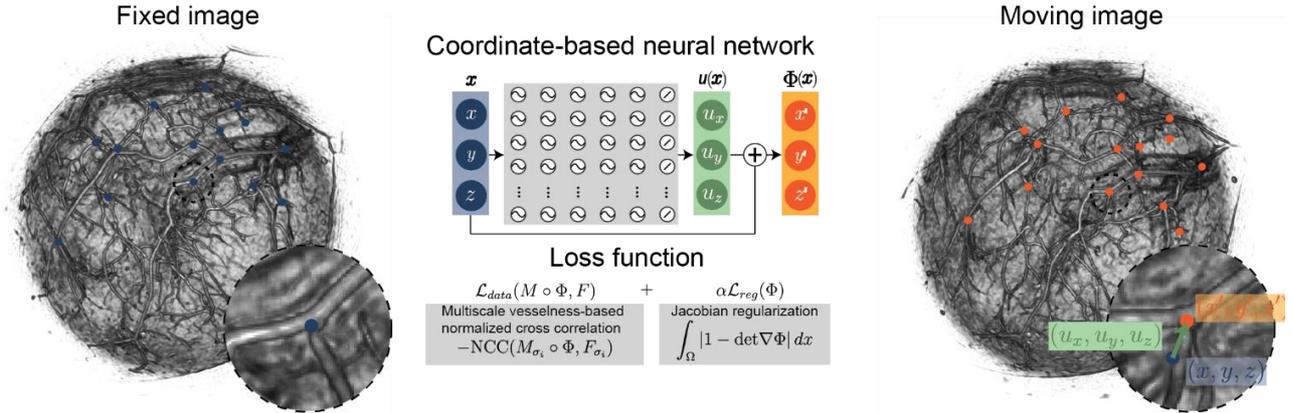
Figure 3: Overview of MUVINN frame work [50]

MUVINN is optimized for vessel image registration due to the implementation of the Frangi vesselness filter. Frangi vesselness filtering is applied to PA images to enhance the signals from the vessels. The multiscale image registration is implemented by using a higher σ value at the start of the training and stepwise decreasing the σ during the training. By doing so, the network can align the images more efficiently and accurately by focusing on the prominent vessels first and then the smaller vessels, which appear later when the σ is reduced. This approach has shown improved robustness to noise.

For each experiment, five values of sigma were defined $\sigma = \{12, 9, 5, 3, 2\}$. For each sigma, 4000 training iterations were conducted, so the total number of training iterations was 20000. Intensities of the Frangi-filtered images were adaptively modulated to improve image contrast. Portions of the image with low variability were enhanced to accentuate structures that were less visible in the image, while portions with high variability in intensity were attenuated. Adaptive modulation of intensity improved the homogeneity of voxel intensities, mitigating non-uniform illumination issues.

There are multiple benefits to using MUVINN for image registration of vessels. Firstly, it is simpler than other networks as it is unsupervised and does not require large training data like CNN. Like the original implementation of INR, MUVINN optimizes for each new image pair iteratively, so it does not generalize to unseen images. The network is only trained for one specific image pair, so the trained network only describes the transformation for the image pair that it is optimized for. However, there is still a generalizing aspect, as the network predicts the transformation for points that do not appear in training [39]. Secondly, it does not require any information about the transformation model beforehand. Thirdly, there is no restriction in different image sizes between the images in the image pair. MUVINN implicitly represents the displacement field continuously instead of explicitly using a discrete set of voxels.

5.1.1   Loss function

There are two terms in the loss function. The first term is a loss function regarding the data, and the second term is a loss function regarding the regularization. NCC between multiscale Frangi vesselness filtered images is used as a loss function. A drawback of using NCC as a loss term is that it treats all image components equally and may not capture the most effective features [48]. However, MUVINN solves this by applying multiscale Frangi vesselness filtering during the training. The multiscale approach is visualized in *Figure 4.* Starting from the top, each row shows the optimization results for each sigma from high to low. *Figure 4a* shows the size of the local patches on which the normalized cross-correlation is calculated. This patch length depends on the current sigma. The lower the sigma, the smaller the patch length. *Figure 4b* shows the training loss curve and c and d show the MIPs at the current phase's start and end. After the first iterations at sigma equal to 7, the framework succeeds in reducing the loss drastically. By the end of the first 5000 iterations, image alignment is already improved.

MUVINN network is easily differentiable as it is based on INR. A huge advantage is that MUVINN does not require numerical methods to compute regularization, while other traditional models do because regularization depends on the derivatives of the image [39]. Jacobian regularization is used to find smooth deformation fields without distortions. Jacobian regularization can penalize large expansion and shrinkage of vascular features. Weighing parameter α of 1 was used during the experiments.

5.1.2   Activation function

A sinusoidal representation network (SIREN) is implemented by using sinusoidal activation functions. It uses a sine as a periodic activation function: $\sin(\omega \cdot \mathrm{Wx} + \mathrm{b})$. Due to its use of the sine function, the output is a continuous value. This makes SIREN suitable for representing complex natural signals and their derivatives. The sinusoidal activation function has a superior ability to represent high-frequency information compared to ReLU [51]. Standard MLP based on ReLU activation functions struggles to represent high-frequency details, while SIREN can capture these details.

Angular frequency ω is a hyper-parameter of SIREN that can tune the sine function's frequency. This ω depends on the presence of high-frequency details in the data. The higher the ω, the faster the MLP converges for the high-frequency components. Siren reports $\omega=30$ to work well across a wide range of different tasks [39]. An angular frequency of $\omega=30$ is implemented during the experiments.
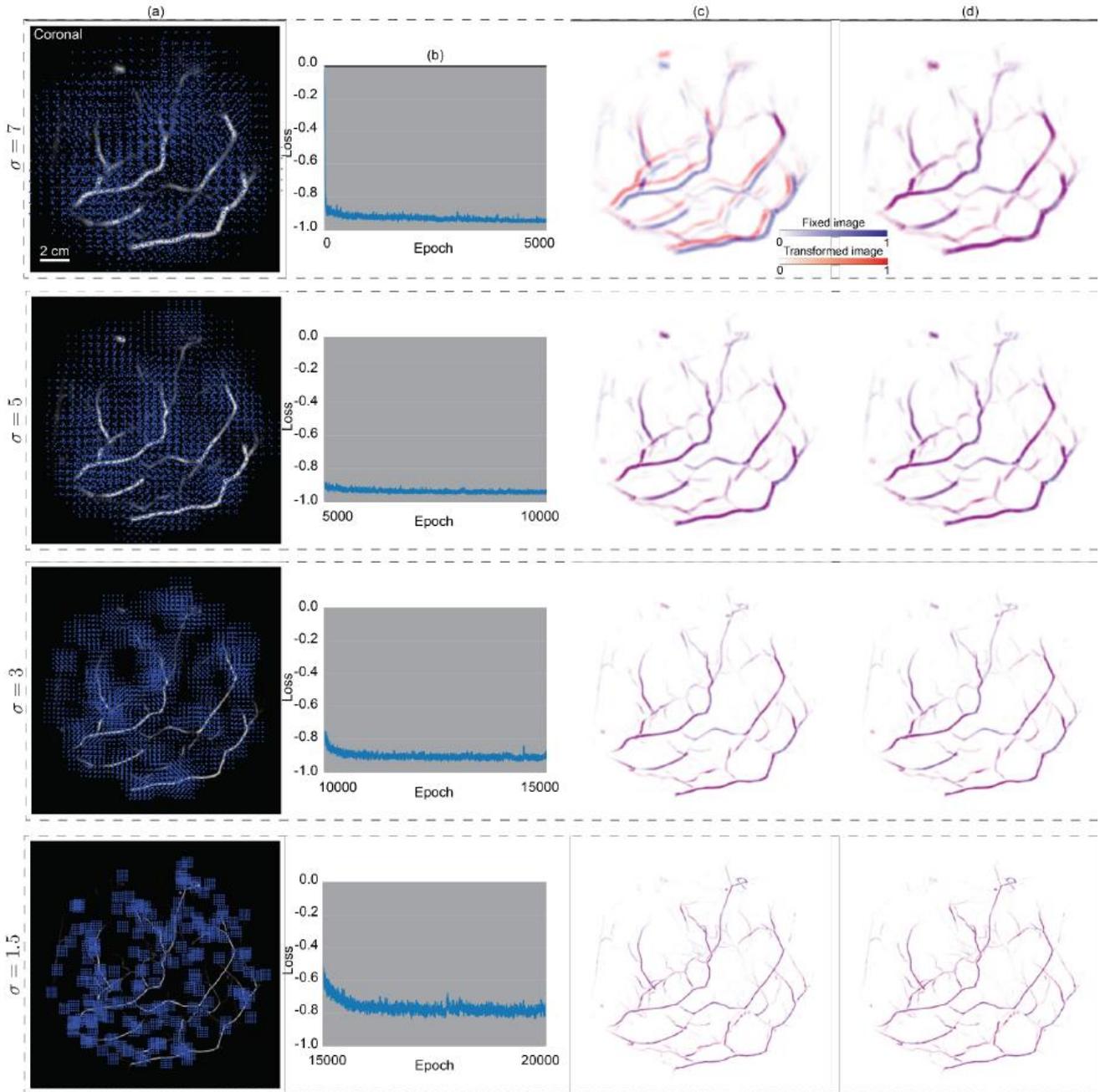
Figure 4: Visualization of the coarse-to-fine strategy on image registration of a photoacoustic image pair. From top to bottom the different phases are shown in order. (a) Point sampling in the fixed image domain; (b) Training loss curve; (c) MIPs of the overlay at the end of the current phase [50]

5.1.3 Implementation and training details

The image coordinates were normalized in the range of [-1, 1] for all three dimensions. A segmentation mask of vessels and the skin was obtained of the fixed image using the Frangi vesselness filter and adaptive thresholding, where the threshold map exponentially decayed as a function of depth. The mask was dilated using a cubic structuring element of 3 x 3 x 3 to include the vessel edges.

A random sampler randomly selected a batch of 200 points from the mask for each epoch. Every point in the sample mask has an equal chance to be selected, and a sample can be selected multiple times. A cubical patch of 5 x 5 x 5 was used for each coordinate from the random sampler to get the neighboring points as

well, leading to a total number of $200 \cdot 5^3 = 25000$ coordinates for each training epoch. When the scale of the Frangi changes, the side length of the patch was changed as well according to the scale of the Frangi filtering in the iteration, which was $2.5 \cdot \sigma/100$. This allows the network to compute normalized cross-correlation in smaller neighborhoods around the sampled coordinates during the optimization as the scale of Frangi decreases, see *Figure 4.*

During the training, the fixed and the moving images are evaluated at the input and transformed coordinates by trilinear interpolation. A trilinear interpolator returns a value, a weighted average of the surrounding voxels, with the distance to each voxel taken as weight.

Adam optimizer is implemented in MUVINN with a learning rate of 5e-5 with a learning rate decay of 5% at each point when the scale of Frangi decreased.

A different number of epochs were used for the synthetically deformed data set compared to the serial image data set due to increased complexity in distortions in the serial image data set. One training consists of 5000 epochs for the synthetically deformed data set and 20000 epochs for a serial image data set. All training and tests were done on a Windows 11 machine with an Intel Core i9-11900K @ 3.5GHz, 128 GB RAM and NVIDIA RTX3090 24GB.

5.2 Experimental evaluation of the framework with synthetically deformed data set

The main purpose of MUVINN is to align the PA images over time. It should be able to align images when repositioning errors occur. Before acquiring a data set of repositioned images with a volunteer, it is necessary to validate the algorithm with synthetically deformed images. By using synthetically deformed images, no new volunteers are needed. Synthetically deformed images are beneficial for preliminary tests because they are easily-made, cost-efficient, and safe. It makes it possible to experiment the algorithm in a more controlled manner and to optimize the algorithm before acquiring a new serial image data with a volunteer. However, the disadvantage is that it is unknown how realistic this deformation is because it is synthetic.

This chapter will describe how the synthetically deformed data set is created, evaluate how the framework works on the synthetically deformed data set, and compare its performance to another image registration framework, Elastix.

5.2.1 Photoacoustic imager

For the experiments conducted in this thesis, Twente Photoacoustic Mammoscope 3 (PAM3) is used as a photoacoustic imager.

Twente Photoacoustic Mammoscope 3 (PAM3) is the 3rd generation imager of the PAMMOTH project. With this imager, it is possible to acquire photoacoustic and speed of sound reconstructions. The device has high throughput, possesses no carcinogenic potential, uses no contrast agent, and causes no pain or discomfort to patients [52]. It has the shape of a bed on which the patient can lie down in a prone position. The patient can put her breast in a cup-shaped hole. This hole is the imaging tank with 512 transducers and two laser heads. These are tuneable lasers that consist of an optical parametric oscillator [53].

In general, five different wavelengths are used according to the protocol. These wavelengths are 720, 755, 797, 833, and 870 nm. Out of these five different wavelengths, 797 nm and 800 nm were used for the synthetically deformed image data, and 720 nm and 870 nm were used for serial PA image data.

Between the imaging tank and the breast, a 3D-printed transparent nylon cup supports the breast and improves the light distribution. There are eight different cup sizes for variable breast sizes. Water is used as a conductive medium for the US and PA imaging. The holes in the nylon cup allow water coupling [54]. The images of the imaging tank, breast-supporting cup, and the PAMMOTH imager can be seen in *Figure 5*.
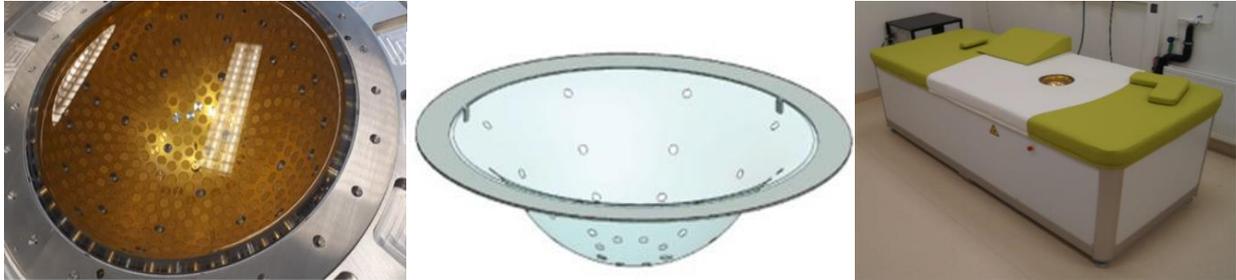
Figure 5: Imaging tank (Left), Nylon cup (Middle), Twente Photoacoustic Mammoscope 3 (Right) [54]

5.2.2   Synthetically deformed data set

Synthetically deformed images are generated by applying a random deformation field to the existing PA image. These images allow evaluation of the accuracy of image registration in a controlled scenario.

For generating a random deformation field, a random displacement can be applied to each voxel in x, y, and z directions, which forms a displacement field together. This displacement field can be combined with a rotation of the image. However, this displacement field can be rough as it is random, so smoothing the displacement field with a Gaussian filter is needed.

A toolkit called MONAI is used for creating synthetically deformed data. MONAI is an open-source deep-learning framework designed for medical imaging tasks. Rand3DElasticd function is one of its key functions, which can apply random 3D elastic deformations to images. By choosing the deformation's magnitude range and the rotation range, the user can control the intensity of the deformation and apply it to the image [55]. The definition of each parameter of the function is provided in *Table 1*.

Table 1: Parameter definitions of Rand3DElasticd function [55]

Parameter	Definition
sigma_range	Defines the range from which the standard deviation of the Gaussian kernel is sampled. This affects the smoothness of the displacement field.
maginitude_range	Defines the range for randomizing the magnitude of the elastic deformation. The two tuples are minimum and maximum values for the magnitude.
rotate_range	Controls the range of random rotation to be applied to the elastic deformation. Three tuples are given. Each tuple specifies the minimum and maximum rotation angle in degrees in order of x, y, and z-axis.
prob	Determines the probability of applying the elastic deformation.
mode	Specifies which interpolation is used to perform the elastic deformation.

Creating a synthetically deformed data set involves constructing synthetically generated deformed images in three different intensity levels: low, medium, and high. Within each intensity level, there are five randomly deformed images. The original, undistorted image serves as the fixed image, while the deformed images

serve as the moving images. The fixed image is considered as the ground truth for registration evaluation. This approach enables us to quantify the accuracy of image registration.

In this dataset, two PA images from two different volunteers acquired for previous studies were modified in three different intensity levels of deformation and range of rotation angle. Each level shares the same parameter configurations for the transformation. As the level increases, the deformation's intensity and the rotation angle range also increase, which results in a greater deformation. The parameter setting used for creating a random deformation for each level can be seen in *Table 2*. Each volunteer contributes to one fixed image and fifteen moving images, resulting in a total of thirty image pairs. *Table 3* presents an overview of all the image pairs within the synthetically deformed dataset.

Table 2: Parameter setting of Rand3DElasticd function for each level

Levels	Parameter settings
Low	monai.transforms.Rand3DElasticd(keys=['fixed'], sigma_range=(5,5), magnitude_range=(50,50), rotate_range=(np.pi/36, np.pi/36, np.pi/36), prob=1, mode=['bilinear'])
Medium	monai.transforms.Rand3DElasticd(keys=['fixed'], sigma_range=(5,5), magnitude_range=(100,100), rotate_range=(np.pi/18, np.pi/18, np.pi/18), prob=1, mode=['bilinear'])
High	monai.transforms.Rand3DElasticd(keys=['fixed'], sigma_range=(5,5), magnitude_range=(150,150), rotate_range=(np.pi*3/36, np.pi*3/36, np.pi*3/36), prob=1, mode=['bilinear'])

Table 3: List of images in the synthetically deformed data set

Distortion intensity level	Volunteer A	Volunteer B
Low	1a, 2a , 3a, 4a, 5a	1b, 2b, 3b, 4b, 5b
Medium	6a, 7a, 8a, 9a, 10a	6b, 7b, 8b, 9b, 10b
High	11a, 12a, 13a, 14a, 15a	11b, 12b, 13b, 14b, 15b

5.2.3 Results MUVINN

The same training parameters were used for all the image pairs for image registration. The details about the parameters used can be found in the *Implementation and training details*.

The image registration results are evaluated based on the visual inspection and image evaluation metrics.

The fixed and moving MIPs are displayed as overlay images before and after image registration for visualization. The displayed images are processed using Frangi vesselness filtering for visualizing vessels. The fixed image is blue, the moving image is red, and their overlap results in a shade of purple. One image pair is shown for each level of distortion. In total, three image pairs are shown.

Various image evaluation metrics, such as PSNR, SSIM, DSC, and NCC, are used for quantitative evaluation. Definitions of these metrics are explained in the *Evaluation metrics*. The absolute metric values pre- and post-registration can be seen in *Appendix C*. The computation time can be found in *Table 14* of *Appendix C*.

The structures exhibited excellent alignment for every image pair. The structures display consistency and accurate overlap, indicated by the purple shade in *Figure 6*.

Two structures were observed that were misaligned in the highest level of distortion images of Volunteer B, which can be seen in *Figure 7*.

Algorithm: MUVINN
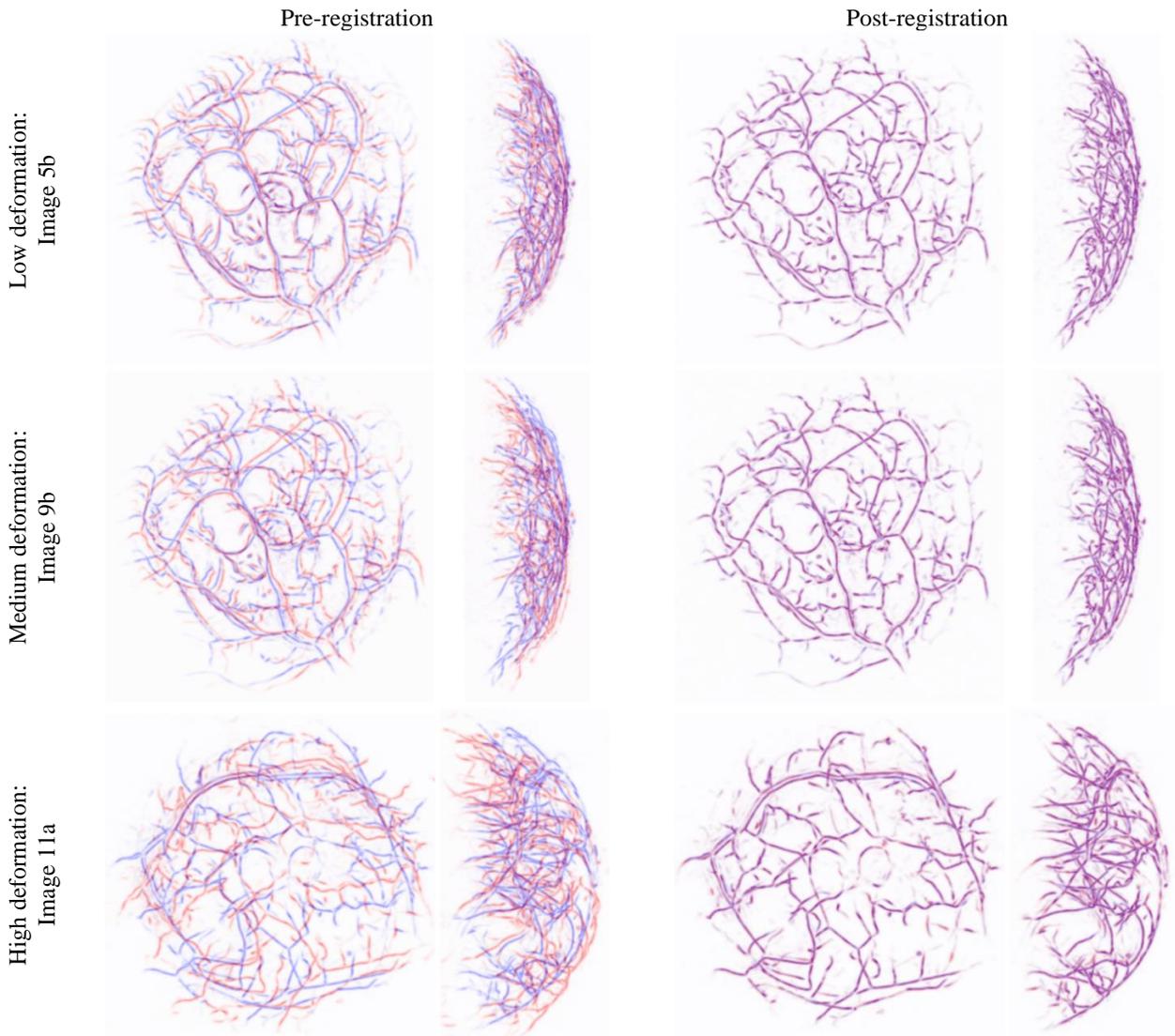
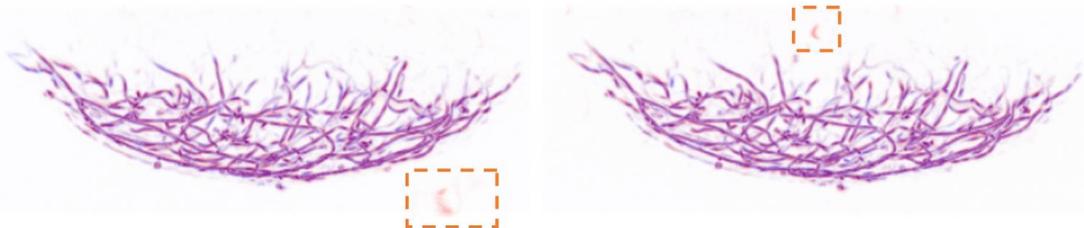*Figure 6: Coronal and sagittal MIPs of pre- and post-registration of synthetically deformed images*

Figure 7: Misalignments after registration

All analyzed image pairs have improved in all three evaluated metrics; see *Table 4*.

The highest mean post-registration PSNR was found in the groups with the lowest distortion intensity, and the PSNR decreased when the distortion level increased. However, the largest mean percentage change was

not in the lowest intensity groups. All image pairs have shown an increase in post-image registration PSNR, indicating that the registered image has an improved similarity compared to the unregistered image. It suggests that there is less voxel-wise difference between the registered image and the reference image.

The mean percentage change in SSIM increased when the distortion intensity level was increased except between the medium and high levels in volunteer B. However, the absolute values showed that the mean post-registration SSIM decreased when the distortion intensity level was increased. The increase in SSIM after the image registration indicates that the registered image has a higher structural similarity to the reference image than the unregistered image. It suggests that registration with MUVINN increases the alignment of structural information and patterns.

There were large changes in initial DSC levels between the different intensity levels of the distortion. The higher the distortion level, the lower the initial DSC values. The increase in DSC shows that the segmented regions of the reference image and the moving image have a higher overlap after the image registration. Generally, the percentage change in DSC increased with escalating distortion levels, reflecting the reduced overlap between segmentations in the unregistered and reference images.

The percentage change in metrics usually increased when the distortion level was increased. However, looking at the absolute values of metrics in *Appendix C*, mean post-registration metric values decreased when the distortion level was increased. Thus, a larger increase in the metrics in the images of higher distortion levels does not mean that these images were better aligned than those from lower distortion intensity levels. The absolute values provided more information than the percentage difference between the pre- and post-registration values of the evaluation metrics. For example, two structures were misaligned in the highest level of distortion for volunteer B, which were in image pair 12b and 13b, shown in *Figure 6*. The percentage change of metrics showed that these two image pairs had the largest increase in metrics among their group. However, the absolute values showed that these image pairs had the worst absolute evaluation metric values among their group after registration. The large difference in evaluation metrics between the pre- and post-registration can be explained by the low pre-registration values, which also explains that the registration was more challenging compared to other pairs.

Table 4: The percentage change of PSNR, SSIM, and DSC before and after the image registration Acronyms: PSNR = peak signal-to-noise ratio, DSC= Dice similarity coefficient

	Volunteer A				Volunteer B			
	Image pairs	ΔPSNR (%)	ΔSSIM (%)	ΔDSC (%)	Image pairs	ΔPSNR (%)	ΔSSIM (%)	ΔDSC (%)
Low deformation	1a	8.23	32.62	923.09	1b	8.52	32.40	937.45
	2a	9.95	46.28	2595.27	2b	8.28	32.11	673.77
	3a	8.34	32.88	630.58	3b	9.71	40.25	2159.20
	4a	9.72	44.11	2123.77	4b	11.82	57.36	2891.96
	5a	8.86	39.37	3087.45	5b	9.24	39.35	2039.06
	Mean	9.02	39.05	1872.03	Mean	9.51	40.29	1740.29
	Standard deviation	0.70	5.61	949.25	Standard deviation	1.26	9.18	821.35
Medium deformation	6a	8.51	38.25	5078.21	6b	11.37	57.32	1211.52
	7a	9.83	47.99	4941.41	7b	10.12	49.64	4230.04
	8a	8.83	41.12	5208.39	8b	9.75	48.36	5579.42
	9a	8.22	36.39	4401.94	9b	11.65	61.94	2931.40
	10a	9.72	48.10	9422.72	10b	10.03	48.77	2385.43
	Mean	9.02	42.37	5810.53	Mean	10.58	53.21	3267.56
	Standard deviation	0.65	4.87	1826.82	Standard deviation	0.77	5.46	1509.27

Algorithm: MUVINN

High deformation	11a	8.73	46.23	13901.46	11b	8.35	40.55	9950.23
	12a	9.67	50.67	5654.30	12b	11.31	62.15	10152.39
	13a	7.75	36.78	2057.23	13b	10.30	53.67	14815.37
	14a	9.36	49.77	11967.41	14b	9.90	49.14	4488.26
	15a	9.68	48.84	6714.32	15b	8.77	42.47	5310.47
	Mean	9.04	46.46	8058.94	Mean	9.73	49.60	8943.34
	Standard deviation	0.73	5.06	4313.28	Standard deviation	1.07	7.84	3741.71

5.2.4 Results Elastix

Elastix was used to compare the performance of MUVINN with another automatic image registration method.

The input images were pre-processed with the Frangi vesselness filter to make the results comparable to MUVINN. The same initial parameters were used for all the image pairs. For Elastix, two transformation models are used to correct for both linear and non-linear deformations, first rigid transformation and subsequently B-spline transformation model. A multi-resolution scheme with four pyramid levels was used. Each pyramid level is a down-sampled version of the original image, in which the first level is the coarsest image. Then, each following level represents the image in a better resolution. The number of iterations for each pyramid was equal to 500 for rigid and equal to 1000 for B-spline transformations. Finally, for the similarity term of the loss function, the AdvancedNormalizedCorrelation and the TransformRigidityPenalty were used as regularization. Details about the parameters can be seen in *Appendix B*.

For visual inspection, the fixed and moving MIPs are displayed as overlay images before and after the image registration. The fixed image is displayed in blue, the moving image is displayed in red, and their overlap results in a shade of purple. One image pair is shown for each level of distortion, which is the same image pair as shown in *Results MUVINN*. The same image evaluation metrics are used as in the previous experiment. The absolute values pre- and post-registration can be found in *Appendix C*.

All the image pairs showed a better alignment after registration, but almost every image pair had several visible misalignments, see *Figure 8*. In the lowest distortion level, some vessels were even more misaligned compared to pre-registration. Although there were many misalignments, no visible artifacts were detected.

Pre-registration   Post-registration

Low deformation: Image 5b
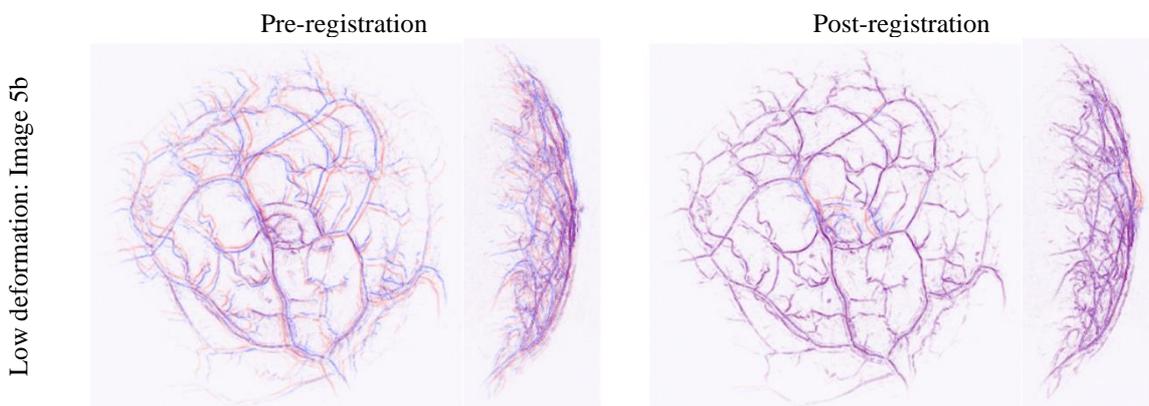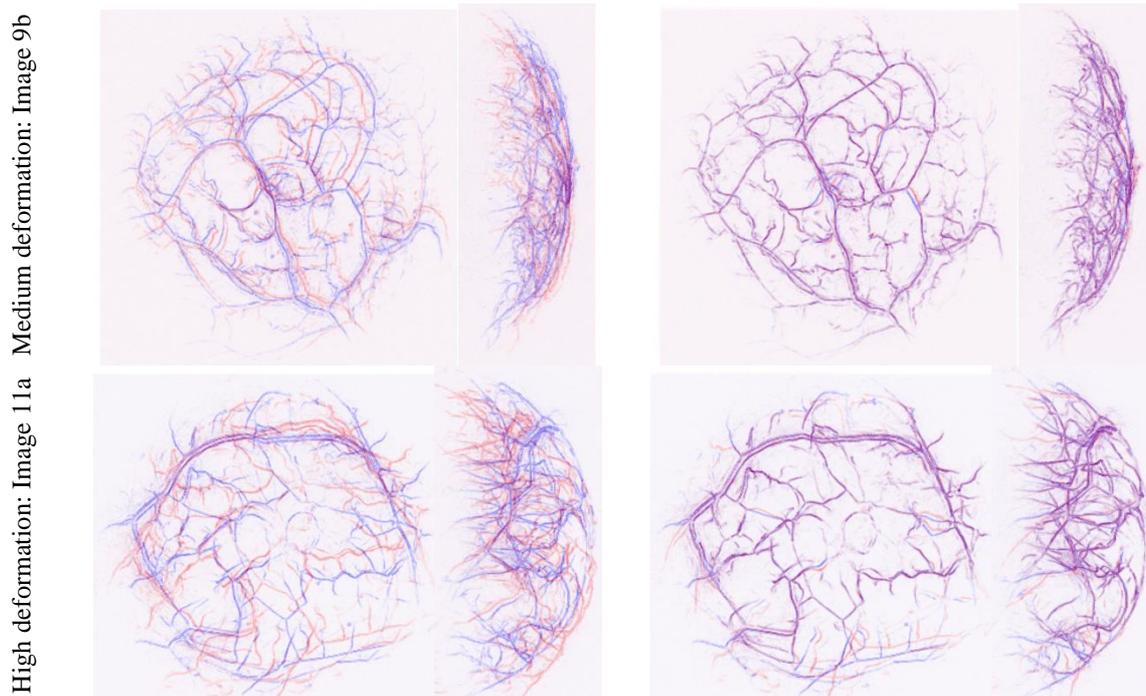
Figure 8: Coronal and sagittal MIPs of pre- and post-registration of synthetically deformed images

All analyzed image pairs have shown improvements in all three evaluated metrics. The percentage change of each metric can be seen in *Table 5*.

The results showed that PSNR and SSIM increased less when the level of distortion was increased, except between the medium and the highest distortion of Volunteer B for SSIM.

For DSC, it was the other way around. The higher the level of distortion, the larger the percentage increase of DSC. This was due to the low initial DSC values in images of higher-order distortion. However, the post-registration DSC values were still lowest in the image pairs of the highest order of distortion.

Additionally, within the same level of distortions, variations in metrics were quite significant between the two volunteers after the registration. Elastix showed large standard deviation values in post-registration metric values of each level of distortion compared to MUVINN. These values can be found in *Table 6* in *Comparison between MUVINN and Elastix*.

Table 5: The percentage change of PSNR, SSIM, and DSC before and after the image registration

	Volunteer A				Volunteer B			
	Image pairs	ΔPSNR (%)	ΔSSIM (%)	ΔDSC (%)	Image pairs	ΔPSNR (%)	ΔSSIM (%)	ΔDSC (%)
Low deformation	1a	5.46	24.77	616.76	1b	7.33	31.06	832.12
	2a	7.04	37.72	2157.20	2b	6.40	30.22	510.96
	3a	6.07	27.04	895.28	3b	6.91	35.43	1718.21
	4a	8.05	40.48	2124.36	4b	7.90	46.87	2847.53
	5a	8.27	38.90	3124.99	5b	7.07	36.06	1631.53
	Mean	6.98	33.78	1783.72	Mean	7.12	35.93	1508.07
	Standard deviation	1.09	6.53	917.13	Standard deviation	0.49	5.93	812.98

Medium deformation	6a	6.35	31.64	4105.00	6b	6.02	39.60	868.81
	7a	6.58	37.50	4004.54	7b	6.87	39.93	3722.21
	8a	4.84	27.56	2905.63	8b	7.05	40.13	5355.68
	9a	6.13	30.27	3499.02	9b	5.99	42.89	2014.36
	10a	7.01	39.52	8469.52	10b	6.15	37.76	4116.53
	Mean	6.18	33.30	4596.74	Mean	6.42	40.06	3215.52
	Standard deviation	0.73	4.50	1982.71	Standard deviation	0.45	1.65	1587.37
High deformation	11a	4.49	27.64	8547.96	11b	5.00	29.33	6741.79
	12a	4.80	30.52	3409.75	12b	5.87	40.24	7646.81
	13a	4.79	25.49	3560.28	13b	5.51	35.29	11190.56
	14a	4.73	30.70	7213.75	14b	4.95	32.52	2614.17
	15a	5.42	33.04	4757.27	15b	4.76	29.02	3457.02
	Mean	4.84	29.48	5497.80	Mean	5.22	33.28	6330.07
	Standard deviation	0.31	2.63	2045.01	Standard deviation	0.41	4.17	3085.13

5.2.5   Comparison between MUVINN and Elastix

Overall, every image pair exhibits improved alignment after registration for both MUVINN and Elastix. However, there are notable misalignments in almost all image pairs for Elastix. Within these misaligned regions, identical vessel structures are usually in close proximity, indicating that they are the same structure, yet fail to achieve perfect overlap compared to post-registration images using MUVINN. MUVINN had only two image pairs in the highest order of distortion, exhibiting one misalignment.

Both frameworks also demonstrated improved evaluation metrics following the registration. However, the improvements achieved by MUVINN are notably higher in all metrics.

Interestingly, the evaluation metrics reveal a considerable discrepancy between the two volunteers after registration in Elastix and MUVINN. Looking at the standard deviation within the same level of distortion, Elastix showed larger variance in all levels, except for DSC in the highest intensity of distortion. The standard deviation values of each distortion intensity level can be found in *Table 6*.

A higher standard deviation implies more significant variability or inconsistency in the corresponding metric. The higher standard deviation of PSNR and SSIM can indicate inconsistent image similarity based on the intensity between the image pairs. A high standard deviation of DSC suggests that the quality of structure alignment varies among the different image pairs. This indicates that the performance of Elastix is more sensitive to variations in vessel structures during the image registration process than MUVINN, leading to inconsistent registration quality and reduced reliability.

Table 6: Standard deviation of post-registration PSNR, SSIM, and DSC of both volunteers

Intensity of distortion	Framework	PSNR	SSIM	DSC
Low	MUVINN	0.49316	0.019173	0.020354
	Elastix	0.61178	0.028253	0.062242
Medium	MUVINN	0.28324	0.011063	0.029097
	Elastix	0.55309	0.027672	0.058283
High	MUVINN	0.40067	0.015419	0.033706
	Elastix	0.50250	0.027093	0.018640

In terms of computational efficiency, it is hard to compare because different devices were used for computations using each framework. For MUVINN, a Windows 11 machine with an Intel Core i9-11900K @3.5GHz, 128 GB RAM and NVIDIA RTX3090 24GB was used. For Elastix, Google Colab's CPU was used. The default CPU for Colab is an Intel Xeon CPU @2.50GHz and 13GB of RAM [56].

Elastix exhibited an average computation time of 11 minutes and 24 seconds, while MUVINN exhibited an average computation time of 5 minutes and 22 seconds. However, these values cannot be compared to each other without knowing the amount of impact using a different device can cause. The computation time of both frameworks can be seen in *Table 14* of *Appendix C.*

5.2.6 Discussion & conclusion of experiment with synthetically deformed data

Using synthetically deformed images has certain advantages, such as enabling precise control over distortions, providing the ground truth, generating images easily, and eliminating the need for volunteers. However, several limitations make it unsuitable as a validation method of the algorithm in real use.

These synthetically deformed images do not represent changes in images over time, as they do not fully capture the complexity and variability in real data. These synthetic deformations cannot account for variations in human anatomy, tissue properties, and changes in field-of-view. It is a simplified presentation of a deformation. Although they can be used as an initial testing ground to validate the algorithm before human trials, as how they are used in this research, their results will not be sufficient to validate the algorithm for real-world use.

Some improvements can be made to the experiments with serial image data. It is better to use the same device for computations so the computational efficiency of different frameworks can be compared to each other. Due to the limitation of available hardware, a different device was used for Elastix compared to MUVINN.

This experiment uses multiple metrics to evaluate the image registration result. Three categories of image evaluation metrics are used: intensity-based, perception-based, and overlap-based. For each metric, the absolute values and percentage difference between pre- and post-registration are used for evaluation.

Combining these three categories of image evaluation metrics provided a more comprehensive insight into image registration results. By combining SSIM, PSNR, and DSC, information about image similarity based on intensity and spatial accuracy could be evaluated. However, TRE would add much value to spatial accuracy analysis as it provides the Euclidean distance between the same landmarks. In this experiment, no notable changes were detected in image quality according to the visualization. No artifacts or distortions were detected after registration.

In conclusion, the visual inspection suggests that the image registration for PA images is successfully done using MUVINN. Vessel structures in the registered image closely match those in the fixed image.

The evaluation metrics collectively indicate that the registered image shows a good alignment considering various aspects by showing a remarkable increase in all three evaluation metrics. The registered image demonstrates close similarity to the reference image, confirming that the image registration for PA images has been successful.

The visual inspection of Elastix suggests it can align the vessels, but the accuracy is visibly lower than MUVINN. Elastix exhibits enhanced evaluation metric values, although notably less impressive than those achieved by MUVINN. Additionally, it demonstrates a notable variance between different volunteers.

A high-quality alignment is required for a comprehensive analysis over time. MUVINN would be a robust solution for vascular image registration. However, this holds true only for synthetically deformed images for this experiment, as these do not represent changes in vasculature over time and do not fully capture the complexity and variability in real data.

5.3   Chapter discussion & conclusion

Previously, quantitative analysis of 3D PA imaging in longitudinal settings was suboptimal due to the need for an image registration framework which can face challenges mentioned in the *Challenges of vascular image registration*. An accurate image registration framework is needed to correct repositioning deformations for comparison between images over time to detect subtle biological changes.

For this issue, a novel image registration MUVINN was developed that uses coordinate-based neural networks and multiscale Frangi vesselness filtering to align longitudinal, three-dimensional PA images. By using a coordinate-based neural network, the framework can represent deformation fields in a continuous domain. This approach allows the deformation field to be calculated at any coordinate in the spatial domain.

MUVINN uses similarity metrics tailored to the vascular image registration task. It uses a similarity term based on vesselness features by incorporating the Frangi vesselness filter. By doing so, only relevant features are considered, and potential noise is neglected. This feature is combined with NCC, which allows the method to be robust to image intensity variations. Like Elastix, a coarse-to-fine strategy is implemented in this framework. Decomposing the image on different scales significantly improved the accuracy of the registration.

Preliminary experiments with synthetically deformed images were conducted to investigate the framework's effectiveness. MUVINN successfully registered all image pairs accurately, which was visually and quantitatively assessed in this chapter. However, it is important to note that these synthetical deformations do not represent the repositioning deformations which occur in real-world scenarios. Thus, further research is needed to validate the framework.

Comparing the results of preliminary experiments for MUVINN and Elastix, MUVINN outperformed Elastix in both visual and quantitative assessments. Both MUVINN and Elastix share the common approach of implementing a multi-resolution strategy during the image registration process. However, there are several drawbacks to using Elastix compared to MUVINN.

Firstly, the input images of Elastix require pre-processing using the Frangi vesselness filter, while MUVINN does not require any pre-processing of the image due to the incorporation of the Frangi vesselness filter during the training. Secondly, the user needs to specify the transformation model in advance for Elastix, while MUVINN does not require any information about the transformation model. The specification of the transformation model may clarify the weak performance of Elastix compared to MUVINN. The parameterization of the transformation model may not be suitable to correct the deformation that occurs in the breast after repositioning. These drawbacks decrease the performance of Elastix and make it less user-friendly compared to MUVINN.

6 Experimental evaluation of the framework with serial image data set

The performance of the algorithm has been assessed using synthetically deformed images in the prior experiment. However, there remains uncertainty about the algorithm's ability to accurately align images over time when there is a real repositioning error. A major drawback of synthetically deformed images lies in their lack of realism compared to a real repositioning error.

A different type of data set needs to be obtained in which real repositioning errors occur to evaluate the algorithm's applicability in the real world. This approach can validate the algorithm using images that contain real repositioning errors, a challenge that synthetically deformed images fail to evaluate.

Initially, it is necessary to construct a data set for assessing the algorithm's performance, which raises the primary research question of this chapter:

- What are the appropriate validation methodologies to assess its accuracy?

This research question results in the following sub-questions:

• What are the inclusion criteria for the volunteer?

• Which experiments are needed to evaluate the performance of the algorithm, and how should the measurements be done to acquire adequate images for these experiments?

When the data set is constructed, the secondary research question to evaluate the algorithm's performance is formulated as follows:

- How does MUVINN perform in a realistic scenario and a challenging scenario?

This results in the following sub-questions:

• How well does the algorithm work when there is a realistic repositioning error?

• How well does the algorithm work when a different wavelength is used?

• How well does the algorithm work when the volunteer is mispositioned?

• How well does the algorithm work when an incorrect cup size is used?

6.1 Volunteer criteria

To establish a data set for algorithm validation, it is crucial to construct a representative data set that closely mirrors a typical data set with a breast cancer patient. This entails finding a volunteer who possesses characteristics similar to the demographic profile often associated with patients diagnosed with breast cancer.

According to the American Cancer Society, breast cancer mainly occurs in middle-aged and older women. The median age at the time of breast cancer diagnosis is 62. Factors such as menopause and obesity can increase the chance of breast cancer, according to Dunneram et al. [57]. Considerations will be given to skin tone to occur to prevent potential artifacts in PA image, which can occur due to certain characteristics.

A single volunteer is selected to construct a data set for this experiment. A 59-year-old woman who was classified with Fitzpatrick skin type 2 and wore brassiere size 80E was selected as the volunteer. The volunteer was informed about the study, and informed consent was obtained.

6.2   Types of measurement

Three types of measurements were performed to evaluate the algorithm:

1.   Repeated measurement of the volunteer after correctly repositioning

2.   Repeated measurement of the volunteer after mispositioning

3.   Repeated measurement of the volunteer with a different cup size

These measurements will be done using Twente Photoacoustic Mammoscope 3 (PAM3) imager as described in *Photoacoustic imager*. Four experiments will be done with these three types of measurements. With these experiments, the algorithm's performance in image registration can be evaluated across various scenarios.

6.2.1   Measurement 1: Correct repositioning

In the first measurement, images with realistic repositioning errors will be acquired. For each image, the volunteer must leave the imaging system before the repositioning so she is fully repositioned. The operator will correctly position the volunteer, so the repositioning error is solely due to the breast repositioning. Five images of the volunteer will be acquired for the first measurement.

Two wavelengths are used to see if wavelength will influence image registration. Three of the five images will be acquired with two wavelengths and the other two with one wavelength. The two wavelengths are 720 nm and 870 nm. The intensities of the vessels will change when a different wavelength is used due to changes in laser energy output and the absorption coefficient of the skin, oxygenated and de-oxygenated vessels at the two wavelengths. Experiments 1 and 2 will be done using the images from measurement 1.

6.2.2   Measurement 2: Mispositioning

In measurement 2, images with a large magnitude of deformation will be acquired. The operator will intentionally misposition the volunteer by placing the breast of the volunteer not precisely at the center of the cup but slightly more downwards. Mispositioning will shift the field-of-view, resulting in the absence of certain vessels between the images. The vessels in the bottom will be absent in the mispositioned images due to the downward shift of the breast, which results in more challenging images for the algorithm. The *Operator's protocol* in the *Appendix D* explains how the breast is mispositioned. Experiment 3 will be done using images from measurement 2.

6.2.3   Measurement 3: Different cup sizes

In measurement 3, images of different cup sizes will be acquired. The same breast will be imaged in three cup sizes different from the optimal cup size.

In real-world scenarios, the operator selects the cup size based on the breast size before acquisition. There are situations when the operator selects the wrong cup size, resulting in lower-quality images. However, it is unlikely that a cup of a different size will be used over time for repeated measurements. The imaged breast volume correlates with the chosen cup size, causing a shift in the field-of-view. If the cup size is too small, the breast gets compressed within the cup, leaving the parts close to the chest wall out because they will not fit inside the cup. Conversely, when the cup size is too large, the breast will be elongated and not fit properly, resulting in gaps between the breast and the cup. The breast will not be stabilized in the cup, resulting in larger motions. The maximum depth and the breast volume inside each cup are listed in *Table 7*. The geometrical curves describing the eight cup sizes are shown in *Figure 9*.

The images with smaller cup sizes will show what happens with the image when the field-of-view decreases in a controlled manner. The images with larger cup size will show what happens with the image when the breast is slightly hanging in the cup. The image without a cup will show what happens when the breast is not fixed at all. The quality of the image will decrease due to uneven illumination, noise and motion artifacts.

In total, eight images will be acquired: Two images of the optimal breast-supporting cup size, two images of the two sizes smaller cup, two images of the three sizes smaller cup, two images of the one size larger cup, and one image without a cup.

Table 7: Maximum depth and submerged volume [54]

Size	Max depth (mm)	Submerged volume (mL)
4	52	496
5	58	599
6	64	717
7	71	856
8	71	1013
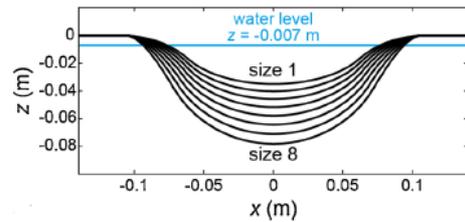
Figure 9: Geometrical curves describing the eight cup sizes [54]

6.3   Types of experiments

Four types of experiments were conducted using the measurements acquired as described in the subchapter *Types of measurement*.

1.   Aligning images with repositioning errors of a correctly repositioned breast

2.   Aligning images with repositioning errors of a mispositioned breast

3.   Aligning images with different illumination

4.   Aligning images with different in field-of-view

6.3.1   Experiment 1: Aligning images with repositioning errors

In experiment 1, correctly repositioned volunteer images of the same wavelength will be aligned. This experiment aims to assess the algorithm's ability to align real repositioning errors.

This experiment has already been conducted during this study with two existing repeated measurements. The outcome of this previous experiment showed that the algorithm successfully aligned vessels that were present in both images while leaving the vessels visible in only one of the images. Due to a small and limited data set, validating the algorithm was not possible. Hence, a new data set was needed. With the new data set, the algorithm is expected to work the same way as in the previous experiment with the smaller data set.

In this experiment, one of the images from the first measurement will be the fixed image, and the other four will function as the moving images, forming four image pairs. Sub-question 3 can be answered with this experiment.

6.3.2   Experiment 2: Aligning images with different illumination

In experiment 2, images at 870 nm will be aligned with different images at 720 nm. This experiment aims to assess the algorithm's ability to align images when there is a difference in illumination wavelength.

Besides the repositioning error, there will be a change in the intensity of the vessels between the two images due to a change in laser energy output and absorption coefficient. Laser wavelength of 720 nm can reach a higher depth than a larger wavelength due to higher energy. The three images acquired with two wavelengths from the first measurement will be aligned in this experiment.

One of the three images will be the fixed image, and the other two will be the moving image, which will form two image pairs.

In this experiment, image pairs of two different wavelengths will be used to compute the displacement field. However, the moving images of the same wavelength will be transformed using the acquired displacement field to enable accurate comparison between registered images. See *Figure 10* for a schematic diagram.

This experiment aims to find out whether the algorithm can match the correct points and align the images when there is an optical fluence difference between the images. If the transformation is successful, it indicates that the same points are matched to each other even when there is a difference in optical fluence due to the wavelength difference. Sub-question 4 can be answered with this experiment.
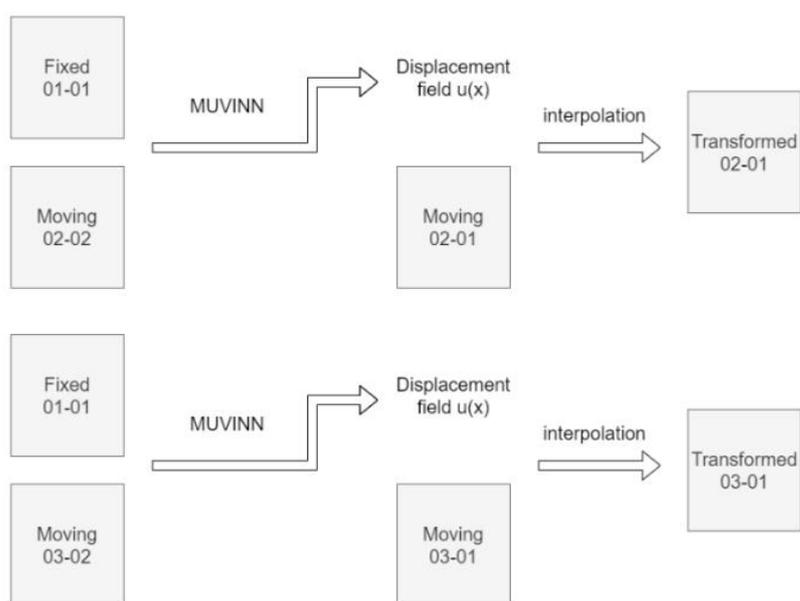
Figure 10: Schematic diagram of Experiment 2

6.3.3   Experiment 3: Aligning images with mispositioning errors

In experiment 3, the algorithm will be challenged by aligning mispositioned images to correctly positioned images. The objective is to assess if the algorithm can minimize the operator's dependency by aligning the mispositioned breast to the correctly positioned breast. The expectation is that the algorithm will align vessels visible in both images and keep the vessels not visible in both images.

In this experiment, one image from the first measurement will be used as the fixed image, and the two images from the second measurement will function as the moving images. Sub-question 5 can be answered with this measurement.

6.3.4   Experiment 4: Aligning images with a shift in field-of-view

In experiment 4, adding a change in the field-of-view to the repositioning error will challenge the algorithm. The objective is to assess whether the cup size error influences the image registration outcome.

This experiment comprises three different parts. The first part involves aligning images of two cup sizes smaller than the optimal size. In the second part, the algorithm aligns images of three cup sizes smaller. The final part focuses on aligning images of one cup size larger.

When there is a cup size difference of three, almost half of the peripheral volume will not be imaged, which is challenging for image registration. The algorithm is expected to align the vessels only visible in both images. This experiment will show what the algorithm does to the vessels in the moving image which are not visible in the fixed image.

In this experiment, two images of the optimal cup size will be used as the fixed image, and the images of wrong cup sizes will function as the moving images. Sub-question 6 can be answered with this measurement.

6.4   Summary of acquired data

Sixteen images are acquired in total, from which three are acquired in two different wavelengths. The image name contains two terms. The first term is the acquisition order, and the second determines the wavelength. 01 in the second term means that the image is acquired with a wavelength of 720 nm, and 02 indicates that the image is acquired with a wavelength of 870 nm. *Table 8* lists all acquired images for each experiment. Due to some errors, image 10 could not be reconstructed, so this image was not available. *Figure 11* summarizes all the experiments and measurements mentioned above.

Table 8: List of acquired images for Serial image data set

Image name	Experiment	Wavelength (nm)	Breast side	Description
01-01 02-01 03-01 04-01 05-01	1	720	L	Correctly repositioning
01-02 02-02 03-02	2	870		Correctly repositioning
06-01 07-01	3	720		Mispositioning
08-01 09-01	4		R	Correctly repositioning in an optimal cup size (cup size = 7)
10-01 11-01				Correctly repositioning in a two size smaller cup (cup size = 5)
12-01 13-01				Correctly repositioning in a three size smaller cup (cup size = 4)
14-01 15-01				Correctly repositioning in a one size larger cup (cup size = 8)
16-01				Positioning without a cup
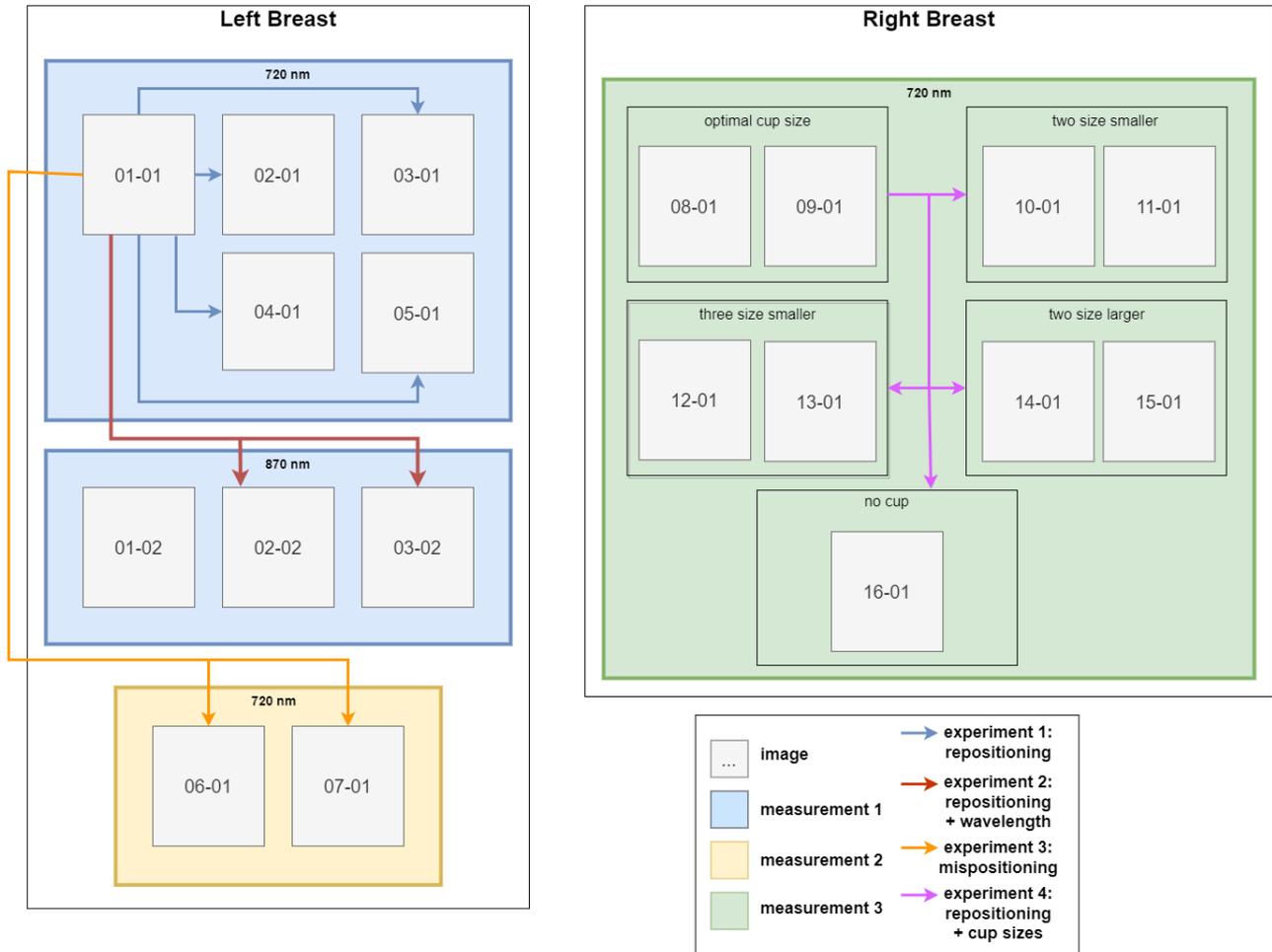
Figure 11: Schematic summary of the volunteer experiments

6.4.1 Safety specifications

Following maximum exposure duration regulations is vital to ensure the volunteer's safety. The first two measurements are done with the left breast. The third measurement is done with the right breast.

One measurement takes 2 minutes when one wavelength is used and 2 minutes and 40 seconds when two wavelengths are used. Sixteen images were acquired in total. Seven images of the left breast were taken. From these seven images, three images have two different wavelengths and four images have one wavelength. Nine images are taken from the right breast. All of them are acquired with one wavelength. This results in 16 minutes of exposure for each breast. This follows the safety measures, which allow 20 minutes of exposure to each breast. A detailed schedule can be seen in the *Operator's protocol* in *Appendix D*.

6.5 Results

Eighteen image pairs were included in total, divided into four different experiments. All image pairs used the same initial parameters. The image registration results are evaluated based on the visual inspection and image evaluation metrics.

For visual inspection, the fixed and moving MIPs are displayed as overlay images for pre- and post-image registration. For the MIPs, images are processed using Frangi vesselness filtering to enhance the vessel structures. The fixed image is displayed in blue, the moving image is displayed in red, and their overlap

results in a shade of purple. One image pair is shown for each experiment, and in case the experiment is divided into subsections, one is shown for each subsection. In total, six image pairs are shown in the results.

Various image evaluation metrics, such as PSNR, SSIM, and DSC, are used for quantitative evaluation, like in *Results MUVINN*. Additionally, NCC and TRE are also calculated for volunteer measurement. The landmarks chosen to compute TRE were the branching points of the vessels. The matching voxels were manually picked using an open-source software called Mevislab. Definitions of these metrics are explained in *Evaluation metrics*. The pre- and post-registration evaluation metrics values of all image pairs can be found in *Appendix E*.

The correctly repositioned images of the same wavelength were aligned in the first experiment. The structures exhibited excellent alignment for all image pairs, as shown in *Figure 12*. One small misalignment was observed in the first image pair, as shown in *Figure 17a*. No significant inconsistencies or distortions were observable in the transformed image.

TRE analysis of the first experiment showed an initial maximum mean displacement of 10.97 ± 3.78 mm, which decreased to a maximum mean displacement of 0.89 ± 0.48 mm post-registration. This means that the mean displacement error is less than two voxels, as the voxel size is 0.4^3 mm.
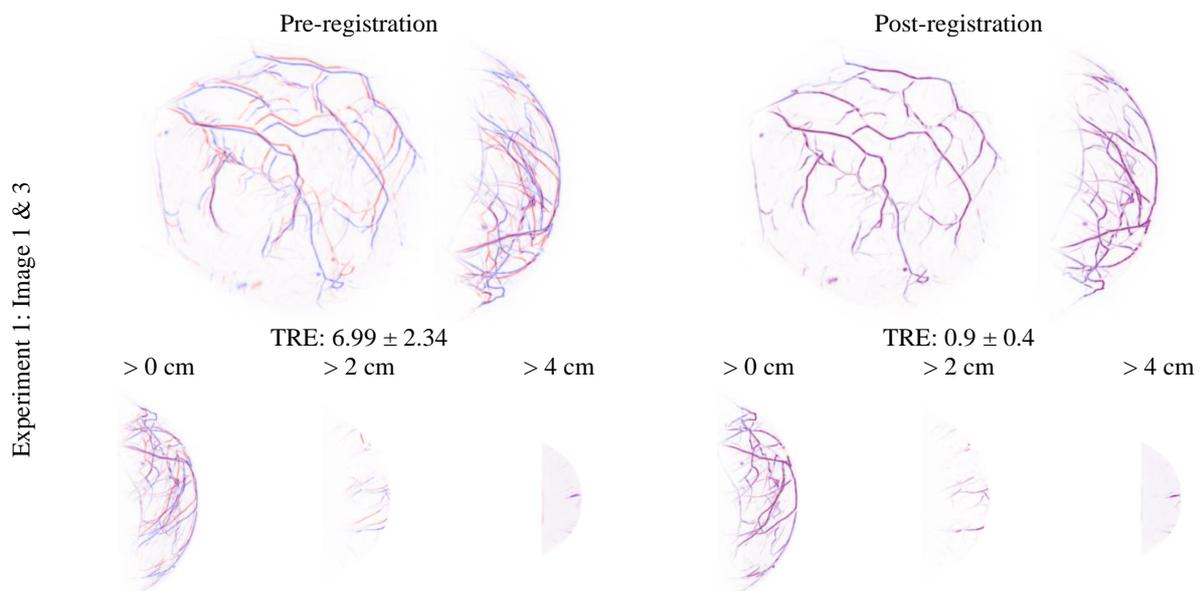
Figure 12: MIPs and depth images of before and after image registration of Experiment 1

Correctly repositioned images with two different wavelengths were aligned in the second experiment. Varying the wavelengths led to different intensities between the images due to different absorption of tissue between the two wavelengths, which can be seen in *Figure 13*. The structures exhibited excellent alignment, but the first image pair had the same misalignment as the previous experiment, as shown in *Figure 17a*.

The initial maximum mean displacement was 10.97 ± 3.78 mm, which decreased to 0.87 ± 0.49 mm post-registration. This post-registration displacement was very similar to the value of the same image pair from experiment 1, which was 0.89 ± 0.49 mm. The other image pair also showed a post-registration displacement value similar to experiment 1.

Experimental evaluation of the framework with serial image data set

Fixed image (λ= 720nm)

Moving image (λ= 860 nm)

Experiment 2: Image 1 & 3

Pre-registration

TRE: 6.99 ± 2.34

Post-registration

TRE: 0.94 ± 0.45
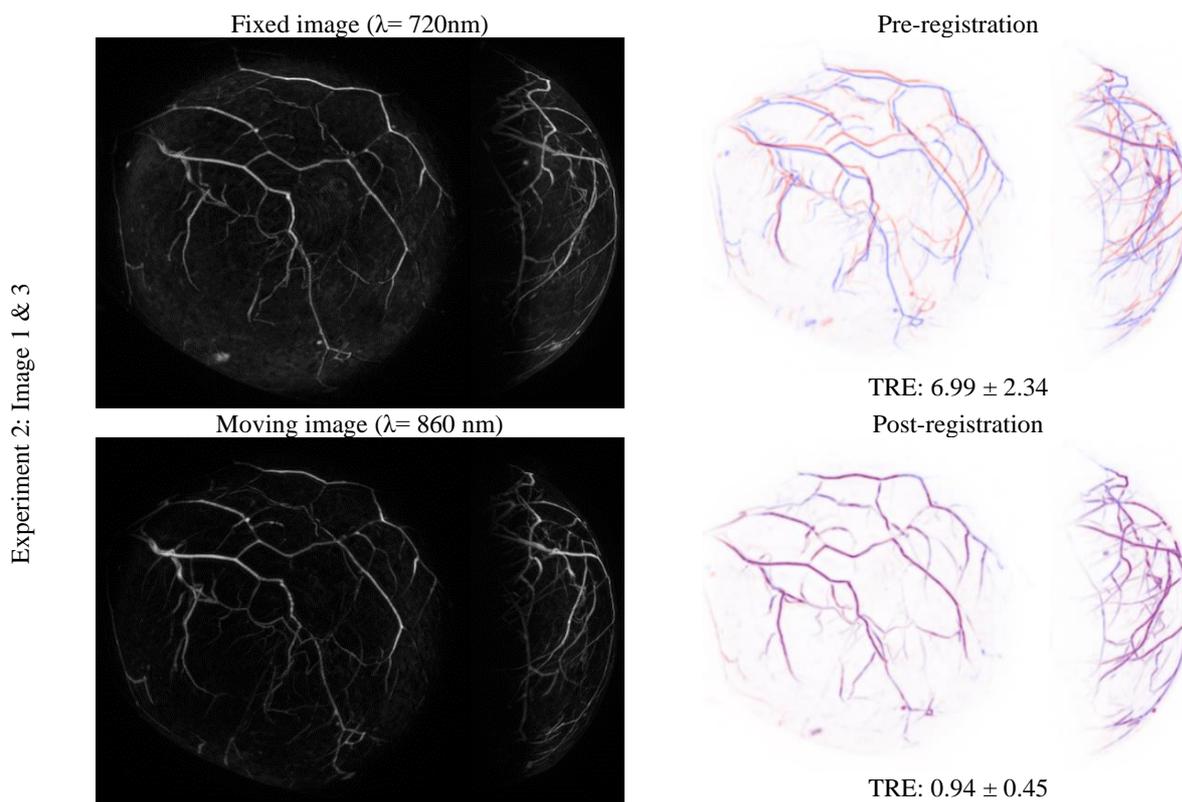
Figure 13: MIPs of Fixed and Moving images without any pre-processing (Left), MIPs before and after image registration of Experiment 2 (Right)

The volunteer was intentionally mispositioned in the third experiment to create a more challenging scenario. The largest initial mean displacement value was found in experiment 3, indicating how challenging this experiment is compared to all other experiments. The initial maximum mean displacement was 33.31 ± 5.16 mm, more than three times larger than the value of the correctly repositioned image pair of the first measurement.

The mispositioning led to changes in the field-of-view, which was visible as some vessels were only present in one of the images. The bottom part of the breast was not imaged in the moving image. Common vessels were aligned, and uncommon vessels, which were only present in one of the images, were still visible and were kept intact after the registration, as shown in *Figure 14*. However, there was also an inconsistency next to just misalignments. The inconsistency can be seen in *Figure 17c*, where small speckles were visible.

Nonetheless, the vessel structures were aligned well after the image registration. *Figure 15* shows the initial distance differences between the landmarks between the first two measurements. There is clearly a larger difference in distance between the landmarks of the fixed image and the moving image in experiment 3 compared to experiment 1. The post-registration result demonstrates the success of the image registration result, as indicated by a reduced mean displacement value of 2.15 mm from 29.14 mm, as shown in *Table 22* in *Appendix E.*

Experimental evaluation of the framework with serial image data set

Pre-registration   Post-registration

TRE: 24.98 ± 6.75   TRE: 1.97 ± 2.62

> 0 cm   > 2 cm   > 4 cm   > 0 cm   > 2 cm   > 4 cm
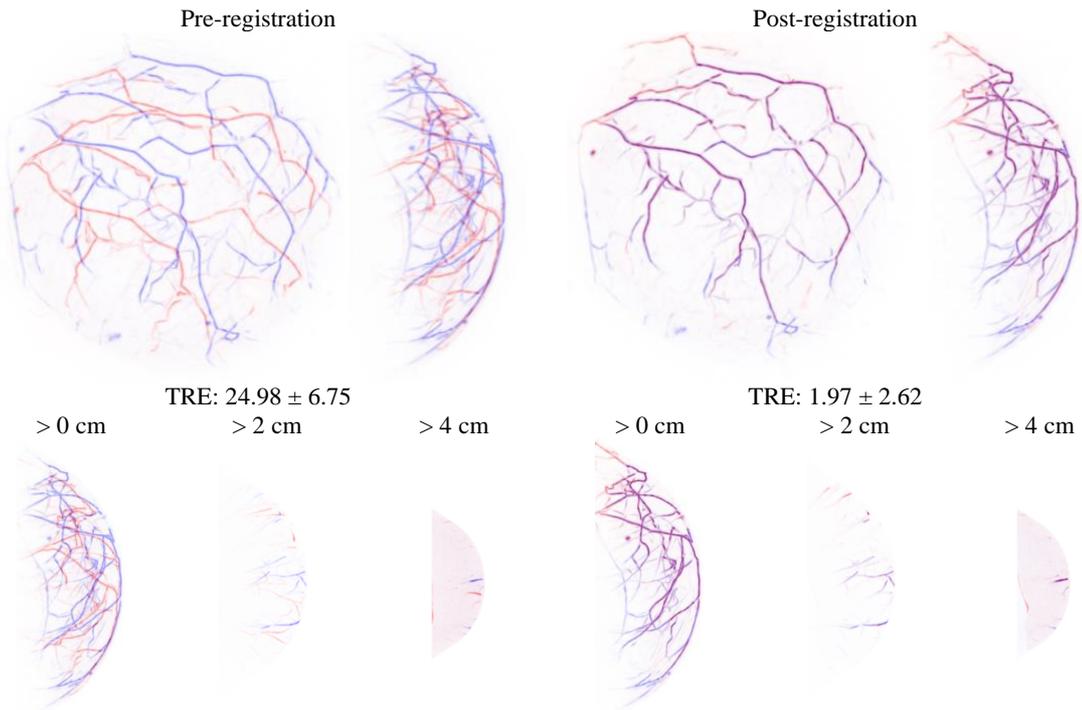
Figure 14: MIPs and depth images of before and after image registration of Experiment 3

Experiment 1

Pre-registration   Post-registration

Experiment 3

Pre-registration   Post-registration
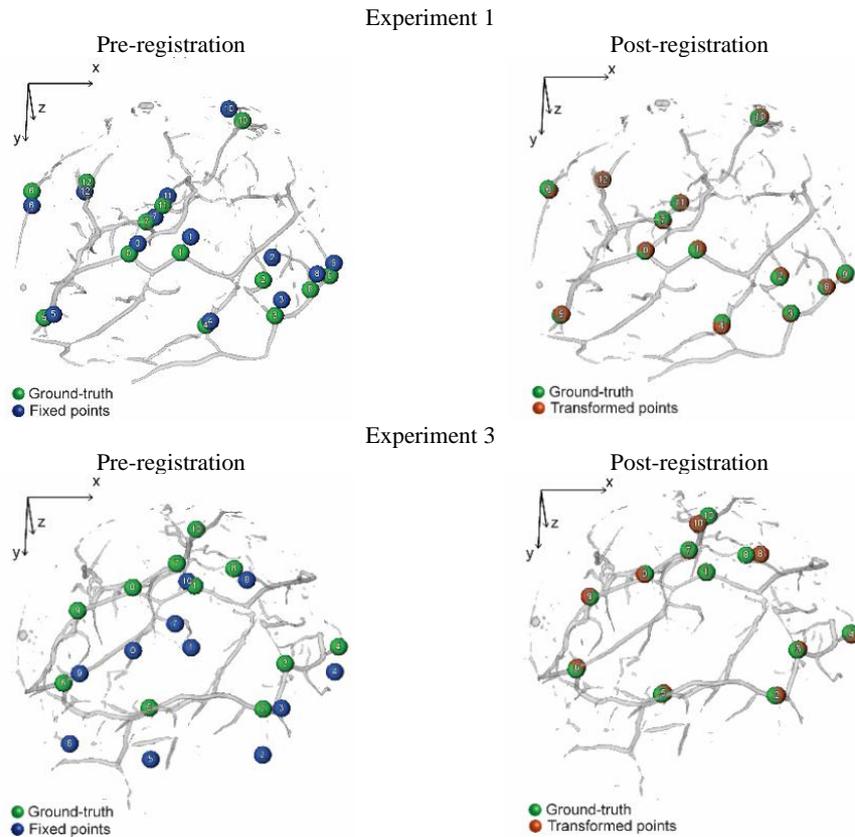
Figure 15: Figure: 3D rendered vascular network of the moving image; before co-registration: green spheres represent the ground truth, namely the annotated points on the moving image, and blue spheres are the annotated points on the fixed image; after co-registration: red spheres are the annotated points on the fixed image after transformation to the moving coordinate system [50]

Experimental evaluation of the framework with serial image data set

The wrong sizes of breast-supporting cups were used in the fourth experiment, which led to changes in the field-of-view. This change was visible as some vessels were only present in one image. In images 11 to 13, vessels close to the chest wall fell out of the boundaries of the cup, so they were not imaged. In images 14 and 15, some vessels in the center were less visible compared to reference images 8 and 9, while some peripheral vessels closer to the chest wall were more visible, see *Figure 16*. In *Figure 16*, the blue arrows show examples of the vessels that are more visible in the reference image, and the red arrows show examples of vessels that are more visible in the moving image.

Like in the third experiment, there was an inconsistency next to just misalignments. For example, in *Figure 17e*, a small vessel appears as a series of dashed segments instead of a continuous and unbroken vessel.

As expected, TRE analysis showed that the most significant mean displacement occurred during the second part of the experiment. This part involved the largest deformation of the moving image due to the substantial change in breast-supporting cup size. The difference in cup size resulted in a cup volume shrinkage from 856 mL to 496 mL and a depth difference of 19 mm. In *Figure 16*, the difference in depth is quite visible in the sagittal images of experiments 4a and 4b before the registration. The largest initial mean displacement occurred in experiment 4b, which was 28.11 ± 5.25 mm. This value decreased to 1.45 ± 0.9 mm, confirming a successful registration.
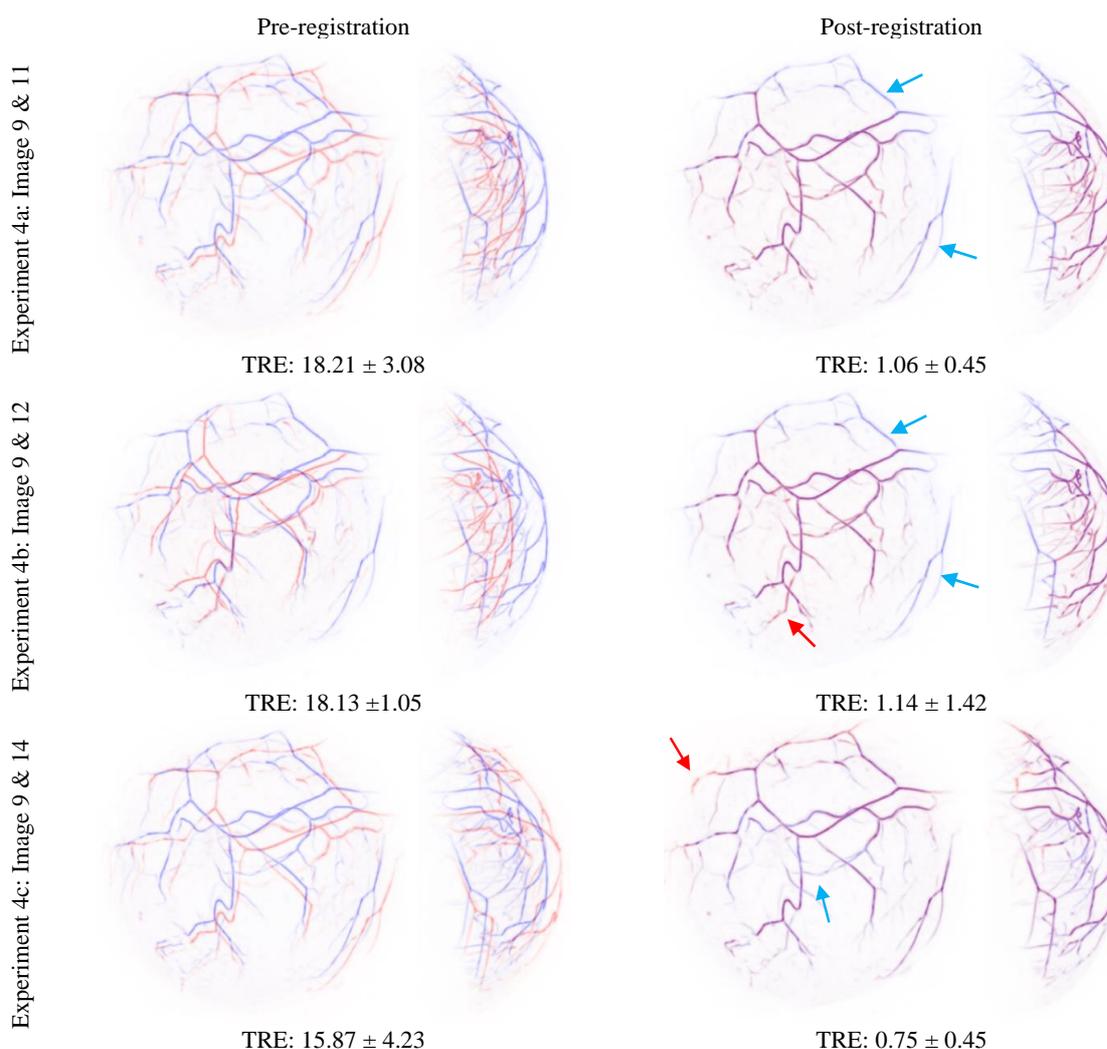
Figure 16: MIPs of before and after image registration of Experiment 4

Experimental evaluation of the framework with serial image data set

a.
01-01 – 02-01/02-02

b.
01-01 – 06-01

c.
01-01 – 07-01

d.
01-01 – 07-01

e.
08-01 – 12-01

f.
08-01 - 13-01
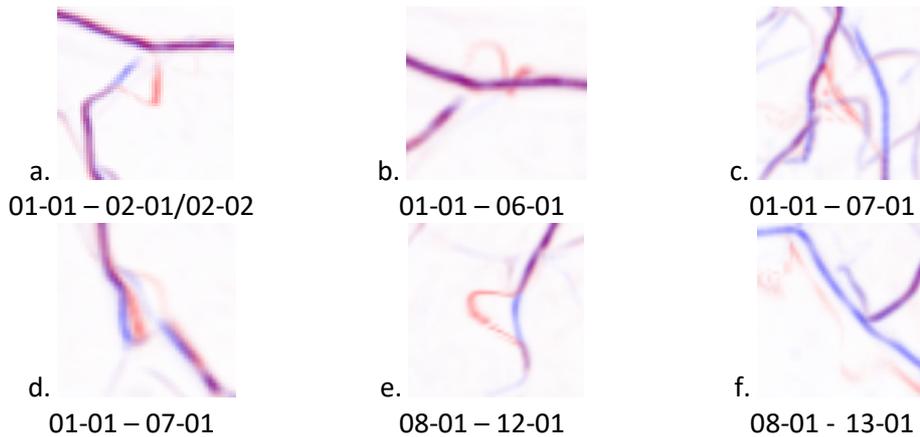
Figure 17: Misalignments and inconsistencies

The percentage change in the evaluation metric is listed in *Table 9*. The mean and standard deviations are calculated for every part or subpart of the experiment for every metric. All registered images have shown improvements in all five evaluated metrics. The percentage change after registration can be significantly higher in some experiments compared to others. For example, the more challenging experiments, like experiment 4, have shown large positive changes in all metrics compared to other experiments. Still, the absolute post-registration metric values are worse than those of experiment 1. The discrepancy in the initial metrics between the reference images in challenging and less challenging scenarios is high. The initial metric values of the reference image are lower in a challenging scenario compared to a less challenging scenario. This leads to a significant percentage change in experiments 3 and 4, compared to experiments 1 and 2, although they show lower post-registration metrics values. So, it is also good to use absolute post-registration metric values to have a more reliable view of the results.

Experiment 1 showed the highest increase in PSNR. It showed the highest mean post-registration PSNR and the highest mean PSNR change, while the initial mean PSNR value was also the highest compared to other experiments. Experiment 3 showed the smallest increase in PSNR compared to other experiments.

Although experiments 1 and 2 showed a higher post-registration SSIM, the largest increase in mean SSIM was found in experiments 4a and 4b, which had moving images using smaller breast-supporting cups. The first two parts of experiment 4 had significantly lower initial SSIM values than other experiments, which clarifies the large increase. The mean initial SSIM value of experiment 4b is 0.7543, while the mean initial SSIM value of experiment 1 is 0.8582. The low initial value in experiments 4a and 4b is probably due to a large decrease in the breast-supporting cup volume and depth. There is no depth difference between the cups used in other experiments, even when a larger cup is used.

While experiment 1 exhibited the best overlap, resulting in the highest post-registration DSC, the largest mean DSC change occurred in experiment 4b. The skin was included in the segmentations for a better registration result, which led to significantly lower initial DSC values in experiment 4 than in other experiments. When a differently sized breast-supporting cup is used, the surfaces of the cups do not match.

The highest mean NCC change was found in experiments 4a and 4b, which had moving images of smaller breast-supporting cups. The initial correlation between the images is significantly lower when a smaller breast-supporting cup is used for the moving image than when a larger cup is used. As explained before, this is probably caused by the difference in depth between the cups. Although moving images of smaller cup

sizes have shown the largest increase in NCC, the post-registration NCC is higher in all other experiments except for experiment 3. Experiment 1 shows the highest post-registration NCC

Table 9: The percentage change of PSNR, SSIM, DSC, NCC, and TRE before and after the image registration Acronyms: PSNR = peak signal-to-noise ratio, NCC = normalized cross-correlation, DSC= Dice similarity coefficient, TRE = target registration error.

	Image pair	ΔPSNR (%)	ΔSSIM (%)	ΔDSC (%)	ΔNCC (%)	ΔTRE (%)
Experiment 1	01-01 – 02-01	10.174	6.605	91.286	99.539	-91.89
	01-01 – 03-01	14.571	12.160	463.049	447.983	-87.12
	01-01 – 04-01	15.490	12.619	426.040	447.475	-89.71
	01-01 – 05-01	16.344	13.729	572.296	505.554	-89.48
	Mean	14.145	11.278	388.168	375.138	-89.55
	Standard deviation	2.12579	2.4666	160.676	143.877	1.511
Experiment 2	Image pair	ΔPSNR (%)	ΔSSIM (%)	ΔDSC (%)	ΔNCC (%)	ΔTRE (%)
	01-01 – 02-02	9.368	6.189	360.066	93.323	-92.07
	01-01 – 03-02	13.994	11.888	550.336	437.341	-86.55
	Mean	11.681	9.039	455.201	265.332	-89.31
	Standard deviation	2.313	2.850	95.135	172.009	2.76
Experiment 3	Image pair	11.681	9.0385	455.201	265.332	-89.31
	01-01 – 06-01	10.222	11.129	459.093	367.705	-92.11
	01-01 – 07-01	8.259	9.269	461.175	320.424	-93.00
	Mean	9.2405	10.199	460.134	344.0645	-92.56
	Standard deviation	0.98150	0.9300	1.0410	23.6405	0.4450
Experiment 4a	Image pair	ΔPSNR (%)	ΔSSIM (%)	ΔDSC (%)	ΔNCC (%)	ΔTRE (%)
	08-01 – 11-01	13.607	17.980	804.30	819.840	-92.11
	09-01 – 11-01	13.604	18.221	714.17	795.425	-93.01
	Mean	13.606	18.101	759.24	807.633	-92.56
	Standard deviation	0.0015000	0.12050	45.065	12.2075	0.4500
Experiment 4b	Image pair	ΔPSNR (%)	ΔSSIM (%)	ΔDSC (%)	ΔNCC (%)	ΔTRE (%)
	08-01 – 12-01	14.055	20.357	1369.95	1263.45	-93.61
	09-01 – 13-01	13.310	20.343	748.331	1189.68	-93.71
	08-01 – 12-01	11.708	17.636	1412.34	1162.00	-94.84
	09-01 – 13-01	12.898	19.758	1499.50	1251.16	-95.27
	Mean	12.993	19.524	1257.53	1216.57	-94.36
	Standard deviation	0.84987	1.1162	297.672	42.1163	0.7148
Experiment 4c	Image pair	ΔPSNR (%)	ΔSSIM (%)	ΔDSC (%)	ΔNCC (%)	ΔTRE (%)
	08-01 – 14-01	13.76	15.26	1001.05	614.01	-86.40
	09-01 – 14-01	13.31	15.57	1543.58	625.71	-95.27
	08-01 – 15-01	11.35	13.96	787.00	553.46	-94.66
	09-01 – 15-01	12.05	14.71	1147.87	588.64	-89.99
	Mean	12.62	14.87	1119.88	595.46	-91.58
	Standard deviation	0.9625	0.6117	276.243	27.703	3.623

An additional experiment was done using a measurement without any breast-supporting cup to challenge the framework further. This scenario will not happen when a PAM3 imager is used, as a breast-supporting cup is always used during the measurement with PAM3. However, this could be used for a preliminary test towards multi-modal imaging with MRI, as breast MRI also does not use a breast-supporting cup.

Not using a breast-supporting cup led to a very noisy image with a low signal due to the motion of the volunteer, uneven illumination, and difficulties with reconstruction due to the different shape of the breast. Next to the increased noise and decreased signals, the positioning of the vasculature changed drastically

compared to other measurements, as the breast was hanging. The comparison image between the correctly repositioned breast with a breast-supporting cup and the breast without any support can be seen in *Figure 18*.

The algorithm did not successfully align the vasculature of a breast without any support to the vasculature of a correctly positioned breast, *Figure 18*. There are two potential causes. The primary potential cause is the large deformation between the two images. Another potential cause could be the low signal-to-noise ratio.

A different set of training parameters were used for this experiment, as the original set of parameters did not work. Various parameters were tested based on a trial and error approach. However, none of these parameters resulted in the alignment of the vasculature. There might be better parameters, which results in better results. However, this could not be found during our experiment.
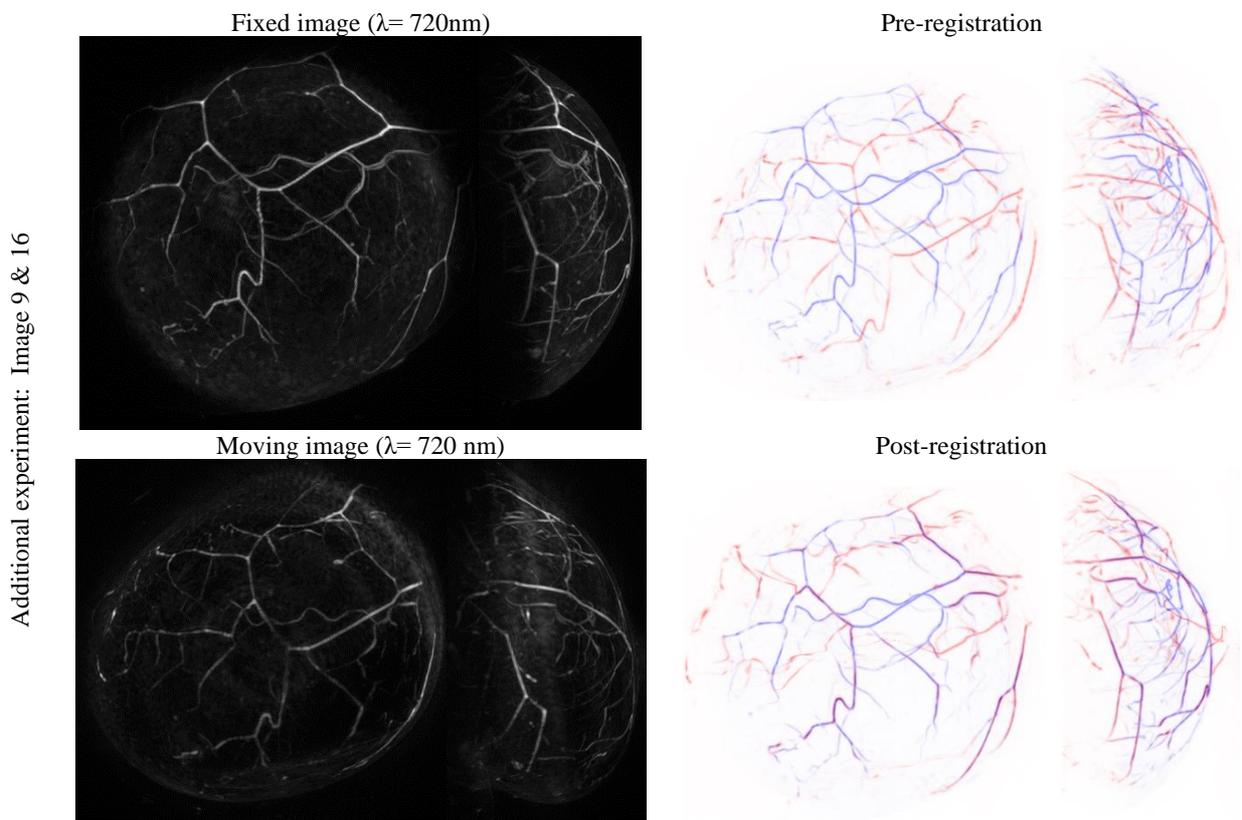
Figure 18: MIPs of Fixed and Moving images without any pre-processing (Left), MIPs before and after image registration of Additional experiment with no breast-supporting cup (Right)

6.5.1 Comparison study

Elastix is used again to compare the results of MUVINN with another conventional framework. The data sets from the two most probable situations were selected for the comparison study, which are from the first and the third experiments. The change in cup size, like in experiment 4, is unlikely to happen in a real scenario.

To ensure a fair comparison, the images were pre-processed using Frangi vesselness filtering with the same sigma values used for MUVINN (σ= {12,9,5,3,2}), and adaptive intensity modulation was performed. The initial parameters were selected based on the accurate registration results and similar computational times of MUVINN.

Experimental evaluation of the framework with serial image data set

A rigid transformation followed by a non-rigid B-spline was chosen as the transformation model to correct for both linear and non-linear deformations to achieve image registration. The coordinates were sampled inside the cup to optimize transformation only inside the breast. A multi-resolution scheme with four pyramid levels was used with a number of iterations for each pyramid equal to 400 for rigid and equal to 1200 for B-spline. For the similarity term of the loss function, the AdvancedNormalizedCorrelation and the TransformRigidityPenalty were used as regularization. The details about the parameters can be found in *Appendix B*.

In experiment 1, Elastix has shown better alignment than before the registration. However, every image pair had multiple misalignments. *Figure 19* shows the best registration result of experiment 1, which still contains multiple misalignments which were not visible in the results of MUVINN. In *Figure 19*, the misalignments are highlighted with a dashed box in the full coronal view of the MIP. The misalignments are zoomed in in the bottom part of the figure with a border of the same color as the dashed box in the full coronal view.

In contrast, MUVINN showed an accurate alignment. No misalignment was detected in the same image pair, and it outperformed Elastix in every evaluation metric. TRE values of Elastix show a minor mean decrease of 13.11% compared to pre-registration, while MUVINN shows a decrease of 89.55%.
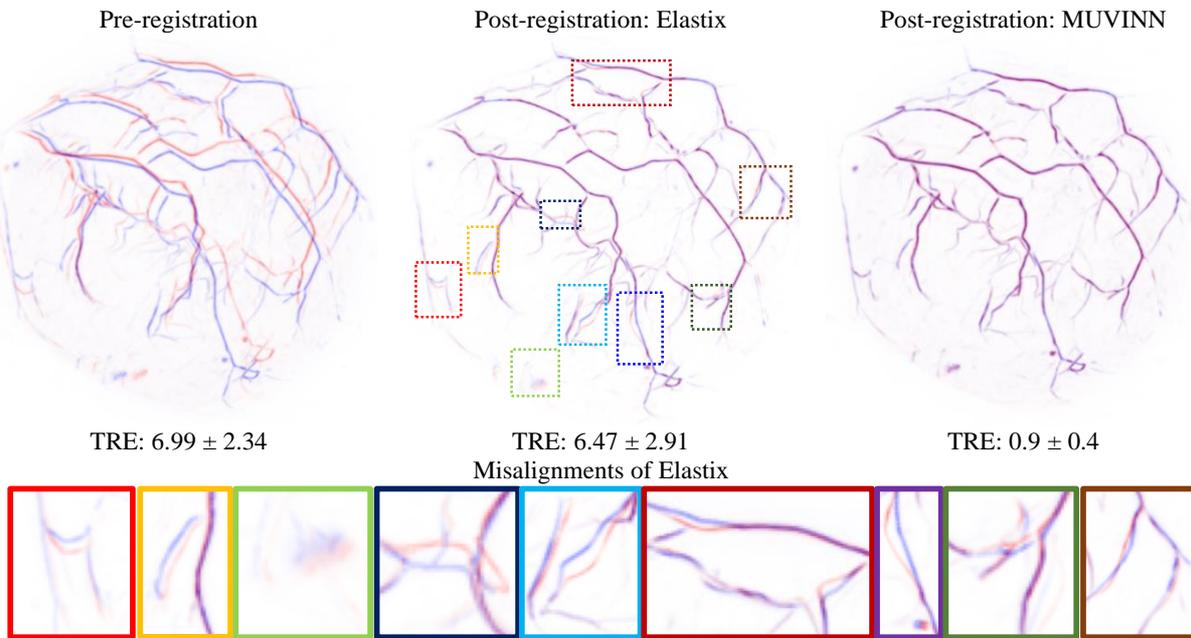
Figure 19: Comparison of registration result between Elastix and MUVINN in experiment 1; Coronal MIPs of the image pair (Top), Misalignments of Elastix highlighted (Bottom)

Table 10 shows pre- and post-registration values of all evaluation metrics used. Compared to the post-registration values of MUVINN *Evaluation metrics* in *Appendix E*, they are all lower.

Table 10: PSNR, SSIM, DSC, NCC, and TRE before and after the image registration with Elastix in experiment 1

Image pairs		PSNR	SSIM	DSC	NCC	TRE (mm) [mean, std]
01-01 –	Before	39.95	0.8903	0.3329	0.3649	[10.97, 3.78]
02-01	After	39.34	0.8724	0.4044	0.3649	[8.66, 3.57]

01-01 – 03-01	Before	38.75	0.8492	0.1193	0.1372	[6.99, 2.34]
	After	41.12	0.9093	0.6293	0.4832	[6.42, 2.977]
01-01 – 04-01	Before	38.80	0.8501	0.1317	0.1419	[6.22, 1.79]
	After	40.34	0.8972	0.5372	0.3759	[4.38, 1.18]
01-01 – 05-01	Before	38.55	0.8432	0.1029	0.1288	[5.99, 1.23]
	After	39.82	0.8836	0.5071	0.3759	[8.00, 1.18]
Mean	Before	39.01	0.8582	0.1717	0.1932	7.54
	After	40.16	0.8906	0.5195	0.4000	6.87
Standard deviation	Before	0.5493	0.01872	0.09363	0.09924	2.01
	After	0.6599	0.01391	0.08027	0.04826	1.65

In experiment 3, Elastix has shown worse results than before the registration. The MIP in *Figure 20* shows that none of the vessels were aligned, and the mean TRE value increased from 29.14 to 31.70. The MIPs at different depths showed that the vessels were also not aligned in depth.

The metrics of experiment 3 remained similar to their initial values, see *Table 11*. These results show that Elastix is not able to register images in more challenging scenarios than just repositioning.

In contrast, MUVINN proved that it can register mispositioned images to correctly repositioned images. The full MIPs and MIPs at different depths show that the vessels are accurately aligned. All the evaluation metrics have shown an improvement after the registration. The mean TRE decreased by 92.56% post-registration.
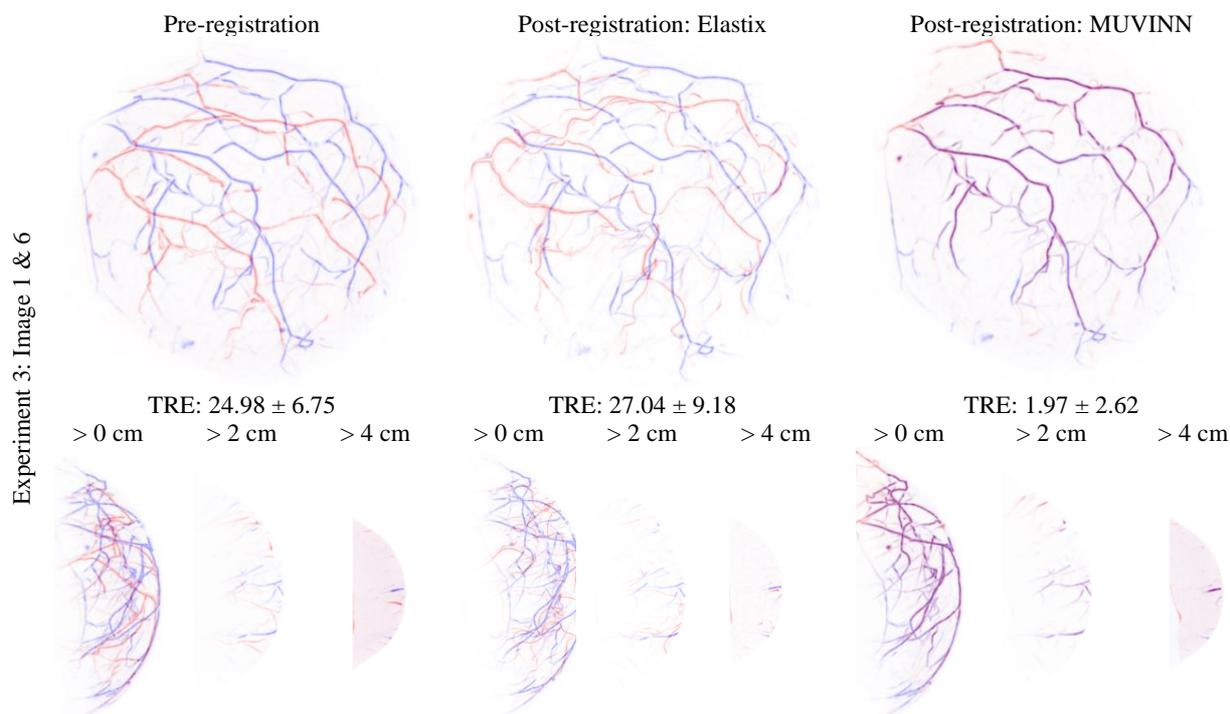

Figure 20: Comparison of registration result between Elastix and MUVINN in experiment 3

Table 11: PSNR, SSIM, DSC, NCC, and TRE before and after the image registration with Elastix in experiment 3

Image pairs		PSNR	SSIM	DSC	NCC	TRE (mm) [mean, std]
01-01 – 06-01	Before	37.96	0.8239	0.08669	0.1236	[24.98, 6.75]
	After	38.33	0.8290	0.05883	0.1239	[29.04, 7.98]
01-01 – 07-01	Before	38.27	0.8226	0.07674	0.1209	[33.31, 5.16]
	After	38.53	0.8252	0.03973	0.1238	[34.02, 5.89]
Mean	Before	38.12	0.8233	0.08172	0.1223	29.14
	After	38.43	0.8271	0.04928	0.1239	31.53
Standard deviation	Before	0.1829	0.002140	0.01602	0.001114	3.25
	After	0.1000	0.001900	0.009550	0.00005	2.49

The comparison study with Elastix reveals that MUVINN demonstrates more accurate registration results for PA breast images compared to Elastix.

6.6   Chapter discussion & conclusion

Although the breast-supporting cup improves the light distribution and stabilizes the breast to prevent motion artifacts, it is not enough to make the images reproducible. This is visible in the images and also in the initial evaluation metrics. For example, the initial TRE values range from 5.99 to 10.97 mm in correctly positioned breast images. Due to the deformability of the breast tissue, repositioning the breast will introduce repositioning errors and may change the field-of-view.

The image registration framework MUVINN has shown good alignment of the structures. All evaluation metrics indicated an improvement in alignment after the registration. The framework could register images in challenging situations, such as after increasing the magnitude of deformation by purposefully mispositioning the breast or changing the field-of-view using a breast-supporting cup of the wrong size. The alignment could also be seen visually, which was visible in MIP images, MIPs at different depths, and 3D renderings.

Like in experiments with synthetically deformed data set, MUVINN showed superior results compared to other conventional image registration framework, Elastix. Two factors could potentially cause Elastix's suboptimal performance. Firstly, the chosen transformation model, rigid and B-spline, might not be suitable for capturing the deformations caused by breast repositioning. Secondly, using a relatively low number of iterations may hinder the optimization process. The computation time for Elastix was lower than MUVINN. However, increasing the number of iterations by 300 for both transformation models did not improve the results, while the computational time increased to 23 minutes and 59 seconds on average.

For future studies, including additional image registration frameworks for a more comprehensive comparison would be beneficial. Demons and Voxelmorph could be valuable options to consider. Demons is an algorithm that can compute the deformation field based on intensity differences and gradients. Unlike Elastix, it does not require parametrization of the transformation. Demons also use a multiscale approach like MUVINN and Elastix [43]. VoxelMorph is an unsupervised deep-learning framework for image registration of medical images. It uses an architecture similar to U-Net. VoxelMorph aligns images based on their intensity values using similarity metrics like mean squared error or local normalized cross-correlation [58].

The amount of epochs has been doubled for serial image data experiments compared to the synthetically deformed data experiments due to the higher complexity of the image registration problem. This led to a higher computational time compared to the synthetically deformed data experiments. The mean

computational time was 20 minutes and 3 seconds with a margin of ± 41 seconds. Each pair of images requires new training, as a new neural network needs to be optimized. This has several benefits as the network is unsupervised and does not need a training set. However, this also leads to higher computational times than a supervised technique.

For all serial image data experiments, identical initial training parameters were used. Selecting the right parameter was performed based on a trial and error approach, involving multiple attempts to identify a set of parameters that worked for every case during the experiments.

Some parameters were more critical than others. The number of iterations and the choice of scales were critical. A higher number of iterations usually gave better image registration results. However, it was at the expense of computational time. By increasing the sigma (σ) values, the algorithm will prioritize aligning large structures before fine-tuning the alignment for more minor details. High σ also produces more regular displacement fields, but there is a risk of fine structures being neglected.

Conversely, a very low σ can lead to a wrong displacement field with irregular distortions in the transformed image. Thus, finding an effective combination of scales was crucial for the performance of MUVINN. Similarly, when dealing with distortions in the transformed image, adjusting the learning rate or increasing the patch length to capture the neighbouring points could work to prevent these distortions. Decreasing the learning rate could prevent distortions until a certain limit. The network's learning capability will be constrained when it is too low. By increasing the patch length, the network can have a more overall view of a larger area instead of focusing on details in a small area. During the parameter tests, the transformation of the image was monitored, and the parameters were adjusted when distortions started to happen. Like other networks, the explainability is limited and finding the right parameter can be challenging.

These parameters will likely not work for every case, but they worked in more challenging scenarios, such as purposely mispositioning the breast in the cup or using the wrong cup size. There were better training parameters for each experiment, which showed a higher image registration quality. However, the parameters used in this research have shown the best general performance without changing the parameters for each experiment.

The framework did not succeed in aligning the vasculature of a breast without any support to the vasculature of a correctly positioned breast. There are two potential causes. The primary potential cause is the large deformation between the two images. Another potential cause could be the low signal-to-noise ratio. To test if the second cause was true, image pair 01-01 – 02-01 was used with a modified moving image. Both images are imaged with a correctly positioned breast. The moving image was modified by adding noise, and the signal of the vasculature was decreased. The intensity of the signal and the noise were similar to the image with no breast-supporting cup. The framework could still align the image pair well, meaning the second cause could be eliminated. These results gave some insights into the framework. As expected, the framework is robust to noise and intensity shifts due to the use of multiscale approach and normalized cross-correlation as a similarity metric. However, drastic changes in the shape and positioning of the vasculature can lead to misalignment.

In conclusion, the results suggest that MUVINN is a robust network for image registration of correctly repositioned images, whether they share the same wavelength or differ in wavelengths, and for challenging scenarios like mispositioning and using a wrong breast-supporting cup size. There was no need to adjust the initial parameters for any of these experiments, even in different scenarios.

The renderings and evaluation metrics collectively indicate that the registered image shows a good alignment by showing a remarkable positive change in all five evaluation metrics. It performed the best in experiment 1, where repositioned images were aligned without introducing other errors.

7 General discussion

For diseases that change over time, like tumors, it is crucial to accurately register images from different time points to track disease progression. During the treatment of breast cancer, for example, with NAC, efficacy can vary between individuals. In some cases, the therapy might have minimal or no impact at all. The variability in response necessitates the importance of monitoring chemotherapy. Monitoring the disease by means of longitudinal imaging is necessary to spare the non-responders by preventing further disease progression and avoiding unnecessary exposure to toxicity. Longitudinal image registration helps doctors to track the same structures of interest over time and during the course of NAC and make informed decisions.

There are several imaging modalities, such as MRI, US, PET, etc., which can be used for monitoring NAC. However, all these methods have certain limitations like challenges in tracking biomarkers related to NAC efficacy, lack of reproducibility, and the invasive nature of certain imaging technologies for patients. Despite the advantages these modalities offer, their limitations emphasize the need for more efficient and comprehensive imaging techniques for assessing the effectiveness of NAC. Compared to these standard-of-care imaging modalities, PA imaging has advantages for repetitive measurements. It could be a valuable alternative to these modalities by visualizing functional responses in the tumor microenvironment in a non-invasive manner.

However, one important requirement for longitudinal imaging is that the images between the imaging sessions are comparable. In PA imaging, repositioning the breast can cause geometric misalignment of vasculature and changes in the field-of-view. Image registration is required for comprehensive quantitative and qualitative analysis with PA imaging.

A few image registration algorithms are proposed for PA imaging, but these methods did not meet the goal of this research, as most of them were aimed at multimodal image registration [59, 60]. Other image registration methods, which focused on unimodal PA imaging registration, used a parametric approach using transformation models (e.g., rigid, affine, or B-splines) [61, 62]. The parametric approach can be problematic in the case of complex and unknown deformations like breast repositioning deformation.

There are also several deep-learning approaches available for medical image registration. However, most of these require a pre-training phase, which requires a large data set. Compared to other standard-of-care imaging modalities, there are no large data sets of PA imaging available. Therefore, there is a need for an image registration framework that can align in vivo 3D PA images with complex and non-linear deformations.

Our image registration framework is developed for vascular image registration to address these issues, which uses coordinate-based neural networks and multiscale Frangi vesselness filtering. The goal of this framework is unimodal intra-patient image registration of 3D PA images for longitudinal imaging. A list of requirements of a robust framework for an image registration framework for monitoring NAC is made, which are listed in *Requirements of a robust image registration framework.* To validate if the framework fulfils some of these requirements, experiments are set up by creating data sets with geometric misalignments in a realistic scenario by repositioning the volunteer, and in challenging scenarios, by introducing more significant change in field-of-view and deformations by mispositioning the volunteer and using a wrong-sized breast-supporting cup. MUVINN proved that it can correct complex non-rigid deformations and demonstrated a significant reduction of displacement and increase in image similarity in high accuracy, which could be assessed visually and quantitatively. By successfully aligning images with changes in the field-of-view, such as in experiments 3 and 4, the framework showed its robustness to appearing and disappearing vessels, which can happen as an effect of treatment or disease progression during monitoring. This proved that the framework is robust to pathological changes. Jacobian determinant map was computed to assess if the vessel's diameter changed. The Jacobian determinant is the determinant of the Jacobian

matrix, which was already computed for the regularization, as mentioned in the *Loss function*. The Jacobian determinant map represents the ratio of the volume change of a local region before and after the transformation at each voxel. It shows local volumetric shrinkage and expansion regions in the image. Visualizing this local distortion of the transformation can be helpful in detecting and correcting registration errors [63]. The Jacobian determinant map was tested in a simplified test where a small sphere was aligned to a large sphere. Here, the map showed that there was a volumetric increase inside the sphere indeed. However, when it was used for PA images with no diameter change, it showed that there was volumetric change, while there was no volumetric change. Although the Jacobian determinant map has proven to work for visualizing volume changes in large structures such as organs and tumors [63, 64], it appears that it is not an appropriate method to visualize changes in the volume of a small structure like a vessel.

MUVINN is INR-based and can represent deformation fields in a continuous domain, which allows the deformation field to be known at any coordinate, unlike frameworks which require transformation models. This approach also allows the integration of any similarity metrics as long as they are differentiable. For PA image registration, the similarity term was tailored for vascular images by incorporating Frangi vesselness filtering, which enhances signals from the vasculature and decreases potential noise. Combining this feature with normalized cross-correlation makes the method robust to variations in image intensities compared to other intensity-based similarity metrics like MSE. Additionally, Jacobian regularization in the loss function contributed to the regular deformation field by penalizing large vascular expansion and shrinkage.

A coarse-to-fine strategy is implemented, which has shown good results in other registration frameworks [37]. By decomposing the image in different scales, the framework performs more efficiently even in the presence of noise or variations in structure, leading to improved robustness of the framework. The improved performance could also be seen in MUVINN when the coarse-to-fine strategy was applied, see *Comparison of the framework with and without a multiscale approach* in *Appendix F*.

The combination of all these features contributed to a robust image registration framework capable of registering serial PA images in different conditions, as demonstrated in the *Results*. The framework could align vascular structures: 1) after displacement in different magnitudes, 2) in different intensities, and 3) in different field-of-view.

Like many other deep learning frameworks, it is challenging to understand how the network makes certain decisions depending on the parameters. Parameter tuning is necessary for optimal performance of MUVINN, which is done by a trial and error approach, explained in more detail in *Chapter discussion* of Chapter 5. Although the final parameters worked for all experiments, there were some inaccuracies in some vessels after registration in experiments 3 and 4, as shown in *Figure 17*. It could be that these inconsistencies occurred because these small structures did not have a significant influence on the loss function during optimization. Increasing the number of iterations or points sampled in each iteration may solve this issue, but the computational expenses would also increase. The average computational time of MUVINN was 20 minutes and 3 seconds, with a margin of ± 41 seconds. These are acceptable computational times for disease treatment monitoring, but they are too long if this framework gets integrated into real-time applications.

7.1   Other applications and future works

The main goal of this framework is to enable longitudinal PA imaging for monitoring the treatment effect of NAC for breast cancer patients. Some other potential applications are using the framework in image mosaicking. Image mosaicking in medical imaging refers to the process of combining multiple overlapping images to produce a single, larger image. This application is necessary when the region of interest is larger than the field-of-view of the imaging device. In experiment 4, the algorithm aligned the images with different field-of-views, where some vessels were only present in one of the two images.

Another potential application is using this framework to align images of the same conditions for frame averaging. In frame averaging, multiple independent images are averaged or summed. This leads to increased signal, which is consistent across the images, while the random noise tends to cancel out or reduce, resulting in clearer images. This may help to enhance smaller vessels that could not be identified with a human eye in MIPs or 3D renderings but were still aligned using the framework.

The framework shows potential for a wider range of applications if it could be used across different imaging modalities or for multimodal image registration. Identifying tumor locations on the PA image becomes feasible if the framework can be used for multimodal imaging like PA-MRI registration. Combining these two modalities can offer a more comprehensive view, combining high-resolution images of the vasculature to assess the tumor microenvironment with the tumor and surrounding breast tissue. This could improve monitoring of the disease and enable the discovery of new image features indicating disease progression. An additional PA image without a breast-supporting cup was acquired for a preliminary test towards PA-MRI image registration, as a breast-supporting cup is not used in breast MRI. Unfortunately, the framework struggled to align the image with and without the cup. More investigation is needed to accomplish multimodal image registration with this framework.

If this framework proves to work with other vascular images than PA images, it could be used for applications such as vascular image registration of MRI images to quantify vascular growth over time. This is essential in treatments with tissue-engineered grafts for the treatment of congenital heart disease [64].

There are still important points to investigate to ensure the framework's efficacy and relevance. The most important step is testing the framework with its target demographic: breast cancer patients undergoing NAC. It is still largely unknown how the therapy changes the morphology of the vasculature and breast. Thus, evaluating the framework in actual treatment monitoring conditions is essential to confirm its efficacy.

Another future direction for improving the framework includes automated parameter optimization and a user-friendly interface, enhancing the framework's usability in a practical setting.

8 General conclusion

Reflecting on the overarching research question of this thesis: How can we develop and validate an accurate and robust image registration framework to monitor the effects of neoadjuvant chemotherapy in breast cancer using photoacoustic images?

We've developed a novel image registration framework, MUVINN, for unimodal intra-patient registration of three-dimensional PA images, using coordinate-based neural networks and multiscale Frangi vesselness filtering. This framework demonstrated significant improvements in image alignment under various unfavorable conditions, such as after deformation in different magnitudes, in different intensities, and in different field-of-view. In *Requirements of a robust image registration framework*, essential criteria for a robust image registration framework for longitudinal PA imaging were outlined. The framework showed a high accuracy with a reasonable average computation time of 20 minutes and 3 seconds. The framework is robust to noise due to a multiscale approach and intensity variations due to the use of normalized cross-correlation in its loss function and intensity modulation. The framework's capability to correct complex non-linear local deformations and adapt to pathological changes was proven in experiments.

In summary, this framework is a promising tool for quantitatively monitoring disease progression and treatment response in breast cancer using photoacoustic. It has been proven to work for unimodal image registration of PA images with a Twente Photoacoustic Mammoscope 3 imager. However, more research is needed for multimodal image registration, such as PA-MRI image registration.

9 References

1. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.: Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 71, 209–249 (2021). https://doi.org/10.3322/CAAC.21660

2. Asaoka, M., Gandhi, S., Ishikawa, T., Takabe, K.: Neoadjuvant Chemotherapy for Breast Cancer: Past, Present, and Future, https://journals.sagepub.com/doi/full/10.1177/1178223420980377, (2020)

3. Piotrzkowska-Wróblewska, H., Dobruch-Sobczak, K., Klimonda, Z., Karwat, P., Roszkowska-Purska, K., Gumowska, M., Litniewski, J.: Monitoring breast cancer response to neoadjuvant chemotherapy with ultrasound signal statistics and integrated backscatter. PLoS One. 14, (2019). https://doi.org/10.1371/JOURNAL.PONE.0213749

4. Reig, B., Heacock, L., Lewin, A., Cho, N., Moy, L.: Role of MRI to Assess Response to Neoadjuvant Therapy for Breast Cancer. Journal of Magnetic Resonance Imaging. 52, (2020). https://doi.org/10.1002/JMRI.27145

5. Cerussi, A., Hsiang, D., Shah, N., Mehta, R., Durkin, A., Butler, J., Tromberg, B.J.: Predicting response to breast cancer neoadjuvant chemotherapy using diffuse optical spectroscopy. Proc Natl Acad Sci U S A. 104, 4014–4019 (2007). https://doi.org/10.1073/PNAS.0611058104/ASSET/0B981F35-8A97-48D2-8EBB-B5FD0734903B/ASSETS/GRAPHIC/ZPQ0060751390004.JPEG

6. Razansky, D., Manohar, S.: Photoacoustics: a historical review. Advances in Optics and Photonics, Vol. 8, Issue 4, pp. 586-617. 8, 586–617 (2016). https://doi.org/10.1364/AOP.8.000586

7. Michoux, N., Van den Broeck, S., Lacoste, L., Fellah, L., Galant, C., Berlière, M., Leconte, I.: Texture analysis on MR images helps predicting non-response to NAC in breast cancer. BMC Cancer. 15, 1–13 (2015). https://doi.org/10.1186/S12885-015-1563-8/TABLES/5

8. Zhu, Q., Tannenbaum, S., Kurtzman, S.H., DeFusco, P., Ricci, A., Vavadi, H., Zhou, F., Xu, C., Merkulov, A., Hegde, P., Kane, M., Wang, L., Sabbath, K.: Identifying an early treatment window for predicting breast cancer response to neoadjuvant chemotherapy using immunohistopathology and hemoglobin parameters. Breast Cancer Res. 20, (2018). https://doi.org/10.1186/S13058-018-0975-1

9. Kim, Y.S., Chang, J.M., Moon, H.G., Lee, J., Shin, S.U., Moon, W.K.: Residual Mammographic Microcalcifications and Enhancing Lesions on MRI After Neoadjuvant Systemic Chemotherapy for Locally Advanced Breast Cancer: Correlation with Histopathologic Residual Tumor Size. Ann Surg Oncol. 23, 1135–1142 (2016). https://doi.org/10.1245/S10434-015-4993-2

10. Skarping, I., Förnvik, D., Heide-Jørgensen, U., Rydén, L., Zackrisson, S., Borgquist, S.: Neoadjuvant breast cancer treatment response; tumor size evaluation through different conventional imaging modalities in the NeoDense study. Acta Oncol. 59, 1528–1537 (2020). https://doi.org/10.1080/0284186X.2020.1830167

11. Kong, X., Zhang, Q., Wu, X., Zou, T., Duan, J., Song, S., Nie, J., Tao, C., Tang, M., Wang, M., Zou, J., Xie, Y., Li, Z., Li, Z.: Advances in Imaging in Evaluating the Efficacy of Neoadjuvant Chemotherapy for Breast Cancer. Front Oncol. 12, (2022). https://doi.org/10.3389/FONC.2022.816297

12. Romeo, V., Accardo, G., Perillo, T., Basso, L., Garbino, N., Nicolai, E., Maurea, S., Salvatore, M.: Assessment and prediction of response to neoadjuvant chemotherapy in breast cancer: A comparison of imaging modalities and future perspectives, https://www.mdpi.com/2072-6694/13/14/3521/htm, (2021)

13. Ochi, T., Tsunoda, H., Matsuda, N., Nozaki, F., Suzuki, K., Takei, H., Yamauchi, H.: Accuracy of morphologic change measurements by ultrasound in predicting pathological response to neoadjuvant chemotherapy in triple-negative and HER2-positive breast cancer. Breast Cancer. 28, 838–847 (2021). https://doi.org/10.1007/S12282-021-01220-5/FIGURES/6

14. O'Flynn, E.A.M., deSouza, N.M.: Functional magnetic resonance: biomarkers of response in breast cancer. Breast Cancer Res. 13, 204 (2011). https://doi.org/10.1186/BCR2815

References

15. Cheng, Q., Huang, J., Liang, J., Ma, M., Ye, K., Shi, C., Luo, L.: The Diagnostic Performance of DCE-MRI in Evaluating the Pathological Response to Neoadjuvant Chemotherapy in Breast Cancer: A Meta-Analysis, www.frontiersin.org, (2020)

16. Al Ewaidat, H., Ayasrah, M.: A Concise Review on the Utilization of Abbreviated Protocol Breast MRI over Full Diagnostic Protocol in Breast Cancer Detection, https://doi.org/10.1155/2022/8705531, (2022)

17. Wang, J., Chu, Y., Wang, B., Jiang, T.: <p>A Narrative Review of Ultrasound Technologies for the Prediction of Neoadjuvant Chemotherapy Response in Breast Cancer</p>. Cancer Manag Res. 13, 7885–7895 (2021). https://doi.org/10.2147/CMAR.S331665

18. Yun, G., Kim, S.M., Yun, B. La, Ahn, H.S., Jang, M.: Reliability of automated versus handheld breast ultrasound examinations of suspicious breast masses. Ultrasonography. 38, 264 (2019). https://doi.org/10.14366/USG.18055

19. Baumgartner, A., Tausch, C., Hosch, S., Papassotiropoulos, B., Varga, Z., Rageth, C., Baege, A.: Ultrasound-based prediction of pathologic response to neoadjuvant chemotherapy in breast cancer patients. The Breast. 39, 19–23 (2018). https://doi.org/10.1016/j.breast.2018.02.028

20. Kumar, A., Srivastava, V., Singh, S., Shukla, R.C.: Color Doppler ultrasonography for treatment response prediction and evaluation in breast cancer. Future Oncol. 6, 1265–1278 (2010). https://doi.org/10.2217/FON.10.93

21. Barr, R.G.: Sonographic breast elastography: a primer. J Ultrasound Med. 31, 773–783 (2012). https://doi.org/10.7863/JUM.2012.31.5.773

22. Ma, Y., Zhang, S., Li, J., Li, J., Kang, Y., Ren, W.: Comparison of strain and shear-wave ultrasounic elastography in predicting the pathological response to neoadjuvant chemotherapy in breast cancers. European Society of Radiology. 27, 2282–2291 (2016). https://doi.org/10.1007/s00330-016-4619-5

23. Park, J.S., Moon, W.K., Lyou, C.Y., Cho, N., Kang, K.W., Chung, J.K.: The assessment of breast cancer response to neoadjuvant chemotherapy: comparison of magnetic resonance imaging and 18F-fluorodeoxyglucose positron emission tomography. Acta Radiol. 52, 21–28 (2011). https://doi.org/10.1258/AR.2010.100142

24. Dialani, V., Chadashvili, T., Slanetz, P.J.: Role of Imaging in Neoadjuvant Therapy for Breast Cancer. Ann Surg Oncol. 22, 1416–1424 (2015). https://doi.org/10.1245/S10434-015-4403-9/TABLES/1

25. Manohar, S., Dantuma, M.: Current and future trends in photoacoustic breast imaging. Photoacoustics. 16, 100134 (2019). https://doi.org/10.1016/J.PACS.2019.04.004

26. Kuniyil Ajith Singh, M., Dantuma, M., Kalloor Joseph, F., Manohar, S., Steenbergen, W.: Breast imaging using an LED-based photoacoustic and ultrasound imaging system: a proof-of-concept study. In: Photons Plus Ultrasound: Imaging and Sensing 2021. p. 89. SPIE (2021)

27. Nyayapathi, N., Xia, J.: Photoacoustic imaging of breast cancer: a mini review of system design and image features. J Biomed Opt. 24, 1 (2019). https://doi.org/10.1117/1.jbo.24.12.121911

28. Oraevsky, A.A., Clingman, B., Zalev, J., Stavros, A.T., Yang, W.T., Parikh, J.R.: Clinical optoacoustic imaging combined with ultrasound for coregistered functional and anatomical mapping of breast tumors. Photoacoustics. 12, 30–45 (2018). https://doi.org/10.1016/J.PACS.2018.08.003

29. Lin, L., Tong, X., Hu, P., Invernizzi, M., Lai, L., Wang, L. V.: Photoacoustic Computed Tomography of Breast Cancer in Response to Neoadjuvant Chemotherapy. Advanced Science. 8, 2003396 (2021). https://doi.org/10.1002/ADVS.202003396

30. Zhu, Q., Tannenbaum, S., Kurtzman, S.H., DeFusco, P., Ricci, A., Vavadi, H., Zhou, F., Xu, C., Merkulov, A., Hegde, P., Kane, M., Wang, L., Sabbath, K.: Identifying an early treatment window for predicting breast cancer response to neoadjuvant chemotherapy using immunohistopathology and hemoglobin parameters. Breast Cancer Res. 20, (2018). https://doi.org/10.1186/S13058-018-0975-1

References

31. Wu, L.A., Chang, R.F., Huang, C.S., Lu, Y.S., Chen, H.H., Chen, J.Y., Chang, Y.C.: Evaluation of the treatment response to neoadjuvant chemotherapy in locally advanced breast cancer using combined magnetic resonance vascular maps and apparent diffusion coefficient. Journal of Magnetic Resonance Imaging. 42, 1407–1420 (2015). https://doi.org/10.1002/JMRI.24915

32. Peter, J.L., Zhang, J.E., Treeby, B., Cox, B., Pedley, B., Beard, P., Laufer, J., Johnson, P., Zhang, E.: In vivo preclinical photoacoustic imaging of tumor vasculature development and therapy. https://doi.org/10.1117/1.JBO.17.5.056016. 17, 056016 (2012). https://doi.org/10.1117/1.JBO.17.5.056016

33. Liapis, E., Karlas, A., Klemm, U., Ntziachristos, V.: Chemotherapeutic effects on breast tumor hemodynamics revealed by eigenspectra multispectral optoacoustic tomography (eMSOT). Theranostics. 11, 7813 (2021). https://doi.org/10.7150/THNO.56173

34. Chen, X., Diaz-Pinto, A., Ravikumar, N., Frangi, A.F.: Deep learning in medical image registration. Progress in Biomedical Engineering. 3, 012003 (2021). https://doi.org/10.1088/2516-1091/ABD37C

35. Deshpande, A., Jamilpour, N., Jiang, B., Michel, P., Eskandari, A., Kidwell, C., Wintermark, M., Laksari, K.: Automatic segmentation, feature extraction and comparison of healthy and stroke cerebral vasculature. Neuroimage Clin. 30, 102573 (2021). https://doi.org/10.1016/j.nicl.2021.102573

36. Can, A., Stewart, C. V., Roysam, B., Tanenbaum, H.L.: A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. IEEE Trans Pattern Anal Mach Intell. 24, 347–364 (2002). https://doi.org/10.1109/34.990136

37. Klein, S., Staring, M.: Elastix: The Manual. Elastix (2020)

38. Wolterink, J.M., Zwienenberg, J.C., Brune, C.: Implicit Neural Representations for Deformable Image Registration. Proceedings of Machine Learning Research-Under Review. 1–11 (2022)

39. Zwienenberg, J., Thesis, M.S., Brune, C., Wolterink, J.: Neural implicit representations for diffeomorphic medical image registration, (2021)

40. Leshno, M., Lin, V.Y., Pinkus, A., Schocken, S.: Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. Neural Networks. 6, 861–867 (1993). https://doi.org/10.1016/S0893-6080(05)80131-5

41. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: A survey. IEEE Trans Med Imaging. 32, 1153–1190 (2013). https://doi.org/10.1109/TMI.2013.2265603

42. Chen, M., Tustison, N.J., Jena, R., Gee, J.C.: Image Registration: Fundamentals and Recent Advances Based on Deep Learning. In: Neuromethods. pp. 435–458. Humana, New York, NY (2023)

43. Nithiananthan, S., Schafer, S., Uneri, A., Mirota, D.J., Stayman, J.W., Zbijewski, W., Brock, K.K., Daly, M.J., Chan, H., Irish, J.C., Siewerdsen, J.H.: Demons deformable registration of CT and cone-beam CT using an iterative intensity matching approach. Med Phys. 38, 1785 (2011). https://doi.org/10.1118/1.3555037

44. Luengo-Gil, G., González-Billalabeitia, E., Chaves-Benito, A., García Martínez, E., García Garre, E., Vicente, V., Ayala de la Peña, F.: Effects of conventional neoadjuvant chemotherapy for breast cancer on tumor angiogenesis. Breast Cancer Res Treat. 151, 577–587 (2015). https://doi.org/10.1007/S10549-015-3421-4/FIGURES/3

45. Mitrofanova, I., Zavyalova, M., Riabov, V., Cherdyntseva, N., Kzhyshkowska, J.: The effect of neoadjuvant chemotherapy on the correlation of tumor-associated macrophages with CD31 and LYVE-1. Immunobiology. 223, 449–459 (2018). https://doi.org/10.1016/J.IMBIO.2017.10.050

46. Sara, U., Akter, M., Uddin, M.S., Sara, U., Akter, M., Uddin, M.S.: Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study. Journal of Computer and Communications. 7, 8–18 (2019). https://doi.org/10.4236/JCC.2019.73002

References

47. Tsai, D.M., Lin, C.T.: Fast normalized cross correlation for defect detection. Pattern Recognit Lett. 24, 2625–2631 (2003). https://doi.org/10.1016/S0167-8655(03)00106-5

48. Xiao, H., Teng, X., Liu, C., Li, T., Ren, G., Yang, R., Shen, D., Cai, J.: A review of deep learning-based three-dimensional medical image registration methods. Quant Imaging Med Surg. 11, 4895 (2021). https://doi.org/10.21037/QIMS-21-175

49. Zou, K.H., Warfield, S.K., Bharatha, A., Tempany, C.M.C., Kaus, M.R., Haker, S.J., Wells, W.M., Jolesz, F.A., Kikinis, R.: Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. Acad Radiol. 11, 178–189 (2004). https://doi.org/10.1016/S1076-6332(03)00671-8

50. De Santi, B., Kim, L., Bulthuis, R.F.G., Lucka, F., Manohar, S.: Automated three-dimensional image registration for longitudinal photoacoustic imaging (under review). J Biomed Opt. (2023)

51. Sitzmann, V., Martel, J.N.P., Bergman, A.W., Lindell, D.B., Wetzstein, G.: Implicit Neural Representations with Periodic Activation Functions.

52. THE PAMMOTH PROJECT AND 3RD GENERATION IMAGER, https://www.utwente.nl/en/tnw/m3i/research/Photoacoustic and Ultrasound Imaging (breast%2C liver%2C thyroid)/1.2 PAMMOTH/

53. Dantuma, M., Lucka, F., Kruitwagen, S.C., Javaherian, A., Alink, L., van Meerdervoort, R.P.P., Nanninga, M., Root, T.J.P.M.O. 't, De Santi, B., Budisky, J., Bordovsky, G., Coffy, E., Wilm, M., Kasponas, T., Aarnink, S.H., de Geus-Oei, L.F., Brochin, F., Martinez, T., Michailovas, A., Kobold, W.M., Jaros, J., Veltman, J., Cox, B., Manohar, S.: Fully three-dimensional sound speed-corrected multi-wavelength photoacoustic breast tomography. (2023)

54. Schoustra, S.M., op 't Root, T.J.P.M., Pompe van Meerdervoort, R.P., Alink, L., Steenbergen, W., Manohar, S.: Pendant breast immobilization and positioning in photoacoustic tomographic imaging. Photoacoustics. 21, 100238 (2021). https://doi.org/10.1016/J.PACS.2020.100238

55. MONAI Documentation: Rand3DElastic Transformation., https://docs.monai.io/en/stable/transforms.html#monai.transforms.Rand3DElastic

56. CPU Architecture, https://colab.research.google.com/github/d2l-ai/d2l-tvm-colab/blob/master/chapter_cpu_schedules/arch.ipynb#scrollTo=0FVSOz_AZ-Xb

57. Dunneram, Y., Greenwood, D.C., Cade, J.E.: Diet, menopause and the risk of ovarian, endometrial and breast cancer. Proceedings of the Nutrition Society. 78, 438–448 (2019). https://doi.org/10.1017/S0029665118002884

58. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A. V: VoxelMorph: A Learning Framework for Deformable Medical Image Registration. IEEE Trans Med Imaging. 38, 1788–1800 (2019). https://doi.org/10.1109/TMI.2019.2897538

59. Gehrung, M., Tomaszewski, M., McIntyre, D., Disselhorst, J., Bohndiek, S.: Co-registration of optoacoustic tomography and magnetic resonance imaging data from murine tumour models. Photoacoustics. 18, 100147 (2020). https://doi.org/10.1016/J.PACS.2019.100147

60. Zhang, S., Liang, Z., Tang, K., Li, X., Zhang, X., Mo, Z., Wu, J., Huang, S., Liu, J., Zhuang, Z., Qi, L., Chen, W.: In vivo co-registered hybrid-contrast imaging by successive photoacoustic tomography and magnetic resonance imaging. Photoacoustics. 31, 100506 (2023). https://doi.org/10.1016/J.PACS.2023.100506

61. Yu, Q., Liao, Y., Liu, K., He, Z., Zhao, Y., Li, F., Shan, T.: Registration of photoacoustic tomography vascular images: Comparison and analysis of automatic registration approaches. Front Phys. 10, 1045192 (2022). https://doi.org/10.3389/FPHY.2022.1045192/BIBTEX

62. Bise, R., Zheng, Y., Sato, I., Toi, M.: Vascular registration in photoacoustic imaging by low-rank alignment via foreground,background and complement decomposition. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 9902 LNCS, 326–334 (2016). https://doi.org/10.1007/978-3-319-46726-9_38/FIGURES/6

References

63. Riyahi, S., Choi, W., Liu, C.J., Zhong, H., Wu, A.J., Mechalakos, J.G., Lu, W.: Quantifying local tumor morphological changes with Jacobian map for prediction of pathologic tumor response to chemo-radiotherapy in locally advanced esophageal cancer. Phys Med Biol. 63, 145020 (2018). https://doi.org/10.1088/1361-6560/AACD22

64. Suh, J.W., Scheinost, D., Qian, X., Sinusas, A.J., Breuer, C.K., Papademetris, X.: SERIAL NONRIGID VASCULAR REGISTRATION USING WEIGHTED NORMALIZED MUTUAL INFORMATION. Proceedings / IEEE International Symposium on Biomedical Imaging: from nano to macro. IEEE International Symposium on Biomedical Imaging. 2010, 25 (2010). https://doi.org/10.1109/ISBI.2010.5490422

65. Kaise, H., Shimizu, F., Akazawa, K., Hasegawa, Y., Horiguchi, J., Miura, D., Kohno, N., Ishikawa, T.: Prediction of pathological response to neoadjuvant chemotherapy in breast cancer patients by imaging. Journal of Surgical Research. 225, 175–180 (2018). https://doi.org/10.1016/J.JSS.2017.12.002

66. Tudorica, A., Oh, K.Y., Y-C Chui, S., Roy, N., Troxell, M.L., Naik, A., Kemmer, K.A., Chen, Y., Holtorf, M.L., Afzal, A., Springer Jr, C.S., Li, X., Huang, W.: Early Prediction and Evaluation of Breast Cancer Response to Neoadjuvant Chemotherapy Using Quantitative DCE-MRI 1. Transl Oncol. 9, 8–17 (2016). https://doi.org/10.1016/j.tranon.2015.11.016

67. Iwasa, H., Kubota, K., Hamada, N., Nogami, M., Nishioka, A.: Early prediction of response to neoadjuvant chemotherapy in patients with breast cancer using diffusion-weighted imaging and gray-scale ultrasonography. Oncol Rep. 31, 1555 (2014). https://doi.org/10.3892/OR.2014.3025

68. Singh, G., Kumar, P., Parshad, R., Seith, A., Thulkar, S., Hosten, N.: Role of color Doppler indices in predicting disease-free survival of breast cancer patients during neoadjuvant chemotherapy. Eur J Radiol. 75, (2010). https://doi.org/10.1016/J.EJRAD.2009.12.027

69. Groheux, D., Biard, L., Giacchetti, S., Teixeira, L., Hindié, E., Cuvier, C., Vercellino, L., Merlet, P., De Roquancourt, A., De Cremoux, P., Resche-Rigon, M., Espié, M.: ^{18}F-FDG PET/CT for the Early Evaluation of Response to Neoadjuvant Treatment in Triple-Negative Breast Cancer: Influence of the Chemotherapy Regimen. J Nucl Med. 57, 536–543 (2016). https://doi.org/10.2967/JNUMED.115.163907

70. Parameters, https://elastix.lumc.nl/doxygen/parameter.html

71. Du, Y., Han, H.: Multiscale Approach for Bounded Deformation Image Registration. Fractal and Fractional 2022, Vol. 6, Page 681. 6, 681 (2022). https://doi.org/10.3390/FRACTALFRACT6110681

72. Xu, J., Chen, E.Z., Chen, X., Chen, T., Sun, S.: Multi-scale Neural ODEs for 3D Medical Image Registration. 12904 LNCS, 213–223 (2021). https://doi.org/10.1007/978-3-030-87202-1_21

10 Appendix

A. NAC efficacy studies with various imaging modalities

Table 12: NAC efficacy studies with various imaging modalities summary

Study	Modality	Population	When is the image taken?	Features	Results
[9]	Mammography	207 patients with stage II or III breast cancers	Two times: • Before NAC • One day before surgery after completion of NAC	• Microcalcifications	Of 11 patients with pCR, microcalcifications on post-NAC mammography were decreased in five patients (45.5 %) and stable in 6 patients (54.5 %). No patients with pCR had increased microcalcifications on post-NAC mammography.
[65]	MRI	107 breast cancer patients	Six times: • Before NAC • After administering fluorouracil, epirubicin (FEC), and cyclophosphamide, which were four cycles • After NAC	• Maximum tumor diameter	The tumor size was reduced by an average of 38% after FEC administration and 59% after completing NAC. The tumor size slightly increased in three patients who received the FEC regimen.
[66]	DCE-MRI	28 patients with grade 2 to 3 invasive breast tumors	Four times: • Before NAC • After the first cycle of NAC • At the midpoint of NAC (usually after 3 or 4 cycles of NAC or before the change of NAC agents) • After completion of NAC	• The longest diameter (LD) of the tumor Two pharmacokinetic values, which are • K^{trans}: contrast agent plasma/interstitium transfer rate constant • τi: mean intracellular water lifetime	The pCRs show a significant increase in τi and a decrease in K^{trans} compared to non-pCRs after the first NAC cycle. After the first cycle, K^{trans} and τi provide excellent (C > 0.9) early discrimination of pCR and non-pCR, and after the third cycle as well (0.8 < C < 0.9). However, LD is a poor early predictor of response after the first and second cycles (C < 0.7).

[67]	DWI-MRI	24 breast cancer patients	Two times: • Before NAC • After the first cycle of NAC	• Diffusion coefficient (ADC) values	The increase in mean ADC value was larger in responders than in non-responders. This was noted sooner than a reduction in tumor diameter. There was a significant correlation between the change in ADC values and response rate ($p=0.016$).
[67]	US	24 breast cancer patients	Two times: • Before NAC • After the first cycle of NAC	• Tumor size	A decrease in tumor size was noted, but the correlation between the change in tumor size and the response rate was low ($p=0.083$).
[68]	Color Doppler imaging	50 breast cancer patients	Two times: • Before NAC • After 2-4 cycles of NAC	• PSV: peak systolic velocity • $RI = \frac{PSV-EDV}{PSV}$ * EDV: end-diastolic velocity • $PI = \frac{PSV-EDV}{mean}$	Patients with increased PSV velocity after chemotherapy had a greater likelihood of recurrence and metastasis than patients in whom PSV decreased after chemotherapy. However, neither RI nor PI correlated with the clinical response.
[22]	Elastography	75 patients	Seven times: • Before biopsy • One day before each cycle of NAC (5 cycles) • Before surgery	• E_{max}: maximum elasticity • E_{mean}: mean elasticity • R: Strain ratio	E_{max}, E_{mean}, and R significantly declined among the different residual cancer burden groups, suggesting a correlation between the change in these parameters and response rate. Change in E_{max} before the second cycle showed the best predictive performance for assessing NAC efficacy.
[69]	PET	23 patients receiving eight cycles (conventional dose), 55 patients receiving six cycles (intensified dose)	Two times: • Before NAC • After the second cycle of NAC	• Change in SUV* after two cycles of NAC *SUV: [tracer concentration]/[injected activity]/[patient body weight]	Baseline tumor uptake was higher in patients who achieved pCR, and residual tumor uptake was lower in patients who achieved pCR. The decrease in tumor uptake before NAC and after the 2nd cycle of NAC was more pronounced in patients who achieved pCR. The decrease in tumor SUV_{max} was less pronounced in the conventional dose group compared to the intensified dose group.

Appendix

B. Elastix parameter settings

Table 13: Parameter setting used for image registration with Elastix in our comparison study

Rigid transformation	B-Spline transformation
(AutomaticParameterEstimation "true")	(AutomaticParameterEstimation "true")
(AutomaticScalesEstimation "true")	(CheckNumberOfSamples "true")
(CheckNumberOfSamples "true")	(DefaultPixelValue 0.000000)
(DefaultPixelValue 0.000000)	(FinalBSplineInterpolationOrder 3.000000)
(FinalBSplineInterpolationOrder 3.000000)	(FinalGridSpacingInPhysicalUnits 16.000000)
(FixedImagePyramid	(FixedImagePyramid
"FixedSmoothingImagePyramid")	"FixedSmoothingImagePyramid")
(ImageSampler "RandomCoordinate")	(GridSpacingSchedule 2.803221 1.988100 1.410000
(Interpolator "LinearInterpolator")	1.000000)
(MaximumNumberOfIterations 500)	(ImageSampler "RandomCoordinate")
(MaximumNumberOfSamplingAttempts 8.000000)	(Interpolator "LinearInterpolator")
(Metric "AdvancedNormalizedCorrelation")	(MaximumNumberOfIterations 1000)
(MovingImagePyramid	(MaximumNumberOfSamplingAttempts 8.000000)
"MovingSmoothingImagePyramid")	(Metric "AdvancedNormalizedCorrelation"
(NewSamplesEveryIteration "true")	"TransformRigidityPenalty")
(NumberOfResolutions 4.000000)	(Metric0Weight 1.000000)
(NumberOfSamplesForExactGradient 4096.000000)	(Metric1Weight 1.000000)
(NumberOfSpatialSamples 2048.000000)	(MovingImagePyramid
(Optimizer "AdaptiveStochasticGradientDescent")	"MovingSmoothingImagePyramid")
(Registration "MultiResolutionRegistration")	(NewSamplesEveryIteration "true")
(ResampleInterpolator "FinalBSplineInterpolator")	(NumberOfResolutions 4.000000)
(Resampler "DefaultResampler")	(NumberOfSamplesForExactGradient 4096.000000)
(ResultImageFormat "nii")	(NumberOfSpatialSamples 4096.000000)
(Transform "EulerTransform")	(Optimizer "AdaptiveStochasticGradientDescent")
(WriteIterationInfo "false")	(Registration
(WriteResultImage "true")	"MultiMetricMultiResolutionRegistration")
	(ResampleInterpolator "FinalBSplineInterpolator")
	(Resampler "DefaultResampler")
	(ResultImageFormat "nii")
	(Transform "BSplineTransform")
	(WriteIterationInfo "false")
	(WriteResultImage "true")

C. Synthetically deformed data set

i. Computation time

Table 14: Computation time for MUVINN and Elastix image registration in experiments with synthetically deformed data set

	Computation time MUVINN (seconds)		Computation time Elastix (seconds)	
Intensity	Volunteer A	Volunteer B	Volunteer A	Volunteer B
Low (5)	322	320	687	675
	325	324	690	678
	330	320	688	680
	327	320	685	673
	323	320	688	681
Medium (5)	324	318	687	673
	325	318	688	675
	326	318	687	683
	325	312	688	673
	325	319	688	675
High (5)	324	319	695	680
	322	321	686	676
	326	317	689	718
	325	322	687	677
	325	323	686	673

ii. Evaluation metrics

Table 15: The absolute values of PSNR, SSIM, and DSC before and after the image registration of synthetically deformed images with low deformation intensity

Volunteer A					Volunteer B				
Image pairs		PSNR	SSIM	DSC	Image pairs		PSNR	SSIM	DSC
1a	Before	31.540	0.56465	0.063629	1b	Before	32.238	0.58786	0.058591
	After MUVINN	34.134	0.74884	0.65099		After MUVINN	34.985	0.77831	0.60786
	After Elastix	33.263	0.70451	0.45607		After Elastix	34.602	0.77044	0.54614
2a	Before	30.673	0.49942	0.023989	2b	Before	32.358	0.59141	0.079907
	After MUVINN	33.724	0.73055	0.64656		After MUVINN	35.037	0.78132	0.61829
	After Elastix	32.832	0.68780	0.54148		After Elastix	34.430	0.77016	0.48820
3a	Before	31.498	0.56333	0.050856	3b	Before	31.597	0.54607	0.026822
	After MUVINN	34.126	0.74856	0.64445		After MUVINN	34.666	0.76584	0.60598
	After Elastix	33.409	0.71566	0.50616		After Elastix	33.780	0.73953	0.48768
4a	Before	30.774	0.50777	0.029074	4b	Before	30.305	0.46471	0.020067
	After MUVINN	33.766	0.73176	0.64654		After MUVINN	33.888	0.73126	0.60038
	After Elastix	33.252	0.71334	0.64671		After Elastix	32.699	0.68250	0.59148
5a	Before	31.179	0.52964	0.019481	5b	Before	31.913	0.55442	0.0278540
	After MUVINN	33.942	0.73814	0.62095		After MUVINN	34.861	0.77259	0.59581
	After	33.757	0.73567	0.62826		After	34.169	0.75436	0.48230

	Elastix					Elastix		
Mean	Before	31.133	0.53296	0.037406	Mean	Before	31.6822	0.548894	0.042648
	After MUVINN	33.938	0.73957	0.64190		After MUVINN	34.687	0.76586	0.60566
	After Elastix	33.303	0.71140	0.55574		After Elastix	33.936	0.74340	0.51916
Standard deviation	Before	0.35818	0.027192	0.016966	Standard deviation	Before	0.737947	0.04572	0.022898
	After MUVINN	0.17272	0.007889	0.010689		After MUVINN	0.41962	0.018091	0.0076070
	After Elastix	0.29779	0.015591	0.072292		After Elastix	0.67770	0.032531	0.043049

Table 16: The absolute values of PSNR, SSIM, and DSC before and after the image registration of synthetically deformed images with medium deformation intensity

Volunteer A		PSNR	SSIM	DSC	Volunteer B		PSNR	SSIM	DSC
Image pairs		PSNR	SSIM	DSC	Image pairs		PSNR	SSIM	DSC
6a	Before	31.211	0.52923	0.01161	6b	Before	30.154	0.45361	0.042668
	After MUVINN	33.868	0.73163	0.60144		After MUVINN	33.582	0.71362	0.55960
	After Elastix	33.193	0.69666	0.48820		After Elastix	31.968	0.63324	0.41337
7a	Before	30.516	0.48348	0.01212	7b	Before	31.051	0.49300	0.012582
	After MUVINN	33.515	0.71549	0.61117		After MUVINN	34.193	0.73772	0.54482
	After Elastix	32.523	0.66477	0.49747		After Elastix	33.183	0.68984	0.48091
8a	Before	31.028	0.51485	0.011198	8b	Before	31.210	0.49878	0.0093070
	After MUVINN	33.768	0.72654	0.59443		After MUVINN	34.254	0.73998	0.52859
	After Elastix	32.529	0.65673	0.33657		After Elastix	33.410	0.69896	0.50776
9a	Before	31.314	0.53675	0.013078	9b	Before	30.118	0.43937	0.018075
	After MUVINN	33.888	0.73206	0.58878		After MUVINN	33.627	0.71149	0.54793
	After Elastix	33.232	0.69925	0.47068		After Elastix	31.923	0.62782	0.38217
10a	Before	30.545	0.48265	0.0062790	10b	Before	31.134	0.49805	0.0097611
	After MUVINN	33.513	0.71478	0.59793		After MUVINN	34.257	0.74097	0.53651
	After Elastix	32.686	0.67338	0.53808		After Elastix	33.0478	0.68609	0.41158
Mean	Before	30.923	0.50939	0.010857	Mean	Before	30.733	0.47656	0.018479
	After MUVINN	33.710	0.72410	0.59875		After MUVINN	33.983	0.72876	0.54349
	After Elastix	32.833	0.67816	0.46620		After Elastix	32.706	0.66719	0.43916
Standard deviation	Before	0.33328	0.022620	0.002374	Standard deviation	Before	0.49049	0.025042	0.012491
	After MUVINN	0.16544	0.0075760	0.0074860		After MUVINN	0.30989	0.013287	0.010505
	After Elastix	0.31589	0.0170200	0.068487		After Elastix	0.63209	0.030273	0.047165

Appendix

Table 17: The absolute values of PSNR, SSIM, and DSC before and after the image registration of synthetically deformed images with high deformation intensity

Volunteer A					Volunteer B				
Image pairs		PSNR	SSIM	DSC	Image pairs		PSNR	SSIM	DSC
11a	Before	30.728	0.48109	0.0038246	11b	Before	31.729	0.52661	0.0048195
	After MUVINN	33.410	0.70351	0.53551		After MUVINN	34.379	0.74015	0.48437
	After Elastix	32.108	0.61408	0.33075		After Elastix	33.316	0.68105	0.32974
12a	Before	30.239	0.45933	0.0093357	12b	Before	30.111	0.43138	0.0047031
	After MUVINN	33.163	0.69208	0.53720		After MUVINN	33.515	0.69947	0.48218
	After Elastix	31.689	0.59952	0.32766		After Elastix	31.880	0.60497	0.36434
13a	Before	31.213	0.52296	0.0100506	13b	Before	30.467	0.45803	0.0030372
	After MUVINN	33.632	0.71533	0.53818		After MUVINN	33.606	0.70383	0.45301
	After Elastix	32.707	0.65628	0.36788		After Elastix	32.146	0.61968	0.342917
14a	Before	30.450	0.46677	0.0045249	14b	Before	30.981	0.48780	0.011353
	After MUVINN	33.300	0.69910	0.54604		After MUVINN	34.048	0.72748	0.52091
	After Elastix	31.889	0.61005	0.33094		After Elastix	32.514	0.64643	0.30814
15a	Before	30.243	0.46618	0.0080749	15b	Before	31.396	0.50927	0.0087115
	After MUVINN	33.170	0.69384	0.55025		After MUVINN	34.149	0.72557	0.47133
	After Elastix	31.882	0.62022	0.39222		After Elastix	32.892	0.65705	0.30987
Mean	Before	30.575	0.47927	0.0071620	Mean	Before	30.937	0.48262	0.0065250
	After MUVINN	33.335	0.70077	0.54144		After MUVINN	33.939	0.7193	0.48236
	After Elastix	32.055	0.62003	0.34989		After Elastix	32.550	0.64184	0.33100
Standard deviation	Before	0.36584	0.022967	0.0025300	Standard deviation	Before	0.59043	0.034358	0.003049
	After MUVINN	0.174280	0.0083170	0.0057020		After MUVINN	0.32870	0.015320	0.022241
	After Elastix	0.35195	0.019338	0.025827		After Elastix	0.51325	0.026983	0.021091

D. Operator's protocol

Measurement 1

1. Fit the breast in different cup sizes to see which cup fits the best with the volunteer's breast

2. Put the volunteer on the bed and place her breast correctly in the cup with the chosen cup size

3. Acquire a photoacoustic image of 720 nm and 870 nm simultaneously

4. When measuring is done, let the volunteer get off the imager completely by letting her stand up next to the imager

5. Repeat the process from step two until three images are acquired

6. Put the volunteer on the bed and place her breast correctly in the cup with the chosen cup size

7. Acquire a photoacoustic image of 720 nm

8. Repeat the process from step six once

*There is no need for waiting or warming up the volunteer. Getting off the imager completely and then repositioning is enough.

Measurement 2

1. Use the same cup size as measurement 1.

2. Put the volunteer on the bed and place her breast slightly downwards from the center of the cup so that there is a distance between the cup's edge and the breast's edge. When the breast is correctly positioned, there is no distance between the edge of the breast and the edge of the cup. See *Figure 21* for definitions of the edge of the cup and the edge of the breast. The distance between the cup's edge and the breast's edge can be chosen based on the assisting tool for the placement of the breast. The center point of the breast was placed on the stripe pointed by the orange arrow in *Figure 22* instead of the red point in the middle of the measuring tool like experiment 1.

3. Acquire a photoacoustic image of 720 nm

4. When measuring is done, let the volunteer get off the imager completely by letting her stand up next to the imager

5. Repeat the process once more

Figure 21: Correctly positioned breast and mispositioned breast

Figure 22: Measuring tool

Measurement 3

1. Choose the optimal cup size

2. Put the volunteer on the bed and place her breast correctly in the cup

3. Acquire a photoacoustic image of 720 nm

4. When measuring is done, let the volunteer get off the imager completely by letting her stand up next to the imager
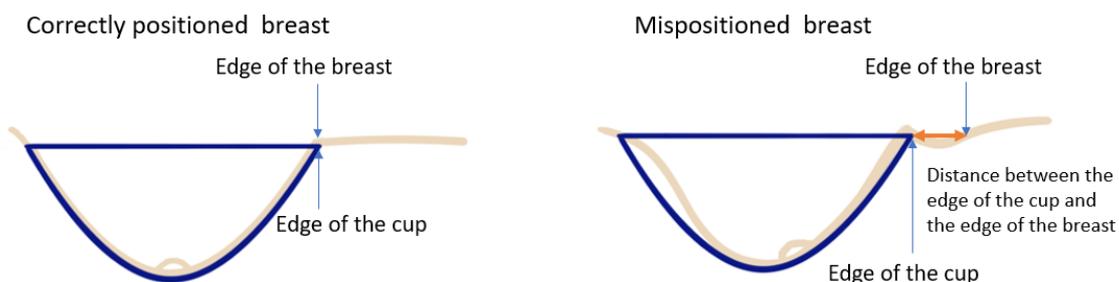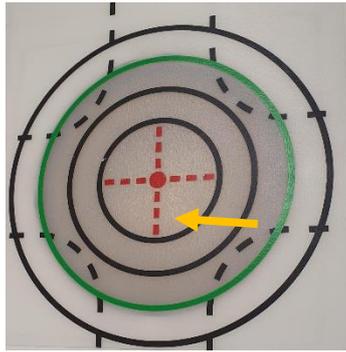
5. Repeat the process once more from steps 2 to 5

6. Change the cup size to the cup sizes in the *Table 18* and repeat the process from steps 2 to 5

Table 18: Number of acquisitions for each cup size used for experiment 4

Cup sizes	Number of acquisition
Optimal cup size (Size = 7)	2
Two sizes smaller than the optimal cup size (Size = 5)	2
Three sizes smaller than the optimal cup size (Size = 4)	2
One size larger than the optimal cup size (Size = 8)	2
No breast-supporting cup	1

Measurement time schedule

Measurement 1 - Correct repositioning:

- **A:** 3 repeated measurements, 720 and 870 nm: 160*3 = 480 seconds
- **B:** 2 repeated measurements, 720 nm: 120*2 = 240 seconds

Measurement 2 - Mispositioning:

- **C:** 2 repeated measurements, 720 nm: 120*2 = 240 seconds

Total time estimation for this breast: *16 minutes of exposure for this breast*

Appendix

1. The volunteer arrives in the room, coffee and explaining the device, the aim of the experiments and the protocol: **15 minutes**
2. Test different cup size on the volunteer and pick the best one: **5 minutes**
3. Volunteer positioning + laser glasses: **4 minutes**
4. Measurement A (160 seconds) + stand-up and lie again (3 minutes) + Measurement A (160 seconds) + stand-up and lie again (3 minutes) + Measurement A (160 seconds) + stand-up (2 minutes): **16 minutes**
5. Relax + change device settings: **5 minutes**
6. Measurement B (2 minutes) + stand-up and lie again (3 minutes) + Measurement B (2 minutes) + stand-up (2 minutes): **9 minutes**
7. Misposition patient (5 minutes) + Measurement C (2 minutes) + stand-up and misposition (6 minutes) + Measurement C (2 minutes) + stand-up (2 minutes)
8. Relax: **5 minutes**

Total: **59 minutes**

Measurement 3 - Different cup sizes (correct cup = 7):

- **D**: 2 repeated measurements, cup 7, 720 nm: 120*2 = 240 seconds

- **E**: 2 repeated measurements, cup 5, 720 nm: 120*2 = 240 seconds

- **F**: 2 repeated measurements, cup 4, 720 nm: 120*2 = 240 seconds

- **G**: 2 repeated measurements, cup 8, 720 nm: 120*2 = 240 seconds

- **H**: 1 measurement without a cup, 720 nm: 120 seconds

Total time estimation for this breast: *18 minutes of exposure for this breast*

1. Volunteer positioning with cup 7 + laser glasses: **4 minutes**
2. Measurement D (2 minutes) + stand-up and lie again (3 minutes) + Measurement D (2 minutes) + stand-up (2 minutes): **9 minutes**
3. Change cup: **3 minutes**
4. Measurement E (2 minutes) + stand-up and lie again (3 minutes) + Measurement E (2 minutes) + stand-up (2 minutes): **9 minutes**
5. Change cup: **3 minutes**
6. Measurement F (2 minutes) + stand-up and lie again (3 minutes) + Measurement F (2 minutes) + stand-up (2 minutes): **9 minutes**
7. Change cup: **3 minutes**
8. Measurement G (2 minutes) + stand-up and lie again (3 minutes) + Measurement G (2 minutes) + stand-up (2 minutes): **9 minutes**
9. Take off cup: **1 minute**
10. Measurement H (2 minutes): **2 minutes**
11. Wrapping up

Total: **52 minutes**

E. Serial image data set

i. Computation time

Table 19: Computation time for MUVINN and Elastix image registration in experiments with serial image data set

Image pairs	Computation time MUVINN (seconds)	Computation time Elastix (seconds)
01-01 – 02-01	1251	1131
01-01 – 03-01	1265	961
01-01 – 04-01	1259	980
01-01 – 05-01	1250	1013
01-01 – 02-02	1143	-
01-01 – 03-02	1195	-
01-01 – 06-01	1242	961
01-01 – 07-01	1269	948
08-01 – 11-01	1192	-
09-01 – 11-01	1178	-
08-01 – 12-01	1179	-
09-01 – 12-01	1181	-
08-01 – 13-01	1182	-
09-01 – 13-01	1184	-
08-01 – 14-01	1177	-
09-01 – 14-01	1163	-
08-01 – 15-01	1165	-
09-01 – 15-01	1172	-

ii. Evaluation metrics

Table 20: The absolute values of PSNR, SSIM, and DSC before and after the image registration in Experiment 1

Image pairs		PSNR	SSIM	DSC	NCC	TRE (mm) [mean, std]
01-01 – 02-01	Before	39.95	0.8903	0.3329	0.3649	[10.97, 3.78]
	After	44.02	0.9491	0.6369	0.7281	[0.89, 0.48]
01-01 – 03-01	Before	38.75	0.8492	0.1193	0.1372	[6.99, 2.34]
	After	44.40	0.9525	0.6717	0.7520	[0.9, 0.4]
01-01 – 04-01	Before	38.80	0.8501	0.1317	0.1419	[6.22, 1.79]
	After	44.80	0.9573	0.6930	0.7769	[0.64, 0.33]
01-01 – 05-01	Before	38.55	0.8432	0.1029	0.1288	[5.99, 1.23]
	After	44.85	0.9590	0.6919	0.7799	[0.63, 0.26]
Mean	Before	39.01	0.8582	0.1717	0.1932	7.54
	After	44.52	0.9545	0.6733	0.7592	0.765
Standard deviation	Before	0.5495	0.01872	0.09363	0.09924	2.01
	After	0.3395	0.003913	0.02270	0.02098	0.130

Table 21: The absolute values of PSNR, SSIM, and DSC before and after the image registration in Experiment 2

Image pairs		PSNR	SSIM	DSC	NCC	TRE (mm) [mean, std]
01-01 – 02-02	Before	39.95	0.8903	0.1037	0.3649	[10.97, 3.78]
	After	43.69	0.9454	0.4769	0.7054	[0.87, 0.49]
01-01 – 03-02	Before	38.75	0.8492	0.0815	0.1372	[6.99, 2.34]
	After	44.17	0.9502	0.5302	0.7374	[0.94, 0.45]
Mean	Before	39.35	0.8698	0.0926	0.2511	8.98
	After	43.93	0.9478	0.5036	0.7214	0.905
Standard deviation	Before	0.5995	0.02055	0.01110	0.1139	1.99
	After	0.2405	0.002400	0.02665	0.01600	0.0350

Table 22: The absolute values of PSNR, SSIM, and DSC before and after the image registration in Experiment 3

Image pairs		PSNR	SSIM	DSC	NCC	TRE (mm) [mean, std]
01-01 – 06-01	Before	37.96	0.8239	0.08669	0.1236	[24.98, 6.75]
	After	41.84	0.9156	0.4847	0.5779	[1.97, 2.62]
01-01 – 07-01	Before	38.27	0.8226	0.07674	0.1209	[33.31, 5.16]
	After	41.43	0.8989	0.4306	0.5084	[2.33, 2.25]
Mean	Before	38.12	0.8233	0.08172	0.1223	29.14
	After	41.64	0.9073	0.4577	0.5432	2.15
Standard deviation	Before	0.1550	0.0006500	0.004975	0.001350	4.17
	After	0.2045	0.008350	0.02705	0.03475	0.18

Table 23: The absolute values of PSNR, SSIM, and DSC before and after the image registration in Experiment 4

	Image pairs		PSNR	SSIM	DSC	NCC	TRE (mm) [mean, std]
Experiment 4a	08-01 – 11-01	Before	37.69	0.7831	0.04290	0.07383	[17.3, 3.71]
		After	42.82	0.9239	0.5325	0.6791	[0.82, 0.56]
	09-01 – 11-01	Before	37.67	0.7802	0.05107	0.07606	[18.21, 3.08]
		After	42.67	0.9209	0.5397	0.6716	[1.23, 0.46]
	Mean	Before	37.68	0.7817	0.04699	0.07495	17.76
		After	42.75	0.9224	0.5361	0.6754	1.03
	Standard deviation	Before	0.01000	0.001450	0.004085	0.001115	0.46
		After	0.07500	0.001500	0.003600	0.003750	0.21
Experiment 4b	08-01 – 12-01	Before	37.21	0.7560	0.03411	0.04903	[20.66, 2.78]
		After	42.44	0.9099	0.5014	0.6685	[1.32, 1.02]
	09-01 – 12-01	Before	37.19	0.7521	0.04912	0.03818	[18.13, 1.05]
		After	42.14	0.9051	0.4167	0.4924	[1.14, 1.42]
	08-01 – 13-01	Before	37.41	0.7564	0.03121	0.04771	[28.11, 5.25]
		After	41.79	0.8898	0.4720	0.6021	[1.45, 0.9]
	09-01 – 13-01	Before	37.37	0.7526	0.03171	0.04750	[21.36, 4.84]
		After	42.19	0.9013	0.5072	0.6418	[1.01, 0.64]
	Mean	Before	37.30	0.7543	0.03654	0.04561	22.065
		After	42.14	0.9015	0.4743	0.6012	1.23
	Standard deviation	Before	0.09631	0.001938	0.007347	0.004327	3.69
		After	0.2318	0.007424	0.03585	0.06711	0.17
Experiment 4c	08-01 – 14-01	Before	38.15	0.8166	0.05423	0.09962	[9.19, 2.26]
		After	43.40	0.9412	0.5971	0.7113	[1.25, 0.46]
	09-01 – 14-01	Before	38.09	0.8120	0.03639	0.09625	[15.87, 4.23]
		After	43.16	0.9384	0.5981	0.6985	[0.75, 0.45]
	08-01 – 15-01	Before	37.96	0.8133	0.06106	0.09543	[15.17, 3.23]
		After	42.27	0.9268	0.5416	0.6236	[0.81, 0.42]
	09-01 – 15-01	Before	37.92	0.8101	0.04462	0.09420	[14.79, 1.68]
		After	42.49	0.9293	0.5568	0.6487	[1.48, 1.56]
	Mean	Before	38.03	0.8130	0.04908	0.096375	13.76
		After	42.83	0.9339	0.5734	0.6705	1.07
	Standard deviation	Before	0.09354	0.002370	0.009367	0.002011	2.663
		After	0.4645	0.006023	0.02479	0.03579	0.3043

F. Comparison of the framework with and without a multiscale approach

As described in Chapter 4, *Algorithm: MUVINN*, MUVINN is based on implicit neural representations. The original implementation from reference [38] was adapted to work on photoacoustic images by implementing Frangi vesselness filtering and a coarse-to-fine strategy. There are several benefits of using a multiscale approach. Using a multiscale approach reduces the amount of data to be processed by starting at a lower resolution. Aligning the structures first at a course level provides a good starting point for the alignment of finer structures. This speeds up the registration process and is more computationally efficient. Without a multiscale approach, there is a larger chance that the algorithm might converge to a local minimum of the similarity measure, failing to reach the global minimum and leading to suboptimal registration [71, 72].

The original algorithm was applied to the data set from experiment 1 with images of normal repositioning using the same training parameters. For a fair comparison, the images were pre-processed using multiscale Frangi vesselness filtering following standard deviation values of (σ= {7,5,3,1.5}). The difference between the MUVINN and the original implementation is that the original implementation does not use a multiscale approach and that the patch length for the neighbouring points of the sampled point was fixed at 0.05 instead of decreasing the patch length when the scale is reduced.

The original implementation failed to align two of the four image pairs. The MIPs of the overlayed images before and after registration of the image pair in which registration failed are shown in *Figure 23.* The transformed image was highly irregular and not correctly aligned despite the decrease in the training loss.

Not implementing the multiscale approach showed that the network failed to align features of different scales at the same time in a consistent way, as it failed to align two out of four image pairs. There was a decrease in the training loss curve, so the algorithm converged to a local minimum probably. For this reason, a coarse-to-fine approach was implemented in MUVINN, where the network gradually optimizes structures at different scales. *Figure 23* shows the MIPs after registration with MUVINN for the same image pair which the original implementation of INR failed to align.
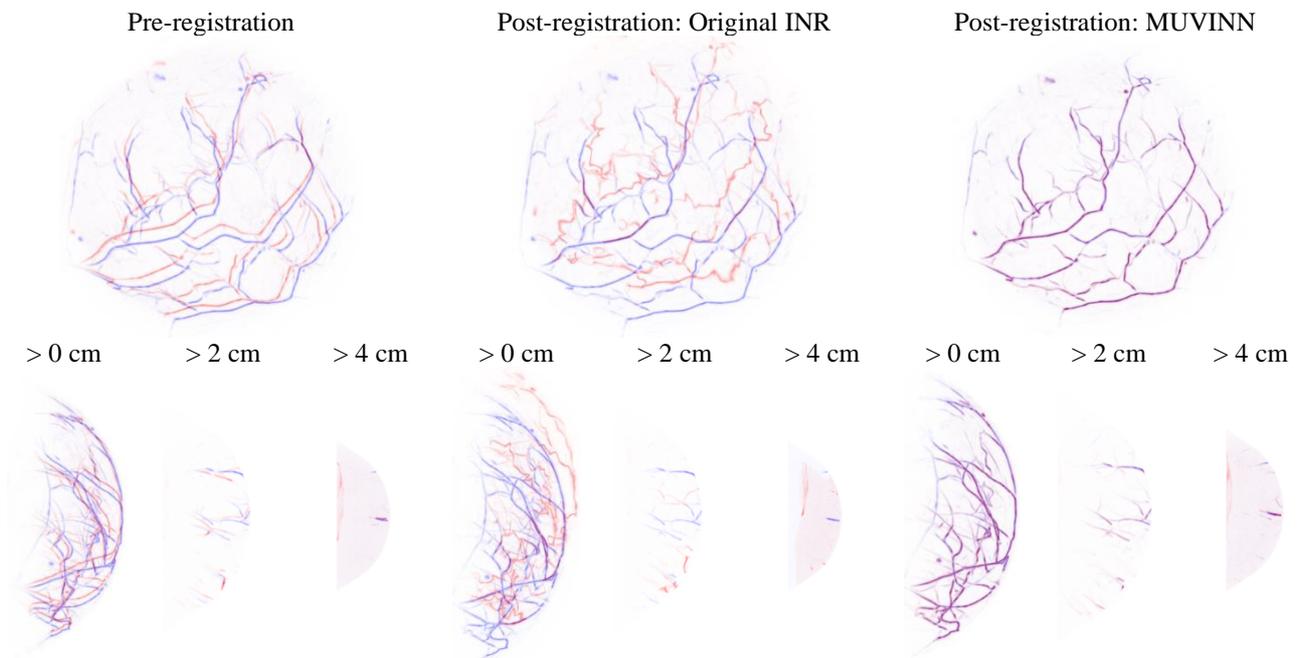
Pre-registration	Post-registration: Original INR	Post-registration: MUVINN

Figure 23: MIPs and depth images of before and after image registration of image pair 01-01 − 05-01 with the original implementation of INR and MUVINN