

Deformable multi-modal image registration for the correlation between optical measurements and histology images

Lianne Feenstra,^{a,b,*} Maud Lambregts^{a,b}, Theo J. M. Ruers,^{a,b}
and Behdad Dashtbozorg^a

^aNetherlands Cancer Institute, Image-Guided Surgery, Department of Surgical Oncology, Amsterdam, The Netherlands

^bUniversity of Twente, Department of Nanobiophysics, Faculty of Science and Technology, Enschede, The Netherlands

ABSTRACT. **Significance:** The accurate correlation between optical measurements and pathology relies on precise image registration, often hindered by deformations in histology images. We investigate an automated multi-modal image registration method using deep learning to align breast specimen images with corresponding histology images.

Aim: We aim to explore the effectiveness of an automated image registration technique based on deep learning principles for aligning breast specimen images with histology images acquired through different modalities, addressing challenges posed by intensity variations and structural differences.

Approach: Unsupervised and supervised learning approaches, employing the VoxelMorph model, were examined using a dataset featuring manually registered images as ground truth.

Results: Evaluation metrics, including Dice scores and mutual information, demonstrate that the unsupervised model exceeds the supervised (and manual) approaches significantly, achieving superior image alignment. The findings highlight the efficacy of automated registration in enhancing the validation of optical technologies by reducing human errors associated with manual registration processes.

Conclusions: This automated registration technique offers promising potential to enhance the validation of optical technologies by minimizing human-induced errors and inconsistencies associated with manual image registration processes, thereby improving the accuracy of correlating optical measurements with pathology labels.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JBO.29.6.066007](https://doi.org/10.1117/1.JBO.29.6.066007)]

Keywords: multi-modal image registration; histology; deformations; validation; registration algorithm; optical techniques

Paper 240113GR received Apr. 24, 2024; revised May 24, 2024; accepted May 29, 2024; published Jun. 12, 2024.

1 Introduction

Optical technologies revolutionized the field of oncologic surgery in recent years by providing non-invasive and innovative ways to monitor the assessment of resection margins during surgical procedures. By providing real-time visualization of tissue characteristics, optical technologies, such as diffuse reflectance spectroscopy (DRS),^{1,2} fluorescence lifetime imaging,³ and

*Address all correspondence to Lianne Feenstra, l.feenstra@nki.nl

hyperspectral imaging,^{4,5} can help to assess if all cancerous tissues are removed while minimizing damage to surrounding healthy tissues. This will lower the number of positive resection margins and thereby reduce the need for additional treatments, such as surgical re-excision or radiotherapy. However, the accuracy and reliability of these technologies are crucial to ensure safety and efficacy during surgical procedures. Therefore, validation of optical technologies is essential to establish their clinical significance and provide healthcare professionals with the confidence to use them in their practice. This involves assessing the clinical outcomes of these technologies with the current ground truth.⁶

Ground truth validation of optical tissue measurements refers to the process of comparing the acquired measurements with the gold standard histopathological analysis of tissue samples. This is provided by hematoxylin and eosin (H&E)-stained tissue sections, from which the measured tissue can be defined microscopically.⁷ To ensure an accurate correlation between the performed measurement locations and their corresponding H&E sections, it is essential to track the locations to locate them back microscopically.⁸ Subsequently, a registration between a specimen snapshot image (with tracked measurement locations) and its corresponding H&E sections (microscopic histology image) is necessary to validate the measured tissue types against ground truth. This process is crucial for establishing the reliability and reproducibility of optical techniques used for tissue diagnosis and especially for the accurate development of tissue classification algorithms.

Most of the time, the registration of a specimen snapshot image with the corresponding histology image encounters some challenges. The histopathological processing of tissue specimens, which involves steps such as fixation, dehydration, clearing, embedding, and cutting, causes tissue deformation in H&E sections. This deformation may include shrinkage, stretching, compression, tearing, and even loss of tissue.⁹ Also, other factors have an influence on deformation. For example, breast tissue, due to the presence of fat, is more prone to shrinkage or compression during processing compared with muscle tissue. Similarly, the size and thickness of tissue sections can affect the degree of tissue deformation, with thicker sections more susceptible to distortion. Besides, over-staining or prolonged staining influences the amount of deformation, while under-staining may lead to poor visualization of tissue structures.¹⁰ When validating optical measurements, it is crucial to take these tissue deformations into account.¹¹ Especially, when using labeled optical data for the development of machine learning models, incorrectly labeled data will influence the performance of tissue classification and ultimately impact clinical outcomes.

However, labeling and validating optical measurements with histopathology is a subject that has received limited attention in the existing research literature.¹²⁻¹⁴ In some studies, the aspect of tissue deformation is not even taken into consideration.^{15,16} In the method proposed by de Boer et al.,¹⁷ a manual point-based deformable registration between specimen snapshot and H&E sections is performed by looking for identical landmarks in both images. The proposed method addressed the need for a deformable registration method when labeling and validating optical measurements since it demonstrated a higher accuracy compared with the method that neglected such deformations. Besides, a manual point-based registration lacks objectivity since the identification of corresponding landmarks can vary among different users. This can lead to inconsistent results and reduce the reliability and accuracy of the registration. Also, this labor-intensive process can be time-consuming, particularly when dealing with large datasets. This approach is also not suitable for images with a limited number of paired distinguishable landmarks, which is often the case when registering multi-modal images.

In general, multi-modal image registration is a complex task that encounters various difficulties. One major challenge arises from the differences in intensity and contrast among images acquired in different modalities. These variations make it challenging to establish accurate paired landmarks among images. Another obstacle is the structural dissimilarity among modalities, resulting in differences in shape, size, and appearance of visible corresponding structures. Non-linear deformations and limited overlapping information further complicate the registration process.¹⁸ When registering specimen snapshots with histology images, microscopic artifacts such as tears, holes, and loss of tissue introduce additional complexities to accomplish an accurate registration. Overcoming these challenges involves the development of advanced algorithms capable of handling variations in intensity, contrast, shape, and deformations.

Automating the registration process, using advanced algorithms and computational techniques, shows potential to address current limitations and enhance the overall registration efficiency, thereby improving the accuracy of validating optical technologies.¹⁹ Therefore, the purpose of this study is to develop a two-dimensional (2D) multi-modal image registration model that is also able to compensate for tissue deformations automatically. The proposed approach is based on the VoxelMorph model, which has been adapted to the needs of a multi-modal 2D registration between specimen snapshot images and microscopic histology images. With this deformable multi-modal image registration model, we aim to achieve a faster and more precise method for labeling optical measurements with a ground truth, which can lead to a more accurate development of tissue classification algorithms, enhancing their practical use in clinical settings.

2 Material and Methods

2.1 Materials

The dataset used in this study consists of 113 breast tissue slices, each of which comprises three distinct images: a snapshot image of the breast tissue slice captured by a camera, a corresponding microscopic H&E histology image, and a manual registered histology image. This study was approved by the Institutional Review Board of NKI-AVL and registered under number IRBm 20-006. Example images of one tissue slice are demonstrated in Fig. 1. The manual registration is performed by manual selection of ~ 60 paired control points followed by a deformable registration using a nonrigid local weighted mean transformation, as described by de Boer et al.¹⁷ Despite the possibility of some misalignment and registration errors, we regard the obtained manual registered histology image in this study as a ground truth image.

2.2 Method

In this paper, a multi-modal image registration technique was developed, which is capable of automatically addressing tissue deformations, leading to a precise alignment of a snapshot image of a breast tissue slice and the corresponding histology image. The proposed multi-modal image registration methods are based on the VoxelMorph single-modality medical image registration framework, which will be explained in the Sec. 2.2.1. In this study, we intended to expand the application of VoxelMorph, as we are dealing with multi-modality images. The development of multi-modal deformable image registration involves a series of steps, beginning with dataset preparation. This is followed by two different deep-learning approaches for multi-modal image registration using unsupervised and supervised learning models, which will be evaluated separately.

2.2.1 VoxelMorph implementation

The VoxelMorph framework uses an unsupervised deep-learning model for deformable medical image registration. The model is initially designed to work with three-dimensional (3D) medical image volumes, such as magnetic resonance imaging or computer tomography scans, and can

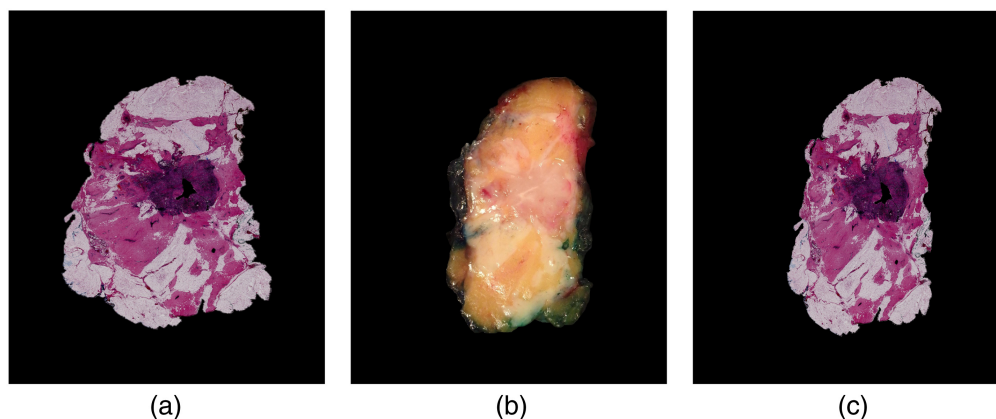


Fig. 1 Dataset example: (a) microscopic histology image, (b) snapshot image of the corresponding breast tissue slice captured by a camera, and (c) the manually registered histology image.

register two volumes of different shapes and sizes without requiring any explicit ground truth registration fields or anatomical landmarks.²⁰ The architecture of VoxelMorph is based on a deep convolutional neural network ($g_\theta(F, M)$), similar to the UNet model.²¹ The network uses a moving image M and fixed image F as input and computes a dense displacement field (DDF, φ) based on a set of learnable parameters θ . The network uses this set of parameters to compute the kernels of the convolutional layers and employs a spatial transformation function to evaluate the similarity between the predicted image ($M(\varphi)$) and the fixed image (F). This allows the model to refine its estimation of the optimal spatial transformation function and update its parameters.²² The generated DDF represents the displacement of each pixel in the moving image relative to the corresponding pixel in the fixed image. This dense map of vectors, with the same dimensions as the moving image, describes the spatial transformation required to align M with F and results in the predicted image ($M(\varphi)$).

The network is trained on an image dataset by minimizing the loss function (L) in each epoch, as described in Eq. (1)

$$L(F, M, \varphi) = L_{\text{sim}}(F, M, (\varphi)) + \lambda L_{\text{smooth}}(\varphi). \quad (1)$$

The loss function L consists of two components: L_{sim} penalizes the difference between the fixed (F) and moving (M) images, and L_{smooth} is a regularization on the dense deformation field (φ). The regularization parameter (λ) defines the weights of the two components. The VoxelMorph network is compatible with any differentiable loss function L .²³

2.2.2 Data preparation

Data augmentation is used to increase the number of images in the training set, as well as variations in deformations, which could improve the learning process of the network. Synthetic deformed images were generated from the existing dataset to simulate more deformation variations that occur during the pathology process. The augmented images were generated using randomly created DDFs, in which a number between -1 and $+1$ was generated for every pixel in both the x - and y -directions, resulting in displacement fields Δx and Δy . The Δx and Δy displacement are then convolved with a Gaussian filter with defined filter size F and standard deviation σ . Here, σ is serving as the elasticity coefficient. A scaling factor range α is then applied to the DDF to control the intensity of the deformation.²⁴ The deformation variables (σ , α , F) are chosen randomly within a specified range ([70 90], [11,000 13,000], and [350 450], respectively) based on the chosen level of deformation intensity, resulting in a total of 565 deformed specimen snapshots and histology images. Figure 2 illustrates some examples of artificial deformations for different deformation intensity levels.

Since the input of the VoxelMorph network consists of a two-channel image representing fixed F and moving M images, it is required to convert both RGB specimen snapshot and histology images into one-channel grayscale images. The weighted average of all color channels (red, green, blue) was used, and it determined the final grayscale representation [Eq. (2)], allowing for selective emphasis on certain colors and structures in the histology image

$$\text{Weighted average} = 0.299I_R + 0.587I_G + 0.114I_B, \quad (2)$$

where I denotes the intensity level of the color channels red (R), green (G), and blue (B) at each pixel in the image. To ensure a similar intensity level between both images, the specimen snapshot images were converted to grayscale using saturation values only (Fig. 3). This conversion method enhances the visual correspondence between connective and tumor tissues and is hypothesized to improve the performance of the model. At last, the computational effort and training time for the networks were reduced by resizing the histology and snapshot input images to 256×192 pixels.

2.2.3 Unsupervised learning model

In the unsupervised learning approach (Fig. 4), the input to the model comprises pairs of synthetic deformed histology images (F), which imitate the deformations during the pathology process, together with the snapshot specimen images (M). The trained network is similar to

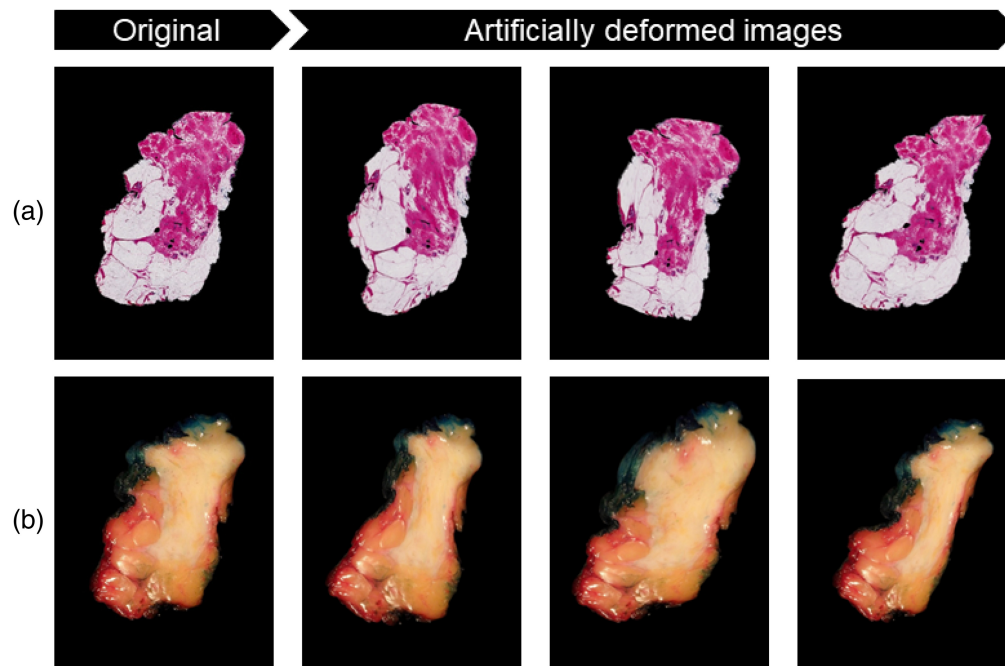


Fig. 2 Examples of synthetic deformation applied to histology (a) and specimen snapshot images (b). In this study, artificially deformed histology images were used for training the unsupervised model, whereas artificially deformed specimen snapshot images were used to train the supervised model.

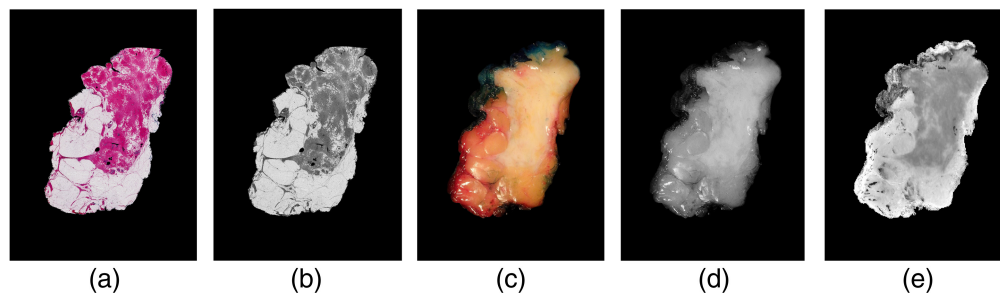


Fig. 3 Example of the preprocessing of the used input images: (a) original RGB histology image, (b) grayscale converted histology image, (c) original RGB specimen snapshot image, (d) grayscale converted specimen snapshot image, and (e) converted specimen snapshot image using saturation values only.

the original VoxelMorph model (as explained in Sec. 2.2.1) and entails training $g_{\theta}(F, M)$ using the input images F and M to compute optimal learnable parameters θ . M will be transformed using the estimated DDF (φ) in combination with a generated spatial transformer function, resulting in the predicted image ($M(\varphi)$).

In the unsupervised learning approach, the main objective is to align images from different modalities without access to ground truth labels. The input images have variations in intensity and structural visibility, so it cannot be assumed that the relationship between intensities in these two images is linear. Making use of the mean squared error (MSE) as a loss function, for example, is inappropriate. Therefore, mutual information (MI) is used as a loss function (L) to quantify the statistical dependence between the two images based on their joint distribution. MI measures the amount of information shared between the two images. In the context of the developed model, the goal is to find a deformation field that maximizes MI between the two input images. To compute MI, a histogram-based MI (HMI) was used, which computes the probability distribution of the intensity values between the two input images and estimates the joint probability distribution between their histograms.²⁵ Specifically, HMI is defined as

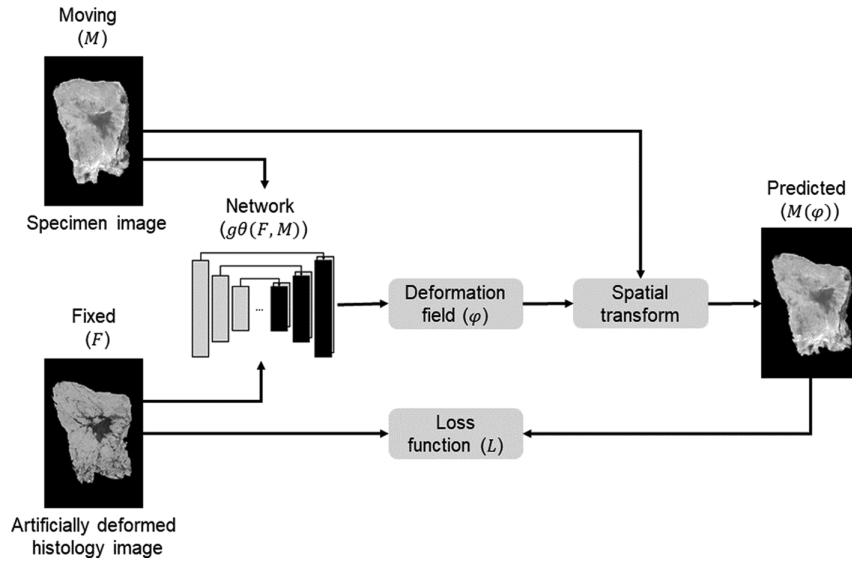


Fig. 4 Unsupervised learning model: the specimen snapshot image (M) and the artificially deformed histology image (F) are used as input images for the unsupervised deep convolutional neural network ($g_{\theta}(F, M)$). MI is used as a loss function (L). The network outputs a DDF (φ), which defines the mapping from moving image coordinates to the fixed image and is used to register M with F . This results in the predicted image ($M(\varphi)$).

$$\text{HMI}(F, M) = \sum_{i,j} p(i, j) \log \frac{p(i, j)}{p(i)p(j)}, \quad (3)$$

where $p(i, j)$ is the joint probability of the intensity values i and j in images F and M , and $p(i)$ and $p(j)$ are the marginal probabilities of intensity values i and j in images F and M , respectively. By replacing L_{sim} in Eq. (1) with HMI [Eq. (3)], the predicted image ($M(\varphi)$) was optimized by maximizing MI between F and M .

2.2.4 Supervised learning model

The VoxelMorph model was originally designed for unsupervised image registration, allowing it to learn without the need for ground truth labels. However, in this study, our dataset consists of manually registered histology images, which can be utilized to train the model in a supervised approach. To achieve this, the moving images (M) consist of artificially deformed snapshot specimen images. Consequently, the fixed images (F) consist of the manually registered histology images, whereas the ground truth labels (γ) include the original snapshot specimen images. The VoxelMorph network ($g_{\theta}(F, M)$) is trained by the loss function (L) to transform M to F using the predicted DDF (φ) in combination with the spatial transformer function. The modified model with an example of our data is illustrated in Fig. 5.

In the supervised learning approach, the valuable information of the previously manually registered histology images was used to measure the discrepancy between the predicted registration and the ground truth label. In this case, the MSE was used as a loss function as L_{sim} in Eq. (1), which is described in Eq. (4)²³

$$\text{MSE}(\gamma, M(\varphi)) = \frac{1}{n} \sum_{i=1}^n [\gamma_i - M(\varphi)_i]^2 \quad (4)$$

where n is the total number of samples, γ_n is the ground truth image for the n 'th sample (original snapshot specimen image), and $M(\varphi)_n$ the predicted value for the n 'th sample (predicted registered image). The MSE measures the average squared difference between the predicted registered image ($M(\varphi)$) and the ground truth image (γ). Since the ground truth and predicted images share the same modality and have similar intensity distributions and local contrast, the MSE is more

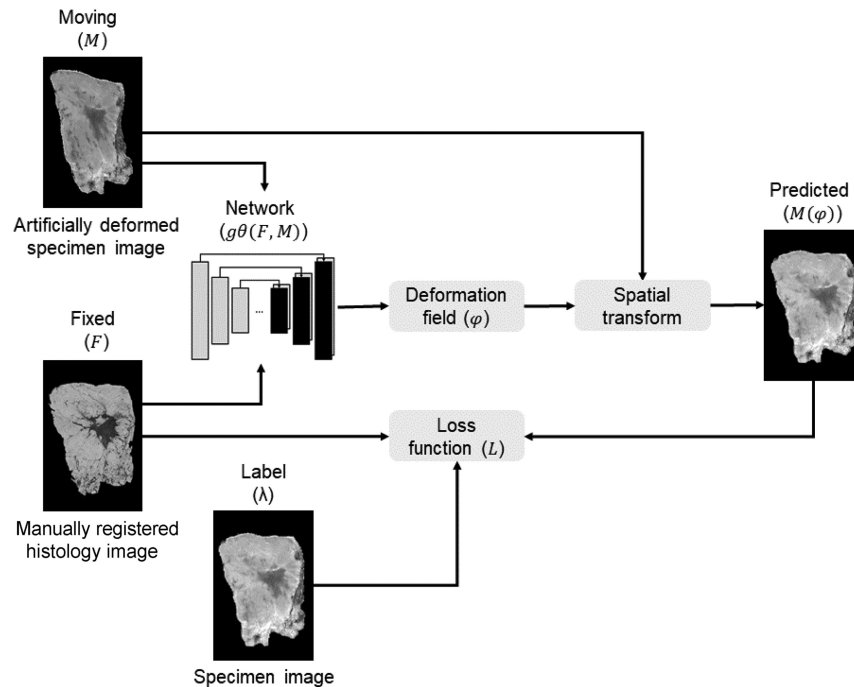


Fig. 5 Supervised learning model: the artificially deformed specimen snapshot image (M) and the manual registered histology image (F) are used as input images for the supervised deep convolutional neural network ($g_{\theta}(F, M)$). The specimen snapshot images are used as ground truth labels (γ). MSE is used as a loss function (L). The network outputs a DDF (φ), which defines the mapping from moving image coordinates to the fixed image and is used to register M with F . This results in the predicted image ($M(\varphi)$).

suitable for this context. It directly penalizes the pixel-wise differences, driving the model to produce outputs that closely match the ground truth.

2.2.5 Training

We utilized Python (version 3.10.4) along with the TensorFlow²⁶ and Keras²⁷ libraries for data manipulation and analysis. The augmented dataset was split into three subsets, whereas 360 paired deformed snapshot specimen images will be allocated for the training set, 90 paired images for the validation set, and 115 paired images for the test set. To train the network, the ADAM [79] optimizer with a learning rate of 0.001 was used. The configuration involved setting the number of epochs to 200, with 100 steps per epoch and a batch size of 16.

2.2.6 Evaluation matrices

To assess the performance of automatic deformable registration models described in Secs. 2.2.3 and 2.2.4, various evaluation metrics were employed. This evaluation was carried out for all images in the test set, both before and after applying the registration models. The Dice score was used to measure the degree of overlap between two binary images (A and B) by comparing the number of common pixels in the two images with the total number of pixels in the reference image (B), as described in Eq. (5). This metric is especially used to evaluate the overlap of the boundaries of the images.

$$\text{Dice}(A, B) = 2 \frac{|A \cap B|}{|A| + |B|}. \quad (5)$$

The *HMI* was used to measure the similarity between two images by comparing their histograms [Eq. (3)]. The *HMI* between two images is the amount of information shared between their

histograms. Specifically, it measures how much the joint histogram of the two images deviates from the product of their individual histograms. As a result, the optimal alignment of the two images can be determined.

For both obtained Dice and *HMI* metrics, statistical analysis was performed between the unregistered and registered results using IBM SPSS statistics v27 (SPSS Inc., Chicago, Illinois, United States). Statistical analysis for non-normally distributed data was performed using the Mann–Whitney U test, whereas a p -value ≤ 0.05 was considered statistically significant.

3 Results

3.1 Evaluation Unsupervised and Supervised Models

Figure 6 visualizes the Dice score and MI for the results of unsupervised and supervised approaches compared with the manual registration.

The violin plots in Fig. 6 show the distribution of evaluation metrics for the same pair of specimen images from the test set. In the case of the unsupervised and supervised approaches, metric values were calculated between the fixed (F) and predicted ($M(\varphi)$) images. For the manual registration, metric values between the manual registered histology image (F) and the specimen snapshot image (label γ) are reported. The width of these plots shows the relative frequency in which each value occurs, and it becomes wider when the value occurs more frequently and has a higher probability.

The distribution of Dice scores ranges from 0.90 to 0.99 (median 0.98 ± 0.02) and 0.77 to 0.92 (median 0.92 ± 0.04) for the unsupervised and supervised approaches, respectively. MI for the unsupervised method is distributed in a range between 0.45 and 1.05 (median 0.76 ± 0.18), which is slightly higher compared with the supervised method where the values are distributed between 0.44 and 0.94 (median 0.74 ± 0.13).

Figures 7 and 8 display multiple registration examples from the test set for the unsupervised and supervised approaches applied on the same paired specimen images.

4 Discussion

When achieving precise registration between (point-based) optical measurements and histopathology, the development of tissue classification algorithms can be optimized, thereby improving

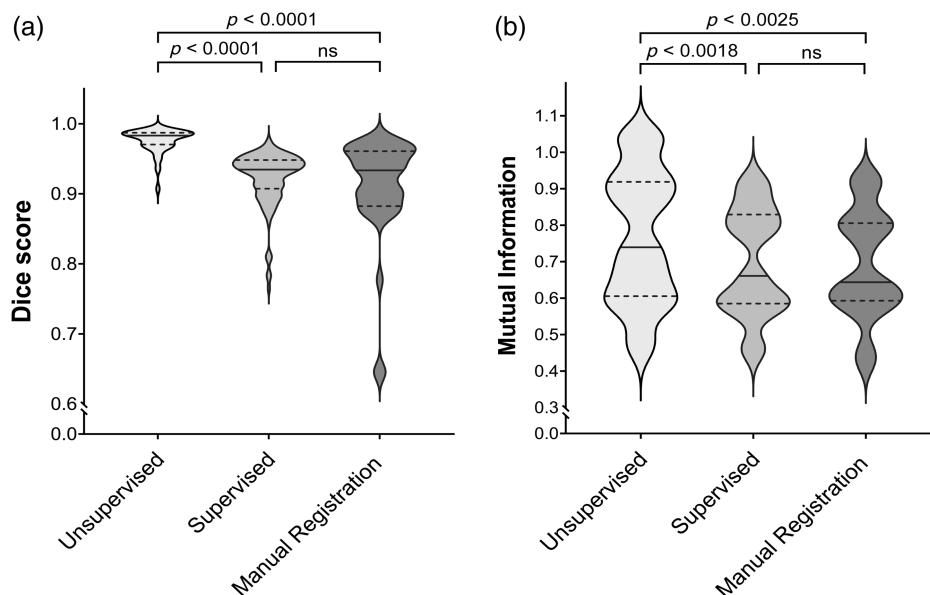


Fig. 6 Evaluation of the automatic deformable image registration method for the unsupervised, supervised, and manual approaches where (a) Dice score and (b) MI values are displayed for 115 specimen pairs after the registration. The solid line represents the median, whereas the dashed lines represent the interquartile range. ns, not significant.

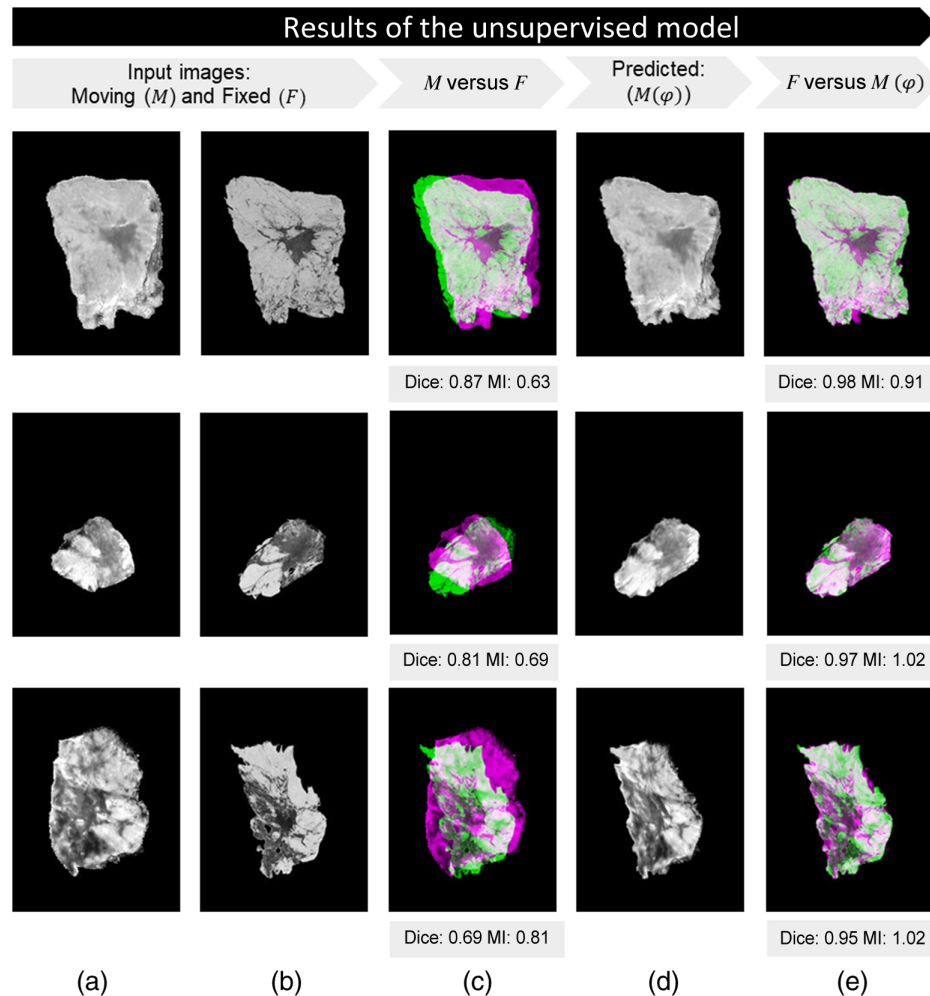


Fig. 7 Results of the unsupervised model: (a) specimen snapshot image (M); (b) artificially deformed histology image (F); (c) unregistered images: overlap between M and F ; (d) predicted image $M(\varphi)$; and (e) registered images: overlap between F and $M(\varphi)$. Dice and MI are shown for the unregistered and registered examples.

the effectiveness of optical technologies in clinical practice. However, registration difficulties arise with deformed multi-modality images, such as histology and tissue specimen images. Utilizing sophisticated algorithms and computational methods to optimize deformable registration processes holds promise in overcoming current inaccuracies in the validation of optical technologies. In this paper, we explored both unsupervised and supervised implementations based on the VoxelMorph model to achieve deformable registration among 2D multi-modal images. We aimed to develop a faster and more accurate registration method for eventually labeling optical measurements with ground truth. We used a previously acquired in-house dataset of manually registered breast specimen images to train the models.

The efficacy of the developed models was assessed through the computation of both Dice scores and MI for all 115 registered images, the overlap between F and $M(\varphi)$, in the test set (Fig. 6). The unsupervised method outperformed the other approaches significantly. Specifically, as indicated by the Dice score, a more accurate overlap between the general shapes of the input images was achieved. MI functioned as a metric for assessing the similarity among distinct image modalities. As illustrated by the violin plot, the unsupervised dataset demonstrated a prominently increased distribution within the 0.6 to 1.0 range, indicating an improved alignment of internal structures compared with the other approaches. Unlike mono-modal image registration, where the ground truth transformation is available, multi-modal images do not have a direct one-to-one correspondence due to differences in imaging modalities. This makes it difficult to define an

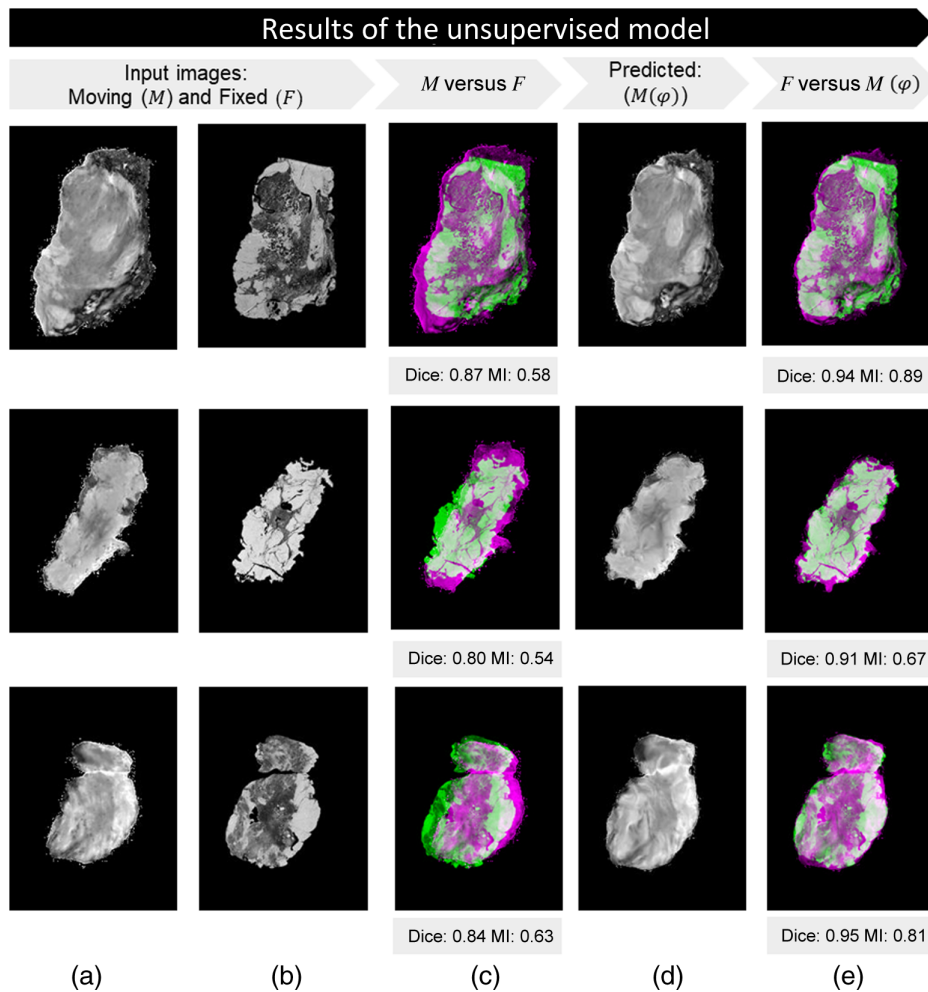


Fig. 8 Results of the supervised model: (a) artificially deformed specimen snapshot image (M); (b) the manual registered (ground truth) histology image (F); (c) unregistered images: overlap between M and F ; (d) predicted image $M(\varphi)$; and (e) registered images: overlap between F and $M(\varphi)$. Dice and MI are shown for the unregistered and registered examples.

objective reference for evaluating the accuracy of registration. The dataset used in this study is unique since it contains manually registered ground truth and histology images, which are not commonly available in similar datasets. This makes the dataset particularly well suited for developing and testing multi-modal image registration algorithms and other image analysis techniques. However, training a model in a supervised manner, with labels derived from manually registered ground truth images, showed only a slight improvement in comparison to the manual registration approach. This can be explained due to the fact that this model was trained with images that possibly contain small manual registration errors.

Besides the use of labels, the main difference between the training of supervised and unsupervised models also involves the use of loss function (L). The choice of loss function significantly impacts the performance and applicability of the image registration model. In our study, the MI-based loss function (used in the unsupervised model) demonstrated superior performance compared with the MSE, as demonstrated in Fig. 6. MI is useful for multi-modal registration because it focuses on shared information between images rather than intensity similarity, making it robust to variations in intensity and contrast. However, it is computationally intensive due to the need to estimate joint probability distributions, and it can be less intuitive and harder to optimize than simpler loss functions such as MSE. MSE is easy and efficient to compute for linear relationships between images, provides a direct measure of alignment error when ground truth labels are available, and is straightforward to interpret and optimize.

The adoption of algorithms for automating the deformable registration processes represents a paradigm shift in image registration, offering distinct advantages over manual methods.^{13,14,17} Our results demonstrate that the unsupervised algorithm achieved superior significant performance when compared with the ground truth manual point-based registration, which emphasizes the use of applicability of computational techniques for multi-modal image registration. In comparison, manual registration methods, including the corresponding pre-processing steps, can be prone to human errors and inconsistencies, making the automated approach a significantly more reliable option. Recently, there has been a growing acknowledgment among studies regarding the essential requirement to account for tissue deformations when correlating optical measurements with a ground truth pathology label.^{15,28} Multi-modal registration is often complicated by the lack of corresponding landmarks between images. Therefore, the use of fiducial markers is investigated but involves invasive procedures, such as the placement of burn marks on the tissue surface, that could potentially inflict damage on delicate tissue structures.¹¹ Besides, manual tasks are characterized by their labor-intensive nature, which demands considerable time to ensure accurate alignment. By contrast, the inherent efficiency of automatic registration accelerates the alignment process, minimizing the potential for discrepancies and enhancing the overall quality of results.

While advancements in registration algorithms have significantly improved the accuracy and robustness of image alignment, the selection of appropriate evaluation metrics remains a challenging and nuanced task. The complexities inherent to multi-modal registration pose a range of difficulties in identifying evaluation metrics that accurately assess the quality of registration outcomes. Multi-modal registration often involves non-linear transformations to account for differences in anatomical structures and intensities across modalities. Conventional metrics such as MSE or MI, which are effective for linear transformations, may inadequately capture the intricate deformations and intensity variations inherent to multi-modal registration. Our findings indicate precise registration that effectively compensates for deformation, not only for the entire shape but also for the internal structures. However, this achievement is not reflected correctly in the calculated MI values since this is influenced by the inherent variations in contrast among multi-modal images and the employed preprocessing procedures of the images. Additional metrics, such as target registration error, could be considered to provide a better assessment of the model's performance.

The complex task of registering microscopic histology images with their corresponding tissue slices in RGB encounters challenges arising from the fundamental differences among these imaging modalities. Microscopic histology images, revealing details at the cellular level, are typically acquired through staining and specialized imaging techniques. By contrast, RGB images offer a macroscopic perspective of tissue slices under conventional optics, capturing color information at a larger scale. The presence of tears and holes disrupts the natural continuity of cellular structures in histology images, introducing gaps and inconsistencies that make the registration process challenging. Developed registration techniques therefore struggle to establish trustworthy correspondences among regions that are distorted by these artifacts. Tears introduce non-local deformations, while holes disrupt the continuity of anatomical features, making it challenging for algorithms to accurately match corresponding areas in RGB images. Therefore, it is essential to acknowledge that the suboptimal performance of this developed model can, at times, be influenced by the degree of deformation and the presence of artifacts in the histology images.

Besides, another issue during the preparation of microscopic H&E sections can arise. The surface of the processed tissue is often not completely flat when embedded in paraffin. To compensate, the tissue block is trimmed until a complete tissue section is visible. However, this process can result in sections being taken from deeper parts of the specimen, along planes that do not necessarily correspond to the 2D surfaces of the original tissue. While the proposed deformable 2D registration approach works well for the validation of optical techniques with multi-millimeter probed depth, such as DRS, it may pose challenges and inaccuracies for superficial imaging methods. In such cases, processing tissue on a flat surface could result in more accurate 3D registration. However, this method has its limitations, particularly the need for a different specimen fixation approach than conventional methods, which may not always be feasible, even in research settings.

The presented approach has the potential to optimize the registration efficiency, for breast tissue specifically, ultimately leading to an enhancement in the precision of correlating optical

measurements with a correct pathology label used for the development of tissue classification algorithms. Further research should also focus on exploring the suitability of this developed model for deformation problems in histology images, which occur across different tissue types. This can potentially result in improved optical technology validation for multiple organ domains, which ensures the dependable integration of optical technologies into clinical practice.

5 Conclusion

Our efforts have resulted in the development of an automated multi-modal image registration technique based on deep learning principles. This method effectively aligns snapshot breast specimen images with corresponding histology images, achieving a high degree of precision. Notably, the performance of the unsupervised model exceeds that of a previous manual approach, presenting a faster and significantly more accurate registration method. This advancement holds the promise of improving the validation of optical technologies across diverse organ domains, ensuring the reliable integration of optical tools into clinical practice.

Disclosures

The authors have no relevant financial interests in this article and no potential conflicts of interest to disclose.

Code and Data Availability

Data and software underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Acknowledgments

The authors would like to thank all surgeons and nurses from the Department of Surgery, all pathologist assistants from the Department of Pathology for their assistance in processing specimens, the NKI-AVL Core Facility Molecular Pathology & Biobanking (CFMPB) for supplying NKI-AVL biobank material, and all students that participated in this research for their time and effort. Research at the Netherlands Cancer Institute is supported by institutional grants from the Dutch Cancer Society and the Dutch Ministry of Health, Welfare, and Sport. The authors gratefully acknowledge the financial support to this research by the Dutch Cancer Society (Grant No. KWF 13443). The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of The Netherlands Cancer Institute (IRBm 20-006) for studies involving humans.

References

1. L. L. de Boer et al., "Towards the use of diffuse reflectance spectroscopy for real-time in vivo detection of breast cancer during surgery," *J. Transl. Med.* **16**, 1–14 (2018).
2. D. Veluponnar et al., "Diffuse reflectance spectroscopy for accurate margin assessment in breast-conserving surgeries: importance of an optimal number of fibers," *Biomed. Opt. Express* **14**(8), 4017–4036 (2023).
3. A. Alfonso-Garcia et al., "Mesoscopic fluorescence lifetime imaging: fundamental principles, clinical applications and future directions," *J. Biophotonics* **14**(6), 202000472 (2021).
4. L.-J. S. Jong et al., "Discriminating healthy from tumor tissue in breast lumpectomy specimens using deep learning-based hyperspectral imaging," *Biomed. Opt. Express* **13**, 2581 (2022).
5. E. Kho et al., "Feasibility of ex vivo margin assessment with hyperspectral imaging during breast-conserving surgery: from imaging tissue slices to imaging lumpectomy specimen," *Appl. Sci.* **11**(19), 8881 (2021).
6. B. C. Wilson, M. Jermyn, and F. Leblond, "Challenges and opportunities in clinical translation of biomedical optical spectroscopy and imaging," *J. Biomed. Opt.* **23**, 030901 (2018).
7. W. A. Wells et al., "Validation of novel optical imaging technologies: the pathologists' view," *J. Biomed. Opt.* **12**, 051801 (2007).
8. L. Feenstra et al., "Point projection mapping system for tracking, registering, labeling, and validating optical tissue measurements," *J. Imaging* **10**(2), 37 (2024).
9. B. Pritt et al., "The effect of tissue fixation and processing on breast cancer size," *Hum. Pathol.* **36**(7), 756–760 (2005).
10. E. McInnes, "Artefacts in histopathology," *Comp. Clin. Pathol.* **13**, 100–108 (2005).
11. J. Unger et al., "Method for accurate registration of tissue autofluorescence imaging data with corresponding histology: a means for enhanced tumor margin assessment," *J. Biomed. Opt.* **23**, 1 (2018).

12. V. Naranjo et al., “Stained and infrared image registration as first step for cancer detection,” in *IEEE-EMBS Int. Conf. Biomed. and Health Inf. (BHI)*, pp. 420–423 (2014).
13. M. Halicek et al., “Deformable registration of histological cancer margins to gross hyperspectral images using demons,” *Proc. SPIE* **10581**, 105810N (2018).
14. G. Lu et al., “Hyperspectral imaging for cancer surgical margin delineation: registration of hyperspectral and histological images,” *Proc. SPIE* **9036**, 90360S (2014).
15. J. E. Phipps et al., “Automated detection of breast cancer in resected specimens with fluorescence lifetime imaging,” *Phys. Med. Biol.* **63**(1), 015003 (2018).
16. A. M. Laughney et al., “Scatter spectroscopic imaging distinguishes between breast pathologies in tissues relevant to surgical margin assessment,” *Clin. Cancer Res.* **18**, 6315–6325 (2012).
17. L. L. De Boer et al., “Method for coregistration of optical measurements of breast tissue with histopathology: the importance of accounting for tissue deformations,” *J. Biomed. Opt.* **24**, 075002 (2019).
18. F. Alam and S. U. Rahman, “Challenges and solutions in multimodal medical image subregion detection and registration,” *J. Med. Imaging Radiat. Sci.* **50**(1), 24–30 (2019).
19. X. Pei et al., “A review of the application of multi-modal deep learning in medicine: bibliometrics and future directions,” *Int. J. Comput. Intell. Syst.* **16**, 1–20 (2023).
20. G. Balakrishnan et al., “VoxelMorph: a learning framework for deformable medical image registration,” *IEEE Trans. Med. Imaging* **38**(8), 1788–1800 (2019).
21. O. Ronneberger, P. Fischer, and T. Brox, “U-net: convolutional networks for biomedical image segmentation,” *IEEE Access* **9**, 16591–16603 (2015).
22. M. Jaderberg et al., “Spatial transformer networks,” in *Adv. Neural Inf. Process. Syst.*, 2015-January, pp. 2017–2025 (2015).
23. G. Balakrishnan et al., “An unsupervised learning model for deformable medical image registration,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recognit.*, pp. 9252–9260 (2018).
24. P. Y. Simard, D. Steinkraus, and J. C. Platt, “Best practices for convolutional neural networks applied to visual document analysis,” in *Proc. Int. Conf. Doc. Anal. and Recognit., ICDAR 2003-January (May 2014)*, pp. 958–963 (2003).
25. J. P. Pluim, J. B. Maintz, and M. A. Viergever, “Mutual-information-based registration of medical images: a survey,” *IEEE Trans. Med. Imaging* **22**(8), 986–1004 (2003).
26. M. Abadi et al., “TensorFlow: large-scale machine learning on heterogeneous distributed systems,” <https://arxiv.org/abs/1603.04467> (2016).
27. “Github - keras-team/keras: Deep learning for humans,” <https://github.com/keras-team/keras> (accessed 03/28/2024).
28. Y. Zhang et al., “Explainable liver tumor delineation in surgical specimens using hyperspectral imaging and deep learning,” *Biomed. Opt. Express* **12**(7), 4510–4529 (2021).

Biographies of the authors are not available.