

GuidedMorph: Two-Stage Deformable Registration for Breast MRI

Yaqian Chen, Hanxue Gu, Haoyu Dong, Qihang Li, Yuwen Chen, Nicholas Konz, Lin Li, Maciej A. Mazurowski

Abstract— Accurately registering breast MR images from different time points enables the alignment of anatomical structures and tracking of tumor progression, supporting more effective breast cancer detection, diagnosis, and treatment planning. However, the complexity of dense tissue and its highly non-rigid nature pose challenges for conventional registration methods, which primarily focus on aligning general structures while overlooking intricate internal details. To address this, we propose GuidedMorph, a novel two-stage registration framework designed to better align dense tissue. In addition to a single-scale network for global structure alignment, we introduce a framework that utilizes dense tissue information to track breast movement. The learned transformation fields are fused by introducing the Dual Spatial Transformer Network (DSTN), improving overall alignment accuracy. A novel warping method based on the Euclidean distance transform (EDT) is also proposed to accurately warp the registered dense tissue and breast masks, preserving fine structural details during deformation. The framework supports paradigms that require external segmentation models and with image data only. It also operates effectively with the VoxelMorph and TransMorph backbones, offering a versatile solution for breast registration. We validate our method on ISPY2 and internal dataset, demonstrating superior performance in dense tissue, overall breast alignment, and breast structural similarity index measure (SSIM), with notable improvements by over 13.01% in dense tissue Dice, 3.13% in breast Dice, and 1.21% in breast SSIM compared to the best learning-based baseline.

Index Terms— Image registration, Deep learning, Breast MRI

I. INTRODUCTION

In recent decades, breast MRI has become a critical tool in the early detection of breast cancer [1], particularly for high-risk populations [2]. Its superior sensitivity and level of detail compared to traditional imaging methods such as mammography and ultrasound enable it to capture sufficient information

Yaqian Chen, Hanxue Gu, Haoyu Dong, Yuwen Chen, and Nicholas Konz are with the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27703, USA

Qihang Li is with the Department of Biostatistics & Bioinformatics, Duke University, Durham, NC 27703, USA

Lin Li was with the Department of Biostatistics & Bioinformatics, Duke University, Durham, NC 27703, USA. Now, Lin Li is with the Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, Gainesville, FL 32611

Maciej A. Mazurowski is with the Departments of Electrical and Computer Engineering, Radiology, Biostatistics & Bioinformatics, and Computer Science, Duke University, Durham, NC 27703, USA

to identify early-stage cancers that other modalities might miss [3]–[6]. While breast MRI excels in sensitivity for early cancer detection, it suffers from high false-positive rates, particularly in initial screenings, necessitating several subsequent scans for accurate assessment [7].

Thus, during diagnostic assessments, it is common for radiologists to track certain changes in the breast over time through sequential scans. In this context, dense tissue alignment is particularly important due to its clinical significance [8]–[10]; rapid changes or growth in these regions are often associated with a heightened risk of developing breast cancer [11], [12]. Consequently, accurately registering dense tissue across MRI scans is essential for capturing subtle transformations that may signal the onset or progression of malignancy.

Despite the wide array of medical image registration techniques currently available, achieving accurate breast MRI registration—especially the precise alignment of dense tissue—continues to be a significant challenge [13], [14]. This difficulty stems largely from the complexity of dense tissue, which can vary considerably (as shown in Fig. 1), and the high non-rigidity of the breast (as opposed to other commonly registered modalities like brain MRI [15], [16]).

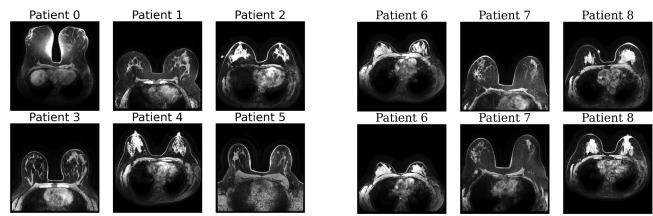


Fig. 1. **Inter-patient and intra-patient variability in breast MRI scans.** Middle slices from different MRI volumes illustrate differences across patients (*left*) and within the same patient (*right*).

In this paper, we propose a two-stage registration framework, **GuidedMorph**, that is particularly designed for the challenging tasks such as breast MRI registration. In addition to the structure of VoxelMorph [17] and its variants [18], [19], our method introduces a dedicated step to improve alignment in information-rich regions, such as dense tissue. It also integrates a publicly available VNet-based segmentation model for dense tissue and breast boundaries [20], [21], using pre-trained weights that require no fine-tuning or extra annotations. To further simplify the pipeline, we propose a non-learning **Dense Tissue Extraction** strategy that identifies complex dense

tissue regions based on Frangi vesselness filter. Furthermore, with the observation that traditional warping methods tend to lose fine structural details (See 11) when applied to binary masks, especially for sparse or anatomically complex regions such as dense tissue, we propose a more robust alternative. A novel warping method based on the Euclidean distance transform (EDT) is proposed, which converts binary masks into continuous-valued distance maps prior to deformation. The novel transformation method consistently demonstrates performance improvements by preserving structural integrity.

The main contributions of this work are summarized as follows:

- We propose GuidedMorph, a novel dual-scale alignment architecture, combining a global structure alignment network with a dense tissue tracking module that utilizes the Dual Spatial Transformer Network to improve the fidelity of non-rigid dense tissue.
- This registration pipeline is flexible, supporting both scenarios where external segmentation masks are available and cases where only image data is provided.
- A new detail-preserving mask warping algorithm is proposed, utilizing the EDT to preserve fine structural details during mask deformation. This approach is significantly effective in preserving thin and fine dense tissue structures, leading to a 2.44% improvement in performance compared to conventional mask warping.
- Compared to the best learning-based baseline, our method achieves the state-of-the-art, improving the performance by over 13.01% in dense tissue Dice, 3.13% in breast Dice, and 1.21% in breast SSIM on both internal and external datasets.

We will release our code upon paper acceptance.

II. RELATED WORK

A. Medical Image Registration

Medical image registration has increasingly attracted attention due to its essential role in aligning anatomical structures across different imaging modalities, time points, or patients, enabling tasks such as longitudinal disease tracking, atlas construction, and image-guided interventions [22]–[24]. Traditional registration approaches are broadly categorized into rigid, affine, and deformable methods [25]. Rigid and affine registrations, which are simpler in nature, are well-supported by numerous existing tools, such as Automated Image Registration (AIR) [26], Correspondence of Closest Gradient Voxels (COCGV) [27], and Advanced Normalization Tools (ANTs) [28]. However, the assumption of rigidity can be limiting for realistic medical image registration, particularly for breast tissue. To mitigate this limitation, recent medical image registration research has shifted toward deformable registration algorithms [29], [30], which model non-linear anatomical variability and enable finer alignment, particularly in soft-tissue applications such as breast MRI [17]–[19]. Early deformable methods relied on optimization-based strategies [31], [32], conducting optimization iteratively for every registration pair. In order to shorten the inference time, recent studies such as VoxelMorph [17] and TransMorph [18] have

demonstrated promising results by accelerating registration through unsupervised training paradigms.

B. Non-Deep Learning Medical Image Registration Methods

Traditional non-deep learning-based deformable registration algorithms typically optimize a similarity metric between fixed and moving images through iterative transformation updates via three main components: a deformation model, an objective function, and an optimization method [29]. The deformation model defines the type of transformation applied to m , ranging from physical models such as elastic body models [33]–[38], viscous fluid flow models [39]–[42], and diffusion models [43]–[46] to models based on interpolation theory [47]–[50]. The objective function typically includes similarity metrics such as Cross-Correlation (CC) [51], Mean Squared Error (MSE) [52], and Mutual Information (MI) [53], as well as a regularization term used to balance the accuracy of the transformation with the smoothness of the deformation field. The optimization strategy varies depending on the number of degrees of freedom of the deformation model, which can range from six for global rigid transformations to millions for non-parametric dense transformations. The increased dimensionality of the state space enriches the descriptive power of the model, but also necessitates more complex and time-consuming optimization algorithms [29]. This trade-off between model complexity and computational efficiency is a key consideration in non-learning-based registration approaches [54], and as we will show in the next section, the high model expressivity yet scalable training of deep learning makes it particularly suited for realistic registration problems.

C. Deep Learning-based Medical Image Registration Methods

To address the aforementioned complexity-efficiency trade-off of non-learning-based registration approaches, many deep learning-based registration methods have been proposed [55]–[58]. Different from traditional methods that focus on iteratively minimizing a predefined function eq. (1), deep neural network (DNN)-based methods are designed to learn the optimal registration field given a training dataset. This approach effectively learns a comprehensive global representation of image registration, facilitating the alignment of previously unseen pairs of volumes and significantly reducing registration time via the single-forward-pass nature of neural networks.

In 2019, Balakrishnan et al. introduced a registration framework called VoxelMorph, which combined convolutional neural networks (CNNs) with spatial transformer networks (STN) [17]. This framework integrates the image feature extraction capabilities of CNNs with the spatial transformation abilities of STNs, enabling robust and efficient alignment of image pairs. Compared to traditional registration algorithms, this method offered significantly higher flexibility and drastically shortened registration time. However, the default single-scale U-Net architectures that utilize a direct spatial transformer loss often perform poorly for registration outside of the

cranial area, falling short of achieving state-of-the-art results in registration tasks with large displacement [16].

To overcome these limitations, some other approaches have implemented multilevel schemes that effectively handle both small/local and large/global deformations [34], [59], [60]. However, many multilevel registration networks merely down-sample image features without directing the network's focus to the region of interest [61]. This is especially crucial for breast registration, where dense tissue occupies a small portion of the volume. Therefore, in this paper, we propose a two-stage VoxelMorph-based registration method that balances global alignment with the registration of smaller, complex yet crucial regions. Note that crucial regions can either be defined by the user via mask input or automatically extracted based on frangi vesselness filter [62], ensuring a balanced and accurate registration process shown in Sec. III.

D. Registration Methods for Breast MRI

Breast MRI registration is a challenging task and has received limited attention in recent literature. Early studies, such as Zuo *et al.* (1996) [63], introduced breast MRI registration based on rigid transformations, acknowledging the need to eliminate patient movement-induced misalignment and distortions in imaging results. Subsequent research has focused on refining and extending these methods to improve registration accuracy. Commonly used techniques are based on biomechanical modeling [64]–[66]. These methods utilize finite element methods (FEMs) for solving differential equations to predict the displacement of tissue within the breast. However, these models are limited in their ability to accurately capture and correct the large deformations that the breast can undergo in practice [13], and it is computationally complicated to develop a biomechanical model for each patient individually.

Some deep-learning-based registration models have also been utilized to address the challenge of breast MRI registration [13], [67], [68]. However, these algorithms primarily evaluate their performance based on tumor detection, often neglecting the accuracy of registration in normal breast MRI cases and rarely comparing their effectiveness against existing registration methods in terms of entire-volume similarities and overall registration quality.

III. METHOD

This work proposes a novel network framework that demonstrates superior performance on breast MRI deformable registration. The following sections present (a) the problem definition (Sec. III-A), (b) an overview of the GuidedMorph framework (Sec. III-B), and additional methodological details.

A. Problem Definition

Given a pair of breast MRI scans, our registration method aims to align a moving (or *source*) volume to another fixed (*target*) volume. Formally, this is defined by the optimization problem of finding

$$\hat{\phi} = \underset{\phi}{\operatorname{argmin}} E_{\text{sim}}(m \circ \phi, f) + \lambda R(\phi) \quad (1)$$

where m and f represent the moving volumes and fixed volumes, respectively, ϕ is the *deformation field* used to deform m to align with f via the composition $m \circ \phi$. $E_{\text{sim}}(m \circ \phi, f)$ measures the similarity between fixed and warped moving volumes, the commonly used measurements include Cross-Correlation (CC) [51] and Mean Squared Error (MSE) [52]. Finally, $R(\phi)$ acts as a regularization function that enforces the smoothness of the deformation field, and the hyperparameter λ governs the regularization trade-off between fidelity to the data with the smoothness of the transformation.

B. GuidedMorph Overview

Our proposed registration framework (summarized in Fig. 2) is a two-stage registration method, where the first stage performs global alignment of the entire MRI volume, while the second stage refines the alignment in critical regions of interest, specifically dense tissue. The framework can be built upon different convolutional networks, including U-Net with a single stage similar to VoxelMorph [17], and Transformer whose single-stage structure resembles TransMorph [18]. However, in contrast to traditional single-scale network architectures that struggle to capture fine details in the volume, the newly introduced second stage in our GuidedMorph learns to output deformation fields for dense tissue and allows the motion from this fine but crucial region to guide the overall registration refinement.

The first stage of our pipeline takes a concatenated moving+fixed volume pair as its input. The fixed and moving volumes are two volumes belonging to the spatial domain $\Omega \subset \mathbb{R}^3$, which are denoted as m and f respectively. A convolutional neural network (CNN) learns the displacement field $\mathbf{u}_{\text{volume}}$ for overall structure alignment between m and f . The detailed architecture for CNN is described in the Implementation Sec. V-A. The $\mathbf{u}_{\text{volume}}$ is a matrix defined over a 4-dimensional spatial domain. Specifically, when $m, f \in \mathbb{R}^{W \times H \times D}$ (where W , H , and D stand for width, height, and depth of the volume, respectively), then the $\mathbf{u}_{\text{volume}} \in \mathbb{R}^{3 \times W \times H \times D}$. The three channels of $\mathbf{u}_{\text{volume}}$ store the displacement along the x , y , z axes, respectively.

After aligning the overall structure in the first stage, our second stage receives a concatenated volume pair supplemented by their corresponding extracted masks; these masks provide additional position information for dense tissue in volumes. In our algorithm, we provide three ways for mask generation: by employing our dense tissue extraction strategy based on Frangi vesselness filter shown in Sec. III-C; through the learning-based dense tissue segmentation; or through the learning-based breast segmentation algorithms detailed in Sec. IV-C. Masks generated from learning-based dense tissue segmentation provide precise position information for dense tissue, while breast mask and filter-extracted masks give less detailed information; however, filter-based extraction requires no external segmentation model, and breast masks are easy to annotate. The CNN in this stage learns the displacement field between the masked crucial regions in the fixed volume (f_{loc}) and stage 1 output moving volume ($m_{1,\text{loc}}$). The remaining voxels in the volume are then warped according to the movement of these crucial regions, ensuring that the entire volume

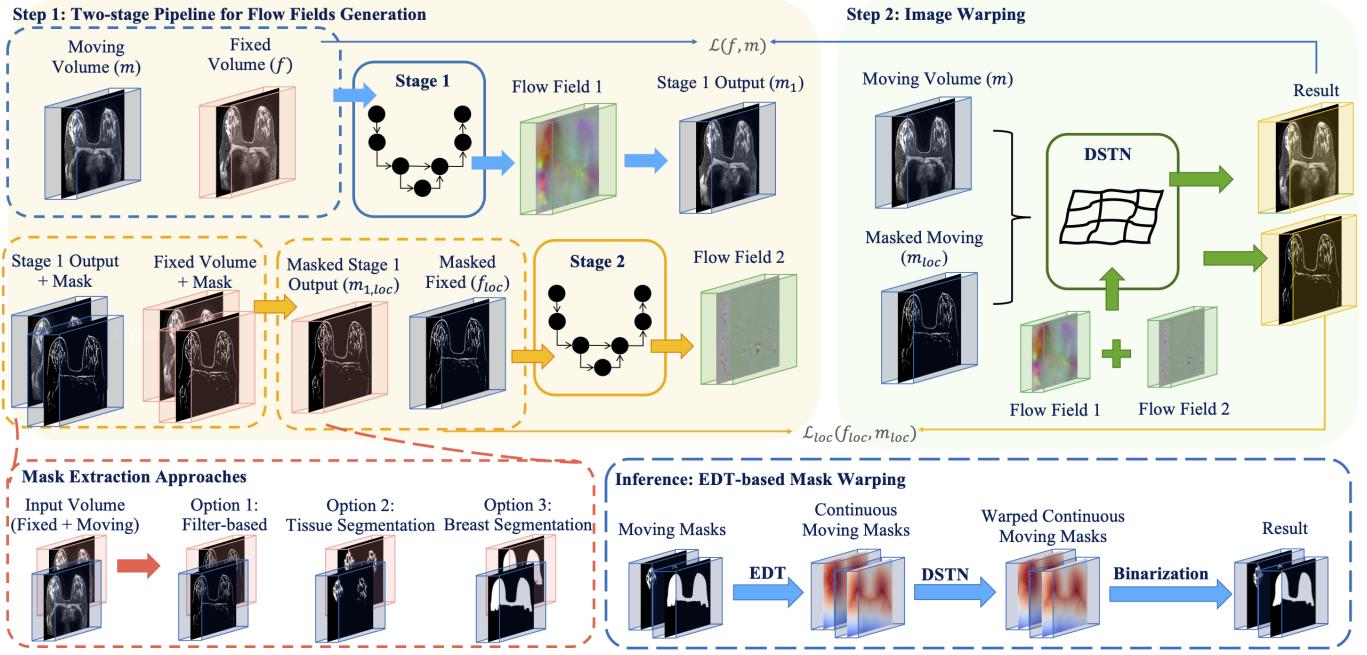


Fig. 2. **GuidedMorph method overview:** the top branch shows Stage 1 (blue), which aligns the moving and fixed volumes to estimate a global flow field. The bottom branch shows Stage 2 (yellow), where extracted features—obtained via filter-based methods, dense tissue segmentation, or breast segmentation—are concatenated with the stage 1 output to guide localized refinement. The framework is supervised by both global image similarity loss and localized feature-based loss. EDT-based mask warping pipeline at the bottom right demonstrates our novel mask warping algorithm during inference stage.

is harmoniously transformed. The output displacement field from this stage, denoted as \mathbf{u}_{mask} , shares the same shape as \mathbf{u}_{volume} .

The deformation fields ϕ for GuidedMorph consist of two parts: $\phi_{volume} = Id + \mathbf{u}_{volume}$ and $\phi_{mask} = Id + \mathbf{u}_{mask}$, where Id is the identity transform. DSTN, which will be described in detail in Sec. III-D, is responsible for fusing the ϕ_{volume} and ϕ_{mask} into a single ϕ , and further apply the spatial transformation to m and m_{loc} respectively. Notably, in the inference stage for binary mask warping, we introduce a EDT prior to warping to preserve structural fidelity during deformation; details are provided in Sec. III-E.

Stochastic gradient descent is utilized to optimize the respective parameters θ_1 and θ_2 for both CNN models simultaneously by minimizing a loss function shown in Eq. (7). The detailed implementation for the loss function is shown in Sec. III-F. We also employ dynamic loss weights to adjust the network focus over time. Particularly, the network initially concentrates on overall structure alignment given by stage 1 and then emphasizes dense tissue registration (stage 2). The detailed implementation for this dynamic focus is shown in Sec. V-A.

C. Dense Tissue Extraction Strategy

As detailed in Sec. GuidedMorph Overview III-B, we propose three approaches to extraction masks which offers dense tissue position information. The first approach utilizes our novel dense tissue extraction strategy, while the second and third approaches rely on learning-based breast and dense tissue segmentation model respectively. Although publicly available

segmentation models and weights exist for breast and dense tissue [69], running inference with the provided code requires significant time and effort. To overcome this limitation, we introduce a novel dense tissue extraction strategy based on the Frangi vesselness filter [62].

The pipeline for filter-based dense tissue extraction, along with its comparison to the segmentation model result, is illustrated in Fig. 3. The extraction process contains three steps, specifically Filtration, Threshold, and Binarization.

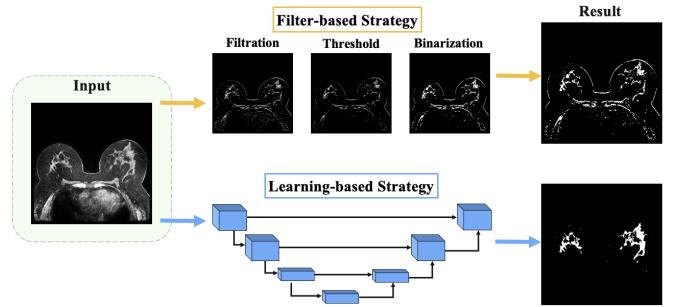


Fig. 3. **Filter-based and learning-based breast dense tissue extraction overview and comparison:** The top branch (yellow) shows the three steps of the filter-based strategy: Filtration, Thresholding, and Binarization. The bottom branch (blue) shows the learning-based strategy. The outputs from the two strategies are compared in the Result column.

The filtration step employs the Frangi vesselness filter [62] to extract tube-like structures in the volume. The Frangi vesselness filter, which is based on the Hessian matrix and its eigenvalues, is one of the mostly used methods to enhance

the tube-like structures including blood vessels, neurites, and tissue. However, the Frangi vesselness filter retains noise in the chest, contours, as well as dense tissue when it is applied to breast MRI.

To address this, we clip the low-intensity parts of the filtered volume in the threshold step to remove the noisy components in the chest. The 90th-percentile threshold is chosen based on both visual and empirical testing, as shown in Fig. 6 and Fig. 7, respectively. The final binarization step binarizes the extracted structures, converting them into a usable binary mask to resemble the output for dense tissue segmentation model.

Although this thresholding process may affect some dense tissue details and some contour noise may still be retained, as shown in the result section of two braches in Fig. 3 and the dense tissue extraction results in Fig. 4, our experiments show that the rough dense tissue extraction based on filter still provides decent dense tissue information which improves the breast and dense tissue alignment significantly (Sec. VI). More detailed information on threshold selection is shown in Sec. V-A.2.

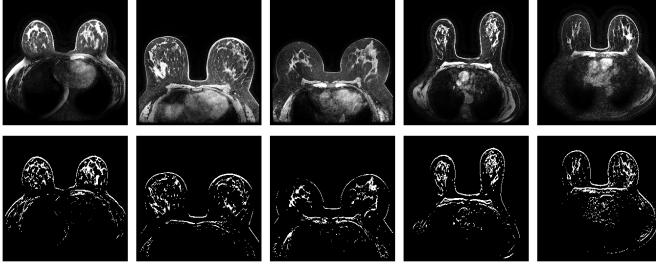


Fig. 4. Extracted masks based on the Frangi vesselness filter: The top row shows the middle slices of the original volumes, and the bottom row shows the corresponding middle slices of the extracted masks.

D. Dual Spatial Transformer Networks (DSTN)

The most intuitive way to apply both \mathbf{u}_{volume} and \mathbf{u}_{mask} to volumes such as m and m_{loc} is to apply the two transformations sequentially, each followed by interpolation. In practice, this is implemented by first applying \mathbf{u}_{volume} via a spatial transformer network (STN), which includes both the transformation and the interpolation steps, to generate a half-warped volume denoted as m_1 (as shown in Eq. (2)).

$$m_1 = \sum_{\mathbf{q} \in \mathbb{Z}^3(\mathbf{p}')} m(\mathbf{q}) \prod_{d \in \{x,y,z\}} (1 - |\mathbf{p}'_d - \mathbf{q}_d|) \quad (2)$$

Here, we adopt similar notation to [17], where \mathbf{p} represents each voxel ($\mathbf{p} \in \mathbb{R}^3$) inside the moving volume m , $\mathbf{p}' = \mathbf{p} + \mathbf{u}_{volume}(\mathbf{p})$, $\mathbb{Z}^3(\mathbf{p})$ is the set of neighboring voxels of a given \mathbf{p} , and d iterates over \mathbb{R}^3 .

Then, to generate the final warped volume (m_2), m_1 is put into the regular STN with \mathbf{u}_{mask} (in Eq. (4)), where \mathbf{p}_1 is some voxel in m_1 , and $\mathbf{p}'' = \mathbf{p}_1 + \mathbf{u}_{mask}(\mathbf{p}_1)$.

$$m_2 = \sum_{\mathbf{q} \in \mathbb{Z}^3(\mathbf{p}'')} m(\mathbf{q}) \prod_{d \in \{x,y,z\}} (1 - |\mathbf{p}''_d - \mathbf{q}_d|) \quad (3)$$

However, this method requires two rounds of resampling with interpolation, potentially blends neighboring values and

blurs edge-rich structures such as dense tissue. To mitigate this problem, we first approximate the two-round transformation and interpolation by composing the two transformations, and then perform a single interpolation named as DSTN. This approximation can be formally expressed as:

$$m \circ \phi \approx \sum_{\mathbf{q} \in \mathbb{Z}^3(\mathbf{p}_{final})} m(\mathbf{q}) \prod_{d \in \{x,y,z\}} (1 - |\mathbf{p}_{final,d} - \mathbf{q}_d|) \quad (4)$$

where $\mathbf{p}_{final} = \mathbf{p} + \mathbf{u}_{volume}(\mathbf{p}) + \mathbf{u}_{mask}(\mathbf{p} + \mathbf{u}_{volume}(\mathbf{p}))$, representing applying both \mathbf{u}_{volume} and \mathbf{u}_{mask} to the original coordinate \mathbf{p} . Fig. 5 demonstrate the difference between DSTN and directly applying STN twice. The effectiveness of DSTN has been empirically and visually validated through a comparison between training with the two-round STN and with DSTN (see Sec. VII).

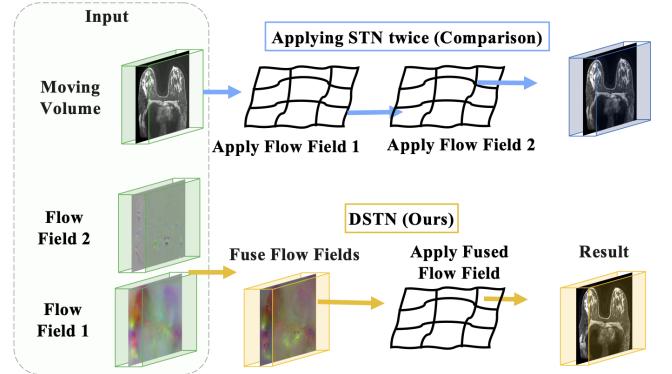


Fig. 5. Pipeline comparison between DSTN and applying STN twice: As illustrated in top branch (blue), the two-round STN approach applies Flow Field 1 to the moving volume, followed by Flow Field 2. In contrast, the bottom branch (yellow) illustrates the DSTN pipeline, where the two flow fields are fused before being applied to the moving volume.

E. EDT-Based Warping Algorithm

During the implementation of our registration framework, we observed that conventional warping techniques—such as nearest-neighbor interpolation—tend to degrade fine structural details in binary masks, as demonstrated in Fig. 11. This issue was notably absent when the same interpolation methods were applied to breast MRI images. We believe that such interpolation schemes inherently smooth binary structures, leading to the loss of high-frequency boundary information—a concern previously highlighted in other literature [70], [71]. To mitigate this, our approach transforms the binary mask into a signed distance map using a EDT [72] prior to warping. This produces a continuous-valued representation that preserves structural fidelity throughout the deformation process.

The signed distance at each voxel is defined as the difference between its distance to the nearest background and foreground boundary. Let $M : \Omega \subset \mathbb{R}^3 \rightarrow \{0, 1\}$ denote a binary mask defined on a volumetric domain Ω , where $M(x) = 1$ indicates foreground (object) voxels and $M(x) = 0$ indicates background. The signed distance transform $D : \Omega \rightarrow \mathbb{R}$ assigns to each voxel $x \in \Omega$ a value equal to its Euclidean

distance to the nearest boundary voxel, with a sign that indicates whether x lies inside or outside the object. Formally, it is defined as (5)

$$D(x) = \text{dist}_{\text{edt}}(x, \partial M^+) - \text{dist}_{\text{edt}}(x, \partial M^-), \quad (5)$$

where ∂M^+ and ∂M^- denote the boundaries of the foreground and background regions, respectively.

The function $\text{dist}_{\text{edt}}(x, S)$ denotes the Euclidean distance from voxel x to the nearest point in the set $S \subset \Omega$, defined in eq. (6)

$$\text{dist}_{\text{edt}}(x, S) = \min_{y \in S} \|x - y\|_2 \quad (6)$$

After converting the binary mask into a signed distance map, we apply the deformation field using a spatial transformer with bilinear interpolation, followed by re-binariization of the warped distance map using zero as the threshold to recover the final mask. The pipeline for this algorithm is shown at the bottom left of Fig. 2. The effectiveness of this algorithm is demonstrated both qualitatively and quantitatively in Sec. VII-4.

F. Loss Function

The loss function for GuidedMorph consists of three components: volume similarity, local region similarity, and regularization for penalizing fused spatial variations in ϕ . Prior registration loss functions typically penalize differences between fixed masks and warped moving masks [17], [73], [74], yet we found that these differences often exhibit discontinuities and fluctuate sharply with even small transformations. Instead, we penalize the differences between the volumes under the masks, offering a smoother and more stable optimization process. Our new loss function is then defined as

$$\begin{aligned} \mathcal{L}(f, f_{loc}, m, m_{loc}, \phi) &= L_{\text{sim}}(f, m \circ \phi) + \lambda L_{\text{smooth}}(\phi) \\ &\quad + \alpha L_{\text{sim}}(f_{loc}, m_{loc} \circ \phi) \end{aligned} \quad (7)$$

Where the masked fixed and moving volumes are f_{loc} and m_{loc} , respectively. λ and α are smoothness regularization and volume similarity strength hyperparameters, respectively. The two most common similarity functions for L_{sim} are the mean square difference (Eq. (8)) and local cross correlation (Eq. (9)), defined as

$$MSE(f, m \circ \phi) = \frac{1}{|\Omega|} \sum_{p \in \Omega} [f(p) - [m \circ \phi](p)]^2 \quad (8)$$

and

$$CC(f, m \circ \phi) = \frac{\sum_{\mathbf{p} \in \Omega} (N(\mathbf{p}))^2}{\sum_{\mathbf{p} \in \Omega} (\sum_{\mathbf{p}_i} \bar{f}(\mathbf{p}_i)^2) (\sum_{\mathbf{p}_i} \bar{m}(\mathbf{p}_i)^2)} \quad (9)$$

respectively, where $\bar{f}(\mathbf{p}_i)$, $\bar{m}(\mathbf{p}_i)$, and $N(\mathbf{p})$ are defined in (10), (11), and (12).

$$\bar{f}(\mathbf{p}_i) = f(\mathbf{p}_i) - \hat{f}(\mathbf{p}) \quad (10)$$

$$\bar{m}(\mathbf{p}_i) = (m \circ \phi)(\mathbf{p}_i) - (\hat{m} \circ \phi)(\mathbf{p}) \quad (11)$$

$$N(\mathbf{p}) = \sum_{\mathbf{p}_i} \bar{f}(\mathbf{p}_i) \bar{m}(\mathbf{p}_i) \quad (12)$$

IV. DATASET

A. ISPY2 Dataset

The ISPY2 dataset is a public dataset introduced by [75] that contains Dynamic Contrast-Enhanced (DCE) breast MRI data for 711 patients diagnosed with early-stage breast cancer. These patients were adaptively randomized to drug treatment arms between 2010 and 2016. The images, acquired from over 22 clinical centers, focus on patients undergoing neoadjuvant chemotherapy (NAC). Each patient underwent four MRI exams at specific timepoints: pretreatment (“T0”), early-treatment (“T1”), mid-treatment (“T2”), and post-treatment (“T3”) during their NAC.

For this study, we randomly selected 100 patients from the ISPY2 dataset, each with four pre-contrast breast MRI studies. This subset was divided into training and test sets, consisting of 80 and 20 patients, respectively. To simulate a clinical scenario in which physicians monitor sequential changes across multiple examinations, we constructed volume pairs for registration using the timepoints (“T0”, “T1”), (“T1”, “T2”), and (“T2”, “T3”). This approach resulted in a total of 240, and 60 pairs for the training, and test sets, respectively.

Before use, all volumes were resampled to a $128 \times 256 \times 256$ grid with 1 mm isotropic voxels. Data preprocessing included (1) clipping intensities to the 1st and 99th percentiles to minimize outlier effects, and (2) linearly scaling these values to a 0–1 range. Additionally, affine volume registration was performed using the Advanced Normalization Tools (ANTs).

B. Internal Breast Dataset

For the internal dataset, we collected breast MRI scans from our institution (The research protocol was approved by the Health System Institutional Review Board (IRB) Pro00112340), spanning from March 2014 to December 2021. To align with the dataset selected from ISPY2, we randomly selected 100 patients, each with four pre-contrast breast MRI studies. This subset was divided into training and test sets, consisting of 80 and 20 patients, respectively. Furthermore, the four studies were ordered by the study date to simulate the “T0,” “T1,” “T2,” and “T3” timepoints in the ISPY2 dataset, and registration pairs were constructed accordingly. This approach resulted in a total of 240 and 60 pairs for the training and test sets, respectively. The preprocessing steps were identical to those applied to the ISPY2 dataset.

C. Mask Generation

To generate masks for the ISPY2 and internal dataset, we followed [21], [69] to train a patch-based VNet [76] on the Duke Breast MRI dataset [69], [77]. The patch size is set to 96 and we utilize sliding window-based inference. We augment the data with random brightness changes. The model achieves 0.95 and 0.88 Dice segmentation coefficient (DSC) on the breast and dense tissue respectively, and was selected by validation performance.

V. EXPERIMENTS

A. Implementation

The proposed deformable registration method was developed using the PyTorch library in Python. Our method is built on a 3D convolutional neural network (CNN) which we detail later in this section, trained at an initial learning rate of 5×10^{-4} . To achieve stable optimization, we employed the AdamW optimizer with momentum, using a batch size of 1.

To adjust the balance between similarity penalty and regularization across different loss functions (Eq. 7), we chose the hyper-parameter settings as follows. For the MSE loss function, which has a higher similarity penalty, we set $\alpha = 1$ and $\lambda = 0.08$ to ensure sufficient regularization. Conversely, for the CC loss function where the similarity penalty is lower, we adjusted the weights to $\alpha = 1$ and $\lambda = 8$, thus maintaining a consistent balance between the similarity penalty and regularization across both metrics. The model was trained on a single 48GB NVIDIA RTX A6000 GPU for 500 epochs, the CPU running times are evaluated on AMD Ryzen Threadripper 3970X, with 32 cores, 64 threads, and 128 MB L3 cache. Detailed information on the 3D CNN architecture, and dynamic loss functions can be found in the subsequent subsections. Since training with loss functions based on MSE and CC yields comparable performance, as shown in Sec. VII, in the experiment, we conduct the baseline implementation and result comparison using only MSE for faster convergence.

1) 3D CNN Architecture: We adopt the CNN architecture from both VoxelMorph [17] and TransMorph [18] for our GuidedMorph (Voxel) and GuidedMorph (Trans), respectively. For GuidedMorph (Voxel), we follow the implementation in [17], using a kernel size of 3 in both the encoder and decoder stages, and a stride of 2. Each convolution is followed by a LeakyReLU activation with a slope of 0.2 to introduce non-linearity. In the decoding phase, the network alternates between upsampling and deconvolution layers, incorporating skip connections that directly pass features from the encoder to the corresponding layers involved in generating the registration. The innermost layer performs convolutions on a volume of size $(1/16)^3$ relative to the input dimensions.

For GuidedMorph (Trans), we adopt the Swin Transformer-based architecture introduced in [18]. The network begins by dividing the input volumes into non-overlapping 3D patches of size $2 \times P \times P$, with P typically set to 4, resulting in a patch embedding of $8C$ dimensions. These patches are flattened and linearly projected into token embeddings of dimension $C = 96$. The tokens pass through four sequential stages of Swin Transformer blocks, each operating at decreasing spatial resolutions and increasing feature depths. To recover spatial precision lost during tokenization, we include two convolutional paths—one using the original resolution and the other using a downsampled version of the input. Each path extracts high-resolution local features, which are concatenated with the Transformer outputs and passed to the decoder through skip connections. The decoder comprises successive upsampling layers with $3 \times 3 \times 3$ convolutions and LeakyReLU activations (slope 0.2). The output deformation field is generated by a final convolutional layer and used by the spatial transformer

module to warp the moving image.

2) Dense Tissue Extraction Strategy: As mentioned in Sec. III-C, this section is utilized to demonstrate the technical implementation details of our filter-based mask extraction algorithm, specifically the threshold selection to reduce noise and better resemble the true dense tissue mask. As shown in Fig. 6, the noise is quite obvious even when we set the threshold up to 60 percentile. With the increase of the threshold, there is an obvious noise reduction, but it also lead to the loss of certain detailed information in dense tissue. To further confirm which threshold provides the best performance on our task, we conduct experiments on our GuidedMorph (Trans), with thresholds for filter-based dense tissue extraction ranging from 60 to 90 percentile. The performance is evaluated using the Dice coefficient between the registered source and target dense tissue masks. Since the result shows that a threshold of 90 yields the best performance, so we adopt it in our filter-based dense tissue extraction.

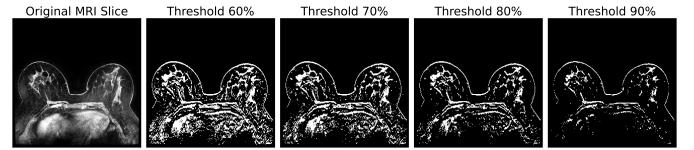


Fig. 6. Visualization of the threshold step across different percentile thresholds.

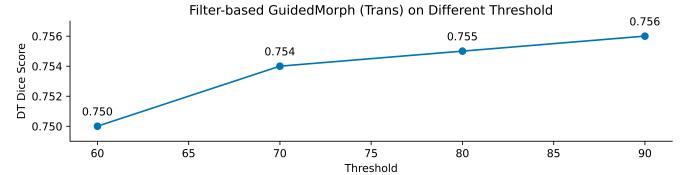


Fig. 7. Quantitative evaluation of filter-based dense tissue extraction across different percentile thresholds.

3) Dynamic Loss Functions: To achieve the dynamic region focus we mentioned in the Sec. III-B, we double the α parameter every 100 epochs for both MSE and CC similarity settings to enhance the model's focus on aligning masked critical sub-regions. This approach not only ensures alignment across the entire structure during the early stages of training but also guarantees that the network thoroughly captures and enhances details within the dense tissue as training progresses.

B. Baseline methods

For comparison, we selected two traditional optimization-based registration algorithms, **SyN** and **Nifty-reg** [31], [78] and three learning-based registration algorithms including **VoxelMorph**, **CycleMorph** and **TransMorph** [17]–[19] due to their superior performance in previous registration tasks [17], [18]. These were chosen as representatives for previous registration algorithms that have demonstrated state-of-the-art registration performance. While VoxelMorph and TransMorph propose multiple architectures, including *-diff*, *-Bayes*, and *-bspl*, we chose the architectures that have demonstrated the

best performance on brain label Dice coefficients in their respective works [17], [18]—namely, "VoxelMorph-2" (denoted as "VolumeMorph" hereafter) and "TransMorph."

For the hyper-parameters, we adhered to the default hyper-parameter settings as specified in their respective publications or GitHub repositories. To ensure a fair comparison, all pre-processing methods applied to our GuidedMorph model were also uniformly implemented across these baseline methods. To demonstrate the effectiveness of two-stage registration and eliminate the effect given by the extra segmentation mask input, we also built single-stage network incorporates the segmentation masks. The network is denoted as "TransMorph + DT" for those using a dense tissue mask and "TransMorph + B" for those using breast mask in the result table VI. The TransMorph is utilized in the experiment since it demonstrates the best performance among all deep-learning-based baselines.

The only difference between "TransMorph + DT" and "Ours + DT" is that "Ours + DT"'s flow field is the composition generated by two stages of the network, whereas "TransMorph + DT" uses the flow field solely from the first stage. The same distinction applies to "TransMorph + B" and "Ours + B". All other settings, including hyperparameters, loss functions, and input variables, are kept the same to ensure a fair comparison.

C. Evaluation

Due to the high accuracy of the segmentation algorithm validated to achieve a Dice coefficient exceeding 88% for breast dense tissue and 95% for overall breast- we are able to directly obtain dense tissue and breast masks for the test dataset. Using these masks, we evaluate the registration performance by comparing the deformed segmentation map with the fixed mask, employing the Dice coefficient score. The method for calculating this score is detailed in Eq. (13).

$$Dice(f_{mask}, m_{mask} \circ \phi) = 2 \cdot \frac{|f_{mask} \cap (m_{mask} \circ \phi)|}{|f_{mask}| + |m_{mask} \circ \phi|} \quad (13)$$

The Dice coefficient approaching 1 indicates that the registered moving volume and the fixed volume are perfectly aligned, while a Dice coefficient of 0 demonstrates no overlap. To further eliminate potential bias from the evaluation process given by segmentation model, we also include the SSIM measurements on the breast (excluding the chest part due to our task focus) as an alternative similarity metric for the registered moving volume and the fixed volume shown in Eq. (14).

$$SSIM(f, m \circ \phi) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (14)$$

Where the μ_x, μ_y are the average intensities of volume f and $m \circ \phi$, σ_x^2, σ_y^2 are the variances and σ_{xy} is the covariance of two volume f and $m \circ \phi$. c_1 and c_2 are the constants used to stabilize the division with weak denominator.

To measure the consistency of the deformation fields, we presented the percentages of non-positive values in the determinant of the Jacobian matrix associated with the deformation fields. In the result Sec. VI the matrix is denoted as % of $|J_\phi| \leq 0$. Additionally, we have displayed the running times for each method to demonstrate their efficiency.

VI. MAIN RESULTS

The results section includes both qualitative and quantitative evaluation result for the breast registration on ISPY2 and our internal breast dataset. While loss function based on CC and MSE gives comparable performance shown in VII-5, the result shown in the below sections are evaluated with MSE. Since TransMorph demonstrates the best performance among all learning-based baselines, we primarily evaluate our GuidedMorph (Trans) in both qualitative and quantitative analyses. Additionally, we include a comparison between the single-stage baseline and our GuidedMorph (Voxel) in Tab. II on the ISPY2 dataset.

A. Qualitative Evaluation

Fig. 8 presents the qualitative evaluation of GuidedMorph (Ours), based on TransMorph architecture, and filter-based dense tissue extraction method, compared against TransMorph on ISPY2 and our internal breast dataset, respectively. For each result figure, rows from top to bottom show slices at the 1/5th, 2/5th, 1/2nd, 3/5th, and 4/5th positions from different MRI volumes. Columns, from left to right, display the source slice, target slice, Trans-registered slice, Trans-target difference, green box zoom-in from Trans-target difference, our registered slice, our model's difference, green box zoom-in from our model's difference. While our model's differences show lighter colors (indicating smaller discrepancies) compared to the Trans-target difference, the results demonstrate the overall superiority of our model. We also add green boxes in the result figure to highlight regions within the images, emphasizing the improved alignment of dense tissue achieved by GuidedMorph (Trans) over TransMorph. Compared to the general difference color, the green boxes emphasize more on the dense tissue area due to the focus of our study.

B. Quantitative Evaluation

To provide a comprehensive comparison of GuidedMorph with other established registration methods, Table I details the quantitative evaluation metrics used on ISPY2 and the internal breast dataset, including the Dice score for dense tissue ($Dice_{DT}$), Dice score for breast tissue ($Dice_B$), SSIM for breast, the percentage of non-positive values in the Jacobian determinant, and the runtime on both GPU and CPU platforms. Detailed descriptions of the evaluation metrics and algorithm implementations are presented in Sec. V-C and V-A, respectively.

The results demonstrate that GuidedMorph, in its filter-based mode without any segmentation masks input (Our (Filter)), achieves higher $Dice_{DT}$, $Dice_B$, and SSIM compared to other registration algorithms that do not rely on segmentation models. The percentage of non-positive values in the Jacobian determinant for the Our model is comparable to that of previous state-of-the-art methods. When utilizing segmentation masks for both dense and breast tissues (Ours + DT and Ours + B), our GuidedMorph further surpasses the performance of single-layer TransMorph equipped with the same masks in terms of Dice scores, SSIM, and the stability

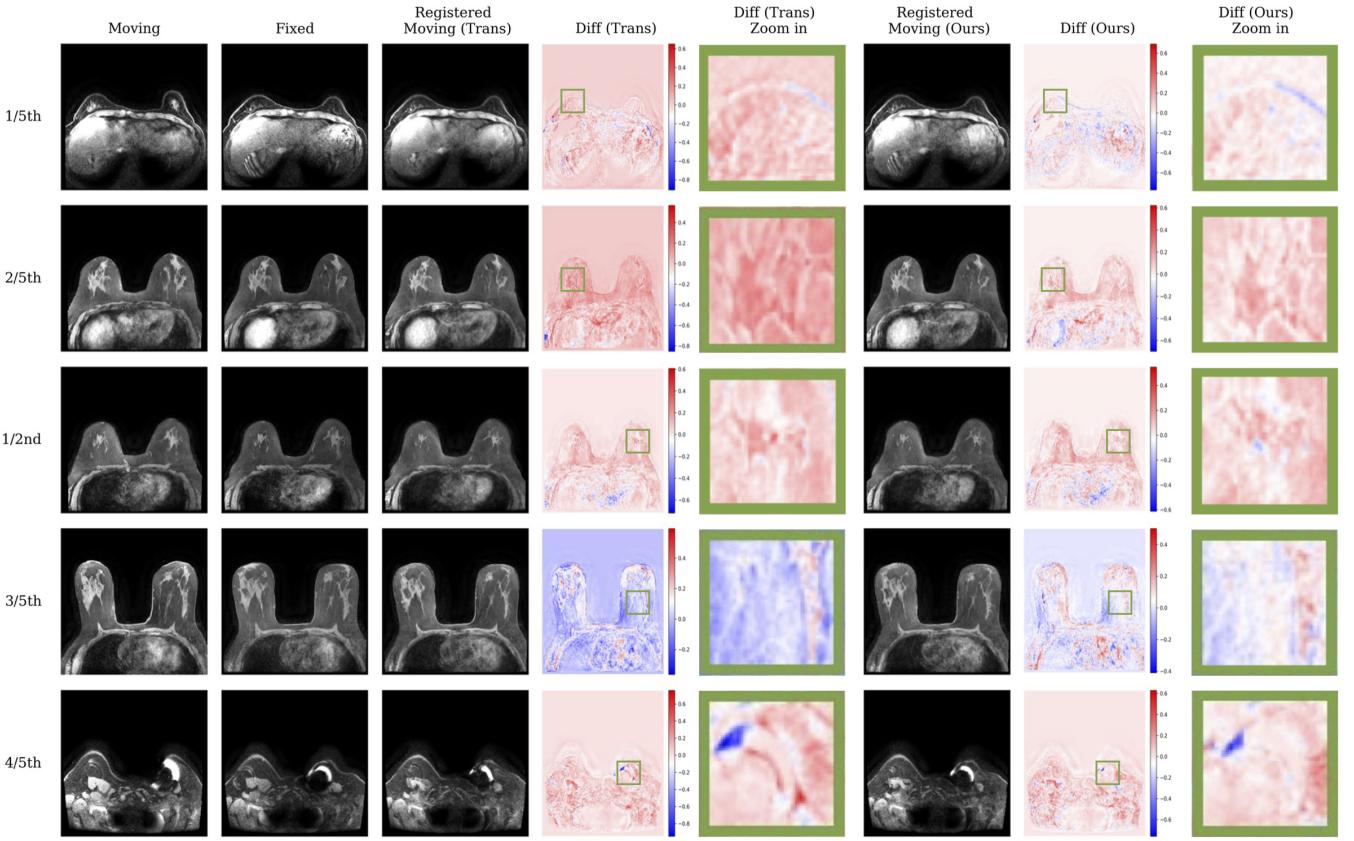


Fig. 8. Qualitative comparison of our model versus the TransMorph baseline: Rows, from top to bottom, show slices at the 1/5th, 2/5th, 1/2nd, 3/5th, and 4/5th positions from different MRI volumes. Columns, from left to right, display the source slice, target slice, Trans-registered slice, Trans-target difference, green box zoom-in from Trans-target difference, our registered slice, our model's difference, green box zoom-in from our model's difference. The green boxes highlight regions where our method achieves better alignment, and the zoom-in columns offer enhanced visualization of these areas.

of the transformation as indicated by the lower percentage of non-positive values in the Jacobian determinant.

Specifically, averaged across both datasets, our (Filter) achieves an improvement of **12.2%** in Dice_{DT} , **5.51%** in Dice_B , and **1.38%** in SSIM compared to the best learning-based baseline that do not rely on segmentation models. Compared to the best performing optimization-based algorithm, our model significantly reduces the inference time and improves Dice_{DT} by **5.29%** while maintaining comparable Dice_B and breast SSIM.

For our model with segmentation mask input, we observe a **21.16%** improvement in dense tissue Dice coefficient, **3.13%** in breast Dice coefficient, and **2.11%** in overall breast SSIM compared to the single stage baseline with same mask input.

We also demonstrate that our framework is compatible not only with TransMorph but also with VoxelMorph as the single-stage backbone, by comparing filter-based GuidedMorph (Voxel) with the original VoxelMorph. As a result, our model achieves improvements of **13.01%** in Dice_{DT} , **4.91%** in Dice_B , and **1.21%** in SSIM over VoxelMorph.

VII. ABLATION STUDIES

In this section, we design experiments to evaluate the components and choices made in our proposed registration model in Sec. III. All experiments are conducted using our

GuidedMorph (Trans) in the filter-based setting on the ISPY2 dataset. The study is divided into five key parts:

(1) **Hyper-parameter Analysis** demonstrates the performance of TransMorph and filter-based GuidedMorph (Trans) under a varied range of hyper-parameters.

(2) **Network Components** presents two additional experiments to demonstrate the efficacy of the first and second stages in our registration model, as described in the GuidedMorph Overview (Section III-B).

(3) In section **DSTN Model**, we compare our proposed DSTN with an alternative option that performs the STN twice, as mentioned in Section III-D, to demonstrate the effectiveness of DSTN approximation.

(4) **EDT-based Warping** compares our warping algorithm described in Sec. III-E with the vanilla warping algorithm using either bilinear or nearest interpolation under various experimental settings.

(5) Finally, in section **Loss Function**, we investigate the impact of different loss functions on model performance by presenting results using both MSE and NCC. Our findings, summarized in a comparative table, indicate that the choice of loss function does not significantly affect the final outcomes.

All experiments are conducted on the ISPY2 dataset.

1) Hyper-parameter Analysis: In this section, we evaluate the performance of TransMorph and GuidedMorph (Trans)

TABLE I

QUANTITATIVE COMPARISON BETWEEN OUR GUIDEDMORPH (TRANS), AND BASELINE REGISTRATION MODELS ON THE ISPY2 AND INTERNAL DATASET UNDER THREE EVALUATION SETTINGS

ISPY2 Dataset							
Method	Dice _{DT}	Dice _B	SSIM	% of J _φ ≤ 0	GPU sec	CPU sec	Seg.
Affine	0.467 ± 0.136	0.850 ± 0.062	0.835 ± 0.053	-	-	-	✗
SyN	0.686 ± 0.135	0.907 ± 0.049	0.885 ± 0.039	0.006 ± 0.018	-	255.411 ± 34.201	✗
NiftyReg	0.718 ± 0.128	0.908 ± 0.050	0.897 ± 0.037	0.213 ± 0.211	-	45.011 ± 9.594	✗
VoxelMorph	0.672 ± 0.127	0.881 ± 0.059	0.891 ± 0.035	2.244 ± 0.962	0.106 ± 0.025	3.577 ± 0.410	✗
CycleMorph	0.670 ± 0.126	0.881 ± 0.059	0.886 ± 0.036	2.387 ± 0.921	0.095 ± 0.026	3.224 ± 0.252	✗
TransMorph	0.681 ± 0.126	0.883 ± 0.058	0.894 ± 0.034	2.138 ± 0.981	0.208 ± 0.027	5.799 ± 1.255	✗
TransMorph + B	0.706 ± 0.124	0.912 ± 0.050	0.902 ± 0.033	0.332 ± 0.248	0.210 ± 0.041	5.818 ± 0.157	✓
TransMorph + DT	0.725 ± 0.118	0.904 ± 0.054	0.897 ± 0.035	0.434 ± 0.361	0.211 ± 0.047	5.854 ± 0.138	✓
Ours (Filter)	0.756 ± 0.120	0.908 ± 0.050	0.904 ± 0.033	0.783 ± 0.430	0.412 ± 0.046	11.003 ± 0.147	✗
Ours + B	0.738 ± 0.119	0.945 ± 0.041	0.938 ± 0.024	0.338 ± 0.288	0.416 ± 0.043	11.203 ± 0.175	✓
Ours + DT	0.934 ± 0.062	0.908 ± 0.052	0.899 ± 0.034	0.366 ± 0.281	0.420 ± 0.047	11.192 ± 0.157	✓

Internal Dataset							
Method	Dice _{DT}	Dice _B	SSIM	% of J _φ ≤ 0	GPU sec	CPU sec	Seg.
Affine	0.402 ± 0.225	0.772 ± 0.123	0.878 ± 0.0561	-	-	-	✗
SyN	0.578 ± 0.233	0.855 ± 0.099	0.905 ± 0.051	0.005 ± 0.008	-	257.039 ± 69.360	✗
NiftyReg	0.612 ± 0.223	0.876 ± 0.071	0.915 ± 0.046	0.339 ± 0.442	-	36.106 ± 8.792	✗
VoxelMorph	0.584 ± 0.209	0.814 ± 0.112	0.911 ± 0.042	3.225 ± 1.859	0.103 ± 0.025	3.492 ± 0.576	✗
CycleMorph	0.581 ± 0.211	0.819 ± 0.111	0.908 ± 0.044	3.600 ± 1.658	0.094 ± 0.057	2.505 ± 0.273	✗
TransMorph	0.600 ± 0.204	0.818 ± 0.110	0.909 ± 0.042	3.438 ± 1.873	0.212 ± 0.053	4.945 ± 0.087	✗
TransMorph + B	0.609 ± 0.214	0.888 ± 0.077	0.919 ± 0.040	0.864 ± 0.638	0.222 ± 0.059	6.326 ± 0.589	✓
TransMorph + DT	0.640 ± 0.207	0.849 ± 0.104	0.915 ± 0.042	1.027 ± 0.704	0.222 ± 0.056	6.487 ± 0.532	✓
Ours (Filter)	0.680 ± 0.201	0.885 ± 0.103	0.924 ± 0.038	1.472 ± 1.126	0.412 ± 0.047	10.022 ± 0.184	✗
Ours + B	0.690 ± 0.192	0.942 ± 0.055	0.954 ± 0.027	0.656 ± 0.539	0.412 ± 0.054	10.373 ± 0.194	✓
Ours + DT	0.883 ± 0.136	0.869 ± 0.092	0.919 ± 0.042	0.751 ± 0.613	0.413 ± 0.058	10.546 ± 0.191	✓

(1) without additional segmentation model (Filter), (2) with a breast segmentation mask as auxiliary input (+B), and (3) with a dense tissue segmentation mask as auxiliary input (+DT). For each setting, we report the Dice coefficient for dense tissue (Dice_{DT}, our primary evaluation metric), overall breast Dice (Dice_B), structural similarity index (SSIM), percentage of non-positive Jacobian determinant (% of |J_φ| ≤ 0), and runtime (GPU and CPU). The best performance in each setting is highlighted in bold, and Dice_{DT} is bolded to emphasize its importance in this study. “Seg.” indicates whether the method requires an external segmentation model as input (✓) or not (✗).

TABLE II

QUANTITATIVE COMPARISON BETWEEN OUR GUIDEDMORPH (VOXEL) IN THE FILTER-BASED SETTING AND THE BASELINE VOXELMORPH MODEL ON THE ISPY2 DATASET

ISPY2 Dataset							
Method	Dice _{DT}	Dice _B	SSIM	% of J _φ ≤ 0	GPU sec	CPU sec	Seg.
VoxelMorph	0.672 ± 0.127	0.881 ± 0.059	0.891 ± 0.035	2.244 ± 0.962	0.106 ± 0.025	3.577 ± 0.410	✗
Ours (Filter)	0.660 ± 0.204	0.854 ± 0.098	0.922 ± 0.038	1.284 ± 0.921	0.215 ± 0.066	6.660 ± 0.195	✗

TABLE III

HYPERPARAMETER α ANALYSIS

α	Dice _{DT}
0.01	0.755 ± 0.120
0.1	0.757 ± 0.122
1	0.756 ± 0.120
10	0.746 ± 0.121
100	0.732 ± 0.128

TABLE IV
VARIATION OF MODULE

Ablation	Dice _{DT}
First Stage Only/ TransMorph (Filter)	0.705 ± 0.126
Second Stage Only	0.719 ± 0.124
STN Twice	0.735 ± 0.123
Our (Filter)	0.756 ± 0.120

based on filter-based mask extraction under various values of the hyper-parameters α and λ , as defined in Eq. (7). For α , we explore a range of values from 0.02 to 0.10 for both TransMorph and Our (Filter) based on TransMorph architecture, with the results presented in Figure 9. Similarly, for λ , we assess its impact using values of 0.01, 0.1, 1, 10, and

TABLE V
MASK WARPING ALGORITHM

Method (EDT + Interp)	Dice _{DT}
no EDT (Bilinear)	0.455 ± 0.143
no EDT (Nearest-Neighbor)	0.738 ± 0.119
with EDT (Nearest-Neighbor)	0.740 ± 0.119
with EDT (Bilinear)	0.756 ± 0.120

TABLE VI
LOSS FUNCTION ON PERFORMANCE

Loss Type	Dice _{DT}
MAE	0.756 ± 0.120
CC	0.757 ± 0.118

100, the result is shown in Table III. As a result, our proposed GuidedMorph demonstrates strong stability without significant performance variation across various hyperparameter settings, with the largest p-value of 0.276. Moreover, our model exhibits superior performance compared to TransMorph for all values of λ , as shown in Fig. 9.

2) Network Components: To demonstrate the effectiveness of combining the two stages, in this section we demonstrate

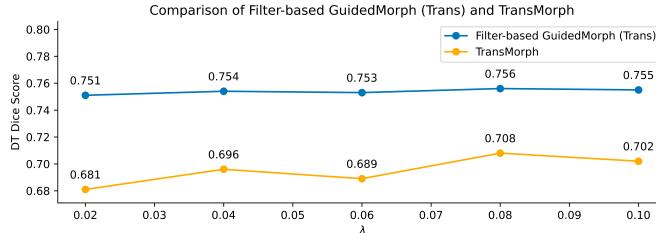


Fig. 9. Hyperparameter analysis with comparison between our GuidedMorph (blue), and baseline TransMorph (yellow).

the performance of the first stage only and the second stage only, as shown in Tab. IV. For this demonstration, the experiment utilizes the filter-based GuidedMorph (Trans) setting. Notably, the difference between the first-stage-only model and TransMorph is that the former uses the same loss function as GuidedMorph—including crucial-region similarity as one of the polynomial terms. Therefore, the "First Stage Only" model is effectively equivalent to TransMorph (Filter). As a result, combining the two stages improves the performance by **7.23%** compared with utilizing the first stage only and by **5.15%** compared with utilizing the second stage only.

3) **DSTN Model:** This section is designed to demonstrate the effectiveness comparing DSTN with applying the STN twice (see Sec. III-D). The experiment is also conducted on the filter-based GuidedMorph (Trans) setting, and the final performance evaluated by $Dice_{DT}$ is shown in IV. As shown, the proposed DSTN improved the performance by **2.86%**. We also present a visual comparison between the registration results from the two methods, as shown in Fig. 10. Although the differences are not apparent in the registered moving volumes, the difference maps—computed between the fixed and registered moving volumes—generally show lighter colors for DSTN, indicating a smaller gap between our result and the target. The performance differences are especially noticeable in the dense tissue regions, as highlighted by the green boxes.

4) **EDT-based Warping:** In this section, we aim to evaluate the different mask warping algorithm for our inference mode to prove the effectiveness of our proposed EDT-based Warping method. Tab. V compares the inference result with and without utilizing the EDT with both bilinear and nearest as interpolation strategy. The result demonstrate that EDT-based Warping method is less sensitive to interpolation algorithm compared to conventional warping algorithm and achieves the improvement of **2.44%** compared to warping algorithm without EDT with any interpolation strategy. The visual comparison between our method with EDT (ours) and conventional method without EDT (w/o EDT) is shown in Fig. 11. Columns, from left to right, display: the manually modified dense tissue segmentation overlaid on the registered moving slice (ground truth), the nearest-neighbor warped segmentation without EDT, our EDT-warped segmentation, the difference between ground truth and no EDT (with zoomed-in region), and the difference between ground truth and ours (with zoomed-in region). With fewer red and blue points in the difference maps, the results demonstrate that introducing EDT into the mask deformation mitigates both over-segmentation and under-segmentation.

5) **Loss Function:** To evaluate the impact of different loss functions on the performance of our registration model, we conducted experiments using both mean squared error (MSE) and normalized cross-correlation (CC) as loss functions in our filter-based GuidedMorph (Trans). The results, as shown in Table VI, indicate that the choice between MSE and CC does not significantly affect the final outcome with p-value being 0.963. The experiment result is also consistent with those reported in previous studies [17], [18].

VIII. DISCUSSION AND CONCLUSIONS

In this work, we propose the two-stage registration algorithm GuidedMorph. Unlike conventional single-stage registration algorithms that primarily focus on aligning general structures while overlooking intricate internal details, we introduce an additional network that utilizes dense tissue position information to allow the motion from fine but crucial structures—such as dense tissue—to guide the overall registration tracking. Our work provides three approaches for dense tissue information mask extraction: learning-based dense tissue segmentation, breast segmentation, and Frangi vesselness filter-based mask extraction.

Dense tissue segmentation-based (Ours + DT) setting offers the best dense tissue Dice coefficient, improving the dense tissue Dice by 33.40% compared to the single-stage baseline with the same mask input. With the significant improvement in aligning the dense tissue, the model didn't compromise with the overall breast Dice, demonstrate with an averaged improvement of 1.4% in breast Dice coefficient, and 0.33% in overall breast SSIM on two datasets.

Utilizing the breast segmentation mask as the second-stage input further improves performance, with a 4.85% increase in breast Dice coefficient and a 3.90% improvement in overall breast SSIM, compared to the single-stage baseline using the breast segmentation mask as auxiliary information.

The Frangi vesselness filter-based mask extraction strategy, similar to breast segmentation in providing a rough dense tissue extraction, demonstrates comparable performance in dense tissue Dice when using the breast mask as the second-stage input, and similar breast Dice performance when using the dense tissue mask as additional input. However, the Frangi vesselness filter-based strategy offers our model an option to achieve decent performance without relying on any additional segmentation model. It also provides a relatively fair comparison with previous unsupervised registration models, as neither requires additional model training or inference for segmentation. Experiments in the main result section demonstrate ours (Filter) achieves an improvement of 12.2% in dense tissue Dice, 5.51% in breast Dice, and 1.38% in SSIM compared to the best deep learning-based baseline without relying on external segmentation model input.

To further eliminate the influence of additional information provided by the filter and to purely evaluate the effectiveness of our two-stage framework, we also compare our (Filter) model with TransMorph under filter setting using the same loss function, hyperparameters, and mask warping algorithm. This comparison is demonstrated in the "First Stage Only/

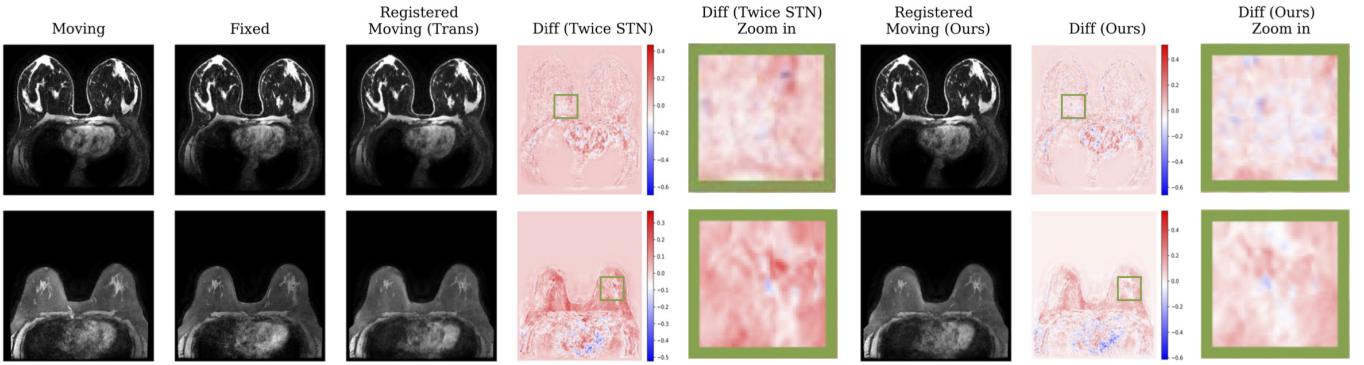


Fig. 10. Comparison between DSTN and applying STN twice: Rows show randomly selected slices.

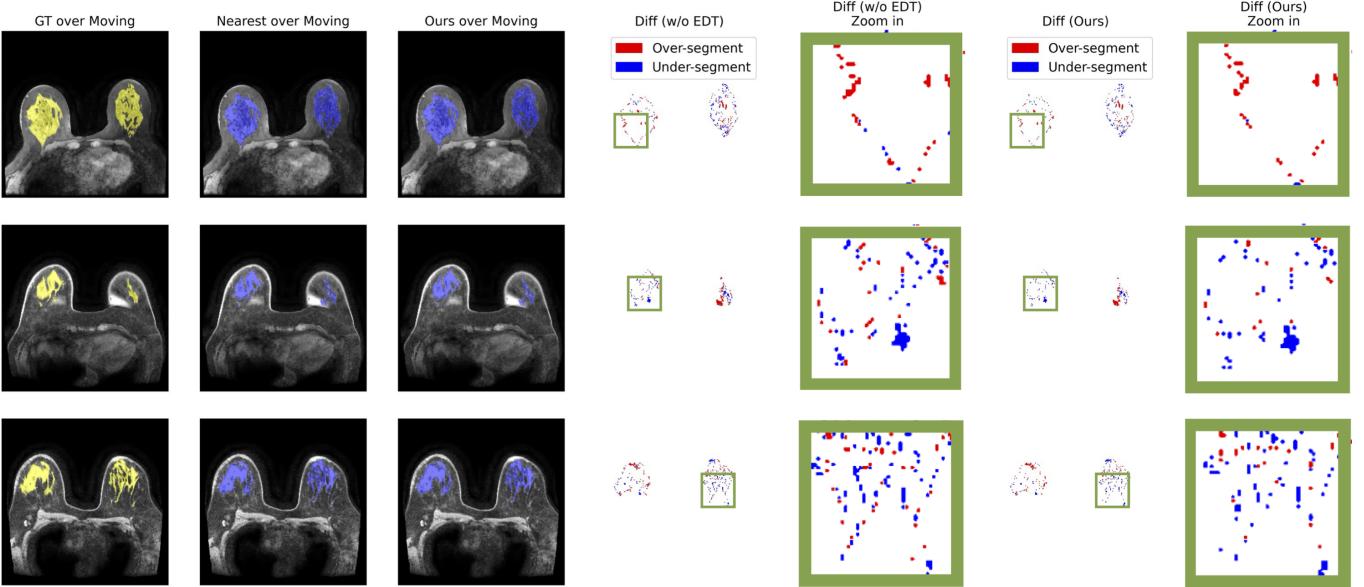


Fig. 11. Comparison between our method with EDT (ours) and conventional method without EDT (w/o EDT): Columns, from left to right, show: ground truth (manually modified segmentation, yellow) overlaid on the registered moving image, segmentation warped without EDT (nearest interpolation, blue), segmentation warped with EDT (blue), the difference map between ground truth and no EDT, and its zoomed-in view, followed by the difference map between ground truth and our EDT-based method, and its zoomed-in view. Red and blue indicate over-segmentation and under-segmentation, respectively.

TransMorph (Filter)” experiment in the ablation study section (see Sec. VII), which shows that the additional stage alone in our framework improves dense tissue Dice by 7.23%.

Our framework has also been tested with VoxelMorph as the backbone, showing an improvement of 13.01% in dense tissue Dice, 4.91% in breast Dice, and 1.21% in breast SSIM compared to single-stage VoxelMorph, demonstrating the framework’s versatility.

Apart from the two-stage framework, we evaluated two image warping algorithms—applying STN twice and DSTN—and demonstrated that DSTN improves performance by 2.86% compared to STN twice. We also introduced EDT prior to warping the binary mask during inference to preserve the mask’s fine structural details during deformation (see Sec. III-E). The EDT-based warping algorithm achieved a 2.4% improvement compared to conventional mask warping.

In the broader field of medical image registration, GuidedMorph’s two-stage framework offers a promising solution for

3D image registration involving large deformations—scenarios in which conventional learning-based methods often fall short [79], [80]. Furthermore, GuidedMorph demonstrates particular effectiveness in breast MRI registration, achieving state-of-the-art performance compared to both learning-based and traditional optimization-based algorithms. This advancement holds significant clinical relevance, as accurate breast registration is essential for longitudinal monitoring [81], treatment response evaluation [82], and multi-sequence data integration in breast cancer care [83].

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Institute Of Biomedical Imaging and Bioengineering of the National Institutes of Health under Award Number R01EB031575. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

- [1] R. M. Mann, N. Cho, and L. Moy, "Breast mri: state of the art," *Radiology*, vol. 292, no. 3, pp. 520–536, 2019.
- [2] H. Mehrabian, L. Richmond, Y. Lu, and A. L. Martel, "Deformable registration for longitudinal breast mri screening," *Journal of digital imaging*, vol. 31, pp. 718–726, 2018.
- [3] C. Kuhl, R. K. Schmutzler, C. C. Leutner, A. Kempe, E. Wardemann, A. Hocke, M. Maringa, U. Pfeifer, D. Krebs, and H. H. Schild, "Breast mr imaging screening in 192 women proved or suspected to be carriers of a breast cancer susceptibility gene: preliminary results," *Radiology*, vol. 215, no. 1, pp. 267–279, 2000.
- [4] C. D. Lehman, C. Isaacs, M. D. Schnall, E. D. Pisano, S. M. Ascher, P. T. Weatherall, D. A. Bluemke, D. J. Bowen, P. K. Marcom, D. K. Armstrong *et al.*, "Cancer yield of mammography, mr, and us in high-risk women: prospective multi-institution breast cancer screening study," *Radiology*, vol. 244, no. 2, pp. 381–388, 2007.
- [5] C. C. Riedl, N. Luft, C. Bernhart, M. Weber, M. Bernathova, M.-K. M. Tea, M. Rudas, C. F. Singer, and T. H. Helbich, "Triple-modality screening trial for familial breast cancer underlines the importance of magnetic resonance imaging and questions the role of mammography and ultrasound regardless of patient mutation status, age, and breast density," *Journal of clinical oncology*, vol. 33, no. 10, p. 1128, 2015.
- [6] E. Warner, H. Messersmith, P. Causer, A. Eisen, R. Shumak, and D. Plewes, "Systematic review: using magnetic resonance imaging to screen women at high risk for breast cancer," *Annals of internal medicine*, vol. 148, no. 9, pp. 671–679, 2008.
- [7] S. L. Heller and L. Moy, "Mri breast screening revisited," *Journal of Magnetic Resonance Imaging*, vol. 49, no. 5, pp. 1212–1221, 2019.
- [8] N. S. Winkler, S. Raza, M. Mackesy, and R. L. Birdwell, "Breast density: clinical implications and assessment methods," *Radiographics*, vol. 35, no. 2, pp. 316–324, 2015.
- [9] M. Nothacker, V. Duda, M. Hahn, M. Warm, F. Degenhardt, H. Madjar, S. Weinbrenner, and U.-S. Albert, "Early detection of breast cancer: benefits and risks of supplemental breast ultrasound in asymptomatic women with mammographically dense breast tissue. a systematic review," *BMC cancer*, vol. 9, pp. 1–9, 2009.
- [10] S. S. Kaplan, "Clinical utility of bilateral whole-breast us in the evaluation of women with dense breast tissue," *Radiology*, vol. 221, no. 3, pp. 641–649, 2001.
- [11] N. F. Boyd, L. J. Martin, M. Bronskill, M. J. Yaffe, N. Duric, and S. Minkin, "Breast tissue composition and susceptibility to breast cancer," *Journal of the National Cancer Institute*, vol. 102, no. 16, pp. 1224–1237, 2010.
- [12] C. Maloney, S. Paul, J. Lieberenz, M. A. Levy, L. Stempel, and R. Alvarado, "Breast density changes over time: Frequency, patterns, and practice implications," 2022.
- [13] M. G. French, G. D. M. Talou, T. P. B. Gamage, M. P. Nash, P. M. Nielsen, A. J. Doyle, J. E. Iglesias, Y. Balbastre, and S. I. Young, "Deformomorphic multi-resolution deep learning registration for applications in breast mri," *arXiv preprint arXiv:2309.13777*, 2023.
- [14] S. Ham, M. Kim, S. Lee, C.-B. Wang, B. Ko, and N. Kim, "Improvement of semantic segmentation through transfer learning of multi-class regions with convolutional neural networks on supine and prone breast mri images," *Scientific Reports*, vol. 13, no. 1, p. 6877, 2023.
- [15] H. Mehrabian, L. Richmond, Y. Lu, and A. L. Martel, "Deformable registration for longitudinal breast mri screening," *Journal of digital imaging*, vol. 31, pp. 718–726, 2018.
- [16] M. P. Heinrich and L. Hansen, "Voxelmorph++ going beyond the cranial vault with keypoint supervision and multi-channel instance optimisation," in *International Workshop on Biomedical Image Registration*. Springer, 2022, pp. 85–95.
- [17] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "Voxelmorph: A learning framework for deformable medical image registration," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.
- [18] J. Chen, E. C. Frey, Y. He, W. P. Segars, Y. Li, and Y. Du, "Transmorph: Transformer for unsupervised medical image registration," *Medical image analysis*, vol. 82, p. 102615, 2022.
- [19] B. Kim, D. H. Kim, S. H. Park, J. Kim, J.-G. Lee, and J. C. Ye, "Cy-clemorph: cycle consistent unsupervised deformable image registration," *Medical image analysis*, vol. 71, p. 102036, 2021.
- [20] A. Abdollahi, B. Pradhan, and A. Alamri, "Vnet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data," *Ieee Access*, vol. 8, pp. 179 424–179 436, 2020.
- [21] Y. Chen, L. Li, H. Gu, H. Dong, D. L. Nguyen, A. D. Kirk, M. A. Mazurowski, and E. S. Hwang, "Breast density in mri: an ai-based quantification and relationship to assessment in mammography," *arXiv preprint arXiv:2504.15192*, 2025.
- [22] M. B. Imran, S. A. Meo, M. Yousuf, S. Othman, and A. Shahid, "Medical image registration: basic science and clinical implications," *Journal of Ayub Medical College Abbottabad*, vol. 22, no. 2, pp. 199–204, 2010.
- [23] A. Roy, P. K. Roy, A. Mitra, P. K. Maji, S. R. Choudhury, and S.-L. Peng, "Enhancing patient care with modality-based image registration in modern healthcare," in *Smart Medical Imaging for Diagnosis and Treatment Planning*. Chapman and Hall/CRC, pp. 95–136.
- [24] M. V. Wyawahare, P. M. Patil, H. K. Abhyankar *et al.*, "Image registration techniques: an overview," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 2, no. 3, pp. 11–28, 2009.
- [25] J. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical image analysis*, vol. 2, no. 1, pp. 1–36, 1998.
- [26] R. P. Woods, S. T. Grafton, C. J. Holmes, S. R. Cherry, and J. C. Mazziotta, "Automated image registration: I. general methods and intrasubject, intramodality validation," *Journal of computer assisted tomography*, vol. 22, no. 1, pp. 139–152, 1998.
- [27] J. L. Ostuni, R. L. Levin, J. A. Frank, and C. DeCarli, "Correspondence of closest gradient voxels—a robust registration algorithm," *Journal of Magnetic Resonance Imaging*, vol. 7, no. 2, pp. 410–415, 1997.
- [28] B. B. Avants, N. Tustison, G. Song *et al.*, "Advanced normalization tools (ants)," *Insight j*, vol. 2, no. 365, pp. 1–35, 2009.
- [29] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE transactions on medical imaging*, vol. 32, no. 7, pp. 1153–1190, 2013.
- [30] L. Mansilla, D. H. Milone, and E. Ferrante, "Learning deformable registration of medical images with anatomical constraints," *Neural Networks*, vol. 124, pp. 269–279, 2020.
- [31] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," *Medical image analysis*, vol. 12, no. 1, pp. 26–41, 2008.
- [32] S. Ourselin, A. Roche, G. Subsol, X. Pennec, and N. Ayache, "Reconstructing a 3d structure from serial histological sections," *Image and vision computing*, vol. 19, no. 1-2, pp. 25–31, 2001.
- [33] C. Broit, *Optimal registration of deformed images*. University of Pennsylvania, 1981.
- [34] R. Bajcsy and S. Kovačić, "Multiresolution elastic matching," *Computer vision, graphics, and image processing*, vol. 46, no. 1, pp. 1–21, 1989.
- [35] J. C. Gee and R. K. Bajcsy, "Elastic matching: Continuum mechanical and probabilistic analysis," *Brain warping*, vol. 2, pp. 183–197, 1998.
- [36] R. D. Rabbitt, J. A. Weiss, G. E. Christensen, and M. I. Miller, "Mapping of hyperelastic deformable templates using the finite element method," in *Vision Geometry IV*, vol. 2573. SPIE, 1995, pp. 252–265.
- [37] X. Pennec, R. Stefanescu, V. Arsigny, P. Fillard, and N. Ayache, "Riemannian elasticity: A statistical regularization framework for non-linear registration," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2005, pp. 943–950.
- [38] M. Droske and M. Rumpf, "A variational approach to nonrigid morphological image registration," *SIAM Journal on Applied Mathematics*, vol. 64, no. 2, pp. 668–687, 2004.
- [39] G. E. Christensen, R. D. Rabbitt, and M. I. Miller, "Deformable templates using large deformation kinematics," *IEEE transactions on image processing*, vol. 5, no. 10, pp. 1435–1447, 1996.
- [40] Y. Wang and L. H. Staib, "Physical model-based non-rigid registration incorporating statistical shape information," *Medical image analysis*, vol. 4, no. 1, pp. 7–20, 2000.
- [41] E. D'agostino, F. Maes, D. Vandermeulen, and P. Suetens, "A viscous fluid model for multimodal non-rigid image registration using mutual information," *Medical image analysis*, vol. 7, no. 4, pp. 565–575, 2003.
- [42] M.-C. Chiang, A. D. Leow, A. D. Klunder, R. A. Dutton, M. Barysheva, S. E. Rose, K. L. McMahon, G. I. De Zubicaray, A. W. Toga, and P. M. Thompson, "Fluid registration of diffusion tensor images using information theory," *IEEE transactions on medical imaging*, vol. 27, no. 4, pp. 442–456, 2008.
- [43] J.-P. Thirion, "Image matching as a diffusion process: an analogy with maxwell's demons," *Medical image analysis*, vol. 2, no. 3, pp. 243–260, 1998.
- [44] X. Pennec, P. Cachier, and N. Ayache, "Understanding the "demon's algorithm": 3d non-rigid registration by gradient descent," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 1999, pp. 597–605.

- [45] T. Vercauteren, X. Pennec, E. Malis, A. Perchant, and N. Ayache, "Insight into efficient image registration techniques and the demons algorithm," in *Biennial International Conference on Information Processing in Medical Imaging*. Springer, 2007, pp. 495–506.
- [46] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Non-parametric diffeomorphic image registration with the demons algorithm," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2007, pp. 319–326.
- [47] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 11, no. 6, pp. 567–585, 1989.
- [48] ———, "Thin-plate splines and the atlas problem for biomedical images," in *Information Processing in Medical Imaging: 12th International Conference, IPMI'91 Wye, UK, July 7–12, 1991 Proceedings 12*. Springer, 1991, pp. 326–342.
- [49] K. Rohr, H. S. Stiehl, R. Sprengel, T. M. Buzug, J. Weese, and M. Kuhn, "Landmark-based elastic registration using approximating thin-plate splines," *IEEE Transactions on medical imaging*, vol. 20, no. 6, pp. 526–534, 2001.
- [50] D. Rueckert, P. Aljabar, R. A. Heckemann, J. V. Hajnal, and A. Hammers, "Diffeomorphic registration using b-splines," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2006: 9th International Conference, Copenhagen, Denmark, October 1–6, 2006. Proceedings, Part II 9*. Springer, 2006, pp. 702–709.
- [51] P. Bourke, "Cross correlation," *Cross Correlation*, Auto Correlation—2D Pattern Identification, vol. 596, 1996.
- [52] R. A. Fisher *et al.*, "012: A mathematical examination of the methods of determining the accuracy of an observation by the mean error, and by the mean square error." 1920.
- [53] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical review E*, vol. 69, no. 6, p. 066138, 2004.
- [54] J. Yang, Z. Huang, S. Quan, Z. Cao, and Y. Zhang, "Ransacs for 3d rigid registration: A comparative evaluation," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 10, pp. 1861–1878, 2022.
- [55] C. Liu, L. Ma, Z. Lu, X. Jin, and J. Xu, "Multimodal medical image registration via common representations learning and differentiable geometric constraints," *Electronics Letters*, vol. 55, no. 6, pp. 316–318, 2019.
- [56] X. Liu, D. Jiang, M. Wang, and Z. Song, "Image synthesis-based multimodal image registration framework by using deep fully convolutional networks," *Medical & Biological Engineering & Computing*, vol. 57, pp. 1037–1048, 2019.
- [57] S. Miao, S. Piat, P. Fischer, A. Tuysuzoglu, P. Mewes, T. Mansi, and R. Liao, "Dilated fcn for multi-agent 2d/3d medical image registration," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.
- [58] S. Miao, Z. J. Wang, and R. Liao, "A cnn regression approach for real-time 2d/3d registration," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1352–1363, 2016.
- [59] B. D. De Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Işgum, "A deep learning framework for unsupervised affine and deformable image registration," *Medical image analysis*, vol. 52, pp. 128–143, 2019.
- [60] A. Hering, B. Van Ginneken, and S. Heldmann, "mlvирнет: Multilevel variational image registration network," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22*. Springer, 2019, pp. 257–265.
- [61] A. Hering, S. Häger, J. Moltz, N. Lessmann, S. Heldmann, and B. Van Ginneken, "Cnn-based lung ct registration with multiple anatomical constraints," *Medical Image Analysis*, vol. 72, p. 102139, 2021.
- [62] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI'98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings 1*. Springer, 1998, pp. 130–137.
- [63] C. S. Zuo, A. Jiang, B. L. Buff, T. G. Mahon, and T. Z. Wong, "Automatic motion correction for breast mr imaging," *Radiology*, vol. 198, no. 3, pp. 903–906, 1996.
- [64] S. Krishnan, T. L. Chenevert, M. A. Helvie, and F. L. Lonly, "Linear motion correction in three dimensions applied to dynamic gadolinium enhanced breast imaging," *Medical physics*, vol. 26, no. 5, pp. 707–714, 1999.
- [65] B. Eiben, V. Vavourakis, J. H. Hipwell, S. Kabus, C. Lorenz, T. Buelow, N. R. Williams, M. Keshtgar, and D. J. Hawkes, "Surface driven biomechanical breast image registration," in *Medical Imaging 2016: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 9786. SPIE, 2016, pp. 282–291.
- [66] A. Lee, V. Rajagopal, J.-H. Chung, P. Bier, P. M. Nielsen, and M. P. Nash, "Biomechanical modelling for breast image registration," in *Medical Imaging 2008: Visualization, Image-Guided Procedures, and Modeling*, vol. 6918. SPIE, 2008, pp. 299–306.
- [67] G. Zhang, W. Guo, L. Kong, Z. Gong, D. Zhao, C. He, and C. Guo, "Unsupervised learning-based registration for t1 and t2 breast mri images," in *The Fourth International Symposium on Image Computing and Digital Medicine*, 2020, pp. 225–228.
- [68] J. Ying, R. Cattell, T. Zhao, L. Lei, Z. Jiang, S. M. Hussain, Y. Gao, H.-H. S. Chow, A. T. Stopeck, P. A. Thompson *et al.*, "Two fully automated data-driven 3d whole-breast segmentation strategies in mri for mr-based breast density using image registration and u-net with a focus on reproducibility," *Visual Computing for Industry, Biomedicine, and Art*, vol. 5, no. 1, p. 25, 2022.
- [69] C. O. Lew, M. Harouni, E. R. Kirksey, E. J. Kang, H. Dong, H. Gu, L. J. Grimm, R. Walsh, D. A. Lowell, and M. A. Mazurowski, "A publicly available deep learning model and dataset for segmentation of breast, fibroglandular tissue, and vessels in breast mri," *Scientific Reports*, vol. 14, no. 1, p. 5383, 2024.
- [70] R. Jayroe Jr, "Nearest neighbor, bilinear interpolation and bicubic interpolation geographic correction effects on landsat imagery," *Tech. Rep.*, 1976.
- [71] Y. Wang, W. Wan, R. Wang, and X. Zhou, "An improved interpolation algorithm using nearest neighbor from vtk," in *2010 International Conference on Audio, Language and Image Processing*. IEEE, 2010, pp. 1062–1065.
- [72] S. F. Frisken, R. N. Perry, A. P. Rockwood, and T. R. Jones, "Adaptively sampled distance fields: A general representation of shape for computer graphics," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 2000, pp. 249–254.
- [73] Z. Xu and M. Niethammer, "Deepatlas: Joint semi-supervised learning of image registration and segmentation," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*. Springer, 2019, pp. 420–429.
- [74] Y. Li, Y. Fu, I. J. Gayo, Q. Yang, Z. Min, S. U. Saeed, W. Yan, Y. Wang, J. A. Noble, M. Emberton *et al.*, "Semi-weakly-supervised neural network training for medical image registration," *arXiv preprint arXiv:2402.10728*, 2024.
- [75] H. Wang and D. Yee, "I-spy 2: a neoadjuvant adaptive clinical trial designed to improve outcomes in high-risk breast cancer," *Current breast cancer reports*, vol. 11, pp. 303–310, 2019.
- [76] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*. Ieee, 2016, pp. 565–571.
- [77] A. Saha, M. R. Harowicz, L. J. Grimm, C. E. Kim, S. V. Ghate, R. Walsh, and M. A. Mazurowski, "A machine learning approach to radiogenomics of breast cancer: a study of 922 subjects and 529 dce-mri features," *British journal of cancer*, vol. 119, no. 4, pp. 508–516, 2018.
- [78] M. Modat, G. R. Ridgway, Z. A. Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin, "Fast free-form deformation using graphics processing units," *Computer methods and programs in biomedicine*, vol. 98, no. 3, pp. 278–284, 2010.
- [79] C. Liu, K. He, D. Xu, H. Shi, H. Zhang, and K. Zhao, "Regfsc-net: Medical image registration via fourier transform with spatial reorganization and channel refinement network," *IEEE Journal of Biomedical and Health Informatics*, 2024.
- [80] K. A. Eppenhof, M. W. Lafarge, M. Veta, and J. P. Pluim, "Progressively trained convolutional neural networks for deformable image registration," *IEEE transactions on medical imaging*, vol. 39, no. 5, pp. 1594–1604, 2019.
- [81] M. W. Tong, J. Y. Hon, M. M. S. Andreassen, S. Loubrie, A. E. Rodríguez-Soto, T. M. Seibert, R. Rakow-Penner, and A. M. Dale, "Longitudinal registration of t1-weighted breast mri: A registration algorithm (flire) and clinical application," *Magnetic Resonance Imaging*, vol. 113, p. 110222, 2024.
- [82] S. Rankin, "Mri of the breast," *The British journal of radiology*, vol. 73, no. 872, pp. 806–818, 2000.
- [83] S. Nayak and S. Mondal, "Breast mri registration using gorilla troops optimization," in *Human-centric smart computing: Proceedings of ICHCSC 2022*. Springer, 2022, pp. 1–14.