Automated Registration for Dual-View X-Ray Mammography Using Convolutional Neural Networks

William C. Walton, *Member, IEEE*, Seung-Jun Kim, *Senior Member, IEEE*, and Lisa A. Mullen

***Abstract*— *Objective:* Automated registration algorithms for a pair of 2D X-ray mammographic images taken from two standard imaging angles, namely, the craniocaudal (CC) and the mediolateral oblique (MLO) views, are developed. *Methods:* A fully convolutional neural network, a type of convolutional neural network (CNN), is employed to generate a pixel-level deformation field, which provides a mapping between masses in the two views. Novel distance-based regularization is employed, which contributes significantly to the performance. *Results:* The developed techniques are tested using real 2D mammographic images, slices from real 3D mammographic images, and synthetic mammographic images. Architectural variations of the neural network are investigated and the performance is characterized from various aspects including image resolution, breast density, lesion size, lesion subtlety, and lesion Breast Imaging-Reporting and Data System (BI-RADS) category. Our network outperformed the state-of-the-art CNN-based and non-CNN-based registration techniques, and showed robust performance across various tissue/lesion characteristics. *Conclusion:* The proposed methods provide a useful automated tool for co-locating lesions between the CC and MLO views even in challenging cases. *Significance:* Our methods can aid clinicians to establish lesion correspondence quickly and accurately in the dual-view X-ray mammography, improving diagnostic capability.**

***Index Terms*— convolutional neural network, image registration, lesion correspondence, mammography, X-ray.**

I. Introduction

Breast cancer is one of the leading causes of death for women worldwide, with half a million lives lost annually, including more than 40,000 in the United States alone [1]. Early detection has been shown to be critical for less invasive treatment of breast cancer and for saving lives [2]. Hence, tools and techniques that can aid clinicians in early detection of breast cancer are invaluable.

X-ray-based mammography is the main imaging modality used for annual breast cancer screening in asymptomatic women. It is also used for more specialized diagnostic exams, which are performed when suspicious symptoms are present [3]. Conventional mammography involves 2D full-field digital mammography (FFDM) or, in recent years, digital breast tomosynthesis (DBT). DBT involves obtaining several low dose mammographic images across an arc. Reconstruction generates multiple contiguous 1 mm thick slices through the breast, and synthesized 2D images of the entire breast [4].

Mammographic imaging typically involves imaging the breast from at least two different angles. The most frequently used views are the craniocaudal (CC) and the mediolateral oblique (MLO) views. Each view involves physically positioning and compressing the breast between the detector and a compression paddle closer to the X-ray source. The CC view is obtained at an angle of 0 degree from the top to the bottom of the compressed breast and the MLO view is obtained at an angle in the range of 45 to 50 degrees from medial, near the center of the chest, toward the axilla [5]. The breast lesions may be visible in both views or only in one view depending on the lesion location in the breast and the density of the breast tissue. When the breast tissue is very dense, e.g., when it is made up of mostly fibrous and glandular components, it can obscure lesions, as the background breast tissue will have similar X-ray attenuation compared to the lesion. This is in contrast to fatty breast tissue, where lesions have much greater density compared to the surrounding tissue, making the lesions readily visible.

Currently, radiologists screen for breast cancer by analyzing each image for abnormalities, and then searching for the correspondences in the other views [6]. Seeing a lesion in both views is an important feature signaling that the lesion is more likely to be real rather than a false alarm. Additionally, it supports better characterization and localization of the lesion, which is critical. In comparing the two views, radiologists consider certain geometrical features such as the distance between the lesion and the nipple, the clock position of the lesion with respect to the nipple, and the size, shape, and textural composition of the lesion. The position of the patient during the image acquisition procedure is also factored in. In essence, finding correspondences between lesions is predominantly a manual process for the radiologists.

Therefore, automated registration algorithms can significantly enhance the workflow of the radiologists and potentially contribute to improving diagnostic accuracy. Studies show that a significant portion of missed lesions are detected retrospectively, suggesting that an automated algorithm to help locate the lesions in both views could have considerably increased the detectability in earlier exams [7]. This, in turn, could help with determining malignancy, or whether to employ other imaging modalities or a biopsy. Registration will also be

W. C. Walton and S.-J. Kim are with the CSEE Dept. at the University of Maryland, Baltimore County, Baltimore, MD. Walton is also with the Johns Hopkins University Applied Physics Laboratory, Laurel, MD. L. A. Mullen is with the Breast Imaging Division at Johns Hopkins Medicine, Baltimore, MD. This work was supported in part by NSF grant 1631838.

instrumental for guiding targeted ultrasound (US) biopsies and surgical procedures, which require accurate lesion positions. Furthermore, computer-aided diagnosis (CAD) algorithms that involve joint processing (or fusion) of multiple breast images can benefit from an accurate registration module [8], [9].

However, the registration of mammographic images is particularly challenging due to the non-rigid and heterogeneous nature of the breast tissue and the distortions that occur during image acquisition [7]. Conventional registration techniques often fail to account for the complexities involved in how the anatomical features in the compressed breast are projected onto 2D X-ray images [10]. While medical image registration techniques often aim at obtaining one-to-one correspondences, mammographic images involve one-to-many mappings, as a pixel in one mammogram image may correspond to a locus of points in the other [11]. Validation of the registration results can be challenging as it requires the ground truth provided by the experts. Radiologists generally record the truth only for candidate lesion locations, and not for other tissue areas.

Our goal is to develop deep learning-based registration algorithms for two-view X-ray mammography to help clinicians establish lesion correspondence quickly and accurately. Recent advances in machine learning techniques using deep neural networks, in particular, convolutional neural networks (CNNs), achieved remarkable improvement in computer vision tasks. However, challenges still remain in achieving the desired level of accuracy and the best approach has not yet been identified [12]–[14]. In fact, most CNN-based image registration methods have focused on the imaging modalities that capture slices from the same viewing angle, such as the Magnetic Resonance Imaging (MRI) or Computed Tomography (CT) scans [15]–[20]. Very limited research on CNN-based mammographic image registration techniques has been reported in the literature, especially without the use of other imaging modalities [21]–[24].

Our approach is to employ a fully convolutional neural network (FCN) [25], which processes a pair of images from the CC and MLO views, to generate a deformation field that provides a mapping between the two views. A key idea is to incorporate the associated lesion location masks into the training loss function, in the form of a regularizer that captures the distance between the registered lesions. It turns out that our distance-based regularization significantly enhances the network's ability to match the corresponding lesion tissue between the two views. In the operational stage, given a CC and MLO pair, a deformation field is inferred without lesion masks, providing a mapping between masses in the two views.

In our conference precursor [21], we used 2D-projected images of 3D handwritten digit shapes to perform preliminary tests of the CNN-based registration algorithms. In the tests involving real X-ray images, the lesion distance-based regularization was not employed. In this paper, careful performance analysis is carried out in terms of different regularizers, the choice of tissue texture similarity measures, and architectural variations of the CNNs. Furthermore, the methods are tested on different X-ray image types, including conventional 2D X-ray mammography, slices from DBT data (3D mammography), and in silico (i.e., computer-modeled) phantom-based synthetic mammographic images. Performance is characterized for different image resolutions and breast densities, as well as with respect to the lesion attributes such as the size, level of subtlety, and Breast Imaging-Reporting and Data System (BI-RADS) category [26].

In our experiments with 2D X-ray imagery, our techniques achieved registration success rates of up to 90.4%. Lesion correspondence between the CC and MLO views was established reliably even for dense tissue cases. Furthermore, when the network was trained on 2D X-ray imagery and then tested on slices from DBT X-ray imagery, up to 96.7% registration success rates were achieved. Our network outperformed the state-of-the-art CNN-based and non-CNN-based deformable image registration techniques. Experiments with synthetic mammogram images also revealed that they can improve the performance when used to augment real training images.

The rest of this paper is organized as follows. In Sec. II, a brief review of the related works is given. The registration problem is formulated in Sec. III, and our proposed methods are put forth in Sec. IV. The experimental results are presented in Sec. V. Some discussions are given in Sec. VI and conclusions are provided in Sec. VII.

II. Related Works

Medical image registration has been an area of active research [14], [27]–[30]. Prior to deep learning, diffeomorphic non-rigid registration techniques achieved state-of-the-art performance [31]–[33]. Recently, CNN-based approaches gained much attention [15]–[20], [34]–[36]. They typically aim to learn a deformation field that provides a mapping between two or more images, in self-supervised [15], or semi-supervised manners [16], [36]. Spline-based interpolation was employed in [17], [19], [34], whereas the deformation vectors of individual pixels were estimated directly in [15], [16], [37]. Some ingested full image frames [36], while others operated on patches [15], [18], [20]. Despite the significant developments, however, existing CNN-based registration techniques have mostly focused on images taken from the same viewing angle, and not many addressed the registration of breast tissue.

Breast image registration poses unique challenges due to the inhomogeneous, anisotropic, soft-tissue, and non-rigid nature of breast tissue [7], which, combined with the physical compression and patient position alteration, results in significant diversity in the tissue appearances and displacement patterns. Thus, it is often advocated to incorporate other data modalities such as the MRI, DBT, and US [38]–[40], as well as various modeling assumptions [41]. When it comes to the registration of the CC and MLO views using only X-ray images, early non-CNN efforts focused on finding correspondence between lesions in different views [42]–[44].

Most CNN-based techniques that process multiple mammographic views have been geared towards malignancy classification [45]. Two CNNs were employed to process a full mammographic X-ray image as well as their patches for malignancy detection [46]. A Siamese CNN architecture was employed for classifying matching versus non-matching pairs of lesions between the CC and MLO views [9]. The symmetry

information in the CC and MLO views of both left and right breasts was exploited for cancer screening [47]. While these works highlight the benefit of processing two or more mammographic views, they do not specifically tackle the problem of multi-view tissue/lesion registration.

To our knowledge, only limited research has been done on CNN-based techniques for registering the CC and MLO views. A U-Net was employed in [22] to register mammographic images from the same view (e.g., CC to CC or MLO to MLO). In [23], an affine transformation was learned for registration of CC and MLO views in a semi-supervised manner using a spatial transformer module [48]. Densely connected CNN blocks with shared weights were employed to obtain discriminative features and find correspondence between detected masses in CC and MLO views in [24].

Some recent CNN-based medical image registration techniques share important characteristics with our proposed CC-to-MLO registration method, such as the use of FCNs with skip connections and the incorporation of additional ground truth mask images for training. In [15], voxel-wise registration of MRI slices of brain tissue was proposed using a FCN with a skip architecture. The method did not use other ground truth labels, but took a self-supervised approach based on an intensity-based similarity measure with total variation regularization. The registration of cardiac features in MR images was tackled using a U-Net architecture in [35]. A diffeomorphic parameterization of the deformation field was adopted and the ground truth segmented shapes of anatomical features were employed to aid training. The algorithm in [16] also involved a U-Net architecture, which, in addition to the input image pair, optionally incorporated a pair of binary segmentation masks of anatomical features into the regularization term of the training cost. The authors of [36] similarly utilized binary anatomical label images in addition to input image pairs in a U-Net-like architecture to perform multi-modal image registration of MR and US images of the prostate region. They adopted a multi-scale Dice similarity measure, based on Gaussian-blurred versions of the fixed and warped label images for training but did not use intensity-based similarity measures.

III. Problem Formulation

A moving (source) image $I_m(\boldsymbol{x})$ and a fixed (target) image $I_f(\boldsymbol{x})$ are defined with 2D pixel coordinates $\boldsymbol{x} \in \Omega \subset \mathbb{R}^2$. Our goal is to learn a function $D_{\boldsymbol{\theta}}(I_f, I_m) = d(\boldsymbol{x})$, represented by a CNN with parameter vector $\boldsymbol{\theta}$, which yields a deformation field $d : \Omega \to \Omega$ that warps the moving image to match the fixed one. That is, it is desired that $(I_m \circ d)(\boldsymbol{x})$ is similar to $I_f(\boldsymbol{x})$ in terms of a suitable similarity measure S. In order to govern the nature of the resulting deformation field based on prior knowledge, a regularizer $R(d)$ is also incorporated. The loss function is then defined as

$$L(I_f, I_m) := -S(I_f, I_m \circ D_{\boldsymbol{\theta}}(I_f, I_m)) + \lambda R(D_{\boldsymbol{\theta}}(I_f, I_m)) \tag{1}$$

where $\lambda \geq 0$ is a weight to balance the similarity and the regularization terms. The CNN training amounts to solving

$$\min_{\boldsymbol{\theta}} \ \mathbb{E}_{\mathcal{D}}\{L(I_f, I_m)\} \tag{2}$$

where $\mathbb{E}_{\mathcal{D}}\{\cdot\}$ represents taking an average with respect to the data set $\mathcal{D}$ of the fixed and moving image pairs (I_f, I_m).

Although (1) is formulated to obtain a pixel-wise mapping $d(\boldsymbol{x})$ for each image pair, our goal is not so much to achieve precise pixel-level registration, as to establish a useful correspondence between regions of interest, such as lesions.[1] That is, our objective is that when a clinician selects a candidate lesion location $\boldsymbol{x}$ in one view, the trained network can present a likely location $d(\boldsymbol{x})$ of the lesion in the other view accurately.

A. Similarity Measure

Two alternative similarity measures are considered in our work. The similarity measure based on the sum absolute error (SAE) is the negative of the average difference in the individual pixel intensities, defined as

$$S_{SAE}(I_1, I_2) := -\frac{1}{|\Omega|} \sum_{\boldsymbol{x} \in \Omega} |I_1(\boldsymbol{x}) - I_2(\boldsymbol{x})| \tag{3}$$

where $|\Omega|$ is the number of pixels in the images.

The second one is normalized cross-correlation (NCC) which is also known as the Pearson correlation coefficient. Upon defining the mean intensity of image I as $\overline{I} := |\Omega|^{-1} \sum_{\boldsymbol{x} \in \Omega} I(\boldsymbol{x})$, the NCC is defined as

$$S_{NCC}(I_1, I_2) := \frac{\sum\limits_{\boldsymbol{x} \in \Omega} (I_1(\boldsymbol{x}) - \overline{I_1})(I_2(\boldsymbol{x}) - \overline{I_2})}{\left[\sum\limits_{\boldsymbol{x} \in \Omega} (I_1(\boldsymbol{x}) - \overline{I_1})^2\right]^{\frac{1}{2}} \left[\sum\limits_{\boldsymbol{x} \in \Omega} (I_2(\boldsymbol{x}) - \overline{I_2})^2\right]^{\frac{1}{2}}} \tag{4}$$

which yields a value between -1 and 1.

The SAE measure is suitable for comparing images that have similar intensity distributions, and is more robust to large intensity differences due to outliers than the sum of squared errors, since the latter places far more emphasis on the pixels with large absolute residuals [49]. The NCC metric is more suitable when the intensity and contrast distributions vary significantly over images as the denominator in (4) effectively normalizes the measure [50].

B. Regularization

1) Total Variation Regularization: Regularization allows for the incorporation of prior knowledge on the learned deformation fields and also prevents overfitting when the number of parameters in $\boldsymbol{\theta}$ is large, which is often the case with deep CNNs. Non-rigid deformation field-based registration algorithms commonly employ a form of smoothness-promoting regularization on the deformation field. For instance, Tikhonov regularization can be employed to penalize the ℓ_2-norm of the Jacobian of the deformation field, enforcing smoothness [51]. When sharp transitions are expected, Tikhonov regularization may not be suitable. In our context, the anisotropic total

[1] In this work, we focus on masses, corresponding to one of the two main categories of breast lesions. Masses and calcifications are commonly treated separately in mammographic image analysis research.

variation regularization (TVR) is considered, which is the ℓ_1-norm of the deformation field Jacobian $\nabla d(\boldsymbol{x})$, given by

$$R_{TVR}(d) := \|\nabla d(\boldsymbol{x})\|_1 \tag{5}$$

where $\|\cdot\|_1$ is the sum of the absolute values of all entries in the Jacobian matrix. It can handle large, non-smooth displacements, which can occur in mapping the anatomical features between the CC and the MLO images [52].

2) Incorporating Ground Truth Lesion Masks: The second regularization we considered utilizes the ground truth lesion locations in the CC and MLO views, which are provided through the lesion masks available with the CC/MLO image pairs. Lesion masks have been exploited in CNN-based registration algorithms by capturing the amount of overlap of known anatomical features after registration [16], [36]. In the dual-view X-ray mammography, however, the amount of overlap may not provide a strong enough supervision signal due to the severe distortions in the mammographic views. (Indeed, there may be no overlap at all.) Instead, we propose a new regularization function, termed distance-based regularization (DBR), which penalizes the (normalized) distance between the lesion locations after registration.

Let $\Lambda_f^{(n)} : \Omega \to \{1, 0\}$ be the mask image that has the pixel intensity of 1 within the n-th lesion, and 0 outside, in the fixed view I_f. In the paired moving image I_m, $\Lambda_m^{(n)} : \Omega \to \{1, 0\}$ represents the corresponding lesion mask. Define the centroid $\mu(\Lambda) \in \mathbb{R}^2$ of a mask Λ as

$$\mu(\Lambda) := \frac{\sum_{\{\boldsymbol{x}:\Lambda(\boldsymbol{x})=1\}} \boldsymbol{x}}{\sum_{\boldsymbol{x}\in\Omega} \Lambda(\boldsymbol{x})}. \tag{6}$$

Then, the DBR function is defined as

$$R_{DBR}(d; \{\Lambda_f^{(n)}, \Lambda_m^{(n)}\}) = \frac{1}{N}\sum_{n=1}^{N} \frac{\|\mu(\Lambda_f^{(n)}) - \mu(\Lambda_m^{(n)} \circ d)\|_1}{\|\mu(\Lambda_f^{(n)}) - \mu(\Lambda_m^{(n)})\|_1} \tag{7}$$

where N is the number of lesions in the given image pair (I_f, I_m). Note that in general there can be zero, one, or more than one annotated lesions in (I_f, I_m). The regularization function is simply set to zero if there are no lesions annotated. For a term inside the sum in (7), the numerator is the distance between the centroids of the lesions in the fixed view and in the moving view *after* the warping is done according to d. Hence, if the displaced lesion pixels are close to the lesion location in the other view, the penalty will be low. The ℓ_1-norm-based Manhattan distance is adopted as it is less sensitive to the outliers than e.g. the Euclidean distance[2]. On the other hand, the denominator is the distance between the lesion centroids *before* registration. Thus, the regularizer penalizes more heavily the case where the ground truth displacement is small, and more leniently when a large displacement is expected. This provides the balance necessary for training the network to work well in all cases. It is also emphasized that the DBR is used only in the training, and not in the operational testing stage.

[2]The modified Hausdorff distance, which computes the distance accounting for the entire set of points, was also tested. The results are presented in App. H in the Supplementary Material.

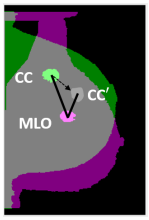


Fig. 1: Lesion distances before and after registration. The CC and the MLO views are superimposed and the ground truth lesion masks in the CC/MLO views, as well as the warped CC lesion (marked as CC′), are indicated.

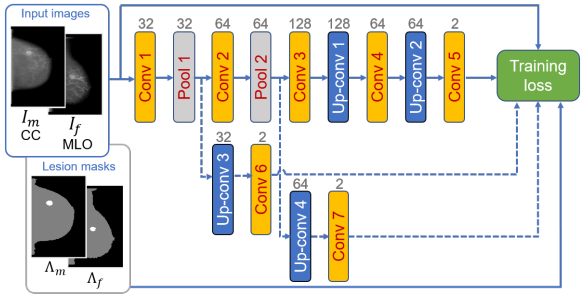


Fig. 2: The proposed CNN architecture.

Fig. 1 illustrates the example lesion masks in an overlapped CC/MLO image. The lesions marked as CC and MLO are the ground truth lesions, and the one marked with CC′ is the warped version of the CC lesion. In this example, it can be seen that the distance between the MLO and the CC′ lesions has been reduced by registration, but they still do not overlap. Thus, a regularization function based on the amount of lesion overlap will not provide any supervision signal in this case, while our DBR function can still capture useful information.

IV. Proposed CNN Architectures

A CNN is often used for image classification, where the input is an image and the output is a class label for the image. The CNN architecture processes the input image through multiple layers of convolution, pooling, and element-wise nonlinear operations, to produce discriminative features. The features are then fed to fully-connected layers to produce the label. In our work, the CNN is adopted to generate a deformation field $d(\boldsymbol{x})$ based on the input images $I_f(\boldsymbol{x})$ and $I_m(\boldsymbol{x})$, where the inputs and the outputs are defined over the same domain $\boldsymbol{x} \in \Omega$. That is, the inputs and the outputs are "images" of the same size. In FCNs, instead of fully-connected final layers, upsampling convolution (or up-convolution) layers are employed to yield an output that is of the same size as the input [25], [53]. Therefore, FCNs are natural candidates for CNN-based image registration [14]–[16].

Fig. 2 shows our CNN architecture. Similar to [25], [15], two variants are considered. One is the serial architecture, consisting of a single path of layers from the input to the output, which is depicted by the solid arrows in Fig. 2. The other variant is the skip architecture, which has additional branches for tapping the features at various depths of layers, as indicated by the dashed arrows in Fig. 2. The number on top of each layer in Fig. 2 represents the number of the feature maps at the output of the layer. The convolution layers also include batch normalization and nonlinear activation using the rectified linear units (ReLUs), except for the final convolution layers

in all paths (i.e., Conv 5, 6, and 7 layers). The kernel sizes and the strides at the individual layers depend on the input image size. The table of kernel sizes and strides is included in App. A in the Supplementary Material.

The input to both architectures is the pair of CC/MLO images (which are input as two channels). The network output is $\Delta d(\boldsymbol{x})$, which captures the *relative* displacement of pixel $\boldsymbol{x}$ in the moving image. The deformation field is given as

$$d(\boldsymbol{x}) := \boldsymbol{x} + \Delta d(\boldsymbol{x}). \tag{8}$$

During the training, only the displacements arriving at the pixels within the image boundary are actually employed. When DBR is used, the available lesion masks are incorporated.

A. Serial Architecture

The serial architecture contains five convolution layers, two pooling layers, and two up-convolution layers. The final layer results in two feature maps that correspond to the vertical and the horizontal displacements in $\Delta d(\boldsymbol{x})$. The training loss for the serial architecture is given by

$$\begin{aligned} L_0(I_f, I_m) = &- S(I_f, I_m \circ d) + \alpha R_{TVR}(d) \\ &+ \beta R_{DBR}(d; \{\Lambda_f^{(n)}, \Lambda_m^{(n)}\}) \end{aligned} \tag{9}$$

where α and β are nonnegative weights for balancing the regularization terms.

B. Skip Architecture

The branching paths in the skip architecture capture the higher-resolution features from the early, shallow layers, which are combined with the lower-resolution yet larger-scale features obtained by the deeper layers. The skip connections help with predicting fine details in the output [25]. In our implementation, two additional branches are taken at the outputs of the first and the second pooling layers, which are appropriately upsampled to match the resolution of the output of the main path. Let us denote the deformation fields from the first and the second skip paths as $d_1(\boldsymbol{x})$ and $d_2(\boldsymbol{x})$, respectively. Define the loss function L_p for the p-th skip path in the same way as in (9), with d replaced by d_p, where $p \in \{1, 2\}$. Then, the skip architecture is trained based on the overall loss function $\overline{L}$ that averages the individual paths' losses using nonnegative weights $\{\mu_p\}_{p=0}^2$ as

$$\overline{L}(I_f, I_m) = \sum_{p=0}^{2} \mu_p L_p(I_f, I_m). \tag{10}$$

V. Experiments

A. Experiment Setup

1) Data Sets: Three X-ray image data sets were utilized in our experiments. The primary data set is the Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM), a publicly available set of digitized scanned-film mammography data, curated by trained mammographers [54]. The data set includes the CC and MLO X-ray image pairs for each breast along with corresponding binary image masks indicating the lesion locations. Information describing each image view, lesion type, pathology, and diagnosis is also provided.

The second data set consists of a limited number of de-identified DBT images with accompanying lesion location and diagnostic information, obtained as part of a research effort in Johns Hopkins Medicine (JHM) [IRB00185772, 12/3/2018].

The third data set involves synthetic mammogram images generated using software tools developed through the Virtual Imaging Clinical Trial for Regulatory Evaluation (VICTRE) project in the United States Food and Drug Administration (FDA) [55]. In-silico X-ray images were generated from 3D phantoms, simulating physical compression of the breast, different imaging angles, and insertion of lesions.

Table I summarizes some details regarding the images used in the experiments, which will be explained later.

2) Preprocessing: Prior to ingestion to the networks, the images were preprocessed. First, the images were re-oriented so that the chest is on the left and the nipple is on the right side of the image frame. Then, artifacts such as the burned-in annotations were removed using a simple histogram-based technique that detects bright pixels in the narrow regions along the top, bottom, and right sides of the image. Next, the breast tissue boundary was extracted in order to generate a breast tissue mask, which is useful for limiting processing to only the breast tissue areas, and thus, improving the training process. The nipple location and its distance to the chest wall were also detected to help distinguish between multiple lesions in a breast when such cases arise. In our experiments, however, only single-lesion cases were used due to the limited availability of ground truth for evaluating multiple-lesion cases. For the MLO images, an extra step is applied to detect and mask the pectoral muscle, which is outside the breast tissue area, to prevent the networks from processing this area. For the training data, slight rotations (up to $\pm 15°$) were applied as a means of data augmentation to generate extra training images. More details on the last two steps are provided in App. B of the Supplementary Material. The images were resampled to the resolutions of 330×220, 660×440, and 990×660, for comparison of the registration performance under different image and mini-batch sizes. The original image sizes are shown in Table I. The pixel intensities were also normalized to the range of $[0, 1]$.

3) Training: The networks were implemented in MATLAB. For training, Adam optimizer was employed with an initial learning rate of 0.001 and random weight initialization [56]. Around 60 to 100 epochs were used for training, with the mini-batch sizes ranging from 8 to 32, based in part on the available GPU memory. For the skip architecture, the weights for the main path and the first and second skip branches were set to $\mu = 0.22$, $\mu = 0.13$, and $\mu = 0.65$, respectively. Here again, although our training formulation accommodates multiple lesions per image pair [cf. (7)], we used only the images with single lesions, due to the lack of ground truth information on lesion correspondence. Matching multiple lesions per image pair for training is left for future research.

TABLE I: Image sizes and counts in data sets.

Data sets	Average original image size	Original image resolution (μm)	Number of training pairs	Augmentations per image	Total training pairs	Validation / test pairs
CBIS-DDSM	5280 × 3131	N/A	496	8	4464	146 / -
JHM DBT slices	2457 × 1975	70	-	-	-	- / 60
FDA synthetic	2000 × 1500	76	2250	4	11250	103 / 95

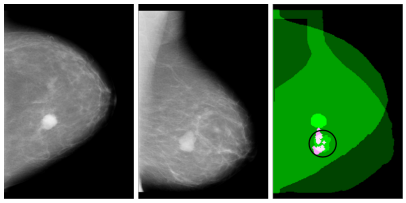


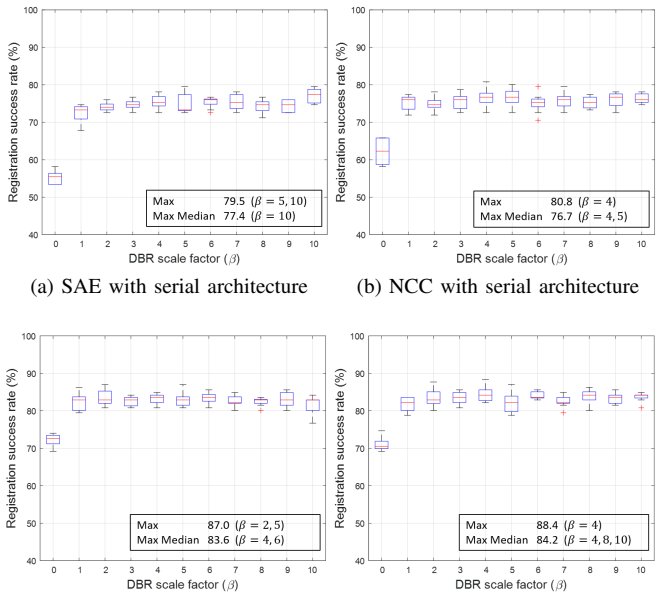
Fig. 3: Determining the registration success. (a) Input CC view with a lesion. (b) Input MLO view with the corresponding lesion. (c) The projected pixels in magenta fall inside the ROI indicated by the black circle in the CC/MLO overlay.

4) Testing: To assess the registration performance, a metric based on the ground truth lesion locations is defined. This is motivated by the clinical usage of registration, where clinicians desire to quickly establish correspondence between the candidate lesion locations in different views. Specifically, for a given pair of test images, the deformation field computed by the trained network is applied to the moving image. In particular, the pixels in the lesion locations in the moving image are translated to hopefully match the lesion pixels in the fixed image. A region of interest (ROI) is defined in the fixed image as a disc centered around the ground truth lesion, with a radius equal to 7.5% of the height of the image. The registration is deemed successful if any of the translated pixels fall inside the ROI. The performance metric is the percentage of the image pairs with successful registration. A sensitivity analysis, in terms of the ROI size and the fraction of overlapping pixels, is provided in App. C in the Supplementary Material. See also the related discussion in Sec. VI.

In the case of the skip architecture, the union of the pixels translated by the three deformation fields is used. The union is preferred to the average as the centroid of the union turns out to be usually closer to the target lesion location than the centroid of the average. Although the union results in a somewhat larger displacement area, it is still substantially smaller than the entire breast tissue area, and thus is useful for the clinicians for finding the lesion locations.

Fig. 3 illustrates the test process. Fig. 3(a) is a CC view with a visibly distinct lesion. Fig. 3(b) is the corresponding MLO view. In Fig. 3(c), the masks for the CC and the MLO view lesions are superimposed, where the translated lesion pixels are depicted in magenta. As some of (in fact, in this example, most of) the magenta pixels fall inside the ROI indicated by the black circle, the registration is counted successful.

It is noted that medical image registration algorithms are often assessed using the mean-square error (MSE) or the Dice metrics [7]. However, for breast image registration, there are significant distortions and occlusions in different views due to the non-rigid nature of the breast and the physical compression process. Thus, direct pixel-wise comparison may be too strict to reflect the registration performance meaningfully. Our performance metric takes advantage of the ground truth lesion masks provided with curated mammography data sets, and falls in with the fact that even rough lesion location correspondence in an automated fashion can be very helpful in clinical settings. More discussion on the metric can be found in Sec. VI.



(a) SAE with serial architecture (b) NCC with serial architecture

(c) SAE with skip architecture (d) NCC with skip architecture

Fig. 4: Registration success rates versus β. Each box indicates the range of 25 to 75 percentiles with the red line showing the median. The outliers depicted by the red plus markers are roughly outside the $\pm 2.7\sigma$ range.

B. CBIS-DDSM Data Set

Several sets of experiments were conducted with the CBIS-DDSM data. We maintained the CBIS-DDSM's established division of training and test data. For training, 496 CC and MLO image pairs containing single masses were selected, which we increased to 4,464 pairs using augmentation, as noted in Table I. Similarly, 146 pairs of test images, with single masses, were used from the designated test set. As the breast tissue and lesions occur in a wide variety of sizes, shapes, and characteristics, instead of further sub-dividing the data set into separate training and validation sets, the test data set was used also for determining the best model parameters and the optimal validation performances are reported. It is later verified that the model does not overfit by testing the trained networks on a completely different data set in Sec. V-C.

1) Parameter Tuning: Parameters α and β in (9) were tuned based on the registration success rate. Since DBR was found to have a greater influence on the performance than TVR, the DBR parameter β was first tuned without TVR (that is, with $\alpha = 0$). Then, α was optimized in search of further

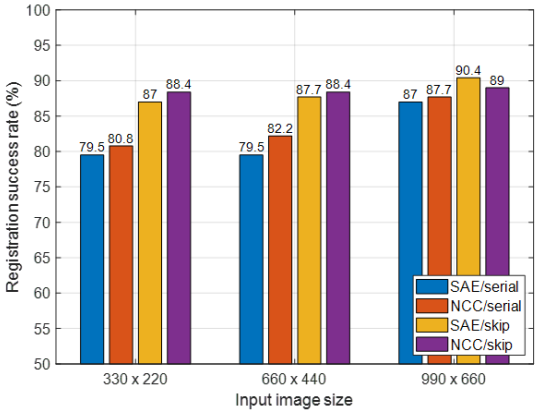


Fig. 5: Performance versus input image size.

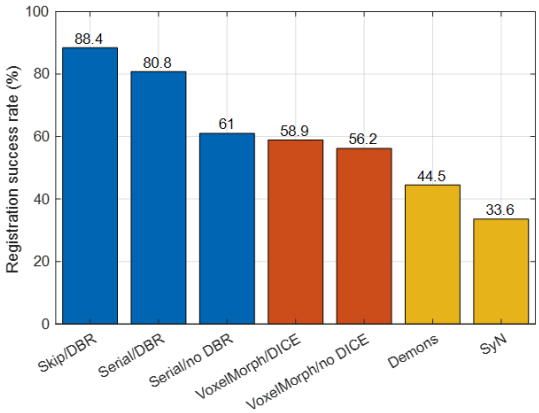


Fig. 6: Performance comparison with existing algorithms.

performance improvement. We used the input resolution of 330×220, as our preliminary trials found that the results gave a good indication of the trends at higher resolutions. Furthermore, numerous parameter combinations could be experimented with lower training burden.

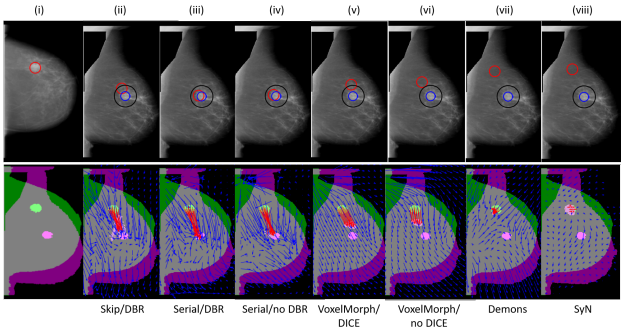
Fig. 4 depicts the registration success rates at different levels of DBR by adjusting β for the serial and the skip architectures and using the SAE and NCC similarity measures. For each β value, nine training trials were performed with different random initializations of the network weights. The blue boxes in Fig. 4 represent $[25, 75]$ percentile ranges of the resulting success rates, and the red lines indicate the medians. It can be seen that the performance improves significantly with DBR compared to the cases with the similarity measure alone ($\beta = 0$). This is because DBR guides the networks to recognize the lesions and put more effort toward registering them correctly.

It turns out that for optimal DBR settings, incorporating TVR does not yield significant improvement in performance beyond what was obtained with DBR. For this reason, we set $\alpha = 0$ henceforth for simplicity. Performance plots involving the tuning of α for the optimal β settings are presented in App. D in the Supplementary Material.

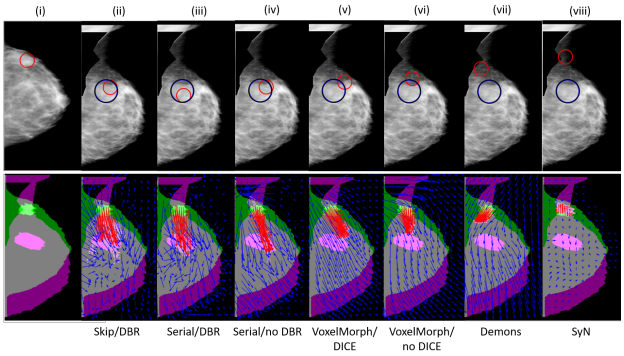
2) Performance of Proposed Networks: Fig. 5 shows the highest registration success rates observed from the proposed algorithm with different input image resolutions, similarity measures, and network architectures using optimal DBR parameters. It can be seen that as the input image size increases, the registration performance also improves in general. The NCC measure generally yields better success rates than SAE. Between the serial and the skip architectures, it is clear that the skip architecture outperforms the serial architecture, which can be attributed to the multiple resolutions in the deformation fields obtained from different skip levels. However, the skip architecture has higher computation and memory requirements.

3) Comparison with Existing Algorithms: The registration performance of our algorithm was compared to those of three existing non-rigid medical image registration methods. First, Thirion's Demons algorithm as implemented in the MATLAB `imregdemons` function was compared [32], [57]. Demons algorithm is a diffeomorphic image registration technique, which performs an iterative optimization for each image pair using a multi-resolution pyramid approach. The symmetric image normalization (SyN) method was also tested using the implementation in the Advanced Normalization Tools (ANTs) software package [33]. The SyN method is another diffeomorphic technique, based on maximizing cross-correlation within the space of diffeomorphic maps. Finally, the VoxelMorph algorithm was tested, which is based on the U-Net CNN architecture. It can also incorporate the ground truth anatomical features into the training based on a Dice metric [16].

The registration performances achieved with the input image size of 330×220 are shown in Fig. 6. For our technique, NCC was employed and the results both with and without DBR ($\beta = 4$ and $\beta = 0$, respectively) are shown. TVR was not enabled ($\alpha = 0$). For VoxelMorph, we experimented with the MSE and NCC similarity metrics and a range of diffusion regularizer weights [16]. This was done both with and without the Dice-based regularization. The results using MSE, with a diffusion regularization weight of $\lambda = 0.01$, with and without the Dice-based regularization, were selected as these yielded the highest registration success rates. It can be seen that our proposed techniques significantly outperform the existing methods. Comparing the CNN-based methods (our methods and VoxelMorph) with the traditional diffeomorphic registration algorithms, one can clearly observe the superiority of the CNN-based approaches. Note that the diffeomorphic methods are computationally more intensive than the CNN-based ones due to their iterative optimization. The CNN-based methods take 30 msec or less for registering images of size 330×220, while Demons algorithm takes about 500 msec and SyN around 5 sec., using an Intel Xeon CPU @ 2.20 GHz. More importantly, it can be seen that our proposed techniques perform much better than VoxelMorph, especially with DBR. As discussed in Sec. III-B.2, due to the significant distortions in mammographic images, the network training may not generate deformation fields that can move the source lesion pixels far enough to actually overlap with the target lesions. In such cases, the Dice metric will not provide useful signals for training, while our DBR metric can still quantify the relative quality of registration by means of the distance between the moved and the target lesions. Interestingly, our serial network architecture without DBR still outperforms the VoxelMorph algorithm with Dice regularization, although the margin becomes narrower compared to the case with DBR. This shows that our optimized CNN architecture already has



(a) Fatty tissue example



(b) Dense tissue example

Fig. 7: Exemplar registration results.

merits over the VoxelMorph architecture for our application. Note that VoxelMorph was designed for brain MRI data, which involve much more correlated images with far less deformation due to the same viewing angle.

Some typical registration results obtained from the proposed and the existing methods are depicted in Fig. 7, based on the input size of 330×220. Fig. 7(a) shows the case of fatty tissue and Fig. 7(b) dense tissue. In either plot, the top panel in column (i) is the CC view with the lesion encircled in red. The bottom panel in the same column displays the overlay of the CC and the MLO masks, highlighting the locations of the lesions in green for CC and in magenta for MLO, respectively. Columns (ii)–(viii) show exemplar registration results from the various algorithms and architectures tested.

The top panels in columns (ii)–(viii) depict the MLO images with the ground truth lesion location encircled in blue and the evaluation metric ROIs indicated by black circles. (In Fig. 7(b), the size of the MLO lesion is similar to the ROI size, rendering the blue and the black circles to almost coincide.) The red circles indicate the centroid locations of the displaced CC lesion pixels with their diameters set equal the lesion size in the CC view. In the bottom panels, the projected lesion pixels are shown in magenta. For the skip architecture, the union of the displaced pixels from all three branches are depicted. The deformation vectors are displayed in quiver plots, showing the individual pixel displacements. The arrows for the lesion pixels are shown in red with higher density. For the skip architecture, the arrows corresponding to the average of the deformation vectors are shown for simplicity of visualization.

It can be seen from Fig. 7 that our registration networks place the CC lesion pixels closer to the target MLO lesion

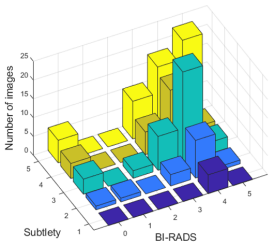


Fig. 8: Image counts by BI-RADS and subtlety categories.

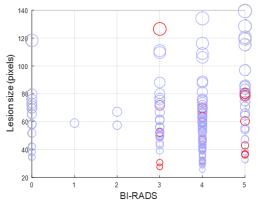


Fig. 9: Lesion sizes for BI-RADS categories. The circle diameters are proportional to lesion sizes. Blue and red denote successful and unsuccessful registrations, respectively, using the serial architecture.

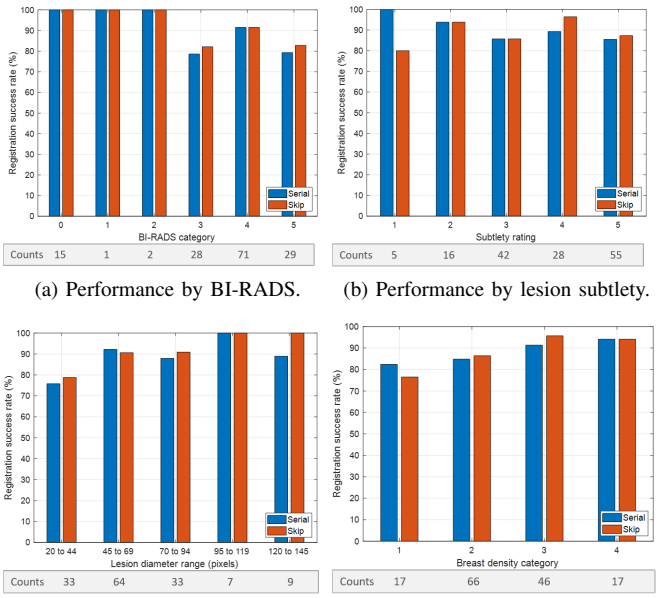
location, compared to the other algorithms. Overall, this is consistent with the registration performance results in Fig. 6. Even in the dense tissue case, where the lesions are more difficult to discern, our algorithms are seen to more closely land the displaced CC lesions at the location of the MLO lesions. Note that VoxelMorph does move the CC lesions in the direction toward the MLO lesions, but often falls short of making the full distance.

4) Performance by Lesion/Tissue Characteristics: The registration performance of the proposed method was also evaluated based on the lesion BI-RADS assessment categories, lesion subtlety ratings, lesion sizes, and the density of breast tissue. BI-RADS is a categorization system in breast imaging, including mammography. BI-RADS assessment categories range from 0 to 6, with 0 denoting incomplete assessment, 1 negative, 2 benign, 3 probably benign, 4 suspicious, 5 highly suggestive of malignancy, and 6 biopsy proven malignancy. Only BI-RADS categories 0 to 5 were available in our subset of the CBIS-DDSM data set.

BI-RADS also provides an assessment of breast tissue density with four density descriptors: 1 almost entirely fatty, 2 scattered fibroglandular, 3 heterogeneously dense, and 4 extremely dense [26]. Dense tissue makes it more difficult to identify masses, as dense tissue can overlap and obscure breast lesions. The CBIS-DDSM data set also contains subtlety ratings, which are not part of BI-RADS, but were generated by experienced radiologists. The subtlety ratings range from 1 (subtle) to 5 (more obvious). We also categorized the lesions by their diameters into five groups. The registration performances are based on the 990×660 resolution images with NCC and DBR. TVR was not used ($\alpha = 0$).

To aid the interpretation of the results, the distribution of the test data set is presented first. Fig. 8 shows the number of images with each BI-RADS and subtlety category. It can be seen that most lesions in the CBIS-DDSM test set are in the higher BI-RADS categories and appear more obvious. Fig. 9 shows the lesion size distributions for each BI-RADS category. The blue and red circles denote successful and unsuccessful registrations, respectively, using the serial architecture.

Fig. 10a shows the success rates for different BI-RADS categories using both the serial and the skip architectures. The number of images in each category is indicated at the bottom of the figure. The success rates are seen to be uniformly above 78% in all cases. Note that the images with ratings 1 or 2 are too few to draw meaningful conclusions. Fig. 10b shows the



(a) Performance by BI-RADS. (b) Performance by lesion subtlety.

(c) Performance by lesion size. (d) Performance by breast density.

Fig. 10: Registration success rates by lesion and tissue characteristics.

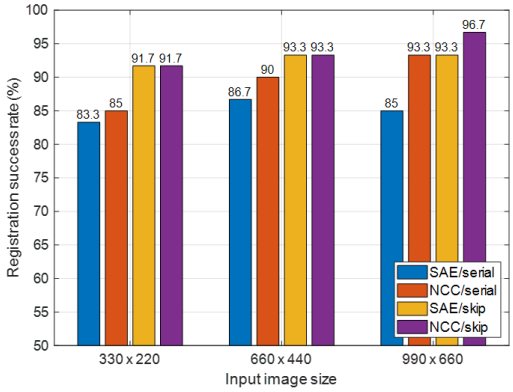


Fig. 11: Registration success rates for JHM DBT slices.

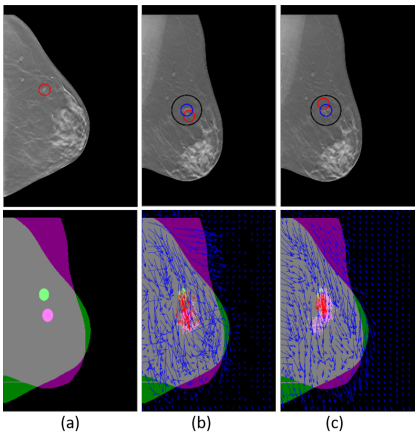


Fig. 12: Exemplar registration results for DBT slices based on model trained on CBIS-DDSM images. (a) Input CC image (top) and the CC/MLO mask overlay (bottom). (b) Result from serial architecture. (c) Result from skip architecture.

performance versus subtlety categories. Again, it is seen that rates above 80% are achieved across the board.

Fig. 10c depicts the success rates by the diameters of the lesions[3] It can be seen that the registration success rate generally increases with the lesion size. In the largest diameter range of $[120, 145]$ pixels, only one of nine lesions was misregistered by the serial architecture, and a close examination of this case makes us question the accuracy of the ground truth based on the nipple-to-lesion distance in each view. In fact, our algorithms actually appear to map the CC lesion close to the expected location in the MLO view. Regarding the slightly lower success rates in the smallest-size group with the diameter range of $[20, 44]$ pixels, the registration algorithm had difficulty especially when the small lesions were located far away in the two views.

Fig. 10d shows the registration performance for different categories of breast tissue density. It can be seen that high registration success rates are maintained even for dense tissue (category 4). This is encouraging since the dense tissue presents greater challenges for radiologists in discerning the lesions. Further examination of the extremely dense cases in our test set revealed that there were not as many small lesions and not as many lesions with large separation distances between the two views. The lack of cases with ground truth for very small lesions in extremely dense tissue is conceivable in that these would be difficult for clinicians to identify visually. Hence, these factors give some explanation for the high registration success rates in this category. The somewhat lower performance in the fatty breast tissue category is attributed to a few image pairs with very small lesions and yet another case with questionable ground truth. In general it is encouraging that our algorithm performs relatively well across the board.

Several additional factors should be considered in interpreting or comparing the registration results, in particular with respect to Figs. 10a and 10d. For instance, BI-RADS and breast density assessments can vary among clinicians [58], [59]. Additionally, the BI-RADS assessment category for lesions may be updated at different stages of the examination process [60]. The BI-RADS scale is also nonlinear [61]. These and other factors are discussed in detail in App. E in the Supplementary Material.

[3]The pixel resolution in microns was not available for the CBIS-DDSM data set. Hence, the diameters were measured in pixels.

C. Experiments with Other Data Sets

1) Experiments with DBT Slices: We also applied the networks, which were trained on CBIS-DDSM images, to a completely different data set in order to see how well the networks can generalize. The JHM DBT data set is a 3D mammography data set containing the CC and MLO cubes for 60 patient cases. Each case involved a single mass-type lesion. One slice from each CC cube and another from the corresponding MLO cube were extracted such that the slices intersect with a lesion. (The result was not very sensitive to which slice in a cube was used as the DBT slices are highly correlated across slices.) The images were then resampled to the input resolution for which the networks were trained. Note that the CC/MLO slices from DBT cubes were used only for testing, and not for training.

Fig. 11 shows the success rates achieved on DBT data based on the the same configurations and models used for Fig. 5. It

TABLE II: Registration success rates by lesion size for JHM DBT slices.

Diameter (pixels)	$22 \sim 44$	$45 \sim 69$	$70 \sim 94$	$95 \sim 119$	$120 \sim 145$
Avg. diam. (cm)	0.6	1.1	1.6	2.1	2.5
Serial arch.	71.4%	93.3%	93.8%	100%	100%
Skip arch.	71.4%	100%	100%	100%	100%
Image counts	7	15	16	13	9

TABLE III: Registration success rates by breast density for JHM DBT slices.

Breast density category	1	2	3	4
Serial architecture	0%	90%	97.3%	100%
Skip architecture	0%	100%	97.3%	100%
Image counts	1	20	37	2

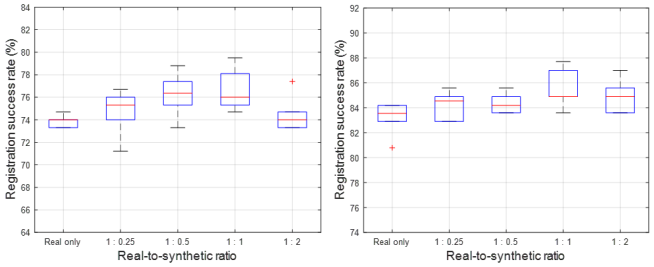
can be seen that the success rates are generally comparable to those achieved on the CBIS-DDSM data in Fig. 5. This shows that the models trained using CBIS-DDSM images are not overfit, and that the registration performance is quite robust.

Fig. 12 shows exemplar registration results using both networks with NCC for the 990×660 resolution test images. Columns (b) and (c) depict the results from the serial and the skip architectures, respectively. As can be seen in the top panels of columns (b) and (c), both architectures displace the CC lesion indicated by the red circle to the vicinity of the MLO lesion indicated by the blue circle. However, it was generally observed that the deformation vectors in the lesion areas for the DBT data tend to vary less uniformly in terms of the directions and magnitudes, compared to those for the CBIS-DDSM data. This seems to indicate increased uncertainty for the networks that were trained in one data set and used on another. Indeed, the DBT data involves completely different imaging processes than those of the digitized scanned-film images in the DDSM data set.

Tables II and III show the performance on the DBT data by lesion size and breast density, respectively. An image resolution of 990×660 pixels was used with the NCC measure. From Table II, it can be seen that the networks perform better with larger lesions, similar to the CBIS-DDSM case. The average success rate of above 90% was achieved. From Table III, it is seen that the DBT data almost entirely consists of scattered fibroglandular (category 2) and heterogeneously dense tissue (category 3), and the success rates are maintained above 90% in these categories. The BI-RADS assessments for the DBT data were not available.

2) Experiments with Synthetic X-ray Images: Here, our goal is to see if further improvement in performance can be achieved by augmenting the real training images with synthetic ones [55]. The CBIS-DDSM training set was combined with the synthetic X-ray images generated from 3D breast phantoms, designed to represent scattered fibroglandular tissue with single mass-type lesions. The lesion diameters were in the range of around 1.0 cm to 1.2 cm (approximately 23 to 28 pixels at the 330×220 resolution) due to the constraints in the software and computing platform.

We first conducted training and testing using only the synthetic data set. The training set contained $2,250$ synthetic image pairs plus 4-fold augmentations, totaling $11,250$ image pairs. The sizes of the validation set and the test set were 103 and 95 image pairs, respectively. The 330×220 image resolution was used with SAE and DBR. A registration success rate of 82.5% was obtained using the serial architecture. This



(a) Serial architecture. (b) Skip architecture.

Fig. 13: Registration success rates when trained using a mixture of real and synthetic X-ray images.

is comparable to the performance achieved on the real data set for similar lesion sizes; see Fig. 10c. Fig. S8 shows an example synthetic image pair, with the registration results using the serial and the skip architectures.

Then, the CBIS-DDSM training set with 4464 image pairs was combined with the synthetic images in different ratios. We experimented with the real to synthetic image ratios of $1 : x$, with x equal to $0, 0.25, 0.5, 1$, and 2, resulting in 4464, 5580, 6696, 8928, and 13392 training image pairs in total, respectively. The trained networks were tested on 146 real test image pairs as before.

Fig. 13 shows the resulting registration success rates for both architectures. It can be observed that the synthetic images can indeed help boost the performance for the mix ratios up to $1 : 1$, but beyond that the performance begins to degrade as the discrepancy between the real and the synthetic data kicks in. Fig. S9 shows the registration examples. Columns (b) and (c) represent the results from the serial architecture without and with the $1 : 1$ synthetic data mixing, respectively. Similarly, columns (d) and (e) correspond to the skip architecture case. In both cases, the improvement in performance using the mixed data set for training is visible in achieving closer displacements to the target location.

VI. Discussion

There are some limitations to our algorithms and experiments. First, our experiments involved limited data. However, the results from applying our models to a different imaging modality and the improvement observed from using computer-generated data for augmentation indicate promising directions to address data limitation—a general challenge in this area.

The ROI-based criterion for registration success may be deemed less precise for assessing registration accuracy than the conventional pixel-level metrics such as the MSE or Dice. Our ROI-based metric is motivated by the mammography radiologists' practices and the need to capture the degree of usefulness in aiding them to establish the lesion correspondence. A radiologist first finds a lesion in one view. Then, with the given knowledge of the size and other features such as textural composition, spiculations, and various geometrical features (such as the relative distance/orientation toward the nipple), she tries to find the matching lesion in the other view. Thus, a ROI that indicates the approximate location of the lesion can already be very helpful to the radiologist. In fact, in breast image registration, there are significant distortions and occlusions in different views, and assessments based on

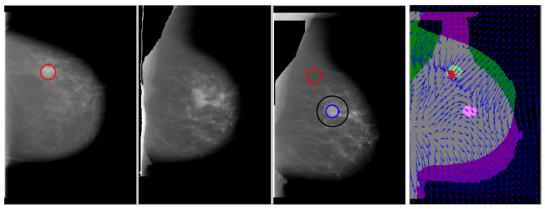


Fig. 14: Demons algorithm results. The warped CC view in panel (b) appears to move the lesion near the desired location in the MLO view in (c) (encircled in blue). Yet, the deformation field in (d) indicates that the bright pixels in (b) were moved from non-lesion locations.

matching the lesions in the pixel level may not adequately capture the level of helpfulness from the clinicians' standpoint and can even be misleading. This can be verified from the example in Fig. 7(b), for which the Dice metric is evaluated to be only 0.13 for our serial network with DBR. Still, the ROI can aid the radiologist to quickly locate the matching lesion. More justification based on a sensitivity analysis is provided in App. C of the Supplementary Material.

Our method tries to displace individual pixels in the moving image and it is not guaranteed that a continuous region is mapped to a continuous region. The TVR constraint can promote this, but as we mentioned in Sec. V-B.1, strong TVR actually comes at odds with the registration performance; see also App. D in the Supplementary Material. Furthermore, the lesions at different views often have very different sizes, as was the case again in Fig. 7(b). Mapping a smaller lesion in one view toward a larger one in another view inevitably results in dispersing the pixels.

Related, it was observed in Sec. V-B.3 that our algorithm yields deformation fields that are not as smooth as those from other algorithms, except for the deformation vectors for the lesion locations. This is because, with DBR, the training of our networks places more emphasis on registering the lesion tissue than the normal tissue areas. Hence, the network inherently learns to detect the lesions and register the corresponding pixels better. In fact, our experiments revealed that smoother deformation fields are sometimes obtained by trying to match the overall shape of one breast scene to the other. That is, it was observed that in some cases algorithms move pixels from non-lesion locations to form regions of high intensity in the target lesion areas. An instance of this is illustrated in Fig. 14 based on Demons algorithm, which incorporates a form of smoothness regularization [32], [57]. Fig. 14(a) shows the CC view with the lesion location. Fig. 14(b) depicts its warped version, which appears to show that the projected CC lesion pixels are near the desired lesion location in the MLO view shown in Fig. 14(c). However, the deformation field shown in Fig. 14(d) indicates that the pixels forming the bright region in Fig. 14(b) are actually moved from areas unrelated to the CC lesion. In fact, the displaced CC lesion, indicated by the red circle in Fig. 14(c), has not moved very far from its original location.

Our experiments were also limited to mass-type lesions. Indeed, there are other common abnormalities such as calcifications, architectural distortions, and asymmetries [26]. However, due to the distinct attributes of the different abnormalities, it is common to design an algorithm for a specific lesion type [45], [62]. We also excluded the multiple lesion case from our study, similar to other recent studies [24]. First, the single lesion occurrences are much more frequent [63]. While the CBIS-DDSM and other curated mammography data sets contain cases with multiple lesions, the ground truth related to matching the lesions is usually not available. Establishing such ground truth in sufficient quantities requires significant expertise and effort. Although beyond the scope of the present work, some indication of the performance of our networks on multi-lesion cases can be viewed for a few examples provided in App. F in the Supplementary Material.

Although we do try to capture the correlations in the tissue areas other than the lesions through the SAE or the NCC measures, the resulting registration in the non-lesion areas is seen rather weak. The limitation can be ascribed to the modest number of training samples and the large distortions inherent in the mammographic views. Additional regularization based on geometrical priors such as the nipple distance or angular positions may be useful [21], [42], [43], but this is left for future research.

VII. Conclusions

An automated registration method for the CC and MLO views of 2D X-ray mammography has been proposed based on CNNs with serial and skip architectures. A custom regularization technique using binary masks of ground truth lesion locations was incorporated to significantly enhance the registration performance. The proposed networks were tested using a real mammography data set (the CBIS-DDSM data set), and the performance was characterized from various aspects. Our method outperformed state-of-the-art CNN-based and non-CNN-based image registration techniques in the dual-view mammography registration task. We also tested the trained networks for registering DBT slices, which verified the robust performance of the networks across related mammographic imaging modalities. Finally, the networks were trained using the real data set mixed with the computer-generated synthetic mammography data set to achieve even better performance. Our method has a potential for aiding the radiologists to quickly establish correspondence for the lesions in different mammographic views. Future research directions include utilizing multiple lesions per image pairs and producing confidence estimates.

Acknowledgments

Insightful discussions with Drs. D. Porter and S. Harvey (MD) on breast imaging and applications, as well as the assistance from Drs. A. Badal and K. Cha on the FDA VICTRE phantom software are gratefully acknowledged.

References

[1] R. L. Siegel *et al.*, "Cancer statistics, 2020," *CA: A Cancer Journal for Clinicians*, vol. 70, no. 1, pp. 7–30, 2020.
[2] O. Ginsburg *et al.*, "Breast cancer early detection: A phased approach to implementation," *Cancer*, vol. 126, pp. 2379–2393, 2020.
[3] H. R. Peppard *et al.*, "Digital breast tomosynthesis in the diagnostic setting: Indications and clinical applications," *Radiographics*, vol. 35, no. 4, pp. 975–990, 2015.

[4] T. Nguyen *et al.*, "Overview of digital breast tomosynthesis: Clinical cases, benefits and disadvantages," *Diag. Interv. Imag.*, vol. 96, no. 9, pp. 843–859, 2015.

[5] G. Eklund, "The art of mammographic positioning," in *Radiological Diagnosis of Breast Diseases*. Springer, 2000, pp. 75–88.

[6] Z. Gandomkar and C. Mello-Thoms, "Visual search in breast imaging," *The British Journal of Radiology*, vol. 92, no. 1102, p. 20190057, 2019.

[7] Y. Guo *et al.*, "Breast image registration techniques: A survey," *Medical and Biological Eng. and Comp.*, vol. 44, no. 1-2, pp. 15–26, 2006.

[8] D. Porter *et al.*, "Multimodality machine learning for breast cancer detection: Synergistic performance with upstream data fusion of digital breast tomosynthesis and ultrasound," presented at the Machine Learning for Health Care Conf., Stanford, CA, Aug. 2018.

[9] S. Perek *et al.*, "Mammography dual view mass correspondence," *arXiv:1807.00637*, 2018.

[10] J. Hipwell *et al.*, "A new validation method for X-ray mammogram registration algorithms using a projection model of breast X-ray compression," *IEEE Trans. Med. Imag.*, vol. 26, no. 9, pp. 1190–1200, 2007.

[11] Y. Kita *et al.*, "Correspondence between different view breast X-rays using curved epipolar lines," *Comput. Vis. Imag. Understanding*, vol. 83, no. 1, pp. 38–56, 2001.

[12] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Imag. Anal.*, vol. 42, pp. 60–88, 2017.

[13] J.-G. Lee *et al.*, "Deep learning in medical imaging: General overview," *Korean J. Radiol.*, vol. 18, pp. 570–584, 2017.

[14] G. Haskins *et al.*, "Deep learning in medical image registration: A survey," *Mach. Vis. Appl.*, vol. 31, no. 8, 2020.

[15] H. Li and Y. Fan, "Non-rigid image registration using self-supervised fully convolutional networks without training data," in *Proc. IEEE Int. Symp. Biomedical Imaging (ISBI'18)*, Washington, D.C., 2018.

[16] G. Balakrishnan *et al.*, "VoxelMorph: A learning framework for deformable medical image registration," *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1788–1800, 2019.

[17] B. de Vos *et al.*, "End-to-end unsupervised deformable image registration with a convolutional neural network," in *Proc. Int. Workshop Deep Learn. Med. Imag. Analysis*, 2017.

[18] H. Sokooti *et al.*, "Nonrigid image registration using multi-scale 3D convolutional neural networks," in *Proc. MICCAI*, 2017.

[19] K. Eppenhof and others., "Deformable image registration using convolutional neural networks," in *Proc. SPIE Med. Imag.*, 2018.

[20] X. Yang *et al.*, "Quicksilver: Fast predictive image registration – a deep learning approach," *NeuroImage*, vol. 158, pp. 378–396, 2017.

[21] W. Walton *et al.*, "Towards CNN-based registration of craniocaudal and mediolateral oblique 2-D X-ray mammographic images," in *Proc. EMBC*, Berlin, Germany, 2019.

[22] J. Li *et al.*, "Mammography registration for unsupervised learning based on CC and MLO views," in *Proc. 3rd Int. Conf. Artificial Intelligence Pattern Recogn.*, 2020, pp. 157–162.

[23] S. Famouri *et al.*, "A deep learning approach for efficient registration of dual view mammography," in *Proc. Workshop on Art. Neural Networks in Pattern Recogn.*, 2020, pp. 162–172.

[24] M. AlGhamdi and M. Abdel-Mottaleb, "DV-DCNN: Dual-view deep convolutional neural network for matching detected masses in mammograms," *Comp. Methods Prog. Biomed.*, Aug. 2021.

[25] J. Long *et al.*, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn. (CVPR)*, 2015.

[26] C. D'Orsi *et al.*, *ACR BI-RADS Atlas: Breast Imaging Reporting and Data System*. American College of Radiology, 2013.

[27] R. Shams *et al.*, "A survey of medical image registration on multicore and the GPU," *IEEE Sig. Process. Mag.*, vol. 27, no. 2, pp. 50–60, 2010.

[28] F. Oliveira *et al.*, "Medical image registration: A review," *Comput. Methods Biomech. Biomed. Eng.*, vol. 17, no. 2, pp. 73–93, 2014.

[29] A. P. Keszei *et al.*, "Survey of non-rigid registration tools in medicine," *Journal of Digital Imaging*, vol. 30, no. 1, pp. 102–116, 2017.

[30] H. R. Boveiri *et al.*, "Medical image registration using deep neural networks: A comprehensive review," *Comput. Electr. Eng.*, vol. 87, pp. 1–24, 2020.

[31] M. F. Beg *et al.*, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms," *Int. J. Comput. Vision*, vol. 61, no. 2, pp. 139–157, 2005.

[32] T. Vercauteren *et al.*, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, no. 1, pp. S61–S72, 2009.

[33] B. Avants *et al.*, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Med. Imag. Anal.*, vol. 12, no. 1, pp. 26–41, 2008.

[34] J. Krebs *et al.*, "Robust non-rigid registration through agent-based action learning," in *Proc. MICCAI*, 2017.

[35] M.-M. Rohé *et al.*, "SVF-Net: Learning deformable image registration using shape matching," in *Proc. MICCAI*, 2017, pp. 266–274.

[36] Y. Hu *et al.*, "Weakly-supervised convolutional neural networks for multimodal image registration," *Med. Imag. Anal.*, vol. 49, pp. 1–13, 2018.

[37] I. Yoo *et al.*, "ssEMnet: Serial-section electron microscopy image registration using a spatial transformer network with learned features," in *Proc. Int. Workshop Deep Learn. Med. Imag. Analysis*, 2017.

[38] N. Ruiter *et al.*, "Model-based registration of X-ray mammograms and MR images of the female breast," *IEEE Trans. Nuclear Science*, vol. 53, no. 1, pp. 204–211, 2006.

[39] C. Curtis *et al.*, "Semiautomated multimodal breast image registration," *Int. J. Biomed. Imag.*, vol. 2012, 2012.

[40] D. Porter *et al.*, "Breast cancer detection/diagnosis with upstream data fusion and machine learning," in *Proc. IWBI*, 2020.

[41] J. Hipwell *et al.*, "A review of biomechanically informed breast image registration," *Physics in Medicine & Biology*, vol. 61, no. 2, 2016.

[42] S. Paquerault *et al.*, "Improvement of computerized mass detection on mammograms: Fusion of two-view information," *Int. J. Med. Phys. Res. Practice*, vol. 29, no. 2, pp. 238–247, 2002.

[43] S. v. Engeland *et al.*, "Finding corresponding regions of interest in mediolateral oblique and craniocaudal mammographic views," *Medical Physics*, vol. 33, no. 9, pp. 3203–3212, 2006.

[44] M. Samulski and N. Karssemeijer, "Matching mammographic regions in mediolateral oblique and cranio caudal views: A probabilistic approach," in *Proc. SPIE Med. Imag.*, 2008.

[45] D. Abdelhafiz *et al.*, "Deep convolutional neural networks for mammography: Advances, challenges and applications," *BMC Bioinformatics*, vol. 20, no. 11, pp. 1–20, 2019.

[46] P. Teare *et al.*, "Malignancy detection on mammography using dual deep convolutional neural networks and genetically discovered false color input enhancement," *J. Dig. Imag.*, vol. 30, no. 4, pp. 499–505, 2017.

[47] K. Geras *et al.*, "High-resolution breast cancer screening with multi-view deep convolutional neural networks," *arXiv:1703.07047*, 2018.

[48] M. Jaderberg *et al.*, "Spatial transformer networks," in *Advances in Neural Information Processing Systems 28 (NIPS'15)*, 2015.

[49] C. Wachinger. MICCAI'10 Tutorial: Intensity-based deformable registration - Similarity measures. [Online]. Available: http://campar.in.tum.de/DefRegTutorial/WebHome

[50] K. Briechle *et al.*, "Template matching using fast normalized cross correlation," in *Proc. SPIE*, vol. 4387, 2001, pp. 95–102.

[51] V. Vishnevskiy *et al.*, "Isotropic total variation regularization of displacements in parametric image registration," *IEEE Trans. Med. Imag.*, vol. 36, no. 2, pp. 385–395, 2016.

[52] ——, "Total variation regularization of displacements in parametric image registration," in *Proc. ABD-MICCAI*, 2014.

[53] O. Ronneberger *et al.*, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015.

[54] R. Lee *et al.*, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific Data*, vol. 4, 2017.

[55] A. Badano *et al.*, "Evaluation of digital breast tomosynthesis as replacement of full-field digital mammography using an in silico imaging trial," *JAMA Network Open*, vol. 1, no. 7, pp. 1–12, 2018.

[56] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv:1412.6980*, 2014.

[57] J.-P. Thirion, "Image matching as a diffusion process: An analogy with Maxwell's demons," *Med. Imag. Anal.*, vol. 2, no. 3, pp. 243–260, 1998.

[58] G. Wengert *et al.*, "Inter-and intra-observer agreement of BI-RADS-based subjective visual estimation of amount of fibroglandular breast tissue with magnetic resonance imaging," *Eur. Radiol.*, vol. 26, no. 11, pp. 3917–3922, 2016.

[59] W. Alomaim *et al.*, "Variability of breast density classification between US and UK radiologists," *J. Med. Imag. Rad. Sci.*, vol. 50, no. 1, pp. 53–61, 2019.

[60] M. M. Eberl *et al.*, "BI-RADS classification for management of abnormal mammograms," *J. Amer. Board Family Med.*, vol. 19, no. 2, pp. 161–164, 2006.

[61] S. J. Magny *et al.*, "Breast Imaging Reporting and Data System (BI-RADS)," *StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing*, 2020.

[62] T. G. Debelee *et al.*, "Survey of deep learning in breast cancer image analysis," *Evolving Systems*, vol. 11, no. 1, pp. 143–163, 2020.

[63] E. O. Cohen *et al.*, "Multiple bilateral circumscribed breast masses detected at imaging: Review of evidence for management recommendations," *Amer. J. Roentgenol.*, vol. 214, no. 2, pp. 276–281, 2020.